

Unit 2.3 Graded Assignment:

Muhammad Khan (2303.khi.deg.027)

Qadeer Hussain (2303.KHI.DEG.006)

Daily Assignment :

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.

Answer:

First of all we Import the pandas library.

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

```
[8]: df= pd.read_csv("Iris.csv")  
df
```

```
[8]:
```

	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

`read_csv()` is a function provided by the pandas library that reads a CSV file and returns a dataframe object.

```
avg_sepallenthCm = df.groupby('Species')['SepallenthCm'].mean()  
avg_sepallenthCm
```

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

`groupby('species')` this method groups the data in the DataFrame `df` by the values in the species column. It creates a `GroupBy` object that groups the data based on unique values in the 'species' column.

`['SepallenthCm']` is used to specify the column `SepallenthCm` for which we want to calculate the average. It selects only the 'SepallenthCm' column from the grouped data.

`mean()` is calculates the mean (average) of the selected column, which is `SepallenthCm` in this case, for each group in the `GroupBy` object.