2.3 Assignment Pandas

Peer Members:

- Syed Muhammad Raqim Ali Shah (2303.KHI.DEG.008)
- Maaz Javaid Siddique (2303.KHI.DEG.004)

Question:

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:

In this first picture first we import pandas by using **import pandas as pd** and
get the data that we downloaded in local directory **Iris.csv** by using **read_csv**
that read the csv file and assign in the data variable.

```
import pandas as pd
data=pd.read_csv('/home/maazjavaidsiddique/Downloads/Iris.csv')
data
```

	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

After the loading of the data, we want to take average of SepalLengthCm on the basis of each Species. So first we group Species by using groupby and find the mean of SepalLengthCm by using mean() and save in the average_of_sepal_length. The output show that the average of SepalLengthCm by Species.

```
average_of_sepal_length=data.groupby('Species')['SepalLengthCm'].mean()
average_of_sepal_length

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