Case Study: Cyclistic Bike Share

1. A clear statement of the business task (ask)

Business Task

The objective of this analysis is to examine how casual riders and annual members use Cyclistic's bike-share service differently. By identifying key differences in riding behavior between these two customer segments, this analysis aims to generate insights that can inform data-driven marketing strategies designed to convert casual riders into annual members.

Key Stakeholders

- Lily Moreno, Director of Marketing
- Cyclistic Marketing Analytics Team
- Cyclistic Executive Team

Business Relevance

Annual members are significantly more profitable than casual riders, making membership growth a critical factor in Cyclistic's long-term success. A deeper understanding of user behavior patterns will enable Cyclistic to target the right customers with tailored marketing initiatives, improve conversion rates, and maximize overall revenue.

2. A description of all data sources used (prepare)

Data Source

This analysis uses **Cyclistic historical bike trip data** made publicly available by **Motivate** International Inc. under a data license agreement. The dataset contains detailed, trip-level records of bike-share usage in Chicago and is appropriate for analyzing customer behavior patterns.

For this case study, 12 monthly Excel (.xlsx) files were used, each representing one month of bike trip data from January 2025 through December 2025.

Data Structure and Contents

Each monthly dataset is organized in a long format, where each row represents a single bike trip. All files share a consistent schema and include the following variables:

- ride_id: Unique identifier for each trip
- rideable_type: Type of bike used
- started_at: Trip start date and time
- ended_at: Trip end date and time

- start_station_name and start_station_id: Trip start location
- end_station_name and end_station_id: Trip end location
- start_lat and start_lng: Start station coordinates
- end_lat and end_lng: End station coordinates
- member_casual: Rider type (casual rider or annual member)

The presence of start and end timestamps enables time-based analysis, while the member_casual variable allows direct comparison between casual riders and annual members, which is central to the business task.

Data Organization and Storage

The original monthly files were preserved as raw data to maintain data integrity. For analysis purposes, the 12 monthly datasets were combined programmatically into a single analytical dataset, ensuring consistent structure and enabling year-level analysis across all rides. This approach avoids manual handling of large files while supporting scalable and reproducible analysis.

Data Credibility and Limitations (ROCCC Evaluation)

- **Reliable:** The data is sourced from Cyclistic's operational systems and published by Motivate International Inc.
- **Original:** The dataset represents first-party data collected directly from bike-share usage.
- **Comprehensive:** A full year of trip data captures usage patterns across weekdays, weekends, and seasons.
- **Current:** The data reflects recent bike-share activity from 2025.
- **Cited:** The dataset is publicly available and licensed for analytical use.

Limitations:

- Personally identifiable information is excluded to protect user privacy.
- Demographic details such as age, gender, or home location are not available.
- Individual riders cannot be tracked across multiple trips, limiting user-level behavioral analysis.

Despite these limitations, the dataset is well-suited for identifying usage differences between casual riders and annual members, directly supporting the objectives of this analysis.

3. Documentation of any cleaning or manipulation of data (process)

To prepare the data for analysis, several cleaning and transformation steps were performed to ensure accuracy, consistency, and usability. All data processing was conducted

programmatically using R, which allowed scalable and reproducible handling of large datasets.

Data Combination

The original dataset consisted of 12 separate CSV files, each representing one month of Cyclistic bike trip data for the year 2025. All files were imported into R and combined into a single dataset using row-binding techniques. This created a unified, year-level dataset suitable for comprehensive analysis while preserving the original raw files unchanged.

Datetime Parsing and Validation

The started_at and ended_at columns were converted from character format into proper datetime objects using R's lubridate package. This step ensured that timestamps were correctly interpreted and enabled accurate time-based calculations, such as ride duration and day-of-week analysis.

Feature Engineering

Two new variables were created to support the business analysis:

- **Ride Length (**ride_length**)**
  Ride duration was calculated as the difference between ended_at and started_at, measured in minutes. This variable was essential for comparing usage behavior between casual riders and annual members.
- **Day of Week (**day_of_week**)**
  A new column was created to capture the day on which each ride started. The variable follows the convention Sunday = 1 through Saturday = 7, consistent with the project guidelines and the Excel WEEKDAY(date, 1) function.

