Air Quality and Pollution Data Analysis

Nilakrishna M

Rajagiri College of Social Sciences, Kalamassery

msccs218@rajagiri.edu

Sr. Jinsy Jose

Asst.Professor

Rajagiri College of Social Sciences,Kalamassery

srjinsijose@rajagiri.edu

1. Abstract

Today India is the second most polluted country in the world. It is posing a heavy threat to the country's health and economy. Almost all of India's 1.4 billion people are exposed to unhealthy levels of ambient PM 2.5 – the most harmful pollutant - emanating from multiple sources. Here I present the analysis of government air quality data of India from 2015–2020 from the Central Pollution Control Board (CPCB). The dataset contains air quality data and AQI (Air Quality Index) at hourly and daily level of various stations across multiple cities in India. I analyzed the data to get a better understanding of the major causes of air pollution in India and how the air quality varies over the years.

Keywords: Air Quality Index, Air Pollution, Exploratory data analysis, Time series analysis, Data visualization

2. Introduction

Air pollution in India is a serious environmental issue. Of the 30 most polluted cities in the world, 21 were in India in 2019. As per a study based on 2016 data, at least 140 million people in India breathe air that is 10 times or more over the WHO safe limit and 13 of the world's 20 cities with the highest annual levels of air pollution are in India. Air Quality Index (AQI) is a number used to communicate the level of pollution in the air and it essentially tells you the level of pollution in the air in a given city on a given day. In this paper I analyzed the dataset containing air quality data and AQI at hourly and daily level of various stations across multiple cities in India using Python and Tableau. I've done exploratory data analysis and various data visualization techniques to get a better understanding of the data. Ahmedabad and Delhi turns out to be the most polluted city in our country. Of the pollutants, PM 2.5 exceeds the standards the most, followed by PM 10, NO2, CO, and Ozone.

3. Literature review

Several research has been done on Air pollution in India. I got various insights by referring the papers published by other researchers.

Anikender Kumar, PramilaGoyal (2011) presented the study that forecasts the daily AQI value for the city Delhi, India using previous record of AQI and meteorological parameters with the help of Principal Component Regression (PCR) and Multiple Linear Regression Techniques. They perform the prediction of daily AQI of the year 2006 using previous records of the year 2000-2005 and different equations.

Aditya C R (et al.2018) employed the machine algorithms to detect and forecast the PM2.5 concentration level on the basis of dataset containing atmospheric conditions in a specific city. They also predicted the PM2.5 concentration level for a particular date.

Nidhi Sharma (et al.2018) had gone through the detailed data analysis of air pollutants from 2009-2017 and also proposed the critical observation of 2016-1017 air pollutants trend in Delhi, India [14]. They have predicted the future trends of various pollutants as Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), Suspended Particulate Matter (PM), Ozone (O3), Carbon Monoxide (CO) and Benzene. By using data analytics Time series Regression forecasting they have predicted the future values of the pollutants mentioned earlier on the of previous records.

Mohamed Shakir and N.Rakesh (2018) have analysed the proportion of various air pollutants (NO, NO2, CO, PM10 and SO2) with respect to the time of the day and the day of the week and estimated the effect of environmental parameters as temperature, wind speed and humidity on the air pollutants mentioned above with the help of WEKA tool.

R. Gunasekaran (et al.2012) the main objective of this study is to monitor the air quality of Salem Swadeswari College, Tamil Nadu area for the period of April 2011 to March 2011 and it has been shown that this area has no serious pollution issues related to the pollutants as Sulfur Dioxide, Oxides of Nitrogen and Suspended Particulate Matter because their annual average concentration are within the range of national standards.

4. Implementation

We are using python and Tableau to do exploratory data analysis and visualization techniques on the data to get various insights**.**

a) Python

Python is a popular programming language in scientific computing, because it has many data-oriented feature packages that can speed up and simplify data processing, thus saving time. Python is a multi-functional, maximally interpreted programming language with several advantages that are often used to streamline massive, and complex data sets.

b) Tableau

Tableau is one of the most emerging and has widespread usage both in industry and education purposes. Tableau is primarily used for data visualization and reporting, which will provide us a spectacular understanding about the data and its reachability. Anyone without having the prior knowledge of coding can efficiently work on Tableau, for thus is its simplicity in performance. We will be applying the tableau visualization on our dataset to get better idea and clarity on the scope of data.

c) Dataset

The dataset contains air quality data of India from 2015-2020 and AQI (Air Quality Index) at hourly and daily level of various stations across multiple cities in India. The attributes are
City, Date, Pollutants like PM2.5, PM10, NO, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, Xylene , AQI, Air quality. The cities in this dataset are Ahmedabad, Aizawl, Amaravati, Amritsar, Bengaluru, Bhopal, Brajrajnagar, Chandigarh, Chennai, Coimbatore, Delhi, Ernakulam, Gurugram, Guwahati, Hyderabad, Jaipur, Jorapokhar, Kochi, Kolkata, Lucknow, Mumbai, Patna, Shillong, Talcher, Thiruvananthapuram, Visakhapatnam.

d) Exploratory Data Analysis

Exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. EDA is applied to investigate the data and summarize the key insights. It will give you the basic understanding of your data, it's distribution, null values and much more. You can either explore data using graphs or through some python functions.

e) Time Series Analysis

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, we record data points at consistent intervals over a set period of time.

5. Results and Discussion

Bar plot of Air Quality Index in various cities is plotted. From this we find that

- Ahmedabad and Delhi are the most polluted cities in India
- Amaravati is the least polluted city.

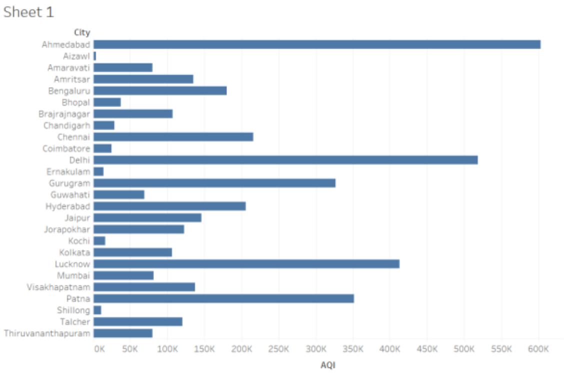

Fig 1 : Bar plot of Air Quality Index in various cities

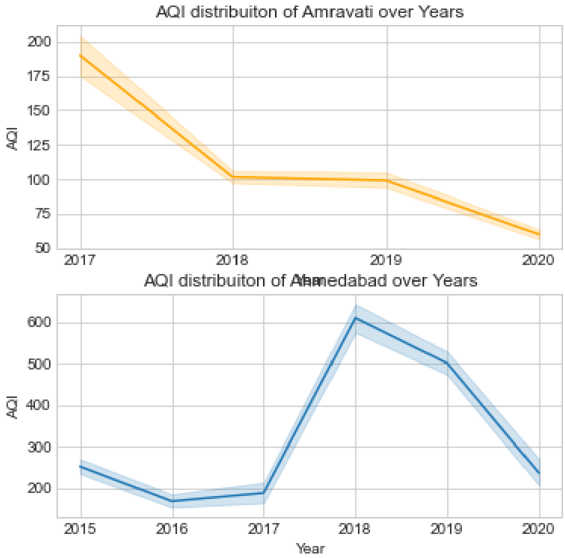

Fig 2: AQI distribution of Ahmedabad and Amaravati from 2015 to 2020:

