Manuel Alejandro Preciado Morán A01639643

En este documento se encuentran las 3 practicas de iris

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

# Dataset url
path = "/home/alex/TC1002S/NotebooksStudents/A01639643/iris/iris.csv"
header = ["sepal_length", "sepal_width", "petal_length", "petal_width", "Clas
# Load the dataset
dataset = pd.read_csv(path, names=header)

# Print the dataset
dataset
```

Out[ ]:

	sepal_length	sepal_width	petal_length	petal_width	Class
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
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

In [ ]:
```
#Print the number of rows(observations)
Nrows = dataset.shape[0]
Nrows
```

Out[ ]: 150

In [ ]:
```
#Print the number of columns
Ncol = dataset.shape[1]
Ncol
```

Out[ ]: 5

In [ ]:
```
# count1 = 0
# count2 = 0
# count3 = 0
# for i in range(Nrows):
#     if sC[i] == "Iris-setosa":
#         count1 += 1
#     elif sC[i] == "Iris-virginica":
#         count2 += 1
#     elif sC[i] == "Iris-versicolor":
#         count3 += 1
# count1
# count2
# count3
sC = dataset["Class"]
counts = pd.Series(sC).value_counts()
print("Iris-setosa: ", counts.get("Iris-setosa"))
print("Iris-virginica: ", counts.get("Iris-virginica"))
print("Iris-versicolor: ", counts.get("Iris-versicolor"))
```

```
Iris-setosa:  50
Iris-virginica:  50
Iris-versicolor:  50
```

Load the iris.csv file in your computer and understand the dataset

How many observations (rows) are in total? 149 How many variables (columns) are in total? What do they represent? 5 How many observations are for each type of flower? Iris Setosa: 50 Iris Versicolour: 50 Iris Virginica: 50

What is the type of data for each variable? float object

What are the units of each variable? cm

PART 2

Exploring the content of the data set

Use the shape method to determine the numbers of rows and columns in a data frame. This can be used to confirm that we have actually obtained the data the we are expecting.

Based on what we see below, the data set being read here has Nr rows, corresponding to Nr observations, and Nc columns, corresponding to Nc variables in this particular data file.

In [ ]:
```
print(dataset.shape)
print("Number of rows", dataset.shape[0])
print("Number of columns/variables", dataset.shape[1])
```

```
(150, 5)
Number of rows 150
Number of columns/variables 5
```

If we want to show the entire data frame we would simply write the following:

In [ ]:
```
dataset
```

Out[ ]:

	sepal_length	sepal_width	petal_length	petal_width	Class
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
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

As you can see, we have a 2-Dimensional object where each row is an independent observation and each coloum is a variable.

Now, use the the head() function to show the first 5 rows of our data frame

In [ ]:
```
dataset.head()
```

Out[ ]:

	sepal_length	sepal_width	petal_length	petal_width	Class
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

Also, you can use the the tail() function to show the last 5 rows of our data frame

In [ ]:
```
dataset.tail()
```

Out[ ]:

	sepal_length	sepal_width	petal_length	petal_width	Class
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

The columns in a Pandas data frame have names, to see the names, use the columns method:

To gather more information regarding the data, we can view the column names with the following function

In [ ]:
```
dataset.columns
```

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

Be aware that every variable in a Pandas data frame has a data type. There are many different data types, but most commonly you will encounter floating point values (real numbers), integers, strings (text), and date/time values. When Pandas reads a text/csv file, it guesses the data types based on what it sees in the first few rows of the data file. Usually it selects an appropriate type, but occasionally it does not. To confirm that the data types are consistent with what the variables represent, inspect the dtypes attribute of the data frame.

In [ ]:
```
dataset.dtypes
```

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

Summary statistics, which include things like the mean, min, and max of the data, can be useful to get a feel for how large some of the variables are and what variables may be the most important.

In [ ]:
```
# Summary statistics for the quantitative variables
print("MEAN \n",dataset[["sepal_length", "sepal_width", "petal_length", "peta
print("MIN \n", dataset[["sepal_length", "sepal_width", "petal_length", "peta
print("MAX \n",dataset[["sepal_length", "sepal_width", "petal_length", "petal
```

```
MEAN
 sepal_length    5.843333
sepal_width     3.057333
petal_length    3.758000
```