Analyzing the Gases for proportions of Acetone, Alcohol, Isopropanol using Electronic Nose and Gas Chromatography

Praveen Samuel Jillella

Amulya Geereddy

Table of Content:

Executive Summary:

Air is a mixture of gases and can be polluted by some gases (such as carbon monoxide, hydrocarbons, and nitrogen oxides), and ash. The World health Organization said that 2.4 billion people died because of the direct problems of air pollution. Therefore, this study is to identify and predict the quality of Air and what proportions of gases are present in the air. To help us to predict the quality of Air and factors impacting the quality (WHO). In this Research we want to study about the functionality of Electronic Nose when we pass different gases and different chemical gases like Methanol, Ethanol and Isopropanol in different proportions and study the resultant values to identify the quality and proportions of chemicals passed and if we can identify the Quality of air passed and what possible chemical gases are present in the Air. This helps us to study about the Electronic Nose sensors and its ability to identify the gas quality when it is exposed to Chemicals with different concentrations and statistically analyze the data (Gas Quality).

Introduction:

Electronic Nose:

An electronic nose is an electronic sensing device intended to detect odors or flavors. The expression "electronic sensing" refers to the capability of reproducing human senses using sensor arrays and pattern recognition systems.

Working of Electronic Nose(elprocus):

Essentially the instrument consists of head space sampling, a chemical sensor array, and pattern recognition modules, to generate signal patterns that are used for characterizing odors.

Electronic noses include three major parts:

1. Sample delivery system,
2. detection system,
3. computing system.

Sample delivery system:

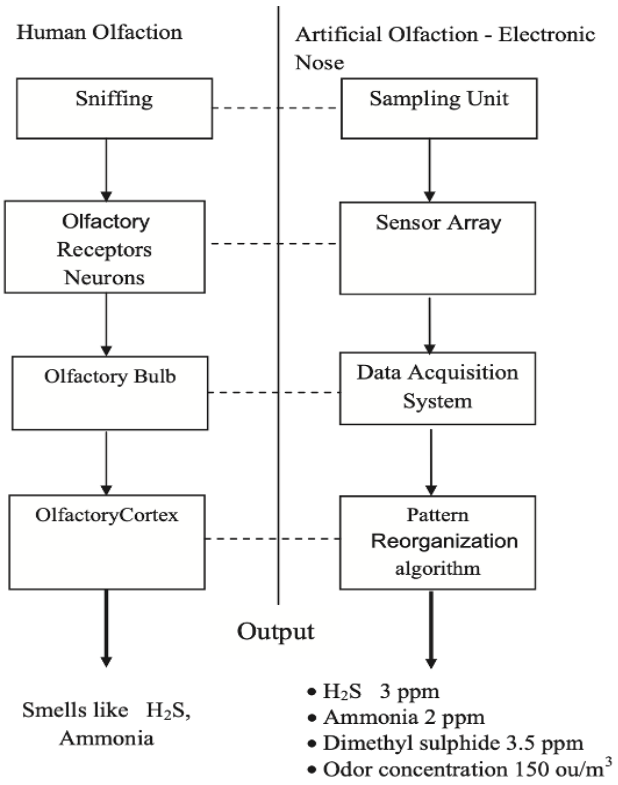
It enables the generation of the headspace (volatile compounds) of a sample, which is the fraction analyzed. The system then injects this headspace into the detection system of the electronic nose. The sample delivery system is essential to guarantee constant operating conditions.

Detection system:

It consists of a sensor set, is the "reactive" part of the instrument. When in contact with volatile compounds, the sensors react, which means they experience a change of electrical properties.

Computing System:

To analyze information from the detection system and provide a pattern recognition output that describes the odor or aroma.



(Reference 1)

In most electronic noses, each sensor is sensitive to all volatile molecules but each in their specific way. However, in bio-electronic noses, receptor proteins which respond to specific odor molecules are used. Most electronic noses use chemical sensor arrays that react to volatile compounds on contact: the adsorption of volatile compounds on the sensor surface causes a physical change of the sensor. A specific response is recorded by the electronic interface transforming the signal into a digital value.

The best-known electronic nose is the breath analyzer. As drivers breathe into the device, a chemical sensor measures the amount of alcohol in their breath. This chemical reaction is then converted into an electronic signal, allowing the police officer to read off the result. Alcohol is easy to detect, because the chemical reaction is specific, and the concentration of the measured gas is high (uc).

Literature Review:

There are studies that identify the quality of gas using Electronic Nose. We were able to analyze and understand the functionality of Electronic Nose and ways we can use for Analysis and different implementations of Statistical Models. Data fusion technology based on the multi-sensor system can obtain the holistic properties of samples. Combining the multi-data-fusion-attention network (MDFA-Net) with the electronic nose (e-nose) and hyperspectral system to identify the egg quality and designing the feature adaptive learning (FAL) unit to select effective information and enhance the ability of feature expression based on the FAL unit to identify the fusion information

of e-nose and hyperspectral system and evaluating the results by implementing deep learning network models to predict the quality of Egg (QinglunZhang 2022).

Evaluating the performance of a specific neural network, used for pattern classification of the electronic nose (e-nose) (Levenberg–Marquardt (LM) 2005). Optical sensors have also been employed in some research laboratories. The end of an optical fiber is coated with a material that reacts with odorants to produce a shift in luminescence. The problem of finding an optimal architecture/configuration is extremely difficult to address because each specific architecture configuration has a unique set of optimal parameters (Levenberg–Marquardt (LM) 2005). If the sensors function independently, one does not expect to observe a dependency among their responses. Under this circumstance, the main eigenvalues would indeed represent the important sensors. the capability of the LM neural network training algorithm in solving the problem of odor recognition using an electronic nose (Levenberg–Marquardt (LM) 2005).
There are studies that sheds light on the practical applicability of electronic nose for the effective industrial odor and gaseous emissions measurement. The applications categorization is based on gaseous pollutants released from the industries. Calibration and calibration transfer methodologies. An electronic nose has the strengths of both olfactometry and analytical instruments since it has ability to determine (after proper training) odor as well as odorant concentration. Most of the cited efforts for measurement or monitoring of industrial odors show satisfactory results. The interference caused by environmental parameters such, as humidity and temperature need to be understood and controlled. Given the complexity that is present in the industrial gaseous emissions, equal attention needs to be paid to all the three parts of the system, viz., sampling technique, sensor array and data ana- lysis techniques. There are various odorous gases such as sulphides group (di ethyl sulphide, diphenyl sulphide, amines, aldehydes and ketones, acids, organic hetero- cycles etc.) having very low odor threshold detection limit as well as PEL i.e., permissible exposure limit (S. Deshmukh et al. / Talanta 144 (2015).
The compounds have different complexities associated that need to be considered while making application specific electronic nose devices. Further, the data analysis should not be merely restricted to qualitative and quantitative analysis of the samples. An electronic nose, when deployed in network mesh, can provide impending solution to on- site and online monitoring of odorous emissions from industry and thus enable better control measures. Hence, on-site calibration algorithms and transfer methodologies need to be focused. Given the advantages electronic nose offers on other techniques, it can be successfully used in this field. However, it has shortcomings but when trained for specific industries, it can be very a useful instrument for measuring the odor intensity and concentration as well as odorant concentration.
(S. Deshmukh et al. / Talanta 144 (2015)). Feature selection algorithms using principal component analysis (PCA) and Best First (BC) coupled with correlation-based feature subset selection (CfsSubsetEval) method were in used to obtain the most efficient feature subsets. Results for the BC feature selection method identified 3 optimized sensors (S2, S6, and S11), suggesting that aromatic compounds relate more to the identification of the samples. (Lim et al. Medicine (2020) 99).
Feature extraction from the sensor signals is a key procedure to further improve the performance of an E-nose, but the evaluation of a feature extraction method is influenced by the type of sensors, parameters of experiments, detection targets, demands of specific application, and so on. Different methods are suitable for different situations, and we must choose the method according to the actual conditions. Although there are no widely approved evaluation criteria for various features,

we can give some advice on feature extraction according to the previous research. (Jia Yan 2015). The sensor array includes sixteen metal oxide (MOX) sensors of four different types (TGS-2600, TGS2602, TGS- 2610, TGS-2620) of the sensor of four units each and is exposed to two different binary gas mixtures: ethylene–CO and ethylene–methane. we learn that the sensors are calibrated before deployment, so the sensory data obtained after deployment during the data collection span will contain negligible drift and carry the sensor array features. Sensor drift and noise are typically caused by sensor design errors and non-ideal causes. Hence, we can use raw dataset timeseries measurements, and virtual sensor drifts to synthesize drifted measurements (DL-based gas identification and quantification.). PCA was developed to assess groupings among the samples according to the roasting intensity level. The principal components (PC) one and two were selected based on them summing > 60% of sensors 2021, 21, 2016 5 of 15 data variability, which is the cutoff point considered to test the significance of the PCA. (JiaYan 2015). Usage of BN is introduced to maintain the stability of convolution parameters and ReLU activation function to realize the nonlinear mapping of convolution. From another paper from we have understood the combination of SPME-GC-MS(Ying Le 2022) and E-nose technology is conducive to the comprehensive study of food flavor from the macro and micro level, which is currently the main method for detecting food volatile flavor substances. In Addition to it we learnt Pattern discrimination enables straightforward and easy interpretation of qualitative and semi-quantitative data.

The combination of the E-nose, (Ying Le 2022) SPME-GC-MS and HS-GC-IMS technology could make up for their respective limitations, more comprehensively reflected the changes of volatile components. The high efficiency and sensitivity of HS-GC-IMS offered the possibility of using it as a visualized tool for monitoring the flavor change. During the research, it was found that most of the production of flavor substances was related to the degradation and metabolism of flavor precursors, such as amino acids and fatty acids.

(Zhenjiang 212013, China). the economical E-nose serves as a quick method for initial screening while the sensitive GC-IMS could provide more details for subsequent further analysis. E-nose data was shown to be effective in clustering different edible oils and distinguishing between pure and adulterated oils using linear discriminant analysis (LDA), albeit with poor performance in quantitative analysis of adulteration rates by partial least squares (PLS) (Nanjing 2022).

1. PCA is often used to reduce data dimensionality, enabling visualization of information within a dataset. In addition, PCA allows calculation of variables that best describe differences among the samples and ranking according to their contribution, namely principal components (PCs).
2. LDA is a statistical method that enables grouping of samples by maximizing variance among classes and minimizing variance within classes, thus enabling resolution among classes to be optimized (Nanjing 2022). Combining non-destructive analysis technology (multi-source information) can overcome the above problems. The present review focused on applying multi-source and non-destructive information on herbs and spices quality authentication, including vibration spectroscopy and electronic sensor technologies. (YulinXu 2022).

Theory and Hypothesis Development (Research Methodology):

The Research was conducted in the University of South Florida laboratory under guidance of Professor Takshi on Electronic Nose by passing normal gas and different proportions of Methanol, Isopropanol and Ethanol and the readings wear recorded.

Apparatus Used:

Electronic Nose, Methanol (137 ppm, 164 ppm, 292 ppm), Isopropanol (217 ppm,117 ppm,143 ppm), Ethanol (200 ppm, 123ppm,161 ppm)

Data Collection:

The Electronic Nose is exposed to Normal gas and after the values of the Electronic Nose array are stabilized then the Baseline values are Recorded. And the Electronic Nose is exposed to Methanol, Isopropanol and Ethanol gases and the values are recorded once the Electronic Nose Array values started stabilizing. The data was collected multiple times and we tried to Analyze the data collected separately and combined with the old data.

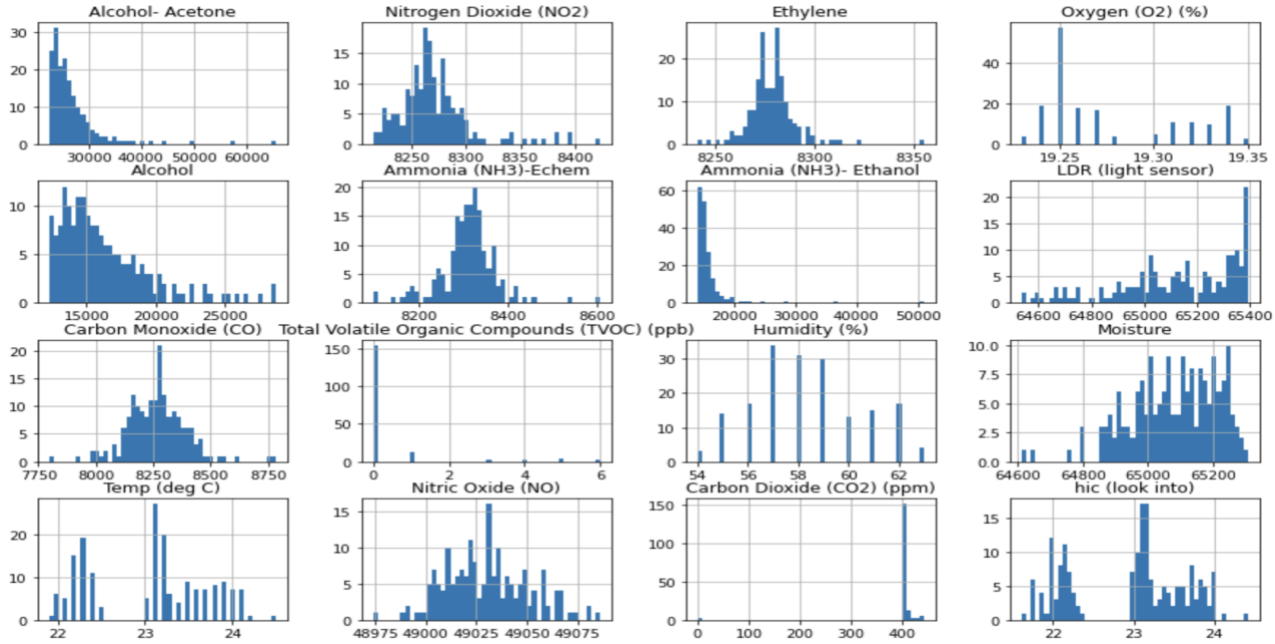
Data Description:

From the Data Collected the data columns with label Ammonia, Nitrogen Dioxide, Ethylene, Carbon Monoxide, Alcohol-Acetone, Alcohol, Ammonia and Nitric Oxide are the Gas sensor's data. The range of the values recorded for the gas sensors listed above can be between 0 to 65535. These values have no units to them. The baseline sheet helps to define what would be the estimated minimum values detected by them. **(Minimum values from the baseline are listed below in the "Baseline" sheet.)**. We also have variables like Carbon Dioxide, Total Volatile organic Compounds, Oxygen and Humidity we can consider for further analysis. In column **O**, the **oxygen** would normally be **20% ±2%.** All other columns in the data collected can be analyzed or ignored based on the interest and we can observe the difference between the listed values in gas sensors can be caused due to their sensitivity to the exposed gases.

Data Analysis:

Data has been collected as 3 sets at 3 different Atmospheric conditions so when we are describing data and doing data Analysis, we tried to Analyze the data for each set of data collected for Baseline, Ethanol, Methanol, and Isopropanol.

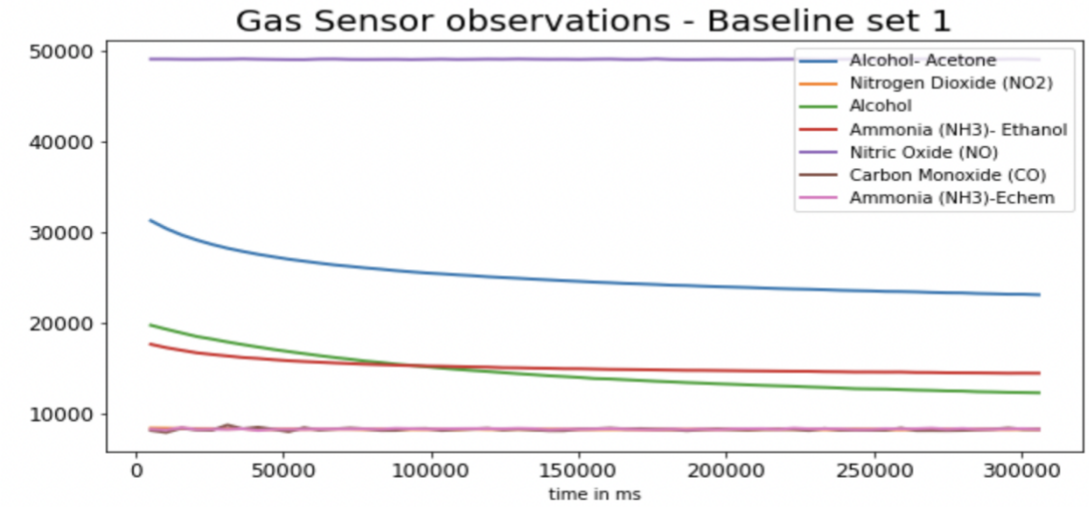
BASELINE:



We can observe from the Baseline data that the **Alcohol-Acetone, Alcohol, Ammonia-Ethanol** is "**right skewed"**, sometimes this type of distribution is also called "positively" skewed. **Oxygen, TVOC, Humidity, hic**, **Moisture, Carbon Dioxide** the data distribution is **"random". Nitrogen Dioxide, Ammonia, Nitric Oxide, Carbon Monoxide, Temp** are "**Normally distributed**". **LDR** is "**left skewed**". **Ethylene** is Bi-Model.

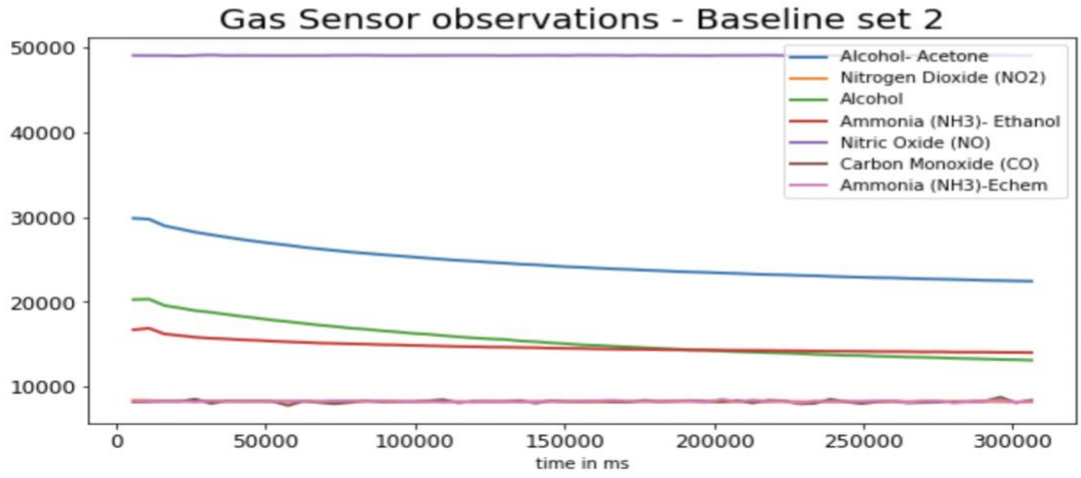
Data Description for Baseline set 1 and its Atmospheric conditions:

```
LDR (light sensor)    65236.406780
Moisture              64988.728814
Oxygen (O2) (%)          19.256610
Humidity (%)             56.322034
Temp (deg C)             23.789831
dtype: float64
```



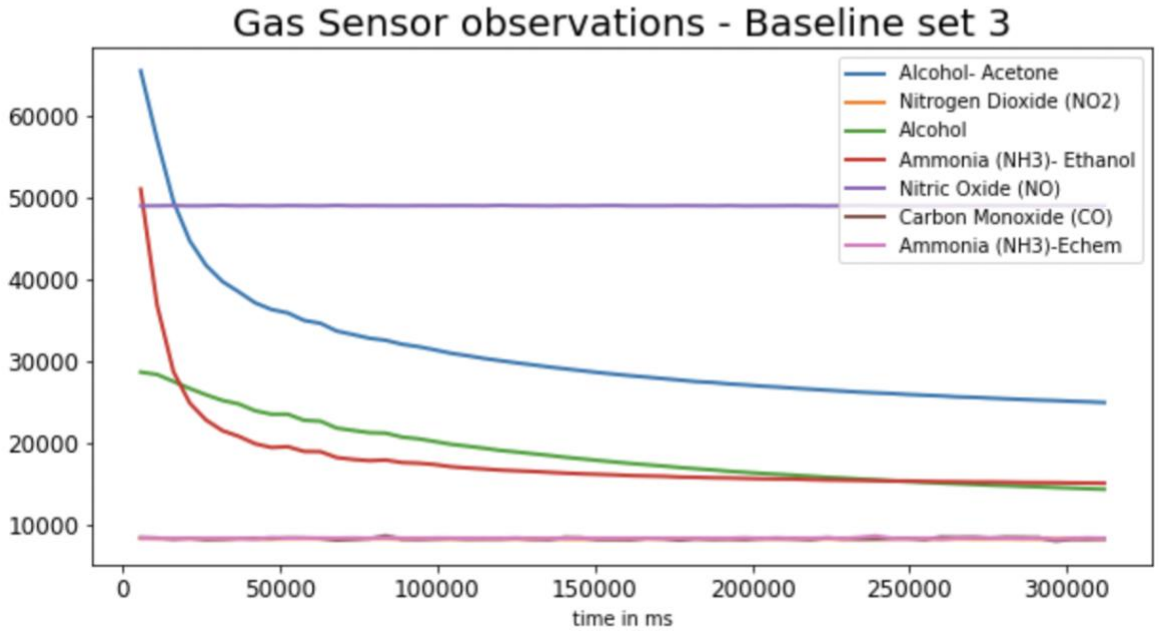
Data Description for Baseline set 2 and its Atmospheric conditions:

```
LDR (light sensor)    64938.305085
Moisture              65047.000000
Oxygen (O2) (%)          19.326102
Humidity (%)             60.864407
Temp (deg C)             23.152542
dtype: float64
```



Data Description for Baseline set 3 and its Atmospheric conditions:

```
LDR (light sensor)    65199.666667
Moisture              65200.050000
Oxygen (O2) (%)          19.249667
Humidity (%)             57.950000
Temp (deg C)             22.250000
dtype: float64
```



Models and Algorithms Used:

We used PCA and t-SNE Algorithms in our analysis on the data.

