Kaviyadevi M 20106064

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

```
In [2]: data=pd.read_csv(r"C:\Users\user\Downloads\6_Salesworkload1 - 6_Salesworkload1.cs
        data
```

Out[2]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLeas
0	10.2016	1.0	United Kingdom	88253.0	London (I)	1.0	Dry	3184.764	0.
1	10.2016	1.0	United Kingdom	88253.0	London (I)	2.0	Frozen	1582.941	0.
2	10.2016	1.0	United Kingdom	88253.0	London (I)	3.0	other	47.205	0.
3	10.2016	1.0	United Kingdom	88253.0	London (I)	4.0	Fish	1623.852	0.
4	10.2016	1.0	United Kingdom	88253.0	London (I)	5.0	Fruits & Vegetables	1759.173	0.
...	...	...	...	...	...	...	...	...	.
7653	6.2017	9.0	Sweden	29650.0	Gothenburg	12.0	Checkout	6322.323	0.
7654	6.2017	9.0	Sweden	29650.0	Gothenburg	16.0	Customer Services	4270.479	0.
7655	6.2017	9.0	Sweden	29650.0	Gothenburg	11.0	Delivery	0	0.
7656	6.2017	9.0	Sweden	29650.0	Gothenburg	17.0	others	2224.929	0.
7657	6.2017	9.0	Sweden	29650.0	Gothenburg	18.0	all	39652.2	0.

7658 rows × 14 columns

Data Preprocessing

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

Out[3]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLease	S
0	10.2016	1.0	United Kingdom	88253.0	London (I)	1.0	Dry	3184.764	0.0	3985
1	10.2016	1.0	United Kingdom	88253.0	London (I)	2.0	Frozen	1582.941	0.0	827
2	10.2016	1.0	United Kingdom	88253.0	London (I)	3.0	other	47.205	0.0	4384
3	10.2016	1.0	United Kingdom	88253.0	London (I)	4.0	Fish	1623.852	0.0	3094
4	10.2016	1.0	United Kingdom	88253.0	London (I)	5.0	Fruits & Vegetables	1759.173	0.0	1655

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

Out[4]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLease
7653	6.2017	9.0	Sweden	29650.0	Gothenburg	12.0	Checkout	6322.323	0.0
7654	6.2017	9.0	Sweden	29650.0	Gothenburg	16.0	Customer Services	4270.479	0.0
7655	6.2017	9.0	Sweden	29650.0	Gothenburg	11.0	Delivery	0	0.0
7656	6.2017	9.0	Sweden	29650.0	Gothenburg	17.0	others	2224.929	0.0
7657	6.2017	9.0	Sweden	29650.0	Gothenburg	18.0	all	39652.2	0.0

```
In [12]: data.describe()
```

Out[12]:

	Time index	StoreID	Dept_ID	HoursLease	Sales units	Turnover	Custome
count	7650.000000	7650.000000	7650.000000	7650.000000	7.650000e+03	7.650000e+03	0.
mean	5.000000	61995.220000	9.470588	22.036078	1.076471e+06	3.721393e+06	Nal
std	2.582158	29924.581631	5.337429	133.299513	1.728113e+06	6.003380e+06	Nal
min	1.000000	12227.000000	1.000000	0.000000	0.000000e+00	0.000000e+00	Nal
25%	3.000000	29650.000000	5.000000	0.000000	5.457125e+04	2.726798e+05	Nal
50%	5.000000	75400.500000	9.000000	0.000000	2.932300e+05	9.319575e+05	Nal
75%	7.000000	87703.000000	14.000000	0.000000	9.175075e+05	3.264432e+06	Nal
max	9.000000	98422.000000	18.000000	3984.000000	1.124296e+07	4.271739e+07	Nal

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

         (7658, 14)
```

```
In [8]:  print(np.size(data))

         107212
```

Data Visualization

```
In [13]:  data=data[['StoreID','Dept_ID']]
          data
```
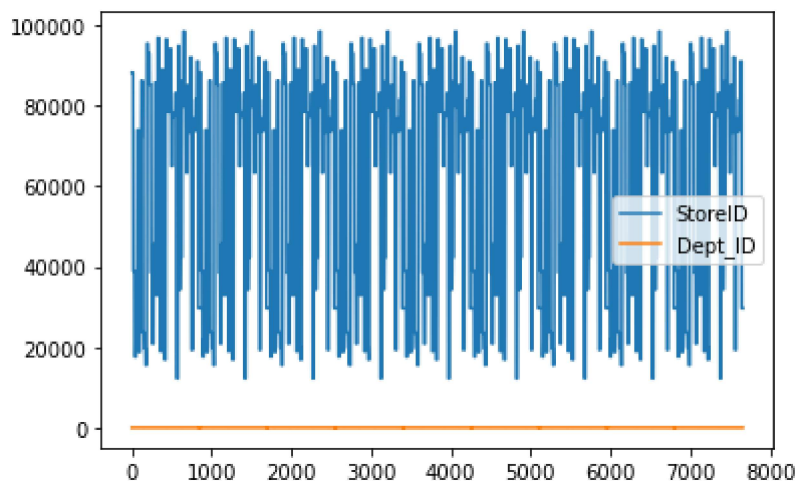
Out[13]:

	StoreID	Dept_ID
0	88253.0	1.0
1	88253.0	2.0
2	88253.0	3.0
3	88253.0	4.0
4	88253.0	5.0
...	...	...
7653	29650.0	12.0
7654	29650.0	16.0
7655	29650.0	11.0
7656	29650.0	17.0
7657	29650.0	18.0

7658 rows × 2 columns

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

Out[14]: <AxesSubplot:>
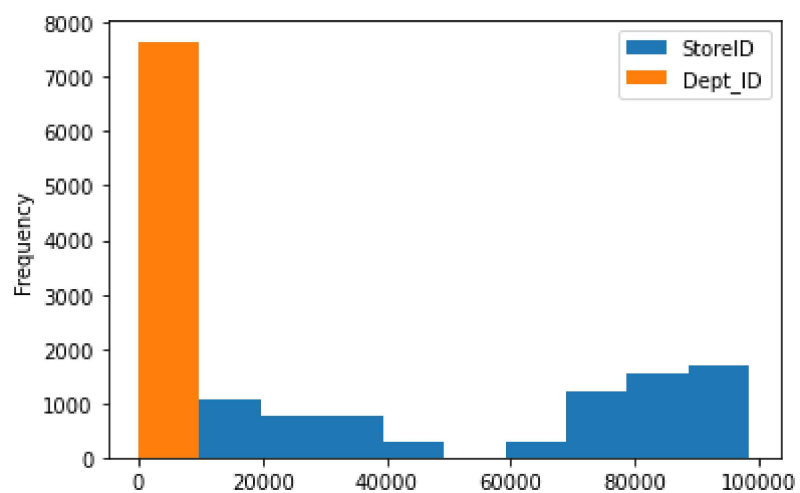
In [15]: `data.plot.bar()`

Out[15]: `<AxesSubplot:>`
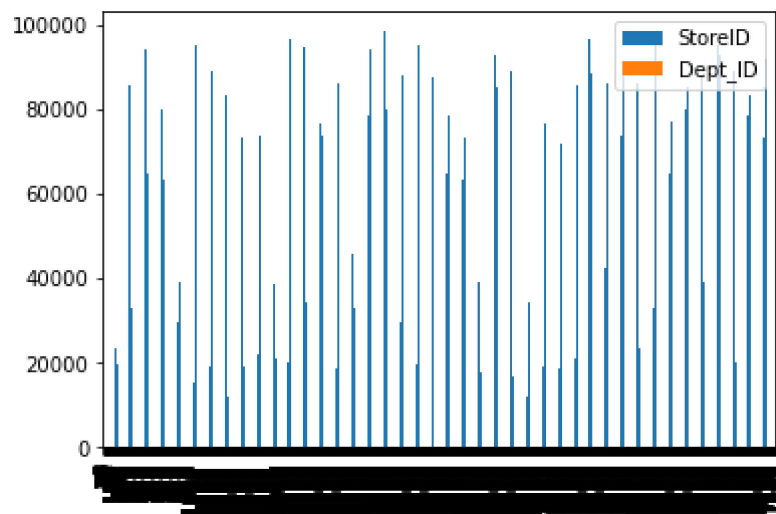
In [16]: `data.plot.area()`

Out[16]: `<AxesSubplot:>`
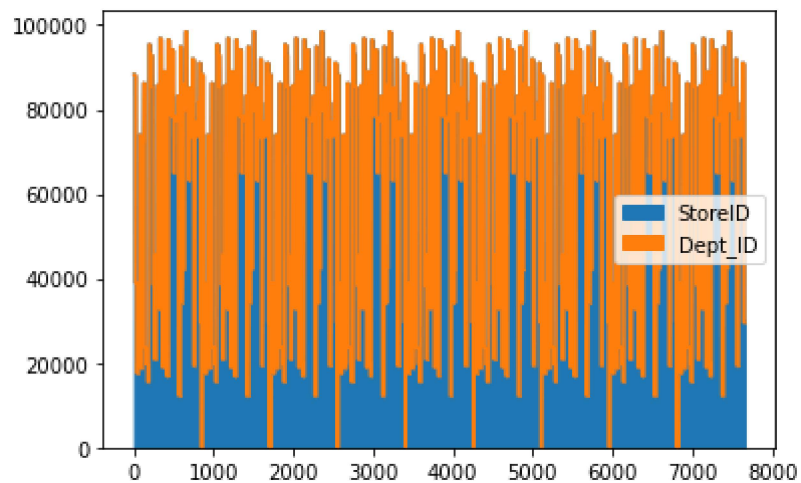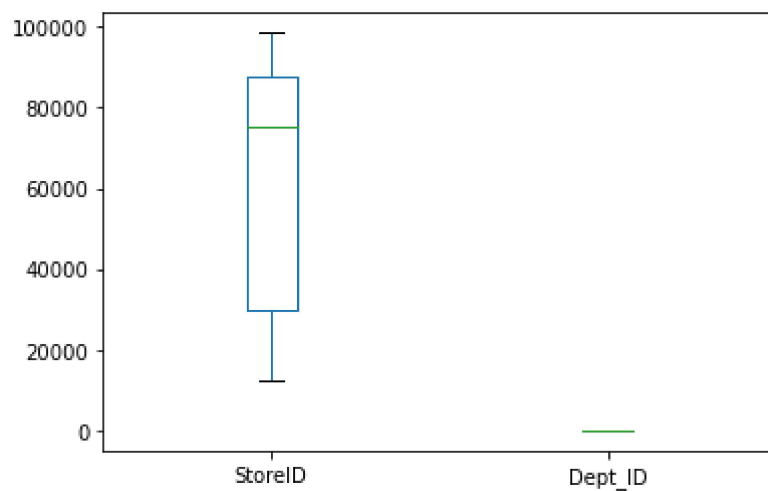
In [17]: `data.plot.hist()`

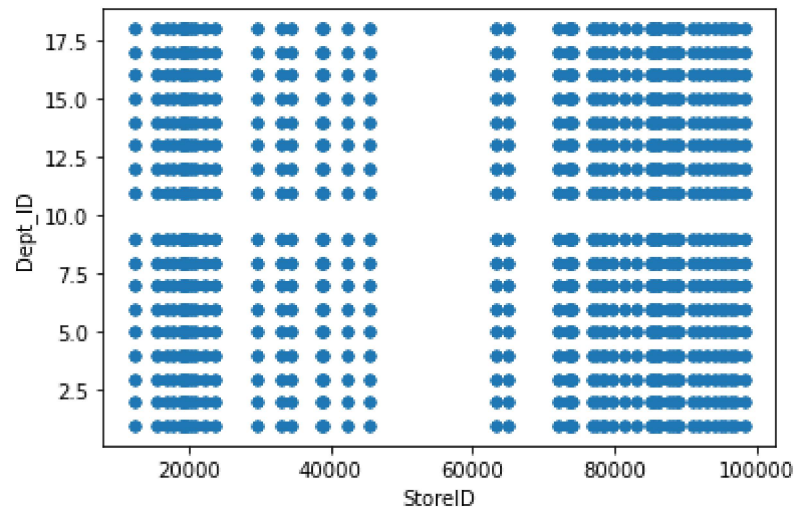Out[17]: `<AxesSubplot:ylabel='Frequency'>`

In [18]: `data.plot.box()`

Out[18]: `<AxesSubplot:>`

```
In [19]: data.plot.scatter(x="StoreID",y="Dept_ID")
```

Out[19]: <AxesSubplot:xlabel='StoreID', ylabel='Dept_ID'>


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

Out[20]: array([<AxesSubplot:ylabel='StoreID'>, <AxesSubplot:ylabel='Dept_ID'>],
            dtype=object)

```
In [ ]:
```