Cyclistic Case Study

Introduction

Cyclistic, a Chicago based bike-share company, is interested in examining the habits and usage patterns of its members compared with those of casual users. Insights from this study will inform marketing strategies to convert casual riders into annual members.

Data Sources and Preparation

Trip data was sourced from the official Divvy repository on AWS S3: divvy-tripdata.s3.amazonaws.com. Initial organization and compilation occurred in a dedicated Kaggle notebook: Cyclistic Case Data. Datasets include rider demographics (gender, birth year for 2014–2019) and detailed trip and station data for 2020–2025, all organized and combined in a dedicated Kaggle notebook titled “Cyclistic Case Data.”

Preprocessing Steps

• Converted data types in relevant columns (e.g., dates, categoricals) to optimize storage and performance.

• Trimmed leading/trailing spaces from start/end station names for consistency.

• Filtered millions of NA values and normalized timestamps in started_at/ended_at columns.

Because the original tables contain a substantial amount of NA's, a secondary table was created that excludes rows with NA values.

Analysis and Findings

Initial exploration focused on comparing overall ride volume between members and casual riders. Bike type preferences were then examined by rider type to identify which bikes casual users favor among the available options. Time patterns were analyzed by parsing started_at into day-of-week and month fields, allowing comparison of usage over weekdays vs. weekends and across seasons. Casual riders appear at a higher rate in recreational windows such as weekends and summer months and also tend to take longer rides than members. Station level traffic was evaluated for higher traffic among casual riders, but given the existence of more than

2,000 stations, the analysis narrowed to the top 10 stations for each rider type, highlighting a small set of high-traffic locations for casual riders. The second dataset focused more on demographics, which enabled exploration of age and gender patterns in usage across rider types. Unlike the trip and station tables, the rider table included a dependent category in addition to member and casual classifications; because dependents could not be reliably reclassified into either group, rows labeled as dependents and those with missing values were removed from the demographic analysis. The cleaned results suggest that males are the predominant users of the service, with the highest concentration in the 25–34 age range.

Conclusions and Recommendations

Expanding marketing efforts around the top high traffic stations identified for casual riders may strengthen conversion opportunities where engagement is already high. Targeted campaigns toward men can help sustain the current core user base, while tailored messaging and incentives for women may broaden overall membership. Since casual usage peaks in warmer months and on weekends, introducing seasonally focused offerings such as a summer membership tier or limited time membership promotions at the start of peak season may effectively convert high usage casual riders into members.

References

Create elegant data visualisations using the grammar of graphics. (n.d.). https://ggplot2.tidyverse.org/

Make Dealing with Dates a Little Easier. (n.d.). https://lubridate.tidyverse.org/

Posit Cloud - Do, share, teach, and learn data science. (n.d.). https://posit.cloud/learn/recipes

Perplexity AI. (2025, December 30). https://www.perplexity.ai