Activity: work with the iris dataset

Marisol Rodríguez Mejía | A01640086

```
# Import the packages that we will be using
import pandas as pd


# Define where you are running the code: colab or local
RunInColab          = True     # (False: no  | True: yes)

# If running in colab:
if RunInColab:
    # Mount your google drive in google colab
    from google.colab import drive
    drive.mount('/content/drive')

    # Find location
    #!pwd
    #!ls
    #!ls "/content/drive/My Drive/Colab Notebooks/MachineLearningWithPython/"

    # Define path del proyecto
    Ruta            = "/content/drive/My Drive/Colab Notebooks/MachineLearningWithPython/"

else:
    # Define path del proyecto
    Ruta            = ""
```

```
Mounted at /content/drive
```

```
# url string that hosts our .csv file
url = Ruta + "datasets/iris/iris.csv"

# Read the .csv file and store it as a pandas Data Frame
df = pd.read_csv(url)
```

1. Calculate the statistical summary for each quantitative variables. Explain the results.

- Identify the name of each column

```
df.columns
```

```
Index(['5.1', '3.5', '1.4', '0.2', 'Iris-setosa'], dtype='object')
```

- Identify the type of each column

```
df.dtypes
```

```
5.1            float64
3.5            float64
1.4            float64
0.2            float64
Iris-setosa     object
dtype: object
```

- Minimum, maximum, mean, average, median, standard deviation

```
df.describe()
```

	5.1	3.5	1.4	0.2
count	149.000000	149.000000	149.000000	149.000000
mean	5.848322	3.054362	3.773826	1.206040
std	0.828594	0.435810	1.760543	0.760354

2. Are there missing data? There is no missing data

```
df.isnull().sum()
```

```
5.1             0
3.5             0
1.4             0
0.2             0
Iris-setosa     0
dtype: int64
```

3. Create a new dataset containing only the petal width and length and the type of Flower

```
df2 = df.copy()
df2.drop(df2.columns[[2, 3]], axis = 1, inplace = True)
df2.head()
```

	5.1	3.5	Iris-setosa
0	4.9	3.0	Iris-setosa
1	4.7	3.2	Iris-setosa
2	4.6	3.1	Iris-setosa
3	5.0	3.6	Iris-setosa
4	5.4	3.9	Iris-setosa

4. Create a new dataset containing only the setal width and length and the type of Flower

```
df3 = df.copy()
df3.drop(df3.columns[[0, 1]], axis = 1, inplace = True)
df3.head()
```

	1.4	0.2	Iris-setosa
0	1.4	0.2	Iris-setosa
1	1.3	0.2	Iris-setosa
2	1.5	0.2	Iris-setosa
3	1.4	0.2	Iris-setosa
4	1.7	0.4	Iris-setosa

5. Create a new dataset containing the setal width and length and the type of Flower encoded as a categorical numerical column

```
df3.iloc[:, -1] = df3.iloc[:, -1].replace({"Iris-setosa": 1, "Iris-versicolor": 2, "Iris-virginica": 3})
df3.head()
```

	1.4	0.2	Iris-setosa
0	1.4	0.2	1
1	1.3	0.2	1
2	1.5	0.2	1
3	1.4	0.2	1
4	1.7	0.4	1