Leer DATOS CON PANDAS

Import libraries

In [1]:
```
# Import the packages that we will be using
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Leer datos Cargar datos

In [2]:
```
from sklearn import datasets
iris = datasets.load_iris()
```

In [3]:
```
# Define the col names for the iris dataset
col_names = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'Flower']

# Dataset url
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"

# Load the dataset from URL
iris_df = pd.read_csv(url, header=None, names=col_names)
print(iris_df.head())
```

```
   sepal_length  sepal_width  petal_length  petal_width       Flower
0           5.1          3.5           1.4          0.2  Iris-setosa
1           4.9          3.0           1.4          0.2  Iris-setosa
2           4.7          3.2           1.3          0.2  Iris-setosa
3           4.6          3.1           1.5          0.2  Iris-setosa
4           5.0          3.6           1.4          0.2  Iris-setosa
```

In [4]:
```
# Mostrar las primeras filas del DataFrame para verificar que los datos se hayan carga
iris_df.head()
```

Out[4]:

	sepal_length	sepal_width	petal_length	petal_width	Flower
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

In [5]:
```
#Get the number of rows and columns
iris_df.shape
```

Out[5]:
```
(150, 5)
```

In [6]:
```
#get number of rows
num_rows = iris_df.shape[0]

#Print the number of rows
print('The number of rows is :' + str(num_rows))
```

```
The number of rows is :150
```

In [7]:
```
#Get number of cos
Ncols = iris_df.shape[1]

#Print the number of columns
print('The number of columns is :' + str(Ncols))
```

```
The number of columns is :5
```

EXPLORAR DATAFRAME

In [8]:
```
iris_df.columns
```

Out[8]:
```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'Flower'], dtype
='object')
```

In [9]:
```
#Get the data types of each column
iris_df.dtypes
```

Out[9]:
```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
Flower           object
dtype: object
```

In [10]:
```
#Obtener un nuevo dataframe con algunas columnas
iris_df2 = iris_df[['sepal_length', 'sepal_width', 'Flower']]

iris_df2.head()
```

Out[10]:

	sepal_length	sepal_width	Flower
0	5.1	3.5	Iris-setosa
1	4.9	3.0	Iris-setosa
2	4.7	3.2	Iris-setosa
3	4.6	3.1	Iris-setosa
4	5.0	3.6	Iris-setosa

In [11]:
```
iris_df2.head(10)
```

Out[11]:

	sepal_length	sepal_width	Flower
0	5.1	3.5	Iris-setosa
1	4.9	3.0	Iris-setosa
2	4.7	3.2	Iris-setosa
3	4.6	3.1	Iris-setosa
4	5.0	3.6	Iris-setosa
5	5.4	3.9	Iris-setosa
6	4.6	3.4	Iris-setosa
7	5.0	3.4	Iris-setosa
8	4.4	2.9	Iris-setosa
9	4.9	3.1	Iris-setosa

In [12]:
```
iris_df2.tail(3)
```

Out[12]:

	sepal_length	sepal_width	Flower
147	6.5	3.0	Iris-virginica
148	6.2	3.4	Iris-virginica
149	5.9	3.0	Iris-virginica

Estadisticas de las variables

Calcular y visualizar estadisticas descriptivas

In [13]:
```
#calculate and visualize the mean and the standard deviation of the data
iris_df2.describe()
```

Out[13]:

	sepal_length	sepal_width
count	150.000000	150.000000
mean	5.843333	3.054000
std	0.828066	0.433594
min	4.300000	2.000000
25%	5.100000	2.800000
50%	5.800000	3.000000
75%	6.400000	3.300000
max	7.900000	4.400000

In [14]:
```
iris_df.describe()
```

Out[14]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

In [15]:
```
#IDENTIFICAR ROWS/cols woth mission information

iris_df.isnull()
```

Out[15]:

	sepal_length	sepal_width	petal_length	petal_width	Flower
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
145	False	False	False	False	False
146	False	False	False	False	False
147	False	False	False	False	False
148	False	False	False	False	False
149	False	False	False	False	False

150 rows × 5 columns

In [16]:
```
#IMPRIMIR ESTADISTICA DESCRIPTIVA
iris_df.describe()
```

Out[16]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000