

Tartu Ülikool

Arvutiteaduse instituut

Informaatika õppekava

Rene Aleksander Truuts

**Ülevaade Microsoft Power BI ja PostgreSQL
integratsioonist ja õppematerjali koostamine**

Bakalaureusetöö (9 EAP)

Juhendaja: MSc Vambola Leping

Tartu 2024

Ülevaade Microsoft Power BI ja PostgreSQL integratsioonist ja õppematerjali koostamine

Lühikokkuvõte:

Käesolev bakalaureusetöö tutvustab Microsoft Power BI ja PostgreSQL, eesmärgiga pakkuda Tartu Ülikooli tudengitele ja teistele eestikeelseid (alg)teadmisi nii andmebaasidest kui ka andmebaasist andmete visualiseerimise võimalusest. Peamine eesmärk on arendada põhjalikke õppematerjale, mis illustreerivad Microsoft Power BI analüütilisi võimalusi ja PostgreSQLi tugevaid andmebaasihalduse funktsionaalsusi. Töö on jaotatud teoreetiliseks uuringuks ja praktiliseks osaks. Teoreetiline osa vaatab üle olemasoleva kirjanduse ja tehnoloogia, samas kui praktiline osa hõlmab SQL skriptide loomist andmete käsitlemiseks PostgreSQLis ja visualiseerimiseks Power BI-s. Töö käigus loodud õppematerjal koondati kokku mitmele töölehele. Need materjalid on mõeldud õppijate abistamiseks teoreetiliste teadmiste praktilisel rakendamisel. Arendatud õppesisu on eeldatavasti tõenäoliselt kasulik ka tulevastes ülikooli kursustes, mis seonduvad andmebaaside õpetamisega.

Märksõnad:

Andmete visualiseerimine, PostgreSQL, Microsoft Power BI, õppematerjalid, SQL, integratsioon.

CERCS:

P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Overlook of Microsoft Power BI and PostgreSQL integration and creating a study material

Abstract:

This thesis explores the integration of Microsoft Power BI and PostgreSQL, aiming to enhance the learning resources available for University of Tartu students. The primary goal is to develop comprehensive educational materials that illustrate the practical applications and synergies between Microsoft Power BI's analytical capabilities and PostgreSQL's robust database management functionalities. The thesis is structured into a theoretical study and a practical component. The theoretical part surveys the existing literature and technology frameworks, while

the practical aspect involves creating detailed guides and exercises for data manipulation and visualization using SQL scripts in PostgreSQL and Power BI. These materials are designed to assist learners in effectively synthesizing theoretical knowledge through hands-on tasks, reflecting the integration's potential to improve data processing and decision-making processes. The educational content developed is anticipated to serve future courses, enhancing students' skills in database management and data visualization critically needed in the data-driven industries.

Keywords:

Data Visualization, PostgreSQL, Microsoft Power BI, Educational Material, SQL, Integration.

CERCS:

P170 Computer science, numerical analysis, systems, control

Sisukord

Sisukord	4
Sissejuhatus	5
Probleemi tutvustus	5
Bakalaureusetöö eesmärk ja teema olulisus	5
Metoodika	6
Teoreetiline uuring	6
Praktiline osa	6
SQL skriptid PostgreSQLis	6
Andmete visualiseerimine Power BI-s	7
1. Andmete visualiseerimine kui tööriist	8
1.1 Andmete visualiseerimise ajaloost	8
1.2 Miks andmeid visualiseerida?	9
1.3 Suurandmete visualiseerimise areng alates 1940. aastatest	10
1.4 Enne visualiseerimist: Andmete töötlemine	12
2. Lühikävaade Microsoft Power BI-st ja andmebaasidest	13
2.1. Kävaade andmebaaside loomise variantidest	13
2.2. SQL andmebaaside eelised Microsoft Exceli ees	14
2.3. Kävaade PostgreSQLis	14
2.4. Kävaade Microsoft Power BI-st	15
3. Töölehe koostamine ja tagasisideanalüüs	17
3.1 Kasutatud tööriistad ja ülesanded	17
3.2 Ülesannete struktuur	17
3.4 Hindamisskeem	18
3.5 Ülesanded	18
3.6 Tagasiside	20
3.7 Kokkuvõte	23
Viidatud kirjandus	24
Lisa	27
1. Materjalid	27
2. Tagasisideküsitlus	28
3. Litsents	31

Sissejuhatus

Probleemi tutvustus

Autor on tööalaselt kokku puutunud paljude erinevate andmetöötlus- ja visualiseerimisprogrammidega, millest mitmed on Microsoft Power BI (vabavara Microsofti litsentsi omajatele) sarnased. Töö kirjutamise ajal ei leidnud autor eestikeelset õppematerjali, mis tutvustaks PostgreSQL ja Microsoft Power BI integratsioonivõimalusi. Leidub üksikuid tasulisi võimalusi, kus tutvustatakse Power BI-d, kuid spetsiifiliselt integratsiooni kohta käivaid materjale ei ole kõigile kättesaadavalt. Samuti pole Tartu Ülikoolis vastavaid õppematerjale Power BI ja PostgreSQL ühistest kasutusvõimalustest. Töö lugemisel ja ülesannete sooritamisel tulevad kasuks algteadmised andmebaasidest ja tabelarvutustest.

Eesti kontekstis on selliste teadmiste levitamine eriti oluline, kuna paljud organisatsioonid ja haridusasutused võivad veel mitte olla teadlikud nende tööriistade potentsiaalst või kuidas neid oma andmetöötlusprotsessides maksimaalselt ära kasutada. Antud töö soovib pakkuda neile vajalikke ressursse ja teadmisi, et nad saaksid teha paremaid otsuseid ning parandada oma andmetöötluse tulemuslikkust.

Bakalaureusetöö eesmärk ja teema olulisus

Bakalaureusetöö eesmärgiks on tutvustada lugejale andmete visualiseerimise ja töötlemise võimalusi, lisaks ka programmi Microsoft Power BI kui ka PostgreSQL SQL-dialekti.

Töö käigus koostatakse ka praktiline õppematerjal, mis sisaldab juhiseid ja ülesandeid, et õppijad saaksid teoreetilisi teadmisi praktika abil kinnistada. Selle lähenemise eesmärk on mitte ainult tutvustada lugejatele andmete visualiseerimise ja töötlemise teoreetilisi aspekte, vaid ka pakkuda käegakatsutavaid oskusi ja tööriistu, mida nad saavad edaspidises töös rakendada.

Töö juurde käivad ka lisades toodud ülesanded ja vastused. Power BI ja Postgre on kirjeldatud 2023. aasta seisuga, seega hilisemal lugemisel võib olla tehtud nii keeles kui programmis muudatusi.

Metoodika

Teoreetiline uuring

Teoreetilise osa ülevaate eesmärk on koguda teavet PostgreSQL ja Microsoft Power BI kohta, keskendudes eriti nende süsteemide kasutusalaadele. Kirjanduse allikateks on akadeemilised artiklid, raamatud, tehnilised dokumentatsioonid ja veebiallikad.

Praktiline osa

Praktiline osa keskendub PostgreSQLi ja Microsoft Power BI integratsiooni juhendi koostamisele. Selleks luuakse mitmeid praktilisi ülesandeid, kus eeldatakse andmete töötlemist SQLiga ning nende visualiseerimist Power BI-s. Iga ülesanne kavandatakse samm-sammult, alates lihtsatest päringutest kuni keerukamate skriptideni. Töös kasutatakse Tartu Ülikooli Arvutiteaduse instituudi kursusel “Andmebaasid” tuttavat “male” andmebaasi ning sellest tulenevalt ka PostgreSQLi, sest ka see on antud kursusel kasutusel. Autor otsustas selle kasuks, sest siis saab loodud töölehte tulevikus ära kasutada näiteks kursusel lisaülesandena või mõnes muus kursuses, mida tudengid saavad võtta peale eelmainitud kursuse läbimist. Samuti lubas selle andmebaasi kasutamine taaskasutada mitut juba loodud installeerimis- ja ühendamise juhendit.

SQL skriptid PostgreSQLis

Praktilise osa esimeses pooles tuleb tegijal luua SQL skripte PostgreSQLis abil DBeaveri keskkonnas. See etapp on mõeldud selleks, et tutvustada või kinnistada andmebaasis kasutatavate päringute põhialuseid ehk pärida andmebaasi sisu efektiivselt. Skriptid kujundatakse nii, et need demonstreeriks mitmekülgseid andmetöötlusvõtteid alates lihtsatest SELECT lausetest kuni keerukamate JOIN operatsioonide, EXTRACT() funktsioonide, alampäringute ja andmete grupeerimiste kasutamiseni. Samuti käsitletakse erinevate andmetüüpidega (nt tekst numbriks) töötamist.

Andmete visualiseerimine Power BI-s

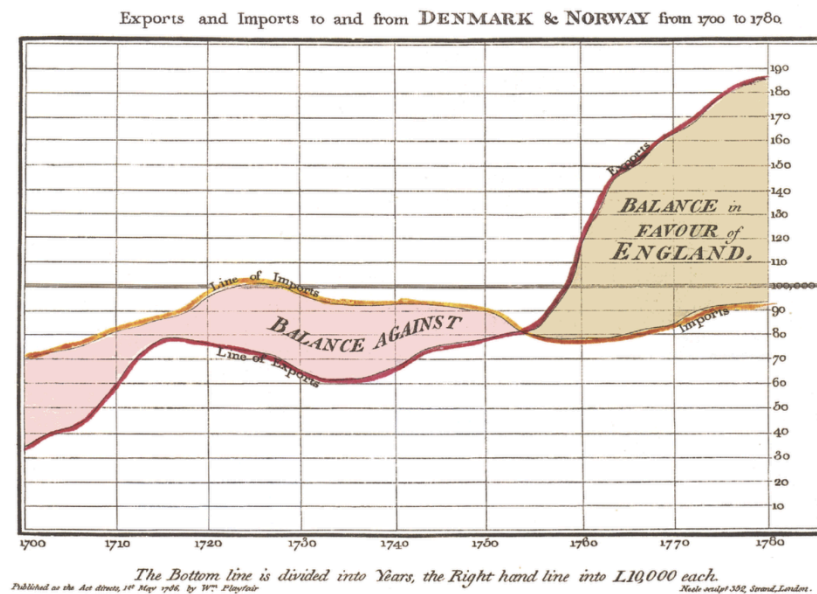
Järgmise sammuna tuleb töölehel kasutada Microsoft Power BI-d. Selles osas õpetatakse, kuidas importida andmeid Power BI-sse ning luua erinevaid visualiseerimistüüpe varem loodud materjalide abil. Ülioluline osa on andmeallikate ja info seadistamine, kus juhendit tuleb väga detailselt järgida. Power BI võimaldab kasutada otseühendust andmebaasiga, seega saavad selles osas kokku eeltehtud skriptid ja visualiseerimistarkvara. Autor kasutas ise Power BI teadmuse saamiseks erinevaid veebiallikaid ja Richard Nõmme 2020. aasta bakalaureusetööd “Ülevaade Microsoft Power BI-st ja õppematerjali koostamine”.

1. Andmete visualiseerimine kui tööriist

1.1 Andmete visualiseerimise ajaloost

Andmete visualiseerimine ja graafiline esitamine on olnud ajaloo jooksul oluline vahend keerulise teabe mõistmiseks ja edastamiseks. Visualiseeringuid on tehtud juba ürgajast [6]. Varaseimad teadaolevad andmete visualiseerimise näited hõlmavad sumerlaste savižetoonide kasutamist kaupade koguste väljendamiseks 3000 eKr ja egiptlaste hieroglüüfide kasutamist põllumajandusandmete kujutamiseks [8]. 2. sajandil lõi Kreeka astronoom Ptolemaios teose Geographia, tuntud maailma kaardi, mis kasutas ruudustikku geograafiliste andmete täpseks esitamiseks [11].

Kaasaegsete andmete visualiseerimise tehnikate arengut saab omistada valgustusajastul tehtud edusammudele. 17. sajandil tutvustas Šoti insener William Playfair joondigramme, tulpdiagramme ja sektordigramme, mida tänapäeval endiselt laialdaselt kasutatakse [8]. Need uuenduslikud visualiseerimise meetodid võimaldasid majandus- ja statistikaandmeid tõhusalt esitada. 18. sajand märkis andmete visualiseerimise valdkonna kiiret arengut. Tuntud medõde ja statistik Florence Nightingale lõi roosdiagrammi, et illustreerida Krimmi sõja ajal sõdurite suremust [7]. Charles Joseph Minardi töö, mis ühendas erinevaid andmetüüpe Napoleoni Vene kampaania kujutamiseks 1812. aastal, peetakse selle ajastu üheks kõige uuenduslikumaks andmete visualiseerimise näiteks [10]. 20. sajandil tulid uued tehnikad ja tehnoloogiad, mis on andmete visualiseerimise valdkonda suuresti mõjutanud. Arvutite tulekuga on interaktiivsed ja dünaamilised visualiseerimised muutunud üha tavalisemaks, võimaldades kasutajatel andmeid tõhusamalt uurida ja manipuleerida [9]. Andmete visualiseerimise valdkond on pidevas arengus, pakkudes uuenduslikke tehnikaid ja tööriistu, mis vastavad pidevalt kasvavale andmemahitudle ja keerukusele.



Joonis 2. Willam Playfairi diagram aastast 1786 [12]

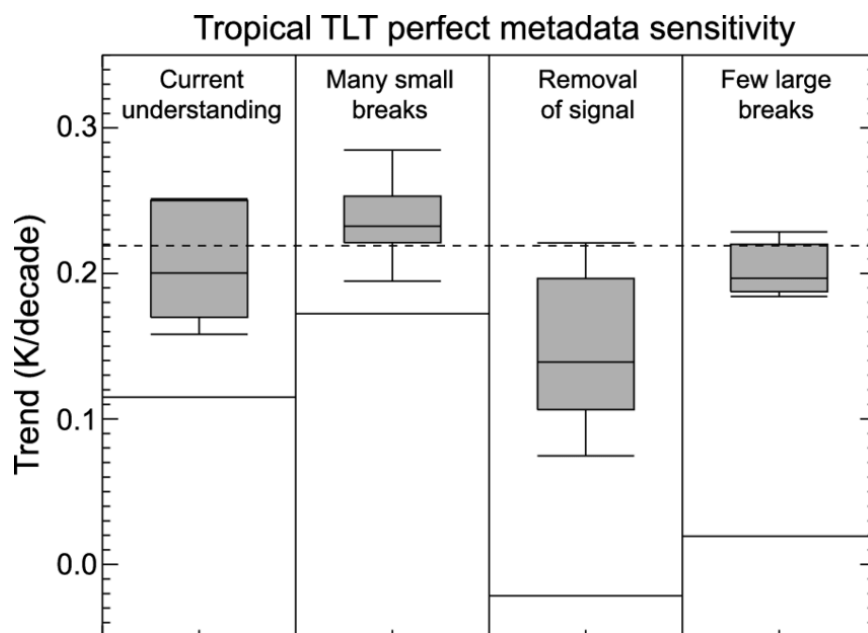
1.2 Miks andmeid visualiseerida?

Esimene oluline eelis andmete visualiseerimisel on nende parem mõistmine. Visualiseeringud võimaldavad meil töödelda keerulist teavet tõhusamalt, kasutades ära meie aju kaasasündinud võimet tunnetada mustreid ja suhteid visuaalselt [4]. Suure koguse andmete korral võib olla keeruline mõista ja tõlgendada aluseks olevaid tendentse ja mustreid. Andmete muutmisel graafilisteks kujutisteks, nagu tulp- või joondigrammid, muutub sama teave oluliselt kättesaadavamaks. See paranenud mõistmine võimaldab kiiresti haarata andmekogumi kriitilisi aspekte ning teha järeldusi tõhusamalt ka neil, kes pole detailselt andmetega kursis.

Teine oluline eelis andmete visualiseerimisel on mustrite tuvastamine. Andmete esitamine visuaalses vormis võimaldab meil kergemini tuvastada suundumusi, erandeid ja korrelatsioone, mis ei pruugi olla ilmsed traditsioonilise arvulise analüüsi kaudu. Kiire mustrite äratundmine soosib uute vaatenurkade ja hüpoteeside tekkimist, mis võib viia läbimurreteni andmete mõistmises [5]. Näiteks aegridade visualiseerimine võib esile tuua tsüklilise käitumise, mida andmetele arvkuju peale vaadates on märksa keerulisem tuvastada.

1.3 Suurandmete visualiseerimise areng alates 1940. aastatest

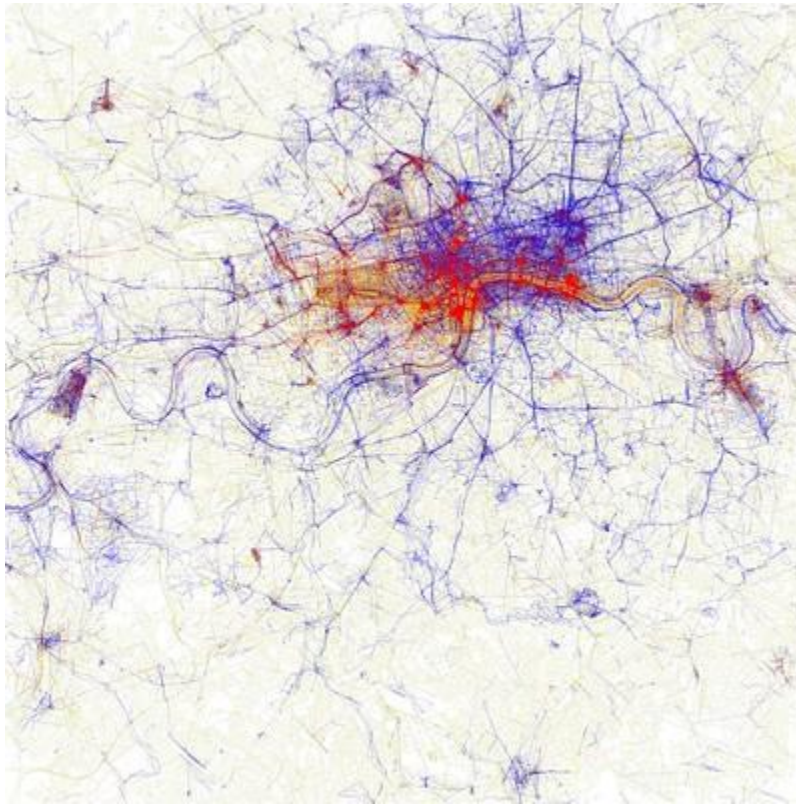
Alates 1940. aastatest on andmete visualiseerimine läbinud olulisi muutusi, peegeldades tehnoloogia arengut ja andmekoguste suurenemist. Varajane periood (1940-1960) oli märgitud staatiliste graafikute ja diagrammide domineerimisega, mida kasutati teaduslikes ja ärilistes aruannetes. Selle arendamisesse andis olulise panuse John Tukey, luues 1950. aastatel analüütilisi meetodeid nagu varajane versioon kastjoonisest, mis võimaldas statistilisi andmeid kompaktselt visualiseerida [16].



Joonis 3. John Tukey näitel loodud kastjoonis. [17]

Seitsmekümnendatel ja kaheksakümnendatel arenes arvutigraafika ja dünaamiline visualiseerimine väga kiiresti. Esimesed interaktiivsed visualiseerimisvahendid lubasid kasutajatel andmetega "mängida" ning avastada andmetes peituvaid seoseid ja mustreid, nagu näiteks Shneidermani uuring aastal 1977, kus ta tõi välja, et programmi abil võib olla parem andmeid mõista tavalise joonise abil. [18]

Interneti laialdane levik 1990. aastatel tõi endaga kaasa visualiseerimistarkvarade, nagu SAP BO esmaversioonid, kasutusele võtmise. Sellel perioodil nähti ka GIS-tarkvara (geoinfotarkvara) ja muude spetsiifiliste tööriistade levikut, mis aitasid erinevate keerukate ruumiliste ja ajaliste andmete visualiseerimisega. 21. sajandi alguse ja interneti laialdase levikuga on erinevate koosanalüüsitavate andmete hulk massiliselt kasvanud. Heaks näiteks on asukohainfot kasutavad kaardid erinevatel teemadel, näiteks joonisel 4 toodud pilt, kus autor on ühele kaardile pannud nii kohalike kui turistide poolt üles laaditavate fotode geoinfo.



Joonis 4. Turistide ja kohalike geoinfoga kaart. [19]

Aja edenemisega on visualiseerimistööriistad muutunud üha võimsamaks. Tööriistad nagu Microsoft Power BI, Tableau võimaldavad kasutajatel mitte ainult andmeid visualiseerida, vaid ka neid analüüsida, muuta ja tõlgendada reaalajas, toetades otsustusprotsesse nii era- kui avalikus sektoris.

1.4 Enne visualiseerimist: Andmete töötlemine

Andmete töötlemine on kriitiline samm andmeanalüüsis. Enne analüüsi alustamist tuleb veenduda, et andmed sobivad edasiseks kasutamiseks – olgu selleks modelleerimine, visualiseerimine või midagi muud. Andmetöötamise protsess algab andmete puhastamisega, mis hõlmab vigade, duplikaatide ja ebatäpsete andmete vajadusel eemaldamist.

Olulise sammuna muudetakse andmed enne analüüsi endale sobivasse formaati. Vastavalt vajadusele teisendatakse andmeid, näiteks logaritmilise muundamise, standardiseerimise või ümberkodeerimise abil (näiteks arv 9 sõneks „9“). Transformatsioon on oluline selleks, et andmed oleksid sobilikud analüütilisteks mudeliteks ja visualiseerimistööriistadeks ja sobiksid kokku teiste andmetega, millega neid äkki koos tahetakse analüüsida [20]. Tihti on üheks osaks ka andmete rikastamine mille käigus lisatakse juba olemasolevatele andmetele veel informatsiooni, näiteks erinevate andmebaasitabelite ühendamise muu info saamiseks.

Andmete segmenteerimine jagab andmed gruppidesse, mis esindavad sarnaseid omadusi või käitumist, võimaldades andmete analüüsijatel kohandada uurimisalgoritme või -meetodeid konkreetsete rühmade vajaduste ja otsitava info järgi. See on eriti kasulik näiteks turundusanalüüsis ja kliendisuhete halduses (segmenteerimisel) [21]. Lõppetapina on andmete töötlemises tavaline osa andmete normaliseerimine, näiteks selliste korduvuste vähendamine andmetes, mis lisakonteksti ei loo. See on oluline suurte süsteemide jõudluse parandamiseks. mis on oluline suurandmete süsteemide jõudluse ja efektiivsuse parandamisel.

Antud töös on kasutatud andmete visualiseerimise tunnetuse saamiseks koos vahendeid PostgreSQL ja Microsoft Power BI.

2. Lühüülevaade Microsoft Power BI-st ja andmebaasidest

2.1. Ülevaade andmebaaside loomise variantidest

Andmebaasi loomiseks on mitmeid erinevaid meetodeid, mis sõltuvad sellest, millist tüüpi andmebaasi kasutaja soovib luua ja milliseid vahendeid tal selleks on. Mõned võimalused on järgmised:

1. Andmebaasi loomine käsitsi - kõige aeganõudvam meetod, kuid võimaldab kasutajal täielikult juhtida andmebaasi struktuuri ja sisu. Selle meetodi korral loob kasutaja andmebaasi tabelid ja väljad ning sisestab andmed ise.
2. Andmebaasi loomine andmebaasi haldussüsteemi abil - see meetod on kiirem ja lihtsam kui esimene, sest kasutab andmebaasi haldussüsteemi, mis hõlbustab andmebaasi loomise protsessi. Andmebaasi haldussüsteemid pakuvad kasutajale mitmeid vahendeid andmebaasi struktuuri ja sisu haldamiseks, näiteks tabelite loomine ja andmete sisestamine [3].
3. Andmebaasi loomine andmebaasi haldussüsteemi kaudu skriptiga - see variant on sarnane teisele, kuid kasutab skripti andmebaasi loomiseks. Skript on kood, mis automatiseerib andmebaasi loomise protsessi ja võimaldab kasutajal andmebaasi loomiseks kasutada erinevaid käske.
4. Andmebaasi loomine andmebaasi haldussüsteemi kaudu andmebaasi mudelite abil - see meetod on sarnane teisele ja kolmandal, kuid kasutab andmebaasi mudelit andmebaasi loomiseks. Andmebaasi mudel on dokument, mis sisaldab andmebaasi struktuuri ja sisu kirjeldust. Mudelit saab kasutada andmebaasi loomiseks ja haldamiseks ilma skripte kirjutamata.
5. Andmebaasi loomine andmebaasi mudelite ja skriptide kombinatsiooniga - kombineerib eelmiste meetodite eeliseid, sest kasutab nii andmebaasi mudelit kui ka skripti andmebaasi loomiseks. See võimaldab kasutajal andmebaasi struktuuri ja sisu täielikult juhtida.

2.2. SQL andmebaaside eelised Microsoft Exceli ees

Andmeteaduse spetsialistid eelistavad andmete haldamisel Exceli asemel sageli andmebaase. Valikul mängivad rolli mitmed tegurid, näiteks laiendatavus, andmete terviklikkus, ligipääsetavus ja tõhusus. Andmebaasid suudavad hõlpsasti hakkama saada suurte andmemahutudega, ületades Exceli, mis ei ole piisavalt võimekas suurte andmekogumite korraga töötlemisel [13]. Organisatsioonid, mis vajavad suurte andmekogumite talletamist, analüüsimist ja töötlemist, võivad Exceli kasutamisel kokku puutuda probleemidega [14].

Andmete terviklikkuse osas säilitavad andmebaasid täpsuse ja järjepidevuse, rakendades selliseid andmebaasi funktsioone nagu andmete valideerimine, muutmise piirangud ja andmeviite terviklikkus. Excel aga ei paku kõiki neid võimalusi, mis võivad inimlike vigade tõttu põhjustada probleeme andmete sisestamisel või muutmisel [15]. Lisaks võimaldavad andmebaasid mitme kasutaja juurdepääsu, võimaldades erinevatel kasutajatel andmeid samaaegselt lugeda ja kirjutada. Exceli failid võivad aga tekitada takistusi andmete jagamisel ja koostöös, eriti kui tegemist on suurte failisuurustega [14].

Lisaks pakuvad andmebaasid tõhusaid päringuid, indekseerimist ja aruannete koostamise tööriistu, mis aitavad andmeteadlastel kiiresti teostada keerulisi analüüse ja hankida konkreetseid andmealasid. Exceli sisseehitatud funktsioonid jäävad selles valdkonnas hätta, pakkudes piiratud võimalusi, mis viivad aeglasema ja vähem efektiivse andmetöötluseni.

2.3. Ülevaade PostgreSQList

PostgreSQL on avatud lähtekoodiga objekt-relatsiooniline andmebaasihaldussüsteem, mida tuntakse sageli lihtsalt "Postgres" nimega. See loodi 1986. aastal California Berkeley Ülikooli tudengi, hilisema professori Michael Stonebrakeri poolt [1]. Tollel ajal töötas Stonebraker andmebaasihaldussüsteemiga Ingres, mida kasutasid paljud ülikoolid ja riigiasutused. Ta tundis, et varasemalt populaarsel andmehaldussüsteemil Ingresil oli puudu mitmeid vajalikke omadusi, seetõttu otsustas ta luua uue andmebaasihaldussüsteemi. Tulemuseks oli Postgres, mis sai oma nime Ingresi mänguliselt tõlgendades.

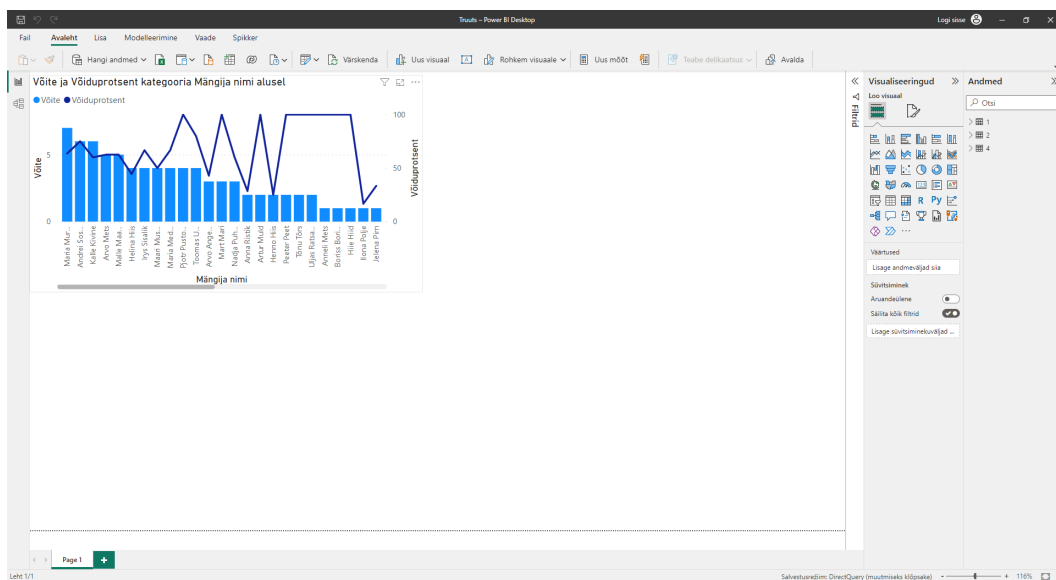
Postgres loodi Ingresist tugevama ja paindlikuma haldussüsteemina. See sisaldas mitmeid uusi omadusi, mida teistes andmebaasides sel ajal ei leidunud, nagu võime salvestada keerukaid andmetüüpe ja toetada kasutaja määratud funktsioone ja operaatoreid. Nende omaduste lisamine tegi Postgresist sobiliku andmebaasisüsteemi paljudele rakendustele, alates lihtsatest andmehoidlatest keerukate analüüside ja mudeliteni.

Aastate jooksul on Postgres jätkuvalt arenenud [23]. 1995. aastal moodustasid arendajad PostgreSQL Global Development Group'i, et tarkvara edasi arendada ja sellele uuendusi toota [1]. Tänapäeval on PostgreSQL laialdaselt kasutusel erinevates tööstusharudes ja rakendustes. See on tuntud oma usaldusväärsuse, paindlikkuse ja jõudluse poolest ning seda toetab suur ja aktiivne kasutajate ja arendajate kogukond. Seda kasutavad paljud suured organisatsioonid, sealhulgas NASA, USA valitsus ja Euroopa Kosmoseagentuur, samuti paljud väikeettevõtted ja individuaalsed kasutajad.

2.4. Ülevaade Microsoft Power BI-st

Microsoft Power BI on ärianalüüsi tarkvara, mis esmakordselt avaldati 2013. aastal. See loodi organisatsioonide andmete paremaks mõistmiseks ja analüüsimiseks ning teadlikumate otsuste tegemiseks. Power BI võimaldab kasutajatel ühendada erinevaid andmeallikaid, sealhulgas Excelit, ning visualiseerida andmeid mitmel viisil. See hõlmab raportite, juhtpaneelide ja teiste visualisatsioonide loomist, mida on lihtne teistega jagada. Power BI integreerub ka teiste Microsofti toodetega, näiteks Office 365-ga [2].

Alates avaldamisest on Power BI läbinud palju uuendusi. See on hõlmanud uute funktsioonide lisamist, näiteks reaajas andmestriimingut, ning kasutajakogemuse parandamist. Tulemusena on Power BI muutunud järjest populaarsemaks organisatsioonide seas üle maailma. Lisaks selle kasulikkusele ärikontekstis on Power BI-d kasutatud ka paljudes teistes valdkondades, näiteks hariduses ja tervishoius. See on võimaldanud nendel valdkondadel saada andmetest tulemuslikke järeldusi ja teha paremaid otsuseid. Üldiselt on Power BI jätkuvalt kasvamas. Kuna andmed saavad maailmas järjest olulisemaks, siis tõenäoliselt jääb Power BI organisatsioonidele, kes tahavad oma andmetest mõistmist saada ja sellest tulemuslikke järeldusi teha, väärtuslikuks tööriistaks.



Joonis 1. Ekraanikuva Microsoft Power BI rakendusest.

3. Töölehe koostamine ja tagasisideanalüüs

3.1 Kasutatud tööriistad ja ülesanded

Autor otsustas töölehe loomise kasuks üle tavaliste juhendite seetõttu, et seda saaks tulevikus kasutada ka õppe- või lisamaterjalidena erinevatel kursustel. Ülesannete läbimine annab töölehe läbijale võimaluse kinnistada teoreetilisi teadmisi praktiliste harjutuste kaudu, mis aitab neil omandatud teadmisi paremini mõista. Samuti võimaldab see tudengitel arendada oma probleemilahendusoskusi ja analüütilist mõtlemist, mida on vaja andmebaaside ja andmete visualiseerimisega seotud ülesannete täitmisel. Tööleht on ka tulevikus kergesti kohandatav ja uuendatav vastavalt tehnoloogia arengule ja kursuse sisu muutumisele.

Töölehes kasutatakse eelräägitud PostgreSQL andmebaasisüsteemi ja Microsoft Power BI analüüsitööriista. PostgreSQL võimaldab päringuid koostada. Power BI abil luuakse interaktiivseid visualiseeringuid. Lisaks ülesannetele on loodud ka juhend ühenduste loomiseks erinevate vajaminevate programmide vahel.

3.2 Ülesannete struktuur

Tööleht koosneb kahest peamisest osast:

- **SQL skriptide loomine (Ülesanded 1-6):** Need harjutused keskenduvad SQL funktsioonide, näiteks CASE, SUM ning grupeerimiskäskude nagu GROUP BY ja ORDER BY kasutamisele ning aitavad töölehe täitjatel mõista, kuidas efektiivselt andmeid kombineerida ja analüüsida.
- **Andmete visualiseerimine Power BI-s (Ülesanded 7-10):** Siin kasutatakse loodud SQL skripte visualiseerimiseks, õppides andmeid efektiivselt esitlema.

3.4 Hindamisskeem

Hindamiskriteeriumid põhinevad iga ülesande korrektusel täitmisel ning loodud visualiseerimiste asjakohasusel. Iga ülesande hindamiseks vajalikud punktid on toodud ülesannete juures Vastuste lehel. Mõned näited:

- **SQL päringute täpsus:** Töölehe täitjad peavad näitama oskust kasutada erinevaid SQL-i funktsioone, andmeid õigesti rühmitada ja järjestada.
- **Visualiseerimise efektiivsus:** Hinnatakse, kui hästi visualiseerimised suudavad edastada analüüsitulemusi.
- **Päringu loetavus ja selgus:** Päringu struktuur peab olema arusaadav ka kõrvaltlugejale.

3.5 Ülesanded

Autor üritas luua töölehe ülesanded nii, et need oleksid lühiajalise tutvumise järel jõukohased ka töölehe täitjale, kes ei ole varem SQLi ja Power BI-ga kokku puutunud.

Ülesandeid lahendatakse “Male” andmebaasil, mis sisaldab kokku nelja tabelit- partiid, turniirid, isikud ja klubid.

Esimene ülesanne keskendub mängijate võitude, kaotuste ja viikide arvutamisele, kasutades CASE ja SUM funktsioone. Tulemused tuleb grupeerida mängija nime järgi.

Teine ülesanne visualiseerib igal turniiril iga klubi võidud, viigid ja kaotused. Läbija peab ühendama mitu tabelit JOIN klauslitega ja kasutama GROUP BY ning ORDER BY klausleid, et andmed oleksid selgelt esitatud.

Kolmas ülesanne arvutab mängijate võiduprotsendi, kasutades matemaatilisi teisendusi nagu NUMERIC ja funktsioone nagu ROUND() ja CAST. Töö hõlmab andmete agregatsiooni ja tulemuste järjestamist võiduprotsendi alusel.

Neljas ülesanne keskendub mängijate võiduprotsendi arvutamisele **mustal poolel mängides**. Täitja peab kasutama funktsioone ROUND() ja CAST AS (numeric) võiduprotsendi leidmisel.

Tudengid peavad tagama, et kuvatakse ainult mängijad, kellel on rohkem kui null mängu, ja järjestama tulemused võiduprotsendi järgi. Vastuses peab olema ka mängija nimi.

Viies ülesanne mõõdab mängijate keskmist mänguaega minutites, kasutades funktsioone `EXTRACT HOUR FROM` ja `MINUTE FROM`. Ülesanne nõuab keskmise mänguaja arvutamist igal turniiril eraldi ning mängijate järjestamist keskmise mänguaja järgi.

Kuues ülesanne loob iga klubi mängijate edetabeli, arvutades mängijate võidud klubi lõikes. Tudengid peavad kasutama `SUM` funktsiooni võitude arvutamiseks ja rühmitama tulemused klubi ja mängija nime järgi, järjestades iga klubi mängijad võitude arvu järgi tähestikulises järjekorras.

Seitsmendas ülesandes tuleb visualiseerida esimese ülesande SQL-skripti tulemused joondiaagrammil, kus horisontaalteljel on mängijate perenimed ja vertikaalteljel on nende võitude arv.

Kaheksandas ülesandes tuleb visualiseerida teise ülesande SQL-skripti tulemused diagrammil, mis ei ole tulpdiaagramm.

Üheksandas ülesandes tuleb visualiseerida kolmanda ülesande SQL-skripti tulemused vabalt valitud visualiseeringuga, millel on mängijate võiduprotsendid..

Kümnendas ülesandes tuleb visualiseerida neljanda ülesande SQL-skripti tulemused joondiaagrammil, kus horisontaalteljel on mängijate nimed ja vertikaalteljel on nende võiduprotsendid.

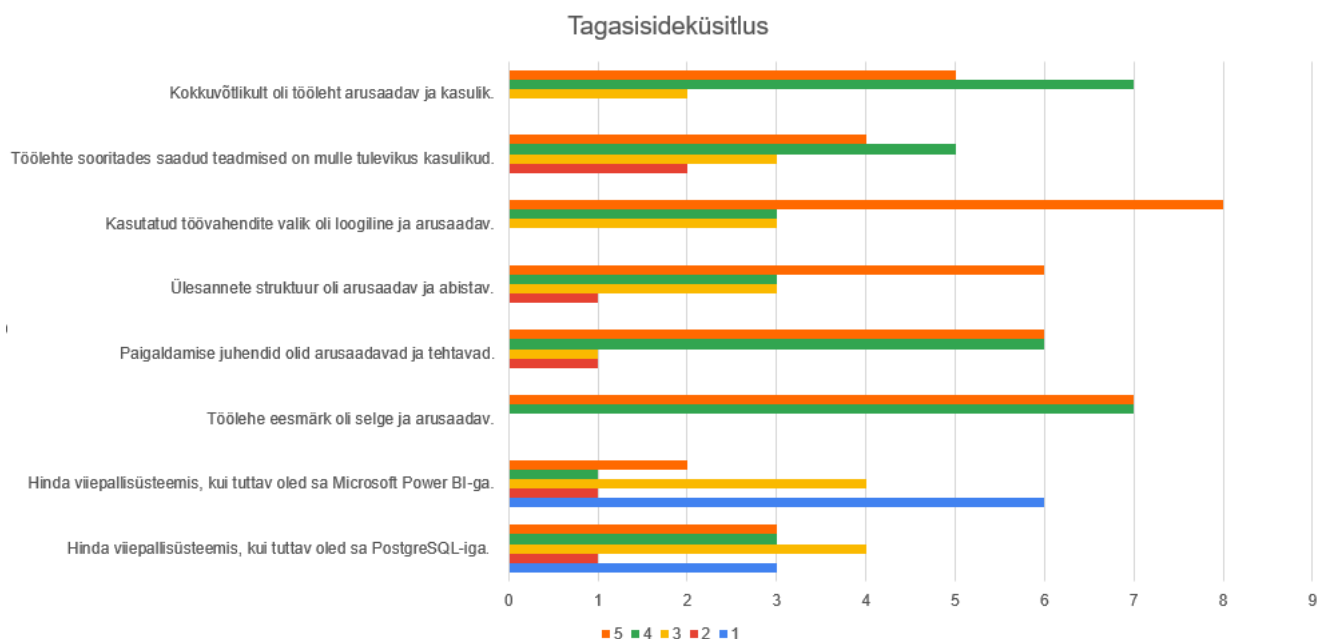
3.6 Tagasiside

Koos ülesannetega saadeti vastajatele ka küsimused töölehe sisu ja ülesehituse hindamiseks. Küsimustikus oli kümme küsimust ja üks vaba tagasisideväli, mida täitis 14 vastajat. Vormis esitati järgmised küsimused / laused:

1. Palun vali enda vanus.
2. Hinda viiepallisüsteemis, kui tuttav oled sa PostgreSQL-iga.
3. Hinda viiepallisüsteemis, kui tuttav oled sa Microsoft Power BI-ga?
4. Töölehe eesmärk oli selge ja arusaadav?
5. Paigaldamise juhendid olid arusaadavad ja tehtavad?
6. Ülesannete struktuur oli arusaadav ja abistav?
7. Kasutatud töövahendite valik oli loogiline ja arusaadav?
8. Töölehte sooritades saadud teadmised on mulle tulevikus kasulikud?
9. Kokkuvõtlikult oli tööleht arusaadav ja kasulik?
10. Kui kaua aega kulus sul SQLi, DBeaveri ja PowerBI paigaldamisele ja ühenduse saavutamisele mõlemas programmis?
11. Siia võib lisada vaba tagasisidet.

Vastuseid sai hinnata 1-5 palli süsteemis, va. küsimusi 1, 10 ja 11. Küsimusele 1 toodud vastusevahemikud olid 15-19, 19-25, 25-30 ja 30+. Küsimusele 10 sai vastuseid valida järgnevast loetelust: “Kuni pool tundi”, “Kuni tund”, “Tund kuni 2 tundi”, “Rohkem kui 2 tundi”.

Joonis 5. Tagasisideküsitluse statistika, kus x-teljel on vastajate arv, kes nii arvasid.



Umbes kolmandik osalejatest olid Power BI-ga varem tuttavad. Veel veidi üle kolmandiku olid Power BI-st varem kuulnud, kuid mitte süvitsi, mis näitas, et nad vajasisid võib-olla rohkem aega juhendite ja funktsioonidega tutvumiseks, kuid samuti tõestab, miks töölehe loomine vajalik oli- teadlikkus programmist on madal.

PostgreSQLi puhul oli osalejate seas rohkem oskajaid. Kümme vastajat olid selle andmebaasisüsteemiga varasemalt töötanud (vähemalt hinne 3). Tervikust 22% olid edasijõudnud kasutajad. Neli vastajat ei olnud PostgreSQLiga varasemalt kokku puutunud.

Töölehe täitjate vastuste analüüs näitab, et enamus vastajatest hindas töölehte positiivselt. Kõik vastajad leidsid, et tööleht oli arusaadav ja kasulik, neist vaid 2 andsid hindeks “3”. Tuleviku kasulikkuse osas leidis 5 täitjat, et see ei ole neile väga kasulik (“2” ja “3”). See on tõenäoliselt seotud nende vastajate varasema SQL tasemega- neile olid ülesanded liiga lihtsad.

Kasutatud töövahendite osas leidis 78% osalejatest, et need olid loogiliselt valitud ja lihtsasti mõistetavad. Neist 72% oli sellega täielikult nõus ja 28% osaliselt. Neutraalne hoiak, mida väljendas 22% vastanutest, oli osati seotud vastajate varem installeeritud tarkvarast

tulenevate probleemidega, mis viidates vajadusele kaasata juhenditesse mitmekülgsemat tarkvara tuge.

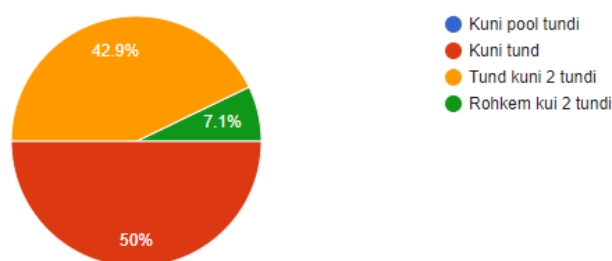
Ülesannete struktuur oli vastajatele sobilik - 42% tundis, et need olid väga abistavad ja arusaadavad ning aitasid kaasa ülesannete mõistmisele ("5"), vaid 1 vastaja vastas "2". Vastajad tõid esile ülesannete selge jaotuse ja samm-sammulise lähenemise, samuti olid abiks toodud näidisvastused.

Paigaldusjuhendite puhul kinnitas 85% vastajatest, et need olid selged ja järgitavad. Ülejäänud kahest vastajast jäi 1 neutraalseks ja 1 vastas, et juhenditest oli pigem keeruline aru saada. Üks vastaja lisas, et juhendites oleks võinud olla toodud ka mõni teine populaarne SQL skriptidele mõeldud programm, nagu näiteks Microsoft SQL Server, et pakkuda alternatiive levinumatele tööriistadele.

Töölehe eesmärk oli selge kõigile vastajatele. Vastajate positiivne tagasiside eesmärgi selguse kohta kinnitab, et sissejuhatus oli piisavalt informatiivne ja motiveeriv.

Kui kaua aega kulus sul SQLi, DBeaveri ja PowerBI paigaldamisele ja ühenduse saavutamisele mõlemas programmis?

14 responses



Joonis 6. Vastajate ajakulu paigaldustele ja ühenduste saavutamisele.

Pooltel vastajatel kulus paigalduseks kuni tund, mis oli hinnanguliselt ka see aeg, mis kulus töölehe loojal. Kuuel kulus tund kuni 2 tundi ja ühel rohkem kui 2 tundi, mis näitab, et

juhendid võiksid kohati olla selgemad. Ühe lahendusena oleks võibolla mõeldav ülesannete tegemine (ja andmebaasi ülessättimine) mõnes pilvesüsteemis.

3.7 Kokkuvõte

Töö esimeses peatükis tutvustati andmete kogumise ja visualiseerimise ajalugu, lisaks ka andmete visualiseerimise põhjuseid ja arengut läbi aja. Töö teises peatükis tutvustati PostgreSQL'i ja Microsoft Power BI'd. Kolmandas peatükis tutvustatakse loodud töölehte ja tagasisideküsitluse tulemusi. Lõputöö tegemise käigus loodi 6 SQL ülesannet male andmebaasile, samuti 4 visualiseerimisülesannet. Töö vastuste juures on ära toodud ka metoodiline juhend hindamiseks ja loodud SQL- ülesannete lahendused. Tagasisideküsitluse tulemused kinnitavad, et tööleht oli osalejate seas üldiselt edukas ning ülesannete sooritamisega saadi kenasti hakkama. Sellegipoolest ilmnas ka mitmeid aspekte, mida tulevikus parendada saaks. Esiteks tuleks täiendada juhendamaterjale, eriti installeerimisprotsesside osas. Hetkel kasutusel olevad kirjalikud juhised võiksid täiendust saada veel detailsema videoõpetusega, kus oleks lisaks ühenduse loomisele ka installeerimine sammhaaval tehtud. See aitaks vähendada segadust ja tõsta kasutajate iseseisvat suutlikkust tarkvara paigaldamisel. Teiseks, kuigi kasutatud tehnoloogilised lahendused on hetkel asjakohased, võiks tulevikus kaaluda alternatiivsete programmide lisamist. See mitmekesistaks õppematerjali ning pakuks kasutajatele võimalust valida erinevate tööriistade vahel, mis võib paremini sobida nende eelistustega ja äkki on ka arvutis olemas.

Viidatud kirjandus

- [1] The PostgreSQL Global Development Group. History of PostgreSQL. 1996-2022. <https://www.postgresql.org/docs/current/history.html> [10.12.2023]
- [2] Microsoft. What is Power BI? 29.01.2023. <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview> [10.12.2023]
- [3] Tooming K. , Taimre L. Mitmetabelilise andmebaasi loomine ja sellesse andmete tõmbamine tekstifailidest. <https://sisu.ut.ee/ajalooandmebaasid/mitmetabelilise-andmebaasi-loomine-ja-sellesse-andmete-t%C3%B5mbamine-tekstifailidest> [10.12.2023]
- [4] Healey, C.K. , Enns, J.T. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics*, vol 18, no.7, 07.2012. p.1170-1188. <https://www.csc2.ncsu.edu/faculty/healey/download/tvcg.12a.pdf> [25.03.2024]
- [5] Cleveland, W.S. , McGill, R. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, vol 79, no.387, 09.1984. p.531 – 554. <http://faculty.washington.edu/aragon/classes/hcde511/s12/readings/cleveland84.pdf> [25.03.2024]
- [6] Friendly, M. A brief history of data visualization. *Handbook of Data Visualization*. 01.2008, p. 15-56. https://www.researchgate.net/publication/226400313_A_Brief_History_of_Data_Visualization [20.03.2024]
- [7] Wikipedia täiendajad. Florence Nightingale. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Florence_Nightingale [20.03.2024]
- [8] Friendly, M. , Denis, D. J. Milestones in the history of thematic cartography, statistical graphics, and data visualization. 2001. <https://www.datavis.ca/milestones/index.php> [04.03.2024]

- [9] Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings of the 1996 IEEE Symposium on Visual Languages*. p. 336-343. 1996. <https://ieeexplore-ieee-org.ezproxy.utlib.ut.ee/stamp/stamp.jsp?tp=&arnumber=545307> [01.04.2024]
- [10] Friendly, M. Visions and re-visions of Charles Joseph Minard. *Journal of Educational and Behavioral Statistics*, vol 27, no.1. p.31-51. <https://www.datavis.ca/papers/jeps.pdf> [27.03.2024]
- [11] Harley, J.B. Cartography in Prehistoric, Ancient, and medieval Europe and the Mediterranean. 1987. *Chapter 1: The Map and the Development of the History of Cartography*. p.1-42. https://press.uchicago.edu/books/HOC/HOC_V1/HOC_VOLUME1_chapter1.pdf [03.04.2024]
- [12] Playfair, W. Exports and Imports to and from Denmark & Norway from 1700 to 1780. *Commercial and Political Atlas*, 1786. https://en.wikipedia.org/wiki/William_Playfair#/media/File:Playfair_TimeSeries-2.png [31.03.2024]
- [13] Microsoft. Excel specifications and limits. 2021. <https://support.microsoft.com/en-us/office/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3> [25.03.2024]
- [14] IBM. Spreadsheets are holding you back. 2020. <https://www.ibm.com/downloads/cas/LYZMAZD6> [04.04.2024]
- [15] Lutkevich, B. , Biscobing, J. Relational Database. 06.2021. <https://www.techtarget.com/searchdatamanagement/definition/relational-database> [22.03.2024].
- [16] Friendly, M. The Golden Age of Statistical Graphics. *Statistical Science* 23(4), 502–535. <https://www.jstor.org/stable/20697655> [18.03.2024]

[17] Cleveland, W.S. Tukey box plot for tropical average TLT trends,1994.

https://www.researchgate.net/figure/Tukey-box-plot-Cleveland-1994-for-tropical-average-TLT-trends-for-1979-2003-from-a_fig1_228574575 [19.03.2024]

[18] Schneiderman, B. ,Mayer, R.,McKay, D., Heller, P. Experimental investigations of the utility of detailed flowcharts in programming. 01.06.1977, *Communications of the ACM*, vol 20, no. 6. p. 373-381. <https://dl.acm.org/doi/pdf/10.1145/359605.359610> [15.03.2024]

[19] Fischer, E. Locals and Tourists #1 (GTWA #2): London. Flickr, 05.06.2010.

<https://www.flickr.com/photos/walkingsf/4671589629/in/album-72157624209158632/>
[20.03.2024]

[20] Kelleher, J. D. , Tierney, B. Data Science. *Chapter 5: Standard Data Science Tasks*. MIT Press, 2018.

[21] Chen, M. , Mao, S. , Zhang, Y. , Leung, V.C.M. Big Data: Related Technologies, Challenges and Future Prospects. *Chapter 6.3.1: Application of Big Data in Enterprises*. 05.05.2014.

[22] The PostgreSQL Global Development Group. PostgreSQL- About. 2024.

<https://www.postgresql.org/about/> [12.05.2024]

Lisa

1. Materjalid

Juhend PostgreSQL, DBeaveri ja Power BI installimiseks:

https://docs.google.com/document/d/1JROVjWE15uKPo_G0Yg2YoQ8Jz7kU9VFI/edit?usp=sharing&ouid=100360935309244295279&rtpof=true&sd=true

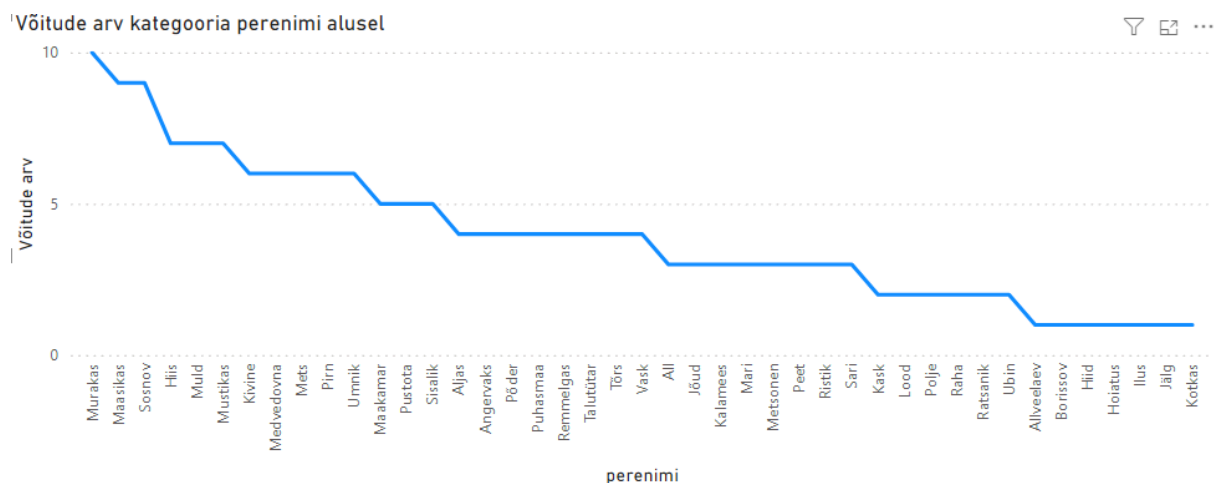
Ülesanded lahendamiseks:

<https://docs.google.com/document/d/1bSZKxT3vi0ic1taGMzVBaViCKCmGEmRt/edit?usp=sharing&ouid=100360935309244295279&rtpof=true&sd=true>

Vastused SQL ülesannetele:

<https://docs.google.com/document/d/1jmvj0rvufKPDt39oHVpJkPBPTcR5rneL/edit?usp=sharing&ouid=100360935309244295279&rtpof=true&sd=true>

Näidisvastus ühele Power BI ülesandele:



2. Tagasisideküsitlus

Palun vali enda vanus. *

- ☐ 15-19
- ☐ 19-25
- ☐ 25-30
- ☐ 30+

Hinda viiepallisüsteemis, kui tuttav oled sa PostgreSQL-iga. *

	1	2	3	4	5	
Väga halb- eelnev kogemus puudub.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Väga hea- tunnen ennast PostgreSQL-iga mugavalt.

Hinda viiepallisüsteemis, kui tuttav oled sa Microsoft Power BI-ga. *

	1	2	3	4	5	
Väga halb- eelnev kogemus puudub.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Väga hea- tunnen ennast Power BI-ga mugavalt.

Töölehe eesmärk oli selge ja arusaadav. *

	1	2	3	4	5	
Ei nõustu üldse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Nõustun täielikult

Paigaldamise juhendid olid arusaadavad ja tehtavad. *

	1	2	3	4	5	
Ei nõustu üldse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Nõustun täielikult

Ülesannete struktuur oli arusaadav ja abistav. *

	1	2	3	4	5	
Ei nõustu üldse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Nõustun täielikult

Kasutatud töövahendite valik oli loogiline ja arusaadav. *

	1	2	3	4	5	
Ei nõustu üldse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Nõustun täielikult

Töölehte sooritades saadud teadmised on mulle tulevikus kasulikud. *

	1	2	3	4	5	
Ei nõustu üldse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Nõustun täielikult

Kokkuvõtlikult oli tööleht arusaadav ja kasulik. *

	1	2	3	4	5	
Ei nõustu üldse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Nõustun täielikult

Kui kaua aega kulus sul SQLi, DBeaveri ja PowerBI paigaldamisele ja ühenduse saavutamisele mõlemas programmis? *

- ☐ Kuni pool tundi
- ☐ Kuni tund
- ☐ Tund kuni 2 tundi
- ☐ Rohkem kui 2 tundi

Siia võib lisada lisatagasisidet.

Your answer

3. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Rene Aleksander Truuts,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose “Ülevaade Microsoft Power BI ja PostgreSQL integratsioonist ja õppematerjali koostamine”, mille juhendaja on Vambola Leping 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.

Rene Truuts 15.05.2024