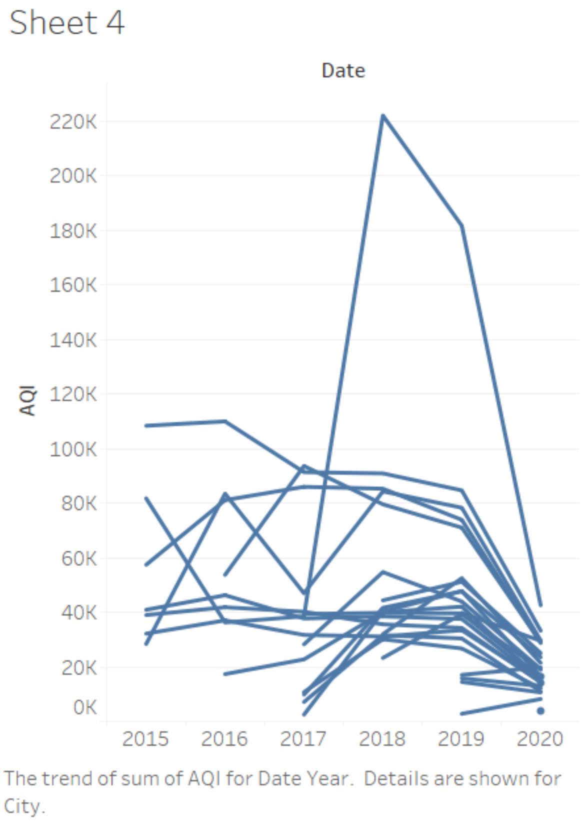

Fig 3: AQI distribution of all cities over the years

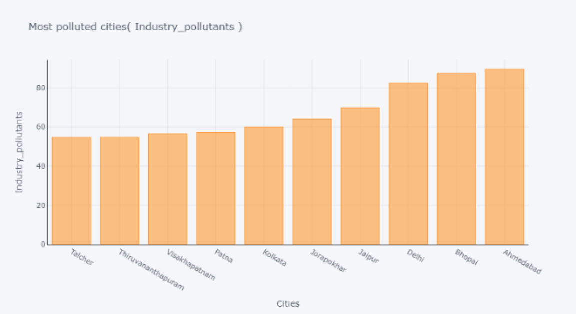

Fig 4: Most polluted cities due to industrial pollutants

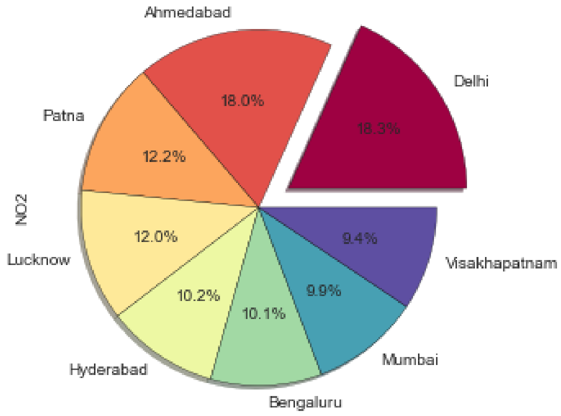

Fig 5 : Distribution of NO2 in various cities.

- From the above plots it is clear that during COVID19 lockdown there is gradual decrease in vehicular pollution contents, industrial pollution content.
- Delhi is the most polluted city in terms of vehicular pollution contents.
- Ahmedabad is the most polluted city in terms of industrial pollution content.
- Of the pollutants, PM 2.5 exceeds the standards the most, followed by PM 10, NO2, CO, and Ozone.
- The cities with AQI less than 50 have a relatively good air quality.
- Amaravati and Hyderabad are the least polluted cities.

6. Conclusion

There is great need to control the air pollution as it is impacting the environment and human health seriously. The concentration of air pollutants like have to be controlled to save the environment. Long-term health effects from air pollution include heart disease, lung cancer, and respiratory diseases such as emphysema. To control air pollution, proper rules and regulations should be implemented by the government, awareness among the people, control the growth of population, number of vehicles, industries and energy consumption. We need to take pollution issue seriously because ignorance is certainly not the proper way to go. The stakes are really high and world needs to wake up and start acting right now because environmental issues are constantly growing in number and size.

7. References

[1]. Sharma, Disha, and Denise Mauzerall^{+}. "Analysis of Air Pollution Data in India between 2015 and 2019." Aerosol and Air Quality Research 22, no. 2 (2022).

[2]. Anikender Kumar, PramilaGoyal, "Forecasting of air quality in Delhi using principal component regression technique", Atmospheric Pollution Research, 2 (2011) 436-444.

[3] . Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu, "Detection and Prediction of Air Pollution using Machine Learning Models", International Journal of Engineering Trends and Technology (IJETT) – volume 59 Issue 4 – May 2018

[4]. . Nidhi Sharma , ShwetaTaneja , VaishaliSagar , Arshita Bhatt, "Forecasting air pollution load in Delhi using data analysis tools", ScienceDirect, 132 (2018) 1077– 1085.

[5]. Mohamed Shakir, N. Rakesh, "Investigation on Air Pollutant Data Sets using Data Mining Tool", IEEE Xplore Part Number:CFP18OZV-ART; ISBN:978-1- 5386-1442-6.

[6]. R. Gunasekaran, K. Kumaraswamy, P.P. Chandrasekaran, R. Elanchezhian, “MONITORING OF AMBIENT AIR QUALITY IN SALEM CITY, TAMIL NADU”, International Journal of Current Research, ISSN: 0975-833X, Vol. 4, Issue, 03, pp.275- 280, March, 2012.

[7].https://en.wikipedia.org/wiki/Air_pollution_in_India