

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

David Jaanson

**Eesti tutvumiskuulutuste analüüs Nelli Teataja
näitel**

Bakalaureusetöö (9 EAP)

Juhendaja: Uku Kangur, MSc

Tartu 2025

Analysis of Estonian personal ads using the example of Nelli Teataja

Abstract:

This bachelor's thesis explores personal advertisements published in Estonian newspapers through a cultural and social analytical lens. The aim of the study is to construct a structured dataset based on dating ads from Nelli Teataja and to examine demographic, character-based, and value-driven patterns within them. Data collection and processing were conducted using natural language processing, machine learning techniques, and statistical analysis. The research focuses on identifying the gender, age, location, partner preferences, and personality traits mentioned by ad authors. Frequently highlighted traits include free, kind, slim, calm, and athletic. The findings indicate that most advertisers are older individuals whose preferences reflect distinct social and cultural values. The analysis also reveals regional and gender-based differences in dating behavior.

Keywords:

personal ads, cultural analysis, natural language processing, social behavior, Estonian press

CERCS:

P170 Computer science, numerical analysis, systems, control

Eesti tutvumiskuulutuste analüüs Nelli Teataja näitel

Lühikokkuvõte:

Käesolev bakalaureusetöö uurib Eesti ajalehtedes avaldatud tutvumiskuulutusi kultuurilise ja sotsiaalse analüüsi vaatenurgast. Töö eesmärk on luua struktureeritud andmestik Nelli Teatajas ilmunud tutvumiskuulutustest ning analüüsida, milliseid demograafilisi, iseloomulikke ja väärtuspõhiseid mustreid kuulutustes esineb. Andmete kogumiseks ja töötlemiseks kasutati loomuliku keele töötlust, masinõppe vahendeid ning statistilisi meetodeid. Eraldi vaadeldi kuulutuste postitajate sugu, vanust, elukohta, partnerieelistusi ja esile toodud iseloomuomadusi. Partneri kirjeldustes sagedamini esile omadused nagu vaba, heatahtlik, sale, rahulik ja sportlik. Tulemused näitavad, et kuulutuste avaldajad on enamasti vanemaealised inimesed. Töös tuuakse esile ka piirkondlikud ning soopõhised erinevused partneriotsingu käitumises.

Võtmesõnad:

Tutvumiskuulutused, kultuurianalüüs, loomuliku keele töötlus, sotsiaalne käitumine, Eesti ajakirjandus

CERCS:

P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Sisukord

Sissejuhatus	5
1. Teoreetiline ülevaade	6
2. Metodoloogia	8
2.1. Andmestik ja andmete eeltöötlus.....	8
2.2. Andmeomaduste kaeve.....	10
2.2.1. ID ja Sugu omaduse määramine.....	11
2.2.2. “Otsib sugu” omaduse määramine.....	11
2.2.3. Inimese vanuse määramine.....	12
2.2.4 Asukoha määramine.....	12
2.3. Iseloomuomaduste ekstraktimine ja korrastamine.....	13
2.4. Iseloomuomaduste klastrite loomine.....	13
3. Tulemused	15
3.1 Ühe muutuja analüüs.....	15
3.2 Kahe muutuja analüüs.....	21
4. Tulemuste piiravad faktorid	37
Kokkuvõte	38
Viited	39
Litsents	42

Sissejuhatus

Inimestevaheliste suhete loomine ja sobiva kaaslase leidmine on olnud läbi aegade üks olulisemaid sotsiaalseid vajadusi. Ajaloos on selleks kasutatud mitmesuguseid meediume ja vahendeid – alates suulistest soovitustest kuni tänapäevaste digitaalsete kohtingurakendusteni. Iga ajastu suhtlemisvahendid peegeldavad valitsevaid sotsiaalseid norme ja ootusi, mõjutades seda, kuidas inimesed oma soove väljendavad ning milliseid omadusi nad potentsiaalses partneris otsivad.

Üks märkimisväärne viis tutvuste loomiseks on olnud ajalehtede tutvumiskuulutused. Need kuulutused on pakkunud inimestele võimalust oma ootusi ja soove avalikult väljendada ning sobivat kaaslast leida. Kuid tutvumiskuulutused ei ole pelgalt praktiline vahend partneri leidmiseks – need peegeldavad ka ajastu sotsiaalseid norme, väärtushinnanguid ja ootusi.

Tutvumiskuulutuste analüüs pakub ainulaadset võimalust uurida inimeste sotsiaalseid ja kultuurilisi eelistusi läbi aja. Need kuulutused annavad tunnistust sellest, millised omadused on olnud soovitatavad, millised väärtushinnangud on pidanud paika ja kuidas need on ajas muutunud. Erinevates ajalehtedes avaldatud kuulutused kajastavad mitmekesiselt pilti ajastu ideaalidest ning pakuvad rikkalikku uurimismaterjali sotsiaalse ja kultuurilise ajaloo mõtestamiseks.

Käesolev uurimistöö peamine eesmärk on viia läbi põhjalik kultuurianalüüs Nelli Teataja tutvumiskuulutuste põhjal. Analüüsi kaks põhiküsimust on:

1. Kes on Nelli Teataja tutvumiskuulutuse avaldaja ning millised demograafilised, sotsiaalsed ja kultuurilised tegurid teda defineerivad?
2. Milliseid omadusi ja väärtusi otsitakse potentsiaalses partneris ning kuidas need kajastavad ajastu sotsiaalseid norme ja ideaale?

Töö koosneb viiest peamisest peatükist. Esimene peatükk annab ülevaate töö teemast, eesmärkidest ning uurimisküsimustest. Teises peatükis esitatakse teoreetiline raamistik ning käsitletakse varasemaid uurimusi, mis on seotud tutvumiskuulutuste ja online-tutvumisega. Kolmas peatükk keskendub metodoloogiale, sealhulgas andmestiku koostamisele, andmete eeltöötlemisele ning iseloomuomaduste määramisele. Neljandas peatükis esitatakse analüüsi tulemused, kus vaadeldakse nii üksik- kui ka kahekordseid muutujaid. Viiendas peatükis arutletakse tulemuste piirangute üle ning kokkuvõttepeatükis võetakse töö olulisemad järeldused kokku ja tehakse ettepanekuid edasiseks uurimiseks.

1. Teoreetiline ülevaade

Käesolev teoreetiline ülevaade käsitleb varasemaid uuringuid tutvumiskuulutuste ja *online*-tutvumise valdkonnas, tuues esile, kuidas inimeste eelistused ja käitumismustrid on aja jooksul muutunud, samuti, kuidas need muutused kajastavad ühiskondlikke ja kultuurilisi suundi.

Tallinna Ülikooli ELU projekti raames 2021. aastal uuriti, milliseid tutvumiskuulutusi avaldati Eesti ajakirjanduses 1930ndatel aastatel [1]. Projektis analüüsiti selliseid väljaandeid nagu Postimees, Film ja Elu, Abieluleht ja Päevaleht. Tulemused näitasid, et toonased kuulutused olid sageli humoorikad ja otsekohesed. Näiteks: "*Kes teist julgeb katsuda taltsutada isekat ja seiklushimulist noormeest, kes ei mõista muud, kui vaid lollusi teha?*".

Samuti ilmses, et majanduslik kindlustatus ja praktilised kaalutlused mängisid tähtsat rolli, mis väljendasid kuulutuste otsekohesuse:

"Lgp. daamid! Kes teist tahaks vaesele noormehele jõuluvanaks olla ja laenata 10-20 kr. Kui meeldin, võib kohe järgneda abielu."

Projekti tulemused näitasid ka, kuidas huumor ja eneseironia aitasid kuulutustel eristuda ja luua isiklikumat sidet lugejaga.

Samuti on uuritud, kuidas inimese põhiolemuse ja isiksuseomaduste tasandid mõjutavad tehnoloogiate kasutamist. Artikkel "Who uses mobile apps to meet strangers: The roles of core traits and surface characteristics" käsitleb individuaalseid tunnuseid ja nende mõju WeChati tutvumisfunktsioonide kasutamisele [2]. Uuring näitas, et isiksuse põhiolemuse ja pinnaliste tunnuste vahel on tugevad seosed. Ekstraversus, mida seostatakse soojuse, seltsivuse ja enesekindlusega, suurendab tõenäosust, et inimene kasutab sotsiaalse avastamise funktsioone. Samuti on leitud, et inimesed, kellele meeldib elus rohkem riske võtta, kasutavad tõenäolisemalt tutvumisrakendusi seksuaalsete kogemuste otsimiseks. Lisaks on avatus uutele kogemustele, mis viitab fantaasiarikkusele ja soovile katsetada uusi asju, positiivselt seotud sotsiaalsete avastamise funktsioonide kasutamisega.

See uurimus rõhutab, et praktiliste rakenduste arendamisel ja reklaamkampaniates on oluline arvestada kasutajate isiksuseomadustega.

Bakalaureusetöös "Rahvusäärmuslus Rate.ee profiilides", uuritakse, kuidas kajastatakse paremäärmuslikku meelsust Rate.ee suhtlusportaali kasutajaprofiilides [3]. Uuringus

analüüsi 174 erinevat profiili, mille huvirubriikides oli vähemalt üks paremäärmuslik sümbol või akronüüm. Uuringu tulemused viitavad sellele, et paremäärmuslikud ideoloogiad on oma kohalolekuga teatud internetikeskkondades olemas ja võivad mõjutada ka suhtlusportaalide kasutajate profiilide sisu.

Ameerika kontekstis on isikuomaduste ja partnerieelistuste muutumist ajas uurinud Larry M. Lance oma 1998. aasta artiklis “Gender Differences in Heterosexual Dating: A Content Analysis of Personal Ads” [4]. Selles uuringus analüüsi 1 433 heteroseksuaalset tutvumiskuulutust, et mõista, milliseid omadusi otsivad mehed ja naised potentsiaalsetes partnerites ning kuidas need ootused on muutunud võrreldes varasemate uuringutega. Analüüs näitas, et sooline erinevus näiteks vanuse või sotsiaalse staatuse eelistustes on muutumas – näiteks muutub mittedisheemiline üha olulisemaks mõlemast soost kuulutajate jaoks. See viitab, et sotsiaalsed ja kultuurilised muutused kajastuvad selgelt ka partneriotsingus ja isiklikus enesetutvustuses.

Online-tutvuse valdkonda käsitleb Günter J. Hitsch'i, Ali Hortacsu ja Dan Ariely 2005. aasta uurimus “What Makes You Click: An Empirical Analysis of Online Dating” [5]. Uurimus põhineb suure internetipõhise tutvumisportaaali kasutajaandmetel, kus analüüsi nii kasutajate demograafilisi tunnuseid (nt sissetulek, haridus, välimus) kui ka nende käitumist platvormil – näiteks kellele kirjutatakse, kas saadakse vastuseid, ja milliste omaduste alusel eelistusi tehakse. Uuringus leiti, et kasutajate käitumine võimaldab tuletada nende tegelikke eelistusi ja et sobivuste kujunemine järgib selgeid mustreid, kus näiteks sissetulek ja haridus mängivad olulist rolli. Samuti võrreldi online-kooskõlasid Ameerika Ühendriikide abieluandmetega, tuues esile nii sarnasusi kui ka erinevusi tegelikus ja virtuaalses paarivalikus.

2. Metodoloogia

2.1. Andmestik ja andmete eeltöötlus

Oma uurimistöös kasutan andmeid, mis pärinevad Digiteeritud Eesti Artiklite (DEA) kollektsioonist [6]. Need on kergesti ligipääsetavad veebikeskkonnas. Kolleksioon sisaldab Eesti ajalehtede ja ajakirjade täistekste, mis on digitaliseeritud ja salvestatud veebivaatlejatesse. Ligipääs nendele andmetele toimub läbi Tartu Ülikooli hallatud JupyterHub keskkonna, kus saab jooksutada Pythonit ja R-i. Pärast kasutaja autentimist on võimalik pääseda ligi nende metaandmetele, teha oma analüüse ja salvestada tulemused.

Andmestikule mugavaks ligipääsuks on loodud R-pakett digar.txts, mille abil saab tervikkolleksioonist eraldada konkreetseid alamhulki ning teha täistekstilisi otsinguid [7]. See võimaldab keskenduda vaid uurimisteamiga seotud artiklitele, jättes kõrvale mitte relevantseid tekstid ja parandades seeläbi töötluse täpsust ning tõhusust. Tähtis on mainida, et kõik artiklid ei ole avalikult kättesaadavad, seetõttu ei avalda me ka antud töös kogutud andmestiku.

