Case study: How does a bike-share navigate speedy success?

Google Data Analytics Capstone Project

by Avinash Mishra

---

1. Ask Phase

About the company

Cyclistic is a bike-share offering company based in Chicago. They have fleet of **5,824 bicycles** that are geotracked and locked into a **network of 692 stations** across the city.

Stakeholder

- **Lily Moreno**, Director of Marketing

Current strategy

Cyclistic serves two types of riders,

- **Casual riders,** who rent for a day or as per their need.
- **Cyclistic members**, who buy annual membership.

Uptil now they relied upon **general awareness campaign** for their marketing strategy.

Goal

To design data driven marketing strategies aimed at converting casual riders into annual members.

Ask Phase

The analysis aims to answer the following key business questions.

1. How do annual members and casual riders use Cylistic bikes differently?
2. Why would casual riders buy Cyclistic annual membership?
3. How can Cyclistic use digital media to influence casual riders to become members?

---

2. Prepare Phase- Data Source and Structure

Purpose: Understand what data is available for analysis. Whether it is reliable or not? Whether it is complete?

Data Source:

The data for this analysis comes from **Divvy Trip Data**, provided by **Motivate International Inc.** under this license.

Link: https://divvy-tripdata.s3.amazonaws.com/index.html

Time Period:

November 2025 - October 2025 (12 Months)

Data Field:

Total Dataset Size: >1gb

Purpose Of Data:

This dataset has records of individual rides from Cyclistic bike sharing program.

Each record provides information about:

- Bike type
- Start and End time
- Start and End station
- Starting latitude and Ending latitude
- Starting longitude and Ending longitude
- Membership type

Data Limitations:

- 12 CSV files for 12 months, need to combine in one
- Missing values

I downloaded CSV files for November 2024 to October 2025—12 files in total. I will combine them into a single dataset for further analysis.

Data Collection and Organization Workflow

I downloaded the Divvy Trip Data from the official Divvy Trip Data repository.

To ensure consistency and reproducibility, I followed a structured file management process:

1. **Downloaded** 12 monthly CSV files (Nov 2024–Oct 2025) in **.zip** format.
2. **Extracted** each **.zip** file into a folder.
3. **Stored** all extracted CSVs in Extracted folder.
4. **Combined** the 12 CSVs into one consolidated dataset using R for cleaning and analysis.

This approach ensured all data files were organized, traceable, and ready for analysis.

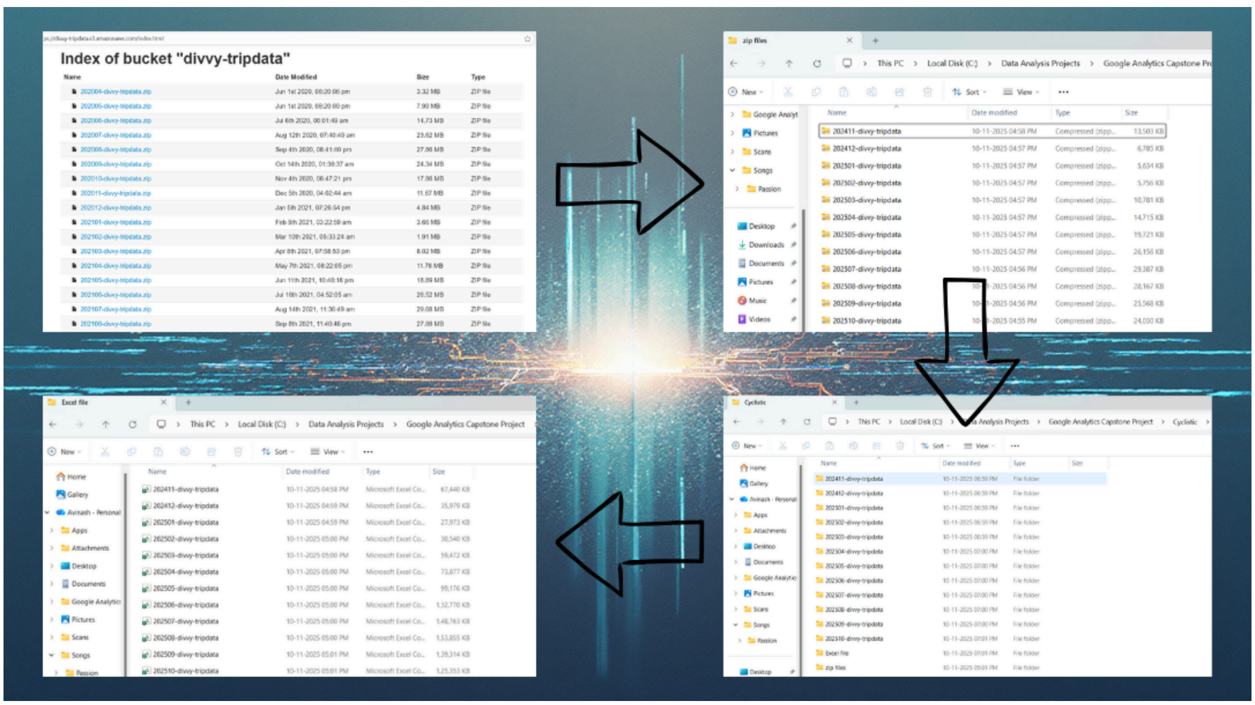


Figure: Data download and organization workflow — from raw .zip files to a single combined dataset.

---

3. Process Phase - Data Cleaning and Preparation

Purpose: To clean, transform, and prepare the combined dataset, so its accurate, consistent, and ready for analysis.

Tools Used:

- **R programming** (R Studio)
- Libraries: `tidyverse`, `dpylr`, `lubridate`, `janitor`

Steps Followed:

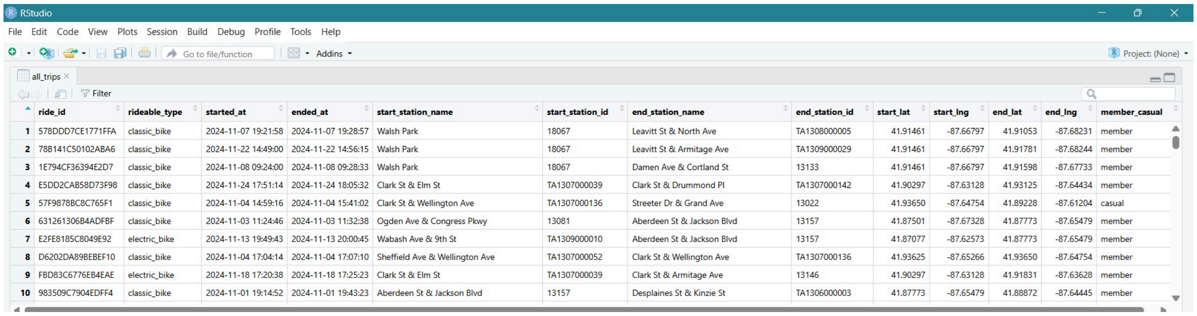
1. Import and combine the data

   After unzipping all 12 months CSVs, I combined them into a single dataset using R.

   ```
    library(tidyverse)
    library(lubridate)
    library(janitor)
   ```

   ```
   path ← "C:/Data Analysis Projects/Google Analytics Capstone Project/Cyc
   listic/CSVs"
   ```

   ```
   files ← list.files(path = path, pattern = "*.CSV", full.names = TRUE) ## read
   s
   all_trips ← files %>% map_df(read_csv) ## combines
   ```



	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
1	578DDD7CE1771FFA	classic_bike	2024-11-07 19:21:58	2024-11-07 19:28:57	Walsh Park	18067	Leavitt St & North Ave	TA1308000005	41.91461	-87.66797	41.91053	-87.68231	member
2	78B141C50102ABA6	classic_bike	2024-11-22 14:49:00	2024-11-22 14:56:15	Walsh Park	18067	Leavitt St & Armitage Ave	TA1309000029	41.91461	-87.66797	41.91781	-87.68244	member
3	1E794CF36394E2D7	classic_bike	2024-11-08 09:24:00	2024-11-08 09:28:33	Walsh Park	18067	Damen Ave & Cortland St	13133	41.91461	-87.66797	41.91598	-87.67733	member
4	E5DD2CAB58D73F98	classic_bike	2024-11-24 17:51:14	2024-11-24 18:05:32	Clark St & Elm St	TA1307000039	Clark St & Drummond Pl	TA1307000142	41.90297	-87.63128	41.93125	-87.64434	member
5	57F9878BC8C765F1	classic_bike	2024-11-04 14:59:16	2024-11-04 15:41:02	Clark St & Wellington Ave	TA1307000136	Streeter Dr & Grand Ave	13022	41.93650	-87.64754	41.89228	-87.61204	casual
6	631261306B4ADFBF	classic_bike	2024-11-03 11:24:46	2024-11-03 11:32:38	Ogden Ave & Congress Pkwy	13081	Aberdeen St & Jackson Blvd	13157	41.87501	-87.67328	41.87773	-87.65479	member
7	E2FE8185C8049E92	electric_bike	2024-11-13 19:49:43	2024-11-13 20:00:45	Wabash Ave & 9th St	TA1309000010	Aberdeen St & Jackson Blvd	13157	41.87077	-87.62573	41.87773	-87.65479	member
8	D6202DA89BEBEF10	classic_bike	2024-11-04 17:04:14	2024-11-04 17:07:10	Sheffield Ave & Wellington Ave	TA1307000052	Clark St & Wellington Ave	TA1307000136	41.93625	-87.65266	41.93650	-87.64754	member
9	FBD83C6776EB4EAE	electric_bike	2024-11-18 17:20:38	2024-11-18 17:25:23	Clark St & Elm St	TA1307000039	Clark St & Armitage Ave	13146	41.90297	-87.63128	41.91831	-87.63628	member
10	983509C7904EDFF4	classic_bike	2024-11-01 19:14:52	2024-11-01 19:43:23	Aberdeen St & Jackson Blvd	13157	Desplaines St & Kinzie St	TA1306000003	41.87773	-87.65479	41.88872	-87.64445	member

```
glimpse(all_trips)
```

```
> glimpse(all_trips)
Rows: 5,569,279
Columns: 13
$ ride_id            <chr> "578DDD7CE1771FFA", "78B141C50102ABA6", "1E794CF36394E2D7", "E5DD2CAB58D73F98", "57F9878BC8C765F1", "631261306B4AD…
$ rideable_type      <chr> "classic_bike", "classic_bike", "classic_bike", "classic_bike", "classic_bike", "classic_bike", "electric_bike", "…
$ started_at         <dttm> 2024-11-07 19:21:58, 2024-11-22 14:49:00, 2024-11-08 09:24:00, 2024-11-24 17:51:14, 2024-11-04 14:59:16, 2024-11-…
$ ended_at           <dttm> 2024-11-07 19:28:57, 2024-11-22 14:56:15, 2024-11-08 09:28:33, 2024-11-24 18:05:32, 2024-11-04 15:41:02, 2024-11-…
$ start_station_name <chr> "Walsh Park", "Walsh Park", "Walsh Park", "Clark St & Elm St", "Clark St & Wellington Ave", "Ogden Ave & Congress …
$ start_station_id   <chr> "18067", "18067", "18067", "TA1307000039", "TA1307000136", "13081", "TA1309000010", "TA1307000052", "TA1307000039"…
$ end_station_name   <chr> "Leavitt St & North Ave", "Leavitt St & Armitage Ave", "Damen Ave & Cortland St", "Clark St & Drummond Pl", "Stree…
$ end_station_id     <chr> "TA1308000005", "TA1309000029", "13133", "TA1307000142", "13022", "13157", "13157", "TA1307000136", "13146", "TA13…
$ start_lat          <dbl> 41.91461, 41.91461, 41.91461, 41.90297, 41.93650, 41.87501, 41.87077, 41.93625, 41.90297, 41.87773, 41.85497, 41.9…
$ start_lng          <dbl> -87.66797, -87.66797, -87.66797, -87.63128, -87.64754, -87.67328, -87.62573, -87.65266, -87.63128, -87.65479, -87.…
$ end_lat            <dbl> 41.91053, 41.91781, 41.91598, 41.93125, 41.89228, 41.87773, 41.87773, 41.93650, 41.91831, 41.88872, 41.90954, 41.9…
$ end_lng            <dbl> -87.68231, -87.68244, -87.67733, -87.64434, -87.61204, -87.65479, -87.65479, -87.64754, -87.63628, -87.64445, -87.…
$ member_casual      <chr> "member", "member", "member", "member", "casual", "member", "member", "member", "member", "member", "member", "cas…
>
```

Result: A single dataframe all_trips containing ~5.5 million rows of ride data.

---

2. Inspect and Clean Column Names

   Cleaned all column names for consistency and readability.

```
all_trips <- clean_names(all_trips)
colnames(all_trips)
```

```
> colnames(all_trips)
 [1] "ride_id"            "rideable_type"      "started_at"         "ended_at"           "start_station_name" "start_station_id"
 [7] "end_station_name"   "end_station_id"     "start_lat"          "start_lng"          "end_lat"            "end_lng"
[13] "member_casual"
```

---

3. Remove missing or invalid data

- Filtered out records with missing station name.

```
all_trips <- all_trips %>%
  drop_na(start_stattion_name, end_station_name)
```

---

4. Create New Columns

```
all_trips <- all_trips %>%
  mutate(
      ride_length = as.numeric(difftime(ended_at, started_at, units = "min
s)),
      day_of_week = wday(started_at, label = TRUE),
      month = format(started_at, "%b"),
      year = format(started_at, "%Y")
      )
```

New Columns Created

- **ride_length:** -duration of each time in minutes.
- **day_of_week:** -weekday label (Mon-Sun).
- **month and year:** -for time-based trend analysis.

---

5. Filter Outliers

   Removed extreme ride durations to avoid distortion (less than 1 minute and more than 24 hours).

```
all_trips <- all_trips %>%
 filter(ride_length > 1 & ride_length < 1440)
```

6. Final Dataset Check

```
summary(all_trips$ride_length)
table(all_trips$member_casual)
```

```
> summary(all_trips$ride_length)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
   1.000    5.665    9.661   14.866   16.802 1439.976 
> table(all_trips$member_casual)

 casual  member 
1921888 3499201 
> 
```

7. Saved cleaned dataframe as `all_trips_combined.csv` for further analysis.

```
write_csv(all_trips, "C:/Data Analysis Projects/Google Analytics Capstone Project/Cyclistic/CSVs/all_trips_combined.csv")
```

4. Analyze Phase - Exploring Data And Finding Insights

Purpose: To explore the cleaned dataset, identify key trends and patterns in rider behavior, and answer the core business question.

- How do annual members and casual riders use Cyclistic bike differently?

Tools Used:

- **R Programming** (R Studio)
- Libraries: `tidyverse`, `lubridate`, `ggplot2`

Step 1: Summarize Basic Metrics

```
summary(all_trips)
```

```
> summary(all_trips)
   ride_id          rideable_type       started_at          ended_at         start_station_name start_station_id   end_station_name  
 Length:5421089     Length:5421089     Length:5421089     Length:5421089     Length:5421089     Length:5421089     Length:5421089    
 Class :character   Class :character   Class :character   Class :character   Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character  



 end_station_id        start_lat       start_lng         end_lat         end_lng       member_casual       ride_length       day_of_week       
 Length:5421089     Min.   :41.64   Min.   :-87.89   Min.   :41.49   Min.   :-88.07   Length:5421089     Min.   :   1.000   Length:5421089    
 Class :character   1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66   Class :character   1st Qu.:   5.665   Class :character  
 Mode  :character   Median :41.90   Median :-87.64   Median :41.90   Median :-87.64   Mode  :character   Median :   9.661   Mode  :character  
                    Mean   :41.90   Mean   :-87.65   Mean   :41.90   Mean   :-87.65                      Mean   :  14.866                     
                    3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63                      3rd Qu.:  16.802                     
                    Max.   :42.07   Max.   :-87.52   Max.   :42.21   Max.   :-87.42                      Max.   :1439.976                     
                                                     NA's   :130     NA's   :130                                                              
    month                year     
 Length:5421089     Min.   :2024  
 Class :character   1st Qu.:2025  
 Mode  :character   Median :2025  
                    Mean   :2025  
                    3rd Qu.:2025  
                    Max.   :2025  
```

Now, key summary values

```
summary_stats <- all_trips %>%
  summarise(
    total_rides = n(),
    avg_ride_length = mean(ride_length, na.rm = TRUE),
    median_ride_length = median(ride_length, na.rm = TRUE),
    max_ride_length = max(ride_length, na.rm = TRUE),
    min_ride_length = min(ride_length, na.rm = TRUE)
  )

View(summary_stats)
```

	total_rides	avg_ride_length	median_ride_length	max_ride_length	min_ride_length
1	5421089	14.86576	9.660683	1439.976	1.000017

Step 2: Compare Members VS Casual Riders

```
member_summary <- all_trips %>%
  group_by(member_casual) %>%
  summarise(
    number_of_rides = n(),
    average_duration = mean(ride_length, na.rm = TRUE),
    median_duration = median(ride_length, na.rm = TRUE)
  )

View(member_summary)
```

	member_casual	number_of_rides	average_duration	median_duration
1	casual	1921888	19.93524	11.90434
2	member	3499201	12.08142	8.71330

Key Findings

- Members took nearly twice as many rides as casual riders.
- Casual riders averaged 8 minutes longer per ride than members.

Step 3: Analyze by Day of Week

```
## Weekdays are not ordered, first we will order them in sequence

all_trips$day_of_week <- factor(
  all_trips$day_of_week,
  levels= c("Sun","Mon","Tue","Wed","Thu","Fri","Sat"),
  ordered = TRUE
)
```

```
## Then create another dataframe weekday_summary containing number of ri
des and average
## duration grouped by type of member and day of week

weekday_summary ← all_trips %>%
  group_by(member_casual, day_of_week) %>%
  summarise(
    number_of_rides = n(),
    average_duration = mean(ride_length, na.rm = TRUE)
  )

View(weekday_summary)
```

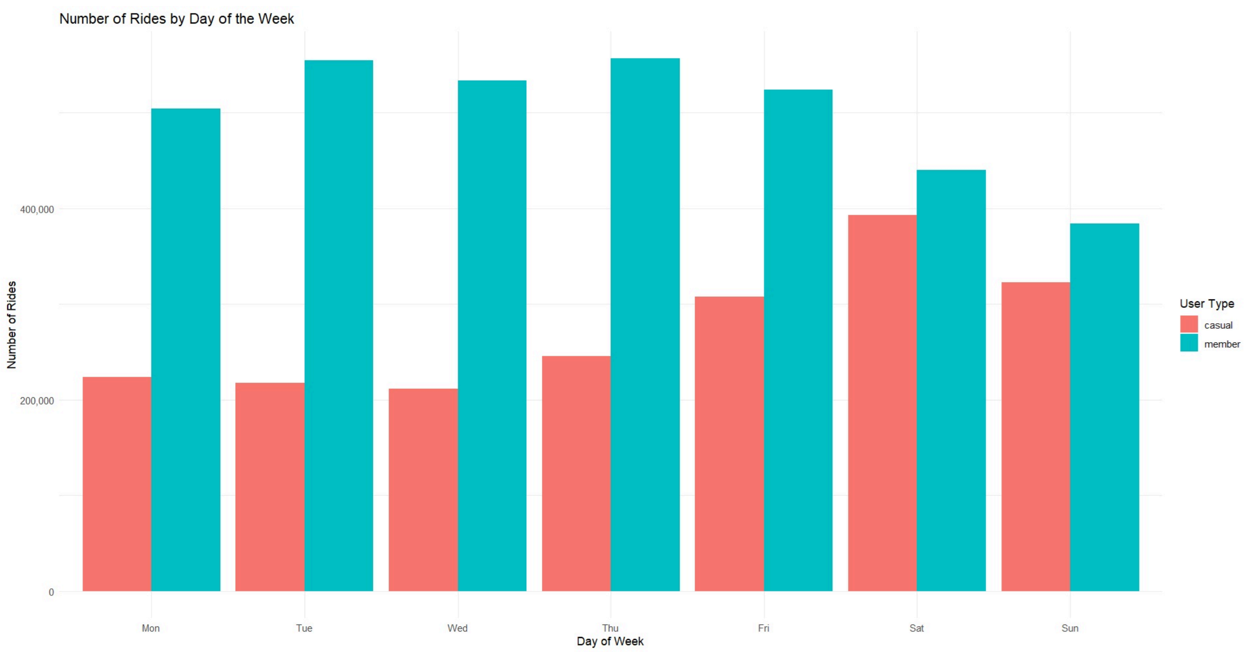
	member_casual	day_of_week	number_of_rides	average_duration
1	casual	Mon	224072	19.65651
2	casual	Tue	217852	17.61196
3	casual	Wed	211407	16.39209
4	casual	Thu	245690	17.54566
5	casual	Fri	307546	19.60518
6	casual	Sat	392878	22.43452
7	casual	Sun	322443	23.11200
8	member	Mon	504255	11.66846
9	member	Tue	555152	11.68352
10	member	Wed	533962	11.51982
11	member	Thu	557255	11.70310
12	member	Fri	524348	12.01234
13	member	Sat	440023	13.18368
14	member	Sun	384206	13.35949

Visualization:

```
library(ggplot2)

ggplot(weekday_summary, aes(x=day_of_week, y=number_of_rides, fill=mem
ber_casual))+
  geom_col(position = "dodge")+
  scale_y_continuous(labels = scales::comma)+     # ← FIXES THE Y-AXIS NUM
BERS
  labs(
```

```
  title="Number of Rides by Day of the Week",
  x = "Day of Week",
  y = "Number of Rides",
  fill = "User Type"
 )+
 theme_minimal()
```



Insights

- Casual riders prefer weekends, suggesting leisure-focused usage.
- Members ride more on weekdays, indicating commute-based behavior.
- Members take over twice as many rides as casual riders, demonstrating high membership satisfaction.

Step 4: Analyze By Month (Seasonal Trend)

```
## Correcting Sequence of Month
all_trips$month ← factor(
```

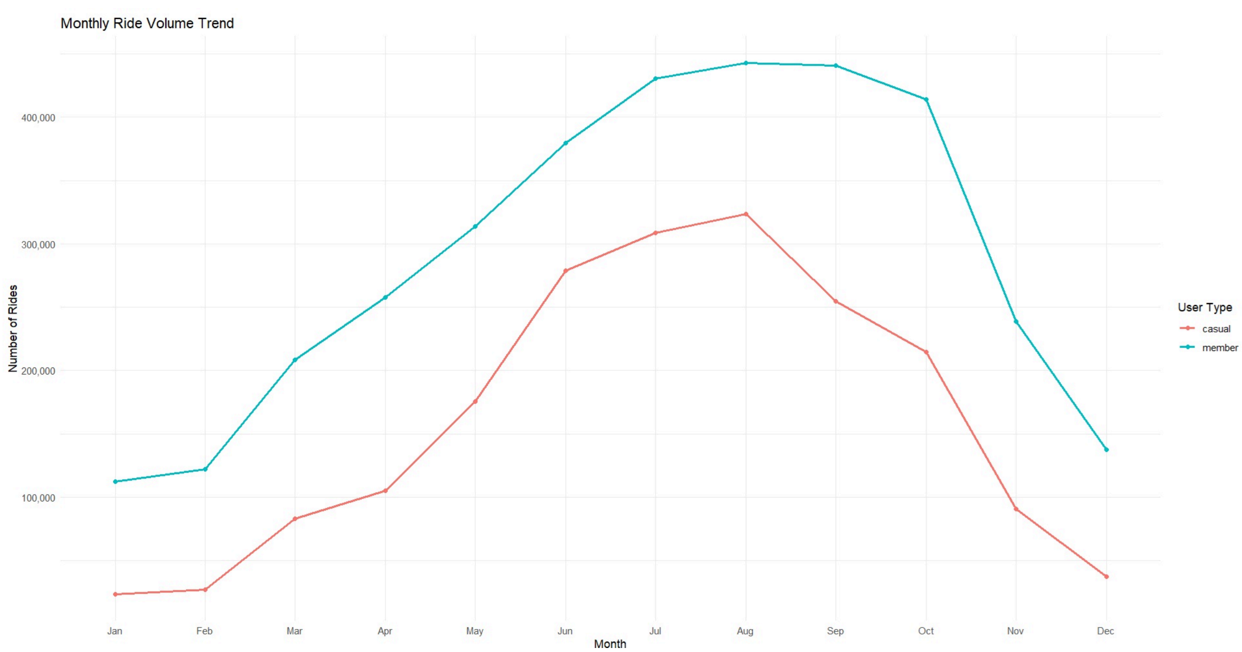
```
  all_trips$month,
  levels = c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec"),
  ordered = TRUE
)

## Analyzing
month_summary <- all_trips %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides = n(), avg_duration = mean(ride_length, na.rm = TRUE))
```

	member_casual	month	number_of_rides	avg_duration
1	casual	Jan	23405	11.973313
2	casual	Feb	27003	12.381231
3	casual	Mar	82864	17.885366
4	casual	Apr	105260	18.655632
5	casual	May	175655	20.831747
6	casual	Jun	278702	21.908679
7	casual	Jul	308446	21.438596
8	casual	Aug	323533	21.710086
9	casual	Sep	254727	19.551890
10	casual	Oct	214343	18.057923
11	casual	Nov	90590	16.373552
12	casual	Dec	37360	13.836939
13	member	Jan	112331	9.979503
14	member	Feb	122097	9.957846
15	member	Mar	208458	11.129125
16	member	Apr	257921	11.273498
17	member	May	313996	11.900135
18	member	Jun	379523	12.846527
19	member	Jul	430394	13.154966
20	member	Aug	443130	13.078290
21	member	Sep	440954	12.616358
22	member	Oct	414055	12.098295
23	member	Nov	238773	10.927697
24	member	Dec	137569	10.610598

Visualizing:

```
ggplot(month_summary, aes(x = month, y = number_of_rides, color = member
_casual, group = member_casual)) +
  geom_line(size = 1) +
  scale_y_continuous(labels = scales::comma)+
  geom_point() +
  labs(
    title = "Monthly Ride Volume Trend",
    x = "Month",
    y = "Number of Rides",
    color = "User Type"
  ) +
  theme_minimal()
```



Insight:

- Ridership increased during summer for both casual riders and members.
- Members show more consistent usage throughout the year.
- It starts to rise in May, peaks in August, then declines from September onward reaching its lowest point in December and January.

Step 5. Analyze by Bike Type

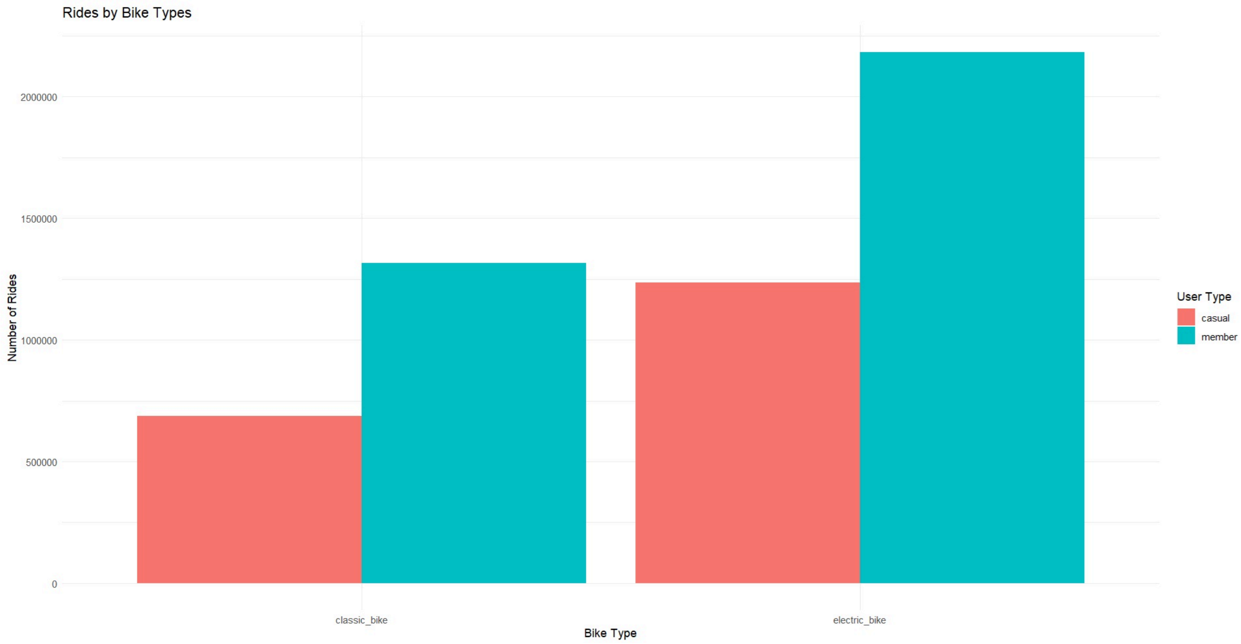
```
bike_summary <- all_trips %>%
  group_by(member_casual, rideable_type) %>%
  summarise(number_of_rides = n(), avg_duration = mean(ride_length, na.rm =
TRUE))

View(bike_summary)
```

	member_casual	rideable_type	number_of_rides	avg_duration
1	casual	classic_bike	685823	28.94343
2	casual	electric_bike	1236065	14.93710
3	member	classic_bike	1314996	13.31668
4	member	electric_bike	2184205	11.33774

Visualization:

```
ggplot(bike_summary, aes(x=rideable_type, y=number_of_rides, fill = member
_casual))+
  geom_col(position = "dodge")+
  labs(
    title = "Rides by Bike Types",
    x = "Bike Type",
    y = "Number of Rides",
    fill = "User Type"
  )+
  theme_minimal()
```



Insights:

- Electric bike is most favored by both type of riders.

Step 6: Insights of Analyze Phase

Observations	Interpretation
Members took nearly twice as many rides as casual riders.	Indicates that annual members have higher engagement and rely heavily on the service for regular transportation.
Casual riders averaged 8 minutes longer per ride than members.	Shows that casual riders use the service for leisure or recreational trips rather than quick point-to-point travel.
Casual riders prefer weekends, suggesting leisure-focused usage.	Weekend spikes indicate tourism and recreational behavior.
Members ride more on weekdays, indicating commute-based behavior.	Suggests members depend on Cyclistic for routine daily travel, likely commuting to work or school.

Observations	Interpretation
Members take over twice as many rides as casual riders.	High ride frequency shows strong satisfaction with the membership model and consistent product usage.
Ridership increased during summer for both casual riders and members.	Seasonal demand trend — warmer months encourage more outdoor activity, increasing overall bike usage.
Members show more consistent usage throughout the year.	Members provide stable, predictable revenue, reinforcing the value of annual subscriptions.
Usage rises in May, peaks in August, then declines through winter.	Clear seasonal cycle — high demand during summer; low demand during cold months.
Electric bikes are favored by both rider types.	Indicates strong user preference for convenience and lower physical effort — opportunity to expand electric bike fleet.

---

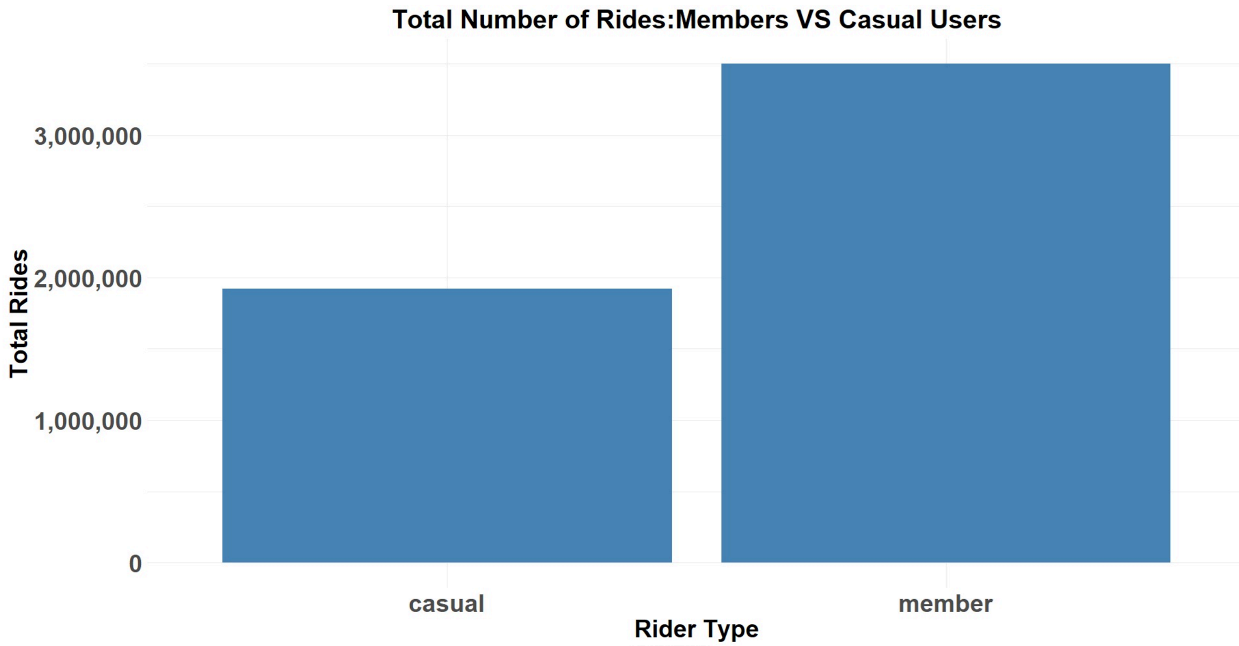
5. Share Phase - Communicating Insights Effectively

Purpose: To present insights visually and clearly so stakeholders understand how members and casual riders behave differently and what actions Cyclistic should take.

Insight of analysis is presented for following key questions.

1. Total numbers of Rides by Rider Type.
2. Average ride duration for each Rider Type.
3. Ride frequency by day of week.
4. Monthly seasonal trend.
5. Bike type preference.

1. **Total numbers of Rides by Rider Type.**

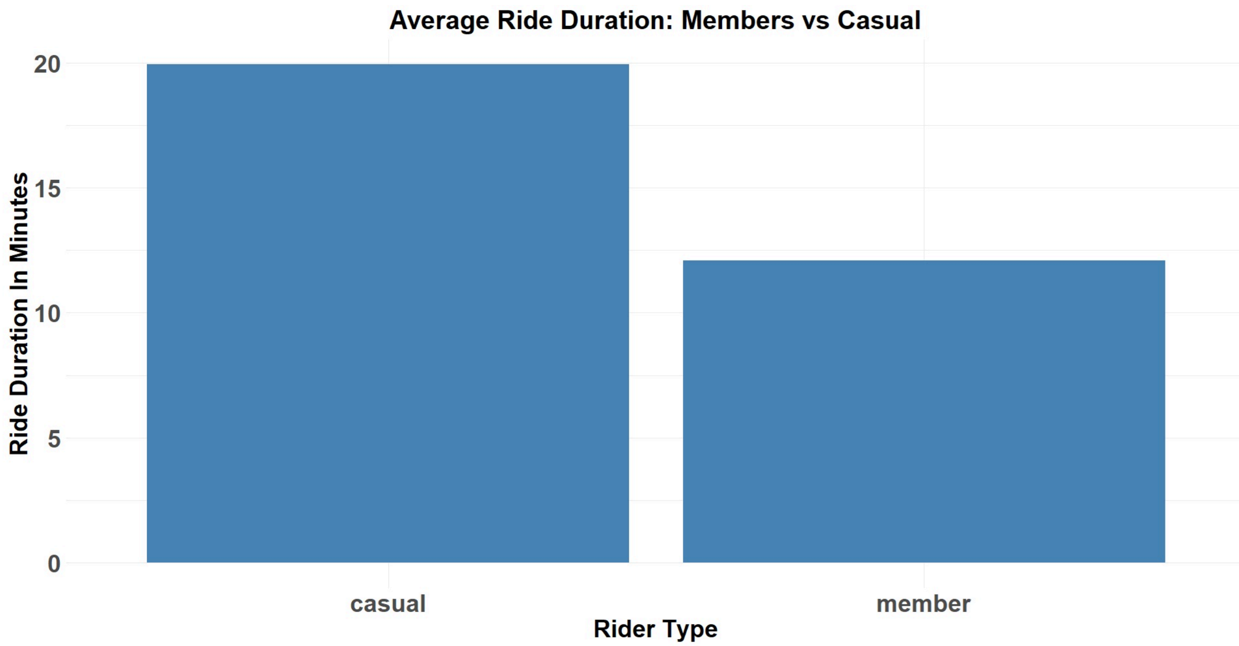


Insight:

Annual members take significantly more rides than casual riders, indicating higher engagement and reliance on Cyclistic for routine travel.

---

2. **Average ride duration for each Rider Type.**

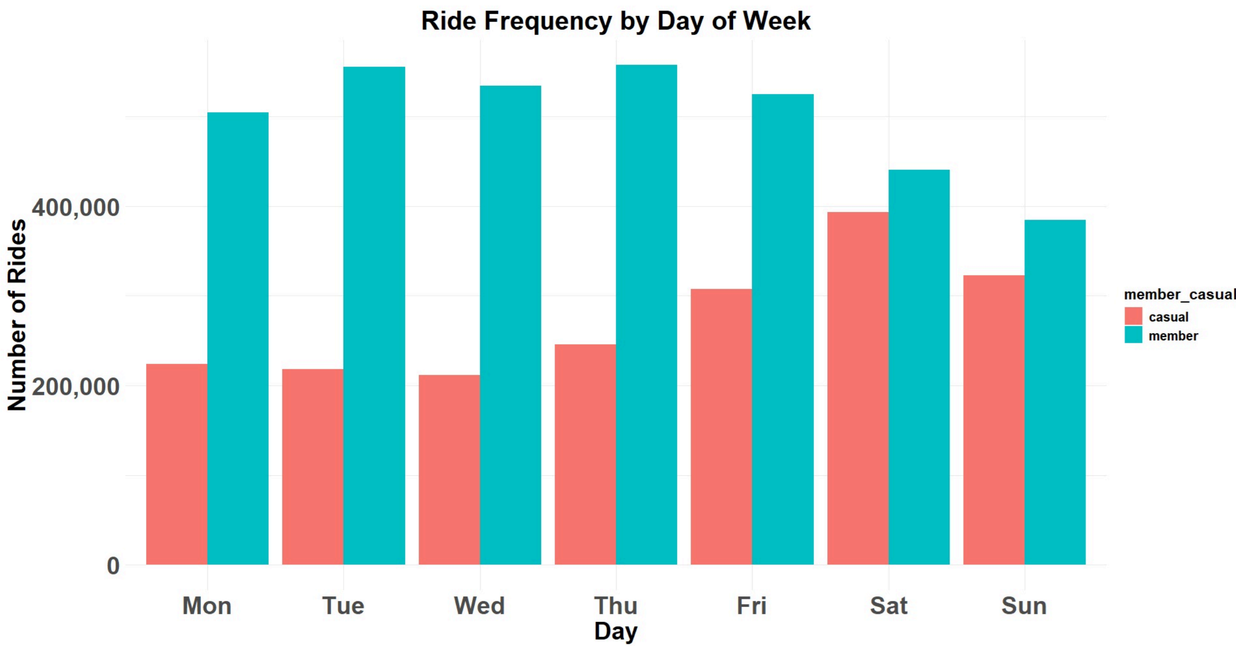


Insight:

Casual riders take much longer trips compared to members, suggesting leisure-focused activity rather than commuting.

---

3. **Ride Frequency by Day of Week**

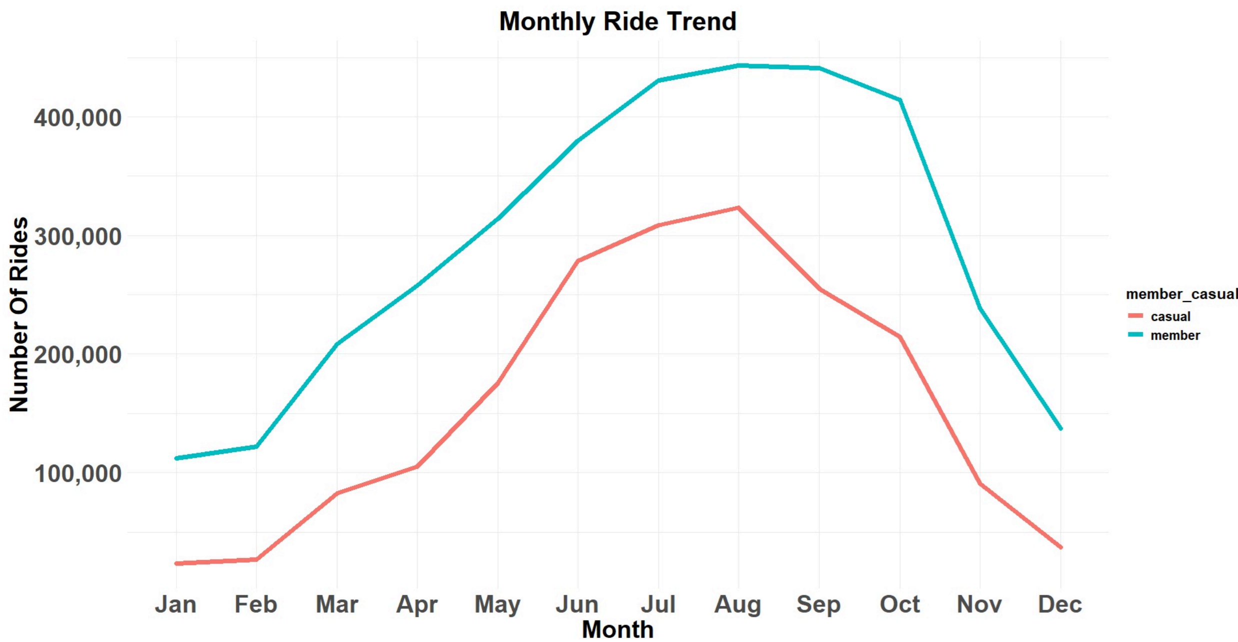


Insight:

- Members ride mainly on weekdays, consistent with commute behavior.
- Casual riders peak on weekends, indicating recreational usage.

---

4. **Monthly seasonal trend**

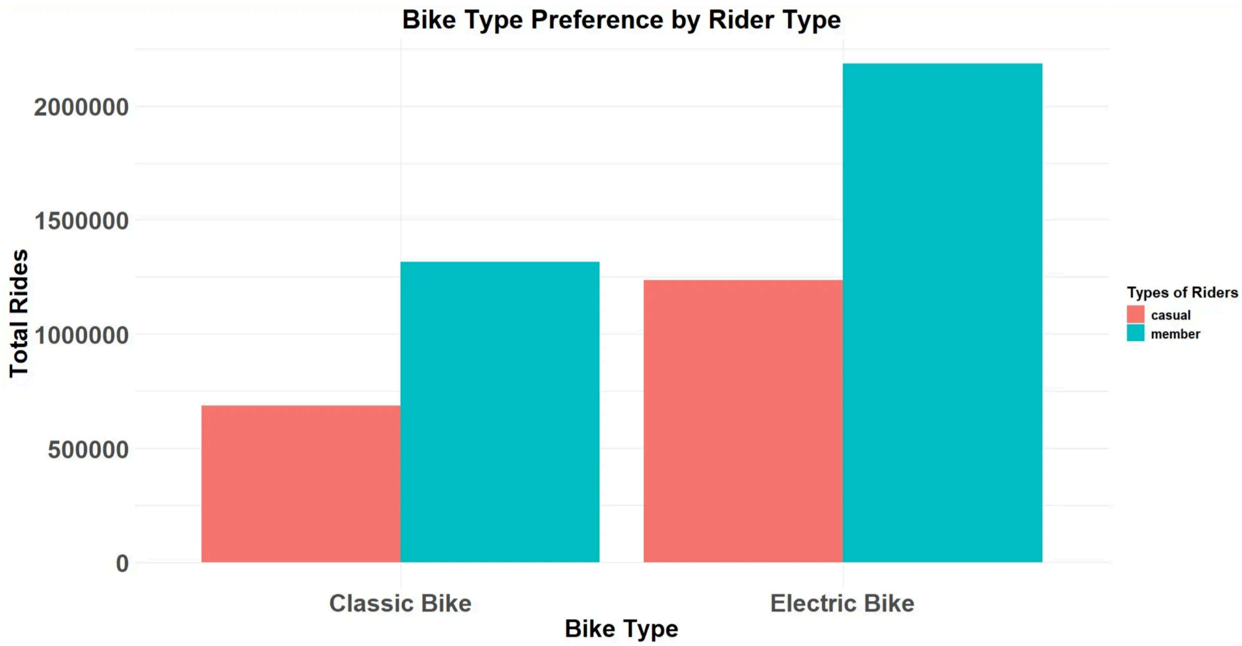


Insight:

- Ridership increases significantly during summer months for both groups, showing strong seasonal patterns.
- Members, however, maintain more consistent usage year-round.

---

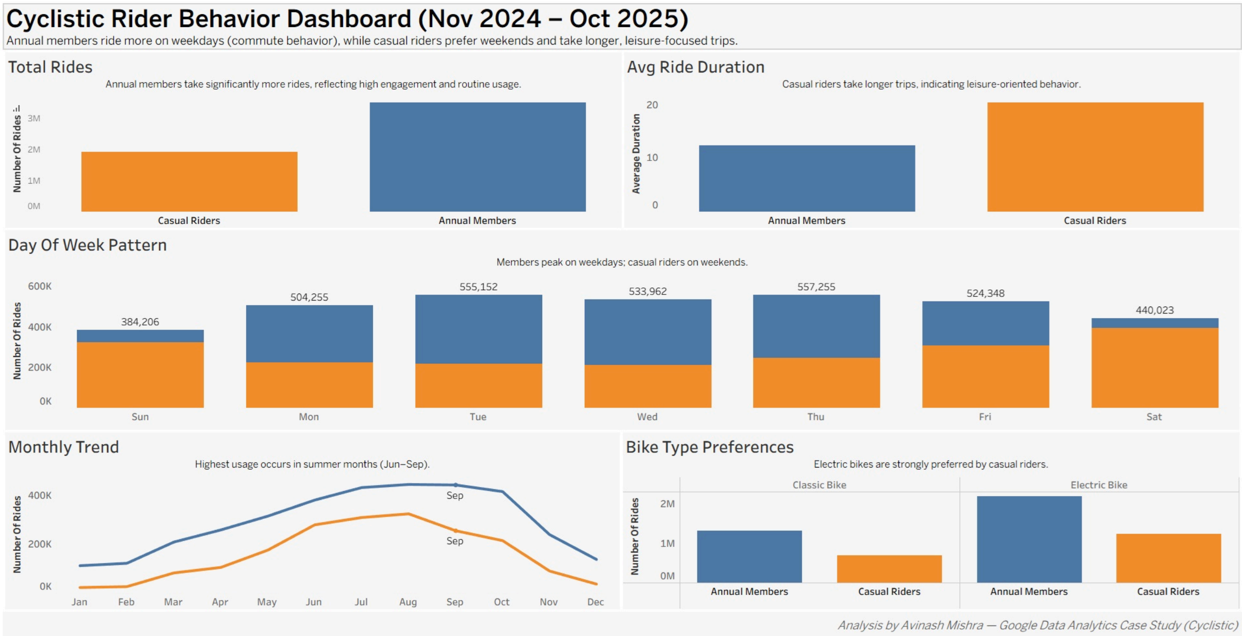
5. **Bike type preference**



Insight:

Electric bikes are popular among both groups, but especially casual riders — likely due to convenience and reduced effort for longer leisure rides.

Tableau Dashboard:


Cyclistic Rider Behavior Dashboard (Nov 2024 – Oct 2025)
Annual members ride more on weekdays (commute behavior), while casual riders prefer weekends and take longer, leisure-focused trips.
Total Rides
Annual members take significantly more rides, reflecting high engagement and routine usage.
Number Of Rides
3M
2M
1M
0M
Casual Riders
Annual Members
Avg Ride Duration
Casual riders take longer trips, indicating leisure-oriented behavior.
Average Duration
20
10
0
Annual Members
Casual Riders
Day Of Week Pattern
Members peak on weekdays; casual riders on weekends.
Number Of Rides
600K
400K
200K
0K
384,206
504,255
555,152
533,962
557,255
524,348
440,023
Sun
Mon
Tue
Wed
Thu
Fri
Sat
Monthly Trend
Highest usage occurs in summer months (Jun–Sep).
Number Of Rides
400K
200K
0K
Sep
Sep
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Bike Type Preferences
Electric bikes are strongly preferred by casual riders.
Classic Bike
Electric Bike
Number Of Rides
2M
1M
0M
Annual Members
Casual Riders
Annual Members
Casual Riders
Analysis by Avinash Mishra — Google Data Analytics Case Study (Cyclistic)