Uber Rides Data Analysis - Documentation and Recommendations

Prepared by :

- Parth Rana (ranaparth906@gmail.com)

Documentation for Each Step in the Code

1. Importing Libraries:

  - Import necessary libraries: `pandas` for data manipulation, `numpy` for numerical operations, `matplotlib.pyplot` for plotting, and `seaborn` for statistical data visualization.

2. Loading the Dataset:

  - Read the Uber dataset from a CSV file into a DataFrame named `uber`.
  - Display the first 10 rows of the DataFrame to understand the structure and content of the dataset.

3. Basic Data Exploration:

  - Use `describe()` to get summary statistics of the dataset.
  - Use `info()` to get a concise summary of the DataFrame, including the data types and non-null values.
  - Check for missing values using `isnull().sum()`.
  - Check for duplicate rows using `duplicated().sum()`.
  - Count the number of unique values in each column using `nunique()`.

4. Handling Missing Values:

  - Display the count of each unique value in the 'PURPOSE' column using `value_counts()`.
  - Fill missing values in the 'PURPOSE' column using forward fill method (`ffill`).
  - Display the first 10 rows to verify the changes.

5. Converting Date Columns:

  - Convert 'START_DATE' and 'END_DATE' columns to datetime format using `pd.to_datetime()`.
  - Display the first 10 rows to verify the changes.

6. Extracting Date and Time Components:

  - Extract date, time, month, and year from 'START_DATE' and 'END_DATE' columns.
  - Add these components as new columns in the DataFrame.

- Display the first 10 rows to verify the changes.

7. Dropping Original Date Columns:
  - Drop 'START_DATE' and 'END_DATE' columns from the DataFrame.
  - Rearrange the remaining columns in a specific order.

8. Handling Missing Values in 'Month' Column:
  - Display the value counts of the 'month' column.
  - Fill missing values in the 'month' column using the mode of the column.
  - Convert 'month' column to integer type.

9. Handling Missing Values in 'Year' Column:
  - Fill missing values in the 'year' column using the mode of the column.
  - Convert 'year' column to integer type.

10. Extracting Hour and Minute Components:
  - Extract hours and minutes from 'start_time' and 'end_time' columns.
  - Add these components as new columns in the DataFrame.

11. Rearranging Columns:
  - Rearrange the columns to drop 'start_time' and 'end_time' columns and organize the remaining columns in a specific order.

12. Calculating Duration:
  - Calculate the duration of each trip in minutes by subtracting 'START_DATE' from 'END_DATE'.
  - Add the calculated duration as a new column 'duration'.

13. Plotting Heatmap of Correlations:
  - Create a heatmap to visualize the correlations between numerical variables in the dataset.

14. Distribution of Months:
  - Plot the distribution of rides across different months using a count plot.

15. Distribution of Categories:
  - Count the occurrences of each category and create a bar plot to visualize the distribution of categories.

16. Distribution of Purpose:
   - Count the occurrences of each purpose and create a bar plot to visualize the distribution of purposes.
   - Rotate x-axis labels for better readability.

17. Count Plot of Purpose by Category:
   - Create a count plot to show the distribution of purposes by category, with the x-axis representing the purpose and the hue representing the category.

18. Distribution of Miles:
   - Plot the distribution of miles traveled using a distribution plot.

19. Distribution of Duration:
   - Plot the distribution of trip durations using a distribution plot.

20. Monthly Analysis:
   - Map month numbers to month names.
   - Calculate the value counts of rides for each month and the maximum miles traveled in each month.
   - Create a line plot to visualize the monthly ride counts and maximum miles.

21. Day Time vs Miles:
   - Plot the relationship between the time of day and the miles traveled using a bar plot.

22. Day Distribution:
   - Plot the distribution of rides across different days using a count plot.

23. Saving Cleaned Data:
   - Save the cleaned and processed DataFrame to a new CSV file named 'UberDatasetCleaned.csv'.

Recommendations Based on the Dataset

Based on the analysis of the Uber dataset, here are some recommendations for improving operations and customer experience:

1. Optimize Ride Allocation During Peak Months:
   - The analysis shows a variation in the number of rides across different months. It would be beneficial to allocate more drivers during peak months to handle the increased demand and reduce wait times for customers.

2. Target Marketing Campaigns by Purpose:
   - The dataset indicates the purposes for which customers use Uber rides. Tailored marketing campaigns can be designed to target specific purposes, such as business travel or leisure, to attract more customers.

3. Improve Services During Peak Hours:
   - The analysis of trip start times can help identify peak hours when demand is highest. Enhancing service availability during these hours can improve customer satisfaction and reduce waiting times.

4. Monitor and Reduce Trip Duration:
   - The analysis of trip durations reveals patterns that can be used to improve route planning and reduce trip times. Implementing better route optimization algorithms can help achieve this goal.

5. Enhance Services in High-Demand Categories:
   - Certain categories, such as business or personal travel, may show higher demand. Focusing on enhancing services for these categories, such as providing premium vehicles or additional amenities, can increase customer loyalty.

6. Seasonal Promotions:
   - Given the variation in ride counts across months, seasonal promotions can be introduced to boost ridership during off-peak months. Discounts or special offers can attract more customers during these periods.

7. Analyze High-Mileage Trips:
   - High-mileage trips could indicate longer journeys or potential inefficiencies in route planning. Analyzing these trips can help in understanding customer needs and optimizing routes

to save time and fuel.

8. Focus on Popular Routes and Destinations:
   - Identifying popular start and stop locations can provide insights into common routes and destinations. Uber can focus on these areas to enhance service quality, such as by setting up dedicated pick-up/drop-off points.

9. Customer Feedback Integration:
   - Continuously collecting and analyzing customer feedback can help identify pain points and areas for improvement. This can lead to better service adjustments and higher customer satisfaction.

10. Driver Training Programs:
   - Implementing training programs for drivers based on the analysis of trip data can help improve service quality. Topics such as efficient driving, customer service, and handling peak hours can be covered.

By implementing these recommendations, Uber can enhance its operational efficiency, improve customer satisfaction, and increase overall ridership.