

Profesora: Dra. María de los Ángeles Rodríguez
06MIOT_10_A_2019-20_Sistemas y tecnologías para la gestión de la información
Actividad Guiada 2
Estudiante: Arnold Alexander Rivas Ortiz

Solución a Actividad Guiada 2 -Manejo de conjuntos de datos.

El repositorio de la actividad es: <https://github.com/arnoldarivaso/arnoldarivaso>

Carga de datos de csv(estaciones_bici.csv) en un DataFrame

Respuesta: Utilizando el lenguaje de programación Python con la librería Pandas, se consiguió cargar en una variable de tipo de DataFrame todos los registros de este archivo csv.

```
# import numpy as np
import pandas as pd

estaciones = pd.read_csv('estaciones_bici.csv', sep=";")

# print(estaciones)
```

El resultado fue una impresión por consola:

```
$ C:/Users/danie/AppData/Local/Programs/Python/Python38-32/python.exe c:/Users/danie/Desktop/Trabajos/actividad_2/actividad2.py
   Unnamed: 0  id  available  connected  download_date  estacion  free  open  ticket  total  updated
0           0    5c6050a42554172704fccdc0      9         1  2019-02-10 17:25:37.787      64    11     1     0    20  2019-02-10 17:21:13.000
1           1    5c6050a42554172704fccdc1      6         1  2019-02-10 17:25:37.787      73    14     1     1    20  2019-02-10 17:24:13.000
2           2    5c605be225541729b7d50885     20         1  2019-02-10 18:13:39.827      63     0     1     1    20  2019-02-10 18:09:16.000
3           3    5c605be225541729b7d50886      6         1  2019-02-10 18:13:39.827      64    14     1     0    20  2019-02-10 18:12:15.000
4           4    5c605be225541729b7d50887      9         1  2019-02-10 18:13:39.827      65    10     1     1    19  2019-02-10 18:09:16.000
...         ...         ...         ...         ...         ...         ...         ...         ...         ...         ...
27542    27542    5c61face25541729b7d57419      0         1  2019-02-11 23:44:00.786     260    20     1     0    20  2019-02-11 23:42:16.000
27543    27543    5c61face25541729b7d5741a     15         1  2019-02-11 23:44:00.786     261     4     1     0    19  2019-02-11 23:39:16.000
27544    27544    5c61face25541729b7d5741b      1         1  2019-02-11 23:44:00.786     268     9     1     1    10  2019-02-11 23:42:16.000
27545    27545    5c61face25541729b7d5741c      1         1  2019-02-11 23:44:00.786     269    14     1     0    15  2019-02-11 23:39:16.000
27546    27546    5c61face25541729b7d5741d      8         1  2019-02-11 23:44:00.786     276    12     1     1    20  2019-02-11 23:42:16.000
```

- Obtención de número de estaciones con total de 30:

Respuesta:

Para filtrar el DataFrame por la cantidad de estaciones que tienen de total 30, se utilizó el siguiente algoritmo:

```
totales = estaciones[estaciones["total"] == 30]

estations = totales.estacion.drop_duplicates()
print(estations.count())
```

Profesora: Dra. María de los Ángeles Rodríguez
06MIOT_10_A_2019-20_Sistemas y tecnologías para la gestión de la información
Actividad Guiada 2
Estudiante: Arnold Alexander Rivas Ortiz

El resultado del código anterior fue:

```
$ C:/Users/danie/AppData/Local/Programs/Python/Python38-32/python.exe c:/Users/danie/Desktop/Trabajos/actividad_2/actividad2.py  
11
```

- Número de estación con la media más alta de bicis disponibles

Respuesta:

```
estations = totales.estacion.drop_duplicates()  
# print(estations.count())  
  
# print(estaciones)  
  
medias = estaciones[["available", "estacion"]]  
aux = 0  
estacion_max_mediu = 0  
for i in medias.estacion.drop_duplicates():  
    media = medias[medias["estacion"].isin([i])].median()  
    # print("Estacion: ", i, " Media: ", media.available)  
    if media.available > aux:  
        aux = media.available  
        estacion_max_mediu = media.estacion  
  
print("Estacion: ", estacion_max_mediu, " Media: ", aux)  
  
# media = estaciones.estacion.median()  
# print(medias)
```

El resultado del algoritmo anterior fue:

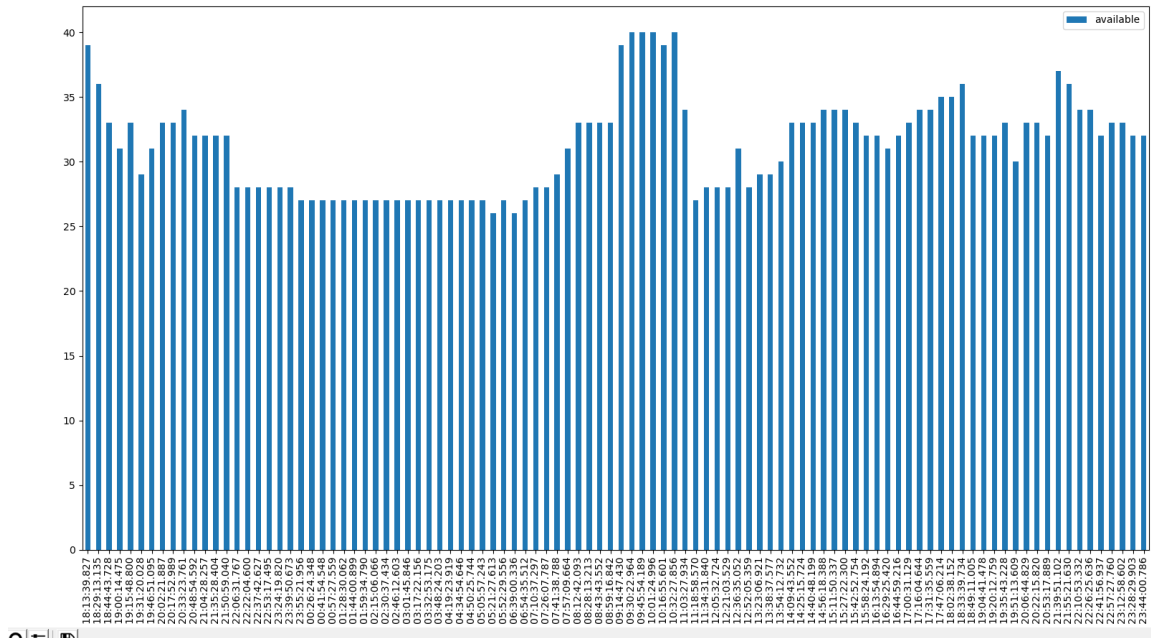
```
$ C:/Users/danie/AppData/Local/Pr  
Estacion: 50.0 Media: 32.0
```

- Realizar el histograma de la estación de bicis disponibles:

Para sacar el histograma de la estación de bicis disponibles se utiliza el siguiente código:

```
data_estacion_max_mediu = data_estacion_max_mediu[["download_date", "avail  
able"]]  
data_estacion_max_mediu.set_index('download_date', inplace=True)  
data_estacion_max_mediu.plot(kind="bar")
```

El resultado es:



- Realizar gráfica con la línea temporal de bicis disponibles

Respuesta: Para lograr esta grafica también se utiliza pandas, con la función plop. A continuación se muestra la forma.

```
# print("Estacion: ", station_max_mediu, " Media: ", aux)
data_estacion_max_mediu = data_estacion_max_mediu[["download_date", "available"]]
data_estacion_max_mediu.set_index('download_date', inplace=True)
data_estacion_max_mediu.plot()
data_estacion_max_mediu.plot(kind="bar")

plt.show()
```

Y el Resultado del código fue:

Profesora: Dra. María de los Ángeles Rodríguez
06MIOT_10_A_2019-20_Sistemas y tecnologías para la gestión de la información
Actividad Guiada 2
Estudiante: Arnold Alexander Rivas Ortiz

Figure 1

