




G2M Insight For Cab Investment Firm



Submitted By: Farheen Fatima
Date : 14-Aug-2022



Agenda

- Problem Statement
- Dataset
- EDA
- Hypothesis and Testing
- Conclusion

Problem Statement

- XYZ is a private equity firm in U.S. Due to remarkable growth in the cab industry in last few years and multiple key players in the market, it is planning for an investment in the cab industry.
- Objective : To provide actionable insights to help XYZ firm in identifying the right company for making investment.
- Cab companies are i) Pink Cab
II) Yellow Cab.

Exploratory Data Analysis

To perform the analysis, following steps has been taken:

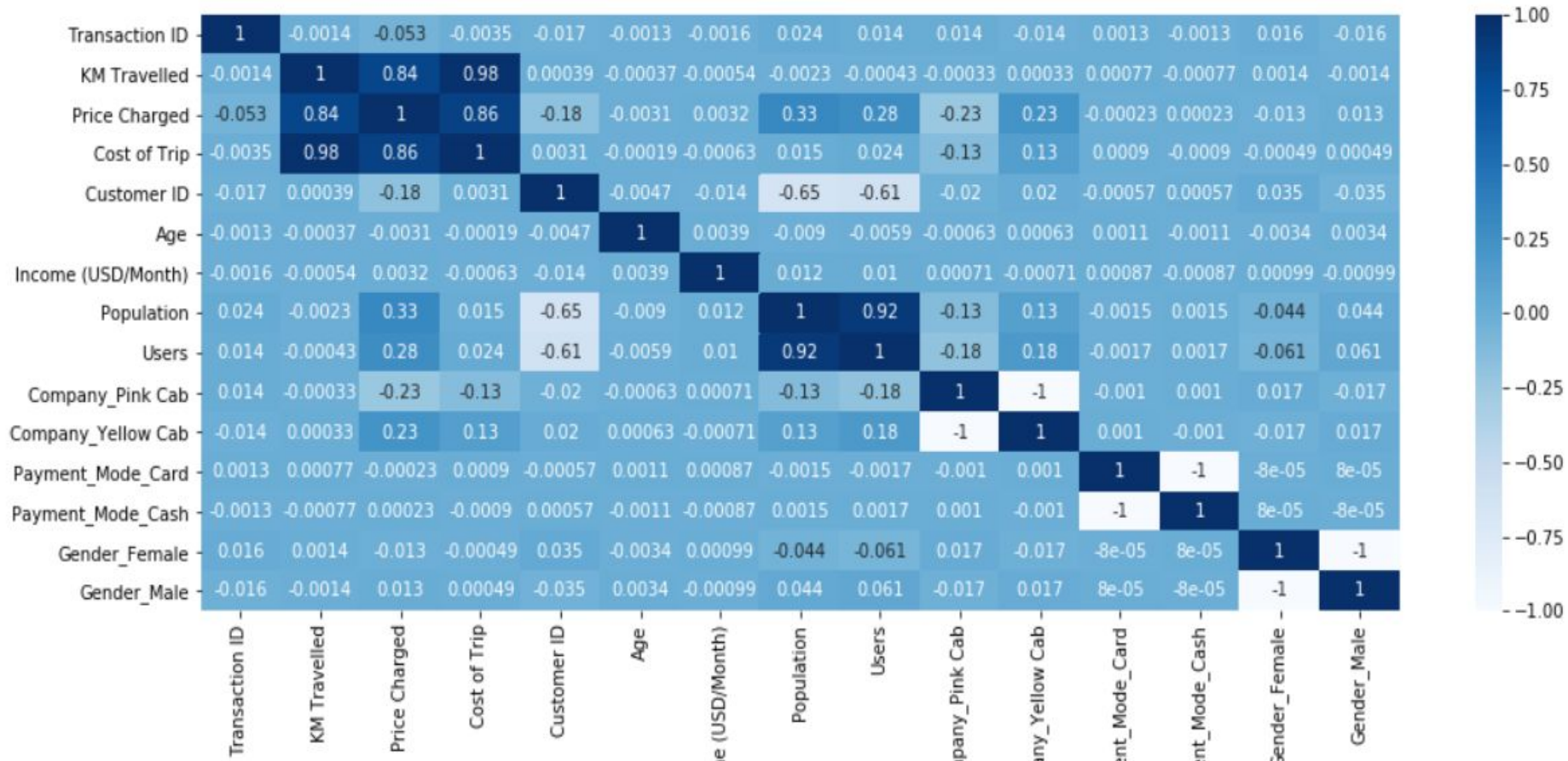
- Merged all four datasets
- Checked for the null values - null values not found in any of the dataset.
- Converted categorical features into numerical features.
- Calculated the correlation to separate the irrelevant features from the dataset.

Dataset

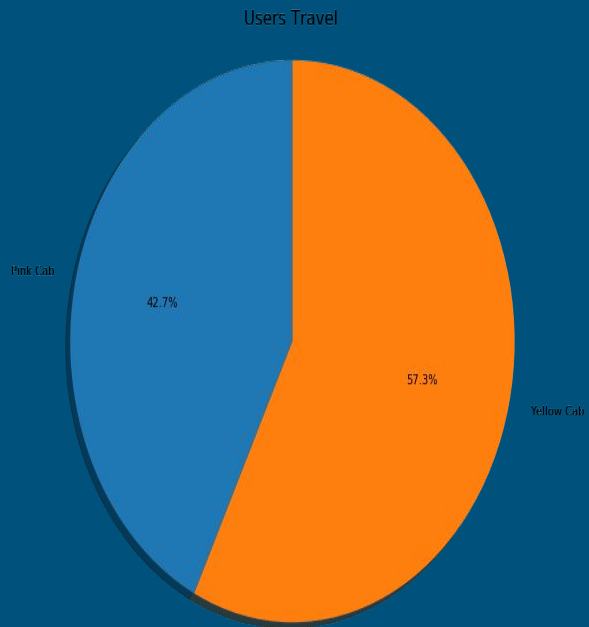
There are four datasets:

1. Cab_Data.csv - this file indicates details of transaction for 2 cab companies
2. Customer_ID.csv - this is a mapping table that contains a unique identifier which links the customer's demographic details
3. Transaction_ID.csv - this is a mapping table that contains transaction to customer mapping and payment mode.
4. City.csv - this file contains list of US cities, their population and number of cab users.

HeatMap

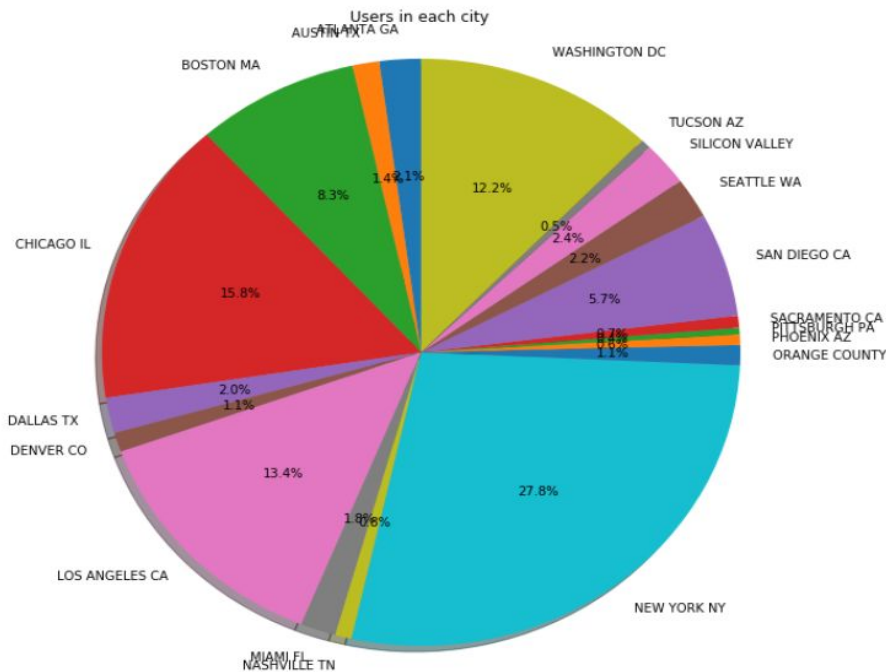


Users by Company



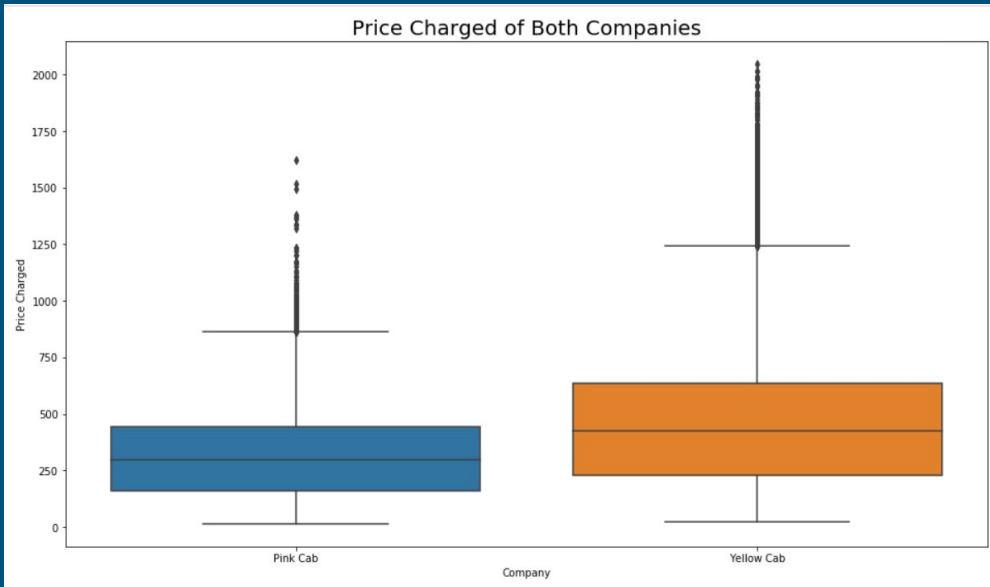
Yellow Cab has more users than pink cab company.

Users in each city



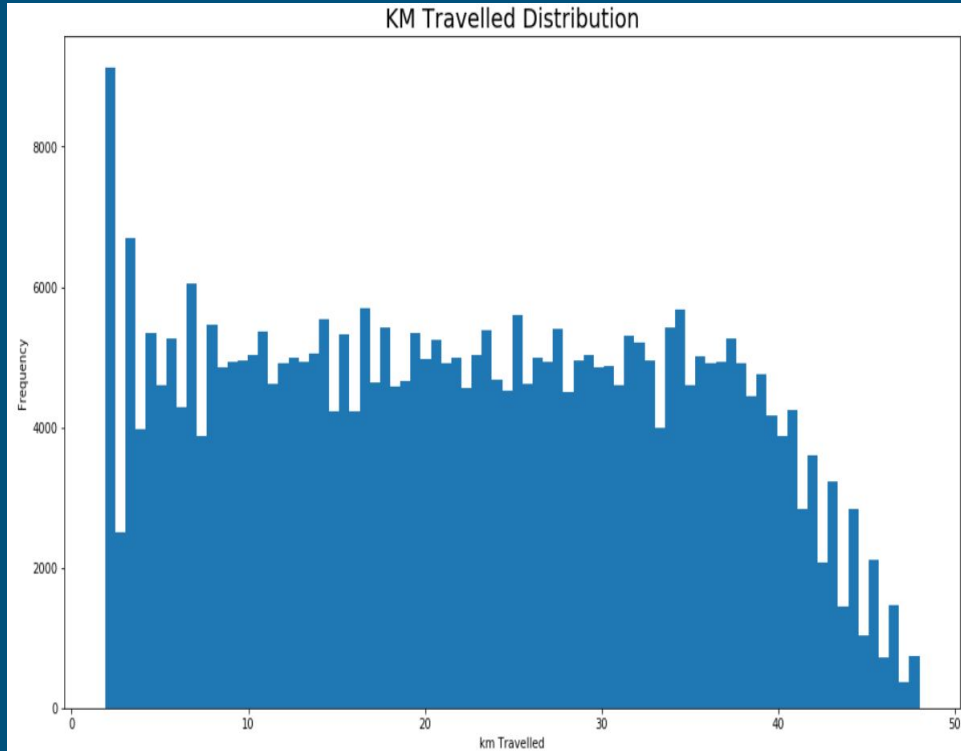
- New York city has the highest cab users with 28%.
- Chicago has the second highest cab users with 15.8%

Price charged for both companies



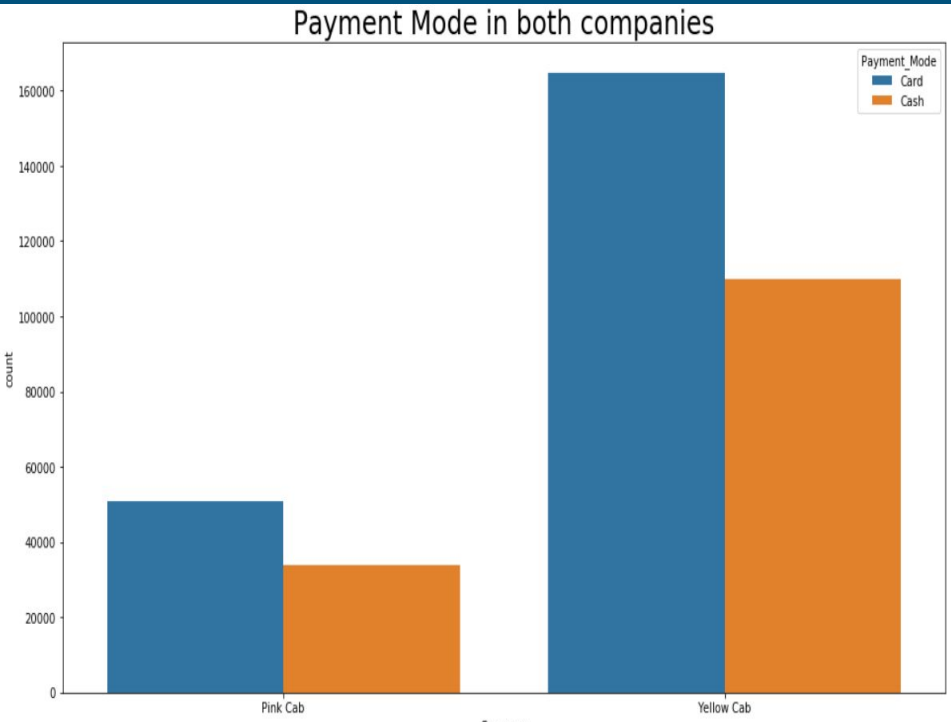
Yellow cab is expensive than Pink cab.

KM Travelled



Distance travelled by both the companies lies in the range 2 KM to 48 KM.

Payment Mode



- Most users preferred card payment in both the companies.

Hypothesis Testing

Hypothesis 1: Difference in Margin by Gender

- P values is 0.1151530
 - We accept null Hypothesis (H_0) that there is no difference regarding gender for Pink Cab.
-
- P value is 6.0604
 - We accept alternative Hypothesis(H_1) that there is a difference regarding gender for Yellow cab.

Hypothesis 2: Difference in Margin by Age

- P value is 0.79004
 - We accept null Hypothesis (H_0) as there is no difference regarding age for Pink cab.
-
- P value is 6.328
 - We accept alternative Hypothesis(H_1) as there is a difference regarding age for Yellow Cab.

Hypothesis 3: Difference in Margins by Payment Mode

- P value is 0.79004
 - We accept null Hypothesis (H_0) as there is no difference in payment mode for Pink cab.
-
- P value is 0.2933
 - We accept null hypothesis (H_0) as there is no difference in Payment mode for Yellow cab.

Conclusion

Based on the above analysis and hypothesis results:

- Yellow cab company is better than Pink cab company for the following reasons:
 - ❑ More users
 - ❑ Profit Margin

Thank you!