Tekstianalüüsi meetodid

Tutvumiskuulutuste analüüs nõuab spetsiifilisi tekstianalüüsi meetodeid, et eraldada kuulutused muust tekstist ning töödelda neid süstemaatiliselt. Peamiselt kasutatavad meetodid hõlmavad:

- Tekstikaeve ja andmete puhastamine - Tekstikaeve on protsess, mis keskendub struktureerimata teksti töötlemisele ja analüüsimisele [8]. Selle etapi peamiseks eesmärgiks on tutvumiskuulutuste eristamine ajalehtede muust sisust, kasutades regulaaravaldisi, masinõppe meetodeid ja semantilisi mudeleid. Tekstide puhastamine on kriitiline samm, et luua analüüsiv andmestik.
- Statistilised ja arvutuslikud meetodid - Tutvumiskuulutuste muustrite ja sageduste analüüsiks kasutatakse statistilisi lähenemisi, näiteks sagedusanalüüsi ja masinõpet.
- Loomulik keele töötlust (NLP) - NLP meetodid, nagu nimekasutusentiteetide tuvastamine (NER), sentimentide analüüs ja semantilise sarnasuse määramine, aitavad tuvastada ja kategoriseerida teksti peamisi omadusi, näiteks vanusegrupid ja omadused [9].

Kultuurianalüütika ja andmete tõlgendamine

Kultuurianalüütika lähtub sotsiaalsete tekstide, sealhulgas tutvumiskuulutuste, sisuanalüüsi ja tähenduste määratlemisest. Seda tööd tehakse sageli kvalitatiivsete ja kvantitatiivsete meetodite kombinatsiooni abil:

1. Kvalitatiivne uurimine - keskendub tekstide sisu tõlgendamisele ja kategooriate tuvastamisele, nagu soovitud omadused, suhete vormid ja väärtushinnangud [10].
2. Kvantitatiivne uurimine - kvantifitseerib andmeid, näiteks millised vanusegrupid on kõige aktiivsemad, millised omadused korduvad kõige sagedamini jne. See lähenemine pakub empiirilist tuge kvalitatiivsetele leidudele [10].

Andmete puhastamine ja eeltöötlus

Selleks, et saaksime tekstiandmeid analüüsida on meil vaja neid puhastada ning standardiseerida. Esiolgu andmed sisaldasid HTML-märgistusi, metaandmeid ja muid sümboleid, mis tuli eemaldada. Kasutasin järgmisi samme:

1. Andmete lugemine: Kõigepealt lugesin sisse andmed, et saada ülevaade kogu kollektiooni sisust. Valisin alustuseks juhuslikult mõne tuntuma Eesti ajalehe (näiteks Päevaleht aastatel 1970-2010).
2. Teksti puhastamine: Tekstide puhastamiseks kasutasin erinevaid regulaaravaldisi, et eemaldada, järgmisi elemente:
 - HTML-märgistused: Kasutasin funktsioone, et eemaldada kõik HTML-i märgistused, mis olid jäänud tekstidesse, näiteks “<p>” ja “<p dir=""auto"">”.
 - Metaandmete markerid: Eemaldasid ebavajalikke metaandmete tähiseid, nagu "@@@@_NationalLibraryofEstonia_DigitalCollectionsData_SeeUsageRights_in_data.digar.ee_@@@@" ja "inflating:?", kasutades regulaaravaldisi, et tagada, et ainult artiklite sisu jääks alles.
 - Muu müra: Samuti eemaldasid igasugused ebavajalikud sümbolid nagu näiteks "■" ja "•".
 - Eemaldasid kõik liigsed tühikud, reavahed ja tabulaatorid, et muuta tekst ühtlaseks ja töötlusvõimeliseks.

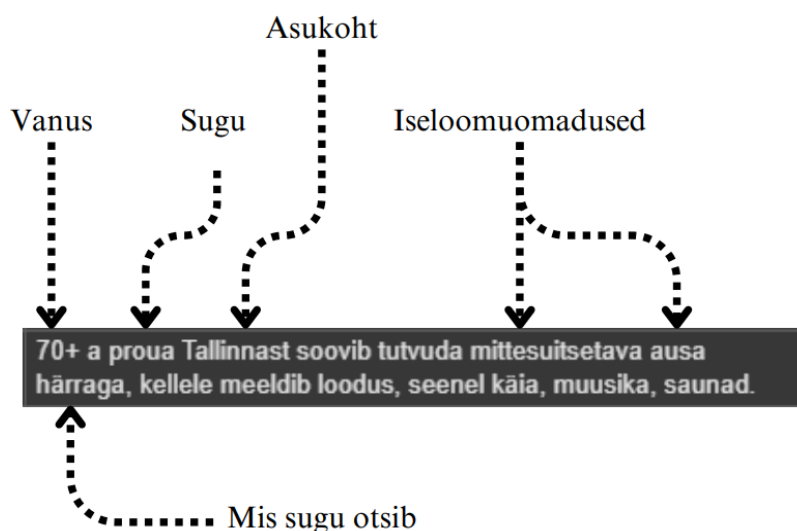
Vaatlusandmete valimine

Andmete valiku tegemine oli oluline osa uurimistööst, kuna oli vaja piirata valimit arvutuslike piirangute tõttu, et töödelda sobiva suurusega andmeid. Samuti tuleb arvestada ajalehtede formaadi erinevusi, mis võivad mõjutada andmete struktuuri ja töötlemist. Suuremad ajalehed, nagu Postimees ja Päevaleht, ei sobinud, kuna need ei sisaldanud piisavalt tutvumiskuulutusi. Algselt leidsin vaid üksikud kuulutused, mis ei andnud piisavat materjali analüüsiks.

Nelli Teataja valimine andmete kogumise allikaks oli põhjendatud selle suure mahu ja sisukuse tõttu. Just Nelli Teataja pakkus kõige rohkem tutvumiskuulutusi, mille kvaliteet ja kogus olid vajalikud, et koostada usaldusväärne andmestik. Lisaks oli Nelli Teataja kuulutuste struktuur selgelt määratletud, näiteks oli sugu alati kohe esitatud, mis lihtsustas andmete analüüsi.

2.2. Andmeomaduste kaeve

Kuulutuste analüüsimiseks loodi struktureeritud tabel, mis sisaldab kuulutuse ID-d, teksti, sugu, otsitavat sugu, asukohta ja iseloomuomadusi. Joonisel 1 on toodud näide ühest kuulutusest, kus nooltega on tähistatud sugu, otsitav sugu, asukoht ja iseloomuomadused. Kuulutuse ID-d joonisel ei kuvata, kuna see eemaldati andmete puhastamise käigus ja salvestati eraldi lahtrisse.



Joonis 1. Näide tutvumiskuulutuse tekstist

2.2.1. ID ja Sugu omaduse määramine

Andmete töötlemise protsess algas puhastatud Nelli Teataja tekstifaili lugemisega, mille järel ühendati kõik andmed üheks pikaks tekstijadaks. See samm lihtsustas artiklite otsimist ja tagas, et kogu sisu oli ühes reas kergesti töödeldav. Tutvumiskuulutuste eraldamiseks kasutasin regulaaravaldisi, kuna Nelli Teatajas algasid need kuulutused pealkirjadega “NAINE MEEST”, “MEES NAIST”, “MEHED” või “NAISED”.

Täpsemaks eraldamiseks kasutasin R-i stringr paketi funktsioone, mille abil leidsin vastavad kuulutused ja jagasin need eraldi artikliteks. Seejärel töötlesin iga artiklit üksikult, et tuvastada seal sisalduv tekst ja viiekohaline ID-kood, mis oli seotud kuulutusega. Kuna igas kuulutuses esines ID, oli oluline see välja võtta ja siduda konkreetse kuulutusega. Samuti määrasin kuulutuse omaniku soo, lähtudes pealkirjast: kui kuulutus oli esitatud rubriigis “MEES NAIST” või “MEHED”, järeldasin, et autor oli mees; kui pealkirjaks oli “NAINE MEEST” või “NAISED”, siis oli autor naine.

Lõpuks ühendasin mõlemad kuulutustekogumid – nii need, mis kuulusid rubriiki “NAINE MEEST” ja “NAISED”, kui ka need, mis olid pealkirja “MEES NAIST” ja “MEHED” all – üheks terviklikuks tabeliks, mis võimaldas kõiki kuulutusi koos analüüsida.

2.2.2. “Otsib sugu” omaduse määramine

Määramaks, mis sugu inimene kuulutusest otsib, kasutasin OpenAI GPT-3.5 Turbo API-t, et analüüsida iga kuulutuse teksti ning tuvastada, kas inimene otsib meest, naist või sõpra/suhtluskaaslast. Tavaline regulaaravaldis ei sobinud, kuna kuulutuste keel oli sageli mitmekesine ja sisaldas keerulisi lauseid, mis ei järginud kindlat mustrit. Selle teostamiseks lisasin andmestikku uue “Otsib_sugu” klassi, mille algväärtus oli NA. Seejärel töötlesin iga kuulutuse teksti, genereerides selle põhjal API päringu.

API päringu küsimuse vormistasin järgmiselt: “*Mis soost inimest otsitakse kuulutuses (vastus anna ühe lausena, kas 'otsib meest', 'otsib naist' või 'otsib sõpra', kui sugu pole antud, ära kasuta teisi väljendeid)?*”. Saadud vastuse lisasin andmetabelisse vastava rea veergu “Otsib_sugu”. Kuna API-päringutes võib esineda vigu, kasutasin tryCatch funktsiooni, et veahaldus oleks sujuv. Vea korral salvestasin väärtuseks NA.

Pärast andmete esmast töötlemist korrastasin vastuseid, asendades ebatäpsused ning standardiseerides väljundid. Selleks lõin funktsiooni, mis kontrollis iga “Otsib_sugu” rea

teksti, otsides märksõnu: “meest”, “naist” ja “sõpra”. Iga vastava sõna korral, määrati sellele õigustatud väärtus.

2.2.3. Inimese vanuse määramine

Vanuse määramiseks kasutasin sarnast lähenemist nagu soo tuvastamisel, kuna tavaline regulaaravaldis ei sobinud. Vanuse mainimine kuulutustes võib olla esitatud erinevates vormides, näiteks numbritena või sõnalistes vormides.. Lõin R-is tsükli, mis töötles igati kuulutuse teksti ning edastas selle OpenAI API-le küsimusega: “Mis on tekstis oleva inimese vanus, kes otsib tutvust? Kui vanust pole, siis pane 'NA' (vastus anna ainult numbrina)”. Saadud vastuse lisasin tabelisse veergu “inimese_vanus”.

Peale vanuse esmast tuvastamist tuli andmeid veel puhastada. Selleks kasutasin regulaaravaldisi, et eemaldada mittevajalikud sümbolid ja jättes alles ainult numbrid või numbrite vahemikud. Kasutasin selleks ”gsub("[^0-9\\+]", "", data\$inimese_vanus)”, mis eemaldab kõik muud sümbolid peale numbrite, miinusmärgi ja plussmärgi.

2.2.4 Asukoha määramine

Asukoha määramiseks kasutasin TartuNLP BERT-põhist nimetuvastuse mudelit (NER), et leida ja ekstraktida kuulutustest kohanimed [11]. Selleks saatsin iga kuulutuse teksti TartuNLP API-le ning filtreerisin saadud tulemustest ainult LOC (*location*) tüüpi märgendiga sõnad. Leitud kohanimed liitsin kokku ning salvestasin uude veergu “Asukoht”.

Andmete täiendavaks korrastamiseks eemaldasid kordused ja normaliseerisin asukohanimed, rakendades nendele lemmatiseerimist EstNLTK tööriista abil [12]. See vähendas sama tähendusega, kuid erinevates käändes esinevate asukohtade esinemist (nt “Tallinnas” -> “Tallinn”).

2.3. Iseloomuomaduste ekstraktimine ja korrastamine

Iseloomuomaduste määramiseks kasutasin EstNLTK morfoloogilist analüüsi, et tuvastada kuulutustes omadussõnu, mis kirjeldavad kuulutuse postitanud isikut või tema poolt soovitud partnerit. Töötlesin iga kuulutuse teksti järgmiselt:

1. Kasutasin “Tekst” klassi, et töödelda teksti morfoloogilise analüüsi kihiga.
2. Otsisin kõik sõnad, mille sõnaliik sisaldab tähist “A” (adjektiiv).
3. Valisin ja lemmatiseerisin need omadussõnad, et tagada ühtne vorm.
4. Koostasid neist komaga eraldatud loendi ja lisasin tabelisse uue veeruna.

Seejärel eemaldasid korduvad sõnad, et jätta alles ainult unikaalsed omadused.

Kõik saadud tulemused salvestasin erinevates etappides CSV-failina, et säilitada töövoos korrektsus. Lõpuks ühendasin erinevad andmestikud kasutades, et moodustada lõplik kuulutuste andmefail, mis sisaldab kõiki täiendatud ja korrastatud andmeid.

