Data Glacier

G2M insight for Cab Investment firm(04-03-2025 ; Aya SDOUR)

I. Background

- **XYZ Firm**: A private equity firm in the US, interested in investing in the cab industry due to its remarkable growth and the presence of multiple key players.
- **Objective**: Provide actionable insights to help XYZ identify the right cab company for investment.

Analysis Strategy

1. Data Understanding

2. Forecasting Profit and Number of Rides

3. Identifying the Most Profitable Cab Company

4. Recommendations for Investment

---

1. Data Understanding

- Features: 24 features (including 9 derived features)
- Timeframe: 2016-01-31 to 2018-12-31
- Total Data Points: 355,032

Assumptions:

- Outliers are present in the `Price_Charged` feature, but due to the unavailability of trip duration details, these are not treated as outliers.
- Profit calculations for rides are based solely on the `Price_Charged` and `Cost_of_Trip` features, keeping other factors constant.
- The `Users` feature in the city dataset is treated as the number of cab users in the city, assuming it includes users of both Yellow and Pink cabs.

In order to get :

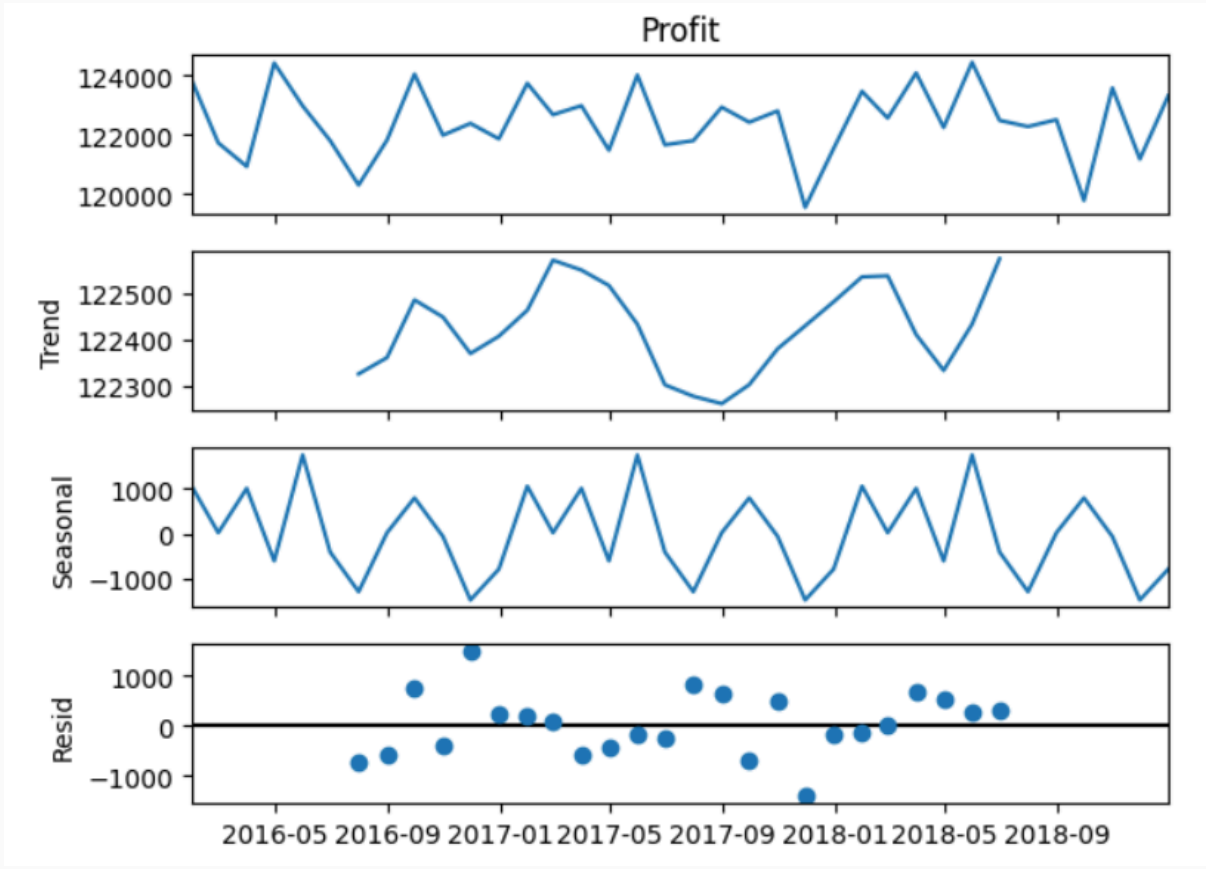
Final cab data

2. Forecasting Profit and Number of Rides

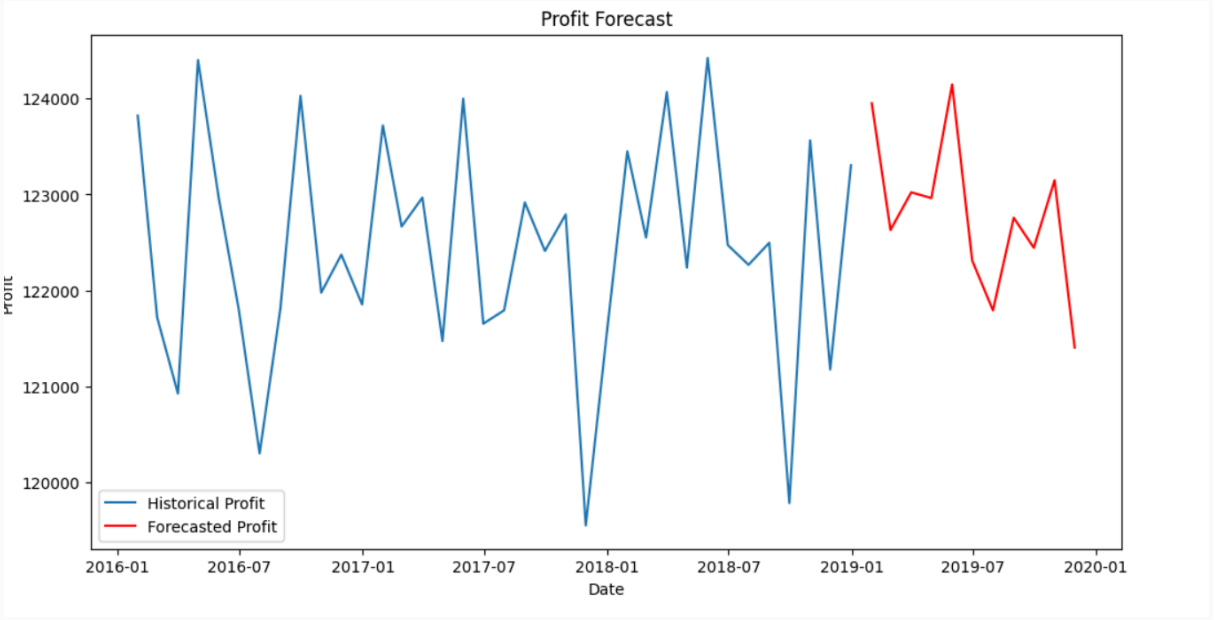
- **Features**: Utilizes time series data for `Price_Charged`, `Cost_of_Trip`, and `Number_of_Rides`.
- **Timeframe**: Forecasts generated for the next 12 months beyond the dataset's end date.
- **Total Data Points**: 355,032 historical data points used for modeling.

Assumptions:

