

Ford GoBike Data Exploration

Sadikovic Samir

Project Overview

The goal of this project is to perform an exploratory data analysis (EDA) on the Ford GoBike dataset to uncover insights into the usage patterns, user demographics, and other interesting aspects of the bike-sharing service in the San Francisco Bay Area.

Dataset

The Ford GoBike dataset contains information about bike-sharing services in the San Francisco Bay Area. The dataset includes details about trips, such as start and end times, start and end stations, trip duration, user information like age and gender, and user type (subscriber or customer).

Data Source

The dataset is provided by Ford GoBike and is available on [\[link to the dataset source if available online\]](#).

Data Cleaning and Preparation

1. **Handling Missing Values:** Missing values in critical columns were handled by either removing rows with missing data or imputing values where appropriate.
2. **Data Type Conversion:** Columns were converted to appropriate data types, such as converting `start_time` and `end_time` to datetime objects.
3. **Feature Engineering:** New features such as `age` and `start_hour` were created to aid in the analysis.

User Demographics

- The majority of users are in their 20s and 30s.
- There are very few users over 60 years old.
- Male users dominate the dataset, followed by female users, and a small portion of users identify as other genders.

Popular Stations and Routes

- The station "Market St at 10th St" is the most popular start station.
- Certain routes are more popular with subscribers, while others are preferred by customers.

Trip Start Times

- Peak usage times are around 8 AM and 5 PM, which correspond to typical commuting hours.
- The highest number of trips occur on weekdays, especially on Thursday, with

Key Insights for Presentation

Main Points for the Presentation

- **Trip Duration:** Most trips are short. This suggests that the service is primarily used for short, regular commutes.
- **User Demographics:** The main users are young adults, with the majority being male. This information can be useful for targeting and marketing strategies.
- **Popular Stations and Routes:** Centrally located and well-connected stations are the most popular. This can be considered for planning and expanding the service.
- **Usage Patterns by Time of Day:** Peak usage during commuting hours indicates the main usage times, which can help with resource planning and service optimization.

These findings provide valuable insights into the usage patterns and demographic trends of the Ford GoBike service. They can help optimize the service and target specific user groups more effectively.

Future Work

- **Deeper Analysis of Usage Patterns:** Investigate usage patterns across different days of the week and months of the year.
- **Impact of Weather:** Analyze how weather conditions affect bike usage.
- **Comparison with Other Cities:** Compare the usage patterns with bike-sharing services in other cities.

Instructions to Run the Analysis

1. **Clone the Repository:** `git clone [repository link]`
2. **Install Dependencies:** `pip install -r requirements.txt`
3. **Run the Jupyter Notebook:** Open `Part_II_slide_deck_template.ipynb` in Jupyter Notebook and run all cells to see the analysis and visualizations.

Acknowledgements

- Ford GoBike for providing the dataset.
- [Any other acknowledgements]