

# **Assignment - 1**

"CSE 476"

## "Data Mining Lab"

B.Sc. Engineering in Computer Science and Engineering
Department of CSE
Bangladesh University of Business and Technology

## **Submitted By**

Name: Habibullah

**ID:** 18192103080

## Under the supervision of

Khan Md. Hasib Assistant Professor

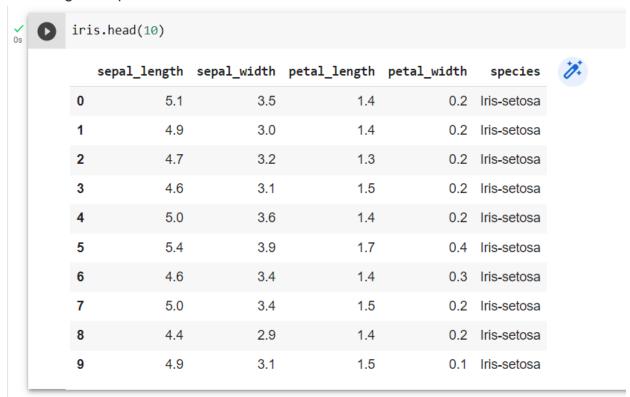
Department of CSE Bangladesh University of Business and Technology CO1. Apply data preprocessing steps (such as: Viewing your data, Handling duplicates, Column cleanup, DataFrame slicing, selecting, and extracting) in the following dataset - https://www.kaggle.com/datasets/selinraja/irish-data.

### 1. Importing pandas Library

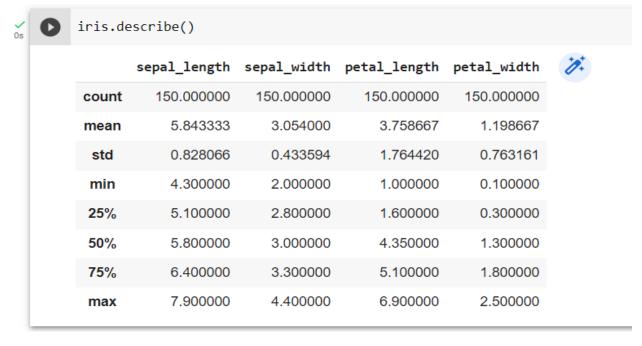
### 2. Uploading the dataset & Viewing the first and last portion of data

• iri	is = pd.read_csv is	(" <u>/content/Ir</u>	is_Data.csv")			
	sepal_length	sepal_width	petal_length	petal_width	species	0
0	5.1	3.5	1.4	0.2	Iris-setosa	
1	4.9	3.0	1.4	0.2	Iris-setosa	
2	4.7	3.2	1.3	0.2	Iris-setosa	
3	4.6	3.1	1.5	0.2	Iris-setosa	
4	5.0	3.6	1.4	0.2	Iris-setosa	
14	5 6.7	3.0	5.2	2.3	Iris-virginica	
14	6.3	2.5	5.0	1.9	Iris-virginica	
14	6.5	3.0	5.2	2.0	Iris-virginica	
14	8 6.2	3.4	5.4	2.3	Iris-virginica	
14	9 5.9	3.0	5.1	1.8	Iris-virginica	
150	rows × 5 columns					

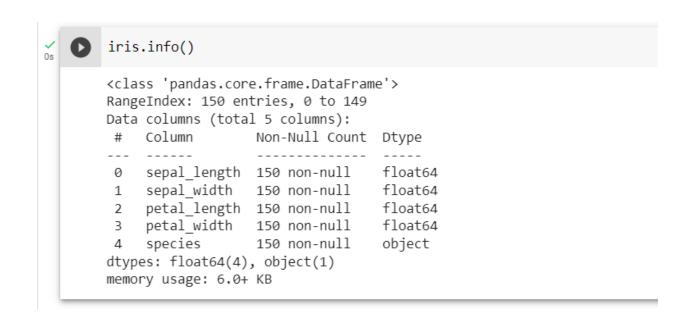
3. Viewing the top/first 10 rows of the dataset.



4. Showing the description of the whole dataset with sepal and petal length and width.



5. Showing the info of the dataset.



## 6. Dropping the duplicate data

disp:	display(iris.drop_duplicates())				
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica
147 rd	ows × 5 columns				

### 7. Column cleanup

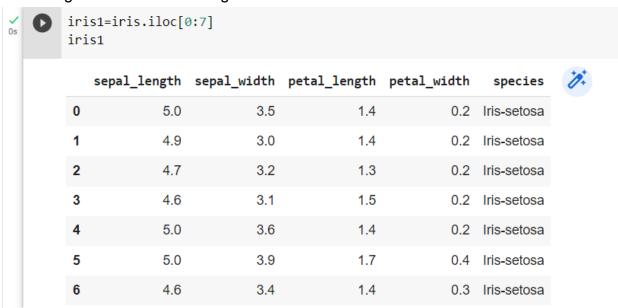
```
[25] for x in iris.index:
           if iris.loc[x, "sepal_length"] > 5:
             iris.loc[x, "sepal_length"] = 5
           iris.head(10)
             sepal_length sepal_width petal_length petal_width
                                                                          species
         0
                       5.0
                                      3.5
                                                      1.4
                                                                    0.2 Iris-setosa
         1
                                                                    0.2 Iris-setosa
                       4.9
                                      3.0
                                                      1.4
                                      3.2
                                                                    0.2 Iris-setosa
                       4.7
                                                      1.3
         3
                       4.6
                                      3.1
                                                      1.5
                                                                    0.2 Iris-setosa
                                      3.6
                       5.0
                                                                    0.2 Iris-setosa
                                                      1.4
         5
                                      3.9
                                                                    0.4 Iris-setosa
                       5.0
                                                      1.7
                       4.6
                                      3.4
                                                      1.4
                                                                    0.3 Iris-setosa
         7
                       5.0
                                      3.4
                                                      1.5
                                                                    0.2 Iris-setosa
                                      2.9
                                                                    0.2 Iris-setosa
                       4.4
                                                      1.4
         9
                       4.9
                                      3.1
                                                      1.5
                                                                    0.1 Iris-setosa
```

8. Showing the unique data of a specific column.

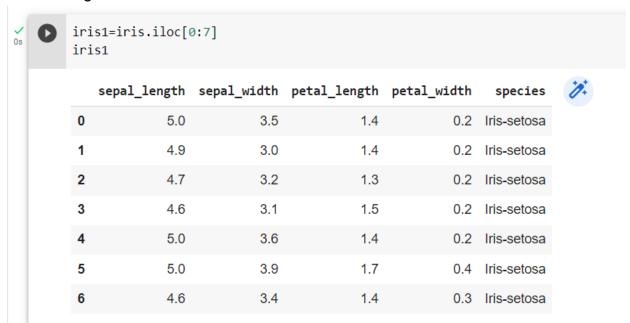
```
print("Species")
print(iris['species'].unique())

Species
['Iris-setosa' 'Iris-versicolor' 'Iris-virginica']
```

## 9. Showing the data frame slicing.



#### 10. Showing the data frame selection.



### 11. Showing the data frame slicing.

	sepal_length	sepal_width	petal_length	petal_width
0	5.0	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
145	5.0	3.0	5.2	2.3
146	5.0	2.5	5.0	1.9
147	5.0	3.0	5.2	2.0
148	5.0	3.4	5.4	2.3
149	5.0	3.0	5.1	1.8

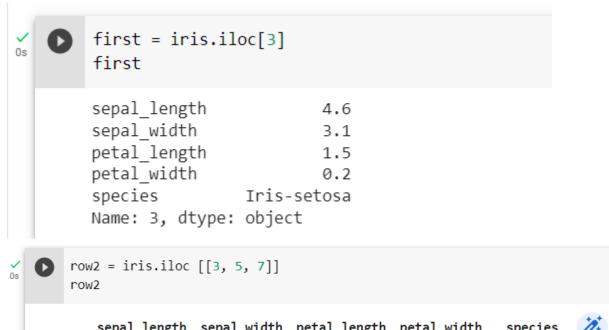
150 rows × 4 columns

copy=iris[['sepal\_length','sepal\_width','petal\_length']]
copy

	sepal_length	sepal_width	petal_length
0	5.0	3.5	1.4
1	4.9	3.0	1.4
2	4.7	3.2	1.3
3	4.6	3.1	1.5
4	5.0	3.6	1.4
145	5.0	3.0	5.2
146	5.0	2.5	5.0
147	5.0	3.0	5.2
148	5.0	3.4	5.4
149	5.0	3.0	5.1

150 rows × 3 columns

## 12. Showing the data frame extracting.



	sepal_length	sepal_width	petal_length	petal_width	species
3	4.6	3.1	1.5	0.2	Iris-setosa
5	5.0	3.9	1.7	0.4	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa