

```
In [1]: import pandas as pd
import numpy as np
from sklearn import preprocessing
df=pd.read_csv("C:\\Users\\Lenovo\\Desktop\\Sem 6\\DS & BDA\\practical\\Mall_Customers.csv")
```

```
In [2]: df.mean(numeric_only=True)
```

```
Out[2]: CustomerID          100.50
Age              38.85
Annual Income (k$)  60.56
Spending Score (1-100)  50.20
dtype: float64
```

```
In [3]: df['Age'].mean()
```

```
Out[3]: 38.85
```

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

```
Out[4]: 0    18.50
1     29.75
2     11.25
3     30.00
dtype: float64
```

```
In [5]: df.median(numeric_only=True)
```

```
Out[5]: CustomerID          100.5
Age              36.0
Annual Income (k$)  61.5
Spending Score (1-100)  50.0
dtype: float64
```

```
In [6]: df['Age'].median()
```

```
Out[6]: 36.0
```

```
In [7]: df.median(axis=1,numeric_only=True)[0:4]
```

```
Out[7]: 0    17.0
1    18.0
2    11.0
3    19.5
dtype: float64
```

```
In [8]: df.mode()
```

```
Out[8]:
```

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Female	32.0	54.0	42.0
1	2	NaN	NaN	78.0	NaN
2	3	NaN	NaN	NaN	NaN
3	4	NaN	NaN	NaN	NaN
4	5	NaN	NaN	NaN	NaN
...	...	...	...	...	...
195	196	NaN	NaN	NaN	NaN
196	197	NaN	NaN	NaN	NaN
197	198	NaN	NaN	NaN	NaN
198	199	NaN	NaN	NaN	NaN
199	200	NaN	NaN	NaN	NaN

200 rows × 5 columns

```
In [9]: df['Age'].mode()
```

```
Out[9]: 0    32
Name: Age, dtype: int64
```

```
In [10]: df.loc[:, 'Age'].mode()
```

```
Out[10]: 0    32
Name: Age, dtype: int64
```

```
In [11]: df.min()
```

```
Out[11]: CustomerID          1
Genre          Female
Age             18
Annual Income (k$)  15
Spending Score (1-100)  1
dtype: object
```

```
In [12]: df.loc[:, 'Age'].min(skipna=False)
```

```
Out[12]: 18
```

```
In [13]: df.std(numeric_only=True)
```

```
Out[13]: CustomerID          57.879185
Age          13.969007
Annual Income (k$)  26.264721
Spending Score (1-100)  25.823522
dtype: float64
```

```
In [14]: df.loc[:, 'Age'].std()
```

```
Out[14]: 13.969007331558883
```

```
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
```

```
In [19]: from sklearn import preprocessing
enc=preprocessing.OneHotEncoder()
enc_df=pd.DataFrame(enc.fit_transform(df[['Genre']]).toarray())
```

```
In [20]: enc_df
```

```
Out[20]:
```

	0	1
0	0.0	1.0
1	0.0	1.0
2	1.0	0.0
3	1.0	0.0
4	1.0	0.0
...	...	...
195	1.0	0.0
196	1.0	0.0
197	0.0	1.0
198	0.0	1.0
199	0.0	1.0

200 rows × 2 columns

```
In [21]: df_encode=df_u.join(enc_df)
```

```
In [22]: df_encode
```

```
Out[22]:
```

	CustomerID	Genre	Age	Income	Spending Score (1-100)	0	1
0	1	Male	19	15	39	0.0	1.0
1	2	Male	21	15	81	0.0	1.0
2	3	Female	20	16	6	1.0	0.0
3	4	Female	23	16	77	1.0	0.0
4	5	Female	31	17	40	1.0	0.0
...	...	...	...	...	...	...	...
195	196	Female	35	120	79	1.0	0.0
196	197	Female	45	126	28	1.0	0.0
197	198	Male	32	126	74	0.0	1.0
198	199	Male	32	137	18	0.0	1.0
199	200	Male	30	137	83	0.0	1.0

200 rows × 7 columns

```
In [23]: csv_url='https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
```

```
In [24]: col_names = ['Sepal_Length', 'Sepal_Width', 'Petal_Length', 'Petal_Width', 'Species']
```

```
In [25]: iris=pd.read_csv(csv_url,names=col_names)
```

```
In [26]: iris
```

```
Out[26]:
```

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	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
...	...	...	...	...	...
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 [27]: irisSet=(iris['Species']=='Iris-setosa')
print('Iris-setosa')
```

Iris-setosa

```
In [28]: print(iris[irisSet].describe())
```

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width
count	50.00000	50.000000	50.000000	50.00000
mean	5.00600	3.418000	1.464000	0.24400
std	0.35249	0.381024	0.173511	0.10721
min	4.30000	2.300000	1.000000	0.10000
25%	4.80000	3.125000	1.400000	0.20000
50%	5.00000	3.400000	1.500000	0.20000
75%	5.20000	3.675000	1.575000	0.30000
max	5.80000	4.400000	1.900000	0.60000

```
In [29]: irisVer=(iris['Species']=='Iris-versicolor')
print('Iris-versicolor')
```

Iris-versicolor

```
In [30]: print(iris[irisVer].describe())
```

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width
count	50.000000	50.000000	50.000000	50.000000
mean	5.936000	2.770000	4.260000	1.326000
std	0.516171	0.313798	0.469911	0.197753
min	4.900000	2.000000	3.000000	1.000000
25%	5.600000	2.525000	4.000000	1.200000
50%	5.900000	2.800000	4.350000	1.300000
75%	6.300000	3.000000	4.600000	1.500000
max	7.000000	3.400000	5.100000	1.800000

```
In [31]: irisVir=(iris['Species']=='Iris-virginica')
```

```
In [32]: print('Iris-virginica')
```

Iris-virginica

```
In [33]: print(iris[irisVir].describe())
```

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width
count	50.00000	50.000000	50.000000	50.00000
mean	6.58800	2.974000	5.552000	2.02600
std	0.63588	0.322497	0.551895	0.27465
min	4.90000	2.200000	4.500000	1.40000
25%	6.22500	2.800000	5.100000	1.80000
50%	6.50000	3.000000	5.550000	2.00000
75%	6.90000	3.175000	5.875000	2.30000
max	7.90000	3.800000	6.900000	2.50000

```
In [ ]:
```