Import labary

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

Import dataset

```
In [2]: data=pd.read_csv(r"c:\Users\user\Downloads\8_dataset.csv")
        data
```

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.'
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

Print head first 20 rows

In [3]:

```
data.head(20)
```

Out[3]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_m
0	842302	M	17.99	10.38	122.80	1001.0	0.1
1	842517	M	20.57	17.77	132.90	1326.0	0.08
2	84300903	M	19.69	21.25	130.00	1203.0	0.10
3	84348301	M	11.42	20.38	77.58	386.1	0.14
4	84358402	M	20.29	14.34	135.10	1297.0	0.10
5	843786	M	12.45	15.70	82.57	477.1	0.12
6	844359	M	18.25	19.98	119.60	1040.0	0.09
7	84458202	M	13.71	20.83	90.20	577.9	0.1
8	844981	M	13.00	21.82	87.50	519.8	0.12
9	84501001	M	12.46	24.04	83.97	475.9	0.1
10	845636	M	16.02	23.24	102.70	797.8	0.08
11	84610002	M	15.78	17.89	103.60	781.0	0.09
12	846226	M	19.17	24.80	132.40	1123.0	0.09
13	846381	M	15.85	23.95	103.70	782.7	0.08
14	84667401	M	13.73	22.61	93.60	578.3	0.1
15	84799002	M	14.54	27.54	96.73	658.8	0.1
16	848406	M	14.68	20.13	94.74	684.5	0.09
17	84862001	M	16.13	20.68	108.10	798.8	0.1
18	849014	M	19.81	22.15	130.00	1260.0	0.09
19	8510426	B	13.54	14.36	87.46	566.3	0.09

20 rows × 33 columns

Print tail last 7 rows

In [12]:
```
data.tail(7)
```

Out[12]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_me
562	925622	M	15.22	30.62	103.40	716.9	0.104
563	926125	M	20.92	25.09	143.00	1347.0	0.109
564	926424	M	21.56	22.39	142.00	1479.0	0.111
565	926682	M	20.13	28.25	131.20	1261.0	0.097
566	926954	M	16.60	28.08	108.30	858.1	0.084
567	927241	M	20.60	29.33	140.10	1265.0	0.117
568	92751	B	7.76	24.54	47.92	181.0	0.052

7 rows × 33 columns

To print statistical data

In [13]:
```
data.describe()
```

Out[13]:

	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 × 32 columns

To print rows and coloum

In [14]:
```
np.shape(data)
```

Out[14]:
```
(569, 33)
```

To print no. of elements

In [15]:
```
np.size(data)
```

Out[15]:
```
18777
```

To print missing values

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

Out[16]:

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

569 rows × 33 columns

Filla a value 60 in missing place

In [17]:

```
data.fillna(value=60)
```

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 [19]:

```
import matplotlib.pyplot as pp
```

In [20]:

```
dd=data[['id','area_mean']]
dd
```

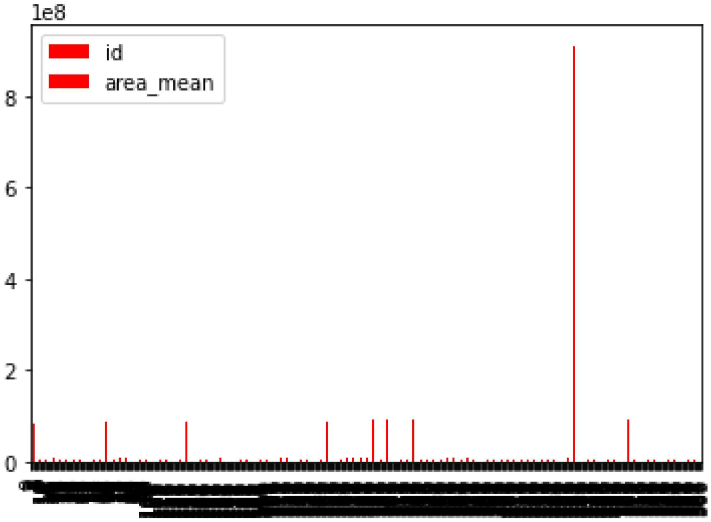
Out[20]:

