Day 6

bottle Dataset

In [1]:

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

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	
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

Mean,median,mode,describe

In [3]:

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

Out[3]:

	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 [4]:

```
print(data.mean())
```

```
Depthm     341.490000
T_degC       7.850421
dtype: float64
```

In [5]:

```
print(data.median())
```

```
Depthm      200.00
T_degC        8.18
dtype: float64
```

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	C
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297	1

In [34]:

```
data.fillna(value=1)
```

Out[34]:

	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 [8]:

```
print(data.mode())
```

```
    Depthm  T_degC
0        0   11.50
1       10   12.64
2       20     NaN
3       30     NaN
4       50     NaN
5       75     NaN
6      100     NaN
7      125     NaN
8      150     NaN
9      200     NaN
10     250     NaN
11     300     NaN
12     400     NaN
13     500     NaN
14     600     NaN
15     700     NaN
```

In [9]:

```
print(data.describe())
```

```
            Depthm      T_degC
count   500.000000  499.000000
mean    341.490000    7.850421
std     355.166886    2.911584
min       0.000000    2.780000
25%      55.000000    5.030000
50%     200.000000    8.180000
75%     598.500000   10.450000
max    1352.000000   12.660000
```

Sum,cumsum,count,min,max

In [10]:

```
print(data.sum())
```

```
Depthm    170745.00
T_degC      3917.36
dtype: float64
```

In [12]:

```
print(data.cumsum())
```

```
     Depthm   T_degC
0         0    10.50
1         8    20.96
2        18    31.42
3        37    41.87
4        57    52.32
..      ...      ...
495  167253  3900.22
496  168045  3904.72
497  168845  3909.20
498  169745  3913.41
499  170745  3917.36

[500 rows x 2 columns]
```

In [13]:

```
print(data.count())
```

```
Depthm    500
T_degC    499
dtype: int64
```

In [14]:

```
print(data.min())
```

```
Depthm    0.00
T_degC    2.78
dtype: float64
```

In [15]:

```
print(data.max())
```

```
Depthm    1352.00
T_degC      12.66
dtype: float64
```

covariance and correlation (spearman and pearsons)

In [44]:

```
data1=data['Depthm'][0:10]
data1
```

Out[44]:

```
0     0
1     8
2    10
3    19
4    20
5    30
6    39
7    50
8    58
9    75
Name: Depthm, dtype: int64
```

In [43]:

```
data2=data['T_degC'][0:10]
data2
```

Out[43]:

```
0    10.50
1    10.46
2    10.46
3    10.45
4    10.45
5    10.45
6    10.45
7    10.24
8    10.06
9     9.86
Name: T_degC, dtype: float64
```

In [46]:

```
from numpy import cov
print(cov(data1,data2))
```

```
[[ 5.87433333e+02 -4.73133333e+00]
 [-4.73133333e+00  4.68400000e-02]]
```

In [45]:

```
print(pearsonr(data1,data2))
```

```
(-0.9019777193232044, 0.00035830661732570235)
```

In [49]:

```
from scipy.stats import spearmanr
print(spearmanr(data1,data2))
```

```
SpearmanrResult(correlation=-0.9660917830792959, pvalue=5.551570051025844e
-06)
```

In [ ]:

```

```