

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Hanna Maria Mägi
Tartu busside hilinemiste ja peatumiste analüüs
Bakalaureusetöö (9EAP)

Juhendaja:
Pelle Jakovits, Ph.D

Tartu 2025

Tartu busside hilinemiste ja peatumiste analüüs

Lühikokkuvõte:

Käesolevas töös analüüsitakse Tartu linna bussiliikluse hilinemisi aastatel 2019–2024. Töö eesmärk oli hinnata, kui täpselt vastavad ennustused tegelikele saabumisaegadele ning tuvastada ajast kinnipidamise seisukohalt probleemsed liinid, perioodid ja peatused. Andmestik saadi Cumulocity keskkonnast ning töötlus viidi läbi Pythonis. Tulemused näitavad, et teatud liinidel ja perioodidel kaldub tegelik hilinemine oluliselt prognoosidest kõrvale. Töös loodi ka puhastatud ja struktureeritud andmestik, mida saab kasutada tulevikus edasiarenduste ja linnapoolse planeerimise tarbeks.

Võtmesõnad: Ühistransport, buss, hilinemine, IoT, andmeanalüüs, Tartu, Cumulocity, GTFS.

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Analysis of Tartu bus delays and stops

Abstract:

This thesis analyzes bus delays in the city of Tartu during the years 2019–2024. The aim of the study was to evaluate how accurately predicted arrival times reflect actual arrival times and to identify problematic routes, time periods, and stops in terms of schedule adherence. The dataset was obtained from the Cumulocity platform and processed using Python. The results show that, for certain routes and time periods, actual delays deviate significantly from predictions. A cleaned and structured dataset was also created as part of the work, which can be used for future developments and urban planning by the city.

Keywords: Public transport, bus, delay, IoT, data analysis, Tartu, Cumulocity, GTFS.

CERCS: P170 Computer science, numerical analysis, systems, control

Sisukord

1.	Sissejuhatus.....	5
2.	Ülevaade valdkonnast ja varasematest töödest	7
2.1	Valdkondlik taust	7
2.2	Varasemad tööd Tartu linna kohta	7
2.3	Hilinemiste ja ajastuse analüüs teaduskirjanduses.....	8
3.	IoT kasutus.....	10
3.1	IoT platvormid	10
3.2	IoT Tartu bussiliikluses.....	12
4.	Metoodika	14
4.1	Andmeallikad.....	14
4.2	Andmete puhastamine ja eeltöötlus	17
4.3	Tehniline lähenemine.....	19
4.3.1	Google Colab	19
4.3.2	Ubuntu server.....	19
4.3.3	PyCharm	19
4.3.4	Python	20
4.3.5	Pandas	20
4.3.6	Tehisaru kasutamine	20
5.	Tartu busside analüüs.....	21
5.1	Koosolek Tartu esindajatega.....	21
5.2	Uurimisküsimused	22
5.3	Andmed peale eeltöötlust.....	23
5.4	Tulemused.....	26
5.4.1	Kuidas on sõidugraafikute regulaarne muutmine septembris ja jaanuaris mõjutanud busside hilinemisi?	26
5.4.2	Kas busside hilinemiste ennustamine vastab reaalsusele?.....	29
5.4.3	Millised liinid on kõige ebausaldusväärsemad ajast kinnipidamise osas?.....	33
5.4.4	Millistes peatustes on hilinemine kõige suurem?	39
5.4.5	Kui kaugel peatuvad bussid peatustest?.....	43
5.4.6	Kui kvaliteetsed on busside poolt edastatud GPS andmed? Kas leidub varieeruvust erinevate busside seadmete kvaliteedis?.....	48
6.	Tuleviku edasiarendused.....	53

7. Kokkuvõte.....	55
Viidatud kirjandus.....	57
Lisad.....	58
Litsents.....	61

1. Sissejuhatus

Tartu ühistranspordi täpsus ja usaldusvärsus on oluline nii linnaelanike igapäevase liikumisvõimaluse kui ka linna kui teenuseosutaja maine seisukohalt. Reisijate jaoks on busside õigel ajal saabumine oluline tegur transpordiviisi valikul, eriti juhul, kui liikumine sõltub täpsetest ümberistumistest või ajakriitilistest kohustustest.

Tartu linnas on esinenud probleeme busside hilinemisega. Probleemi kogevad otseselt Tartu elanikud, kes planeerivad oma liikumisi bussigraafikute alusel. Kui busside tegelik liikumine ei vasta prognoosidele ega sõiduplaanile, väheneb usaldus ühistranspordi vastu ning suureneb tõenäosus, et eelistatakse muid transpordiviise. Kuna ühistransport on Tartu linna hallatav avalik teenus, mõjutab selle töökindlus otseselt ka linna usaldusvärsuse tajumist.

Tartu linnal on olemas nii reaajas bussiaegade jälgimise rakendus¹ kui ka juurdepääs GTFS-formaadis (General Transit Feed Specification formadis) sõiduplaaniandmetele². Olemasolevates tööprotsessides ei ole seni põhjalikult analüüsitud busside hilinemisi aastate lõikes ega nende peatumistäpsust ruumilises tähenduses. Samuti ei ole varasemalt uuritud busside asukoha andmeid ning nende peatumise kaugust ametlikest peatustest. Üheks peamiseks takistuseks selliste analüüside läbiviimisel on olnud andmemahud. Cumulocity keskkonda salvestatud bussiaegade koguneb aastate lõikes kümnete gigabaitide kaupa, mis eeldab spetsiaalseid töötlusvahendeid ning automatiseeritud andmevoogude käsitlemist.

Käesoleva töö eesmärk on analüüsida Tartu bussiliikluse ajast kinnipidamist aastatel 2019–2024 ning hinnata ennustussüsteemi täpsust võrreldes tegelike saabumisaegadega. Analüüsi käigus selgitatakse välja, millised liinid kalduvad enim hilinema ning millised peatused või liinilõigud vajaksid täiendavat tähelepanu. Töö tulemusena valmib puhastatud ja struktureeritud andmestik, mis on edaspidi taaskasutatav ka teiste analüütikute või linnaplaneerijate poolt.

Töös kasutati tehisintellekti (TI) rakendusi abivahendina andmeanalüüsi ja tekstiloome protsessis. Andmete esmaseks analüüsiks rakendati Google Colabi keskkonda, mis põhineb Jupyter Notebook arhitektuuril ning sisaldab sisseehitatud tehisintellekti funktsionaalsust. Andmetöötluse automatiseerimiseks ja keerukamate loogikatega koodilõikude genereerimiseks kasutati Google Colab [1] keskkonnas kättesaadavat koodisoovituste mudelit.

¹ <https://tartu.ee/et/tartu-bussiaegade-mobiilirakendus>

² <https://peatu.ee/gtfs/>

Lisaks rakendati OpenAI poolt arendatud ChatGPT-4o mini [2] keelemudelit, et toetada andmetöötluskriptide koostamist serverikeskkonnas ning viimistleda lõputöö tekstilist osa. Tekstilise osa juures esitati keelemudelile valmis kirjutatud lõik ning paluti selle sõnastust akadeemilisemaks muuta. TI-mudelite abil loodud või muudetud sisu on lõputöös autoripoolselt kontrollitud, vajadusel muudetud ning korrektselt vormistatud vastavalt teadustöö nõuetele.

Töö on jaotatud kuueks peatükiks.

Käesoleva töö teises peatükis antakse ülevaade varasematest teadustöödest ja tehnoloogilistest alustest, mis käsitlevad Internet of Things (IoT) ehk asjade internet (samuti eestipärase nimetusega värgvõrk) kasutust, ühistranspordi andmeanalüüsi ning targa linna lahendusi. Kolmas peatükk keskendub IoT-tehnoloogia rakendamisele Tartu bussiliikluses ning tutvustab Cumulocity ja ArcGIS platvormide rolli andmete kogumisel ja visualiseerimisel. Neljandas peatükis kirjeldatakse töö metoodikat, sealhulgas kasutatud andmeallikaid, töötlusprotsessi, tarkvaralisi vahendeid ja analüüsi ülesehitust. Viiendas peatükis esitatakse töö käigus saadud tulemused, mis vastavad püstitatud uurimisküsimustele. Kuuendas peatükis arutletakse tulevikusuundade ja võimalike süsteemsete muudatuste üle ning seitsmendas peatükis tehakse kokkuvõtte töö olulisematest järeldustest ja panusest.

2. Ülevaade valdkonnast ja varasematest töödest

Ühistranspordi töökindlus ja täpsus on kujunenud üheks keskseks teemaks linnalise liikuvuse arendamisel, eriti kontekstis, kus soovitakse vähendada sõltuvust eraautodest. Usaldusväärse ühistranspordi tagamiseks on oluline mõista, millised tegurid mõjutavad bussiaegadest kinnipidamist ning kuidas neid andmepõhiselt hinnata. Käesolevas peatükis käsitletakse ühistranspordi täpsuse uurimisega seotud teaduskirjandust ning tutvustatakse peamisi tehnoloogilisi platvorme ja andmeformaate standardeid, mida kasutatakse hilinemiste ja peatusandmete analüüsimisel.

2.1 Valdkondlik taust

Ühistransport moodustab olulise osa linnade liikuvusvõrgustikust, mõjutades otseselt linnaelanike liikumisvõimalusi, ajakulu ning keskkonnajalajälge. Linnade kasv ja vajadus pakkuda autoliiklusele keskkonnasõbralikke alternatiive on tõstnud ühistranspordi töökindluse ja täpsuse esiplaanile. Uurimused näitavad, et isegi lühiajalised hilinemised võivad mõjutada kasutajate rahulolu ning vähendada ühistranspordi atraktiivsust võrreldes individuaalse transpordiga [3]. Seetõttu on täpne ja pidev monitooring ning analüüs busside tegeliku ja planeeritud liikumise osas muutunud kriitiliselt tähtsaks.

Uute andmekogumistehnoloogiate, eriti värvvõrgu platvormide kasutuselevõtt, on muutnud võimalikuks reaajas toimiva transpordisüsteemi detailse jälgimise. Cumulocity tüüpi platvormid võimaldavad koguda sündmuspõhist infot bussi liikumise ja peatumiste kohta iga paari sekundi järel. Nende andmete alusel on võimalik läbi viia varasemast oluliselt täpsemat ja dünaamilisemat analüüsi, kaasates nii ajalise kui ruumilise mõõtme. Lisaks sellele on üldiselt laialt levinud GTFS andmestik, mis sisaldab ühistranspordi ametlikke sõiduplaane, peatuseid ja marsruute ning mida kasutatakse võrdlusbaasina tegelike andmete analüüsimisel.

2.2 Varasemad tööd Tartu linna kohta

Tartu linna ühistranspordisüsteemi analüüs on seni olnud piiratud. Varasemates lõputöödes on peamiselt keskendutud jalgrattaringlusele või nutikatele linnaelementidele. Näiteks on uuritud COVID-19 mõju Tartu Smart Bike'i kasutusele [4], milles keskenduti rattareiside kestusele ja sagedusele. Kuigi see töö kasutab samuti liikuvusandmeid, ei hõlma see bussitranspordi hilinemiste või ajagraafikute täpsuse uurimist.

Peeter Vahe bakalaureusetöös [5] on uuritud Tartu Rattaringluses rataste lahti tegemiseks kasutatavaid kaarte ning kuidas teha nendest funktsioneerivad duplikaate. Uuriti ja selgitati

kaartidel olevaid andmeid ning nende ainulaadseid omadusi. Nimetatud bakalaureusetöö näitas, kuidas toimub kloonimisprotsess ja kaartide autentimine, pakuti soovitusi, kuidas Tartu Rattaringluse autentimisprotsessi parandada.

Erald Keshi uuris [6] erinevaid meetodeid, mille abil saab hinnata ainult bussi sisenemisega seotud AFC-süsteemidest pärit väljumise andmeid. Töös on toodud, et ainult sisenemisega seotud automaatsed piletihindade kogumise süsteemid, nagu näiteks Tartus kasutatav süsteem, ei hõlma selliseid andmeid nagu reisijate sihtkohad või ühistranspordist väljumiste ajad. Töö autori sõnul piirab see lähte- ja sihtkohtade maatriksite koostamist ja selliste andmete saamist nagu busside täituvus ja tiptunnid. Üks võimalus selle piirangu kõrvaldamiseks on kasutada tõenäosuslikke hindamismeetodeid, mida vastav magistr töö uuris.

Tartu osales rahvusvahelises koostööprojekti SmartEnCity [7], mis kestis perioodil 2016-2022 ja mida rahastati Euroopa Liidu teadusuuringute ja innovatsiooni raamprogrammist Horisont 2020 [8]. Projekti peamine eesmärk oli viia ellu targa ja säästva linnakeskkonna terviklahendus, mis inspireeriks elanikke tegema keskkonnateadlikke otsuseid ning oleks hiljem erinevates Euroopa piirkondades rakendatav. SmartEnCity projekt ja sellele eelnenud planeerimisalased initsiatiivid on keskendunud laiemalt Tartu linna keskkonnasäästlikkuse tõstmisele, sealhulgas transpordi optimeerimisele.

Varasemates töodes ei ole põhjalikult uuritud busside reaalsel liikumist ega nende peatumiskohtade täpsust ruumilises tähenduses. Bussiasukoha andmete analüüsi on seni takistanud nende suur maht ja keeruline struktuur, kuna andmed edastatakse sündmuspõhiselt ning kogunevad kümnete miljonite ridadena aastas. Seetõttu on hiline mis te ja saabumisprognooside täpsuse hindamine jäänud seni käsitlemata. Käesolev töö üritab seda puudujääki leevendada, pakkudes esmakordselt detailset vaadet Tartu ühistranspordi reaalaja andmete kvaliteedile ja täpsusele.

2.3 Hilinemiste ja ajastuse analüüs teaduskirjanduses

Rahvusvaheliselt on ühistranspordi hiline mis te analüüsi käsitletud mitmete metoodiliste lähenemiste kaudu, sealhulgas statistiline modelleerimine, trajektoori- ja ajaseeriaanalüüs, regressioonimudelid ning masinõppepõhised prognoosimudelid [9]. Uurimustes on keskendutud ajagraafikust kõrvalekaldumise sageduse ja kestuse kvantifitseerimisele ning selle seostele kellaaja, nädalapäeva, ilmastiku, liinikoormuse ja marsruudimadustega. Samuti on leitud, et tiptundidel ja suurtel magistraalidel esinevad hiline mis te on oluliselt

intensiivsemad, mida on analüüsitud nii punktuaalselt (nt konkreetse peatuse kontekstis) kui ka süsteemselt (kogu liinivõrgu ulatuses).

Hilinemiste hindamisel kasutatakse laialdaselt andmeallikaid nagu GTFS andmed reaalajas, mis võimaldavad võrrelda reaalajas liikumist planeeritud sõiduplaaniga. Selline lähenemine aitab tuvastada nii individuaalseid hilinemisi kui ka mustreid kogu liinivõrgus. Hilinemiste süstemaatilist ja juhuslikku jaotumist on uuritud Seattle'i ühistranspordivõrgus [10], näidates, kuidas GTFS-realtime andmestiku põhjal saab eristada korduvaid süsteemseid probleeme juhuslikest häiretest. See lähenemine võimaldab paremini hinnata teenuse töökindlust ja kasutajate kogemust kogu süsteemi lõikes.

Viimastel aastatel on suurenenud huvi ka peatuspaari tasemel hilinemiste ahela analüüsi vastu. Selle abil uuritakse, kuidas hilinemine ühes peatuses kandub edasi järgmistesse, võimaldades paremini hinnata süsteemset mõju kogu liini või reisi ajakavale [11]. Selline analüüs aitab lisaks punktuaalsele täpsusele avastada ka nõrku kohti transpordivõrgus.

Masinõppemeetodite, näiteks närvivõrkude, otsustuspuude ja gradient boostingu, abil on loodud ka hilinemisprognoosimise mudeleid, mis suudavad arvestada kompleksseid mittelineaarseid seoseid ning pakkuda ajas dünaamilisi ennustusi [12]. Sellised lähenemised on eriti efektiivsed, kui kasutada suurt andmemahku reaalajas sisendina.

Kokkuvõttes näitab teaduskirjandus, et hilinemiste analüüs ühistranspordis on mitmetasandiline ja nõuab mahukat ning struktureeritud andmetöötlust. Sellised analüüsid on olulised nii sõidugraafikute optimeerimisel kui ka teenuse usaldusväärsuse parandamisel reisijate vaates. Käesolev töö toetub nendele põhimõtetele, keskendudes positsioneerimisandmetest ja hilinemisväärtustest tuletatud keskmiste hinnangute analüüsimisele peatuste tasemel, et hinnata Tartu linna bussiliikluse ajakohasust ja täpsust erinevate liinide ja perioodide lõikes.

3. IoT kasutus

IoT on tehnoloogia, millega kasutajad saavad oma seadmeid internetiga ühendada ning edastada nende seadmete kaudu infot. Käesolevas peatükis käsitletakse, kuidas IoT-tehnoloogiat rakendatakse Tartu bussiliikluses ning milliseid platvorme andmete kogumiseks ja visualiseerimiseks kasutatakse.

3.1 IoT platvormid

IoT platvormid on tarkvararaamistikud, mis pakuvad vahendeid ja teenuseid IoT seadmete haldamiseks, analüüsimiseks ja andmete integreerimiseks. IoT platvormid on loodud lihtsustamaks IoT rakenduste arendamist ja kasutuselevõttu, pakkudes standardiseeritud viisi erinevate seadmete ja andurite genereeritud andmete haldamiseks ja töötlemiseks.

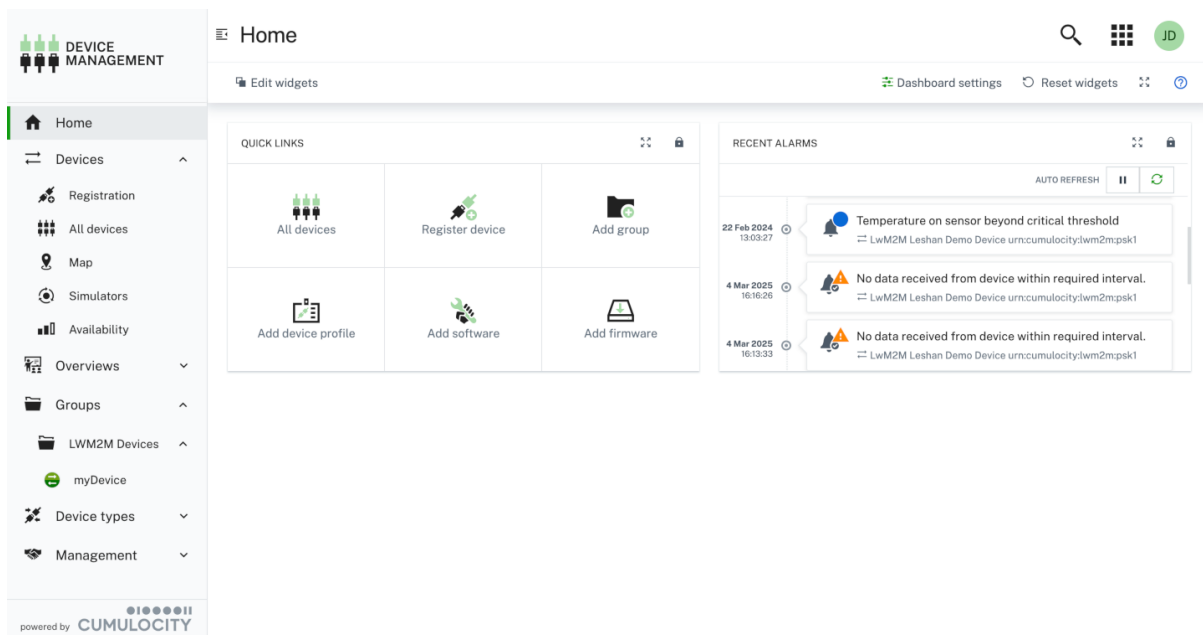
IoT platvormid sisaldavad tavaliselt erinevaid komponente, nagu seadme haldamise tööriistad, andmete haldamise tööriistad, analüütikatööriistad ja kommunikatsiooniprotokollid. Need komponendid töötavad koos, et võimaldada arendajatel seadmeid ühendada, andmeid hallata ja rakendusi luua, mis saavad kasutada IoT-seadmete genereeritud andmeid.

Lisaks seadmete ja andmete haldamisele pakuvad IoT platvormid ka turvalisuse ja privaatsuse funktsioone, et tagada IoT-seadmete genereeritud andmete turvalisus ja kaitstus. Nad võivad pakuda ka funktsioone nagu kaugjälgimine, ennustav hooldus ja reaajas hoiatused, mis aitavad organisatsioonidel optimeerida oma tegevusi ja suurendada tõhusust.

IoT platvormid pakuvad organisatsioonidele vajalikke vahendeid ja teenuseid laia valiku seadmete ja andurite genereeritud andmete haldamiseks, analüüsimiseks ja integreerimiseks.

Cumulocity

Cumulocity on pilvepõhine IoT platvorm, mis võimaldab organisatsioonidel ühendada, hallata ja analüüsida oma IoT seadmeid ja andmeid. See asutati 2010. aastal Düsseldorfis Saksamaal ja omandati 2017. aastal Software AG poolt [13].



Joonis 1. Cumulocity veebiliidese kuvatõmmis [14].

Platvorm pakub erinevaid IoT-ga seotud teenuseid (vt joonis 1), sealhulgas seadmete haldamine, andmehaldus, reaaliajase analüüsi ja rakenduste võimaldamine. Cumulocity abil saavad ettevõtted hõlpsasti ja turvaliselt ühendada oma seadmed pilvega ning hallata neid kaugjuhtimisega kesksest asukohast. See muudab ettevõtetel nende seadmete jälgimise ja juhtimise lihtsaks ning võimaldab koguda ja analüüsida nende andmeid.

Üks Cumulocity peamisi omadusi on selle seadmete halduse võimekus. Platvorm pakub ühtset vaadet kõigile ühendatud seadmetele, võimaldades kasutajatel neid reaaliajase hallata ja jälgida. Samuti pakub see tööriistu seadmete kaugkonfigureerimiseks ja värskendamiseks, mis teeb suure seadmete hulga haldamise lihtsaks.

Cumulocity pakub ka võimsaid andmehalduse ja analüüsi võimekusi. Platvorm suudab koguda andmeid laiast valikust allikatest, sealhulgas sensoritest, seadmetest ja teistest IoT süsteemidest. Seejärel kasutab masinõppealgoritme ja muid edasijõudnud analüütikatööriistu andmete muutmiseks väärtuslikeks teadmisteks, mida ettevõtted saavad oma tegevuse parandamiseks kasutada.

Lisaks pakub Cumulocity rakenduste võimaldamise võimekust, mis võimaldab ettevõtetel luua kohandatud rakendusi oma IoT seadmetele. Need rakendused võivad automatiseerida protsesse, parandada efektiivsust ja pakkuda uusi teenuseid klientidele.

Cumulocity üks võtme-eelistest on turvafunktsioonid. Platvorm on kavandatud olema väga turvaline, kaitstes andmeid ja seadmeid volitamata juurdepääsu eest täiustatud krüpteerimis- ja autentimismehhanismidega.

3.2 IoT Tartu bussiliikluses

Viimastel aastatel on asjade internet muutunud oluliseks komponendiks ühistranspordisüsteemide arendamisel üle kogu maailmas. Tartu on välja arendanud hästi toimiva ühistranspordivõrgu, mis põhineb peamiselt bussiliinidel ja elektriratastel. Süsteemi tõhususe, töökindluse ja reisijate rahulolu parandamiseks on rakendatud mitmeid IoT-lahendusi.

Üheks keskseks lahenduseks on GPS-seadmete paigaldamine bussidesse, mis võimaldab jälgida sõidukite asukohta ja liikumist reaalajas. See loob aluse marsruutide ja sõidugraafikute dünaamiliseks kohandamiseks vastavalt tegelikele liiklusoludele. Busside liikumist puudutavad andmed edastatakse Cumulocity IoT-platvormi, kus neid töödeldakse ja arhiveeritakse. Reaalajas kogutav info võimaldab pakkuda reisijatele täpsemat ja ajakohasemat teavet busside saabumiste ning liikumise kohta, mis on eriti oluline muutlike ilmastikutingimustega perioodidel, näiteks talvekuudel.

Käesoleva töö aluseks olev andmestik on kogutud Cumulocity platvormi kaudu ajavahemikus 11.10.2019 kuni 09.08.2024. Iga buss saatis ligikaudu iga 10 sekundi järel sündmuspõhiseid kirjeid (ingl *event*), mis sisaldasid mitmesuguseid tehnilisi ja logistilisi andmeid. Kirjeldatud kirjed sisaldavad muu hulgas seadme unikaalset identifikaatorit, sündmuse ajatempliga seotud informatsiooni ning bussi geograafilist asukohta. Lisaks edastatakse detailsed andmed konkreetse bussireisi kohta.

Kõige mahukam andmeväli on seotud järgmiste peatuste teabega. Iga alamüksus sisaldab info peatuse eeldatava saabumisaja, peatusesse saabumise hilinemine, eeldatava väljumisaja ja väljumise hilinemise kohta. Selline andmestruktuur võimaldab rekonstrueerida bussi eeldatava liikumisteedkonna kogu ülejäänud marsruudil, mis loob eeldused detailseks analüüsiks peatuste tasandil ning võimaldab hinnata ajagraafikute prognoositavuse täpsust.

IoT-tehnoloogia rakendamine Tartu ühistranspordisüsteemis on seega oluline töövahend linna transpordiplaneerimisel ja teenuse kvaliteedi parandamisel. Nutikad piletisüsteemid, GPS-seadmed, turvakaamerad ja reaalajas positsioneerimine võimaldavad paremat ülevaadet

liinivõrgu toimimisest ning aitavad tagada ohutu, mugava ja tõrgeteta liikumise nii linnaelanikele kui ka külalistele.

4. Metoodika

Käesolevas peatükis kirjeldatakse andmete kogumise, puhastamise ja analüüsimise etappe, mis viidi läbi Tartu linna busside hilinemiste ja peatumiste uurimiseks. Samuti tuuakse välja kasutatud tehnilised vahendid, tööriistad ja lähenemisviisid, mis võimaldasid andmestikust tuletada sisulisi järeldusi.

4.1 Andmeallikad

Vahemikus 11.10.2019–09.08.2024 edastasid Tartu linna bussid oma liikumisandmeid iga kümne sekundi tagant Cumulocity keskkonda. Ühes kuus kogunes keskmiselt ligikaudu 0,5GB andmeid, mis salvestusid .json-formaadis failidena. Aasta andmete salvestamiseks oleks olnud vaja 6GB ruumi ning vahemiku 2019-2024 töötlemata kujul andmete salvestamiseks ligikaudu 30GB ruumi. Andmete mahukusest tulenevalt oli otstarbekas töötlemise käigus neid esmalt õgvendada ning eemaldada kõik mittekriitilised väljad, et võimaldada analüüsi tõhusamat läbiviimist.

Andmete formaat oli otseselt seotud Cumulocity platvormi tööpõhimõttega. Nimelt edastavad busside sensorid oma sündmusi Cumulocity süsteemi kaudu ning nende andmete kättesaamiseks on võimalik kasutada üksnes REST (inglise keeles Representational State Transfer) API päringuid. 2025. aasta veebruaris laaditi Cumulocityst alla kogu ajalooline andmestik päevade kaupa tihendatud .zst failidena. Iga fail vastas ühele kuupäevale ning sisaldas kõigi Tartu busside selle päeva jooksul edastatud sündmusi. Failides oli kasutusel JSON Lines formaat, kus iga rida tähistas ühte JSON-objekti (vt joonis 2).

```
{"name": "Gilbert", "wins": [{"straight", "7♣"}, {"one pair", "10♥"}]}
```

```
{"name": "Alexa", "wins": [{"two pair", "4♠"}, {"two pair", "9♠"}]}
```

```
{"name": "May", "wins": []}
```

```
{"name": "Deloise", "wins": [{"three of a kind", "5♠"}]}
```

Joonis 2. Näide JSON Lines faili ülesehitusest [15].

Tihendatud failide vorminguks valiti .zst ehk Zstandard. Zstandard on Meta (endise nimega Facebook) loodud andmetihenduse algoritm, mille eesmärk on pakkuda samaaegselt kõrget pakkimissuhet ja väga kiiret dekompressiooni. Selle eelisteks on suur efektiivsus suurte failide puhul ning toetus mitmetele kaasaegsetele töötlusraamistikele.

Zstandard valiti selle töö jaoks välja järgmistel põhjustel:

- see pakub paremat tihendust kui tavapärased .zip või .gzip formaadid;
- failide hilisem lahtipakkimine on märgatavalt kiirem, mis võimaldab andmeid töödelda voona;
- formaat on hästi toetatud ka Pythonis (zstandard teek), võimaldades suurt hulka faile töödelda serverites.

Alternatiivsete formaatide (nt .zip või .tar.gz) kasutamine oleks tähendanud aeglasemat töötlemist ja suuremat kettaruumi kulu. Kuna andmemaht aastate lõikes oli suur (keskmiselt 10 MB päevas, s.t üle 3,5 GB aastas), siis osutus Zstandard ainsaks praktiliseks lahenduseks. Näiteid informatsiooni kohta, mida iga buss Cumulocitysse esitas on leitav all lisades (vt Lisa 1).

Info, mida iga bussi kohta Cumulocitysse edastatakse:

- **creationTime** – millal antud Cumulocity event loodi
- **source**
 - **name** – Cumulocity keskkonnas seadme nimi, millelt info tuli
 - **self** – link Cumulocity keskkonna seadmele, millelt info tuli
 - **id** – Cumulocity seadme id, millelt info tuli
- **type** - loodud event tüüp
- **lastUpdated** – millal antud sündmus viimati uuendati
- **self** – link sellele konkreetsele Cumulocity sündmusele
- **time** - kellaeg
- **id** – antud Cumulocity sündmuse id
- **text** – kirjeldab ära, missuguse sündmusega on tegemist
- **entities []**
 - **tripUpdate** – bussi teekonna informatsioon
 - **stopTimeUpdates []** – nimekiri, mis sisaldab endas iga peatuse kohta käivat informatsiooni
 - **stopSequence** – mitmenda peatusega on tegu
 - **arrival**
 - **delay** – mitu sekundit hilinetakse peatusesse
 - **time** – mis kellaks buss peaks peatusesse saabuma
 - **stopId** – peatuse id
 - **departure**
 - **delay** – mitu sekundit peatusest väljudes hilinetakse
 - **time** – mis kell buss peaks peatusest lahkuma
 - **trip**
 - **routeId** – bussi marsruudi id
 - **directionId** – bussi suuna id
 - **tripId** – antud reisi id
 - **startTime** – reisi alustamise kellaeg
 - **startDate** – reisi alustamise kuupäev
 - **id** – bussi id
 - **vehicle**
 - **currentStatus** – näitab, mis staatuses buss on
 - **stopId** – näitab, mis on peatus, kuhu buss sõidab
 - **position**
 - **bearing** – mis suunas buss liigub

- **latitude** – bussi laiuskraad
 - **speed** – bussi liikumiskiirus
 - **longitude** – bussi pikkuskraad
- **currentStopSequence** – näitab mitmenda stopSequence-i juures buss hetkel on
- **timestamp** – kellaeg UNIX formaadis
- **vehicle**
 - **id** – bussi id
 - **label** – näitab, kas buss hetkel edastab andmeid
- **header**
 - **gtfsRealTimeVersion** – protokoll, mille alusel andmeid edastatakse
 - **timestamp** - kellaeg
 - **incrementality** – näitab, kas edastatakse kõik andmed korraga või mitte

Need väljad võimaldavad hiljem määrata sõidu kulgemise ajajoone, hinnata saabumiste ja lahkumiste täpsust ning tuvastada hilinemisi.

Lisaks Cumulocity keskkonnast kogutud andmetele kasutati tulemuste visualiseerimise ja tõlgendamise etapis *stops.txt* faili, mis on kättesaadav peatus.ee GTFS-andmestiku kaudu³. See fail sisaldab kõigi Eesti ühistranspordipeatuste metaandmeid.

Info, mida iga bussi kohta *stops.txt* sisaldab:

- **stop_id** - peatuse id antud andmebaasis
- **stop_code** - peatuse id, sama mis busside edastatud `vehicle.stopId`
- **stop_name** - peatuse nimetus
- **stop_lat** - peatuse asukoht laiuskraadil
- **stop_lon** - peatuse asukoht pikkuskraadil
- **zone_id** - tsooni id
- **alias** - peatuse alternatiivne nimetus nt Tabasalu
- **stop_area** - peatuse asukoht, kas vald või linn
- **stop_desc** - peatuse kirjeldus, enamasti suuna jaoks
- **lest_x** - projektsioonikoordinaat (L-EST97)
- **lest_y** - projektsioonikoordinaat (L-EST97)
- **zone_name** - peatuse tsooni nimetus nt Kohila
- **authority** - peatuse vastutav võimuorgan nt Tartu MV

³ <https://peatus.ee/gtfs/>

Failis olev väli `stop_code` võimaldas seostada Cumulocityst saadud `stopId` väärtused tegelike peatuste nimedega. Kuigi see andmestik ei olnud osa analüüsitavast baasandmest, osutus see vajalikuks tulemuste hilisemal esitamisel, et muuta peatuste vahelised hilinemised arusaadavamaks ka visuaalselt ja sisuliselt.

4.2 Andmete puhastamine ja eeltöötlus

Andmeanalüüsiga alustamist oli esmalt vajalik andmete kättesaamine Cumulocity keskkonnast alla laaditud `.zst` failidest. Kuna failid olid tihendatud ja struktureeritud JSON Lines formaadis, tuli iga rida käsitleda kui eraldiseisvat sündmust (event). Failide sisu struktuuri mõistmiseks analüüsiti esmalt tüüpilist Cumulocity event-objekti, et määrata, millised väljad on analüüsi seisukohalt olulised.

Analüüsi käigus selgus, et kõige enam andmeruumi hõivas `stopTimeUpdates` väli. See väli sisaldas info loetelu kõigi busside järgmiste peatuste kohta, sh eeldatav saabumisaeg ja hiline mine igas peatuses. Lähemal vaatlusel ilmn es, et andmestruktuur dubleeris suurel määral hiline miste väärtusi: sagedasti arutati välja hiline mine ainult järgmisesse peatusesse ning sama väärtus kanti edasi ka kõigile järgnevaile peatustele. Sellest tulenevalt ei olnud kõigi järgnevate peatuste prognoosid analüütilises mõttes kasulikud, kuna need ei pakkunud uut ega täiendavat teavet võrreldes esimese peatuse ennustusega.

Kuna kõik järgnevate peatuste hiline miste prognoosid olid otseselt sõltuvad eelneva peatuse hiline misest ning sisuliselt kordasid sama väärtust, otsustati andmemahtu optimeerides jätta alles ainult järgmise peatuse hiline mise ennustus. See võimaldas oluliselt vähendada töödeldava andmestiku mahtu ilma, et see mõjutaks analüüsi kvaliteeti või tulemuste usaldusväärsust.

Algsetest andmeväljadest eemaldati järgmised väljad, mis näitasid infot Cumulocity sündmuse kohta, kuid mitte bussi hiline mise või liikumise kohta. Niisugused väljad ei olnud vajalikud Tartu bussiliikluse hiline mise analüüsi jaoks: `creationTime`, `source`, `type`, `lastUpdated`, `self`, `id`, `text`, `bearing` ja `header`. Alles jäeti üksnes need väljad, mis võimaldasid arvutada reaals et saabumisaega, hinnata hiline misi ning ühendada andmed sõidu- ja peatusinfoga.

Puhastatud andmekogu struktuur näeb välja järgnev (vt Lisa 2):

- **start_date** - kuupäev, millal reis algas (vormingus AAAA-KK-PP)
- **trip_id** - reisi identifikaator antud andmestikus
- **route_id** - liini identifikaator, millega reis seotud on

- **direction_id** - suuna ID, tavaliselt 1 või 2
- **start_time** - plaanitud väljumisaeg reisi alguses (HH:MM:SS)
- **vehicle_id** - unikaalne sõiduki ID süsteemis
- **vehicle_label** - bussi numbrimärk
- **current_stop_sequence** - peatuse järjekorranumber liinil, kuhu buss parasjagu liigub
- **vehicle_timestamp** - aeg, millal sõiduki andmed edastati (UNIX timestamp)
- **status** - staatuse väli, nt kas buss on liikumas, lähenemas või peatunud peatuses
- **latitude** - sõiduki asukoha laiuskraad
- **longitude** - sõiduki asukoha pikkuskraad
- **bearing** - suund kraadides, kuhu sõiduk liigub (0–359)
- **speed** - sõiduki kiirus m/s
- **stop_id** - peatuse ID antud andmebaasis (ühilduv GTFS `stops.txt` faili stop_code-iga)
- **stop_sequence** - peatuse järjekorranumber marsruudil
- **arrival_delay** - saabumise hiline mine sekundites võrreldes graafikuga
- **arrival_time** - planeeritud saabumisaeg peatusesse (UNIX timestamp)
- **departure_delay** - väljumise hiline mine sekundites võrreldes graafikuga
- **departure_time** - planeeritud väljumisaeg peatusest (UNIX timestamp)

Töö käigus tuvastati andmestikus ka mitmeid ebanormaalseid väärtusi. Näiteks esines juhtumeid, kus busside hiline mine oli süsteemis märgitud mitmetunnisena, mis ei olnud liiklusolusid ja sõidugraafikut arvestades realistlik. Selliste kõrvalekallete vältimiseks rakendati andmete puhastamisel filter, mille alusel kaasati analüüsi ainult need juhtumid, kus hiline mine või varasem saabumine jäi 15 minuti (900 sekundi) piiresse. See lävend võimaldas välistada väärtused, mis olid tõenäoliselt põhjustatud süsteemiveast, kommunikatsioonihäiretest või muust tehnilisest anomaaliast.

Aasta 2023 andmestiku põhjal moodustasid need äärmuslikud juhtumid 0,72% kõigist ridadest. Kokku 86 227 078 reast esines 621 269 juhul olukord, kus buss oli märgitud hiline ma või saabuma rohkem kui 15 minutit varem võrreldes planeeritud ajaga.

Pärast 09.08.2024 kuupäeva ei edastatud Tartu linna bussid enam uusi andmeid Cumulocity platvormi, kuna linn vahetas andmete edastamise platvormi. Cumulocity keskkonnast laeti 2025. aasta veebruaris alla kogu olemasolev busside andmestik. Andmed salvestati päevade kaupa .zst formaadis failidena, kus iga rida esindas ühte JSON-objekti.

Selline struktuur oli töötlemise seisukohalt kasulik, kuna võimaldas andmeid lugeda ja töödelda järjestikku ridade kaupa ilma et kogu failimaht oleks tulnud korraga mällu laadida. See võimaldas suurte andmekogumite analüüsi ka piiratud ressursiga keskkondades näiteks Google Colabis.

4.3 Tehniline lähenemine

See peatükk kirjeldab tööriistu ja meetodeid, mida kasutati busside andmete analüüsimiseks.

4.3.1 Google Colab

Andmete esmaseks töötlemiseks kasutati Google Colabi keskkonda. Colab on pilvepõhine Jupyter Notebooki keskkond, mis võimaldab käivitada Python-koodi otse veebibrauseris. Selle suurim eelis oli võimalus töödelda suuremahulisi andmeid ilma kohalikku arvutit koormamata. Kuna failid olid suured ja töötlusprotsessid mitmeastmelised, võimaldas Colab mugavalt kasutada mälumahukaid andmeid ja paralleelselt testida erinevaid töövõtteid.

4.3.2 Ubuntu server

Viimases andmetöötlustetapis kasutati ka lokaalselt seadistatud Ubuntu serverit. Server võimaldas töödelda suuremaid andmehulki stabiilses keskkonnas ja jooksutada skripte, mille täitmine Colabi sessiooni piirangu tõttu katkeda võis. Lisaks kasutati serverit .zst failide süsteemseks haldamiseks ja ajaliselt mahukamate töötlustsüklite läbiviimiseks. Ubuntu terminali tööriistu ja automatiseeritud töövooge kasutati eeltöödeldud andmete teisaldamiseks ja struktureerimiseks.

4.3.3 PyCharm

Lõpliku andmetöötluste automatiseerimine ning korduvate skriptide struktureeritud haldamine toimus PyCharm arenduskeskkonnas. PyCharm võimaldas mugavat versioonihaldust, failide ja kaustade süsteemset organiseerimist ning skriptide modulaarsuse paremat kontrolli. See oli eriti oluline andmevoogude töötluste skriptide puhul, kus oli vaja töödelda sadade failide kaupa andmeid kaustapõhiselt.

Töö käigus oli vajalik töödelda andmeid ka kaugserveris, millele ligipääs toimus SSH gateway kaudu. PyCharm võimaldab seadistada ühenduse Remote Development funktsiooni abil, mille kaudu sai redigeerida ja käivitada skripte otse serveris ilma failide lokaalset kopeerimist vajamata. Selline lähenemine tagas sujuva töövoogu pilves paiknevate andmekogumite töötlemisel ning võimaldas serveripõhist ressursikasutust ilma Colabi piiranguteta.

4.3.4 Python

Töö peamine programmeerimiskeel oli Python, kuna see sobib hästi andmetöötluks, failiformaatide käsitlemiseks ning võimaldab kasutada võimsaid analüüsiteeke. Python võimaldas üles ehitada kogu töölustorustiku, alates .zst failide lahtipakkimisest kuni puhastatud CSV-failide salvestamiseni. Samuti võimaldas Python automatiseerida korduvaid andmeanalüüsi samme ja hallata andmete eeltöötlust etapiviisiliselt. Litsents

4.3.5 Pandas

Andmete analüüsi ja puhastamise tuumikuks oli Pandas teek. Selle abil loeti JSON-ridu sõnastikeks, teisendati need tabelkujulisteks andmeraamistikeks (DataFrame), filtreeriti välja ebaolulised või vigased kirjed ning arvutati vajalikke ajavahemikke ja hilinemisi. Pandase abil seoti ka erinevad andmeallikad (nt stopId ja peatuse nimi) ning moodustati lõplikud CSV-tabelid analüüsiks.

4.3.6 Tehisaru kasutamine

Töö käigus kasutati ChatGPT (versioon GPT-4 [2]) tehisaru abivahendina Python-skriptide loomisel ja andmetöötluks automatiseerimisel. ChatGPT aitas kirjutada skripte, mis töötlesid GTFS ja Cumulocity andmeid, filtreeris andmeid, arvutasid saabumisaegade erinevusi ning sidusid peatusi marsruudiandmete alusel.

Esitatud küsimused olid näiteks:

- „Kuidas eemaldada Pandases read, kus tegelik ja ennustatud saabumisaeg erinevad üle 15 minuti?“;
- „Kuidas määrata eelmine ja järgmine peatus GTFS marsruudi järgi?“;
- „Kirjuta skript, mis käib läbi kõik kaustas olevad .zst failid ja salvestab need CSV-d.“.

Tehisaru toetas tehniliste lahenduste leidmist ja kiirendas töövoogu, võimaldades keskenduda analüüsile. Kõik tehisaru loodud koodide sisu on autori poolt üle kontrollitud.

5. Tartu busside analüüs

Käesolevas peatükis esitatakse Tartu linna bussiliikluse andmete analüüs. Analüüs tugineb aastatel 2019–2024 kogutud andmestikule, mille hulka kuuluvad reaajas edastatud sündmused Cumulocity platvormilt ja ametlikud sõiduplaaniandmed GTFS-formaadis.

Metoodika protsessis olid mitmed olulised etapid. Esmalt uuriti andmeid, et mõista nende struktuuri ja potentsiaalseid analüüsivõimalusi. Seejärel toimus andmete puhastamine ja esialgne analüüs, mille käigus tuvastati olulised ebanormaalsed väärtused ning määrati analüüsi lähtekohad. Koostöös Tartu linna esindajatega täpsustati uurimisküsimused, mis olid vajalikud töö eesmärkide saavutamiseks.

Pärast seda keskenduti suurte andmemahtude töötlemisele, mis hõlmas andmete struktureerimist, puhastamist ja analüüsiks vajalike vaheandmete loomist. Seejärel viidi läbi andmete analüüs, mille tulemused on jaotatud vastavalt uurimisprotsesside erinevatele etappidele ja esitatud vastavates alapeatükkides. Analüüsi lõpus tehti tulemuste valideerimine, et hinnata analüüsitulemuste usaldusväärsust, lähtudes nii andmete kvaliteedist kui ka analüüsimeetodite sobivusest.

Andmetöötlemise ja visualiseerimise automatiseerimiseks kasutati tekstipõhise tehisintellekti tööriista ChatGPT (mudel GPT-4), mille abil koostati skripte analüüsi läbiviimiseks. Tekstiroboti abi kasutati eelkõige visualiseeringute loomiseks, andmeväljade töötlemiseks ning väiksemate CSV-failide genereerimiseks. Näiteks küsiti soovitusi, kuidas teatud veergudega andmestikust koostada sobivaid graafikuid või kuidas optimeerida andmete filtreerimisprotsesse. Kõik loodud skriptid on kasutaja poolt kontrollitud ja vajadusel muudetud vastavalt töö eesmärkidele.

5.1 Koosolek Tartu esindajatega

Enne andmete analüüsimist toimusid kaks koosolekut Tartu linna esindajatega. Kohtumised toimusid töö planeerimise algfaasis, enne andmete lõplikku puhastamist ja analüüsi. Nende kohtumiste eesmärk oli välja selgitada, missugused oleksid Tartu linna peamised vajadused seoses ühistranspordi andmete analüüsiga selleks, et töö oleks praktiline väljund Tartu linna vajadustega.

Esimesel kohtumisel viibisid linnatranspordi juhataja Roman Meeksa, Tartu välisprojektide projektijuht Jaanus Tamm, Tartu andmehalduse juht Getter Kartau, töö juhendaja Pelle Jakovits ning Hanna Maria Mägi.

Teine kohtumine keskendus rohkem tehnilistele küsimustele ning seal osalesid veel lisaks ühistranspordi peaspetsialist Madis Oona ning Tartu ühistranspordi kvaliteedijuht Oskar Veversiga.

Koosolekute tulemusel selgus, et Tartu linnas on juba olemas spetsialistid, kes analüüsivad ning korrigeerivad Tartu bussiaegu kaks korda aastas: septembris ja jaanuaris. Nende analüüs keskendub põhiliselt liinide keskmiste hilinemiste võrdlemisele. Sellest tulenevalt tekkis ettepanek, et käesolev lõputöö võiks keskenduda mitte ainult olemasoleva olukorra kirjeldamisele, vaid ka muutuste ja redigeeringute analüüsimisele ajas. Lisaks selgus, et Tartu linnal puudub hetkel hea ülevaade, kui kaugel bussid tegelikult oma peatustest peatuvad. Tartu linna eesmärk on muuta ühistransport võimalikult ligipääsetavaks vaegnägijatele ja teistele erivajadustega reisijatele. Busside tegelike peatumiskohtade täpsus on oluline mitte üksnes ligipääsetavuse vaatenurgast, vaid ka näiteks navigatsioonirakenduste, automaatse reisijate teavitamise ning liinide optimeerimise seisukohalt. Täpsem arusaam busside positsioneerimisest aitab kaasa terviklikuma reisijakogemuse loomisele.

Kohtumised võimaldasid luua tööle praktilise ja rakendusliku aluse, mis lähtus otseselt Tartu esindajate vajadustest. Samuti aitasid need seada realistlikke ja mõõdetavaid eesmärke, mida on võimalik tulevikus Tartu linna transpordisüsteemi arendamisel rakendada. Koosolekud aitasid paika panna antud töö uurimisküsimused.

5.2 Uurimisküsimused

Andmete eelanalüüsi ning Tartu linna esindajate koosoleku tulemusena pandi paika järgnevad uurimisküsimused.

K1: Kuidas on sõidugraafikute regulaarne muutmine septembris ja jaanuaris mõjutanud busside hilinemisi?

Tartu linnas uuendatakse bussigraafikuid kaks korda aastas, eesmärgiga paremini vastata muutuvatele sõiduvajadustele. Selle küsimuse eesmärk on analüüsida, kas need graafikumuudatused on aidanud vähendada hilinemisi. Uurides hilinemiste mustreid enne ja pärast graafikute jõustumist, saab hinnata graafikumuudatuste tõhusust. See on oluline transpordikorralduse optimeerimiseks ja reisijate rahulolu tagamiseks.

K2: Kas busside hilinemiste ennustamine vastab reaalsusele?

Paljudes süsteemides kuvatakse reaajas ennustusi busside saabumise kohta, kuid nende täpsus võib varieeruda. See küsimus võrdleb ennustatud ja tegelikke saabumisaegu, et hinnata

süsteemi usaldusväärsust. Täpsed ennustused on reisijatele kriitilise tähtsusega, sest ebatäpsed ennustused mõjutavad negatiivselt ühistranspordi kasutaja kogemust ja usaldust.

K3: Millised liinid on kõige ebausaldusväärsemad ajast kinni pidamise osas?

Tartu bussiliinide täpsus varieerub, kuid mõned liinid hilinevad järjepidevamalt kui teised. Selle küsimuse eesmärk on välja selgitada, millised liinid on kõige sagedamini graafikust maas ning kui suured on nende hilinemised keskmiselt. Tulemused aitavad linnal paremini suunata tähelepanu just nendele liinidele, kus ajast kinnipidamine on suurim probleem.

K4: Millistes peatustes on hilinemine kõige suurem?

Erinevates peatustes võivad bussid hilineda erineval määral, sõltuvalt liiklusoludest, fooride olemasolust või liini eripäradest. Selle küsimusega uuritakse, millised peatused on kõige rohkem hilinemistega seotud. Nende tuvastamine aitab tuua esile kriitilised kohad linnaliikluses, kus oleks kasulik muuta graafikuid või parendada taristut.

K5: Kui kaugel peatuvad bussid peatustest?

Busside peatumise täpsus on oluline reaalajas info ja reisijakogemuse seisukohalt. See küsimus keskendub sellele, kui suured on erinevused busside tegeliku peatumiskoha ja ametliku peatuse asukoha vahel. Suured hälbed võivad viidata GPS-andmete ebatäpsusele või praktilistele probleemidele peatumiskohtade juures ning mõjutada usaldust reaalaja süsteemide vastu.

K6: Kui kvaliteetsed on busside poolt edastatud GPS andmed? Kas leidub varieeruvust erinevate busside seadmete kvaliteedis?

Reaalajas jälgitavate busside asukoht põhineb GPS-andmetel, mille täpsus ja usaldusväärsus on kriitilise tähtsusega nii lõppkasutaja kui ka andmeanalüüsi seisukohalt. Käesolevas töös on üheks uurimisküsimuseks hinnata, kui kvaliteetsed on busside poolt edastatud GPS-andmed ning kas erinevate sõidukite lõikes esineb varieeruvust andmete täpsuses või usaldusväärsuses. Sellise varieeruvuse olemasolul võib see viidata andmeside probleemidele, bussidesse paigaldatud seadmete tehnilisele seisukorrale või muudele süsteemsetele teguritele. Andmekvaliteedi hindamine aitab kaasa üldise süsteemitöökindluse ja analüüsitulemuste usaldusväärsuse parendamisele.

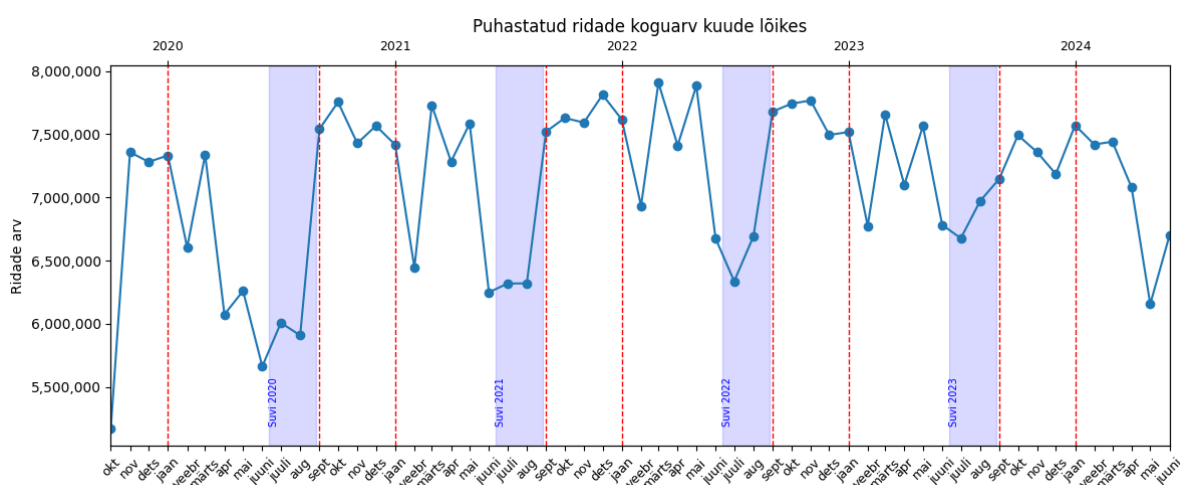
5.3 Andmed peale eeltöötlust

Pärast andmete puhastamist ja struktuuri ühtlustamist on võimalik andmestikku analüüsida. Selles peatükis antakse ülevaade eeltöödeldud andmete üldisest mahust ja struktuurist kuude

lõikes. Tuuakse esile mustrid, hooajalisus ning võimalikud anomaaliad, mis võivad mõjutada edasist analüüsi. Andmete analüüsiks kasutatud koodi repositooriumi link on lisades (vt Lisa 3).

Joonisel 3 on kujutatud Cumulocityst kogutud andmemaht ajavahemikus 11.10.2019 kuni 30.06.2024. Sinise taustaga on tähistatud busside suvegraafiku kehtivusperioodid ning punaste joontega kuud, mil Tartu linn on sõidugraafikuid üle vaadanud. Jooniselt on näha, et suvekuudel, kui kehtib lühendatud sõidugraafik, on andmete hulk märgatavalt väiksem. Keskmiselt koguneb sel perioodil iga kuu ligikaudu miljon kirjet vähem võrreldes tavagraafiku kuudega. See tuleneb tõenäoliselt väiksemast liinide arvust ja harvematest väljumistest, millega tuleb ka hilisemates analüüsides arvestada.

Vaatlusalustest aastatest on kõige väiksem andmemaht 2020. aastal, kuna andmete kogumine algas alles oktoobris. Aastad 2021 kuni 2023 on andmehulgalt ühtlasemad. Nende aastasisene varieeruvus püsib sarnasel tasemel. Igast aastast eristuvad selgelt veebruar ja aprill, mil andmete kogus on järjepidevalt madalam kui ülejäänud kuudel. See võib viidata pühade, koolivaheaegade või seadmete katkestustega seotud mõjudele.



Joonis 3. Puhastatud andmete hulk vahemikus 11.10.2019–30.06.2024.

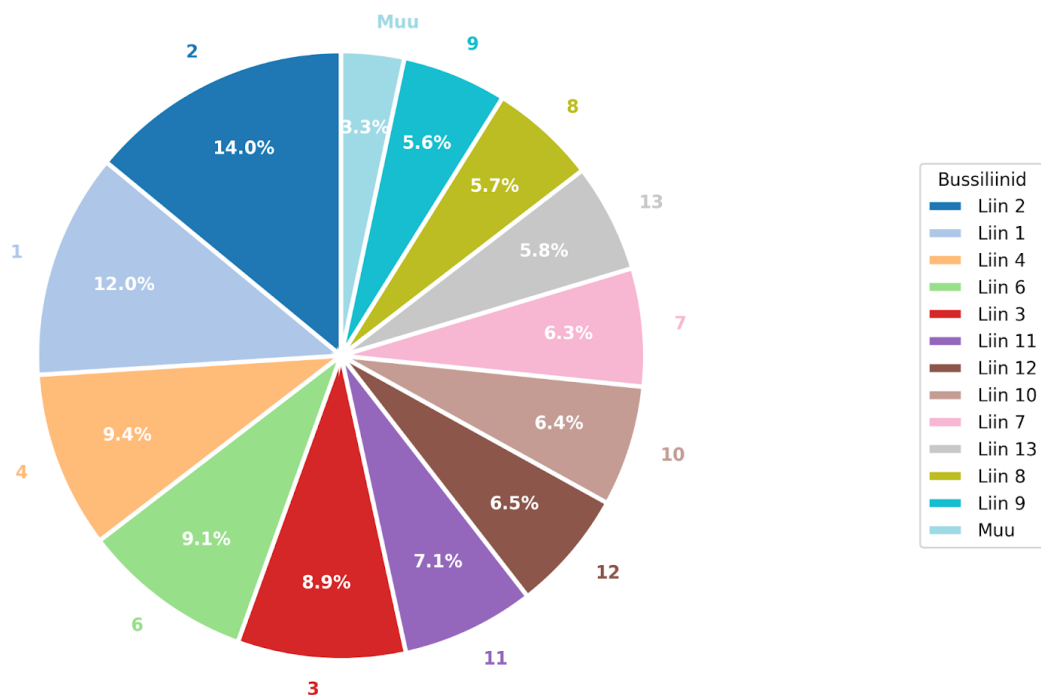
Järgnevalt vaadeldi, millised bussiliinid on andmestikus esindatud ning kui palju ridu nende kohta on kogutud. Kokku tuvastati 39 erinevat liinitähist, millest enamikul on määratletud kaks suunda. Esineb ka erandeid. Näiteks liin 3 sisaldab kolme suunda ning liinidel nagu 31R, 69A, 9A ja E1 on tähelised või muul viisil eristuvad tähised, mis viitavad laiendatud või spetsiaalsetele marsruutidele. Erandlik on ka Sotsiaaltransport, millel puudub tavaline

suunajaotus (*direction_id* = 0). Liinide esindatus andmetes on ebäühtlane, osa liinidest esineb märkimisväärselt rohkem kui teised, mis võib tuleneda nende sagedasemast kasutusest.

Joonisel 4 selgub, et andmetes esinevad kõige rohkem liinid 2, 1, 4 ja 6, millest igaüks moodustab üle 9% kogu andmestikust. Kõige silmapaistvam on liin 2, mis moodustab umbes 14% kõigist andmeridest. Järgneb liin 1 (12%), liin 4 (9,4%) ja liin 6 (9,1%). Need liinid on Tartu põhiliinid, mida kasutatakse sagedamini, millel on rohkem peatusi, tänu millele saadab iga buss rohkem andmeid andmebaasi.

Väiksema osakaaluga liinid, mille osakaal jääb alla 2%, on grupeeritud “Muu” alla. Nende hulka kuuluvad näiteks liinid 69A, E1 ja sotsiaaltransport, mis kokku moodustavad vaid umbes 3,3% andmetest.

Joonisel 4 on selgelt näha, kuidas suur osa andmetest koondub vähestele liinidele, mis peegeldab ka linna liiklusvoogude koondumist kindlatele marsruutidele. See jaotus aitab edasises analüüsis fokuseerida just suure andmemahuga liinidele, mille hilinevismustrid on statistiliselt usaldusväärsemad.



Joonis 4 . Bussiliinide osakaal kogutud andmetes.

5.4 Tulemused

Käesolevas peatükis esitatakse töö käigus läbi viidud analüüside olulisemad tulemused. Iga alapeatükk vastab ühele uurimisküsimusele ning toob välja vastavad järeldused kas visuaalse analüüsi või arvutuslike näitajate põhjal. Tulemuste tõlgendamisel võetakse arvesse nii andmestiku tehnilisi piiranguid kui ka linna transpordisüsteemi eripärasid.

5.4.1 Kuidas on sõidugraafikute regulaarne muutmine septembris ja jaanuaris mõjutanud busside hilinemisi?

Üheks käesoleva töö eesmärgiks oli analüüsida, kas busside hilinemistes esineb ajaliselt korduvaid mustreid, mis oleksid seotud Tartu linna ühistranspordi sõidugraafikute muudatuste perioodidega. Linna esindajatega toimunud kohtumistel toodi välja, et sõiduaegade ja liinivõrgu ülevaatus toimub kaks korda aastas: jaanuari alguses ja septembri alguses.

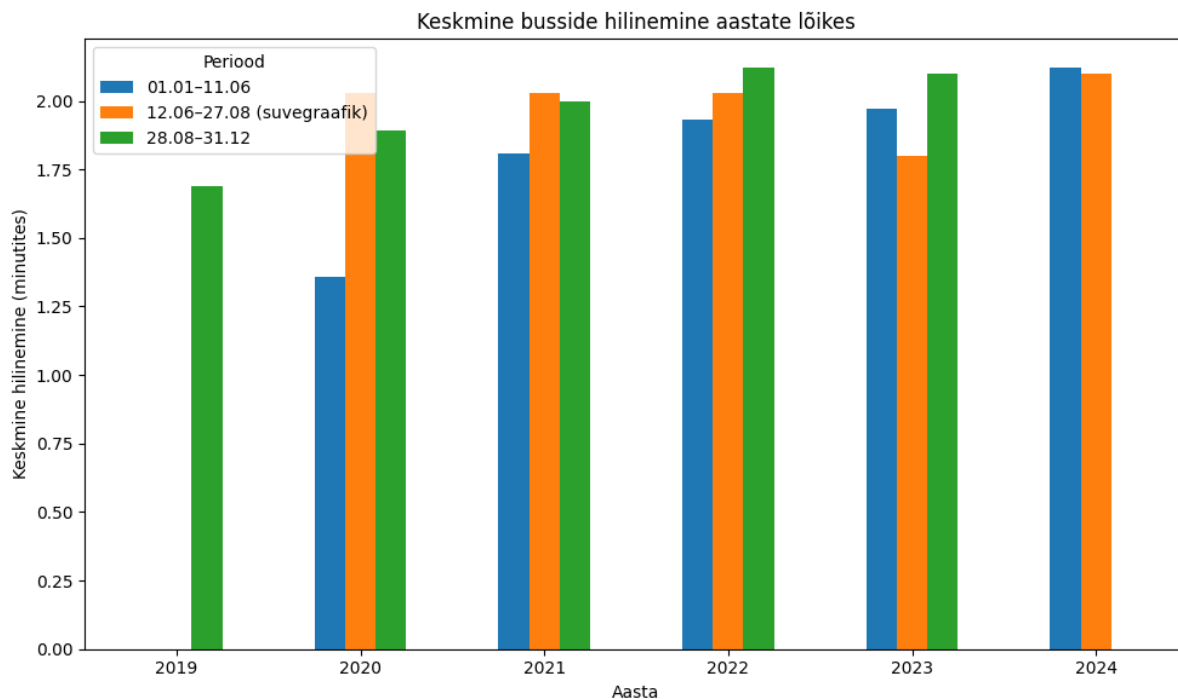
Selle põhjal jaotati aastad kahte võrdlevasse ajavahemikku:

jaanuar-august - graafiku muudatus jaanuarist;

september-detsember - graafikumuudatus septembris.

Andmete eelanalüüsi tõttu toodi välja ka busside suvegraafiku vahemik 12.06-27.08, et tavagraafikul sõitvad bussid ei mõjutaks jaanuar-augusti vahemiku keskmiste hilinemiste analüüsi.

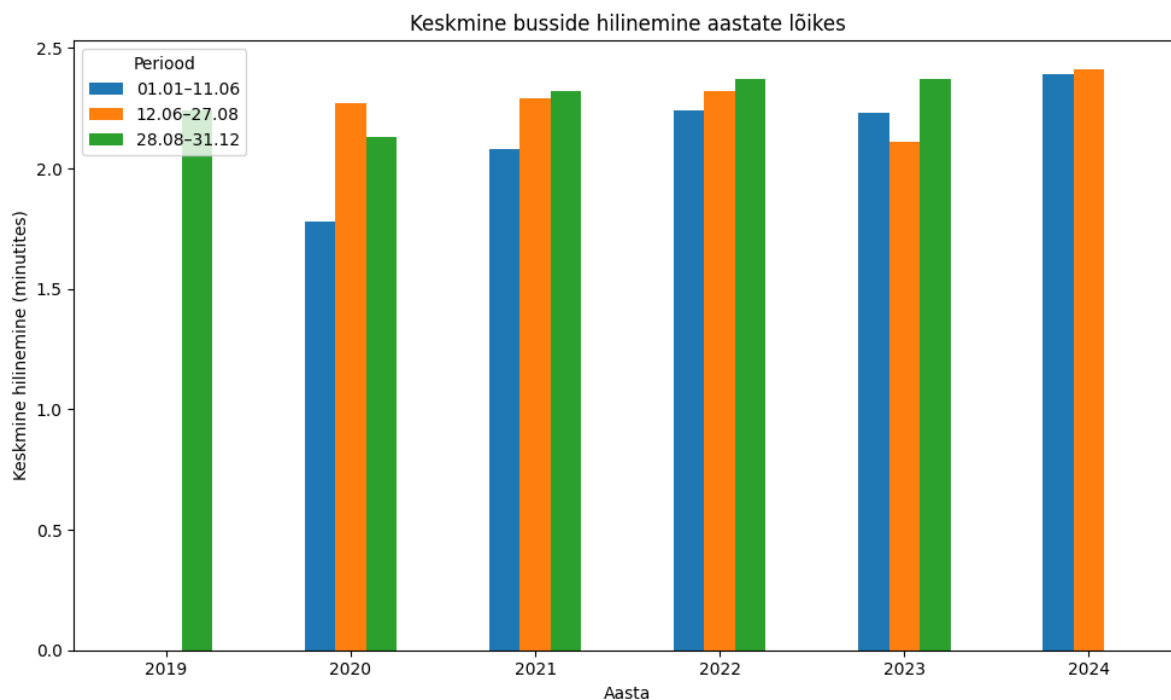
Joonisel 5 on kujutatud busside keskmisi hilinemisi minutites kolme perioodi lõikes aastatel 2019–2024. Keskmise arvutamisel on võetud aluseks peatustesse jõudmise aeg tervikuna: kas buss jõudis varem, täpselt või hiljem ja selle põhjal on saadud keskmine. Tulemused näitavad, et septembri algusest aasta lõpuni on hilinemised järjepidevalt suuremad kui aasta esimeses pooles. See tendents kordub kõigil vaadeldud aastatel, viidates sellele, et graafiku muutustele järgnevatel kuudel võib bussiaegade prognoositavus ajutiselt langeda. Septembris võib ühe põhjusena välja tuua ka selle, et algab kool, puhkuste periood saab läbi ning linnaliikluses on pärast suvekuid rohkem autosid. See omakorda võib põhjustada viivitusi busside ajagraafikutest kinnipidamisel. Graafikult tuleb ka välja, et keskmiselt on aastast aastasse busside hilinemine kasvanud 15 sekundi võrra.



Joonis 5 . Keskmsed busside hiline mised aastate lõikes vahemikus 2019–2024.

Joonisel 6 on esitatud üksnes need andmepunktid, kus buss saabus peatusale hiljem kui sõiduplaanis ettenähtud aeg, ehk käsitletud on ainult tegelikke hiline mised. Andmestik ei hõlma täpselt ega varem saabunud busse. Analüüsist ilmneb, et perioodil septembrist detsembrini on keskmised hiline mised kõigil vaadeldud aastatel suuremad kui aasta esimeses pooles, mis võib olla seotud suvepuhkuste lõppemise, koolide alguse ja sellest tuleneva liikluskoormuse kasvuga.

Märkimisväärne on ka see, et aastate jooksul on iga ajaperioodi keskmine hiline mine kasvanud. Eriti selgelt on seda näha aasta esimeses pooles (sinine tulp), kus keskmine hiline mine on tõusnud 1 minuti ja 45 sekundi pealt 2 minuti ja 15 sekundini. Joonise põhjal võib järeldada, et kui buss ei jõua peatusesse õigeaegselt, siis on keskmine hiline mised pikkus ligikaudu 2 minutit. See on Tartu linna poolt lubatud 3 minutilise vahemiku sees.



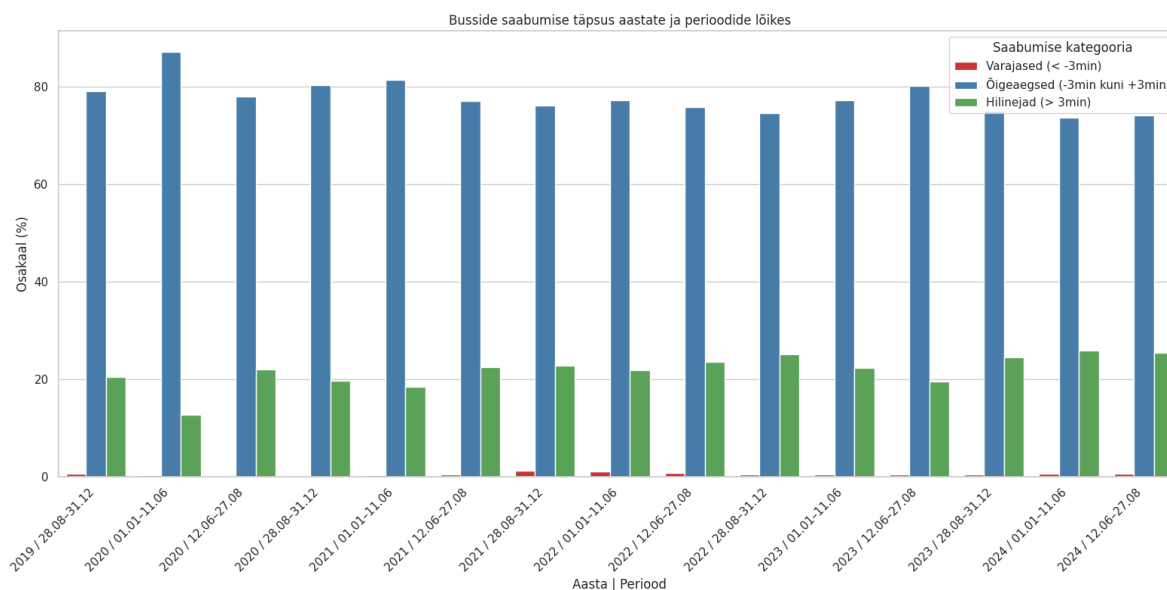
Joonis 6. Keskmesed busside hiline mised perioodil 2019–2024 ainult hiline misi võrreldes.

Joonisel 7 on kujutatud busside saabumise täpsus ajavahemikus 2019 kuni 2024, jagatuna kolme kategooriasse: varajased, õigeaegsed ja hilinejad. Õigeaegseteks on loetud bussid, mis saabusid peatustesse kuni kolme minuti jooksul enne või pärast graafikus olnud aega, lähtudes Tartu linna poolt kehtestatud hiline mispiirist.

Tulemused kinnitavad varasematel joonistel ilmnenud mustrit: perioodil september kuni detsember on hilinejate osakaal järjepidevalt suurem kui aasta esimeses pooles. Üllatuslikult moodustavad üle kolme minuti hiline mised märkimisväärse osa kõigist hiline mistest. Kuna analüüsist on eelnevalt välja jäetud juhtumid, kus hiline mine ületas 15 minutit, võib järeldada, et suur osa bussidest hilineb vahemikus kolm kuni viisteist minutit.

Eriti silmatorkav on üle kolme minuti hiline miste kasv aasta esimeses pooles. Näiteks 2023. aastal hilines ligikaudu 20 protsenti bussidest oma peatustesse rohkem kui kolm minutit, mis tähendab, et pea viiendik bussidest ei püsinud kehtestatud ajapiirides.

Joonis 7 näitab samuti, et busside varajasi saabumisi esineb äärmiselt harva. See viitab sellele, et hiline mised on süsteemselt kaldu hilisema saabumise suunas. Samuti püsivad suvise graafiku hiline mised aastate lõikes suhteliselt stabiilsena, välja arvatud 2023. aastal, mil täheldati suuremat varieeruvust.



Joonis 7. Busside saabumise täpsuse protsendiline jaotus peatusesse saabumise põhjal.

Analüüsi tulemused näitavad, et Tartu linna ühistranspordi sõidugraafikute regulaarsetel muudatustel jaanuaris ja septembris ei ole olnud märgatavat mõju busside hilinemiste vähenemisele. Kuigi graafikuid korrigeeritakse kaks korda aastas, on hilinemised nii aasta alguses kui ka lõpus püsivad stabiilselt kõrgel tasemel ning mitmel aastal on näha isegi kasvutrendi. Kõikidel vaadeldud aastatel on hilinemised suuremad perioodil september kuni detsember, mis järgneb sügisesele graafikumuudatusele. Seda võib seletada suurenenud liikluskoormuse ja kooliaasta algusega, kuid fakt, et hilinemiste tase on püsunud või suurenenud ka pärast muudatusi, viitab sellele, et senised graafikumuudatused ei ole olnud piisavalt tõhusad. Samuti nähtub andmetest, et aastate lõikes on kasvanud keskmine hilinemine ka aasta esimeses pooles. Eriti ilmneb see 2023. aasta andmetest, kus ligikaudu 20 protsenti bussidest hilines rohkem kui kolm minutit, ületades Tartu linna poolt kehtestatud ajapiiri. Kokkuvõttes võib järeldada, et graafikumuudatustele ei ole järgnenud süsteemset paranemist hilinemiste osas, mistõttu tuleks senist lähenemist graafikute koostamisele ja valideerimisele kriitiliselt hinnata.

5.4.2 Kas busside hilinemiste ennustamine vastab reaalsusele?

Hilinemiste ennustamise täpsuse hindamiseks võrreldi aastate lõikes busside prognoositud ja tegelikke saabumisaegu. Prognoositud hilinemine tähistab bussi poolt edastatud eeldatavat saabumisaega järgmisesse peatusesse ning see väärtus saadetakse tavaliselt kohe pärast eelmisest peatusest lahkumist. Tegelik hilinemine seevastu põhineb bussi tegelikul saabumisel sihtpeatusesse, mille kohta saadetakse uus sündmus koos reaalse saabumisaega.

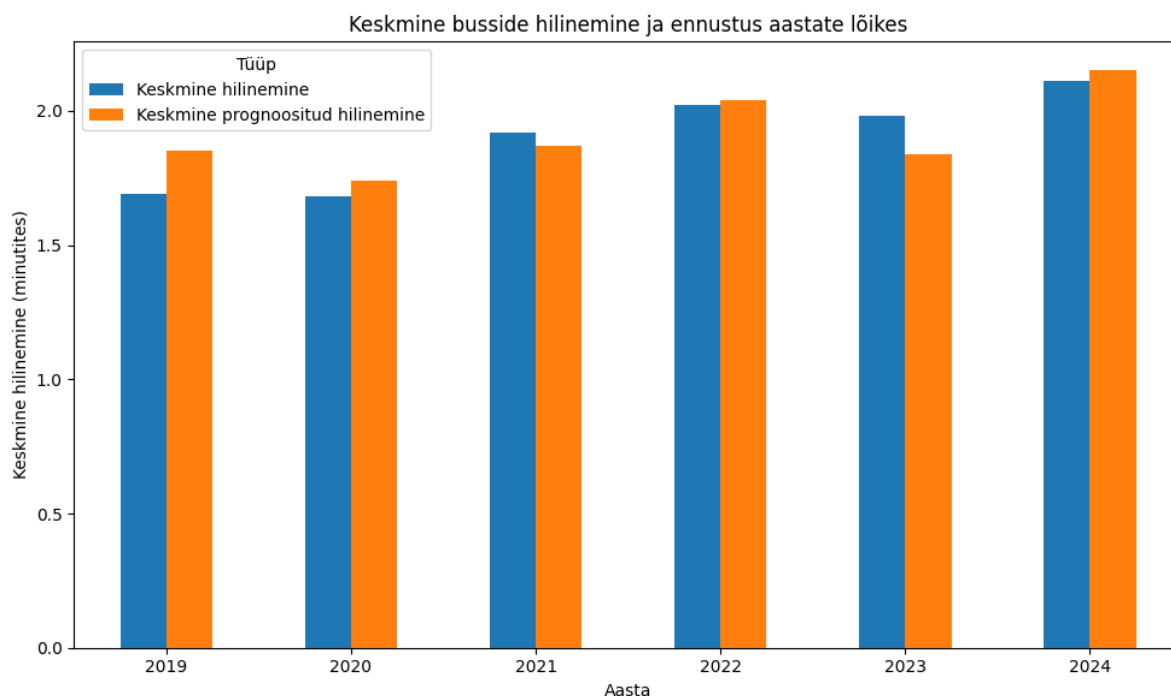
Näiteks olukorras, kus buss lahkub Riimäe peatusest ja liigub Pauluse peatuse suunas, arvutab süsteem lahkumise hetkel eeldatava saabumisaja Pauluse peatusesse ning edastab selle kui prognoositud hilinemise. Kui buss jõuab Pauluse peatusesse, saadetakse uus sündmus, mis kajastab bussi tegelikku saabumisaega. Nende kahe andmepunkti võrdlemisel on võimalik hinnata, kui täpne oli algne prognoos ning tuletada hilinemise ennustamise kvaliteet.

Prognoositud ja tegelike saabumisaegade tuletamiseks grupeeriti ühe päeva andmestik `trip_id` ja `vehicle_id` alusel, mille tulemusel saadi unikaalsed bussireiside kaupa grupeeritud andmekogumid. Iga reisi sees tehti täiendav grupeerimine `stop_id` väärtuse järgi, mis võimaldas käsitleda iga peatust eraldi.

Peatusepõhistes gruppides moodustasid andmejadad, mis algasid info edastusega pärast eelmisest peatusest lahkumist ja lõppesid sihtpeatusesse jõudmisega. Eelmisest peatusest lahkudes oli võimalik määrata järgmise peatusesse jõudmise prognoositud aeg, kui vaadata *delay* veeru väärtust eelmisest peatusest lahkudes. Tegelik saabumisaeg tuletati esimese STOPPED_AT staatusega rea põhjal. See rida tähistas bussi realselt peatusesse jõudmist.

Kuna buss ei pruugi igas peatuses peatuda või võib peatus olla nii lühike, et sündmust STOPPED_AT ei registreerita, tuli arvestada ka alternatiivsete sündmustega. Kui vastavas peatuses puudus STOPPED_AT sündmus, kasutati INCOMING_AT staatusega andmepunkti. Kui puudusid mõlemad, määrati tegelik saabumisaeg grupi viimasest IN_TRANSIT_TO staatusega reast, mis esindas bussi viimast teadaolevat liikumist enne peatust.

Joonisel 8 on kujutatud busside hilinemiste keskmised väärtused aastatel 2019–2024. Horisontaalteljel on aastad ja vertikaalteljel keskmine hilinemine minutites. Iga aasta kohta on esitatud nii tegelik hilinemine (sinine tulp) kui ka prognoositud hilinemine (oranž tulp).



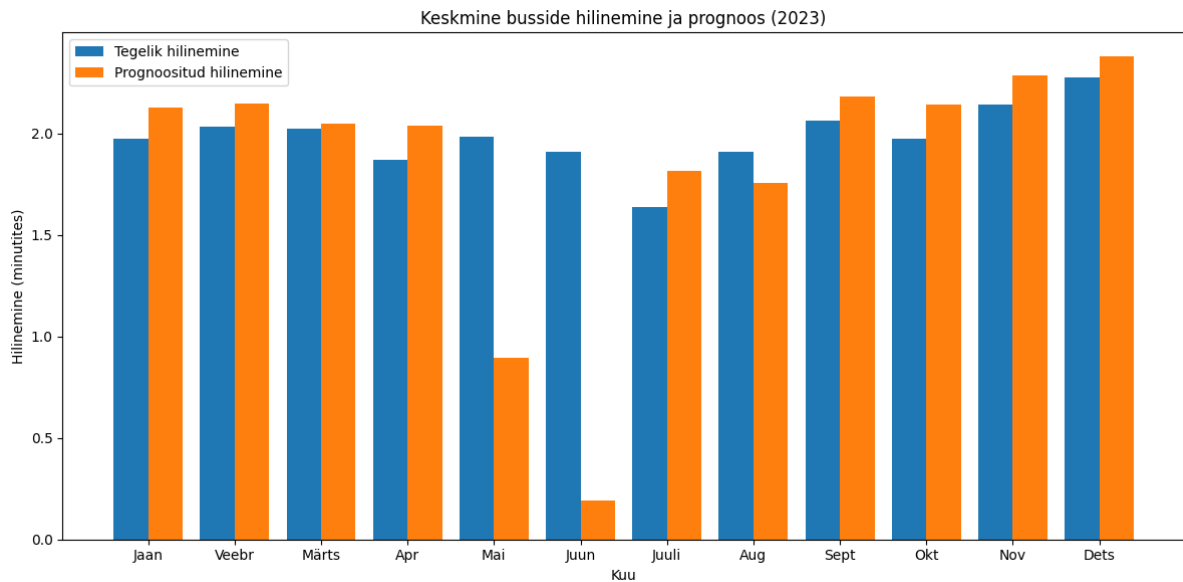
Joonis 8. Busside keskmine hiline mine ja keskmine ennustatud hiline mine vahemikus 2019-2024.

Jooniselt 8 on näha, et kuigi prognoositud ja tegeliku hiline mine väärtused on üldjoontes lähedased, esineb aastate lõikes varieeruvust. Mõnel aastal (nt 2022) jäävad väärtused teineteisele väga lähedale, samas kui teistel (nt 2023) on vahe oluliselt suurem. See viitab, et hiline mine ennustamise täpsus võib aastati kõikuda. Üldiselt võib järeldada, et busside prognoositud hiline mine ning reaalne peatusesse jõudmine on omavahel kooskõlas.

Joonisel 9 on kujutatud 2023. aasta busside keskmisi hiline mise kuude lõikes. Andmetest ilmneb, et alates septembrist hakkavad hiline mised kasvama ning detsembris saavutatakse aasta kõrgeim tase. Sellele võivad kaasa aidata pühade-eelsed liiklusummikud ja suurenenud autoliiklus kaubanduskeskuste suunal, mis mõjutavad paratamatult ka ühistranspordi liikumist. Vastupidise tendentsina on kõige väiksem keskmine hiline mine juulis, kesksuvisel perioodil, mil liikluskoormus on üldiselt madalam ning sõidugraafikud hõredamad.

Huvitava kõrvalekalde moodustavad prognoositud hiline mised mais ja juunis, mis on märgatavalt madalamad kui teistel kuudel. Sellele võib olla mitmeid seletusi. Üheks võimalikuks põhjuseks on hooajaline muutus liiklusoludes ja reisijate arvu vähenemine enne koolivaheaega. Samuti võib süsteem ennustuste koostamisel olla kasutanud varasemate kuude andmeid, mis ei kajasta piisavalt täpselt kevadperioodi tegelikke olusid. Võimalik on ka, et

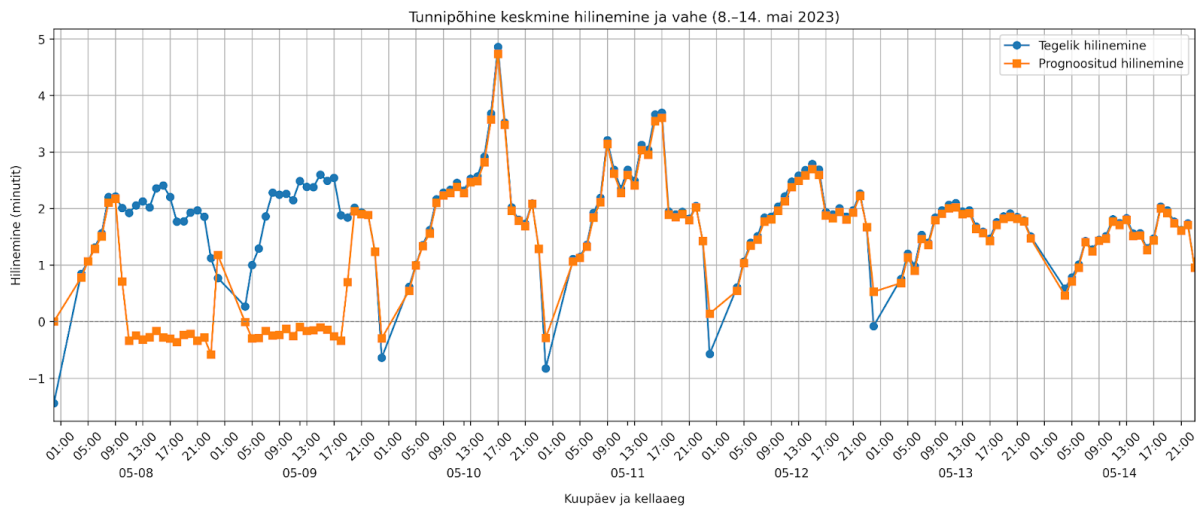
prognoosimudel hindas hilinemisi antud kuudel liiga optimistlikult või ei suutnud arvestada mõne lühiajalise muutujaga. Need tegurid kokku võivad selgitada, miks prognoositud hilinemised mais ja juunis tavapärasest madalamaks kujunesid.



Joonis 9. Keskmine busside hiline mine ning prognoositud hiline mine kuude lõikes aastal 2023.

Et hinnata hiline miste ennustuste täpsust ajas, koostati päevapõhine võrdlus 2023. aasta ühe nädala andmete põhjal. Joonisel 10 on kujutatud igapäevased keskmised prognoositud ja tegelikud hiline mised minutites. Horisontaalteljel on kuupäevad, vertikaalteljel hiline mise suurus. Graafikult on näha, et prognoositud ja tegelik hiline mine liiguvad üldjoontes sarnaselt, kuid esineb ka perioode, kus nende vahele tekib märgatav erinevus.

Nädalavalik langes maisse, mida joonise 9 põhjal iseloomustavad suuremad erinevused prognoositud ja tegelike saabumisaegade vahel. Konkreetne erinevus ilmneb näiteks esmaspäeval, 8. mail kell 10:00 kuni teisipäeva, 9. mai kella 18:00 vahemikus. Sellel perioodil prognoosisid bussid, et nad jõuavad peatustesse keskmiselt pool minutit enne graafikut, kuid tegelikkuses saabusid nad sageli hoopis hiline misega, ulatudes keskmiselt kahe minutini. Sellised erinevused viitavad olukordadele, kus ennustumudel ei suuda liiklusoludest või muudest välistest teguritest tingitud hiline misi õigesti ette näha.



Joonis 10. Päevapõhine prognoositud ja tegelik hilinemine 2023 mais.

Analüüsi tulemused näitavad, et busside prognoositud ja tegelik hilinemine on üldjoontes omavahel kooskõlas, kuid esineb perioode ja olukordi, kus nende vaheline erinevus on märgatav. Aastate lõikes võib täheldada, et ennustuste täpsus varieerub: mõnel aastal on prognoosid väga lähedased tegelikkusele, samas kui teistel aastatel, nagu näiteks 2023, on erinevused suuremad. Kuude lõikes ilmnes, et mais ja juunis on prognoositud hilinemised oluliselt madalamad kui tegelikud, mis viitab võimalikule mudeli alahindamisele kevadperioodi liiklusolude suhtes. Nädalapõhine analüüs kinnitab seda tähelepanekut, tuues välja konkreetseid ajavahemikke, kus süsteem ennustas varasemat saabumist, kuid bussid hilinesid realselt keskmiselt kuni kaks minutit.

Need tulemused viitavad sellele, et kuigi olemasolev süsteem suudab üldise suundumuse enamasti adekvaatselt edasi anda, esineb täpsuses ajutisi ja hooajalisi kõikumisi. Seetõttu oleks soovitatav üle vaadata, milliseid andmeid kasutatakse prognoosimudeli sisendina ning kaaluda mudeli täiendamist, et see suudaks paremini arvestada lühiajalisi muutusi liiklusoludes, hooajalisust ja päevaseid tipptundide eripärasid. Samuti näitab analüüs, et ennustussüsteemi tõhususe hindamisel ei piisa ainult aasta keskmistest, vaid oluline on jälgida ka nädalapõhiseid ja kuupõhiseid erinevusi.

5.4.3 Millised liinid on kõige ebausaldusväärsemad ajast kinnipidamise osas?

Joonisel 11 on kujutatud 2023. aasta iga kuu kolm kõige suurema keskmise hilinemisega bussiliini. Keskmine hilinemine on mõõdetud sekundites ning visualiseering annab ülevaate kuupõhistest erinevustest liinide täpsuses. Analüüsist ilmneb, et mitmed liinid esinevad mitmetel kuudel, viidates potentsiaalsetele hilinemise mustritele.

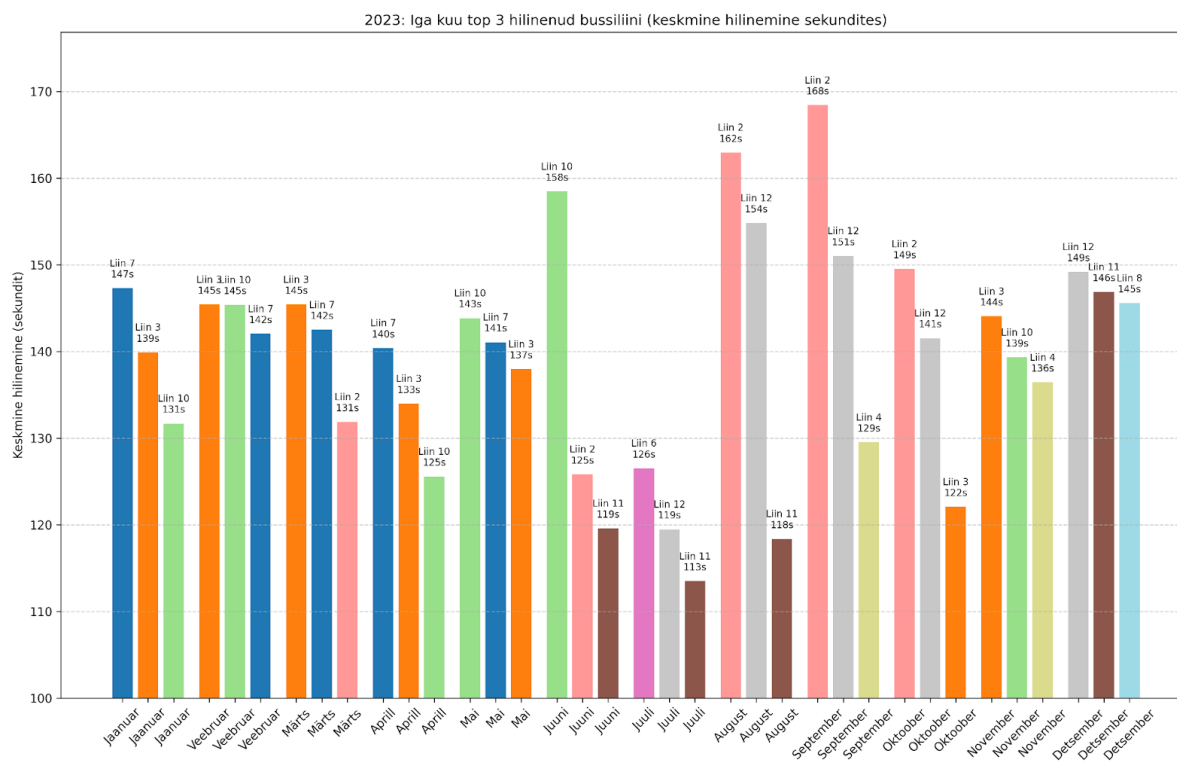
Kõige silmapaistvam on liin 2, mis esineb esikolmikus viiel erineval kuul, sealhulgas kõrgeima keskmise hilinemisega kuudel augustis (162 sekundit) ja septembris (168 sekundit). See viitab sellele, et liin 2 hilinemised ei ole juhuslikud, vaid pigem püsivad ning seotud tõenäoliselt kas marsruudi, liiklusolude või graafiku planeerimisega.

Aasta esimeses pooles esinevad kõige rohkem liin 7 ja liin 3, mis esinevad järjepidevalt jaanuarist maini. Samuti paistab silma liin 10, mis esineb hajutatult mitmetel kuudel (sh veebruaris, aprillis, mais ja novembris), viidates võimalikule ebastabiilsusele aastaringelt.

Liin 11 ei esine küll kuude lõikes pidevalt, kuid selle ilmnemisel (juulis, augustis ja detsembris) on keskmised hilinemised kõrged, ulatudes kuni 146 sekundini. Selline esinemismuster võib viidata väiksema koormusega liinile, kus üksikud hilinemised mõjutavad keskmist tugevamalt. Samas sarnaneb liin 12 muustrilt liin 2-ga, esinedes sagedasti aasta teises pooles ja detsembris, mil keskmised hilinemised on samuti kõrged.

Tulemused näitavad, et hilinemised on kuude lõikes koondunud eriti suve lõppu ja sügisesse. Augustis ja septembris registreeritakse mitme liini puhul aasta suurimad hilinemised, mis langevad ajaliselt kokku kooliaasta alguse ja liikluskoormuse kasvuga. Lisaks näitab detsember taas hilinemiste suurenemist, mis võib olla seotud pühade-eelse liiklusintensiivsuse kasvuga.

Kokkuvõttes viitab analüüs, et liin 2 on 2023. aasta andmetel kõige ebausaldusväärsem liin keskmise hilinemise põhjal. Lisaks vajaksid eritähelepanu ka liinid 7, 3 ja 10, mille hilinemismustrid korduvad mitmel järjestikusel kuul. Selliste liinide puhul oleks soovitatav hinnata graafikute vastavust tegelikule liiklusolukorrale ning kaaluda täiendavaid meetmeid punktuaalsuse parandamiseks.



Joonis 11. Kõige suuremate keskmiste hilinemistega liinide igakuine võrdlus aastal 2023.

Joonisel 12 on esitatud 2023. aasta iga kuu kohta kolm bussiliini, millel oli vastaval kuul kõige väiksem keskmine hilinemine. Visualiseeritud andmed võimaldavad tuvastada liine, mis püsivad ajagraafikutes kõige täpsemini, ning annavad ülevaate nende jaotumisest aasta lõikes.

Kõige parema tulemuse annab liin 21, mis on esikolmikus kõigil kaheteistkümnel kuul, sealjuures mitmel korral ka kui kõige väiksema hilinemisega liin. Näiteks aprillis oli liin 21 keskmine hilinemine 25 sekundi ning mais 26 sekundi suurusjärgus, mis on üks madalamaid väärtusi kogu vaatlusaasta jooksul. Selline ühtlus viitab hästi planeeritud sõidugraafikule ja võimalikult väikesele kokkupuutumisele liikluskorralduse või muude hilinemist põhjustavate teguritega.

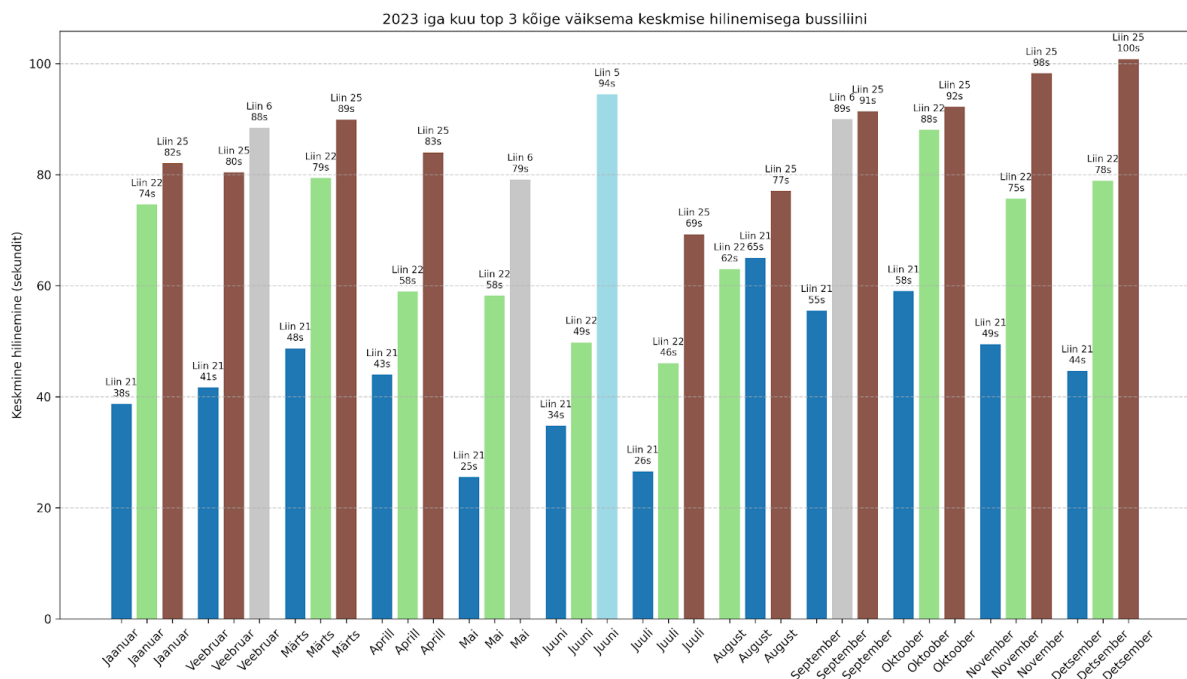
Liin 22 on samuti märkimisväärselt stabiilne, esinedes esikolmikus üheksal erineval kuul. Tema keskmine hilinemine jääb enamikul juhtudest alla 60 sekundi ning püsib ühtlaselt madalal ka sügiskuudel, mil üldine liikluskorraldus on kõrgem.

Lisaks paistab silma liin 25, mis on esikolmikus üheksal kuul ning seejuures domineerib eriti aasta teises pooles. Näiteks novembris ja detsembris on liin 25 keskmine hilinemine üle 90 sekundi, kuid sellegipoolest jääb see teistest liinidest madalamaks. Siinkohal tuleb arvestada,

et need hilinemised on küll suuremad, kuid arvestades aasta üldist hilinemise kasvu, on liin 25 siiski suhteliselt täpne.

Mõned kuud toovad esile ka üksikuid liine, mis mujal ei esine, näiteks liin 6 mais ja juunis ning liin 5 juunis. Need võivad viidata konkreetse perioodi ajutisele täpsusele, kuid üldise järjepidevuse puudumine viitab pigem juhuslikule sobitusele ajagraafikusse.

Tulemuste põhjal võib järeldada, et liin 21 on 2023. aastal kõige täpsema ajakohasusega liin, arvestades tema järjepidevat esinemist väikseima hilinemisega liinide seas. Ka liinid 22 ja 25 paistavad silma kui üldiselt täpsed ja stabiilsed, eriti aastate lõikes, kus üldine hilinemistase on suurenenud. Need liinid võiksid toimida võrdlusbaasina teiste liinide analüüsimisel ning nende täpsust tagavaid tegureid oleks soovitatav uurida süvitsi ka planeerimise ja graafikuoptimeerimise kontekstis.



Joonis 12. Kõige madalamate keskmiste hilinemistega liinide igakuine võrdlus aastal 2023.

Joonisel 13 on esitatud 2023. aasta 20 kõige suuremate keskmiste hilinemistega bussiliini kuude lõikes. Andmed on sorteeritud kogu aasta keskmise hilinemise alusel kahanevas järjekorras, võimaldades tuvastada järjepidevalt hilinevaid liine ning nende muutuvat käitumist aastasisese dünaamika kontekstis.

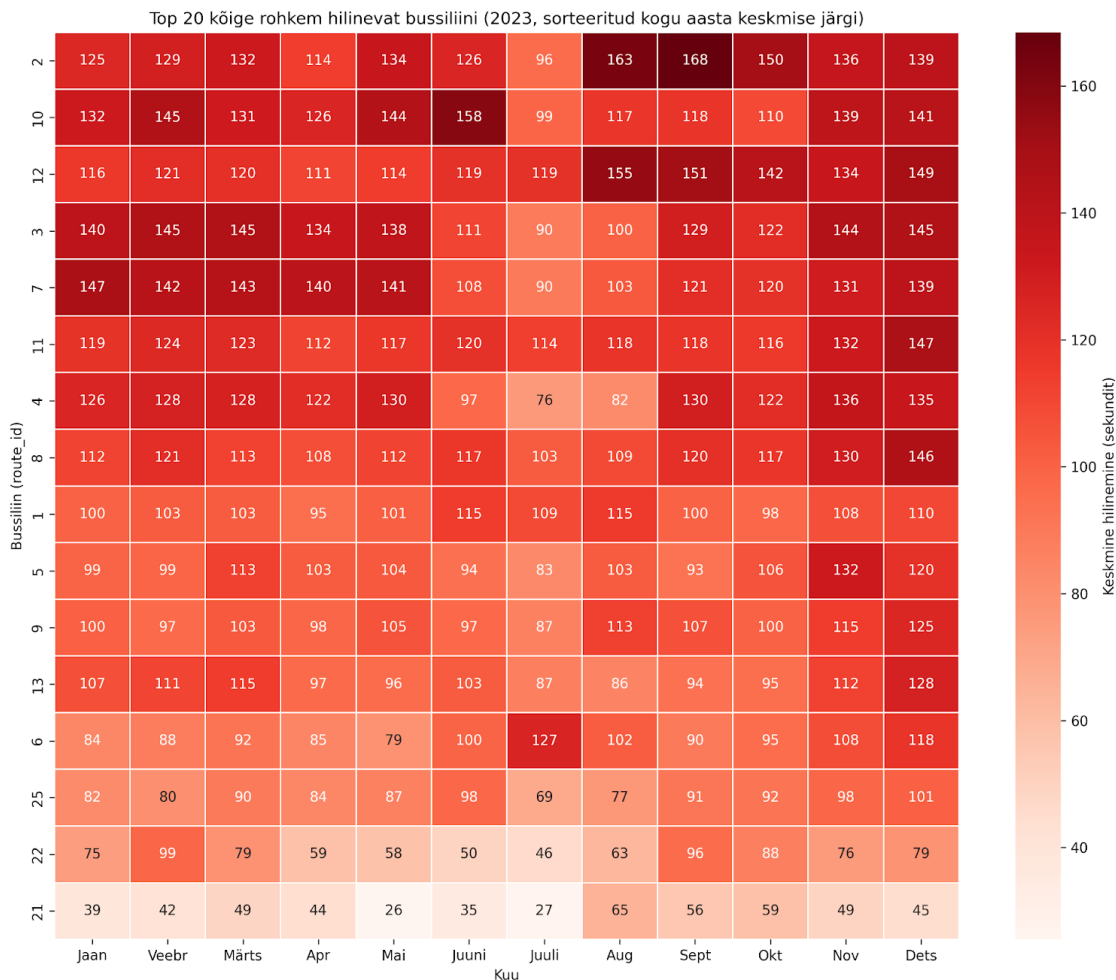
Kõige enam paistab välja liin 2, mis on kogu aasta lõikes kõige suurema keskmise hilinemisega liin. Eriti kõrgeks tõusevad selle liini hilinemised augustis (163 sekundit) ja septembris (168 sekundit), mis langeb kokku suve lõpu ja kooliaasta algusega, mil üldine liikluskoormus linnas kasvab. Samuti on liin 2 ainus, mille hilinemised ületavad 160 sekundi piiri mitmel järjestikusel kuul, viidates süstemaatilisele ülekoormusele või ebapiisavalt planeeritud graafikule.

Ka liinid 10, 12 ja 3 kuuluvad püsivalt esiviisikusse, näidates suurt keskmist hilinemist enamikes kuudes. Liin 10 paistab silma maikuus (158 s), liin 12 aga augustis ja septembris (vastavalt 155 ja 151 sekundit), mis viitab võimalikule marsruudi haavatavusele hooajaliste liiklusmuutuste suhtes.

Märkimisväärne on ka see, et mitme liini puhul on hilinemised juulis ja augustis madalamad, kuid seejärel tõusevad oluliselt alates septembrist, mis langeb kokku sügisese graafikumuudatusega. Näiteks liin 4 näitab madalat hilinemist juulis (76 s), kuid septembris tõuseb see 130 sekundini.

Samas võimaldab joonis jälgida ka liinide, nagu liin 21 ja liin 22, järjepidevalt madalat hilinemisnäitajat. Liin 21, mille hilinemised püsivad vahemikus 26 kuni 65 sekundit, paistab silma kogu vaatlusaasta vältel oma täpsusega. See kooskõlastub ka varasema joonise (Joonis 11) tulemustega, mis näitasid liini 21 esinemist väikseima hilinemisega liinide seas igal kuul.

Kokkuvõttes viitab analüüs sellele, et mõned liinid (nt liin 2, 10, 12 ja 3) esinevad järjepidevalt suurima hilinemisega liinide seas, samas kui liin 21 ja 22 püsivad täpsuse osas kindlal tasemel. Selline muster näitab, et probleemsete liinide hilinemised ei ole juhuslikud, vaid süstemaatilised ning vajaksid põhjalikumat käsitlust näiteks graafiku optimeerimise, marsruudi paranduste või liiklusvoogude ümberkorraldamise kaudu.



Joonis 13. 2023. aasta kõige suurema keskmise hilinemisega bussiliinide jaotus kuude lõikes.

Andmete põhjal läbi viidud liinide võrdlev analüüs näitab selgelt, et busside ajakohasuses esineb märkimisväärsed erinevusi nii liinide kui ka kuude lõikes. Tulemused viitavad sellele, et osa liine kaldub järjepidevalt sõiduplaanidest kõrvale, samas kui teised suudavad püsida ajas täpsed.

Kõige ebausaldusväärsemaks liiniks osutus liin 2, mille keskmine hilinemine oli aasta lõikes kõrgeim ning mis esines pidevalt ka kuupõhistes tippude hulgas. Samuti paistsid kõrgete hilinemiste poolest silma liinid 3, 10 ja 12, mille tulemused olid stabiilselt kõrged nii kogu aasta kui ka mitmete üksikute kuude lõikes. Eriti ilmnes hilinemiste kasv suve lõpus ja sügisel, mis langeb kokku graafikumuudatuste ja kooliaasta algusega.

Vastukaaluks näitasid liinid 21 ja 22 erakordset ajakohasust, püsides järjepidevalt kõige väiksema hilinemisega liinide hulgas. Eriti kerkis esile liin 21, mis esines kõigil kuudel

väikseima hilinemisega liinide seas. Ka liin 25 hoidis aastaringselt stabiilset, suhteliselt täpset taset, eriti aasta teises pooles.

Kokkuvõttes võimaldab analüüs välja tuua liinid, mis vajavad suuremat tähelepanu graafikute täpsustamisel ja planeerimisel, ning samas ka liinid, mille toimiv lahendus võiks olla eeskujuks ülejäänud liinivõrgu optimeerimisel. Süsteemsete hilinemismustrite tuvastamine loob aluse sihipärasemate parandusmeetmete rakendamiseks nii liinipõhiselt kui ka linnaülestel liikluskorraldusotsuste tasandil.

5.4.4 Millistes peatustes on hilinemine kõige suurem?

Joonisel 14 on esitatud igal aastal kolm peatust, kus busside keskmine hilinemine mõõdetuna minutites on olnud suurim. Tulemused põhinevad aastatel 2019–2024 kogutud andmetel ning iga tulba kohal on esitatud ka vaatluste koguarv (n), mis võimaldab hinnata andmemahtu, mille põhjal keskmine on arvutatud.

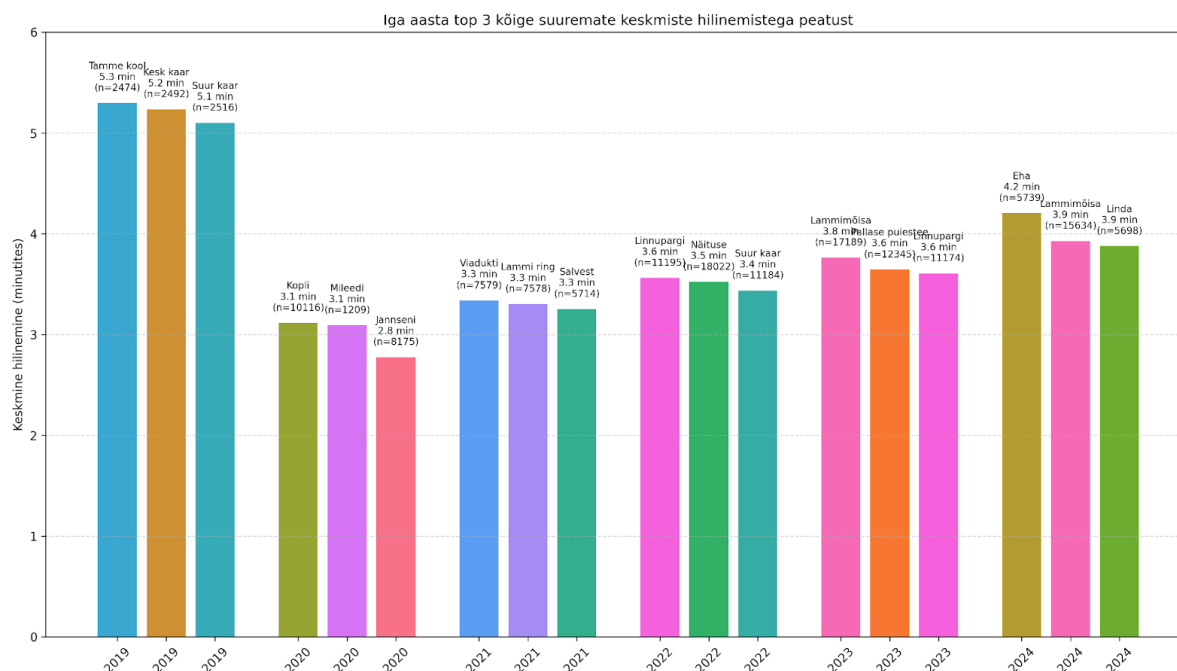
Kõige kõrgema keskmise hilinemisega peatused ilmnesid 2019. aastal, mil Tamme kool, Keskkaar ja Suurkaar keskmiselt 5-minutilised hilinemised. Need väärtused on märgatavalt kõrgemad kui järgnevatel aastatel ja viitavad tõenäoliselt andmete puudulikkusele aastal 2019.

Alates 2020. aastast stabiliseeruvad keskmised hilinemised vahemikus 2,8 kuni 4,2 minutit. Lammimõisa peatus on joonisel esindatud kolmel järjestikusel aastal (2022–2024), olles nii 2023 kui 2024 aastal esikolmikus keskmise hilinemisega vastavalt 3,8 ja 3,9 minutit. Peatuse korduv esinemine viitab süsteemsele ajast kõrvalekaldumisele konkreetses asukohas.

Ka Linnupargi, Näituse, Eha ja Linda peatused paistavad silma püsiva esinemisega kõrge hilinemisega peatuste seas. Nende asukoht ja võimalik kokkupuude tiheda liiklusvooga võib olla seotud korduvate hilinemiste põhjuseks.

Samas näitab 2020. aasta madalamaid hilinemisi, kus peatused nagu Jannseni registreerivad hilinemisi alla 3 minuti (2,8 min), mis võib viidata nii muutustele liiklusoludes kui ka ajutisele graafikute sobitumisele.

Kokkuvõttes võib järeldada, et kõige suuremad keskmised hilinemised on aastate lõikes koondunud konkreetsetesse peatustesse, eelkõige Lammimõisa, Linnupargi ja Suurkaar peatusesse. Nende korduv esinemine viitab vajadusele hinnata nende peatuste ümbruse liiklusvoogusid, peatusse jõudmise marsruute ning võimalikke graafikust tulenevaid kitsaskohti. Selline info on oluline nii graafikute optimeerimisel kui ka reisijate ootuste juhtimisel.



Joonis 14. 3 kõige suurema keskmiste hilinemistega peatust aastast (2019–2024).

Joonisel 15 on esitatud 2023. aasta igal kuul üks peatus, millel oli vastaval kuul Tartu bussiliikluses suurim keskmine hiline. Keskmised väärtused on väljendatud minutites ning lisatud on ka vastava peatuse vaatluste arv (n), mis võimaldab hinnata andmete statistilist representatiivsust.

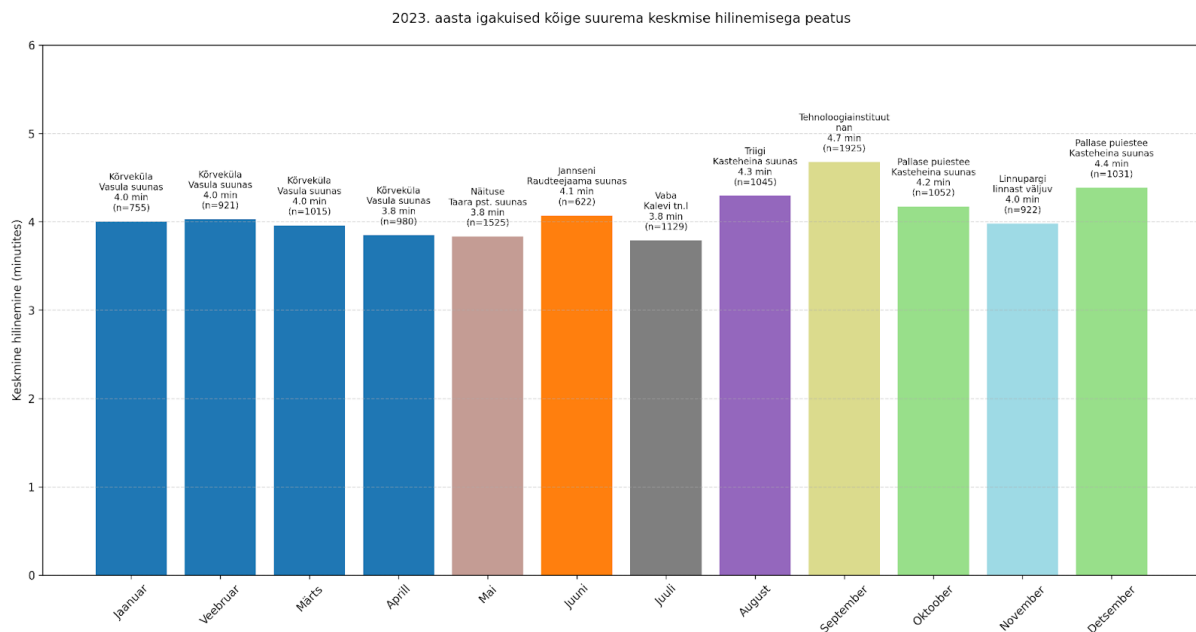
Aasta esimeses pooles domineerib tugevalt Kõrveküla peatus, mis oli kõige suurema hilinemisega peatus jaanuaris, veebruaris, märtsis ja aprillis. Kõigil nendel kuudel ulatus keskmine hiline 3,8 kuni 4,0 minutini. Kõrveküla korduv esinemine viitab tõenäolisele süsteemsele probleemile marsruudil või piirkonna liiklusoludes.

Mai ja juuni toovad esile peatused Näituse ja Jannseni, mille keskmised hiline on vastavalt 3,8 ja 4,1 minutit. Jannseni peatus on kajastatud ka varem analüüsitud aastapõhistes andmetes, mis viitab selle jätkuvale ajast kõrvale kaldumisele.

Juulis on tipus Vaba peatus 3,8 minutiga, kuid alates augustist hakkavad esinema uued peatused. Triigi peatus registreerib augustis keskmise hiline 4,3 minutit, samas kui Tehnoloogiainstituut on 4,7 minutiga septembris kõige kõrgema hiline väärtusega kogu aasta lõikes. See peatus väärib eritähelepanu, kuna tema kõrge hiline langeb ajaliselt kokku sügise alguse ja graafikumuudatustega.

Sügiskuudel (oktoober–detsember) kerkib esile Pallase puiestee peatus, olles tippilineja kahel järjestikusel kuul, keskmise hilinemisega 4,2 ja 4,0 minutit. Lisaks esineb novembris taas Linnupargi peatus, mis on olnud ka varasematel aastatel kõrge hilinemisega peatus.

Tulemused näitavad, et mõned peatused nagu Kõrveküla ja Pallase puiestee esinevad korduvate kõrgete hilinemistega. See viitab, et ajast kõrvalekaldumine ei ole juhuslik ega üksikjuhtum, vaid seotud kas konkreetse marsruudi, piirkondliku liiklusolukorra või graafikute sobimatusega tegelikele oludele.



Joonis 15. 2023. aasta igakuised kõige suurema keskmise hilinemisega peatused.

Joonisel 16 on esitatud 2023. aasta top 20 peatust, mille keskmine hilineimine oli kõrgeim, koos kuupõhiste väärtustega minutites. Tulemused võimaldavad analüüsida mitte ainult aastast kogupilti, vaid ka kuude kaupa varieeruvust konkreetsetes peatuses, tuues esile hooajalised kõikumised ja võimalikud süsteemsed probleemid.

Kõige järjepidevamalt on kõrgete hilinemistega peatus Pallase puiestee, mille hilinemised kasvavad märgatavalt alates augustist, saavutades tippväärtuse novembris (4,5 min) ja püüdes kõrgel ka detsembris (4,4 min). Sarnane muster esineb ka Linnupargil, kus hilineimine ulatub detsembris 4,3 minutini.

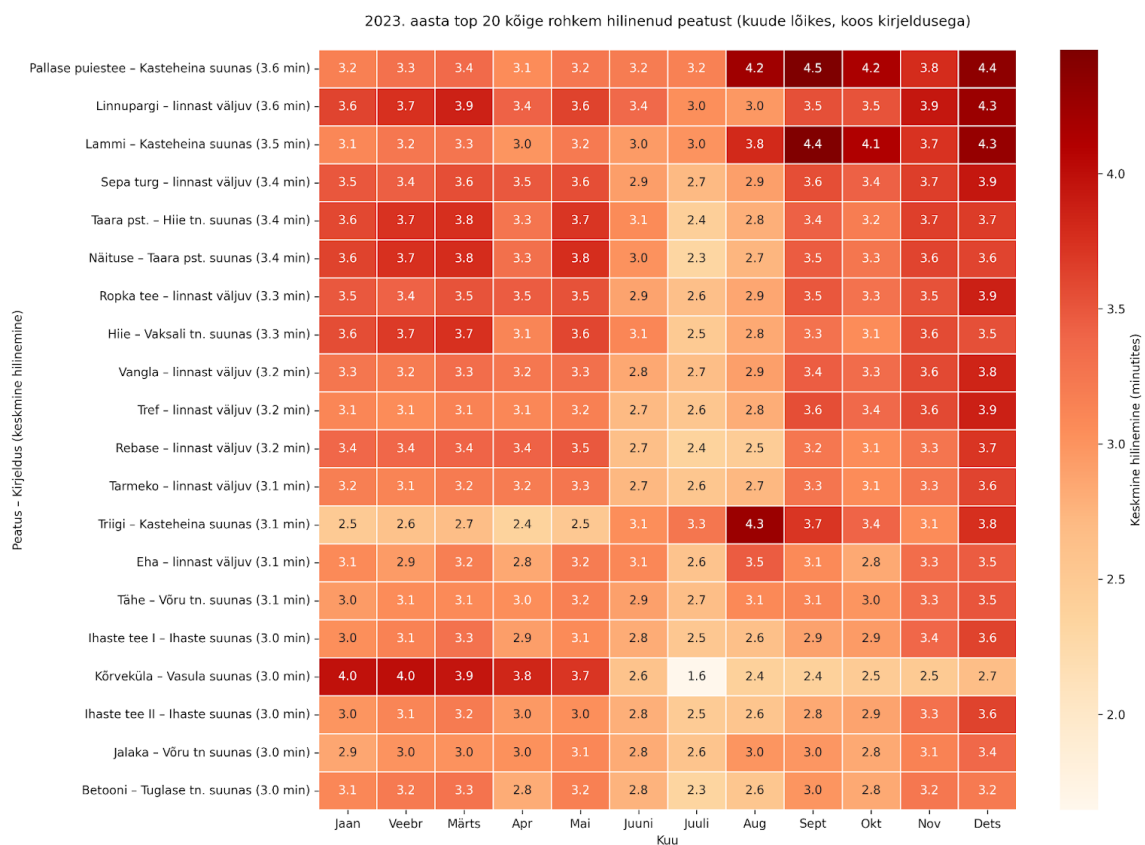
Sügisperioodi iseloomustabki üldine hilinemiste tõus mitmes peatuses. Näiteks Lammi, Sepa turg, Salvest ja Näituse näitavad samuti selget tõusu hilissügisel. Paljud peatused saavutavad

oma aastase maksimumi just oktoobris, novembris või detsembris, mis langeb kokku nii kooliaasta täismahus tööperioodiga kui ka pühade-eelse liikluskoormuse kasvuga.

Tähelepanu väärib ka Triigi peatus, mis on üldiselt väiksema hilinemisega, kuid registreerib augustis järsu tõusu 4,3 minutini, mis võib viidata ajutisele töökorralduse või liikluskorralduse probleemile.

Teisest küljest ilmneb Kõrveküla peatuses selgelt langustrend: jaanuaris ja veebruaris on keskmine hilinemine 4 minutit, kuid hiljem langeb see märgatavalt, jõudes juulis 1,6 minutini ja septembris vaid 2,4 minutini. Selline mustri muutus võib viidata graafiku kohandamisele või liiklusolude paranemisele konkreetsetel perioodidel.

Peatuste Eha, Tähe, Jalaka ja Betooni puhul on hilinemised madalamad ja stabiilsemad, jäädes valdavalt 2,5 kuni 3,5 minuti vahele. Samas nende järjepidev esinemine top 20 hulgas viitab sellele, et nende asukohad on üldise süsteemi toimimise seisukohalt kitsaskohad.



Joonis 16. 2023. aasta 20 kõige suurema keskmise hilinemisega peatust kuude lõikes.

Analüüs jooniste 14, 15 ja 16 alusel näitab, et busside suurimad hilinemised ei ole juhuslikud, vaid koonduvad teatud kindlatesse peatustesse ning korduvad eri aastatel ja kuudel. Aastatel

2019–2024 ilmnes kõige suurem keskmine hilinemine 2019. aastal, kuid tõenäoliselt oli see seotud sellega, et sellel aastal oli vähem andmeid kogutud. Alates 2020. aastast on hilinemised stabiliseerunud vahemikus 2,8 kuni 4,2 minutit ning mitmed peatused on korduvalt olnud suurima keskmise hilinemisega peatuste hulgas.

Kõige järjepidevamad probleemsed peatused on Lammimõisa, Linnupargi, Pallase puiestee, Kõrveküla, Näituse ja Jannseni. Eriti kerkib esile Lammimõisa peatus, mis on esikolmikus kolmel järjestikusel aastal, ning Pallase puiestee, mille hilinemised tõusevad märgatavalt sügisel ja saavutavad tippaseme novembris ja detsembris (vastavalt 4,5 ja 4,4 minutit). Samuti paistab silma Tehnoloogiainstituut, mille septembrikuine keskmine hilinemine (4,7 minutit) on aasta kõrgeim.

Kõrveküla peatus on probleemne just aasta esimeses pooles, kuid näitab seejärel langustrendi, mis võib viidata edukale graafikumuudatusele või piirkondlikele liiklusolude paranemisele. Sarnased hooajalised mustrid ilmnevad mitmes peatuses, kus hilinemised kasvavad järsult just sügiskuudel (september–detsember), mil ühistranspordi koormus ja üldine liiklusintensiivsus tõusevad.

Seega võib vastuseks uurimisküsimusele järeldada, et Tartu linnas esinevad suurimad hilinemised eeskätt peatustes Lammimõisa, Linnupargi, Pallase puiestee, Tehnoloogiainstituut, Kõrveküla ja Näituse. Nende korduv esinemine kõrgeima hilinemisega peatuste seas viitab probleemile, mille põhjuseks võivad olla kas marsruudipõhised liiklusummikud, graafikute sobimatus tegelike oludega või ajast sõltuvad koormustipud. Selliste peatuste ümbruse liiklusvoogude, liinikorralduse ja ajastuse analüüs on oluline eeldus täpsemate ja realistlikumate graafikute kujundamiseks.

5.4.5 Kui kaugel peatuvad bussid peatustest?

Joonisel 17 on kujutatud busside peatumise kauguse jaotus ametlikust peatuspunktist vahemikus 0 kuni 28 meetrit. Andmed pärinevad 2023. aasta maikuu vaatluste põhjal ning visualiseeritud on vaid need juhtumid, mille kaugus jääb alla või võrdub 28 meetriga. Kauguse arvutamisel bussi peatumiskoha ja ametliku peatuse vahel kasutati Haversine'i meetodit, mis arvestab Maa kera kuju ja sobib täpselt väikeste vahemaade arvutamiseks geograafiliste koordinaatide alusel.

Vahemik 0-28 meetriton määratud statistilise puhastusmeetodina kasutatud interkvartiilivahemiku (IQR) põhjal. IQR-meetod on andmepõhine viis äärmuslike väärtuste

tuvastamiseks ja kõrvaldamiseks, tuginedes jaotuse kesksetele kvartiilidele. Meetodi kohaselt loetakse normatiivseks vahemikuks andmed, mis jäävad vahemikku: $Q1 - 1.5 \cdot IQR$ kuni $Q3 + 1.5 \cdot IQR$.

Kuna miinumkaugus ei saa olla negatiivne, keskendutakse ainult ülemise piiri määratlemisele. Mai 2023 andmetel olid väärtused järgmised:

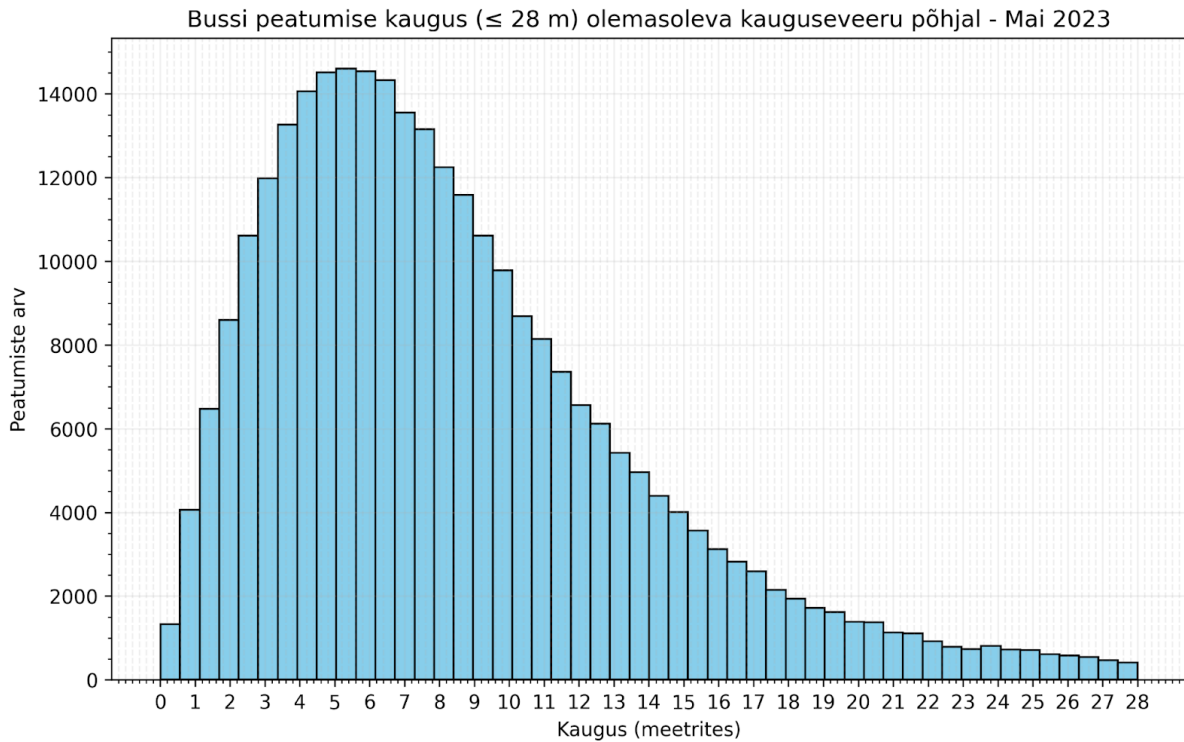
- Esimene kvartiil (Q1): 4,50 meetrit;
- Kolmas kvartiil (Q3): 12,2 meetrit;
- Interkvartiilivahemik (IQR): 7,7 meetrit;
- Ülemine lävend ($Q3 + 1.5 \times IQR$): 27,6 meetrit.

Seega on kõik kaugused, mis ületavad 27,6 meetrit, käsitletavad äärmuslike kõrvalekalletena ning neid ei arvestata üldjaotuse analüüsis.

Statistiline ülevaade kogu andmestikust:

- Minimaalne kaugus: 0,01 meetrit;
- Maksimalne kaugus: 11 601,17 meetrit;
- Keskmise kaugus: 11,21 meetrit;
- Kehtivaid andmepunkte kokku: 3 642 605;
- Andmepunkte $\leq 27,60$ m: 3 370 377;
- Osakaal alla lävendi: 92,53%.

Histogramm (vt joonis 17) näitab, et suurim hulk peatuseid toimub 5–8 meetri kaugusel, kusjuures tipp on 6 meetri juures, kus peatuste arv ületab 14 000. Jaotus on positiivselt kaldus, viidates sellele, et pikemad kaugused on vähem levinud, kuid siiski esinevad. Valdav osa bussipeatusi jääb alla 15 meetri kaugusesse, mis kinnitab, et GPS-andmete ruumiline täpsus on piisav praktiliseks analüüsiks.



Joonis 17. Peatuses bussi peatumise kaugus peatusest (meetrites).

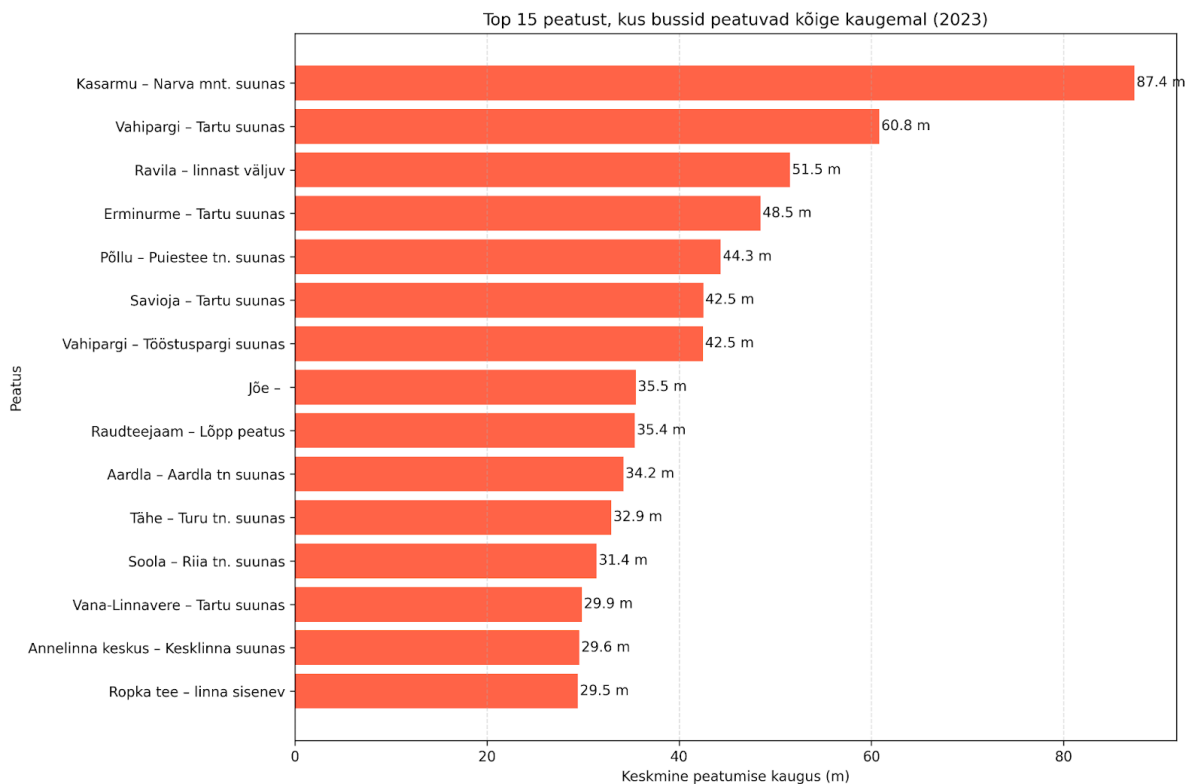
Joonisel 18 on kujutatud 2023. aasta 15 peatust, kus busside keskmine peatumiskaugus ametlikust peatuspunktist oli kõige suurem. Kaugused on esitatud meetrites ning iga tulp esindab vastava peatuse keskmist väärtust, arvatud kõigi aasta jooksul kogutud peatumisjuhtumite põhjal.

Kõige suurema keskmise kaugusega peatus on Kasarmu, kus bussid peatuvad keskmiselt 87,4 meetri kaugusel määratud punktist. See väärtus ületab oluliselt nii kogu andmestiku keskmise (11,21 m) kui ka soovitusliku usalduspiiri (27,60 m), mis tuletati interkvartiilimeetodil. Selline kõrvalekalle viitab tõsisele nihkele kas peatuse koordinaatides, GPS-signaali ebatäpsuses või sõidukite tegelikus peatumiskäitumises.

Teise ja kolmanda koha hõivavad Vahipargi (60,8 m) ja Ravila (51,5 m). Mõlemad ületavad märkimisväärselt usaldusväärse kaugusepiiri ning kuuluvad selgelt väljajäävate ehk ebatüüpiliste peatuste hulka. Sama tendents kordub ka järgnevatel peatustel nagu Erminurme, Põllu ja Savioja, kus peatumiskaugused jäävad vahemikku 42–49 meetrit.

Ülejäänud peatuste hulgas on mitmeid, mis jäävad 30–40 meetri vahemikku, näiteks Jõe, Raudteejaam, Tähe ja Soola. Kuigi need väärtused on madalamad kui esiviisikus, ületavad need

siiski märgatavalt nii keskmise kui ka tüüpilise peatumiskauguse ning viitavad süsteemsele nihkele.



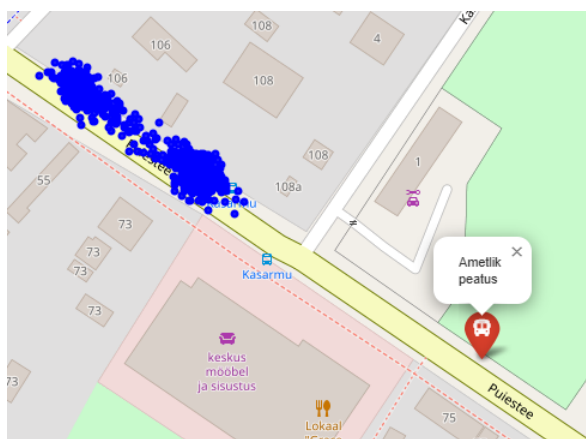
Joonis 18. 15 peatust, kus bussid peatuvad kõige kaugemal võrreldes ametliku peatuse asukohaga.

Joonisel 19 on kujutatud kõik busside peatumiskohad Kasarmu peatuses (Narva maantee suunas) 2023. aasta jooksul. Punkt pilvest ilmneb, et busside tegelik peatumine on oluliselt nihkes võrreldes ametliku peatuspunktiga, mis on esitatud kaardil punase märgina. Enamik sinistest punktidest on koondunud peatusest märkimisväärselt lääne suunas, ehk tee lõigule, mis jääb ametlikust kohast üle 80 meetri eemale. Kvantitatiivne analüüs näitas, et Kasarmu peatuses oli keskmine peatumiskaugus 87,4 meetrit, mis on kõrgeim kogu 2023. aasta andmestikus.

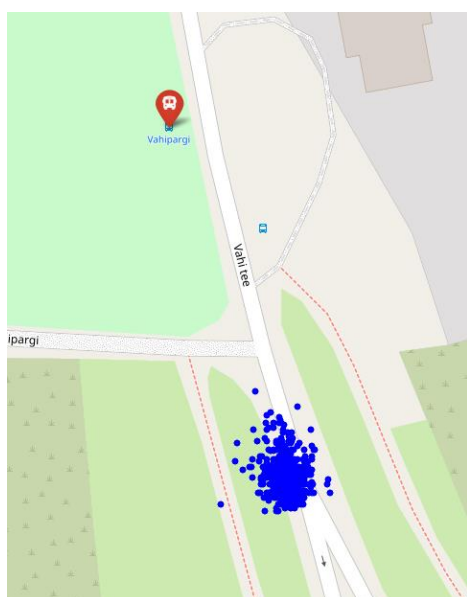
Oluline on märkida, et peatuskohtade koordinaadid pärinevad avalikust andmestikust peatus.ee kaudu, mis sisaldab Eesti ühistranspordi ametlikke GTFS-andmeid. Kasarmu peatuse puhul viitavad aga visuaalsed tulemused sellele, et OpenStreetMapi [16] aluskaart annab tegeliku peatuskoha asukoha osas täpsema ülevaate kui peatus.ee kaudu saadud longitude ja latitude väärtused. See viitab vajadusele kontrollida ametlike andmete täpsust, eriti kui neid kasutatakse ruumiliseks analüüsiks või visualiseerimiseks.

Joonisel 20 on esitatud Vahipargi peatuse (Tööstuspargi suunas) kõik peatumiskohad samal perioodil. Punktipilv näitab selget hajutust ning busside peatumised on valdavalt koondunud lõuna suunas teeäärsel lõigul, eemal ametlikust peatuspunktist. Vahipargi peatus registreeris samuti väga kõrge keskmise peatumiskauguse: 60,8 meetrit, mis asetab selle teisele kohale 2023. aasta kõrgeimate kauguste pingereas.

Mõlema peatuse puhul viitavad visualiseeringud sellele, et busside tegelik peatumiskäitumine ei lange kokku ametlike peatuspunktide asukohaga, mis võib olla põhjustatud kas GPS-i täpsusprobleemidest, teoludest, liikluskorraldusest või valesti määratud koordinaatidest GTFS-andmetes.



Joonis 19. Aasta 2023 kõik peatumised Kasarmu peatuses.



Joonis 20. Aasta 2023 kõik peatumised Vahipargi peatuses.

Bussid peatuvad valdavalt alla 15 meetri kaugusel ametlikust peatuspunktist, mis jääb GPS-mõõtemääramatuse piiresse ja sobib hästi praktiliseks analüüsiks. Samas leidub mitmeid peatusi, kus busside tegelik peatumine on püsivalt rohkem kui 40 meetrit eemal ametlikust punktist. Nende kõrvalekallete põhjuseks võivad olla põhjustatud:

- ebatäpselt määratud peatuskoordinaadid GTFS-andmetes,
- liikluskorralduslikud vajadused (nt tee kitsus, remont või ohutus),
- GPS-seadmete mõõteviga või hilinevad sündmused.

Sellised juhtumid tuleks eraldi käsitleda ning vajadusel korrigeerida ametlikke peatuskohti, et tagada usaldusväärne alus peatumiste analüüsiks ning ruumiliseks visualiseerimiseks.

5.4.6 Kui kvaliteetsed on busside poolt edastatud GPS andmed? Kas leidub varieeruvust erinevate busside seadmete kvaliteedis?

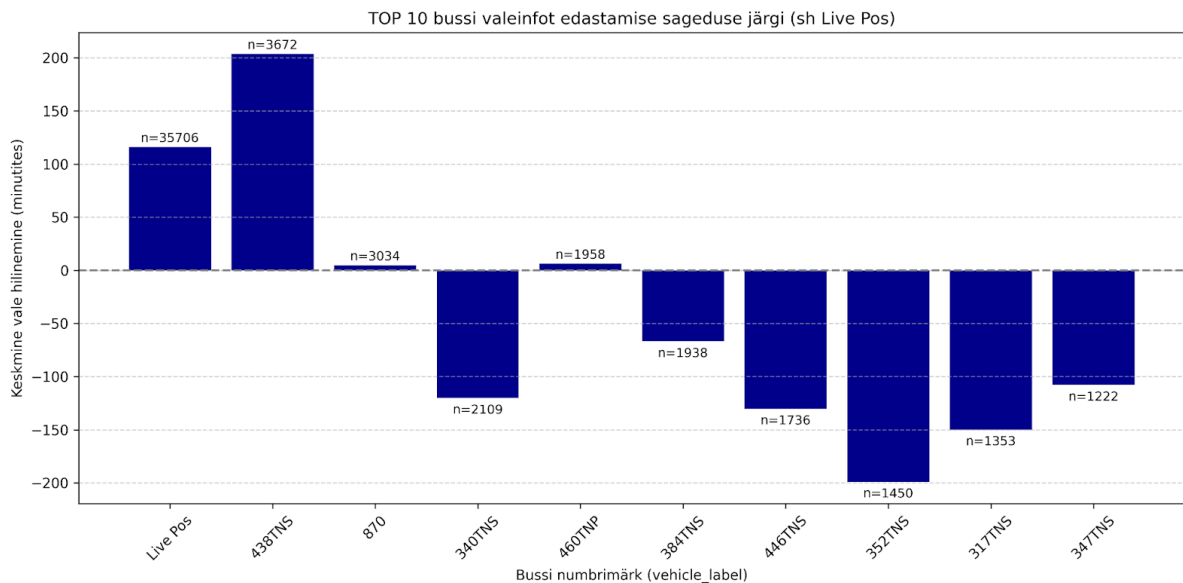
Joonisel 21 on kujutatud kümme bussi (tuvastatud vehicle_label alusel), mille puhul tuvastati kõige rohkem andmeridu, kus hilinemise absoluutväärtus ületas 15 minutit. Sellised juhtumid tähistavad antud analüüsi kontekstis valeinfot, kuna need viitavad anomaalsetele või mittetöötavatele reaalse hilinevamisandmetele. Tulbad visualiseerivad iga bussi keskmist vale hilinevamis minutites, s.o kui suurelt ennustus tegelikkusest keskmiselt erineb.

Kõige rohkem valeinfot sisaldavaid ridu pärines üldisest sündmusetüübist Live Pos, mille puhul ei olnud võimalik bussi numbrit tuvastada. Sellised sündmused genereeritakse Cumulocity süsteemi poolt siis, kui andmete struktuur on puudulik või seadmeidentifikaator puudub. See tähendab, et sõiduk on küll edastanud sündmuse, kuid süsteem pole suutnud seda siduda konkreetse bussiga. Live Pos ei ole seega mitte konkreetse bussi viga, vaid pigem platvormipoolne veateade, mis viitab seadme registreerimis- või andmestruktuuri probleemile.

Individuaalsete busside hulgas paistab silma 435 TNS, mille puhul esines 3672 valeinfoga rida, kus keskmine vale hilinevamine oli ligikaudu 195 minutit. See on märkimisväärselt kõrge ja viitab tugevale süsteemsele rikkele, kas seadme tarkvaralisele anomaaliale, katkestustele andmevoos või valeajastatud sündmuste logimisele.

Väga suur hulk valeinfot oli ka bussil 870, mille keskmine viga jäi nullilähedaseks. See viitab pigem juhuslikult esinevatele kõrvalekalletele, mis küll sageli esinesid, kuid mille kvantitatiivne mõju oli piiratud. Bussi numbrimärk 870 võib tähendada, et antud bussi seade pole korrektselt registreeritud ning numbrimärk on jäänud poolikuks.

Samal ajal registreerisid mitmed bussid süsteemseid alahindamisi, nende keskmine vale hiline mine oli negatiivne ja ulatus kuni -150 minutini. See tähendab, et süsteem ennustas nende busside saabumist kordades varem kui need tegelikult peatusesse jõudsid, mis võis kasutajatele tekitada väga eksitava mulje. Sageli viitavad sellised mustrid kas katkendlikule GPS-i signaalile, puudulikule ajasünkroniseerimisele või andmeedastuse hilinemisele Cumulocity keskkonnas.

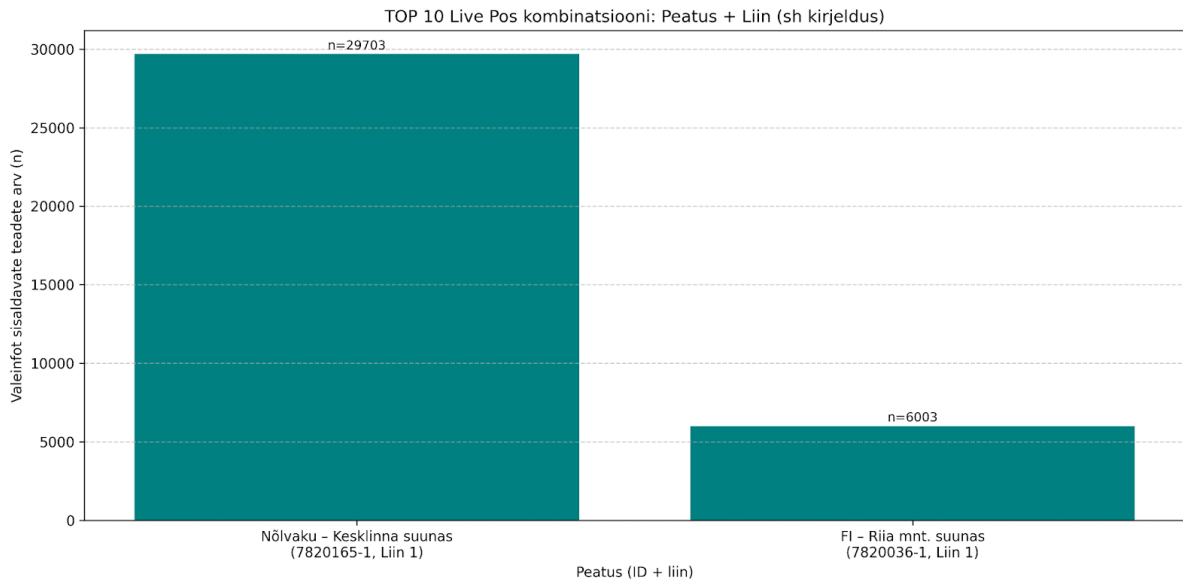


Joonis 21. 10 busi järjestatuna valeinfo edastamise sageduse põhjal.

Joonisel 22 on kujutatud peatuse ja liini kombinatsioonid, kus esines kõige rohkem Live Pos sündmusi. Andmed näitavad, et probleem on seotud ainult liiniga 1, kusjuures Nõlvaku (Kesklinna suunas) lõpp-peatuses registreeriti pea 30 000 sellist juhtumit.

Kuna Nõlvaku on liin 1 lõpp-peatus, võib probleem olla seotud sellega, et buss alustab või lõpetab seal sõitu ning sündmus saadetakse enne, kui busi ID on süsteemis täielikult registreeritud.

Kokkuvõttes viitab graafik, et Live Pos sündmused ei ole juhuslikud, vaid koonduvad konkreetselt liinile 1, mistõttu tuleks selle liini seadistusi ja andmeedastust eraldi hinnata.

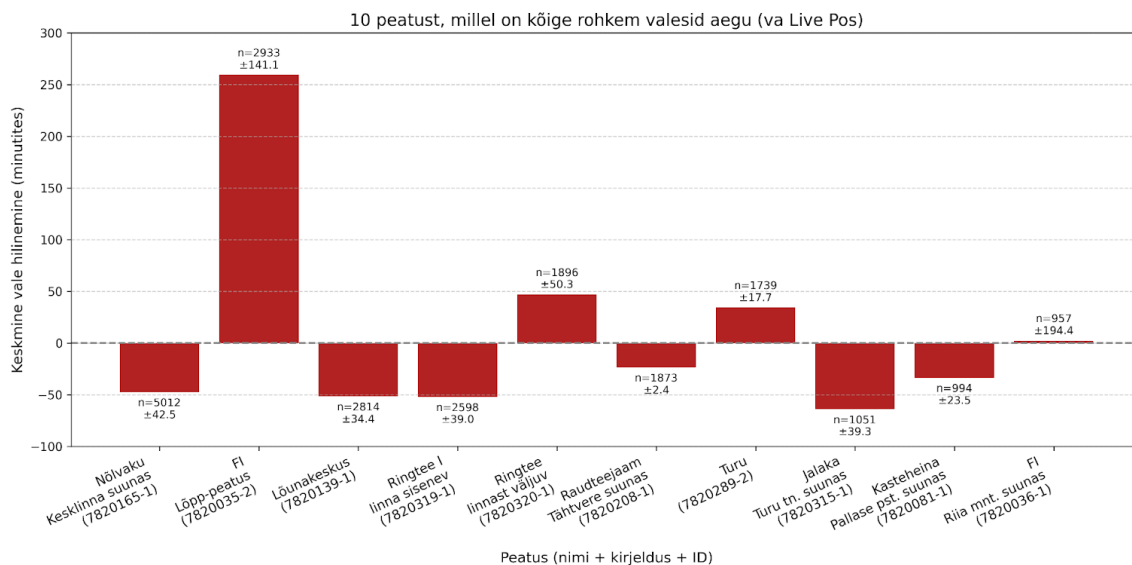


Joonis 22. Enim valeinfot sisaldavate teadete arv võrdluses peatuste ja liinidega.

Joonisel 23 on kujutatud kümme peatust, kus esines enim vale hilinemise juhtumeid. Andmetest jäeti Live Pos andmed kõrvale. Silmapaistvalt kõrge keskmine viga on Fi lõpp-peatuses (keskmiselt 141 minutit) ja Fi Riia mnt. suunas (keskmiselt 194 minutit), mis viitab tõenäolisele süsteemsele probleemile selles piirkonnas või marsruudis.

Mitmed teised peatused, nagu Lõunakeskus, Ringtee, Raudteejaam ja Jalaka, näitavad negatiivseid keskmisi, mis tähendab, et busside saabumisi prognoositi oluliselt varem kui need tegelikult toimusid.

Kuna Nõlvaku, Fi ja Kasteheina on mitmete liinide lõpp-peatused, siis võib järeldada, et bussid ei edasta täpset infot oma liikumiste kohta reise alguses ja lõpus.



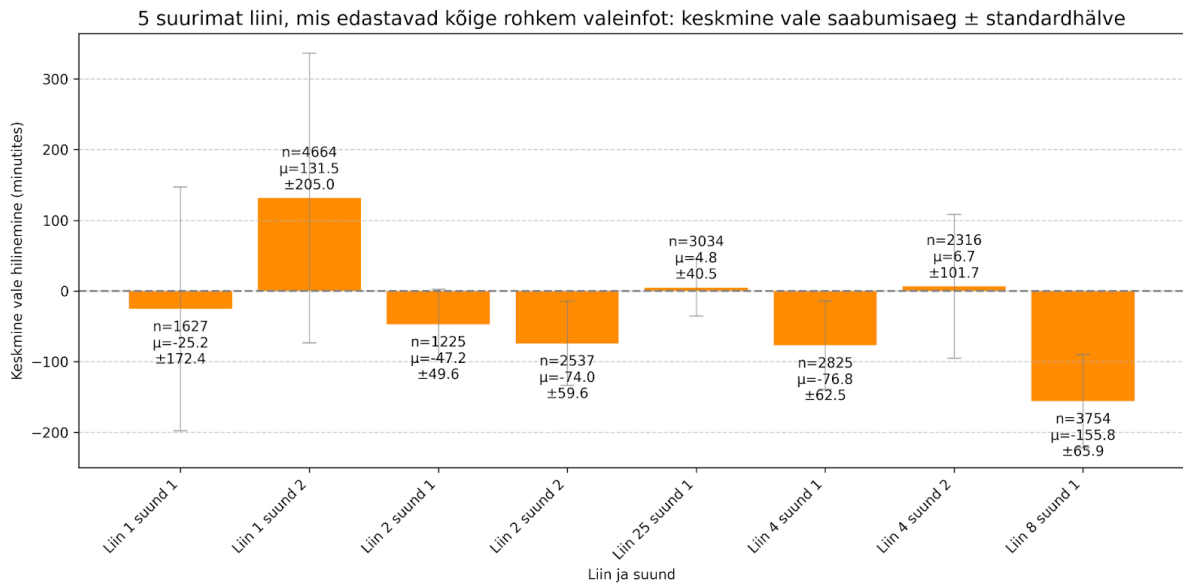
Joonis 23. 10 peatust, millel on kõige rohkem valesid aegu.

Joonisel 24 on kujutatud viit liini ja suunda, kus esines kõige rohkem vale saabumisinfot. Kõige probleemsem on liin 1, mõlemas suunas.

Need väärtused viitavad tugevale süsteemsele probleemile just liin 1 puhul. Märkimisväärne on ka, et liin 1 edastab suunal 2 palju suuremat hilinemist samal ajal kui teised liinid ennustavad suurelt varajast jõudmist.

Teised liinid näitavad valdavalt negatiivseid keskmisi, mis tähendab, et süsteem ennustas bussi saabumist liiga vara. Kõrged standardhälbed viitavad andmete ebastabiilsusele.

Kokkuvõttes vajab liin 1 tehnilist tähelepanu, kuna sellel on nii suurim valeinfo hulk kui ka kõige ulatuslikum kõrvalekalle.



Joonis 24. 5 suurimat liini, mis edastavad kõige rohkem valeinfot.

Andmete põhjal on selge, et busside GPS-andmete kvaliteet on ebaühtlane ning sõltub suuresti konkreetsest sõidukist ja marsruudist. Näiteks edastas buss 435 TNS kokku 3672 vale hilinemise juhtumit, mille keskmine kõrvalekalle tegelikust saabumisajast oli 195 minutit. Selline ulatuslik ja püsiv viga viitab bussi seadme probleemile, mis võib olla seotud seadme tarkvara, ajasünkroniseerimise või andmeedastuse usaldusväärsusega.

Tähelepanu väärivad ka bussid 340 TNS ja 352 TNS, mille puhul esines peatusesse saabumise suurt alahindamist. Sellised negatiivsed kõrvalekalded viitavad tugevale prognoosimismudeli või andmevoo veale.

Lisaks esineb suur hulk tuvastamata busside andmeid, mida tähistatakse sündmuse tüübina Live Pos. Selliseid ridu oli 35 706 ja need kõik tulid liin 1 kahest peatusest: Nõlvaku ja Fi. Sellest tulenevalt peaks põhjalikumalt uurima, mis on liin 1 seadistustes valesi, et ta järjepidevalt annab valesid hinnanguid nendesse peatustesse hilinemise osas.

Kui Live Pos andmed välja jätta, ka siis on liin 1 kõige problemaatilisem. Selles liinis esines kõige rohkem valeinfot mõlemas sõidusuunas. Suunal 1 oli keskmine vale hilinemine -155 minutit, samas kui suunal 2 oli see 131 minutit. Selline vastuoluline ja ebastabiilne käitumine viitab vajadusele liinipõhiseks tehniliseks kontrolliks, sest probleem ei piirdu üksnes seadmega, vaid võib olla seotud ka liini töökorralduse või seadistustega.

6. Tuleviku edasiarendused

Töö käigus ilmnis mitmeid kitsaskohti, mille lahendamine võimaldaks tulevikus suurendada andmeanalüüsi täpsust ning tugevdada Tartu linna ühistranspordi planeerimise ja juhtimise võimekust. Parandused võiksid hõlmata nii tehnilist andmekogumist kui ka andmetöötluste ja -kasutuse strateegiat.

Üheks olulisemaks piiranguks osutus GPS-andmete ruumiline ebatäpsus, mis mõjutab busside tegelike peatumiskohtade hindamise täpsust. Analüüs näitas, et peatusesündmused registreeruvad sageli mitukümmend meetrit ametlikust peatuspunktist eemal, mis vähendab ruumilise analüüsi usaldusväärsust. Samuti ilmnis, et mitme peatuse ametlikud koordinaadid ei ühtinud järjepidevalt busside tegelike peatumiskohtadega, mis viitab võimalikule lahknevusele GTFS-andmestiku ja reaalse liikluspildi vahel.

Teiseks tuleks arendada hilinemiste ennustussüsteemi. Praegune prognoos tugineb ainult bussi enda liikumiskiirusele ja varasematele sündmustele, jättes arvesse võtmata mitmeid olulisi mõjutegureid. Tulevikulahendused võiksid hõlmata reaalajas liiklusvoo andmeid, ilmastikutingimusi, kalendripõhiseid erisusi (nt koolivaheajad, pühad), samuti liinispetsiifilisi ajaloolisi hilinemismustreid. Sellise andmestiku toel oleks võimalik rakendada masinõppemudeleid, mis suudaksid pakkuda märgatavalt täpsemaid saabumisprognoose kui staatiline loogikapõhine süsteem.

Lisaks tuleks tähelepanu pöörata andmete ligipääsetavusele ja visualiseerimise võimalustele. Andmestike sidumine selliste platvormidega nagu ArcGIS Online võimaldaks reaalajas seiret, andmepõhist aruandlust ning interaktiivset visualiseerimist. Automatiseeritud andmepuhastus-, filtreerimis- ja visualiseerimismehhanismid vähendaksid manuaalset töökoormust ning muudaksid analüüsi tulemused kättesaadavamaks ka mitte-tehnilistele kasutajatele, näiteks transpordiplaneerijatele või linnaelanikele.

Et käesolevas töös kogutud andmestik oleks sobiv ArcGISi platvormile edastamiseks, on vajalik selle täiendav agregatsioon. Praegusel kujul on andmemahd liiga suur, et toetada sujuvat veebipõhist visualiseerimist. Samas sisaldavad andmed kõiki vajalikke välju (sh peatumisaeg, hilinemine, peatuskood, sõiduki ID ja geokoordinaadid), et võimaldada ruumilist analüüsi busside peatumiste ja hilinemiste kohta. Sobiva eeltöötluste järel oleks võimalik integreerida andmestik ArcGISi keskkonda nii, et analüüsitulemused oleksid dünaamiliselt jälgitavad ja praktiliseks kasutuseks valmis.

Lisatäpsuse tagamiseks võiks andmestikku rikastada ka sõidukipõhiste diagnostikaandmetega, nagu signaalikadu, seadme restardid või ühenduse katkestuste logid. Need võimaldaksid tuvastada mitte ainult tulemuse (vale hiline mine), vaid ka selle võimalikud tehnilised põhjused.

7. Kokkuvõte

Käesolev bakalaureusetöö uuris Tartu linna bussiliikluse hilinemiste ja peatumiste täpsust ajavahemikus 11.10.2019–09.08.2024, kasutades Cumulocity IoT platvormi sündmusandmeid ja GTFS formaadis peatuste andmestikku. Töö eesmärk oli analüüsida, millisel määral vastavad busse ennustatud saabumisaegad tegelikule liikluspildile, tuvastada ajaliselt ebausaldusväärseid liinid ja peatused ning hinnata GPS-andmete ruumilist täpsust.

Tulemused näitavad, et sõidugraafikute muudatustel jaanuaris ja septembris ei ole olnud üheselt positiivset mõju hilinemiste vähendamisele. Vastupidi, mitmetel aastatel suurenes hilinemiste maht just graafikumuudatustele järgnevatel kuudel. Prognoositavate ja tegelike saabumisaegade võrdlus näitas, et kuigi üldised keskmised võivad kattuda, esinevad kuude ja nädalate lõikes märkimisväärsed kõrvalekalded. Eriti 2023. aasta mais ja juunis olid ennustatud hilinemised oluliselt väiksemad kui tegelikud, mis viitab vajadusele täiustada olemasolevaid prognoosimudeleid.

Liinide lõikes osutus kõige ebausaldusväärsemaks liin 2, mille hilinemised olid järjepidevalt suured. Täpsemad ja stabiilsemad olid liinid 21 ja 22. Peatuste analüüs näitas, et suuremad hilinemised koondusid aastate lõikes kindlatesse peatustesse nagu Lammimõisa, Pallase puiestee ja Kõrveküla. Samuti ilmnisid korduvad mustrid kuude lõikes, mis viitab hooajalistele ja asukohaspetsiifilistele kitsaskohtadele.

GPS-andmete ruumilise täpsuse hindamisel selgus, et enamik busse peatus vähem kui 15 meetri kaugusel ametlikust peatuspunktist. Siiski esines mitmeid erandeid, näiteks Kasarmu ja Vahipargi peatused, kus busside keskmine peatumiskaugus ületas 60 meetrit. Sellised kõrvalekalded viitavad tõenäoliselt ebatäpsele peatuste koordinaatide määratlusele GTFS andmestikus.

Andmekvaliteedi analüüs tõi esile mitmed probleemid, sealhulgas liinil 1, kus esines kõige rohkem sündmusi, mille põhjal ei olnud võimalik sõidukit korrektselt tuvastada. Samuti paistis silma buss 435TNS, mille puhul esines arvukalt valehilinemise juhtumeid. Live Pos sündmused, kus sõiduki identiteet jäi määramata, koondusid samuti liin 1 lõpp-peatustesse.

Töö käigus tuli töödelda väga mahukat andmestikku, mis koosnes miljonitest sündmuseridadest ning ulatus kokku sadadesse gigabaitidesse. Sellise andmemahu töötlemine nõudis mitmeastmelist puhastust, andmete filtreerimist ja optimeerimist, et võimaldada usaldusväärne ja skaleeritav analüüs piiratud arvutusressursside tingimustes.

Töö tulemusena loodi puhastatud ja struktureeritud andmestik, mis võimaldab tulevikus täpsemaid analüüse ning toetab tõenduspõhiste otsuste tegemist Tartu linna transpordi korralduses. Töö pakkus ka tehnilisi ja metoodilisi soovitusi andmekvaliteedi parandamiseks, prognoosimudelite täiendamiseks ning reaalaja andmete parema visualiseerimise võimaldamiseks.

Töö panustab Tartu ühistranspordi andmepõhise juhtimise arengusse, pakkudes senisest täpsemat ülevaadet liinide töökindlusest, peatuskohtade täpsusest ja reaalajas süsteemide toimivusest. Selle rakenduspotentsiaal ulatub linnaplaneerimise, transpordikorralduse ja kasutajakogemuse parandamise valdkondadeni.

Viidatud kirjandus

- [1] “colab.google,” colab.google. Accessed: May 15, 2025. [Online]. Available: <http://0.0.0.0:8080/>
- [2] “Hello GPT-4o.” Accessed: May 15, 2025. [Online]. Available: <https://openai.com/index/hello-gpt-4o/>
- [3] “Route design for last-in, first-out deliveries with backhauling - ScienceDirect.” Accessed: May 15, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0968090X17300128?via%3Dihub>
- [4] M. Pourmoradnasseri, “Exploring the Impact of COVID-19 on Tartu Smart Bike Usage”.
- [5] P. Vahe, “Tartu Smart Bike Share Access Cards Authentication Analysis”.
- [6] M. Pourmoradnasseri and A. Hadachi, “Alighting Estimation in Entry-Only AFC Systems; a Case Study of Tartu City”.
- [7] “SmartEnCity.” Accessed: Mar. 31, 2023. [Online]. Available: <https://smartencity.eu/about/>
- [8] “Horisont 2020 (8RP),” Eesti Teadusagentuur. Accessed: May 15, 2025. [Online]. Available: <https://etag.ee/valiskoostoo/euroopa-horisont/horisont-2020/>
- [9] G. E. Sánchez-Martínez, H. N. Koutsopoulos, and N. H. M. Wilson, “Real-time holding control for high-frequency transit with dynamics,” *Transportation Research Part B: Methodological*, vol. 83, pp. 1–19, Jan. 2016, doi: 10.1016/j.trb.2015.11.013.
- [10] “Measurement and classification of transit delays using GTFS-RT data | Public Transport.” Accessed: May 15, 2025. [Online]. Available: <https://link.springer.com/article/10.1007/s12469-022-00291-7>
- [11] P. Li and J. He, “Geometric design safety estimation based on tire–road side friction,” *Transportation Research Part C: Emerging Technologies*, vol. 63, pp. 114–125, Feb. 2016, doi: 10.1016/j.trc.2015.12.009.
- [12] “(PDF) Real-Time Passenger Train Delay Prediction Using Machine Learning: A Case Study With Amtrak Passenger Train Routes,” *ResearchGate*, Dec. 2024, doi: 10.1109/OJITS.2022.3194879.
- [13] “Software AG Acquires Cumulocity for IoT Cloud PaaS.” Accessed: May 15, 2025. [Online]. Available: https://www.vdcresearch.com/News-events/iot-blog/Software-AG-Acquires-Cumulocity-for-IoT-Cloud-PaaS.html?utm_source=chatgpt.com
- [14] “Home dashboard - Cumulocity documentation.” Accessed: May 15, 2025. [Online]. Available: <https://cumulocity.com/docs/>
- [15] “JSON Lines Examples.” Accessed: May 16, 2025. [Online]. Available: <https://jsonlines.org/examples/>
- [16] “OpenStreetMap,” OpenStreetMap. Accessed: May 16, 2025. [Online]. Available: <https://www.openstreetmap.org/>

Lisad

Lisa 1. Link näidis json failidele enne töötlemist

Link OneDrive näidis json failie: [Töötlemata bussiandmed](#)

Lisa 2. Link puhastatud andmekogule

Link bussandmete puhastatud andmekogule OneDriveis: [Tartu busside hilineviste analüüs](#)

Puhastatud andmekogus on olemas kõik Cumulocityst tulnud busside andmed vahemikus 2019-2024. Need on kokku pakitud .gz formaadis.

Lisa 3. Link analüüsi GitHubi lähtekoodile

Link analüüsi lähtekoodile GitHubis:

https://github.com/HannaMariaMagi/Tartu_busside_hilinemiste_ja_asukohtade_analuus/tree/master

Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina,

Hanna Maria Mägi

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

Tartu busside hilinemiste ja peatumiste analüüs,

mille juhendaja on Pelle Jakovits,

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 3.0, 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.

Hanna Maria Mägi

15.05.2025