Intro to Pandas DataFrame:

Pandas DataFrame:

- DataFrame is a two-dimensional labelled data structure
- A DataFrame comprises of rows and columns.

In [1]:

```
import pandas as pd
```

In [2]:

```
df1 = pd.DataFrame({"Name" : ['Bob','Sam','Anne'], "Marks" : [76,25,92]})
df1
```

Out[2]:

| | Name | Marks |
|---|---|---|
| **0** | Bob | 76 |
| **1** | Sam | 25 |
| **2** | Anne | 92 |

DataFrame In-Built Function:

- head():

```
If you want to look at the first five rows of a dataframe then you can use head function.
```

In [6]:

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

In [7]:

```
iris.head()
```

Out[7]:

| | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|---|---|---|---|---|---|
| **0** | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| **1** | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| **2** | 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| **3** | 4.6 | 3.1 | 1.5 | 0.2 | setosa |
| **4** | 5.0 | 3.6 | 1.4 | 0.2 | setosa |

- tail():

```
If you want to look at the last few rows present in a dataframe then you can use tail funct
ion.
```

In [8]:

```
iris.tail()
```

Out[8]:

|  | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|---|---|---|---|---|---|
| **145** | 6.7 | 3.0 | 5.2 | 2.3 | virginica |
| **146** | 6.3 | 2.5 | 5.0 | 1.9 | virginica |
| **147** | 6.5 | 3.0 | 5.2 | 2.0 | virginica |
| **148** | 6.2 | 3.4 | 5.4 | 2.3 | virginica |
| **149** | 5.9 | 3.0 | 5.1 | 1.8 | virginica |

- shape:

```
If we want to know how many rows or columns are persent in a dataframe then we can use the
shape.
```

In [11]:

```
iris.shape
```

Out[11]:

```
(150, 5)
```

- describe():

```
If we want a basic describtion of a dataframe then we can use the describe function.
```

In [12]:

```
iris.describe()
```

Out[12]:

|  | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width |
|---|---|---|---|---|
| **count** | 150.000000 | 150.000000 | 150.000000 | 150.000000 |
| **mean** | 5.843333 | 3.057333 | 3.758000 | 1.199333 |
| **std** | 0.828066 | 0.435866 | 1.765298 | 0.762238 |
| **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 |

.iloc() and .loc():

```
This two methods are used for the purpose to extract individual rows and columns.
```

- .iloc():

```
It will help us to extract rows and columns with respect to their index. [.iloc -> i refers
to index]
```

In [13]:

```
iris.head()
```

Out[13]:

|   | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|---|---|---|---|---|---|
| **0** | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| **1** | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| **2** | 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| **3** | 4.6 | 3.1 | 1.5 | 0.2 | setosa |
| **4** | 5.0 | 3.6 | 1.4 | 0.2 | setosa |

In [14]:

```
iris.iloc[0:3 , 0:2]
```

Out[14]:

|   | Sepal.Length | Sepal.Width |
|---|---|---|
| **0** | 5.1 | 3.5 |
| **1** | 4.9 | 3.0 |
| **2** | 4.7 | 3.2 |

- .loc():

```
Here we can't give the index value for the columns rather we have to give the names.
```

In [15]:

```
iris.loc[5:10 , ('Sepal.Length','Species')]
```

Out[15]:

|   | Sepal.Length | Species |
|---|---|---|
| **5** | 5.4 | setosa |
| **6** | 4.6 | setosa |
| **7** | 5.0 | setosa |
| **8** | 4.4 | setosa |
| **9** | 4.9 | setosa |
| **10** | 5.4 | setosa |

Droping Columns:

```
It is used to drop the columns.
```

In [16]:

```
iris.drop('Species',axis = 1)
```

Out[16]:

|  | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width |
|---|---|---|---|---|
| **0** | 5.1 | 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** | 6.7 | 3.0 | 5.2 | 2.3 |
| **146** | 6.3 | 2.5 | 5.0 | 1.9 |
| **147** | 6.5 | 3.0 | 5.2 | 2.0 |
| **148** | 6.2 | 3.4 | 5.4 | 2.3 |
| **149** | 5.9 | 3.0 | 5.1 | 1.8 |

150 rows × 4 columns

Dropping Rows:

```
It is used to drop the rows.
```

In [17]:

```
iris.drop([1,2,3],axis = 0)
```

Out[17]:

|  | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|---|---|---|---|---|---|
| **0** | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| **4** | 5.0 | 3.6 | 1.4 | 0.2 | setosa |
| **5** | 5.4 | 3.9 | 1.7 | 0.4 | setosa |
| **6** | 4.6 | 3.4 | 1.4 | 0.3 | setosa |
| **7** | 5.0 | 3.4 | 1.5 | 0.2 | setosa |
| **...** | ... | ... | ... | ... | ... |
| **145** | 6.7 | 3.0 | 5.2 | 2.3 | virginica |
| **146** | 6.3 | 2.5 | 5.0 | 1.9 | virginica |
| **147** | 6.5 | 3.0 | 5.2 | 2.0 | virginica |
| **148** | 6.2 | 3.4 | 5.4 | 2.3 | virginica |
| **149** | 5.9 | 3.0 | 5.1 | 1.8 | virginica |

147 rows × 5 columns

In [ ]: