

Data Visualization

Bizet Thibaut



Summary

- I/ Data visualization
- II/ Map visualization
- III/ Machine Learning



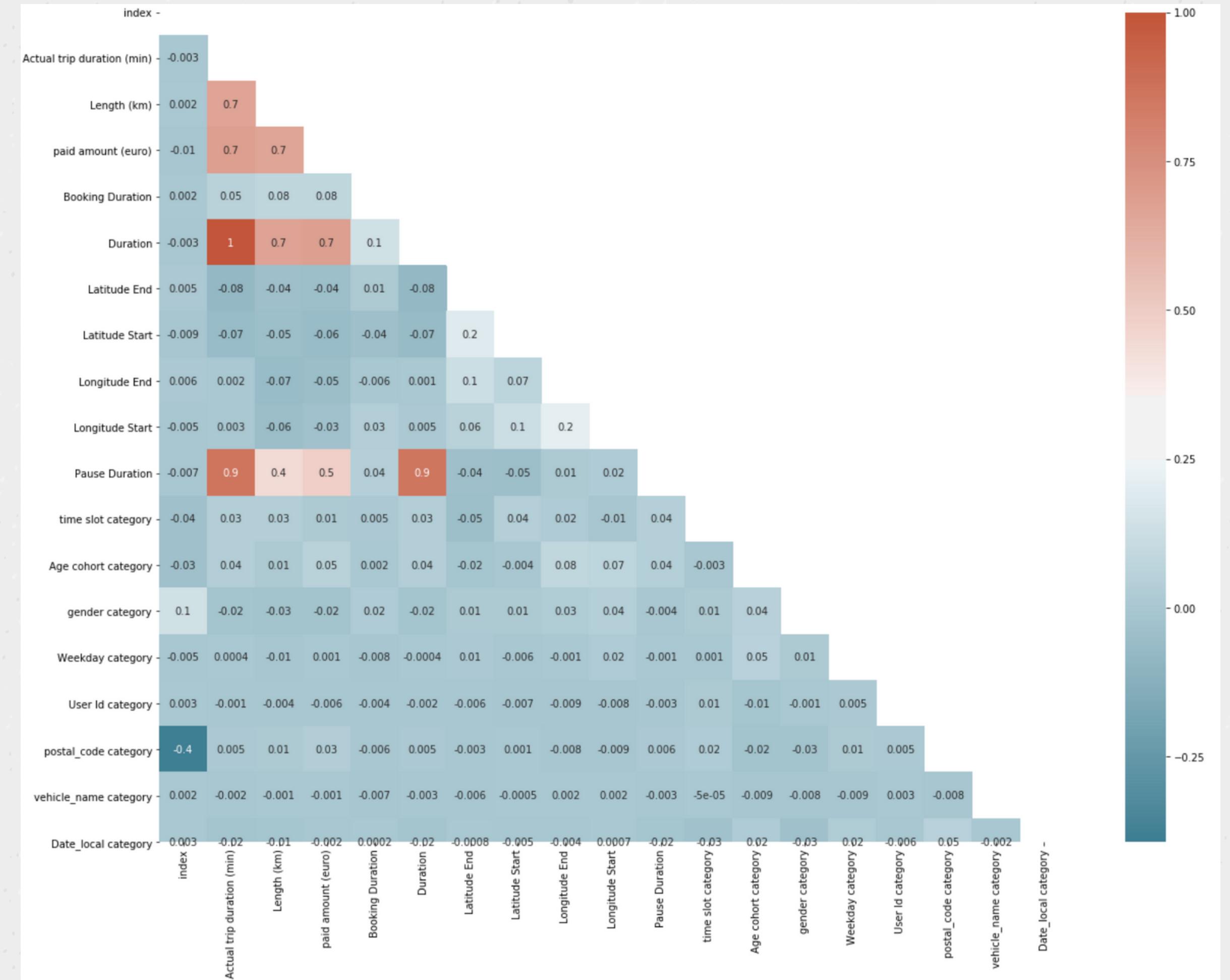
Data Visualization

- Matplotlib
- Seaborn

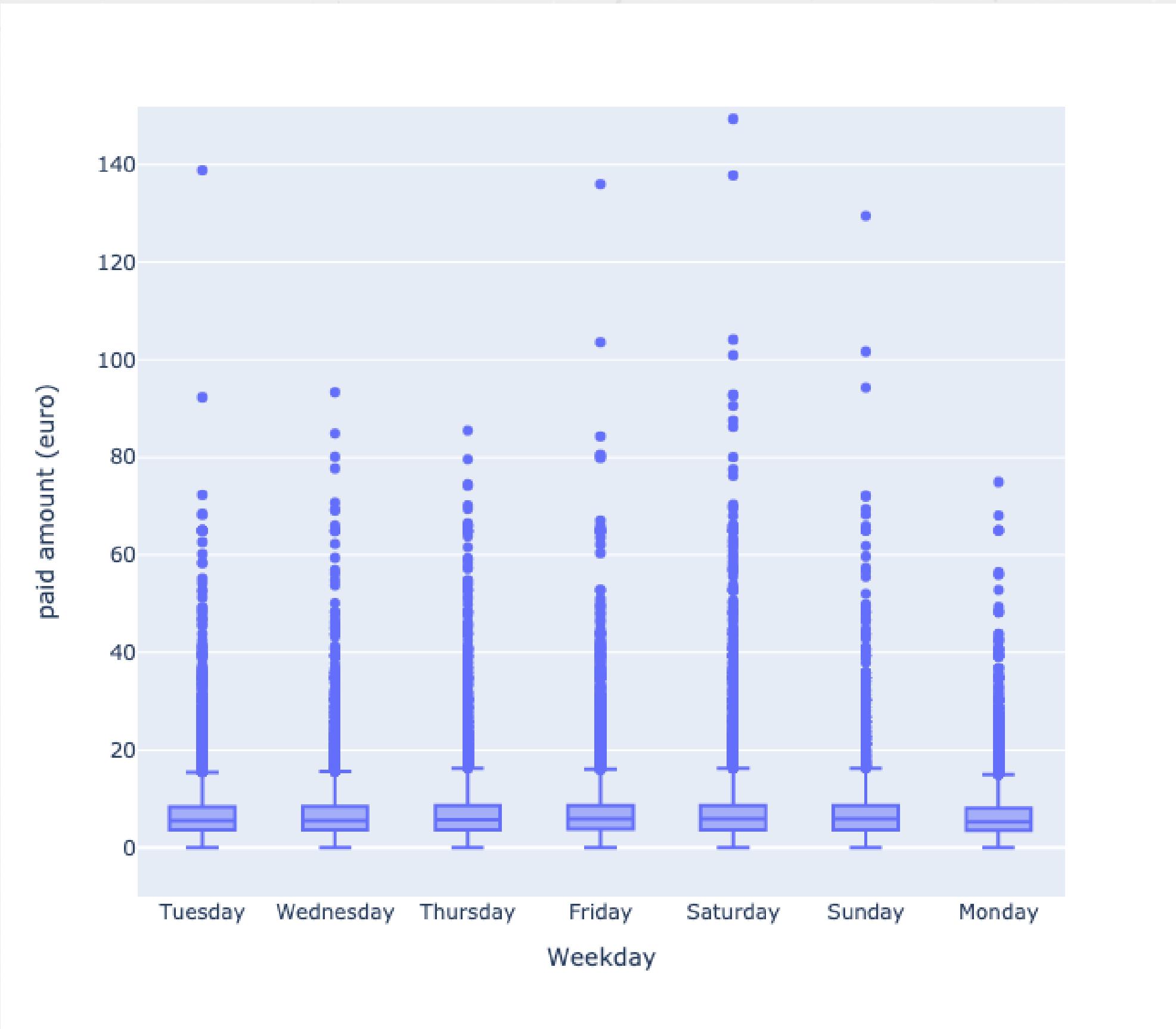


Correlation Matrix

First correlations are from what explains the duration and the price of a trip

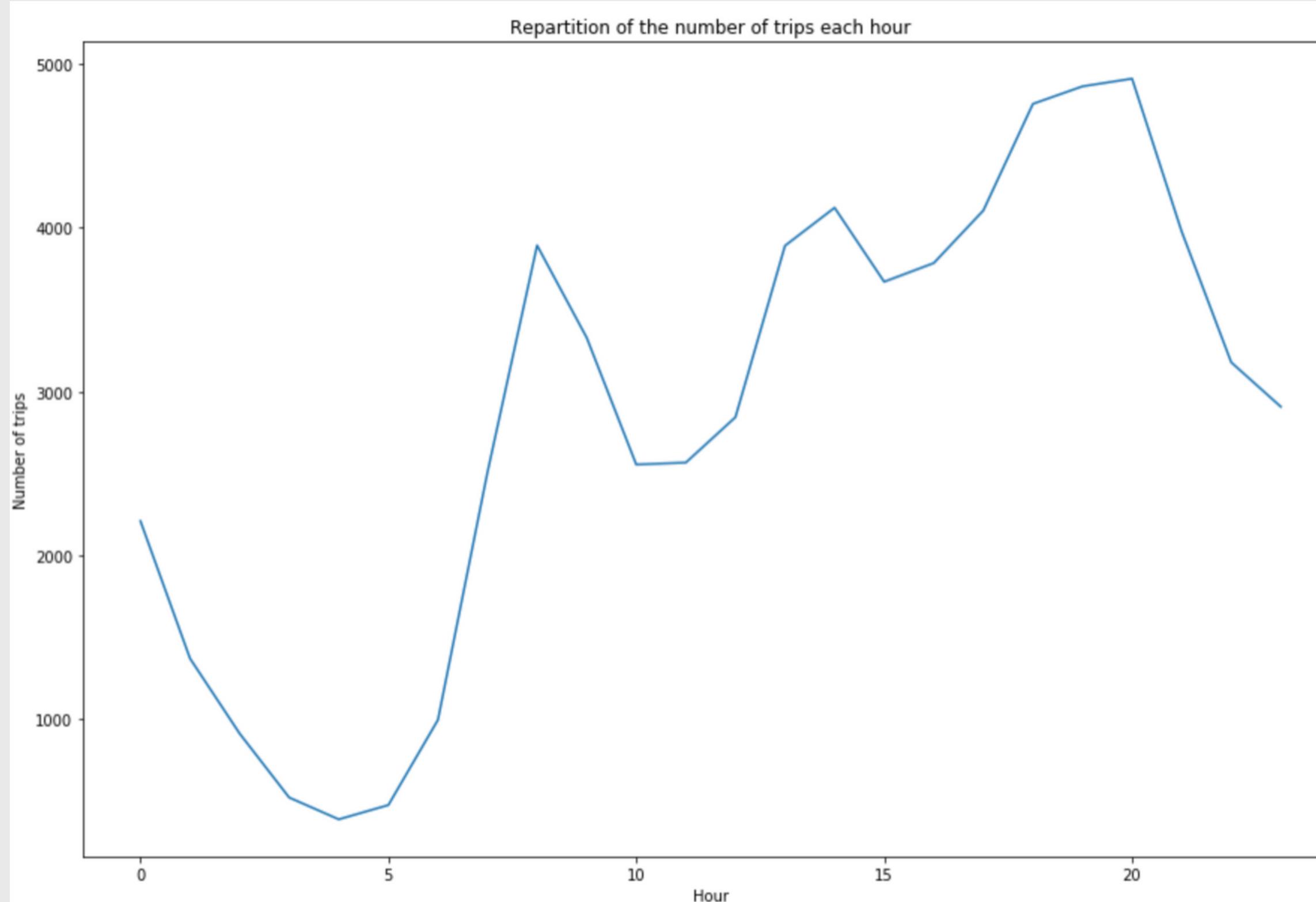


Price Distribution per day



The price distribution per day has a median quite close for each days

Number of trips per hour



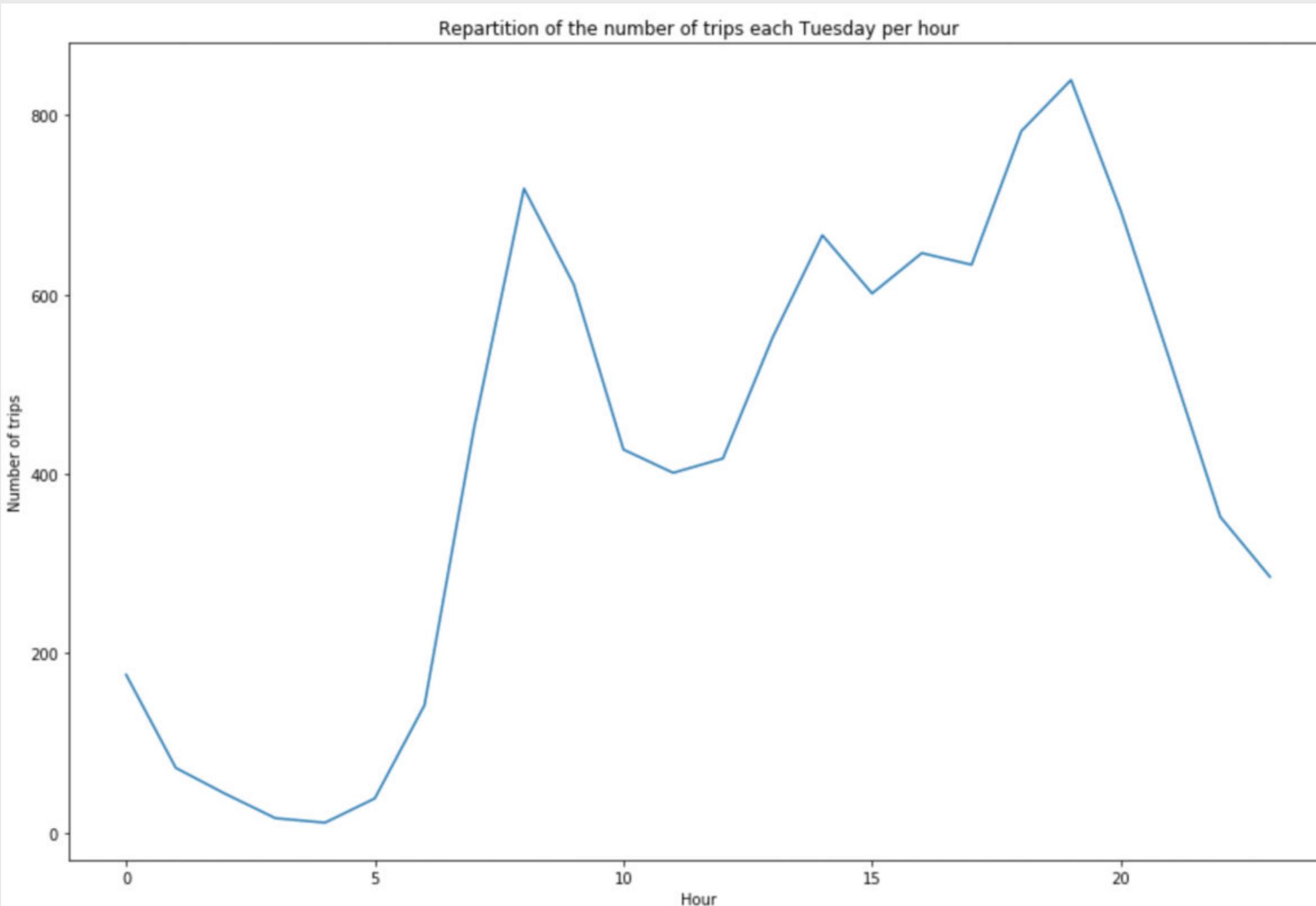
The dataset is a car rental service per minute. So, it's obvious people don't drive at late or early morning hours.

It's why the time period 00H-7H is very quiet.

We have 3 peaks :

- 8H, we can think users go to work
- 13-14H, to reach a place where people will pass their afternoon
- 19H-20H , we can think people come back home

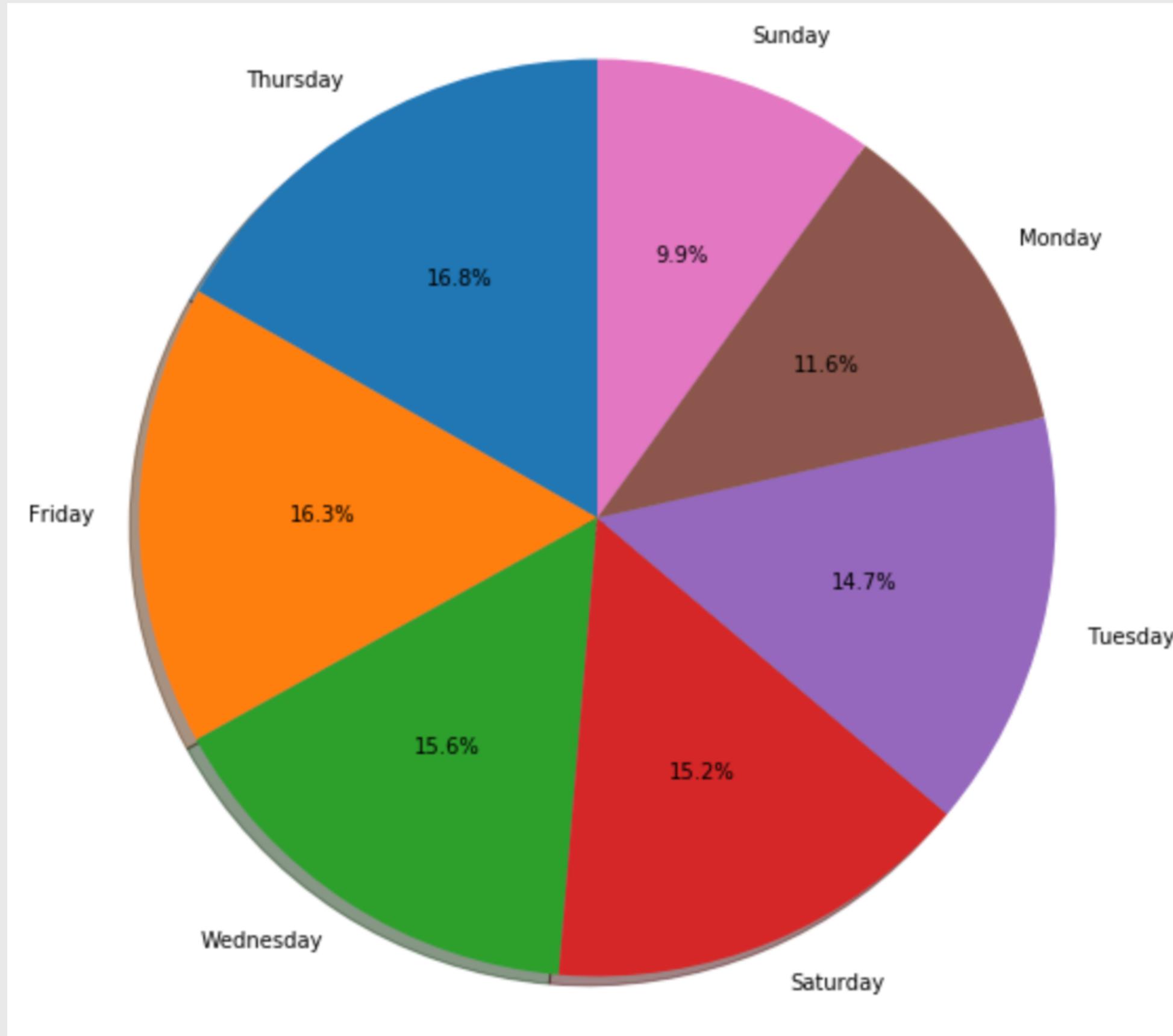
Number of trips per day & hour



The service experienced two different trends during a typical week :

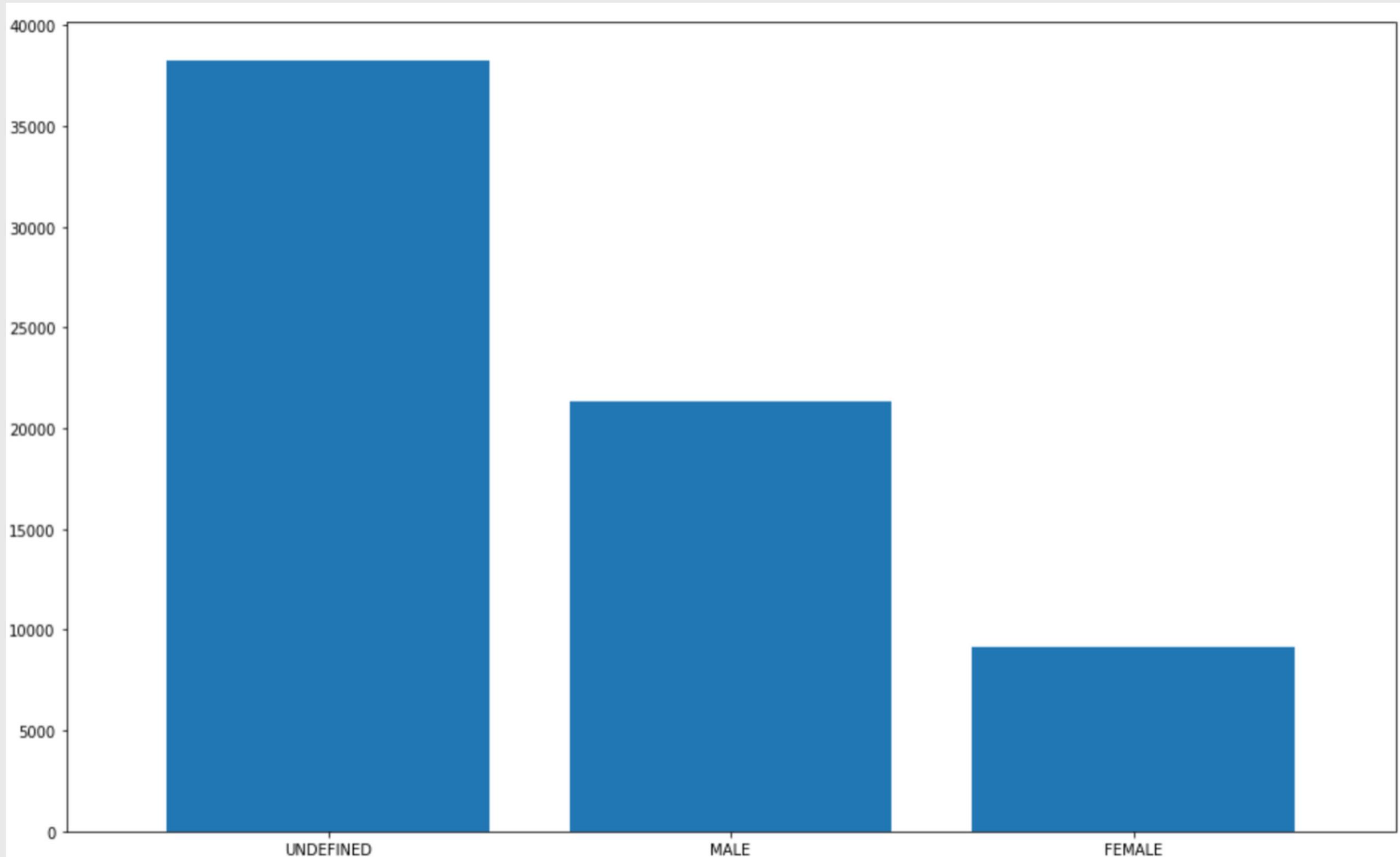
- Work days (Monday -> Friday)
- Week-end (Saturday -> Sunday)

Distribution of trips per day.



As you can see, the use of the service is homogeneous. There is a small drop during the week-end.

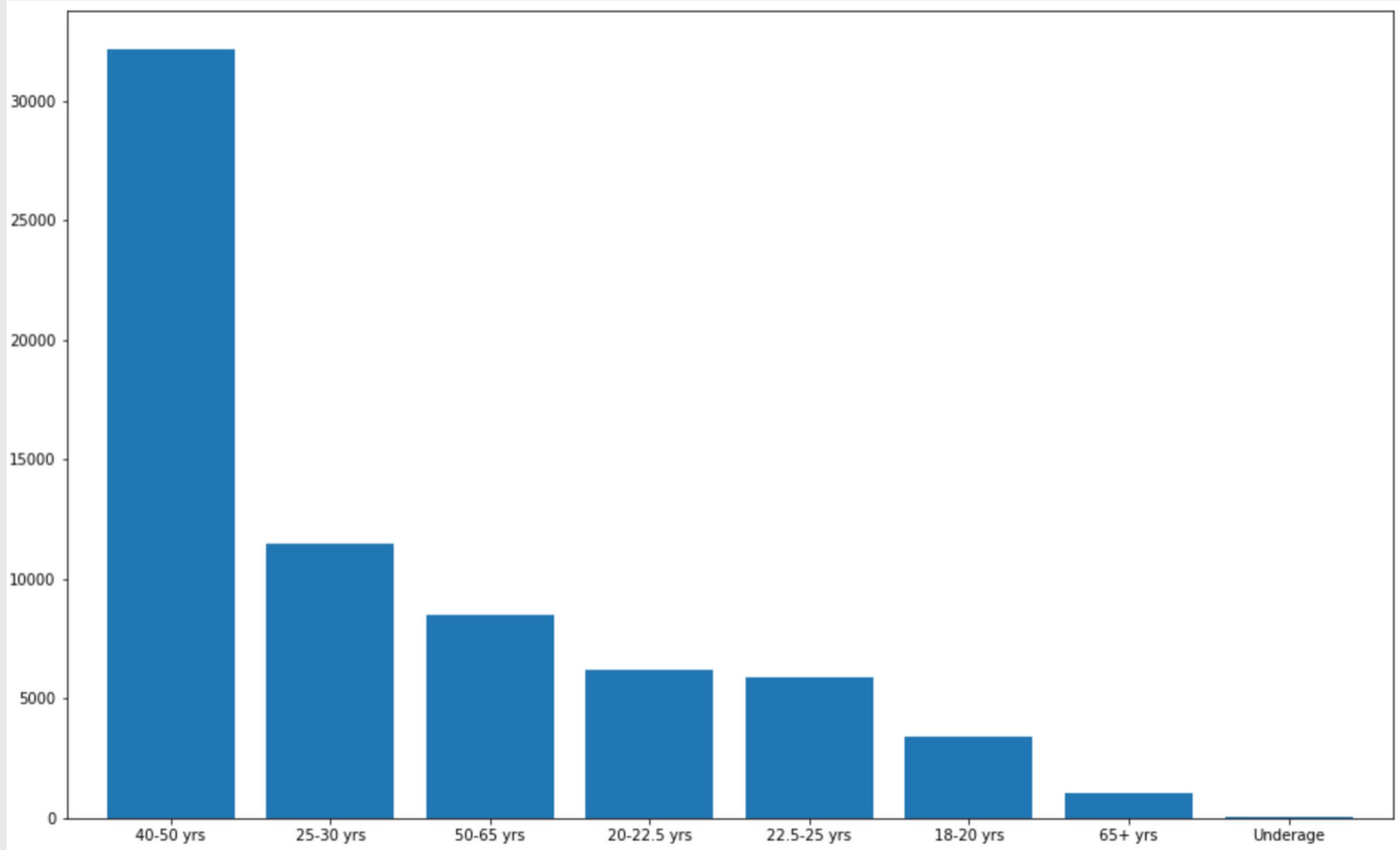
Customers sexe



More than the half of the data don't know the sexe of the client.

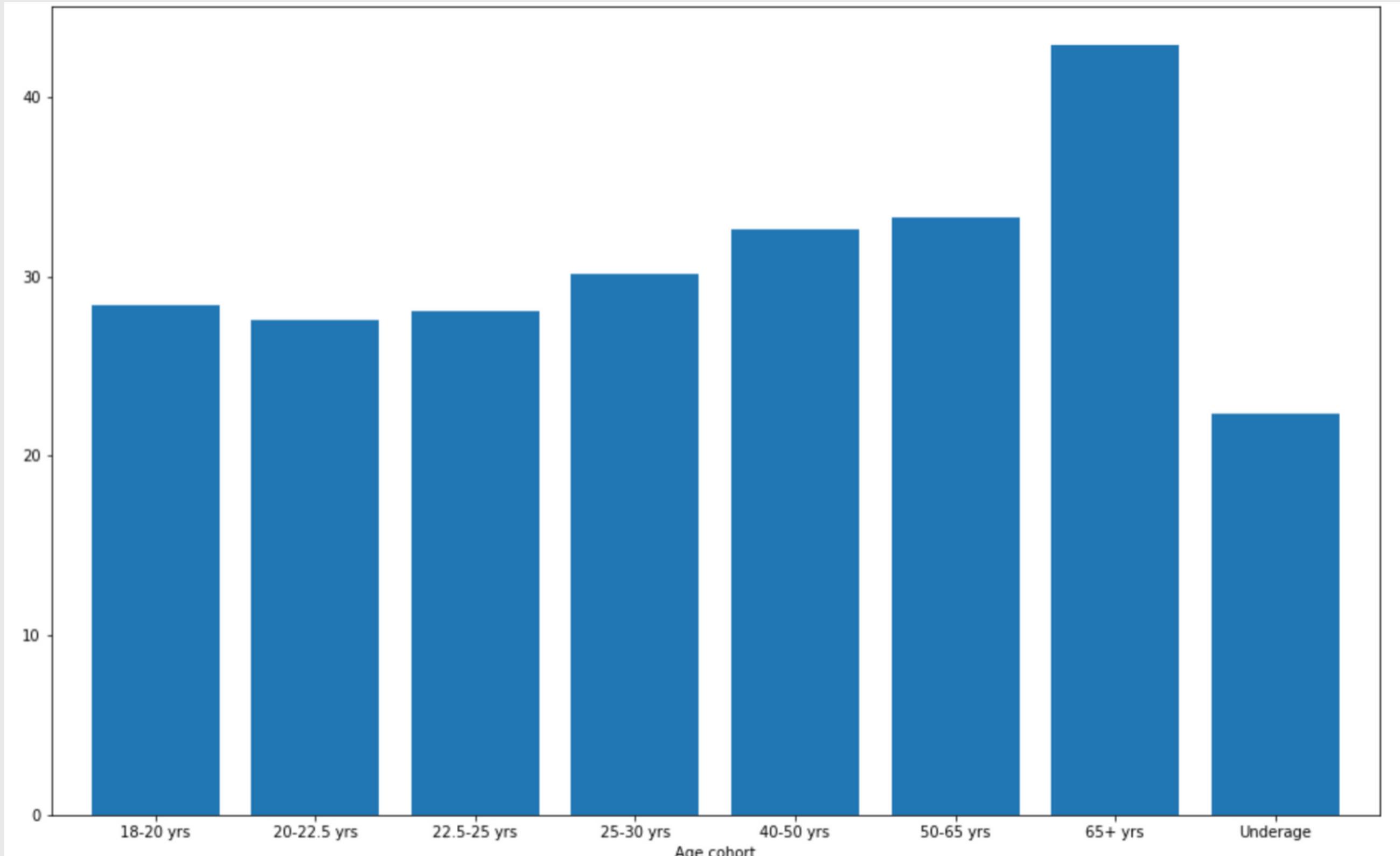
So, we cannot use this feature to understand a trend.

Customers Group



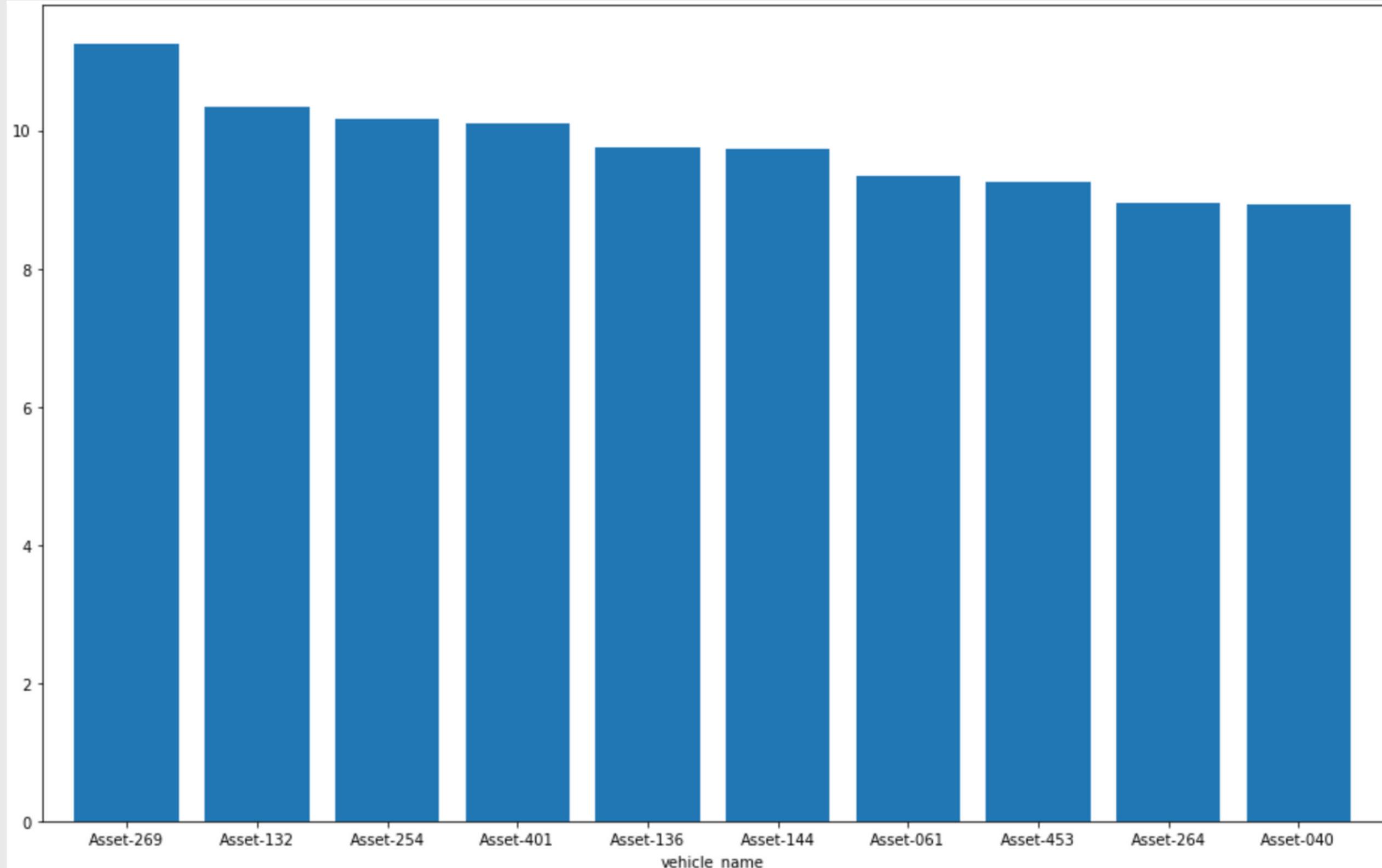
The service is not likely used by the millennials. But more by the generation X. We can think that millennials and generation Z are less present because they don't have a driving license. And prefer services such as Uber

Customers Group



This is the average actual trip duration for each age cohort. Which is quite similar for each group. excepting the group "65+" which is a bit higher

Vehicle fleet



Top 10 of cars the more used during the month. We can use these data to stay awake about what car is more likely to get an issue due to her use

Key values

Total of kilometers travelled : 467292.11393452197 Km
Revenue of the month : 494180.6 €
Price per kilometer : 1.0575410653500685 €

Average Booking Duration : 6.944951965065503 min
Average Pause Duration : 5.656999154050059 min
Average Duration : 38.333108949338516 min
So, customers aren't moving with their car
during 32.874847526117556 % of their trip

Average kilometers travelled per man : 7.35 Km
Average kilometers travelled per woman : 6.86 Km



Map Visualization

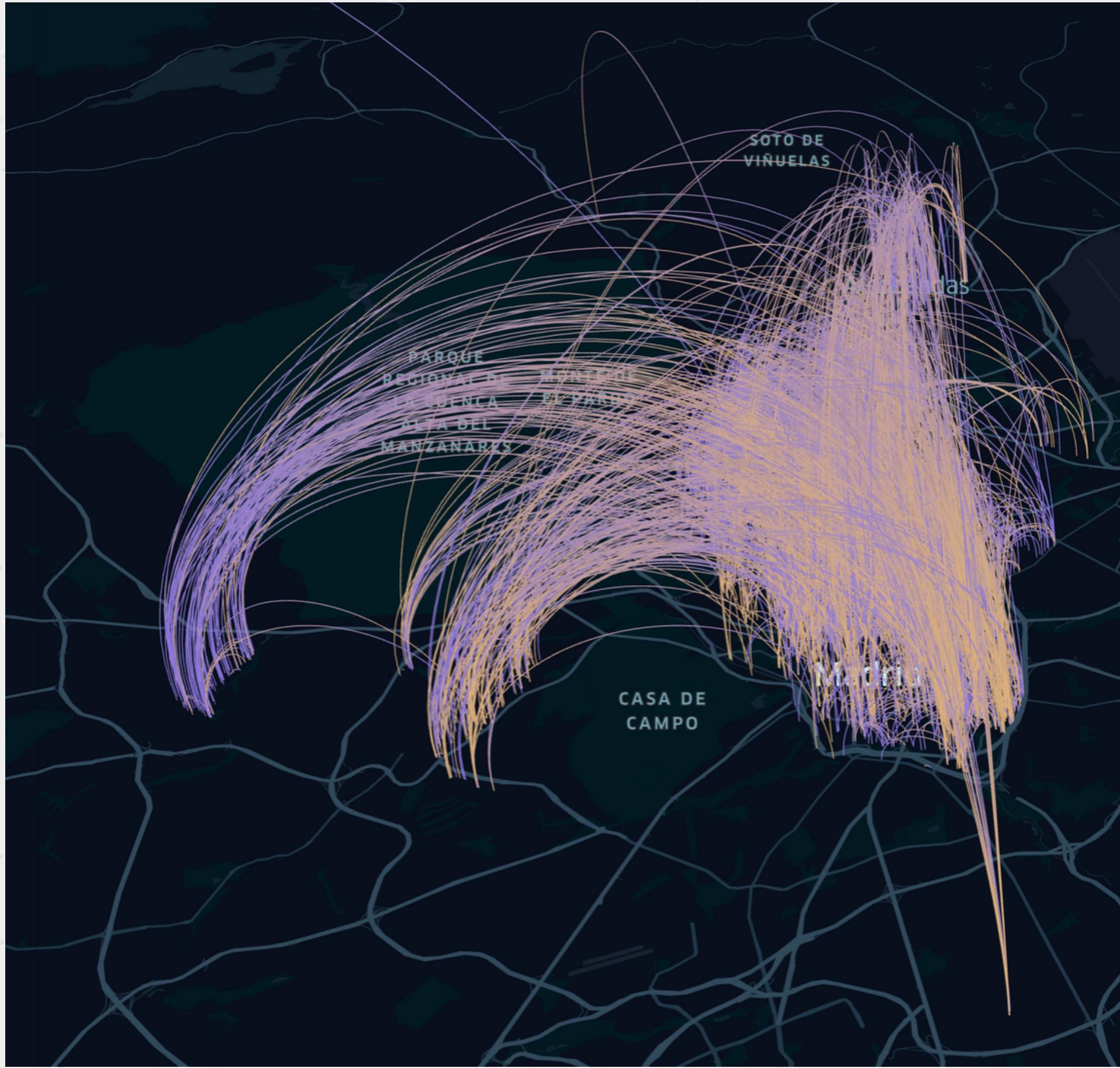
- Folium Heatmap
- Open Route Service API (to get direction)
- Kepler (Uber DataViz Tool)



Global Overview

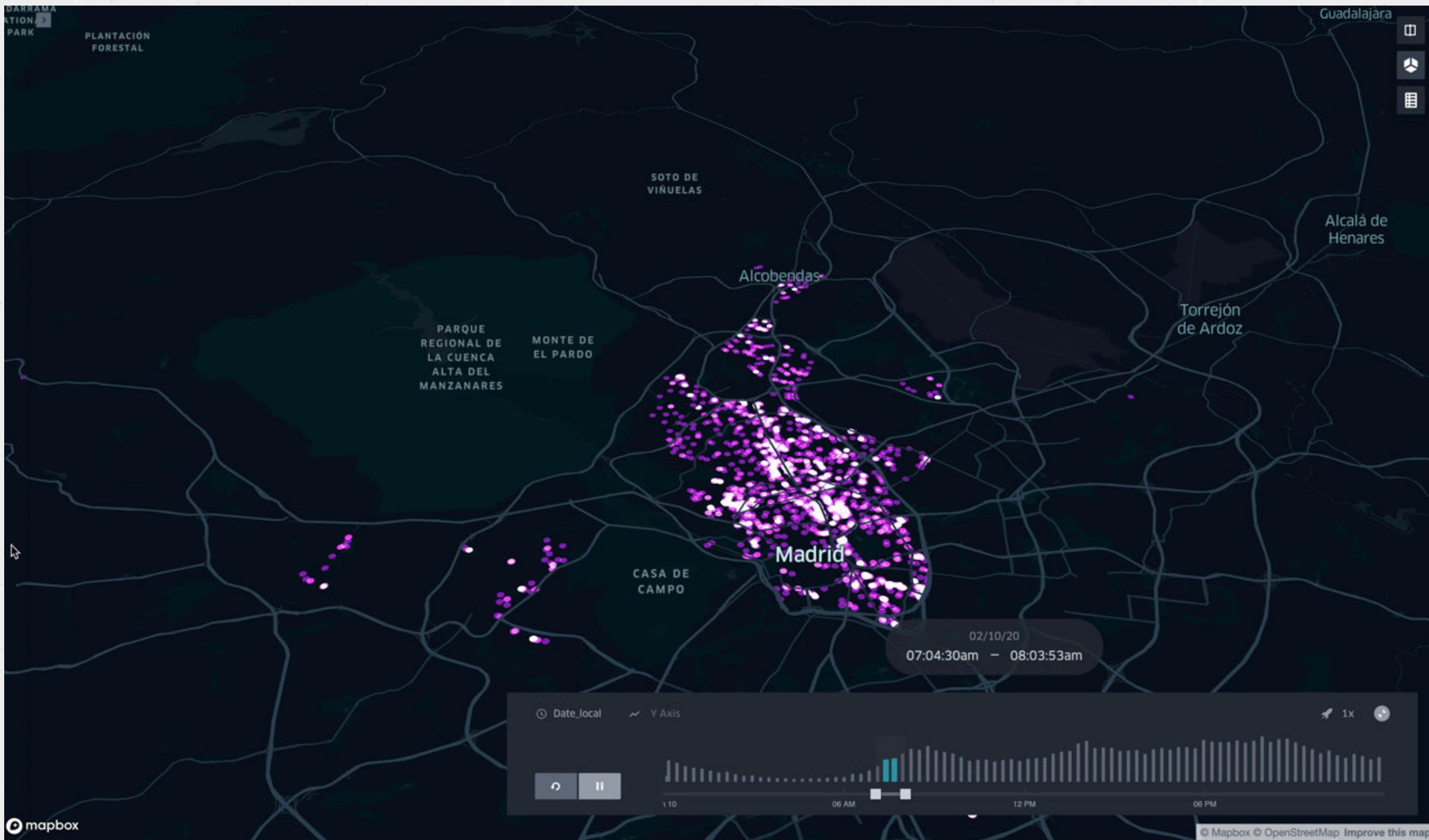


Global Overview

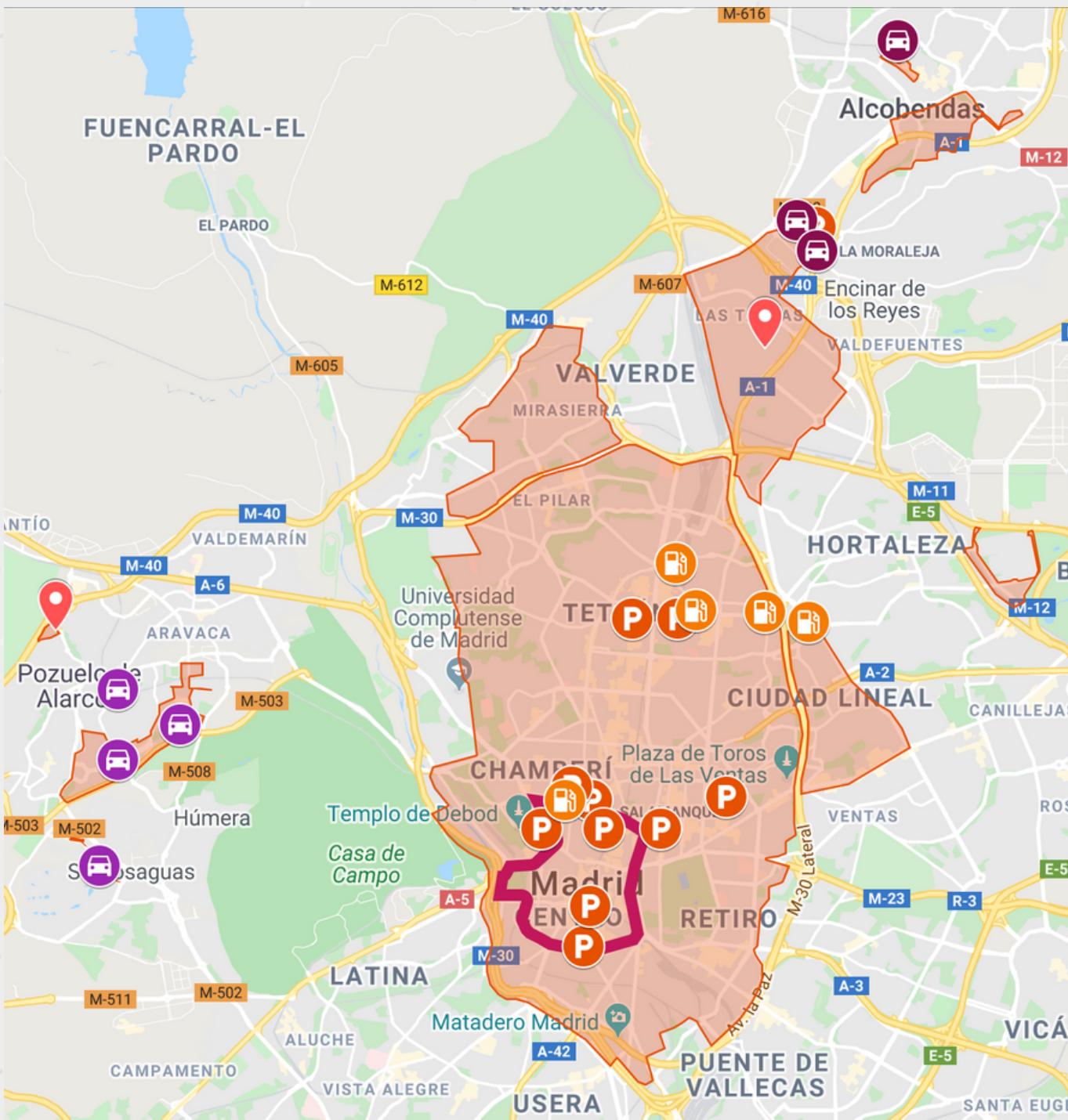


Customers mainly navigate between
7 different areas

End Position Vehicle

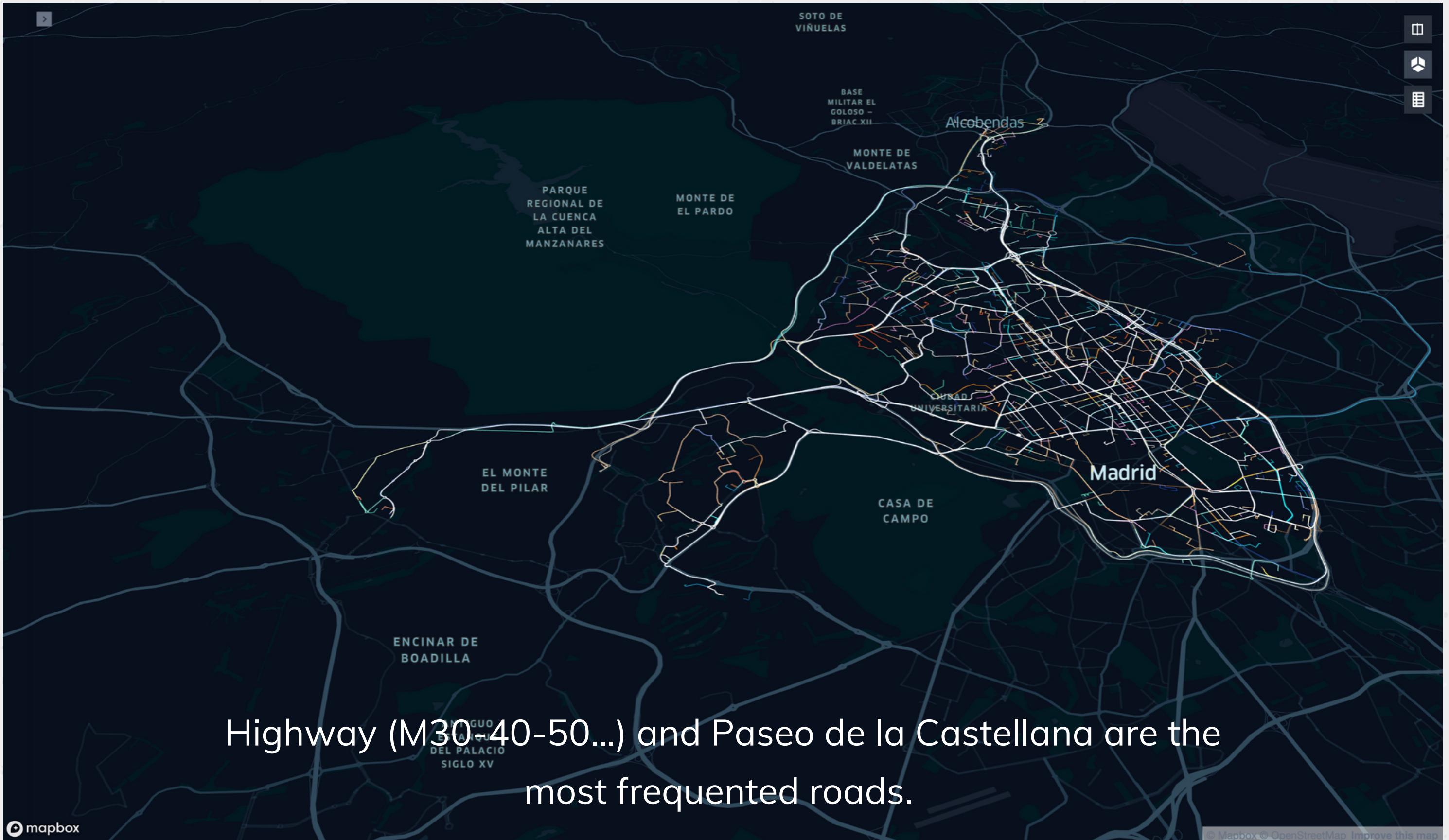


End Position Vehicle

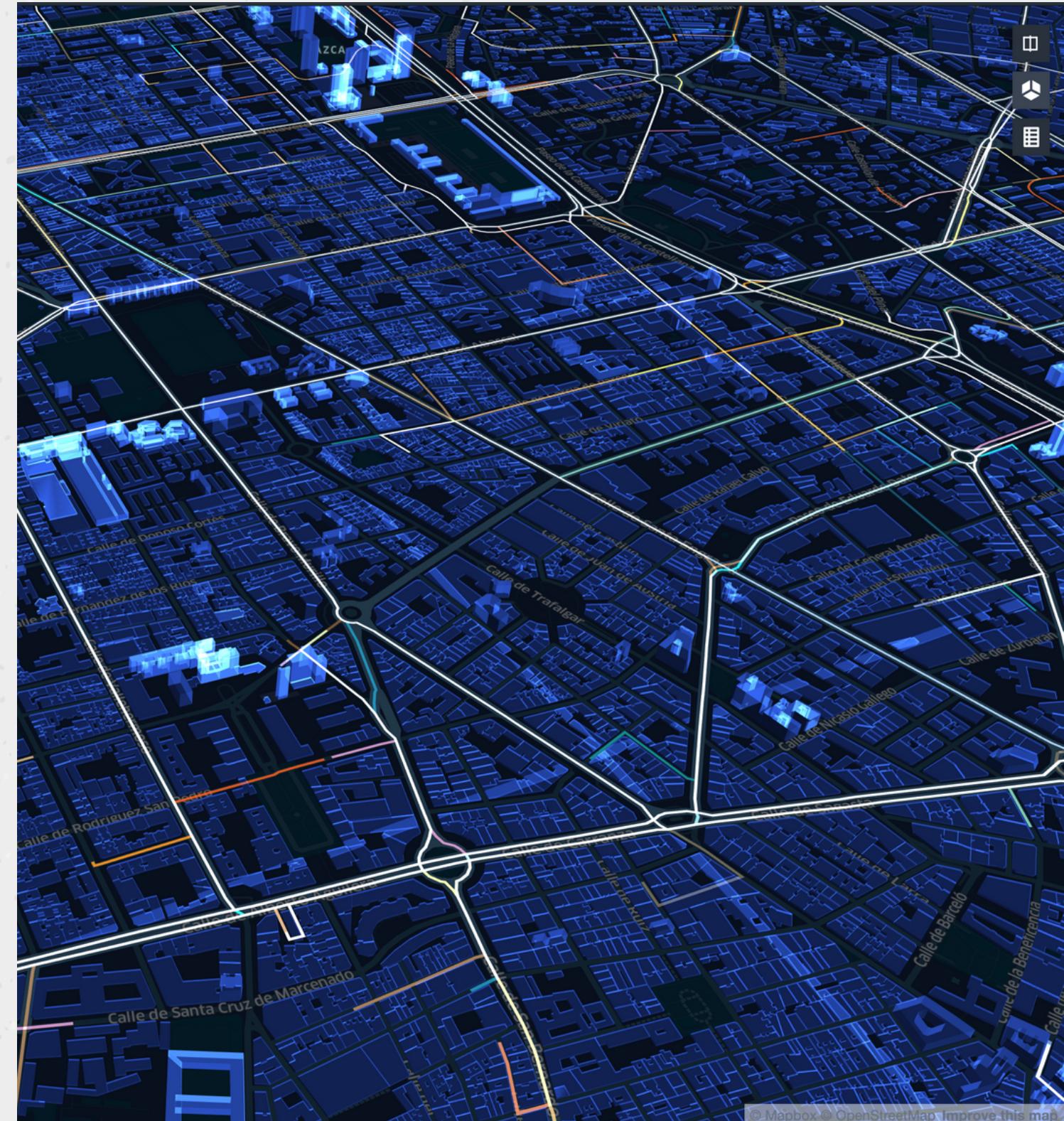


The heatmap show some red points which are representing the location of Wible's car park in Madrid. Furthermore, we can see again a concentration around the Azca district and Paseo de la Castellana. We can therefore think that a large part of our clientele is made up of rich and middle class people

Vehicle itinerary from 8:00 to 8:10



Vehicle itinerary from 8:00 to 8:10



Machine Learning

Predict the amount of a rental :

- OneHotEncoding
- Removing Outliers with std
- Mean Normalization
- Random Forest Regressor
- Feature Importance
- Randomized Search CV



With all features

```
model.score(X_test,y_test) :  
model.score(X_train,y_train) :
```

```
CPU times: user 4min 54s, sys: 1.43 s, total: 4min 55s  
Wall time: 46.6 s  
0.8340652385220476  
0.822536709420586
```

Using feature importance

Variable: Actual trip duration (min)	Importance: 0.96
Variable: Longitude End	Importance: 0.02
Variable: Latitude End	Importance: 0.01
Variable: Length (km)	Importance: 0.0
Variable: Duration	Importance: 0.0
Variable: Booking Duration	Importance: 0.0
Variable: Pause Duration	Importance: 0.0
Variable: Longitude Start	Importance: 0.0
Variable: Latitude Start	Importance: 0.0

```
CPU times: user 7.55 s, sys: 65.2 ms, total: 7.62 s  
Wall time: 1.17 s  
0.815773274828652  
0.8266125045986523
```

Using RandomizedSearchCV

```
CPU times: user 165 µs, sys: 1 µs, total: 166 µs  
Wall time: 171 µs  
0.8428240340551716  
0.8421032426098815
```

END

Bizet Thibaut

