2025

# Cyclistic Case Study – Final Report

## GOOGLE DATA ANALYTICS CERTIFICATE – CAPSTONE PROJECT

JULIO FERREIRA

## Table of Contents

# Introduction

Cyclistic is a bike-share program in Chicago with a fleet of 5,824 bicycles and 692 docking stations, operating since 2016. The program offers flexible pricing plans, such as single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or day passes are classified as casual riders, while those with annual memberships are considered Cyclistic members.

Financial analysts have found that annual members are significantly more profitable than casual riders. While flexible pricing has attracted a broad range of users, Lily Moreno, Director of Marketing, believes future growth depends on increasing the proportion of annual members. Rather than targeting entirely new customers, Moreno emphasizes that casual riders are an ideal audience for conversion, as they are already familiar with Cyclistic's system and have demonstrated clear demand.

In this context, the marketing analytics team was assigned the task of analyzing historical trip data to uncover differences between casual riders and annual members, to explore possible motivations that could drive casual riders to purchase memberships, and to determine how digital media strategies can support this conversion. The analysis must provide evidence-based insights to guide the executive team's strategic decisions.

---

# Ask Phase

The starting point of the project is to clearly define the business task and the guiding questions that will direct the analysis. Cyclistic's challenge is not only to understand past usage patterns but to generate actionable insights that can support a targeted marketing campaign aimed at converting casual riders into members.

The business task is defined as follows:

- Use Cyclistic's historical bike trip data to identify differences in service usage between annual members and casual riders.
- Determine factors that could motivate casual riders to purchase annual memberships.
- Explore how digital media can be leveraged to encourage this conversion.

From this task, three guiding questions emerge to drive the analysis:

1. How do annual members and casual riders differ in terms of trip frequency, duration, and usage patterns?
2. What motivations or factors could prompt casual riders to purchase annual memberships (e.g., convenience, time of day, day of the week, trip duration, popular stations)?
3. How can Cyclistic use digital media to influence casual riders to become members?

Answering these questions will provide a framework for designing campaigns that are data-driven and aligned with Cyclistic's profitability goals.

---

## Prepare Phase

To answer the business questions, the first step was to collect and prepare the data. The datasets used in this analysis come from the Divvy bike trip system, the official source powering Cyclistic's operations. Specifically, the selected data includes trips from Q1 2019 and Q1 2020, downloaded directly from the Divvy system and stored locally in CSV format for accessibility and ease of processing. Each dataset contains one row per trip and includes key fields such as:

- trip_id
- start_time and end_time
- start_station_id and start_station_name
- end_station_id and end_station_name
- user_type (casual or member)
- start_lat and start_lng (latitude and longitude)
- end_lat and end_lng (latitude and longitude)

Before analysis, the dataset's integrity, credibility, and completeness were evaluated. During data exploration and cleaning, several issues were identified that could distort the results if not addressed:

1. **Negative and zero durations** – 117 trips with negative duration and 93 trips with zero duration were detected. All originated and ended at the same station (HQ QR, station_id = 675), indicating system test rides or technical unlock events.

2. **Extreme durations** – 7,943 trips had durations shorter than 1 minute (7,461 trips) or longer than 24 hours (1,440 minutes) (482 trips). These values are unrealistic and would bias averages and distributions.
3. **Duplicates** – No duplicate trips were found.
4. **Missing values** – Some start and end stations lacked coordinates (40 and 60 respectively). These cases were retained, as they were not significant enough to bias aggregate-level analysis.

Trips with negative, zero, or extreme durations were replaced with NA values and excluded from descriptive calculations, totaling **8,153 treated rides**.

After cleaning, the dataset was evaluated against the **ROCCC** criteria (Reliable, Objective, Complete, Consistent, Current) and meets all dimensions:

- **Reliable** – obtained from the official Divvy system.
- **Objective** – reflects raw trip records without subjective interpretation.
- **Complete** – covers all trips from the selected periods.
- **Consistent** – standardized after cleaning with harmonized formats.
- **Current** – provides recent-enough data (2019–2020) to generate relevant insights.

All data is anonymized to ensure privacy and security and stored in CSV format for accessibility. At this stage, the data is credible, clean, and ready for analysis.

---

# Process Phase

With the data prepared, the next step involved cleaning, transformation, and feature engineering to enable meaningful analysis. The tools chosen were **R** and the **tidyverse** package, due to their efficiency in handling large datasets and performing reproducible transformations.

The main steps carried out during this phase included:

## 1. Standardization of Columns

The 2019 and 2020 datasets were originally structured using different naming conventions and field definitions. To combine them into a unified analytical dataset, column names were aligned across both periods.

Redundant columns — such as *bikeid*, *gender*, and *birthyear* from the 2019 dataset, and *rideable_type* from the 2020 dataset — were removed to maintain consistency and reduce noise. Additionally, all station IDs were converted to character format to avoid issues with numerical interpretation or loss of leading zeros.

## 2. Ride Duration Calculation

Ride duration was standardized across datasets by creating the variable *duration_mins*, expressed in minutes.

- For 2019, durations were computed from the original *tripduration* field.
- For 2020, durations were calculated directly from the timestamp difference (*end_time* – *start_time*).

Invalid values were handled in two passes: extreme durations (<1 min or >1,440 min) were set to NA, followed by negative and zero durations. This ensures that aggregate metrics are not biased by system errors or incomplete records.

## 3. Completion of Station Coordinates

Latitude and longitude columns were added to the 2019 dataset using station information from 2020. Unique station reference tables were created and merged via left joins to fill start and end coordinates. After the merge, 40 start-station coordinates and 60 end-station coordinates remained missing, corresponding to stations that no longer existed in 2020. No imputation was performed, preserving data integrity for analysis.

## 4. Treatment of HQ QR Station

The station “HQ QR” (ID 675) appears across the dataset but does not correspond to a physical docking location. This station was handled during the visualization stage in Tableau, where it was excluded from geospatial maps to prevent distortions. Trips using HQ QR were preserved in the dataset for completeness.

## 5. Feature Engineering

To enable temporal, behavioral, and usage-pattern analyses, new variables were created:

- *day_of_week*
- *month*
- *hour*
- *part_of_day*

These allow for the exploration of daily, weekly, and hourly usage patterns.

## 6. Data Integrity Verification

A duplicate check was performed on both datasets using trip identifiers before merging: no duplicate trip IDs were found. This confirms the uniqueness of trip records and ensures a clean consolidation process.

## 7. User Type Standardization

The *user_type* field contained different labels across datasets. To produce consistent classifications for analysis, values were recoded as follows:

- Subscriber / member → Member
- Customer / casual → Casual

This ensures alignment with Cyclistic's two primary user segments: casual riders and annual members.

## 8. Dataset Consolidation

The 2019 and 2020 datasets were merged using a row bind operation, producing a single dataset (all_trips). This unified dataset allowed subsequent cleaning, feature engineering, and analysis to be applied consistently across both years.

## Final Dataset Summary

After all transformations, the consolidated dataset contains **791,956 trips and 19 variables**, including trip identifiers, timestamps, start and end stations with coordinates, ride duration in minutes, user type, and temporal variables (day of week, month, hour, year and part of day). This dataset is now clean, reliable, and ready for user behavior analysis.

```
Rows: 791,956
Columns: 19
$ trip_id            <chr>
$ start_time         <dttm>
$ end_time           <dttm>
$ start_station_id   <chr>
$ start_station_name <chr>
$ end_station_id     <chr>
$ end_station_name   <chr>
$ user_type          <chr>
$ start_lat          <dbl>
$ start_lng          <dbl>
$ end_lat            <dbl>
$ end_lng            <dbl>
$ duration_mins      <dbl>
$ year               <dbl>
$ day_of_week        <ord>
$ month              <ord>
$ hour               <int>
$ part_of_day        <chr>
$ station_name       <chr>
```

*Figure 1: Dataset Structure (glimpse output)*

### Additional Contextual Data

A complementary dataset of major Chicago tourist attractions was incorporated in Tableau, including geographic coordinates collected from Wikipedia, LatLong.net, Distancesto.com, and CoordinatesFinder. This allows analysis of potential overlaps between casual riders' trips and popular tourist destinations, supporting insights for tourism-driven marketing strategies.

---

## Analyze Phase

With the data prepared and transformed, the next step was to perform a comprehensive analysis to uncover behavioral differences between casual riders and annual members. The goal of this phase is not only to calculate descriptive statistics but also to interpret them in a way that directly addresses the business task: understanding how these two groups use Cyclistic differently and identifying factors that may motivate casual riders to convert to annual memberships.

To achieve this, the analysis was structured into six main dimensions: overall user distribution, weekly usage patterns, daily usage trends, ride duration, temporal evolution, and geographic patterns.

## 1. Overall User Distribution

The first step is to understand the **composition of Cyclistic's riders**. A pie chart was created to show the total number of rides completed by casual and annual members.

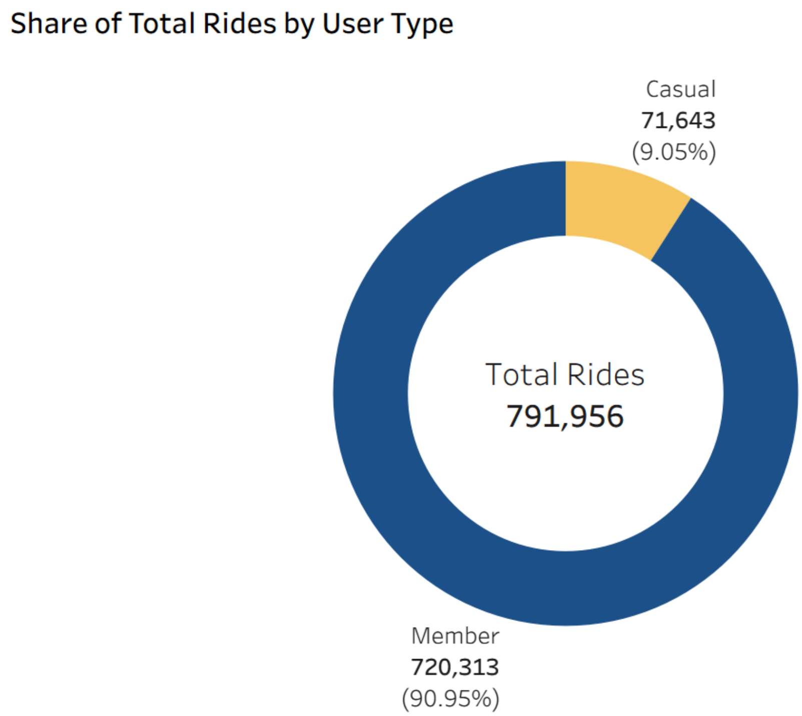


*Figure 2: Composition of Cyclistic Riders*

- **Insights:**
    - Annual members account for 90.05% of all rides (720,313), confirming their role as Cyclistic's most profitable and loyal segment.
    - Casual riders, however, represent 9.05% (71,643), highlighting a meaningful opportunity for membership conversion campaigns.

## 2. Weekly Usage Patterns

To explore differences in **usage across the week**, bar charts were created comparing casual and member riders by day of the week.

- Absolute volume of rides shows the raw number of trips per weekday, and percentage distribution highlights the share of trips across weekdays, making it easier to see relative differences between groups.

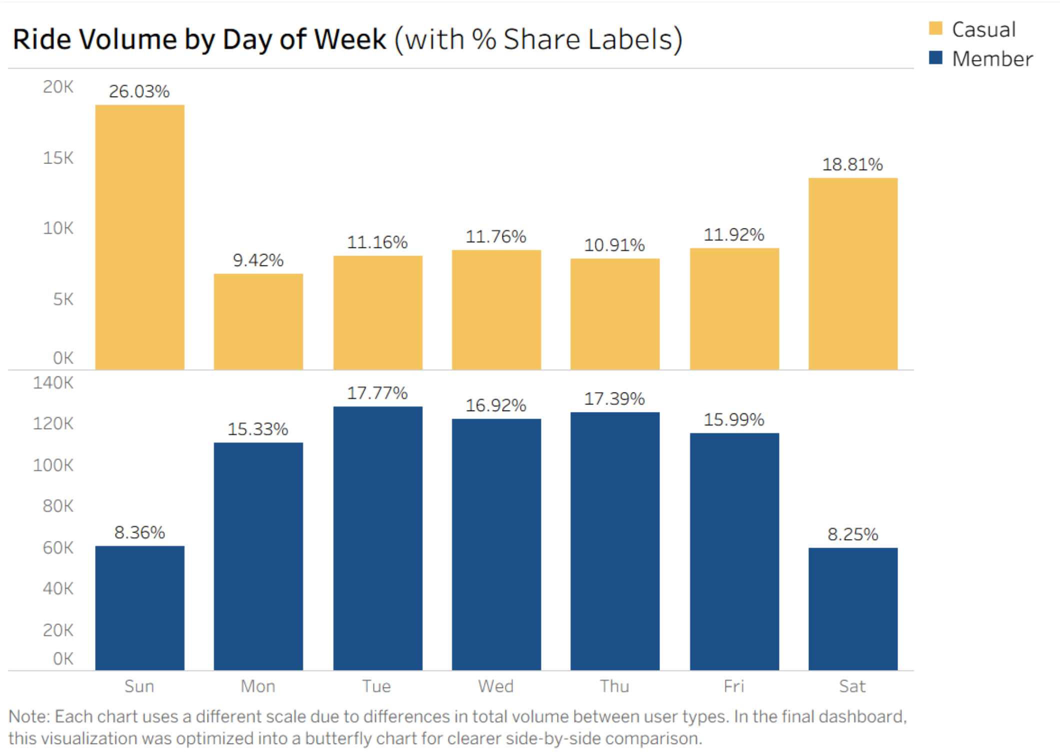


*Figure 3: Usage Across the Week*

- **Insights:**
  - Casual riders show a clear weekend-oriented pattern, with Sunday (26%) and Saturday (19%) concentrating almost half of their total rides. This behavior highlights the leisure or recreational nature of casual users.
  - Members, on the other hand, demonstrate a stable weekday pattern, peaking between Tuesday and Thursday (17%-18%) and dropping on weekends to around 8%. This consistency aligns with their likely use for daily commuting.

## 3. Daily Usage Trends (Hourly Patterns)

The next step is to analyze the **time of day** when rides occur. A line chart and a heatmap were used to capture hourly usage trends.

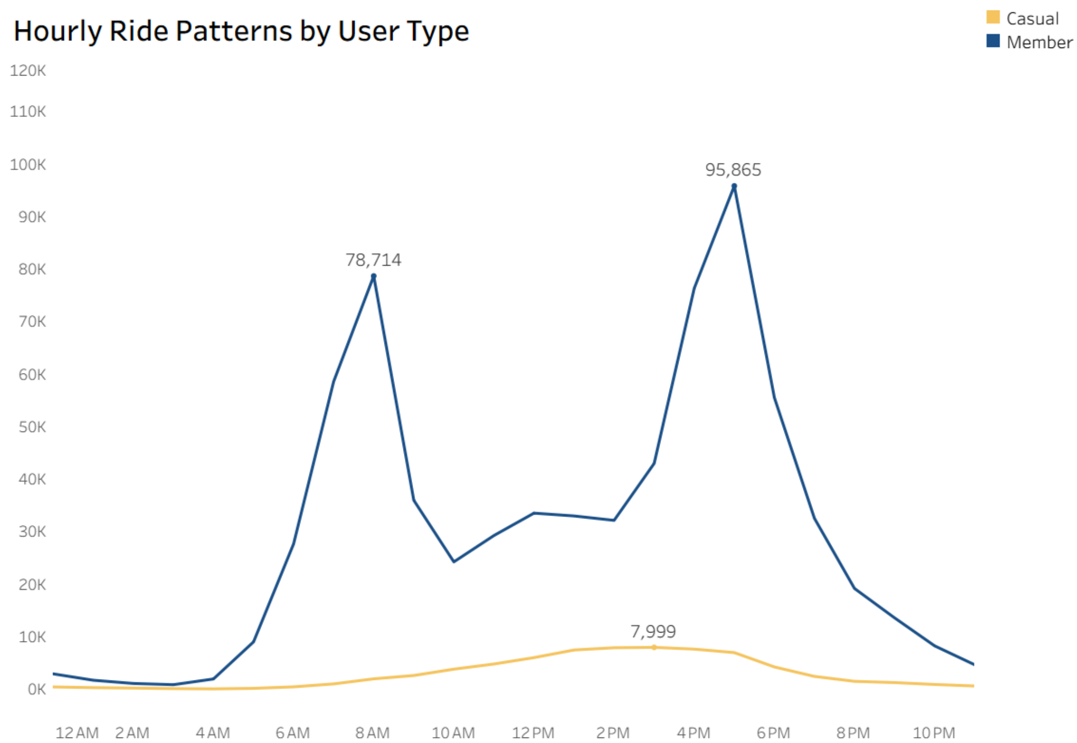


*Figure 4.1: Usage Across the Day*

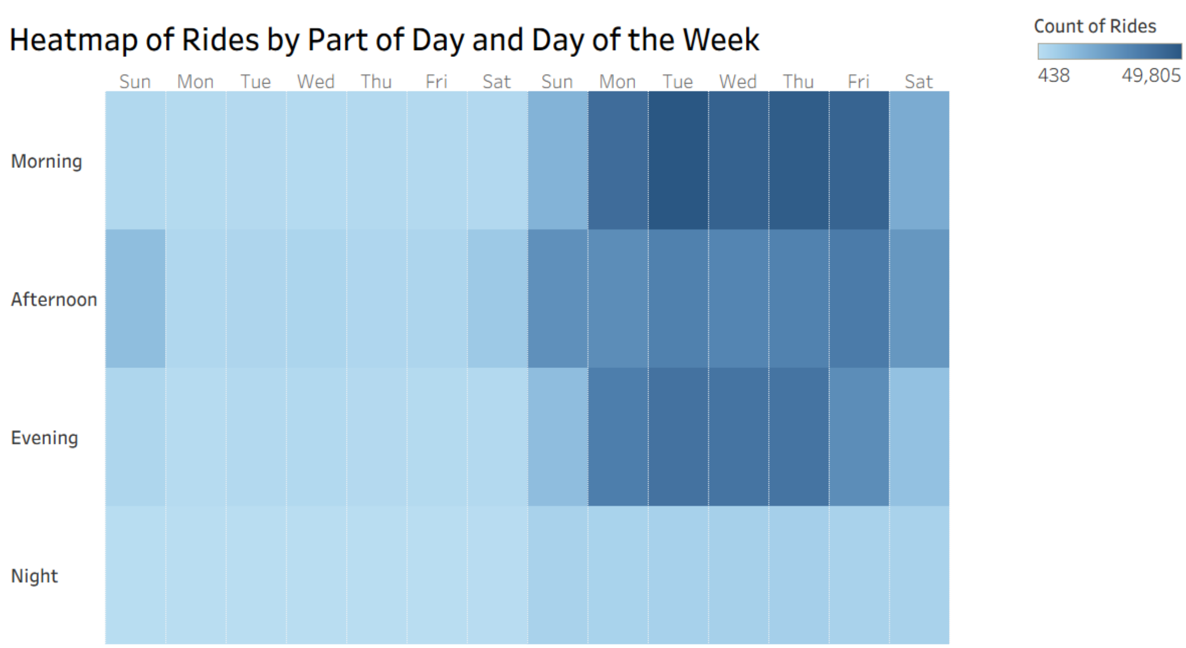


Figure 4.2: Usage by Time of Day and Weekday

- **Insights:**
  - Members display clear peaks during weekday mornings (7–9 AM) and late afternoons (4–6 PM), reflecting structured commuting behavior. Casual riders, in contrast, ride more evenly throughout the day, with moderate increases around midday and early evening, emphasizing leisure-oriented usage.
  - Peak ride volumes differ markedly: members dominate peak periods (95,865 rides), whereas casuals maintain moderate levels (7,999), reinforcing their flexible, non-routine riding habits.
  - The heatmap confirms these temporal trends, showing consistent weekday patterns for members and recreational weekend/afternoon peaks for casual riders, suggesting opportunities for targeted weekend campaigns.

## 4. Ride Duration

To understand **how long users ride**, two visualizations were created:

- A line chart comparing the average ride duration across weekdays for casual and member riders.
- A bar chart displaying the distribution of rides by duration categories for each group.

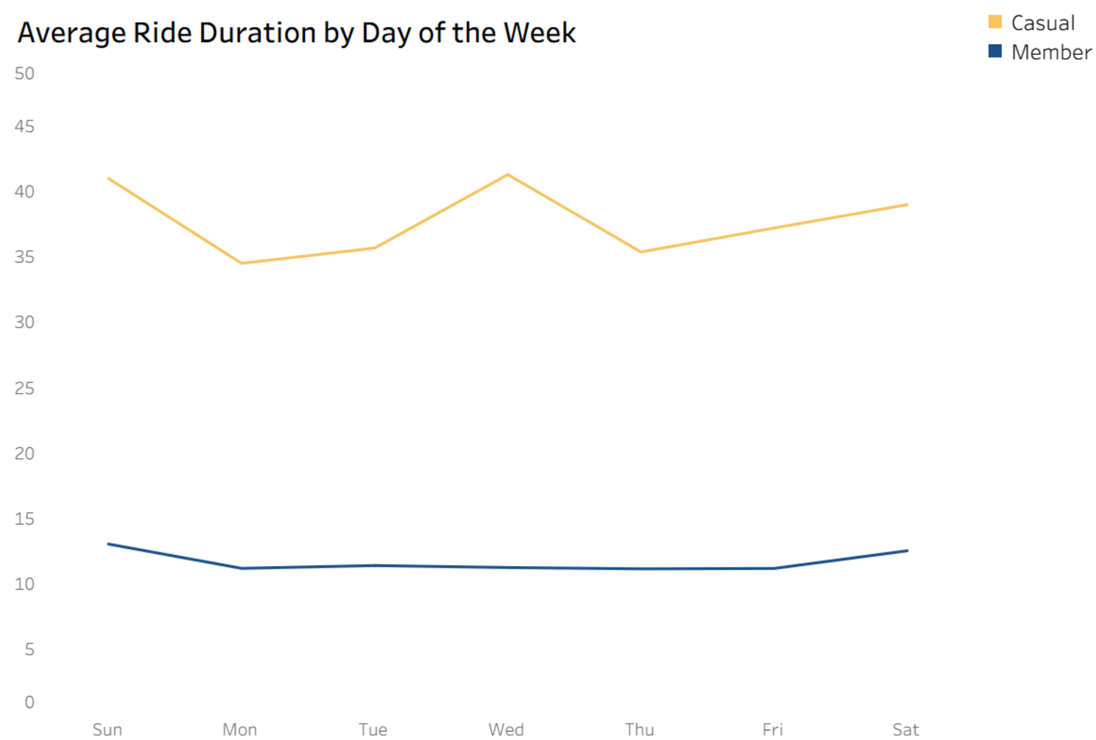


*Figure 5.1: Average Usage*

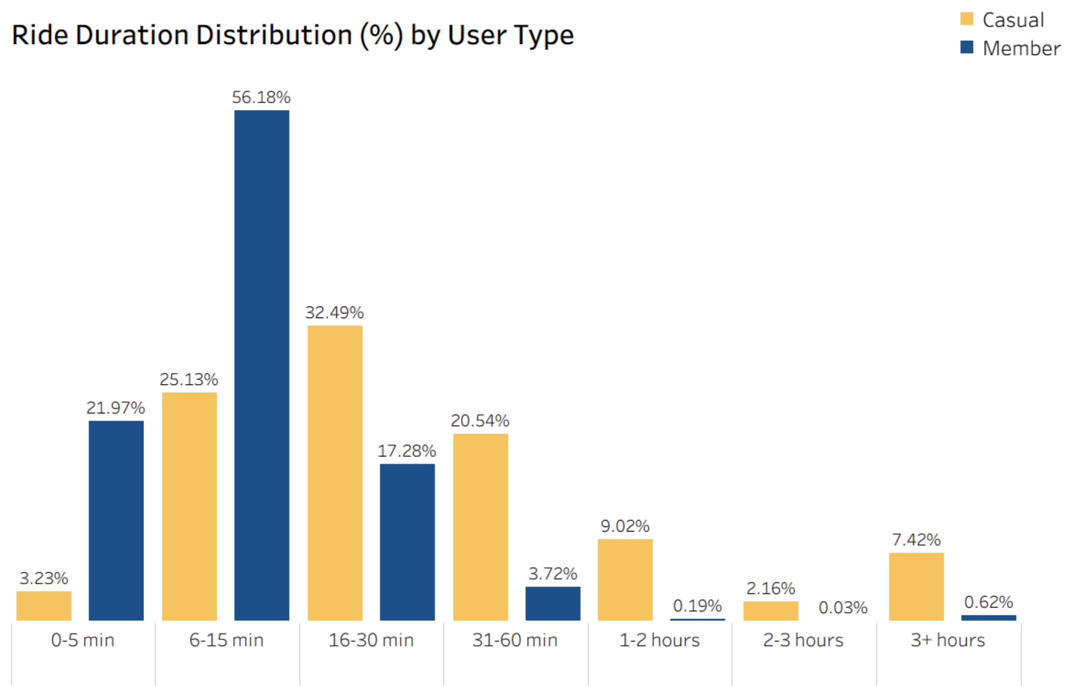


Figure 5.2: Usage Distribution

- **Insights:**
  - Casual riders take significantly longer trips than members across all days, averaging around 36–41 minutes versus 11–13 minutes for members. A midweek peak among casuals (Wednesday) may reflect favorable conditions or midweek leisure activity, but overall patterns confirm leisure-oriented use rather than commuting.
  - Duration distribution further supports this behavior: about 72% of casual trips last more than 15 minutes, compared to only 22% among members. Members, in turn, concentrate over three quarters of their rides below 15 minutes, confirming their preference for short, routine trips.
  - Weekend trends emphasize the contrast in riding purpose — casual riders' trip duration increases slightly on Saturdays and Sundays (around 39–41 minutes), while members remain stable near 12 minutes, suggesting stronger recreational use among casuals during weekends.

## 5. Temporal Evolution (Monthly Performance)

Since the dataset includes the first quarter of 2019 and 2020, a **line chart** was used to compare monthly performance by user type.

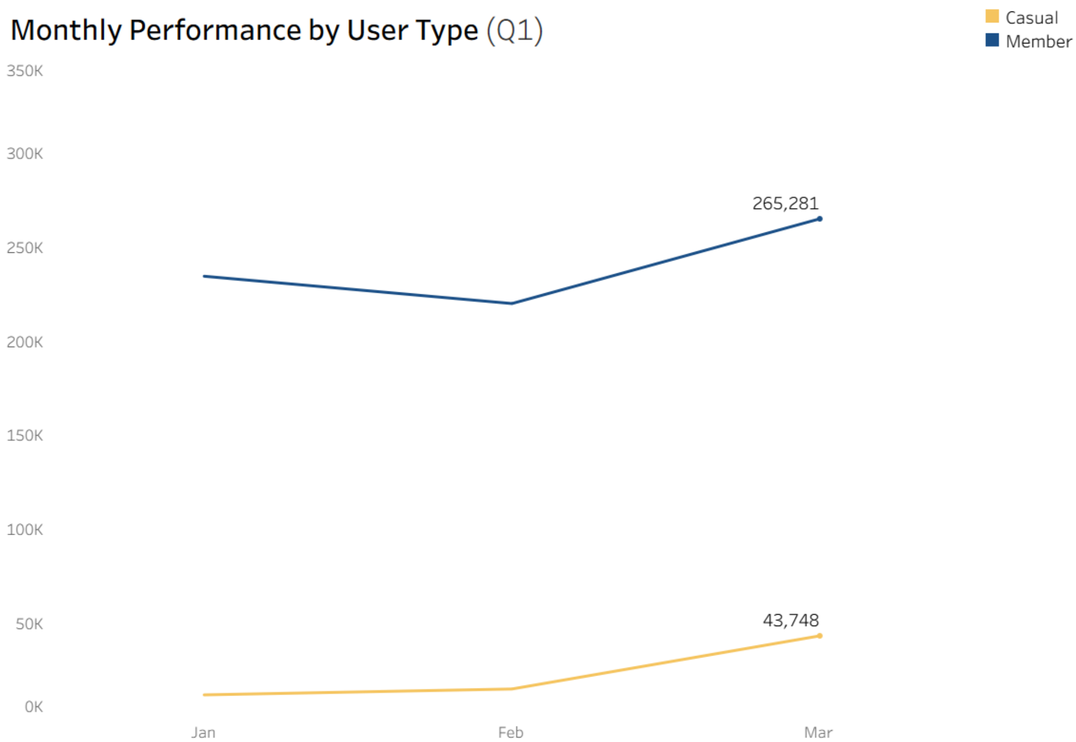


*Figure 6: Monthly Performance*

- **Insights:**
  - Both user groups show growth from January to March, reflecting the start of the biking season as weather improves.
  - Members maintain a consistently higher ride volume, but casual riders' trips grow almost threefold (from 12K to 44K), signaling strong seasonal engagement.
  - This sharp rise among casuals highlights an opportunity for targeted early-season campaigns, converting new or occasional riders as usage increases.

## 6. Geographic Patterns and Tourist Influence

Finally, geographic maps were generated to show the **spatial distribution of rides**. Separate maps display start and end stations for casual and member riders. An additional map with Chicago's main tourist attractions was added for context.

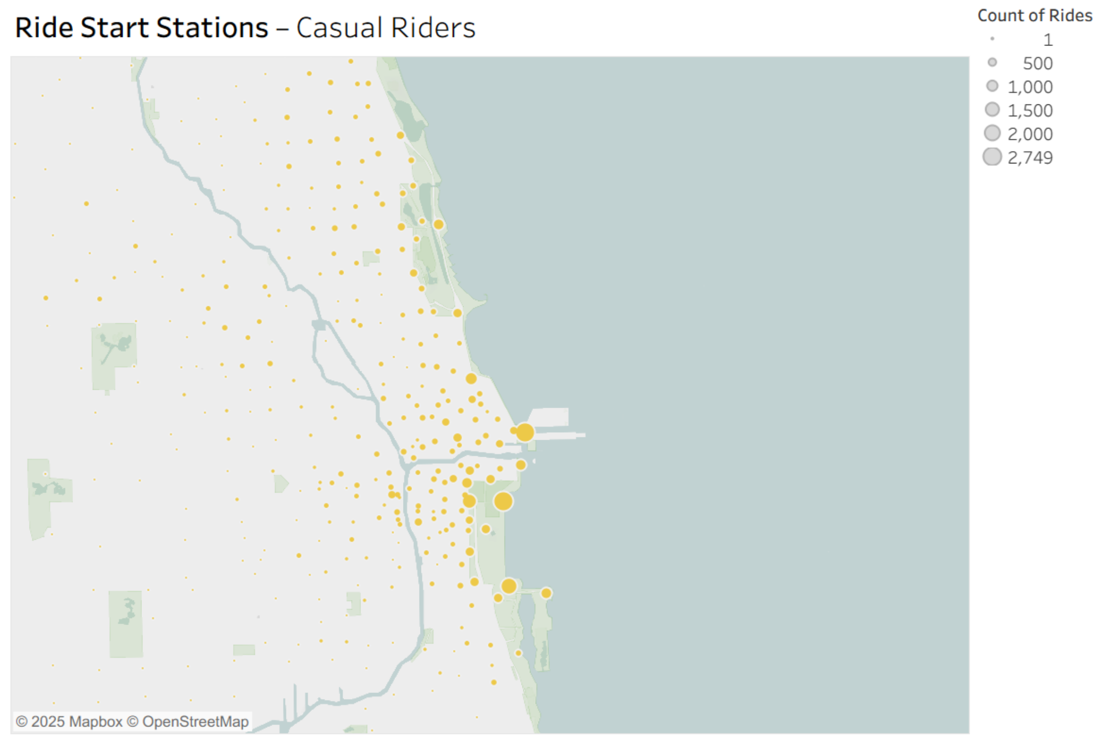


*Figure 7.1: Casual Riders - Start Stations*

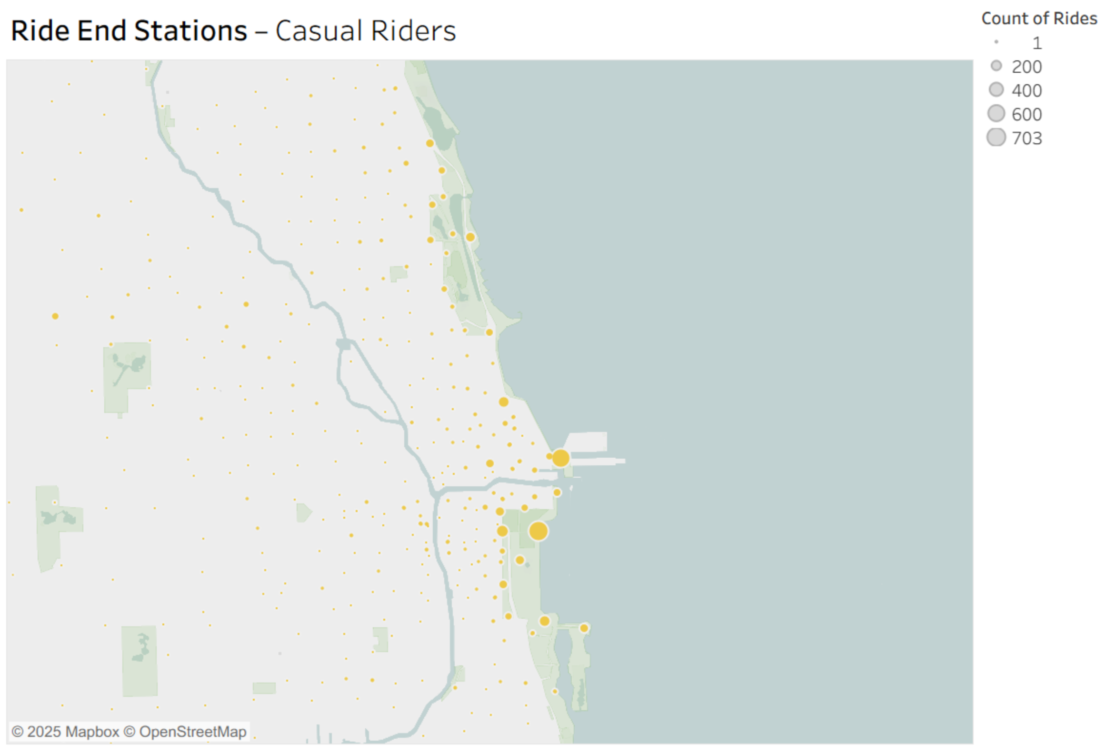


*Figure 7.2: Casual Riders - End Stations*

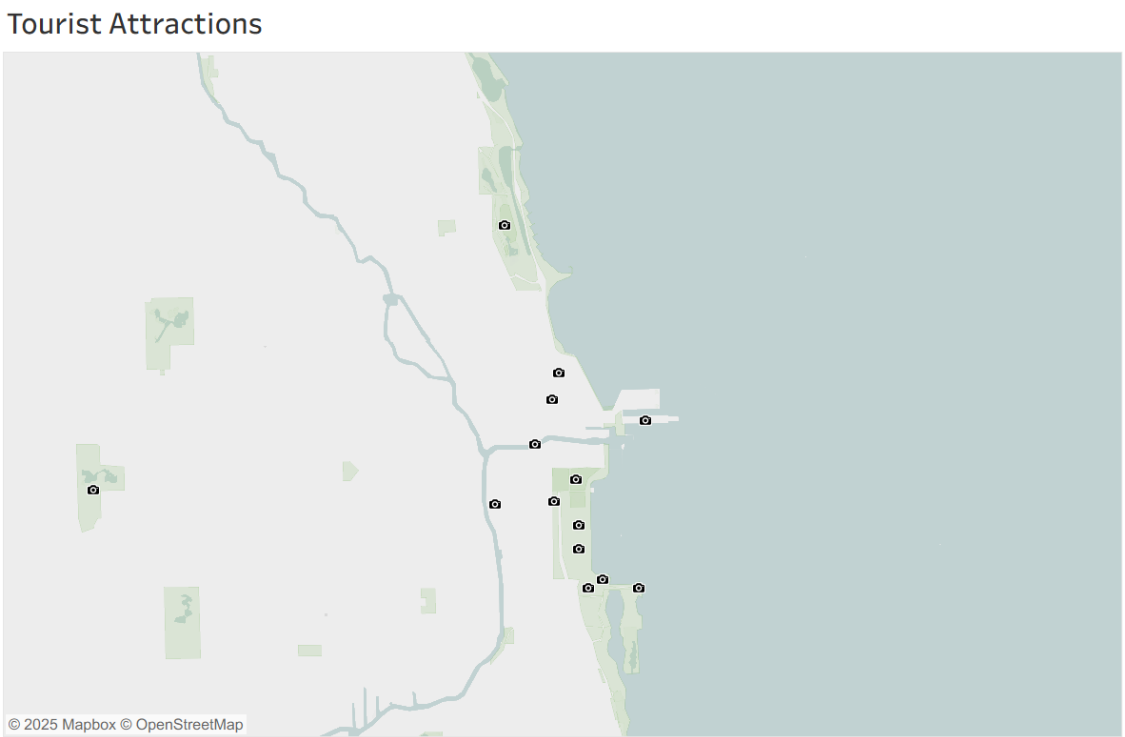


*Figure 7.3: Main Touristic Attractions in Chicago*

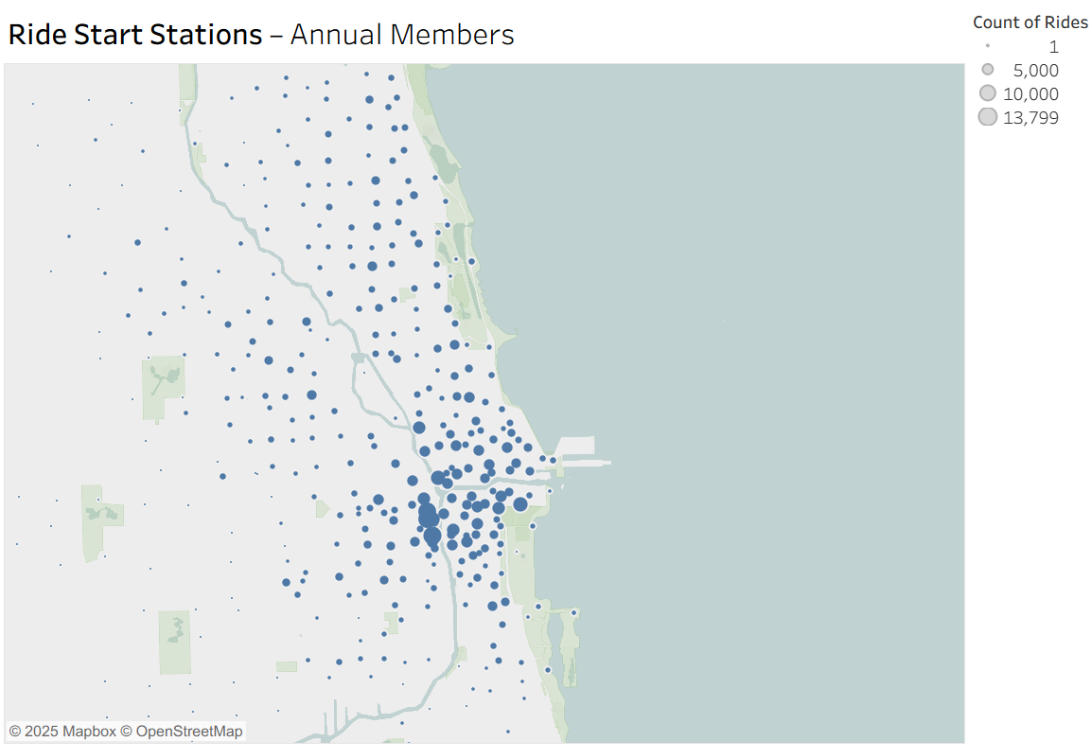


*Figure 7.4: Members - Start Stations*

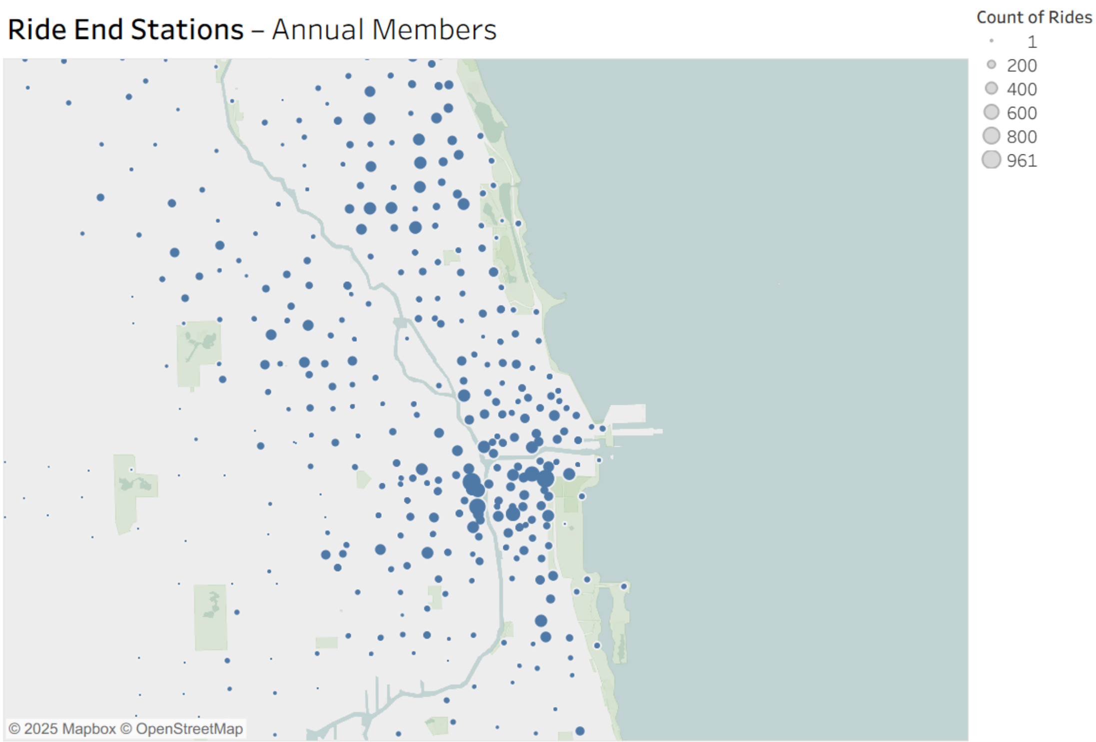


*Figure 7.5: Members - End Stations*

- **Insights:**
  - Annual members show a more widespread distribution, aligned with residential and commuting hubs, reflecting routine travel patterns. Casual riders tend to concentrate around popular recreational and touristic areas, highlighting leisure-oriented trips.
  - These spatial differences reinforce distinct usage behaviors — members depend on the service for consistent daily mobility, while casual users engage more around leisure and seasonal demand, suggesting opportunities for location-based marketing and tailored engagement strategies.

## Summary of Key Insights

The Analyze Phase confirms clear behavioral differences between Cyclistic's two customer segments:

- **Annual members**: Shorter, more frequent rides; weekday and rush-hour usage; stable performance across the first quarter; concentration in residential and business areas.

- **Casual riders**: Longer rides; peak activity on weekends and in March, during early spring; frequent use of stations near tourist attractions.

These findings directly answer the first guiding question: members and casual riders use Cyclistic bikes in fundamentally different ways.

They also begin to address the second guiding question: the main motivations for casual riders to convert may include cost savings on long rides, convenience for commuting, and broader usability beyond leisure.

The analysis provides a strong foundation for the next phase, where marketing strategies can be designed to leverage these insights, particularly through digital media campaigns tailored to casual riders' behavior.

---

## Share Phase

The Share Phase focused on communicating insights clearly and effectively to Cyclistic's stakeholders. To achieve this, the findings from the Analyze Phase were consolidated into an interactive **Tableau dashboard**, designed to summarize and visualize key differences between casual riders and annual members.

The dashboard integrates multiple perspectives — including overall user distribution, temporal patterns, ride duration, and spatial behavior — allowing decision-makers to explore user dynamics and identify actionable opportunities for marketing strategies aimed at converting casual riders into members.

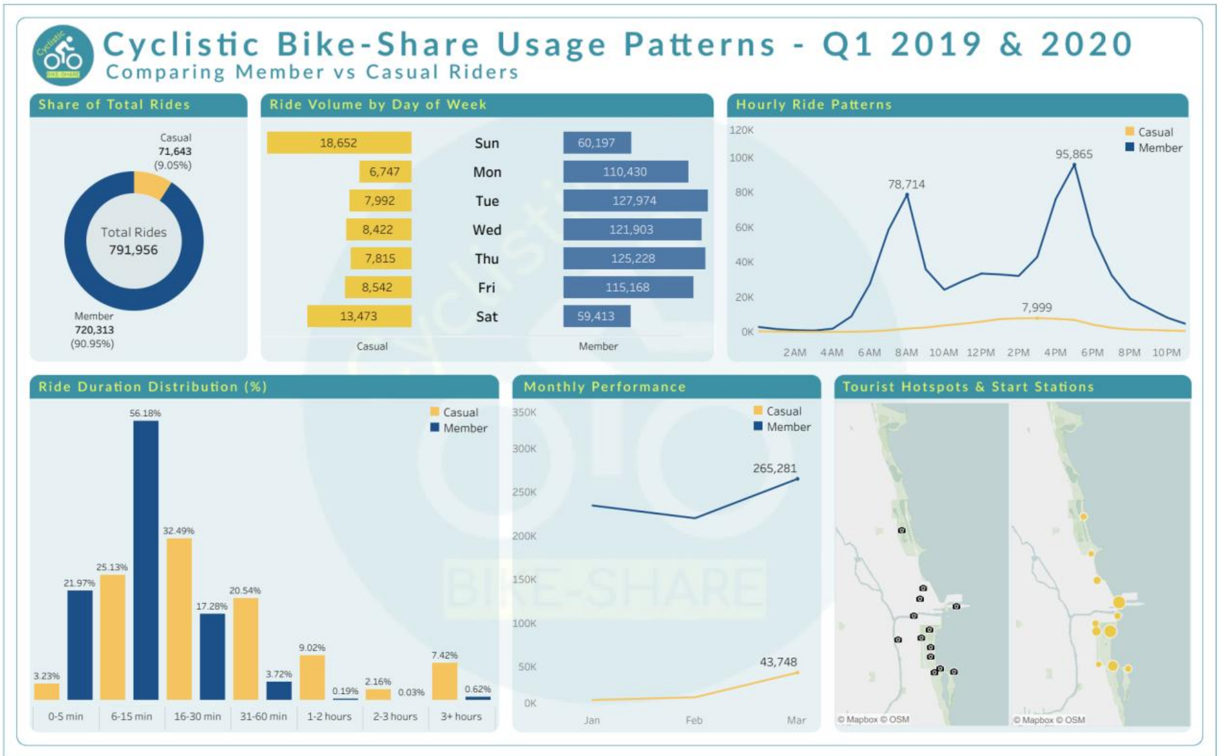


*Figure 8: Final Dashboard*

---

# Act Phase

Cyclistic's riding data reveals clear behavioral differences between annual members and casual riders, highlighting distinct usage patterns, motivations, and peak demand periods. The following top three recommendations translate these insights into actionable strategies designed to increase casual-to-member conversions while reinforcing Cyclistic's position as a reliable and accessible mobility option in Chicago.

## Foundational Requirement: First-Party Data Collection

Before implementing any marketing or membership strategy, Cyclistic must introduce a basic first-party data system (e.g., email opt-in, app login, or QR/SMS ride confirmation). This lightweight identification is essential for tracking returning riders, measuring campaign performance, and enabling personalized offers. It serves as a structural prerequisite for the recommendations that follow.

## Top Three Strategic Recommendations

### 1. Membership Strategy Optimization

A restructured membership portfolio can guide casual riders toward long-term commitment by offering options that reflect their observed riding behaviors while positioning the annual membership as the highest-value alternative.

- **Seasonal (6-Month) Summer Membership:** Data shows a sharp rise in ridership toward the end of the first quarter. While the current dataset cannot confirm seasonality, this uptick suggests a timely opportunity to test a shorter, lower-commitment summer plan designed for riders who may be more active as weather conditions improve.
- **Weekend Membership:** Casual riders demonstrate concentrated activity on Saturdays and Sundays. A weekend-only plan directly targets this behavior, offering a flexible, low-commitment option that meets their habits while gently nudging them toward full membership.
- **Recreation Bundle (e.g., 5 rides/month):** For casual riders who take short, recurring trips, a light-touch monthly bundle creates a smooth progression toward membership without requiring major behavioral change.

All three plans can be intentionally positioned through decoy pricing, making the annual membership appear the most cost-effective when presented alongside the shorter options — increasing perceived value and driving conversion. Coordinated digital campaigns across social media, email, and in-app messaging will further amplify adoption by highlighting the benefits of each plan during high-potential periods, leveraging observed riding behaviors to optimize timing and messaging.

### 2. Events and On-Ground Engagement

Strategic in-person initiatives can transform occasional recreational use into a more frequent and deliberate riding habit — a key step toward increasing membership intent among casual riders.

- **Recreational Group Rides:** Curated rides along lakefront trails, weekend sunrise tours, or neighborhood “discovery rides” turn a casual activity into a shared experience, strengthening Cyclistic’s brand presence in the city’s most popular riding zones.
- **Fitness-Oriented Events:** Multi-day challenges, mileage goals, and community fitness rides appeal to health-focused users who already ride for exercise on

weekends. By tying these events to achievement badges or limited-time rewards, Cyclistic introduces a motivating progression system similar to fitness apps.

- **Location-Based Pop-Ups:** Temporary pop-up stations at high-traffic recreational points — such as lakefront parks, Navy Pier, or major city events — reinforce Cyclistic's visibility where casual riders naturally gather, making the brand feel present, accessible, and community-driven.

Together, these initiatives help casual riders associate Cyclistic with active, social, and recurring experiences — moving them from occasional leisure use to habitual engagement that more naturally leads to membership consideration. By designing these experiences to include targeted offers and limited-time incentives, Cyclistic can further encourage casual riders to consider membership.

## 3. Strategic Partnerships to Expand Reach and Everyday Utility

Partnerships placed where casual riders naturally spend time can strengthen Cyclistic's visibility, reduce barriers to use, and position the service as a practical, everyday mobility option — not just a weekend recreation choice.

- **Hotels and Tourism Operators:** Collaborations with hotels, visitor centers, and tour providers place Cyclistic in front of riders at the exact moment they decide how to move around the city. Guest-focused incentives — such as ride credits, temporary access plans, or priority docks — can increase first-time use and, through sustained exposure, raise the likelihood of future membership consideration.
- **Local Mobility & Wellness Partners:** Partnering with fitness studios, parks, universities, and municipal programs reinforces Cyclistic as an eco-friendly transportation and recreation alternative. Exclusive perks tied to these partners (e.g., discounted first-month membership or event-based ride bundles) help convert already motivated riders into long-term members.
- **Business & Employer Engagement:** Encouraging local businesses to promote bike commuting through corporate memberships or employer-sponsored ride credits embeds Cyclistic into daily routines — a consistent behavior strongly associated with annual membership adoption.

Across all partnerships, coordinated promotional campaigns — through social media, in-app prompts, and location-based messaging — ensure that riders encounter Cyclistic at high-intent moments, increasing engagement and driving membership conversions.

---

# Conclusion

Cyclistic now has a clearer understanding of how annual members and casual riders differ — not only in when and where they ride, but in how these behaviors signal distinct motivations. By strengthening its first-party data foundation, Cyclistic can move from broad, city-wide marketing to targeted, evidence-based engagement that speaks directly to each rider segment.

The strategies outlined in this report — optimizing the membership portfolio, activating on-ground and community-based events, and forming strategic partnerships — create multiple pathways that nudge casual riders toward deeper engagement and, ultimately, annual membership. Together, they reinforce Cyclistic's positioning as a reliable, accessible, and purpose-driven mobility service for both everyday commuters and recreational riders.

This Capstone project demonstrates how data-driven decision-making can shape product strategy, marketing execution, and long-term rider loyalty. As Cyclistic continues to expand its data capabilities and gather ongoing feedback, it will be well-positioned to refine these initiatives, enhance rider experience, and sustain growth through informed, measurable, and user-centered actions.