

Statistics

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

File directory

```
In [2]: data=pd.read_csv(r"C:\Users\user\Desktop\Vicky\9 bottle.csv")
```

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

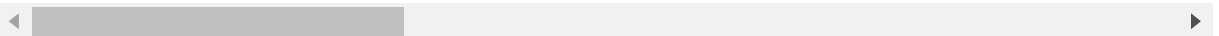
All Mathematical function

```
In [3]: data.describe()
```

```
Out[3]:
```

	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



To display the top portion of the dataset

In [4]:

data.head()

Out[4]:

	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

To display the mean median mode of the dataset for only numerical value

```
In [5]: data1=data[["Depthm","Btl_Cnt"]]
print(data1.mean())
print(data1.mode())
print(data1.median())
print()
```

```
Depthm      226.831951
Btl_Cnt     432432.000000
dtype: float64
```

```
      Depthm  Btl_Cnt
0         10.0        1
1          NaN        2
2          NaN        3
3          NaN        4
4          NaN        5
...         ...      ...
864858      NaN    864859
864859      NaN    864860
864860      NaN    864861
864861      NaN    864862
864862      NaN    864863
```

```
[864863 rows x 2 columns]
```

```
Cst_Cnt      16848.00000
Btl_Cnt     432432.00000
Depthm       125.00000
T_degC        10.06000
Salnty        33.86300
```

```
      ...
DIC2      2265.88500
TA1       2244.32500
TA2       2247.50500
pH2         7.94665
pH1         7.92885
Length: 70, dtype: float64
```

To display the total of each columns

```
In [6]: print(data1.sum())
```

```
Depthm      196178562
Btl_Cnt     373994436816
dtype: int64
```

To displayn the minimum value

```
In [7]: print(data1.min())
```

```
Depthm      0
Btl_Cnt      1
dtype: int64
```

To displayn theCummulative sum

In [8]: `print(data1.cumsum())`

	Depthm	Btl_Cnt
0	0	1
1	8	3
2	18	6
3	37	10
4	57	15
...
864858	196178530	373990977370
864859	196178532	373991842230
864860	196178537	373992707091
864861	196178547	373993571953
864862	196178562	373994436816

[864863 rows x 2 columns]

To count the total number of values in columns

In [9]: `print(data.count())`

Cst_Cnt	864863
Btl_Cnt	864863
Sta_ID	864863
Depth_ID	864863
Depthm	864863
...	...
TA1	2084
TA2	234
pH2	10
pH1	84
DIC Quality Comment	55
Length: 74, dtype: int64	

In [11]: `print(data1.cov())`

	Depthm	Btl_Cnt
Depthm	9.988777e+04	-1.275262e+07
Btl_Cnt	-1.275262e+07	6.233241e+10

In [12]: `from scipy.stats import spearmanr`
`from scipy.stats import pearsonr`

In [*]: `data2=[["Cst_Cnt", "R_PRES"]]`
`data2`

In [*]: `print(spearmanr(data1,data2))`

In []:

In []: