Exploratory Data Analysis
G2M Case Study

April 17, 2023

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

Case Study : Cab Company

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 : come up with actionable insights to help company XYZ identify the right company to make their investment.

DATA EXPLORATION

4 individual data sets with time period of data from **31/01/2016** to **31/12/2018.**

Below are the list of datasets which were provided for the analysis:

- **Cab_Data.csv**
- **Customer_ID.csv**
- **Transaction_ID.csv**
- **City.csv**

The 4 datasets were successfully merged to obtain a comprehensive dataset which would enable easy and effective analysis.

HYPOTHESIS

- Hypothesis 1: Cab company B has a larger market share than cab company A.

- Hypothesis 2: Cab company B has a higher customer retention rate than cab company A.

- Hypothesis 3: Cab company B has a lower cost per ride than cab company A.

- Hypothesis 4: Cab company B has a higher average ride duration than cab company A.

Where : Pink cab = cab company A, Yellow cab = cab company B
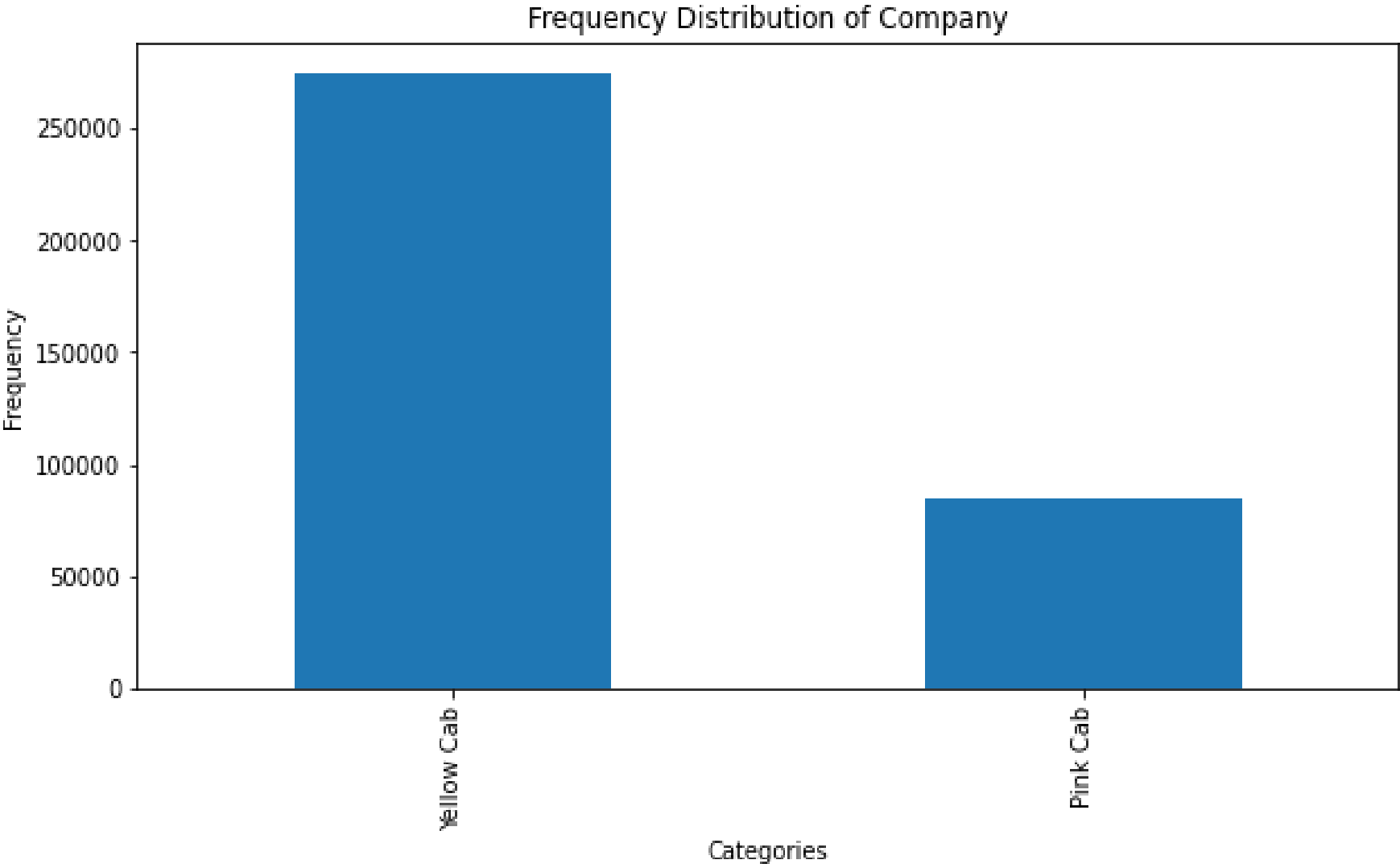
ANALYSIS

After the data was successfully merged, the following analysis were carried out on the dataset;

- Descriptive analysis : Examined the data types of each column, carried out search for missing values and got results of measures of central tendency (mean, STD, mode etc)

- Correlation analysis : Plotted a correlation heat map for the numerical variables. Obtained a high correlation between cost of trip, price charged and km travelled.

- Chi-squared test of Independence : this test was carried out on the categorical variables to test their independence on one another with the help of contingency tables.
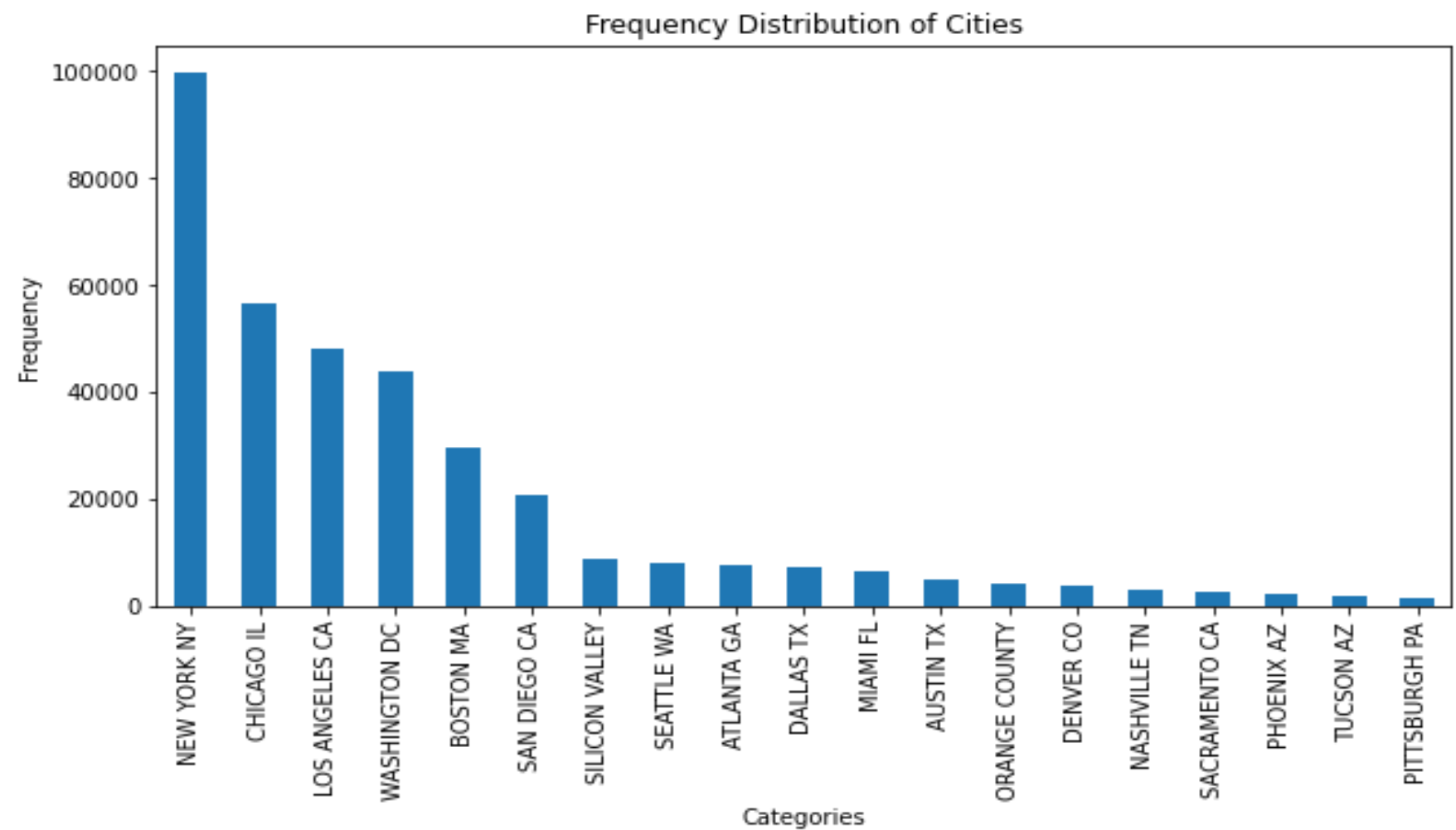
ANALYSIS

- Time Series analysis : helped to understand the underlying causes of trend and systemic patterns over time. Obtained insightful visualizations for better decision making.
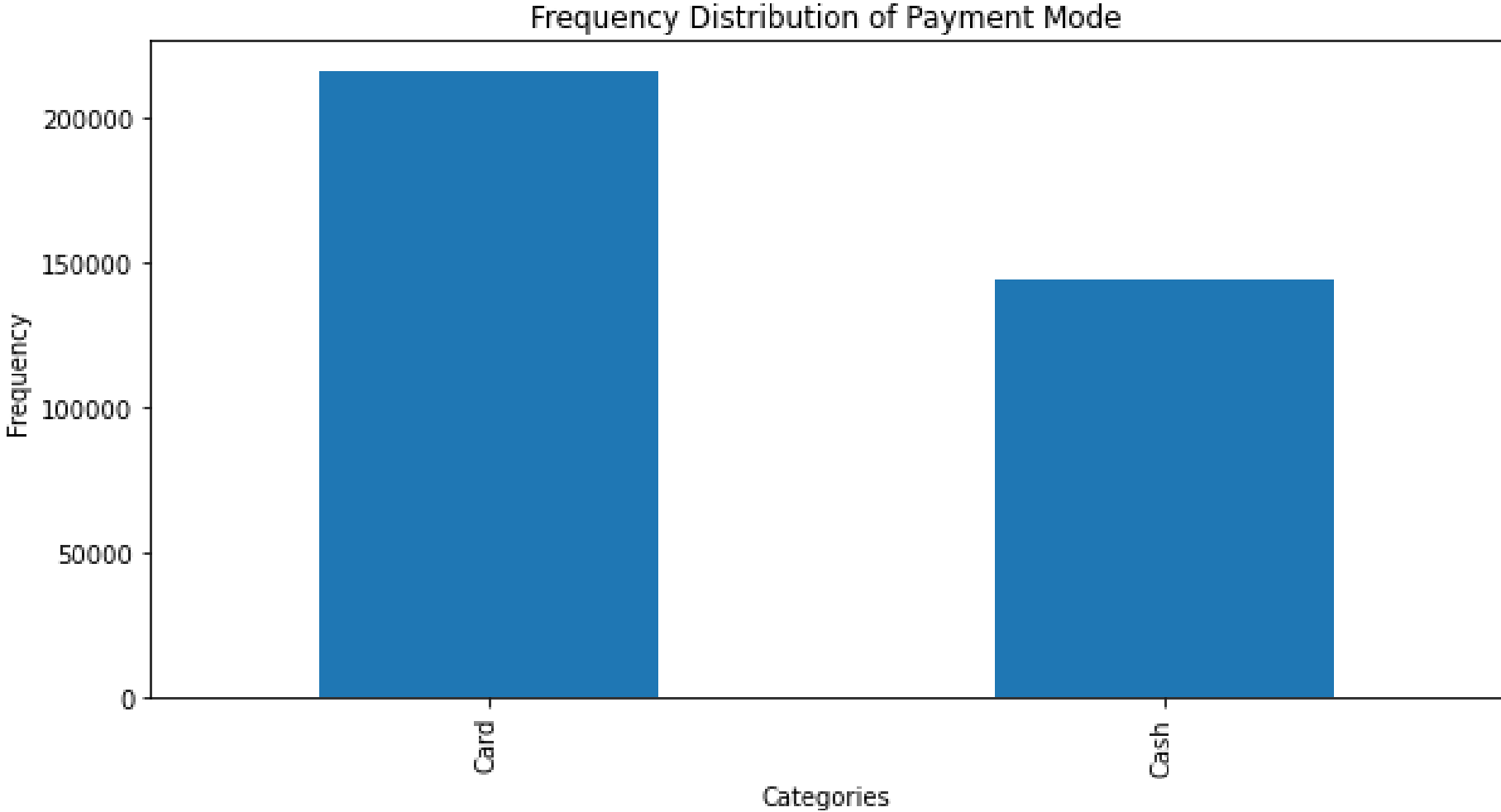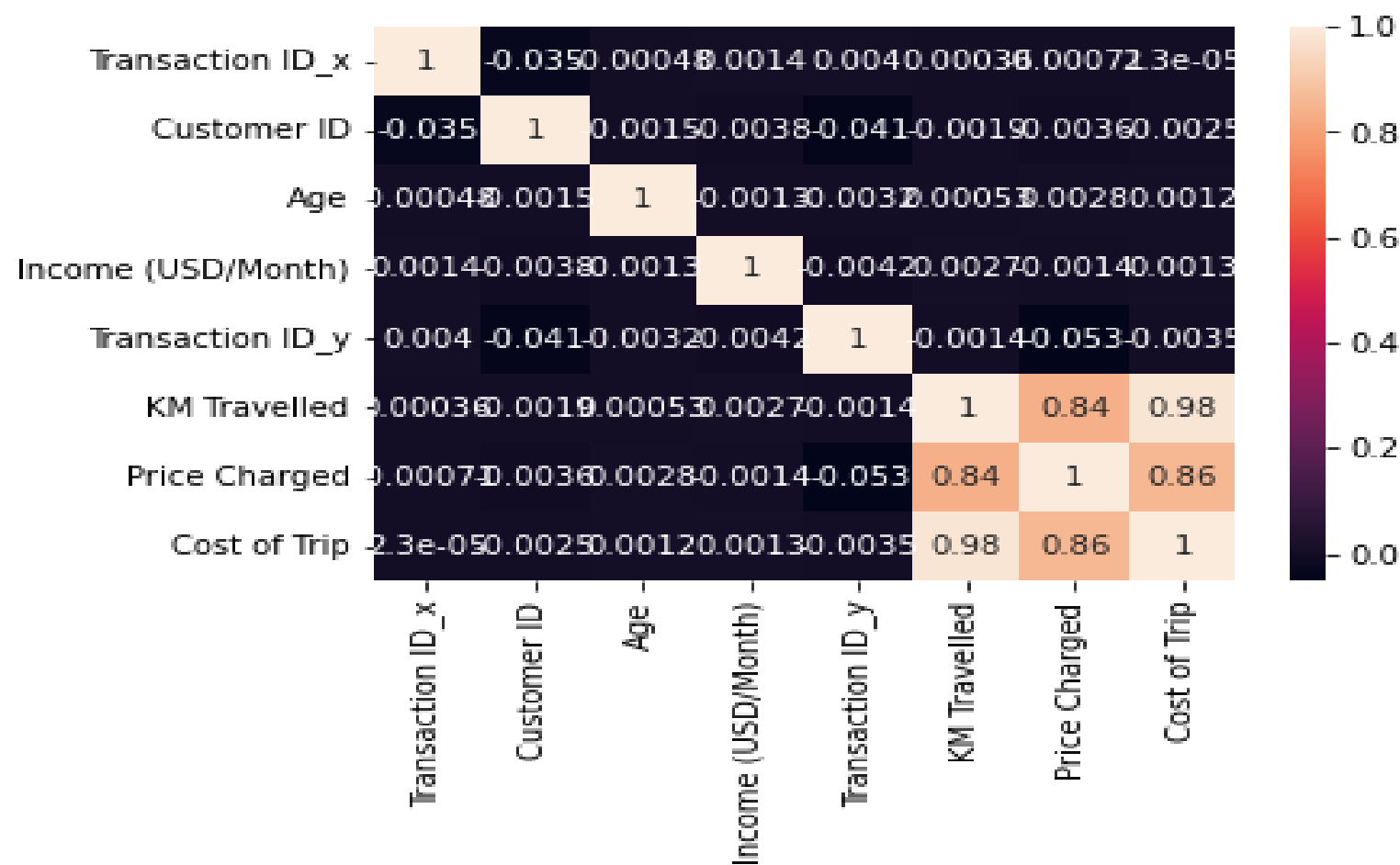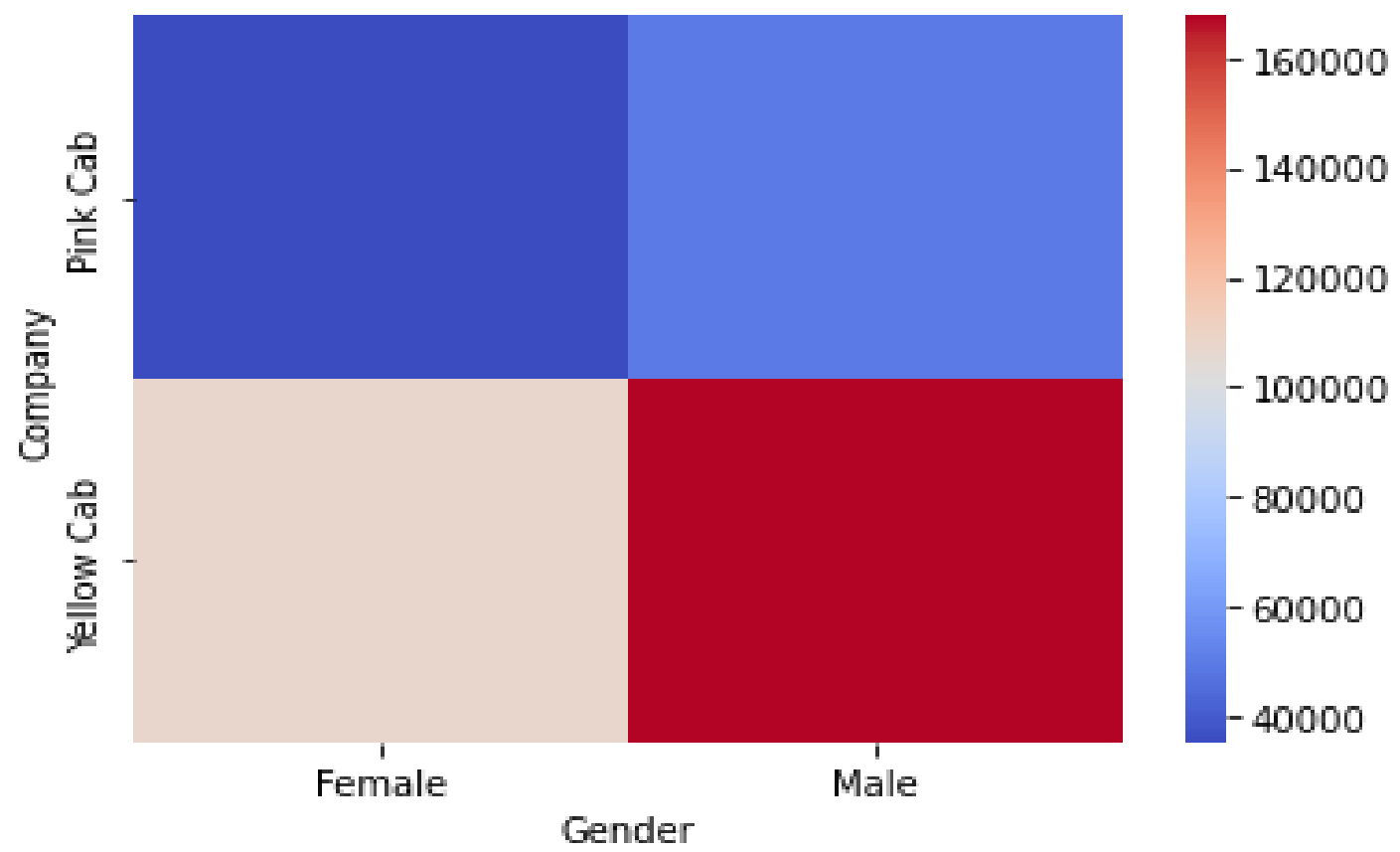
FREQUENCY DIST OF COMAPNY

Frequency Distribution of Company

FREQUENCY DIST OF CITY

Frequency Distribution of Cities

FREQ DIST OF PAYMENT METHOD

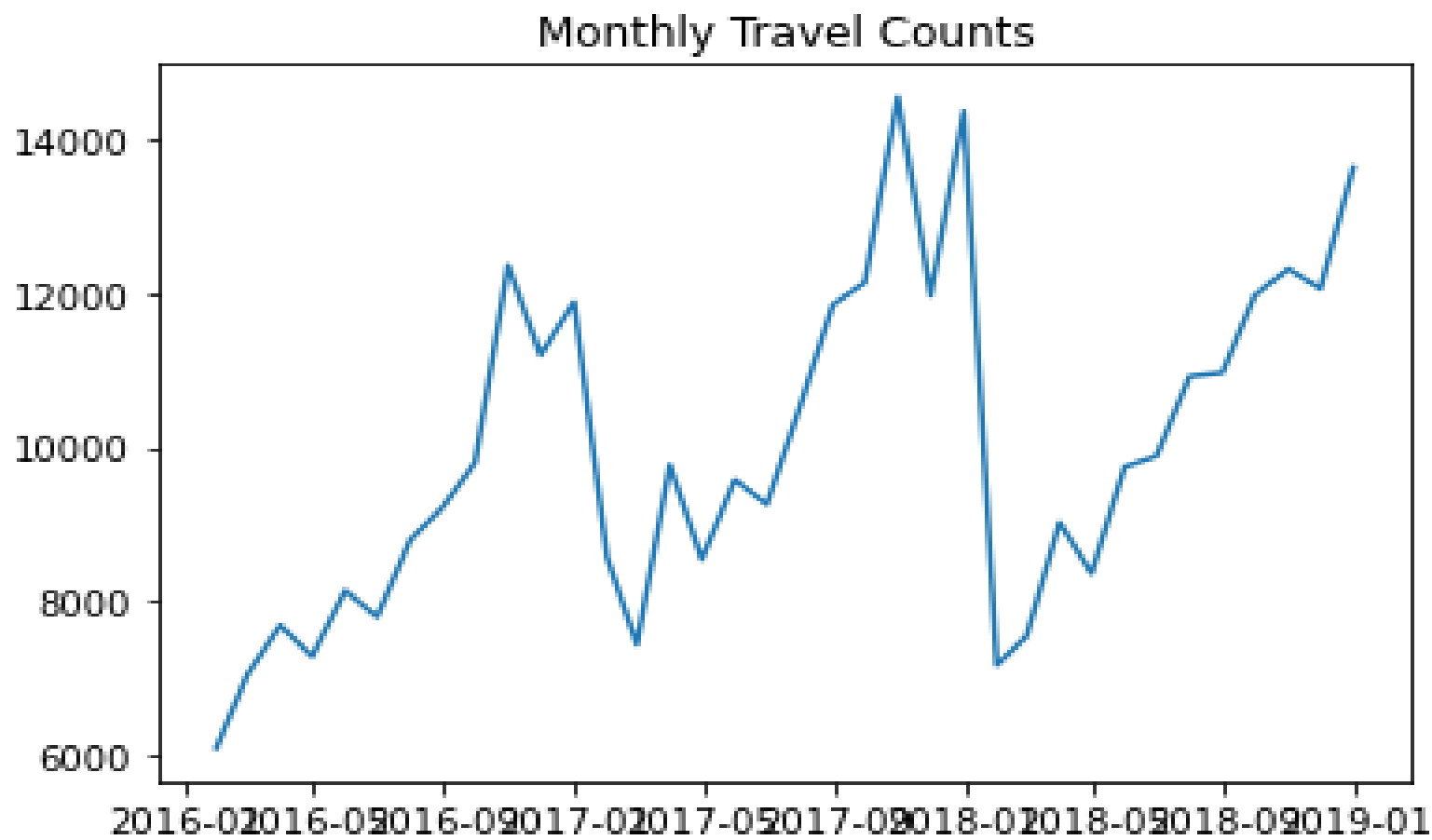Frequency Distribution of Payment Mode

CORRELATION ANALYSIS ON NUMERICAL VARIABLES

CORRELATION OF GENDER AND COMPANY

MONTHLY TRAVEL COUNT
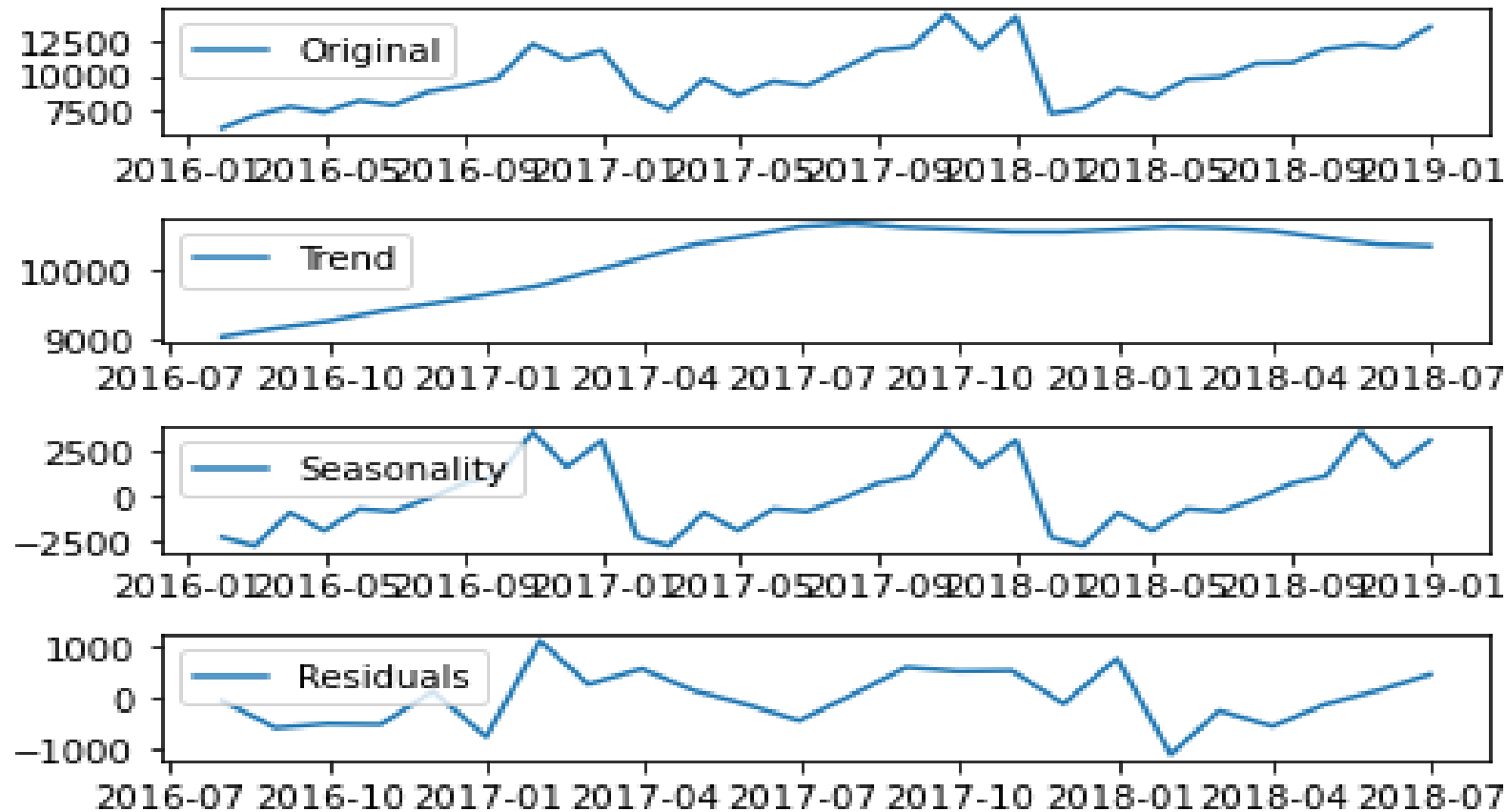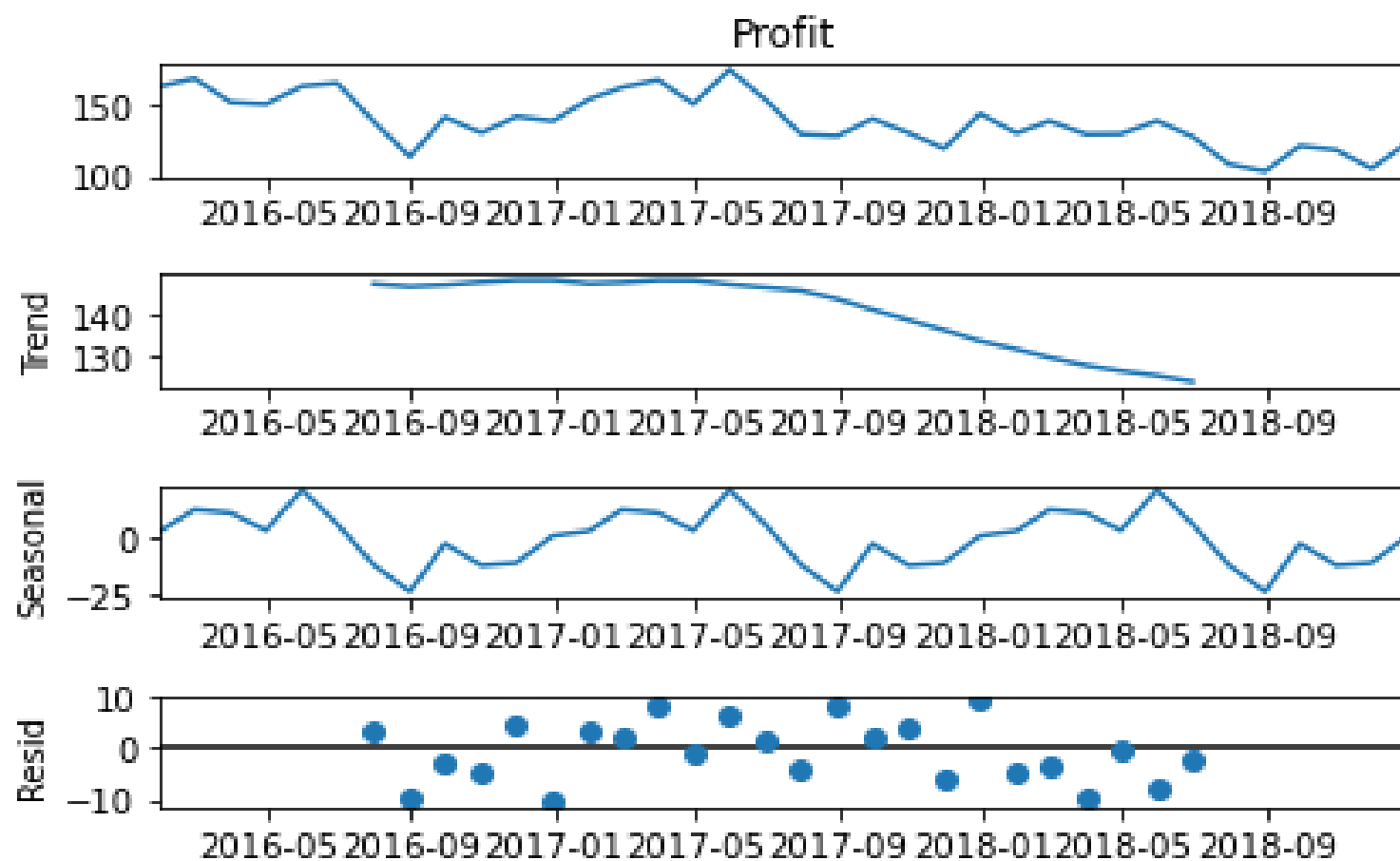
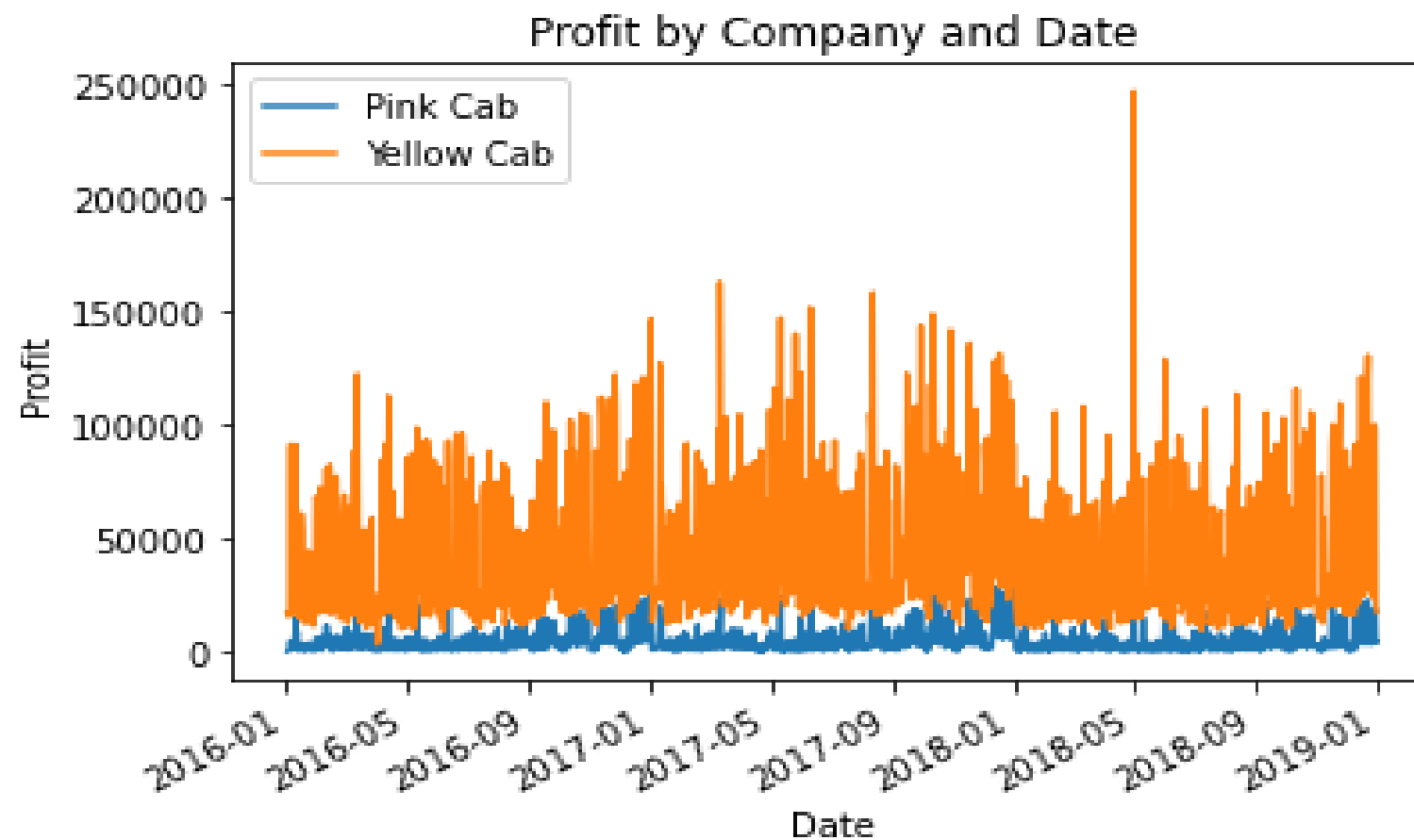

Monthly Travel Counts

DECOMPOSED TIME SERIES

MONTHLY PROFIT

PROFIT BY COMPANY AND DATE

RECOMMENDATION

From the analysis carried out, we've been able to come up with these insights as to why the yellow cab is best for investment.

- The highest number of trips were taken by users aged between 21-30, and this age group also generated the highest profit for the yellow cab company. This could suggest that the company should focus more on this age group in its marketing campaigns and customer acquisition efforts.

- The month of December had the highest number of trips and generated the highest profit for the yellow cab company. The company could leverage this information to plan special holiday packages and promotions for its customers during this period to maximize revenue.

- The time series analysis revealed that the yellow cab company's profits unlike the pink cab company have been generally increasing over time, with some seasonality and trend components.

Thank You