NAME: MOHAMMAD HUSSAM(2033.KHI.DEG.020)

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

&

AQSA TAUHEED(2303.KHI.DEG.011)

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.

```python
[1]: import pandas as pd

[2]: df = pd.read_csv("./Iris.csv")

[3]: df
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

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:

```
[6]: new_df = df[["SepalLengthCm","Species"]]
```

```
[7]: new_df
```

[7]:

	SepalLengthCm	Species
0	5.1	Iris-setosa
1	4.9	Iris-setosa
2	4.7	Iris-setosa
3	4.6	Iris-setosa
4	5.0	Iris-setosa
...	...	...
145	6.7	Iris-virginica
146	6.3	Iris-virginica
147	6.5	Iris-virginica
148	6.2	Iris-virginica
149	5.9	Iris-virginica

150 rows × 2 columns

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:

```
[10]: groups = df.groupby('Species')
      mean_sepal_length = groups['SepalLengthCm'].mean()
```

```
[11]: mean_sepal_length
```

```
[11]: Species
      Iris-setosa        5.006
      Iris-versicolor    5.936
      Iris-virginica     6.588
      Name: SepalLengthCm, dtype: float64
```

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:

```
[12]: for species, mean_length in mean_sepal_length.items():
          print(species, ":", mean_length)

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

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
```

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