**R CODE FOR ANALYSIS**

# First I need all the libraries for my analysis.

library(“tidyverse”)

library(“ggplot2”)

library(“lubridate”)

library(“geosphere”

library(“gridExtra”)

library(“ggmap”)

# Then I have to load my data for the analysis.

aug\_2023 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/aug\_2023.csv”)

jul\_2023 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/jul\_2023.csv”)

jun\_2023 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/jun\_2023.csv”)

may\_2023 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/may\_2023.csv”)

apr\_2023 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/apr\_2023.csv”)

mar\_2023 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/mar\_2023.csv”)

feb\_2023 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/feb\_2023.csv”)

jan\_2023 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/jan\_2023.csv”)

dec\_2022 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/dec\_2022.csv”)

nov\_2022 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/nov\_2022.csv”)

oct\_2022 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/oct\_2022.csv”)

sep\_2022 <- read.csv(“D:/ADIMULAM/Project/Bike-Share Navigate Speedy Success/Data/sep\_2022.csv”)

#Check the Colnames

colnames(aug\_2023), colnames(jul\_2023), colnames(jun\_2023), colnames(may\_2023), colnames(apr\_2023), colnames(mar\_2023), colnames(feb\_2023), colnames(jan\_2023), colnames(dec\_2023), colnames(nov\_2023), colnames(oct\_2023), colnames(sep\_2023).

# Convert ride\_id and rideable\_type to character

aug\_2023 <- mutate(aug\_2023, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

jul\_2023 <- mutate(jul\_2023, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

jun\_2023 <- mutate(jun\_2023, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

may\_2023 <- mutate(may\_2023, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

apr\_2023 <- mutate(apr\_2023, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

mar\_2023 <- mutate(mar\_2023, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

feb\_2023 <- mutate(feb\_2023, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

jan\_2023 <- mutate(jan\_2023, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

dec\_2022 <- mutate(dec\_2022, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

nov\_2022 <- mutate(nov\_2022, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

oct\_2022 <- mutate(oct\_2022, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

sep\_2022 <- mutate(sep\_2022, ride\_id = as.character(ride\_id),

rideable\_type = as.character(rideable\_type))

#Combine into one big data frame

all\_months <- bind\_rows(aug\_2023, jul\_2023, jun\_2023, may\_2023, apr\_2023, mar\_2023, feb\_2023, jan\_2023, dec\_2022, nov\_2022, oct\_2022, sep\_2022)

# Phase 3 : Process

#Inspect the new table that has been created

all\_months\_clean <- drop\_na(all\_months)

colnames(all\_months\_clean)

nrow(all\_months\_clean)

dim(all\_months\_clean)

head(all\_months\_clean)

str(all\_months\_clean)

summary(all\_months\_clean)

#Reassign to the desired values

all\_months\_clean <- all\_months\_clean %>% mutate(member\_casual = recode(member\_casual, "Subscriber" = "member", "Customer" = "casual"))

table(all\_months\_clean$member\_casual)

#Add columns that list the date, month, day and year of each ride.

all\_months\_clean$date <- as.Date(all\_months\_clean$started\_at)

all\_months\_clean$month <- format(as.Date(all\_months\_clean$date), "%m")

all\_months\_clean$day <- format(as.Date(all\_months\_clean$date), "%d")

all\_months\_clean$year <- format(as.Date(all\_months\_clean$date), "%Y")

all\_months\_clean$day\_of\_week <- format(as.Date(all\_months\_clean$date), "%A")

#Add a ride\_length calculation to all\_months\_clean

all\_months\_clean$ride\_length <- difftime(all\_months\_clean$ended\_at, all\_months\_clean$started\_at)

#ride\_distance travelled

> all\_months\_clean$ride\_distance <- distGeo(matrix(c(all\_months\_clean$start\_lng, all\_months\_clean$start\_lat), ncol = 2), matrix(c(all\_months\_clean$end\_lng, all\_months\_clean$end\_lat), ncol = 2))

> all\_months\_clean$ride\_distance <- all\_months\_clean$ride\_distance/1000

#Convert ride\_length from factor to numeric

> all\_months\_clean$ride\_speed = c(all\_months\_clean$ride\_distance)/as.numeric(c(all\_months\_clean$ride\_length), units = "hours")

> all\_months\_clean <- all\_months\_clean[!(all\_months\_clean$start\_station\_name == "HQ QR" | all\_months\_clean$ride\_length<0),]

# Phase 4 : Analyze

#Calculate the average distance for both casual and member type users.

> userType\_means <- all\_months\_clean %>%

group\_by(member\_casual) %>%

summarise(mean\_time = mean(ride\_length), mean\_distance = mean(ride\_distance))

> membervstime <- ggplot(userType\_means) +

+ geom\_col(mapping = aes(x = member\_casual, y = mean\_time, fill = member\_casual), show.legend = FALSE) +

+ labs(title = "Mean travel time by User type", x = "User Type", y = "Mean time in sec")

>

> memvsdis <- ggplot(userType\_means) +

+ geom\_col(mapping = aes(x = member\_casual, y = mean\_distance, fill = member\_casual), show.legend = FALSE) +

+ labs(title = "Mean travel distance by User Type", x = "User Type", y = "Mean distance in Km", caption = "Data by Motivate International Inc")

> grid.arrange(membervstime, memvsdis, ncol = 2)

#Number of rides

> all\_months\_clean %>%

mutate(weekday = wday(started\_at, label = TRUE)) %>%

group\_by(member\_casual, weekday) %>%

summarise(number\_of\_rides = n(), average\_duration = mean(ride\_length), .groups = 'drop') %>%

arrange(member\_casual, weekday) %>%

ggplot(aes(x = weekday, y = number\_of\_rides, fill = member\_casual)) +

geom\_col(position = "dodge") +

labs(title = "Number of rides by User type during the week", x = "Days of the week", y = "Number of rides", caption = "Data by Motivate International Inc", fill = "User type") +

theme(legend.position = "top")

Analysis :

* It seems that the members are highly users almost the week, but on wednesday and thursday are highly users.
* Casual users are high only on saturday. It seems they are normal casual riders on weekends.

#Creating a new data frame with bike-type

> with\_bike\_type <- all\_months\_clean %>% filter(rideable\_type == "classic\_bike" | rideable\_type == "electric\_bike")

> with\_bike\_type %>%

group\_by(member\_casual, rideable\_type) %>%

summarise(totals = n(), .groups = "drop") %>%

ggplot() +

geom\_col(aes(x=member\_casual, y = totals, fill = rideable\_type), position = "dodge") +

labs(title = "Bike type usage by user type", x = "User type", y = NULL, fill = "Bike type") +

scale\_fill\_manual(values = c("classic\_bike" = "#746F72", "electric\_bike" = "#FFB100")) +

theme\_minimal()+

theme(legend.position = "top")

#Both user types during a week

> with\_bike\_type %>%

mutate(weekday = wday(started\_at, label = TRUE)) %>%

group\_by(member\_casual, rideable\_type, weekday) %>%

summarise(totals = n(), .groups = "drop") %>%

ggplot(aes(x=weekday, y = totals, fill=rideable\_type)) +

geom\_col(, position = "dodge") +

facet\_wrap(~member\_casual) +

labs(title = "Bike type usage by user type during a week", x = "User type", y = NULL, caption = "Data by Motivate International Inc") +

scale\_fill\_manual(values = c("classic\_bike" = "#746F72", "electric\_bike" = "#FFB100")) +

theme\_minimal() +

theme(legend.position = "none")

Analysis:

* Here we can see that the annual members use both electrical and normal types.
* But the casual members are using most of them as electrical.
* From the analysis the casual riders need more speed and time saving rides.

#Let's check the coordinates of rides.

> coordinates\_table <- all\_months\_clean %>%

+ filter(start\_lng != end\_lng & start\_lat = end\_lat) %>%

> coordinates\_table <- all\_months\_clean %>%

+ filter(start\_lng != end\_lng & start\_lat != end\_lat) %>%

+ group\_by(start\_lng, start\_lat, end\_lng, end\_lat, member\_casual, rideable\_type) %>%

+ summarise(total = n(), .groups = "drop") %>%

+ filter(total > 250)

> casual <- coordinates\_table %>% filter(member\_casual == "casual")

> member <- coordinates\_table %>% filter(member\_casual == "member")

> chi\_bb <- c(

+ left = -87.700424,

+ bottom = 41.790769,

+ right = -87.554855,

+ top = 41.990119

+ )

> chicago\_stamen <- get\_stamenmap(

+ bbox = chi\_bb,

+ zoom = 12,

+ maptype = "toner"

+ )

#Then plot the data

> ggmap(chicago\_stamen, darken = c(0.8, "white")) +

geom\_curve(casual, mapping = aes(x = start\_lng, y = start\_lat, xend = end\_lng, yend = end\_lat, alpha = total, color = rideable\_type), size = 0.5, curvature = .2, arrow = arrow(length = unit(0.2, "cm"), ends = "first", type = "closed")) +

coord\_cartesian() +

labs(title = "Most popular routes by casual users", x = NULL, y = NULL, color = "User type", caption = "Data by Motivate International Inc") +

theme(legend.position = "none")

Analysis:

* As we can see the casual riders are only using the vehicles at in town only.
* They are using this for normal purposes only.
* Mostly they are using this in their free time only.

> ggmap(chicago\_stamen, darken = c(0.8, "white")) +

geom\_curve(member, mapping = aes(x = start\_lng, y = start\_lat, xend = end\_lng, yend = end\_lat, alpha = total, color = rideable\_type), size = 0.5, curvature = .2, arrow = arrow(length = unit(0.2, "cm"), ends = "first", type = "closed")) +

coord\_cartesian()+

labs(title = "Most popular routes by annual members", x = NULL, y=NULL, caption = "Data by Motivate International Inc") +

theme(legend.position = "none")