Cyclistic Case Study
Converting Casual Riders into Annual Members

A 17-Month Behavioral Analysis Using R, Tableau, and Public Divvy Data
Time Period: January 2024 – May 2025
Total Trips Analyzed: 5,111,422
Tools Used:

- R (Data Cleaning & Feature Engineering)
- Tableau (Interactive Dashboards)
- Kaggle (Data Hosting)
- GitHub (Code & Documentation)

Prepared by: Vishvam Shastri
Published on: July 31, 2025

Over a 17-month period, from January 2024 to May 2025, Cyclistic riders completed over 5.11 million trips across Chicago's neighborhoods, business districts, parks, and lakefront paths. Within this extensive dataset lay a persistent marketing challenge: casual riders were generating substantial seasonal activity (contributing 42% of summer rides but only 19% in winter), but very few were converting into annual members. The primary goal of this case study was to investigate ride behavior patterns across casual and member riders in order to generate actionable recommendations for improving membership conversion.

The data used for this case study was obtained from Divvy's public bike-share system data, which is made available by Lyft on behalf of the City of Chicago. The dataset includes anonymized ride-level information covering all the selected months. This source meets the ROCCC framework: it is reliable, having been published directly by the system operator; original, as it originates from Divvy's internal trip logs; comprehensive, covering millions of rides across multiple months and seasons; current, including recent data up to mid-2025; and cited, with attribution clearly stated in alignment with licensing requirements. The data is publicly available under Divvy's Data License Agreement, which permits analysis and publication with attribution, subject to specific terms and conditions.

The analysis began with assembling and consolidating 17 monthly CSV files into a single structured dataset using R. The map_dfr() function was used to efficiently combine all files row-wise, resulting in a dataset with over 5.11 million records. Before any analysis could begin, rigorous data cleaning and preparation were necessary. The data was filtered to remove 2.1 million records (22% of raw data), including:

- 209,464 incomplete records
- 46,807 rides shorter than 1 minute
- 200 trips with implausible speeds over 60 km/h
- 22,858 "short loop" station returns (same start/end in <2 minutes)

Each cleaning step was documented in a structured audit log for reproducibility and transparency.

To prepare the data for deeper analysis, additional features were engineered. Calendar-related fields, including day of the week, hour of day, and seasonal labels, were derived from ride timestamps. A weekend/weekday flag was introduced to distinguish between leisure and commuter patterns (casual riders accounted for 38% of weekend rides vs. 23% for members). Spatial clustering was made possible by rounding latitude and longitude to three decimal places. Ride durations were categorized into buckets labeled as Short (under 10 minutes), Medium (10-30 minutes), Long (30-60 minutes), and Very Long (over 60 minutes). These enhancements enabled a more refined breakdown of patterns across user types, time intervals, and geography.

The cleaned dataset revealed several important behavioral differences. Casual riders exhibited a pronounced seasonal trend, with ride volume increasing significantly between May and October months (peaking at 228,234 rides in July 2024). While a higher number of rides still

came from members, casual riders had a higher proportion of their activity concentrated during summer months. Weekday ride comparisons revealed that casual activity peaked between 4 PM and 6 PM (accounting for 28% of their daily rides), especially in the warmer months, indicating recreational or flexible post-work usage. In contrast, member usage was more evenly distributed across the year, with clear peaks around 8 AM and 5 PM on weekdays (representing 32% of all member rides), underscoring the commuter nature of their behavior.

Spatial analysis further emphasized this behavioral divide. Casual riders frequently began rides at locations near tourist and recreational zones, such as "Streeter Dr & Grand Ave" (54,781 rides) and "DuSable Lake Shore Dr & Monroe St" (37,333 rides). Members, on the other hand, preferred stations situated near business centers and transit hubs, including "Clinton St & Washington Blvd" (32,263 rides), "Canal St & Adams St", and "Clinton St & Lake St". These station-level insights can help prioritize promotional campaigns in zones with higher conversion potential.

Bike type preferences also differed between the two groups. Casual riders demonstrated a higher preference for electric bikes, which made up approximately 37 percent of their rides (vs. 36% for members). Members leaned more heavily toward classic bikes (64% of their rides). This distinction may reflect comfort and convenience priorities among casual users, especially during longer or scenic rides.

Ride duration analysis confirmed that casual riders consistently took longer trips on average. For example, casual rides averaged 24 minutes vs. 12 minutes for members, with median durations of 13.4 minutes vs. 8.8 minutes. A boxplot comparison of ride durations confirmed this trend, even after removing outliers. These findings point to a clear opportunity: marketing campaigns can emphasize the cost benefits of membership, particularly for users who frequently take longer rides (a casual rider taking 10 monthly 30-minute rides at $5.10/ride would spend $612/year vs. a $108 annual membership - saving $504 annually).

The analysis also identified a high-value segment among casual users: "power casuals" - riders taking 5+ same-route rides in a single day. These users showed extreme repeat behavior at tourist hubs, with 135 repeat rides in one day observed at Streeter Dr & Grand Ave. These power users are ideal candidates for real-time promotions (e.g., "You've taken 3 rides today - upgrade now to save $38.50 on today's rides alone!"), as they could save $1,738 annually by switching from pay-per-ride to membership.

Insights were visualized across eleven interactive dashboards built in Tableau, each focused on a key theme such as ride volume trends (showing casual riders contributing 58% of summer rides), temporal usage (3-7 PM weekday peaks), bike type preferences, route patterns, and spatial clusters. Visual tools included grouped bar charts, boxplots, treemaps, heatmaps, slope charts, and dual-maps.

The findings from this case study point to several strategic opportunities for Cyclistic's marketing team:

- Launch targeted May-October campaigns when casual ridership is highest (summer casual rides are 2.3x winter volumes), emphasizing cost savings: "Streeter Dr riders save $972/year with membership"
- Place membership kiosks at top 5 casual stations (representing 18% of all casual rides) with localized messaging: "5+ rides/week? Stop paying $1,846 - join for $108!"
- Bundle e-bike access with membership (37% casual adoption vs. 36% member)
- Implement real-time upsells for power casuals after 3+ same-day rides: "Today's rides would cost $38.50 - free as a member!"
- Highlight cost savings for frequent/long riders: "30-minute riders save $504 annually"

By combining methodical data preparation with targeted analytics and visualization, this project transformed over five million ride records into actionable strategies. The result is a data-backed roadmap that enables Cyclistic to engage casual users more effectively and grow its annual membership base with precision and purpose.

While this case study focused on identifying behavioral differences between casual and member riders to support membership conversion strategies, the dataset holds further untapped potential for ongoing analysis. Future investigations could explore weather effects on ride patterns by integrating meteorological data, study station-level imbalances to optimize bike rebalancing operations, or apply predictive modeling to anticipate high-demand locations and times. Additional segmentation analysis like tourist vs. local usage patterns, or first-time vs. repeat casual riders could also refine targeting strategies. With deeper machine learning techniques or time-series forecasting, Cyclistic can further align operational decisions and marketing campaigns with dynamic rider behavior trends across seasons, neighborhoods, and rider types.