Basic Analysis using numpy and pandas

Fitness Dataset

To import library

In [1]:

```python
import numpy as np
```

In [2]:

```python
import pandas as pd
```

To import dataset

In [3]:

```python
d=pd.read_csv(r"C:\Users\user\Downloads\fit.csv")
d
```

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 Top 10 record

In [4]:

```
d.head(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 get last record

In [5]:

```
d.tail(20)
```

Out[5]:

	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

Statistics Analysis

In [6]:

```
d.describe()
```

Out[6]:

	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 row and column

In [7]:

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

Out[7]:

```
(9, 5)
```

Find Number of Elements

In [8]:

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

Out[8]:

```
45
```

Find Missing Value

In [9]:

```
d.isna()
```

Out[9]:

	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 value

In [10]:

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

Out[10]:

	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 [11]:

```python
d["Sum of Total Sales"]
```

Out[11]:

```
0        75
1       160
2       101
3       127
4       179
5       167
6       171
7       170
8      1150
Name: Sum of Total Sales, dtype: int64
```

In [12]:

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

Out[12]:

	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 [13]:

```python
import matplotlib.pyplot as pp
```
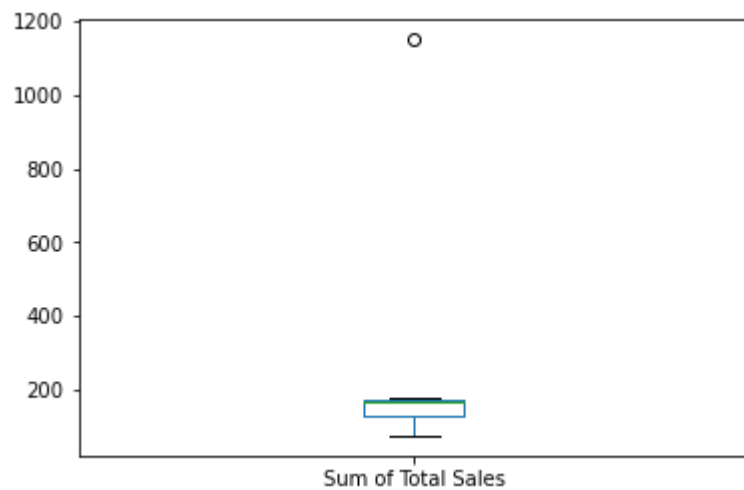
In [14]:

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

Out[14]:

```
<AxesSubplot:>
```
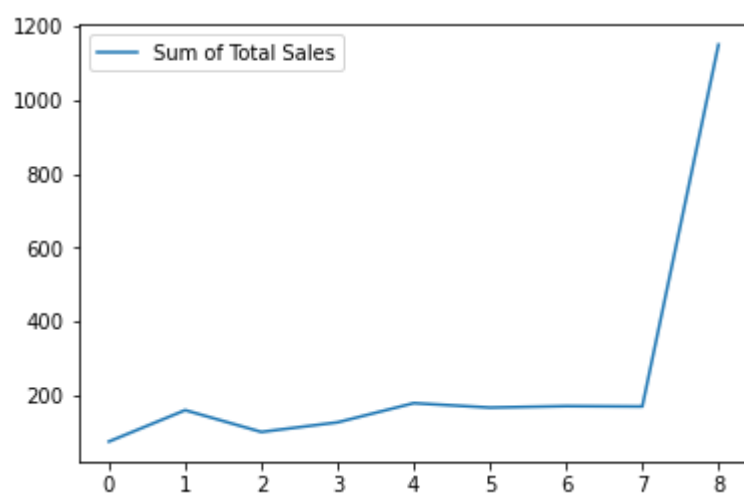

In [15]:

```
data.plot.box()
```
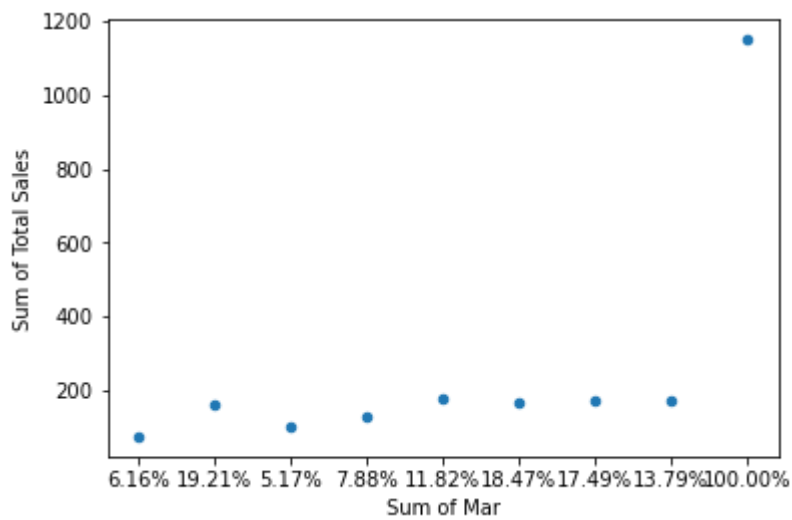
Out[15]:

```
<AxesSubplot:>
```

In [16]:

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

Out[16]:

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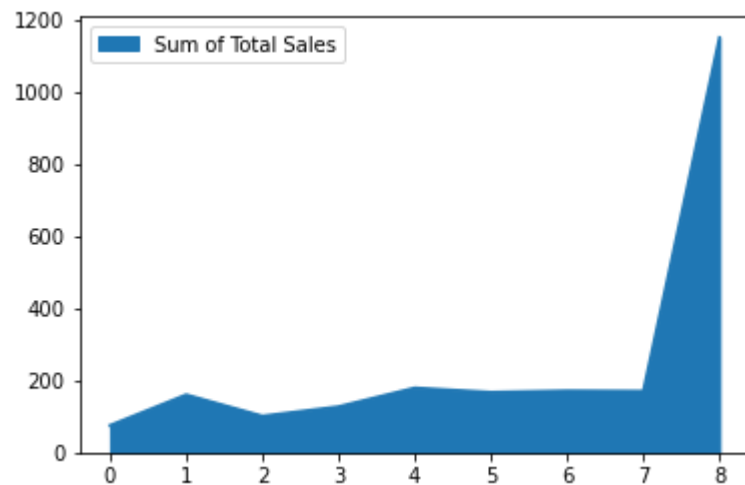

In [17]:

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

Out[17]:

```
<AxesSubplot:>
```

In [18]:

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

Out[18]:

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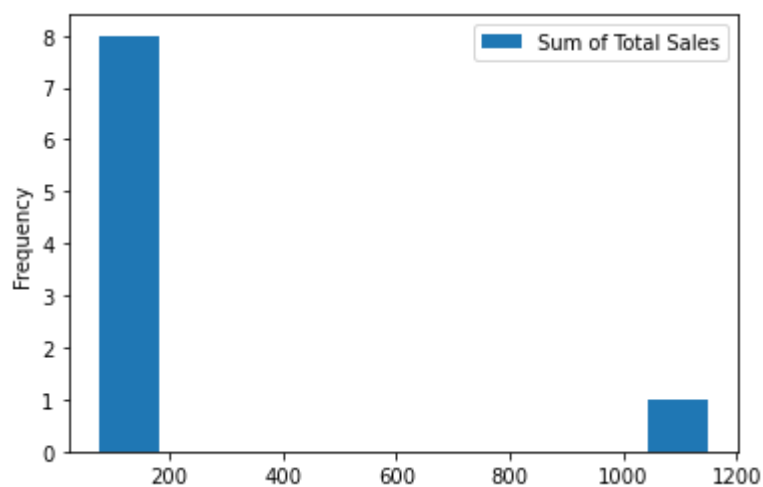

In [19]:

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

Out[19]:

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