Project: Predicing Breast-Cancer using Linear Regression

Problem Statement:

The goal is to predict whether a breast cancer diagnosis is benign or malignant based on the given attributes in the dataset etc:using Linear Regression.

Data Collection: Use the Kaggle Breast-Cancer dataset, Download it from Kaggle.

https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset

(Data Wrangling and Cleaning)

```
import pandas as pd
data = pd.read_csv(r"C:\Users\Kruthi Mule\Downloads\archive\breast-
cancer.csv")

data
```

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

     smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
0            0.11840           0.27760         0.30010
```

```
0.14710
1           0.08474           0.07864         0.08690
0.07017
2           0.10960           0.15990         0.19740
0.12790
3           0.14250           0.28390         0.24140
0.10520
4           0.10030           0.13280         0.19800
0.10430
..              ...               ...             ...
...
564         0.11100           0.11590         0.24390
0.13890
565         0.09780           0.10340         0.14400
0.09791
566         0.08455           0.10230         0.09251
0.05302
567         0.11780           0.27700         0.35140
0.15200
568         0.05263           0.04362         0.00000
0.00000

     ...  radius_worst  texture_worst  perimeter_worst  area_worst  \
0    ...        25.380          17.33           184.60      2019.0
1    ...        24.990          23.41           158.80      1956.0
2    ...        23.570          25.53           152.50      1709.0
3    ...        14.910          26.50            98.87       567.7
4    ...        22.540          16.67           152.20      1575.0
..   ...           ...            ...              ...         ...
564  ...        25.450          26.40           166.10      2027.0
565  ...        23.690          38.25           155.00      1731.0
566  ...        18.980          34.12           126.70      1124.0
567  ...        25.740          39.42           184.60      1821.0
568  ...         9.456          30.37            59.16       268.6

     smoothness_worst  compactness_worst  concavity_worst  \
0             0.16220            0.66560           0.7119
1             0.12380            0.18660           0.2416
2             0.14440            0.42450           0.4504
3             0.20980            0.86630           0.6869
4             0.13740            0.20500           0.4000
..                ...                ...              ...
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     concave points_worst  symmetry_worst  fractal_dimension_worst
0                  0.2654          0.4601                  0.11890
```

```
1                  0.1860          0.2750                  0.08902
2                  0.2430          0.3613                  0.08758
3                  0.2575          0.6638                  0.17300
4                  0.1625          0.2364                  0.07678
..                    ...             ...                      ...
564                0.2216          0.2060                  0.07115
565                0.1628          0.2572                  0.06637
566                0.1418          0.2218                  0.07820
567                0.2650          0.4087                  0.12400
568                0.0000          0.2871                  0.07039

[569 rows x 32 columns]
```

```
data.head()
```

```
         id diagnosis  radius_mean  texture_mean  perimeter_mean
area_mean  \
0    842302         M        17.99         10.38          122.80
1001.0
1    842517         M        20.57         17.77          132.90
1326.0
2  84300903         M        19.69         21.25          130.00
1203.0
3  84348301         M        11.42         20.38           77.58
386.1
4  84358402         M        20.29         14.34          135.10
1297.0

   smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
0          0.11840           0.27760          0.3001
0.14710
1          0.08474           0.07864          0.0869
0.07017
2          0.10960           0.15990          0.1974
0.12790
3          0.14250           0.28390          0.2414
0.10520
4          0.10030           0.13280          0.1980
0.10430

   ...  radius_worst  texture_worst  perimeter_worst  area_worst  \
0  ...         25.38          17.33           184.60      2019.0
1  ...         24.99          23.41           158.80      1956.0
2  ...         23.57          25.53           152.50      1709.0
3  ...         14.91          26.50            98.87       567.7
4  ...         22.54          16.67           152.20      1575.0

   smoothness_worst  compactness_worst  concavity_worst  concave
points_worst  \
```

```
0            0.1622             0.6656           0.7119
0.2654
1            0.1238             0.1866           0.2416
0.1860
2            0.1444             0.4245           0.4504
0.2430
3            0.2098             0.8663           0.6869
0.2575
4            0.1374             0.2050           0.4000
0.1625

   symmetry_worst  fractal_dimension_worst
0          0.4601                  0.11890
1          0.2750                  0.08902
2          0.3613                  0.08758
3          0.6638                  0.17300
4          0.2364                  0.07678

[5 rows x 32 columns]
```

```
data.tail()
```

```
         id diagnosis  radius_mean  texture_mean  perimeter_mean
area_mean  \
564  926424         M        21.56         22.39          142.00
1479.0
565  926682         M        20.13         28.25          131.20
1261.0
566  926954         M        16.60         28.08          108.30
858.1
567  927241         M        20.60         29.33          140.10
1265.0
568   92751         B         7.76         24.54           47.92
181.0

     smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
564          0.11100           0.11590         0.24390
0.13890
565          0.09780           0.10340         0.14400
0.09791
566          0.08455           0.10230         0.09251
0.05302
567          0.11780           0.27700         0.35140
0.15200
568          0.05263           0.04362         0.00000
0.00000

     ...  radius_worst  texture_worst  perimeter_worst  area_worst  \
564  ...        25.450          26.40           166.10      2027.0
```

```
565  ...        23.690          38.25           155.00      1731.0
566  ...        18.980          34.12           126.70      1124.0
567  ...        25.740          39.42           184.60      1821.0
568  ...         9.456          30.37            59.16       268.6

     smoothness_worst  compactness_worst  concavity_worst  \
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     concave points_worst  symmetry_worst  fractal_dimension_worst
564                0.2216          0.2060                  0.07115
565                0.1628          0.2572                  0.06637
566                0.1418          0.2218                  0.07820
567                0.2650          0.4087                  0.12400
568                0.0000          0.2871                  0.07039

[5 rows x 32 columns]
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 32 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   id                       569 non-null    int64
 1   diagnosis                569 non-null    object
 2   radius_mean              569 non-null    float64
 3   texture_mean             569 non-null    float64
 4   perimeter_mean           569 non-null    float64
 5   area_mean                569 non-null    float64
 6   smoothness_mean          569 non-null    float64
 7   compactness_mean         569 non-null    float64
 8   concavity_mean           569 non-null    float64
 9   concave points_mean      569 non-null    float64
 10  symmetry_mean            569 non-null    float64
 11  fractal_dimension_mean   569 non-null    float64
 12  radius_se                569 non-null    float64
 13  texture_se               569 non-null    float64
 14  perimeter_se             569 non-null    float64
 15  area_se                  569 non-null    float64
 16  smoothness_se            569 non-null    float64
 17  compactness_se           569 non-null    float64
 18  concavity_se             569 non-null    float64
 19  concave points_se        569 non-null    float64
 20  symmetry_se              569 non-null    float64
 21  fractal_dimension_se     569 non-null    float64
```

```
 22  radius_worst             569 non-null    float64
 23  texture_worst            569 non-null    float64
 24  perimeter_worst          569 non-null    float64
 25  area_worst               569 non-null    float64
 26  smoothness_worst         569 non-null    float64
 27  compactness_worst        569 non-null    float64
 28  concavity_worst          569 non-null    float64
 29  concave points_worst     569 non-null    float64
 30  symmetry_worst           569 non-null    float64
 31  fractal_dimension_worst  569 non-null    float64
dtypes: float64(30), int64(1), object(1)
memory usage: 142.4+ KB

data.describe()

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

       smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
count       569.000000        569.000000      569.000000
569.000000
mean          0.096360          0.104341        0.088799
0.048919
std           0.014064          0.052813        0.079720
0.038803
min           0.052630          0.019380        0.000000
0.000000
25%           0.086370          0.064920        0.029560
0.020310
50%           0.095870          0.092630        0.061540
0.033500
75%           0.105300          0.130400        0.130700
0.074000
```

```
max           0.163400          0.345400        0.426800
0.201200

       symmetry_mean  ...  radius_worst  texture_worst
perimeter_worst  \
count     569.000000  ...    569.000000     569.000000
569.000000
mean        0.181162  ...     16.269190      25.677223
107.261213
std         0.027414  ...      4.833242       6.146258
33.602542
min         0.106000  ...      7.930000      12.020000
50.410000
25%         0.161900  ...     13.010000      21.080000
84.110000
50%         0.179200  ...     14.970000      25.410000
97.660000
75%         0.195700  ...     18.790000      29.720000
125.400000
max         0.304000  ...     36.040000      49.540000
251.200000

        area_worst  smoothness_worst  compactness_worst
concavity_worst  \
count   569.000000        569.000000         569.000000
569.000000
mean    880.583128          0.132369           0.254265
0.272188
std     569.356993          0.022832           0.157336
0.208624
min     185.200000          0.071170           0.027290
0.000000
25%     515.300000          0.116600           0.147200
0.114500
50%     686.500000          0.131300           0.211900
0.226700
75%    1084.000000          0.146000           0.339100
0.382900
max    4254.000000          0.222600           1.058000
1.252000

       concave points_worst  symmetry_worst  fractal_dimension_worst
count            569.000000      569.000000               569.000000
mean               0.114606        0.290076                 0.083946
std                0.065732        0.061867                 0.018061
min                0.000000        0.156500                 0.055040
25%                0.064930        0.250400                 0.071460
50%                0.099930        0.282200                 0.080040
75%                0.161400        0.317900                 0.092080
max                0.291000        0.663800                 0.207500
```

```
[8 rows x 31 columns]
```

```
data.ndim
```

```
2
```

```
data.columns
```

```
Index(['id', 'diagnosis', 'radius_mean', 'texture_mean',
'perimeter_mean',
       'area_mean', 'smoothness_mean', 'compactness_mean',
'concavity_mean',
       'concave points_mean', 'symmetry_mean',
'fractal_dimension_mean',
       'radius_se', 'texture_se', 'perimeter_se', 'area_se',
'smoothness_se',
       'compactness_se', 'concavity_se', 'concave points_se',
'symmetry_se',
       'fractal_dimension_se', 'radius_worst', 'texture_worst',
       'perimeter_worst', 'area_worst', 'smoothness_worst',
       'compactness_worst', 'concavity_worst', 'concave points_worst',
       'symmetry_worst', 'fractal_dimension_worst'],
      dtype='object')
```

Handling Missing Data

```
data.isnull()
```

```
        id  diagnosis  radius_mean  texture_mean  perimeter_mean
area_mean  \
0    False      False        False         False           False
False
1    False      False        False         False           False
False
2    False      False        False         False           False
False
3    False      False        False         False           False
False
4    False      False        False         False           False
False
..     ...        ...          ...           ...             ...
...
564  False      False        False         False           False
False
565  False      False        False         False           False
False
566  False      False        False         False           False
False
567  False      False        False         False           False
```

```
False
568  False      False        False         False           False
False

     smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
0              False             False           False
False
1              False             False           False
False
2              False             False           False
False
3              False             False           False
False
4              False             False           False
False
..               ...               ...             ...
...
564            False             False           False
False
565            False             False           False
False
566            False             False           False
False
567            False             False           False
False
568            False             False           False
False

     ...  radius_worst  texture_worst  perimeter_worst  area_worst  \
0    ...         False          False            False       False
1    ...         False          False            False       False
2    ...         False          False            False       False
3    ...         False          False            False       False
4    ...         False          False            False       False
..   ...           ...            ...              ...         ...
564  ...         False          False            False       False
565  ...         False          False            False       False
566  ...         False          False            False       False
567  ...         False          False            False       False
568  ...         False          False            False       False

     smoothness_worst  compactness_worst  concavity_worst  \
0               False              False            False
1               False              False            False
2               False              False            False
3               False              False            False
4               False              False            False
..                ...                ...              ...
564             False              False            False
```

```
565              False                False             False
566              False                False             False
567              False                False             False
568              False                False             False
```

```
     concave points_worst  symmetry_worst  fractal_dimension_worst
0                   False           False                    False
1                   False           False                    False
2                   False           False                    False
3                   False           False                    False
4                   False           False                    False
..                    ...             ...                      ...
564                 False           False                    False
565                 False           False                    False
566                 False           False                    False
567                 False           False                    False
568                 False           False                    False

[569 rows x 32 columns]
```

```
data.isnull().sum()
```

```
id                         0
diagnosis                  0
radius_mean                0
texture_mean               0
perimeter_mean             0
area_mean                  0
smoothness_mean            0
compactness_mean           0
concavity_mean             0
concave points_mean        0
symmetry_mean              0
fractal_dimension_mean     0
radius_se                  0
texture_se                 0
perimeter_se               0
area_se                    0
smoothness_se              0
compactness_se             0
concavity_se               0
concave points_se          0
symmetry_se                0
fractal_dimension_se       0
radius_worst               0
texture_worst              0
perimeter_worst            0
area_worst                 0
smoothness_worst           0
compactness_worst          0
```

```
concavity_worst            0
concave points_worst       0
symmetry_worst             0
fractal_dimension_worst    0
dtype: int64
```

```
data.duplicated()
```

```
0      False
1      False
2      False
3      False
4      False
       ...
564    False
565    False
566    False
567    False
568    False
Length: 569, dtype: bool
```

```
data.duplicated().sum()
```

```
0
```

```
data.fillna(data.mean(), inplace=True)
```

```
data
```

```
           id diagnosis  radius_mean  texture_mean  perimeter_mean
area_mean  \
0      842302         M        17.99         10.38          122.80
1001.0
1      842517         M        20.57         17.77          132.90
1326.0
2    84300903         M        19.69         21.25          130.00
1203.0
3    84348301         M        11.42         20.38           77.58
386.1
4    84358402         M        20.29         14.34          135.10
1297.0
..        ...       ...          ...           ...             ...
...
564    926424         M        21.56         22.39          142.00
1479.0
565    926682         M        20.13         28.25          131.20
1261.0
566    926954         M        16.60         28.08          108.30
858.1
567    927241         M        20.60         29.33          140.10
1265.0
```

```
568     92751         B         7.76         24.54           47.92
181.0

     smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
0            0.11840           0.27760         0.30010
0.14710
1            0.08474           0.07864         0.08690
0.07017
2            0.10960           0.15990         0.19740
0.12790
3            0.14250           0.28390         0.24140
0.10520
4            0.10030           0.13280         0.19800
0.10430
..               ...               ...             ...
...
564          0.11100           0.11590         0.24390
0.13890
565          0.09780           0.10340         0.14400
0.09791
566          0.08455           0.10230         0.09251
0.05302
567          0.11780           0.27700         0.35140
0.15200
568          0.05263           0.04362         0.00000
0.00000

     ...  radius_worst  texture_worst  perimeter_worst  area_worst  \
0    ...        25.380          17.33           184.60      2019.0
1    ...        24.990          23.41           158.80      1956.0
2    ...        23.570          25.53           152.50      1709.0
3    ...        14.910          26.50            98.87       567.7
4    ...        22.540          16.67           152.20      1575.0
..   ...           ...            ...              ...         ...
564  ...        25.450          26.40           166.10      2027.0
565  ...        23.690          38.25           155.00      1731.0
566  ...        18.980          34.12           126.70      1124.0
567  ...        25.740          39.42           184.60      1821.0
568  ...         9.456          30.37            59.16       268.6

     smoothness_worst  compactness_worst  concavity_worst  \
0             0.16220            0.66560           0.7119
1             0.12380            0.18660           0.2416
2             0.14440            0.42450           0.4504
3             0.20980            0.86630           0.6869
4             0.13740            0.20500           0.4000
..                ...                ...              ...
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
```

```
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     concave points_worst  symmetry_worst  fractal_dimension_worst
0                  0.2654          0.4601                  0.11890
1                  0.1860          0.2750                  0.08902
2                  0.2430          0.3613                  0.08758
3                  0.2575          0.6638                  0.17300
4                  0.1625          0.2364                  0.07678
..                    ...             ...                      ...
564                0.2216          0.2060                  0.07115
565                0.1628          0.2572                  0.06637
566                0.1418          0.2218                  0.07820
567                0.2650          0.4087                  0.12400
568                0.0000          0.2871                  0.07039

[569 rows x 32 columns]
```

Feature Engineering

```
data = pd.get_dummies(data, drop_first=True)

data
```

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

     smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
0            0.11840           0.27760         0.30010
0.14710
1            0.08474           0.07864         0.08690
0.07017
2            0.10960           0.15990         0.19740
0.12790
3            0.14250           0.28390         0.24140
0.10520
4            0.10030           0.13280         0.19800
0.10430
```

```
..               ...               ...             ...
...
564          0.11100           0.11590         0.24390
0.13890
565          0.09780           0.10340         0.14400
0.09791
566          0.08455           0.10230         0.09251
0.05302
567          0.11780           0.27700         0.35140
0.15200
568          0.05263           0.04362         0.00000
0.00000

     symmetry_mean  ...  texture_worst  perimeter_worst  area_worst  \
0           0.2419  ...          17.33           184.60      2019.0
1           0.1812  ...          23.41           158.80      1956.0
2           0.2069  ...          25.53           152.50      1709.0
3           0.2597  ...          26.50            98.87       567.7
4           0.1809  ...          16.67           152.20      1575.0
..             ...  ...            ...              ...         ...
564         0.1726  ...          26.40           166.10      2027.0
565         0.1752  ...          38.25           155.00      1731.0
566         0.1590  ...          34.12           126.70      1124.0
567         0.2397  ...          39.42           184.60      1821.0
568         0.1587  ...          30.37            59.16       268.6

     smoothness_worst  compactness_worst  concavity_worst  \
0             0.16220            0.66560           0.7119
1             0.12380            0.18660           0.2416
2             0.14440            0.42450           0.4504
3             0.20980            0.86630           0.6869
4             0.13740            0.20500           0.4000
..                ...                ...              ...
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     concave points_worst  symmetry_worst  fractal_dimension_worst  \
0                  0.2654          0.4601                  0.11890
1                  0.1860          0.2750                  0.08902
2                  0.2430          0.3613                  0.08758
3                  0.2575          0.6638                  0.17300
4                  0.1625          0.2364                  0.07678
..                    ...             ...                      ...
564                0.2216          0.2060                  0.07115
565                0.1628          0.2572                  0.06637
566                0.1418          0.2218                  0.07820
567                0.2650          0.4087                  0.12400
```

```
568                0.0000          0.2871                  0.07039

     diagnosis_M
0           True
1           True
2           True
3           True
4           True
..           ...
564         True
565         True
566         True
567         True
568        False

[569 rows x 32 columns]
```

```
data.describe()
```

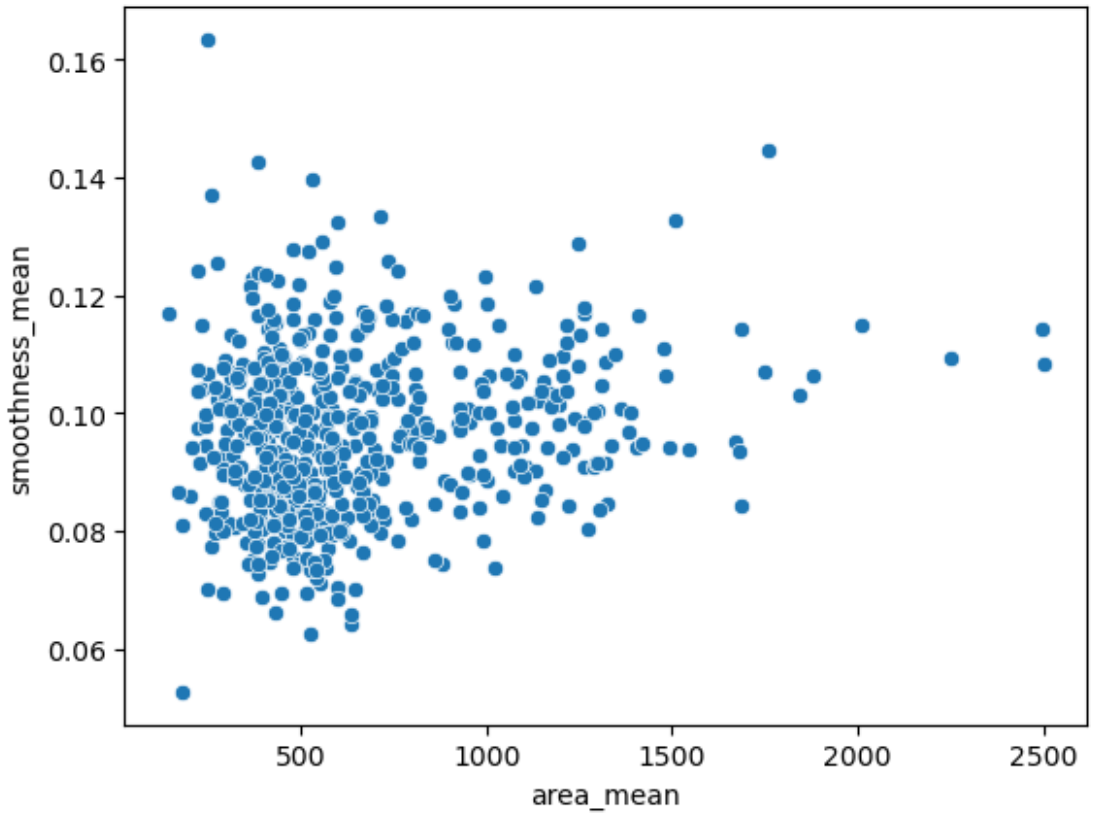
```
                 id  radius_mean  texture_mean  perimeter_mean
area_mean  \
count  5.690000e+02   569.000000    569.000000      569.000000
569.000000
mean   3.037183e+07    14.127292     19.289649       91.969033
654.889104
std    1.250206e+08     3.524049      4.301036       24.298981
351.914129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

       smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
count       569.000000        569.000000      569.000000
569.000000
mean          0.096360          0.104341        0.088799
0.048919
std           0.014064          0.052813        0.079720
0.038803
min           0.052630          0.019380        0.000000
0.000000
25%           0.086370          0.064920        0.029560
0.020310
```

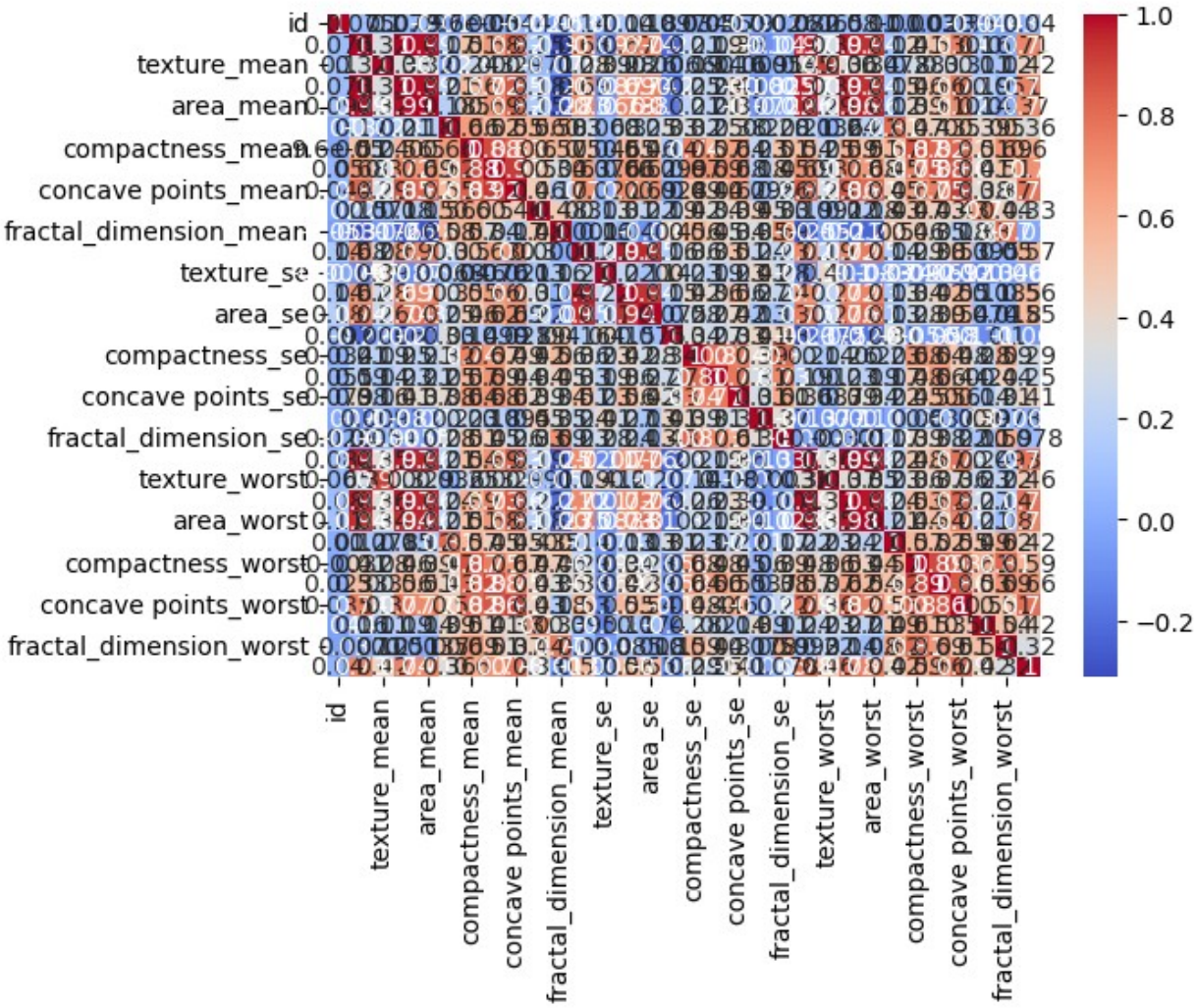
```
50%           0.095870          0.092630        0.061540
0.033500
75%           0.105300          0.130400        0.130700
0.074000
max           0.163400          0.345400        0.426800
0.201200

       symmetry_mean  ...  radius_worst  texture_worst
perimeter_worst  \
count     569.000000  ...    569.000000     569.000000
569.000000
mean        0.181162  ...     16.269190      25.677223
107.261213
std         0.027414  ...      4.833242       6.146258
33.602542
min         0.106000  ...      7.930000      12.020000
50.410000
25%         0.161900  ...     13.010000      21.080000
84.110000
50%         0.179200  ...     14.970000      25.410000
97.660000
75%         0.195700  ...     18.790000      29.720000
125.400000
max         0.304000  ...     36.040000      49.540000
251.200000

        area_worst  smoothness_worst  compactness_worst
concavity_worst  \
count   569.000000        569.000000         569.000000
569.000000
mean    880.583128          0.132369           0.254265
0.272188
std     569.356993          0.022832           0.157336
0.208624
min     185.200000          0.071170           0.027290
0.000000
25%     515.300000          0.116600           0.147200
0.114500
50%     686.500000          0.131300           0.211900
0.226700
75%    1084.000000          0.146000           0.339100
0.382900
max    4254.000000          0.222600           1.058000
1.252000

       concave points_worst  symmetry_worst  fractal_dimension_worst
count            569.000000      569.000000               569.000000
mean               0.114606        0.290076                 0.083946
std                0.065732        0.061867                 0.018061
min                0.000000        0.156500                 0.055040
```

```
25%                0.064930         0.250400                0.071460
50%                0.099930         0.282200                0.080040
75%                0.161400         0.317900                0.092080
max                0.291000         0.663800                0.207500

[8 rows x 31 columns]
```

Data Analysis and Visualization

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.scatterplot(x='area_mean', y='smoothness_mean', data=data)
plt.show()

sns.heatmap(data.corr(), annot=True, cmap='coolwarm')
plt.show()
```






```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 32 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   id                       569 non-null    int64  
 1   radius_mean              569 non-null    float64
 2   texture_mean             569 non-null    float64
 3   perimeter_mean           569 non-null    float64
 4   area_mean                569 non-null    float64
 5   smoothness_mean          569 non-null    float64
 6   compactness_mean         569 non-null    float64
 7   concavity_mean           569 non-null    float64
 8   concave points_mean      569 non-null    float64
 9   symmetry_mean            569 non-null    float64
 10  fractal_dimension_mean   569 non-null    float64
```

```
 11  radius_se                569 non-null    float64
 12  texture_se               569 non-null    float64
 13  perimeter_se             569 non-null    float64
 14  area_se                  569 non-null    float64
 15  smoothness_se            569 non-null    float64
 16  compactness_se           569 non-null    float64
 17  concavity_se             569 non-null    float64
 18  concave points_se        569 non-null    float64
 19  symmetry_se              569 non-null    float64
 20  fractal_dimension_se     569 non-null    float64
 21  radius_worst             569 non-null    float64
 22  texture_worst            569 non-null    float64
 23  perimeter_worst          569 non-null    float64
 24  area_worst               569 non-null    float64
 25  smoothness_worst         569 non-null    float64
 26  compactness_worst        569 non-null    float64
 27  concavity_worst          569 non-null    float64
 28  concave points_worst     569 non-null    float64
 29  symmetry_worst           569 non-null    float64
 30  fractal_dimension_worst  569 non-null    float64
 31  diagnosis_M              569 non-null    bool
dtypes: bool(1), float64(30), int64(1)
memory usage: 138.5 KB
```

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import OneHotEncoder, LabelEncoder

file_path = r"C:\Users\Kruthi Mule\Downloads\archive\breast-
cancer.csv"
data = pd.read_csv(file_path)

# Matrix
if len(data.select_dtypes(include=[np.number]).columns) > 1:
    matrix = data.select_dtypes(include=[np.number]).values
    matrix_transpose = np.transpose(matrix)
    matrix_product = np.dot(matrix, matrix_transpose)
    print("\nMatrix product:\n", matrix_product)
```

```
Matrix product:
 [[7.09477812e+11 7.09659089e+11 7.10068239e+13 ... 7.80778382e+11
  7.81021958e+11 7.81250972e+10]
 [7.09659089e+11 7.09840530e+11 7.10249489e+13 ... 7.80977880e+11
  7.81221602e+11 7.81450784e+10]
 [7.10068239e+13 7.10249489e+13 7.10664225e+15 ... 7.81430622e+13
  7.81672583e+13 7.81899375e+12]
 ...
 [7.80778382e+11 7.80977880e+11 7.81430622e+13 ... 8.59245751e+11
  8.59512932e+11 8.59763833e+10]
```

```
 [7.81021958e+11 7.81221602e+11 7.81672583e+13 ... 8.59512932e+11
  8.59780853e+11 8.60032697e+10]
 [7.81250972e+10 7.81450784e+10 7.81899375e+12 ... 8.59763833e+10
  8.60032697e+10 8.60286075e+09]]
```

```
# One-Hot Encoding
encoder = OneHotEncoder()
encoded_data =
encoder.fit_transform(data.select_dtypes(include=[object])).toarray()
print("\nOne-Hot Encoded data:\n", encoded_data)
```

```
One-Hot Encoded data:
 [[0. 1.]
 [0. 1.]
 [0. 1.]
 ...
 [0. 1.]
 [0. 1.]
 [1. 0.]]
```

```
# Label Encoding
label_encoder = LabelEncoder()
for column in data.select_dtypes(include=[object]).columns:
    data[column] = label_encoder.fit_transform(data[column])

print("\nLabel Encoded data:\n", data)
```

```
Label Encoded data:
            id  diagnosis  radius_mean  texture_mean
perimeter_mean  \
0      842302          1        17.99         10.38          122.80
1      842517          1        20.57         17.77          132.90
2    84300903          1        19.69         21.25          130.00
3    84348301          1        11.42         20.38           77.58
4    84358402          1        20.29         14.34          135.10
..        ...        ...          ...           ...             ...
564    926424          1        21.56         22.39          142.00
565    926682          1        20.13         28.25          131.20
566    926954          1        16.60         28.08          108.30
567    927241          1        20.60         29.33          140.10
568     92751          0         7.76         24.54           47.92

     area_mean  smoothness_mean  compactness_mean  concavity_mean  \
0       1001.0          0.11840           0.27760         0.30010
1       1326.0          0.08474           0.07864         0.08690
2       1203.0          0.10960           0.15990         0.19740
3        386.1          0.14250           0.28390         0.24140
4       1297.0          0.10030           0.13280         0.19800
..         ...              ...               ...             ...
```

```
564     1479.0          0.11100           0.11590         0.24390
565     1261.0          0.09780           0.10340         0.14400
566      858.1          0.08455           0.10230         0.09251
567     1265.0          0.11780           0.27700         0.35140
568      181.0          0.05263           0.04362         0.00000

     concave points_mean  ...  radius_worst  texture_worst
perimeter_worst  \
0                0.14710  ...        25.380          17.33
184.60
1                0.07017  ...        24.990          23.41
158.80
2                0.12790  ...        23.570          25.53
152.50
3                0.10520  ...        14.910          26.50
98.87
4                0.10430  ...        22.540          16.67
152.20
..                   ...  ...           ...            ...
...
564              0.13890  ...        25.450          26.40
166.10
565              0.09791  ...        23.690          38.25
155.00
566              0.05302  ...        18.980          34.12
126.70
567              0.15200  ...        25.740          39.42
184.60
568              0.00000  ...         9.456          30.37
59.16

     area_worst  smoothness_worst  compactness_worst  concavity_worst
\
0        2019.0           0.16220            0.66560           0.7119

1        1956.0           0.12380            0.18660           0.2416

2        1709.0           0.14440            0.42450           0.4504

3         567.7           0.20980            0.86630           0.6869

4        1575.0           0.13740            0.20500           0.4000

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

564      2027.0           0.14100            0.21130           0.4107

565      1731.0           0.11660            0.19220           0.3215

566      1124.0           0.11390            0.30940           0.3403
```

```
567      1821.0           0.16500            0.86810           0.9387

568       268.6           0.08996            0.06444           0.0000


     concave points_worst  symmetry_worst  fractal_dimension_worst
0                  0.2654          0.4601                  0.11890
1                  0.1860          0.2750                  0.08902
2                  0.2430          0.3613                  0.08758
3                  0.2575          0.6638                  0.17300
4                  0.1625          0.2364                  0.07678
..                    ...             ...                      ...
564                0.2216          0.2060                  0.07115
565                0.1628          0.2572                  0.06637
566                0.1418          0.2218                  0.07820
567                0.2650          0.4087                  0.12400
568                0.0000          0.2871                  0.07039

[569 rows x 32 columns]
```

```python
# mean
mean_values = data.mean()  # Mean calculation

# Display the results
print("Mean values:\n", mean_values)
```

```
Mean values:
 id                         3.037183e+07
diagnosis                  3.725835e-01
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
dtype: float64
```

```
# median
median_values = data.median()  # Median calculation

# Display the results
print("Median values:\n", median_values)
```

```
Median values:
 id                         906024.000000
diagnosis                       0.000000
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
dtype: float64

# mode
mode_values = data.mode().iloc[0]  # Mode can have multiple values;
picks the first row

# Display the results
print("Mode values:\n", mode_values)

Mode values:
 id                         8670.000000
diagnosis                     0.000000
radius_mean                  12.340000
texture_mean                 14.930000
perimeter_mean               82.610000
area_mean                   512.200000
smoothness_mean               0.100700
compactness_mean              0.114700
concavity_mean                0.000000
concave points_mean           0.000000
symmetry_mean                 0.160100
fractal_dimension_mean        0.056670
radius_se                     0.220400
texture_se                    0.856100
perimeter_se                  1.778000
area_se                      16.640000
smoothness_se                 0.005080
compactness_se                0.011040
concavity_se                  0.000000
concave points_se             0.000000
symmetry_se                   0.013440
fractal_dimension_se          0.001784
radius_worst                 12.360000
texture_worst                17.700000
perimeter_worst             101.700000
area_worst                  284.400000
smoothness_worst              0.121600
compactness_worst             0.148600
concavity_worst               0.000000
concave points_worst          0.000000
symmetry_worst                0.222600
fractal_dimension_worst       0.074270
Name: 0, dtype: float64
```

Linear Regression

```
from sklearn.linear_model import LinearRegression

x = data[['texture_mean', 'perimeter_mean', 'area_mean']]
```

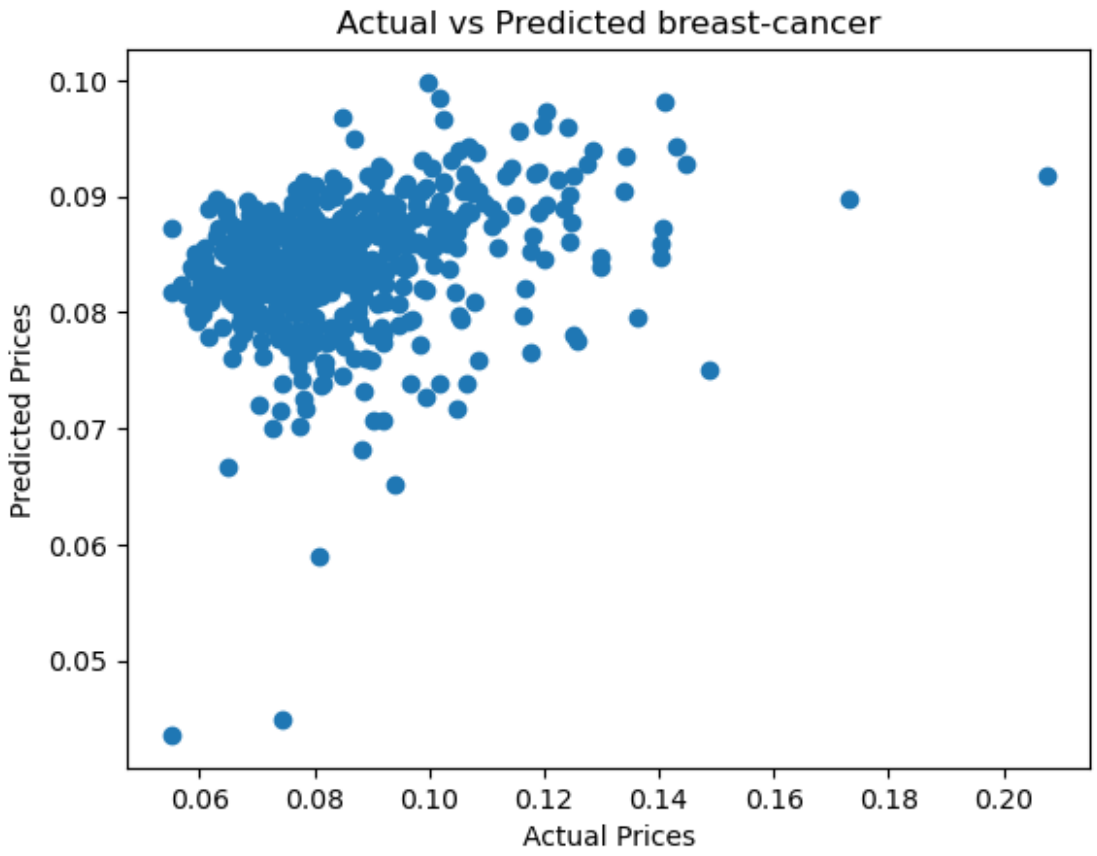
```
y = data['fractal_dimension_worst']

model = LinearRegression()

model.fit(x, y)

print(f'Intercept: {model.intercept_}')
print(f'Coefficients: {model.coef_}')
```

```
Intercept: 0.015935031786171824
Coefficients: [ 4.49897191e-04  1.27454699e-03 -8.83911909e-05]
```

```
predicted_prices = model.predict(x)

comparison = pd.DataFrame({'Actual': y, 'Predicted':
predicted_prices})
print(comparison.head())
```

```
    Actual  Predicted
0  0.11890   0.088640
1  0.08902   0.076110
2  0.08758   0.084852
3  0.17300   0.089855
4  0.07678   0.079934
```

```
plt.scatter(y, predicted_prices)
plt.xlabel('Actual Prices')
plt.ylabel('Predicted Prices')
plt.title('Actual vs Predicted breast-cancer')
plt.show()
```



Data Visulization

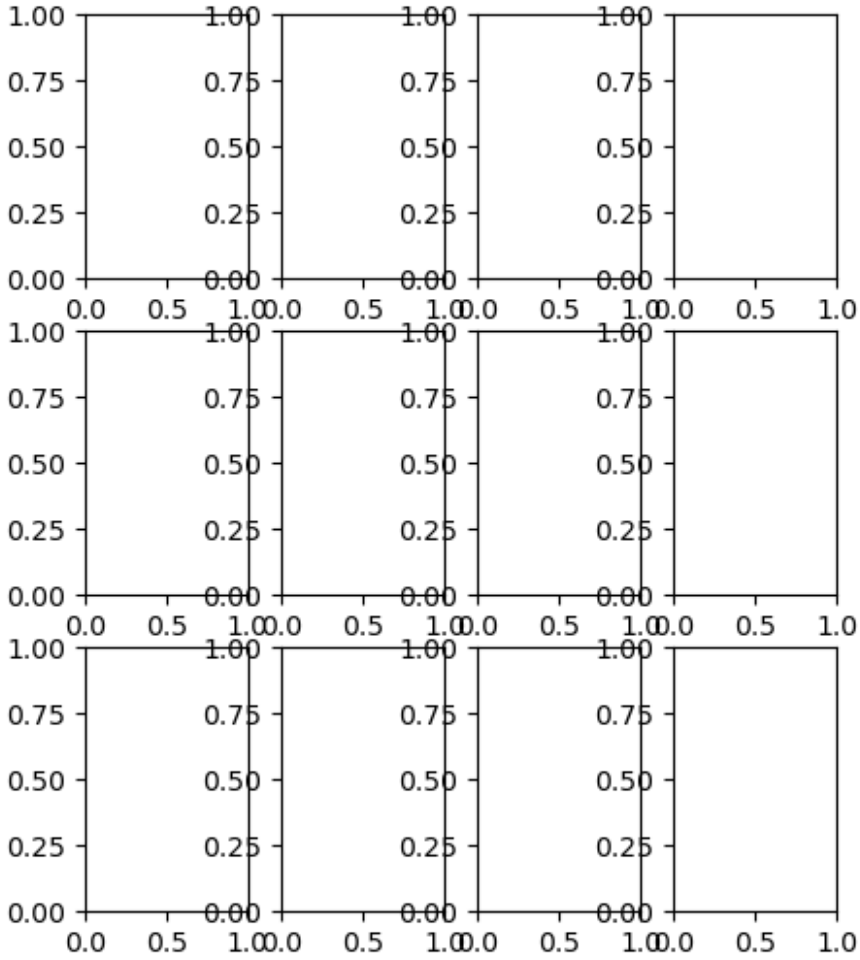
```
import numpy as np

import pandas as pd

df=pd.read_csv(r"C:\Users\Kruthi Mule\Downloads\archive\breast-cancer.csv")
df
```

```
           id diagnosis  radius_mean  texture_mean  perimeter_mean
area_mean  \
0      842302         M        17.99         10.38          122.80
1001.0
1      842517         M        20.57         17.77          132.90
1326.0
2    84300903         M        19.69         21.25          130.00
1203.0
3    84348301         M        11.42         20.38           77.58
386.1
4    84358402         M        20.29         14.34          135.10
1297.0
..        ...       ...          ...           ...             ...
```

```
...
564    926424         M        21.56         22.39          142.00
1479.0
565    926682         M        20.13         28.25          131.20
1261.0
566    926954         M        16.60         28.08          108.30
858.1
567    927241         M        20.60         29.33          140.10
1265.0
568     92751         B         7.76         24.54           47.92
181.0

     smoothness_mean  compactness_mean  concavity_mean  concave
points_mean  \
0            0.11840           0.27760         0.30010
0.14710
1            0.08474           0.07864         0.08690
0.07017
2            0.10960           0.15990         0.19740
0.12790
3            0.14250           0.28390         0.24140
0.10520
4            0.10030           0.13280         0.19800
0.10430
..               ...               ...             ...
...
564          0.11100           0.11590         0.24390
0.13890
565          0.09780           0.10340         0.14400
0.09791
566          0.08455           0.10230         0.09251
0.05302
567          0.11780           0.27700         0.35140
0.15200
568          0.05263           0.04362         0.00000
0.00000

     ...  radius_worst  texture_worst  perimeter_worst  area_worst  \
0    ...        25.380          17.33           184.60      2019.0
1    ...        24.990          23.41           158.80      1956.0
2    ...        23.570          25.53           152.50      1709.0
3    ...        14.910          26.50            98.87       567.7
4    ...        22.540          16.67           152.20      1575.0
..   ...           ...            ...              ...         ...
564  ...        25.450          26.40           166.10      2027.0
565  ...        23.690          38.25           155.00      1731.0
566  ...        18.980          34.12           126.70      1124.0
567  ...        25.740          39.42           184.60      1821.0
568  ...         9.456          30.37            59.16       268.6
```

```
     smoothness_worst  compactness_worst  concavity_worst  \
0             0.16220            0.66560           0.7119
1             0.12380            0.18660           0.2416
2             0.14440            0.42450           0.4504
3             0.20980            0.86630           0.6869
4             0.13740            0.20500           0.4000
..                ...                ...              ...
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     concave points_worst  symmetry_worst  fractal_dimension_worst
0                  0.2654          0.4601                  0.11890
1                  0.1860          0.2750                  0.08902
2                  0.2430          0.3613                  0.08758
3                  0.2575          0.6638                  0.17300
4                  0.1625          0.2364                  0.07678
..                    ...             ...                      ...
564                0.2216          0.2060                  0.07115
565                0.1628          0.2572                  0.06637
566                0.1418          0.2218                  0.07820
567                0.2650          0.4087                  0.12400
568                0.0000          0.2871                  0.07039

[569 rows x 32 columns]
```

```
import matplotlib.pyplot as plt

fig, axs = plt.subplots(3, 4, figsize=(5,6))
```



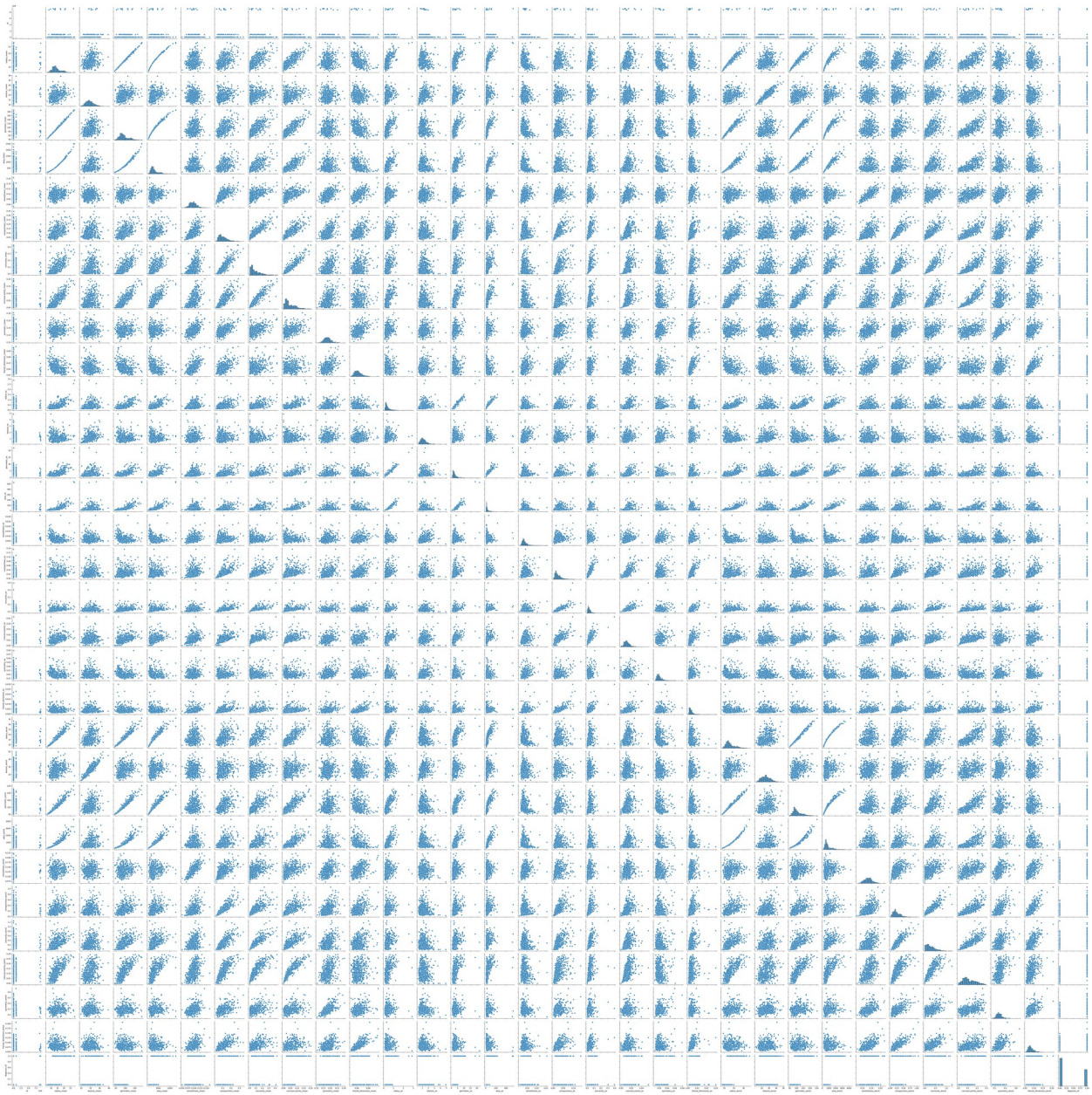
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 32 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   id                       569 non-null    int64  
 1   diagnosis                569 non-null    object 
 2   radius_mean              569 non-null    float64
 3   texture_mean             569 non-null    float64
 4   perimeter_mean           569 non-null    float64
 5   area_mean                569 non-null    float64
 6   smoothness_mean          569 non-null    float64
 7   compactness_mean         569 non-null    float64
 8   concavity_mean           569 non-null    float64
 9   concave points_mean      569 non-null    float64
 10  symmetry_mean            569 non-null    float64
 11  fractal_dimension_mean   569 non-null    float64
 12  radius_se                569 non-null    float64
```

```
 13  texture_se               569 non-null    float64
 14  perimeter_se             569 non-null    float64
 15  area_se                  569 non-null    float64
 16  smoothness_se            569 non-null    float64
 17  compactness_se           569 non-null    float64
 18  concavity_se             569 non-null    float64
 19  concave points_se        569 non-null    float64
 20  symmetry_se              569 non-null    float64
 21  fractal_dimension_se     569 non-null    float64
 22  radius_worst             569 non-null    float64
 23  texture_worst            569 non-null    float64
 24  perimeter_worst          569 non-null    float64
 25  area_worst               569 non-null    float64
 26  smoothness_worst         569 non-null    float64
 27  compactness_worst        569 non-null    float64
 28  concavity_worst          569 non-null    float64
 29  concave points_worst     569 non-null    float64
 30  symmetry_worst           569 non-null    float64
 31  fractal_dimension_worst  569 non-null    float64
dtypes: float64(30), int64(1), object(1)
memory usage: 142.4+ KB
```

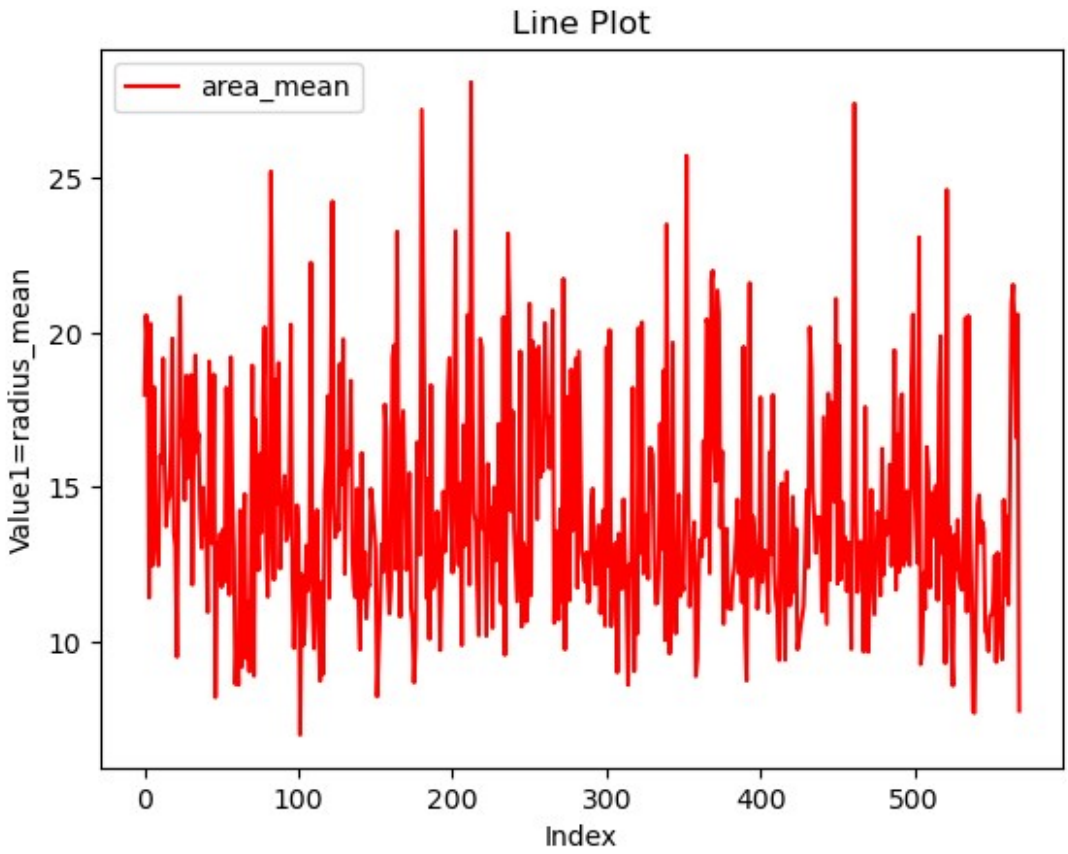
```
import seaborn as sns
sns.pairplot(data)
```

```
<seaborn.axisgrid.PairGrid at 0x15b96d57d40>
```



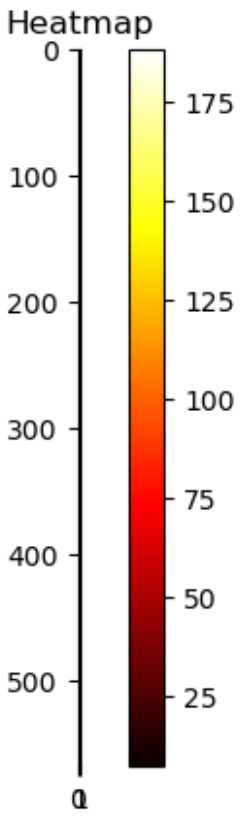
Line Plot

```
plt.plot(df.index, df['radius_mean'], label='area_mean', color='r')
plt.title('Line Plot')
plt.xlabel('Index')
plt.ylabel('Value1=radius_mean')
plt.legend()
plt.show()
```



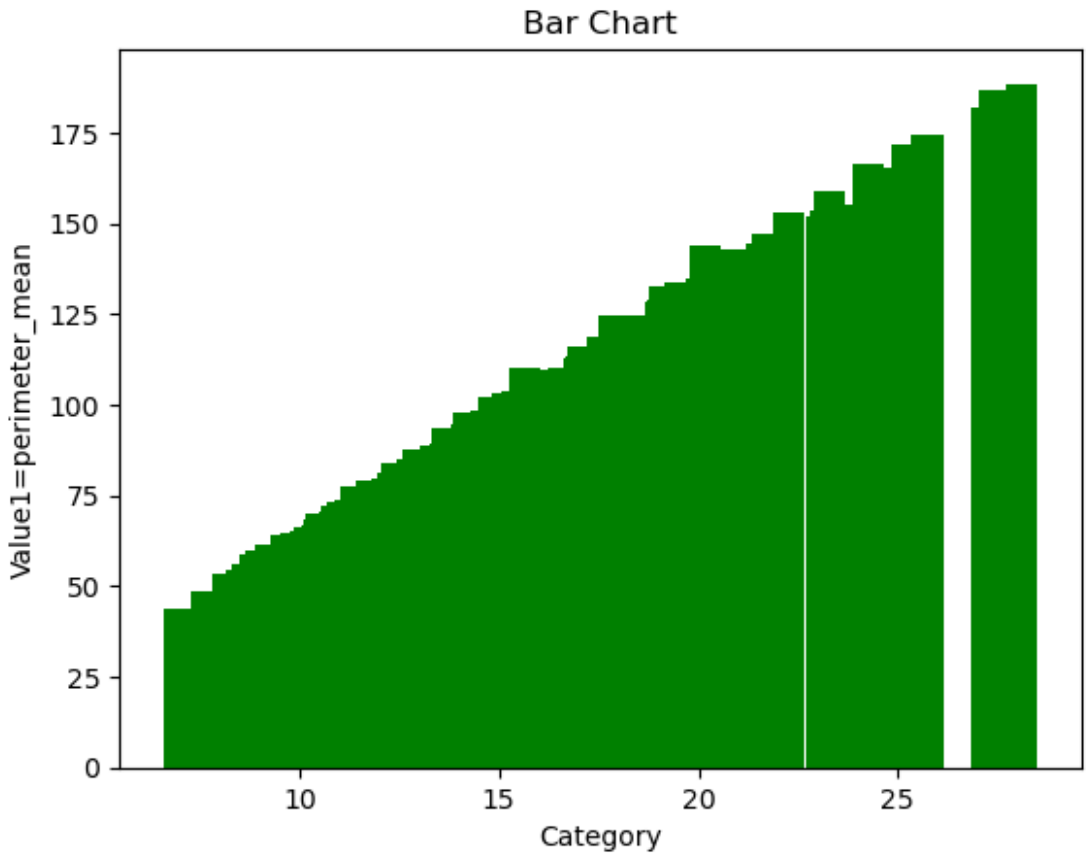
Heatmap

```
data =df[['radius_mean', 'perimeter_mean']].values
plt.imshow(data, cmap='hot')
plt.title('Heatmap')
plt.colorbar()
plt.show()
```



Bar Chart

```
plt.bar(df['radius_mean'], df['perimeter_mean'], color='g')
plt.title('Bar Chart')
plt.xlabel('Category')
plt.ylabel('Value1=perimeter_mean')
plt.show()
```

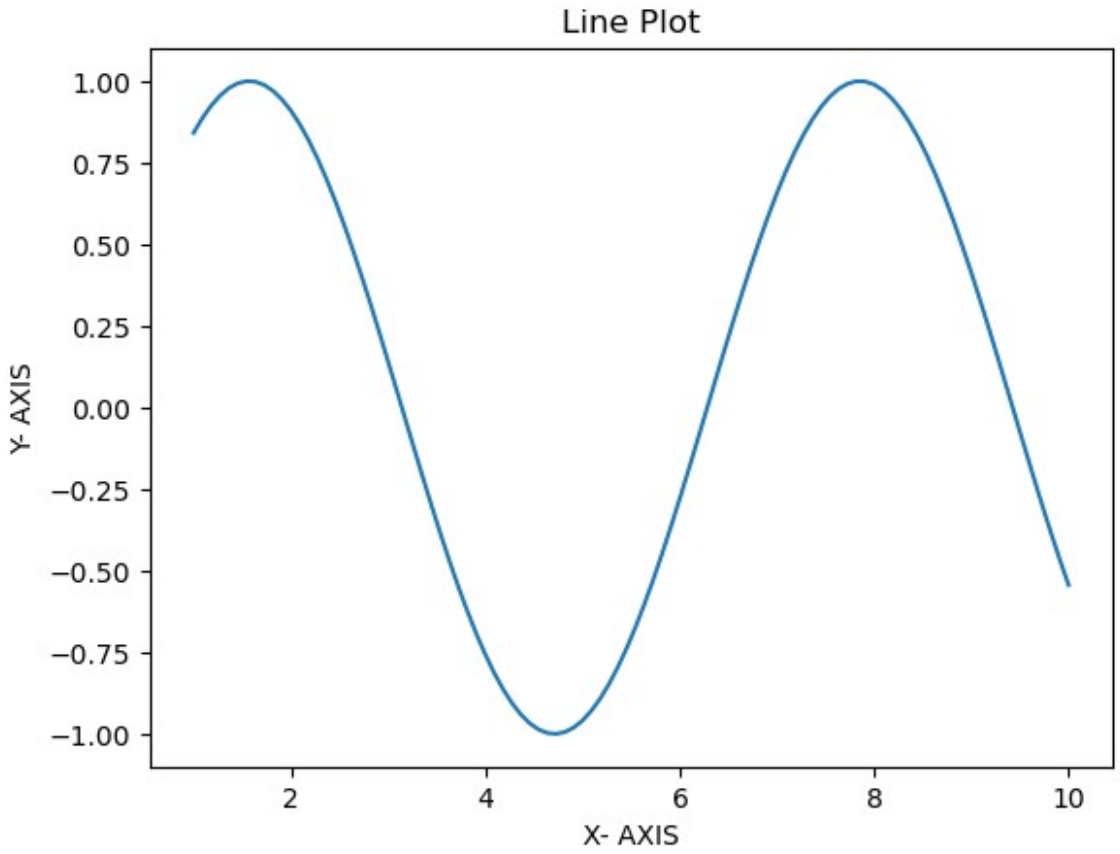



```
import  matplotlib.pyplot as plt
import numpy as np

# Data
x = np.linspace(1, 10, 100)
y = np.sin(x)
print("x=",x)
print("                              ")
print("Y",y)
# Line Plot
plt.plot(x, y)
plt.title('Line Plot')
plt.xlabel("X- AXIS")
plt.ylabel("Y- AXIS")
plt.show()
```

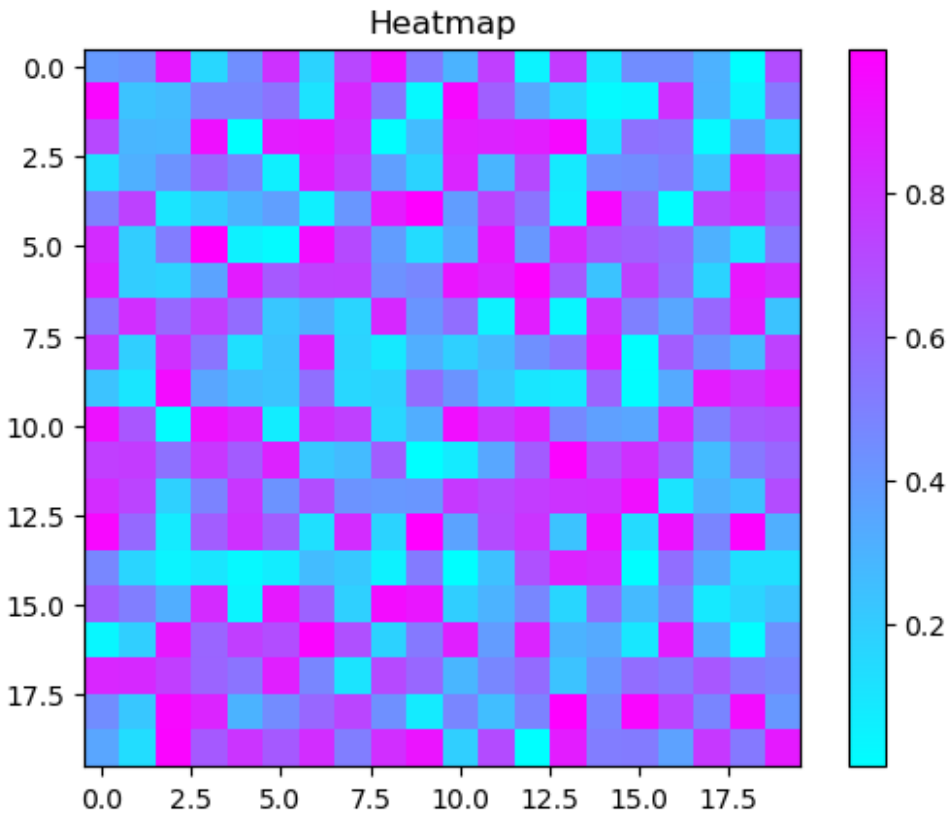
```
x= [ 1.          1.09090909  1.18181818  1.27272727  1.36363636
1.45454545
  1.54545455  1.63636364  1.72727273  1.81818182  1.90909091  2.
  2.09090909  2.18181818  2.27272727  2.36363636  2.45454545
2.54545455
  2.63636364  2.72727273  2.81818182  2.90909091  3.
3.09090909
```

```
  3.18181818  3.27272727  3.36363636  3.45454545  3.54545455
3.63636364
  3.72727273  3.81818182  3.90909091  4.          4.09090909
4.18181818
  4.27272727  4.36363636  4.45454545  4.54545455  4.63636364
4.72727273
  4.81818182  4.90909091  5.          5.09090909  5.18181818
5.27272727
  5.36363636  5.45454545  5.54545455  5.63636364  5.72727273
5.81818182
  5.90909091  6.          6.09090909  6.18181818  6.27272727
6.36363636
  6.45454545  6.54545455  6.63636364  6.72727273  6.81818182
6.90909091
  7.          7.09090909  7.18181818  7.27272727  7.36363636
7.45454545
  7.54545455  7.63636364  7.72727273  7.81818182  7.90909091  8.
  8.09090909  8.18181818  8.27272727  8.36363636  8.45454545
8.54545455
  8.63636364  8.72727273  8.81818182  8.90909091  9.
9.09090909
  9.18181818  9.27272727  9.36363636  9.45454545  9.54545455
9.63636364
  9.72727273  9.81818182  9.90909091 10.        ]

Y [ 0.84147098  0.88704699  0.92529707  0.95590534  0.978619
0.99325047
  0.99967891  0.99785123  0.98778253  0.96955595  0.94332203
0.90929743
  0.86776314  0.8190622   0.76359681  0.70182505  0.63425707
0.56145091
  0.48400786  0.40256749  0.31780241  0.23041267  0.14112001
0.05066187
 -0.04021468 -0.1307591  -0.22022362 -0.30786935 -0.39297247 -
0.47483011
 -0.55276624 -0.6261372  -0.69433703 -0.7568025  -0.8130177  -
0.86251837
 -0.9048957  -0.93979971 -0.96694212 -0.98609877 -0.99711147 -
0.99988924
 -0.99440916 -0.98071647 -0.95892427 -0.92921254 -0.89182665 -
0.84707537
 -0.79532828 -0.73701276 -0.67261042 -0.60265314 -0.52771868 -
0.44842592
 -0.36542971 -0.2794155  -0.19109366 -0.10119362 -0.01045784
0.0803643
  0.17052273  0.25927286  0.34588171  0.429634    0.50983804
0.58583144
  0.6569866   0.72271585  0.78247636  0.83577457  0.88217031
0.92128041
```

```
  0.95278186  0.9764145   0.99198316  0.99935926  0.99848187
0.98935825
  0.97206374  0.94674118  0.9135997   0.87291301  0.82501713
0.77030762
  0.70923631  0.64230758  0.57007418  0.49313267  0.41211849
0.32770071
  0.24057653  0.15146548  0.06110351 -0.0297631  -0.1203839  -
0.21001048
 -0.29790263 -0.38333447 -0.46560043 -0.54402111]
```

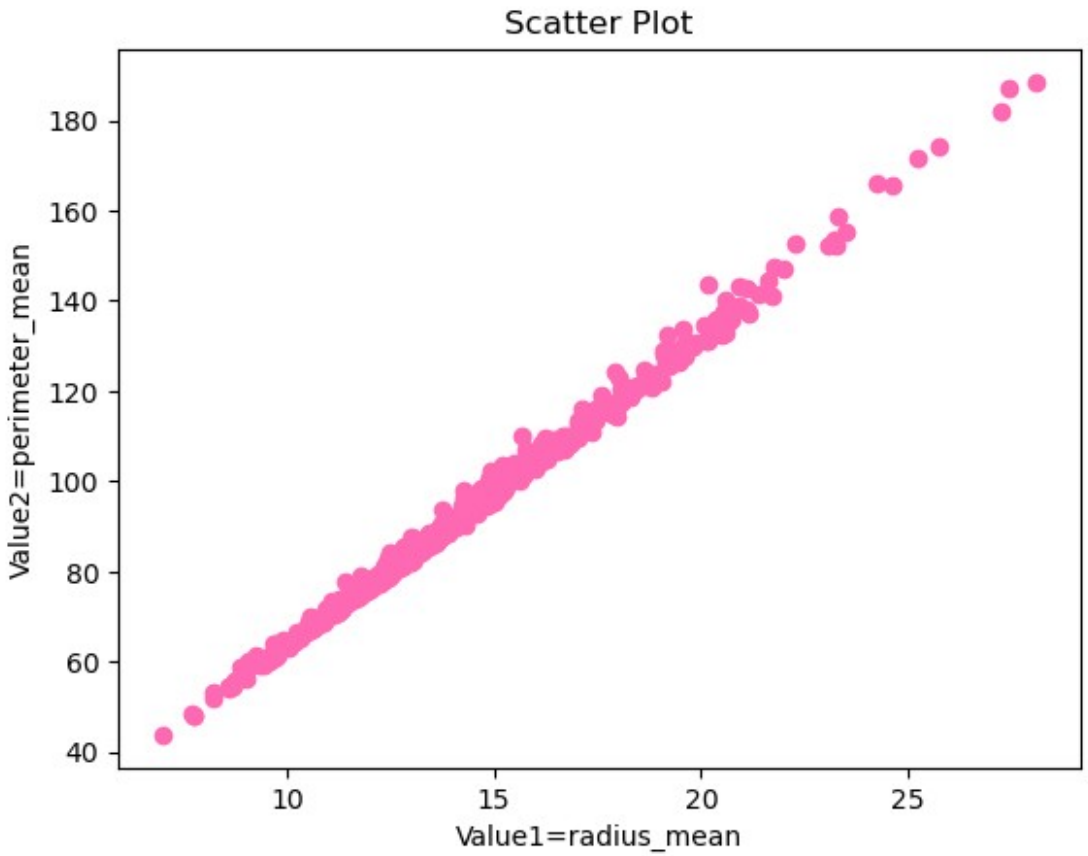



```python
import matplotlib.pyplot as plt
import numpy as np

# Data
data = np.random.rand(20, 20)

# Heatmap
plt.imshow(data, cmap='cool', interpolation='nearest')
plt.title('Heatmap')
plt.colorbar()
plt.show()
```



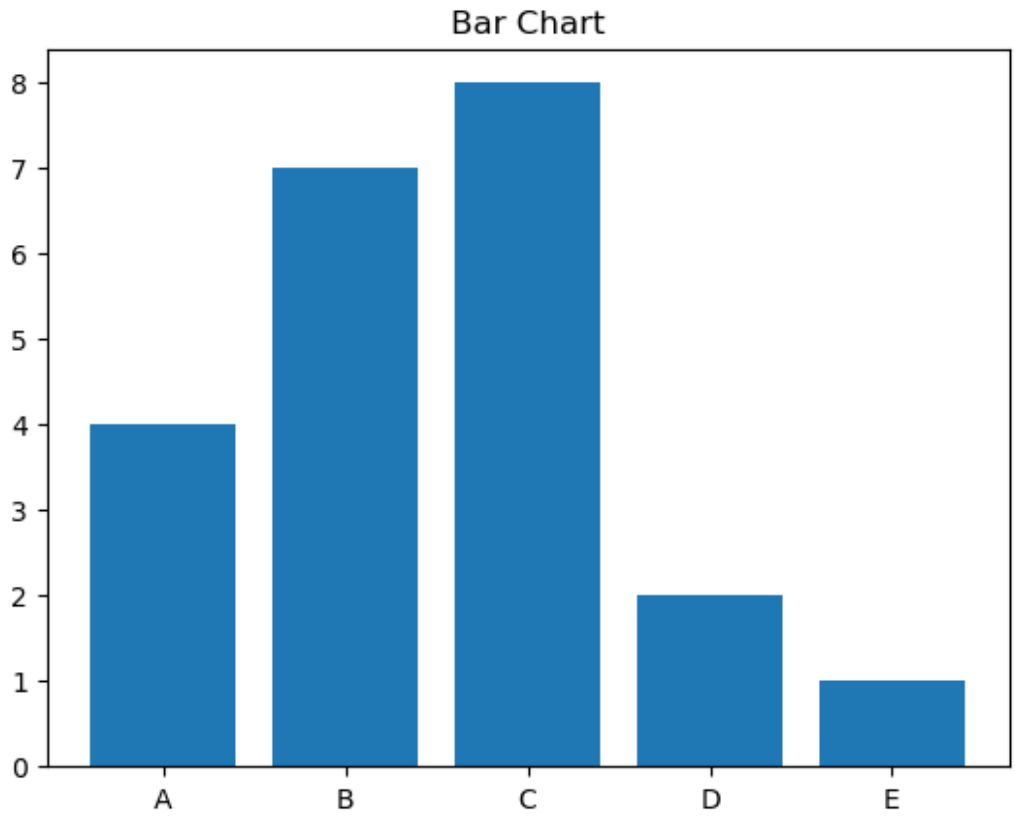
Scatter Plot

```
plt.scatter(df['radius_mean'], df['perimeter_mean'], color="#ff69b4")
plt.title('Scatter Plot')
plt.xlabel('Value1=radius_mean')
plt.ylabel('Value2=perimeter_mean')
plt.show()
```



Bar Chart

```
import matplotlib.pyplot as plt

# Data
categories = ['A', 'B', 'C', 'D', 'E',]
values = [4, 7, 8, 2, 1]

# Bar Chart
plt.bar(categories, values)
plt.title('Bar Chart')
plt.show()
```


Bar Chart
8
7
6
5
4
3
2
1
0
A
B
C
D
E