

```
In [20]:
```

```
import numpy as np
import pandas as pd
from numpy import linalg as la
```

```
In [2]:
```

```
d=pd.read_csv(r"C:\Users\user\Downloads\8_BreastCancerPrediction - 8_BreastCancerPredic
d
```

```
Out[2]:
```

	<b>id</b>	<b>diagnosis</b>	<b>radius_mean</b>	<b>texture_mean</b>	<b>perimeter_mean</b>	<b>area_mean</b>	<b>smoothness_mean</b>	<b>compactness_mean</b>	<b>concavity_mean</b>	<b>concave points_mean</b>	<b>symmetry_mean</b>	<b>fractal_dimension_mean</b>	<b>radius_se</b>	<b>texture_se</b>	<b>perimeter_se</b>	<b>area_se</b>	<b>smoothness_se</b>	<b>compactness_se</b>	<b>concavity_se</b>	<b>concave points_se</b>	<b>symmetry_se</b>	<b>fractal_dimension_se</b>	<b>radius_worst</b>	<b>texture_worst</b>	<b>perimeter_worst</b>	<b>area_worst</b>	<b>smoothness_worst</b>	<b>compactness_worst</b>	<b>concavity_worst</b>	<b>concave points_worst</b>	<b>symmetry_worst</b>	<b>fractal_dimension_worst</b>
<b>0</b>	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
<b>1</b>	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
<b>2</b>	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
<b>3</b>	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000			
<b>4</b>	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000			
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...			
<b>564</b>	926424	M	21.56	22.39	142.00	1479.0	0.11100	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000			
<b>565</b>	926682	M	20.13	28.25	131.20	1261.0	0.09780	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000			
<b>566</b>	926954	M	16.60	28.08	108.30	858.1	0.08455	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000				
<b>567</b>	927241	M	20.60	29.33	140.10	1265.0	0.11780	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000				
<b>568</b>	92751	B	7.76	24.54	47.92	181.0	0.05263	0.12030	0.08510	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000				

569 rows × 32 columns



```
In [3]:
```

```
d.head()
```

```
Out[3]:
```

	<b>id</b>	<b>diagnosis</b>	<b>radius_mean</b>	<b>texture_mean</b>	<b>perimeter_mean</b>	<b>area_mean</b>	<b>smoothness_mean</b>	<b>cor</b>
<b>0</b>	842302	M	17.99	10.38	122.80	1001.0	0.11840	-0.00000
<b>1</b>	842517	M	20.57	17.77	132.90	1326.0	0.08474	-0.00000
<b>2</b>	84300903	M	19.69	21.25	130.00	1203.0	0.10960	-0.00000
<b>3</b>	84348301	M	11.42	20.38	77.58	386.1	0.14250	-0.00000
<b>4</b>	84358402	M	20.29	14.34	135.10	1297.0	0.10030	-0.00000

5 rows × 32 columns



```
In [4]:
```

```
d.tail()
```

Out[4]:

**id** **diagnosis** **radius\_mean** **texture\_mean** **perimeter\_mean** **area\_mean** **smoothness\_mean** **concave\_point\_mean**

<b>564</b>	926424	M	21.56	22.39	142.00	1479.0	0.11100
<b>565</b>	926682	M	20.13	28.25	131.20	1261.0	0.09780
<b>566</b>	926954	M	16.60	28.08	108.30	858.1	0.08455
<b>567</b>	927241	M	20.60	29.33	140.10	1265.0	0.11780
<b>568</b>	92751	B	7.76	24.54	47.92	181.0	0.05263

5 rows × 32 columns

In [5]:

d.describe()

Out[5]:

**id** **radius\_mean** **texture\_mean** **perimeter\_mean** **area\_mean** **smoothness\_mean** **com-**

<b>count</b>	5.690000e+02	569.000000	569.000000	569.000000	569.000000	569.000000
<b>mean</b>	3.037183e+07	14.127292	19.289649	91.969033	654.889104	0.096360
<b>std</b>	1.250206e+08	3.524049	4.301036	24.298981	351.914129	0.014064
<b>min</b>	8.670000e+03	6.981000	9.710000	43.790000	143.500000	0.052630
<b>25%</b>	8.692180e+05	11.700000	16.170000	75.170000	420.300000	0.086370
<b>50%</b>	9.060240e+05	13.370000	18.840000	86.240000	551.100000	0.095870
<b>75%</b>	8.813129e+06	15.780000	21.800000	104.100000	782.700000	0.105300
<b>max</b>	9.113205e+08	28.110000	39.280000	188.500000	2501.000000	0.163400

8 rows × 31 columns

In [6]:

`d.isna()`

Out[6]:

	<b>id</b>	<b>diagnosis</b>	<b>radius_mean</b>	<b>texture_mean</b>	<b>perimeter_mean</b>	<b>area_mean</b>	<b>smoothness_mean</b>	<b>comp</b>
<b>566</b>	False	False	False	False	False	False	False	False
<b>567</b>	False	False	False	False	False	False	False	False
<b>568</b>	False	False	False	False	False	False	False	False

569 rows × 32 columns

In [7]:

```
d.dropna()
```

Out[7]:

	<b>id</b>	<b>diagnosis</b>	<b>radius_mean</b>	<b>texture_mean</b>	<b>perimeter_mean</b>	<b>area_mean</b>	<b>smoothness_mean</b>	<b>comp</b>
<b>0</b>	842302	M	17.99	10.38	122.80	1001.0	0.11840	
<b>1</b>	842517	M	20.57	17.77	132.90	1326.0	0.08474	
<b>2</b>	84300903	M	19.69	21.25	130.00	1203.0	0.10960	
<b>3</b>	84348301	M	11.42	20.38	77.58	386.1	0.14250	
<b>4</b>	84358402	M	20.29	14.34	135.10	1297.0	0.10030	
...	...	...	...	...	...	...	...	...
<b>564</b>	926424	M	21.56	22.39	142.00	1479.0	0.11100	
<b>565</b>	926682	M	20.13	28.25	131.20	1261.0	0.09780	
<b>566</b>	926954	M	16.60	28.08	108.30	858.1	0.08455	
<b>567</b>	927241	M	20.60	29.33	140.10	1265.0	0.11780	
<b>568</b>	92751	B	7.76	24.54	47.92	181.0	0.05263	

569 rows × 32 columns

## Visualization

In [10]:

```
d=d[['perimeter_mean','area_mean']]
d
```

Out[10]:

	<b>perimeter_mean</b>	<b>area_mean</b>
<b>0</b>	122.80	1001.0
<b>1</b>	132.90	1326.0
<b>2</b>	130.00	1203.0
<b>3</b>	77.58	386.1
<b>4</b>	135.10	1297.0

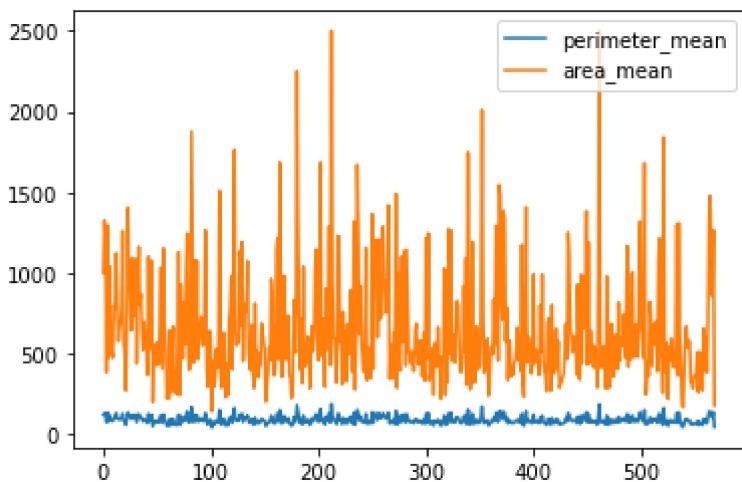
```
perimeter_mean  area_mean
```

	perimeter_mean	area_mean
...	...	...
564	142.00	1479.0
565	131.20	1261.0
566	108.30	858.1
567	140.10	1265.0
568	47.92	181.0

569 rows × 2 columns

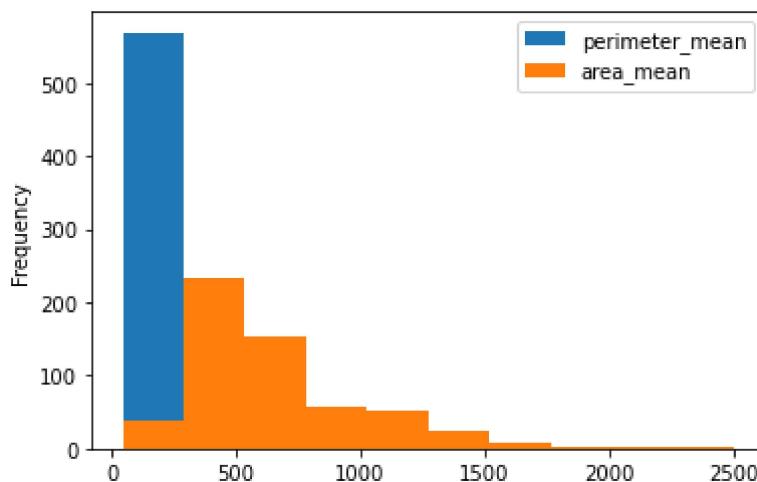
```
In [11]: d.plot.line()
```

```
Out[11]: <AxesSubplot:
```



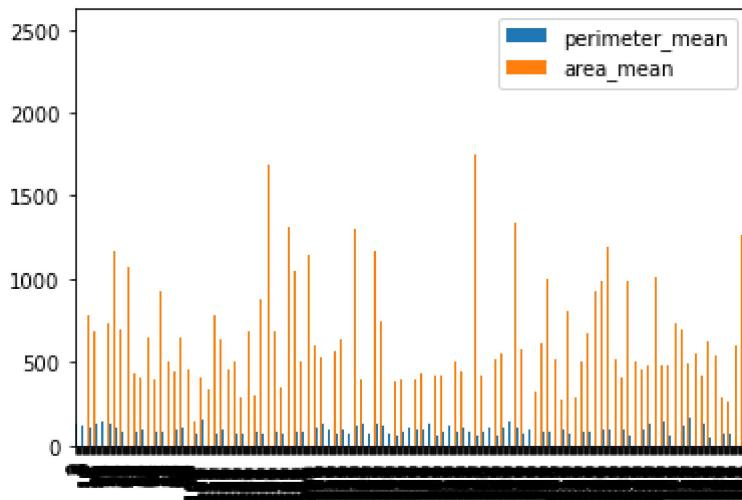
```
In [12]: d.plot.hist()
```

```
Out[12]: <AxesSubplot:ylabel='Frequency'>
```



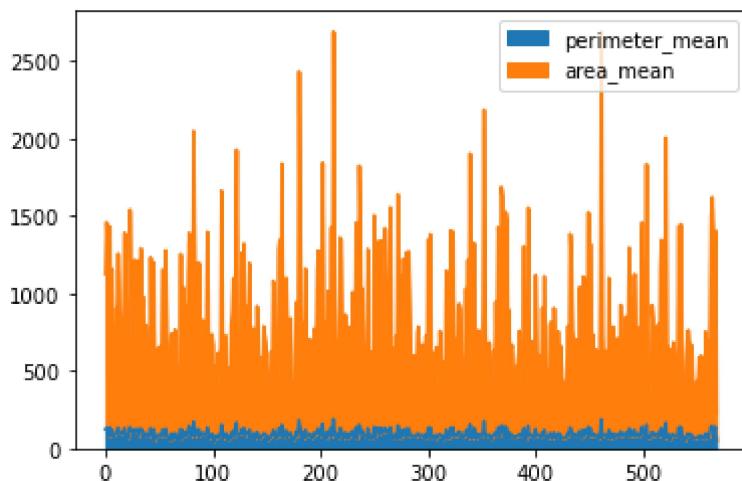
```
In [13]: d.plot.bar()
```

```
Out[13]: <AxesSubplot:>
```



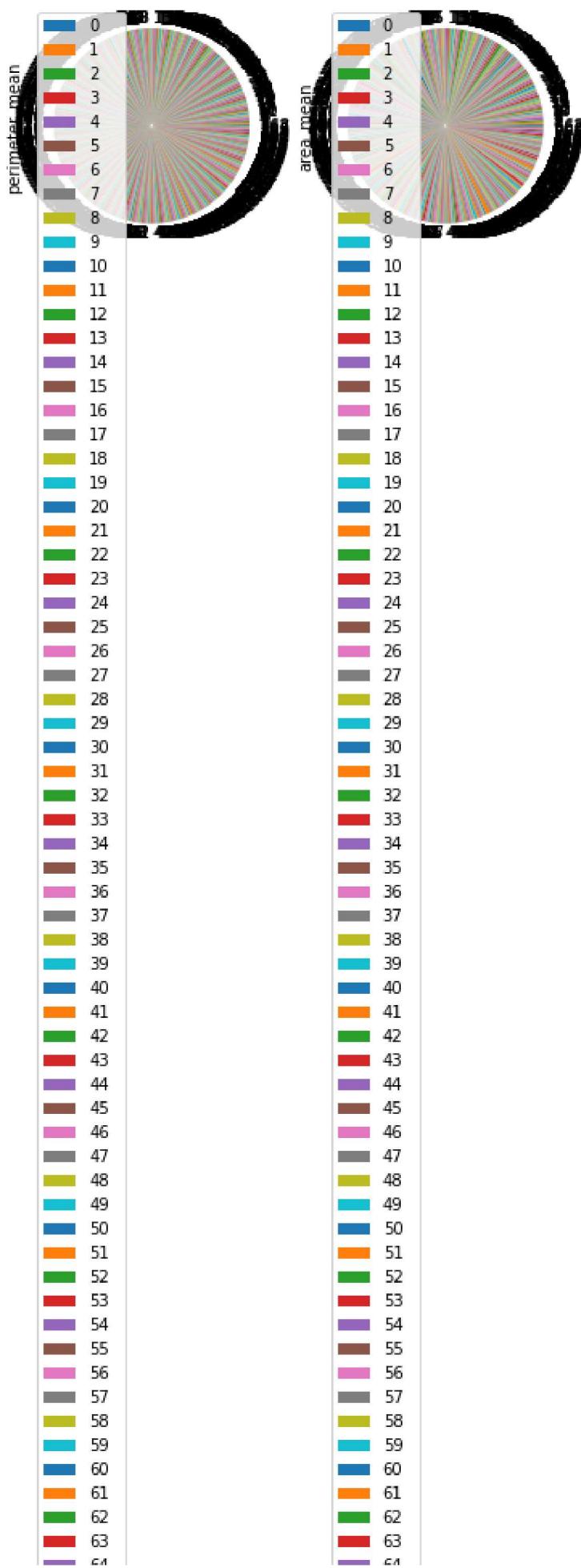
```
In [14]: d.plot.area()
```

```
Out[14]: <AxesSubplot:>
```

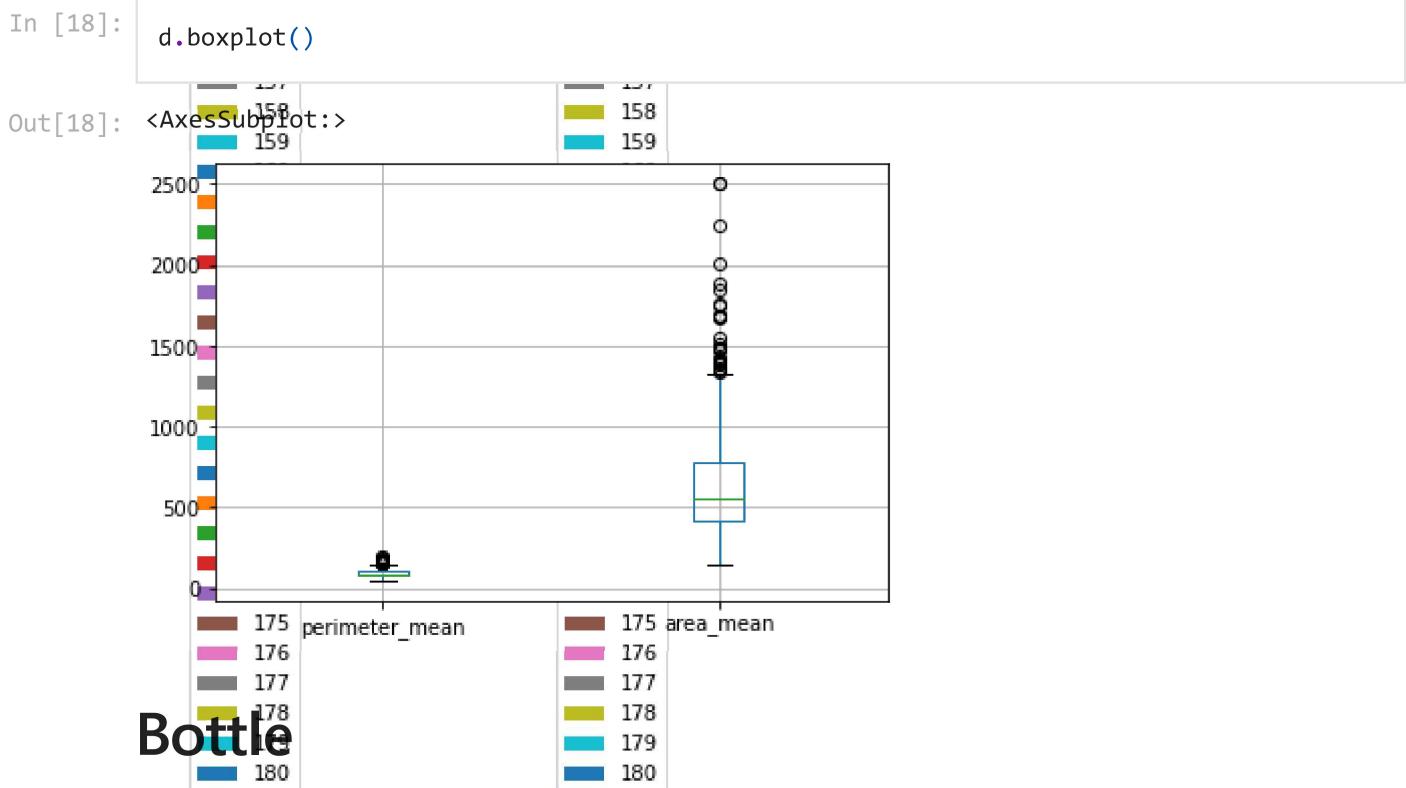
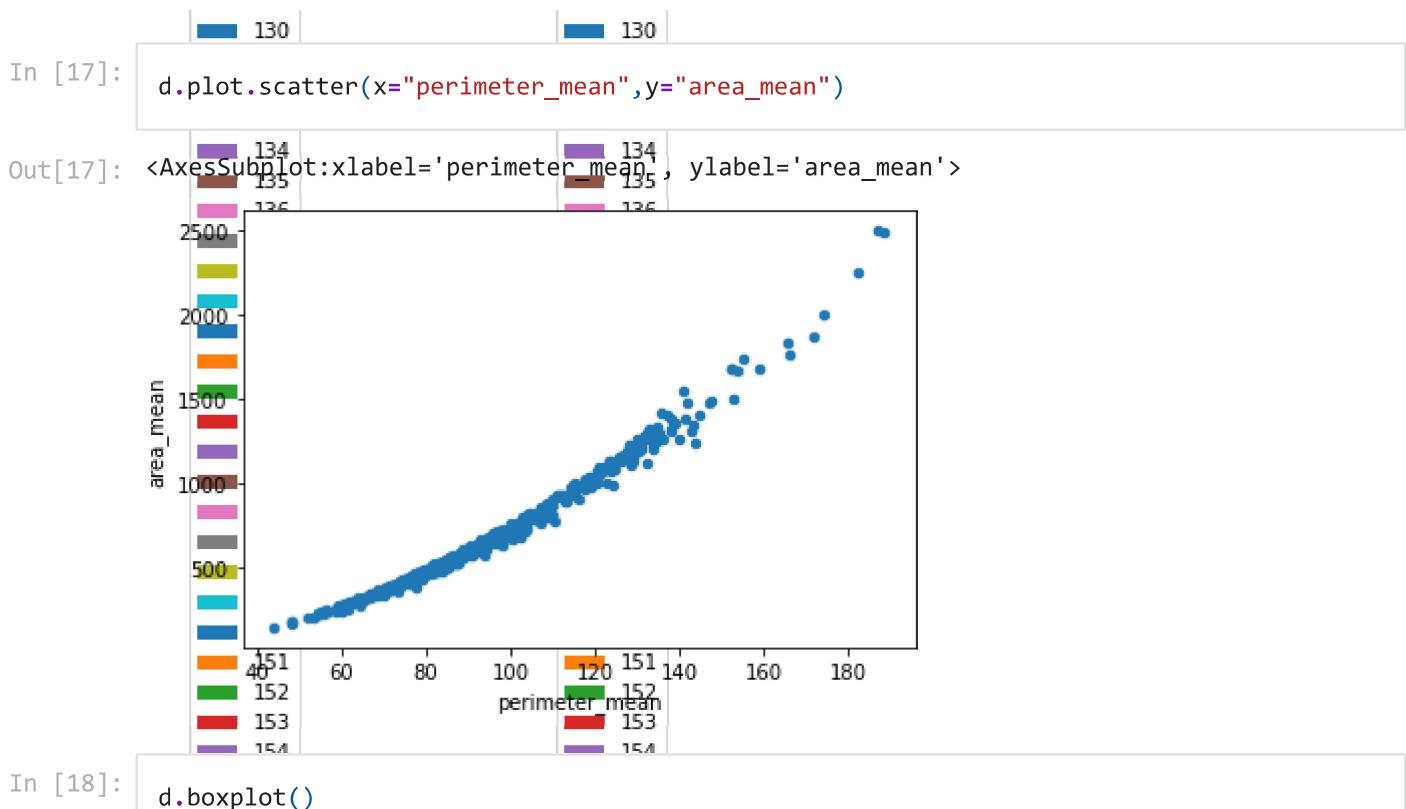


```
In [16]: d.plot.pie(subplots =True)
```

```
Out[16]: array([<AxesSubplot:ylabel='perimeter_mean'>,
   <AxesSubplot:ylabel='area_mean'>], dtype=object)
```



64	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100
101	101
102	102
103	103
104	104
105	105
106	106
107	107
108	108
109	109
110	110
111	111
112	112
113	113
114	114
115	115
116	116
117	117
118	118
119	119
120	120
121	121
122	122
123	123
124	124
125	125
126	126
127	127
128	128
129	129



Bottle

In [64]:

```
a=pd.read_csv(r"C:\Users\user\Downloads\9_bottle.csv")
a
```

C:\ProgramData\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3165: DtypeWarning: Columns (47,73) have mixed types. Specify dtype option on import or set low\_memory=False.

```
has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
190
191
192
193
194
```

Out[64]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat	R
	195			195							
	196			196							
	197			197							
	198			198							
<b>0</b>	1	1	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0000A-3	0	10.500	33.4400	NaN	25.64900	NaN	...
<b>1</b>	1	2	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0008A-3	8	10.460	33.4400	NaN	25.65600	NaN	...
<b>2</b>	1	3	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0010A-7	10	10.460	33.4370	NaN	25.65400	NaN	...
<b>3</b>	1	4	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0019A-3	19	10.450	33.4200	NaN	25.64300	NaN	...
<b>4</b>	1	5	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0020A-7	20	10.450	33.4210	NaN	25.64300	NaN	...
...	...	...	...	...	...	...	...	...	...	...	...
<b>864858</b>	34404	864859	093.4 026.4	20- 1611SR- MX-310- 2239- 09340264- 0000A-7	0	18.744	33.4083	5.805	23.87055	108.74	...
<b>864859</b>	34404	864860	093.4 026.4	20- 1611SR- MX-310- 2239- 09340264- 0002A-3	2	18.744	33.4083	5.805	23.87072	108.74	...
<b>864860</b>	34404	864861	093.4 026.4	20- 1611SR- MX-310- 2239- 09340264- 0005A-3	5	18.692	33.4150	5.796	23.88911	108.46	...
	259			259							
	760			760							

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat	R
	200	261	262	263	264						
864861	34404	864862	093.4 026.4	1611SR- MX-310- 2239- 09340264- 0010A-3	20-	10	18.161	33.4062	5.816	24.01426	107.74 ...
864862	34404	864863	093.4 026.4	1611SR- MX-310- 2239- 09340264- 0015A-3	20-	15	17.533	33.3880	5.774	24.15297	105.66 ...
864863	34404	864863	093.4 026.4	1611SR- MX-310- 2239- 09340264- 0015A-3	20-	15	17.533	33.3880	5.774	24.15297	105.66 ...
	279	280	281	282	283						

864863 rows × 74 columns

In [65]:

a.head()

Out[65]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat	R_PHAEO
	286	287	288	289	290						
0	1	1	054.0 056.0	4903CR- HY-060- 0930- 05400560- 0000A-3	19-	0	10.50	33.440	NaN	25.649	NaN ... NaN
1	1	2	054.0 056.0	4903CR- HY-060- 0930- 05400560- 0008A-3	19-	8	10.46	33.440	NaN	25.656	NaN ... NaN
2	1	3	054.0 056.0	4903CR- HY-060- 0930- 05400560- 0010A-7	19-	10	10.46	33.437	NaN	25.654	NaN ... NaN
3	1	4	054.0 056.0	4903CR- HY-060- 0930- 05400560- 0019A-3	19-	19	10.45	33.420	NaN	25.643	NaN ... NaN
4	1	5	054.0 056.0	4903CR- HY-060- 0930-	19-	20	10.45	33.421	NaN	25.643	NaN ... NaN

0	1	1	054.0 056.0	4903CR- HY-060- 0930- 05400560- 0000A-3	19-	0	10.50	33.440	NaN	25.649	NaN ... NaN
1	1	2	054.0 056.0	4903CR- HY-060- 0930- 05400560- 0008A-3	19-	8	10.46	33.440	NaN	25.656	NaN ... NaN
2	1	3	054.0 056.0	4903CR- HY-060- 0930- 05400560- 0010A-7	19-	10	10.46	33.437	NaN	25.654	NaN ... NaN
3	1	4	054.0 056.0	4903CR- HY-060- 0930- 05400560- 0019A-3	19-	19	10.45	33.420	NaN	25.643	NaN ... NaN
4	1	5	054.0 056.0	4903CR- HY-060- 0930-	19-	20	10.45	33.421	NaN	25.643	NaN ... NaN



	391	Cst_Cnt	Btl_Cnt	391	Depthm	T_degC	Salnty	O2ml_L
	392		392					
<b>count</b>	864863.000000	864863.000000	864863.000000	853900.000000	817509.000000	696201.000000	81217	2
<b>mean</b>	17138.790958	432432.000000	226.831951	10.799677	33.840350	3.392468	2	2
<b>std</b>	10240.949817	249664.587267	316.050259	4.243825	0.461843	2.073256	2	2
<b>min</b>	1.000000	1.000000	0.000000	1.440000	28.431000	-0.010000	2	2
<b>25%</b>	8269.000000	216216.500000	46.000000	7.680000	33.488000	1.360000	2	2
<b>50%</b>	16848.000000	432432.000000	125.000000	10.060000	33.863000	3.440000	2	2
<b>75%</b>	26557.000000	648647.500000	300.000000	13.880000	34.196900	5.500000	2	2
<b>max</b>	34404.000000	864863.000000	5351.000000	31.140000	37.034000	11.130000	25	25

In [68]:

a.isna()

Out[68]:

	416	417	418	419	416	417	418	419	Depth	ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat	R_PH
0	False	True	False	True	True	...	...										
1	False	True	False	True	True	True	...										
2	False	True	False	True	True	True	...										
3	False	True	False	True	True	True	...										
4	False	True	False	True	True	True	...										
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
864858	False	False	False	False	False	False											
864859	False	False	False	False	False	False											
864860	False	False	False	False	False	False											
864861	False	False	False	False	False	False											
864862	False	False	False	False	False	False											
864863	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	

Tn [69]:

a.dronna()

0 rows x 74 columns

450  
457

458  
459

460  
461

462  
463

464  
465

466  
467

468  
469

470  
471

472  
473

474  
475

476  
477

478  
479

480  
481

482  
483

484  
485

486  
487

488  
489

490  
491

492  
493

494  
495

496  
497

498  
499

500  
501

502  
503

504  
505

506  
507

508  
509

510  
511

512  
513

514  
515

516  
517

In [55]:

```
a=a[['Depthm','T_degC']].head(100)
```

a

Out[55]:

467  
468  
469

470  
471

472  
473

474  
475

476  
477

478  
479

480  
481

482  
483

484  
485

486  
487

488  
489

490  
491

492  
493

494  
495

496  
497

498  
499

500  
501

502  
503

504  
505

506  
507

508  
509

510  
511

512  
513

514  
515

516  
517

In [56]:

```
a.plot.line()
```

Out[56]:

497  
498  
499

500  
501

502  
503

504  
505

506  
507

508  
509

510  
511

512  
513

514  
515

516  
517

518  
519

520  
521

522  
523

524  
525

526  
527

528  
529

530  
531

532  
533

534  
535

536  
537

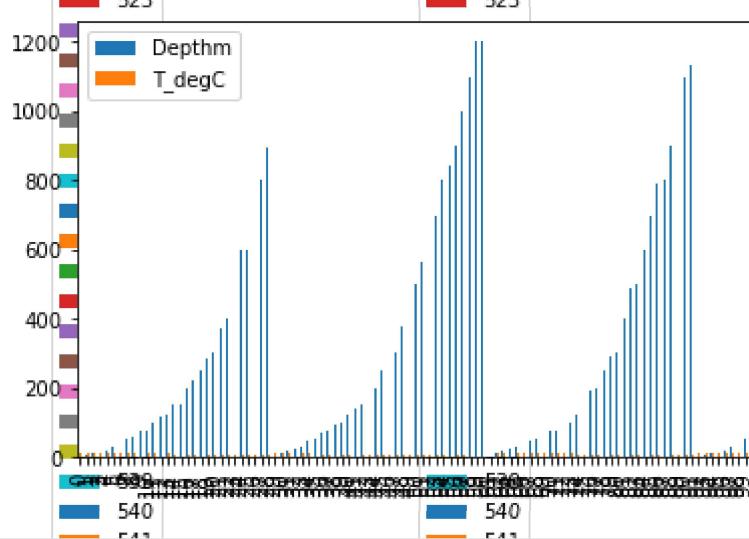
538  
539

540  
541

In [57]:

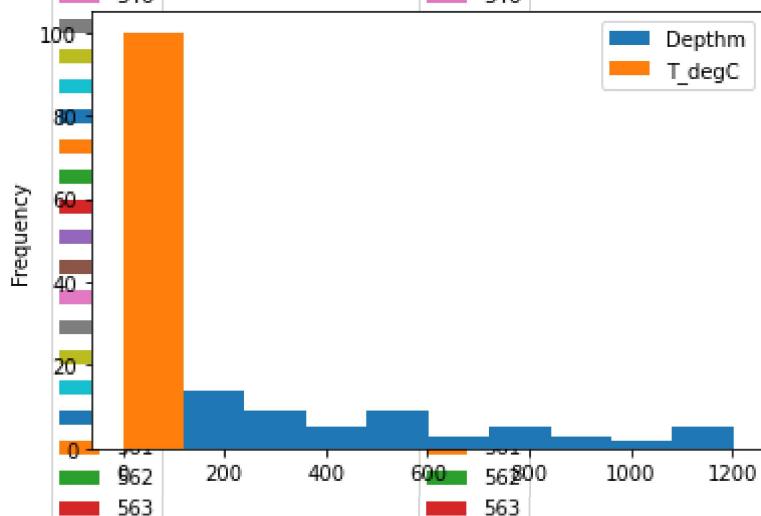
```
a.plot.bar()
```

Out[57]: <AxesSubplot: 522>



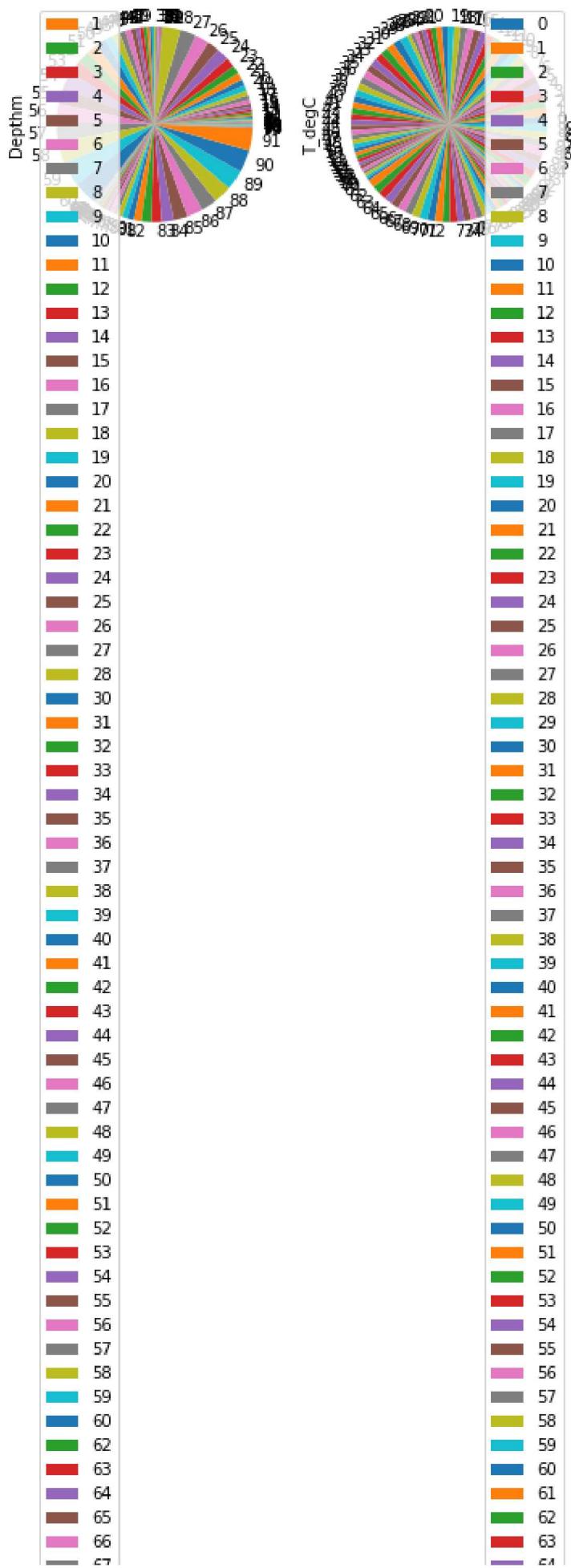
In [58]: `a.plot.hist()`

```
Out[58]: <AxesSubplot: xlabel='Frequency'> 545  
546 546
```



```
In [59]: a.plot.pie(subplots=True)
```

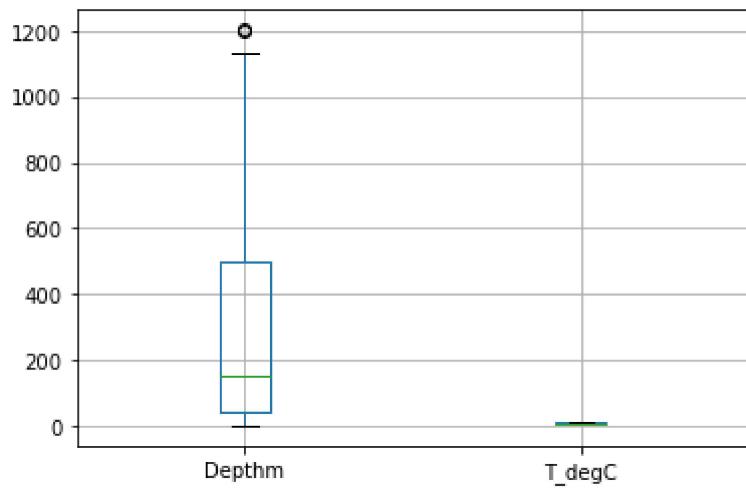
```
Out[59]: array([<AxesSubplot:ylabel='Depth'm>, <AxesSubplot:ylabel='T_degC'>],  
               dtype=object)
```





```
In [60]: a.boxplot()
```

```
Out[60]: <AxesSubplot:>
```



```
In [61]: a.plot.scatter(x="Depthm",y="T_degC")
```

```
Out[61]: <AxesSubplot:xlabel='Depthm', ylabel='T_degC'>
```

