

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

```
In [86]: df=pd.read_csv(r"C:\Users\Admin\Downloads\6_Salesworkload1 - 6_Salesworkload1.csv")
df
```

Out[86]:

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

7658 rows × 14 columns



```
In [89]: df.isnull()
```

```
Out[89]:
```

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLease	Sales units
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...
7653	False	False	False	False	False	False	False	False	False	False
7654	False	False	False	False	False	False	False	False	False	False
7655	False	False	False	False	False	False	False	False	False	False
7656	False	False	False	False	False	False	False	False	False	False
7657	False	False	False	False	False	False	False	False	False	False

7658 rows × 14 columns



```
In [90]: df.shape
```

```
Out[90]: (7658, 14)
```

```
In [91]: df.size
```

```
Out[91]: 107212
```

```
In [92]: df.ndim
```

```
Out[92]: 2
```

In [93]: `df.head()`

Out[93]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLease	
0	10.2016	1.0	United Kingdom	88253.0	London (I)	1.0	Dry	3184.764	0.0	39
1	10.2016	1.0	United Kingdom	88253.0	London (I)	2.0	Frozen	1582.941	0.0	8
2	10.2016	1.0	United Kingdom	88253.0	London (I)	3.0	other	47.205	0.0	43
3	10.2016	1.0	United Kingdom	88253.0	London (I)	4.0	Fish	1623.852	0.0	30
4	10.2016	1.0	United Kingdom	88253.0	London (I)	5.0	Fruits & Vegetables	1759.173	0.0	16

In [94]: `df.tail()`

Out[94]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLeas	
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.

In [95]: `df.loc[1:3]`

Out[95]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLease	Sal un
1	10.2016	1.0	United Kingdom	88253.0	London (I)	2.0	Frozen	1582.941	0.0	8272
2	10.2016	1.0	United Kingdom	88253.0	London (I)	3.0	other	47.205	0.0	43840
3	10.2016	1.0	United Kingdom	88253.0	London (I)	4.0	Fish	1623.852	0.0	30942

```
In [106]: df2=df[df['Dept_ID']>=18]
df2
```

Out[106]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLease
16	10.2016	1.0	United Kingdom	88253.0	London (I)	18.0	all	40086.486	0.0
33	10.2016	1.0	United Kingdom	38976.0	Manchester	18.0	all	74401.374	0.0
50	10.2016	1.0	United Kingdom	17647.0	Liverpool	18.0	all	60727.659	1896.0
67	10.2016	1.0	United Kingdom	22117.0	Birmingham	18.0	all	35535.924	0.0
84	10.2016	1.0	United Kingdom	73949.0	Leicester	18.0	all	49061.73	510.0
...
7589	6.2017	9.0	Belgium	76852.0	Brussels (II)	18.0	all	49675.395	0.0
7606	6.2017	9.0	Belgium	73762.0	Antwerp	18.0	all	56066.952	0.0
7623	6.2017	9.0	Sweden	81473.0	Stockholm	18.0	all	72563.526	0.0
7640	6.2017	9.0	Sweden	90992.0	Malmö	18.0	all	40133.691	0.0
7657	6.2017	9.0	Sweden	29650.0	Gothenburg	18.0	all	39652.2	0.0

450 rows × 14 columns



```
In [107]: df.iloc[1:3]
```

Out[107]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLease	Sales units
1	10.2016	1.0	United Kingdom	88253.0	London (I)	2.0	Frozen	1582.941	0.0	82725
2	10.2016	1.0	United Kingdom	88253.0	London (I)	3.0	other	47.205	0.0	438400



```
In [108]: df.index
```

Out[108]: RangeIndex(start=0, stop=7658, step=1)

```
In [104]: df.columns
```

Out[104]: Index(['MonthYear', 'Time index', 'Country', 'StoreID', 'City', 'Dept_ID', 'Dept. Name', 'HoursOwn', 'HoursLease', 'Sales units', 'Turnover', 'Customer', 'Area (m2)', 'Opening hours'], dtype='object')

```
In [109]: df.describe()
```

```
Out[109]:
```

	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	(
mean	5.000000	61995.220000	9.470588	22.036078	1.076471e+06	3.721393e+06	N
std	2.582158	29924.581631	5.337429	133.299513	1.728113e+06	6.003380e+06	N
min	1.000000	12227.000000	1.000000	0.000000	0.000000e+00	0.000000e+00	N
25%	3.000000	29650.000000	5.000000	0.000000	5.457125e+04	2.726798e+05	N
50%	5.000000	75400.500000	9.000000	0.000000	2.932300e+05	9.319575e+05	N
75%	7.000000	87703.000000	14.000000	0.000000	9.175075e+05	3.264432e+06	N
max	9.000000	98422.000000	18.000000	3984.000000	1.124296e+07	4.271739e+07	N



```
In [110]: df1=df2[['Time index', 'Dept_ID']]
df1
```

```
Out[110]:
```

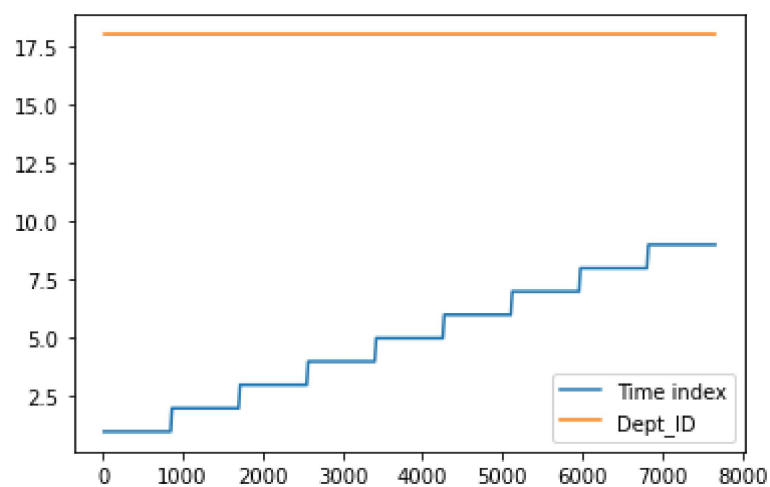
	Time index	Dept_ID
16	1.0	18.0
33	1.0	18.0
50	1.0	18.0
67	1.0	18.0
84	1.0	18.0
...
7589	9.0	18.0
7606	9.0	18.0
7623	9.0	18.0
7640	9.0	18.0
7657	9.0	18.0

450 rows × 2 columns

```
In [111]: import matplotlib.pyplot as pp
```

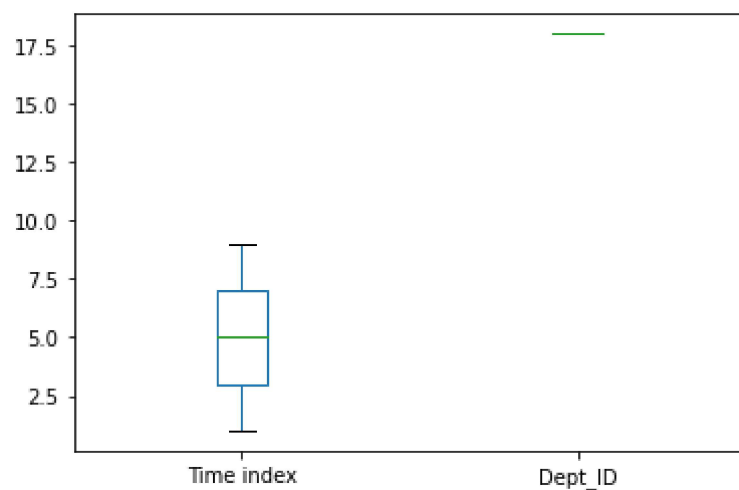
```
In [112]: df1.plot.line()
```

```
Out[112]: <matplotlib.axes._subplots.AxesSubplot at 0x142d4edb790>
```



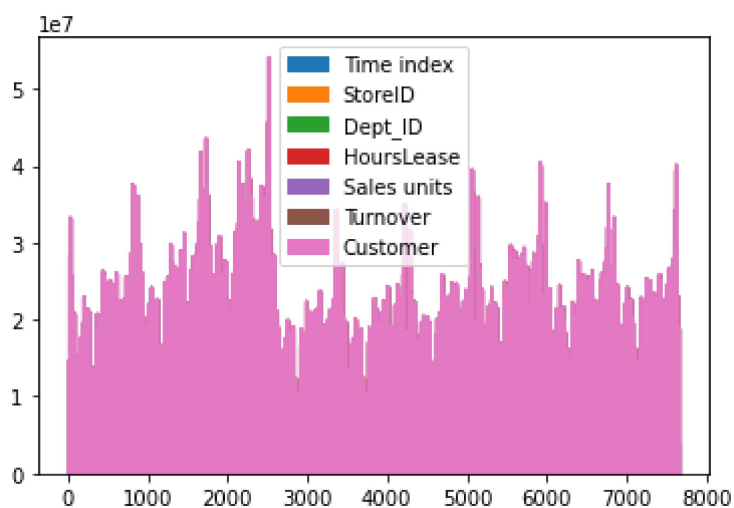
```
In [113]: df1.plot.box()
```

```
Out[113]: <matplotlib.axes._subplots.AxesSubplot at 0x142d53fa310>
```



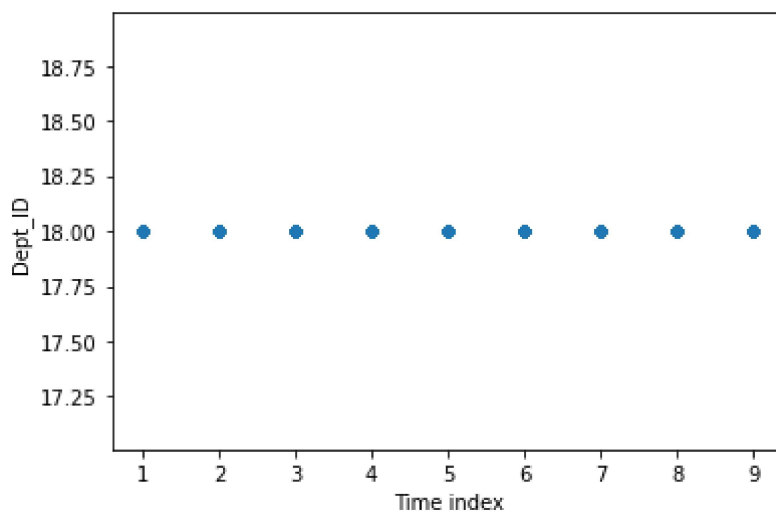
```
In [116]: df.plot.area()
```

```
Out[116]: <matplotlib.axes._subplots.AxesSubplot at 0x142d5ecc7c0>
```



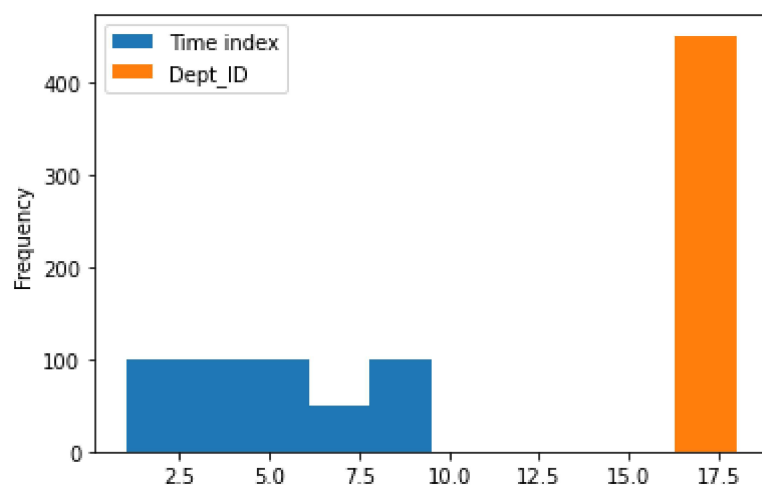
```
In [117]: df1.plot.scatter(x='Time index',y='Dept_ID')
```

```
Out[117]: <matplotlib.axes._subplots.AxesSubplot at 0x142d5f747c0>
```



```
In [118]: df1.plot.hist()
```

```
Out[118]: <matplotlib.axes._subplots.AxesSubplot at 0x142d5fe1cd0>
```



```
In [119]: df1.plot.pie(subplots=True)
```

```
Out[119]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x00000142D60136D0>,  
<matplotlib.axes._subplots.AxesSubplot object at 0x00000142D6095280>],  
      dtype=object)
```

