

Data set 8

In [1]:

```
import numpy as np
import pandas as pd
```

In [3]:

```
a=pd.read_csv(r"C:\Users\user\Downloads\8_BreastCancerPrediction.csv")
```

To print top rows:

In [4]:

```
a.head()
```

Out[4]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness
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	(

5 rows × 33 columns



To print Last rows:

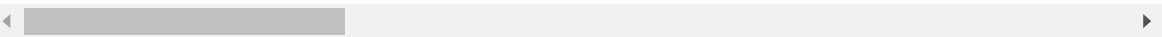
In [5]:

```
a.tail()
```

Out[5]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness
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	(

5 rows × 33 columns



Statistical Summary:

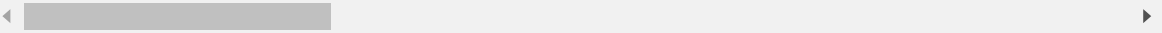
In [6]:

```
a.describe()
```

Out[6]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_
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	0.054575
std	1.250206e+08	3.524049	4.301036	24.298981	351.914129	0.014611
min	8.670000e+03	6.981000	9.710000	43.790000	143.500000	0.049946
25%	8.692180e+05	11.700000	16.170000	75.170000	420.300000	0.053079
50%	9.060240e+05	13.370000	18.840000	86.240000	551.100000	0.055647
75%	8.813129e+06	15.780000	21.800000	104.100000	782.700000	0.100279
max	9.113205e+08	28.110000	39.280000	188.500000	2501.000000	0.163459

8 rows × 32 columns



To print no of rows and columns

In [7]:

```
a.shape
```

Out[7]:

(569, 33)

To print no of elements

In [8]:

```
a.size
```

Out[8]:

18777

Missing no of values

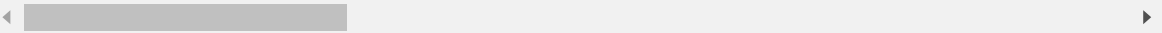
In [9]:

```
a.isna()
```

Out[9]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_
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	

569 rows × 33 columns



In [10]:

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

In [11]:

```
b=a[['radius_mean','perimeter_mean']]
b
```

Out[11]:

	radius_mean	perimeter_mean
0	17.99	122.80
1	20.57	132.90
2	19.69	130.00
3	11.42	77.58
4	20.29	135.10
...
564	21.56	142.00
565	20.13	131.20
566	16.60	108.30
567	20.60	140.10
568	7.76	47.92

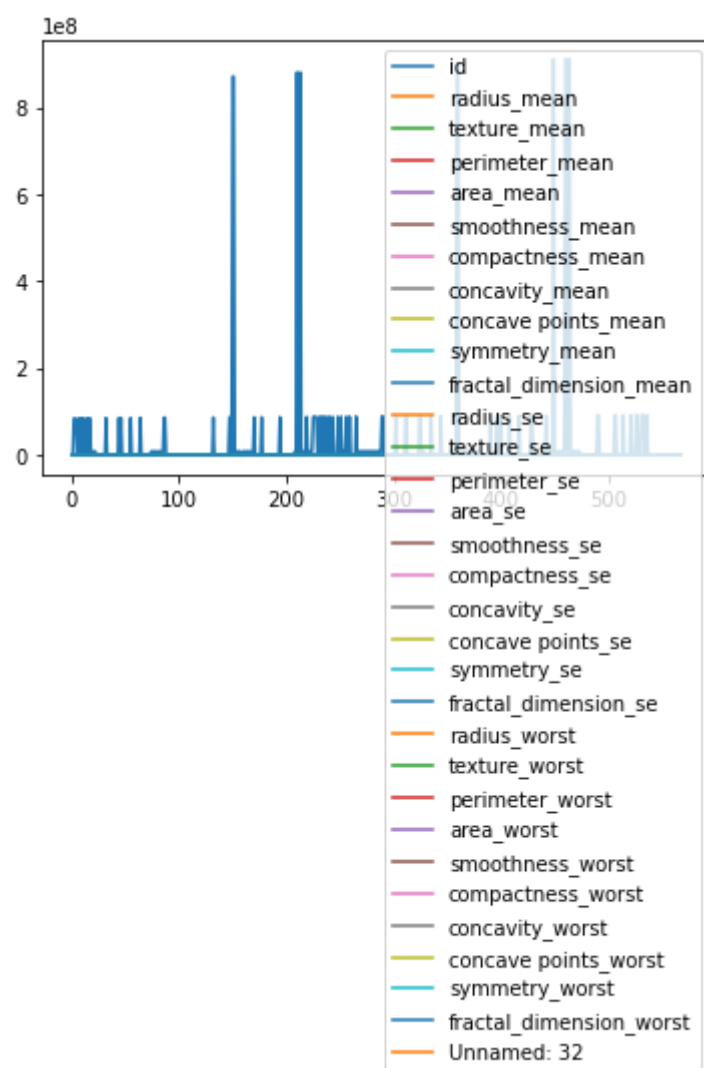
569 rows × 2 columns

In [12]:

a.plot.line()

Out[12]:

<AxesSubplot:>

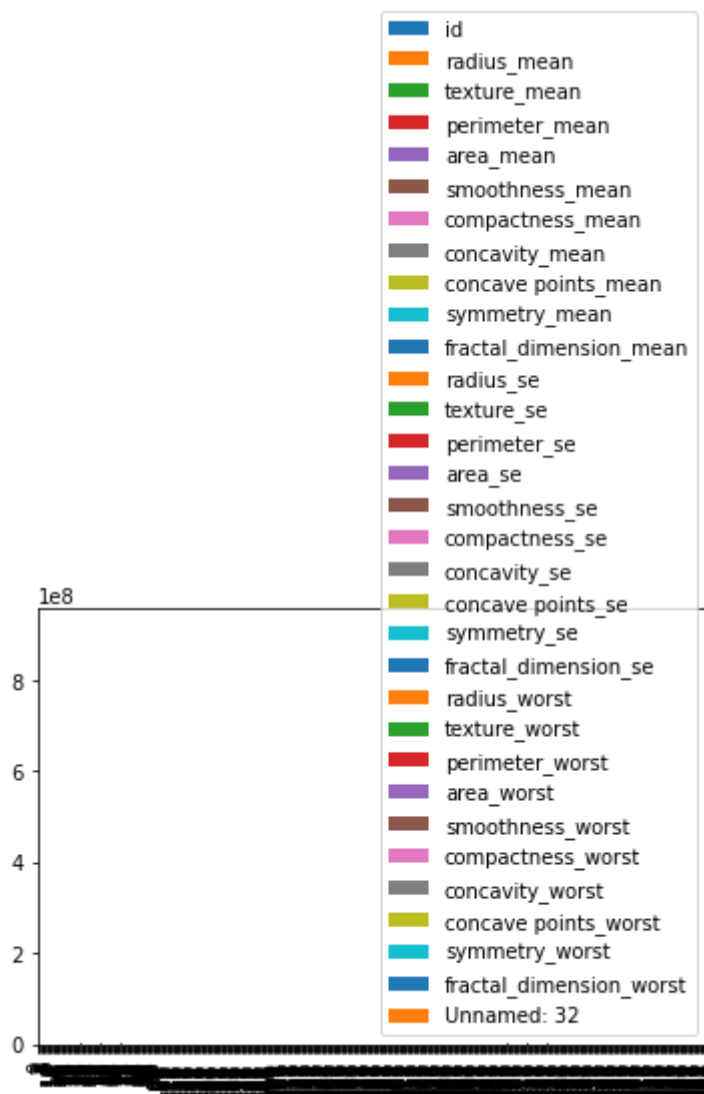


In [13]:

```
a.plot.bar()
```

Out[13]:

<AxesSubplot:>

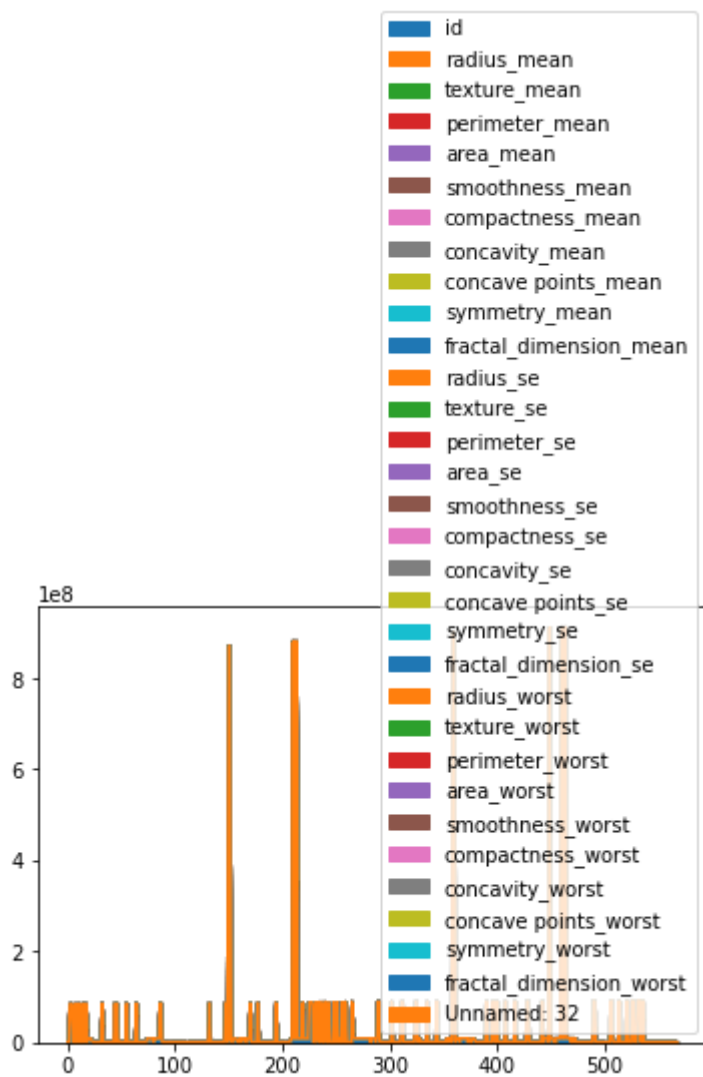


In [14]:

a.plot.area()

Out[14]:

<AxesSubplot:>

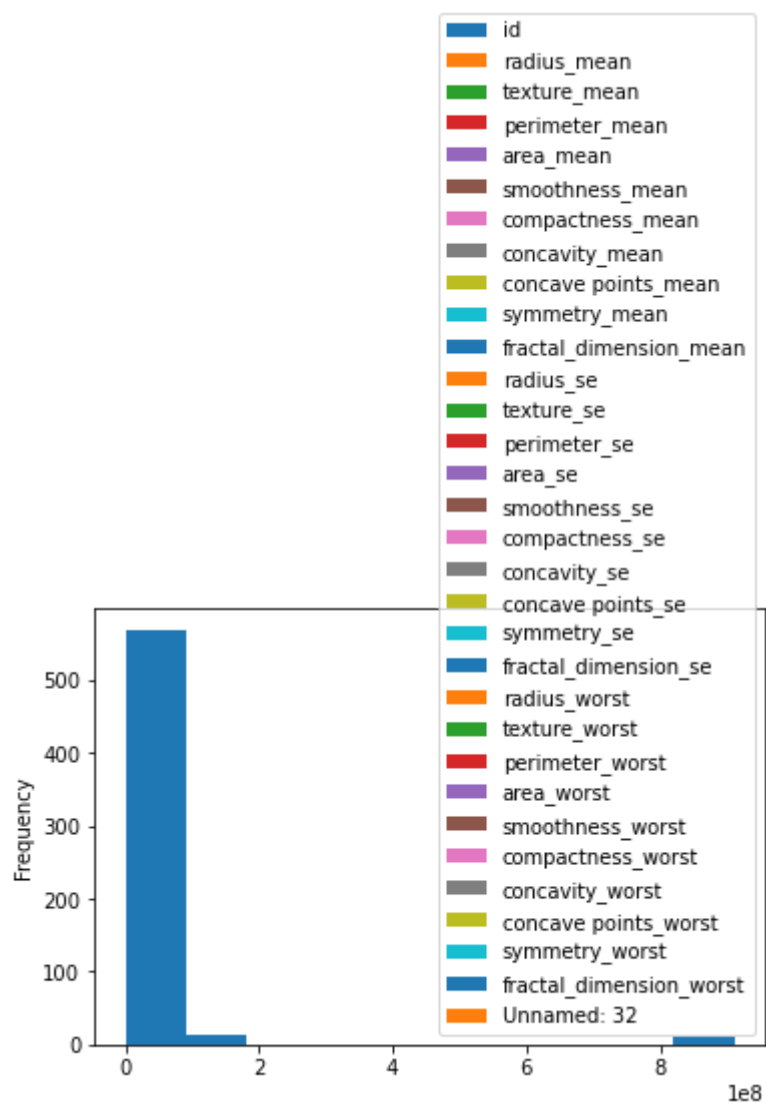


In [15]:

```
a.plot.hist()
```

Out[15]:

<AxesSubplot:ylabel='Frequency'>

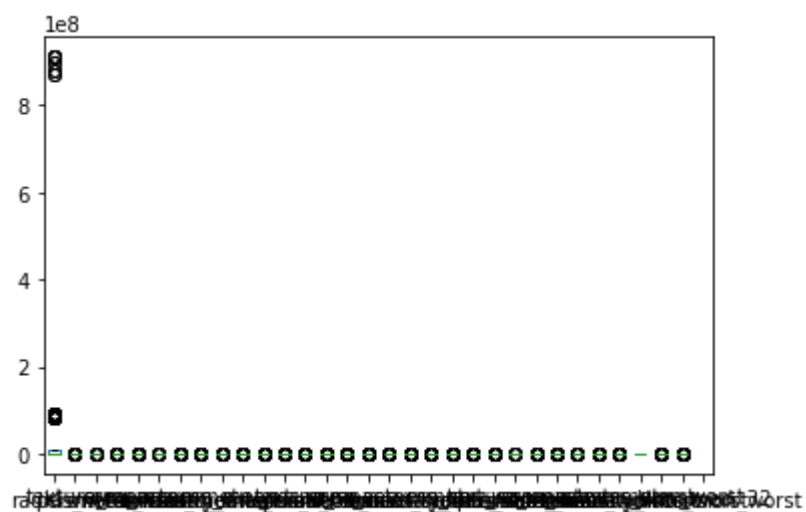


In [16]:

```
a.plot.box()
```

Out[16]:

<AxesSubplot:>

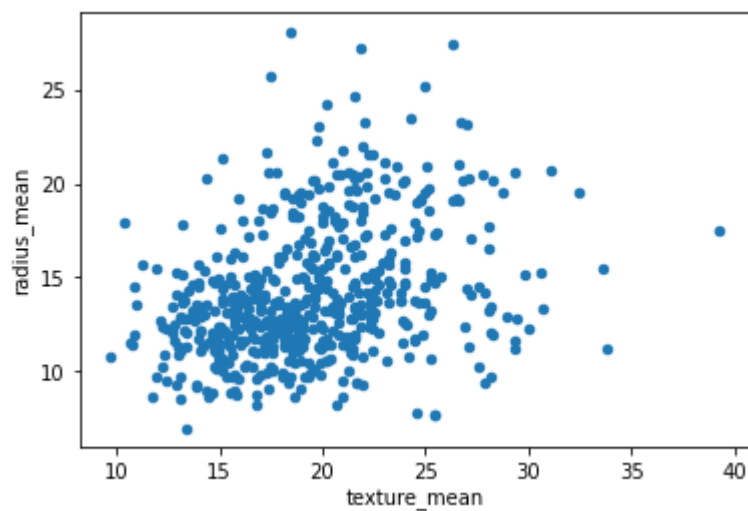


In [17]:

```
a.plot.scatter(x='texture_mean',y='radius_mean')
```

Out[17]:

<AxesSubplot:xlabel='texture_mean', ylabel='radius_mean'>

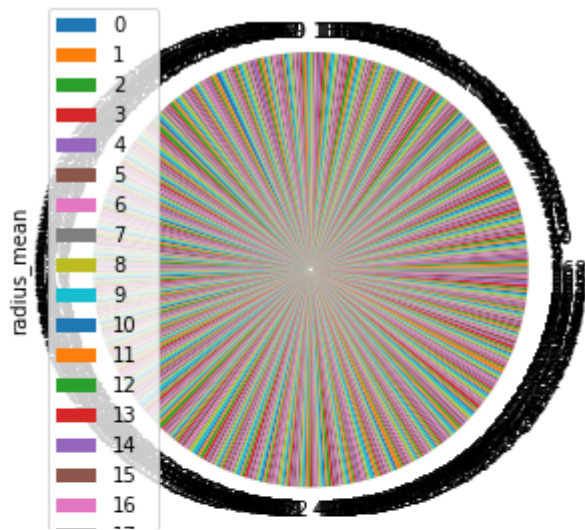


In [19]:

```
a.plot.pie(y='radius_mean',figsize=(5,5))
```

Out[19]:

<AxesSubplot:ylabel='radius_mean'>



In []: