Importing Libraries

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

Importing Datasets

In [2]:
```
df=pd.read_csv("madrid_2001.csv")
df
```

Out[2]:

	date	BEN	CO	EBE	MXY	NMHC	NO_2	NOx	OXY	O_3	PM10	
0	2001-08-01 01:00:00	NaN	0.37	NaN	NaN	NaN	58.400002	87.150002	NaN	34.529999	105.000000	N
1	2001-08-01 01:00:00	1.50	0.34	1.49	4.10	0.07	56.250000	75.169998	2.11	42.160000	100.599998	
2	2001-08-01 01:00:00	NaN	0.28	NaN	NaN	NaN	50.660000	61.380001	NaN	46.310001	100.099998	N
3	2001-08-01 01:00:00	NaN	0.47	NaN	NaN	NaN	69.790001	73.449997	NaN	40.650002	69.779999	N
4	2001-08-01 01:00:00	NaN	0.39	NaN	NaN	NaN	22.830000	24.799999	NaN	66.309998	75.180000	N
...	...	...	...	...	...	...	...	...	...	...	...	
217867	2001-04-01 00:00:00	10.45	1.81	NaN	NaN	NaN	73.000000	264.399994	NaN	5.200000	47.880001	N
217868	2001-04-01 00:00:00	5.20	0.69	4.56	NaN	0.13	71.080002	129.300003	NaN	13.460000	26.809999	N
217869	2001-04-01 00:00:00	0.49	1.09	NaN	1.00	0.19	76.279999	128.399994	0.35	5.020000	40.770000	
217870	2001-04-01 00:00:00	5.62	1.01	5.04	11.38	NaN	80.019997	197.000000	2.58	5.840000	37.889999	
217871	2001-04-01 00:00:00	8.09	1.62	6.66	13.04	0.18	76.809998	206.300003	5.20	8.340000	35.369999	

217872 rows × 16 columns

Data Cleaning and Data Preprocessing

In [3]:
```
df=df.dropna()
```

In [4]:
```
df.columns
```

Out[4]:
```
Index(['date', 'BEN', 'CO', 'EBE', 'MXY', 'NMHC', 'NO_2', 'NOx', 'OXY', 'O_3',
       'PM10', 'PXY', 'SO_2', 'TCH', 'TOL', 'station'],
      dtype='object')
```

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 29669 entries, 1 to 217871
Data columns (total 16 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   date     29669 non-null  object 
 1   BEN      29669 non-null  float64
 2   CO       29669 non-null  float64
 3   EBE      29669 non-null  float64
 4   MXY      29669 non-null  float64
 5   NMHC     29669 non-null  float64
 6   NO_2     29669 non-null  float64
 7   NOx      29669 non-null  float64
 8   OXY      29669 non-null  float64
 9   O_3      29669 non-null  float64
 10  PM10     29669 non-null  float64
 11  PXY      29669 non-null  float64
 12  SO_2     29669 non-null  float64
 13  TCH      29669 non-null  float64
 14  TOL      29669 non-null  float64
 15  station  29669 non-null  int64  
dtypes: float64(14), int64(1), object(1)
memory usage: 3.8+ MB
```

In [6]:
```
data=df[['CO' ,'station']]
data
```

Out[6]:

	CO	station
1	0.34	28079035
5	0.63	28079006
21	0.43	28079024
23	0.34	28079099
25	0.06	28079035
...	...	...
217829	4.48	28079006

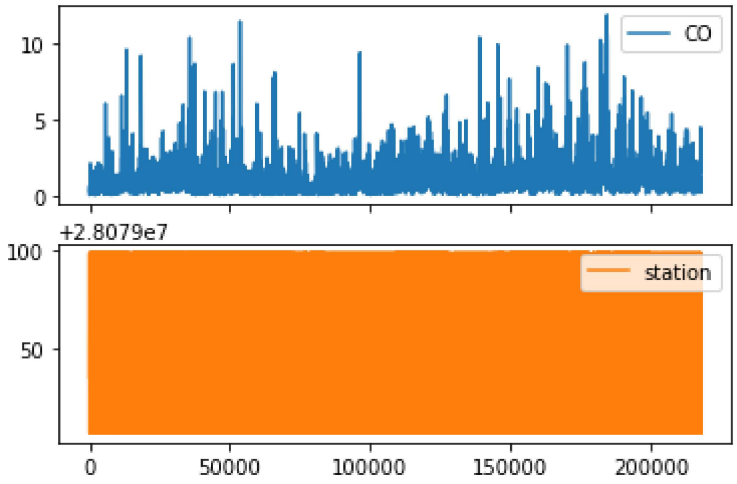
	CO	station
217847	2.65	28079099
217849	1.22	28079035
217853	1.83	28079006
217871	1.62	28079099

29669 rows × 2 columns

Line chart

In [7]:
```
data.plot.line(subplots=True)
```

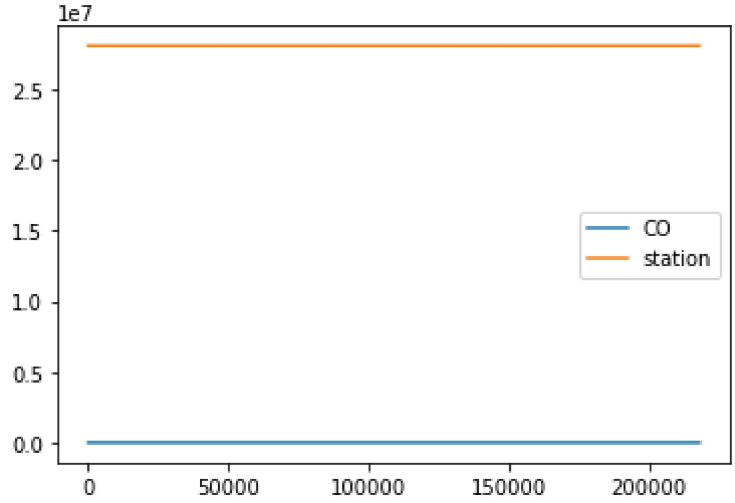
Out[7]:
```
array([<AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



Line chart

In [8]:
```
data.plot.line()
```

Out[8]:
```
<AxesSubplot:>
```

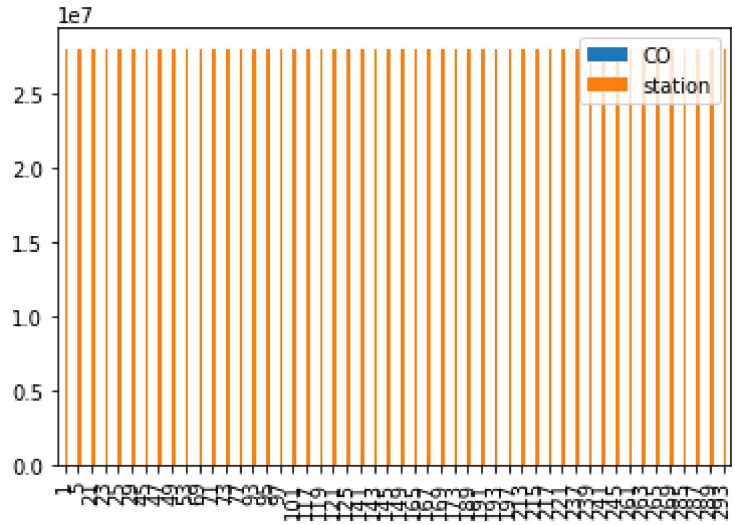


Bar chart

In [9]:
```
b=data[0:50]
```

In [10]:
```
b.plot.bar()
```

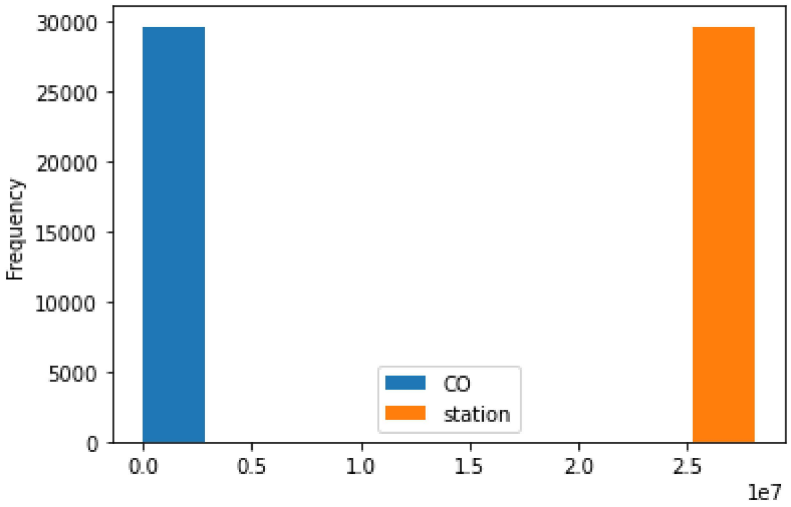
Out[10]: <AxesSubplot:>



Histogram

In [11]:
```
data.plot.hist()
```

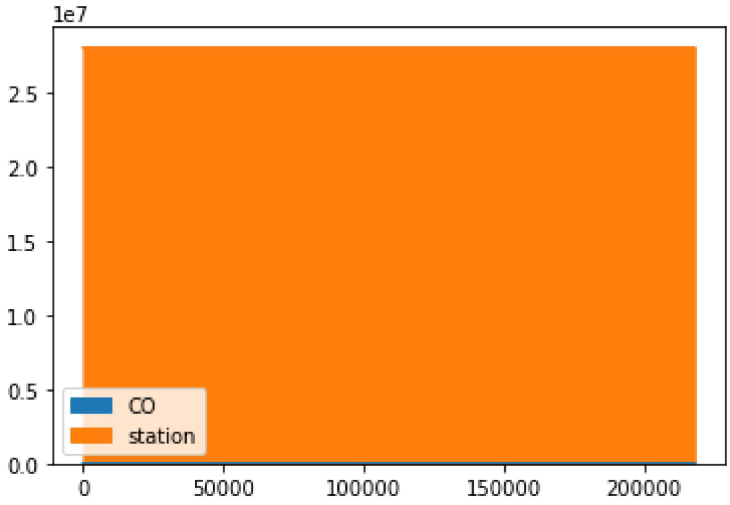
Out[11]: <AxesSubplot:ylabel='Frequency'>



Area chart

In [12]:
```
data.plot.area()
```

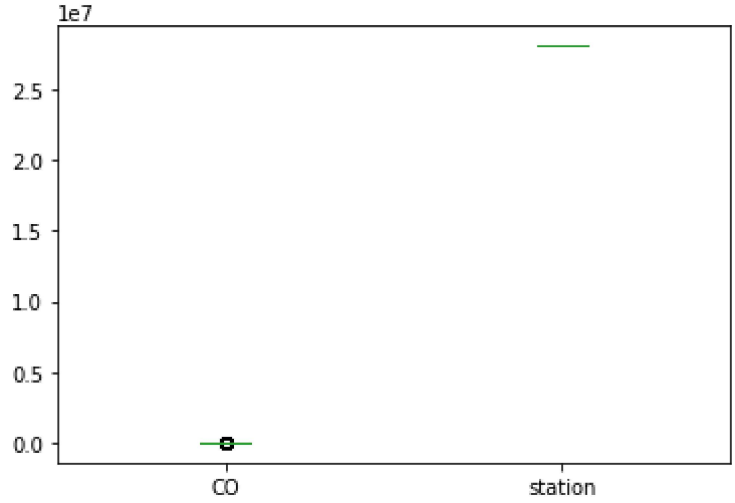
Out[12]: <AxesSubplot:>



Box chart

In [13]:
```
data.plot.box()
```

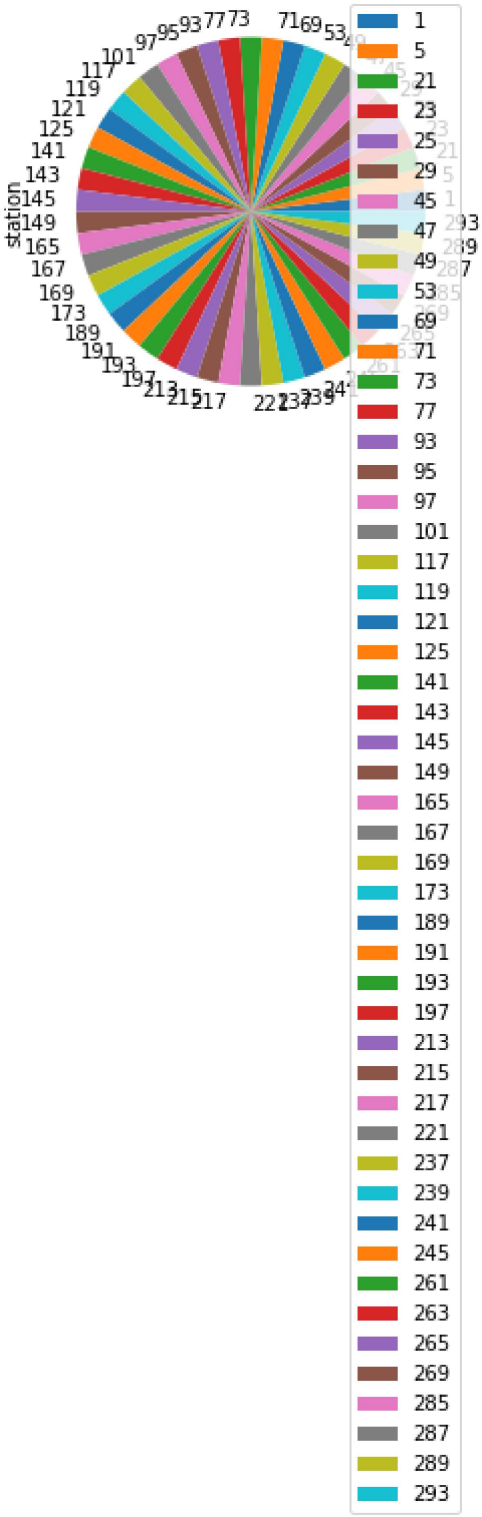
Out[13]: <AxesSubplot:>



Pie chart

In [14]:
```
b.plot.pie(y='station' )
```

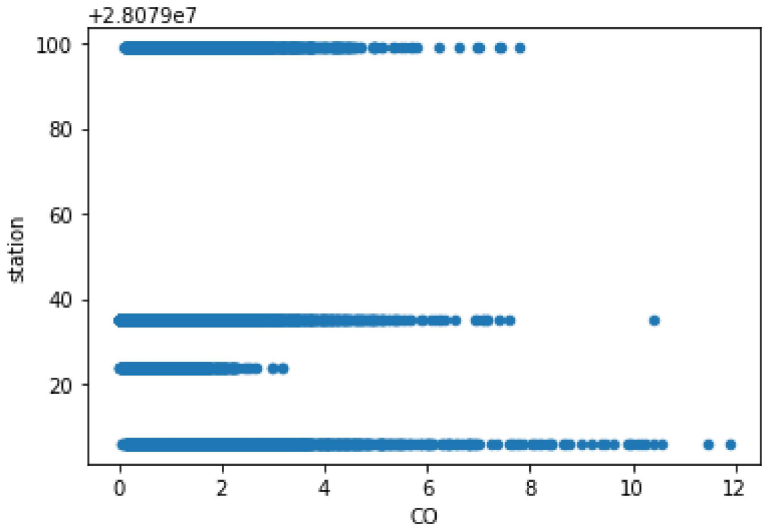
Out[14]:
```
<AxesSubplot:ylabel='station'>
```



Scatter chart

In [15]:
```
data.plot.scatter(x='CO' ,y='station')
```

Out[15]: <AxesSubplot:xlabel='CO', ylabel='station'>



In [16]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 29669 entries, 1 to 217871
Data columns (total 16 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   date     29669 non-null  object 
 1   BEN      29669 non-null  float64
 2   CO       29669 non-null  float64
 3   EBE      29669 non-null  float64
 4   MXY      29669 non-null  float64
 5   NMHC     29669 non-null  float64
 6   NO_2     29669 non-null  float64
 7   NOx      29669 non-null  float64
 8   OXY      29669 non-null  float64
 9   O_3      29669 non-null  float64
 10  PM10     29669 non-null  float64
 11  PXY      29669 non-null  float64
 12  SO_2     29669 non-null  float64
 13  TCH      29669 non-null  float64
 14  TOL      29669 non-null  float64
 15  station  29669 non-null  int64  
dtypes: float64(14), int64(1), object(1)
memory usage: 3.8+ MB
```

In [17]:
```
df.describe()
```

Out[17]:

	BEN	CO	EBE	MXY	NMHC	NO_2	NO
count	29669.000000	29669.000000	29669.000000	29669.000000	29669.000000	29669.000000	29669.00000
mean	3.361895	1.005413	3.580229	8.113086	0.195222	67.652292	163.00485
std	3.176669	0.863135	3.744496	7.909701	0.192585	34.003120	147.49177
min	0.100000	0.000000	0.140000	0.210000	0.000000	1.180000	1.28000
25%	1.280000	0.470000	1.390000	3.040000	0.080000	44.299999	68.08999
50%	2.510000	0.760000	2.600000	5.830000	0.140000	64.449997	123.69999
75%	4.420000	1.270000	4.580000	10.640000	0.250000	86.540001	213.19999

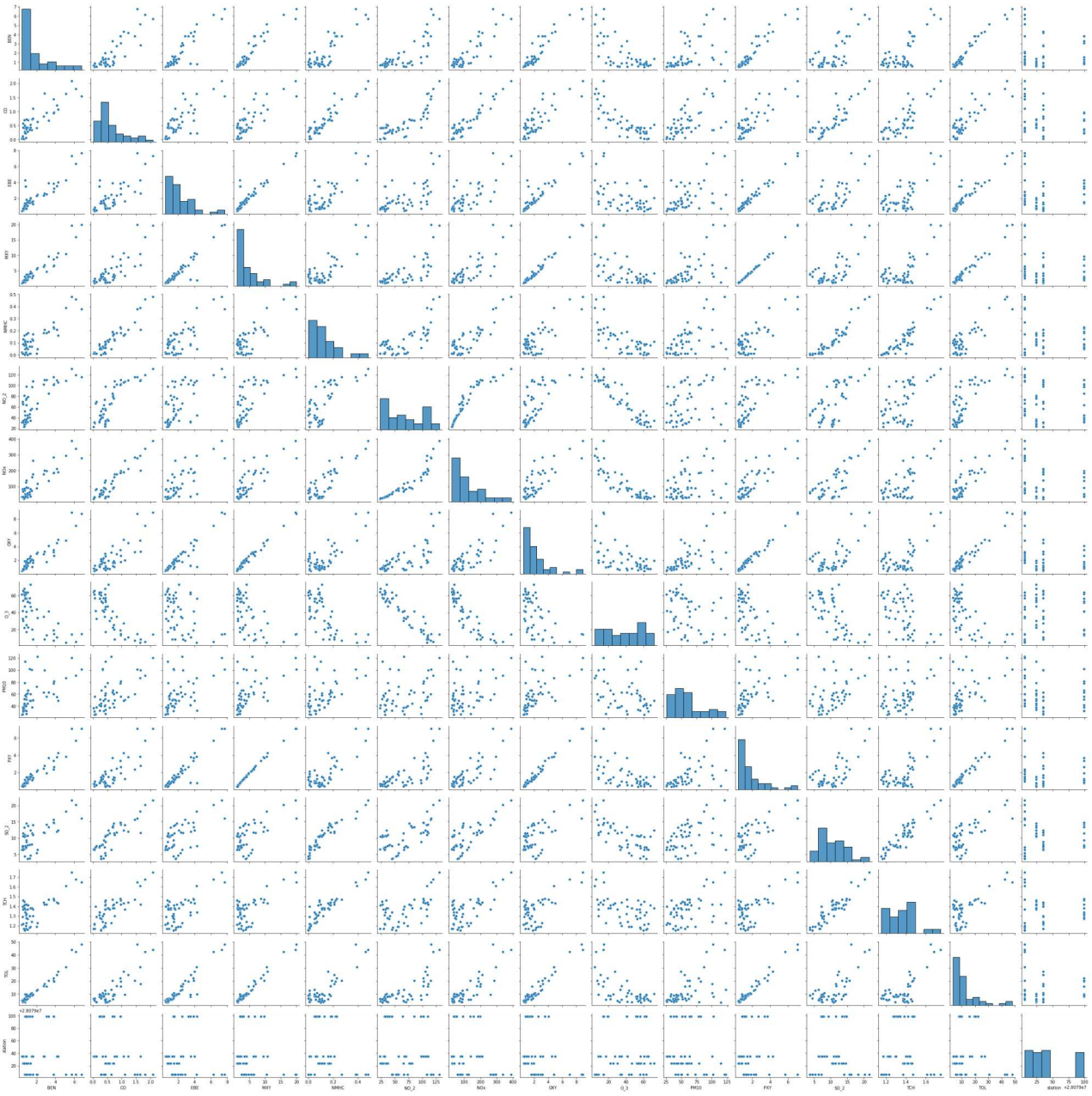
	BEN	CO	EBE	MXY	NMHC	NO_2	NO
max	54.560001	11.890000	77.260002	150.600006	2.880000	292.700012	1940.00000

In [18]:
```
df1=df[['BEN', 'CO', 'EBE', 'MXY', 'NMHC', 'NO_2', 'NOx', 'OXY', 'O_3',
       'PM10', 'PXY', 'SO_2', 'TCH', 'TOL', 'station']]
```

EDA AND VISUALIZATION

In [19]:
```
sns.pairplot(df1[0:50])
```

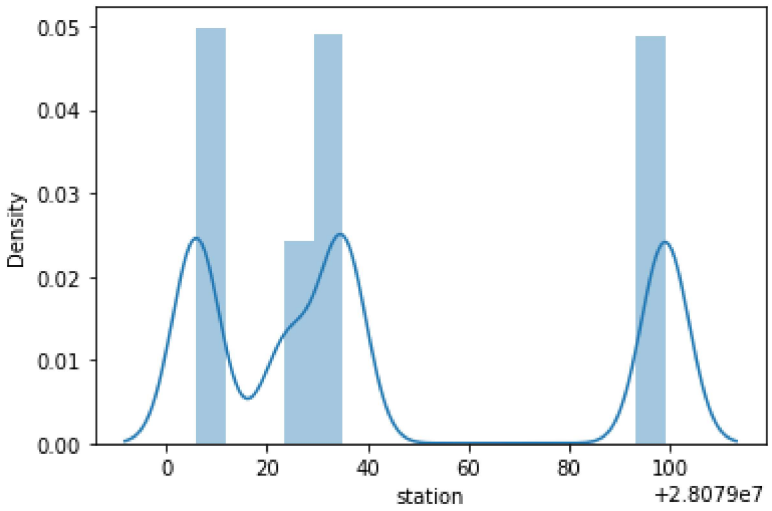
Out[19]: <seaborn.axisgrid.PairGrid at 0x147d48b2490>



In [20]:
```
sns.distplot(df1['station'])
```

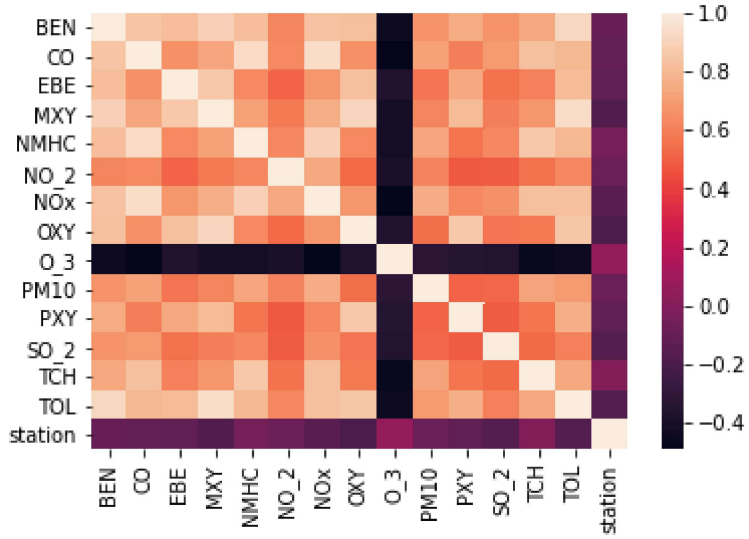
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning:
`distplot` is a deprecated function and will be removed in a future version. Please adap
t your code to use either `displot` (a figure-level function with similar flexibility) o
r `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[20]: <AxesSubplot:xlabel='station', ylabel='Density'>



In [21]:
```
sns.heatmap(df1.corr())
```

Out[21]: <AxesSubplot:>



TO TRAIN THE MODEL AND MODEL BULDING

In [22]:
```
x=df[['BEN', 'CO', 'EBE', 'MXY', 'NMHC', 'NO_2', 'NOx', 'OXY', 'O_3',
       'PM10', 'PXY', 'SO_2', 'TCH', 'TOL']]
y=df['station']
```

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=0.3)
```

Linear Regression

In [24]:
```
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[24]:
```
LinearRegression()
```

In [25]:
```
lr.intercept_
```

Out[25]:
```
28079003.438920423
```

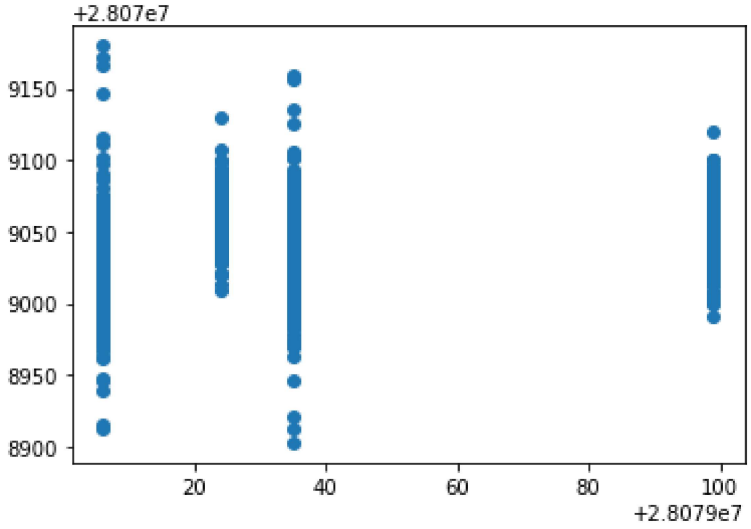
In [26]:
```
coeff=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
coeff
```

Out[26]:

	Co-efficient
BEN	7.075677
CO	-13.965353
EBE	0.667502
MXY	-0.073467
NMHC	77.806104
NO_2	0.125633
NOx	-0.093374
OXY	-3.046324
O_3	-0.028070
PM10	-0.050195
PXY	1.272918
SO_2	-0.302553
TCH	39.753855
TOL	-1.225787

In [27]:
```
prediction =lr.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[27]:
```
<matplotlib.collections.PathCollection at 0x147e3fe9580>
```



ACCURACY

In [28]:
```
lr.score(x_test,y_test)
```

Out[28]: 0.15635997591267914

In [29]:
```
lr.score(x_train,y_train)
```

Out[29]: 0.16831590812021358

Ridge and Lasso

In [30]:
```
from sklearn.linear_model import Ridge,Lasso
```

In [31]:
```
rr=Ridge(alpha=10)
rr.fit(x_train,y_train)
```

Out[31]: Ridge(alpha=10)

Accuracy(Ridge)

In [32]:
```
rr.score(x_test,y_test)
```

Out[32]: 0.1557138659362748

In [33]:
```
rr.score(x_train,y_train)
```

Out[33]: 0.16809856624085817

Accuracy(Lasso)

In [34]:
```
la=Lasso(alpha=10)
la.fit(x_train,y_train)
```

Out[34]:
```
Lasso(alpha=10)
```

In [35]:
```
la.score(x_train,y_train)
```

Out[35]:
```
0.039641720127249425
```

Accuracy(ElasticNet)

In [36]:
```
la.score(x_test,y_test)
```

Out[36]:
```
0.03959231329198354
```

In [37]:
```
from sklearn.linear_model import ElasticNet
en=ElasticNet()
en.fit(x_train,y_train)
```

Out[37]:
```
ElasticNet()
```

In [38]:
```
en.coef_
```

Out[38]:
```
array([ 5.00888111,  0.        ,  0.67831515, -0.30022739,  0.08150739,
        0.06357557, -0.03139477, -2.47602946, -0.03612483,  0.08255532,
        0.8134146 , -0.32427339,  1.23699365, -0.68984559])
```

In [39]:
```
en.intercept_
```

Out[39]:
```
28079048.922354154
```

In [40]:
```
prediction=en.predict(x_test)
```

In [41]:
```
en.score(x_test,y_test)
```

Out[41]:
```
0.09743364213542827
```

Evaluation Metrics

In [42]:
```
from sklearn import metrics
print(metrics.mean_absolute_error(y_test,prediction))
```

```
print(metrics.mean_squared_error(y_test,prediction))
print(np.sqrt(metrics.mean_squared_error(y_test,prediction)))
```

```
30.29926288849478
1206.0630969111587
34.728419153643586
```

Logistic Regression

In [43]:
```
from sklearn.linear_model import LogisticRegression
```

In [44]:
```
feature_matrix=df[['BEN', 'CO', 'EBE', 'MXY', 'NMHC', 'NO_2', 'NOx', 'OXY', 'O_3',
       'PM10', 'PXY', 'SO_2', 'TCH', 'TOL']]
target_vector=df[ 'station']
```

In [45]:
```
feature_matrix.shape
```

Out[45]: (29669, 14)

In [46]:
```
target_vector.shape
```

Out[46]: (29669,)

In [47]:
```
from sklearn.preprocessing import StandardScaler
```

In [48]:
```
fs=StandardScaler().fit_transform(feature_matrix)
```

In [49]:
```
logr=LogisticRegression(max_iter=10000)
logr.fit(fs,target_vector)
```

Out[49]: LogisticRegression(max_iter=10000)

In [50]:
```
observation=[[1,2,3,4,5,6,7,8,9,10,11,12,13,14]]
```

In [51]:
```
prediction=logr.predict(observation)
print(prediction)
```

```
[28079035]
```

In [52]:
```
logr.classes_
```

Out[52]: array([28079006, 28079024, 28079035, 28079099], dtype=int64)

In [53]:
```
logr.score(fs,target_vector)
```

Out[53]:
```
0.8087229094340894
```

In [54]:
```
logr.predict_proba(observation)[0][0]
```

Out[54]:
```
1.724527777144498e-43
```

In [55]:
```
logr.predict_proba(observation)
```

Out[55]:
```
array([[1.72452778e-43, 2.43756289e-56, 9.99998565e-01, 1.43537418e-06]])
```

Random Forest

In [56]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [57]:
```
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[57]:
```
RandomForestClassifier()
```

In [58]:
```
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
            'n_estimators':[10,20,30,40,50]
}
```

In [59]:
```
from sklearn.model_selection import GridSearchCV
grid_search =GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

Out[59]:
```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2, 3, 4, 5],
                         'min_samples_leaf': [5, 10, 15, 20, 25],
                         'n_estimators': [10, 20, 30, 40, 50]},
             scoring='accuracy')
```

In [60]:
```
grid_search.best_score_
```

Out[60]:
```
0.7291987673343605
```

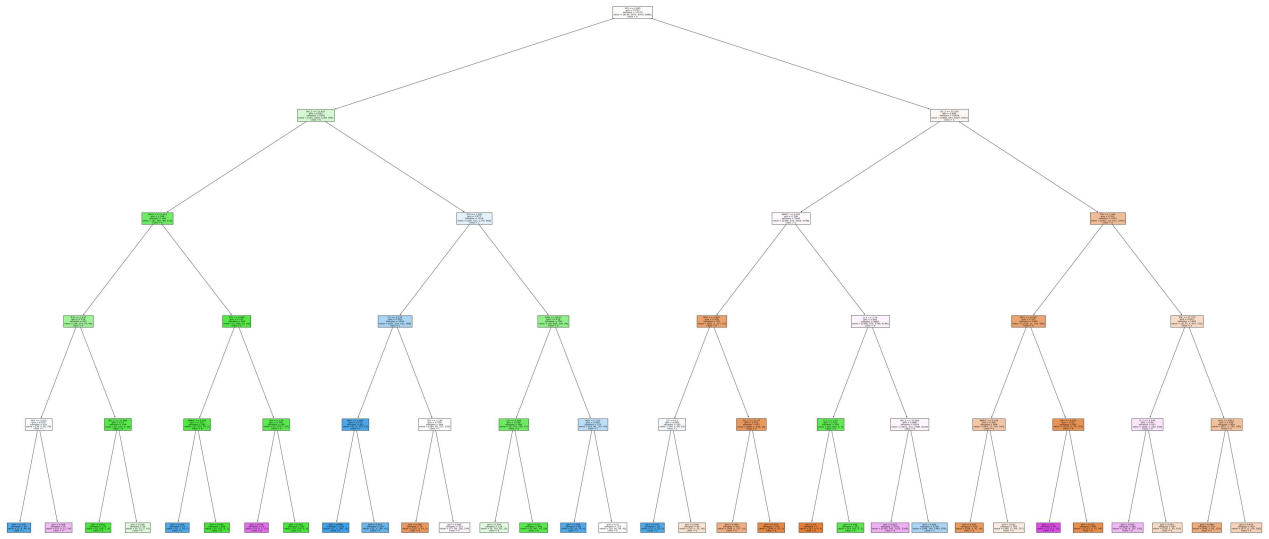
In [61]:
```
rfc_best=grid_search.best_estimator_
```

In [62]:
```
from sklearn.tree import plot_tree

plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['a','b','c','d']
```

Out[62]:
```
[Text(2232.0, 1993.2, 'PXY <= 1.005\ngini = 0.733\nsamples = 13170\nvalue = [6136, 2872,
5874, 5886]\nclass = a'),
 Text(1116.0, 1630.8000000000002, 'NO_2 <= 21.425\ngini = 0.651\nsamples = 2542\nvalue =
[247, 1878, 1258, 594]\nclass = b'),
 Text(558.0, 1268.4, 'PM10 <= 11.815\ngini = 0.344\nsamples = 786\nvalue = [41, 967, 88,
112]\nclass = b'),
 Text(279.0, 906.0, 'TCH <= 1.235\ngini = 0.578\nsamples = 301\nvalue = [36, 279, 63, 8
4]\nclass = b'),
 Text(139.5, 543.5999999999999, 'OXY <= 0.625\ngini = 0.675\nsamples = 103\nvalue = [36,
6, 60, 59]\nclass = c'),
 Text(69.75, 181.19999999999982, 'gini = 0.226\nsamples = 36\nvalue = [1, 6, 48, 0]\ncla
ss = c'),
 Text(209.25, 181.19999999999982, 'gini = 0.568\nsamples = 67\nvalue = [35, 0, 12, 59]\n
class = d'),
 Text(418.5, 543.5999999999999, 'NO_2 <= 17.865\ngini = 0.17\nsamples = 198\nvalue = [0,
273, 3, 25]\nclass = b'),
 Text(348.75, 181.19999999999982, 'gini = 0.053\nsamples = 170\nvalue = [0, 250, 1, 6]\n
class = b'),
 Text(488.25, 181.19999999999982, 'gini = 0.538\nsamples = 28\nvalue = [0, 23, 2, 19]\nc
lass = b'),
 Text(837.0, 906.0, 'PXY <= 0.645\ngini = 0.147\nsamples = 485\nvalue = [5, 688, 25, 28]
\nclass = b'),
 Text(697.5, 543.5999999999999, 'NMHC <= 0.035\ngini = 0.112\nsamples = 285\nvalue = [0,
414, 24, 2]\nclass = b'),
 Text(627.75, 181.19999999999982, 'gini = 0.203\nsamples = 19\nvalue = [0, 2, 24, 1]\ncl
ass = c'),
 Text(767.25, 181.19999999999982, 'gini = 0.005\nsamples = 266\nvalue = [0, 412, 0, 1]\n
class = b'),
 Text(976.5, 543.5999999999999, 'TCH <= 1.24\ngini = 0.191\nsamples = 200\nvalue = [5, 2
74, 1, 26]\nclass = b'),
 Text(906.75, 181.19999999999982, 'gini = 0.359\nsamples = 19\nvalue = [5, 0, 1, 21]\ncl
ass = d'),
 Text(1046.25, 181.19999999999982, 'gini = 0.035\nsamples = 181\nvalue = [0, 274, 0, 5]
\nclass = b'),
 Text(1674.0, 1268.4, 'TCH <= 1.305\ngini = 0.677\nsamples = 1756\nvalue = [206, 911, 11
70, 482]\nclass = c'),
 Text(1395.0, 906.0, 'CO <= 0.235\ngini = 0.593\nsamples = 1000\nvalue = [187, 102, 912,
388]\nclass = c'),
 Text(1255.5, 543.5999999999999, 'MXY <= 1.665\ngini = 0.173\nsamples = 412\nvalue = [3,
4, 575, 53]\nclass = c'),
 Text(1185.75, 181.19999999999982, 'gini = 0.068\nsamples = 253\nvalue = [1, 1, 385, 12]
\nclass = c'),
 Text(1325.25, 181.19999999999982, 'gini = 0.321\nsamples = 159\nvalue = [2, 3, 190, 41]
\nclass = c'),
 Text(1534.5, 543.5999999999999, 'TCH <= 1.195\ngini = 0.704\nsamples = 588\nvalue = [18
4, 98, 337, 335]\nclass = c'),
 Text(1464.75, 181.19999999999982, 'gini = 0.268\nsamples = 73\nvalue = [107, 0, 19, 1]
\nclass = a'),
 Text(1604.25, 181.19999999999982, 'gini = 0.666\nsamples = 515\nvalue = [77, 98, 318, 3
34]\nclass = d'),
 Text(1953.0, 906.0, 'NOx <= 94.23\ngini = 0.476\nsamples = 756\nvalue = [19, 809, 258,
94]\nclass = b'),
 Text(1813.5, 543.5999999999999, 'CO <= 0.385\ngini = 0.326\nsamples = 584\nvalue = [0,
743, 126, 51]\nclass = b'),
 Text(1743.75, 181.19999999999982, 'gini = 0.594\nsamples = 136\nvalue = [0, 102, 83, 2
5]\nclass = b'),
 Text(1883.25, 181.19999999999982, 'gini = 0.18\nsamples = 448\nvalue = [0, 641, 43, 26]
\nclass = b'),
 Text(2092.5, 543.5999999999999, 'MXY <= 1.03\ngini = 0.645\nsamples = 172\nvalue = [19,
66, 132, 43]\nclass = c'),
 Text(2022.75, 181.19999999999982, 'gini = 0.206\nsamples = 58\nvalue = [0, 10, 76, 0]\n
class = c'),
 Text(2162.25, 181.19999999999982, 'gini = 0.72\nsamples = 114\nvalue = [19, 56, 56, 43]
\nclass = b'),
 Text(3348.0, 1630.8000000000002, 'SO_2 <= 25.295\ngini = 0.699\nsamples = 10628\nvalue
```

```
= [5889, 994, 4616, 5292]\nclass = a'),
 Text(2790.0, 1268.4, 'NMHC <= 0.045\ngini = 0.706\nsamples = 7649\nvalue = [2958, 974,
4005, 4198]\nclass = d'),
 Text(2511.0, 906.0, 'MXY <= 2.925\ngini = 0.356\nsamples = 762\nvalue = [962, 0, 217, 5
2]\nclass = a'),
 Text(2371.5, 543.5999999999999, 'CO <= 0.2\ngini = 0.658\nsamples = 105\nvalue = [54,
0, 58, 39]\nclass = c'),
 Text(2301.75, 181.19999999999982, 'gini = 0.197\nsamples = 25\nvalue = [1, 0, 33, 3]\nc
lass = c'),
 Text(2441.25, 181.19999999999982, 'gini = 0.636\nsamples = 80\nvalue = [53, 0, 25, 36]
\nclass = a'),
 Text(2650.5, 543.5999999999999, 'SO_2 <= 11.27\ngini = 0.271\nsamples = 657\nvalue = [9
08, 0, 159, 13]\nclass = a'),
 Text(2580.75, 181.19999999999982, 'gini = 0.407\nsamples = 287\nvalue = [364, 0, 127, 1
0]\nclass = a'),
 Text(2720.25, 181.19999999999982, 'gini = 0.114\nsamples = 370\nvalue = [544, 0, 32, 3]
\nclass = a'),
 Text(3069.0, 906.0, 'O_3 <= 3.74\ngini = 0.693\nsamples = 6887\nvalue = [1996, 974, 378
8, 4146]\nclass = d'),
 Text(2929.5, 543.5999999999999, 'O_3 <= 1.24\ngini = 0.222\nsamples = 358\nvalue = [65,
502, 0, 7]\nclass = b'),
 Text(2859.75, 181.19999999999982, 'gini = 0.0\nsamples = 7\nvalue = [11, 0, 0, 0]\nclas
s = a'),
 Text(2999.25, 181.19999999999982, 'gini = 0.196\nsamples = 351\nvalue = [54, 502, 0, 7]
\nclass = b'),
 Text(3208.5, 543.5999999999999, 'NO_2 <= 73.065\ngini = 0.668\nsamples = 6529\nvalue =
[1931, 472, 3788, 4139]\nclass = d'),
 Text(3138.75, 181.19999999999982, 'gini = 0.624\nsamples = 3791\nvalue = [1143, 287, 13
23, 3169]\nclass = d'),
 Text(3278.25, 181.19999999999982, 'gini = 0.605\nsamples = 2738\nvalue = [788, 185, 246
5, 970]\nclass = c'),
 Text(3906.0, 1268.4, 'TCH <= 1.565\ngini = 0.531\nsamples = 2979\nvalue = [2931, 20, 61
1, 1094]\nclass = a'),
 Text(3627.0, 906.0, 'MXY <= 8.765\ngini = 0.392\nsamples = 1494\nvalue = [1758, 11, 17
8, 368]\nclass = a'),
 Text(3487.5, 543.5999999999999, 'NMHC <= 0.105\ngini = 0.565\nsamples = 589\nvalue = [5
44, 11, 125, 243]\nclass = a'),
 Text(3417.75, 181.19999999999982, 'gini = 0.305\nsamples = 206\nvalue = [264, 0, 41, 1
6]\nclass = a'),
 Text(3557.25, 181.19999999999982, 'gini = 0.622\nsamples = 383\nvalue = [280, 11, 84, 2
27]\nclass = a'),
 Text(3766.5, 543.5999999999999, 'OXY <= 4.035\ngini = 0.23\nsamples = 905\nvalue = [121
4, 0, 53, 125]\nclass = a'),
 Text(3696.75, 181.19999999999982, 'gini = 0.19\nsamples = 20\nvalue = [1, 0, 2, 26]\ncl
ass = d'),
 Text(3836.25, 181.19999999999982, 'gini = 0.201\nsamples = 885\nvalue = [1213, 0, 51, 9
9]\nclass = a'),
 Text(4185.0, 906.0, 'TOL <= 37.275\ngini = 0.619\nsamples = 1485\nvalue = [1173, 9, 43
3, 726]\nclass = a'),
 Text(4045.5, 543.5999999999999, 'O_3 <= 9.145\ngini = 0.64\nsamples = 627\nvalue = [35
6, 7, 192, 434]\nclass = d'),
 Text(3975.75, 181.19999999999982, 'gini = 0.625\nsamples = 360\nvalue = [138, 6, 137, 2
92]\nclass = d'),
 Text(4115.25, 181.19999999999982, 'gini = 0.591\nsamples = 267\nvalue = [218, 1, 55, 14
2]\nclass = a'),
 Text(4324.5, 543.5999999999999, 'NOx <= 435.1\ngini = 0.556\nsamples = 858\nvalue = [81
7, 2, 241, 292]\nclass = a'),
 Text(4254.75, 181.19999999999982, 'gini = 0.446\nsamples = 379\nvalue = [440, 2, 63, 11
0]\nclass = a'),
 Text(4394.25, 181.19999999999982, 'gini = 0.619\nsamples = 479\nvalue = [377, 0, 178, 1
82]\nclass = a')]
```



Conclusion

Accuracy

In [64]:
```
lr.score(x_train,y_train)
```

Out[64]: 0.16831590812021358

In [65]:
```
rr.score(x_train,y_train)
```

Out[65]: 0.16809856624085817

In [66]:
```
la.score(x_train,y_train)
```

Out[66]: 0.039641720127249425

In [67]:
```
en.score(x_test,y_test)
```

Out[67]: 0.09743364213542827

In [68]:
```
logr.score(fs,target_vector)
```

Out[68]: 0.8087229094340894

In [69]:
```
grid_search.best_score_
```

Out[69]: 0.7291987673343605

Logistic Regression is suitable for this dataset