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

# Pre-processing

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

In [ ]:
```
```

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

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 | 0.1 |
| **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 |

5 rows × 33 columns

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

Out[4]:

| | id | diagnosis | radius_mean | texture_mean | perimeter_mean | area_mean | smoothness_m |
|---|---|---|---|---|---|---|---|
| **564** | 926424 | M | 21.56 | 22.39 | 142.00 | 1479.0 | 0.1 |
| **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 | 0.1 |
| **568** | 92751 | B | 7.76 | 24.54 | 47.92 | 181.0 | 0.05 |

5 rows × 33 columns

In [5]:
```
data.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 | 569.0000 |
| **mean** | 3.037183e+07 | 14.127292 | 19.289649 | 91.969033 | 654.889104 | 0.0963 |
| **std** | 1.250206e+08 | 3.524049 | 4.301036 | 24.298981 | 351.914129 | 0.0140 |
| **min** | 8.670000e+03 | 6.981000 | 9.710000 | 43.790000 | 143.500000 | 0.0526 |
| **25%** | 8.692180e+05 | 11.700000 | 16.170000 | 75.170000 | 420.300000 | 0.0863 |
| **50%** | 9.060240e+05 | 13.370000 | 18.840000 | 86.240000 | 551.100000 | 0.0958 |
| **75%** | 8.813129e+06 | 15.780000 | 21.800000 | 104.100000 | 782.700000 | 0.1053 |
| **max** | 9.113205e+08 | 28.110000 | 39.280000 | 188.500000 | 2501.000000 | 0.1634 |

8 rows × 32 columns

In [6]:
```
print(np.shape(data))
```

```
(569, 33)
```

In [7]:
```
print(np.size(data))
```

```
18777
```

In [8]:
```
data.isnull()
```

Out[8]:

| ıre_mean | perimeter_mean | area_mean | smoothness_mean | compactness_mean | concavity_mean | p |
|---|---|---|---|---|---|---|
| False | False | False | False | False | False | |
| False | False | False | False | False | False | |
| False | False | False | False | False | False | |
| False | False | False | False | False | False | |
| False | False | False | False | False | False | |
| ... | ... | ... | ... | ... | ... | |
| False | False | False | False | False | False | |
| False | False | False | False | False | False | |
| False | False | False | False | False | False | |
| False | False | False | False | False | False | |
| False | False | False | False | False | False | |

In [9]:
```
data.fillna(value=0)
```

Out[9]:

| | 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

# Visualization

In [12]:
```
data=data[['texture_mean','area_mean']]
data
```

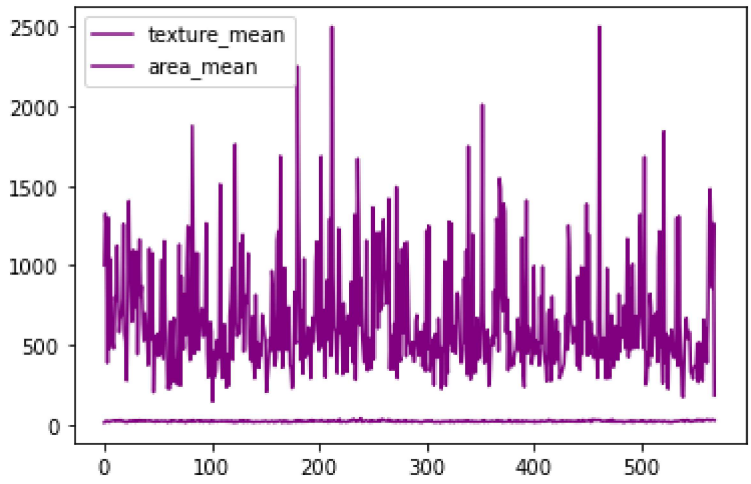
Out[12]:

| | texture_mean | area_mean |
|---|---|---|
| **0** | 10.38 | 1001.0 |
| **1** | 17.77 | 1326.0 |
| **2** | 21.25 | 1203.0 |
| **3** | 20.38 | 386.1 |
| **4** | 14.34 | 1297.0 |
| **...** | ... | ... |
| **564** | 22.39 | 1479.0 |
| **565** | 28.25 | 1261.0 |
| **566** | 28.08 | 858.1 |
| **567** | 29.33 | 1265.0 |
| **568** | 24.54 | 181.0 |

569 rows × 2 columns

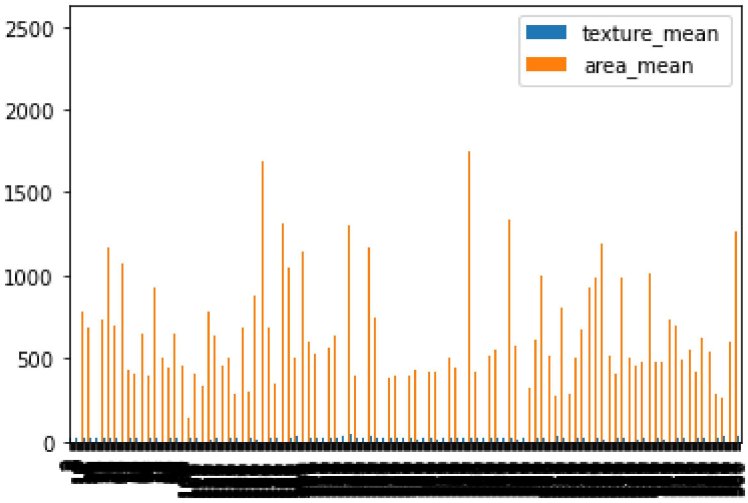
In [14]:
```
data.plot.line(color='purple')
```

Out[14]: <AxesSubplot:>



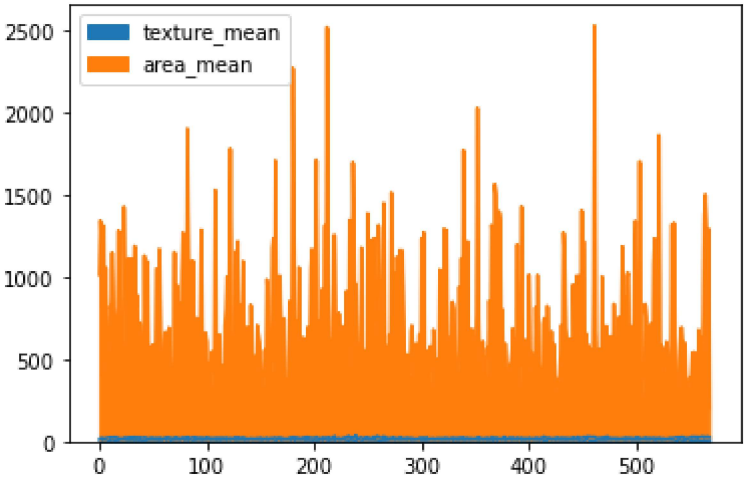
In [15]:
```
data.plot.bar()
```

Out[15]: <AxesSubplot:>



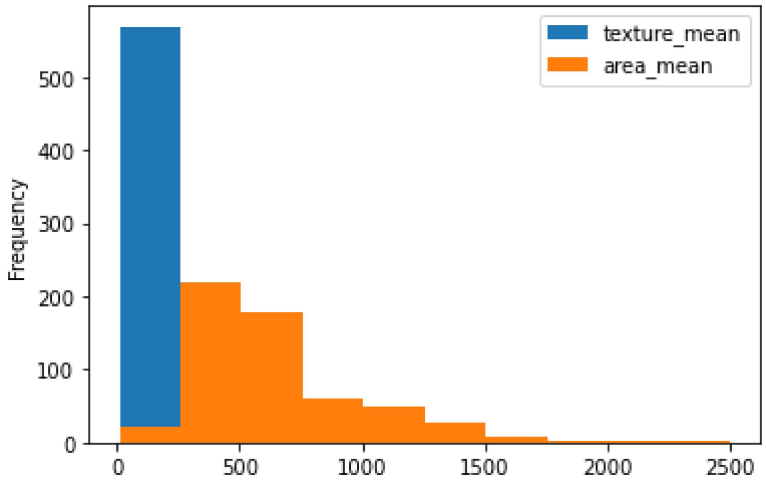

In [16]:
```
data.plot.area()
```

Out[16]: <AxesSubplot:>



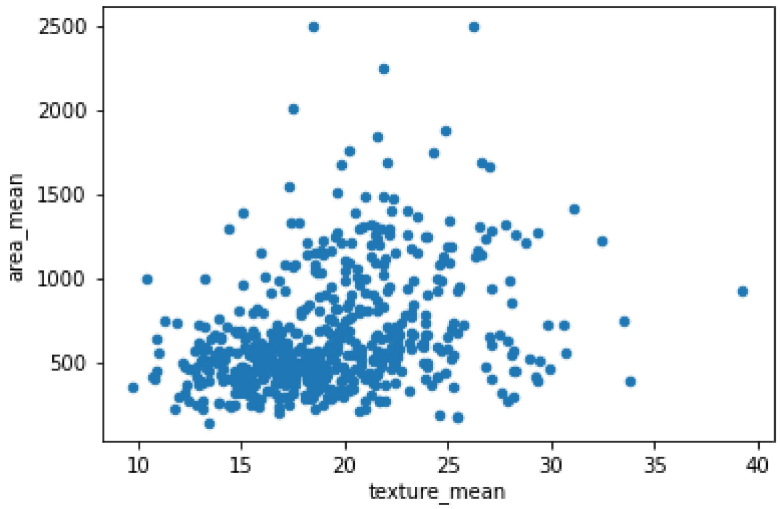
In [17]:
```
data.plot.hist()
```

Out[17]:
```
<AxesSubplot:ylabel='Frequency'>
```



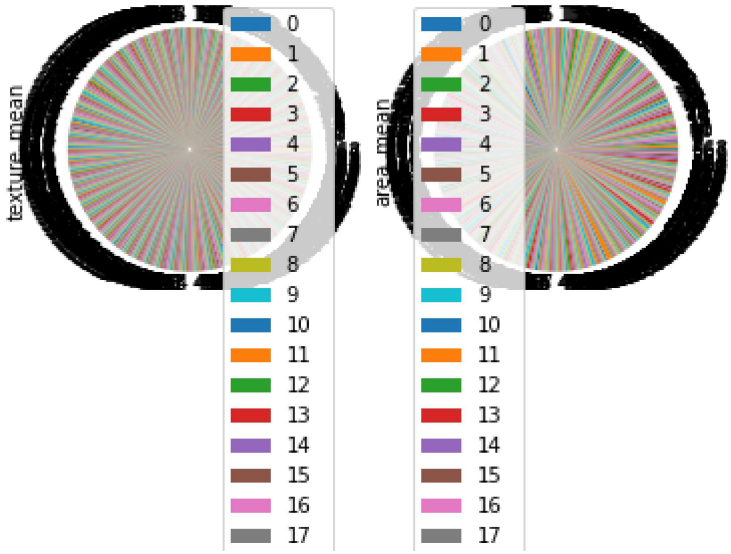

In [20]:
```
data.plot.scatter(x='texture_mean',y='area_mean')
```

Out[20]:
```
<AxesSubplot:xlabel='texture_mean', ylabel='area_mean'>
```



In [21]:
```
data.plot.pie(subplots=True)
```

Out[21]:
```
array([<AxesSubplot:ylabel='texture_mean'>,
       <AxesSubplot:ylabel='area_mean'>], dtype=object)
```




In [ ]: