G2M insight for Cab Investment firm

Company Name : XYZ
Location: UAE
Team: Data Science
Date: 23-March-2022

Agenda

- Executive Summary
- Problem Statement
- Approach
- EDA
- Building Models
- Recommendations

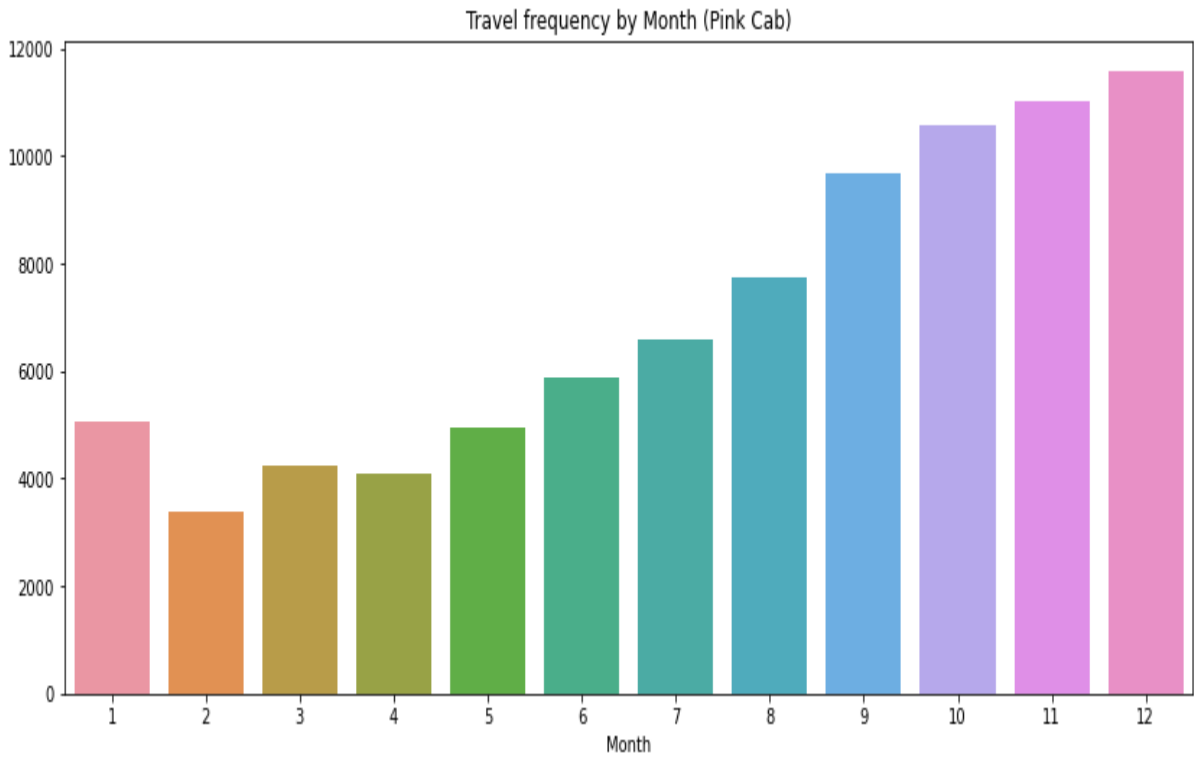
Description:

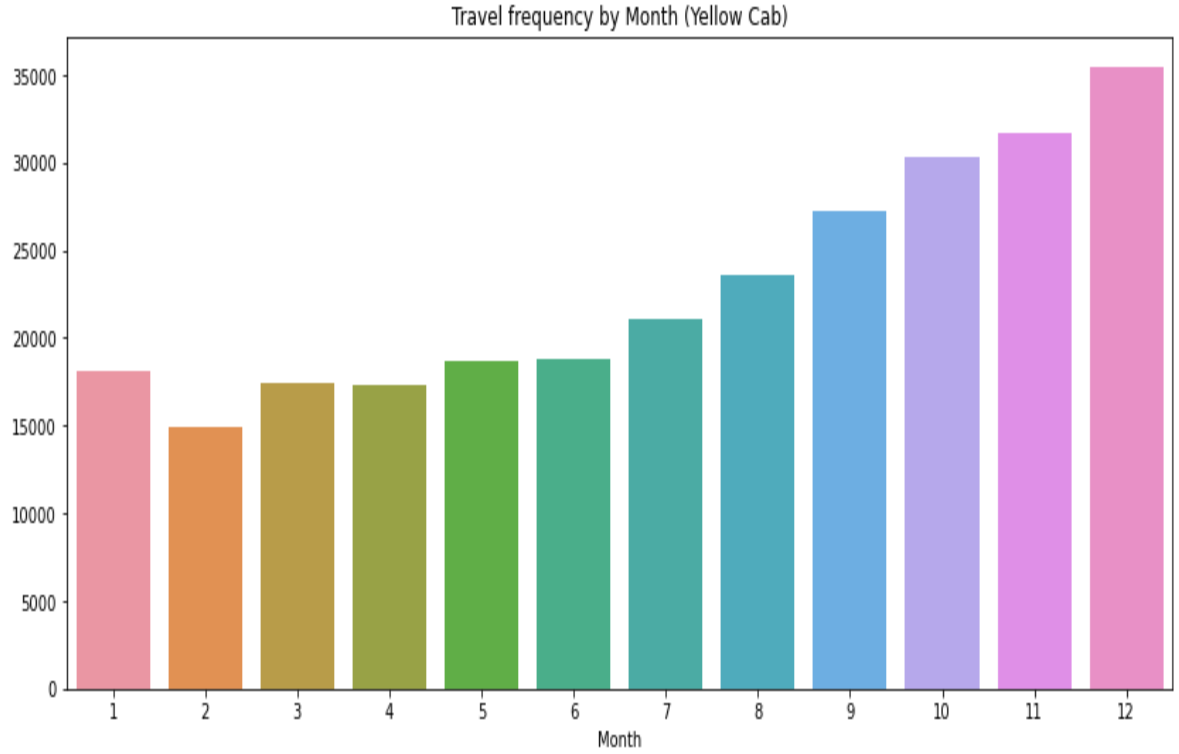
- XYZ is a private firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision. Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment Dataset 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 transaction for 2 cab companies Customer_ID.csv – this is a mapping table that contains a unique identifier which 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 list of US cities, their population and number of cab users.

- **The Analysis include :**
  - **Data Understanding,**
  - **Data Visualization,**
  - **Creating multiple hypothesis,**
  - **Building models and finding the best fit model based on Accuracy.**

EXPLORATORY DATA ANALYSIS

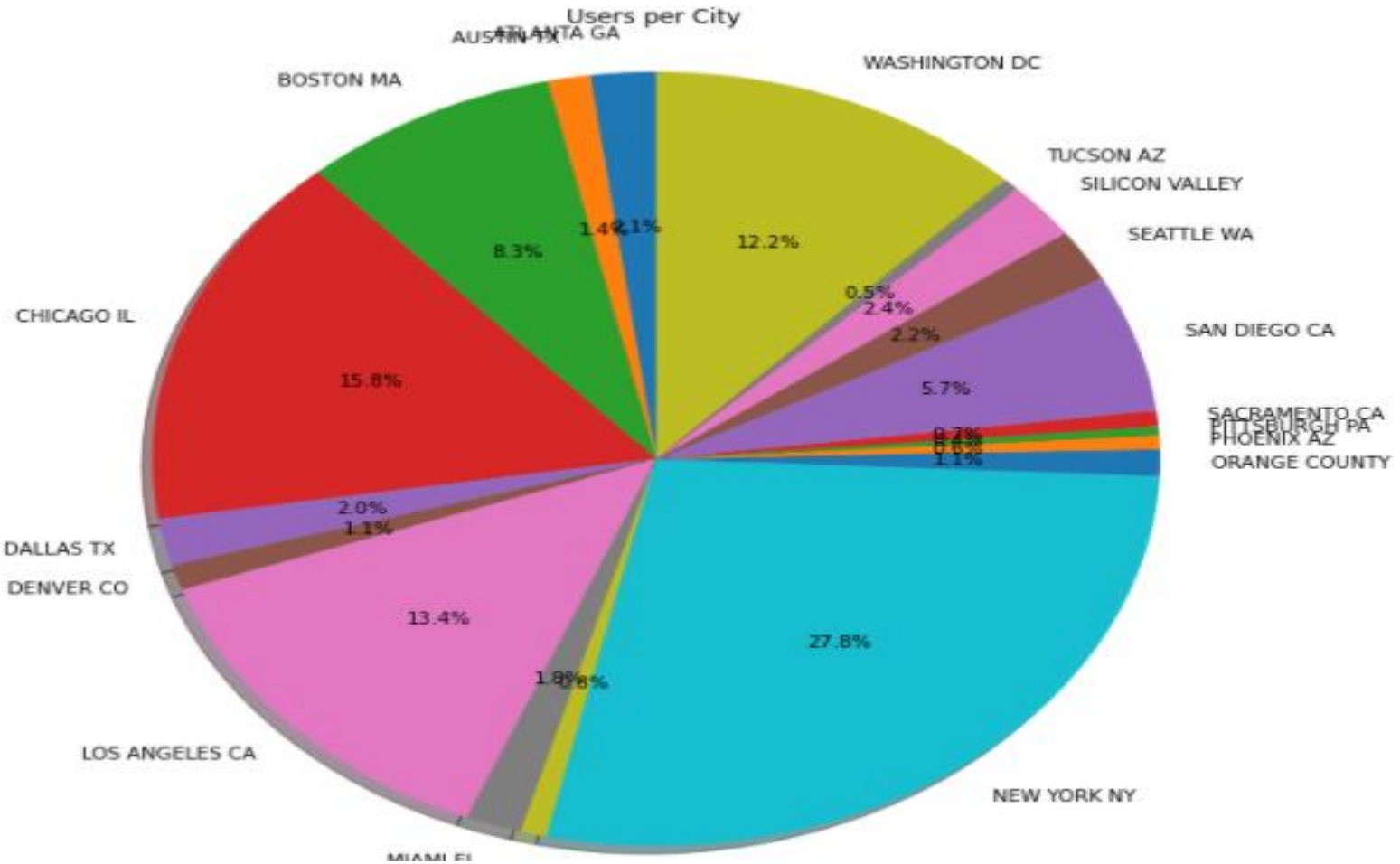
Travel Frequency per Month:





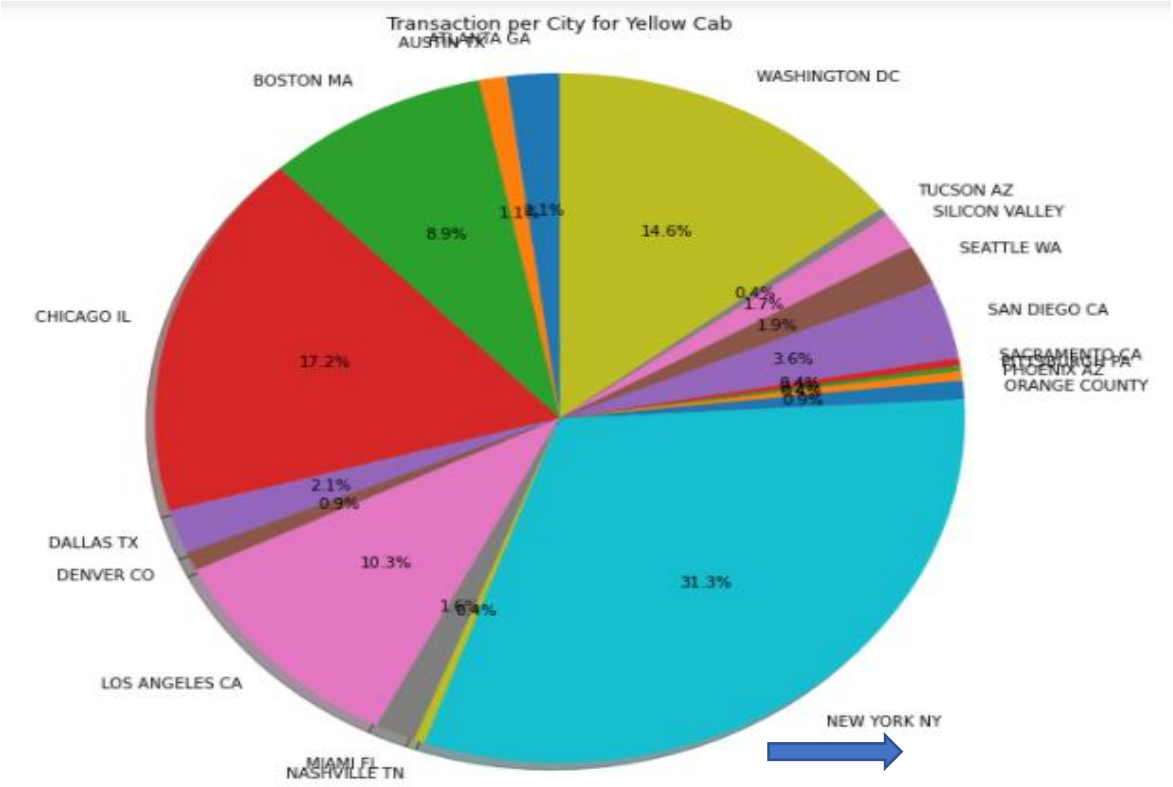
- Yellow Cab has higher travels (35000) in the month of December which is the holiday season compared to Pink Cab (11000).

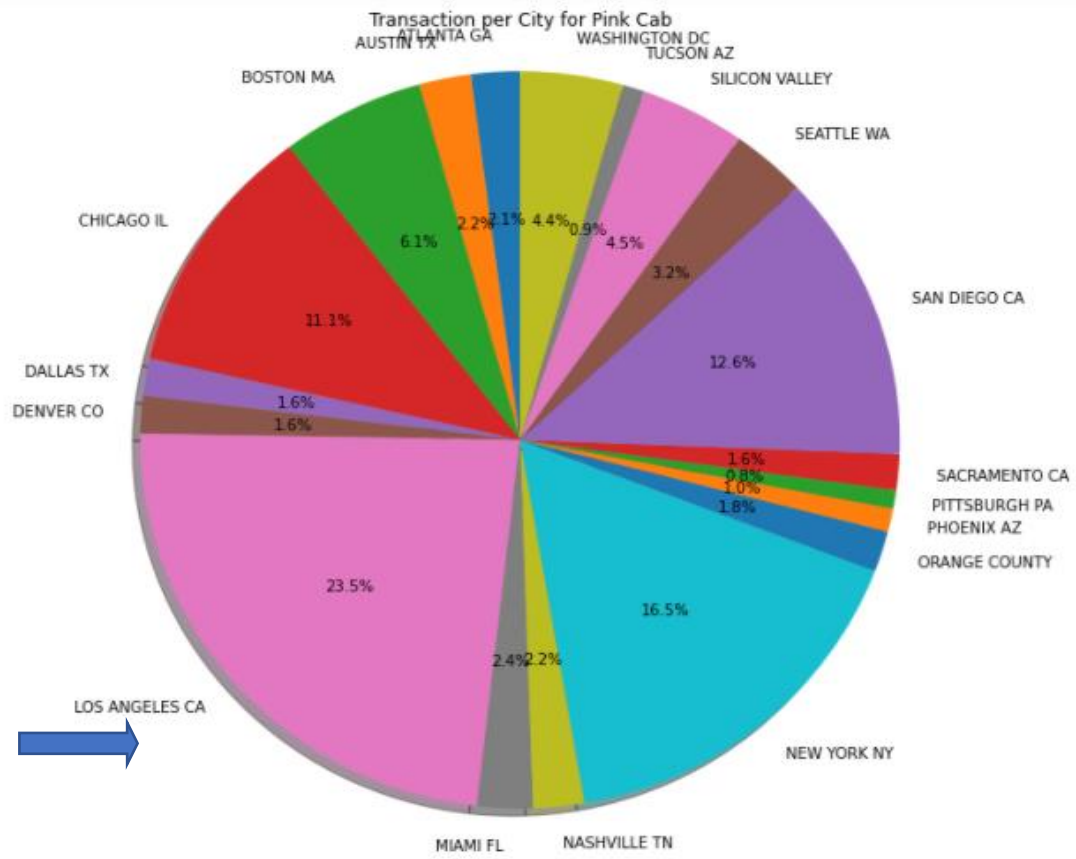
Cab Users per City:



- New York City has the highest Cab users with 28% followed by Chicago with 16% and Los Angeles with 13%

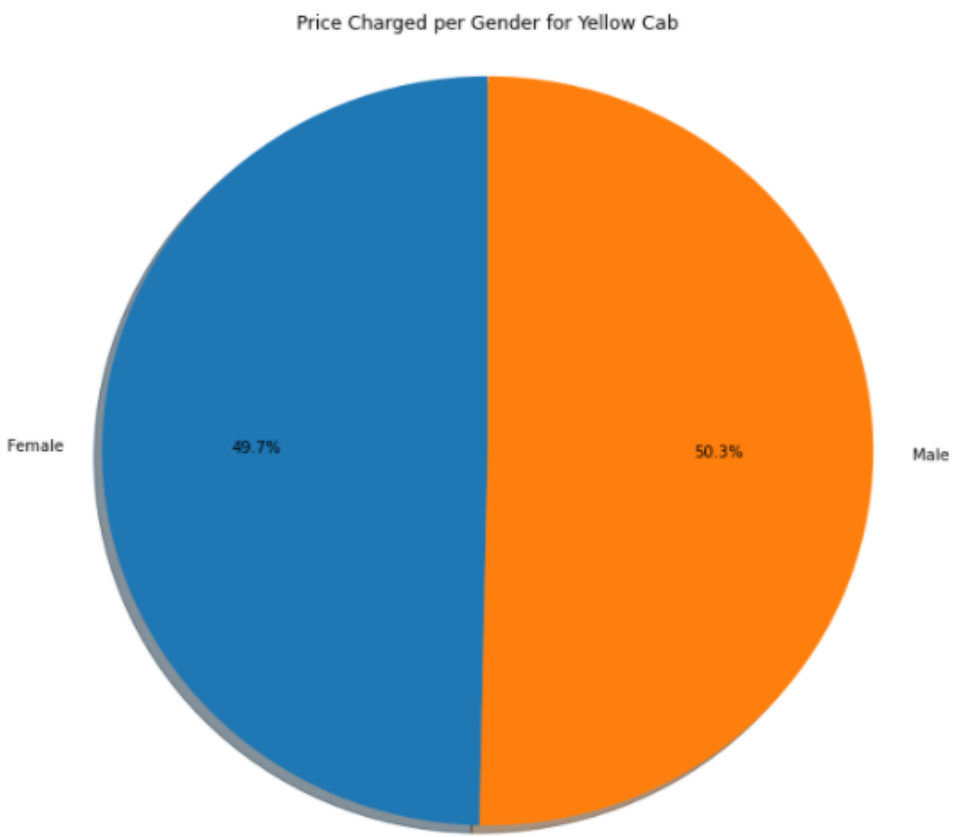
Transaction per City for both Cabs:

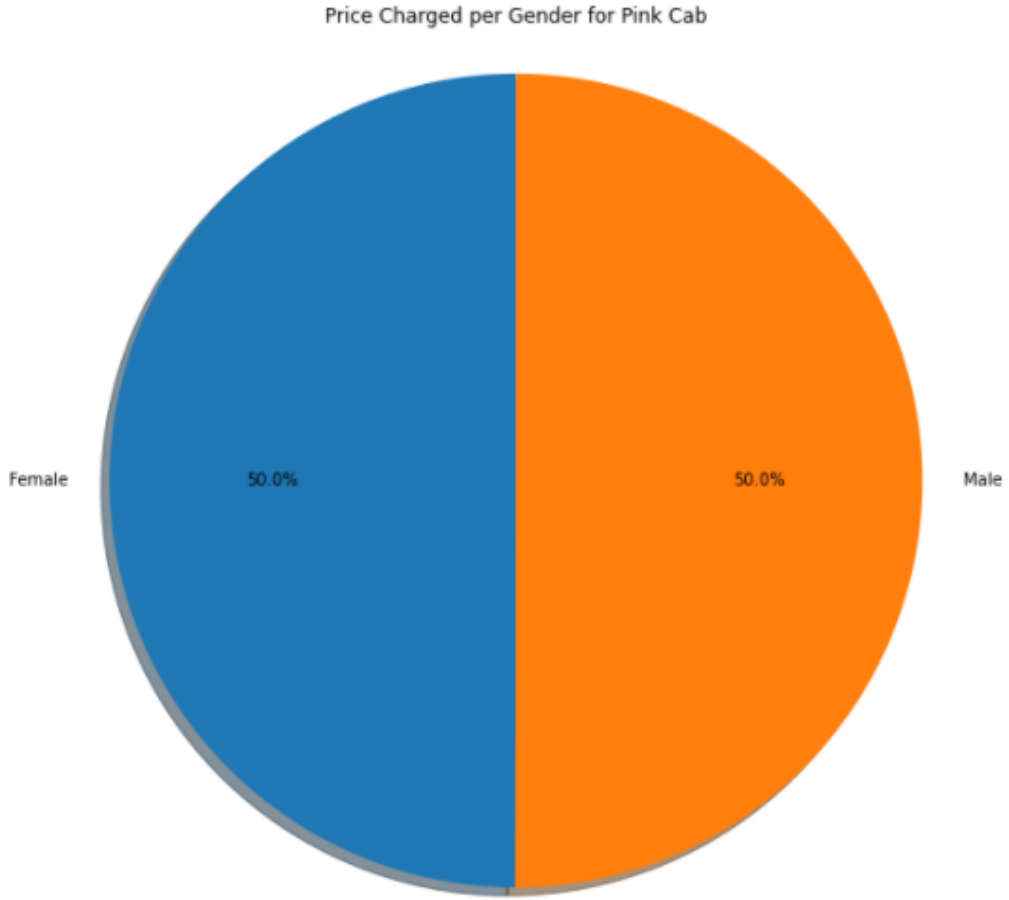




- Transaction for Yellow Cab is highest in New York City(31%) and New York City has the highest Cab Users of 28% .
- Transaction for Pink Cab is highest in Los Angeles City.

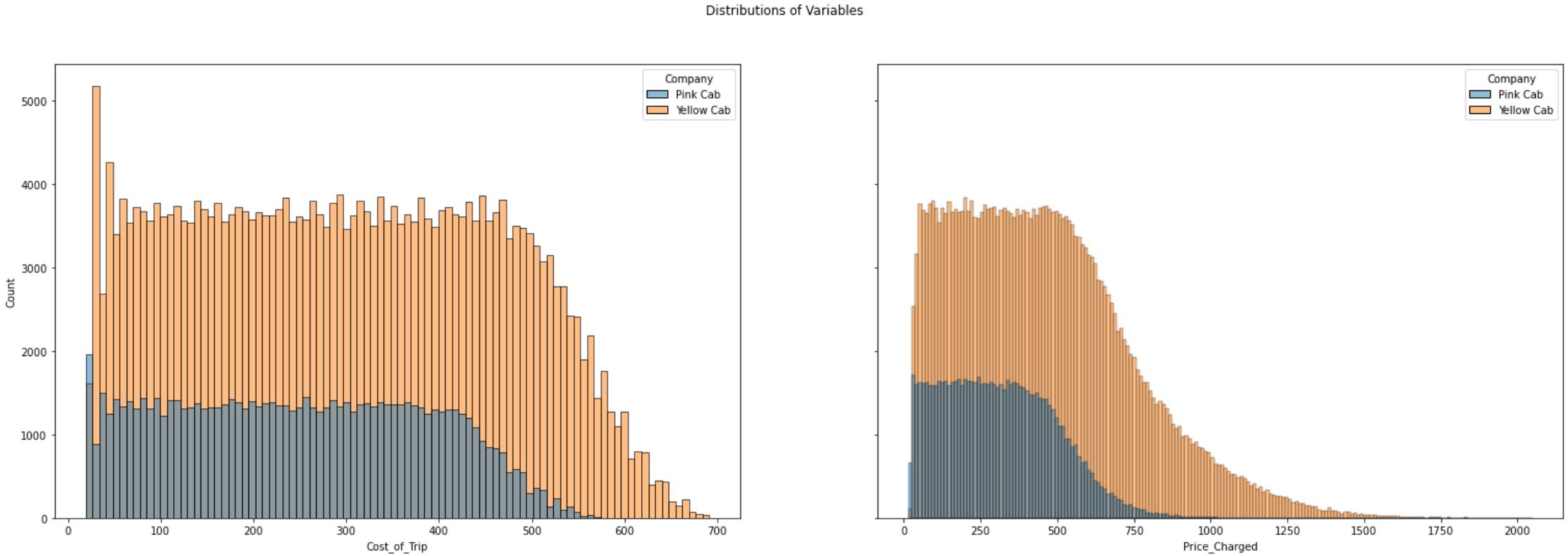
Price Charged per Gender for both Cabs:



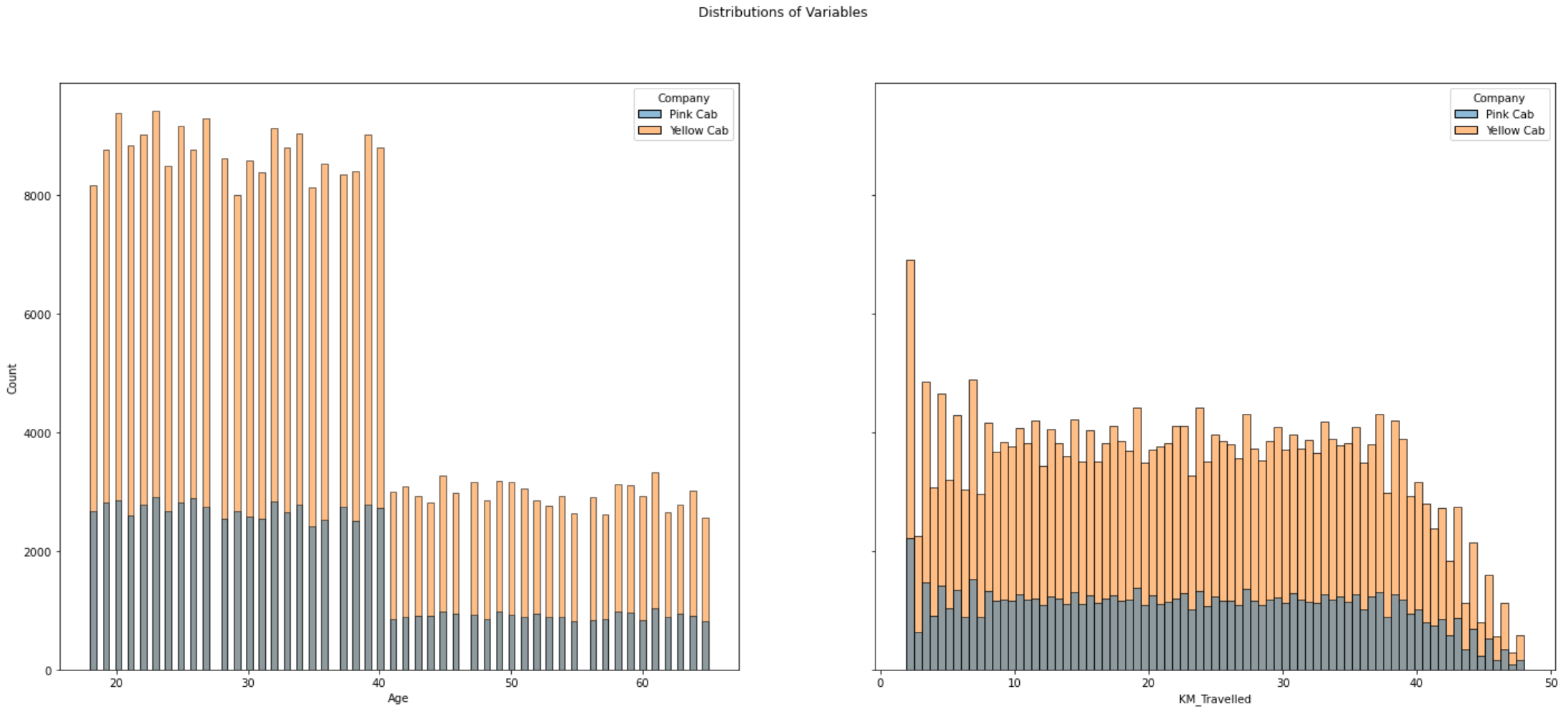


- Yellow Cab charge less from Female Customers. Pink Cab charges the same for both Male and Female Customers.

