PREDICTING RESPONSE TIMES OF THE PARIS FIRE BRIGADE VEHICLES

Vincent Haguet

Capstone project - Coursera

Introduction

PARIS FIRE BRIGADE IS USING DATA TO OPTIMIZE THEIR DISPLACEMENTS

THEY NEED TO BE ABLE TO PREDICT THEIR RESPONSE TIME

THEN THEY NEED TO PICK UP THE BEST VEHICLE TO GO ON SITE AS FAST AS POSSIBLE

The data output

There are four different components to be predicted:

The ID of the chosen vehicle

The time between selection of the vehicle and the departure

The time between departure of the vehicle and the arrival on site

The time between selection of the vehicle and the arrival on site (sum of the first two)

Large data set with many features of different types

The continuous variables

The categorical variables

The labels variables

Other types of variables

Supplementary file available but hard to use

Features that will be used:

Alert reason category

Emergency vehicle type

Estimated distance

Estimated duration

Floor of intervention

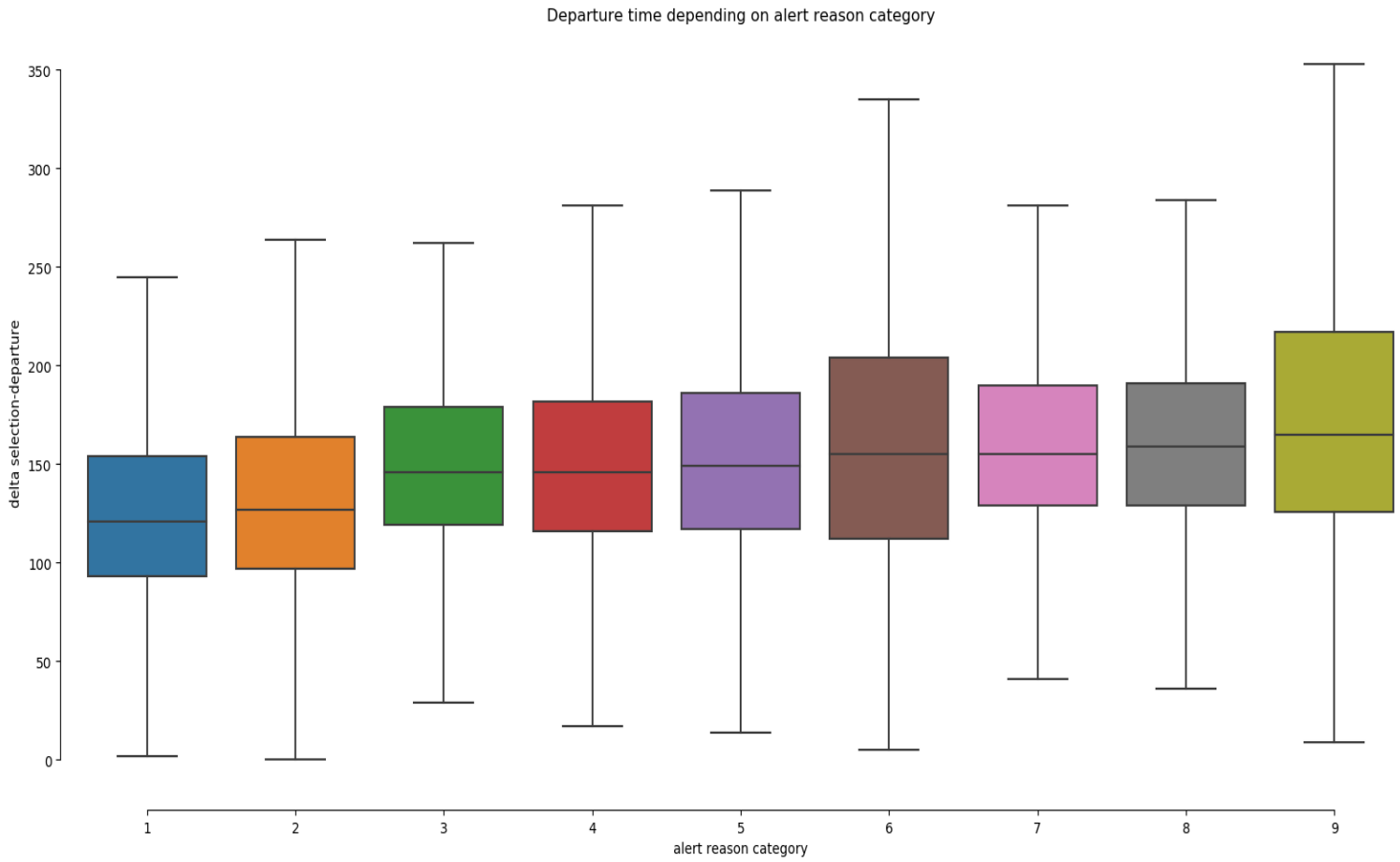
Hour of intervention

Month of intervention

Intervention on public roads

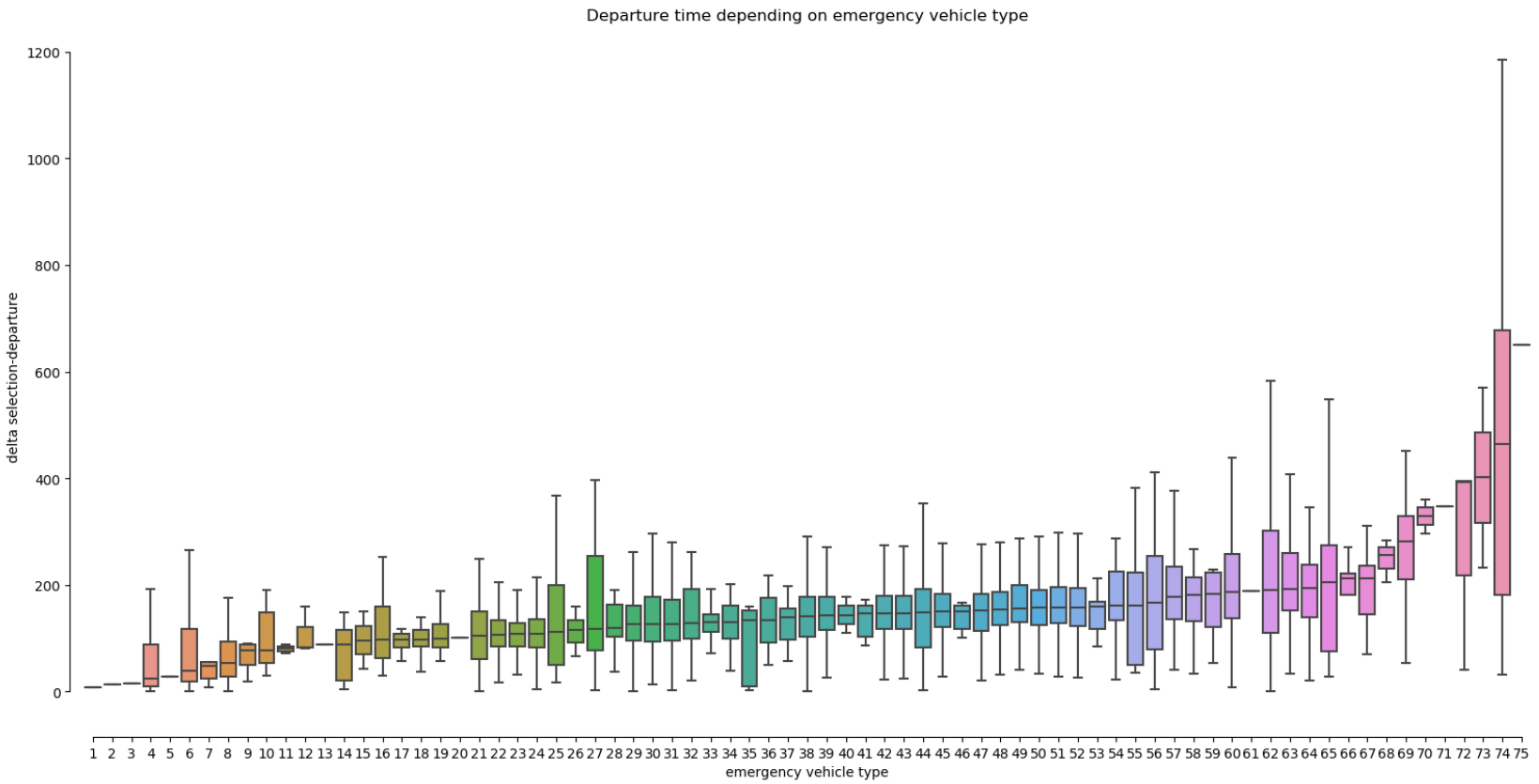
The data input

The data set is issued from the ENS website which proposes many challenges. A full description of the data set is available here: https://paris-fire-brigade.github.io/data-challenge/challenge.html



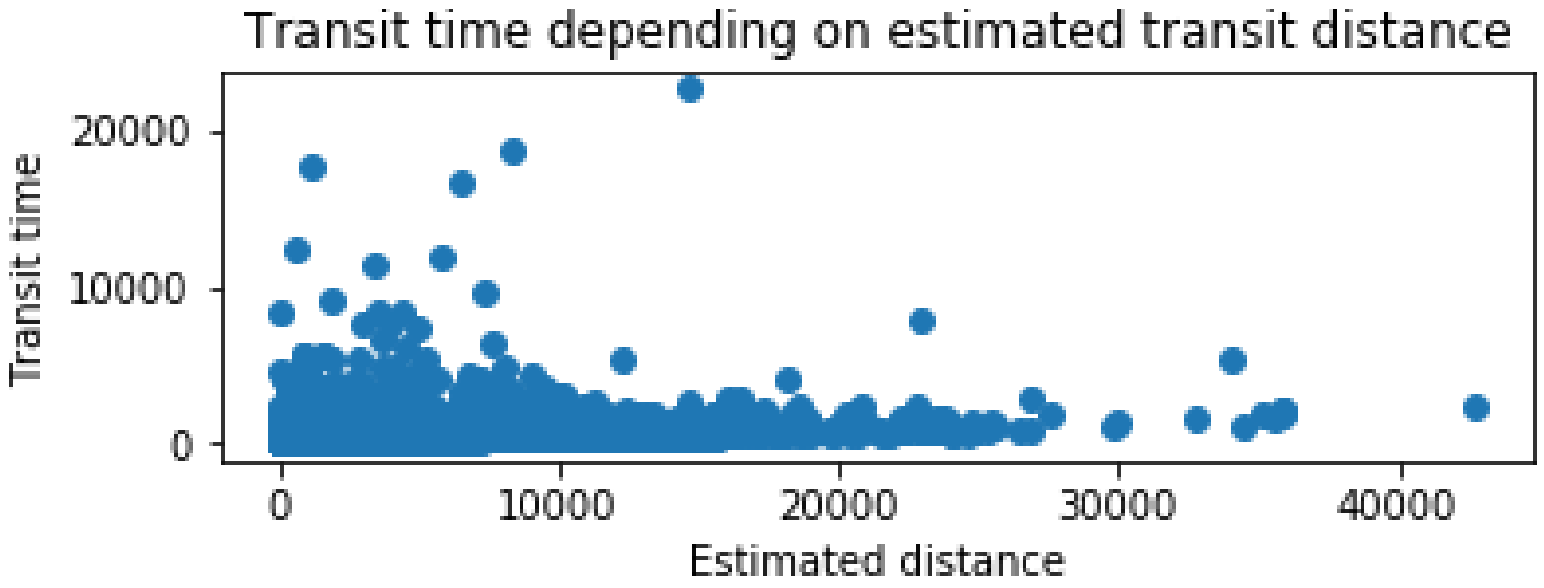
Data visualization

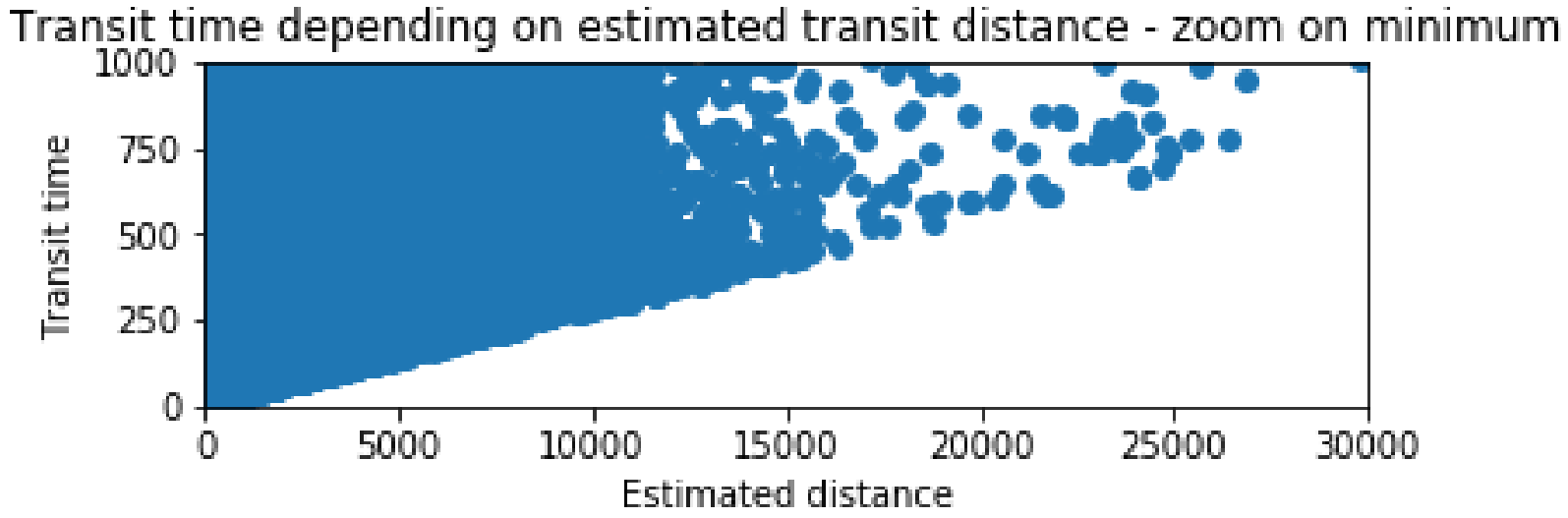
Influence of the alert category on the departure time



Data visualization

Influence of the vehicle type on the departure time





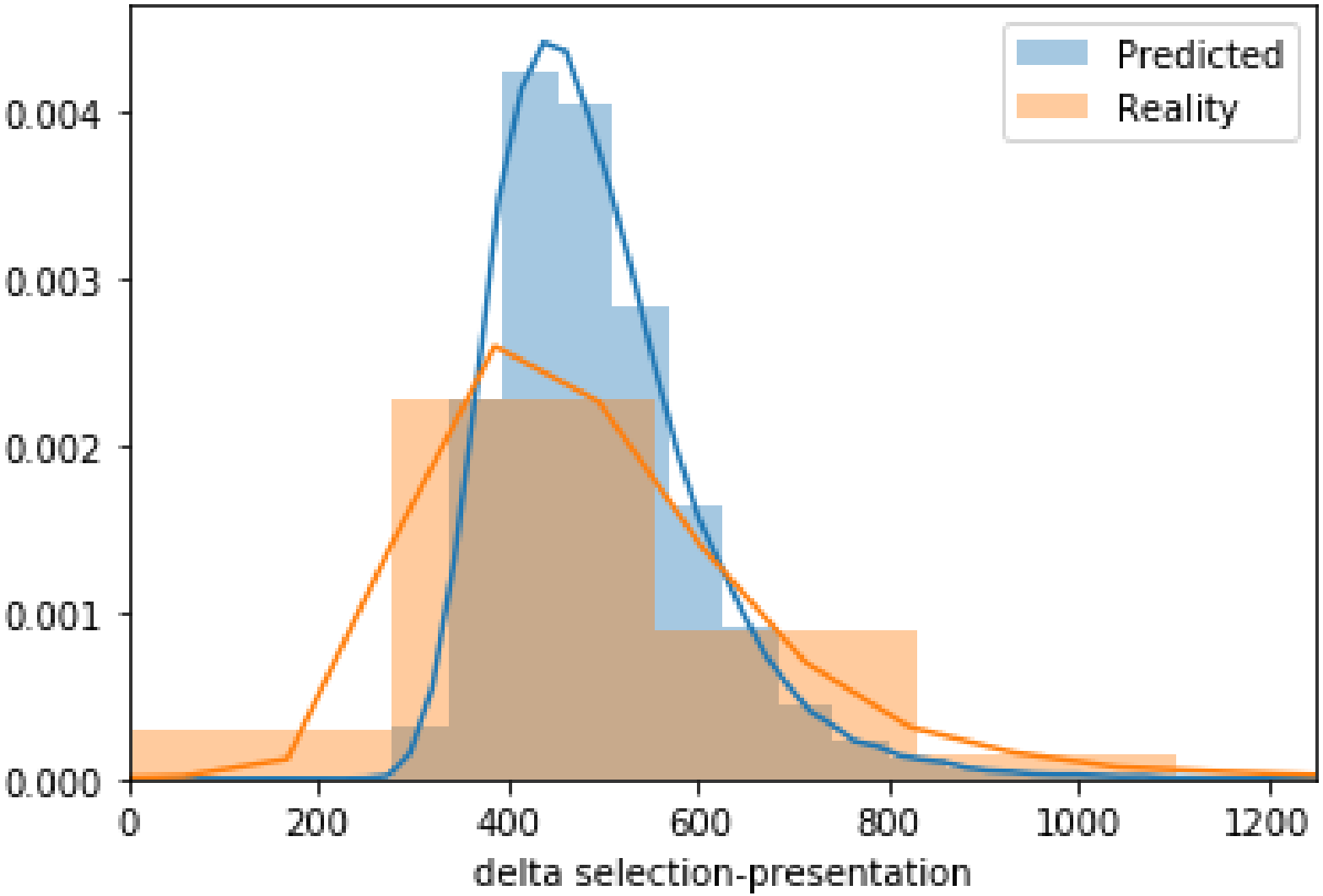
Data Visualization

Influence of the estimated distance on the transit time. A linear relation can be observed.

Choosing the model
Departure time

Different types of models can be used:

- KNeighborsClassifiers, accuracy score: ~8%
- DecisionTrees, accuracy score: ~12%
- Random forest, accuracy score: ~14%
- Linear regression, R^2 score: 0,38



Results

Here are the prediction plotted on a distplot.

We can see that the model badly predicts the extreme values. Since we removed many features the model can't predict extreme values. It also undertrained compared to the complexity of the dataset.

Mean average error

Another way of evaluating the precision of the model is to calculate the mean absolute error.

Mean absolute error: 114 seconds

Average response time: 495 seconds

Mean absolute error/ average response time: 23%

Conclusion

The model is simple and has a low success rate but can approximately predict the outcome with a 23% relative average error, which is not so bad given the complexity of the problem.

Many improvements can be done such as:

Implementing other features	Frequency analysis on each feature	Gradient boosting techniques	….