How Does a Bike-Share Navigate Speedy Success?

Customer: Lily Moreno

Company: Cyclistic

Industry: Bike-share company

Author: Omar Dissouki

Publish Date: July 20th 2024

Contents

The Story of Lily Moreno

Cyclistic is a bike-share company that is based in Chicago. Riders use the service for leisure and commuting. The company operates over 5,800 bicycles and 600 docking stations.

Lily Moreno, the director of marketing at Cyclistic, has the goal of converting casual riders into annual members to ensure steady growth. As a junior analyst, I have been tasked with understanding user behavior with the goal of developing targeted marketing strategies.

Steps:

Data Collection

Data Source

Historical bike trip data will be gathered from Divvy Trip Data for the period from July 2022 to June 2023.

Dataset Variables

Variable	Description
Ride_id	Unique identifier for each ride in the dataset
Rideable_type	Type of bike, either classic or electric
Started_at	Time when ride started
Ended_at	Time when ride ended
Start_station_name	Name of the station where ride started
Start_station_id	Unique identifier for each station
End_station_name	Name of the station where ride ended
End_station_id	Unique identifier for each station
Start_lat	Latitude of the ride starting location
Start_lng	Longitude of the ride starting location
End_lat	Latitude of the ride ending location
End_lng	Longitude of the ride ending location
Member_casual	Type of customer, either subscribed or casual

Data Cleaning

Data cleaning involved using Excel and R to standardize data formats, correct inconsistencies, and remov null and duplicate entries. Average ride length and number of rides taken each day of the week for each month was calculated using Excel. The data was further refined in R, where I consolidated monthly files, converted date fields, categorized times of day, and filtered out erroneous data.

Data Analysis

Seasonality and pattern analysis

Usage frequency trends

- July 2022 - December 2022: Decrease in usage during colder months and school year starts. Casual riders peak on Fridays and Saturdays
- January 2023 - June 2023: Increase in bike rides as the year progressed into warmer months, particularly on Sundays. Peaks around summer time.

Ride Duration Trends

- July 2022 - December 2022: Members average 12 minutes, casual riders 21 minutes. Longer rides on weekends.
- January 2023 - June 2023: Consistent patterns, with casual riders maintaining longer durations.

Overall analysis

July 2022 to June 2023

- Decrease in usage from July to December then an increase in usage from January to June 2023.
- Sundays preferred for both groups.
- Casual riders' peak usage hours are in the afternoons and evenings, while members show more consistent usage throughout the day.

Key takeaways

Rider Patterns:

- Casual riders primarily use bikes on weekends, indicating a preference for leisurely biking activities.
- Members tend to ride on weekdays, suggesting a regular commuting behavior.

Ride Length:

- Casual riders have longer average ride durations, approximately 21 minutes per ride.
- Members average around 12 minutes per ride.

Peak Times:

- Casual riders' peak usage hours are in the afternoons and evenings, likely reflecting recreational use.
- Members show more consistent usage throughout the day, possibly due to commuting needs.

Seasonality:

- During the summer months, both casual riders and members significantly increase their ride frequency. Conversely, there is a decline in usage during the winter months due to the harsh Chicago weather.

Station Popularity:

- Casual riders prefer certain bike stations, particularly those located near tourist attractions and recreational areas such as the Navy Pier, Millennium Park and The Lakefront Trail.

- Members prefer using the bikes near locations such as W Hubbard St which has multiple business offices nearby and the Riverwalk area due to the presence of local restaurants, bars, and cafes. These spots offer opportunities for dining and socializing.

Recommendations

1. Developing promotions and advertising campaigns using digital media specifically aimed at casual riders especially during summer and peak times, emphasizing:
    a. Cost savings, by sending personalized emails or notifications to casual riders showing how much they would have, depending on their usage, if they had an annual membership.
    b. Convenience of having the annual membership, i.e. being able to prebook bikes during peak hours.
    c. Exclusive benefits of annual memberships, such as free rides, access to special events and Cyclistic merch.
2. Providing a free trial to casual riders allowing them to experience the benefits of membership, while implementing a referral program where existing members can earn rewards for bringing in new annual members.
3. Improving user experience by:
    a. Including more electric bikes in high demand areas, such as the Streeter Dr & Grand Ave area.
    b. Removing unused docked bikes which accounts for 2.4% of the total number of trips.

Lessons Learned

- Data analytics can be leveraged to make targeted marketing more effective.
- Connecting the dots by analyzing the user behavior is an essential part in developing a successful marketing strategy.
- Regularly studying user behavior would be useful to produce more accurate insights that can be used to develop a more successful business model.

Conclusion

To conclude, this study showed how data analytics can be used to understand customer behavior and develop effective marketing strategies. By identifying differences between the usage of casual riders and annual members, the marketing team would be able to implement successful targeted marketing that would result in an increase in membership and overall growth. The lessons learned from this analysis would help me in my future projects. The steps taken in this project can be applied in most user behavior studies to improve user experience, which would result in a more successful business model.

In depth analysis

July 2022

Usage Frequency

- Members: Trip frequency is nearly consistent throughout the week. Mondays has the lowest number of trips.

- Casual Riders: Casual riders had most of their trips on Saturdays and Sundays.

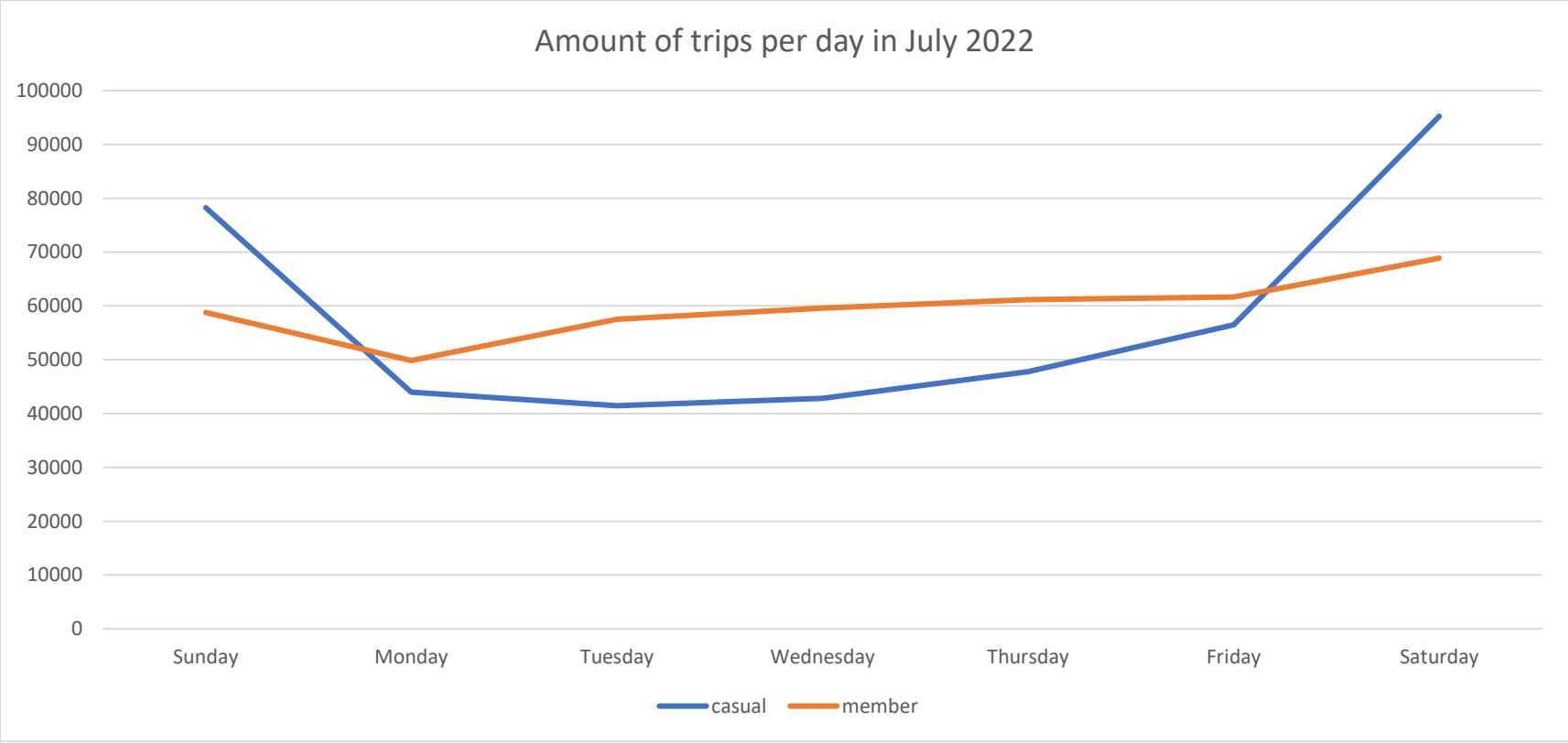

Figure 1. Line comparison Casual vs Member rides per day in July 2022

Ride Duration

Both Members and Casuals had most trips on Saturday and Sunday

- Members: The average ride duration for members was less than 15 minutes, nearly half the length of casual riders' rides.
- Casual Riders: The average ride duration was over 20 minutes every day of the week.

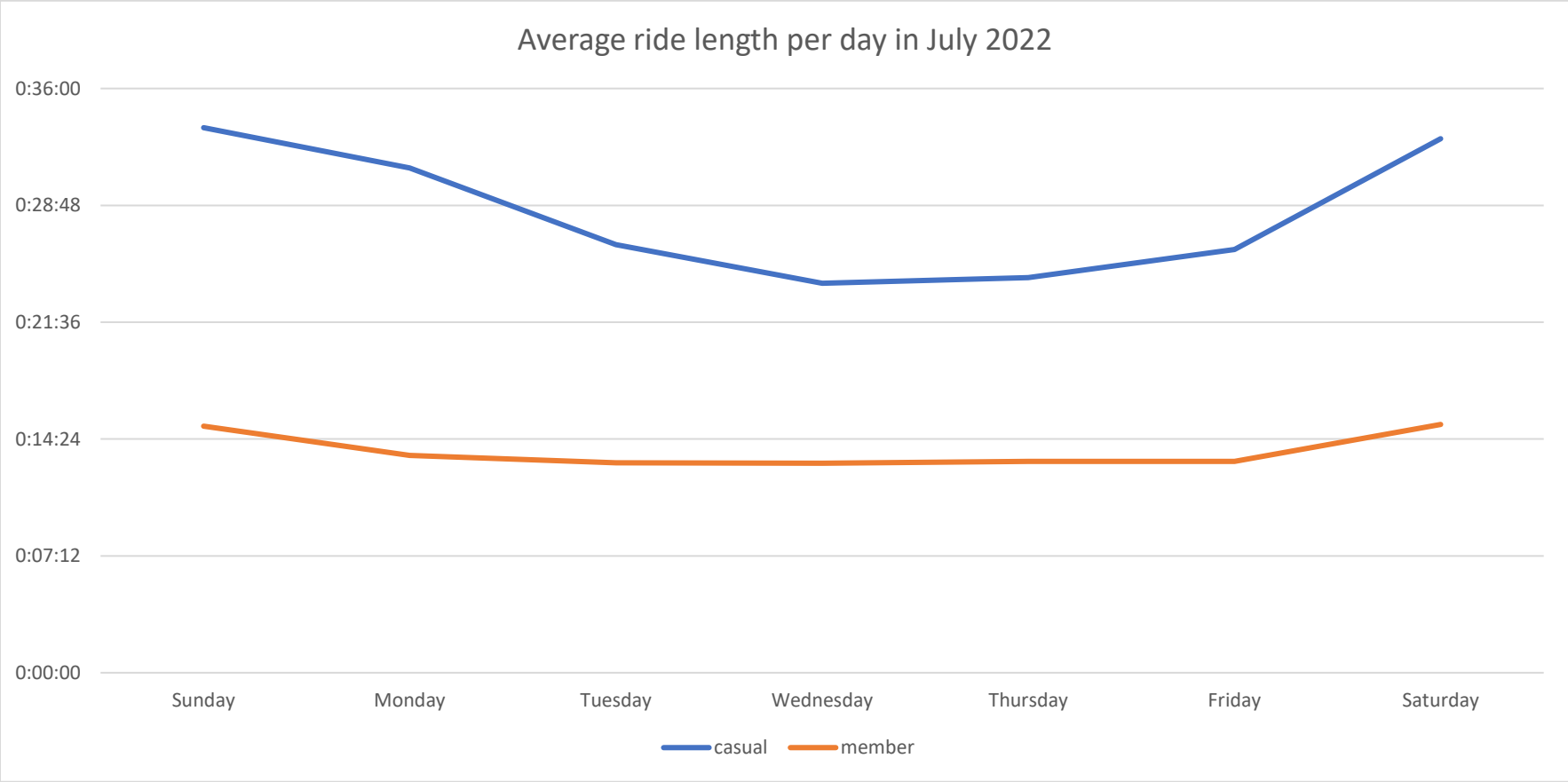

Figure 2. Line comparison Casual vs Member ride length per day in July 2022

Rideable types

Electric and classic bikes were equally popular and had a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used with a maximum of around 10,000 trips on Saturdays and Sundays.

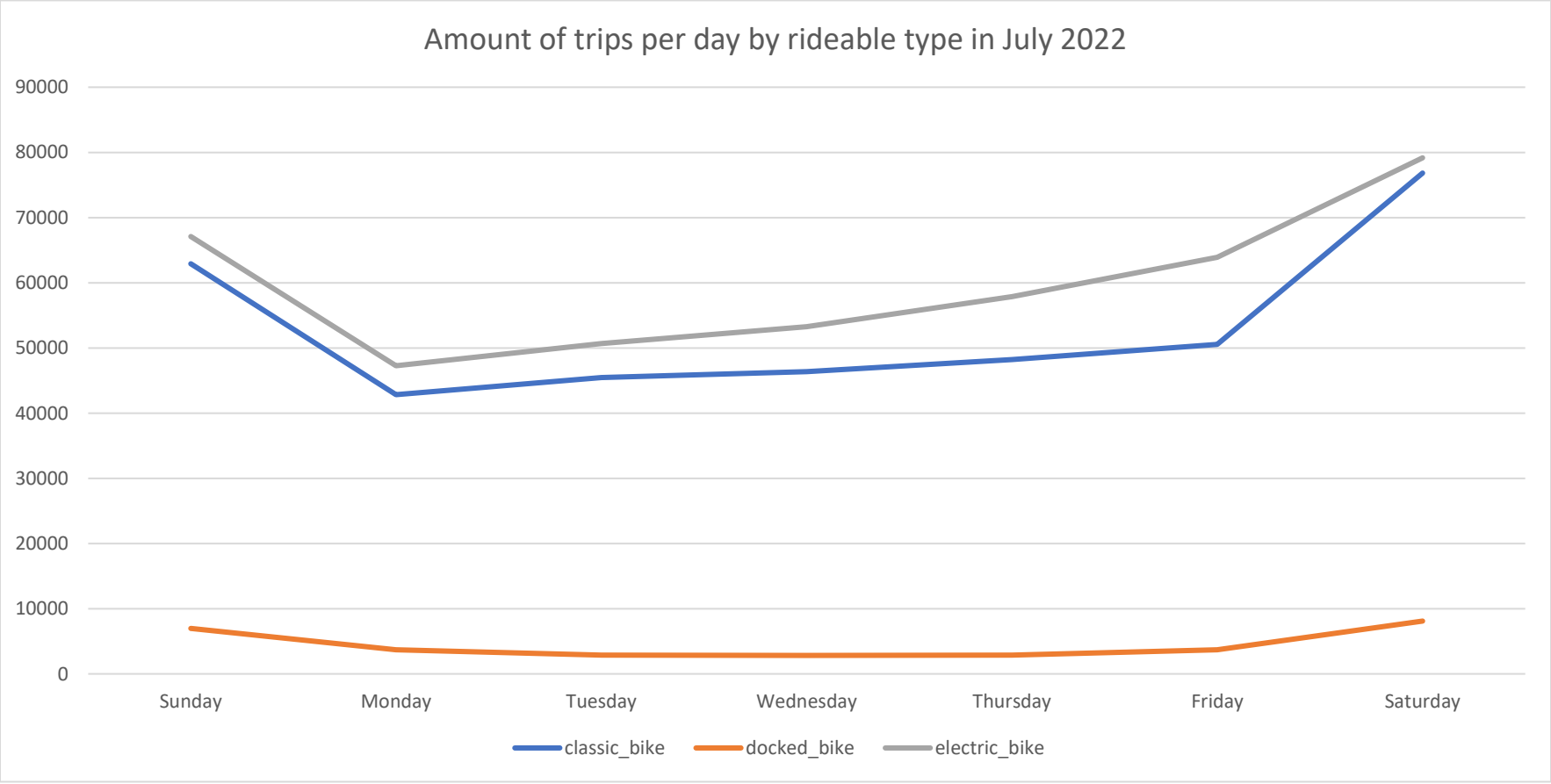

Figure 3. Line comparison number of rides per day by rideable type in July 2022

August 2022

Usage Frequency

- Members: They had more trips during the week compared to last month, Peak usage was on Tuesdays and Wednesdays.

- Casual Riders: They were consistent during the week and peaked on Fridays and Saturdays.

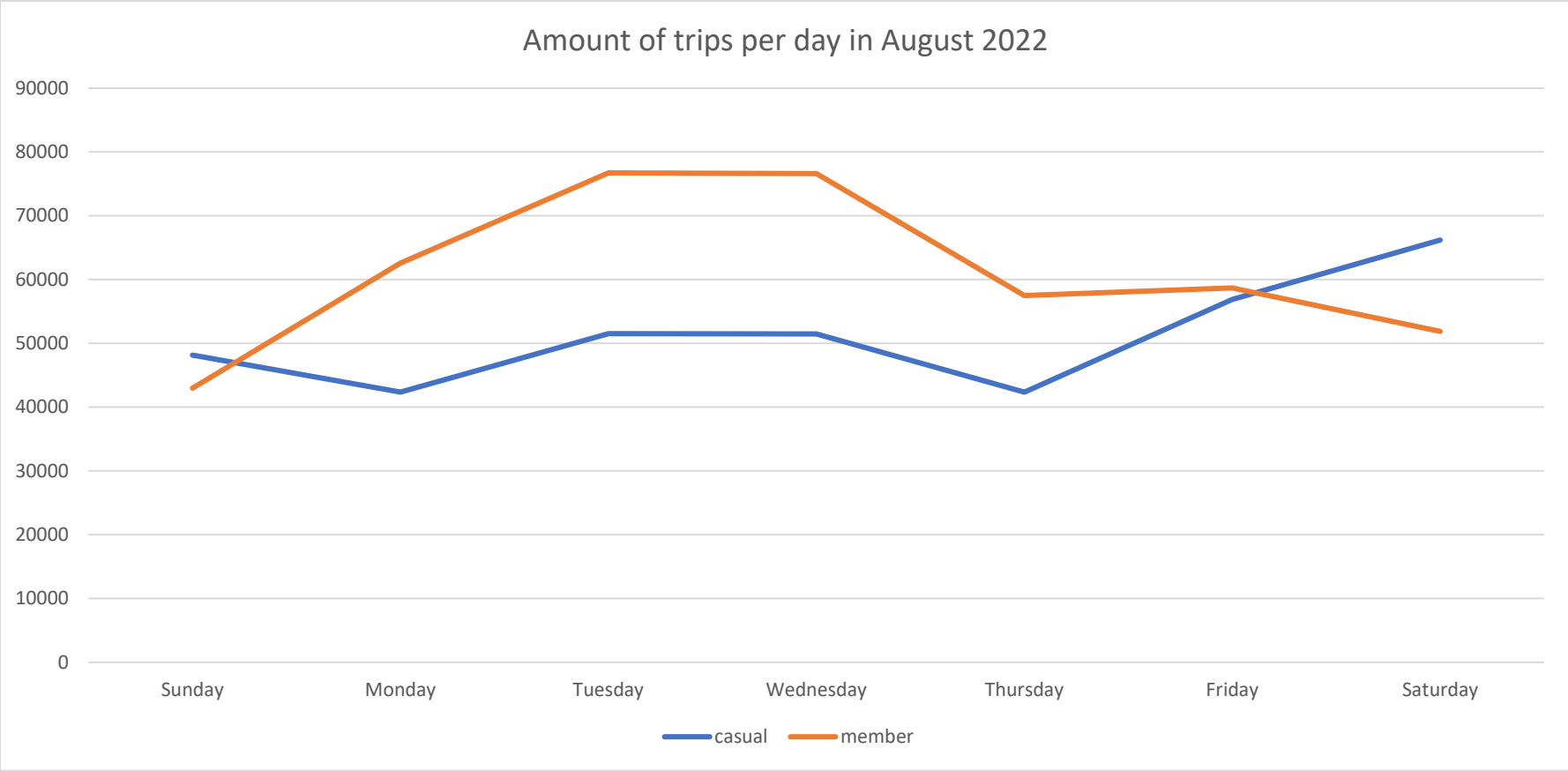

Figure 4. Line comparison Casual vs Member rides per day in August 2022

Ride Duration

- Members: On Saturdays and Sundays, members had their longest rides. However, even on these days, their average ride duration remained less than 15 minutes.
- Casual Riders: Casual riders had rides over 30 minutes on Fridays, Saturdays, and Sundays, suggesting that they use the bikes for leisure activities during weekends.

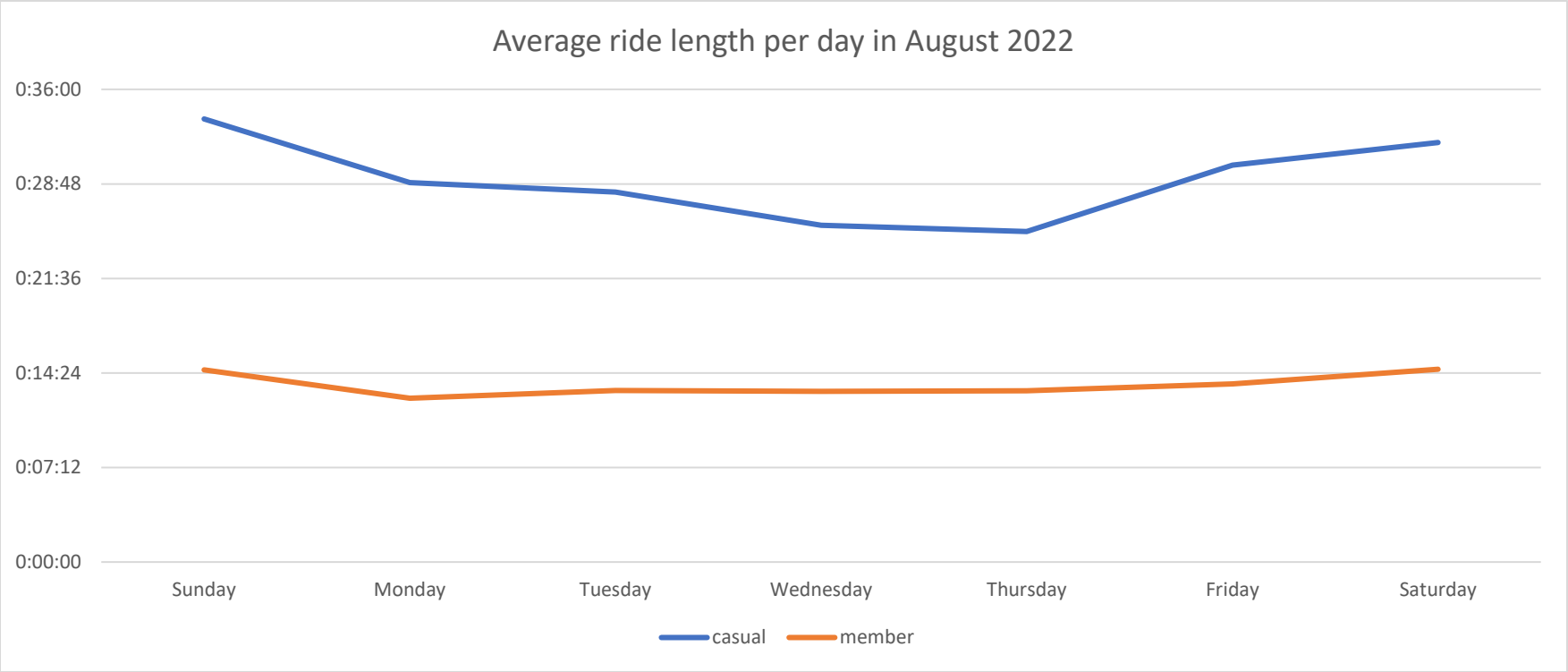

Figure 5. Line comparison Casual vs Member ride length per day in August 2022

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used with a maximum of around 5,000 trips each day of the week throughout the month.

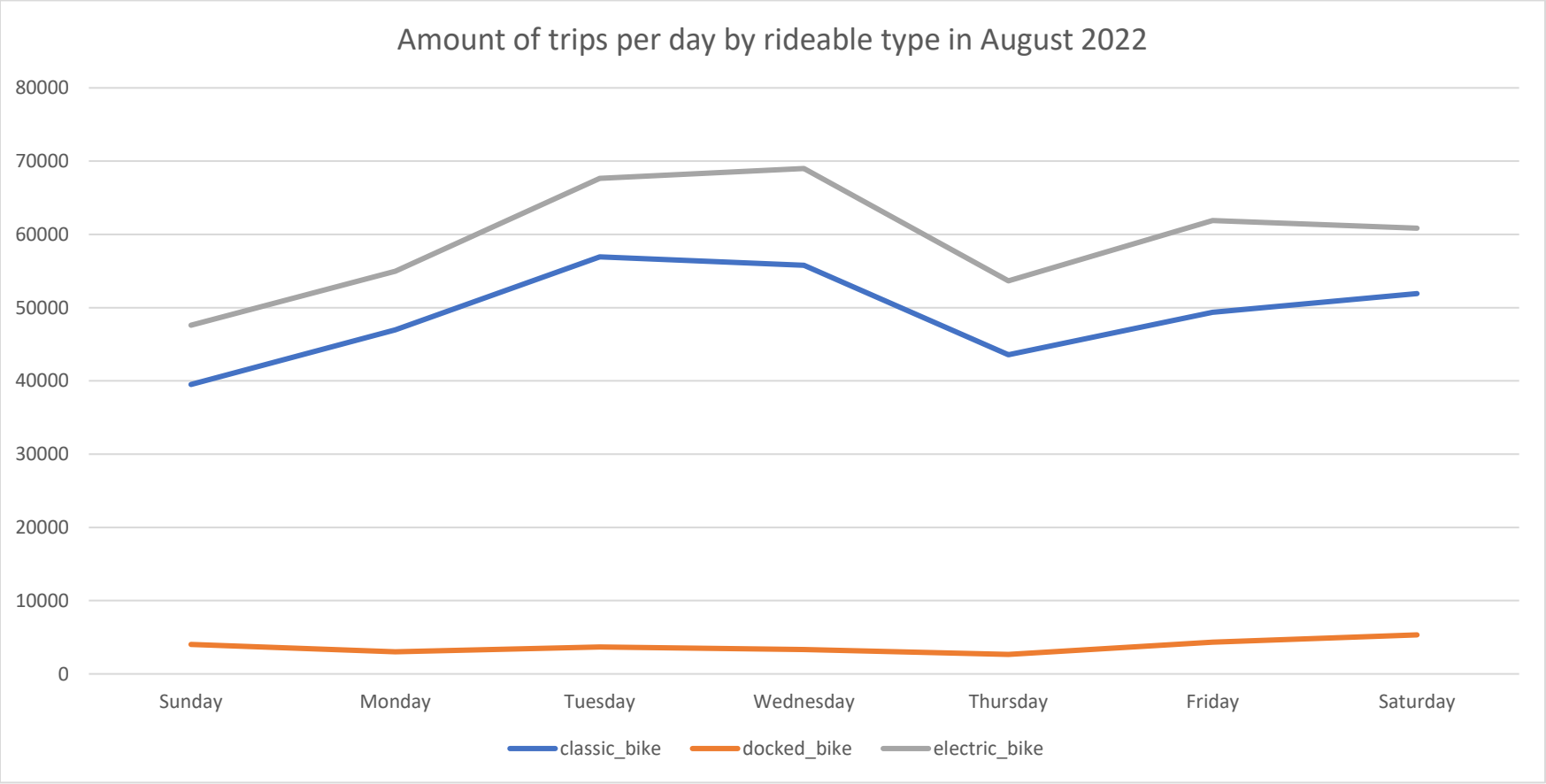

Figure 6. Line comparison number of rides per day by rideable type in August 2022

September 2022

Usage Frequency

- Members: Significant decrease in the number of trips throughout the whole week compared to the previous month. Peak usage was on Thursday.
- Casual Riders: Casual riders reduced their usage during the week, with about 70% decrease on Tuesday and a 55% decrease on Wednesday. However, Friday and Saturday usage remained relatively stable.

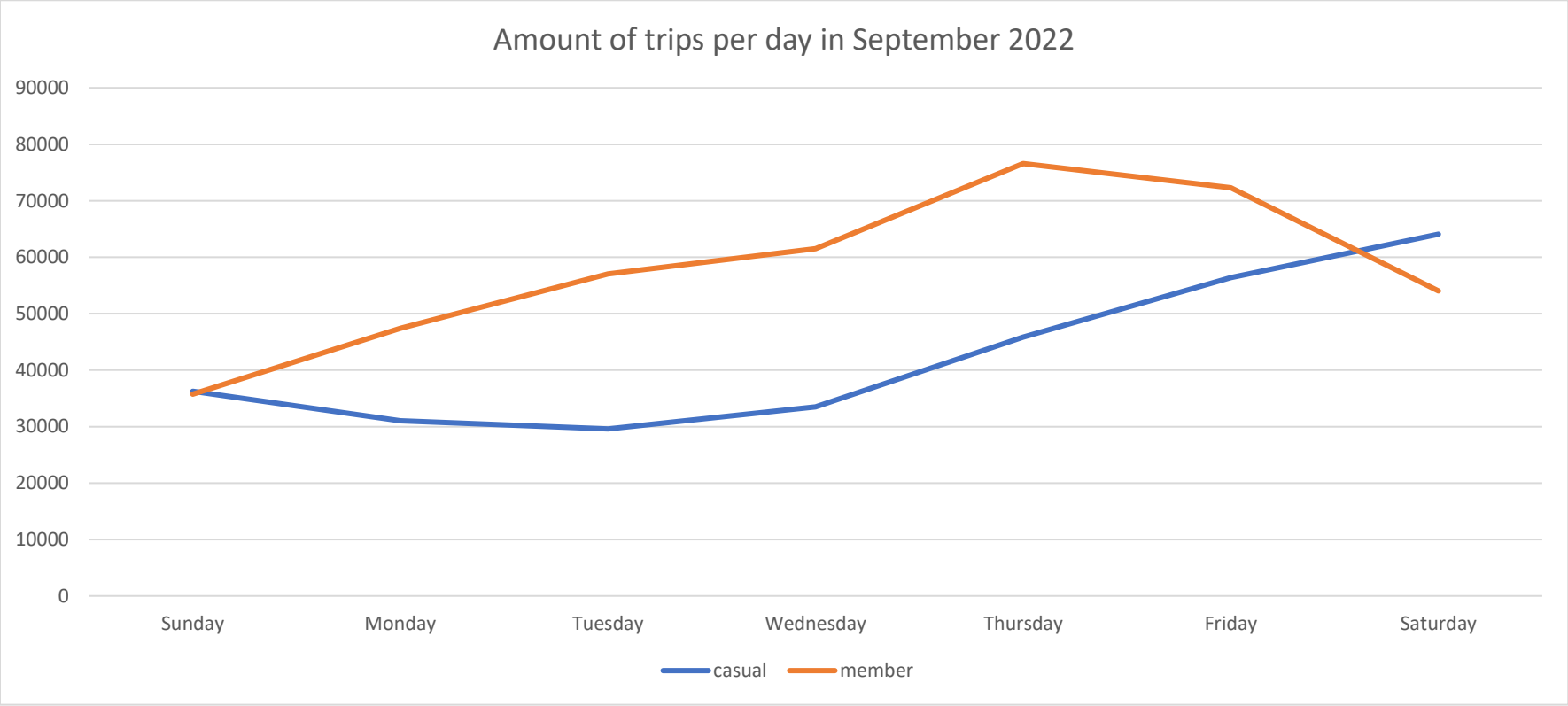

Figure 7. Line comparison Casual vs Member rides per day in September 2022

Ride Duration

- Members: On Saturdays and Sundays, members had rides shorter than 15 minutes.
- Casual Riders: Casual riders had trips over 30 minutes on Saturdays, and Sundays. However, trips during the week were between 21-28 minutes.

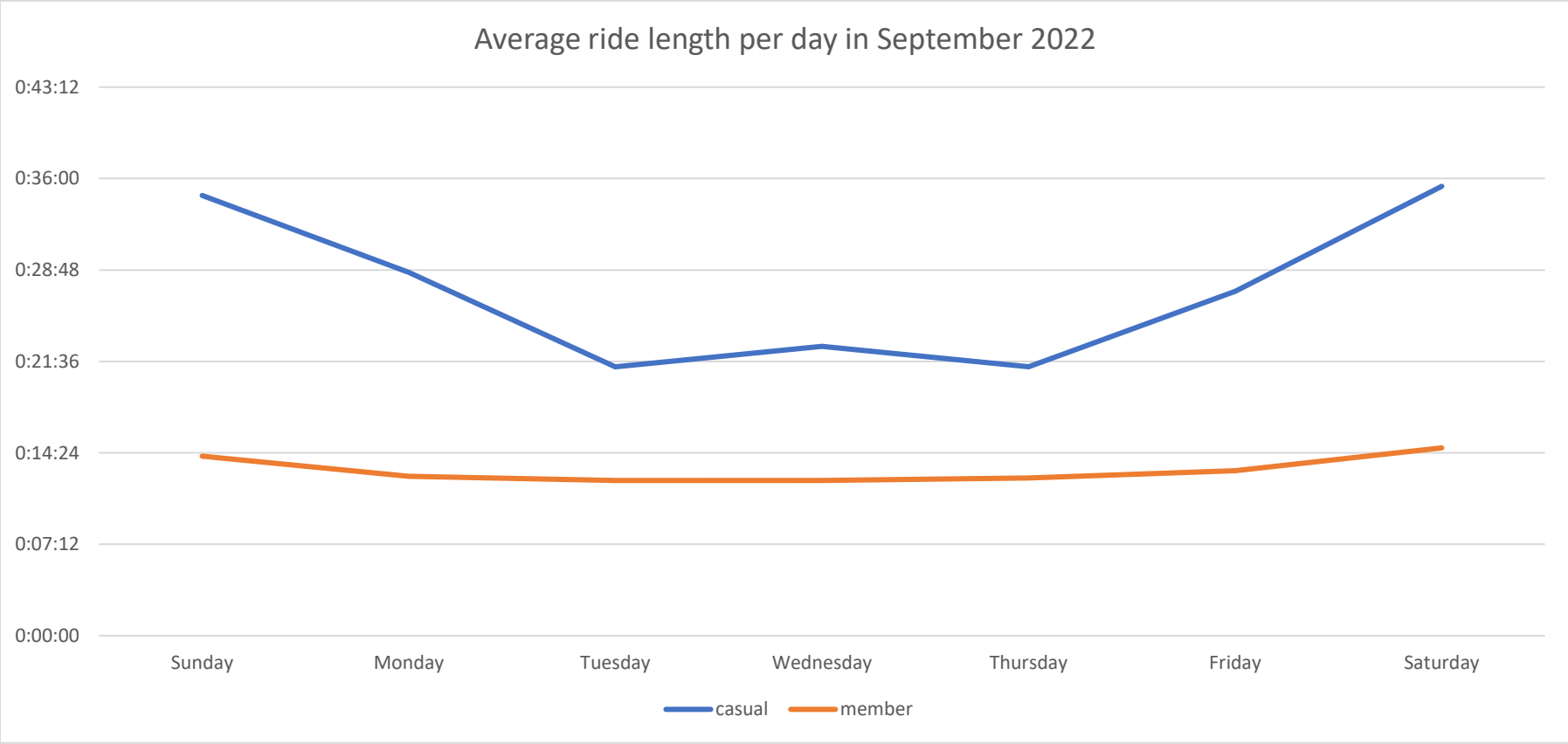

Figure 8. Line comparison Casual vs Member ride length per day in September 2022

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used with a maximum of around 5,000 trips on Saturdays throughout the month.

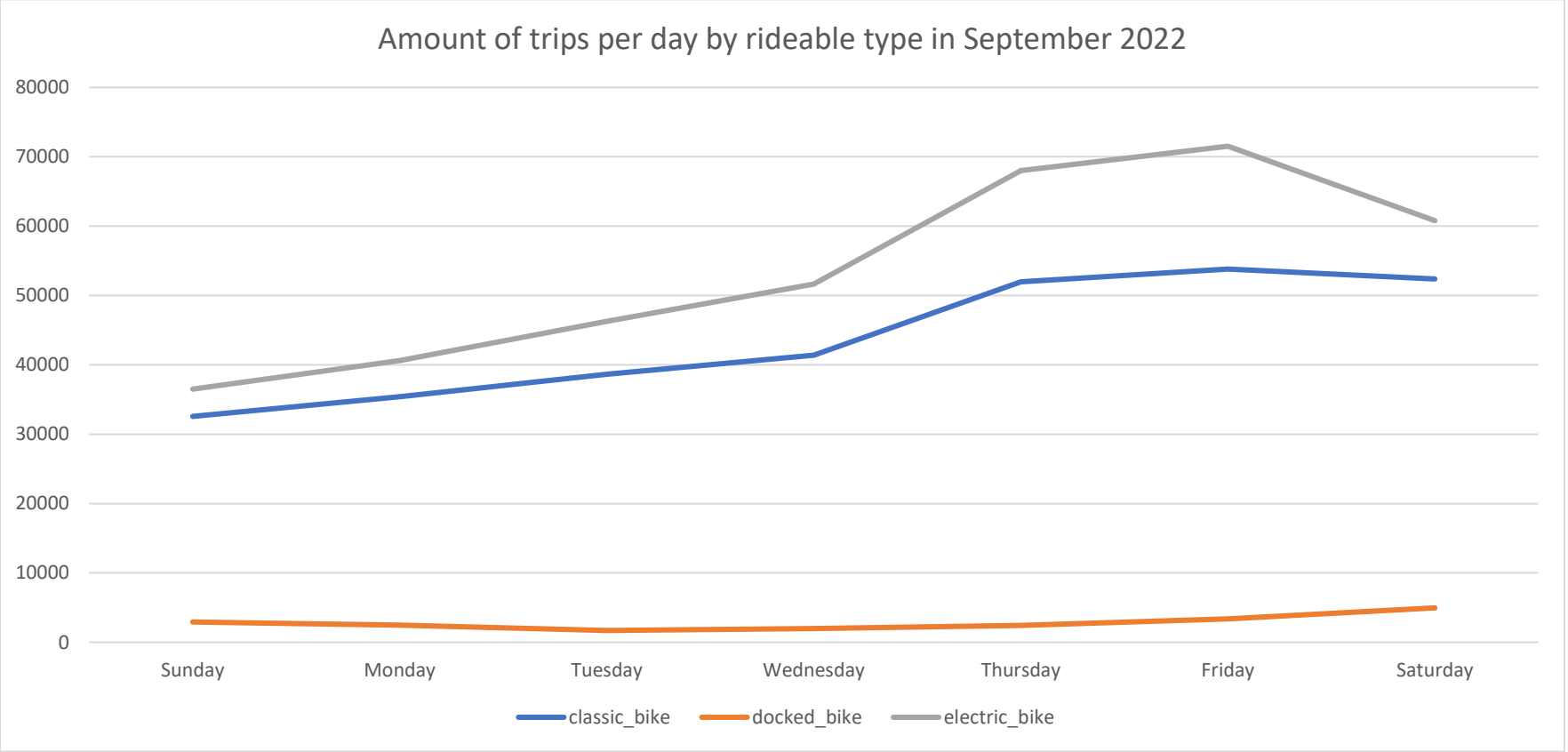

Figure 9. Line comparison number of rides per day by rideable type in September 2022

October 2022

Usage Frequency

- Members: There was a change in usage patterns, where members used the bikes during the week, especially Tuesdays and Wednesdays, more than weekends
- Casual Riders: Casual riders reduced their bike usage, particularly during the week. This decrease might be attributed to the onset of the cold weather in fall.

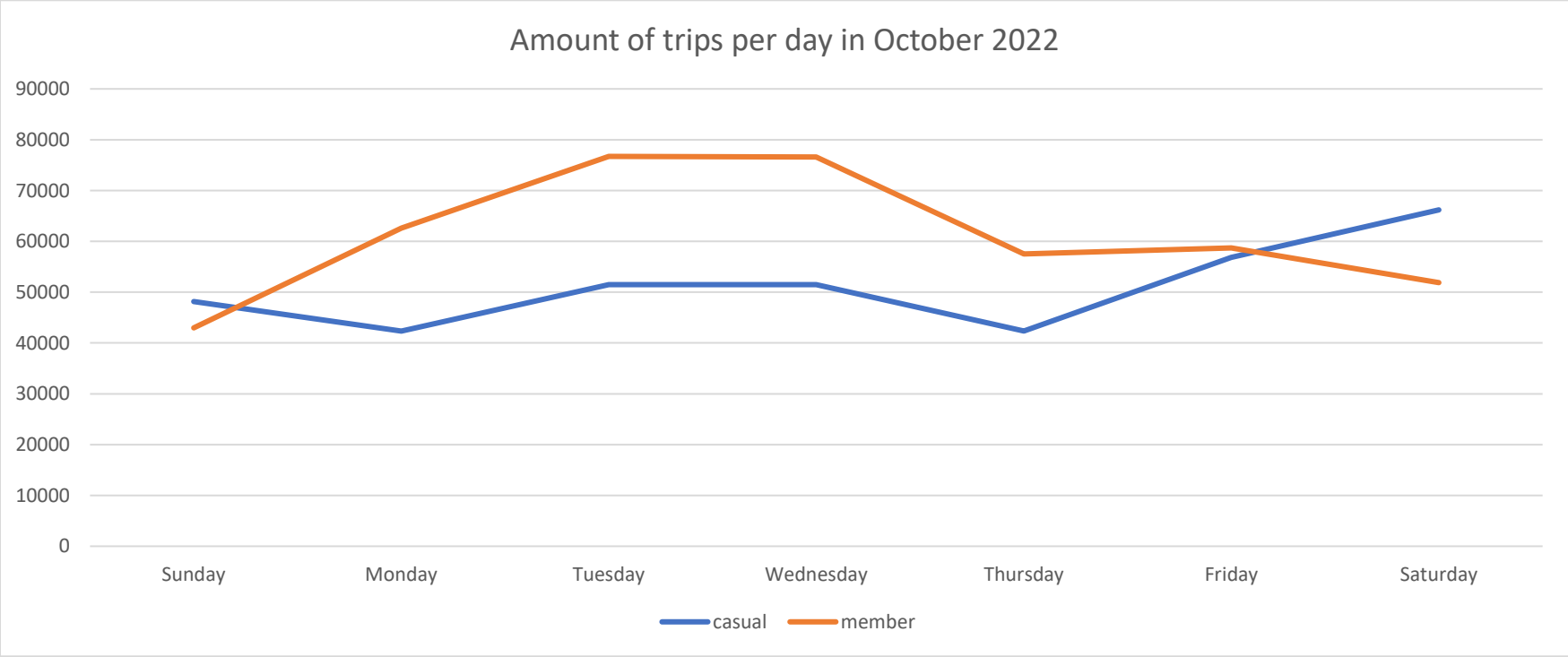

Figure 10. Line comparison Casual vs Member rides per day in October 2022

Ride Duration

- Members: The ride length for members remained consistent.

- Casual Riders: Casual riders continued to take longer trips, around the double, compared to members.

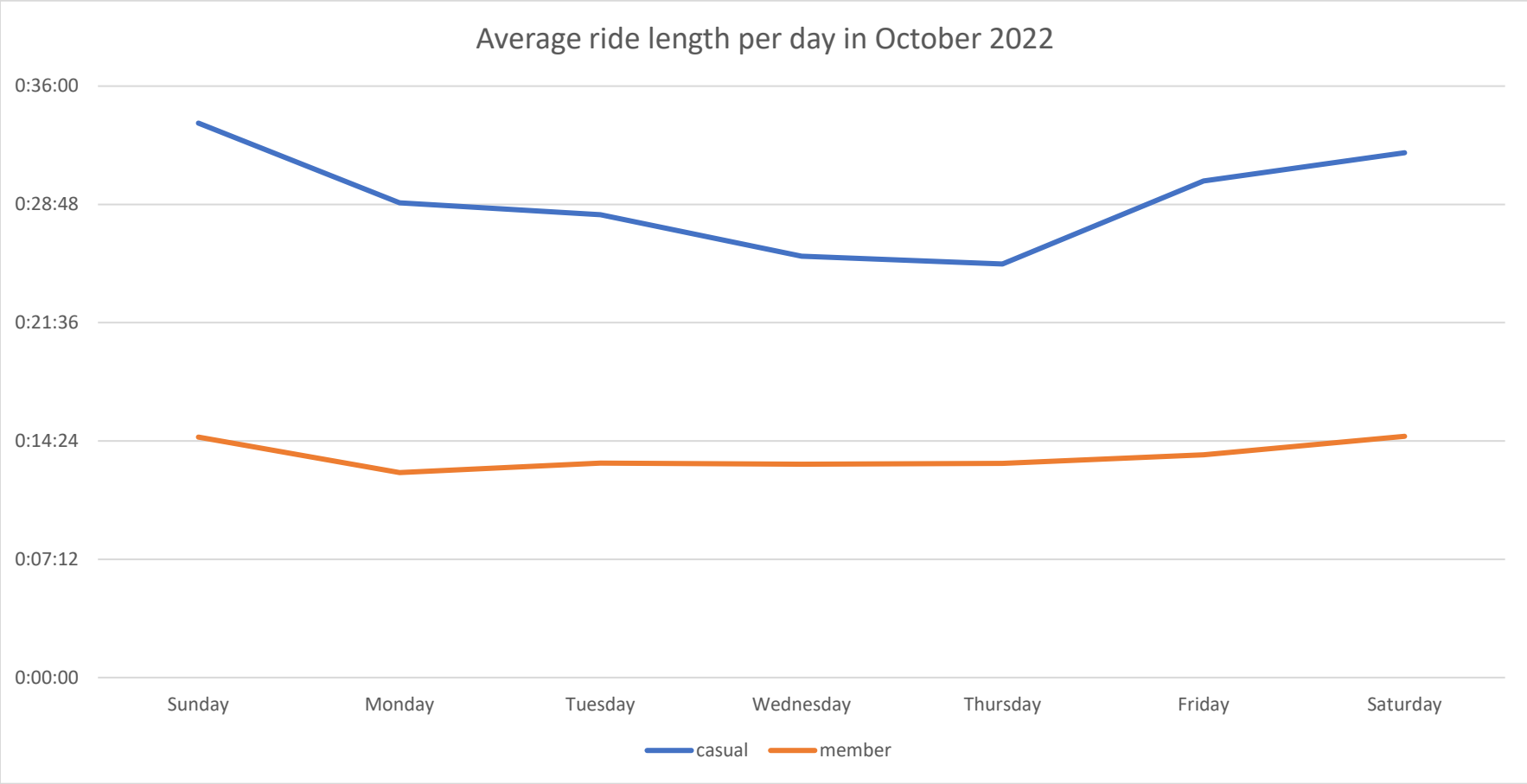

Figure 11. Line comparison Casual vs Member rides per day in October 2022

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used with a maximum of around 5,000 trips on Saturdays throughout the month.

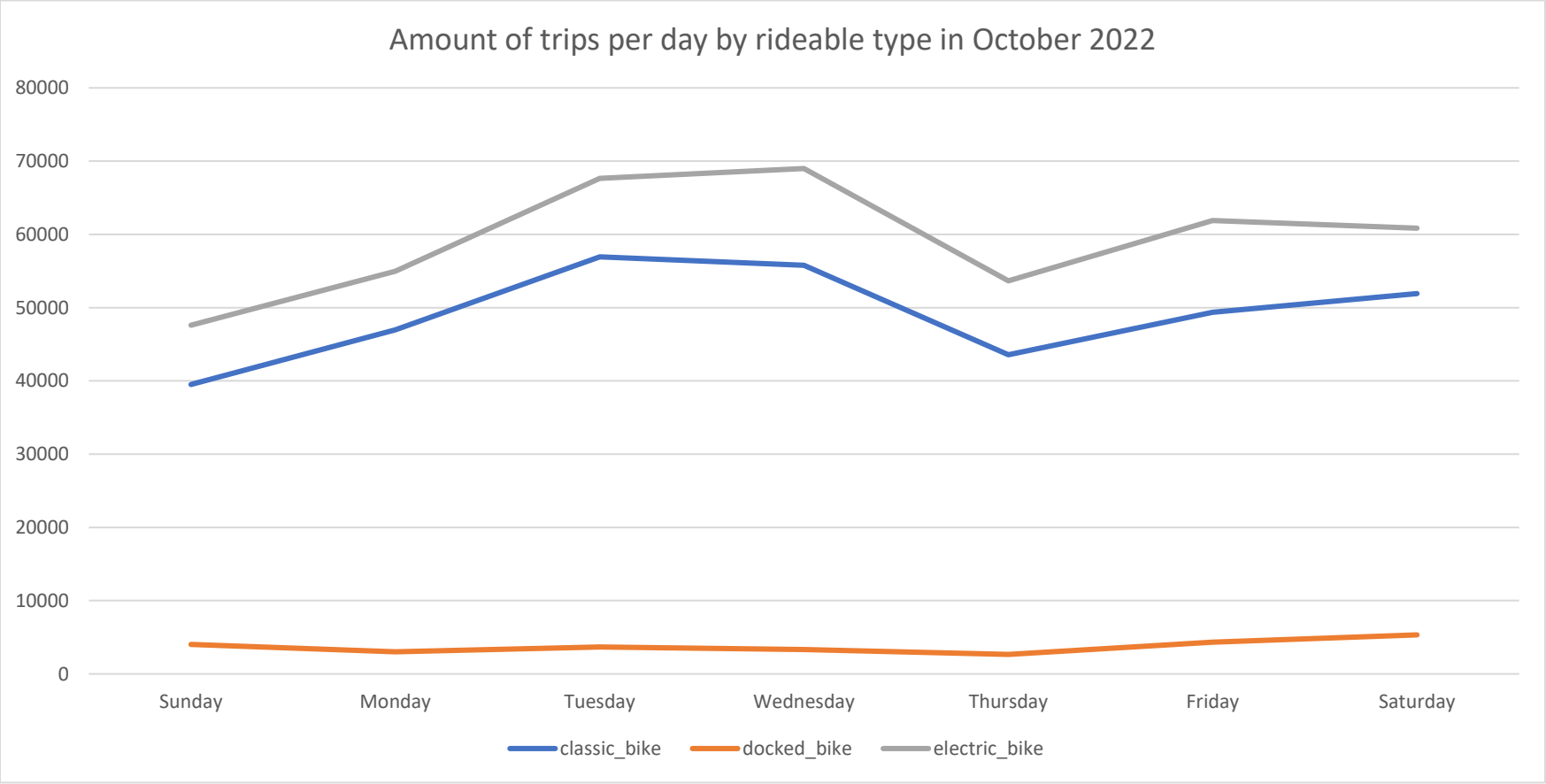

Figure 12. Line comparison number of rides per day by rideable type in October 2022

November 2022

Usage Frequency

- Members: Tuesdays and Wednesdays remained the peak days for usage, however there was a significant decrease in the number of trips taken over the weekend. The cause might be the cold weather discouraging users from biking.
- Casual Riders: Casual riders showed consistent bike usage, with an average of around 12,500 trips per day. The higher number of trips was taken during the weekdays Tue-Thu.

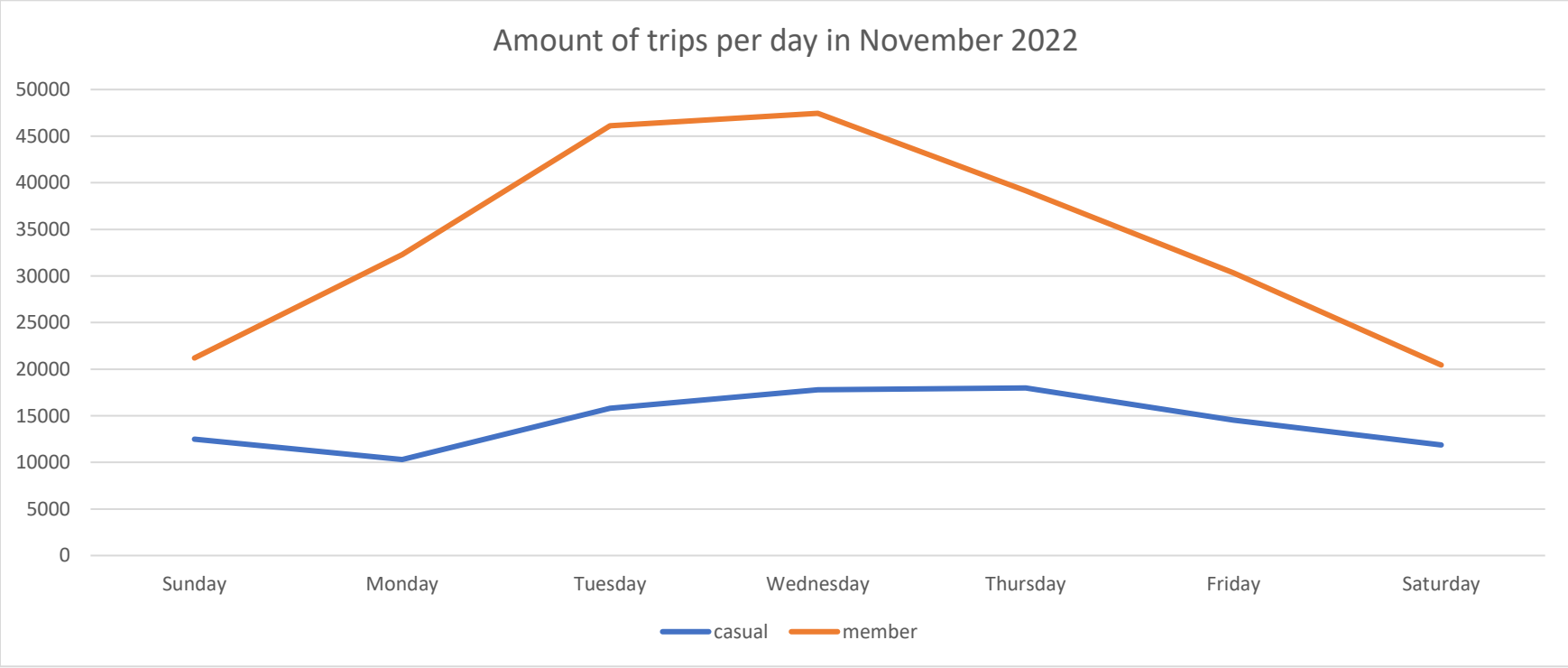

Figure 13. Line comparison Casual vs Member rides per day in November 2022

Ride Duration

- Members: The ride length for members remained consistent.

- Casual Riders: Casual riders continued to take longer trips compared to members. Sundays had a longer average ride length.

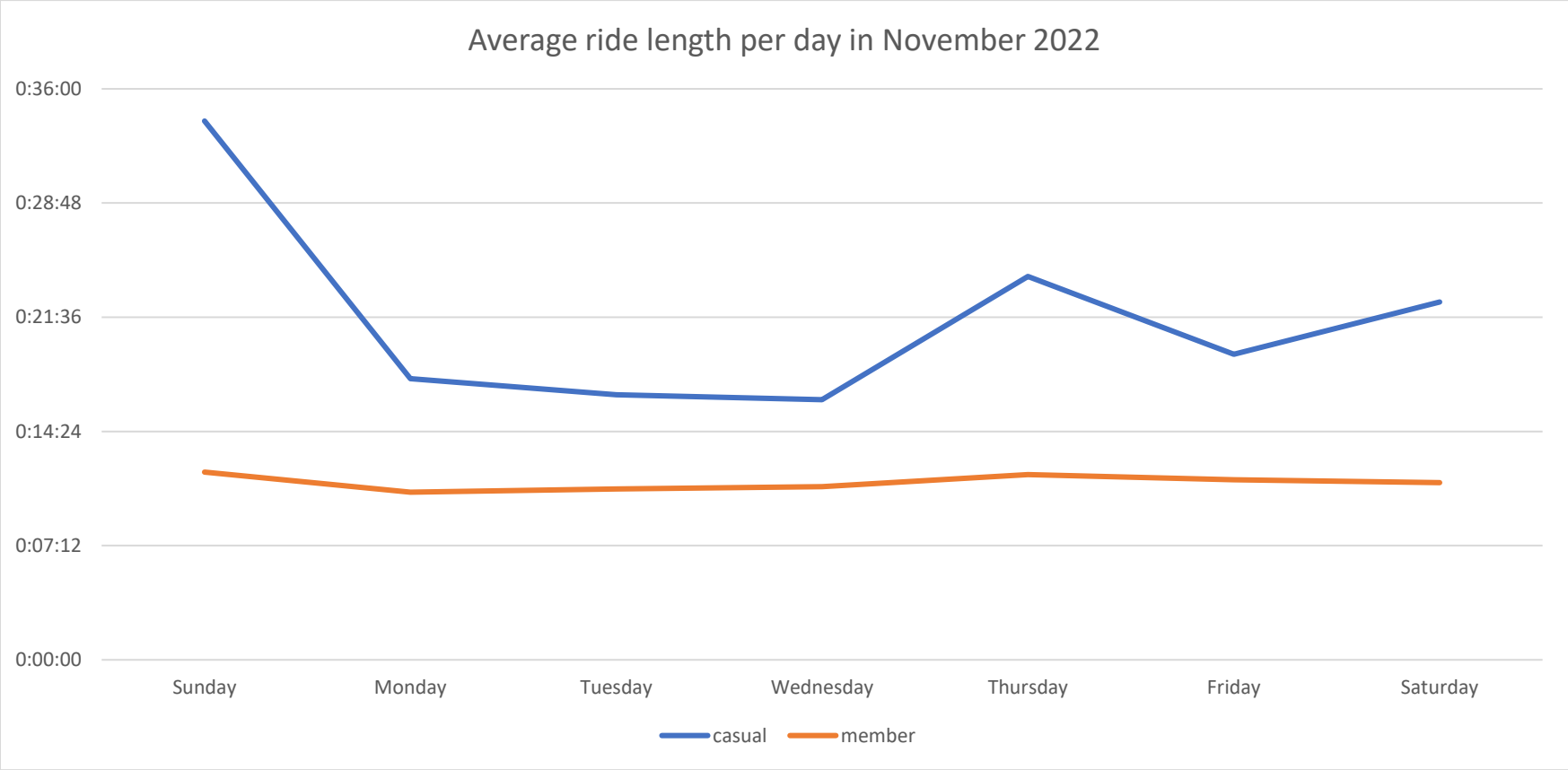

Figure 14. Line comparison Casual vs Member ride length per day in November 2022

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used with a maximum of around 5,000 trips on Saturdays throughout the month.

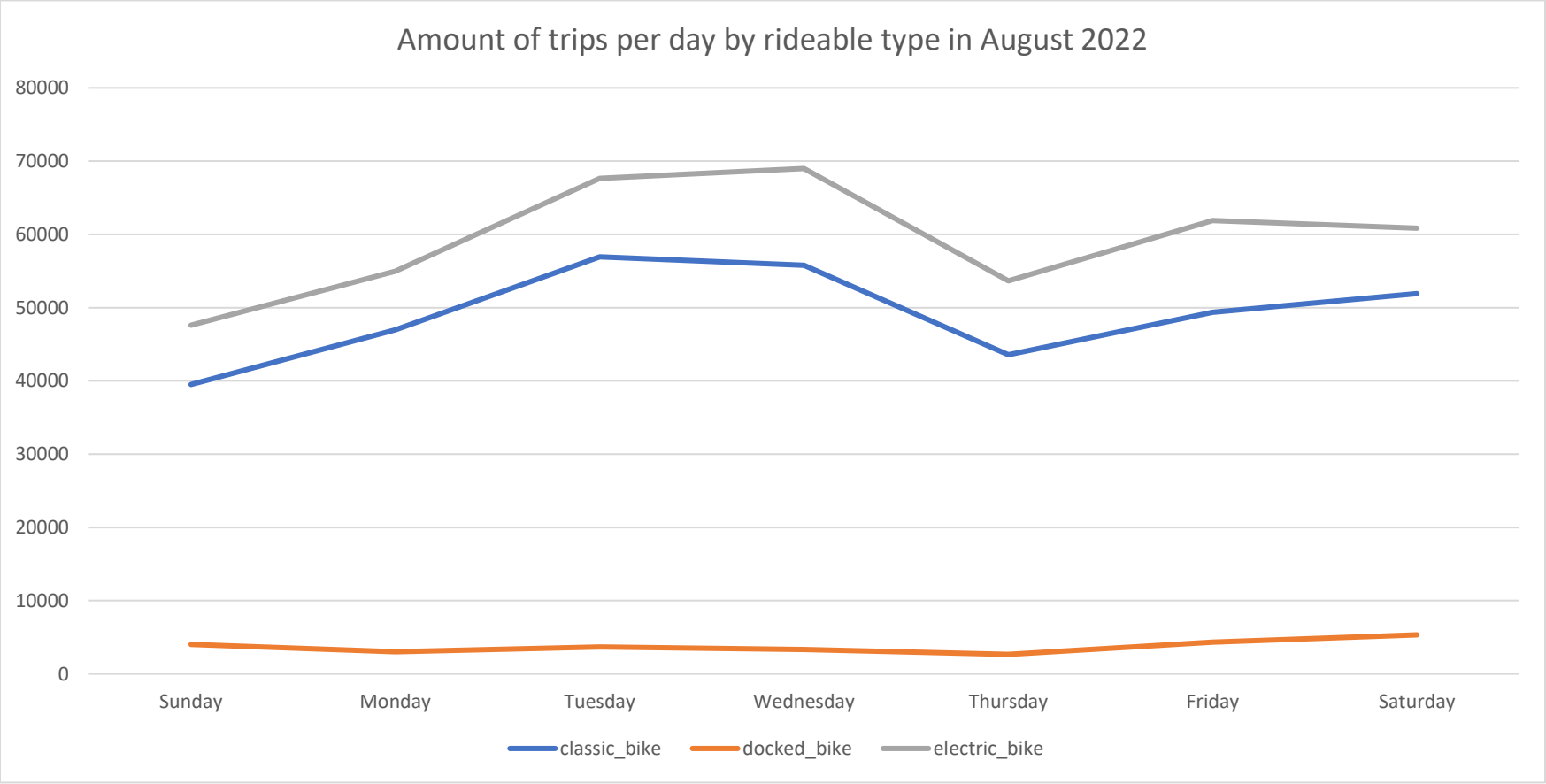

Figure 15. Line comparison number of rides per day by rideable type in November 2022

December 2022

N.B.

Data showed the lowest number of trips for both members and casual riders in December 2022

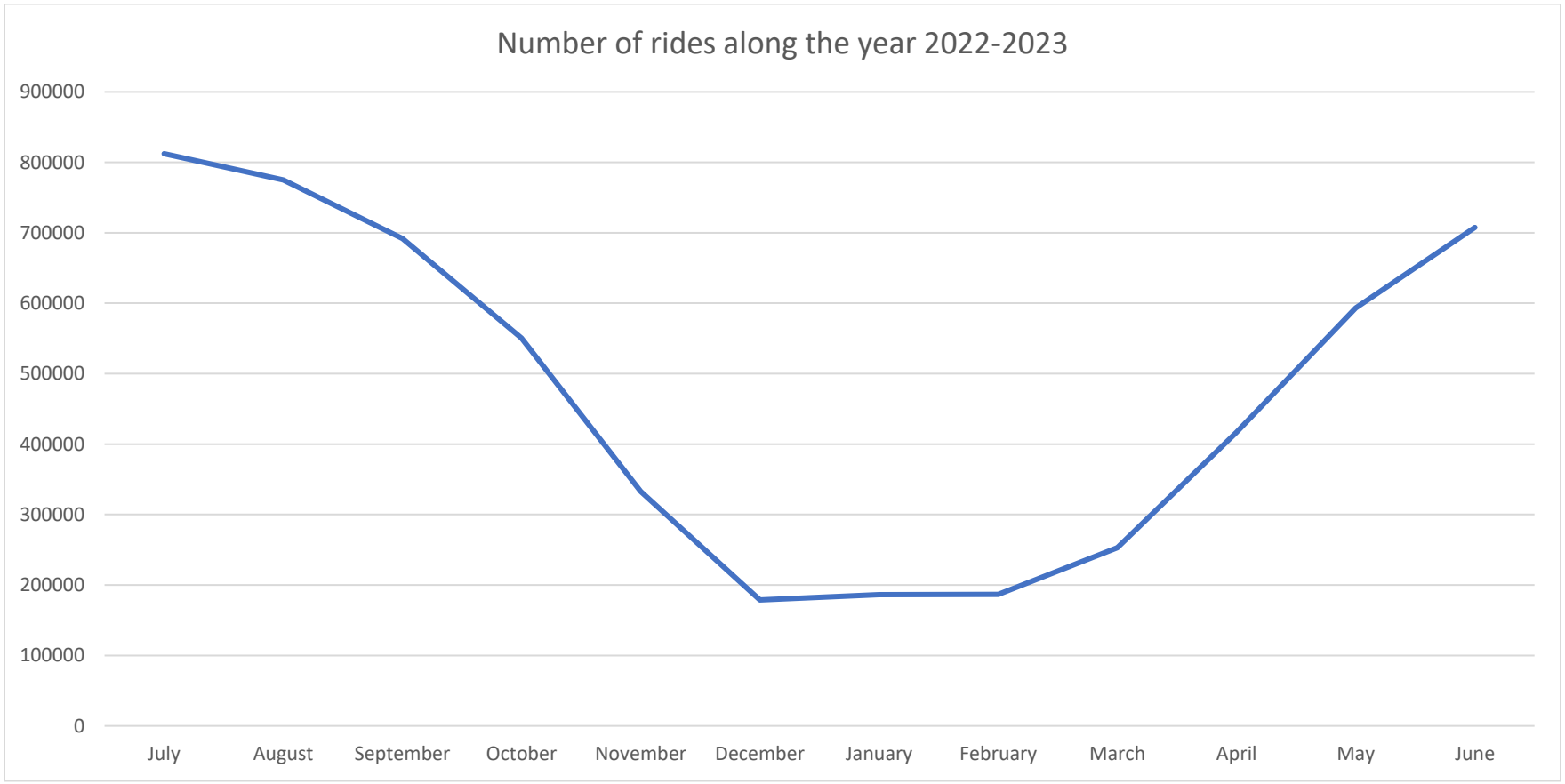

Figure 16. Number of rides along the year (July 2022-June 2023)

Usage Frequency

- Members: The highest number of trips taken by members was on Thursday, with a total of 27,000 trips. This number is significantly less than the highest recorded last month, which was on Wednesday with around 47,000 trips.
- Casual Riders: Riders limited their use of bikes, with the maximum number of trips taken on Thursdays reaching only around 8,000 trips. This number is less than the lowest recorded day from the previous month, which was Monday with 10,000 trips.

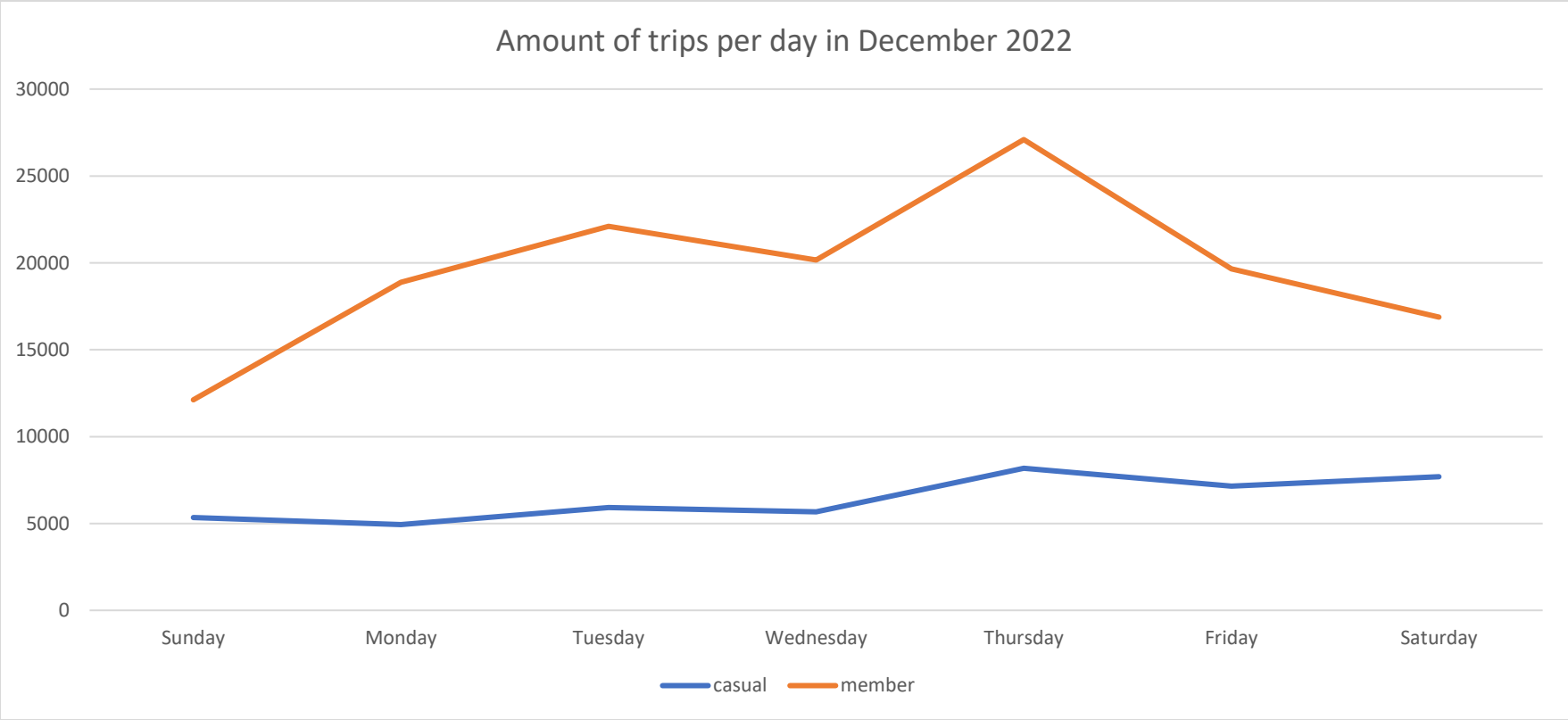

Figure 17. Line comparison Casual vs Member rides per day in December 2022

Ride Duration

- Members: The ride length for members significantly decreased to around 10 minutes on average all days of the week.
- Casual Riders: Casual riders are still taking longer trips compared to members, however, they are in the 15-29 mins range.

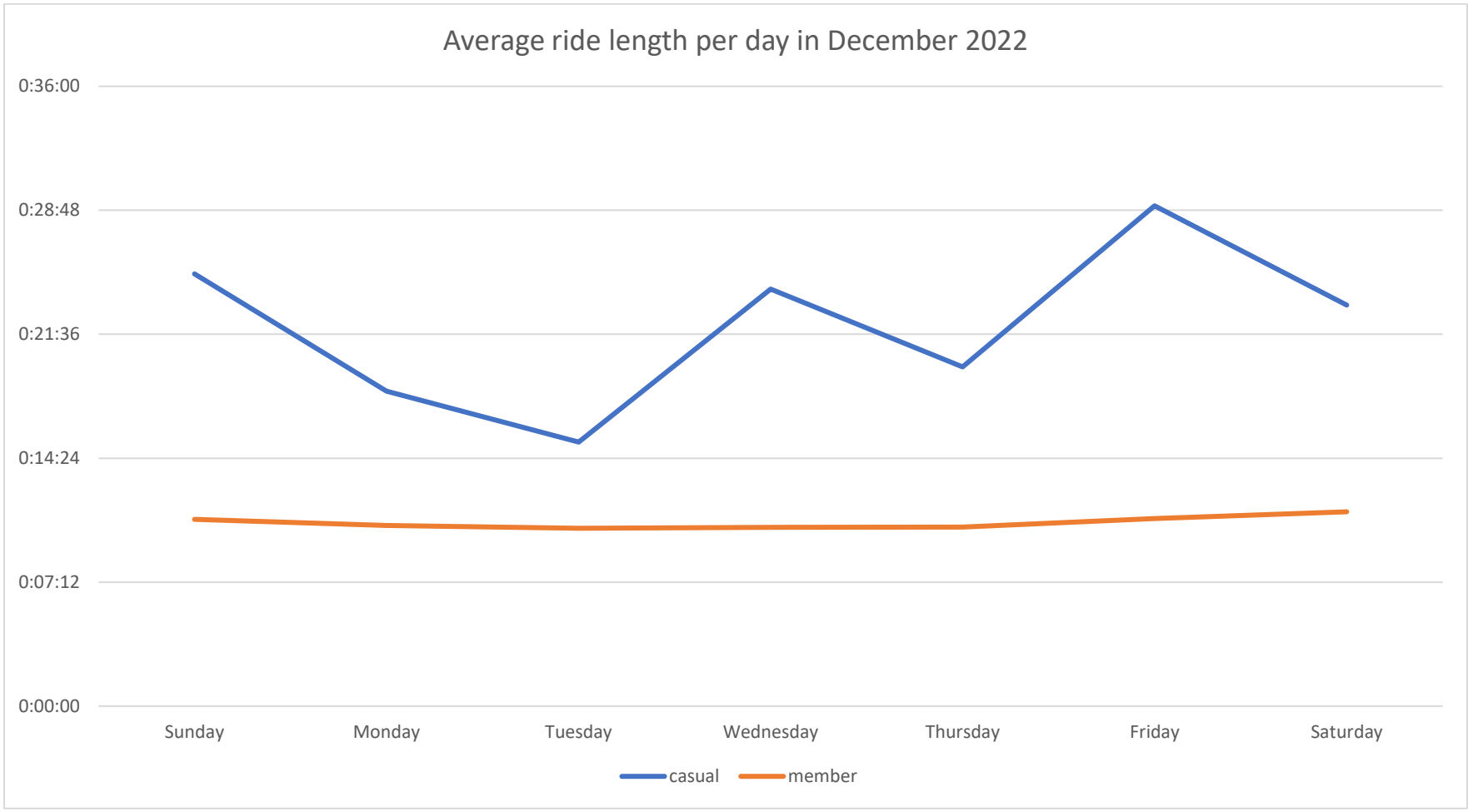

Figure 18. Line comparison Casual vs Member ride length per day in December 2022

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used throughout the month, which could be caused by the lack of interest in looking for a docking station in cold weather in addition to the potential presence of snow, which make people less interested in biking.

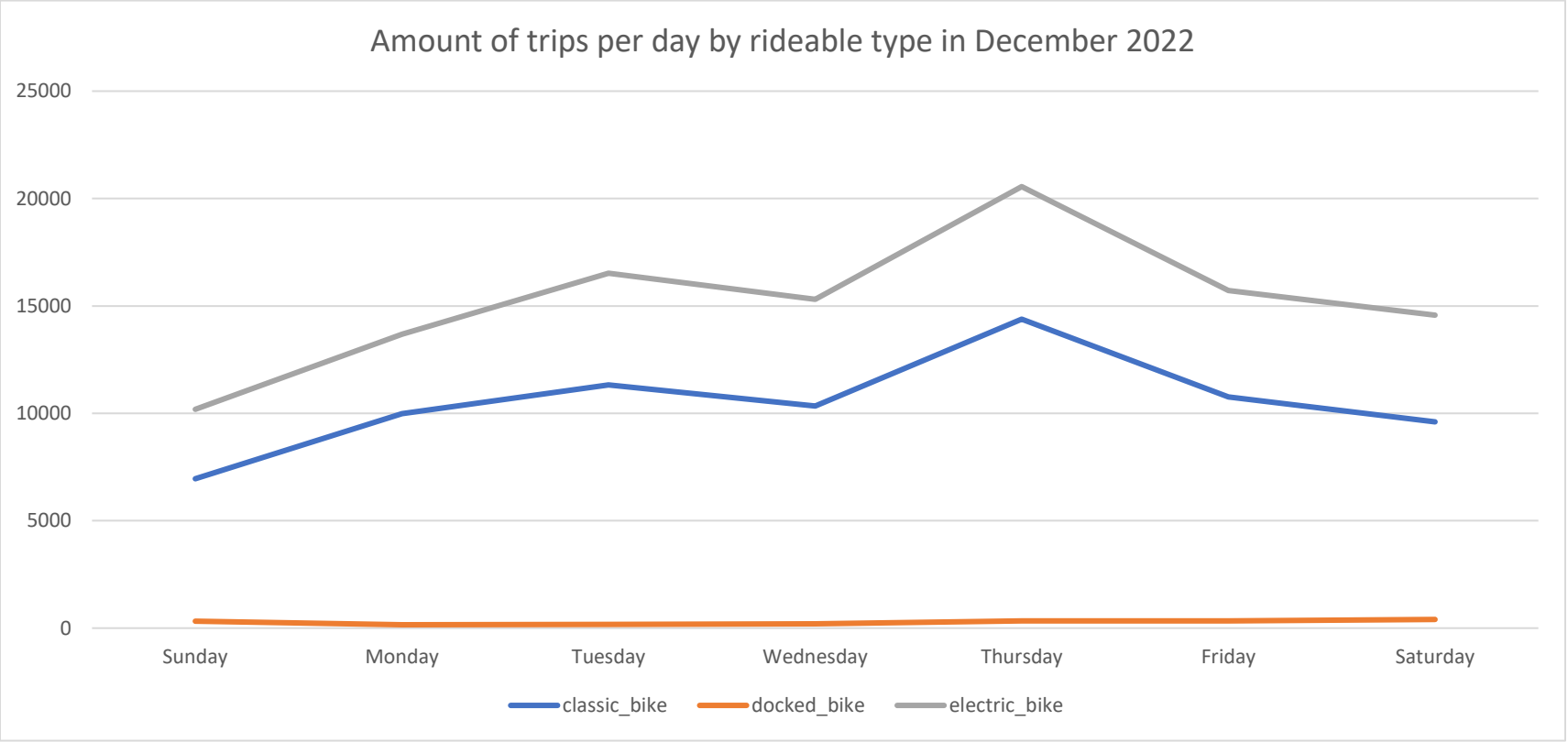

Figure 19. Line comparison number of rides per day by rideable type in December 2022

January 2023

Usage Frequency

- Members: Members utilized bikes 3x more often than casual riders. Maximum trips of 29,000 on Tuesdays

- Casual Riders: Casual riders used the bike-sharing service less frequently compared to members.

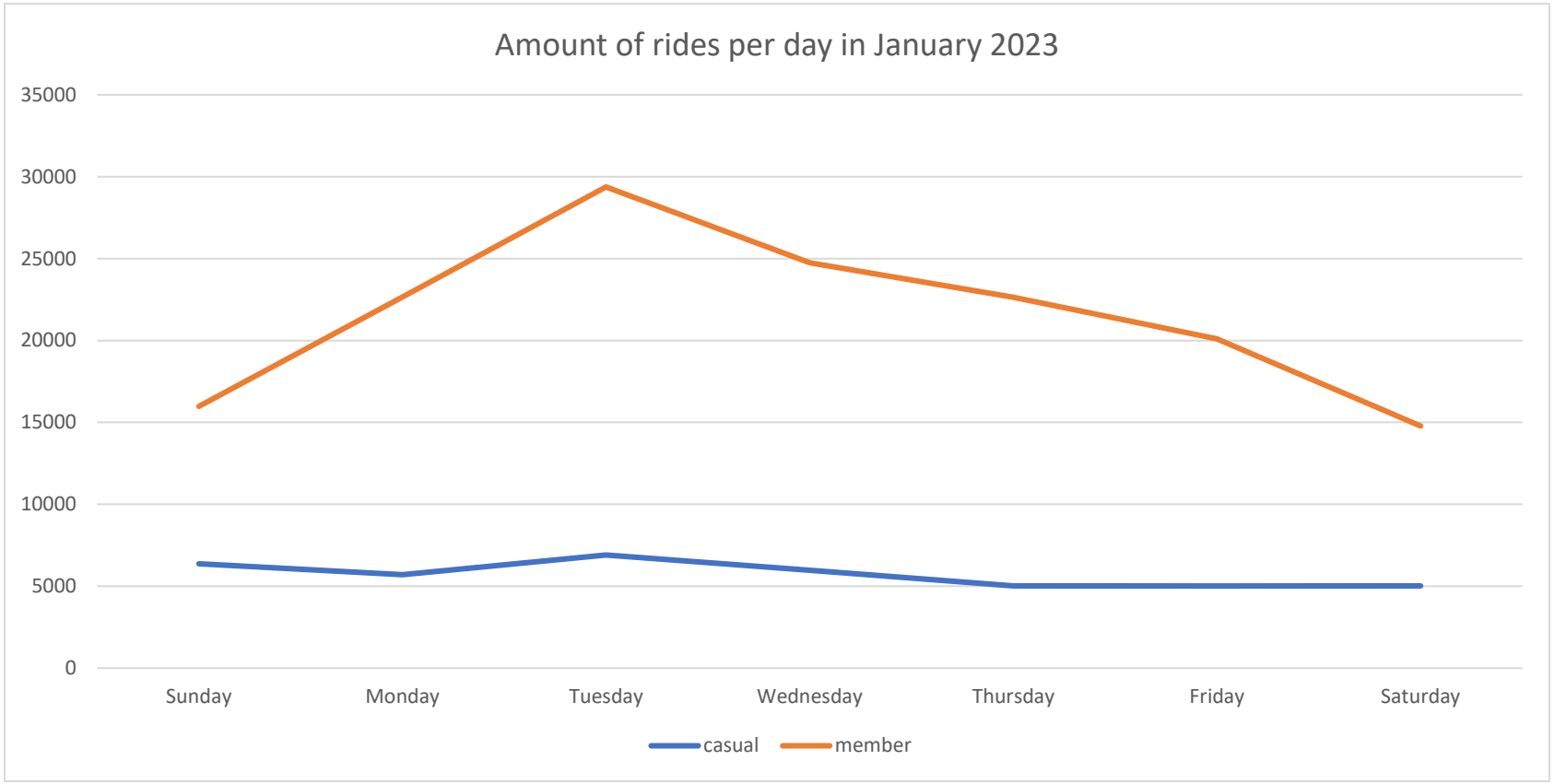

Figure 20. Line comparison Casual vs Member rides per day in January 2023

Ride Duration

- Members: The average ride length for members remained relatively consistent and short throughout the weeks.
- Casual Riders: In contrast, casual riders took rides with an average duration twice as long as members, indicating their preference for longer rides.

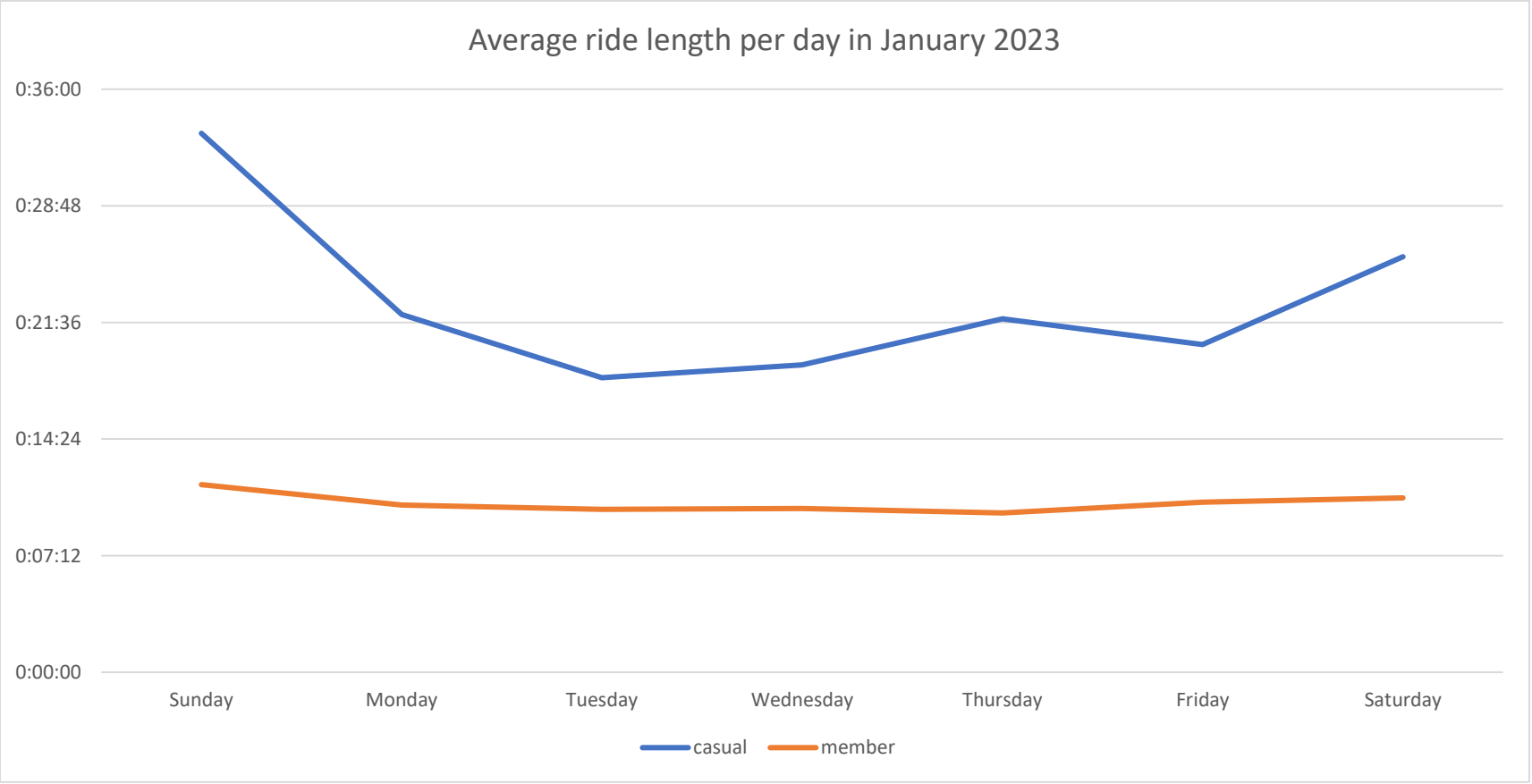

Figure 21. Line comparison Casual vs Member ride length per day in January 2023

Rideable types

Electric and classic bikes were equally popular and had a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used throughout the month, which could be caused by the lack of interest in looking for a docking station in cold weather in addition to the potential presence of snow, which makes people less interested in biking.

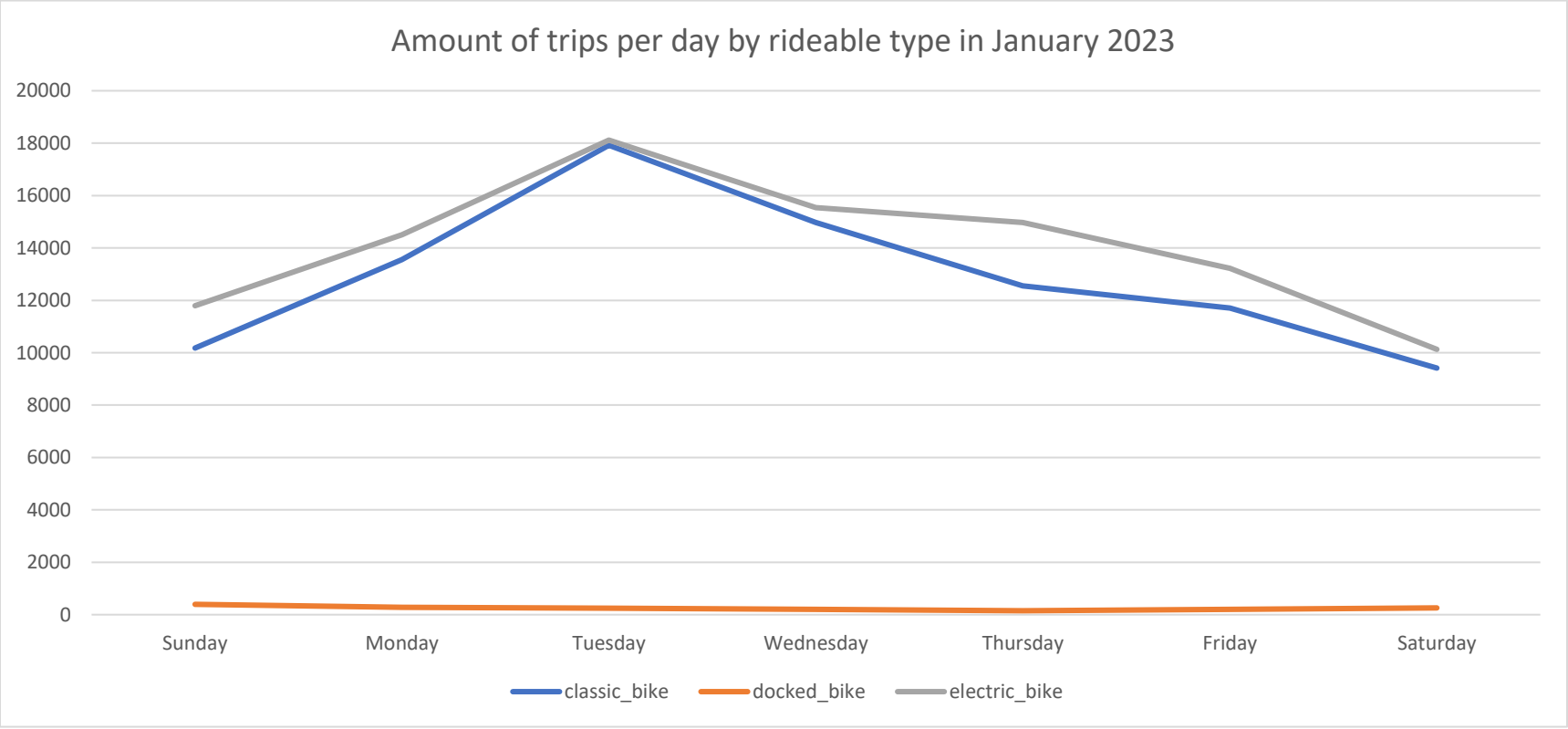

Figure 22. Line comparison number of rides per day by rideable type in January 2023

February 2023

Usage Frequency

- Members: Bike usage by members increased by 30% and 20% on Sundays and Saturdays respectively, while it decreased by around 20% from Tuesday to Friday compared to January.
- Casual Riders: Number of trips increased on Sundays by 80% compared to January. But the rest of the week stayed the same.

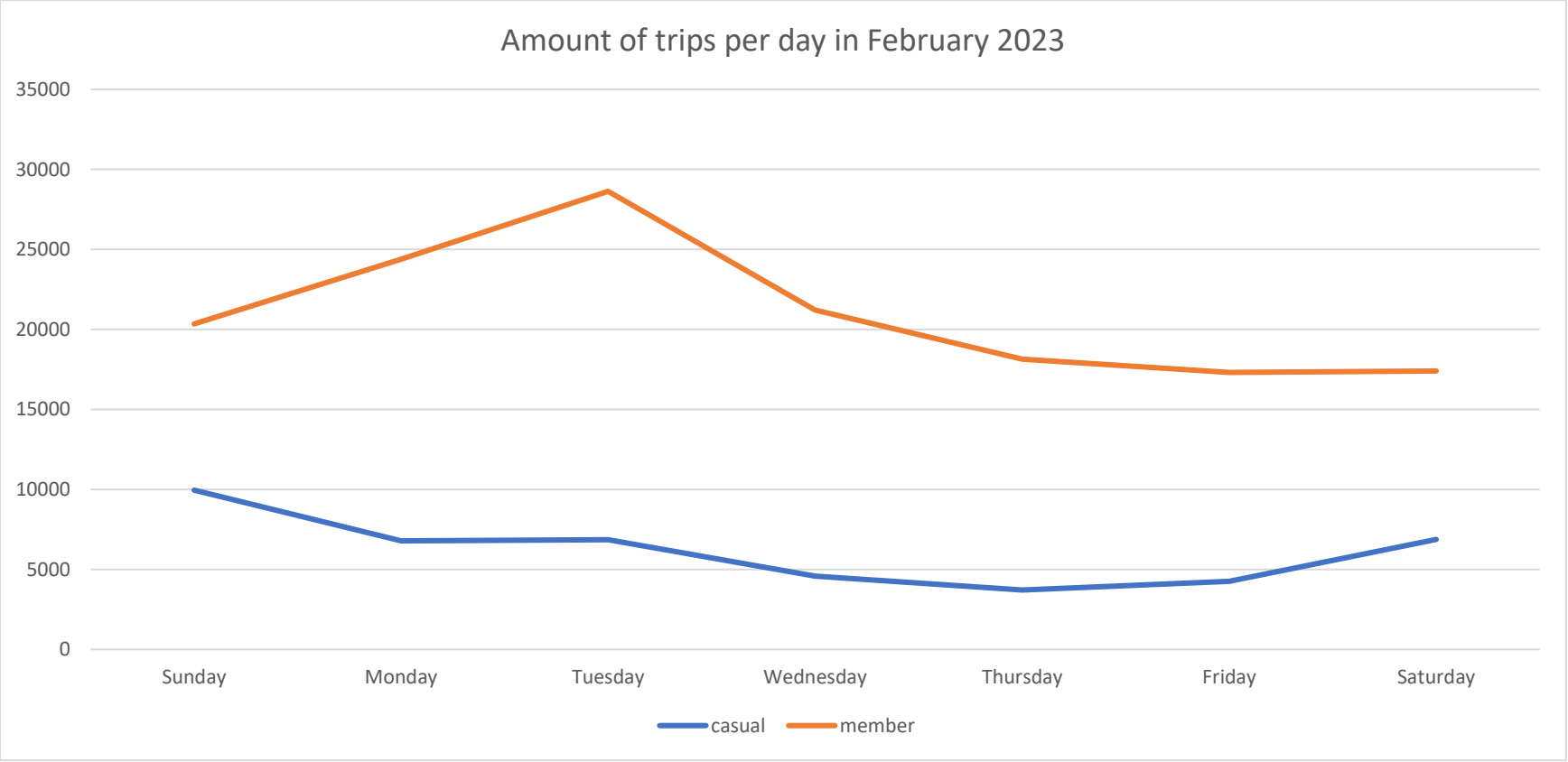

Figure 23. Line comparison Casual vs Member rides per day in February 2023

Ride Duration

- Members: The average ride length for members remained relatively consistent and short throughout the weeks.
- Casual Riders: Casual riders took the longest rides on Saturday and Sunday.

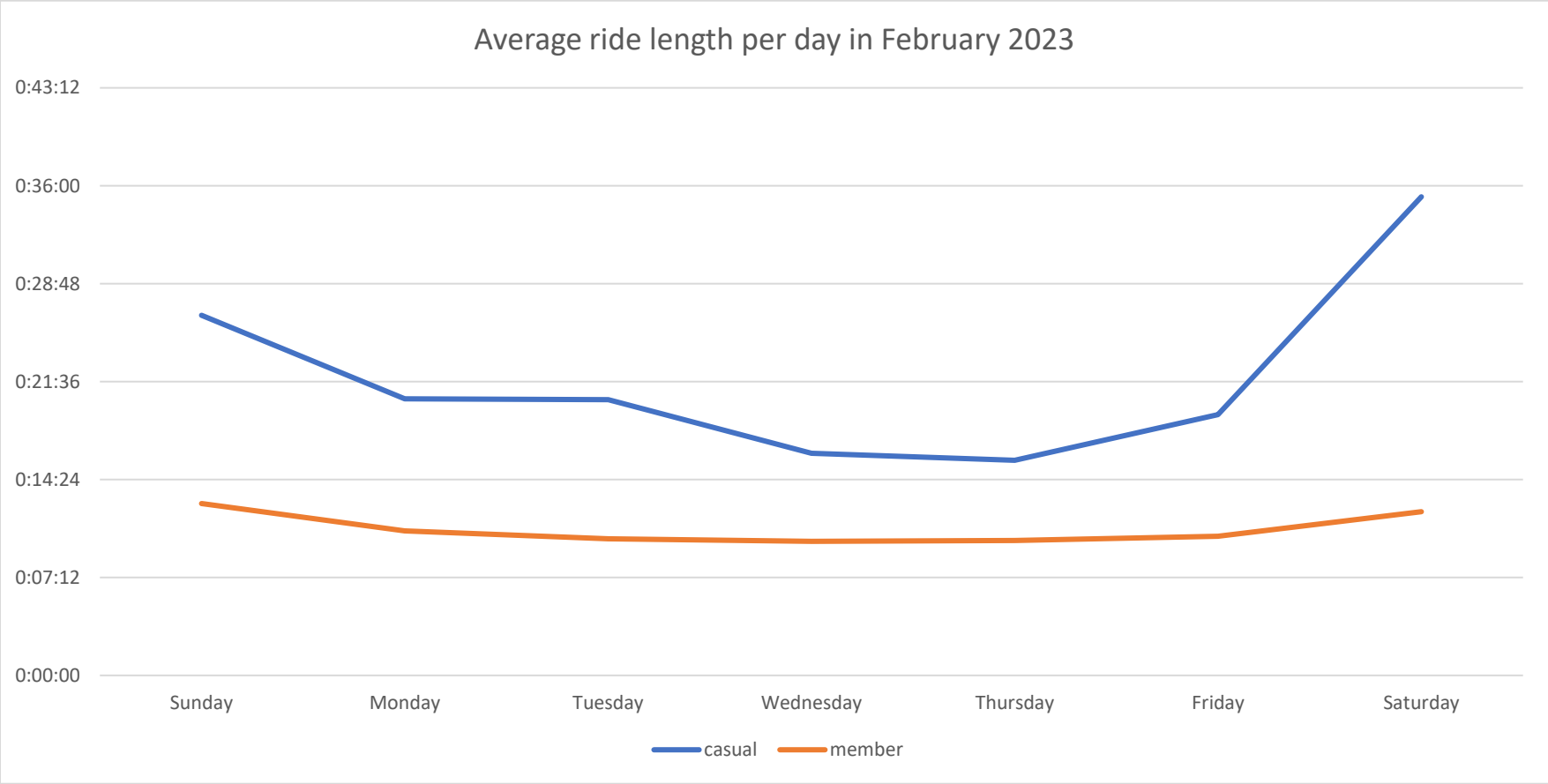

Figure 24. Line comparison Casual vs Member ride length per day in February 2023

Rideable types

Electric and classic bikes were equally popular and had a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used throughout the month.

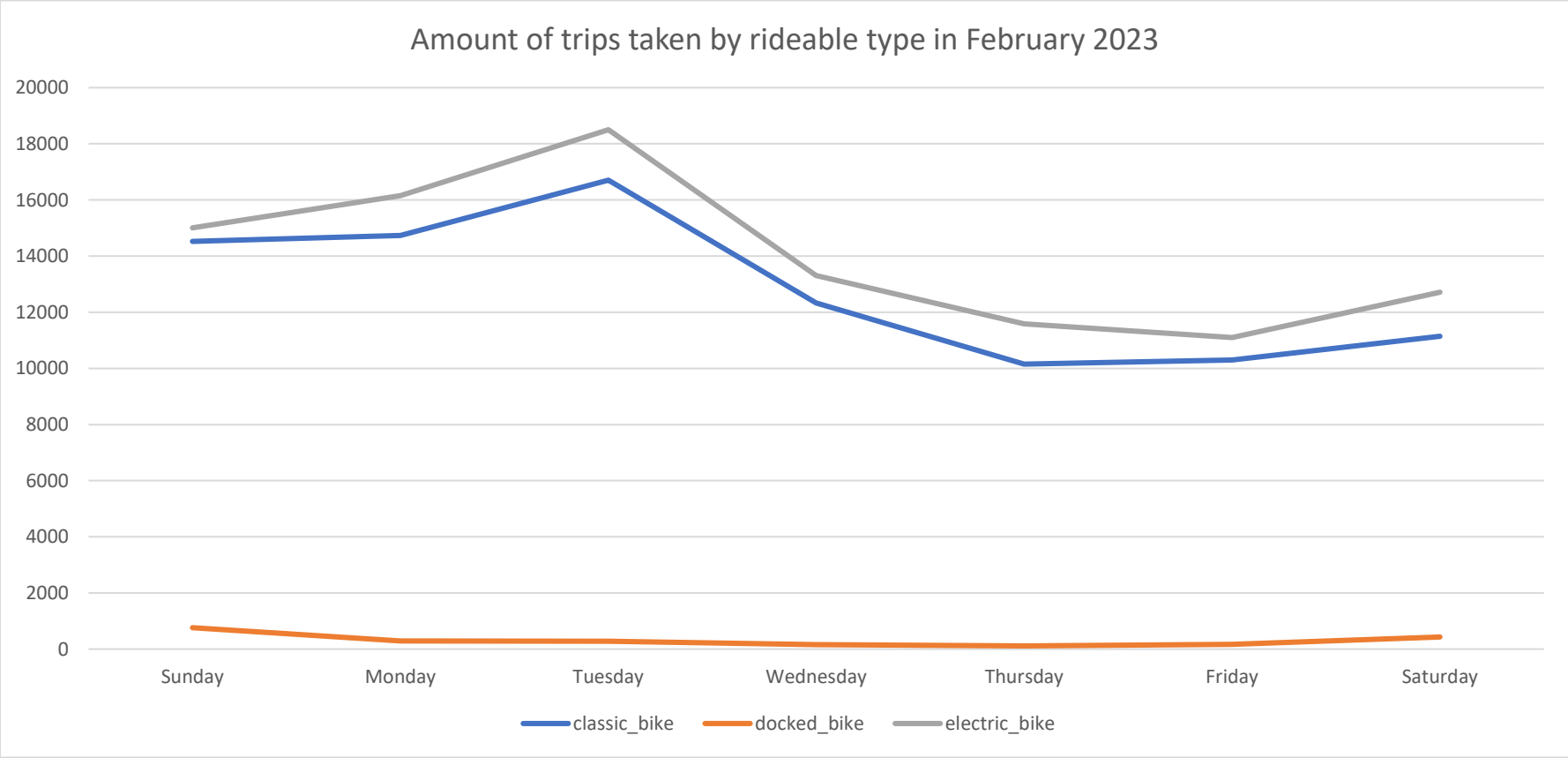

Figure 25. Line comparison number of rides per day by rideable type in February 2023

March 2023

Usage Frequency

- Members: Members' bike usage increased notably on Wednesday, Thursday, and Friday.
- Casual Riders: Casual riders' bike usage saw significant increases on Wednesday, Thursday, and Friday, a moderate increase on Tuesday, remained the same on Monday and Saturday, and a notable decrease on Sunday.

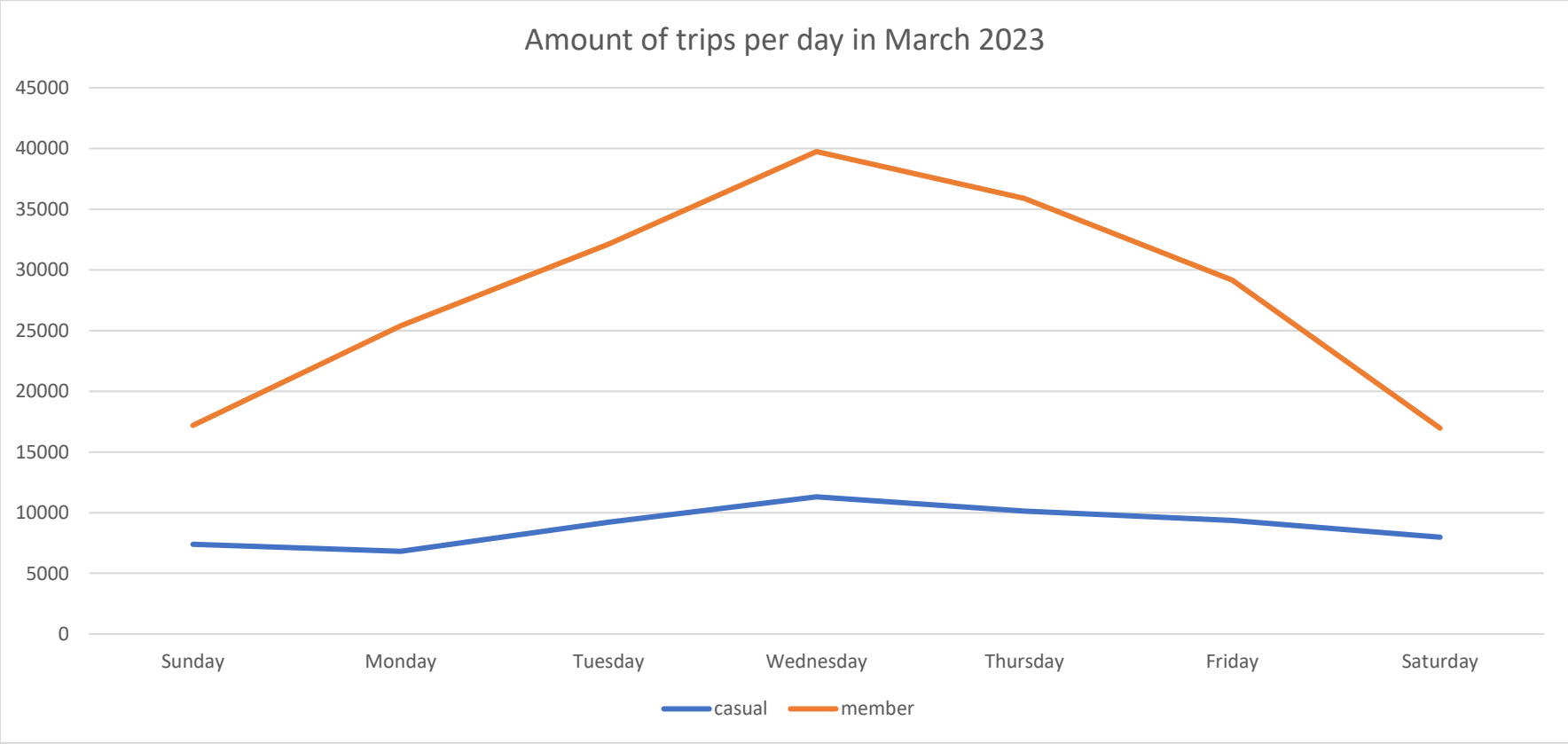

Figure 26. Line comparison Casual vs Member rides per day in March 2023

Ride Duration

- Members: The average ride length for members relatively decreased throughout the weeks.

- Casual Riders: Casual riders also took shorter rides throughout the weeks

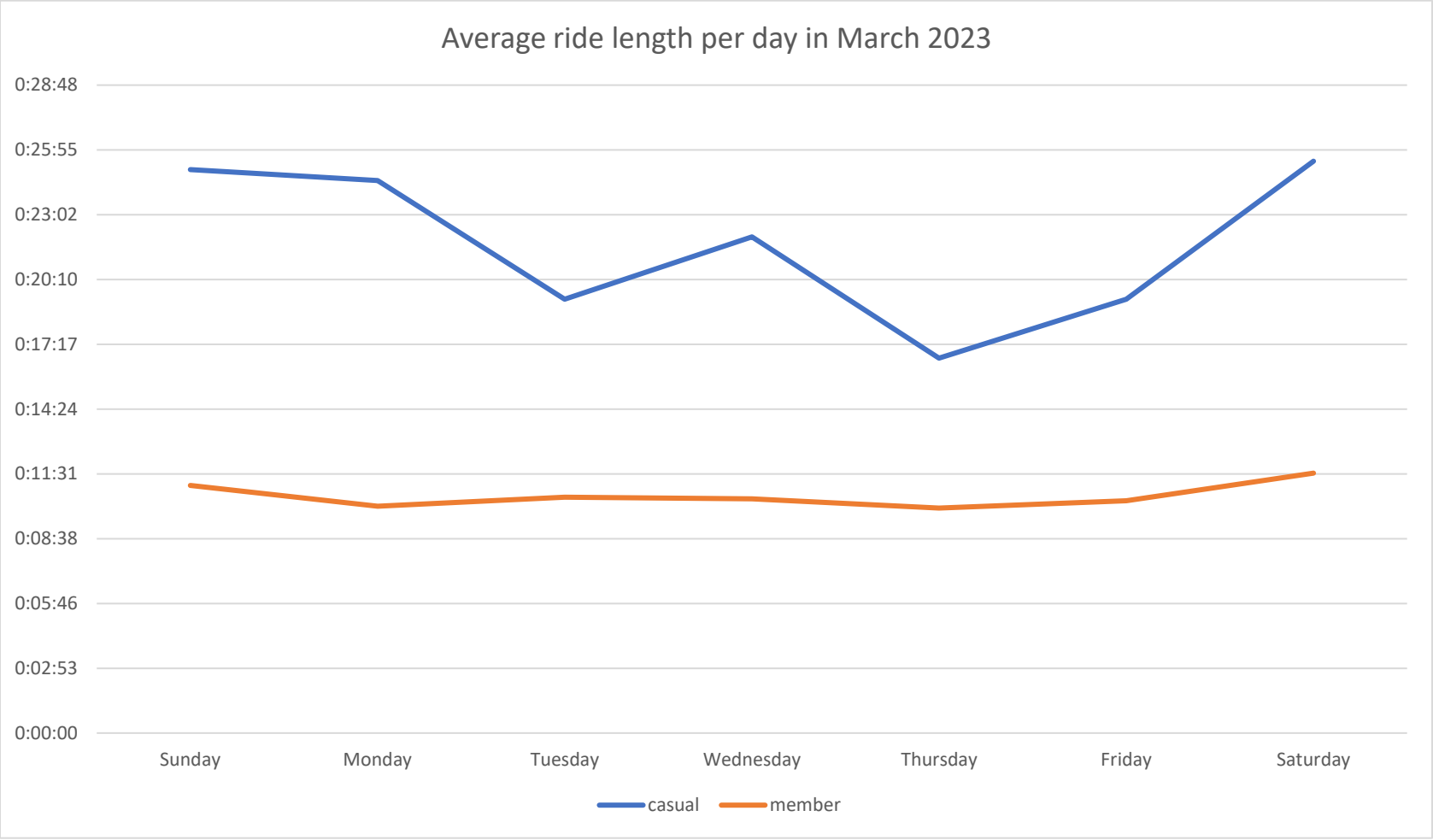

Figure 27. Line comparison Casual vs Member ride length per day in March 2023

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used throughout the month.

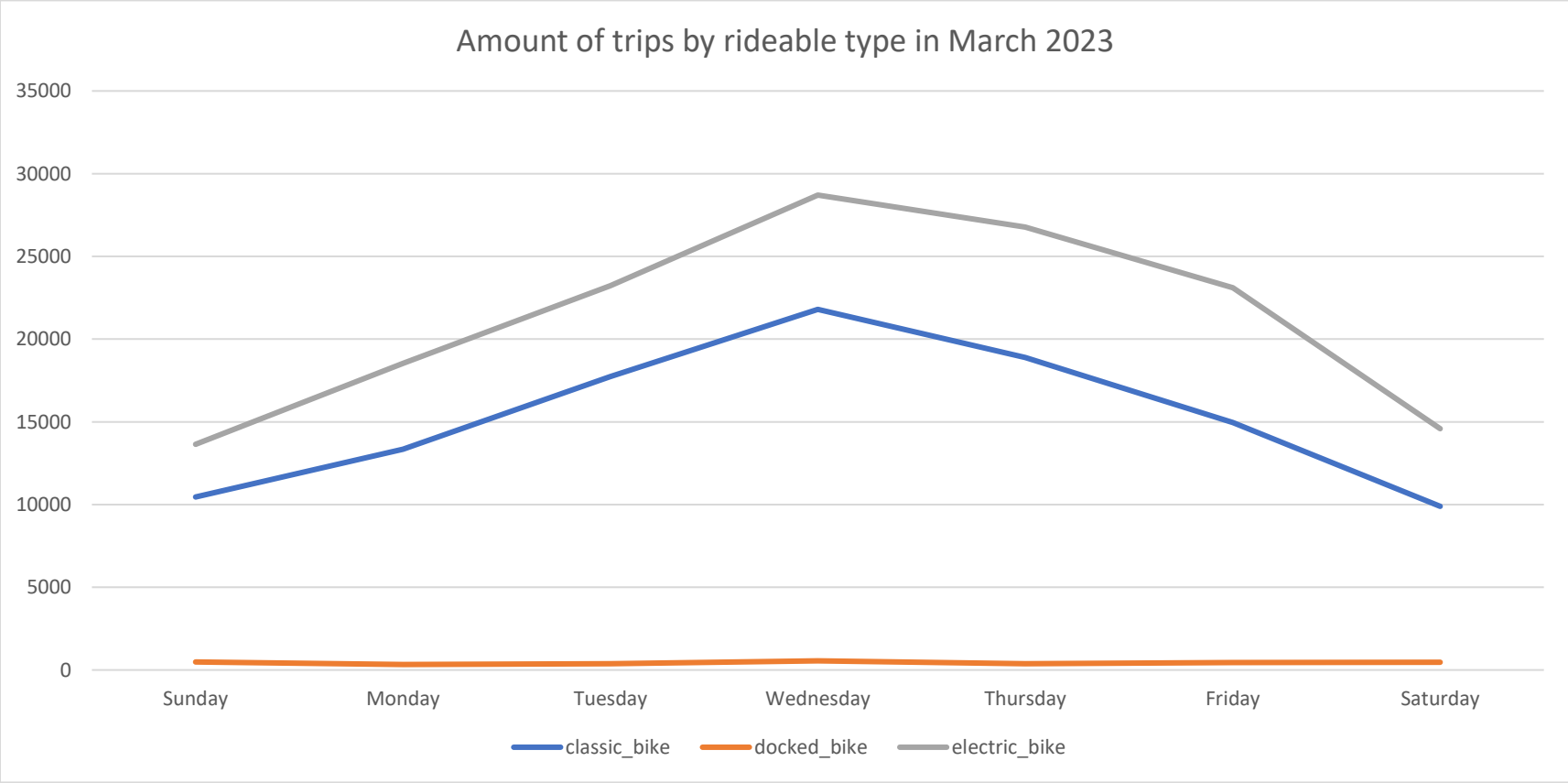

Figure 28. Line comparison number of rides per day by rideable type in March 2023

April 2023

Usage Frequency

- Members: Members' bike usage saw impressive increases on Friday, Saturday, and Sunday, notable increases on Monday and Thursday, and moderate increases on Tuesday and Wednesday.
- Casual Riders: Casual riders' bike usage saw an incredible increase on Saturday, impressive increases on Friday and Sunday, notable increases on Thursday and Tuesday, and moderate increases on Wednesday and Monday.

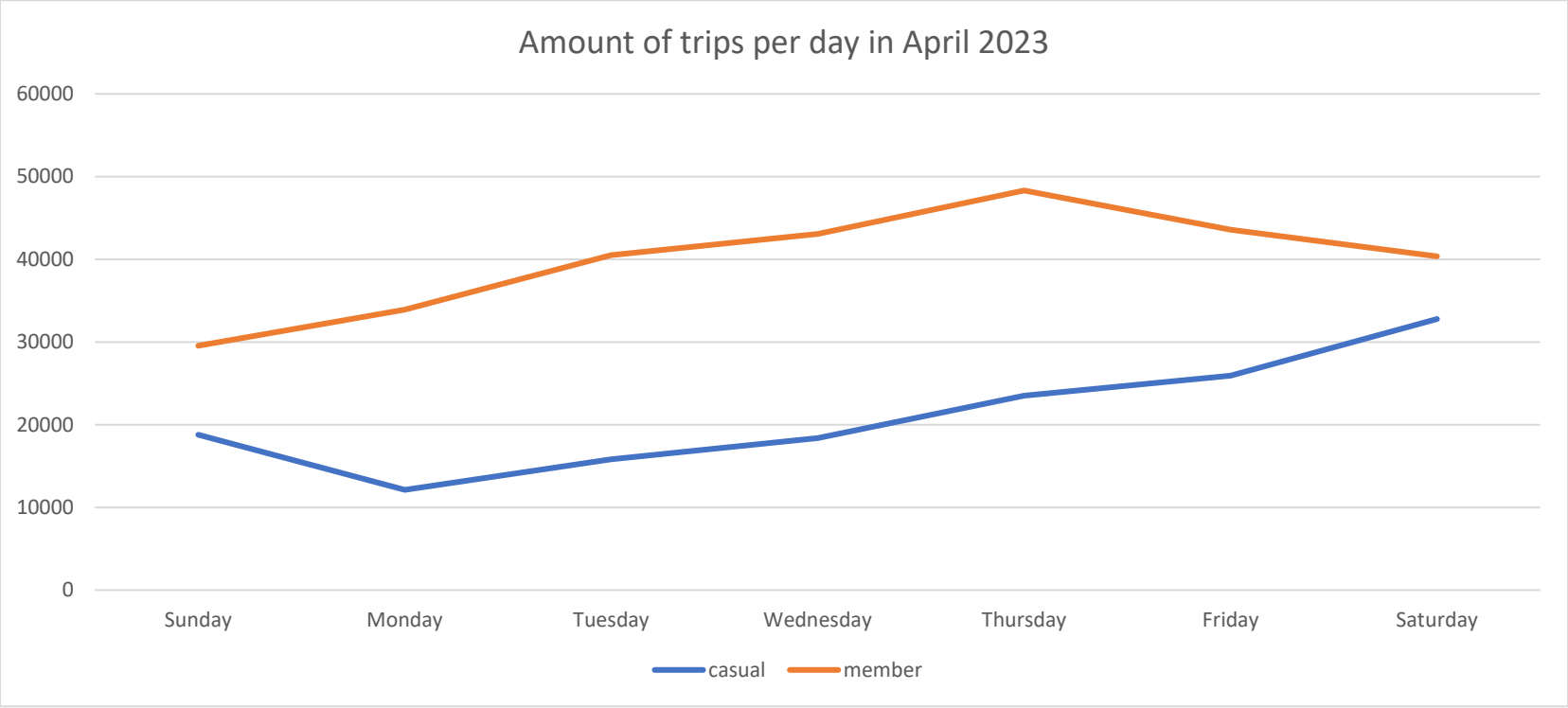

Figure 29. Line comparison Casual vs Member rides per day in April 2023

Ride Duration

- Members: The average ride length for members relatively increases throughout the weeks.
- Casual Riders: Casual riders took shorter rides throughout the weeks

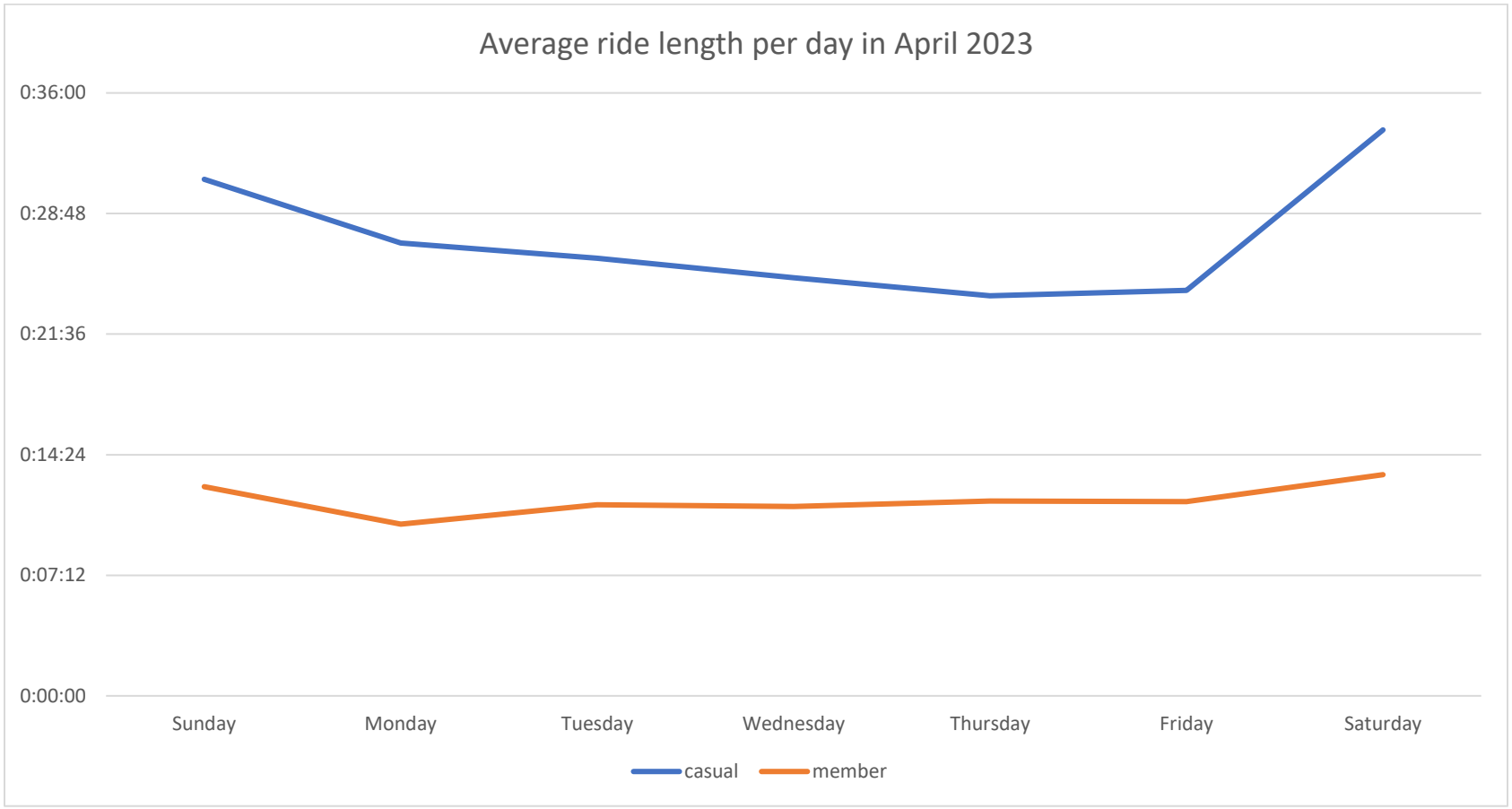

Figure 30. Line comparison Casual vs Member ride length per day in April 2023

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used throughout the month.

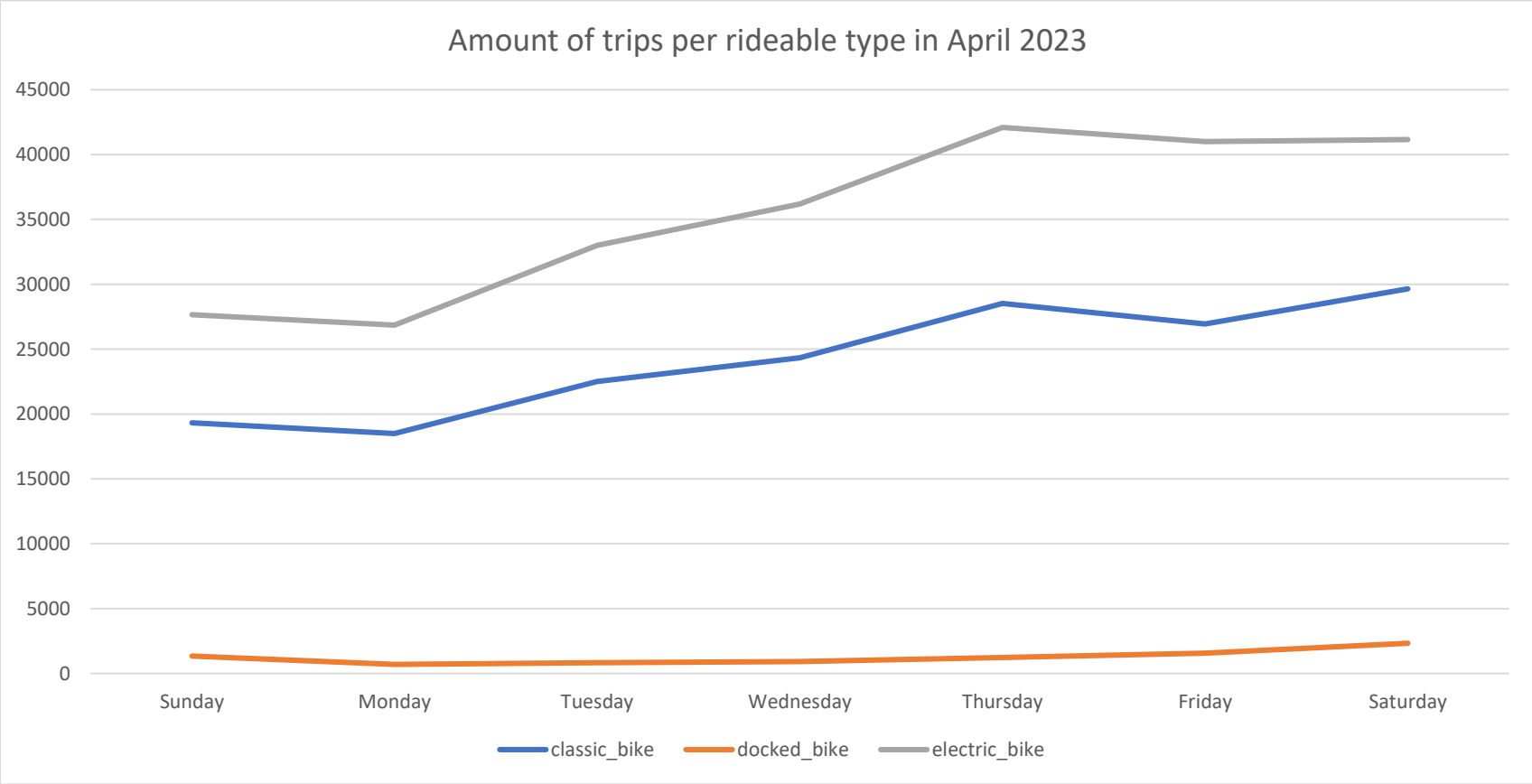

Figure 31. Line comparison number of rides per day by rideable type in April 2023

N.B.

These findings indicate a significant surge in bike rides for both members and casual riders during April 2023, with weekends being the most popular time for biking. Thursdays consistently had the fewest rides for both groups.

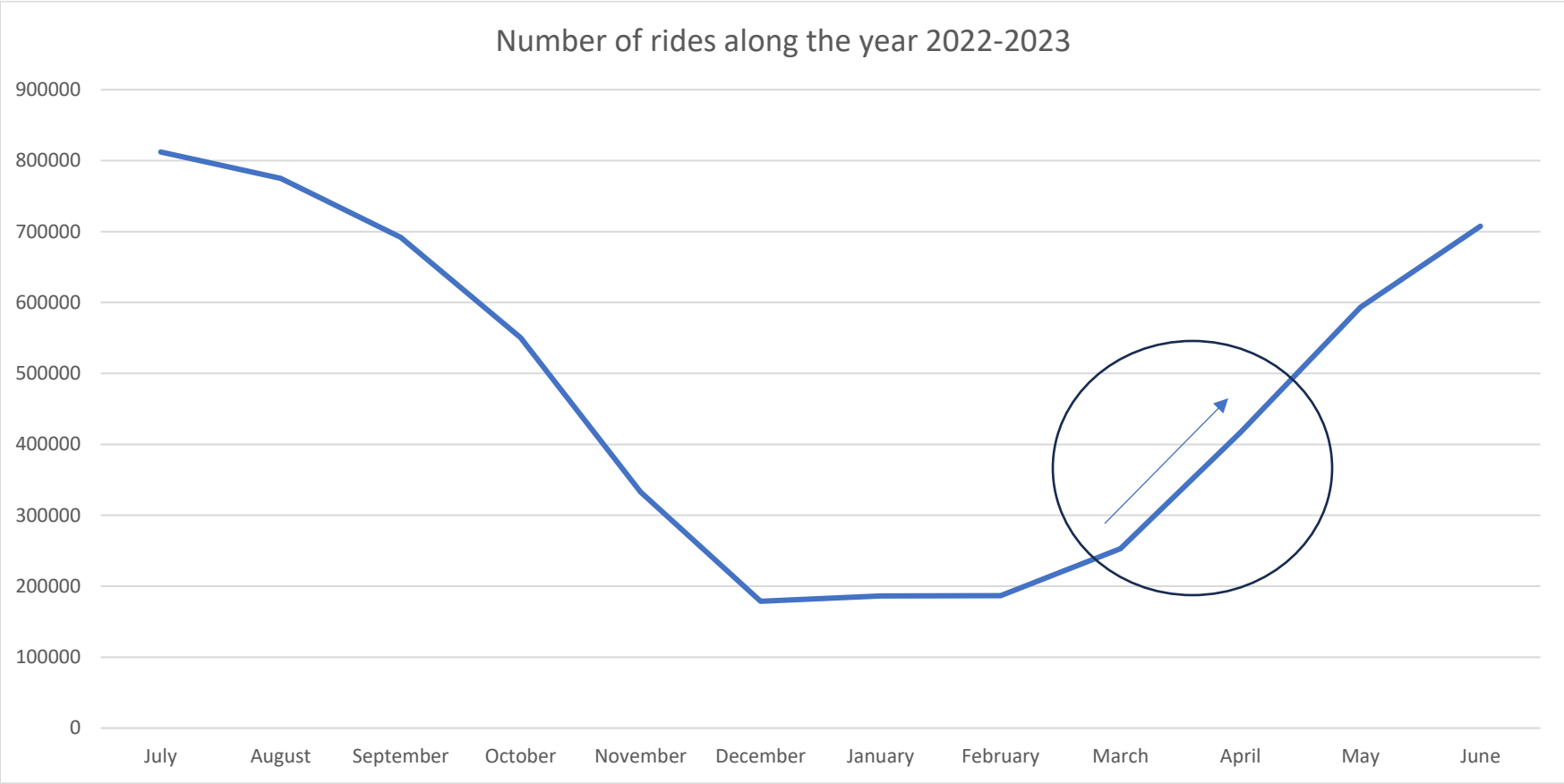

Figure 32. Number of rides along the year (July 2022-June 2023)

May 2023

Usage Frequency

- Members: Members' bike usage saw significant increases on Monday, Tuesday, and Wednesday, moderate increases on Sunday and Thursday, and slight increases on Saturday and Friday.
- Casual Riders: Casual riders' bike usage saw impressive increases on Sunday and Monday, significant increases on Tuesday and Wednesday and moderate increases on Thursday, Friday, and Saturday.

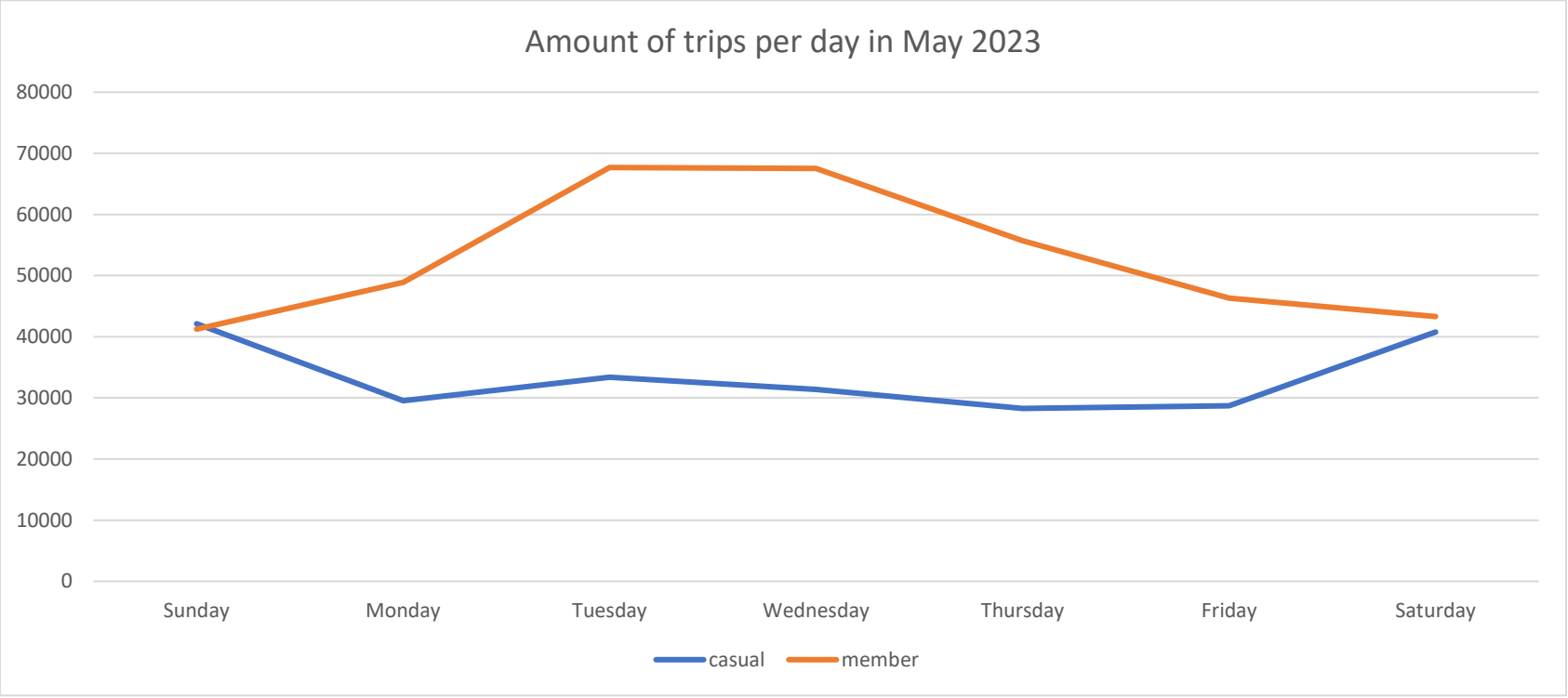

Figure 33. Line comparison Casual vs Member rides per day in May 2023

Ride Duration

- Members: The average ride length for members stayed relatively the same as last month.
- Casual Riders: Casual riders took slightly longer rides throughout the weeks

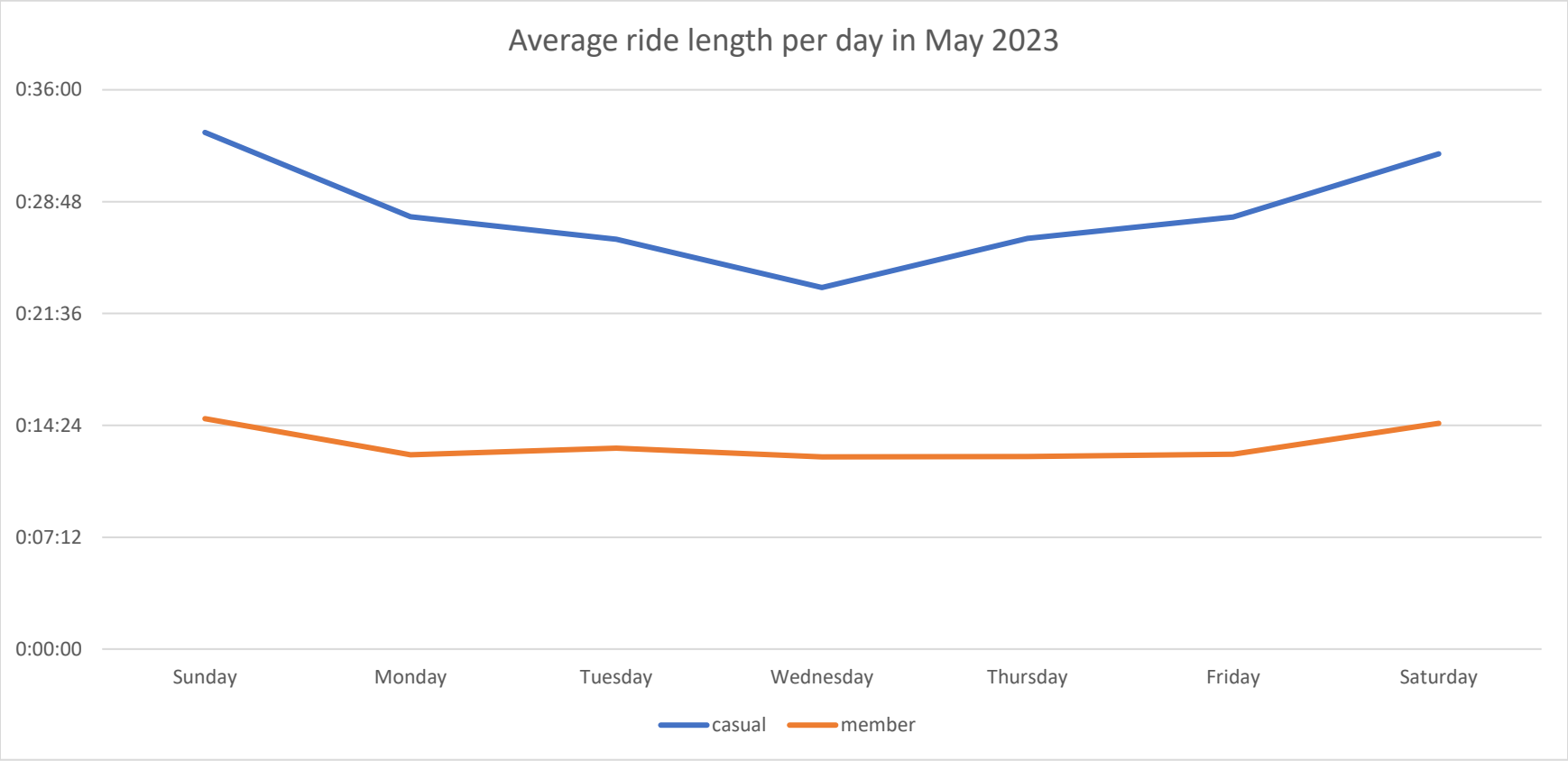

Figure 34. Line comparison Casual vs Member ride length per day in May 2023

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used throughout the month.

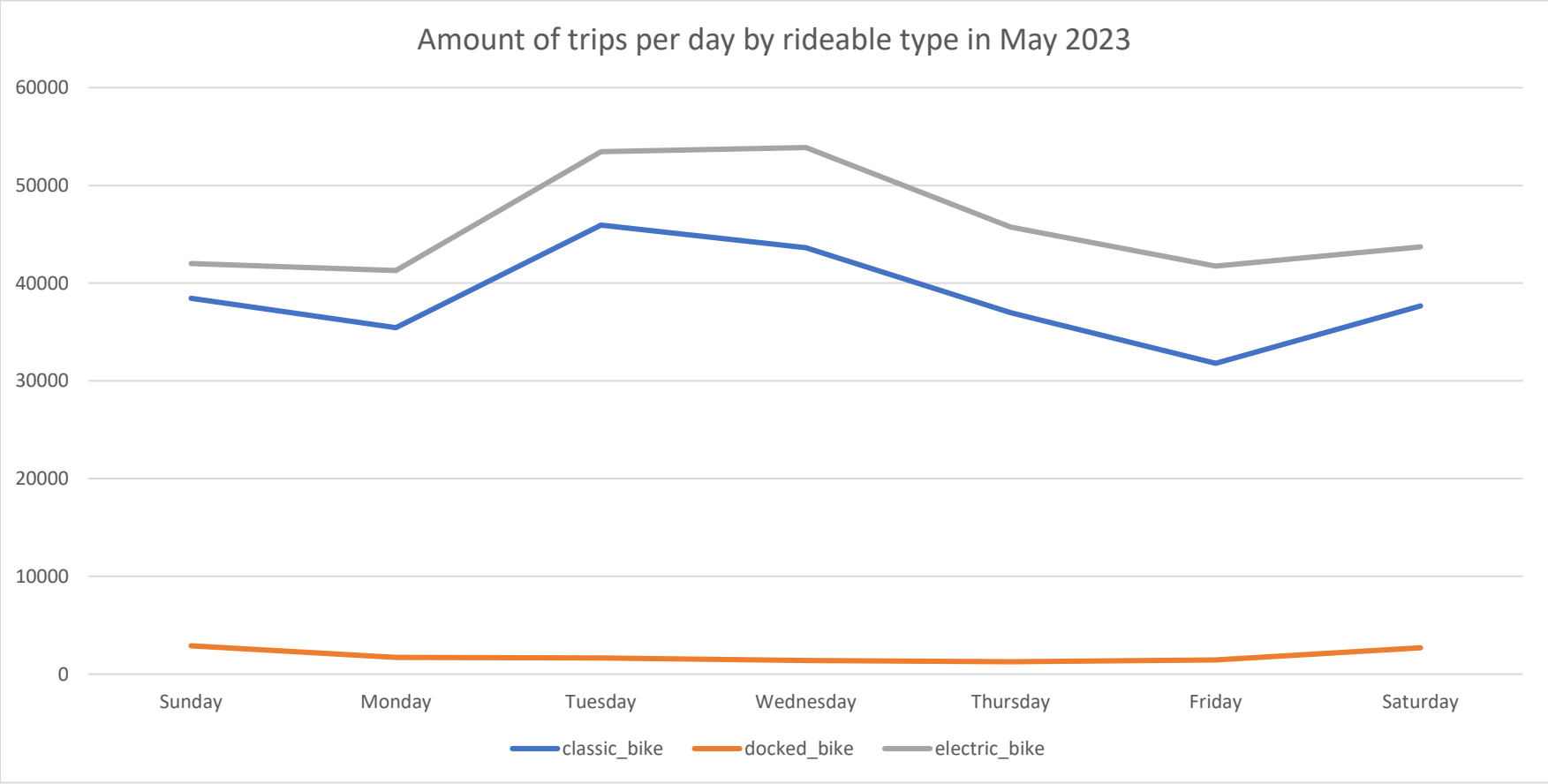

Figure 35. Line comparison number of rides per day by rideable type in May 2023

June 2023

Usage Frequency

- Members: Thursday and Friday stood out as peak days with over 70,000 trips each. Saturday saw a notable 40% increase, Monday and Sunday experienced a slight increase. However, Tuesday and Wednesday saw a decrease in the number of trips.
- Casual Riders: Casual riders' bike usage saw remarkable increases on Friday, substantial increases on Saturday and Thursday, significant increases on Tuesday, and slight increases on Monday, Wednesday, and Sunday.

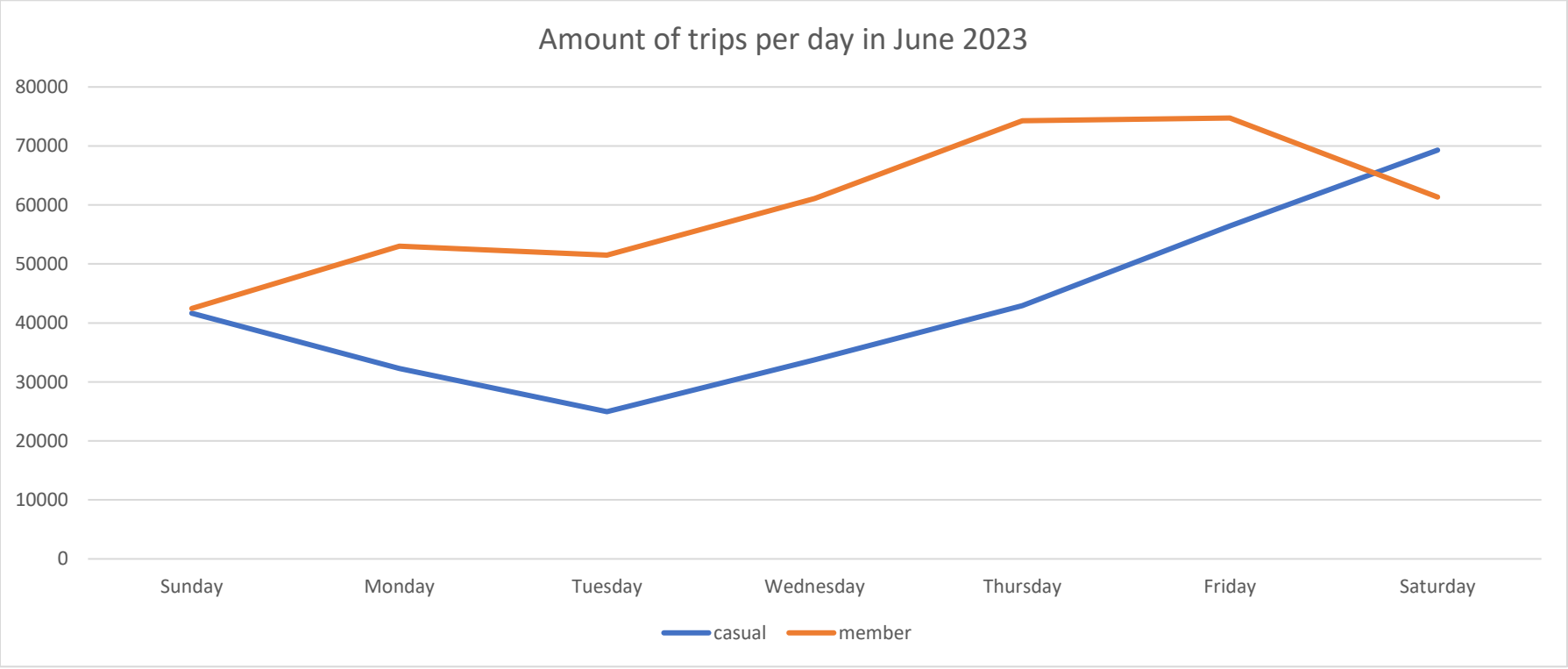

Figure 36. Line comparison Casual vs Member rides per day in June 2023

Ride Duration

- Members: The average ride length for members stayed relatively the same as last month.
- Casual Riders: The average ride length for casual riders also stayed relatively the same as last month.

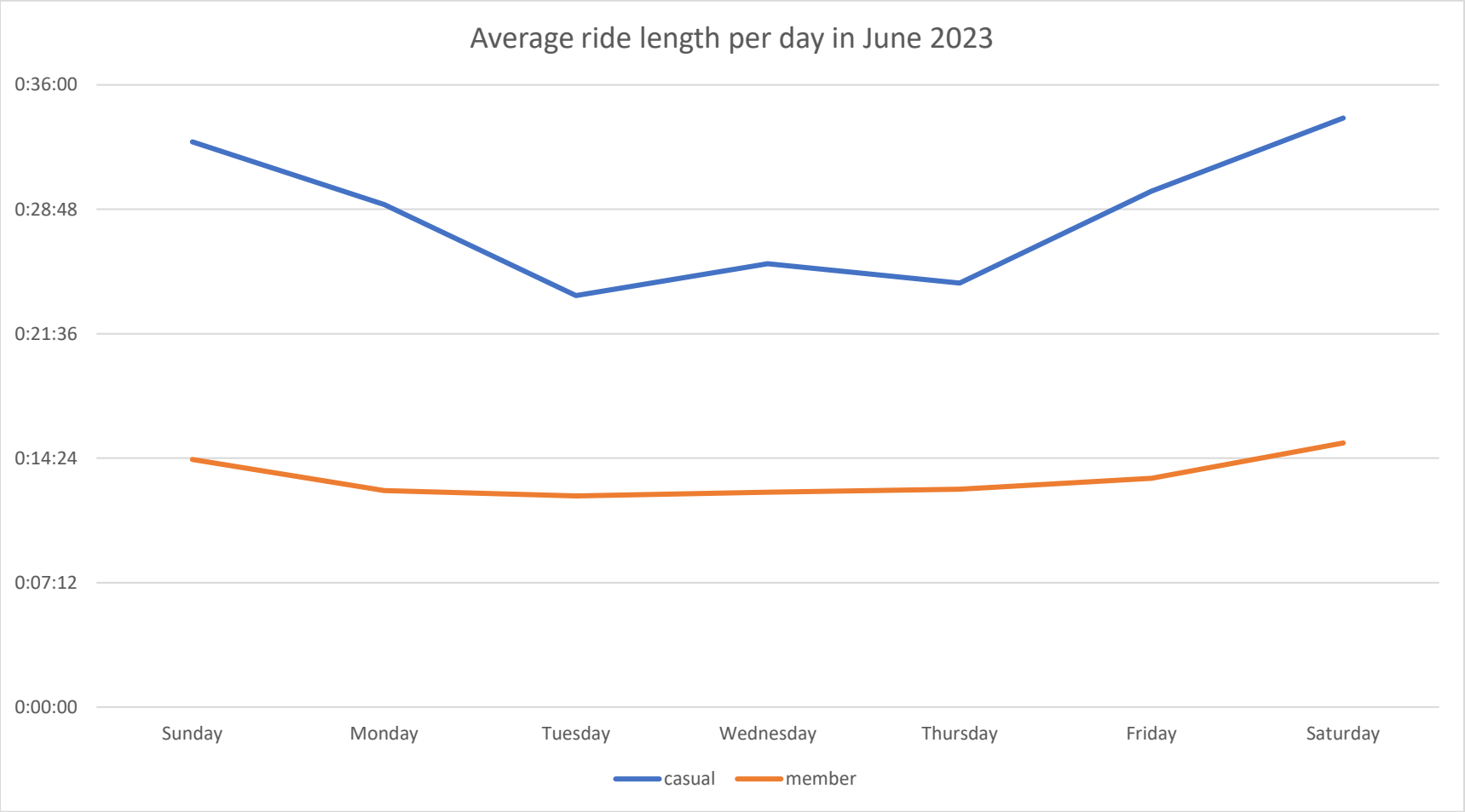

Figure 37. Line comparison Casual vs Member ride length per day in June 2023

Rideable types

Electric is more popular than classic bikes. They both follow a similar usage pattern throughout the week, suggesting that users interchange between them frequently.

Docked bikes were barely used throughout the month.

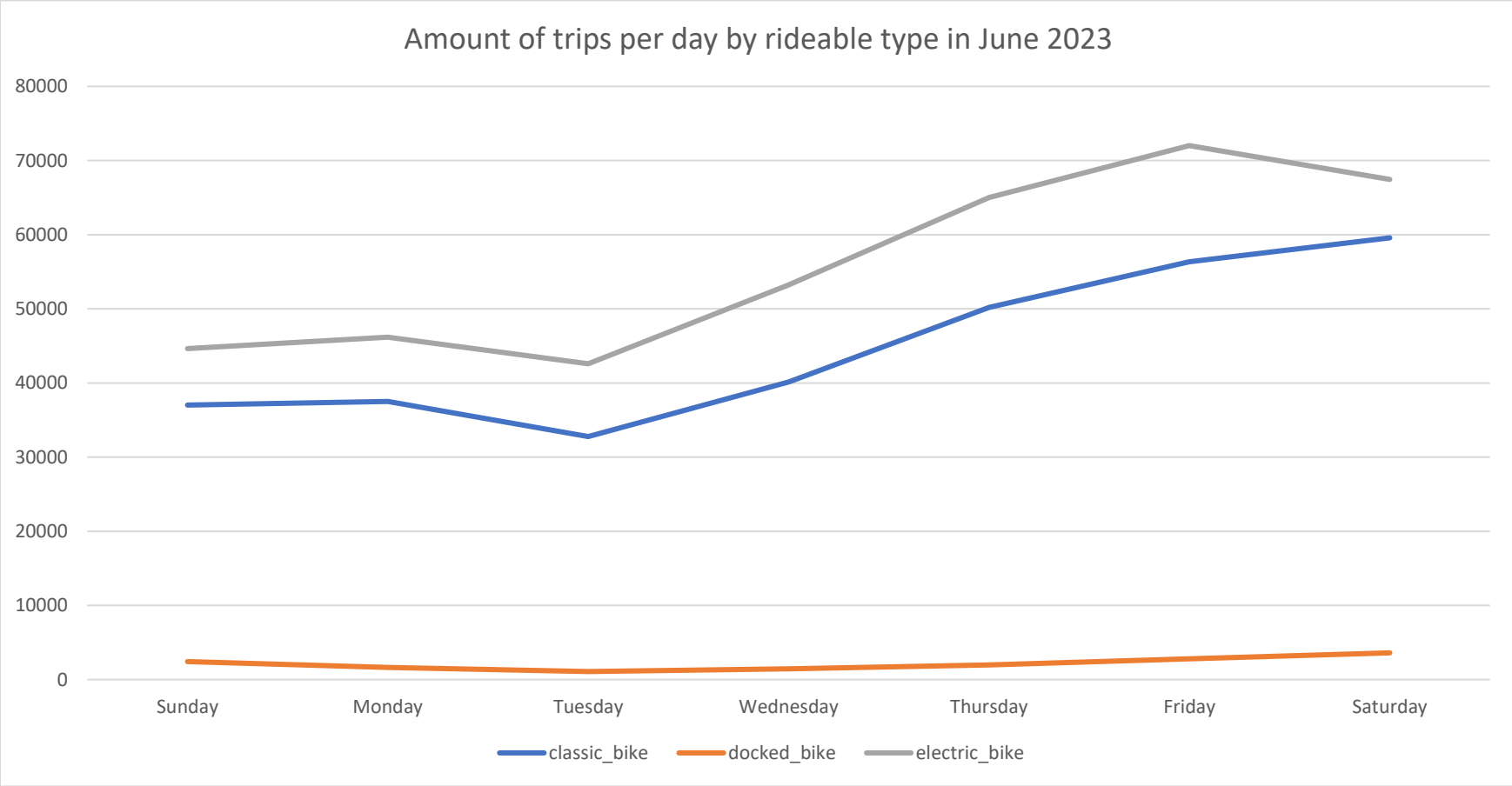

Figure 38. Line comparison number of rides per day by rideable type in June 2023

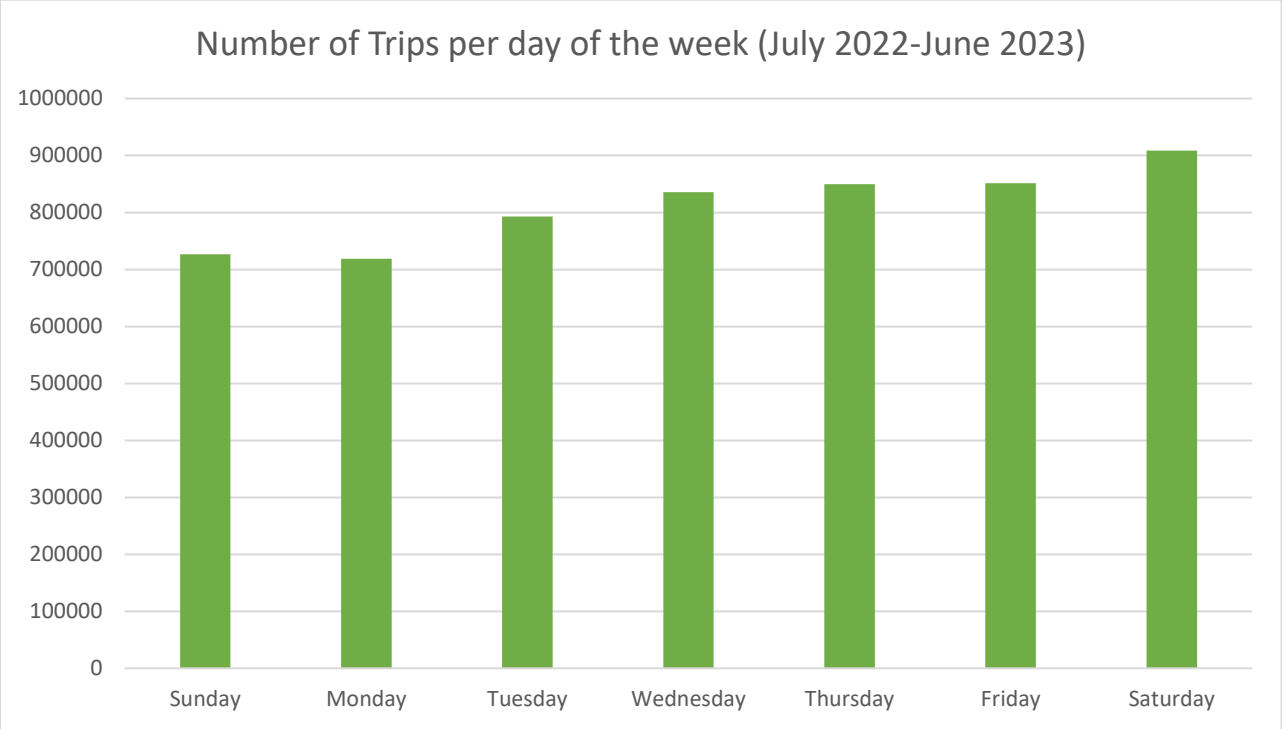

Figure 39. Bar chart for the number of trips taken per day of the week (July 2022-June 2023)

Figure 40. Number of trips by the time of the day

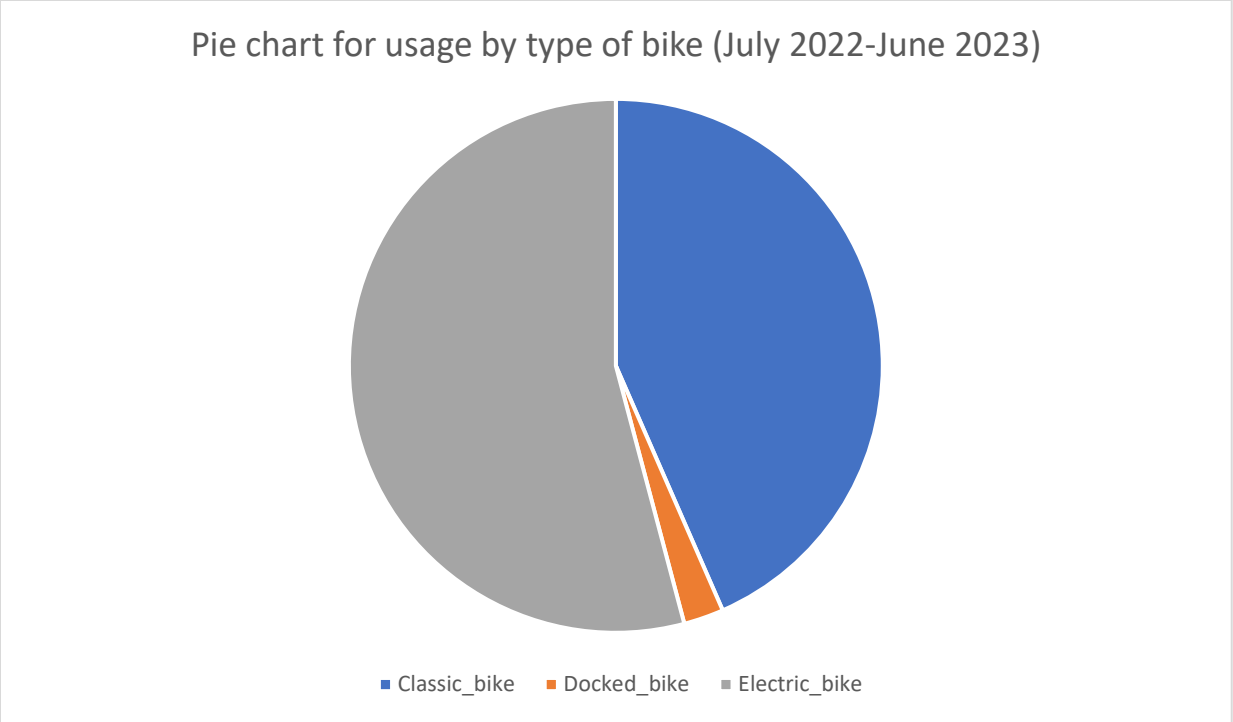

Figure 41. Pie chart of the usage by type of bike (July 2022-June 2023)

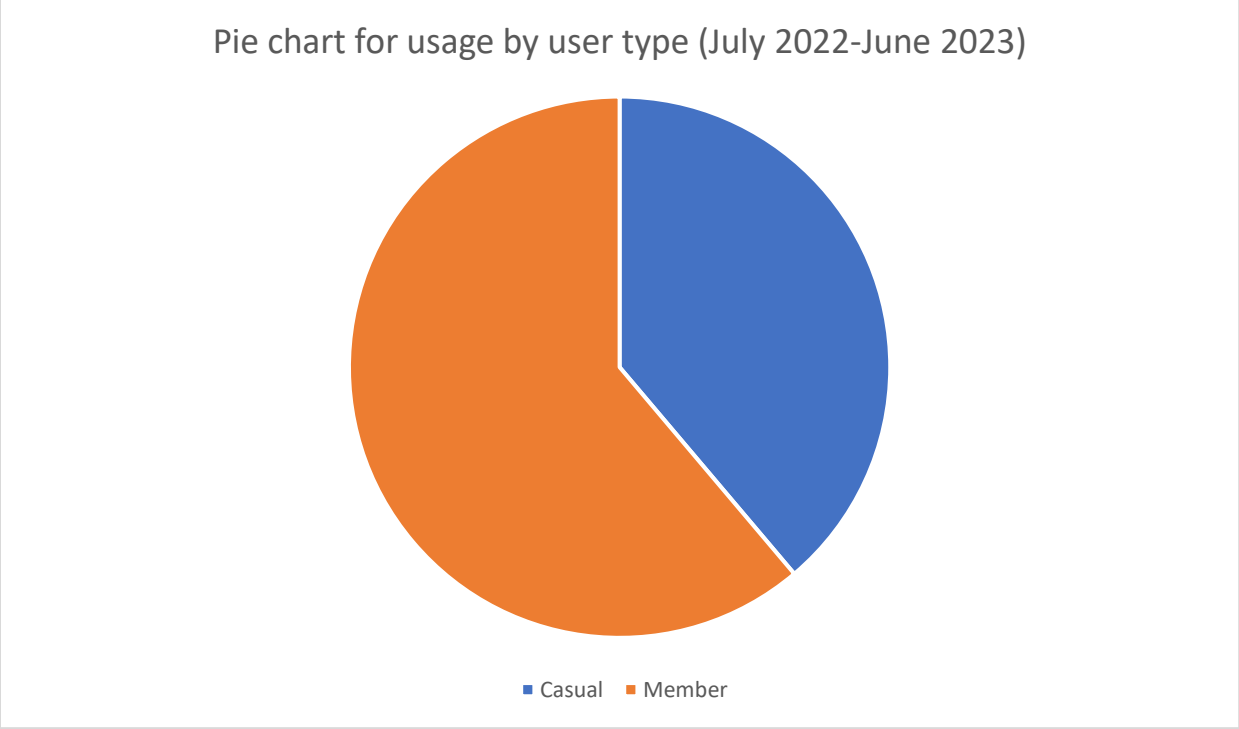

Figure 42. Pie chart of the usage by user type (July 2022-June 2023)

Appendix A: Data Source

The bike trip data used in this case study was obtained from Coursera, Divy Trip Data. The dataset includes anonymized information on bike trips taken by both members and casual riders from July 2022 to June 2023. It contains variables such as trip dates, start and end locations, trip durations, bike types, and user types.

Appendix B: Data Preprocessing

Before conducting the analysis, the raw data underwent thorough preprocessing to ensure its accuracy and consistency. Data cleaning and transformation techniques were applied to handle missing values, remove duplicates, and format the data for analysis.

Appendix C: R Packages and Code

The R programming language was utilized for data analysis and visualization. This appendix provides the code snippets used in the analysis and highlights the R packages employed for various tasks, including data manipulation, visualization, and statistical analysis.

- **lubridate:** Date-time manipulation.
- **tidyverse:** Collection of R packages of the following.
    - **dplyr:** Data manipulation.
    - **tidyr:** Data manipulation.
    - **ggplot2:** Data visualization.
    - **readr:** Reading CSV files.

```
##R code##
library(tidyverse)  # Includes ggplot2, dplyr, tidyr and readr
```

```
library(lubridate)  # For date format
library(scales)     # Allows the use of commas for better readability of numbers

# Read cvs for each month and plug into dataframes (df)
jul22_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Jul22_df.csv")
aug22_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Aug22_df.csv")
sep22_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Sep22_df.csv")
oct22_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Oct22_df.csv")
nov22_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Nov22_df.csv")
dec22_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Dec22_df.csv")
jan23_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Jan23_df.csv")
feb23_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Feb23_df.csv")
mar23_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Mar23_df.csv")
```

```
apr23_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Apr23_df.csv")
may23_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/May23_df.csv")
jun23_df <- read.csv("C:/Users/Omar/OneDrive/Desktop/Coursera/Data Analytics Professional Certificate/Case Study/Jun23_df.csv")

# Put all 12 month in one 7/22 -6/23 (1 year) data frame
# N.B. rbinds needs to have the same number of columns USE df = subset(df, select = -c(X,X.1,X.2) ) TO REMOVE extra Columns
cyclistic_df <- rbind(jul22_df, aug22_df, sep22_df, oct22_df, nov22_df, dec22_df, jan23_df, feb23_df, mar23_df, apr23_df, may23_df, jun23_df)

# Put the dataframe in another dataframe (data) for cleaning
data <- cyclistic_df

# Change dates format and make it date instead of string FORMAT %m/%d/%Y %H:%M for (mm/dd/yyyy hh:mm)
# For started_at
data$started_at <- as.POSIXct(data$started_at, format = "%m/%d/%Y %H:%M" )

# For ended_at
```

```
data$ended_at <- as.POSIXct(data$ended_at, format = "%m/%d/%Y %H:%M" )

# Calculate time difference in minutes (diff = ended_at - started_at)
data$time_difference_mins <- as.numeric(difftime(data$ended_at, data$started_at, units =
"mins"))

# Creates a day field from day started with FORMAT %Y-%m-%d for yyyy-mm-dd)
data$day <- format(data$started_at, format = "%Y-%m-%d")

# Turn started_at date into to Sunday-Saturday
data$weekday <- weekdays(data$started_at)

# Turn started_at to January-June   Format %B gives full name of the month of the date
data$month <- format(data$started_at, format = "%B")

# When did it start categorize them into Morning, Afternoon, Evening and Night
day_time_start <- function(time) {
  if (is.na(time)) {
    return(NA)
  }

  hour <- hour(time)
```

```
  if (hour >= 5 && hour < 12) {
    return("Morning")
  } else if (hour >= 12 && hour < 17) {
    return("Afternoon")
  } else if (hour >= 17 && hour < 21) {
    return("Evening")
  } else {
    return("Night")
  }
}

# Create a new field for time trip starts
data$day_time_start <- sapply(data$started_at, day_time_start)

# When did it end categorize them into Morning, Afternoon, Evening and Night
day_time_end <- function(time) {

  if (is.na(time)) {
    return(NA)
  }
  hour <- hour(time)
```

```
  if (hour >= 5 && hour < 12) {
    return("Morning")
  } else if (hour >= 12 && hour < 17) {
    return("Afternoon")
  } else if (hour >= 17 && hour < 21) {
    return("Evening")
  } else {
    return("Night")
  }
}

# Create a new field for time trip ends
data$day_time_end <- sapply(data$ended_at, day_time_end)

Season <- function(month) {
  if (month %in% c("December", "January", "February")) {
    return("Winter")
  } else if (month %in% c("March", "April", "May")) {
    return("Spring")
  } else if (month %in% c("June", "July", "August")) {
    return("Summer")
  } else {
```

```
    return("Fall")
  }
}

# Create a new field for time trip starts
data$Season <- sapply(data$month, Season)

# CLEANING DATA
# Remove null data and put it in a new dataframe (Cleaned_df)
cleaned_df <- na.omit(data)

# Remove duplicates
cleaned_df <- cleaned_df[!duplicated(cleaned_df), ]

# Remove if ride duration is less than 0
cleaned_df <- cleaned_df[!(cleaned_df$time_difference_mins <= 0), ]

# Remove useless fields
cleaned_df <- cleaned_df[, !(names(cleaned_df) %in% c("ride_id",
"start_station_name","end_station_name", "start_lat","start_lng","end_lat","end_lng"))]

# Remove useless data frames for faster analysis
```

```
remove(Jul22_df, aug22_df, sep22_df, oct22_df, nov22_df, dec22_df, jan23_df, feb23_df,
mar23_df, apr23_df, may23_df, jun23_df)

# NOW WORKING WITH CLEANED_DF FOR ANALYSIS

# Make the months in order
month_order <- c("July", "August", "September","October",
"November","December","January", "February", "March", "April", "May", "June")
cleaned_df$month <- factor(cleaned_df$month, levels = month_order)

# Make the weekdays in order
weekday_order <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday",
"Sunday")
cleaned_df$weekday <- factor(cleaned_df$weekday, levels = weekday_order)

# Make the Seasons in order
season_order <- c("Winter", "Spring", "Summer", "Fall")
cleaned_df$Season <- factor(cleaned_df$Season, levels = season_order)

#########################################################################

# Bar graph for day_time for each month in scientific notation
```

```
ggplot(cleaned_df, aes(month, fill = day_time_start),levels = month_order) +
  geom_bar(position= "dodge")

# Bar graph for amount of Trips Started by Time of Day Per Month
ggplot(cleaned_df, aes(x = month, fill = day_time_start),levels = month_order) +
  geom_bar(position = "dodge") +
  scale_y_continuous(labels = comma) +
  labs(x = "Month", y = "Number of Trips", fill = "Day Time Start") +
  theme_minimal() +
  ggtitle("Trips Started by Time of Day Per Month")


# Bar graph for amount of Trips Ended by Time of Day Per Month
ggplot(cleaned_df, aes(x = month, fill = day_time_end),levels = month_order) +
  geom_bar(position = "dodge") +
  scale_y_continuous(labels = comma) +
  labs(x = "Month", y = "Number of Trips", fill = "Day Time End") +
  theme_minimal() +
  ggtitle("Trips Ended by Time of Day Per Month")


# Faceted bar chart
```

```
long_df <- cleaned_df %>%
  pivot_longer(cols = c(day_time_start, day_time_end), names_to = "event", values_to =
"time_of_day")

ggplot(long_df, aes(x = time_of_day, fill = time_of_day)) +
  geom_bar() +
  facet_wrap(~ event, scales = "free_y") +
  labs(x = "Time of Day", y = "Number of Trips", fill = "Time of Day") +
  theme_minimal() +
  ggtitle("Trips Started and Ended by Time of Day")

##################################################

# Bar graph for amount of trips per month by seasons
ggplot(cleaned_df, aes(x = Season, fill = Season)) +
  geom_bar(position = "dodge", stat = "count") +
  labs(title = "Seasonal Bike Usage", x = "Seasons", y = "Number of Trips") +
  theme_minimal()


# Bar graph for day_time for each month but fill
ggplot(cleaned_df, aes(month, fill = day_time_start)) +
```

```
  geom_bar(position= "fill")



# Bar graph, which type of rides are used every month
ggplot(cleaned_df, aes(x = rideable_type)) +
  geom_bar() +
  scale_y_continuous(labels = comma) +
  facet_wrap(~month)

# NOW WORKING WITH TEST_DF FOR ANALYSIS
test_df<- cleaned_df

# How many are members and causals July 22 - June 23
test_df %>%
  group_by(member_casual) %>%
  count()



# How many rides based on type of ride
test_df %>%
  group_by(rideable_type) %>%
  count()
```

```
# Total rides by Casual and Member per month
test_df$month <- factor(test_df$month, levels = month_order)
test_df %>%
  group_by(member_casual, month) %>%
  count() %>%
  print(n = 24)


# Rides per month
test_df$month <- factor(test_df$month, levels = month_order)
test_df %>%
  group_by(month) %>%
  count() %>%
  print(n = 24)


# Total rides by Casual and Member per weekday
test_df$weekday <- factor(test_df$weekday, levels = weekday_order)
test_df %>%
  group_by(member_casual, weekday) %>%
```

```
  count()



# Total rides per weekday
test_df$weekday <- factor(test_df$weekday, levels = weekday_order)
test_df %>%
  group_by(weekday) %>%
  count()

# Rides per month
test_df$Season <- factor(test_df$Season, levels = season_order)
test_df %>%
  group_by(Season) %>%
  count() %>%
  print(n = 24)

# When the rides start by member_casual
day_time_order <- c("Morning", "Afternoon", "Evening", "Night")
test_df$day_time_start <- factor(test_df$day_time_start, levels = day_time_order)
test_df %>%
  group_by(member_casual, day_time_start) %>%
  count()
```

```
# When the rides start for all
test_df$day_time_start <- factor(test_df$day_time_start, levels = day_time_order)
test_df %>%
  group_by(day_time_start) %>%
  count()


# When the rides end by member_casual
day_time_order <- c("Morning", "Afternoon", "Evening", "Night")
test_df$day_time_end <- factor(test_df$day_time_end, levels = day_time_order)
test_df %>%
  group_by(member_casual, day_time_end) %>%
  count()


# When the rides end for all
test_df$day_time_end <- factor(test_df$day_time_end, levels = day_time_order)
test_df %>%
  group_by(day_time_end) %>%
  count()
```

```
# Ride length by member_casual
test_df <- test_df %>%
  mutate(category = case_when(
    time_difference_mins < 60 ~ "Less than an hour",
    time_difference_mins >= 60 & time_difference_mins <= 1440 ~ "Between an hour and 24 hours",
    time_difference_mins > 1440 ~ "More than 24 hours"
  ))
category_order <- c("Less than an hour", "Between an hour and 24 hours", "More than 24 hours")
test_df$category <- factor(test_df$category, levels = category_order)
test_df %>%
  group_by(member_casual, category) %>%
  count()


# Ride length for all
test_df <- test_df %>%
  mutate(category = case_when(
    time_difference_mins < 60 ~ "Less than an hour",
```

```
    time_difference_mins >= 60 & time_difference_mins <= 1440 ~ "Between an hour and 24
hours",
    time_difference_mins > 1440 ~ "More than 24 hours"
  ))
category_order <- c("Less than an hour", "Between an hour and 24 hours", "More than 24
hours")
test_df$category <- factor(test_df$category, levels = category_order)
test_df %>%
  group_by(category) %>%
  count()



# Avg ride length for all casuals and members grouped by casual_member
test_df %>%
  group_by(member_casual) %>%
  summarize(avg_ride_length_mins = mean(time_difference_mins))



# Avg ride length for all casuals and members grouped by casual_member and rideable_type
test_df %>%
  group_by(member_casual, rideable_type) %>%
  summarize(avg_ride_length_mins = mean(time_difference_mins))
```

```
# Avg ride length for all casuals and members grouped by casual_member and rideable_type
test_df %>%
  group_by(rideable_type) %>%
  summarize(avg_ride_length_mins = mean(time_difference_mins))



# Histogram of ride length by Member Type
ggplot(test_df, aes(x = category, fill = member_casual)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "Histogram of ride length by Member Type",
       x = "Ride Length",
       y = "Number of trips") +
  theme_minimal() +
  scale_y_continuous(labels = comma)



# Bar graph for total count of members and casuals
ggplot(test_df, aes(x = member_casual, fill = member_casual)) +
  geom_bar() +
  labs(title = "Total trips comparison between Members and Casuals",
```

```
      x = "Member/Casual",
      y = "Count") +
  theme_minimal()+
  scale_y_continuous(labels = comma)


# Histogram graph for ride_length for Members and Casuals
ggplot(test_df, aes(x = time_difference_mins, fill = member_casual)) +
  geom_histogram(binwidth = 10, position = "dodge", color = "black") +
  labs(title = "Histogram graph for ride_length for Members and Casuals",
      x = "Time Difference (minutes)",
      y = "Frequency") +
  theme_minimal() +
  xlim(0, 60)  # Set the desired x-axis range here


# Violin Plot of ride length (1 hour) for Members and Casuals
ggplot(test_df, aes(x = member_casual, y = time_difference_mins, fill = member_casual)) +
  geom_violin() +
  labs(title = "Violin Plot of ride length (1 hour) for Members and Casuals",
      x = "Member/Casual",
      y = "Ride length (minutes)") +
```

```
  theme_minimal() +
  ylim(0, 60)  # Set the desired y-axis range here
```

[Note: The R code provided above showcases the initial data analysis and visualization process conducted by the author, Omar Dissouki. The code helps in understanding the trends and patterns observed in bike usage for members and casual riders from January to June 2023. The code can be further optimized and refined for future data analysis and reporting purposes.]

Appendix D: Visualizations

This section presents a collection of visualizations used to illustrate the key findings from the data analysis. Various plots, such as bar charts, line graphs, and heatmaps, were used to showcase the trends and patterns observed in bike usage throughout the year.

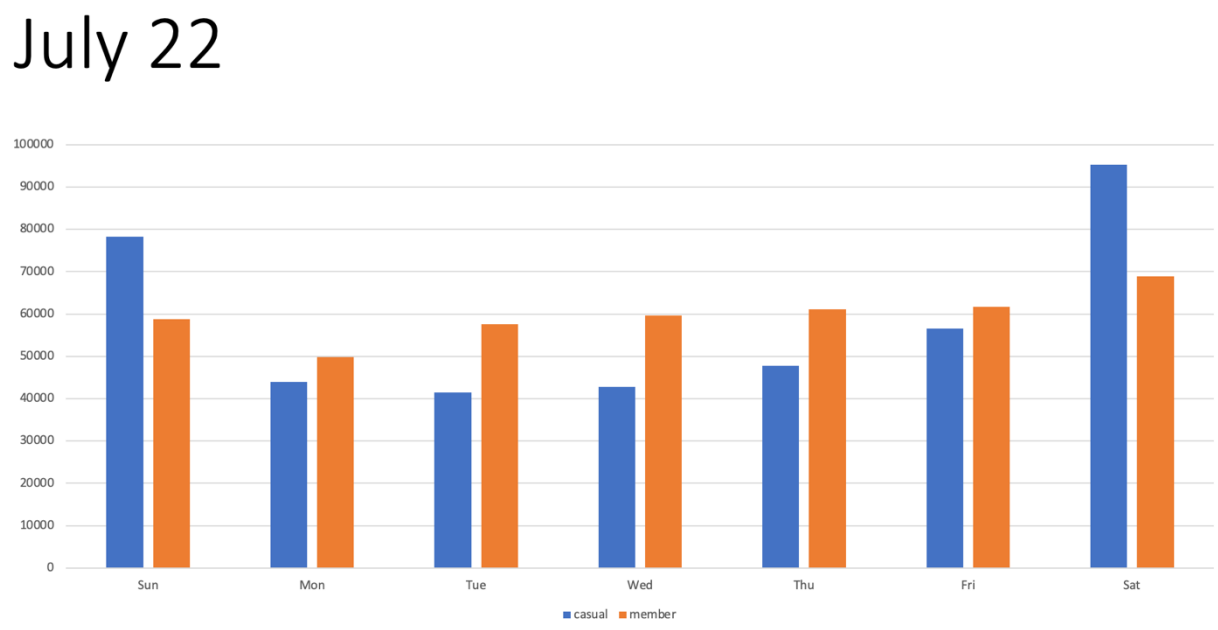

Figure 43 Number of trips taken in July 2022 by Members and Casual riders

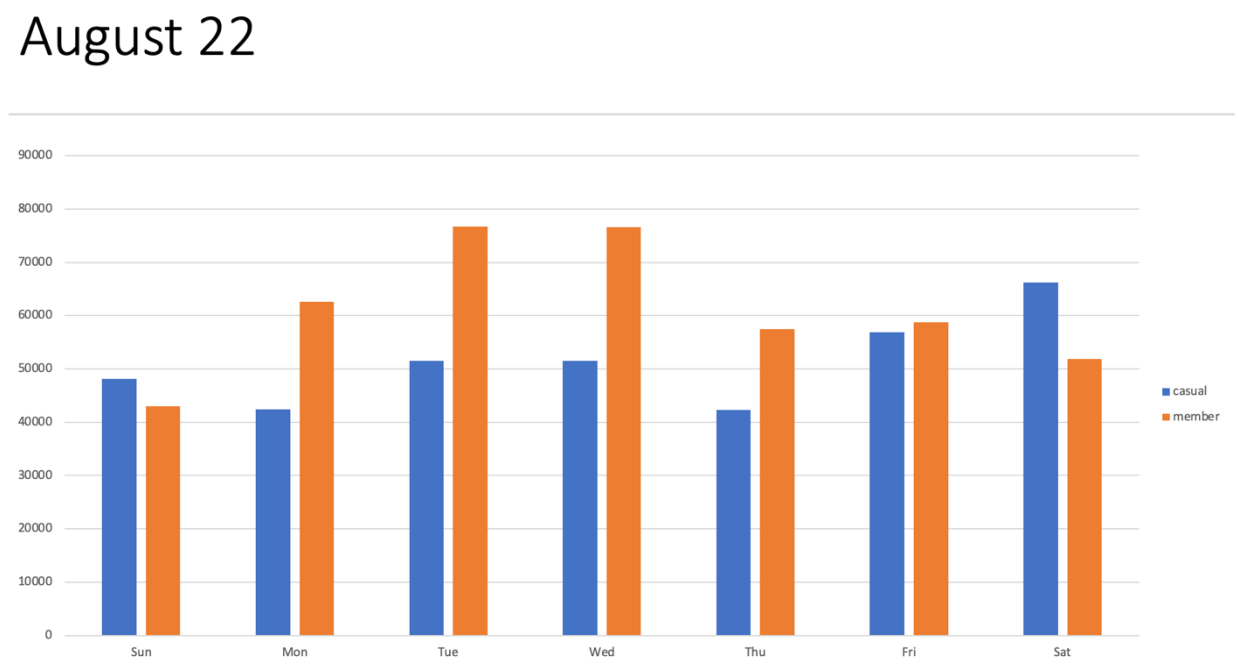

Figure 44 Number of trips taken in August 2022 by Members and Casual riders

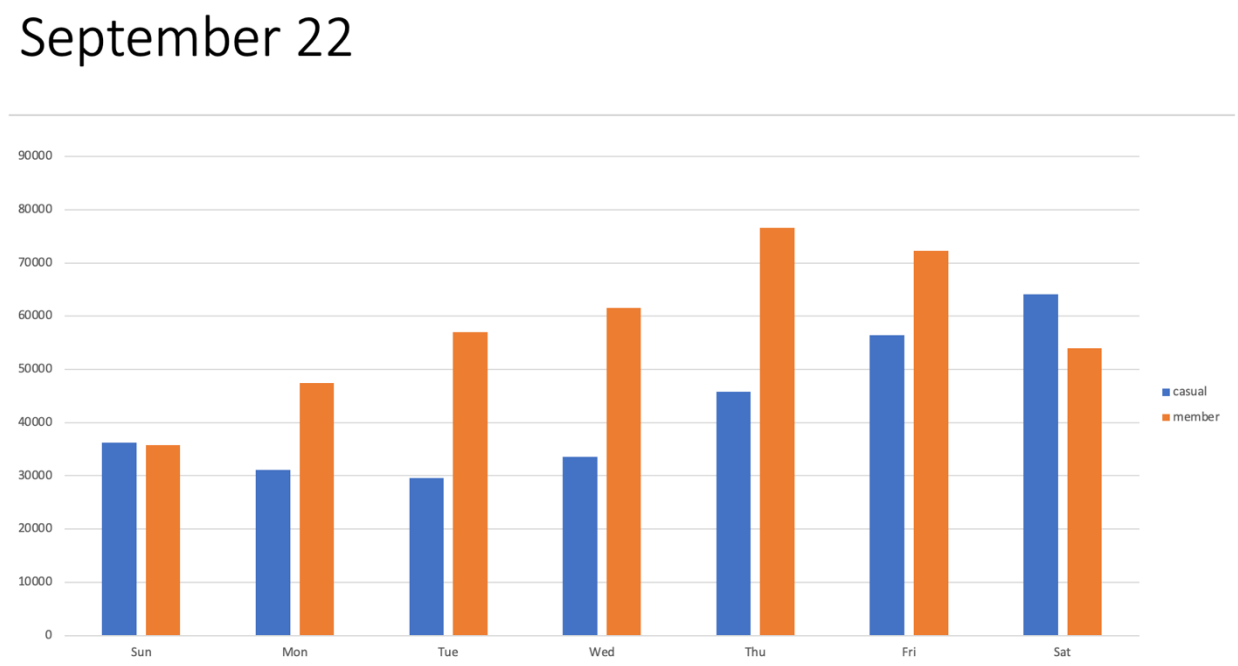

Figure 45 Number of trips taken in September 2022 by Members and Casual riders

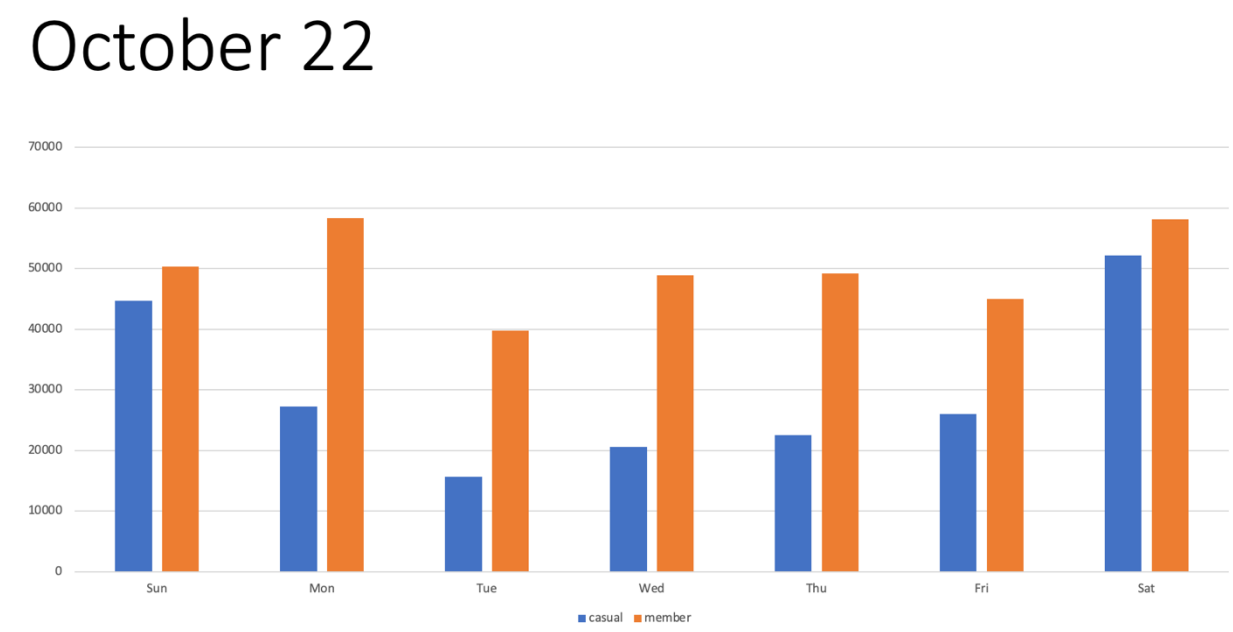

Figure 46 Number of trips taken in October 2022 by Members and Casual riders

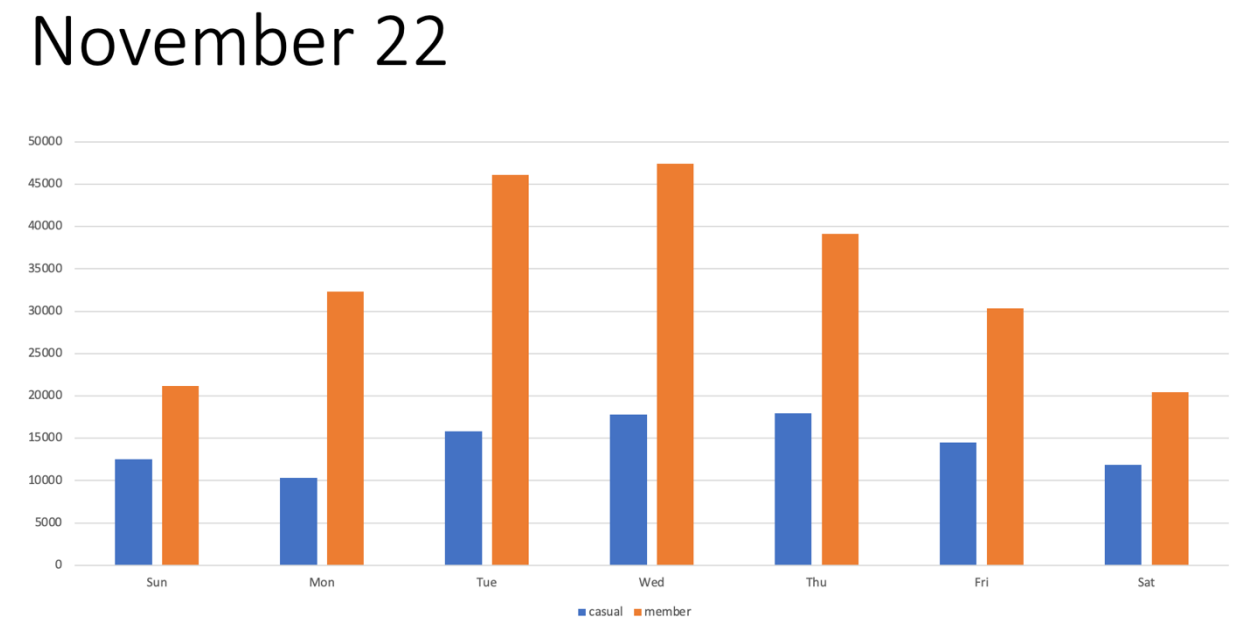

Figure 47 Number of trips taken in November 2022 by Members and Casual riders

December 22

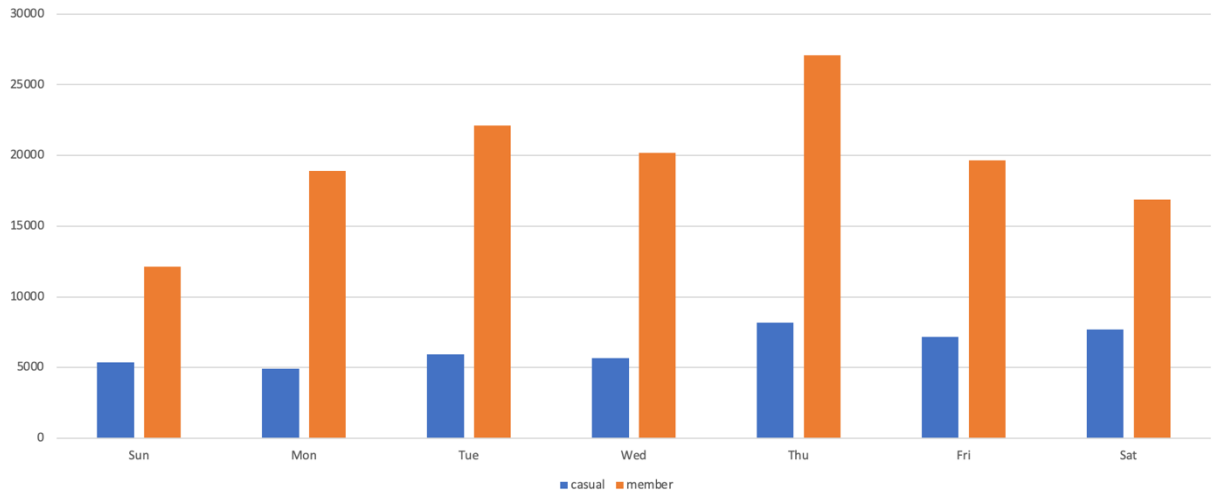

Figure 48 Number of trips taken in December 2022 by Members and Casual riders

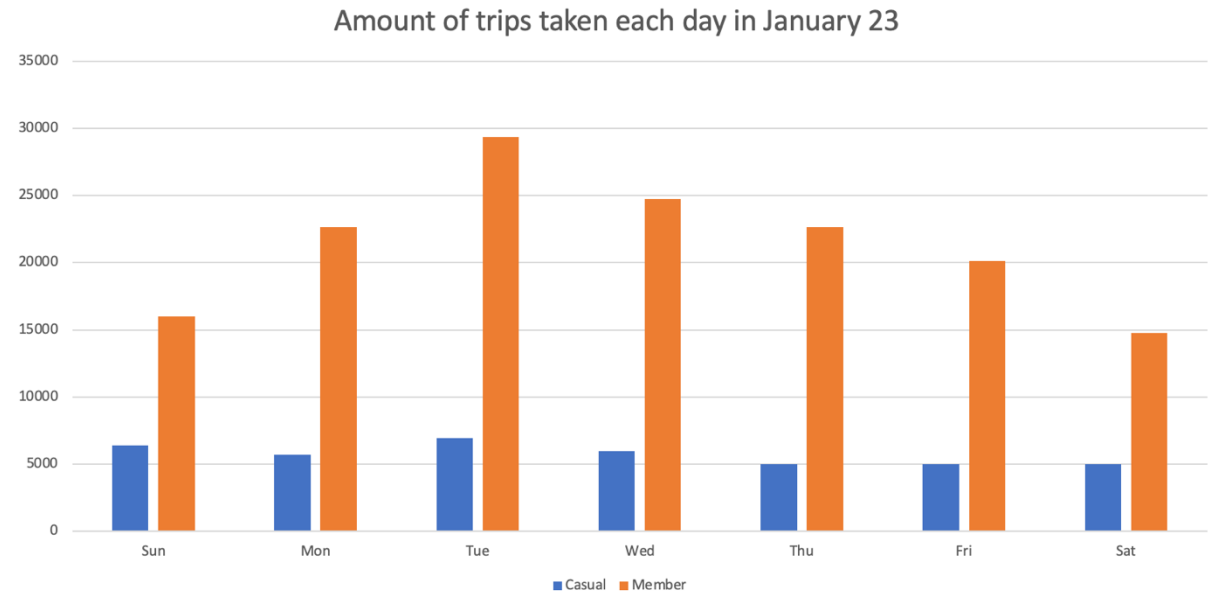

Figure 49 Number of trips taken in January 2023 by Members and Casual riders

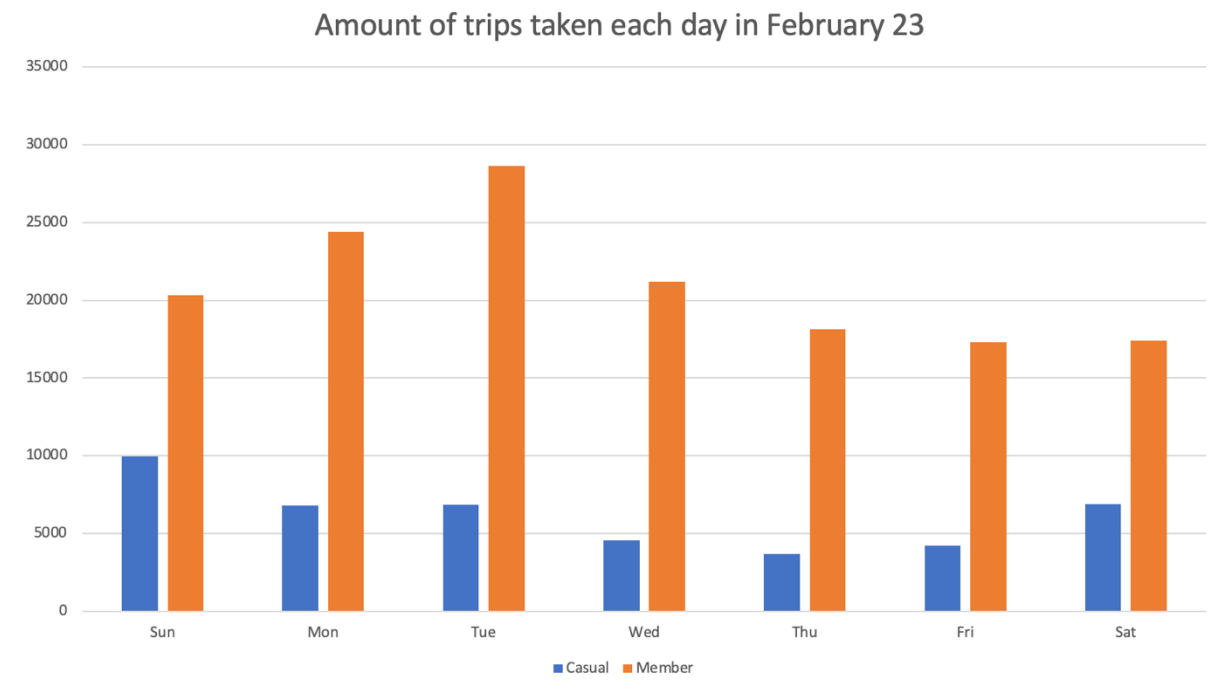

Figure 50 Number of trips taken in February 2023 by Members and Casual riders

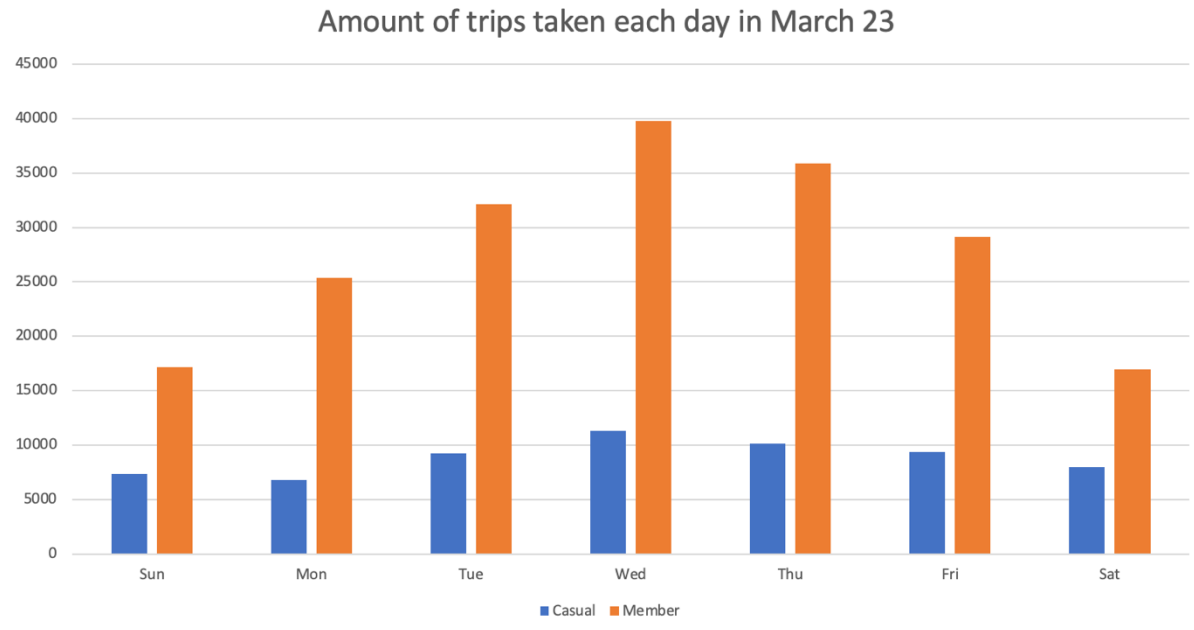

Figure 51 Number of trips taken in March 2023 by Members and Casual riders

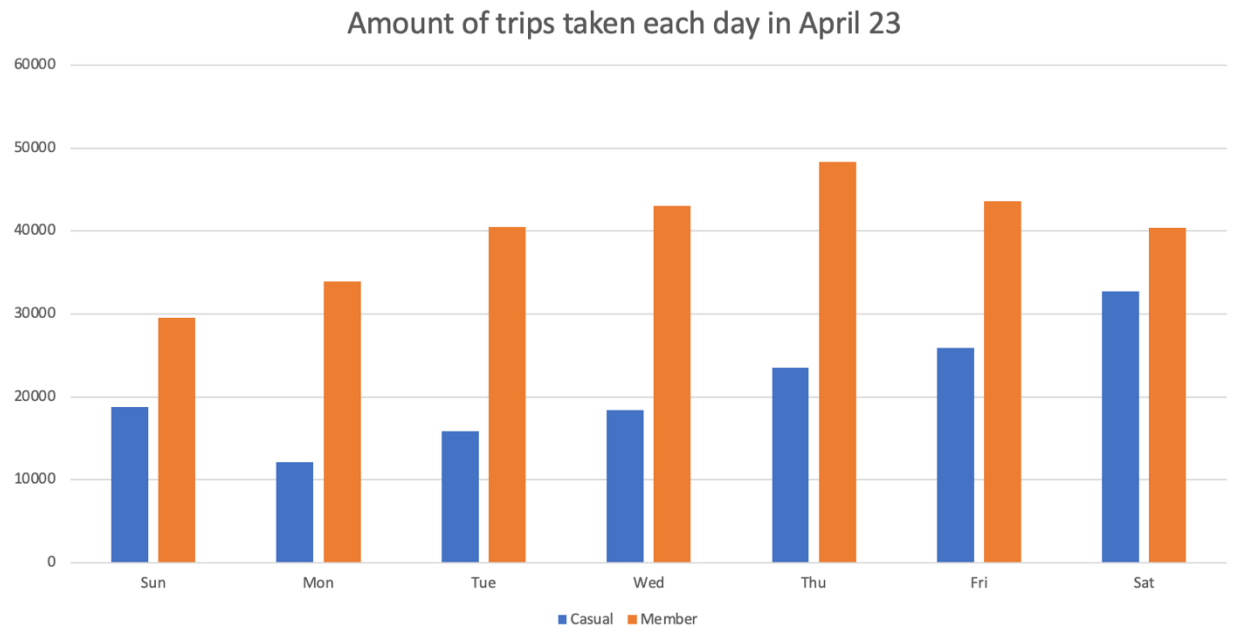

Figure 52 Number of trips taken in April 2023 by Members and Casual riders

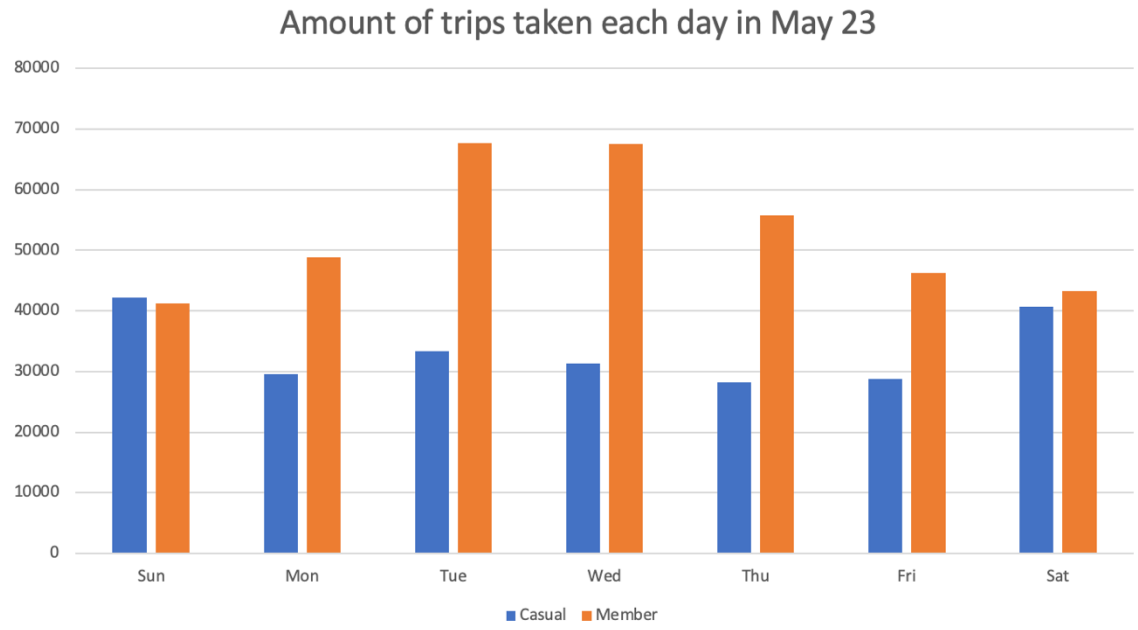

Figure 53 Number of trips taken in May 2023 by Members and Casual riders

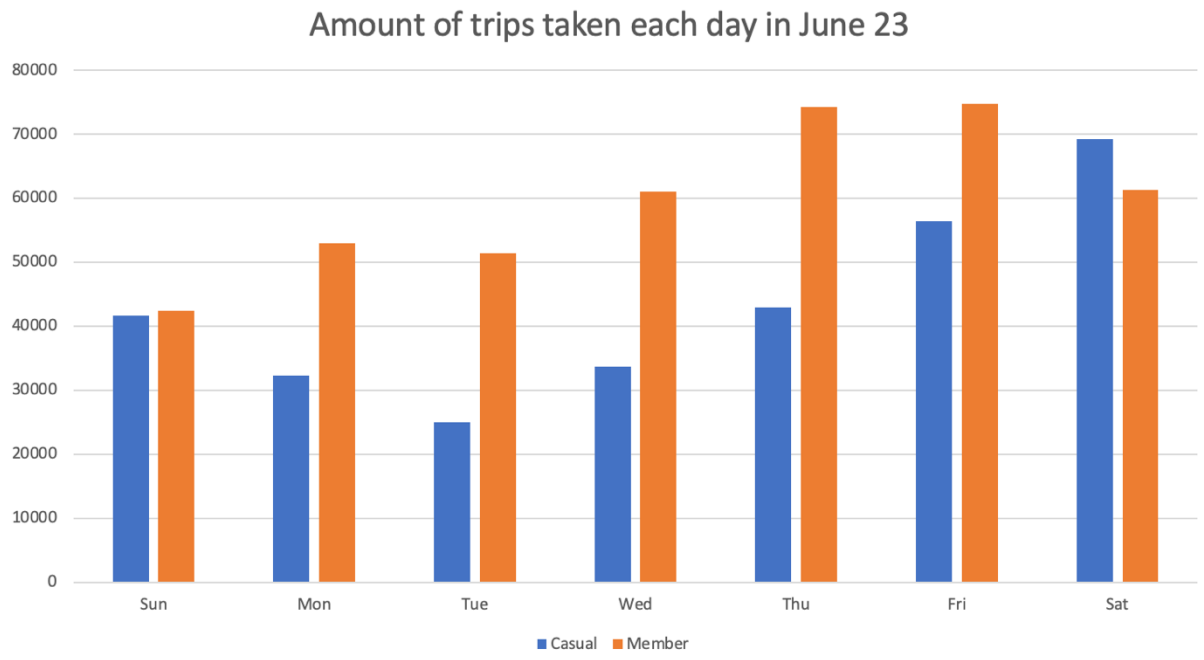

Figure 54 Number of trips taken in June 2023 by Members and Casual riders

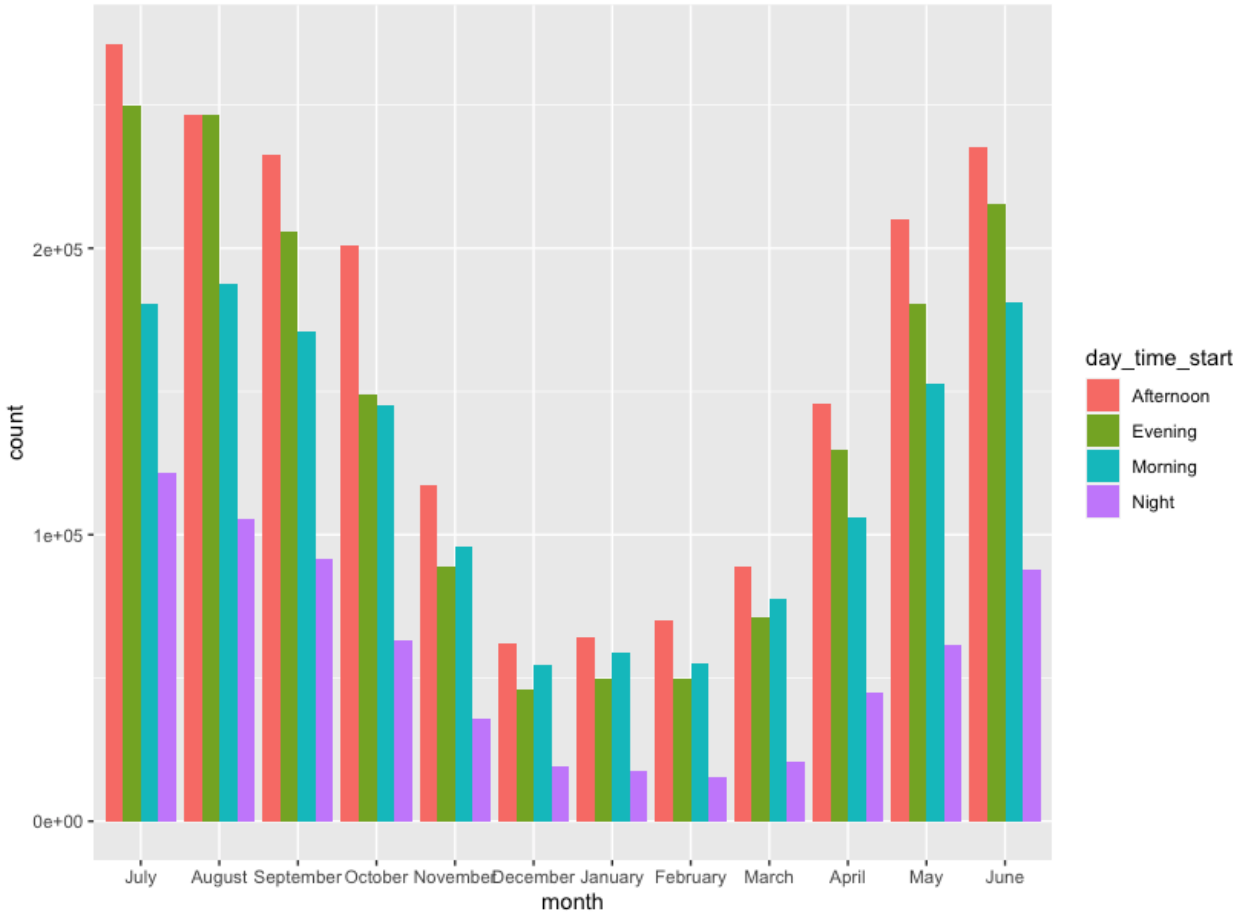

Figure 55 Histogram showing the frequency of riders during each 4 period in the day for each month

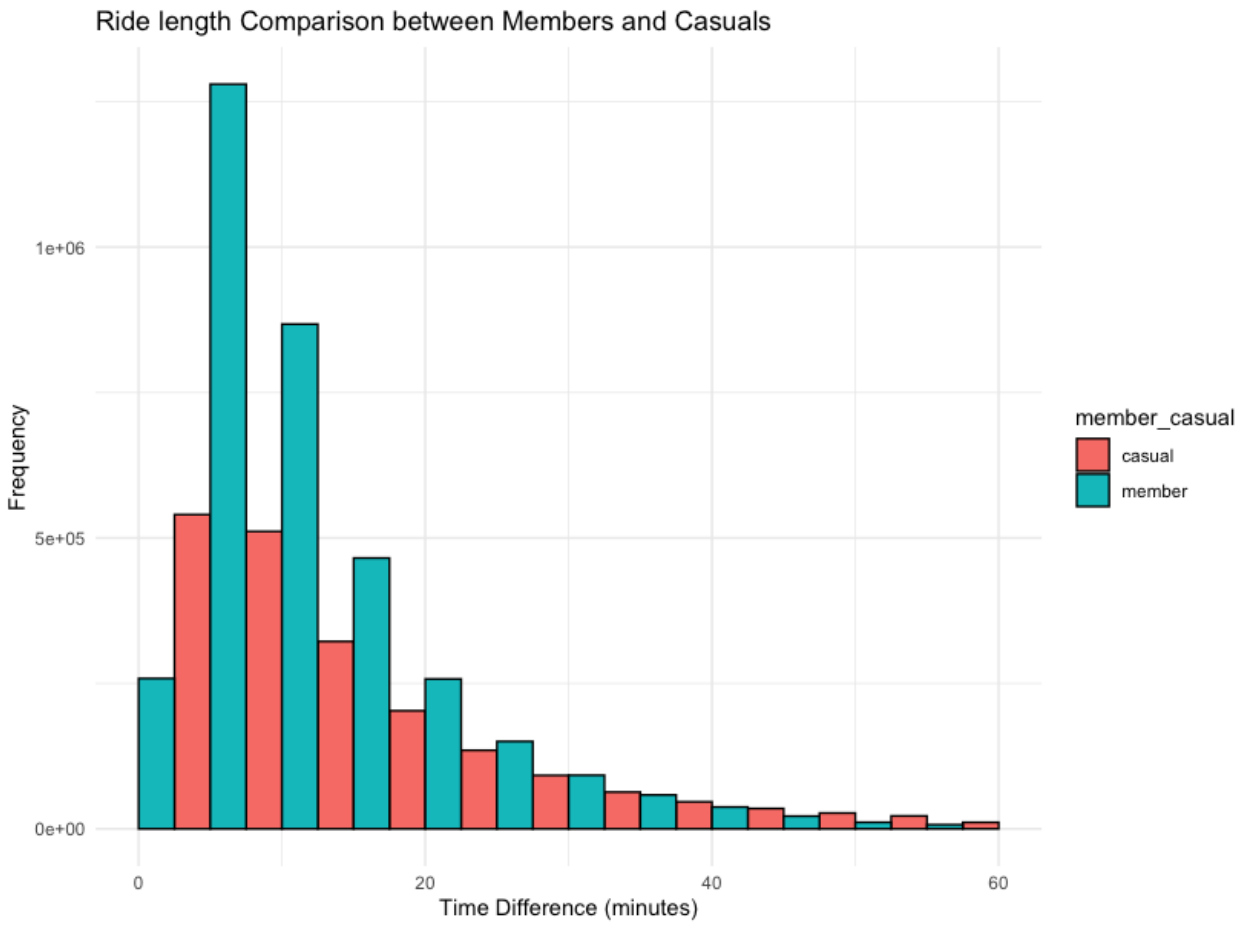

Figure 56 Histogram showing the ride length comparison for Members and Casual riders

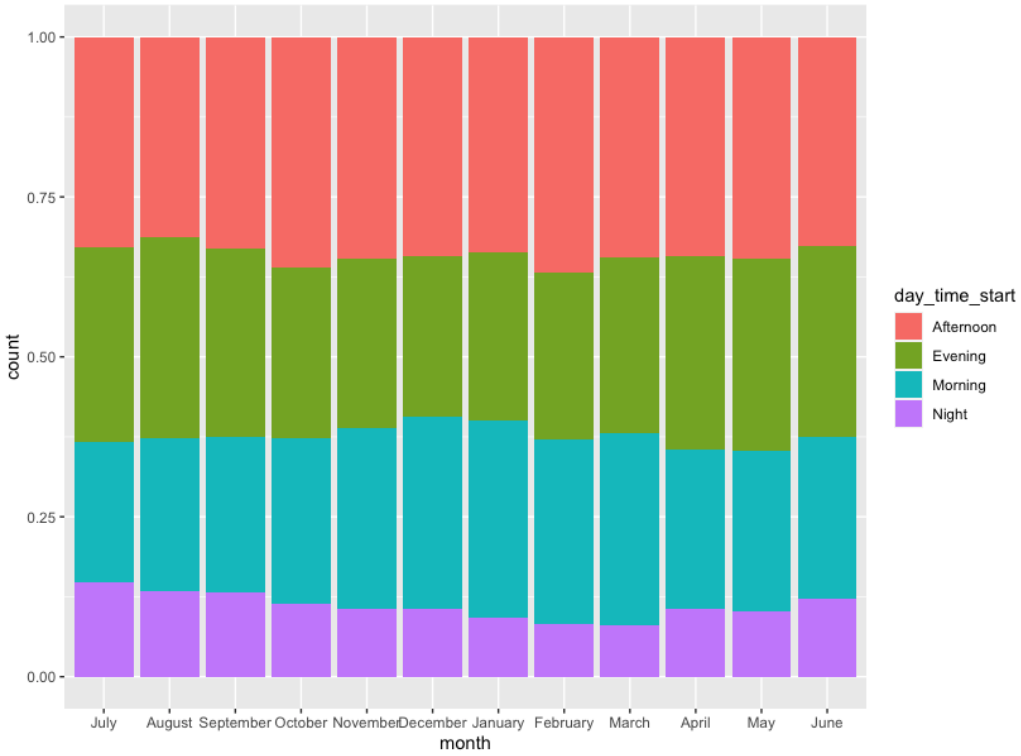

Figure 57 Bar graph showing the percentage of each of the 4 times of the day for each month

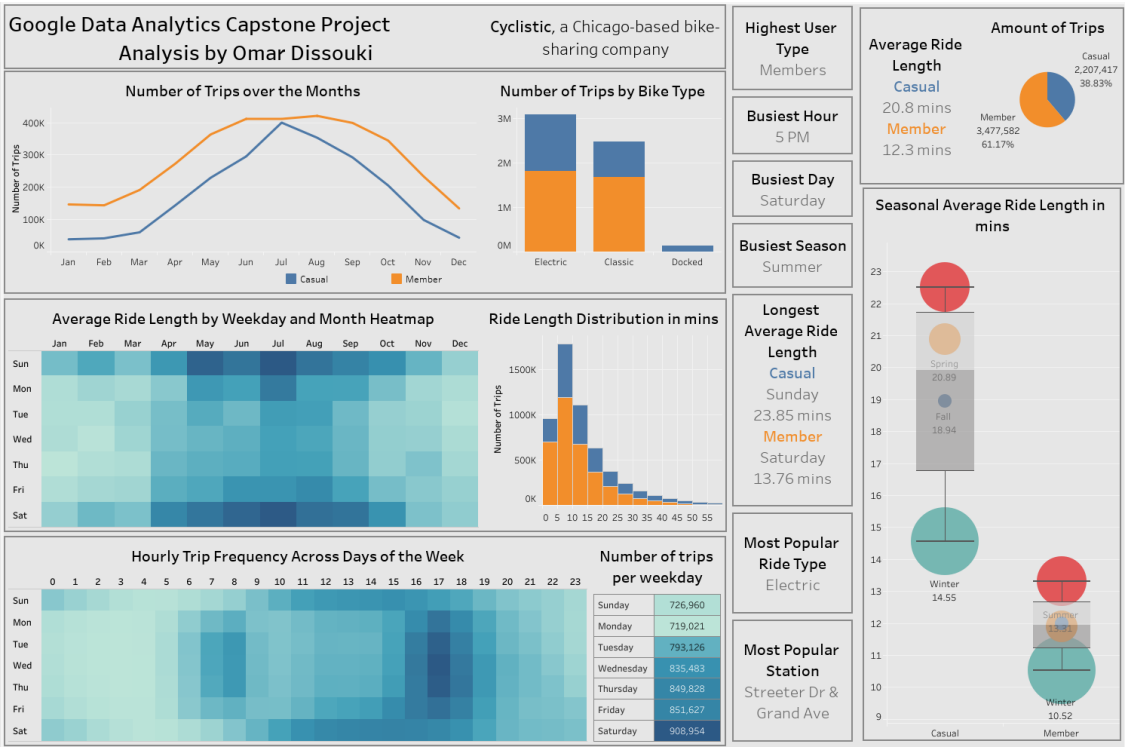

Figure 58. Tableau Dashboard

For interactive dashboard, click [here]

Appendix E: Limitations

Every analysis has its limitations, and this case study is no exception. In this section, I outline the potential limitations and challenges encountered during the analysis. This will help in interpreting the results and understanding the scope of the findings.

- This analysis was done on 12 months only
- The brand knowledge is increasing so more people are knowing the brand everyday
- Didn't take into consideration any promotion happening at any time that might have attracted more rides at a certain time
- Didn't take into consideration any of the advertisement by marketing team

Appendix F: References

Coursera Google Data Analytics Professional Certification

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

ChatGPT for wording and coding help