Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Cab Investment Firm EDA Preprocessing

Prepared by- Mohammad Tohin Bapari

Data Analyst Intern, Data Glaicer

Date: 15 June 2024

- What to Execute
- What Are The Problems
- How I Approached
- Key Expletory Data Analysis
- What I Have Found
- My Recommendations

Data Glacier

Your Deep Learning Partner

Project Summery

- XYZ, a private firm in the United States, is considering investing in the rapidly growing cab industry. To make an informed investment decision, XYZ has commissioned an exploratory data analysis (EDA) project focusing on two key players in the market. The firm has provided multiple datasets that cover various aspects of the cab companies' operations and customer profiles over a period from January 31, 2016, to December 31, 2018.
- The objective of this project is to analyse given datasets to derive actionable insights that will guide XYZ in identifying the most promising cab company for their investment.
- The deliverable for this project is a comprehensive presentation to XYZ's executive team. This presentation will be evaluated based on the quality of analysis, the effectiveness of the visualizations, and the value of the insights and recommendations provided.

Data Glacier
Your Deep Learning Partner

What Are The Problems

- The cab industry has seen significant growth in recent years, with multiple companies vying for market share. XYZ aims to capitalize on this growth by investing in one of the leading cab companies. However, to ensure a sound investment decision, XYZ needs a thorough understanding of the market dynamics, customer behaviour, and the performance of the two potential investment targets.

- The problem at hand is to analyse the provided datasets to answer the following key questions:
  - What are the key performance metrics of the two cab companies?
  - How do the customer profiles and behaviours differ between the two companies?
  - Which company shows better financial performance and growth potential?
  - What are the significant trends and patterns in the cab industry that could influence the investment decision?

Data Glacier
Your Deep Learning Partner

Given Dataset

Data Sets

- I 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** – Overview of two cab companies.

➤ **Customer_ID.csv** – Demographic details of two cab companies' customers.

➤ **Transaction_ID.csv** – Mapping of transactions to customers and payment modes for two cab companies.

➤ **City.csv** – List of US cities with population and number of cab users.

Data Glacier
Your Deep Learning Partner

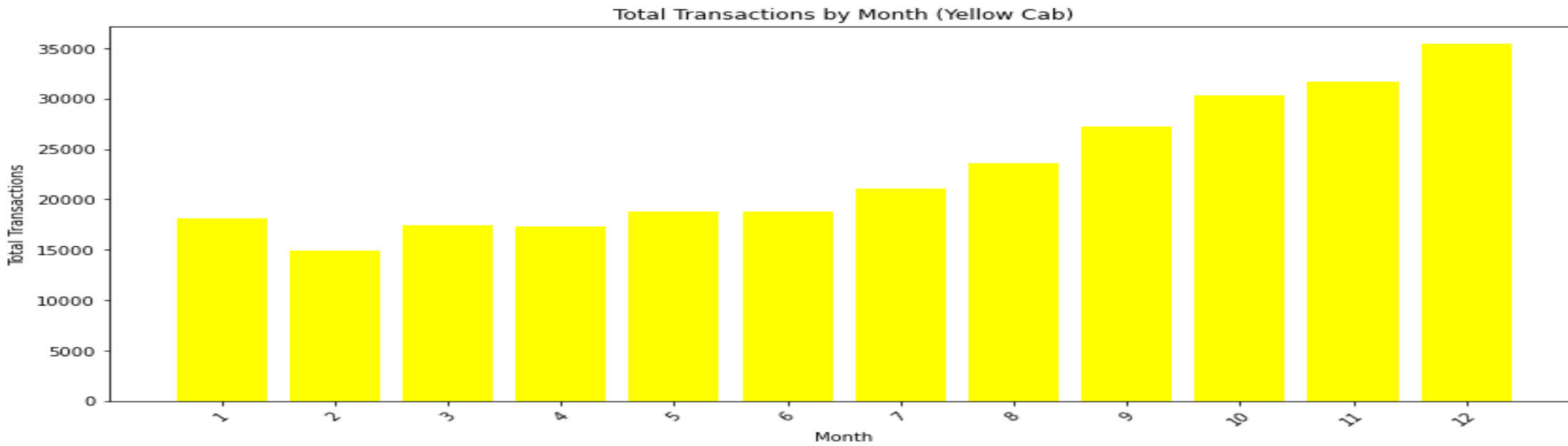
How I Approached

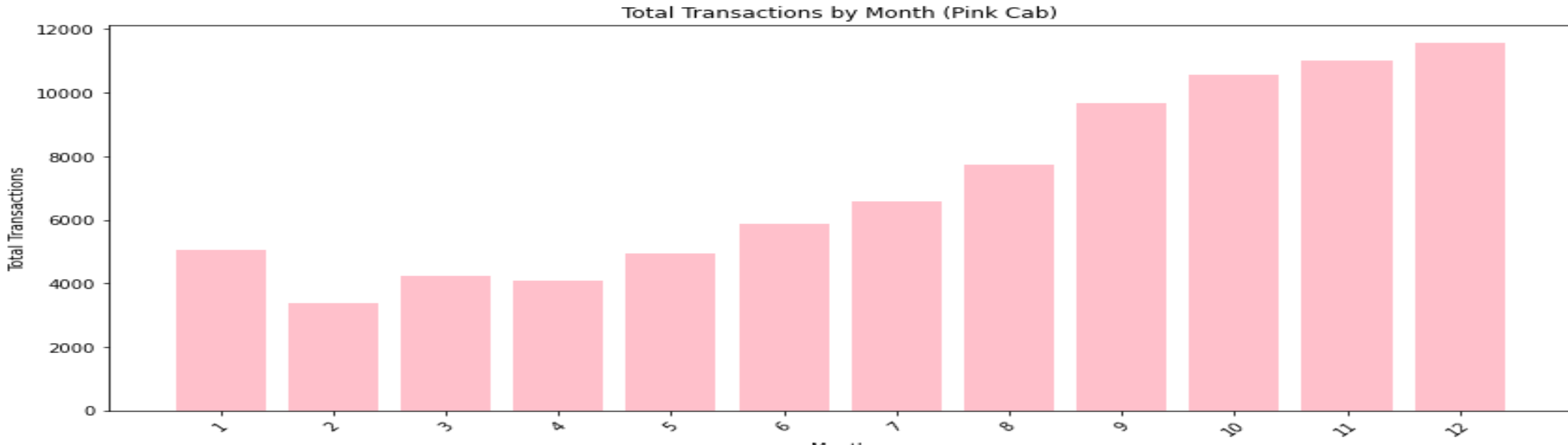
- Load Data: Load the datasets Cab_Data.csv, Customer_ID.csv, Transaction_ID.csv, City.csv into pandas DataFrames.
- Inspect Data: Examine the first few rows and data types of each dataset to understand their structure and content.
- Data Cleaning: Handle missing values. Correct data types if necessary.
- Data Integration: Merge the datasets to create a comprehensive dataset called master_data for analysis for both the cab companies.
- Data Extraction: Split the datasets to create a comprehensive dataset for analysis for both the cab companies.
- Exploratory Analysis: Descriptive statistics to understand distributions. Visualizations to identify patterns and insights.
- Feature Engineering: Create new features if necessary (e.g., customer lifetime value, average trip cost).
- Analysis for Recommendations: Identify key metrics and trends. Compare the performance of the two cab companies.
- Prepare Presentation: Create visualizations and summaries to communicate findings. Formulate recommendations based on the analysis.

Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

Total Transactions by Month

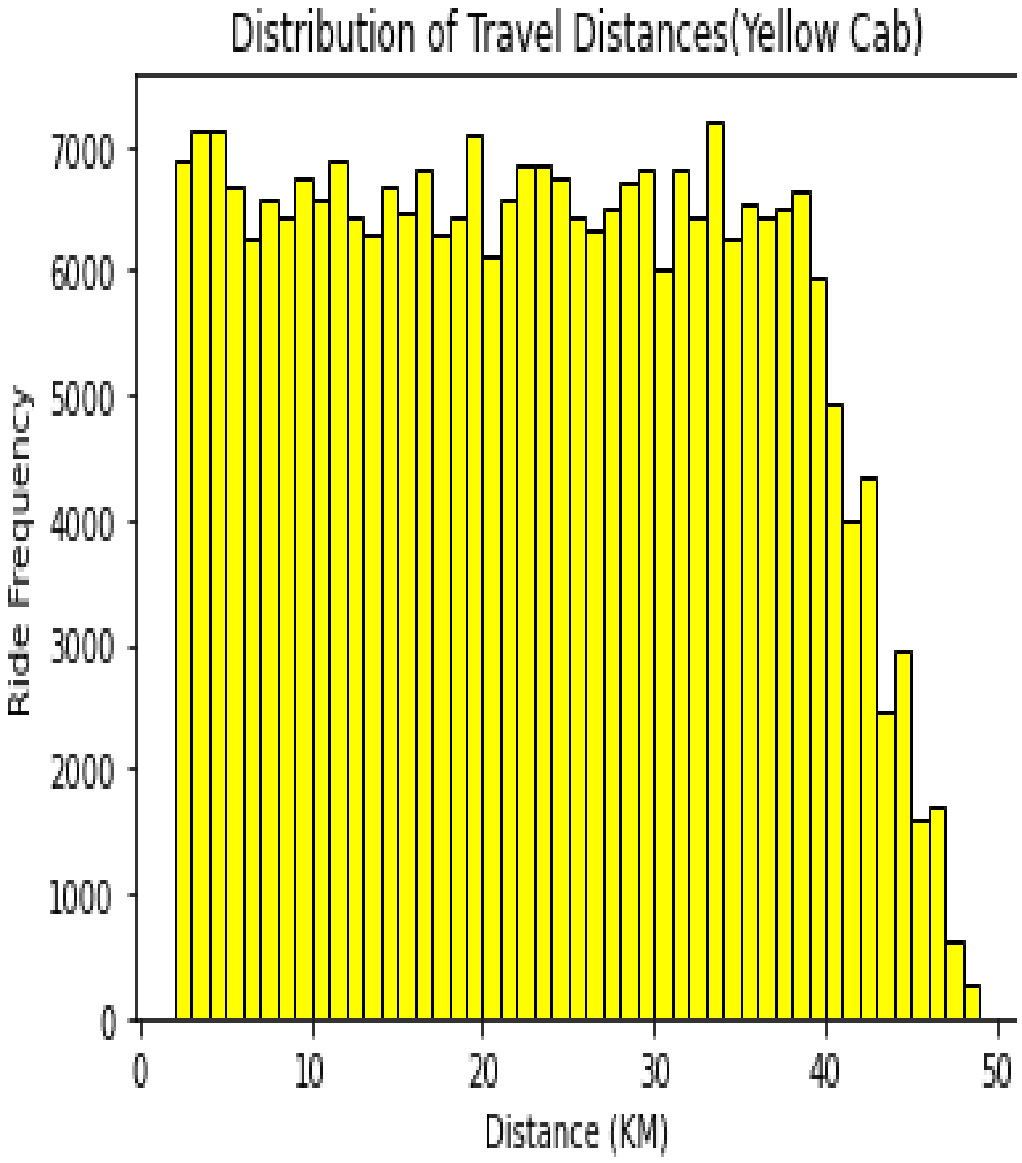


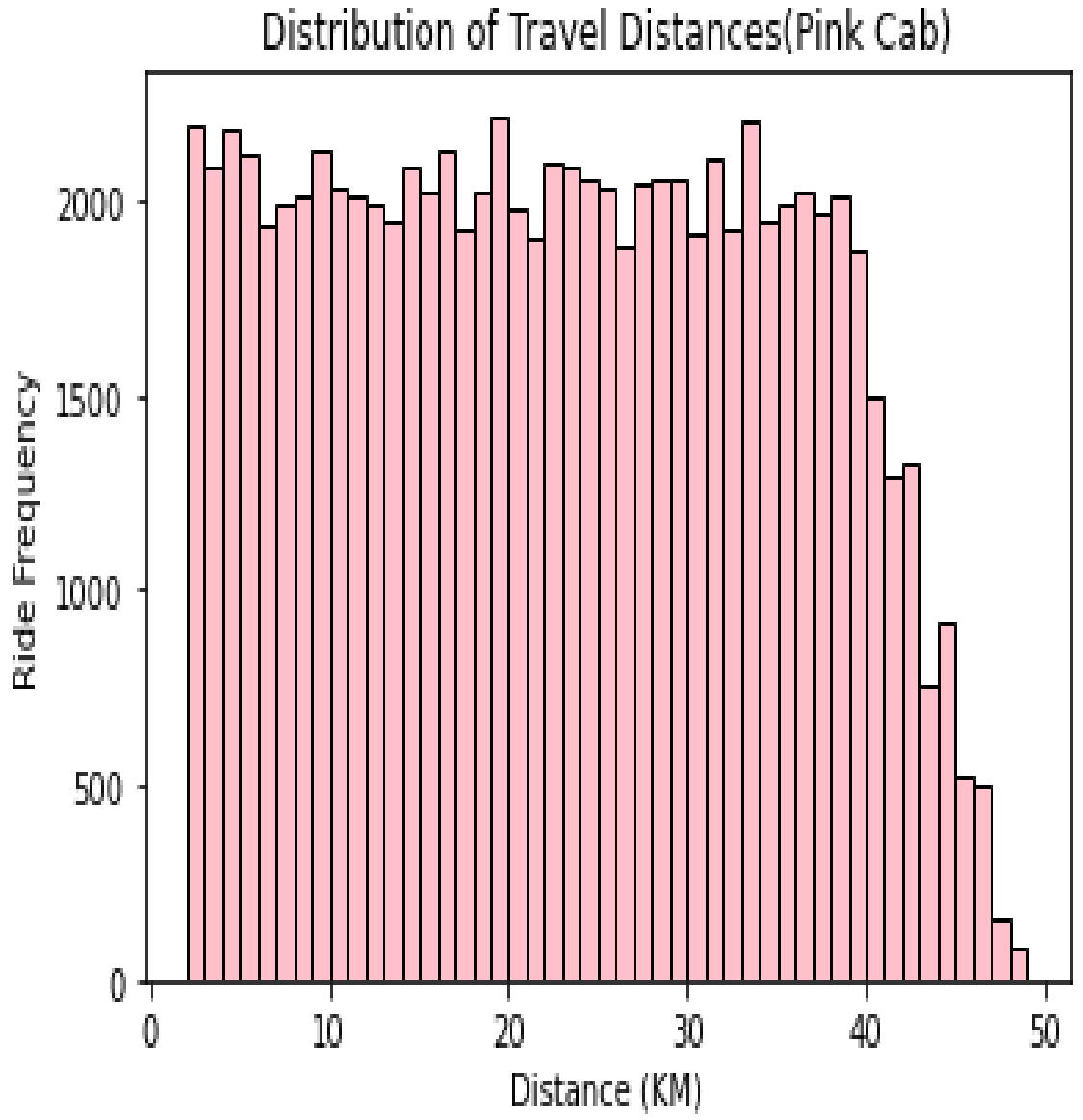


Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

Distribution of Travel Distances

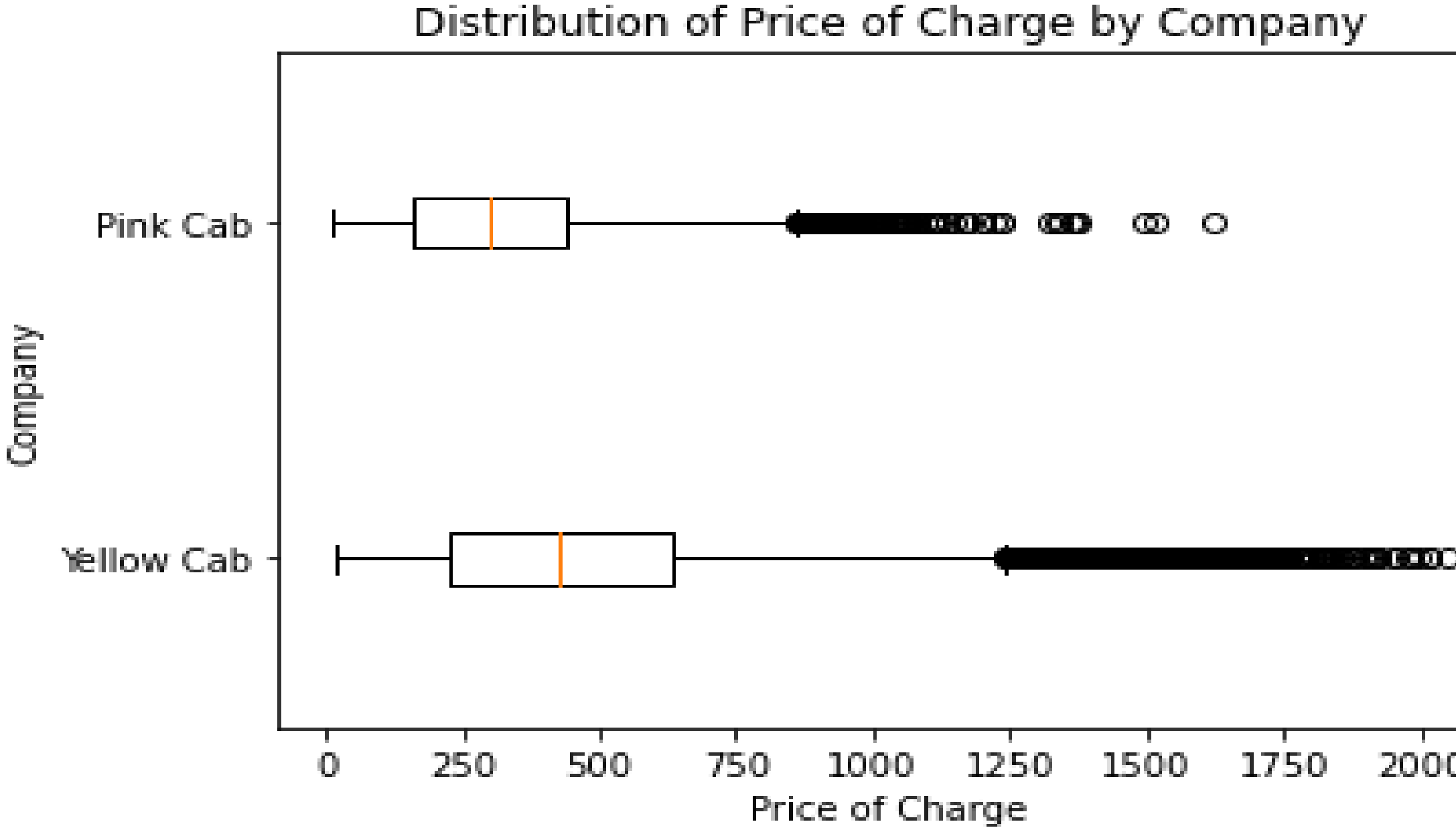




Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

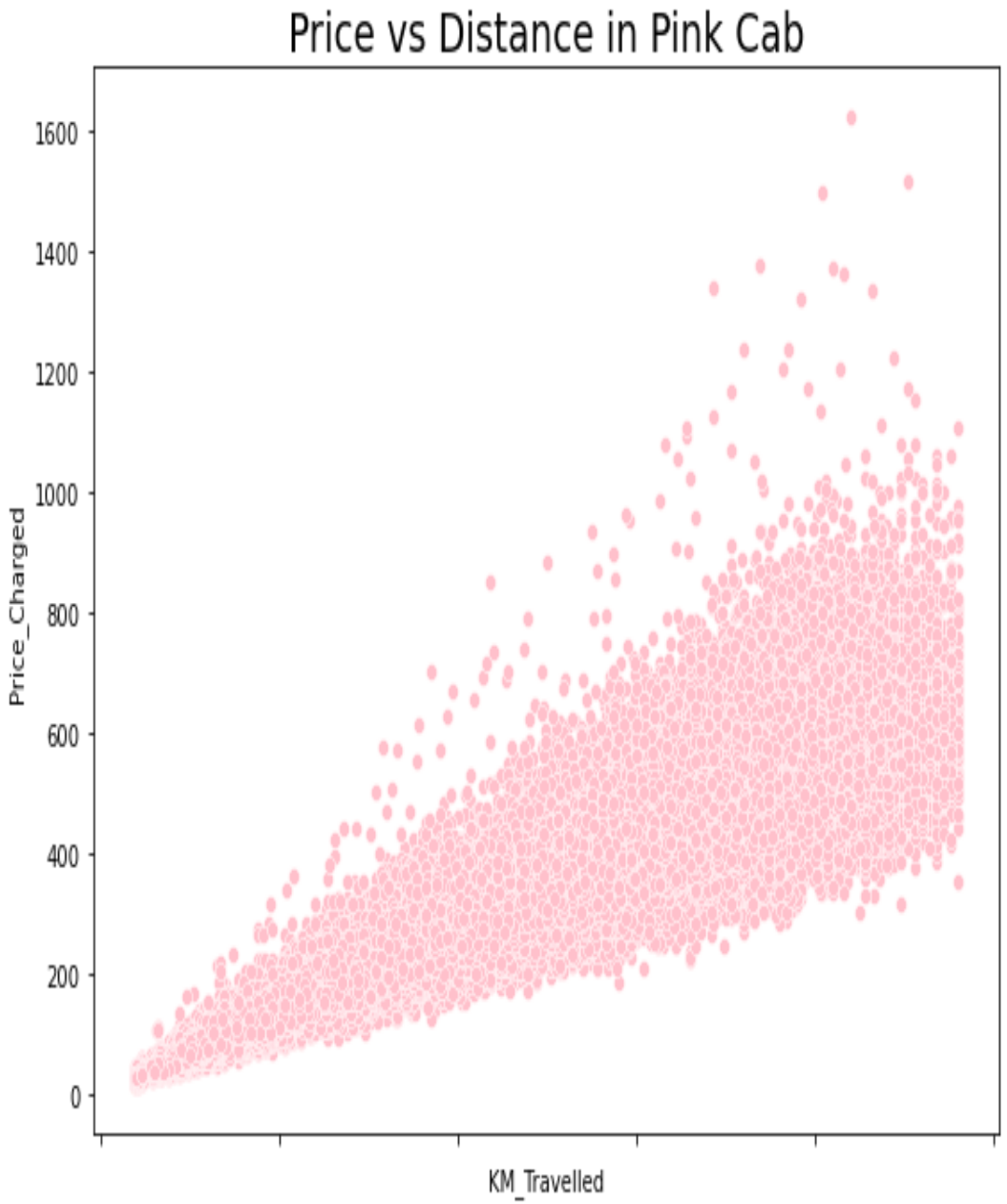
Distribution of Price of Charge by Compan

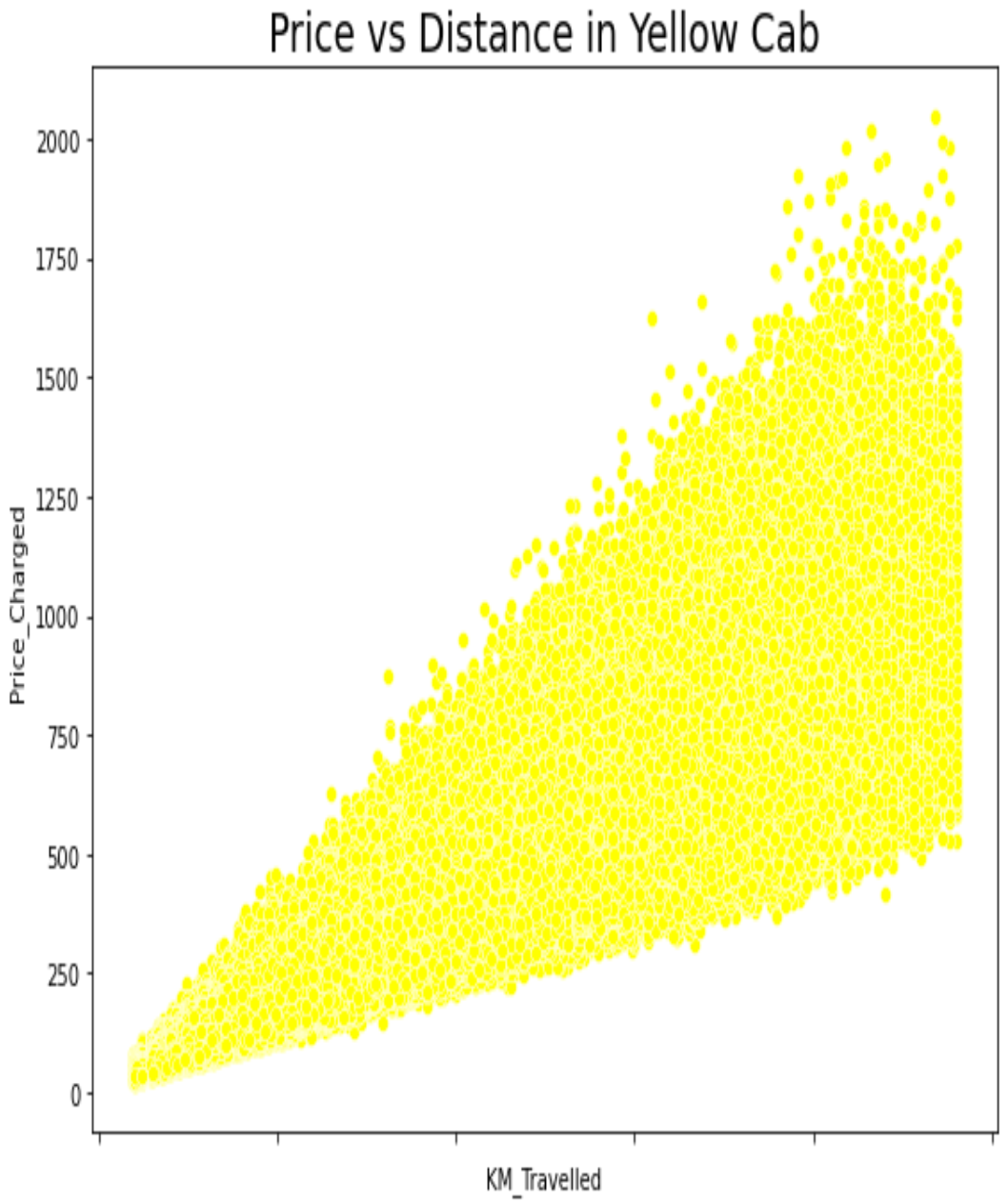


Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

Price vs Distance

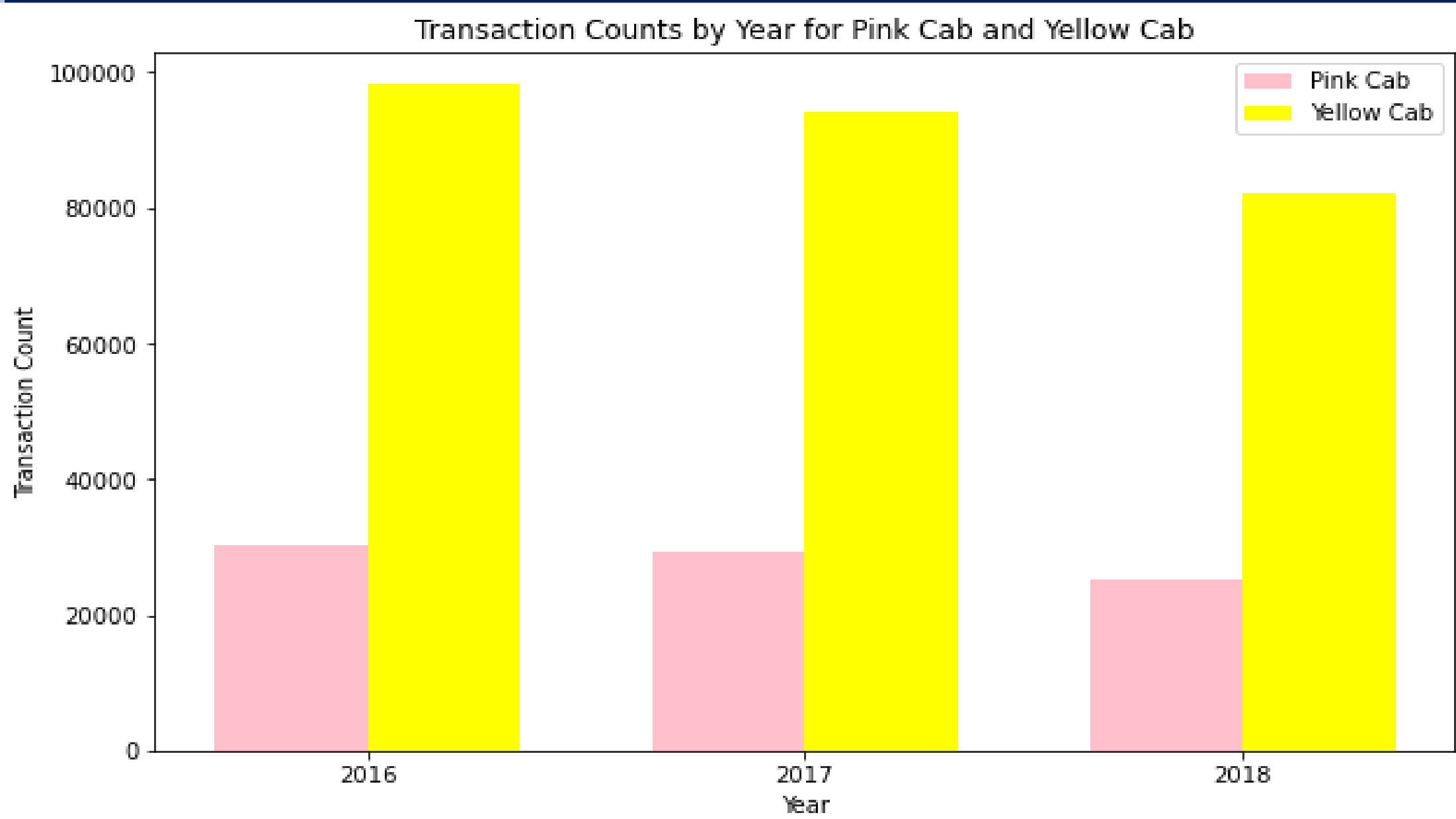




Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

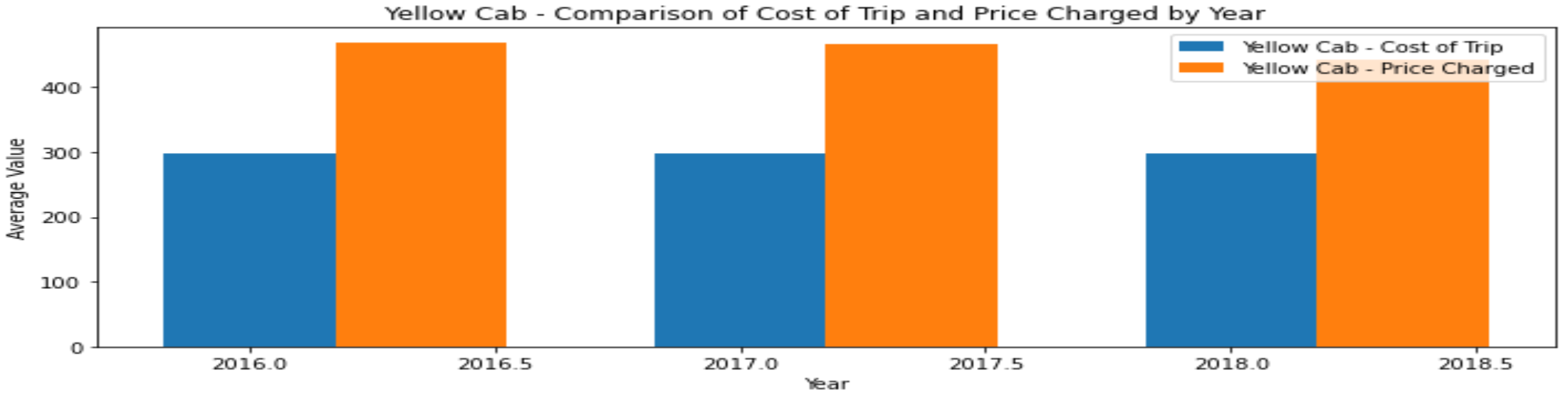
Transaction Counts by Year

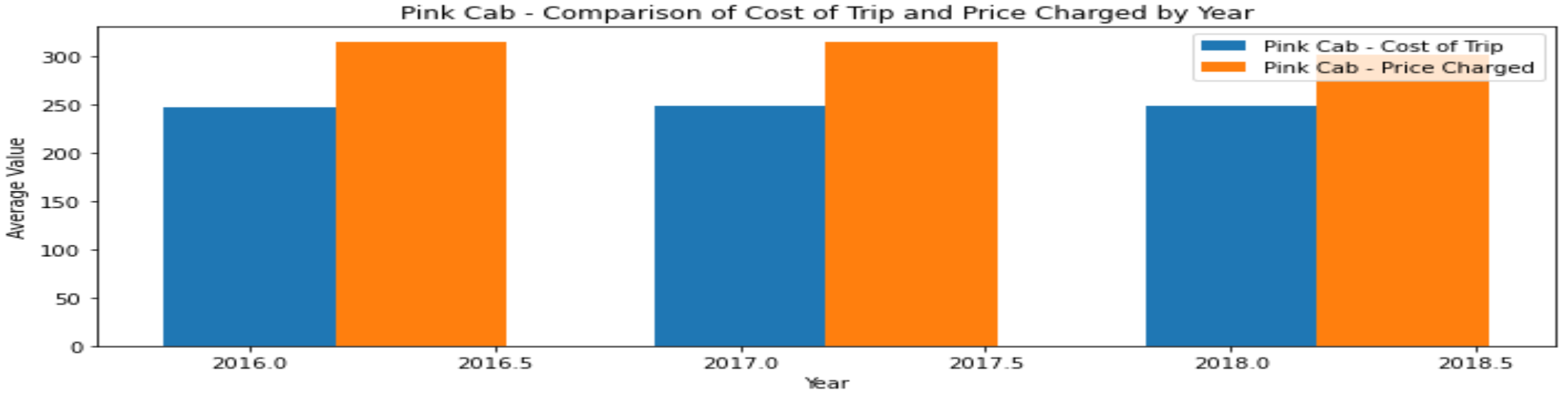


Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

Comparison of Cost of Trip and Price Charged by Year

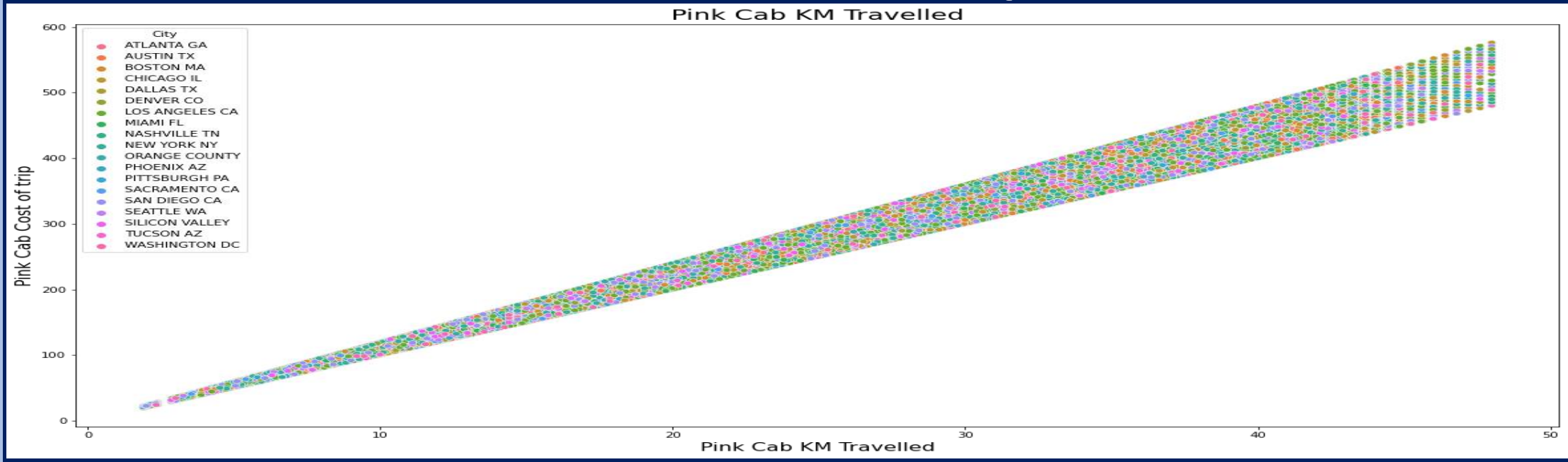


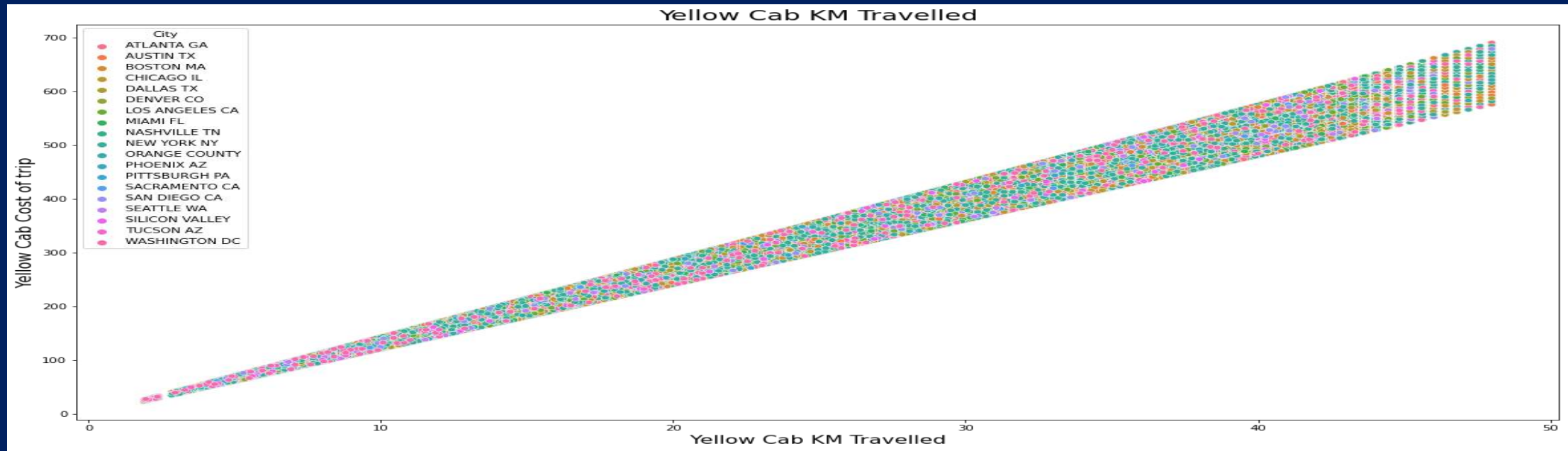


Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

KM Travelled vs. Cost of Trip Both Cabs

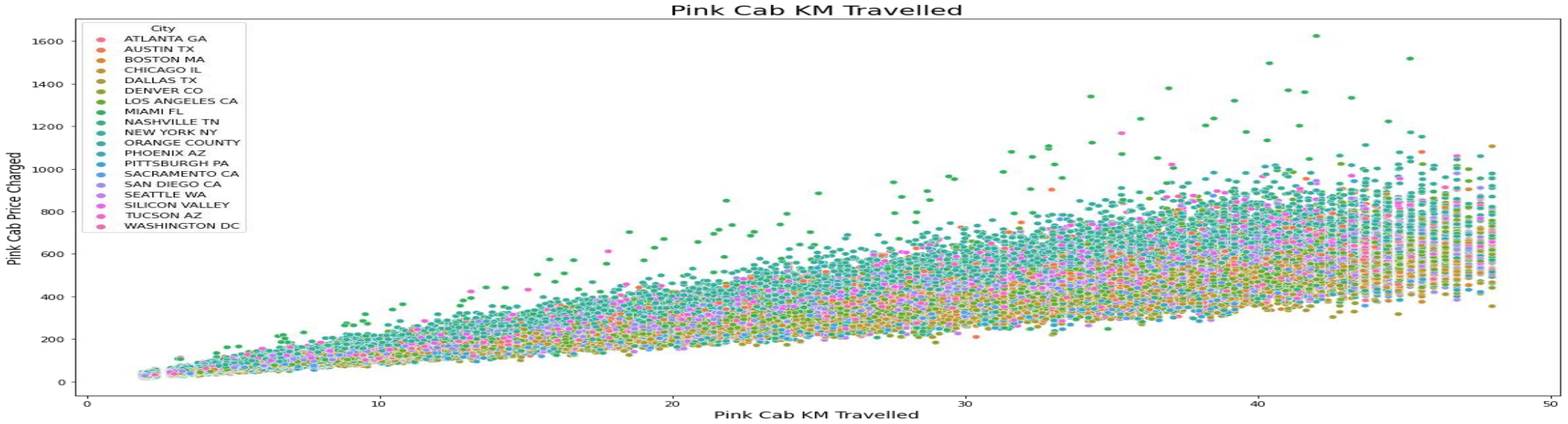


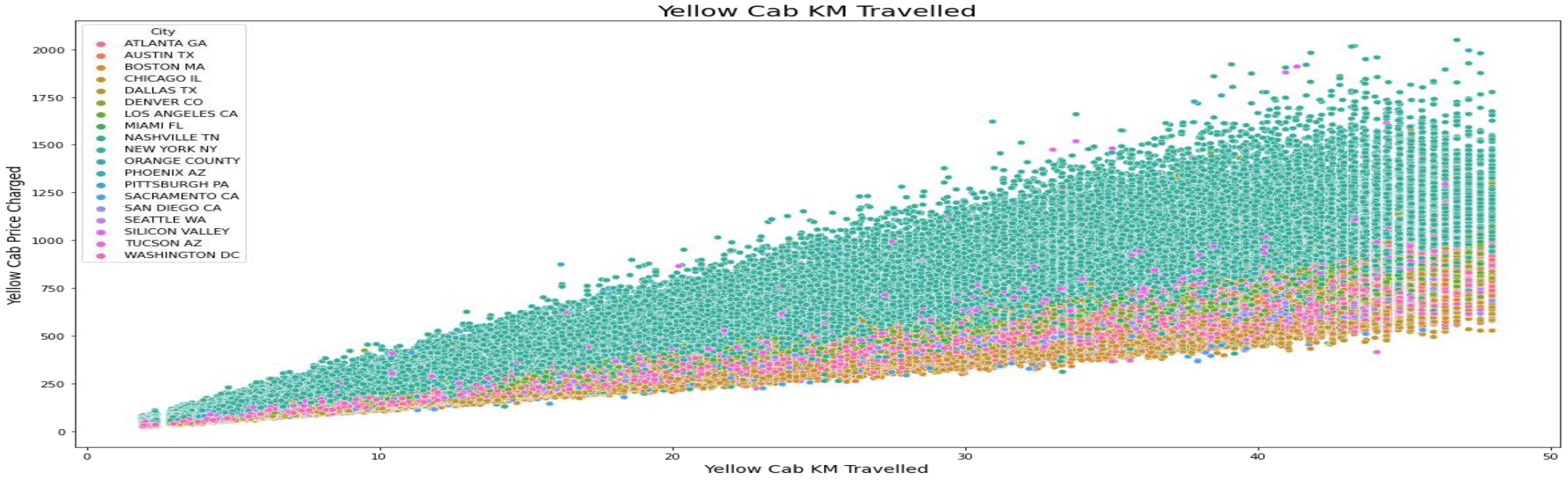


Data Glacier
Your Deep Learning Partner

KM Travelled vs. Price Change Both Cabs

Key Expletory Data Analysis



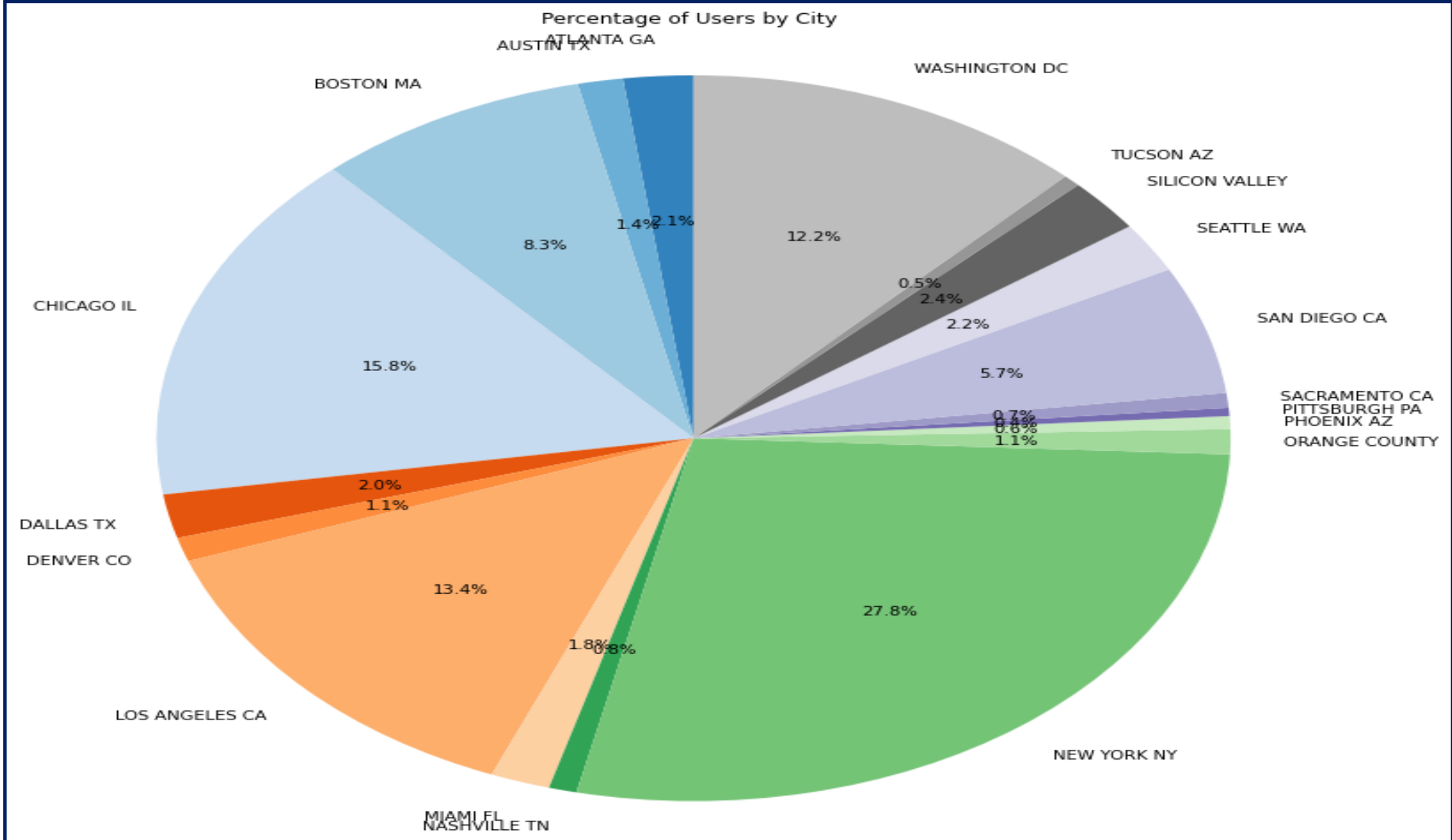


Data Glacier

Your Deep Learning Partner

Key Expletory Data Analysis

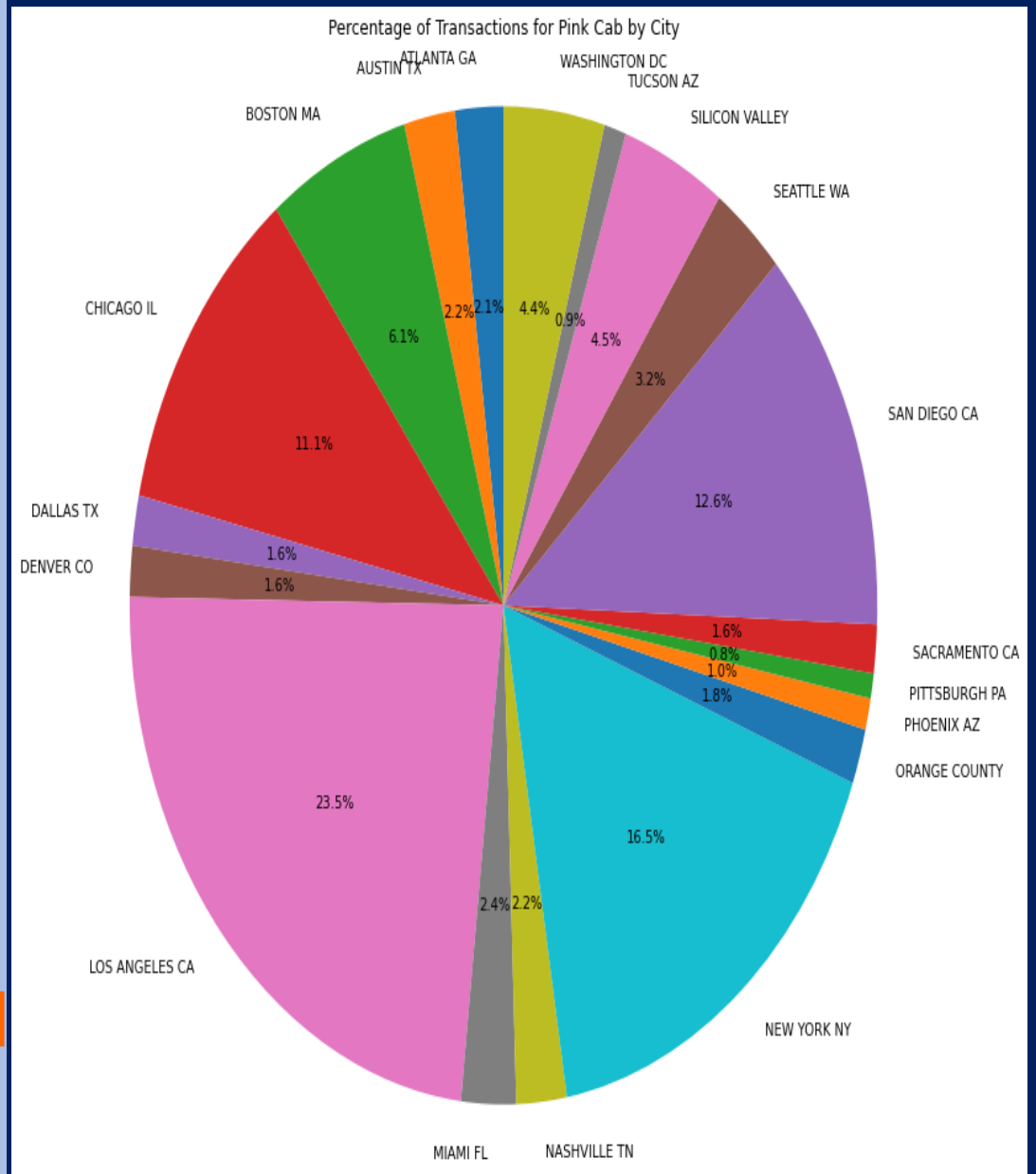
Percentage of Users by City

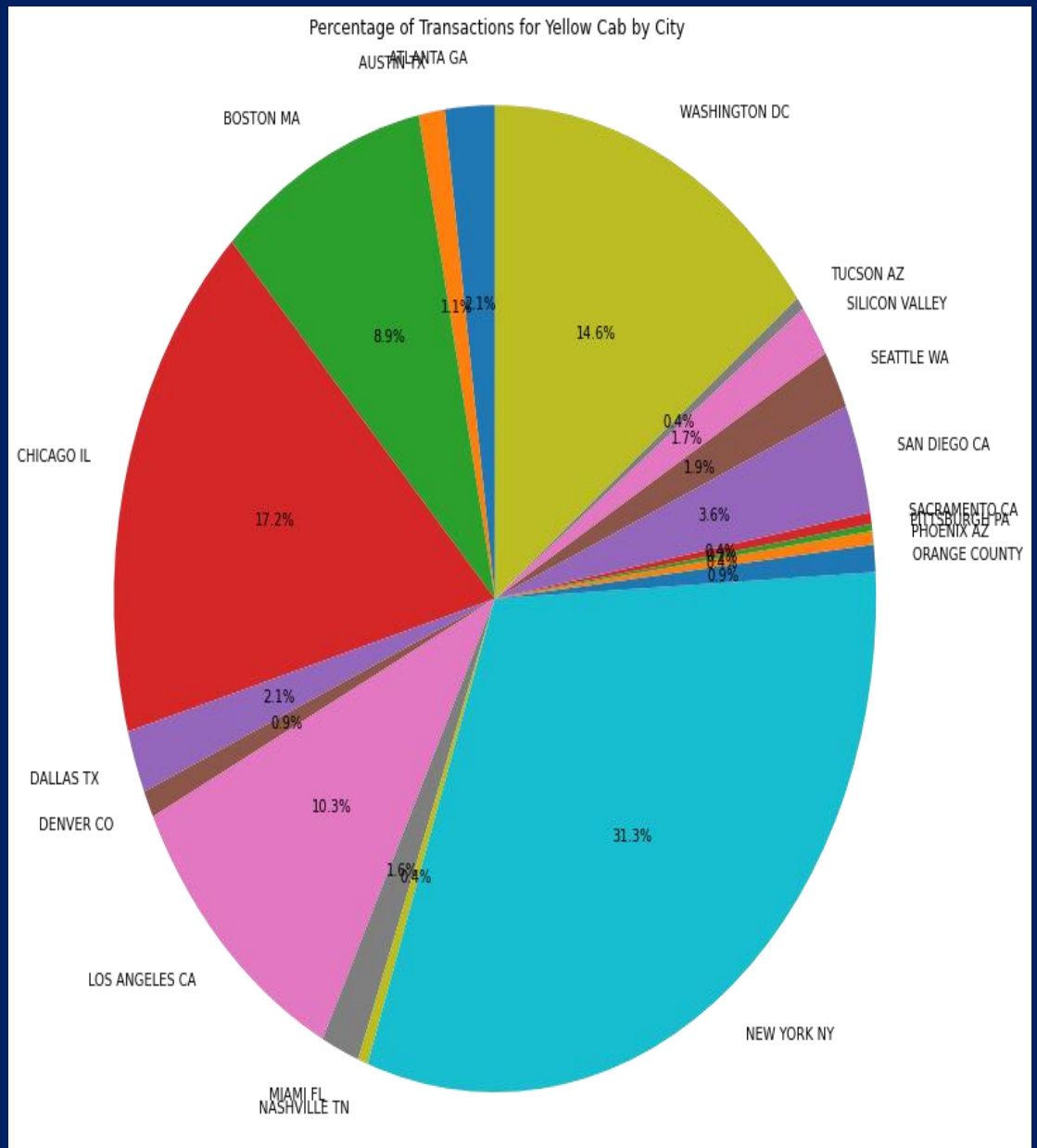


Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

Percentage of Transactions

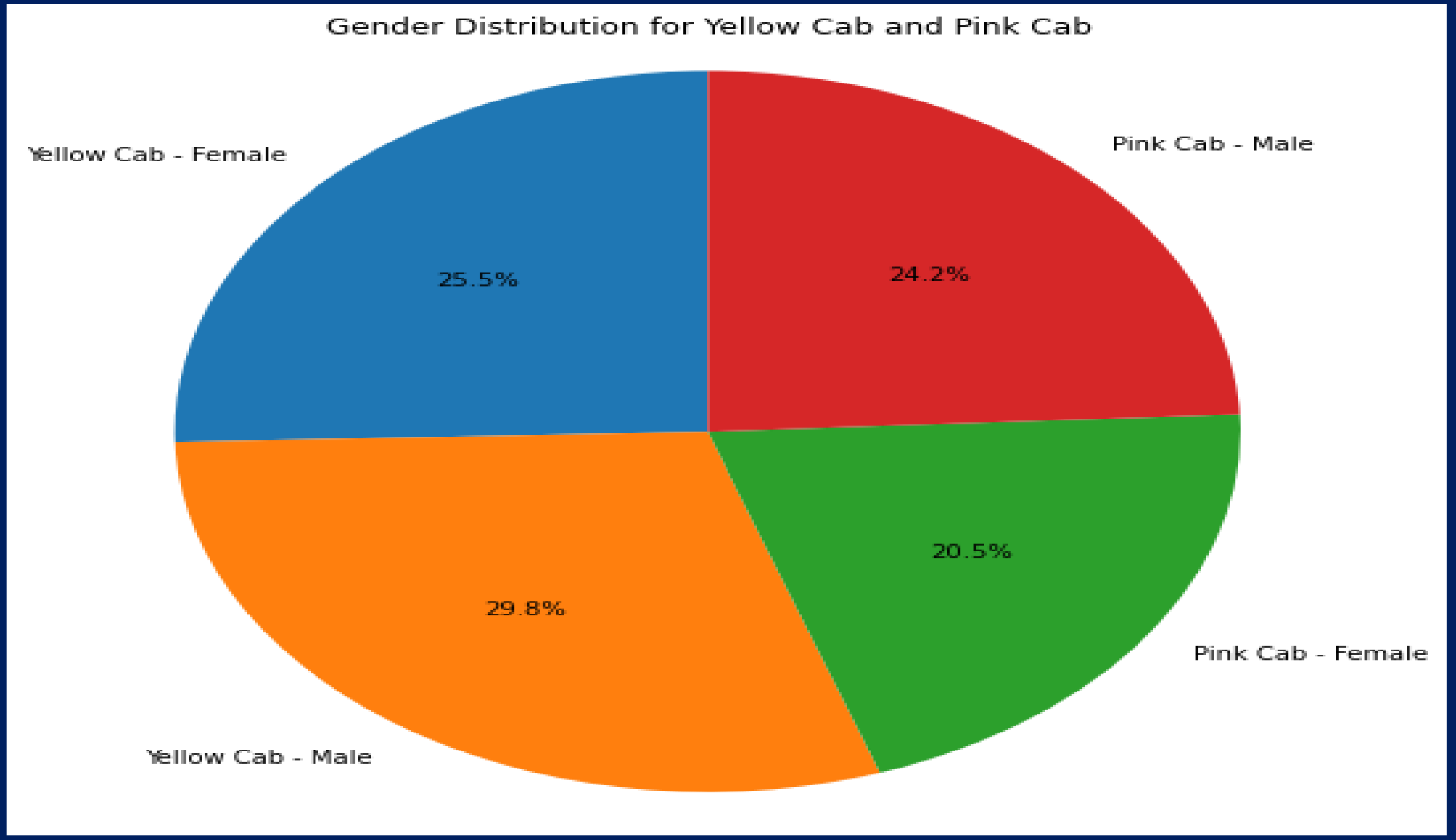




Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

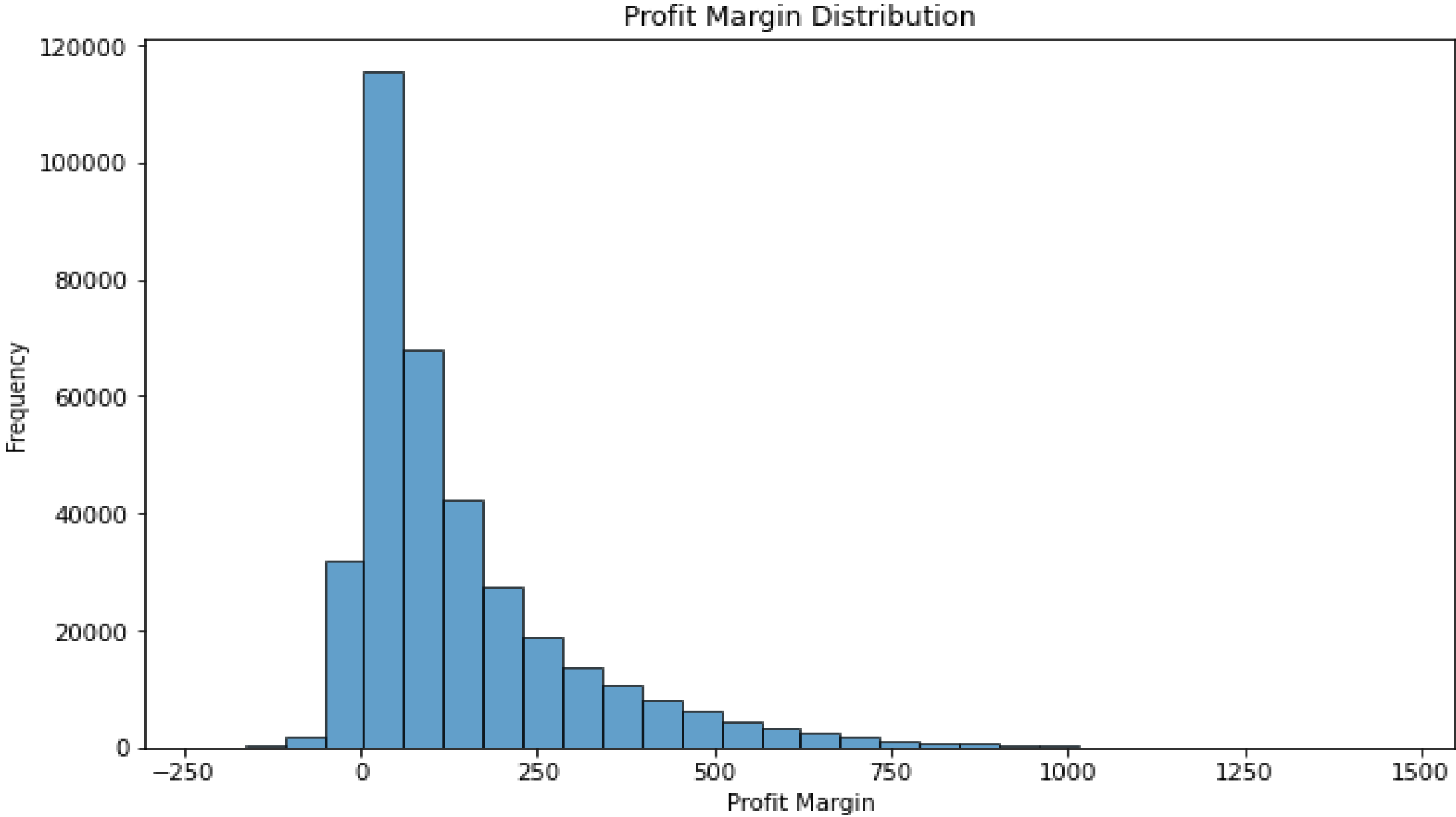
Gender Distribution for both Cabs



Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

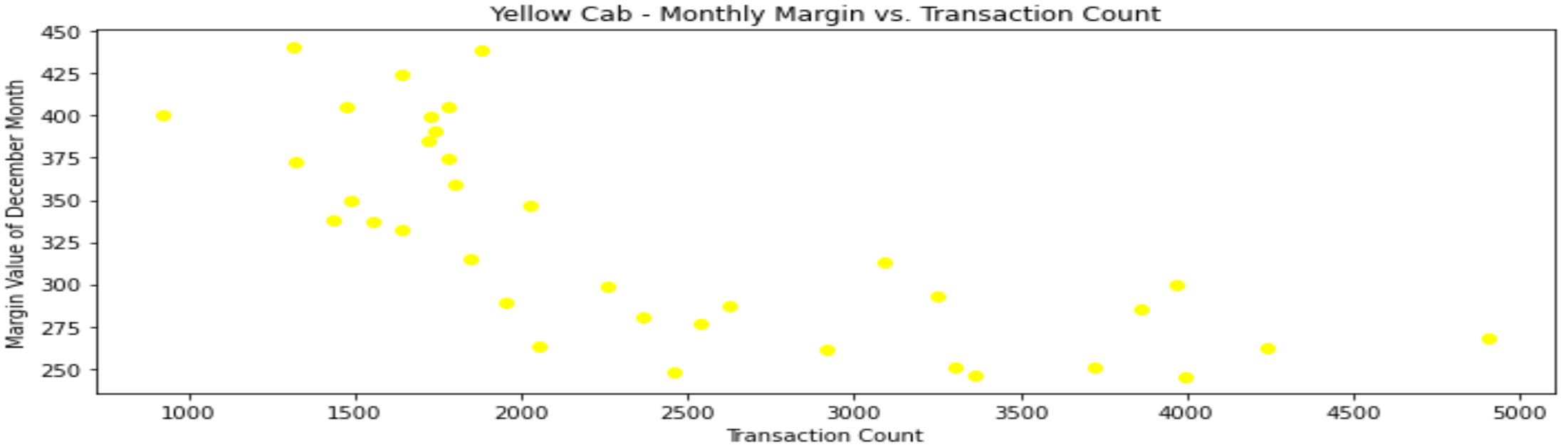
Profit Margin Distribution

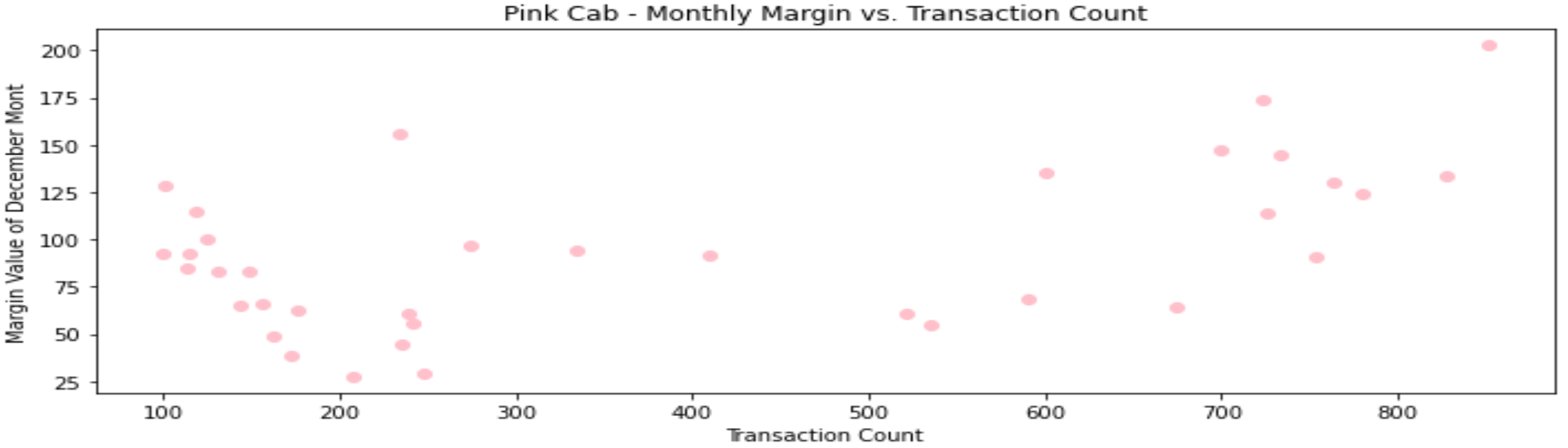


Data Glacier
Your Deep Learning Partner

Key Expletory Data Analysis

Monthly Margin vs. Transaction Count

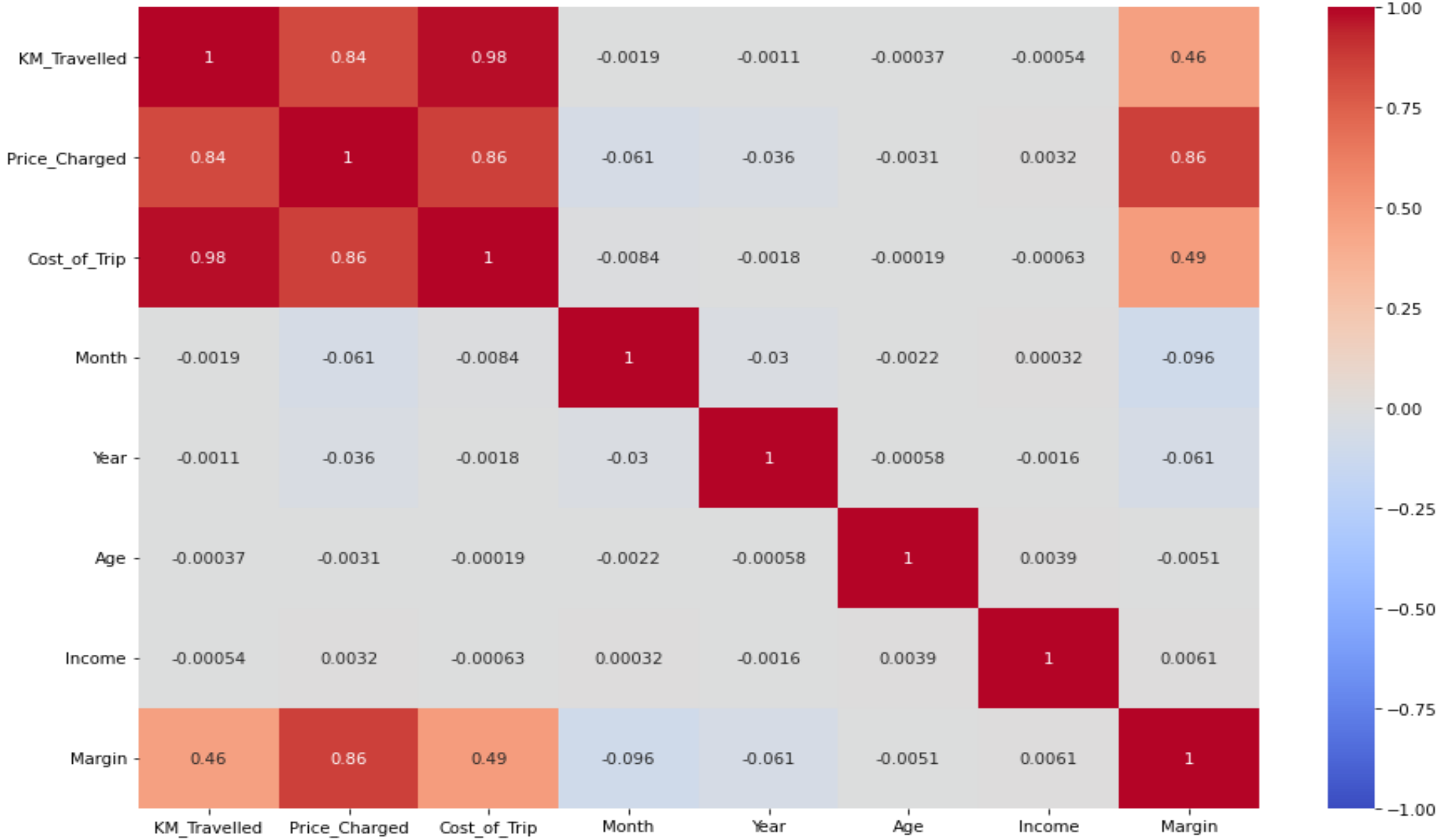




Data Glacier
Your Deep Learning Partner

Key Explatory Data Analysis

Correlation with the Features



Data Glacier
Your Deep Learning Partner

Key Findings

After a thorough evaluation of the two cab companies, Yellow Cab and Pink Cab, we have determined that Yellow Cab is the superior investment choice based on several critical factors. Here is a summary of our findings:

- **Customer Reach**
  - City-wise Distribution: Yellow Cab has a broader customer base across various cities. The distribution of users per city indicates higher transaction volumes for Yellow Cab compared to Pink Cab, signifying a more extensive reach and market penetration.
- **Profit Analysis**
  - City-wise Profitability: Yellow Cab consistently achieves higher profit margins across different cities. The profitability data indicates that Yellow Cab is more effective in optimizing its operations and maximizing revenue.
  - Profit per Trip: There is a notable increase in profit per trip for Yellow Cab with the distance traveled. This indicates an efficient pricing strategy that maximizes revenue from longer trips.

Data Glacier
Your Deep Learning Partner

Recomondations

Demand Trends

- Demand Consistency: Yellow Cab exhibits a steady demand across various cities, suggesting a reliable customer base and stable market demand.
- Monthly Growth: There is a noticeable increase in the number of Yellow Cab users over the months, indicating growing popularity and customer retention.
- Payment Modes: Yellow Cab users utilize both cash and digital payment methods, unlike Pink Cab, which has a lower cash usage rate. This flexibility in payment options likely contributes to its higher customer satisfaction and retention.

Seasonal Trends

- Distance Consistency: Trips taken with Yellow Cab demonstrate consistent distances traveled from 2016 to 2018. In contrast, Pink Cab experiences significant fluctuations in trip distances, indicating possible inconsistencies in service reliability or customer preferences.

Data Glacier
Your Deep Learning Partner

Recomondations

Based on the evaluation of customer reach, profit analysis, demand trends, and seasonal consistency, we recommend investing in Yellow Cab. The company has demonstrated superior performance in key areas, indicating a stable and profitable investment opportunity. The consistent demand, broad customer base, and effective pricing strategies position Yellow Cab as a strong market leader with promising growth potential.

Data Glacier
Your Deep Learning Partner

Thank You