

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import seaborn as sns
```

In [5]:

```
df = pd.read_csv("F:\\DSE\\3rd year engineering\\5th sem\\6th sem\\DSBDA\\dataset\\Mall_
```

In [7]:

```
df.head()
```

Out[7]:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

In [9]:

```
df.mean()
```

C:\Users\hp\AppData\Local\Temp\ipykernel\_19488\3698961737.py:1: FutureWarning: The default value of numeric\_only in DataFrame.mean is deprecated. In a future version, it will default to False. In addition, specifying 'numeric\_only=None' is deprecated. Select only valid columns or specify the value of numeric\_only to silence this warning.

```
df.mean()
```

Out[9]:

```
CustomerID      100.50
Age              38.85
Annual Income (k$)  60.56
Spending Score (1-100)  50.20
dtype: float64
```

In [11]:

```
df.median()
```

C:\Users\hp\AppData\Local\Temp\ipykernel\_19488\530051474.py:1: FutureWarning: The default value of numeric\_only in DataFrame.median is deprecated. In a future version, it will default to False. In addition, specifying 'numeric\_only=None' is deprecated. Select only valid columns or specify the value of numeric\_only to silence this warning.

```
df.median()
```

Out[11]:

```
CustomerID      100.5
Age              36.0
Annual Income (k$)  61.5
Spending Score (1-100)  50.0
dtype: float64
```

In [12]:

```
df.std()
```

C:\Users\hp\AppData\Local\Temp\ipykernel\_19488\3390915376.py:1: FutureWarning: The default value of numeric\_only in DataFrame.std is deprecated. In a future version, it will default to False. In addition, specifying 'numeric\_only=None' is deprecated. Select only valid columns or specify the value of numeric\_only to silence this warning.

```
df.std()
```

Out[12]:

```
CustomerID      57.879185
Age              13.969007
Annual Income (k$)  26.264721
Spending Score (1-100)  25.823522
dtype: float64
```

In [13]:

```
df['Age'].mode()
```

Out[13]:

```
0    32
Name: Age, dtype: int64
```

In [14]:

```
df.min()
```

Out[14]:

```
CustomerID      1
Genre           Female
Age             18
Annual Income (k$)  15
Spending Score (1-100)  1
dtype: object
```

In [15]:

```
df.std(axis=1,numeric_only=True)[0:4]
```

Out[15]:

```
0    15.695010
1    35.074920
2     8.057088
3    32.300671
dtype: float64
```

In [16]:

```
df.groupby(['Genre'])['Age'].mean()
```

Out[16]:

```
Genre
Female    38.098214
Male      39.806818
Name: Age, dtype: float64
```

In [17]:

```
df_u=df.rename(columns={'Annual Income (k$)':'Income'},inplace=False)
```

In [18]:

```
(df_u.groupby(['Genre']).Income.mean())
```

Out[18]:

```
Genre
Female    59.250000
Male      62.227273
Name: Income, dtype: float64
```

## part B

In [24]:

```
iris=pd.read_csv('F:\\DSE\\3rd year engineering\\5th sem\\6th sem\\DSBDA\\dataset\\iris.
```

In [28]:

iris

Out[28]:

	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
...	...	...	...	...	...
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

150 rows × 5 columns

In [35]:

irisSet=(iris['species']=='Iris-setosa')

In [36]:

print('Iris-setosa')

Iris-setosa

In [37]:

print(iris[irisSet].describe())

	sepal_length	sepal_width	petal_length	petal_width
count	0.0	0.0	0.0	0.0
mean	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN

In [38]:

irisVer=(iris['species']=='Iris-versicolor')

In [39]:

```
print('Iris-versicolor')
```

Iris-versicolor

In [40]:

```
print(iris[irisVer].describe())
```

	sepal_length	sepal_width	petal_length	petal_width
count	0.0	0.0	0.0	0.0
mean	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN

In [ ]: