

```
In [1]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt
```

```
In [2]: data=pd.read_csv(r"C:\Users\user\Desktop\Vicky\8_BreastCancerPrediction.csv")
```

```
In [3]: data.shape
```

```
Out[3]: (569, 33)
```

In [4]: `data.head`

```
Out[4]: <bound method NDFrame.head of
ean  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
```

```
... texture_worst  perimeter_worst  area_worst  smoothness_worst  \
0      ...      17.33      184.60      2019.0      0.16220
1      ...      23.41      158.80      1956.0      0.12380
2      ...      25.53      152.50      1709.0      0.14440
3      ...      26.50      98.87      567.7      0.20980
4      ...      16.67      152.20      1575.0      0.13740
..      ...      ...      ...      ...      ...
564      ...      26.40      166.10      2027.0      0.14100
565      ...      38.25      155.00      1731.0      0.11660
566      ...      34.12      126.70      1124.0      0.11390
567      ...      39.42      184.60      1821.0      0.16500
568      ...      30.37      59.16      268.6      0.08996
```

```
compactness_worst  concavity_worst  concave points_worst  symmetry_worst
\
0      0.66560      0.7119      0.2654      0.4601
1      0.18660      0.2416      0.1860      0.2750
2      0.42450      0.4504      0.2430      0.3613
3      0.86630      0.6869      0.2575      0.6638
4      0.20500      0.4000      0.1625      0.2364
..      ...      ...      ...      ...
564      0.21130      0.4107      0.2216      0.2060
565      0.19220      0.3215      0.1628      0.2572
566      0.30940      0.3403      0.1418      0.2218
567      0.86810      0.9387      0.2650      0.4087
568      0.06444      0.0000      0.0000      0.2871
```

```
fractal_dimension_worst  Unnamed: 32
0      0.11890      NaN
```

1	0.08902	NaN
2	0.08758	NaN
3	0.17300	NaN
4	0.07678	NaN
..
564	0.07115	NaN
565	0.06637	NaN
566	0.07820	NaN
567	0.12400	NaN
568	0.07039	NaN

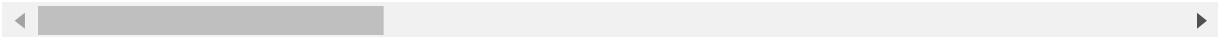
[569 rows x 33 columns]>

```
In [5]: data.tail()
```

Out[5]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_m
564	926424	M	21.56	22.39	142.00	1479.0	0.11
565	926682	M	20.13	28.25	131.20	1261.0	0.09
566	926954	M	16.60	28.08	108.30	858.1	0.08
567	927241	M	20.60	29.33	140.10	1265.0	0.11
568	92751	B	7.76	24.54	47.92	181.0	0.05

5 rows x 33 columns

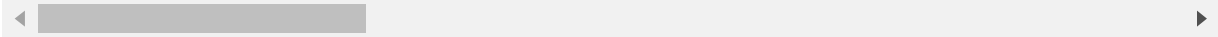


```
In [6]: data.describe()
```

Out[6]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
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	0.014060
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 x 32 columns



```
In [7]: data.isna()
```

Out[7]:

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

569 rows × 33 columns

```
In [8]: data.size
```

Out[8]: 18777

In [9]: `data.isna`

```
Out[9]: <bound method DataFrame.isna of
```

	mean	perimeter_mean	area_mean	\	id	diagnosis	radius_mean	texture
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

	texture_worst	perimeter_worst	area_worst	smoothness_worst	\
0	17.33	184.60	2019.0	0.16220	
1	23.41	158.80	1956.0	0.12380	
2	25.53	152.50	1709.0	0.14440	
3	26.50	98.87	567.7	0.20980	
4	16.67	152.20	1575.0	0.13740	
..	
564	26.40	166.10	2027.0	0.14100	
565	38.25	155.00	1731.0	0.11660	
566	34.12	126.70	1124.0	0.11390	
567	39.42	184.60	1821.0	0.16500	
568	30.37	59.16	268.6	0.08996	

	compactness_worst	concavity_worst	concave points_worst	symmetry_worst
0	0.66560	0.7119	0.2654	0.4601
1	0.18660	0.2416	0.1860	0.2750
2	0.42450	0.4504	0.2430	0.3613
3	0.86630	0.6869	0.2575	0.6638
4	0.20500	0.4000	0.1625	0.2364
..
564	0.21130	0.4107	0.2216	0.2060
565	0.19220	0.3215	0.1628	0.2572
566	0.30940	0.3403	0.1418	0.2218
567	0.86810	0.9387	0.2650	0.4087
568	0.06444	0.0000	0.0000	0.2871

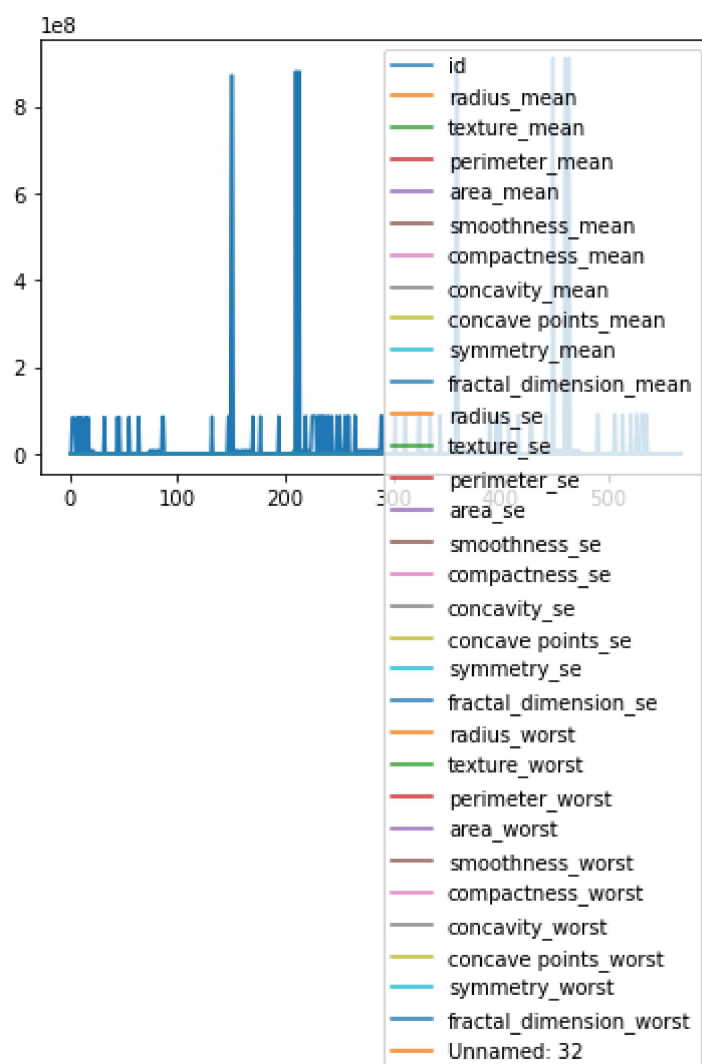
	fractal_dimension_worst	Unnamed: 32
0	0.11890	NaN

1	0.08902	NaN
2	0.08758	NaN
3	0.17300	NaN
4	0.07678	NaN
..
564	0.07115	NaN
565	0.06637	NaN
566	0.07820	NaN
567	0.12400	NaN
568	0.07039	NaN

[569 rows x 33 columns]>

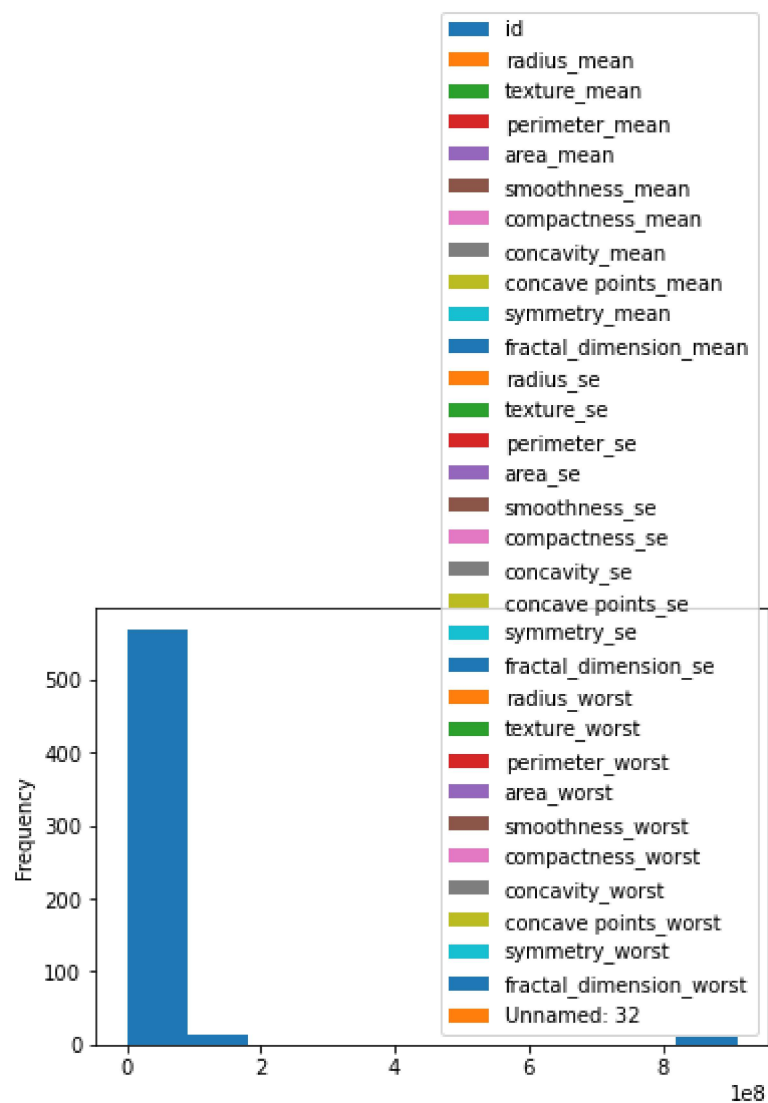
In [10]: data.plot.line()

Out[10]: <AxesSubplot:>



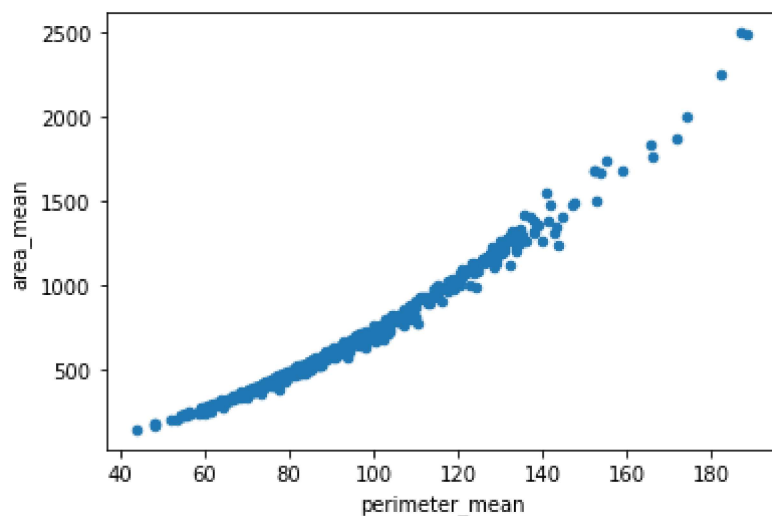

```
In [11]: data.plot.hist()
```

```
Out[11]: <AxesSubplot:ylabel='Frequency'>
```



```
In [14]: data.plot.scatter(x="perimeter_mean",y='area_mean')
```

```
Out[14]: <AxesSubplot:xlabel='perimeter_mean', ylabel='area_mean'>
```

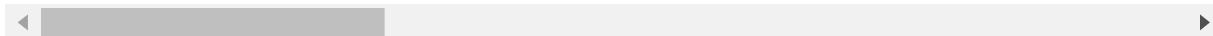


```
In [17]: data.fillna(value=5)
```

```
Out[17]:
```

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

569 rows × 33 columns



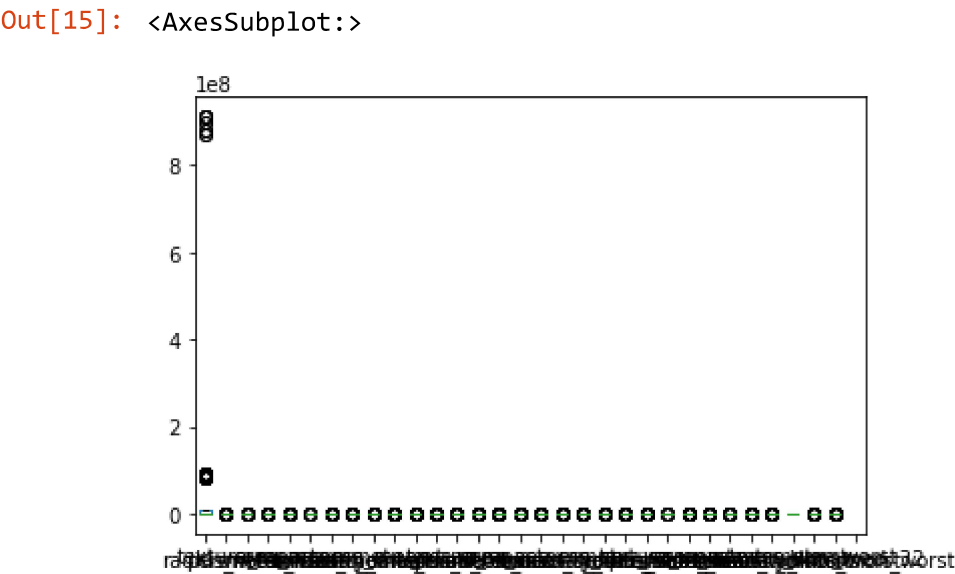
```
In [16]: pd.isna(data)
```

Out[16]:

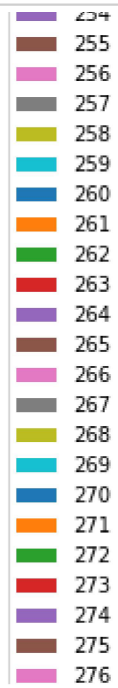
	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...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

569 rows × 33 columns

```
In [15]: data.plot.box()
```

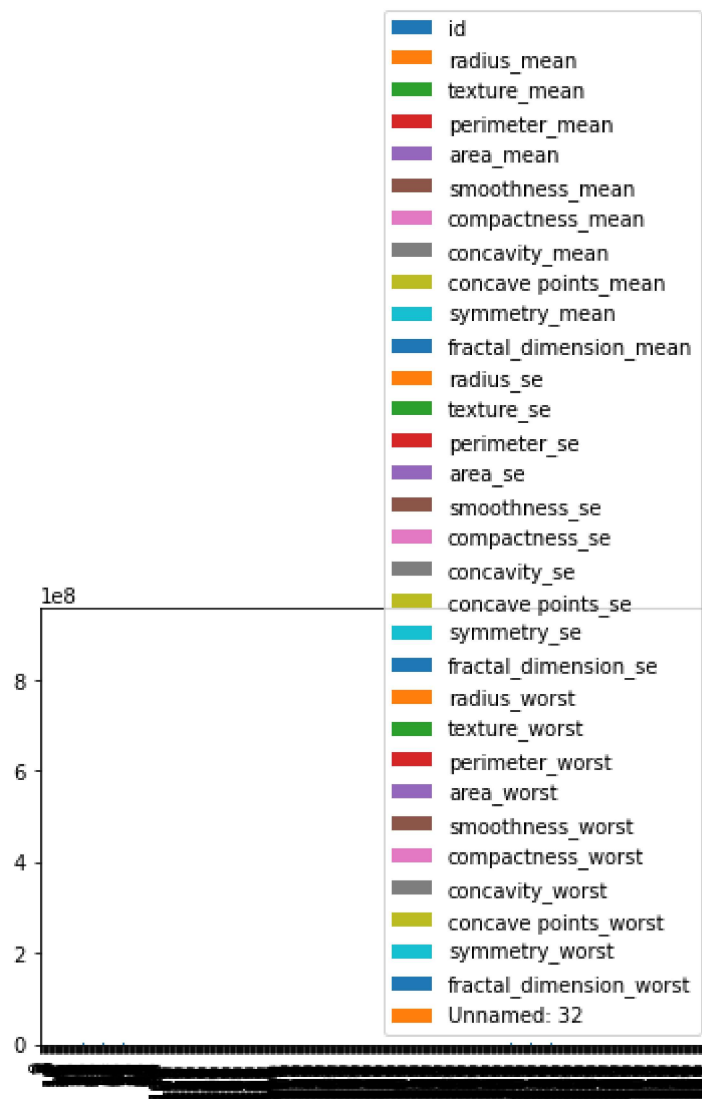


```
In [13]: data.plot.pie(y="area_mean")
```



In [18]: `data.plot.bar()`

Out[18]: `<AxesSubplot:>`



In []: