

## Importing the datasets

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

```
In [2]: df=pd.read_csv(r'C:\Users\user\Desktop\9_bottle.csv')
```

```
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,
```

## Display first 10 values

```
In [3]: df.head(10)
```

Out[3]:

	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	1	6	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0030A-7	30	10.45	33.431	NaN	25.651	NaN	...	
6	1	7	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0039A-3	39	10.45	33.440	NaN	25.658	NaN	...	
7	1	8	054.0 056.0	19- 4903CR- HY-060- 0930- 05400560- 0050A-7	50	10.24	33.424	NaN	25.682	NaN	...	
8	1	9	054.0 056.0	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	...	R
9	1	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

display the bottom 11 values

```
In [4]: df.tail(11)
```

Out[4]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2S
<b>864852</b>	34403	864853	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0321A-3	321	7.538	34.0420	1.984	26.59793	29.5
<b>864853</b>	34403	864854	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0381A-3	381	6.943	34.1104	1.108	26.73575	16.2
<b>864854</b>	34403	864855	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0400A-7	400	6.694	34.1101	1.096	26.76927	15.9
<b>864855</b>	34403	864856	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0440A-3	440	6.312	34.1563	0.718	26.85639	10.3
<b>864856</b>	34403	864857	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0500A-7	500	5.993	34.2160	0.456	26.94518	6.5
<b>864857</b>	34403	864858	093.3 120.0	20- 1611SR- MX-313- 2053- 09331200- 0521A-3	521	5.818	34.2382	0.366	26.98477	5.2
<b>864858</b>	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
<b>864859</b>	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
<b>864860</b>	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

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2S
864861	34404	864862	093.4026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426	107.7
864862	34404	864863	093.4026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297	105.6

11 rows × 74 columns

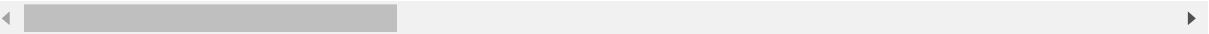
## To show the statistical values

In [5]: df.describe()

Out[5]:

	Cst_Cnt	Btl_Cnt	Depthm	T_degC	Salnty	O2ml_
count	864863.000000	864863.000000	864863.000000	853900.000000	817509.000000	696201.000000
mean	17138.790958	432432.000000	226.831951	10.799677	33.840350	3.39246
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



## shape()

In [6]: np.shape(df)

Out[6]: (864863, 74)

## size(df)

```
In [7]: np.size(df)
```

```
Out[7]: 63999862
```

## isna()

```
In [8]: df.dropna()
```

```
Out[8]:
```

Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat	...	R_F
---------	---------	--------	----------	--------	--------	--------	--------	--------	-------	-----	-----

---

0 rows × 74 columns



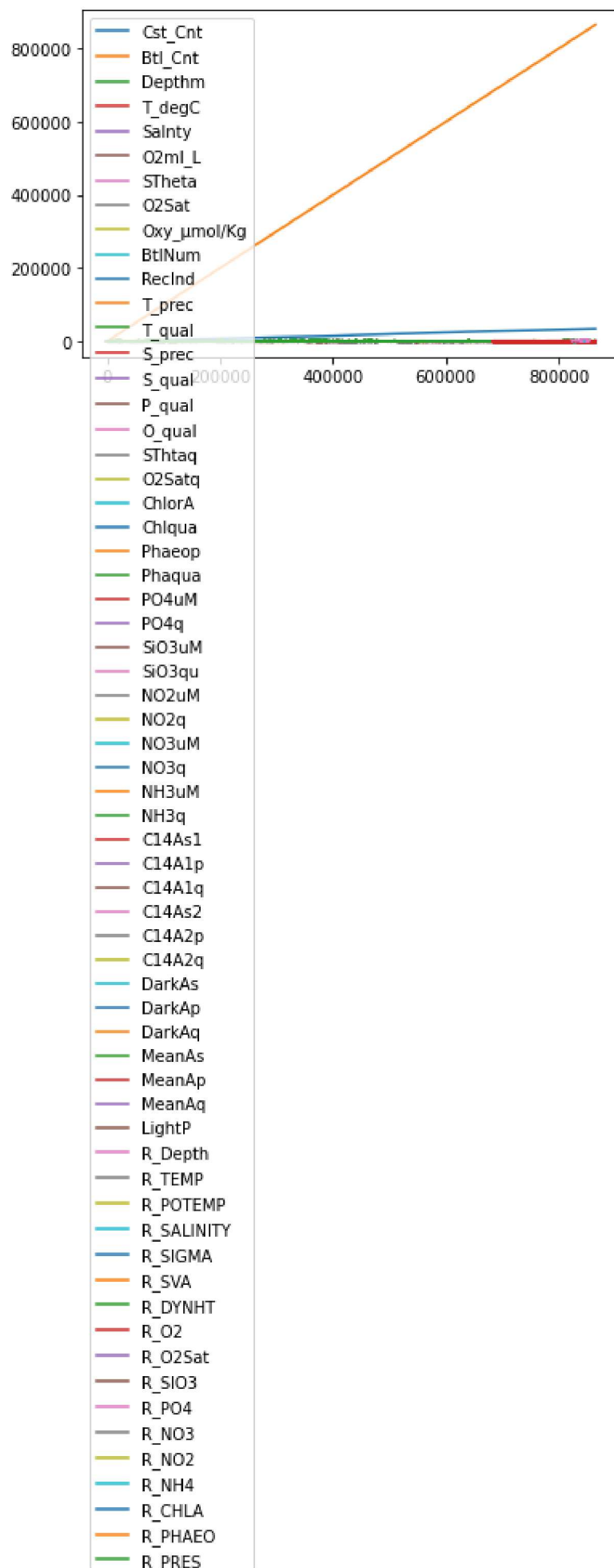
## Visualization



```
In [9]: df.plot.line()
```

```
Out[9]: <AxesSubplot:>
```