Principal component analysis (PCA) is a technique that transforms high-dimensions data into lower-dimensions while retaining as much information as possible.

How does PCA work(PCA)?

PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. It's a two-step process. We can't write a book summary if we haven't read or understood the content of the book.

t-SNE(t-SNE):

t-SNE) t-Distributed Stochastic Neighbor Embedding is a non-linear dimensionality reduction algorithm used for exploring high-dimensional data. It maps multi-dimensional data to two or more dimensions suitable for human observation.

Working of t-SNE:
t-SNE a non-linear dimensionality reduction algorithm finds patterns in the data by identifying observed clusters based on similarity of data points with multiple features. But it is not a clustering algorithm it is a dimensionality reduction algorithm. This is because it maps the multi-dimensional data to a lower dimensional space, the input features are no longer identifiable. Thus, you cannot make any inference based only on the output of t-SNE. So essentially it is mainly a data exploration and visualization technique.

Random Forest Classifier**(Random Forest Classifier):**

Random Forest is a powerful and versatile supervised machine learning algorithm that grows and combines multiple decision trees to create a "forest." It can be used for both classification and regression problems
Random Forest grows multiple decision trees which are merged for a more accurate prediction. The logic behind the Random Forest model is that multiple uncorrelated models (the individual decision trees) perform much better as a group than they do alone. When using Random Forest for classification, each tree gives a classification or a "vote." The forest chooses the classification with the majority of the "votes." When using Random Forest for regression, the forest picks the average of the outputs of all trees.

Feed Forward Neural Network:

Feed-forward neural networks allow signals to travel one approach only, from input to output. There is no feedback (loops) such as the output of some layer does not influence that same layer. Feed-forward networks tends to be simple networks that associates inputs with outputs. It can be used in pattern recognition. This type of organization is represented as bottom-up or top-down. a series of inputs enter the layer and are multiplied by the weights. Each value is then added together to get a sum of the weighted input values.

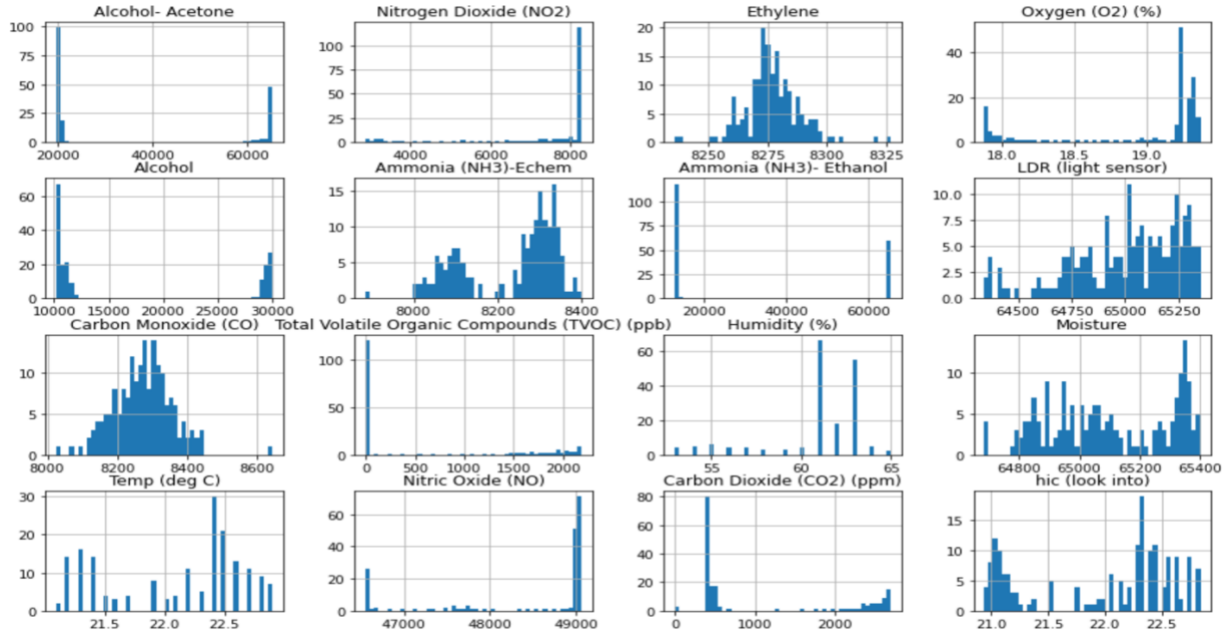
Multiclass Classification Neural Networks:

Classification jobs with more than two class labels are referred to as multi-class classification. Multiclass classification in machine learning, unlike binary classification, does not distinguish between normal and pathological results. Instead, examples are assigned to one of several pre-defined classes.

We have Analyzed the data by implementing PCA and TSNE on each sheet. i.e., we performed PCA analysis for Baseline, Ethanol, Methanol, and Isopropanol. We tried to understand the data for each chemical component, so we started by assigning each proportion of respective chemical components like Ethanol, Methanol and Isopropanol as different target and tried to analyze the data by plotting 2D plot for PCA dimensionality reduction by using 2 components which give more deeper analysis and relation in the data. dimensionality reduction is the technique of representing multi-dimensional data (data with multiple features having a correlation with each other) in 2 or 3 dimensions. And combined the whole data and performed PCA analysis and 10 category multi class model for the whole data to study and Analyze data in various aspects.

Principal Component analysis (PCA) is a multivariate data analysis approach that allows us to summarize and visualize the most important information contained in a multivariate data set. And we implemented t-SNE for every data we implemented PCA to study the It's non-linear Dimensionality reduction by t-SNE. We have Implemented Random Forest Classifier on the whole data set and Feed Forward Neural Network along with Multiclass Classification with Activation layers "RELU" and "SOFTMAX" and 25 Epochs.

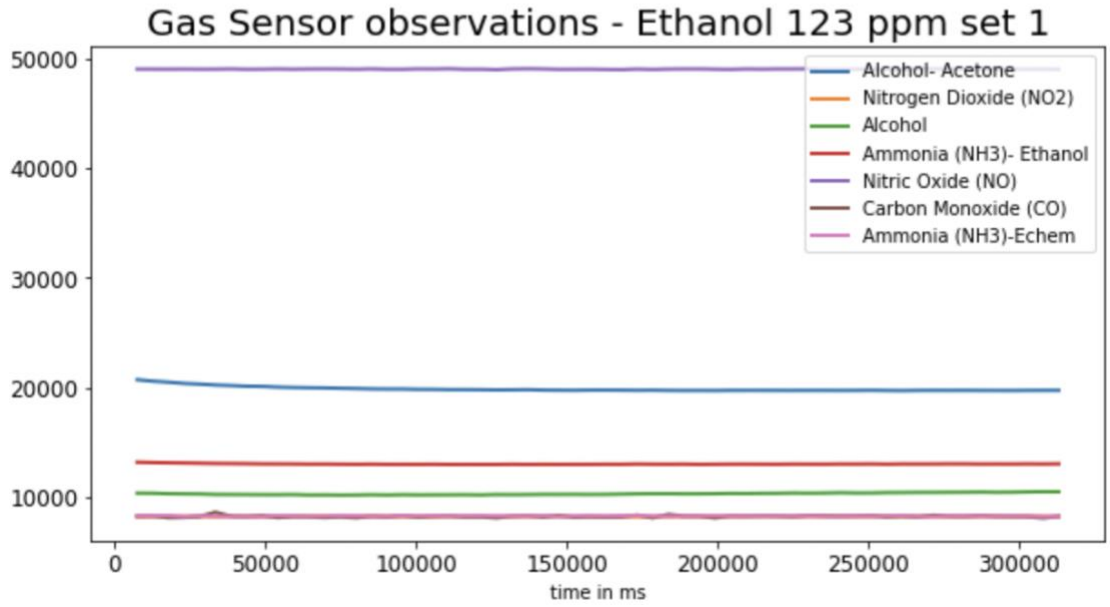
ETHANOL 123 ppm



Ethylene, Carbon Monoxide are Normally Distributed. **Oxygen** is left Skewed. **Ammonia-Echem** is multi modally distributed. All other are randomly distributed.

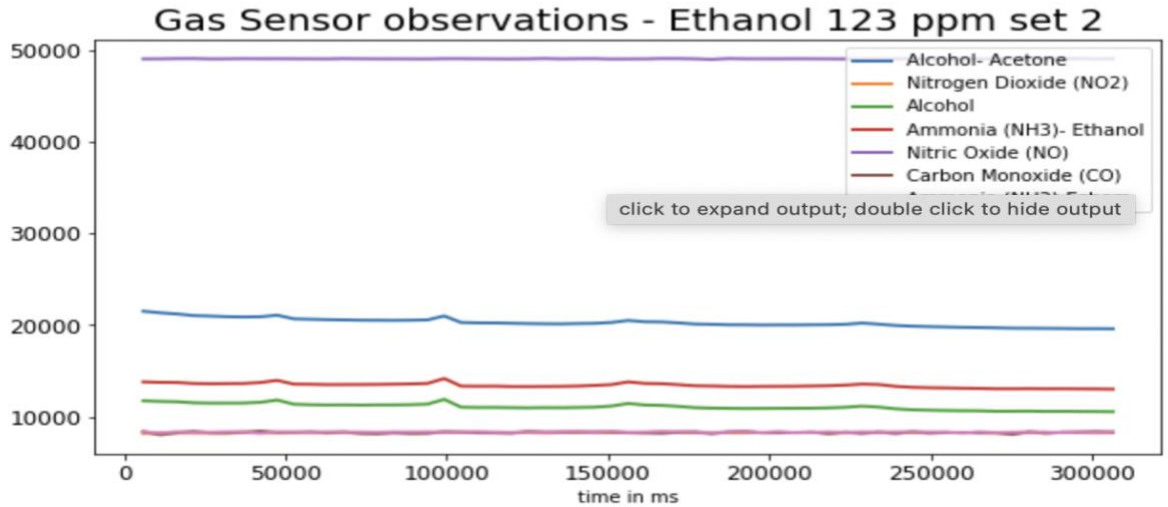
Data Description for Ethanol 123 ppm set 1 and its Atmospheric conditions:

```
LDR (light sensor)    64788.966667
Moisture              65339.166667
Oxygen (O2) (%)          19.327167
Humidity (%)             61.016667
Temp (deg C)             22.555000
dtype: float64
```



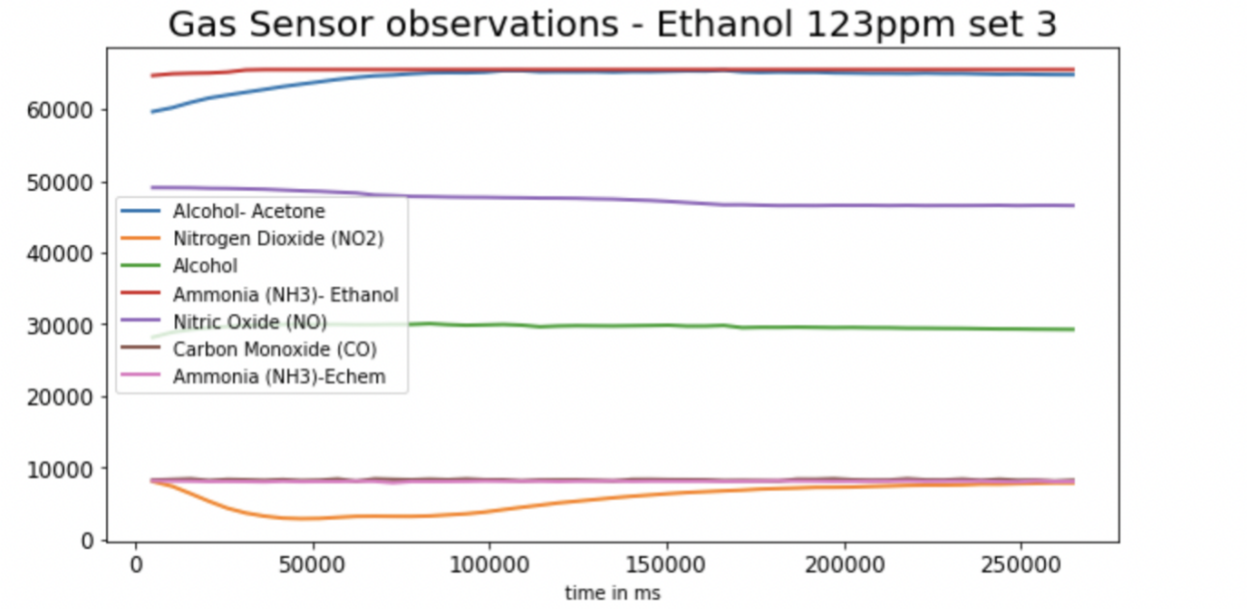
Data Description for Ethanol 123 ppm set 2 and its Atmospheric conditions:

```
LDR (light sensor)    65053.118644
Moisture              64938.203390
Oxygen (O2) (%)          19.234576
Humidity (%)             62.813559
Temp (deg C)             22.366102
dtype: float64
```

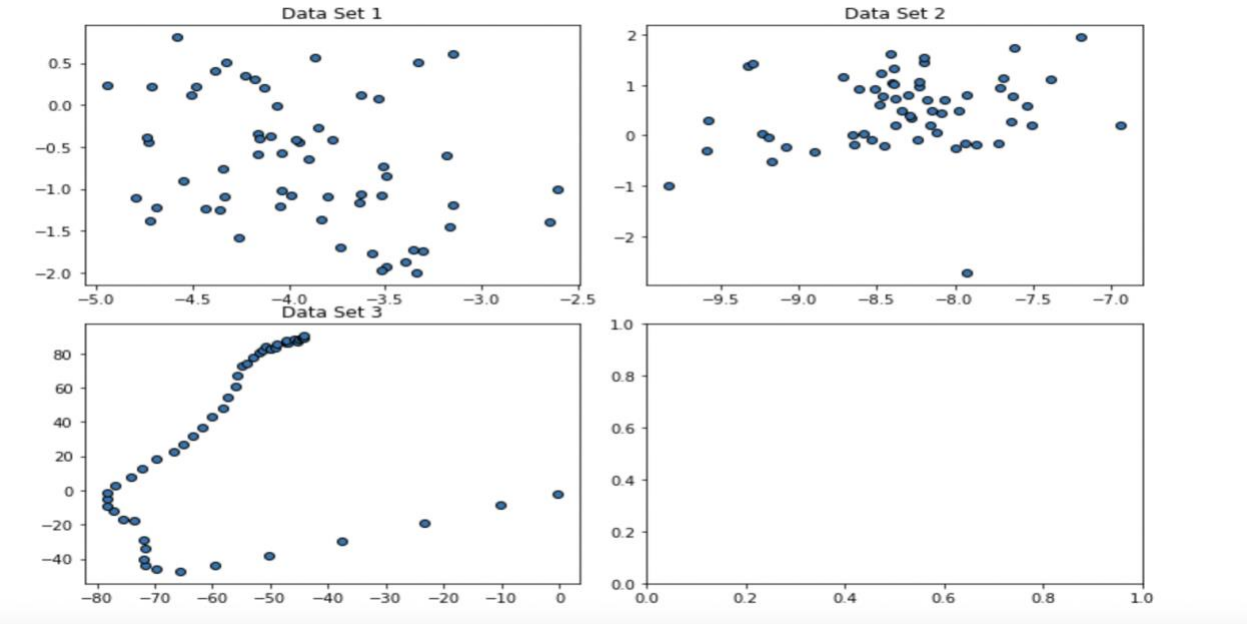


Data Description for Ethanol 123 ppm set 3 and its Atmospheric conditions:

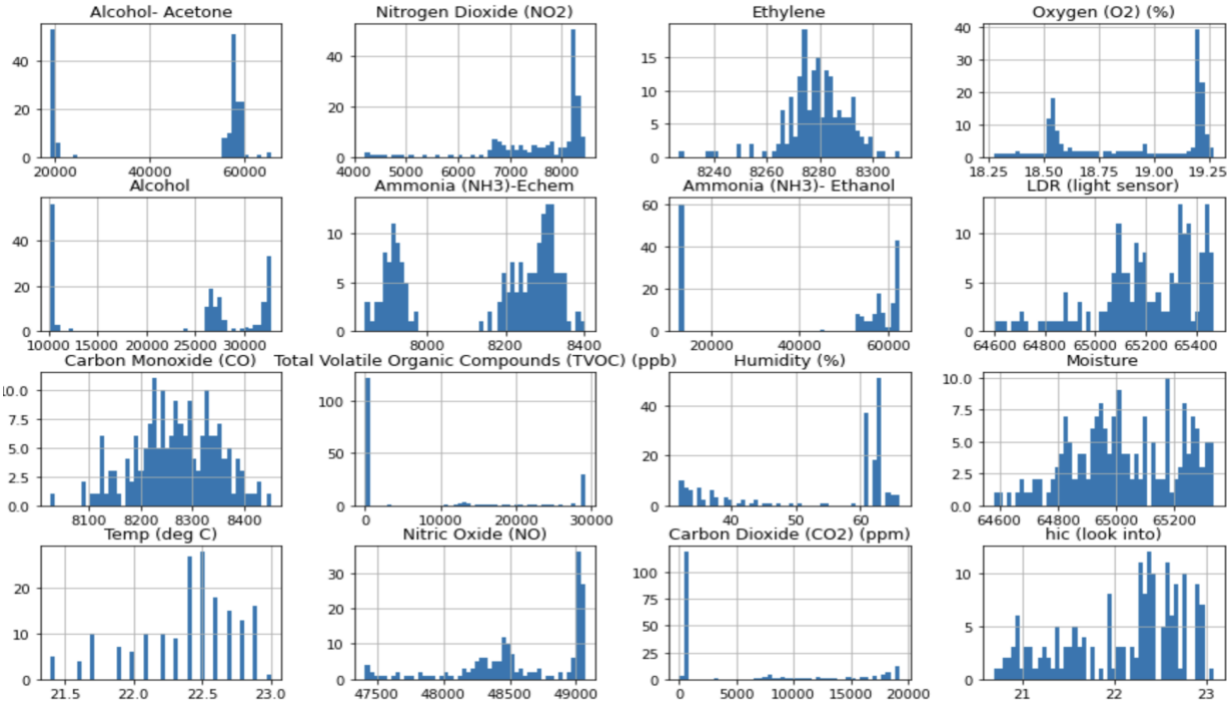
```
LDR (light sensor)    65156.411765
Moisture              65023.098039
Oxygen (O2) (%)          18.345098
Humidity (%)             60.058824
Temp (deg C)             21.325490
dtype: float64
```



After applying PCA on Ethanol 123 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Ethanol 123 ppm PCA Components 1(x-axis) varying from -9.5 to -5 and Component 2(y-axis) varying from 3.5 to 2. With data set 3 we assume the data is not producing the results.



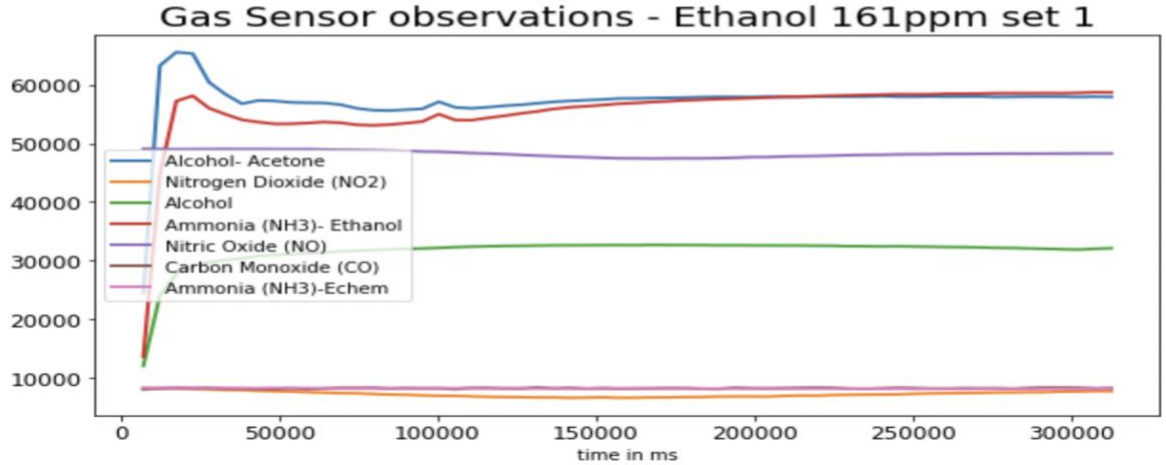
ETHANOL 161 ppm



We can observe from the above ETHANOL 161 ppm histograms that **Ethylene** is "**Normally Distributed"**, **Alcohol-Acetone, Ammonia-Echem, Oxygen, Ammonia-Ethanol** are **Bi-Modally** Distributed components, And **LDR, Carbon Monoxide, Moisture, Temp, hic** data is "**multi model distribution".**

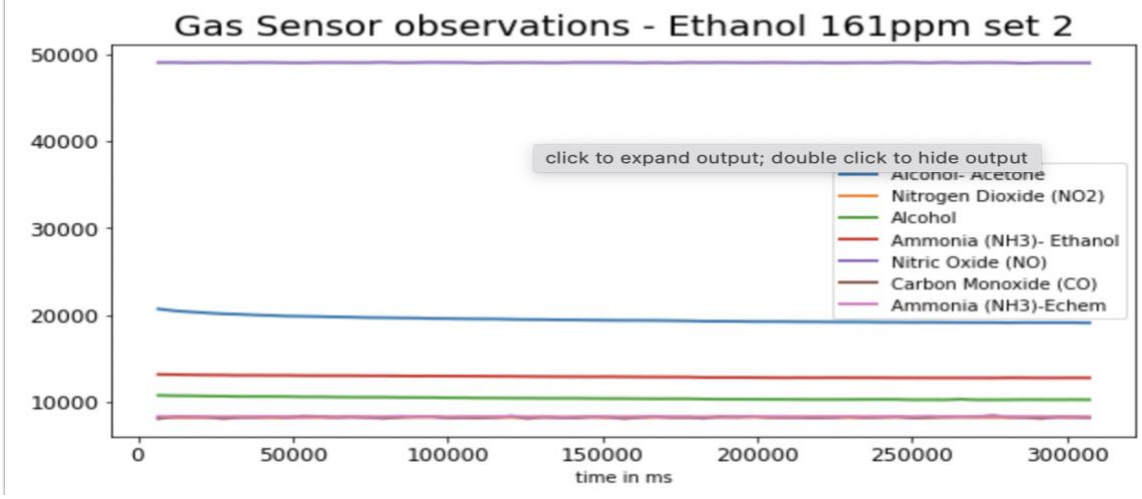
Data Description for Ethanol 161 ppm set 1 and its Atmospheric conditions:

```
LDR (light sensor)     65033.433333
Moisture               64919.700000
Oxygen (O2) (%)           18.853167
Humidity (%)              62.083333
Temp (deg C)              22.595000
dtype: float64
```



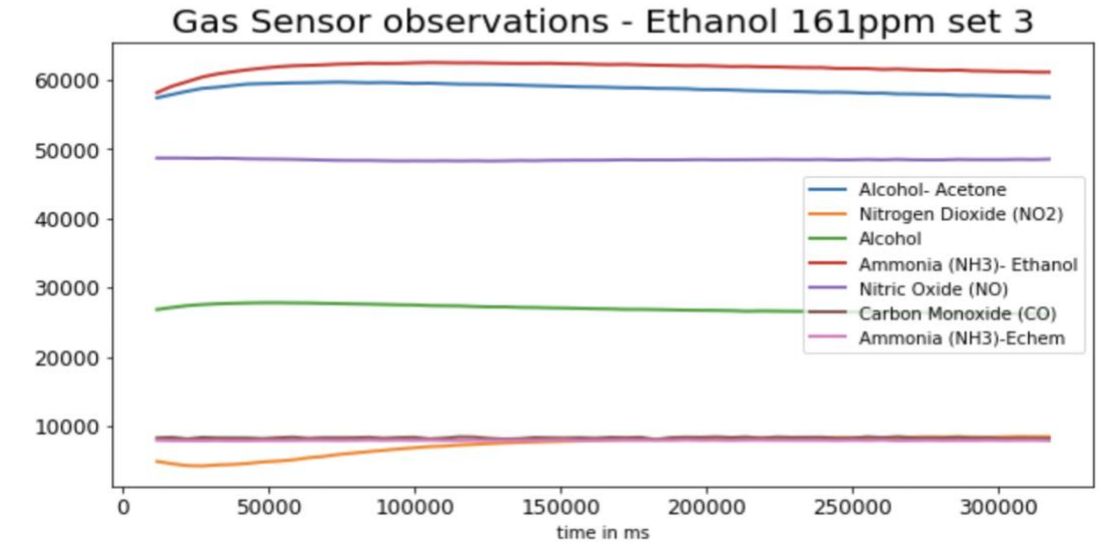
Data Description for Ethanol 161 ppm set 2 and its Atmospheric conditions:

```
LDR (light sensor)     65242.254237
Moisture               65225.406780
Oxygen (O2) (%)           19.200169
Humidity (%)              62.796610
Temp (deg C)              22.513559
dtype: float64
```



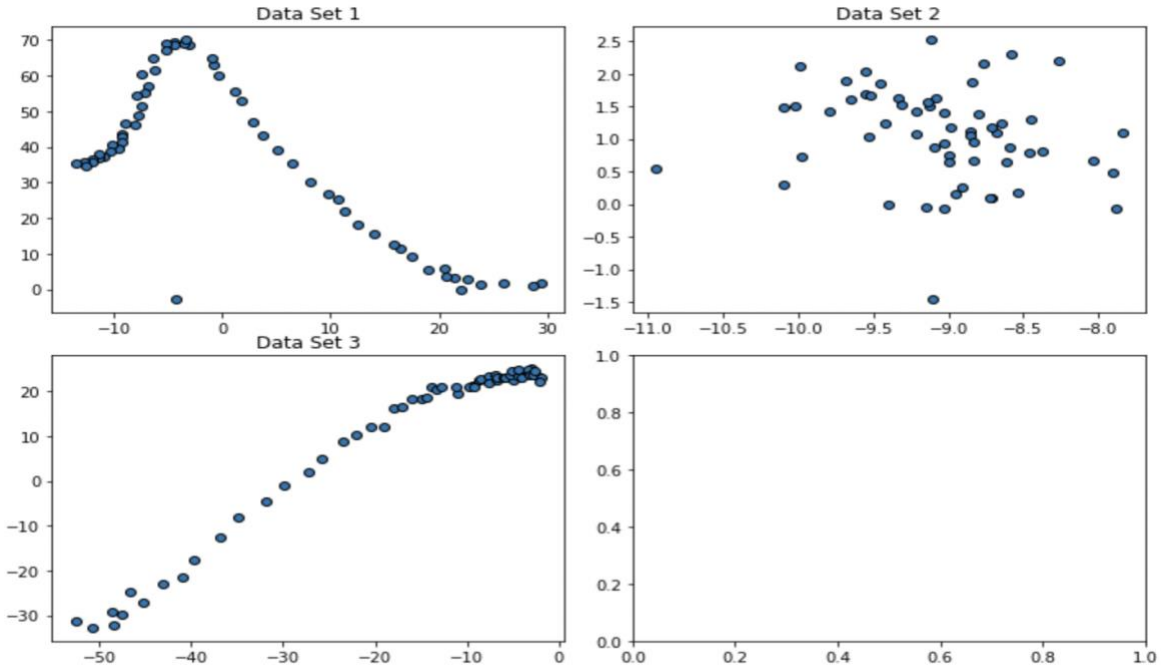
Data Description for Ethanol 161 ppm set 3 and its Atmospheric conditions:

```
LDR (light sensor)     65336.800000
Moisture               64949.016667
Oxygen (O2) (%)           18.624500
Humidity (%)              37.783333
Temp (deg C)              22.055000
dtype: float64
```



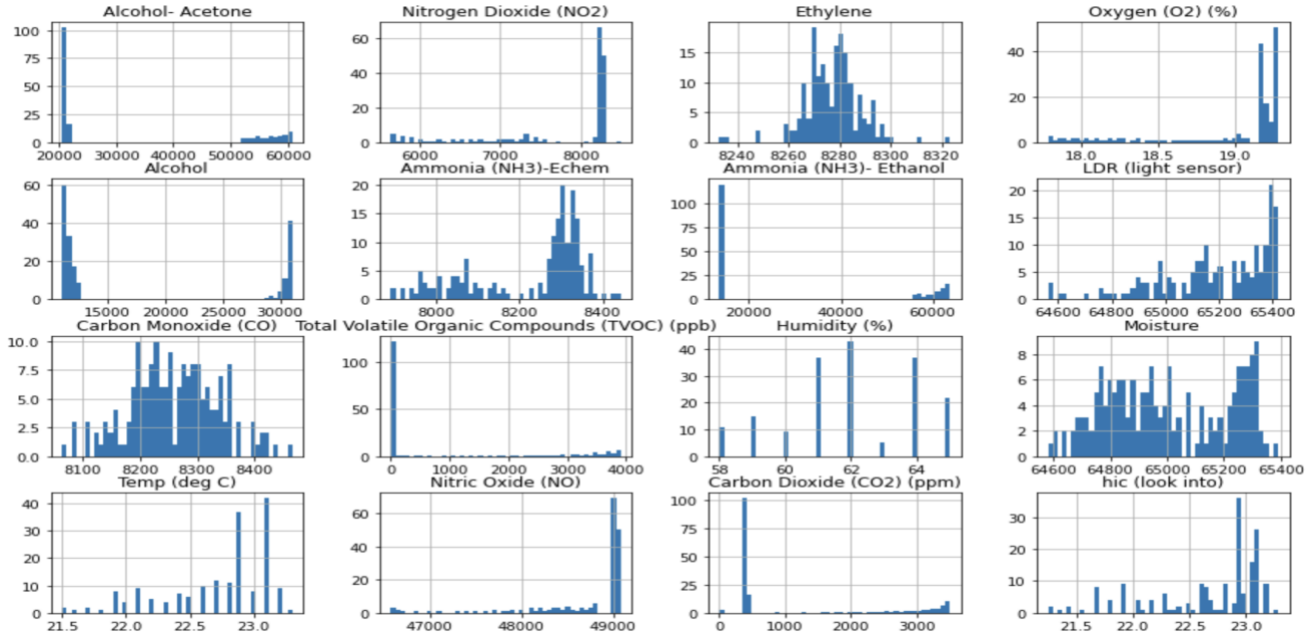
After applying PCA on Ethanol 161 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Ethanol 161 ppm PCA Components 1(x-axis) varying

from -10 to 30 and Component 2(y-axis) varying from 0 to 50. With data set 3 we assume the data is not producing the results.



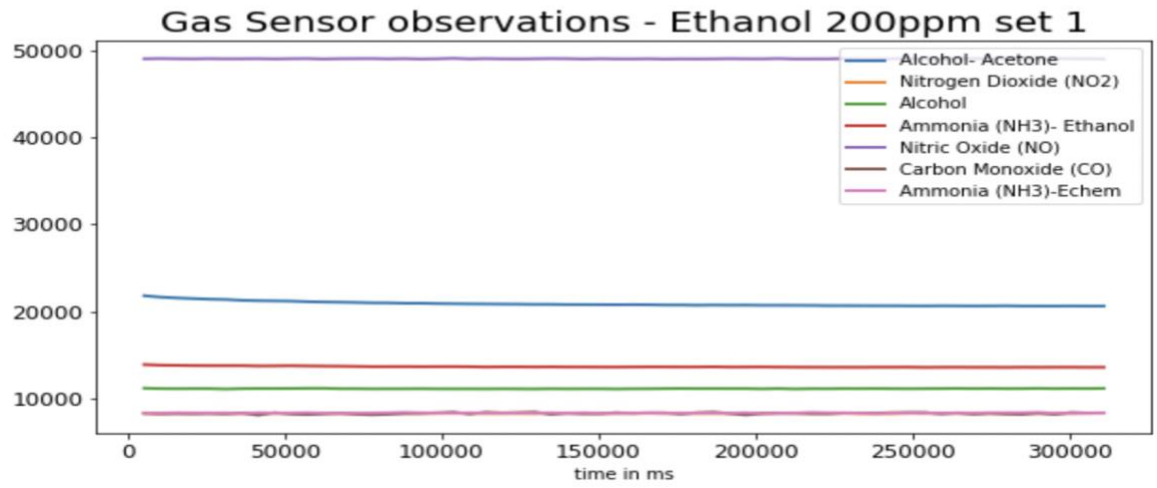
Ethanol 200ppm

We Can observe that **Ethylene** is "**Normally distributed**". And distribution in **Carbon-Monoxide, Moisture** is "**Multi model**". **Nitrogen Dioxide, Oxygen, LDR, Temp, Nitric Oxide, hic** are "**left skewed**". And other components are Randomly distributed.



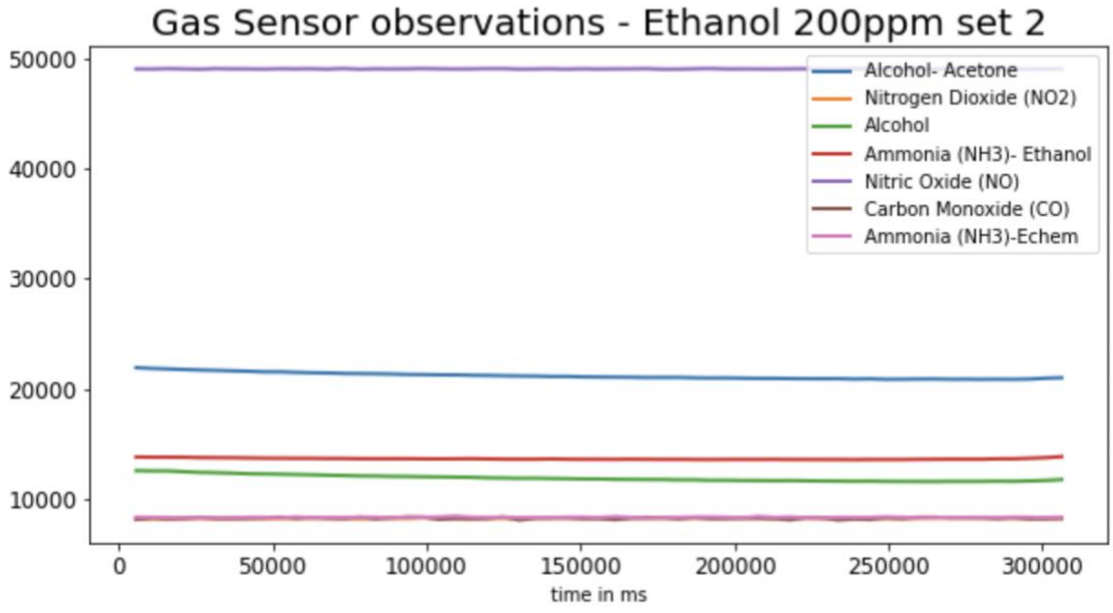
Data Description for Ethanol 200 ppm set 1 and its Atmospheric conditions:

```
LDR (light sensor)     65257.766667
Moisture               65256.566667
Oxygen (O2) (%)           19.178000
Humidity (%)              61.533333
Temp (deg C)              23.105000
dtype: float64
```



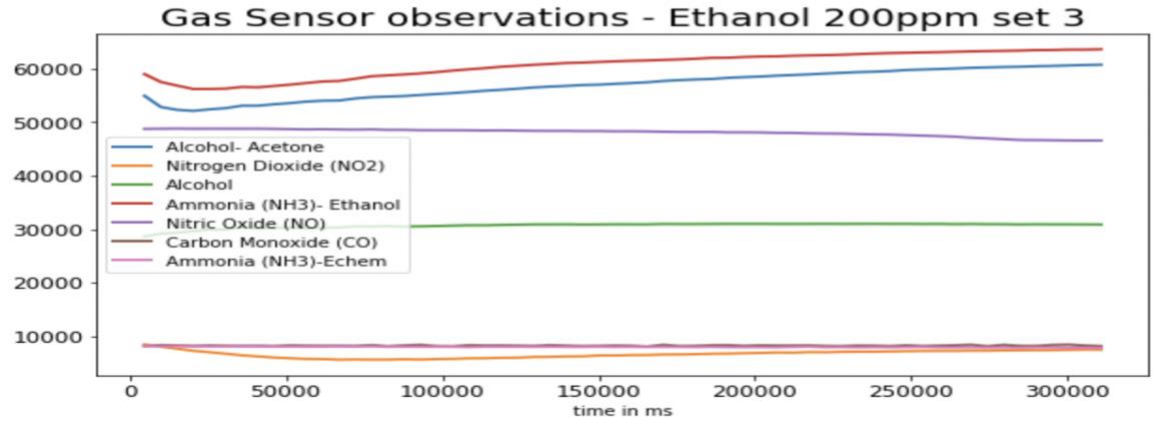
Data Description for Ethanol 200 ppm set 2 and its Atmospheric conditions:

```
LDR (light sensor)     65009.813559
Moisture               64784.423729
Oxygen (O2) (%)           19.256949
Humidity (%)              64.372881
Temp (deg C)              22.805085
dtype: float64
```

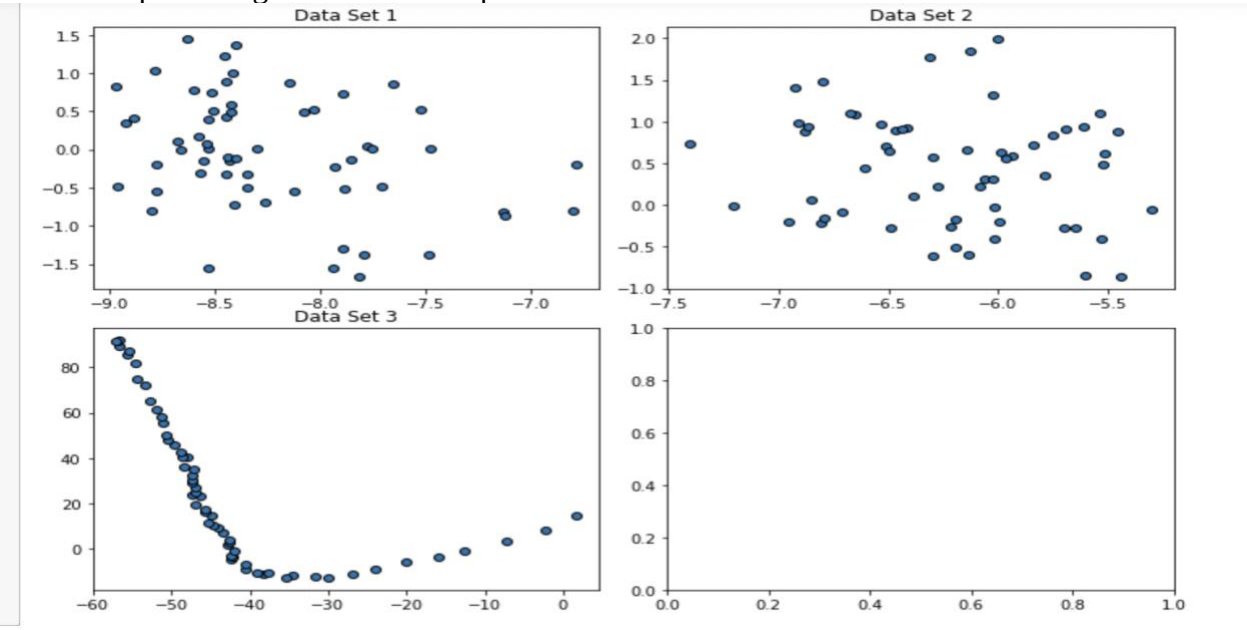


Data Description for Ethanol 200 ppm set 3 and its Atmospheric conditions:

```
LDR (light sensor)     65293.633333
Moisture               64969.483333
Oxygen (O2) (%)           18.424000
Humidity (%)              60.150000
Temp (deg C)              22.248333
dtype: float64
```

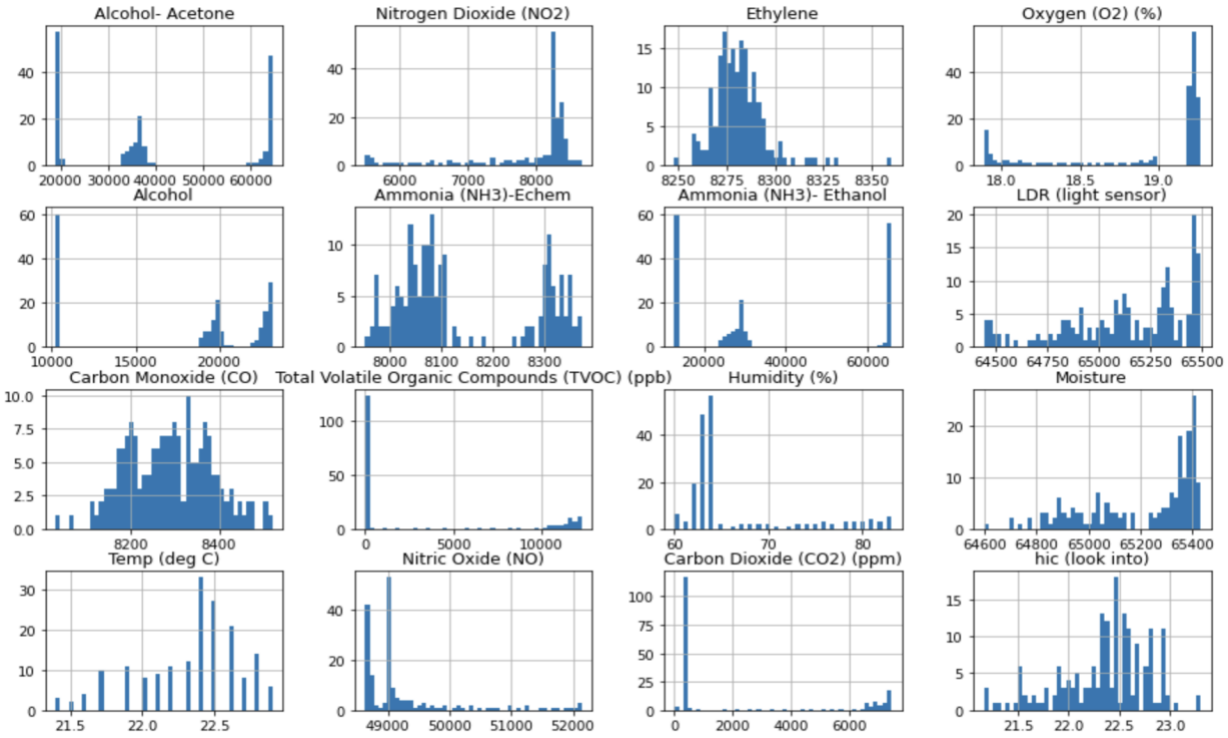


After applying PCA on Ethanol 200 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Ethanol 200 ppm PCA Components 1(x-axis) varying from -9 to -5.5 and Component 2(y-axis) varying from -2 to 2.5. With data set 3 we assume the data is not producing the results as expected.



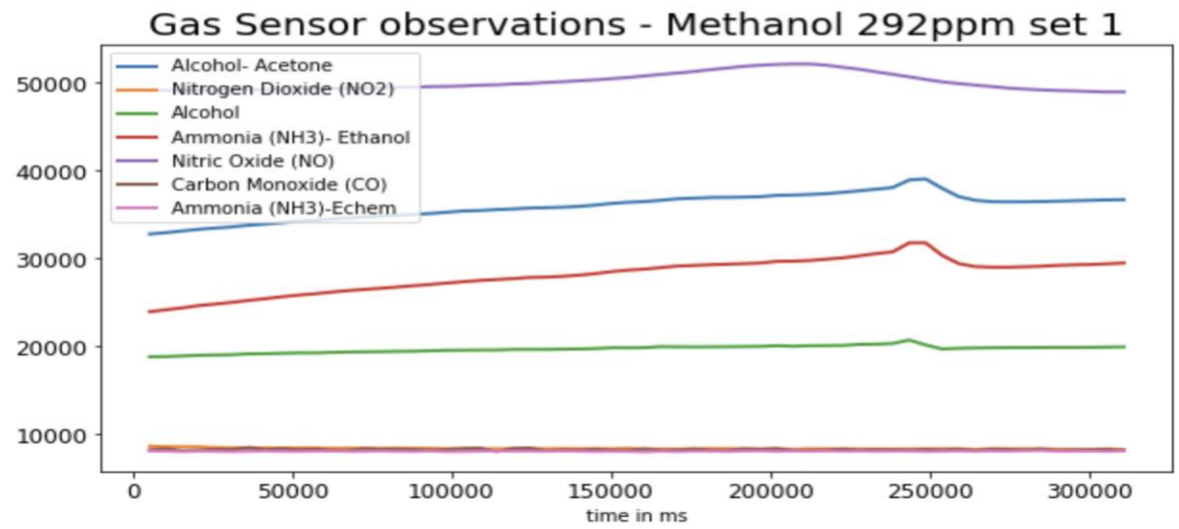
Methanol 292ppm

Ethylene is "**Normally distributed**" and right skewed. **Nitrogen, Oxygen** is "**left skewed**". **Acetone, Ammonia-Echem, Ethanol, LDR, carbon Monoxide, Moisture, hic** is "**multi model distribution"**.



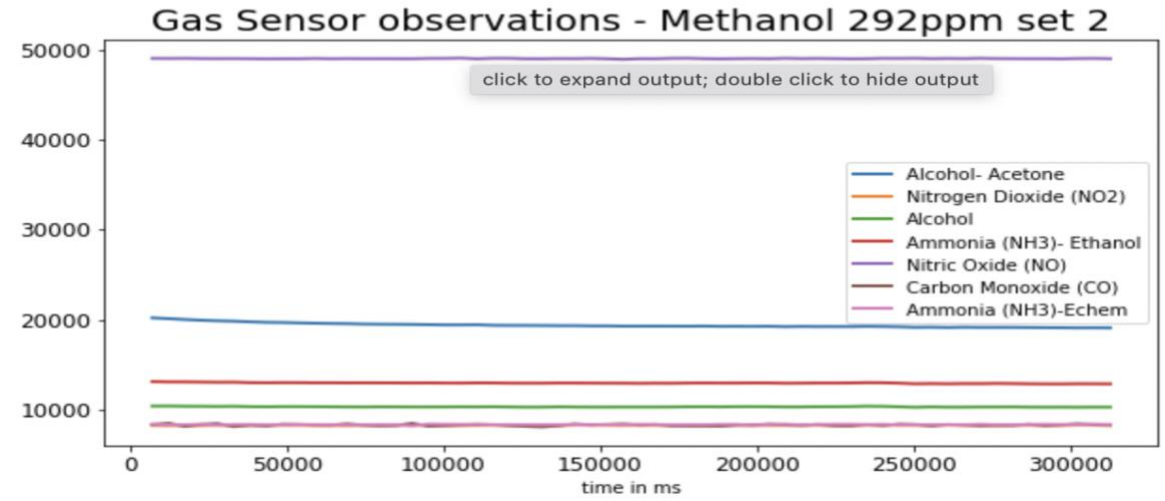
Data Description for Methanol 292 ppm set 1 and its Atmospheric conditions:

```
LDR (light sensor)     64838.733333
Moisture               65378.050000
Oxygen (O2) (%)           19.237000
Humidity (%)              63.933333
Temp (deg C)              22.476667
dtype: float64
```



Data Description for Methanol 292 ppm set 2 and its Atmospheric conditions:

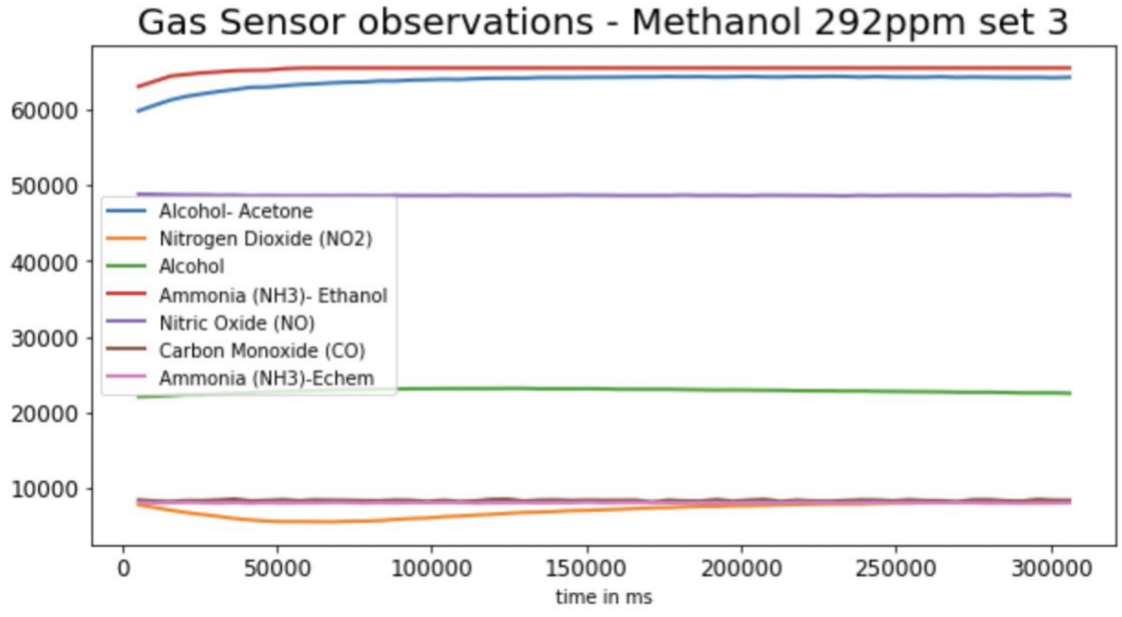
```
LDR (light sensor)     65179.400000
Moisture               64949.133333
Oxygen (O2) (%)           19.200833
Humidity (%)              62.716667
Temp (deg C)              22.456667
dtype: float64
```



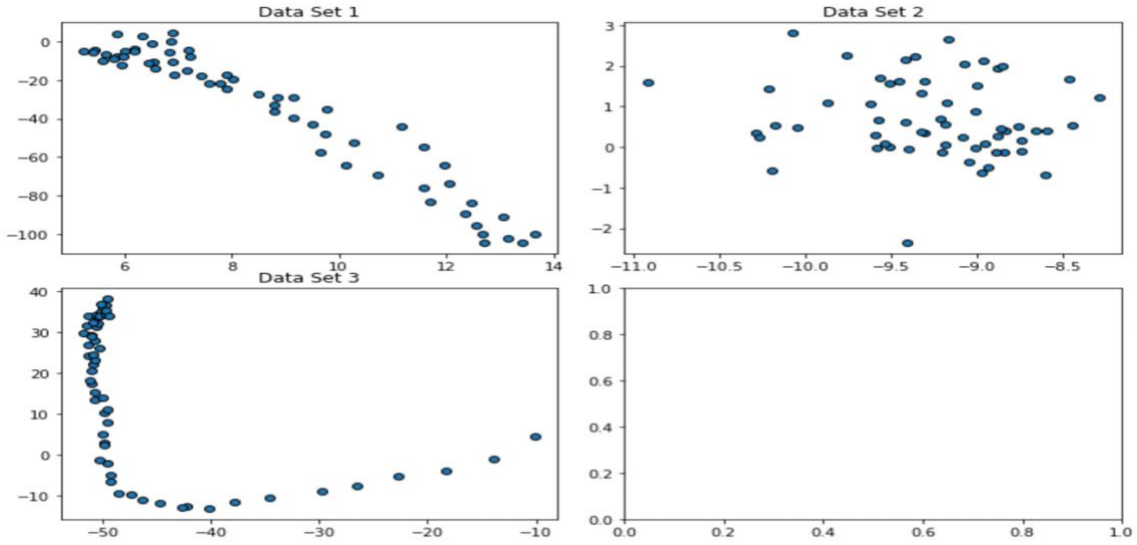
Data Description for Methanol 292 ppm set 3 and its Atmospheric conditions:

```
LDR (light sensor)     65393.271186
Moisture               65314.813559
Oxygen (O2) (%)           18.277119
Humidity (%)
Temp (deg C)
dtype: float64
```

click to expand output; double click to hide output

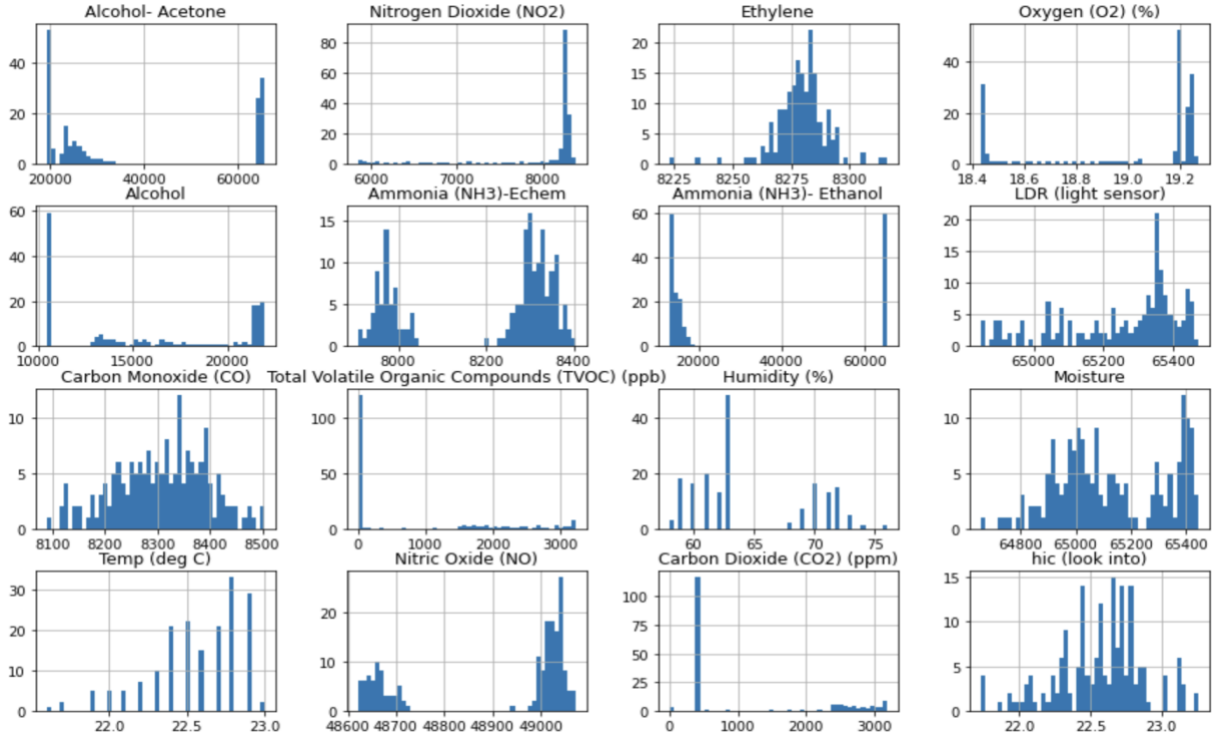


After applying PCA on Ethanol 292 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Ethanol 292 ppm PCA Components 1(x-axis) varying from 6 to 8.5 and Component 2(y-axis) varying from -80 to 3. With data set 3 we assume the data is not producing the results.



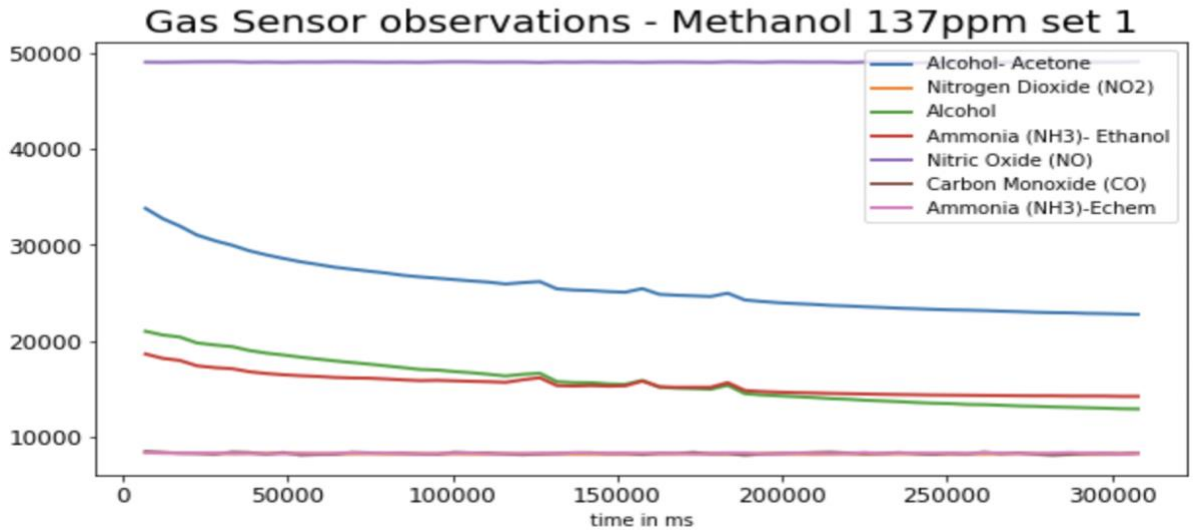
Methanol 137 ppm

We can observe that **Ethylene** is "**Normally Distributed**", **Ammonia, Moisture** are "**Bi model"**, Rest are Randomly or Multi Modally distributed.



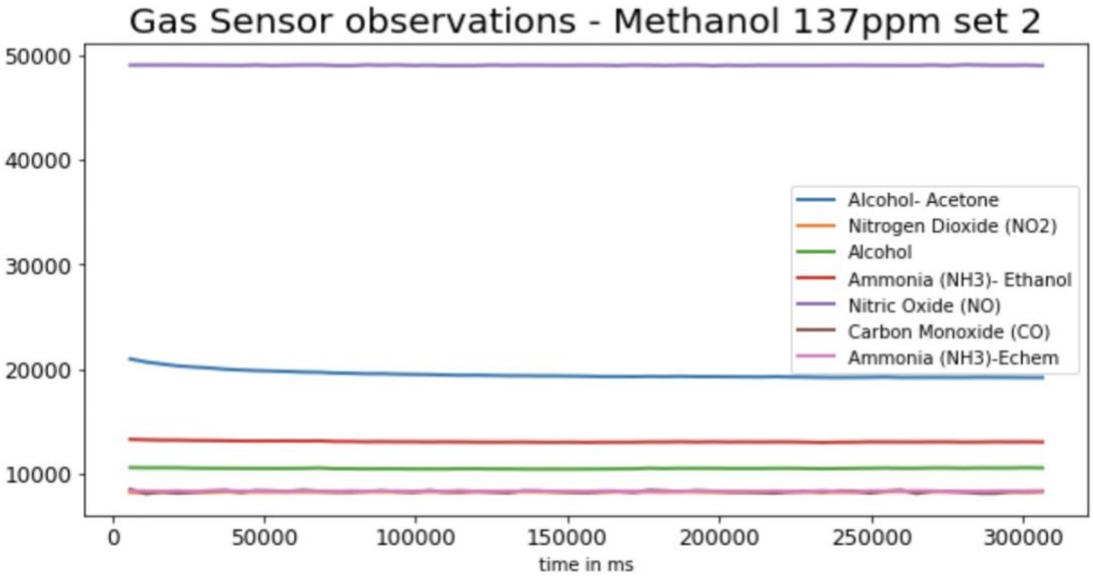
Data Description for Methanol 137 ppm set 1 and its Atmospheric conditions:

```
LDR (light sensor)     65222.237288
Moisture               65022.389831
Oxygen (O2) (%)           19.239492
Humidity (%)              60.016949
Temp (deg C)              22.777966
dtype: float64
```



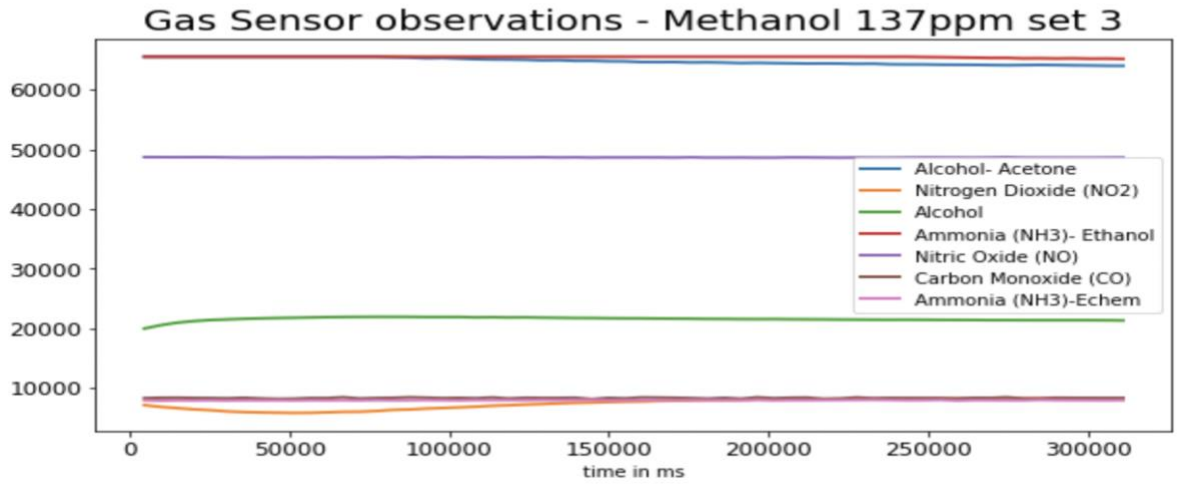
Data Description for Methanol 137 ppm set 2 and its Atmospheric conditions:

```
LDR (light sensor)     65208.474576
Moisture               64993.101695
Oxygen (O2) (%)           19.193390
Humidity (%)              62.796610
Temp (deg C)              22.506780
dtype: float64
```

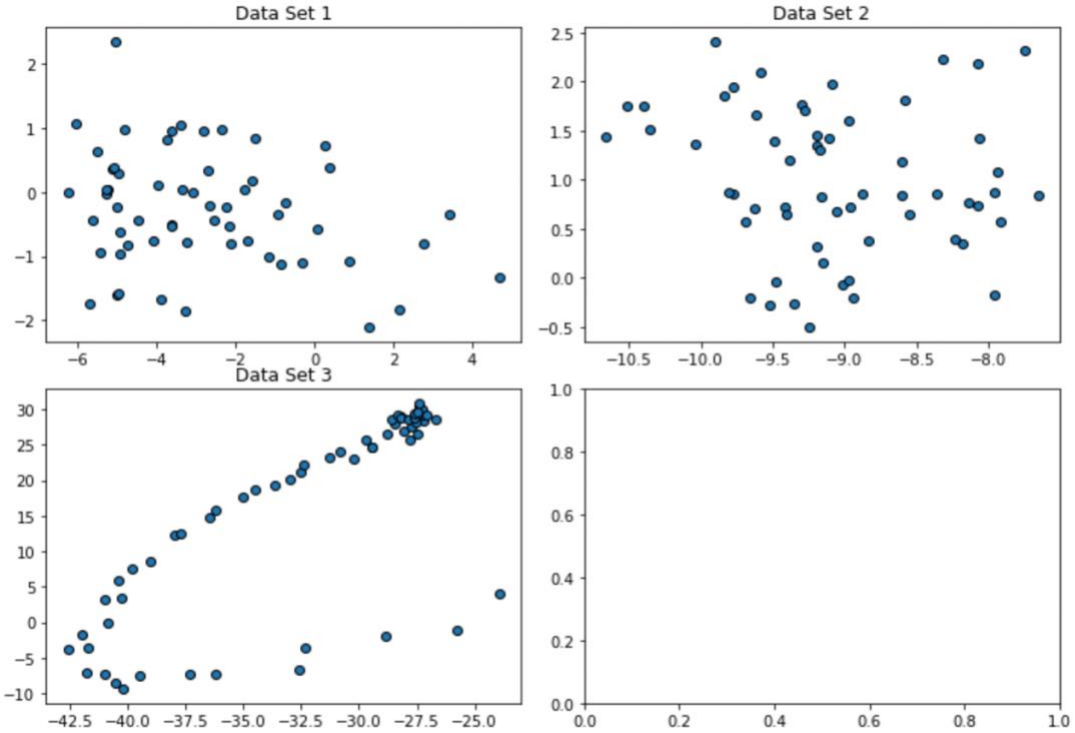


Data Description for Methanol 137 ppm set 3 and its Atmospheric conditions:

```
LDR (light sensor)     65307.233333
Moisture               65374.000000
Oxygen (O2) (%)           18.577333
Humidity (%)              70.950000
Temp (deg C)              22.420000
dtype: float64
```

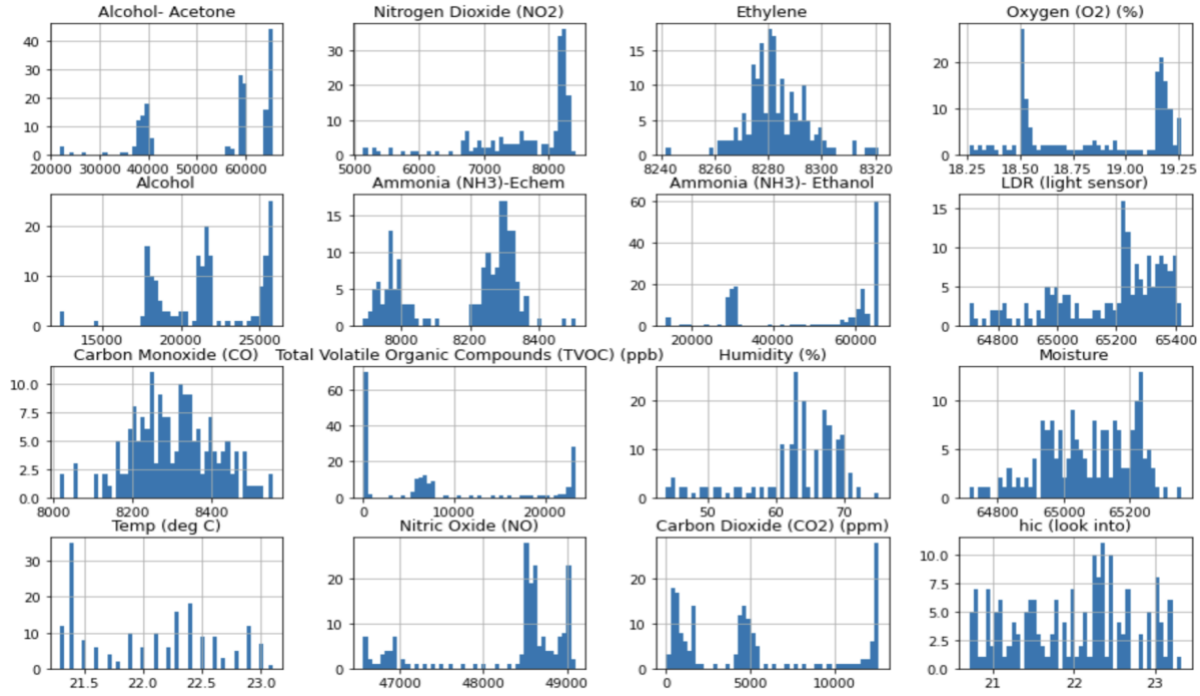


After applying PCA on Methanol 137 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Methanol 137 ppm PCA Components 1(x-axis) varying from -10.5 to 4 and Component 2(y-axis) varying from -2 to 3. With data set 3 we assume the data is not producing the results.



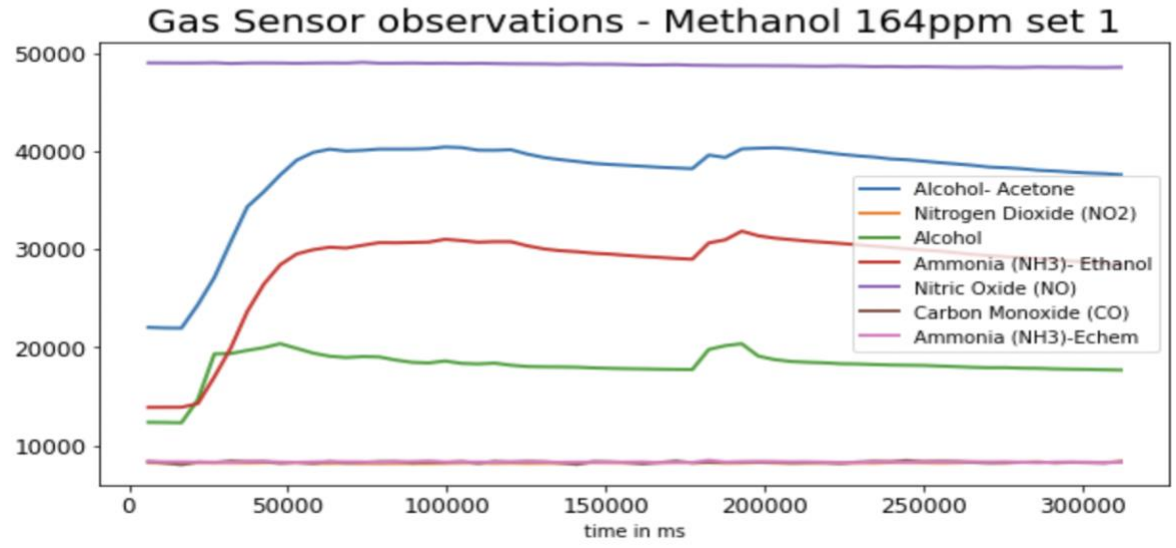
METHANOL 164ppm

From the below Methanol 164 ppm histograms we can observe that **Ethylen** is "**Normally Distributed"**. **Nitrogen Dioxide** is "**left skewed"**, **Oxygen, Ammonia-Echem, Carbon-Monoxide, Ammonia-Ethanol** are "Bi Model", Rest are randomly or Multi Modal Distribution.



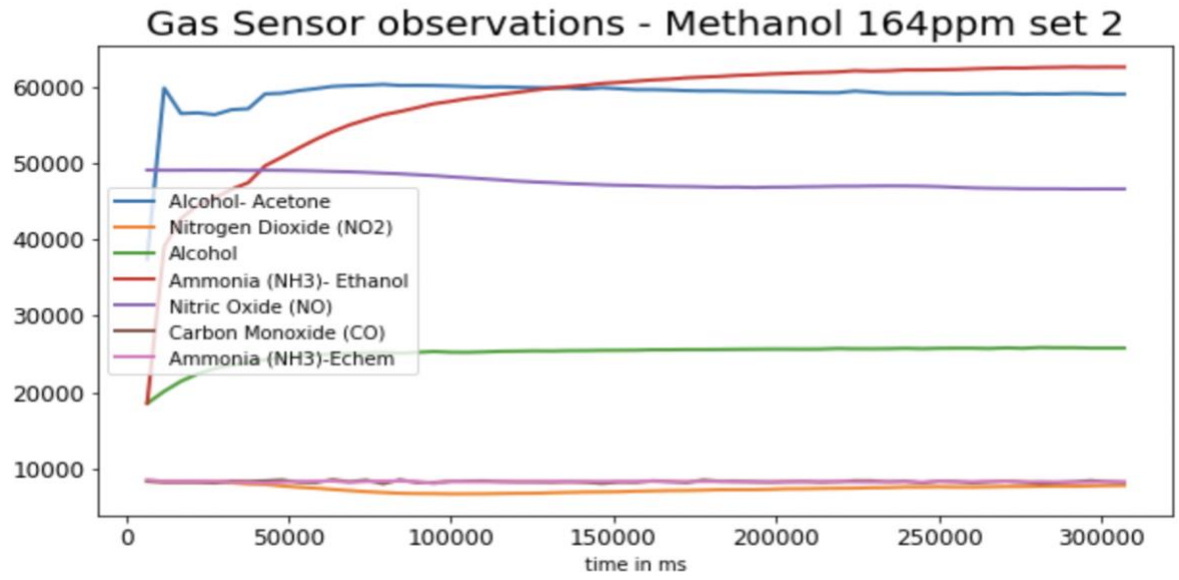
Data Description for Methanol 164 ppm set 1 and its Atmospheric conditions:

```
LDR (light sensor)     65211.883333
Moisture               64978.750000
Oxygen (O2) (%)           19.174500
Humidity (%)              62.850000
Temp (deg C)              22.158333
dtype: float64
```



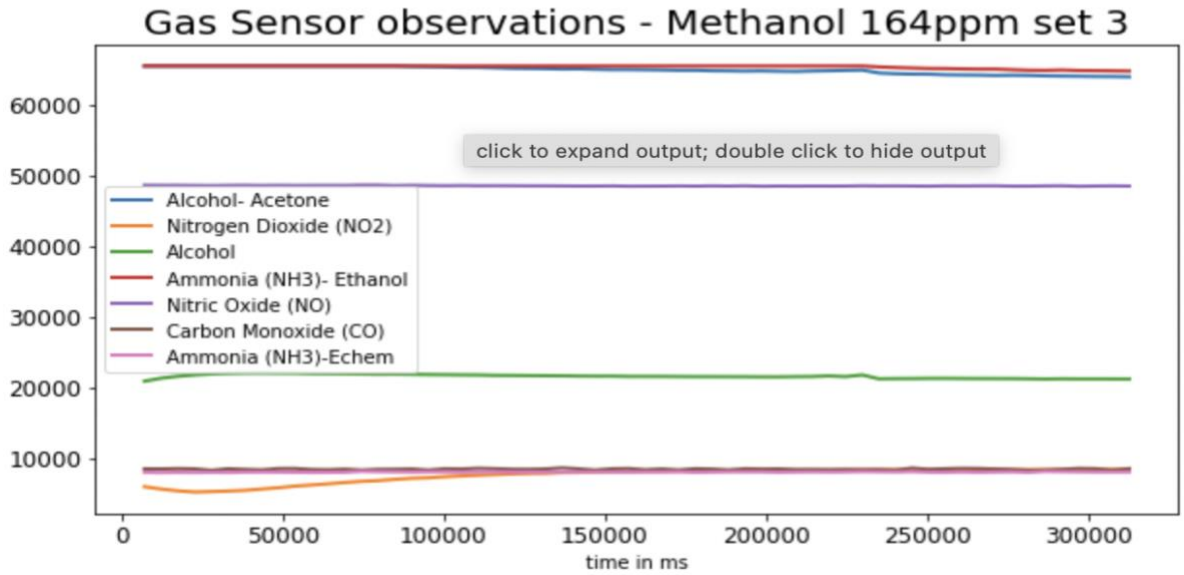
Data Description for Methanol 164 ppm set 2 and its Atmospheric conditions:

```
LDR (light sensor)    65086.694915
Moisture              65080.864407
Oxygen (O2) (%)          18.791695
Humidity (%)             58.169492
Temp (deg C)             21.408475
dtype: float64
```

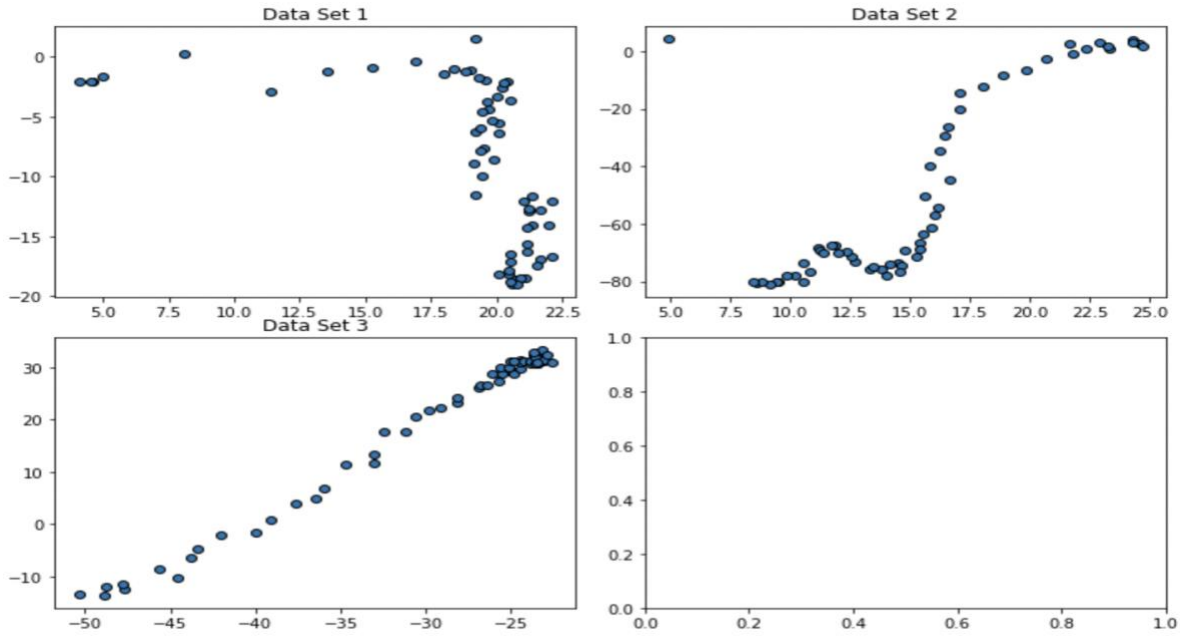


Data Description for Methanol 164 ppm set 3 and its Atmospheric conditions:

```
LDR (light sensor)     65246.933333
Moisture               65173.483333
Oxygen (O2) (%)           18.602333
Humidity (%)              68.283333
Temp (deg C)              22.601667
dtype: float64
```

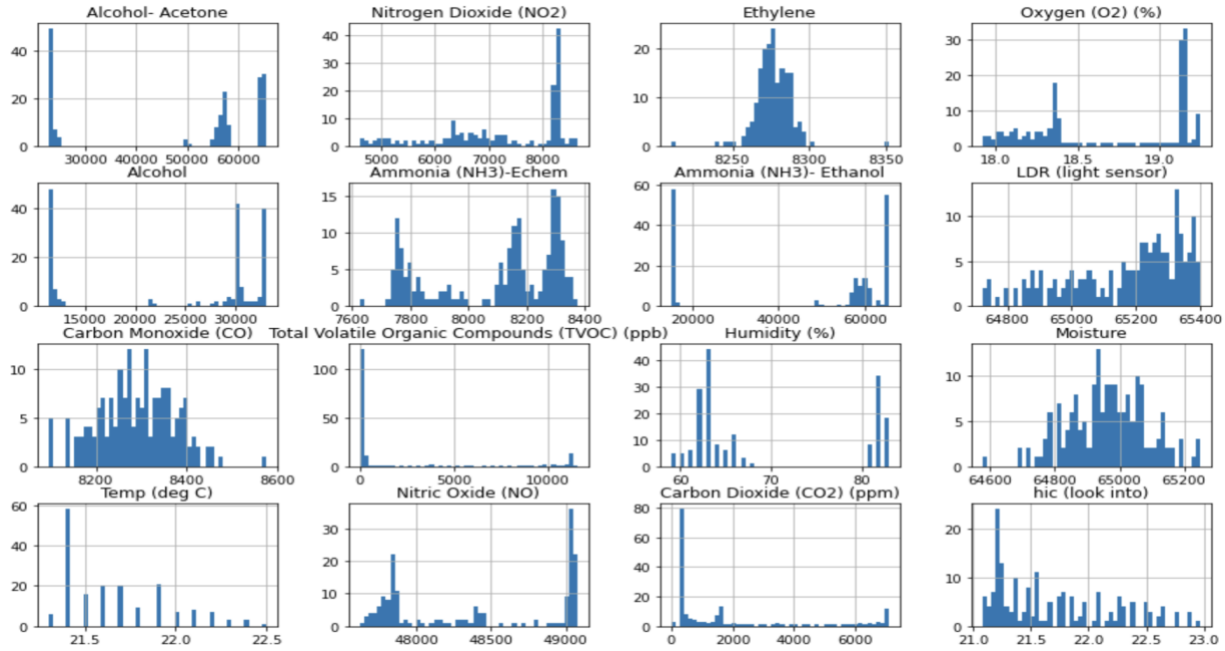


After applying PCA on Methanol 164 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Methanol 164 ppm PCA Components 1(x-axis) varying from 5 to 25 and Component 2(y-axis) varying from -70 to 0. With data set 3 we assume the data is not producing the results.



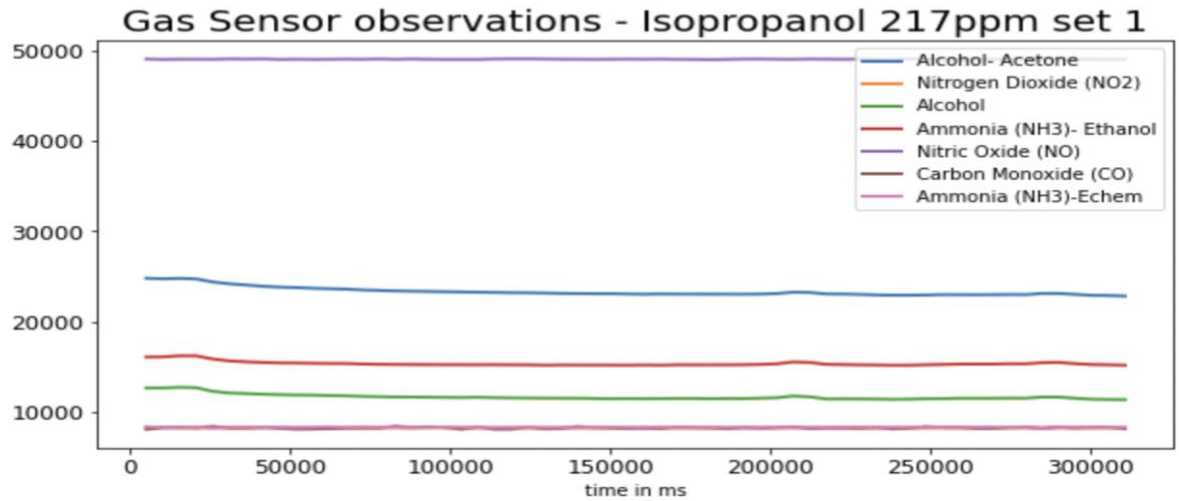
ISOPROPANOL 217 ppm

We can observe the distribution of Isopropanol 217 ppm in the below histograms and indicates **Ethylene is "Normally Distributed".** And **Nitrogen Dioxide, Alcohol-Acetone, Oxygen, Alocohol, Ammonia, Ammonia Ethanol, Carbon Monoxide, Humidity Moisture** are **"Multi Modal Distributed".**



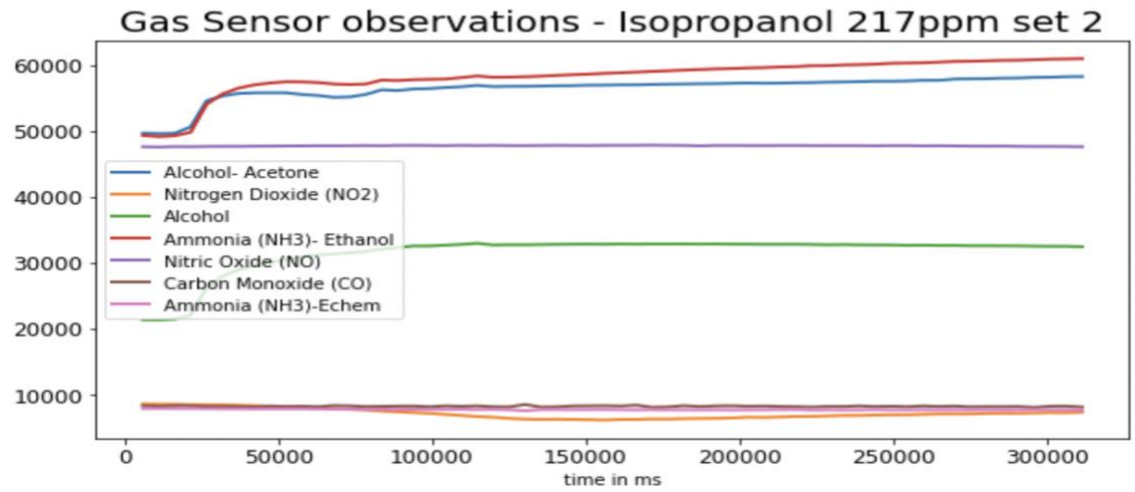
Data Description for Isopropanol 217ppm set 1 and its Atmospheric conditions:

```
LDR (light sensor)     65174.400000
Moisture               64966.500000
Oxygen (O2) (%)           19.145333
Humidity (%)              62.700000
Temp (deg C)              21.566667
dtype: float64
```



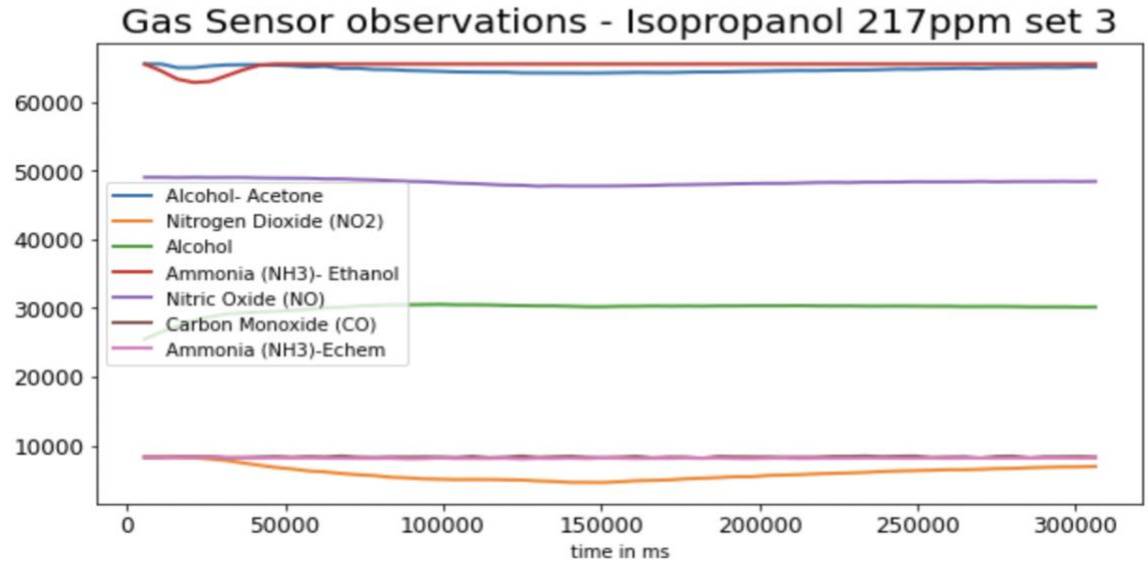
Data Description for Isopropanol 217ppm set 2 and its Atmospheric conditions:

```
LDR (light sensor)     65055.883333
Moisture               64985.900000
Oxygen (O2) (%)           18.240333
Humidity (%)              82.166667
Temp (deg C)              21.906667
dtype: float64
```

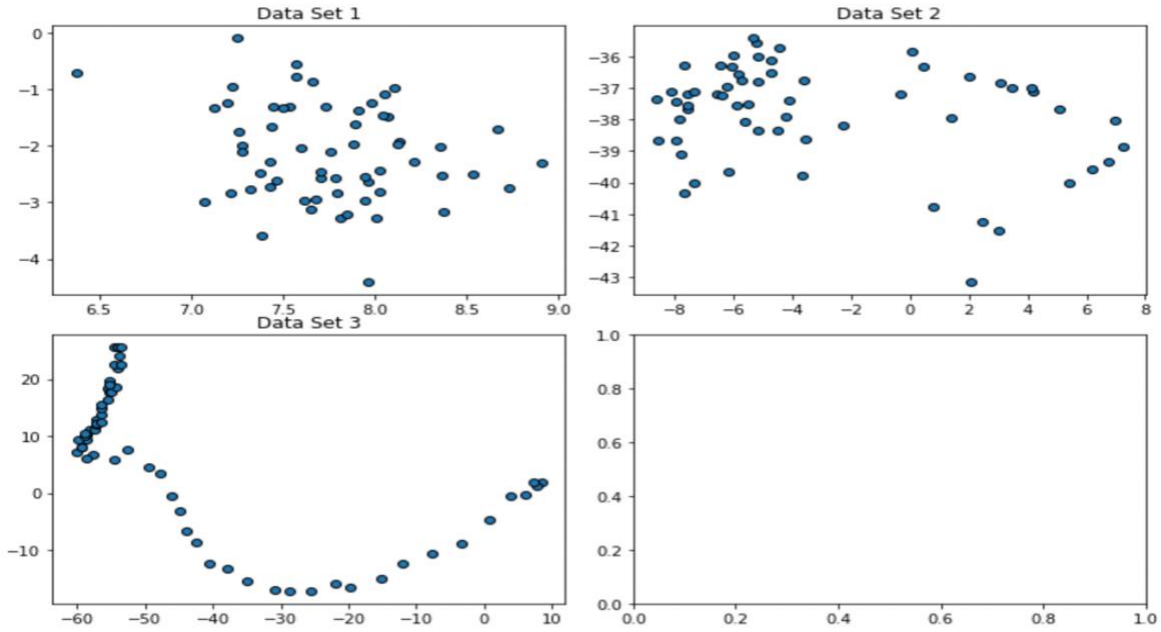


Data Description for Isopropanol 217ppm set 3 and its Atmospheric conditions:

```
LDR (light sensor)     65281.694915
Moisture               64938.932203
Oxygen (O2) (%)           18.659322
Humidity (%)              63.254237
Temp (deg C)              21.516949
dtype: float64
```

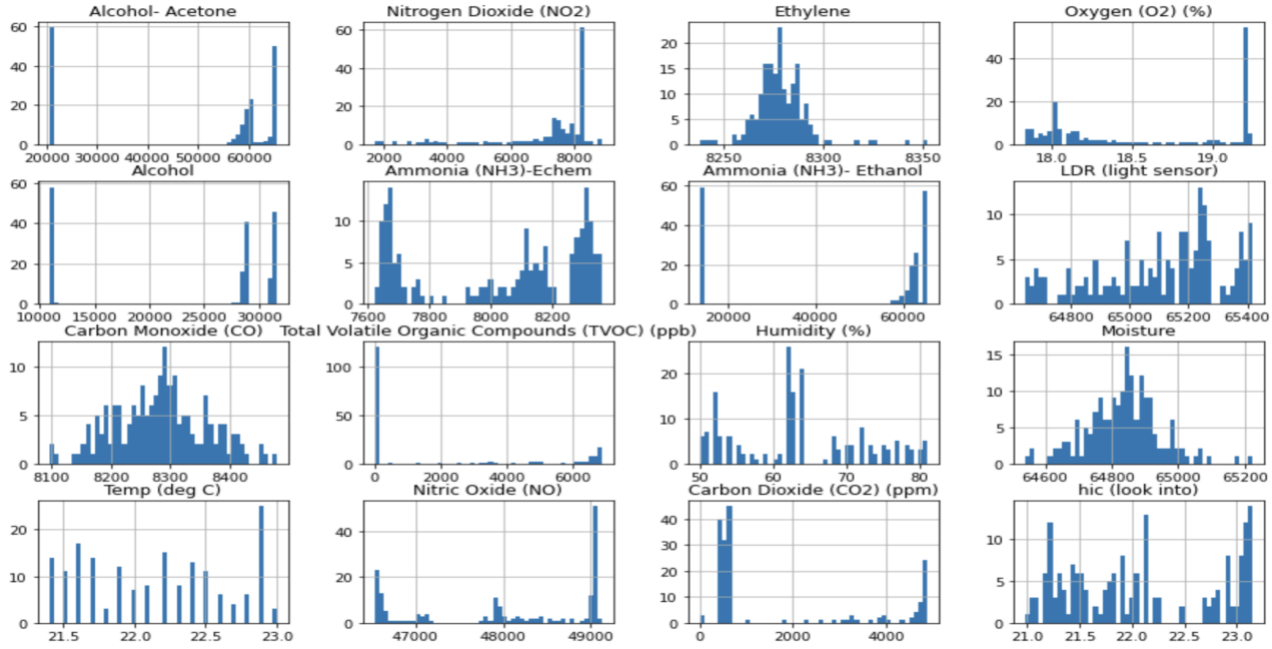


After applying PCA on Isopropanol 217 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Isopropanol 217 ppm PCA Components 1(x-axis) varying from -8 to 8 and Component 2(y-axis) varying from -37 to 0 . With data set 3 we assume the data is not producing the results.



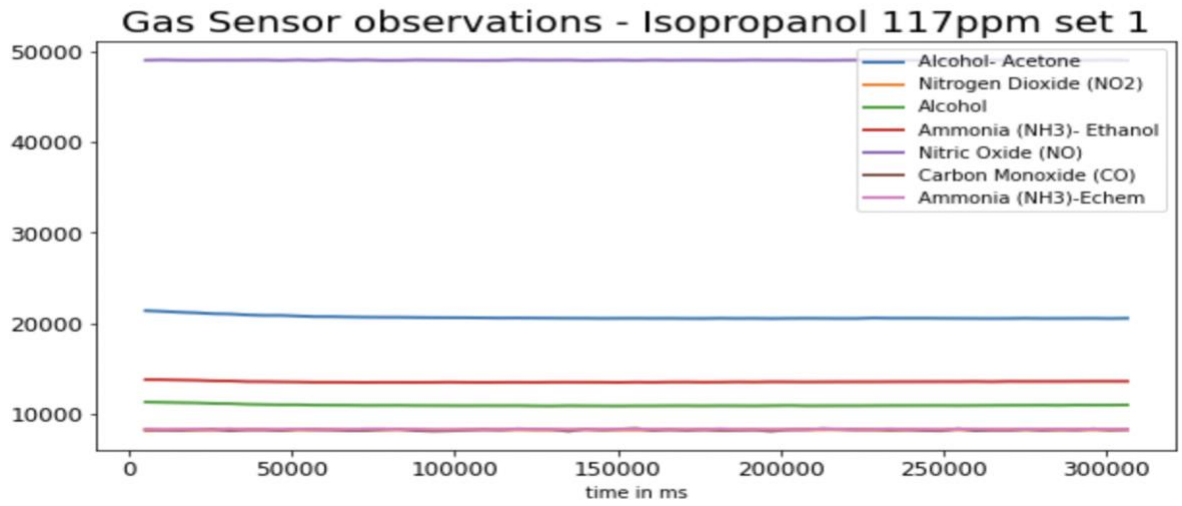
ISOPROPANOL 117 PPM

The distribution of ISOPROPANOL 117pm is Moisture , Ethylene, Carbon Monoxide being **"Normally Distributed"** and rest are distributed **multi modally or Randomly.**



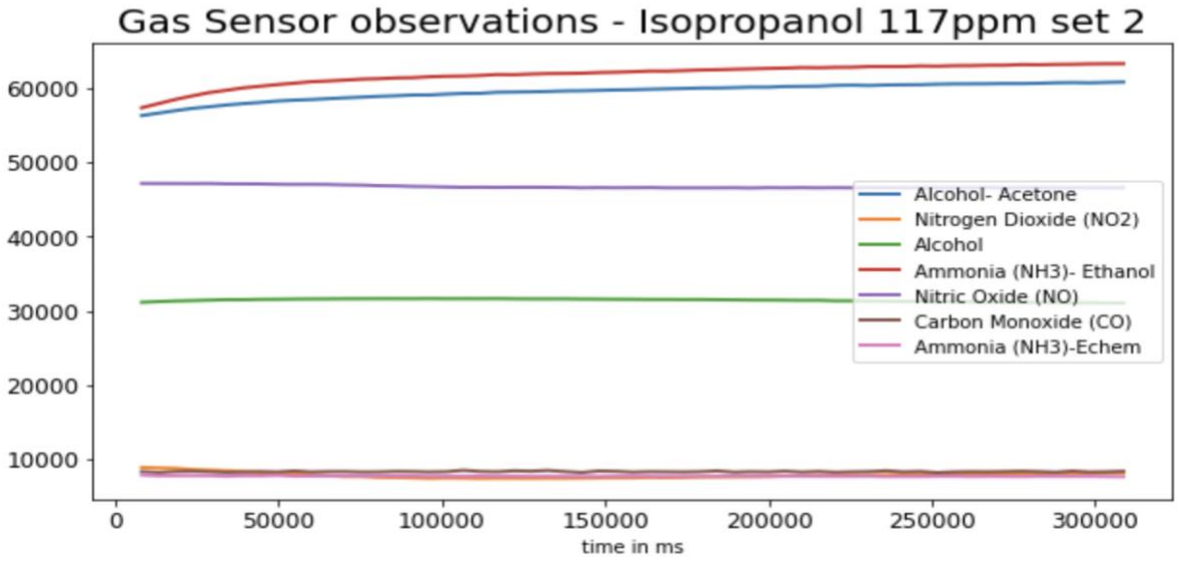
Data Description for Isopropanol 117ppm set 1 and its Atmospheric conditions:

```
LDR (light sensor)     65035.762712
Moisture               64833.254237
Oxygen (O2) (%)           19.212881
Humidity (%)              62.949153
Temp (deg C)              21.876271
dtype: float64
```



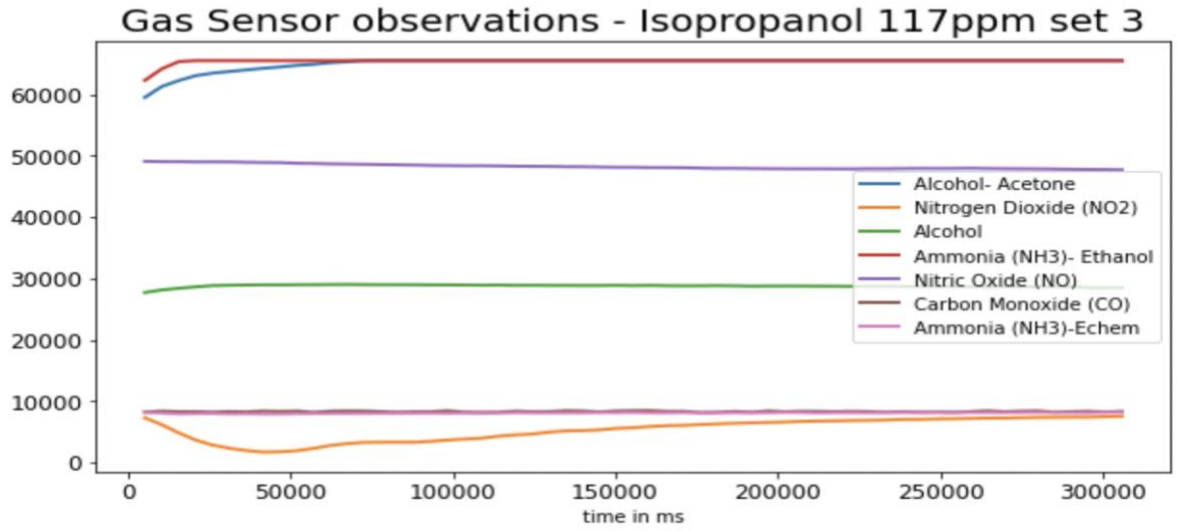
Data Description for Isopropanol 117ppm set 2 and its Atmospheric conditions:

```
LDR (light sensor)     65034.423729
Moisture               64799.627119
Oxygen (O2) (%)           17.960508
Humidity (%)              74.050847
Temp (deg C)              22.720339
dtype: float64
```

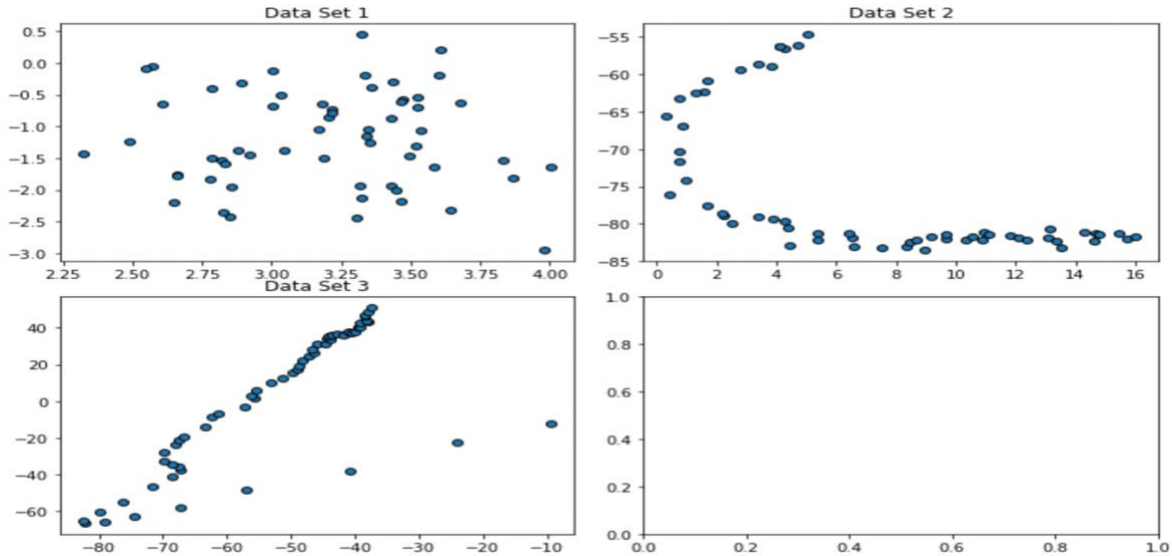


Data Description for Isopropanol 117ppm set 3 and its Atmospheric conditions:

```
LDR (light sensor)     65251.525424
Moisture               64874.322034
Oxygen (O2) (%)           18.471017
Humidity (%)              53.983051
Temp (deg C)              21.859322
dtype: float64
```

