# Basic Analysis using numpy and pandas

# Fitness dataset

To import library

In [1]:

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

To import dataset

In [2]:

```
d=pd.read_csv(r"C:\Users\user\Downloads\3_Fitness-1.csv")
d
```

Out[2]:

| | Row Labels | Sum of Jan | Sum of Feb | Sum of Mar | Sum of Total Sales |
|---|---|---|---|---|---|
| **0** | A | 5.62% | 7.73% | 6.16% | 75 |
| **1** | B | 4.21% | 17.27% | 19.21% | 160 |
| **2** | C | 9.83% | 11.60% | 5.17% | 101 |
| **3** | D | 2.81% | 21.91% | 7.88% | 127 |
| **4** | E | 25.28% | 10.57% | 11.82% | 179 |
| **5** | F | 8.15% | 16.24% | 18.47% | 167 |
| **6** | G | 18.54% | 8.76% | 17.49% | 171 |
| **7** | H | 25.56% | 5.93% | 13.79% | 170 |
| **8** | Grand Total | 100.00% | 100.00% | 100.00% | 1150 |

To get top 10 record

In [3]:

```
d.head(10)
```

Out[3]:

| | Row Labels | Sum of Jan | Sum of Feb | Sum of Mar | Sum of Total Sales |
|---|---|---|---|---|---|
| **0** | A | 5.62% | 7.73% | 6.16% | 75 |
| **1** | B | 4.21% | 17.27% | 19.21% | 160 |
| **2** | C | 9.83% | 11.60% | 5.17% | 101 |
| **3** | D | 2.81% | 21.91% | 7.88% | 127 |
| **4** | E | 25.28% | 10.57% | 11.82% | 179 |
| **5** | F | 8.15% | 16.24% | 18.47% | 167 |
| **6** | G | 18.54% | 8.76% | 17.49% | 171 |
| **7** | H | 25.56% | 5.93% | 13.79% | 170 |
| **8** | Grand Total | 100.00% | 100.00% | 100.00% | 1150 |

To get last 10

In [4]:

```
d.tail(10)
```

Out[4]:

| | Row Labels | Sum of Jan | Sum of Feb | Sum of Mar | Sum of Total Sales |
|---|---|---|---|---|---|
| **0** | A | 5.62% | 7.73% | 6.16% | 75 |
| **1** | B | 4.21% | 17.27% | 19.21% | 160 |
| **2** | C | 9.83% | 11.60% | 5.17% | 101 |
| **3** | D | 2.81% | 21.91% | 7.88% | 127 |
| **4** | E | 25.28% | 10.57% | 11.82% | 179 |
| **5** | F | 8.15% | 16.24% | 18.47% | 167 |
| **6** | G | 18.54% | 8.76% | 17.49% | 171 |
| **7** | H | 25.56% | 5.93% | 13.79% | 170 |
| **8** | Grand Total | 100.00% | 100.00% | 100.00% | 1150 |

To describe statistics Analysis

In [5]:

```
d.describe()
```

Out[5]:

| | Sum of Total Sales |
|---|---|
| **count** | 9.000000 |
| **mean** | 255.555556 |
| **std** | 337.332963 |
| **min** | 75.000000 |
| **25%** | 127.000000 |
| **50%** | 167.000000 |
| **75%** | 171.000000 |
| **max** | 1150.000000 |

To get rows and columns

In [6]:

```
np.shape(d)
```

Out[6]:

```
(9, 5)
```

To get number of elements

In [7]:

```
np.size(d)
```

Out[7]:

```
45
```

To get the missing value

In [8]:

```
d.isna()
```

Out[8]:

| | Row Labels | Sum of Jan | Sum of Feb | Sum of Mar | Sum of Total Sales |
|---|---|---|---|---|---|
| **0** | False | False | False | False | False |
| **1** | False | False | False | False | False |
| **2** | False | False | False | False | False |
| **3** | False | False | False | False | False |
| **4** | False | False | False | False | False |
| **5** | False | False | False | False | False |
| **6** | False | False | False | False | False |
| **7** | False | False | False | False | False |
| **8** | False | False | False | False | False |

To drop the missing elements

In [9]:

```
d.dropna(axis=1,how='any')
```

Out[9]:

| | Row Labels | Sum of Jan | Sum of Feb | Sum of Mar | Sum of Total Sales |
|---|---|---|---|---|---|
| **0** | A | 5.62% | 7.73% | 6.16% | 75 |
| **1** | B | 4.21% | 17.27% | 19.21% | 160 |
| **2** | C | 9.83% | 11.60% | 5.17% | 101 |
| **3** | D | 2.81% | 21.91% | 7.88% | 127 |
| **4** | E | 25.28% | 10.57% | 11.82% | 179 |
| **5** | F | 8.15% | 16.24% | 18.47% | 167 |
| **6** | G | 18.54% | 8.76% | 17.49% | 171 |
| **7** | H | 25.56% | 5.93% | 13.79% | 170 |
| **8** | Grand Total | 100.00% | 100.00% | 100.00% | 1150 |

In [10]:

```
d["Row Labels"]
```

Out[10]:

```
0              A
1              B
2              C
3              D
4              E
5              F
6              G
7              H
8    Grand Total
Name: Row Labels, dtype: object
```

In [13]:

```
data=d[['Sum of Mar','Sum of Total Sales']]
data
```

Out[13]:

| | Sum of Mar | Sum of Total Sales |
|---|---|---|
| **0** | 6.16% | 75 |
| **1** | 19.21% | 160 |
| **2** | 5.17% | 101 |
| **3** | 7.88% | 127 |
| **4** | 11.82% | 179 |
| **5** | 18.47% | 167 |
| **6** | 17.49% | 171 |
| **7** | 13.79% | 170 |
| **8** | 100.00% | 1150 |

In [14]:

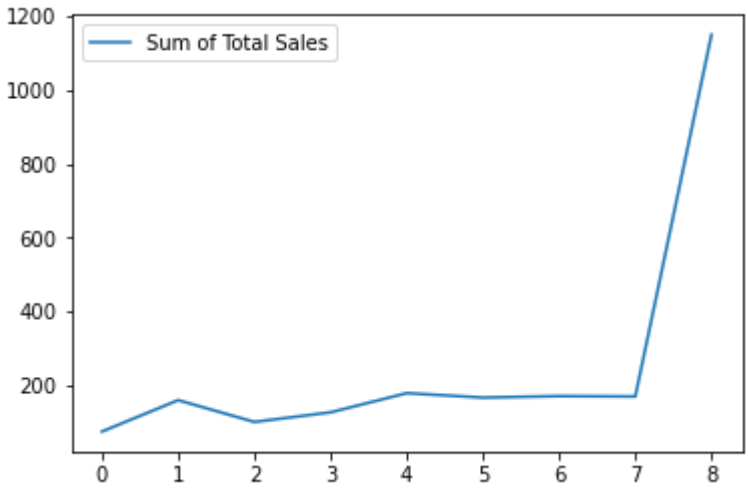
```
import matplotlib.pyplot as pp
```

In [15]:

```
data.plot.line()
```

Out[15]:

<AxesSubplot:>

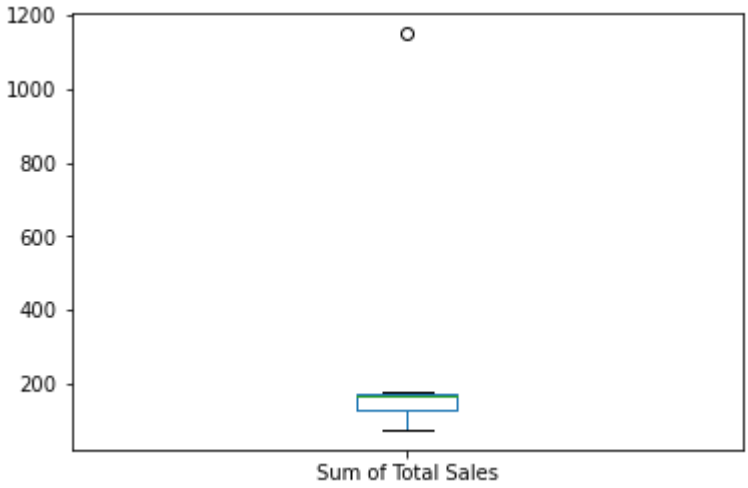



In [16]:

```
data.plot.box()
```

Out[16]:

<AxesSubplot:>

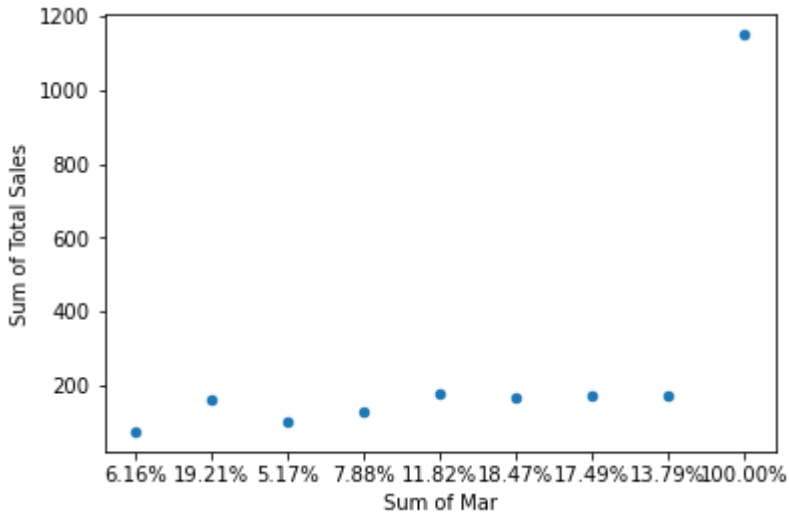


In [18]:

```
data.plot.scatter(x="Sum of Mar",y="Sum of Total Sales")
```

Out[18]:

```
<AxesSubplot:xlabel='Sum of Mar', ylabel='Sum of Total Sales'>
```

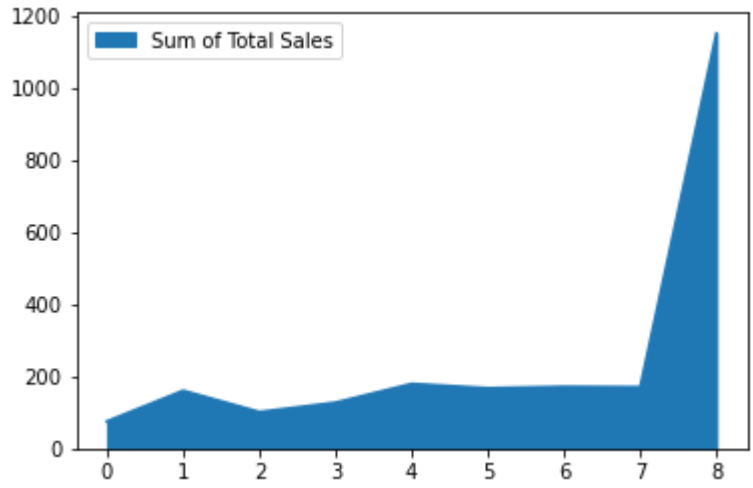



In [19]:

```
data.plot.area()
```

Out[19]:

```
<AxesSubplot:>
```

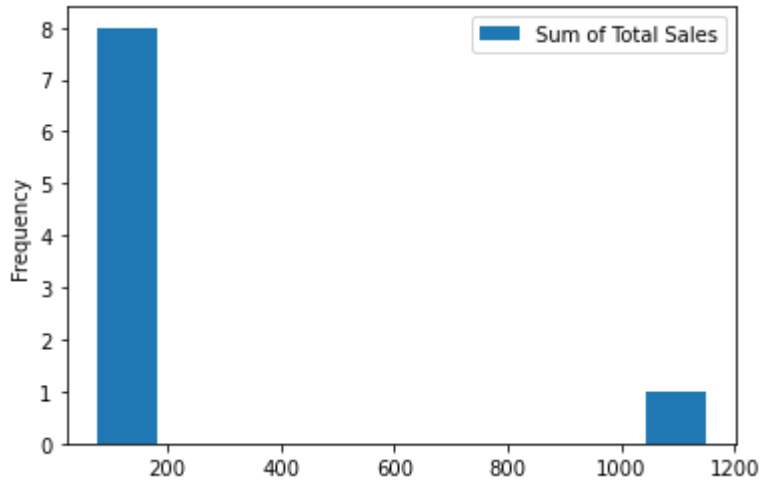


In [20]:

```
data.plot.hist()
```

Out[20]:

```
<AxesSubplot:ylabel='Frequency'>
```

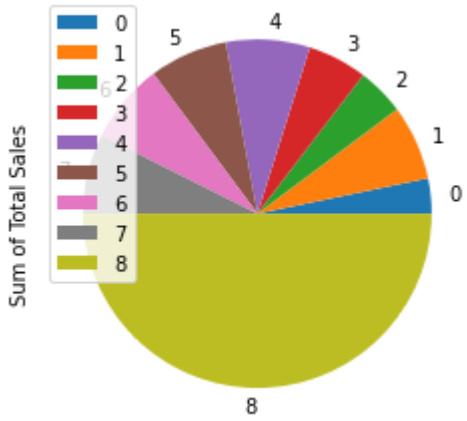



In [21]:

```
d.plot.pie(y="Sum of Total Sales")
```

Out[21]:

```
<AxesSubplot:ylabel='Sum of Total Sales'>
```




In [ ]:

```

```