	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

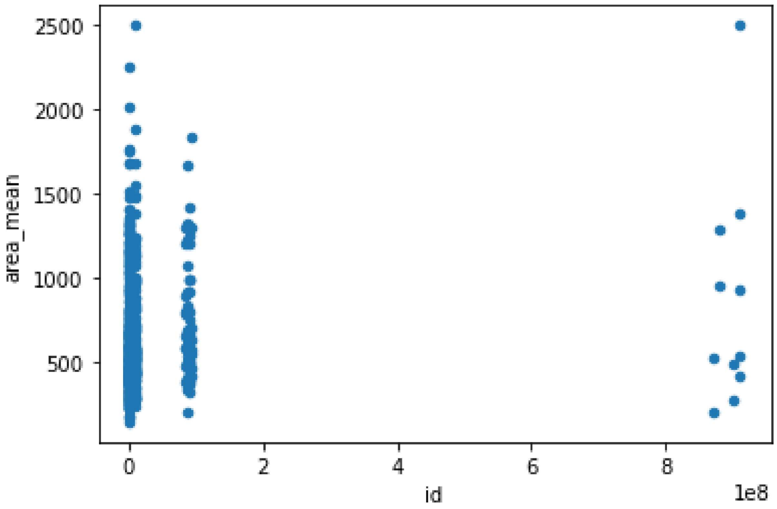
```
In [21]: dd.plot.bar(color='r')
```

```
Out[21]: <AxesSubplot:>
```



```
In [22]: dd.plot.scatter(x='id',y='area_mean')
```

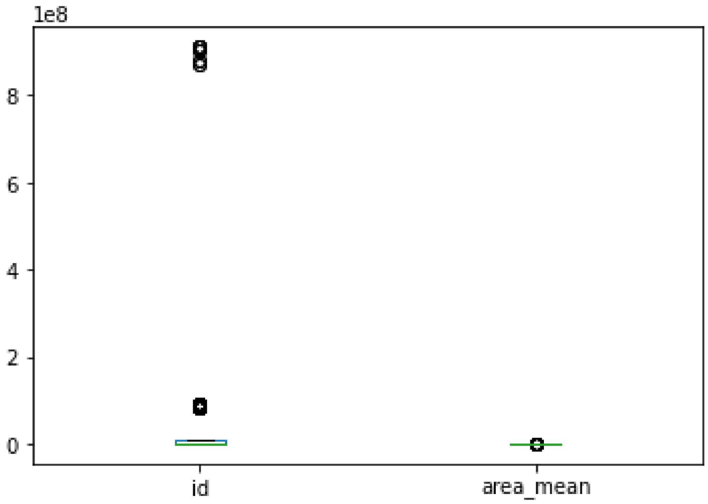
```
Out[22]: <AxesSubplot:xlabel='id', ylabel='area_mean'>
```



```
In [23]: dd.plot.pie(y='id')
```

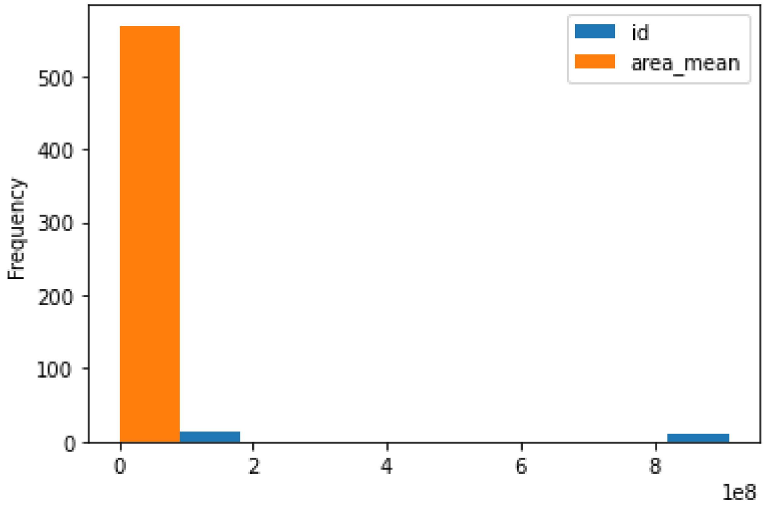
```
In [24]: dd.plot.box()
```

```
Out[24]: <AxesSubplot:>
```



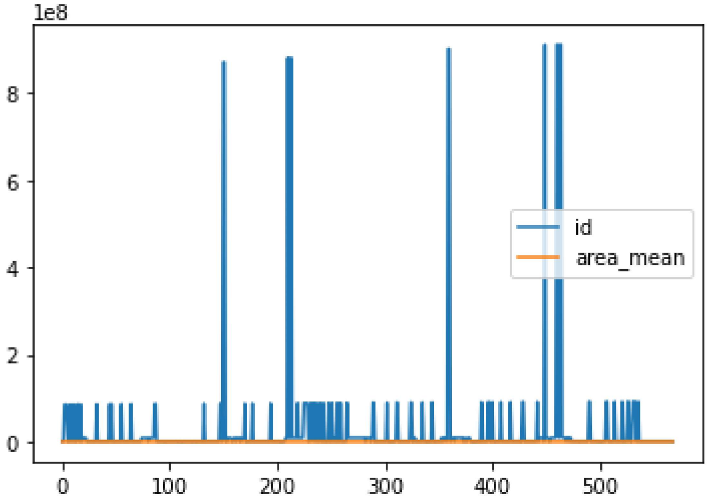
In [25]:
```
dd.plot.hist()
```

Out[25]:
```
<AxesSubplot:ylabel='Frequency'>
```



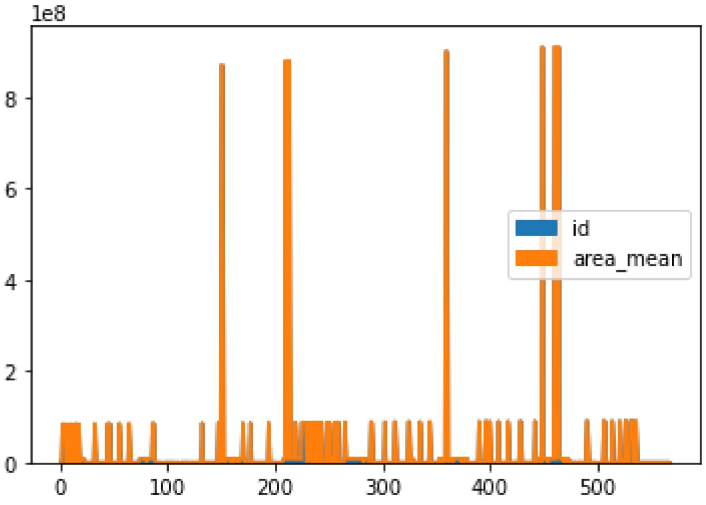
In [26]:
```
dd.plot.line()
```

Out[26]:
```
<AxesSubplot:>
```



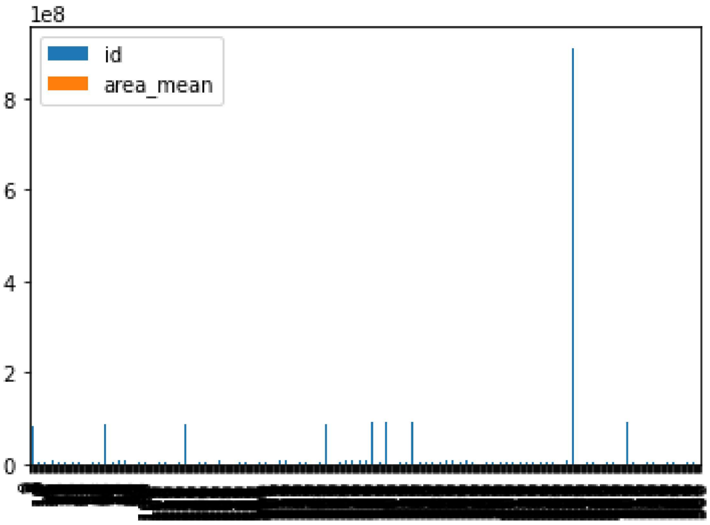
```
In [27]: dd.plot.area()
```

```
Out[27]: <AxesSubplot:>
```



```
In [28]: dd.plot.bar()
```

```
Out[28]: <AxesSubplot:>
```



```
In [ ]:
```