

TARTU ÜLIKOOL  
Arvutiteaduse instituut  
Informaatika

**Kris Porovarde**

**Eesti avaandmete portaali probleemide kaardistamine  
ja andmeanalüüsi töölaua loomine**

**Bakalaureusetöö (9 EAP)**

Juhendaja: Jaak Vilo

Tartu 2024

# **Eesti avaandmete portaali probleemide kaardistamine ja andmeanalüüsi töölaua loomine**

## **Lühikokkuvõte:**

Avaandmed on andmed, mis on kõigile kasutamiseks ja jagamiseks, mida saab kasutada nii äriliste kui ka mittetulunduslikeks ettevõtmisteks. Eesti avaandmete teabevärv pakub võimalust pääseda ligi Eesti avaandmetele, et neid tarbida või visualiseerida. Töö esimene osa annab ülevaate avaandmete seisust Eestis koos pakutud andmete analüüsi vahendite kirjeldusega. Hetkel portaalis olev visualiseerimise tööriist pole ühtlane, tihti ei avane või lausa puudub valikust. Ühtlasi on see aegunud ja baseerub vananenud tehnoloogiatel. Nende probleemide lahendamiseks arendati käesolevas bakalaureusetöös graafiline kasutajaliides, millega saab avaandmeid visualiseerida ja analüüsida. Rakendus loodi kasutades tehnoloogiaid: Python, Streamlit, BeautifulSoup, PyGWalker. Lisaks kaardistati bakalaureusetöö käigus hetkeprobleemid avaandmete teabevärva portaalis ning pakuti välja nendele vastavaid soovitusi.

## **Võtmesõnad:**

Avaandmed, Eesti avaandmete teabevärv, Python, Streamlit, API

**CERCS:** P175

# **Mapping problems in the Estonian open data portal and creating a data analytics dashboard**

## **Abstract:**

Open data is data that can be used and shared by everyone for both commercial and non-commercial purposes. The Estonian Open Data Portal provides an opportunity to access Estonian Open Data to be consumed or visualized. Currently, the visualization tool available on the portal is not consistent, often doesn't open, or is missing altogether from the options. Additionally, it is outdated, and the user interface and functionalities are based on outdated technologies. To address these issues, a graphical user interface was developed in this bachelor's thesis, enabling the visualization and analysis of open data. The application was developed using technologies such as Python, Streamlit, BeautifulSoup, and PyGWalker. In addition, during the bachelor's thesis, the current problems in the open data portal were identified, and corresponding recommendations were provided.

## **Keywords:**

Open data, Estonian open data portal, Python, Streamlit, API

**CERCS:** P175

# Sisukord

<b>Sissejuhatus.....</b>	<b>4</b>
<b>1. Eesti riigi avaandmete portaali ja ökosüsteemi ülevaade.....</b>	<b>6</b>
1.1 Avaandmed.....	6
1.1.1 Avaandmete taaskasutus ja litsentsimine.....	6
1.1.2 Avaandmete võimalused ja tähtsus.....	7
1.2 Eesti avaandmete teabevärv.....	8
1.2.1 Portaali hetkeseis ja puudujäägid.....	8
1.2.2 Avaandmete portaali rakendusliides.....	9
1.3 Standardsete tööluarakenduste ülevaade.....	10
1.3.1 Põllumajanduse Registrate ja Informatsiooni Ameti statistikamoodul.....	10
1.3.2 Statistikaameti andmebaas.....	11
1.3.2 Tervisestatistika ja terviseuuringute andmebaas (TAI).....	11
1.4 Kaardistatud avaandmete portaali probleemid.....	12
<b>2. Universaalse andmeanalüüsi tööluarakenduse loomine.....</b>	<b>14</b>
2.1 Eesmärk.....	14
2.2 Kasutatud tehnoloogiad ja meetodid.....	14
2.2.1 Arenduskeel - Python.....	14
2.2.2 Andmestike kogumine - BeautifulSoup, Selenium.....	15
2.2.3 Andmete puhastamine.....	16
2.2.4 Andmete visualiseerimine.....	17
2.3 Loodud veebirakenduse kirjeldus.....	17
2.3.1 Avaandmed API ülevaade.....	17
2.3.2 Andmestike allalaadimine.....	20
2.3.3 Andmestike puhastamise meetodite rakendamine.....	21
2.3.4 Andmete visualiseerimine veebirakenduse abil.....	21
2.4 Tööluarakenduse ülevaade.....	22
<b>3. Valminud tulemuse analüüs.....</b>	<b>27</b>
3.1 Soovitused avaandmete ökosüsteemi arendamiseks.....	27
3.2 Arutelu.....	28
<b>Kokkuvõte.....</b>	<b>29</b>
<b>Viidatud kirjandus.....</b>	<b>30</b>
<b>Lisad.....</b>	<b>32</b>
Lisa 1. Lähtekood.....	32
Lisa 2. Loodud rakenduse kuvatõmmised.....	33
<b>Litsents.....</b>	<b>41</b>

# Sissejuhatus

Avaandmed on andmed, mis on kõigile kasutamiseks, taaskasutamiseks ja jagamiseks, mida saab kasutada nii äriliste kui ka mittetulunduslikeks ettevõtmisteks, uuringute läbiviimiseks ja andmepõhiste otsuste tegemiseks. Avaandmete teabevärv pakub võimalust pääseda ligi Eesti avaandmetele, et neid analüüsida või visualiseerida. Lisaks leiab teabevärvast Eesti avaandmete põhjal loodud kasutuslugusid [1].

Viimastel aastatel on avaandmete kättesaadavus märkimisväärselt paranenud, mis pakub suurt innovatsiooni- ja arengupotentsiaali erinevates sektorites. Vaatamata sellele, et avaandmete teabevärva pakub suurel hulgal andmestikke, on siiski paljud neist andmekogumitest alakasutatud või ebakvaliteetsed. See tekitab omakorda väljakutse isikule, kes püüab avaandmeid kasutada teatud eesmärgil. Seega eksisteerib vajadus vahendite ja lahenduste järele, mis võimaldavad avaandmete tõhusat uurimist, analüüsi ja visualiseerimist.

Eesti avaandmete teabevärv sisaldab suurel hulgal erinevate valdkondade andmestikke. Küll aga jääb märkimisväärne osa nendest andmekogumist kasutamata näiteks teadlikkuse puudumise või andmete kvaliteediprobleemide tõttu. Paljud andmestikud on ebakvaliteetsed, mittetäielikud või aegunud, mistõttu on need praktiliseks kasutamiseks vähem väärtuslikud. Tõhusate tööriistade puudumine nende andmekogumite visualiseerimiseks ja analüüsimiseks süvendab probleemi veelgi, takistades olemasolevate andmete läbitöötlust.

Vaatamata Eesti avaandmete teabevärvas leiduvate andmekogumite rohkusele on nende kasutamisega seotud teatud piirangud ja väljakutsed. Paljud andmekogumid omavad ebakõlasid struktuuris ja kvaliteedis, mistõttu on keeruline neid sujuvalt standardiseerida ja integreerida. Ühtlasi on andmestikud teabevärvas hoiustatud erinevates formaatides, mis muudab andmete kättesaamist keeruliseks. Lisaks, kuigi portaalis on mitmesuguseid andmekogumeid, ei pruugi kõik neist olla asjakohased või kasulikud konkreetsetel teadusuuringutel või analüütilistel eesmärkidel. Seetõttu peab arendatav lahendus neid piiranguid arvesse võtma ja töötama välja mehhanismid erinevate andmekogumite tõhusaks käsitlemiseks, tagades samal ajal visualiseerimise asjakohasust.

Nende probleemide lahendamiseks arendasime käesolevas bakalaureusetöös välja graafilise kasutajaliidese, millega saab Eesti avaandmeid visualiseerida. Lahendus hõlmab Pythoni abil andmestikke kogumist ja hoiustamist, millele järgneb andmete eeltöötluse etapid andmestike filtreerimiseks, tasakaalustamiseks ja kvaliteedi tõstmiseks. Seejärel arendasime välja kasutajasõbraliku graafilise kasutajaliidese, mis võimaldab kasutajatel andmekogumeid analüüsida ja visualiseerida. Lisaks kaardistasime probleemid avaandmete portaalis ning pakume nendele vastavaid soovitusi.

Töö on struktureeritud järgmiselt: esiteks annab töö ülevaate avaandmete seisust Eestis, teiseks kirjeldab töö loodud rakenduse disaini otsused ja funktsionaalsuse ning kolmandaks, võetakse kokku avaaandmetega seotud probleemid ning annab soovitusel, mida tulevikus järgida paremini avaandmete publitseerimisel.

Tööl on kaasas ka 2 lisa. Lisa 1 kirjeldab loodud rakenduse lähtekoodi asukohta ning saadaval oleva lehekülje veebilinki. Lisa 2 hõlmab endas loodud rakenduse kuvatõmmiseid.

# 1. Eesti riigi avaandmete portaali ja ökosüsteemi ülevaade

Esimeses peatükis anname ülevaate avaandmete olulisusest ja kasutamisest. Tutvustame avaandmete tausta, keskendudes avaandmete mõistele, kasutus võimalustele ja tähtsusele. Seejärel vaatleme Eesti avaandmete teabevärava portaali, käsitledes portaali hetkeseisu, puudujääke ning kättesaadavust. Lisaks analüüsime olemasolevaid lahendusi, milleks on Põllumajanduse Registrate ja Informatsiooni Ameti statistikamoodul ning Statistikaameti andmebaas. Peatükk annab ülevaate avaandmete valdkonnast Eestis ja võimaldab paremat arusaamist sellest, kuidas avaandmeid kasutatakse.

## 1.1 Avaandmed

Avalikult kättesaadavalt teavet, mille üldist kasutamist ei ole seadusega või mingil muul määral piiratud, nimetatakse avaandmeteks. See tähendab, et avaandmed on üldmõiste, mille all üldiselt mõistetakse avatud vormingutes olevaid andmeid, mida on lubatud kõigil igasugusel eesmärgil vabalt kasutada ja jagada, v.a isikuandmed, mis avaldatakse avaandmetena, nende puhul kehtivad isikuandmete kaitse üldmääruse nõuded. Avaliku sektori ja avalikult rahastatud avaandmete laiemat kasutamist ärilisel või mitteärilisel eesmärgil tuleb ühiskonnas julgustada, tagades minimaalsed õiguslikud, tehnilised, rahalised või üldised piirangud, eelkõige innovatsiooni ja majanduskasvu huvides [1].

### 1.1.1 Avaandmete taaskasutus ja litsentsimine

Avaandmete taaskasutamiseks on vaja kahte komponenti: andmeid ja avatust. Üldjuhul antakse avaandmed taaskasutamisele tingimusteta ehk litsentsita, kuid kui tingimuste seadmine on avalikes huvides vajalik, peavad vastavad tingimused olema objektiivsed, proportsionaalsed ja mittediskrimineerivad. Taaskasutamisele seatud tingimused, ei tohi taaskasutamist põhjendamatult piirata ega takistada konkurentsi. Tingimused on kirjas näiteks litsentsides, kuid neid võib olla palju ning need on tavaliselt pikad õiguslikud tekstid, mistõttu võib litsentside maailm ning nende vahel valimine tunduda keeruline [1].

Litsents on millegi kasutusõigus ehk üks isik annab teisele isikule kokkulepitud tingimustel õiguse millegi kasutamiseks, mis tähendab, et kui keegi tahab kellegi poolt loodut või tema

valduses olevat kasutada, peab tal selleks olema luba. Litsentside kaudu on võimalik otsest loa küsimist ning andmist vältida, näiteks võidakse litsentsiga seada piiranguid andmete kasutamisele või nõuda informatsiooni kasutamisel algallikale viitamist. Avaandmete puhul annavad litsentsid kasutajatele kindluse ning teadmise, kuidas andmeid kasutada ja jagada võib. Litsents määrab ära, millistel tingimustel, viisil ja ulatuses saavad andmete taaskasutajad, näiteks idufirmad ja üksikisikud, avaliku sektori poolt avalikustatud andmeid kasutada ja jagada. Avaandmed on küll kõigile kasutamiseks, kuid mitte kõikide avaandmete baasil võib luua kommertslahendusi [1].

### 1.1.2 Avaandmete võimalused ja tähtsus

Avaliku sektori andmete avalikustamisest saavad kasu nii avalik sektor, kodanikud, vabasektor kui ka erasektor. Avalik sektor muutub läbipaistvamaks, andmete avamine võib lihtsustada avaliku sektori asutuste omavahelist ja piiriülest koostööd. Avatud andmed võimaldavad tõsta andmete kvaliteeti ja riigi usaldusväärsust, tekivad paremad võimalused kodanike kaasamiseks. Ühtlasi soodustavad avaandmed võimalusi uuenduslike tehnoloogiliste lahenduste loomiseks ja rakendamiseks [2].

Erasektoris saab avaandmeid kasutada uute teenuste loomiseks ja olemasolevate teenuste uuendamiseks. Avaandmeid on võimalik erasektoris ühendada ja siduda oma äriiga seonduvate andmete kui ka teenustega. Näiteks kodanikule saab luua teenuseid, kus kodanik jagab teenusepakkujale tema kohta käivaid isikuandmeid ja kombineerida neid avaandmete ja erasektori andmetega [2].

Teaduslikke aluspõhimõtteid, nagu läbipaistvus, vastastikune hindamine, reprodutseeritavus ja jälgitavus, on peaaegu võimatu rakendada ilma mudelite ja andmetele juurdepääsuta. Kuna selliste põhimõtete järgimine viib kvaliteetsema teaduseni, siis on oluline, et andmed oleksid avalikud. Avalike andmete puhul on võimalik ära hoida olukordi, kus ilmneb sisse vigasid, kas autori enda poolse eksimuse tõttu või keskkonnast tulenevatest muutujatest. Lisaks on andmete kogumine ressursimahukas töö, mistõttu andmete avalikustamine võimaldab säästa aega ja raha teadlaste ja muude isikute seas, kes seetõttu saavad keskenduda uurimisküsimusele, mitte

andmete kogumisele. Kokkuvõttes andmete avalikustamine ei ole ainult kasulik algse teadlase tunnustamiseks, vaid ka teaduse terviklikuks arenguks. [3].

## 1.2 Eesti avaandmete teabevärv

Eesti avaandmete teabevärv [4] pakub nii eraisikutele kui ka äriühingutele võimaluse ligi pääseda Eestis kättesaadavatele avaandmetele, et neid visualiseerida või analüüsida. Avaandmeid jagavad enamasti avaliku sektori asutused, kuid ka ettevõtted, teadusasutused jt organisatsioonid, kes oma töö käigus koguvad ja toodavad andmeid, mis võiksid avalikkusele kasulikud olla. 2024. aasta kevade seisuga on portaalis 2232 teabevaldajat ehk kõik asutused, kes on teabevärvas andmeid jaganud ning kokku on 1836 andmestikku, mis jagunevad erinevateks valdkondadeks, näiteks tervis, keskkond, transport jt. Teabevaldajate arv on suurem kui andmekogumite, sest teabevaldajate hulka kuuluvad ka asutused, kes pole veel ühtegi andmestikku avaldanud [4].

Portaalis on andmestike alamleht, mis sisaldab tabelit, kus on saadaval kõik andmestikud. Tabelis on välja toodud andmestiku nimi, teabevaldaja, lisamise kuupäev, viimati uuendamise kuupäev ja uuenemissagedus. Andmete värskendamise sagedus sõltub andmete iseloomust ning nende avaldajast, kuid tavaliselt uuenevad selles portaalis olevad andmed aeglaselt. Andmestiku nimega on seotud hüperlink alamlehele, kus leidub lisainfot andmestiku kohta ning, kas võimalus andmestik alla laadida või viide andmekogumi algportaalile. Lisaks on mõnel puhul võimalus andmeid visualiseerida, mis võimaldab andmete paremat analüüsimist.

### 1.2.1 Portaali hetkeseis ja puudujäägid

Andmestiku alamlehel olev visualiseerimise tööriist (vt joonis 1) pole aga ühtlane, tihti see ei avane või lausa puudub valikust andmestiku juures. Isegi kui andmestiku visualiseerimine töötab, siis see on aegunud ja baseerub vananenud tehnoloogiatel. Peamiste probleemide hulka kuuluvad halb kasutajaliides, võimalus erinevaid graafikuid joonistada ja jõudlus. Mis tähendab, et märkimisväärne osa andmestikest ei leia kasutust ja analüüsimata. Ühtlasi on paljud andmestikud ebakvaliteetsed, mittetäielikud või aegunud, teisisõnu vajavad andmetöötlust, et need praktilises kasutuses väärtuslikud oleksid. Sestap on avaandmete lehel alamleht kasutuslood. Kasutuslugude lehel koondab teabevärv kokku avaandmete põhjal loodud



rakendusi. Rakenduste autoriteks on nii asutused kui ka eraisikud ning rakendusi saab lisada igaüks, kas kasutajaloomisega või teabevärava halduriga otse ühendust võttes.

Table	Keskmine määratud töövõimetoetus (eurodes)	KUU	Määramiste arv
Count			
MAAKOND (NB! Andmetes on "Kõ	MAAKOND (NB! Andmetes on "Kõ	Töövõime ulatus (NB! Andmetes	Totals
Töövõime ulatus (NB! Andmetes			
	Harju maakond	Kokku	92
		Osaline töövõime	92
		Puuduv töövõime	92
	Hiiu maakond	Kokku	92
		Osaline töövõime	92
		Puuduv töövõime	92
	Ida-Viru maakond	Kokku	92
		Osaline töövõime	92
		Puuduv töövõime	92
	Järva maakond	Kokku	92
		Osaline töövõime	92
		Puuduv töövõime	92
	Jõgeva maakond	Kokku	92
		Osaline töövõime	92
		Puuduv töövõime	92

Joonis 1. Näide töötavast visualiseerijast lehel <https://avaandmed.eesti.ee/>.

### 1.2.2 Avaandmete portaali rakendusliides

Avaandmete portaali rakendusliides ehk API (ingl *application programming interface*) annab võimaluse andmetike lisamiseks, muutmiseks, kustutamiseks ja metaandmete saamiseks. Portaali API kaudu saab üles laadida andmestikke, lisada ja muuta metaandmeid, mis tagab andmete kvaliteedi parandamise ilma inimese otsese sekkumiseta. Seega on API abil võimalik teha mitmeid erinevaid tegevusi, näiteks saada lihtsalt infot või saata ise mingi väärtus serverile. Tegevuse defineerib kasutaja poolt valitud API *endpoint*. Kasutades API-s päringut GET on võimalik saada infot portaali lisatud andmestike kohta ning olenevalt parameetritest, kas ühe kindla andmestiku või kõikide andmestike kohta. Andmestiku hindamiseks on olemas päring POST. Kuna see päring muudab serveris andmeid, siis tuleb anda päringu kehana kaasa info, mida ja kuidas serveris muuta. Portaali lisatud andmestiku seisundi muutmiseks on eraldi päring PUT, mis eeldab, et kasutaja on autentitud. DELETE päringu abil on võimalik autentitud kasutajal või organisatsioonil enda poolt lisatud andmestikke kustutada [5].

```

"data": [
  {
    "id": "ef10fd46-502e-4f74-a1dc-4afbf3e42930",
    "nameEt": "INSPIRE (HB) - Eesti vääriselupaigad (WFS)",
    "nameEn": "INSPIRE (HB) - Eesti vääriselupaigad (WFS)",
    "slug": "inspire-(hb)-eesti-vaariselupaigad-(wfs)",
    "descriptionEt": "Maa-ameti poolt pakutav INSPIRE allalaadimisteenus, mille kaudu tehakse kättesaadavaks järgmine andmekiht: vääriselupaigad.",
    "descriptionEn": "Maa-ameti poolt pakutav INSPIRE allalaadimisteenus, mille kaudu tehakse kättesaadavaks järgmine andmekiht: vääriselupaigad.",
    "toBeDeletedAt": null,
    "createdAt": "2024-03-25T02:01:37.577Z",
    "updatedAt": "2024-03-25T02:01:37.631Z",
    "user": null,
    "organization": {
      "id": "df05fbfc-2811-4088-b9db-e901b0514709",
      "regCode": "70003098",
      "name": "Maa-amet",
      "slug": "maa-amet",
      "contact": "Maa-amet",
      "contactEmail": "maamet@maamet.ee",
      "description": "Maa-amet on Eesti riigi kompetentsikeskus maaküsimustes ning suurim ruumiandmete tootja, haldaja, levitaja ja ruumiandmeteenuste pakkuja.",
      "isPublicBody": true,
      "notifications": []
    },
    "name": "INSPIRE (HB) - Eesti vääriselupaigad (WFS)"
  }
],

```

Joonis 2. Näide API päring GET <https://avaandmed.eesti.ee/api/datasets/> vastusest.

## 1.3 Standardsete töölaarakenduste ülevaade

### 1.3.1 Põllumajanduse Registrate ja Informatsiooni Ameti statistikamoodul

Põllumajanduse Registrate ja Informatsiooni Ametil (PRIA) on aastate jooksul kogunenud hulgaliselt avaandmeid, mida sooviti jagada nii klientide kui ka koostööpartneritega, sellepärast arendati välja statistikamoodul. PRIA poolt arendatud statistikamoodul osutus 2023. aastal Majandus- ja Kommunikatsiooniministeeriumi poolt välja antud parima avaandmete visualiseerija tiitli võitjaks. Statistikamoodul pakub infot maaeluga seotud valdkondade kohta ning selle peamine eesmärk on kajastada andmestikke võimalikult lihtsalt, visualiseeritult ning automatiseeritult. Tegu on pidevalt areneva mooduliga, mis toetab maal elavaid inimesi oma ettevõtmiste planeerimisel ning aitab tutvustada PRIA tegevust ja kogutavat informatsiooni erinevatele sihtgruppidele [6].

Statistikamoodulist leiab statistikat PRIA poolt menetletud toetuste, põllumajandusloomade registris olevate loomade ning PRIA avalike andmete kohta. Moodulis leiab statistikat erinevate valdkonda kohta nagu näiteks põllud, loomad, kalandus ja muu. Andmed on esitatud taotleja tegevuskoha aadressi järgi ning kuvamisel on andmed eelmise päeva seisuga. Andmeid uuendatakse iga tööpäeva hommikuks hiljemalt kell 9.00. Statistikamoodul pakub võimalust

filtreerida andmeid maakondade, liikide jt parameetrite põhjal. Moodul on üles ehitatud kasutades *Power BI* tarkvara, mille abil kuvatakse erinevaid andmevälju [7].

*Power BI* on Microsofti poolt välja töötatud tarkvarateenuste ja rakenduste kogum, mis töötavad koos, et muuta töötlemata andmestikud sidusateks, visuaalselt kaasahaaravateks ja interaktiivseteks ülevaadeteks. Andmed võivad olla Exceli tabelarvutuse kujul või pilvepõhiste andmeladude kogumitena. *Power BI* ühendub mugavalt andmestikega, et neid visualiseerides kuvada [8].

### 1.3.2 Statistikaameti andmebaas

Eesti Statistikaamet on loonud statistika andmebaasi, mis hõlmab endas kõigile kättesaadavaid andmeid. Andmestikud jagunevad erinevateks kategooriateks nagu keskkond, majandus, rahvastik jm. Valides statistika andmebaasis kõigepealt endale sobiv andmestik, avaneb vaade, kus on täpsem ülevaade andmestikus leiduva informatsiooni kohta. Statistika andmebaas võimaldab kasutajal valida antud andmestikust muutujad, mida soovitakse tabelis kujutada. Seejärel, kui kasutaja on valinud muutujad, ilmneb tabel, kus on valitud andmed kuvatud. Kasutaja saab ise valida, millist tabelit kuvatakse, näiteks kas sorditud tabel, püsttulp joonis vm. Küll aga see eeldab, et andmed on sobilikud vastavale kujule, vastasel juhul tuleb veateade, et on valitud liiga palju või liiga vähe väärtusi [9].

Statistika andmebaas võimaldab ka andmete modifitseerimist ning tabeli metaandmete ümberkirjutamist, täpsemalt joonise pealkirja muutmist, sildi paigutust ja legendi modifitseerimist. Peale vajaliku tabeli saavutamist on võimalik jooniseid salvestada ja alla laadida, endale sobilikus formaadis, kas PX, XML vm. Lisaks on ka võimalik teha API päring tabeli kohta, et unikaalset tabelit enda rakendusega ühendada. Portaali on loonud Rootsi Statistikaameti arendatud PxWeb platvormi baasil [9].

### 1.3.2 Tervisestatistika ja terviseuuringute andmebaas (TAI)

Tervisestatistika ja terviseuuringute andmebaas on Eesti suurim tervisega seotud avalik statistika kogu, mida haldab Tervise Arengu Instituut. Andmebaas koosneb põhivaldkondadest, mis jagunevad alamvaldkondadeks. Alamvaldkondades olevatel andmetabelitel on unikaalne kood,

mis kergendab andmete otsimist. Andmebaasis olevaid andmekogumeid on võimalik nii visualiseerida netis, kui ka alla laadida erinevates failivormingutes, näiteks .px, .xlsx, .csv jt. Andmebaasis olevad andmestikud saavad sageli uuendusi, järgides avaldamiskalendri graafikut, mis on saadaval nende veebilehel. Iga andmestiku viimane uuendamise kuupäev on välja toodud andmetabeli loetelus pealkirja järel [10].

Andmebaasi visualiseerimise tööriist on arendatud tarkvaraga PX-Web (versioon 1.0). PX-Web on Rootsi Statistikaameti poolt välja töötatud lahendus dünaamiliste tabelite moodustamiseks px-failide põhjal [10]. Statistilisi andmeid, mida soovitakse levitada, salvestatakse PX-failis, mis asub kohalikus veebiserveris või SQL-serveri andmebaasis, millel on ühine Põhjamaade metamudel. PX-Web on võimeline kombineerima kahte tüüpi andmeallikaid üheks andmestikuks. PX-Web rakendus koosneb kahest osast :

1. Administreerimisliides - seda kasutavad PX-Webi administraatorid oma rakenduse haldamiseks ja hooldamiseks.
2. Kasutajaliides - Rakenduse osa, millega lõppkasutaja suhtleb, mis kuvab graafikud.

PX-Web võimaldab kasutajal valida andmebaasi, tabeli, muutuja ja väärtuste vahel, et saavutada soovitud vaade [11].

## 1.4 Kaardistatud avaandmete portaali probleemid

Lahenduskäigu jooksul, uurides ning analüüsides erinevaid andmestikke avaandmete portaalis, jõuti järelduseni, et portaalis võiks nii mõnigi asi olla teisiti. Alljärgnev on loetelu avaandmete portaalis kaardistatud probleemidest, mis hõlmab nii portaali tehnilist poolt, kui ka andmekogumite üldisi omadusi, mis on saadaval avaandmete portaalis.

1. **Visualiseerimise võimalus** - avaandmete portaalis kasutusel olev visualiseerimise võimalus on mittetäielik. Visualiseerija paljudel juhtudel ei tööta, pole valikus või kui avaneb, siis on limiteeritud. Kasutajakogemus pole kõige parem ning olemasolevad funktsionaalsused pole piisavad, et viia läbi tõhus andmete analüüs.
2. **Andmestike kättesaadavus** - avaandmete API analüüsi põhjal leiti, et avaandmete API kaudu on kättesaadavad, vaid 3,1% andmestikest. Avaandmete kasulikkus tuleb välja

analüüsides ning visualiseerides, mis eelkõige eeldab andmestikud on kättesaadavad, seega leiti, et tegu on hetkel probleemse kohaga.

3. **Andmestike sisu erikujud** - avaandmete analüüsimiseks on sageli vaja ühendada mitu andmekogumit, mis tihtipeale ei pruugi olla väljastatud sama organisatsiooni poolt. Mistõttu on oluline, et andmestikud on sisult samasuguse ülesehitusega ning ühtlase kujuga. Paraku esineb juhtumeid, kus on näiteks kuupäevad, maakonnad või aastad erikujul, mis tähendab, et andmekogumite ühendamist ei saa teostada ning sügavam analüüs jääb teostamata.
4. **Andmete kasulikkus, kvaliteediprobleemid** - andmete kasulikuks visualiseerimiseks ja analüüsiks on oluline, et vaadeldavad andmestikud ei omaks suuri kvaliteediprobleeme. Avaandmete portaalis leidub palju andmestikke, mis on tasakaalustamata või mittetäielikud. Ühtlasi kuna portaalis on kättesaadaval palju andmestikke ja mida aeg edasi seda rohkem neid sinna lisandub, siis on oluline mõelda välja, kuidas filtreerida välja aegunud ja kasutatud andmekogumid.
5. **Failide kodeering** - avaandmete portaalis kättesaadavad failid on paljudes erinevates kodeeringutes, näiteks UTF-8 ja iso-8859-13. Seega tekib olukordi kus faile sisselugedes, pole kuvatud terviklik või korrektne informatsioon.
6. **Faili formaadid** - avaandmete portaalis on kokku üle 14 erineva failivormingu nagu näiteks XML, JSON, CSV jne.

## 2. Universaalse andmeanalüüsi töölaudarakenduse loomine

Käesolev peatükk keskendub kasutatud tehnoloogiatele, lahenduskäigu kirjeldusele ning töölaudarakenduse ülevaatele. Esimeses osas antakse ülevaade rakenduse eesmärgist ning seejärel tutvustatakse kasutatud tehnoloogiaid ja meetodeid, mis valiti lõputöö raames rakenduse loomiseks. Peatüki teises pooles käsitletakse rakenduse lahenduskäiku, kirjeldades avaandmete API ülevaadet, andmetike allalaadimist, puhastamist ja visualiseerimist veebirakenduse abil. Seejärel antakse detailne ülevaade välja töötatud rakendusest ning selle põhjal tehtud analüüsist. Viimases alamosas tutvustatakse valminud rakendust, tuues esile selle olulisemad funktsionaalsused ning võimekused.

### 2.1 Eesmärk

Käesoleva bakalaureusetöö käigus arendas töö autor välja graafilise kasutajaliidese, millega saab valitud andmestiku andmeid visualiseerida ja analüüsida. Pakkudes nii standardset ülevaadet andmestikust kui ka võimalust muuta andmestikku ja ühendada teise andmestikuga, et luua unikaalseid graafikuid vastavalt olukorrale. Lahenduskäigu jooksul võetakse arvesse eelmises peatükis välja toodud avaandmete portaali kaardistatud probleeme.

### 2.2 Kasutatud tehnoloogiad ja meetodid

#### 2.2.1 Arenduskeel - Python

Lahenduse loomisel kasutatakse nii ees- kui ka tagapoolse (ingl *front-end*, *back-end*) süsteemi loomiseks programmeerimiskeelt Python. Põhjuseks see, et Pythonis saadaval olevate teekide (ingl *library*) abil on hea andmeid koguda, töödelda ning visualiseerida.

Viimase 20 aastaga on Python tõusnud keskpärasest teadusliku andmetöötluse keelest üheks kõige olulisemaks keeleks andmeteaduse, masinõppe ja üldise tarkvaraarenduse jaoks akadeemilistes ringkondades ja tööstuses. Andmete analüüsiks, interaktiivseks andmetöötluseks ning andmete visualiseerimiseks saab tuua välja paralleelse teiste avatud lähtekoodi programmeerimiskeeltega nagu R, MATLAB, SAS jt. Viimastel aastatel on Pythoni täiustatud

avatud lähtekoodi teekidega, näiteks pandas ja scikit-learn, mis on muutnud selle populaarseks valikuks andmeanalüüsi ülesannete jaoks. Koos Pythoni üldise tugevusega tarkvaratehnika valdkonnas on see suurepärane valik peamise keelena andmeanalüüsi rakenduse loomisel [12].

### 2.2.2 Andmestike kogumine - BeautifulSoup, Selenium

Andmestike kogumiseks kasutatakse meetodit *Web scraping* (inglise keelest tõlgitud veebikoorimine). *Web scraping* on protsess, mille käigus eraldatakse ja kombineeritakse veebist huviorbiidil olevaid andmeid. Antud protsessis imiteerib tarkvaraagent, mida tuntakse ka kui netirobot (ingl *Web robot*), veebiserverite ja inimese vahelist interaktsiooni tavapärase veebi külastamise käigus. Samm-sammult saavutab robot juurdepääsu nii vajaminevatele veebilehtedele, analüüsib nende sisu huvipakkuvate andmete leidmiseks, kättesaamiseks ning kogub andmeid vastavalt eesmärgile [13].

Järgnev lõik tugineb Glez-Peña jt artiklile [13], mille kohaselt *Web scraping* API-d ja raamistikud käsitlevad laialt levinud ülesandeid, millega veebiandmete kaabitsad tegelevad, eesmärgiga saavutada spetsiifilisi otsingu eesmärgi. Esmaseks eesmärgiks on tagada juurdepääs veebilehele. Selle jaoks loob *Web scraper* ühenduse vastava veebisaidiga HTTP-protokolli (ingl *Hypertext Transfer Protocol*) kaudu. HTTP-s on olekuta tekstipõhine Interneti protokoll, mis koordineerib päringu-vastuse operatsioone kliendi ja veebiserveri vahel. HTTP-s kasutatakse kõige enam päringumeetodeid GET, mille abil saavutatakse ressursi päringuid, ja POST, mida kasutatakse vormide täitmiseks ja failide üleslaadimiseks. Kasutajaagent on oluline päringu päises, sest selle põhjal otsustab server, milline programm sisule juurde pääseb. Lisaks peab netirobot järgima veebilehe n-õ kasutustingimusi, mis on kirjeldatud veebisaidi failis “robots.txt”. Selles failis on igal veebiserveril vastavalt kokkuleppele info netirobotitele, milles veebisaidi omanik kirjeldab, millistele ressurssidele võib või ei tohi automaatsete protsessidega ligi pääseda. Pärast HTML dokumendi hankimist võib *Web scraper* eraldada huvipakkuva sisu, kasutades regulaaravaldisi üksinda või koos lisaloogikaga. Alternatiivse lahendusena on olemas ka HTML-i parsimise raamistikud, mis töötavad veebilehtede DOM (ingl *Document Object Model*) struktuuri põhjal. Kui veebilehelt on sisu kätte saadud, siis viimaseks etapiks on andmete struktureerimine vastavalt, et oleks võimalik läbi viia andmete analüüs või andmete hoiustamine.

Veebikoorimine saavutatakse kasutades Pythoni teeki BeautifulSoup, Selenium ja Requests. BeautifulSoup võimaldab andmete kättesaamist HTML ja XML failidest [14]. Selenium pakub lahendust jäljendada kasutaja ja veebilehe vahelist suhtlust [15]. Pythoni teegi Requests abil on lihtne saata HTTP/1.1 päringuid, ilma lisaandmete või parameetrite lisamiseta [16].

### 2.2.3 Andmete puhastamine

Andmete puhastamine (ingl *data cleaning*) hõlmab endas vigaste andmete leidmist, eemaldamist ja ebaühtlaste andmete tuvastamist, et parandada andmekogumi kvaliteeti. Andmete kvaliteedi probleemid esinevad üksikutes andmekogudes nagu failides ja andmebaasides. Peamised põhjused, miks sellised probleemid tekivad, on puuduvad andmed, väärandmed või andmete sisestamisel tehtud kirjavead [17].

Enne kui asuda andmete puhastamise kallale, on tarvilik läbi viia esmane andmete analüüs, et leida erinevad vead ja tuvastada ebakõlad, mis vajavad eemaldamist. Lisaks andmete käsitsi kontrollimisele tuleb kasutada ka analüüsiprogramme, et saada metaandmeid andmete omaduste kohta ning ühtlasi tuvastada erinevaid probleeme [17]. Kui probleemid on kaardistatud, saab andmekogumite parandamiseks kasutada erinevaid meetodeid.

Käesoleva lõputöö raames kasutatakse andmete puhastamiseks meetodeid duplikaatide eemaldamine ja *Accusation*. Duplikaat väärtuste puhul tuleb leida nende duplikaat väärtused andmestikust ning seejärel need eemaldada, et saavutada parem jõudlus. Dasari ja Varma artikli [18] alusel on Accusation meetod, mida kasutatakse puuduvate väärtuste puhul ning sellel on neli erinevat viisi, kuidas seda teostada.

- Esimene võimalus on eemaldada andmekogumist read, kus esineb puuduvaid väärtusi.
- Teine lähenemine seisneb sarnaste väärtuste leidmises samast andmekogumist ning nende kasutamises puuduvate väärtuste asendamiseks.
- Kolmas meetod hõlmab puuduvatele väärtustele uute väärtuste määramist, kasutades keskmise või mediaani arvutamise meetodeid.
- Neljas lähenemine on käsitleda puuduvaid atribuute sõltuvate atribuutidena ja seejärel regressiooni abil puuduv atribuut leida.



## 2.2.4 Andmete visualiseerimine

Andmete visualiseerimiseks luuakse antud lahenduse raames veebirakendus, kasutades Pythoni teeki Streamlit ja PyGWalker. Streamlit on avatud lähtekoodiga Pythoni teek, mis võimaldab luua ja juurutada (ingl *deploy*) veebirakendusi masinõppe ja andmeteaduse jaoks. Streamliti on sisse ehitud erinevad meetodid, mille abil on võimalik visualiseerida andmestike ja muuta neid vastavalt kasutaja olukorrale [19].

PyGWalker on Pythoni teek, mis võimaldab luua erinevaid jooniseid ja diagramme vaid mõne lihtsa nupuvajutusega, ilma et oleks vaja kodeerimisoskusi. Selle intuitiivne kasutajaliides ja mitmekülgne visualiseerimisvõimaluste valik koos võimalusega integreerida Streamlitiga muudavad selle interaktiivseks tööriistaks andmete visualiseerimisel ja jagamisel [20].

## 2.3 Loodud veebirakenduse kirjeldus

### 2.3.1 Avaandmed API ülevaade

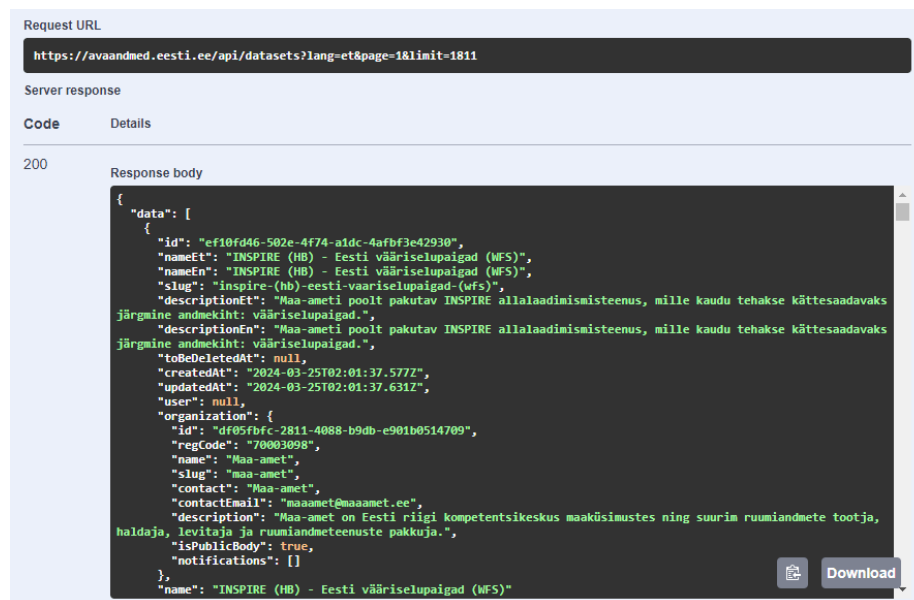
Tööprotsessi alustati eesmärgiga saavutada ülevaade avaandmete portaalis olevatest andmestikkude kättesaadavusest. Täpsemalt uuriti avaandmete API kaudu, mitu andmestikku eksisteerib, kui paljud neist on kättesaadavad ning kes on peamised panustaja avaandmete üleslaadimisel.

Kõigepealt küsisime avaandmete portaali API käest avaldatud andmestikke arvu (vt joonis 3). Tulemuseks saadi, et avaandmete andmebaasis on 1811 andmestikku.



Joonis 3. Näide API päring GET <https://avaandmed.eesti.ee/api/datasets/total>.

Järgmise sammuna päriti API käest infot kõigi 1811 andmestiku kohta (vt joonis 4). Päringu vastuseks oli JSON (ingl *JavaScript Object Notation*) tüüpi objekt, mille sisuks oli massiiv andmestiku objektidest. Iga andmestiku kohta oli saadaval näiteks väli id, nimi, organisatsioon jt. Kasutusele võeti ainult andmestiku nimi ja id.



Joonis 4. Näide päringust <https://avaandmed.eesti.ee/api/datasets?lang=et&page=1&limit=1811>.

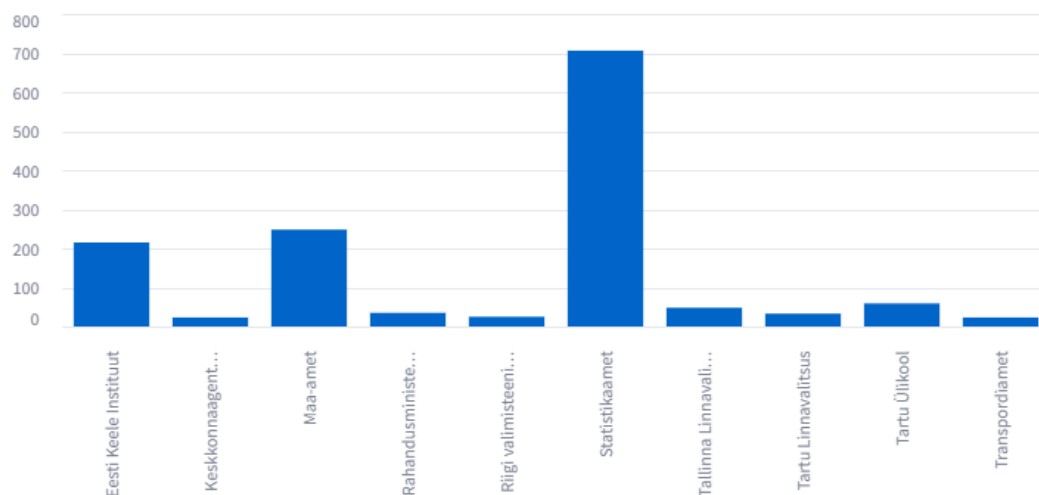
Andmestiku id abil tehti uus API päring (vt näidet joonisel 5), mis tagastas kõik olemasoleva info iga individuaalse andmestiku kohta. Seejärel filtreeriti välja vajaminevad väljad, milleks olid

“id”, “name”, “url”, “description”, “categories”, “files”, “published at”, “updated at” ja “organization”. Peale seda lisati iga andmestiku filtreeritud info kokku üheks andmestikuks, et selle põhjal teostada detailsem ülevaade avaandmete portaali andmebaasist.

```
{
  "data": {
    "id": "ef10fd46-502e-4f74-a1dc-4afb3e42930",
    "status": "completed",
    "nameEt": "INSPIRE (HB) - Eesti vääriselupaigad (WFS)",
    "nameEn": "INSPIRE (HB) - Eesti vääriselupaigad (WFS)",
    "slug": "inspire-(hb)-eesti-vaariselupaigad-(wfs)",
    "descriptionEt": "Maa-ameti poolt pakutav INSPIRE allalaadimismisteenus, mille kaudu tehakse kättesaadavaks järgmine andmekiht: vääriselupaigad.",
    "descriptionEn": "Maa-ameti poolt pakutav INSPIRE allalaadimismisteenus, mille kaudu tehakse kättesaadavaks järgmine andmekiht: vääriselupaigad.",
    "maintainer": "Maa-amet",
    "maintainerEmail": "inspire@maaamet.ee",
    "maintainerPhone": null,
    "citations": [
      {
        "url": "https://inspire.geoportaal.ee/geoserver/HB_public/wfs?request=GetCapabilities&service=wfs&version=2.0.0",
        "name": "WFS teenuse GetCapabilities",
        "urlBroken": false
      }
    ],
    "conformities": [
      {
        "releaseDate": "2008-12-04T00:00:00",
        "specification": "Komisjoni määrus (EÜ) nr 1205/2008, 3. detsember 2008, millega rakendatakse Euroopa Parlamendi ja nõukogu direktiivi 2007/2/EÜ seoses metaandmetega",
        "specificationEn": "Komisjoni määrus (EÜ) nr 1205/2008, 3. detsember 2008, millega rakendatakse Euroopa Parlamendi ja nõukogu direktiivi 2007/2/EÜ seoses metaandmetega"
      },
      {
        "releaseDate": "2009-10-20T00:00:00",
        "specification": "Komisjoni määrus (EÜ) nr 976/2009, 19. oktoober 2009, millega rakendatakse Euroopa Parlamendi ja nõukogu direktiivi 2007/2/EÜ seoses võrguteenustega",
        "specificationEn": "Komisjoni määrus (EÜ) nr 976/2009, 19. oktoober 2009, millega rakendatakse Euroopa Parlamendi ja nõukogu direktiivi 2007/2/EÜ seoses võrguteenustega"
      }
    ],
    "southLatitude": "57.51",
    "northLatitude": "59.68",
    "westLongitude": "21.77",
    "eastLongitude": "28.21",
    "url": null,
    "language": "et",
    "dataFrom": "2020-02-03T00:00:00.000Z",
    "dataTo": null,
    "updateIntervalUnit": "unknown",
    "updateIntervalFrequency": 0,
    "access": "public",
    "availableTo": null,
    "organizationId": "df05fbfc-2811-4088-b9db-e901b0514709",
    "userId": null,
    "landingPage": null,
    "qualifiedAttribution": null,
    "wasGeneratedBy": null,
    "wasGeneratedByEn": null,
    "spatialResolution": null,
    "geoportalIdentifier": "d19c96f2-cfdf-4652-98b1-5f0e115dc477",
    "geoportalKeywords": "infoFeatureAccessService, elupaigad, biotoobid, vääriselupaik, INSPIRE, wfs",
    "geoportalKeywordsEn": null,
    "lineage": null,
    "pixelSize": null,
    "resourceType": "service",
    "topicCategories": [
      {
        "spatialRepresentationType": null,
        "spatialDataServiceType": "download",
        "temporalResolution": null,
        "maturity": null,
        "version": 0,
        "versionNotes": null,
        "isActual": true,
        "toBeDeletedAt": null,
        "publishedAt": "2024-03-25T02:01:37.631Z",
        "translatedAt": null,
        "createdAt": "2024-03-25T02:01:37.577Z",
        "updatedAt": "2024-03-25T02:01:37.631Z",
        "licence": {
          "id": 1,
          "name": "Creative Commons CC0 1.0 Üldine"
        },
        "description": "õigus vabalt edasi arendada, muuta, muudesse töödesse inkorporeerida, taaskasutada ja jagada mistahes eesmärgil",
        "user": null,
        "organization": {
          "id": "df05fbfc-2811-4088-b9db-e901b0514709",
          "regCode": "70003098",
          "name": "Maa-amet",
          "slug": "maa-amet",
          "contact": "Maa-amet",
          "contactEmail": "maaamet@maaamet.ee",
          "description": "Maa-amet on Eesti riigi kompetentsikeskus maaküsimustes ning suurim ruumandmete tootja, haldaja, levitaja ja ruumandmete teenuste pakkuja.",
          "isPublicBody": true,
          "notifications": [],
          "coordinateReferenceSystems": [],
          "keywords": [
            {
              "id": 4283,
              "name": "ruumandmed",
              "language": "et",
              "keywordId": 4283,
              "keywordEmsCategory": {
                "id": 6474,
                "emsCategoryId": 47,
                "keywordId": 4283
              }
            }
          ],
          "categories": [
            {
              "id": 47,
              "name": "GEOGRAAFIA. GEODEESIA. KARTOGRAAFIA"
            }
          ],
          "regions": [
            {
              "id": 17,
              "name": "Muu",
              "coordinates": null
            }
          ],
          "files": [],
          "parentDatasets": [],
          "childDatasets": [],
          "mapRegions": [
            {
              "id": 17,
              "name": "Muu"
            }
          ],
          "isContentAllowed": true,
          "name": "INSPIRE (HB) - Eesti vääriselupaigad (WFS)"
        }
      }
    ]
  }
}
```

Joonis 5. Näide päringust <https://avaandmed.eesti.ee/api/datasets/{id}>.

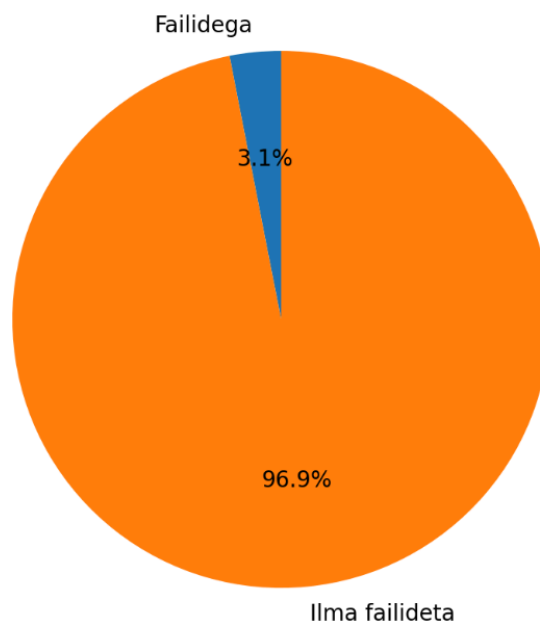
Avaandmete portaali andmebaasi ülevaate loomisel leiti, et suurimad panustajad ehk organisatsioonid, kes on avaldanud kõige rohkem andmestikke avaandmete portaali, on Statistikaamet, Maa-amet ja Eesti Keele Instituut (vt joonis 6). Teine märkimisväärne avastus oli, et kui vaadata kõiki 1818 andmestikku, siis ainult 56. andmestikul oli väli “files”, mille abil on võimalik avaandmete API kaudu andmestikku alla laadida. Seega on avaandmete API kaudu kättesaadavad vaid 3,1% andmestikest (vt joonis 7). Siinkohal tasub ära mainida, et API päringuid teostati eri aegadel, millest tulenevalt on lõpptulemusel välja toodud statistika 1818 andmestiku kohta ja mitte 1811 nagu mainiti peatüki alguses.



Joonis 6. Avaandmete portaali suurimad panustajad.

Andmestike hulk, millel on väli failidega: 56

Andmestike hulk, mille failide väli on tühi: 1762



Joonis 7. Sektordiagramm, mis näitab välja “files” olemasolu andmestikes.

### 2.3.2 Andmestike allalaadimine

Andmestike kogumisel polnud tingimata eesmärgiks analüüsida kõiki andmestikke, mis olid saadaval, sest autoril pole ressursse, et hallata ja testida üle 1800 andmestiku. Töö skoobis

piirdusime eesmärgiga koguda umbes 10 andmestikku, mis on pigem sarnastest valdkondadest, et neid võrrelda ja analüüsida. Veebikoorimise käigus kogusime andmestikud avaandmete portaali alamlehel, kus olid ainult failiformaadi CSV failid. Tulemus saavutati kasutades Pythoni teeki BeautifulSoup, Selenium ja Requests. Selenium jäljendas kasutaja käitumist Eesti avaandmete portaali veebilehel, navigeerides seal iga andmestiku alamlehele. BeautifulSoup teeki abil saadi kätte andmed vastavalt lehtedelt, antud olukorras lingid kust on võimalik andmestikke alla laadida ning teeki Requests abil tõmmati failid alla. Lisaks failide alla laadimise käiguse rakendati funktsioon, millega muudeti iga faili kodeering UTF-8 kodeeringuks, et hiljem andmete sisse lugemisel ja kuvamisel ei tekiks ebakõlasid.

### 2.3.3 Andmestike puhastamise meetodite rakendamine

Peale andmete kogumist oli järgmiseks etapiks andmestike puhastamine ja korrigeerimine enne kui neid kujutatakse või visualiseeritakse. Uurides andmestike põgusalt selgus, et peamiseks probleemiks on read või veerud, mis sisaldavad tühjasid (ingl *null*) väärtusi. Kuna praeguse bakalaureusetöö raames ei olnud võimalik luua kõikidele andmestikele imputeerimise lahendusi või minna detailidesse, miks on andmed puudu, otsustasime selle probleemi esmaseks lahendamiseks eemaldada andmestikust veerud, mis sisaldasid üle 70% tühjasid väärtusi. Tulevikus tasub detailsemalt uurida puuduvate andmete põhjuseid ja vaadata, milliseid neist on piisavalt lihtne kas imputeerida või aktsepteerida hilisemates analüüsides andmete puudumist.

### 2.3.4 Andmete visualiseerimine veebirakenduse abil

Andmestike puhastamisele järgnes veebirakenduse loomine, mille abil saab andmeid visualiseerida. Veebirakenduse jaoks kasutas autor Pythoni teeki Streamlit ja PyGWalker. Rakenduse loomisel oli peamine rõhk suunatud kasutajakogemuse optimeerimisele, teisisõnu keskenduti sellele, et kasutajaliides oleks interaktiivne ning kasutajal oleks võimalus valida, millist visualiseeringut soovib teostada.

Streamliti abil valmis rakenduse põhiosa. Rakendus loeb sisse andmestikud, pakub võimalust kasutajal valida, millist andmestikku soovib kuvada ning seejärel kuvab valitud andmestiku. Andmestiku visualiseerimiseks kasutati Streamliti sisse ehitatud funktsioone nagu “.write()” ja “.bar\_chart()”. Funktsioon “.write()” võimaldab kuvada nii tekstikujul kui ka tabelina

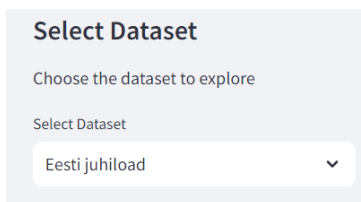
informatsiooni, antud juhul andmestike kirjeldamiseks. Meetodi “.bar\_chart()” kaudu teostati võimalus kasutajal visualiseerida andmestikus leiduvaid väärtusi tulpdiagrammi viisil.

Järgmiseks lisasime veebirakendusesse PyGWalker komponenti, mis võimaldab kasutajal ise muuta nii valitud andmestikus olevaid andmeid, kui ka kujutada andmeid erinevatel joonistel või graafikutel. PyGWalker komponent on tööriist, mille abil on kasutajale vabad käed. Kasutaja saab ise langetada otsuseid vastavalt sellele, millist tulemust või visualiseeringut ta soovib saavutada.

Lisaks sisaldab veebirakendus ka graafikute salvestamist, olgu selleks siis kasutaja enda loodud visualiseering või rakenduse standardvaade. See võimalus lisab kasutajale paindlikkust ning võimaldab talletada olulisi andmeid või tulemusi hilisemaks analüüsiks või jagamiseks..

## 2.4 Töölauarakenduse ülevaade

Valminud rakendus on disainitud eesmärgil pakkuda kasutajale mugavat ja lihtsat interaktsiooni. Neid aspekte silmas pidades on rakendus loodetavalt oma ülesehituselt lihtne ja loogiline. Avades kasutajaliidese saab kasutaja valida rippmenüüst omale sobiva andmestiku (vt joonis 8). Seejärel saab kasutaja valida, milliseid veerge antud andmestikust soovib visualiseerida (vt joonis 9). Kui kasutaja on valinud nii andmestiku ja veerud, mida soovib kuvada, avaneb andmestik ekraanile. Kuvatud on andmestik terviklikult minimaliseeritud tabeli kujul, kui kasutaja soovib saab avada ka uues aknas täissuuruses. Ühtlasi on kõik veerud sorteeritavad nii kasvavas kui ka kahanevas järjestuses. Lisaks on esitatud ka tabel, kus leiab andmestiku üldistavaid mõõtmeid, näiteks ridade arv, keskmine, suurim väärtus jm (vt joonis 10 ja 11).



Joonis 8. Näide andmestiku valimisest rakenduses.

## Eesti juhiloa Dataset

Select Columns to Show

Loa number x

Loa tüüp x

Kehtiv alates x

Kehtiv kuni x

Kategooriad x

Sünniaasta x

Sugu x

Taotluse büroo x

Väljastaja büroo x

Kiirluba x

Joonis 9. Näide valitud andmestiku veergude valimisest.

Selected Columns

	Loa number	Loa tüüp	Kehtiv alates	Kehtiv kuni	Kategooriad	Sünniaasta	Sugu	Taotluse büroo	Väljastaja büroo	Kiirluba
0	EV4	JUHILUBA	2019-02	2029-02	B B1 C1	1,977	M	E-TEENUS	POSTIGA	E
1	EV1	JUHILUBA	2016-03	2026-03	A A1 B B1 C C1	1,955	M	E-TEENUS	POSTIGA	E
2	EV8	JUHILUBA	2023-09	2033-09	AM B B1 C C1	1,973	M	E-TEENUS	POSTIGA	E
3	EV4	JUHILUBA	2019-09	2029-09	B B1 C C1	1,957	M	Tartu	Tartu	E
4	ET9	JUHILUBA	2014-12	2024-12	B B1	1,940	N	Saue	Saue	E
5	EV4	JUHILUBA	2019-01	2029-01	B B1	1,980	N	E-TEENUS	POSTIGA	E
6	EV7	JUHILUBA	2022-01	2032-01	AM B B1	1,999	N	E-TEENUS	POSTIGA	E
7	EV8	ESMANE	2023-08	2025-08	AM B B1	2,005	N	E-TEENUS	POSTIGA	E
8	EV0	JUHILUBA	2015-07	2025-07	B B1	1,976	N	E-TEENUS	POSTIGA	E
9	EV0	JUHILUBA	2015-07	2025-07	A A1 A2 B B1	1,988	M	E-TEENUS	POSTIGA	E

Joonis 10. Näide andmestiku minimaliseeritud tabeli kujust.

Data Summary

	Loa number	Loa tüüp	Kehtiv alates	Kehtiv kuni	Kategooriad	Sünniaasta	Sugu	Taotluse büroo	Väljastaja büroo	Kiirluba
count	682230	682230	682230	682230	682230	682,230	682230	682230	682230	682230
missing	0	0	0	0	0	0	0	0	0	0
unique	13	2	120	121	597	None	3	21	20	3
sum	None	None	None	None	None	347,906,519	None	None	None	None
top	EV8	JUHILUBA	2022-08	2025-03	B B1	None	M	E-TEENUS	POSTIGA	E
freq	93178	653181	8050	8493	282815	None	381750	472038	467752	672853
mean	None	None	None	None	None	1,975.7362	None	None	None	None
std	None	None	None	None	None	15.835	None	None	None	None
min	None	None	None	None	None	1,921	None	None	None	None
25%	None	None	None	None	None	1,964	None	None	None	None

Joonis 11. Näide andmestiku ülevaatest.

Tihti peale ei piisa ainult algandmetest ning tekib vajadus modifitseerida ning lisada uusi veerge. Seega on rakenduses kasutajal ka võimalik muuta andmestikku. Kasutaja saab valida, millist funktsiooni soovib kasutada, näiteks summa või suhe. Seejärel valib vastava veeru või veerud, millel seda rakendada ning uue veeru nime. Eelmainitud valikute põhjal luuakse ja salvestatakse andmestikku uus veerg (vt joonis 12).

### Modify the dataset and Create a custom Chart

Select Modification Function  
Aggregate

Select Columns to Aggregate  
Choose an option

Enter Name for Aggregated Column  
Total

Aggregate Data

Joonis 12. Näide andmestikku muutmisest.

Ühtlasi on võimalus kasutajal ühendada valitud andmestik mingi muu teise andmestikuga, valides vastavad veerud, mida on võimalik ühendada ja ühildamismeetod, mille abil soovitakse ühendada. Kui eelmainitud valikud on korrektsed ning on võimalik kahte andmestikku omavahel ühendada, siis teostatakse süsteemi poolt kahe andmestiku ühendamise ja kuvatakse see ekraanile. Seejärel on kasutajal võimalus analüüsida ning valida, kas salvestada ühendatud andmestik uueks hetke andmestikuks, et seda saaks visualiseerida (vt joonis 13).

### Merge Datasets

#### First/current dataset

Select Column to Merge On

group

#### Second dataset

Select Second Dataset

Test2

Select Column from Second Dataset to Merge On

group

#### Select Join Type

Options:

- ☒ inner  
☐ outer  
☐ left  
☐ right

Merge Datasets

Set Merged as  
Current Dataset

Datasets successfully merged!

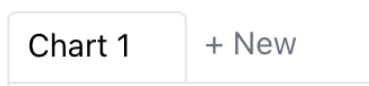
	employee	group	skills
0	Bob	Accounting	math
1	Bob	Accounting	spreadsheets
2	Jake	Engineering	coding
3	Jake	Engineering	linux
4	Lisa	Engineering	coding
5	Lisa	Engineering	linux
6	Sue	HR	spreadsheets
7	Sue	HR	organization

Joonis 13. Näide kahe andmestiku ühendamisest ja selle väljundist.



Järgnev alapeatüki osa põhineb Kanaries dokumendile [21], mis annab ülevaate andmestiku visualiseerimiseks kasutatud Pythoni teegi PyGwalkeri kohta. PyGWalker komponent võimaldab kasutajal ise luua, muuta ja uurida diagramme, samaaegselt tehese muudatusi nii andmestikule kui ka graafikutele.

Iga diagrammi haldab komponendi ülaosas asuv vahemenüü, mille abil saab luua uue diagrammi, vajutades nupule “New” (vt joonis 14).









Joonis 14. Näide uue diagrammi loomise nupust.

Tööriistariba asub komponendi ülaosas nagu näha on joonisel 15.



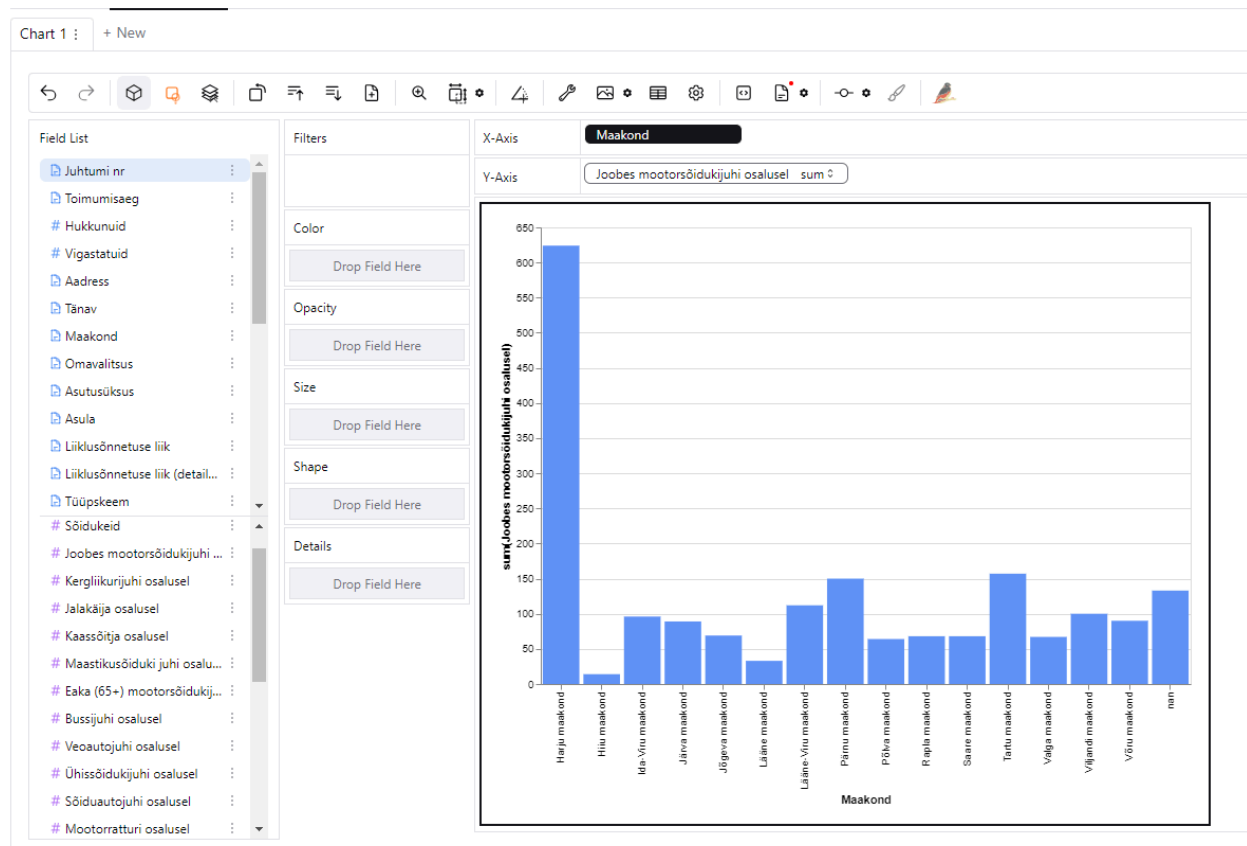
Joonis 15. PyGWalker tööriistariba.

Tööriistariba põhifunktsionaalsused on kujutatud tabelis 1.

Undo/Redo	Mark Type	Sort Order	Layout Mode	Exploration Mode	Export
					
Võta operatsioon tagasi või muuda seda.	Valige erinevate diagrammide vahel	Sorteeri kasvavalt või kahanevalt	Diagrammi suuruse muutmine	Uurige andmeid, kas punktirežiim või pintsliirežiim	Ekspordi diagramm kas PNG- või SVG-failina

Tabel 1. Tööriistariba ikoonide selgitused

Graafiku loomiseks tuleb kõigepealt valida rippmenüüst “Mark Type” vastav diagrammitüüp, näiteks tulpdiagramm, joondigramm või sektordiagramm. Järgmiseks tuleb valida vasakpoolsest külghüppmenüüst sobivad veerud, mida soovitakse kuvada. Ning seejärel lohistada need vastavalt, kas x-telje või y-telje sisendiks (vt joonis 16).



Joonis 16. Näide PyGWalkeri abil loodud tulpdiagrammist.

### 3. Valminud tulemuse analüüs

Töö viimases peatükis keskendume valminud tulemuse analüüsile, pakume soovitusi ja edasiseid võimalusi avaandmete ökosüsteemi arendamiseks. Esmalt anname soovitused Eesti avaandmete portaali haldajatele, mis käsitlevad töö jooksul portaalis kaardistatud probleeme. Seejärel süveneme arutelluse rakenduse võimalustest ja piirangutest, analüüsides hetke puudujääke ja võimalusi. Peatükk võimaldab sügavamat mõistmist saavutatud tulemuste praktilisest väärtusest ning nende rakendamise perspektiividest tulevikus.

#### 3.1 Soovitused avaandmete ökosüsteemi arendamiseks

Lõputöö raames kaardistatud probleemide põhjal loodi andmeanalüüsi töölauearakendus, mis parandas andmete visualiseerimist ja lahendas praegusi probleeme Eesti avaandmete portaalis. Siiski ei ole võimalik kõiki probleeme lahendada ainult ühe rakenduse abil, sest paljud probleemid tulenevad andmestike loomise etapist. Teisisõnu probleemiks pole tingimata avaandmete keskkond või tööriist, vaid andmestike ebaühtlus ning kvaliteet.

Andmestike visualiseerimine ja analüüs võiksid oluliselt paraneda, kui kehtestataks universaalsed formaadid teatud tunnustele ja nende väärtustele. Kõige suuremat kasu saab andmestikest siis, kui seda analüüsida koos mõne teise sarnase valdkonna andmestikuga, selleks andmestikke sobivalt ühendades. Andmestike ühendamiseks on head valikud üldised tunnused näiteks kuupäevad, maakonnad, asutused jms. Andmekogumite ühendamine eeldab, et väärtused ning tunnused on samal kujul, et ei tekiks ebakõlasid uues andmestikus. Seega Eesti avaandmete portaalis saadavate andmestike tõhusamaks muutmine eeldab universaalsete formaatide ja väärtuste ning klassifikaatorite kehtestamist. Tuleb paika panna soovituslikud kujud, mida kõik portaali panustajad järgima peaksid.

Ühtlasi on probleemiks see, et andmestikud on erinevates kodeeringutes, mis raskendab nende ühtlustamist ja analüüsimist. Erinevad kodeeringud tekitavad raskusi andmestike ühildamisel ja kogumisel, sest tuleb tegeleda täiendava kodeerimise muutmisega. Andmestikud erinevates kodeeringutes nõuavad lisatööd, et muuta need ühtseks ja analüüsitavaks, mis on kriitiline samm andmete tõhusaks kasutamiseks.

Lisaks sellele on tõsine probleem ka andmestikke kättesaadavusega. Eelmises peatükis välja toodud avaandmete API analüüsi käigus selgus, et avaandmete API kaudu on kättesaadavad, vaid 3,1% andmestikest. Kuna avaandmete kasulikkus ilmneb analüüsi ja visualiseerimise kaudu on vajalik, et andmestikud oleksid saadaval.

## 3.2 Arutelu

Peatükis 2.2.4 kirjeldati ja toodi näiteid töölaarakenduse hetke võimekustest, mis on saadaval kasutamiseks. Üheks andmeanalüüsi töölaarakenduse funktsionaalsuseks on kahe erineva andmestiku ühendamine valitud ühildamismeetodi abil. Hetkel eeldab see funktsionaalsus, et kasutaja on teinud analüüsi andmekogumitele ning teab, milliseid andmekogumeid ja kuidas ta neid ühendada soovib. Seda protsessi võiks aga automatiseerida, ehk rakendus võiks ise soovitada valitud andmestikule, milliste veergudega seda ühendada saab, tehes seda läbi tunnuste ja väärtuste võrdluse. Sellist automaatset funktsionaalsust ei jõudnud autor süsteemis implementeerida ning seega jääb järgmiseks sammuks.

Teiseks võimalikuks järgmiseks sammuks rakenduse lokaliseerimine ehk tõlkimine teise keelde. Teatavasti ei kasuta Eesti avaandmete portaali ainult eestlased, seega töölaarakenduse lokaliseerimine potentsiaalselt laiendaks rakenduse kasutajaskonda. Lokaliseerimine aitaks ka paremini mõista andmestikke erinevates geograafilistes piirkondades, kus keelebarjäär võib olla takistuseks andmete tõhusale kasutamisele ja analüüsile. Rakenduse mitmekeelsus suurendaks rakenduse universaalsust ja kasutajasõbralikkust, muutes selle atraktiivsemaks laiemale kasutajaskonnale.

# Kokkuvõte

Töö esimeses peatükis kirjeldati Eesti riigi avaandmete portaali ja ökosüsteemi hetkeseisu, käsitleti portaali puudujääke ja kättesaadavust ning ühtlasi analüüsiti olemasolevaid standardseid töölaarakendusi. Teises peatükis anti ülevaade kasutatud tehnoloogiatest, lahenduskäigu kirjeldusest ning valminud töölaarakenduse ülevaatest. Kolmandas peatükis toodi välja soovitused ja edasised võimalused avaandmete ökosüsteemi arendamiseks ning analüüsiti töö raames valminud tulemust.

Bakalaureusetöö rakenduslikuks eesmärgiks oli luua universaalne töölaarakenduse prototüüp, millega avaandmeid visualiseerida ja analüüsida. Ühtlasi oli eesmärgiks kaardistada hetkeprobleemid Eesti avaandmete teabevärava portaalis, millele pakkuda välja vastavad soovitused, et arendada Eesti riigi avaandmete ökosüsteemi.

Käesoleva bakalaureusetöö käigus arendas töö autor välja graafilise kasutajaliidese, millega saab valitud andmestiku andmeid visualiseerida ja analüüsida. Pakkudes nii standardset ülevaadet andmestikust kui ka võimalust teostada erinevaid modifitseerivaid funktsioone andmestikul, milleks on veergude lisamine ja teise andmestikuga ühendamine, et luua unikaalseid graafikuid. Lisaks bakalaureusetöö käigus kaardistas autor Eesti riigi avaandmete portaali hetkeprobleemid ning pakkus välja vastavad soovitused.

Loodud graafilise kasutajaliidese edasiarendusena on vaja täiustada kahe andmestiku ühendamise funktsionaalsust, arendades edasi selle funktsiooni võimekust pakkuda loogilisi andmestiku ühenduskohti, et saavutada parem kasutajakogemus. Samuti on võimalikuks sammuks lisada graafilisele kasutajaliidesele lokaliseerimine, mis potentsiaalselt laiendaks rakenduse kasutajaskonda.

# Viidatud kirjandus

- [1] Eesti avaandmete teabevärv, <https://avaandmed.eesti.ee/instructions/open-data-licensing-guide> (21.03.2024)
- [2] Vabariigi Valitsus, "Eesti avaliku teabe masinloetav avalikustamise roheline raamat", <https://joinup.ec.europa.eu/sites/default/files/document/2014-04/Estonian%20Open%20Data%20Greenbook.pdf> (21.03.2024)
- [3] Pfenninger S, DeCarolis J, Hirth L, Quoilin S, Staffell I, "The importance of open data and software: Is energy research lagging behind?." Energy Policy 101 (2017): 211-215. <https://www.sciencedirect.com/science/article/pii/S0301421516306516#s0010> (21.03.2024)
- [4] Eesti avaandmete teabevärv, <https://avaandmed.eesti.ee/> (03.03.2024)
- [5] Majandus- ja Kommunikatsiooniministeerium, Digiriigi arengu osakond, "Inimkeeles API võimalustest avaandmete portaali API näitel", August 2021, <https://www.dropbox.com/s/y4gpr9eo40pn764/API%20juhend.pdf?dl=0> (26.03.2024)
- [6] Põllumajanduse Registrate ja Informatsiooni Amet, <https://www.pria.ee/uudised/pria-statistikamoodul-palvis-parima-avaandmete-visualiseerija-titli> (20.03.2024)
- [7] Põllumajanduse Registrate ja Informatsiooni Amet, <https://www.pria.ee/infokeskus/statistika> (20.03.2024)
- [8] Microsoft, <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview> (20.03.2024)
- [9] Statistikaamet, <https://andmed.stat.ee/et> (21.03.2024)
- [10] Tervisestatistika ja terviseuuringute andmebaas, <https://statistika.tai.ee/Resources/Info/TSTUA.html> (01.04.2024)
- [11] Statistics Sweden, "PX-Web system description", <https://www.scb.se/globalassets/vara-tjanster/px-programmen/systemdescription-pxweb.pdf> (01.04.2024)
- [12] McKinney W, Python for data analysis, " O'Reilly Media, Inc."; 2022 Aug 12. [https://books.google.ee/books?hl=en&lr=&id=EAKBEAAAQBAJ&oi=fnd&pg=PP30&dq=python+for+data+analysis&ots=PJUsNsah50&sig=\\_n-WenvZKXYwYnAOw5HUv76qkFY&redir\\_e=sc=y#v=onepage&q&f=false](https://books.google.ee/books?hl=en&lr=&id=EAKBEAAAQBAJ&oi=fnd&pg=PP30&dq=python+for+data+analysis&ots=PJUsNsah50&sig=_n-WenvZKXYwYnAOw5HUv76qkFY&redir_e=sc=y#v=onepage&q&f=false) (01.04.2024)

- [13] Glez-Peña D., Lourenço A., López-Fernández H, Reboiro-Jato M., Fdez-Riverola F., “Web scraping technologies in an API world”, Briefings in Bioinformatics, Volume 15, Issue 5, September 2014, Pages 788–797. <https://doi.org/10.1093/bib/bbt026> (04.12.2023)
- [14] Beautiful Soup Documentation, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (10.04.2024)
- [15] Selenium, Documentation v4.0, “The Selenium Browser Automation Project”, <https://www.selenium.dev/documentation/> (10.04.2024)
- [16] Requests: HTTP for Humans, <https://requests.readthedocs.io/en/latest/> (10.04.2024)
- [17] Rahm E., Do H. H., “Data Cleaning: Problems and Current Approaches”, University of Leipzig, Germany.  
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.8661&rep=rep1&type=pdf> (04.12.2023)
- [18] Dasari D., Varma P. S., "Employing Various Data Cleaning Techniques to Achieve Better Data Quality using Python," 2022 6th International Conference on Electronics, 2022, pp. 1379-1383. <https://ieeexplore-ieee-org.ezproxy.utlib.ut.ee/document/10009079> (04.12.2023)
- [19] Streamlit, “Streamlit documentation”, <https://docs.streamlit.io/> (20.04.2024)
- [20] Kanaries, “How to use PyGWalker with Streamlit”,  
<https://docs.kanaries.net/pygwalker/use-pygwalker-with-streamlit> (18.04.2024)
- [21] Kanaries, “Create Data Visualizations with Graphic Walker”,  
<https://docs.kanaries.net/graphic-walker/data-viz/create-data-viz> (18.04.2024)

# Lisad

## Lisa 1. Lähtekood ja rakenduse saadavus

Valminud lahenduse lähtekood on saadaval Githubis aadressil:

- [https://github.com/krisporovarde/avaandmed\\_dashboard](https://github.com/krisporovarde/avaandmed_dashboard)

Loodud graafiline kasutajaliides on saadaval leheküljel:

- <https://avaandmeddashboard-rpjpjvhl6eqt8gwc4cjhsd.streamlit.app/>



Dashboard

- ▶ Estonian Open Data Portal overview
- ▶ Example of a dataset usage

Choose Dataset for exploration

Dataset:

Eesti juhiload

Advanced options:

☐ Add Column

☐ Merge Datasets

☐ Pivot Table

Eesti juhiload Dataset

Select Columns to Show

Loa number

Loa tüüp

Kehitiv alates

Kehitiv kuni

Kategooriad

Sünilaasta

Sugu

Kehitiv büroo

Väljastaja büroo

Kirjutaja

Data Summary

	Loa number	Loa tüüp	Kehitiv alates	Kehitiv kuni	Kategooriad	Sünilaasta	Sugu	Taotluse bün
count	682230	682230	682230	682230	682230	682,230	682230	682230
missing	0	0	0	0	0	0	0	0
unique	13	2	120	121	597	None	3	21
sum	None	None	None	None	None	347,906,519	None	None
top	EV8	JUHILUBA	2022-08	2025-03	BIB1	None	M	E-TEENUS
freq	93178	653181	8050	8493	282815	None	381750	472038
mean	None	None	None	None	None	1,975,762	None	None
std	None	None	None	None	None	15,835	None	None
min	None	None	None	None	None	1,921	None	None
25%	None	None	None	None	None	1,964	None	None

Create a custom chart

Manage app

## Create a custom chart

A screenshot of the Estonian Open Data Portal dashboard. The top navigation bar is grey with the word 'Dashboard' in white. Below it, there are two red triangle icons followed by the text 'Estonian Open Data Portal overview' and 'Example of a dataset usage'. The main content area has a light grey background. On the left, there's a section titled 'Choose Dataset for exploration' in bold black text. Below this title, it says 'Dataset:' followed by a white button with the text 'Eesti juhtlaid' and a right-pointing chevron. To the right of this, under the heading 'Advanced options:', there are three options, each with an unchecked checkbox: 'Add Column', 'Merge Datasets', and 'Pivot Table'.

[illegible]

35

Dashboard

▶ Estonian Open Data Portal overview

▶ Example of a dataset Usage

Choose Dataset for exploration

Dataset:

Eesti juhiload

Advanced options:

☐ Add Column

☒ Merge Datasets

☐ Pivot Table

Share

☆

🔍

⌂

📄

Merge Datasets

First/current dataset

Select Column to Merge On

Loa number

Second dataset

Select Second Dataset

Värbamisfaasis kilinillised uuringud Eestis

Select Column from Second Dataset to Merge On

EU CT nr

Select Join Type

Options:

☒ inner

☐ outer

☐ left

☐ right

Merge Datasets

Set Merged as Current Dataset

<

Manage app

Dashboard

▶ Estonian Open Data Portal overview

▶ Example of a dataset usage

Choose Dataset for exploration

Dataset:

Eesti juhiload

Advanced options:

☐ Add Column

☐ Merge Datasets

☒ Pivot Table

Share ☆ ⋮



## Create Pivot Tables

Select Index Column for Pivot Table

Loa number

Select Column to Display

Loa number

Select Aggregation Function

sum

Generate Pivot Table

< Manage app

×

Dashboard

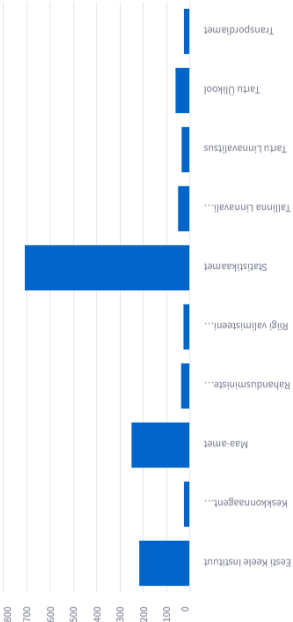
▲ Estonian Open Data Portal overview

▲ Example of a dataset usage

Share ☆ ⋮

3	5041ab02-7ae9-4eb2-a30b-24ee628ef0e6	Eesti topograafia andmekogu tuletsikhid (WMS)
4	32bbc19f-f3c4-4610-b469-d577277ee596	Nõlvakallete andmed
5	59b50945-9e99-496d-a3f5-d65722a071b	Masinõppe tuletsikhid (WFS)
6	7da99900-5c52-4ffe-bb9c-e99ffc5d444	Masinõppe tuletsikhid (WMS)
7	5b4fec2-cc5b-4a71-b63d-4e8d8a5e7b0c	Ajalooliste kaartide rakenduse aluskaartide metainfo (WMS)
8	a4720379-b97e-4456-d150-95ce08b6b241	Ajaloolised linnaplaanid (WMS)
9	1a19332e-393b-4fc5-b9e9-da8ffcf6683	Kaardivõrgud ja kaardilehtede jaotused (WMS)

### Top 10 Contributors



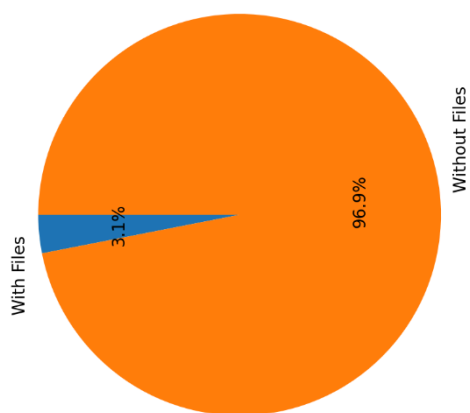
< Manage app

Dashboard

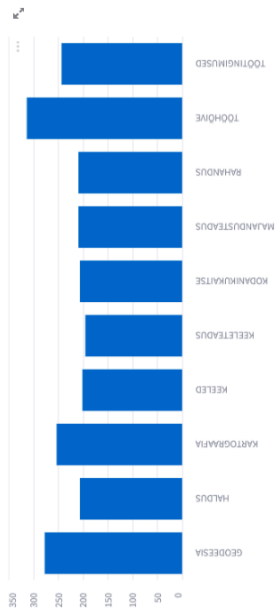
▲ Estonian Open Data Portal overview

▲ Example of a dataset usage

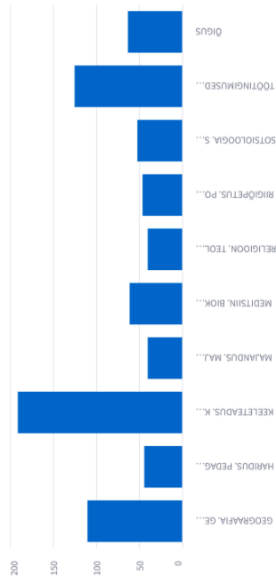
## Files Availability



TOP 10 Categories by individual category



TOP 10 Categories by a group category





# Litsents

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Kris Porovarde,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

### **Eesti avaandmete portaali probleemide kaardistamine ja andmeanalüüsi töölaua loomine**

mille juhendaja on Jaak Vilo,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace 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.

*Kris Porovarde*

**13.05.2024**