Exploratory Data Analysis

G2M Insight For Cab Investment Firm

Submitted by: Chenyu Wang

21-Sep-2022

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Problem Statement

- **Executive Summary:** XYZ is a private firm in US. 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 Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

- **Problem Statement:** XYZ is firm, want to understand the market in order to invest in Cap industry as per G2M Strategy before taking the final decision.

- **Objectives:** To Provide actionable insights that help them identify the right company to make their investment and understand the Cab Industry market in US.

- **Approach** :

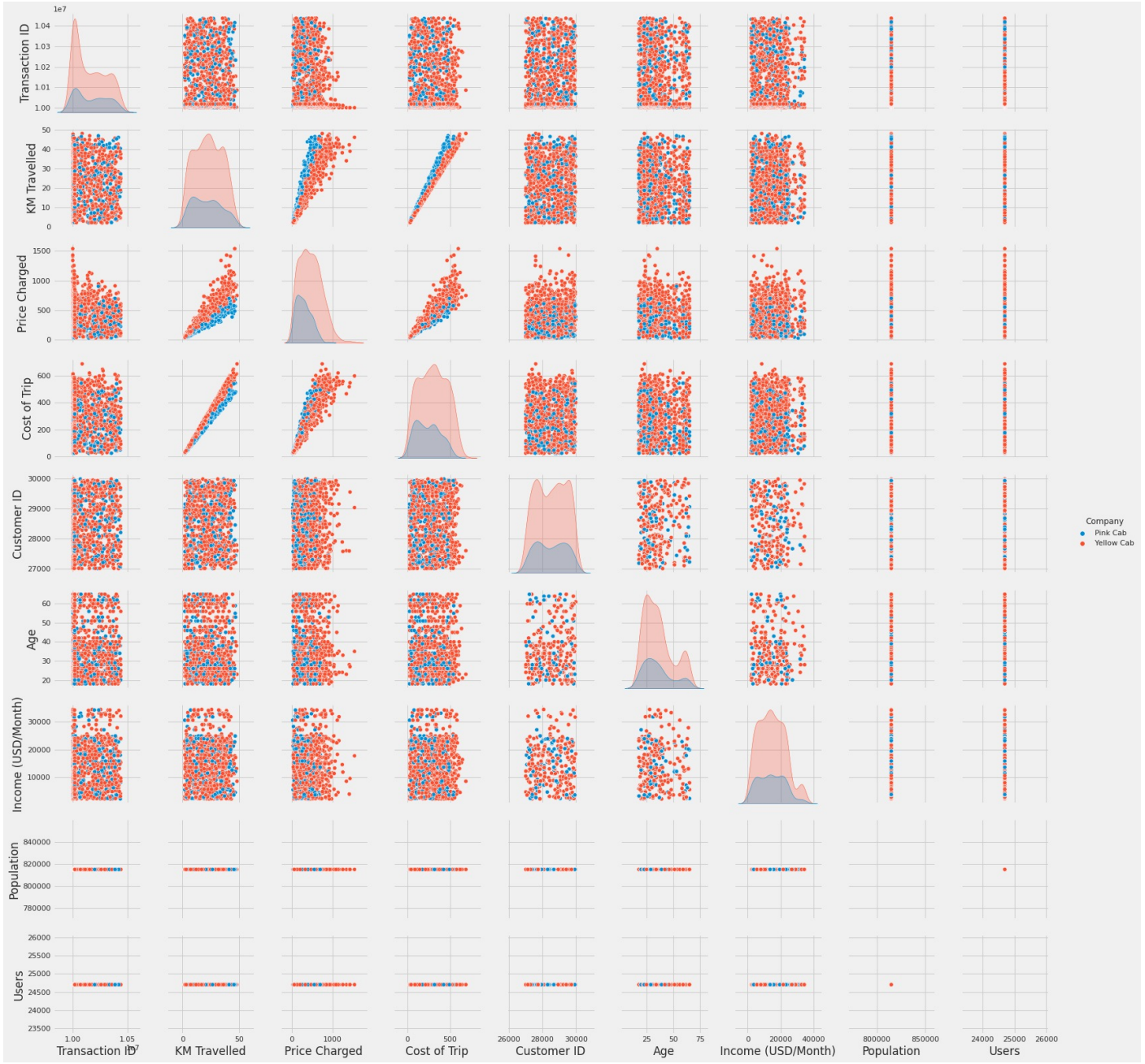
  1. Review the Source Documentation.
  2. Understanding data (the columns and data types).
  3. Identify relationships across the data files.
  4. Cleaning & Preparing data.
  5. Features Analysis.
  6. Summary.

Data Information

There have been provided 4 individual data sets. Time period of data is from 31/01/2016 to 31/12/2018.Below are the list of datasets which are provided for the analysis:

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

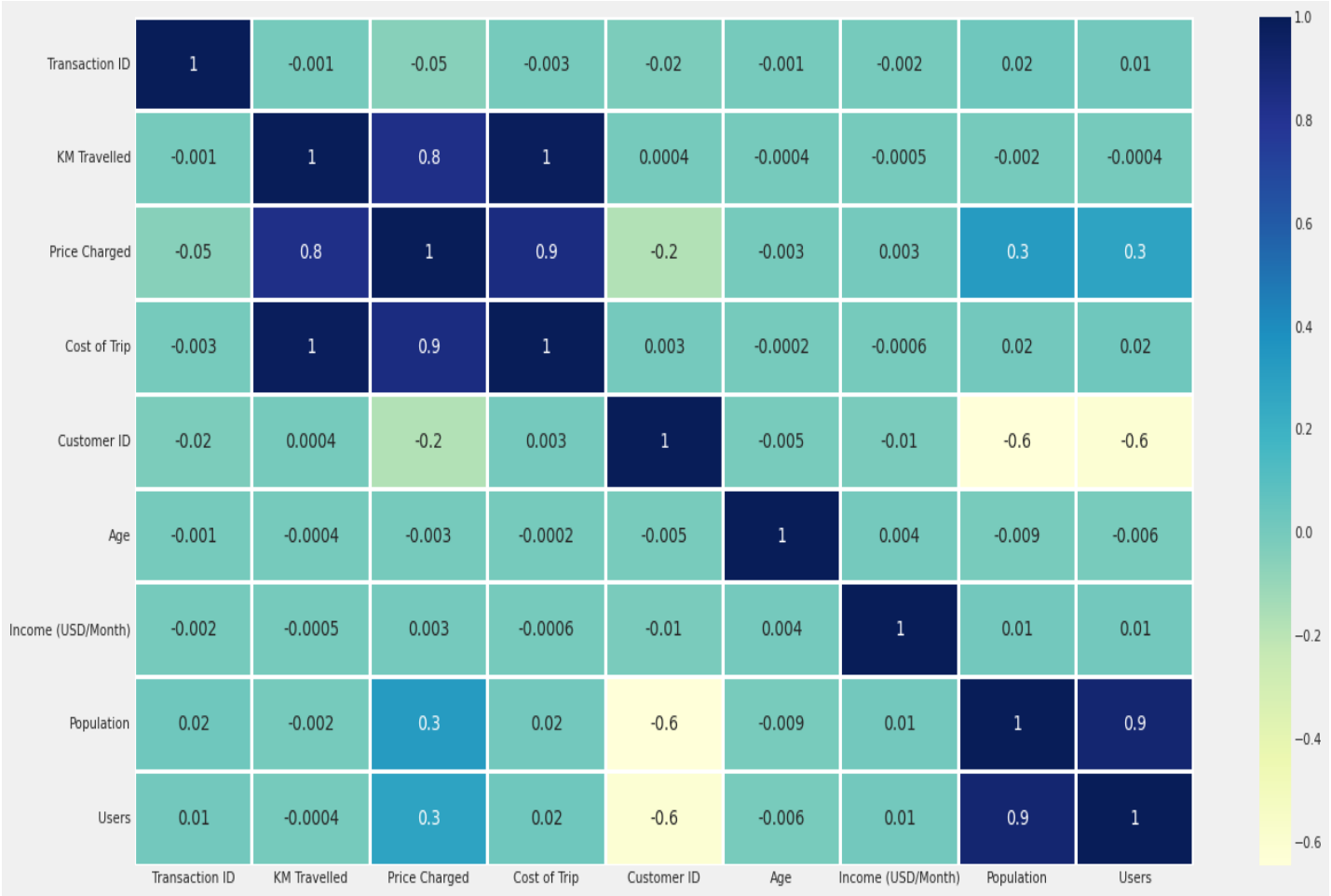
Relationships Between Variables



Correlation

Strong Correlation between

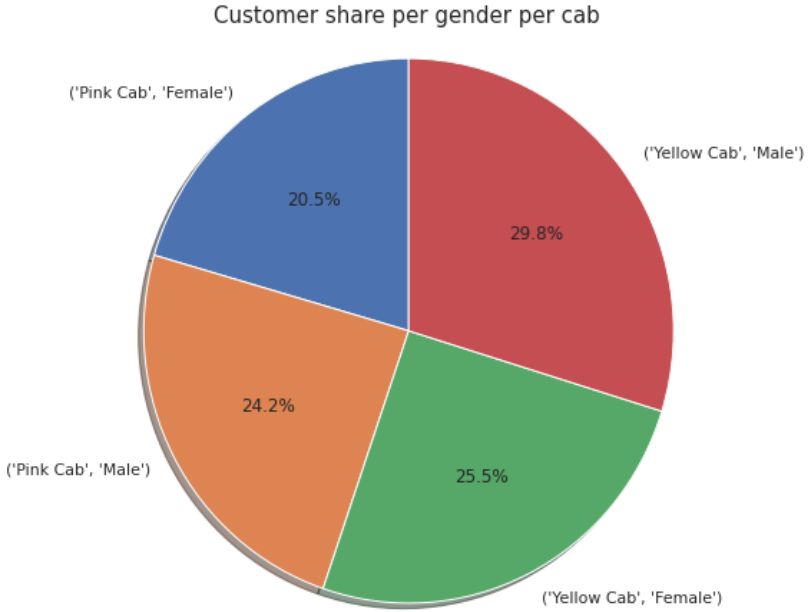
- Population & Users
- Price Charged & Cost of Trip & KM Travelled

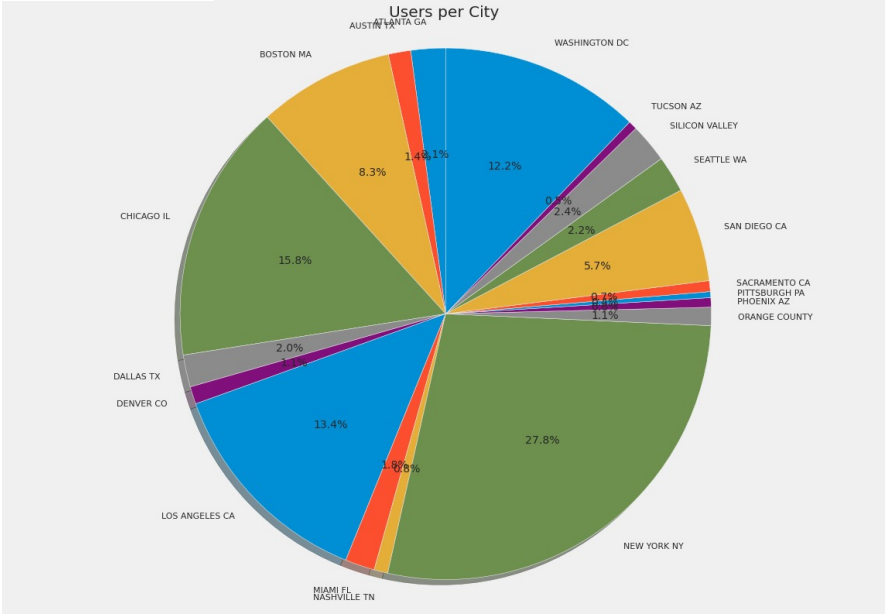


Data Glacier
Your Deep Learning Partner

User Analysis

- Users prefer to ride on Yellow Cab
- Male users prefer to travel in Cab
- New York City has the highest Cab users with 28%
- Chicago has Cab users with 16%