Data Cleaning and Filtering

To improve data quality and analytical accuracy, the following records were removed:

- Trips with negative or zero ride duration, which likely resulted from system or data entry errors
- Records with missing or invalid ride duration values
- Records with missing user type (member_casual)

Trips with missing station names or station IDs were retained, as these fields were not required for the core analysis.

Final Dataset

After cleaning and transformation, the resulting dataset contained only valid bike trips with correctly formatted datetime values and newly engineered variables. This cleaned dataset was used for all subsequent analysis and visualization steps.

4. A summary of your analysis (analyze)

The analysis examined behavioral differences between casual riders and annual members using Cyclistic's 2025 trip data. Key dimensions included ride length, temporal usage patterns, bike type preferences, and station popularity.

Annual members account for a larger share of total rides and demonstrate consistent usage patterns across weekdays and commute hours. Their rides are generally shorter and less variable, indicating that Cyclistic bikes are primarily used for functional, routine transportation such as commuting.

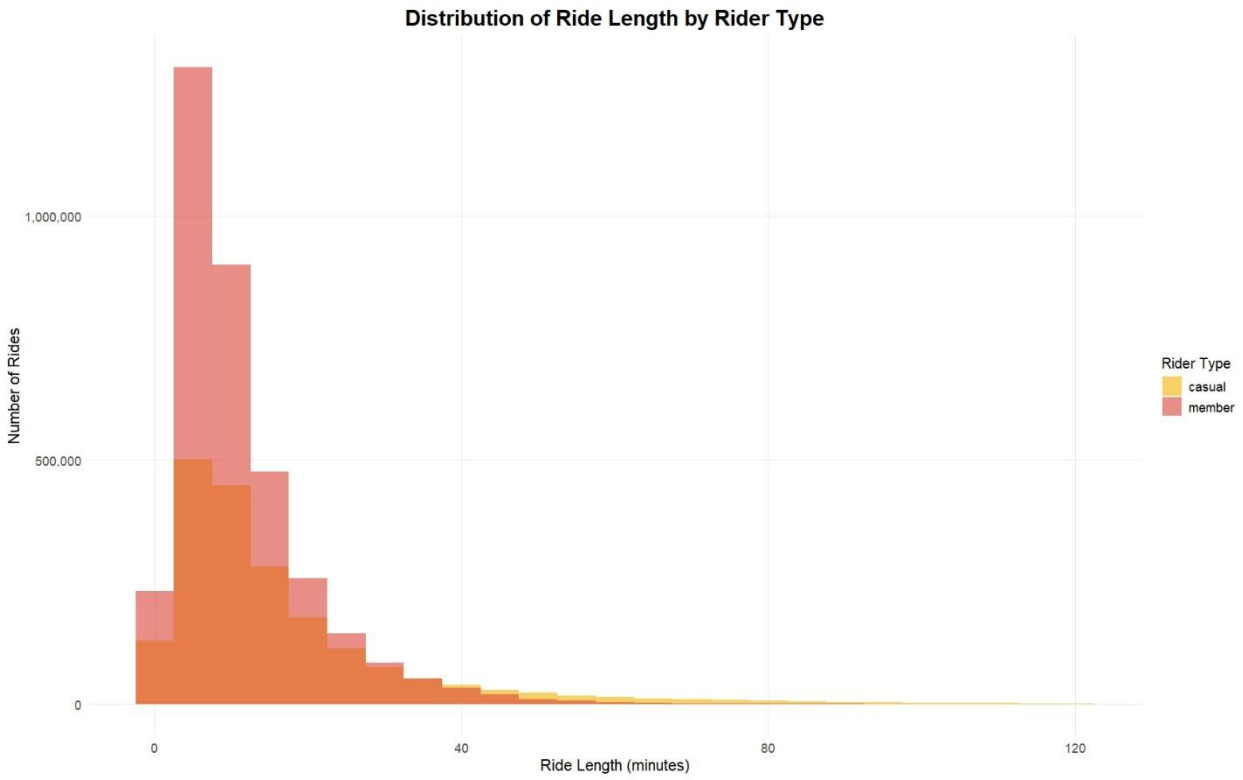
In contrast, casual riders take fewer total rides but exhibit significantly longer and more variable ride lengths. Their activity peaks on weekends, during evenings hours, suggesting a leisure-oriented use of the service. Casual riders are also more sensitive to seasonality compared to members.

Station usage analysis highlights strong geographic differences between the two groups. Members' most frequently used stations are concentrated near residential areas, transit hubs, and business districts, while casual riders favor stations near parks, waterfronts, and recreational destinations.

Overall, the analysis confirms that casual riders and annual members represent two distinct user segments with different motivations, behaviors, and usage patterns. These insights form the foundation for targeted strategies aimed at converting casual riders into annual members.

5. Supporting visualizations and key findings (share)

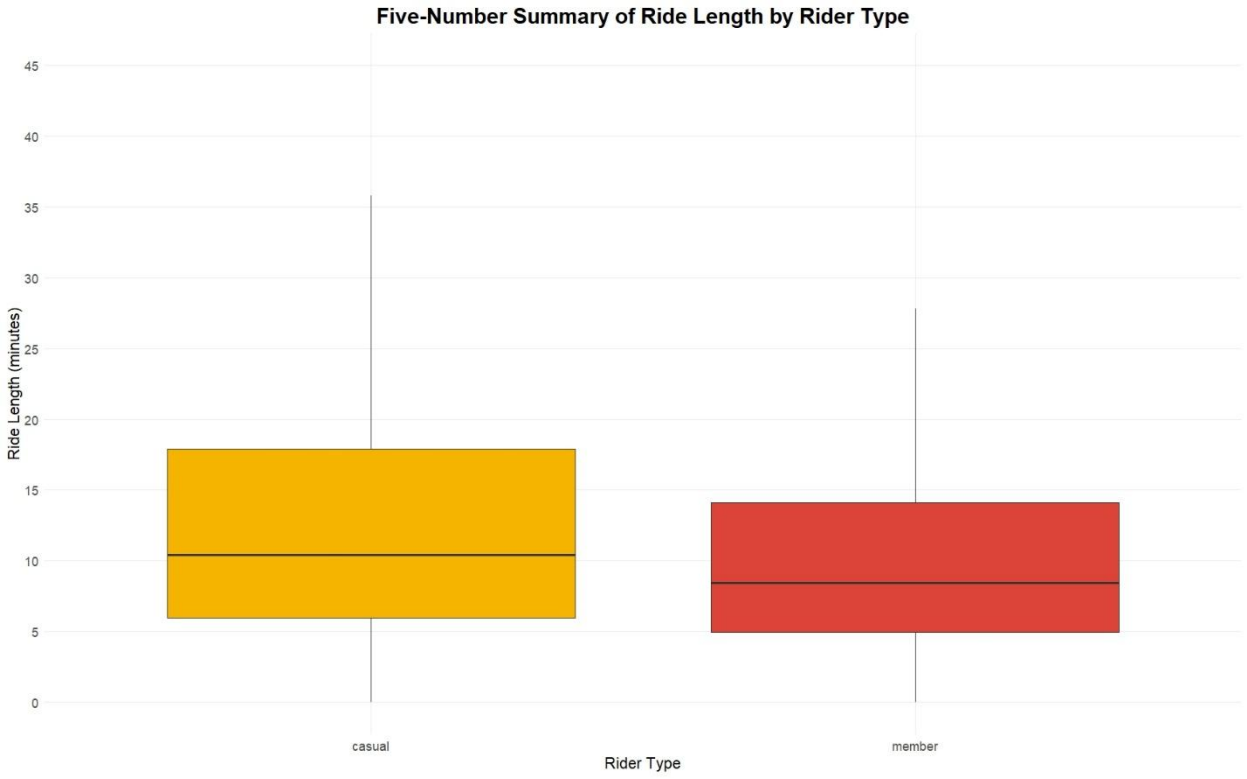
Figure 1: Distribution of Ride Length by Rider Type



Key findings:

- Casual riders show longer and more variable ride durations.
- Annual members' ride lengths are shorter and more consistent.
- Indicates leisure use by casual riders versus functional use by members.

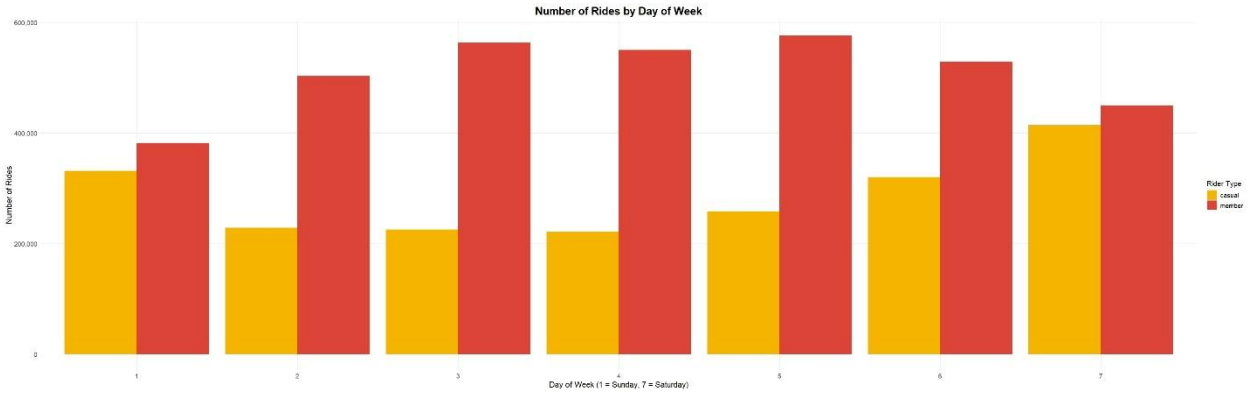
Figure 2: Five-Number Summary of Ride Length (Boxplot)



Key findings:

- Median and upper-quartile ride lengths are higher for casual riders.
- Members exhibit less variability in ride duration.
- Confirms ride length distribution patterns observed in the histogram.

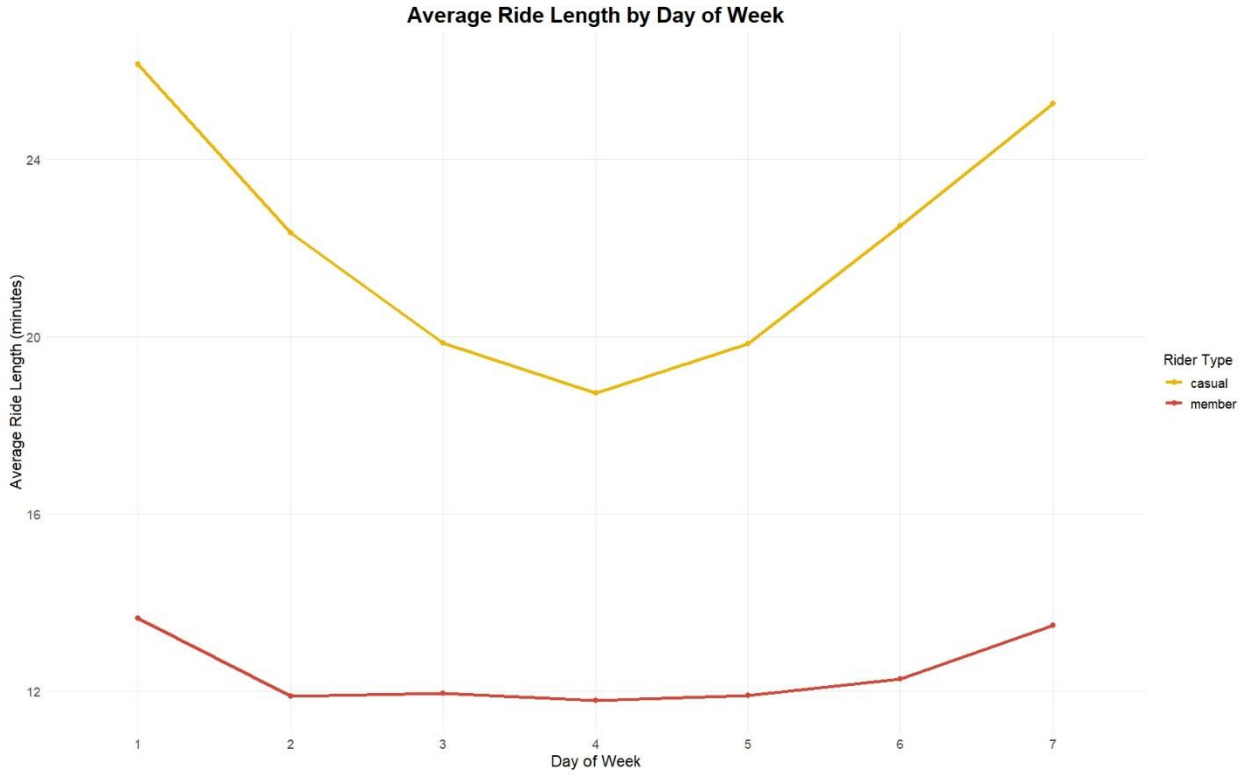
Figure 3: Number of Rides by Day of Week



Key findings:

- Casual rider activity peaks on weekends.
- Member usage remains steady throughout the week.
- Suggests casual riders are driven by free time availability.

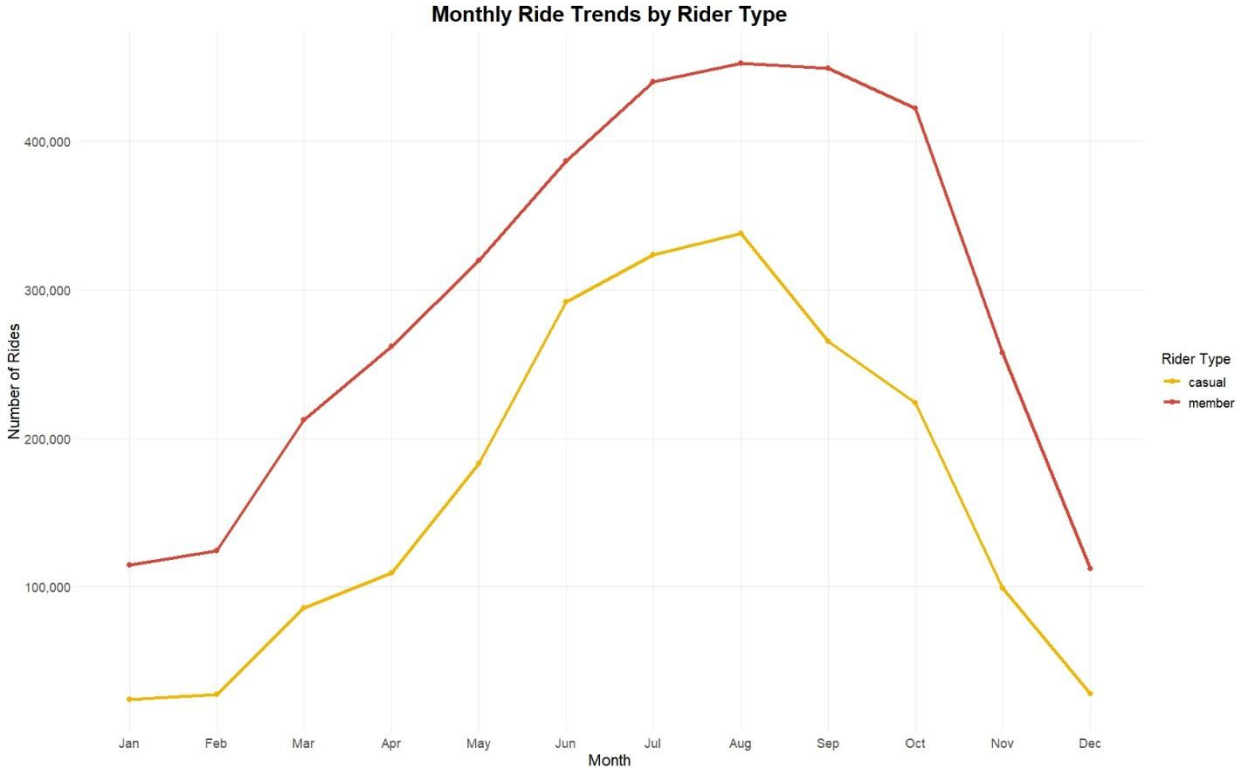
Figure 4: Ride length by day of week



Key findings

- Huge difference in ride length by group
- Members ride length varies slightly from weekdays to weekends
- But casual riders show huge variability across days

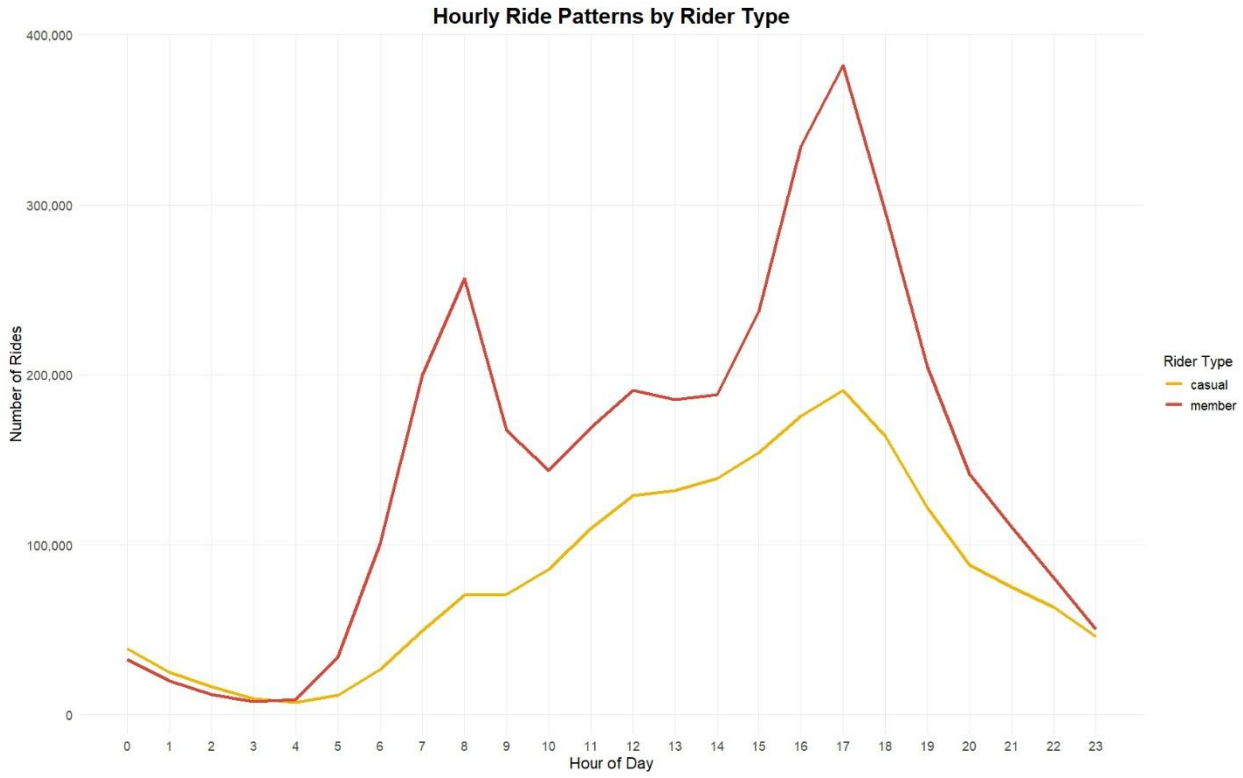
Figure 5: Monthly Ride Trends by Rider Type



Key findings:

- Both rider types peak during summer months.
- Casual riders show sharper seasonal fluctuations.
- Members demonstrate more stable year-round usage.

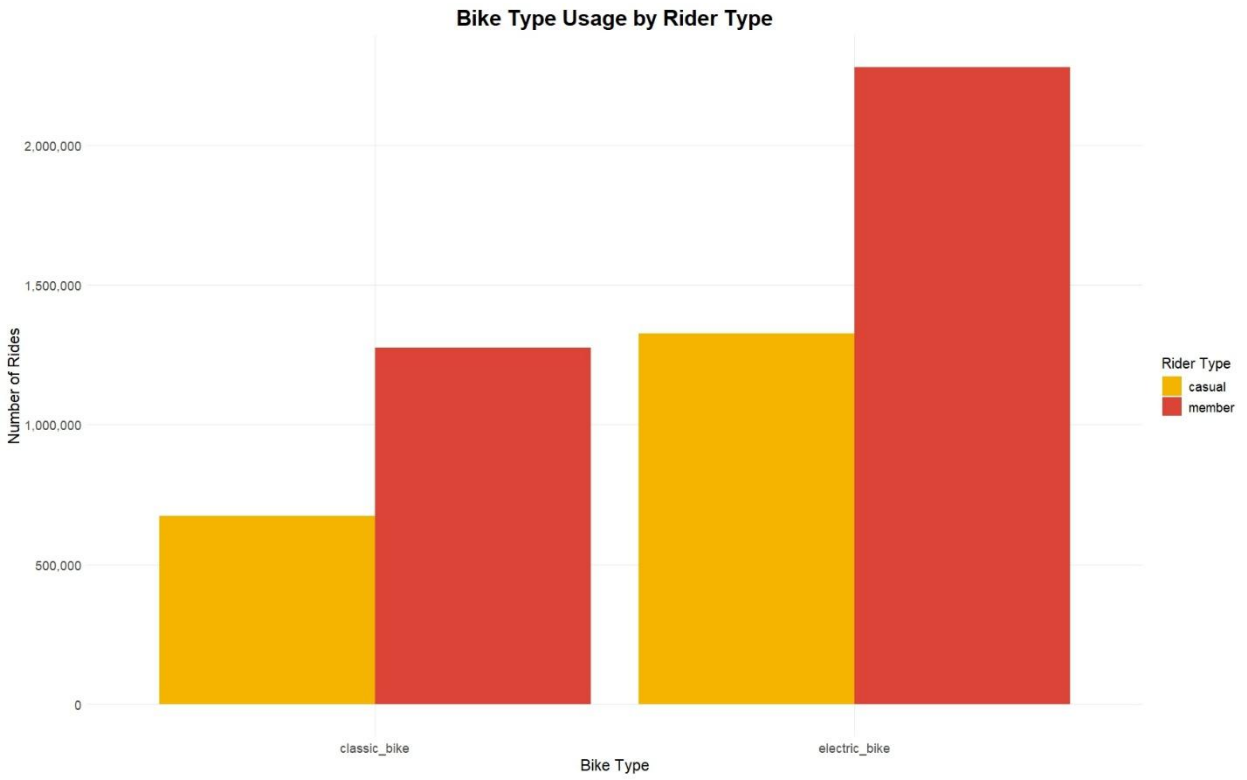
Figure 6: Hourly Ride Patterns



Key findings:

- Members peak during morning and evening commute hours.
- Casual riders peak during evening
- Reinforces commuting vs leisure usage patterns.

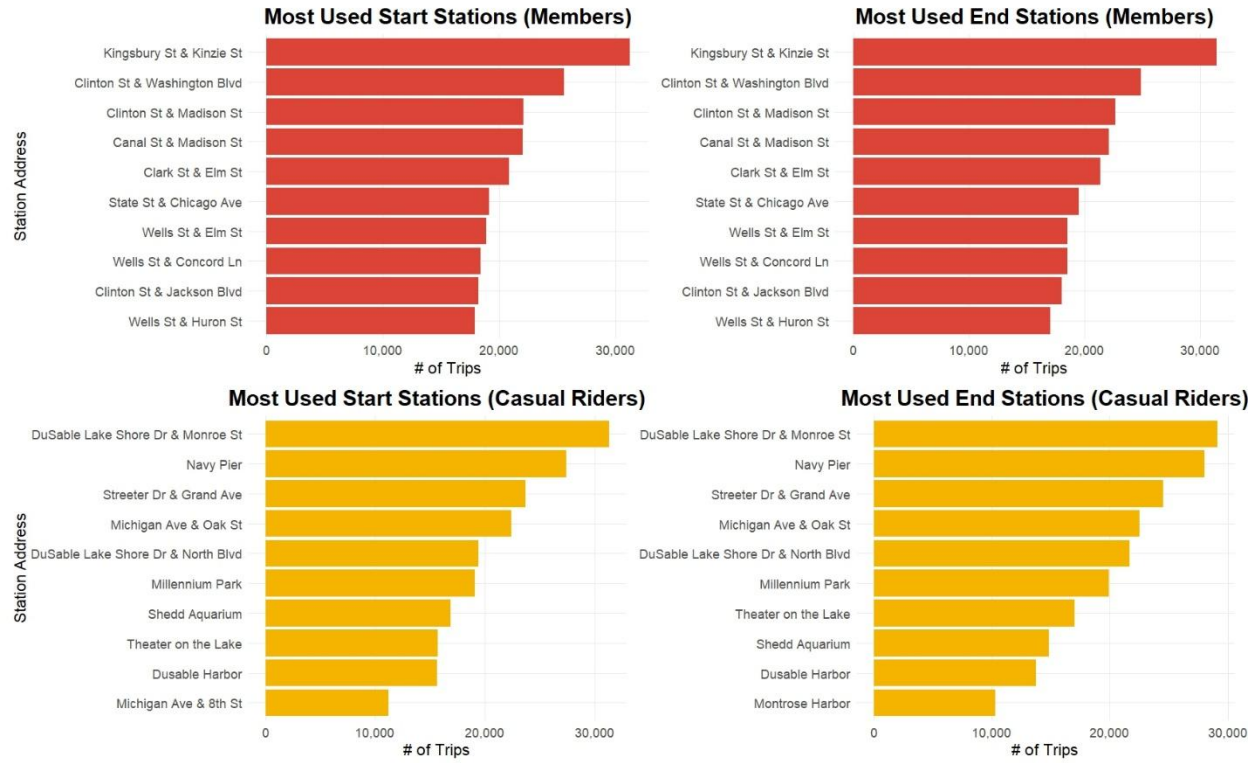
Figure 7: Bike Type Usage by Rider Type



Key findings:

- Both groups prefers electri bike over classis bike.
- Both groups show similar variations in bike type usage.

Figure 8: Most Used Start & End Stations by Rider Type (2×2 Plot)



Key findings:

- Members' top stations cluster around business and transit areas.
- Casual riders favor stations near recreational locations.
- Minimal overlap in top stations highlights distinct travel purposes.

6. Your top three recommendations based on your analysis (act)

Recommendation 1: Target Casual Riders with Weekend-Focused Membership Trials

Casual riders demonstrate high weekend and leisure usage. Cyclistic should offer weekend-only or short-term membership trials promoted at high-traffic recreational stations to encourage conversion without requiring immediate long-term commitment.

Recommendation 2: Promote Membership Value Through Ride-Length Savings

Because casual riders take longer rides, Cyclistic should highlight cost savings for extended trips under annual memberships. Messaging should emphasize "unlimited longer rides" and reduced per-minute costs compared to single-ride pricing.

Recommendation 3: Use Station-Based Marketing at Leisure Hotspots

Marketing campaigns should be deployed at parks, waterfronts, and tourist-heavy stations, where casual rider activity is highest. Digital signage and in-app prompts at trip end can encourage riders to "upgrade".

Final Business Impact

By aligning pricing, messaging, and marketing locations with observed casual rider behavior, Cyclistic can increase membership conversions while preserving its strong commuter base. These data-driven strategies leverage existing usage patterns rather than attempting to change rider behavior.