

Ejercicios No. 3:Archivos

La siguiente lista de ejercicios permiten hacer uso del lenguaje, el interprete y el editor de su preferencia.

1. De la carpeta *scripts_files.ii*. El script *script_measures.py* debe realizar las siguientes operaciones:

- (Lectura de archivos) Se debe cargar los datos del archivo *age_user.csv* utilizando Numpy, este contiene las edades de las personas de la variable data.

```
from numpy import genfromtxt
ages= genfromtxt("age_users.csv",delimiter=";",dtype=int)
```

- (Tipos de datos) Se debe adicionar una columna con las edades al dataframe.
- (Manejo de estructuras de datos) De acuerdo al nombre cree una lista que contenga el genero de las personas y adiconelo al dataframe.
- (Operaciones con estructuras) Filtre el dataset para he indique: ¿Cuántas mujeres tiene mas de 30 años?, ¿Cuántas hombres tiene menos de 40 años?.

```
# example of filter Name and Age
dataframe[(dataframe.Name=="James") & (dataframe.Age<60)]
```

- (Funciones) ¿Cuál es la edad que tiene el 90% de las personas más jóvenes? (percentiles). Ordene los datos de acuerdo a la: la edad, el nombre.
- (Transformación de datos) Cree una nueva columna en el dataframe que indique en que rango de edades se encuentra la edad de cada individuo use la función *cut* de pandas. Los rangos son:(59, 69], (20, 29], (39, 49], (29, 39], (49, 59], (69, 79]

```
pd.cut(x=dataframe.Age, bins=[20, 29, 39, 49])
```

- (Gráficos) Realice un gráfico de barras que presente las edades.
- (Gráficos) Realice un gráfico de barras que presente la cantidad de individuos por estado.

2. En la carpeta *scripts_files.ii*. Cree un script en Python que:

- (Lectura de archivos) Cargue los datos del archivo *Phone_users.csv* que posee información acerca de los usuarios de una compañía telefónica y que se encuentran en mora. Responda las siguientes preguntas de acuerdo a los datos.(El delimitador en lugar de ; es #)

```
import pandas as pd
pd.read_csv("Phone_users.csv", delimiter="#")
```

- (Funciones) Encuentre los indicadores de tendencia central, dispersión y posición de las edades de las personas de este conjunto de datos.
- (Operaciones con estructuras) Cree la columna latitud y longitud a partir de la columna location del dataset. Encuentre los individuos que se encuentren mas cercano a uno de los individuos que usted elija.

```
import pandas as pd
dframe=pd.read_csv("Phone_users.csv", delimiter="#")
long=dframe["Location"].str.split(", ", expand=True)
```

3. Utilice el script *script_weather.py* e implemente las siguientes operaciones.
 - Cargue el archivo *Datos_Climaticos.csv*
 - Elimine o no tenga en cuenta los NAs en los datos.
 - Calcule mediana, media, mediana agrupada, varianza, varianza poblacional, desviación estándar, covarianza y correlación de la temperatura, la humedad y el contaminante PM10.
 - Grafique el histograma de la temperatura, la humedad y el contaminante PM10.
4. Cree un nuevo script y utilizando el conjunto de datos iris de Pydataset
 - Realice una descripción de los datos utilizando la función describe.

```
from pydataset import data
iris = data('iris')
iris.describe()
```

- Determinar la media y la mediana de las diferentes especies presentes en el conjunto de datos.
 - Calcule la desviación estándar.
5. Cree un nuevo script que:
 - Cargue los datos del archivo *Docentes_de_planta.csv*, indexar los datos por facultad e identifique los tipos de datos de cada uno de las columnas del dataframe.
 - Presente la cantidad de profesores que tienen titulo de Doctorado, elimine los NAs y grafique.
 - Presente en una misma tabla los profesores que posee titulo de doctorado, asistente y titular. Elimine los NAs y grafique los datos.
 - Presente los datos de conteo acerca de los programas académicos de CIENCIAS AGROINDUSTRIALES, CIENCIAS BÁSICAS Y TECNOLÓGICAS. Grafique.

```
dataframe.loc['CIENCIAS AGROINDUSTRIALES']
```

Ejercicios No. 4: Gráficos

La siguiente lista de ejercicios permiten hacer uso del lenguaje, el interprete y el editor de su preferencia.

1. De la carpeta *scripts_files.ii*. Se debe crear un script , con nombre *script_base.py* que debe realizar las siguientes operaciones:

- Grafique las funciones *sin* y *cos* utilizando la función *linspace* de numpy

```
import numpy as np
np.linspace(-np.pi, np.pi, 256, endpoint=True)
```

- Utilice el conjunto de datos *trees* desde Pydataset, realice la descripción de los datos y grafique. Personalice títulos, etiquetas y ejes.
- Utilice el conjunto de datos *iris* desde seaborn, realice un grafico de enjambre (*swarmplot*) por cada especie y la longitud de su petalo

```
irisdt=sns.load_dataset("iris")
sns.swarmplot(x="species", y="petal_length", data=irisdt)
```

- Utilizando el dataframe creado en *script_measures.py*, que contiene información [Name, State, Age], cree un grafico *stripplot* x=Name, y=Age.
- Utilizando el conjunto de datos *tips* desde seaborn, realice un gráfico de barras y de boxplot, x=day, y=total_bill.

2. Cree un script que utilice el conjunto de datos llamado iris.

- Separe en nuevas variables las especies: setosa, versicolor, virginica
- Cree histogramas utilizando FacetGrid por cada una de las columnas, sepal_length, sepal_width, petal_length, petal_width.

```
sns.FacetGrid(iris, hue="species", height=5) \
    .map(sns.distplot, "sepal_length") \
    .add_legend();
```

3. Utilizando el conjunto de datos del archivo *Phone_users.csv*, que contiene los usuarios que se encuentran en mora.

- Cree un gráfico dinámico utilizando bokeh de los usuarios y las edades.
- Adicion información utilizando tooltips.

4. Utilizando la ubicación de los registros en el archivo *Phone_users.csv*,

- Cree un script que localice en un mapa los usuarios que se encuentran en mora.
- Situe un marcador en la localizacion 3.3753, -76.5337