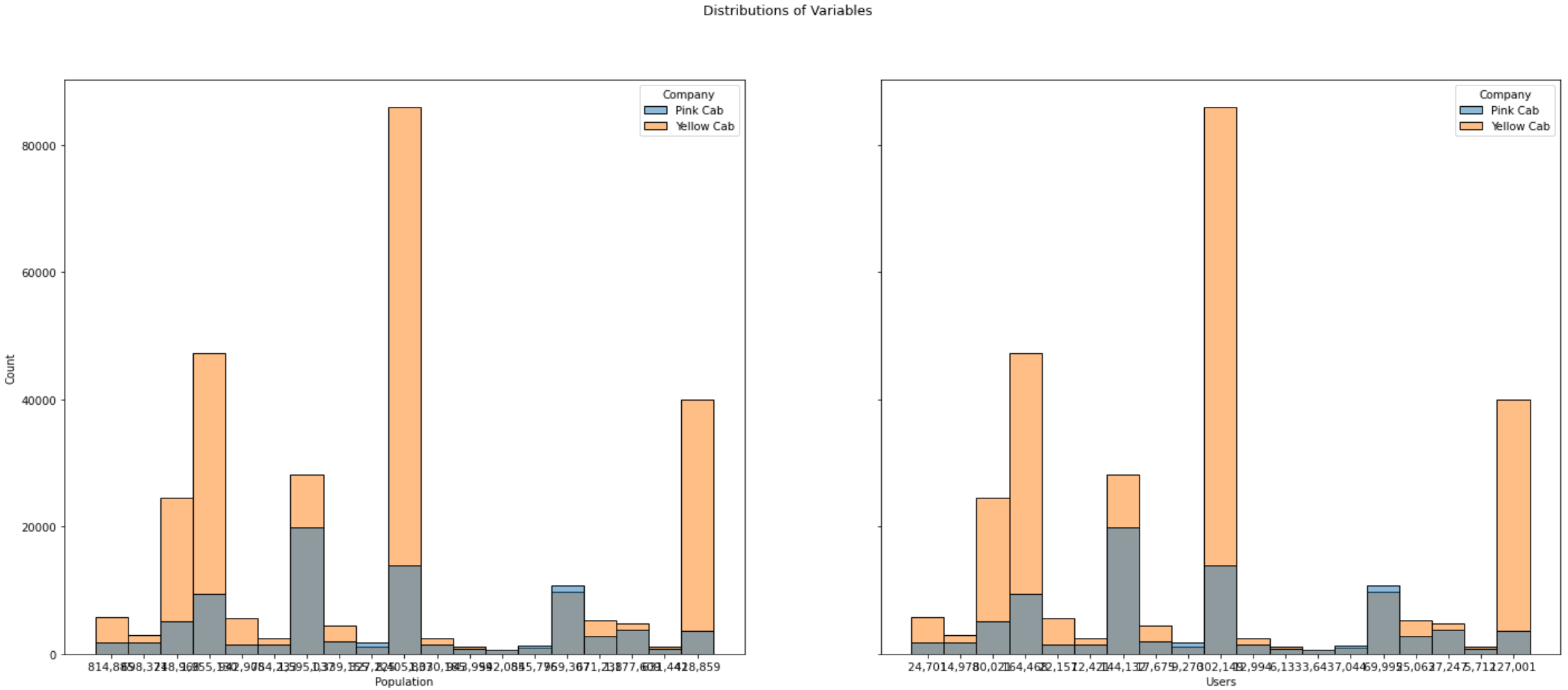
The KDE curves and distribution plots of selected variables related to Cab Firms :



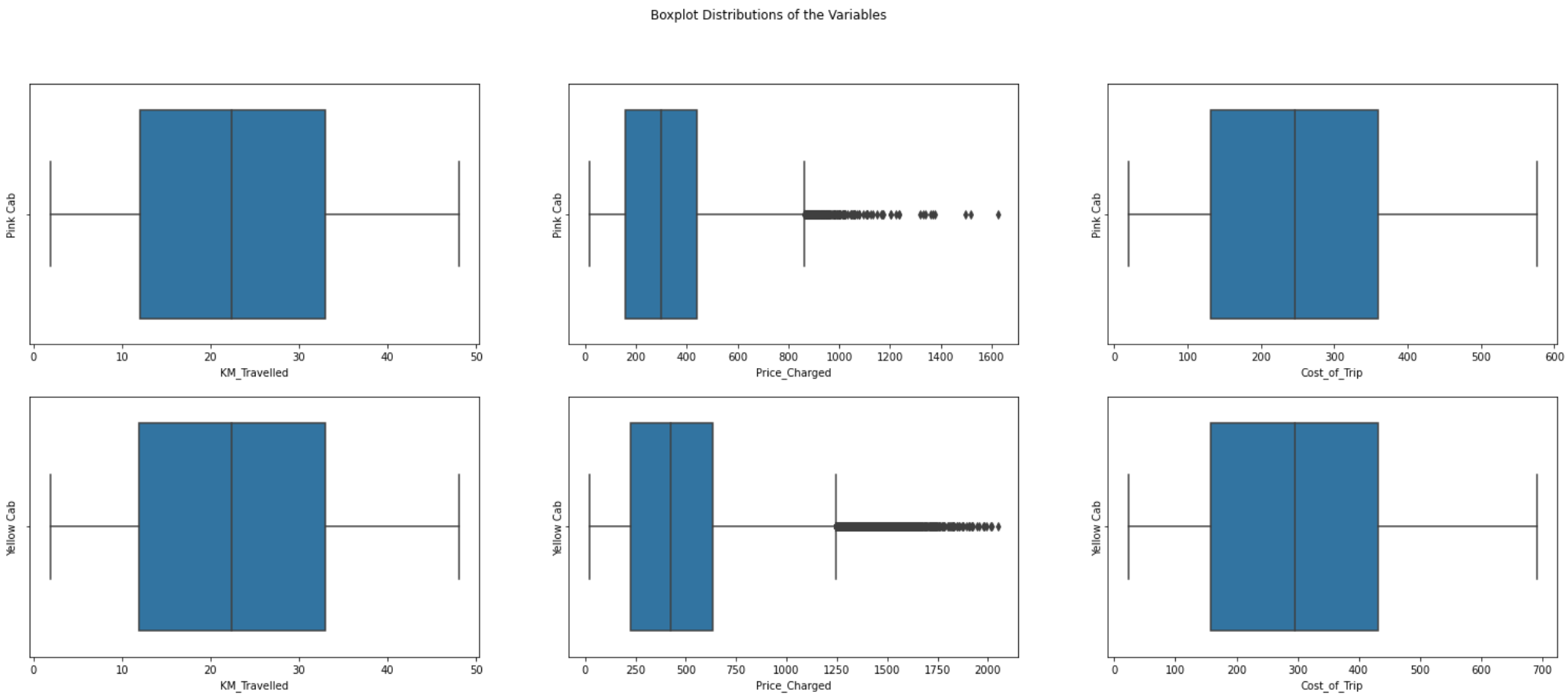
The KDE curves and distribution plots of selected variables related to Cab Firms :



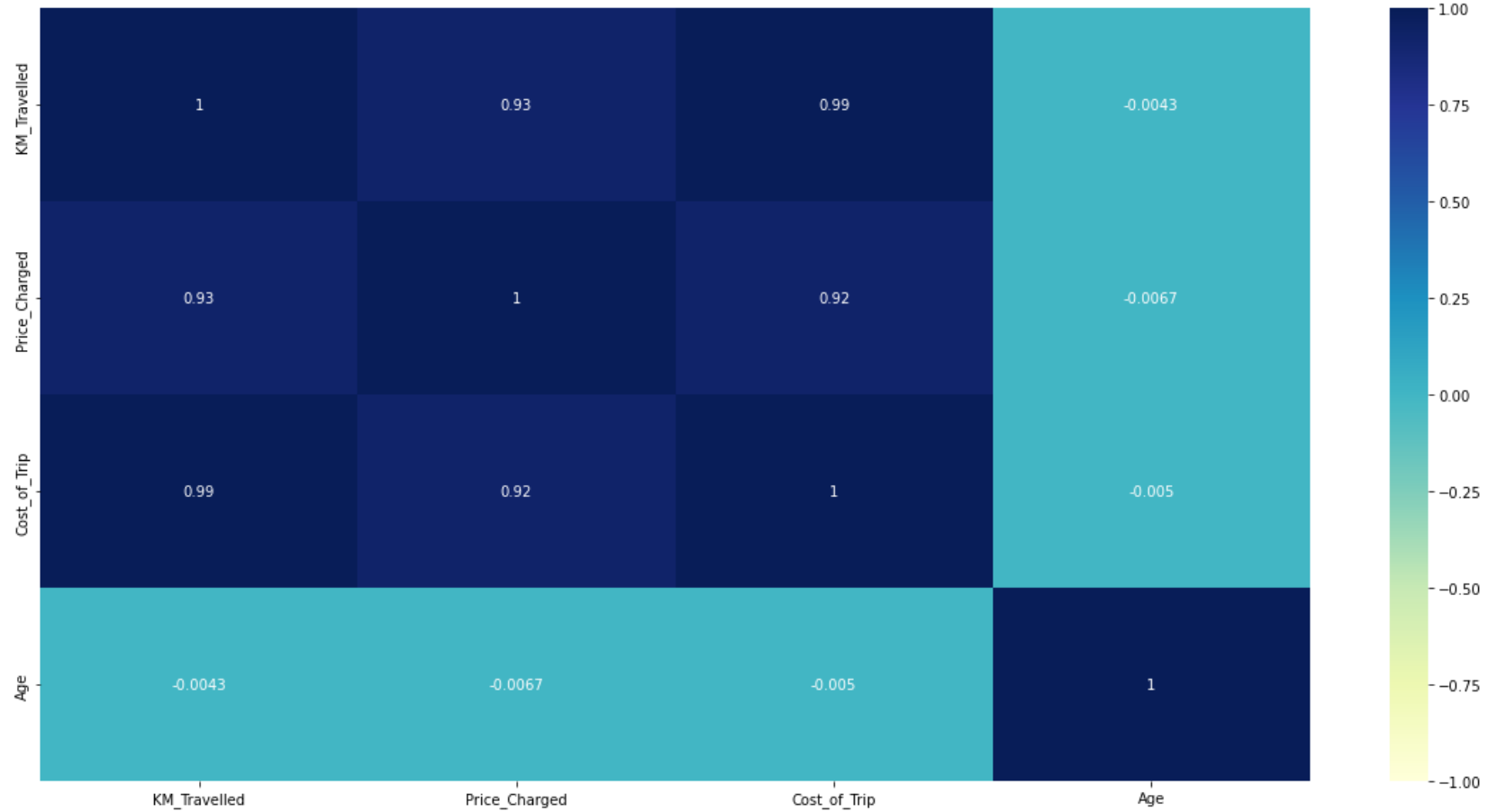
The KDE curves and distribution plots of selected variables related to Cab Firms :



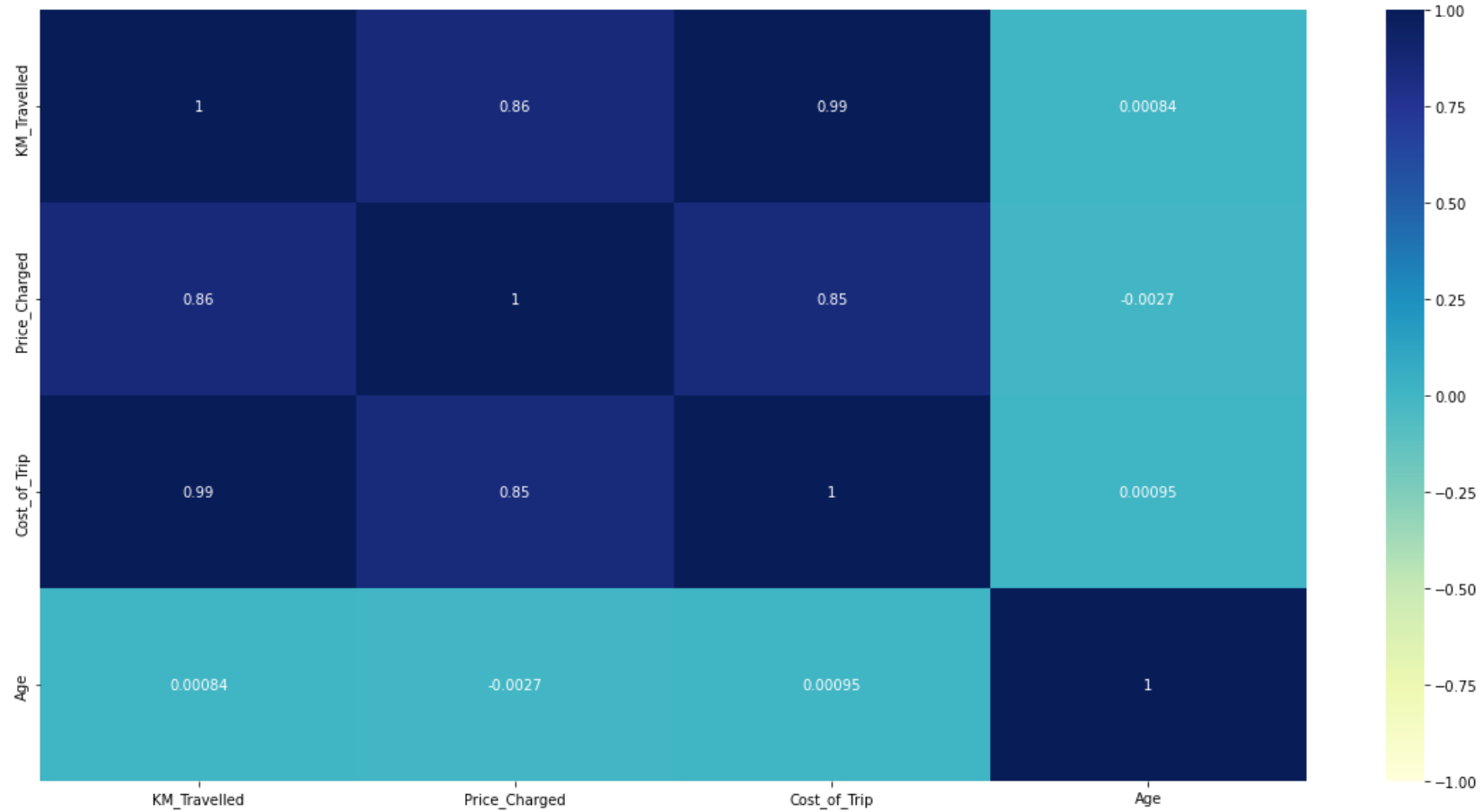
Boxplot distributions of the desired variables were drawn:



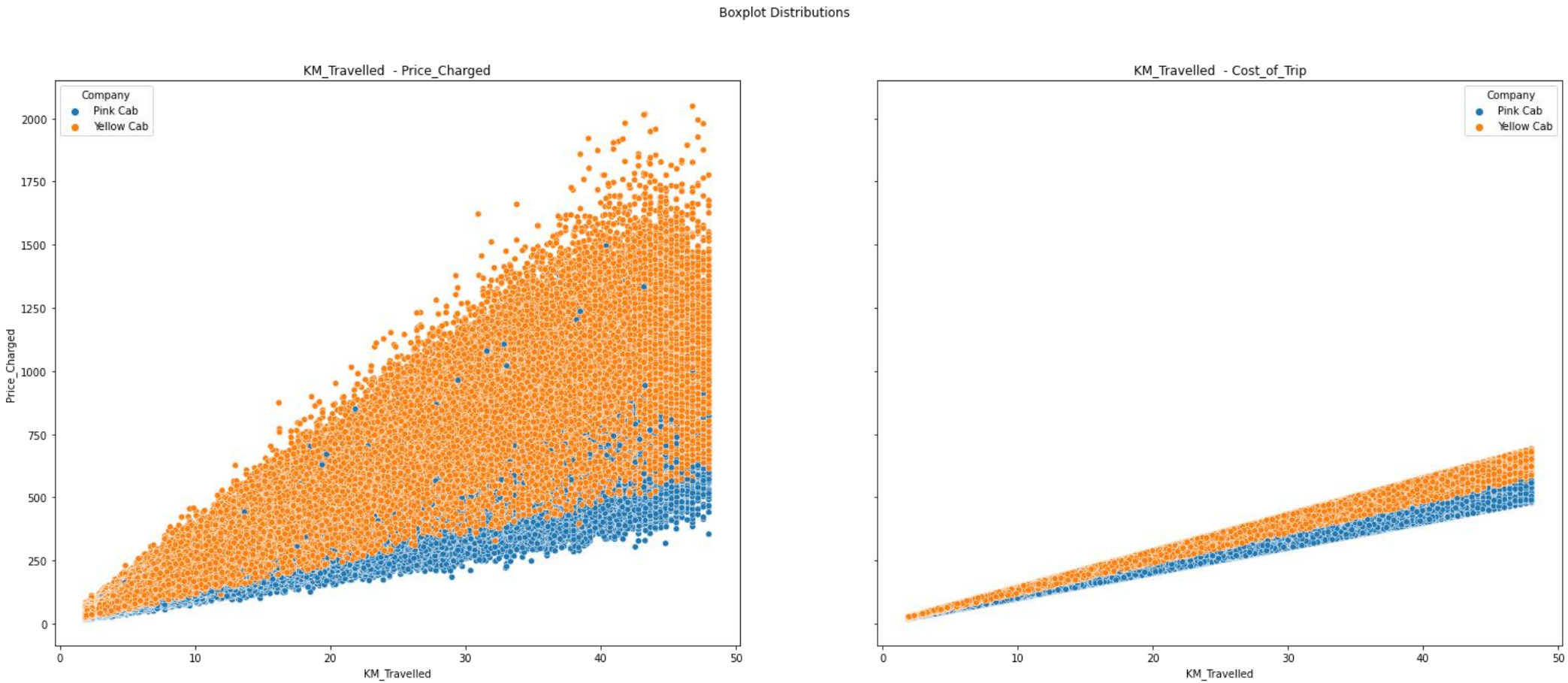
Heatmap of the correlation of MasterData features for Pink Cab Firm:



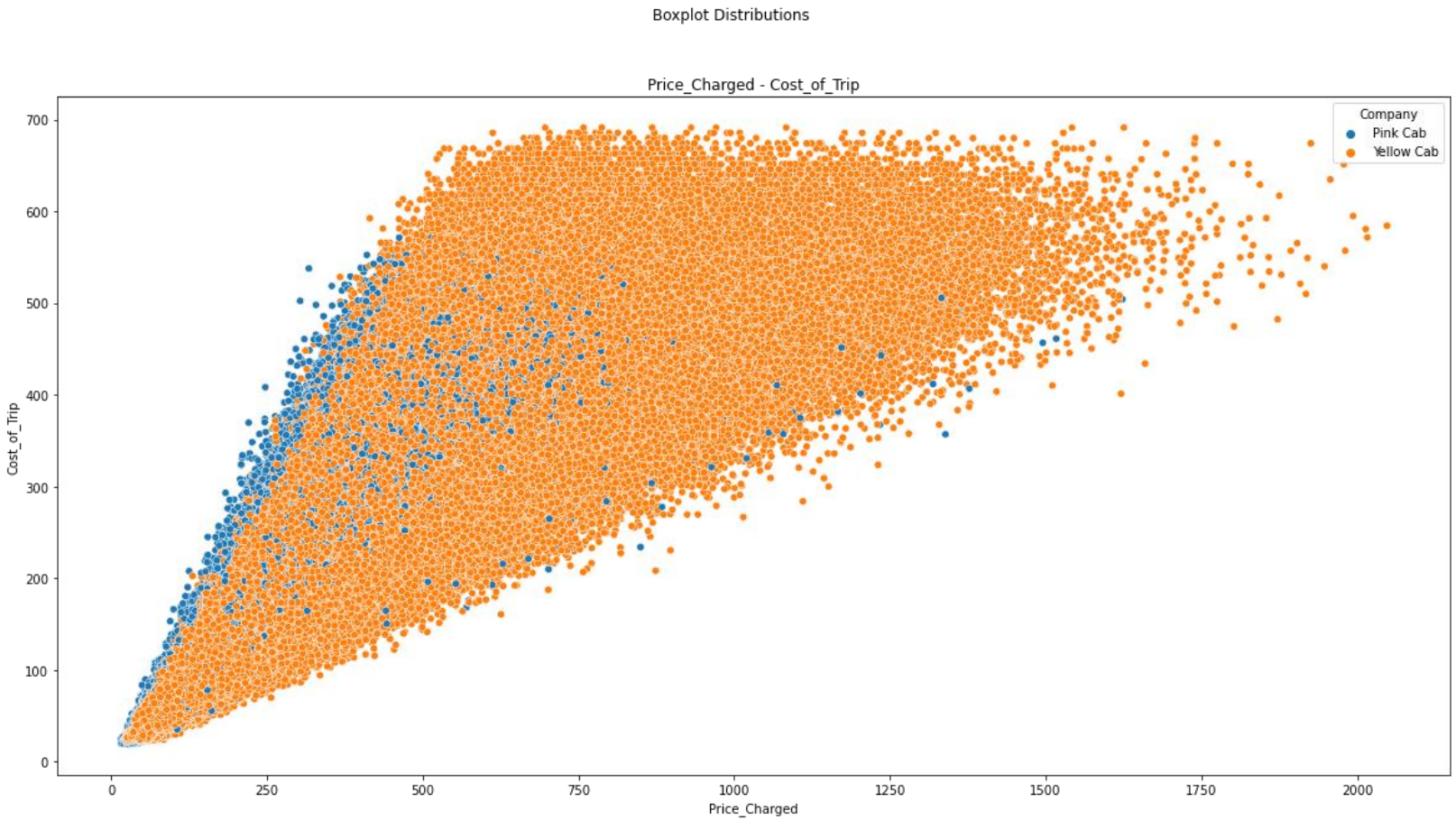
Heatmap of the correlation of MasterData features for Yellow Cab Firm:



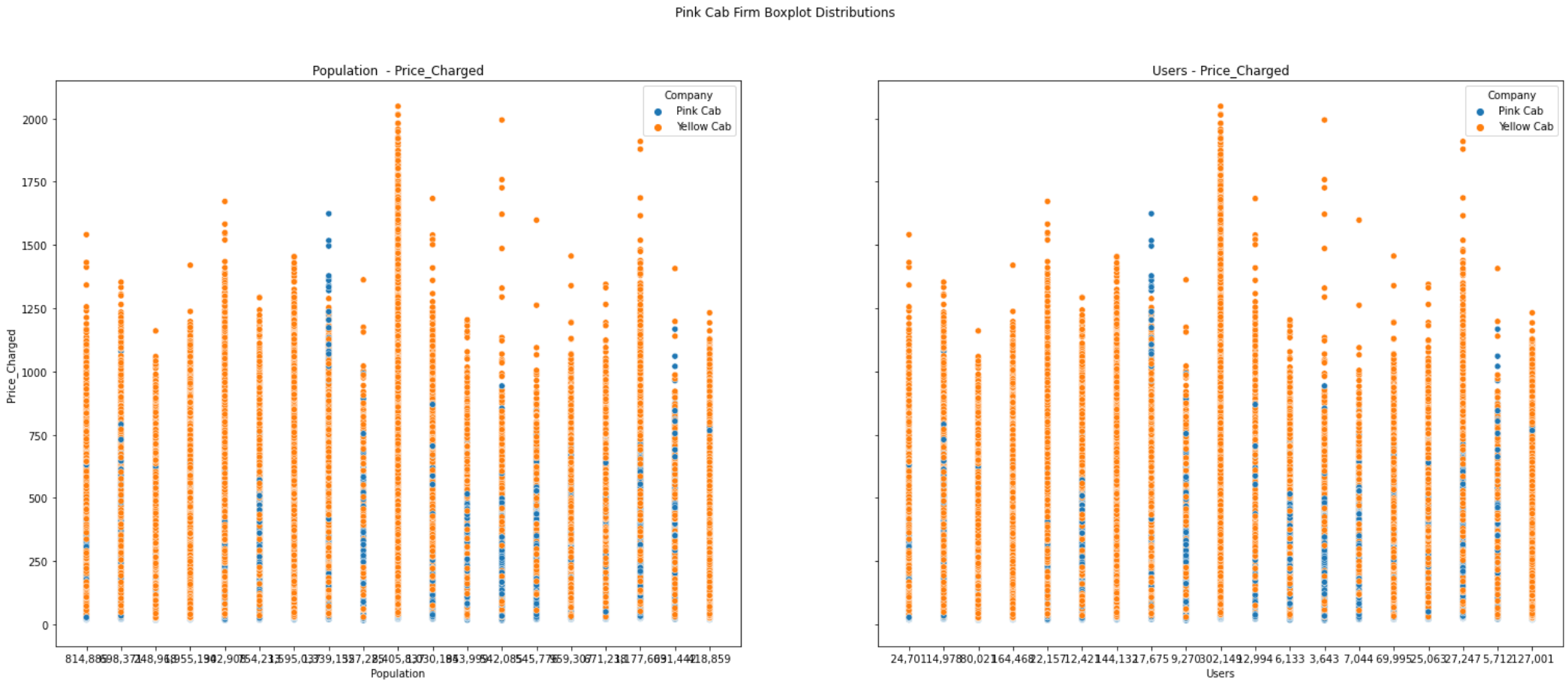
Below are visualized scatter plots of MasterData features to see correlations between them if exists:



Below are visualized scatter plots of MasterData features to see correlations between them if exists:



Below are visualized scatter plots of MasterData features to see correlations between them if exists:



Building Predictive Models using Linear Regression, Decision Tree, Random Forest, Gradient Boosted, Light Gradient Boosted

Best Fit Model: RMSE Value & Accuracy

- **RMSE or root mean square error measures the error which is Prediction values – Actual values.**
- **Lower the RMSE value the better is the Model.**
- **Higher the Accuracy-better the model**

First, split the data into a training set (75%), and test set (25%).

Yellow Cab

	Accuracy
Linear Regression	74.138775439840 81%
Decision Tree	46.992906507436 3%
Random Forest	72.736691604233 3%
Gradient Boosted	78.029176353199 19%
Light Gradient Boosted	76.662302894854 56%

Pink Cab

	Accuracy
Linear Regression	86.049852007617 25%
Decision Tree	70.654788633715 8%
Random Forest	85.378924866032 96%
Gradient Boosted	86.077628397844 86%
Light Gradient Boosted	86.057568659318 62%

Best Fit Model: RMSE Value & Accuracy

As per the above data, Gradient Boosted Model is the best fit model for further deployment.

Recommendation

- Transaction per year: For Yellow Cab Transaction per year from 2016 to 2018 is almost double than Pink Cab.
- Customers for Yellow Cab is highest in New York City which has the highest Cab Users of 28%.

These points are two only examples from the above analysis. Hence, on the basis of our above analysis, Yellow Cab is recommended for investment.

Thank You