Clean, preprocess and visulaize the data

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

In [31]:
```
d=pd.read_csv(r"C:\Users\Admin\Downloads\9_bottle.csv")
d
```

```
C:\Users\Admin\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3
071: 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[31]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
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	108.7
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072	108.7
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911	108.4
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426	107.7

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297	105.6

864863 rows × 74 columns

In [32]:
```
d.head()
```

Out[32]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat	...	R
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 [33]:
```
d.tail()
```

Out[33]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805	23.87055	108.7
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072	108.7
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911	108.4
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426	107.7
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297	105.6

5 rows × 74 columns

In [37]:
```
d.describe()
```

Out[37]:

	Cst_Cnt	Btl_Cnt	Depthm	T_degC	Salnty	O2ml_
count	864863.000000	864863.000000	864863.000000	853900.000000	817509.000000	696201.00000
mean	17138.790958	432432.000000	226.831951	10.799677	33.840350	3.39246
std	10240.949817	249664.587267	316.050259	4.243825	0.461843	2.07325
min	1.000000	1.000000	0.000000	1.440000	28.431000	-0.01000
25%	8269.000000	216216.500000	46.000000	7.680000	33.488000	1.36000
50%	16848.000000	432432.000000	125.000000	10.060000	33.863000	3.44000
75%	26557.000000	648647.500000	300.000000	13.880000	34.196900	5.50000
max	34404.000000	864863.000000	5351.000000	31.140000	37.034000	11.13000

8 rows × 70 columns

In [35]:
```
d.shape
```

Out[35]:
```
(864863, 74)
```

In [36]:
```
d.size
```

Out[36]:
```
63999862
```

In [38]:
```
d.isna
```

Out[38]:
```
<bound method DataFrame.isna of         Cst_Cnt  Btl_Cnt       Sta_ID
Depth_ID  \
0             1        1  054.0 056.0  19-4903CR-HY-060-0930-05400560-0000A-3
1             1        2  054.0 056.0  19-4903CR-HY-060-0930-05400560-0008A-3
2             1        3  054.0 056.0  19-4903CR-HY-060-0930-05400560-0010A-7
3             1        4  054.0 056.0  19-4903CR-HY-060-0930-05400560-0019A-3
4             1        5  054.0 056.0  19-4903CR-HY-060-0930-05400560-0020A-7
...         ...      ...          ...                                     ...
864858    34404   864859  093.4 026.4  20-1611SR-MX-310-2239-09340264-0000A-7
864859    34404   864860  093.4 026.4  20-1611SR-MX-310-2239-09340264-0002A-3
864860    34404   864861  093.4 026.4  20-1611SR-MX-310-2239-09340264-0005A-3
864861    34404   864862  093.4 026.4  20-1611SR-MX-310-2239-09340264-0010A-3
864862    34404   864863  093.4 026.4  20-1611SR-MX-310-2239-09340264-0015A-3

        Depthm  T_degC   Salnty  O2ml_L    STheta   O2Sat  ...  R_PHAEO  \
0            0  10.500  33.4400     NaN  25.64900     NaN  ...      NaN
1            8  10.460  33.4400     NaN  25.65600     NaN  ...      NaN
2           10  10.460  33.4370     NaN  25.65400     NaN  ...      NaN
3           19  10.450  33.4200     NaN  25.64300     NaN  ...      NaN
4           20  10.450  33.4210     NaN  25.64300     NaN  ...      NaN
...        ...     ...      ...     ...       ...     ...  ...      ...
864858       0  18.744  33.4083   5.805  23.87055  108.74  ...     0.18
864859       2  18.744  33.4083   5.805  23.87072  108.74  ...     0.18
864860       5  18.692  33.4150   5.796  23.88911  108.46  ...     0.18
864861      10  18.161  33.4062   5.816  24.01426  107.74  ...     0.31
864862      15  17.533  33.3880   5.774  24.15297  105.66  ...     0.61

