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

In [9]:
```
Pre-processing
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-9-79bb87d727ab> in <module>
----> 1 Pre-processing

NameError: name 'Pre' is not defined
```

In [23]:
```
data=pd.read_csv(r"C:\Users\user\Downloads\9_bottle.csv")
data
```

```
C:\ProgramData\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3
165: DtypeWarning: Columns (47,73) have mixed types.Specify dtype option on i
mport or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[23]:

| | Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | T_degC | Salnty | O2ml_L | STheta | O2S |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0000A-3 | 0 | 10.500 | 33.4400 | NaN | 25.64900 | Na |
| **1** | 1 | 2 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0008A-3 | 8 | 10.460 | 33.4400 | NaN | 25.65600 | Na |
| **2** | 1 | 3 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0010A-7 | 10 | 10.460 | 33.4370 | NaN | 25.65400 | Na |
| **3** | 1 | 4 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0019A-3 | 19 | 10.450 | 33.4200 | NaN | 25.64300 | Na |
| **4** | 1 | 5 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0020A-7 | 20 | 10.450 | 33.4210 | NaN | 25.64300 | Na |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **864858** | 34404 | 864859 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0000A-7 | 0 | 18.744 | 33.4083 | 5.805 | 23.87055 | 108. |
| **864859** | 34404 | 864860 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0002A-3 | 2 | 18.744 | 33.4083 | 5.805 | 23.87072 | 108. |
| **864860** | 34404 | 864861 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0005A-3 | 5 | 18.692 | 33.4150 | 5.796 | 23.88911 | 108. |
| **864861** | 34404 | 864862 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0010A-3 | 10 | 18.161 | 33.4062 | 5.816 | 24.01426 | 107. |

| | Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | T_degC | Salnty | O2ml_L | STheta | O2S |
|---|---|---|---|---|---|---|---|---|---|---|
| **864862** | 34404 | 864863 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0015A-3 | 15 | 17.533 | 33.3880 | 5.774 | 24.15297 | 105. |

864863 rows × 74 columns

In [24]:
```
data.head()
```

Out[24]:

| | Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | T_degC | Salnty | O2ml_L | STheta | O2Sat | ... | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0000A-3 | 0 | 10.50 | 33.440 | NaN | 25.649 | NaN | ... | |
| **1** | 1 | 2 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0008A-3 | 8 | 10.46 | 33.440 | NaN | 25.656 | NaN | ... | |
| **2** | 1 | 3 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0010A-7 | 10 | 10.46 | 33.437 | NaN | 25.654 | NaN | ... | |
| **3** | 1 | 4 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0019A-3 | 19 | 10.45 | 33.420 | NaN | 25.643 | NaN | ... | |
| **4** | 1 | 5 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0020A-7 | 20 | 10.45 | 33.421 | NaN | 25.643 | NaN | ... | |

5 rows × 74 columns

In [25]:
```
data.tail()
```

Out[25]:

| | Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | T_degC | Salnty | O2ml_L | STheta | O2S |
|---|---|---|---|---|---|---|---|---|---|---|
| **864858** | 34404 | 864859 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0000A-7 | 0 | 18.744 | 33.4083 | 5.805 | 23.87055 | 108. |
| **864859** | 34404 | 864860 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0002A-3 | 2 | 18.744 | 33.4083 | 5.805 | 23.87072 | 108. |
| **864860** | 34404 | 864861 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0005A-3 | 5 | 18.692 | 33.4150 | 5.796 | 23.88911 | 108. |
| **864861** | 34404 | 864862 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0010A-3 | 10 | 18.161 | 33.4062 | 5.816 | 24.01426 | 107. |
| **864862** | 34404 | 864863 | 093.4 026.4 | 20-1611SR-MX-310-2239-09340264-0015A-3 | 15 | 17.533 | 33.3880 | 5.774 | 24.15297 | 105. |

5 rows × 74 columns

In [26]:
```
data.describe()
```

Out[26]:

| | Cst_Cnt | Btl_Cnt | Depthm | T_degC | Salnty | O2ml |
|---|---|---|---|---|---|---|
| **count** | 864863.000000 | 864863.000000 | 864863.000000 | 853900.000000 | 817509.000000 | 696201.0000 |
| **mean** | 17138.790958 | 432432.000000 | 226.831951 | 10.799677 | 33.840350 | 3.3924 |
| **std** | 10240.949817 | 249664.587267 | 316.050259 | 4.243825 | 0.461843 | 2.0732 |
| **min** | 1.000000 | 1.000000 | 0.000000 | 1.440000 | 28.431000 | -0.0100 |
| **25%** | 8269.000000 | 216216.500000 | 46.000000 | 7.680000 | 33.488000 | 1.3600 |
| **50%** | 16848.000000 | 432432.000000 | 125.000000 | 10.060000 | 33.863000 | 3.4400 |
| **75%** | 26557.000000 | 648647.500000 | 300.000000 | 13.880000 | 34.196900 | 5.5000 |
| **max** | 34404.000000 | 864863.000000 | 5351.000000 | 31.140000 | 37.034000 | 11.1300 |

8 rows × 70 columns

In [27]:
```
print(np.shape(data))
```

```
(864863, 74)
```

In [28]:
```
print(np.size(data))
```

```
63999862
```

In [8]:
```
data.isnull()
```

Out[8]:

| | Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | T_degC | Salnty | O2ml_L | STheta | O2Sat |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | False | False | False | False | False | False | False | True | False | True |
| **1** | False | False | False | False | False | False | False | True | False | True |
| **2** | False | False | False | False | False | False | False | True | False | True |
| **3** | False | False | False | False | False | False | False | True | False | True |
| **4** | False | False | False | False | False | False | False | True | False | True |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **864858** | False | False | False | False | False | False | False | False | False | False |
| **864859** | False | False | False | False | False | False | False | False | False | False |
| **864860** | False | False | False | False | False | False | False | False | False | False |
| **864861** | False | False | False | False | False | False | False | False | False | False |
| **864862** | False | False | False | False | False | False | False | False | False | False |

864863 rows × 74 columns

```
Visualization
```

In [29]:
```
da=data.head(200)
da
```

Out[29]:

| | Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | T_degC | Salnty | O2ml_L | STheta | O2Sat | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0000A-3 | 0 | 10.50 | 33.440 | NaN | 25.649 | NaN | ... |
| **1** | 1 | 2 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0008A-3 | 8 | 10.46 | 33.440 | NaN | 25.656 | NaN | ... |
| **2** | 1 | 3 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0010A-7 | 10 | 10.46 | 33.437 | NaN | 25.654 | NaN | ... |
| **3** | 1 | 4 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0019A-3 | 19 | 10.45 | 33.420 | NaN | 25.643 | NaN | ... |
| **4** | 1 | 5 | 054.0 056.0 | 19-4903CR-HY-060-0930-05400560-0020A-7 | 20 | 10.45 | 33.421 | NaN | 25.643 | NaN | ... |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **195** | 7 | 196 | 056.7 146.0 | 19-4903CR-HY-063-0506-05671460-0125A-7 | 125 | 11.94 | 33.383 | NaN | 25.348 | NaN | ... |
| **196** | 7 | 197 | 056.7 146.0 | 19-4903CR-HY-063-0506-05671460-0130A-3 | 130 | 11.77 | 33.400 | NaN | 25.393 | NaN | ... |
| **197** | 7 | 198 | 056.7 146.0 | 19-4903CR-HY-063-0506-05671460-0150A-7 | 150 | 10.84 | 33.420 | NaN | 25.577 | NaN | ... |
| **198** | 7 | 199 | 056.7 146.0 | 19-4903CR-HY-063-0506-05671460-0172A-3 | 172 | 9.73 | 33.440 | NaN | 25.782 | NaN | ... |

| | Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | T_degC | Salnty | O2ml_L | STheta | O2Sat | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **199** | 7 | 200 | 056.7 146.0 | 19-4903CR-HY-063-0506-05671460-0200A-7 | 200 | 8.90 | 33.518 | NaN | 25.977 | NaN | ... |

200 rows × 74 columns

In [38]:
```
df=da[['T_degC','Depthm']]
df
```

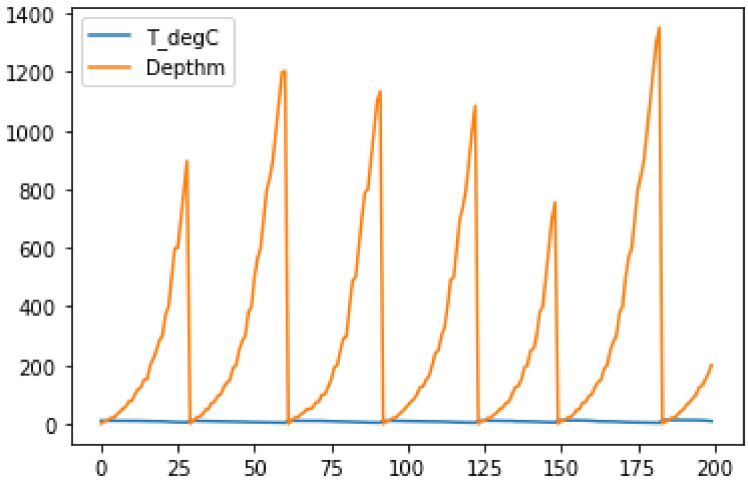
Out[38]:

| | T_degC | Depthm |
|---|---|---|
| **0** | 10.50 | 0 |
| **1** | 10.46 | 8 |
| **2** | 10.46 | 10 |
| **3** | 10.45 | 19 |
| **4** | 10.45 | 20 |
| **...** | ... | ... |
| **195** | 11.94 | 125 |
| **196** | 11.77 | 130 |
| **197** | 10.84 | 150 |
| **198** | 9.73 | 172 |
| **199** | 8.90 | 200 |

200 rows × 2 columns

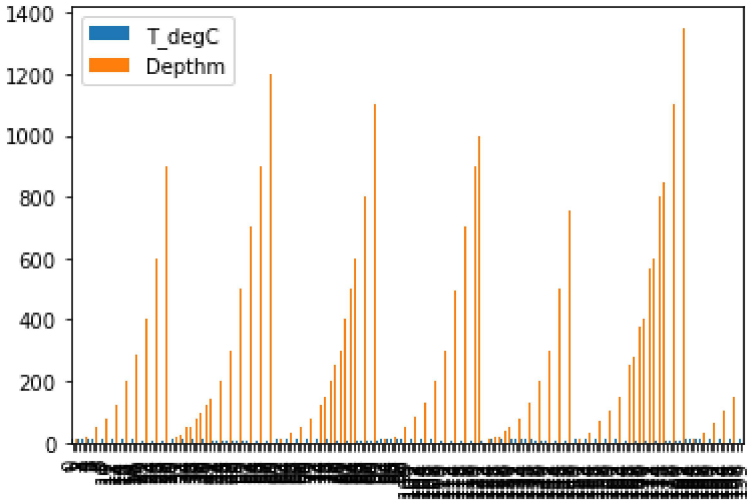
In [39]:
```
df.plot.line()
```

Out[39]: <AxesSubplot:>



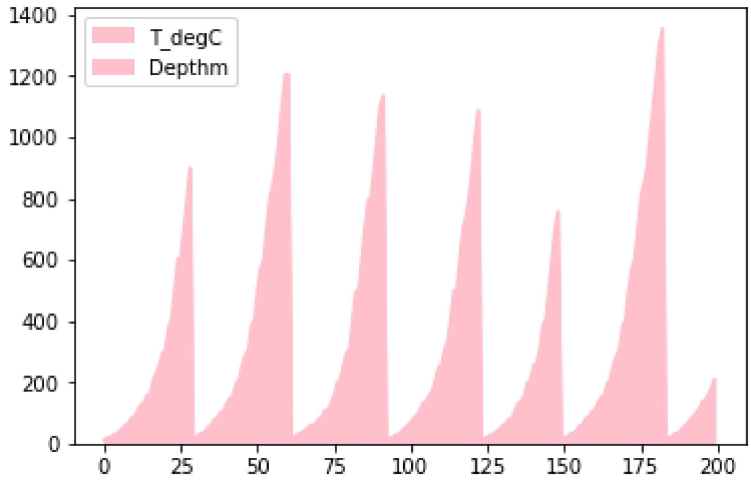

In [40]:
```
df.plot.bar()
```

Out[40]: <AxesSubplot:>



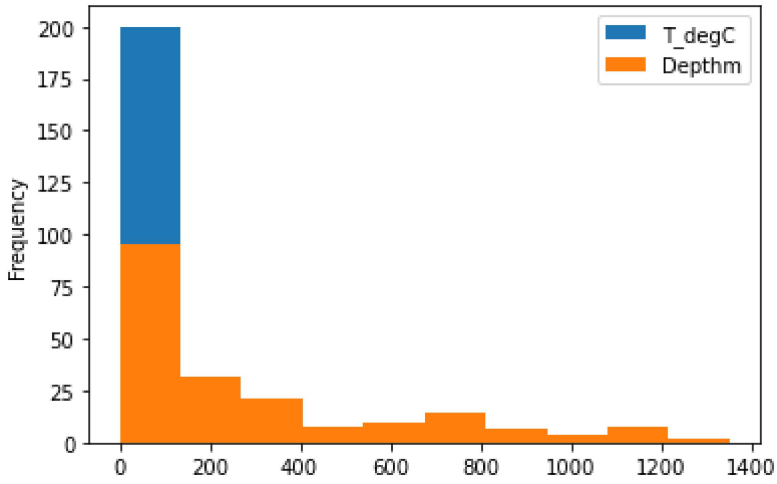
In [41]:
```
df.plot.area(color="pink")
```

Out[41]:
```
<AxesSubplot:>
```



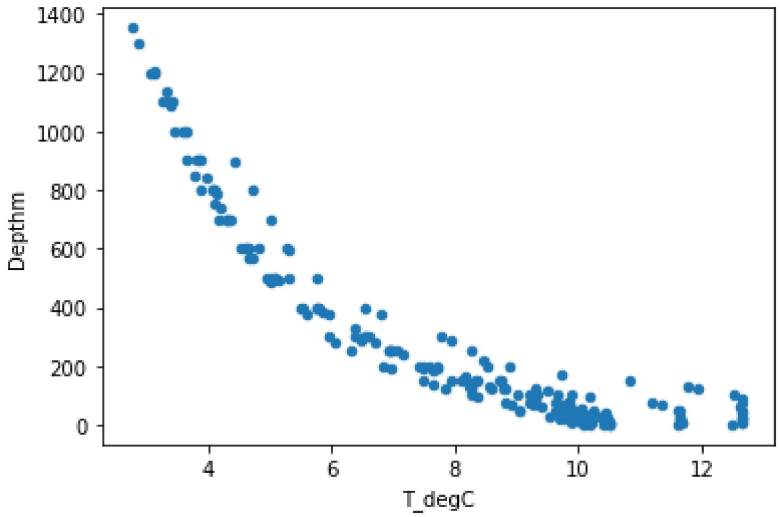

In [42]:
```
df.plot.hist()
```

Out[42]:
```
<AxesSubplot:ylabel='Frequency'>
```

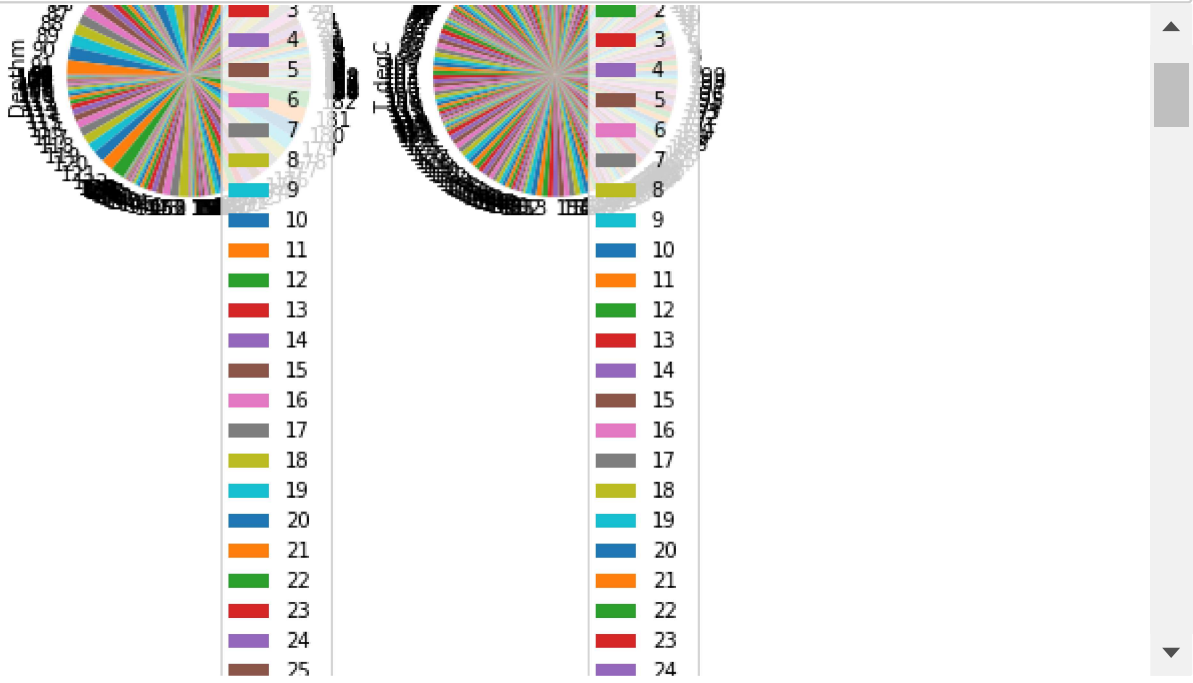


In [45]:
```
df.plot.scatter(x='T_degC',y='Depthm')
```

Out[45]: <AxesSubplot:xlabel='T_degC', ylabel='Depthm'>




In [44]:
```
de=df[['Depthm','T_degC']]
de.plot.pie(subplots=True)
```




In [ ]: