Documentación de Procesamiento de datos.

Realizado por Fernando Hernández Esquivel

Fecha: 15-09-2022

1.-Descargar los archivos de los últimos 12 meses para su limpieza y análisis, del siguiente link:

<https://divvy-tripdata.s3.amazonaws.com/index.html>

2.-Descomprime los archivos.

3.-Guarda los archivos en la carpeta destinada al proyecto.

4.-Cambia el nombre de los archivos para una mejor manipulación.

ejemplo:

202208-divvy-tripdata.csv por 2022\_08.csv.

5.-Abrir el programa estadístico R.

6.-Crear un nuevo documento ya sea Script o Rmarkdown.

7.-Cargar la librería tidyverse, en caso de no tener descargada la librería, descargarla con el siguiente código: install.packages(“tidyverse”).

8.-Carga los archivos extensión .csv.

9.-Une los documentos en un nuevo dataframe.

10.-Revisa la estructura de los datos.

11.-Crea una columna que se llame “ride\_length”. Calcula la extensión de cada viaje restando la columna “started\_at” de la columna “ended\_at”.

12.-Crea una columna llamada “day\_of\_week” y calcula el día de la semana en el que empezó cada viaje.

13.-Retirar las observaciones que en la variable start\_station\_name = “HQ QR” y en la variable ride\_length < 0.

14.-Revisa los valores nulos por columna.

15.-Elimina los valores nulos del dataframe.

16.-Cambiar el tipo de dato a factor de las variables: rideable\_type, member\_casual, day\_of\_week.

17.-Realizar análisis descriptivo.

18.-Crear 2 archivos .csv con los resultados de las consultas para uso externo.

19.-Realizar gráficos con la paquetería ggplot2.

20.-Agregar las 3 recomendaciones al informe.

21.-Guardar el archivo Rmarkdown.

22.-Agregar la documentación al código.