

---

# **Agent-based modelleren van het openbaar vervoer om veranderingen in passagiersstromen te identificeren**

**Sophie Ensing  
Thesis MSc Data Science  
20/06/2019**

---

---

# Introductie

- **Veranderingen in het openbaar vervoer**
- **Impact drukte**
- **Simulatie van boten door de grachten<sup>[1]</sup>**

<sup>[1]</sup> Jeroen van der Does de Willebois. Assessing the impact of quay-wall renovations on the nautical traffic in Amsterdam. Master's thesis, TU Delft, 2019

---

**Hoe kan een agent model gebruikt worden om veranderingen in passagiersstromen in het openbaar vervoer te identificeren?**

**Hoe kan een agent model gebruikt worden om veranderingen in passagiersstromen in het openbaar vervoer te identificeren?**

---

# Agent-based modeling

- **Passagiers**
  - **Regels**
  - **Doelen**
- **Omgeving is het openbaar vervoer netwerk**

---

# Passagiers



**Groep 1:** voorkeur  
voor de kortste route



**Groep 2:** voorkeur voor  
minste overstappen



**Groep 3:** voorkeur  
voor kortste wachttijd

---

**Hoe kan een agent model gebruikt worden om veranderingen in passagiersstromen in het openbaar vervoer te identificeren?**

---

# Sub-netwerk

Tram 24 en 12 + Metro 50, 52 en 53

networkx



maps.amsterdam.nl

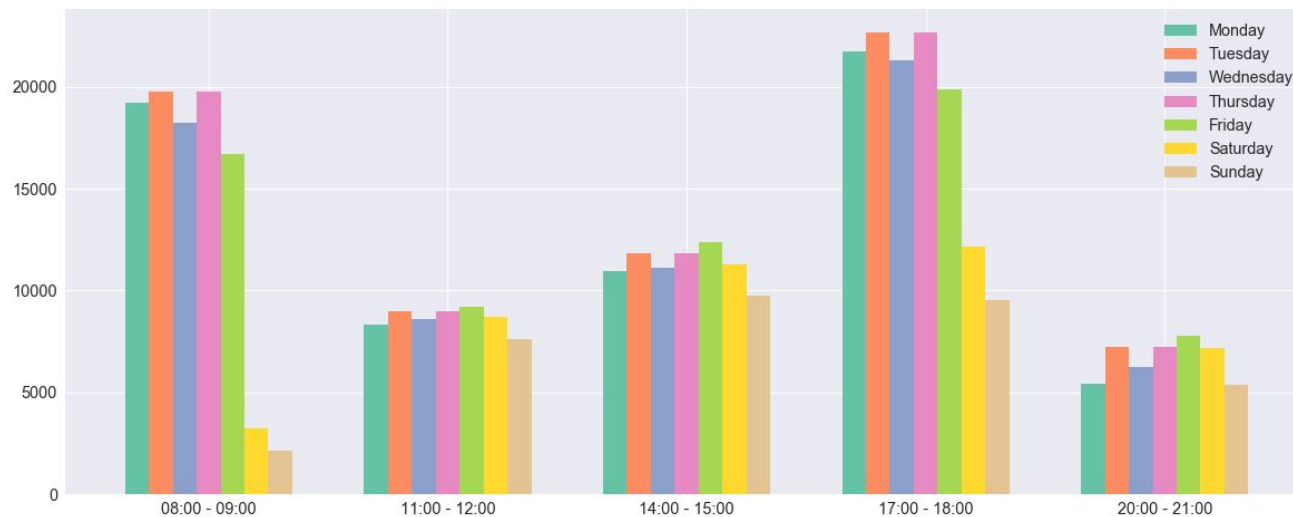




---

# GVB data

Gemiddelde aantal passagiers per uur, per weekdag\*



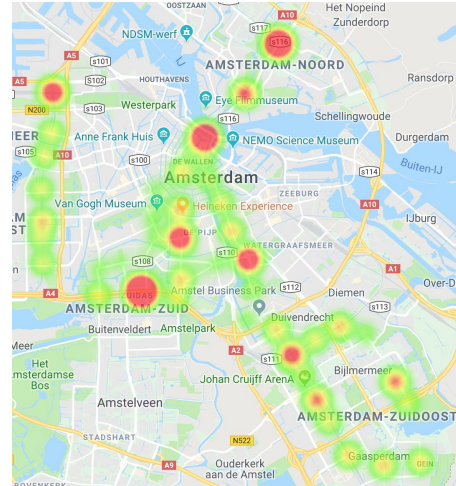
\* aantallen gecorrigeerd voor de gebruikte lijnen in het onderzoek

---

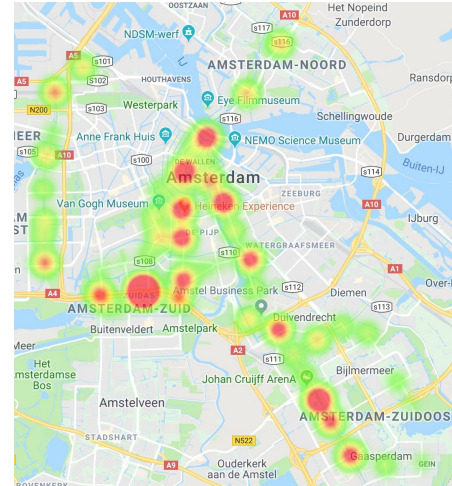
# GVB data

Drukke verdeling op een weekday, 08:00 - 09:00\*

Vertrekhalte



Aankomsthalte



\* aantallen gecorrigeerd voor de gebruikte lijnen in het onderzoek

---

---

# Passagiers

1. **Genereren passagier**
  2. **Toewijzen begin- en eindstation**
  3. **Alle routeopties berekenen**
  4. **Route kiezen op basis van regels**
-

**Hoe kan een agent model gebruikt worden om veranderingen in passagiersstromen in het openbaar vervoer te identificeren?**

---

# Veranderingen in het OV

## Scenario 1

- Al het openbaar vervoer rijdt volgens schema
- Genereren van passagiers op basis van kansverdelingen
- Weekdag/weekenddag
- 08:00 - 09:00, 11:00 - 12:00, 14:00 - 15:00, 17:00 - 18:00, 20:00 - 21:00

## Scenario 2

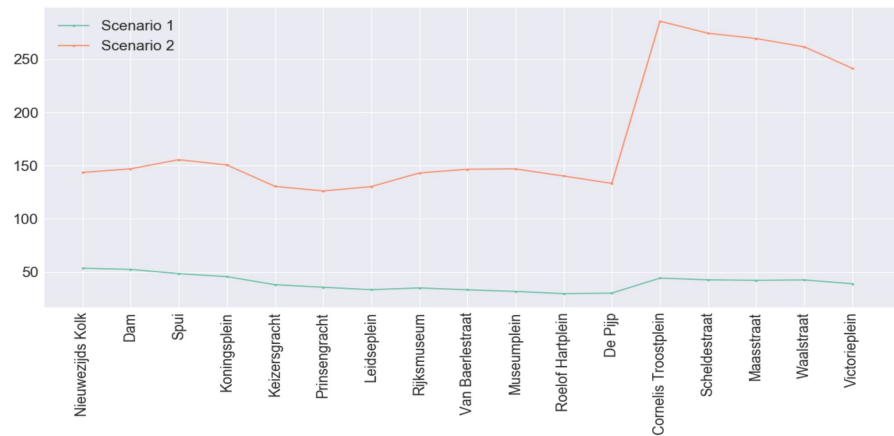
- Scenario 1, maar:
- Tijdens elk uur in de simulatie rijdt tram 12 het eerste half uur niet.

---

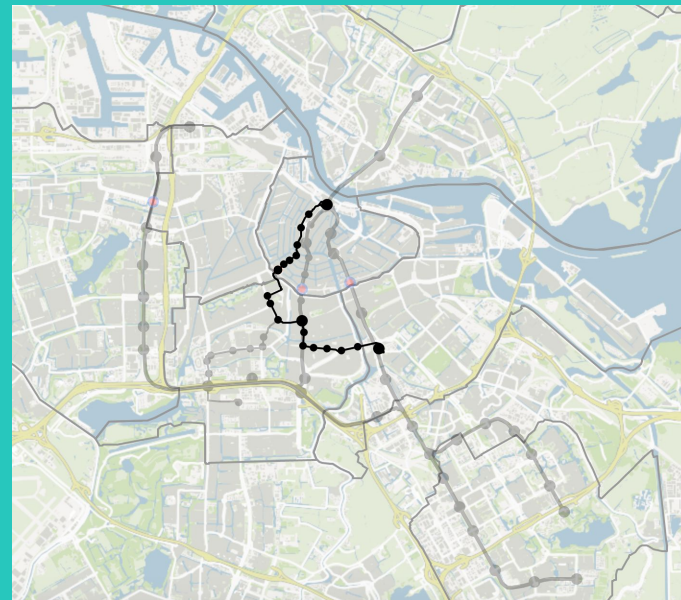
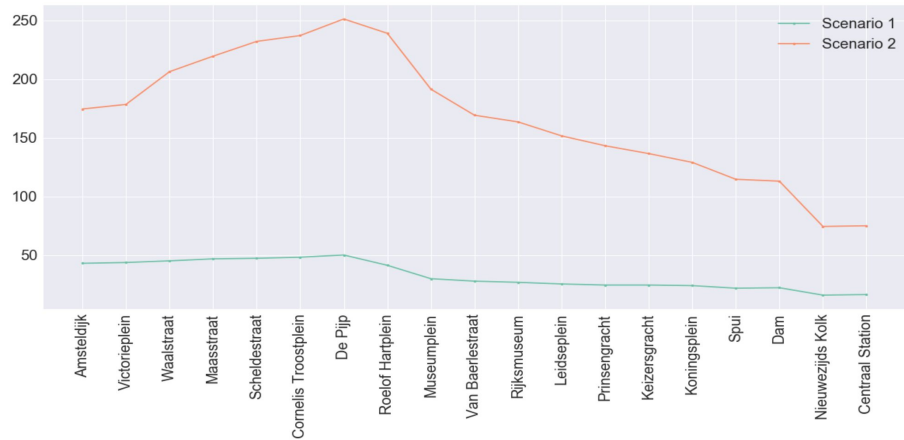
# Resultaten

- **Foutmarge (*Mean Average Error*) berekend voor alle stations en ritten.**
    - Foutmarge tussen 3 en 36 personen gemiddeld
  - **Foutmarge belangrijk als baseline**
  - **Wachttijd, aantal overstappen en ritduur**
    - Wachttijd veel meer outliers
    - Geen groot effect op overstappen en ritduur
  - **Bezetting van alle voertuigen**
    - Invloed op alle lijnen behalve metro 50
-

## Richting Amstelstation

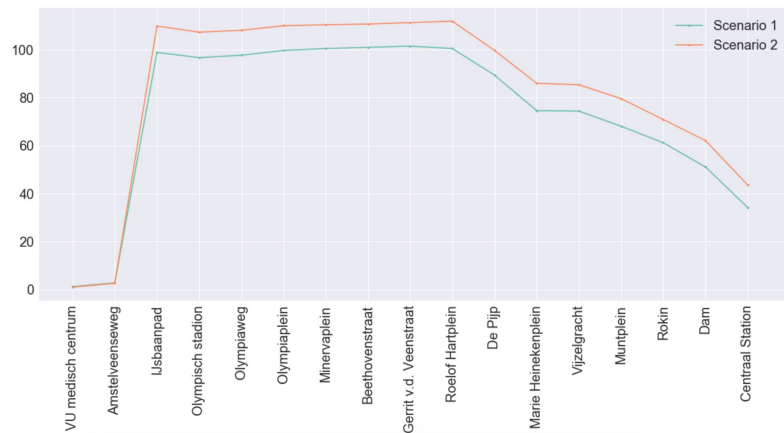


## Richting Centraal station

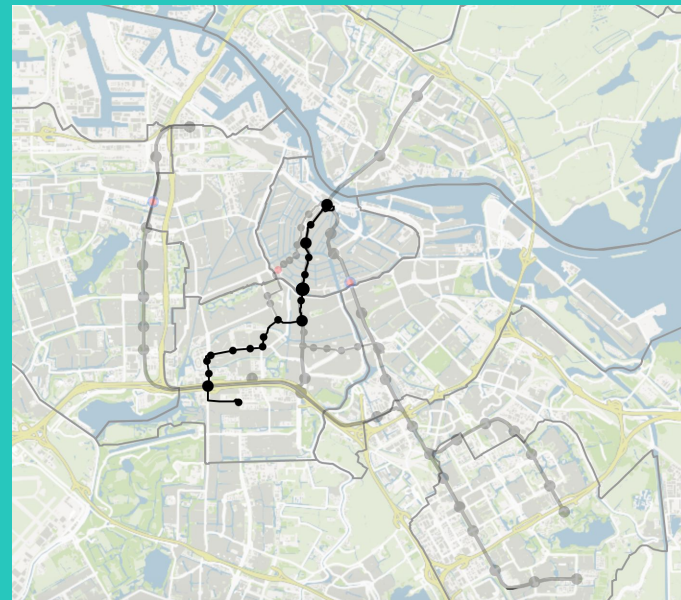
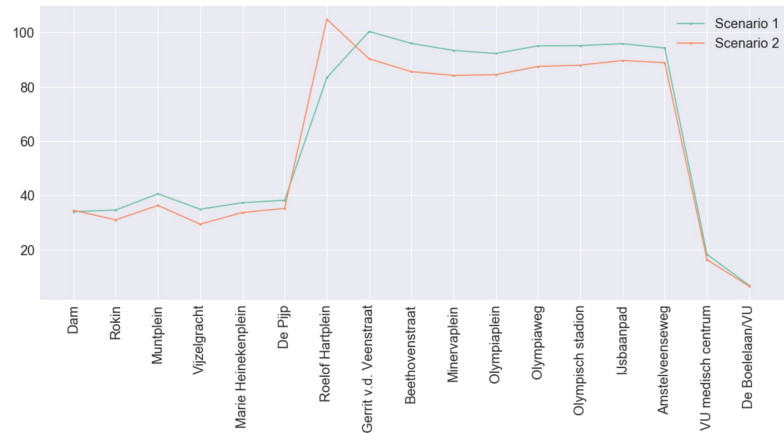


# Tram 12

## Richting Centraal station



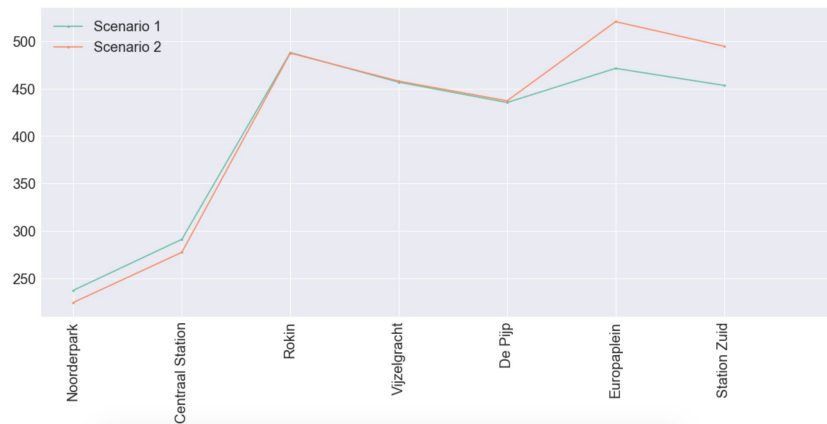
## Richting De Boelelaan/VU



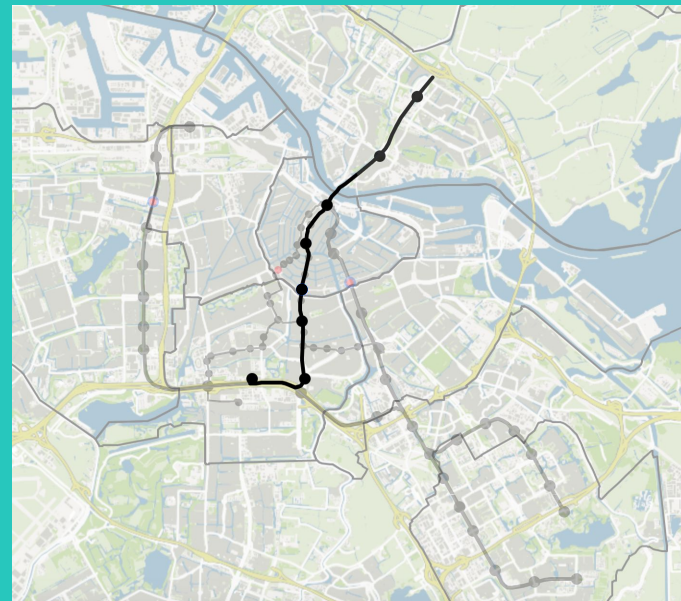
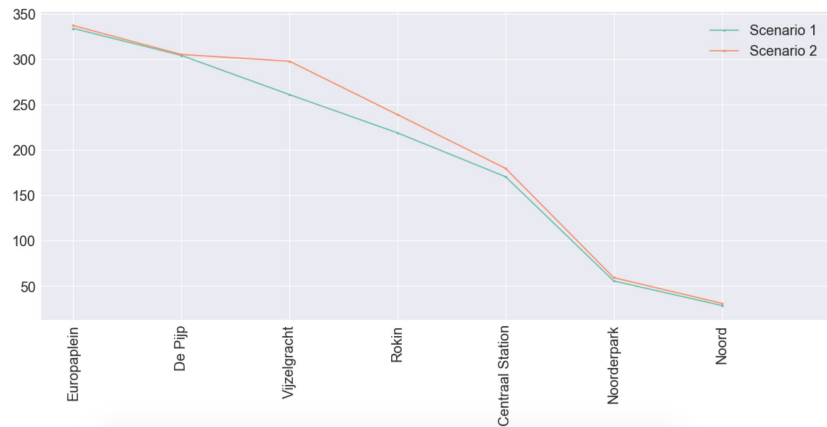
# Tram 24



### Richting Station Zuid

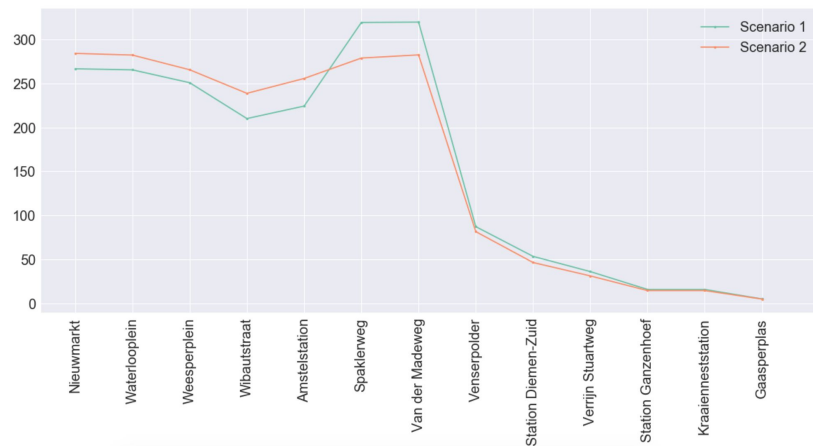


### Richting Station Noord

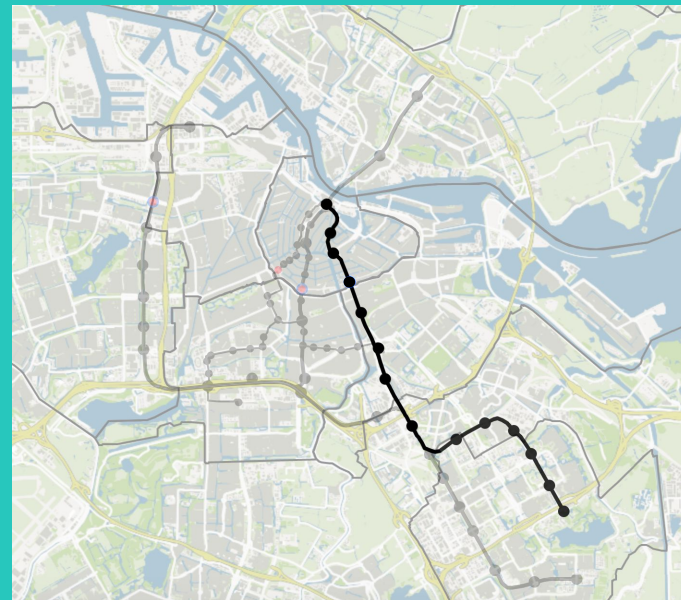
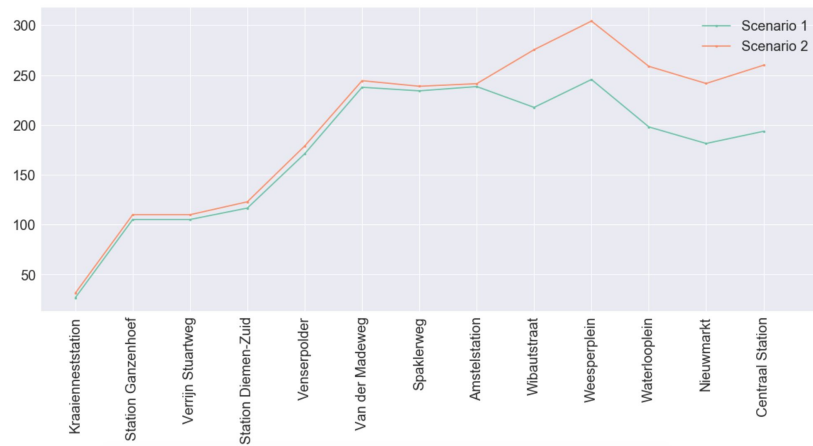


# Metro 52

## Richting Gaasperplas



## Richting Centraal Station



# Metro 53

---

# Discussie

- **Limitaties van de data**
    - Sub-trips
    - Begin en eindstation
  - **Sub-netwerk**
  - **Verdeling passagiers**
  - **Contributie**
    - GVB: kritieke lijnen en veranderingen
    - Gemeente: drukte als resultaat van veranderingen
-

---

# Conclusie

- **Verder onderzoek van belang**
    - Toepassen op het hele netwerk
    - Meer onderzoek naar gedrag
    - Wat is drukte in deze context?
  - **Model biedt duidelijke baseline performance**
-

**Bedankt voor jullie aandacht.**

---

---

# **Voorspellingsmodel Drukke in Amsterdam**

**Don de Lange**  
**Msc Data Science**

---

---

# Probleem

- Amsterdam is **DRUK**



—

**Hoe kan een voorspelling gemaakt  
van drukte binnen Amsterdam  
gegeven worden, met behulp van  
stad gerelateerde data?**



—

Hoe kan een voorspelling gemaakt  
van drukte binnen Amsterdam  
gegeven worden, met behulp van  
stad gerelateerde data?

---

# Gerelateerde Data

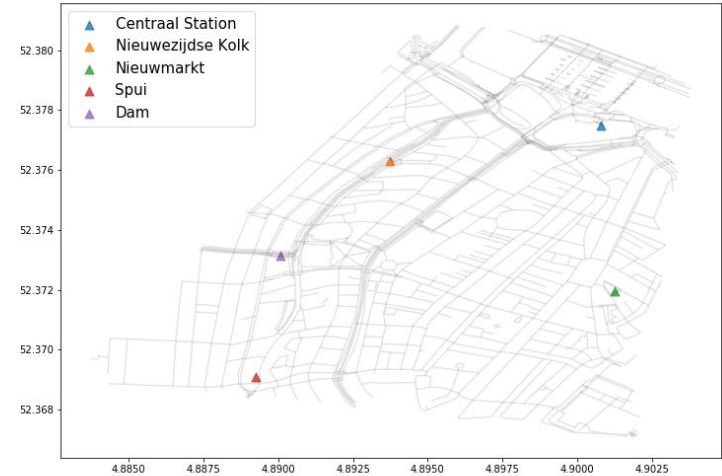
*Metingen per uur, voor iedere datum*

---

---

# GVB

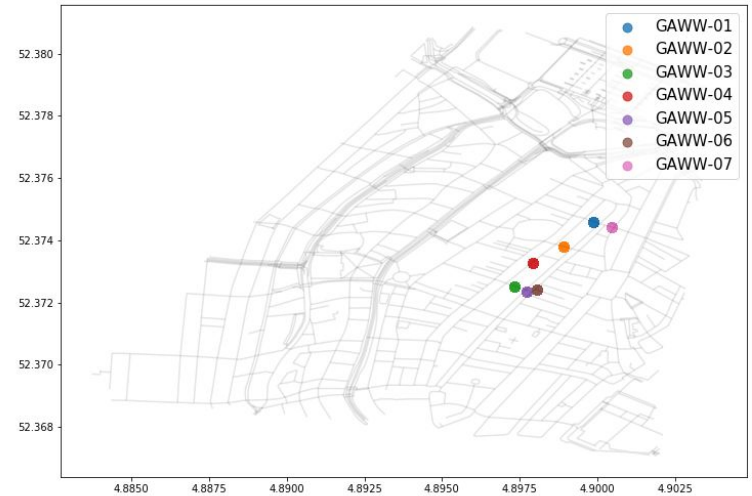
- Per Station
  - Tram en Metro  
Samengevoegd
  - Totaal aantal inchecken en  
uitchecken
  - Coördinaten



---

# CMSA

- Per Sensor ID
  - Wi-Fi sensor + Tel camera's zelfde gebied
  - Totaal voetgangers + apparaten
  - Coördinaten



---

# Evenementen

- Extra **DRUK**
- Per evenement
  - Datum
  - Coördinaten



—

**Drukte Voorspelling → CMSA gegeven drukte**

---

# Voorspellingsmethode

---

---

# Wat wordt er voorspeld

- 2 Methoden

Regressie

Drukte Voorspelling

Telling CMSA

Classificatie

Drukte Niveau

Deel tellingen CMSA

Level 1

0% - 25%

Level 2

25% - 50%

Level 3

50% - 75%

Level 4

75% - 100%

---



---

# Coördinaten gebruik

- Afstand GVB stations tot sensor die wordt voorspeld van belang
    - Stations met lagere afstand meer belangrijker
  - Afstand tussen GVB station en Sensor vooraf berekend en toegevoegd als gewicht
-

---

# Voorspellingsmodellen

- Random Forest
    - Traint **meerdere** beslissingsbomen op **deelverzamelingen** van de data
    - Iedere **zwakke** beslissingsboom wordt **gecombineerd** tot 1 **sterke** beslissingsboom.
  - XGBoost
    - Traint sequentiële beslissingsbomen op **deelverzamelingen** van de data
    - Error functie beslissingsboom minimaliseren
-

---

# Evaluatie

- Regressie
    - **R<sup>2</sup>:**
      - In hoeverre de voorspelling kan worden uitgelegd met de gegeven data
      - Zo hoog mogelijk
  - Classificatie
    - Accuracy: Ratio correct gelabelde voorspellingen
    - **Precision:** Per niveau, correct gelabelde voorspellingen van het totaal voorspellingen met het label
    - Recall: Per label, correct gelabelde voorspellingen van het totaal instanties met het label
-

---

# Voorspelling

- Voorspel drukte op alle sensoren

Methode:

1. Train modellen met cross-validation op 80% van de dagen
  2. Maak voorspellingen op overige 20%
-

---

# Voorspelling Resultaten

Regressie	
Model	R2
Baseline	57.7%
Random Forest	83.3%
<b>XGBoost</b>	<b>85.2%</b>

Classificatie			
Model	Accuracy	Precision	Recall
Baseline	24.1%	24.1%	25%
Random Forest	84.4%	84.4%	84.4%
<b>XGBoost</b>	<b>85.8%</b>	<b>85.8%</b>	<b>85.8%</b>

---

---

# Generalisatie

---

---

# Generaliseer voorspellingen

- Voorspellingen niet getrainde sensor locaties

Methode:

1. Train modellen met cross-validation op 6 sensoren
  2. Maak voorspellingen overgebleven sensor
-

---

# Generalisatie Resultaten

Regressie	
Model	R2
Baseline	58.3%
Random Forest	84.4%
<b>XGBoost</b>	<b>85.5%</b>

Classificatie			
Model	Accuracy	Precision	Recall
Baseline	24.2%	24.2%	25%
<b>Random Forest</b>	<b>84.2%</b>	<b>84.2%</b>	<b>84.2%</b>
<b>XGBoost</b>	<b>84.2%</b>	<b>84.2%</b>	<b>84.2%</b>

---



---

# Discussie

- GVB data kan gebruikt worden om drukte te voorspellen
- Modellen kunnen voorspellingen maken op onbekende punten

## Limitaties

- Onduidelijk hoeveel CMSA sensoren er nodig zijn
  - Onduidelijk welke CMSA sensoren nuttiger zijn dan andere
-

---

---

# Conclusie

- Verder onderzoek CMSA sensoren van belang
  - Verder onderzoek naar generalisatie van belang
  - Onderzoek goed startpunt
-

---

**Bedankt voor jullie  
aandacht**