

Unit 2.3 Graded Assignment: Pandas

Instructions:

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.

Submitted by:

1. Ali Nasir (2303.KHI.DEG.012)
2. Saif ur Rehman (2303.KHI.DEG.007)

Solution:

- In this assignment we have first imported pandas library that I was for data cleaning and manipulation.
- Then we use `pd.head()` to see the first 5 rows of datasets.
- Finally we use group by operation to separate the unique values in the data set i.e. Iris-setosa, iris-virginica, iris-versicolor then we find mean of Sepal length of each species.

```
import pandas as pd
[1] ✓ 0.4s

data = pd.read_csv('Datasets/Iris.csv')
data.sample(5)
[19] ✓ 0.0s

...
  Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
25  26              5.0           3.0           1.6           0.2  Iris-setosa
136 137              6.3           3.4           5.6           2.4  Iris-virginica
52  53              6.9           3.1           4.9           1.5  Iris-versicolor
66  67              5.6           3.0           4.5           1.5  Iris-versicolor
78  79              6.0           2.9           4.5           1.5  Iris-versicolor

dataFrame = data.groupby('Species')['SepalLengthCm'].mean()
dataFrame
[20] ✓ 0.0s

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