Google Case Study I

A Capstone Project Given in the End of Google Data Analytics Certificate

Project Done By: Abdul Rauf Bhutto

Uploaded portfolio project on LinkedIn, GitHub and Kaggle

Contents

GOOGLE CASE STUDY I

I have recently completed the Google Data Analytics Certificate. I learned the six steps of data analysis process (Ask, Prepare, Process, Analyze, Share and Act), and applied technical tools (Google Sheet, SQL , R language and Tableau) in each stage to gather information from data.

Capstone project given at the end of the certificate lets students work on case study. In this project I will be working for the marketing team of Cyclistic, a fictionnal bike share-company in Chicago.

- **I have decided to do the entire capstones project with R language.**

ASK

The director of marketing believes the company's future success depends on maximizing the number of annual memberships.

This is the questions/business task that would guide the future of the marketing program:

- **To understand how annual members and casual riders use our Cyclistic bikes**

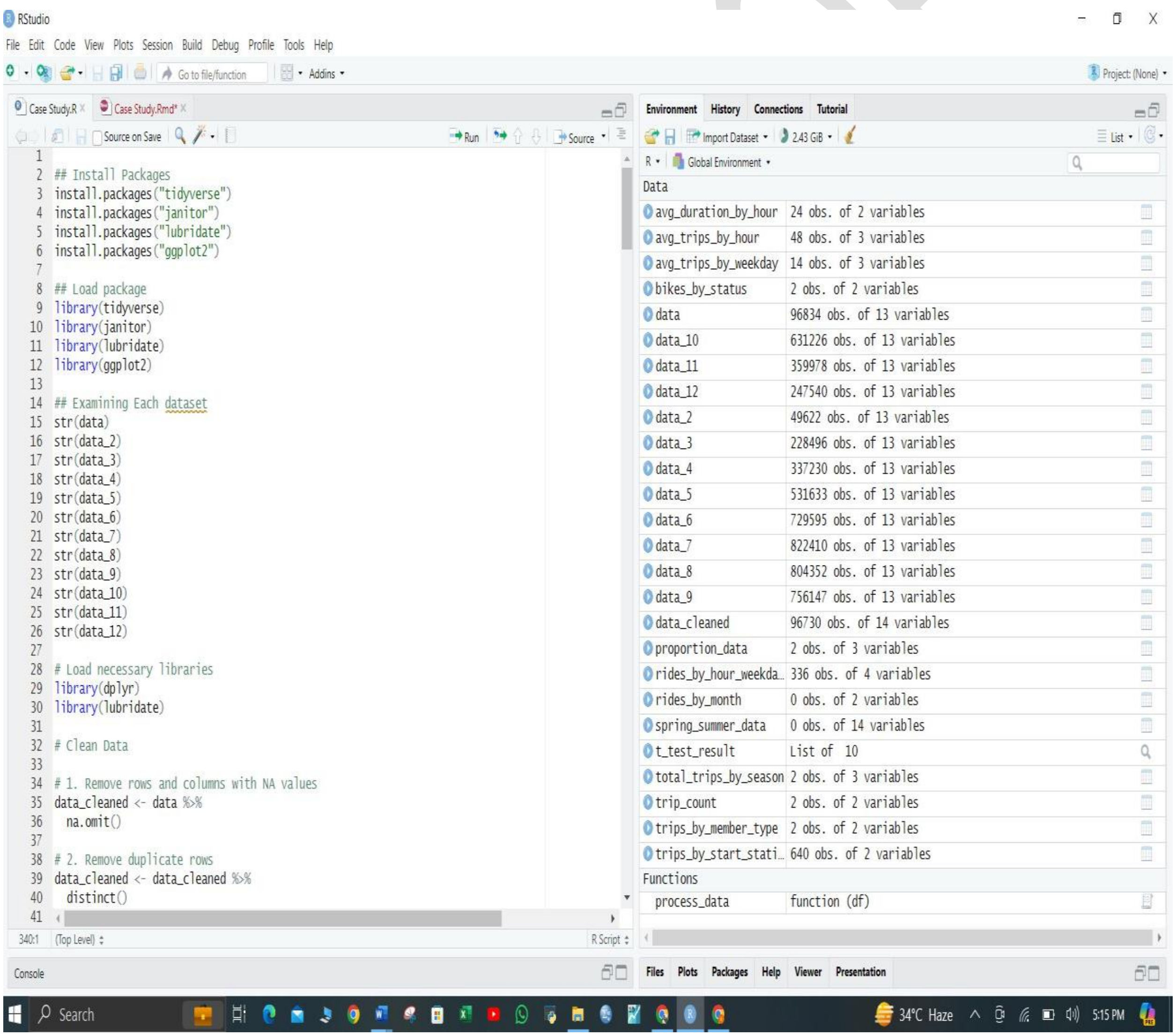
differently especially in Winter?

Objectives

The goal is to identify patterns in usage that can inform targeted marketing strategies, operational adjustments, and service improvements to enhance user experience and increase ridership.

PREPARE

We will be using Cyclistics historical trip data from http://https://divvy-tripdata.s3.amazonaws.com/index.html 2021 January till 2021 December. data has been made available by Motivate International Inc. under this license.



PROCESS

I would be using RStudio for this huge data.

Firstly, we would need to install & load the packages required for this process, which in this case will be: Tidyverse, Janitor & Lubridate.

```
## Install Packages
install.packages("tidyverse")
install.packages("janitor")
install.packages("lubridate")
install.packages("ggplot2")

## Load package
library(tidyverse)
library(janitor)
library(lubridate)
library(ggplot2)
```

Examnine Each Dataset

str(data)

str(data_2)

str(data_3)

str(data_4)

str(data_5)

str(data_6)

str(data_7)

str(data_8)

str(data_9)

str(data_10)

str(data_11)

str(data_12

```
> ## Examining Each dataset
> str(data)
'data.frame':   96834 obs. of  13 variables:
 $ ride_id           : chr  "E19E6F1B8D4C42ED" "DC88F20C2C55F27F" "EC45C94683FE3F27" "4FA453A75AE37
7DB" ...
 $ rideable_type     : chr  "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...
 $ started_at        : chr  "2021-01-23 16:14:19" "2021-01-27 18:43:08" "2021-01-21 22:35:54" "2021
-01-07 13:31:13" ...
 $ ended_at          : chr  "2021-01-23 16:24:44" "2021-01-27 18:47:12" "2021-01-21 22:37:14" "2021
-01-07 13:42:55" ...
 $ start_station_name: chr  "California Ave & Cortez St" "California Ave & Cortez St" "California A
ve & Cortez St" "California Ave & Cortez St" ...
 $ start_station_id  : chr  "17660" "17660" "17660" "17660" ...
 $ end_station_name  : chr  "" "" "" "" ...
 $ end_station_id    : chr  "" "" "" "" ...
 $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
 $ start_lng         : num  -87.7 -87.7 -87.7 -87.7 -87.7 ...
 $ end_lat           : num  41.9 41.9 41.9 41.9 41.9 ...
 $ end_lng           : num  -87.7 -87.7 -87.7 -87.7 -87.7 ...
 $ member_casual     : chr  "member" "member" "member" "member" ...
> str(data_2)
'data.frame':   49622 obs. of  13 variables:
 $ ride_id           : chr  "89E7AA6C29227EFF" "0FEFDE2603568365" "E6159D746B2DBB91" "B32D3199F1C2E
75B" ...
 $ rideable_type     : chr  "classic_bike" "classic_bike" "electric_bike" "classic_bike" ...
 $ started_at        : chr  "2021-02-12 16:14:56" "2021-02-14 17:52:38" "2021-02-09 19:10:18" "2021
-02-02 17:49:41" ...
 $ ended_at          : chr  "2021-02-12 16:21:43" "2021-02-14 18:12:09" "2021-02-09 19:19:10" "2021
-02-02 17:54:06" ...
 $ start_station_name: chr  "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Clark St & Lake
St" "Wood St & Chicago Ave" ...
 $ start_station_id  : chr  "525" "525" "KA1503000012" "637" ...
 $ end_station_name  : chr  "Sheridan Rd & Columbia Ave" "Bosworth Ave & Howard St" "State St & Ran
dolph St" "Honore St & Division St" ...
 $ end_station_id    : chr  "660" "16806" "TA1305000029" "TA1305000034" ...
 $ start_lat         : num  42 42 41.9 41.9 41.8 ...
 $ start_lng         : num  -87.7 -87.7 -87.6 -87.7 -87.6 ...
```

Clean Data

* Removed rows and columns with NA values

* Removed unnecessary columns (delete start_lat, start_lng, end_lat, end_lng)

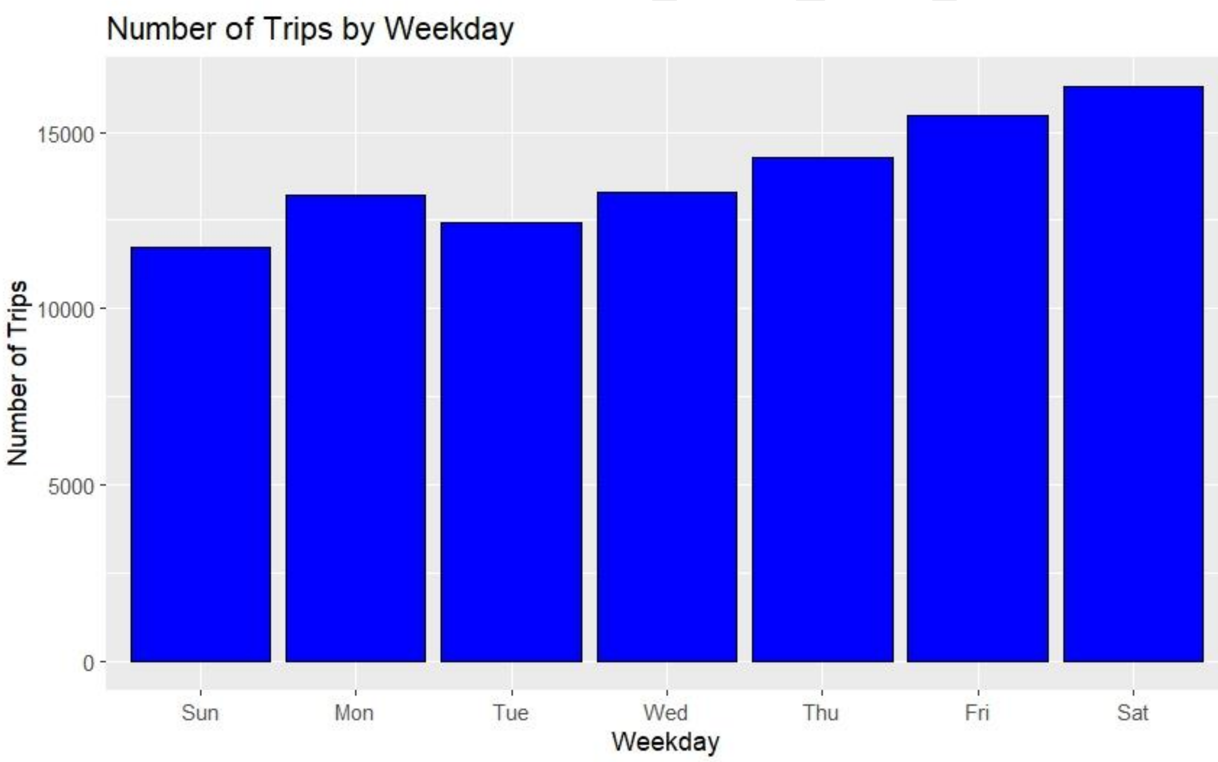
* Created new columns for analysis such as seasons

ANALYSE

I used ggplot2 for data analysis in R. Through analysis we will try to understand casual and members behavior and how their usage differ.

First I would like to know why people use Cyclistic bikes.

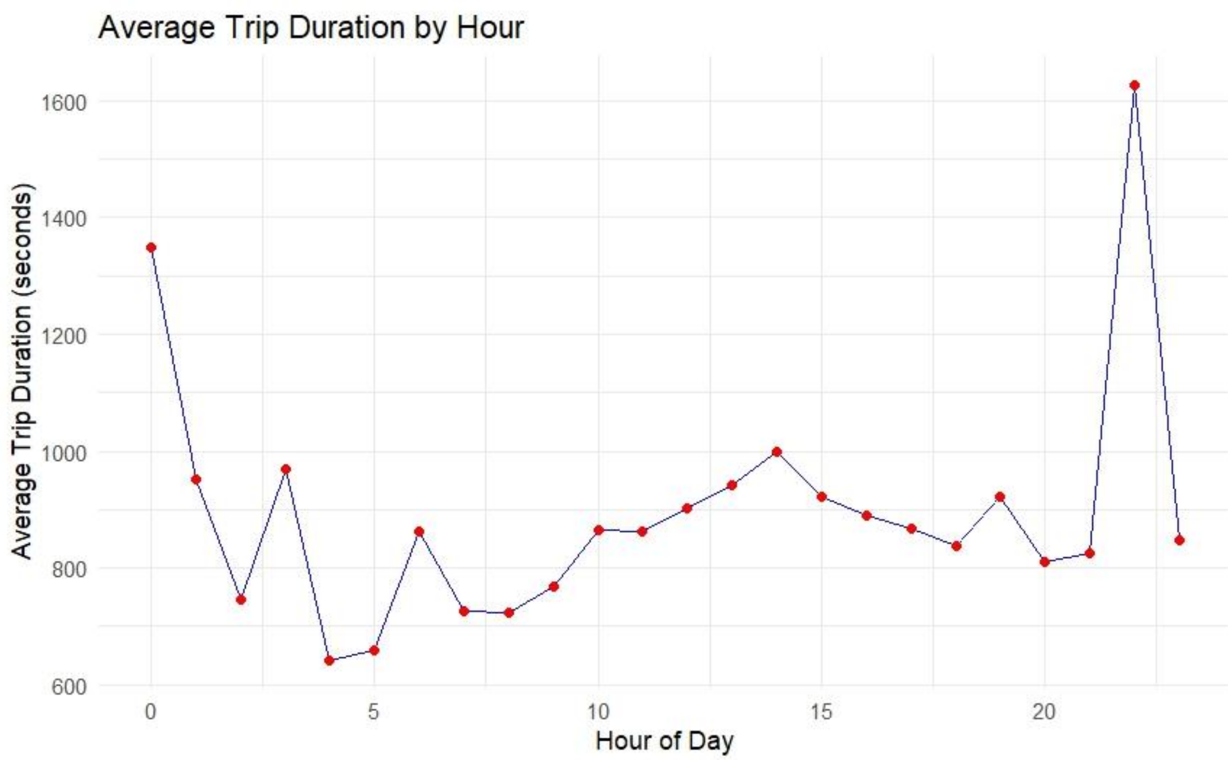
Number of Trips by Weekday



- Weekday ridership is consistent. This suggests that people are using Cyclistic bikes to commute to work.
- Ridership is lower on Sunday as it is holiday, so a few people are using it for leisure activities.

This graph will validate our previous claim that people are using Cyclistic bikes to commute to work.

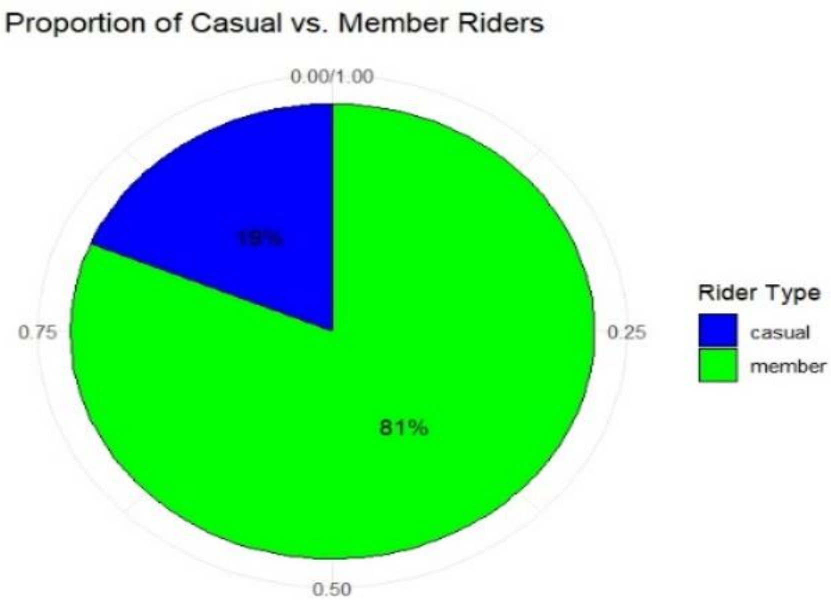
Average trip Duration by Hour



- The average trip duration is highest in the morning hours, between 7am and 10am. This suggests that people are using Cyclistic bikes for commuting to work or school.

- The average trip duration is lowest in the middle of the day, between 11am and 5pm. This could be because people are using bikes for shorter trips, such as errands or leisure rides.

- There is another bump in trip duration in the evening hours, between 5pm and 8pm. This could be people commuting home from work or school, or people using bikes for evening recreation.



Proportion of Casual vs Member Riders

- Casual Riders: Represent 19% of the total riders.
- Members: Represent 81% of the total riders.

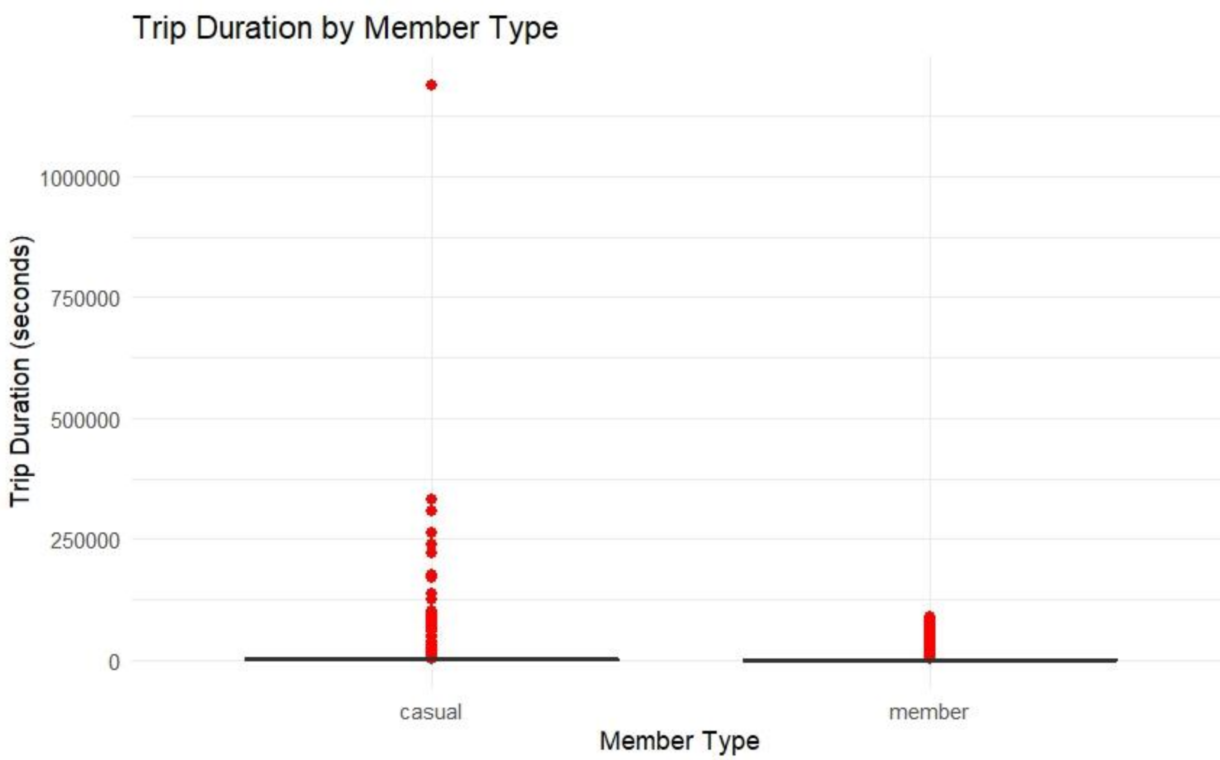
A significant majority of the riders are members, comprising 81% of the total usage. This indicates that the primary user base for Cyclistic bikes is composed of regular, subscribing members.

Casual riders make up a smaller portion, only 19% of the total. This smaller segment indicates that occasional or one-time users are less frequent compared to regular members.

Opportunities for Growth in Casual Segment as believed by the director of marketing that company's future success depends on maximizing the number of annual memberships.

The 19% casual rider segment represents an opportunity for growth. Cyclistic could develop strategies to convert casual riders into members by highlighting the cost savings and convenience of a membership.

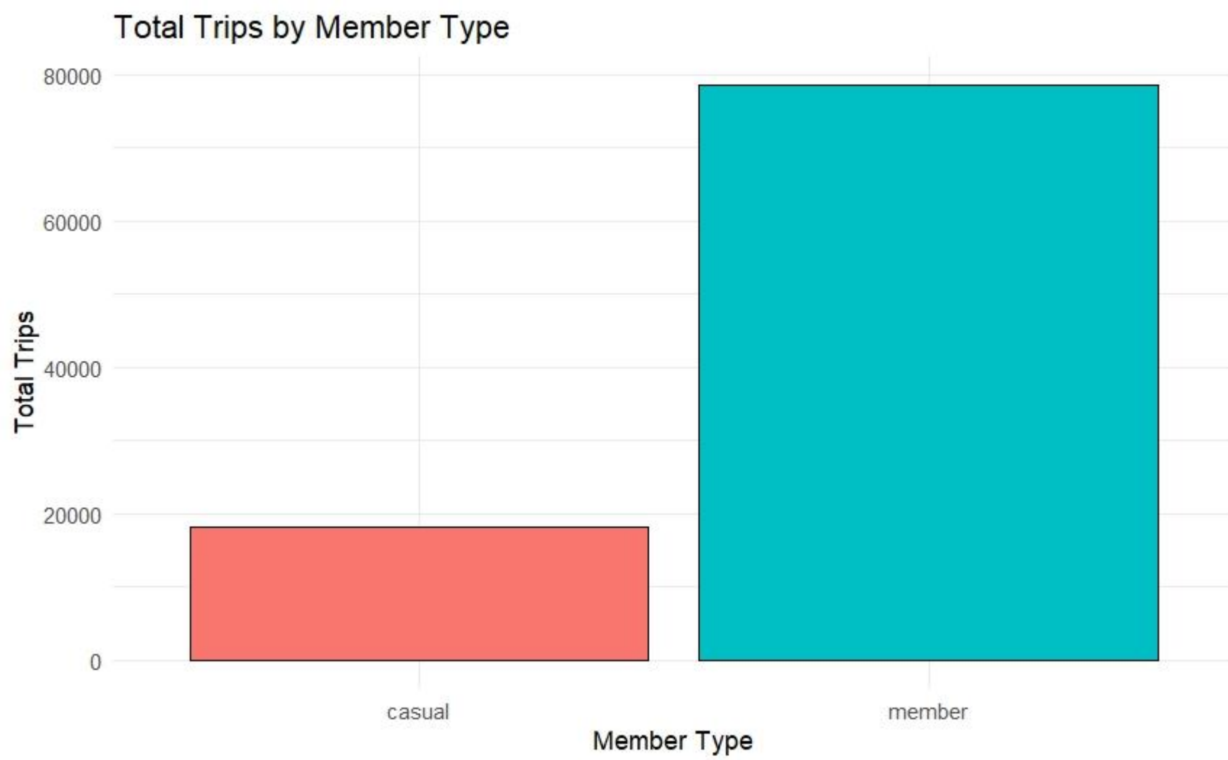
Trip Duration Distribution



Trip Duration Distribution: This graph suggests that there is an opportunity to focus new marketing strategy based on trip duration since casual riders use it for longest time.

- Casual Riders: Exhibit a wide range of trip durations with several outliers, indicating occasional very long trips.
- Members: Have a narrower range of trip durations with fewer outliers, indicating more consistent and shorter trips.

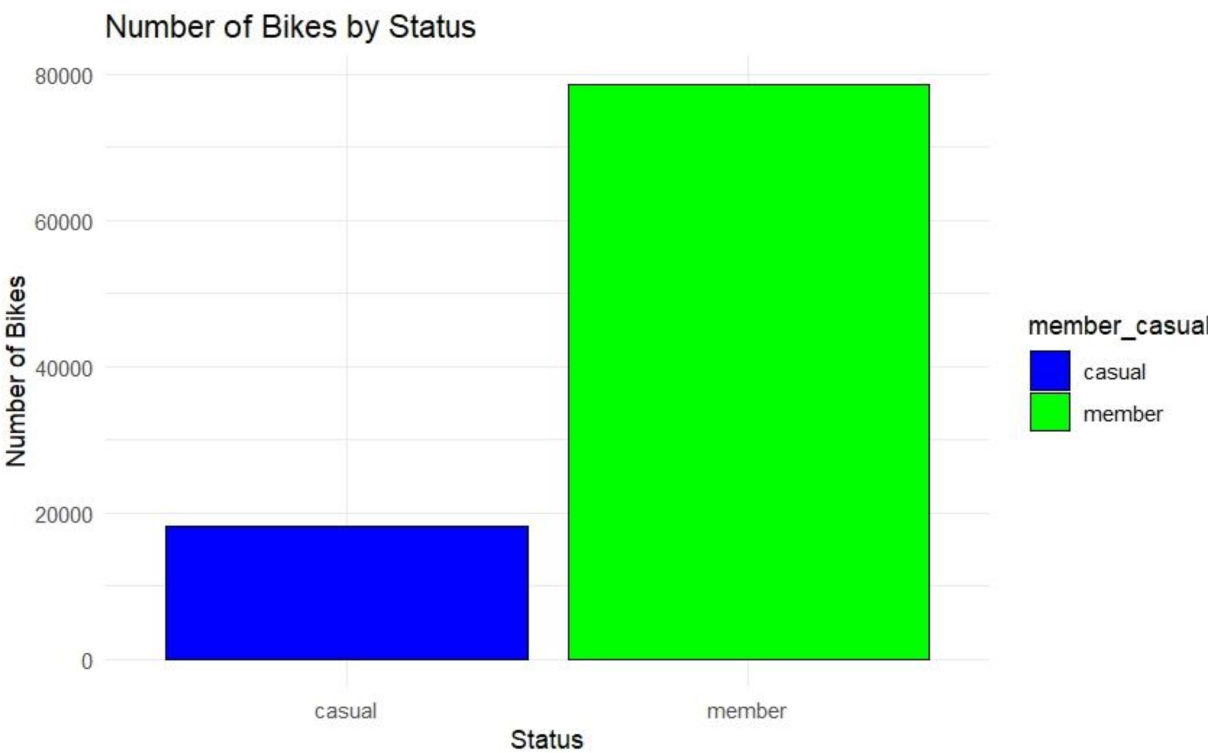
Total Trips by Member Type



- Annual members have taken approximately 80,000 trips.
- Casual riders have taken approximately 20,000 trips.

Annual members use the service much more frequently than casual riders. This suggests that annual members rely on Cyclistic bikes more regularly, possibly for commuting or regular activities.

Number of Bikes by Status



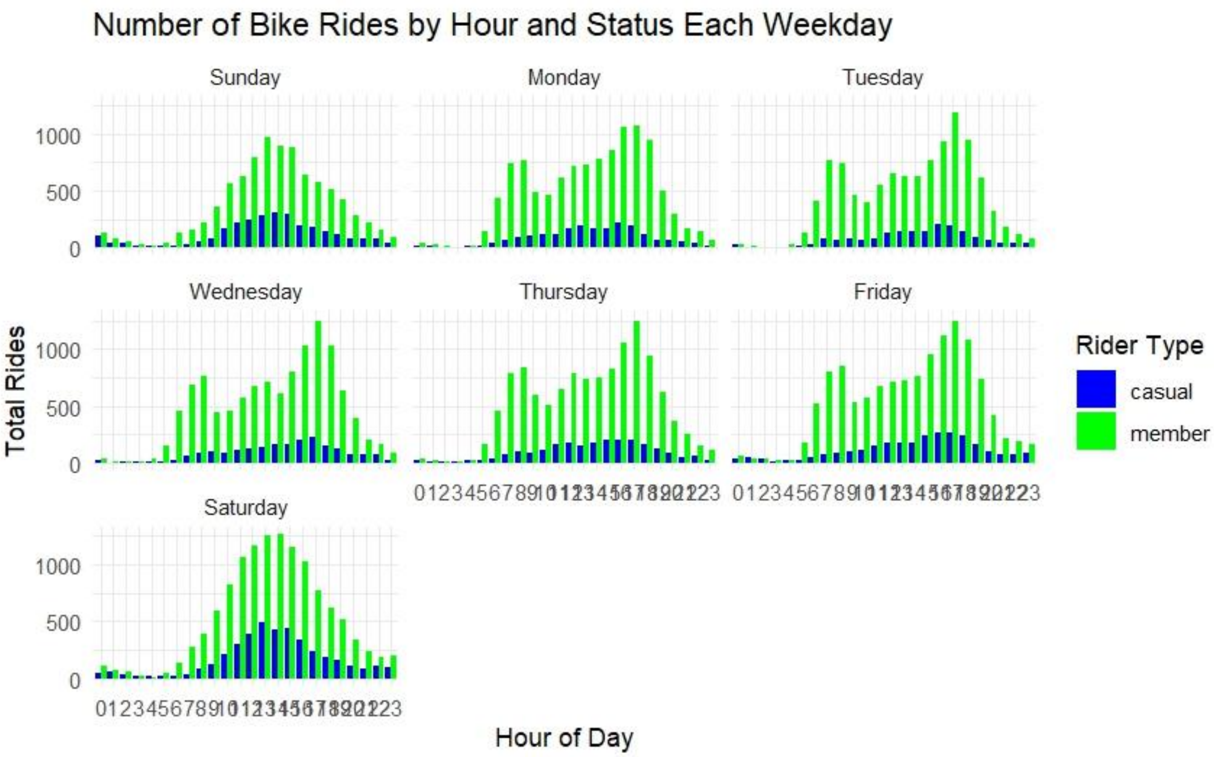
- Casual riders have used around 20,000 bikes.
- Members have used around 80,000 bikes.

Annual members use significantly more bikes than casual riders, indicating higher engagement and frequency of use.

The higher number of bikes used by members suggests they are more likely to use Cyclistic bikes regularly, possibly for daily commutes or regular errands.

Casual Use: Casual riders might use bikes sporadically, indicating lower commitment and possibly using the service for leisure or occasional trips.

Number of Bikes by Hour and Status Each Weekday



- Members (green) consistently have a higher number of rides across all days and hours.
- Peak times: Both groups have specific peak times:

Weekdays:

Members have morning peaks around 7-9 AM and evening peaks around 4-6 PM, likely reflecting commute times.

Casual riders have a smaller peak around the same times but also show higher usage during the mid-day.

Weekends:

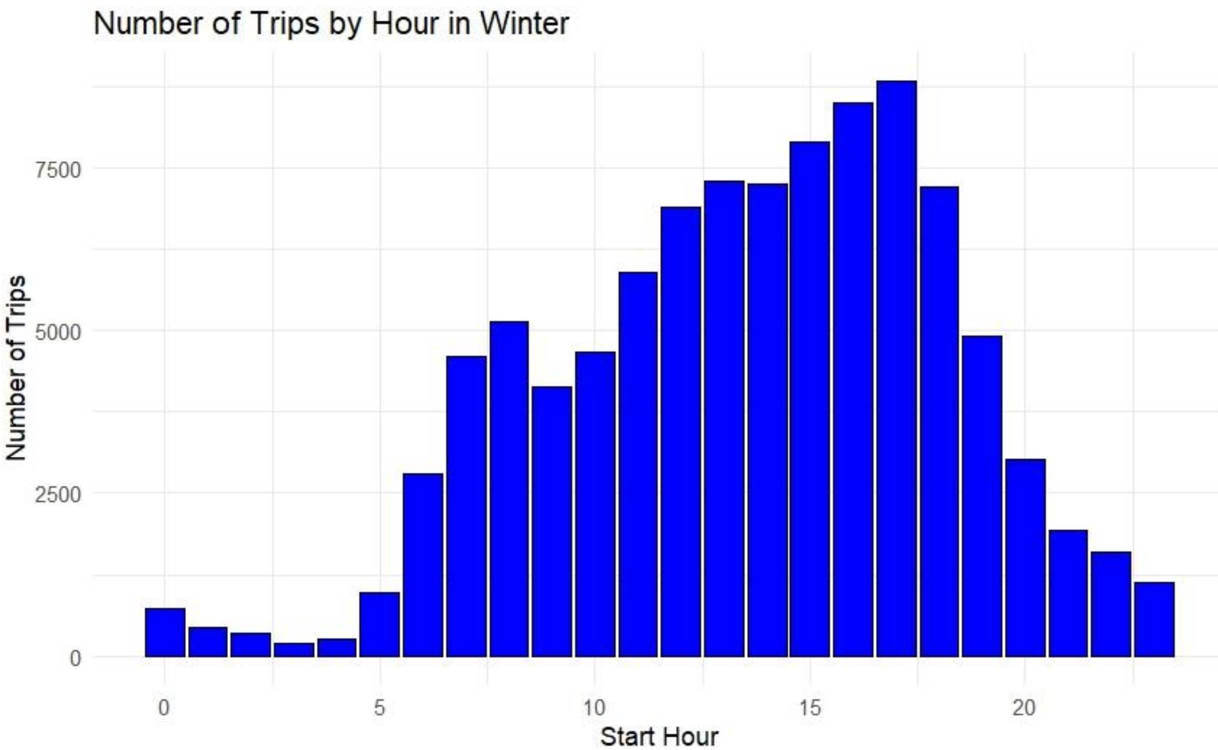
Both groups have morning peaks around 10 AM - 2 PM, indicating more recreational use.

Off-peak use: Members have higher utilization during off-peak hours as well, indicating more consistent use.

Members use the bikes more intensively, indicating a reliance on them for their daily commutes and other activities.

Casual Riders use their peaks during weekends and off-peak weekday hours, suggesting that they use the bikes for recreational purposes or special needs.

Number of Trips by Hour in Winter



Morning Commute (6 - 9 AM): Significant increase, smaller peak around 8-9 AM.

Midday (9 AM - 3 PM): High and stable usage, likely for errands and leisure.

Afternoon Peak (3 - 6 PM): Highest number of trips, peak around 3-4 PM, likely evening commutes.

Evening (6 - 11 PM): Gradual decline, still substantial usage for evening activities.

Late Night (11 PM - 12 AM): Low usage, minimal activity by midnight.

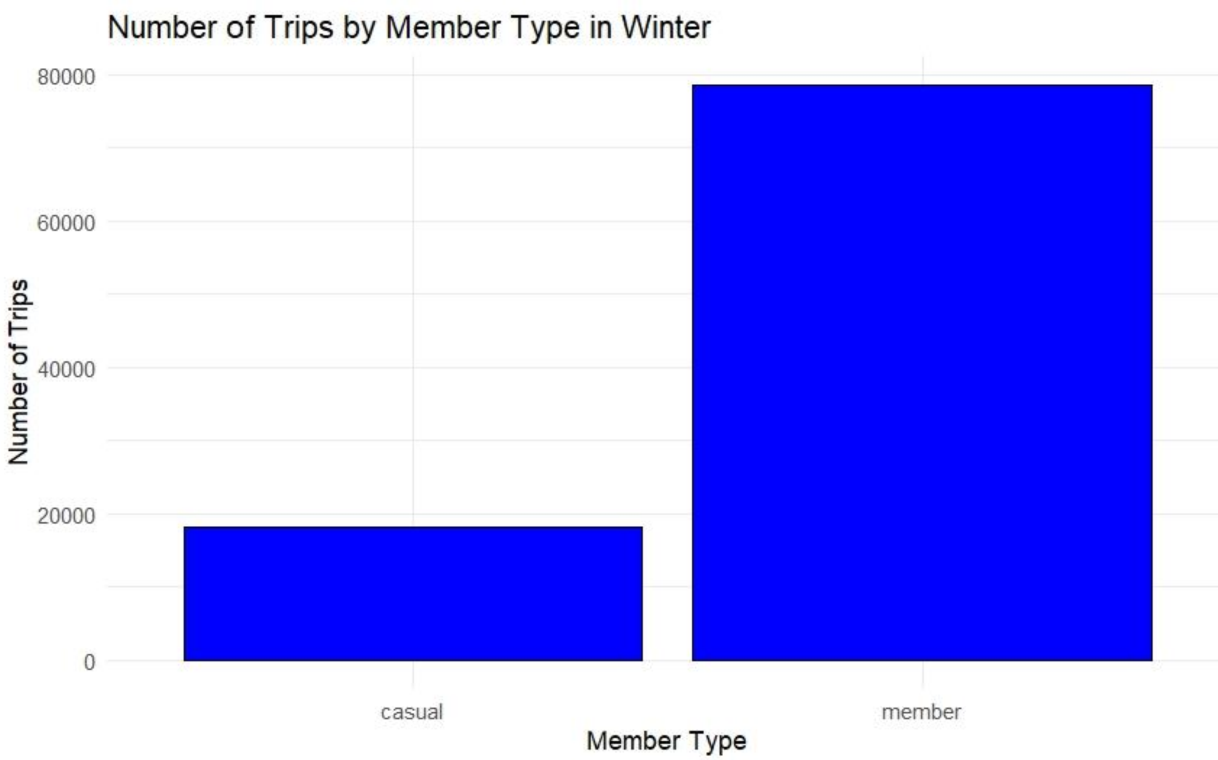
Differences in Usage:

Annual Members: Consistent usage, especially during commute hours, with structured patterns reflecting work/school schedules.

Casual Riders: Higher usage during non-commute hours for leisure and recreation, more sporadic patterns.

Annual members primarily use bikes for commuting, while casual riders use them more for leisure activities.

Number of Trip by Member Type in Winter

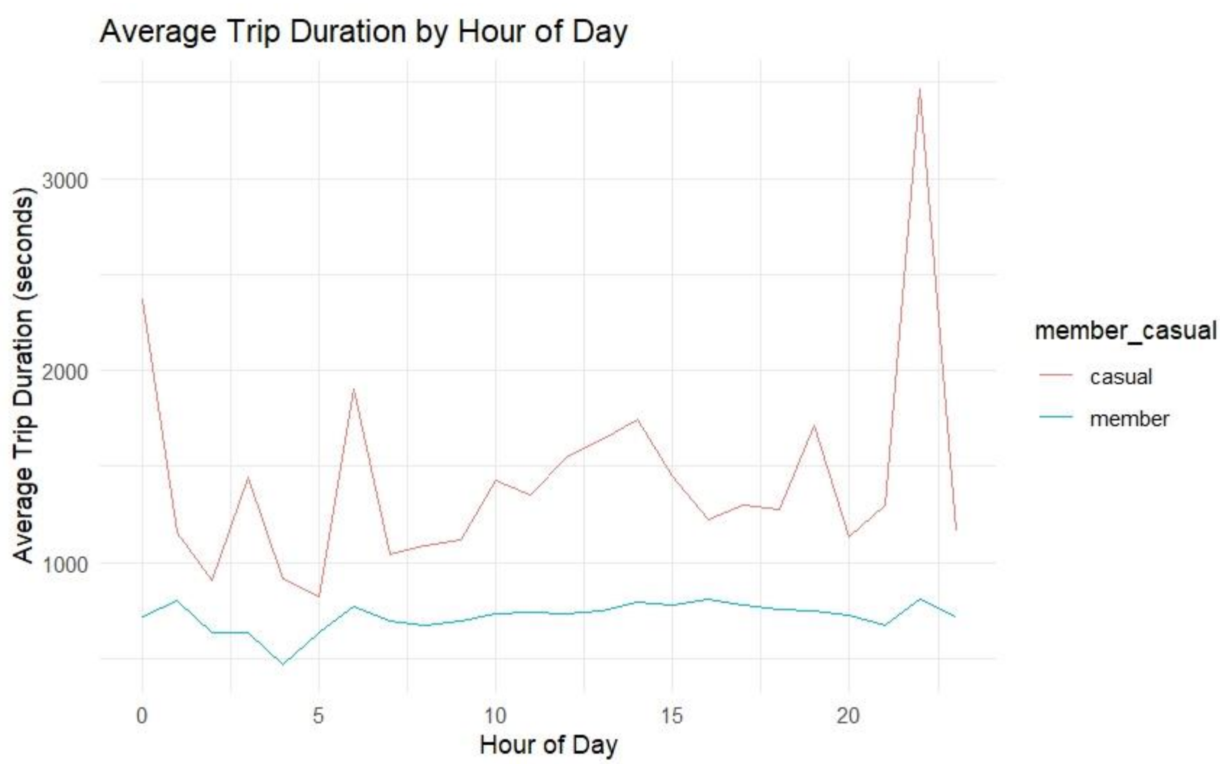


- Members: Significantly higher number of trips, with around 80,000 trips.
- Casual Riders: Much lower number of trips, with around 20,000 trips.

Annual Members uses Cyclistic bikes more frequently during winter, suggesting a higher reliance on bikes for regular commuting and possibly leisure activities.

Casual Riders use the bikes less frequently in winter, indicating that casual usage drops significantly compared to annual members.

Average Trip Duration by Hour of Day



Casual Riders: Average trip duration is generally higher compared to members, with significant spikes observed around the early morning (0-1 AM) and late evening (around 8 PM).

Members: Consistently lower and more stable average trip duration throughout the day.

Casual Riders: Experience high average trip durations in the early morning hours, peaking at around 3,000 seconds (50 minutes) around midnight.

Average trip duration decreases after early morning, with fluctuations throughout the day, and another peak in the late evening.

Members: Display a stable pattern with average trip durations generally staying below 1,000 seconds (16-17 minutes) across all hours.

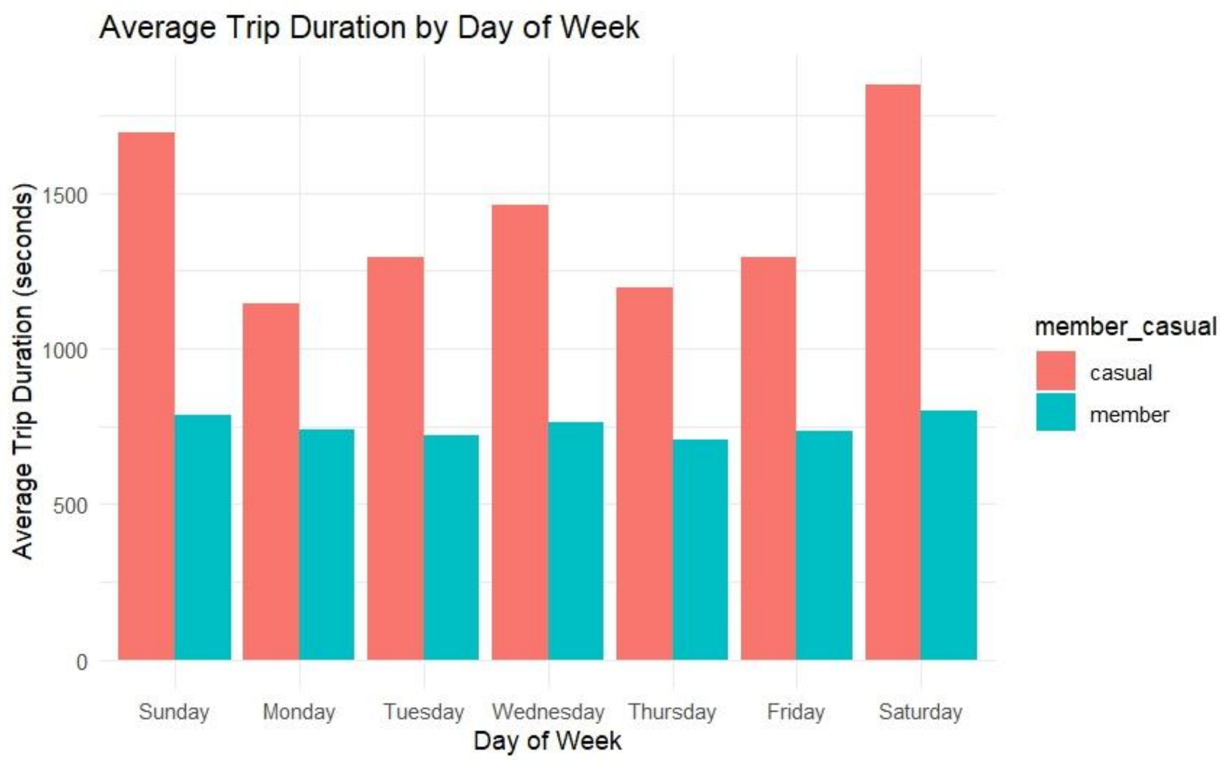
Minimal fluctuation, indicating regular and consistent usage for shorter durations.

Casual Riders: Likely use bikes for longer, possibly leisure-oriented trips, especially noticeable during non-peak hours (early morning and late evening).

Higher variability in trip duration suggests diverse usage patterns.

Members: Tend to use bikes for shorter, more consistent trips, likely for commuting or regular short-distance travel.

Average Trip Duration by Day of Week



Casual Riders: Average trip durations are consistently higher than those of members across all days of the week.

Members: Average trip durations are relatively lower and more stable compared to casual riders.

Casual Riders: Longer average trip durations, peaking on Sundays and Saturdays.

Generally higher average trip durations on weekends compared to weekdays.

Members: Consistent and shorter average trip durations throughout the week.

Slightly longer trip durations on weekends compared to weekdays, but the increase is not as pronounced as with casual riders.

Day-Specific Observations: Sunday: Highest average trip duration for casual riders, indicating increased leisure or recreational usage.

Saturday: Second highest average trip duration for casual riders.

Weekdays: Casual riders have a relatively higher but variable trip duration compared to members.

Members: Display stable trip durations throughout the week, with minor increases on weekends.

Casual Riders: Tend to use bikes more for longer, leisure-oriented trips during weekends, reflected in higher average trip durations on Saturdays and Sundays.

Weekday usage also indicates longer trips, possibly for exploration or recreational purposes rather than commuting.

Members: Use bikes for shorter, more consistent trips, likely for commuting and routine activities.

Weekend trips are slightly longer, suggesting some recreational use, but still much shorter on average than casual riders' trips.

SHARE

After analyzing the data trying to answer to understand how annual members and casual riders use our Cyclistic bikes differently particularly in winter? It is time to share the data I would use Tableau for sharing data.

ACT

Top Three Recommendations Based on Analysis

Targeted Marketing Campaigns:

Promote weekend and leisure packages to casual riders, highlighting scenic routes and recreational activities. Special offers or discounts could encourage higher usage.

Emphasize the convenience and cost-effectiveness of membership for daily commutes and regular short trips, potentially offering referral bonuses to increase memberships.

Operational Adjustments:

Increase bike availability and maintenance support during peak hours (morning and afternoon) to accommodate the higher number of trips.

Enhance services during weekends to cater to the longer trip durations of casual riders, ensuring bike availability in popular recreational areas.

Service Improvements:

Develop flexible membership plans that cater to the needs of different user groups, such as a weekend pass for casual riders or a commuter pass for daily users.

Implement a feedback system to continuously gather user input on trip experiences and areas for improvement, ensuring that services remain aligned with user needs and preferences.