## **Bangladesh University of Business & Technology**

# Dhaka, Commerce College Road Mirpur-2, Dhaka-1216



Course Code : CSE 476

Course Title : Data Mining Lab

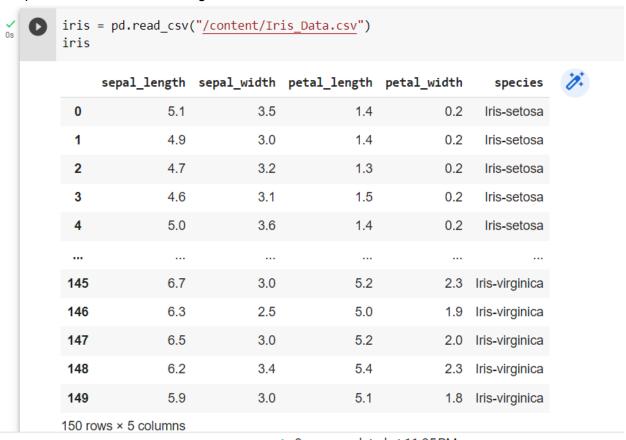
## **Assignment - 1**

There Bad II eath
(han Md. Hasib
ssistant Professor
ept of CSE, BUBT

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

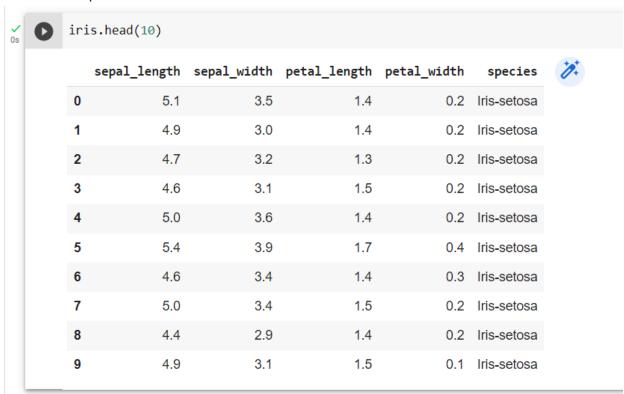
#### 1.Import library

## 2. Upload the dataset & Viewing the data

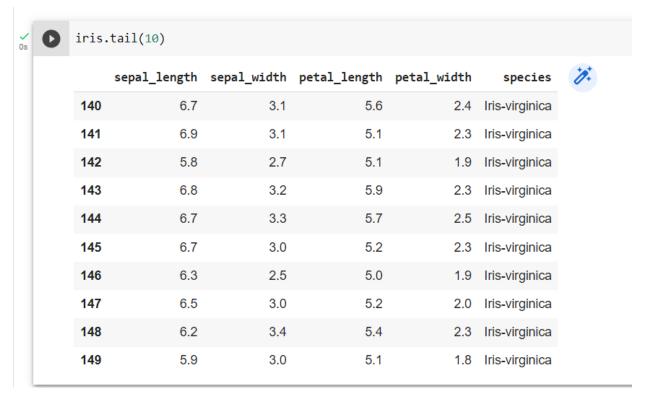


✓ 0s completed at 11:35 PM

3. View the top 10 rows of the dataset.



4. View the bottom 10 rows of the dataset.



## 5. Showing the description of the whole dataset.





	separ_rengen	separ_width	pecar_rengen	petar_wracii
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

## 6. Showing the info ot the dataset.



iris.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype
0	sepal_length	150 non-null	float64
1	sepal_width	150 non-null	float64
2	petal_length	150 non-null	float64
3	petal_width	150 non-null	float64
4	species	150 non-null	object

dtypes: float64(4), object(1)

memory usage: 6.0+ KB

## 7. Dropping the duplicate data



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 rows × 5 columns

#### 8. Column cleanup

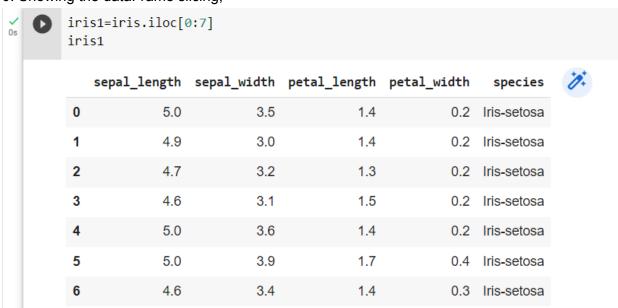
```
v [25] for x in iris.index:
           if iris.loc[x, "sepal_length"] > 5:
             iris.loc[x, "sepal_length"] = 5
           iris.head(10)
             sepal_length sepal_width petal_length petal_width
                                                                          species
          0
                       5.0
                                      3.5
                                                      1.4
                                                                    0.2 Iris-setosa
          1
                       4.9
                                      3.0
                                                                    0.2 Iris-setosa
                                                      1.4
          2
                       4.7
                                      3.2
                                                                    0.2 Iris-setosa
                                                      1.3
          3
                       4.6
                                      3.1
                                                      1.5
                                                                    0.2 Iris-setosa
                       5.0
                                      3.6
                                                      1.4
                                                                    0.2 Iris-setosa
          5
                       5.0
                                      3.9
                                                                    0.4 Iris-setosa
                                                      1.7
                       4.6
                                      3.4
                                                      1.4
                                                                    0.3 Iris-setosa
          7
                       5.0
                                      3.4
                                                      1.5
                                                                    0.2 Iris-setosa
                       4.4
                                      2.9
                                                      1.4
                                                                    0.2 Iris-setosa
          9
                       4.9
                                      3.1
                                                      1.5
                                                                    0.1 Iris-setosa
```

9. Showing the unique data of specific column.

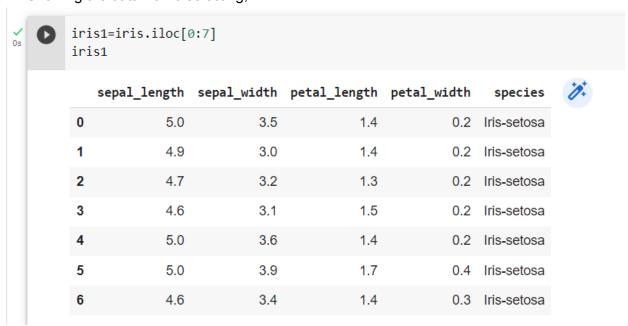
```
print("Species")
print(iris['species'].unique())

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

## 10. Showing the dataFrame slicing,



## 11. Showing the dataFrame selecting,



## 12. Showing the dataFramee 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

1

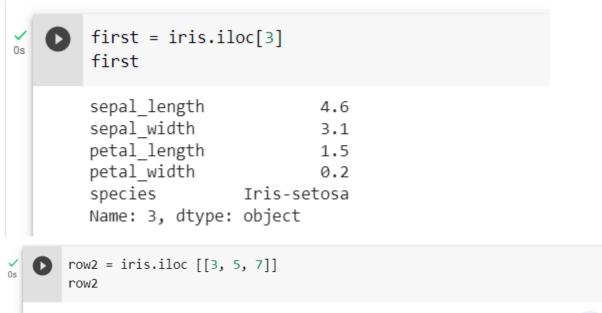
150 rows × 4 columns

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

	sepal_length	sepal_width	petal_length	1
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

## 13. Showing the dataFramee 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