G2M Case Study

Cab market exploration

2nd January 2022

Agenda

Executive Summary
Problem Statement
Approach
EDA
EDA Summary
Recommendations

Data Glacier
Your Deep Learning Partner

Summary

- **Problem Statement:** The company XYZ is interested in the Cab Market and is considering investing. They are considering two companies: “Pink Cab” and “Yellow Cab”. Our task is to investigate which is the better option.
- **Approach:** We are provided with a database detailing the combined transactions of both companies. Our approach is to:
  - Clean and unify the database, creating a “master data”.
  - Perform an exploratory analysis on the database and understand it.
  - Come up with hypotheses that are relevant to our problem; test them.
- **Exploratory Analysis:** The companies showed a strong seasonal component in terms of transactions and profit, with an almost flat trend (very stable, predictable). Yellow Cab, however, has greater numbers in total profit, profit per trip, number of transactions. It also dominates in the cities with the greatest population.
- **Recommendation:** Both companies seem like reliable investments, but Yellow Cab is the better option, given the data and the metrics we worked with.

Problem Statement

The situation:

XYZ is interested in investing in the Cab Market as a part of their GTM strategy. They consider two companies: “Pink Cab” and “Yellow Cab”, both operating in the USA.

We have access to transaction information from the last three years.

The objective:

To study and explore the provided dataset in order to obtain relevant information, so XYZ can invest with knowledge of the market.

Preliminary Data description

The files:

- **Cab_Data.csv:** The “main file”, containing information about every transaction. Records the date, city, company, expenses, etc. of each transaction.
- **City.csv:** Contains information about the total population and user count in each city.
- **Customer_ID.csv:** Age, gender and income of the clients.
- **Transaction_ID.csv:** Links every transaction with a client and a paying method.

All the provided files are **joined into master_data.csv**, which contains all the fields from all the listed files.

The fields:

The fields we have found more relevant are:

- **Date of Travel**: The timestamp of a transaction. This field needed formatting, in order to make time series clear.
- **Company**: Necessary to assign transactions to either company.
- **Profit** (per trip)**:** The net gain of each transaction. This is a designed field, calculated by resting **Cost of Travel** to **Price Charged**.

Though not studied directly, other fields have been used to build more complex metrics, such as **Trips per Year and Person**, which is also split into cities and uses information about each city’s population.

Seasonality in profit per transaction

We want to study the profit an average transaction makes. The plot in this slide shows the average profit per trip, averaged over each month. As reference, the amount of transactions in the corresponding months has also been plotted.

Takeaways:

- **Profit per trip decreases when transaction count increases** (law of demand).
- **There is a slight reduction of profit per trip** in the last year. Since the number of transactions is similar, this reduction can also be appreciated in the total profit*.

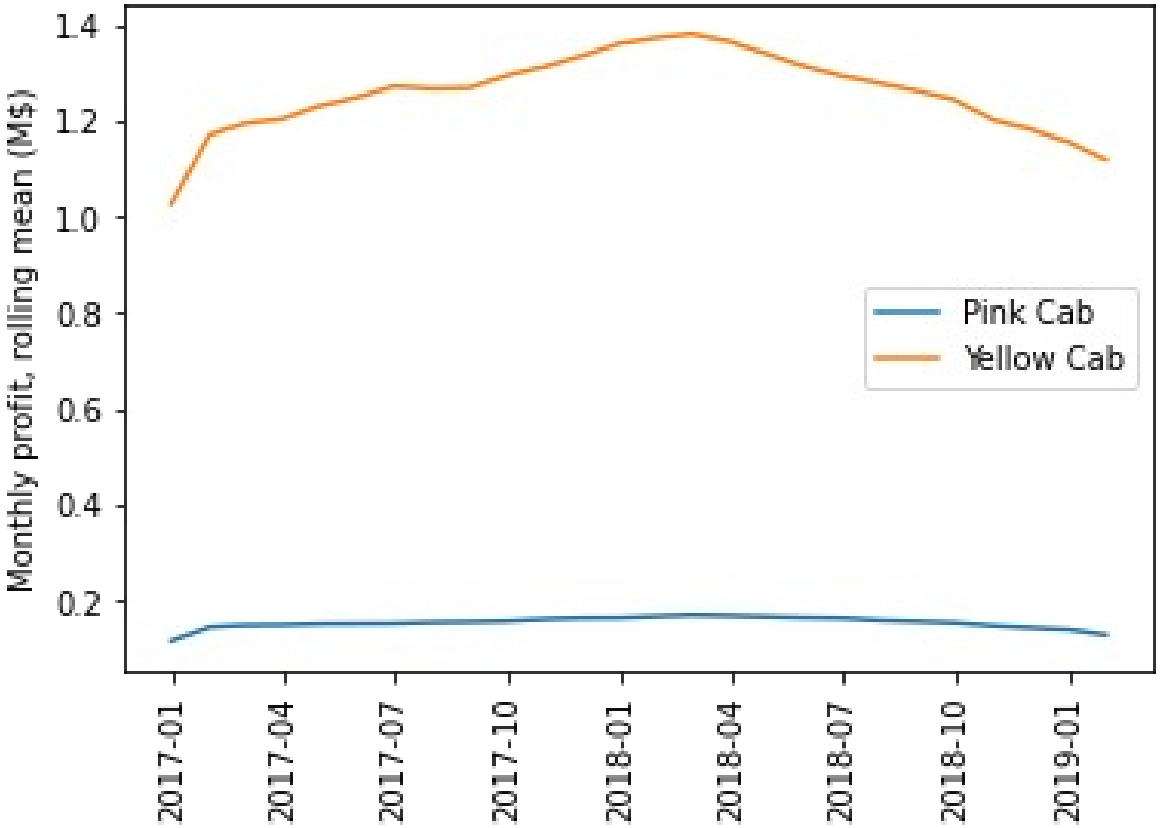
*See next slide

Company Growth

In profit terms, how has each company been growing?
As seen in previous plots, the main component in the transaction and profit series is the seasonal, which obscures the trend behind it.
To make it more apparent, I have taken a rolling mean of period=12 months, which “cancels” the seasonal component.

Takeaways:

- **Both companies’ trends are quite flat.** Pink Cab’s trend is incredibly flat, almost an horizontal line. Yellow Cab’s trend has a bit more of variance, decreasing slightly between April 2018 and January 2019. Anyhow, there is too few points to determine if it is a genuine fall or a yearly fluctuation.

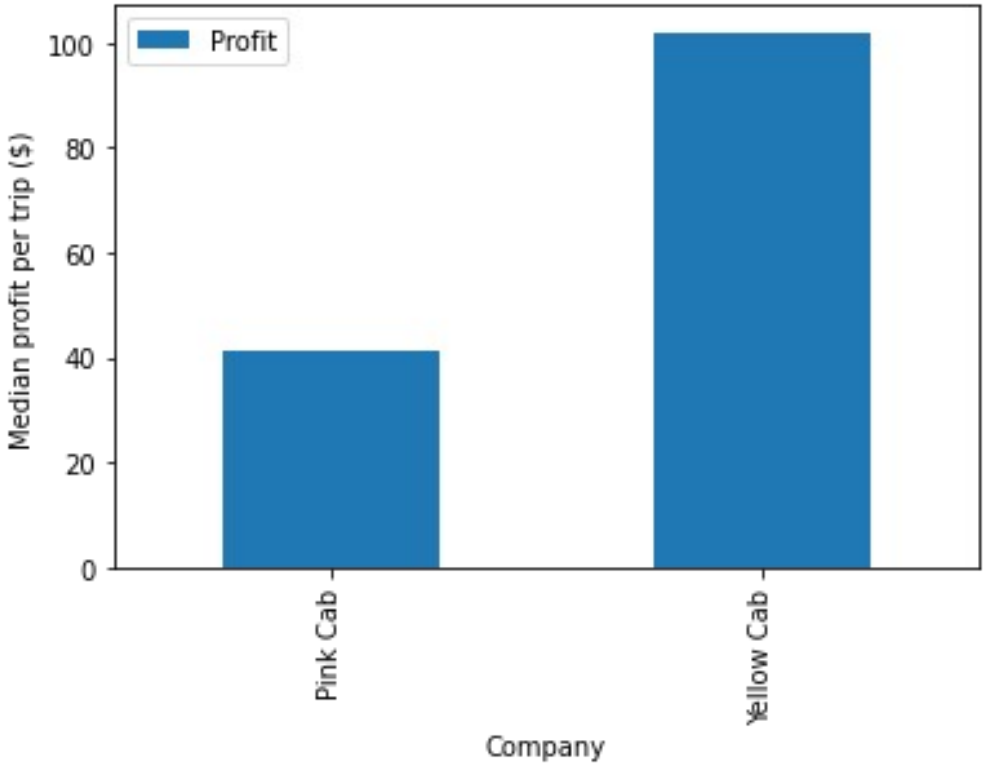


Company Profit per Trip

What is the profit a trip generates, if we consider Pink Cab and Yellow Cab separately? The bar plot answers that question. We can see that Yellow Cab makes more than twice as much money as Pink Cab.

Takeaways:

- **Yellow Cab is more stable.** Consider an increase of 20$ in cost per trip, which could be due increase in gas price, taxes or something unexpected. Pink Cab would lose ~50% of its profit, while Yellow Cab would only lose ~20%.
- **Pink Cab could be more profitable**. In the same fashion, an increase of 20$ in the profit per trip would mean 50% growth to Pink Cab, while only a 20% growth to Yellow Cab.

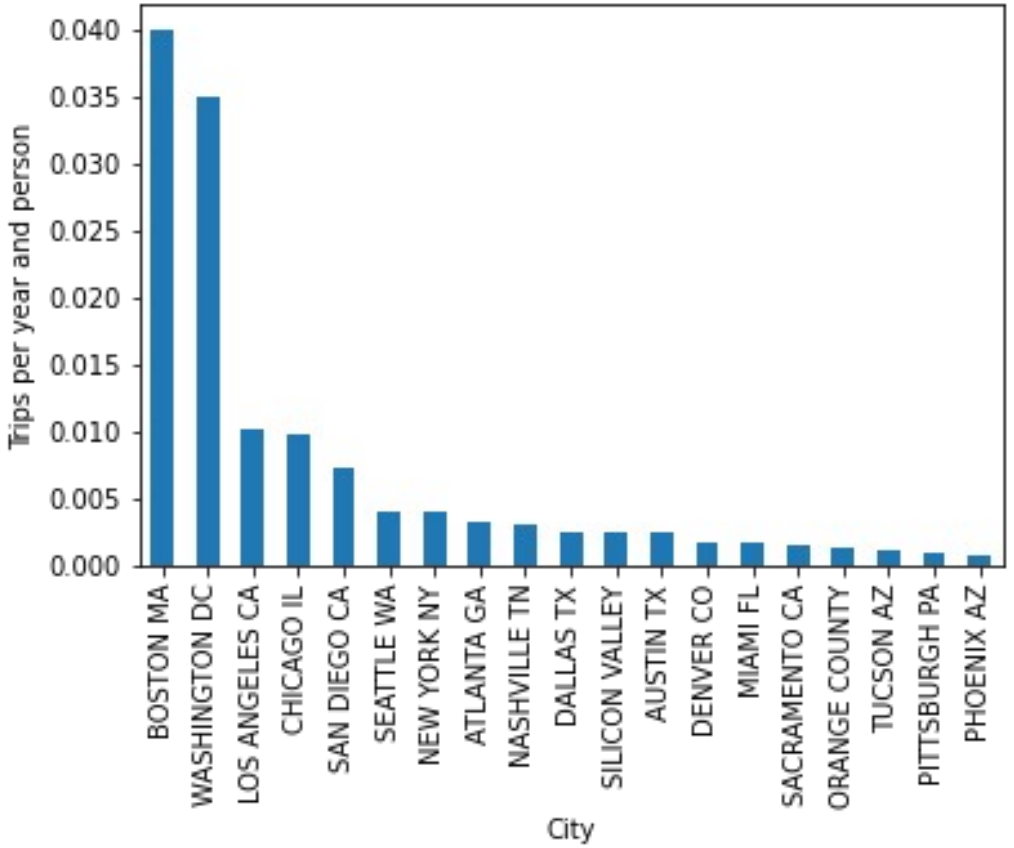


Trip count in relation to population

The plot in the right shows the amount of trips the average person takes in a year, split by city.

Takeaways:

- **The absolute numbers are low:** Consider Boston, the city with most trips per person and year, with a value of approx. 0.04. This means only one person in 25 takes a single trip in a year.
- **Boston and Washington tower over the other cities**: Even though their numbers are still low, they are more than thrice those of the other cities. This could be signaling a predisposition of those cities, e.g. if they have a poor public transport system.

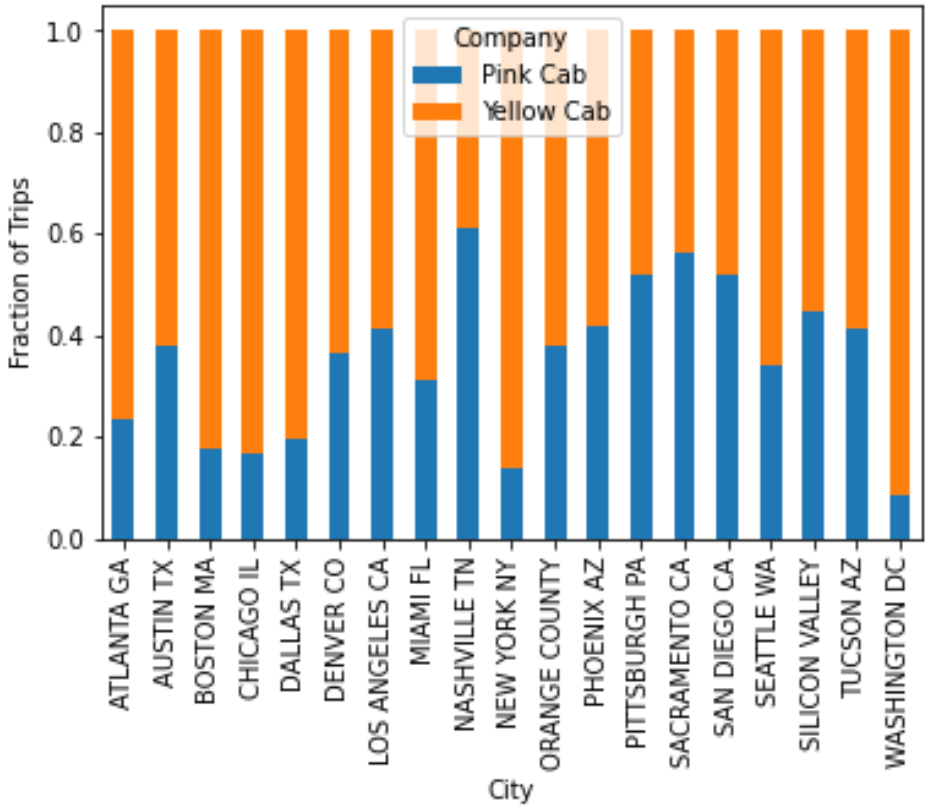


Company presence along cities

The plot at the right shows the fraction of trips that each company provides in each city. As expected from the profit analysis (see slide number 6), Yellow Cab has a greater share in most cities.

Takeaways:

- **Yellow Cab dominates in the most populated cities:** The company provides more than 80% of the trips in New York and Chicago, and a majority in LA, though not for much.
- **Pink Cab's best cities are among the least profitable**: The cities with a greatest Pink Cab trip fraction are Nashville, Sacramento and Pittsburgh. The population of those cities is under a million, and their trip per person rate is also among the lowest (see previous slide). They do have an approx. 50% share in San Diego, though, whose population is over a million and has a decent rate.



Recommendations and conclusion

In broad terms, the previous slides have shown that:

- Yellow cab has a **higher trip count**, a **higher profit per trip** and a **higher total profit** than Pink Cab.
- Yellow cab has a **stronger presence in profitable cities**.
- Excluding seasonality, **both companies are stable**, i.e. their trends are quite flat.

For the given data and with these considerations **Yellow Cab is a better option** than Pink Cab.

If an investment is made on Yellow Cab, the trend of its profits should be observed regularly; in case the slightly decaying trend of last year’s profit signals a genuine decay and not a yearly fluctuation.

Thank You