

Basic Analysis using numpy and pandas

Bottle dataset

To import library

In [1]:

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

To import dataset

In [2]:

```
d=pd.read_csv(r"E:\Dataset\9_bottle.csv")  
d
```

```
C:\ProgramData\Anaconda3\lib\site-packages\IPython\core\interactiveshell.p  
y:3165: DtypeWarning: Columns (47,73) have mixed types.Specify dtype optio  
n on import or set low_memory=False.  
    has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[2]:

Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	(
0	1	1	054.0 056.0 19-4903CR-HY-060-0930-05400560-0000A-3	0	10.500	33.4400	NaN	25.64900	
1	1	2	054.0 056.0 19-4903CR-HY-060-0930-05400560-0008A-3	8	10.460	33.4400	NaN	25.65600	
To get top 10 record									
2	1	3	054.0 056.0 19-4903CR-HY-060-0930-05400560-0010A-7	10	10.460	33.4370	NaN	25.65400	
3	1	4	054.0 056.0 19-4903CR-HY-060-0930-05400560-0019A-3	19	10.450	33.4200	NaN	25.64300	
4	1	5	054.0 056.0 19-4903CR-HY-060-0930-05400560-0020A-7	20	10.450	33.4210	NaN	25.64300	
...
864858	34404	864859	093.4 026.4 20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805	23.87055	1
864859	34404	864860	093.4 026.4 20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072	1
864860	34404	864861	093.4 026.4 20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911	1
864861	34404	864862	093.4 026.4 20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426	1

In [3]: Cst_Cnt Btl_Cnt Sta_ID Depth_ID Depthm T_degC Salnty O2ml_L STheta (

d.head(10)

				20-							
				1611SR-							
864862	34404	864863	093.4	MX-310-	15	17.533	33.3880	5.774	24.15297	1	
			026.4	2239-							
				09340264-							
				0015A-3							

864863 rows × 74 columns

Out[3]:

Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat	...
0	1	1	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.50	33.440	NaN	25.649	NaN	...
1	1	2	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.46	33.440	NaN	25.656	NaN	...
2	1	3	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.46	33.437	NaN	25.654	NaN	...
3	1	4	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.45	33.420	NaN	25.643	NaN	...
4	1	5	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.45	33.421	NaN	25.643	NaN	...
5	1	6	19-4903CR-HY-060-0930-05400560-0030A-7	30	10.45	33.431	NaN	25.651	NaN	...
6	1	7	19-4903CR-HY-060-0930-05400560-0039A-3	39	10.45	33.440	NaN	25.658	NaN	...
7	1	8	19-4903CR-HY-060-0930-05400560-0050A-7	50	10.24	33.424	NaN	25.682	NaN	...
8	1	9	19-4903CR-HY-060-0930-05400560-0058A-3	58	10.06	33.420	NaN	25.710	NaN	...

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat	...
To get last 10	10	10	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0075A-7	75	9.86	33.494	NaN	25.801	NaN	...

10 rows × 74 columns

In [4]:

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

Out[4]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	(
864853	34403	864854	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0381A-3	381	6.943	34.1104	1.108	26.73575	
864854	34403	864855	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0400A-7	400	6.694	34.1101	1.096	26.76927	
864855	34403	864856	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0440A-3	440	6.312	34.1563	0.718	26.85639	
864856	34403	864857	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0500A-7	500	5.993	34.2160	0.456	26.94518	
864857	34403	864858	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0521A-3	521	5.818	34.2382	0.366	26.98477	
864858	34404	864859	093.4 026.4	20- 1611SR- MX-310- 2239- 09340264- 0000A-7	0	18.744	33.4083	5.805	23.87055	1
864859	34404	864860	093.4 026.4	20- 1611SR- MX-310- 2239- 09340264- 0002A-3	2	18.744	33.4083	5.805	23.87072	1
864860	34404	864861	093.4 026.4	20- 1611SR- MX-310- 2239- 09340264- 0005A-3	5	18.692	33.4150	5.796	23.88911	1
864861	34404	864862	093.4 026.4	20- 1611SR- MX-310- 2239- 09340264- 0010A-3	10	18.161	33.4062	5.816	24.01426	1

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	(
Describe statistics Analysis				20-1611SR-093.4						
				MX-310-026.4	15	17.533	33.3880	5.774	24.15297	1
				2239-09340264-0015A-3						

In [5]:

```
d.describe()
```

Out[5]:

	Cst_Cnt	Btl_Cnt	Depthm	T_degC	Salnty	O2
count	864863.000000	864863.000000	864863.000000	853900.000000	817509.000000	696201.000000
mean	17138.790958	432432.000000	226.831951	10.799677	33.840350	3.390000
std	10240.949817	249664.587267	316.050259	4.243825	0.461843	2.070000
min	1.000000	1.000000	0.000000	1.440000	28.431000	-0.070000
25%	8269.000000	216216.500000	46.000000	7.680000	33.488000	1.360000
50%	16848.000000	432432.000000	125.000000	10.060000	33.863000	3.440000
75%	26557.000000	648647.500000	300.000000	13.880000	34.196900	5.500000
max	34404.000000	864863.000000	5351.000000	31.140000	37.034000	11.130000

8 rows × 70 columns

To get rows and columns

In [6]:

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

Out[6]:

(864863, 74)

To get number of elements

In [7]:

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

Out[7]:

63999862

To get the missing value

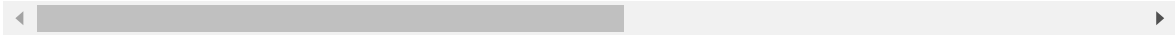
In [8]:

```
d.isna()
```

Out[8]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
0	False	False	False	False	False	False	False	True	False	Tru
1	False	False	False	False	False	False	False	True	False	Tru
2	False	False	False	False	False	False	False	True	False	Tru
3	False	False	False	False	False	False	False	True	False	Tru
4	False	False	False	False	False	False	False	True	False	Tru
...
864858	False	False	False	False	False	False	False	False	False	Fals
864859	False	False	False	False	False	False	False	False	False	Fals
864860	False	False	False	False	False	False	False	False	False	Fals
864861	False	False	False	False	False	False	False	False	False	Fals
864862	False	False	False	False	False	False	False	False	False	Fals

864863 rows × 74 columns



To drop the missing elements

In [9]:

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

Out[9]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	Reclnd	R_Depth	R_PRES
0	1	1	054.0 056.0	19-4903CR-HY-060- 0930-05400560- 0000A-3	0	3	0.0	0
1	1	2	054.0 056.0	19-4903CR-HY-060- 0930-05400560- 0008A-3	8	3	8.0	8
2	1	3	054.0 056.0	19-4903CR-HY-060- 0930-05400560- 0010A-7	10	7	10.0	10
3	1	4	054.0 056.0	19-4903CR-HY-060- 0930-05400560- 0019A-3	19	3	19.0	19
4	1	5	054.0 056.0	19-4903CR-HY-060- 0930-05400560- 0020A-7	20	7	20.0	20
...
864858	34404	864859	093.4 026.4	20-1611SR-MX-310- 2239-09340264- 0000A-7	0	7	0.0	0
864859	34404	864860	093.4 026.4	20-1611SR-MX-310- 2239-09340264- 0002A-3	2	3	2.0	2
864860	34404	864861	093.4 026.4	20-1611SR-MX-310- 2239-09340264- 0005A-3	5	3	5.0	5
864861	34404	864862	093.4 026.4	20-1611SR-MX-310- 2239-09340264- 0010A-3	10	3	10.0	10
864862	34404	864863	093.4 026.4	20-1611SR-MX-310- 2239-09340264- 0015A-3	15	3	15.0	15

864863 rows × 8 columns

In [10]:

```
d["Sta_ID"]
```

Out[10]:

```
0      054.0 056.0
1      054.0 056.0
2      054.0 056.0
3      054.0 056.0
4      054.0 056.0
```

...

```
864858  093.4 026.4
864859  093.4 026.4
864860  093.4 026.4
864861  093.4 026.4
864862  093.4 026.4
```

Name: Sta_ID, Length: 864863, dtype: object

In [11]:

```
data=pd.DataFrame(d[['Depthm', 'T_degC']][0:500])
data
```

Out[11]:

	Depthm	T_degC
0	0	10.50
1	8	10.46
2	10	10.46
3	19	10.45
4	20	10.45
...
495	700	4.90
496	792	4.50
497	800	4.48
498	900	4.21
499	1000	3.95

500 rows × 2 columns

In [12]:

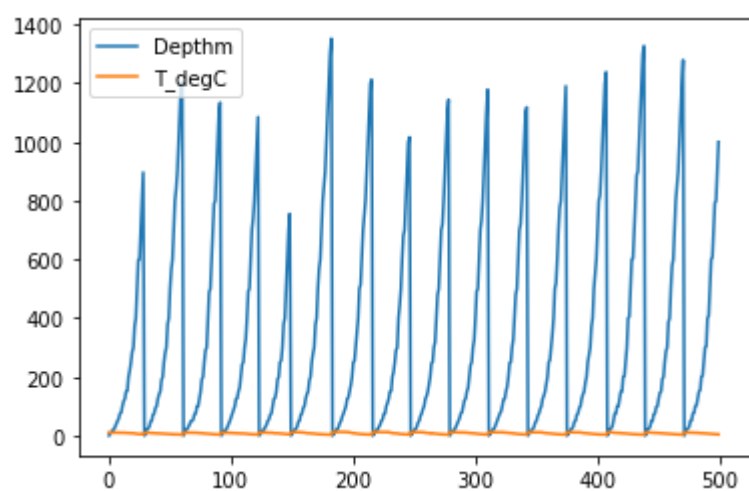
```
import matplotlib.pyplot as pp
```

In [13]:

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

Out[13]:

<AxesSubplot:>

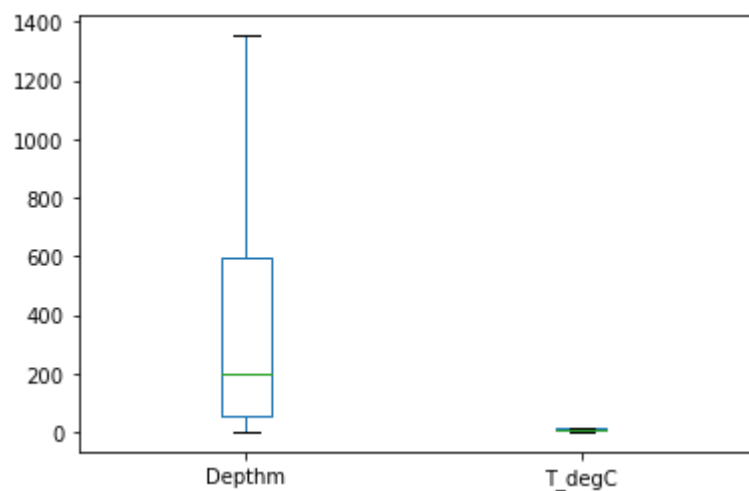


In [14]:

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

Out[14]:

<AxesSubplot:>

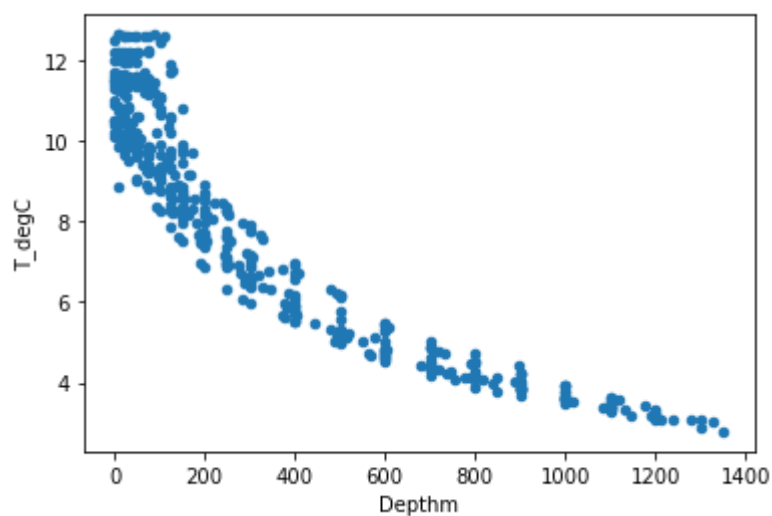


In [15]:

```
data.plot.scatter(x="Depthm",y="T_degC")
```

Out[15]:

<AxesSubplot:xlabel='Depthm', ylabel='T_degC'>

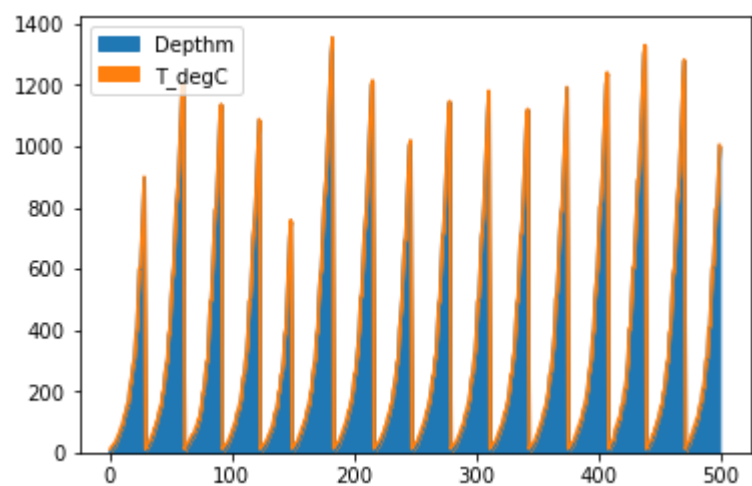


In [16]:

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

Out[16]:

<AxesSubplot:>

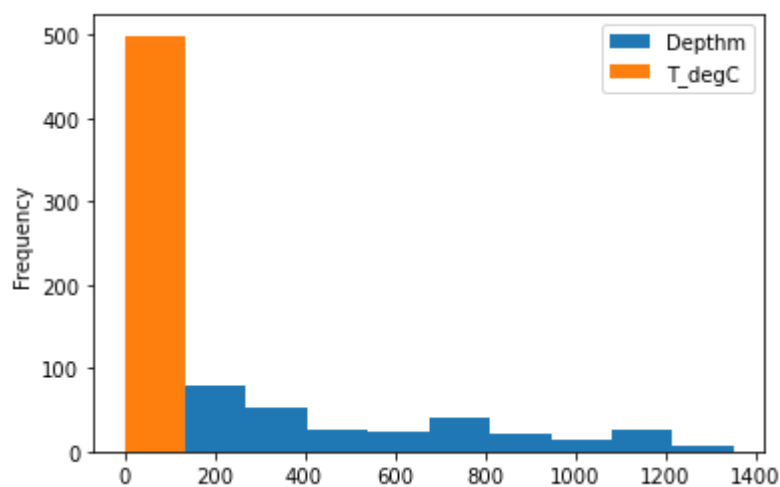


In [17]:

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

Out[17]:

<AxesSubplot:ylabel='Frequency'>

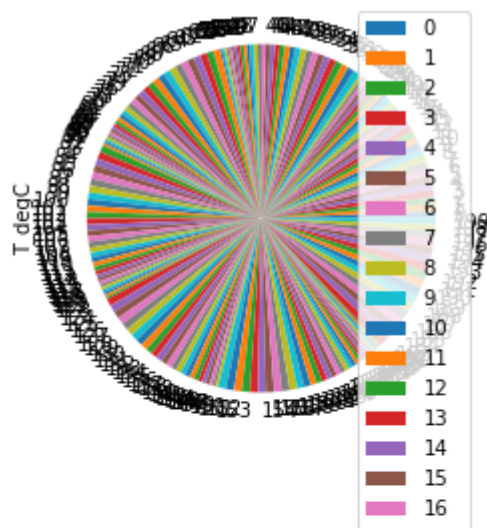


In [18]:

```
data=pd.DataFrame(d[['Depthm','T_degC']][0:200])  
data.plot.pie(y="T_degC")
```

Out[18]:

<AxesSubplot:ylabel='T_degC'>



In []: