Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis
G2M Cab Industry Analysis for XYZ Firm

Yash Parmar: yparmar2024@gmail.com
April 14th, 2025

Agenda

Background
Data Exploration
EDA
EDA Summary
Recommendations

Background – G2M (Cab Industry) Case Study

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. The outcome of my delivery will be a presentation to XYZ's Executive team.

- Objective: To provide actionable insights that will assist XYZ in identifying the right cab company for investment

- The Analysis Approach:
  - Data Understanding: Reviewing and preparing datasets for analysis
  - Forecasting: Analyzing profit trends and the number of rides for each cab type
  - Profitability Analysis: Identifying the most profitable cab company
  - Recommendations: Proposing an investment decision backed by data-drive insights

Data Exploration

Dataset Information
- **Time Frame:** 31/01/2016 to 31/12/2018
- **Cab_Data.csv:** 7 features & 359,392 data points
  - Transaction ID (Qualitative), Date of Travel (Quantitative), Company (Qualitative), City (Qualitative), KM Travelled (Quantitative), Price Charged (Quantitative), Cost of Trip (Quantitative)
- **Customer_ID.csv:** 4 features & 49,171 data points
  - Customer ID (Qualitative), Gender (Qualitative), Age (Quantitative), Income USD/Month (Quantitative)
- **Transaction_ID.csv:** 3 features & 440,098 data points
  - Transaction ID (Qualitative), Customer ID (Qualitative), Payment_Mode (Qualitative)
- **City.csv:** 3 features & 20 data points
  - City (Qualitative), Population (Quantitative), Users (Quantitative)

Data Summary

Data Summary:
- All of the datasets do not contain any outliers, and are consistent with each dataset
- Cab_Data contains dataset for each cab trip as described in the features
- Customer_ID contains dataset for each person who took the cab trip as described in the features
- Transaction_ID contains dataset for each person and how they payed for the cab trip they took as described in the features
- City contains dataset for different cities and how many people take cab trips out of people living in the city as described in the features

Data Merging

Data Merging:
- Customer_ID.csv and Transaction_ID.csv must be joined since they share a column "Customer ID", we will name this the "Transactions.csv" file for now
- Transactions.csv and Cab_Data.csv must be joined since they share a column "Transaction ID", we will name this the "Customer_Info.csv" file for now
- City.csv and master.csv must be joined since they share a column "City", we will name this the "master.csv" file for now
- From this, we have joined all of the datasets together from either the "Customer ID", "Transaction ID", or "City" column to get a master dataset, this means that each Transaction ID has a Customer ID, Date of Travel, Company, KM Travelled, Price Charged, Cost of Trip, Gender, Age, Income USD/Month, Payment_Mode, City, Population and Users

*All duplicates, NaN values, or outliers will be removed in the Jupyter Notebook*

Total KM Travelled by Companies

From this Bar Graph, we can see that the Yellow Cab has about 3 times the amount of kilometers travelled than the Pink Cab. This shows that Pink Cab is either a newer company or isn't as popular as Yellow Cab. Pink Cab could be a high risk, high reward investment since it's a newer company, but Yellow Cab would be a much safer investment since it's clearly made its name in the market of Cabs

Total KM Travelled by Pink Cab and Yellow Cab
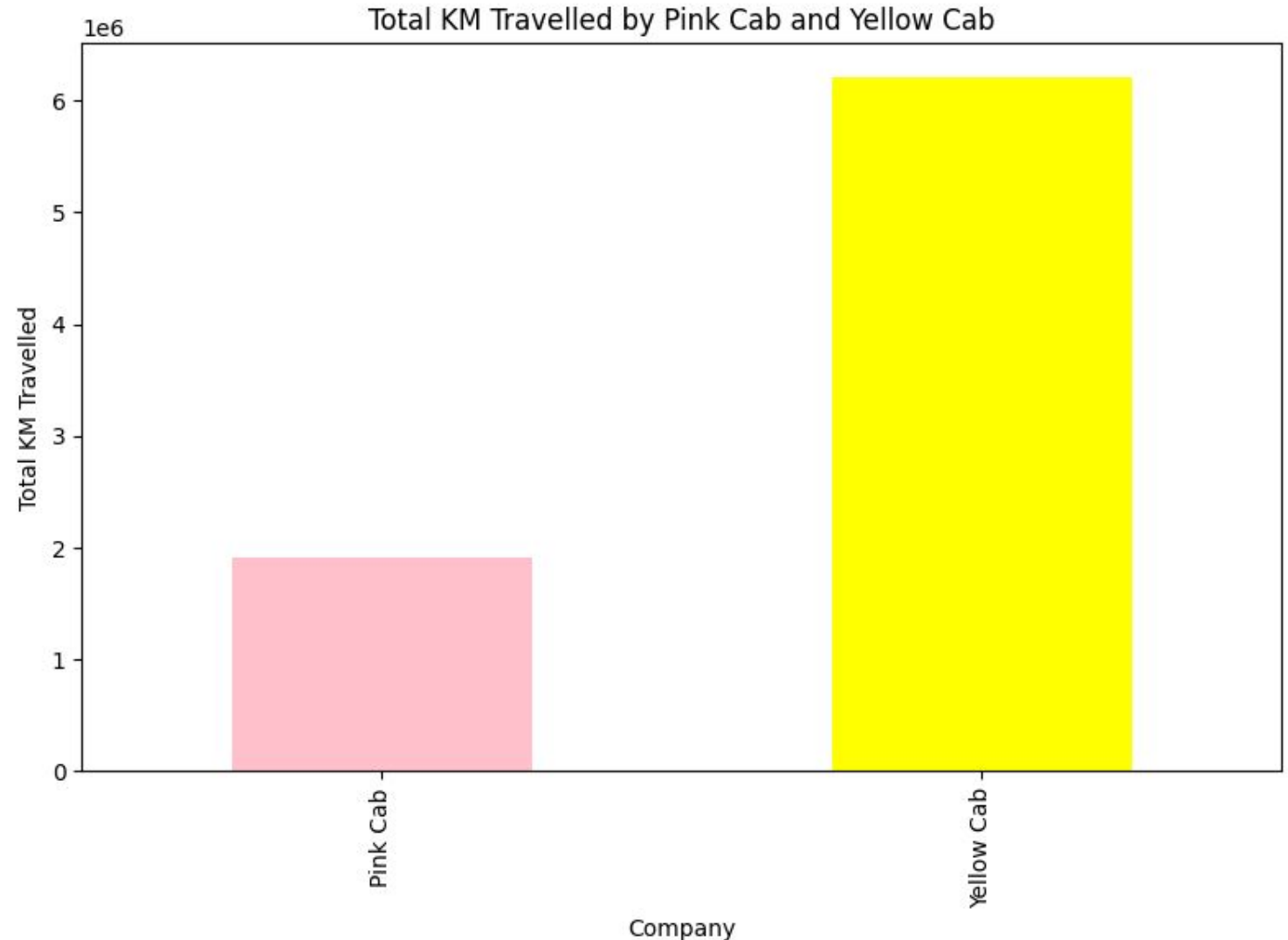
From this Bar Graph, we can see that the Yellow Cab company generally has more kilometers travelled in 13/20 cities. Generally, the Yellow Cab dominates more in more popular places, showing that it's a more popular choice than the Pink Cab. Further proving the previous analysis that Pink Cab is a newer company, meaning it could be one of the better upcoming companies, but the Yellow Cab is a safer investment.


KM Travelled by City for Pink Cab and Yellow Cab
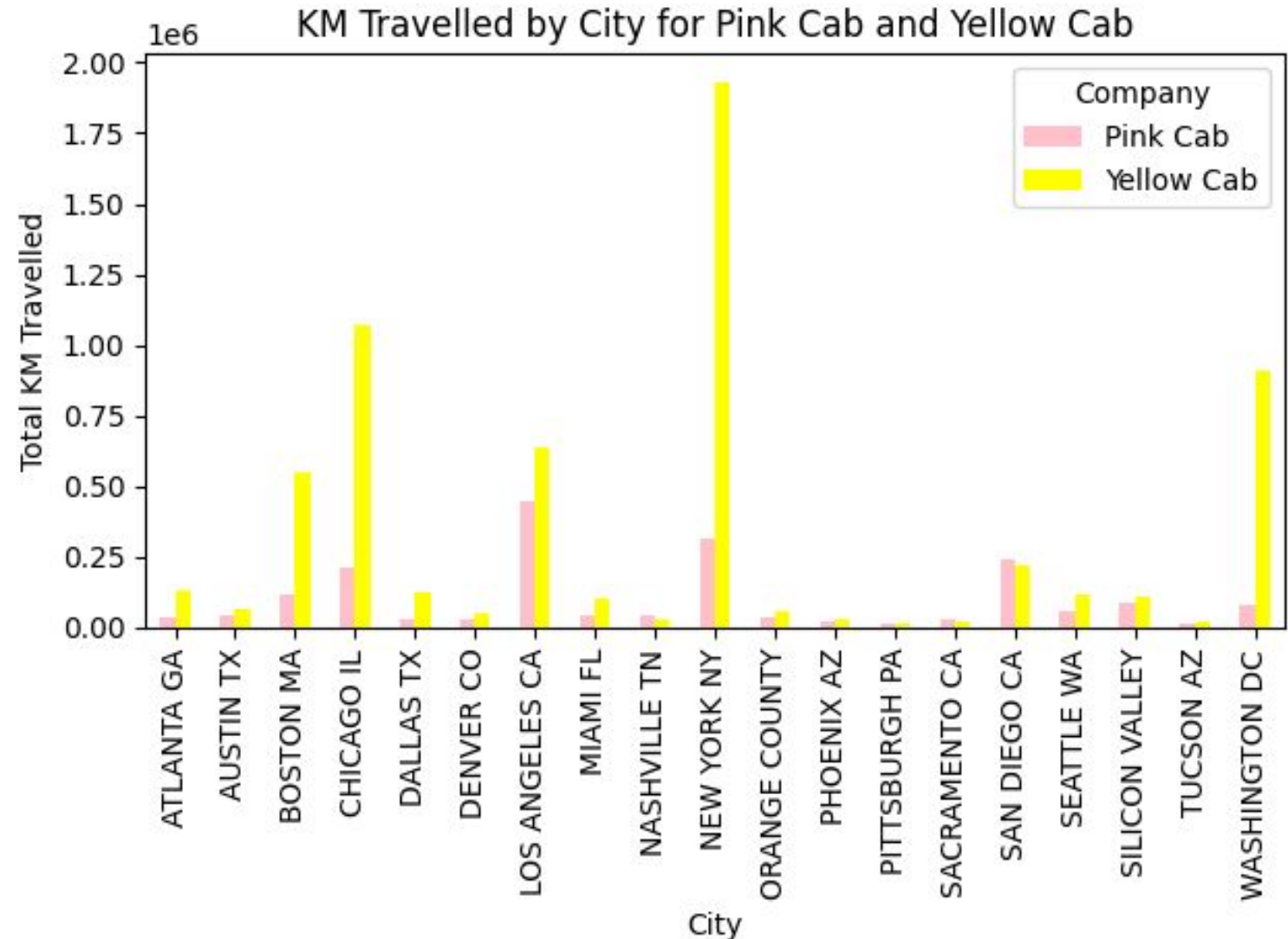
Payment Method for Companies

From this Pie Chart, we can see that this didn't provide much insight since both distributions are about similar. With hopes of checking which company would be more automated, and generally a better investment, with card payments than cash payments allowing for customer satisfaction and quicker payment transactions.



Payment Method Distribution - Pink Cab

Cash 40.1%

Card 59.9%
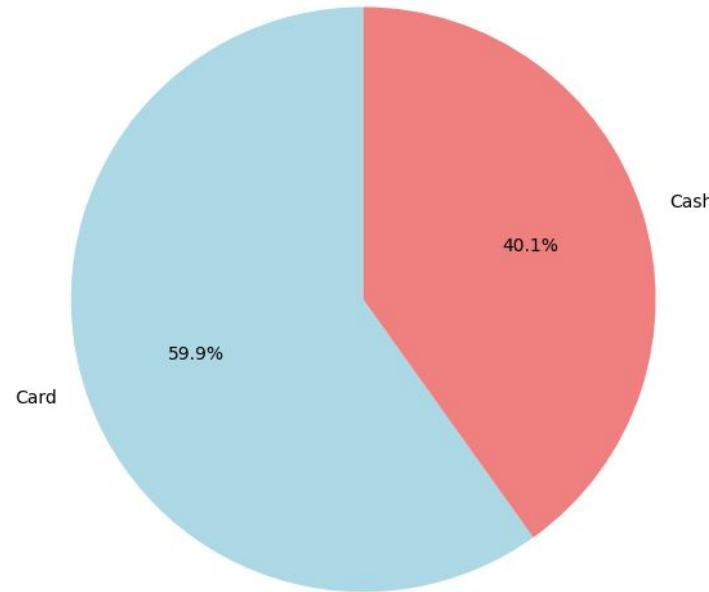
Payment Method Distribution - Yellow Cab

Cash 40.0%

Card 60.0%
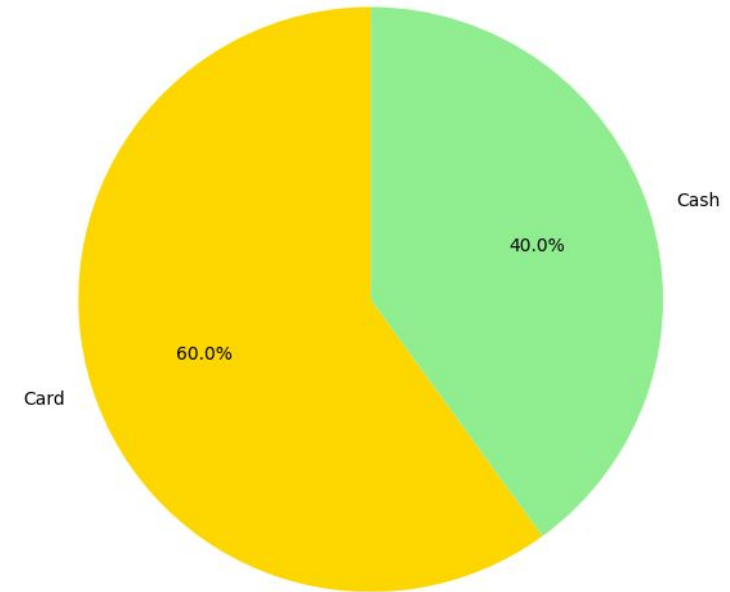
Average Passenger Income for Companies

From this Bar Chart, we can see that the income of passengers between both companies is about the same. With hopes of finding a difference in the average passenger income, to highlight the passengers with higher income taking the more "luxurious" cab. Since the income is the same, we can assume that the passengers aren't preferring one cab over the other due to customer satisfaction, so we can state that both cabs provide about the same experience.

Average Income of Passengers for Pink Cab and Yellow Cab
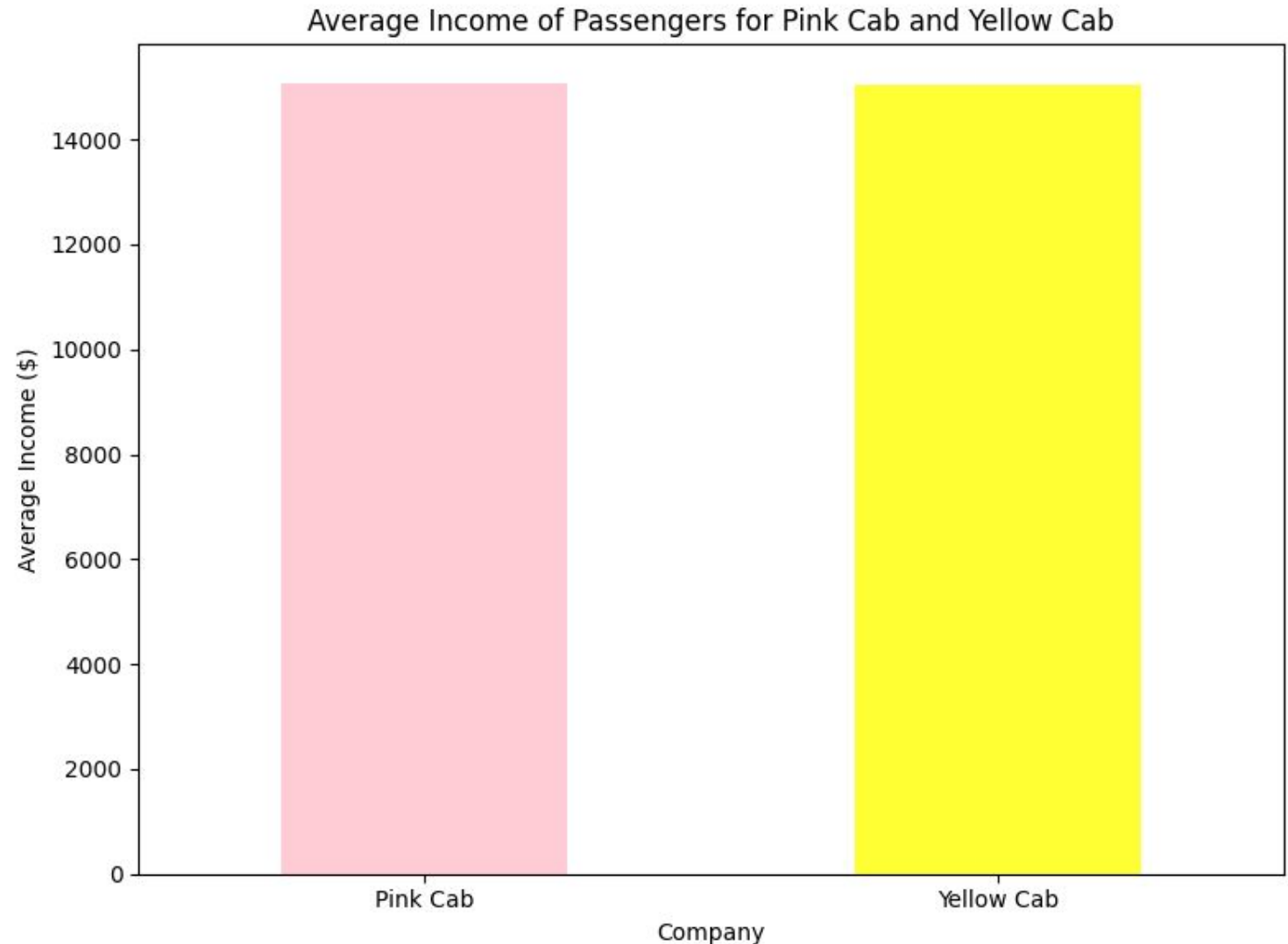
Price Charged and KM Travelled for Companies

From this Scatter Plot, we can see that both scatter plots don't differ much, other than the more scattered in outliers in the Pink Cab than the more clumped up outliers in the Yellow Cab. The R^2 value is 0.861 for the Pink Cab, and 0.739 for the Yellow Cab; this means that the Pink Cab generally is closer to the line of best fit than the Yellow Cab, offering more proper prices per kilometer. We can see that Pink Cab, the assumed newer company, is giving their customers fair prices since their customers aren't regulars, compared to Yellow Cab which has regular customers since they have been seemingly been in the market for longer. Pink Cab has no intention of scamming their customers by increasing prices for KM travelled, which shows by the low amount of outliers, which decreases the risk for this high risk, high reward investment.

Price Charged vs KM Travelled (Pink Cab)

Price Charged vs KM Travelled (Yellow Cab)
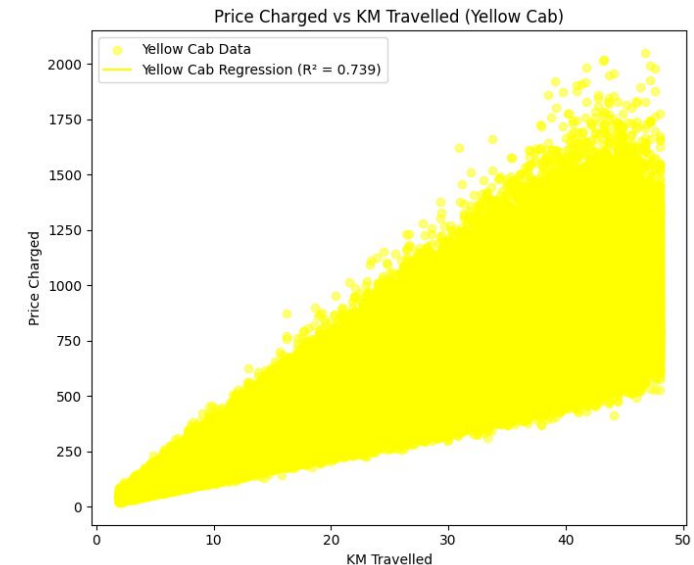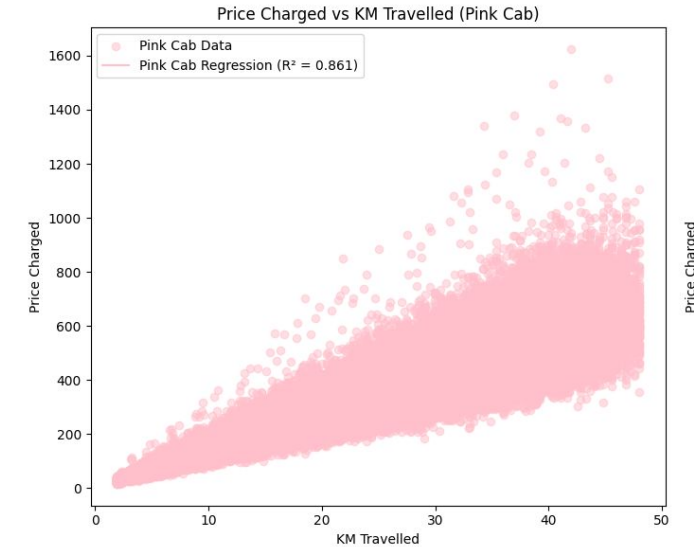
Total Profit for Companies

From this Bar Chart, we can see that the Yellow Cab clearly makes more than the Pink Cab, about 9 times more. Clearly, Yellow Cab is a good investment to make since they make more money than Pink Cab allowing for the investor's portfolio to increase as well. Pink Cab has potential to increase but most likely should not be the first choice of investment.

Total Profits for Pink Cab and Yellow Cab
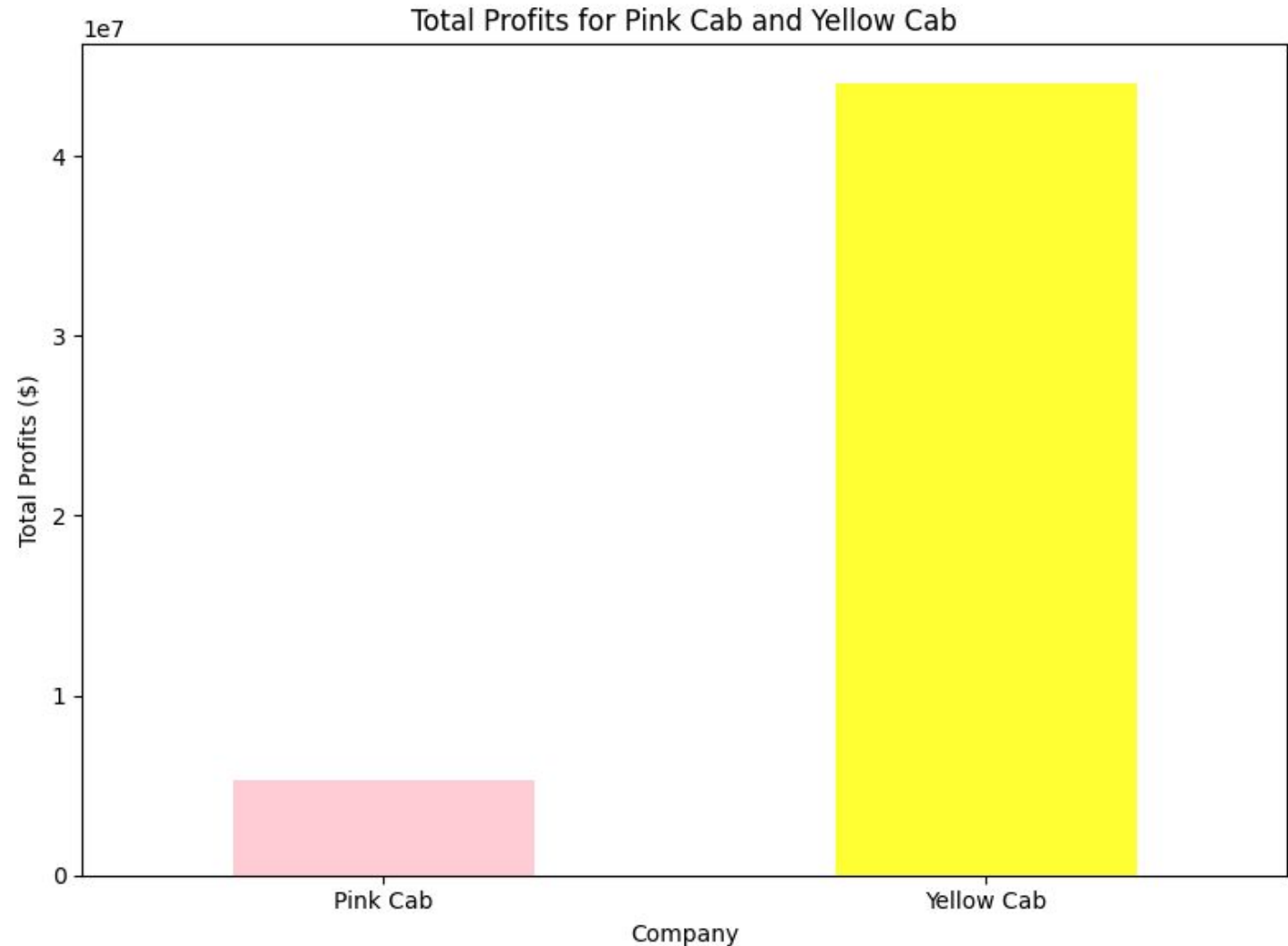
Recommendation for Investing

Let's review all of the data analysis and what we've recovered from it:

- Market Reach:
  - Yellow Cab operates in more cities (13 vs. 7) and dominates in key urban areas. Yellow Cab has a larger market footprint than Pink Cab.
- Pricing Model:
  - Pink Cab has a higher R^2 value (0.861 vs. 0.739), showing more consistent pricing relative to kilometers traveled. Pink Cab offers fair and competitive pricing to attract new customers while Yellow Cab relies on a loyal customer base.
- Profitability:
  - Yellow Cab generates significantly higher profits (9:1) than Pink Cab. Yellow Cab's profitability reflects its strong market position.

Through statistical analysis of both companies, evaluating key factors such as market reach, profitability, pricing models, and more, I confidently recommend investing in Yellow Cab. Yellow Cab holds a strong position in the market, consistent financial performance, and ability to keep its customer base and get more which makes it a safer and more reliable choice for long-term growth and stability.

Thank You

Data Glacier
Your Deep Learning Partner