**Dynamic-Pricing-Optimization-for-Ride-Sharing-Services**

**Dynamic-Pricing-Strategy**

"An end-to-end data science project implementing and evaluating a dynamic pricing strategy for a ride-sharing service, featuring exploratory data analysis, predictive modelling, and testing."

**Dynamic Pricing**

Dynamic pricing is a strategic pricing strategy that involves adjusting the prices of goods or services in real-time based on various factors such as demand, supply, competition, and other market conditions. This approach contrasts with traditional fixed pricing, where prices remain constant over time. Dynamic pricing aims to optimize revenue and profit by responding dynamically to fluctuations in market dynamics.

**Dataset Description:**

The dataset for the dynamic pricing project is a comprehensive collection of information related to a ride-sharing service. Each entry in the dataset represents a specific ride and includes various attributes that capture both the characteristics of the ride itself and the historical context of the service.

1.Number\_of\_Riders:

-Represents the count of riders involved in each ride.

-Numeric variable.

2.Number\_of\_Drivers:

-Indicates the count of drivers available during each ride.

-Numeric variable.

3.Location\_Category:

-Categorizes the location of the ride into different categories such as Urban, Suburban, or Rural.

-Categorical variable.

4.Customer\_Loyalty\_Status:

-Captures the loyalty status of the customer, such as Silver or Regular.

-Categorical variable.

5.Number\_of\_Past\_Rides:

-Records the number of rides the customer has taken in the past.

-Numeric variable.

6.Average\_Ratings:

-Represents the average ratings given by customers, providing insights into the service quality.

-Numeric variable.

7.Time\_of\_Booking:

-Specifies the time of day when the ride was booked (e.g., Morning, Afternoon, Evening, Night).

-Categorical variable.

8.Vehicle\_Type:

-Identifies the type of vehicle used for the ride, such as Premium or Economy.

-Categorical variable.

9.Expected\_Ride\_Duration:

-Indicates the expected duration of the ride in minutes.

-Numeric variable.

10.Historical\_Cost\_of\_Ride:

-Represents the historical cost of the ride, providing a baseline for comparison with dynamically adjusted prices.

-Numeric variable.

**Exploratory Data Analysis (EDA)**

The exploratory data analysis (EDA) for the dynamic pricing project involved a comprehensive exploration of the dataset to understand the relationships, patterns, and key insights that could inform subsequent stages of the project.

1. **Summary Statistics:**
   * The summary statistics provided a quick overview of the central tendencies and dispersion of numerical features in the dataset. This included key statistics such as mean, standard deviation, minimum, and maximum values.
2. **Expected Ride Duration vs. Historical Cost:**
   * Explored the relationship between the expected ride duration and the historical cost of the ride using a scatter plot. The plot included a trendline to visualize potential trends or patterns.
3. **Historical Cost Distribution by Vehicle Type:**
   * Investigated the distribution of historical costs based on different vehicle types using a box plot. This visualization provided insights into how the type of vehicle may impact the cost of rides.
4. **Correlation Matrix:**
   * Computed and visualized the correlation matrix to identify potential relationships between numerical features. This step is crucial for understanding the interdependencies between different factors.
5. **Historical Cost Distribution by Customer Loyalty Status:**
   * Examined how historical costs vary based on customer loyalty status using a box plot. This analysis provides insights into whether loyal customers experience different pricing dynamics.
6. **Historical Cost Distribution by Time of Booking:**
   * Investigated the distribution of historical costs based on the time of booking. This analysis helps understand how temporal factors influence ride costs.
7. **Historical Cost Distribution by Location Category:**
   * Explored the distribution of historical costs across different location categories (Urban, Suburban, Rural) using a box plot. This analysis provides insights into regional pricing dynamics.
8. **Number of Riders vs. Number of Drivers:**
   * Visualized the relationship between the number of riders and the number of drivers using a scatter plot. This analysis provides insights into the balance between demand and supply.
9. **Average Ratings vs. Historical Cost:**
   * Explored the relationship between average ratings and historical costs using a scatter plot. Understanding how customer ratings relate to costs can provide insights into the perceived value of the service.

**Implementing Dynamic Pricing Strategy**

1. **Demand and Supply Multipliers:**
   * **Demand Multiplier:** Calculated based on percentiles for high and low demand. If the number of riders is above the 75th percentile, the demand multiplier is set relative to the 75th percentile; otherwise, it's relative to the 25th percentile.
   * **Supply Multiplier:** Calculated similarly for high and low supply using percentiles for the number of drivers.
2. **Profitability Analysis:**
   * The code calculates the adjusted ride cost based on the demand and supply multipliers. It then computes the profit percentage for each ride by comparing the adjusted cost with the historical cost.
   * Rides with a positive profit percentage are identified as profitable, and those with a negative profit percentage are labeled as loss rides.
3. **Profitability Visualization:**
   * The code creates a donut chart to visually represent the distribution of profitable and loss rides. This provides a quick overview of the impact of dynamic pricing on ride profitability.
4. **Dynamic Pricing Adjustment based on Expected Ride Duration:**
   * The code creates a scatter plot illustrating the relationship between the expected ride duration and the adjusted ride cost. This allows for the visualization of how the dynamic pricing strategy adjusts costs based on the estimated duration of the ride.
5. **Time-Based Pricing Adjustment:**
   * The code introduces a time-based pricing adjustment by defining peak hours (17, 18, 19) and applying a time multiplier. Rides booked during peak hours receive a higher cost adjustment.
6. **Loyalty-Based Pricing Adjustment:**
   * An additional example is provided for a loyalty-based pricing adjustment. A discount is applied to rides for customers with a loyalty status of 'Silver'. This reflects a strategy of providing discounts to loyal customers.

**Training Predictive Model**

The training of the predictive model is a crucial step in the dynamic pricing project, involving the preparation of the data, conversion of categorical features, splitting the dataset for training and testing, and ultimately training a machine learning model.

1. **Data Preprocessing:**
   * **Identification of Features:** The code identifies numeric and categorical features in the dataset. Numeric features are those of type float or int, while categorical features are of type object.
   * **Handling Missing Values (Numeric Features):** Missing values in numeric features are filled with the mean of each respective column.
   * **Outlier Detection and Handling (Numeric Features):** The code utilizes the Interquartile Range (IQR) method to detect and handle outliers in numeric features. Values outside 1.5 times the IQR are replaced with the mean.
   * **Handling Missing Values (Categorical Features):** Missing values in categorical features are filled with the mode (most frequent value) of each respective column.
2. **Data Splitting:**
   * The dataset is split into training and testing sets using the train\_test\_split function from scikit-learn.
   * Features (x) include "Number\_of\_Riders," "Number\_of\_Drivers," "Vehicle\_Type," and "Expected\_Ride\_Duration."
   * The target variable (y) is "adjusted\_ride\_cost."
   * The test size is set to 20%, and a random seed ensures reproducibility.
3. **Model Training:**
   * A Random Forest Regressor is chosen as the predictive model. The model is instantiated and trained using the training set (x\_train, y\_train).

**Key Considerations:**

* **Random Forest Regressor:** The choice of a Random Forest Regressor for the predictive model indicates a decision to use an ensemble learning technique capable of handling non-linearity and capturing complex relationships in the data.
* **Data Reshaping:** The target variable (y) is reshaped into a 1D array, ensuring compatibility with the scikit-learn model.

**Key Considerations:**

* **User Input Prediction:** The ability to predict prices using user-input values showcases the practical application of the trained model for real-world scenarios.
* **Evaluation Metrics:** MAE, MSE, and RMSE provide a comprehensive assessment of how well the model generalizes to new, unseen data.
* **Feature Importances:** Understanding feature importances helps in interpreting the model and identifying key factors influencing predictions.