library(tidyverse)

library(lubridate)

library(ggplot2)

getwd()

setwd('E:/DA\_Course/CaseStudy\_Cyclistics/Extraction/')

##adding datasets

aug2022 <- read\_csv("202208-divvy-tripdata.csv")

sep2022 <- read\_csv("202209-divvy-tripdata.csv")

oct2022 <- read\_csv("202210-divvy-tripdata.csv")

nov2022 <- read\_csv("202211-divvy-tripdata.csv")

dec2022 <- read\_csv("202212-divvy-tripdata.csv")

jan2023 <- read\_csv("202301-divvy-tripdata.csv")

feb2023 <- read\_csv("202302-divvy-tripdata.csv")

mar2023 <- read\_csv("202303-divvy-tripdata.csv")

apr2023 <- read\_csv("202304-divvy-tripdata.csv")

may2023 <- read\_csv("202305-divvy-tripdata.csv")

jun2023 <- read\_csv("202306-divvy-tripdata.csv")

jul2023 <- read\_csv("202307-divvy-tripdata.csv")

str(aug2022)

library(dplyr)

feb2023<- mutate(feb2023,started\_at=as\_datetime(started\_at))

feb2023<- mutate(feb2023,ended\_at=as\_datetime(ended\_at))

##stacking all data into one large dataframe

alltrips <- bind\_rows(aug2022,sep2022,oct2022,nov2022,dec2022,jan2023,feb2023,mar2023,apr2023,may2023,jun2023,jul2023)

## CLEANING

colnames(alltrips)

nrow(alltrips)

dim(alltrips)

head(alltrips)

str(alltrips)

alltrips<- mutate(alltrips,started\_at=as\_datetime(started\_at),ended\_at=as\_datetime(ended\_at))

summary(alltrips)

table(alltrips$member\_casual)

## lets say our company wanted to retag casuals to customers and members as subscribers

alltrips<- alltrips %>%

mutate(member\_casual=recode(member\_casual,"casual"="customer","member"="subscriber"))

table(alltrips$member\_casual)

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

head(alltrips)

alltrips$month <-format(as.Date(alltrips$started\_at),"%m")

alltrips$year <- format(as.Date(alltrips$started\_at),"%Y")

alltrips$day <- format(as.Date(alltrips$started\_at),"%d")

alltrips$day\_of\_week <- format(as.Date(alltrips$started\_at),"%A")

alltrips$ride\_length<-difftime(alltrips$ended\_at,alltrips$started\_at)

alltrips$ride\_length<-seconds\_to\_period(alltrips$ride\_length)

is.numeric(alltrips2$ride\_length)

alltrips %>% filter(alltrips$ride\_length<0) %>% select(ride\_length) ##There seem to be negative values in ride\_length,turns out, it was because of some maintenance job

alltrips$ride\_length <- period\_to\_seconds(alltrips$ride\_length)

mean(alltrips$ride\_length)

alltrips$ride\_length <-as.numeric(as.character(alltrips$ride\_length))

alltrips2<-(alltrips[!(alltrips$ride\_length<0),])

str(alltrips2)

is.numeric(alltrips2$ride\_length)

alltrips2<- data.frame(alltrips2)

##Analysing Data

mean(alltrips2$ride\_length)

median(alltrips2$ride\_length)

max(alltrips2$ride\_length)

min(alltrips2$ride\_length)

summary(alltrips2$ride\_length)

aggregate(alltrips2$ride\_length ~ alltrips2$member\_casual,FUN=mean)

aggregate(alltrips2$ride\_length ~ alltrips2$member\_casual,FUN=median)

aggregate(alltrips2$ride\_length ~ alltrips2$member\_casual,FUN=max)

aggregate(alltrips2$ride\_length ~ alltrips2$member\_casual,FUN=min)

aggregate(alltrips2$ride\_length ~ alltrips2$member\_casual + alltrips2$day\_of\_week,FUN=mean)

alltrips2$day\_of\_week <- ordered(alltrips2$day\_of\_week, levels=c("Sunday","Monday","Tuesday","Wednesday","Thursday","Friday","Saturday"))

wday(alltrips2$started\_at,label=TRUE)

alltrips2 %>%

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

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

summarise(no\_of\_rides=n(),

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

arrange(member\_casual,weekday)

##Visualizing

#on the basis of no. of rides

alltrips2 %>%

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

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

summarise(no\_of\_rides=n(),average\_duration=mean(ride\_length)) %>%

ggplot(aes(x=weekday,y=no\_of\_rides,fill=member\_casual)) +

geom\_col(position="dodge")

#on the basis of average trip duration

alltrips2 %>%

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

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

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

ggplot(aes(x=weekday,y=average\_duration,fill=member\_casual)) + geom\_col(position="dodge")

##Export to csv

counts <- aggregate(alltrips2$ride\_length ~ alltrips2$member\_casual + alltrips2$day\_of\_week,FUN=mean)

write.csv(counts,file="E:/DA\_Course/CaseStudy\_Cyclistics/avg\_ride\_length\_r.csv")