Google project capstone

Daniel B. Ojugbele

2024-04-15

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

## installing and loading the required packages

install.packages(“tidyverse”)

library(tidyverse)

install.packages(“janitor”)

library(janitor)

install.packages(“lubridate”)

library(lubridate)

install.packages(“ggplot2”)

library(ggplot2)

## getting the working directory

getwd()

## setting the working directory

setwd(“C:/Users/USER/Desktop/Divvy Exercise/csv”)

## importing data to data-frame in R

library(readr) df1<-read\_csv(“202004-divvy-tripdata.zip”) View(df1)

library(readr) df2<-read\_csv(“202005-divvy-tripdata.zip”)

View(df2)

library(readr) df3<-read\_csv(“202006-divvy-tripdata.zip”)

View(df3)

library(readr) df4<-read\_csv(“202007-divvy-tripdata.zip”) View(df4)

library(readr) df5<-read\_csv(“202008-divvy-tripdata.zip”) View(df5)

library(readr) df6<-read\_csv(“202009-divvy-tripdata.zip”) View(df6)

library(readr) df7<-read\_csv(“202010-divvy-tripdata.zip”) View(df7)

library(readr) df8<-read\_csv(“202011-divvy-tripdata.zip”) View(df8)

library(readr) df9<-read\_csv(“202012-divvy-tripdata.zip”) View(df9)

library(readr) df10<-read\_csv(“202101-divvy-tripdata.zip”) View(df10)

library(readr) df11<-read\_csv(“202102-divvy-tripdata.zip”) View(df11)

library(readr) df12<-read\_csv(“202103-divvy-tripdata.zip”) View(df12)

## Merging data into a single data frame

bikerides<-rbind(df1,df2,df3,df4,df5,df6,df7,df8,df9,df10,df11,df12)

dim(bikerides)

## cleaning and processing data

bikerides<-janitor::remove\_empty(dat = bikerides, which = c(“cols”)) bikerides<-janitor::remove\_empty(dat= bikerides, which = c(“rows”)) dim(bikerides)

## converting start at and end at column to date data

bikeridesstarted\_at) bikeridesended\_at)

## getting the start time and end time components from date-time

bikeridesstarted\_at) bikeridesended\_at)

## getting the start and end time

bikeridesstarted\_at) bikeridesended\_at)

## getting the ride length and changing diff date to numeric format, the duration is in second

bikeridesended\_at, bikerides$started\_at, units=c("hours")) ## changing diff date to numeric bikerides$length\_hr<-as.numeric(bikerides$length\_hr) ## getting the trip duration in mins bikerides$length\_min<-difftime(bikeridesstarted\_at, units=c(“mins”))

bikerideslength\_min)

## we will now remove row which have a trip duration equal to or less than 0

bikerides\_v2 <- na.omit(bikerides[!(bikerides$start\_station\_name == "HQ QR" | bikerides$length\_min <= 0), ])

dim(bikerides\_v2) colSums(is.na(bikerides\_v2))

## getting the structure of my data frame str(bikerides)

## now we will assign days of the week to start and end date

bikerides\_v2start\_date, label = TRUE, abbr = FALSE) bikerides\_v2end\_date, label = TRUE, abbr = FALSE) dim(bikerides\_v2) View(bikerides\_v2)

## We export the end result as CSV to work on Tableau write.csv(bikerides\_v2, “google project 1.csv”, row.names = FALSE)

## We get the avg trip durations as well as the maximun trip duration

cat(‘average duration of rides is :’, mean(bikerides\_v2$length\_min), "mins") cat(' maximum duration of rides is : ', round(max(bikerides\_v2$length\_min)/60/24), “days”)