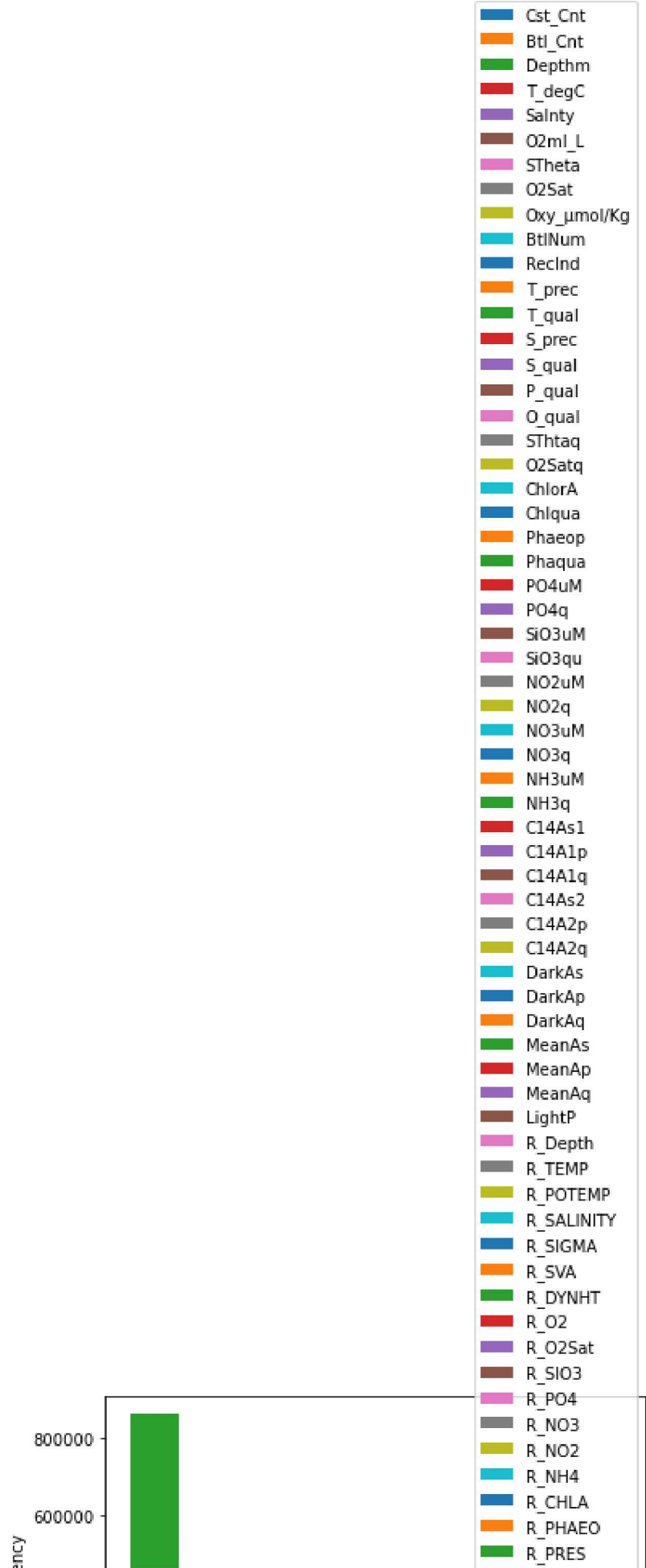


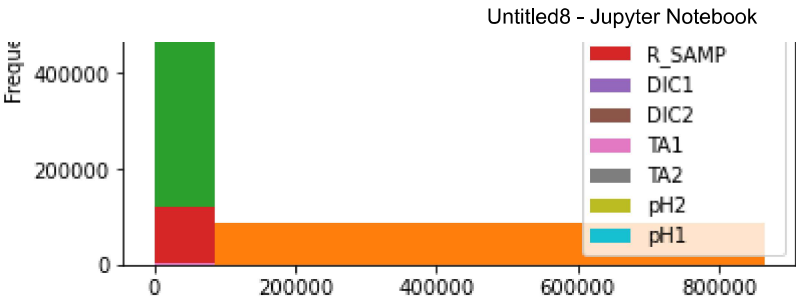


```
In [10]: df.plot.hist()
```

```
Out[10]: <AxesSubplot:ylabel='Frequency'>
```







In [ ]: