**R code for bike share data**

Max Johnson

08/23/2023

library(tidyverse)

library(lubridate)

library(janitor)

library(dplyr)

library(ggplot2)

trip22\_07 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202207-divvy-tripdata.csv")

trip22\_08 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202208-divvy-tripdata.csv")

trip22\_09 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202209-divvy-publictripdata.csv")

trip22\_10 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202210-divvy-tripdata.csv")

trip22\_11 <- read.csv ("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202211-divvy-tripdata.csv")

trip22\_12 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202212-divvy-tripdata.csv")

trip23\_01 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202301-divvy-tripdata.csv")

trip23\_02 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202302-divvy-tripdata.csv")

trip23\_03 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202303-divvy-tripdata.csv")

trip23\_04 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202304-divvy-tripdata.csv")

trip23\_05 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202305-divvy-tripdata.csv")

trip23\_06 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202306-divvy-tripdata.csv")

trip23\_07 <- read.csv("C:/Users/maxj3/OneDrive - University of Texas at Arlington/summer/Newexcel folder/capstone/new data/202306-divvy-tripdata.csv")

#Combine into one data frame

trips\_23fill<- rbind(trip22\_07, trip22\_08, trip22\_09, trip22\_10, trip22\_11, trip22\_12, trip23\_01, trip23\_02, trip23\_03, trip23\_04, trip23\_05, trip23\_06,trip23\_07)

View(trips\_23fill)

trips\_23fill <-trips\_23fill %>%

select(-c(start\_lat, start\_lng, end\_lat, end\_lng, start\_station\_id,end\_station\_id, end\_station\_name))

trips\_23fill$date <- as.Date(trips\_23fill$started\_at)

trips\_23fill$month <- format(as.Date(trips\_23fill$date), "%m")

trips\_23fill$day <- format(as.Date(trips\_23fill$date), "%d")

trips\_23fill$year <- format(as.Date(trips\_23fill$date), "%Y")

trips\_23fill$day\_of\_week <- format(as.Date(trips\_23fill$date),"%A")

trips\_23fill$ride\_length <- (as.double(difftime(trips\_23fill$ended\_at, trips\_23fill$started\_at))) /60

str(trips\_23fill) #confirm data type is double [True]

trips\_23fill$ride\_length <- as.numeric(as.character(trips\_23fill$ride\_length)) #change datatype to numeric for further analysis

trips\_23fill$started\_at <- as.POSIXct(trips\_23fill$started\_at)

trips\_23fill$time <- format(trips\_23fill$started\_at, format = "%H:%M")

# Convert to POSIXct object

trips\_23fill$day\_of\_week <- ordered(trips\_23fill$day\_of\_week, levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))

trips\_23fill<- trips\_23fill[!(trips\_23fill$ride\_length<0),] # remove blanks

trips\_23fill <- subset( trips\_23fill, select = -start\_station\_name )

trips\_23fill %>%

mutate(day\_of\_week = wday(started\_at, label = TRUE)) %>% #creates weekday field using wday()

group\_by(member\_casual, day\_of\_week) %>% #groups by usertype and weekday

trips\_23fill$day\_of\_week <- ordered(trips\_23fill$day\_of\_week, levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))

trips\_23fill$day\_of\_week <- format(as.Date(trips\_23fill$date), "%A")

trips\_23fill %>% #total rides broken down by weekday

group\_by(member\_casual, day\_of\_week) %>%

summarize(number\_of\_rides = n()) %>%

arrange(member\_casual, day\_of\_week) %>%

ggplot(aes(x = day\_of\_week, y = number\_of\_rides, fill = member\_casual,)) + geom\_col(position = "dodge") +

labs(x='Day of Week', y='Total Number of Rides', title='Rides per Day of Week', fill = 'Type of Membership') +

scale\_y\_continuous(breaks = c(100000,200000,300000, 400000, 500000), labels = c("100k","200k","300k", "400K", "500K"))

#Pie Chart

data <- data.frame(member\_casual = c("casual", "member"),

n = c(2545425, 3953523))

data$percentage <- (data$n / total) \* 100 # Calculate percentages

pie\_chart1 <- ggplot(data = data, aes(x = "", y = n, fill = member\_casual)) +

geom\_bar(stat = "identity", width = 1) +

coord\_polar(theta = "y") +

theme\_void() +

labs(fill = "Member Type", title = "Total Users") +

geom\_text(aes(label = paste0(round(percentage, 1), "%")),

position = position\_stack(vjust = 0.5))

print(pie\_chart1)

#Average ride times

men<-data.frame(aggregate(trips\_23fill$ride\_length ~ trips\_23fill$member\_casual, FUN = mean))

colnames(men) <- c("member\_casual", "average\_ride\_length")

ggplot(men, aes(fill=member\_casual,y= average\_ride\_length, x = member\_casual))+

geom\_bar(stat = "identity", width = .5)+

labs(title="Average Ride Times", y="Average Ride Length", x="Member Type")

#Rides Per month

trips\_23fill %>%

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

summarize(total\_rides = n()) %>%

ggplot(aes(x= month, y= total\_rides, fill= member\_casual))+

geom\_bar(stat = "identity",position = "dodge")+

labs(title= "Total Rides by Month", x= "Month", y= "Total Rides")+

scale\_y\_continuous(breaks =c (100000,200000,300000,400000,500000), labels = c("100K","200K", "300K","400K","500K"))

#Average ride times by week

trips\_23fill %>%

group\_by(day\_of\_week,member\_casual) %>%

summarize(average= mean(ride\_length)) %>%

ggplot(aes(x=day\_of\_week,y=average, fill= member\_casual))+geom\_bar(stat = "identity", position ="dodge")+

labs(title = "Average Ride Time by Week", x= "Day of Week", y="Average Ride Time")