EDA REPORT

This report presents an exploratory data analysis (EDA) of the cab industry, focusing on two companies: Pink Cab and Yellow Cab. The analysis includes various aspects such as gender distribution, payment methods, average users' salary, age distribution, and pricing over the years. Additionally, a decision tree model has been employed to classify the factors influencing the choice between Pink Cab and Yellow Cab.

A diagram of a number of women

Description automatically generated with medium confidence

A screenshot of a graph

Description automatically generated

**A screenshot of a graph

Description automatically generatedGender Distribution**

The gender distribution shows a slightly higher number of male users compared to female users for both companies. Yellow Cab has a higher number of users in both gender categories compared to Pink Cab.

* **Pink Cab**:
  + Female: 14,819
  + Male: 17,511
* **Yellow Cab**:
  + Female: 18,394
  + Male: 21,502

**Payment Method**

The analysis of payment methods reveals a preference for card payments over cash for both cab companies. However, Yellow Cab users significantly prefer card payments more than Pink Cab users.

* **Pink Cab**:
  + Card: 50,719
  + Cash: 33,992
* **Yellow Cab**:
  + Card: 164,785
  + Cash: 109,896

**Average Users' Salary**

The average salary of users is fairly similar for both cab companies, with Pink Cab users having a slightly higher average salary.

* **Pink Cab**: $15,034.10
* **Yellow Cab**: $14,983.90

**Age Distribution**

The age distribution of users is shown in the histogram. Both companies have a similar age distribution, with most users being in the 30-50 age range.

**Average Price Charged**

The average price charged by year shows a clear distinction between the two companies. Yellow Cab consistently charges more on average compared to Pink Cab.

* **2016**:
  + Pink Cab: $314.7
  + Yellow Cab: $467.9
* **2017**:
  + Pink Cab: $316.1
  + Yellow Cab: $466.8
* **2018**:
  + Pink Cab: $302.0
  + Yellow Cab: $441.1
* **2019**:
  + Pink Cab: $330.8
  + Yellow Cab: $392.2

**Total Cab Rides by Cities**

The total cab rides by cities indicate that some cities have a higher usage of cab services.

**Cab User Ratio by Cities**

The cab user ratio by cities shows the distribution of users for both cab companies across different cities.

**Cost of Trip Forecast**

The cost of trip forecast indicates the expected cost trends for both companies over the years. It provides insights into pricing strategies and market trends.

**Median KM Travelled per Trip**

The median kilometers travelled per trip for users of both companies show the typical trip distances covered by the cabs.

**Decision Tree Analysis**

The decision tree model helps in understanding the factors influencing the choice between Pink Cab and Yellow Cab. Key features such as Price Charged, Gender, Income, and Cost of Trip are used to make the classification.

**Key Insights from the Decision Tree:**

* **Price Charged**: This is the most significant factor affecting the choice. Users are more likely to choose Yellow Cab when the price charged is higher.
* **Gender**: Female users are more likely to choose Pink Cab compared to male users.
* **Income**: Higher income users tend to prefer Yellow Cab.
* **Cost of Trip**: Higher cost of trip generally leads to choosing Yellow Cab.

**Conclusion**

The exploratory data analysis provides a comprehensive overview of the cab industry, highlighting significant patterns and preferences among users of Pink Cab and Yellow Cab. The decision tree model further clarifies the key factors that influence the choice of cab service. Understanding these insights can help in formulating better marketing strategies and improving service offerings for both companies.

**Recommendations**

1. **Targeted Marketing**: Utilize the insights from gender and income distribution to tailor marketing campaigns.
2. **Pricing Strategies**: Consider dynamic pricing models based on user preferences and historical data.
3. **Service Improvements**: Enhance services in cities with higher usage to maintain user satisfaction and loyalty.
4. **Payment Method Promotions**: Encourage the use of preferred payment methods through promotions and discounts.