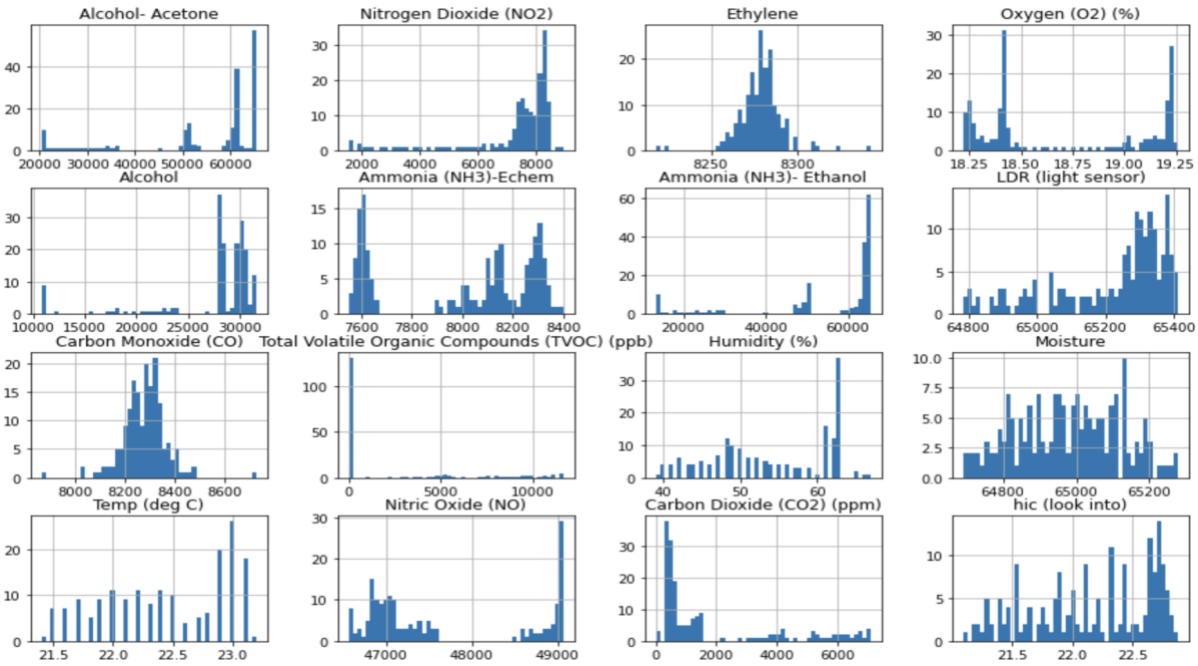


After applying PCA on Isopropanol 117 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Isopropanol 117 ppm PCA Components 1(x-axis) varying from 2 to 18 and Component 2(y-axis) varying from -70 to 1 . With data set 3 we assume the data is not producing the results.



ISOPROPANOL 143PPM

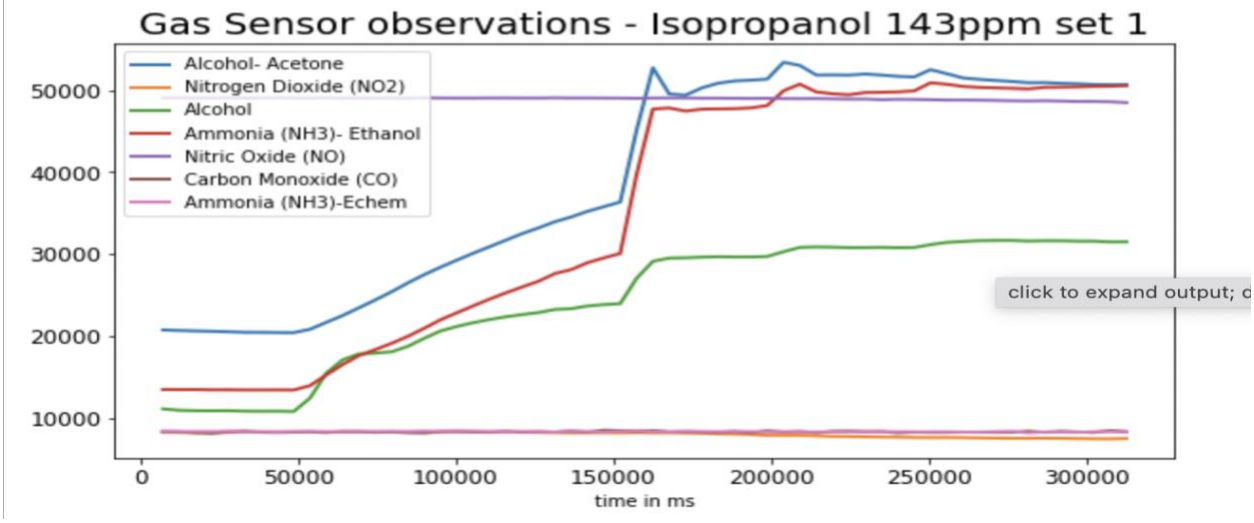
From the histograms we can see that **Carbon Monoxide** and **Ethylene** are "**Normally Distributed"**. All other elements are multi modally distributed or Randomly Distributed.



Description for Isopropanol 143ppm set 1 and its Atmospheric conditions:

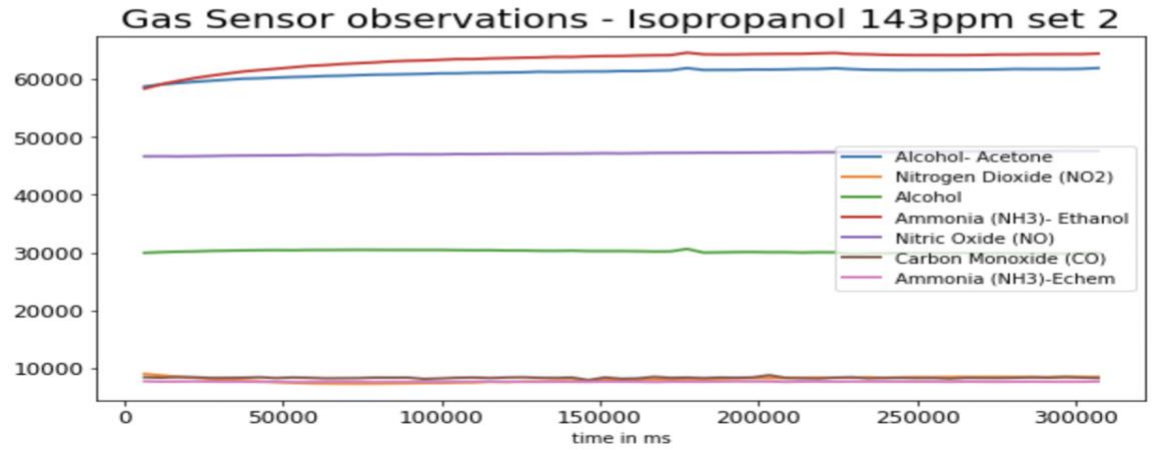
Data

```
LDR (light sensor)     65199.300000
Moisture               64862.333333
Oxygen (O2) (%)           19.196667
Humidity (%)              62.366667
Temp (deg C)              22.013333
dtype: float64
```



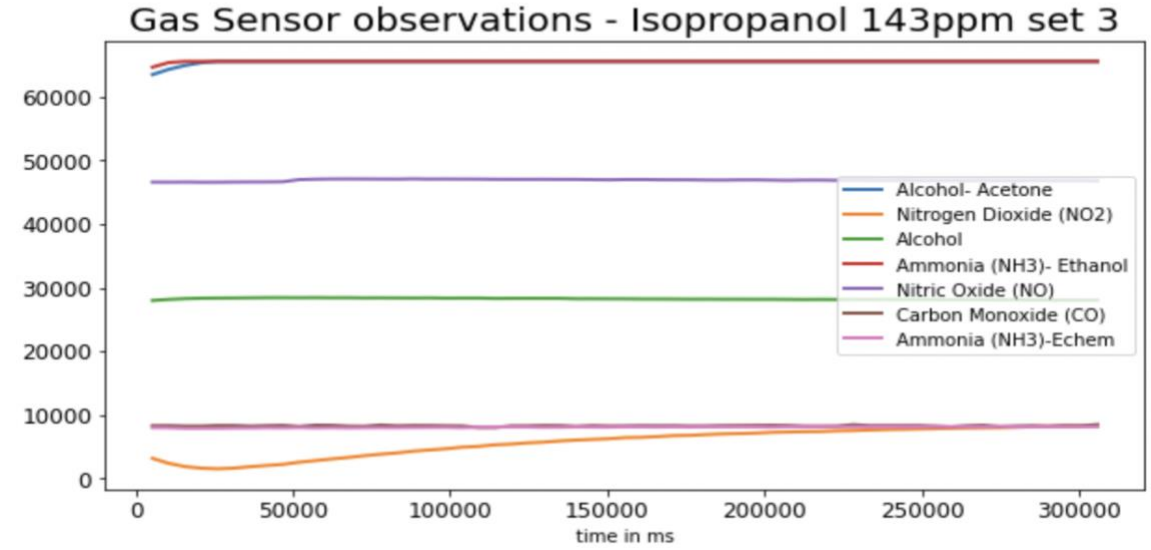
Data Description for Isopropanol 143 ppm set 2 and its Atmospheric conditions:

```
LDR (light sensor)     65153.881356
Moisture               65012.372881
Oxygen (O2) (%)           18.389153
Humidity (%)              53.525424
Temp (deg C)              22.957627
dtype: float64
```

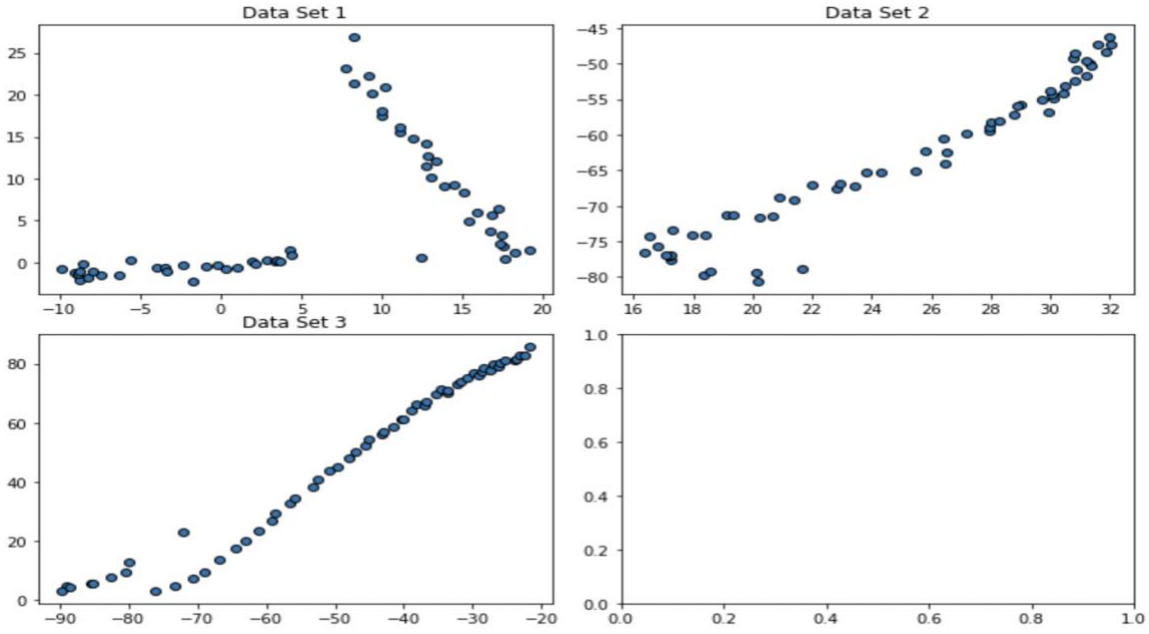


Data Description for Isopropanol 143 ppm set 3 and its Atmospheric conditions:

```
LDR (light sensor)     65297.983051
Moisture               65061.423729
Oxygen (O2) (%)           18.498136
Humidity (%)              46.711864
Temp (deg C)              22.369492
dtype: float64
```



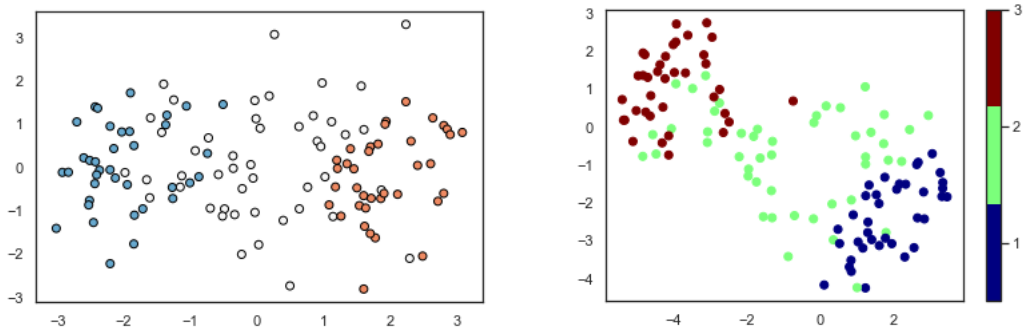
After applying PCA on Isopropanol 143 ppm, by comparing the data with baseline we can observe from data set 1 and data set 2 the range of Isopropanol 143 ppm PCA Components 1(x-axis) varying from -10 to 32 and Component 2(y-axis) varying from -65 to 20. With data set 3 we assume the data is not producing the results.



Ethanol PCA/tSNE

We have first identified each proportion of the Ethanol as 3 targets. We used the below code to assign data frames to ethanol 123 ppm as Target 1, ethanol 161 ppm as Target 2, and ethanol 200 ppm as Target 3.

We have color coded the below scatter plot with three different colors i.e., blue, white, and brown, to differentiate the data after performing the PCA on the data and then plotting the scatter plot for the data explains the relation between each target variables. We can observe from the below plot that the data is scattered randomly and explains there is no correlation between the variables.

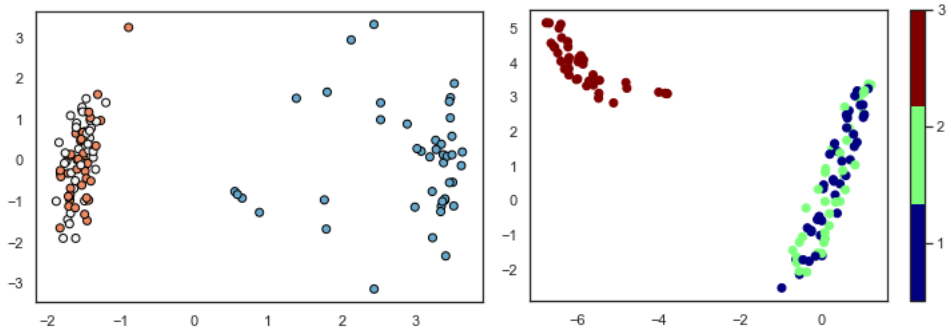


TSNE for Ethanol:

We color coded Ethanol 123 ppm as blue, Ethanol 161 ppm as green and Ethanol 200 ppm as brown in the TSNE Algorithm.

Methanol PCA/tSNE

We have first identified each proportion of the Methanol as 3 targets. We used the below code to assign data frames to Methanol 292 ppm as Target 1, Methanol 137ppm Target 2, and Methanol 164 ppm as Target 3.

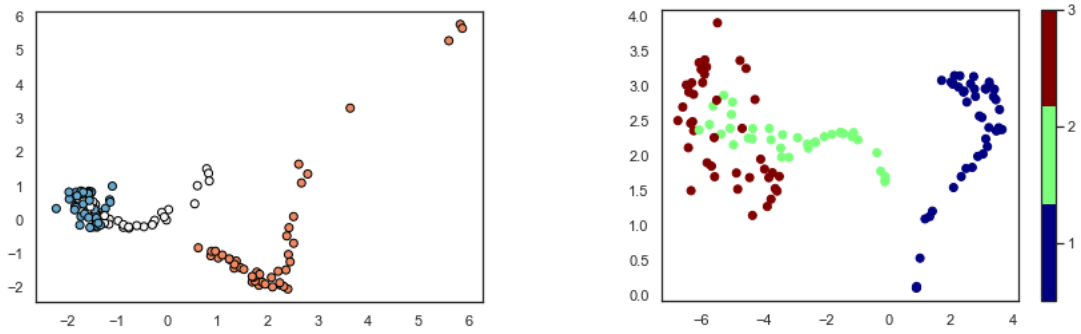


TSNE for Methanol:

We color coded Methanol 292 ppm as blue, Methanol 137 ppm as green and Methanol 164 ppm as brown in the TSNE Algorithm.

Isopropanol PCA/tSNE

We have first identified each proportion of the Isopropanol as 3 targets. We used the below code to assign data frames to Isopropanol 217 ppm as Target 1, Isopropanol 117 ppm as Target 2, Isopropanol 143 ppm as Target 3.



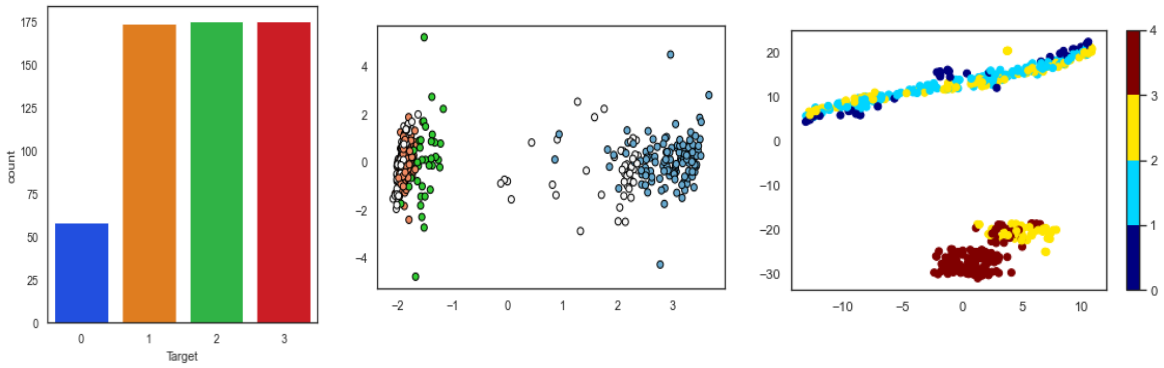
TSNE for Isopropanol:

On the Right chart above we color coded Isopropanol 217 ppm as blue, Ethanol 117 ppm as green and Ethanol 143S ppm as brown in the TSNE Algorithm.

Data with 4 elements as Category

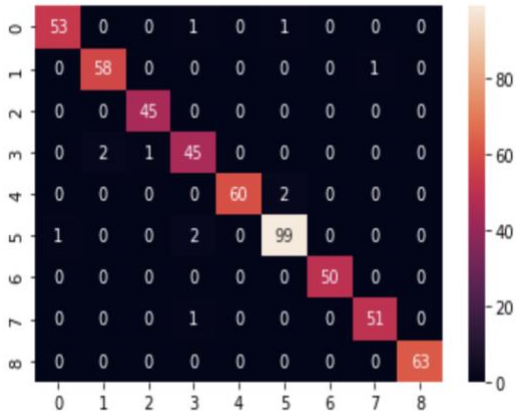
3 component analysis using PCA algorithm and tSNE Algorithm.
We have joined each proportion of the chemical components of ethanol, Methanol, isopropanol as 3 targets. We used the below code to assign data frames to Ethanol as Target 1, Methanol as Target 2, and Isopropanol as Target 3 and Baseline as Target 0.

The below image demonstrates the distribution of data in 4 categories. We have first identified each proportion of 3 targets. We used the below code to assign data frames to Target 0 in blue, Target 1 in light blue, Target 2 in yellow, and Target 3 in brown colors in Tsne.



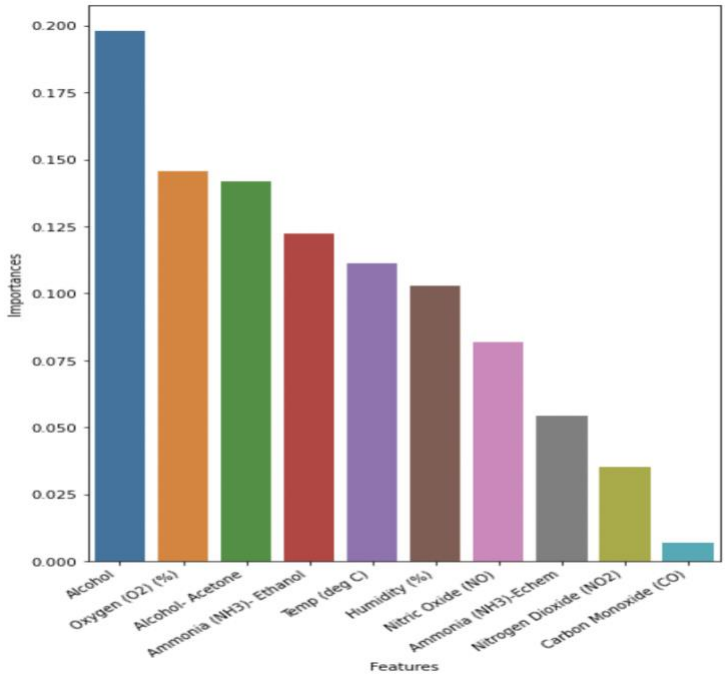
Multiclass classification for Baseline, Ethanol, Methanol, and Isopropanol with Multiple ppm levels

We have implemented Multiclass Classification on the Data combining data sets 1, 2 and 3 for all the components Baseline, Ethanol and Isopropanol and Implemented Random Forest Algorithm on the data and can observe from the below chart F1 score, precision and recall with 98% Accuracy indicating the model performance as good. As from the below Heat Map for Confusion Matrix of the Model we can observe all the true positives for all the 10 categorical variables and shows positive correlation ship between many indicating the strong correlation between multiple components.



	precision	recall	f1-score	support
0	1.00	0.98	0.99	52
1	0.96	1.00	0.98	54
2	1.00	1.00	1.00	49
3	0.98	0.94	0.96	50
4	1.00	1.00	1.00	64
5	0.96	0.97	0.97	105
7	1.00	0.98	0.99	57
8	0.98	0.98	0.98	50
9	0.96	0.98	0.97	55
accuracy			0.98	536
macro avg	0.98	0.98	0.98	536
weighted avg	0.98	0.98	0.98	536

We can observe from the below chart that these columns are playing major role in identifying the patterns and identifying the gas passed eventually. And it clearly indicates Alcohol playing major Role is Identifying the gas followed by Oxygen and Alcohol-Acetone.



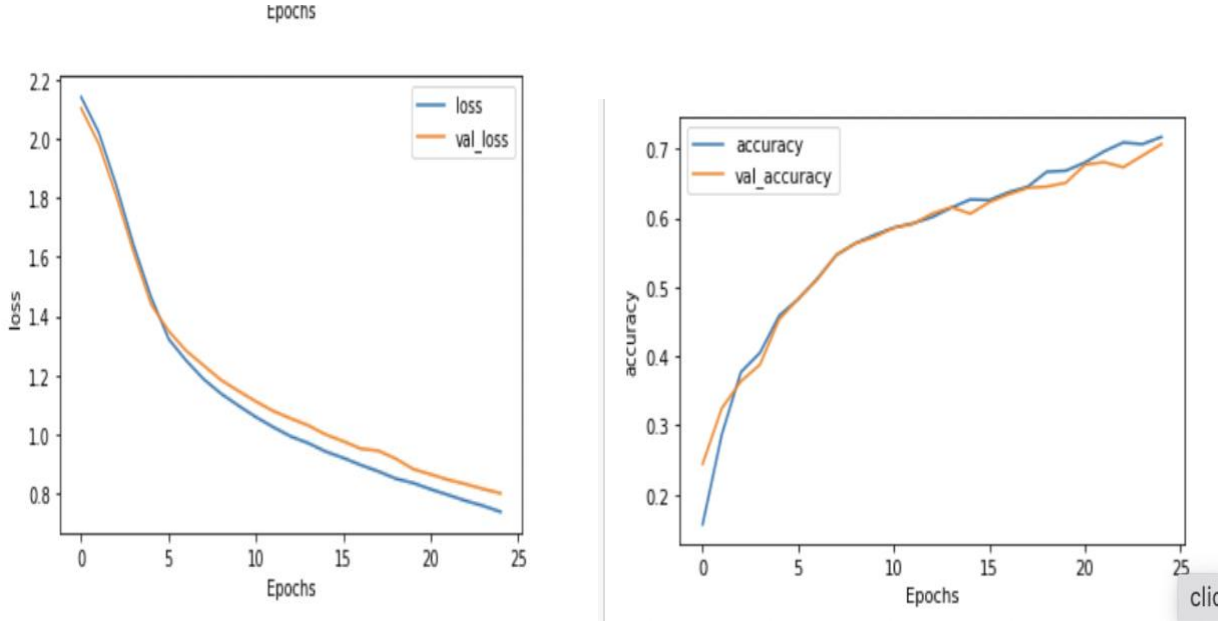
From this we can say that Neural Network Model have been more reliable technique to identify the gas and we have performed it on the whole datasets, we tried to improve the Accuracy of the gases identification from 40 to 70% by adding multiple layers i.e. 4 layers with Activation as ReLU as SoftMax to improve the performance of the Model, And also ran the data in 25 epochs to improve the performance of the Neural Network model and achieved 70% Accuracy.

```
model = tf.keras.Sequential([
    tf.keras.layers.Dense(9, input_dim = in_dim, activation = 'relu'),
    tf.keras.layers.Dense(15, activation = 'relu'),
    tf.keras.layers.Dense(5, activation = 'relu'),
    tf.keras.layers.Dense(9, activation = 'softmax')
])
##compile the model
model.compile(loss = 'categorical_crossentropy', optimizer = 'adam', metrics = ['accuracy'])
```

Result:

```
Epoch 25/25
250/250 - 2s - loss: 0.7348 - accuracy: 0.7174 - val_loss: 0.7978
val_accuracy: 0.7071
```

The below graphs indicate the Accuracy and Loss functions for the Models:
Accuracy being measure of how accurate the model's prediction is compared to true data.
loss function is a function that compares the target and predicted output values; measures how well the neural network models the training data. When training, we aim to minimize this loss between the predicted and target outputs. From the below graphs we can see that blue color is train data and orange color being test data. We can observe that there is less Bias in the Accuracy and loss in between test and train data.



Conclusion and Future Work:

We attempted to use data science tools to measure and detect Methanol, Ethanol, and Isopropanol gases in three concentrations in this USF-built electronic nose. Baseline data, or data recorded with

only normal air passing by, was collected. This has been used to compare other gases to. We discovered that a few sensor readings had a high variance, particularly alcohol-acetone and alcohol-ammonia (NH3)-ethanol. This could be caused by a change in atmospheric conditions such as humidity or temperature. We're also trying to figure out if this is due to a problem with the sensors' ability to record data or if there's any residual gas in the chamber. More data collection and analysis can help to resolve this ambiguity. After performing a principal component analysis on each dataset of one gas with a specific concentration and transforming it with respect to its respective baseline data set, we can see that a few gases show obvious clustering with some distributed points. This could be due to the above-mentioned reasons, and more data can help to resolve this. In addition, we divided the data into 9 categories based on the baseline (ethanol 123ppm, ethanol 200ppm, ethanol 161ppm, methanol 292ppm, methanol 137ppm, methanol 164ppm, isopropanol 217ppm, isopropanol 117ppm, isopropanol 143ppm). To orchestrate machine learning models, an ensemble model (Random Forest classifier) and a FNN (Feedforward Neural network) algorithm were used. After combining sensor and atmospheric data, normalize it with a standard scalar. The results have been very encouraging. Additional investigation is required to determine the validity of the data, and the use of advanced statistical models can result in an electronic nose that is ready for the market.

Limitations:

- Consumption of more time for Data Collection and small data sets to analyze the trends in data.
- Data Collection was performed for baseline and other chemical gases on different days which could impact the quality of data.
- Quality or performance of Sensors could impact the ability to get the right data as we could observe the deviation in Dataset 3 from Dataset 1, Dataset 2 indicating the possibility of sensors functioning or other factors.
- Due to limited data, we were not able to identify the reasons for the deviation in the new data collected.

References:

Application of electronic nose for industrial odors and gaseous emissions measurement and monitoring – An overview
Link : https://www.sciencedirect.com/science/article/pii/S0039914015300904

Performance of the Levenberg–Marquardt neural network training method in electronic nose applications
Link: https://www.sciencedirect.com/science/article/pii/S0925400505000961

Deep learning-based gas identification and quantification with auto-tuning of hyper-parameters
Link: https://link.springer.com/article/10%2E1007/s00500-021-06222-1

Quality grade classification of China commercial moxa floss using electronic nose
Link: https://journals.lww.com/md-journal/Fulltext/2020/08140/Quality_grade_classification_of_China_commercial.38.aspx

Integrating a Low-Cost Electronic Nose and Machine Learning Modelling to Assess Coffee Aroma Profile and Intensity
Link: https://www.proquest.com/docview/2508562107?accountid=14745&forcedol=true&pq-origsite=primo

Methane detection to 1 ppm using machine learning analysis of atmospheric pressure plasma optical emission spectra

Link: https://iopscience.iop.org/article/10.1088/1361-6463/ac5770

Electronic Nose Feature Extraction Methods: A Review
Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4701255/

An adaptive learning method for the fusion information of electronic nose and hyperspectral system to identify the egg quality
Link: https://www-sciencedirect-com.ezproxy.lib.usf.edu/science/article/pii/S0924424722004599

Analysis of the changes of volatile flavor compounds in a traditional Chinese shrimp paste during fermentation based on electronic nose, SPME-GC-MS and HS-GC-IMS
Link: https://www.sciencedirect.com/science/article/pii/S2213453022001422?via%3Dihub

Discrimination of different oil types and adulterated safflower seed oil based on electronic nose combined with gas chromatography-ion mobility spectrometry
Link: https://www-sciencedirect-com.ezproxy.lib.usf.edu/science/article/pii/S0889157522004227

Recent trends of multi-source and non-destructive information for quality authentication of herbs and spices
Link: https://www.sciencedirect.com/science/article/abs/pii/S030881462201901X

https://www.cacgas.com.au/gas-quality-matters
https://www.uc.edu/content/dam/refresh/cont-ed-62/olli/new-tech-sept22.pdf
https://www-sciencedirect-com.ezproxy.lib.usf.edu/science/article/pii/S0924424722004599
https://towardsdatascience.com/principal-component-analysis-pca-explained-visually-with-zero-math-1cbf392b9e7d
https://iopscience.iop.org/article/10.1088/1361-6463/ac5770

Appendix:

```
import numpy as np
import pandas as pd
import xlrd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.manifold import TSNE
import scipy
from scipy import fft
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
df_base_line = pd.read_excel('DATALOG_Real.xlsm', sheet_name='Baseline',
header = 2)
df_base_line.columns
gas_sensors = {'Ammonia (NH3)-Echem', 'Nitrogen Dioxide (NO2)','Carbon
Monoxide (CO)','Alcohol- Acetone',
            'Alcohol','Ammonia (NH3)- Ethanol','Nitric Oxide (NO)'}
cols = {'Ammonia (NH3)-Echem', 'Nitrogen Dioxide (NO2)', 'Ethylene',
      'Carbon Monoxide (CO)', 'Alcohol- Acetone', 'Alcohol',
      'Ammonia (NH3)- Ethanol',  'LDR (light sensor)', 'Moisture',
      'Nitric Oxide (NO)', 'Carbon Dioxide (CO2) (ppm)',
      'Total Volatile Organic Compounds (TVOC) (ppb)', 'Oxygen (O2) (%)',
      'Humidity (%)', 'Temp (deg C)', 'hic (look into)', }
```

Baseline Data Description

This Code has been used for all elements i.e., Baseline, Ethanol, Methanol for all 3 proportions.

```
df_base_line[cols].hist(bins = 50, figsize = (15, 10))
df_baseline_set1 = df_base_line[:59]
df_baseline_set1 = df_baseline_set1.set_index('ms')
ax = df_baseline_set1[gas_sensors].plot(linewidth=2, fontsize=12,figsize=(10, 5));
# Additional customizations
ax.set_xlabel('time in ms');
ax.legend(fontsize=10);
ax.set_title('Gas Sensor observations - Baseline set 1', fontsize = 20)
```

```
df_baseline_set1[['LDR (light sensor)','Moisture','Oxygen (O2) (%)','Humidity (%)', 'Temp (deg C)']].mean()
```

PCA analysis

```
feature_columns = ['Ammonia (NH3)-Echem', 'Nitrogen Dioxide (NO2)','Carbon Monoxide (CO)','Alcohol- Acetone',
                   'Alcohol','Ammonia (NH3)- Ethanol','Nitric Oxide (NO)','Oxygen (O2) (%)',
                   'Humidity (%)', 'Temp (deg C)']
scalar = StandardScaler()
pca = PCA(n_components=2)
set_sc = scalar.fit_transform(df_base_line[feature_columns])
base_set_pca = pca.fit_transform(set_sc)
```

Fit Scaling and PCA for set 1 data

```
scalar1 = StandardScaler()
pca1 = PCA(n_components=2)
set1_sc = scalar1.fit_transform(df_baseline_set1[feature_columns])
base_set1_pca = pca1.fit_transform(set1_sc)
```

Fit Scaling and PCA for set 2 data

```
scalar2 = StandardScaler()
pca2 = PCA(n_components=2)
set2_sc = scalar2.fit_transform(df_baseline_set2[feature_columns])
base_set2_pca = pca2.fit_transform(set2_sc)
```

Fit Scaling and PCA for set 3 data

```
scalar3 = StandardScaler()
pca3 = PCA(n_components=2)
set3_sc = scalar3.fit_transform(df_baseline_set3[feature_columns])
base_set3_pca = pca3.fit_transform(set3_sc)
```

Ethenol 123ppm PCA

This Code has been used for all elements i.e., Baseline, Ethanol, Methanol for all 3 proportions.

```
#define subplots
fig, ax = plt.subplots(2, 2, figsize=(10,7))
fig.tight_layout()
```

```
df_ethanol_123_set1_x = scalar1.transform(df_ethanol_123_set1[feature_columns])
df_ethanol_123_set1_x = pca1.transform(df_ethanol_123_set1_x)
kwarg_params = {'linewidth': 1, 'edgecolor': 'black'}
ax[0, 0].scatter(df_ethanol_123_set1_x[:, 0], df_ethanol_123_set1_x[:, 1], **kwarg_params)
ax[0, 0].title.set_text('Data Set 1')

df_ethanol_123_set2_x = scalar2.transform(df_ethanol_123_set2[feature_columns])
df_ethanol_123_set2_x = pca2.transform(df_ethanol_123_set2_x)
kwarg_params = {'linewidth': 1, 'edgecolor': 'black'}
ax[0, 1].scatter(df_ethanol_123_set2_x[:, 0], df_ethanol_123_set2_x[:, 1], **kwarg_params)
ax[0, 1].title.set_text('Data Set 2')

df_ethanol_123_set3_x = scalar3.transform(df_ethanol_123_set3[feature_columns])
df_ethanol_123_set3_x = pca3.transform(df_ethanol_123_set3_x)
kwarg_params = {'linewidth': 1, 'edgecolor': 'black'}
ax[1, 0].scatter(df_ethanol_123_set3_x[:, 0], df_ethanol_123_set3_x[:, 1], **kwarg_params)
ax[1, 0].title.set_text('Data Set 3')
```

DCT

Ethanol 200ppm

This Code has been used for all elements i.e., Baseline, Ethanol, Methanol for all 3 proportions.

```
plt.figure(figsize = (20,20))
plt.subplot(3, 1, 1)
df_ethanol_200_set1_sc = scalar1.transform(df_ethanol_200_set1[feature_columns])
df_ethanol_200_set1_dct = fft.dct(df_ethanol_200_set1_sc)
kwarg_params = {'linewidth': 1, 'edgecolor': 'black'}
plt.plot(pd.DataFrame(df_ethanol_200_set1_dct, columns = feature_columns),linewidth=2)
plt.title('Data Set 1')

plt.subplot(3, 1, 2)
df_ethanol_200_set2_sc = scalar2.transform(df_ethanol_200_set2[feature_columns])
df_ethanol_200_set2_dct = fft.dct(df_ethanol_200_set2_sc)
kwarg_params = {'linewidth': 1, 'edgecolor': 'black'}
plt.plot(pd.DataFrame(df_ethanol_200_set2_dct, columns = feature_columns),linewidth=2)
plt.title('Data Set 2')

plt.subplot(3, 1, 3)
df_ethanol_200_set3_sc = scalar3.transform(df_ethanol_200_set3[feature_columns])
df_ethanol_200_set3_dct = fft.dct(df_ethanol_200_set3_sc)
kwarg_params = {'linewidth': 1, 'edgecolor': 'black'}
```

```
plt.plot(pd.DataFrame(df_ethanol_200_set3_dct, columns = feature_columns),linewidth=2)
plt.title('Data Set 3')

plt.show()
```

Multiclass classification for Baseline , Ethanol, Methanol and Isopropanol with Multiple ppm levels

```
df_base_line['Target'] = 0
df_ethanol_123['Target'] = 1
df_ethanol_200['Target'] = 2
df_ethanol_161['Target'] = 3
df_methanol_292['Target'] = 4
df_methanol_137['Target'] = 5
df_methanol_164['Target'] = 5
df_isopropanol_217['Target'] = 7
df_isopropanol_117['Target'] = 8
df_isopropanol_143['Target'] = 9
target_column = ['Target']
df_final =
pd.concat([df_base_line,df_ethanol_123,df_ethanol_200,df_ethanol_161,df_methanol_292,
df_methanol_137,
df_methanol_164,df_isopropanol_217,df_isopropanol_117,df_isopropanol_143])
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test =
train_test_split(df_final[feature_columns],df_final[target_column], test_size=0.3)
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
X_train_sc = scalar.fit_transform(X_train)
X_test_sc = scalar.transform(X_test)
#training a DescisionTreeClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
rf_model = RandomForestClassifier(max_depth = 20,)
rf_model.fit(X_train_sc, y_train)
rf_predictions = rf_model.predict(X_test_sc)
# creating a confusion matrix
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
confusion_matrix(y_test, rf_predictions)
print(classification_report(y_test, rf_predictions))
rf_importance = rf_model.feature_importances_
feature_importance_rf = sorted(zip(X_train.columns, rf_importance), reverse = True)
feature_importance_rf_df = pd.DataFrame(feature_importance_rf, columns = ['Feature',
'Importance'])
feature_importance_rf_df.sort_values(by=['Importance'], inplace=True, ascending=False)
```

```
plt.figure(figsize = (7, 10))
ax = sns.barplot(x = feature_importance_rf_df['Feature'][:25],
        y = feature_importance_rf_df['Importance'][:25])
ax.set_xticklabels(ax.get_xticklabels(), rotation = 40, ha = 'right')
plt.xlabel('Features')
plt.ylabel('Importances')
```

Neural Networks:

```
import pandas as pd
import numpy as np
import os
import shutil
import pickle as pk
import matplotlib.pyplot as plt

from sklearn.preprocessing import OneHotEncoder
from sklearn.model_selection import train_test_split

from keras import models
from keras import layers
from keras.callbacks import EarlyStopping, ModelCheckpoint
from keras.models import load_model
from sklearn import preprocessing
from keras.utils.np_utils import to_categorical
l_encode = preprocessing.LabelEncoder()
l_encode.fit(df_final[target_column])
Y = l_encode.transform(df_final[target_column])
Y = to_categorical(Y)
Y
]  from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(df_final[feature_columns],Y,
test_size=0.3)
_train_sc = scalar.fit_transform(X_train)
X_test_sc = scalar.transform(X_test)
from keras.models import Sequential
from keras.layers import Dense
in_dim = len(df_final[feature_columns].columns)

import tensorflow as tf
model = tf.keras.Sequential([
    tf.keras.layers.Dense(9, input_dim = in_dim, activation = 'relu'),
    tf.keras.layers.Dense(15, activation = 'relu'),
    tf.keras.layers.Dense(5, activation = 'relu'),
    tf.keras.layers.Dense(9, activation = 'softmax')
```

```
])

##compile the model
model.compile(loss = 'categorical_crossentropy', optimizer = 'adam', metrics =
['accuracy'])
history = model.fit(X_train_sc, y_train, epochs = 25, batch_size = 5,
                    validation_data=(X_test_sc, y_test),
                    verbose=2)

import matplotlib.pyplot as plt

##plot the scores from history
def plot_graphs(history, string):
  plt.plot(history.history[string])
  plt.plot(history.history['val_'+string])
  plt.legend([string, 'val_'+string])
  plt.xlabel("Epochs")
  plt.ylabel(string)
  plt.show()

plot_graphs(history, "accuracy")
plot_graphs(history, "loss")
```