

Real Time Air Quality Monitoring and Weather Forecasting System

Team Members:- Shayoun Bhaduri, Pranjal Srivastava, Ayush Misra

Mentor:- Er. Mrityunjai Rai.

The problem statement given to our team was Real Time Air Quality Monitoring and Weather Forecasting System. The problem statement was fun and exciting to work upon. I would like to thank **IBM** and **Smart Internz** to provide us with a wonderful opportunity like this so that we can learn and implement at the same time.

We have divided our problem statement into 3 parts for better integration and faster work delivery:-

1. Air Quality Monitoring
2. Weather Forecasting
3. User Interface or Front End

Our team has focused on delivering accuracy till the last value available.

Pre-requisites for the model

For making a working model of the Problem Statement , we require following pr-requisities:-

- **Jupyter Notebook**:- This is where we will run our code
- **Weather Forecasting DataSet**:- This is the dataset on which we will work upon. The dataset used in the model has the temperature recordings from the years 1901 to 2017 (upto 116 years) and using this dataset we will forecast the weather for the year of 2018.
Link for the dataset:- <https://github.com/smartinternz02/SBSPS-Challenge-9571-Real-Time-Air-Quality-Monitoring-Weather-Forecasting-System/blob/main/Team-Hyperlearn/Weather%20Forecasting%20Model/Weather%20forecasting%20%20Data%20in%20India%20from%201901%20to%202017.csv>
- **Air Quality Index dataset** :- This is the dataset on which we will work upon for the Air Quality Model. This dataset consists of the data of the major cities of india and their Air Quality Index(AQI).
Link for the dataset:-https://raw.githubusercontent.com/smartinternz02/SBSPS-Challenge-9571-Real-Time-Air-Quality-Monitoring-Weather-Forecasting-System/main/Team-Hyperlearn/Real%20Time%20Air%20Quality%20Monitoring/city_day.csv
- **Geopandas library**:-Installation of its environment on the system. This library will help us to plot the Air QualityIndex of various cities of india on the Indian Map.
- **Matplot Library** :- This library will plot all the graphs for the weather forecasting model and air quality model.
- **Plotly Library** :- This library adds the animation and effects to the models.
- **Pandas and Numpy Libraries**:- These libraries are used to work upon the time series datasets.
- Basic Machine Learning Algorithms such as **K-NN, K-means, Decision-Tree**

Weather Forecasting System

This model shows the various trends in the temperature that is spread across over a period of 116 years. The main highlights of the model include it's capability to work on a large dataset

and its accuracy (approx 96.03).

The model uses **plotly** and **matplotlib** graphs to analyse and study the temperature trends.

Analysis of Weather Forecasting Model

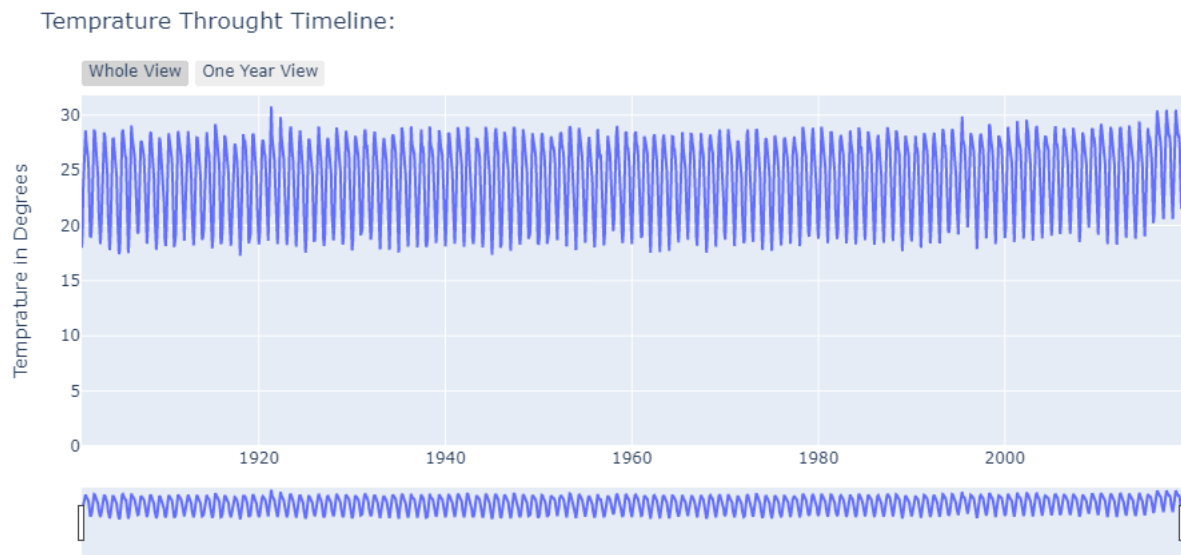


Fig1.

The Fig1. shows the temperature trend throughout the timeline. The timeline taken here spreads from 1900 to 2050. This gives us a broad idea about how the temperature has changes or is changing .

This plotting is helpful to understand the change in the trend of temperature for each year between 1900 to 2050.

Next, a **Box Plot** is plotted for the tempreature of the timeline.

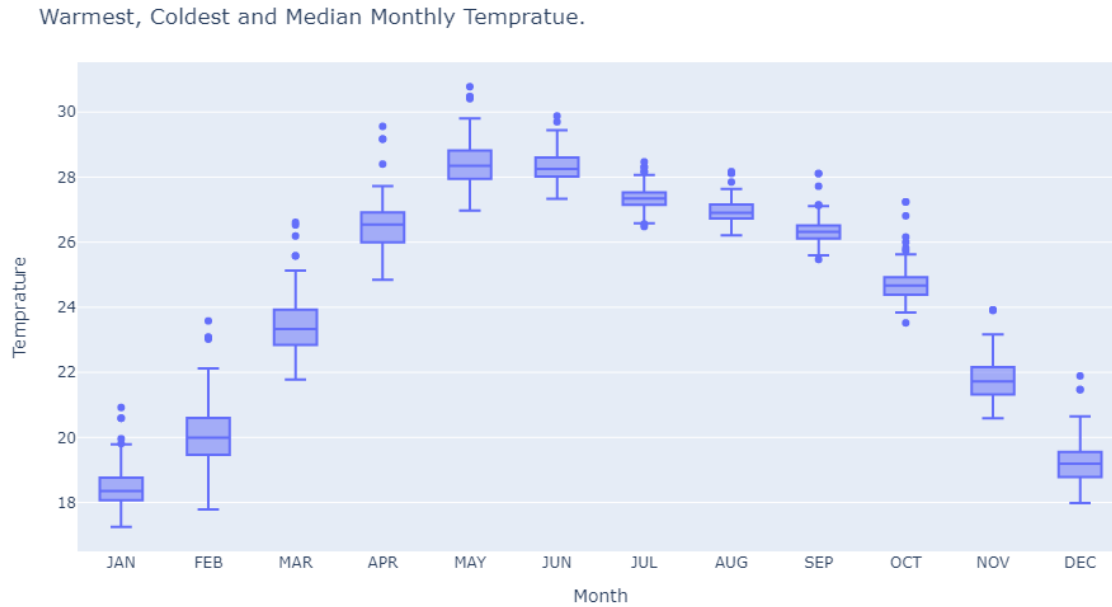


Fig 2.

This box plot plots the entire temperature data into two sections:-

1. **Warmest**
2. **Coldest**

There is also a **median monthly temperature**. This temperature is the median of all the temperature that was recorded for that particular month.

This median temperature helps to increase the accuracy of the model and its prediction.

Using the box plot, a temperature cluster has been created which helps to classify or categorize the temperatures into three main clusters:-

1. Hottest
2. Coldest
3. Neither too hot nor too cold

Temprature clusters.

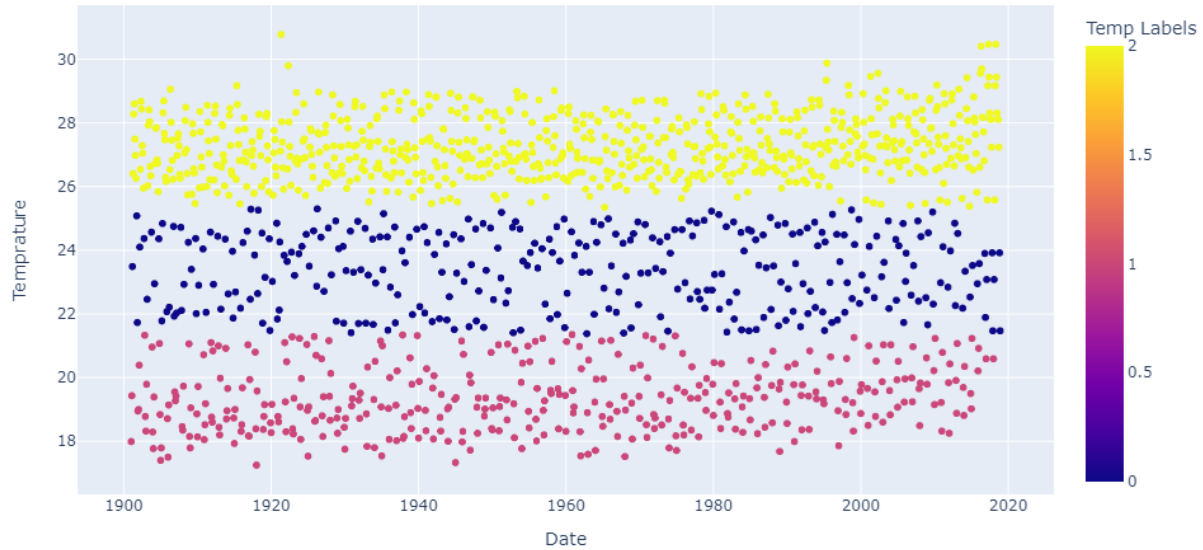


Fig 3.

Fig 3 shows the three clusters that houses the entire temperarues of the dataset. This cluster graphs helps in the study of the hottest temperature of the particular year and the coldest temperature of the particular year.

This graph is useful for the study and analysis of the variation that happens in the temperature over the period of time.

Yearly Mean Temprature :

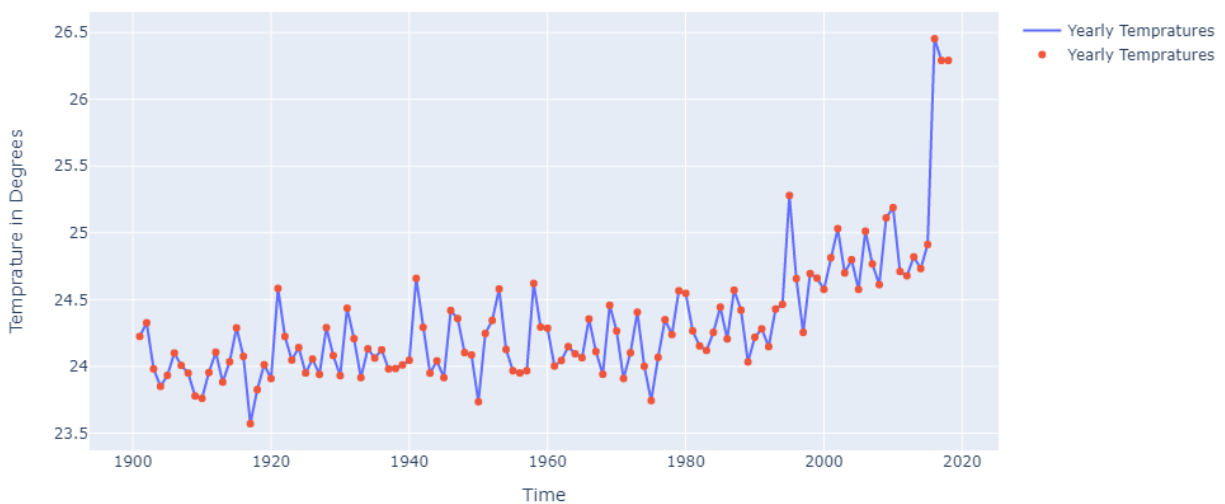


Fig 4.

The graph in fig 4 shows the line graph of the yearly mean temperature for each year. This graph is very vital for the research and study of the mean temperature for a particular year. The main use of this graph is that it helps to analyse and study the minutest variation in the temperature.

A similar trend line has been recorded after observing this line graph.

Fig 5 showcases the trend line graph that has been recorded with the help of graph in fig 4.

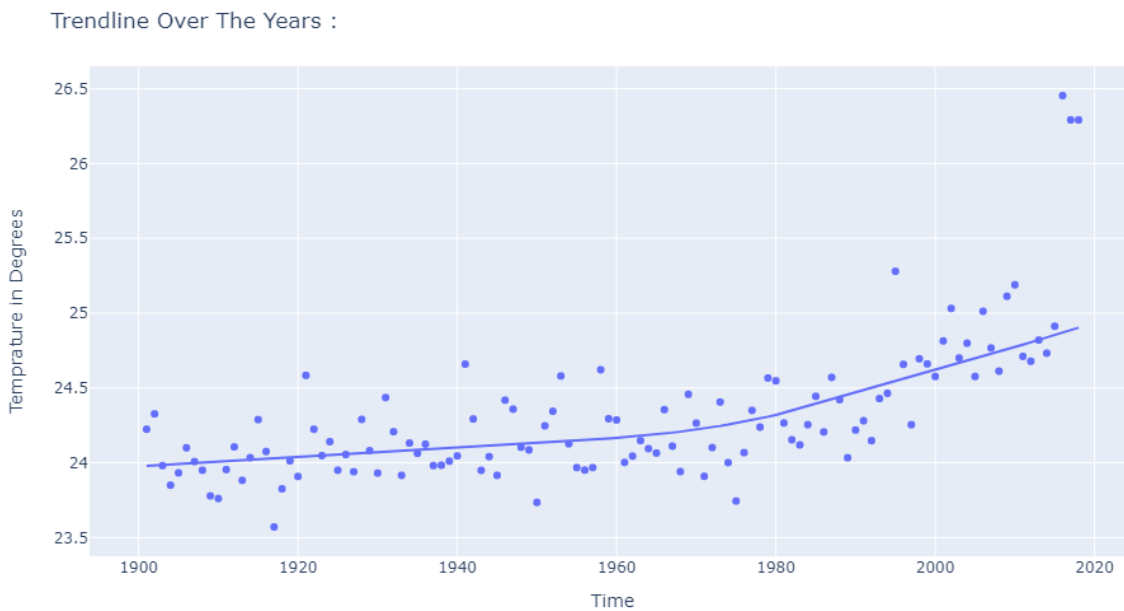


Fig 5.

Observing the yearly mean and its trendline meant that there were some changes happening in the temperature over a period of year. To understand the change more closely, a monthly analysis of all the 116 years has been plotted.

This analysis gives the variation happening over a day of the month, thus giving the exact information about when and at what time the variation occurred.

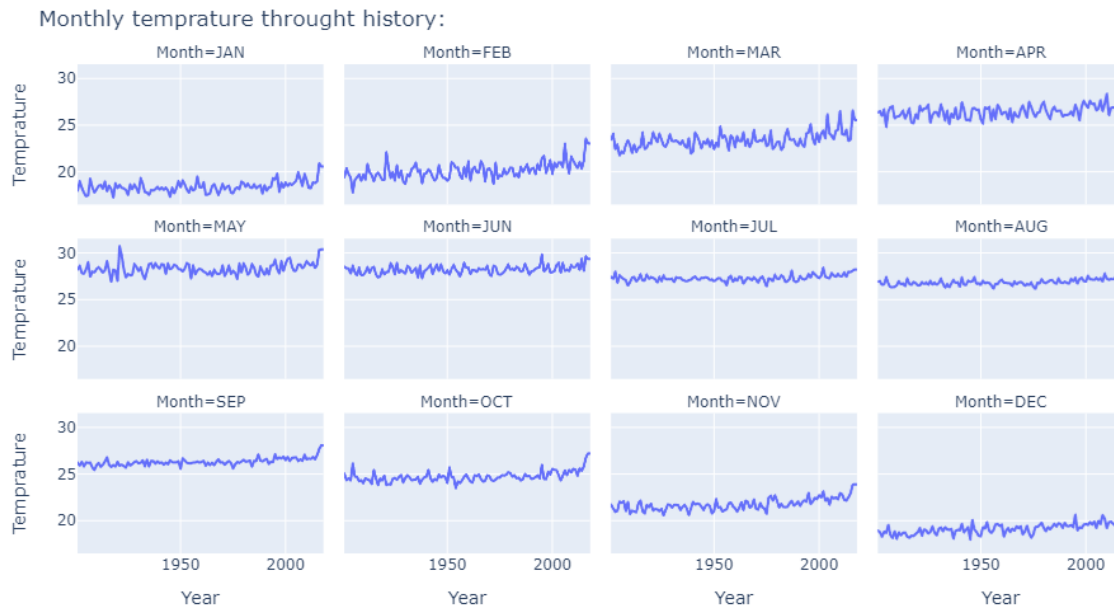


Fig 6.

After monthly analysis, it is very important to understand how the temperature changes throughout the various seasons of the year.

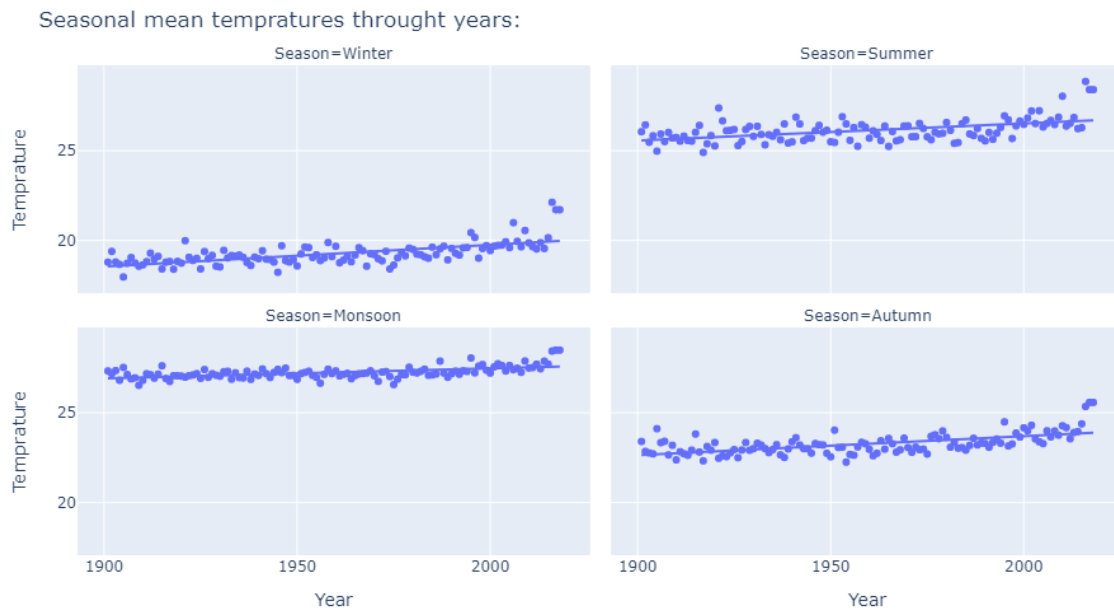


Fig 7.

Observing these analysis a scatter plot has been recorded for the entire dataset . This plot helps to undersatnd how much variation occurs in a particular year.

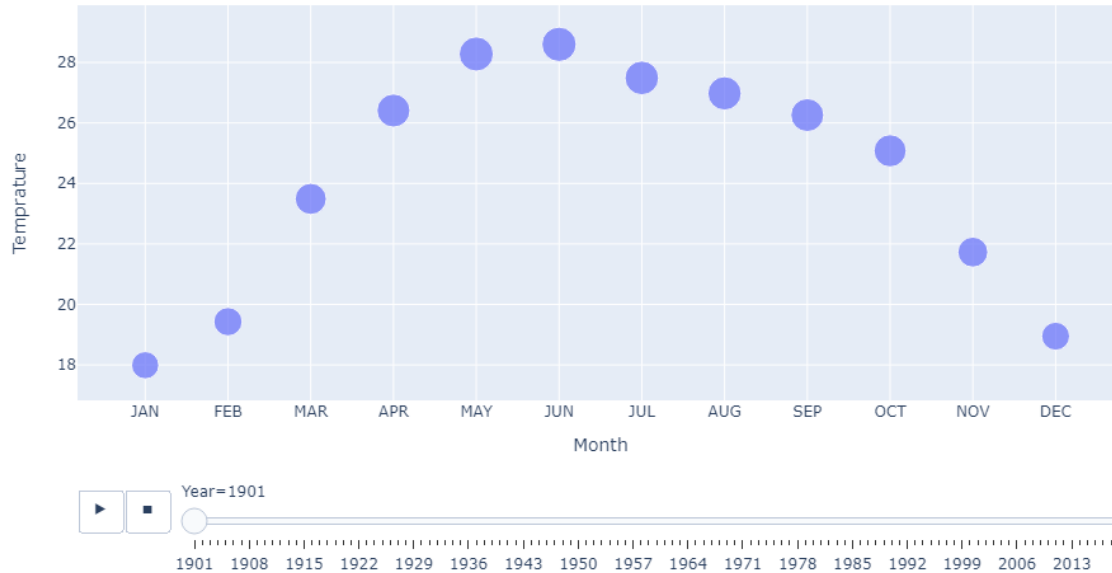


Fig 8.

Forecasting the Weather

After analysing and observing the weather from the year 1901 upto 2017, now the model predicts the weather for 2018.

The model uses **Decision Tree** algorithm for prediction, since it can handle such large dataset easily and in addition to that decision tree works on by sorting the entire dataset thus making it easier to work upon.

Fig 9 shows a line graph that showcases the temperature for the year 2018.

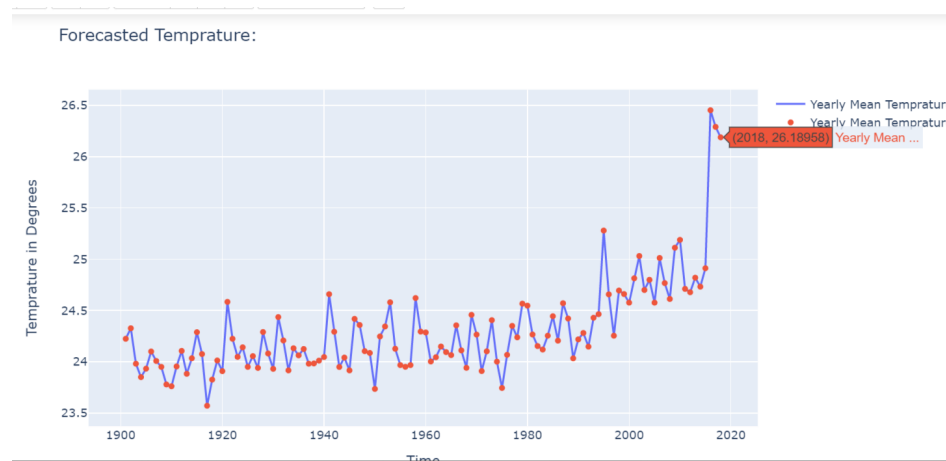


Fig 9.

The fig 9 shows the final output of the **Weather Forecasting Model**.

The model holds a accuracy of 96.03 which shows that the forecasted tempertature is almost equal to the actual temperature for the year 2018.

Air Quality Monitoring

After the weather forecasting model, Air Quality is being observed.

After through analysis and research through various graphs and plots, the model plots and shows the cities of India with their AQI levels.

Fig 1 is a simple bar graph that shows the cities with their AQI levels.

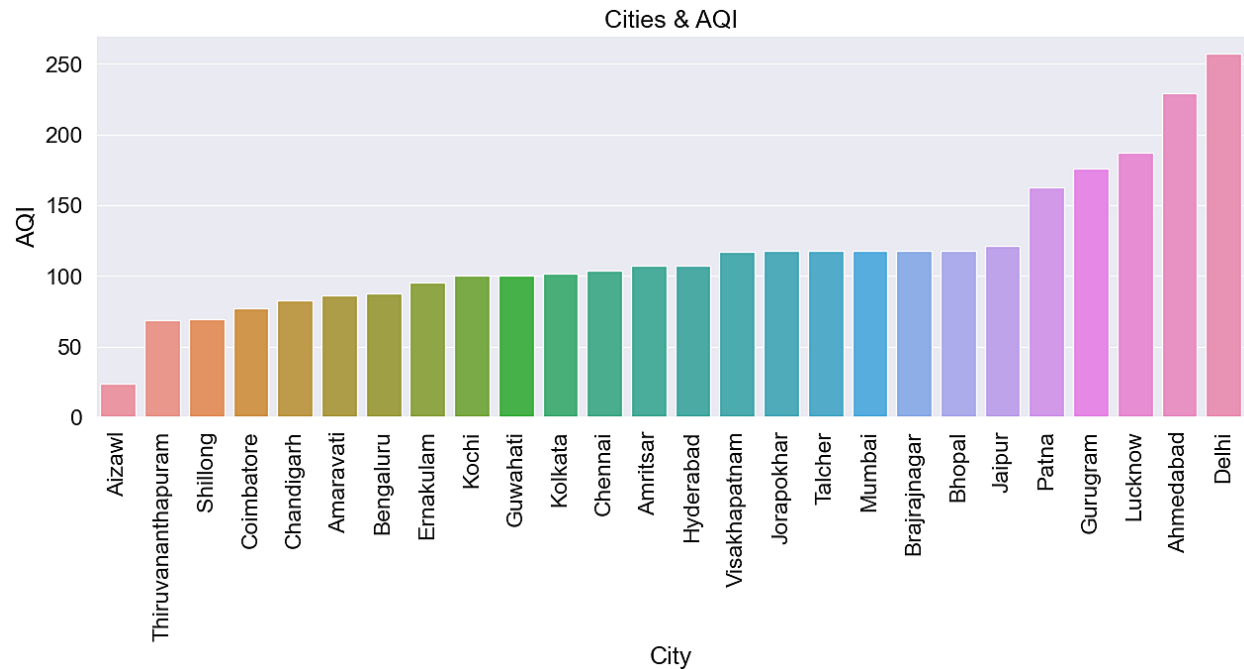


Fig 1.

This bar graph helps to understand which city has highest AQI level and which city has least.

Fig 2 shows the year wise AQI for the cities in the dataset.

This graph plots the AQI level for the cities from 2015 to 2020 year . This grph shows the variation in the level of AQI in a span of 5 years.

On observing, it can also be seen the Ahemdabad has the highest AQI for the year 2018, which is a serious threat to the environment and ozone layer.

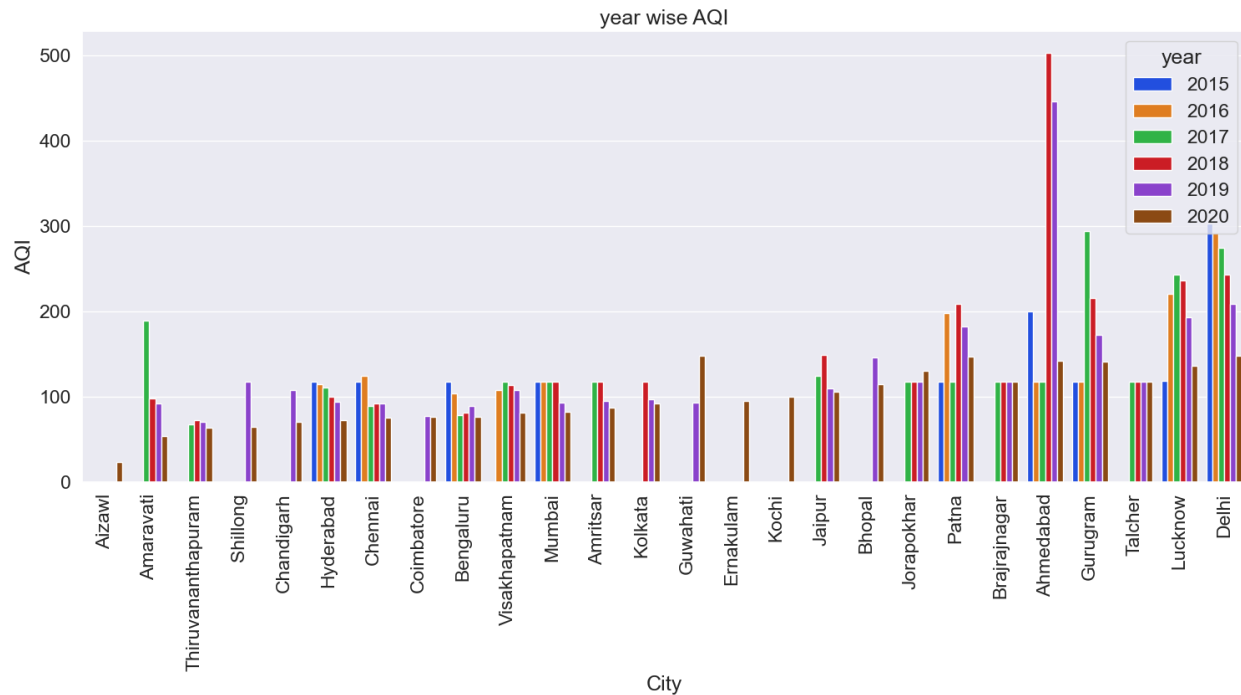


Fig 2.

On observing the variation in the AQI level for a year , it is not clear that at what time the variation happens ?, or at which month or which part of the year, the AQI rises.

To answer these questions a monthly analysis for the AQI of the cities has been performed in the fig 3.

This analysis, not only records the time at which the AQI rises, but also serves as a relevant source of study and research as to when and how AQI level can be kept under control.

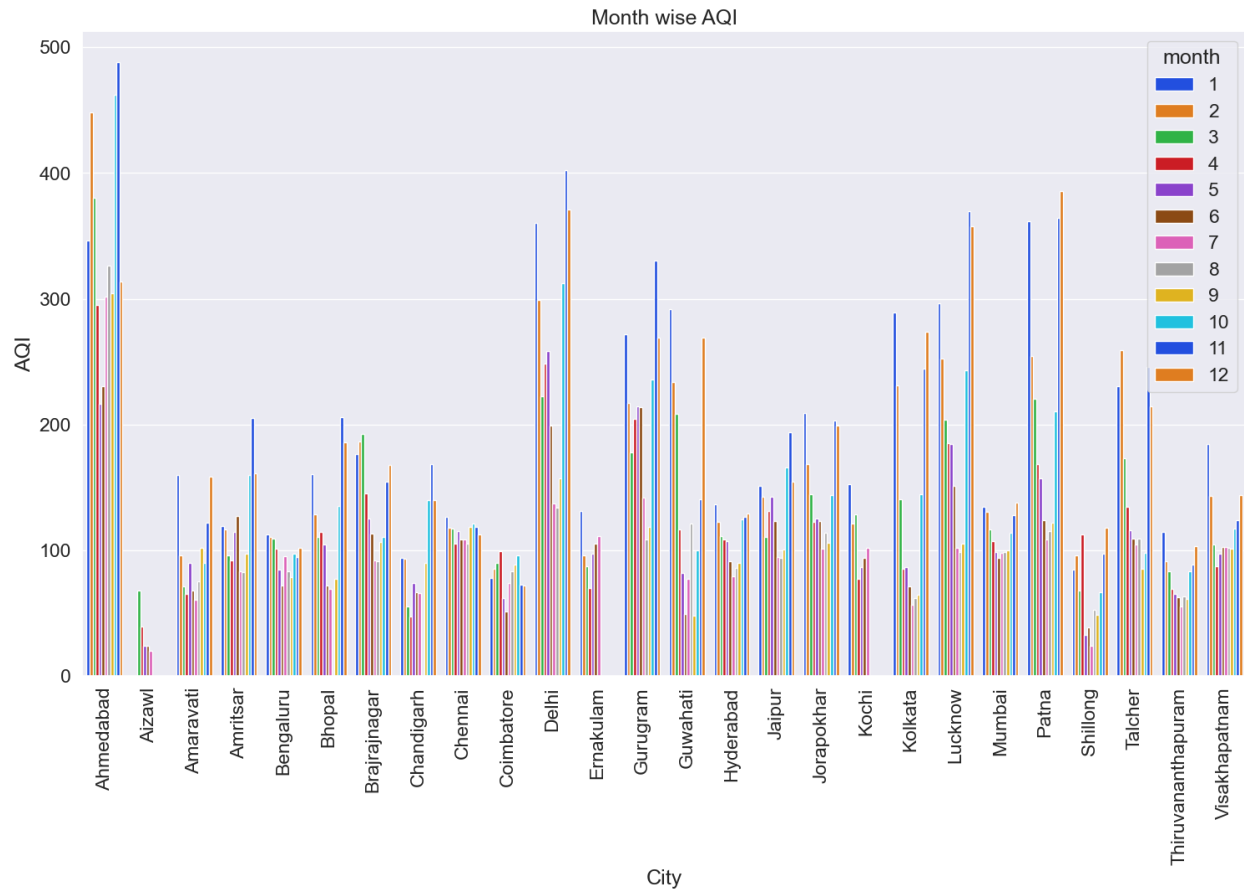


Fig 3.

