

statistics

day 7 - kaviyadevi

importing libraries

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

importing dataset

```
In [2]: data=pd.read_csv(r"C:\Users\user\Downloads\3_Fitness-1 - 3_Fitness-1.csv")
data
```

```
Out[2]:
```

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179
5	F	8.15%	16.24%	18.47%	167
6	G	18.54%	8.76%	17.49%	171
7	H	25.56%	5.93%	13.79%	170
8	Grand Total	100.00%	100.00%	100.00%	1150

importing numeric values from data set

```
In [3]: df=data[["Sum of Jan","Sum of Feb","Sum of Mar","Sum of Total Sales"]]
df
```

```
Out[3]:
```

	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	5.62%	7.73%	6.16%	75
1	4.21%	17.27%	19.21%	160
2	9.83%	11.60%	5.17%	101
3	2.81%	21.91%	7.88%	127
4	25.28%	10.57%	11.82%	179
5	8.15%	16.24%	18.47%	167
6	18.54%	8.76%	17.49%	171
7	25.56%	5.93%	13.79%	170
8	100.00%	100.00%	100.00%	1150

a. Find mean,median,mode and describe

```
In [4]: print(df.mode())
```

	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	100.00%	10.57%	100.00%	75
1	18.54%	100.00%	11.82%	101
2	2.81%	11.60%	13.79%	127
3	25.28%	16.24%	17.49%	160
4	25.56%	17.27%	18.47%	167
5	4.21%	21.91%	19.21%	170
6	5.62%	5.93%	5.17%	171
7	8.15%	7.73%	6.16%	179
8	9.83%	8.76%	7.88%	1150

```
In [5]: print(df.mean())
```

```
Sum of Total Sales    255.555556
dtype: float64
```

```
In [6]: print(df.median())
```

```
Sum of Total Sales    167.0
dtype: float64
```

In [7]: `print(df.describe())`

```

                Sum of Total Sales
count          9.000000
mean         255.555556
std          337.332963
min           75.000000
25%          127.000000
50%          167.000000
75%          171.000000
max          1150.000000

```

b.) Find sum,cumsum,count,min and max values

In [8]: `print(df.sum())`

```

Sum of Jan          5.62%4.21%9.83%2.81%25.28%8.15%18.54%25.56%100...
Sum of Feb          7.73%17.27%11.60%21.91%10.57%16.24%8.76%5.93%1...
Sum of Mar          6.16%19.21%5.17%7.88%11.82%18.47%17.49%13.79%1...
Sum of Total Sales                                     2300
dtype: object

```

In [9]: `print(df.cumsum())`

```

                                Sum of Jan  \
0                                5.62%
1                                5.62%4.21%
2                                5.62%4.21%9.83%
3                                5.62%4.21%9.83%2.81%
4                                5.62%4.21%9.83%2.81%25.28%
5                                5.62%4.21%9.83%2.81%25.28%8.15%
6                                5.62%4.21%9.83%2.81%25.28%8.15%18.54%
7                                5.62%4.21%9.83%2.81%25.28%8.15%18.54%25.56%
8    5.62%4.21%9.83%2.81%25.28%8.15%18.54%25.56%100...

                                Sum of Feb  \
0                                7.73%
1                                7.73%17.27%
2                                7.73%17.27%11.60%
3                                7.73%17.27%11.60%21.91%
4                                7.73%17.27%11.60%21.91%10.57%
5                                7.73%17.27%11.60%21.91%10.57%16.24%
6                                7.73%17.27%11.60%21.91%10.57%16.24%8.76%
7                                7.73%17.27%11.60%21.91%10.57%16.24%8.76%5.93%
8    7.73%17.27%11.60%21.91%10.57%16.24%8.76%5.93%1...

                                Sum of Mar  Sum of Total Sales
0                                6.16%           75
1                                6.16%19.21%          235
2                                6.16%19.21%5.17%          336
3                                6.16%19.21%5.17%7.88%          463
4                                6.16%19.21%5.17%7.88%11.82%          642
5                                6.16%19.21%5.17%7.88%11.82%18.47%          809
6                                6.16%19.21%5.17%7.88%11.82%18.47%17.49%          980
7                                6.16%19.21%5.17%7.88%11.82%18.47%17.49%13.79%         1150
8    6.16%19.21%5.17%7.88%11.82%18.47%17.49%13.79%1...         2300

```

In [10]: `print(df.count())`

```

Sum of Jan          9
Sum of Feb          9
Sum of Mar          9
Sum of Total Sales  9
dtype: int64

```

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

```

Sum of Jan          100.00%
Sum of Feb           10.57%
Sum of Mar          100.00%
Sum of Total Sales    75
dtype: object

```

```
In [12]: print(df.max())
```

```
Sum of Jan      9.83%
Sum of Feb      8.76%
Sum of Mar      7.88%
Sum of Total Sales    1150
dtype: object
```

c.) Find covariance and correlation

covariance

```
In [13]: df.cov()
```

```
Out[13]:
```

	Sum of Total Sales
Sum of Total Sales	113793.527778

corelation

```
In [18]: from scipy.stats import pearsonr
print(pearsonr(data["Sum of Total Sales"],data["Sum of Total Sales"]))

(1.0, 0.0)
```

```
In [20]: from scipy.stats import spearmanr
print(spearmanr(data["Sum of Total Sales"],data["Sum of Total Sales"]))

SpearmanrResult(correlation=1.0, pvalue=0.0)
```