Heamnath N

20104028

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

Importing csv

In [2]:
```python
df=pd.read_csv("8_BreastCancerPrediction.csv")
df
```

Out[2]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
0	842302	M	17.99	10.38	122.80	1001.0	0.11840
1	842517	M	20.57	17.77	132.90	1326.0	0.08474
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960
3	84348301	M	11.42	20.38	77.58	386.1	0.14250
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030
...	...	...	...	...	...	...	...
564	926424	M	21.56	22.39	142.00	1479.0	0.11100
565	926682	M	20.13	28.25	131.20	1261.0	0.09780
566	926954	M	16.60	28.08	108.30	858.1	0.08455
567	927241	M	20.60	29.33	140.10	1265.0	0.11780
568	92751	B	7.76	24.54	47.92	181.0	0.05263

569 rows × 33 columns

Sum

In [3]:
```python
df.sum()
```

Out[3]:
```
id                    17281572085
diagnosis             MMMMMMMMMMMMMMMMMMBBBMMMMMMMMMMMMMMMMBMMMMMMMMM...
radius_mean           8038.429
texture_mean          10975.81
perimeter_mean        52330.38
area_mean             372631.9
smoothness_mean       54.829
```

```
compactness_mean                                                    59.37002
concavity_mean                                                     50.526811
concave points_mean                                               27.834994
symmetry_mean                                                      103.0811
fractal_dimension_mean                                            35.73184
radius_se                                                         230.5429
texture_se                                                        692.3896
perimeter_se                                                     1630.7877
area_se                                                         22951.798
smoothness_se                                                      4.006317
compactness_se                                                    14.497061
concavity_se                                                      18.147525
concave points_se                                                 6.712002
symmetry_se                                                       11.688568
fractal_dimension_se                                               2.1593
radius_worst                                                     9257.169
texture_worst                                                   14610.34
perimeter_worst                                                 61031.63
area_worst                                                     501051.8
smoothness_worst                                                 75.31773
compactness_worst                                               144.67681
concavity_worst                                                 154.875247
concave points_worst                                            65.210941
symmetry_worst                                                  165.053
fractal_dimension_worst                                         47.76517
Unnamed: 32                                                      0.0
dtype: object
```

Mean

In [4]:
```python
df.mean()
```

Out[4]:
```
id                                 3.037183e+07
radius_mean                        1.412729e+01
texture_mean                       1.928965e+01
perimeter_mean                     9.196903e+01
area_mean                          6.548891e+02
smoothness_mean                    9.636028e-02
compactness_mean                   1.043410e-01
concavity_mean                     8.879932e-02
concave points_mean                4.891915e-02
symmetry_mean                      1.811619e-01
fractal_dimension_mean             6.279761e-02
radius_se                          4.051721e-01
texture_se                         1.216853e+00
perimeter_se                       2.866059e+00
area_se                            4.033708e+01
smoothness_se                      7.040979e-03
compactness_se                     2.547814e-02
concavity_se                       3.189372e-02
concave points_se                  1.179614e-02
symmetry_se                        2.054230e-02
fractal_dimension_se               3.794904e-03
radius_worst                       1.626919e+01
texture_worst                      2.567722e+01
perimeter_worst                    1.072612e+02
area_worst                         8.805831e+02
smoothness_worst                   1.323686e-01
compactness_worst                  2.542650e-01
concavity_worst                    2.721885e-01
concave points_worst               1.146062e-01
symmetry_worst                     2.900756e-01
```

```
fractal_dimension_worst        8.394582e-02
Unnamed: 32                             NaN
dtype: float64
```

Median

In [5]:
```
df.median()
```

Out[5]:
```
id                         906024.000000
radius_mean                    13.370000
texture_mean                   18.840000
perimeter_mean                 86.240000
area_mean                     551.100000
smoothness_mean                 0.095870
compactness_mean                0.092630
concavity_mean                  0.061540
concave points_mean             0.033500
symmetry_mean                   0.179200
fractal_dimension_mean          0.061540
radius_se                       0.324200
texture_se                      1.108000
perimeter_se                    2.287000
area_se                        24.530000
smoothness_se                   0.006380
compactness_se                  0.020450
concavity_se                    0.025890
concave points_se               0.010930
symmetry_se                     0.018730
fractal_dimension_se            0.003187
radius_worst                   14.970000
texture_worst                  25.410000
perimeter_worst                97.660000
area_worst                    686.500000
smoothness_worst                0.131300
compactness_worst               0.211900
concavity_worst                 0.226700
concave points_worst            0.099930
symmetry_worst                  0.282200
fractal_dimension_worst         0.080040
Unnamed: 32                           NaN
dtype: float64
```

Mode

In [6]:
```
df.mode()
```

Out[6]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
0	8670	B	12.34	14.93	82.61	512.2	0.1007
1	8913	NaN	NaN	15.70	87.76	NaN	NaN
2	8915	NaN	NaN	16.84	134.70	NaN	NaN
3	9047	NaN	NaN	16.85	NaN	NaN	NaN
4	85715	NaN	NaN	17.46	NaN	NaN	NaN

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
...	...	...	...	...	...	...	...
564	911157302	NaN	NaN	NaN	NaN	NaN	NaN
565	911296201	NaN	NaN	NaN	NaN	NaN	NaN
566	911296202	NaN	NaN	NaN	NaN	NaN	NaN
567	911320501	NaN	NaN	NaN	NaN	NaN	NaN
568	911320502	NaN	NaN	NaN	NaN	NaN	NaN

569 rows × 33 columns

Describe

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

Out[7]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	comp
count	5.690000e+02	569.000000	569.000000	569.000000	569.000000	569.000000	
mean	3.037183e+07	14.127292	19.289649	91.969033	654.889104	0.096360	
std	1.250206e+08	3.524049	4.301036	24.298981	351.914129	0.014064	
min	8.670000e+03	6.981000	9.710000	43.790000	143.500000	0.052630	
25%	8.692180e+05	11.700000	16.170000	75.170000	420.300000	0.086370	
50%	9.060240e+05	13.370000	18.840000	86.240000	551.100000	0.095870	
75%	8.813129e+06	15.780000	21.800000	104.100000	782.700000	0.105300	
max	9.113205e+08	28.110000	39.280000	188.500000	2501.000000	0.163400	

8 rows × 32 columns
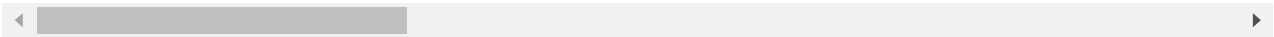
CumSum

In [8]:
```python
df.cumsum()
```

Out[8]:

	id	diagnosis	ra
0	842302	M	
1	1684819	MM	
2	85985722	MMM	

	id	diagnosis	ra
3	170334023	MMMM	
4	254692425	MMMMM	
...	...		...
564	17278698457	MMMMMMMMMMMMMMMMMMMMBBBMMMMMMMMMMMMMMMMBMMMMMMMM...	
565	17279625139	MMMMMMMMMMMMMMMMMMMMBBBMMMMMMMMMMMMMMMMBMMMMMMMM...	
566	17280552093	MMMMMMMMMMMMMMMMMMMMBBBMMMMMMMMMMMMMMMMBMMMMMMMM...	
567	17281479334	MMMMMMMMMMMMMMMMMMMMBBBMMMMMMMMMMMMMMMMBMMMMMMMM...	
568	17281572085	MMMMMMMMMMMMMMMMMMMMBBBMMMMMMMMMMMMMMMMBMMMMMMMM...	

569 rows × 33 columns

Count

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

```
Out[9]:  id                        569
         diagnosis                 569
         radius_mean               569
         texture_mean              569
         perimeter_mean            569
         area_mean                 569
         smoothness_mean           569
         compactness_mean          569
         concavity_mean            569
         concave points_mean       569
         symmetry_mean             569
         fractal_dimension_mean    569
         radius_se                 569
         texture_se                569
         perimeter_se              569
         area_se                   569
         smoothness_se             569
         compactness_se            569
         concavity_se              569
         concave points_se         569
         symmetry_se               569
         fractal_dimension_se      569
         radius_worst              569
         texture_worst             569
         perimeter_worst           569
         area_worst                569
         smoothness_worst          569
         compactness_worst         569
         concavity_worst           569
         concave points_worst      569
         symmetry_worst            569
         fractal_dimension_worst   569
         Unnamed: 32                 0
         dtype: int64
```

Min

```
In [10]:   df.min()
```

```
Out[10]:   id                          8670
           diagnosis                      B
           radius_mean                6.981
           texture_mean                9.71
           perimeter_mean             43.79
           area_mean                  143.5
           smoothness_mean          0.05263
           compactness_mean         0.01938
           concavity_mean               0.0
           concave points_mean          0.0
           symmetry_mean              0.106
           fractal_dimension_mean   0.04996
           radius_se                 0.1115
           texture_se                0.3602
           perimeter_se               0.757
           area_se                    6.802
           smoothness_se           0.001713
           compactness_se          0.002252
           concavity_se                 0.0
           concave points_se            0.0
           symmetry_se             0.007882
           fractal_dimension_se    0.000895
           radius_worst                7.93
           texture_worst              12.02
           perimeter_worst            50.41
           area_worst                 185.2
           smoothness_worst         0.07117
           compactness_worst        0.02729
           concavity_worst              0.0
           concave points_worst         0.0
           symmetry_worst            0.1565
           fractal_dimension_worst  0.05504
           Unnamed: 32                  NaN
           dtype: object
```

Max

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

```
Out[11]:   id                     911320502
           diagnosis                      M
           radius_mean                28.11
           texture_mean               39.28
           perimeter_mean             188.5
           area_mean                 2501.0
           smoothness_mean           0.1634
           compactness_mean          0.3454
           concavity_mean            0.4268
           concave points_mean       0.2012
           symmetry_mean              0.304
           fractal_dimension_mean   0.09744
           radius_se                  2.873
           texture_se                 4.885
           perimeter_se               21.98
```

```
area_se                         542.2
smoothness_se                  0.03113
compactness_se                  0.1354
concavity_se                     0.396
concave points_se              0.05279
symmetry_se                    0.07895
fractal_dimension_se           0.02984
radius_worst                     36.04
texture_worst                    49.54
perimeter_worst                  251.2
area_worst                      4254.0
smoothness_worst                0.2226
compactness_worst                1.058
concavity_worst                  1.252
concave points_worst             0.291
symmetry_worst                  0.6638
fractal_dimension_worst         0.2075
Unnamed: 32                        NaN
dtype: object
```

Covariance

In [12]:
```python
from numpy import cov
```

In [13]:
```python
cov(df['radius_mean'],df['texture_mean'])
```

Out[13]:
```
array([[12.41892013,  4.90758156],
       [ 4.90758156, 18.49890868]])
```

Pearson

In [14]:
```python
from scipy.stats import pearsonr
pearsonr(df['radius_mean'],df['texture_mean'])
```

Out[14]: (0.323781890927733, 2.360374375922593e-15)

Spearman

In [15]:
```python
from scipy.stats import spearmanr
spearmanr(df['radius_mean'],df['texture_mean'])
```

Out[15]: SpearmanrResult(correlation=0.3409562685372812, pvalue=5.900189597213798e-17)