

Cyclistic Bike Share Company

Sanjay chauhan 2/11/2022

```
library(tidyverse)
## -- Attaching packages ----- tidyverse
1.3.1 --
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.1      v forcats 0.5.1
## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter() ##
x dplyr::lag()      masks stats::lag()
library(janitor)
##
## Attaching package: 'janitor'
## The following objects are masked from 'package:stats':
##
##      chisq.test, fisher.test library(lubridate)
##
## Attaching package: 'lubridate'
## The following objects are masked from 'package:base':
##
##      date, intersect, setdiff, union
library(ggplot2) library(scales)
##
## Attaching package: 'scales'
## The following object is masked from 'package:purrr':
##
##      discard
## The following object is masked from 'package:readr': ##
##      col_factor
library(forcats) library(stringr)
library(dplyr)
knitr::opts_chunk$set(echo=FALSE)
```

Introduction

This report explores the Cyclistic bike share company performance during the month of January, 2021. Specifically it focuses on “How do members and casual riders use bikes differently”. Moreover, the other elements and factors impacting performance of the company also kept in focus while following analysis process. *Data Source:* <https://bit.ly/3BggueF>

Exploratory Data Analysis

Now, we look at the data provided by the company for analysis

```
## Rows: 96834 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (9): ride_id, rideable_type, started_at, ended_at, start_station_name,
s...
## dbl (4): start_lat, start_lng, end_lat, end_lng
##
## i Use `spec()` to retrieve the full column specification for this data. ##
i Specify the column types or set `show_col_types = FALSE` to quiet this
message.
## [1] 96834      13
```

Data wrangling process

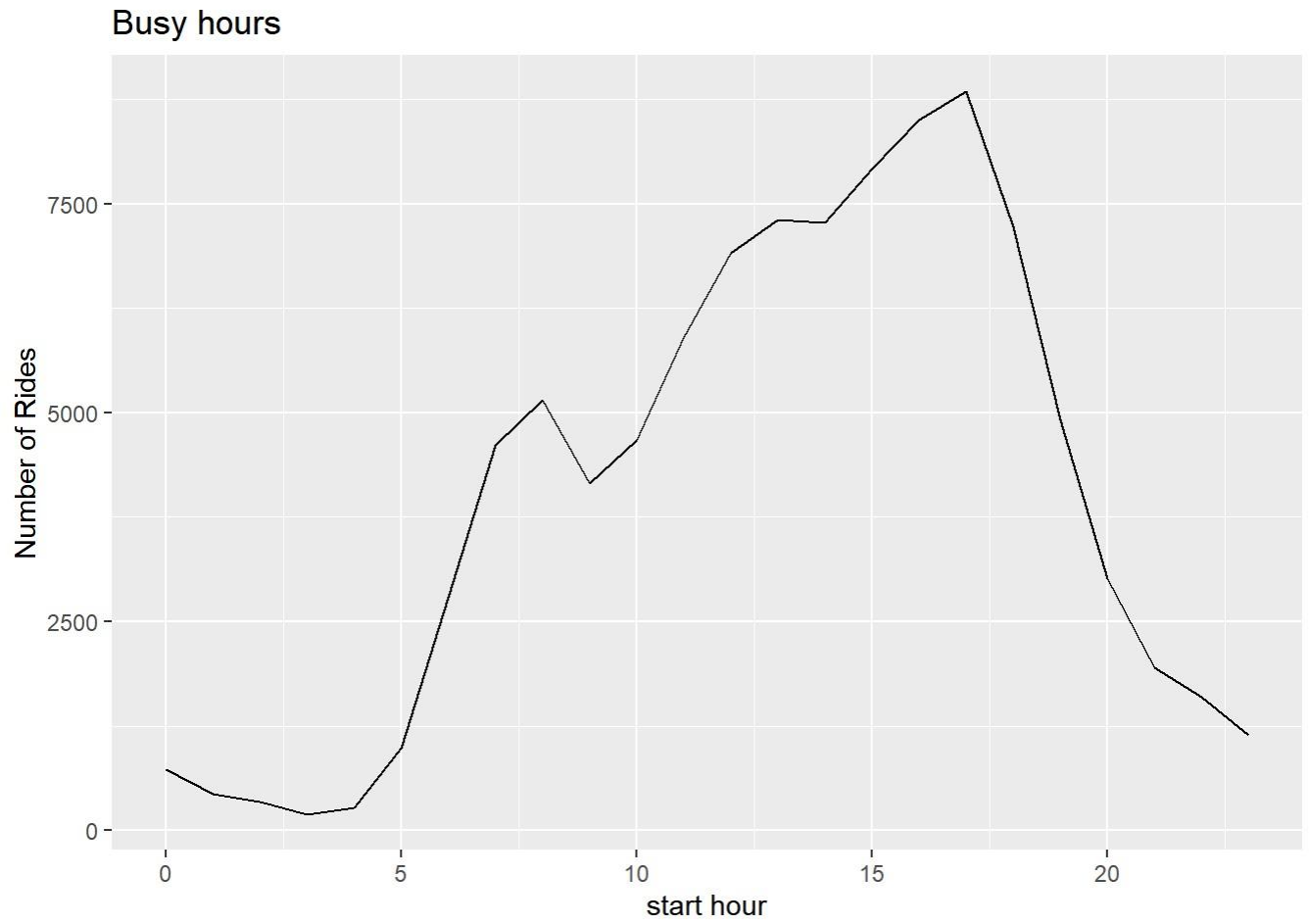
```
## # A tibble: 96,834 x 13
##   ride_id rideable_type started_at ended_at start_station_n~
start_station_id
##   <chr>    <chr>          <chr>      <chr>    <chr>          <chr>
## 1 E19E6F1B~ electric_bike 1/23/2021~ 1/23/20~ California Ave ~ 17660
## 2 DC88F20C~ electric_bike 1/27/2021~ 1/27/20~ California Ave ~ 17660
## 3 EC45C946~ electric_bike 1/21/2021~ 1/21/20~ California Ave ~ 17660
## 4 4FA453A7~ electric_bike 1/7/2021 ~ 1/7/202~ California Ave ~ 17660
## 5 BE5E8EB4~ electric_bike 1/23/2021~ 1/23/20~ California Ave ~ 17660
## 6 5D8969F8~ electric_bike 1/9/2021 ~ 1/9/202~ California Ave ~ 17660
## 7 09275CC1~ electric_bike 1/4/2021 ~ 1/4/202~ California Ave ~ 17660
## 8 DF7A32A2~ electric_bike 1/14/2021~ 1/14/20~ California Ave ~ 17660
## 9 C2EFC623~ electric_bike 1/9/2021 ~ 1/9/202~ California Ave ~ 17660
## 10 B9F73448~ classic_bike 1/24/2021~ 1/24/20~ California Ave ~ 17660
## # ... with 96,824 more rows, and 7 more variables: end_station_name <chr>,
## #   end_station_id <chr>, start_lat <dbl>, start_lng <dbl>, end_lat <dbl>,
## #   end_lng <dbl>, member_casual <chr>
## Rows: 96,834
## Columns: 13
## $ ride_id          <chr> "E19E6F1B8D4C42ED", "DC88F20C2C55F27F",
"EC45C94683~
## $ rideable_type    <chr> "electric_bike", "electric_bike",
"electric_bike", ~
## $ started_at       <dtm> 2021-01-23 16:14:00, 2021-01-27 18:43:00,
2021-01-~
## $ ended_at         <dtm> 2021-01-23 16:24:00, 2021-01-27 18:47:00,
2021-01-~
## $ start_station_name <chr> "California Ave & Cortez St", "California Ave &
Cor~
## $ start_station_id <chr> "17660", "17660", "17660", "17660", "17660",
"17660~
## $ end_station_name  <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "Wood St &
Augu~
## $ end_station_id    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "657",
"13258",~
## $ start_lat        <dbl> 41.90034, 41.90033, 41.90031, 41.90040,
41.90033, 4~
```

```
## $ start_lng      <dbl> -87.69674, -87.69671, -87.69664, -87.69666,
87.696~
## $ end_lat        <dbl> 41.89000, 41.90000, 41.90000, 41.92000,
41.90000, 4~
## $ end_lng        <dbl> -87.72000, -87.69000, -87.70000, -87.69000,
87.700~
## $ member_casual  <chr> "member", "member", "member", "member",
"casual", "~
```

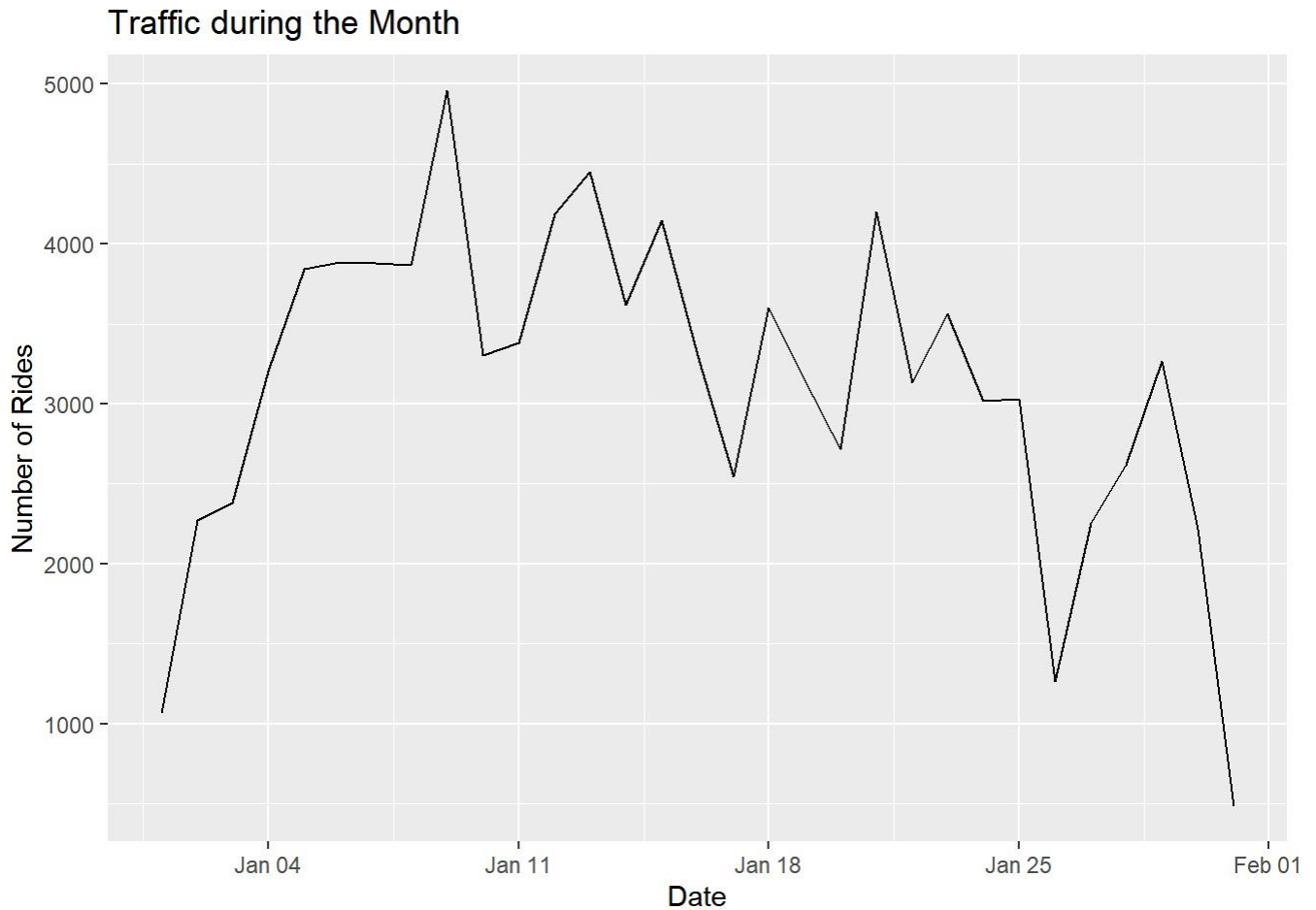
Additional columns for diving deep in the data

```
## Rows: 96,834
## Columns: 17
## $ ride_id        <chr> "E19E6F1B8D4C42ED", "DC88F20C2C55F27F",
"EC45C94683~
## $ rideable_type  <chr> "electric_bike", "electric_bike",
"electric_bike", ~
## $ started_at     <dtm> 2021-01-23 16:14:00, 2021-01-27 18:43:00,
2021-01-~
## $ ended_at       <dtm> 2021-01-23 16:24:00, 2021-01-27 18:47:00,
2021-01-~
## $ start_station_name <chr> "California Ave & Cortez St", "California Ave &
Cor~
## $ start_station_id <chr> "17660", "17660", "17660", "17660", "17660",
"17660~
## $ end_station_name <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "Wood St &
Augu~
## $ end_station_id  <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "657",
"13258",~
## $ start_lat       <dbl> 41.90034, 41.90033, 41.90031, 41.90040,
41.90033, 4~
## $ start_lng       <dbl> -87.69674, -87.69671, -87.69664, -87.69666,
87.696~
## $ end_lat         <dbl> 41.89000, 41.90000, 41.90000, 41.92000,
41.90000, 4~
## $ end_lng         <dbl> -87.72000, -87.69000, -87.70000, -87.69000,
87.700~
## $ member_casual  <chr> "member", "member", "member", "member",
"casual", "~
## $ start_hour      <int> 16, 18, 22, 13, 2, 14, 5, 15, 9, 19, 12, 15,
15, 15~
## $ end_hour        <int> 16, 18, 22, 13, 2, 15, 5, 15, 10, 19, 13, 15,
15, 1~
## $ start_date       <date> 2021-01-23, 2021-01-27, 2021-01-21, 2021-01-
07, 20~
## $ day_of_week     <chr> "Saturday", "Wednesday", "Thursday",
"Thursday", "S~
```

A visualization Projecting most busiest hours in a day



Plotting the busiest phase during the month



Member vs Casual behaviour Analysis

Now, we will declutter the data i.e. the a new data frame has been formed taking care the need of the further analysis process.

Comparison of Average, median, max, min and count between the members and the casual riders.

```
## rides2$member_casual rides2$duration
## 1 casual 26.48130 mins
## 2 member 12.11419 mins
## rides2$member_casual rides2$duration
## 1 casual 13 mins
## 2 member 9 mins
## rides2$member_casual rides2$duration
## 1 casual 19826 mins
## 2 member 1226 mins
## rides2$member_casual rides2$duration
## 1 casual 1 mins
## 2 member 1 mins
## # A tibble: 2 x 2
```

```
## member_casual      n
## <chr>              <int>
## 1 member           68345
## 2 casual           14629
```

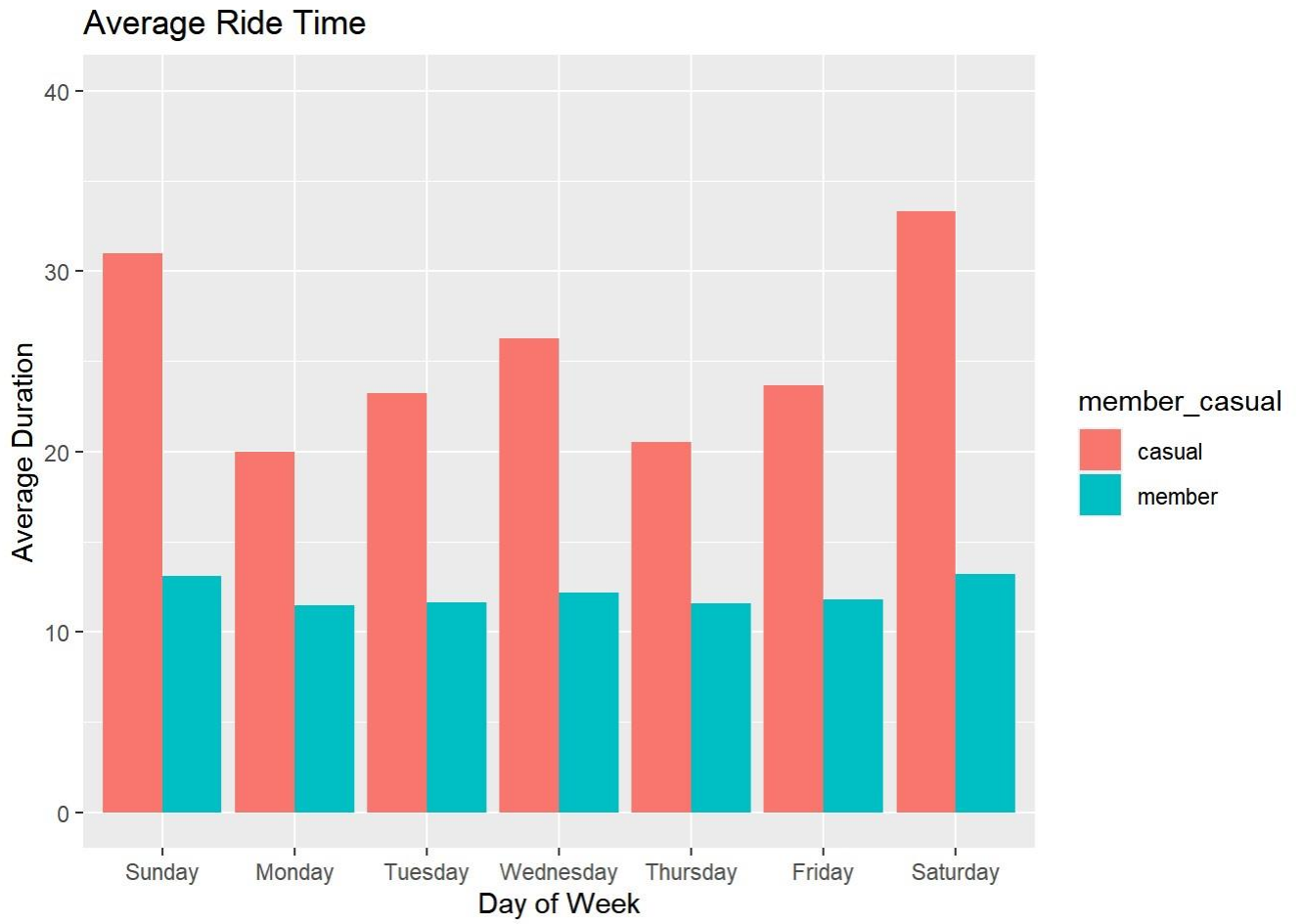
Weekly difference Average riding duration between members and casual riders.

```
##      rides2$member_casual rides2$day_of_week rides2$duration
## 1      casual            Sunday    30.96141 mins
## 2      member            Sunday    13.06661 mins
## 3      casual            Monday     19.96681 mins
## 4      member            Monday     11.45661 mins
## 5      casual            Tuesday    23.24932 mins
## 6      member            Tuesday    11.64322 mins
## 7      casual            Wednesday  26.25465 mins
## 8      member            Wednesday  12.19187 mins
## 9      casual            Thursday   20.49736 mins
## 10     member            Thursday   11.59947 mins
## 11     casual            Friday     23.64801 mins
## 12     member            Friday     11.81728 mins
## 13     casual            Saturday   33.30996 mins
## 14     member            Saturday   13.18710 mins A
```

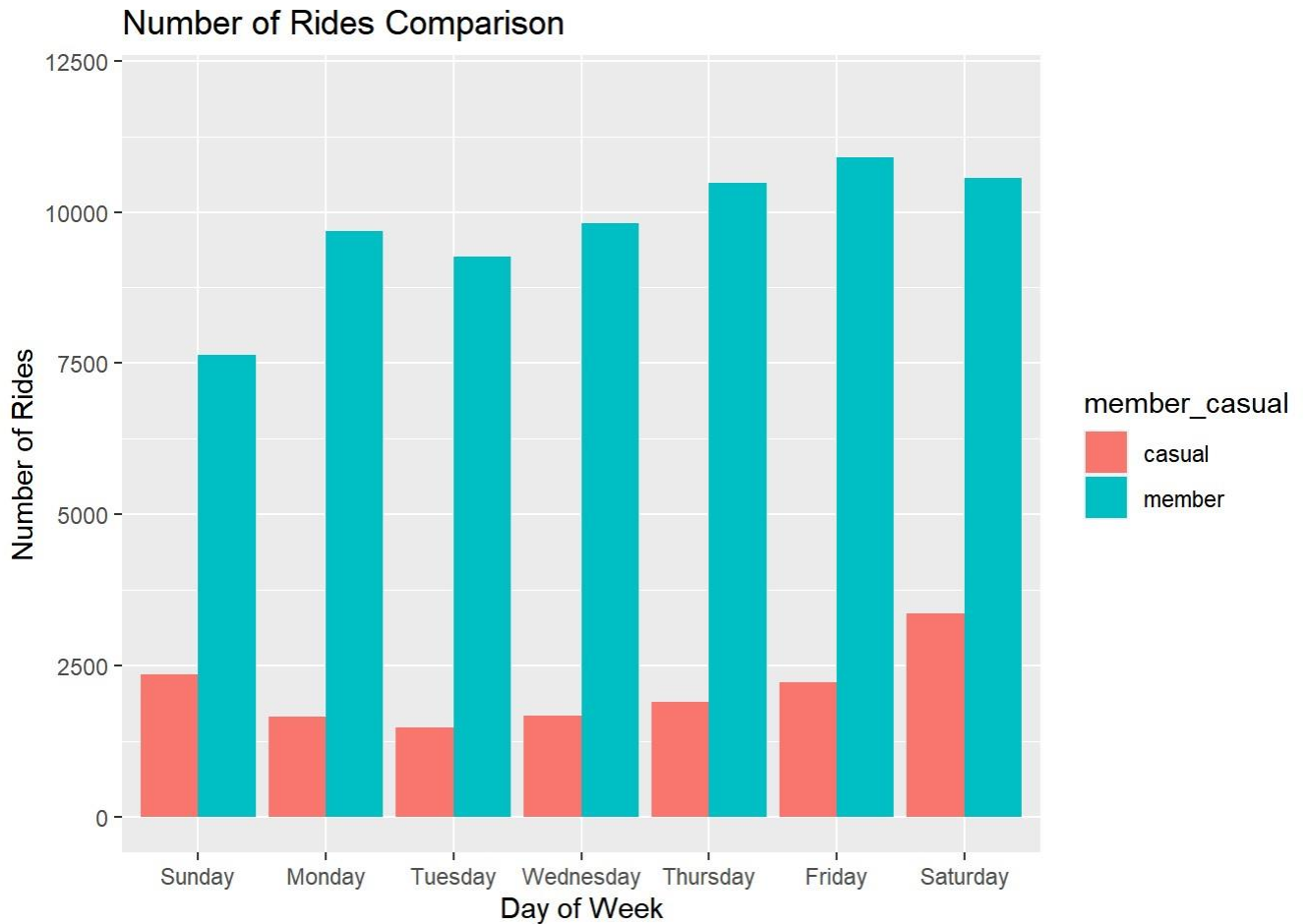
descriptive summary of Members and Casual Riders.

```
## `summarise()` has grouped output by 'member_casual'. You can override
using the `.groups` argument.
```

A visualization of average ride time difference of member and casual riders on the bases of week days.



A plot of Number of Rides taken by member and casual riders during weekdays.

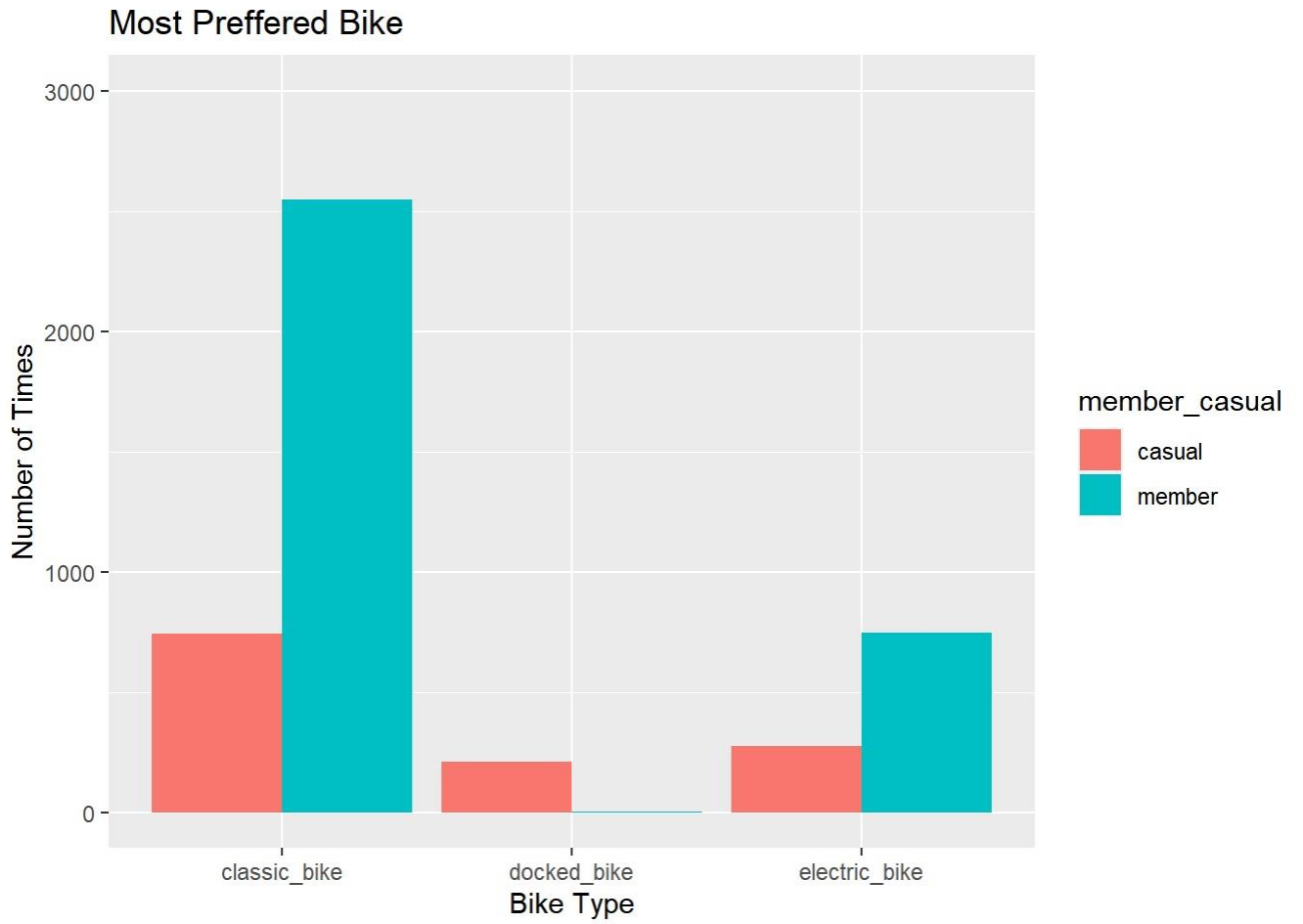


Bike Type Analysis

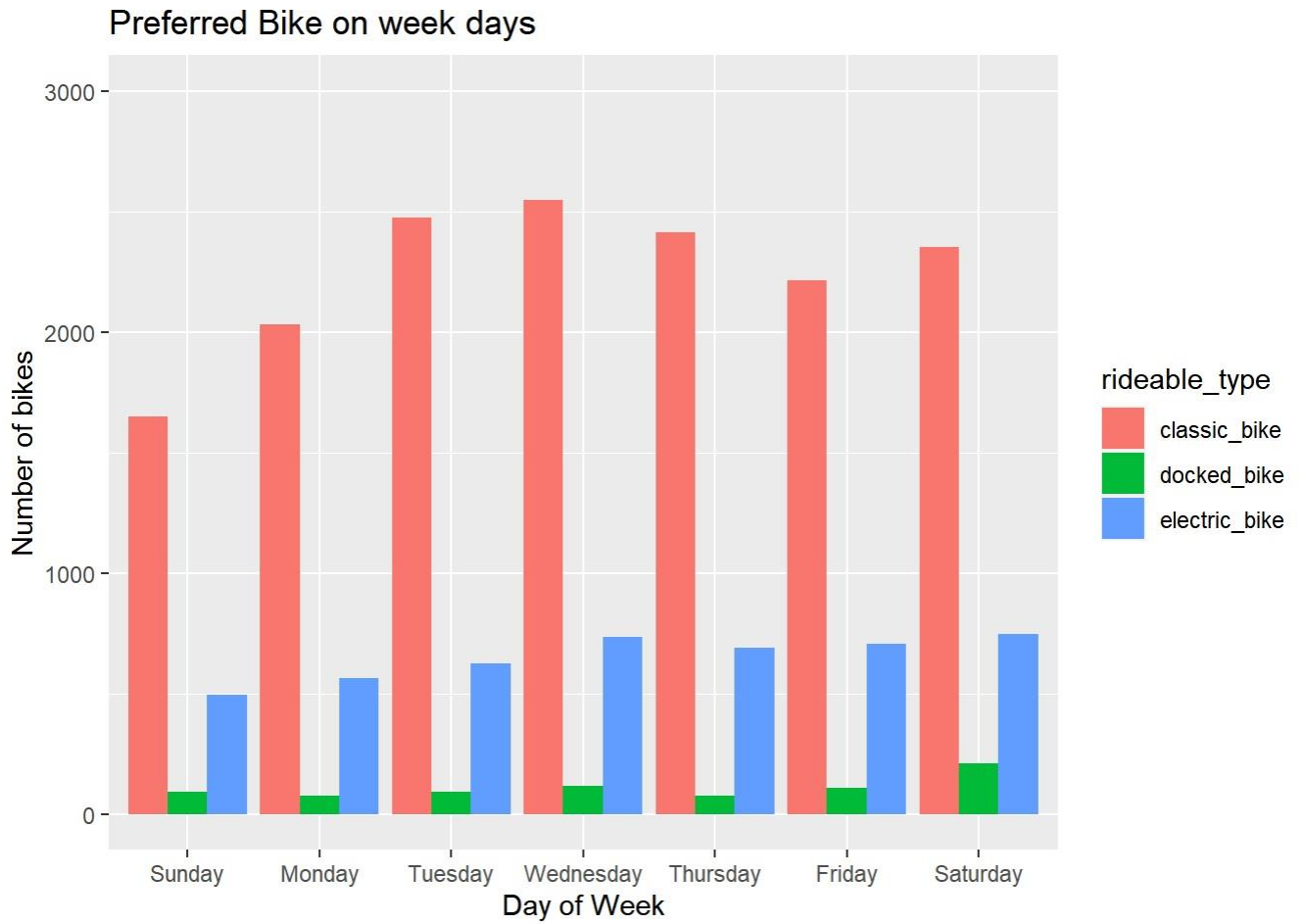
A descriptive summary showing comparison between each type of week.

```
## `summarise()` has grouped output by 'rideable_type', 'member_casual',  
'start_date'. You can override using the `.groups` argument.
```

A visualization showcasing the most preferred bike during the month also representing difference in bike preference between member and casual riders.

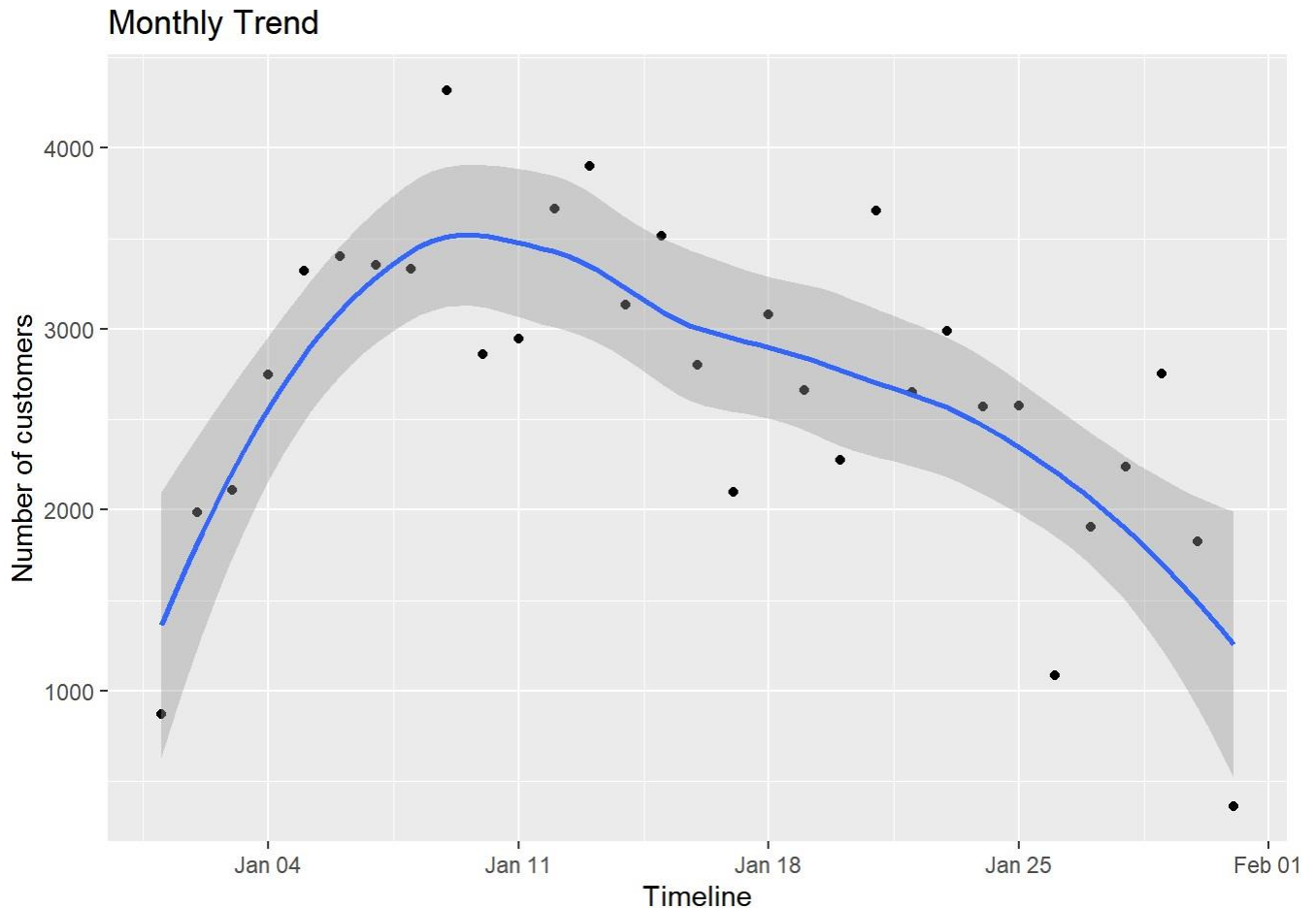


Visual representation of preferred bike on each day of week



A Scatter plot showcasing the trend of performance of the company during the Month

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Conclusion

In this analysis, we explored three factors in the dataset provided by the Cyclist Bike Share company and these three factors significantly impacting company's performance:

- Members and Casual Riders behavior analysis □ Most Popular bike among all
- company's Performance trend during the month

From our exploratory analysis of Members and Casuals Riders' pattern of using biked, we saw that Casual users ride time is far higher than the members in fact it is almost double during the weekends and the reason could be that they use bikes for leisure while the members use for up down from workplace to their homes and vice-versa. And if we look at the number of rides between both we see that members represent far higher numbers. It clearly indicates that members use our service more than once or twice each day while many users do not take a single ride each day.

Now if we look at the Bike type analysis, we see that Classic bikes are much more popular than the rest while Docked bikes are least popular. In fact, a vast majority of members do not use docked bikes.

Finally, the Performance also showcasing a **Downward Trend**. Company touches peak around first 10 to 15 days and then we can witness a continuous fall in company's performance.

What Next?

After witnessing analysis and conclusion. We have to decide what next? How can we improve company's performance?

So with the Members and Casual Riders we see that we start losing customers after the first half of the month and most likely they are casual riders. So the target is to turn casual riders into members.

What we can do is we can introduce the furthermore numbers of their preferred bikes and those are Classics and electrics. Although the casual riders use docked bikes but we can replace them by Classic and electric bikes as they are much more popular than the docked ones.

We can also introduce some great deals, discount and offers to the first particular number of riders as we saw that riders stop using our services after mid phase of the month. With exciting packages for members we can turn casual riders into members and can engage them for latter phase of the month as well.