03/08/2023 (P6)

In [161]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [163]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\madrid_2006.csv")
df
```

Out[163]:

	date	BEN	CO	EBE	MXY	NMHC	NO_2	NOx	OXY	O_3	PM10	PM25	PXY	SO_2	TCH
0	2006-02-01 01:00:00	NaN	1.84	NaN	NaN	NaN	155.100006	490.100006	NaN	4.880000	97.570000	40.259998	NaN	33.779999	NaN
1	2006-02-01 01:00:00	1.68	1.01	2.38	6.36	0.32	94.339996	229.699997	3.04	7.100000	25.820000	NaN	2.48	11.890000	1.59
2	2006-02-01 01:00:00	NaN	1.25	NaN	NaN	NaN	66.800003	192.000000	NaN	4.430000	34.419998	NaN	NaN	19.719999	NaN
3	2006-02-01 01:00:00	NaN	1.68	NaN	NaN	NaN	103.000000	407.799988	NaN	4.830000	28.260000	NaN	NaN	21.129999	NaN
4	2006-02-01 01:00:00	NaN	1.31	NaN	NaN	NaN	105.400002	269.200012	NaN	6.990000	54.180000	NaN	NaN	11.050000	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
230563	2006-05-01 00:00:00	5.88	0.83	6.23	NaN	0.20	112.500000	218.000000	NaN	24.389999	93.120003	NaN	NaN	7.400000	1.50
230564	2006-05-01 00:00:00	0.76	0.32	0.48	1.09	0.08	51.900002	54.820000	0.61	48.410000	29.469999	15.640000	0.50	8.840000	1.32
230565	2006-05-01 00:00:00	0.96	NaN	0.69	NaN	0.19	135.100006	179.199997	NaN	11.460000	64.680000	35.000000	NaN	12.110000	1.51
230566	2006-05-01 00:00:00	0.50	NaN	0.67	NaN	0.10	82.599998	105.599998	NaN	NaN	94.360001	NaN	NaN	4.890000	1.47
230567	2006-05-01 00:00:00	1.95	0.74	1.99	4.00	0.24	107.300003	160.199997	2.01	17.730000	52.490002	27.920000	1.70	9.170000	1.50

230568 rows × 17 columns

In [169]:
```
df=df.dropna()
```

In [170]:
```
df.columns
```

Out[170]:
```
Index(['date', 'BEN', 'CO', 'EBE', 'MXY', 'NMHC', 'NO_2', 'NOx', 'OXY', 'O_3',
       'PM10', 'PM25', 'PXY', 'SO_2', 'TCH', 'TOL', 'station'],
      dtype='object')
```

In [171]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 24758 entries, 5 to 230567
Data columns (total 17 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   date     24758 non-null  object 
 1   BEN      24758 non-null  float64
 2   CO       24758 non-null  float64
 3   EBE      24758 non-null  float64
 4   MXY      24758 non-null  float64
 5   NMHC     24758 non-null  float64
 6   NO_2     24758 non-null  float64
 7   NOx      24758 non-null  float64
 8   OXY      24758 non-null  float64
 9   O_3      24758 non-null  float64
 10  PM10     24758 non-null  float64
 11  PM25     24758 non-null  float64
 12  PXY      24758 non-null  float64
 13  SO_2     24758 non-null  float64
 14  TCH      24758 non-null  float64
 15  TOL      24758 non-null  float64
 16  station  24758 non-null  int64  
dtypes: float64(15), int64(1), object(1)
memory usage: 3.4+ MB
```

In [172]:
```
data=df[['EBE', 'MXY', 'PXY']]
data
```

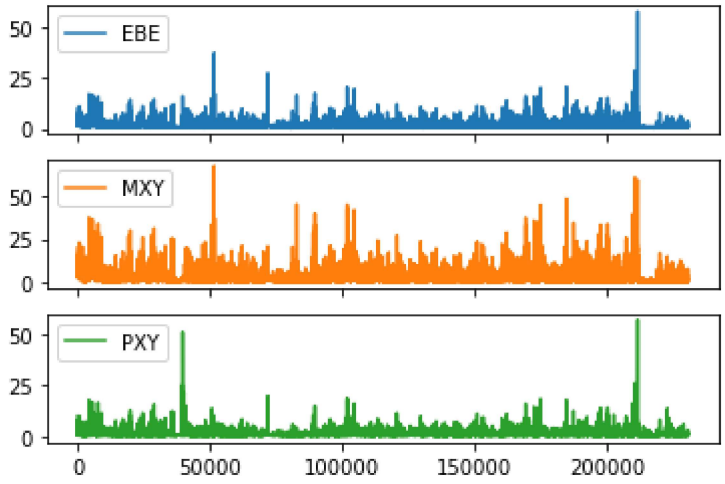
Out[172]:

	EBE	MXY	PXY
5	9.98	19.959999	10.11
22	1.24	2.670000	0.79
25	2.64	9.660000	4.46
31	7.92	17.139999	8.06
48	1.24	2.740000	0.82
...	...	...	...
230538	0.37	0.430000	1.00
230541	1.53	2.200000	1.21
230547	3.71	7.960000	3.36
230564	0.48	1.090000	0.50
230567	1.99	4.000000	1.70

24758 rows × 3 columns

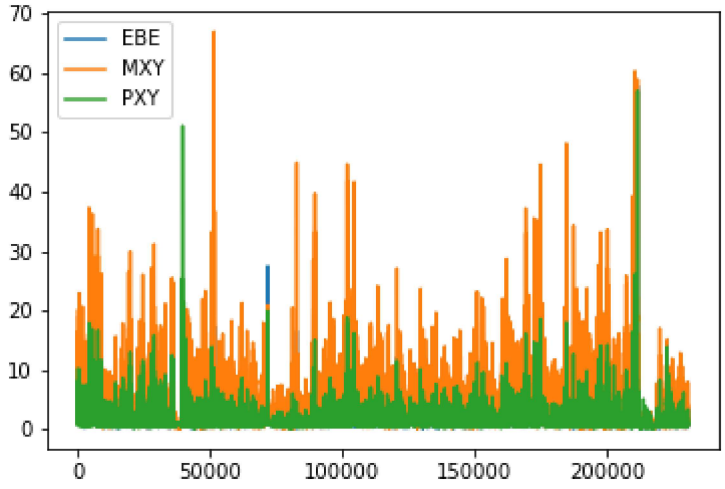
In [173]:
```
data.plot.line(subplots=True)
```

Out[173]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



In [174]:
```
data.plot.line()
```

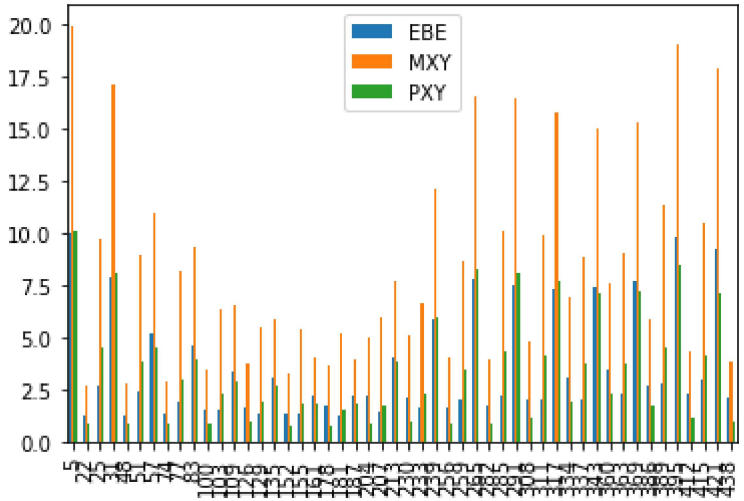
Out[174]:
```
<AxesSubplot:>
```



In [175]:
```
b=data[0:50]
```

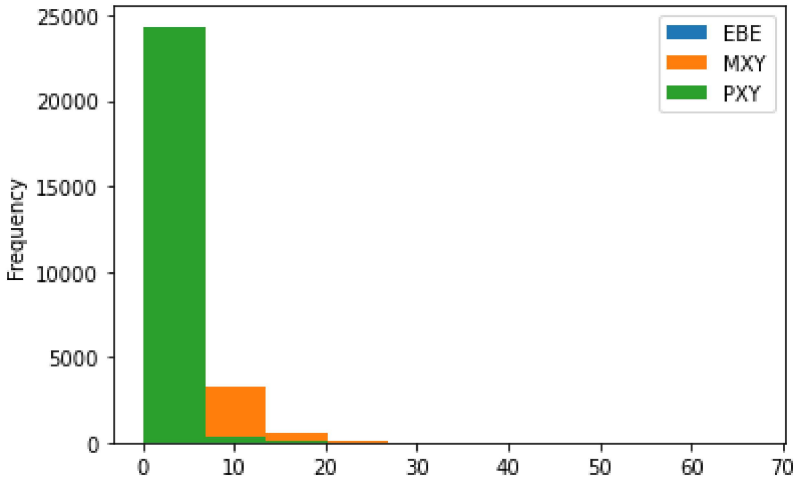
In [176]:
```
b.plot.bar()
```

Out[176]: <AxesSubplot:>



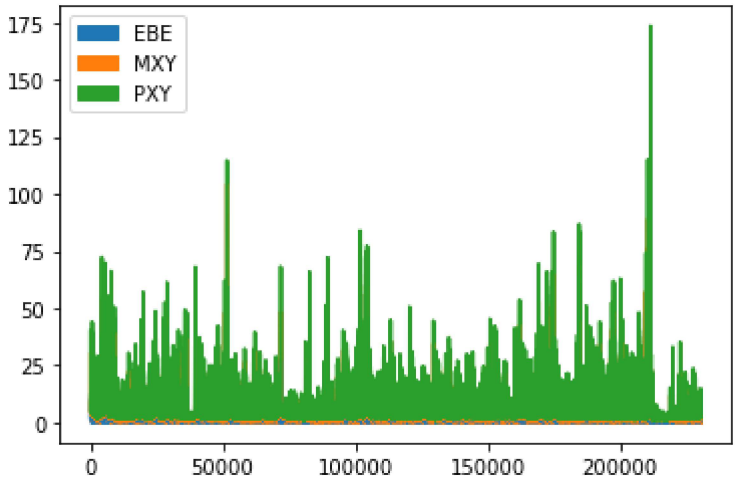
In [177]:
```
data.plot.hist()
```

Out[177]: <AxesSubplot:ylabel='Frequency'>



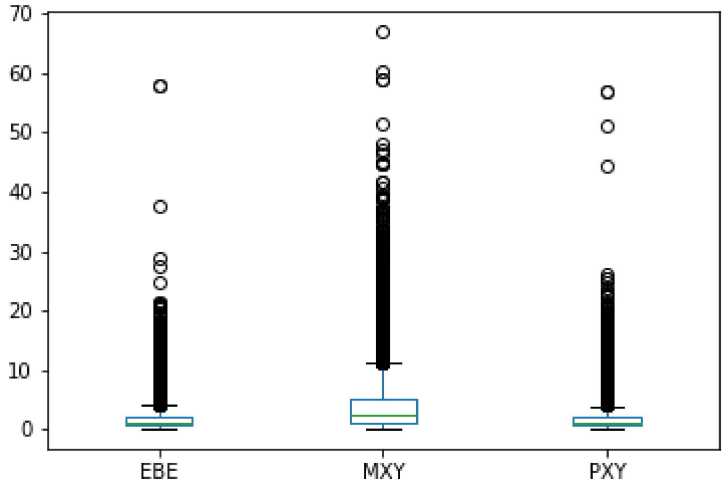
In [178]:
```
data.plot.area()
```

Out[178]: <AxesSubplot:>



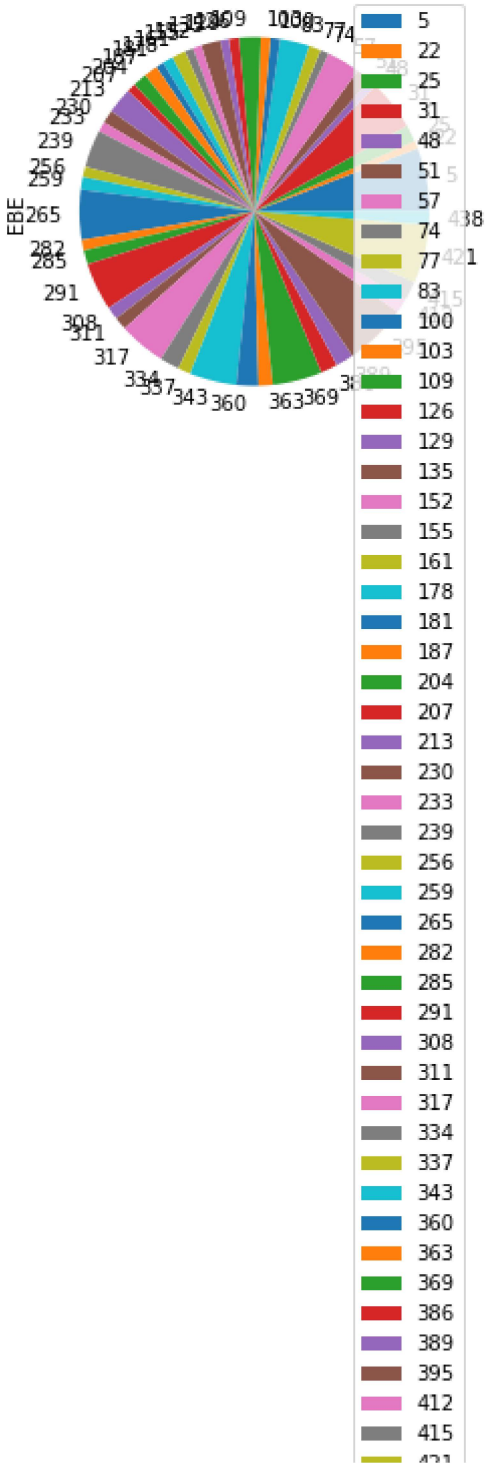
In [179]:
```
data.plot.box()
```

Out[179]: <AxesSubplot:>



In [180]:
```
b.plot.pie(y='EBE' )
```

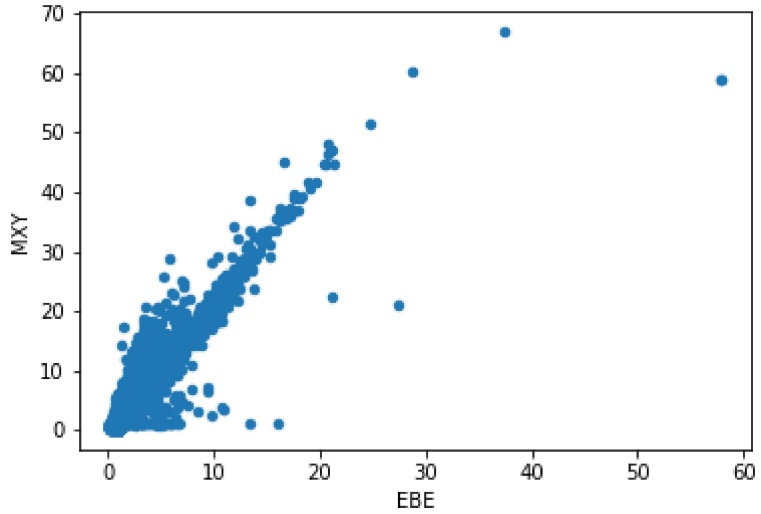
Out[180]:
```
<AxesSubplot:ylabel='EBE'>
```


EBE
5
22
25
31
48
51
57
74
77
83
100
103
109
126
129
135
152
155
161
178
181
187
204
207
213
230
233
239
256
259
265
282
285
291
308
311
317
334
337
343
360
363
369
386
389
395
412
415
421

421
438

In [181]:
```
data.plot.scatter(x='EBE' ,y='MXY')
```

Out[181]: <AxesSubplot:xlabel='EBE', ylabel='MXY'>



In [182]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 24758 entries, 5 to 230567
Data columns (total 17 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   date     24758 non-null  object 
 1   BEN      24758 non-null  float64
 2   CO       24758 non-null  float64
 3   EBE      24758 non-null  float64
 4   MXY      24758 non-null  float64
 5   NMHC     24758 non-null  float64
 6   NO_2     24758 non-null  float64
 7   NOx      24758 non-null  float64
 8   OXY      24758 non-null  float64
 9   O_3      24758 non-null  float64
 10  PM10     24758 non-null  float64
 11  PM25     24758 non-null  float64
 12  PXY      24758 non-null  float64
 13  SO_2     24758 non-null  float64
 14  TCH      24758 non-null  float64
 15  TOL      24758 non-null  float64
 16  station  24758 non-null  int64  
dtypes: float64(15), int64(1), object(1)
memory usage: 3.4+ MB
```

In [183]:
```
df.describe()
```

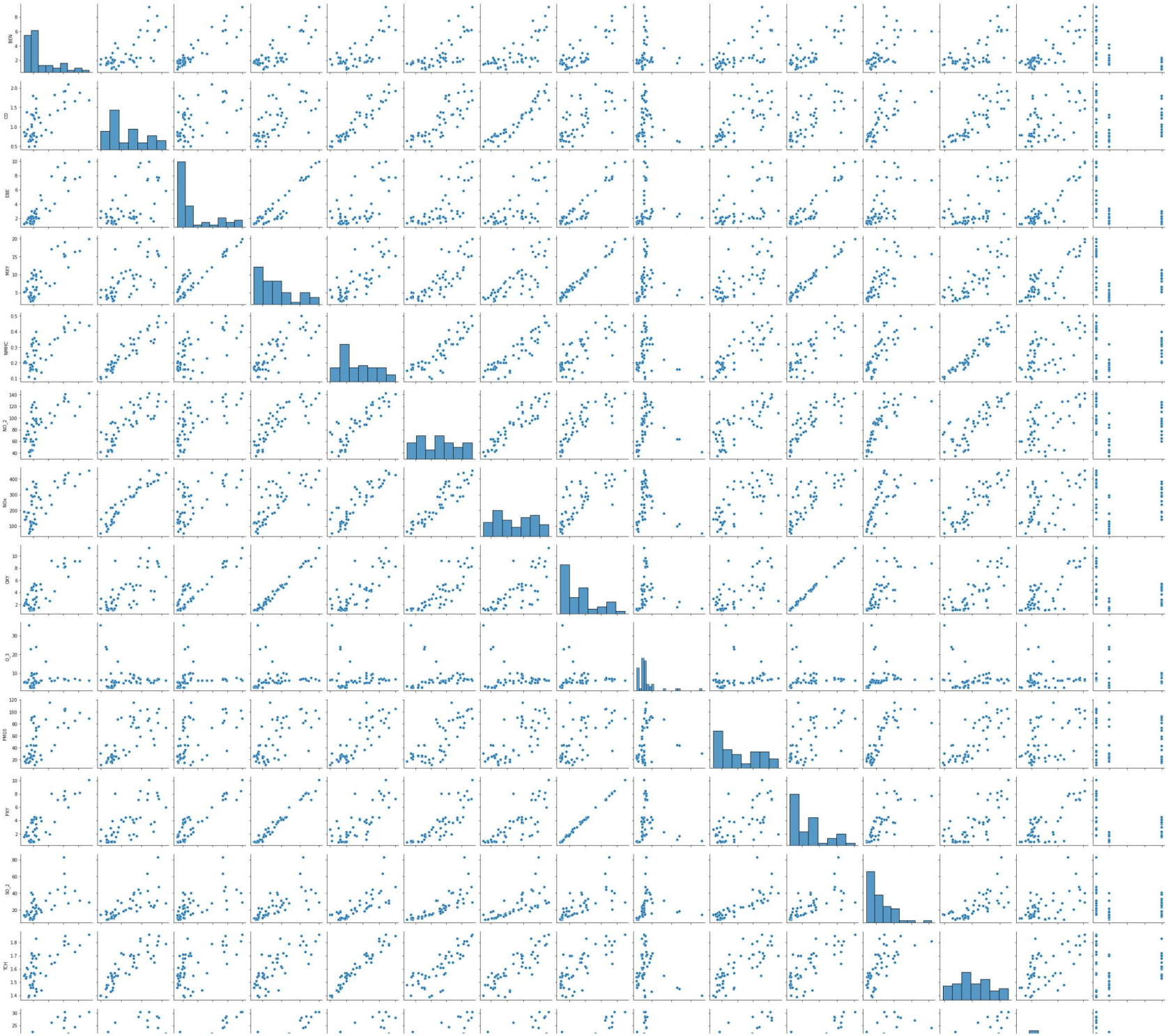
Out[183]:

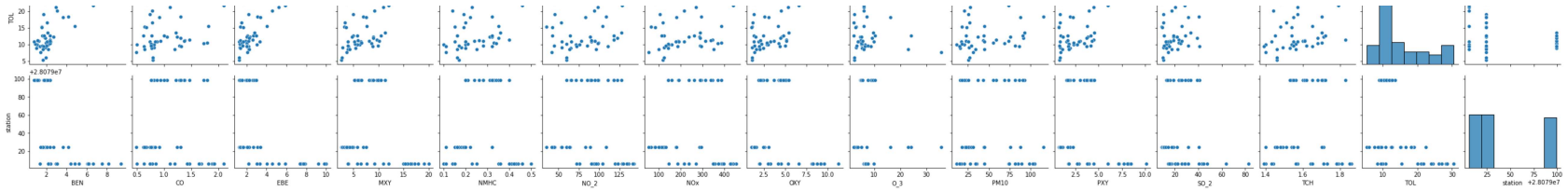
	BEN	CO	EBE	MXY	NMHC	NO_2	NOx	OXY	O_3
count	24758.000000	24758.000000	24758.000000	24758.000000	24758.000000	24758.000000	24758.000000	24758.000000	24758.000000
mean	1.350624	0.600713	1.824534	3.835034	0.176546	58.333481	116.419090	1.990347	37.155685
std	1.541636	0.419048	1.868939	4.069036	0.126683	40.529382	117.557064	1.931620	31.127681
min	0.110000	0.000000	0.170000	0.150000	0.000000	1.680000	2.020000	0.190000	0.310000
25%	0.450000	0.360000	0.810000	1.060000	0.100000	28.450001	36.882501	0.960000	12.110000
50%	0.850000	0.500000	1.130000	2.500000	0.150000	52.959999	85.180000	1.260000	28.675000
75%	1.680000	0.720000	2.160000	5.090000	0.220000	79.347498	158.300003	2.470000	53.029999
max	45.430000	7.250000	57.799999	66.900002	2.020000	461.299988	1680.000000	63.000000	178.899994

In [184]:
```
df1=df[['BEN', 'CO', 'EBE', 'MXY', 'NMHC', 'NO_2', 'NOx', 'OXY', 'O_3',
 'PM10', 'PXY', 'SO_2', 'TCH', 'TOL', 'station']]
```

In [185]:
```
sns.pairplot(df1[0:50])
```

Out[185]:
```
<seaborn.axisgrid.PairGrid at 0x1b170cdca30>
```

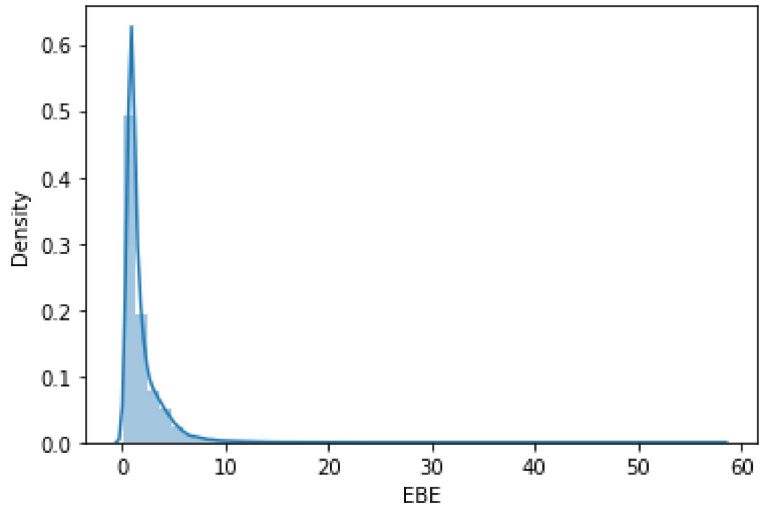

BEN
CO
EBE
MXY
NMHC
NO_2
NOx
OXY
O_3
PM10
PXY
SO_2
TCH



In [186]:
```
sns.distplot(df1['EBE'])
```

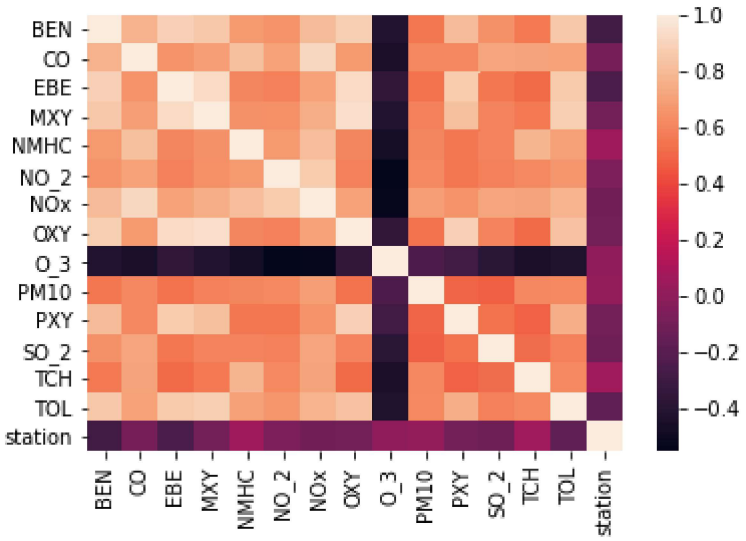
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a dep
recated function and will be removed in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an axes-level function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

Out[186]: <AxesSubplot:xlabel='EBE', ylabel='Density'>



In [187]:
```
sns.heatmap(df1.corr())
```

Out[187]: <AxesSubplot:>



In [188]:
```
x=df[['BEN', 'CO', 'EBE', 'MXY', 'NMHC', 'NO_2', 'NOx', 'OXY', 'O_3',
 'PM10', 'PXY', 'SO_2', 'TCH', 'TOL']]
y=df['station']
```

In [189]:
```
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)
```

In [190]:
```
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[190]: LinearRegression()

In [191]:
```
lr.intercept_
```

Out[191]: 28079026.2561639

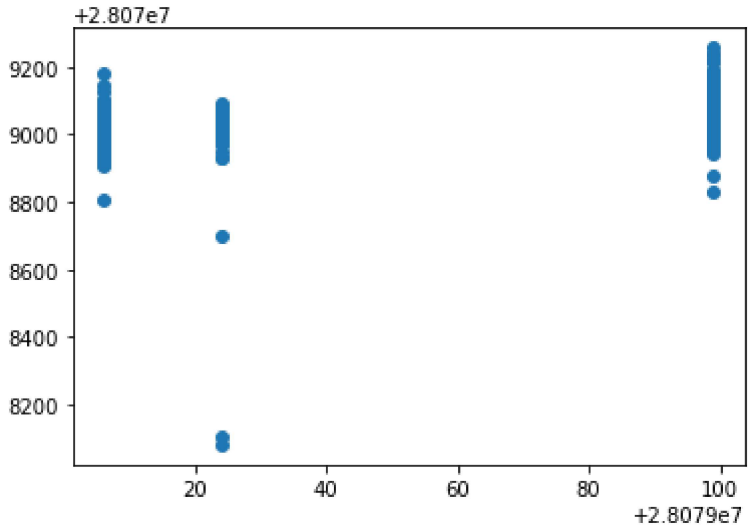
In [192]:
```
coeff=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
coeff
```

Out[192]:

	Co-efficient
BEN	-21.893539
CO	-7.439464
EBE	-24.811502
MXY	6.532366
NMHC	127.075045
NO_2	-0.042926
NOx	0.019741
OXY	10.767984
O_3	-0.037720
PM10	0.118907
PXY	6.860635
SO_2	-0.512966
TCH	13.701935
TOL	-0.018435

In [193]:
```
prediction =lr.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[193]: <matplotlib.collections.PathCollection at 0x1b17ee8c3d0>



In [194]:
```
lr.score(x_test,y_test)
```

Out[194]: 0.2762628434964486

In [195]:
```
lr.score(x_train,y_train)
```

Out[195]: 0.4181657194466114

In [196]:
```
from sklearn.linear_model import Ridge,Lasso
```

In [197]:
```
rr=Ridge(alpha=10)
rr.fit(x_train,y_train)
```

Out[197]: Ridge(alpha=10)

In [198]:
```
rr.score(x_test,y_test)
```

Out[198]: 0.27378982126458806

In [199]:
```
rr.score(x_train,y_train)
```

Out[199]:
```
0.41752121857497326
```

In [200]:
```
la=Lasso(alpha=10)
la.fit(x_train,y_train)
```

Out[200]:
```
Lasso(alpha=10)
```

In [201]:
```
la.score(x_train,y_train)
```

Out[201]:
```
0.062109979229048484
```

In [202]:
```
la.score(x_test,y_test)
```

Out[202]:
```
0.05849392873330106
```

In [203]:
```
from sklearn.linear_model import ElasticNet
en=ElasticNet()
en.fit(x_train,y_train)
```

Out[203]:
```
ElasticNet()
```

In [204]:
```
en.coef_
```

Out[204]:
```
array([-9.39800478,  0.        , -9.38981297,  3.68157567,  0.42838989,
       -0.01386303,  0.01633955,  2.57991111, -0.11699899,  0.29745146,
        2.3894965 , -0.39224048,  0.52916397, -0.854195  ])
```

In [205]:
```
en.intercept_
```

Out[205]:
```
28079052.203664407
```

In [206]:
```
prediction=en.predict(x_test)
```

In [207]:
```
en.score(x_test,y_test)
```

Out[207]:
```
0.220680994144546
```

In [208]:
```
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)))
```

```
32.08977616142971
1285.8284012148508
35.85844950935345
```

In [209]:
```
from sklearn.linear_model import LogisticRegression
```

In [210]:
```
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 [211]:
```
feature_matrix.shape
```

Out[211]:
```
(24758, 14)
```

In [212]:
```
target_vector.shape
```

Out[212]:
```
(24758,)
```

In [213]:
```
from sklearn.preprocessing import StandardScaler
```

In [214]:
```
fs=StandardScaler().fit_transform(feature_matrix)
```

In [215]:
```
logr=LogisticRegression(max_iter=10000)
logr.fit(fs,target_vector)
```

Out[215]:
```
LogisticRegression(max_iter=10000)
```

In [216]:
```
observation=[[1,2,3,4,5,6,7,8,9,10,11,12,13,14]]
```

In [217]:
```
prediction=logr.predict(observation)
print(prediction)
```

```
[28079099]
```

In [218]:
```
logr.classes_
```

Out[218]:
```
array([28079006, 28079024, 28079099], dtype=int64)
```

In [219]:
```
logr.score(fs,target_vector)
```

Out[219]:
```
0.8741416915744405
```

In [220]:
```
logr.predict_proba(observation)[0][0]
```

Out[220]:
```
3.5557727473608076e-15
```

In [221]:
```
logr.predict_proba(observation)
```

Out[221]:
```
array([[3.55577275e-15, 7.80743173e-29, 1.00000000e+00]])
```

In [222]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [223]:
```
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[223]:
```
RandomForestClassifier()
```

In [224]:
```
parameters={'max_depth':[1,2,3,4,5],
 'min_samples_leaf':[5,10,15,20,25],
 'n_estimators':[10,20,30,40,50]
}
```

In [225]:
```
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[225]:
```
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 [226]:
```
grid_search.best_score_
```

Out[226]:
```
0.8745527986151183
```

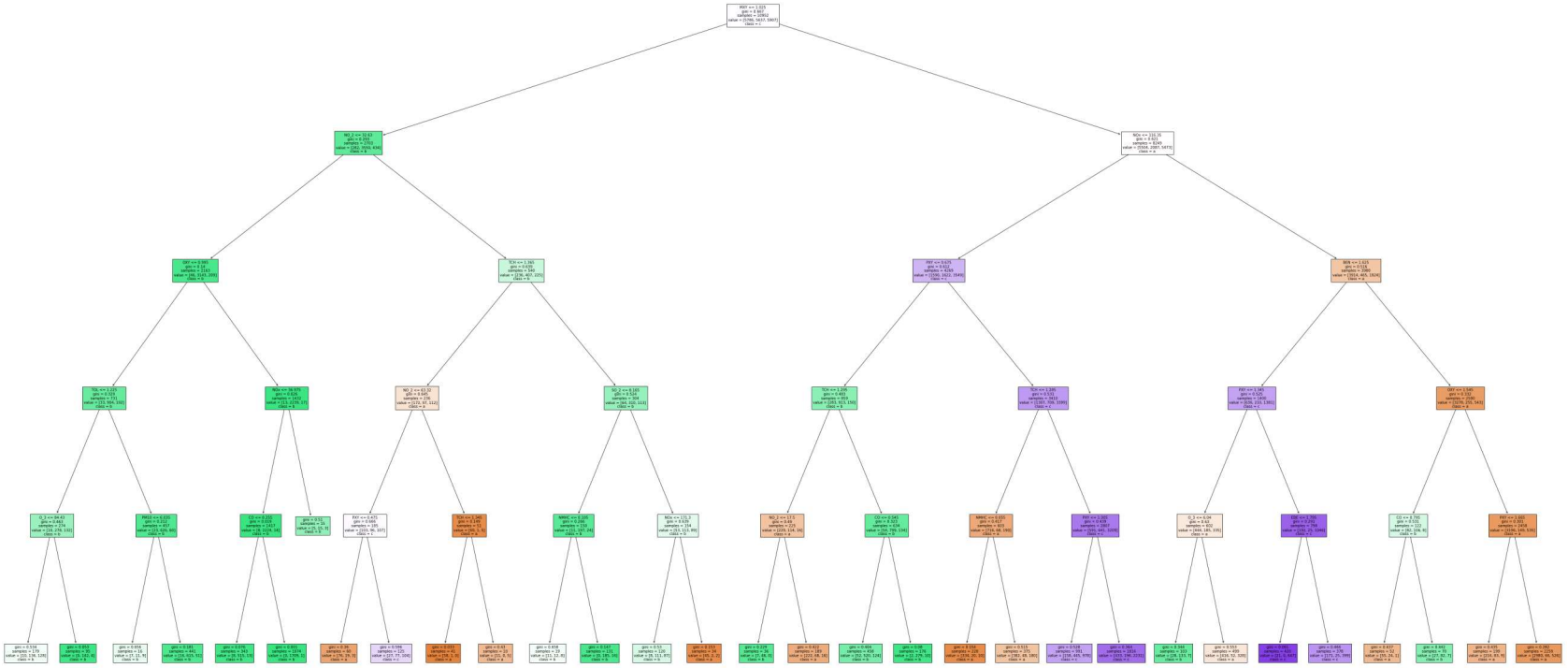
In [227]:
```
rfc_best=grid_search.best_estimator_
```

In [228]:
```
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'],filled=True)
```

Out[228]: [Text(2148.3, 1993.2, 'MXY <= 1.025\ngini = 0.667\nsamples = 10952\nvalue = [5786, 5637, 5907]\nclass = c'),
 Text(1023.0000000000001, 1630.8000000000002, 'NO_2 <= 32.63\ngini = 0.293\nsamples = 2703\nvalue = [282, 3550, 434]\nclass = b'),
 Text(558.0, 1268.4, 'OXY <= 0.995\ngini = 0.14\nsamples = 2163\nvalue = [46, 3143, 209]\nclass = b'),
 Text(297.6, 906.0, 'TOL <= 1.225\ngini = 0.329\nsamples = 731\nvalue = [33, 904, 192]\nclass = b'),
 Text(148.8, 543.5999999999999, 'O_3 <= 84.43\ngini = 0.463\nsamples = 274\nvalue = [10, 278, 132]\nclass = b'),
 Text(74.4, 181.19999999999982, 'gini = 0.534\nsamples = 179\nvalue = [10, 136, 128]\nclass = b'),
 Text(223.20000000000002, 181.19999999999982, 'gini = 0.053\nsamples = 95\nvalue = [0, 142, 4]\nclass = b'),
 Text(446.40000000000003, 543.5999999999999, 'PM10 <= 6.035\ngini = 0.212\nsamples = 457\nvalue = [23, 626, 60]\nclass = b'),
 Text(372.0, 181.19999999999982, 'gini = 0.656\nsamples = 16\nvalue = [7, 11, 9]\nclass = b'),
 Text(520.8000000000001, 181.19999999999982, 'gini = 0.181\nsamples = 441\nvalue = [16, 615, 51]\nclass = b'),
 Text(818.4000000000001, 906.0, 'NOx <= 36.975\ngini = 0.026\nsamples = 1432\nvalue = [13, 2239, 17]\nclass = b'),
 Text(744.0, 543.5999999999999, 'CO <= 0.255\ngini = 0.019\nsamples = 1417\nvalue = [8, 2224, 14]\nclass = b'),
 Text(669.6, 181.19999999999982, 'gini = 0.076\nsamples = 343\nvalue = [8, 515, 13]\nclass = b'),
 Text(818.4000000000001, 181.19999999999982, 'gini = 0.001\nsamples = 1074\nvalue = [0, 1709, 1]\nclass = b'),
 Text(892.8000000000001, 543.5999999999999, 'gini = 0.51\nsamples = 15\nvalue = [5, 15, 3]\nclass = b'),
 Text(1488.0, 1268.4, 'TCH <= 1.365\ngini = 0.639\nsamples = 540\nvalue = [236, 407, 225]\nclass = b'),
 Text(1190.4, 906.0, 'NO_2 <= 63.32\ngini = 0.645\nsamples = 236\nvalue = [172, 97, 112]\nclass = a'),
 Text(1041.6000000000001, 543.5999999999999, 'PXY <= 0.475\ngini = 0.666\nsamples = 185\nvalue = [103, 96, 107]\nclass = c'),
 Text(967.2, 181.19999999999982, 'gini = 0.36\nsamples = 60\nvalue = [76, 19, 3]\nclass = a'),
 Text(1116.0, 181.19999999999982, 'gini = 0.596\nsamples = 125\nvalue = [27, 77, 104]\nclass = c'),
 Text(1339.2, 543.5999999999999, 'TCH <= 1.345\ngini = 0.149\nsamples = 51\nvalue = [69, 1, 5]\nclass = a'),
 Text(1264.8000000000002, 181.19999999999982, 'gini = 0.033\nsamples = 41\nvalue = [58, 1, 0]\nclass = a'),
 Text(1413.6000000000001, 181.19999999999982, 'gini = 0.43\nsamples = 10\nvalue = [11, 0, 5]\nclass = a'),
 Text(1785.6000000000001, 906.0, 'SO_2 <= 8.165\ngini = 0.524\nsamples = 304\nvalue = [64, 310, 113]\nclass = b'),
 Text(1636.8000000000002, 543.5999999999999, 'NMHC <= 0.105\ngini = 0.266\nsamples = 150\nvalue = [11, 197, 24]\nclass = b'),
 Text(1562.4, 181.19999999999982, 'gini = 0.658\nsamples = 19\nvalue = [11, 12, 8]\nclass = b'),
 Text(1711.2, 181.19999999999982, 'gini = 0.147\nsamples = 131\nvalue = [0, 185, 16]\nclass = b'),
 Text(1934.4, 543.5999999999999, 'NOx <= 171.3\ngini = 0.639\nsamples = 154\nvalue = [53, 113, 89]\nclass = b'),
 Text(1860.0000000000002, 181.19999999999982, 'gini = 0.53\nsamples = 120\nvalue = [8, 111, 87]\nclass = b'),
 Text(2008.8000000000002, 181.19999999999982, 'gini = 0.153\nsamples = 34\nvalue = [45, 2, 2]\nclass = a'),
 Text(3273.6000000000004, 1630.8000000000002, 'NOx <= 116.35\ngini = 0.621\nsamples = 8249\nvalue = [5504, 2

```
087, 5473]\nclass = a'),
 Text(2678.4, 1268.4, 'PXY <= 0.675\ngini = 0.612\nsamples = 4269\nvalue = [1590, 1622, 3549]\nclass = c'),
 Text(2380.8, 906.0, 'TCH <= 1.295\ngini = 0.483\nsamples = 859\nvalue = [283, 913, 150]\nclass = b'),
 Text(2232.0, 543.5999999999999, 'NO_2 <= 17.5\ngini = 0.49\nsamples = 225\nvalue = [229, 114, 16]\nclass =
a'),
 Text(2157.6000000000004, 181.19999999999982, 'gini = 0.229\nsamples = 36\nvalue = [7, 46, 0]\nclass = b'),
 Text(2306.4, 181.19999999999982, 'gini = 0.422\nsamples = 189\nvalue = [222, 68, 16]\nclass = a'),
 Text(2529.6000000000004, 543.5999999999999, 'CO <= 0.545\ngini = 0.323\nsamples = 634\nvalue = [54, 799, 13
4]\nclass = b'),
 Text(2455.2000000000003, 181.19999999999982, 'gini = 0.404\nsamples = 458\nvalue = [52, 520, 124]\nclass =
b'),
 Text(2604.0, 181.19999999999982, 'gini = 0.08\nsamples = 176\nvalue = [2, 279, 10]\nclass = b'),
 Text(2976.0, 906.0, 'TCH <= 1.285\ngini = 0.531\nsamples = 3410\nvalue = [1307, 709, 3399]\nclass = c'),
 Text(2827.2000000000003, 543.5999999999999, 'NMHC <= 0.055\ngini = 0.417\nsamples = 603\nvalue = [716, 68,
190]\nclass = a'),
 Text(2752.8, 181.19999999999982, 'gini = 0.154\nsamples = 228\nvalue = [334, 20, 10]\nclass = a'),
 Text(2901.6000000000004, 181.19999999999982, 'gini = 0.515\nsamples = 375\nvalue = [382, 48, 180]\nclass =
a'),
 Text(3124.8, 543.5999999999999, 'PXY <= 1.005\ngini = 0.439\nsamples = 2807\nvalue = [591, 641, 3209]\nclas
s = c'),
 Text(3050.4, 181.19999999999982, 'gini = 0.528\nsamples = 991\nvalue = [158, 445, 978]\nclass = c'),
 Text(3199.2000000000003, 181.19999999999982, 'gini = 0.364\nsamples = 1816\nvalue = [433, 196, 2231]\nclass
= c'),
 Text(3868.8, 1268.4, 'BEN <= 1.625\ngini = 0.516\nsamples = 3980\nvalue = [3914, 465, 1924]\nclass = a'),
 Text(3571.2000000000003, 906.0, 'PXY <= 1.345\ngini = 0.525\nsamples = 1400\nvalue = [636, 210, 1381]\nclas
s = c'),
 Text(3422.4, 543.5999999999999, 'O_3 <= 6.04\ngini = 0.63\nsamples = 602\nvalue = [444, 185, 335]\nclass =
a'),
 Text(3348.0000000000005, 181.19999999999982, 'gini = 0.344\nsamples = 103\nvalue = [28, 133, 7]\nclass =
b'),
 Text(3496.8, 181.19999999999982, 'gini = 0.553\nsamples = 499\nvalue = [416, 52, 328]\nclass = a'),
 Text(3720.0000000000005, 543.5999999999999, 'EBE <= 1.795\ngini = 0.291\nsamples = 798\nvalue = [192, 25, 1
046]\nclass = c'),
 Text(3645.6000000000004, 181.19999999999982, 'gini = 0.061\nsamples = 420\nvalue = [21, 0, 647]\nclass =
c'),
 Text(3794.4, 181.19999999999982, 'gini = 0.466\nsamples = 378\nvalue = [171, 25, 399]\nclass = c'),
 Text(4166.400000000001, 906.0, 'OXY <= 1.545\ngini = 0.332\nsamples = 2580\nvalue = [3278, 255, 543]\nclass
= a'),
 Text(4017.6000000000004, 543.5999999999999, 'CO <= 0.795\ngini = 0.531\nsamples = 122\nvalue = [82, 106, 8]
\nclass = b'),
 Text(3943.2000000000003, 181.19999999999982, 'gini = 0.437\nsamples = 52\nvalue = [55, 24, 1]\nclass = a'),
 Text(4092.0000000000005, 181.19999999999982, 'gini = 0.442\nsamples = 70\nvalue = [27, 82, 7]\nclass = b'),
 Text(4315.200000000001, 543.5999999999999, 'PXY <= 1.665\ngini = 0.301\nsamples = 2458\nvalue = [3196, 149,
```

```
535]\nclass = a'),
 Text(4240.8, 181.19999999999982, 'gini = 0.435\nsamples = 199\nvalue = [216, 83, 9]\nclass = a'),
 Text(4389.6, 181.19999999999982, 'gini = 0.282\nsamples = 2259\nvalue = [2980, 66, 526]\nclass = a')]
```



Conclusion

In [ ]:

```
linear Regression=0.4181657194466114

Ridge Regression=0.41752121857497326

Lasso Regression=0.062109979229048484

ElasticNet Regression=0.220680994144546

Logistic Regression=0.8741416915744405

Random Forest=0.8745527986151183

Logistic Regression is Suitable for this Dataset=0.8745527986151183
```