airvisualizeaf

May 1, 2025

[37]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[38]:
```
df=pd.read_csv('AirQuality.csv',delimiter=';')
```

[39]:
```
df.shape
```

[39]:
```
(9471, 17)
```

[40]:
```
df.head()
```

[40]:

	Date	Time	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	\
0	10/03/2004	18.00.00	2,6	1360.0	150.0	11,9	1046.0	
1	10/03/2004	19.00.00	2	1292.0	112.0	9,4	955.0	
2	10/03/2004	20.00.00	2,2	1402.0	88.0	9,0	939.0	
3	10/03/2004	21.00.00	2,2	1376.0	80.0	9,2	948.0	
4	10/03/2004	22.00.00	1,6	1272.0	51.0	6,5	836.0	

	NOx(GT)	PT08.S3(NOx)	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)	T	RH	\
0	166.0	1056.0	113.0	1692.0	1268.0	13,6	48,9	
1	103.0	1174.0	92.0	1559.0	972.0	13,3	47,7	
2	131.0	1140.0	114.0	1555.0	1074.0	11,9	54,0	
3	172.0	1092.0	122.0	1584.0	1203.0	11,0	60,0	
4	131.0	1205.0	116.0	1490.0	1110.0	11,2	59,6	

	AH	Unnamed: 15	Unnamed: 16
0	0,7578	NaN	NaN
1	0,7255	NaN	NaN
2	0,7502	NaN	NaN
3	0,7867	NaN	NaN
4	0,7888	NaN	NaN

[41]:
```
df=df.rename(columns={'T':'Temperature'})
df=df.rename(columns={'RH':'Relative Humidity'})
df=df.rename(columns={'AH':'Absolute Humidity'})
```

```
df
```

[41]:

	Date	Time	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	\
0	10/03/2004	18.00.00	2,6	1360.0	150.0	11,9	
1	10/03/2004	19.00.00	2	1292.0	112.0	9,4	
2	10/03/2004	20.00.00	2,2	1402.0	88.0	9,0	
3	10/03/2004	21.00.00	2,2	1376.0	80.0	9,2	
4	10/03/2004	22.00.00	1,6	1272.0	51.0	6,5	
…	…	…	…	…	…	…	
9466	NaN	NaN	NaN	NaN	NaN	NaN	
9467	NaN	NaN	NaN	NaN	NaN	NaN	
9468	NaN	NaN	NaN	NaN	NaN	NaN	
9469	NaN	NaN	NaN	NaN	NaN	NaN	
9470	NaN	NaN	NaN	NaN	NaN	NaN	

	PT08.S2(NMHC)	NOx(GT)	PT08.S3(NOx)	NO2(GT)	PT08.S4(NO2)	\
0	1046.0	166.0	1056.0	113.0	1692.0	
1	955.0	103.0	1174.0	92.0	1559.0	
2	939.0	131.0	1140.0	114.0	1555.0	
3	948.0	172.0	1092.0	122.0	1584.0	
4	836.0	131.0	1205.0	116.0	1490.0	
…	…	…	…	…	…	
9466	NaN	NaN	NaN	NaN	NaN	
9467	NaN	NaN	NaN	NaN	NaN	
9468	NaN	NaN	NaN	NaN	NaN	
9469	NaN	NaN	NaN	NaN	NaN	
9470	NaN	NaN	NaN	NaN	NaN	

	PT08.S5(O3)	Temperature	Relative Humidity	Absolute Humidity	\
0	1268.0	13,6	48,9	0,7578	
1	972.0	13,3	47,7	0,7255	
2	1074.0	11,9	54,0	0,7502	
3	1203.0	11,0	60,0	0,7867	
4	1110.0	11,2	59,6	0,7888	
…	…	…	…	…	
9466	NaN	NaN	NaN	NaN	
9467	NaN	NaN	NaN	NaN	
9468	NaN	NaN	NaN	NaN	
9469	NaN	NaN	NaN	NaN	
9470	NaN	NaN	NaN	NaN	

	Unnamed: 15	Unnamed: 16
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

```
…           …          …
9466          NaN          NaN
9467          NaN          NaN
9468          NaN          NaN
9469          NaN          NaN
9470          NaN          NaN

[9471 rows x 17 columns]
```

[42]:
```
df=df.drop(['Unnamed: 15', 'Unnamed: 16'],axis=1)
df
```

[42]:
```
             Date      Time CO(GT)  PT08.S1(CO)  NMHC(GT) C6H6(GT)  \
0      10/03/2004  18.00.00    2,6       1360.0     150.0     11,9
1      10/03/2004  19.00.00      2       1292.0     112.0      9,4
2      10/03/2004  20.00.00    2,2       1402.0      88.0      9,0
3      10/03/2004  21.00.00    2,2       1376.0      80.0      9,2
4      10/03/2004  22.00.00    1,6       1272.0      51.0      6,5
…           …        …    …           …        …      …
9466         NaN       NaN    NaN          NaN       NaN      NaN
9467         NaN       NaN    NaN          NaN       NaN      NaN
9468         NaN       NaN    NaN          NaN       NaN      NaN
9469         NaN       NaN    NaN          NaN       NaN      NaN
9470         NaN       NaN    NaN          NaN       NaN      NaN

      PT08.S2(NMHC)  NOx(GT)  PT08.S3(NOx)  NO2(GT)  PT08.S4(NO2)  \
0            1046.0    166.0        1056.0    113.0        1692.0
1             955.0    103.0        1174.0     92.0        1559.0
2             939.0    131.0        1140.0    114.0        1555.0
3             948.0    172.0        1092.0    122.0        1584.0
4             836.0    131.0        1205.0    116.0        1490.0
…               …      …           …      …           …
9466            NaN      NaN           NaN      NaN           NaN
9467            NaN      NaN           NaN      NaN           NaN
9468            NaN      NaN           NaN      NaN           NaN
9469            NaN      NaN           NaN      NaN           NaN
9470            NaN      NaN           NaN      NaN           NaN

      PT08.S5(O3) Temperature Relative Humidity Absolute Humidity
0          1268.0        13,6              48,9            0,7578
1           972.0        13,3              47,7            0,7255
2          1074.0        11,9              54,0            0,7502
3          1203.0        11,0              60,0            0,7867
4          1110.0        11,2              59,6            0,7888
…           …          …                …               …
9466          NaN         NaN               NaN               NaN
9467          NaN         NaN               NaN               NaN
```

```
9468          NaN          NaN                NaN                NaN
9469          NaN          NaN                NaN                NaN
9470          NaN          NaN                NaN                NaN

[9471 rows x 15 columns]
```

[43]:
```
df['CO(GT)']=df['CO(GT)'].str.replace(',','.').astype(float)
df['C6H6(GT)']=df['C6H6(GT)'].str.replace(',','.').astype(float)
df['Temperature']=df['Temperature'].str.replace(',','.').astype(float)
df['Relative Humidity']=df['Relative Humidity'].str.replace(',','.').
 ↪astype(float)
df['Absolute Humidity']=df['Absolute Humidity'].str.replace(',','.').
 ↪astype(float)
```

[44]:
```
df=df.drop_duplicates()
```

[45]:
```
df1=df.iloc[:9466,:]
df1
```

[45]:
```
            Date      Time  CO(GT)  PT08.S1(CO)  NMHC(GT)  C6H6(GT)  \
0     10/03/2004  18.00.00     2.6       1360.0     150.0      11.9
1     10/03/2004  19.00.00     2.0       1292.0     112.0       9.4
2     10/03/2004  20.00.00     2.2       1402.0      88.0       9.0
3     10/03/2004  21.00.00     2.2       1376.0      80.0       9.2
4     10/03/2004  22.00.00     1.6       1272.0      51.0       6.5
…            …         …       …            …         …         …
9353  04/04/2005  11.00.00     2.4       1163.0    -200.0      11.4
9354  04/04/2005  12.00.00     2.4       1142.0    -200.0      12.4
9355  04/04/2005  13.00.00     2.1       1003.0    -200.0       9.5
9356  04/04/2005  14.00.00     2.2       1071.0    -200.0      11.9
9357         NaN       NaN     NaN          NaN       NaN       NaN

      PT08.S2(NMHC)  NOx(GT)  PT08.S3(NOx)  NO2(GT)  PT08.S4(NO2)  \
0            1046.0    166.0        1056.0    113.0        1692.0
1             955.0    103.0        1174.0     92.0        1559.0
2             939.0    131.0        1140.0    114.0        1555.0
3             948.0    172.0        1092.0    122.0        1584.0
4             836.0    131.0        1205.0    116.0        1490.0
…               …        …             …         …            …
9353         1027.0    353.0         604.0    179.0        1264.0
9354         1063.0    293.0         603.0    175.0        1241.0
9355          961.0    235.0         702.0    156.0        1041.0
9356         1047.0    265.0         654.0    168.0        1129.0
9357            NaN      NaN           NaN      NaN           NaN

      PT08.S5(O3)  Temperature  Relative Humidity  Absolute Humidity
0          1268.0         13.6               48.9             0.7578
```

```
1           972.0         13.3              47.7             0.7255
2          1074.0         11.9              54.0             0.7502
3          1203.0         11.0              60.0             0.7867
4          1110.0         11.2              59.6             0.7888
…             …            …                 …                …
9353       1269.0         24.3              23.7             0.7119
9354       1092.0         26.9              18.3             0.6406
9355        770.0         28.3              13.5             0.5139
9356        816.0         28.5              13.1             0.5028
9357          NaN          NaN               NaN              NaN

[9358 rows x 15 columns]
```

[46]:
```
df.isna().sum()
```

[46]:
```
Date                1
Time                1
CO(GT)              1
PT08.S1(CO)         1
NMHC(GT)            1
C6H6(GT)            1
PT08.S2(NMHC)       1
NOx(GT)             1
PT08.S3(NOx)        1
NO2(GT)             1
PT08.S4(NO2)        1
PT08.S5(O3)         1
Temperature         1
Relative Humidity   1
Absolute Humidity   1
dtype: int64
```

[47]:
```
df=df.fillna(df.mean(numeric_only=True))
df=df.dropna()
```

[48]:
```
df.isna().sum()
```

[48]:
```
Date                0
Time                0
CO(GT)              0
PT08.S1(CO)         0
NMHC(GT)            0
C6H6(GT)            0
PT08.S2(NMHC)       0
NOx(GT)             0
PT08.S3(NOx)        0
NO2(GT)             0
```

```
PT08.S4(NO2)          0
PT08.S5(O3)           0
Temperature           0
Relative Humidity     0
Absolute Humidity     0
dtype: int64
```

[49]:
```
df['Absolute Humidity']=df['Absolute Humidity'].multiply(100)
```

[50]:
```
def remove_outliers(column):
    Q1=column.quantile(0.25)
    Q3=column.quantile(0.75)
    IQR=Q3-Q1
    threshold=1.5*IQR
    outlier_mask=(column<Q1-threshold) | (column>Q3+threshold)
    return column[~outlier_mask]
```

[51]:
```
df.columns
```

[51]:
```
Index(['Date', 'Time', 'CO(GT)', 'PT08.S1(CO)', 'NMHC(GT)', 'C6H6(GT)',
       'PT08.S2(NMHC)', 'NOx(GT)', 'PT08.S3(NOx)', 'NO2(GT)', 'PT08.S4(NO2)',
       'PT08.S5(O3)', 'Temperature', 'Relative Humidity', 'Absolute Humidity'],
      dtype='object')
```

[52]:
```
col_name=['PT08.S1(CO)', 'C6H6(GT)','PT08.S2(NMHC)', 'NOx(GT)', 'PT08.S3(NOx)',␣
 ↪'NO2(GT)', 'PT08.S4(NO2)','PT08.S5(O3)', 'Temperature', 'Relative Humidity',␣
 ↪'Absolute Humidity']
for col in col_name:
    df[col]=remove_outliers(df[col])
```

[53]:
```
df['Date']=pd.to_datetime(df["Date"], format="%d/%m/%Y" , errors='coerce')
df['Year']=df['Date'].dt.year
df['Month']=df['Date'].dt.month
```

[54]:
```
df
```

[54]:

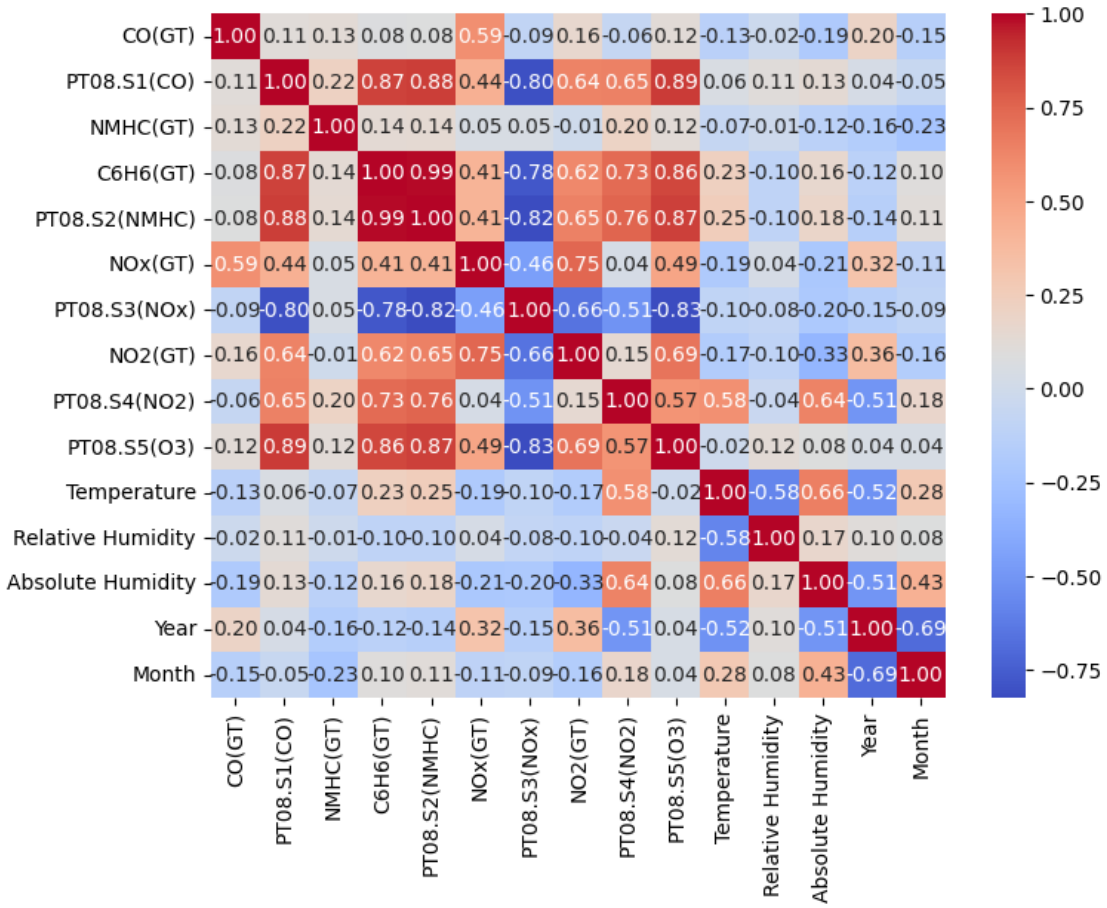
	Date	Time	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	\
0	2004-03-10	18.00.00	2.6	1360.0	150.0	11.9	
1	2004-03-10	19.00.00	2.0	1292.0	112.0	9.4	
2	2004-03-10	20.00.00	2.2	1402.0	88.0	9.0	
3	2004-03-10	21.00.00	2.2	1376.0	80.0	9.2	
4	2004-03-10	22.00.00	1.6	1272.0	51.0	6.5	
…	…	…	…	…	…	…	
9352	2005-04-04	10.00.00	3.1	1314.0	-200.0	13.5	
9353	2005-04-04	11.00.00	2.4	1163.0	-200.0	11.4	
9354	2005-04-04	12.00.00	2.4	1142.0	-200.0	12.4	
9355	2005-04-04	13.00.00	2.1	1003.0	-200.0	9.5	

```
9356 2005-04-04  14.00.00     2.2       1071.0    -200.0      11.9

      PT08.S2(NMHC)  NOx(GT)  PT08.S3(NOx)  NO2(GT)  PT08.S4(NO2)  \
0            1046.0    166.0        1056.0    113.0        1692.0
1             955.0    103.0        1174.0     92.0        1559.0
2             939.0    131.0        1140.0    114.0        1555.0
3             948.0    172.0        1092.0    122.0        1584.0
4             836.0    131.0        1205.0    116.0        1490.0
…               …       …             …         …            …
9352         1101.0    472.0         539.0    190.0        1374.0
9353         1027.0    353.0         604.0    179.0        1264.0
9354         1063.0    293.0         603.0    175.0        1241.0
9355          961.0    235.0         702.0    156.0        1041.0
9356         1047.0    265.0         654.0    168.0        1129.0

      PT08.S5(O3)  Temperature  Relative Humidity  Absolute Humidity  Year  \
0          1268.0         13.6               48.9              75.78  2004
1           972.0         13.3               47.7              72.55  2004
2          1074.0         11.9               54.0              75.02  2004
3          1203.0         11.0               60.0              78.67  2004
4          1110.0         11.2               59.6              78.88  2004
…             …            …                  …                 …    …
9352       1729.0         21.9               29.3              75.68  2005
9353       1269.0         24.3               23.7              71.19  2005
9354       1092.0         26.9               18.3              64.06  2005
9355        770.0         28.3               13.5              51.39  2005
9356        816.0         28.5               13.1              50.28  2005

      Month
0         3
1         3
2         3
3         3
4         3
…       …
9352      4
9353      4
9354      4
9355      4
9356      4

[9357 rows x 17 columns]
```

[55]:
```
df['yearr']=df.Year.astype(str)
df['month']=df.Month.astype(str)
```

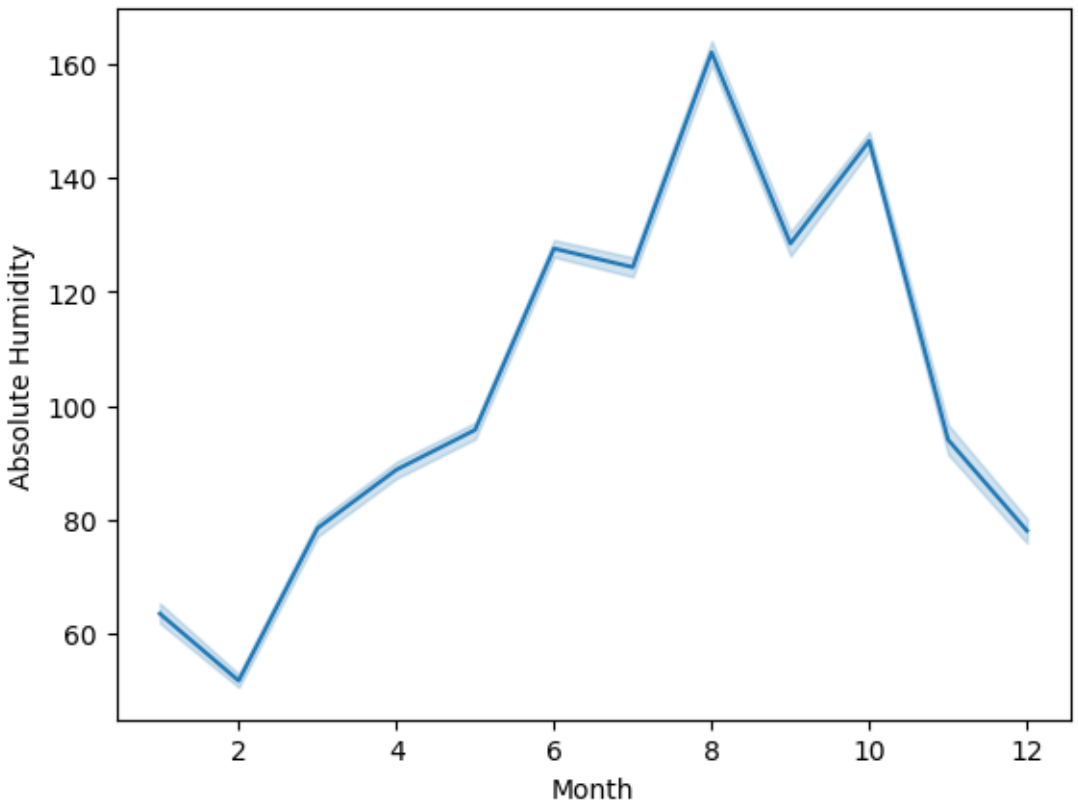
[56]:
```
plt.figure(figsize=(8,6))
sns.heatmap(df.corr(numeric_only=True), annot=True,cmap='coolwarm',fmt='.2f')
```

[56]: <Axes: >



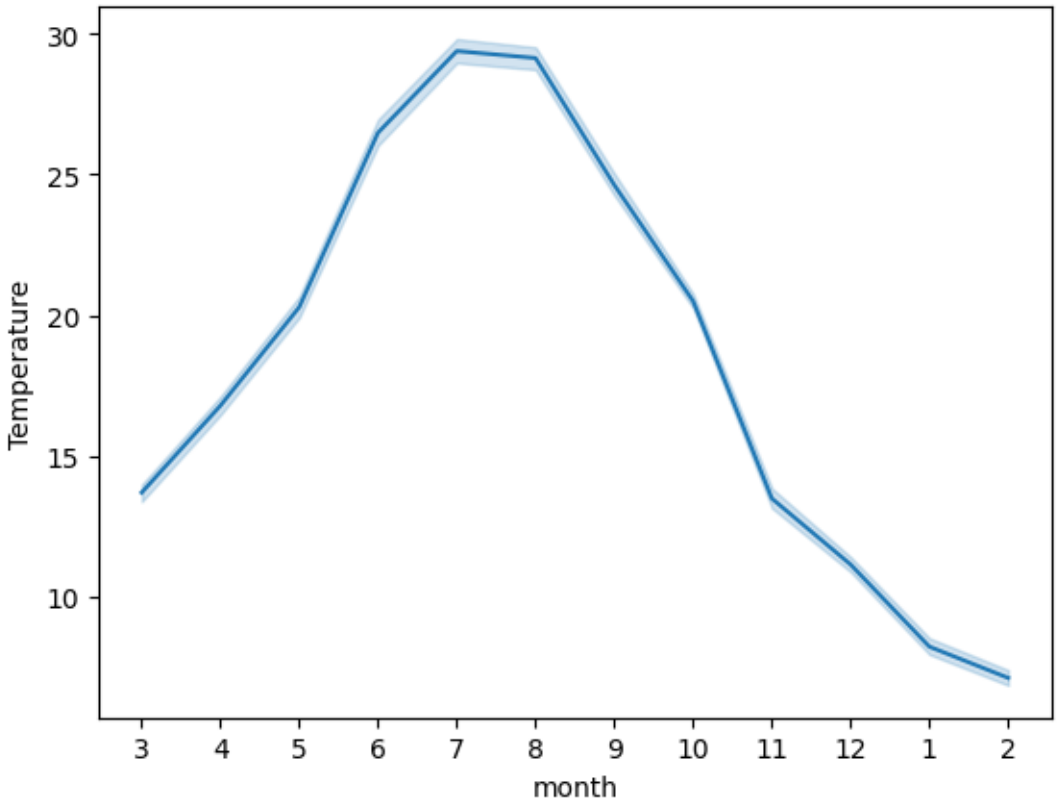
[57]:
```
sns.lineplot(df,x="Month",y="Absolute Humidity")
```

[57]: <Axes: xlabel='Month', ylabel='Absolute Humidity'>



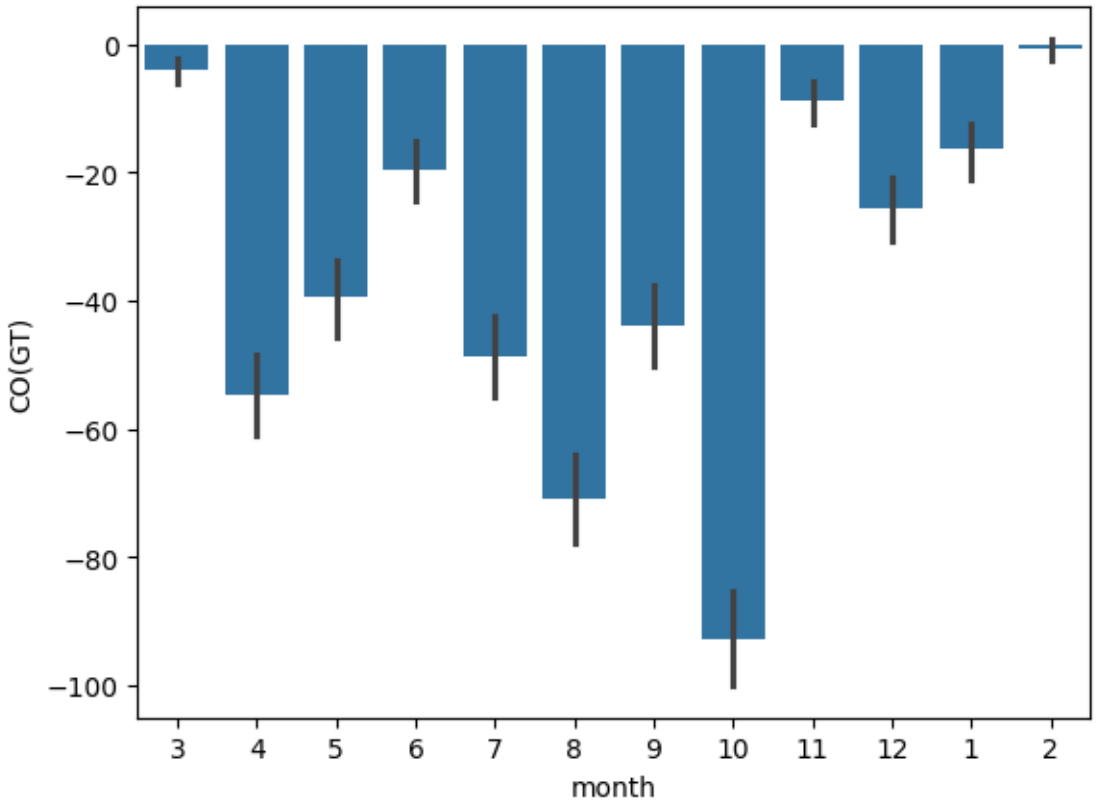
[58]:
```
sns.lineplot(df,x='month',y='Temperature')
```

[58]:
```
<Axes: xlabel='month', ylabel='Temperature'>
```



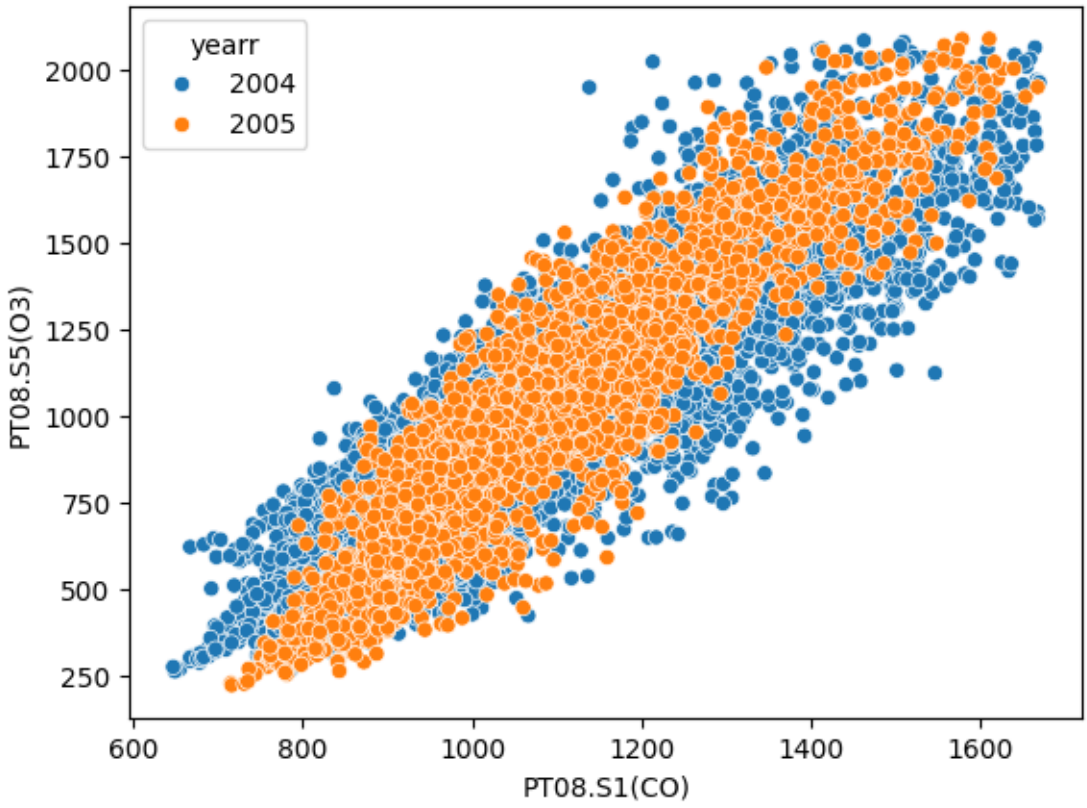
[68]:
```
sns.barplot(df,x=df.month,y=df['CO(GT)'])
```

