

Report on How Does a Bike-Share Navigate Speedy Success

Rokhsona Pervin

Introduction:

This is a case study of a Bike Share Company by the marketing analyst team at Cyclistic. The goal of this project is to increase the number of annual memberships. To meet our goals, we selected datasets for the year 2020-2021. We analyzed our data to find any difference between causal rides and annual members. we also demonstrate how causal riders and member riders used different types of bikes during the weekday and weekends. Finally, we are trying to figure it out what our monthly trends tell to predict for future decisions.

Data Preparation:

We choose data for the year November, 2020 to October, 2021. After installing all the required packages and library, we uploaded our data.

Then, we combined our 12 months dataset for analysis.

```
all<-rbind(N2020,D2020,J2021,F2021,M2021,A2021,MAY2021,JUN2021,
           JUL2021,AUG2021,SEP2021,OCT2021)
names(all)
```

```
## [1] "ride_id"           "rideable_type"      "started_at"
## [4] "ended_at"          "start_station_name" "start_station_id"
## [7] "end_station_name"  "end_station_id"     "start_lat"
## [10] "start_lng"         "end_lat"            "end_lng"
## [13] "member_casual"     "ride_length"        "ride_length_formatted"
## [16] "days_of_week"     "group_day"          "Month"
## [19] "Year"
```

We Notice that there are some are missing values and negative values in our data. So, we deleted those missing values and removed them.

```
A<-complete.cases(all) # To check missing values
all.N<-all[A,]
newall<-all.N[,c(2,13,14,16:19)]
names(newall)
```

```
## [1] "rideable_type" "member_casual" "ride_length"   "days_of_week"
## [5] "group_day"     "Month"         "Year"
head(newall)
```

```
##   rideable_type member_casual ride_length days_of_week group_day Month Year
## 1 electric_bike      casual 0.006712963         1         1    11 2020
## 2 electric_bike      casual 0.007858796         1         1    11 2020
## 3 electric_bike      casual 0.020150463         1         1    11 2020
## 4 electric_bike      casual 0.006423611         1         1    11 2020
## 5 electric_bike      casual 0.023229167         1         1    11 2020
```

```
## 6 electric_bike          casual 0.034270833          7          1    11 2020
all.NN<-newall[newall$ride_length>0,] # for deleting missing value

head(all.NN)

##   rideable_type member_casual ride_length days_of_week group_day Month Year
## 1 electric_bike          casual 0.006712963          1          1    11 2020
## 2 electric_bike          casual 0.007858796          1          1    11 2020
## 3 electric_bike          casual 0.020150463          1          1    11 2020
## 4 electric_bike          casual 0.006423611          1          1    11 2020
## 5 electric_bike          casual 0.023229167          1          1    11 2020
## 6 electric_bike          casual 0.034270833          7          1    11 2020

## creating categorical variables using factor function
bb<-within(all.NN, {
  days_of_week<-factor(days_of_week, labels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
  group_day<-factor(group_day, labels= c("Weekend", "Week_Day"))
  Month<-factor(Month, labels=c("Nov", "Dec", "Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct"))
  Year<-factor(Year)}

str(bb)

## 'data.frame':   5335502 obs. of  7 variables:
##  $ rideable_type: chr  "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...
##  $ member_casual: chr  "casual" "casual" "casual" "casual" ...
##  $ ride_length  : num  0.00671 0.00786 0.02015 0.00642 0.02323 ...
##  $ days_of_week : Factor w/ 7 levels "Sunday","Monday",...: 1 1 1 1 1 7 7 7 7 7 ...
##  $ group_day    : Factor w/ 2 levels "Weekend","Week_Day": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Month        : Factor w/ 12 levels "Nov","Dec","Jan",...: 11 11 11 11 11 11 11 11 11 11 ...
##  $ Year         : Factor w/ 2 levels "2020","2021": 1 1 1 1 1 1 1 1 1 1 ...

head(bb)

##   rideable_type member_casual ride_length days_of_week group_day Month Year
## 1 electric_bike          casual 0.006712963      Sunday Week_Day   Sep 2020
## 2 electric_bike          casual 0.007858796      Sunday Week_Day   Sep 2020
## 3 electric_bike          casual 0.020150463      Sunday Week_Day   Sep 2020
## 4 electric_bike          casual 0.006423611      Sunday Week_Day   Sep 2020
## 5 electric_bike          casual 0.023229167      Sunday Week_Day   Sep 2020
## 6 electric_bike          casual 0.034270833    Saturday Week_Day   Sep 2020
```

Analysis:

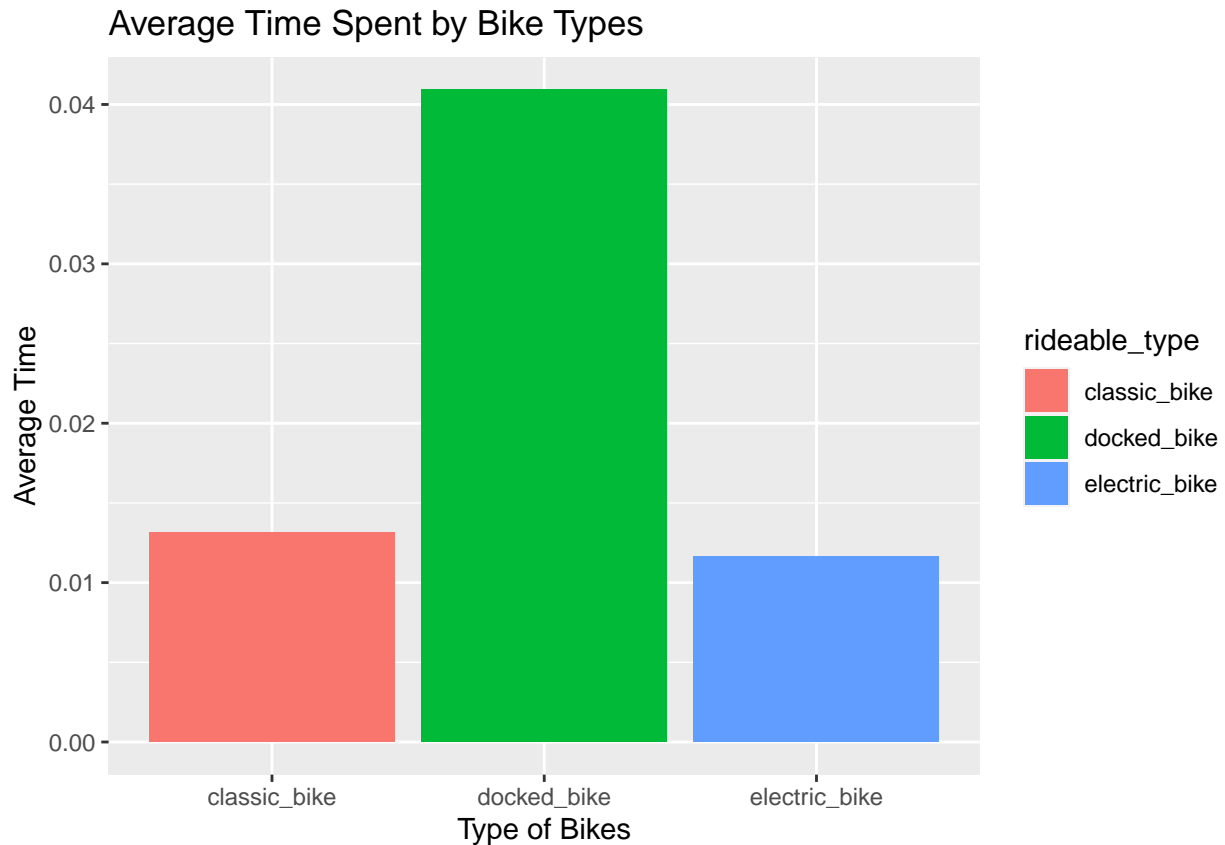
We are interested in the average time spent for each bike type. Therefore, we created a plot of average time spent for all three types Bikes.

```
bb %>%

  group_by(rideable_type) %>%

  summarise(mean_ride_length = mean(ride_length)) %>%

  ggplot(aes(x = rideable_type, y = mean_ride_length, fill=rideable_type)) +
  geom_bar(stat='identity')+labs(title="Average Time Spent by Bike Types",
                                x ="Type of Bikes", y = "Average Time")
```



We see on an average Docked_bike is used more time length compared to classic_bike and electric bike. Classic_bike and Electric bike user seems a little difference.

Now to see if there are any differences between Causal and Member groups among types of bike uses we created a plot based on those two groups

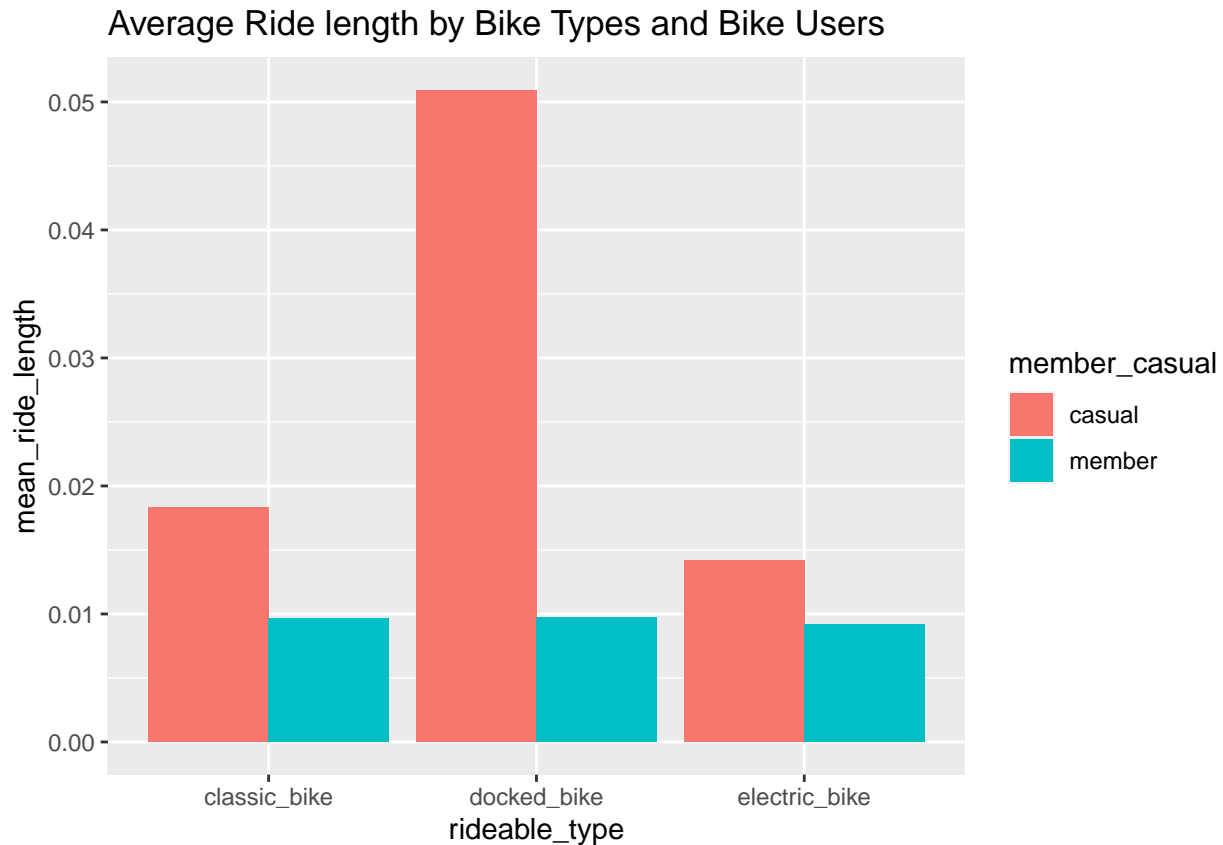
```
bb %>%

  group_by(member_casual,rideable_type) %>%

  summarise(mean_ride_length = mean(ride_length)) %>%

  ggplot(aes(x = rideable_type, y = mean_ride_length, fill=member_casual)) +
  geom_bar(stat = "identity",position="dodge") +theme_gray()+
  labs(title="Average Ride length by Bike Types and Bike Users")

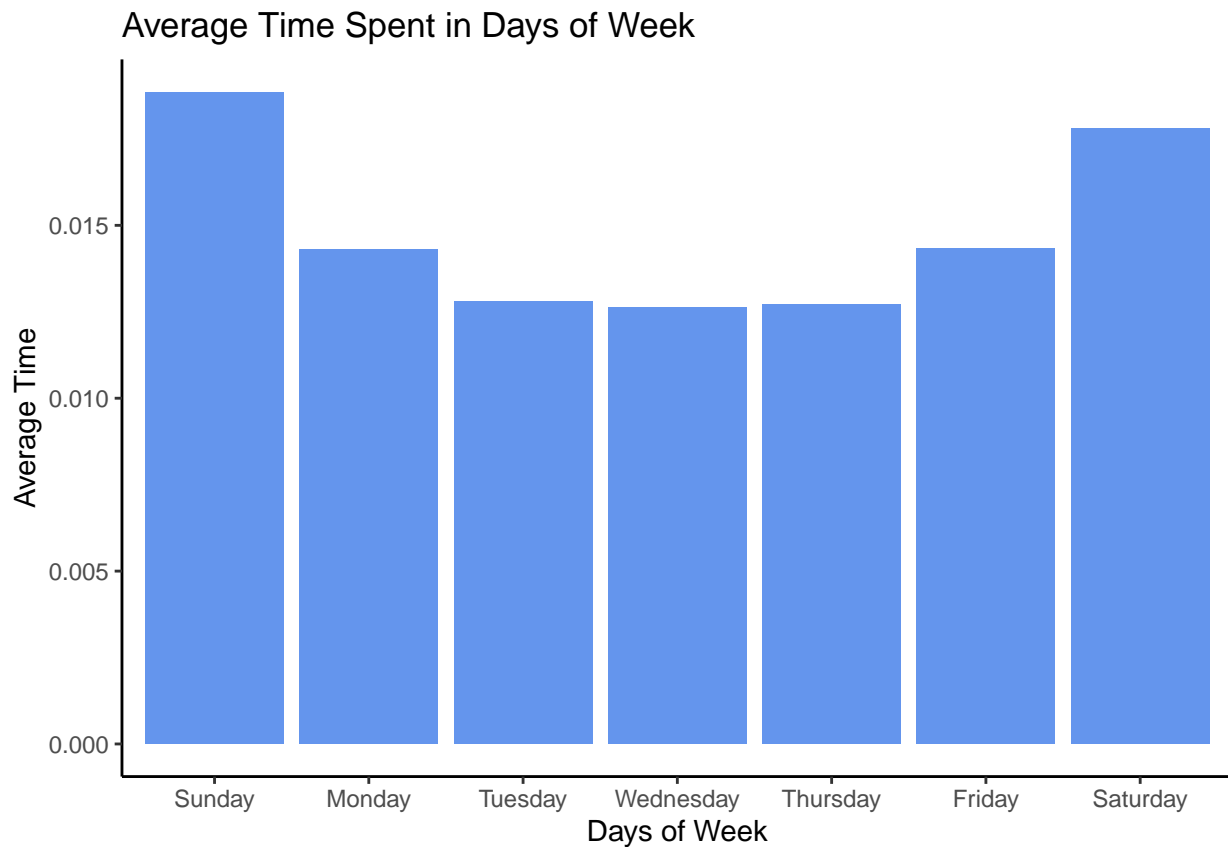
## `summarise()` has grouped output by 'member_casual'. You can override using the `.groups` argument.
```



We see that on an average Casual riders use Docked bikes more frequently than Classic and Electric Bike types. On the other hand, Members use all three bikes almost the same.

Next, we would like to see if there is any difference among different days of use. So, we created a plot of average_time by Days of week.

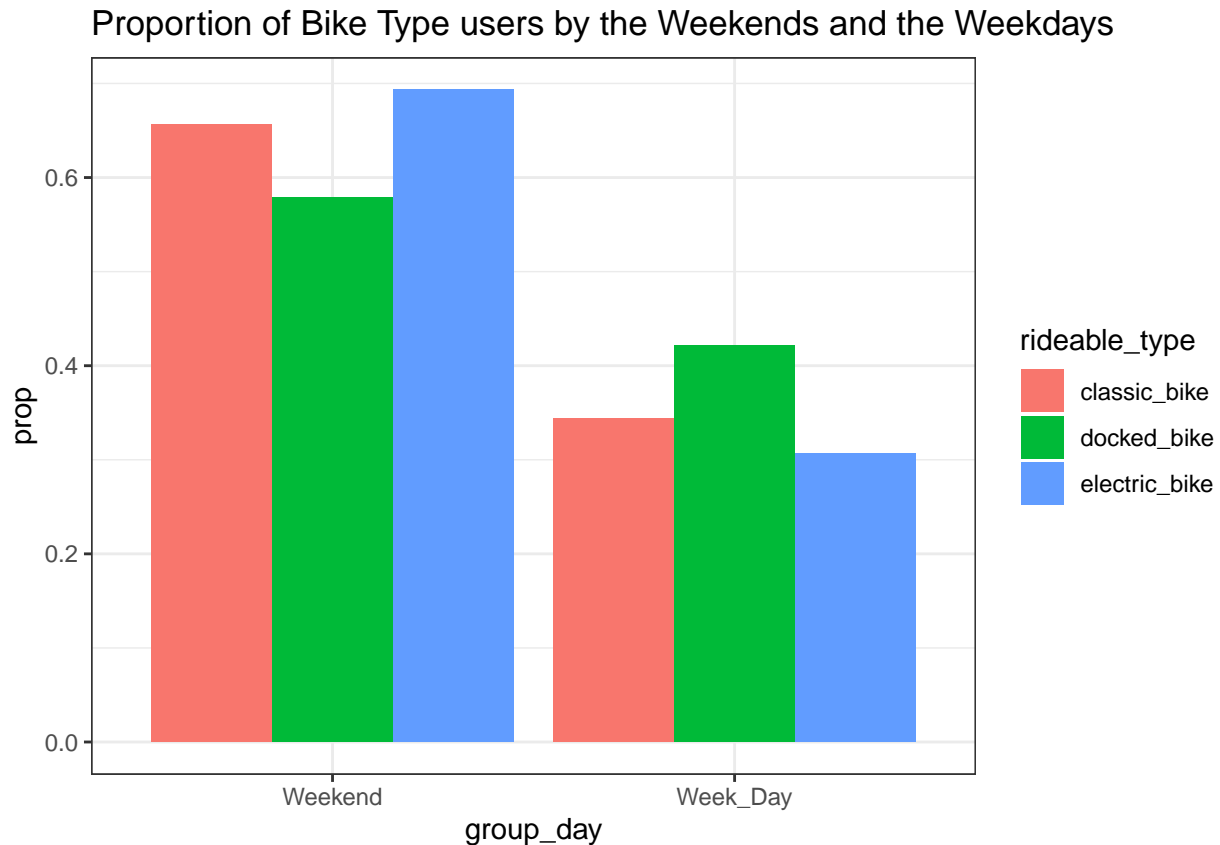
```
bb %>%
  group_by(days_of_week) %>%
  summarise(mean_ride_length = mean(ride_length)) %>%
  ggplot(aes(x = days_of_week, y = mean_ride_length)) +
  geom_bar(stat='identity', fill="cornflowerblue")+labs(title="Average Time Spent in Days of Week",
    x = "Days of Week", y = "Average Time")+theme_classic()
```



It revealed that on an average time more bikes were used on the weekends compared to the weekdays.

We are also interested in proportion of bike used by weekend and weekdays among three types of Bikes.

```
ggplot(bb, aes(x=group_day, fill=rideable_type)) +
  geom_bar(aes(y = ..prop.., group =rideable_type ), position="dodge")+theme_bw()+
  labs(title="Proportion of Bike Type users by the Weekends and the Weekdays")
```



We see Electric bikes are used more the weekends than the weekdays.

