

Name: Lisha Ingawale

Roll No:2193140

Class: LY CSE IS 1

Assignment 1:

1. Getting data to work with: Download the sample dataset locally for any application (Kaggle)
2. Setting up the working directory.
3. Unpacking the data. Decompress the file locally.
4. Looking at the data. Display the top (10) and bottom (10) of the file.
5. Measuring the length of the data set. Count the number of lines in the file.
6. Encode the categorical data
7. Plot a graph and give your insights for the application selected cases.

The screenshot shows a Google Colab notebook titled "BDA_ASSIG_1.ipynb". The code cell contains the following Python code:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

iris = pd.read_csv('iris.csv')
iris.head(5)
```

The output of the code is a preview of the first 5 rows of the Iris dataset:

	Id	SepallengthCm	SepalwidthCm	PetallengthCm	PetalwidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

Below the table, the output of the `iris.info()` command is shown:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype  
---  -
0    Id              150 non-null   int64  
1    SepallengthCm   150 non-null   float64
2    SepalwidthCm    150 non-null   float64
```

colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBlk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

+ Code + Text

```
[ ] iris.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column          Non-Null count  Dtype
---  --
 0   Id              150 non-null   int64
 1   SepalLengthCm  150 non-null   float64
 2   SepalWidthCm   150 non-null   float64
 3   PetalLengthCm  150 non-null   float64
 4   PetalWidthCm   150 non-null   float64
 5   Species         150 non-null   object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

```
[ ] iris.describe()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	6.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBlk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

+ Code + Text

```
[ ] iris.groupby('Species').min()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
Species					
Iris-setosa	1	4.3	2.3	1.0	0.1
Iris-versicolor	51	4.9	2.0	3.0	1.0
Iris-virginica	101	4.9	2.2	4.5	1.4

```
[ ] iris.groupby('Species').max()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
Species					
Iris-setosa	50	5.8	4.4	1.9	0.6
Iris-versicolor	100	7.0	3.4	5.1	1.8
Iris-virginica	150	7.9	3.8	6.9	2.5









