Introduction

In the bustling city of Chicago, where traffic congestion often reigns supreme, Cyclistic, a dynamic bike-sharing company, has carved a niche for itself. With a diverse customer base that includes single-ride pass users, full-day pass enthusiasts, and annual members, Cyclistic offers an eco-friendly and convenient solution to urban transportation needs. However, amidst the cacophony of whirring wheels and the hum of city life, Cyclistic's Director of Marketing, Lily Moreno, faces a pivotal challenge. While annual memberships are undeniably the crown jewel of the business, the majority of Cyclistic's clientele comprises casual riders.

This case study delves into the world of Cyclistic, following my journey as a junior data analyst tasked with deciphering the habits and preferences that distinguish casual riders from annual members. Armed with historical trip data and a keen analytical eye, the analyst embarks on a quest to unearth valuable insights. The goal is clear: to create a marketing strategy that transforms casual riders into dedicated annual members, securing Cyclistic's future growth and success.

Through data cleaning, exploratory analysis, and the art of visual storytelling, this case study demonstrates the power of data analytics in driving business decisions. In the end, it culminates in three strategic recommendations, poised to reshape Cyclistic's marketing landscape.

Key stakeholders

● Cyclistic executive team.

● Lily Moreno (The director of marketing and manager)

● Cyclistic marketing analytics team

Ask

The questions to address here are:

● How do annual members and casual riders use Cyclistic bikes differently?

● Why would casual riders buy Cyclistic annual memberships?

● How can Cyclistic use digital media to influence casual riders to become members?

The insights gained by exploring these questions will help the marketing team to increase company profitability.

Data

This Project is to analyze Cyclistic's historical trip data provided by Motivate International Inc. under this license.(https://www.divvybikes.com/data-license-agreement). The link contains data from 2014 but for our purposes we will be using the data from the first quarter of 2023.

Original and Validated Data: The data I am working with is original and has been validated. This means that it is less likely to contain errors, inconsistencies, or inaccuracies that are common in some third-party datasets. Original data is collected firsthand, providing a higher level of reliability.

First-Party Data: First-party data is data collected directly from the source, in this case, Divvy's bike-sharing system. This type of data is highly reliable and trustworthy because it comes directly from the organization that operates the service, minimizing concerns about data quality and credibility.

Objective Information: The dataset exclusively comprises objective information related to Divvy's customers bike-sharing rides. This is crucial for data analysis because objective data is less susceptible to biases and subjective interpretation.

No Concerns Regarding Bias: Since the data comes from Divvy's own customers and is considered objective, there are no major concerns about bias in the dataset. This is important for ensuring the validity of your analysis and the accuracy of your findings.

Three-Month Period: The dataset covers a three-month period in 2020 Q1. This is a reasonable timeframe for conducting an analysis of ridership and behavior patterns.

PREPARE (Data cleaning and Manipulation)

- Downloaded Cyclistic trip 2023 Q1 data in CSV file and cleaned the data
- Checked the data for error
- Text wrapped and aligned all the fields so that it was easier to read and understand.
- Checked for duplicate data
- Checked for null values.
- Removed unwanted columns.(Ride ID, Station ID)
- Changed format of started_at and ended_at columns to DATETIME.
- Calculated Ride Length in minutes. That is the length of each ride by subtracting the column "started_at" from the column "ended_at". While calculating the Start date and End date, for some of the records the start date was greater than the end date. Deleted the error records.
- Calculate the day of the week that each ride started. That is 1 = Sunday and 7 = Saturday.

rideable_type	started_at	ended_at	start_station_name	end_station_name	start_lat	start_lng	end_lat	end_lng	member_casual	ride_length	day_of_week
electric_bike	2023-01-21 20:05:42	2023-01-21 20:16:33	Lincoln Ave & Fullerton Ave	Hampden Ct & Diversey Ave	41.924073935	-87.646278381	41.93	-87.64	member	10m 51s	7
classic_bike	2023-01-10 15:37:36	2023-01-10 15:46:05	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799568	-87.594747	41.809835	-87.599383	member	8m 29s	3
electric_bike	2023-01-02 07:51:57	2023-01-02 08:05:11	Western Ave & Lunt Ave	Valli Produce - Evanston Plaza	42.008571	-87.69048283333330	42.039742	-87.699413	casual	13m 14s	2
classic_bike	2023-01-22 10:52:58	2023-01-22 11:01:44	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799568	-87.594747	41.809835	-87.599383	member	8m 46s	1
classic_bike	2023-01-12 13:58:01	2023-01-12 14:13:20	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799568	-87.594747	41.809835	-87.599383	member	15m 19s	5
electric_bike	2023-01-31 07:18:03	2023-01-31 07:21:16	Lakeview Ave & Fullerton Pkwy	Hampden Ct & Diversey Ave	41.926068902	-87.638858199	41.93	-87.64	member	3m 13s	3
electric_bike	2023-01-15 21:18:36	2023-01-15 21:32:36	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799553633	-87.594616652	41.809835	-87.599383	member	14m	1
classic_bike	2023-01-25 10:49:01	2023-01-25 10:58:22	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799568	-87.594747	41.809835	-87.599383	member	9m 21s	4
electric_bike	2023-01-25 20:49:47	2023-01-25 21:02:14	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799587488	-87.594670296	41.809835	-87.599383	member	12m 27s	4
classic_bike	2023-01-06 16:37:19	2023-01-06 16:49:52	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799568	-87.594747	41.809835	-87.599383	member	12m 33s	6
classic_bike	2023-01-05 17:31:57	2023-01-05 17:41:46	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799568	-87.594747	41.809835	-87.599383	member	9m 49s	5
classic_bike	2023-01-03 17:32:34	2023-01-03 17:41:56	Kimbark Ave & 53rd St	Greenwood Ave & 47th St	41.799568	-87.594747	41.809835	-87.599383	member	9m 22s	3
electric_bike	2023-01-09 19:11:35	2023-01-09 19:19:15	Broadway & Waveland Ave	Hampden Ct & Diversey Ave	41.949081182	-87.648604989	41.93	-87.64	member	7m 40s	2
electric_bike	2023-01-03 20:25:53	2023-01-03 20:35:50	Broadway & Waveland Ave	Hampden Ct & Diversey Ave	41.94910574	-87.64862752	41.93	-87.64	casual	9m 57s	3
electric_bike	2023-01-12 22:12:32	2023-01-12 22:17:07	Lincoln Park Conservatory	Hampden Ct & Diversey Ave	41.924067736	-87.635828733	41.93	-87.64	member	4m 35s	5
classic_bike	2023-01-09 21:09:30	2023-01-09 21:16:08	Clark St & Columbia Ave	Warren Park West	42.00445062934	-87.6724024047	42.001785	-87.688829	member	6m 38s	2
electric_bike	2023-01-21 09:13:54	2023-01-21 09:16:24	Lakeview Ave & Fullerton Pkwy	Hampden Ct & Diversey Ave	41.925889015	-87.638755322	41.93	-87.64	member	2m 30s	7
electric_bike	2023-01-05 17:28:08	2023-01-05 17:43:24	Western Ave & Lunt Ave	Valli Produce - Evanston Plaza	42.0086135	-87.6905225	42.039742	-87.699413	casual	15m 16s	5
electric_bike	2023-01-17 17:17:48	2023-01-17 17:33:50	McClurg Ct & Ohio St	Hampden Ct & Diversey Ave	41.89287138	-87.617195845	41.93	-87.64	member	16m 2s	3
classic_bike	2023-01-03 18:18:33	2023-01-03 19:07:22	McClurg Ct & Ohio St	Clark St & Elmdale Ave	41.892592119709700	-87.61728912591930	41.990860448125600	-87.66972362995150	member	48m 49s	3
electric_bike	2023-01-02 17:32:56	2023-01-02 17:47:58	Clarendon Ave & Gordon Ter	Clark St & Elmdale Ave	41.957727909	-87.649412751	41.990860448125600	-87.66972362995150	member	15m 2s	2
electric_bike	2023-01-01 18:04:58	2023-01-01 18:09:00	Clark St & Berwyn Ave	Clark St & Elmdale Ave	41.977920413	-87.668058395	41.990860448125600	-87.66972362995150	member	4m 2s	1
classic_bike	2023-01-02 12:45:48	2023-01-02 12:53:06	Clark St & Berwyn Ave	Clark St & Elmdale Ave	41.978030622486600	-87.66856491565700	41.990860448125600	-87.66972362995150	casual	7m 18s	2
classic_bike	2023-01-17 17:15:15	2023-01-17 17:44:38	Clark St & Newport St	Warren Park West	41.94454	-87.654678	42.001785	-87.688829	member	29m 23s	3
electric_bike	2023-01-11 17:03:25	2023-01-11 17:13:34	Avondale Ave & Irving Park Rd	Campbell Ave & Irving Park Rd	41.953178167	-87.731947541	41.95	-87.69	casual	10m 9s	4
electric_bike	2023-01-16 18:16:20	2023-01-16 18:24:22	Wilton Ave & Diversey Pkwy	Hampden Ct & Diversey Ave	41.932496666666700	-87.65268	41.93	-87.64	member	8m 2s	2
electric_bike	2023-01-12 07:36:18	2023-01-12 07:44:11	Komensky Ave & 55th St	Pulaski Rd & 51st St	41.79	-87.72	41.8	-87.72	member	7m 53s	5
electric_bike	2023-01-17 12:46:15	2023-01-17 12:54:55	Clark St & Newport St	Hampden Ct & Diversey Ave	41.944490552	-87.654772282	41.93	-87.64	member	8m 40s	3
electric_bike	2023-01-24 17:53:39	2023-01-24 17:58:58	Clark St & Columbia Ave	Clark St & Elmdale Ave	42.004458785	-87.672376752	41.990860448125600	-87.66972362995150	casual	5m 19s	3
electric_bike	2023-01-13 15:21:08	2023-01-13 15:26:19	Lincoln Ave & Fullerton Ave	Hampden Ct & Diversey Ave	41.924267888	-87.646338344	41.93	-87.64	member	5m 11s	6
electric_bike	2023-01-21 18:16:25	2023-01-21 18:24:26	Lincoln Ave & Fullerton Ave	Hampden Ct & Diversey Ave	41.92431128	-87.646572113	41.93	-87.64	member	8m 1s	7
classic_bike	2023-01-11 17:22:28	2023-01-11 17:39:45	Central St & Girard Ave	Valli Produce - Evanston Plaza	42.064313	-87.686152	42.039742	-87.699413	member	17m 17s	4
electric_bike	2023-01-18 07:51:13	2023-01-18 07:54:13	Komensky Ave & 55th St	Pulaski Rd & 51st St	41.79	-87.72	41.8	-87.72	member	3m	4
electric_bike	2023-01-19 07:48:40	2023-01-19 07:52:17	Komensky Ave & 55th St	Pulaski Rd & 51st St	41.79	-87.72	41.8	-87.72	member	3m 37s	5
electric_bike	2023-01-06 12:33:24	2023-01-06 12:42:58	Troy St & Elston Ave	Campbell Ave & Irving Park Rd	41.945234537	-87.706618905	41.95	-87.69	casual	9m 34s	6
classic_bike	2023-01-07 13:16:50	2023-01-07 13:21:20	Dorchester Ave & 49th St	Greenwood Ave & 47th St	41.805772	-87.592464	41.809835	-87.599383	member	4m 30s	7
electric_bike	2023-01-07 04:35:39	2023-01-07 04:38:38	Dorchester Ave & 49th St	Greenwood Ave & 47th St	41.805816889	-87.592506051	41.809835	-87.599383	member	2m 59s	7
electric_bike	2023-01-07 16:42:49	2023-01-07 17:03:08	Millennium Park	Hampden Ct & Diversey Ave	41.881034374	-87.624126554	41.93	-87.64	member	20m 19s	7
electric_bike	2023-01-14 10:50:48	2023-01-14 10:57:01	Clark St & Newport St	Hampden Ct & Diversey Ave	41.94445783333330	-87.65474183333330	41.93	-87.64	casual	6m 13s	7
electric_bike	2023-01-19 19:17:46	2023-01-19 19:18:14	Hampden Ct & Diversey Ave	Hampden Ct & Diversey Ave	41.93	-87.64	41.93	-87.64	casual	28s	5
electric_bike	2023-01-26 12:36:49	2023-01-26 12:37:37	Hampden Ct & Diversey Ave	Hampden Ct & Diversey Ave	41.93	-87.64	41.93	-87.64	member	48s	5

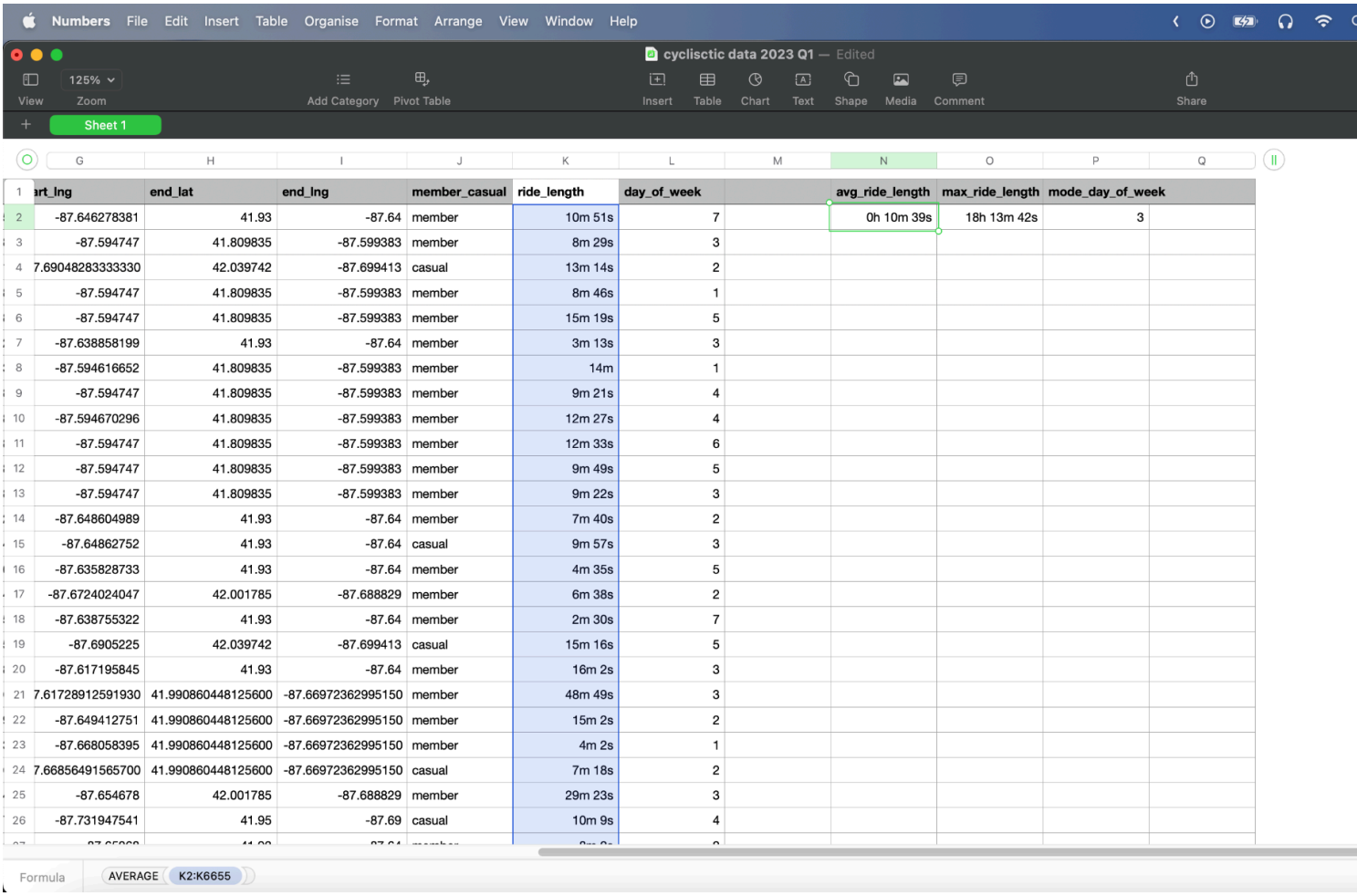
Analyze

The data has been cleaned. Now I had to analyze it in order to answer the relevant questions.

Analysis using SQL

Run a few calculations using SQL and Spreadsheets to get a better sense of the data layout:

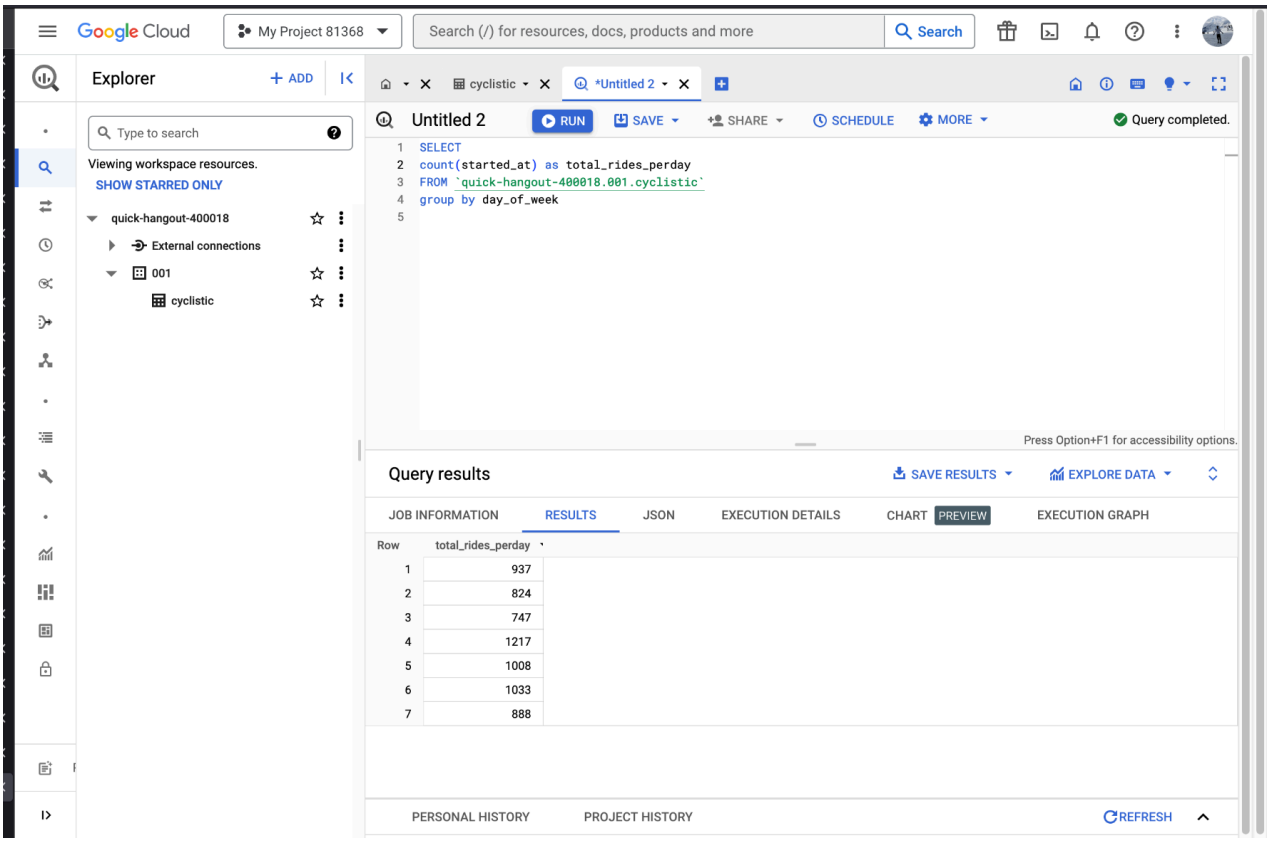
- Calculate the mean of ride_length = 10min 39sec
- Calculate the max ride_length = 18hrs 13min 42sec
- Calculate the mode of day_of_week = 3 (Tuesday)

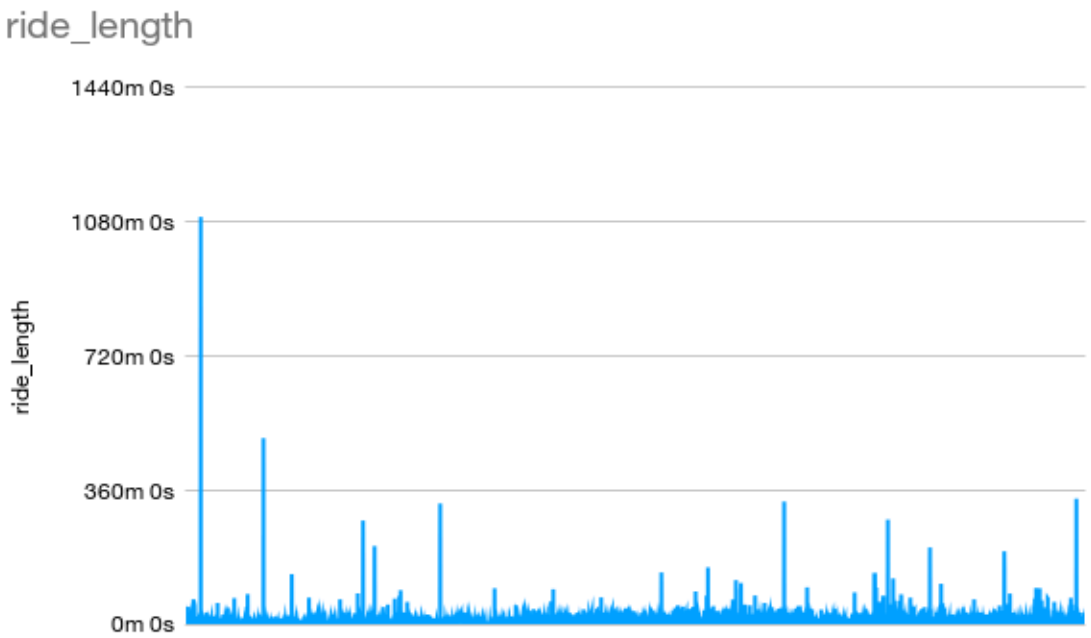


	art_lng	end_lat	end_lng	member_casual	ride_length	day_of_week		avg_ride_length	max_ride_length	mode_day_of_week
2	-87.646278381	41.93	-87.64	member	10m 51s	7		0h 10m 39s	18h 13m 42s	3
3	-87.594747	41.809835	-87.599383	member	8m 29s	3				
4	7.69048283333330	42.039742	-87.699413	casual	13m 14s	2				
5	-87.594747	41.809835	-87.599383	member	8m 46s	1				
6	-87.594747	41.809835	-87.599383	member	15m 19s	5				
7	-87.638858199	41.93	-87.64	member	3m 13s	3				
8	-87.594616652	41.809835	-87.599383	member	14m	1				
9	-87.594747	41.809835	-87.599383	member	9m 21s	4				
10	-87.594670296	41.809835	-87.599383	member	12m 27s	4				
11	-87.594747	41.809835	-87.599383	member	12m 33s	6				
12	-87.594747	41.809835	-87.599383	member	9m 49s	5				
13	-87.594747	41.809835	-87.599383	member	9m 22s	3				
14	-87.648604989	41.93	-87.64	member	7m 40s	2				
15	-87.64862752	41.93	-87.64	casual	9m 57s	3				
16	-87.635828733	41.93	-87.64	member	4m 35s	5				
17	-87.6724024047	42.001785	-87.688829	member	6m 38s	2				
18	-87.638755322	41.93	-87.64	member	2m 30s	7				
19	-87.6905225	42.039742	-87.699413	casual	15m 16s	5				
20	-87.617195845	41.93	-87.64	member	16m 2s	3				
21	7.61728912591930	41.990860448125600	-87.66972362995150	member	48m 49s	3				
22	-87.649412751	41.990860448125600	-87.66972362995150	member	15m 2s	2				
23	-87.668058395	41.990860448125600	-87.66972362995150	member	4m 2s	1				
24	7.66856491565700	41.990860448125600	-87.66972362995150	casual	7m 18s	2				
25	-87.654678	42.001785	-87.688829	member	29m 23s	3				
26	-87.731947541	41.95	-87.69	casual	10m 9s	4				

Formula AVERAGE K2:K6655

- Total Rides Per Day = 1 refers to Sunday and 6 refers to Saturday





Visualization Using R

Create a pivot table to calculate and visualize the data quickly using R for below:

- Calculate the average ride_length for members and casual riders.
- Calculate the average ride_length for users by day_of_week.
- Calculate the number of rides for users by day_of_week

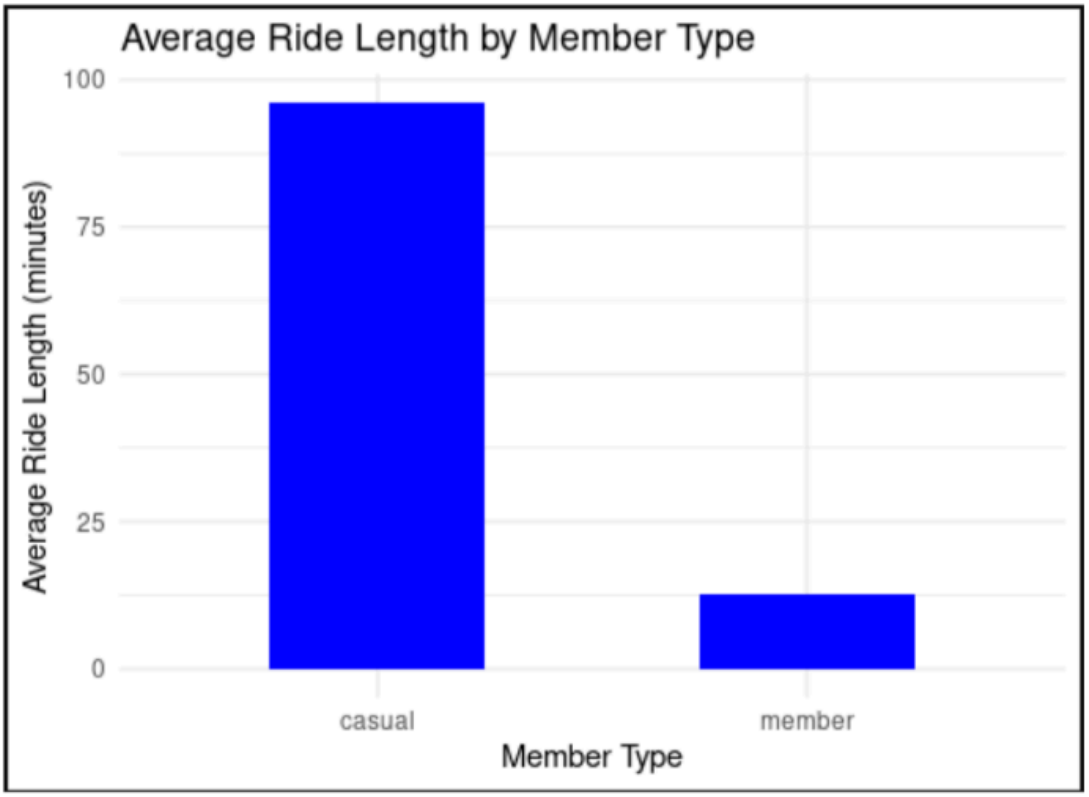
Calculate the average ride_length for members and casual riders. Try rows = member_casual; Values = Average of ride_length

```
library(dplyr)
library(ggplot2)
library(tidyverse)
library(readr)
# Read the CSV file
data <- read_csv("cyclistic_data.csv";)
# Calculate the average ride_length_in_minute
average_ride_length_by_member <- data %>%
group_by(member_casual) %>%
summarize(average_ride_length = mean(ride_length))
# Create a bar plot
ggplot(average_ride_length_by_member, aes(x = member_casual, y =
average_ride_length)) +
geom_bar(stat = "identity", fill = "blue", width = 0.5) +
```

```
labs(
title = “Average Ride Length by Member Type”,
x = “Member Type”,
y = “Average Ride Length (minutes)”
) +
theme_minimal()
```

Response:

Customer Type	Average Ride Length in Minute
Casual	96.0
member	12.7



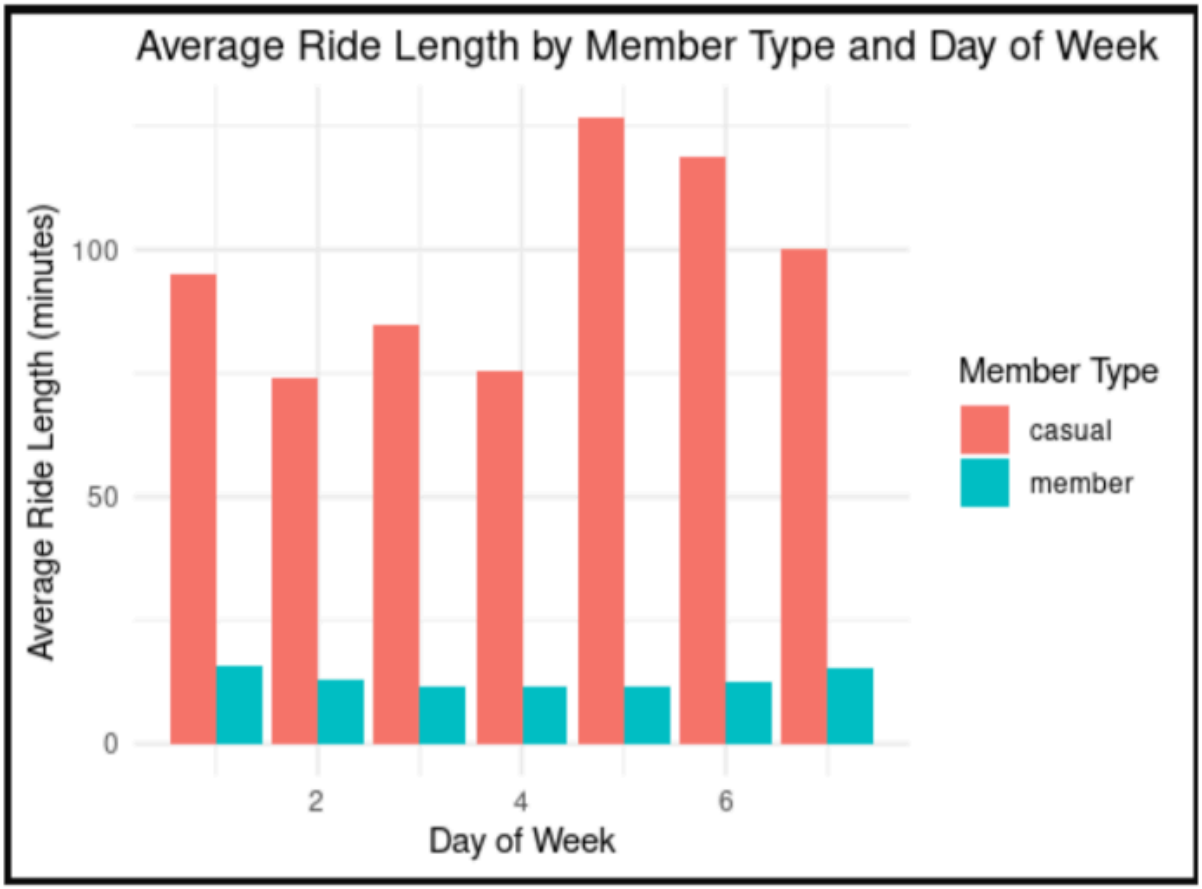
Calculate the average ride length for users by day per week. Try columns = day_of_week;Rows = member_casual; Values = Average of ride_length.

```
library(tidyr)
library(dplyr)
library(ggplot2)
# Read the CSV file
data <- read_csv(“cyclistic_data.csv”)
```

```
# Calculate the average ride_length_in_minute for users by day_of_week
result<- data %>%
group_by(member_casual, day_of_week) %>%
summarize(average_ride_length = mean(ride_length_))

# Create a bar graph
ggplot(result, aes(x = day_of_week, y = average_ride_length, fill = member_casual)) +
geom_bar(stat = “identity”, position = “dodge”) +
labs(
title =”Average Ride Length by Member Type and Day of Week”,
x = “Day of Week”,
y = “Average Ride Length (minutes)”,
fill = “Member Type”
) +
theme_minimal()
```

Result:



Calculate the number of rides for users by day_of_week by adding Count of trip_id to Values.

```
library(dplyr)
library(ggplot2)
```

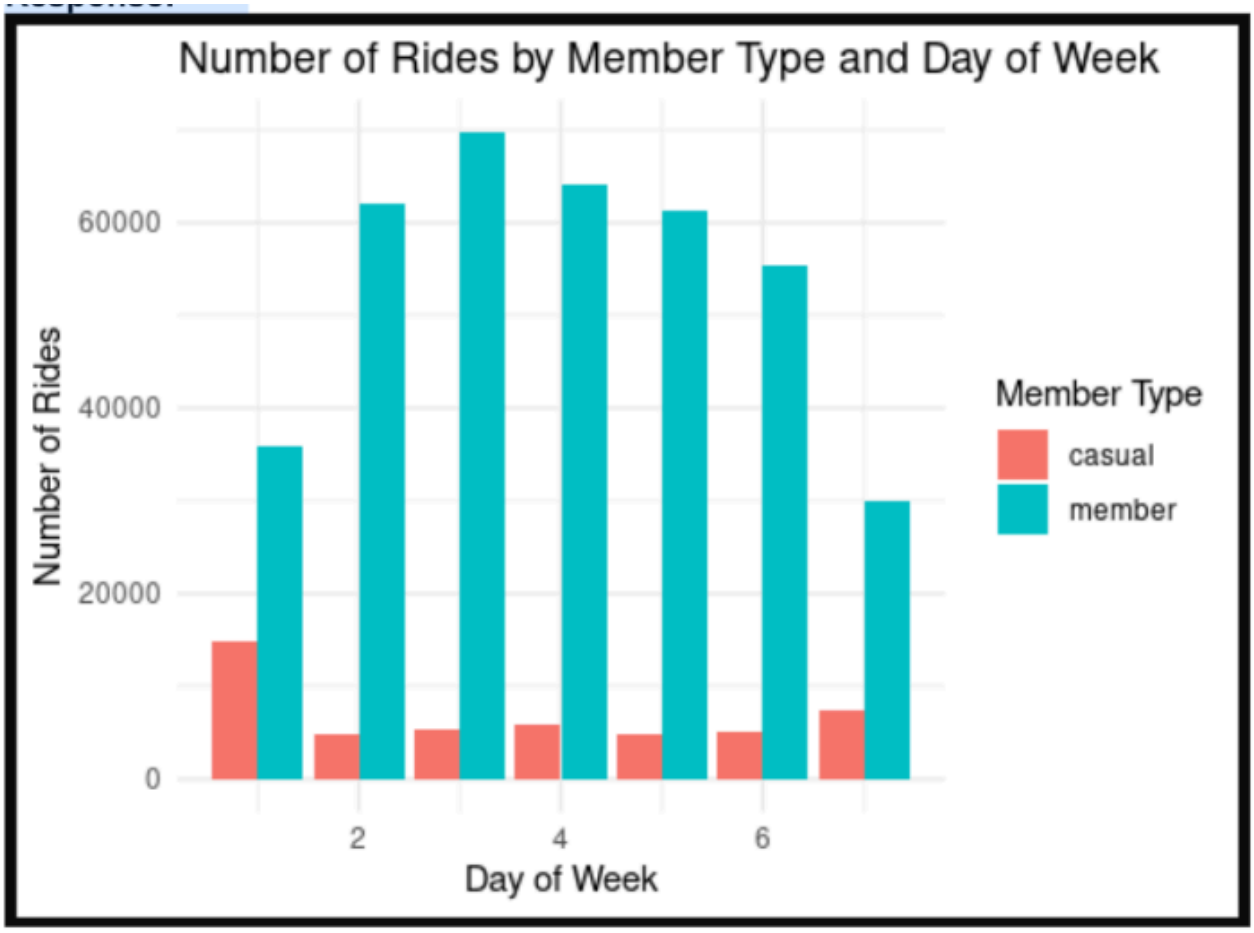
```
# Read the CSV file

data <- read_csv("cyclistic_data.csv")

# Calculate the number of rides for users by day_of_week
ride_counts <- data %>%
group_by(member_casual, day_of_week) %>%
summarize(ride_count = n())

# Create a bar graph
ggplot(ride_counts, aes(x = day_of_week, y = ride_count, fill = member_casual)) +
geom_bar(stat ="identity", position = "dodge") +
labs(
title = "Number of Rides by Member Type and Day of Week",
x = "Day of Week",
y = "Number of Rides",
fill = "Member Type"
) +
theme_minimal()
```

Response:

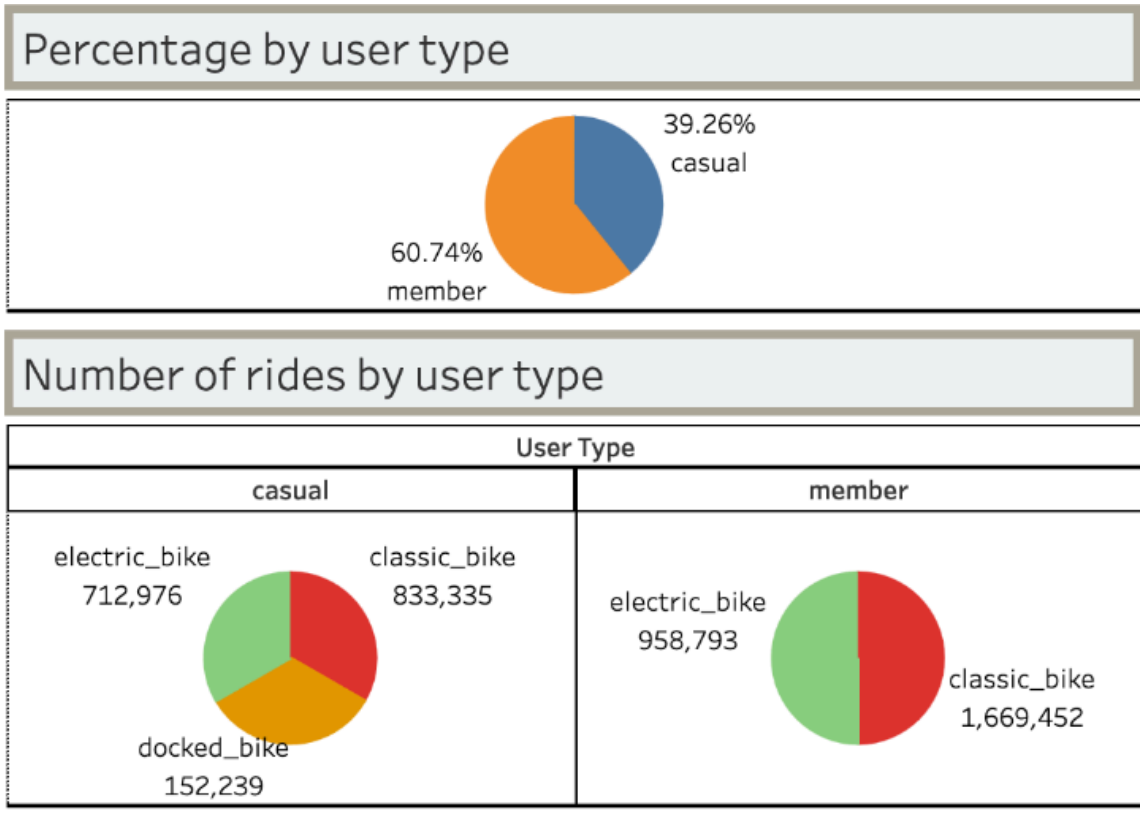


Summary of R Analysis

● Compare the Ride Length of customers by members and casual riders. The Casular riders average ride length is more than the Members.

● Calculated the average length for users by day_of_week. The Casual riders average ride length by day of the week is more than the Members.

● Calculated the number of rides for users by day_of_week by adding the Count of trip_id. The number of rides by members is more than the Casual riders.



Key Findings:

- Casual Riders ride longer on average than members.
- Number of Annual Members is more than the number of casual members.
    - **Undoubtedly, annual members have a higher ride count, but it is noteworthy that casual members travel twice the distance compared to annual members.**
- People use Cyclistic most on Saturdays and Sundays in general.
- Casual members demonstrate greater travel activity on weekends, whereas annual members exhibit higher travel volumes on weekdays, potentially indicating a weekday commuting pattern.

- There are few stations with a higher volume of people starting and ending their commute than others.
- The share of electric and classic bikes are almost equal, which can in turn help the company in promoting the benefits and expanding its fleet of electric bikes

Recommendations:

A. Promote longer distance rides on weekdays
    a. Recognize and leverage the fact that casual riders tend to ride for longer distances on weekdays.
    b. Consider special promotions or challenges that reward frequent weekday riders.

B. Increase Awareness of Annual Subscription:
    a. Create a comprehensive marketing strategy that targets both new clients and casual riders, informing them about the annual subscription option.
    b. Use various marketing channels such as email campaigns, in-app notifications, and physical signage at bike stations to increase awareness.
    c. Highlight the cost savings and convenience of annual membership compared to pay-per-ride for casual riders.

C. Leverage Testimonials for Community Building:
    a. Showcase the benefits of annual memberships by obtaining testimonials from existing members who have experienced positive life changes through Cyclistic's system.
    b. These testimonials can be used in marketing materials, on the website, and in
    promotional videos to build trust and a sense of community among potential members.

D. Offer Limited-Time Discounts:
    a. To incentivize new riders to join the annual membership program, offer limited-time discounts or promotions.
    b. Create a sense of urgency by specifying the end date of the discount period.
    c. Highlight the cost-effectiveness of taking advantage of the discount.

E. Launch a Social Media Campaign:

a. Utilize social media platforms to run a targeted campaign that emphasizes the advantages of annual membership.
b. Share success stories, customer testimonials, and engaging content that showcases the benefits and convenience of annual membership.
c. Encourage social sharing and engagement to reach a wider audience by offering giveaways

F. Engage with the Community:
    a. Organize community events, group rides, or contests that bring together casual riders and annual members.
    b. Use these events to promote the benefits of annual membership and create a sense of belonging within the Cyclistic community.

G. Monitor and Analyze Results:
    a. Implement tracking mechanisms to monitor the effectiveness of the marketing strategies.
    b. Continuously analyze conversion rates, engagement metrics, and customer feedback to make data-driven adjustments to the campaign.

H. Personalize Communication and Targeted Advertisements:
    a. Tailor marketing messages to address the specific needs and preferences of casual riders.
    b. Use data analytics to understand the behavior and preferences of casual riders and send personalized recommendations and offers.
    c. Place advertisements at such bike docks where a large number of riders visit.

I. Brand Partnership:
    a. Partner with other health-driven brands to enhance the company's reach.
    b. Cyclistic could partner with various health-tracking companies to make customized health bands for them

-Ayan Nayyer