

# Avoimen datan hyödyntäminen kokeilukiihdyttämössä

Datan avaamisen perusteet 24.5.2023

Saska Lohi  
Kaupunkiympäristön toimiala

Helsinki

# Lyhyt esitys siitä mitä

- Kokeilukiihdyttämö tarkoittaa
- Miksi avoin data on yleensä hyvä juttu kokeiluissa
- Case-esittelyt

# Kokeilukiihdyttämö 1/2

- Kanslian vetämä palvelu jo vuodesta 2019
- Vuosittain ollut 1-2 kokeilukampanjaa, joissa on jonkinlainen aiheeteema. Ensimmäisessä oli esimerkiksi tekoäly, sillä erilaiset tekoälyratkaisut olivat pinnalla ja kaupunki halusi oppia niistä
- Kokeilujen tarkoituksena onkin hankkia **oppeja**
  
- Kympin edustajana kokeilukiihdyttämön kehitysryhmässä ollut Saska
- Kampanja kampanjalta toimintaa on kehitetty parempaan suuntaan
  
- Kymp veti paikkatietoteemaisen kampanjan 2021, vetojuhtana Ville Jussila



# Kokeilukiihdyttämö 2/2

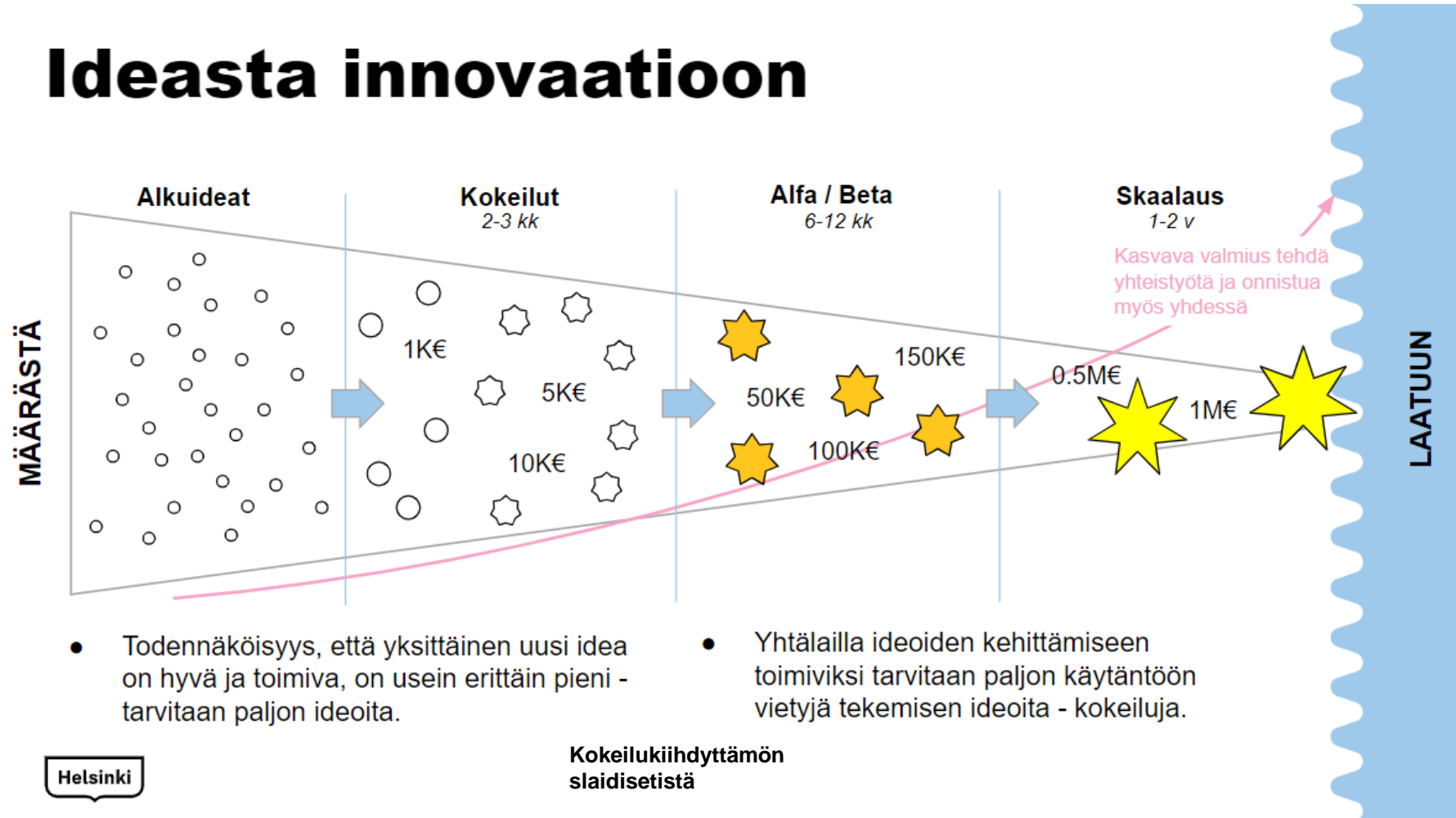
- Konseptia kehitetty ja se onkin jo melko vakiintunut, kampanjat rullaavat samankaltaisella rakenteella
- Kokeilujen budjetit 5000-10 000 euroa per kokeilu, rahoitusta per kampanja ollut yleensä n. 100k euroa
- Kokeilut toteuttanut yrityskumppani per kokeilu (tai useamman yhdistelmä)
- Kokeilut aina ajallisesti n. 3 kk aktiivista tekemistä, valmisteluun ym. menee pidempään
  - Esim. hakuaika huhti-toukokuussa,
  - päätökset ja hankinnat kesäkuussa
  - Itse kokeilun tekeminen syksyllä, valmista jouluksi



# Taustaa

- Kuvattu ja selkeä prosessi ”ideasta innovaatioon”

## Ideasta innovaatioon



# Avoim data kokeiluissa

Helsinki

# Miksi avoin data on monesti näppärää kokeilutoiminnassa?

- Monesti helppo saatavuus, ei tarvitse erikseen irrottaa jostain järjestelmästä
- Ei yleensä maksa mitään
- Ei yleensä tietosuojan alaista → ei mene aikaa vaikutustenarviointiin
- Yleensä dokumentoitu ainakin jollain tasolla
- Voidaan hyödyntää monia eri datalähteitä ja luoda yhdistelmiä
- Tulokset voidaan myös avoimesti jakaa

# Haasteita voi myös olla

- Jos dataa onkin valtavasti, niin sen käsittelyyn menee todella paljon aikaa ja vaivaa, pitää kuratoida
- Datalähteiden valinta voi viedä aikaa, jos mahdollisuuksia on paljon
- Dokumentaatio voi olla puutteellista
- Rajapinnat eivät aina välttämättä toimi kovin hyvin (Helsingin kaupunki tässä monesti poikkeus, meillä on hyvä tilanne!)
- Jos käytetään esim. tekoälyn kouluttamiseen että vältetään tietosuojaongelmat, mutta jos avoimena julkaistu data onkin erilaista verrattuna siihen dataan mihin kehitettyä ratkaisua tulisi soveltaa, niin se ei välttämättä vastaakaan käyttötarkoitukseen





### Robotti tuottamaan analyysia paikkatietoaineistoista

Kokeilussa tavoitteena oli rakentaa Helsingin paikkatietopalvelulle automaattisen palvelun konsepti ja kehitysmalli, jossa käytetään chatbot-tyyppistä käyttöliittymää ja yksinkertaista itsekäyttöpalvelua.

Saska Lohi Teemu Kiiveri

Päättynyt Digitalisaatio Chatbot Automaatio Paikkatieto

Näytä kokeilu



### Sovellus allekirjoitusten tunnistamiseen PDF-asiakirjoista

Tietosuojasyistä Helsingin kaupunki ei julkaise internetissä asiakirjoja, joissa esiintyy henkilöiden allekirjoituksia tai muita henkilötietoja. Tiedostojen tarkistaminen on aikaa vievää työtä. Tavoitteena oli kehittää työkalu, joka automaattisesti tunnistaa PDF-asiakirjoista allekirjoitukset.

Niko Latvakoski

Päättynyt Tekoäly Konenäkö

Näytä kokeilu



### Tekoälyavusteinen kävelykierrosopas

Varsinkin koronapandemian aikaan lähiliikunnan merkitys virkistäytymiseen ja ulkoiluun on kasvanut huomattavasti. Kokeilussa luotiin virtuaalisen matkaoppaan web-palvelun prototyyppi, joka osaa tekoälyä hyödyntäen luoda reitin erilaisten käyttäjän valitsemien kohdeteemojen, halutun keston ja alku- ja loppupisteen perusteella.

Saska Lohi

Päättynyt Liikkuminen Tekoäly Kävelykierrokset

Näytä kokeilu



### Suojateiden tunnistaminen ilmakuvista koneoppimisen avulla

Kokeilussa selvitettiin mahdollisuutta luoda tekoälypohjainen työkalu, joka tunnistaa, käsittelee ja luokittelee tiemerkintöjä kaupungin avoimista ilmakuvista. Suojateiden tunnistaminen valikoitui kokeilun tiemerkintädataksi.

Saska Lohi

Päättynyt Tekoäly Konenäkö Liikenne Suojatie Älyliikenne

Näytä kokeilu

Helsinki

# Tekoälykokeilun loppuraportti

Suojateiden tunnistaminen ilmakuvista  
koneoppimisen avulla

Helsinki



Kokeilukiihdyttämö

Tukea digitalisaatiokokeiluihin  
kaupungin työntekijöille

# Suojateiden tunnistaminen ilmakuvista koneoppimisen avulla

Kaupunkiympäristön toimiala: Fanny Taxell, Saska Lohi,  
Virpi Vertainen, Minna Leinonen, Esko Laiho

Integrify: Katja Lairikko, Alex Jung

Helsinki

Helsinki

# Kokeilun taustaa

- Kaupungin aineisto suojateistä oli vanhentunut, uusin versio siitä oli vuodelta 2014
- Haluttiin tutkia voisiko ilmakuvista koneoppimisen perusteella louhia suojatiet ja laatia päivitetyn datasetin

## 2. Kokeilun eteneminen

HANKINTA

1. Datan kerääminen

2. Koneoppimisen mallin luominen

3. Mallin arviointi ja kehitys

AINEISTO VISUALISOINTI

- 1. Datan kerääminen.** Kartta.hel.fi kaupungin karttapalvelusta kerättiin ohjelmallisesti n. 3000 ilmakuvaa Helsingin alueelta. Mallin tarkkuuden parantamiseksi olisi ollut parempi jos saatavilla olisi ollut enemmän dataa. Tilannetta auttoi se, että pystyttiin hyödyntämään myös Espoon ja Vantaan kaupungin ilmakuvia. Mitä enemmän ja laadukkaampaa dataa on saatavilla, sitä parempi malli saadaan aikaiseksi.

**2. Koneoppimisen mallin luonti:** Data jaettiin opetusdataan ja testidataan. Opetusdataan eli jokaiseen ilmakuvaan merkittiin manuaalisesti suojateiden paikka.

Malli luotiin hyödyntämällä opetusdataa suojateiden tunnistamiseen eri koneoppimisen arkkitehtuurien avulla joita hienosäädettiin eri parametrien avulla. Näitä olivat RetinaNet, faste R-CNN ja Mask R-CNN. Parhaan tarkkuuden tuotti RetinaNet.

### 3. Mallin arviointi ja kehitys:

Testidatan avulla mallin toimivuus validoitiin. Malli pystyy tunnistamaan suojatien noin 70-80% tarkkuudella. Tätä tarkkuutta voidaan parantaa opetusdatan määrän lisäämisellä.

# 5. Opit tekoälyn kehittämisestä

- Tekoäly toimii enemmänkin ”tukiälynä” – Ihminen on korvaamaton
- Ihmisen rooli korostuu etenkin datan muokkaamisessa ja validoinnissa. Ihminen pystyy opettamaan mallin toimintaa ja ohjaamaan oikeaan suuntaan.

”Paras malli on yhtä hyvä kuin käytetty opetusdata”

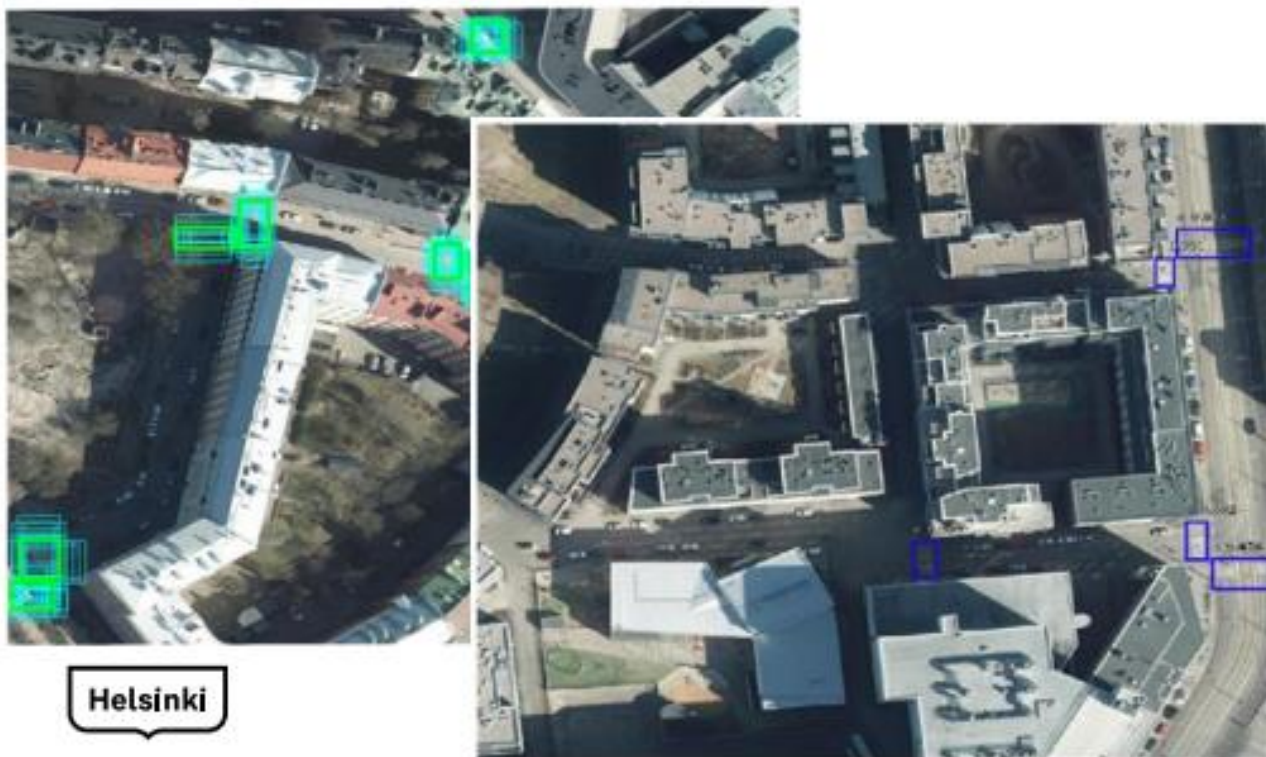
”60 % projektin ajasta kannattaa käyttää datan esitarkistukseen”  
– Työ, tekoäly ja ihminen kurssi

# 3. Kokeilun tuotokset

Kokeilun lopputuoteoksena syntyi koneoppimisen malli jonka avulla pystytään tunnistamaan suojatiet ilmakuvista ja listaamaan niiden koordinaatit.

1. Tunnistetaan ilmakuvista todennäköiset suojatiet ja annetaan niille todennäköisyysarvot sekä määritellään tunnistuskehykset

RetinaNet



Mask R-CNN



# Avoim data tässä kokeilussa

- Auttoi pääsemään nopeasti alkuun, muissa kokeiluissa kesti pitkään datan saamisen kanssa
- Saatiin paljon dataa mallin kouluttamista varten
- Pystyttiin testaamaan mallin toimintaa, kun testidataa oli saatavilla opetusdatan lisäksi
- Tuotokset julkaistiin avoimesti githubissa!
- Kokeilun pohjalta syntyi yksi gradu Aalto-yliopistossa, siinä vietiin tunnistaminen pidemmälle suojateiden kuntoluokitukseen asti



# Tekoäly- ja ohjelmistorobotiikkakokeilun loppuraportti - syksy 2020

Sovellus allekirjoitusten tunnistamiseen PDF-asiakirjoista

Helsinki



Tukea digitalisaatiokokeiluihin kaupungin työntekijöille

# Sovellus allekirjoitusten tunnistamiseen PDF-asiakirjoista

**Niko Latvakoski** (Kymp / Maka / Aska)  
**Aleksander Alafuzoff** (Siili Solutions Oy)  
**Susa Eräranta** (Kymp / Maka / Aska)  
**Katariina Hirvonen** (Kymp / Maka / Aska)  
**Juuso Ala-Outinen** (Kymp / Maka / Aska)  
**Saska Lohi** (Kymp / Hatu / Kepa)

Tietosuoja-asioiden osalta myös:

**Jussi Vaanola** (Kymp / Hatu)  
**Tetti Kunnas** (Kymp / Hatu)

Helsinki

Helsinki

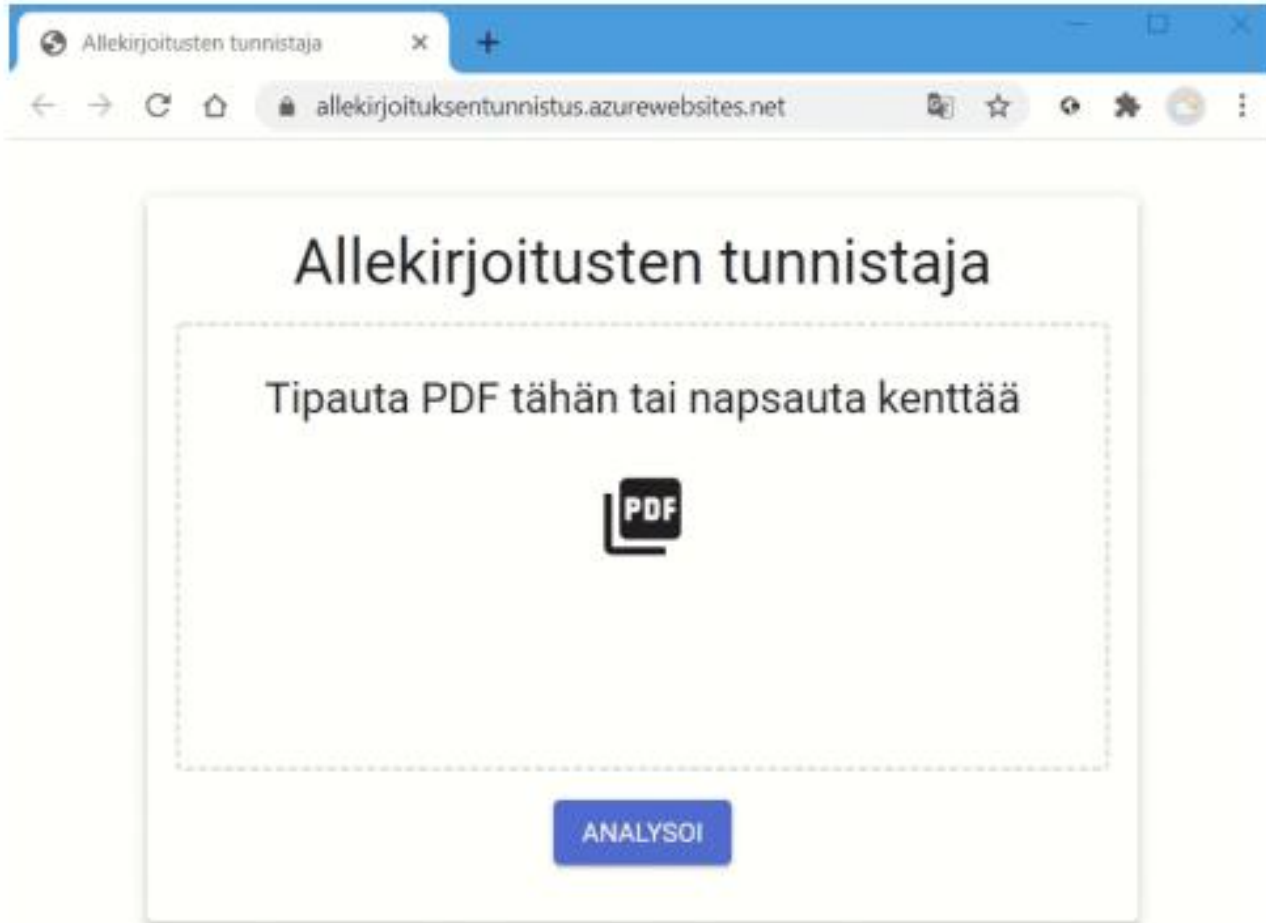
# Kokeilun tarve

- Kaupunki julkaisee avoimena datana paljon pdf-muotoisia asiakirjoja (tässä tapauksessa kaavadokumentteja), jotka sisältävät allekirjoituksia, jotka luetaan henkilötiedoksi
- Allekirjoituksia ei voi julkaista
- Tällä hetkellä dokumentteja kahlataan käsin ennen julkaisua, ja allekirjoitukset poistetaan
- → voisiko tämän tehdä automaattisesti?

# 1. Kokeilun onnistuminen

- Tavoitteena oli kehittää työkalu, joka automaattisesti tunnistaa PDF-asiakirjoista allekirjoitukset. Lisäksi haluttiin saada selville, kuinka kuvantunnistukseen perustuva tekoäly käytännössä toimii, ja millaisia seikkoja tekoälyn opettamisessa tulee ottaa huomioon. Lisäksi tavoitteena oli oppia ohjelmistokehityksestä ja projektinhallinnasta.
- Tavoitteet saavutettiin erinomaisesti. Tuloksena oli toivotun kaltainen sovellus, jota on mahdollista kehittää ja laajentaa varsinaisen kokeiluhankkeen päätyttyäkin. Kokeilussa opittiin paljon syvien neuroverkkojen toiminnasta, opettamisesta ja testaamisesta.

# 3. Kokeilun tuotokset



# 9. Kokeilun data

- Aineistona käytettiin julkisia, kaupungin internetsivuilta ladattuja asiakirjoja, jotka olivat allekirjoituksen tunnistamisen kannalta edustavia (allekirjoitukset oli pyyhitty pois).
- Allekirjoitusaineisto ladattiin Wikimediasta ja sopivat allekirjoitukset validoitiin (viereinen kuva).
- Sopivia neuroverkkomalleja opetettiin Jupyter Notebooks-ohjelmointityökalulla opetusaineiston pohjalta ja suoritusta arvioitiin validointiaineistolla. Sovellusta on testattu monenlaisilla testitiedostoilla.



# Avoim data tässä kokeilussa

- Auttoi pääsemään nopeasti alkuun, kun pystyimme hyödyntämään dataa joka ei ole tietosuojaan alaista, eikä sen käyttö vaatinut selvittelyä
- Yhdisteltiin eri avoimia datalähteitä:
  - Wikimediasta kuolleiden henkilöiden allekirjoituksia
  - kaupungin aineistoista dokumentit joihin allekirjoitukset lisättiin
  - Yhdistelmäaineistolla voitiin kouluttaa tekoälyä hienosti
- Malli ja sovellus saatiin toimimaan hienosti! Ei kuitenkaan ole käytössä, koska kaupungilta ei löytynyt tällaiselle palvelulle omistajatahoa

# Tekoäly- ja ohjelmistorobotiikkakokeilun loppuraportti - syksy 2021

Tekoälyavusteinen kävelykierrosopas

Helsinki



Tukea digitalisaatiokokeiluihin kaupungin työntekijöille



# Tekoälyavusteinen kävelykierrosopas - HelsWalks

## Tiimi

Saska Lohi (Kymp)  
Laura Kauria (Kymp)  
Sari Saresto (Kaupunginmuseo)  
Jaani Lahtinen (Gispositio Oy)  
Tero Pietilä (Gispositio Oy)  
Ari Mutanen (Altoros Oy)

Helsinki

Helsinki

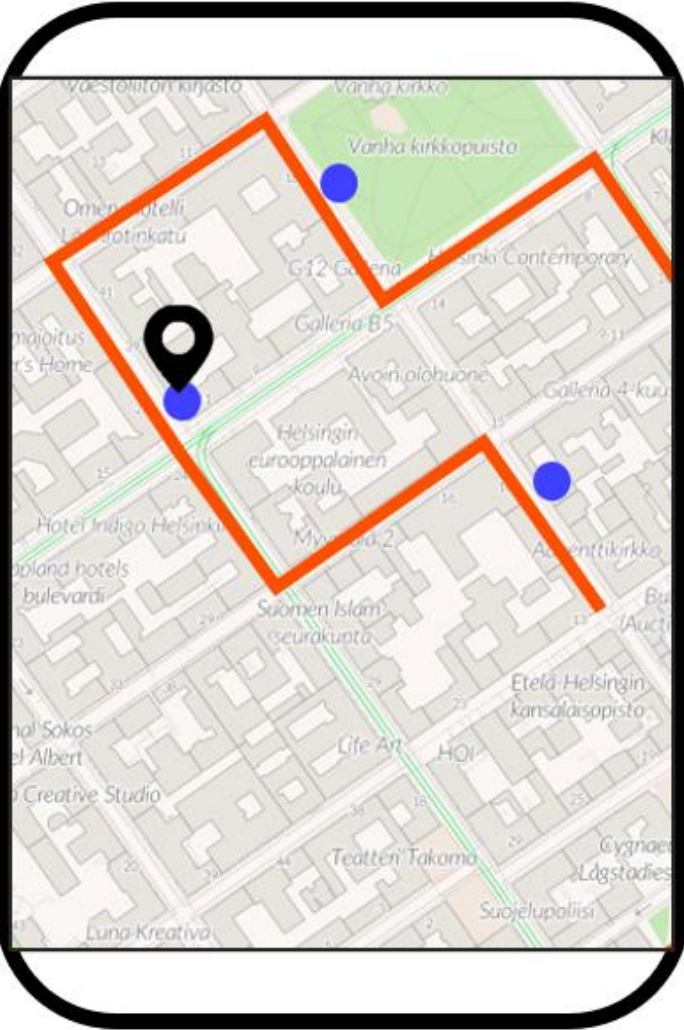
# Kokeilun tavoite ja eteneminen

- Tavoitteena oli luoda ”virtuaalinen kävelykierrosopas” joka johdattaisi käyttäjän mielenkiintoiselle kävelyreitille, jonka varrelle sijoittuu kohteita joista saa myös lisätietoa
- Lisäksi tavoitteena oli tutkia sitä miten erilaisia aineistoja pystytään yhdistelemään uudenvälisellä tavalla ja hyödyntämään niiden luomaa kokonaisuutta mielekkäiden teemojen luomisessa
- Virtuaalisen kävelykierrosoppaan hyötyjä:
  - Kierroksen voi tehdä milloin vain, ei sidottu oppaan aikatauluihin
  - Ryhmään mahtuu aina! Ja korona-aikana voi tehdä yksin kävelyn
  - Ei vaadi reittien ennakkokuratoointia, vaan reitit muodostuvat lennosta
    - Reitit myös sijoittuvat ympäri kaupunkia sen mukaan missä käyttäjä on!
  - Mahdollistaa monipuolisia kustomoituja teemoja eri käyttäjille ja mielenkiinnonkohteille

# Mistä lähdettiin

# Swaippaa

# Lisätietoja



Näytä ehdotus



## Lisätietoja



# Kävelykierrosoppaan kehittäminen

Tekninen alusta ja  
datalähteet:

Sijaintitietoa ja yhteyksiä tarjoavat palvelut

Yhteyksiä tarjoavat palvelut



Virtuaalipalvelin: Azure. Tietokanta & Sovelluspalvelin (Windows)

Virtuaalipalvelin Azure, WWW-palvelut (Windows)

Apache Tomcat

Java

Geoserver

Tekoälysovellus

R

Apache WWW

Javascript

React

OpenLayers

PHP

Postgres

PSQL

Postgis

PGRouting

Postgres tietokanta

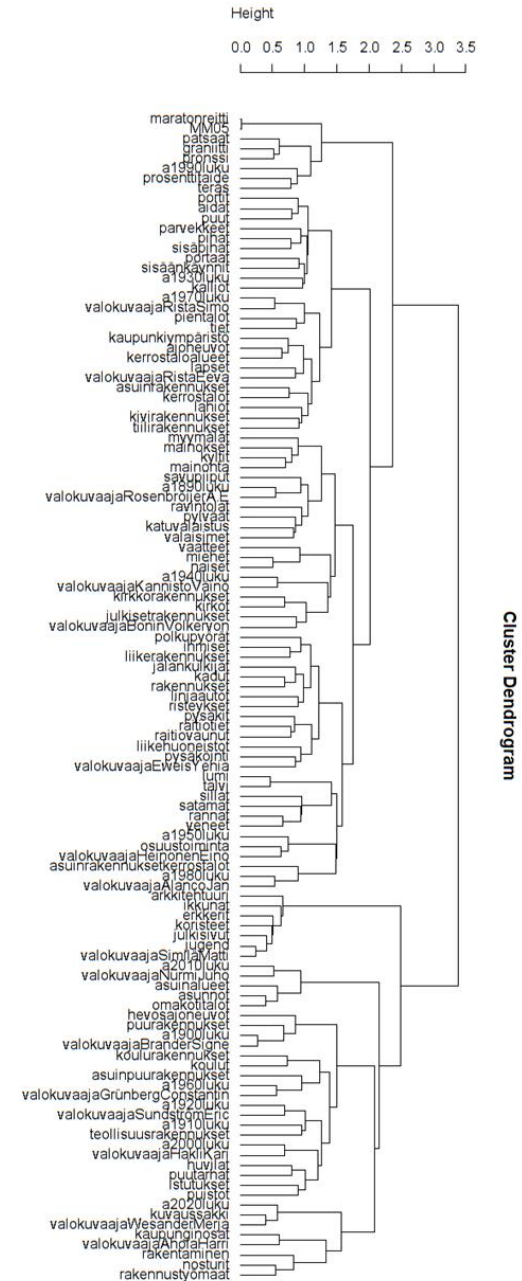
Lähteitä jouduttiin karsimaan, dataa on paljon!

Helsinki

Alustavissa selvityksissä projektille tarjolla useita laadukkaita aineistolähteitä, joista valitaan soveltuvin kokonaisuus, ei siis kaikkia tässä kuvattuja.

# Kävelykierrosoppaan kehittäminen

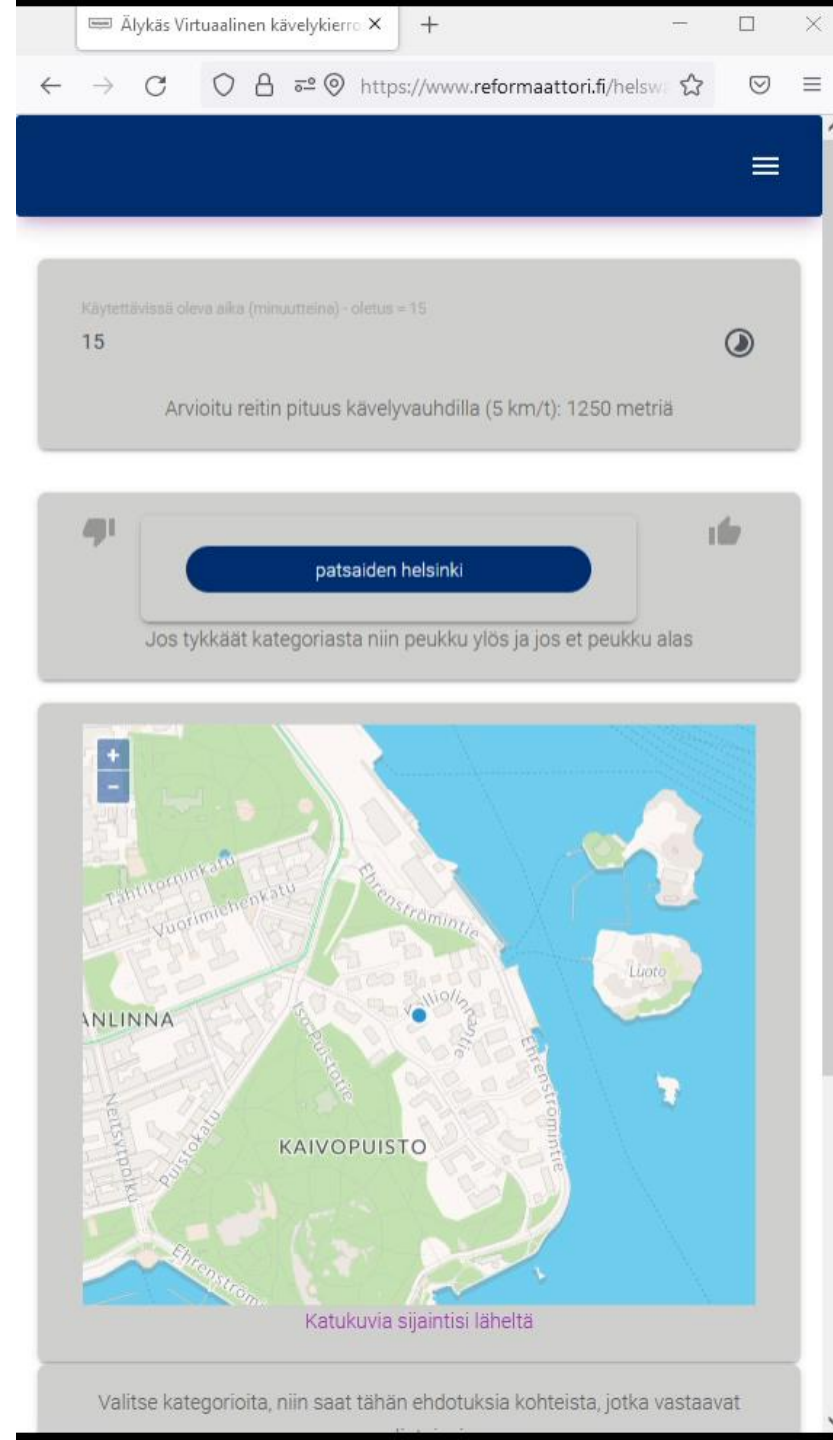
- Aloitettiin veistosdatalla (löytyy esim. palvelukartta.hel.fi -> veistokset ja julkinen taide)
- Lisäksi rakennusdataa kaupungin rajapinnoista
- Yhdistettiin rakennusdata finna-tietokannasta löytyviin aineistoihin rakennustunnusten ja osoitteiden perusteella
- Pyrittiin yhdistämään datasetit avainsanojen perusteella; osoittautui mahdottomaksi koska avainsanoja/luokitteluja todella paljon erilaisia



# Tuotokset

- Luotiin valmis prototyyppi, joka toteuttaa idean virtuaalisesta kävelykierrosoppaasta
- Data vaatisi kuitenkin vielä jonkin verran työstöä ja kuratointia, nyt mukaan tulee sangen mielenkiintoisiakin kohteita
- Rakennuksia on valtava määrä ja ne dominoivat mukaan päätyviä kohteita -> vaatisi jonkin verran painotuksia että mikä tulee mukaan

Helsinki



# Avoim data tässä kokeilussa

- Tarvitsimme paljon kohteita koko Helsingin alueelta
- Löysimme monia eri datalähteitä, joita pystyimme yhdistelemään
  
- Ongelmana kuitenkin luokitusten heterogeenisuus, jouduimme luomaan omaa luokittelua
- Lisäksi kohteita tuli valtavasti → kävelykierros ei ole kiva jos koko ajan joutuu pysähtelemään

# BONUS: RAKKI-projekti

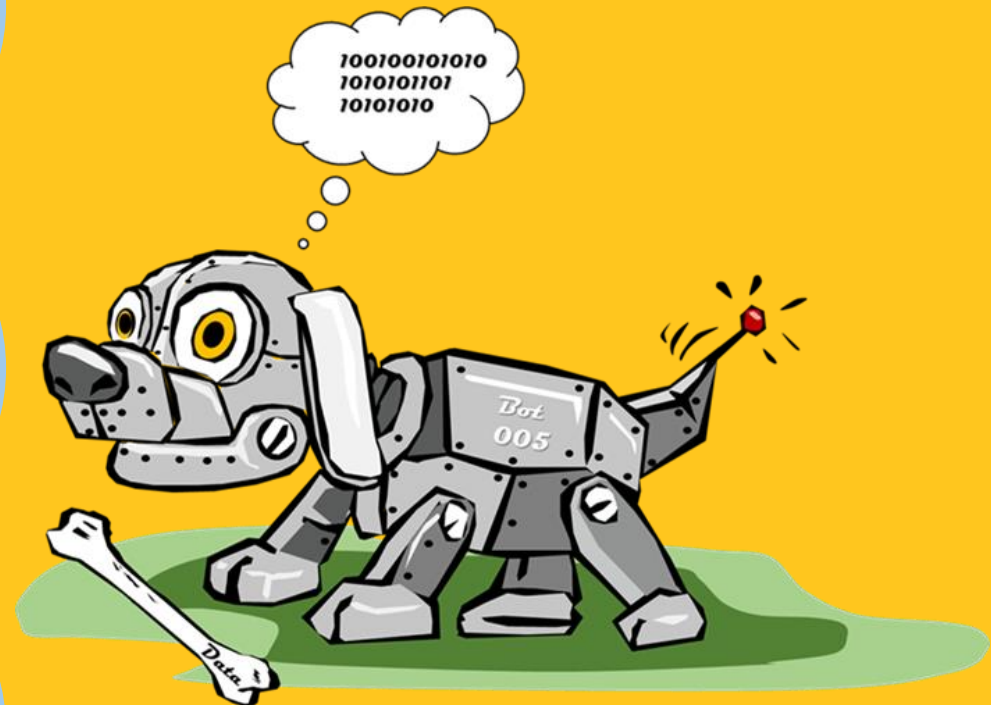
## Helsingin kaupungin tiimi:

KYMP: Teemu Kiiveri, Saska Lohi ja Timo lipponen

## Yrityskumppanit

IBM

Helsinki



### **RAKKI-PROJEKTI**

Robotti aineistojen koneellisen käsittelyn integrointiin

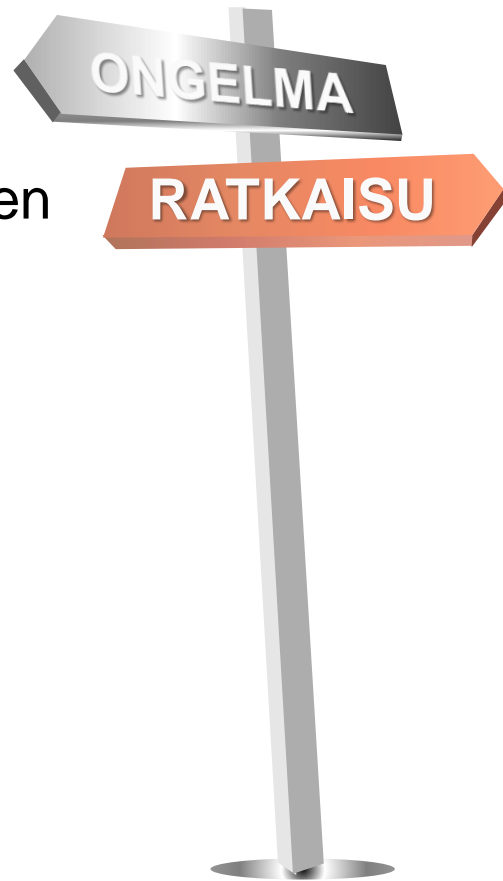


# Kokeilun tavoite ja eteneminen



## ASIAKKAAN ONGELMA

Paikkatietoanalyysien tekeminen on hidasta ja aikaa vievää, varsinkin jos niitä ei tee joka päivä eikä ole rutiinia. Halutun aineiston saaminen paikkatietosintuntijalta kestää kauan tai ei onnistu lainkaan heidän työkuormansa takia.



## RATKAISU ONGELMAAN

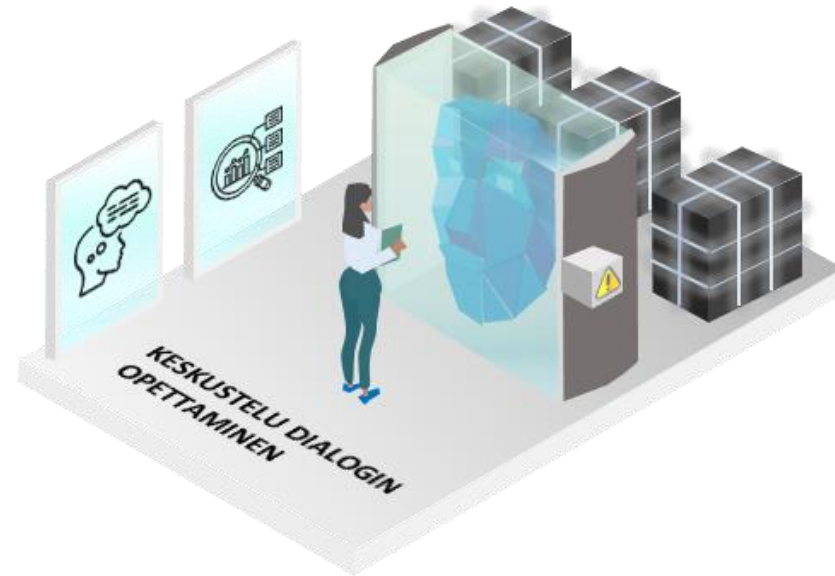


Tavoitteena oli rakentaa Helsingin paikkatietopalvelulle automaattisen palvelun konsepti ja kehitysmalli, jossa käytettiin chatbot tyyppistä käyttöliittymää ja yksinkertaista itsekäyttöpalvelua.

# Ratkaisun kehittäminen

## ESIMERKKIKYSYMYKSET JA ODOTETTU VASTAUS OLIVAT SEURAAVANLAISIA:

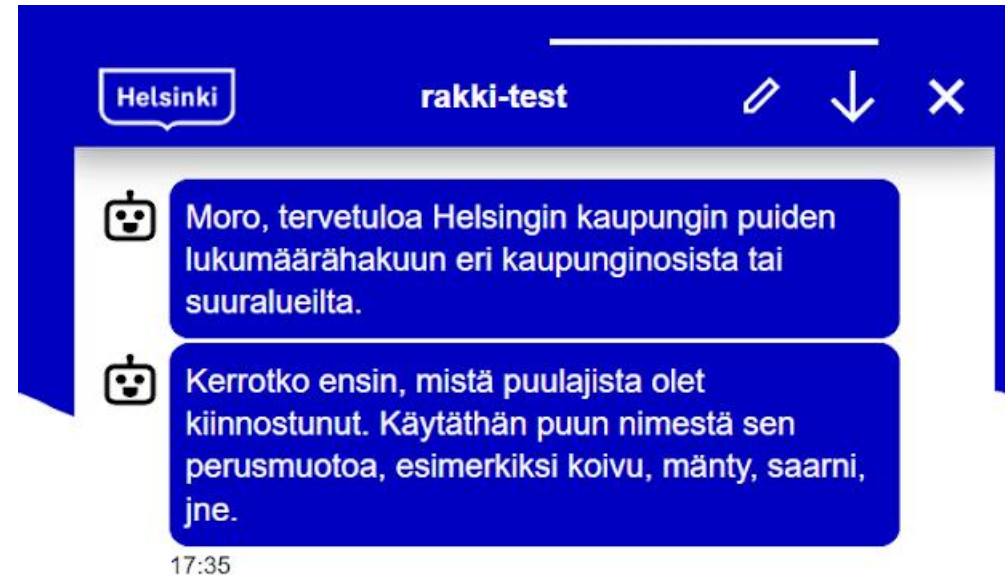
- Montako Lehmusta (Puu) on Eirassa?
- Vastauksena Chatbotissa esitetään suoraan puiden lukumäärän
- Chat botin tietopyyntö toteutetaan siten, että dialogi ohjaa ja kertoo mitä pitää tehdä ja mitä tietoja milloinkin tarvitaan. Samalla mallilla voidaan myöhemmin dialogia laajentaa ja käyttää useita pisteaineistoja, aluejakoja ja tehdä useampia rajoituksia.



# Tuotokset (Chatbot käyttöliittymä)

Projektin pääasiallisena tuotoksena on Proof of Concept (PoC) -tasoinen Demobotti, joka toimii Helsingin testiympäristössä.

- Demobotilla voidaan kokeilla Chatbot-alustan ja -käyttöliittymän toimivuutta tavoitteisiin nähden ja todeta kannattaako kehitystyötä jatkaa.



# Tuotokset (Itsekäyttöpalvelu)

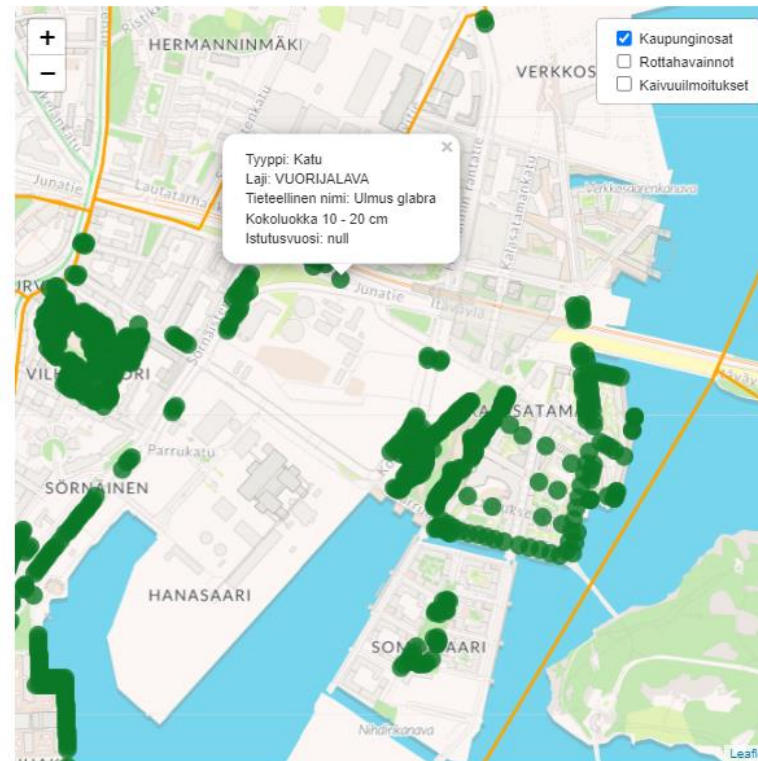
## Proof of Concept (PoC) - tasoinen itsepalvelukäyttö- liittymä, joka toimii Helsingin sisäverkon testiympäristössä.

- Demopalvelulla testattiin miten katu- ja puistopuiden esiintymistä eri kaupunginosissa voitaisiin havainnollistaa käyttäjälle. Lisäksi tehtiin kokeeksi top5-listaukset yleisimmistä puulajeista sekä yleisimmistä istutusvuosista
- Karttakäyttöliittymä tarjoaa visuaalisemman käyttökokemuksen chatbottiin verrattuna, mutta käyttäjää ei juurikaan opasteta palvelun käyttöön.



### Hae kaupunginosan puustotiedot

Valitse tasovalinnasta (oikea yläkulma) mitä näytetään kartalla taustakartan lisäksi.



Kaupunginosassa SÖRNÄINEN on puita puurekisterissä yhteensä 1047 kappaletta.

Niistä katupuita on 485 ja puistopuita 562 kpl.

#### Kaupunginosan top 5 puistopuulajit:

METSÄMÄNTY: 170 kpl  
VUORIJALAVA: 85 kpl  
METSÄVAAHTERA: 57 kpl  
TERVALEPPÄ: 53 kpl  
PUISTOLEHMUS: 39 kpl

#### Kaupunginosan top 5 katupuulajit:

PUISTOLEHMUS: 350 kpl  
LEHTIKUUSI: 25 kpl  
VUORIJALAVA: 17 kpl  
KEISARINLEHMUS: 16 kpl  
VAAHTERA: 15 kpl

#### Kaupunginosan puiden top 5 istutusvuodet

null: 680 kpl  
2020: 213 kpl  
2016: 100 kpl  
2022: 26 kpl  
2017: 25 kpl

Klikkaa nappia.

Valitse kaupunginosa: 10 SÖRNÄINEN

Hae

# Avoim data tässä kokeilussa

- Saimme datan nopeasti ja helposti käyttöön
- Kokeilussa luotiin uudenlainen käyttöliittymä dataan

# Kiitos