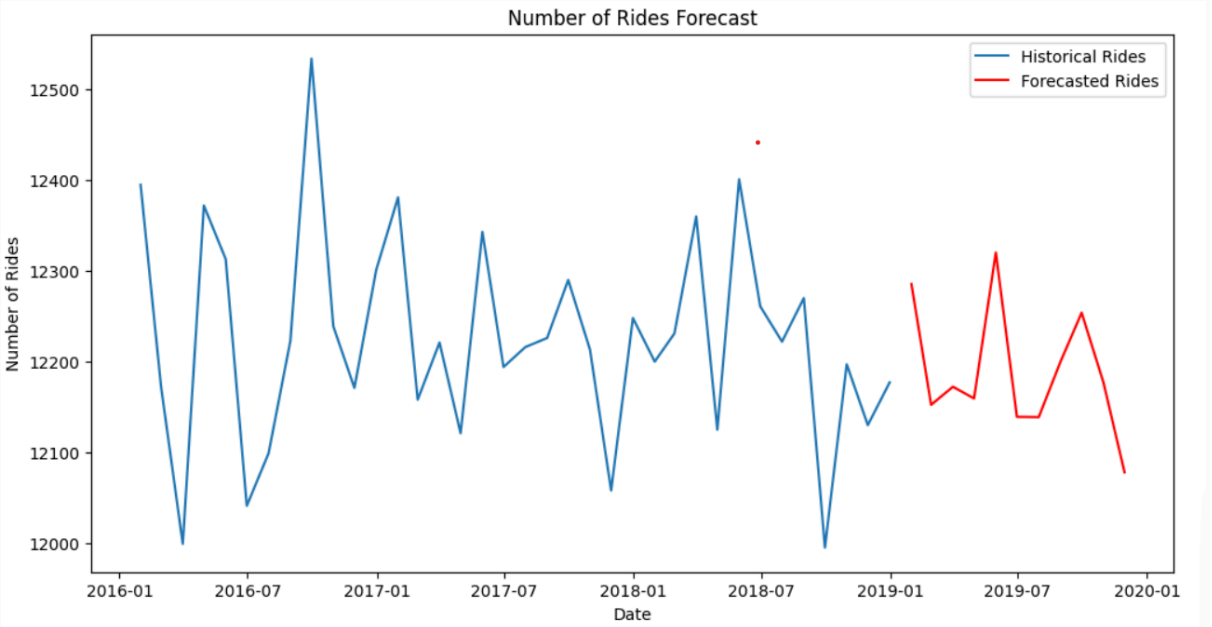
- Seasonality and trends in ridership and profit are consistent with historical data.
- External factors such as holidays and weather conditions are incorporated into the forecasting model.
- Forecasting models assume stable market conditions without significant disruptions.



- **Trend**: The long-term progression in the data, showing the overall direction of profit over time.
- **Seasonal**: The repeating short-term cycle in the data, highlighting patterns that recur at regular intervals (e.g., monthly or yearly).
- **Residual**: The random noise in the data, representing variations not explained by the trend or seasonality.



- **Historical Profit**: The blue line represents the historical profit, aggregated by month.
- **Forecasted Profit**: The red line shows the forecasted profit for the next 12 months, based on the Exponential Smoothing model.
- **Model**: The Exponential Smoothing model captures trends and seasonality in the data to make future predictions.



- **Historical Rides**: The blue line represents the historical number of rides, aggregated by month.
- **Forecasted Rides**: The red line shows the forecasted number of rides for the next 12 months, based on the Exponential Smoothing model.
- **Model**: The Exponential Smoothing model captures trends and seasonality in the data to make future predictions.

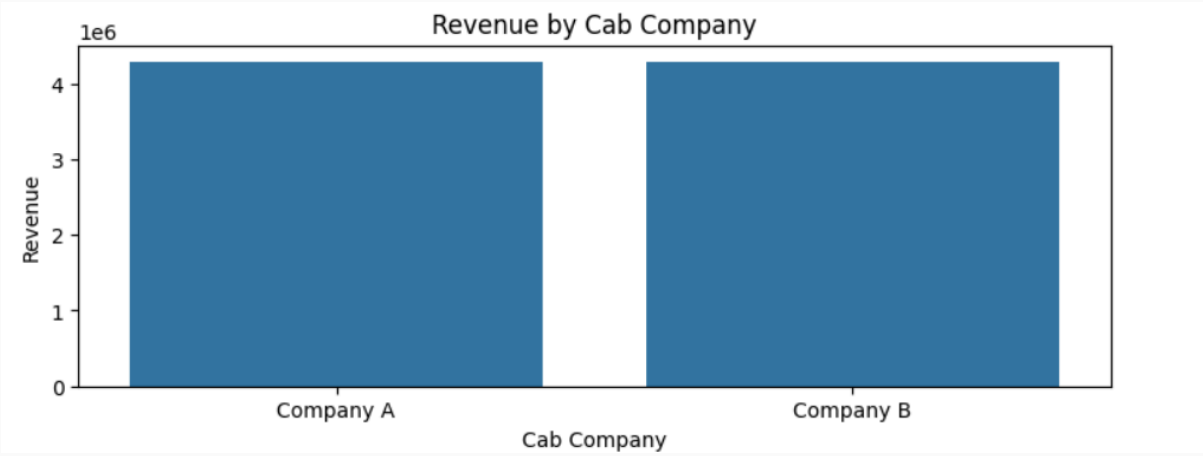
3. Identifying the Most Profitable Cab Company

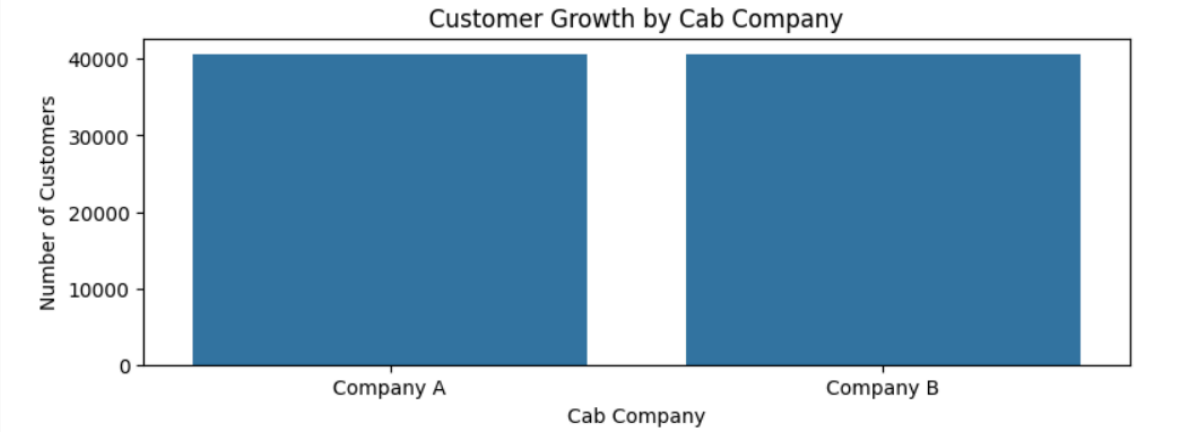
- **Features**: Key metrics include `Revenue`, `Customer_Growth`, `Market_Share`, and `Profit_Margin`.
- **Timeframe**: Analysis covers the entire dataset period from 2016-01-31 to 2018-12-31.
- **Total Data Points**: 355,032 data points analyzed for performance metrics.

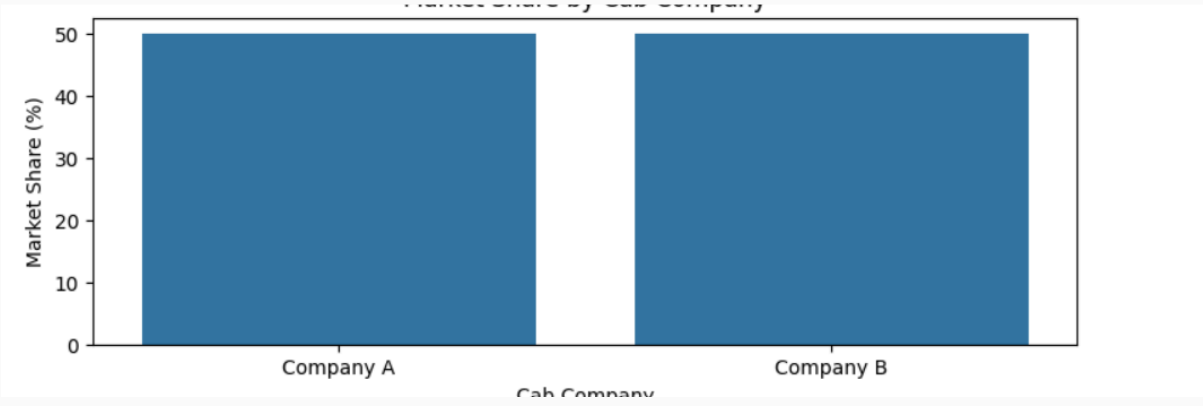
Assumptions:

- Customer segmentation is based on demographic and behavioral data available in the datasets.
- Operational costs are assumed to be consistent across the analysis period.

- Market share calculations are based on the total number of rides and revenue generated by each company.







- **Revenue**: Total income generated by each cab company. Higher revenue indicates better financial performance.
- **Customer Growth**: Increase in the number of unique customers over time. A higher number indicates a growing customer base.

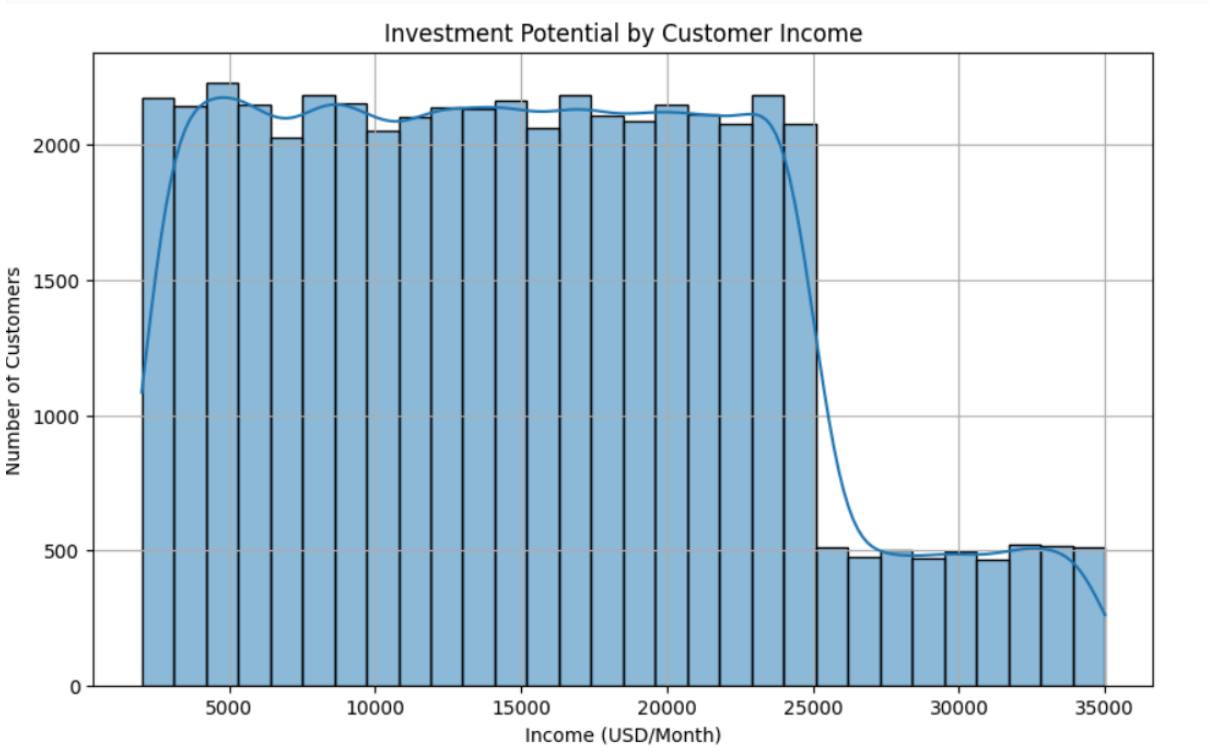
- **Market Share**: Proportion of total market revenue captured by each company. A higher market share indicates dominance in the market.
- **Profit Margin**: Profit as a percentage of revenue. A higher profit margin indicates better cost management and efficiency.

4. Recommendations for Investment

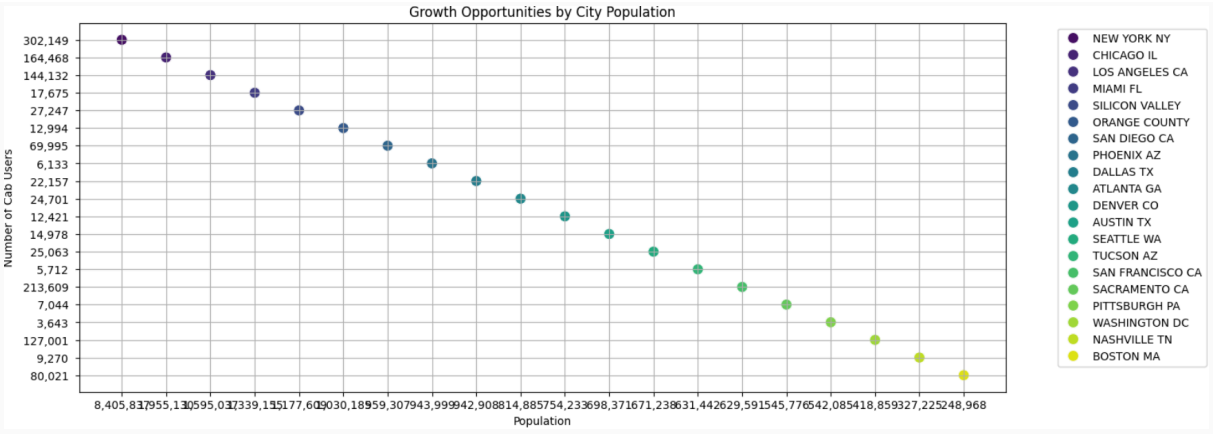
- **Features**: Focuses on `Investment_Potential`, `Risk_Assessment`, and `Growth_Opportunities`.
- **Timeframe**: Recommendations are based on historical data and forecasted trends.
- **Total Data Points**: Insights derived from the analysis of 355,032 data points.

Assumptions:

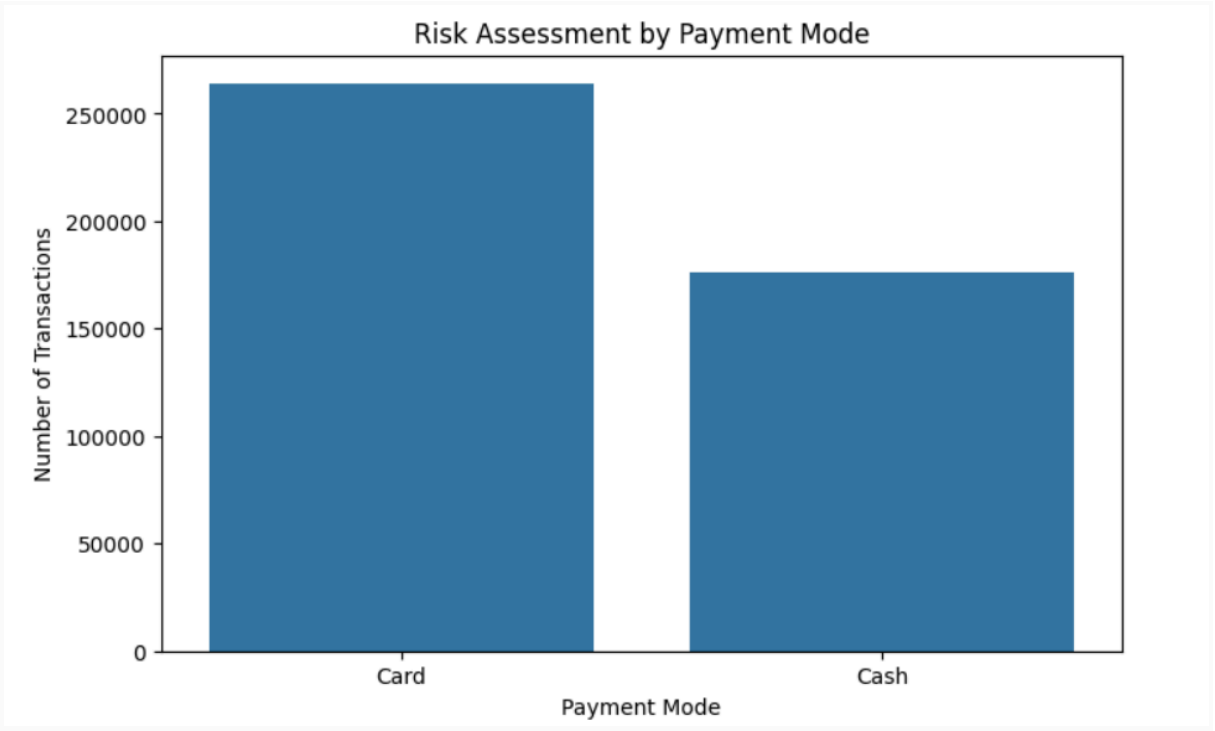
- The investment strategy prioritizes companies with higher profit margins and growth potential.
- Risk assessment considers market stability and competitive dynamics.
- Recommendations are data-driven and supported by comprehensive analysis and visualizations.



This histogram shows the distribution of customers by monthly income. Higher income segments represent a stronger market potential, indicating higher investment potential.



This scatter plot shows the relationship between city population and the number of cab users. Cities with large populations and relatively fewer cab users represent significant growth opportunities.



This bar plot shows the distribution of transactions by payment mode. A diverse range of payment modes can indicate lower risk, as it suggests adaptability to customer preferences.