

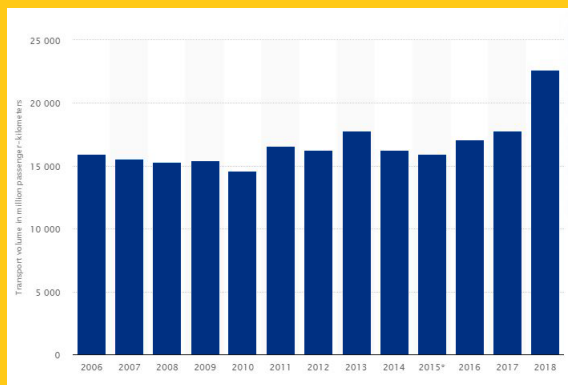
**CASTAWAY NL**

Vertrek	Naar / Opmerkingen	Spoor	Trein	09 57
10:08	Maastricht Rijdt niet	15	 Intercity sein- en wisselstoring	
10:08	Nijmegen Rijdt niet	14	 Intercity sein- en wisselstoring	
10:10	Almere Oostvaarders Rijdt niet	3	 Sprinter sein- en wisselstoring	
10:10	Den Helder Rijdt niet	5	 Intercity sein- en wisselstoring	
10:11	Breda Rijdt niet	19 <sup>b</sup>	 Sprinter	
10:13	Schiphol Rijdt niet	7	 Intercity sein- en wisselstoring	
10:14	Den Haag Centraal Rijdt niet	9	 Intercity	

# PROBLEM

NS (Nederlandse Spoorwegen):

- 1 million travelers per day
- 4,800 train journeys per day
- 3,000 railway cars
- Over 260,000 seats
- 380 stations
- 35 million passengers in September 2018
- 36,7 million passengers in September 2019



Volume of passenger rail transport in the Netherlands 2006-2018  
(in million passenger-kilometers)  
Source: Statista 2020



# INSIGHTS



*All NS train stations in The Netherlands*



# INSIGHTS



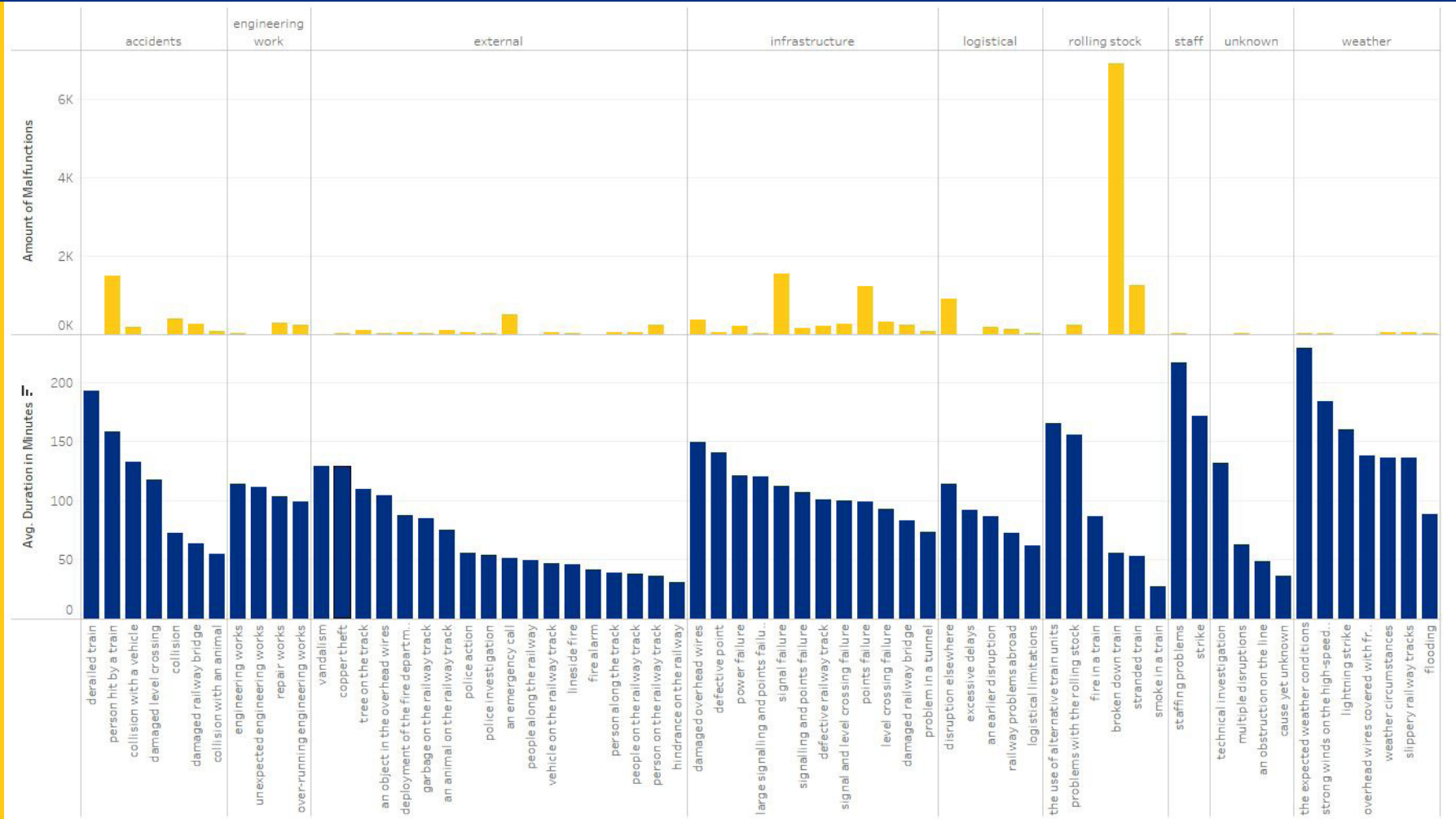
*Accumulated time taken to solve a malfunction (2019)*

# INSIGHTS



Average duration of malfunction  
35min. - 180min.

# INSIGHTS



# PRODUCT

## Features used:

- Station Name
- Month
- Day of the Week
- Holiday (is it a national holiday or not?)
- Time Type (peak or off-peak)
- Malfunction Cause
- Rain
- Wind
- Temperature

**PRODUCT**

**CASTAWAY DEMO**



# NEXT STEPS

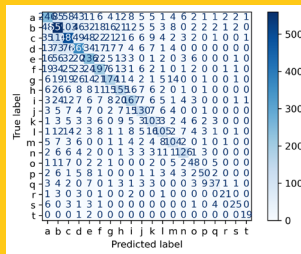
## Improve results of models

### Classification Strategy:

- Random Forest Classifier (duration grouped into 15 minute intervals)

Accuracy = 0.59

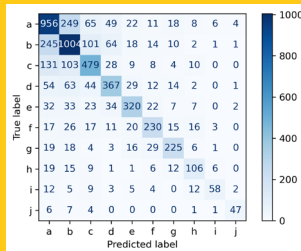
Kappa = 0.55



- Random Forest Classifier (duration grouped into 30 minute intervals)

Accuracy = 0.66

Kappa = 0.59



### Regression Strategy:

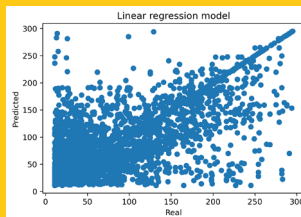
- K Neighbors Regressor

R2 = 0.52

MAE = 27 (minutes)

MSE = 2169.97

RMSE = 46.58



# Improve data collection

- Get historic data from as far back as possible.
- Connect to NS API to update the dataset with new data every day or week.
- Collect data that allows to dismiss programmed malfunctions, such as engineering projects.
- Collect data on distributions of passenger amounts throughout the day, the week, and the year.
- Find new features that influence the duration of a malfunction.

#### Data sources:

- Rijden de Treinen. “Open data”. RIJDENDETREINEN.nl.  
<https://www.rijdendetreinen.nl/> (accessed November 28, 2020).
- Koninklijk Nederlands Meteorologisch Instituut. “Daggegevens van het weer in Nederland”. KNMI.nl.  
<http://projects.knmi.nl/klimatologie/daggegevens/selectie.cgi> (accessed November 28, 2020).

**THANK YOU**