[68]:
```
<Axes: xlabel='month', ylabel='CO(GT)'>
```



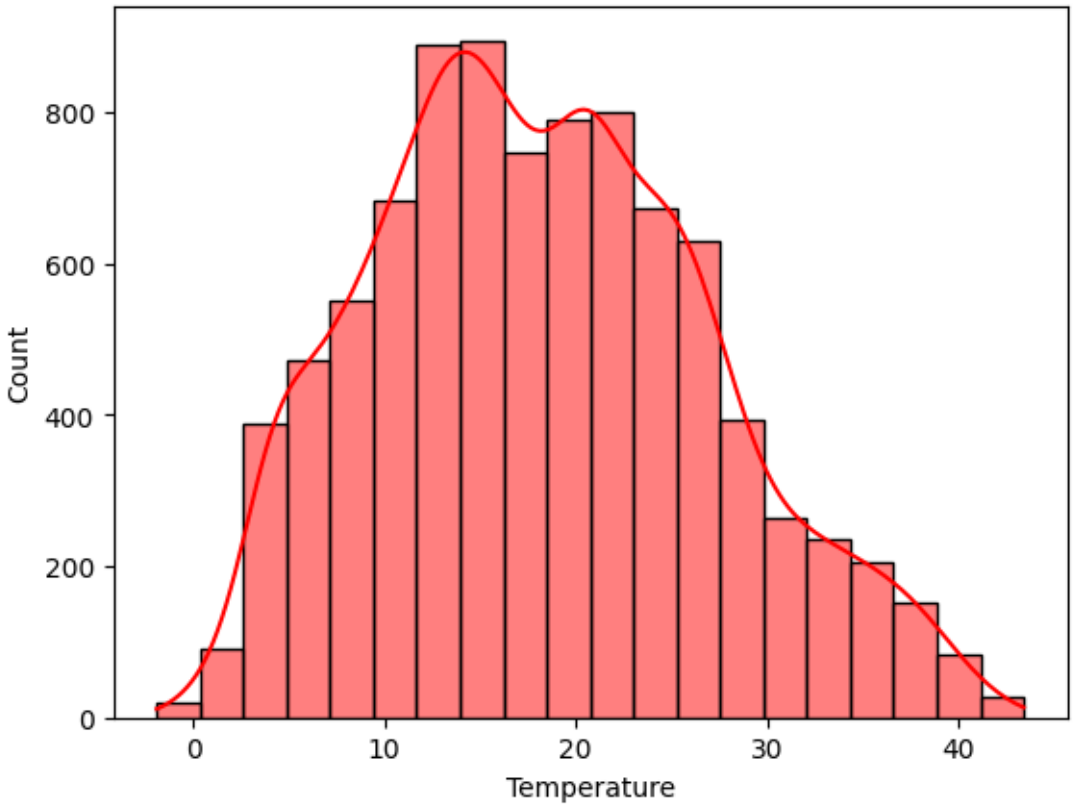
[60]:
```
sns.scatterplot(df,x='PT08.S1(CO)',y='PT08.S5(O3)',hue='yearr')
```

[60]: <Axes: xlabel='PT08.S1(CO)', ylabel='PT08.S5(O3)'>



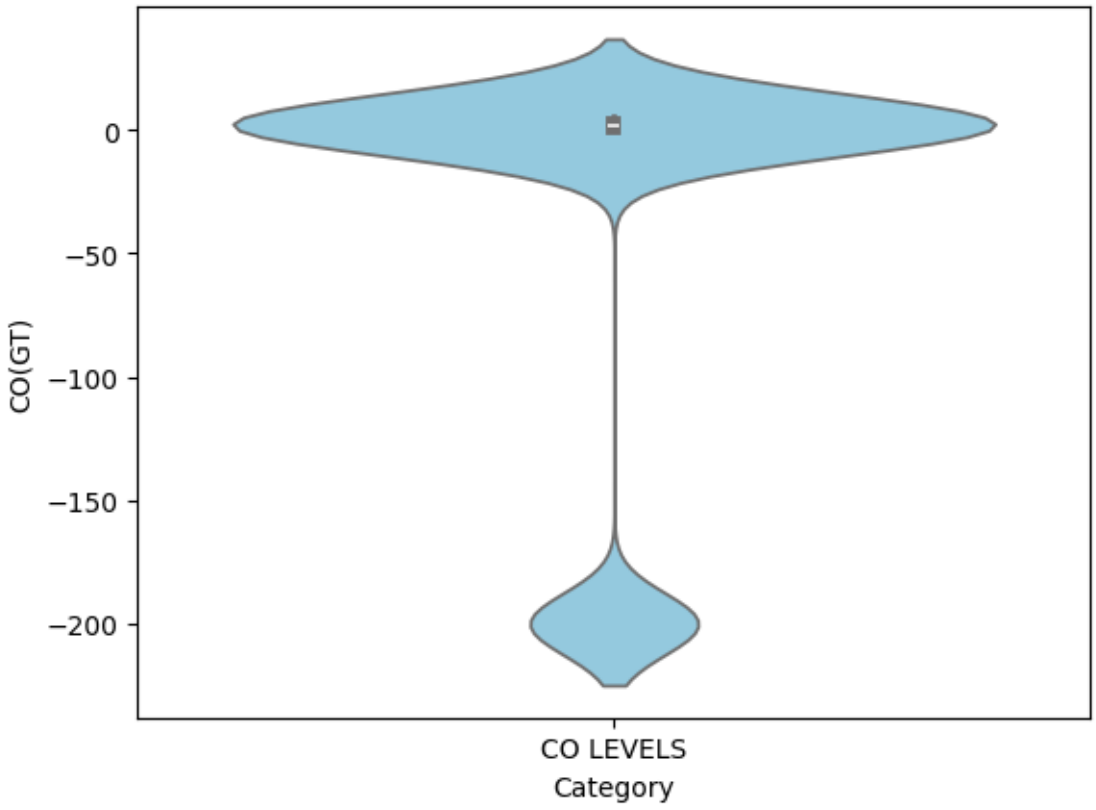
[61]:
```
sns.histplot(df["Temperature"],        bins=20,kde=True,color="red")
```

[61]: <Axes: xlabel='Temperature', ylabel='Count'>



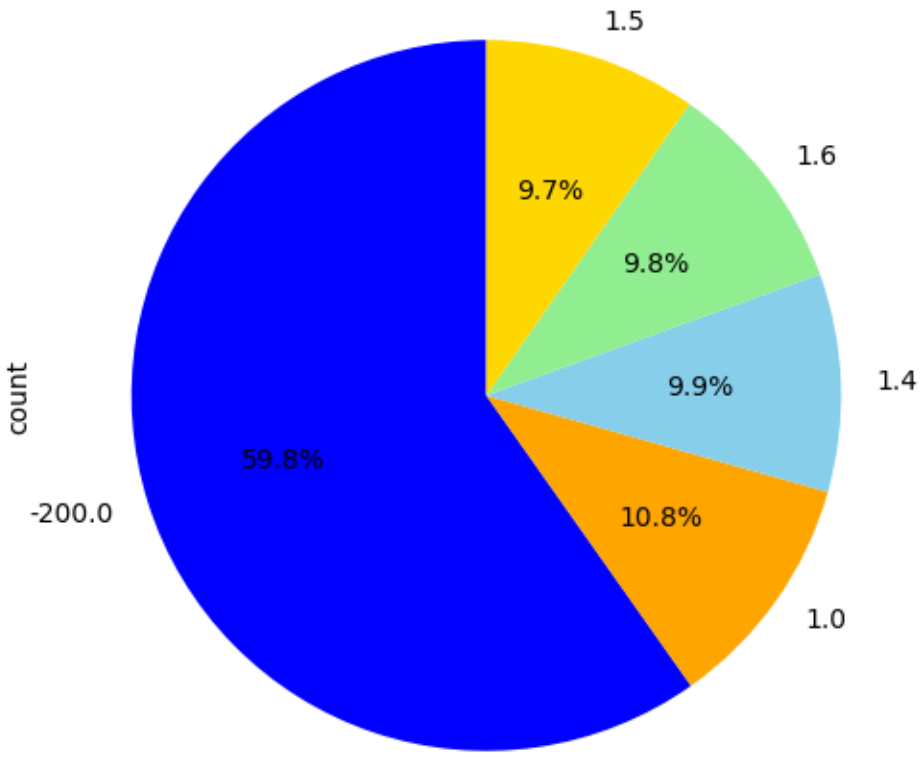
[62]:
```
df['Category']='CO LEVELS'
sns.violinplot(x=df['Category'],y=df['CO(GT)'],color='skyblue')
```

[62]: <Axes: xlabel='Category', ylabel='CO(GT)'>



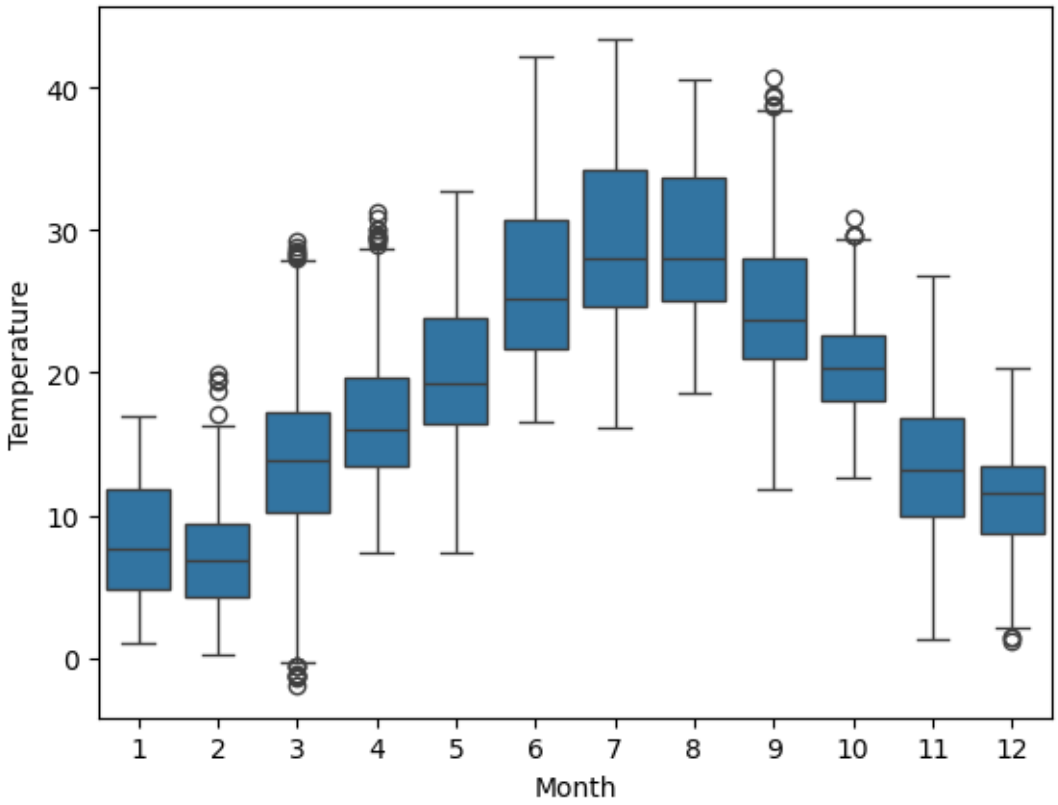
[66]:
```
co_counts=df['CO(GT)'].value_counts().nlargest(5)
plt.figure(figsize=(8,6))
co_counts.plot.pie(autopct='%1.1f%%',labels=co_counts.
 ↪index,startangle=90,colors=['blue','orange','skyblue','lightgreen','gold'])
```

[66]:
```
<Axes: ylabel='count'>
```



[67]:
```
sns.boxplot(df,x="Month",y='Temperature')
```

[67]: <Axes: xlabel='Month', ylabel='Temperature'>


Temperature
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
Month

[ ]: