Clean, preprocess and visulaize the data

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

In [2]:
```
d=pd.read_csv(r"C:\Users\Admin\Downloads\8_BreastCancerPrediction - 8_BreastCar
d
```

Out[2]:

	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	0.1
3	84348301	M	11.42	20.38	77.58	386.1	0.1
4	84358402	M	20.29	14.34	135.10	1297.0	0.1
...	...	...	...	...	...	...	
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 × 32 columns

In [3]:
```
d.head()
```

Out[3]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_me
0	842302	M	17.99	10.38	122.80	1001.0	0.118
1	842517	M	20.57	17.77	132.90	1326.0	0.084
2	84300903	M	19.69	21.25	130.00	1203.0	0.109
3	84348301	M	11.42	20.38	77.58	386.1	0.142
4	84358402	M	20.29	14.34	135.10	1297.0	0.100

5 rows × 32 columns

In [4]:
```
d.tail()
```

Out[4]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_me
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 × 32 columns

In [5]:
```
d.describe()
```

Out[5]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mea
count	5.690000e+02	569.000000	569.000000	569.000000	569.000000	569.00000
mean	3.037183e+07	14.127292	19.289649	91.969033	654.889104	0.09636
std	1.250206e+08	3.524049	4.301036	24.298981	351.914129	0.01406
min	8.670000e+03	6.981000	9.710000	43.790000	143.500000	0.05263
25%	8.692180e+05	11.700000	16.170000	75.170000	420.300000	0.08637
50%	9.060240e+05	13.370000	18.840000	86.240000	551.100000	0.09587
75%	8.813129e+06	15.780000	21.800000	104.100000	782.700000	0.10530
max	9.113205e+08	28.110000	39.280000	188.500000	2501.000000	0.16340

8 rows × 31 columns

In [6]:
```
d.shape
```

Out[6]: (569, 32)

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

Out[8]: 18208

In [9]:
```
d.isna
```

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

In [11]:
```
d.dropna(axis=1,how="any")
```

Out[11]:

	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 × 32 columns

In [12]:
```
d["id"]
```

Out[12]:
```
0        842302
1        842517
2      84300903
3      84348301
4      84358402
         ...
564      926424
565      926682
566      926954
567      927241
568       92751
Name: id, Length: 569, dtype: int64
```

In [18]:
```
d1=d[["id","area_mean"]]
d1
```

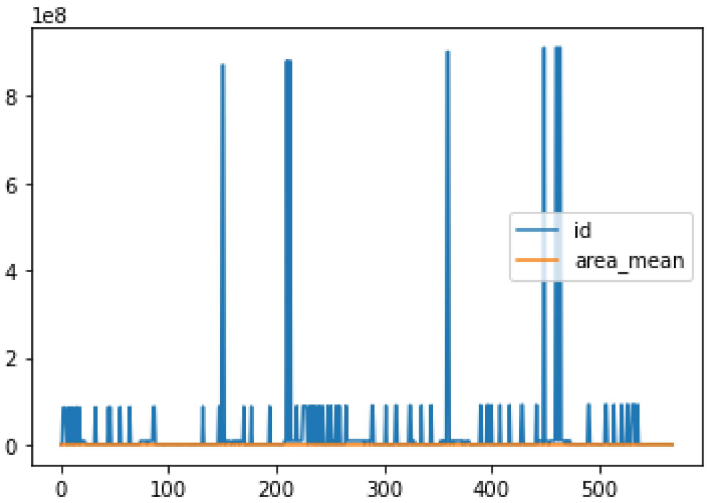
Out[18]:

	id	area_mean
0	842302	1001.0
1	842517	1326.0
2	84300903	1203.0
3	84348301	386.1
4	84358402	1297.0
...	...	...
564	926424	1479.0
565	926682	1261.0
566	926954	858.1
567	927241	1265.0
568	92751	181.0

569 rows × 2 columns

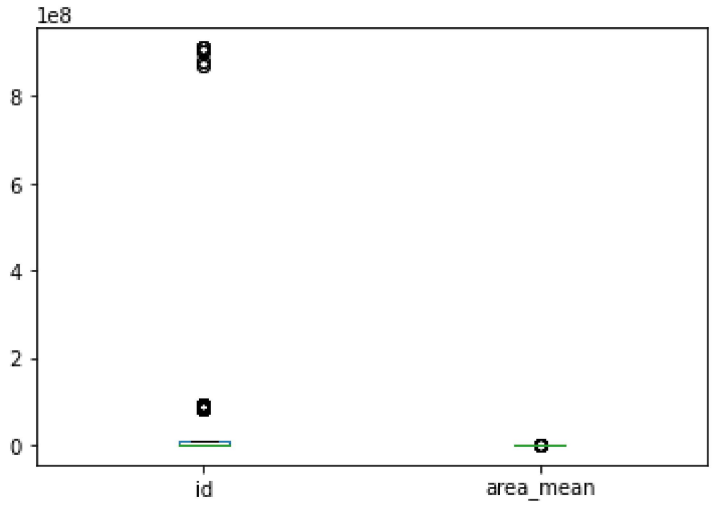
In [19]:
```
d1.plot.line()
```

Out[19]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x19133243e80>
```



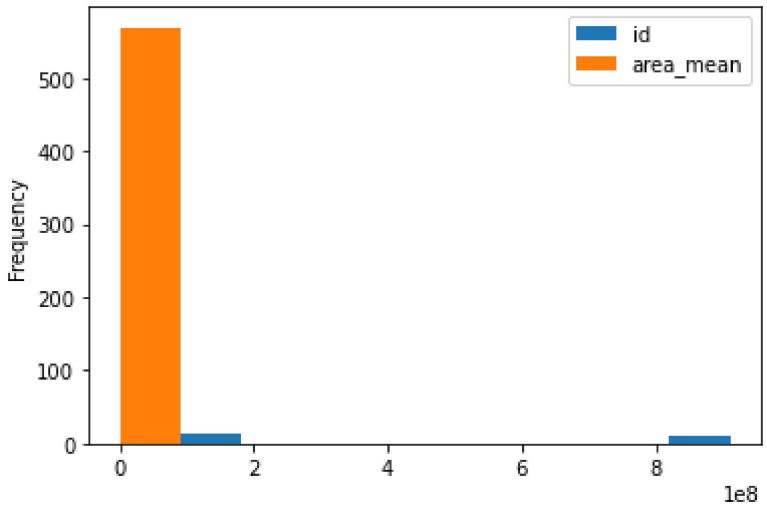
In [20]:
```
d1.plot.box()
```

Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x191332e5040>



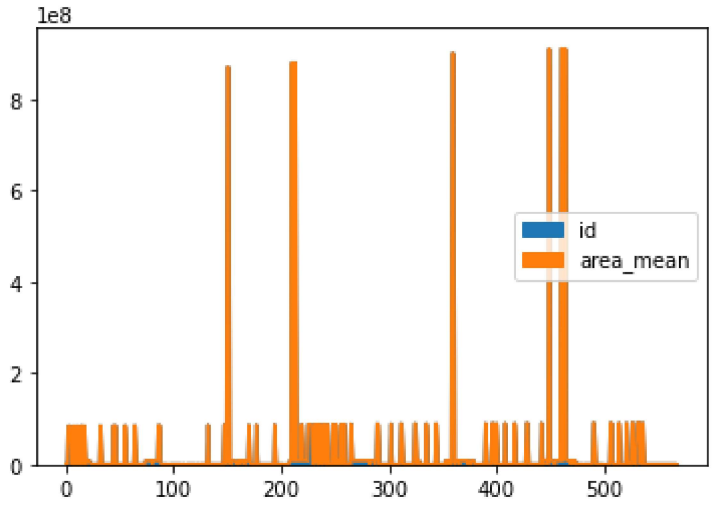
In [21]:
```
d1.plot.hist()
```

Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x19133397730>



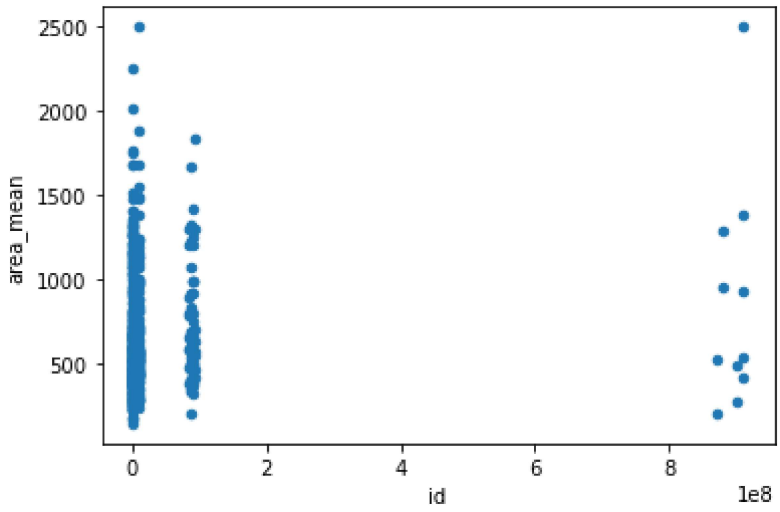
In [22]:
```
d1.plot.area()
```

Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x19133433220>



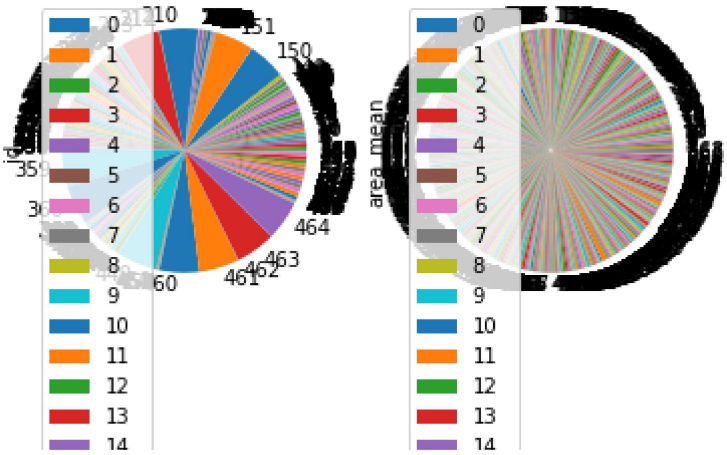
In [24]:
```
d1.plot.scatter(x="id",y="area_mean")
```

Out[24]: <matplotlib.axes._subplots.AxesSubplot at 0x1913353ce20>



In [28]:
```
d1.plot.pie(subplots=True)
```

Out[28]:
```
array([<matplotlib.axes._subplots.AxesSubplot object at 0x000001913627C820
>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x0000019135D9FDC0
>],
      dtype=object)
```



In [ ]: