Data-Driven User Behavior Analysis Using Shiny

Manpreet Kaur M, Ashish Juttu, Tanaka M
Masters in Applied Data Science, Clarkson University

December 2, 2024

Abstract

This project analyzes six months of Divvy bike-sharing data to uncover user behavior and operational patterns. Key metrics such as trip durations, popular stations, and user types are examined to provide insights into peak usage times, routes, and geospatial hotspots. Findings support strategies to improve user engagement and operational efficiency. A Shiny dashboard is used for dynamic, interactive visualizations, making data exploration accessible to stakeholders.

Introduction

Bike-sharing systems like Divvy promote sustainable urban transportation. This project analyzes six months of Divvy's data to uncover trends and differences in user behavior between casual users and members. The study focuses on trip durations, station popularity, and geospatial trends to support service quality improvement, resource optimization, and customer satisfaction.

Objectives:

- Understand usage differences between casual users and members.
- Identify high-demand routes and peak usage times.
- Provide data-driven recommendations for operational optimization.

Dataset and Methods

The dataset, sourced from Divvy's public repository, includes rideable type, user type, trip duration, timestamps, and geospatial data. Preprocessing involved handling missing values, filtering irrelevant records, and aggregating data for visualization. Challenges like memory constraints were addressed using optimized packages (`data.table`, `dplyr`) and reactive caching in Shiny.

Shiny Application Features

The Shiny app provides:

- **Overview:** Key metrics (total rides, trip durations, user breakdown).

- **User Behavior:** Time-series trends with filtering by user type and bike type.
- **Station Insights:** Geospatial heatmaps for popular start and end stations.
- **Custom Visualizations:** Dynamic charts based on user-selected dimensions.

Challenges and Solutions

Handling a 6M+ record dataset required memory-efficient methods and pre-aggregated data for visualization scalability. Rendering interactive dashboards with minimal lag was achieved through caching and limiting real-time computations. Cross-system reproducibility was ensured through optimized coding practices.

Conclusion

This project demonstrates how interactive dashboards can deliver actionable insights into bike-sharing usage patterns. The methodologies and findings highlight the importance of data preprocessing and visualization design in supporting operational excellence.

References

[1] Divvy Bikes Data. Available at `https://www.kaggle.com/datasets/laerozzo/divvy-data-09-2023-09-2024`.

[2] Ashish Juttu. *Data-Driven User Behavior Analysis.* Available at `https://www.kaggle.com/code/ashishjuttu/data-driven-user-behaviour-analysis`.