Selle protsessi tulemusena sain struktureeritud ja korrastatud tabeli, kus igale kuulutusele on lisatud tunnused: “ID”, “Tekst”, “Sugu”, “Otsib_sugu”, “Asukoht”, “inimese_vanus” ja “Iseloomuomadused”. Need väljad võimaldavad teha edasise analüüsi ning uurida kuulutuste sisulisi eripärasid.

2.4. Iseloomuomaduste klastrite loomine

Tabelis esinevad iseloomuomadused suure sagedusega ja paljud neist on semantiliselt sarnased (näiteks “noor”, “nooruslik” ja “nooremapoolne”). Sellel põhjusel oli otstarbekas sõnad rühmitada klastritesse, mis võimaldas:

1. **Vähendada andmete mõõtmelisust:** 544 erineva iseloomuomaduse asemel sain 96 sisukat klastrit.
2. **Tuvastada üldisemaid mustreid:** Sarnase tähendusega omaduste kombineerimine näitab selgemaid suundumusi.
3. **Lihtsustada tulemuste tõlgendamist:** Klastrite pealkirjad annavad kiire ülevaate peamistest iseloomutüüpidest.
4. **Võimaldada kvantitatiivset analüüsi:** Klastrite sagedused ja seosed muude tunnustega on usaldusväärsemad.

Klastrite loomise protsess algas sõnavektorite eraldamisega. Selleks kasutasin Eesti keelele spetsialiseeritud Word2Vec mudelit “Pretrained word and multi-sense embeddings for Estonian” [13]. Mudel (cbow_750_5_10_20.word_vectors) sisaldas 750-mõõtmelisi sõnavektoreid, mis olid treenitud suurel eesti keele tekstikogul. Katsetasin kolme erinevat klastrimisalgoritmi:

1. **K-means**: iteratiivne, tsentroidil põhinev klasterdamisalgoritm, mis jagab andmestiku sarnasteks rühmadeks nende tsentroidide vahelise kauguse alusel. Tsentroid ehk klasteri kese on olenevalt andmete omadustest kõigi klasteri punktide keskmine või mediaan [14].
2. **DBSCAN**: *Density-Based Spatial Clustering of Applications with Noise*. Leiab suure tihedusega põhivalemeid ja laiendab nende põhjal klastreid. Sobib andmete jaoks, mis sisaldavad sarnase tihedusega klastreid [15].
3. **Gaussi segud (GMM)**: paindlikum lähenemine. Erinevalt K-means-ist, mis määrab iga punkti ühele klasterile, kasutab GMM andmete klasterdamiseks tõenäosuslikku lähenemist, mis võimaldab klasteritel olla mitmekesisema kuju ja pehmete piiridega [16].

Kõige ühtlasemalt ja loogilisemalt jagas omadussõnad rühmadesse K-means meetod.

Pärast masinõppemeetoditega klastrite loomist viisin läbi käsitsi kontrolli, et tagada iga omadussõna paiknemise sobivas klasteris. Iga klasteri jaoks määratlesin pealkirja, mis kirjeldas klasterisse kuuluvate omaduste ühist tähendust.

3. Tulemused

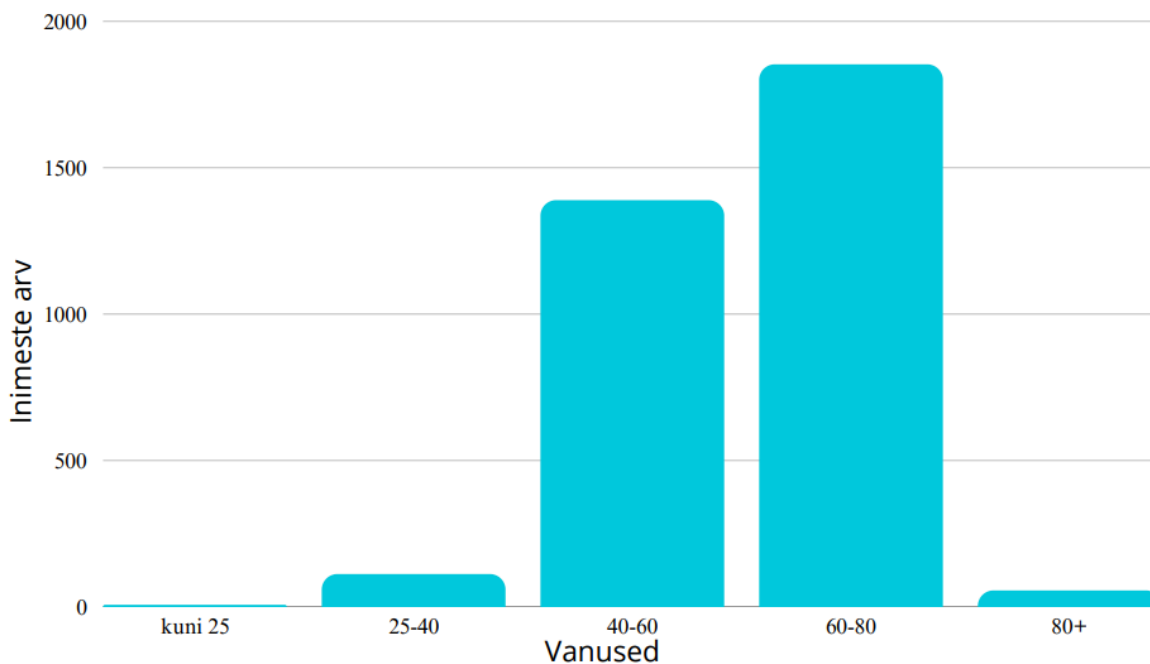
3.1 Ühe muutuja analüüs

Mis on otsijate vanused?

Analüüsisin kuulutuste postitajate vanust ning jagasin nad viide vanuserühma. Suurim grupp kuulutuste esitajatest on vanuses 60-80 aastat, moodustades 1853 kuulutust. Järgnevad vanuserühmad on 40-60 aastat (1389 kuulutust) ja 25-40 aastat (112 kuulutust). Noorima vanuserühma (kuni 25 aastat) esindajaid on vaid 7, ning kõige vanemasse kategooriasse (80+) kuulub 56 kuulutust. Joonis 2 illustreerib selgelt, kuidas enamik kuulutusi pärineb vanematelt inimestelt.

Selline jaotus võib olla seotud mitmete teguritega. Esiteks, kuigi nooremad inimesed kasutavad tihti *online*-tutvumisplatvorme ja mobiilirakendusi, on vanemad inimesed enamasti harjunud traditsiooniliste meediavahenditega, nagu ajalehed, et leida endale kaaslast. Tutvumiskenduste populaarsus on tõepoolest nooremate seas kasvanud, kuid vanemaealised inimesed võivad veel eelistada kuulutuste kaudu suhtlemist, kuna see meetod on neile tuttav ja mugav. Samuti on vanemaealiste seas suurem tõenäosus otsida täiendavat kaaslast, eriti lehestumise või tühja pesaga seonduvate sotsiaalsete muutuste tõttu [17].

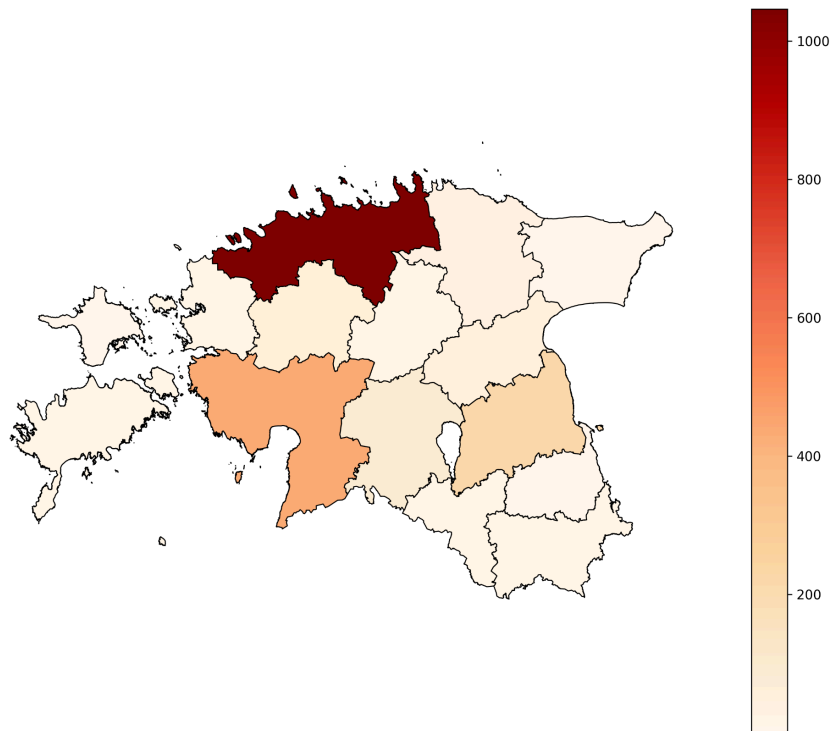
Samuti on oluline mainida, et 60-80 vanusegrupis otsitakse tihti täiendavat vajaduspõhist kaaslast, kuna paljud inimesed selles vanuses võivad olla lehestunud või otsivad kedagi, kellega jagada igapäevaelu. Vanemaealiste jaoks on kuulutused sageli võimalus leida inimest, kellega jagada mitte ainult romantilisi, vaid ka praktilisi vajadusi, näiteks igapäevaelu toetust või seltskonda. Nooremate seas on seevastu tutvumiskuulutuste kasutamine harvem ja pigem ajendatud soovist leida romantilisi suhteid, kui mitte praktilisi lahendusi.



Joonis 2. Vanuserühmade jaotus

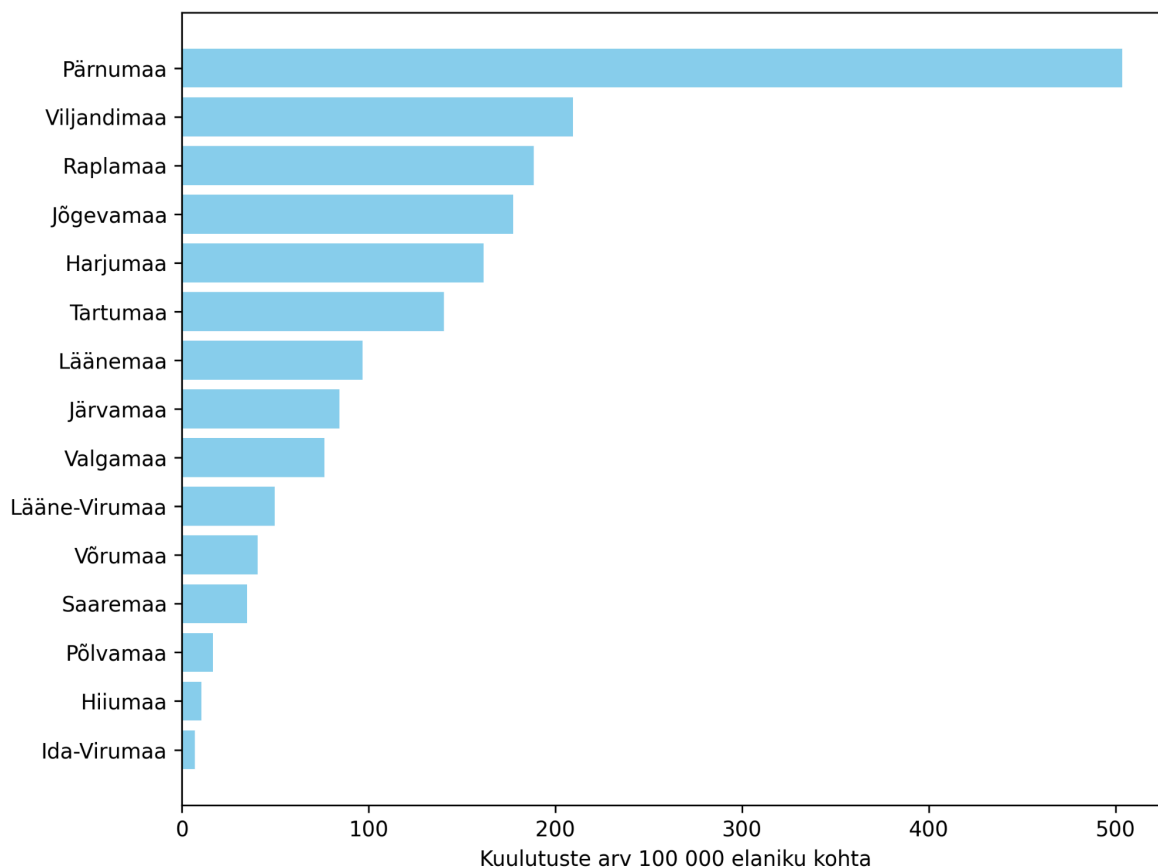
Kust kaaslast otsitakse?

Kuulutustes mainitud asukohtade põhjal selgub, et enim otsitakse tutvusi Harjumaa (1046 kuulutust), järgneb Pärnumaa (441 kuulutust) ja Tartumaa (227 kuulutust). Ülejäänud maakondades on kuulutuste arv oluliselt väiksem, kusjuures Hiiumaalt on registreeritud vaid üks kuulutus. Joonis 3 näitab selgelt, et suurem osa kuulutusi pärineb tihedama asustusega piirkondadest, samas kui väiksemates maakondades on kuulutuste arv väga väike.



Joonis 3. Kuulutuste arv maakondade lõikes

Samas ei pruugi absoluutarvud anda kogu pilti. Kui vaadelda kuulutuste arvu 100 000 elaniku kohta, kerkivad esile maakonnad, näiteks Pärnumaa, Viljandimaa ja Raplammaa, kus *per capita* kuulutuste määr on kõrgem kui suurima rahvaarvuga Harjumaal, mis langeb alles viiendale kohale [18]. Seega annab *per capita* analüüs parema ülevaate sellest, millistes maakondades olid kuulutused võrreldes rahvaarvuga populaarsemad.



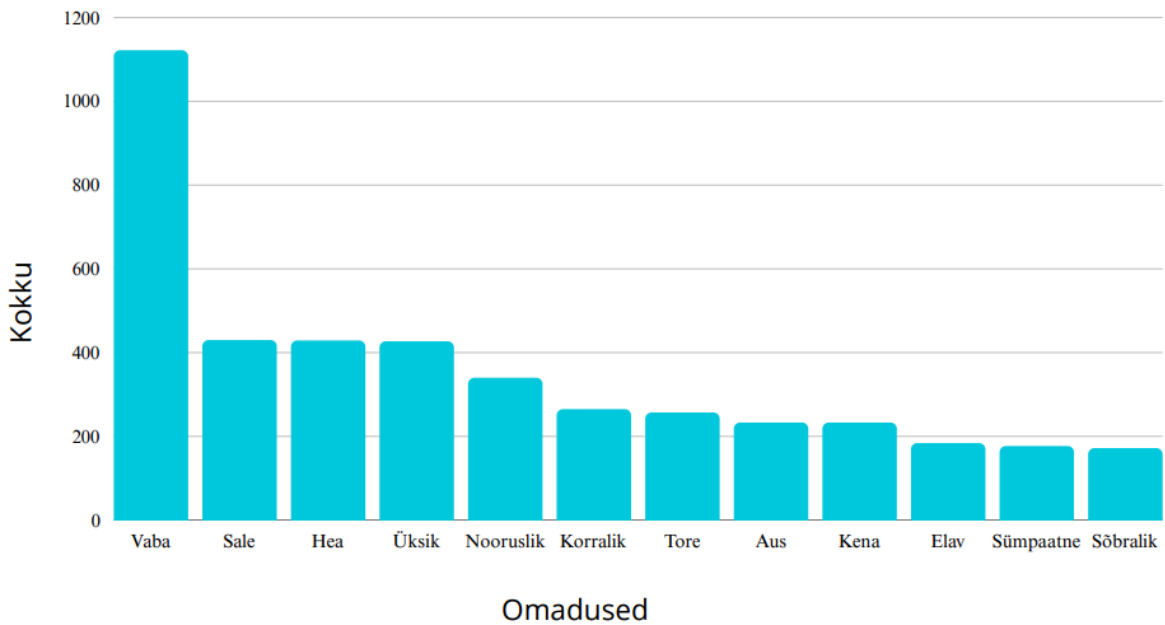
Joonis 4. Kuulutuste arv 100 000 elaniku kohta maakonnas

Millised omadusi otsijad mainivad?

Kuulutustes oli kokku mainitud 555 erinevat omadust. Enim kasutatud omadussõna olid "vaba" (1122 korda), mis viitab sellele, et paljud kuulutuste postitajad rõhutavad oma vallalisust või vabadust. Teised sagedased omadused olid "sale" (430), "hea" (429), mis viitab sellele, et füüsiline vorm, iseloomuomadused ja sotsiaalne staatus on sageli esile tõstetud.

Samuti mainiti tihti "nooruslik" (340), "korralik" (265), "tore" (257), "aus" (233), "elav" (184), "sümpaatne" (177) ja "sõbralik" (172), mis viitab sellele, et kuulutustes peetakse oluliseks nii iseloomu kui ka elustiili.

Joonis 5 illustreerib selgelt, millised iseloomujooned ja omadused on kuulutustes kõige sagedamini esile tõstetud.



Joonis 5. Kõige sagedamini mainitud isikuomadused tutvumiskuulutustes.

Postitajate sugu

Kuulutuste postitajate seas on naisi veidi rohkem kui mehi. Kokku on 2351 naist ja 2210 meest, mis tähendab, et naised on kuulutuste avaldamisel aktiivsemad. Kui vaadata Eesti rahvastiku soolist jaotust, siis 2025. aasta seisuga on Eestis 649 753 meest ja 720 242 naist, mis tähendab, et naised moodustavad umbes 52,5% rahvastikust [19]. Kuigi kuulutuste postitajate seas on naisi rohkem, ei ole see suurenenud osakaal siiski kaugel riiklikust soolisest jaotusest, kus naised on samuti ülekaalus.

Tabel 1. Kuulutuste postitajate sooline jaotus

	Mees	Naine
Kokku	2210	2351

Keda otsitakse?

Tutvumiskuulutustes väljendatakse erinevaid eelistusi, kuid üldiselt otsitakse kõige sagedamini meest (2217 kuulutajat) või naist (2020 kuulutajat), samas kui osa kuulutajatest soovib lihtsalt vestluskaaslast või sõpra (324 kuulutajat).

Tulpdiagramm illustreerib, et meeste ja naiste otsijate arv on üsna tasakaalus, kuid leidub ka neid, kes ei otsi tingimata romantilist suhet, vaid pigem seltskonda ja vestlust.

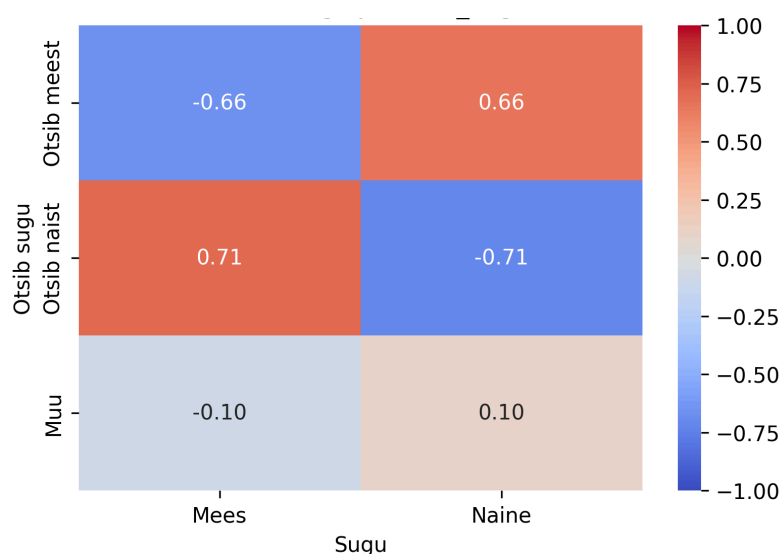
Tabel 2. Meeste ja naiste otsijate jaotus tutvumiskuulutustes

	Otsib meest	Otsib naist	Otsib suhtluskaaslast
Kokku	2217	2020	324

3.2 Kahe muutuja analüüs

Korrelatsioon soo ja otsitava soo vahel

Korrelatsioonide analüüs näitab, et meeste puhul on tugevalt eelistatud naiste otsimine, samas kui naised eelistavad sageli mehi. Positiivne korrelatsioon väärtusega 0.66 ("Otsib meest", "Naine") ja 0.71 ("Otsib naist", "Mees") viitab sellele, et sooline eelistus partnerite suhtes järgib traditsioonilisi heteroseksuaalseid suhteid.



Joonis 6. Korrelatsioonid soo ja otsitava soo vahel

Korrelatsioon asukoha ja otsib sugu omaduse vahel

Analüüsides asukoha ja otsitava soo vahelisi korrelatsioone, ilmnevad järgmised statistiliselt olulised seosed:

1. Harjumaa

- Otsib meest: positiivne korrelatsioon (0.11) - Harjumaal otsitakse mehi sagedamini

2. Jõgevamaa

- Otsib naist: positiivne korrelatsioon (0.05) - Jõgevamaal otsitakse naisi veidi sagedamini

3. Pärnumaa

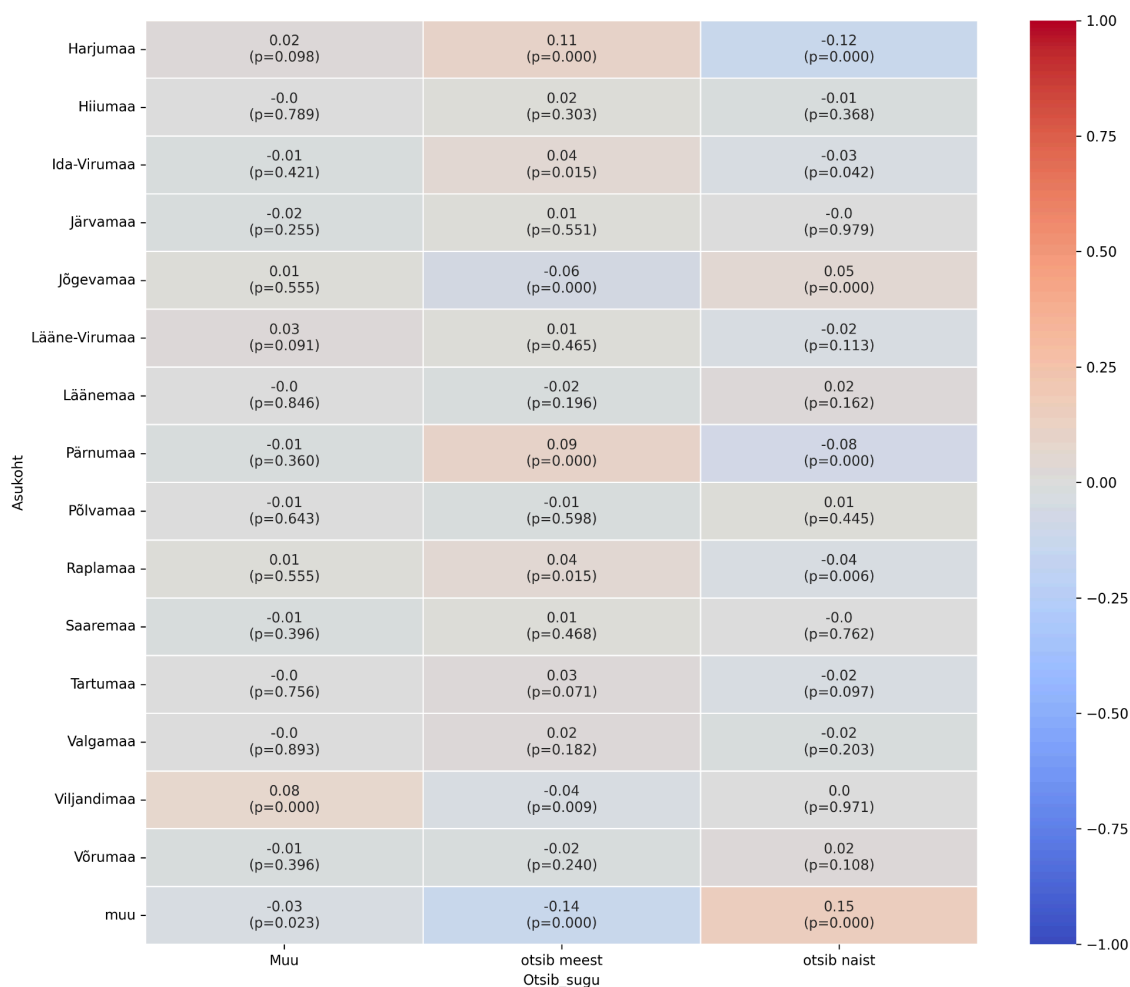
- Otsib meest: positiivne korrelatsioon (0.09) - meeste otsimine on sagedasem

4. Viljandimaa

- Muu kategooria: positiivne korrelatsioon (0.08) - Viljandimaal on sõbra/suhtluskaaslaste otsimine suhteliselt levinum

Ülejäänud piirkondades pole korrelatsioon piisavalt tugev, et nende põhjal järeldusi teha.

Need mustrid võivad peegeldada piirkondlikke erinevusi soolises jaotuses ja partneriotsingu eelistustes. Praeguste tulemuste põhjal võib linnapiirkondades (Harjumaa ja Pärnumaa) võib olla rohkem meeste otsimisega avaldatud kuulutusi, samas kui maapiirkondades (Viljandimaa ja Jõgevamaa) domineerib naiste otsimine.



Joonis 7. Korrelatsioonid ja p-väärtused asukoha ja otsitava soo vahel

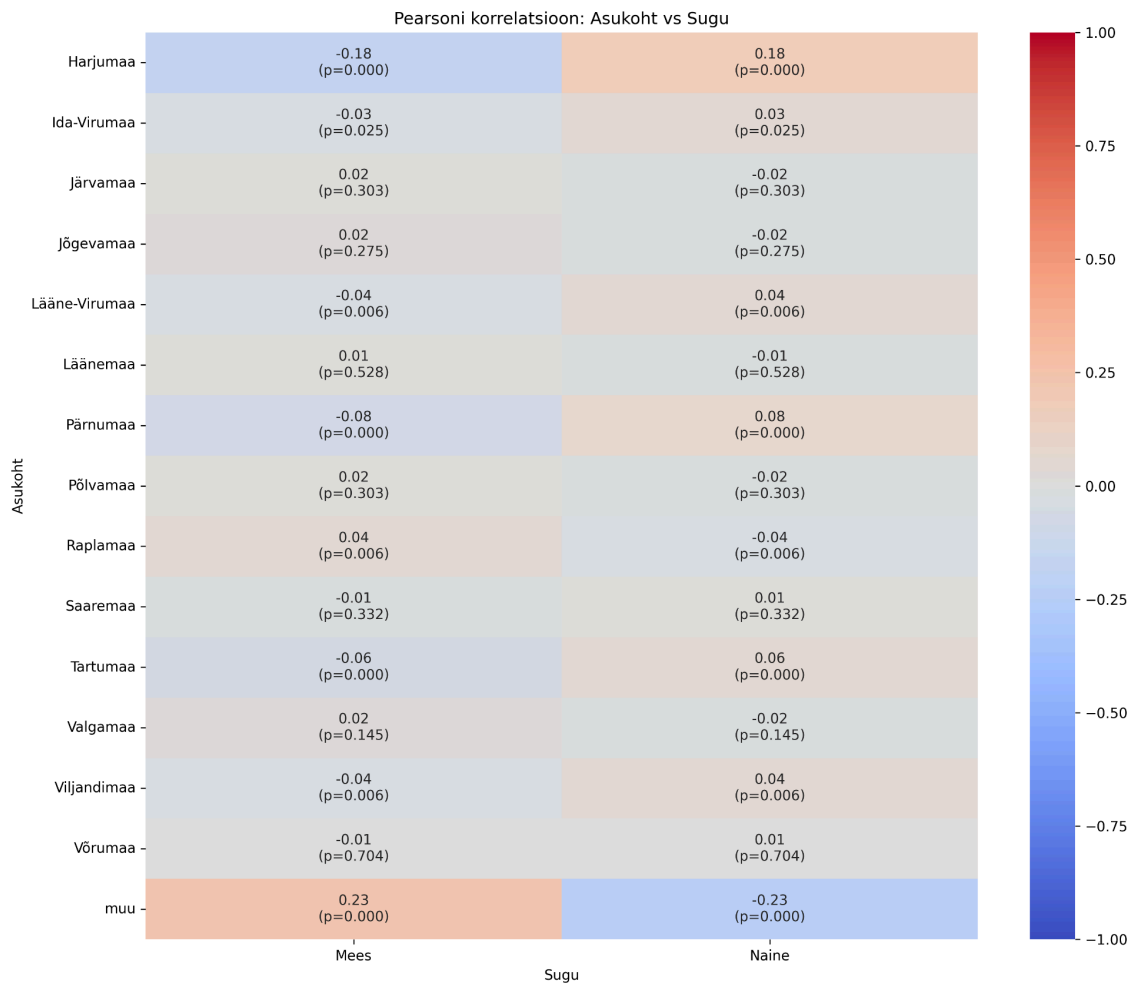
Korrelatsioon asukoha ja soo omaduse vahel

Andmete analüüs näitab, et asukoha ja soo vahel esineb märkimisväärseid korrelatsioone, mis viitavad soospetsiifilistele mustritele erinevates piirkondades. Suurema populatsiooniga maakondades on kuulutustes esindatud mõlemad sood suhteliselt võrdselt, samas kui väiksemates ja äärealadel paiknevates kohtades domineerivad sagedamini mehed. Näiteks Harjumaal on seos naistega (+0.18) ja meestega (-0.18), Pärnumaal (+0.08) ja Tartumaal (+0.06) on samuti kerge positiivne seos naistega. See võib viidata traditsioonilistele soorollidele, kus mehed on ajalooliselt olnud rohkem seotud maaelu ja füüsilise tööga, samal ajal kui naised on liikunud rohkem linnadesse, kus on paremad haridus- ja töövõimalused [20].

Kultuurilisest perspektiivist võib seda mustrit seostada ka sotsiaalsete ootustega ja demograafiliste suundumustega. Linnastumine on sageli seotud suurema soolise võrdsusega, kuna linnakeskkond pakub rohkem võimalusi iseseisvuseks ja karjääri arendamiseks, mis võib soodustada naiste aktiivsemat osalust avalikus ruumis. Seevastu maapiirkondades võib traditsiooniliste väärtuste püsimine mõjutada meeste ülekaalu kuulutustes, peegeldades soolist jaotust ning tööjaotuse eripära.

Veelgi enam, mehed võivad äärealadel olla aktiivsemad partneriotsingutes, kuna naiste vähesus nendes piirkondades suurendab motivatsiooni suhete loomiseks. Naistel aga võib suuremates linnades olla laiem valik ja vähem sotsiaalset survet püsisuhte loomiseks. Need leiud kinnitavad, et geograafiline asukoht ei ole pelgalt füüsiline ruum, vaid on tihedalt seotud kultuuriliste normide ja soolise dünaamikaga.

Joonis 8 näitab täpsemalt asukoha ja soo omaduse vahelisi seoseid.



Joonis 8. Korrelatsioonid ja p-väärtused asukoha ja soo vahel

Korrelatsioon asukoha ja vanuse vahel

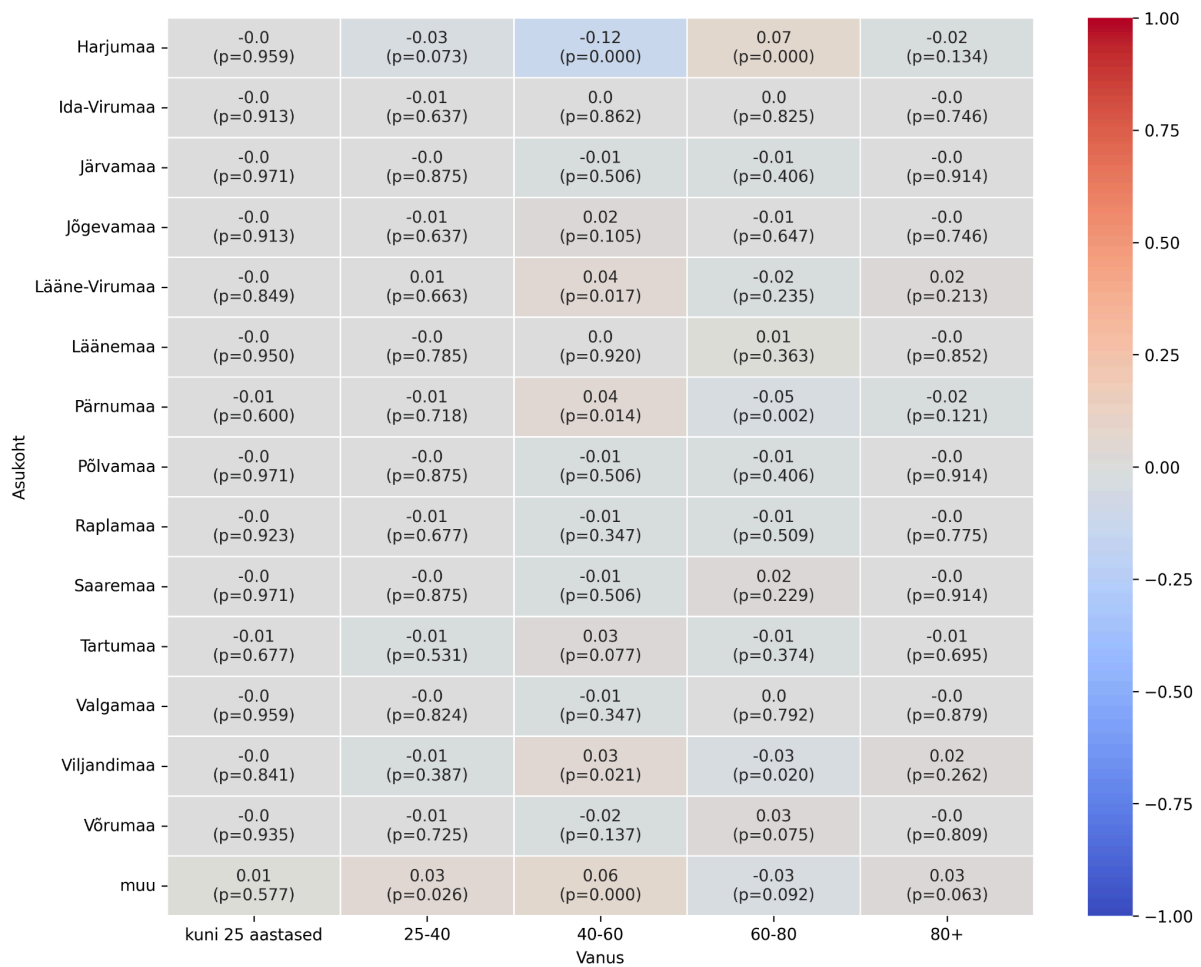
Analüüsid asukoha ja vanuse vahelisi korrelatsioone tutvumiskuulutustes, ilmnevad mõned huvitavad mustrid, kuigi üldised seosed on üsna nõrgad.

Harjumaa näitab märkimisväärsed korrelatsioone. Negatiivne seos vanuserühmaga 40–60 aastat (-0,12), mis viitab, et keskealised inimesed on selles piirkonnas kuulutustes vähem esindatud. Positiivne seos vanuserühmaga 60–80 aastat (0,07), mis näitab vanemate aktiivsust tutvumiskuulutuste avaldamisel.

Ülejäänud piirkonnad ei näita tugevaid korrelatsioone ühegi vanuserühmaga, mis viitab sellele, et vanuseline jaotus on nendes piirkondades ühtlasem.

Linna ja maapiirkondade erinevused: Harjumaa kui suurim linnastunud piirkond võib olla atraktiivne vanematele inimestele, kes otsivad seltskonda või abikaasat, samas kui keskealised võivad kasutada rohkem teisi tutvumisvahendeid.

Kuigi asukoha ja vanuse vahelised korrelatsioonid ei ole väga tugevad, on siiski märgata piirkondlikke erinevusi. Linnastunud piirkondades (eriti Harjumaa) on vanemad inimesed aktiivsemad tutvumiskuulutuste avaldajad, samas kui maapiirkondades on vanuselised erinevused vähem olulised.



Joonis 9. Korrelatsioonid ja p-väärtused asukoha ja vanuse vahel

Korrelatsioon vanuse ja otsib sugu omaduse vahel

Analüüsid vanuserühmade ja otsitava soo vahelisi korrelatsioone, ilmnevad mõned statistiliselt olulised seosed.

1. Otsib naist

- 25–40-aastased: Positiivne korrelatsioon (0.13) – nooremad täiskasvanud otsivad sagedamini naisi.
- 40–60-aastased: Positiivne korrelatsioon (0.19) – keskealised inimesed otsivad oluliselt sagedamini naisi.

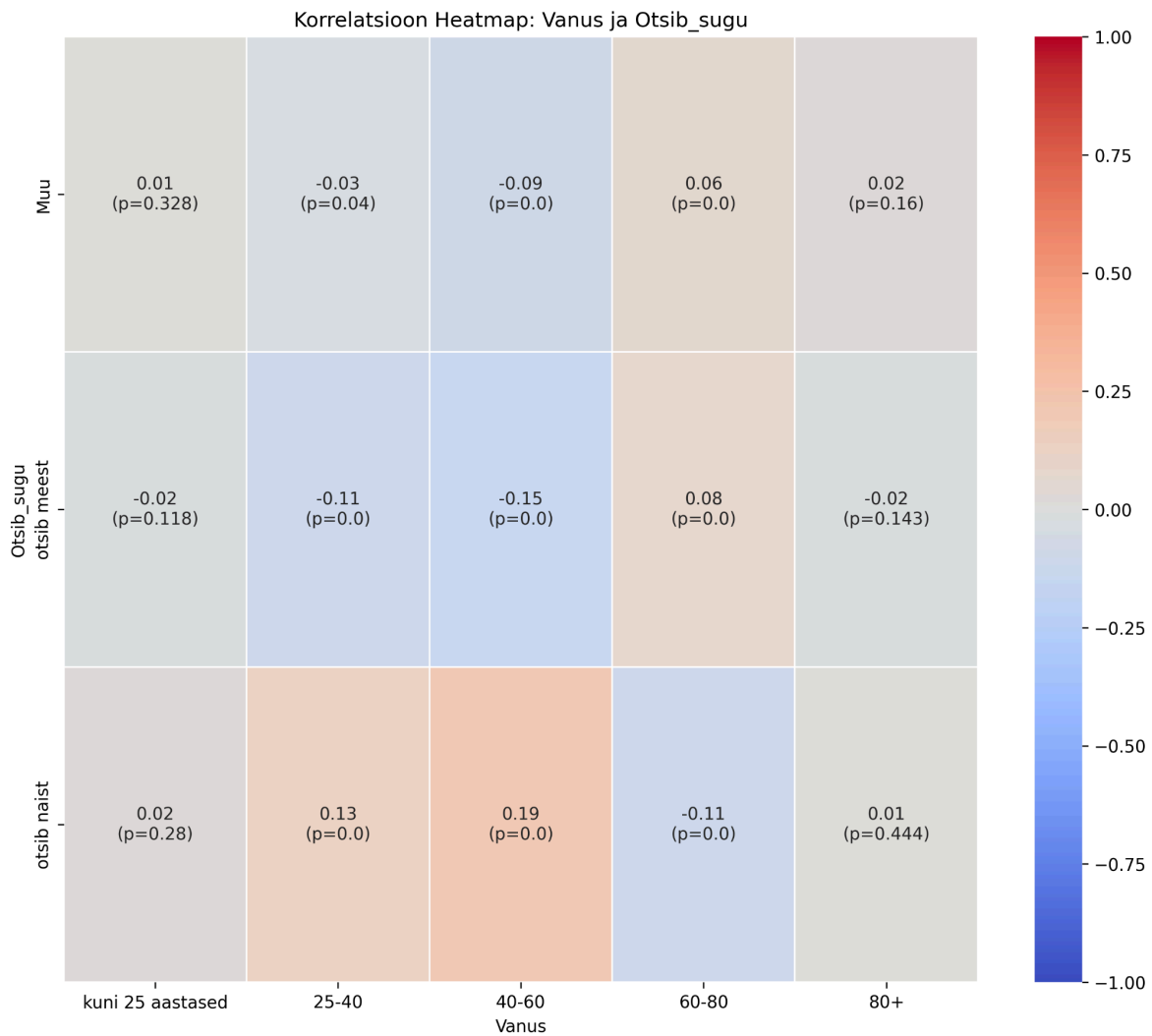
2. Otsib meest

- 60–80-aastased: Nõrk positiivne korrelatsioon (0.08) – vanemad inimesed võivad otsida mehi veidi sagedamini.

3. Muu (sõpra/suhtluskaaslast)

- 60–80-aastased: Nõrk positiivne korrelatsioon (0.06) – vanemad inimesed võivad otsida sõpru/suhtluskaaslast veidi rohkem.

Kuni 25 aastastel noortel pole piisavalt suuri korrelatsioone, et teha järeldusi. Nooremad täiskasvanud (25–40) näitavad mõõdukat huvi naiste otsimise vastu, mis võib peegeldada traditsiooniliste suhete eelistusi. Keskealised (40–60) inimesed eelistavad selgelt naiste otsimist, samas kui meeste otsimine on neil haruldasem. Vanemad (60–80) inimesed kalduvad pigem meeste otsimisele, kuid see seos on nõrk. "Muu" kategooria (sõprussuhted) ei näita tugevaid seoseid, kuid vanemad inimesed võivad olla veidi aktiivsemad sõprade otsijatena.



Joonis 10. Korrelatsioonid ja p-väärtused vanuse ja otsitava soo vahel

Korrelatsioon vanuse ja soo omaduse vahel

Korrelatsioonide analüüs vanuse ja soo omaduse vahel näitab, et nende kahe tunnuse vahelised seosed on üldiselt nõrgad, kuid mõned huvitavad mustrid on siiski ilmnenud.

Meeste puhul on korrelatsioonid vanuserühmadega järgmised:

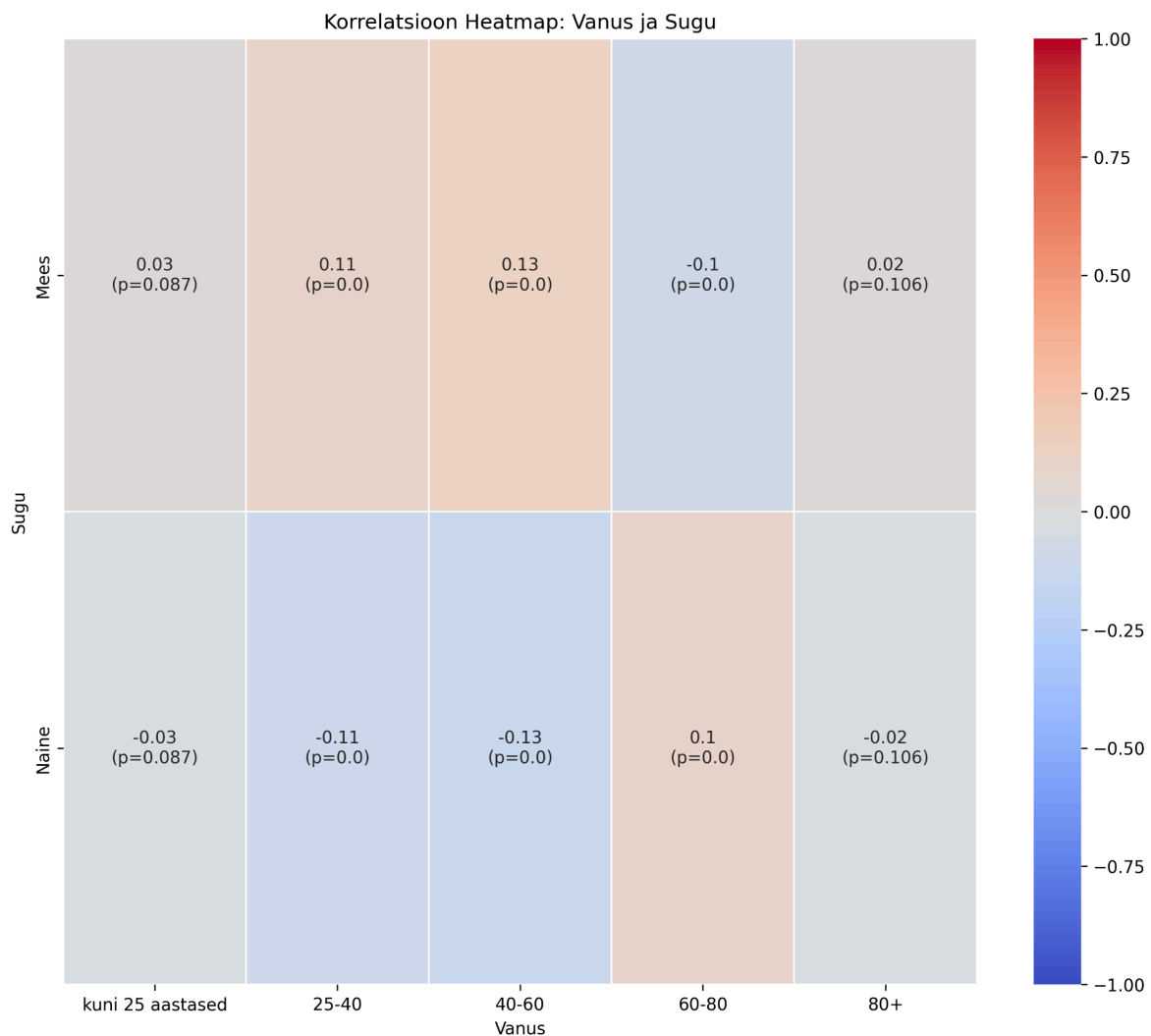
- Kuni 25 aastased: 0.03
- 25–40 aastased: 0.11
- 40–60 aastased: 0.13
- 60–80 aastased: -0.10
- 80+ aastased: 0.02

Negatiivsed väärtused näitavad, et meeste osakaal väheneb vanemates rühmades (eriti 60-80 aastased), kuid on stabiilsem nooremates ja vanemates rühmades.

Naiste puhul on korrelatsioonid:

- Kuni 25 aastased: -0.03
- 25–40 aastased: -0.11
- 40–60 aastased: -0.13
- 60–80 aastased: 0.10
- 80+ aastased: -0.02

Kuigi korrelatsioonid on nõrgad, näitab analüüs, et vanus ja sugu mõjutavad kuulutuste esitamise aktiivsust. Vanemad inimesed (60-80-aastased) on traditsioonilistes tutvumiskanalites aktiivsemad, samas kui nooremad ja kõige vanemad (80+ aastased) generatsioonid kasutavad neid harva.



Joonis 11. Korrelatsioonid ja p-väärtused soo ja vanuse vahel

Korrelatsioon vanuse ja iseloomuomaduste vahel

Korrelatsioonide analüüs näitab selgeid erinevusi eri vanuserühmade väärtushinnangutes ja eelistustes:

- Nooremad generatsioonid rõhutavad välimust ja noorust
- Keskealised inimesed otsivad stabiilsust ja turvalisust
- 60-80-aastased hindavad kõrgelt heatahtlikkust ja rahulikkust
- Väga vanas eas (80+) muutuvad prioriteedid uuesti, kus alkoholivaba eluviis ja tervislikkus muutuvad olulisemaks

Need mustrid peegeldavad eluetappidele omaseid väärtusi ja vajadusi, kus noored keskenduvad välisele atraktiivsusele, keskealised pereelu ja karjääri tasakaalustamisele, ning vanemad generatsioonid harmoonilisele kaaslasele ja tervisele.

Nooremate vanuserühmade (kuni 25 ja 25-40 aastased) puhul on märgatav positiivne korrelatsioon selliste omadustega nagu:

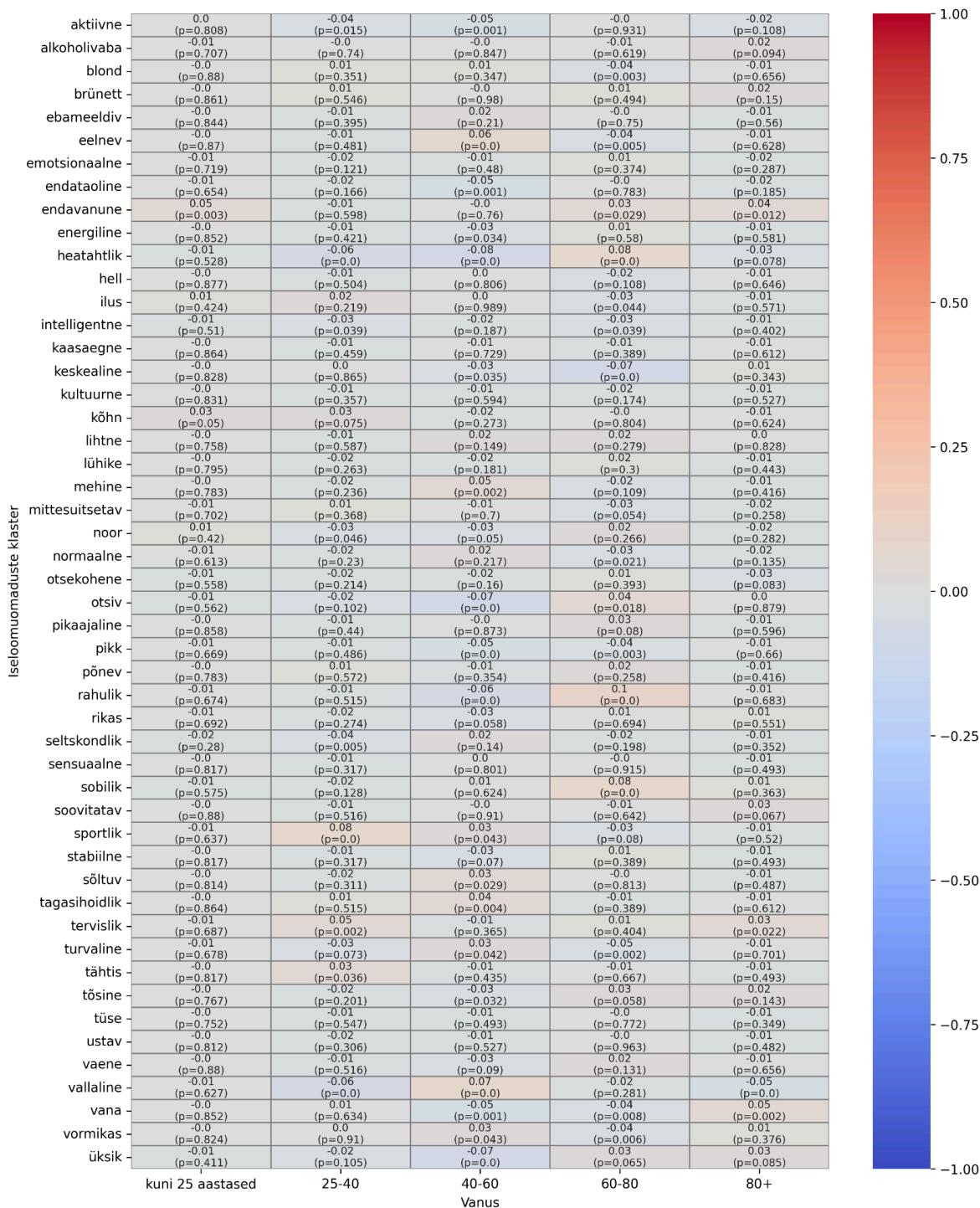
- sportlik (0.08 25-40 aastastel)
- endavanune (0.05 kuni 25 aastastel)

Keskealiste vanuserühma (40-60 aastased) puhul on tugevaimad positiivsed korrelatsioonid järgmiste omadustega:

- vallaline (0,07)
- mehine (0.05)
- eelnev (0.06)

Vanemate vanuserühmade (60-80 ja 80+ aastased) puhul on kõige märkimisväärsamad korrelatsioonid:

- heatahtlik (0,08 60-80-aastastel, -0,03 80+)
- sobilik (0,08 60-80-aastastel, 0.01 80+)
- rahulik (0,10 60-80-aastastel, -0.01 80+)
- vana (-0.04 80-80-aastastel, 0.05 80+)



Joonis 12. Korrelatsioonide väärtused iseloomuomaduste ja vanuse vahel.

Korrelatsioon asukoha ja iseloomuomaduste vahel

Analüüsid asukoha ja iseloomuomaduste vahelisi korrelatsioone, ilmnevad järgmised statistiliselt olulised seosed:

1. Järvamaa

- Soovitatav (0.24) - kõige tugevam positiivne seos, mis näitab, et Järvamaa kuulutustes rõhutatakse erilisel soovitatavust ja hea maine omadusi

2. Harjumaa

- Rahulik (0.18) - hindavad rahulikku loomust
- Emotsionaalne (0.09) - emotsionaalsusele antakse tähelepanu
- Sportlik (0.08) - aktiivne elustiil on oluline
- Pikk (0.09) - füüsilistele omadustele (eriti pikkusele) rõhutatakse

3. Pärnumaa

- Aktiivne (0.10) - aktiivsust esile tõstetakse
- Kõhn (0.08) - kehaehitusele antakse tähelepanu
- Noor (0.09) - nooruslikkust rõhutatakse
- Tõsine (0.09) - tõsidust väärtustatakse

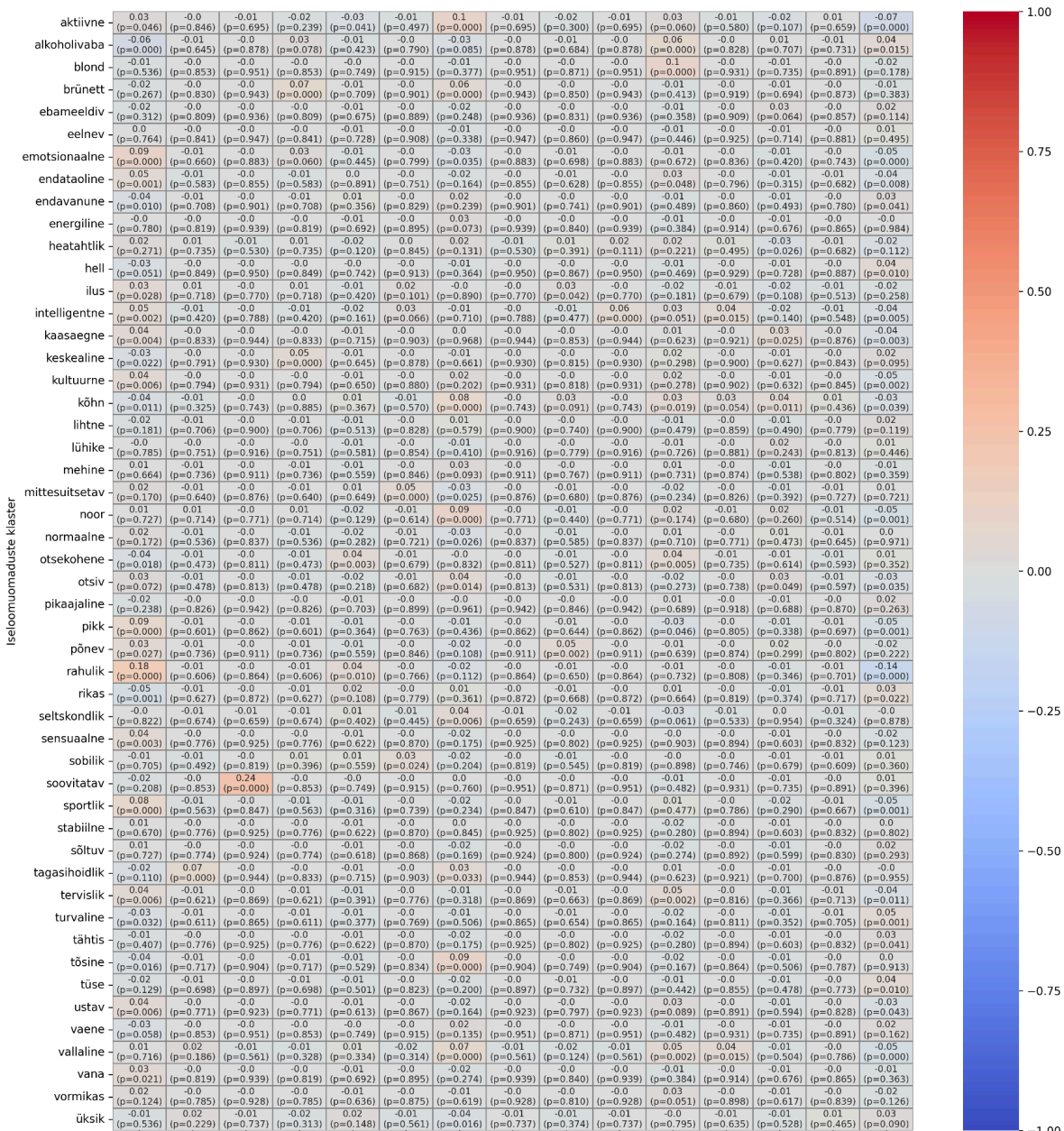
4. Tartumaa

- Blond (0.10) - blondidele antakse erilist tähelepanu

5. Jõgevamaa

- Pikk (0.07) - nõrk positiivne seos pikkuse ja Jõgevamaa vahel

Ülejäänud piirkondades pole korrelatsioonid piisavalt suured, et teha põhjanevaid järeldusi.



Joonis 13. Korrelatsioonide väärtused iseloomuomaduste ja asukoha vahel.

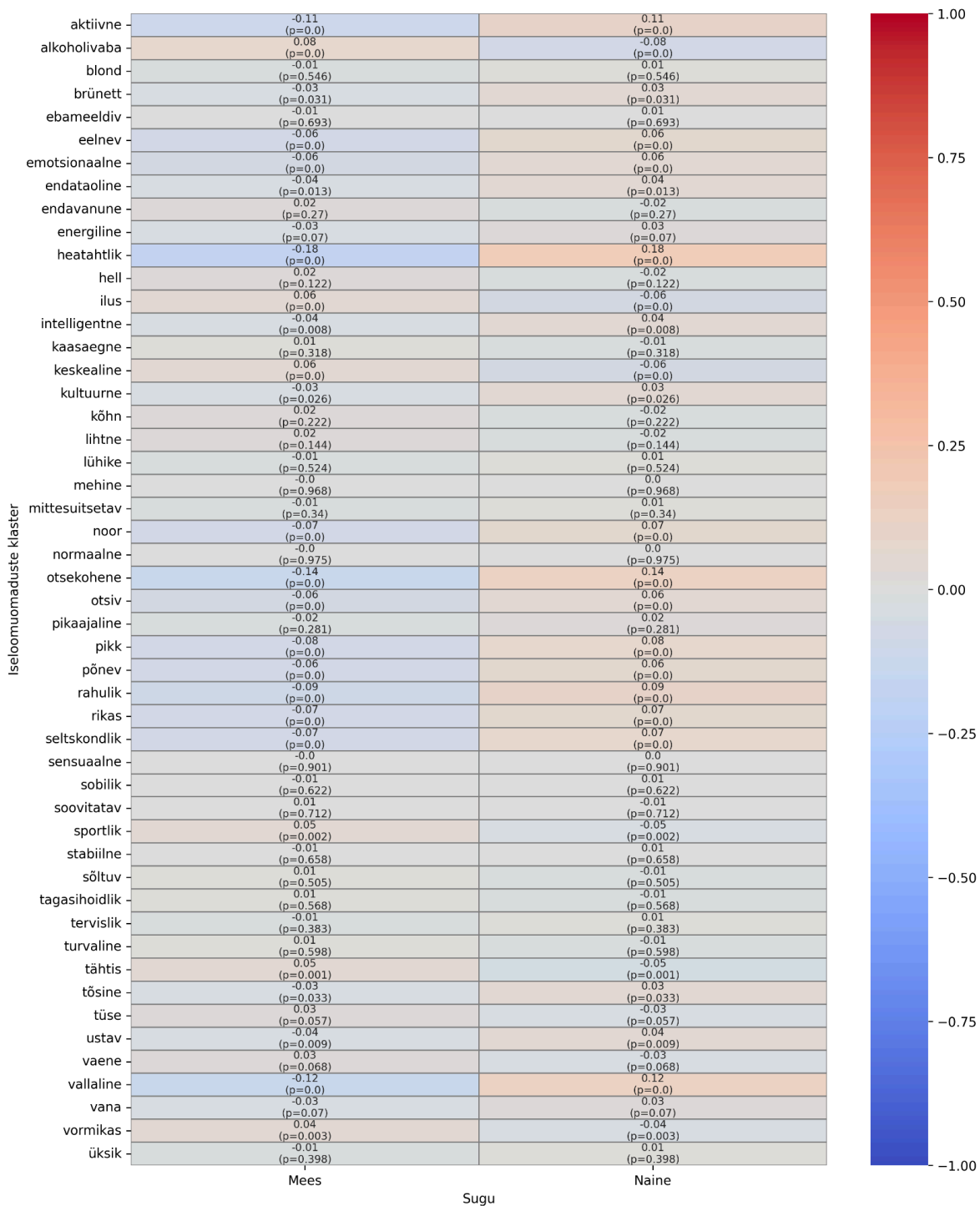
Korrelatsioon soo ja iseloomuomaduste vahel

Üldiselt näitab analüüs, et naised kirjeldavad tutvumiskuulutustes rohkem omadusi nagu heatahtlikkus, otsekohesus ja rahulikkus, samas kui mehed rõhutavad veidi rohkem füüsilisi omadusi nagu välimus (iludus ja pikkus) ja sportlikkus.

Korrelatsioonid on üldiselt mõõdukad, mis viitab sellele, et kuigi teatud mustrid on olemas, ei ole soolised erinevused iseloomuomaduste kirjeldamisel väga suured.

Kõige tugevamad omaduste korrelatsioonide erisused sugude vahel ilmnevad järgmiste omaduste puhul:

- Otsekohene (-0.14 meestele, 0.14 naistele)
- Heatahtlik (-0.18 meestele, 0.18 naistele)
- Vallaline (-0.12 meestele, 0.12 naistele)



Joonis 14. Korrelatsioonide väärtused iseloomuomaduste ja soo vahel.

Korrelatsioon otsitava soo ja iseloomuomaduste vahel

Analüüs näitab, et otsitav sugu mõjutab oluliselt seda, milliseid omadusi inimesed enda või potentsiaalse partneri kohta esile tõstavad. Meeste otsijad rõhutavad rohkem empaatiat, noorust ja füüsilist atraktiivsust, samas kui naiste otsijad hindavad enamasti praktilisemaid ja elustiiliga seotud omadusi. Vaatlesin ainult omadusi, mille korrelatsioon oli üle 0.05 ning mille p-väärtus oli alla 0.05.

Peamised mustrid meeste otsijate kirjeldustel:

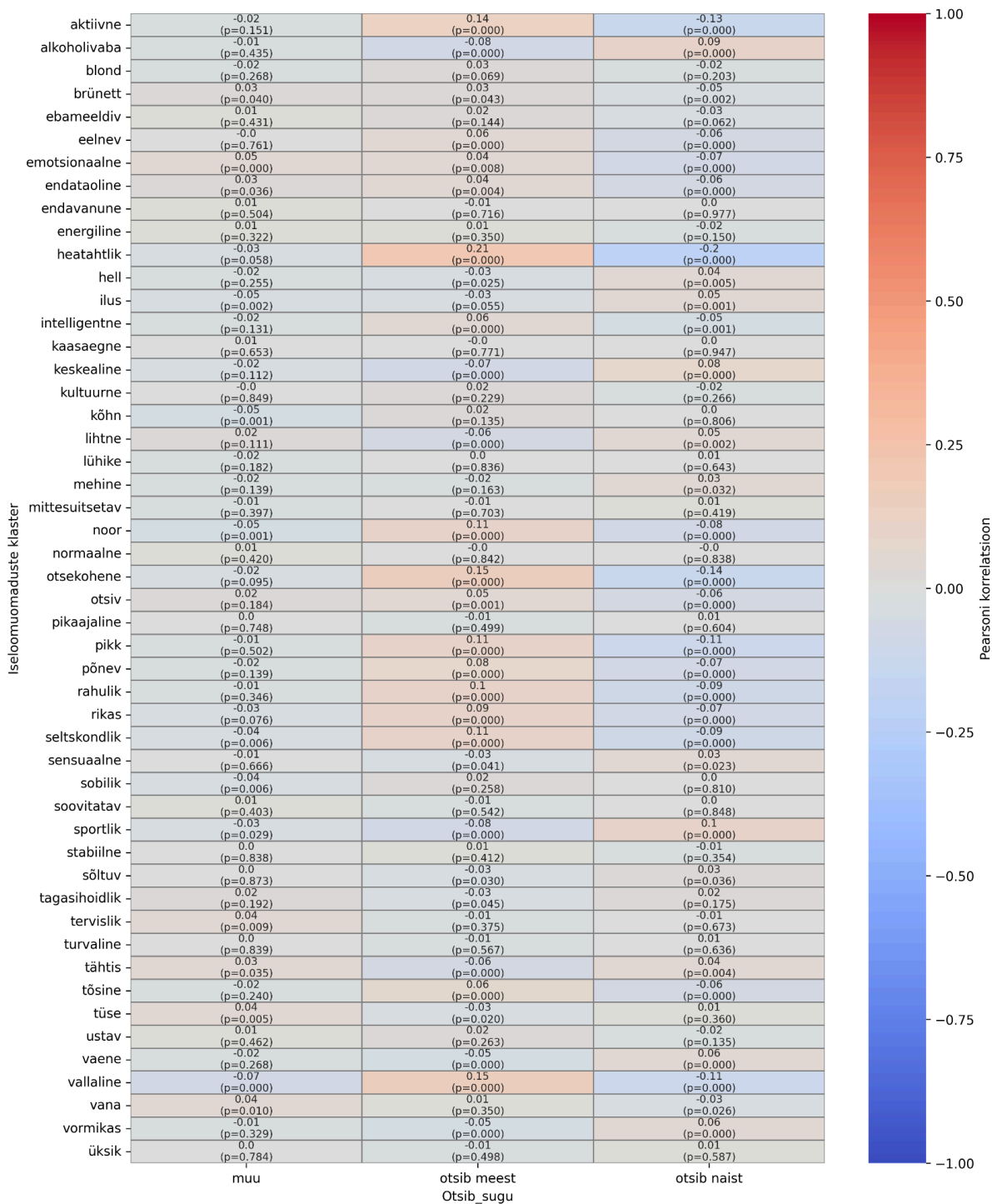
- Heatahtlikkus (0.21), vallalisus (0.15), aktiivsus (0.14) ja otsekoheesus (0.15) on omadused, mida meeste otsijad eriti esile tõstavad
- Pikkus (0.11) ja nooruslikkus (0.11) on samuti positiivselt seotud meeste otsijatega, mis peegeldab füüsiliste atraktiivsuse tähtsust

Naiste otsijate eelistused:

- Sportlikkus (0.10), mittediitsetav (0.09) on naiste otsijate enim esiletõstetumad omadused. Mittediitsetamise tähtsus naiste partnerieelistustes kattub Lance'i uurimuse [4] tulemustega, mis viitavad sarnastele eelistustele ka teistes kultuurilistes kontekstides.
- Vormikas (0.06) ja ilus (0.05) on füüsilised omadused mida enim esile tõstetakse

Muu kategooria:

- Emotsionaalsus (0.05) on positiivselt seotud nende kuulutustega, kus inimesed otsivad rohkem sõpra või suhtluspartnerit kui romantilist suhet. See võib viidata soovile leida kaaslast, kellega jagada igapäevaseid emotsioone.



Joonis 15. Korrelatsioonide väärtused iseloomuomaduste ja otsitava soo vahel.

4. Tulemuste piiravad faktorid

Uurimistöös esineb mitu tegurit, mis võivad mõjutada tulemuste usaldusväärsust ja täpsust. Andmete puhastamise protsessi võib mõjutada tekstianalüüsi ebatäielikkus, kuna kuulutuste keeleline mitmekesisus muudab sõnade tõlgendamise keerukaks. Lisaks võib API-de (nagu OpenAI GPT-3.5 Turbo) kasutamine teksti analüüsimisel tuua kaasa ebatäpsusi, kuna mudelid ei pruugi alati konteksti adekvaatselt tõlgendada, eriti kui tegemist on ajalooliste või regionaalsete keelevormidega.

Teiseks, andmestiku esindatus võib olla piiratud, kuna analüüs põhines peamiselt ühe ajalehe (Nelli Teataja) kuulutustel. Kuigi see võimaldas suurema ühtsuse ja struktureerituse, võib teiste ajalehtede puudumine muuta tulemusi kitsamaks, mis ei pruugi olla täielikult üldistatavad kogu Eesti kontekstile.

Lisaks võivad tulemusi moonutada postitajate endi tekitatud vead, näiteks kuulutuse vale rubriigi valimine. Andmestikus leidsime juhtumeid, kus mees oli oma kuulutuse paigutanud rubriiki 'NAISED' või naine rubriiki 'MEHED'. Sellised vead võivad tekkida nii postitaja eksimusest, ajalehe toimetuse segadusest kui ka digiteerimise käigus tekkinud vigadest. Kuigi sellised juhtumid olid üksikud, võivad nad statistilistes analüüsides põhjustada väikeseid kõrvalekaldeid, eriti sooliste jaotuste või partneriotsingu mustrite osas. Selliste vigade mõju vähendamiseks oleks edaspidi vaja täiendavaid kontrollimehhanisme.

Kokkuvõte

Käesolev bakalaureusetöö keskendus tutvumiskuulutuste analüüsile Eesti ajalehtedes, kasutades digitaalseid meetodeid ja kultuurianalüüsi. Töö eesmärk oli selgitada, kuidas inimesed end ja potentsiaalseid partnereid kirjeldavad, millised demograafilised ja geograafilised mustrid ilmnevad ning kuidas need peegeldavad ühiskondlikke norme. Uurimine põhines Nelli Teataja kuulutustel, mida analüüsiti masinõppe ja statistiliste meetodite abil.

Peamised tulemused näitasid, et kuulutuste postitajate hulgas domineerisid vanemad inimesed (60–80-aastased), kelle hulgas naised olid veidi aktiivsemad kui mehed. Enim esindatud piirkonnad olid Harjumaa, Pärnumaa ja Tartumaa, kus linnastumise tase on kõrge. Iseloomuomaduste hulgas esile tõsteti enim vabadust/vallalisust, heatahtlikkust ja nooruslikkust, kusjuures soolised erinevused olid märgatavad: naised rõhutasid rohkem emotsionaalseid omadusi, samas kui mehed hindasid füüsilist välimust.

Tulevikusuundumused antud valdkonnas võiksid hõlmata laiemat ajaloolist võrdlust (nt erinevate aastakümnete või -sajandite kuulutuste võrdlus), sügavamalt kvalitatiivset analüüsi keeleliste nüansside kohta ning teiste meediavormide (nagu online-tutvumisportaalide) kaasamist, et mõista muutusi partneriotsingute kultuuris. Samuti oleks kasulik uurida, kuidas sotsiaalmajanduslikud tegurid (nt haridus, elustiil) mõjutavad kuulutuste sisu, ning laiendada andmekogusid, hõlmates rohkem piirkondi ja allikaid. Sellised täiendavad uuringud aitaksid paremini mõista nii ajaloolisi kui ka kaasaegseid suhtlemisviise.

Viited

[1] Sõjajärgses Eestis otsisid mehed tutvumiskuulutuste kaudu tigeid, aga rikkaid naisi. 2021.

<https://naine.postimees.ee/7333104/ajalugu-sojajargses-eestis-otsisid-mehed-tutvumiskuulutuste-kaudu-tigeid-aga-rikkaid-naisi>

[2] Lixuan Zhang, Iryana Pentina, Wendy Fox-Kirk. Who uses mobile apps to meet strangers: The roles of core traits and surface characteristics. (2017). https://www.researchgate.net/publication/320970872_Who_uses_mobile_apps_to_meet_strangers_The_roles_of_core_traits_and_surface_characteristics

[3] Rain Pruul. RAHVUSÄÄRMUSLUS RATE.EE PROFIILIDES. (2007). <https://dspace.ut.ee/server/api/core/bitstreams/f239d57c-8c4c-498f-8753-baa77949c09b/content>

[4] Larry M. Lance. Gender Differences in Heterosexual Dating: A Content Analysis of Personal Ads. (1998). <https://journals.sagepub.com/doi/abs/10.1177/106082659800600303>

[5] Günter J. Hitsch, Ali Horta, Dan Ariely. What Makes You Click: An Empirical Analysis of Online Dating. (2005). https://home.uchicago.edu/~hortacsu/online_dating_feb2005.pdf

[6] Eesti Rahvusraamatukogu. DIGAR Eesti artiklid. <https://dea.digar.ee>

[7] Peeter Tinit. Accessing National Library texts. (2021). <https://github.com/peeter-t2/digar.txts>

[8] Kristel Uibo. Tekstikaave. (2019). <https://kristel.gitbooks.io/sissejuhatus-tekstikaavesse/content/tekstikaave.html>

[9] Daniel Nelson. Mis on NLP (loomuliku keele töötlemine)?. (2024). <https://www.unite.ai/et/what-is-natural-language-processing/>

- [10] Sirje Virkus. Infokäitumise, info hankimise ja otsingu ning infopädevuse uurimise meetodid. (2010). https://www.tlu.ee/~sirvir/Infootsingu%20teooria/Infokaitumise,%20info%20hankimise%20ja%20otsingu%20ning%20infopadevuse%20uurimise%20meetodid/lk_32.html
- [11] Hasan Tanvir, Claudia Kittask, Sandra Eiche, Kairit Sirts. EstBERT: A Pretrained Language-Specific BERT for Estonian. (2021). <https://aclanthology.org/2021.nodalida-main.2/>
- [12] Sven Laur, Siim Orasmaa, Dage Särg, Paul Tammo. EstNLTK 1.6: Remastered Estonian NLP Pipeline. (2020). <https://aclanthology.org/2020.lrec-1.884/>
- [13] Eleri Aedmaa. Pretrained word and multi-sense embeddings for Estonian. (2023). <https://github.com/eleriaedmaa/embeddings>
- [14] Eda Kavlakoglu, Vanna Winland. What is k-means clustering?. (2024). <https://www.ibm.com/think/topics/k-means-clustering>
- [15] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. (1996). <https://dl.acm.org/doi/10.5555/3001460.3001507>
- [16] Geoffrey McLachlan, David Peel. Finite Mixture Models. (2000). <https://onlinelibrary.wiley.com/doi/book/10.1002/0471721182>
- [17] Arvutimaailm. Tutvumine kolib Internetti: kolmandik Eesti noortest kasutab kohtingurakendusi. (2018). <https://www.am.ee/Tutvumine-kolib-internetti>
- [18] Siseministeerium. Eelarvejaotus. (2024). <https://siseministeerium.ee/sites/default/files/documents/2024-05/Lisa%201%20Eelarvejaotus%202024.pdf>
- [19] Eesti Statistika. Rahvaarv. (2025). <https://stat.ee/et/avasta-statistikat/valdkonnad/rahvastik/rahvaarv>

[20] Mikk Salu. Haridus ja töö tõi noored naised maalt linna. (2013).
<https://www.postimees.ee/1089994/haridus-ja-too-toi-noored-naised-maalt-linna>

Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, David Jaanson,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **Eesti tutvumiskuulutuste analüüs Nelli Teataja näitel**, mille juhendaja on Uku Kangur, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;
2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni;
3. olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile;
4. kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

David Jaanson

15.05.2025