Lab No. 02 Objective Investigating the dataset using pandas

In [24]:
```python
import pandas as pd
```

In [29]:
```python
df = pd.read_csv('data/iris.csv')
df.head(1)
```

Out[29]:

	sepal.length	sepal.width	petal.length	petal.width	variety
0	5.1	3.5	1.4	0.2	Setosa

In [30]:
```python
df.index
```

Out[30]: RangeIndex(start=0, stop=150, step=1)

In [31]:
```python
df.columns
```

Out[31]: Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',
       'variety'],
      dtype='object')

In [32]:
```python
df.shape
```

Out[32]: (150, 5)

In [33]:
```python
type(df)
```

Out[33]: pandas.core.frame.DataFrame

In [34]:
```python
type(df["sepal.length"])
```

Out[34]: pandas.core.series.Series

In [35]:
```python
df.values
```

In [36]:
```python
act_cols = df.columns
```

In [37]:
```python
new_cols = ['sl', 'sw', 'pl', 'pw', 'flowers']
df.columns = new_cols
df.head(1)
```

Out[37]:

	sl	sw	pl	pw	flowers
0	5.1	3.5	1.4	0.2	Setosa

In [38]:
```python
# df.columns=act_cols
# df.head(1)
```

```
In [39]:  type(df["sl"])
```

Out[39]: pandas.core.series.Series

```
In [40]:  type(df["sw"])
```

Out[40]: pandas.core.series.Series

```
In [41]:  type(df["pl"])
```

Out[41]: pandas.core.series.Series

```
In [42]:  type(df["pw"])
```

Out[42]: pandas.core.series.Series

```
In [43]:  df.count()
```

Out[43]: sl   150
          sw   150
          pl   150
          pw   150
          flowers   150
          dtype: int64

```
In [44]:  df.describe()
```

Out[44]:

	sl	sw	pl	pw
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

```
In [45]:  df.sort_values('flowers', ascending=False)[1:5]
```

Out[45]:

	sl	sw	pl	pw	flowers
111	6.4	2.7	5.3	1.9	Virginica
122	7.7	2.8	6.7	2.0	Virginica
121	5.6	2.8	4.9	2.0	Virginica
120	6.9	3.2	5.7	2.3	Virginica

```python
In [46]: df.sort_index(ascending=False)[0:3]
```

Out[46]:

	sl	sw	pl	pw	flowers
149	5.9	3.0	5.1	1.8	Virginica
148	6.2	3.4	5.4	2.3	Virginica
147	6.5	3.0	5.2	2.0	Virginica

```python
In [47]: df[140:145]
```

Out[47]:

	sl	sw	pl	pw	flowers
140	6.7	3.1	5.6	2.4	Virginica
141	6.9	3.1	5.1	2.3	Virginica
142	5.8	2.7	5.1	1.9	Virginica
143	6.8	3.2	5.9	2.3	Virginica
144	6.7	3.3	5.7	2.5	Virginica

```python
In [48]: df.iloc[140:145][["sl", "pw"]]
```

Out[48]:

	sl	pw
140	6.7	2.4
141	6.9	2.3
142	5.8	1.9
143	6.8	2.3
144	6.7	2.5

```python
In [49]: # df[140:145][["sl", "pw"]]
```

```python
In [50]: df.values[10:13]
```

Out[50]: array([[5.4, 3.7, 1.5, 0.2, 'Setosa'],
          [4.8, 3.4, 1.6, 0.2, 'Setosa'],
          [4.8, 3.0, 1.4, 0.1, 'Setosa']], dtype=object)

Task

1. Create a **DataFrame** from iris bunch data type dataset available with sci-kit learn
2. Explore more DataFrame / Series methods that may help in getting insight into a given dataset using pandas