Summary of Python Project: Uber Data Analysis

Introduction

The project focuses on analyzing Uber trip data to uncover insights about traffic patterns, payment methods, and trip distances.

Data Preparation

- **Dataset**: The dataset used is `yellow_tripdata_2013-12.parquet`.

- **Initial Inspection:** The dataset initially contains 13,971,118 rows and several columns, including `congestion_surcharge` and `airport_fee`, which are entirely null.

- **Data Cleaning:**

  - Removed columns with all null values (`congestion_surcharge` and `airport_fee`).

  - Dropped duplicate rows, reducing the dataset to 13,970,973 rows.

Analysis

1. Traffic Congestion Analysis:

  - Traffic congestion peaks between 8 AM and 12 PM, indicating heavy morning rush hours.

2. Payment Method Analysis:

  - **Most Preferred Payment Method:** Credit cards are the most preferred payment method among Uber users.

  - **Visualization:** A bar plot shows the number of trips by each payment method, emphasizing the dominance of credit card payments.

3. Trip Distance Analysis:

  - **Distance Categorization**: Trips are categorized into distance bins (0-5 miles, 5-10 miles, etc.).

  - **Findings:** Most Uber trips fall in the 0-5 miles category.

- **Visualization:** A bar plot illustrates the number of trips in each distance category.

Key Insights

- **Traffic Patterns:** Congestion is highest in the mornings, likely due to commuters.

- **Payment Preferences:** Credit cards are favored for their convenience and security.

- **Trip Distances :** Short trips (0-5 miles) are the most common among Uber users.

Implications

- **For Businesses :** Enhancing services around peak traffic times and optimizing payment processing systems to favor credit card transactions can improve customer satisfaction.

- **For Future Research:** Further analysis could explore seasonal variations, the impact of weather on trip patterns, or demographic preferences for payment methods.