We'll use data from "Iris"(phool) dataset

Import Libraries

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
phool = sns.load_dataset("iris")
phool.head()
```

Out[2]:

	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

Saving DataFrame into CSV File

In [3]:
```
phool.to_csv("kashti.csv")
```

Saving DataFrame into Excel File

Library for excel *pip install openpyxl*

In [6]:
```
phool.to_excel("kashti.xlsx")
```

Basic Satistics

In [8]:
```
phool.describe()
```

Out[8]:

	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

In [9]:
```
phool.head()
```

Out[9]:

	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

Droping few column and make a new Data Set

In [11]:
```
new_phool = phool.drop(['species'],axis=1)
```

In [12]:
```
new_phool.head()
```

Out[12]:

	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

Mean of Dataset

In [13]:
```
phool.mean()
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_8104\2868387341.py:1: FutureWarning: Dropp
ing of nuisance columns in DataFrame reductions (with 'numeric_only=None') is depreca
ted; in a future version this will raise TypeError.  Select only valid columns before
calling the reduction.
  phool.mean()
```

Out[13]:
```
sepal_length    5.843333
sepal_width     3.057333
petal_length    3.758000
petal_width     1.199333
dtype: float64
```

Mean using Groupby

In [14]:
```
phool.groupby(['species','sepal_length','sepal_width']).mean()
```

Out[14]:

			petal_length	petal_width
species	**sepal_length**	**sepal_width**		
setosa	**4.3**	**3.0**	1.1	0.1
	4.4	**2.9**	1.4	0.2
		3.0	1.3	0.2
		3.2	1.3	0.2
	4.5	**2.3**	1.3	0.3
...	**...**	**...**	...	...
virginica	**7.7**	**2.6**	6.9	2.3
		2.8	6.7	2.0
		3.0	6.1	2.3
		3.8	6.7	2.2
	7.9	**3.8**	6.4	2.0

127 rows × 2 columns

Value Counts of column variables

In [16]:
```
phool.value_counts(['species'])
```

Out[16]:
```
species
setosa        50
versicolor    50
virginica     50
dtype: int64
```