



Data Glacier

Your Deep Learning Partner

G2M Case Study

Virtual Internship

20-Jun-2022

Background –G2M(cab industry) case study

- XYZ is a private equity 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.
- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

The exploratory data analysis has been divided into three parts:

- Data Understanding and Processing
- Data Analysis
- EDA Summary

Data Understanding

After viewing and understanding the data file names we can find that we have common field names such as:

Transaction ID > common between Cab_Data & Transaction_ID datasets

Customer ID > common between Customer_ID & Transaction_ID datasets

As a result we merge these data based on the common field variables and discover the dataset types

- 12 features
- Total data points : 359,392

We can learn from the categorical variables the following,

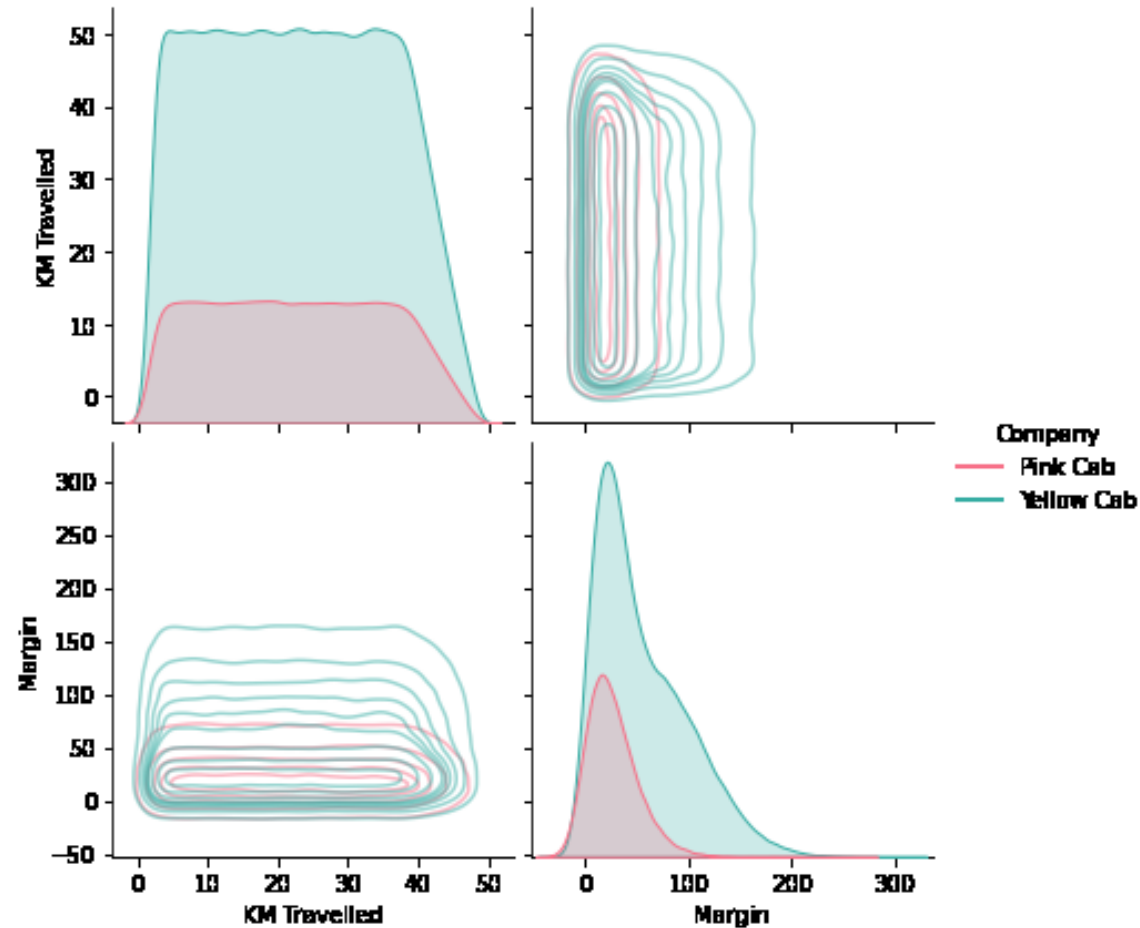
- Operating companies: **ATLANTA GA and Yellow Cab**
- Payment Methods: **Card and Cash**
- Trip services across: **19 cities**
- No. of Customers served: ~ **45,000**

Data Understanding

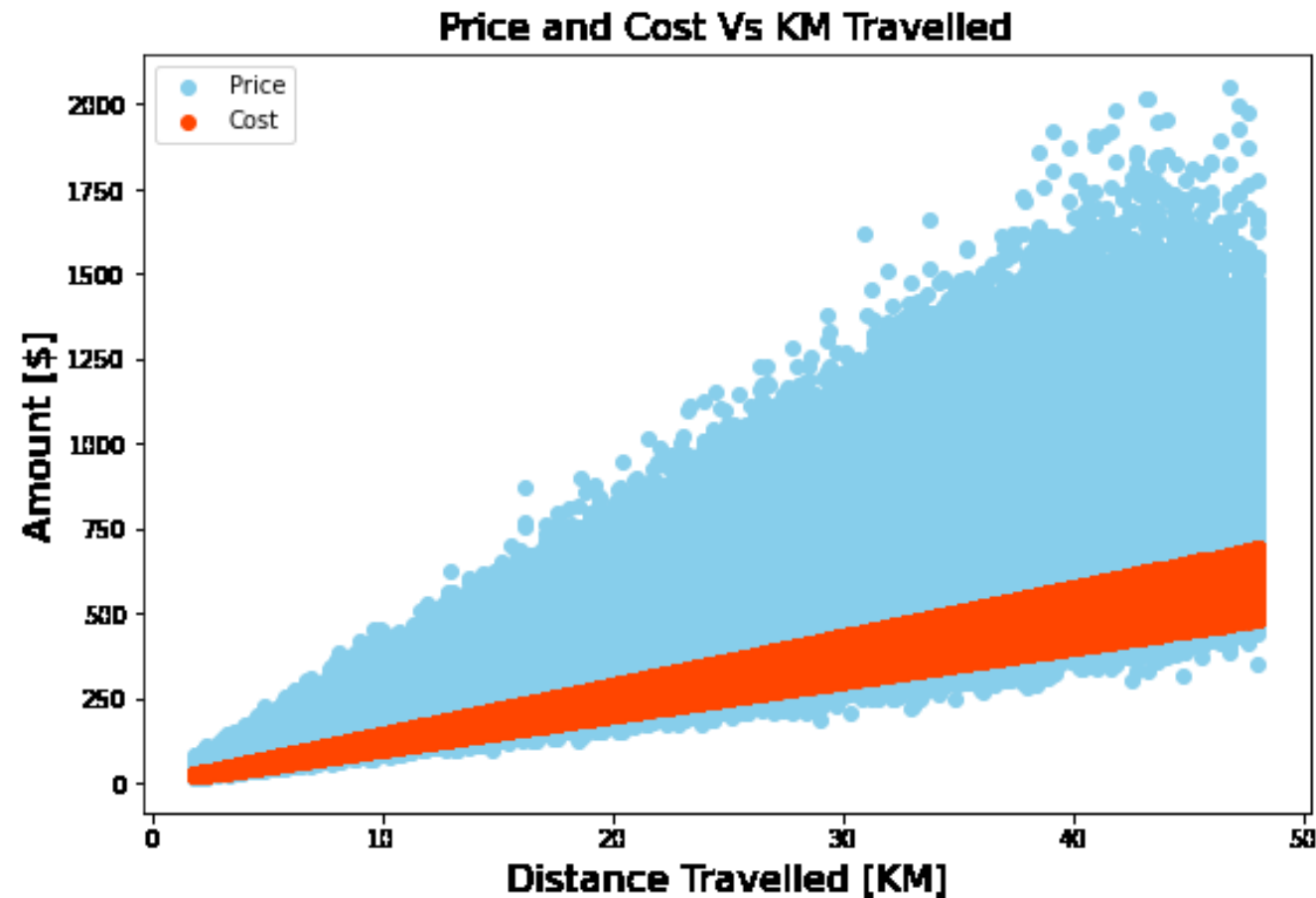
Statistics of numerical features:

| | KM Travelled | Price Charged | Cost of Trip | Age | Income (USD/Month) |
|--------------|--------------|---------------|--------------|---------|--------------------|
| count | 359,392 | 359,392 | 359,392 | 359,392 | 359,392 |
| mean | 23 | 423 | 286 | 35 | 15,049 |
| std | 12 | 274 | 158 | 13 | 7,969 |
| min | 2 | 16 | 19 | 18 | 2,000 |
| 25% | 12 | 206 | 151 | 25 | 8,424 |
| 50% | 22 | 386 | 282 | 33 | 14,685 |
| 75% | 33 | 584 | 414 | 42 | 21,035 |
| max | 48 | 2,048 | 691 | 65 | 35,000 |

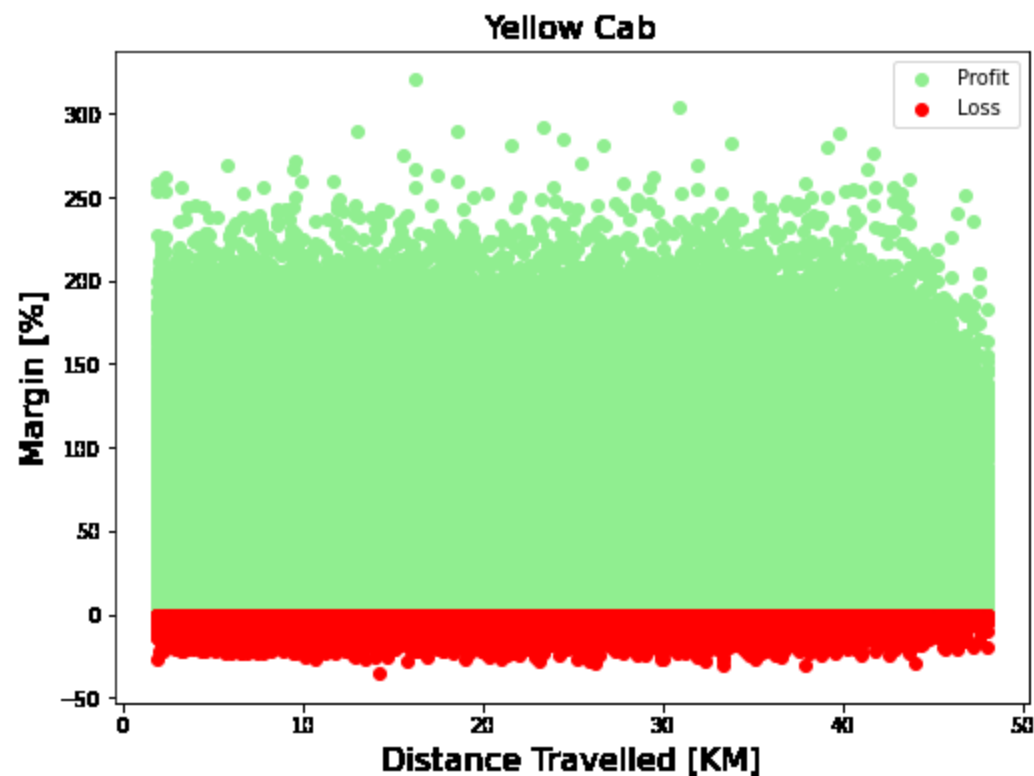
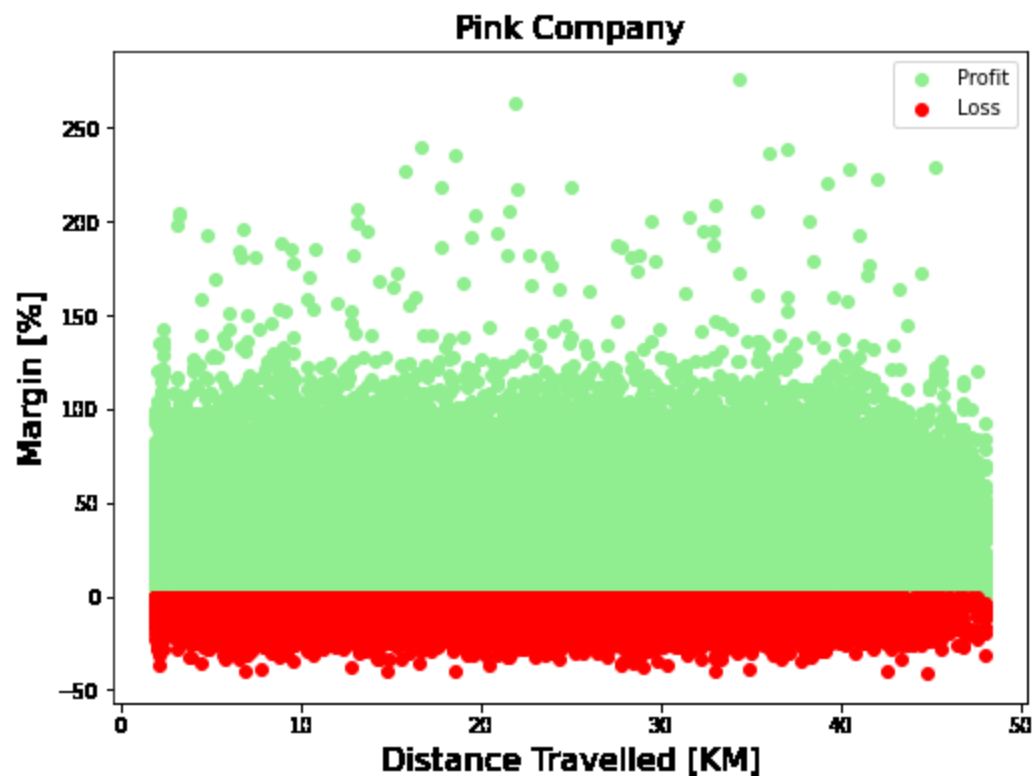
Data Analysis – Profit Margin Distribution



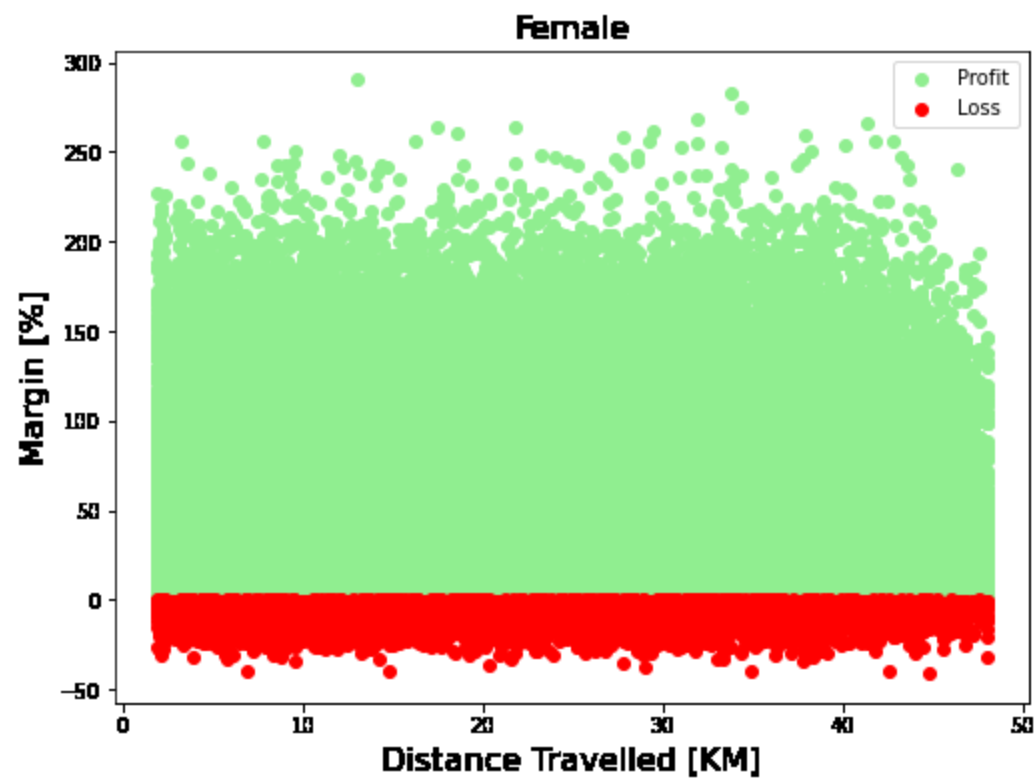
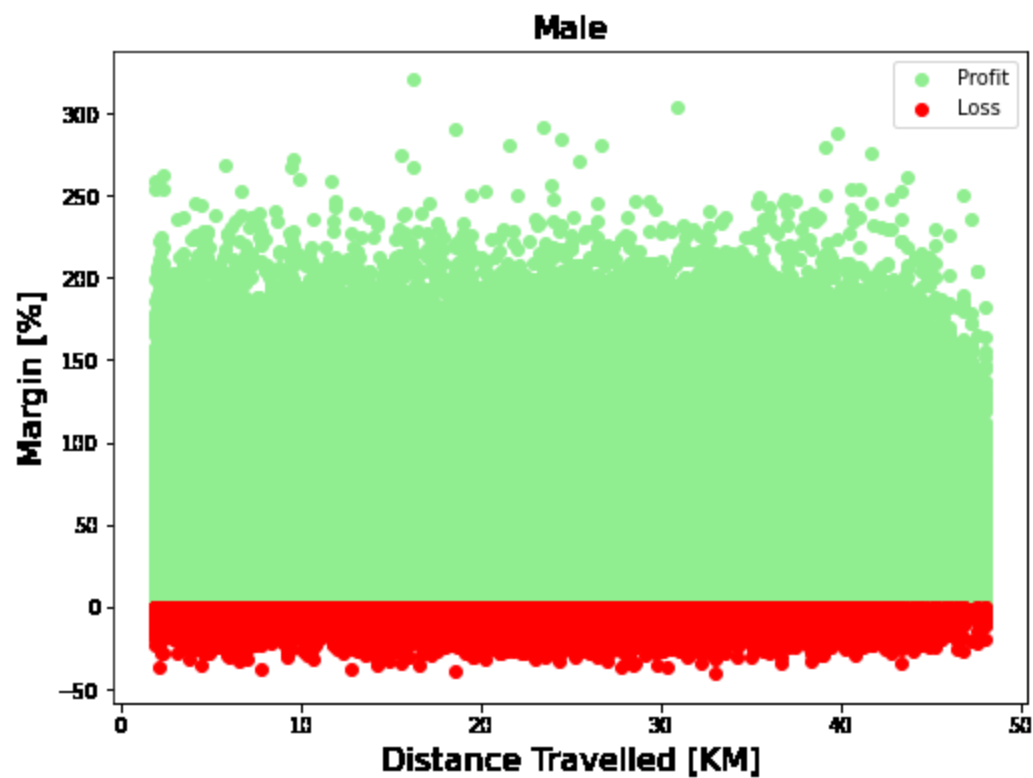
Data Analysis – Price and Cost Vs. KM Travelled



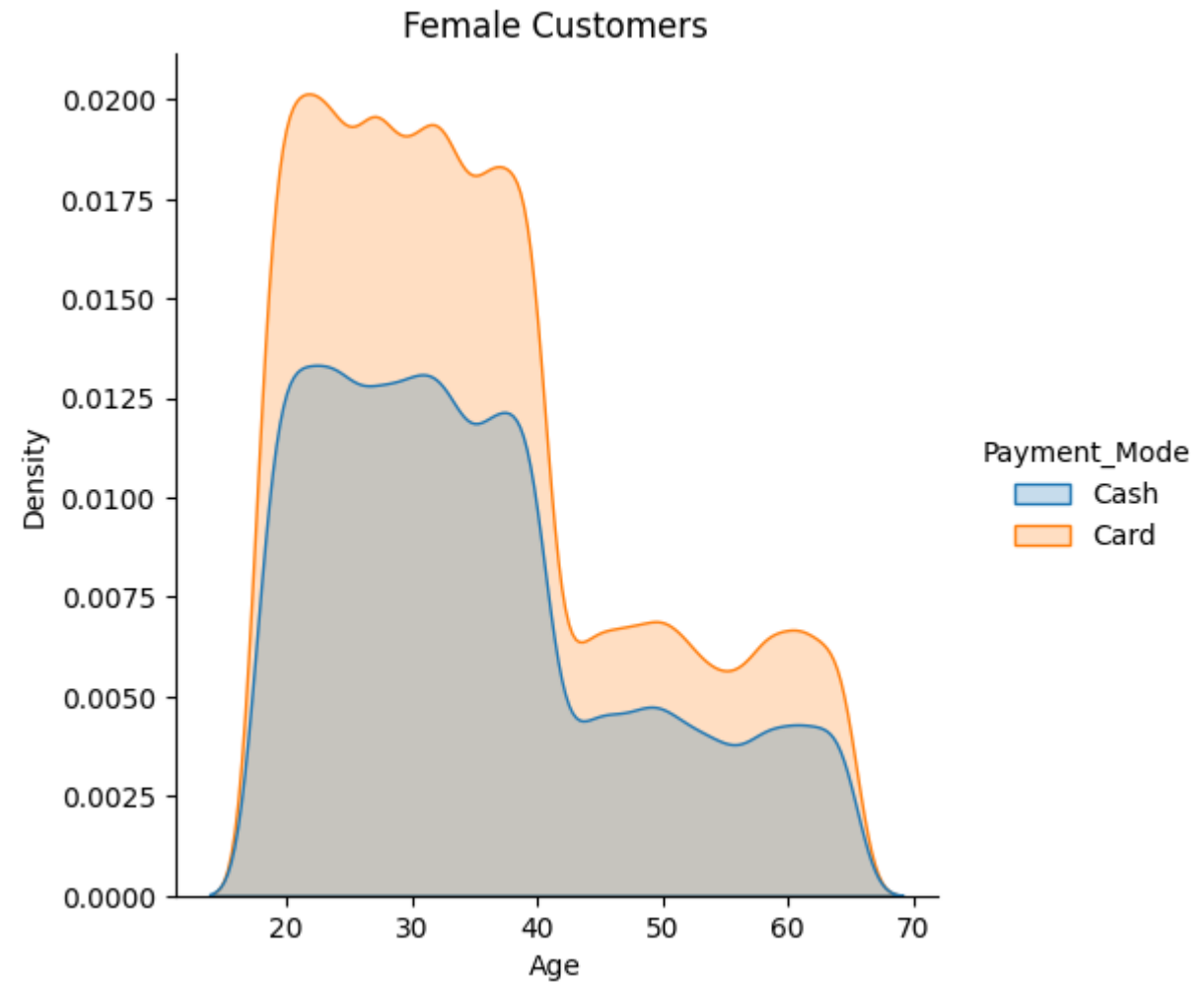
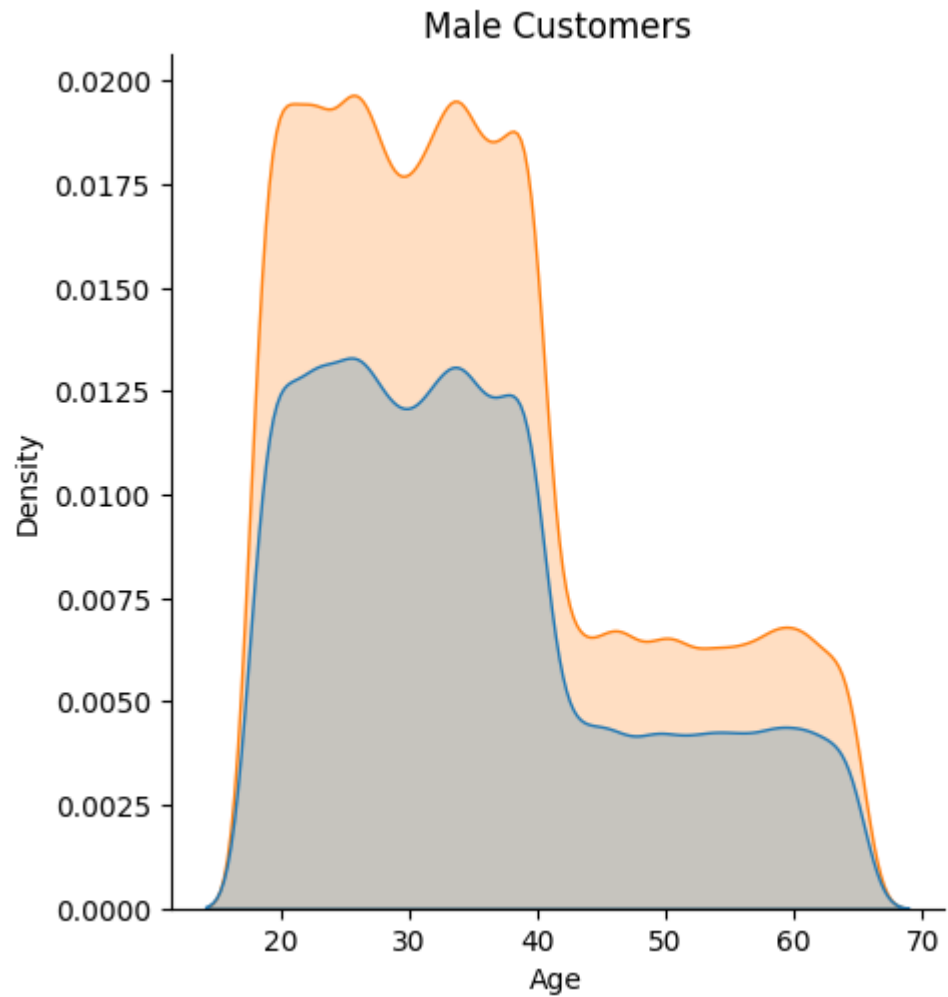
Profit Margin Analysis



Profit Margin Analysis



Age Analysis



EDA Summary

- The yellow cab company can achieve higher margins than the pink cab company.
- The yellow cab company has significantly provided more trips than the pink cab company. We could observe that by serving more customers the margins achieved can be increased.
- We could observe that among both cab companies the margin distribution is uniform to a great extent along the travelled distance domain, where the yellow cab company can achieve higher profit margins and less loss margins than the Pink Company.
- We could find that the margins are equally spread between males and females and therefore both genders contribute proportionally in equal terms towards the margin creation.
- We could find that for both genders males and females, we have youth customers (< 40 years old) are more than double the number of old customers (> 40 years old)
- We could observe that almost all customers tend to use payment cards more than using cash with an approximate factor of 1.5. This observation is consistent among both genders and across the whole age span.

Thank You



Data Glacier

Your Deep Learning Partner