Challenge of the Month: January

Air Quality Dataset Analysis

Tousif Azim

The dataset contains instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device.

Summary of the dataset:

Source of dataset: Kaggle

	CO_GT	PT08_S1_CO	C6H6_GT	PT08_S2_NMHC	Nox_GT	PT08_S3_Nox
count	7674.000000	7674.000000	7674.000000	7674.000000	7674.000000	7674.000000
mean	2.152750	1110.580746	10.275735	947.198121	255.313869	826.920071
std	1.453252	213.927135	7.279297	259.700177	209.472100	251.068817
min	0.100000	647.000000	0.200000	387.000000	2.000000	322.000000
25%	1.100000	953.000000	4.800000	752.000000	107.000000	657.000000
50%	1.800000	1087.000000	8.900000	934.000000	201.000000	807.000000
75%	2.900000	1235.000000	14.000000	1116.750000	326.000000	949.000000
max	11.900000	2040.000000	63.700000	2214.000000	1479.000000	2683.000000

| PT08_S3_Nox	NO2_GT	PT08_S4_NO2	PT08_S5_O3	T	RH	
7674.000000	7674.000000	7674.000000	7674.000000	7674.000000	7674.000000	7674.0
826.920071	114.811932	1444.752723	1043.512936	17.770425	49.060076	0.9
251.068817	46.918066	342.727584	396.752410	8.670011	17.072161	0.3
322.000000	2.000000	551.000000	221.000000	-1.900000	9.200000	0.1
657.000000	82.000000	1215.250000	759.000000	11.500000	36.200000	0.7
807.000000	114.000000	1444.752723	1013.000000	17.550000	49.060076	0.9
949.000000	141.000000	1659.000000	1287.000000	23.500000	61.800000	1.2
2683.000000	340.000000	2775.000000	2523.000000	44.600000	88.700000	2.1

Co-Relations between gasses:

Pie chart

Each slice of the pie chart represents the mean of concentration of a gas throughout the year. Concentration of tungsten oxide (PTO8_S4_NO2) is maximum.
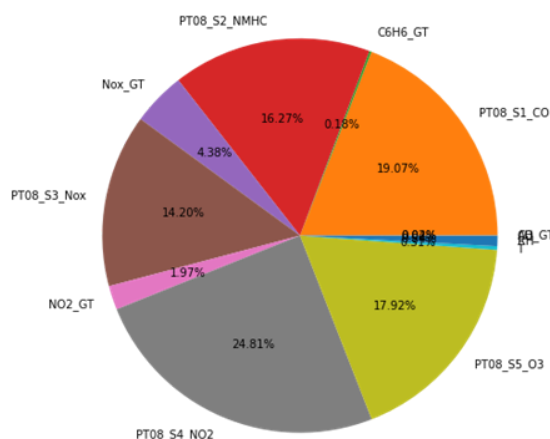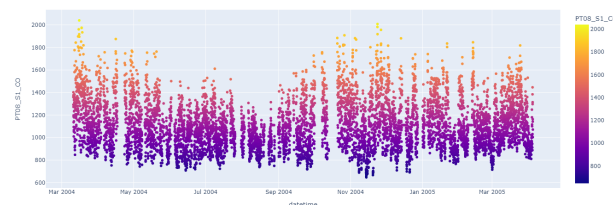

Heat Map

Few gasses are highly correlated. This heat map shows that Non Metanic HydroCarbons has correlation coefficient of 0.98 with titania.Concentration of carbon oxide has correlation coefficient of 0.9 with Non Methane HydroCarbons.
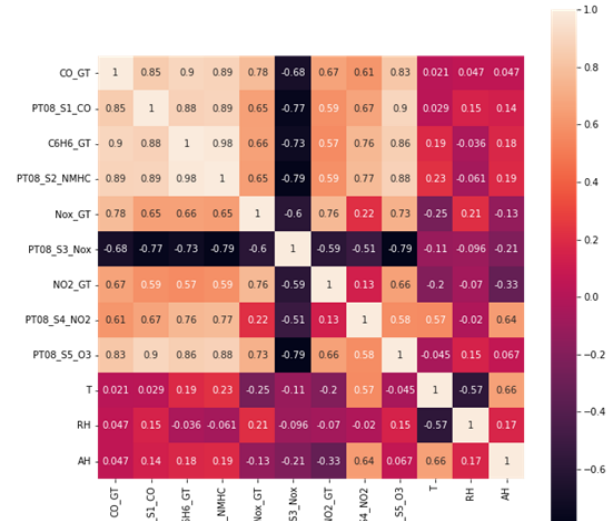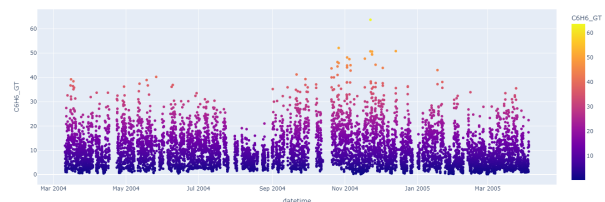

Concentration of oxides over time: (Scatter plots)

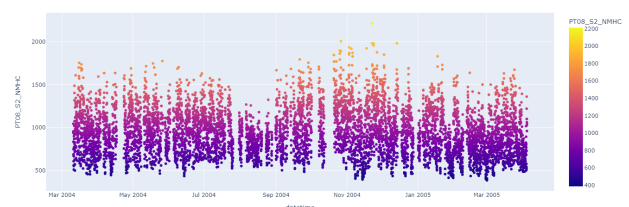Tin oxides over the 2 years

Concentration of tin oxides is maximum in Nov-March(Winters) and minimum in Jul - Sep (Rainy season).
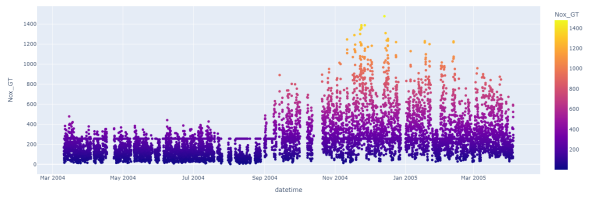
Concentration of benzene over the 2 years

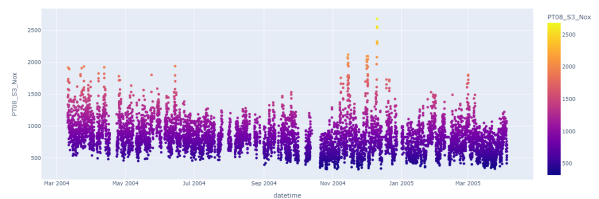Titania(nominally NMHC targeted)over the 2 years

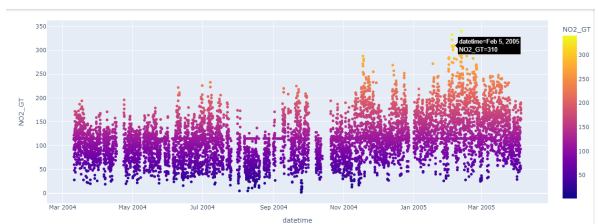## Concentration of Nitric oxide over the 2 years

Concentration of nitric oxides is maximum in Nov-Jan(Winters) and less in Mar- Sep (summer season).

Tungsten oxide (nominally NOx targeted) over the 2 years
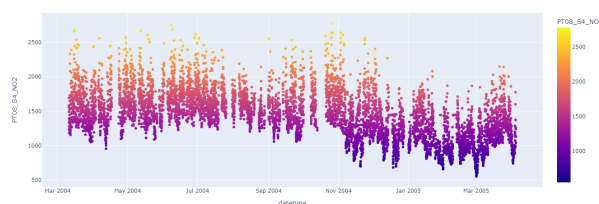

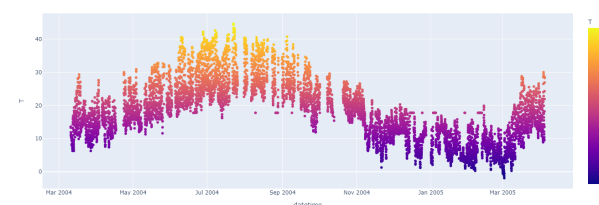## Nitrogen dioxide over the 2 years.

Concentration of nitrogen dioxide is maximum in february.

Tungsten Oxide(nominally NO2 targeted) over the two years
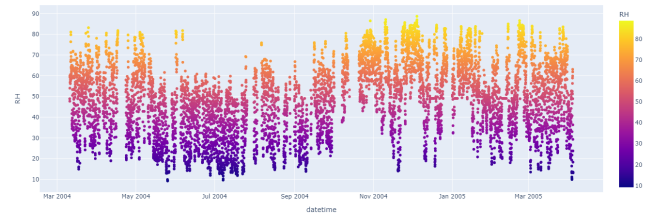

Concentration of tungsten oxide doesn't have any specific relation with date or season.

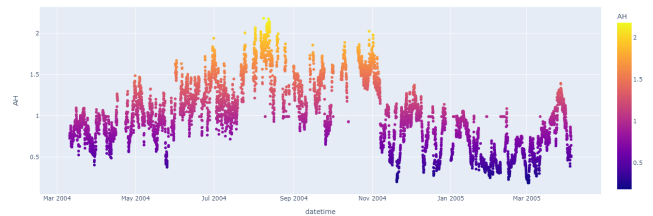## Temperature over the year

This graph clearly shows variation in temperature throughout the year. Temperature starts rising in March and is highest in June. Temperature starts reducing after September and is lowest in December.
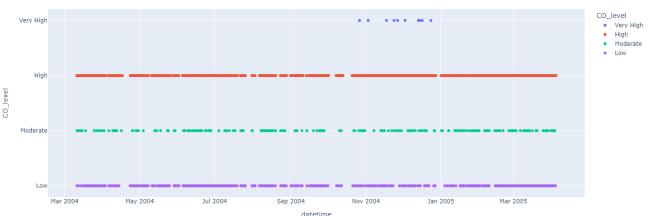
Relative humidity V/S datetime

Absolute humidity V/S datetime
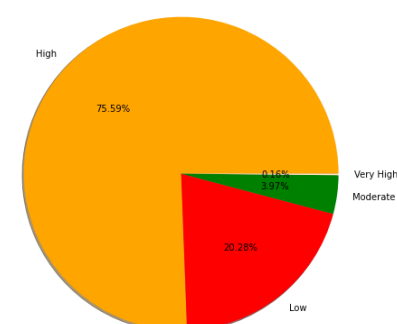

CO level throughout the year

Note: Scatter plot might not show any data point but only a graph sheet in the output cell. But when the plot is downloaded, the correct plot is downloaded with all the data points.(Example attached below)
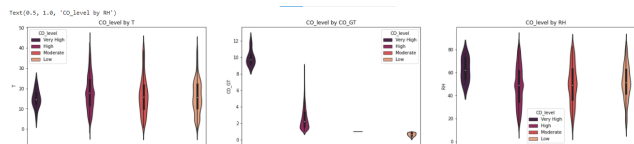

Features of Air depending upon the level of Carbon oxide concentration
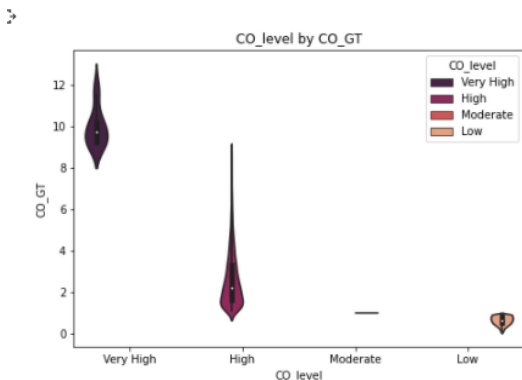
Pie chart

This pie chart shows that most of the year, the concentration of Carbon oxide is high.

Violin Plot

Showing the relation between level of CO concentration and Temperature, Relative humidity , Absolute humidity , I have plotted three violin plots as subplots.

CO_level by Temperature

 The white dots in the middle show the median. Density of the data(CO) is between 10-12 when the CO_level is very high. Density of CO concentration is between 2-6 when the CO_level is high. Density of CO conc. is between 0-1 when the CO_level is low.

CO_level by Relative humidity
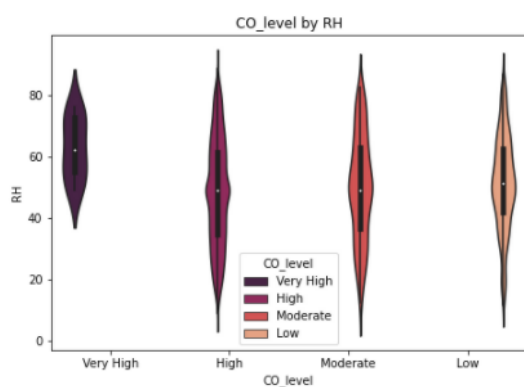
This graph shows that Relative humidity is not directly dependent on CO_level.

In the following cells, I have shown the features of different gasses according to the levels of carbon.
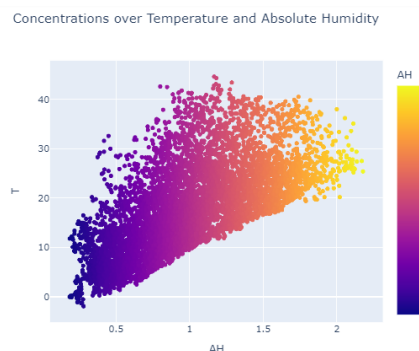
Relation between relative humidity/ absolute humidity and temperature and concentration of humidity.
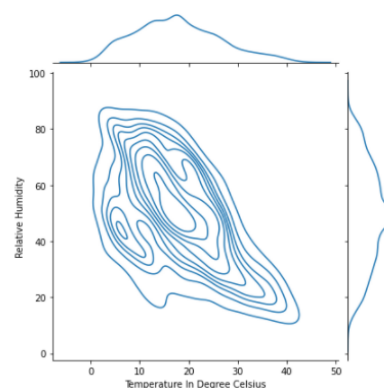
Relation between relative humidity and temperature
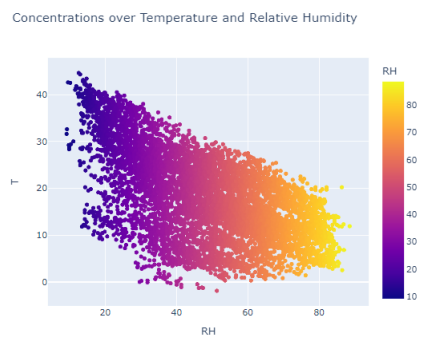
This scatterplot is showing that more the temperature, lesser the relative humidity.It means relative humidity is indirectly proportional to the temperature.

Relation between absolute humidity and temperature

Absolute humidity is not dependent on temperature .

Joint Plot

Jointplot displays a bivariate density curve on the main plot, showing the relation between relative humidity (dependent variable here) and temperature. This plot also displays density curves on the margins.Graphs on the margin show the distribution of temperature and the third plot placed on the right margin shows distribution of humidity.