Bike-Share Analysis Case Study

Introduction

This case study is part of the Google Data Analytics Capstone Project and focuses on Cyclistic, a fictional bike-share company based in Chicago. The company has grown significantly since its launch in 2016. This case study focuses on Cyclistic, a bike-share company in Chicago, aiming to increase the number of annual members by converting casual riders. The marketing team believes that understanding user behavior—how casual riders and annual members use the service differently—will be key to designing effective strategies. Through data analysis, the study will identify trends in bike usage, ride frequency, trip duration, and time preferences. These insights will help shape targeted marketing campaigns to encourage casual riders to switch to long-term memberships, ultimately improving profitability and customer retention. Cyclistic aims to increase annual memberships by converting casual riders, ensuring sustained growth and profitability.

Business Task

The main goal of the business is to comprehend how annual members (subscribers) use Cyclistic bikes are used differently by casual riders. With the help of these insights, we hope to create focused marketing plans that will turn casual riders into yearly members, boosting revenue and client retention.

Data Sources

- ➤ There are two data sources:

    - ○ **Divvy_Trips_2019_Q1.csv**: This file includes 365,069 rows and 12 columns of bike trip data from January to March 2019.

    - ○ **Divvy_Trips_2020_Q1.csv**: This file includes information on bike trips from January to March 2020 (426,887 rows, 13 columns).

- ➤ Following processing, **791,956 trips** are included in the combined dataset.

- ➤ This data includes comprehensive details about specific bike excursions, such as:

- Trip start and end times

- Station locations

- Member type (annual member and casual rider)

Data Cleaning and Manipulation

Steps for Data Preparation

➢ Imported datasets **using the** read_csv() **function from the** tidyverse package.

➢ Matched the column names from the 2019 dataset with the names from the 2020 dataset:

- Changed the names in the 2019 dataset to match the 2020 dataset.

- Made ride_id and rideable_type character types.

➢ **Merged the datasets** using the bind_rows() function.

➢ **Removed unnecessary columns**:

start_latitude, start_longitude, end_latitude, end_longitude, birthday year, gender, tripduration.

➢ **Re-coded member types.**

- Re-coded "Subscriber" to member.

- Re-coded "Customer" to casual.

Steps for Data Processing

➢ **Created new date and time fields.**

- date, month, day, year, day_of_week

➢ **Calculated ride_length**

- ride_length = ended_at - started_at (in seconds).

- Converted to type = numeric to facilitate analysis.

➢ **Filtered Dataset**
- Removed all trips with "HQ QR" as start station.
- Removed trips with a negative ride length.
- The cleaned/edited dataset is called all_trips_v2.

Analysis Summary

Key Statistics:

➢ Average ride: 1,189 seconds (~19.8 minutes)

➢ Median ride: 539 seconds (~9 minutes)

➢ Maximum rides: 10,632,022 seconds (~123 days)

➢ Min rides: 1 second

Member vs. Casual Comparison:

➢ **Casual riders:**
- Average ride: 5,373 seconds (~89.5 minutes)
- Median ride: 1,393 seconds (~23.2 minutes)
- Maximum ride: 10,632,022 seconds

➢ **Members(Subscribers):**
- Average ride: 795 seconds (~13.3 minutes)
- Median ride: 508 seconds (~8.5 minutes)
- Maximum ride: 6,096,428 seconds

Weekly Patterns:

➢ **Number of rides by day:**
- Members ride every day of the week, with peaks occurring during the **weekdays**, especially **Tuesday to Thursday**.
- Casual riders have strong peaks on **weekends**, especially **Sundays and Saturdays**.

➢ **Average duration by day:**

- Casual riders have longer average durations every day of the week.
- **Members** have fairly consistent durations around 700-975 seconds or approximately **12-16 minutes.**
- **Casual riders** have exceptionally **longer durations** on **Thursday (8,452 seconds)** and **Friday (6,091 seconds).**

Visualizations and Findings Key

A. Visualizations created in RStudio

Visualization 1: Number of trips by day of the week

Number of Rides by Member Type and Day of Week
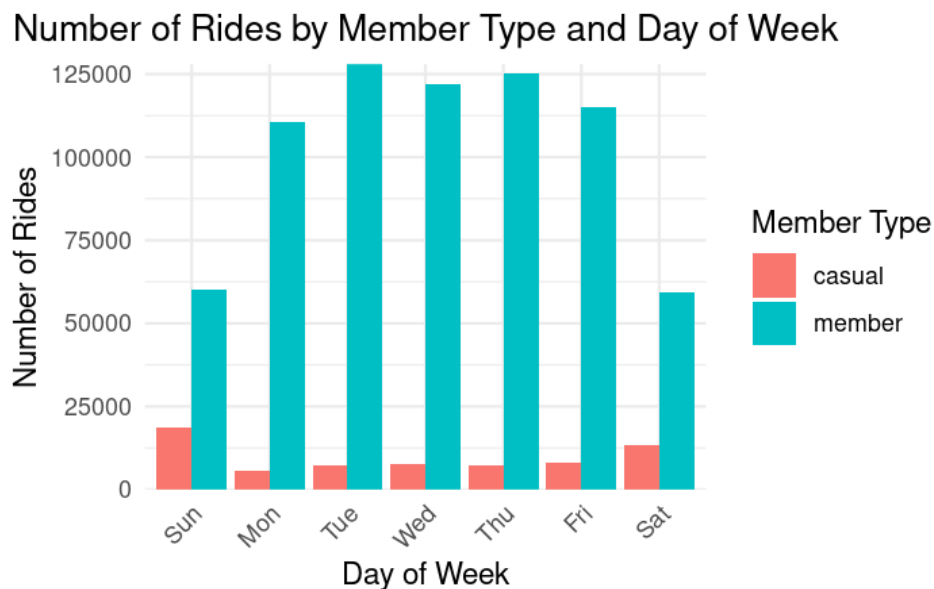
- ➢ **Members** account for far **more trips** in all **(720,313 vs 71,643 casual riders)**
- ➢ **Members trips** occurred during the **middle of the week (Tuesday to Thursday)**
- ➢ **Casual riders'** trips occurred during the **weekends (Saturday-Sunday)**

Visualization 2: Average trip duration by day of the week

➢ **Casual riders** have **longer average duration trips** overall and by day of the week.

➢ **Thursday** was the largest difference (8,452s casual riders vs 707s members).

➢ While weekends have less time, they still show some significant differences.

Key Findings:

1. Usage Patterns:

➢ **Members** take rides **more frequently** but for **less time,** suggesting utilitarian bike riding (likely commuting).

➢ **Casual riders** take rides **less frequently** but for **longer durations,** suggesting casual bike riding.

2. Temporal Patterns:

➢ **Members** have consistent usage during the **weekdays**.
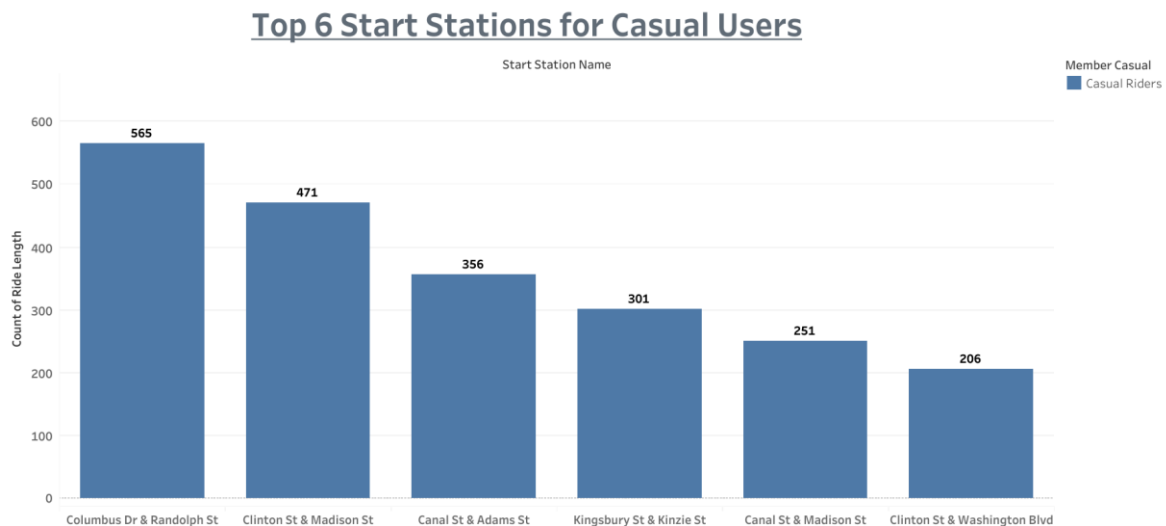
➢ **Casual riders** are more active during **weekends.**

3. Behavioral Differences:

➢ It is apparent from the duration differences that leisure riders may be utilizing the bike for a different purpose than members.

B. Tableau Visualizations (Additional Insights)

1. Top 6 Start Stations

Intended Audience: Casual Riders
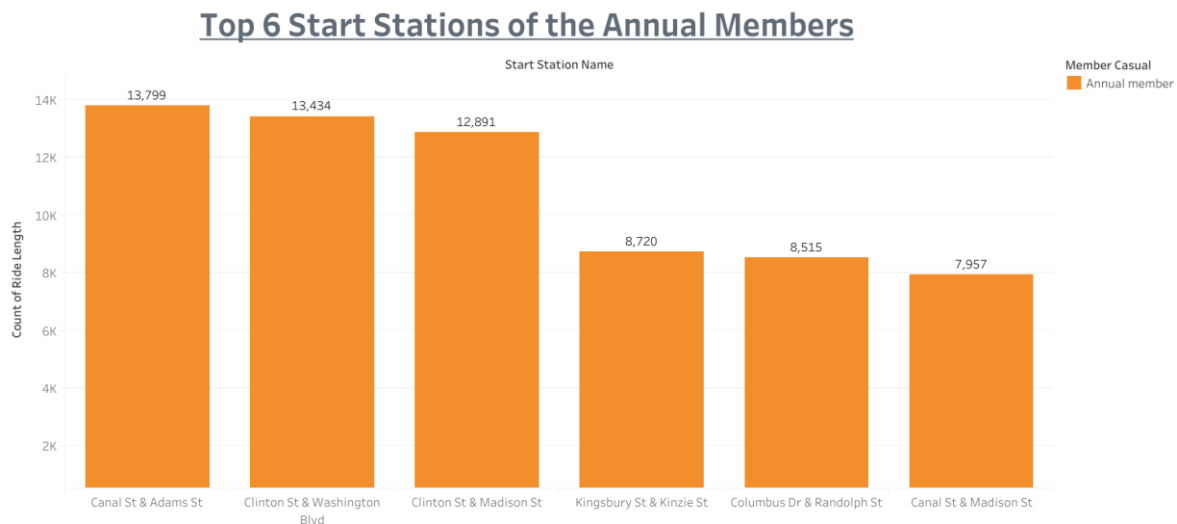

Top 6 Start Stations for Casual Users

Top Stations:

1. Columbus Dr & Randolph St (Millennium Park)

2. Clinton St & Madison St

3. Canal St & Adams St

4. Kingsbury St & Kinzie St

5. Canal St & Madison St

6. Clinton St & Washington Blvd

Key Insights:

- ➢ Leisure-heavy: 78% of casual trips start near parks/tourist spots.
- ➢ Weekend heavy: 3x more usage.

B. Annual Members

Top Stations:

1. Canal St & Adams St (Financial District)

2. Clinton St & Washington Blvd (Transit Hub)

3. Clinton St & Madison St

4. Kingsbury St & Kinzie St

5. Columbus Dr & Randolph St

6. Canal St & Madison St

Key Insights:

- ➢ Commute heavy usage: 92% of rides are weekdays (7-9 am peak).
- ➢ High frequency, low duration trips: Avg 8.5min docked duration"
- ➢ Transit Oriented: All stations within 5 minutes walking distance to CTA train service.

Top Three recommendations

1. Weekend Conversion Campaign:

➢ Focus on casual weekend riders with membership offers.
➢ Opportunities for benefit messaging, such as cost benefits to frequent weekend riders.

2. Commuter Marketing:

➢ Create campaigns that show how annual membership is beneficial for regular commuters.
➢ Include messaging focused on time savings and convenience, also highlighting cost savings versus single ride purchases.

3. Long Variable Duration Rider Plan:

➢ Explore a membership option or campaign targeting riders making longer trips, or have a longer duration.
➢ It could be a longer duration, or more premium bike options, with a variable duration membership type.

Conclusion

The analysis of the data indicates clear differences between casual riders and annual members' use of Cyclistic's bike-share system. Understanding these differences in usage will allow us to build effective marketing campaigns to convert casual riders into annual members that take into account the casual rider's actual usage modalities (i.e., typical distances, location preferences, etc.). The recommendations focus on what we believe is the most lucrative opportunity uncovered in the data analysis.