



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

G2M insight for Cab Investment firm - Sali El-loh

09/21/2023

# Agenda

Executive Summary  
Problem Statement  
Approach  
EDA  
EDA Summary  
Recommendations

# Executive Summary

- Project Objective:  
XYZ seeks insights for investing in the cab industry. This analysis covers four datasets spanning January 2016 to December 2018.
- Key Findings:
- Company Performance: Cab Company Yellowcab outperforms Pinkcab consistently in users, revenue, and margin.
- Seasonal Trends: Cab usage exhibits seasonality with peak demand during certain periods.
- Margin vs. Customer Count: While more customers mean higher revenue, margin growth isn't always proportional.
- Customer Segments: Distinct customer segments exist with different preferences and behaviors

# Problem Statement

1. Market Analysis
2. Consumer Behavior
3. Competitive Landscape
4. Financial Considerations
5. Risk and Challenges
6. Exit Strategy

# EDA Approach and Summary

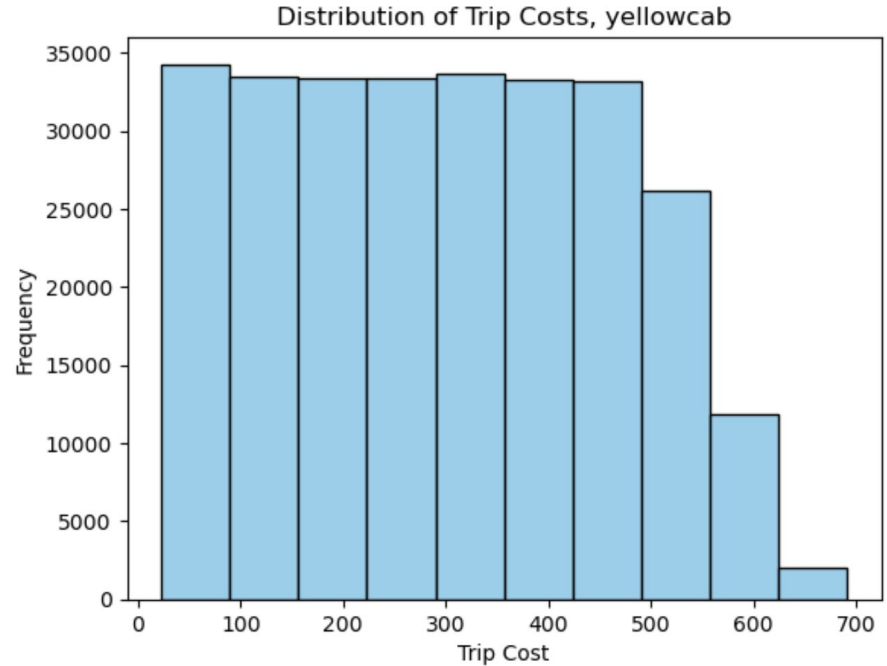
1. Data Ingestion
  - Gathered and reviewed provided datasets.
  - Understood data structures and relationships.
2. Data Preprocessing
  - Cleaned and transformed data.
  - Handled duplicates, missing values, and outliers.
3. Hypothesis Generation
  - Formulated hypotheses for investigation.
4. Exploratory Analysis
  - Analyzed cab company data.
  - Explored user behaviors and trends.
5. Recommendations
  - Proposed investing in Yellowcab company
  - Identified seasonal patterns and customer segments.
  - Emphasized pricing and cost optimization.

# Data

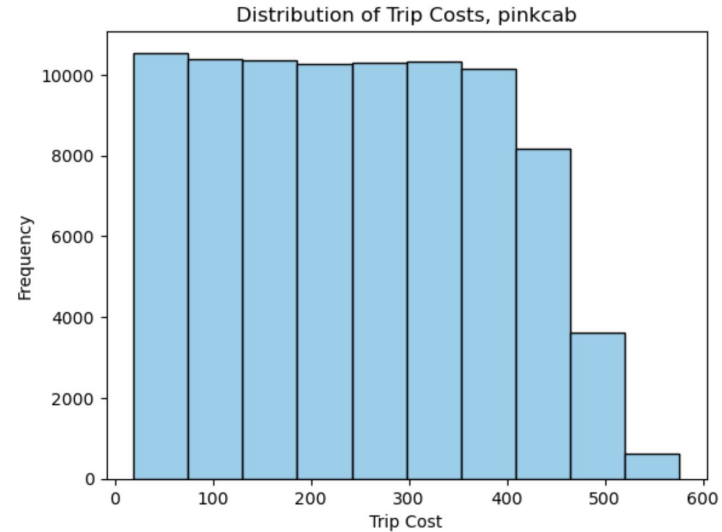
1. **Cab\_Data.csv** – this file includes details of transaction for 2 cab companies
2. **Customer\_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details
3. **Transaction\_ID.csv** – this is a mapping table that contains transaction to customer mapping and payment mode
4. **City.csv** – this file contains list of US cities, their population and number of cab users
5. Time period of data from 31/01/2016 to 31/12/2018.

# Analyze Each File Separately: Cab Data

Trip cost distribution for yellowcab

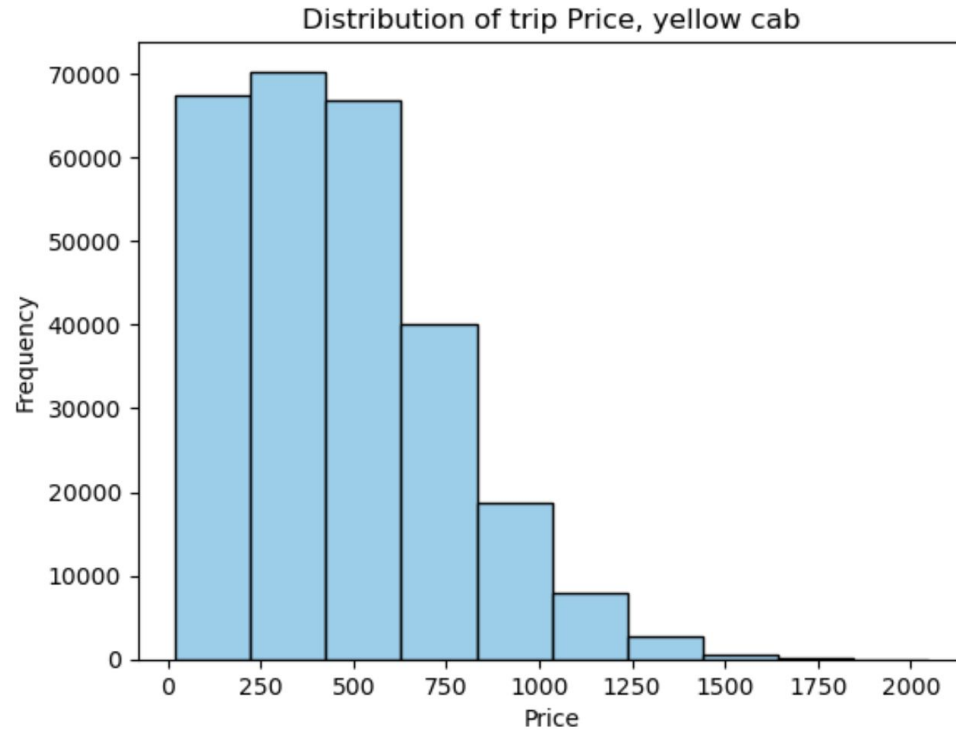


# Trip cost distribution for pinkcab



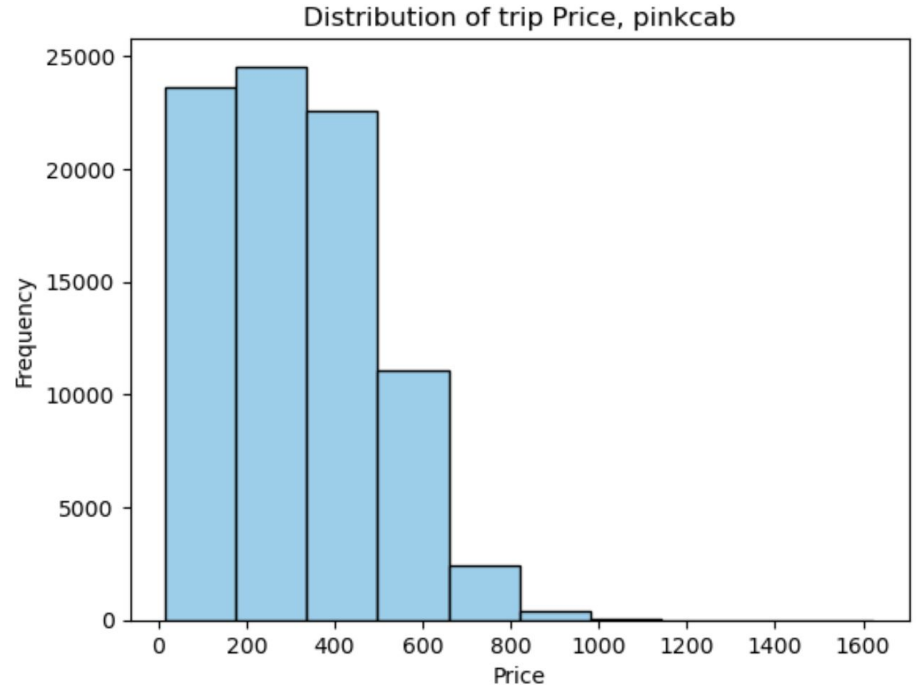


# Distribution of Trip Price, yellow cab



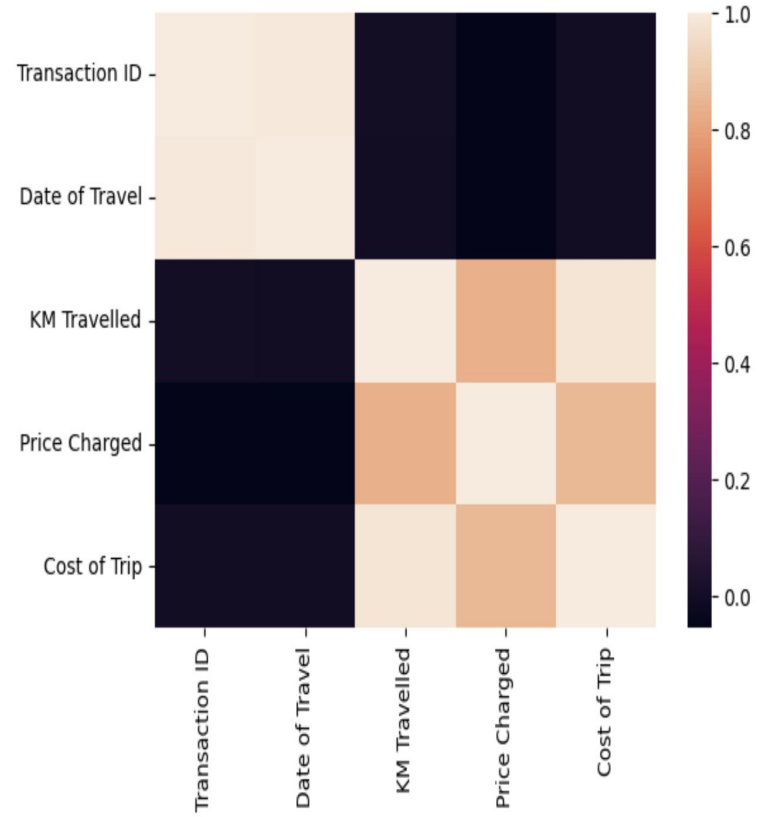
# Distribution of Trip Price, pinkcab

From this data, we can conclude that On average, the trip costs are the same for both companies, but yellow cab is charging much more for the trips. Also, yellowcab company has much more Transaction IDs, meaning it is picking up more people



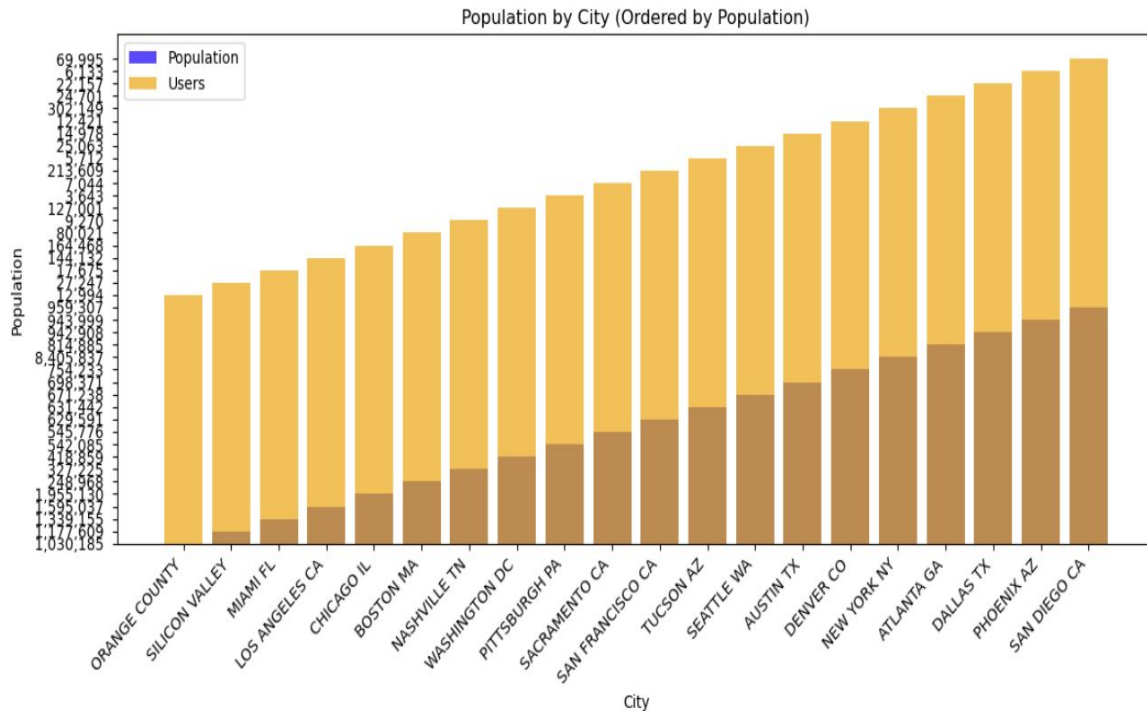
# Correlation Analysis for Transaction Data

On average, there is high correlation between the KM travelled and the price charged and the cost of the trip



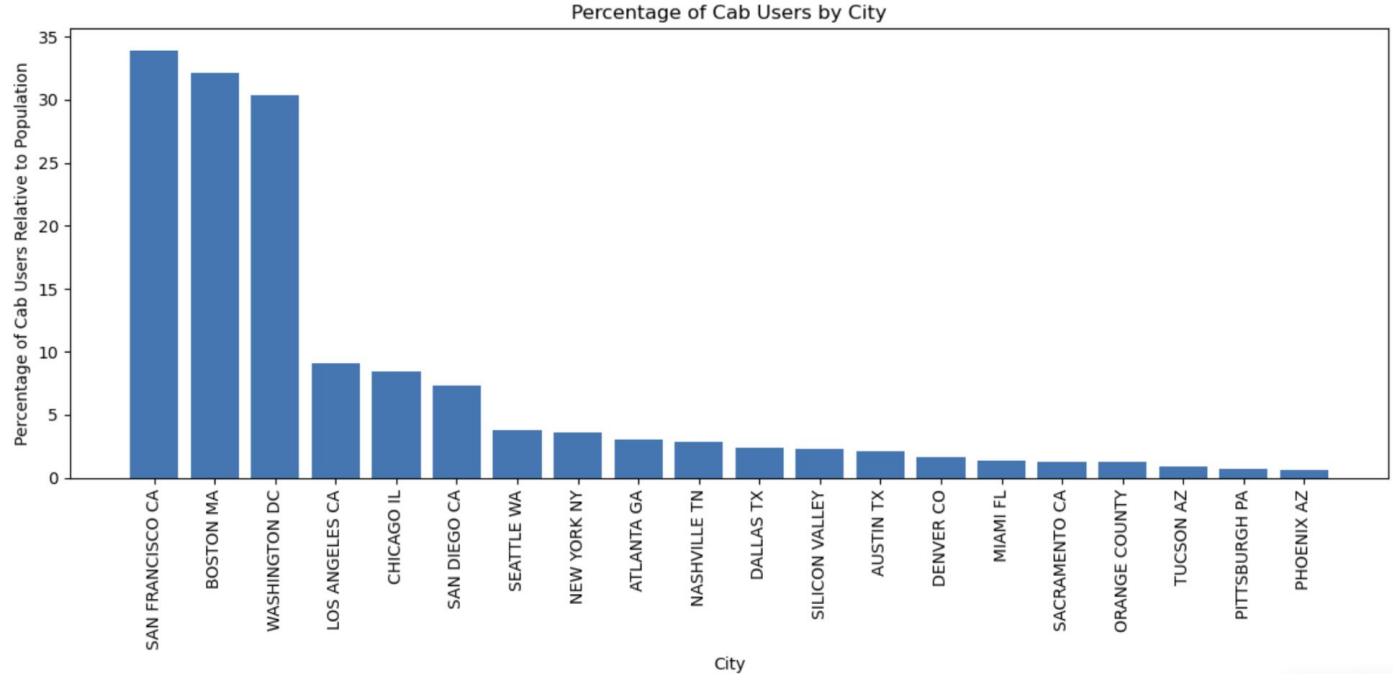
# Analyze City Data

This is a visualization of the users relative to the population

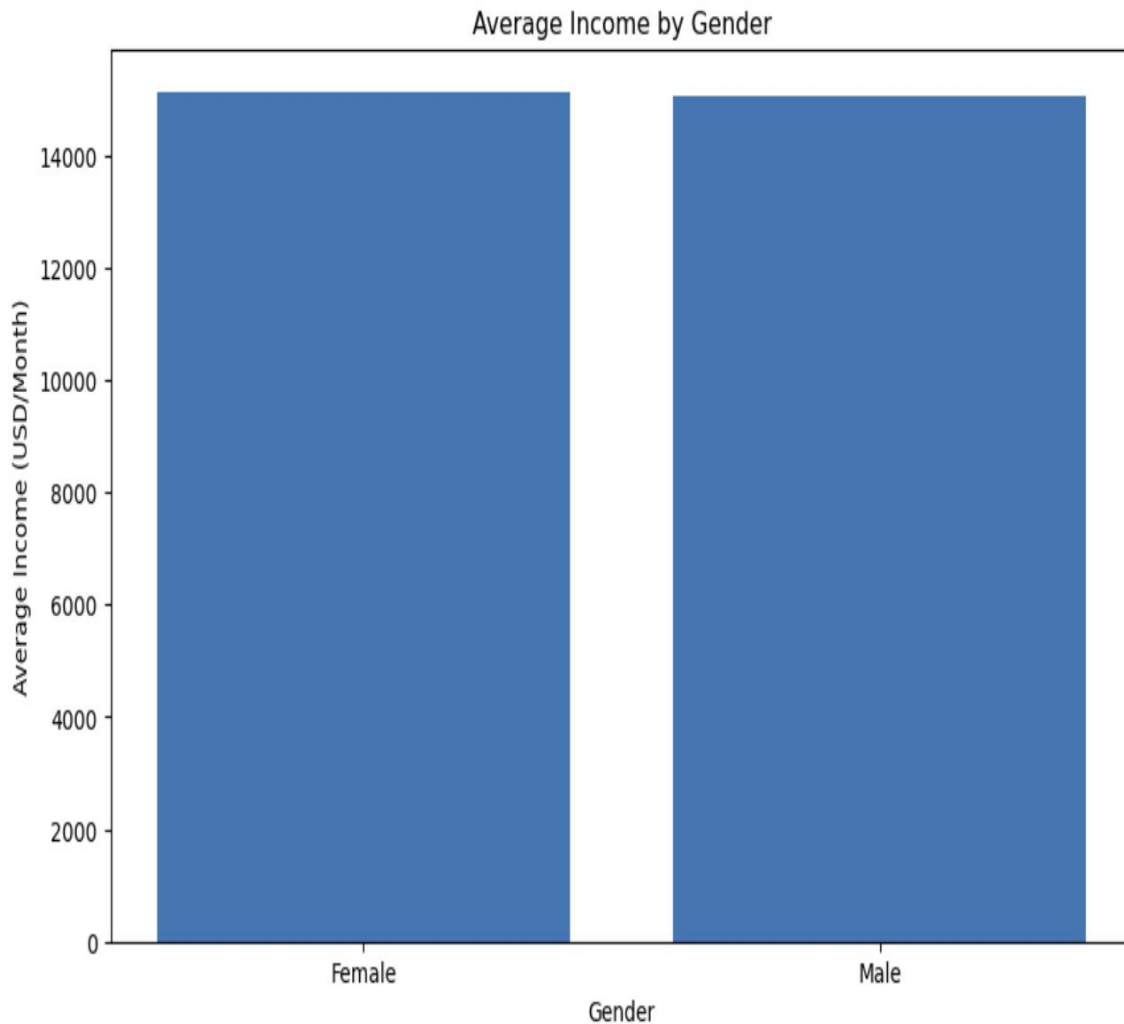


# Percentage of cab users relative to population

The percentage of cab user is highest in the cities san francisco, boston, and washington DC

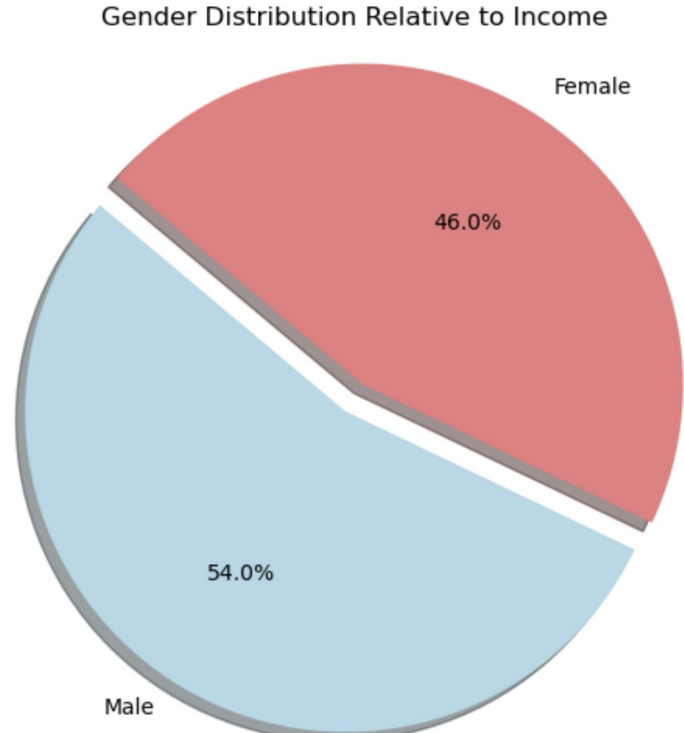


The average income for each gender is the same, so you can target both genders equally



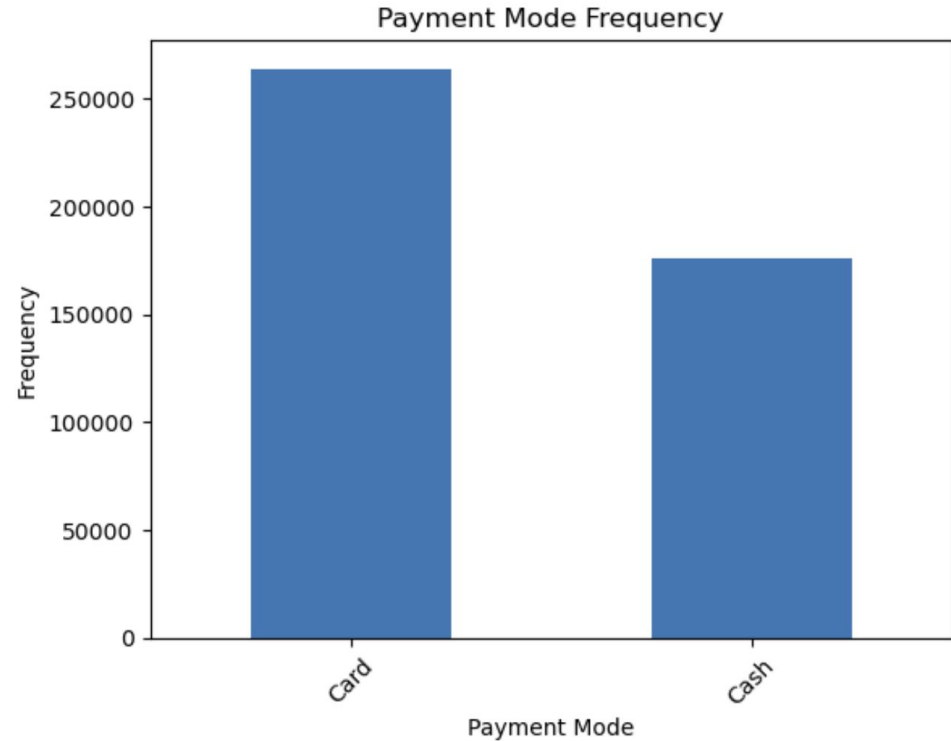
# Analyze Customer ID Data

The average income is similar, but the distribution of male to female is different. This shows that men are more likely to use cabs than woman



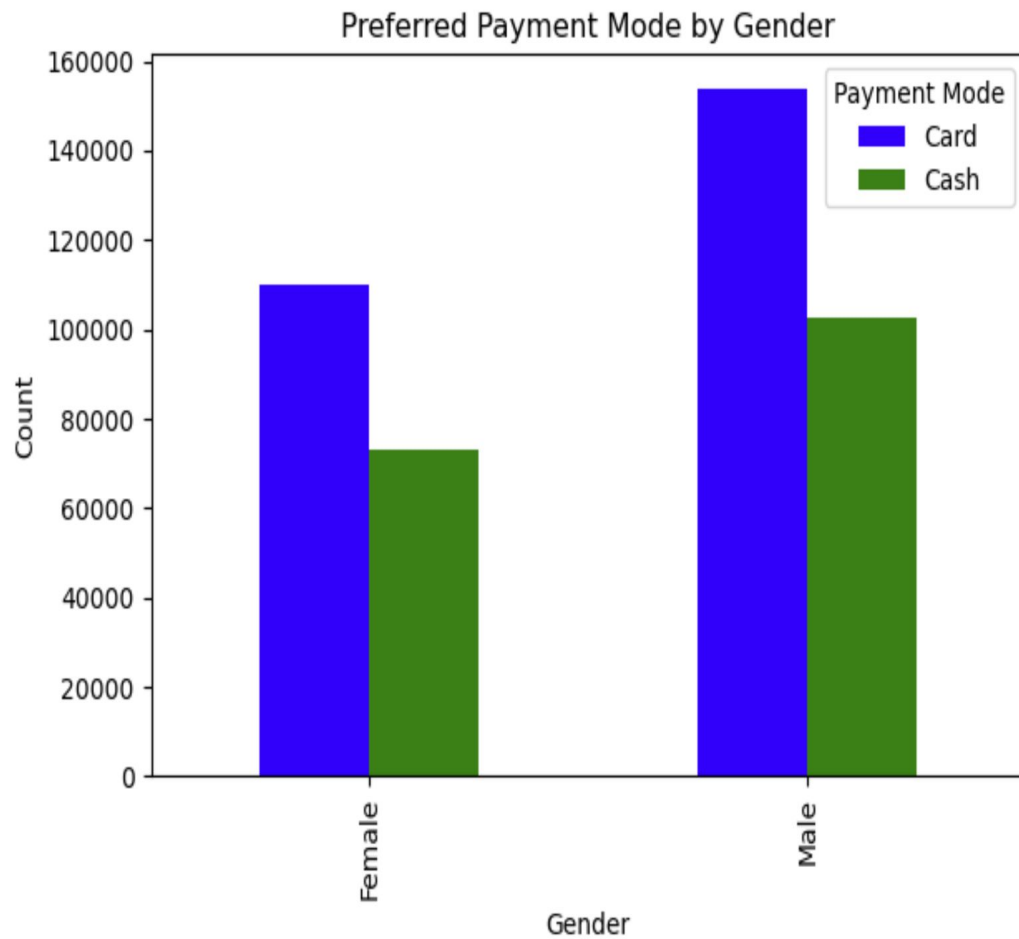
# Analyze Transaction ID Data

For all companies, the transaction mostly used is card.





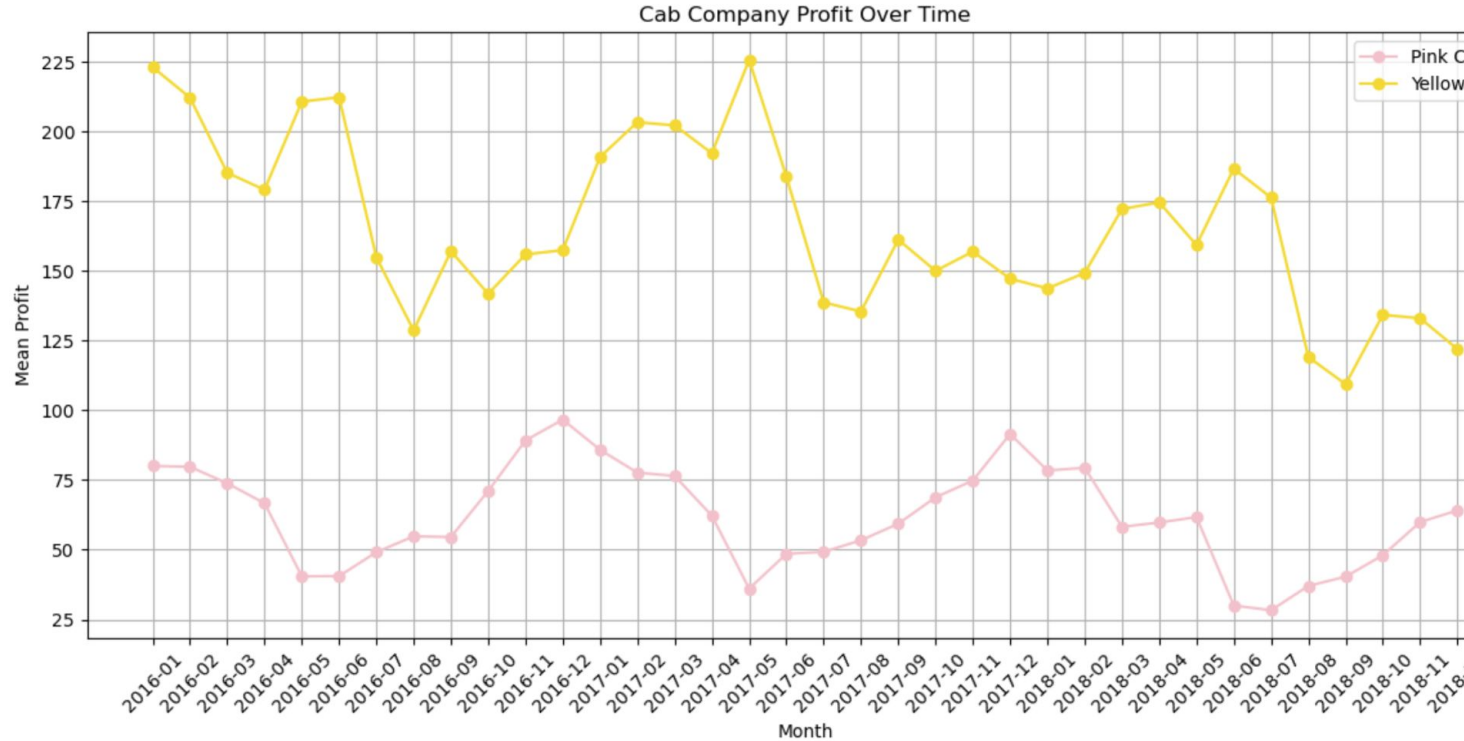
Both men and women  
prefer user card over  
cash



# Merged Data Analysis

# Profit Analysis

- Yellow cab company has a higher profit value on average per month
- Yellowcab and pink cab had opposite profit amounts for the same months, which shows rivalry
- Pinkcab is more consistent in its profit while yellowcab is not



# Recommendations

What company you invest on depends if you want more risk or safer side, yellowcab is more risky due to its inconsistency in profit gain, but it has on average higher profit when compared to pinkcab

Concentrate your services in cities like San Fransisco, Boston, and washington DC, since the can user to population ratio is the highest, meaning gaining more profit

Target your trips towards more male dominated events, such as sports arenas etc. as they are more likely to take a cab than woman, and since the income is similar, there is no profit loss

Make sure you are up to date with technology, as more people are likely to use card over cash when paying for their trips, also both Men and women both prefer cards over cash,

# Sali El-loh

## Exploratory Data Analysis

## Thank You