Assignment No 3 : Descriptive Statistics - Measures of Central Tendency and Variability

---

Import Required Libraries

In [1]:
```
import pandas as pd
import numpy as np
```

Read csv into Dataframe

In [2]:
```
df = pd.read_csv('iris.csv')
```

In [3]:
```
df
```

Out[3]:

	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
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

Data Preprocessing

In [4]:
```
df.shape
```

Out[4]:
```
(150, 6)
```

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
Id               0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

In [6]:
```
df.dtypes
```

Out[6]:
```
Id                 int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

In [7]:
```
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 [8]:
```
df.drop(columns = 'Id', inplace = True)
df
```

Out[8]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	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
...	...	...	...	...	...

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
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 rows × 5 columns

In [9]:
```
df.describe()
```

Out[9]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
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

Calculation of Mean for each Feature

In [10]:
```
np.mean(df['SepalLengthCm'])
```

Out[10]:
```
5.843333333333335
```

In [11]:
```
np.mean(df['SepalWidthCm'])
```

Out[11]:
```
3.0540000000000007
```

In [12]:
```
np.mean(df['PetalLengthCm'])
```

Out[12]:
```
3.7586666666666693
```

In [13]:
```
np.mean(df['PetalWidthCm'])
```

Out[13]:
```
1.1986666666666672
```

In [14]:
```
np.mean(df)
```

Out[14]:
```
SepalLengthCm    5.843333
SepalWidthCm     3.054000
PetalLengthCm    3.758667
PetalWidthCm     1.198667
dtype: float64
```

Calculation of Standard Deviation for each feature

In [15]:
```
np.std(df)
```

Out[15]:
```
SepalLengthCm    0.825301
SepalWidthCm     0.432147
PetalLengthCm    1.758529
PetalWidthCm     0.760613
dtype: float64
```

Calculating minimum value

In [16]:
```
np.min(df)
```

Out[16]:
```
SepalLengthCm           4.3
SepalWidthCm              2
PetalLengthCm             1
PetalWidthCm            0.1
Species         Iris-setosa
dtype: object
```

Calculating maximum value

In [17]:
```
np.max(df)
```

Out[17]:
```
SepalLengthCm              7.9
SepalWidthCm               4.4
PetalLengthCm              6.9
PetalWidthCm               2.5
Species         Iris-virginica
dtype: object
```

Calculating 1st quantile (25th percentile)

In [18]:
```
df.quantile(0.25)
```

Out[18]:
```
SepalLengthCm    5.1
SepalWidthCm     2.8
PetalLengthCm    1.6
PetalWidthCm     0.3
Name: 0.25, dtype: float64
```

Calculating 2nd quantile (50th percentile)

In [19]:
```
df.quantile(0.50)
```

Out[19]:
```
SepalLengthCm    5.80
SepalWidthCm     3.00
PetalLengthCm    4.35
PetalWidthCm     1.30
Name: 0.5, dtype: float64
```

Calculating 3rd quantile (75th percentile)

In [20]:
```
df.quantile(0.75)
```

Out[20]:
```
SepalLengthCm    6.4
SepalWidthCm     3.3
PetalLengthCm    5.1
PetalWidthCm     1.8
Name: 0.75, dtype: float64
```