## Bangladesh University of Business & Technology Dhaka,

Commerce College Road
Mirpur-2, Dhaka-1216



Course Code: CSE 476

**Course Name: Data Mining Lab** 

## Assignment 01

Submitted By	Submitted To
Sadeka Jafrin Id:18192103069 Intake:41 sec:03	Khan Md. Hasib Assistant Professor Dept 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.

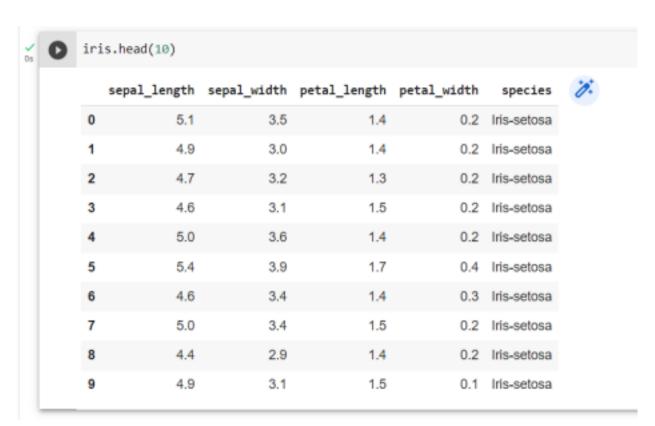
## 1.Import library

2. Upload the dataset & Viewing the data

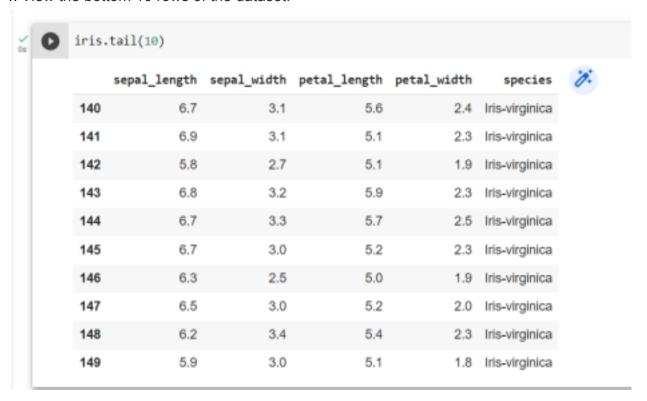
iris iris	= pd.read_csv	("/content/Ir:	is_Data.csv")		
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
145	6.7	3.0	5.2	2.3	Iris-virginica
146	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
150 rd	ws × 5 columns		. Oo oomulata		

Os completed at 11:35PM

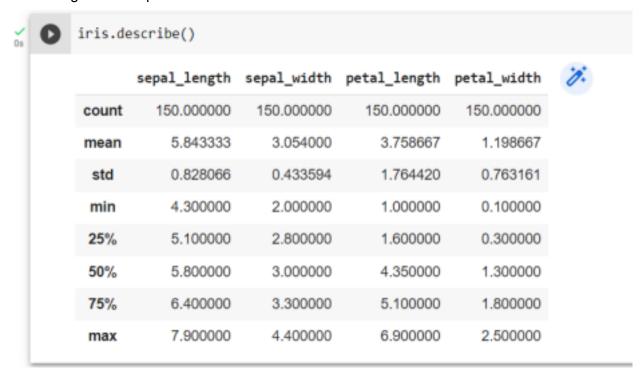
3. View the top 10 rows of the dataset.



4. View the bottom 10 rows of the dataset.



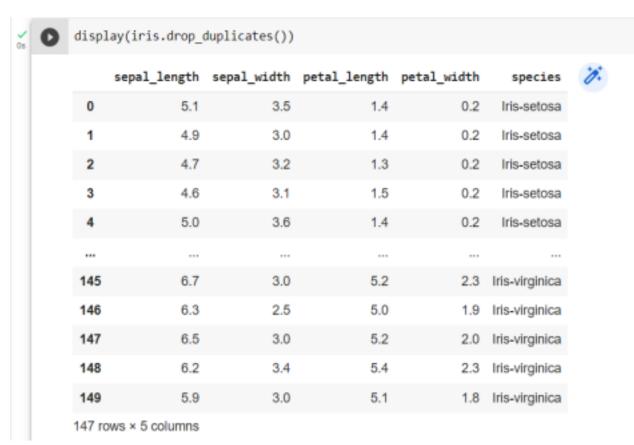
5. Showing the description of the whole dataset.



6. Showing the info ot the dataset.



7. Dropping the duplicate data



## 8. Column cleanup

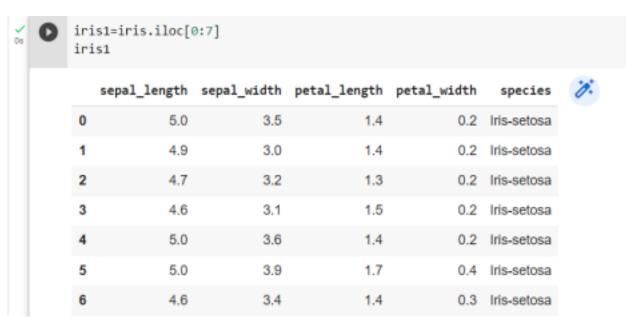
```
[25] for x in iris.index:
           if iris.loc[x, "sepal_length"] > 5:
             iris.loc[x, "sepal_length"] = 5
           iris.head(10)
             sepal_length sepal_width petal_length petal_width
                                                                          species
        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

9. Showing the unique data of specific column.

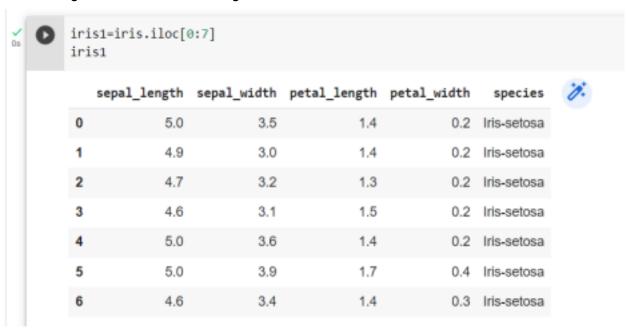
```
print("Species")
print(iris['species'].unique())

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

10. Showing the dataFrame slicing,



11. Showing the dataFrame selecting,



12. Showing the dataFramee slicing.

	iris:	?=iris.loc[:,':	sepal_length'	'petal_width'	]	
		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 ows × 4 columns	3.0	5.1	1.8	
			anath! !cons!	uddth! Instal	loogth!!!	
0	copy=	iris[['sepal_l	engtn , sepai	_width , petal	_tength []]	
		sepal_length	sepal_width	petal_length	·	
	0	sepal_length 5.0	sepal_width	petal_length	V.	
	0				J.	
		5.0	3.5	1.4	J.	
	1	5.0 4.9	3.5 3.0	1.4 1.4	J.	
	1 2	5.0 4.9 4.7	3.5 3.0 3.2	1.4 1.4 1.3	J.	
	1 2 3	5.0 4.9 4.7 4.6	3.5 3.0 3.2 3.1	1.4 1.4 1.3 1.5	J.	
	1 2 3 4	5.0 4.9 4.7 4.6 5.0	3.5 3.0 3.2 3.1 3.6	1.4 1.4 1.3 1.5	J.	
	1 2 3 4 	5.0 4.9 4.7 4.6 5.0	3.5 3.0 3.2 3.1 3.6	1.4 1.4 1.3 1.5 1.4	J.	
	1 2 3 4 	5.0 4.9 4.7 4.6 5.0 	3.5 3.0 3.2 3.1 3.6 	1.4 1.4 1.3 1.5 1.4 	J.	
	1 2 3 4  145 146	5.0 4.9 4.7 4.6 5.0  5.0	3.5 3.0 3.2 3.1 3.6  3.0 2.5	1.4 1.4 1.3 1.5 1.4  5.2 5.0	J.	
	1 2 3 4  145 146 147	5.0 4.9 4.7 4.6 5.0  5.0 5.0	3.5 3.0 3.2 3.1 3.6  3.0 2.5 3.0	1.4 1.4 1.3 1.5 1.4  5.2 5.0	J.	

13. Showing the dataFramee extracting.