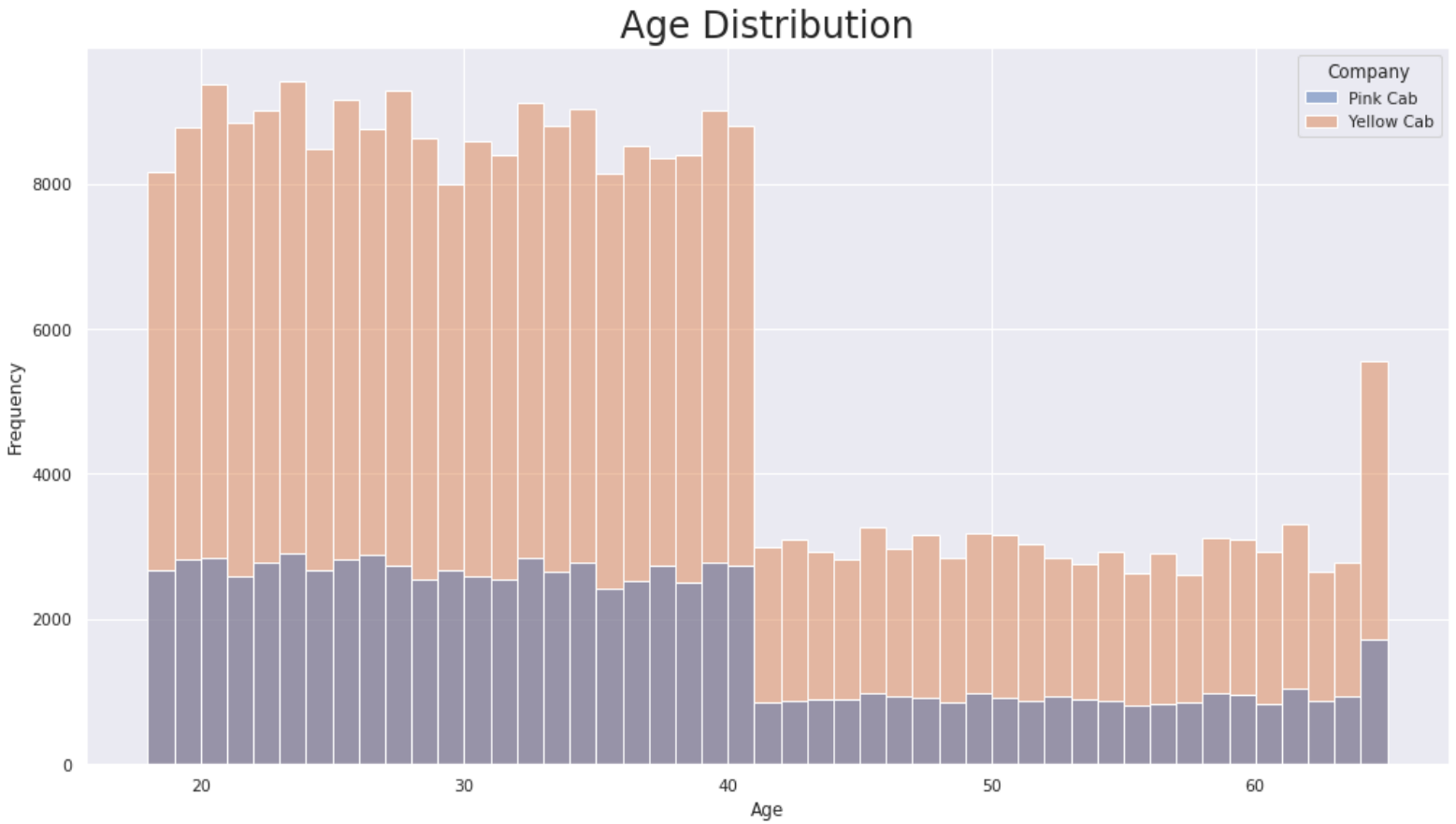




Data Glacier
Your Deep Learning Partner

User Analysis

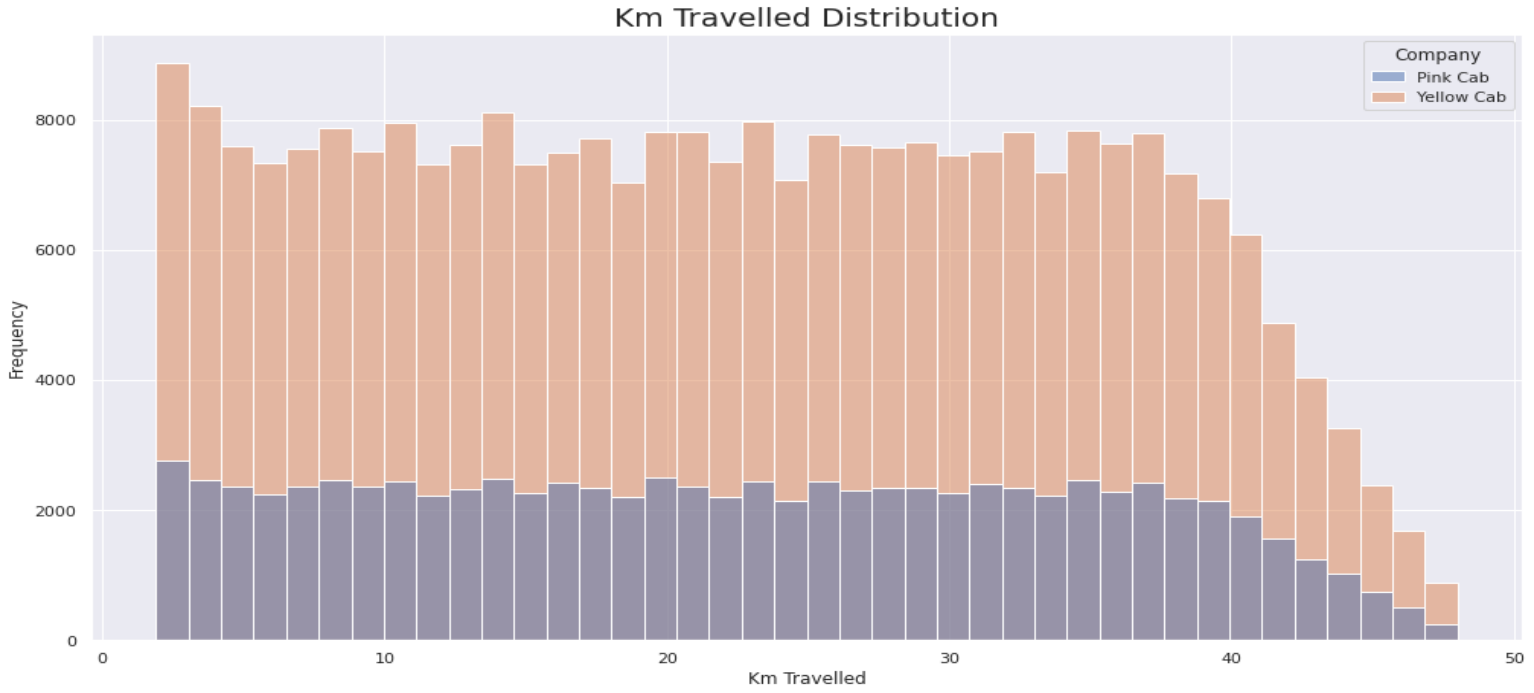
- Users under 40 years old prefer more to travel in Cab

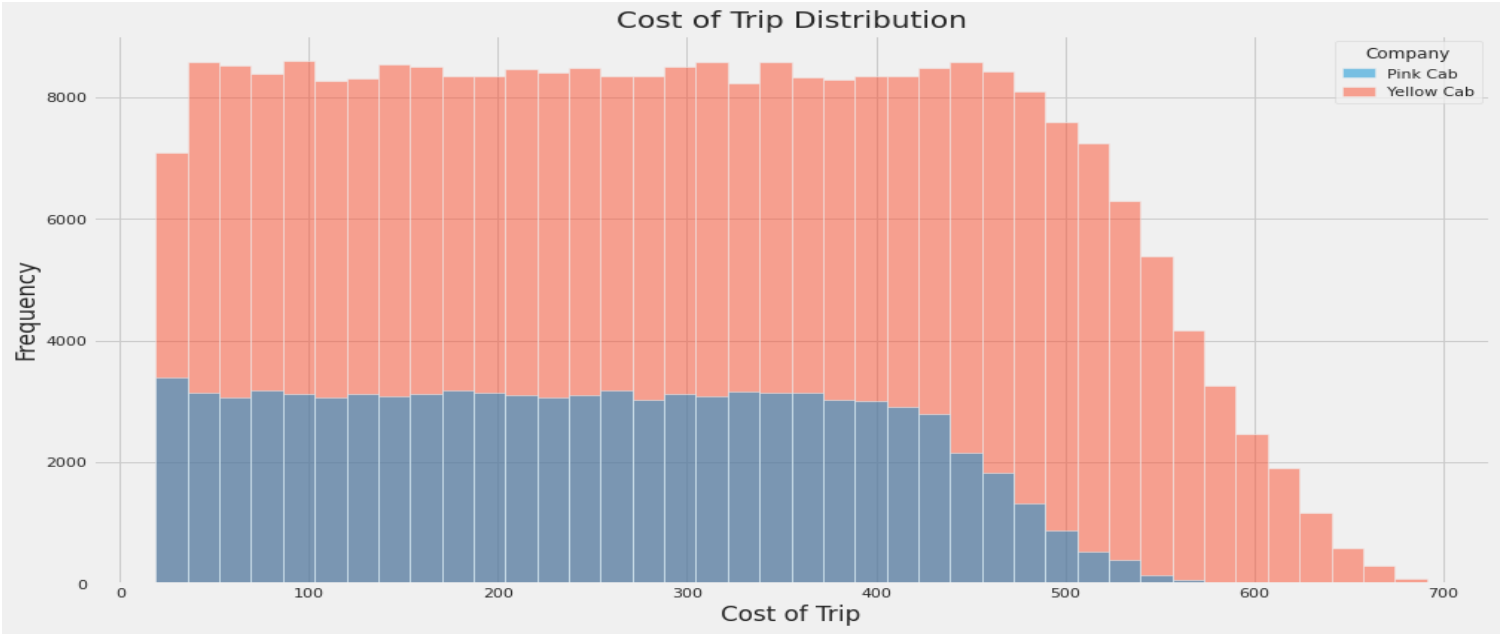


Data Glacier
Your Deep Learning Partner

Travel Distribution Analysis

- Most of Trip KM varies from 2 to 40 KM
- Trip KM of Yellow Cab is more than Pink Cab
- Trip cost of Yellow Cab is higher than Pink Cab

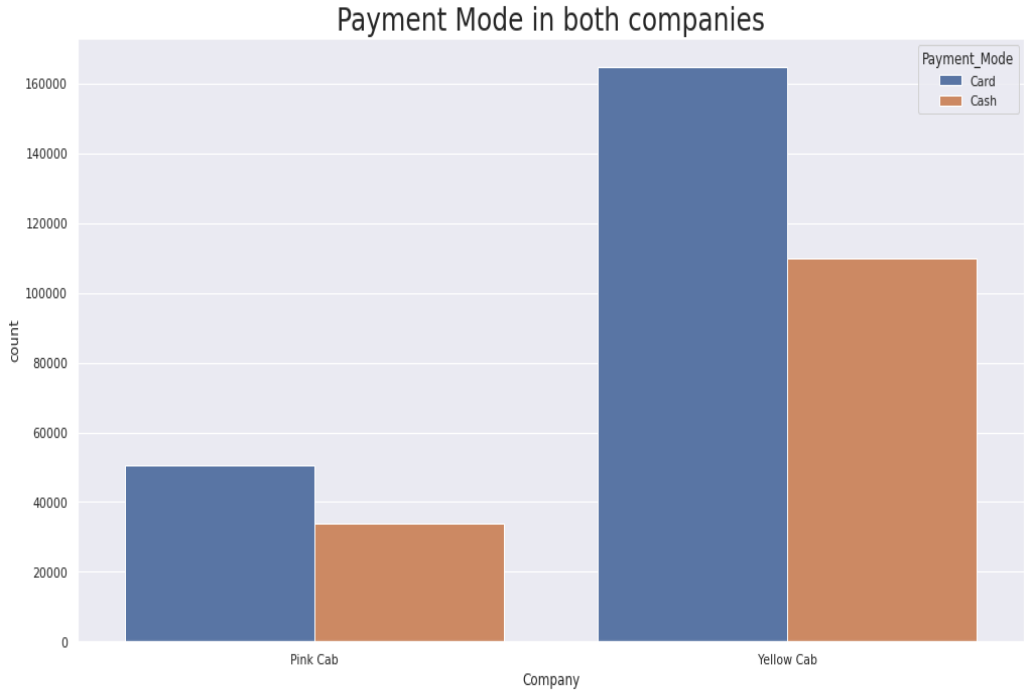




Data Glacier
Your Deep Learning Partner

Transaction Analysis

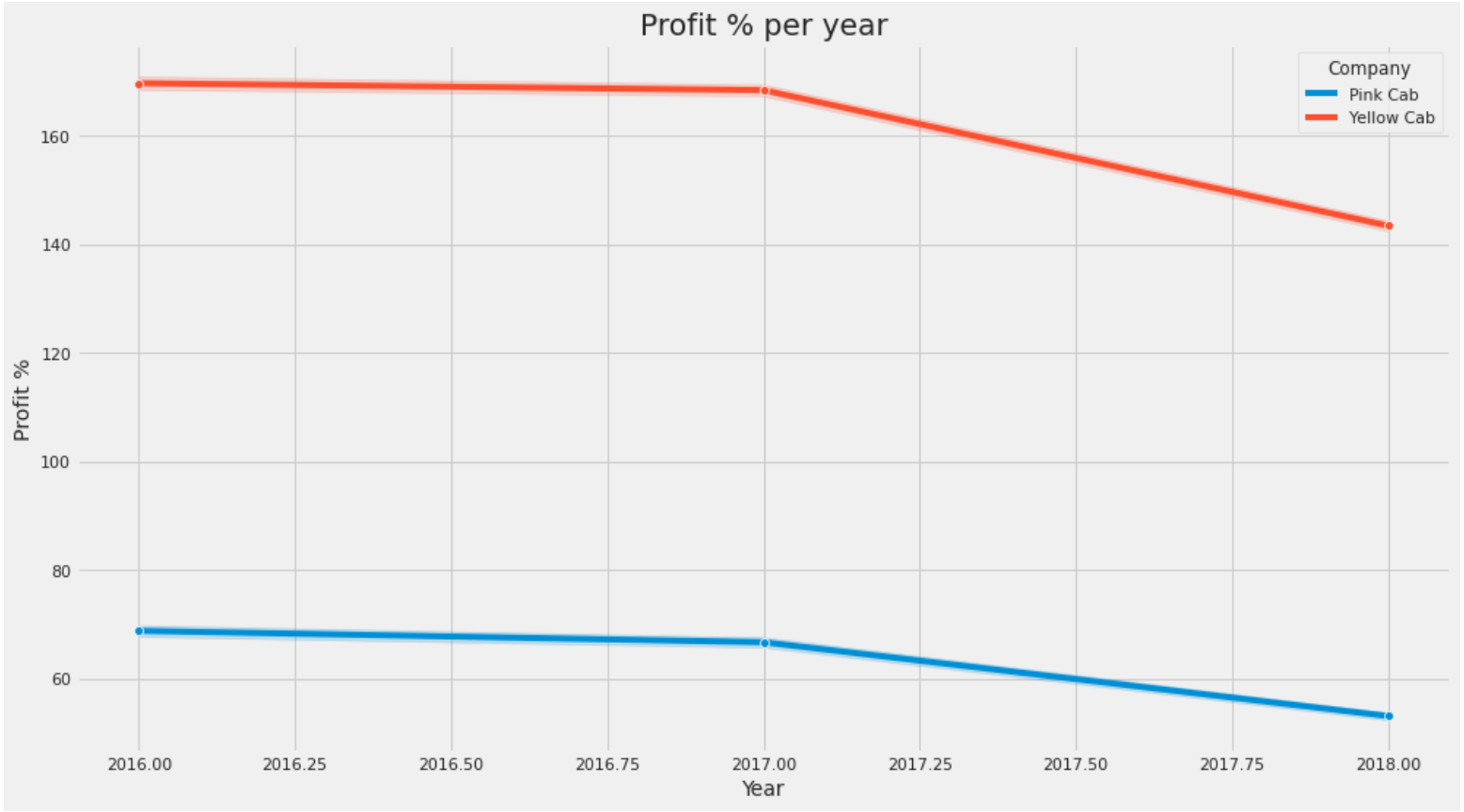
- Users prefer to use Card as Payment method
- The Avg Price Charged for Yellow Cab is higher than Pink Cab



Data Glacier
Your Deep Learning Partner

Profit Analysis

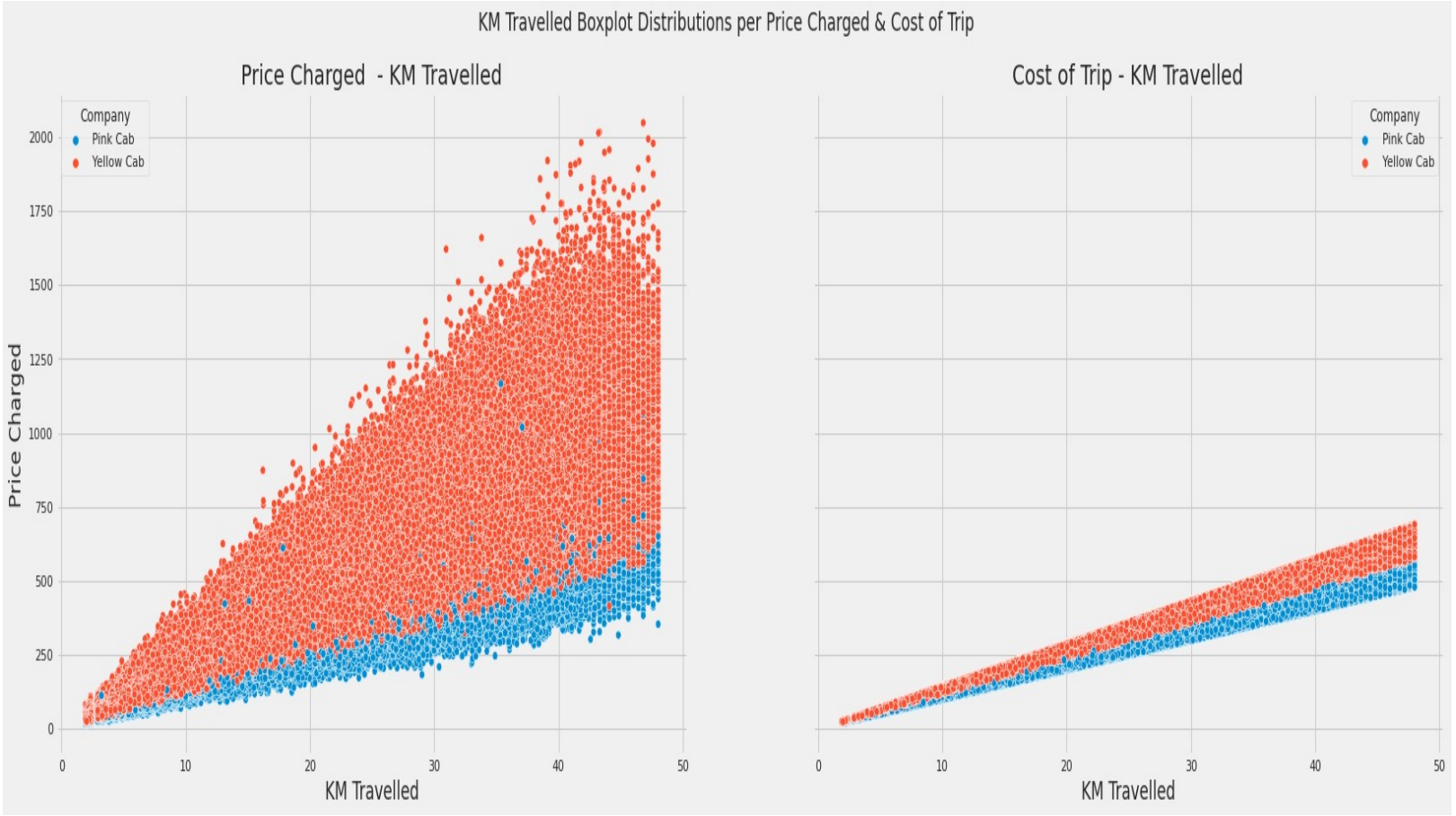
- Both companies' profit decreased relative to last year
- The profit of Yellow Cab is higher than Pink Cab’s profit



Data Glacier
Your Deep Learning Partner

Profit Analysis

- A linear relationship between KM travelled and Price Charged
- A linear relationship between KM travelled and Cost of Trip
- Yellow Cab has high price charges and cost compared to Pink.



Data Glacier
Your Deep Learning Partner

Recommendations

- From the above analysis, it seems that Yellow Cab is better than Pink Cab.
- Each variable distribution of Yellow Cab is much higher than Pink Cab.
- Yellow Cab Firm has more users than Pink Cab Firm in market Share.
- Yellow Cab Firm has higher profit than Pink Cab.
- Yellow Cab has more transactions than Pink Cab in market Share.
- the Yellow Cab Firm is highly recommended.

Exploratory Data Analysis

Data Glacier

Your Deep Learning Partner

Thank You