Deena 20104016

Basic Analysis using Numpy and Pandas

Import Libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Importing Dataset

In [2]:
```
df=pd.read_csv("3_Fitness-1.csv")
df
```

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 display first 10 rows

In [3]:
```
df.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

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
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 display last 5 rows

In [4]:
```
df.tail(5)
```

Out[4]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
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

Satistical Summary

In [5]:
```
df.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 find shape and size

In [6]:
```
df.shape
```

Out[6]:
```
(9, 5)
```

In [7]:
```
df.size
```

Out[7]: 45

To fill the null values

In [8]:
```
df.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 fill missing values

In [9]:
```
df.dropna()
```

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

coloumns

In [10]:
```
df.columns
```

Out[10]:
```
Index(['Row Labels', 'Sum of Jan', 'Sum of Feb', 'Sum of Mar',
       'Sum of Total Sales'],
      dtype='object')
```

to print a particular coloumn

In [11]:
```
data=df[["Sum of Jan","Sum of Total Sales"]]
data
```

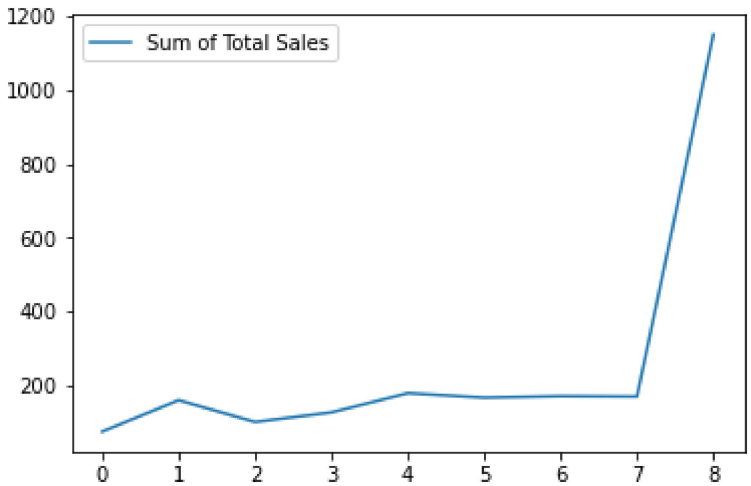
Out[11]:

	Sum of Jan	Sum of Total Sales
0	5.62%	75
1	4.21%	160
2	9.83%	101
3	2.81%	127
4	25.28%	179
5	8.15%	167
6	18.54%	171
7	25.56%	170
8	100.00%	1150

line plot

In [12]:
```
data.plot.line()
```

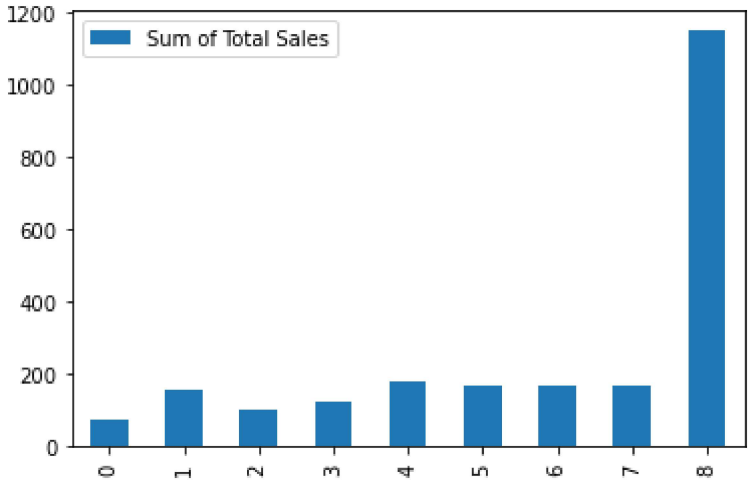
Out[12]:
```
<AxesSubplot:>
```



bar plot

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

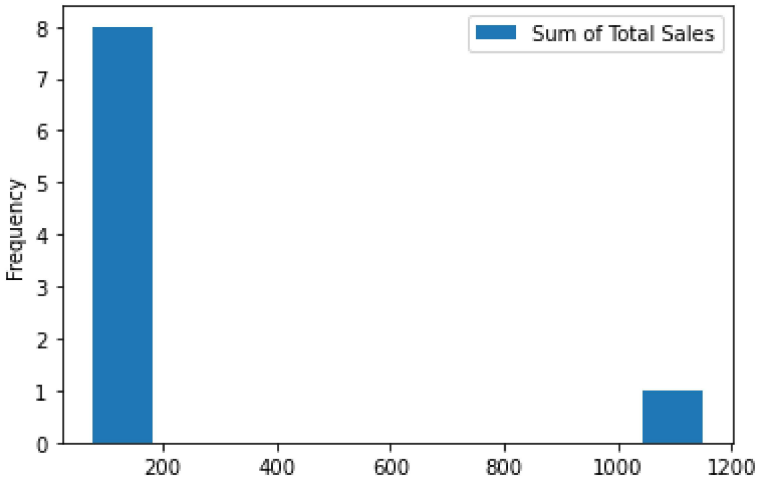
Out[13]: <AxesSubplot:>



hist plot

In [14]:
```
data.plot.hist()
```

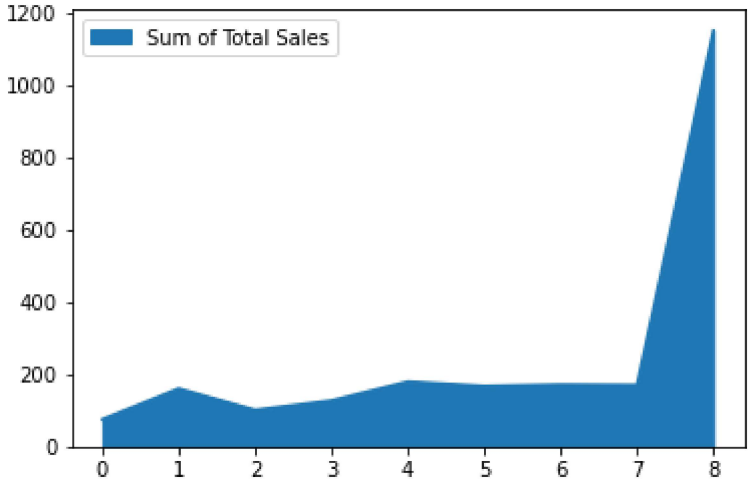
Out[14]: <AxesSubplot:ylabel='Frequency'>



Area plot

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

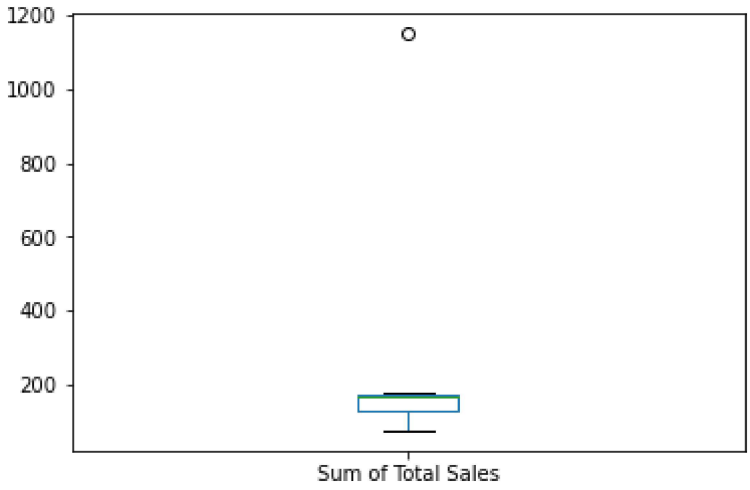
Out[15]: <AxesSubplot:>



Box plot

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

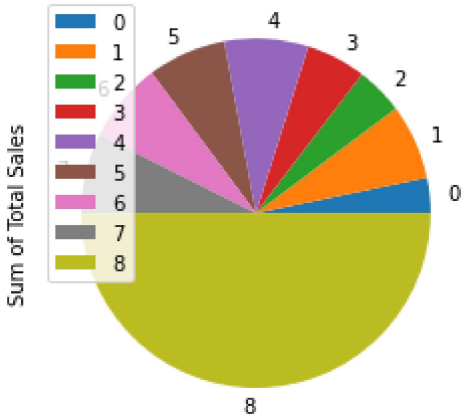
Out[16]: <AxesSubplot:>



pie plot

In [18]:
```
data.plot.pie(y='Sum of Total Sales')
```

Out[18]: <AxesSubplot:ylabel='Sum of Total Sales'>



In [19]:
```
data.plot.scatter(x="Sum of Jan",y="Sum of Total Sales")
```

Out[19]: <AxesSubplot:xlabel='Sum of Jan', ylabel='Sum of Total Sales'>