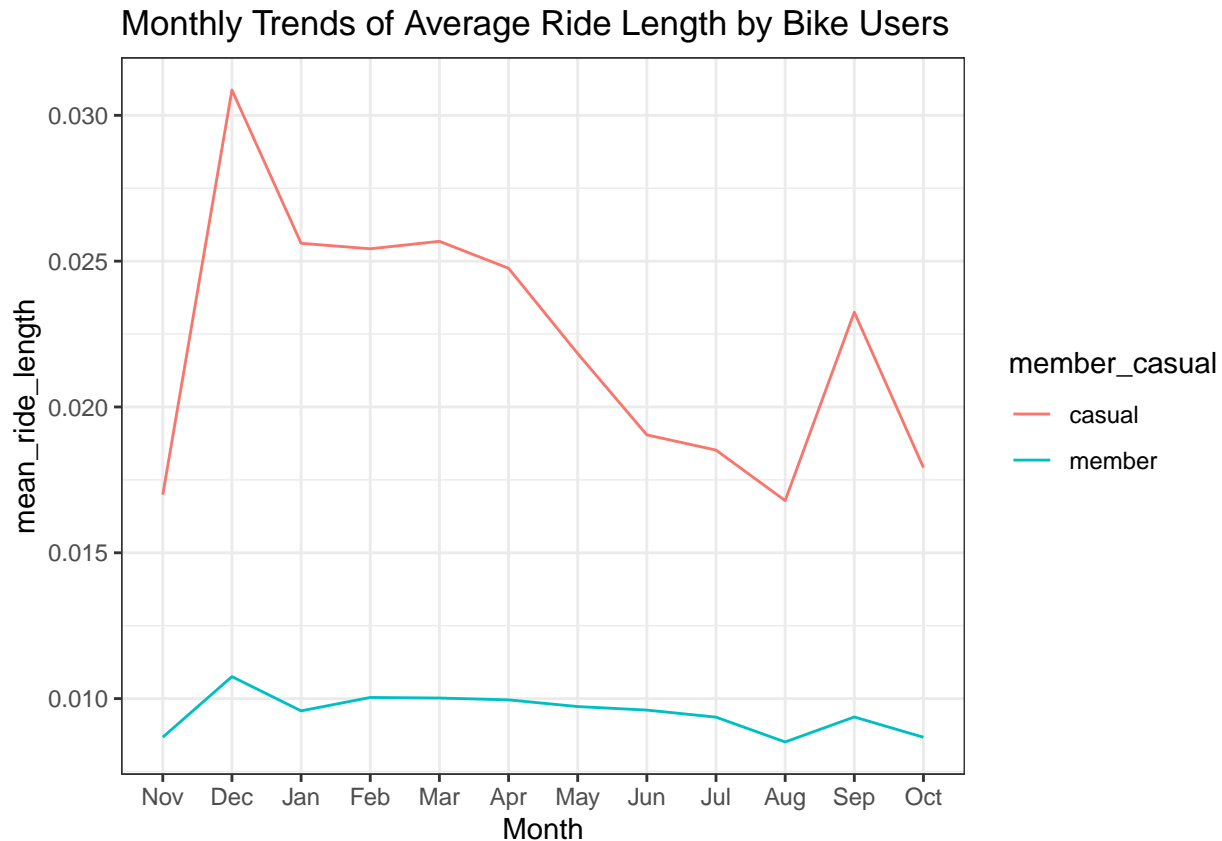
Finally, to see average ride length Monthly trend between Causal and Members users, we created a graph of mean ride length based on Months for the years 2020-2021.

```
bb %>%
```

```
group_by(member_casual, Month) %>%

summarise(mean_ride_length = mean(ride_length)) %>%
ggplot(aes(x=Month, y=mean_ride_length, group=member_casual)) +
geom_line(aes(color=member_casual))+
scale_x_discrete(limits=levels(bb$Month))+ theme_bw()+
labs(title="Monthly Trends of Average Ride Length by Bike Users")
```

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



It revealed that, on average Casual rides seem higher than Members rides. Moreover, On Decembers and in Septembers look pick time.

Conclusions:

1. On an average Docked bike is used more time length compared to classic bike and electric bike. Classic bike and Electric bike user seem a little difference.
2. We see on an average during weekends bike is used more than weekdays. Moreover, casual riders use docked bike more compared to classic bike in the Weekends. Member riders on the other hand seems no big difference among the three bike types.
3. We see that over the twelve-month period casual riders uses way too more bikes compared to member riders. It also demonstrates that months December and September seems high peak.

Recommendations:

1. Since we see the classic bike is used more compared to other bike types, so, increasing number of classic bikes will be a better idea.
2. We also found out that the use of the bikes by both casual and Member during Decembers and Septembers is more on an average than the other months of the year. Therefore, it would be profitable if there is any discount can offer for customers the other months.
3. Finally, the analysis reveals that more bikes used in the weekend are found than in weekdays. Increasing the number of bikes on weekend can leads the company to maximize profits.