



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis and Modelling for XYZ investment

Name: Juan Carlos

Location: Spain

Team: Data and Analytics

20-March-2021

Agenda

Business Problem

Approach

EDA insights

Model and forecast

Recommendations

What's the problem we want to tackle?

The taxicabs industry is growing quickly in last years and might be a good opportunity of investment.

We have two options for investment:

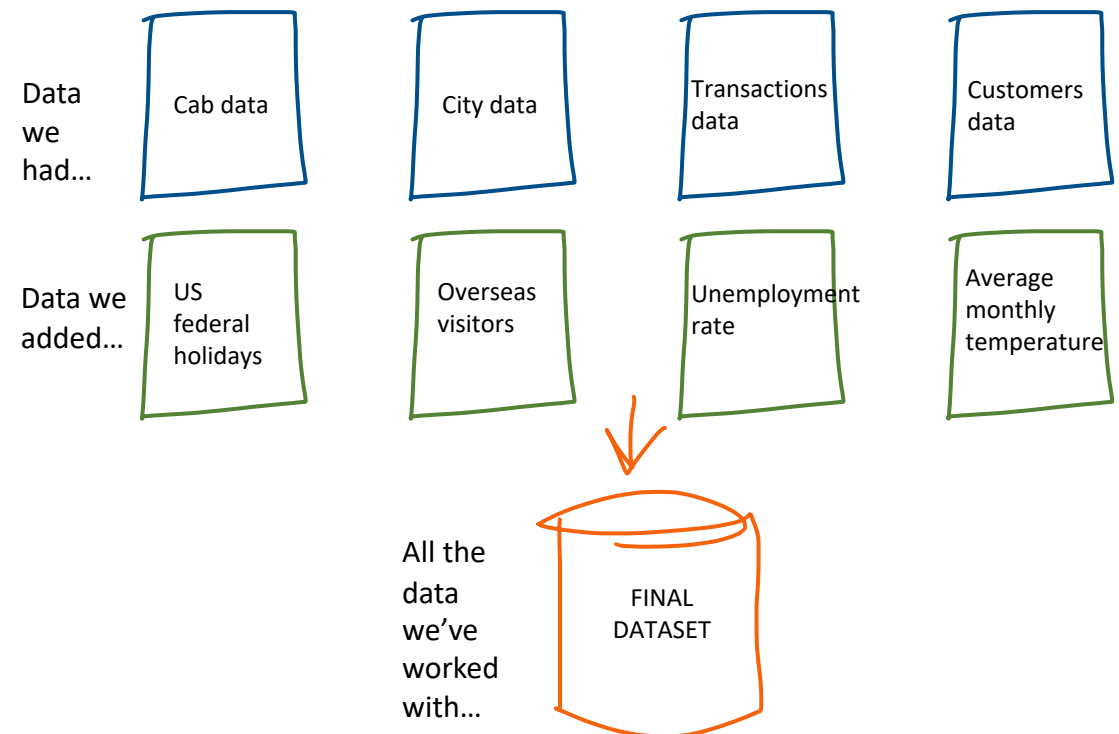
Yellow Cab

Pink Cab

XYZ is a US company interested in investing in the taxicabs market. Which of those two cab companies is a better option?

How are we approaching the question with data?

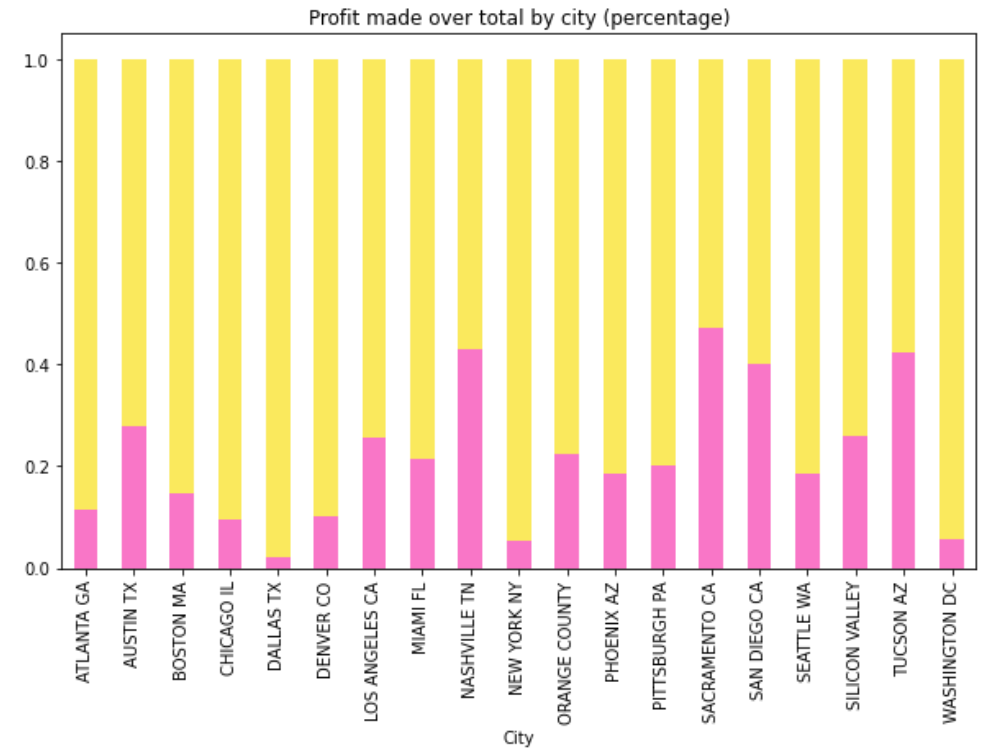
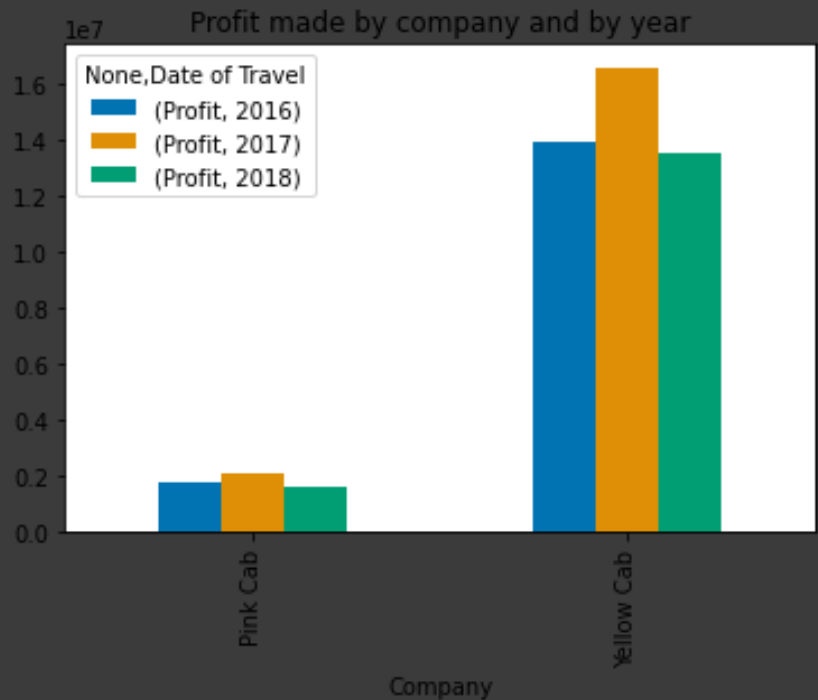
We used the provided data + external data to build a final dataset that allows us to draw interesting conclusions and answer the question.



Exploring the data: first insights

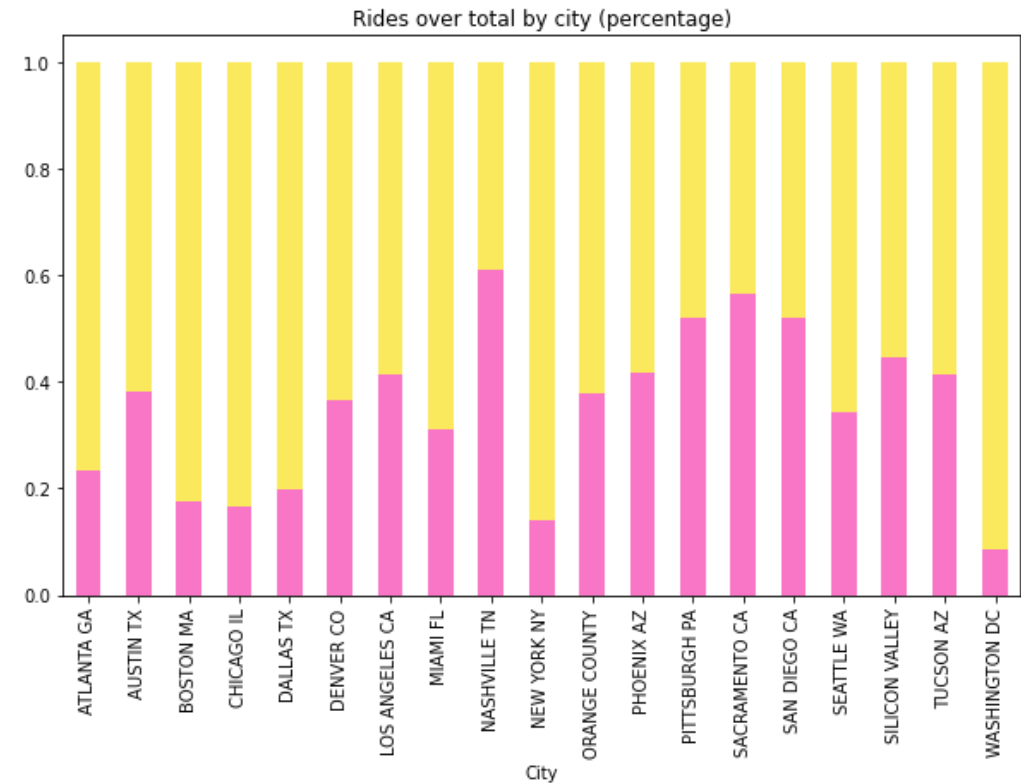
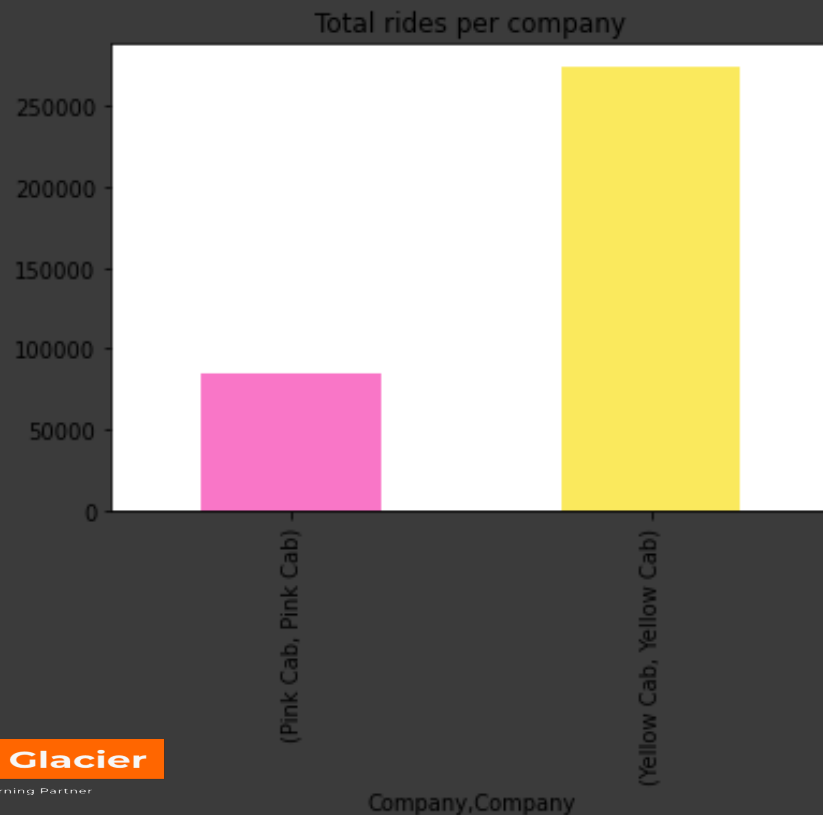
- ✓ 76 % of total rides are run by **Yellow Cab**.
- ✓ 28 % of total rides are run in New York, followed by 16 % in Chicago.
- ✓ 57 % of our sample are men, 43 % are women.
 - ✓ 60 % of total rides were paid by card.
- ✓ The bigger the population, the higher the profit is.
- ✓ Prices tend to be higher in most visited cities. Also means bigger cities.
 - ✓ Unemployment rate has a slightly negative correlation with population and monthly visitors, meaning unemployment tends to be a bit slower in bigger cities.
- ✓ Average monthly temperature tends to be negatively correlated with the number of users in a city.

Exploring the data: Profit made by company



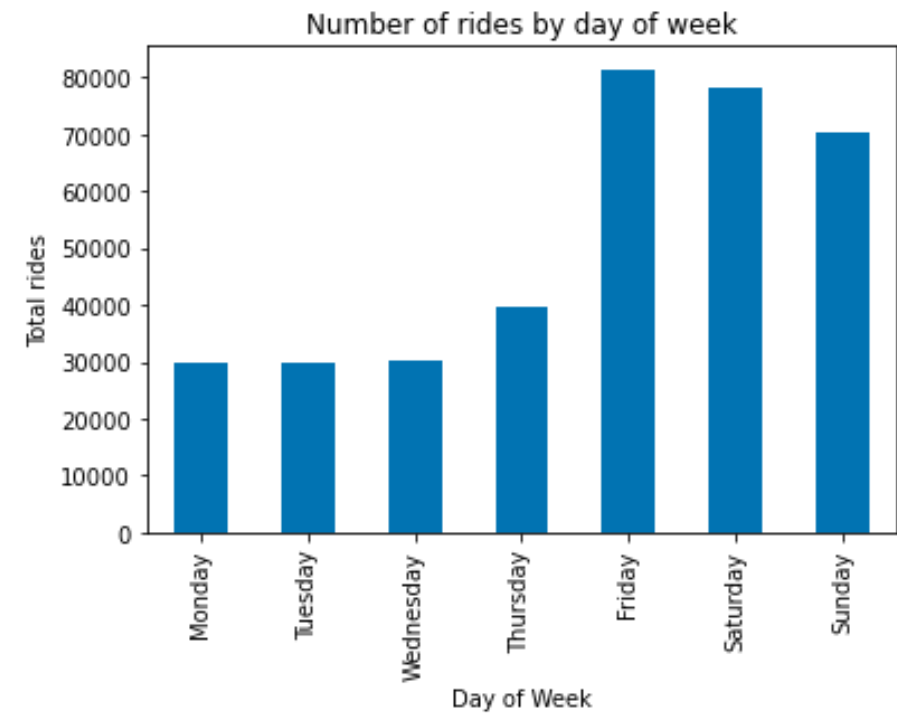
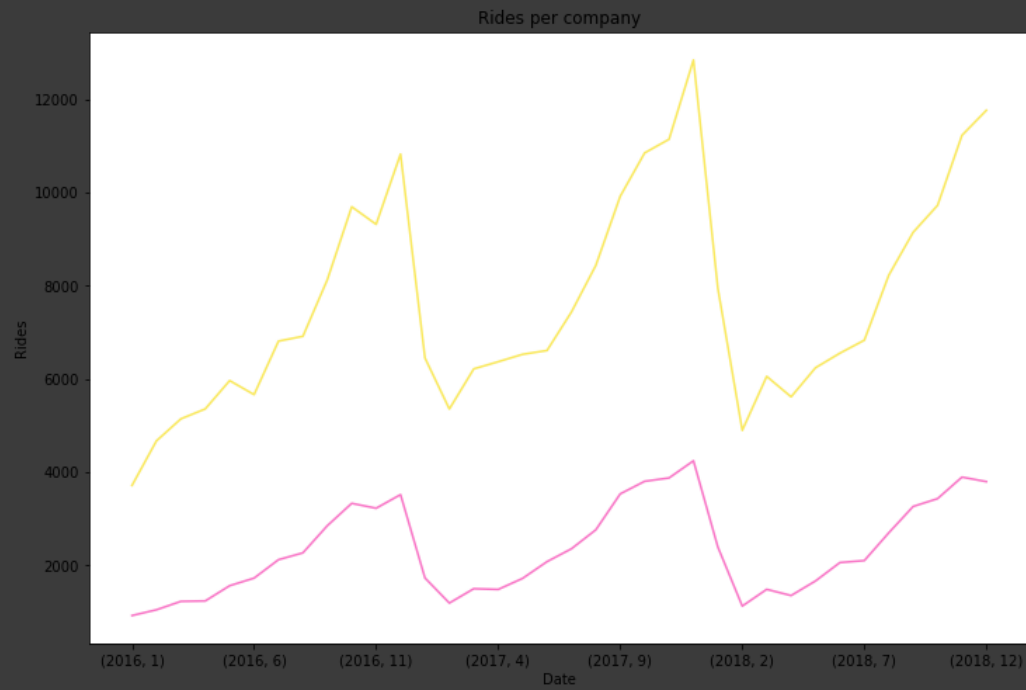
- ✓ **Yellow Cab** makes more profit by far along the three years (left plot).
- ✓ Sacramento, Nashville, Tucson and San Diego have a more balanced profit distribution (around 40-45 % for **Pink cab**).
- ✓ Around 80 % of profit distribution in the rest of cities falls in the **Yellow Cab** side.
- ✓ **Pink Cab** makes 60 USD profit per ride on average, while **Yellow Cab** makes 160 USD profit per ride on average (2.5 times higher).

Exploring the data: Rides run by company



- ✓ Almost 275,000 rides are run by **Yellow Cab**, against nearly 85,000 run by **Pink**.
- ✓ 60 % of rides in Nashville are run by **Pink Cab**, followed by Sacramento, Pittsburgh and San Diego.
- ✓ The rest of cities are dominated by **Yellow Cab** holding between 60-70 % of total rides. In the biggest cities, like NY, Chicago, Boston, Washington it goes up to 80-90 % of rides run.

Exploring the data: Number of rides along time



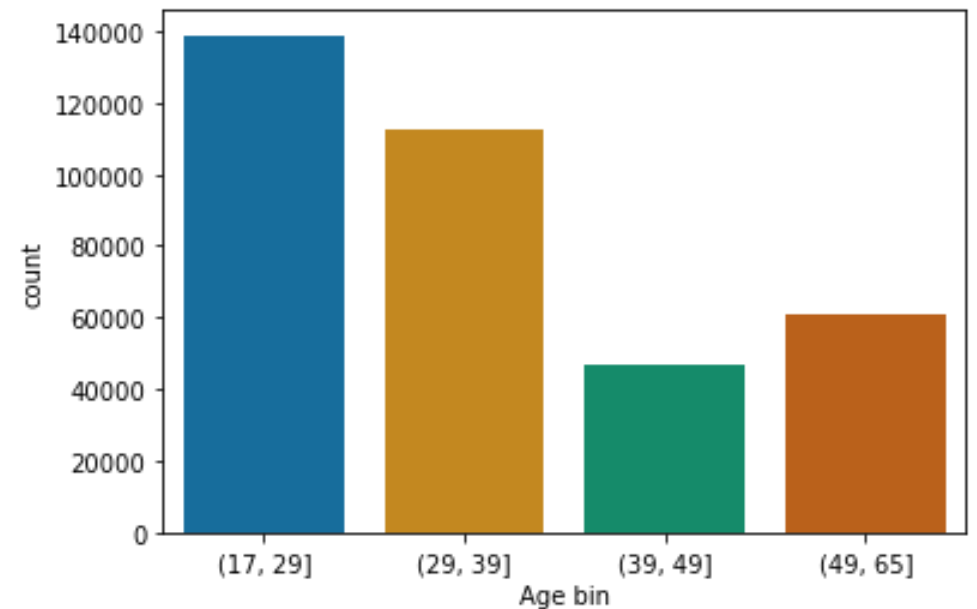
- ✓ It rises around February, shortly after Christmas celebrations. It stabilizes a bit in spring or the beginning of summer and steeply increases from summer until the end of the year. The peak is Christmas and New Year's and sharply decreases after that until February as mentioned before.
- ✓ From Monday to Wednesday, the number of rides remains the same. It starts increasing on Thursdays and Fridays double the number of rides. Fridays are the busiest days. It decreases a little bit on Saturday and then on Sunday but remains much higher than on Mondays.

Exploring the data: Gender differences

- ✓ There are 50,000 more men than women in our sample.
- ✓ Men generate 5 % more profit on average than women.
140.18 USD against 133.32 USD.
- ✓ 60 % of the customers pay via card, without distinction between female and male customers.
- ✓ Slightly more men use **Yellow** than women, 77 % against 75.5 %.
- ✓ The average age is 35 years old.

Exploring the data: Age differences

- ✓ The only difference is that our sample is split in two: people under 40 and older than 40. The former group is more numerous.



Exploring the data: Other details

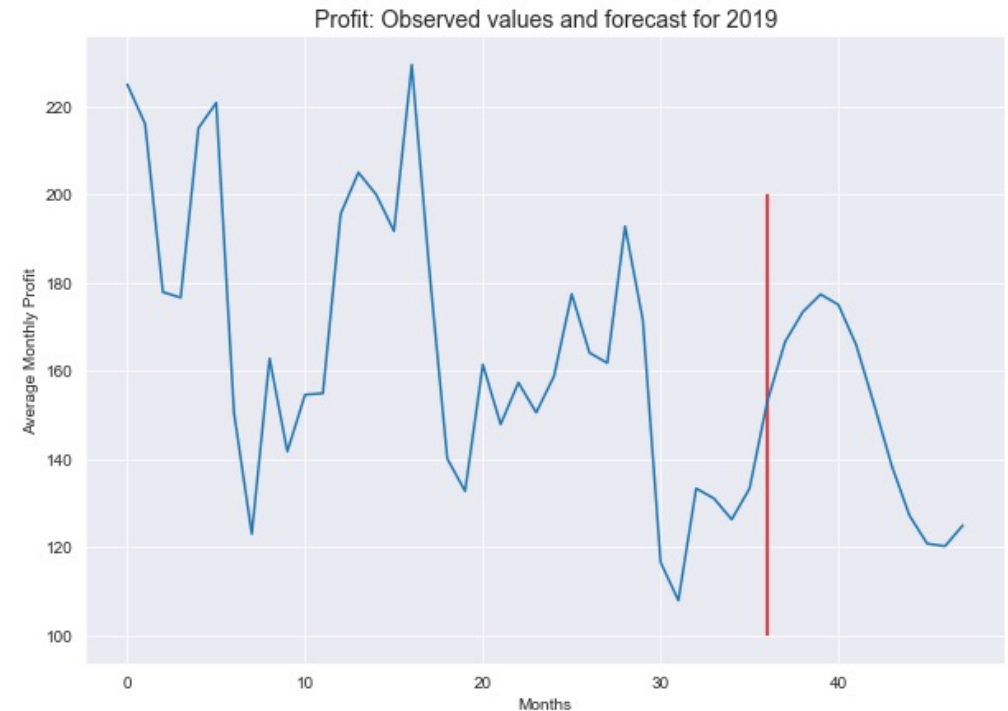
- ✓ There is a positive correlation between monthly visitors and profit and visitors and population. It seems to be related to large-city factor.
- ✓ There is a negative correlation between monthly rides and unemployment rate, maybe due to numerous visitors in bigger cities holding a lower unemployment rate on average.
- ✓ It seems that larger cities tend to have lower temperatures on average.

Yellow Cab

What to expect in
the next future?
Statistics came to
shed a light on us

- ✓ Upward trends: Monthly visitors
- ✓ Downward trends: KM travelled, Price charged, cost of trip, unemployment rate

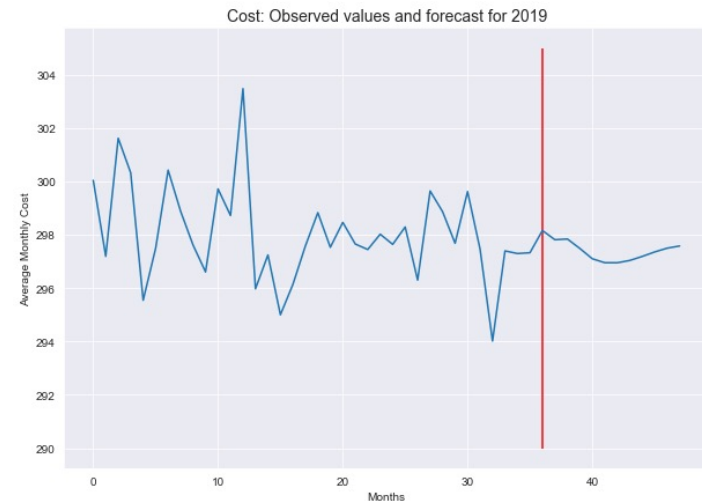
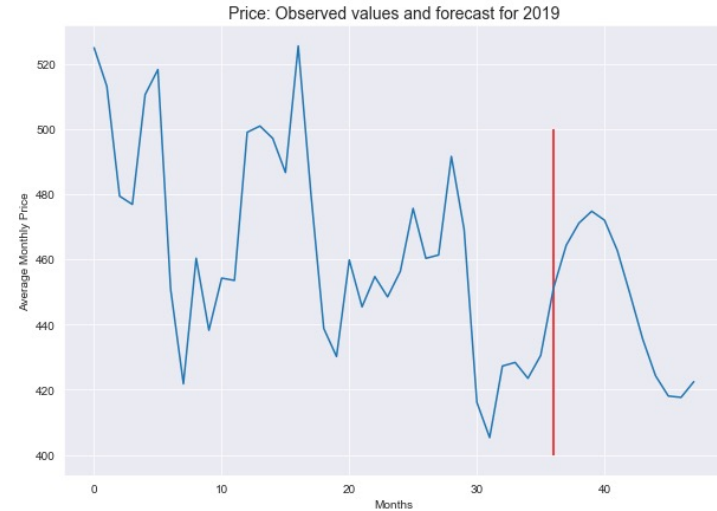
What about profit? It holds a downward trend regarding the average monthly profit made.



Yellow Cab

What to expect in
the next future?
Statistics came to
shed a light on us

What about prices and cost of trips? Prices follow the same oscillation than profit, while cost remains more constant.



The end of the story... What decision should we make?

- ✓ **Yellow Cab** takes the lead almost everywhere and in every examined factor. Typically, the larger the city, the bigger the advantage is.
- ✓ Rides and profit sharply decrease after Christmas and New Year's Eve and start increasing steeply in spring/summer, reaching the peak by the end of the year.
- ✓ Nonetheless, there is a downward trend and so our prediction model expects the same for 2019.
- ✓ **RECOMMENDATION:** given the data analysed and forecasts, we recommend to invest in **Yellow Cab**.

Data Glacier
appreciate your
attention and
trust

Thank You!