

## Members

Huzefa Anver (2303.KHI.DEG.002)

Syed Mohammad Anjil Hussain Rizvi (2303.KHI.DEG.031)

---

1. Import pandas and refer to it as pd
2. pd.read\_csv to read the csv file downloaded earlier from Kaggle
3. File is in the current directory so we don't need to specify the path
4. Save the dataframe in iris\_ds variable and printing it to see it is successfully loaded

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

```
[22]: iris_ds = pd.read_csv("Iris.csv")  
iris_ds
```

```
[22]:
```

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

150 rows × 6 columns

5. As we have to calculate Avg SepalLength of each species so, with  
`iris_ds["Species"].unique()` we are finding every unique value in species column and putting it into an array.
6. Saving the unique species array into a species variable and then we are iterating through its elements using a for loop and storing each element inside `"i"`.
7. Inside the loop :
  - a. 1<sup>st</sup> step : Conditioning used to separate whole data of each species and then separating the sepal length column, saving it to `"sepal_l"`
  - b. 2<sup>nd</sup> step: calculating the mean of the `"sepal_l"`
  - c. 3<sup>rd</sup> step: appending a dictionary inside a list declared above as `avg_spl_l_each_species` containing species name and average length.
8. Printing the final list

```
[30]: species = iris_ds["Species"].unique()  
species|
```

```
[30]: array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

```
[31]: avg_spl_l_each_species = []  
for i in species:  
    sepal_l = iris_ds[iris_ds["Species"] == i]["SepalLengthCm"]  
    avrg = sepal_l.mean()  
    avg_spl_l_each_species.append({"Species":i,"AverageSepalLength":avrg})
```

```
[35]: avg_spl_l_each_species
```

```
[35]: [{'Species': 'Iris-setosa', 'AverageSepalLength': 5.006},  
      {'Species': 'Iris-versicolor', 'AverageSepalLength': 5.936},  
      {'Species': 'Iris-virginica', 'AverageSepalLength': 6.5879999999999998}]
```

9. Putting the list into the new data frame to give it a nice visualization.

```
[36]: pd.DataFrame(avg_spl_l_each_species)
```

```
[36]:
```

	Species	AverageSepalLength
0	Iris-setosa	5.006
1	Iris-versicolor	5.936
2	Iris-virginica	6.588