**NAME: AQSA TAUHEED (2303-KHI-DEG-011)**

**PAIRING WITH : MAVIA ALAM KHAN (2303.KHI.DEG.017)**

**&**

**MOHAMMAD HUSSAM(2033.KHI.DEG.020)**

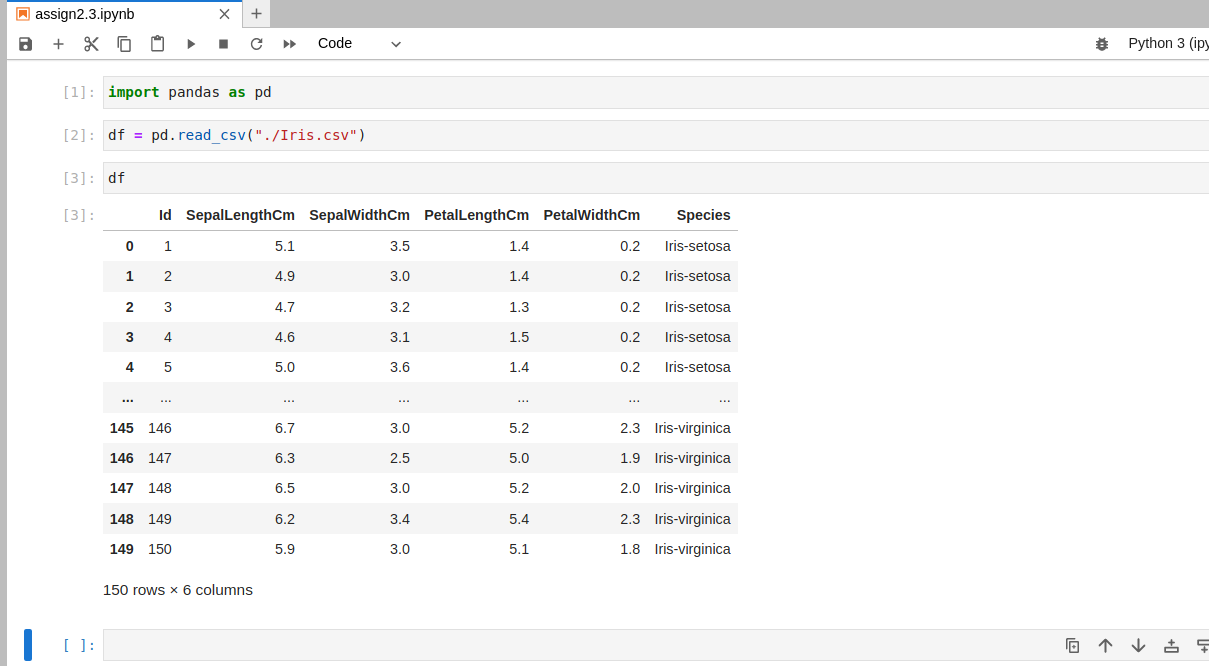
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ASSIGNMENT 2.3**

Download the Iris dataset from  
https://www.kaggle.com/datasets/uciml/iris  
and write a program that loads the CSV file and answers  
what is the average sepal length for each of three iris species.  
  
**Solution:**

**STEP-1:**

We downloaded the dataset from Iris website which is CSV file.



import pandas as pd

We have to import pandas library as pd , here pd used as alias for word pandas .

df = pd.read\_csv("./Iris.csv")

After importing the pandas library, we can load the contents of a Iris.csv file into a DataFrame, for this we use function pd.read\_csv that takes a ‘Iris.csv’ file as input and returns a DataFrame object. By assigning the output of this function to the variable **df**, we're creating a new DataFrame object that contains the data from the that file.

**STEP-2:**



we're using **df[["SepalLengthCm","Species"]]** to select only the SepalLengthCm and Species columns from the original df DataFrame. By assigning the result to a variable new\_df, we're creating a new DataFrame object that contains only these two columns, as show in image above.

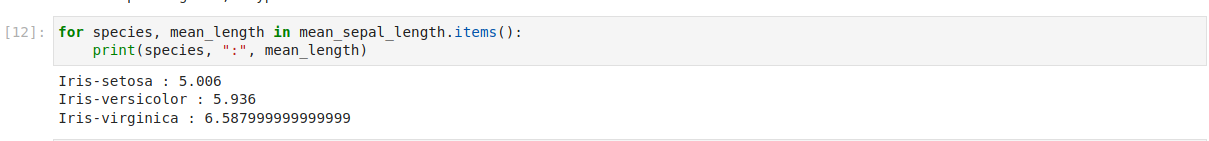
**STEP -3:**



We first grouped the rows of the df DataFrame by their values in the Species column using the groupby() method and assigns the result to the variable groups.Then, the mean value of the SepalLengthCm column for each group in groups is calculated using the mean() method, which returns a pandas Series object containing the mean sepal length for each unique value of the Species column. This Series object is assigned to the variable mean\_sepal\_length ,after this we printed mean\_sepal\_length as shown in image above.

The output for mean\_ sepal\_length form is a **pandas** **Series** object with two columns as shown above in image.

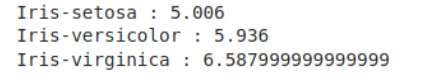
**STEP-4:**



The for loop is used iterating over the mean\_sepal\_length series using the items() method, which returns an iterator over the (index, value) pairs of the series.For each iteration, the index is assigned to the species variable and the value is assigned to the mean\_length variable. This allows us to loop through the different species in the mean\_sepal\_length series.

**print(species, ":", mean\_length)** would print out the species name and its corresponding mean sepal length

**OUTPUT:**



Iris-setosa : 5.006  
Iris-versicolor : 5.936  
Iris-virginica : 6.587999999999999