



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas

Skirtingų socialinių grupių vartojimo išlaidų struktūrų tyrimas

Baigiamasis magistro studijų projektas

Saulė Bielevičiūtė
Projekto autorė

Prof. dr. Jurgita Bruneckienė
Vadovė

Doc. dr. Mantas Landauskas
Vadovas

Kaunas, 2026



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas

Skirtingų socialinių grupių vartojimo išlaidų struktūrų tyrimas

Baigiamasis magistro studijų projektas
Didžiųjų verslo duomenų analitika (6213AX001)

Saulė Bielevičiūtė

Projekto autorė

Prof. dr. Jurgita Bruneckienė

Vadovė

Doc. dr. Mantas Landauskas

Vadovas

Doc. dr. Ineta Zykienė

Recenzentė

Prof. dr. Robertas Alzbutas

Recenzentas

Kaunas, 2026



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas
Saulė Bielevičiūtė

Skirtingų socialinių grupių vartojimo išlaidų struktūrų tyrimas

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autoriaus ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Saulė Bielevičiūtė

Patvirtinta elektroniniu būdu

Bielevičiūtė, Saulė. Skirtingų socialinių grupių vartojimo išlaidų struktūrų tyrimas. Magistro studijų baigiamasis projektas / vadovė prof. dr. Jurgita Bruneckienė, vadovas doc. dr. Mantas Landauskas; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypties grupė): Taikomoji matematika (Matematikos mokslai).

Reikšminiai žodžiai: namų ūkiai, skirtingos socialinės grupės, vartojimo išlaidos, vartojimo išlaidų struktūra, klasterinė analizė, mišrūs duomenys.

Kaunas, 2026. 70 p.

Santrauka

Temos aktualumas. Namų ūkių vartojimo išlaidos yra svarbi ekonominės elgsenos išraiška, atspindinti ne tik gyventojų ekonominę padėtį, bet ir jų gyvenimo lygį, vartojimo prioritetus bei socialinius skirtumus. Skirtingi namų ūkiai pasižymi nevienodomis socialinėmis, demografinėmis ir ekonominėmis charakteristikomis, todėl jų vartojimo struktūra gali reikšmingai skirtis. Mokslinėje literatūroje vartojimo išlaidos dažniausiai analizuojamos vertinant pavienių veiksnių poveikį vartojimui, tačiau mažiau dėmesio skiriama duomenimis grindžiamam socialinių grupių identifikavimui ir detalesnei vartojimo struktūrų analizei identifikuotų grupių viduje. **Tyrimo objektas** – namų ūkiai ir jų vartojimo išlaidų struktūra. **Tyrimo tikslas** – identifikuoti skirtingas namų ūkių socialines grupes, remiantis namų ūkių socialinėmis, demografinėmis ir ekonominėmis charakteristikomis, bei įvertinti jų vartojimo išlaidų struktūrų skirtumus. **Tyrimo metodai** – mokslinės literatūros analizė ir sintetinimas, duomenų paruošimo ir požymių atrankos metodai, mišraus tipo duomenų klasterizavimo metodai, klasterių kokybės vertinimo metodai, statistiniai ir entropija pagrįsti vartojimo struktūrų palyginimo metodai, išvadų generavimas.

Siekiant įgyvendinti iškeltą tikslą darbas suskirstytas į tris pagrindines dalis: literatūros apžvalgą, metodologinę dalį ir empirinį tyrimą. Literatūros apžvalgoje analizuota socialinių grupių samprata, namų ūkių vartojimo išlaidų ekonominė reikšmė, vartojimo skirtumus lemiantys veiksniai bei mišraus tipo duomenų klasterizavimo metodų taikymo galimybės socialinių grupių identifikavimui. Metodologinėje dalyje aprašyti duomenų paruošimo etapai, požymių atrankos metodai, socialinių grupių identifikavimo metodika bei vartojimo išlaidų struktūrų analizės metodai. Empirinėje tyrimo dalyje atlikta namų ūkių duomenų analizė, vykdytas požymių informatyvumo vertinimas, Gower matricos svorių optimizavimas, taikyti k-prototipų, k-medoidų ir hierarchinės klasterizacijos metodai, o vartojimo struktūrų skirtumams tirti naudoti statistiniai bei entropija grindžiami metodai.

Tyrimas parodė, kad remiantis Lietuvos 2021 m. namų ūkių biudžetų statistinio tyrimo duomenų socialinėmis, demografinėmis ir ekonominėmis charakteristikomis, taikant k-medoidų PAM klasterizavimo algoritmą ir optimizuotą Gower atstumo matricą, galima identifikuoti šešias ekonomiškai interpretuojamas socialines grupes. Pastebėta, kad pašalinus perteklinius kintamuosius ir optimizavus požymių svorius klasterizavimo kokybę – vidutinį silueto koeficientą galima padidinti nuo 0,24 iki 0,62. Vartojimo išlaidų analizė atskleidė, kad socialinės grupės pasižymėjo nevienodu vidiniu vartojimo išlaidų pasiskirstymu. Todėl grupių viduje, nagrinėjant skirtingas išlaidų kategorijas, nustatytas skirtingas vidinio neapibrėžtumo lygis ($NH = 0,069-0,796$). Vis tik daugumoje grupių tarpusavio informacijos rodikliai išliko žemi (daugiausia $NMI < 0,18$), todėl daugeliu atvejų nustatytos priklausomybės atitiko silpnų ryšių lygį ir neatskleidė nuoseklios tarpkategorinių ryšių struktūros.

Bielevičiūtė, Saulė. A Study of Consumption Expenditure Patterns Across Different Social Groups. Master's Final Degree Project / supervisor prof. dr. Jurgita Bruneckienė, supervisor doc. dr. Mantas Landauskas; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics (Mathematical Sciences).

Keywords: households, different social groups, consumption expenditure, consumption expenditure structure, cluster analysis, mixed-type data.

Kaunas, 2026. 70.

Summary

Relevance of the topic. Household consumption expenditure represents an important expression of economic behavior, reflecting not only the economic conditions of households but also their living standards, consumption priorities, and social inequalities. Since households differ in their social, demographic, and economic characteristics, substantial variation in consumption structures may be observed across different population groups. Existing studies mainly focus on the effects of individual determinants on consumption, while less attention is paid to data-driven identification of social groups and detailed analysis of consumption structures within such groups. **Object of the study** – households and their consumption expenditure structure. **Aim of the study** – to identify different social groups of households based on social, demographic, and economic household characteristics and to evaluate differences in their consumption expenditure structures. **Methods** – scientific literature review and synthesis; data preparation and feature selection methods; mixed-type data clustering techniques; cluster quality assessment methods; statistical and entropy-based approaches for comparing consumption structures; conclusion generation.

To achieve the research aim, the study was divided into three main parts: literature review, methodological framework, and empirical analysis. The literature review examined the concept of social groups, the economic significance of household consumption expenditure, determinants of consumption differences, and the applicability of mixed-type data clustering methods for social group identification. The methodological section described data preparation procedures, feature selection techniques, the methodology for social group identification, and methods used for the analysis of consumption expenditure structures. The empirical section involved household data analysis, assessment of feature informativeness, optimization of Gower distance matrix weights, application of k-prototypes, k-medoids, and hierarchical clustering methods, as well as the use of statistical and entropy-based approaches to investigate differences in consumption structures.

The findings demonstrated that six economically interpretable social groups could be identified using social, demographic, and economic characteristics from the 2021 Lithuanian Household Budget Survey data, applying the k-medoids PAM clustering algorithm together with an optimized Gower distance matrix. It was observed that removing redundant variables and optimizing feature weights improved clustering quality, increasing the average silhouette coefficient from 0.24 to 0.62. The analysis of household consumption expenditure revealed that social groups exhibited different patterns of internal expenditure distribution. Consequently, varying levels of internal uncertainty were identified across expenditure categories within groups ($NH = 0.069\text{--}0.796$); however, mutual information values remained low in most groups (mostly $NMI < 0.18$), indicating only weak intercategory relationships.

Turinys

Lentelių sąrašas	7
Paveikslų sąrašas	8
Santrumpų ir terminų sąrašas	9
Įvadas.....	10
1. Literatūros apžvalga	11
1.1. Socialinių grupių samprata	11
1.2. Namų ūkių vartojimo išlaidų samprata ir ekonominė reikšmė.....	13
1.3. Vartojimo išlaidoms įtaką darantys veiksniai.....	15
1.3.1. Bronfenbrennerio ekologinis veiksnių modelis.....	15
1.3.2. Mikrolygmens veiksniai ir jų įtaka vartojimo išlaidoms.....	16
1.4. Socialinių grupių identifikavimo metodų apžvalga.....	20
1.5. Vartojimo išlaidų struktūrų analizės metodų apžvalga	24
1.6. Skirtingų socialinių grupių vartojimo išlaidų struktūrų tyrimų stoka	26
2. Metodologija	27
2.1. Duomenų rinkinys ir jo paruošimas	27
2.2. Namų ūkių duomenų rinkinio paruošimas	29
2.3. Skirtingų socialinių grupių identifikavimo metodika.....	30
2.3.1. Klasterizavimo metodai.....	31
2.3.2. Klasterių skaičiaus nustatymo ir klasterizavimo kokybės vertinimo metodai	35
2.4. Socialinių grupių vartojimo išlaidų struktūrų palyginimo metodai.....	37
2.4.1. Vartojimo išlaidų struktūrų statistinio palyginimo metodai	37
2.4.2. Entropija grindžiamo ryšio metodas.....	39
2.5. Programinė įranga ir naudojamos bibliotekos.....	41
3. Skirtingų socialinių grupių vartojimo išlaidų struktūrų tyrimas	42
3.1. Duomenų šaltinių aprašomoji analizė bei apjungimas	42
3.2. Požymių atranka	44
3.3. Skirtingų socialinių grupių identifikavimas	47
3.3.1. Namų ūkių klasterizavimas po pirminės požymių atrankos.....	47
3.3.2. Gower atstumų matricos optimizacija	50
3.3.3. Namų ūkių klasterizavimas po atgalinio požymių šalinimo.....	51
3.3.4. Klasterinės analizės rezultatų vertinimas ir profiliavimas.....	53
3.4. Vartojimo išlaidų struktūrų tyrimas.....	55
3.4.1. Vartojimo išlaidų struktūrų statistinio palyginimo analizė	55
3.4.2. Entropija grindžiamo ryšio metodo taikymas socialinių grupių vartojimo išlaidų analizei..	57
3.5. Socialinių grupių vartojimo išlaidų tyrimo apibendrinimas, apribojimai ir rekomendacijos...	63
Išvados	65
Literatūros sąrašas	66
Priedai.....	71

Lentelių sąrašas

1 lentelė. Socialinės grupės struktūriniai elementai (sudaryta pagal Burns'ą ir bendraautorius [4])	12
2 lentelė. Namų ūkių vartojimo išlaidų kategorijos pagal pirmojo lygmens COICOP klasifikavimą	28
3 lentelė. <i>Python</i> bibliotekos ir jų paskirtis	41
4 lentelė. Gower svorių optimizavimo rezultatų suvestinė (\bar{S} – vidutinis silueto koeficientas)	51
5 lentelė. Taikytų klasterizavimo metodų rezultatų suvestinė	53
6 lentelė. Identifikuotų skirtingų socialinių grupių profilių suvestinė	54
7 lentelė. Normaliojo skirstinio bendrieji informaciniai ryšiai	59
8 lentelė. Eksponentinio skirstinio bendrieji informaciniai ryšiai	60

Paveikslų sąrašas

1 pav. ES vartojimo išlaidų dinamika, % (sudaryta pagal Eurostat 2024 m. duomenis).....	14
2 pav. Bronfenbrennerio ekologinis modelis (sudaryta pagal Crawfordo analizę [16]).....	15
3 pav. Engelio dėsnio grafinė interpretacija lyginant skirtingų pasaulio šalių maisto išlaidų dalį bendrose vartojimo išlaidose 2023 m. (sudaryta pagal Hannahą [21])	17
4 pav. Klasterizavimo technikų mišraus tipo duomenims sisteminė schema (sudaryta pagal Ghattasą ir kt. [35]).....	22
5 pav. Tyrimo eigos schema.....	42
6 pav. Engelio dėsnio empirinio patikrinimo grafikas	44
7 pav. ϕK koreliacijos koeficientų matrica.....	45
8 pav. K-prototipų algoritmo klasterių skaičiaus nustatymas	48
9 pav. K-medoidų algoritmo klasterių skaičiaus nustatymas.....	49
10 pav. Vidutinio silueto koeficiento reikšmės kitimo dinamika taikant atgalinį požymių šalinimą.....	52
11 pav. Normaliojo skirstinio imčių histogramos ir entropijos rodikliai	58
12 pav. Normaliojo skirstinio porinių ryšių rezultatai	58
13 pav. Eksponentinio skirstinio imčių histogramos ir entropijos rodikliai	59
14 pav. Eksponentinio skirstinio porinių ryšių rezultatai.....	60

Santrumpų ir terminų sąrašas

Santrumpos:

COICOP (angl. *Classification of Individual Consumption by Purpose*) – tarptautinis individualaus vartojimo išlaidų pagal paskirtį klasifikatorius, naudojamas namų ūkių vartojimo išlaidų klasifikavimui [1].

PAM (angl. *Partitioning Around Medoids*) – k-medoidų klasterizavimo algoritmas.

NH (angl. *Normalized Entropy*) – normalizuota entropija.

NMI (angl. *Normalized Mutual Information*) – normalizuota tarpusavio informacija.

Terminai:

Namų ūkis – atskirai gyvenantis vienas asmuo arba grupė viename būste gyvenančių asmenų, kurie dalijasi išlaidas ir bendrai apsirūpina gyventi būtinomis priemonėmis [2].

Namų ūkio galva – asmuo, kuris skiria daugiausia savo pajamų šeimos poreikiams tenkinti. Kai tokio asmens išskirti negalima (pvz., visas namų ūkis ūkininkauja ir pajamų negalima priskirti kuriam nors ūkio nariui), namų ūkio galva laikomas asmuo, kurį nurodo namų ūkis [2].

Klasteris – panašių stebinių grupė.

Gower atstumas – atstumo matas mišrių tipų duomenims.

Įvadas

Namų ūkių vartojimo išlaidos laikomos svarbiu gyventojų ekonominės padėties ir vartojimo elgsenos rodikliu, atspindinčiu gyvenimo sąlygas, prioritetus bei vartojimo ypatumus. Vartojimo išlaidos sudaro vieną didžiausių bendrojo vidaus produkto sudedamųjų dalių, todėl jų pokyčiai gali turėti reikšmingą poveikį ekonomikos augimui, vidaus paklausai bei gyventojų gerovei. Be poveikio makroekonominiams procesams, vartojimo išlaidos taip pat plačiai naudojamos gyvenimo lygiui, ekonominei gerovei ir socialiniams skirtumams vertinti. Dėl šios priežasties vartojimo struktūros analizė laikoma svarbia priemone, leidžiančia geriau suprasti gyventojų ekonominę elgseną bei jos formavimosi dėsningumus.

Mokslinėje literatūroje vartojimo skirtumai dažniausiai analizuojami vertinant atskirų veiksmų, tokių kaip pajamos, išsilavinimas, užimtumas, turtas ar amžius, poveikį namų ūkių vartojimo lygiui bei išlaidų pasiskirstymui tarp skirtingų vartojimo kategorijų. Tačiau vartojimo elgsena formuojasi veikiamą daugelio tarpusavyje susijusių socialinių, demografinių ir ekonominių veiksmų, todėl pavienių charakteristikų analizė ne visada leidžia visapusiškai paaiškinti egzistuojančius vartojimo skirtumus. Kadangi namų ūkiai dažniausiai pasižymi ne viena, o keliomis kartu pasireiškiančiomis charakteristikomis, atsiranda poreikis vertinti jų derinius, leidžiančius identifikuoti tarpusavyje panašias socialines grupes bei atskleisti tarp jų egzistuojančius vartojimo skirtumus.

Skirtingų socialinių grupių identifikavimas bei jų vartojimo išlaidų struktūrų analizė yra svarbūs ne tik teoriniu, bet ir praktiniu požiūriu. Vartojimo struktūros skirtumų analizė gali prisidėti prie gilesnio gyventojų elgsenos modelių bei socialinės diferenciacijos procesų supratimo. Tokia informacija gali būti naudinga formuojant socialinės politikos priemones, vertinant gyventojų pažeidžiamumą ekonominių sukrėtimų metu ir planuojant tikslines paramos priemones skirtingoms gyventojų grupėms.

Tyrimo tikslas – identifikuoti skirtingas namų ūkių socialines grupes, remiantis namų ūkių socialinėmis, demografinėmis ir ekonominėmis charakteristikomis, bei įvertinti jų vartojimo išlaidų struktūrų skirtumus.

Tyrimo uždaviniai:

1. Išanalizuoti namų ūkių vartojimo išlaidų ekonominę reikšmę, vartojimo skirtumus lemiančius veiksnius bei socialinių grupių formavimosi teorinius aspektus.
2. Apžvelgti mišraus tipo duomenų klasterizavimo bei vartojimo išlaidų struktūrų statistinės analizės metodų taikymo galimybes socialinių grupių identifikavimui ir jų skirtumų vertinimui.
3. Parengti socialinių grupių identifikavimo ir vartojimo išlaidų struktūrų analizės metodologiją.
4. Empiriškai identifikuoti namų ūkių socialines grupes ir įvertinti jų vartojimo išlaidų struktūrų skirtumus, taikant klasterinės analizės, statistinius ir entropija grindžiamus metodus.
5. Apibendrinti tyrimo rezultatus ir įvertinti jų praktinį pritaikomumą socialinės politikos formavimo bei vartojimo elgsenos tyrimų kontekste.

Tyrimo metodai – mokslinės literatūros analizė ir sintetinimas, duomenų paruošimo ir požymių atrankos metodai, mišraus tipo duomenų klasterizavimo metodai, klasterių kokybės vertinimo metodai, statistiniai ir entropija pagrįsti vartojimo struktūrų palyginimo metodai, išvadų generavimas.

1. Literatūros apžvalga

Šiame skyriuje apžvelgiama socialinių grupių samprata, analizuojama namų ūkių vartojimo išlaidų ekonominė reikšmė bei pagrindiniai veiksniai, lemiantys vartojimo skirtumus tarp skirtingų socialinių grupių. Literatūros kontekste aptariami duomenų analizės metodai, taikomi socialinių grupių identifikavimui bei vartojimo išlaidų struktūros skirtumų vertinimui.

1.1. Socialinių grupių samprata

Socialinių grupių samprata socialiniuose moksluose pirmiausia siejama su gyventojų bendrais požymiais, socialiniais ryšiais ir bendru priklausymo jausmu. Visuotinėje lietuvių enciklopedijoje socialinė grupė apibrėžiama kaip žmonių visuma, kurią gali vienyti bendras buvimas erdvėje ir laike, veikla, ekonominiai, demografiniai, psichologiniai bei kiti požymiai [3]. Šis apibrėžimas leidžia socialines grupes suprasti kaip daugiamates socialines struktūras, kurių formavimuisi reikšmingi struktūriniai, elgsenos ir kultūriniai veiksniai. Enciklopedijoje pažymima, kad pagal tam tikrus požymius išskiriamos žmonių visumos gali būti laikomos socialinėmis kategorijomis, todėl grupių identifikavimas visada remiasi tam tikru klasifikavimo principu. Be to, socialinė grupė neegzistuoja be socialinių ryšių tarp jos narių, nes grupės nariai yra susiję tiesioginiais arba netiesioginiais ryšiais, o bendrumą palaiko formalios ar neformalios narystės formos, bendri interesai bei tarpusavio priklausomybė. Socialinių grupių tyrimuose taip pat akcentuojama socialinės sąveikos reikšmė, kadangi sociologai, psichologai ir kitų socialinių mokslų atstovai grupes analizuoja kolektyvinės sąmonės, bendrumo, integracijos, sutelktumo bei grupės dinamikos aspektais [3].

Šiuolaikinėje sociologinėje literatūroje socialinė grupė apibrėžiama ne tik kaip asmenų visuma, bet ir kaip organizuotas socialinis darinys, turintis aiškią vidinę struktūrą, bendras normas ir funkcinis veikimo principus. Burns'o ir bendraautorijų teigimu, grupė apima asmenų arba socialinių veikėjų visumą, kuriai būdingas bendras grupinis identitetas, bendros taisyklės bei aiškiai apibrėžtas narystės ir įsipareigojimo grupei pagrindas [4]. Autoriai taip pat pabrėžia, kad grupei būdinga savita vidinė tvarka, specifiniai gebėjimai ir funkcinės galios, o jos veikimas grindžiamas bendromis taisyklėmis, prieiga prie išteklių bei narių gebėjimu šiuos principus taikyti praktikoje. Toks požiūris leidžia socialinę grupę suvokti kaip organizacinę struktūrą, kuriai būdingas tam tikras darbo pasidalijimas, bendri tikslai ir koordinuota veikla.

Burns'o ir bendraautorijų apibrėžti socialinės grupės struktūriniai elementai (žr. 1 lentelę) rodo, kad socialinės grupės struktūrą formuoja ne vien formalus priklausymas, bet ir tarpusavyje susijusių organizacinių elementų visuma [4]. Grupės identitetas leidžia jos nariams suvokti bendrą priklausymą ir išskirti savo grupę kitų socialinių vienetų kontekste. Narystės taisyklės apibrėžia įsitraukimo ir priklausymo grupei sąlygas, o bendros vertybės ir įsitikinimai formuoja priimtino elgesio normas bei bendrą socialinės tikrovės suvokimą. Ne mažiau svarbios yra socialinių santykių organizavimo taisyklės, nustatančios narių vaidmenis, atsakomybės ribas ir tarpusavio sąveikos formas. Taip pat pabrėžiama, kad socialinės grupės išsiskiria savo veiklos organizavimu vietas ir laiko požiūriu, kadangi grupės gali būti atpažįstamos pagal nustatytas susitikimų vietas, veiklos laiką ar periodiškumą. Kaip pavyzdį autoriai pateikia religines bendruomenes, kuriose pamaldų laikas yra institucionalizuotas – musulmonų bendruomenėse pagrindine pamaldų diena laikomas penktadienis, žydų bendruomenėse – šeštadienis, o krikščionių bendruomenėse – sekmadienis [4].

1 lentelė. Socialinės grupės struktūriniai elementai (sudaryta pagal Burns'ą ir bendraautorius [4])

Nr.	Elementas	Elemento savybė
I	Grupės identiteto taisyklės	Nusako, kas sudaro grupės identitetą – simbolika ir kiti ženklai, leidžiantys nariams atpažinti ir suvokti savo priklausymą konkrečiai socialinei grupei.
II	Narystės ir įsitraukimo taisyklės	Apibrėžia, kas gali priklausyti grupei, kokiomis sąlygomis tampama grupės nariu, kaip vyksta naujų narių priėmimas ir įsitraukimas.
III	Bendros vertybinės orientacijos	Nusako, kokias vertybes grupė laiko priimtinomis, svarbiomis ar pageidaujamos, kas grupėje laikoma teigiamu arba neigiamu elgesiu.
IV	Bendri įsitikinimai ir pažintiniai modeliai	Apima bendrą grupės narių supratimą apie aplinką, žinojimą, įsitikinimus ir interpretavimo modelius.
V	Socialinių santykių ir struktūros taisyklės	Apibrėžia tarpusavio santykius, vaidmenų paskirstymą, hierarchiją, atsakomybės ribas ir sąveikos modelius grupėje.
VI	Veiklos ir procedūrų taisyklės	Nusako, kokios veiklos vykdomos grupėje, kaip priimami sprendimai, koordinuojama veikla ir organizuojami procesai.
VII	Aplinkos taisyklės	Apibrėžia grupės santykį su išorine aplinka, prisitaikymo prie aplinkos principus ir veiklos kontekstą.
VIII	Pokyčių ir transformacijos taisyklės	Nusako, kaip grupė prisitaiko prie pokyčių, atnaujina savo struktūrą, normas ar veikimo principus.
IX	Išteklių ir technologijų valdymo taisyklės	Apima materialinių, technologinių ir kitų išteklių naudojimą, paskirstymą ir kontrolę grupės veikloje.
X	Laiko ir vietos organizavimo taisyklės	Nusako, kur ir kada vyksta grupės veikla, susitikimai, bendravimas ir kolektyvinė sąveika.

Nors socialinės grupės gali būti apibrėžiamos pagal bendrus identiteto, narystės ir organizacinius principus, jų formavimasis visuomenėje taip pat siejamas su individų ar namų ūkių socialine padėtimi. Socialinė padėtis gali būti suprantama kaip individo ar namų ūkio vieta visuomenėje, kurią lemia socialinių, ekonominių ir demografinių veiksnių visuma, apimanti pajamas, išsilavinimą, užimtumą, profesiją ar turimus išteklius [5]. Skirtinga socialinė padėtis lemia nevienodas gyvenimo sąlygas, galimybes naudotis ištekliais bei prisitaikyti prie socialinių ir ekonominių pokyčių. Dėl šių skirtumų gali formuotis skirtingos socialinės grupės, atspindinčios panašią jų narių socialinę ir ekonominę situaciją. Pavyzdžiui, ribotais ekonominiais ir socialiniais ištekliais pasižyminčios gyventojų grupės dažniau siejamos su socialiai pažeidžiamomis grupėmis, o didesniais ekonominiais, profesiniais ar socialiniais ištekliais pasižyminčios grupės gali būti siejamos su aukštesnio socialinio statuso grupėmis [5].

Struktūrinis socialinės grupės aiškinimas papildomas socialinės tapatybės teorija, kuri grupinį priklausymą aiškina ne tik kaip formalų priskyrimą tam tikrai socialinei kategorijai, bet ir kaip psichologinį tapatinimąsi su konkrečia socialine grupe. Socialinės psichologijos literatūroje pabrėžiama, kad priklausymas grupei turi reikšmingą poveikį vertybių formavimuisi, elgsenos ir sprendimų priėmimo principams, kurie gali pasireikšti ne tik individualiu, bet ir kolektyviniu lygmeniu. Tyrimai rodo, kad identifikavimosi su savąja grupe stiprumas yra susijęs su tuo, kiek grupinė narystė tampa psichologiškai ir socialiai reikšminga [6]. Namų ūkių kontekste tai reiškia, kad bendros socialinės normos, vertybinės orientacijos ir grupinė priklausomybė gali formuoti bendrus vartojimo, taupymo ir kitų ekonominių sprendimų modelius, pasireiškiančius namų ūkio lygmeniu. Charness'as ir Chenas, apžvelgdami Tajfelio ir Turnerio socialinės tapatybės teoriją, nurodo, kad grupinės tapatybės formavimąsi sudaro trys pagrindiniai komponentai – kategorizavimas, identifikacija ir palyginimas [7]. Kategorizavimas nusako procesą, kurio metu individai save ir kitus

priskiria tam tikroms socialinėms kategorijoms, identifikacija atspindi psichologinį susiejimą su konkrečia grupe, o palyginimas apima savo grupės vertinimą kitų grupių kontekste. Tajfelio ir kt. socialinės tapatybės teorija rodo, kad grupinė priklausomybė siejama su bendrumo jausmo stiprėjimu bei bendrų elgesio normų formavimusi ir skatina perimti grupei būdingus sprendimų priėmimo principus. Tačiau socialinė tapatybė nėra statiška – jos reikšmė priklauso nuo situacijos, socialinio konteksto ir aplinkos veiksnių. Literatūroje pabrėžiama, kad socialinė tapatybė yra daugiasluoksnė, todėl skirtingose situacijose gali išryškėti skirtingi identiteto aspektai, o pati grupinė priklausomybė išlieka jautri kontekstiniams pokyčiams [7]. Dėl šios priežasties namų ūkio elgsena gali atspindėti skirtingų socialinių tapatybių, normų ir vaidmenų sąveiką, kuri kinta priklausomai nuo socialinės, ekonominės ir kultūrinės aplinkos.

1.2. Namų ūkių vartojimo išlaidų samprata ir ekonominė reikšmė

Namų ūkių vartojimo išlaidos mokslinėje literatūroje apibrėžiamos kaip piniginės išlaidos, skirtos prekėms ir paslaugoms įsigyti, siekiant patenkinti einamuosius bei ilgalaikius vartojimo tikslus [9]. Galutinės namų ūkių vartojimo išlaidos apima įvairių prekių ir paslaugų įsigijimą, įskaitant maisto ir ne maisto produktus bei įvairias paslaugas. Nors vartojimo išlaidos ir vartojimas nėra tapačios sąvokos, jos itin glaudžiai siejasi. Vartojimo išlaidos atspindi pinigines išlaidas prekėms ir paslaugoms įsigyti, o vartojimas apibūdina faktiškai suvartotų prekių ir paslaugų kiekį bei jų teikiamą naudą per tam tikrą laikotarpį. Dėl glaudaus šių sąvokų ryšio vartojimo išlaidos ekonominiuose tyrimuose dažnai naudojamos kaip netiesioginis realaus vartojimo matas. Browningas ir bendraautoriai pabrėžia, kad namų ūkių vartojimo išlaidų duomenys yra vienas svarbiausių empirinių šaltinių analizuojant vartojimo elgseną, taupymo sprendimus, pajamų nelygybę bei ekonominių sukrėtimų poveikį namų ūkiams [9]. Atkreipiamas dėmesys, jog namų ūkių vartojimo išlaidos apima tik vartojimo tikslams skirtas išlaidas ir neįtraukia investicinio pobūdžio išlaidų, susijusių su turto kaupimu ar kapitalo formavimu. Bendras vartojimo išlaidų lygis paprastai apskaičiuojamas agreguojant atskiras išlaidų kategorijas pagal deklaruotų išlaidų struktūrą.

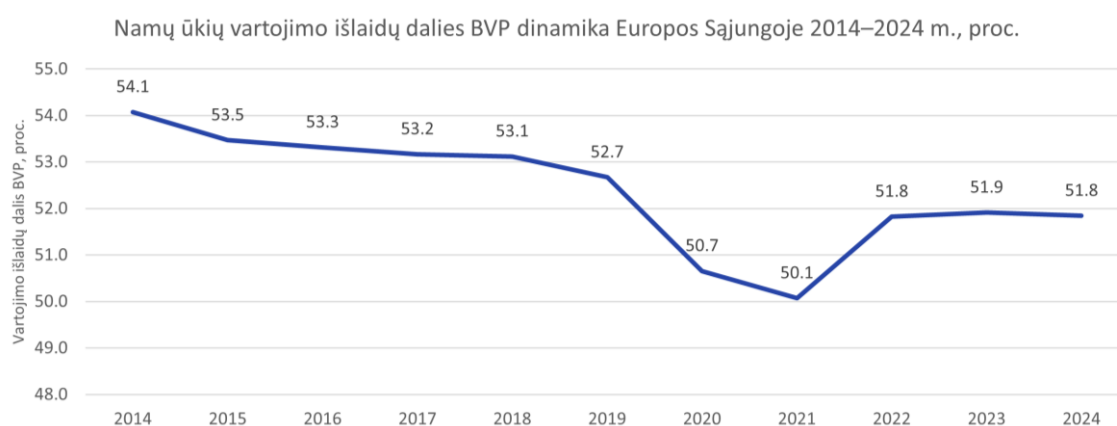
Siekiant užtikrinti vartojimo išlaidų palyginamumą tarptautiniu mastu, statistinėje praktikoje taikomas individualaus vartojimo pagal paskirtį klasifikatorius (COICOP) [1]. Šis klasifikatorius naudojamas nacionalinių sąskaitų sistemoje, namų ūkių biudžetų tyrimuose, vartotojų kainų indeksų skaičiavimuose bei tarptautiniuose makroekonominiuose palyginimuose. COICOP taikymas leidžia vartojimo išlaidas suskirstyti į homogeniškas prekių ir paslaugų grupes bei objektyviai vertinti vartojimo struktūros skirtumus tarp gyventojų grupių bei šalių.

Makroekonominiu požiūriu namų ūkių vartojimo išlaidos yra viena iš pagrindinių bendrojo vidaus produkto (toliau – BVP) sudedamųjų dalių. BVP apskaičiuojamas kaip namų ūkių vartojimo išlaidų, investicijų, valstybės išlaidų bei grynojo eksporto suma [10]:

$$BVP = C + I + G + (EX - IM); \quad (1)$$

čia C – namų ūkių vartojimo išlaidos; I – investicijos; G – valstybės išlaidos; EX – prekių ir paslaugų eksportas; IM – prekių ir paslaugų importas.

Namų ūkių vartojimo reikšmę makroekonomikoje patvirtina ir naujausi Europos Sąjungos (toliau – ES) statistiniai duomenys. Eurostat 2024 m. paskelbti duomenys¹ rodo, kad 2024 m. galutinės namų ūkių vartojimo išlaidos sudarė 51,8 % ES bendrojo vidaus produkto, t. y. buvo tik 0,1 procentinio punkto mažesnės nei 2023 m. (51,9 %), tačiau vis dar išliko didžiausia galutinės paklausos sudedamąją dalimi. Taip pat nustatyta, kad lyginant su 2014 m., kai šis rodiklis siekė 54,1 % BVP, namų ūkių vartojimo dalis 2024 m. sumažėjo 2,3 procentinio punkto. 2014–2021 m. laikotarpiu stebimas nuoseklus vartojimo dalies mažėjimas nuo 54,1 % iki 50,1 %, o nuo 2022 m. fiksuojamas dalinis atsigavimas – rodiklis stabilizavosi 51,8–51,9 % intervale. 1 paveiksle pateikti Eurostat duomenys rodo, kad net ir infliacijos, energetinės krizės bei geopolitinio neapibrėžtumo laikotarpiu namų ūkių vartojimas išlieka viena didžiausių ir mažiausiai svyruojančių Europos ekonomikos sudedamųjų dalių.



1 pav. ES vartojimo išlaidų dinamika, % (sudaryta pagal Eurostat 2024 m. duomenis¹)

Kadangi ekonomikoje vartojimas laikomas vienu svarbiausių agreguotos paklausos komponentų, glaudžiai susijusiu su bendruoju vidaus produktu, taupymu ir naudingumo teorija, namų ūkių vartojimo išlaidos sudaro reikšmingą bendros paklausos dalį, kurios pokyčiai gali reikšmingai paveikti gamybos apimtį, investicijų lygį, užimtumą ir ekonomikos ciklų dinamiką. Vartojimo reikšmę ekonominiams procesams patvirtina ir empiriniai tyrimai. Khurana bei Beero savo tyrime nustatė, kad privataus galutinio vartojimo išlaidos (angl. *Private Final Consumption Expenditure*, PFCE), atspindinčios namų ūkių galutinį vartojimą nacionalinių sąskaitų sistemoje, yra statistiškai reikšmingai susijusios su bendruoju vidaus produktu ir kitais makroekonominiais rodikliais [11]. Taip pat tyrimo rezultatai atskleidė, kad 1980–2020 m. laikotarpiu PFCE Indijoje padidėjo beveik keturis kartus, patvirtindami didėjančią vartojimo reikšmę ekonominiuose procesuose.

Be makroekonominės reikšmės, namų ūkių vartojimo išlaidos plačiai naudojamos gyventojų gerovės vertinimui. Nors gyvenimo lygis dažnai vertinamas pajamų rodikliais, dalis tyrėjų vartojimo išlaidas laiko tikslesniu gerovės rodikliu [12]. Jie teigia, kad vartojimo struktūros analizė gali padėti įvertinti ne tik namų ūkių finansines galimybes, bet ir jų prieigą prie sveikatos priežiūros, švietimo, kultūros ir laisvalaikio paslaugų, todėl padeda geriau suprasti gyvenimo kokybės formavimosi procesus. Vartojimo išlaidų rodikliai taip pat reikšmingi vertinant pajamų nelygybę, skurdo lygį ir socialinius skirtumus tarp gyventojų grupių. Ypač besivystančiose šalyse vartojimo išlaidos vienam gyventojui dažnai naudojamos kaip materialinės gerovės rodiklis [13]. Be to, vartojimo išlaidų duomenys plačiai

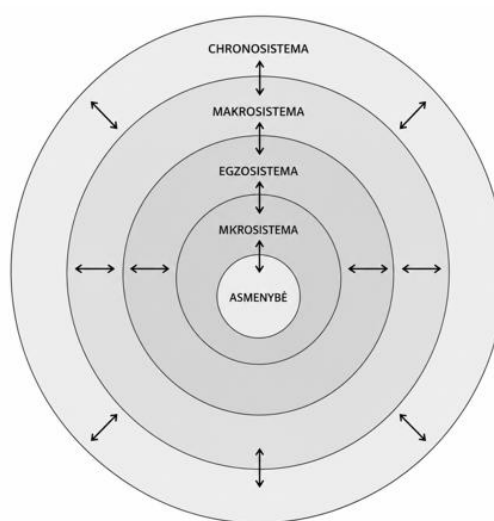
¹ https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Household_consumption_by_purpose#Composition_in_2024_compared_with_2023

taikomi analizuojant ekonomikos augimo, pajamų paskirstymo ir skurdo mažinimo procesus bei vertinant socialinės apsaugos, mokesčių ir pajamų perskirstymo politikos efektyvumą.

1.3. Vartojimo išlaidoms įtaką darantys veiksniai

1.3.1. Bronfenbrennerio ekologinis veiksnių modelis

Vartojimo išlaidų struktūros skirtumai tarp skirtingų socialinių grupių formuojasi veikiant daugialypiems tarpusavyje susijusiems veiksniams, apimantiems individualias, socialines, ekonomines, institucines bei laiko dimensijas. Empiriniai tyrimai rodo, kad vartojimo nelygybė, kaip ir pajamų nelygybė, gali atspindėti realius gyvenimo lygio skirtumus tarp individų, namų ūkių, socialinių grupių ar regionų [14]. Siekiant sistemiskai paaiškinti šių skirtumų formavimąsi, analizei pasitelkiamas Bronfenbrennerio ekologinis veiksnių modelis, leidžiantis vartojimo elgseną analizuoti kaip daugialypių tarpusavyje susijusių aplinkos sistemų sąveikos rezultatą (žr. 2 pav.).



2 pav. Bronfenbrennerio ekologinis veiksnių modelis (sudaryta pagal Crawfordo analizę [15])

Bronfenbrennerio ekologinis modelis grindžiamas prielaida, kad individo elgsena formuojasi nuolatinėje sąveikoje su jį supančia aplinka. Pagal šį požiūrį individą veikia kelios tarpusavyje susijusios aplinkos sistemos – nuo artimiausios aplinkos, tokios kaip šeima ar socialiniai ryšiai, iki platesnių visuomeninių, kultūrinių ir ekonominių veiksnių. Crawfordas, analizuodamas šią teorinę sistemą, pažymi, kad aplinkos poveikis pasireiškia skirtinguose tarpusavyje susijusiuose lygmenyse, kurie organizuojami hierarchiškai viena kitą apimančiose struktūrose [15]. Vartojimo kontekste toks požiūris vartojimo elgseną aiškina kaip įvairių aplinkos sistemų sąveikos rezultatą, kai skirtingi socialiniai, ekonominiai ir kultūriniai veiksniai formuoja individų elgseną, vertybines nuostatas ir vartojimo įpročius. Šiuo požiūriu vartojimo sprendimus ir vartojimo išlaidų struktūrą gali veikti mikro-, mezo-, egzo-, makro- ir chrono-sistemos.

Artimiausią poveikį žmogaus elgsenai daro mikrosistema, apibrėžiama kaip veiklų, vaidmenų ir tarpasmeninių santykių visuma, patiriama kasdienėje aplinkoje. Vartojimo išlaidų analizės kontekste šis lygmuo apima namų ūkių demografines ir socialines charakteristikas, tokias kaip šeiminių padėtis, namų ūkio sudėtis, vaikų skaičius, amžius, užimtumas, išsilavinimas bei gyvenimo būdo ypatumai. Kadangi Bronfenbrennerio teorijoje šeima laikoma viena svarbiausių mikrosistemos sudedamųjų dalių, galima daryti prielaidą, kad namų ūkio struktūra yra reikšminga vartojimo elgsenos

formavimosi aplinka [15]. Empiriniai tyrimai taip pat rodo, kad namų ūkių sudėtis ir jos narių amžius reikšmingai siejasi su vartojimo išlaidų pasiskirstymu tarp skirtingų kategorijų [8].

Mezosistema apima skirtingų artimos aplinkos sistemų tarpusavio sąveiką, pavyzdžiui, ryšius tarp šeimos, darbo aplinkos, ugdymo institucijų ar socialinių tinklų [15]. Vartojimo išlaidų analizės kontekste šis lygmuo apibūdina, kaip skirtingų socialinių vaidmenų ir aplinkų sąveika gali formuoti kasdienes namų ūkių poreikius bei vartojimo prioritetus. Literatūroje pabrėžiama, kad žmonės disponuoja ribotais laiko, energijos ir kitais ištekliais, todėl didesnis išteklių paskirstymas vienoje gyvenimo srityje gali reikšmingai paveikti kitas sritis [16].

Tuo tarpu egzosistema apima aplinkas, kuriose pats asmuo ar grupė tiesiogiai nedalyvauja, tačiau jų poveikis pasireiškia netiesiogiai. Šiame lygmenyje vartojimo elgseną gali veikti darbo rinkos sąlygos, darbdavių politika, aplinkos infrastruktūros išplėtojimas, vietos valdžios sprendimai bei socialinės paramos sistemos, galinčios paveikti disponuojamus išteklius ir vartojimo galimybes [15]. Prie šių netiesioginių poveikį darančių veiksnių gali būti priskiriamos ir finansų rinkų sąlygos. Jappelli ir Pagano parodė, kad namų ūkių likvidumo apribojimai gali būti susiję su vartojimo ir taupymo elgsena [17]. Ribotos skolinimosi galimybės gali skatinti didesnę taupymą ir kartu riboti vartojimo galimybes, palyginti su situacijomis, kai kredito rinkos yra lengviau prieinamos.

Makrosistema apima platesnius kultūrinius, socialinius ir ekonominius veiksnius, kurie daro įtaką kitų sistemų veikimui ir jų tarpusavio sąveikai. Vartojimo išlaidų analizėje šiame lygmenyje reikšmingi ekonominės politikos sprendimai, mokesčių sistema, infliacija, kultūrinės vartojimo normos, socialinė stratifikacija, švietimo sistema bei bendras ekonominio išsivystymo lygis [15].

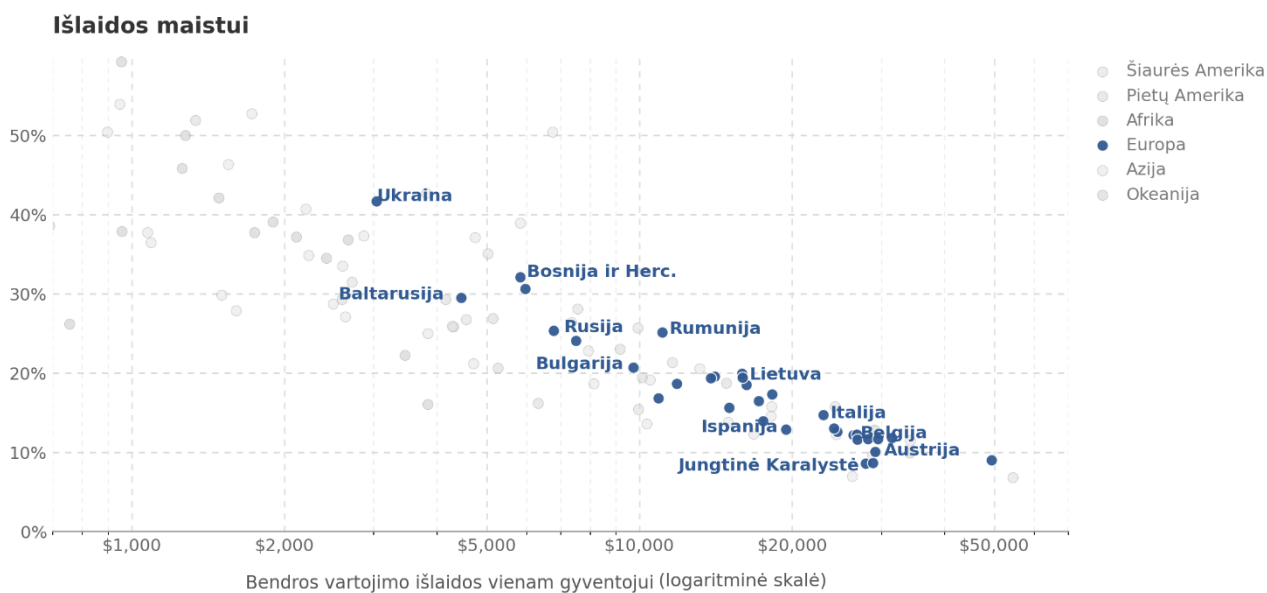
Galiausiai chronosistema į modelį įtraukia laiko dimensiją ir leidžia vertinti, kaip vartojimo elgsena kinta veikiama ilgalaikių socialinių ir ekonominių pokyčių. Šiame lygmenyje vartojimo struktūra gali kisti dėl ekonominių ciklų, technologinių pokyčių, infliacijos šokų, pandemijos, geopolitinių konfliktų bei kitų išorinių sukrėtimų poveikio [15]. Pavyzdžiui, COVID-19 pandemija ir 2022 m. prasidėjusios Rusijos karinės invazijos į Ukrainą sukelti ekonominiai ir energetiniai pokyčiai parodė, kad tokie socialiniai ir ekonominiai sukrėtimai lemia vartojimo prioritetų, išlaidų paskirstymo bei vartojimo elgsenos pokyčius [18].

Bronfenbrennerio teorijoje taip pat pabrėžiama socioekonominės aplinkos reikšmė. Crawfordas pažymi, kad Bronfenbrenneris į savo modelį įtraukė šeimos socioekonominį statusą kaip vieną iš pagrindinių aplinkos veiksnių, o empiriniai tyrimai parodė, kad neigiami artimos aplinkos procesai gali stipriau paveikti žemesnio socioekonominio statuso grupes, tuo tarpu teigiami aplinkos procesai gali būti labiau susiję su palankesniais rezultatais aukštesnio statuso grupėse [15]. Toks požiūris rodo, kad vartojimo išlaidų struktūra priklauso ne tik nuo disponuojamų pajamų, bet ir nuo platesnės socialinės, ekonominės, kultūrinės bei institucinės aplinkos poveikio, kuris formuoja vartojimo elgseną, įpročius ir vartojimo prioritetus.

1.3.2. Mikrolygmens veiksniai ir jų įtaka vartojimo išlaidoms

Nors teorinis Bronfenbrennerio modelis apima mikro-, mezo-, egzo-, makro- ir chrono-sistemų poveikį, šiame tyrime pagrindinis dėmesys skiriamas mikrolygmens veiksniams, nusakomiems namų ūkių ar jų narių socialinėmis, demografinėmis ir ekonominėmis charakteristikomis.

Tarp mikrolygmens veiksnių, paaiškinančių vartojimo išlaidų skirtumus, ypač išsiskiria pajamų lygis, kuris laikomas vienu svarbiausių veiksnių, susijusių su namų ūkių vartojimo galimybėmis. Tyrimai rodo, kad mažesnes pajamas gaunantys namų ūkiai santykinai didesnę savo biudžeto dalį skiria būtiniausiems poreikiams, tokiems kaip maistas, būstas ir komunalinės paslaugos, o didesnes pajamas gaunantių namų ūkių vartojimo struktūroje didesnę dalį sudaro diskrecinio vartojimo išlaidos, susijusios su transportu, laisvalaikiu, švietimu ar kultūra [19, 20]. Toks vartojimo modelis atitinka Engelio dėsnį ir leidžia vartojimo struktūrą naudoti kaip netiesioginį gyvenimo lygio bei ekonominės gerovės rodiklį. Engelio dėsnis teigia, kad didėjant namų ūkio pajamoms absoliučios išlaidos maistui gali augti, tačiau jų dalis bendrame vartojimo biudžete mažėja. Kadangi praktiniuose tyrimuose pajamų duomenys ne visada yra tiksliai prieinami arba gali būti nepakankamai reprezentatyvūs realiam gyvenimo lygiui atspindėti, alternatyviai naudojamos bendros namų ūkio vartojimo išlaidos, kurios laikomos faktinio disponuojamų išteklių panaudojimo rodikliu. Tokiu atveju analizuojamas ryšys tarp bendrų vartojimo išlaidų ir atskirų vartojimo kategorijų dalies bendrame biudžete. Tai leidžia įvertinti namų ūkio ekonominę padėtį net ir neturint tiesioginių pajamų duomenų. Engelio kreivių retrospektyvinė analizė rodo, kad šis dėsningumas išlieka viena stabiliausių empirinių vartojimo elgsenos dėsningumų. Tyrime, atliktame remiantis Belgijos darbininkų šeimų biudžetais, nustatyta, kad didėjant vidutinėms šeimos pajamoms nuo 565 iki 1198 frankų, maisto išlaidų dalis bendrame biudžete mažėjo nuo 70,89 % iki 62,42 % [19]. Šiuolaikiniai tarptautiniai duomenys leidžia stebėti analogišką priklausomybę pasauliniu mastu. Hannaho tyrime, analizuojant ryšį tarp bendrų vartojimo išlaidų vienam gyventojui ir maisto išlaidų dalies vartojimo struktūroje (žr. 3 pav.) matoma aiški neigiama priklausomybė: mažesnių vartojimo išlaidų lygio šalyse maisto išlaidos sudaro daugiau kaip 30–40 % visų vartojimo išlaidų, tuo tarpu aukštesnio vartojimo lygio Europos valstybėse ši dalis sumažėja iki maždaug 8–15 % [20]. Pavyzdžiui, Ukrainoje 2023 m. maisto išlaidos sudarė apie 42 % bendrų vartojimo išlaidų, Rumunijoje – apie 24 %, Lietuvoje – apie 19–20 %, o Belgijoje ir Austrijoje – tik apie 9–12 %. Šie skirtumai patvirtina, kad mažėjanti maisto išlaidų dalis bendrame vartojimo biudžete gali būti laikoma vienu iš didesnio disponuojamų išteklių lygio bei aukštesnės materialinės gerovės indikatorių.



3 pav. Engelio dėsnio grafinė interpretacija lyginant skirtingų pasaulio šalių maisto išlaidų dalį bendrose vartojimo išlaidose 2023 m. (sudaryta pagal Hannahą [20])

Greta disponuojamų finansinių išteklių svarbų vaidmenį vartojimo sprendimuose atlieka ir namų ūkių gebėjimas tuos išteklius valdyti. Finansinis raštingumas literatūroje siejamas su efektyvesniu finansinių išteklių paskirstymu gyvenimo ciklo laikotarpiu bei didesniu gebėjimu priimti racionalius finansinius sprendimus [21, 22]. Dinkovos ir kt. atliktame tyrime, paremtame reprezentatyviais Nyderlandų namų ūkių tyrimo paneliniais duomenimis, apimančiais 2009–2017 m. laikotarpį, nustatytas teigiamas ryšys tarp finansinio raštingumo ir vartojimo lygio [21]. Empiriniai rezultatai parodė, kad didesnis finansinis raštingumas yra susijęs su didesniu nediskrecinio vartojimo mastu, ypač didesnėmis maisto išlaidomis. Tame pačiame tyrime nustatyti ir reikšmingi lyčių skirtumai finansinio raštingumo srityje – moterų teisingų atsakymų dalis visuose finansinio raštingumo klausimuose buvo mažesnė nei vyrų, o atsakymų „nežinau“ dalis buvo beveik dvigubai didesnė. Tyrimo rezultatai taip pat parodė, kad porų namų ūkiuose vyro finansinis raštingumas bei subjektyvus finansinių žinių vertinimas buvo teigiamai susiję su nediskrecinio vartojimo mastu, tuo tarpu analogiško statistiškai reikšmingo ryšio žmonos finansiniam raštingumui nenustatyta [21]. Papildomai Kinijos namų ūkių duomenimis paremtas Ma ir kt. tyrimas taip pat patvirtino teigiamą finansinio raštingumo poveikį vartojimo elgsenai [22]. Šie rezultatai rodo, kad finansinis raštingumas gali būti laikomas papildomu veiksniumi, formuojančiu ne tik bendrą vartojimo lygį, bet ir išlaidų paskirstymą tarp skirtingų vartojimo kategorijų.

Namų ūkių vartojimo galimybes gali veikti ir anksčiau sukauptų finansinių išteklių dydis bei jų panaudojimo būdai. Taupymas literatūroje dažnai laikomas vartojimo išlyginimo mechanizmu, leidžiančiu namų ūkiams amortizuoti laikinus pajamų svyravimus ir palaikyti stabilesnį vartojimo lygį. Fukudos tyrime, atliktame naudojant Japonijos namų ūkių panelinius duomenis, buvo analizuojamas namų ūkių prisitaikymas prie neigiamų pajamų šokų, atsirandančių dėl vyro priverstinio darbo praradimo [23]. Tyrime nustatyta, kad santaupų naudojimas buvo vienas iš svarbių namų ūkių prisitaikymo būdų – namų ūkiai mažino lėšų skyrimą taupymui ir naudojo anksčiau sukauptus išteklius vartojimo poreikiams palaikyti. Taip pat nustatyta, kad maždaug 30 % namų ūkių naudojami santaupų išėmimu kaip papildomu pajamų šaltiniu [23]. Šie rezultatai leidžia teigti, kad santaupos atlieka ne tik kapitalo kaupimo funkciją, bet ir gali prisidėti prie vartojimo svyravimų mažinimo bei stabilesnio vartojimo modelio palaikymo ekonominio neapibrėžtumo laikotarpiais.

Vartojimo elgseną formuoja ne tik einamosios pajamos, bet ir anksčiau sukauptas turtas bei jo struktūra. Literatūroje pabrėžiama, kad turtas gali veikti vartojimą dviem pagrindiniais būdais: tiesioginiu turto efektu (angl. *wealth effect*) ir galimybe naudoti turtą kaip finansinių išteklių šaltinį. Campbello ir Cocco tyrime, atliktame naudojant Jungtinės Karalystės namų ūkių duomenis buvo analizuojamas būsto kainų pokyčių poveikis skirtingų gyventojų grupių vartojimui ir nustatyta, kad būstas yra dominuojanti tipinio namų ūkio turto sudedamoji dalis, todėl jo vertės pokyčiai gali turėti reikšmingą poveikį vartojimo elgsenai [24]. Tyrimo rezultatai parodė, kad būsto kainų augimas gali būti susijęs su didesniu vartojimu ne tik dėl padidėjusios suvokiamos turto vertės, bet ir dėl galimybės lengviau skolintis naudojant turimą turtą kaip užtikrinimo priemonę. Taip pat nustatyta, kad poveikis nėra vienodas skirtingoms gyventojų grupėms – stipriausias poveikis nustatytas vyresnio amžiaus būsto savininkams, o jauniems būstą nuomojantiems asmenims statistiškai reikšmingas poveikis beveik nepasireiškė [24]. Papildomai Carrollo ir bendraautorijų atliktame tyrime, remiantis kelių šalių makroekonominiais duomenimis ir analizuojant būsto turto bei vartojimo ryšį ilgesniu laikotarpiu, nustatyta, kad vieno piniginio vieneto būsto vertės pokytis trumpuoju laikotarpiu buvo susijęs su maždaug 2 centų vartojimo padidėjimu, o ilgalaikis poveikis siekė apie 9 centus [25]. Tai leidžia

daryti prielaidą, kad sukaupto turto skirtumai gali būti susiję su ilgalaikiais vartojimo galimybių ir vartojimo stabilumo skirtumais.

Vartojimo išlaidų struktūra taip pat yra glaudžiai susijusi su asmenų padėtimi darbo rinkoje. Darbo rinkos padėties svarbą iliustruoja ir Fukudos tyrimas, kuriame nustatyta, kad vyro priverstinis darbo praradimas reikšmingai sumažina darbo pajamas, o namų ūkio vartojimo išlaidos taip pat mažėja, nors vartojimui perduodama tik apie penktadalis pajamų šoko [23]. Tai rodo, kad darbo praradimas yra reikšmingas vartojimo išlaidų pokyčių veiksnys, tačiau jo poveikis gali būti sušvelninamas nedarbo išmokomis, santaupomis ir sutuoktinio darbo pajamomis. Panašiai Europos Centrinio Banko vartotojų lūkesčių apklausos duomenimis paremtame Dias da Silvos, Rusinovos ir Weißlerio tyrime, apimančiame Belgiją, Prancūziją, Vokietiją, Italiją, Ispaniją ir Nyderlandus, nustatyta, kad netikėtas darbo praradimas buvo susijęs su neigiama vartojimo reakcija, o šis poveikis stipresnis vyresniems darbuotojams ir namų ūkiams, turintiems mažiau likvidaus turto [26]. Todėl darbo rinkos padėtis gali veikti ne tik esamą vartojimo lygį, bet ir namų ūkių galimybes prisitaikyti prie neigiamų ekonominių pokyčių.

Vartojimo elgseną taip pat gali formuoti žmogiškasis kapitalas, kurio viena svarbiausių sudedamųjų dalių yra išsilavinimas. Yükselio ir Başaro tyrime, atliktame naudojant Turkijos namų ūkių biudžetų tyrimo duomenis 2005–2019 m., nustatyta, kad išsilavinimo poveikis vartojimui pasižymi netiesiniu ryšiu [27]. Autorių rezultatai rodo, kad aukštesnis išsilavinimas ilgainiui siejamas su didesnėmis vartojimo išlaidomis, galimai dėl didesnių pajamų ir besikeičiančių vartojimo prioritetų. Analizuojant konkrečias vartojimo kategorijas nustatyta, namų ūkio narių išsilavinimo lygis siejamas su didesnėmis būsto, transporto, komunikacijos, rekreacijos, kultūros ir švietimo išlaidomis; ypač stiprus ryšys nustatytas švietimo išlaidoms, kur darbo jėgos išsilavinimo koeficientas siekė 1,364, tačiau kvadratinis narys buvo neigiamas ($-0,163$), rodantis mažėjančio ribinio poveikio arba apverstos U formos ryšį [27]. Tai leidžia išsilavinimą vertinti kaip veiksni, galintį keisti vartojimo prioritetus ir ilgainiui transformuoti namų ūkių išlaidų struktūrą.

Reikšmingą vaidmenį vartojimo struktūros formavimuisi turi ir gyvenimo ciklo ypatumai – amžius. Tame pačiame Yükselio ir Başaro tyrime nustatyta, kad bendras vartojimas su amžiumi didėja, tačiau vėliau jo augimas lėtėja ir ima mažėti [27]. Tyrime nurodoma, kad namų ūkio galvos amžiaus koeficientas bendroms vartojimo išlaidoms yra teigiamas, o kvadratinis amžiaus narys – neigiamas, todėl amžiaus poveikis vartojimui nėra tiesinis. Atskirų vartojimo kategorijų analizė parodė, kad alkoholio, tabako, komunikacijos ir švietimo išlaidos taip pat pasižymi apverstos U formos ryšiu su amžiumi, o būsto, komunalinių paslaugų, baldų ir namų priežiūros išlaidos labiau išryškėja vyresniuose namų ūkiuose [27]. Amžiaus ir vartojimo sąsają papildomai iliustruoja Campbello ir Cocco tyrimas, kuriame nustatyta, kad būsto kainų poveikis vartojimui skiriasi pagal amžiaus bei būsto nuosavybės grupes – stipriausias poveikis nustatytas vyresnių būsto savininkų grupėje, o jaunų būstą nuomojančių asmenų grupėje jis buvo mažiausias ir statistiškai nereikšmingas [24]. Todėl amžius gali būti siejamas ne tik su besikeičiančiais poreikiais, bet ir su skirtingomis finansinėmis galimybėmis įvairiais gyvenimo ciklo etapais.

Vartojimo modelius taip pat formuoja namų ūkio sudėtis bei jo vidaus struktūra. Yükselio ir Başaro tyrimas taip pat parodė, kad namų ūkio dydžio ir bendrų vartojimo išlaidų ryšys pasižymi nelinijiniu apverstos U formos ryšiu: namų ūkio dydžio koeficientas buvo teigiamas, o kvadratinis dydžio narys – neigiamas [27]. Tai reiškia, kad didėjant namų ūkio dydžiui bendras vartojimas iš pradžių auga, tačiau vėliau augimas lėtėja, galimai dėl masto ekonomijos ir išteklių dalijimosi efektų. Tyrime taip

pat nustatyta, kad vaikų ir vyresnio amžiaus asmenų buvimas keičia išlaidų struktūrą: didesnė mokyklinio amžiaus vaikų dalis namų ūkyje buvo siejama su didesnėmis švietimo, rekreacijos, kultūros ir aprangos išlaidomis, o didesnė vyresnio amžiaus asmenų dalis buvo siejama su didesnėmis maisto bei maitinimo išlaidomis ir mažesnėmis švietimo bei transporto išlaidomis. Be to, vienišų tėvų namų ūkiai pasižymėjo mažesnėmis bendromis vartojimo išlaidomis, o išplėstinės šeimos pasižymėjo šiek tiek didesnėmis bendromis vartojimo išlaidomis, nors didesnis namų ūkio dydis gali sudaryti sąlygas efektyvesniam išteklių paskirstymui ir masto ekonomijos efektams [27]. Namų ūkio sudėties reikšmę papildomai pagrindžia Browningo, Chiappori ir Lewbelio teorinis bei empirinis tyrimas, kuriame modeliuojamos vartojimo masto ekonomijos, suaugusiųjų ekvivalentiškumo skalės ir derybinė galia tarp namų ūkio narių. Autoriai pabrėžia, kad gyvenimas kartu leidžia dalytis vartojimo gėrybėmis ir taip sumažinti kiekvienam nariui tenkančias išlaidas, o namų ūkio ištekliai paskirstomi tarp narių pagal dalijimosi taisyklę [28]. Taigi namų ūkio struktūra gali keisti tiek bendrą vartojimo lygį, tiek konkrečių išlaidų kategorijų prioritetus.

Papildomus vartojimo skirtumus gali paaiškinti ir kiti socialiniai bei demografiniai veiksniai, tarp jų – lytis. Poveikis gali pasireikšti ne bendru vartojimo lygiu, bet atskirų vartojimo kategorijų prioritetais. Pavlisos tyrime, atliktame naudojant 2010 m. Eurostat Prancūzijos namų ūkių biudžetų tyrimo mikrolygmens duomenis, buvo analizuojami skirtingų profesinių grupių ir socialinių charakteristikų ryšiai su vartojimo struktūra [29]. Tyrimo rezultatai parodė statistiškai reikšmingą lyties poveikį reprezentacinėms išlaidoms: namų ūkiuose, kuriuose namų ūkio galva buvo moteris, su išvaizda ir reprezentacija susijusios išlaidos buvo vidutiniškai 12 procentinių punktų didesnės nei namų ūkiuose, kuriuose namų ūkio galva buvo vyras. Tai leidžia daryti prielaidą, kad lyties poveikis dažniau pasireiškia ne bendram vartojimo lygiui, o skirtingų vartojimo kategorijų pasirinkimui.

Literatūroje nustatyta, kad vartojimo struktūrą veikia ne tik disponuojamų pajamų lygis, bet ir finansinis raštingumas, sukauptos santaupos, turtas, užimtumo statusas, išsilavinimas, amžius, namų ūkio sudėtis bei lytis. Šie veiksniai lemia ne tik bendrą vartojimo lygį, bet ir skirtingų išlaidų kategorijų prioritetus bei namų ūkių gebėjimą prisitaikyti prie ekonominių pokyčių. Todėl, siekiant identifikuoti vartojimo modelius ir socialines grupes, svarbu vertinti ne pavienius rodiklius, o jų tarpusavio sąveiką, kuri geriau atspindi realią namų ūkių ekonominę bei socialinę situaciją.

1.4. Socialinių grupių identifikavimo metodų apžvalga

Socialinių grupių identifikavimas empiriniuose tyrimuose yra sudėtingas uždavinys, kadangi socialinė tapatybė dažnai yra daugialypė, dinamiška ir priklausoma nuo konteksto. Šią problemą savo tyrime aptaria Cruwys'as ir kt., pabrėždami, kad grupinė priklausomybė ne visuomet gali būti apibrėžiama remiantis vien objektyviomis socialinėmis ar demografinėmis charakteristikomis, nes svarbią reikšmę turi ir subjektyvus individo tapatinimasis su tam tikromis socialinėmis grupėmis [30]. Vienas iš tokių subjektyvios socialinės tapatybės ir grupinių ryšių identifikavimo bei vizualizavimo metodų yra socialinės tapatybės žemėlapis (angl. *Social Identity Mapping*, SIM), kurio metu dalyviai patys identifikuoja jiems reikšmingas grupes, įvertina jų svarbą ir tarpusavio ryšius. Šis metodas leidžia atskleisti ne tik formalią grupinę priklausomybę, bet ir subjektyviai suvokiamą socialinės tapatybės struktūrą bei grupių tarpusavio santykius. Vis dėlto SIM metodas yra gana imlus laiko ir duomenų rinkimo sąnaudų požiūriu, todėl jos taikymas didelės apimties tyrimuose gali būti sudėtingas.

Nors subjektyvus socialinės tapatybės vertinimas gali būti svarbus siekiant suprasti individo socialinius ryšius bei grupinį identitetą, namų ūkių ir vartojimo elgsenos tyrimuose socialinių grupių identifikavimas dažniausiai remiasi objektyviomis socialinėmis ir ekonominėmis charakteristikomis, tačiau socialinė diferenciacija paprastai negali būti paaiškinta vienu rodikliu. Namų ūkių struktūrą formuoja pajamos, turtas, išsilavinimas, užimtumas, gyvenimo būdo elementai, vartojimo modeliai, gyvenamo būsto sąlygos, šeimos sudėtis ir kiti tarpusavyje susiję socialiniai indikatoriai. Hennigo ir Liao socioekonominės stratifikacijos tyrime pabrėžiama, kad mišraus tipo kintamieji – skaitiniai, ranginiai ir kategoriniai – yra būdingi socialinės stratifikacijos tyrimams, nes socialinių grupių klasifikavimas dažniausiai remiasi ne vienu požymiu, o kelių skirtingų matavimo skalių rodiklių deriniu [31]. Tokiu atveju socialinių grupių identifikavimas gali būti suprantamas kaip daugiamatės panašumo struktūros identifikavimas, o ne kaip mechaninis namų ūkių suskirstymas pagal vieną pasirinktą kriterijų.

Kadangi socialinės grupės dažniausiai nėra tiesiogiai stebimos duomenyse, o jų ribos iš anksto nėra žinomos, pagrindiniu uždaviniu tampa ne pačių grupių paieška, bet tarpusavio panašumo tarp namų ūkių identifikavimas. Tokiu atveju siekiama nustatyti, ar egzistuoja namų ūkių rinkiniai, kurie pasižymi panašiomis socialinėmis, ekonominėmis ir demografinėmis charakteristikomis bei gali būti interpretuojami kaip potencialios socialinės grupės. Būtent šiame kontekste klasterizavimo metodai tampa svarbūs, kadangi leidžia identifikuoti duomenyse slypinčias struktūras be iš anksto apibrėžtų grupių ribų.

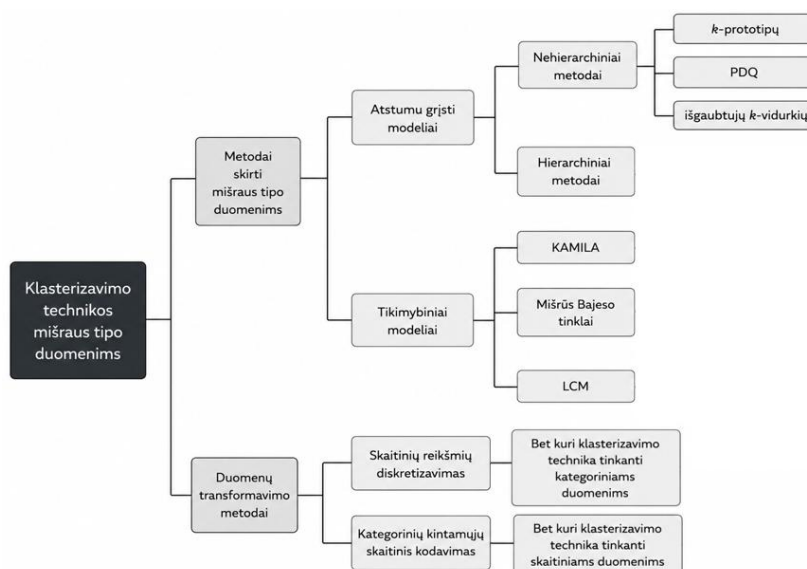
Toks požiūris lemia ir papildomus metodologinius iššūkius, susijusius su socialinio panašumo apibrėžimu. Hennigas ir Liao pabrėžia, kad socialinės grupės samprata nėra vienareikšmiškai apibrėžta, todėl duomenimis grindžiama socialinė stratifikacija turi atsakyti ne tik į tai, kaip objektus suskirstyti, bet ir į tai, kokie indikatoriai turėtų apibrėžti socialinį panašumą [31]. Autoriai taip pat akcentuoja, kad klasterizavimas reikalauja aiškaus ryšio tarp tyrimo tikslo, panašumo sampratos ir pasirinktų kintamųjų apibrėžimo. Vadinasi, socialinių grupių identifikavimas nėra vien techninė procedūra – tyrėjas turi pagrįsti, kodėl vieni rodikliai laikomi socialinę diferenciaciją atskleidžiančiais, o kiti laikomi mažiau reikšmingais.

Dėl šios priežasties vienu svarbiausių mišraus tipo duomenų klasterizavimo etapų tampa kintamųjų pasirinkimas. Hennigo ir Liao analizuotame 2007 m. JAV vartotojų finansų tyrime iš pradinio duomenų rinkinio buvo analizuoti 17 430 vyrų stebėjimai, o socialinei stratifikacijai vertinti naudoti aštuoni mišraus tipo kintamieji: santaupos, pajamos, išsilavinimo metai, einamųjų sąskaitų skaičius, taupomųjų sąskaitų skaičius, būsto statusas, gyvybės draudimas ir profesinė klasė [31]. Šis pavyzdys leidžia daryti prielaidą, kad socialinės grupės turi būti apibrėžiamos remiantis subalansuotu ekonominių, socialinių ir demografinių indikatorių rinkiniu. Kintamųjų pasirinkimo svarbą pabrėžia ir Herro namų ūkių klasifikavimo tyrimas, kuriame analizuoti 2 819 Indonezijos namų ūkių duomenys ir 245 kintamieji, iš kurių 81 % buvo kategoriniai [32]. Toks duomenų rinkinys atskleidžia praktinę socialinių tyrimų problemą: apklausų duomenyse dažnai fiksuojamas itin platus požymių spektras, tačiau ne visi požymiai vienodai prisideda prie prasmingų socialinių grupių formavimo. Todėl kintamųjų atranka, nereikšmingų ar dubliuojančių požymių šalinimas bei svarbiausių grupės skiriančių indikatorių nustatymas tampa būtina klasterizavimo proceso dalimi.

Atsižvelgiant į socialinių grupių daugiamatiškumą ir mišrią analizuojamų duomenų struktūrą, praktikoje kyla ne tik socialinio panašumo apibrėžimo, bet ir tinkamo klasterizavimo metodo pasirinkimo klausimas. Klasterizavimas apibūdinamas kaip uždavinys, kuriame objektai suskirstomi

į santykinai homogeniškas grupes, siekiant maksimaliai padidinti vidinį panašumą ir sumažinti panašumą tarp skirtingų klasterių. Costa ir bendraautorai papildomai pažymi, kad socialinių mokslų duomenų rinkiniuose dažnai kartu egzistuoja skirtingų tipų demografiniai ir socioekonominiai rodikliai, todėl mišraus tipo duomenų klasterizavimas tampa svarbia kiekybinės empirinės metodologijos dalimi [33]. Vis dėlto mišraus tipo duomenų klasterizavimas yra sudėtingesnis nei vien skaitinių ar vien kategorinių duomenų analizė, nes reikia nuspręsti, kaip apskaičiuoti skirtingų tipų kintamųjų panašumą, kaip sujungti skaitinius ir kategorinius požymius bei kokį algoritmą taikyti galutiniam grupių identifikavimui.

Mišraus tipo duomenų klasterizavimo metodų spektras literatūroje pateikiamas keliomis klasifikavimo logikomis. Ghattaso ir kt. sisteminėje schemoje (žr. 4 pav.) metodai išskaidomi pagal tai, ar jie remiasi duomenų transformavimu, ar yra specialiai sukurti mišriems duomenims [34]. Specialiai mišriems duomenims sukurti metodai toliau skaidomi į atstumu grįstus ir tikimybinus modelius. Atstumu grindžiamai kryptčiai priskiriami k -prototipų, PDQ, išgaubtųjų k -vidurkių metodas, hierarchinė klasterizacija naudojant Gower ar kitus atstumo matus. Tikimybiniais modeliais grindžiamai kryptčiai KAMILA, LCM ir mišrūs Bajeso tinklų modeliai.



4 pav. Klasterizavimo technikų mišraus tipo duomenims sisteminė schema (sudaryta pagal Ghattasą ir kt. [34])

Vienas paprasčiausių, tačiau kartu ir vienas mažiausiai efektyvių mišrių duomenų apdorojimo būdų yra jų transformavimas į vieną duomenų tipą [34]. Dažniausiai naudojamos transformacijos yra kategorinių kintamųjų perkodavimas į dvejetainius vektorius arba tęstinių kintamųjų diskretizavimas. Vieno kategorinio požymio perkodavimas į kelis dvejetainius požymius gali padidinti dimensijų skaičių ir pakeisti skirtingų kintamųjų įtaką atstumų skaičiavimui. Prie duomenų transformavimo principais grindžiamų metodų taip pat priskiriami metodai, perkeliantys mišrius duomenis į bendrą reprezentacinę erdvę. Vienas tokių metodų yra FAMD (angl. *Factor Analysis for Mixed Data*) naudojimas, kai mišrūs kintamieji perkeliama į bendrą sumažintos dimensijos erdvę, o klasterizacija atliekama jau naudojant faktorių balus. Costa ir bendraautorai šią kryptį apibūdina kaip sekvenčią metodą, kai pirmiausia taikoma FAMD, siekiant gauti stebinių koordinatės mažesnio matmens erdvėje, o vėliau šiems balams taikomas k -vidurkių klasterizavimas [33].

Vis dėlto vietoje duomenų transformavimo literatūroje plačiai taikomi ir metodai, tiesiogiai pritaikyti mišriems duomenims. Viena svarbiausių jų grupių yra atstumo metrikomis pagrįsti metodai, kuriuose panašumas tarp objektų apskaičiuojamas tiesiogiai naudojant skirtingų tipų požymius. Vienas dažniausiai taikomų sprendimų yra Gower atstumas. Costa ir bendraautorai nurodo, kad populiarī praktika yra pirmiausia apskaičiuoti Gower panašumo arba skirtumo matricą, o vėliau jai taikyti k-medoidų arba hierarchinę klasterizaciją [34]. Preud'homm'as ir bendraautorai išsamiau paaiškina, kad Gower atveju tęstinių kintamųjų skirtumai normalizuojami pagal jų reikšmių intervalą, o kategorinių kintamųjų skirtumai vertinami pagal sutapimo–nesutapimo principą. Taip skaitiniai ir kategoriniai skirtumai apjungiami į bendrą atstumo matricą [35].

Praktiniuose taikymuose Gower atstumas dažnai derinamas su k-medoidų PAM algoritmu arba hierarchine klasterizacija. K-medoidų metodas formuoja klasterius aplink realius stebinius – medoidus – todėl jo rezultatai gali būti lengviau interpretuojami kaip tipiniai grupių atstovai. Bektaso ir Schumanno mobilumo profilių tyrime šią logiką pritaikė Šveicarijos gyventojų segmentavimui. Kadangi duomenys buvo mišraus tipo, autoriai pasirinko Gower atstumą, o klasteriams formuoti – k-medoidų PAM algoritmą [36]. Autoriai pirmiausia atrinko mobilumą apibūdinančius požymius: automobilių skaičių namų ūkyje, kelionių skaičių per dieną, pagrindinį transporto būdą, pusės kainos viešojo transporto kortelės turėjimą, multimodalumą ir kasdienį atstumą. Vėliau buvo testuojami klasterių skaičiai nuo 2 iki 15; geriausias sprendimas gautas su 13 klasterių, kurių vidutinis silueto koeficientas siekė 0,7465. Optimizavus Gower svorius šis rodiklis padidėjo iki 0,8458. Didžiausi svoriai buvo suteikti transporto būdo pasirinkimui, multimodalumui ir pusės kainos kortelės turėjimui, o likusių kintamųjų svoriai išliko artimi pradiniais. Bektaso ir Schumanno rezultatai rodo, kad svorių optimizavimas gali reikšmingai pagerinti klasterių atskyrimą, todėl kintamųjų svarbos vertinimas aktualus ir socialinių grupių identifikavimo uždaviniuose.

Kita plačiai naudojama mišrių duomenų klasterizavimo metodų grupė yra prototipais pagrįsti metodai. Tradicinis k-vidurkių metodas tinkamas skaitiniams duomenims, tačiau negali tinkamai įvertinti kategorinių požymių. Todėl buvo sukurtas k-prototipų metodas, apjungiantis skaitinių ir kategorinių požymių informaciją bendroje klasterizavimo procedūroje. Aschenbruckas ir bendraautorai pažymi, kad k-prototipų metodas yra k-vidurkių plėtinys mišraus tipo duomenims, dažnai pasirenkamas dėl paaiškinamumo, gebėjimo apdoroti didelius duomenų kiekius ir galimybės taikyti paralelinį skaičiavimą [37]. Vis dėlto algoritmo rezultatai jautrūs pradinį prototipų pasirinkimui, todėl praktikoje svarbūs pakartotiniai algoritmo paleidimai.

Dar viena mišrių duomenų klasterizavimo kryptis yra tikimybiniai metodai. Hennigas ir Liao lygina latentinių klasių modelį su disimiliarumu grįstu k-medoidų metodu ir pabrėžia, kad skirtingos metodų šeimos remiasi skirtingomis prielaidomis apie tai, kas laikoma klasteriu [31]. Latentinių klasių modeliuose daroma prielaida, kad stebimi duomenys generuojami iš paslėptų klasių, todėl grupės identifikuojamos remiantis tikimybinio priklausymu. Tuo tarpu atstumo metodai grupes formuoja remdamiesi stebinių tarpusavio panašumu. Dėl šios priežasties skirtingi metodai gali pateikti skirtingas to paties duomenų rinkinio struktūras.

Taigi socialinių grupių identifikavimas mišraus tipo duomenyse yra nelengvas uždavinys, apimantis ne tik grupių formavimą, bet ir socialinio panašumo apibrėžimą. Literatūros analizė rodo, kad reikšmingą vaidmenį šiame procese atlieka pasirinktų kintamųjų rinkinys bei metodų gebėjimas integruoti skirtingų tipų požymių informaciją.

1.5. Vartojimo išlaidų struktūrų analizės metodų apžvalga

Vartojimo išlaidų struktūrų analizė ekonominiuose ir socialiniuose tyrimuose pasižymi metodologine įvairove, nes vartojimo elgsena formuojama iš daugelio tarpusavyje susijusių ekonominių, demografinių ir socialinių veiksnių. Literatūroje vartojimo struktūrų tyrimui taikomi metodai gali būti skirstomi į kelias pagrindines grupes: klasikinės inferencinės statistikos metodus, ekonometrinius vartojimo modelius, mašininio mokymosi metodus bei informacijos teorija grįstus ryšio vertinimo metodus.

Klasikiniai statistinio palyginimo metodai

Vienas dažniausiai taikomų metodologinių požiūrių vartojimo struktūrų analizėje remiasi grupių vidurkių, dispersijų ir pasiskirstymų palyginimu. Tokie metodai leidžia identifikuoti, ar skirtingos socialinės, demografinės ir ekonominės grupės statistiškai reikšmingai skiriasi pagal vartojimo išlaidų rodiklius.

Vienmatė dispersinė analizė (ANOVA) yra vienas plačiausiai taikomų metodų, skirtų nustatyti, ar kelių nepriklausomų grupių vidurkiai skiriasi statistiškai reikšmingai. Šio metodo teorinius pagrindus suformulavo Fišeris, parodęs, kad bendrą duomenų variaciją galima išskaidyti į tarpgrupinę ir vidinę grupių variaciją, o jų santykis leidžia spręsti apie grupių skirtumų reikšmingumą. ANOVA metodas plačiai taikomas įvairiuose tyrimuose analizuojant skirtingų grupių ekonominius ir vartojimo rodiklius. Zamanas vertindamas JAV namų ūkių išlaidų pokyčius prieš ir po COVID-19 laikotarpio, vienfaktorinę ANOVA taikė reikšmingiems skirtumams tarp skirtingų laikotarpių nustatyti, duomenis suskirsčius į keturias grupes pagal tiriamą laikotarpį [38]. Tyrime pažymima, kad vienfaktorinė ANOVA naudojama reikšmingiems skirtumams tarp grupių nustatyti, o prieš analizę turi būti patikrintos pagrindinės metodo prielaidos – stebėjimų nepriklausomumas, dispersijų homogeniškumas ir grupių skirstinių normalumas. Kai reiškinys aprašomas keliomis tarpusavyje susijusiomis kategorijomis, literatūroje dažnai taikoma daugiamatė dispersinė analizė (MANOVA). MANOVA leidžia vienu metu tarp grupių vertinti kelių priklausomų kintamųjų skirtumus bei atsižvelgti į koreliacinę struktūrą tarp analizuojamų požymių, todėl yra tinkama daugiamatėms struktūroms tirti. Hristovas ir bendraautorai, analizuodami namų ūkių maisto vartojimo ir pirkimo elgsenos pokyčius skirtingose Europos šalyse COVID-19 laikotarpiu, šį metodą taikė socialinių ir ekonominių veiksnių sąsajoms su priklausomų kintamųjų visuma įvertinti [39]. Nustačius bendrus skirtumus tarp grupių, papildomai buvo atliekami poriniai grupių palyginimai.

Statistinio palyginimo metodams taip pat priskiriami empirinių pasiskirstymų palyginimo ir atitikimo testai (angl. *goodness-of-fit tests*), tokie kaip *Pirsono*, *Chi kvadrato*, *Kolmogorovo–Smirnov* ar *Andersono–Darlingo* kriterijai. Šie metodai leidžia įvertinti, ar empirinės vartojimo struktūros atitinka teorinius pasiskirstymus arba ar skirtingų socialinių grupių vartojimo pasiskirstymai skiriasi statistiškai.

Ekonometriniai vartojimo struktūrų modeliai

Vartojimo struktūrų analizėje plačiai naudojami ekonometriniai modeliai, leidžiantys vertinti vartojimo priklausomybę nuo pajamų, kainų iki kitų socioekonominių veiksnių. Vienas iš klasikinių metodų yra Engelio kreivių analizė, paremta Ernsto Engelio suformuluotu dėsnium, teigiančiu, kad didėjant namų ūkio pajamoms maisto išlaidų dalis bendrame biudžete mažėja. Vėlesni tyrimai Engel kreives naudojo ne tik vartojimo ryšių aprašymui, bet ir namų ūkių poreikių bei vartojimo skirtumų

vertinimui. Didelis ir bendraautorai pažymi, kad Engelio požiūris grindžiamas prielaida, jog namų ūkių gerovės lygis gali būti vertinamas pagal maistui skiriamų išlaidų dalį, o skirtingų namų ūkių tipų pajamų santykiai leidžia įvertinti jų vartojimo skirtumus [40]. Be to, autoriai pažymi, kad išlaidų maistui dalis mažėja didėjant pajamoms, todėl ši priklausomybė tapo vienu iš svarbiausių vartojimo struktūros analizės pagrindų.

Vėlesniuose tyrimuose vartojimo struktūrų modeliavimui plačiai taikomas išplėstinis linijinės išlaidų sistemos modelis (angl. *Extended Linear Expenditure System*, ELES), leidžiantis vienu metu vertinti biudžeto dalių pasiskirstymą ir vartojimo elastingumus. ELES modelis remiasi prielaida, kad vartotojų paklausa tam tikru laikotarpiu priklauso nuo pajamų bei skirtingų prekių kainų, o vartojimas skaidomas į bazinį vartojimą ir vartojimą, viršijantį būtinąjį poreikį. Po bazinių poreikių patenkinimo papildomos vartojimo išlaidos paskirstomos pagal ribinį polinkį vartoti atskiras prekių ar paslaugų kategorijas. Naujausi empiriniai tyrimai rodo, kad ELES modelis efektyviai naudojamas namų ūkių vartojimo struktūros ir pajamų elastingumo analizei. Fanas ir bendraautorai, taikydami ELES modelį Kinijos namų ūkių vartojimo struktūrai tirti, nustatė, kad maisto vartojimo bazinis poreikis ir ribinis polinkis vartoti išlieka didžiausi, tačiau vartojimo struktūra palaipsniui pereina nuo bazinių poreikių tenkinimo prie vystymosi ir gerovės gerinimo tipo vartojimo [41].

Taip pat dažnai taikomi daugiakriteriniai regresiniai modeliai, tokie kaip daugialypė regresija, logistinė regresija bei panelinių duomenų modeliai, leidžiantys įvertinti socioekonominių charakteristikų poveikį vartojimo struktūros pokyčiams.

Mašininio mokymosi metodai

Didelės apimties vartojimo duomenų analizėje vis dažniau taikomi mašininio mokymosi metodai, leidžiantys identifikuoti sudėtingas netiesines priklausomybes bei paslėptas vartojimo elgsenos struktūras. Tokie metodai kaip sprendimų medžiai, atsitiktiniai miškai, gradientinio stiprinimo algoritmai ar neuroniniai tinklai leidžia analizuoti didelės dimensijos vartojimo duomenis bei prognozuoti vartotojų elgseną [42, 43]. Šių metodų privalumas – gebėjimas vienu metu apdoroti didelį kintamųjų skaičių bei aptikti sudėtingas sąveikas tarp požymių, kurios ne visuomet gali būti lengvai identifikuojamos taikant klasikinius statistinius ar ekonometrinius metodus. Be to, mašininio mokymosi modeliai dažnai nereikalauja iš anksto griežtai apibrėžti funkcinių ryšių tarp kintamųjų, todėl gali būti lankstesni analizuojant sudėtingas duomenų struktūras.

Tačiau mašininio mokymosi metodų taikymas gali būti ribotas tais atvejais, kai analizuojami mažos apimties duomenų rinkiniai, kadangi didžiausias šių metodų efektyvumas dažniausiai pasiekiamas esant dideliame stebėjimų skaičiui. Mažesnėse imtyse gali padidėti persimokymo (angl. *overfitting*) tikimybė, sumažėti modelių stabilumas bei jų gebėjimas patikimai identifikuoti bendrus duomenų dėsningumus. Taip pat, sudėtingesni modeliai neretai pasižymi mažesniu rezultatų interpretuojamumu, todėl jų taikymas gali apsunkinti aiškų atskirų veiksmų poveikio paaiškinimą.

Entropija grindžiami ryšio metodai

Pastaraisiais dešimtmečiais analizėje vis didesnio dėmesio sulaukia informacijos teorija grįsti metodai, leidžiantys vertinti sudėtingas priklausomybes tarp kintamųjų bei identifikuoti duomenyse esančius dėsningumus. Informacijos teorijos pagrindus suformulavęs Claudas Shannonas parodė, kad entropija gali būti interpretuojama kaip atsitiktinio kintamojo neapibrėžtumo matas, o tarpusavio informacija (angl. *mutual information*) – kaip neapibrėžtumo sumažėjimas stebint kitą kintamąjį.

Vėliau Coveris ir Thomasas išplėtojo šią koncepciją parodydami, kad entropija ir tarpusavio informacija leidžia identifikuoti tiek tiesines, tiek netiesines priklausomybes nepriklausomai nuo kintamųjų pasiskirstymo formos.

Tarpusavio informacijos (angl. *mutual information*) metodai šiuolaikinėje statistinėje analizėje laikomi universaliais priklausomybės matais, galinčiais aptikti sudėtingus ir aukštesnio laipsnio ryšius tarp kintamųjų, kurių ne visuomet identifikuoja klasikiniai koreliacijos metodai [44]. Šie metodai gali būti taikomi įvairių tipų duomenims analizuoti, kadangi jiems nereikalinga griežta prielaida apie tiesinį ryšį ar konkrečią duomenų pasiskirstymo formą. Dėl šių savybių informacijos teorija grįsti metodai vis dažniau taikomi situacijose, kai siekiama analizuoti sudėtingas daugiamates duomenų struktūras bei tarpusavio sąveikas tarp požymių.

1.6. Skirtingų socialinių grupių vartojimo išlaidų struktūrų tyrimų stoka

Nagrinėjant mokslinę literatūrą, susijusią su socialinių grupių identifikavimu ir namų ūkių vartojimo išlaidų struktūrų analize, matyti, kad daugumoje tyrimų vartojimo išlaidos analizuojamos bendros imties lygmeniu, daugiausia dėmesio skiriant pavienių socialinių, ekonominių ar demografinių veiksnių poveikiui. Dažniausiai nagrinėjamos pajamų, amžiaus, išsilavinimo ir užimtumo sąsajos su vartojimo lygiu bei atskiromis vartojimo kategorijomis. Nors tokie veiksniai leidžia paaiškinti atskirus vartojimo skirtumus, jų analizė ne visada leidžia atskleisti ryšius, susiformuojančius dėl skirtingų namų ūkių charakteristikų tarpusavio sąveikos. Kadangi namų ūkių vartojimo elgsena formuojasi veikiant socialinių, ekonominių ir demografinių požymių visumai, panašūs vartojimo dėsningumai gali būti susiję ne tik su pavienėmis charakteristikomis, bet ir su jų deriniais. Tuo tarpu tyrimų, kuriuose socialinės grupės būtų identifikuojamos remiantis socialinių, ekonominių ir demografinių požymių kombinacijomis, o jų vartojimo išlaidų struktūros analizuojamos grupių lygmeniu, literatūroje beveik nėra. Dėl šios priežasties aktualu vartojimo išlaidų skirtumus nagrinėti ne tik bendros imties ar pavienių veiksnių lygmeniu, bet ir duomenimis identifiкуotų socialinių grupių kontekste. Panašiomis mikrolygmens charakteristikomis pasižyminčių namų ūkių grupavimas gali padėti identifikuoti imtyje egzistuojančius dėsningumus bei atskleisti gyventojų vartojimo struktūrų skirtumus, kurie vertinant pavienius veiksnius gali likti nepastebėti.

2. Metodologija

Šiame skyriuje pristatoma tyrime taikoma metodologija, aptariamas teorinis jos pagrindimas, pagrindžiamas pasirinktų rodiklių ir metodų taikymas bei aprašoma naudota programinė įranga. Remiantis mokslinėje literatūroje pateikiamais metodologiniais sprendimais (žr. 1.4–1.5 skyrius), šiame darbe taikoma klasterinė analizė, siekiant pagal turimas namų ūkių charakteristikas identifikuoti skirtingas socialines grupes. Vėliau tarp nustatytų grupių atliekama statistinė vartojimo išlaidų analizė, siekiant įvertinti identifikuotų socialinių grupių skirtumus pagal vartojimo struktūrą.

2.1. Duomenų rinkinys ir jo paruošimas

Naudojamas Lietuvos oficialiosios statistikos sistemoje sukauptas 2021 m. namų ūkių biudžetų statistinio tyrimo duomenų rinkinys, kuris šio darbo rengimo metu atitinka naujausią viešai prieinamą informaciją. Nors 2025 m. rugpjūčio 13 d. buvo atnaujintas šio statistinio tyrimo teisinis ir organizacinis reglamentavimas², remiantis Valstybės duomenų agentūros skelbiama informacija, kitas duomenų rinkimo ciklas vykdomas 2026 m., o oficialus statistinių rezultatų publikavimas³ numatomas tik 2027 m. rugsėjo mėnesį. Naudojami duomenys yra ilgalaikės Lietuvos oficialiosios statistikos stebėsenos sistemos dalis – namų ūkių biudžetų tyrimai šalyje vykdomi nuo 1992 m., o tyrimo metodika nuosekliai tobulinama vadovaujantis Eurostat rekomendacijomis [2]. Statistinio stebėjimo vienetas šiame tyrime yra namų ūkis, leidžiantis vartojimo elgseną analizuoti bendro išteklių paskirstymo lygmeniu. Valstybės duomenų agentūra duomenis renka taikant atrankinio statistinio tyrimo principus, užtikrinančius skirtingų socialinių, ekonominių ir geografinių gyventojų grupių reprezentatyvumą. Duomenų rinkimas vykdytas derinant tiesioginę apklausą, namų ūkio išlaidų dienoraštį ir administracinių registrų informaciją [2], todėl išlaikomas metodologinis nuoseklumas, didelis detalumo lygis bei sudaromas patikimas empirinis pagrindas socialinių grupių vartojimo struktūrų analizei.

2021 m. namų ūkių biudžetų statistinio tyrimo duomenų rinkinį sudaro trys tarpusavyje susiję duomenų šaltiniai: individualūs namų ūkių narių duomenys, bendri namų ūkių charakteristikų duomenys ir vartojimo išlaidų duomenys. Visi šie duomenys publikuojami ir viešai prieinami per Valstybės duomenų agentūros portalą⁴. Pirminį duomenų failą sudaro individualių duomenų informacija, apimanti namų ūkių narių demografines ir socialines charakteristikas, tokias kaip amžius, lytis, išsilavinimas, ekonominis aktyvumas, profesinė padėtis ir kiti individualaus lygmens požymiai. Šį failą sudaro 8496 stebėjimai, kuriuose kiekviena eilutė atitinka konkretų tiriamo namų ūkio narį. Antrąjį duomenų failą sudaro namų ūkių charakteristikų informacija, apimanti namų ūkio sudėtį, gyvenamąją vietą, pajamų kvintilinę grupę, būsto sąlygas ir kitus socialinius-ekonominius rodiklius. Šiame faile pateikiami 4334 stebėjimai, kuriuose kiekvienas įrašas reprezentuoja vieną namų ūkį. Trečiąjį duomenų failą sudaro vartojimo išlaidų informacija, kurioje pateikiamos detalizuotos namų ūkių išlaidos, remiantis COICOP klasifikatoriumi. Pažymėtina, kad duomenų faile vartojimo išlaidos papildomai skirstomos į pinigines ir natūrinės vartojimo išlaidas, išreikštas eurai, tačiau šiame darbe naudojamas bendras vartojimo išlaidų dydis, neatskiriant šių komponentų. Taip pat, siekiant užtikrinti rezultatų interpretacinį aiškumą ir palyginamumą tarp skirtingų socialinių grupių, analizėje naudojamas tik bendrų mėnesinių namų ūkių vartojimo išlaidų kintamasis ir 13 pagrindinių vartojimo išlaidų kategorijų, atitinkančių pirmąjį COICOP klasifikatoriaus lygmenį (žr.

² <https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/33157f11787c11f09f35d7a7964762ee?jfwid=-r5oyymmqe8>

³ <https://vda.lrv.lt/public/canonical/1765786371/2187/N%C5%AAB-01-p.pdf>

⁴ <https://open.geodata.gov.lt/pages/c7530ae3b0a44befb5df054f3ca04810>

2 lentelę). Išsamūs kiekviename faile pateikiamų kintamųjų aprašai, kodavimas ir metaduomenys pateikiami viešai prieinamose duomenų specifikacijose: individualių duomenų kintamųjų sąrašas⁵, namų ūkių charakteristikų kintamųjų sąrašas⁶ ir vartojimo išlaidų kintamųjų sąrašas⁷.

2 lentelė. Namų ūkių vartojimo išlaidų kategorijos pagal pirmojo lygmens COICOP klasifikavimą

Pirmojo lygmens kodas	Pavadinimas
00	Visos namų ūkio vartojimo išlaidos (mėnesinės)
01	Maistas ir nealkoholiniai gėrimai
02	Alkoholiniai gėrimai, tabakas ir narkotikai
03	Apranga ir avalynė
04	Būstas, vanduo, elektra, dujos ir kitas kuras
05	Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra
06	Sveikata
07	Transportas
08	Informacija ir ryšiai
09	Poilsis, sportas ir kultūra
10	Švietimo paslaugos
11	Restoranai ir apgyvendinimo paslaugos
12	Draudimas ir finansinės paslaugos
13	Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos

Šiame darbe aprašyti duomenų failai sujungiami naudojant unikalų namų ūkio identifikatorių, leidžiantį kiekvienam namų ūkiui susieti individualius narių požymius, bendrąsias namų ūkio charakteristikas bei faktines vartojimo išlaidas. Kadangi individualių duomenų rinkinyje vienam namų ūkiui gali priklausyti keli įrašai, pirmame duomenų paruošimo etape individualūs duomenys agreguojami namų ūkio lygmeniu, apskaičiuojant tyrimui reikšmingus socialinius indikatorius. Antrajame etape šie agreguoti rodikliai sujungiami su namų ūkių charakteristikomis apibūdinančiais kintamaisiais, o paskutiniame etape prijungiama vartojimo išlaidų informacija, leidžianti suformuoti galutinę analizės matricą.

Pirminiame tyrimo projektavimo etape buvo svarstoma taikyti tik namų ūkio galvos individualius požymius, remiantis klasikine namų ūkių analizės metodologine prieiga, kai pagrindinės socialinės ir ekonominės charakteristikos priskiriamos vienam namų ūkių reprezentuojančiam nariui. Vis dėlto tolimesnė metodologinė analizė parodė, kad toks požiūris gali nepakankamai atspindėti realią namų ūkio vidinę socialinę struktūrą, ypač tais atvejais, kai vartojimo sprendimai bendrai formuojami kelių ekonomiškai aktyvių ar skirtingoms amžiaus grupėms priklausančių narių. Dėl šios priežasties

⁵ https://docs.google.com/spreadsheets/d/1fZCFo9oPAJB87NzIdXejW_zAcD19W8hM/edit?gid=1438669803#gid=1438669803

⁶ https://docs.google.com/spreadsheets/d/1aQ2EFgKDU_CaCx7WT6Ixa-PzlsBMydf/edit?gid=680109068#gid=680109068

⁷ <https://docs.google.com/spreadsheets/d/16bONK6YxbUF7ZyNciQfMj-p8vKk-095Y/edit?gid=539686697#gid=539686697>

galiausiai pasirinkta agreguota namų ūkio lygmens analizė, leidžianti įvertinti visų namų ūkio narių kolektyvines socialines charakteristikas.

2.2. Namų ūkių duomenų rinkinio paruošimas

Kaip pažymi Hanas, Pei ir Tongas, duomenų paruošimas yra kritinis žinių išgavimo proceso etapas, lemiantis tolimesnių analizės metodų stabilumą ir interpretacinę vertę [45]. Prieš taikant neprižiūrimo mokymosi metodus namų ūkių segmentavimui ir formuojant empiriniu pagrindu identifiktuotas socialines grupes, būtina užtikrinti, kad pradinis duomenų rinkinys būtų statistiškai nuoseklus, informatyvus ir tinkamas daugiamatei analizei. Kadangi šiame darbe socialinės grupės nėra apibrėžiamos iš anksto pagal teorines ar administracines klasifikacijas, bet nustatomos duomenimis grįsto klasterizavimo būdu, pradinių požymių kokybė, reprezentatyvumas ir tarpusavio struktūra laikomi vienais svarbiausių veiksnių, lemiančių tolimesnių grupavimo rezultatų patikimumą, interpretacinį aiškumą ir praktinį pritaikomumą vartojimo išlaidų struktūrų tyrime. Nors neprižiūrimo mokymosi metodai, tokie kaip klasterizavimo algoritmai, leidžia identifikuoti latentines struktūras duomenyse be iš anksto apibrėžto tikslinio kintamojo, jų rezultatai yra itin jautrūs pradiniam duomenims. Tai empiriniuose tyrimuose patvirtina Jainas, kuris nurodo, kad net ir keli nereikšmingi arba pertekliniai požymiai gali reikšmingai paveikti klasterių struktūrą ir sumažinti rezultatų kokybę [46]. Dėl šios priežasties didelės apimties duomenų rinkiniuose dažnai pasitaiko požymių, kurie nesuteikia papildomos informacinės vertės, dubliuoja kitų kintamųjų perduodamą informaciją arba dėl mažo kintamumo tampa statistinio triukšmo šaltiniu. Atsižvelgiant į tai, šiame tyrime atliekamas sistemingas duomenų rinkinio paruošimas, apimantis mažai reprezentatyvių kategorijų identifikavimą bei požymių informatyvumo vertinimą. Atrankos tikslas – sumažinti duomenų matmenis neprarandant esminės struktūrinės informacijos.

Požymių pasiskirstymo analizė. Cai ir kt., aprašydami neprižiūrimo mokymosi požymių atranką, pabrėžia, kad požymių vertinimo rodikliai turi atspindėti kiekvieno požymio gebėjimą diferencijuoti skirtingas klases ar klasterius [47]. Straipsnyje kaip vienas iš paprasčiausių ir efektyviausių neprižiūrimo mokymosi požymių atrankos kriterijų išskiriamas didžiausios dispersijos metodas, kuris iš esmės projektuoja stebinius į didžiausios dispersijos dimensijas. Toks dispersijos kriterijus originaliame daugiaklasterinės požymių atrankos (angl. *Multi-Cluster Feature Selection*, toliau – MCFS) metode taikomas duomenims, kuriuose stebiniai apibrėžiami kaip taškai realiųjų skaičių erdvėje, o požymių kintamumas vertinamas pagal jų išsibarstymą duomenų erdvėje. Bendrinis dispersijos slenksčio metodas grindžiamas prielaida, kad mažą arba beveik nulinę dispersiją turintys kintamieji paprastai nepasižymi reikšminga prognozinė ar diskriminacine verte. Esant n stebinių ir d kintamųjų matricai $X \in R^{n \times d}$, j -ojo kintamojo imties dispersija apskaičiuojama pagal formulę [48]:

$$\sigma_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{x}_j)^2; \quad (2)$$

čia X_{ij} – j -ojo kintamojo reikšmė i -ajame stebinyje; \bar{x}_j – j -ojo kintamojo imties vidurkis; n – stebinių skaičius. Jei apskaičiuota dispersija yra mažesnė už iš anksto nustatytą slenkstį θ ($\sigma_j^2 < \theta$), toks kintamasis gali būti šalinamas iš tolimesnės analizės kaip mažai informatyvus.

Kadangi sudarant skirtingas socialines grupes daugiausia analizuojami kokybiniai požymiai, MCFS taikoma didžiausios dispersijos logika adaptuojama per kategorijų dažnių pasiskirstymo analizę. Tam kiekvienai kategorinei požymio reikšmei apskaičiuojamas santykinis dažnis:

$$f_i = \frac{n_i}{N}; \quad (3)$$

čia n_i – konkrečios kategorijos stebinių skaičius; N – bendras stebinių skaičius.

Kokybinių kintamųjų atveju santykinio dažnio logika siejama su labai mažos dispersijos (angl. *near-zero variance*) principu, pagal kurį mažai informatyviais laikomi kintamieji, kurių unikalių reikšmių dalis imties dydžio atžvilgiu yra maža (pvz., neviršija 10 %), o dažniausiai pasitaikančios reikšmės ir antros pagal dažnį reikšmės santykis yra didelis (pvz., apie 20) [49]. Tokiais atvejais kintamojo pašalinimas iš modelio laikomas metodologiškai pagrįstu.

PhiK koreliacijos koeficientas (ϕ_K). Kintamųjų tarpusavio priklausomumo stiprumo vertinimui empiriniuose tyrimuose plačiai taikomi klasikiniai koreliacijos matai, tokie kaip Pirsono ir Spirmeno koreliacijos koeficientai, skirti tiesinių arba monotoniškų priklausomybių identifikavimui. Tradicinių priklausomybės vertinimo metodų taikymas mišraus tipo duomenims dažnai yra ribotas, todėl literatūroje vis dažniau naudojami universalesni metodai. Baakas ir kt. nurodo, kad ϕ_K metodas gali būti nuosekliai taikomas skaitinių, ranginių ir kategorinių kintamųjų priklausomybėms vertinti, todėl yra tinkamas mišraus tipo duomenų analizei [50]. Autoriai taip pat pabrėžia, kad kiekvienas analizuojamas kintamasis šiame metode iš esmės traktuojamas taip, tarsi jo tipas būtų kategorinis, o priklausomybės stiprumo vertinimas grindžiamas Pirsono χ^2 kontingencijos testo logika, nes ϕ_K koreliacija yra išvesta iš Pirsono χ^2 kontingencijos testo. Pagal šią logiką kontingencinėje lentelėje kiekviena eilutė atitinka vieno kintamojo kategorijas, kiekvienas stulpelis – kito kintamojo kategorijas, o kiekvienas langelis nusako stebinių, vienu metu patenkančių į abi kategorijas, skaičių. Gautos ϕ_K reikšmės, kaip nurodo autoriai, yra apribotos intervale [0; 1], kur 0 reiškia ryšio nebuvimą, o 1 atitinka visišką priklausomybę [50].

Vertinant kintamųjų tarpusavio priklausomumą atsižvelgiama ne tik į apskaičiuotą priklausomumo stiprumą, bet ir į jo statistinį reikšmingumą, kadangi aukšta koreliacijos koeficiento reikšmė nebūtinai reiškia statistiškai patikimą ryšį. Kaip pažymi Baakas ir kt., išvadoms apie kintamųjų priklausomybę formuoti būtina vertinti tiek koreliacijos koeficiento dydį, tiek jo statistinį reikšmingumą [50]. Statistinio reikšmingumo vertinimui autoriai siūlo taikyti hibridinį kintamųjų porų priklausomybės vertinimo metodą, kuris grindžiamas *Monte Karlo* imčių modeliavimu ir χ^2 pasiskirstymo koregavimu. Kintamųjų tarpusavio ryšys laikomas statistiškai reikšmingu tik tais atvejais, kai hibridiniu metodu apskaičiuota p-reikšmė yra mažesnė už slenkstinę reikšmę 0,05, atitinkančią 95 % pasikliautinumo lygį.

2.3. Skirtingų socialinių grupių identifikavimo metodika

Kadangi tyrime analizuojami Lietuvos namų ūkių duomenys apima tiek skaitinius, tiek kategorinius požymius, socialinių grupių identifikavimui reikalingi metodai, gebantys vienu metu įvertinti skirtingų tipų kintamuosius. Siekiant identifikuoti aiškiai apibrėžtas namų ūkių grupes ir kiekvieną stebinį priskirti konkrečiam klasteriui, pasirinkta taikyti atstumais grįstus klasterizavimo metodus. Dėl šios priežasties tyrime taikomi k-medoidų, k-prototipų ir hierarchinio klasterizavimo metodai, siekiant nustatyti Lietuvos namų ūkių duomenų struktūrą geriausiai atspindintį grupavimo sprendinį. Atsižvelgiant į mišrų duomenų pobūdį, stebinių panašumui vertinti pasirinkta Gower atstumų matrica bei jos optimizavimas.

2.3.1. Klasterizavimo metodai

Normalizavimas

Atliekant skirtingų socialinių grupių paiešką neprižiūravimo mokymosi metodais, vienas svarbiausių duomenų paruošimo etapų yra skirtingų matavimo skalių suderinimas. Kaip pažymi Jainas, klasterizavimo algoritmų, paremtų atstumų metrikomis, rezultatai yra tiesiogiai priklausomi nuo pradinių požymių mastelio, todėl duomenų transformavimas prieš grupavimą yra būtina metodologinė procedūra [46]. Kadangi šiame tyrime socialinių grupių identifikavimui nuspręsta išbandyti k-prototipų algoritmą, skirtą mišraus tipo duomenų klasterizavimui, normalizavimas taikomas tik skaitiniams požymiams, o kategoriniai kintamieji paliekami pirminėje formoje. Toks sprendimas grindžiamas algoritmo veikimo principu, kadangi k-prototipų metodas vienu metu naudoja euklidinio atstumo komponentą skaitiniams kintamiesiems ir kategorinių požymių nesutapimų metriką. Originaliame algoritmo aprašyme Huangas eksperimentinėje dalyje nurodo, kad taikant k-prototipų algoritmą skaitiniai požymiai buvo transformuoti į intervalą [0; 1], todėl šiame tyrime taip pat taikomas skaitinių požymių mastelio suvienodinimas [51].

Šiame darbe skaitinių požymių transformavimui taikomas Min–Max normalizavimo metodas (*angl. Min–Max normalization*), leidžiantis kiekvieno kintamojo reikšmes perkelti į bendrą intervalą nuo 0 iki 1. Normalizuota kiekvieno stebėjimo reikšmė apskaičiuojama pagal formulę [52]:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}; \quad (4)$$

čia x_i – pradinė kintamojo reikšmė; x_{\min} ir x_{\max} – atitinkamai minimali ir maksimali to požymio reikšmė imtyje.

Gower artumo matricos sudarymas

Remiantis nagrinėta literatūra, tyrime socialinių grupių identifikavimui taip pat nuspręsta taikyti k-medoidų bei hierarchinį klasterizavimą. Prieš taikant šiuos metodus, būtina suformuoti stebinių tarpusavio panašumo struktūrą, leidžiančią kiekybiškai įvertinti, kiek panašūs arba skirtingi yra analizuojami namų ūkiai. Kadangi šiame tyrime naudojami namų ūkių duomenys yra mišraus tipo ir apima tiek kiekybinius, tiek kokybinius požymius, tradiciniai atstumo matai, tokie kaip Euklido, Čebyševio ar Manhatano atstumai, nėra tinkami, nes jie efektyviai apdoroja tik homogeninius skaitinius požymius. Mišrių duomenų analizėje vienas plačiausiai taikomų metodų yra Gower panašumo matas, leidžiantis vieningoje matematinėje sistemoje įvertinti skirtingų tipų kintamuosius.

Šiame tyrime Gower artumo matas naudojamas kaip tarpinis duomenų transformavimo etapas prieš taikant k-medoidų ir hierarchinio klasterizavimo algoritmus. Abu metodai, skirtingai nei k-prototipų algoritmas, remiasi ne tiesiogine požymių matrica, bet objektų tarpusavio atstumų struktūra. Todėl prieš atliekant klasterizavimą apskaičiuojami poriniai atstumai tarp visų namų ūkių ir sudaroma simetrinė Gower atstumo matrica, kuri vėliau naudojama kaip duomenų reprezentavimo forma grupavimo procedūroje. Gower atstumo matas apskaičiuoja porinį nepanašumą tarp dviejų objektų i ir j , kiekvieną požymį vertinant atskirai ir vėliau agreguojant dalinius nepanašumus į bendrą atstumo koeficientą [53]:

$$D_{ij} = \frac{\sum_{k=1}^p w_k d_{ij}^{(k)}}{\sum_{k=1}^p w_k}; \quad (5)$$

čia D_{ij} – bendras atstumas tarp objektų i ir j ; p – analizuojamų požymių skaičius; w_k – k -ojo požymio svoris; $d_{ij}^{(k)}$ – dalinis nepanašumas pagal k -ąjį požymį. Gauta atstumo reikšmė priklauso intervalui $[0; 1]$, kur reikšmė, artima 0, rodo didelį objektų panašumą, o reikšmė, artima 1 – didelį skirtumą.

Skaitinių požymių atveju dalinis nepanašumas apskaičiuojamas remiantis normalizuotu absoliučiu skirtumu [53]:

$$d_{ij}^{(k)} = \frac{|x_{ik} - x_{jk}|}{R_k}; \quad (6)$$

čia x_{ik}, x_{jk} – dviejų objektų reikšmės; $R_k = \max(x_k) - \min(x_k)$ atitinka k -ojo požymio reikšmių intervalą. Toks normalizavimas leidžia skirtingų mastelių kintamuosius transformuoti į bendrą intervalą ir sumažina neproporcingą didelės dispersijos požymių įtaką bendram atstumo įvertinimui.

Kategorinių požymių atveju dalinis nepanašumas apskaičiuojamas pagal sutapimo principą [53]:

$$d_{ij}^{(k)} = \begin{cases} 0, & x_{ik} = x_{jk} \\ 1, & x_{ik} \neq x_{jk} \end{cases}. \quad (7)$$

Tai reiškia, kad identiškos kategorinės reikšmės laikomos visiškai panašiomis, o nesutampančios – visiškai skirtingomis.

Apskaičiavus visus porinius atstumus tarp n namų ūkių, sudaroma simetrinė Gower atstumo matrica:

$$\mathbf{D} = \begin{bmatrix} 0 & D_{12} & \cdots & D_{1n} \\ D_{21} & 0 & \cdots & D_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1} & D_{n2} & \cdots & 0 \end{bmatrix}. \quad (8)$$

Python programavimo aplinkoje Gower atstumų skaičiavimui naudojama „*gower*“ biblioteka, realizuojanti klasikinę mišrių duomenų Gower formulę. Vis dėlto, naujesni empiriniai tyrimai rodo, kad vienodi visų požymių svoriai ne visada leidžia pasiekti optimalią klasterių tarpusavio atskirtį. Bektasas ir Schumannas parodė, kad optimizuojant Gower svorius galima reikšmingai padidinti klasterių vidinį homogeniškumą bei tarpusavio atskirtį, vertinamą silueto kriterijumi [36]. Remiantis šiuo metodologiniu principu, šiame tyrime papildomai naudojama požymių svorių optimizavimo procedūra, kurios metu kiekvienam požymiui priskiriamas individualus svoris w_k , kuris optimizuojamas taip, kad maksimalizuotų klasterizavimo kokybės rodiklius. Metodo taikymą atspindinti schema pateikiama 1 priede.

K-prototipų klasterizavimo metodika

Skirtingai nei k -vidurkių klasterizavimo metodas, kuris remiasi tik Euklido atstumu ir gali būti taikomas tik skaitiniams duomenims, k -prototipų metodas leidžia vienu metu apdoroti tiek skaitinius, tiek kategorinius požymius vienoje klasterizavimo procedūroje. Tam naudojamas mišrus nepanašumo matas, sujungiantis skaitinių požymių kvadratinį Euklido atstumą ir kategorinių požymių nesutapimo informaciją [51]:

$$d_2(X, Y) = \sum_{j=1}^p (x_j - y_j)^2 + \gamma \sum_{j=p+1}^m \delta(x_j, y_j); \quad (9)$$

čia X – pirmasis objektas; Y – antrasis objektas; p – skaitinių požymių skaičius; m – kategorinių požymių skaičius; x_j ir y_j – lyginamų objektų reikšmės; γ – svorio parametras, subalansuojantis skaitinių ir kategorinių požymių indėlių; $\delta(x_j, y_j)$ – kategorinių reikšmių nesutapimo funkcija:

$$\delta(x_j, y_j) = \begin{cases} 0, & x_j = y_j \\ 1, & x_j \neq y_j \end{cases}. \quad (10)$$

K-prototipų algoritmas minimizuoja bendrą nuostolių funkciją [51]:

$$P(W, Q) = \sum_{l=1}^k \left(\sum_{i=1}^n w_{il} \sum_{j=1}^p (x_{ij} - q_{lj})^2 + \gamma \sum_{i=1}^n w_{il} \sum_{j=p+1}^m \delta(x_{ij}, q_{lj}) \right); \quad (11)$$

čia k – klasterių skaičius; n – stebinių skaičius; w_{il} – indikatorinis kintamasis, įgyjantis reikšmę 1, jei i -asis stebinis priskiriamas l -ajam klasteriui, ir 0 priešingu atveju; p – skaitinių požymių skaičius; m – kategorinių požymių skaičius; x_{ij} – i -ojo stebinio j -ojo požymio reikšmė; q_{lj} – l -ojo klasterio prototipo j -ojo požymio reikšmė; γ – svorio parametras, subalansuojantis skaitinių ir kategorinių požymių įtaką; $\delta(x_{ij}, q_{lj})$ – kategorinių požymių nesutapimo funkcija.

K-prototipų algoritmo veikimas grindžiamas iteraciniu optimizavimo principu. Pirmajame etape inicializuojami k klasterių prototipai, kurių parinkimui gali būti taikomi skirtingi inicializavimo metodai, pavyzdžiui, *Huang* ir *Cao* metodai, siekiant pagerinti algoritmo konvergavimą ir rezultatų stabilumą (detalesnis šių metodų aprašymas pateiktas 2 priede). Antrajame etape kiekvienas stebinis priskiriamas tam klasteriui, kuriam apskaičiuotas mišrus atstumas yra mažiausias. Trečiajame etape klasterių prototipai perskaičiuojami: skaitinių požymių reikšmės atnaujinamos naudojant aritmetinį vidurkį, o kategorinių požymių – modą, t. y. dažniausiai pasitaikančią kategoriją. Šis procesas kartojamas tol, kol objektų priskyrimas klasteriams nebekinta arba nuostolių funkcijos pokytis tampa mažesnis už iš anksto nustatytą slenkstinę reikšmę.

K-medoidų klasterizavimo metodika

Tuo tarpu k-medoidų klasterizavimo metodas, skirtingai nei k-vidurkių algoritmas, kuris klasterio centrą apibrėžia kaip matematinį centroidą, klasterio centru pasirenka realų duomenų rinkinio objektą – medoidą. Medoidas apibrėžiamas kaip objektas, kurio bendras atstumas su kitais to paties klasterio objektais yra mažiausias [54]:

$$m_k = \arg \min_{x_i \in C_k} \sum_{x_j \in C_k} D_{ij}; \quad (12)$$

čia m_k – k -ojo klasterio medoidas; C_k – k -ajam klasteriui priklausančių objektų aibė; D_{ij} – Gower atstumas tarp objektų x_i ir x_j . Bendras algoritmo optimizavimo tikslas – minimizuoti visų objektų atstumų iki jų klasterių medoidų sumą:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} D(x_i, m_k); \quad (13)$$

čia $D(x_i, m_k)$ žymi Gower atstumą tarp objekto x_i ir klasterio medoido m_k .

Kadangi šiame tyrime klasterizavimas atliekamas naudojant iš anksto apskaičiuotą Gower atstumo matricą, algoritmas operuoja ne pradinių požymių erdvėje, o objektų tarpusavio atstumų matrica. Tokiu būdu klasterizavimo procedūra gali būti taikoma mišraus tipo duomenims, nepriklausomai nuo atskirų požymių matavimo skalių.

K-medoidų metodas gali būti realizuojamas skirtingais optimizavimo algoritmais, tarp kurių plačiausiai žinomas yra PAM (angl. *Partitioning Around Medoids*) algoritmas [54]. Skirtingai nei bendrinė k-medoidų metodologija, PAM algoritmas pateikia iteracinę medoidų paieškos procedūrą, paremtą nuosekliu medoidų pakeitimo (angl. *swap*) mechanizmu ir tiesioginiu tikslo funkcijos minimizavimu visoje atstumų matricoje.

Hierarchinio klasterizavimo metodika

Hierarchinis klasterizavimas šiame darbe atliekamas taikant aglomeracinę strategiją, kurios metu pradžioje kiekvienas stebinys laikomas atskiru klasteriu, o vėliau iteratyviai jungiami tarpusavyje panašiausi klasteriai. Vienas svarbiausių hierarchinio klasterizavimo etapų yra klasterių tarpusavio atstumo apskaičiavimo būdo pasirinkimas. Šiame tyrime vertinami trys dažnai mokslinėje literatūroje naudojami aglomeracijos metodai: vienetinės jungties (angl. *single linkage*), pilnosios jungties (angl. *complete linkage*) ir vidutinės jungties (angl. *average linkage*) [55, 56, 57].

Vienetinės jungties metodas klasterių atstumą apibrėžia kaip mažiausią atstumą tarp dviejų skirtingiems klasteriams priklausančių objektų:

$$\Delta(A, B) = \min_{x_i \in A, x_j \in B} D_{ij}; \quad (14)$$

čia A ir B – lyginami klasteriai; D_{ij} – Gower atstumas tarp objektų. Metodas leidžia identifikuoti artimiausias objektų grandines, tačiau gali būti jautrus vadinamajam grandinės efektui, kai formuojami pailgi ir mažiau homogeniški klasteriai.

Pilnosios jungties metodas klasterių atstumą apibrėžia kaip didžiausią atstumą tarp dviejų skirtingų klasterių objektų:

$$\Delta(A, B) = \max_{x_i \in A, x_j \in B} D_{ij}. \quad (15)$$

Metodas dažniausiai formuoja kompaktiškesnius ir geriau atskirtus klasterius, tačiau gali būti jautresnis išskirtinėms reikšmėms.

Vidutinės jungties metodas klasterių tarpusavio atstumą apibrėžia kaip visų porinių atstumų tarp dviejų klasterių objektų aritmetinį vidurkį:

$$\Delta(A, B) = \frac{1}{n_A n_B} \sum_{x_i \in A} \sum_{x_j \in B} D_{ij}; \quad (16)$$

čia n_A ir n_B – klasterių dydžiai. Skirtingai nei vienetinės jungties metodas, kuris remiasi tik artimiausių objektų poromis ir gali formuoti grandinės efektą, arba pilnosios jungties metodas, kuris jautriau reaguoja į pavienes ekstremalias reikšmes, vidutinės jungties metodas vertina visų dviejų lyginamų klasterių narių tarpusavio panašumą. Kiekviename iteraciniame žingsnyje algoritmas identifikuoja dviejų mažiausią vidutinį tarpusavio atstumą turinčių klasterių porą ir juos sujungia į naują klasterį. Po kiekvieno sujungimo atstumai tarp naujai suformuoto klasterio ir likusių klasterių

perskaičiuojami pagal vidutinės jungties kriterijų. Procesas kartojamas tol, kol visi objektai sujungiami į vieną hierarchinę struktūrą.

Hierarchinio klasterizavimo rezultatai vizualizuojami naudojant dendrogramą, kuri leidžia grafiškai stebėti klasterių jungimosi eigą ir identifikuoti natūralius grupavimo lygmenis.

2.3.2. Klasterių skaičiaus nustatymo ir klasterizavimo kokybės vertinimo metodai

Klasterių skaičiaus nustatymas

Vienas svarbiausių neprižiūravimo mokymosi uždavinių yra optimalaus klasterių skaičiaus k nustatymas, kadangi, skirtingai nei prižiūravimo mokymosi uždaviniuose, tikrosios objektų klasės iš anksto nėra žinomos. Kaip pažymi Jainas, netinkamai parinktas klasterių skaičius gali lemti tiek perteklinį grupių fragmentavimą, tiek skirtingų latentinių segmentų sujungimą į pernelyg apibendrintas grupes [46]. Dėl šios priežasties šiame tyrime klasterių skaičiaus parinkimas atliekamas derinant grafinius ir statistinius validavimo metodus.

Taikant k -prototipų ir k -medoidų algoritmus, pradinis galimų klasterių skaičius vertinamas naudojant **alkūnės grafiką** (angl. *Elbow method*), kuris leidžia identifikuoti tašką, kuriame didinant klasterių skaičių klasterių vidinio nevienalytiškumo mažėjimas tampa nebeženklaus. Kadangi skirtingi klasterizavimo algoritmai optimizuoja nevienodas tikslo funkcijas, šiame tyrime alkūnės metodas grindžiamas konkretaus algoritmo minimizuojamos nuostolių funkcijos (angl. *cost function*) kitimo analize. K -prototipų algoritmo atveju vertinama bendros nuostolių funkcijos reikšmė, apimanti skaitinių požymių atstumus ir kategorinių požymių nesutapimus, o k -medoidų atveju – bendras objektų skirtumas klasterių medoidų atžvilgiu. Optimalus klasterių skaičius nustatomas grafike ieškant vadinamojo „alkūnės taško“, kuriame didinant klasterių skaičių tolimesnis nuostolių funkcijos mažėjimas tampa santykinai nedidelis.

K -prototipų ir k -medoidų algoritmuose klasterių skaičiaus pagrįstumui vertinti taikomas ir **Calinski–Harabasz** rodiklis, vertinantis bendrą klasterių struktūrą pagal tarpklasterinės ir vidinės sklaidos santykį. Rodiklis apskaičiuojamas pagal formulę [58]:

$$CH = \frac{SS_B}{SS_W} \times \frac{n - k}{k - 1}; \quad (17)$$

čia SS_B – tarpklasterinė kvadratų suma, nusakanti klasterių tarpusavio išsisklaidymą; SS_W – vidinė klasterių kvadratų suma, apibūdinanti objektų sklaidą klasterių viduje; n – bendras stebėjimų skaičius; k – klasterių skaičius. Didesnės **Calinski–Harabasz** rodiklio reikšmės rodo geriau tarpusavyje atskirtus bei homogeniškesnius klasterius, todėl didžiausią rodiklio reikšmę turintis sprendinys laikomas tinkamiausiu.

Papildomai k -prototipų ir k -medoidų metodams klasterių skaičius nustatomas remiantis silueto koeficientu, kuris vertina individualių objektų priskyrimo kokybę, parodydamas, kiek kiekvienas stebiny yra artimas savo klasteriui, palyginti su kitomis grupėmis. Praktiniam skirtingų klasterių skaičiaus scenarijų palyginimui naudojamas silueto koeficiento grafikas, o tinkamiausiu laikomas sprendinys, pasižymintis didžiausia vidutinio silueto koeficiento reikšme.

Tuo tarpu taikant hierarchinį klasterizavimą klasterių skaičius nustatomas analizuojant dendrogramą ir vertinant klasterių sujungimo lygių pokyčius. Praktikoje didesni vertikalūs atstumai tarp gretimų

klasterių sujungimų interpretuojami kaip galimos natūralios ribos tarp santykinai homogeniškų grupių. Horizontaliai kertant dendrogramą ties didžiausiu sujungimo aukščio šuoliu gaunamas preliminarus klasterių skaičiaus įvertinimas [59].

Klasterizavimo kokybės vertinimas

Šiame tyrime klasterizavimo rezultatų vertinimas atliekamas dviem vienas kitą papildančiais etapais. Pirmajame etape visi suformuoti klasterizavimo sprendiniai vertinami pagal vidinį kokybės kriterijų – silueto koeficientą. Antrajame etape detalesnei analizei pasirenkamas didžiausią vidutinę silueto reikšmę pasiekęs sprendinys, kuriam papildomai atliekamas stabilumo vertinimas taikant pakoreguotą Rando indeksą (angl. *Adjusted Rand Index*, ARI), klasterių atskiriamumo analizė naudojant *XGBoost* klasifikatorių bei klasterių profiliavimas.

Pagrindinis šiame tyrime taikomas vidinis klasterizavimo kokybės vertinimo rodiklis yra **silueto koeficientas**. Šis rodiklis pasirinktas dėl galimybės vienu metu įvertinti du esminius klasterizavimo kokybės aspektus – vidinį homogeniškumą ir tarpklasterinę atskirtį. Silueto metodas leidžia objektyviai įvertinti, kiek kiekvienas objektas yra artimas savo klasteriui ir nutolęs nuo kitų klasterių, todėl išlieka vienu patikimiausių klasterizavimo validavimo metodų [60].

Silueto koeficientas kiekvienam objektui i apskaičiuojamas pagal formulę:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)}; \quad (18)$$

čia a_i – vidutinis objekto i atstumas iki kitų to paties klasterio objektų; b_i – mažiausias vidutinis objekto i atstumas iki artimiausio kito klasterio. Silueto koeficiento reikšmės priklauso intervalui $[-1; 1]$. Reikšmės, artimos 1, rodo aiškų objekto priskyrimą konkrečiam klasteriui, reikšmės, artimos 0, signalizuoja ribinę padėtį tarp grupių, o neigiamos reikšmės gali rodyti netinkamą priskyrimą [60].

Bendra klasterizavimo kokybė vertinama apskaičiuojant vidutinį silueto koeficientą:

$$\bar{S} = \frac{1}{n} \sum_{i=1}^n S_i; \quad (19)$$

čia n – bendras stebinių skaičius; S_i – individualus silueto koeficientas. Didžiausia vidutinė silueto reikšmė interpretuojama kaip optimalus kompromisas tarp grupių vidinio vientisumo ir jų tarpusavio skirtumų, todėl naudojama optimaliam klasterių skaičiui bei tinkamiausiam klasterizavimo metodui parinkti.

Antrajame kokybės vertinimo etape atliekamas detalesnis geriausio identifikuoto sprendinio vertinimas. Klasterių stabilumas analizuojamas taikant **ARI**, kuris įvertina, kiek nuosekliai sutampa klasterių struktūros, gautos pakartotinai atliekant klasterizavimą skirtingose duomenų poimtyse. Skirtingai nei paprasti panašumo rodikliai, ARI įvertina atsiktinio sutapimo tikimybę, todėl leidžia objektyviau nustatyti klasterių stabilumą. Šiame darbe naudojama *Python* bibliotekoje realizuota pakoreguoto Rando indekso skaičiavimo funkcija *adjusted_rand_score()*, paremta Huberto ir Arabie pasiūlyta korekcijos schema [61].

Papildomai klasterių atskiriamumas analizuojamas taikant Chen ir Guestrin pasiūlytą **XGBoost** klasifikatorių [62]. Šiame tyrime klasterių etiketės naudojamos kaip tikslinis kintamasis, o namų ūkių charakteristikos – kaip aiškinamieji požymiai. Aukštas klasifikavimo tikslumas leidžia daryti

prielaidą apie aiškesnę grupių struktūrą ir jų tarpusavio atskirtį. XGBoost šiuo atveju taikomas ne kaip alternatyvus klasterizavimo metodas, bet kaip išorinė validavimo priemonė.

Statistiniai kokybės rodikliai apibūdina klasterių struktūrinės savybės, tačiau neatskleidžia jų socialinės ir ekonominės interpretacijos. Dėl šios priežasties papildomai atliekamas klasterių **profilavimas**, kurio metu analizuojamos kiekvienos grupės centrinės tendencijos ir pasiskirstymo charakteristikos. Skaitiniams požymiams apskaičiuojami aritmetiniai vidurkiai, o kategorinių požymių interpretacijai naudojami santykiniai dažniai.

2.4. Socialinių grupių vartojimo išlaidų struktūrų palyginimo metodai

Vartojimo išlaidų struktūrų skirtumų analizei šiame tyrime pasirinkta taikyti statistinio palyginimo metodus bei entropija grindžiamus metodus, skirtus vidinių vartojimo struktūrų neapibrėžtumo ir informacinio ryšio vertinimui. Engelio modeliai bei mašininio mokymosi metodai tyrime netaikomi dėl duomenų struktūros apribojimų ir riboto stebėjimų skaičiaus atskirose grupėse.

2.4.1. Vartojimo išlaidų struktūrų statistinio palyginimo metodai

Pirminiam grupių palyginimui tikslinga taikyti daugiamatę vienfaktorinę dispersinę analizę (MANOVA), o identifikavus bendrus daugiamatės struktūros skirtumus – pereiti prie atskirų vartojimo kategorijų vienmatės vienfaktorinės analizės, taikant ANOVA arba, esant parametrinių prielaidų pažeidimams, neparametrinius alternatyvius metodus. Kadangi (M)ANOVA taikymas grindžiamas parametrinių prielaidų tenkinimu, išsamus šių prielaidų aprašymas pateikiamas 3 priede.

Daugiamatė vienfaktorinę dispersinę analizę (MANOVA)

Daugiamatė dispersinė analizė (angl. *Multivariate Analysis of Variance*, MANOVA) yra parametrinis statistinis metodas, skirtas įvertinti, ar kelios nepriklausomos grupės statistiškai reikšmingai skiriasi pagal kelių tarpusavyje koreliuojančių priklausomų kintamųjų visumą [63]. MANOVA metodu tikrinamos dvi statistinės hipotezės:

- **Nulinė hipotezė H_0** : visų tiriamų grupių priklausomų kintamųjų vidurkių vektoriai yra lygūs;
- **Alternatyvioji hipotezė H_A** : bent vienos grupės vidurkių vektorius statistiškai reikšmingai skiriasi.

Vienmatė vienfaktorinę dispersinę analizę (ANOVA)

Identifikavus bendrus daugiamatės struktūros skirtumus, kiekvienos vartojimo kategorijos skirtumai tarp socialinių grupių papildomai vertinami taikant vienmatę dispersinę analizę (angl. *Analysis of Variance*, ANOVA). ANOVA metodu tikrinamos šios hipotezės:

- H_0 : visų grupių vidurkiai yra lygūs;
- H_A : bent vienos grupės vidurkis skiriasi.

Neparametrinė alternatyva – Kruskalo–Voliso kriterijus

Jeigu bent viena iš esminių parametrinių metodų prielaidų nėra tenkinama ir duomenų transformacijos neleidžia pasiekti reikiamo pasiskirstymo, grupių skirtumų analizei gali būti taikomas *Kruskalo–Voliso* kriterijus (angl. *Kruskal–Wallis test*), kuris laikomas neparametrine vienfaktorinės

ANOVA alternatyva [64]. Šis metodas pagrįstas rangų analize ir nereikalauja normalumo prielaidos. Tikrinamos hipotezės:

- H_0 : visų k populiacijų skirstiniai yra vienodi;
- H_A : bent vienos populiacijos skirstinys skiriasi.

Kadangi *Kruskalo–Voliso* kriterijus interpretuojamas kaip medianų skirtumų testas tik tuo atveju, kai lyginamų grupių skirstinių formos yra panašios, prieš jo taikymą papildomai vertinamas skirstinių homogeniškumas.

Skirstinių homogeniškumo vertinimas

Prieš interpretuojant *Kruskalo–Voliso* kriterijaus rezultatus kaip medianų skirtumus, papildomai vertinama, ar lyginamų grupių empiriniai skirstiniai pasižymi panašia forma. Tam taikomas **Andersono–Darlingo kriterijus**, leidžiantis identifikuoti skirtumus tarp skirstinių, ypač jautriai vertinant skirstinių uodegas. Tikrinamos šios statistinės hipotezės:

- H_0 : visų grupių skirstiniai yra vienodi;
- H_A : bent vienos grupės skirstinys skiriasi.

Jeigu Andersono–Darlingo kriterijus parodo statistiškai reikšmingus skirstinių skirtumus, *Kruskalo–Voliso* kriterijaus rezultatų interpretacija kaip centrinių tendencijų skirtumų tampa metodologiškai nepagrįsta.

Andersono–Darlingo k -imčių bendroji statistikos formulė [65]:

$$A_{kN}^2 = \sum_{i=1}^k n_i \int_{B_N} \frac{\{F_{in}(x) - H_N(x)\}^2}{H_N(x)\{1 - H_N(x)\}} dH_N(x); \quad (20)$$

čia n_i – i -osios imties dydis; $F_{in}(x)$ – i -osios imties empirinė skirstinio funkcija; $H_N(x)$ – jungtinės imties empirinė skirstinio funkcija; o $B_N = \{x \in R: H_N(x) < 1\}$.

Nustačius statistiškai reikšmingus skirtumus tarp grupių, gali būti atliekama porinė socialinių grupių analizė. Dviejų imčių Andersono–Darlingo testas yra tiesioginis bendrosios k -imčių statistikos specialusis atvejis. Kai lyginamos dvi nepriklausomos imtys X_1, \dots, X_m ir Y_1, \dots, Y_n , kurių empirinės skirstinio funkcijos žymimos $F_m(x)$ ir $G_n(x)$, dviejų imčių statistika įgauna tokią formą:

$$A_{mn}^2 = \frac{mn}{N} \int_{-\infty}^{\infty} \frac{\{F_m(x) - G_n(x)\}^2}{H_N(x)\{1 - H_N(x)\}} dH_N(x); \quad (21)$$

čia m ir n – lyginamų imčių dydžiai; $N = (m + n)$ – bendras imties dydis; $F_m(x)$ – pirmosios imties empirinė skirstinio funkcija; $G_n(x)$ – antrosios imties empirinė skirstinio funkcija; $H_N(x)$ – jungtinė empirinė skirstinio funkcija.

Papildomai porinių skirtumų vertinimui taikomas Kolmogorovo–Smirnov dviejų imčių testas. Testo statistika apibrėžiama kaip didžiausias vertikalus skirtumas tarp dviejų empirinių skirstinio funkcijų [66]:

$$D = \sup_x |F_m(x) - G_n(x)|; \quad (22)$$

čia sup – didžiausia funkcijų skirtumo reikšmė visame stebėjimų intervale; $F_m(x)$ ir $G_n(x)$ – lyginamų imčių empirinės skirstinio funkcijos.

2.4.2. Entropija grindžiamo ryšio metodas

Ryšio vertinimui taikomos trys viena kitą papildančios metrikos: Šenono entropija, tarpusavio informacija (angl. *Mutual Information*, MI) bei bendroji tarpusavio informacija (angl. *Total correlation*, TC).

Šenono entropija apibrėžia atsitiktinio kintamojo neapibrėžtumą arba informacinį turinį. Diskrečiuoju atveju Šenono entropija apibrėžiama kaip:

$$H(X) = - \sum_{x \in X} p(x) \log p(x); \quad (23)$$

čia X – atsitiktinis kintamasis; $p(x)$ – tikimybė stebėti reikšmę x .

Kadangi absoliuti entropijos reikšmė priklauso nuo galimų būsenų skaičiaus, skirtingų požymių palyginamumui šiame tyrime taikoma normalizuota Šenono entropija, apskaičiuojama dalijant stebėtą entropijos reikšmę iš maksimalios galimos entropijos [67]:

$$NH(X) = \frac{H(X)}{\log_2(B)}; \quad (24)$$

čia B – histograminio diskretizavimo intervalų skaičius. Normalizuotos entropijos reikšmės priklauso intervalui $[0; 1]$, kur $NH(X) = 0$ reiškia visiškai deterministinį kintamąjį, o $NH(X) = 1$ atitinka maksimalų neapibrėžtumą.

Šenono entropija leidžia įvertinti individualaus kintamojo informacinį turinį, tačiau kintamųjų tarpusavio priklausomybei vertinti būtina nustatyti, kiek informacijos apie vieną kintamąjį suteikia kito kintamojo stebėjimas. Šiam tikslui taikoma tarpusavio informacijos metrika:

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}; \quad (25)$$

čia $p(x, y)$ – jungtinė tikimybė; $p(x)$, $p(y)$ – marginalinės tikimybės. Kai $I(X; Y) = 0$, kintamieji yra statistškai nepriklausomi. Didesnė MI reikšmė rodo stipresnę informacinę ryšį tarp analizuojamų kintamųjų.

MI absoliuti reikšmė priklauso nuo analizuojamų kintamųjų entropijos, todėl tarpusavio priklausomybių palyginamumui šiame tyrime taikoma normalizuota tarpusavio informacija, naudojant aritmetinio vidurkio normalizaciją [68]:

$$NMI = \frac{2I(X; Y)}{H(X) + H(Y)}. \quad (26)$$

Normalizuotos tarpusavio informacijos reikšmės priklauso intervalui $[0; 1]$, kur $NMI = 0$ reiškia visišką statistinę nepriklausomybę, o $NMI = 1$ atitinka visišką deterministinę priklausomybę.

Kadangi socialinių grupių vartojimo struktūra gali apimti ne tik porinius, bet ir kelių vienu metu sąveikaujančių vartojimo kategorijų ryšius, papildomai taikoma daugiamatė tarpusavio informacija (TC) apibrėžiama kaip:

$$TC(X_1, \dots, X_n) = \sum_{i=1}^n H(X_i) - H(X_1, \dots, X_n); \quad (27)$$

čia $H(X_i)$ – atskiro kintamojo entropija; $H(X_1, \dots, X_n)$ – jungtinė visų kintamųjų entropija. Kai $TC = 0$ visi kintamieji yra tarpusavyje nepriklausomi. Didesnė TC reikšmė rodo stipresnę bendrą priklausomybės struktūrą tarp analizuojamų kintamųjų.

Kadangi TC reikšmė didėja kartu su kintamųjų skaičiumi ir jų entropija, skirtingų daugiamačių sistemų palyginamumui šiame tyrime taikoma normalizuota TC:

$$NTC = \frac{TC(X_1, \dots, X_n)}{\sum_{i=1}^n H(X_i)}. \quad (28)$$

NTC reikšmė priklauso intervalui $[0; 1]$, kur $NTC = 0$ – visiška daugiamatė nepriklausomybė; $NTC \rightarrow 1$ egzistuojanti kolektyvinė priklausomybė tarp visų sistemos kintamųjų.

Koncepcijos eksperimentinė validacija

Prieš taikant entropija grįstą metodiką realioms klasterizuotiems namų ūkių duomenims, šiame tyrime atliekamas koncepcijos patvirtinimo eksperimentas (angl. *proof of concept*), kurio tikslas – empiriškai patikrinti, ar pasirinktos metrikos geba identifikuoti skirtingo stiprumo ir formos priklausomybes tarp kintamųjų. Siekiant įvertinti metodikos jautrumą skirtingoms duomenų struktūroms, bazinio kintamojo skirstinys formuojamas naudojant tiek normalųjį, tiek eksponentinį skirstinį.

Eksperimento pagrindui generuojamas bazinis atsitiktinis kintamasis

$$X \sim \begin{cases} N(0; 0,3), & \text{kai analizuojamas simetrinio pobūdžio scenarijus} \\ \text{Exp}(\lambda = 1), & \text{kai analizuojamas asimetrinio pobūdžio scenarijus} \end{cases}; \quad (29)$$

bei atsitiktinis triukšmo komponentas, naudojamas priklausomybės scenarijų sudarymui

$$\varepsilon \sim N(0; 0,3). \quad (30)$$

Tuomet penki sintetiniai scenarijai formuojami kaip deterministinės, nepriklausomos, tiesinės, netiesinės ir silpnos priklausomybės scenarijai:

$$Y = \begin{cases} X, & \text{identiškas scenarijus} \\ Y^*, \quad Y^* \perp X, & \text{nepriklausomas} \\ X + \varepsilon, & \text{tiesinis} \\ X^2 + \varepsilon, & \text{netiesinis} \\ 0,3X + 0,7\varepsilon, & \text{silpnas} \end{cases} \quad (31)$$

Ši eksperimentinė schema skirta empiriškai įvertinti, kaip NH, NMI ir NTC reaguoja į skirtingo stiprumo bei struktūros priklausomybes. Tik po eksperimentinės validacijos pasirinkta metodika taikoma realių namų ūkių klasterių analizei, siekiant nustatyti, ar tarp identifikuotų socialinių grupių egzistuoja ne tik statistiškai reikšmingi vartojimo skirtumai, bet ir daugiamačiai informaciniai ryšiai, atspindintys vartojimo elgsenos struktūrinius skirtumus.

2.5. Programinė įranga ir naudojamos bibliotekos

Šiame darbe visa empirinė analizė ir skaičiavimų realizacija buvo vykdoma naudojant *Python* programavimo kalbą. Naudojama virtualios KTU sistemos⁸ aplinka „*PyTorch*“. 3 lentelėje pateikiamos pagrindinės tyrime naudotos *Python* bibliotekos, nurodoma ir jų paskirtis.

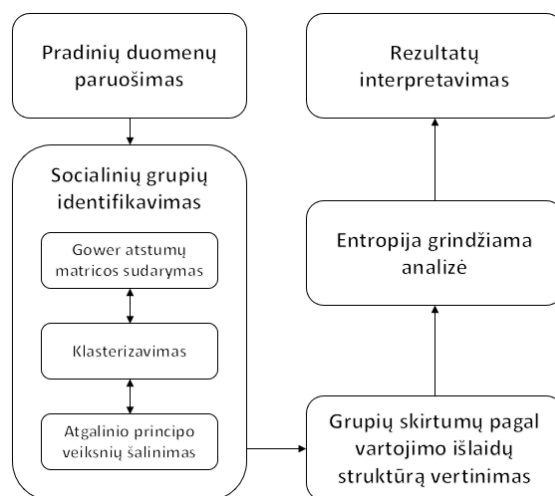
3 lentelė. *Python* bibliotekos ir jų paskirtis

Biblioteka	Paskirtis tyrime
<i>Requests</i>	API užklausų vykdymas ir Lietuvos namų ūkių biudžeto tyrimo duomenų nuskaitymas
<i>Pandas</i>	Namų ūkių duomenų struktūrizavimas, jungimas ir agregavimas
<i>NumPy</i>	Skaitiniai skaičiavimai ir transformacijos
<i>PhiK</i>	Mišraus tipo namų ūkių požymių tarpusavio ryšių analizė
<i>Gower</i>	Mišraus tipo namų ūkių duomenų Gower atstumo matricos skaičiavimas
<i>Kmodes</i>	K-prototipų klasterizavimo realizacija socialinių grupių identifikavimui
<i>Scikit-learn-extra</i>	K-medoidų klasterizavimo realizacija socialinių grupių identifikavimui
<i>Scipy.cluster.hierarchy</i>	Hierarchinio klasterizavimo ir dendrogramų sudarymas socialinių grupių identifikavimui
<i>Scikit-learn</i>	Duomenų transformavimas, klasterių kokybės vertinimas, kryžminė validacija
<i>Matplotlib</i>	Klasterizavimo rezultatų ir statistinių rodiklių vizualizavimas
<i>XGBoost</i>	Identifikuotų socialinių grupių atskiriamumo validavimas
<i>Seaborn</i>	Klasterių profilių sudarymas
<i>SciPy</i>	Statistinių testų vykdymas vartojimo išlaidų skirtumų tarp socialinių grupių analizei
<i>Statsmodels</i>	Dispersinės analizės ir parametrinių prielaidų vertinimas

⁸ <https://ai-notebook.ktu.edu/>

3. Skirtingų socialinių grupių vartojimo išlaidų struktūrų tyrimas

Visa tyrimo eiga gali būti apibrėžiama keliais pagrindiniais etapais – nuo duomenų paruošimo iki gautų rezultatų analizės ir interpretavimo (žr. 5 pav.).



5 pav. Tyrimo eigos schema

3.1. Duomenų šaltinių aprašomoji analizė bei apjungimas

Atsižvelgiant į tyrimo metodologijoje apibrėžtą duomenų struktūrą, tyrimas grindžiamas trimis tarpusavyje susietais 2021 m. namų ūkių biudžetų statistinio tyrimo duomenų šaltiniais, apimančiais bendruosius namų ūkių rodiklius, individualių namų ūkių narių charakteristikas bei vartojimo išlaidų informaciją. 4–7 prieduose pateikiama duomenų aprašomoji statistika.

Namų ūkių lygmens rodiklių aprašomoji statistika (žr. 4 priedą) rodo, kad tiriamų duomenų struktūra pasižymi reikšmingu socialiniu ir ekonominiu heterogeniškumu. Analizuojamas duomenų rinkinys apima 16 kintamųjų, aprašančių namų ūkių charakteristikas, naudojamų socialinių grupių identifikavimo procese. Šiame rinkinyje vienintelis tolydus kintamasis yra namų ūkio užimamas naudingasis plotas. Jo vidutinė reikšmė rodo, kad tipinis namų ūkis užima maždaug 71 m² gyvenamojo ploto, tačiau tarp namų ūkių matyti ryškūs skirtumai. Dalis jų gyvena labai kompaktiškuose būstuose, kur plotas vos viršija minimalias gyvenimo sąlygas ir yra artimas keliolikai kvadratinių metrų, o kita dalis gyvena itin erdviuose patalpose, kurių plotas siekia iki 400 m². Šis išsibarstymas tarp stebimų namų ūkių rodo didelius gyvenimo sąlygų skirtumus ir leidžia šį rodiklį vertinti kaip vieną iš potencialiai informatyviausių struktūrinių požymių, galinčių stipriai prisidėti prie kokybiškų klasterių formavimosi. Aukštą informacinę vertę rodo ir pajamų kvintilinių grupių kintamasis, kurio kategorijų pasiskirstymas yra santykinai tolygus – nuo 16,2 % iki 25,0 %, todėl jis sudaro palankias prielaidas patikimai ekonominės padėties diferenciacijai. Reikšmingi struktūriniai skirtumai taip pat stebimi gyvenamosios vietos, namų ūkio tipo, būsto tipo, namų ūkio galvos amžiaus bei išsilavinimo rodikliuose, kurių kategoriniai pasiskirstymai nėra koncentruoti vienoje dominuojančioje srityje. Pastebima, kad dalis namų ūkių lygmens kintamųjų pasižymi reikšmingu kategoriniu disbalansu, kuris klasterizavimo kontekste gali riboti jų informacinį svorį. Ryškiausias pavyzdys – būsto valdos statusas, kuriame 88,3 % stebimų priskiriami būsto savininkams be finansinių įsipareigojimų kategorijai. Panaši koncentracija stebima išlaikomų vaikų skaičiaus rodiklyje, kur 80,9 % namų ūkių neturi išlaikomų vaikų, bei centralizuoto šalto vandens tiekimo kintamajame, kuriame 79,8 % stebimų priklauso tai pačiai kategorijai. Tokie pasiskirstymai

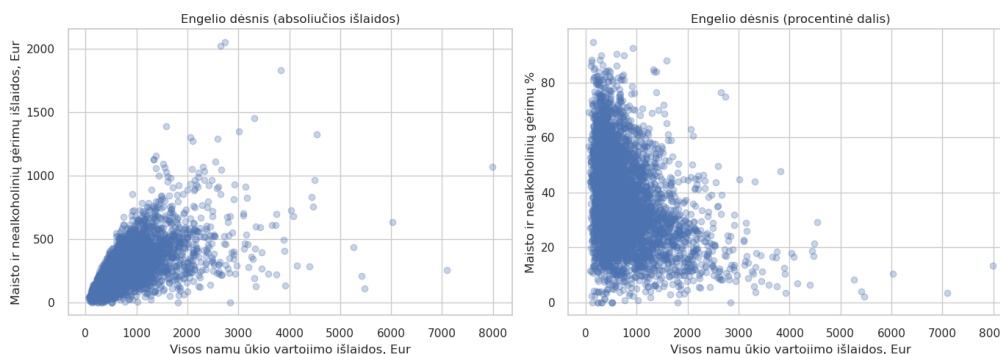
metodologiniu požiūriu mažina kintamųjų gebėjimą formuoti skirtingus klasterius, todėl jų įtraukimas į klasterizavimo modelį yra vertinamas atsargiai, ypač siekiant išvengti požymių, kurie skirti apibūdinti bendrą imties struktūrą, bet ne tarpgrupinius skirtumus.

Papildomą analitinę vertę suteikia individualių namų ūkių narių lygmens duomenys, leidžiantys išplėsti namų ūkio socialinio profilio sampratą papildomomis charakteristikomis. Svarbu paminėti, kad šiame duomenų rinkinyje pagrindinę profesinę veiklą apibūdinančiame kintamajame yra trūkstamų reikšmių, todėl jos priskiriamos jau esamai kategorijai „Nenurodė“, taip išsaugant visų stebėjimų informacinę vertę ir užtikrinant duomenų struktūros vientisumą. Individualaus lygmens duomenų aprašomoji analizė pateikiama 5 priede, tačiau patys individualūs stebėjimai tiesiogiai klasterinėje analizėje nenaudojami – jie pasitelkiami agreguotų namų ūkių rodiklių formavimui. Esminiu analitiniu privalumu laikytina galimybė naudojant tik mikrolygmens duomenis, konstruoti agreguotus rodiklius, atspindinčius namų ūkių socialinę ir ekonominę struktūrą. Tokiu būdu mikrolygmens charakteristikos transformuojamos į sudėtinius rodiklius, tokius kaip: vyrų dalis namų ūkyje, vidutinis namų ūkio narių amžius, dirbančių asmenų dalis namų ūkyje, aukštąjį išsilavinimą turinčių asmenų dalis, aukštos kvalifikacijos užimtųjų dalis bei bendras amžiaus priklausomybės rodiklis. Šių rodiklių sudarymas yra metodologiškai pagrįstas, kadangi rodikliai gali būti interpretuojami kaip vartojimo išlaidų skirtumus tarp namų ūkių lemiantys veiksniai (žr. 1.3 skyrelį). Agreguotų rodiklių aprašomoji analizė pateikiama 6 priede.

Apžvelgiami ir vartojimo išlaidų duomenys, kurie šiame tyrime nėra tiesiogiai įtraukiami į skirtingų socialinių grupių formavimo procesą, tačiau naudojami jau identifikuotų klasterių ekonominei interpretacijai ir tarpusavio struktūrinių skirtumų vertinimui. Duomenys apima bendrųjų vartojimo išlaidų kintamąjį bei 13 vartojimo išlaidų kategorijų. Namų ūkių vartojimo išlaidų aprašomoji statistika (žr. 7 priedą) rodo, kad tipinio namų ūkio bendros mėnesinės vartojimo išlaidos siekia apie 793 eurus, tačiau ši reikšmė, dėl didelės duomenų sklaidos, tik iš dalies atspindi tipinę vartojimo situaciją. Tai patvirtina ir gana aukštas standartinis nuokrypis (615,55 Eur), rodantis ryškius skirtumus tarp namų ūkių vartojimo galimybių bei turimų finansinių išteklių. Tipiškesnę vartojimo lygį apibūdina medianinė reikšmė, siekianti 630,50 Eur, kuri parodo, kad pusė tiriamų namų ūkių per mėnesį išleidžia mažiau nei ši suma. Papildomą ekonominio skirtumo mastą atskleidžia ir itin platus stebimų reikšmių intervalas, kuris prasideda nuo 56,23 Eur per mėnesį per mėnesį ir baigiasi ties 7984,83 Eur per mėnesį. Toks beveik 143 kartus besiskiriantis vartojimo išlaidų diapazonas rodo, kad imtyje vienu metu egzistuoja tiek itin riboto vartojimo, tiek labai aukšto vartojimo lygio namų ūkiai. Bendra vartojimo struktūros analizė rodo, kad didžiausią dalį namų ūkių mėnesinio biudžeto sudaro išlaidos maistui ir nealkoholiniams gėrimams. Vidutiniškai šiai vartojimo kategorijai namų ūkiai per mėnesį skiria apie 258 eurus, o tipinis vartojimo lygis siekia maždaug 218 eurų. Tuo pačiu stebimi gana ryškūs skirtumai tarp atskirų namų ūkių: dalis jų šiai būtinųjų poreikių kategorijai skiria santykinai nedideles sumas, tuo tarpu kitų namų ūkių išlaidos viršija du tūkstančius eurų per mėnesį. Toks platus išlaidų diapazonas tarp tiriamų namų ūkių, atskleidžia skirtingą ekonominį pajėgumą bei nevienodus vartojimo prioritetus. Kadangi maisto išlaidos sudaro didžiausią absoliučią vartojimo struktūros dalį, ši kategorija yra laikoma vienu svarbiausių empirinių rodiklių, leidžiančių vertinti namų ūkių materialinę padėtį ir interpretuoti vartojimo elgsenos skirtumus tarp identifikuotų socialinių grupių. Reikšmingą dalį bendrame vartojimo biudžete taip pat sudaro būsto, vandens, elektros, dujų ir kito kuro išlaidos, kurių vidurkis siekia 130,52 Eur, bei transporto išlaidos, kurių vidutinė reikšmė sudaro 87,64 Eur, tačiau šioje kategorijoje stebima itin didelė variacija ($\sigma = 242,03$ Eur) bei maksimali reikšmė, kuri siekia 5291,96 Eur. Pastebima, kad dalyje vartojimo kategorijų,

tokių kaip švietimo paslaugos, restoranai ir apgyvendinimo paslaugos, apranga ir avalynė ar transportas, medianos yra nulinės, o tai rodo, jog didelė dalis namų ūkių analizuojamu laikotarpiu šiose kategorijose išlaidų nepatyrė arba nefiksavo. Šie rezultatai nestebina, kadangi toks vartojimo išlaidų struktūros pasiskirstymas yra nuoseklus 2021 m. COVID-19 pandemijos laikotarpiui būdingoms ekonominio neapibrėžtumo sąlygoms, kai namų ūkiai demonstravo polinkį riboti nebūtinašias vartojimo išlaidas ir prioritetizuoti būtinųjų poreikių finansavimą.

Kadangi maisto ir nealkoholinių gėrimų išlaidos sudaro didžiausią vartojimo struktūros dalį, papildomai atliekama Engelio dėsnio empirinė patikra, vertinant ryšį tarp bendrų namų ūkių vartojimo išlaidų ir išlaidų maistui bei nealkoholiniams gėrimams (žr. 6 pav.). Gauti rezultatai rodo, kad didėjant bendroms vartojimo išlaidoms išlaidos maistui didėja santykinai lėčiau, todėl jų dalis bendrame namų ūkio biudžete nuosekliai mažėja. Mažesnes vartojimo išlaidas turinčiuose namų ūkiuose maisto išlaidų dalis kai kuriais atvejais siekia 60–85 % visų vartojimo išlaidų, o tai gali būti siejama su didesne disponuojamų finansinių išteklių dalimi, skiriama būtiniausių poreikių tenkinimui, bei atspindėti žemesnę materialinės gerovės lygį ir didesnę socialinę pažeidžiamumą. Gauti rezultatai atitinka Engelio dėsnio teorines prielaidas ir patvirtina, kad vartojimo struktūra gali būti interpretuojama kaip papildomas namų ūkių ekonominės gerovės indikatorius.



6 pav. Engelio dėsnio empirinio patikrinimo grafikas

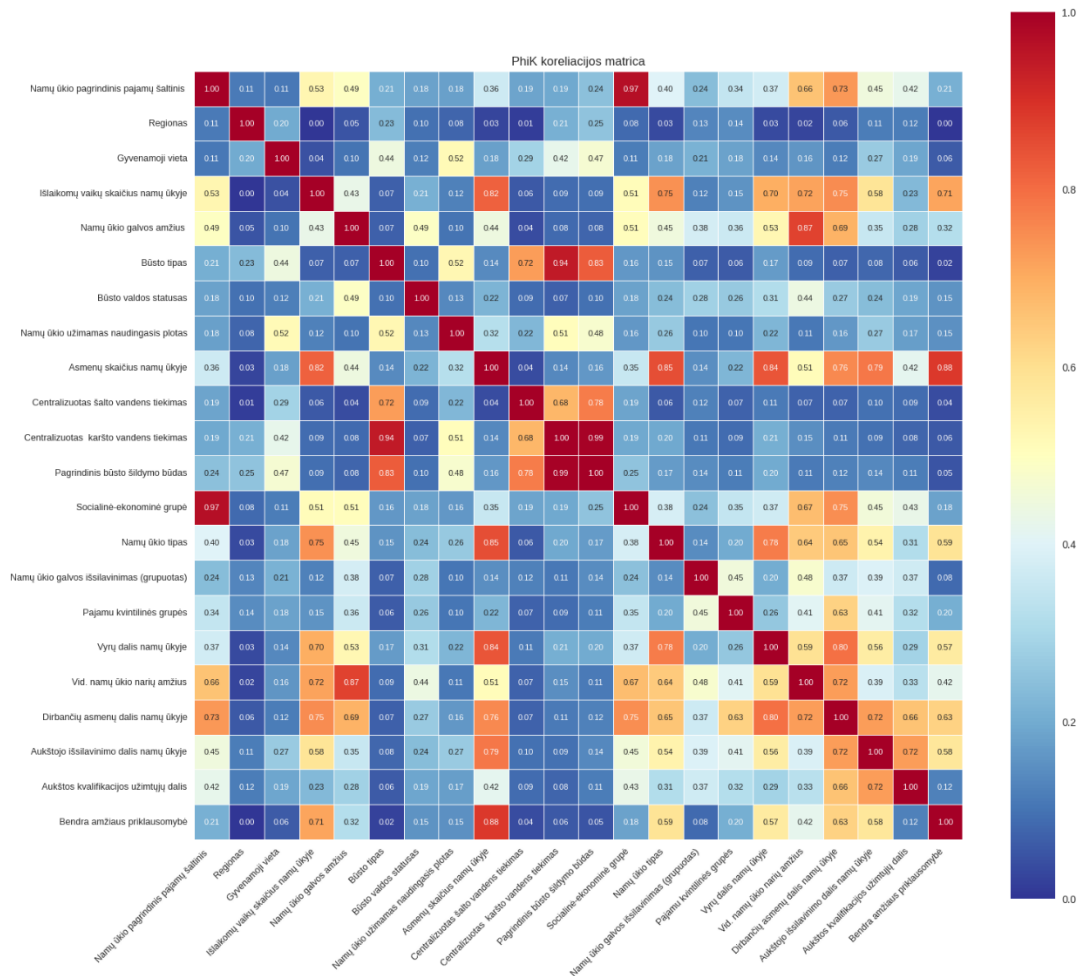
Atsižvelgiant į tai, kad vartojimo išlaidų rodikliai šiame tyrime laikytini ne klasterių formavimo, bet jų ekonominio profiliavimo ir tarpusavio lyginimo priemone, pradiniam klasterizavimo požymių atrankos etapui sudaromas 22 kintamųjų rinkinys, apimantis namų ūkių charakteristikas.

3.2. Požymių atranka

Atlikus kiekvieno kintamojo informatyvumo ir tinkamumo tolimesnei analizei vertinimą, siekta identifikuoti mažo kintamumo požymius, kurie gali turėti ribotą vertę formuojant skirtingas socialines grupes. Skaičiavimų rezultatai pateikti 8 priede. Kategorinių požymių atveju informatyvumas vertintas pagal dominuojančios reikšmės dažnį, o skaitinių – pagal jų dispersiją. Vadovaujantis literatūroje dažnai taikomu beveik nulinio kintamumo principu, mažo kintamumo kategoriniais požymiais laikyti tie kintamieji, kuriuose dažniausiai pasitaikanti reikšmė sudaro daugiau kaip 90 % visų stebinių. Gauti rezultatai rodo, kad nė vienas iš analizuojamų požymių nepasiekė nustatytų šalinimo kriterijų. Didžiausia dominuojančios kategorijos dalis nustatyta būsto valdos statuso kintamajame (88,32 %), tačiau ši reikšmė išliko mažesnė už nustatytą slenkstį. Tuo tarpu skaitinių požymių analizė parodė pakankamą kintamumą visiems rodikliams – didžiausia dispersija nustatyta namų ūkio užimamo naudingo ploto ($\sigma^2 = 1426,545$) ir vidutinio namų ūkio narių amžiaus ($\sigma^2 = 347,769$) kintamuosiuose. Gauti rezultatai leidžia teigti, kad visi analizuojami požymiai išlaiko

pakankamą heterogeniškumą ir gali prisidėti prie tolimesnio namų ūkių skirstymo klasterizavimo procese.

Siekiant įvertinti galimą informacinį kintamųjų persidengimą bei identifikuoti požymius, galinčius daryti neproporcingą įtaką klasterių formavimosi procesui, toliau atliekama ϕ_K koreliacinė analizė. Koreliacijos koeficientų matrica pateikiama 7 paveiksle. Kintamųjų atranka šiame etape grindžiama ne tik statistiniu informacijos persidengimu, bet ir tyrimo tikslu – identifikuoti skirtingas socialines grupes, pasižyminčias nevienoda ekonomine padėtimi, gyvenimo sąlygomis ir vartojimo elgsena. Gauti rezultatai rodo, kad dalis analizuojamų kintamųjų pasižymi itin stipriomis tarpusavio priklausomybėmis, todėl tolimesnėje analizėje tikslinga dalį požymių eliminuoti.



7 pav. ϕ_K koreliacijos koeficientų matrica

Itin stiprus ryšys nustatytas tarp namų ūkio pagrindinio pajamų šaltinio ir socialinės-ekonominės grupės kintamųjų, kur ϕ_K koeficientas siekia 0,97. Toks ryšio stiprumas rodo beveik pilną informacinį persidengimą tarp šių rodiklių, o pats ryšys yra savaime suprantamas, kadangi abiejų kintamųjų pagrindas tiesiogiai susijęs su asmens ar namų ūkio padėtimi darbo rinkoje bei pagrindiniu pragyvenimo šaltiniu. Pavyzdžiui, samdomas darbuotojas dažniausiai pajamas gauna iš darbo užmokesčio, pensinio amžiaus asmuo – iš pensijų, savarankiškai dirbantis asmuo – iš individualios ar verslo veiklos, todėl aukštas šių rodiklių tarpusavio ryšys yra metodologiškai tikėtinas. Socialinių grupių analizės kontekste ekonomiškai svarbiau identifikuoti, iš kokio šaltinio namų ūkis generuoja pagrindines pajamas – darbo užmokesčio, pensijų, verslo veiklos ar kitų pajamų formų – nei vien formaliai priskirti namų ūki tam tikrai socialinei-profesinei kategorijai. Be to, dalis kitų į modelį

įtraukiamų rodiklių, tokių kaip užimtumo santykis ar ekonominės priklausomybės rodiklis, jau netiesiogiai perteikia reikšmingą informaciją apie socialinę-ekonominę padėtį. Dėl šios priežasties pajamų šaltinio kintamasis šiuo atveju laikytinas informatyvesniu ir geriau atitinkančiu tyrimo tikslą, todėl klasterinėje analizėje socialinės-ekonominės grupės kintamasis eliminuojamas.

Toliau iš modelio eliminuojami infrastruktūrinius būsto požymius apibūdinantys kintamieji – centralizuotas šalto vandens tiekimas, centralizuotas karšto vandens tiekimas ir pagrindinis būsto šildymo būdas. Tarp šių rodiklių nustatyti itin stiprūs tarpusavio ryšiai – tarp karšto vandens tiekimo ir šildymo būdo $\phi_K \approx 0,99$, tarp būsto tipo ir karšto vandens tiekimo $\phi_K \approx 0,94$, tarp būsto tipo ir šildymo būdo $\phi_K \approx 0,83$. Tokie rezultatai rodo, kad šie požymiai aprašo tą pačią gyvenamosios infrastruktūros dimensiją. Atsižvelgiant į tai, kad modelyje jau išlaikomas būsto tipas bei namų ūkio užimamas naudingasis plotas, kurie ekonomiškai geriau atspindi gyvenimo sąlygas ir turto lygį, infrastruktūriniai rodikliai eliminuojami kaip pertekliniai.

Demografinėje dimensijoje pašalintas namų ūkio galvos amžiaus grupės kintamasis. Nustatytas labai stiprus ryšys su agreguotu vidutiniu namų ūkio narių amžiumi ($\phi_K \approx 0,87$) rodo reikšmingą informacinį persidengimą tarp šių rodiklių – abu kintamieji iš esmės perteikia labai artimą informaciją apie namų ūkio gyvenimo ciklą ir kartų struktūrą. Toks aukštas ryšys yra statistiškai pagrįstas ir pačia namų ūkių sudėtimi. Aprašomoji analizė (žr. 4 priedą) rodo, kad 40,7 % visų stebimų namų ūkių sudaro vienas gyvenantis asmuo, o dar 28,7 % – pora be išlaikomų vaikų. Tai reiškia, kad beveik septyni iš dešimties namų ūkių (69,4 %) sudaro vieno arba dviejų asmenų namų ūkiai. Vieno asmens namų ūkiuose namų ūkio galvos amžius natūraliai sutampa su vidutiniu namų ūkio amžiumi, o dviejų suaugusių asmenų namų ūkiuose, ypač porų be vaikų atveju, amžiaus skirtumai dažniausiai nėra dideli, todėl ir agreguota vidutinio amžiaus reikšmė išlieka labai artima namų ūkio galvos amžiui. Atsižvelgiant į nustatytą statistinį informacijos persidengimą bei tai, kad agreguotas vidutinio namų ūkio narių amžiaus rodiklis apima visų namų ūkio narių informaciją ir metodologiškai tiksliau atspindi bendrą namų ūkio demografinę struktūrą, namų ūkio galvos amžiaus grupės kintamasis tolimesnėje klasterinėje analizėje eliminuojamas kaip perteklinę informaciją perduodantis kintamasis.

Analogiškai iš klasterizavimo modelio šalinamas asmenų skaičiaus namų ūkyje kintamasis. Nors absoliutus namų ūkio dydis savaime yra interpretuojamas rodiklis, ϕ_K analizė rodo itin stiprų ryšį su bendra amžiaus priklausomybe ($\phi_K \approx 0,88$), aukštojo išsilavinimo dalimi ($\phi_K \approx 0,79$) bei dirbančių asmenų dalimi ($\phi_K \approx 0,76$). Tai rodo, kad namų ūkio dydžio informacija jau netiesiogiai atsispindi agreguotuose struktūriniuose rodikliuose. Kadangi didelė dalis informacijos apie namų ūkio dydį jau atsispindi agreguotuose struktūriniuose rodikliuose, absoliutus narių skaičius šiame tyrime laikytinas pertekliniu.

Iš klasterizavimui sudaromo duomenų rinkinio taip pat šalinamas namų ūkio galvos išsilavinimo kintamasis. Nors statistinis ryšys su agreguotu išsilavinimo rodikliu yra vidutinio stiprumo ($\phi_K \approx 0,39$), tarp šių požymių egzistuoja konceptualus informacijos persidengimas. Namų ūkio galvos išsilavinimas perteikia tik vieno asmens edukacinį profilį, tuo tarpu agreguotas aukštojo išsilavinimo asmenų dalies namų ūkyje rodiklis apima visų namų ūkio narių išsilavinimo struktūrą ir todėl tiksliau atspindi bendrą namų ūkio edukacinį potencialą. Atsižvelgiant į tai, bei siekiant sumažinti modelio kintamųjų triukšmą, namų ūkio galvos išsilavinimo kintamasis tolimesnėje analizėje eliminuojamas.

Galiausiai eliminuojamas vyrų dalies namų ūkyje kintamasis. Kintamųjų tarpusavio priklausomybės analizė rodo gana stiprų ryšį su dirbančių asmenų santykiu namų ūkyje ($\phi_K \approx 0,80$), kas ekonomiškai gali būti paaiškinama istoriškai stipresniu vyrų dalyvavimu darbo rinkoje bei dažnesniu pajamų generavimu aktyviose amžiaus grupėse. Be to, pirminiai klasterizacijos rezultatai parodė, kad šio požymio įtraukimas sistemingai mažina silueto koeficiento reikšmę, didindamas klasterių vidinį nevienalytiškumą, todėl praktiniame modelyje šis kintamasis generavo daugiau papildomo triukšmo nei naujos analitinės informacijos. Atsižvelgiant į tai, bei į ribotą ekonominę ir vartojimo elgsenos interpretacinę vertę šiame tyrime, vyrų dalies namų ūkyje kintamasis sudarant skirtingas socialines grupes eliminuojamas.

Atlikus pirminę požymių atranką, tolimesnei klasterizavimo analizei sudaromas 14 kintamųjų rinkinys; įtraukiami šie kintamieji: namų ūkio pagrindinis pajamų šaltinis, regionas, gyvenamoji vieta, išlaikomų vaikų skaičius namų ūkyje, būsto tipas, būsto valdos statusas, namų ūkio užimamas naudingasis plotas, namų ūkio tipas, pajamų kvintilinė grupė, taip pat iš individualių duomenų sukonstruoti agreguoti rodikliai: vidutinis namų ūkio narių amžius, dirbančių asmenų dalis namų ūkyje, aukštojo išsilavinimo asmenų dalis namų ūkyje, aukštos kvalifikacijos užimtųjų dalis bei bendras amžiaus priklausomybės rodiklis.

3.3. Skirtingų socialinių grupių identifikavimas

Skirtingų socialinių grupių identifikavimas atliekamas taikant nuoseklų kelių etapų klasterizavimo procesą, siekiant nustatyti optimalų namų ūkių grupavimo sprendinį. Tyrimo metu vertinami skirtingi požymių rinkiniai, optimizuojami atstumų matų parametrai ir lyginami skirtingų algoritmų rezultatai, siekiant identifikuoti stabiliausią ir geriausiai duomenų struktūrą atspindintį grupavimo modelį.

3.3.1. Namų ūkių klasterizavimas po pirminės požymių atrankos

Suformavus klasterizavimui skirtą mišraus tipo duomenų rinkinį, tolimesnė analizė pradedama Gower atstumų matricos sudarymu. Gauta atstumų matrica tampa pagrindiniu atskaitos tašku tolimesniam klasterizavimo rezultatų validavimui ir kokybės vertinimui. Svarbu pažymėti, kad ne visi taikomi klasterizavimo algoritmai šią matricą naudoja tiesiogiai. K-prototipų metodas klasterius formuoja tiesiogiai iš pirminių duomenų, o Gower matrica naudojama tik optimaliam klasterių skaičiui nustatyti bei klasterizavimo kokybės rodikliams apskaičiuoti. Kadangi k-prototipų ir k-medoidų metodai priklauso skaidymu paremtiems klasterizavimo algoritmams, prieš atliekant klasterizavimą būtina iš anksto apibrėžti klasterių skaičių k . Siekiant identifikuoti statistiškai pagrįstą ir ekonomiškai interpretuojamą k reikšmę, kiekvienam testuojamam klasterių skaičiaus variantui atliekamas klasterizavimo kokybės vertinimas, remiantis alkūnės grafiko, vidutinio silueto pločio ir *Calinski–Harabasz* indekso kriterijais.

Taikomų klasterizavimo metodų statistiniam palyginimui šiame tyrime pasirenkamas pagrindinis vertinimo kriterijus – vidutinis silueto koeficientas (toliau – \bar{S}), leidžiantis objektyviai įvertinti ir tarpusavyje palyginti skirtingų metodų suformuotų klasterių vidinį homogeniškumą bei jų tarpusavio atskirtį.

K-prototipų algoritmas

K-prototipų algoritmas yra jautrus skaitinių požymių masteliui, todėl prieš klasterizavimą visiems

tęstiniais kintamiesiems taikomas normalizavimas, siekiant išvengti situacijos, kai didesni diapazoną turintys požymiai neproporcingai dominuoja centroidų formavimo procese.

Siekiant įvertinti modelio parametrinį stabilumą, k-prototipų algoritmui atliekama parametru validacija, testuojant skirtingas centroidų inicializavimo strategijas bei algoritmo parametru kombinacijas. Validavimo metu lyginami abu metode galimi centroidų inicializavimo būdai – *Cao* ir *Huang*, keičiamos atsitiktinės pradinės būsenos reikšmės (0, 42, 123), taip pat vertinama didesnio maksimalaus iteracijų skaičiaus įtaka, lyginant standartinę (100) ir išplėstinę (500) iteracijų limitą. Visais atvejais klasterizavimas atliekamas naudojant dviejų klasterių sprendinį, o modelių tarpusavio palyginimui taikomas vidutinis silueto koeficientas.

Atlikta parametru validavimo analizė parodė, kad nepriklausomai nuo pasirinkto centroidų inicializavimo metodo, atsitiktinės pradinės būsenos ar maksimalaus iteracijų skaičiaus, visais testuotais atvejais buvo gauta identiška vidutinio silueto pločio reikšmė lygi 0,29. Tai leidžia teigti, kad taikomas k-prototipų modelis analizuojamų duomenų atveju pasižymi aukštu parametru stabilumu, o klasterizavimo rezultatai nėra jautrūs pasirinktoms algoritmo inicializavimo sąlygoms. Atsižvelgiant į gautą parametru nuoseklumą, tolimesnei analizei pasirinktas bazinis modelio variantas, taikant *Cao* inicializavimo metodą, atsitiktinės pradinės būsenos reikšmę 42 ir 100 iteracijų skaičių.

Pasirinktam klasterizavimo modeliui optimalaus klasterių skaičiaus nustatymas atliekamas taikant skyriaus pradžioje aptartus tris vertinimo kriterijus, o gauti rezultatai pateikiami 8 paveiksle.



8 pav. K-prototipų algoritmo klasterių skaičiaus nustatymas

Gauti rezultatai rodo, kad alkūnės grafikas optimalų sprendinį identifikuoja ties trimis klasteriais, tuo tarpu tiek silueto kriterijus, tiek *Calinski-Harabasz* indeksas didžiausias reikšmes pasiekia taikant dviejų klasterių sprendinį. Atsižvelgiant į vertinimo kriterijų daugumos principą, tolimesnei analizei pasirenkamas 2 klasterių sprendinys. Taikant šį sprendinį, stebiniai pasiskirsto į du klasterius, kuriuos sudaro 2666 ir 1668 stebiniai, o bendras vidutinis silueto koeficientas siekia 0,29. Papildoma individualių klasterių silueto analizė rodo, kad pirmasis klasteris pasižymi pakankamai aiškia vidine struktūra ($S_1 = 0,36$), tuo tarpu antrojo klasterio vidutinė silueto reikšmė yra pastebimai mažesnė ($S_2 = 0,18$). Atsižvelgiant į tai, kad $S < 0,2$ literatūroje yra interpretuojamas kaip nepriimtinos klasterizavimo kokybės požymis [60], galima teigti, jog antrojo klasterio vidinė struktūra išlieka nepakankamai aiški. Tai rodo, kad nors bendras dviejų klasterių sprendinys statistiškai identifikuoja egzistuojančią grupavimo struktūrą, dalies stebinių segmentacija išlieka riboto aiškumo.

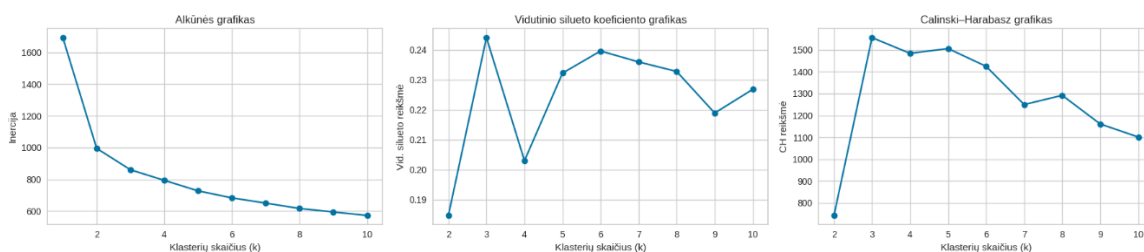
K-medoidų algoritmas

Toliau taikomas k-medoidų klasterizavimo metodas, kuris, priešingai nei k-prototipai, klasterių formavimui tiesiogiai naudoja iš anksto apskaičiuotą Gower atstumų matricą, leidžiančią įvertinti

stebinių tarpusavio nepanašumus, suformuoti panašių objektų grupes ir identifikuoti kiekvieną klasterį reprezentuojančius medoidus.

Siekiant įvertinti šio modelio parametrinį stabilumą, taip pat atliekama algoritmo parametrų validacija, testuojant skirtingas medoidų inicializavimo strategijas – „*k-medoids*++“, „*random*“ ir „*heuristic*“, keičiant atsitiktinės pradinės būsenos reikšmes (0, 42, 123) bei vertinant didesnio maksimalaus iteracijų skaičiaus įtaką, lyginant standartinį (100) ir išplėstinį (500) iteracijų limitą. Validavimas atliekamas su dviejų klasterių sprendiniu. Gautų rezultatų analizė rodo, kad nepriklausomai nuo pasirinktos medoidų inicializavimo strategijos, atsitiktinės pradinės būsenos ar maksimalaus iteracijų skaičiaus, visais testuotais atvejais gaunama identiška vidutinio silueto pločio reikšmė ($\bar{S} = 0,18$). Todėl *k-medoids* modelis analizuojamų duomenų atžvilgiu taip pat pasižymi aukštu parametrų stabilumu. Atsižvelgiant į gautą parametrų nuoseklumą, tolimesnei analizei pasirinktas bazinis modelio variantas, taikant „*k-medoids*++“ inicializavimo metodą, atsitiktinės pradinės būsenos reikšmę 42 ir maksimalų 100 iteracijų skaičių.

Tolimesniam modelio vertinimui optimalaus klasterių skaičiaus nustatymas atliekamas remiantis anksčiau aprašytais klasterizavimo kokybės kriterijais, o gauti rezultatai pateikiami 9 paveiksle.



9 pav. K-medoidų algoritmo klasterių skaičiaus nustatymas

Gauti rezultatai rodo, kad tiek alkūnės grafikas, tiek vidutinio silueto koeficiento, tiek *Calinski-Harabasz* kriterijus didžiausias reikšmes pasiekia taikant trijų klasterių sprendinį. Atsižvelgiant į vieningą visų vertinimo kriterijų rezultatą, tolimesnei analizei pasirinktas 3 klasterių sprendinys. Taikant šį sprendinį, stebiniai pasiskirsto į tris klasterius, kuriuos sudaro 1603, 1841 ir 890 stebinių, o bendras vidutinis silueto koeficientas siekia 0,24. Tačiau papildoma individualių klasterių silueto analizė atskleidžia nevienodą vidinę grupių struktūrą. Antrasis ($S_2 = 0,39$) ir trečiasis ($S_3 = 0,26$) klasteriai pasižymi pakankamai aiškia vidine struktūra ir tarpusavio atskirtimi, tuo tarpu pirmojo klasterio vidutinė silueto reikšmė yra itin žema ($S_1 = 0,07$). Atsižvelgiant į tai, nors bendras *k-medoids* sprendinys statistiškai identifikuoja trijų grupių struktūrą, ne visi susiformavę klasteriai pasižymi vienodu interpretaciniu stabilumu.

Hierarchinio klasterizavimo algoritmas

Galiausiai taikomas hierarchinis klasterizavimas, pasitelkiant Gower atstumo matricą. Klasterių formavimui naudojami skirtingi aglomeracinio jungimo metodai – vienetinės, vidutinės ar pilnosios jungties. Klasterių skaičius nustatomas automatiškai analizuojant dendrogramos aukščius ir ieškant didžiausio aukščio skirtumo tarp gretimų jungimų, t. y. dendrogramos pjovimo taškas nustatomas pagal didžiausią atstumų šuolį. Atitinkamos dendrogramos pateikiamos 9 priede, kur mėlyna horizontali linija žymi automatiškai nustatytą dendrogramos pjovimo aukštį, pagal kurį galutinėje struktūroje atskiriami tarpusavyje labiausiai nutolę klasteriai.

Gauti rezultatai rodo, kad vienietinės jungties metodas generuoja aukščiausią vidutinio silueto pločio reikšmę ($\bar{S} = 0,31$), tačiau suformuota klasterių struktūra pasižymi ryškiu disbalansu – praktiškai visi stebiniai priskiriami vienam dominuojančiam klasteriui ($n_1 = 4333$), o antrąjį klasterį sudaro tik pavienis stebinys. Nors statistiniu požiūriu gauta silueto reikšmė yra aukščiausia tarp testuotų hierarchinio klasterizavimo variantų, tokia klasterių struktūra ekonominės interpretacijos požiūriu laikytina neprasminga ir negali būti laikoma tinkama socialinių grupių segmentavimui. Vidutinės jungties metodas generuoja kiek mažesnę vidutinio silueto pločio reikšmę ($\bar{S} = 0,26$), tačiau ir šiuo atveju klasterių struktūra išlieka metodologiškai problemiška – didžioji dalis stebinių koncentruojasi viename klasteryje ($n_1 = 4331$), o likusius du klasterius sudaro tik 2 ir 1 stebiniai. Tokia struktūra taip pat neleidžia identifikuoti ekonomiškai interpretuojamų bei statistiškai stabilų socialinių grupių. Tuo tarpu pilnosios jungties metodas suformuoja gerokai labiau subalansuotą dviejų klasterių struktūrą ($n_1 = 1525$ ir $n_2 = 2809$), tačiau bendras vidutinis silueto koeficientas ($S = 0,22$) išlieka santykinai žemas ir rodo tik vidutinio stiprumo klasterių vidinį vientisumą ir tarpusavio atskirtį.

Nors hierarchinis klasterizavimas kai kuriais atvejais pateikė aukštesnes \bar{S} reikšmes nei anksčiau nagrinėti metodai, gauti klasteriai pasižymi menku ekonominiu interpretuojamumu ir neleidžia suformuoti tyrimo kontekste tikėtinos 4–6 socialinių grupių struktūros. Ekonominiu požiūriu dviejų ar trijų grupių segmentavimas dažniausiai leidžia identifikuoti tik bendras žemo, vidutinio ir aukšto socioekonominio statuso grupes, tačiau neleidžia pakankamai detalai atskirti pensinio amžiaus, ekonomiškai aktyvių, aukštos kvalifikacijos, priklausomų ar mišrios struktūros namų ūkių.

3.3.2. Gower atstumų matricos optimizacija

Atsižvelgiant į tai, kad klasterizavimo rezultatus reikšmingai lemia ne tik pasirinktas algoritmas, bet ir pačių požymių indėlis į bendrą panašumo struktūrą, tolimesniame etape atliekama Gower matricos svorių optimizacija, kurios tikslas – įvertinti, ar skirtingas požymių svoris bendroje struktūroje gali pagerinti k-medoidų klasterizavimo kokybę. Pradiniame variante visi požymiai laikomi vienodai svarbiais, tačiau socialinių duomenų atveju skirtingi kintamieji nebūtinai turi vienodą informacinę vertę. Dėl šios priežasties kiekvienam požymiui priskiriamas optimizuotas svoris, o Gower atstumo matrica perskaičiuojama pagal naują svorių kombinaciją.

Optimizavimo procedūra vykdoma iteraciniu principu. Pirmiausia kiekvienam požymiui nustatomos leistinos svorių ribos nuo 1 iki 3 imtinai, taip apribojant galimą vieno požymio neproporcingą dominavimą bendroje atstumų struktūroje. Tuomet, taikant L-BFGS-B⁹ optimizavimo algoritmą, ieškoma tokios svorių kombinacijos, kuri maksimizuotų vidutinį silueto koeficientą. Kadangi optimizavimo funkcijos tikslas techniniu požiūriu yra minimizavimas, skaičiavimuose minimizuojama neigiama vidutinio silueto reikšmė. Kiekvienoje iteracijoje pagal einamąjį svorių vektorių sudaroma svertinė Gower atstumo matrica, atliekama k-medoidų klasterizacija su pasirinktu klasterių skaičiumi, apskaičiuojama vidutinio silueto reikšmė ir rezultatas išsaugomas optimizavimo istorijoje. Tokia procedūra leidžia įvertinti, kaip skirtingos požymių svorių kombinacijos veikia klasterių kokybę. Kadangi svorių optimizavimas yra laikui imli procedūra, šiame etape vertinami tik 4–6 klasterių sprendiniai, kurie labiausiai atitinka tyrimo tikslą. Kiekvienam klasterių skaičiui optimizavimo procedūra atliekama atskirai. Bendri svorių optimizavimo rezultatai, apimantys klasterizavimo skaičiavimo trukmę bei klasterizavimo kokybės pokyčius, pateikiami 4 lentelėje. Siekiant detaliau įvertinti atskirų požymių indėlį į galutinę panašumo struktūrą, 10 priede pateikiama

⁹ https://en.wikipedia.org/wiki/Limited-memory_BFGS

optimizuotų svorių suvestinė, parodanti, kuriems požymiams optimizavimo algoritmo metu buvo priskirti santykinai didesni arba mažesni svoriai.

Nors svorių optimizavimas kai kuriems sprendiniams leido nežymiai padidinti vidutinį silueto koeficientą, bendras klasterizavimo kokybės pagerėjimas buvo ribotas. Be to, detalesnė rezultatų analizė parodė, kad dalies individualių klasterių silueto koeficientų reikšmės išliko mažesnės nei 0,2, o tai rodo silpną klasterių tarpusavio atskirtį ir nepakankamą jų vidinį homogeniškumą. Tokie rezultatai leidžia daryti prielaidą, kad pagrindinis apribojimas šiame etape susijęs ne su požymių svorių paskirstymu, bet su pačių kintamųjų rinkinio sudėtimi. Dėl šios priežasties tolimesniame tyrimo etape nuspręsta grįžti prie požymių atrankos procedūros ir iš naujo įvertinti į modelį įtrauktų kintamųjų informatyvumą bei jų skiriamąją galią.

4 lentelė. Gower svorių optimizavimo rezultatų suvestinė (\bar{S} – vidutinis silueto koeficientas)

Klasterių skaičius (k)	Skaičiavimo laikas, min	\bar{S} prieš optimizaciją	\bar{S} po optimizacijos	\bar{S} pokytis	Stipriausias požymis
4	15	0,203	0,389	0,187	Namų ūkio pagrindinis pajamų šaltinis / Būsto tipas
5	111	0,232	0,384	0,152	Dirbančių asmenų dalis namų ūkyje / Būsto tipas
6	124	0,240	0,382	0,142	Dirbančių asmenų dalis namų ūkyje / Būsto tipas / Aukštojo išsilavinimo dalis namų ūkyje

3.3.3. Namų ūkių klasterizavimas po atgalinio požymių šalinimo

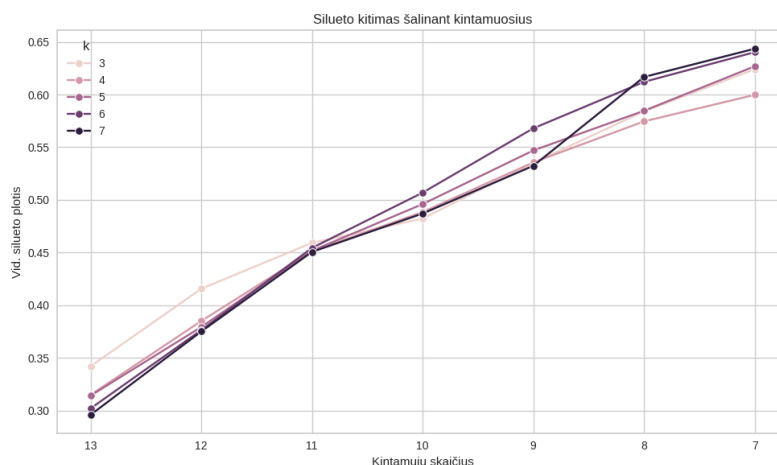
Kadangi Gower matricos svorių optimizavimas nepagerino klasterizavimo rezultatų iki pakankamai interpretuojamo lygmens, tolimesniame etape taikoma papildoma požymių redukcija. Ši procedūra grindžiama atgalinio požymių šalinimo (angl. *backward feature elimination*) principu, plačiai taikomu mašininio mokymosi (angl. *machine learning*) uždaviniuose ir modelių atrankos procedūrose, kai analizė pradedama nuo viso turimo požymių rinkinio, o kiekviename žingsnyje pašalinamas vienas kintamasis, kurio eliminavimas labiausiai pagerina modelio kokybę, šiuo atveju vertinamą pagal vidutinio silueto plotį.

Atgalinis požymių šalinimas

Atgalinis šalinimas atliekamas taikant k-medoidų algoritmą su Gower atstumų matrica, naudojant „*k-medoids++*“ inicializavimą ir fiksuotą atsitiktinės pradinės būsenos reikšmę 42. Siekiant išlaikyti pakankamą požymių informacinę aprėptį, šalinimas vykdomas tik tol, kol modelyje lieka ne mažiau kaip pusė analizuojamų požymių. Kadangi pradinį rinkinį sudaro 14 kintamųjų, mažiausias vertintas požymių skaičius yra 7. Procedūra automatiškai kartojama taikant 3, 4, 5, 6 ir 7 klasterių sprendinius. Bendras skaičiavimo laikas siekė 68 min. 20 sek.

Gauti rezultatai (žr. 11 priedą) rodo, kad visų testuotų variantų atveju pirmieji trys šalinami požymiai sutampa ir atitinka gyvenamąją vietą, būsto tipą bei namų ūkio tipą. Toks nuoseklus eliminavimas eiliškumas leidžia daryti prielaidą, kad šie kintamieji bendroje Gower atstumų struktūroje generuoja papildomą triukšmą ir mažina klasterių tarpusavio atskirtį. Siekiant detaliau įvertinti, kaip nuoseklus požymių eliminavimas veikia klasterizavimo kokybę skirtinguose klasterių skaičiaus scenarijuose, 10 paveiksle pateikiama silueto koeficiento kitimo dinamika, sumažinus požymių skaičių nuo 13 iki 7.

Grafiškai pateikti rezultatai leidžia įvertinti klasterizavimo kokybės jautrumą požymių rinkinio pokyčiams. Matyti, kad nepriklausomai nuo pasirinkto klasterių skaičiaus ($k = 3-7$), pirmųjų požymių eliminavimas lemia nuoseklų ir gana spartų vidutinio silueto pločio augimą – nuo maždaug 0,30–0,34 iki 0,45–0,46, o tai empiriškai patvirtina, jog pradiniam požymių rinkinyje buvo kintamųjų, kurių skiriamoji geba klasterių formavimo procese yra ribota. Toliau mažinant požymių skaičių, silueto reikšmė toliau didėja ir atskirais scenarijais pasiekia net 0,64, rodydama vis didesnę statistinį klasterių išgryninimą.



10 pav. Vidutinio silueto koeficiento reikšmės kitimo dinamika taikant atgalinį požymių šalinimą

Vis dėlto galutinis požymių atrankos sprendimas šiame tyrime nėra grindžiamas vien didžiausia vidutinio silueto pločio reikšme. Vėlesniuose atgalinio požymių šalinimo etapuose algoritmas pradeda eliminuoti rodiklius, kurie ekonominės interpretacijos požiūriu apibūdina esmines namų ūkių socialinės padėties dimensijas. Tarp tokių požymių patenka pajamų kvintilinė grupė, būsto valdos statusas bei namų ūkio užimamas naudingasis plotas, tiesiogiai atspindintys disponuojamų išteklių lygį, sukauptą materialinį kapitalą ir gyvenimo sąlygas. Analogiškai, dalyje scenarijų eliminuojami ir tokie rodikliai kaip aukštos kvalifikacijos užimtųjų dalis bei vidutinis namų ūkio narių amžius, perteikiantys žmogiškojo kapitalo bei gyvenimo ciklo informaciją. Nors statistiniu požiūriu šių kintamųjų eliminavimas didina klasterių tarpusavio atskirtį, toks sprendimas kartu mažina modelio ekonominį ir socialinį paaiškinamumą. Dėl šios priežasties galutinis požymių atrankos sprendimas grindžiamas ne vien didžiausia klasterizavimo kokybės rodiklio reikšme, bet ir kompromisu tarp statistinio tikslumo bei išsaugamo informacinio turinio. Pirmųjų trijų požymių eliminavimas visiems klasterių skaičiaus scenarijams lėmė labai panašią vidutinio silueto pločio reikšmę ($\bar{S} = 0,45$), todėl vien šių požymių pašalinimas nesuteikė aiškaus pranašumo tarp alternatyvių klasterių sprendinių. Tuo tarpu papildomas ketvirtojo požymio pašalinimas leido išryškinti skirtumus tarp skirtingų klasterių skaičiaus scenarijų ir kartu reikšmingai padidinti klasterizavimo kokybę, neprarandant pagrindinių namų ūkių ekonominę ir socialinę diferenciaciją apibūdinančių rodiklių. Atsižvelgiant į tai, tolimesnei analizei pasirinktas 6 klasterių sprendinys su 10 požymių rinkiniu. Po keturių požymių eliminavimo šis sprendinys pasiekė didžiausią vidutinio silueto pločio reikšmę tarp galutinių kandidatų ($\bar{S} = 0,507$), išlaikydamas pagrindinius namų ūkių socialinės padėties aspektus apibūdinančius kintamuosius.

Namų ūkių klasterizavimas

Nors pasirinktas 6 klasterių sprendinys su 10 požymių rinkiniu tarp visų nesvertų modelių pasiekė didžiausią vidutinio silueto koeficiento reikšmę ($\bar{S} = 0,507$), detalesnė individualių klasterių analizė parodė, kad ne visi klasteriai tenkino minimalią priimtina kokybės ribą ($S_i > 0,2$). Tai rodo, kad nors bendras klasterizavimo kokybės rodiklis buvo aukščiausias tarp vertintų alternatyvų, dalis grupių vis dar pasižymėjo nepakankamu vidiniu homogeniškumu ir silpnesne tarpusavio diferenciacija. Dėl šios priežasties tolimesniame etape taikyta Gower atstumo matricos požymių svorių optimizavimo procedūra. Optimizavimo rezultatai parodė reikšmingą klasterizavimo kokybės pagerėjimą – vidutinis silueto koeficientas padidėjo nuo 0,507 iki 0,622. Papildoma individualių klasterių analizė taip pat parodė, kad po optimizavimo visų grupių silueto reikšmės viršijo minimalią ribą ($S_i > 0,2$): $S_0 = 0,2200$, $S_1 = 0,6968$, $S_2 = 0,4545$, $S_3 = 0,2558$, $S_4 = 0,2600$ ir $S_5 = 0,4699$.

Galutinės optimizuotų požymių svorių reikšmės pateikiamos 12 priede, o pagrindinių taikytų klasterizavimo metodų, jų parametrizavimo ir gautų rezultatų palyginamoji suvestinė pateikiama 5 lentelėje. Atsižvelgiant į bendrą klasterizavimo kokybės pagerėjimą bei individualių klasterių stabilumą, optimizuotas 6 klasterių k-medoidų modelis su 10 požymių rinkiniu laikomas tinkamiausiu iš šiame tyrime vertintų sprendinių.

5 lentelė. Taikytų klasterizavimo metodų rezultatų suvestinė

Klasterizavimo metodas	Taikomi svoriai	Požymių skaičius	\bar{S} reikšmė	Visų $S_i > 0,2$	Klasterių skaičius	Stebinių pasiskirstymas klasteriuose					
						1	2	3	4	5	6
k-prototipų	--	14	0,293	Ne	2	1668	2666				
k-medoidų	Ne	14	0,244	Ne	3	1603	1841	890			
k-medoidų	Taip	14	0,390	Ne	4	923	1303	1278	830		
k-medoidų	Taip	14	0,384	Ne	5	502	1280	831	881	840	
k-medoidų	Taip	14	0,382	Ne	6	840	1273	408	667	729	417
vienetinės jungties	Ne	14	0,311	Ne	2	4333	1				
vidutinės jungties	Ne	14	0,262	Ne	3	4331	2	1			
pilnosios jungties	Ne	14	0,220	Ne	2	1525	2809				
k-prototipų	--	10	0,333	Ne	6	294	775	512	493	1725	535
k-medoidų	Ne	10	0,507	Ne	6	423	1901	370	891	524	225
k-medoidų	Taip	10	0,622	Taip	6	422	1902	902	218	323	567

3.3.4. Klasterinės analizės rezultatų vertinimas ir profiliavimas

Siekiant įvertinti identifikuotų klasterių patikimumą ir statistinį pagrįstumą, papildomai atlikta klasterių stabilumo bei tarpusavio atskiriamumo analizė. Išsamūs ARI ir *XGBoost* taikytų metodų parametrai bei gauti rezultatai pateikiami 13 priede.

Klasterių stabilumas vertinamas taikant pakartotinio imties atrinkimo metodą ir naudojant pakoreguotą Rando indeksą (ARI). Analizėje atliekama 30 iteracijų, kiekvienoje jų atsitiktinai atrenkant 80 % pradinės imties ir išlaikant pradinių klasterių proporcijas. Gauti rezultatai parodė labai aukštą sprendinio stabilumą – vidutinė ARI reikšmė siekė 0,991, standartinis nuokrypis sudarė tik 0,011. Tai rodo, kad klasterių struktūra išlieka praktiškai nepakitusi keičiantis duomenų sudėčiai,

todėl identifikuotos grupės gali būti laikomos stabiliomis ir mažai jautriomis atsitiktiniams imties svyravimams.

Papildomai klasterių tarpusavio atskirčiai įvertinti taikytas gradientinio stiprinimo klasifikavimo metodas *XGBoost*, kurio paskirtis buvo patikrinti, ar identifikuotas grupes galima patikimai atkurti remiantis pradinių požymių rinkiniu. Modelio vertinimui taikyta trijų dalių stratifikuota kryžminė validacija. Gauti rezultatai parodė labai aukštą klasterių tarpusavio atskirtį – vidutinis kryžminės validacijos tikslumas siekė 0,996 ($\sigma = 0,002$), o nepriklausomos testavimo imties tikslumas sudarė 0,993. Klaidų analizė parodė, kad neteisingai priskirtų stebinių skaičius buvo minimalus, o didžioji dalis stebinių buvo priskirti atitinkamoms grupėms. Didžiausią reikšmę klasterių diferenciacijai turėjo pagrindinis namų ūkio pajamų šaltinis, dirbančių asmenų dalis, aukštojo išsilavinimo dalis, aukštos kvalifikacijos užimtųjų dalis bei išlaikomų vaikų skaičius namų ūkyje. Siekiant įvertinti rezultatų stabilumą ir patikrinti, ar modelio veikimas nepriklauso nuo konkrečios mokymo ir testavimo imčių sudėties, modelio vertinimas atliktas remiantis trimis skirtingomis mokymo ir testavimo imčių skaidymo realizacijomis. Skirtingos duomenų skaidymo realizacijos buvo generuojamos naudojant skirtingas atsitiktinių pradinių būsenų reikšmes: 42, 123 ir 999, išlaikant nekintančius modelio parametrus bei vertinimo procedūrą. Nustatyta, kad keičiant duomenų padalijimo pradinės būsenos reikšmę klasifikavimo tikslumo reikšmės išliko labai panašios, todėl rezultatai nebuvo jautrūs konkrečios testavimo imties sudėčiai. Tai rodo, kad gauta aukšta klasifikavimo kokybė nėra atsitiktinai susijusi su palankiai suformuota testavimo imtimi. Identifikuoti klasteriai pasižymi aiškia vidine struktūra ir gali būti laikomi ne atsitiktinio duomenų susiskaidymo rezultatu, bet daugiamatę socialinę struktūrą atspindinčiomis grupėmis.

Toliau atliekamas identifikuotų klasterių profiliavimo, siekiant apibrėžti jų socialines ir ekonomines charakteristikas. Klasterių profiliai sudaromi remiantis aprašomosios statistikos rodikliais bei kategorinių požymių pasiskirstymo struktūra, vertinant centrinės tendencijas, sklaidos rodiklius ir požymių variaciją tarp grupių. Remiantis gautais profiliavimo rezultatais, identifikuotos skirtingos socialinės grupės pateiktos 6 lentelėje. Detalesni sudarytų šešių skirtingų socialinių grupių profiliai pateikiami 14 priede.

6 lentelė. Identifikuotų skirtingų socialinių grupių profilių suvestinė

Klasteris	Identifikuota socialinė grupė	Pagrindinės charakteristikos
1 klasteris	Jaunos mažesnių pajamų šeimos su vaikais	Jaunesni namų ūkiai, žemas išsilavinimo ir kvalifikacijos lygis, didesnis išlaikomų vaikų skaičius, žemesnės pajamų kvintilės, santykinai prastesnės gyvenimo sąlygos.
2 klasteris	Pensinio amžiaus namų ūkiai	Vyresnio amžiaus namų ūkiai, labai maža dirbančių asmenų dalis, ryškiausias pajamų šaltinis – pensijos, maža išlaikomų vaikų dalis, žemesnės pajamos.
3 klasteris	Vidutinių–aukštesnių pajamų dirbantys namų ūkiai	Aukšta dirbančių asmenų dalis, vidutinis amžius, neturi vaikų, didesnė aukštesnių pajamų grupių koncentracija.
4 klasteris	Socialiai pažeidžiami namų ūkiai	Dominoja socialinės išmokos ir kiti socialiniai pajamų šaltiniai, žemiausios pajamų grupės, mažas užimtumas, didesnė priklausomybės našta.
5 klasteris	Jaunos ekonomiškai aktyvios šeimos	Jaunesni namų ūkiai su vaikais, aukšta dirbančiųjų dalis, didelė būsto savininkų su paskolomis dalis, aukštesnės pajamos.
6 klasteris	Aukštos kvalifikacijos ir aukštų pajamų namų ūkiai	Didžiausia aukštojo išsilavinimo ir aukštos kvalifikacijos darbuotojų dalis, didžiausios pajamos, aukštas užimtumas, mažas išlaikomų vaikų skaičius.

Gauti klasterių profiliai iš esmės atitinka teorinėje dalyje aptartą socialinių grupių tipologiją, kadangi grupių formavimasi daugiausia lėmė socialinę padėtį apibūdinantys veiksniai – pajamos, užimtumas, išsilavinimas, pagrindinis pajamų šaltinis ir namų ūkio struktūra. Mažesniais ekonominiais ištekliais pasižymintys pensinio amžiaus ir socialiai pažeidžiami namų ūkiai atitinka teoriškai aptariamą didesnės socialinės rizikos grupes, tuo tarpu aukštos kvalifikacijos ir aukštų pajamų namų ūkiai atspindi aukštesnio socialinio statuso grupes.

3.4. Vartojimo išlaidų struktūrų tyrimas

Identifikavus šešias skirtingas socialines grupes, toliau atliekama jų vartojimo išlaidų struktūrų analizė, siekiant įvertinti, ar identifikuotos socialinės grupės pasižymi skirtingais vartojimo elgsenos ypatumais. Šiame etape vertinami vartojimo išlaidų pasiskirstymo skirtumai tarp identifikuotų grupių bei nagrinėjami vartojimo struktūros dėsningumai. Vartojimo išlaidų pasiskirstymas pagal suformuotus klasterius pateikiamas 15 priede ir naudojamas kaip pagrindas tolesnei statistinei bei entropija grindžiamai analizei.

Iš vartojimo išlaidų pasiskirstymo grafikų nustatomi keli bendri dėsningumai. Daugumos vartojimo kategorijų pasiskirstymai pasižymi ryškia teigiama asimetrija – didžioji namų ūkių dalis koncentruojasi mažesnių išlaidų intervale, o didesnių išlaidų reikšmės pasireiškia tik nedidelei namų ūkių daliai ir formuoja ilgesnę dešiniąją pasiskirstymo uodegą. Toks pasiskirstymo pobūdis ryškiausiai matomas nebūtinąjo vartojimo kategorijose, tokiose kaip poilsis, restoranų ir apgyvendinimo paslaugos, transportas, apranga bei būsto įrengimo išlaidos, kur stebima didesnė išlaidų sklaida ir ryškesnės kraštinės reikšmės. Tuo tarpu būtinąjo vartojimo kategorijose, ypač maisto bei būsto, vandens, elektros, dujų ir kito kuro išlaidose, stebima didesnė reikšmių koncentracija siauresniuose intervaluose. Tai rodo santykinai mažesnę šių išlaidų kintamumą tarp namų ūkių. Be to, vartojimo išlaidų pasiskirstymo formos tarp klasterių skiriasi ne tik bendru išlaidų lygiu, bet ir sklaidos pobūdžiu bei atskirų vartojimo kategorijų reikšmingumu.

3.4.1. Vartojimo išlaidų struktūrų statistinio palyginimo analizė

Siekiant statistiškai įvertinti, ar suformuotos socialinės grupės reikšmingai skiriasi pagal vartojimo išlaidų struktūrą, toliau atlikta nuosekli daugiapakopė hipotezių tikrinimo procedūra, apimanti tiek parametrinių, tiek neparametrinių metodų prielaidų vertinimą. Šiame analizės etape faktoriumi (nepriklausomu kintamuoju) laikoma klasterizavimo metu suformuota socialinė grupė (klasteris), o priklausomais kintamaisiais – atskiros vartojimo išlaidų kategorijos, apimančios bendras vartojimo išlaidas bei COICOP klasifikacija pagrįstas pirmojo lygio vartojimo grupes.

Pirmiausia nustatoma, kad analizuojamos vartojimo išlaidų kategorijos gali būti laikomos tolydziais kiekybiniais kintamaisiais, todėl formaliai gali būti įtraukiamos į daugiamatės vienfaktorinės dispersinės analizės (MANOVA) modelį. Toliau vertinama priklausomų kintamųjų tarpusavio tiesiškumo prielaida, sudarant porinių sklaidos diagramų matricą (žr. 16 priedą). Grafinė analizė rodo ryškią asimetriją, didelę nulinių reikšmių koncentraciją, heteroskedastiškumą bei aiškios tiesinės struktūros nebuvimą. Tokie rezultatai leidžia daryti prielaidą, kad nėra tenkinamos kelios svarbios MANOVA metodo taikymo sąlygos, todėl šio metodo taikymas laikomas ribotai pagrįstu. Atsižvelgiant į tai, pereinama prie vienmačių analizių, kiekvieną vartojimo kategoriją nagrinėjant atskirai.

Kiekvienai vartojimo išlaidų kategorijai ir bendrų vartojimo išlaidų kintamajam atskirai vertinama vienmatė vienfaktorinės dispersinės analizės (ANOVA) taikymo galimybė. Pirmoji ANOVA prielaida, susijusi su stebinių nepriklausomumu, laikoma tenkinama, kadangi ją užtikrina pats tyrimo planas – kiekvienas stebiny atitinka atskirą nepriklausomą namų ūkį, o duomenų rinkimo procedūra užtikrina nepriklausomus stebėjimus. Antroji prielaida, susijusi su grupių dispersijų homogeniškumu, tikrinama taikant Leveno ir Bartletto kriterijus. Gauti rezultatai (žr. 17 priedą) rodo, kad visose be išimties analizuojamose vartojimo kategorijose grupių dispersijos skiriasi statistiškai reikšmingai, todėl klasikinės vienfaktorinės dispersinės analizės taikymas tampa ribotas. Atsižvelgiant į tai, tolimesniame etape kiekvienai vartojimo kategorijai taikoma Velčo ANOVA, kuri leidžia korektiškai vertinti grupių vidurkių skirtumus esant nevienodoms dispersijoms. Po modelio įvertinimo papildomai tikrinama trečioji parametrinės analizės prielaida – modelio paklaidų normalumas. Šiam tikslui taikoma kelių testų kombinacija: Šapiro–Vilko (*Shapiro–Wilk*), Kolmogorovo–Smirnovo (*Kolmogorov–Smirnov*) ir Kramero–fon Miseso (*Cramér–von Mises*) kriterijai. Visose vartojimo išlaidų kategorijose gauti rezultatai (žr. 18 priedą) rodo, kad nulinė hipotezė apie Velčo ANOVA modelių paklaidų normalumą skirstinį buvo atmesta ($p < 0,05$), paklaidų skirstiniai nėra suderinami su normaliuoju skirstiniu. Papildomai vizualinė histogramų analizė atskleidžia ryškia dešinės pusės asimetriją. Kadangi normalumo prielaida sistemingai pažeidžiama visose analizuojamose kategorijose, parametrinės dispersinės analizės taikymas nėra tęsiamas.

Atsižvelgiant į nustatytus parametrinių metodų ribotumus, tolimesniame etape pereinama prie neparametrinio vienfaktorinės dispersinės analizės analogo – Kruskalo–Voliso (*Kruskal–Wallis*) kriterijaus 6 nepriklausomoms imtims. Nors šis metodas nereikalauja normalumo prielaidos ir leidžia vertinti, ar kelių nepriklausomų grupių skirstiniai statistiškai reikšmingai skiriasi, vienoda skirstinių forma reikalinga tuo atveju, kai skirtumai interpretuojami kaip medianų skirtumai. Dėl šios priežasties prieš taikant neparametrinę analizę įvertinama, ar grupių skirstiniai skiriasi, taikant Andersono–Darlingo (*Anderson–Darling*) kriterijų kiekvienai vartojimo kategorijai atskirai. Gauti rezultatai (žr. 19 priedą) rodo, kad visose analizuojamose vartojimo išlaidų kategorijose nulinė hipotezė apie vienodus klasterių skirstinius buvo atmesta ($p < 0,05$), todėl identifikuotų socialinių grupių vartojimo išlaidų pasiskirstymai skiriasi statistiškai reikšmingai. Tai reiškia, kad identifikuotos socialinės grupės pasižymi reikšmingai besiskiriančiais vartojimo pasiskirstymais, todėl Kruskalo–Voliso kriterijaus interpretacija tampa metodologiškai dviprasmiška. Atsižvelgiant į tai, neparametrinė dispersinė analizė nėra tęsiama.

Toliau atliktas porinių klasterių skirstinių palyginimas, siekiant identifikuoti, kurios konkrečios socialinių grupių poros pasižymi statistiškai reikšmingai skirtingais vartojimo išlaidų pasiskirstymais. Kadangi ankstesni prielaidų tikrinimo etapai parodė ryškia vartojimo išlaidų skirstinių asimetriją, nulinių reikšmių koncentraciją, nevienodą sklaidą ir skirtingas skirstinių formas tarp klasterių, tolimesnėje analizėje naudojami testai, orientuoti į empirinių skirstinių palyginimą. Kiekvienai klasterių porai atskirai taikomi Kolmogorovo–Smirnovo (KS) ir Andersono–Darlingo (AD) dviejų imčių kriterijai. Šių testų kombinacija leidžia įvertinti skirtumus tarp empirinių skirstinių – nuo bendro pasiskirstymo funkcijų nesutapimo iki jautresnio skirtumų identifikavimo skirstinių centruose, uodegose ir bendroje jų formoje.

Porinių klasterių palyginimų rezultatai (žr. 20 priedą) rodo, kad daugumoje vartojimo kategorijų identifikuotos socialinės grupės pasižymi statistiškai reikšmingai skirtingais vartojimo išlaidų pasiskirstymais ($p < 0,05$). Ryškiausi skirtumai nustatyti bendrų vartojimo išlaidų, restoranų ir apgyvendinimo paslaugų, draudimo ir finansinių paslaugų, informacijos ir ryšių bei aprangos ir

avalynės kategorijose. Šiose kategorijose beveik visos klasterių poros statistiškai reikšmingai skyrėsi pagal abu taikytus kriterijus. Didelė dalis šių kategorijų susijusios su lankstesniais vartojimo pasirinkimais, todėl jų diferenciacija tikėtina susijusi su skirtingu disponuojamų pajamų lygiu, ekonominiu aktyvumu bei gyvenimo ciklo ypatumais. Ryškiausi skirtumai nuosekliai nustatyti tarp 2 ir 5 klasterių. Bendrų vartojimo išlaidų kategorijoje tarp šių grupių nustatytas didžiausias skirtumas ($KS = 0,629$; $p < 0,001$), analogiška tendencija stebima restoranų ir apgyvendinimo paslaugų ($KS = 0,659$; $p < 0,001$) bei informacijos ir ryšių kategorijose ($KS = 0,650$; $p < 0,001$). Tai rodo, kad pensinio amžiaus namų ūkiai ir jaunos ekonomiškai aktyvios šeimos pasižymi iš esmės skirtingais vartojimo modeliais. Tikėtina, kad šie skirtumai atspindi ne tik pajamų lygio, bet ir gyvenimo būdo bei technologinio aktyvumo skirtumus.

Kita vertus, dalis socialinių grupių tam tikrose vartojimo kategorijose pasižymi panašiais vartojimo išlaidų pasiskirstymais. Ryškiausiai tai pastebima tarp 1 ir 6 klasterių, kur poilsio, sporto ir kultūros ($p = 0,852$), aprangos ir avalynės ($p = 0,169$), būsto apstatymo ($p = 0,258$) bei asmens priežiūros išlaidų kategorijose pagrindo atmesti nulinės hipotezės apie vienodus skirstinius nebuvo. Nors šios grupės iš esmės skiriasi išsilavinimo, kvalifikacijos bei pajamų lygiu, panašūs rezultatai leidžia daryti prielaidą, kad atskirose vartojimo srityse išlaidų struktūrą gali lemti ne vien ekonominiai veiksniai, bet ir panašūs vartojimo prioritetai. Panaši tendencija stebima ir tarp 3 bei 6 klasterių. Maisto ir nealkoholinių gėrimų kategorijoje nebuvo pagrindo atmesti nulinės hipotezės apie vienodus skirstinius ($KS p = 0,392$; $AD p = 0,228$). Atsižvelgiant į tai, kad abi grupės pasižymi aukštu ekonominiu aktyvumu ir jose dominuoja namų ūkiai be išlaikomų vaikų, galima daryti prielaidą, kad būtinojo vartojimo struktūra tampa mažiau priklausoma nuo pajamų lygio ir labiau susijusi su gyvenimo būdo ypatumais. Mažiausias skirtumų lygis nustatytas alkoholinių gėrimų, tabako ir narkotikų bei sveikatos išlaidų kategorijose. Alkoholio kategorijoje daugumos klasterių porų atveju nebuvo pagrindo atmesti nulinės hipotezės apie vienodus skirstinius ($p > 0,05$), o reikšmingi skirtumai daugiausia buvo susiję su pensinio amžiaus namų ūkiais. Sveikatos išlaidose taip pat nustatyta daugiau statistiškai panašių grupių porų, ypač tarp 1–3, 2–5, 3–4 ir 5–6 klasterių. Tai leidžia teigti, kad dalis vartojimo kategorijų yra mažiau jautrios socialinei diferenciacijai ir gali būti labiau susijusios su individualiais poreikiais bei amžiaus ypatumais.

Gauti rezultatai patvirtina, kad identifikuoti klasteriai reprezentuoja ne atsitiktinį duomenų susiskaidymą, o vartojimo požiūriu skirtingas socialines grupes. Ryškiausi skirtumai pasireiškia diskrecinio vartojimo kategorijose, tuo tarpu būtinojo vartojimo srityse tarp skirtingų socialinių grupių dažniau stebimi panašūs vartojimo modeliai.

3.4.2. Entropija grindžiamo ryšio metodo taikymas socialinių grupių vartojimo išlaidų analizei

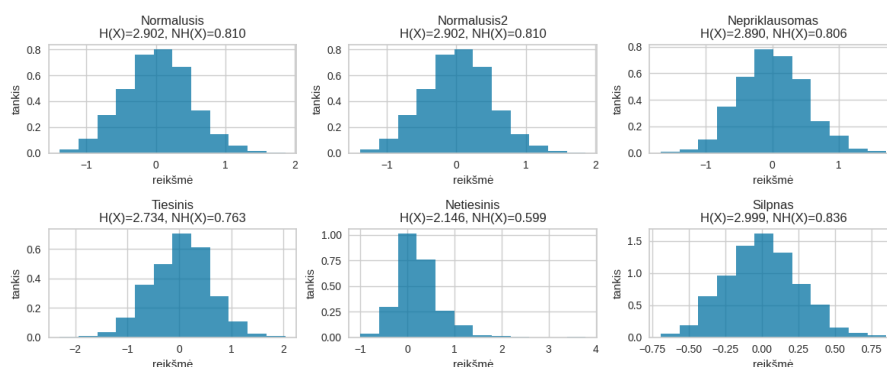
3.4.2.1. Entropija grindžiamo ryšio metodo koncepcinis patikrinimas

Siekiant empiriškai įvertinti **entropija grindžiamo ryšio metodo** gebėjimą identifikuoti skirtingo stiprumo ir formos priklausomybes tarp vartojimo išlaidų kintamųjų, pirmiausia atliktas koncepcinis metodo patikrinimas naudojant dirbtinai generuotus duomenis, kurių priklausomybės struktūra žinoma iš anksto. Simuliacijoms naudotas fiksuotas atsitiktinių skaičių generatoriaus pradinis parametras $seed = 42$, kiekvienam scenarijui generuojant po 1000 stebinių. Entropijos, tarpusavio informacijos ir bendrojo informacinio ryšio skaičiavimui taikytas histograminis diskretizavimas su 12 intervalų, o skaitinio stabilumo užtikrinimui taikytas $\varepsilon = 10^{-12}$ korekcinis parametras. Metodo

validavimas atliktas dviem etapais. Pirmajame etape buvo analizuoti simetriški, normaliuoju skirstiniu generuoti duomenys, o antrajame – asimetriški, eksponentiniu skirstiniu generuoti duomenys. Abiem atvejais buvo suformuoti penki priklausomybės scenarijai, remiantis (31) formule: identiška priklausomybė, nepriklausomumas, stipri linijinė priklausomybė, nelinijinė priklausomybė bei silpna stochastinė priklausomybė.

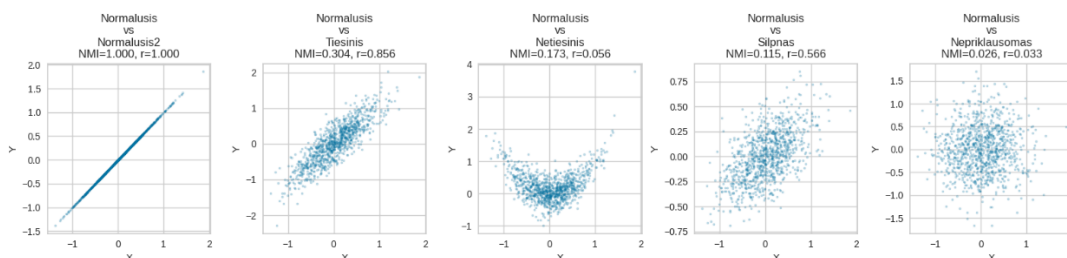
Simetriškų (normaliuoju) duomenų validavimas

Pirmiausia buvo sugeneruoti normaliuoju skirstiniu paremti duomenys. Apskaičiuotos individualios Šenono entropijos rodo, kad normalieji duomenys ir tiesinės priklausomybės scenarijuje sugeneruoti duomenys pasižymi aukščiausiomis normalizuotos entropijos reikšmėmis ($NH = 0,763-0,836$), tuo tarpu netiesinis scenarijus generuoja mažesnę normalizuotą entropiją ($NH = 0,599$), rodančią didesnę informacijos koncentraciją ir mažesnę reikšmių išsibarstymą (žr. 11 pav.).



11 pav. Normaliojo skirstinio imčių histogramos ir entropijos rodikliai

Porinių priklausomybių rezultatai rodo, kad identiškų kintamųjų poroje nustatoma maksimali priklausomybė ($NMI = 1,000$; $r = 1,000$), tai patvirtina metodo teorinį nuoseklumą (žr. 12 pav.). Tiesinės priklausomybės atveju nustatyta sąlyginai aukšta tarpusavio informacija ($NMI = 0,304$), atitinkanti stiprią Pearson koreliaciją ($r = 0,856$). Tuo tarpu netiesinės priklausomybės scenarijuje Pearson koreliacija tampa labai silpna ($r = 0,056$), tačiau entropinis ryšio metodas vis tiek identifikuoja aiškią informacinę priklausomybę ($NMI = 0,173$), tai rodo, kad entropinis metodas gali identifikuoti priklausomybes, kurių klasikiniai tiesiniai koreliacijos matai gali neatskleisti. Silpnos priklausomybės scenarijuje nustatyta $NMI = 0,115$, o nepriklausomų duomenų poroje informacinė priklausomybė išlieka labai maža ($NMI = 0,026$).



12 pav. Normaliojo skirstinio porinių ryšių rezultatai

Papildomai atliktas metodo validavimas daugiamatėje erdvėje, vertinant bendrąją informacinę ryšį (TC). Rezultatai rodo, kad beveik nepriklausomame trijų dimensijų (toliau – 3D) scenarijuje normalizuota bendroji informacija siekia tik 0,089, mišriame 3D scenarijuje padidėja iki 0,139, priklausomame 3D scenarijuje – iki 0,183, o priklausomame keturių dimensijų (toliau – 4D)

scenarijuje pasiekia 0,217 (žr. 7 lentelę). Tai rodo, kad nagrinėtuose scenarijuose didėjant priklausomų dimensijų skaičiui nustatyta didesnė bendros informacinės struktūros reikšmė.

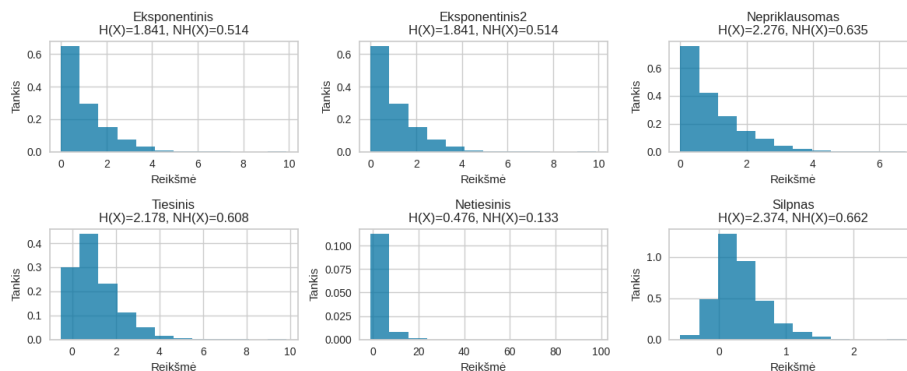
21 priede pateikiami papildomi metodo jautrumo ir stabilumo analizės rezultatai, leidžiantys įvertinti siūlomo entropija grindžiamojo ryšio metodo atsparumą skirtingų parametrizacijų bei pakartotinių simuliacijų atžvilgiu. Atlikta jautrumo analizė, vertinant histograminio diskretizavimo pasirinkimo įtaką normalizuotos tarpusavio informacijos įverčiams, keičiant intervalų skaičių nuo 8 iki 20. Gauti rezultatai rodo, kad visų priklausomybės scenarijų atveju NMI reikšmės išlieka santykinai stabilios, o identišκών kintamųjų poroje NMI nekinta ir visame testuotame intervaliniame diapazone išlieka lygi 1. Tai leidžia teigti, kad siūlomas metodas pasižymi mažu jautrumu histograminio diskretizavimo parametrų pasirinkimui. Taip pat metodo stabilumui įvertinti taikyta 100 pakartojimų simuliacija, paremta *Monte Carlo* principu. Gauti rezultatai rodo nedidelę informacinių įverčių sklaidą tarp pakartojimų: tiesinės priklausomybės atveju nustatytas $\overline{NMI} = 0,308$ ($\sigma = 0,012$), netiesinės – $\overline{NMI} = 0,176$ ($\sigma = 0,012$), silpnos priklausomybės – $\overline{NMI} = 0,118$ ($\sigma = 0,010$), o nepriklausomų kintamųjų atveju – $\overline{NMI} = 0,027$ ($\sigma = 0,003$). Identišκών kintamųjų scenarijuje NMI visų simuliacijų metu išliko lygi 1, o standartinis nuokrypis praktiškai lygus nuliui. Nedidelė rezultatų variacija tarp pakartotinių eksperimentų empiriškai patvirtina aukštą siūlomo metodo statistinį stabilumą ir atsparumą atsitiktinių imčių variacijai.

7 lentelė. Normaliojo skirstinio bendrieji informaciniai ryšiai

Atvejis	Imties tipai	H(X _i) suma	H(X _{1,...,X_d})	TC	Normalizuota TC
Beveik nepriklausomi 3D	Normalusis; Nepriklausomas; Silpnas	8,790	8,012	0,778	0,089
Miksuoti 3D	Normalusis; Tiesinis; Nepriklausomas	8,526	7,345	1,181	0,139
Priklausomi 3D	Normalusis; Tiesinis; Netiesinis	7,782	6,357	1,426	0,183
Priklausomi 4D	Normalusis; Tiesinis; Netiesinis; Silpnas	10,781	8,446	2,335	0,217

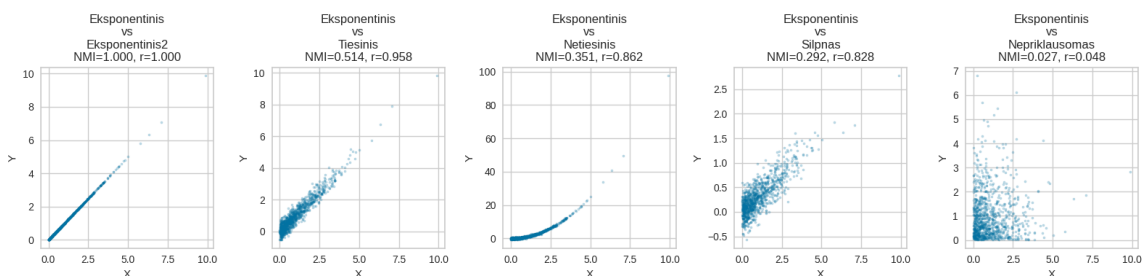
Asimetriškų (eksponentinių) duomenų validavimas

Identiška procedūra taikoma eksponentiniu skirstiniu generuotiems, stipriai asimetriškiems duomenims. Individualios entropijos analizė rodo mažesnes normalizuotos entropijos reikšmes nei simetriškų duomenų atveju (NH = 0,194–0,665), tai rodo didesnę informacijos koncentraciją ir mažesnę reikšmių išsibarstymą (žr. 13 pav.).



13 pav. Eksponentinio skirstinio imčių histogramos ir entropijos rodikliai

Nepaisant skirstinio asimetrijos, metodas išlaiko gebėjimą identifikuoti priklausomybės struktūrą. Identiškų kintamųjų poroje išlieka maksimali priklausomybė ($NMI = 1$), tiesinės priklausomybės atveju nustatyta $NMI = 0,514$, netiesinės – $NMI = 0,351$, silpnos – $NMI = 0,298$, o nepriklausomų kintamųjų poroje informacinė priklausomybė išlieka minimali ($NMI = 0,027$) (žr. 14 pav.). Pažymėtina, kad netiesinės priklausomybės scenarijuje metodas geba identifikuoti stiprią informacinę priklausomybę net ir esant stipriai netiesiniam ir asimetriškam priklausomybės pobūdžiui.



14 pav. Ekspontentinio skirstinio porinių ryšių rezultatai

Daugiamatė analizė taip pat parodė nuoseklų TC didėjimą: nuo 0,119 beveik nepriklausomame 3D scenarijuje iki 0,329 priklausomame 4D scenarijuje (žr. 8 lentelę).

8 lentelė. Ekspontentinio skirstinio bendrieji informaciniai ryšiai

Atvejis	Imties tipai	$H(X_i)$ suma	$H(X_1, \dots, X_d)$	TC	Normalizuota TC
Beveik nepriklausomi 3D	Ekspontentinis; Nepriklausomas; Silpnas	6,162	5,430	0,732	0,119
Miksuoti 3D	Normalusis; Tiesinis; Nepriklausomas	6,025	4,821	1,204	0,200
Priklausomi 3D	Normalusis; Tiesinis; Netiesinis	4,950	3,219	1,731	0,350
Priklausomi 4D	Normalusis; Tiesinis; Netiesinis; Silpnas	7,335	4,919	2,416	0,329

21 priede pateikiami metodo jautrumo histograminio diskretizavimo parametrams bei statistinio stabilumo rezultatai. Histogramų intervalų jautrumo analizė atspindi aukštą metodo stabilumą visame testuotame intervalų diapazone. Taip pat 100 pakartojimų simuliacija rodo, kad tiesinės priklausomybės scenarijuje nustatytas $\overline{NMI} = 0,507$ ($\sigma = 0,018$), netiesinės priklausomybės atveju $\overline{NMI} = 0,477$ ($\sigma = 0,070$), silpnos priklausomybės $\overline{NMI} = 0,273$ ($\sigma = 0,018$), o nepriklausomų kintamųjų scenarijuje informacinė priklausomybė išlieka minimali ($\overline{NMI} = 0,0268$; $\sigma = 0,0040$). Nors netiesinės priklausomybės scenarijuje stebima didesnė rezultatų sklaida, NMI reikšmės nuosekliai išlieka didesnės nei nepriklausomų kintamųjų atveju, todėl metodas išlaiko gebėjimą identifikuoti ir sudėtingesnes netiesines priklausomybės struktūras.

Gauti rezultatai leidžia daryti išvadą, kad siūlomas entropija grindžiamas metodas nuosekliai ir statistiškai stabiliai identifikuoja priklausomybių struktūras tiek simetriškai, tiek asimetriškai pasiskirsčiusiuose skirstiniuose. Papildomai nustatytas mažas jautrumas histogramų diskretizavimo parametrams bei aukštas Monte Carlo stabilumas leidžia pagrįsti metodo taikymą realių namų ūkių vartojimo išlaidų duomenų analizei.

3.4.2.2. Entropija grindžiama vartojimo išlaidų struktūrų analizė

Entropija grindžiama vartojimo išlaidų struktūrų analizė atliekama remiantis dviem rodikliais – normalizuota entropija (NH) bei normalizuota tarpusavio informacija (NMI). Šiuo atveju normalizuota entropija leidžia įvertinti vartojimo išlaidų pasiskirstymo neapibrėžtumą ir vidinį heterogeniškumą socialinėse grupėse, o normalizuota tarpusavio informacija atskleidžia, kiek skirtingos vartojimo kategorijos dalijasi bendra informacija tarpusavyje. Detalūs NH rezultatai pateikiami 22 priede, o NMI rezultatai – 23 priede. Rezultatų interpretacija pagrįsta gautais validavimo rezultatais: nepriklausomų kintamųjų atveju vidutinė NMI reikšmė siekia tik apie 0,027, silpnų ryšių atveju – apie 0,12–0,27, netiesinių ryšių – apie 0,18–0,48, o stiprių tiesinių priklausomybių atveju gali viršyti 0,50 ar net artėti prie 1. Todėl šiame skyriuje ryšiai, kurių NMI mažesnė nei 0,01, nėra interpretuojami kaip reikšmingi tarpusavio ryšiai, nes tokios reikšmės praktiškai atitinka nepriklausomų kintamųjų lygį. Taip pat vertinant rezultatus būtina atsižvelgti į tai, kad didesnės NMI reikšmės ne visada reiškia stiprią ekonominę priklausomybę – kai kuriais atvejais jos gali būti nulemtos didelės nulinių reikšmių koncentracijos tam tikrose išlaidų kategorijose. Be to, tarpusavio informacija apibūdina tik ryšio stiprumą, tačiau neatskleidžia jo krypties, todėl ekonomiškai netikėtų sąsajų atveju vienos kategorijos išlaidos kitą gali veikti tiek teigiamai, tiek neigiamai.

Analizuojant bendrąsias vartojimo išlaidas pirmiausia matyti, kad didžiausiu neapibrėžtumu pasižymi 1 grupė – jaunos mažesnių pajamų šeimos su vaikais (NH = 0,770) ir 5 grupė – jaunos ekonomiškai aktyvios šeimos (NH = 0,760). Šių grupių vartojimo išlaidų struktūros pasižymi didesne vidine įvairove ir mažesniu homogeniškumu. Tikėtina, kad tai susiję su skirtingais gyvenimo ciklo etapais, vaikų poreikiais bei nevienodu pajamų paskirstymu. Tuo tarpu mažiausias bendras vartojimo neapibrėžtumas nustatytas 4 grupėje – socialiai pažeidžiamuose namų ūkiuose (NH = 0,5050), 6 grupėje – aukštos kvalifikacijos ir aukštų pajamų namų ūkiuose (NH = 0,509) bei 3 grupėje – vidutinių–aukštesnių pajamų dirbančiuose namų ūkiuose (NH = 0,522), kas rodo didesnę vartojimo išlaidų struktūrų panašumą tarp tos pačios grupės namų ūkių ir labiau panašias vartojimo galimybes.

Ryškus neapibrėžtumas nustatytas maisto ir nealkoholinių gėrimų kategorijoje. Didžiausios NH reikšmės nustatytos 5 grupėje (NH = 0,796), 1 grupėje (NH = 0,736) ir 4 grupėje (NH = 0,694). Kadangi maistas laikomas būtinojo vartojimo kategorija, tokios aukštos reikšmės nėra visiškai įprastos. Vartojimo ryšių analizė neparodė nuoseklaus maisto išlaidų ryšių modelio tarp grupių. Nors 4 grupėje nustatytas santykinai stipresnis ryšys tarp maisto bei asmens priežiūros išlaidų (NMI = 0,218), kitose grupėse maisto kategorijos sąsajos su kitomis vartojimo sritimis dažniausiai išliko itin silpnos. Tai leidžia daryti prielaidą, kad didesnė entropija šiuo atveju labiau atspindi didesnę vartojimo išlaidų pasiskirstymo nevienodumą grupės viduje, o ne nuoseklų tarpkategorinį vartojimo modelį.

Priešinga tendencija stebima būsto, vandens, elektros, dujų ir kito kuro kategorijoje, kur nustatytas santykinai mažas vartojimo neapibrėžtumas. Pensinio amžiaus namų ūkių grupėje NH siekia tik 0,069, o aukštų pajamų ir aukštos kvalifikacijos grupėje – 0,124. Tokios reikšmės rodo itin mažą vidinę šių išlaidų struktūrų įvairovę. Kadangi būsto išlaidos paprastai sudaro reikšmingą bendrų vartojimo išlaidų dalį, būtų galima tikėtis didesnio išlaidų pasiskirstymo variantiškumo, todėl gauti rezultatai, tarp šių grupių namų ūkių, gali būti siejami su panašiomis būsto išlaidų struktūromis arba tam tikrų išlaidų komponentų dominavimu. Be to, aukštų pajamų ir aukštos kvalifikacijos grupėje būsto išlaidų kategorija nepasižymėjo stipriais ryšiais su kitomis vartojimo sritimis – didžiausias

ryšys fiksuotas su poilsio, sporto ir kultūros kategorija siekė tik $NMI = 0,0677$, todėl ryškesnių sisteminių priklausomybių nenustatyta.

Transporto kategorijoje taip pat stebimas neįprastai mažas neapibrėžtumas – 4 grupėje NH siekia tik 0,062, o 3 grupėje – 0,098. Tačiau šiuo atveju atsiranda tam tikras neatitikimas tarp entropijos ir tarpusavio informacijos rezultatų. Nors vidinė variacija maža, kai kuriose grupėse transporto išlaidos pasižymi sąlyginai dideliu ryšio stiprumu su kitomis kategorijomis. Jaunų ekonomiškai aktyvių šeimų grupėje transporto ryšys (NMI) su poilsio išlaidomis siekia 0,125, su draudimo ir finansinėmis paslaugomis – 0,122, o su restoranų ir apgyvendinimo išlaidomis – 0,111. Kadangi NMI nenusako ryšio krypties, negalima teigti, kad didesnės transporto išlaidos lemia didesnes ar mažesnes kitų kategorijų išlaidas. Tokie rezultatai tik rodo, kad šios kategorijos pasiskirsto panašiai ir gali būti veikiamos bendrų veiksnių, pavyzdžiui, pajamų ar gyvenimo būdo skirtumų.

Dar vienas nestandartinis rezultatas stebimas švietimo paslaugų kategorijoje. Visose grupėse nustatytos labai mažos NH reikšmės – mažiausia pensinio amžiaus namų ūkiuose ($NH = 0,005$), o didžiausia jaunų ekonomiškai aktyvių šeimų grupėje ($NH = 0,192$). Daugelyje grupių ši kategorija taip pat pasižymėjo itin silpnais ryšiais su kitomis vartojimo sritimis. Dalis ryšių net nesiekė 0,01 ribos, todėl nebuvo laikomi reikšmingais. Tačiau socialiai pažeidžiamų namų ūkių grupėje nustatytas santykinai aukštas ryšys tarp švietimo ir poilsio išlaidų ($NMI = 0,170$), kuris ekonomiškai nėra lengvai interpretuojamas. Atsižvelgiant į tai, kad švietimo išlaidose dažnai stebima daug nulinių reikšmių, negalima atmesti galimybės, kad toks rezultatas iš dalies susiformavo dėl išlaidų nebuvimo modelių, o ne dėl realios ekonominės priklausomybės.

Bendrai nustatyta, kad didžiausias vartojimo struktūrų neapibrėžtumas būdingas jaunų šeimų grupėms, o mažiausias – socialiai pažeidžiamiems, pensinio amžiaus bei aukštų pajamų namų ūkiams. Tuo pačiu pastebėta, kad didesnis vartojimo neapibrėžtumas nebūtinai reiškia stipresnius ryšius tarp išlaidų kategorijų. Dalis aukštesnių NMI reikšmių buvo nustatytos ekonomiškai ne visada intuityviose kategorijose, todėl rezultatus būtina vertinti atsargiai, ypač tais atvejais, kai galimą poveikį galėjo turėti didelė nulinių išlaidų koncentracija arba tai, kad tarpusavio informacija neleidžia nustatyti ryšio krypties.

Papildomai atlikta 3–4 vartojimo kategorijų derinių daugiamatė tarpusavio informacijos analizė parodė, kad ryškesni bendri tarpusavio informacijos ryšiai ($NTC > 0,2$) buvo nustatyti tik vidutinių–aukštesnių pajamų dirbančių namų ūkių ir socialiai pažeidžiamų namų ūkių grupėse. Nustatytos normalizuotos bendrosios tarpusavio informacijos reikšmės svyravo nuo 0,203 iki 0,274, kas nurodo vidutinio stiprumo ryšius tarp kelių vartojimo kategorijų. Didžiausia reikšmė nustatyta vidutinių–aukštesnių pajamų dirbančių namų ūkių grupėje ($NTC = 0,274$), o socialiai pažeidžiamų namų ūkių grupėje reikšmės buvo šiek tiek mažesnės ($NTC = 0,203–0,215$). Abiejose grupėse stipriausiuose deriniuose dažniausiai kartojosi maisto ir nealkoholinių gėrimų, būsto apstatymo ir namų priežiūros, draudimo ir finansinių paslaugų, asmens priežiūros, aprangos bei informacijos ir ryšių išlaidų kategorijos, todėl galima daryti prielaidą, kad būtent šios vartojimo sritys sudaro labiausiai tarpusavyje susijusią vartojimo struktūros dalį. Nors nustatyti stipresni tarpusavio informacijos ryšiai daugeliu atvejų ekonomiškai atrodo pagrįsti ir dažniausiai apima kasdienio vartojimo bei gyvenimo būdo kategorijas, rezultatai turėtų būti interpretuojami atsargiai. Išlieka NMI atveju aptarti apribojimai, susiję su nulinių išlaidų koncentracija bei tarpusavio informacijos metodo interpretavimo ypatumais.

3.5. Socialinių grupių vartojimo išlaidų tyrimo apibendrinimas, apribojimai ir rekomendacijos

Atliktame tyrimo socialinių grupių identifikavimui taikytas daugiamatis požiūris, leidžiantis vienu metu vertinti tarpusavyje susijusias namų ūkių socialines, demografines ir ekonomines charakteristikas. Taikant mišrių duomenų klasterizavimo metodus buvo identifikuotos šešios skirtingos socialinės grupės: jaunos mažesnių pajamų šeimos su vaikais, pensinio amžiaus namų ūkiai, vidutinių–aukštesnių pajamų dirbantys namų ūkiai, socialiai pažeidžiami namų ūkiai, jaunos ekonomiškai aktyvios šeimos bei aukštos kvalifikacijos ir aukštų pajamų namų ūkiai. Didžiausią indėlį į grupių diferenciaciją turėjo dirbančių asmenų dalis, aukštojo išsilavinimo dalis, aukštos kvalifikacijos užimtųjų dalis, išlaikomų vaikų skaičius namų ūkyje ir namų ūkio pagrindinis pajamų šaltinis.

Atlikta klasterių validacija leidžia teigti, kad identifikuota struktūra nėra tik atsitiktinio duomenų susiskaidymo pasekmė. Gauti rezultatai parodė labai aukštą klasterių atkūrimo tikslumą – kryžminės validacijos vidutinis tikslumas siekė 0,996, o nepriklausomos testavimo imties tikslumas sudarė 0,993. Papildomas vertinimas taikant skirtingas mokymo ir testavimo imčių skaidymo realizacijas parodė, kad rezultatai išliko stabilūs ir nebuvo jautrūs konkrečios testavimo imties sudėčiai. Tai rodo, kad identifikuotos socialinės grupės išsaugo aiškią vidinę struktūrą ir gali būti laikomos potencialiai socialinę diferenciaciją atspindinčiomis grupėmis.

Vartojimo išlaidų analizė atskleidė, kad grupės skiriasi ne tik bendromis išlaidų apimtimis, bet ir vidiniu vartojimo neapibrėžtumu. Didžiausias vartojimo heterogeniškumas nustatytas jaunų šeimų grupėse. Maisto ir nealkoholinių gėrimų kategorijoje didžiausios normalizuotos entropijos reikšmės nustatytos jaunų ekonomiškai aktyvių šeimų grupėje (NH = 0,7955), jaunų mažesnių pajamų šeimų grupėje (NH = 0,7355) ir socialiai pažeidžiamų namų ūkių grupėje (NH = 0,6935). Tuo tarpu pensinio amžiaus namų ūkių vartojimo struktūra pasižymėjo didesniu homogeniškumu (NH = 0,4341). Tyrimo rezultatai taip pat parodė, kad pensinio amžiaus, socialiai pažeidžiami ir aukštų pajamų namų ūkiai pasižymi mažesniu vartojimo išlaidų pasiskirstymo įvairumu grupės viduje, todėl jų išlaidų pasiskirstymas yra mažiau įvairus.

Šio tyrimo rezultatai gali būti naudingi formuojant socialinės politikos priemones bei vertinant skirtingų gyventojų grupių jautrumą ekonominiams sukrėtimams. Socialiai pažeidžiami namų ūkiai ir jaunos mažesnių pajamų šeimos su vaikais gali būti laikomos vienomis jautriausių grupių ekonominio nestabilumo sąlygomis, nes jų vartojimo struktūroje didesnę reikšmę turi būtinosios išlaidos, o diskrecinio vartojimo galimybės yra ribotos. Gauti rezultatai leidžia manyti, kad šioms grupėms gali būti ypač aktualios tikslinės pajamų palaikymo, būsto išlaidų kompensavimo ir šeimų rėmimo priemonės. Lietuvoje jau taikomos priemonės, skirtos mažinti socialinę atskirtį ir materialinį nepriteklių, pavyzdžiui, socialinės kortelės¹⁰ sunkiau besiverčiantiems gyventojams, būsto šildymo išlaidų kompensavimo parama¹¹ bei papildomos šeimų gerovės stiprinimo priemonės¹². Tačiau gauti rezultatai rodo, kad vien pajamų lygio vertinimas ne visada leidžia pakankamai tiksliai identifikuoti namų ūkių poreikius, kadangi skirtingos socialinės grupės pasižymi nevienodais vartojimo modeliais ir vidine struktūra. Todėl identifikuotos socialinės grupės galėtų papildyti šiuo metu taikomus

¹⁰ <https://socmin.lrv.lt/lt/naujienos/apie-160-tukst-sunkiau-besiverciantiu-zmoniu-kovo-menesi-sulauks-pinigines-paramos-i-socialines-korteles-5Wb/>

¹¹ <https://vda.lrv.lt/lt/naujienos/naujas-irankis-sildymo-kompensacijos-gyventojus-pasieks-greiciau/>

¹² <https://socmin.lrv.lt/lt/naujienos/nuo-birzelio-naujoves-seimu-gerovei-60e/>

paramos skyrimo kriterijus, leidžiant tiksliau nukreipti pagalbą ne tik pagal pajamas, bet ir pagal namų ūkio socialinį profilį, ekonominį aktyvumą, vaikų skaičių bei vartojimo struktūrą. Pavyzdžiui, jaunoms mažesnių pajamų šeimoms su vaikais didesnę reikšmę galėtų turėti šeimos išlaidų mažinimo priemonės, susijusios su vaikų priežiūros ir ugdymo paslaugomis, tuo tarpu pensinio amžiaus namų ūkiams ar socialiai pažeidžiamoms grupėms svarbesnės galėtų būti tikslinės kompensavimo priemonės būtiniausioms išlaidoms padengti. Tuo tarpu jaunų ekonomiškai aktyvių šeimų grupėje nustatytas didelis vartojimo neapibrėžtumas ir didesnė finansinių įsipareigojimų našta leidžia daryti prielaidą, kad ekonominių svyravimų laikotarpiais ši grupė gali būti jautresnė palūkanų normų pokyčiams ar pajamų sumažėjimui. Šie rezultatai gali būti naudingi valstybės institucijoms planuojant prevencines ekonominio pažeidžiamumo mažinimo priemones, susijusias su būsto finansavimo sąlygų stabilizavimu, šeimų finansinio atsparumo didinimu bei laikinu pajamų netekimu susijusių rizikų mažinimu.

Kaip ir visi tyrimai, šis darbas turi keletą apribojimų, į kuriuos svarbu atsižvelgti interpretuojant gautus rezultatus. Tyrimas atliktas remiantis 2021 m. Lietuvos namų ūkių biudžetų statistinio tyrimo duomenimis, todėl identifikuotos socialinės grupės ir jų vartojimo išlaidų struktūros atspindi konkretaus laikotarpio situaciją ir gali būti jautrios tuo metu vyravusioms ekonominėms bei socialinėms aplinkybėms. Kadangi vartojimo struktūra gali kisti dėl infliacijos, darbo rinkos pokyčių, ekonominių ciklų ar kitų išorinių veiksnių, nustatyti dėsningumai nebūtinai išlieka nekintantys ilgesniu laikotarpiu. Nors naudoti oficialiosios statistikos duomenys užtikrina reprezentatyvumą ir aukštą duomenų kokybę, tyrime naudotas charakteristikų rinkinys neapima visų galimų vartojimo skirtumus lemiančių veiksnių, tokių kaip individualios nuostatos, vertybės, finansiniai lūkesčiai ar asmeniniai vartojimo prioritetai. Be to, identifikuotų socialinių grupių sudėtis priklauso nuo pasirinktų požymių, duomenų paruošimo sprendimų bei taikytų klasterinės analizės metodų, todėl taikant kitokius metodinius sprendimus galėtų būti gauti iš dalies skirtingi grupavimo rezultatai.

Tolimesniuose socialinių grupių ir jų namų ūkių vartojimo išlaidų tyrimuose tikslinga išplėsti analizę įtraukiant skirtingų laikotarpių duomenis ir vertinant socialinių grupių vartojimo struktūrų pokyčius laike. Taip pat rekomenduojama atlikti detalesnę vartojimo išlaidų analizę naudojant žemesnio lygmens COICOP klasifikacijos kategorijas, leidžiančias analizuoti ne tik plačias vartojimo sritis (pvz., maistą ar transportą), bet ir konkretesnes išlaidų subkategorijas, tokias kaip atskiri maisto produktai, transporto paslaugos ar laisvalaikio išlaidos.

Išvados

1. Socialinių grupių samprata siejama ne tik su individo priskyrimu tam tikrai socialinei kategorijai ar priklausymu grupei, kuriai būdingos bendros normos, vertybės, elgesio modeliai bei socialinis tapatinimasis su jos nariais, bet ir su panašia socialine padėtimi, apibrėžiama socialinių, demografinių bei ekonominių charakteristikų visuma. Šios charakteristikos taip pat laikomos vienais svarbiausių veiksnių, lemiančių vartojimo skirtumus tarp namų ūkių, iš kurių dažniausiai išskiriami pajamų, užimtumo, išsilavinimo, amžiaus bei namų ūkio sudėties veiksniai.
2. Duomenimis grįstam socialinių grupių identifikavimui gali būti taikomi skirtingi metodai – nuo duomenų transformavimu pagrįstų metodų iki algoritmų, tiesiogiai pritaikytų mišriems duomenims. Kadangi socialinių grupių formavimąsi apibūdina skirtingo tipo charakteristikos, jų identifikavimui svarbu taikyti metodus, gebančius vienu metu apdoroti tiek skaitinius, tiek kategorinius požymius, todėl mišraus tipo duomenų analizėje gali būti taikomi k-prototipų, k-medoidų ir hierarchinės klasterizacijos metodai naudojant Gower atstumo matricą. Tuo tarpu vartojimo išlaidų struktūrų skirtumų vertinimui gali būti taikomi klasikiniai statistiniai bei entropijos teorija grindžiami metodai, leidžiantys vertinti tiek grupių skirtumus, tiek sudėtingesnes priklausomybes tarp vartojimo kategorijų.
3. Parengta socialinių grupių identifikavimo ir vartojimo išlaidų struktūrų analizės metodologija, apimanti duomenų paruošimo, požymių atrankos, mišraus tipo duomenų klasterizavimo, klasterių kokybės vertinimo bei statistinių ir informacijos teorija grindžiamų metodų taikymą vartojimo struktūrų skirtumų analizei. Metodologija pritaikyta 2021 m. Lietuvos namų ūkių biudžetų tyrimo duomenims, apimantiems 4334 namų ūkius, kuriuose analizuoti 22 socialines, demografines ir ekonomines charakteristikas apibūdinantys kintamieji bei 14 vartojimo išlaidų kategorijų kintamųjų.
4. Taikant k-medoidų PAM metodą su optimizuota Gower atstumo matrica identifikuotos šešios socialinės grupės: jaunos mažesnių pajamų šeimos su vaikais, pensinio amžiaus namų ūkiai, vidutinių–aukštesnių pajamų dirbantys namų ūkiai, socialiai pažeidžiami namų ūkiai, jaunos ekonomiškai aktyvios šeimos bei aukštos kvalifikacijos ir aukštų pajamų namų ūkiai. Klasterizavimo kokybę patvirtino vidutinis silueto koeficientas (0,622), o papildoma validacija parodė aukštą klasterių atkūrimo tikslumą – nepriklausomos testavimo imties tikslumas siekė 0,993. Įvertinus identifikuotų socialinių grupių vartojimo išlaidų struktūras paaiškėjo, kad skirtingos grupės pasižymi nevienodais vartojimo elgsenos ir vidinio vartojimo pasiskirstymo ypatumais. Didžiausias vartojimo išlaidų neapibrėžtumas maisto ir nealkoholinių gėrimų kategorijoje nustatytas jaunų ekonomiškai aktyvių šeimų grupėje (NH = 0,796), jaunų mažesnių pajamų šeimų grupėje (NH = 0,736) bei socialiai pažeidžiamų namų ūkių grupėje (NH = 0,694), o pensinio amžiaus namų ūkių grupė pasižymėjo mažesniu vartojimo išlaidų pasiskirstymo neapibrėžtumu (NH = 0,434). Tuo tarpu tarpusavio informacijos rezultatai parodė, kad didesnis vartojimo neapibrėžtumas nebūtinai reiškia stipresnius ryšius tarp vartojimo kategorijų – jaunų ekonomiškai aktyvių šeimų grupėje transporto išlaidos pasižymėjo stipriausiais ryšiais su poilsio (NMI = 0,1247), draudimo ir finansinių paslaugų (NMI = 0,1218) bei restoranų ir apgyvendinimo išlaidomis (NMI = 0,1110), tačiau dalis kitų nustatytų ryšių išliko silpni arba ekonomiškai sunkiai interpretuojami.
5. Identifikuotos socialinės grupės leidžia išskirti skirtingu ekonominiu aktyvumu, pajamų lygiu bei socialiniu pažeidžiamumu pasižyminčius namų ūkius. Gauta grupių struktūra gali prisidėti prie socialinės politikos formavimo, sudarant prielaidas tiksliau vertinti gyventojų jautrumą ekonominiams sukrėtimams bei taikyti diferencijuotas socialinės paramos priemones.

Literatūros sąrašas

1. OFICIALIOS STATISTIKOS PORTALAS. *Individualaus vartojimo išlaidų pagal paskirtį klasifikatorius (COICOP 2018)* [interaktyvus]. Prieiga per: <https://osp.stat.gov.lt/individualaus-vartojimo-islaidu-pagal-paskirti-klasifikatorius-coicop>
2. LIETUVOS STATISTIKOS DEPARTAMENTAS. *Namų ūkių biudžetų statistinio tyrimo metodika*. Vilnius: 2021. PATVIRTINTA Lietuvos statistikos departamento generalinio direktoriaus 2021 m. liepos 8 d. įsakymu Nr. DĮ – 178. Prieiga per: https://osp.stat.gov.lt/documents/10180/130368/NUBT_metodika_2016_092.pdf
3. VISUOTINĖ LIETUVIŲ ENCIKLOPEDIJA. Socialinė grupė [interaktyvus]. Vilnius: Mokslo ir enciklopedijų leidybos centras. Prieiga per: <https://www.vle.lt/straipsnis/socialine-grupe/>
4. BURNS, Tom R.; CORTE, Ugo; MACHADO, Nora. A universal theory of social groups: the actor-system-dynamics approach to agents, rule regimes, and interaction processes. 2014. Prieiga per: https://www.researchgate.net/publication/259998269_A_UNIVERSAL_THEORY_OF_SOCIAL_GROUPS_The_Actor-System-Dynamics_Approach_to_Agents_Rule_Regimes_and_Interaction_Processes
5. CUTTER, S.L., BORUFF, B.J. and SHIRLEY, W.L. Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, 2003, vol. 84, no. 2. pp. 242–261. DOI 10.1111/1540-6237.8402002.
6. LEACH, Colin Wayne, et al. Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification. *Journal of personality and social psychology*, 2008, 95.1: 144. DOI: 10.1037/0022-3514.95.1.144
7. CHARNESS, Gary; CHEN, Yan. Social identity, group behavior, and teams. *Annual Review of Economics*, 2020, 12.1: 691-713. DOI: 10.1146/annurev-economics-091619-032800
8. NAZIR, Aasima, MIR, Showket Ahmad. Effects of Inflation on Household Consumption Expenditure: An Empirical Insights From The Indian Economy [interaktyvus]. *Indian Journal of Human Development*, 2025, 19.1: 118-134. DOI: 10.1177/09737030251346761
9. BROWNING, Martin; CROSSLEY, Thomas F.; WINTER, Joachim. The measurement of household consumption expenditures. *Annu. Rev. Econ.*, 2014, 6.1: 475-501. Prieiga per: <https://doi.org/10.1146/annurev-economics-080213-041247>
10. MIŠEIKIS, Feliksas; ZABULYTĖ, Loreta. Bendrojo vidaus produkto ekonometrinis modelis. *Ekonomika*, 2003, 64: 92-99. DOI: 10.15388/Ekon.2003.17327
11. KHURANA, J.; BEERO, S. K. Analysis of Economic Factors Affecting Household Private Final Consumption Expenditure in India. *International Journal of Science and Research (IJSR)*, 2023, vol. 12, no. 5, p. 44–49. Prieiga per: DOI: 10.21275/SR23306143350
12. DÖNMEZ, Gizem Acet; GÜNEŞ, Hurşit. Household consumption expenditures in turkey: A comparative quantile regression analysis. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 2021, 43.1: 179-196. Prieiga per: DOI: 10.14780/muiibd.960326
13. MOHANTY, Sanjay K., et al. Assessing the impact of measurement error in household consumption on estimates of catastrophic health expenditure in India. *Humanities and Social Sciences Communications*, 2023, 10.1: 722. DOI: 10.1057/s41599-023-02226-4
14. AYYASH, Mohsen; SEK, Siok Kun. Decomposing inequality in household consumption expenditure in Malaysia. *Economies*, 2020, 8.4: 83. DOI: 10.3390/economies8040083

15. CRAWFORD, M. Ecological Systems theory: Exploring the development of the theoretical framework as conceived by Bronfenbrenner. *J Pub Health Issue Pract*, 2020, 4.2: 170. Prieinama per: <https://raggeduniversity.co.uk/wp-content/uploads/2025/02/Crawford2020.pdf>
16. GONG, Yanping, et al. The relationship between work-to-family conflict and conspicuous consumption: an identity theory perspective. *Psychology research and behavior management*, 2023, 39-56. Prieiga per: <https://www.dovepress.com/article/download/80831>
17. JAPPELLI, Tullio; PAGANO, Marco. Saving, growth, and liquidity constraints. *The quarterly journal of economics*, 1994, 109.1: 83-109. Prieiga per: <https://doi.org/10.2307/2118429>
18. LIObIKIENĖ, Genovaitė; MATIUK, Yuliia; KRIKŠTOLAITIS, Ričardas. The concern about main crises such as the Covid-19 pandemic, the war in Ukraine, and climate change's impact on energy-saving behavior. *Energy Policy*, 2023, 180: 113678. Prieiga per: <https://doi.org/10.1016/j.enpol.2023.113678>
19. CHAI, Andreas; MONETA, Alessio. Retrospectives: engel curves. *Journal of Economic Perspectives*, 2010, 24.1: 225-240. DOI: 10.1257/jep.24.1.225
20. HANNAH, R. (2023) *Engel's Law: Richer people spend more money on food, but it makes up a smaller share of their income*. Our World in Data [interaktyvu]. Prieinama per: <https://ourworldindata.org/engels-law-food-spending>
21. DINKOVA, Milena; KALWIJ, Adriaan; ALESSIE, Rob. Know more, spend more? The impact of financial literacy on household consumption. *De Economist*, 2021, 169.4: 469-498. Prieiga per: <https://doi.org/10.1007/s10645-021-09391-4>
22. MA, Yingqi, et al. The Impact of Financial Literacy on Household Consumption. *Frontiers in Business, Economics and Management*, 2022, 5.3: 55-63. DOI: 10.54097/fbem.v5i3.1908
23. FUKUDA, Kenta. Job loss, consumption insurance, and household time allocation. *Labour Economics*, 2024, 91: 102645. Prieiga per: <https://doi.org/10.1016/j.labeco.2024.102645>
24. CAMPBELL, John Y.; COCCO, Joao F. How do house prices affect consumption? Evidence from micro data. *Journal of monetary Economics*, 2007, 54.3: 591-621. Prieiga per: <https://doi.org/10.1016/j.jmoneco.2005.10.016>
25. CARROL, Christopher D.; OTSUKA, Misuzu; SLACALEK, Jirka. How large is the housing wealth effect? A new approach. *NBER Working Paper*, 2006, 12746: 17. Prieiga per: <https://www.jstor.org/stable/20870039?seq=1>
26. DIAS DA SILVA, Alexandre; RUSINOVA, Desislava; WEIBLER, Matthias. *Consumption effects of job loss expectations: New evidence for the euro area*, ECB Working Paper, No. 2817, ISBN 978-92-899-6080-9, European Central Bank (ECB), Frankfurt a. M., Prieiga per: <https://doi.org/10.2866/06611>
27. YÜKSEL, Emre; BAŞAR, Dilek. Household Consumption Expenditures in Türkiye: Socio-Economic Determinants, Spending Patterns, and Policy Perspectives. *Ekonomi Politika ve Finans Araştırmaları Dergisi*, 2025, 10.2: 467-483. Prieiga per: <https://doi.org/10.30784/epfad.1661884>
28. BROWNING, Martin; CHIAPPORI, Pierre-Andre; LEWBEL, Arthur. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *Review of Economic Studies*, 2013, 80.4: 1267-1303. Prieiga per: <https://www.jstor.org/stable/43551559>
29. PAVLISA, Karina. Structures of consumption and professional identity: An analysis of the French household budget survey. *Sociology*, 2024, 58.6: 1451-1471. Prieiga per: <https://doi.org/10.1177/00380385241254310>

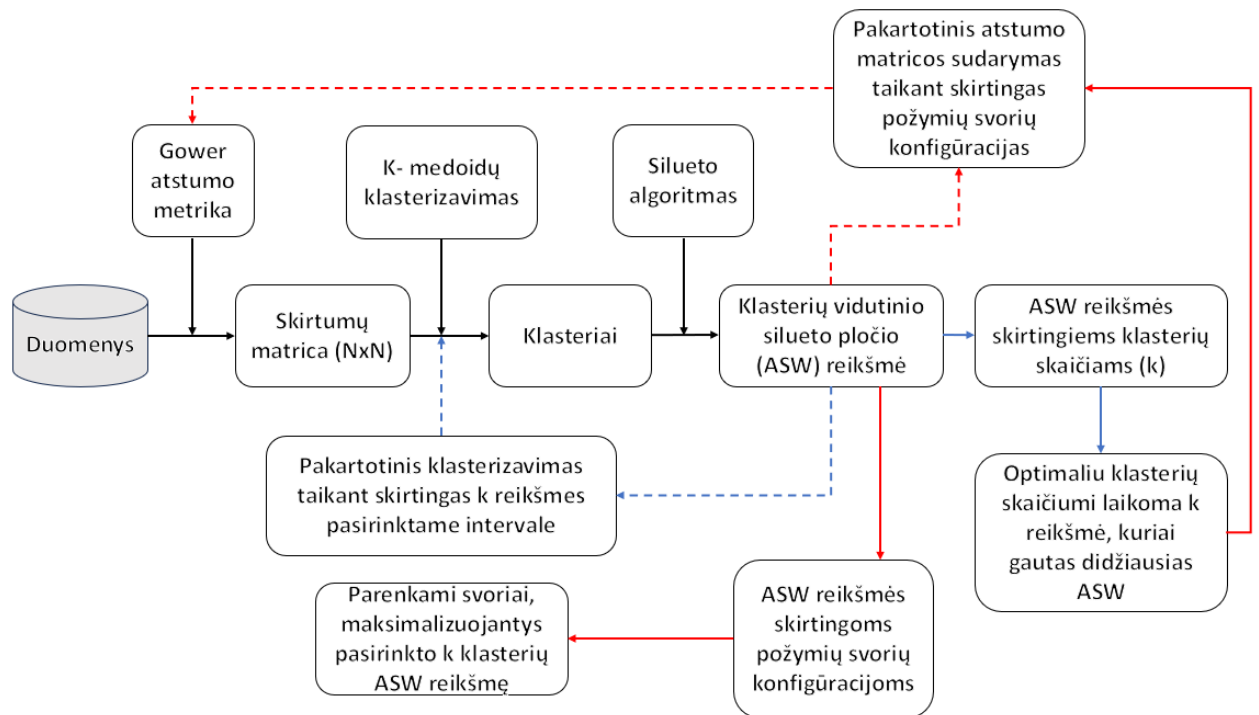
30. CRUWYS, Tegan, et al. Social Identity Mapping: A procedure for visual representation and assessment of subjective multiple group memberships. *British Journal of Social Psychology*, 2016, 55.4: 613-642. DOI: 10.1111/bjso.12155
31. HENNIG, Christian; LIAO, Tim F. How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 2013, 62.3: 309-369. Prieiga per: <https://www.jstor.org/stable/pdf/24771810.pdf>
32. HERR, Alexander. Statistics for categorical surveys—a new strategy for multivariate classification and determining variable importance. *Sustainability*, 2010, 2.2: 533-550. Prieiga per: <https://www.mdpi.com/2071-1050/2/2/533>
33. COSTA, Efthymios; PAPATSOUMA, Ioanna; MARKOS, Angelos. Benchmarking distance-based partitioning methods for mixed-type data. *arXiv preprint arXiv:2203.16287*, 2022. Prieiga per: <https://arxiv.org/pdf/2203.16287>
34. GHATTAS, Badih; SANCHEZ SAN-BENITO, Alvaro. Clustering Approaches for Mixed-Type Data: A Comparative Study. *Journal of Probability and Statistics*, 2025, 2025.1: 2242100. Prieiga per: <https://doi.org/10.1155/jpas/2242100>
35. PREUD'HOMME, Gregoire, et al. Head-to-head comparison of clustering methods for heterogeneous data: a simulation-driven benchmark. *Scientific reports*, 2021, 11.1: 4202. Prieiga per: <https://www.nature.com/articles/s41598-021-83340-8>
36. BEKTAS, Alperen; SCHUMANN, René. How to optimize gower distance weights for the k-medoids clustering algorithm to obtain mobility profiles of the swiss population. In: *2019 6th Swiss Conference on Data Science (SDS)*. IEEE, 2019. p. 51-56. DOI: 10.1109/SDS.2019.000-8
37. ASCHENBRUCK, Rabea; SZEPANNEK, Gero; WILHELM, Adalbert FX. Initialization strategies for clustering mixed-type data with the k-prototypes algorithm: R. Aschenbruck et al. *Advances in Data Analysis and Classification*, 2025, 1-30. Prieiga per: <https://link.springer.com/article/10.1007/s11634-025-00639-4>
38. ZAMAN, Mobasshira. Evaluation of household expenditure in the united states: pre-covid and post-covid statistical analysis. *International journal of research in industrial engineering*, 2023, 12.4: 388-396. Prieiga per: <https://doi.org/10.22105/riej.2023.383801.1364>
39. HRISTOV, Hristo, et al. European household spending and socio-economic impacts on food behavior during the first wave of COVID-19. *Frontiers in nutrition*, 2022, 9: 869091. Prieiga per: <https://doi.org/10.3389/fnut.2022.869091>
40. DUDEL, Christian; GARBUSZUS, Jan Marvin; SCHMIED, Julian. Assessing differences in household needs: a comparison of approaches for the estimation of equivalence scales using German expenditure data. *Empirical Economics*, 2021, 60.4: 1629-1659. DOI: 10.1007/s00181-020-01822-6
41. FAN, Chaozhi, et al. The new economic era analysis of the structure system of Chinese household consumption expenditure based on the ELES model. *Computational Intelligence and Neuroscience*, 2022, 2022.1: 3278194. Prieiga per: <https://doi.org/10.1155/2022/3278194>
42. LEE, En; ONG, Thian Song; LEE, Y. Evaluating household consumption patterns: Comparative analysis using ordinary least squares and random forest regression models. *HighTech and Innovation Journal*, 2024, 5.2: 489-507. Prieiga per: <http://dx.doi.org/10.28991/HIJ-2024-05-02-019>

43. PUTTANAPONG, Nattapong; LIM, Siphath. Predicting Household Expenditure Using Machine Learning Techniques: A Case of Cambodia. *Nakhara: Journal of Environmental Design and Planning*, 2024, 23.3: 421-421. DOI: 10.54028/NJ202423421
44. YOUNG, Alexander L., et al. Mutual information: Measuring nonlinear dependence in longitudinal epidemiological data. *Plos one*, 2023, 18.4: e0284904. DOI: 10.1371/journal.pone.0284904
45. HAN, Jiawei, PEI, Jian, TONG, Hanghang. *Data mining: concepts and techniques*. Morgan kaufmann, 2022, p. 55-79. Prieiga per: <https://www.sciencedirect.com/book/monograph/9780123814791/data-mining-concepts-and-techniques>
46. JAIN, Anil K. Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 2010, 31.8: 651-666. DOI 10.1016/j.patrec.2009.09.011
47. CAI, Deng; ZHANG, Chiyuan; HE, Xiaofei. Unsupervised feature selection for multi-cluster data. In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2010. p. 333-342. DOI: 10.1145/1835804.1835848
48. KAMALOV, F., SULIEMAN, H., ALZAATREH, A., EMARLY, M., CHAMLAL, H., & SAFARALIEV, M. (2025). *Mathematical Methods in Feature Selection: A Review. Mathematics*, 13(6), 996. Prieiga per: <https://doi.org/10.3390/math13060996>
49. KUHN, Max, et al. *Applied predictive modeling*. New York: Springer, 2013, p. 43-45. Prieiga per: <https://uistam.org/wp-content/uploads/2025/08/Applied-Predictive-Modeling.pdf>
50. BAAK, Max, et al. A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. *Computational Statistics & Data Analysis*, 2020, 152: 107043. Prieiga per: <https://doi.org/10.1016/j.csda.2020.107043>
51. GRANITSA, Y., KHUJAYEV, I. ir KADYROV, A. Cluster analysis of regional indicators using DBSCAN algorithm. *Dela Press Conference Series: Economics, Business and Management*, Vol. 001, 025 [interaktyvus]. 2022. DOI: 10.56199/dpcsebm.zzt14298
52. HUANG, Zhexue. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, 1998, 2.3: 283-304. Prieiga per: <https://link.springer.com/article/10.1023/a:1009769707641>
53. HENDERI, Henderi; WAHYUNINGSIH, Tri; RAHWANTO, Efana. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *International Journal of Informatics and Information Systems*, 2021, 4.1: 13-20. Prieiga per: <https://ijjis.org/index.php/IJIS/article/view/73/32>
54. KAUFMAN, Leonard; ROUSSEEUW, Peter J. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009. 32-37 p. DOI:10.1002/9780470316801
55. SCHUBERT, Erich; ROUSSEEUW, Peter J. Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. *Information Systems*, 2021, 101: 101804. Prieiga per: <https://doi.org/10.1016/j.is.2021.101804>
56. JAIN, Anil K.; MURTY, M. Narasimha; FLYNN, Patrick J. Data clustering: a review. *ACM computing surveys (CSUR)*, 1999, 31.3: 264-323. DOI:10.1145/331499.331504
57. DAY, William HE; EDELSBRUNNER, Herbert. Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of classification*, 1984, 1.1: 7-24. Prieiga per: <https://link.springer.com/article/10.1007/bf01890115>

58. MURTAGH, Fionn; CONTRERAS, Pedro. Algorithms for hierarchical clustering: an overview. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 2012, 2.1: 86-97. Prieiga per: <https://doi.org/10.1002/widm.53>
59. CALIŃSKI, Tadeusz; HARABASZ, Jerzy. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 1974, 3.1: 1-27. DOI: 10.1080/03610927408827101
60. FERNÁNDEZ, Alberto; GÓMEZ, Sergio. Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms. *Journal of Classification*, 2008, 25.1: 43-65. Prieiga per: <https://link.springer.com/article/10.1007/s00357-008-9004-x>
61. HARANTOVÁ, Veronika, et al. Two-step cluster analysis of passenger mobility segmentation during the COVID-19 pandemic. *Mathematics*, 2023, 11.3: 583. Prieiga per: <https://doi.org/10.3390/math11030583>
62. HUBERT, Lawrence; ARABIE, Phipps. Comparing partitions. *Journal of classification*, 1985, 2.1: 193-218. Prieiga per: <https://link.springer.com/article/10.1007/BF01908075>
63. CHEN, Tianqi; GUESTRIN, Carlos. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016. p. 785-794. Prieiga per: <https://dl.acm.org/doi/epdf/10.1145/2939672.2939785>
64. JOHNSON, Richard A.; WICHERN, Dean W. *Applied Multivariate Statistical Analysis*. 6th ed. Pearson, 2007. 296-308 p. Prieiga per: <https://www.stat.uchicago.edu/~lekheng/courses/331/books/Johnson-Wichern%20-%20Applied%20Multivariate%20Analysis.pdf>
65. MARCINKO, Tomáš. Consequences of assumption violations regarding one-way ANOVA. *Proceedings of the 8th International Days of Statistics and Economics*, 2014, 116.47: 974-985. Prieiga per: https://msed.vse.cz/static/msed_2014/article/342-Marcinko-Tomas-paper.pdf
66. SCHOLZ, Fritz W.; STEPHENS, Michael A. K-sample Anderson–Darling tests. *Journal of the American Statistical Association*, 1987, 82.399: 918-924. Prieiga per: <https://faculty.washington.edu/fscholz/Papers/ADk.pdf>
67. HODGES JR, J. L. The significance probability of the Smirnov two-sample test. *Arkiv för matematik*, 1958, 3.5: 469-486. <https://link.springer.com/article/10.1007/BF02589501>
68. WU, Yun-Chun; SHIH, Ming-Chieh; TU, Yu-Kang. Using normalized entropy to measure uncertainty of rankings for network meta-analyses. *Medical Decision Making*, 2021, 41.6: 706-713. Prieiga per: <https://doi.org/10.1177/0272989X21999023>
69. MCDAID, Aaron F.; GREENE, Derek; HURLEY, Neil. Normalized mutual information to evaluate overlapping community finding algorithms. *arXiv preprint arXiv:1110.2515*, 2011. Prieiga per: <https://arxiv.org/pdf/1110.2515>
70. WATANABE, Satoshi. Information theoretical analysis of multivariate correlation. *IBM Journal of research and development*, 1960, 4.1: 66-82. Prieiga per: DOI: 10.1147/rd.41.0066

Priedai

1 priedas. Gower atstumo svorių optimizavimo procedūros schema k-medoidų algoritme



Sudaryta pagal Bektasą ir Schumanną [37].

Metodas prasideda nuo pradinių duomenų rinkinio, kuriam apskaičiuojama Gower atstumo metrika. Kiekvienai objektų porai apskaičiuojamas jų tarpusavio nepanašumas ir sudaroma kvadratinė nepanašumo matrica $N \times N$. Sudaryta nepanašumo matrica perduodama k-medoidų klasterizavimo algoritmui, kuris suskirsto objektus į klasterius pagal tarpusavio panašumą. Suformavus klasterius, taikomas silueto algoritmas, leidžiantis įvertinti klasterizavimo kokybę. Šiame etape apskaičiuojamas vidutinis silueto plotis (ASW). Kuo ASW reikšmė didesnė, tuo kokybiškesniu laikomas klasterizavimo rezultatas.

Procedūra vykdoma dviem optimizavimo kryptimis:

- Modelis pakartotinai klasterizuojamas taikant skirtingas klasterių skaičiaus k reikšmes pasirinktame intervale. Kiekvienam variantui apskaičiuojama ASW reikšmė. Vėliau šios reikšmės palyginamos, o optimaliu klasterių skaičiumi laikoma ta k reikšmė, kuriai gaunama didžiausia vidutinio silueto pločio reikšmė.
- Nustačius optimalų klasterių skaičių, pereinama prie požymių svorių optimizavimo. Šiame etape pakartotinai formuojama Gower atstumo matrica, kiekvieną kartą taikant skirtingas požymių svorių konfigūracijas. Kiekvienai svorių kombinacijai atliekama klasterizacija ir apskaičiuojama ASW reikšmė. Galiausiai parenkamas toks svorių rinkinys, kuris maksimalizuoja pasirinkto klasterių skaičiaus ASW reikšmę.

Schemoje naudojamos trijų tipų rodyklės, žyminčios skirtingus proceso etapus:

1. Juoda → pagrindinis duomenų srautas.
2. Mėlyna → klasterių skaičiaus (k) paieška.
3. Raudona → požymių svorių optimizavimas.

2 priedas. K-prototipų algoritmo inicializavimo metodai naudojami pradinių centroidų padėčių nustatymui

Klasterizavimo kontekste, ypač taikant k-prototipų ir k-modų algoritmus, *Huang* ir *Cao* inicializavimo metodai naudojami pradinių centroidų padėčių nustatymui. Šių metodų tikslas – parinkti tinkamus pradinius klasterių centrus, kurie turėtų įtakos algoritmo konvergencijos eigai ir galutinei klasterizavimo kokybei.

Huang inicializavimo metodas centroidus formuoja remdamasis kategorinių požymių dažnių pasiskirstymu. Pirmasis centroidas parenkamas atsitiktinai iš stebėjimų aibės, o vėlesni centroidai atrenkami taip, kad pakankamai skirtųsi nuo jau pasirinktų centrų pagal požymių dažnius. Pagrindinė šio metodo idėja yra užtikrinti centroidų įvairovę ir išvengti pernelyg panašių pradinių klasterių centrų. Toks metodas ypač naudingas tais atvejais, kai kategoriniai požymiai turi daug galimų kategorijų ir siekiama užtikrinti didesnį pradinių sprendinių heterogeniškumą.

Cao inicializavimo metodas pasižymi sudėtingesne logika ir siekia parinkti tarpusavyje gerai atskirtus centroidus. Skirtingai nei *Huang* metodas, jis atsižvelgia į bendrą duomenų struktūrą bei tankio pasiskirstymą. Kiekvienam duomenų taškui apskaičiuojamas tankio rodiklis, įvertinantis tiek artimų stebėjimų skaičių tam tikrame atstume, tiek kategorinių požymių savybes. Didesnio tankio srityse esantys objektai turi didesnę tikimybę būti parinkti kaip pradiniai centroidai. Tokiu būdu siekiama nustatyti centroidus, reprezentuojančius pagrindines duomenų struktūros charakteristikas ir tankiausiai užpildytas duomenų sritis.

Abu metodai sukurti siekiant pagerinti k-prototipų ir k-modų algoritmų konvergenciją, užtikrinant kokybiškesnius pradinius sprendinius, geriau atspindinčius duomenų pasiskirstymą. *Cao* metodas dažniausiai laikomas stabilesniu ir atsparesniu atsitiktinumui, nes detalčiau įvertina bendrą duomenų struktūrą. Tuo tarpu *Huang* metodas paprastai pasižymi mažesnėmis skaičiavimo sąnaudomis ir gali veikti greičiau, tačiau kai kuriais atvejais gali prireikti daugiau iteracijų iki konvergavimo. Taikant šiuos metodus mišraus tipo duomenų klasterizavimui, inicializavimo strategijos pasirinkimas gali reikšmingai paveikti galutinį klasterizavimo rezultatą. Nors *Cao* metodas dažnai rekomenduojamas dėl didesnio stabilumo, galutinis pasirinkimas priklauso nuo konkrečių duomenų charakteristikų, todėl praktikoje tikslinga atlikti empirinį skirtingų inicializavimo metodų palyginimą.

3 priedas. (M)ANOVA metodo taikymo prielaidos

MANOVA metodo taikymas grindžiamas keliomis esminėmis prielaidomis:

1. Priklausomi kintamieji turi būti tolydūs, o grupės tarpusavyje nepriklausomos;
2. Tiesiškumas – tarp priklausomų kintamųjų turi egzistuoti tiesiniai ryšiai, kurie vertinami sudarant sklaidos diagramų matricą;
3. Daugiamatis normalumas – kiekvienoje grupėje priklausomų kintamųjų jungtinis pasiskirstymas turi būti artimas daugiamačiam normaliajam skirstiniui. Vienmačiam normalumui vertinti taikomas Šapiro-Vilko kriterijus kartu su $Q-Q$ grafikais, o daugiamačiam normalumui – *Mardia* asimetrijos ir eksceso kriterijai;
4. Homogeniškumas – visų grupių kovariacinės struktūros turi būti vienodos. Ši prielaida tikrinama taikant *Box M* kriterijų;
5. Daugiamačių išskirčių nebuvimas – stebėjimai neturi būti ekstremalūs daugiamatėje erdvėje, tikrinama vertinant Mahalanobio atstumą.

ANOVA metodo taikymas reikalauja šių prielaidų:

1. Stebėjimų nepriklausomumas, dažniausiai užtikrinamas tyrimo planu;
2. Normalumo prielaida – kiekvienos grupės paklaidos turi būti normaliai pasiskirsčiusios. Ši prielaida tikrinama Šapiro-Vilko, Kolmogorovo–Smirnov ir Andersono–Darlingo kriterijais;
3. Dispersijų homogeniškumas – visų grupių dispersijos turi būti lygios:
– $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$
– H_A : bent dvi dispersijos nelygio

Hipotezės tikrinamos Livyno arba Bartleto kriterijumi. Kadangi Livyno kriterijus mažiau jautrus normalumo pažeidimams, praktiniuose socioekonominiuose tyrimuose dažniausiai pasirenkamas būtent jis. Kai dispersijų homogeniškumo prielaida pažeidžiama, vietoje klasikinio Fišerio kriterijaus taikoma Velčo dispersinė analizė.

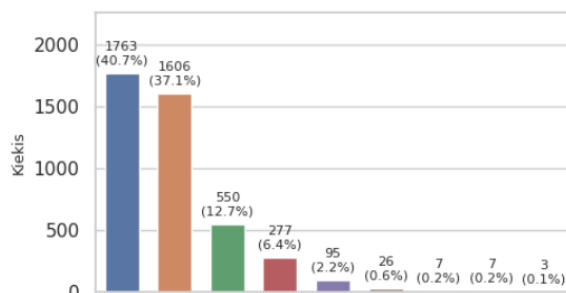
4 priedas. Namų ūkių lygmens rodiklių aprašomoji statistika

Duomenų rinkinį sudaro 4 334 namų ūkių stebiniai ir 17 stulpelių, iš kurių pirmasis naudojamas kaip identifikatorius, o likusieji 16 – analizei naudojami kintamieji:

1. Namų ūkio identifikatorius
2. Asmenų skaičius namų ūkyje
3. Gyvenamoji vieta
4. Regionas
5. Namų ūkio tipas
6. Išlaikomų vaikų skaičius namų ūkyje
7. Socialinė-ekonominė grupė
8. Namų ūkio pagrindinis pajamų šaltinis
9. Namų ūkio galvos amžius
10. Namų ūkio galvos išsilavinimas (grupuotas)
11. Pajamų kvintilinės grupės
12. Būsto tipas
13. Būsto valdos statusas
14. Namų ūkio užimamas naudingasis plotas
15. Centralizuotas šalto vandens tiekimas
16. Centralizuotas karšto vandens tiekimas
17. Pagrindinis būsto šildymo būdas

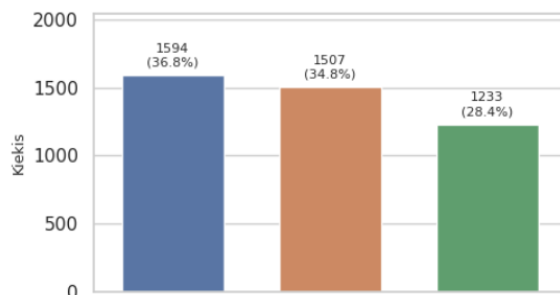
Kintamojo 'Asmenų skaičius namų ūkyje' struktūra

Kategorija	Kiekis	Procentai
1	1763	40.7
2	1606	37.1
3	550	12.7
4	277	6.4
5	95	2.2
6	26	0.6
7	7	0.2
8	7	0.2
9	3	0.1



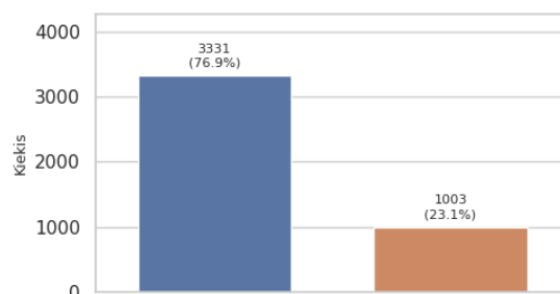
Kintamojo 'Gyvenamoji vieta' struktūra

Kategorija	Kiekis	Procentai
Kaimas	1594	36.8
5 didieji miestai	1507	34.8
Kiti miestai	1233	28.4



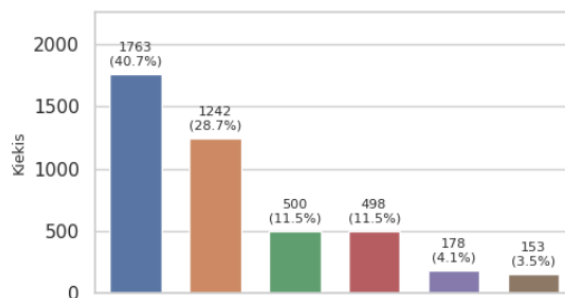
Kintamojo 'Regionas' struktūra

Kategorija	Kiekis	Procentai
Vidurio ir vakarų Lietuvos regionas	3331	76.9
Sostinės regionas	1003	23.1



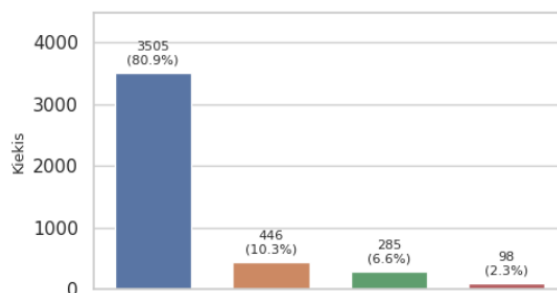
Kintamojo 'Namų ūkio tipas' struktūra

Kategorija	Kiekis	Procentai
Vienas gyvenantis asmuo	1763	40.7
Pora be išlaikomų vaikų	1242	28.7
Kiti namų ūkiai be išlaikomų vaikų	500	11.5
Pora su išlaikomais vaikais	498	11.5
Vienas suaugęs asmuo su išlaikomais vaikais	178	4.1
Kiti namų ūkiai su išlaikomais vaikais	153	3.5



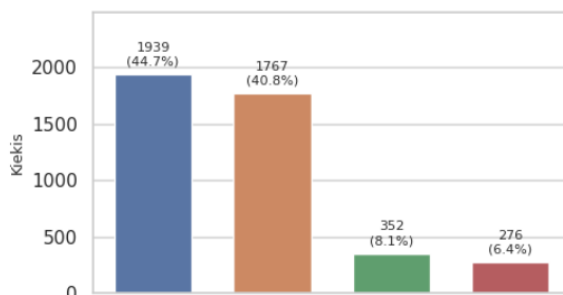
Kintamojo 'Išlaikomų vaikų skaičius namų ūkyje' struktūra

Kategorija	Kiekis	Procentai
Namų ūkiai be išlaikomų vaikų	3505	80.9
Namų ūkiai su 1 išlaikomu vaiku	446	10.3
Namų ūkiai su 2 išlaikomais vaikais	285	6.6
Namų ūkiai su 3 ir daugiau išlaikomų vaikų	98	2.3



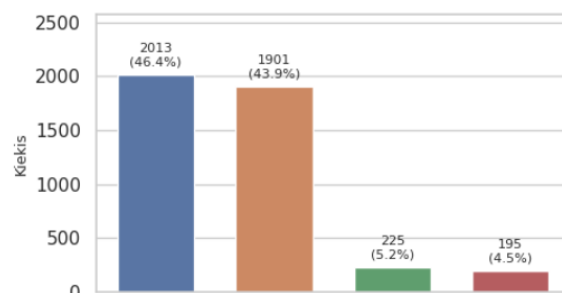
Kintamojo 'Socialinė-ekonominė grupė' struktūra

Kategorija	Kiekis	Procentai
Samdomieji darbuotojai	1939	44.7
Senatvės pensininkai	1767	40.8
Kiti	352	8.1
Savarankiškai dirbantys asmenys	276	6.4



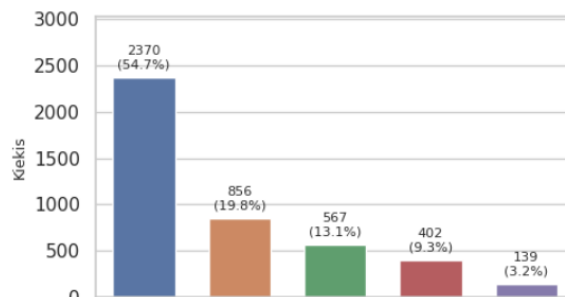
Kintamojo 'Namų ūkio pagrindinis pajamų šaltinis' struktūra

Kategorija	Kiekis	Procentai
Darbo užmokestis	2013	46.4
Pensija	1901	43.9
Pašalpa, stipendija ir kitas pragyvenimo šaltinis	225	5.2
Pajamos iš savo ar šeimos verslo	195	4.5



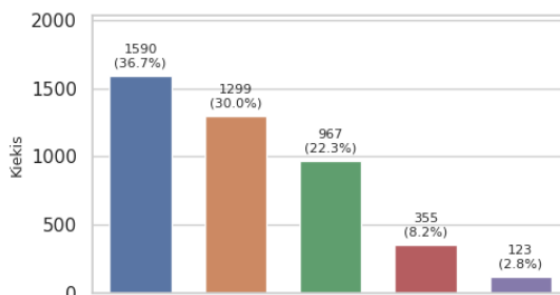
Kintamojo 'Namų ūkio galvos amžius' struktūra

Kategorija	Kiekis	Procentai
60 metų ir vyresni	2370	54.7
50-59 metų	856	19.8
40-49 metų	567	13.1
30-39 metų	402	9.3
iki 30 metų	139	3.2



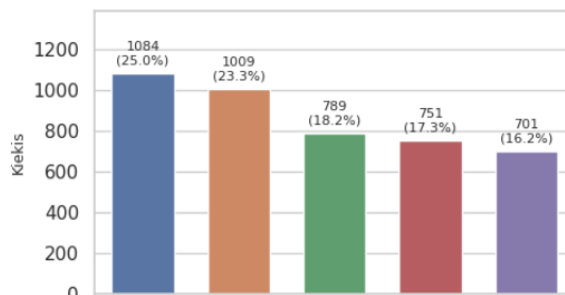
Kintamojo 'Namų ūkio galvos išsilavinimas (grupuotas)' struktūra

Kategorija	Kiekis	Procentai
Aukštesnysis ir aukštasis	1590	36.7
Aukštesniojo ir aukštojo išsilavinimo nesuteikiantis povidurinis	1299	30.0
Vidurinis	967	22.3
Pagrindinis	355	8.2
Pradinis, nebaigė pradinės mokyklos	123	2.8



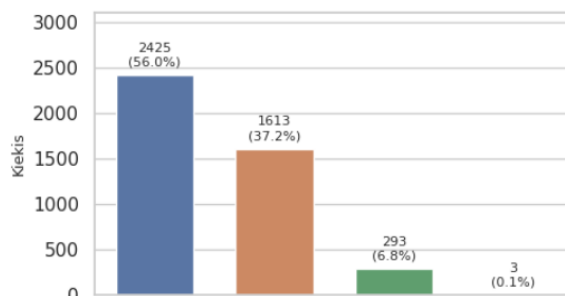
Kintamojo 'Pajamų kvintilinės grupės' struktūra

Kategorija	Kiekis	Procentai
2-a pajamų kvintilinė grupė	1084	25.0
3-a pajamų kvintilinė grupė	1009	23.3
5-a pajamų kvintilinė grupė	789	18.2
1-a pajamų kvintilinė grupė	751	17.3
4-a pajamų kvintilinė grupė	701	16.2



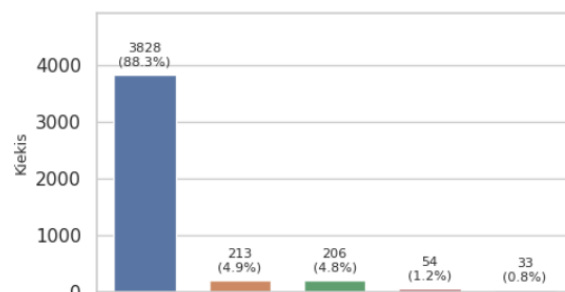
Kintamojo 'Būsto tipas' struktūra

Kategorija	Kiekis	Procentai
Butas daugiabučiame name	2425	56.0
Vieno buto namas	1613	37.2
Butas dviejų butų ar sublokuotame name	293	6.8
Kita gyvenamoji patalpa	3	0.1



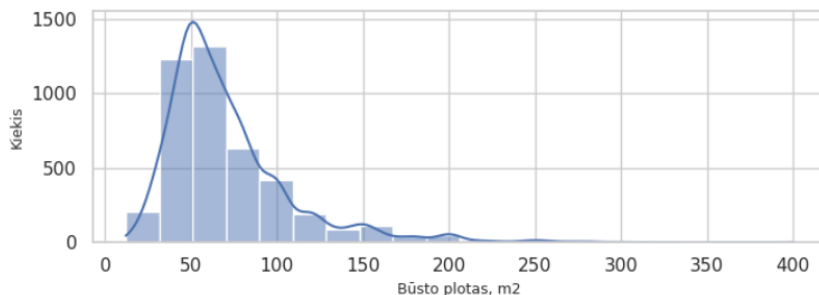
Kintamojo 'Būsto valdos statusas' struktūra

Kategorija	Kiekis	Procentai
Būsto savininkas, neturintis išmokėti paskolos šiam būstui įsigyti	3828	88.3
Gyvena nemokėdamas nuomos	213	4.9
Būsto savininkas, turintis neišmokėtą paskolą šiam būstui įsigyti	206	4.8
Nuomojasi šį būstą mažesne nei rinkos kaina	54	1.2
Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina	33	0.8



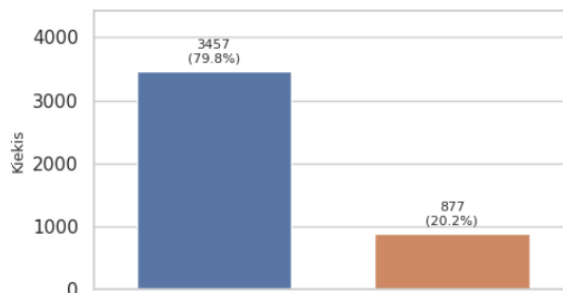
Statistika: Namų ūkio užimamas naudingasis plotas

Statistika	Reikšmė
Kiekis	4334.0
Vid.	71.41
Std	37.77
Min	12.0
25%	49.0
Mediana	62.0
75%	81.0
Max	400.0



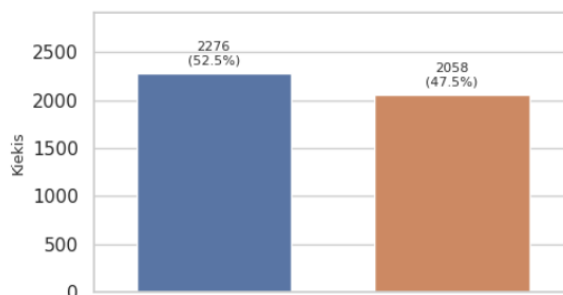
Kintamojo 'Centralizuotas šalto vandens tiekimas' struktūra

Kategorija	Kiekis	Procentai
Taip	3457	79.8
Ne	877	20.2



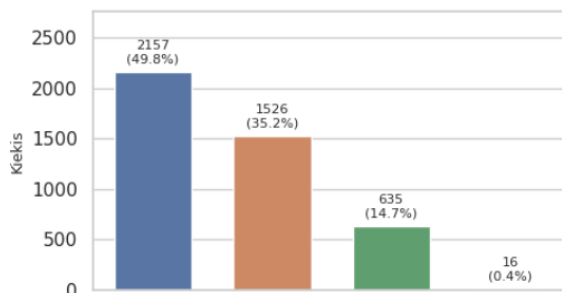
Kintamojo 'Centralizuotas karšto vandens tiekimas' struktūra

Kategorija	Kiekis	Procentai
Ne	2276	52.5
Taip	2058	47.5



Kintamojo 'Pagrindinis būsto šildymo būdas' struktūra

Kategorija	Kiekis	Procentai
Centrinis šildymas iš šilumos tinklu	2157	49.8
Centrinis šildymas iš vietinio šilumos šaltinio	1526	35.2
Krosnis, židinys	635	14.7
Nėra stacionaraus šilumos šaltinio	16	0.4



5 priedas. Individualių namų ūkių narių rodiklių aprašomoji analizė

Duomenų rinkinį sudaro 8 496 namų ūkių narių stebiniai ir 16 stulpelių, iš kurių pirmasis naudojamas kaip identifikatorius:

1. Namų ūkio identifikatorius
2. Gimimo šalis
3. Pagrindinės pilietybės šalis
4. Tėvo gimimo šalis
5. Motinos gimimo šalis
6. Lytis
7. Amžius suėjusių metų skaičiumi
8. Teisinė santuokinė padėtis
9. Gyvena su partneriu
10. Namų ūkio nario išsilavinimas
11. Šiuo metu mokosi
12. Dabartinis formaliojo švietimo ar mokymo veiklos lygis
13. Ekonominio aktyvumo (veiklos) pobūdis
14. Pagrindinis darbas: profesija
15. Užimtumo statusas pagrindinėje darbovietėje
16. Darbo sutartis pagrindiniame darbe

Amžiaus struktūros analizė rodo, kad tiriamųjų vidutinis amžius siekia 50,54 metų ($\sigma = 23,08$), o medianinė reikšmė yra 55 metai. Kadangi medianinis amžius viršija vidurkį, galima daryti prielaidą apie šiek tiek jaunesnių amžiaus grupių įtaką bendram pasiskirstymui, tačiau bendra struktūra išlieka orientuota į vyresnio amžiaus gyventojų segmentą. Ketvirtadalis tiriamųjų yra jaunesni nei 35 metų, o 75 % asmenų amžius neviršija 69 metų. Amžiaus intervalas nuo 0 iki 98 metų rodo, kad imtyje apimamas visas namų ūkių gyvenimo ciklas – nuo vaikų iki vyresnio amžiaus asmenų.

Išsilavinimo struktūra rodo santykinai aukštą žmogiškojo kapitalo lygį. Didžiausia tiriamųjų dalis turi vidutinį išsilavinimą (ISCED 3–4) – 45,3 %, o beveik trečdalis (30,4 %) yra įgiję aukštąjį išsilavinimą (ISCED 5–8). Žemesnio išsilavinimo grupė sudaro tik 13,0 % visų stebinių. Tai leidžia teigti, kad analizuojama imtis pasižymi santykinai aukšta išsilavinimo koncentracija, kuri gali būti siejama su didesnėmis užimtumo galimybėmis bei skirtingais vartojimo prioritetais.

Ekonominio aktyvumo struktūra atskleidžia gana ryškų dirbančiųjų ir pensinio amžiaus gyventojų dominavimą. Dirbantys asmenys sudaro 40,0 % imties, o senatvės arba išankstinės senatvės pensininkai – 32,7 %. Toks pasiskirstymas iš dalies atspindi ir anksčiau nustatą santykinai aukštą vidutinį amžių. Bedarbiai sudaro 5,4 %, studentai ir moksleiviai – 4,8 %, o dėl ilgalaikių sveikatos problemų ekonomiškai neaktyvūs asmenys – 3,6 % visų stebinių. Galima pastebėti, kad ekonomiškai aktyvūs ir ekonomiškai neaktyvūs gyventojai imtyje pasiskirstę netolygiai.

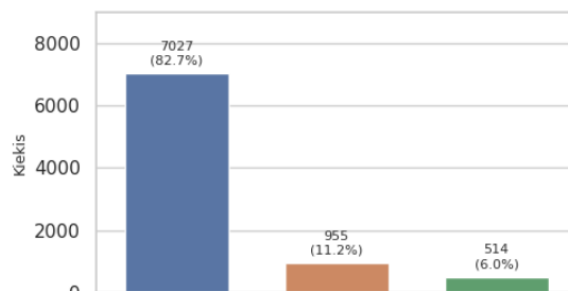
Darbo rinkos charakteristikos rodo gana aukštą užimtumo stabilumą. Tarp dirbančiųjų didžiausią dalį sudaro samdomi darbuotojai (35,1 %), tuo tarpu savarankiškai dirbančių asmenų dalis yra gerokai mažesnė – apie 4,7 %. Papildomai pastebima, kad 34,3 % visų tiriamųjų turi neterminuotą darbo sutartį, o terminuotos darbo sutartys sudaro tik 0,7 %. Tai leidžia daryti prielaidą, kad didelė dalis ekonomiškai aktyvių gyventojų pasižymi santykinai stabilium užimtumu ir nuolatinėmis darbo pajamomis.

Demografinė struktūra rodo nedidelį moterų dominavimą – jos sudaro 56,7 % visų stebinių, tuo tarpu vyrų dalis siekia 43,3 %. Didžiausią dalį sudaro susituokę asmenys (48,2 %), o nevedę arba netekę asmenys sudaro 28,0 %. Taip pat pastebima reikšminga išsituokusių (12,5 %) ir našlių (11,2 %) dalis, kuri gali būti susijusi su santykinai vyresne tiriamųjų amžiaus struktūra.

Migracijos rodikliai rodo gana homogenišką populiacijos sudėtį. Dauguma tiriamųjų yra gimę Lietuvoje (82,7 %), o Lietuvos pilietybę turi net 88,0 % asmenų. Asmenų, gimusių kitose šalyse, dalis išlieka santykinai nedidelė (6,0 %), todėl galima teigti, kad analizuojama imtis pasižymi gana nedidele tarptautinės migracijos įtaka.

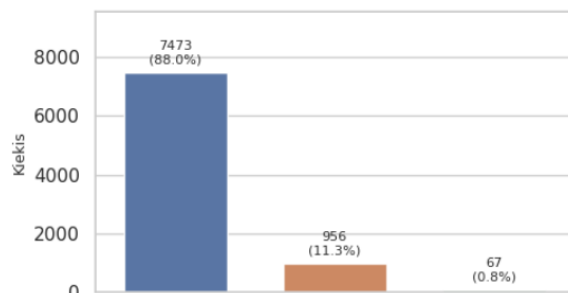
Kintamojo 'Gimimo šalis' struktūra

Kategorija	Kiekis	Procentai
Lietuva	7027	82.7
Netaikoma	955	11.2
Kita šalis	514	6.0



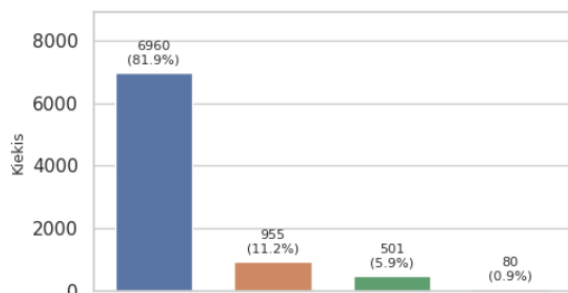
Kintamojo 'Pagrindinės pilietybės šalis' struktūra

Kategorija	Kiekis	Procentai
Lietuvos	7473	88.0
Netaikoma	956	11.3
Kitos šalies	67	0.8



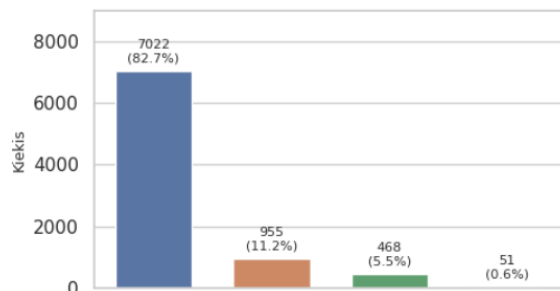
Kintamojo 'Tėvo gimimo šalis' struktūra

Kategorija	Kiekis	Procentai
Lietuva	6960	81.9
Netaikoma	955	11.2
Kita šalis	501	5.9
Nenurodė	80	0.9



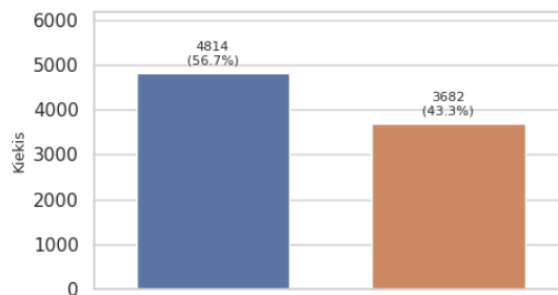
Kintamojo 'Motinos gimimo šalis' struktūra

Kategorija	Kiekis	Procentai
Lietuva	7022	82.7
Netaikoma	955	11.2
Kita šalis	468	5.5
Nenurodė	51	0.6



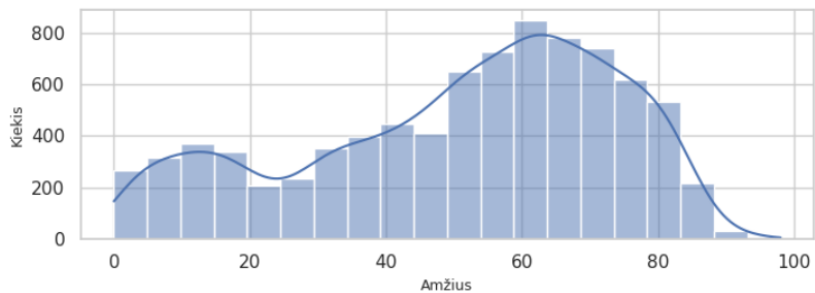
Kintamojo 'Lytis' struktūra

Kategorija	Kiekis	Procentai
Moteris	4814	56.7
Vyras	3682	43.3



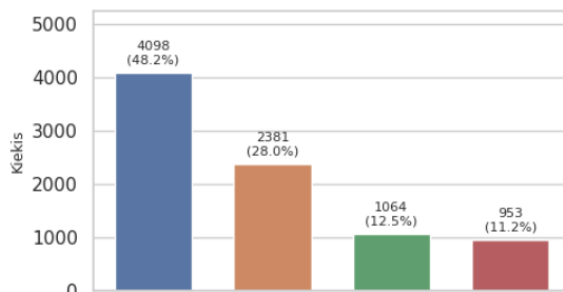
Kintamojo 'Amžius suėjusių metų skaičiumi' statistika

Statistika	Reikšmė
Kiekis	8496.0
Vid.	50.54
Std	23.08
Min	0.0
25%	35.0
Mediana	55.0
75%	69.0
Max	98.0



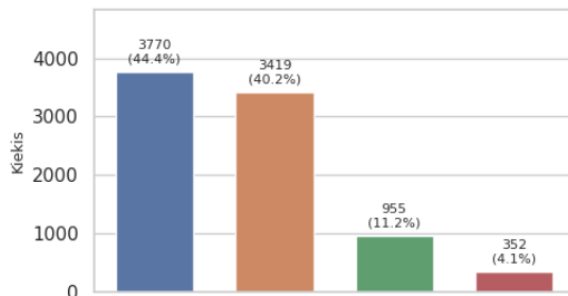
Kintamojo 'Teisinė santuokinė padėtis' struktūra

Kategorija	Kiekis	Procentai
Vedęs (ištekėjusi)	4098	48.2
Nevedęs (netekėjusi)	2381	28.0
Išsituokęs (-usi)	1064	12.5
Našlys (-ė)	953	11.2



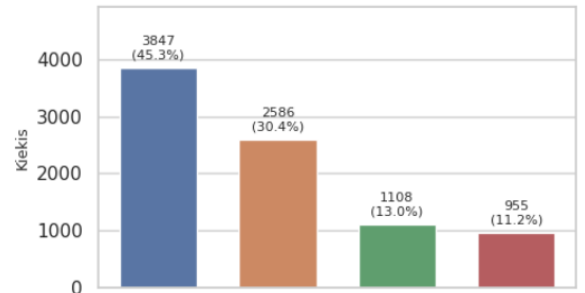
Kintamojo 'Gyvena su partneriu' struktūra

Kategorija	Kiekis	Procentai
Taip, teisiniu pagrindu	3770	44.4
Ne	3419	40.2
Netaikoma	955	11.2
Taip, be teisinio pagrindo	352	4.1



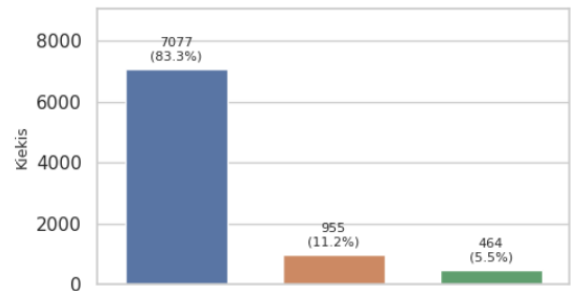
Kintamojo 'Namų ūkio nario išsilavinimas' struktūra

Kategorija	Kiekis	Procentai
Vidutinis (ISCED 3-4)	3847	45.3
Aukščiausias (ISCED 5-8)	2586	30.4
Žemesnis (ISCED 0-2)	1108	13.0
Netaikoma	955	11.2



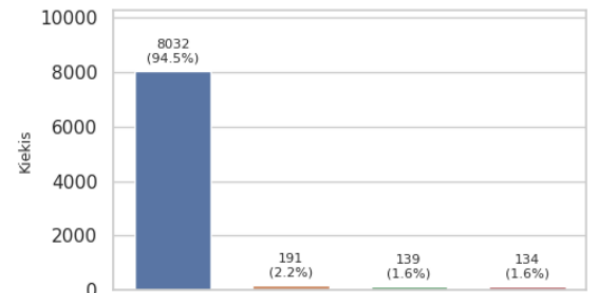
Kintamojo 'Šiuo metu mokosi' struktūra

Kategorija	Kiekis	Procentai
Ne	7077	83.3
Netaikoma	955	11.2
Taip	464	5.5



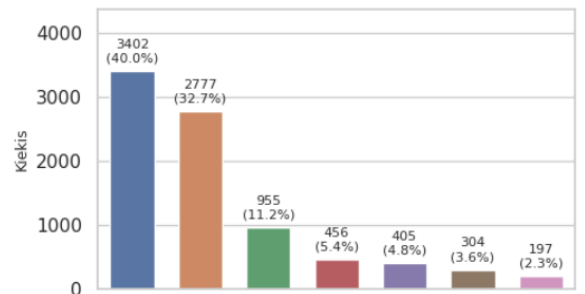
Kintamojo 'Dabartinis formaliojo švietimo ar mokymo veiklos lygis' struktūra

Kategorija	Kiekis	Procentai
Netaikoma	8032	94.5
Vidutinis (ISCED 3-4)	191	2.2
Žemesnis (ISCED 0-2)	139	1.6
Aukščiausias (ISCED 5-8)	134	1.6



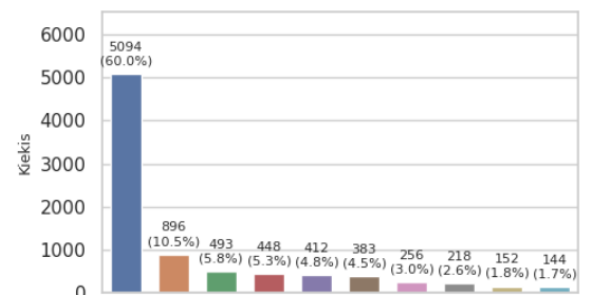
Kintamojo 'Ekonominio aktyvumo (veiklos) pobūdis' struktūra

Kategorija	Kiekis	Procentai
Dirbantis asmuo	3402	40.0
Senatvės ar išankstinės senatvės pensininkas	2777	32.7
Netaikoma	955	11.2
Bedarbis	456	5.4
Mokinys (-ė), studentas (-ė)	405	4.8
Dėl ilgalaikio sveikatos sutrikimo nedirbantis asmuo	304	3.6
Namų šeimininkas (-ė), kitas ekonomiškai neaktyvus asmuo	197	2.3



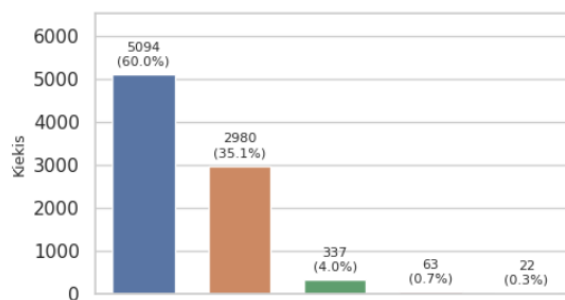
Kintamojo 'Pagrindinis darbas: profesija' struktūra

Kategorija	Kiekis	Procentai
Netaikoma	5094	60.0
Specialistai	896	10.5
Kvalifikuoti darbininkai ir amatininkai	493	5.8
Nenurodė	448	5.3
Paslaugų sektoriaus darbuotojai ir pardavėjai	412	4.8
Irenginių ir mašinų operatoriai ir surinkėjai	383	4.5
Vadovai	256	3.0
Technikai ir jaunesnieji specialistai	218	2.6
Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai	152	1.8
Istaigų tarnautojai	144	1.7



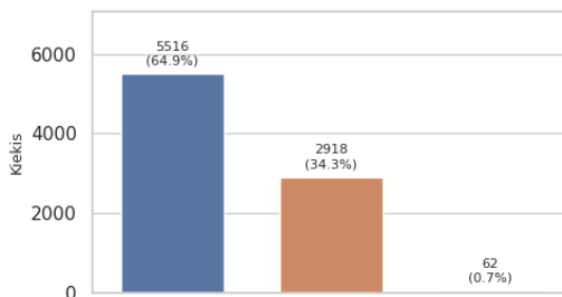
Kintamojo 'Užimtumo statusas pagrindinėje darbovietėje' struktūra

Kategorija	Kiekis	Procentai
Netaikoma	5094	60.0
Samdomasis darbuotojas	2980	35.1
Savarankiškai be samdomųjų darbuotojų dirbantis asmuo	337	4.0
Savarankiškai su samdomaisiais darbuotojais dirbantis asmuo	63	0.7
Dirbantis be atlyginimo šeimos versle	22	0.3



Kintamojo 'Darbo sutartis pagrindiniame darbe' struktūra

Kategorija	Kiekis	Procentai
Netaikoma	5516	64.9
Neterminuota sutartis	2918	34.3
Terminuota sutartis	62	0.7



6 priedas. Individualių duomenų agregavimas ir aprašomoji analizė

Tegul h – namų ūkis; i – asmuo, priklausantis namų ūkiui h ; n_h – namų ūkio h narių skaičius; $I(\cdot)$ – funkcija, kurios reikšmė lygi 1, jei sąlyga tenkinama, ir 0 – priešingu atveju. Tada:

Vyrų dalis namų ūkyje

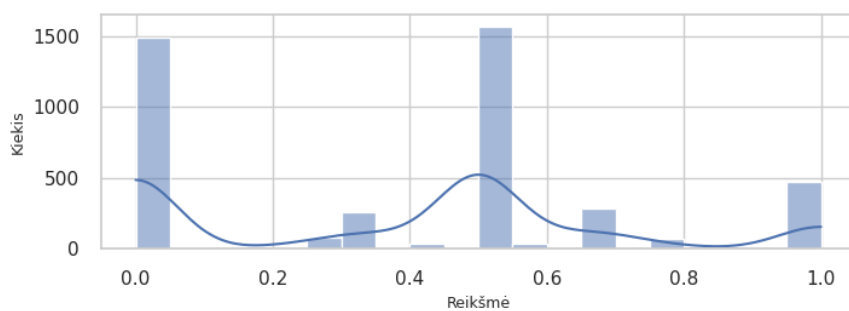
$$MR_h = \frac{\sum_{i \in h} I(M_i = 1)}{n_h};$$

kur M_i – asmens lyties požymis ($M_i = 1$, jei asmuo yra vyras).

Vidutinė vyrų dalis namų ūkyje siekia 0,38 ($\sigma = 0,33$), o medianinė reikšmė lygi 0,50. Kadangi tiek pirmasis, tiek trečiasis kvartiliai atitinkamai siekia 0,00 ir 0,50, galima pastebėti, kad reikšminga dalis namų ūkių pasižymi gana nevienoda lyčių sudėtimi – nuo namų ūkių, kuriuose vyrų nėra, iki tokių, kuriuose visi nariai yra vien tikvyrai. Tai rodo gana nemažą namų ūkių sudėties įvairovę.

Kintamojo 'Vyrų dalis namų ūkyje' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	0.38
Std	0.33
Min	0.0
25%	0.0
Mediana	0.5
75%	0.5
Max	1.0



Vidutinis namų ūkio narių amžius

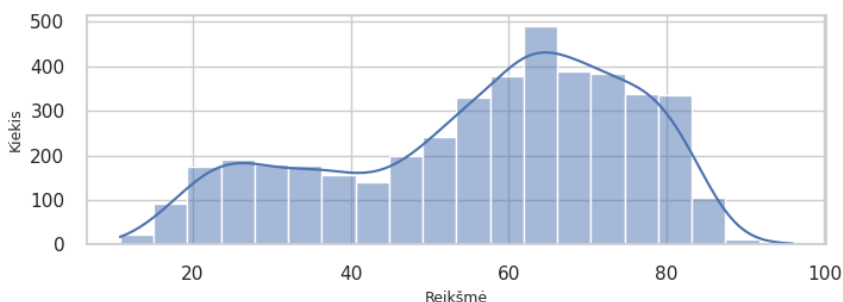
$$A_h = \frac{\sum_{i \in h} age_i}{n_h};$$

kur age_i – i -ojo asmens amžius metais.

Vidutinis namų ūkio narių amžius siekia 56,68 metų ($\sigma = 18,65$), o medianinė reikšmė sudaro 60,58 metų. Pastebima, kad daugiau nei pusėje namų ūkių vidutinis narių amžius viršija 60 metų, o trečiasis kvartilis siekia net 71 metus. Tokie rezultatai leidžia daryti prielaidą, kad imtyje vyrauja santykinai vyresnio amžiaus namų ūkiai, kurių vartojimo struktūra bei socialiniai poreikiai gali skirtis nuo jaunesnių namų ūkių.

Kintamojo 'Vid. namų ūkio narių amžius' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	56.68
Std	18.65
Min	10.8
25%	43.38
Mediana	60.58
75%	71.0
Max	96.0



Dirbančių asmenų dalis namų ūkyje

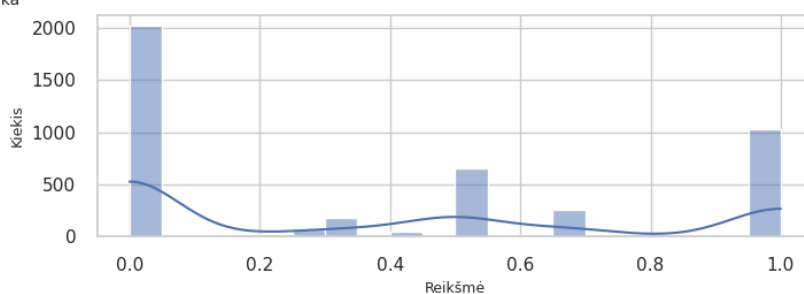
$$ER_h = \frac{\sum_{i \in h} I(E_i = 1)}{n_h};$$

kur E_i – užimtumo požymis ($E_i = 1$, jei asmuo dirba).

Ekonominio aktyvumo rodiklis rodo gana didelę variaciją tarp namų ūkių. Vidutinė dirbančių asmenų dalis namų ūkyje sudaro 0,38 ($\sigma = 0,41$), o medianinė reikšmė siekia 0,33. Tuo tarpu pirmasis kvartilis yra lygus 0,00, o trečiasis – 0,67, kas rodo, kad reikšminga dalis namų ūkių apskritai neturi dirbančių asmenų, tuo tarpu kituose dirbantieji sudaro didžiąją dalį namų ūkio narių. Tokia variacija gali būti siejama su pensinio amžiaus gyventojų dominavimu dalyje namų ūkių arba skirtinga namų ūkių gyvenimo ciklo stadija.

Kintamojo 'Dirbančių asmenų dalis namų ūkyje' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	0.38
Std	0.41
Min	0.0
25%	0.0
Mediana	0.33
75%	0.67
Max	1.0



Aukštojo išsilavinimo asmenų dalis namų ūkyje

$$HER_h = \frac{\sum_{i \in h} I(H_i = 1)}{n_h};$$

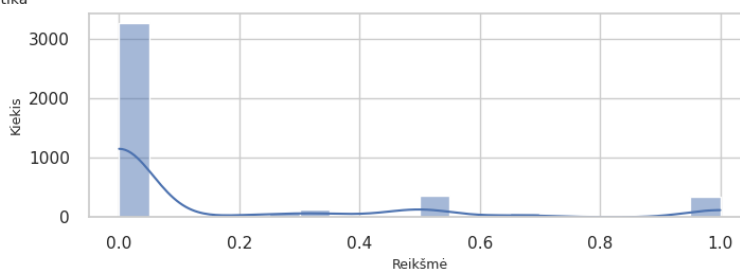
kur H_i – išsilavinimo požymis ($H_i = 1$, jei asmuo turi aukštąjį išsilavinimą).

Aukštasis išsilavinimas atspindi Tarptautinę standartinę švietimo klasifikaciją (toliau – ISCED¹³) ir apima 5–8 ISCED lygius.

Išsilavinimo rodiklis rodo gana ryškią asimetriją. Vidutinė aukštąjį išsilavinimą turinčių asmenų dalis namų ūkyje siekia 0,15 ($\sigma = 0,30$), tačiau tiek medianinė, tiek kvartilinės reikšmės yra lygios 0,00. Tai reiškia, kad daugiau nei pusėje namų ūkių nėra nė vieno aukštąjį išsilavinimą turinčio asmens, o aukštesnė vidutinė reikšmė susidaro dėl santykinai nedidelės dalies namų ūkių, kuriuose aukštąjį išsilavinimą turinčių asmenų koncentracija yra didesnė.

Kintamojo 'Aukštojo išsilavinimo dalis namų ūkyje' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	0.15
Std	0.3
Min	0.0
25%	0.0
Mediana	0.0
75%	0.0
Max	1.0



¹³ <https://iqa.international/isced-levels/>

Aukštos kvalifikacijos užimtųjų dalis

$$ISR_h = \frac{\sum_{i \in h} I(S_i = 1)}{\sum_{i \in h} I(E_i = 1)};$$

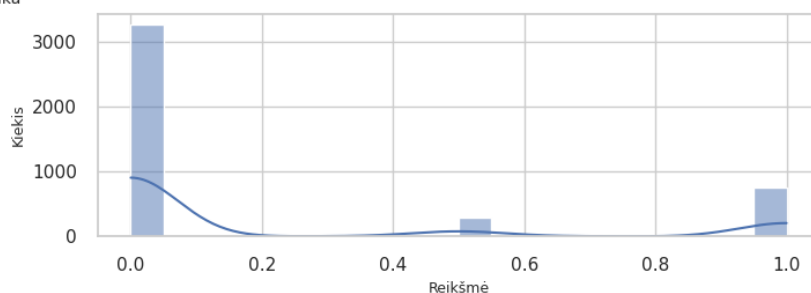
kur S_i – profesinės kvalifikacijos požymis ($S_i = 1$, jei asmuo priklauso ISCO 1-3 profesijų grupėms); E_i – užimtumo požymis. Vardiklyje skaičiuojamas bendras dirbančių asmenų skaičius namų ūkyje, jei namų ūkyje nėra dirbančių asmenų, rodiklio reikšmė prilyginama 0.

Aukštos kvalifikacijos užimtųjų dalies rodiklis sudaromas vadovaujantis ISCO klasifikatoriumi¹⁴ bei Tarptautinė darbo tarnybos (ILO) apibrėžtais aukštesniais kvalifikacijos lygiais¹⁵.

Panaši struktūra, kaip aukštojo išsilavinimo rodiklyje, stebima ir aukštos kvalifikacijos užimtųjų dalies rodiklyje. Vidutinė reikšmė siekia 0,21 ($\sigma = 0,38$), tačiau tiek medianinė, tiek kvartilinės reikšmės taip pat yra lygios 0,00. Tai rodo, kad didelėje dalyje namų ūkių aukštos kvalifikacijos darbuotojų nėra, tačiau egzistuoja mažesnė grupė namų ūkių, kuriuose šių asmenų dalis yra reikšminga.

Kintamojo 'Aukštos kvalifikacijos užimtųjų dalis' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	0.21
Std	0.38
Min	0.0
25%	0.0
Mediana	0.0
75%	0.0
Max	1.0



Bendra amžiaus priklausomybė

$$DR_h = \frac{\sum_{i \in h} I(\text{age } e_i < 20) + \sum_{i \in h} I(\text{age } e_i > 64)}{\sum_{i \in h} I(20 \leq \text{age } e_i \leq 64)};$$

kur $\text{age } e_i$ – i -ojo asmens amžius. Skaitiklyje apibrėžiami ekonomiškai priklausomi asmenys (jaunesni nei 20 metų ir vyresni nei 64 metų), tuo tarpu vardiklyje – darbingo amžiaus asmenys (20-64 metų). Jei namų ūkyje nėra darbingo amžiaus asmenų, rodiklio reikšmė prilyginama 0.

Bendro amžiaus priklausomybės rodiklis sudaromas vadovaujantis Eurostat¹⁶, naudojama 3 tipo metodika, bei Lietuvos pensinio amžiaus rodikliu¹⁷.

Bendro amžiaus priklausomybės rodiklio analizė atskleidžia ypač didelį nevienodumą tarp namų ūkių. Vidutinė rodiklio reikšmė siekia 0,29 ($\sigma = 0,56$), o medianinė reikšmė lygi 0,00. Toks rezultatas rodo, kad daugiau nei pusėje namų ūkių nėra ekonomiškai priklausomų asmenų arba jų dalis yra labai maža. Vis dėlto maksimali reikšmė siekia 6, kas rodo, kad kai kuriuose namų ūkiuose ekonomiškai priklausomų asmenų skaičius ženkliai viršija darbingo amžiaus gyventojų skaičių.

¹⁴ https://ec.europa.eu/eurostat/cache/metadata/en/demo_pop_esms.htm

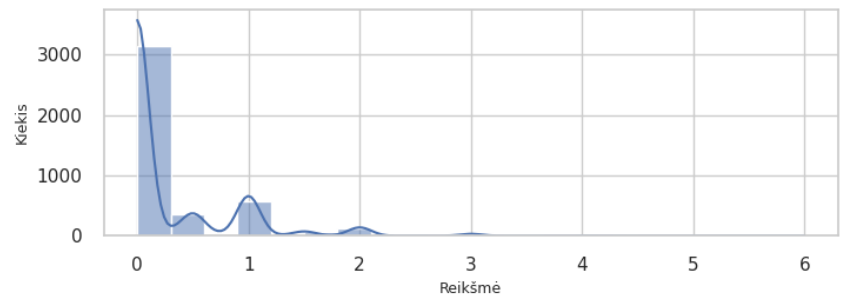
¹⁵ <https://webapps.ilo.org/ilostat-files/ISCO/newdocs-08-2021/ISCO-08/ISCO-08-08%20EN%20Vol%201.pdf>

¹⁶ https://ec.europa.eu/eurostat/cache/metadata/en/demo_pop_esms.htm

¹⁷ <https://sodra.lt/senatves-pensijos-amziaus-lentele>

Kintamojo 'Bendra amžiaus priklausomybė' statistika

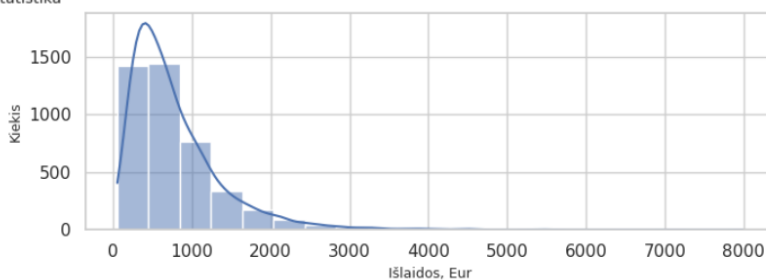
Statistika	Reikšmė
Kiekis	4334.0
Vid.	0.29
Std	0.56
Min	0.0
25%	0.0
Mediana	0.0
75%	0.5
Max	6.0



7 priedas. Namų ūkių vartojimo išlaidų rodiklių aprašomoji statistika

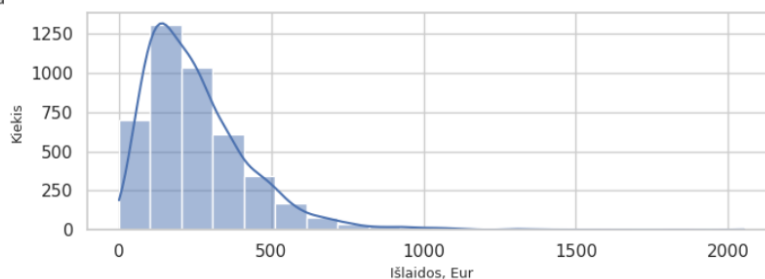
Kintamojo 'Visos namų ūkio vartojimo išlaidos (mėnesinės)' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	792.74
Std	615.55
Min	56.23
25%	382.34
Mediana	630.5
75%	1016.54
Max	7984.83



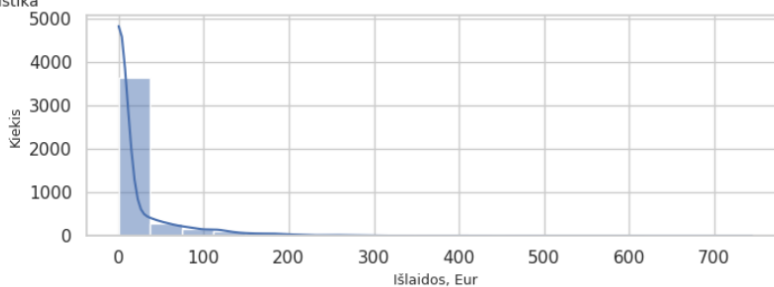
Kintamojo 'Maistas ir nealkoholiniai gėrimai' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	257.53
Std	180.29
Min	0.0
25%	131.14
Mediana	218.4
75%	339.4
Max	2053.07



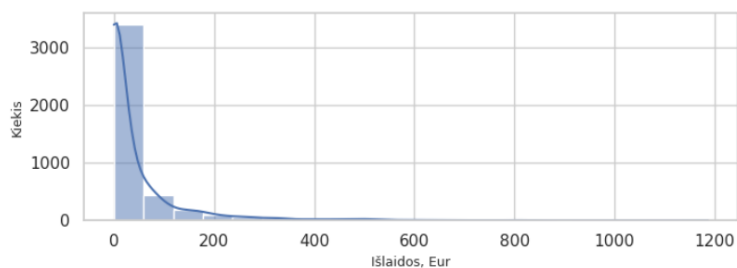
Kintamojo 'Alkoholiniai gėrimai, tabakas ir narkotikai' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	20.19
Std	54.96
Min	0.0
25%	0.0
Mediana	0.0
75%	1.94
Max	744.46



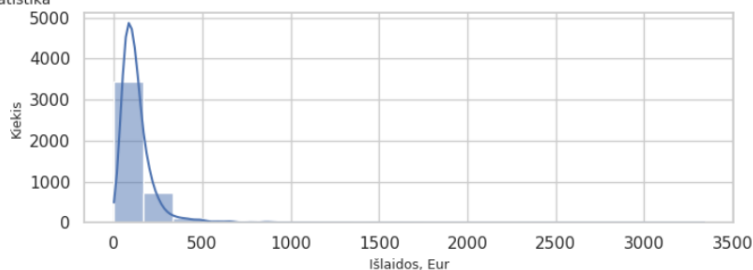
Kintamojo 'Apranga ir avalynė' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	45.83
Std	104.97
Min	0.0
25%	0.0
Mediana	0.0
75%	45.0
Max	1186.25



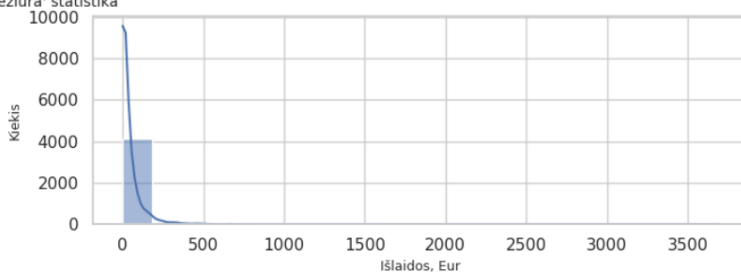
Kintamojo 'Būstas, vanduo, elektra, dujos ir kitas kuras' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	130.52
Std	124.95
Min	0.0
25%	71.22
Mediana	106.33
75%	152.38
Max	3339.3



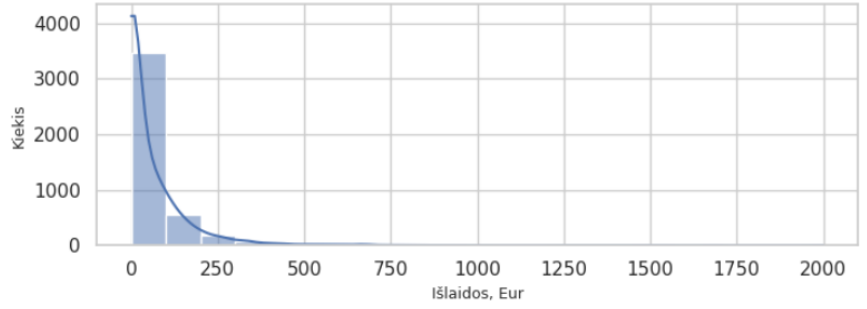
Kintamojo 'Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	41.91
Std	103.08
Min	0.0
25%	0.0
Mediana	13.98
75%	48.47
Max	3691.87



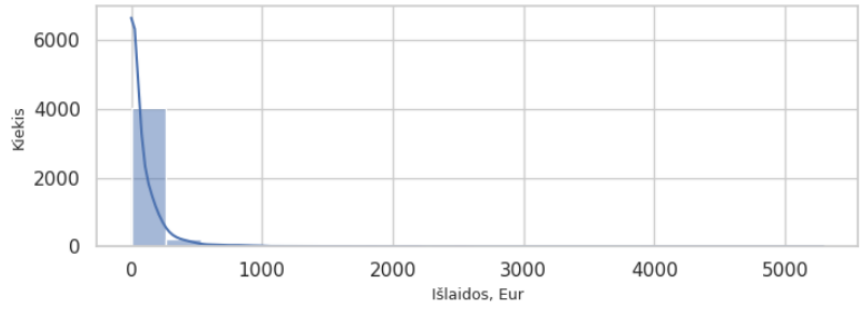
Kintamojo 'Sveikata' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	61.57
Std	127.86
Min	0.0
25%	0.0
Mediana	15.0
75%	79.57
Max	1999.98



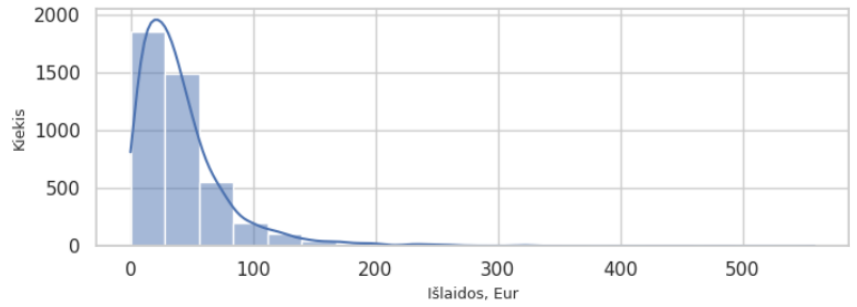
Kintamojo 'Transportas' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	87.64
Std	242.03
Min	0.0
25%	0.0
Mediana	0.0
75%	101.76
Max	5291.96



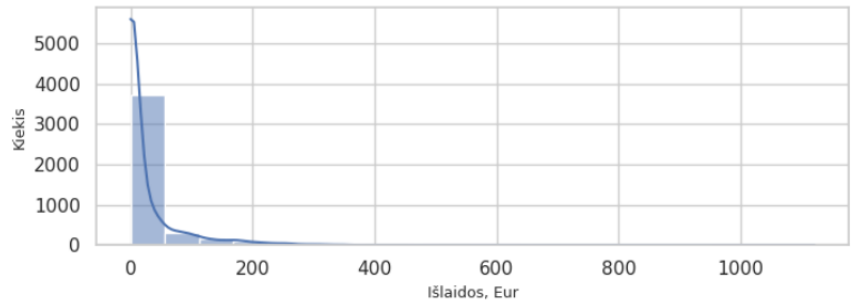
Kintamojo 'Informacija ir ryšiai' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	42.29
Std	41.64
Min	0.0
25%	17.0
Mediana	32.0
75%	53.0
Max	558.32



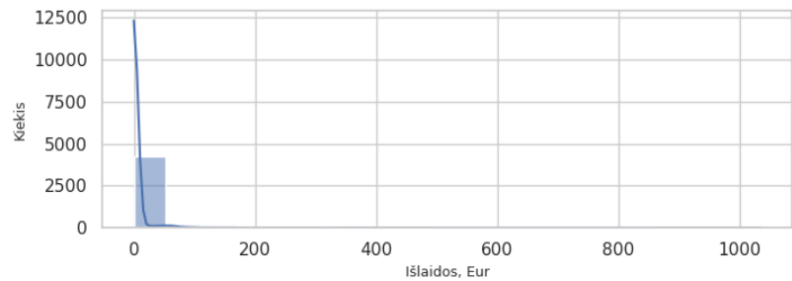
Kintamojo 'Poilsis, sportas ir kultūra' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	27.21
Std	62.19
Min	0.0
25%	0.0
Mediana	2.17
75%	25.0
Max	1121.28



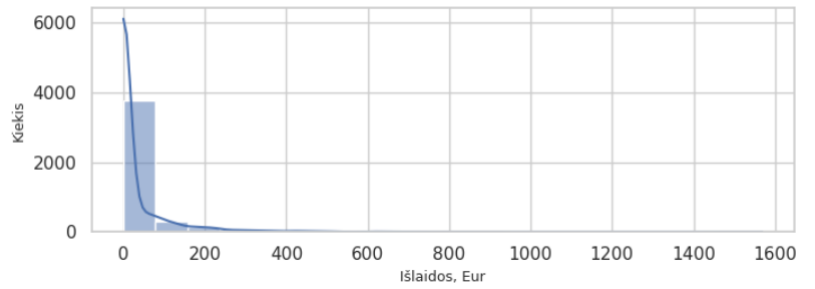
Kintamojo 'Švietimo paslaugos' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	4.73
Std	36.99
Min	0.0
25%	0.0
Mediana	0.0
75%	0.0
Max	1035.0



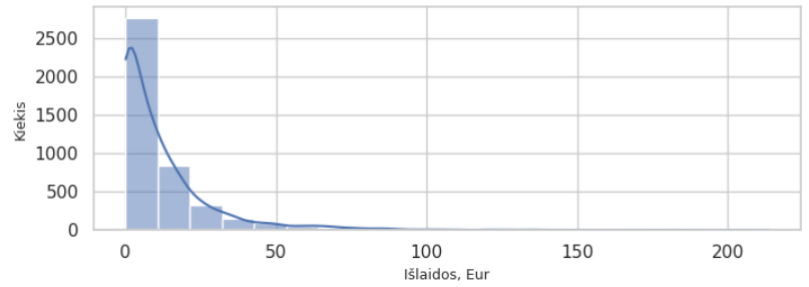
Kintamojo 'Restoranai ir apgyvendinimo paslaugos' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	31.74
Std	91.61
Min	0.0
25%	0.0
Mediana	0.0
75%	10.64
Max	1568.69



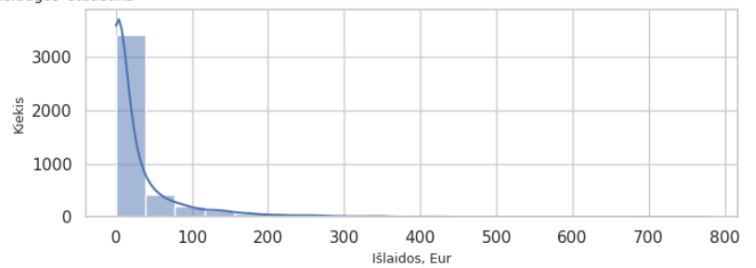
Kintamojo 'Draudimas ir finansinės paslaugos' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	12.17
Std	18.64
Min	0.0
25%	0.0
Mediana	6.4
75%	15.83
Max	213.45



Kintamojo 'Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos' statistika

Statistika	Reikšmė
Kiekis	4334.0
Vid.	29.42
Std	59.39
Min	0.0
25%	0.0
Mediana	6.67
75%	30.85
Max	778.57

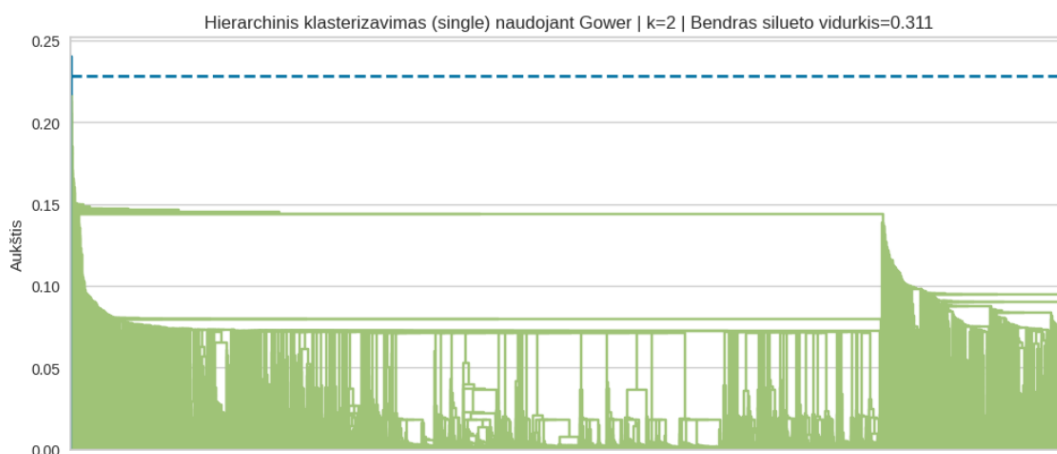


8 priedas. Požymių atranka - kintamųjų informatyvumo tikrinimas

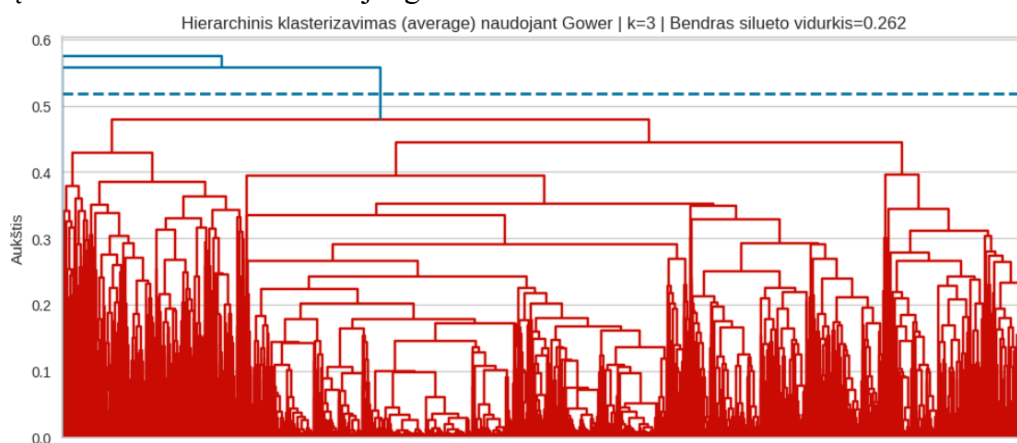
	pavadinimas	tipas	unikalios reikšmės	dispersija	dažniausiai pasikartojanti reikšmė	procentas	šalinti
0	Namų ūkio pagrindinis pajamų šaltinis	kategorinis	4		1	46.45	---
1	Regionas	kategorinis	2		2	76.86	---
2	Gyvenamoji vieta	kategorinis	3		3	36.78	---
3	Išlaikomų vaikų skaičius namų ūkyje	kategorinis	4		4	80.87	---
4	Namų ūkio galvos amžius	kategorinis	5		5	54.68	---
5	Būsto tipas	kategorinis	4		3	55.95	---
6	Būsto valdos statusas	kategorinis	5		1	88.32	---
7	Namų ūkio užimamas naudingasis plotas	skaitinis	159	1426.545			---
8	Asmenų skaičius namų ūkyje	kategorinis	9		1	40.68	---
9	Centralizuotas šalto vandens tiekimas	kategorinis	2		1	79.76	---
10	Centralizuotas karšto vandens tiekimas	kategorinis	2		2	52.51	---
11	Pagrindinis būsto šildymo būdas	kategorinis	4		1	49.77	---
12	Socialinė-ekonominė grupė	kategorinis	4		1	44.74	---
13	Namų ūkio tipas	kategorinis	6		1	40.68	---
14	Namų ūkio galvos išsilavinimas (grupuotas)	kategorinis	5		5	36.69	---
15	Pajamų kvintilinės grupės	kategorinis	5		2	25.01	---
16	Vyrų dalis namų ūkyje	skaitinis	17	0.1065			---
17	Vid. namų ūkio narių amžius	skaitinis	406	347.7691			---
18	Dirbančių asmenų dalis namų ūkyje	skaitinis	20	0.1672			---
19	Aukštojo išsilavinimo dalis namų ūkyje	skaitinis	14	0.0913			---
20	Aukštos kvalifikacijos užimtųjų dalis	skaitinis	8	0.1472			---
21	Bendra amžiaus priklausomybė	skaitinis	18	0.3162			---

9 priedas. Namų ūkių hierarchinio kasterizavimo dendogramos

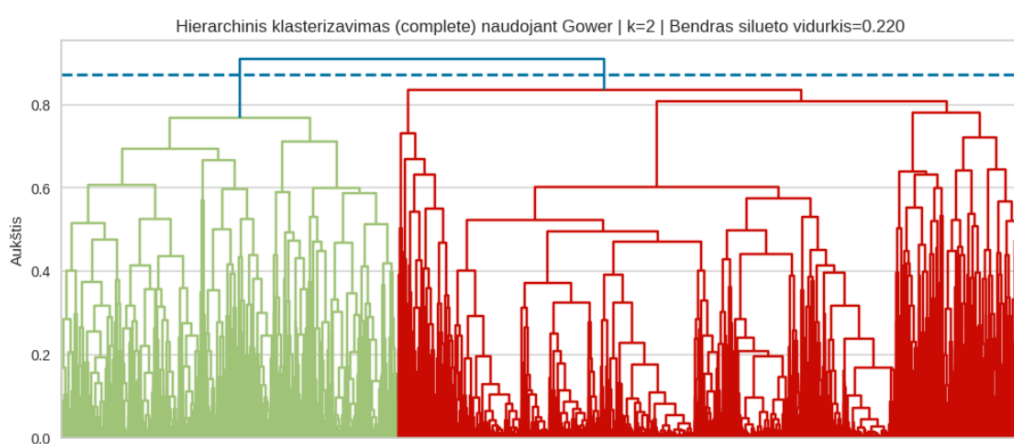
Namų ūkių klasterizavimas vienetinės jungties metodu



Namų ūkių klasterizavimas vidutinės jungties metodu



Namų ūkių klasterizavimas pilnosios jungties metodu



10 priedas. Optimizuoti Gower matricos požymių svoriai

Požymis	Pradinis svoris	Optimizuotas svoris ($k = 4$)	Optimizuotas svoris ($k = 5$)	Optimizuotas svoris ($k = 6$)
Gyvenamoji vieta	1,00	1,00	1,00	1,00
Regionas	1,00	1,00	1,00	1,00
Namų ūkio tipas	1,00	1,00	1,00	1,00
Išlaikomų vaikų skaičius namų ūkyje	1,00	1,00	2,37	1,00
Dirbančių asmenų dalis namų ūkyje	1,00	2,15	3,00	3,00
Namų ūkio pagrindinis pajamų šaltinis	1,00	3,00	2,96	2,99
Vid. namų ūkio narių amžius	1,00	1,00	1,00	1,00
Bendra amžiaus priklausomybė	1,00	1,00	1,00	1,00
Aukštojo išsilavinimo dalis namų ūkyje	1,00	1,00	1,00	3,00
Aukštos kvalifikacijos užimtųjų dalis	1,00	1,00	1,00	2,54
Pajamų kvintilinės grupės	1,00	1,00	1,00	1,00
Būsto valdos statusas	1,00	1,00	1,00	1,00
Namų ūkio užimamas naudingasis plotas	1,00	1,00	1,00	1,00
Būsto tipas	1,00	3,00	3,00	3,00

11 priedas. Požymių šalinimo rezultatai pagal klasterių skaičių (k)

k	Pašalintas požymis	Likusių požymių sk.	Silueto koeficientas
3	Gyvenamoji vieta	13	0,342
3	Būsto tipas	12	0,416
3	Namų ūkio tipas	11	0,459
3	Aukštos kvalifikacijos užimtųjų dalis	10	0,482
3	Regionas	9	0,535
3	Pajamų kvintilinė grupė	8	0,584
3	Būsto valdos statusas	7	0,624
4	Gyvenamoji vieta	13	0,315
4	Būsto tipas	12	0,385
4	Namų ūkio tipas	11	0,451
4	Regionas	10	0,488
4	Pajamų kvintilinė grupė	9	0,536
4	Būsto valdos statusas	8	0,575
4	Vid. namų ūkio narių amžius	7	0,600
5	Gyvenamoji vieta	13	0,314
5	Būsto tipas	12	0,379
5	Namų ūkio tipas	11	0,452
5	Regionas	10	0,496
5	Pajamų kvintilinė grupė	9	0,547
5	Būsto valdos statusas	8	0,585
5	Namų ūkio užimamas naudingasis plotas	7	0,627
6	Gyvenamoji vieta	13	0,302
6	Būsto tipas	12	0,376
6	Namų ūkio tipas	11	0,454
6	Regionas	10	0,507
6	Pajamų kvintilinė grupė	9	0,568
6	Būsto valdos statusas	8	0,612
6	Namų ūkio užimamas naudingasis plotas	7	0,640
7	Gyvenamoji vieta	13	0.296
7	Būsto tipas	12	0.375
7	Namų ūkio tipas	11	0.451
7	Pajamų kvintilinė grupė	10	0.487
7	Regionas	9	0.532
7	Būsto valdos statusas	8	0.617
7	Namų ūkio užimamas naudingasis plotas	7	0.644

12 priedas. Galutinės optimizuotų požymių svorių reikšmės

Kintamasis	Pradinis svoris	Optimizuotas svoris
Dirbančių asmenų dalis namų ūkyje	1,000	1,235
Aukštojo išsilavinimo dalis namų ūkyje	1,000	2,008
Aukštos kvalifikacijos užimtųjų dalis	1,000	3,000
Namų ūkio užimamas naudingasis plotas	1,000	1,000
Vid. namų ūkio narių amžius	1,000	1,000
Bendra amžiaus priklausomybė	1,000	1,000
Pajamų kvintilinė grupė	1,000	1,000
Namų ūkio pagrindinis pajamų šaltinis	1,000	3,000
Išlaikomų vaikų skaičius namų ūkyje	1,000	3,000
Būsto valdos statusas	1,000	1,000

13 priedas. Klasterių stabilumo ir atskiriamumo vertinimo rezultatai

Stabilumas – klasterių stabilumo vertinimo parametrų suvestinė

Parametras	Reikšmė
Stabilumo vertinimo metodas	Pakoreguotas Rando indeksas (ARI)
Iteracijų skaičius	30
Imties dydis iteracijoje	80 %
Imties atrinkimo būdas	Stratifikuotas
Klasterizavimo algoritmas	k-medoidai
Klasterių skaičius	6
Atstumo metrika	Svertinis Gower atstumas
Medoidų optimizavimo metodas	PAM
Inicializavimo metodas	k-medoids++

Klasterių stabilumo vertinimo rezultatai

Rodiklis	Reikšmė
Vidutinė ARI reikšmė	0,991
Standartinis nuokrypis	0,011
Minimali reikšmė	0,955
Maksimali reikšmė	1,000

Atskiriamumas – XGBoost modelio parametrų suvestinė, kai mokymo ir testavimo imčių skaidymui panaudota pradinė būsenos reikšmė = 42

Parametras	Reikšmė
Modelio tipas	XGBoost
Medžių skaičius	100
Maksimalus medžio gylis	4
Mokymosi žingsnis	0,05
Tikslinė funkcija	multi:softprob
Vertinimo kriterijus	mlogloss
Kryžminė validacija	3 dalių stratifikuota
Mokymo imtis	80 %
Testavimo imtis	20 %

XGBoost klasterių atskiriamumo vertinimo rezultatai

Rodiklis	Reikšmė
Kryžminės validacijos tikslumas	0,996
Kryžminės validacijos standartinis nuokrypis	0,002
Nepriklausomos testavimo imties tikslumas	0,993

Sumaišymo matrica naudojant visus modelio požymius

		prognuozuotos klasės					
		1	2	3	4	5	6
faktinės klasės	1	83	0	0	1	0	0
	2	0	381	0	0	0	0
	3	0	0	180	0	0	0
	4	0	1	2	41	0	0
	5	1	0	0	0	64	0
	6	0	1	0	0	0	112

```
Kartojimas 2, random_state=123
CV accuracy: [0.99567474 0.99134948 0.995671 ]
CV vidurkis: 0.9942317398414276
CV std: 0.0020380648016482057
Test accuracy: 0.9953863898500577
precision recall f1-score support
0 0.99 1.00 0.99 84
1 1.00 1.00 1.00 381
2 0.99 1.00 0.99 180
3 1.00 0.98 0.99 44
4 1.00 0.98 0.99 65
5 1.00 0.99 1.00 113
accuracy 1.00 1.00 1.00 867
macro avg 1.00 0.99 0.99 867
weighted avg 1.00 1.00 1.00 867
```

```
[[ 84 0 0 0 0 0]
 [ 0 380 1 0 0 0]
 [ 0 0 180 0 0 0]
 [ 0 0 1 43 0 0]
 [ 1 0 0 0 64 0]
 [ 0 1 0 0 0 112]]
```

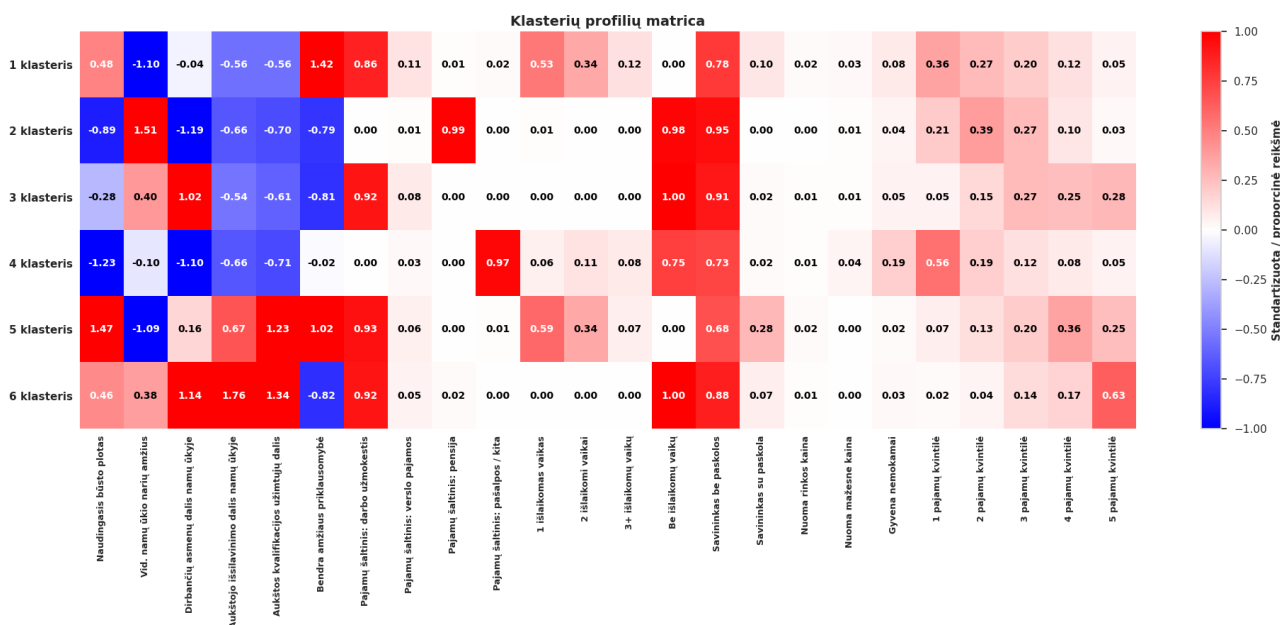
```
Kartojimas 3, random_state=999
CV accuracy: [0.99394464 0.99394464 0.99393939]
CV vidurkis: 0.9939428890985984
CV std: 2.471450774830608e-06
Test accuracy: 0.9965397923875432
precision recall f1-score support
0 0.98 0.99 0.98 84
1 1.00 1.00 1.00 381
2 1.00 1.00 1.00 180
3 1.00 0.98 0.99 44
4 1.00 0.98 0.99 65
5 1.00 1.00 1.00 113
accuracy 1.00 1.00 1.00 867
macro avg 1.00 0.99 0.99 867
weighted avg 1.00 1.00 1.00 867
```

```
[[ 83 1 0 0 0 0]
 [ 0 381 0 0 0 0]
 [ 0 0 180 0 0 0]
 [ 1 0 0 43 0 0]
 [ 1 0 0 0 64 0]
 [ 0 0 0 0 0 113]]
```

```
Bendri rezultatai
CV vidurkiai: [0.995672743250598, 0.9942317398414276, 0.9939428890985984]
CV vidurkių vidurkis: 0.9946157904635413
CV vidurkių std: 0.0007566242871765828
Test accuracy: [0.9930795847750065, 0.9953863898500577, 0.9965397923875432]
Test vidurkis: 0.9950019223375625
Test std: 0.0014385457081022251
```

14 priedas. Skirtingų socialinių grupių profiliai

Klasterių profilių matricoje pateikiamos identifikuotų grupių charakteristikos pagal profiliuojamus požymius. Skaitiniams kintamiesiems naudojamos klasterių vidutinės reikšmės, išreikštos standartizuotomis Z reikšmėmis, o kategoriniams kintamiesiems pateikiamos atitinkamų kategorijų santykinės dalys klasteryje. Spalvų intensyvumas atspindi požymio nuokrypį nuo bendros imties vidurkio arba kategorijos santykinę dalį klasteryje. Teigiamos reikšmės rodo didesnę nei bendroje imtyje požymio reikšmę arba didesnę kategorijos koncentraciją, o neigiamos – mažesnę.



1 klasteris – Jaunos mažesnių pajamų šeimos su vaikais. Šiam klasteriui būdingas santykinai jaunas namų ūkių profilis – vidutinis namų ūkio narių amžius siekia 28,7 metų, o mediana (27 m.) ir kvartilų intervalas nuo 22,3 iki 34,8 metų rodo, kad grupę daugiausia sudaro ankstyvo gyvenimo ciklo namų ūkiai. Vaikų vaidmuo šioje grupėje yra ryškus: daugiau nei pusė namų ūkių (53,3 %) augina vieną vaiką, dar 34,4 % – du vaikus. Nors darbo pajamos sudaro pagrindinį pajamų šaltinį 85,8 % atvejų, užimtumo, išsilavinimo ir kvalifikacijos rodikliai išlieka santykinai žemi – dirbančių asmenų dalis namų ūkyje siekia 44 %, aukštąjį išsilavinimą turinčių asmenų dalis sudaro 4 %, o aukštos kvalifikacijos užimtųjų dalis – 7 %. Ekonominę šios grupės padėtį atspindi ir pajamų struktūra – net 63,0 % namų ūkių priklauso žemiausioms pajamų kvintilėms. Gyvenimo sąlygos taip pat išlieka vidutinės: nors būsto plotas vidutiniškai siekia 78,6 m², medianinė reikšmė yra mažesnė (68 m²), o 50 % stebinių koncentruojasi 50–98 m² intervale.

2 klasteris – Pensinio amžiaus namų ūkiai. Ši grupė aiškiai išsiskiria vyresniu amžiaus profiliu – vidutinis namų ūkio narių amžius siekia 71,6 metų, o pusė stebinių koncentruojasi tarp 67 ir 78 metų. Tai rodo labai homogenišką pensinio amžiaus gyventojų struktūrą. Darbo rinkos ryšys šioje grupėje praktiškai išnykęs – dirbančių asmenų dalis namų ūkyje tesudaro 3 %, o aukštąjį išsilavinimą bei aukštos kvalifikacijos užimtųjų dalis yra beveik nulinė. Pajamų struktūra taip pat vienareikšmė – 99 % namų ūkių pagrindinis pajamų šaltinis yra pensijos. Vaikų šiame klasteryje praktiškai nėra: 98,5 % namų ūkių neturi išlaikomų vaikų. Ekonominiu požiūriu daugiau nei pusė grupės (59,5 %) patenka į žemiausias pajamų kvintiles.

3 klasteris – Vidutinių–aukštesnių pajamų dirbantys namų ūkiai. Šio klasterio profilis atspindi ekonomiškai aktyvius vidutinio ir vyresnio amžiaus namų ūkius. Vidutinis amžius siekia 53,4 metų,

o centriniai 50 % stebinių pasiskirsto 48–61 metų intervale. Grupės išskirtinis bruožas – labai aukštas užimtumo lygis: dirbančių asmenų dalis namų ūkyje siekia 83 %, o darbo užmokestis pagrindiniu pajamų šaltiniu tampa 91,8 % atvejų. Tuo pačiu visi šios grupės namų ūkiai neturi išlaikomų vaikų, todėl disponuojami finansiniai ištekliai mažiau paskirstomi išlaikomų šeimos narių poreikiams. Tai atsispindi ir pajamų struktūroje – daugiau nei pusė (53,8 %) priklauso ketvirtai–penktai pajamų kvintilei.

4 klasteris – Socialiai pažeidžiami namų ūkiai. Ši grupė pasižymi ryškia ekonomine ir socialine atskirtimi. Nors vidutinis amžius siekia 45,1 metų, mediana yra gerokai didesnė (52,5 m.), o platus kvartilų intervalas nuo 29,8 iki 58,8 metų rodo didelį vidinį heterogeniškumą. Ekonominis aktyvumas čia išlieka itin žemas – dirbančių asmenų dalis namų ūkyje siekia 5 %, o aukštojo išsilavinimo ir aukštos kvalifikacijos užimtųjų dalis yra beveik nulinė. Pagrindiniu pajamų šaltiniu 96,8 % atvejų tampa socialinės išmokos ir kitos pajamos, o daugiau nei pusė namų ūkių (56,4 %) priklauso žemiausiai pajamų kvintilei. Papildomą socialinę našta atspindi ir santykinai didelis amžiaus priklausomybės rodiklis (0,44).

5 klasteris – Jaunos ekonomiškai aktyvios šeimos. Šis klasteris, kaip ir pirmasis, pasižymi jaunu amžiaus profiliu – vidutinis amžius siekia 28,9 metų, o pusė stebinių koncentruojasi tarp 22,3 ir 35 metų. Tačiau nuo pirmojo klasterio jis ryškiai skiriasi ekonominiu aktyvumu, išsilavinimu ir kvalifikacija. Aukštojo išsilavinimo asmenų dalis namų ūkyje sudaro 44 % gyventojų, aukštos kvalifikacijos užimtųjų dalis siekia net 88 %, o dirbančių asmenų dalis – 52 %. Vaikų struktūra panaši – 58,5 % augina vieną vaiką, o 34,4 % – du. Geresnę ekonominę padėtį atspindi didesnis būsto plotas – vidutiniškai 87,8 m², o viršutinis kvartilis pasiekia 100 m². Be to, net 28,5 % namų ūkių turi būsto paskolas – didžiausia dalis tarp visų grupių.

6 klasteris – Aukštos kvalifikacijos ir aukštų pajamų namų ūkiai. Šiai grupei būdingas didžiausia aukštojo išsilavinimo ir aukštos kvalifikacijos užimtųjų koncentracija visoje imtyje. Aukštojo išsilavinimo asmenų dalis namų ūkyje sudaro 80 % namų ūkių narių, aukštos kvalifikacijos užimtųjų dalis siekia 93 %, o dirbančių asmenų dalis – 87 %. Vidutinis amžius sudaro 52,9 metų, o 50 % stebinių patenka į 46,4–61,5 metų intervalą. Grupė taip pat pasižymi itin aukšta ekonomine padėtimi – net 63,1 % namų ūkių priklauso aukščiausiai pajamų kvintilei, o darbo užmokestis pagrindinį pajamų šaltinį sudaro 92,4 % atvejų. Šioje grupėje dominuoja namų ūkiai be išlaikomų vaikų, todėl didesni finansiniai ištekliai gali būti nukreipiami vartojimui ir gyvenimo kokybei gerinti.

Tolydžių kintamųjų Q3 (75%) pagal klasterį:

cluster	Namų ūkio užimamas naudingasis plotas	Vid. namų ūkio narių amžius	Dirbančių asmenų dalis namų ūkyje	Aukštojo išsilavinimo dalis namų ūkyje	Aukštos kvalifikacijos užimtųjų dalis	Bendra amžiaus priklausomybė
0	1.0	98.00	34.79	0.50	0.0	1.00
1	2.0	78.75	78.00	0.00	0.0	0.00
2	3.0	85.00	61.00	1.00	0.0	0.00
3	4.0	80.00	58.75	0.00	0.0	0.25
4	5.0	100.00	35.00	0.67	0.5	1.00
5	6.0	95.00	61.50	1.00	1.0	0.00

Tolydžių kintamųjų Q1 (25%) pagal klasterį:

cluster	Namų ūkio užimamas naudingasis plotas	Vid. namų ūkio narių amžius	Dirbančių asmenų dalis namų ūkyje	Aukštojo išsilavinimo dalis namų ūkyje	Aukštos kvalifikacijos užimtųjų dalis	Bendra amžiaus priklausomybė
0	1.0	50.0	22.33	0.33	0.00	0.5
1	2.0	46.0	67.00	0.00	0.00	0.0
2	3.0	49.0	48.00	0.50	0.00	0.0
3	4.0	42.0	29.81	0.00	0.00	0.0
4	5.0	60.0	22.25	0.47	0.33	0.5
5	6.0	50.0	46.42	0.67	0.50	0.0

Tolydžių kintamųjų MIN pagal klasterį:

cluster	Namų ūkio užimamas naudingasis plotas	Vid. namų ūkio narių amžius	Dirbančių asmenų dalis namų ūkyje	Aukštojo išsilavinimo dalis namų ūkyje	Aukštos kvalifikacijos užimtųjų dalis	Bendra amžiaus priklausomybė
0	1.0	24	12.50	0.00	0.00	0.0
1	2.0	12	18.25	0.00	0.00	0.0
2	3.0	16	19.00	0.00	0.00	0.0
3	4.0	13	10.80	0.00	0.00	0.0
4	5.0	29	12.00	0.20	0.20	0.5
5	6.0	25	22.00	0.33	0.33	0.5

Tolydžių kintamųjų MAX pagal klasterį:

cluster	Namų ūkio užimamas naudingasis plotas	Vid. namų ūkio narių amžius	Dirbančių asmenų dalis namų ūkyje	Aukštojo išsilavinimo dalis namų ūkyje	Aukštos kvalifikacijos užimtųjų dalis	Bendra amžiaus priklausomybė
0	1.0	280	56.4	0.80	0.33	0.5
1	2.0	280	96.0	1.00	0.50	1.0
2	3.0	400	79.0	1.00	0.50	0.5
3	4.0	280	82.0	1.00	0.50	0.5
4	5.0	300	52.0	0.75	0.75	1.0
5	6.0	400	78.0	1.00	1.00	1.0

Tolydžių kintamųjų MEAN pagal klasterį:

cluster	Namų ūkio užimamas naudingasis plotas	Vid. namų ūkio narių amžius	Dirbančių asmenų dalis namų ūkyje	Aukštojo išsilavinimo dalis namų ūkyje	Aukštos kvalifikacijos užimtųjų dalis	Bendra amžiaus priklausomybė
0	1.0	78.55	28.71	0.44	0.04	0.07
1	2.0	65.92	71.60	0.03	0.00	0.01
2	3.0	71.53	53.40	0.83	0.04	0.05
3	4.0	62.75	45.06	0.05	0.00	0.00
4	5.0	87.76	28.94	0.52	0.44	0.88
5	6.0	78.36	52.94	0.87	0.80	0.93

Tolydžių kintamųjų MEDIAN pagal klasterį:

cluster	Namų ūkio užimamas naudingasis plotas	Vid. namų ūkio narių amžius	Dirbančių asmenų dalis namų ūkyje	Aukštojo išsilavinimo dalis namų ūkyje	Aukštos kvalifikacijos užimtųjų dalis	Bendra amžiaus priklausomybė
0	1.0	68.0	27.0	0.5	0.0	0.0
1	2.0	57.0	72.5	0.0	0.0	0.0
2	3.0	61.0	55.0	1.0	0.0	0.0
3	4.0	56.0	52.5	0.0	0.0	0.0
4	5.0	76.0	28.0	0.5	0.5	1.0
5	6.0	65.0	56.0	1.0	1.0	1.0

Kategorinis kintamasis: Namų ūkio pagrindinis pajamų šaltinis

cluster	incsour_gr	count	percent
0	1.0 Darbo užmokestis	362	85.78
1	1.0 Pajamos iš savo ar šeimos verslo	45	10.66
2	1.0 Pensija	5	1.18
3	1.0 Pašalpa, stipendija ir kitas pragyvenimo šaltinis	10	2.37
4	2.0 Pajamos iš savo ar šeimos verslo	19	1.00
5	2.0 Pensija	1883	99.00
6	3.0 Darbo užmokestis	828	91.80
7	3.0 Pajamos iš savo ar šeimos verslo	74	8.20
8	4.0 Pajamos iš savo ar šeimos verslo	7	3.21
9	4.0 Pašalpa, stipendija ir kitas pragyvenimo šaltinis	211	96.79
10	5.0 Darbo užmokestis	299	92.57
11	5.0 Pajamos iš savo ar šeimos verslo	20	6.19
12	5.0 Pensija	1	0.31
13	5.0 Pašalpa, stipendija ir kitas pragyvenimo šaltinis	3	0.93
14	6.0 Darbo užmokestis	524	92.42
15	6.0 Pajamos iš savo ar šeimos verslo	30	5.29
16	6.0 Pensija	12	2.12
17	6.0 Pašalpa, stipendija ir kitas pragyvenimo šaltinis	1	0.18

Kategorinis kintamasis: Išlaikomų vaikų skaičius namų ūkyje

cluster	tipas2	count	percent
0	1.0 Namų ūkiai su 1 išlaikomu vaiku	225	53.32
1	1.0 Namų ūkiai su 2 išlaikomais vaikais	145	34.36
2	1.0 Namų ūkiai su 3 ir daugiau išlaikomų vaikų	52	12.32
3	2.0 Namų ūkiai su 1 išlaikomu vaiku	19	1.00
4	2.0 Namų ūkiai su 2 išlaikomais vaikais	5	0.26
5	2.0 Namų ūkiai su 3 ir daugiau išlaikomų vaikų	5	0.26
6	2.0 Namų ūkiai be išlaikomų vaikų	1873	98.48
7	3.0 Namų ūkiai be išlaikomų vaikų	902	100.00
8	4.0 Namų ūkiai su 1 išlaikomu vaiku	13	5.96
9	4.0 Namų ūkiai su 2 išlaikomais vaikais	24	11.01
10	4.0 Namų ūkiai su 3 ir daugiau išlaikomų vaikų	18	8.26
11	4.0 Namų ūkiai be išlaikomų vaikų	163	74.77
12	5.0 Namų ūkiai su 1 išlaikomu vaiku	189	58.51
13	5.0 Namų ūkiai su 2 išlaikomais vaikais	111	34.37
14	5.0 Namų ūkiai su 3 ir daugiau išlaikomų vaikų	23	7.12
15	6.0 Namų ūkiai be išlaikomų vaikų	567	100.00

Kategorinis kintamasis: Pajamų kvintilinės grupės				
cluster	qui_inc	count	percent	
0	1.0	1-a pajamų kvintilinė grupė	152	36.02
1	1.0	2-a pajamų kvintilinė grupė	114	27.01
2	1.0	3-a pajamų kvintilinė grupė	84	19.91
3	1.0	4-a pajamų kvintilinė grupė	51	12.09
4	1.0	5-a pajamų kvintilinė grupė	21	4.98
5	2.0	1-a pajamų kvintilinė grupė	398	20.93
6	2.0	2-a pajamų kvintilinė grupė	733	38.54
7	2.0	3-a pajamų kvintilinė grupė	514	27.02
8	2.0	4-a pajamų kvintilinė grupė	193	10.15
9	2.0	5-a pajamų kvintilinė grupė	64	3.36
10	3.0	1-a pajamų kvintilinė grupė	44	4.88
11	3.0	2-a pajamų kvintilinė grupė	131	14.52
12	3.0	3-a pajamų kvintilinė grupė	242	26.83
13	3.0	4-a pajamų kvintilinė grupė	229	25.39
14	3.0	5-a pajamų kvintilinė grupė	256	28.38
15	4.0	1-a pajamų kvintilinė grupė	123	56.42
16	4.0	2-a pajamų kvintilinė grupė	41	18.81
17	4.0	3-a pajamų kvintilinė grupė	27	12.39
18	4.0	4-a pajamų kvintilinė grupė	17	7.80
19	4.0	5-a pajamų kvintilinė grupė	10	4.59
20	5.0	1-a pajamų kvintilinė grupė	21	6.50
21	5.0	2-a pajamų kvintilinė grupė	42	13.00
22	5.0	3-a pajamų kvintilinė grupė	63	19.50
23	5.0	4-a pajamų kvintilinė grupė	117	36.22
24	5.0	5-a pajamų kvintilinė grupė	80	24.77
25	6.0	1-a pajamų kvintilinė grupė	13	2.29
26	6.0	2-a pajamų kvintilinė grupė	23	4.06
27	6.0	3-a pajamų kvintilinė grupė	79	13.93
28	6.0	4-a pajamų kvintilinė grupė	94	16.58
29	6.0	5-a pajamų kvintilinė grupė	358	63.14

Kategorinis kintamasis: Būsto valdos statusas				
cluster	hd01	count	percent	
0	1.0	Būsto savininkas, neturintis išmokėti paskolos...	328	77.73
1	1.0	Būsto savininkas, turintis neišmokėtą paskolą ...	41	9.72
2	1.0	Nuomininkas ar subnuomininkas, mokantis nuomą ...	7	1.66
3	1.0	Nuomojasi įj būstą mažesne nei rinkos kaina	13	3.08
4	1.0	Gyvena nemokėdamas nuomos	33	7.82
5	2.0	Būsto savininkas, neturintis išmokėti paskolos...	1799	94.58
6	2.0	Būsto savininkas, turintis neišmokėtą paskolą ...	8	0.42
7	2.0	Nuomininkas ar subnuomininkas, mokantis nuomą ...	3	0.16
8	2.0	Nuomojasi įj būstą mažesne nei rinkos kaina	23	1.21
9	2.0	Gyvena nemokėdamas nuomos	69	3.63
10	3.0	Būsto savininkas, neturintis išmokėti paskolos...	821	91.02
11	3.0	Būsto savininkas, turintis neišmokėtą paskolą ...	22	2.44
12	3.0	Nuomininkas ar subnuomininkas, mokantis nuomą ...	7	0.78
13	3.0	Nuomojasi įj būstą mažesne nei rinkos kaina	8	0.89
14	3.0	Gyvena nemokėdamas nuomos	44	4.88
15	4.0	Būsto savininkas, neturintis išmokėti paskolos...	160	73.39
16	4.0	Būsto savininkas, turintis neišmokėtą paskolą ...	5	2.29
17	4.0	Nuomininkas ar subnuomininkas, mokantis nuomą ...	3	1.38
18	4.0	Nuomojasi įj būstą mažesne nei rinkos kaina	8	3.67
19	4.0	Gyvena nemokėdamas nuomos	42	19.27
20	5.0	Būsto savininkas, neturintis išmokėti paskolos...	219	67.80
21	5.0	Būsto savininkas, turintis neišmokėtą paskolą ...	92	28.48
22	5.0	Nuomininkas ar subnuomininkas, mokantis nuomą ...	5	1.55
23	5.0	Gyvena nemokėdamas nuomos	7	2.17
24	6.0	Būsto savininkas, neturintis išmokėti paskolos...	501	88.36
25	6.0	Būsto savininkas, turintis neišmokėtą paskolą ...	38	6.70
26	6.0	Nuomininkas ar subnuomininkas, mokantis nuomą ...	8	1.41
27	6.0	Nuomojasi įj būstą mažesne nei rinkos kaina	2	0.35
28	6.0	Gyvena nemokėdamas nuomos	18	3.17

15 priedas. Vartojimo išlaidų pasiskirstymas klasteriuose

1 klasteris – Jaunos mažesnių pajamų šeimos su vaikais. Šio klasterio vartojimo struktūra rodo, kad didžiausia namų ūkių išlaidų dalis koncentruojasi būtinojo vartojimo kategorijose. Bendros mėnesinės vartojimo išlaidos vidutiniškai siekia 1089,66 Eur, o mediana – 1016,65 Eur. Santykinai didesnės maisto (vidurkis – 337,74 Eur; mediana – 303,86 Eur) ir būsto išlaidos (vidurkis – 155,91 Eur; mediana – 128,52 Eur) gali būti siejamos su vaikų išlaikymo poreikiais bei ribotesniais finansiniais ištekliais. Tuo pačiu mažesnės išlaidos poilsio, restoranų, kultūros ir asmeninių paslaugų kategorijose leidžia daryti prielaidą apie ribotas diskrecinio vartojimo galimybes. Tokia vartojimo struktūra atspindi ankstyvo gyvenimo ciklo namų ūkius, kurių vartojimo prioritetai orientuoti į pagrindinių poreikių užtikrinimą.

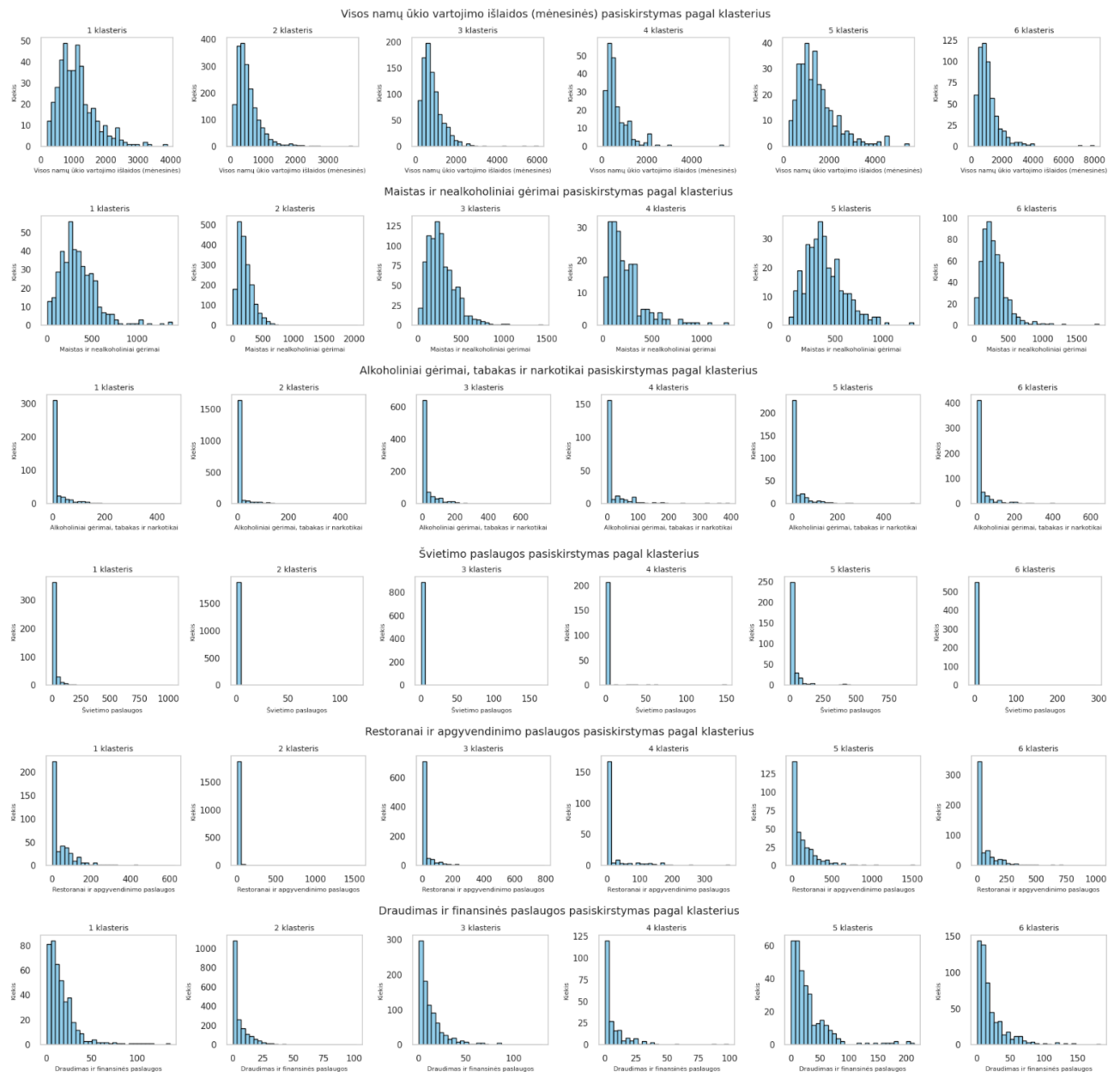
2 klasteris – Pensinio amžiaus namų ūkiai. Pensinio amžiaus namų ūkių vartojimo struktūra pasižymi dideliu išlaidų koncentravimu būtiniausiose vartojimo kategorijose. Bendros vartojimo išlaidos vidutiniškai siekia 531,46 Eur, o mediana – 442,92 Eur – tai mažiausios reikšmės tarp visų klasterių. Santykinai didesnės būsto (vidurkis – 110,21 Eur) ir sveikatos išlaidos (vidurkis – 65,20 Eur; mediana – 25,60 Eur) gali būti siejamos su amžiaus struktūra bei ribotu disponuojamų pajamų lygiu. Tuo tarpu restoranų (vidurkis – 4,28 Eur), transporto (31,89 Eur) ir aprangos (20,14 Eur) kategorijose stebimos mažesnės išlaidos. Tokia struktūra rodo santykinai konservatyvų vartojimo modelį, kuriame prioritetas teikiamas kasdieniams ir būtiniesiems poreikiams.

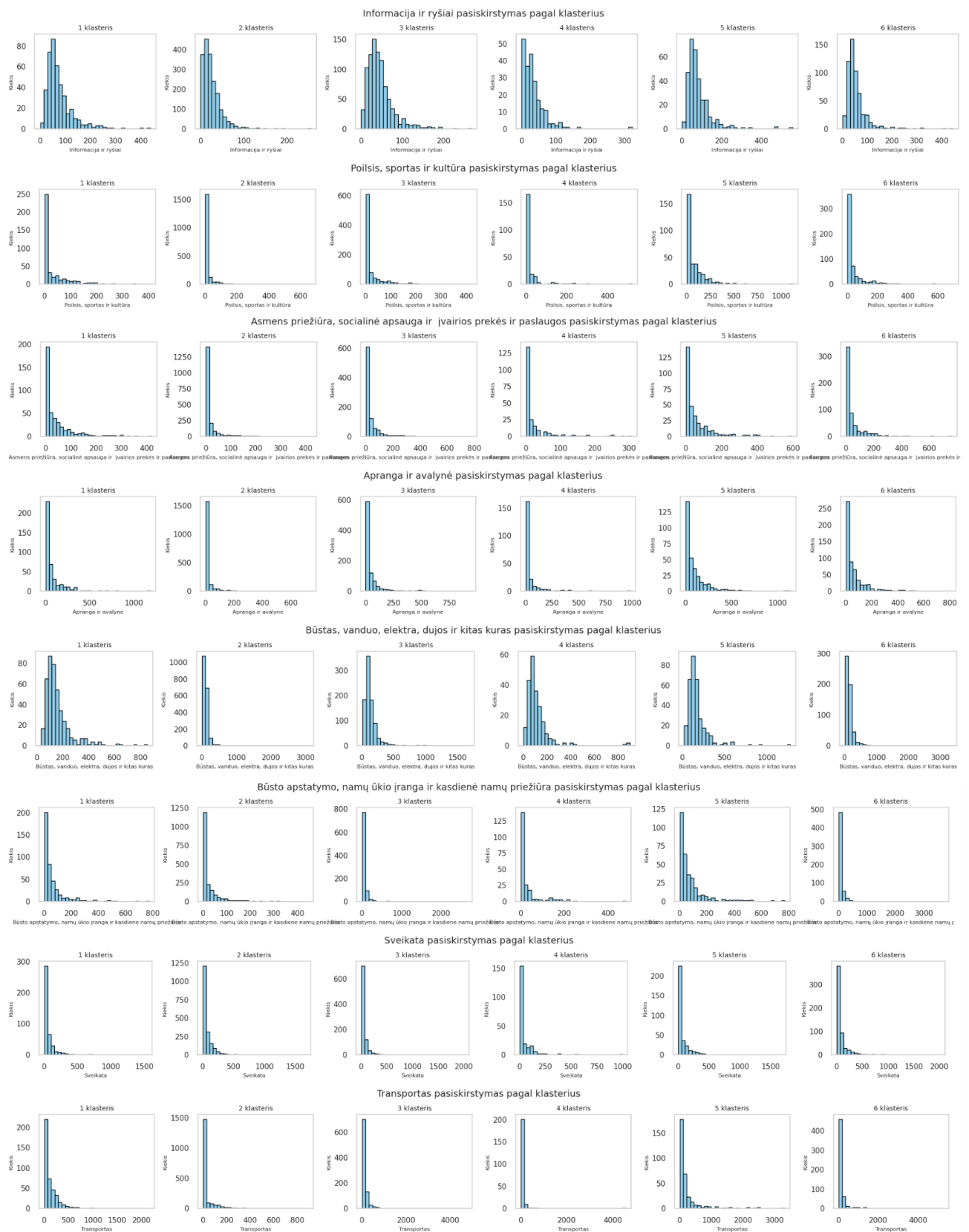
3 klasteris – Vidutinių–aukštesnių pajamų dirbantys namų ūkiai. Šio klasterio vartojimo struktūra atspindi ekonomiškai aktyvius namų ūkius, kuriems nebūdinga išlaikomų vaikų našta. Bendros mėnesinės vartojimo išlaidos vidutiniškai siekia 824,57 Eur, o mediana – 693,36 Eur. Mažesni įsipareigojimai šeimos išlaikymui sudaro galimybes didesnę disponuojamų pajamų dalį skirti transportui (vidurkis – 105,56 Eur), poilsiui, restoranams ir kitoms gyvenimo kokybę didinančioms kategorijoms. Taip pat stebimos santykinai didesnės draudimo ir finansinių paslaugų išlaidos (vidurkis – 12,72 Eur), kurios gali būti siejamos su aktyvesniu finansinių instrumentų naudojimu. Vartojimo struktūra tampa labiau diversifikuota ir mažiau orientuota vien į bazinių poreikių tenkinimą.

4 klasteris – Socialiai pažeidžiami namų ūkiai. Šios grupės vartojimo struktūra pasižymi aiškiu būtinojo vartojimo dominavimu. Bendros vartojimo išlaidos vidutiniškai siekia 659,44 Eur, o mediana – 479,78 Eur. Didesnė išlaidų dalis koncentruojasi maisto (236,50 Eur) ir būsto (114,32 Eur) kategorijose, tuo tarpu daugumoje kitų vartojimo sričių išlaidos išlieka nedidelės. Santykinai mažesnės poilsio, restoranų, kultūros ir aprangos išlaidos leidžia daryti prielaidą apie ribotas galimybes vartojimui, viršijančiam bazinių poreikių tenkinimą. Toks vartojimo modelis atspindi ribotų ekonominių išteklių poveikį vartojimo pasirinkimams.

5 klasteris – Jaunos ekonomiškai aktyvios šeimos. Šio klasterio vartojimo struktūra pasižymi didžiausiu vartojimo lygiu tarp visų grupių. Bendros mėnesinės vartojimo išlaidos vidutiniškai siekia 1458,77 Eur, o mediana – 1305,67 Eur. Didesnės būsto išlaidos (vidurkis – 169,82 Eur) gali būti siejamos su didžiausia būsto paskolų dalimi tarp visų grupių. Tuo pačiu aukštesnis išsilavinimo ir profesinės kvalifikacijos lygis sudaro galimybes daugiau išlaidų skirti restoranų (131,67 Eur), poilsio (77,04 Eur), asmens priežiūros (59,03 Eur) ir kitoms gyvenimo kokybę didinančioms kategorijoms. Nors vaikų buvimas išlaiko didesnę maisto išlaidų poreikį (385,90 Eur), geresnė ekonominė padėtis leidžia platesnes vartojimo galimybes.

6 klasteris – Aukštos kvalifikacijos ir aukštų pajamų namų ūkiai. Šio klasterio vartojimo struktūra pasižymi ne tik didesniu bendru išlaidų lygiu, bet ir didesne vartojimo įvairove. Bendros vartojimo išlaidos vidutiniškai siekia 1069,41 Eur, o mediana – 893,78 Eur. Santykinai didesnės išlaidos transportui (218,73 Eur), sveikatai (78,99 Eur), restoranams (68,48 Eur), draudimui ir finansinėms paslaugoms (20,36 Eur) rodo mažesnę vartojimo orientavimą į bazinių poreikių tenkinimą. Kadangi šioje grupėje dominuoja namų ūkiai be išlaikomų vaikų, didesnė disponuojamų pajamų dalis gali būti nukreipiama gyvenimo kokybei gerinti. Toks vartojimo modelis būdingas aukštesnio socialinio–ekonominio statuso grupėms, kurių vartojimo pasirinkimai tampa įvairesni ir mažiau apriboti finansinių veiksnių.





Visos namų ūkio vartojimo išlaidos (mėnesinės)

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	184.76	3891.77	1089.661232	574.412003	685.6950	1016.645	1346.7675
1	2	64.00	3744.47	531.460394	358.020344	290.7200	442.915	668.1050
2	3	106.41	6031.28	824.572062	544.443818	462.9425	693.355	1046.6725

3	4	56.23	5416.45	659.440183	593.277331	311.6225	479.780	826.2050
4	5	229.22	5474.17	1458.765697	851.861944	863.0300	1305.670	1810.3900
5	6	150.05	7984.83	1069.409189	756.084124	615.6300	893.780	1293.6950

Maistas ir nealkoholiniai gėrimai

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.00	1392.11	337.738128	205.525486	202.3800	303.855	450.8500
1	2	0.00	2053.07	204.907408	144.255679	109.6325	171.770	266.8725
2	3	0.00	1454.35	271.491330	170.238564	148.6100	240.565	360.4050
3	4	9.56	1272.85	236.496101	198.121402	108.1850	178.820	301.6500
4	5	6.22	1326.17	385.901579	205.959938	246.5900	353.100	506.1550
5	6	0.00	1829.28	287.100018	193.275640	160.0550	255.860	369.5800

Alkoholiniai gėrimai, tabakas ir narkotikai

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	468.44	21.830284	51.286790	0.0	0.0	21.3775
1	2	0.0	476.33	9.996420	33.486000	0.0	0.0	0.0000
2	3	0.0	744.46	32.079501	73.025356	0.0	0.0	33.0200
3	4	0.0	395.73	24.821835	57.212001	0.0	0.0	26.6575
4	5	0.0	534.46	28.060464	66.264653	0.0	0.0	33.0200
5	6	0.0	643.18	27.987231	66.892162	0.0	0.0	26.6000

Švietimo paslaugos

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	1035.00	15.957867	65.587081	0.0	0.0	0.000
1	2	0.0	116.66	0.104711	3.076694	0.0	0.0	0.000
2	3	0.0	166.66	0.783603	8.396862	0.0	0.0	0.000
3	4	0.0	150.00	2.535138	13.517793	0.0	0.0	0.000
4	5	0.0	916.63	34.074272	102.815721	0.0	0.0	22.915
5	6	0.0	291.66	2.314162	18.935345	0.0	0.0	0.000

Restoranai ir apgyvendinimo paslaugos

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	624.05	47.785711	74.251552	0.0	13.17	68.7300
1	2	0.0	1568.69	4.281052	50.352870	0.0	0.00	0.0000
2	3	0.0	790.83	25.561208	66.233010	0.0	0.00	17.2025
3	4	0.0	375.89	22.156009	52.368271	0.0	0.00	2.4975
4	5	0.0	1532.88	131.665015	181.041606	0.0	74.79	192.2800
5	6	0.0	1030.90	68.475062	122.797806	0.0	8.33	89.9050

Draudimas ir finansinės paslaugos

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
--	---------	-----	-----	----------	----	-----	-----	-----

0	1	0.0	136.50	16.944336	18.735546	5.6025	12.46	22.6275
1	2	0.0	101.86	5.943297	9.648919	0.0000	1.77	8.3300
2	3	0.0	130.36	12.716641	15.405647	2.0000	8.33	16.7900
3	4	0.0	99.13	7.343624	12.635183	0.0000	2.00	9.8950
4	5	0.0	213.45	29.956440	34.121564	8.4900	19.17	37.5000
5	6	0.0	184.16	20.362963	24.098650	6.0450	12.33	26.2900

Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	418.36	42.359242	63.102179	0.0	16.915	55.6625
1	2	0.0	424.15	15.741246	35.847339	0.0	0.000	14.7700
2	3	0.0	778.57	32.305787	62.907523	0.0	9.520	36.4975
3	4	0.0	307.42	24.510275	51.093742	0.0	3.590	21.5400
4	5	0.0	592.15	59.028421	87.798801	0.0	25.680	78.1250
5	6	0.0	758.30	46.103280	80.490795	0.0	17.290	50.5500

Apranga ir avalynė

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	1186.25	92.489692	169.183312	0.0	32.13	99.1675
1	2	0.0	736.41	20.143938	61.883323	0.0	0.00	11.1225
2	3	0.0	940.14	46.558714	100.055352	0.0	4.91	48.2475
3	4	0.0	986.33	39.573211	103.203421	0.0	0.00	33.2925
4	5	0.0	1139.81	101.933839	149.811747	0.0	52.62	131.3700
5	6	0.0	805.45	66.553827	106.462599	0.0	30.33	80.0000

Būstas, vanduo, elektra, dujos ir kitas kuras

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	32.00	855.33	155.914882	104.267659	93.3625	128.515	180.9975
1	2	0.00	3080.47	110.212781	106.901208	62.3025	91.285	131.9775
2	3	12.12	1686.09	132.919545	113.888298	73.5000	106.825	154.5725
3	4	0.00	897.66	114.320459	113.683929	59.7800	85.880	136.5500
4	5	22.09	1283.99	169.824241	124.854676	101.3450	138.170	193.0250
5	6	15.00	3339.30	159.766772	186.747926	84.8700	123.850	183.5900

Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	773.44	59.311564	94.513656	6.950	28.180	68.9675
1	2	0.0	445.30	24.310710	45.408514	0.000	5.520	29.2825
2	3	0.0	2653.59	46.570831	128.063299	0.000	19.775	54.7525
3	4	0.0	482.65	30.764633	60.531994	0.000	4.325	29.9350
4	5	0.0	770.80	76.404458	110.471441	9.385	40.710	91.6300

5	6	0.0	3691.87	65.213580	175.364165	0.000	27.160	75.4150
---	---	-----	---------	-----------	------------	-------	--------	---------

Sveikata

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	1540.40	58.478957	138.084183	0.0	0.00	69.1100
1	2	0.0	1666.65	65.198780	110.985084	0.0	25.60	90.7500
2	3	0.0	1999.98	48.395044	129.461607	0.0	0.00	56.1275
3	4	0.0	999.99	39.686697	96.648873	0.0	0.00	48.7500
4	5	0.0	1666.65	65.186594	133.756387	0.0	20.00	72.8300
5	6	0.0	1999.98	78.992787	168.326037	0.0	16.67	98.3100

Transportas

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	2158.68	129.679834	206.196653	0.00	65.610	181.3700
1	2	0.0	891.63	31.890226	81.843866	0.00	0.000	13.3300
2	3	0.0	4759.87	105.557982	229.118185	0.00	40.415	140.3325
3	4	0.0	4566.49	62.814358	321.183261	0.00	0.000	50.0675
	5							
4	6	0.0	3305.30	218.734737	424.060711	1.96	86.900	204.9300
5	5.0	0.0	5291.96	149.720212	377.798764	0.00	42.500	143.3600

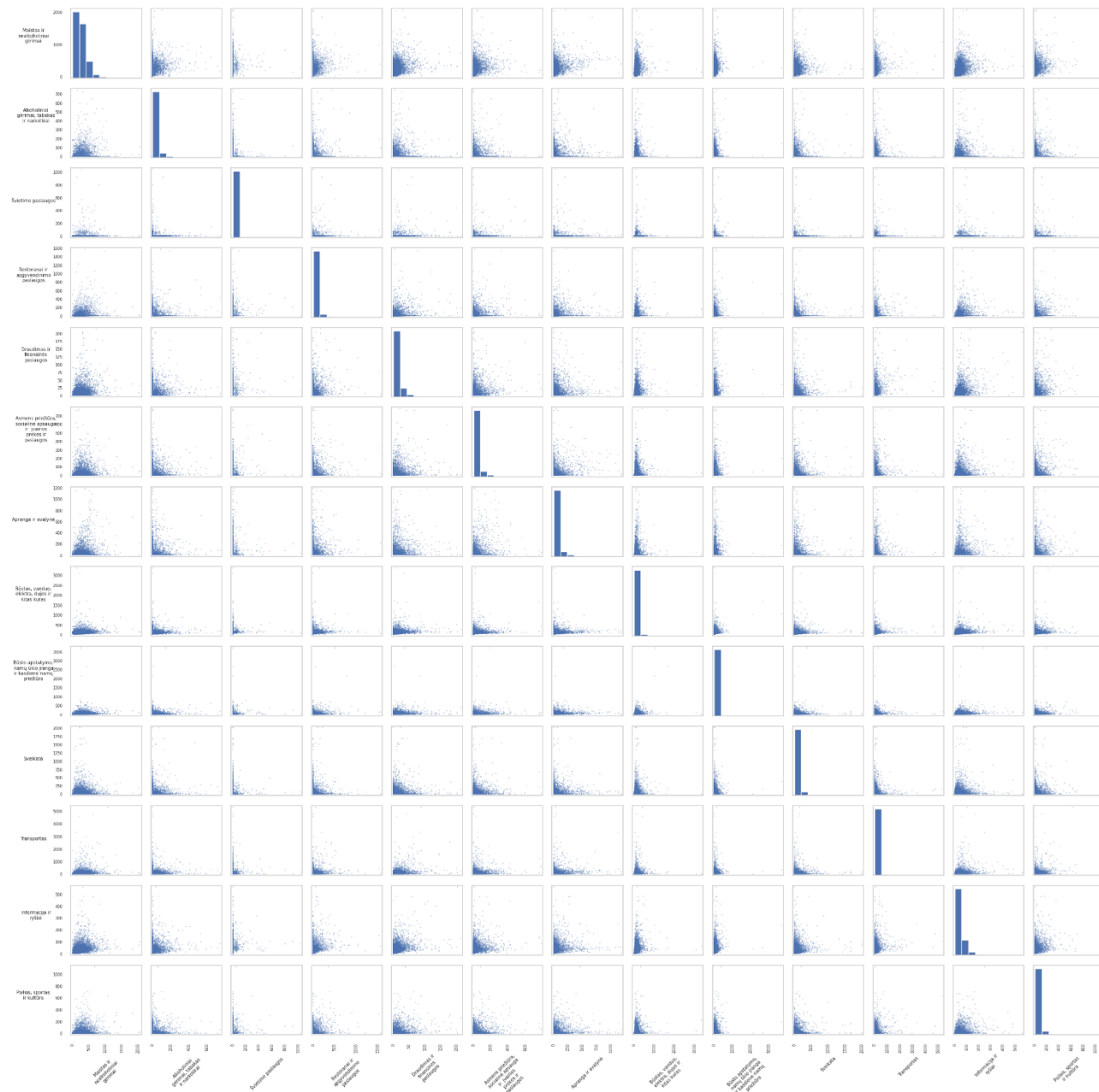
Informacija ir ryšiai

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	438.99	75.322441	54.743220	41.25	60.00	89.9975
1	2	0.0	255.00	24.799096	21.925018	10.00	20.00	33.0000
2	3	0.0	266.99	44.649412	31.514129	24.00	38.00	55.3150
3	4	0.0	321.66	33.802706	33.727016	11.25	26.24	44.0000
4	5	0.0	558.32	80.954644	65.004861	43.00	65.00	99.8300
5	6	0.0	446.65	53.826208	46.330054	29.00	41.00	62.0000

Poilsis, sportas ir kultūra

	cluster	Min	Max	Vidurkis	Sd	25%	50%	75%
0	1	0.0	407.19	35.849171	61.244206	0.000	7.535	46.2025
1	2	0.0	662.60	13.931215	35.997539	0.000	0.000	9.8400
2	3	0.0	421.55	24.983348	50.233911	0.000	3.315	25.2375
3	4	0.0	521.84	20.616009	56.355469	0.000	0.000	16.1300
4	5	0.0	1121.28	77.041950	113.934353	3.615	32.920	110.9600
5	6	0.0	700.21	42.993792	86.666223	0.000	7.920	44.3300

16 priedas. Vartojimo išlaidų kategorijų porinių sklaidos diagramų matrica



17 priedas. Vartojimo išlaidų kintamųjų dispersijos tikrinimas

 KINTAMASIS: Visos namų ūkio vartojimo išlaidos (mėnesinės)

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	64.8734	0.0	False
1	Bartlett	6	851.7202	0.0	False

 KINTAMASIS: Maistas ir nealkoholiniai gėrimai

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	24.6378	0.0	False
1	Bartlett	6	184.8558	0.0	False

 KINTAMASIS: Alkoholiniai gėrimai, tabakas ir narkotikai

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	26.2691	0.0	False
1	Bartlett	6	943.6343	0.0	False

 KINTAMASIS: Apranga ir avalynė

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	56.9225	0.0	False
1	Bartlett	6	1136.1920	0.0	False

 KINTAMASIS: Būstas, vanduo, elektra, dujos ir kitas kuras

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	6.5912	0.000004	False
1	Bartlett	6	367.4821	0.000000	False

 KINTAMASIS: Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	19.2633	0.0	False
1	Bartlett	6	2248.0949	0.0	False

 KINTAMASIS: Sveikata

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	5.5688	0.00004	False
1	Bartlett	6	208.2307	0.00000	False

KINTAMASIS: Transportas

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	49.1986	0.0	False
1	Bartlett	6	3109.9679	0.0	False

KINTAMASIS: Informacija ir ryšiai

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	60.1607	0.0	False
1	Bartlett	6	1346.7403	0.0	False

KINTAMASIS: Poilsis, sportas ir kultūra

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	67.2805	0.0	False
1	Bartlett	6	1401.6909	0.0	False

KINTAMASIS: Švietimo paslaugos

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	61.0136	0.0	False
1	Bartlett	6	11881.3126	0.0	False

KINTAMASIS: Restoranai ir apgyvendinimo paslaugos

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	152.0498	0.0	False
1	Bartlett	6	1790.6487	0.0	False

KINTAMASIS: Draudimas ir finansinės paslaugos

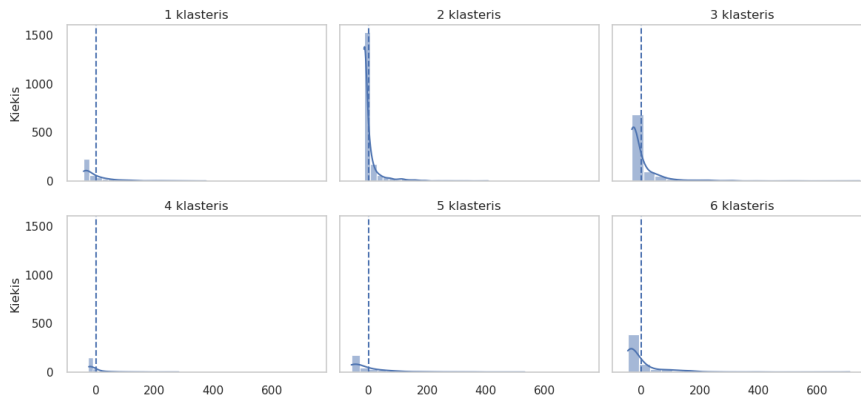
	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	83.5970	0.0	False
1	Bartlett	6	1621.8281	0.0	False

KINTAMASIS: Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos

	Testas	Grupių sk.	Statistika (F)	p-reikšmė	Dispersijos lygios (p > 0.05)
0	Levene (Brown-Forsythe)	6	44.2935	0.0	False
1	Bartlett	6	980.7551	0.0	False

18 priedas. ANOVA modelio liekanų normalumo tikrinimas

Visos namų ūkio vartojimo išlaidos (mėnesinės)



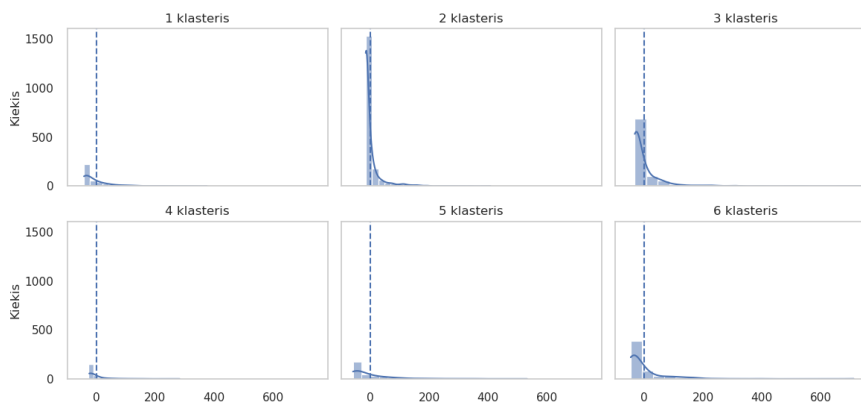
Klasrteris	Testas	Statistika	p-reikšmė
0	1 Shapiro-Wilk	0.6903	0.000
1	1 Lilliefors	0.2510	0.001
2	1 Cramer-von Mises	7.2425	0.000

0	2 Shapiro-Wilk	0.5359	0.000
1	2 Lilliefors	0.3303	0.001
2	2 Cramer-von Mises	64.8660	0.000

0	3 Shapiro-Wilk	0.5406	0.000
1	3 Lilliefors	0.3038	0.001
2	3 Cramer-von Mises	23.0822	0.000

0	4 Shapiro-Wilk	0.5359	0.000
1	4 Lilliefors	0.3157	0.001
2	4 Cramer-von Mises	7.1227	0.000

Maistas ir nealkoholiniai gėrimai



Klasrteris	Testas	Statistika	p-reikšmė
0	1 Shapiro-Wilk	0.6903	0.000
1	1 Lilliefors	0.2510	0.001
2	1 Cramer-von Mises	7.2425	0.000

0	2 Shapiro-Wilk	0.4861	0.000
1	2 Lilliefors	0.3303	0.001
2	2 Cramer-von Mises	64.8660	0.000

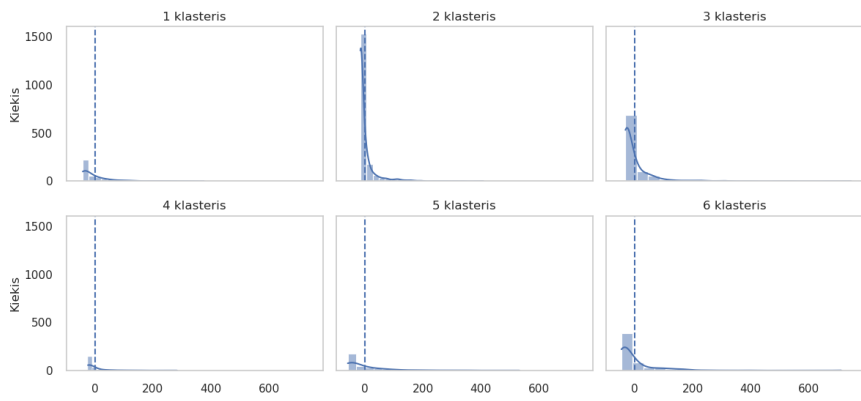
0	3 Shapiro-Wilk	0.5406	0.000
1	3 Lilliefors	0.3038	0.001
2	3 Cramer-von Mises	23.0822	0.000

0	4 Shapiro-Wilk	0.5359	0.000
1	4 Lilliefors	0.3157	0.001
2	4 Cramer-von Mises	7.1227	0.000

0	5 Shapiro-Wilk	0.6851	0.000
1	5 Lilliefors	0.2507	0.001
2	5 Cramer-von Mises	5.5049	0.000

0	6 Shapiro-Wilk	0.5759	0.000
1	6 Lilliefors	0.2834	0.001
2	6 Cramer-von Mises	12.9616	0.000

Alkoholiniai gėrimai, tabakas ir narkotikai



Klasrteris	Testas	Statistika	p-reikšmė
0	1 Shapiro-Wilk	0.6903	0.000
1	1 Lilliefors	0.2510	0.001
2	1 Cramer-von Mises	7.2425	0.000

0	2 Shapiro-Wilk	0.4861	0.000
1	2 Lilliefors	0.3303	0.001
2	2 Cramer-von Mises	64.8660	0.000

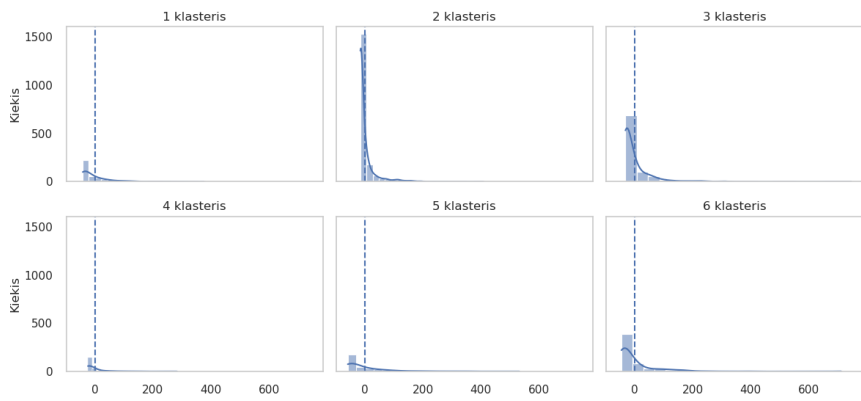
0	3 Shapiro-Wilk	0.5406	0.000
1	3 Lilliefors	0.3038	0.001
2	3 Cramer-von Mises	23.0822	0.000

0	4 Shapiro-Wilk	0.5359	0.000
1	4 Lilliefors	0.3157	0.001
2	4 Cramer-von Mises	7.1227	0.000

0	5 Shapiro-Wilk	0.6851	0.000
1	5 Lilliefors	0.2507	0.001
2	5 Cramer-von Mises	5.5049	0.000

0	6 Shapiro-Wilk	0.5759	0.000
1	6 Lilliefors	0.2834	0.001
2	6 Cramer-von Mises	12.9616	0.000

Apranga ir avalynė



Klasrteris	Testas	Statistika	p-reikšmė
0	1 Shapiro-Wilk	0.6903	0.000
1	1 Lilliefors	0.2510	0.001
2	1 Cramer-von Mises	7.2425	0.000

0	2 Shapiro-Wilk	0.4861	0.000
1	2 Lilliefors	0.3303	0.001
2	2 Cramer-von Mises	64.8660	0.000

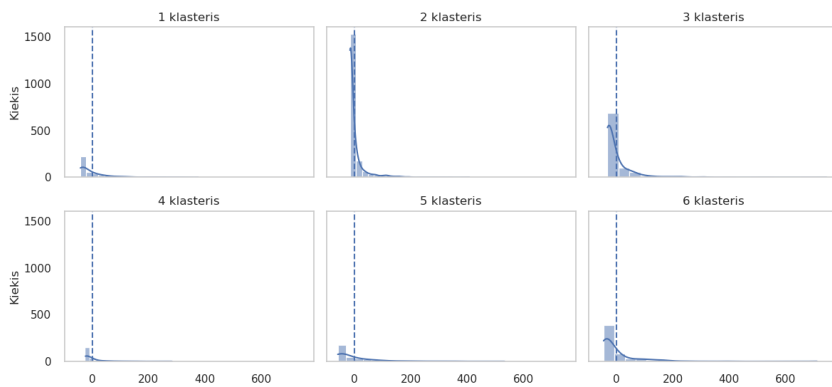
0	3 Shapiro-Wilk	0.5406	0.000
1	3 Lilliefors	0.3038	0.001
2	3 Cramer-von Mises	23.0822	0.000

0	4 Shapiro-Wilk	0.5359	0.000
1	4 Lilliefors	0.3157	0.001
2	4 Cramer-von Mises	7.1227	0.000

0	5 Shapiro-Wilk	0.6851	0.000
1	5 Lilliefors	0.2507	0.001
2	5 Cramer-von Mises	5.5049	0.000

0	6 Shapiro-Wilk	0.5759	0.000
1	6 Lilliefors	0.2834	0.001
2	6 Cramer-von Mises	12.9616	0.000

Būstas, vanduo, elektra, dujos ir kitas kuras



Klasrteris	Testas	Statistika	p-reikšmė
0	1 Shapiro-Wilk	0.6903	0.000
1	1 Lilliefors	0.2510	0.001
2	1 Cramer-von Mises	7.2425	0.000

0	2 Shapiro-Wilk	0.4861	0.000
1	2 Lilliefors	0.3303	0.001
2	2 Cramer-von Mises	64.8660	0.000

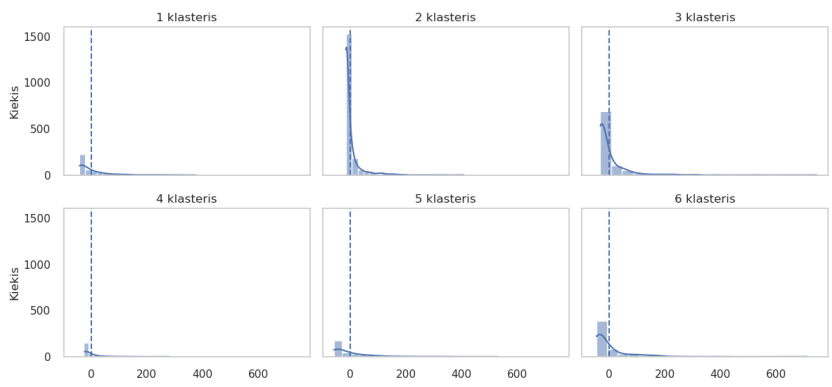
0	3 Shapiro-Wilk	0.5406	0.000
1	3 Lilliefors	0.3038	0.001
2	3 Cramer-von Mises	23.0822	0.000

0	4 Shapiro-Wilk	0.5359	0.000
1	4 Lilliefors	0.3157	0.001
2	4 Cramer-von Mises	7.1227	0.000

0	5 Shapiro-Wilk	0.6851	0.000
1	5 Lilliefors	0.2507	0.001
2	5 Cramer-von Mises	5.5049	0.000

0	6 Shapiro-Wilk	0.5759	0.000
1	6 Lilliefors	0.2834	0.001
2	6 Cramer-von Mises	12.9616	0.000

Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra



Klasrteris	Testas	Statistika	p-reikšmė
0	1 Shapiro-Wilk	0.6903	0.000
1	1 Lilliefors	0.2510	0.001
2	1 Cramer-von Mises	7.2425	0.000

0	2 Shapiro-Wilk	0.4861	0.000
1	2 Lilliefors	0.3303	0.001
2	2 Cramer-von Mises	64.8660	0.000

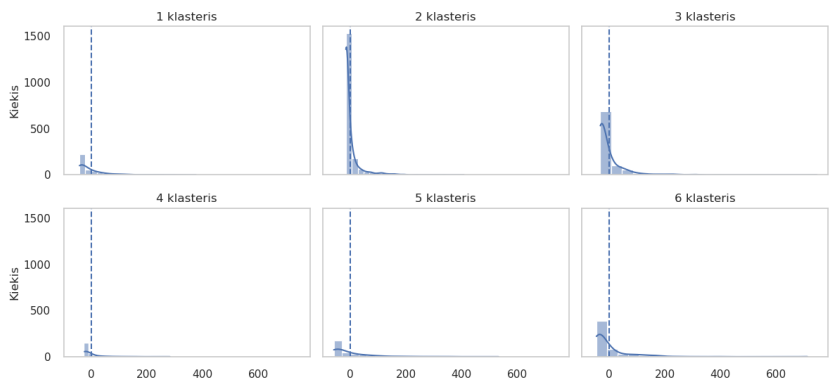
0	3 Shapiro-Wilk	0.5406	0.000
1	3 Lilliefors	0.3038	0.001
2	3 Cramer-von Mises	23.0822	0.000

0	4 Shapiro-Wilk	0.5359	0.000
1	4 Lilliefors	0.3157	0.001
2	4 Cramer-von Mises	7.1227	0.000

0	5 Shapiro-Wilk	0.6851	0.000
1	5 Lilliefors	0.2507	0.001
2	5 Cramer-von Mises	5.5049	0.000

0	6 Shapiro-Wilk	0.5759	0.000
1	6 Lilliefors	0.2834	0.001
2	6 Cramer-von Mises	12.9616	0.000

Sveikata



Klasrteris	Testas	Statistika	p-reikšmė
0	1 Shapiro-Wilk	0.6903	0.000
1	1 Lilliefors	0.2510	0.001
2	1 Cramer-von Mises	7.2425	0.000

0	2 Shapiro-Wilk	0.4861	0.000
1	2 Lilliefors	0.3303	0.001
2	2 Cramer-von Mises	64.8660	0.000

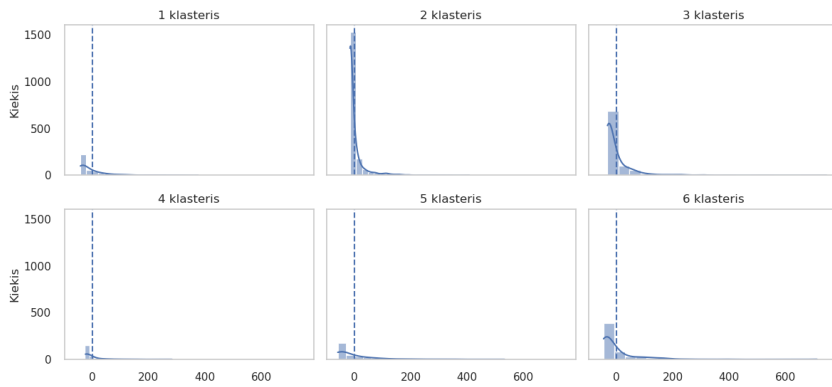
0	3 Shapiro-Wilk	0.5406	0.000
1	3 Lilliefors	0.3038	0.001
2	3 Cramer-von Mises	23.0822	0.000

0	4 Shapiro-Wilk	0.5359	0.000
1	4 Lilliefors	0.3157	0.001
2	4 Cramer-von Mises	7.1227	0.000

0	5 Shapiro-Wilk	0.6851	0.000
1	5 Lilliefors	0.2507	0.001
2	5 Cramer-von Mises	5.5049	0.000

0	6 Shapiro-Wilk	0.5759	0.000
1	6 Lilliefors	0.2834	0.001
2	6 Cramer-von Mises	12.9616	0.000

Transportas



Klasrteris	Testas	Statistika	p-reikšmė
0	1 Shapiro-Wilk	0.6903	0.000
1	1 Lilliefors	0.2510	0.001
2	1 Cramer-von Mises	7.2425	0.000

0	2 Shapiro-Wilk	0.4861	0.000
1	2 Lilliefors	0.3303	0.001
2	2 Cramer-von Mises	64.8660	0.000

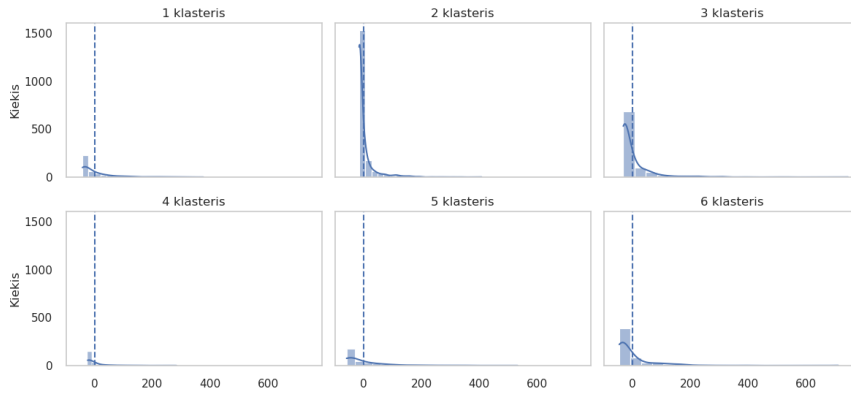
0	3 Shapiro-Wilk	0.5406	0.000
1	3 Lilliefors	0.3038	0.001
2	3 Cramer-von Mises	23.0822	0.000

0	4 Shapiro-Wilk	0.5359	0.000
1	4 Lilliefors	0.3157	0.001
2	4 Cramer-von Mises	7.1227	0.000

0	5 Shapiro-Wilk	0.6851	0.000
1	5 Lilliefors	0.2507	0.001
2	5 Cramer-von Mises	5.5049	0.000

0	6 Shapiro-Wilk	0.5759	0.000
1	6 Lilliefors	0.2834	0.001
2	6 Cramer-von Mises	12.9616	0.000

Informacija ir ryšiai



Klasrteris	Testas	Statistika	p-reikšmė
0	1	Shapiro-Wilk	0.6903
1	1	Lilliefors	0.2510
2	1	Cramer-von Mises	7.2425

0	2	Shapiro-Wilk	0.4861
1	2	Lilliefors	0.3303
2	2	Cramer-von Mises	64.8660

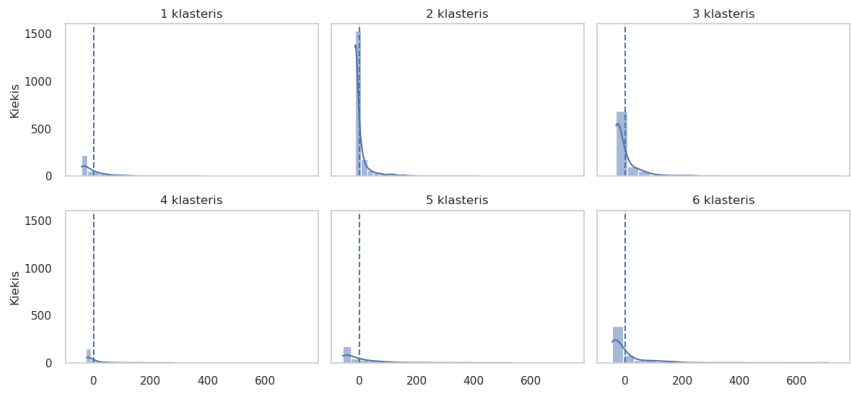
0	3	Shapiro-Wilk	0.5406
1	3	Lilliefors	0.3038
2	3	Cramer-von Mises	23.0822

0	4	Shapiro-Wilk	0.5359
1	4	Lilliefors	0.3157
2	4	Cramer-von Mises	7.1227

0	5	Shapiro-Wilk	0.6851
1	5	Lilliefors	0.2507
2	5	Cramer-von Mises	5.5049

0	6	Shapiro-Wilk	0.5759
1	6	Lilliefors	0.2834
2	6	Cramer-von Mises	12.9616

Poilsis, sportas ir kultūra



Klasrteris	Testas	Statistika	p-reikšmė
0	1	Shapiro-Wilk	0.6903
1	1	Lilliefors	0.2510
2	1	Cramer-von Mises	7.2425

0	2	Shapiro-Wilk	0.4861
1	2	Lilliefors	0.3303
2	2	Cramer-von Mises	64.8660

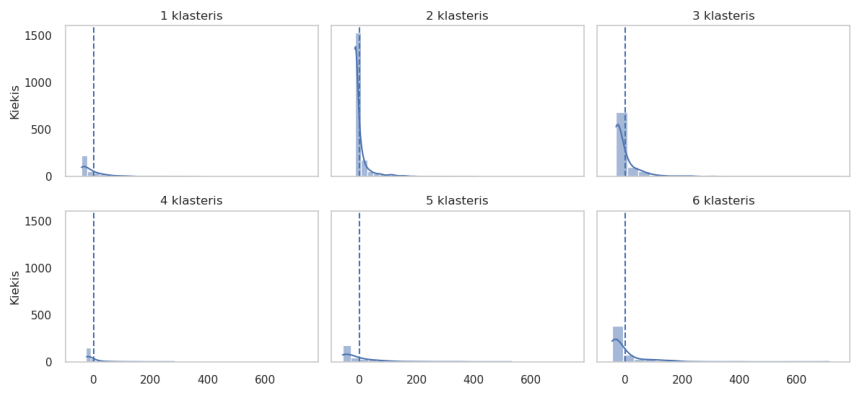
0	3	Shapiro-Wilk	0.5406
1	3	Lilliefors	0.3038
2	3	Cramer-von Mises	23.0822

0	4	Shapiro-Wilk	0.5359
1	4	Lilliefors	0.3157
2	4	Cramer-von Mises	7.1227

0	5	Shapiro-Wilk	0.6851
1	5	Lilliefors	0.2507
2	5	Cramer-von Mises	5.5049

0	6	Shapiro-Wilk	0.5759
1	6	Lilliefors	0.2834
2	6	Cramer-von Mises	12.9616

Švietimo paslaugos



Klasrteris	Testas	Statistika	p-reikšmė
0	1	Shapiro-Wilk	0.6903
1	1	Lilliefors	0.2510
2	1	Cramer-von Mises	7.2425

0	2	Shapiro-Wilk	0.4861
1	2	Lilliefors	0.3303
2	2	Cramer-von Mises	64.8660

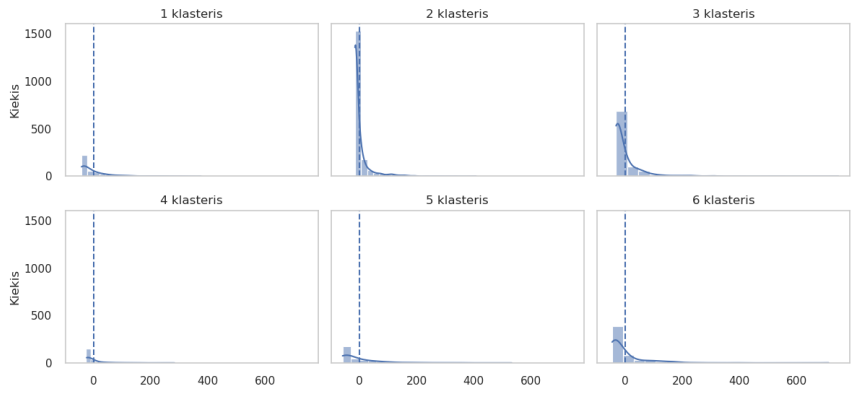
0	3	Shapiro-Wilk	0.5406
1	3	Lilliefors	0.3038
2	3	Cramer-von Mises	23.0822

0	4	Shapiro-Wilk	0.5359
1	4	Lilliefors	0.3157
2	4	Cramer-von Mises	7.1227

0	5	Shapiro-Wilk	0.6851
1	5	Lilliefors	0.2507
2	5	Cramer-von Mises	5.5049

0	6	Shapiro-Wilk	0.5759
1	6	Lilliefors	0.2834
2	6	Cramer-von Mises	12.9616

Restoranai ir apgyvendinimo paslaugos



Klasrteris	Testas	Statistika	p-reikšmė
0	1	Shapiro-Wilk	0.6903
1	1	Lilliefors	0.2510
2	1	Cramer-von Mises	7.2425

0	2	Shapiro-Wilk	0.4861
1	2	Lilliefors	0.3303
2	2	Cramer-von Mises	64.8660

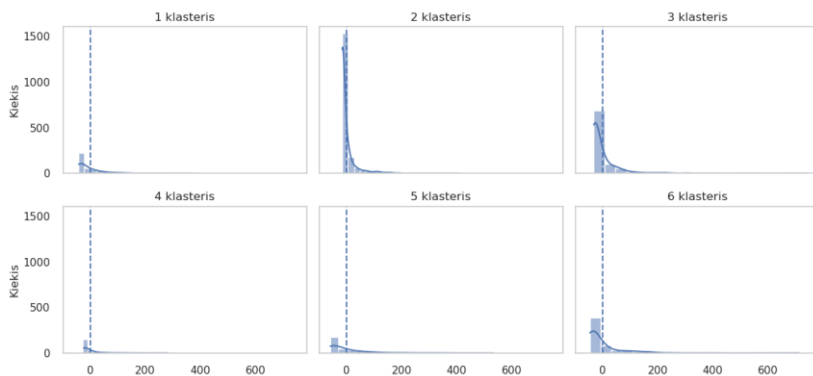
0	3	Shapiro-Wilk	0.5406
1	3	Lilliefors	0.3038
2	3	Cramer-von Mises	23.0822

0	4	Shapiro-Wilk	0.5359
1	4	Lilliefors	0.3157
2	4	Cramer-von Mises	7.1227

0	5	Shapiro-Wilk	0.6851
1	5	Lilliefors	0.2507
2	5	Cramer-von Mises	5.5049

0	6	Shapiro-Wilk	0.5759
1	6	Lilliefors	0.2834
2	6	Cramer-von Mises	12.9616

Draudimas ir finansinės paslaugos



Klasrteris	Testas	Statistika	p-reikšmė
0	1	Shapiro-Wilk	0.6903
1	1	Lilliefors	0.2510
2	1	Cramer-von Mises	7.2425

0	2	Shapiro-Wilk	0.4861
1	2	Lilliefors	0.3303
2	2	Cramer-von Mises	64.8660

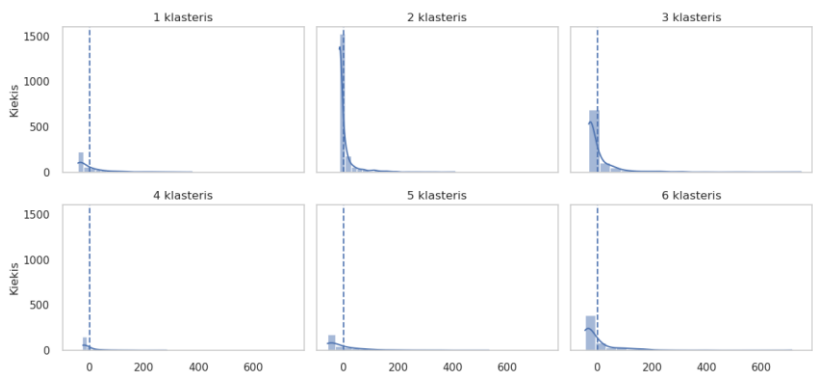
0	3	Shapiro-Wilk	0.5406
1	3	Lilliefors	0.3038
2	3	Cramer-von Mises	23.0822

0	4	Shapiro-Wilk	0.5359
1	4	Lilliefors	0.3157
2	4	Cramer-von Mises	7.1227

0	5	Shapiro-Wilk	0.6851
1	5	Lilliefors	0.2507
2	5	Cramer-von Mises	5.5049

0	6	Shapiro-Wilk	0.5759
1	6	Lilliefors	0.2834
2	6	Cramer-von Mises	12.9616

Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos



Klasrteris	Testas	Statistika	p-reikšmė
0	1	Shapiro-Wilk	0.6903
1	1	Lilliefors	0.2510
2	1	Cramer-von Mises	7.2425

0	2	Shapiro-Wilk	0.4861
1	2	Lilliefors	0.3303
2	2	Cramer-von Mises	64.8660

0	3	Shapiro-Wilk	0.5406
1	3	Lilliefors	0.3038
2	3	Cramer-von Mises	23.0822

0	4	Shapiro-Wilk	0.5359
1	4	Lilliefors	0.3157
2	4	Cramer-von Mises	7.1227

0	5	Shapiro-Wilk	0.6851
1	5	Lilliefors	0.2507
2	5	Cramer-von Mises	5.5049

0	6	Shapiro-Wilk	0.5759
1	6	Lilliefors	0.2834
2	6	Cramer-von Mises	12.9616

19 priedas. Klasterių skirstinių homogeniškumo vertinimas taikant Andersono–Darlingo testas

	Kintamasis	Klasterių sk.	Anderson-Darling	p-reikšmė	Ar H0 atmetama
0	Visos namų ūkio vartojimo išlaidos (mėnesinės)	6	346.5540	0.0002	True
1	Maistas ir nealkoholiniai gėrimai	6	146.8594	0.0002	True
2	Alkoholiniai gėrimai, tabakas ir narkotikai	6	90.3904	0.0002	True
3	Švietimo paslaugos	6	383.5237	0.0002	True
4	Restoranai ir apgyvendinimo paslaugos	6	583.9931	0.0002	True
5	Draudimas ir finansinės paslaugos	6	287.7621	0.0002	True
6	Asmens priežiūra, socialinė apsauga ir įvairi...	6	134.3694	0.0002	True
7	Apranga ir avalynė	6	283.0620	0.0002	True
8	Būstas, vanduo, elektra, dujos ir kitas kuras	6	95.1869	0.0002	True
9	Būsto apstatymo, namų ūkio įranga ir kasdienė ...	6	114.7854	0.0002	True
10	Sveikata	6	33.5684	0.0002	True
11	Transportas	6	276.5660	0.0002	True
12	Informacija ir ryšiai	6	407.6261	0.0002	True
13	Poilsis, sportas ir kultūra	6	126.5406	0.0002	True

20 priedas.

KINTAMASIS: Visos namų ūkio vartojimo išlaidos (mėnesinės)

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.5191	0.0000	317.9359	0.0010	True	True
1	1-3	422	902	0.2520	0.0000	56.1127	0.0010	True	True
2	1-4	422	218	0.4750	0.0000	86.0029	0.0010	True	True
3	1-5	422	323	0.2400	0.0000	26.3714	0.0010	True	True
4	1-6	422	567	0.0997	0.0149	3.4084	0.0135	True	True
5	2-3	1902	902	0.3143	0.0000	196.3734	0.0010	True	True
6	2-4	1902	218	0.1128	0.0126	5.4960	0.0023	True	True
7	2-5	1902	323	0.6286	0.0000	391.2903	0.0010	True	True
8	2-6	1902	567	0.4620	0.0000	311.6694	0.0010	True	True
9	3-4	902	218	0.2742	0.0000	31.5603	0.0010	True	True
10	3-5	902	323	0.4103	0.0000	133.4250	0.0010	True	True
11	3-6	902	567	0.1761	0.0000	35.6919	0.0010	True	True
12	4-5	218	323	0.5689	0.0000	120.0238	0.0010	True	True
13	4-6	218	567	0.4207	0.0000	73.3687	0.0010	True	True
14	5-6	323	567	0.2659	0.0000	42.6726	0.0010	True	True

KINTAMASIS: Maistas ir nealkoholiniai gėrimai

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.3712	0.0000	149.6131	0.0010	True	True
1	1-3	422	902	0.1745	0.0000	25.1338	0.0010	True	True
2	1-4	422	218	0.3127	0.0000	39.3206	0.0010	True	True
3	1-5	422	323	0.1317	0.0031	7.5688	0.0010	True	True
4	1-6	422	567	0.1406	0.0001	13.5049	0.0010	True	True
5	2-3	1902	902	0.2133	0.0000	82.2540	0.0010	True	True
6	2-4	1902	218	0.0823	0.1330	2.6697	0.0263	False	True
7	2-5	1902	323	0.4637	0.0000	198.5545	0.0010	True	True
8	2-6	1902	567	0.2406	0.0000	80.7444	0.0010	True	True
9	3-4	902	218	0.1840	0.0000	13.0645	0.0010	True	True
10	3-5	902	323	0.2986	0.0000	58.9320	0.0010	True	True
11	3-6	902	567	0.0476	0.3924	0.3994	0.2282	False	False
12	4-5	218	323	0.3967	0.0000	61.4389	0.0010	True	True
13	4-6	218	567	0.2168	0.0000	16.9095	0.0010	True	True

KINTAMASIS: Alkoholiniai gėrimai, tabakas ir narkotikai

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.1432	0.0000	57.0642	0.0010	True	True
1	1-3	422	902	0.0619	0.2096	3.9263	0.0086	False	True
2	1-4	422	218	0.0420	0.9492	-0.8905	0.2500	False	False
3	1-5	422	323	0.0603	0.4947	-0.3741	0.2500	False	False
4	1-6	422	567	0.0350	0.9152	-0.3645	0.2500	False	False
5	2-3	1902	902	0.1984	0.0000	169.7462	0.0010	True	True
6	2-4	1902	218	0.1656	0.0000	44.2980	0.0010	True	True
7	2-5	1902	323	0.1711	0.0000	64.3021	0.0010	True	True
8	2-6	1902	567	0.1678	0.0000	92.7401	0.0010	True	True
9	3-4	902	218	0.0579	0.5749	0.1190	0.2500	False	False
10	3-5	902	323	0.0413	0.7913	-0.0259	0.2500	False	False
11	3-6	902	567	0.0431	0.5182	0.8418	0.1475	False	False
12	4-5	218	323	0.0350	0.9951	-1.1873	0.2500	False	False
13	4-6	218	567	0.0356	0.9831	-1.1857	0.2500	False	False
14	5-6	323	567	0.0400	0.8804	-1.1486	0.2500	False	False

KINTAMASIS: Švietimo paslaugos

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.1875	0.0000	448.3435	0.0010	True	True
1	1-3	422	902	0.1763	0.0000	174.0851	0.0010	True	True
2	1-4	422	218	0.1391	0.0067	26.8880	0.0010	True	True
3	1-5	422	323	0.0976	0.0562	10.5063	0.0010	False	True
4	1-6	422	567	0.1596	0.0000	85.8138	0.0010	True	True
5	2-3	1902	902	0.0117	1.0000	16.4209	0.0010	False	True
6	2-4	1902	218	0.0489	0.7175	83.9199	0.0010	False	True
7	2-5	1902	323	0.2827	0.0000	686.2823	0.0010	True	True
8	2-6	1902	567	0.0279	0.8748	51.3826	0.0010	False	True
9	3-4	902	218	0.0383	0.9488	14.3406	0.0010	False	True
10	3-5	902	323	0.2715	0.0000	283.2790	0.0010	True	True
11	3-6	902	567	0.0167	0.9999	5.2078	0.0029	False	True
12	4-5	218	323	0.2344	0.0000	54.8388	0.0010	True	True
13	4-6	218	567	0.0258	0.9998	1.1719	0.1071	False	False
14	5-6	323	567	0.2566	0.0000	152.6964	0.0010	True	True

KINTAMASIS: Restoranai ir apgyvendinimo paslaugos

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.4827	0.0000	837.9400	0.0010	True	True
1	1-3	422	902	0.2536	0.0000	83.4153	0.0010	True	True
2	1-4	422	218	0.2855	0.0000	44.0989	0.0010	True	True
3	1-5	422	323	0.2791	0.0000	46.5272	0.0010	True	True
4	1-6	422	567	0.0923	0.0298	3.0516	0.0186	True	True
5	2-3	1902	902	0.2328	0.0000	375.9538	0.0010	True	True
6	2-4	1902	218	0.2014	0.0000	153.6331	0.0010	True	True
7	2-5	1902	323	0.6591	0.0000	1193.2530	0.0010	True	True
8	2-6	1902	567	0.4581	0.0000	853.1779	0.0010	True	True
9	3-4	902	218	0.0366	0.9649	-0.3236	0.2500	False	False
10	3-5	902	323	0.4320	0.0000	224.1893	0.0010	True	True
11	3-6	902	567	0.2306	0.0000	88.6242	0.0010	True	True
12	4-5	218	323	0.4657	0.0000	110.3519	0.0010	True	True
13	4-6	218	567	0.2623	0.0000	42.8726	0.0010	True	True
14	5-6	323	567	0.2263	0.0000	40.1877	0.0010	True	True

KINTAMASIS: Draudimas ir finansinės paslaugos

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.4085	0.0000	220.4682	0.0010	True	True
1	1-3	422	902	0.1570	0.0000	19.4339	0.0010	True	True
2	1-4	422	218	0.4001	0.0000	69.2498	0.0010	True	True
3	1-5	422	323	0.2369	0.0000	28.3978	0.0010	True	True
4	1-6	422	567	0.0901	0.0363	1.7561	0.0613	True	False
5	2-3	1902	902	0.2877	0.0000	189.6423	0.0010	True	True
6	2-4	1902	218	0.0534	0.6112	0.3638	0.2364	False	False
7	2-5	1902	323	0.5331	0.0000	326.0389	0.0010	True	True
8	2-6	1902	567	0.4279	0.0000	287.5638	0.0010	True	True
9	3-4	902	218	0.2738	0.0000	38.4032	0.0010	True	True
10	3-5	902	323	0.3100	0.0000	87.8847	0.0010	True	True
11	3-6	902	567	0.1676	0.0000	32.2550	0.0010	True	True
12	4-5	218	323	0.5028	0.0000	114.7308	0.0010	True	True
13	4-6	218	567	0.4073	0.0000	78.1428	0.0010	True	True
14	5-6	323	567	0.1755	0.0000	18.2329	0.0010	True	True

KINTAMASIS: Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1–2	422	1902	0.2821	0.0000	125.6699	0.0010	True	True
1	1–3	422	902	0.1236	0.0003	15.0870	0.0010	True	True
2	1–4	422	218	0.2189	0.0000	24.4828	0.0010	True	True
3	1–5	422	323	0.0970	0.0587	2.8088	0.0231	False	True
4	1–6	422	567	0.0490	0.5854	-0.6472	0.2500	False	False
5	2–3	1902	902	0.1790	0.0000	74.2801	0.0010	True	True
6	2–4	1902	218	0.0901	0.0781	3.3899	0.0138	False	True
7	2–5	1902	323	0.3615	0.0000	142.5918	0.0010	True	True
8	2–6	1902	567	0.3116	0.0000	168.1670	0.0010	True	True
9	3–4	902	218	0.1158	0.0164	5.3892	0.0025	True	True
10	3–5	902	323	0.1923	0.0000	29.1337	0.0010	True	True
11	3–6	902	567	0.1391	0.0000	21.7050	0.0010	True	True
12	4–5	218	323	0.3005	0.0000	34.1338	0.0010	True	True
13	4–6	218	567	0.2503	0.0000	29.6409	0.0010	True	True
14	5–6	323	567	0.1201	0.0047	3.2079	0.0162	True	True

KINTAMASIS: Apranga ir avalynė

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1–2	422	1902	0.4237	0.0000	318.3779	0.0010	True	True
1	1–3	422	902	0.1964	0.0000	41.3667	0.0010	True	True
2	1–4	422	218	0.2992	0.0000	44.1620	0.0010	True	True
3	1–5	422	323	0.1197	0.0095	4.1149	0.0073	True	True
4	1–6	422	567	0.0705	0.1694	1.0664	0.1186	False	False
5	2–3	1902	902	0.2412	0.0000	178.4143	0.0010	True	True
6	2–4	1902	218	0.1374	0.0011	19.2505	0.0010	True	True
7	2–5	1902	323	0.4918	0.0000	348.1647	0.0010	True	True
8	2–6	1902	567	0.3942	0.0000	341.1456	0.0010	True	True
9	3–4	902	218	0.1208	0.0107	6.7910	0.0010	True	True
10	3–5	902	323	0.2790	0.0000	69.7259	0.0010	True	True
11	3–6	902	567	0.1769	0.0000	34.7609	0.0010	True	True
12	4–5	218	323	0.3624	0.0000	62.3933	0.0010	True	True
13	4–6	218	567	0.2697	0.0000	39.5105	0.0010	True	True
14	5–6	323	567	0.1459	0.0003	11.5415	0.0010	True	True

KINTAMASIS: Būstas, vanduo, elektra, dujos ir kitas kuras

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.2799	0.0000	93.1304	0.0010	True	True
1	1-3	422	902	0.1789	0.0000	25.4234	0.0010	True	True
2	1-4	422	218	0.3194	0.0000	44.6869	0.0010	True	True
3	1-5	422	323	0.0956	0.0647	2.0635	0.0459	False	True
4	1-6	422	567	0.0777	0.1009	2.1951	0.0406	False	True
5	2-3	1902	902	0.1130	0.0000	29.8045	0.0010	True	True
6	2-4	1902	218	0.0528	0.6250	0.4891	0.2088	False	False
7	2-5	1902	323	0.3316	0.0000	106.1457	0.0010	True	True
8	2-6	1902	567	0.2194	0.0000	79.0895	0.0010	True	True
9	3-4	902	218	0.1590	0.0002	11.2192	0.0010	True	True
10	3-5	902	323	0.2351	0.0000	37.2004	0.0010	True	True
11	3-6	902	567	0.1164	0.0001	13.6771	0.0010	True	True
12	4-5	218	323	0.3485	0.0000	52.0123	0.0010	True	True
13	4-6	218	567	0.2538	0.0000	31.3927	0.0010	True	True
14	5-6	323	567	0.1335	0.0012	7.3539	0.0010	True	True

KINTAMASIS: Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.2903	0.0000	106.4413	0.0010	True	True
1	1-3	422	902	0.1334	0.0001	11.1983	0.0010	True	True
2	1-4	422	218	0.3028	0.0000	40.0350	0.0010	True	True
3	1-5	422	323	0.1325	0.0029	4.4728	0.0054	True	True
4	1-6	422	567	0.0641	0.2583	0.7815	0.1565	False	False
5	2-3	1902	902	0.1830	0.0000	72.3345	0.0010	True	True
6	2-4	1902	218	0.0613	0.4345	-0.1107	0.2500	False	False
7	2-5	1902	323	0.3644	0.0000	128.1574	0.0010	True	True
8	2-6	1902	567	0.2559	0.0000	104.8422	0.0010	True	True
9	3-4	902	218	0.1843	0.0000	18.9995	0.0010	True	True
10	3-5	902	323	0.1997	0.0000	26.6481	0.0010	True	True
11	3-6	902	567	0.0983	0.0022	7.9998	0.0010	True	True
12	4-5	218	323	0.3582	0.0000	49.9049	0.0010	True	True
13	4-6	218	567	0.2655	0.0000	31.6705	0.0010	True	True
14	5-6	323	567	0.1272	0.0023	5.8223	0.0018	True	True

KINTAMASIS: Sveikata

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1–2	422	1902	0.1281	0.0000	17.1202	0.0010	True	True
1	1–3	422	902	0.0682	0.1302	0.2583	0.2500	False	False
2	1–4	422	218	0.0988	0.1111	2.6479	0.0268	False	True
3	1–5	422	323	0.1058	0.0304	4.7290	0.0043	True	True
4	1–6	422	567	0.0878	0.0445	6.0639	0.0015	True	True
5	2–3	1902	902	0.1493	0.0000	50.8907	0.0010	True	True
6	2–4	1902	218	0.2146	0.0000	27.9169	0.0010	True	True
7	2–5	1902	323	0.0647	0.1880	0.3788	0.2329	False	False
8	2–6	1902	567	0.0549	0.1369	2.3328	0.0358	False	True
9	3–4	902	218	0.0787	0.2139	0.9768	0.1293	False	False
10	3–5	902	323	0.1276	0.0008	12.3721	0.0010	True	True
11	3–6	902	567	0.1168	0.0001	18.6697	0.0010	True	True
12	4–5	218	323	0.2038	0.0000	13.9086	0.0010	True	True
13	4–6	218	567	0.1798	0.0001	15.9269	0.0010	True	True
14	5–6	323	567	0.0627	0.3733	-0.3481	0.2500	False	False

KINTAMASIS: Transportas

	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1–2	422	1902	0.4241	0.0000	310.0063	0.0010	True	True
1	1–3	422	902	0.1261	0.0002	13.6334	0.0010	True	True
2	1–4	422	218	0.3488	0.0000	68.2448	0.0010	True	True
3	1–5	422	323	0.0882	0.1074	2.5021	0.0306	False	True
4	1–6	422	567	0.0925	0.0294	4.0203	0.0079	True	True
5	2–3	1902	902	0.3102	0.0000	272.1348	0.0010	True	True
6	2–4	1902	218	0.1029	0.0294	4.6367	0.0047	True	True
7	2–5	1902	323	0.4612	0.0000	309.4811	0.0010	True	True
8	2–6	1902	567	0.3585	0.0000	273.6874	0.0010	True	True
9	3–4	902	218	0.2456	0.0000	38.3162	0.0010	True	True
10	3–5	902	323	0.1668	0.0000	26.5012	0.0010	True	True
11	3–6	902	567	0.0627	0.1227	2.9665	0.0201	False	True
12	4–5	218	323	0.3839	0.0000	78.1686	0.0010	True	True
13	4–6	218	567	0.2863	0.0000	48.8410	0.0010	True	True
14	5–6	323	567	0.1466	0.0002	11.7785	0.0010	True	True

KINTAMASIS: Informacija ir ryšiai

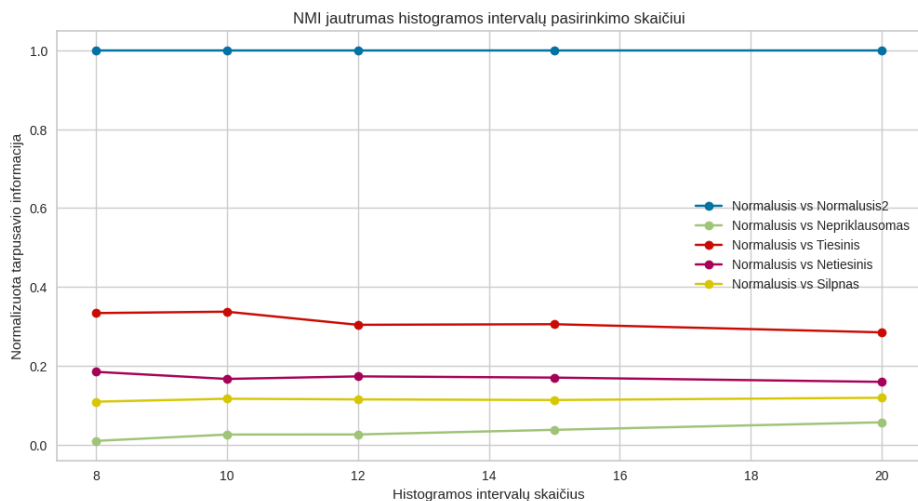
	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.6239	0.0000	445.4608	0.001	True	True
1	1-3	422	902	0.3162	0.0000	107.9220	0.001	True	True
2	1-4	422	218	0.4973	0.0000	114.8050	0.001	True	True
3	1-5	422	323	0.0736	0.2576	0.0495	0.250	False	False
4	1-6	422	567	0.2554	0.0000	51.0993	0.001	True	True
5	2-3	1902	902	0.3632	0.0000	272.9615	0.001	True	True
6	2-4	1902	218	0.1508	0.0002	13.2112	0.001	True	True
7	2-5	1902	323	0.6504	0.0000	396.0332	0.001	True	True
8	2-6	1902	567	0.4602	0.0000	308.4325	0.001	True	True
9	3-4	902	218	0.2261	0.0000	31.9714	0.001	True	True
10	3-5	902	323	0.3589	0.0000	111.2426	0.001	True	True
11	3-6	902	567	0.1170	0.0001	12.9608	0.001	True	True
12	4-5	218	323	0.5265	0.0000	112.5137	0.001	True	True
13	4-6	218	567	0.3233	0.0000	57.5090	0.001	True	True
14	5-6	323	567	0.3022	0.0000	58.9261	0.001	True	True

KINTAMASIS: Poilsis, sportas ir kultūra

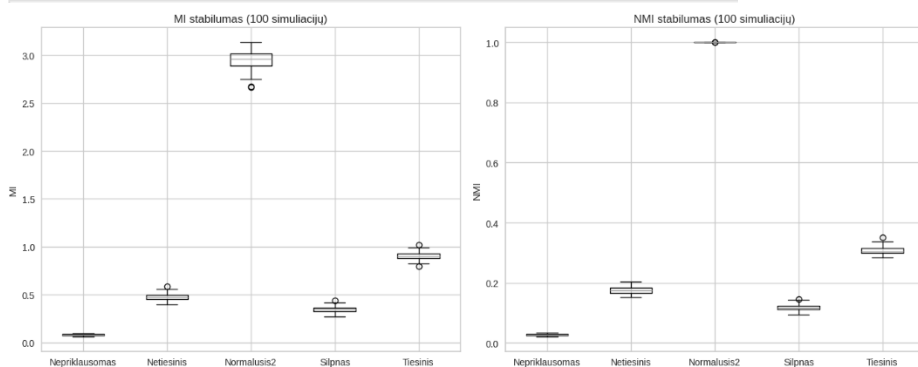
	Klasterių pora	n ₁	n ₂	KS_stat	KS_p	AD_stat	AD_p	KS_sig	AD_sig
0	1-2	422	1902	0.2177	0.0000	73.6685	0.001	True	True
1	1-3	422	902	0.1036	0.0038	10.2352	0.001	True	True
2	1-4	422	218	0.2075	0.0000	22.0357	0.001	True	True
3	1-5	422	323	0.2200	0.0000	31.3317	0.001	True	True
4	1-6	422	567	0.0383	0.8517	-0.7837	0.250	False	False
5	2-3	1902	902	0.1252	0.0000	37.7218	0.001	True	True
6	2-4	1902	218	0.0708	0.2666	-0.2000	0.250	False	False
7	2-5	1902	323	0.4228	0.0000	223.1149	0.001	True	True
8	2-6	1902	567	0.2186	0.0000	88.7356	0.001	True	True
9	3-4	902	218	0.1233	0.0086	7.0411	0.001	True	True
10	3-5	902	323	0.3100	0.0000	90.4759	0.001	True	True
11	3-6	902	567	0.1082	0.0005	12.4354	0.001	True	True
12	4-5	218	323	0.3919	0.0000	74.6067	0.001	True	True
13	4-6	218	567	0.2034	0.0000	22.4866	0.001	True	True
14	5-6	323	567	0.2176	0.0000	33.8771	0.001	True	True

21 priedas. Duomenų validavimo rezultatai

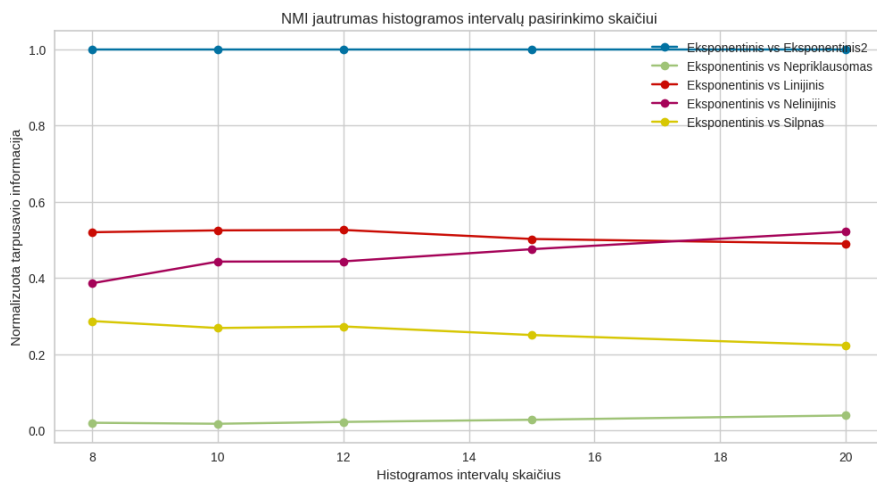
Simetriškų (normalių) duomenų validavimas



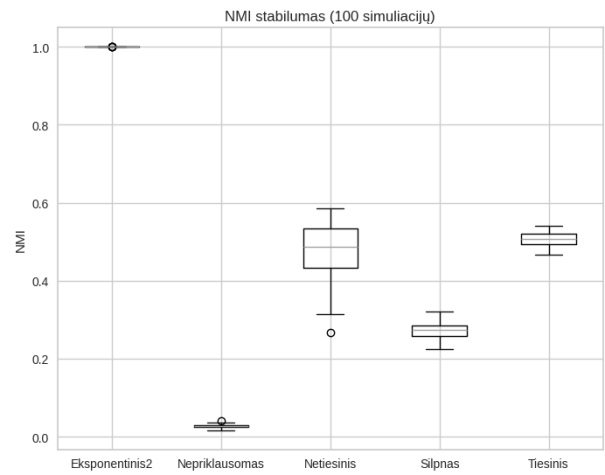
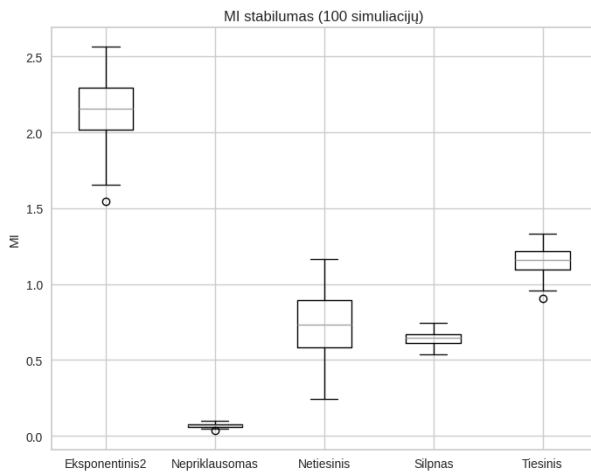
	case	MI_mean	MI_std	NMI_mean	NMI_std
0	Nepriklausomas	0.079986	0.008934	0.027185	2.895909e-03
1	Netiesinis	0.472871	0.034570	0.175501	1.229025e-02
2	Normalusis2	2.945868	0.094304	1.000000	5.549119e-15
3	Silpnas	0.347664	0.029166	0.118263	9.963553e-03
4	Tiesinis	0.906885	0.037074	0.307559	1.225487e-02



Asimetriškų (eksponentinių) duomenų validavimas



	case	MI_mean	MI_std	NMI_mean	NMI_std
0	Eksponentinis2	2.146977	0.216214	1.000000	2.537089e-14
1	Nepriklausomas	0.068289	0.011610	0.026797	4.028695e-03
2	Netiesinis	0.731200	0.210988	0.476474	7.001581e-02
3	Silpnas	0.641573	0.043343	0.272498	1.819373e-02
4	Tiesinis	1.151549	0.085312	0.506751	1.775995e-02



22 priedas. Entropijos tikrinimas vartojimo išlaidoms

00. Visos namų ūkio vartojimo išlaidos (mėnesinės)

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	2.7598	0.7698	1089.6612	573.7310	184.76	3891.77
2	1902	1.9490	0.5437	531.4604	357.9262	64.00	3744.47
3	902	1.8707	0.5218	824.5721	544.1419	106.41	6031.28
4	218	1.8105	0.5050	659.4402	591.9150	56.23	5416.45
5	323	2.7239	0.7598	1458.7657	850.5423	229.22	5474.17
6	567	1.8261	0.5094	1069.4092	755.4171	150.05	7984.83

01. Maistas ir nealkoholiniai gėrimai

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	2.6367	0.7355	337.7381	205.2818	0.00	1392.11
2	1902	1.5563	0.4341	204.9074	144.2178	0.00	2053.07
3	902	2.3823	0.6645	271.4913	170.1442	0.00	1454.35
4	218	2.4863	0.6935	236.4961	197.6665	9.56	1272.85
5	323	2.8517	0.7955	385.9016	205.6409	6.22	1326.17
6	567	2.1192	0.5911	287.1000	193.1051	0.00	1829.28

02. Alkoholiniai gėrimai, tabakas ir narkotikai

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	1.0542	0.2941	21.8303	51.2260	0.0	468.44
2	1902	0.5909	0.1648	9.9964	33.4772	0.0	476.33
3	902	1.0022	0.2796	32.0795	72.9849	0.0	744.46
4	218	1.3127	0.3662	24.8218	57.0806	0.0	395.73
5	323	1.1771	0.3283	28.0605	66.1620	0.0	534.46
6	567	0.9659	0.2694	27.9872	66.8331	0.0	643.18

03. Apranga ir avalynė

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	1.4012	0.3909	92.4897	168.9827	0.0	1186.25
2	1902	0.6602	0.1842	20.1439	61.8671	0.0	736.41
3	902	0.9953	0.2776	46.5587	99.9999	0.0	940.14
4	218	0.7922	0.2210	39.5732	102.9664	0.0	986.33
5	323	1.6230	0.4527	101.9338	149.5797	0.0	1139.81
6	567	1.5043	0.4196	66.5538	106.3687	0.0	805.45

04. Būstas, vanduo, elektra, dujos ir kitas kuras

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	2.1357	0.5957	155.9149	104.1440	32.00	855.33
2	1902	0.2474	0.0690	110.2128	106.8731	0.00	3080.47
3	902	1.1194	0.3122	132.9195	113.8251	12.12	1686.09
4	218	1.8370	0.5124	114.3205	113.4229	0.00	897.66
5	323	1.7713	0.4941	169.8242	124.6613	22.09	1283.99
6	567	0.4448	0.1241	159.7668	186.5832	15.00	3339.30

05. Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	1.3716	0.3826	59.3116	94.4016	0.0	773.44
2	1902	1.1578	0.3230	24.3107	45.3966	0.0	445.30
3	902	0.1620	0.0452	46.5708	127.9923	0.0	2653.59
4	218	1.1470	0.3200	30.7646	60.3930	0.0	482.65
5	323	1.7058	0.4758	76.4045	110.3003	0.0	770.80
6	567	0.2042	0.0570	65.2136	175.2095	0.0	3691.87

06. Sveikata

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	0.7184	0.2004	58.4790	137.9205	0.0	1540.40
2	1902	0.8079	0.2254	65.1988	110.9559	0.0	1666.65
3	902	0.4466	0.1246	48.3950	129.3898	0.0	1999.98
4	218	0.8720	0.2432	39.6867	96.4269	0.0	999.99
5	323	0.8054	0.2247	65.1866	133.5492	0.0	1666.65
6	567	0.8031	0.2240	78.9928	168.1775	0.0	1999.98

07. Transportas

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	1.1932	0.3328	129.6798	205.9522	0.0	2158.68
2	1902	0.8974	0.2503	31.8902	81.8223	0.0	891.63
3	902	0.3495	0.0975	105.5580	228.9911	0.0	4759.87
4	218	0.2239	0.0624	62.8144	320.4458	0.0	4566.49
5	323	1.1271	0.3144	218.7347	423.4038	0.0	3305.30
6	567	0.4670	0.1303	149.7202	377.4655	0.0	5291.96

08. Informacija ir ryšiai

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	2.2412	0.6252	75.3224	54.6783	0.0	438.99
2	1902	1.6299	0.4546	24.7991	21.9193	0.0	255.00
3	902	2.2575	0.6297	44.6494	31.4967	0.0	266.99
4	218	1.7616	0.4914	33.8027	33.6496	0.0	321.66
5	323	2.0224	0.5641	80.9546	64.9042	0.0	558.32
6	567	1.7827	0.4973	53.8262	46.2892	0.0	446.65

09. Poilsis, sportas ir kultūra

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	1.6424	0.4581	35.8492	61.1716	0.0	407.19
2	1902	0.4981	0.1389	13.9312	35.9881	0.0	662.60
3	902	1.2067	0.3366	24.9833	50.2061	0.0	421.55
4	218	0.7592	0.2118	20.6160	56.2261	0.0	521.84
5	323	1.3839	0.3860	77.0420	113.7578	0.0	1121.28
6	567	1.1813	0.3295	42.9938	86.5898	0.0	700.21

10. Švietimo paslaugos

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	0.3425	0.0955	15.9579	65.5093	0.0	1035.00
2	1902	0.0195	0.0054	0.1047	3.0759	0.0	116.66
3	902	0.1181	0.0329	0.7836	8.3922	0.0	166.66
4	218	0.4101	0.1144	2.5351	13.4868	0.0	150.00
5	323	0.6869	0.1916	34.0743	102.6564	0.0	916.63
6	567	0.2129	0.0594	2.3142	18.9186	0.0	291.66

11. Restoranai ir apgyvendinimo paslaugos

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	1.5737	0.4390	47.7857	74.1635	0.0	624.05
2	1902	0.0531	0.0148	4.2811	50.3396	0.0	1568.69
3	902	0.7967	0.2222	25.5612	66.1963	0.0	790.83
4	218	1.1498	0.3207	22.1560	52.2480	0.0	375.89
5	323	1.6159	0.4507	131.6650	180.7611	0.0	1532.88
6	567	1.4147	0.3946	68.4751	122.6895	0.0	1030.90

12. Draudimas ir finansinės paslaugos

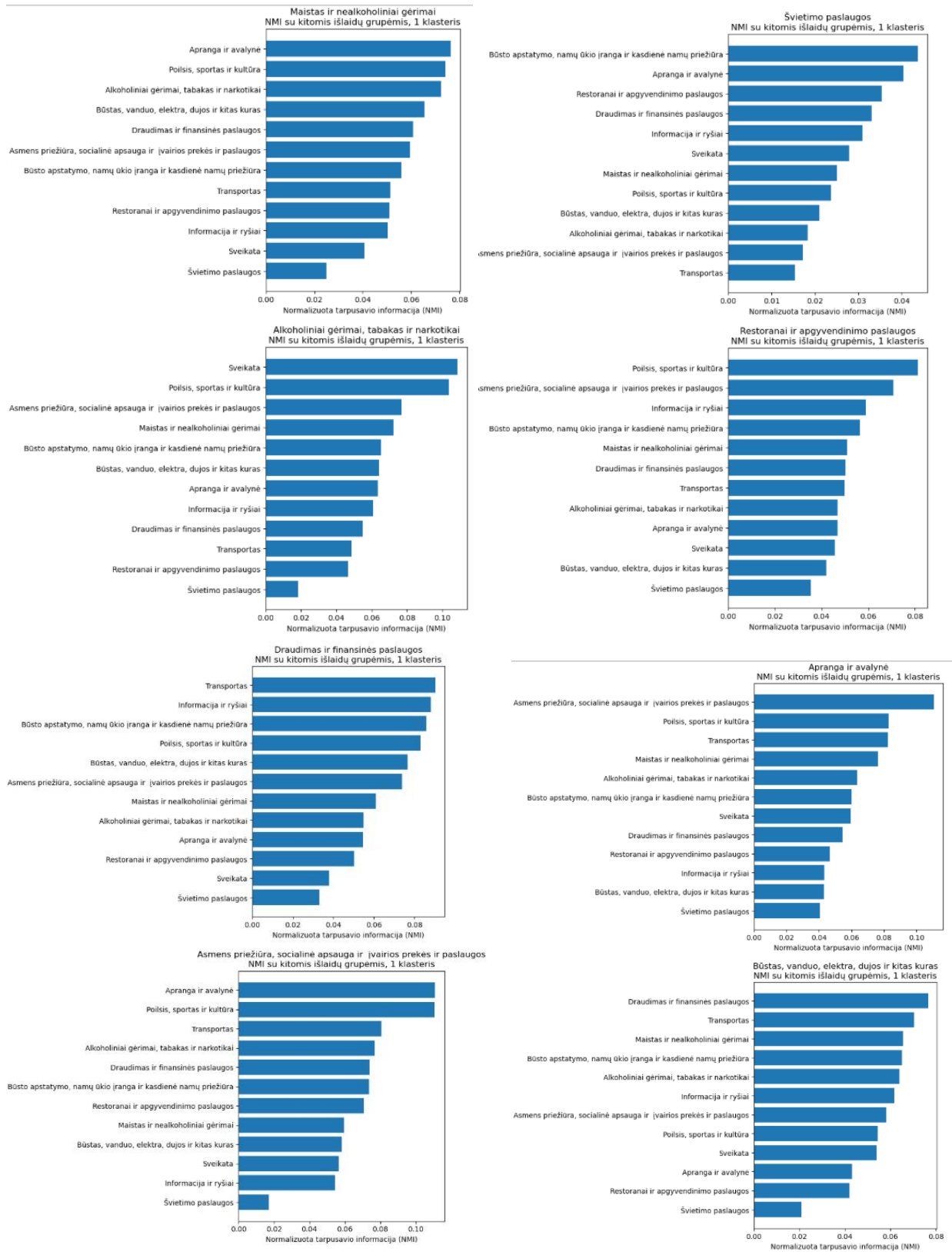
Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	2.0293	0.5661	16.9443	18.7133	0.0	136.50
2	1902	1.2087	0.3371	5.9433	9.6464	0.0	101.86
3	902	1.7456	0.4869	12.7166	15.3971	0.0	130.36
4	218	1.4484	0.4040	7.3436	12.6062	0.0	99.13
5	323	2.1235	0.5923	29.9564	34.0687	0.0	213.45
6	567	1.8760	0.5233	20.3630	24.0774	0.0	184.16

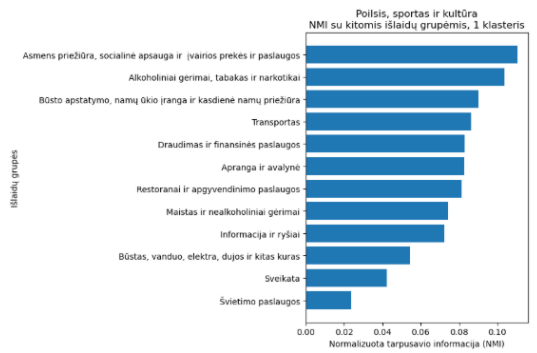
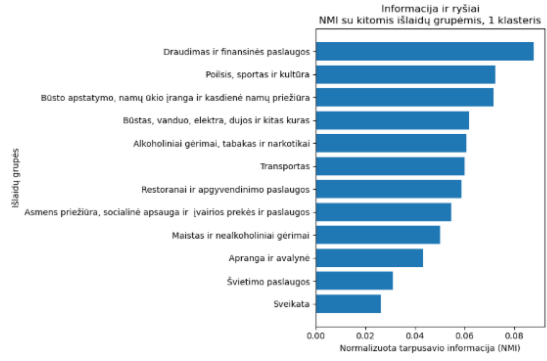
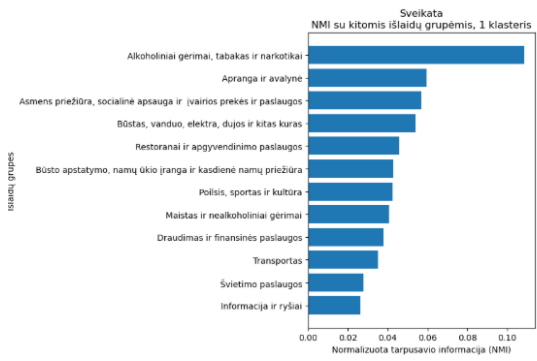
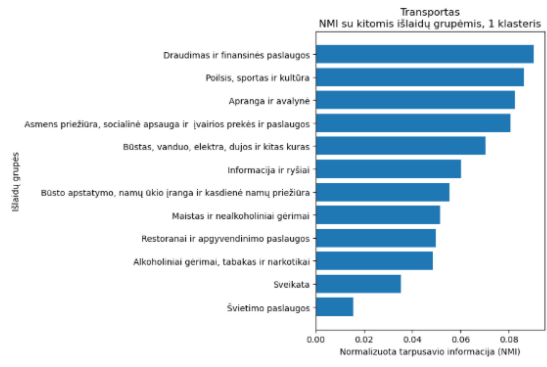
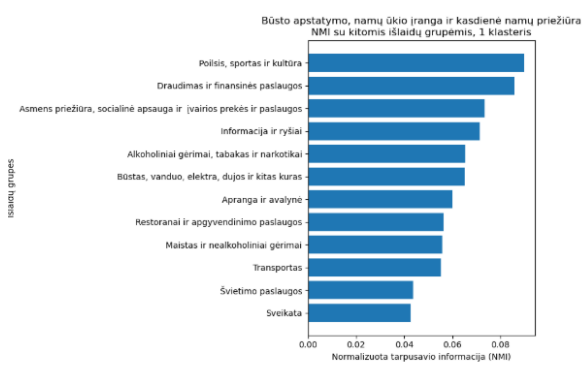
13. Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos

Klasteris	n	H(X)	NH(X)	vidurkis	std	min	max
1	422	1.7969	0.5012	42.3592	63.0274	0.0	418.36
2	1902	0.7751	0.2162	15.7412	35.8379	0.0	424.15
3	902	0.9008	0.2513	32.3058	62.8726	0.0	778.57
4	218	1.3628	0.3802	24.5103	50.9764	0.0	307.42
5	323	1.7881	0.4988	59.0284	87.6628	0.0	592.15
6	567	1.1505	0.3209	46.1033	80.4198	0.0	758.30

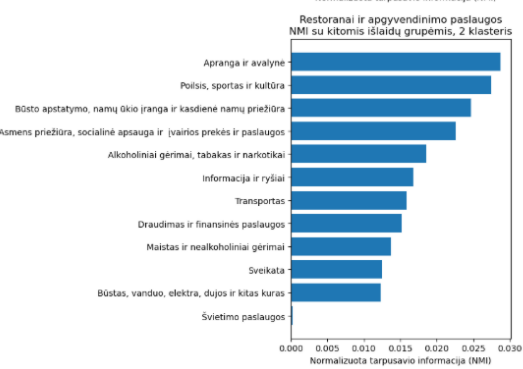
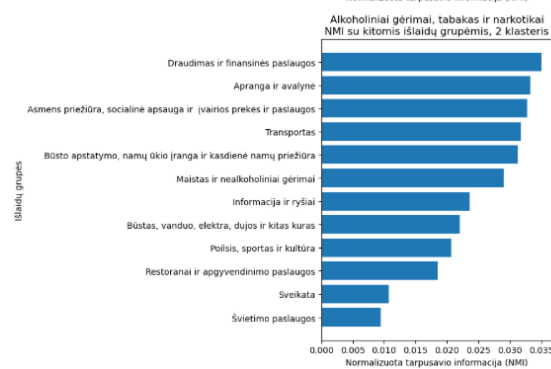
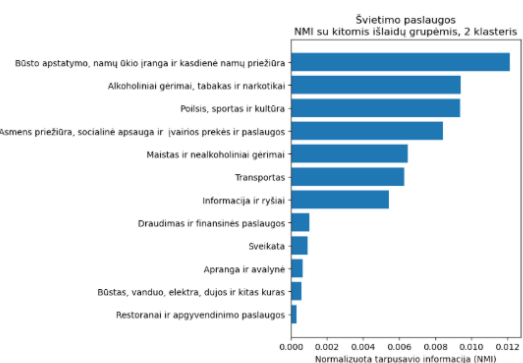
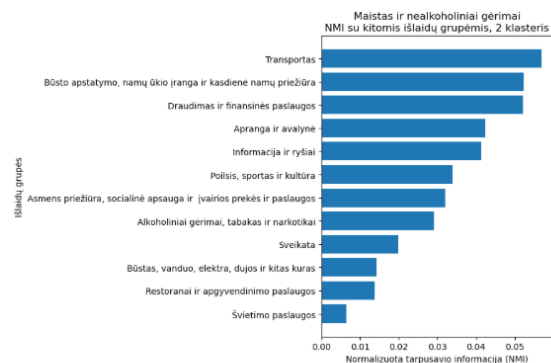
23 priedas. Vartojimo išlaidų ryšiai remiantis normalizuota tarpusavio informacija

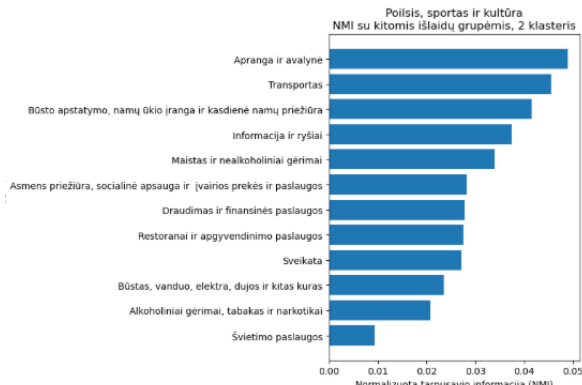
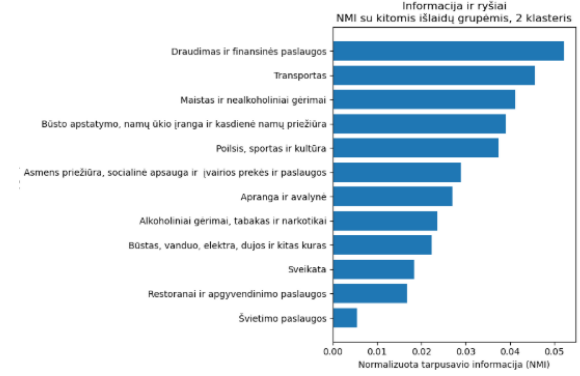
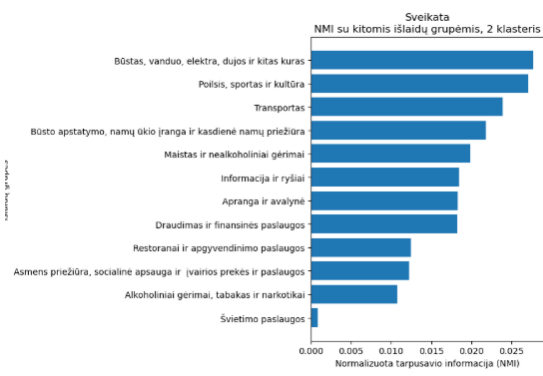
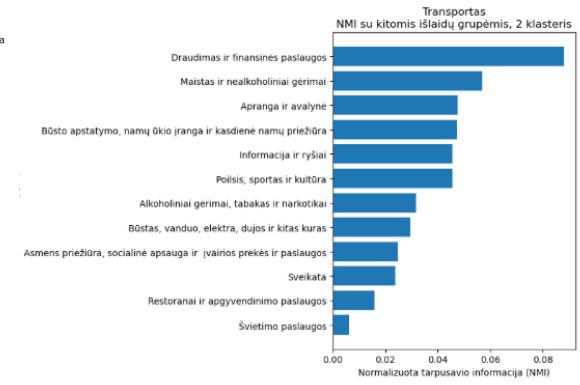
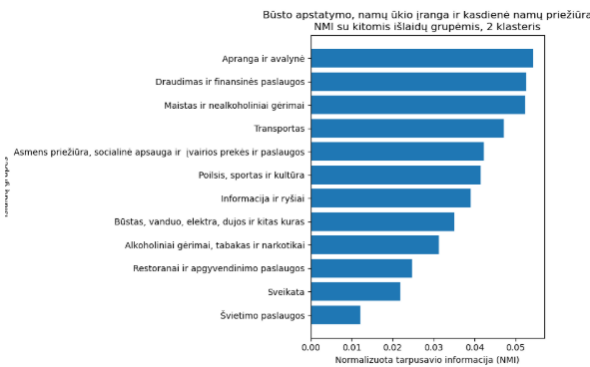
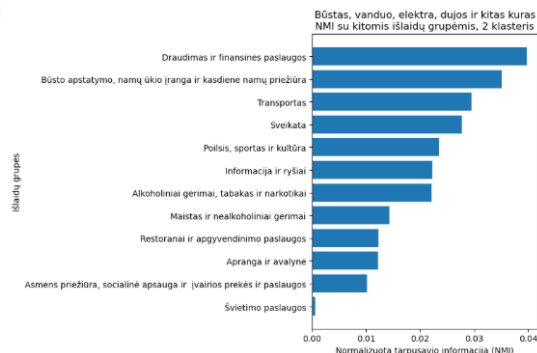
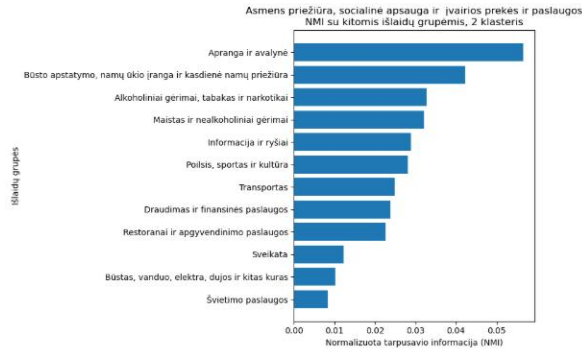
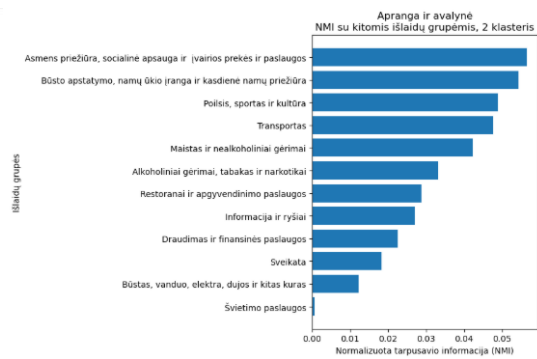
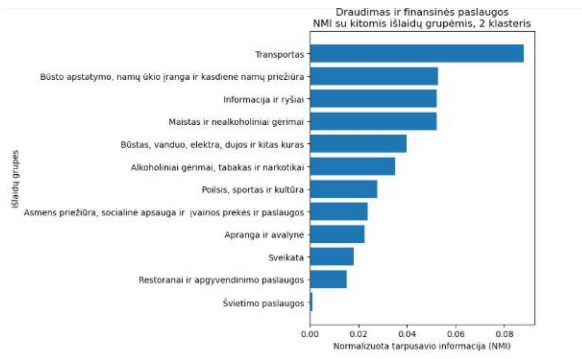
Jaunos mažesnių pajamų šeimos su vaikais (1 klasteris)



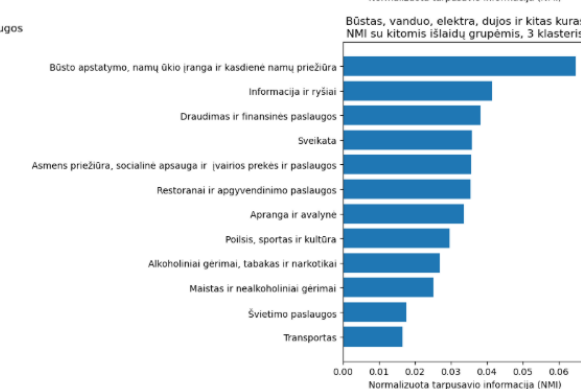
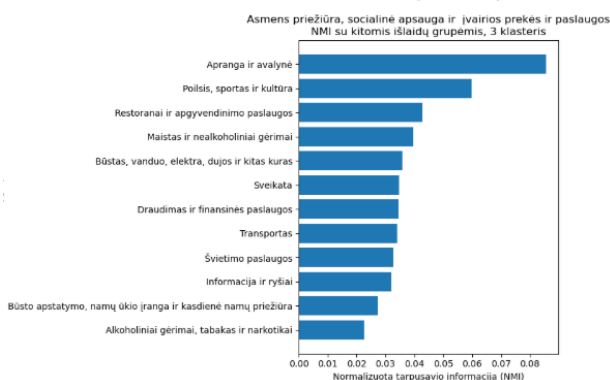
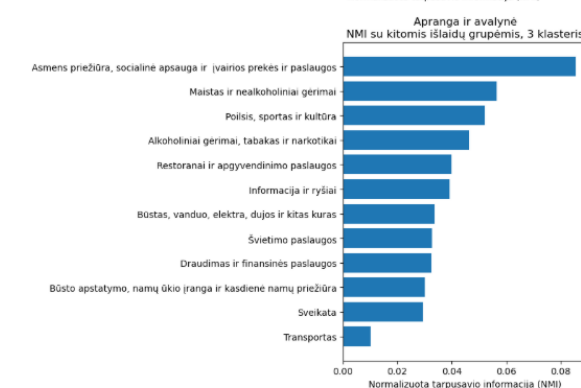
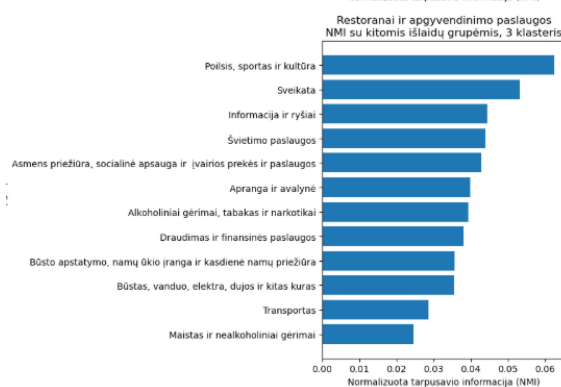
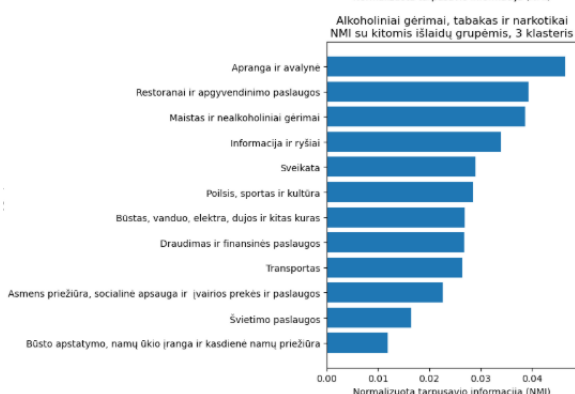
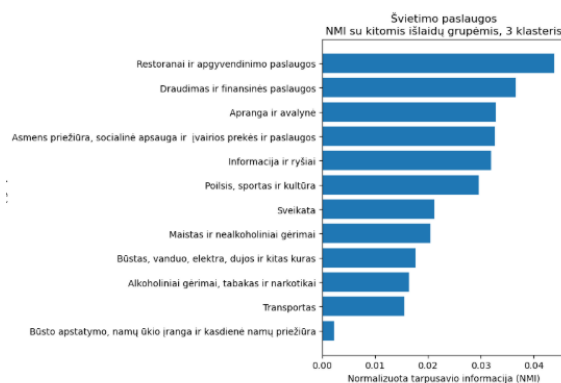
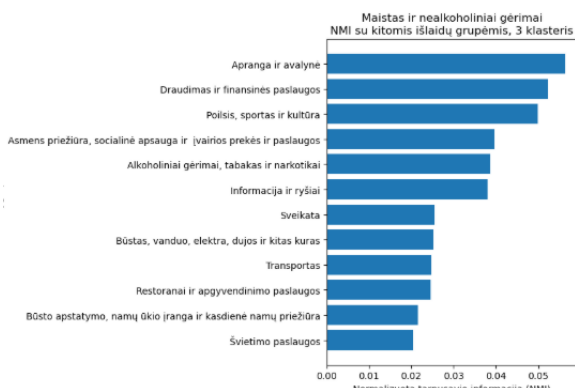


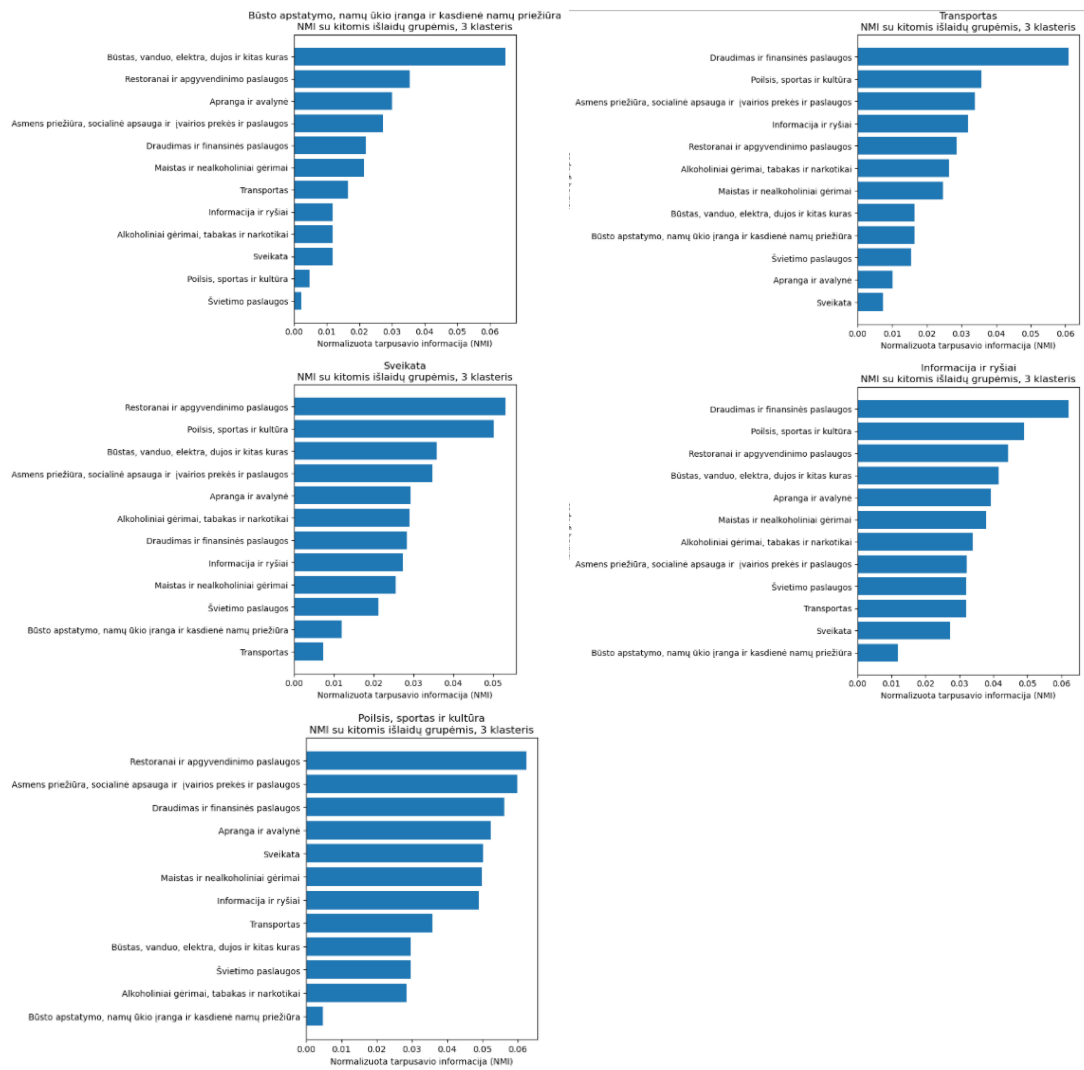
Pensinio amžiaus namų ūkiai (2 klasteris)



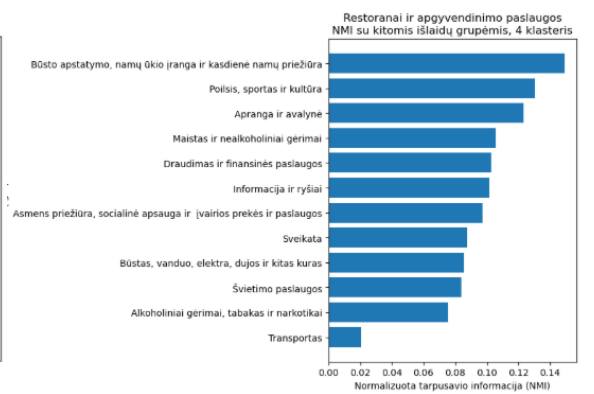
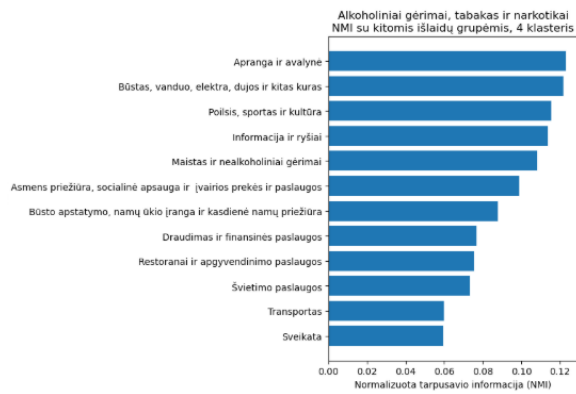
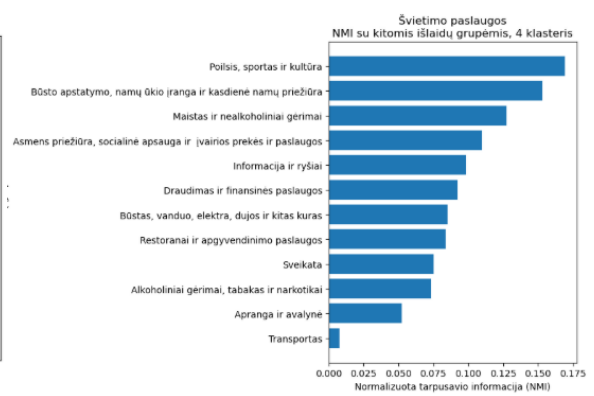
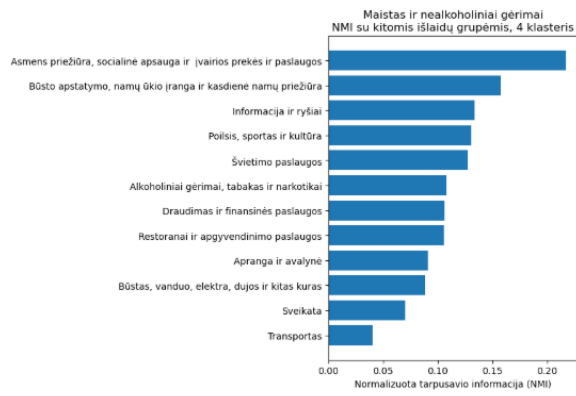


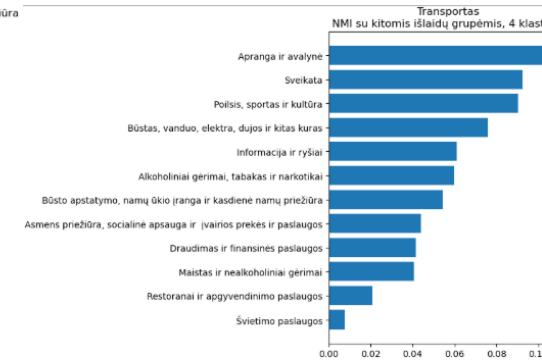
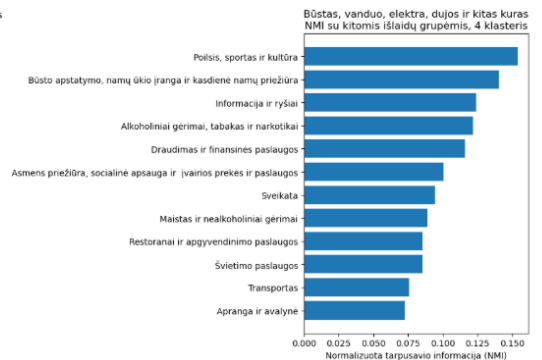
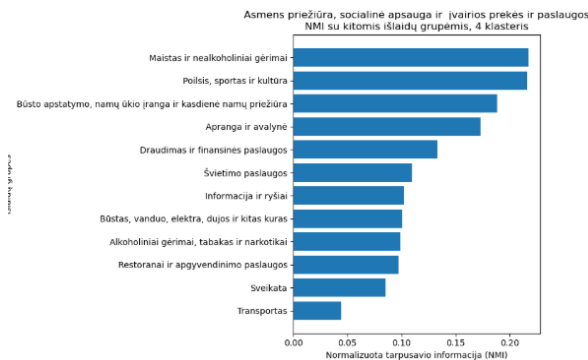
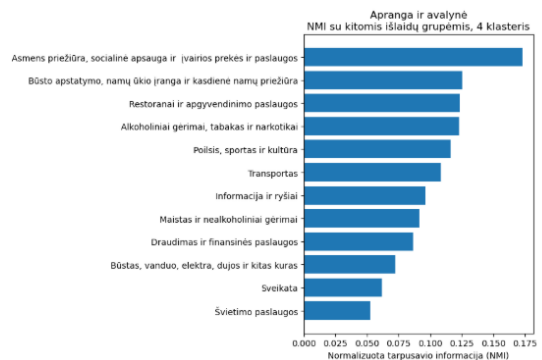
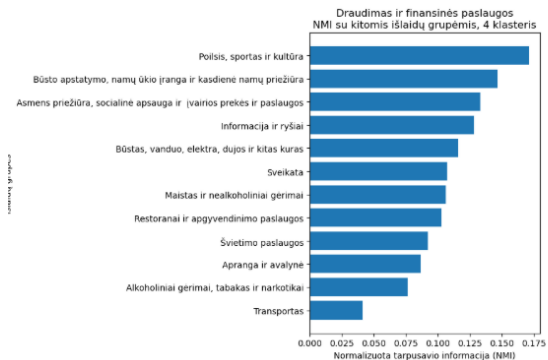
Vidutinių–aukštesnių pajamų dirbantys namų ūkiai (3 klasteris)



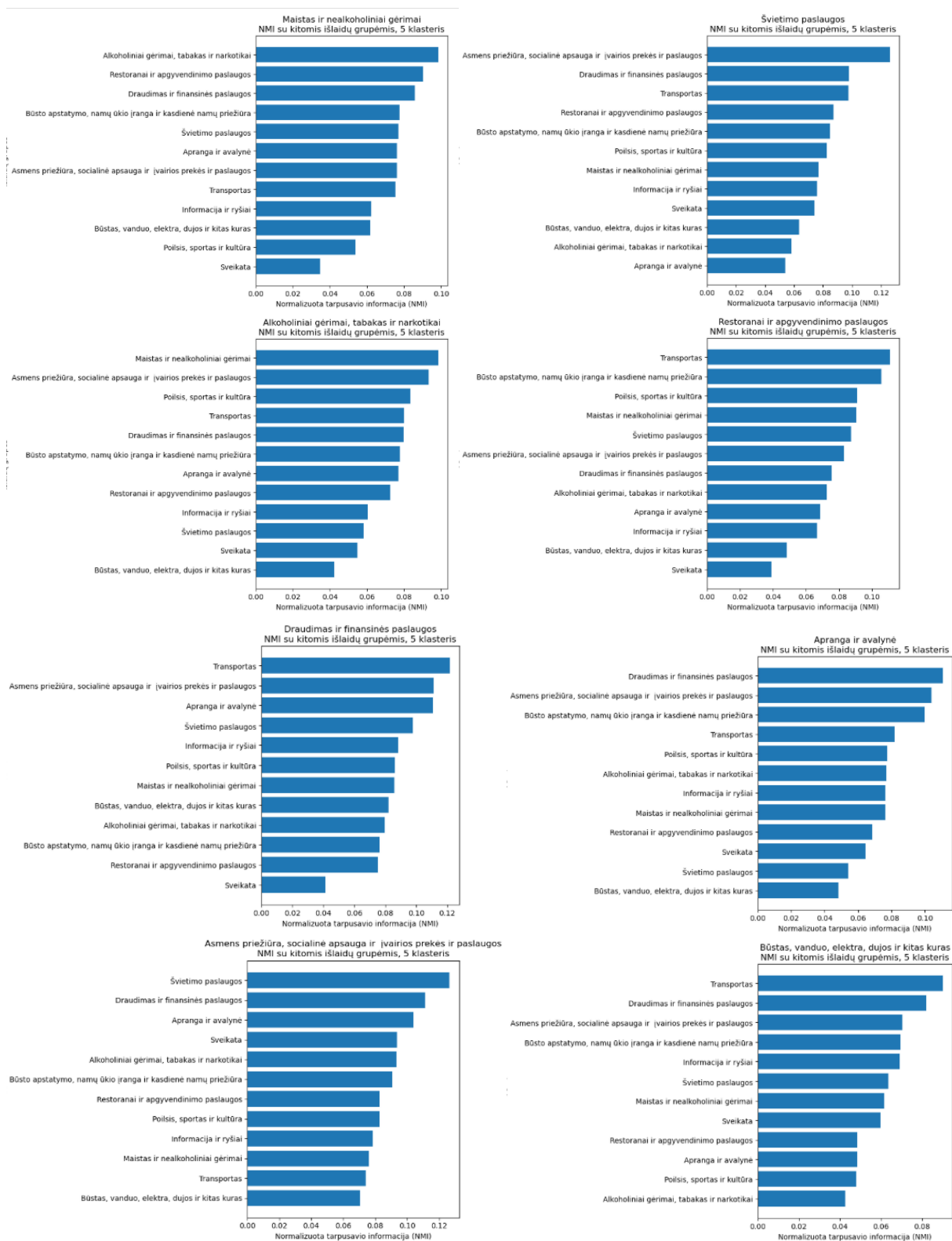


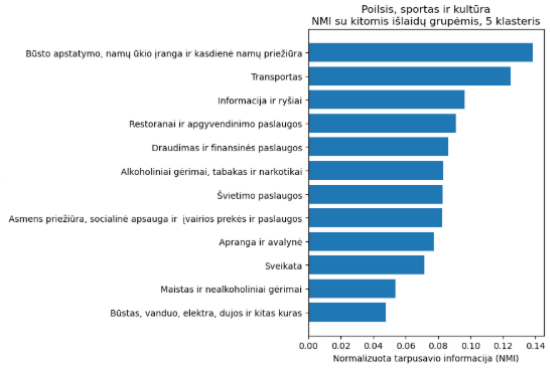
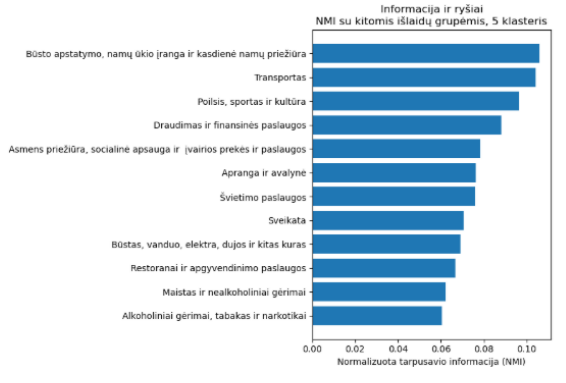
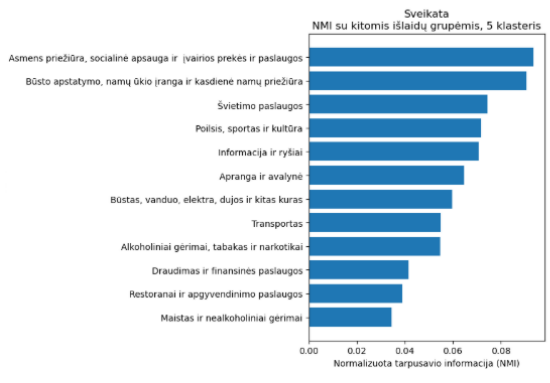
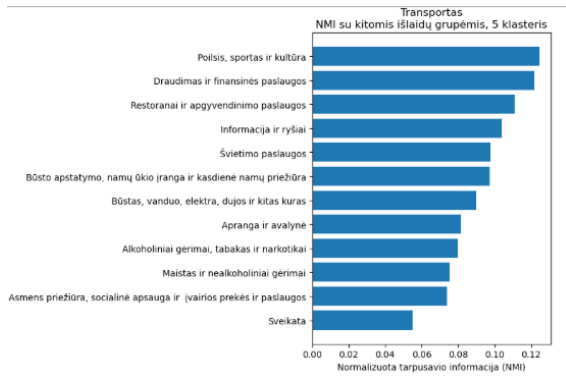
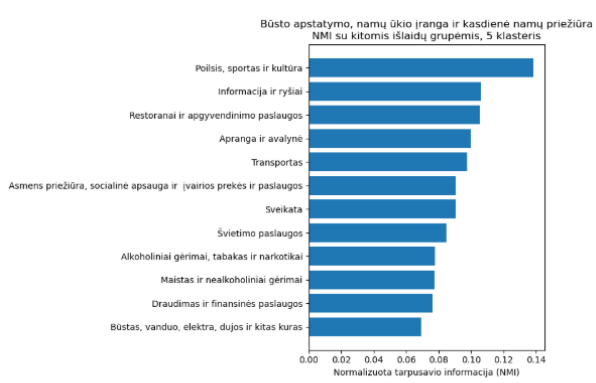
Socialiai pažeidžiami namų ūkiai (4 klasteris)





Jaunos ekonomiškai aktyvios šeimos (5 klasteris)





Aukštos kvalifikacijos ir aukštų pajamų namų ūkiai (6 klasteris)

