Data Glacier

Your Deep Learning Partner

G2M Insight for Cab Investment Firm

Company Name: XYZ

October 2022

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Hypothesis Testing

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.***

- ***Provide actionable insights to help XYZ firm in identifying the right company for making investment.***

- ***The Analysis included:***
  - ***Review the Source Documentation***
  - ***Understand the field names and data types***
  - ***Identify relationships across the files***
  - ***Field/feature transformations***
  - ***Determine which files should be joined versus which ones should be appended***
  - ***Create master data and explain the relationship***
  - ***Identify and remove duplicates***
  - ***Perform other analysis like NA value and outlier detection***

Data Sets

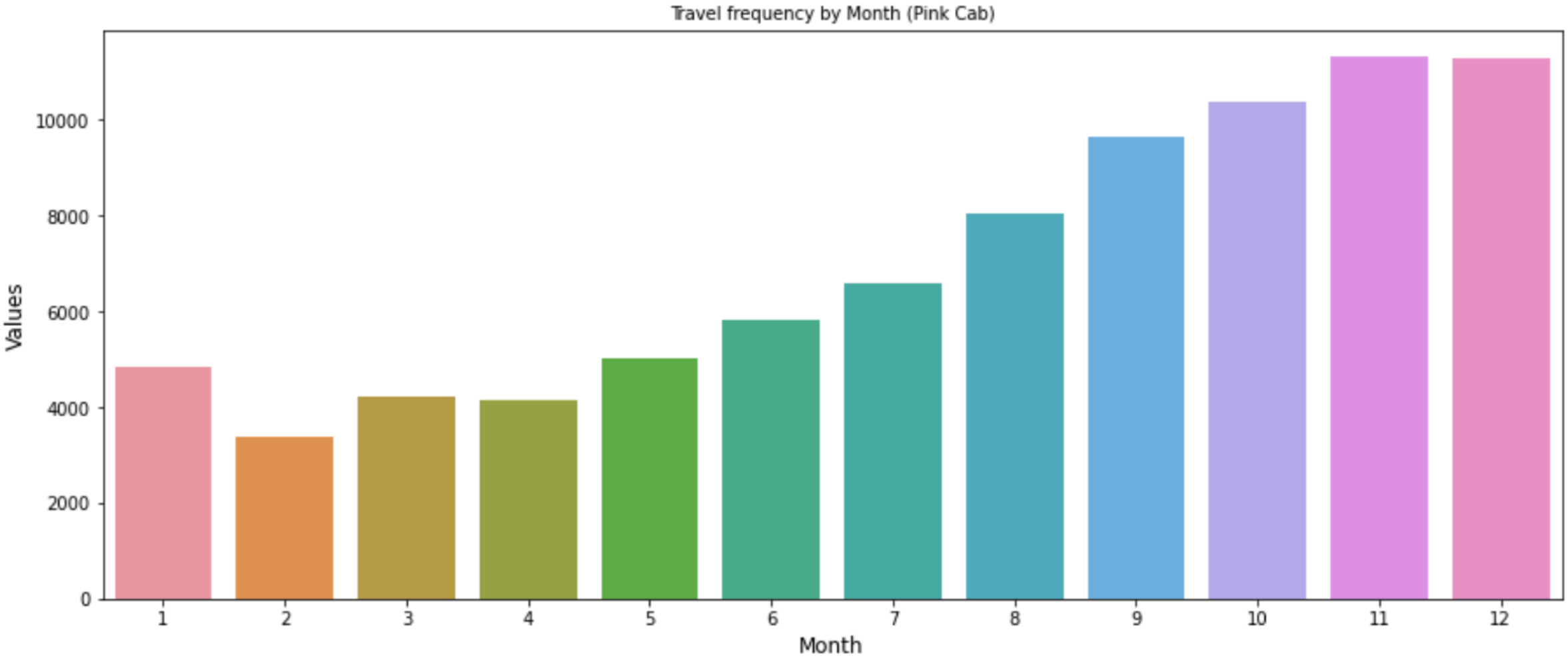
- **Cab_Data.csv – this file includes details of transaction for 2 cab companies**
- **Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details**
- **Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping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

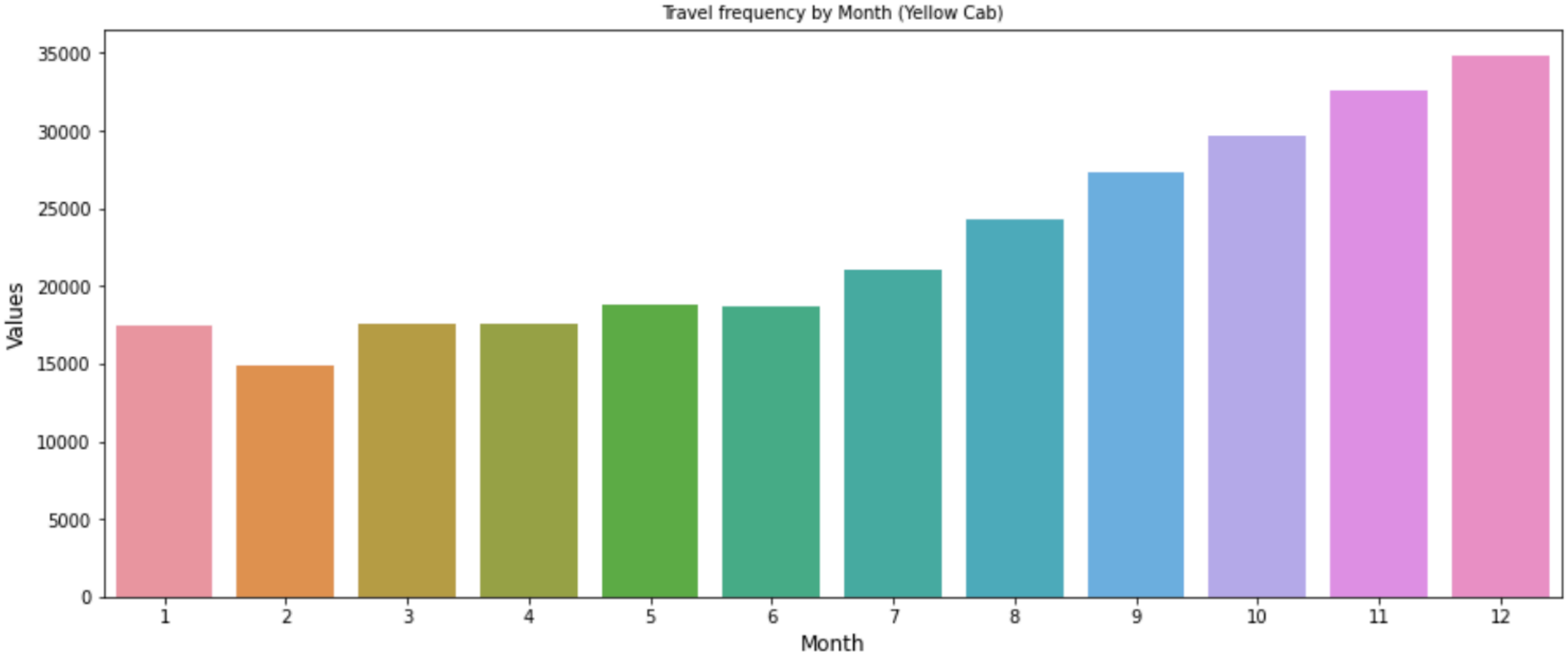
EXPLORATORY DATA ANALYSIS

Cab Companies:

- **Pink Cab**
- **Yellow Cab**

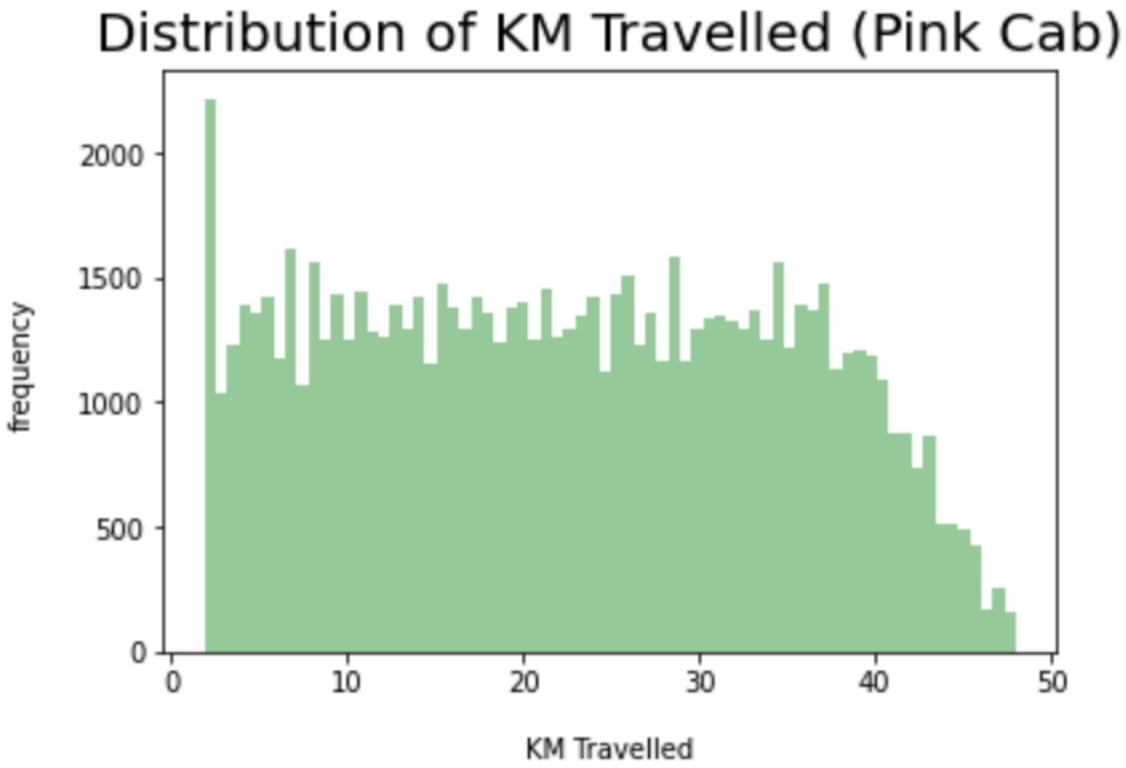
Travel frequency by Month

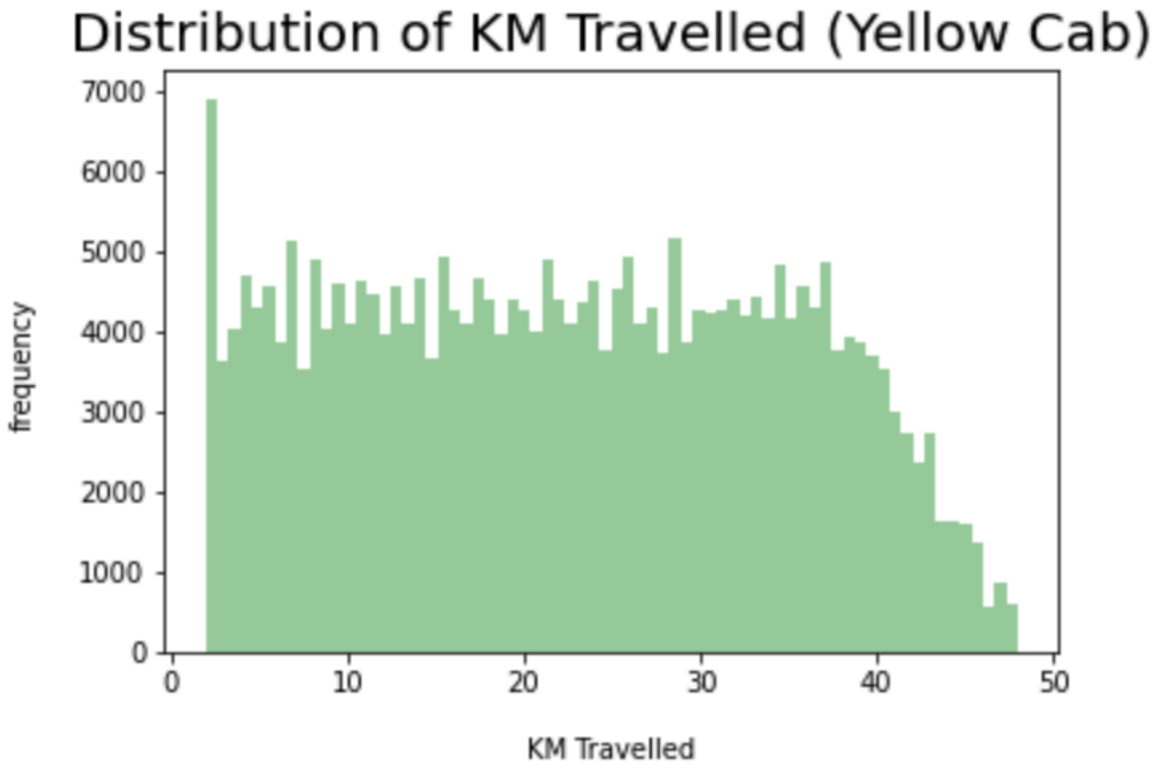




Yellow Cab has 35000 travels at the end of year. Pink Cab has more than 10000 travels at the same time.

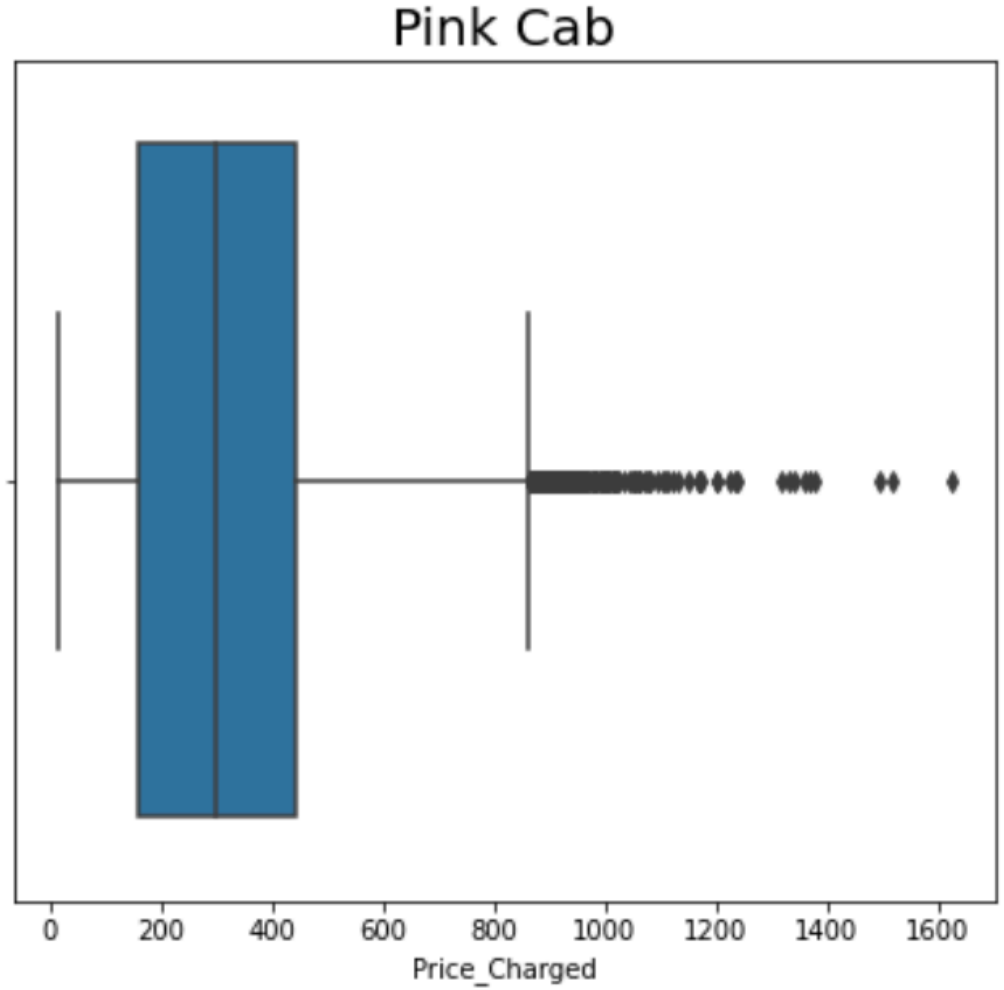
Distribution of KM Travelled

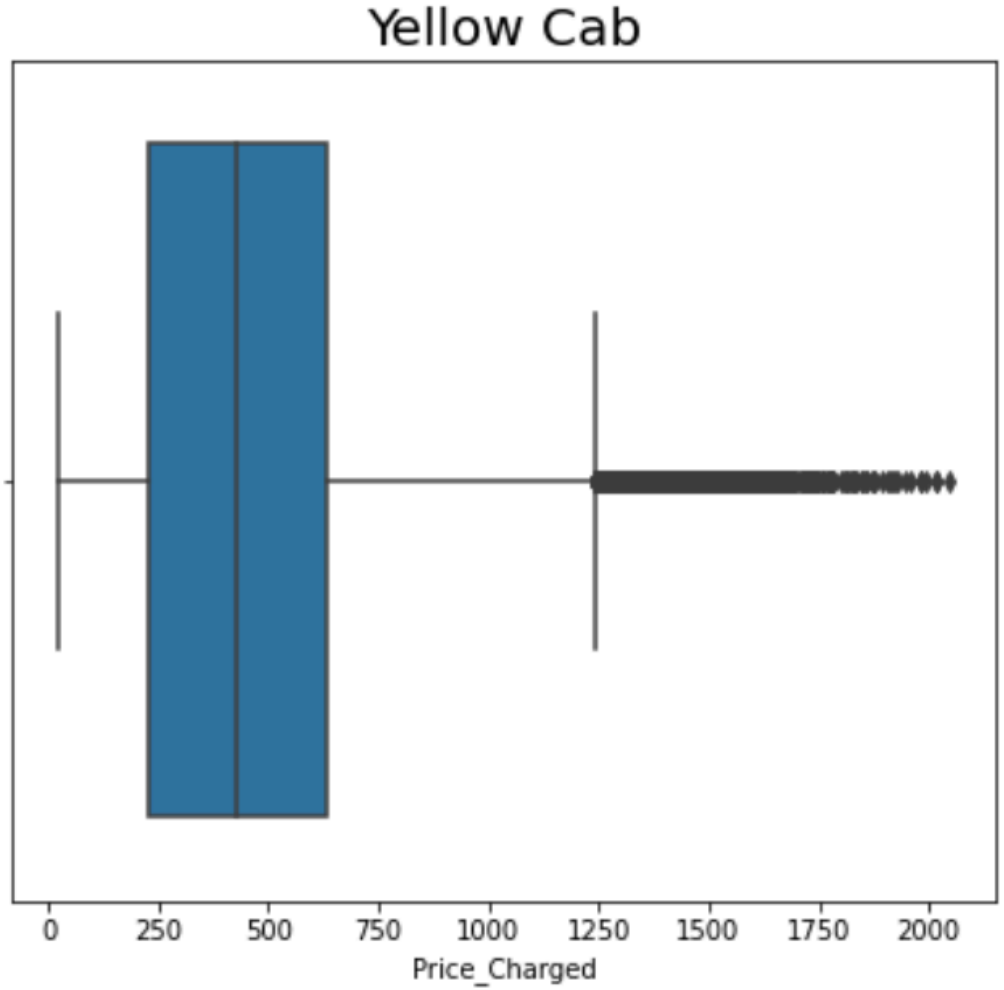




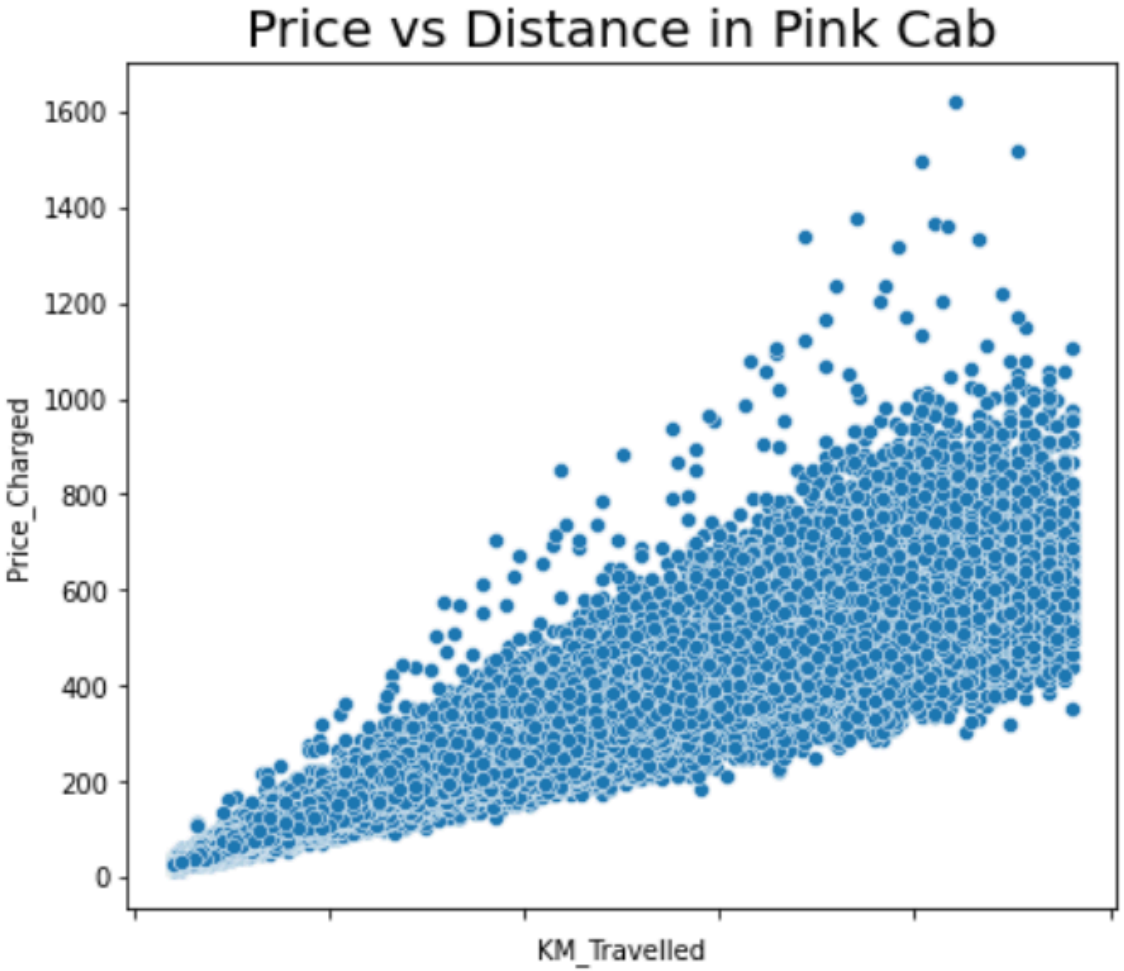
For Both Cabs -Pink and Yellow Cabs- have approximately similar range -2 to 48-.

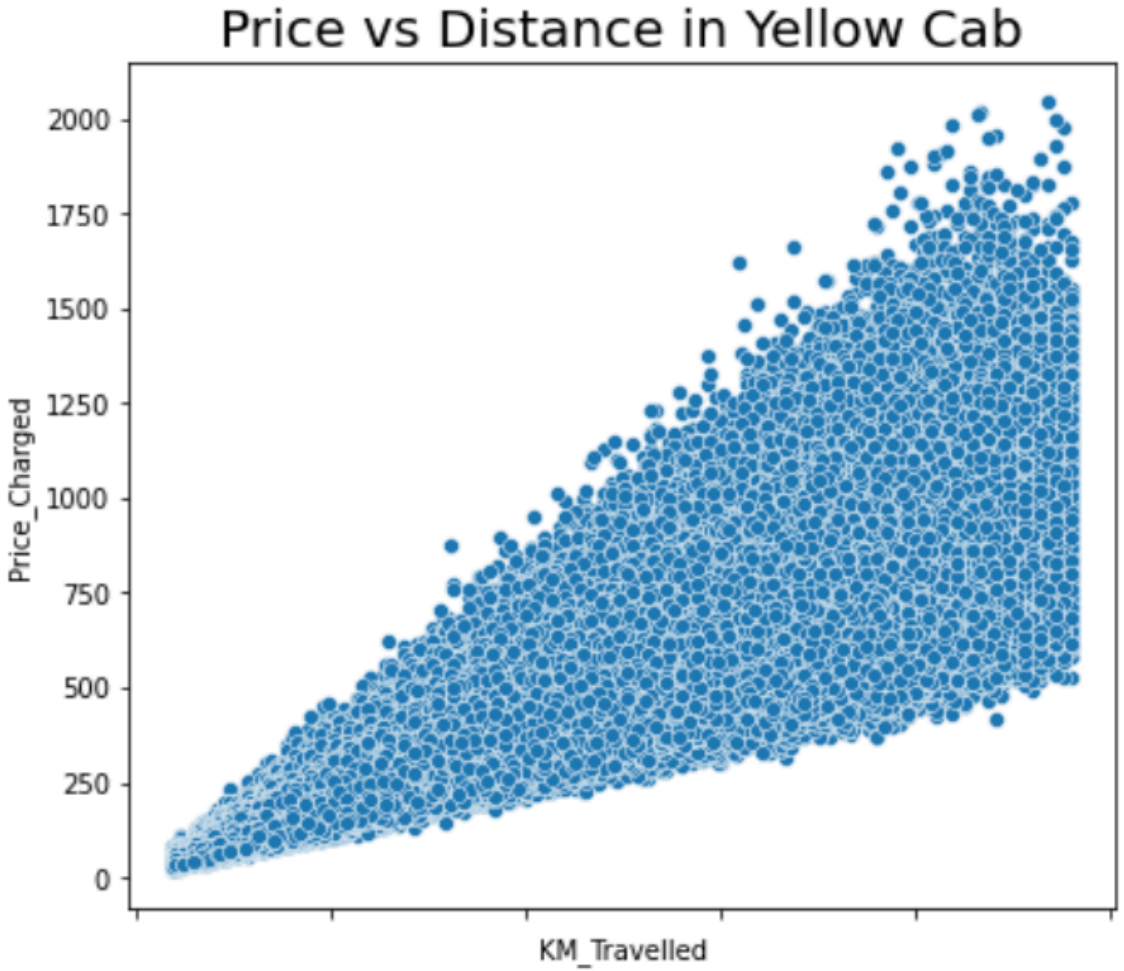
Distribution of Price Charged for both Cabs





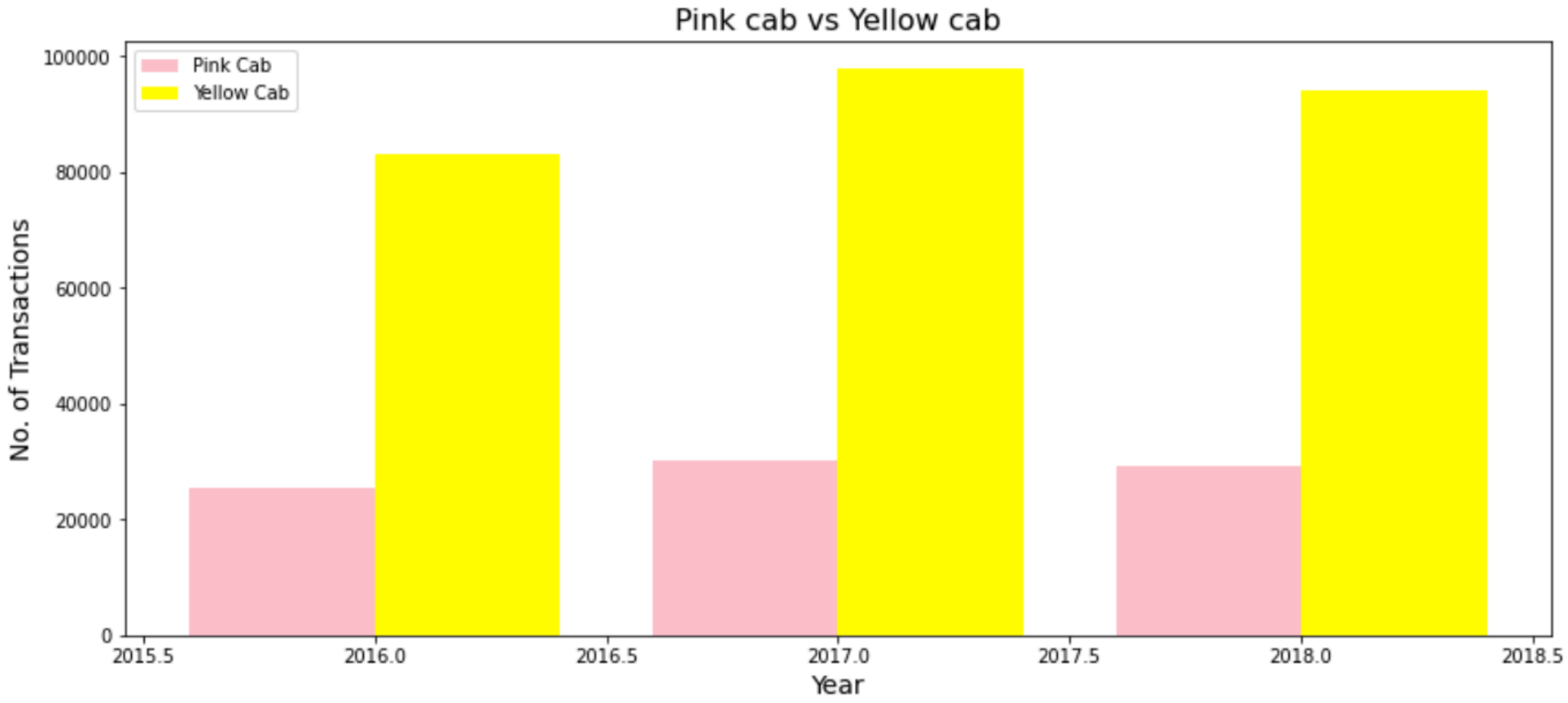
Outliers are shown at above graphs.





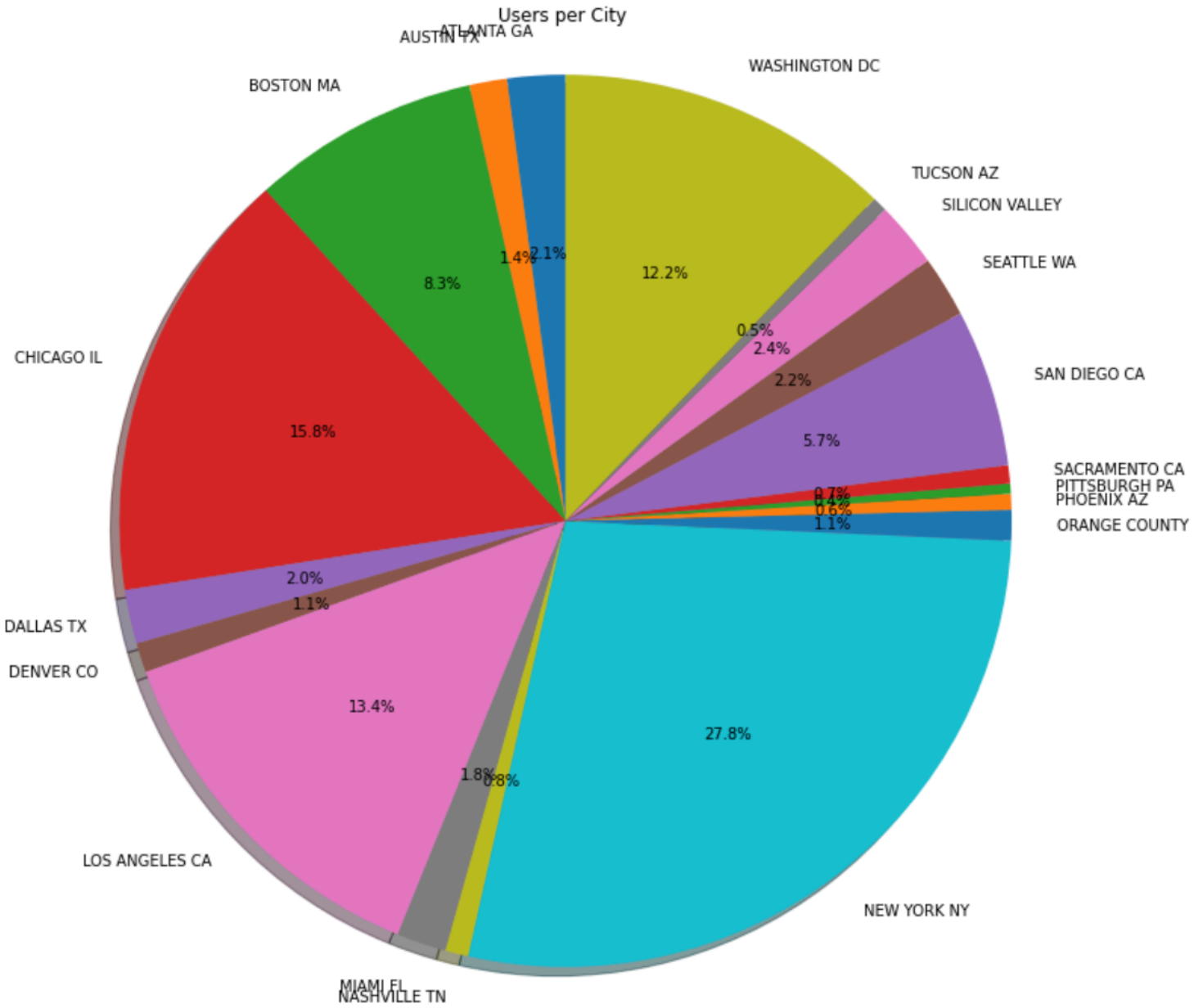
Price and Distance graph of Yellow Cab is more than Pink Cab’s.

Transaction per Year of both Cabs



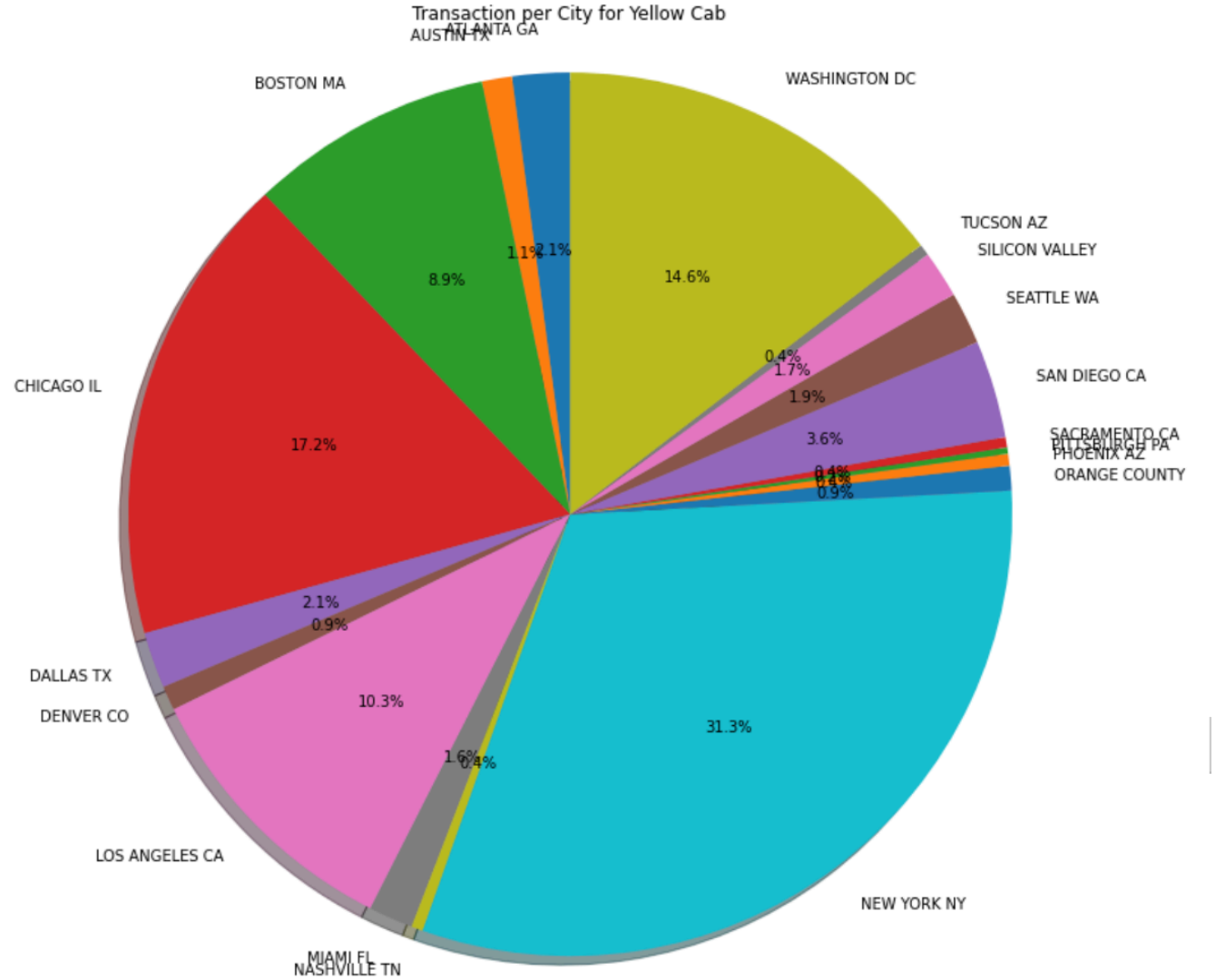
Obviously, this graph shows that Yellow Cab has more transaction per year.

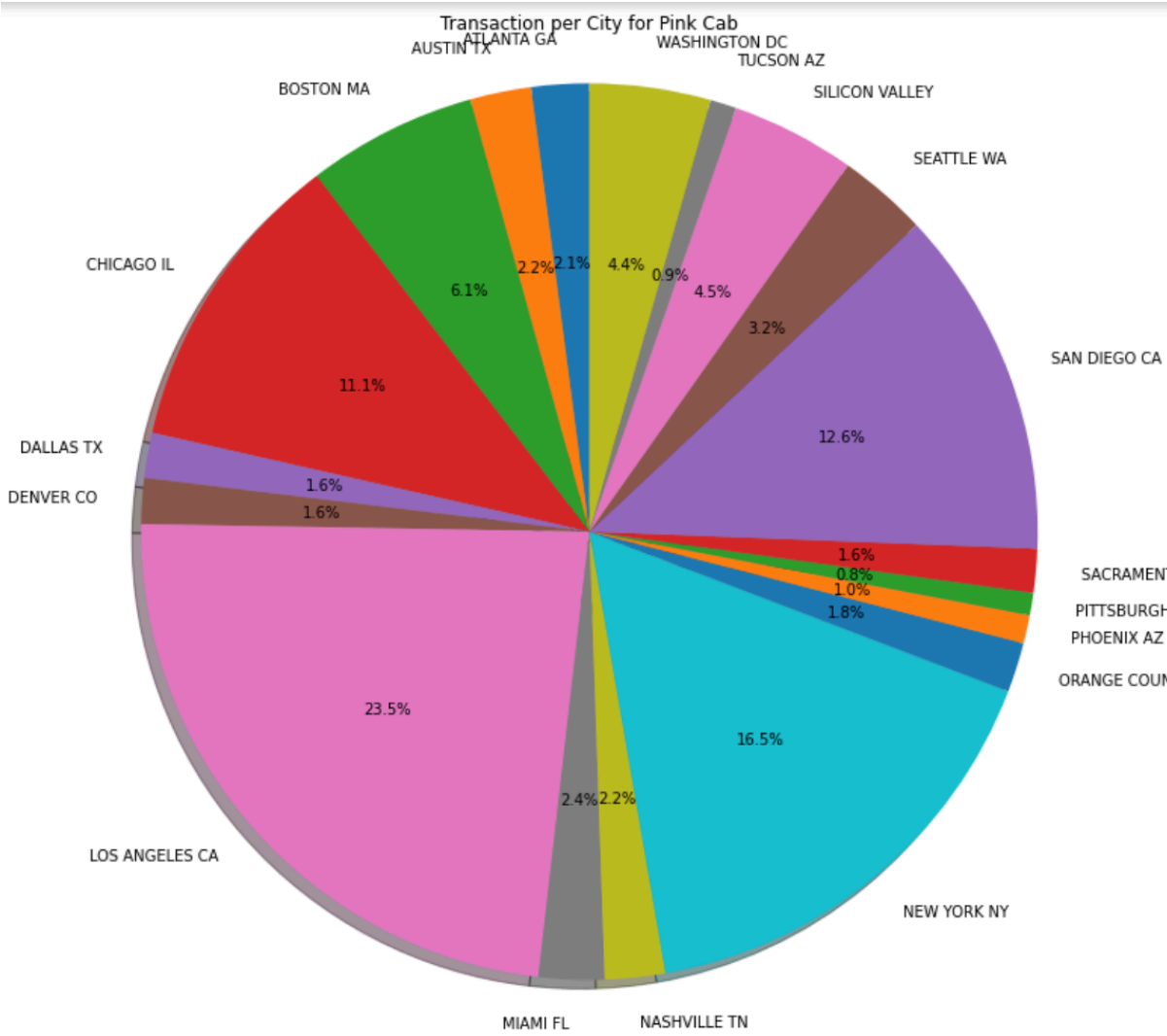
Users Per City



- New York City has highest(27.8%)
- Pittsburgh PA has lowest (0.4%)

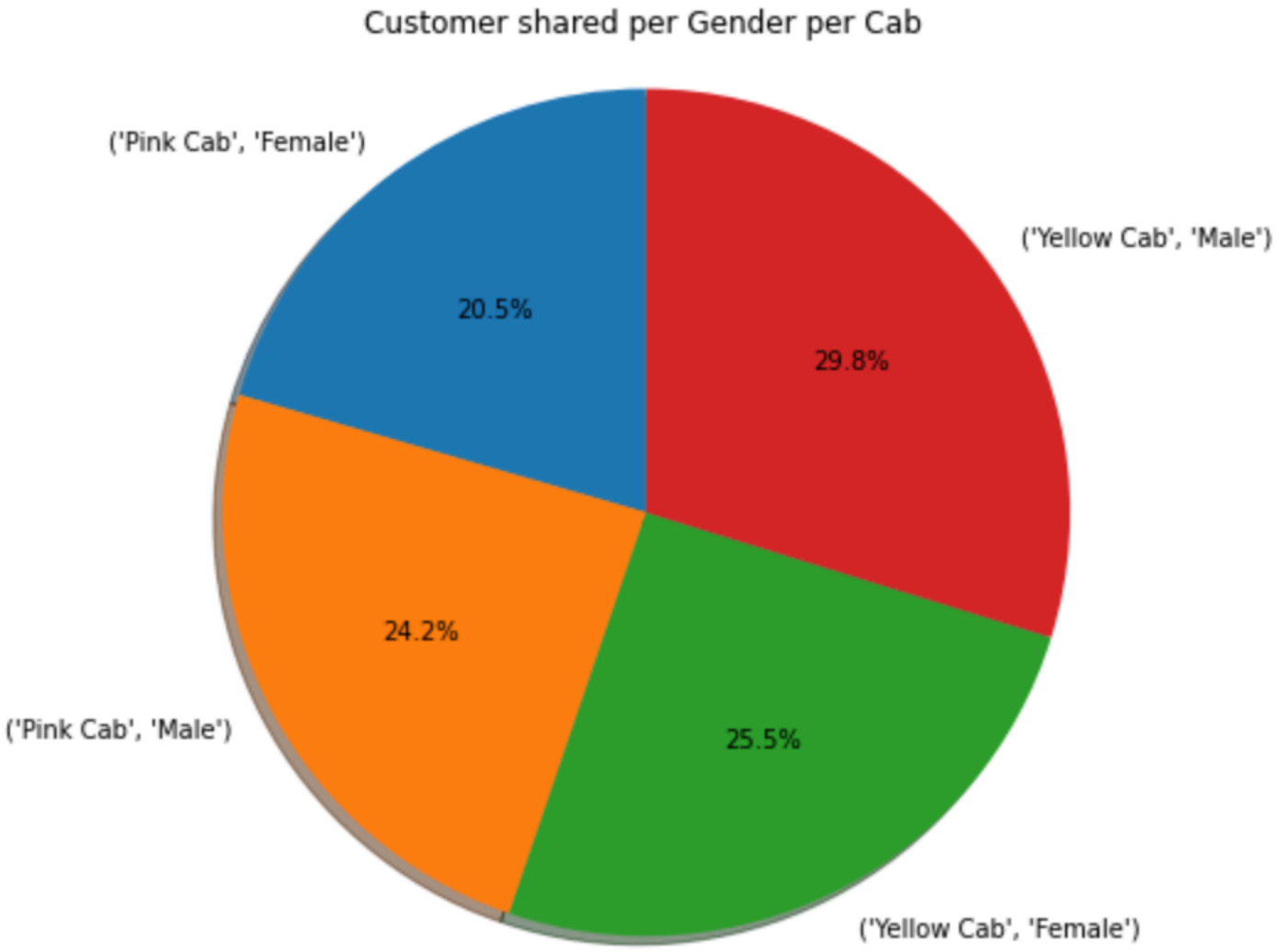
Transactions per City for Yellow and Pink Cabs





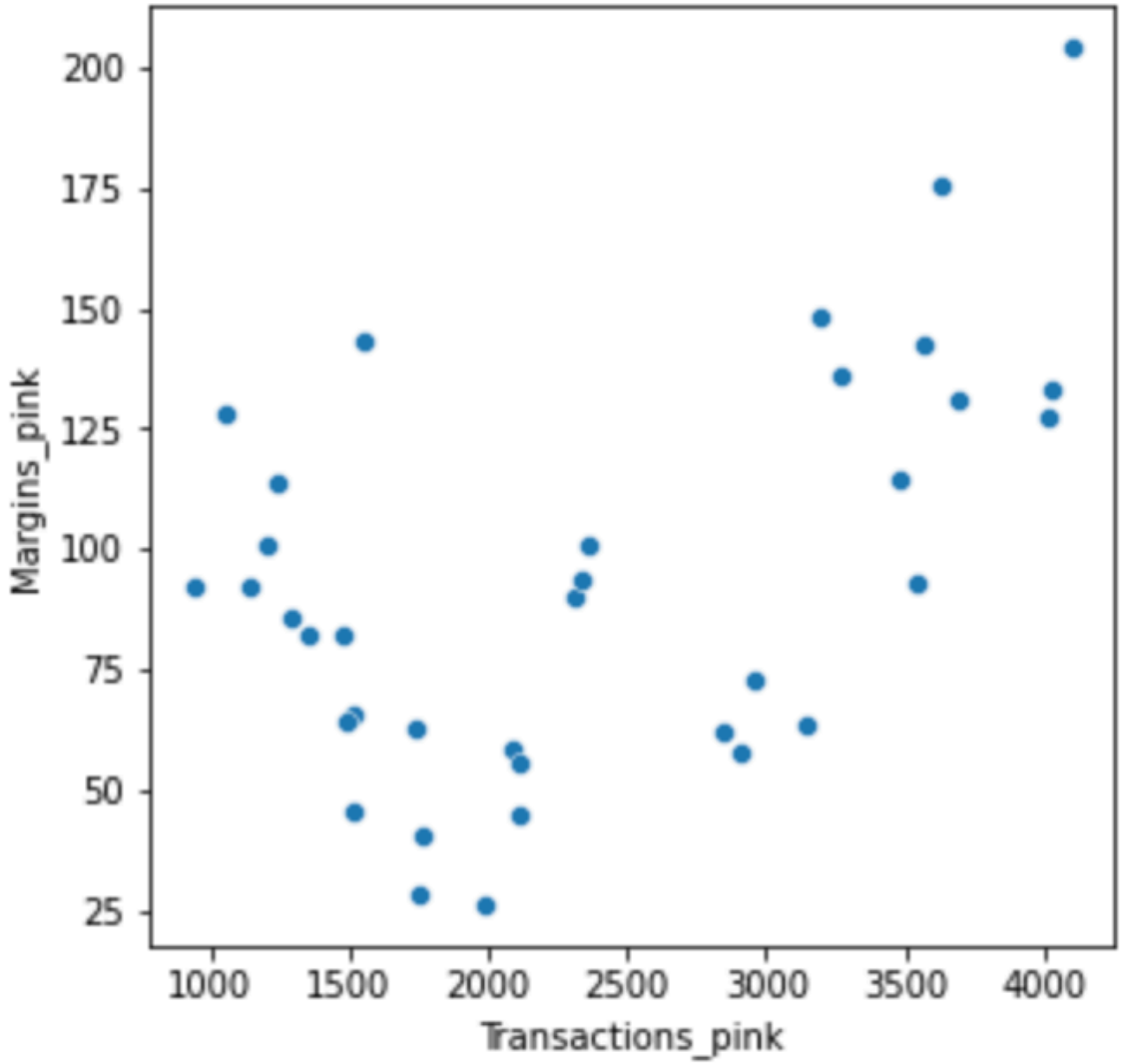
- New York City has 31.3% in the Yellow Cab. This is highest value.
- Los Angeles CA has 23.5% in the Pink Cab.

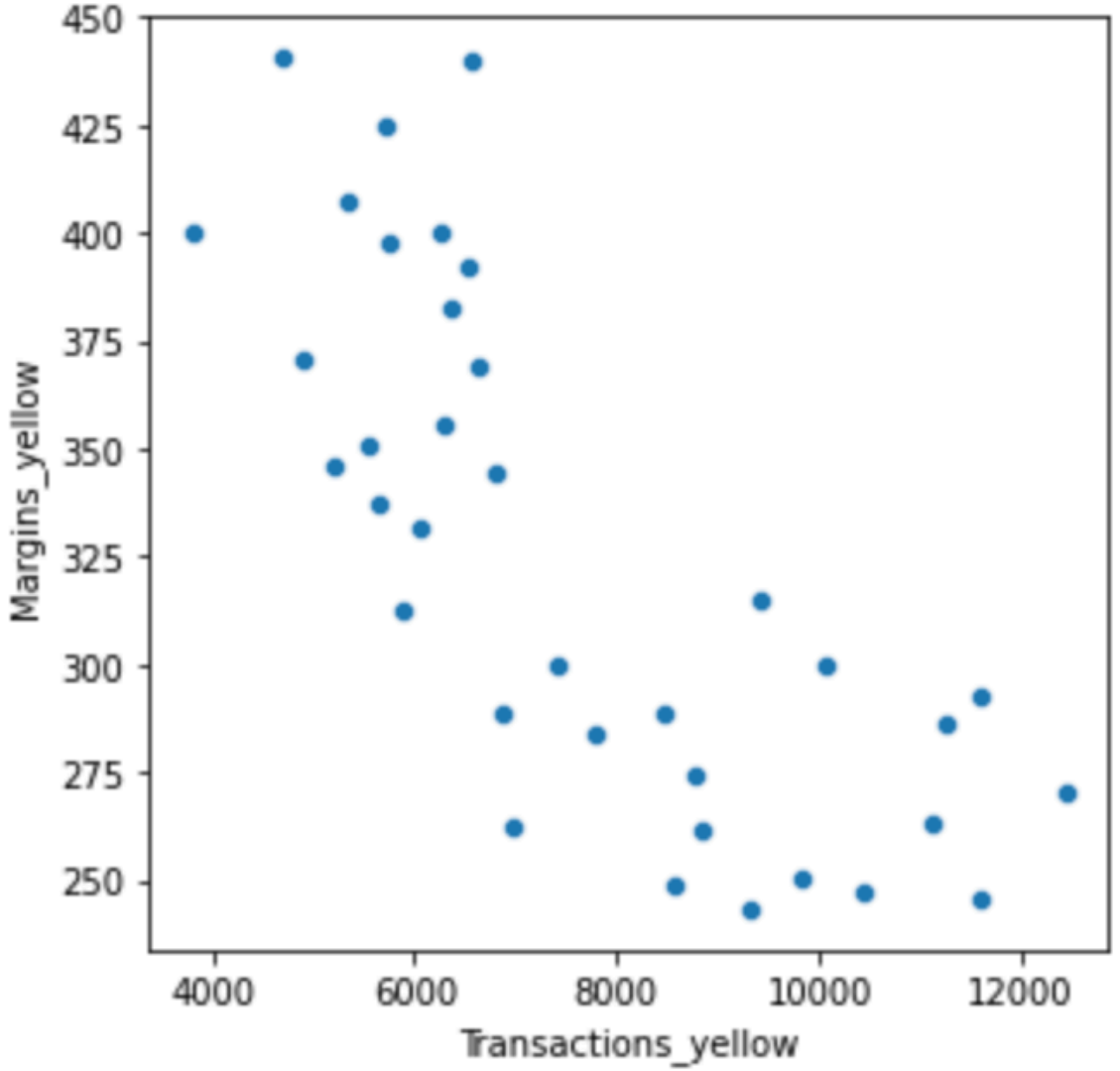
Genders Per Cabs



- For Both Cabs have similar results. Gender Male uses more than Gender Female.

Margins per Transactions





- **Pink Cabs increase margins with increase in number of Transactions.**

- **Yellow Cab decrease Margins with the increase in Transaction.**

Correlation

- This a positive correlation.

Thank You