

Roll No. 412039

AIM: Getting Introduced to Data Analytics Libraries in Python.

Code & Output;

In [1]:

```
import pandas as pd
import numpy as np
```

In [3]:

```
# Reading a dataset
df = pd.read_csv("Downloads/iris.csv")
```

In [4]:

```
df.head()
```

Out[4]:

	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

In [6]:

```
df.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
```

In [7]:

```
# Calculating all required statistical measure
df.describe()
```

Out[7]:

	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	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

In [9]:

```
# Calculating variance
df.var()
```

C:\Users\user\AppData\Local\Temp\ipykernel_4616\1568254755.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

```
df.var()
```

Out[9]:

```
Id                1887.500000
SepalLengthCm      0.685694
SepalWidthCm       0.188004
PetalLengthCm      3.113179
PetalWidthCm       0.582414
dtype: float64
```

In [10]:

```
df.median()
```

C:\Users\user\AppData\Local\Temp\ipykernel_4616\530051474.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

```
df.median()
```

Out[10]:

```
Id                75.50
SepalLengthCm      5.80
SepalWidthCm       3.00
PetalLengthCm      4.35
PetalWidthCm       1.30
dtype: float64
```

In [16]:

```
import statistics  
statistics.mode(df['PetalLengthCm'])
```

Out[16]:

1.5

Conclusion : Thus, we have seen how statistical measures such as mean, median, standard deviation, variance are calculated in Data Analytics using Python.