

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

DATA COLLECTION

In [2]:

```
a=pd.read_csv(r"C:\Users\user\Downloads\9_bottle (1).csv")  
a
```

```
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 |

In [3]: Cst_Cnt Btl_Cnt Sta_ID Depth_ID Depthm T_degC Salnty O2ml_L STheta (

b=a.head(1000)

| | | | | | | | | | | | |
|--------|-------|--------|-------|------------|----|--------|---------|-------|----------|---|--|
| b | | | | 20-1611SR- | | | | | | | |
| 864862 | 34404 | 864863 | 093.4 | MX-310- | 15 | 17.533 | 33.3880 | 5.774 | 24.15297 | 1 | |
| | | | 026.4 | 2239- | | | | | | | |
| | | | | 09340264- | | | | | | | |
| | | | | 0015A-3 | | | | | | | |

864863 rows × 74 columns

Out[3]:

| 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 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 995 | 33 | 996 | 092.0 088.0 19-4903NS-HY-061-0906-09200880-0300A-7 | 300 | 7.22 | 34.040 | NaN | 26.636 | NaN |
| 996 | 33 | 997 | 092.0 088.0 19-4903NS-HY-061-0906-09200880-0379A-3 | 379 | 6.58 | 34.040 | NaN | 26.724 | NaN |
| 997 | 33 | 998 | 092.0 088.0 19-4903NS-HY-061-0906-09200880-0400A-7 | 400 | 6.44 | 34.049 | NaN | 26.750 | NaN |
| 998 | 33 | 999 | 092.0 088.0 19-4903NS-HY-061-0906-09200880-0500A-7 | 500 | 5.85 | 34.113 | NaN | 26.876 | NaN |

| Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | T_degC | Salnty | O2ml_L | STheta | O2Sat |
|---------|---------|--------|------------|--------|--------|--------|--------|--------|-------|
| 999 | 33 | 1000 | 19-4903NS- | 552 | 5.60 | 34.160 | NaN | 26.944 | NaN |
| | | | HY-061- | | | | | | |
| | | | 0906- | | | | | | |
| | | | 09200880- | | | | | | |

DATA CLEANING AND PRE-PROCESSING

1000 rows × 74 columns

In [4]:

```
b.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1000 entries, 0 to 999
```

```
Data columns (total 74 columns):
```

| # | Column | Non-Null Count | Dtype |
|----|-------------|----------------|---------|
| 0 | Cst_Cnt | 1000 non-null | int64 |
| 1 | Btl_Cnt | 1000 non-null | int64 |
| 2 | Sta_ID | 1000 non-null | object |
| 3 | Depth_ID | 1000 non-null | object |
| 4 | Depthm | 1000 non-null | int64 |
| 5 | T_degC | 998 non-null | float64 |
| 6 | Salnty | 970 non-null | float64 |
| 7 | O2ml_L | 0 non-null | float64 |
| 8 | STheta | 968 non-null | float64 |
| 9 | O2Sat | 0 non-null | float64 |
| 10 | Oxy_μmol/Kg | 0 non-null | float64 |
| 11 | BtlNum | 0 non-null | float64 |
| 12 | RecInd | 1000 non-null | int64 |
| 13 | T_prec | 998 non-null | float64 |
| 14 | T_qual | 10 non-null | float64 |
| 15 | S_prec | 970 non-null | float64 |
| 16 | S_qual | 45 non-null | float64 |
| 17 | P_qual | 1000 non-null | float64 |
| 18 | O_qual | 1000 non-null | float64 |
| 19 | SThtaq | 55 non-null | float64 |
| 20 | O2Satq | 1000 non-null | float64 |
| 21 | ChlorA | 0 non-null | float64 |
| 22 | Chlqua | 1000 non-null | float64 |
| 23 | Phaeop | 0 non-null | float64 |
| 24 | Phaqua | 1000 non-null | float64 |
| 25 | PO4uM | 0 non-null | float64 |
| 26 | PO4q | 1000 non-null | float64 |
| 27 | SiO3uM | 0 non-null | float64 |
| 28 | SiO3qu | 1000 non-null | float64 |
| 29 | NO2uM | 0 non-null | float64 |
| 30 | NO2q | 1000 non-null | float64 |
| 31 | NO3uM | 0 non-null | float64 |
| 32 | NO3q | 1000 non-null | float64 |
| 33 | NH3uM | 0 non-null | float64 |
| 34 | NH3q | 1000 non-null | float64 |
| 35 | C14As1 | 0 non-null | float64 |
| 36 | C14A1p | 0 non-null | float64 |
| 37 | C14A1q | 1000 non-null | float64 |
| 38 | C14As2 | 0 non-null | float64 |
| 39 | C14A2p | 0 non-null | float64 |
| 40 | C14A2q | 1000 non-null | float64 |
| 41 | DarkAs | 0 non-null | float64 |
| 42 | DarkAp | 0 non-null | float64 |
| 43 | DarkAq | 1000 non-null | float64 |
| 44 | MeanAs | 0 non-null | float64 |
| 45 | MeanAp | 0 non-null | float64 |
| 46 | MeanAq | 1000 non-null | float64 |
| 47 | IncTim | 0 non-null | object |
| 48 | LightP | 0 non-null | float64 |
| 49 | R_Depth | 1000 non-null | float64 |
| 50 | R_TEMP | 998 non-null | float64 |
| 51 | R_POTEMP | 962 non-null | float64 |
| 52 | R_SALINITY | 970 non-null | float64 |
| 53 | R_SIGMA | 945 non-null | float64 |
| 54 | R_SVA | 945 non-null | float64 |
| 55 | R_DYNHT | 973 non-null | float64 |


```
56 R_O2          0 non-null    float64
57 R_O2Sat       0 non-null    float64
58 R_SIO3        0 non-null    float64
59 R_PO4         0 non-null    float64
60 R_NO3         0 non-null    float64
61 R_NO2         0 non-null    float64
62 R_NH4         0 non-null    float64
63 R_CHLA        0 non-null    float64
64 R_PHAE0       0 non-null    float64
65 R_PRES        1000 non-null int64
66 R_SAMP        0 non-null    float64
67 DIC1          0 non-null    float64
68 DIC2          0 non-null    float64
69 TA1           0 non-null    float64
70 TA2           0 non-null    float64
71 pH2           0 non-null    float64
72 pH1           0 non-null    float64
73 DIC Quality Comment 0 non-null    object
```

dtypes: float64(65), int64(5), object(4)
memory usage: 578.2+ KB

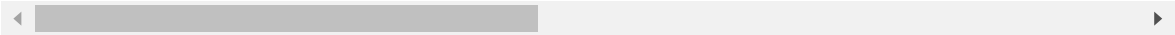
In [5]:

```
b.describe()
```

Out[5]:

| | Cst_Cnt | Btl_Cnt | Depthm | T_degC | Salnty | O2ml_L | STheta (|
|-------|-------------|-------------|-------------|------------|------------|--------|------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 998.000000 | 970.000000 | 0.0 | 968.000000 |
| mean | 16.803000 | 500.500000 | 329.604000 | 8.408657 | 33.668295 | NaN | 26.106232 |
| std | 9.500972 | 288.819436 | 346.635231 | 3.237212 | 0.509149 | NaN | 0.855427 |
| min | 1.000000 | 1.000000 | 0.000000 | 2.780000 | 32.630000 | NaN | 23.706000 |
| 25% | 9.000000 | 250.750000 | 50.000000 | 5.482500 | 33.220500 | NaN | 25.182250 |
| 50% | 16.000000 | 500.500000 | 189.500000 | 8.430000 | 33.748000 | NaN | 26.239000 |
| 75% | 25.000000 | 750.250000 | 515.250000 | 11.342500 | 34.108750 | NaN | 26.888000 |
| max | 33.000000 | 1000.000000 | 1352.000000 | 19.760000 | 34.650000 | NaN | 27.581000 |

8 rows × 70 columns



In [8]:

```
c=b.dropna(axis=1)  
c
```

Out[8]:

| Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | Reclnd | P_qual | O_qual | O2Satq | Chlqua |
|---------|---------|--------|--|--------|--------|--------|--------|--------|--------|
| 0 | 1 | 1 | 19-4903CR-HY-060-0930-05400560-0000A-3 | 0 | 3 | 9.0 | 9.0 | 9.0 | 9.0 |
| 1 | 1 | 2 | 19-4903CR-HY-060-0930-05400560-0008A-3 | 8 | 3 | 9.0 | 9.0 | 9.0 | 9.0 |
| 2 | 1 | 3 | 19-4903CR-HY-060-0930-05400560-0010A-7 | 10 | 7 | 9.0 | 9.0 | 9.0 | 9.0 |
| 3 | 1 | 4 | 19-4903CR-HY-060-0930-05400560-0019A-3 | 19 | 3 | 9.0 | 9.0 | 9.0 | 9.0 |
| 4 | 1 | 5 | 19-4903CR-HY-060-0930-05400560-0020A-7 | 20 | 7 | 9.0 | 9.0 | 9.0 | 9.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 995 | 33 | 996 | 19-4903NS-HY-061-0906-09200880-0300A-7 | 300 | 7 | 9.0 | 9.0 | 9.0 | 9.0 |
| 996 | 33 | 997 | 19-4903NS-HY-061-0906-09200880-0379A-3 | 379 | 3 | 9.0 | 9.0 | 9.0 | 9.0 |
| 997 | 33 | 998 | 19-4903NS-HY-061-0906-09200880-0400A-7 | 400 | 7 | 9.0 | 9.0 | 9.0 | 9.0 |
| 998 | 33 | 999 | 19-4903NS-HY-061-0906-09200880-0500A-7 | 500 | 7 | 9.0 | 9.0 | 9.0 | 9.0 |

| Cst_Cnt | Btl_Cnt | Sta_ID | Depth_ID | Depthm | RecInd | P_qual | O_qual | O2Satq | Chlqua |
|---------|---------|--------|----------------|--|--------|--------|--------|--------|--------|
| 999 | 33 | 1000 | 092.0 088.0 | 19- 4903NS- HY-061- 0906- 09200880- 0552A-3 | 552 | 3 | 9.0 | 9.0 | 9.0 |

In [9]:

```
c.columns
```

1000 rows × 22 columns

Out[9]:

```
Index(['Cst_Cnt', 'Btl_Cnt', 'Sta_ID', 'Depth_ID', 'Depthm', 'RecInd',  
      'P_qual', 'O_qual', 'O2Satq', 'Chlqua', 'Phaqua', 'PO4q', 'SiO3qu',  
      'NO2q', 'NO3q', 'NH3q', 'C14A1q', 'C14A2q', 'DarkAq', 'MeanAq',  
      'R_Depth', 'R_PRES'],  
      dtype='object')
```

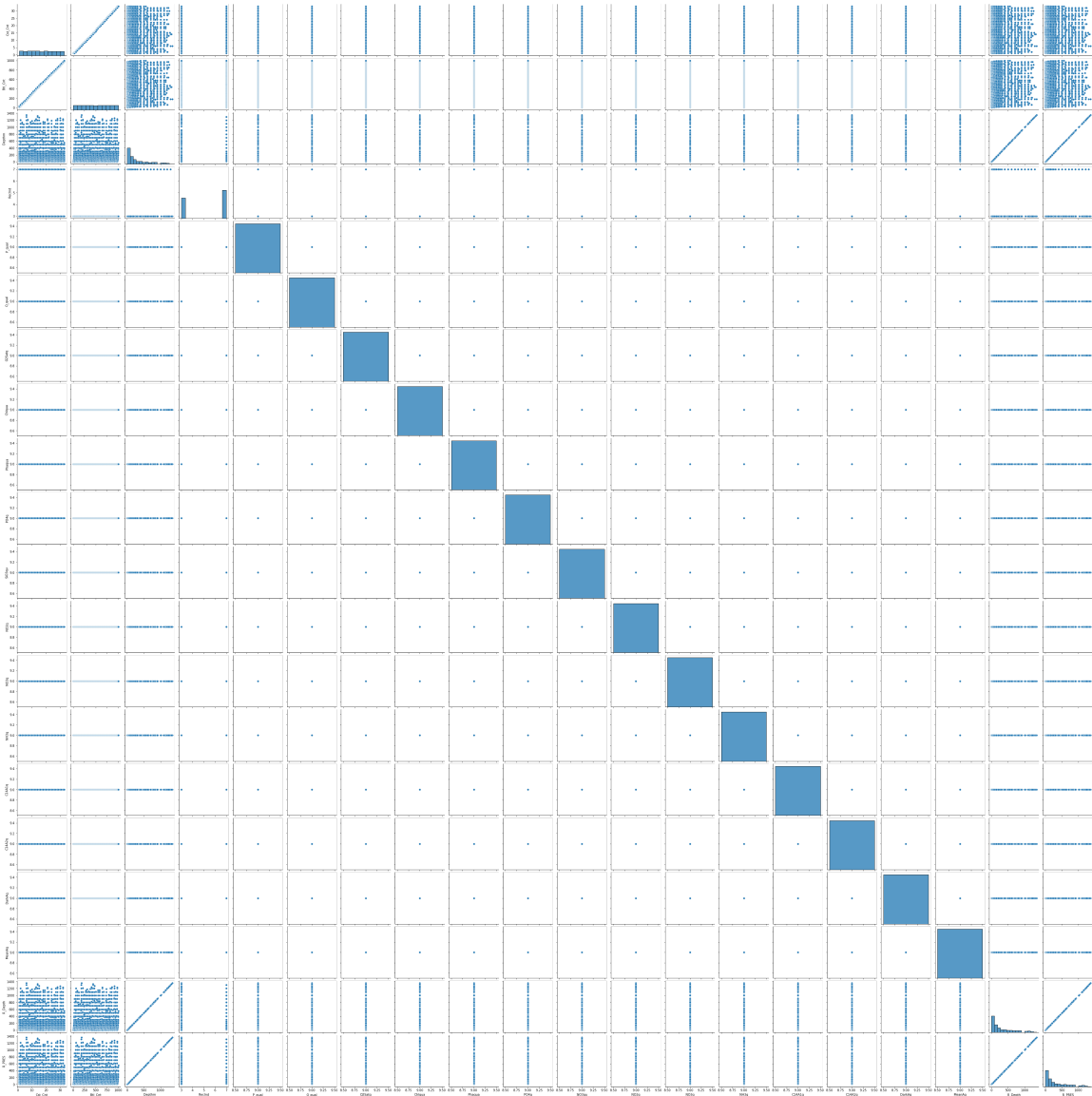
EDA AND VISUALIZATION

In [12]:

```
sns.pairplot(c)
```

Out[12]:

<seaborn.axisgrid.PairGrid at 0x1fb930062e0>



In [14]:

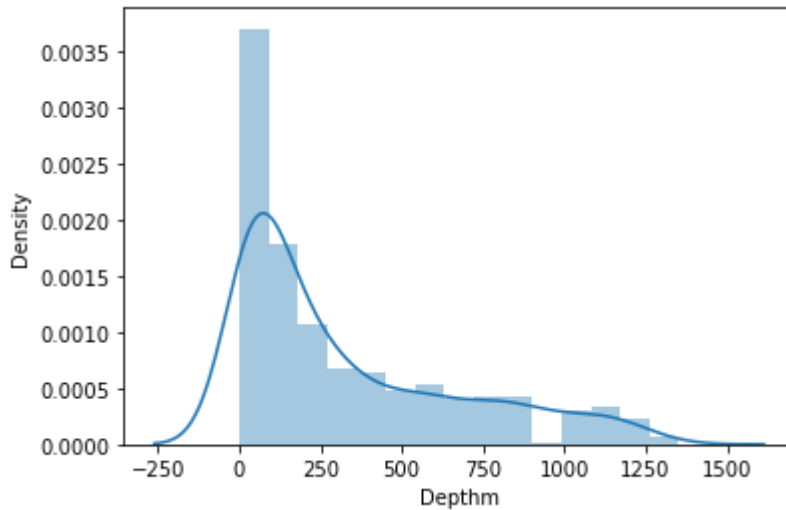
```
sns.distplot(c['Depthm'])
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in
a future version. Please adapt your code to use either `displot` (a figure
-level function with similar flexibility) or `histplot` (an axes-level fun
ction for histograms).

```
warnings.warn(msg, FutureWarning)
```

Out[14]:

<AxesSubplot:xlabel='Depthm', ylabel='Density'>



In [15]:

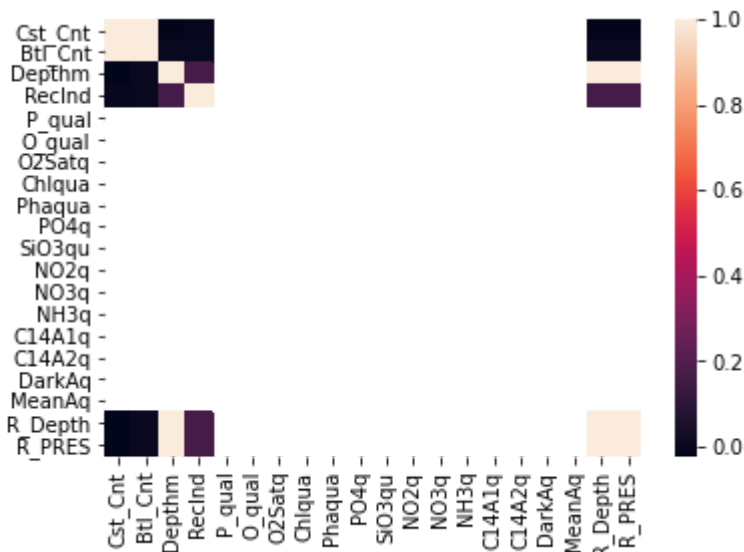
```
f=c[['Cst_Cnt', 'Btl_Cnt', 'Sta_ID', 'Depth_ID', 'Depthm', 'RecInd',  
     'P_qual', 'O_qual', 'O2Satq', 'Chlqua', 'Phaqua', 'PO4q', 'SiO3qu',  
     'NO2q', 'NO3q', 'NH3q', 'C14A1q', 'C14A2q', 'DarkAq', 'MeanAq',  
     'R_Depth', 'R_PRES']]
```

In [16]:

```
sns.heatmap(f.corr())
```

Out[16]:

<AxesSubplot:>



In [22]:

```
x=f[['Cst_Cnt', 'Btl_Cnt', 'RecInd',
      'P_qual', 'O_qual', 'O2Satq', 'Chlqua', 'Phaqua', 'PO4q', 'SiO3qu',
      'NO2q', 'NO3q', 'NH3q', 'C14A1q', 'C14A2q', 'DarkAq', 'MeanAq',
      'R_Depth', 'R_PRES']]
y=f['Depthm']
```

In [23]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.5)
```

In [24]:

```
from sklearn.linear_model import LinearRegression

lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[24]:

LinearRegression()

In [25]:

```
print(lr.intercept_)
```

2.8421709430404007e-13

In [26]:

```
r=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
r
```

Out[26]:

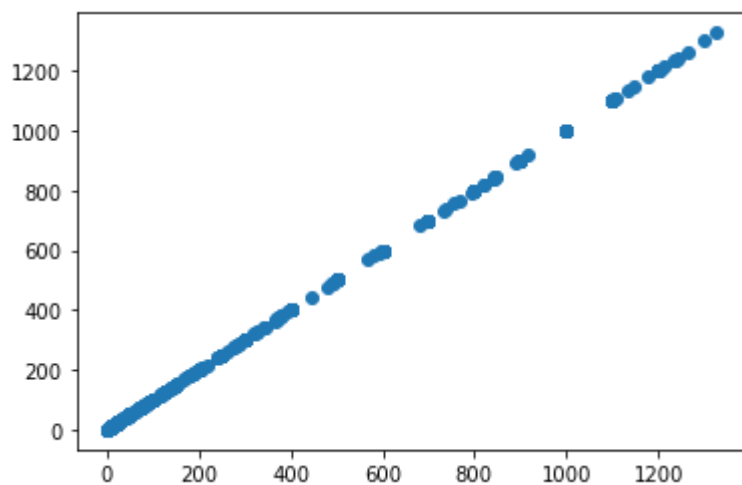
| | Co-efficient |
|---------|---------------|
| Cst_Cnt | -3.642866e-13 |
| Btl_Cnt | 1.225698e-14 |
| RecInd | -8.752498e-15 |
| P_qual | 0.000000e+00 |
| O_qual | 0.000000e+00 |
| O2Satq | 0.000000e+00 |
| Chlqua | 0.000000e+00 |
| Phaqua | 0.000000e+00 |
| PO4q | 0.000000e+00 |
| SiO3qu | 0.000000e+00 |
| NO2q | 0.000000e+00 |
| NO3q | 0.000000e+00 |
| NH3q | 0.000000e+00 |
| C14A1q | 0.000000e+00 |
| C14A2q | 0.000000e+00 |
| DarkAq | 0.000000e+00 |
| MeanAq | 0.000000e+00 |
| R_Depth | 1.000000e+00 |
| R_PRES | -8.039266e-15 |

In [27]:

```
u=lr.predict(x_test)
plt.scatter(y_test,u)
```

Out[27]:

<matplotlib.collections.PathCollection at 0x1fba6df4e20>



In [28]:

```
print(lr.score(x_test,y_test))
```

1.0

In []: