Bike Share data

Shahid

2024-11-14

Introduction

Cyclistic, a leading bike-sharing company, is focused on increasing its profitability by analyzing the behavior of two main user groups: Casual Riders and Annual Members. Casual Riders typically use bikes for short-term trips, while Annual Members are more frequent riders who have purchased a year-long membership.

By analyzing the usage data from 12 months of user activity, this project aims to uncover key differences in the riding patterns between these two groups. The goal is to use this information to develop targeted marketing strategies that can convert Casual Riders into Annual Members, thereby maximizing revenue and fostering long-term customer loyalty.

The analysis will focus on identifying differences in: - Ride frequency - Ride duration - Ride distance - Bike type usage

The insights derived from this analysis will help inform Cyclistic's marketing strategies to encourage Casual Riders to adopt Annual Memberships.

Data Loading and Overview

In this section, we load and inspect the data that will be used for the analysis. We will examine the first few rows of the data to ensure everything is correctly loaded.

```
# Specify the path to your RDS files
# Manually try to load each RDS file
clean_data_combined_v2 <- readRDS("D:/Google Data analytics notes/R
Language/R programming/Cyclist Bike Share Data/clean_data_combined_v2.rds")

summarize_table_3 <- readRDS("D:/Google Data analytics notes/R Language/R
programming/Cyclist Bike Share Data/summarize_table_3.rds")

usertype_meantime <- readRDS("D:/Google Data analytics notes/R Language/R
programming/Cyclist Bike Share Data/usertype_meantime.rds")

summary_table_2 <- readRDS("D:/Google Data analytics notes/R Language/R
programming/Cyclist Bike Share Data/summary_table_2.rds")

summary_table <- readRDS("D:/Google Data analytics notes/R Language/R
programming/Cyclist Bike Share Data/summary_table.rds")
```

```
## Check the data
head(clean_data_combined_v2)
```

```
##                 Id          Type          Start.Time            End.Time
## 1 F96D5A74A3E41399 electric_bike 2023-01-21 20:05:00 2023-01-21 20:16:00
## 2 13CB7EB698CEDB88  classic_bike 2023-01-10 15:37:00 2023-01-10 15:46:00
## 3 BD88A2E670661CE5 electric_bike 2023-01-02 07:51:00 2023-01-02 08:05:00
## 4 C90792D034FED968  classic_bike 2023-01-22 10:52:00 2023-01-22 11:01:00
## 5 3397017529188E8A  classic_bike 2023-01-12 13:58:00 2023-01-12 14:13:00
## 6 58E68156DAE3E311 electric_bike 2023-01-31 07:18:00 2023-01-31 07:21:00
##                  Start.Station     Start.id
End.Station
## 1   Lincoln Ave & Fullerton Ave TA1309000058      Hampden Ct & Diversey
Ave
## 2         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th
St
## 3        Western Ave & Lunt Ave       RP-005 Valli Produce - Evanston
Plaza
## 4         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th
St
## 5         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th
St
## 6 Lakeview Ave & Fullerton Pkwy TA1309000019      Hampden Ct & Diversey
Ave
##         End.id Start.Lattitude Start.Longitude End.Lattitude End.Longitude
## 1       202480        41.92407       -87.64628      41.93000     -87.64000
## 2 TA1308000002        41.79957       -87.59475      41.80983     -87.59938
## 3          599        42.00857       -87.69048      42.03974     -87.69941
## 4 TA1308000002        41.79957       -87.59475      41.80983     -87.59938
## 5 TA1308000002        41.79957       -87.59475      41.80983     -87.59938
## 6       202480        41.92607       -87.63886      41.93000     -87.64000
##   Membership.Type Start.hour End.hour travel_time start_day start_month
## 1          member         20       20          11  Saturday          01
## 2          member         15       15           9   Tuesday          01
## 3          casual         07       08          14    Monday          01
## 4          member         10       11           9    Sunday          01
## 5          member         13       14          15  Thursday          01
## 6          member         07       07           3   Tuesday          01
##   start_year travel_distance travel_time_hours     speed
## 1       2023        14810.14         0.1833333  80782.58
## 2       2023        14798.47         0.1500000  98656.50
## 3       2023        14815.98         0.2333333  63497.07
## 4       2023        14798.47         0.1500000  98656.50
## 5       2023        14798.47         0.2500000  59193.90
## 6       2023        14810.33         0.0500000 296206.67
```

```
head(summarize_table_3)
```

```
## # A tibble: 6 × 4
## # Groups:   Membership.Type [1]
```

```
##   Membership.Type start_day number_of_ride average_duration
##   <chr>           <chr>              <int>            <dbl>
## 1 casual          Friday            305525             20.4
## 2 casual          Monday            229778             20.6
## 3 casual          Saturday          394589             23.7
## 4 casual          Sunday            325966             24.4
## 5 casual          Thursday          266145             18.4
## 6 casual          Tuesday           247379             18.7
```

```
head(usertype_meantime)
```

```
## # A tibble: 2 × 4
##   Membership.Type mean_time mean_distance mean_speed
##   <chr>               <dbl>         <dbl>      <dbl>
## 1 casual               20.9        14807.    111207.
## 2 member               12.3        14807.    144047.
```

```
head(summary_table_2)
```

```
## # A tibble: 5 × 4
## # Groups:   Membership.Type [2]
##   Membership.Type Type          number_of_ride mean_travel_time
##   <chr>           <chr>                  <int>            <dbl>
## 1 casual          classic_bike          848893             25.9
## 2 casual          docked_bike            83267             52.9
## 3 casual          electric_bike        1084459             14.6
## 4 member          classic_bike         1754386             13.1
## 5 member          electric_bike        1782038             11.4
```

```
head(summary_table)
```

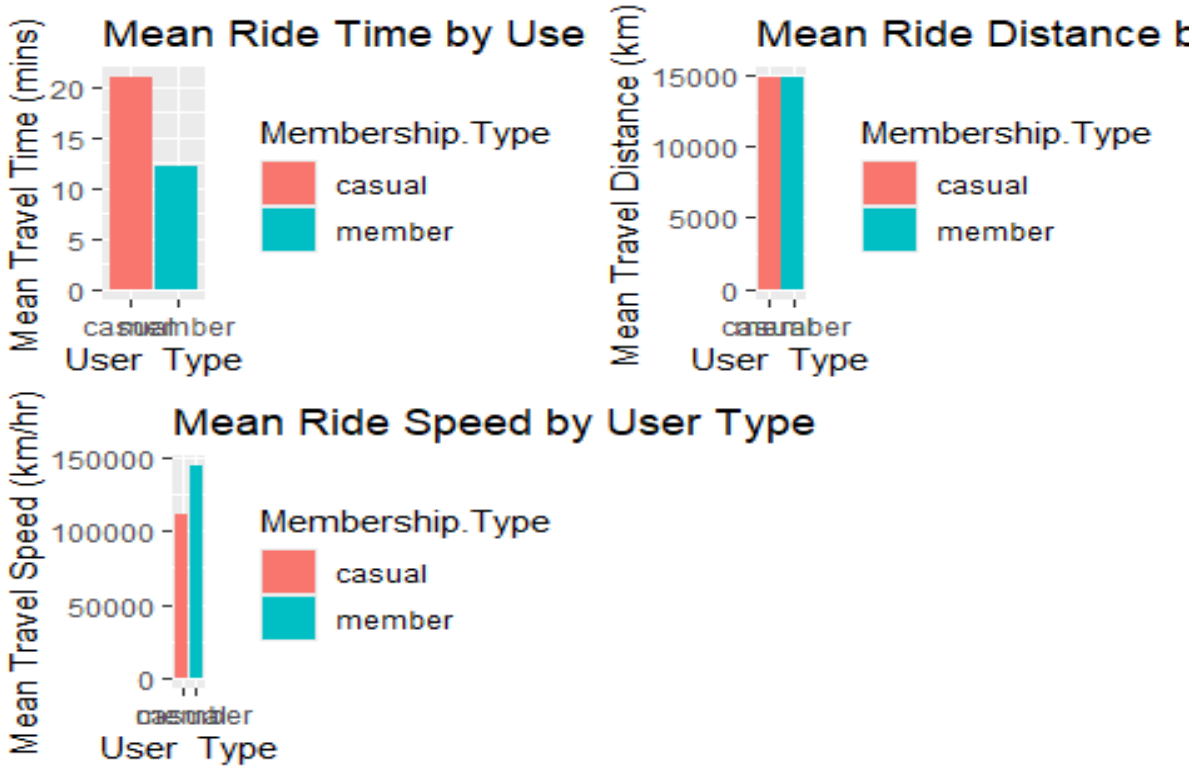
```
## # A tibble: 6 × 4
## # Groups:   Membership.Type [1]
##   Membership.Type start_day number_of_ride average_duration
##   <chr>           <chr>              <int>            <dbl>
## 1 casual          Friday            305525             20.4
## 2 casual          Monday            229778             20.6
## 3 casual          Saturday          394589             23.7
## 4 casual          Sunday            325966             24.4
## 5 casual          Thursday          266145             18.4
## 6 casual          Tuesday           247379             18.7
```

Calculate mean_time, mean_distance, and mean_speed by Membership.Type

Mean Ride Time by User Type

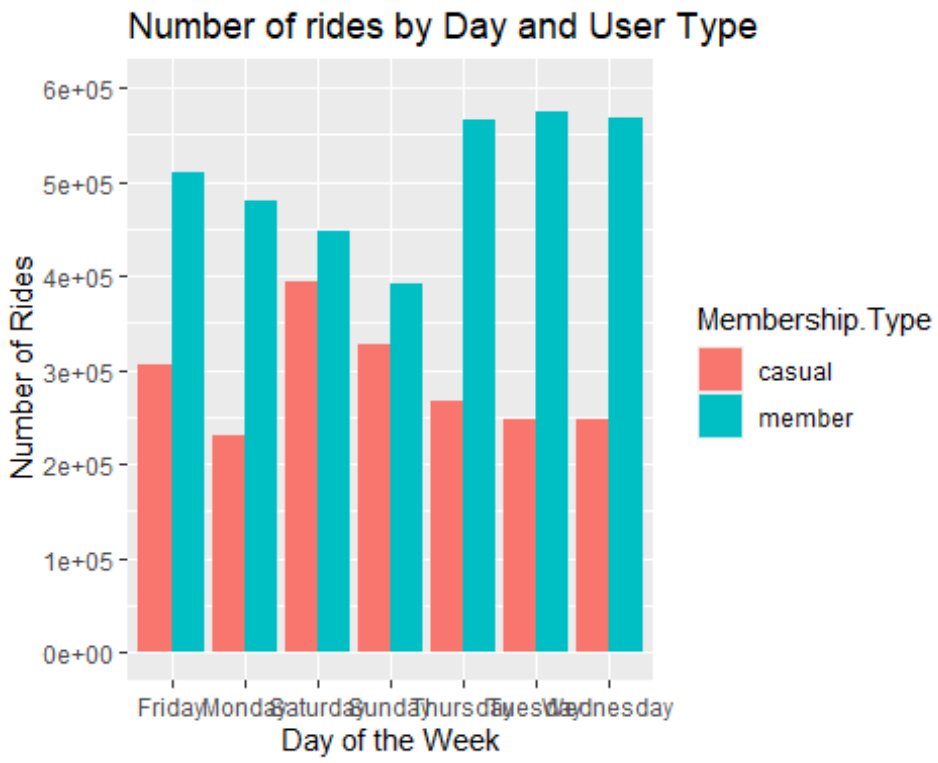
The following plot shows the mean ride time for both Casual Riders and Annual Members. This comparison helps us understand if one group tends to take longer rides than the other

```r
  # Mean Ride Time
  a <- ggplot(data = usertype_meantime) +
  geom_col(mapping = aes(x =Membership.Type, y = mean_time, fill =
Membership.Type), position = "dodge") +
  labs(title = "Mean Ride Time by User Type", x = "User  Type", y = "Mean
Travel Time (mins)")

  # Mean Ride Distance
b <- ggplot(data = usertype_meantime) +
  geom_col(mapping = aes(x = Membership.Type, y = mean_distance, fill =
Membership.Type), position = "dodge") +
  labs(title = "Mean Ride Distance by User Type", x = "User  Type", y = "Mean
Travel Distance (km)")

  # Mean Ride Speed

c <- ggplot(data = usertype_meantime) +
  geom_col(mapping = aes(x = Membership.Type, y = mean_speed, fill =
Membership.Type), position = "dodge") +
  labs(title = "Mean Ride Speed by User Type", x = "User  Type", y = "Mean
Travel Speed (km/hr)")
# This arrange the three plots in grid
grid.arrange(a,b,c,nrow=2,ncol=2)
```



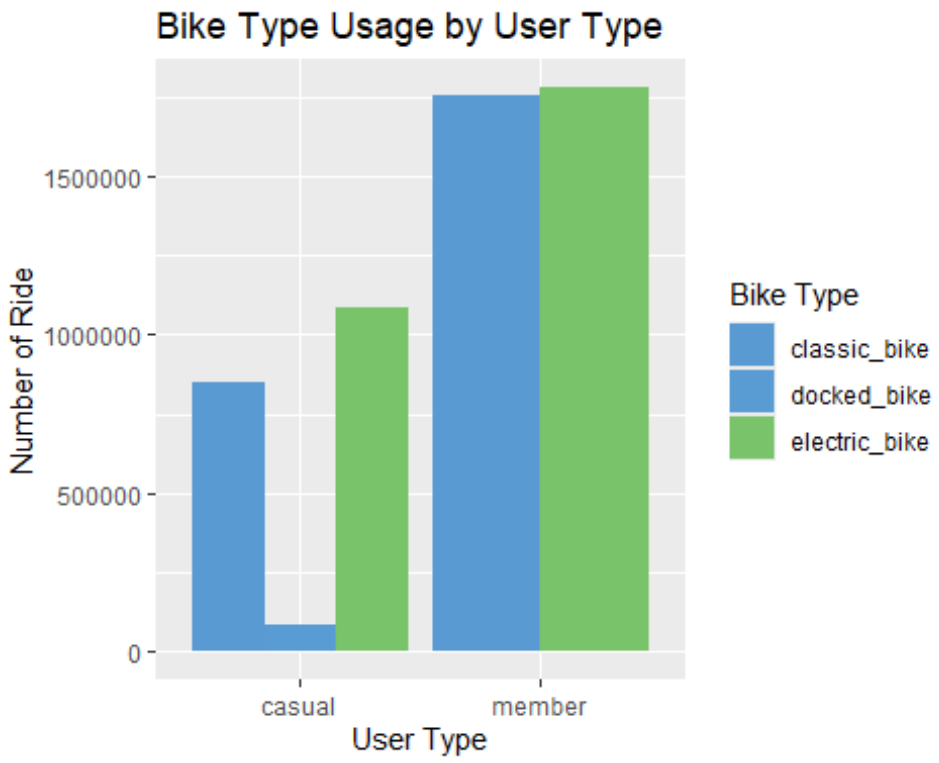
Visualization for Number of Rides Vs User Type

```
ggplot(data = summary_table)+
  geom_col(mapping = aes(x=start_day,y=number_of_ride,fill =
Membership.Type),position = "dodge")+
  labs(title = "Number of rides by Day and User Type",x="Day of the Week",y="
Number of Rides")+
  scale_y_continuous(limits = c(0,600000),breaks = seq(0,600000,by= 100000))
```



Bike Type Usage by User Type

```
  ggplot(data=summary_table_2)+
  geom_col(mapping = aes(x=Membership.Type,y=number_of_ride,fill =
Type),position = "dodge")+
  scale_fill_manual(values =
c(electric_bike="#79c36a","classic_bike"="#599ad3","docked_bike"="#599cd3"))+
  labs(title = "Bike Type Usage by User Type",
       x="User Type",
       y="Number of Ride",
      fill="Bike Type" )
```



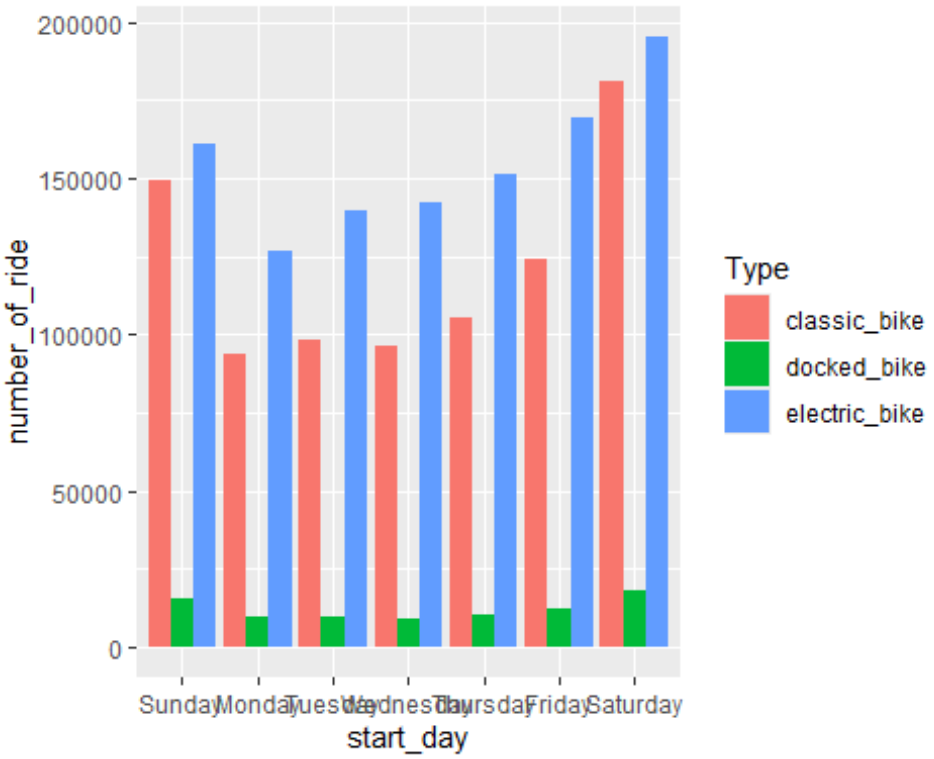
Creating a dataframe containing rideable type, user type, day, and number of rides

```
# Filter the data to include only relevant bike types and group by user type
and day
summarize_table_3 <- clean_data_combined_v2 %>%
  filter(Type == "classic_bike" | Type == "electric_bike" |
Type=="docked_bike") %>% group_by(Membership.Type, Type, start_day) %>%
  summarize(number_of_ride = n(),
    mean_travel_time = mean(travel_time, na.rm = TRUE)
  ) %>%
   # Ensure the start_day column is ordered correctly (days of the week)
  mutate(stsrt_day = factor(start_day, levels = c("Sunday", "Monday",
"Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))) %>%
  arrange(start_day)

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

print(summarize_table_3)

## # A tibble: 35 × 6
## # Groups:   Membership.Type, Type [5]
##    Membership.Type Type      start_day number_of_ride mean_travel_time
```

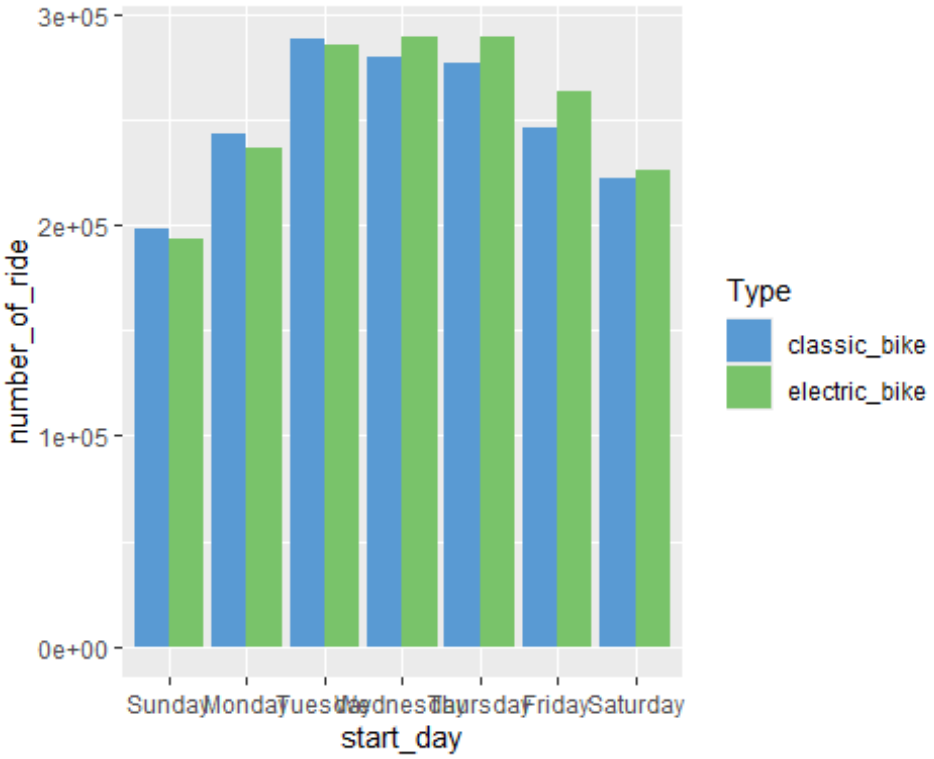
```
stsrt_day
##    <chr>           <chr>     <ord>              <int>            <dbl>
<ord>
##  1 casual          classic_… Sunday            149545             29.4
Sunday
##  2 casual          docked_b… Sunday             15609             54.1
Sunday
##  3 casual          electric… Sunday            160812             16.8
Sunday
##  4 member          classic_… Sunday            198383             14.7
Sunday
##  5 member          electric… Sunday            193083             12.6
Sunday
##  6 casual          classic_… Monday             93888             25.8
Monday
##  7 casual          docked_b… Monday              9490             53.5
Monday
##  8 casual          electric… Monday            126400             14.2
Monday
##  9 member          classic_… Monday            242927             12.5
Monday
## 10 member          electric… Monday            236114             10.7
Monday
## # ℹ 25 more rows
```

Separting the Casual and Annual Member data into two separte data frame for the visualization

```
  casual_data <- summarize_table_3 %>% filter(Membership.Type=="casual")
  annual_member <- summarize_table_3 %>% filter(Membership.Type=="member")

  ggplot(data=casual_data)+
    geom_col(mapping =
aes(x=start_day,y=number_of_ride,fill=Type),position="dodge")
```




```
   ggplot(data=annual_member)+
    geom_col(mapping =
aes(x=start_day,y=number_of_ride,fill=Type),position="dodge")+
     scale_fill_manual(values =
c(electric_bike="#79c36a","classic_bike"="#599ad3","docked_bike"="#599cd3" ))
```



Observations:

There seems to be an unexpected preference towards classic bike compared to electric bike for both casual users and annual members during this time period.

The preference seems to be higher in case of annual members.

The number of rides specific to the day of the week do not provude any significant pattern.

Conclusions:

Casual riders tend to use the bikes more on the weekends, however annual members have been seen to use the bikes equally throughout the week, suggesting that the casual riders probably use the bikes for leisurely weekend rides rather than as a daily commute to work.

The average speed of the casual riders is lower compared to the annual members, further solidifying the idea of a more leisurely approach of biking for the casual riders.

An unexpected preference for the classic bike has been seen for both the casual riders and annual members

Recomendation:

A weekend specific discount for the annual members, to influence the casual riders to change to an annual membership, as they ride mostly on weekends.

Advertisements focusing on the benefits of using the service as a means of daily commute to work.

Conducting a public survey to understand the issues related to the electric bikes.