



LOPPUPROJEKTI

PROJEKTIRYHMÄN TYÖNJAKO

- Datapipelinen ja Ennustemallin teko
- Datan puhdistus ja analysointi
- Datan puhdistus ja analysointi
- Product Owner ja datan visualisointi

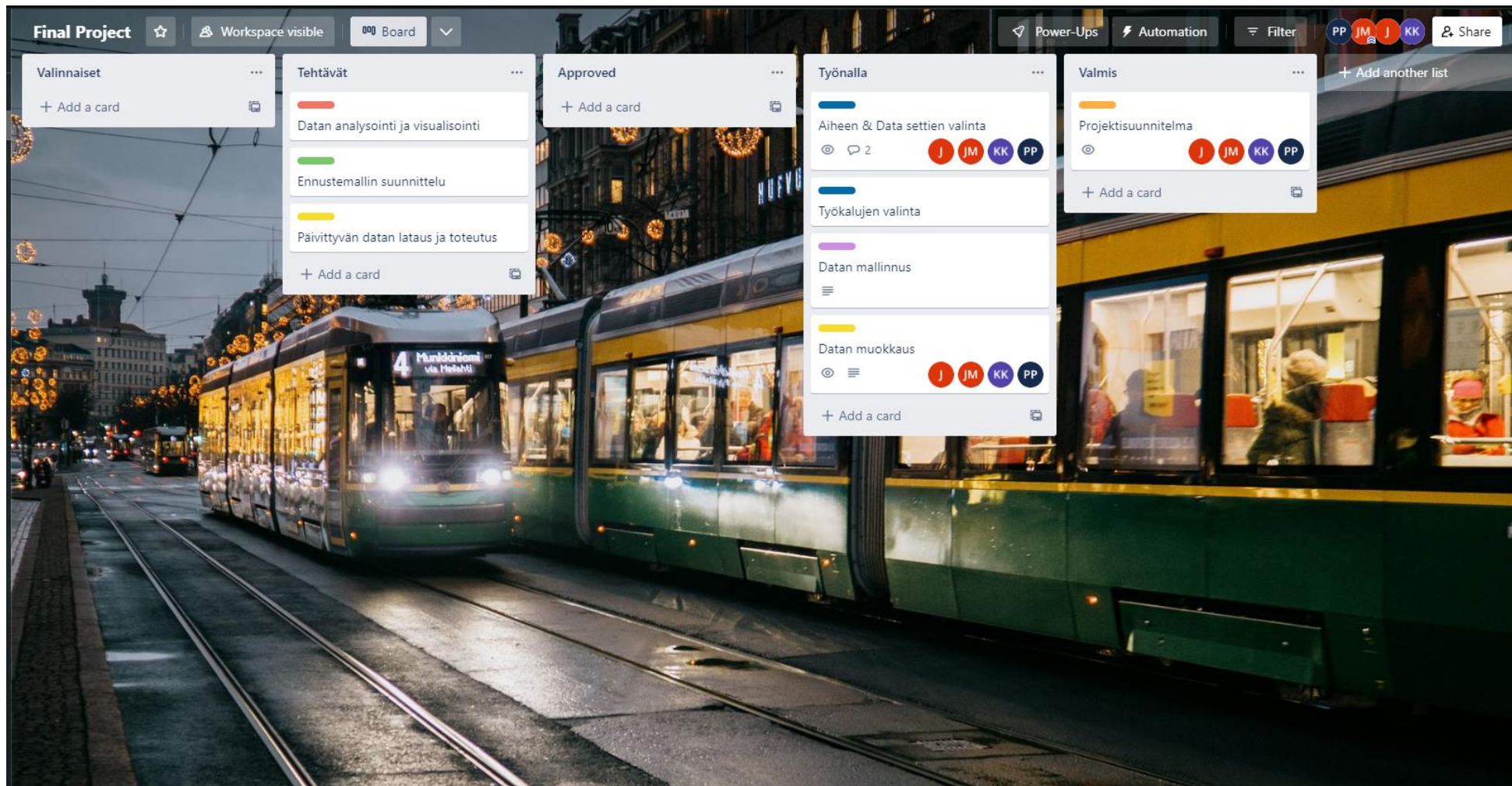
PROJEKTIN ESITTELY

- Sään vaikutus Helsingin seudun henkilö- ja pakettiautojen liikennemääriin ja ajoneuvojen nopeuksiin. Liikenteen mittauspisteiksi valittiin kolme pistettä eri puolelta pk-seutua ja näitä lähellä olevat säähavaintopisteet.
- Ajanjaksoiksi määriteltiin 2019 – 2023.
- Tarkoitus oli lisäksi selvittää onko sään ja aiempien liikennemäärien perusteella mahdollista luoda ennustemallia päivittäisen työmatkan suunnittelun avuksi.
- Datalähteenä käytettiin Ilmatieteenlaitoksen säähavaintodataa ja digitrafficin tieliikennedataa.

TEKNIIKAT

- Datan lataaminen ja muokkaus: Azure Synapse Analytics, Spark Notebook
 - historiadatan lataus csv:nä ja päivittyvä data haettiin API:n kautta
- Datan mallinnus: SQL DBM
- Tietovaraston toteutus: Azure Synapse data pipeline, Azure SQL database (basic)
- Datan analysointi ja visualisointi: PowerBI

PROJEKTIN HALLINTA & KETTERÄT MENETELMÄT



DATAN HAKU (HISTORIA)

[HTTPS://WWW.ILMATIETEENLAITOS.FI/HAVAINTOJEN-LATAUS](https://www.ilmatieteenlaitos.fi/havaintojen-lataus)

- Säädatan haku Ilmatieteenlaitoksen Havaintojen lataus sivulta

1

Valitse haettavat suuret

Säähavainnot	Säteilyhavainnot	Merihavainnot	Ilmanlaatuhavainnot
--------------	------------------	---------------	---------------------

☒ Hetkellishavainnot

☐ Pilvien määrä

☐ Ilmanpaine (msl)

☒ Sademäärä

☐ Suhteellinen kosteus

☒ Sateen intensiteetti

☒ Lumensyvyys

☒ Ilman lämpötila

☐ Kastepistelämpötila

☒ Näkyvyys

☐ Tuulen suunta

☐ Puuskanopeus

☒ Tuulen nopeus

Havaintoväli (tiheimmillään)

☐ 10 min

☒ 1 h

☐ Vuorokausihavainnot

☐ Sademäärä

☐ Lumensyvyys

☐ Ilman lämpötila

☐ Maanpintaminimi

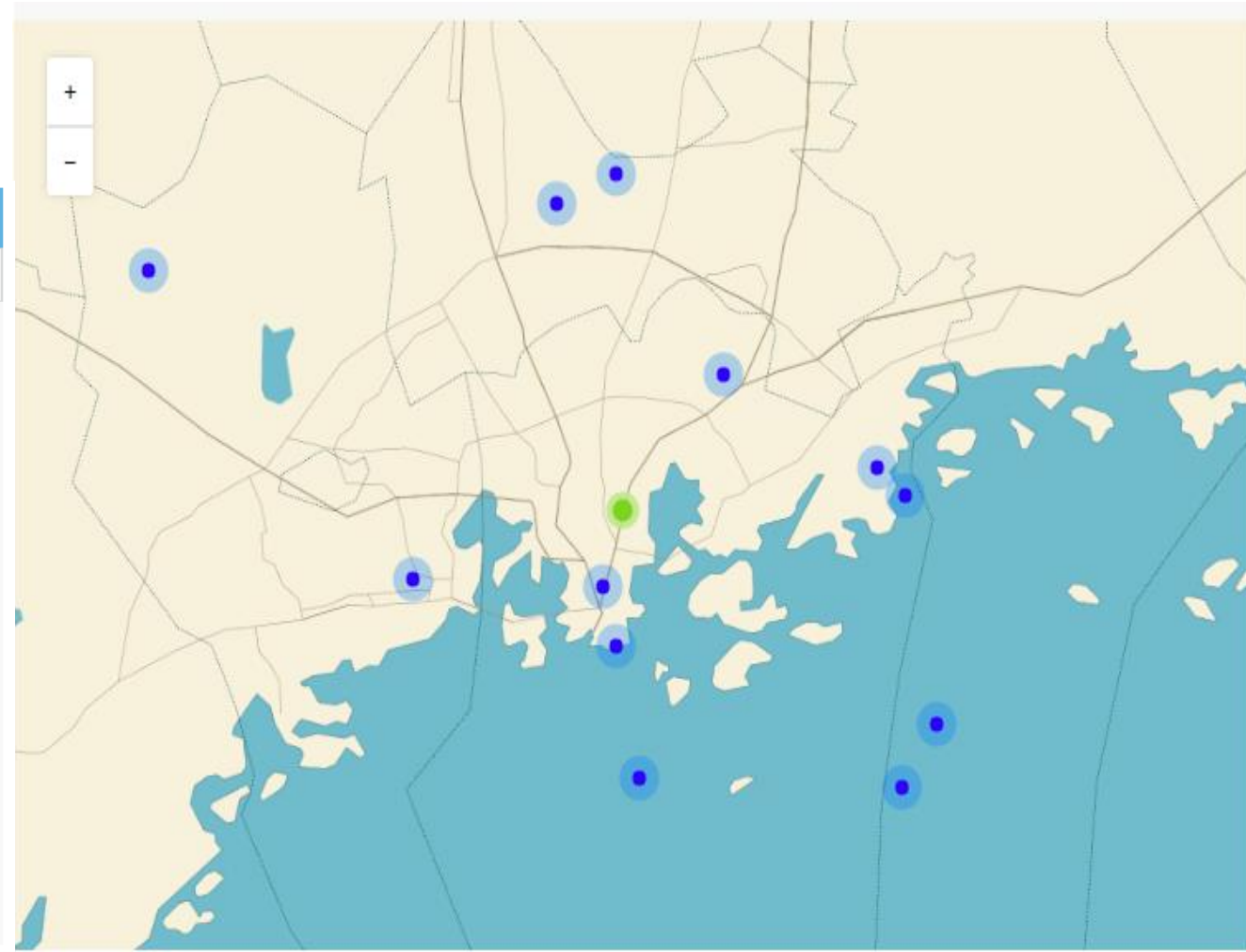
☐ Ylin lämpötila

☐ Alin lämpötila

☐ Kuukausihavainnot

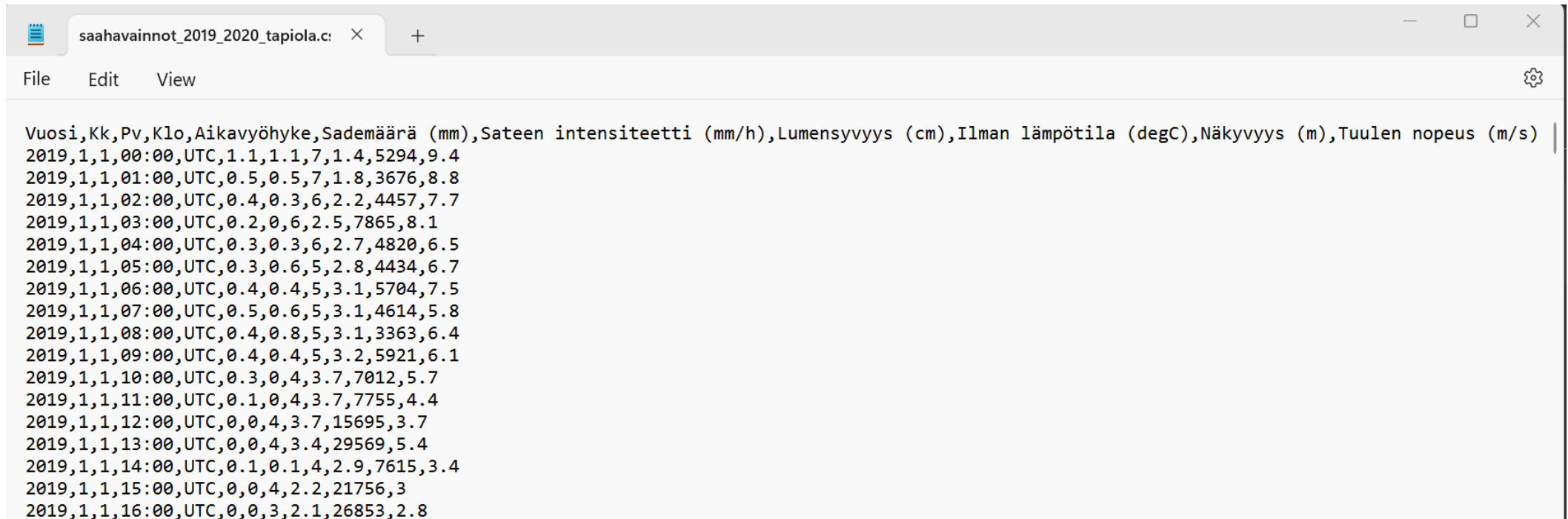
☐ Kuukauden sadesumma

☐ Kuukauden keskilämpötila



DATAN HAKU (HISTORIA)

- Säädatan lataus csv:nä 2 vuoden erissä/mittauspiste



The screenshot shows a text editor window with the title bar "saahavainnot_2019_2020_tapiola.csv". The editor contains a CSV file with the following data:

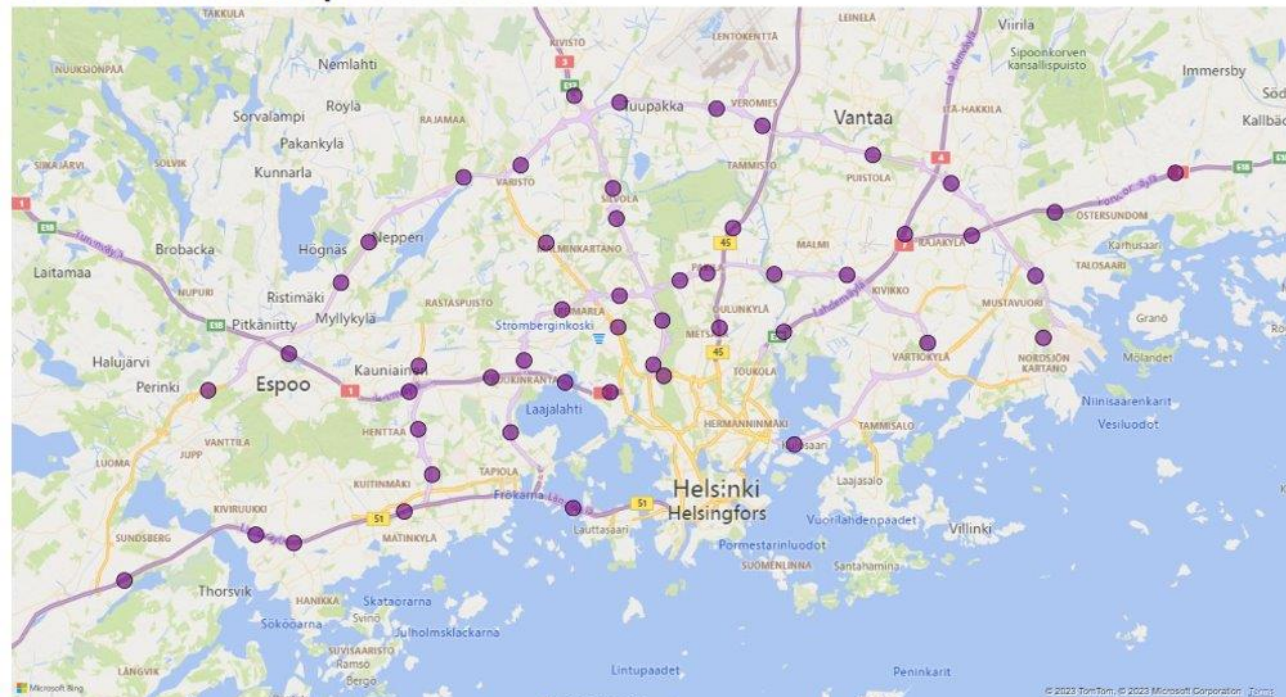
Vuosi	Kk	Pv	Klo	Aikavyöhyke	Sademäärä (mm)	Sateen intensiteetti (mm/h)	Lumensyvyys (cm)	Ilman lämpötila (degC)	Näkyvyys (m)	Tuulen nopeus (m/s)
2019	1	1	00:00	UTC	1.1	1.1	7	1.4	5294	9.4
2019	1	1	01:00	UTC	0.5	0.5	7	1.8	3676	8.8
2019	1	1	02:00	UTC	0.4	0.3	6	2.2	4457	7.7
2019	1	1	03:00	UTC	0.2	0.6	2.5	7865	8.1	
2019	1	1	04:00	UTC	0.3	0.3	6	2.7	4820	6.5
2019	1	1	05:00	UTC	0.3	0.6	5	2.8	4434	6.7
2019	1	1	06:00	UTC	0.4	0.4	5	3.1	5704	7.5
2019	1	1	07:00	UTC	0.5	0.6	5	3.1	4614	5.8
2019	1	1	08:00	UTC	0.4	0.8	5	3.1	3363	6.4
2019	1	1	09:00	UTC	0.4	0.4	5	3.2	5921	6.1
2019	1	1	10:00	UTC	0.3	0.4	3.7	7012	5.7	
2019	1	1	11:00	UTC	0.1	0.4	3.7	7755	4.4	
2019	1	1	12:00	UTC	0.0	0.4	3.7	15695	3.7	
2019	1	1	13:00	UTC	0.0	0.4	3.4	29569	5.4	
2019	1	1	14:00	UTC	0.1	0.1	4	2.9	7615	3.4
2019	1	1	15:00	UTC	0.0	0.4	2.2	21756	3	
2019	1	1	16:00	UTC	0.0	0.3	2.1	26853	2.8	

DATAN HAKU (HISTORIA)

HTTPS://TIE.DIGITRAFFIC.FI/UI/TMS/HISTORY/

- Liikennedatan haku digitrafficin LAM-tilastohaun avulla

Liikenteen mittauspisteet kartalla



LAM-tilastohaku

Aineisto ②

Liikennemäärät

Raportti

Tuntiliikenneraportti

Aika

Alkuaika

2020-01-01

Loppuaika

2023-03-31

LAM-pisteet ②

☒ Yksittäisten pisteiden valinta

154

Listaus

☐ Pistejoukkojen valinta

LAM-pistejoukot

Listaus

Ajoneuvoluokat

Valitse ajoneuvoluokka (voit valita useampia)
Kaikki (yhteensä)
Kevyet ajoneuvot (1,6,7,8)
Raskaat ajoneuvot (2,3,4,5,9)
Henkilo- tai pakettiauto

Suunnat

Kaikki (suunnat summattuna)
Suunta 1
Suunta 2

Kaistat

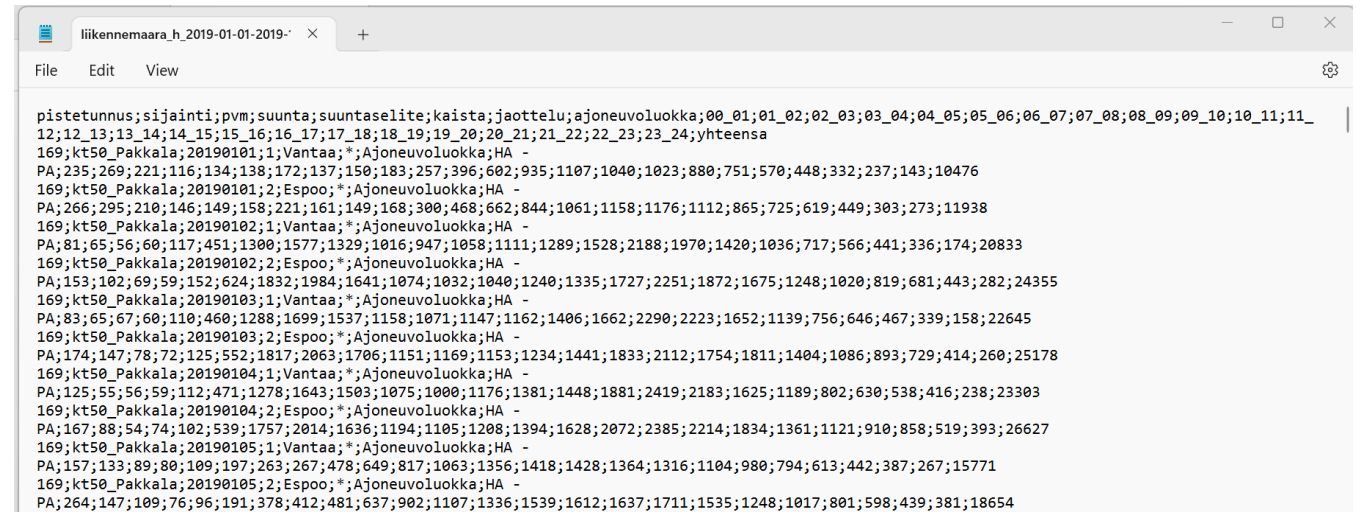
Ei - kaistat summattuna
Kyllä - kaistat eriteltynä

Tyhjennä lomake

Lataa tiedot (CSV)

DATAN HAKU (HISTORIA)

- Liikennedatan lataus csv:nä
- Valitut kentät:
 - pistetunnus
 - sijainti
 - pvm
 - suunta
 - suuntaselite
 - klo
 - Keskinopeus
 - Liikennemäärä



```
liikennemaara_h_2019-01-01-2019-
File Edit View
pistetunnus;sijainti;pvm;suunta;suuntaselite;kaista;jaottelu;ajoneuvoluokka;00_01;01_02;02_03;03_04;04_05;05_06;06_07;07_08;08_09;09_10;10_11;11_12;12_13;13_14;14_15;15_16;16_17;17_18;18_19;19_20;20_21;21_22;22_23;23_24;yhteensa
169;kt50_Pakkala;20190101;1;Vantaa*;Ajoneuvoluokka;HA -
PA;235;269;221;116;134;138;172;137;150;183;257;396;602;935;1107;1040;1023;880;751;570;448;332;237;143;10476
169;kt50_Pakkala;20190101;2;Espoo*;Ajoneuvoluokka;HA -
PA;266;295;210;146;149;158;221;161;149;168;300;468;662;844;1061;1158;1176;1112;865;725;619;449;303;273;11938
169;kt50_Pakkala;20190102;1;Vantaa*;Ajoneuvoluokka;HA -
PA;81;65;56;60;117;451;1300;1577;1329;1016;947;1058;1111;1289;1528;2188;1970;1420;1036;717;566;441;336;174;20833
169;kt50_Pakkala;20190102;2;Espoo*;Ajoneuvoluokka;HA -
PA;153;102;69;59;152;624;1832;1984;1641;1074;1032;1040;1240;1335;1727;2251;1872;1675;1248;1020;819;681;443;282;24355
169;kt50_Pakkala;20190103;1;Vantaa*;Ajoneuvoluokka;HA -
PA;83;65;67;60;110;460;1288;1699;1537;1158;1071;1147;1162;1406;1662;2290;2223;1652;1139;756;646;467;339;158;22645
169;kt50_Pakkala;20190103;2;Espoo*;Ajoneuvoluokka;HA -
PA;174;147;78;72;125;552;1817;2063;1706;1151;1169;1153;1234;1441;1833;2112;1754;1811;1404;1086;893;729;414;260;25178
169;kt50_Pakkala;20190104;1;Vantaa*;Ajoneuvoluokka;HA -
PA;125;55;56;59;112;471;1278;1643;1503;1075;1000;1176;1381;1448;1881;2419;2183;1625;1189;802;630;538;416;238;23303
169;kt50_Pakkala;20190104;2;Espoo*;Ajoneuvoluokka;HA -
PA;167;88;54;74;102;539;1757;2014;1636;1194;1105;1208;1394;1628;2072;2385;2214;1834;1361;1121;910;858;519;393;26627
169;kt50_Pakkala;20190105;1;Vantaa*;Ajoneuvoluokka;HA -
PA;157;133;89;80;109;197;263;267;478;649;817;1063;1418;1428;1364;1316;1104;980;794;613;442;387;267;15771
169;kt50_Pakkala;20190105;2;Espoo*;Ajoneuvoluokka;HA -
PA;264;147;109;76;96;191;378;412;481;637;902;1107;1336;1539;1612;1637;1711;1535;1248;1017;801;598;439;381;18654
```

DATAN HAKU (REAALIAIKA)

```
locations = {
    '100226':154,
    '100240':169,
    '100158':118,
}
df=pd.DataFrame()
for stationid, location in locations.items():

    url = f'https://lamapi.azurewebsites.net/api/Public/getDayById/{stationid}'
    response = requests.get(url)

    if response.status_code == 200:
        print("API request successful.")
        response_data = response.json()
        df2 = pd.json_normalize(response_data,record_path = 'data',meta='direction')
    else:
        print("Error:", response.status_code, response.reason)

    luku = (len(df2)-26)/2
    df2.insert(0,'sijainti',location)
    df2['time'] = pd.to_datetime(df2['time'], format='%H:%M')

    if len(df2)>26:
        df2.loc[len(df2)/2-luku:len(df2)/2-1,'time']+= pd.Timedelta(hours=12)
        df2.loc[len(df2)-luku:len(df2),'time']+= pd.Timedelta(hours=12)

    df2.loc[0,'time']+= pd.Timedelta(hours=12)
    df2.loc[len(df2)/2,'time']+= pd.Timedelta(hours=12)
    df2['time'] = df2['time'].dt.strftime('%H:%M')
    df2['pvm']=date.today()
    df = pd.concat([df, df2], ignore_index=True)
```

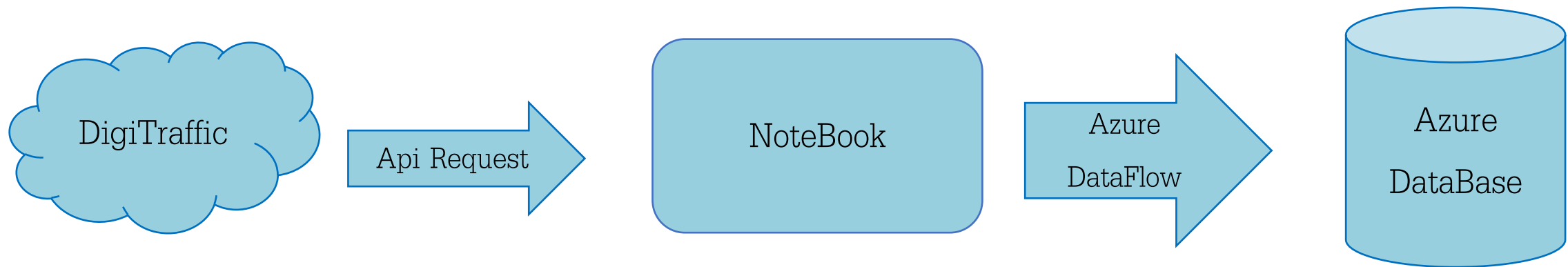
Haettu data

	count	speed	time	direction
0	172	94	12:00	Lahti
1	123	96	01:00	Lahti
2	83	93	02:00	Lahti
3	56	88	03:00	Lahti
4	74	92	04:00	Lahti
5	166	94	05:00	Lahti
6	408	97	06:00	Lahti
7	920	97	07:00	Lahti
8	974	96	08:00	Lahti
9	847	96	09:00	Lahti
10	989	95	10:00	Lahti
11	930	95	11:00	Lahti
12	1103	95	12:00	Lahti
13	122	91	12:00	Helsinki
14	73	92	01:00	Helsinki
15	47	93	02:00	Helsinki
16	62	91	03:00	Helsinki
17	153	91	04:00	Helsinki
18	595	97	05:00	Helsinki
19	2148	96	06:00	Helsinki
20	2424	95	07:00	Helsinki
21	2517	96	08:00	Helsinki
22	1559	95	09:00	Helsinki
23	1110	94	10:00	Helsinki
24	1146	94	11:00	Helsinki
25	1103	94	12:00	Helsinki

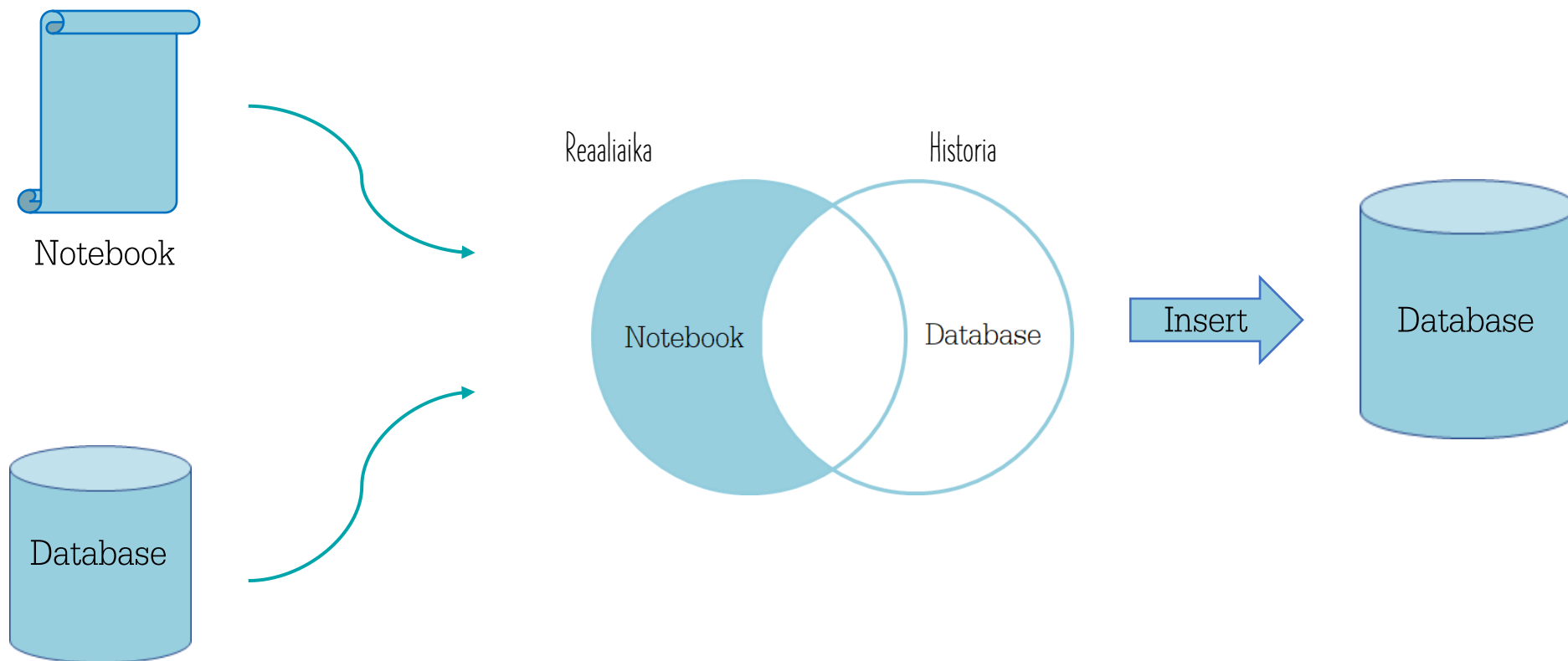
Muokattu data

	sijainti	count	speed	time	direction	pvm
0	154	172	94	00:00	Lahti	2023-04-12
1	154	123	96	01:00	Lahti	2023-04-12
2	154	83	93	02:00	Lahti	2023-04-12
3	154	56	88	03:00	Lahti	2023-04-12
4	154	74	92	04:00	Lahti	2023-04-12
5	154	166	94	05:00	Lahti	2023-04-12
6	154	408	97	06:00	Lahti	2023-04-12
7	154	920	97	07:00	Lahti	2023-04-12
8	154	974	96	08:00	Lahti	2023-04-12
9	154	847	96	09:00	Lahti	2023-04-12
10	154	989	95	10:00	Lahti	2023-04-12
11	154	930	95	11:00	Lahti	2023-04-12
12	154	1103	95	12:00	Lahti	2023-04-12
13	154	122	91	00:00	Helsinki	2023-04-12
14	154	73	92	01:00	Helsinki	2023-04-12
15	154	47	93	02:00	Helsinki	2023-04-12
16	154	62	91	03:00	Helsinki	2023-04-12
17	154	153	91	04:00	Helsinki	2023-04-12
18	154	595	97	05:00	Helsinki	2023-04-12
19	154	2148	96	06:00	Helsinki	2023-04-12
20	154	2424	95	07:00	Helsinki	2023-04-12
21	154	2517	96	08:00	Helsinki	2023-04-12
22	154	1559	95	09:00	Helsinki	2023-04-12
23	154	1110	94	10:00	Helsinki	2023-04-12
24	154	1146	94	11:00	Helsinki	2023-04-12
25	154	1103	94	12:00	Helsinki	2023-04-12

DATAN HAKU (PIPELINE)

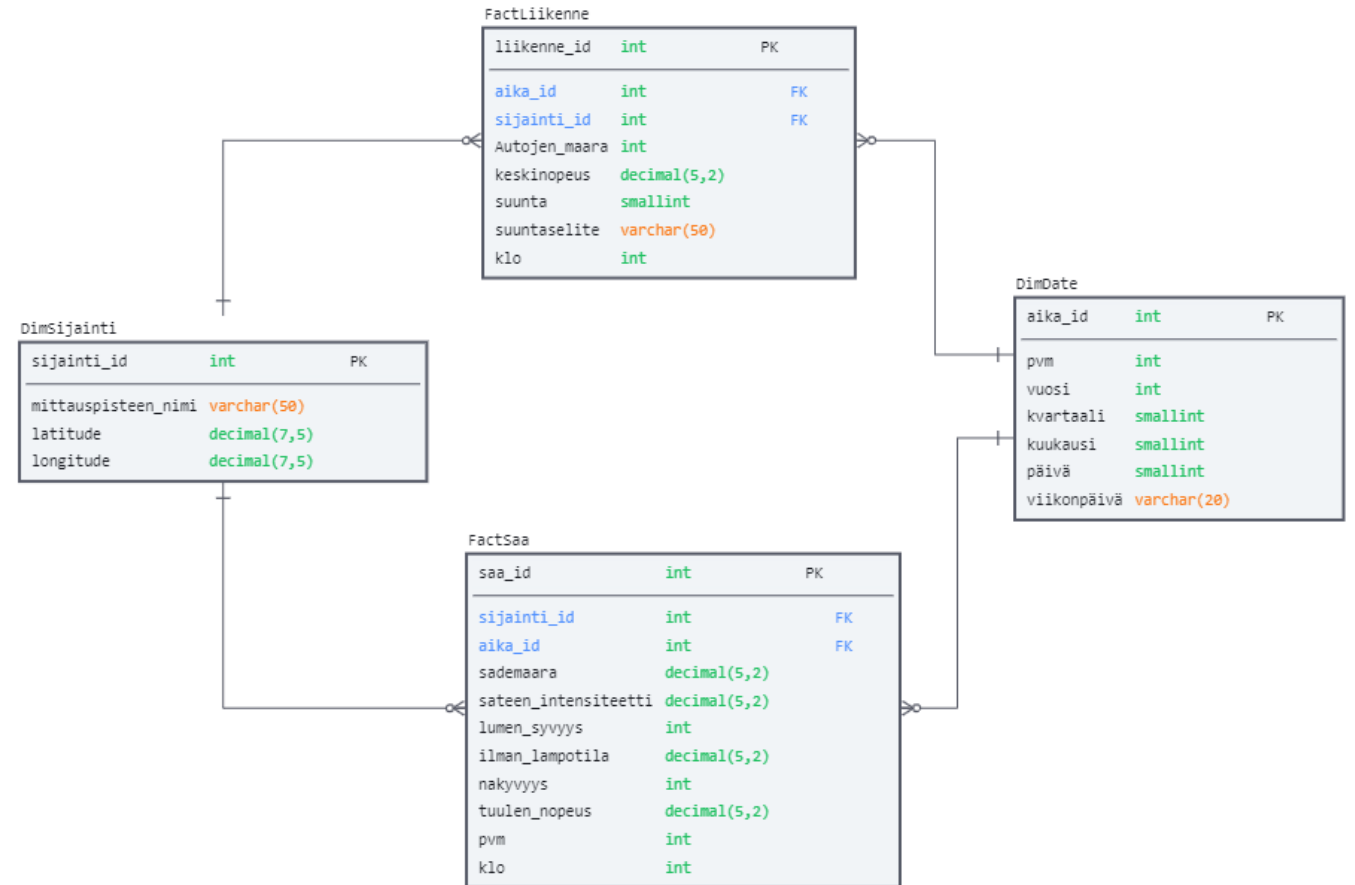


PIPELINE (DATAFLOW)



TIETOVARASTO

- Kaksi Faktataulua: Sää ja Liikenne
- Molemmilla Faktatauluilla Dimensio-
tauluina sijainti ja aika



ANALYSOINTI

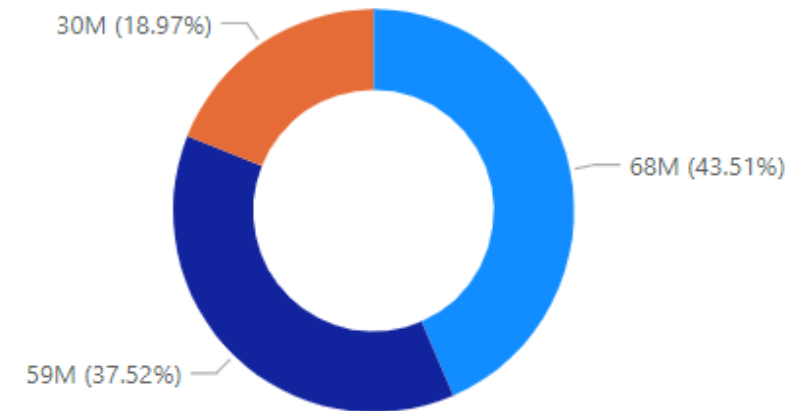
Nopeusrajoitukset

- Kt50_Pakkala -> 80 km/h
- St101_Espoo_Keilaniemi -> 60 km/h
- Vt4_Viikinmäki - > 100 km/h

Keskinopeusdataa ei saatavissa vuodelta 2019

Liikenteen jakautuminen

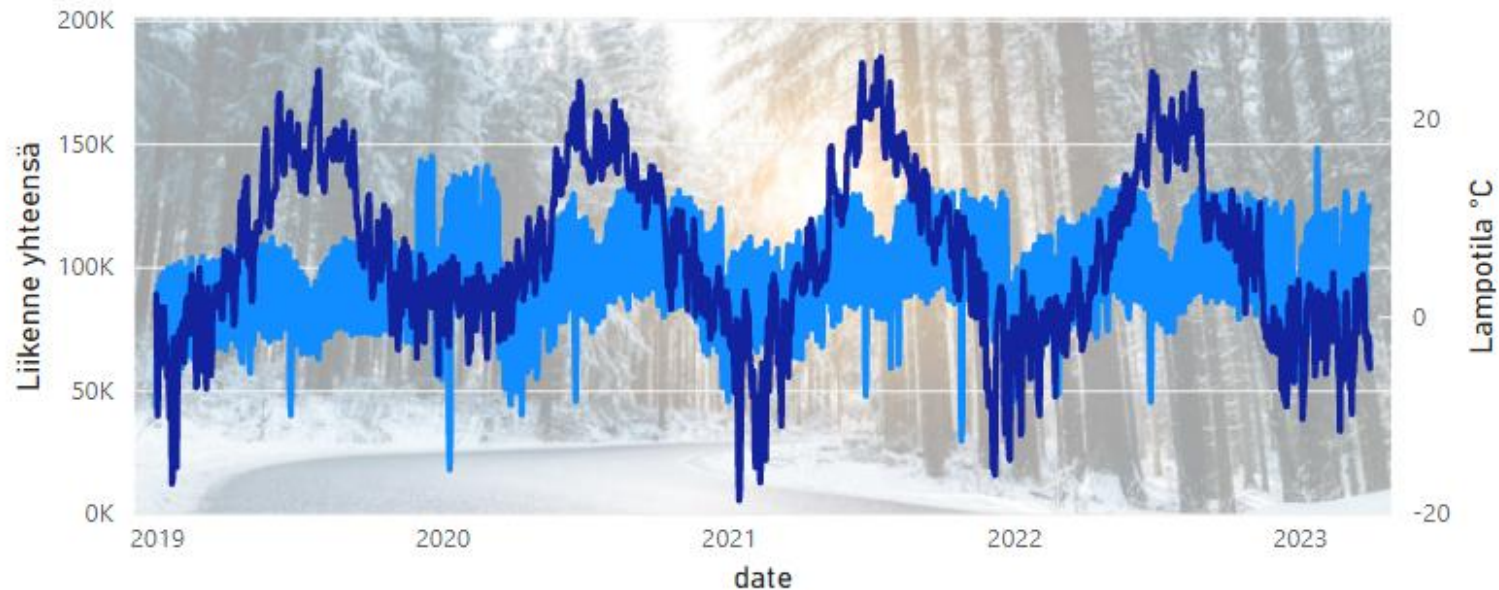
Mittauspiste ● kt50_Pakkala ● vt4_Viikinmäki ● st101_Espoo_Keilaniemi



DATAN ANALYSOINTI

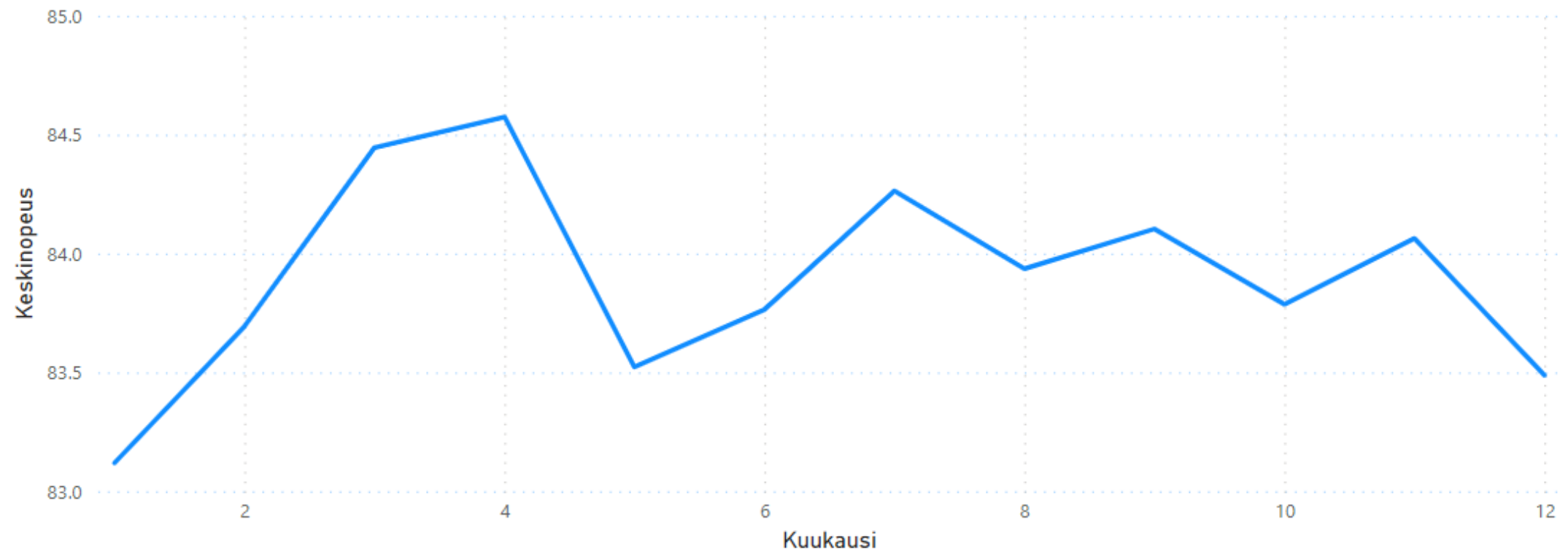


● Liikenne maara ● Keskilämpötila

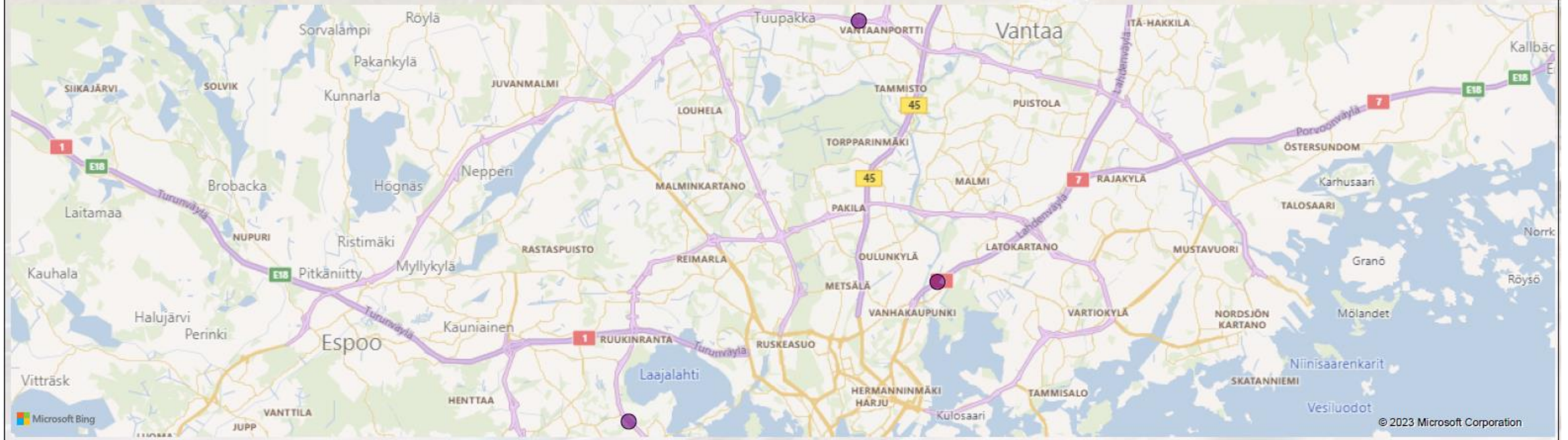


DATAN ANALYSOINTI

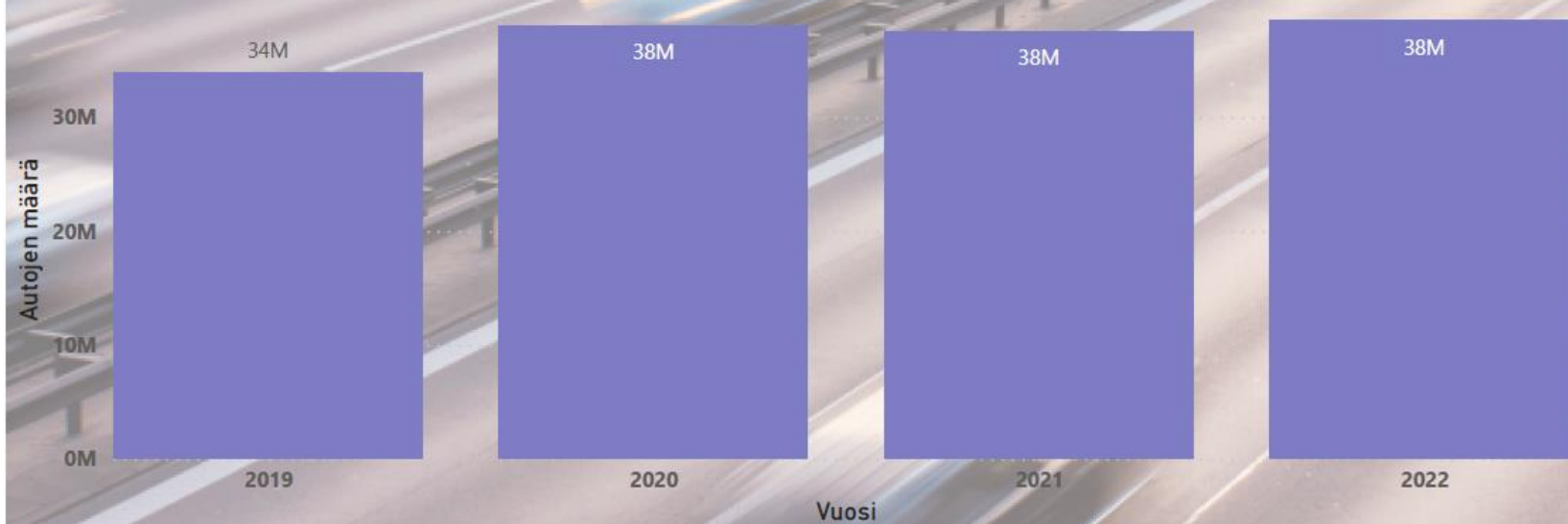
Keskinopeudet kuukausittain Pakkala



Liikennemittauspisteet



Autojen määrät / vuosi



Mittauspiste

kt50_Pakkala	vt4_Viikinmäki
st101_Espoo_Keilaniemi	

Ajosuunta

Espoo	Itäkeskus	Länsiväylä
Helsinki	Lahti	Vantaa

Mittauspiste

kt50_Pakkala

st101_Espoo_Keilaniemi

vt4_Viikinmäki

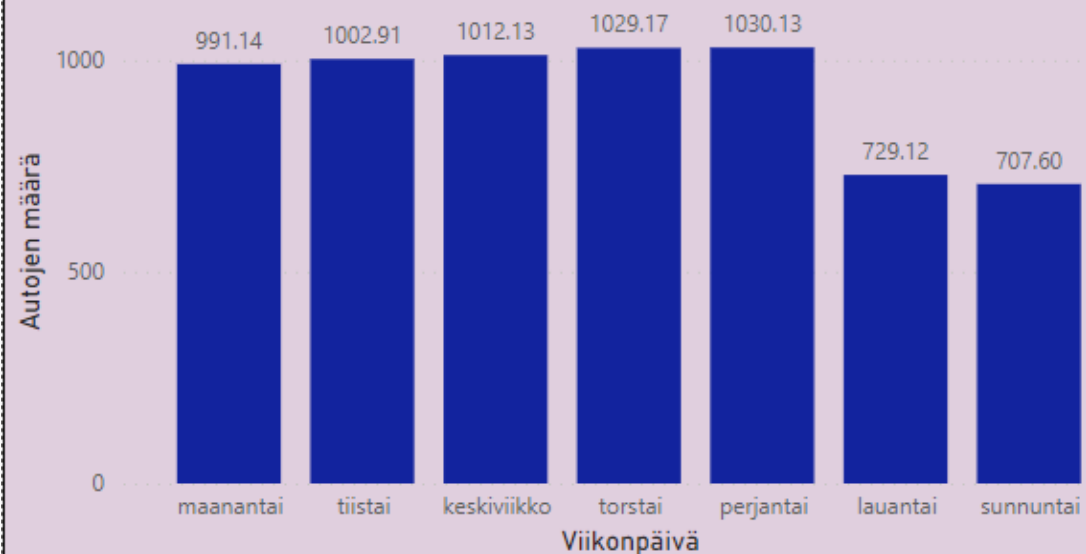
Vuosi

2020

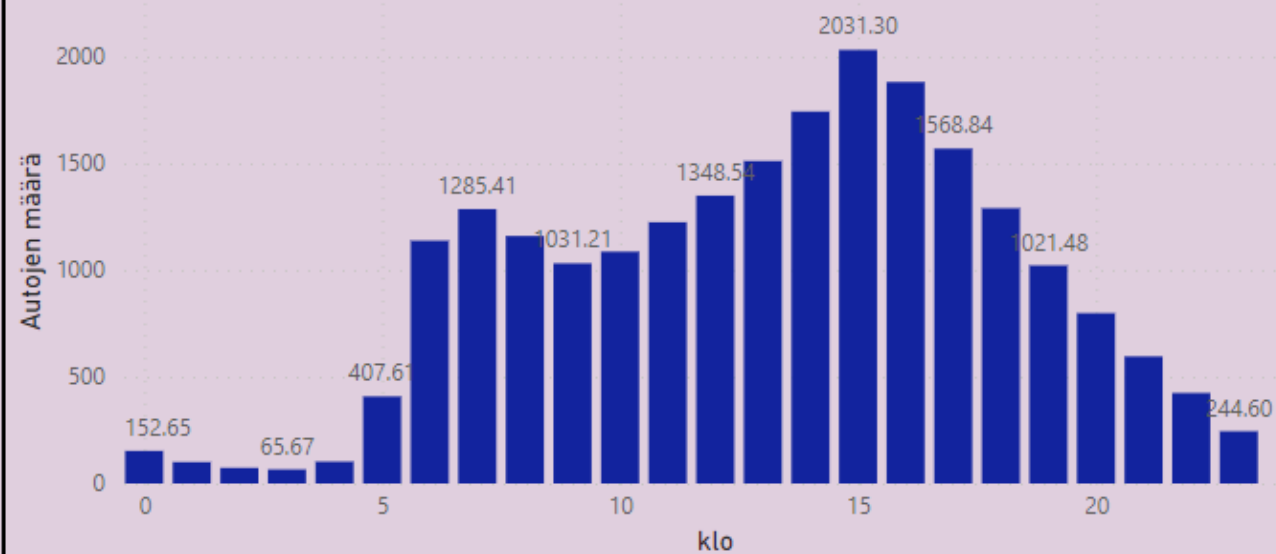
2021

2022

Autojen määrä / viikompäivä



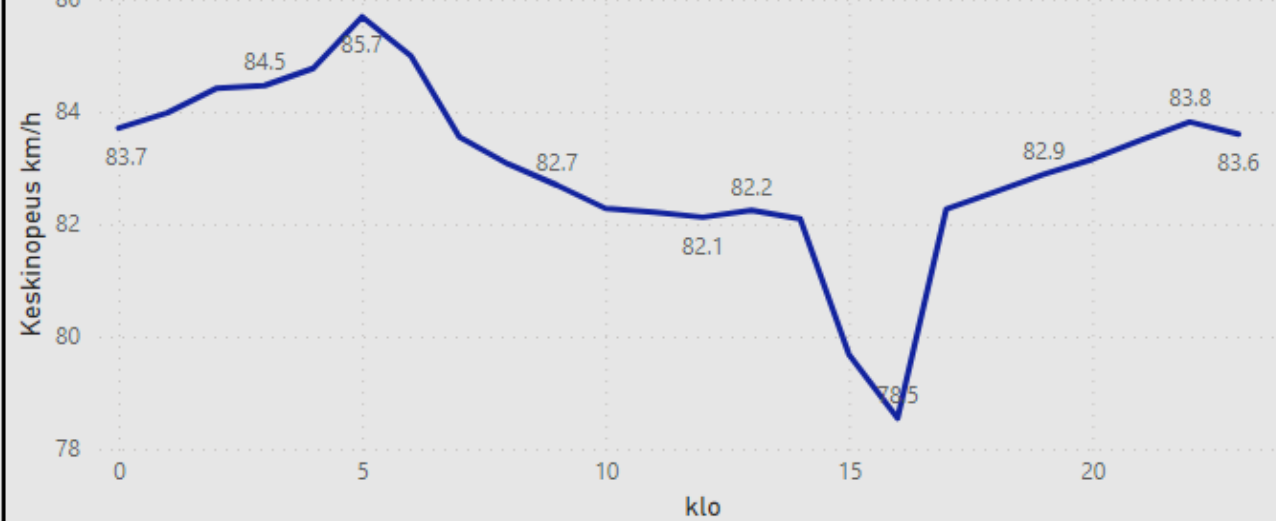
Autojen määrä / kelloaika

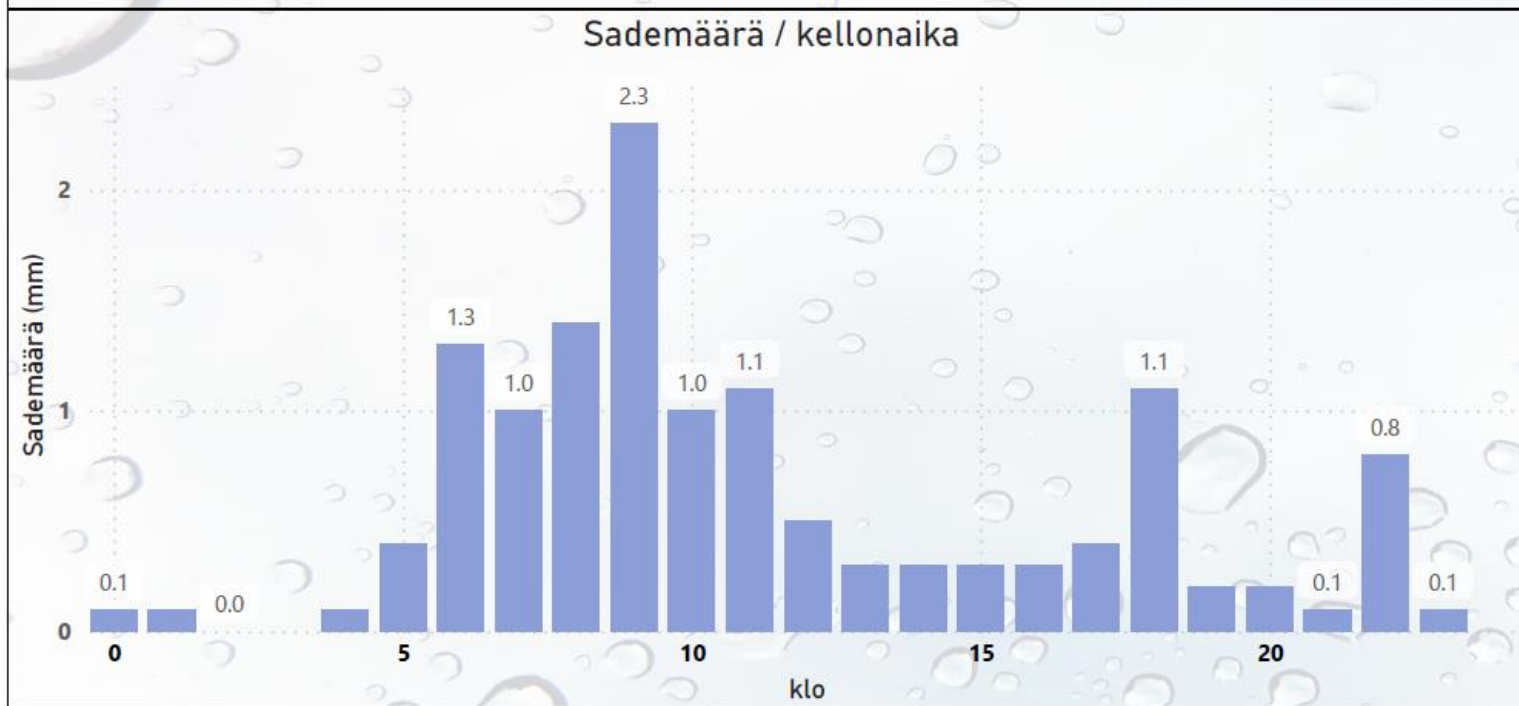
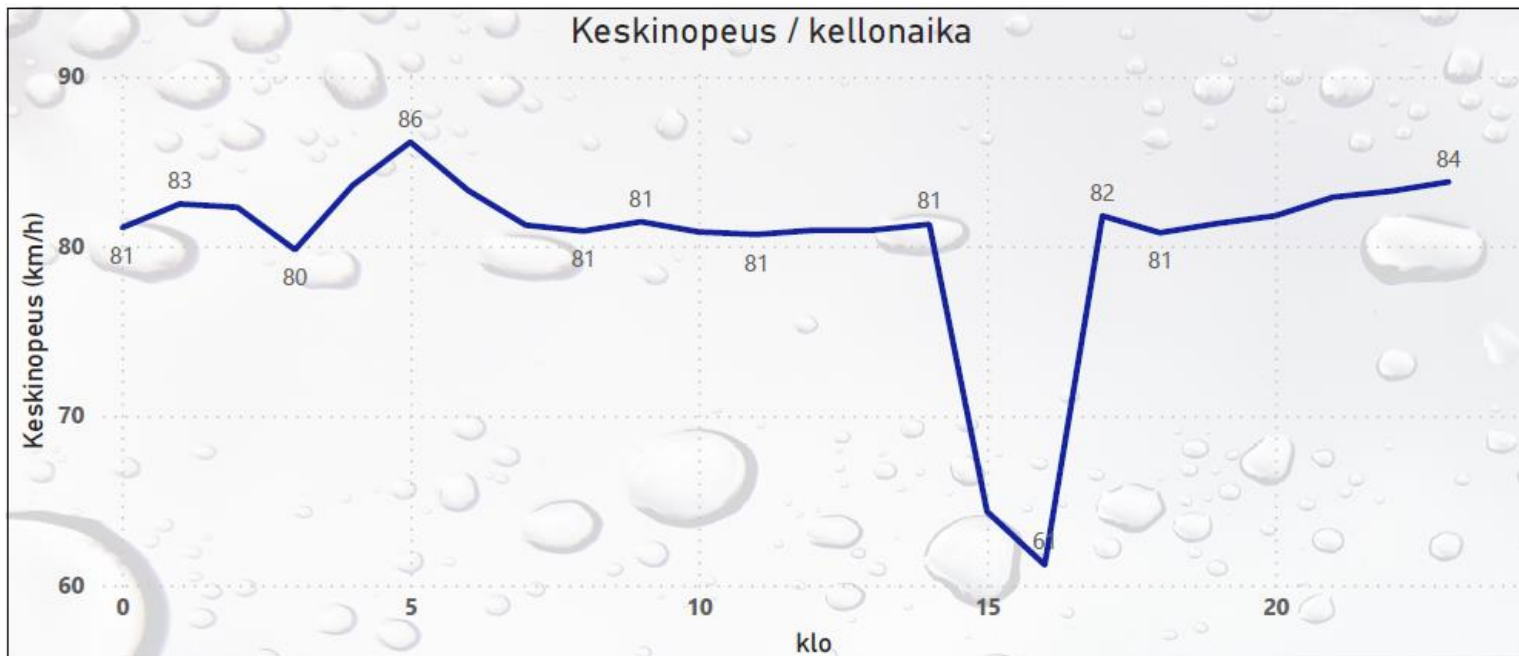


Keskinopeus / viikompäivä



Keskinopeus / kelloaika





Mittauspiste

kt50_Pakkala

st101_Espoo_Keilaniemi

vt4_Viikinmäki

Vuosi

2020

2021

2022

Kuukausi

Jan

Feb

Mar

Apr

May

Jun

Päivä

1

7

2

8

3

9

4

10

5

11

6

12

keskiviikko

Sää

8.4

Average of ilman_lamptila

23040.08

Average of nakyyvyys

6.88

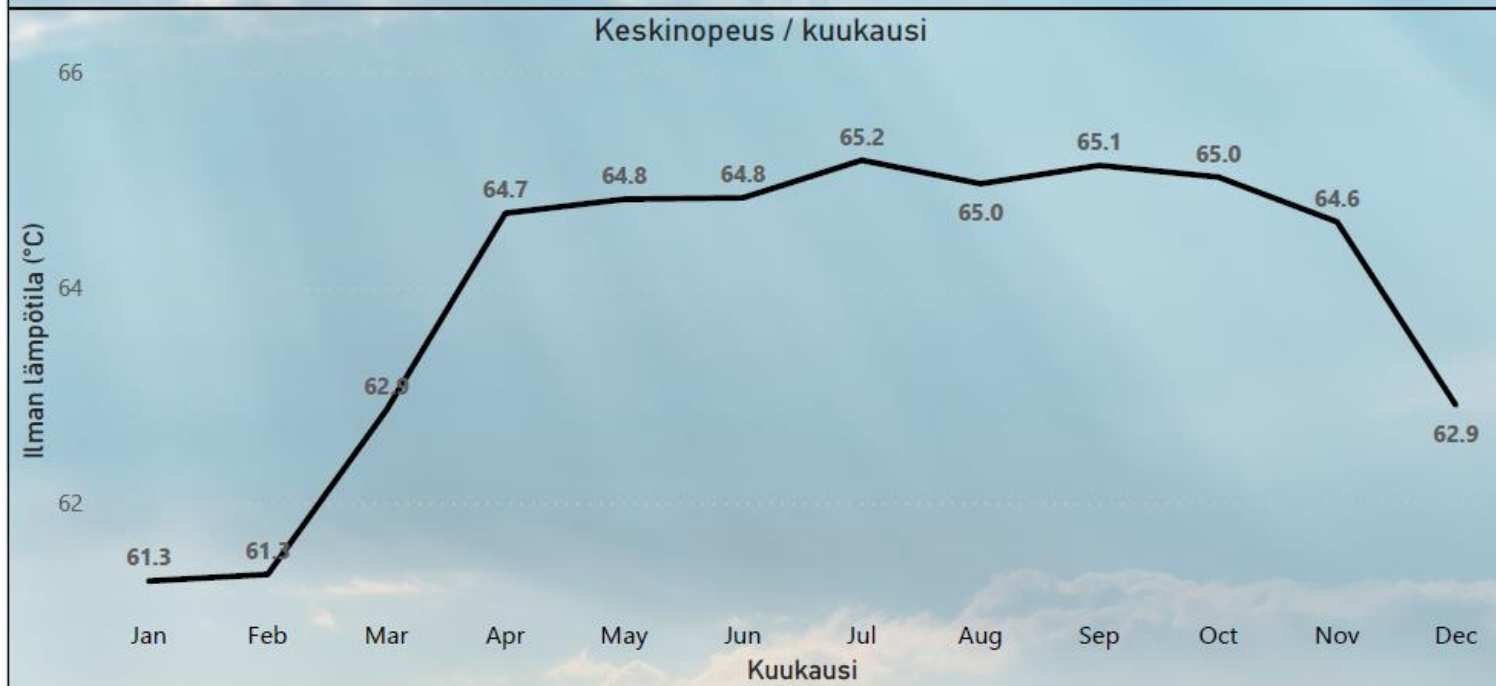
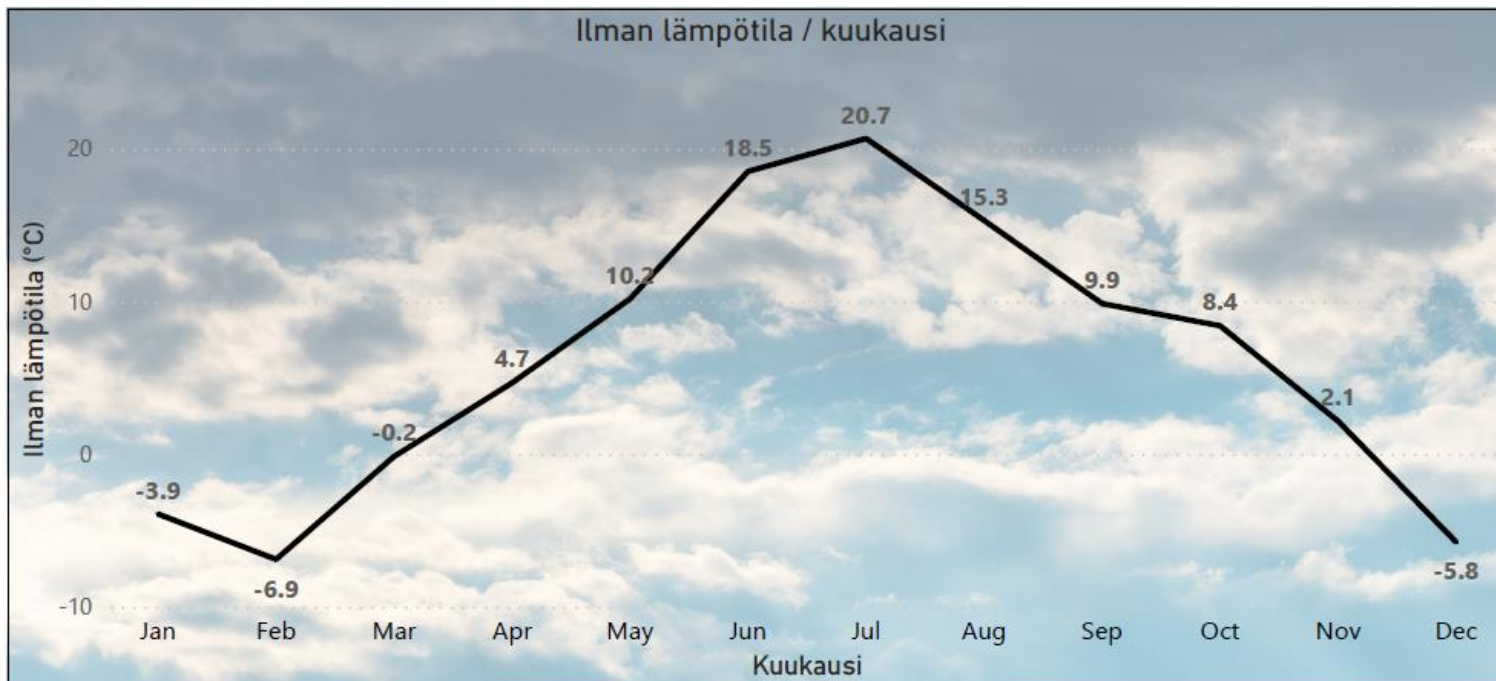
Average of tuulen_nopeus

0.00

Average of lumen_syvyys

Liikenteen määrä

49K



Mittauspiste

kt50_Pakkala

st101_Espoo_Kei...

vt4_Viikinmäki

Vuosi

2020

2021

2022

2023

Ilman
keskilämpötila (°C)

6.2

ENNUSTEMALLI

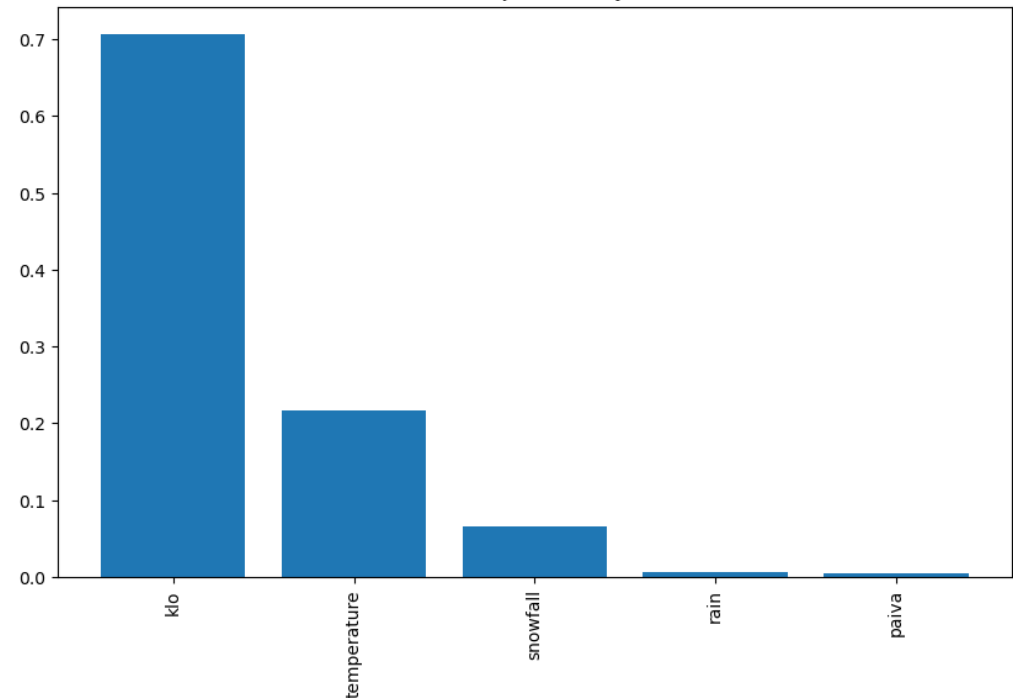
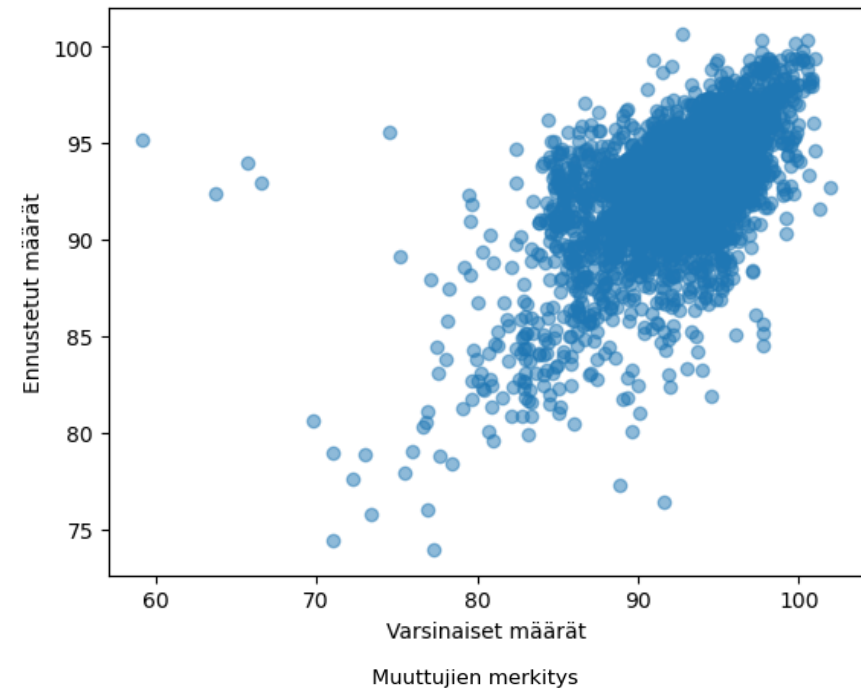
Mallinna käytettiin
RandomForestRegressiota

Mallia kokeiltiin
käyttämällä Viikinmäen
pisteen dataa

Muuttujina käytettiin
Sateen määrää, lämpötilaa,
viikompäivää ja kellonaikaa

Mallilla ennustettiin
liikennemäärää

Mallin keskimääräinen
virhe marginaali
lukumäärien suhteen 15%



PROJEKTIN HAASTEET JA KEHITYSIDEAT

- Azuren hallinta
- Liikennetiedotteiden kerääminen
- Reaali datan haku pipelineen
 - Data päivittyy 1-2h myöhässä

YHTEENVETO

- Suurimmat nopeudet klo: 5 - 6
- Suurimmat liikennemäärät klo: 15-16
- Talvi vaikuttaa nopeuksiin
- Ennustemalli on toimiva
- Lähes saman tuloksen saa kun katsoo ulos ikkunasta

KIITOS
MIELENKIINNOSTA!

Kysymyksiä?