After viewing the increasing trend of AQI over the years and over the months, it is time to plot these readings on the geographical map of India.

For this the library, *geopandas* is required.

But to use that, it is required that its environment must be created.

SO, to create the environment, a set of command is run on the annaconda terminal and then the environment of the jupyter notebook is changes to 'myenv'.

After completing all the pre-requisites for the geopandas library, it is time to plot the graphs.

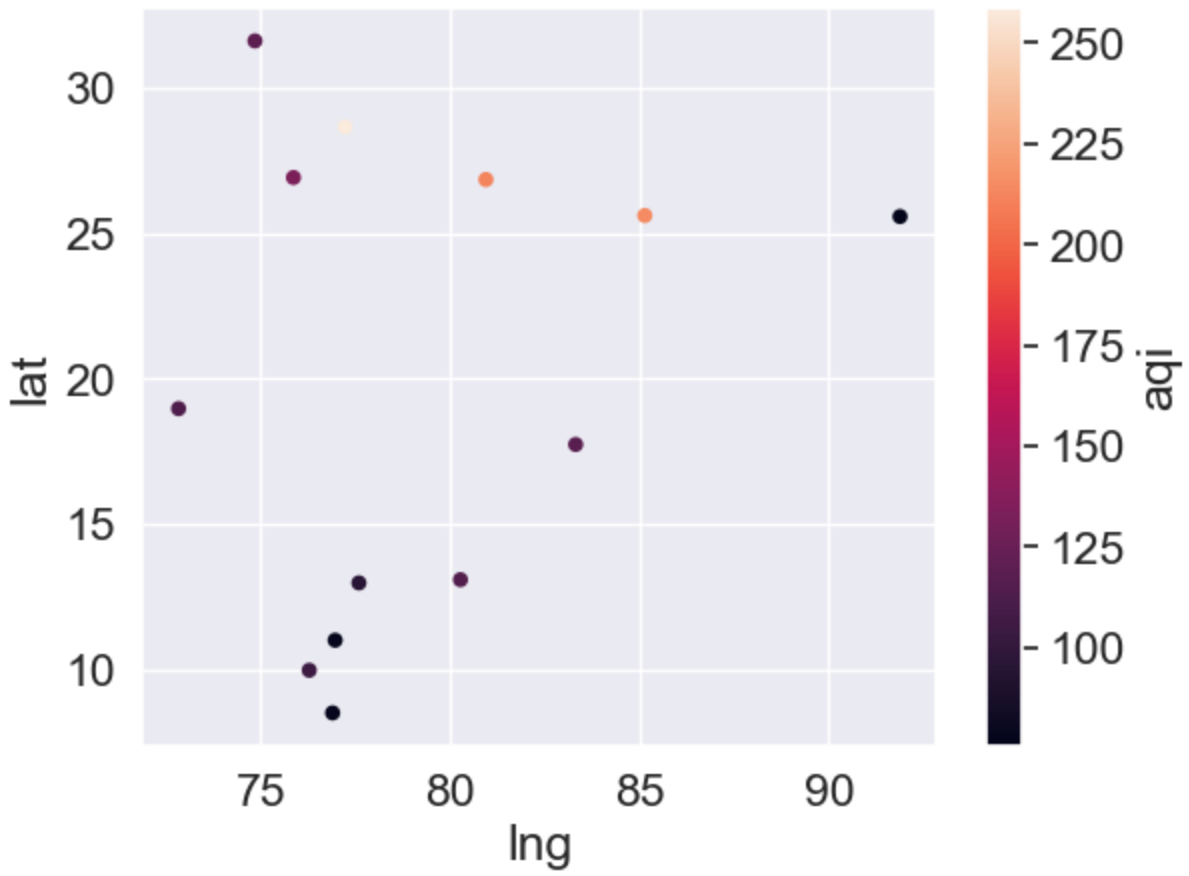


Fig 4

This fig 4 shows the AQI levels of various cities to be plotted on the Geopolitical map of India.



Fig 5.

Fig 5 shows the various major cities of India with their AQI levels.

This graph is a geo scatter plot and this graph helps to find out the AQI of the cities of India.

If given the real time dataset, then using this graph, one can easily find out the AQI level of that particular place at that given instance.

Conclusion

- In the above sections of the report, we have discussed the working and process behind our proposed models .
- A basic conclusion that can be drawn from the study and analysis of weather and Air Quality Index is that, to achieve 100 percent accuracy for prediction, one must analyse and study the trend that the data is following.
- It is very important to understand the trend that the particular temperature or a particular ppm matter follows.
- By understanding this trend, if this model is provided with real time data then, predictions can be made for the future with enhanced accuracy.