ROME BUSINESS SCHOOL

Title: Data Science Languages. Assignment

Please go through all the requirements for this assignment and read them carefully.

PURPOSE OF THE ASSIGNMENT:

Assess the ability of each student to use the python programming languages to manipulate data.

TASK OF THE ASSIGNMENT:

Step 1: load the iris dataset from the file we used during the course lectures. Hint: in case you do not have it any more, you can find it at the following link: http://wpage.unina.it/rafmiele/2022_RBS/iris_per_python/iris.txt

Step 2: build some basic statistics on the 4 numerical variables. Comment the results.

Step 3: build the same statistics on the 4 numerical variables for each group of iris that is defined by the modalities of the 5th variable (called "Species").

Step 4: extract all the iris whose measures fulfil **both** the following conditions:

- Petal Length is higher or equal than 2.5
- Petal Width is smaller than 1.8

Compute the frequency distribution of the target variable (Species) on the selected flowers. Which is the most frequent specie?

Step 5 (final deliverable): describe, in a very small document (no more than one page) the findings you got and the most important code snippets you had to write.

TECHNICAL REQUIREMENTS:

You can use google colab as we did in our lectures or any other python environment you prefer.

GRADING CRITERIA:

1. Ability to write code that answers the question.
2. Ability to explain and comment the results.

ADDITIONAL SOURCES TO BE USED:

Anything you can find on the internet. Do not use "people"!

REFERENCING:

The lectures' slide decks.

PLAGIARISM DISCLAIMER (NOTE FOR THE STUDENT): "I hereby declare that this assignment is my own work and all references created are as per the Harvard Referencing System''.