        R_PRES  R_SAMP  DIC1  DIC2  TA1  TA2  pH2  pH1  DIC Quality Comment
0            0     NaN   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
1            8     NaN   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
2           10     NaN   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
3           19     NaN   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
4           20     NaN   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
...        ...     ...   ...   ...  ...  ...  ...  ...                  ...
864858       0     NaN   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
864859       2     4.0   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
864860       5     3.0   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
864861      10     2.0   NaN   NaN  NaN  NaN  NaN  NaN                  NaN
864862      15     1.0   NaN   NaN  NaN  NaN  NaN  NaN                  NaN

[864863 rows x 74 columns]>
```

In [39]:
```
d.dropna(axis=1,how="any")
```

Out[39]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	RecInd	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 [40]:
```
d["Cst_Cnt"]
```

Out[40]:
```
0             1
1             1
2             1
3             1
4             1
           ...
864858    34404
864859    34404
864860    34404
864861    34404
864862    34404
Name: Cst_Cnt, Length: 864863, dtype: int64
```

In [43]:
```
d1=d[["Depthm","RecInd"]].head(100)
d1
```

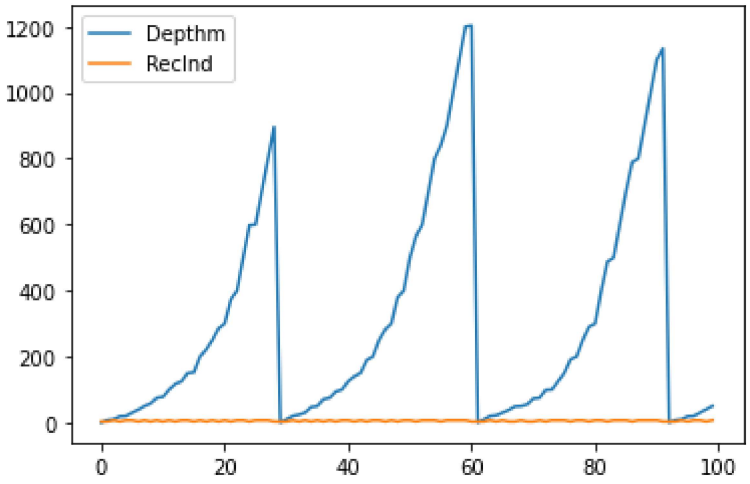
Out[43]:

	Depthm	RecInd
0	0	3
1	8	3
2	10	7
3	19	3
4	20	7
...	...	...
95	19	3
96	20	7
97	30	7
98	40	3
99	50	7

100 rows × 2 columns

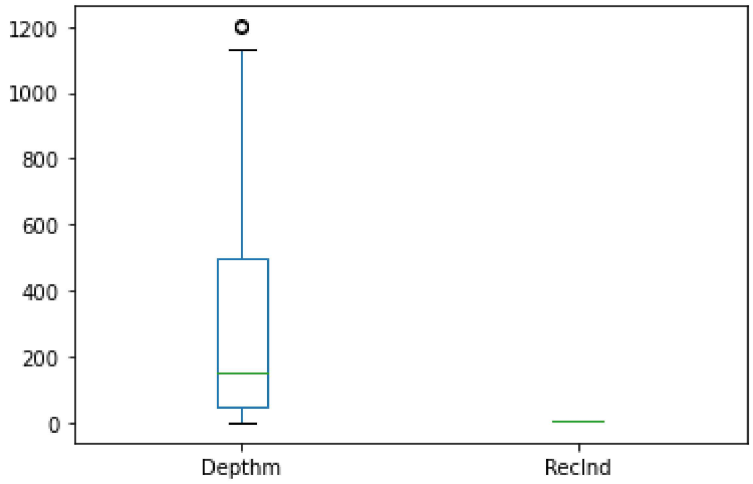
In [44]:
```
d1.plot.line()
```

Out[44]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x19138cb8700>
```



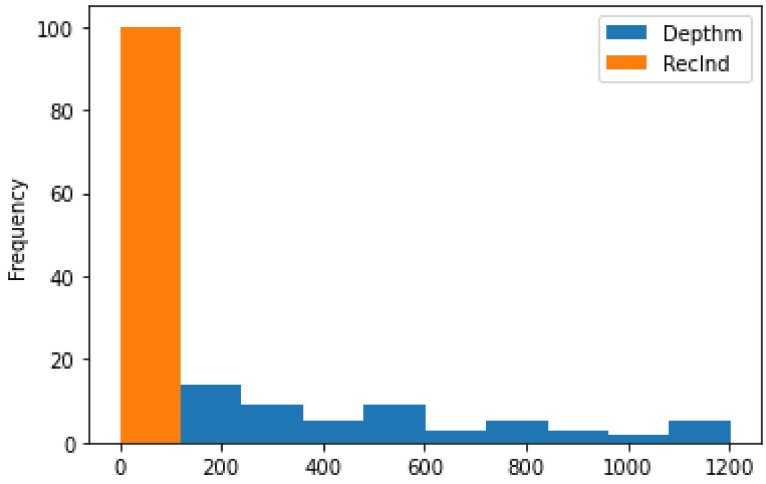
In [45]:
```
d1.plot.box()
```

Out[45]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1912086d8b0>
```



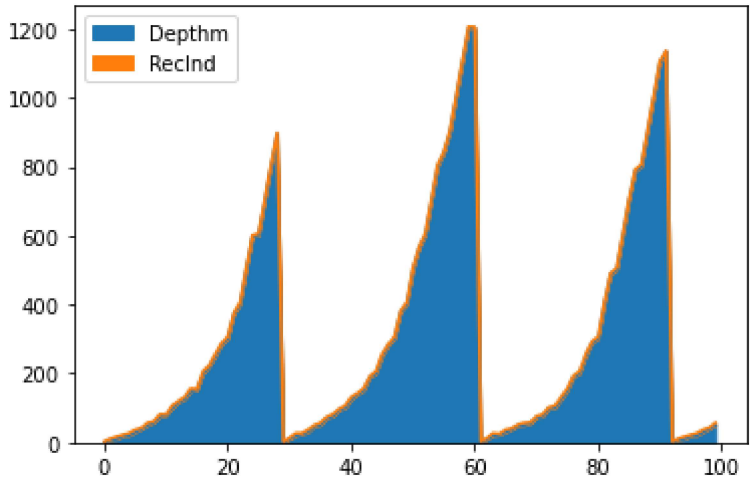
In [46]:
```
d1.plot.hist()
```

Out[46]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x191208b73a0>
```



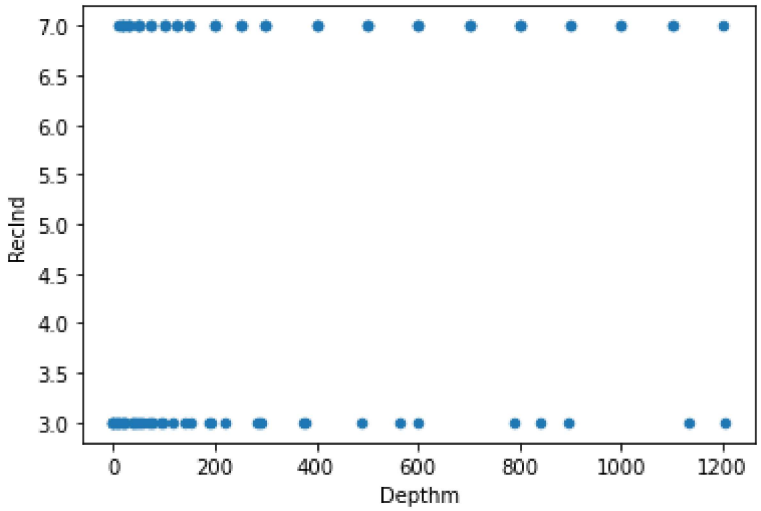
In [47]:
```
d1.plot.area()
```

Out[47]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1912091d250>
```



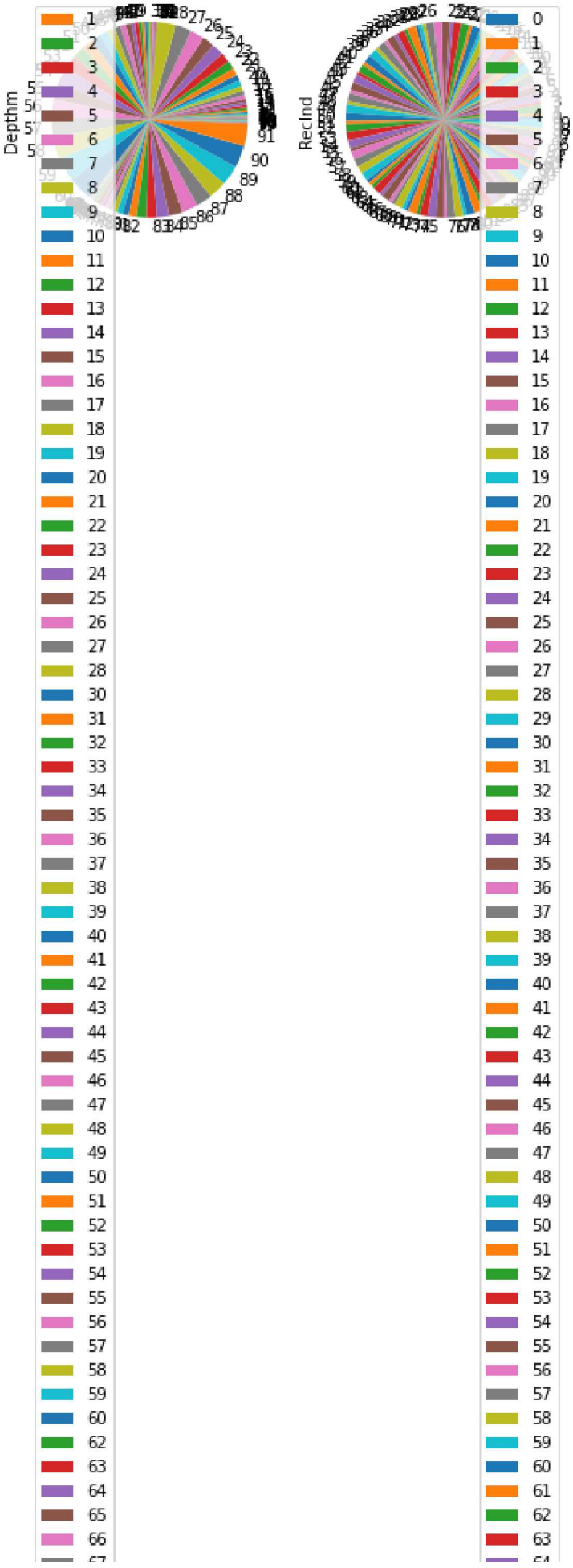
In [49]:
```
d1.plot.scatter(x="Depthm",y="RecInd")
```

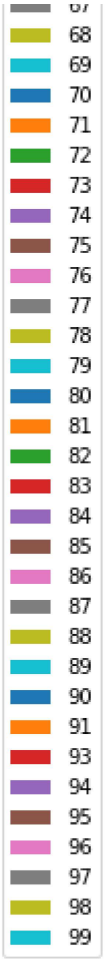
Out[49]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x19120a2c850>
```

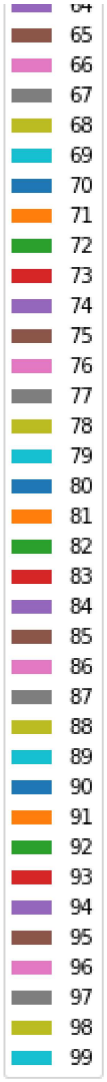


In [50]:
```
d1.plot.pie(subplots=True)
```

Out[50]:
```
array([<matplotlib.axes._subplots.AxesSubplot object at 0x0000019120A80E80>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x0000019120AB9730>],
      dtype=object)
```


Depthm
RecInd
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
62
63
64
65
66
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63


68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
93
94
95
96
97
98
99


65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

In [ ]: