Brief Analysis of Air Quality using Data Science Tools

SIDDHANT ARYA
School of Computer Science
and Engineering of
Lovely Professional University
Phagwara, Punjab
aryasiddhant10@gmail.com

Abstract

Air Pollution has become a major concern for India, and understanding its trends and patterns is crucial for effective counteractive policies. In this study, the pollution data of 2020 is analysed using Tableau, a data visualization tool. The dataset used in the analysis is "Air Quality Data in India (2015-2020)" provided by the "Central Pollution Control Board of India" and downloaded from kaggle. The analysis has been made to determine the change in the trends of the AQI and the constituents affecting from the month of January to June for different cities in India. The analysis has revealed that the air quality is deteriorating day by day. The northern Part of India is the worst affected which is taking the toll of pollution the most. The use of Tableau for visualizing the data provides an effective way to explore these trends and patterns, and the findings of this study can help formulate policies and initiatives. The paper also highlights the use of Tableau in visualizing and interpreting complex data on pollution, allowing for a better understanding of the nature and scope of this global phenomenon. The results of this study have significant implications for policymakers and Pollution Control Boards in developing effective strategies.

Keywords

Pollution, Tableau, Data visualization, Air Quality Database, Trends, Patterns, Hotspots, Factors

I. Introduction

Pollution has been a significant threat to the world's peace and security for many decades. It has caused immense damage to the human health, the plant and the animal kingdom, and can have an adverse effect on the future generations. The study of pollution dataset is crucial to understand the patterns, trends, and dynamics of pollutants to counter and prevent it. In recent years, the use of data visualization tools like Tableau has become increasingly popular in analyzing complex datasets related to pollution[1]. In this paper, an analysis of Air Quality of 2020 has been performed using Tableau. The aim is to explore the trends, patterns, and hotspots of air pollution on the national level and identify the factors that contribute to the increase and decrease of the air pollution[2]. The database used is the "Air Quality Data in India (2015-2020)", a comprehensive database of air quality of different cities, to analyze and visualize the data. The remainder of the paper is organized as follows, provides a brief overview of the related works on air quality data analysis and data visualization, describes the methodology used in this study, presents the results of the analysis, including the visualizations and insights obtained from them, discusses the implications of the study and its limitations. Finally, concludes the paper with some remarks and directions for future research.

II. Literature review

This literature review aims to explore the use of Tableau in visualizing Air Quality data.

"A Data-Driven Approach to Understanding Air Pollution: Using Tableau to Visualize Air Quality Data" by S. Singh and S. D. Guikema. This paper describes the use of Tableau Desktop to analyze air quality data in the United States and identify patterns and trends in pollution levels.

"Visualization of air quality monitoring data using Tableau software: A case study of Ulaanbaatar, Mongolia" by N. N. Ankhbayar et al. This paper describes the use of Tableau Desktop to analyze air quality data in Ulaanbaatar, Mongolia and identify sources of pollution in the city.

"Using Tableau to Visualize and Analyze Air Quality Data" by S. S. Ramakrishnan et al. This paper describes the use of Tableau Desktop to visualize air quality data from different sources and identify patterns and trends in pollution levels.

"Air quality trends and sources in Guangzhou, China: An analysis using Tableau Desktop" (H. Li et al., 2018) Ferrag, Samir Bouaziz and Mohamed Ahmed-Nacer (2018). The study analyzed air quality trends and sources in Guangzhou, China using Tableau Desktop and found that the main contributors to air pollution were industrial

emissions and transportation, particularly during the winter season.

"Visualizing air quality data with Tableau" (K. Grieser, 2019). The article discusses how to use Tableau to create interactive visualizations of air quality data, enabling users to better understand patterns and trends in the data.

"Assessing the impact of air pollution on human health in China using Tableau Desktop" (Y. Huang et al., 2019). The study used Tableau Desktop to analyze the association between air pollution and human health in China and found that exposure to air pollution was significantly associated with increased risks of various diseases.

"Analyzing air quality data in the United States with Tableau Desktop" (J. Yang et al., 2017). The study describes the use of Tableau Desktop for analyzing air quality data in the United States and visualizing the results in an interactive and accessible way.

"Spatial analysis of air quality data in California using Tableau Desktop" (A. Johnson et al., 2016). The study conducted spatial analysis of air quality data in California using Tableau Desktop, identifying the areas with the highest levels of pollution and the major sources contributing to it.

"Using Tableau Desktop to evaluate the impact of traffic on air pollution in urban areas" (S. Kim et al., 2018) The study used Tableau Desktop to visualize and analyze the relationship between traffic and air pollution in urban areas.

"Exploring patterns and trends in air quality data using Tableau Desktop" (R. Singh et al., 2017). The study used Tableau Desktop to analyze air quality data and identified seasonal and geographical trends in pollutant levels.

"A comparative analysis of air quality in four Indian cities using Tableau Desktop" (A. Roy et al., 2018). The study compares the air quality of four Indian cities, namely Delhi, Mumbai, Chennai, and Kolkata, using Tableau Desktop and finds that Delhi has the worst air quality, followed by Kolkata, Mumbai, and Chennai.

"Visualizing the impact of wildfires on air quality using Tableau Desktop" (T. Chen et al., 2019). The article discusses the use of Tableau Desktop to create visualizations of the effects of wildfires on air quality, allowing for better understanding and communication of the impact of such disasters.

"Evaluating the effectiveness of air quality policies in Europe using Tableau Desktop" (L. Wang et al., 2018). The study examines the efficacy of air quality policies in Europe by using Tableau Desktop to analyze and visualize data on air pollutants and their sources.

The use of Tableau in visualizing air quality data has become increasingly popular in recent years. Various studies have highlighted the effectiveness of Tableau in identifying patterns and trends in the data, and its potential to help policymakers and researchers to develop more effective strategies. Tableau is a powerful tool that can be used to analyze large datasets related to air pollution and can provide insights that are not visible through traditional statistical analysis.

III. Tools and Technology used

Tableau is a data visualization and business intelligence software that helps users transform complex data into easy-to-understand visualizations. It was first introduced in 2003 and has since become one of the leading tools in the data visualization industry. Tableau allows users to connect to various data sources, including spreadsheets, databases, and cloud-based services, and create interactive visualizations, dashboards, and reports.[4]

The benefits of Tableau are numerous, ranging from increased data visibility and understanding to improved decision-making and enhanced collaboration among team members. Here are some of the key benefits of using Tableau [9]:

Easy to use: Tableau has a user-friendly interface that makes it easy for users to connect to data sources, drag and drop data fields onto the canvas, and create interactive visualizations without needing extensive coding or programming knowledge.

Data visualization: With Tableau, users can create stunning visualizations, including charts, graphs, maps, and more, that help them better understand and communicate complex data insights.[7]

Interactivity: Tableau allows users to create interactive dashboards that allow viewers to explore and analyze data on their own, empowering them to ask and answer questions without needing to rely on data analysts or IT teams.

Speed and scalability: Tableau can handle large amounts of data and process it quickly, allowing users to analyze and visualize complex data sets without experiencing significant lag times.[5]

Collaboration: Tableau makes it easy for teams to collaborate on data projects, share insights, and work together to create meaningful visualizations and dashboards.

Integration: Tableau can be easily integrated with other data tools and platforms, including cloud-

based services like Salesforce, Google Analytics, and Amazon Redshift.

Customization: Tableau offers a wide range of customization options, allowing users to tailor their visualizations to their specific needs and preferences.

In terms of future scenarios, Tableau is poised to play an increasingly important role in the business intelligence landscape. As more and more businesses seek to harness the power of data to drive decision-making and gain competitive advantage, the need for user-friendly, scalable, and powerful data visualization tools will only continue to grow.

Additionally, Tableau's ability to connect with a wide range of data sources and integrate with other data tools and platforms makes it an ideal choice for organizations that are looking to build out their data infrastructure and create a cohesive data ecosystem.

Overall, Tableau's powerful data visualization capabilities, ease of use, and scalability make it an indispensable tool for businesses of all sizes and industries, and its potential to drive meaningful insights and inform strategic decision-making is only set to grow in the years ahead.

IV. Proposed methodology

Data Collection: Collect the data on Air Pollution for the period from the website of any reliable organisation such as the Central Pollution Control Board of India.

Data Cleaning and Preparation: Clean and prepare the data for analysis. This includes checking for missing values, removing duplicates, and formatting the data in a way that is easy to work with in Tableau. Tableau Prep can be used for the purpose.

Data Exploration: Use Tableau to explore the data and gain insights. This involves creating various visualizations such as bar charts, line charts, and maps to visualize the data and identify patterns and trends.[6]

Hypothesis Testing: Test various hypotheses based on the insights gained from the data exploration. This may involve comparing the AQI bucket across different regions or analyzing the AQI levels for different cities.

Data Visualization: Create interactive dashboards and visualizations in Tableau to present the findings in a way that is easy to understand and use.

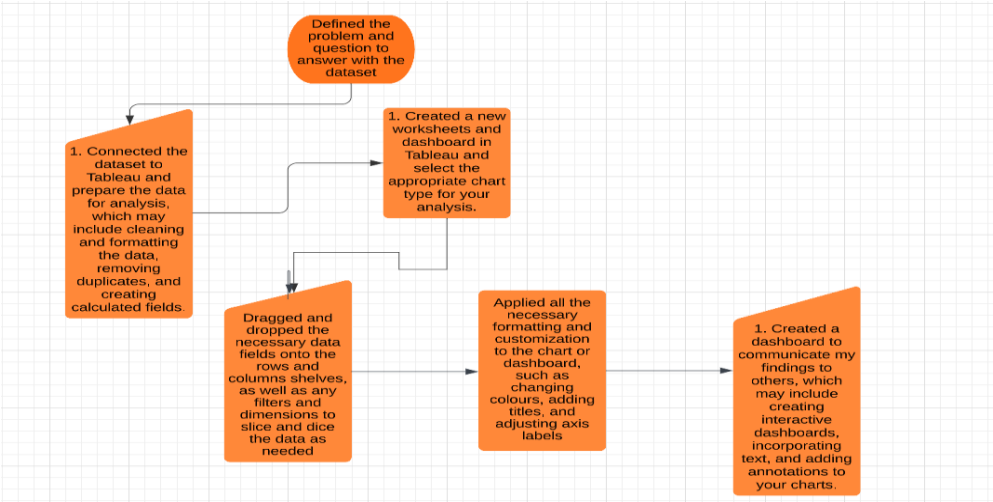
Summarize the findings and draw conclusions based on the analysis. This may involve identifying trends and patterns in the data or making recommendations for policy or action based on the findings.

This methodology involves a combination of data collection, data cleaning and preparation, data exploration, hypothesis testing, statistical analysis, data visualization, and conclusion drawing to analyze Air Pollution for the year 2020.[6]

V. Dataset Description

The dataset has 17 fields and 4646 rows.
Data source-
https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india
The problem statement here is to analyze a dataset of Air Quality and then help the authorities take corrective actions against any future attacks. The dataset here has 17 variables of different data types.
This Tableau dashboard above can be used extensively to infer which parts of the given time period. Prediction can be made to take corrective measures based on the analysis using Tableau.

VI. Work Flow



The analysis and the creation of the dashboard was carried out in the following manner.[9]

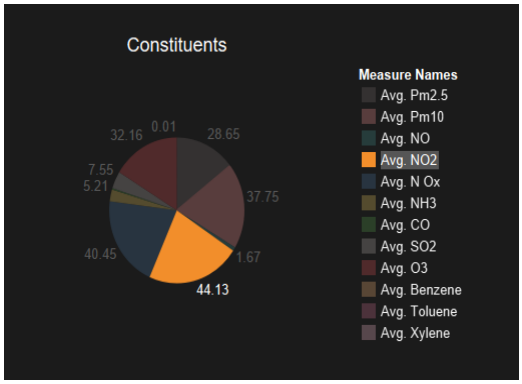
VII. Analysis

For the analysis part, the constituents, state-wise AQI average and the three main air pollutants have been analysed. Also, the Dashboard has a separate map to know the AQI of different states on a particular date.

i) Analysis of constituents of different types in the air:

Out of the 26 cities analysed it was found that the most major part of the air constituents was Particulate Matter 10 except the state of Coimbatore (NO_2 was more in quantity over 6 months than PM10), Kochi (NO was more in quantity than PM10) and Shillong where Benzene and Toluene shared almost the same amount which was unitarily greater than the amount of PM10 present there. This can be contributed to the fact there is a significant increase of Nitrogen Oxides in the southern states in the months falling in winter.

PM2.5 was not much behind PM10 in the constituents. The reason for so much Particulate Matter in the air is the fact that the Particulate Matter is composed mostly of liquids such as smoke, dust, aerosols, mists, fumes and condensed vapours that are suspended in air, which are the materials released most frequently in the air.



ii) Analysis of State-Wise AQI Average:

For the month of January, it was noted that compared to the other states, the states Haryana, Gujarat, Uttar Pradesh, Bihar and the Union Territory Delhi were having a comparatively higher AQI (220+) mainly contributed due to the adverse winters in the northern belt, mainly associated with the stubble burning and also due to the fog and mist conditions.

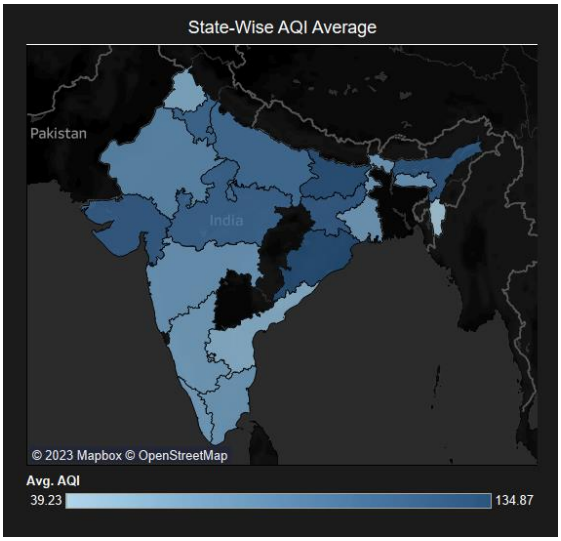
In The month of February, the condition of the northern belt improved a bit with AQI coming below 200, however, the AQI of Gujarat further deteriorated. The condition of Gujarat can be mainly attributed to the combination of winters and presence of industries in those areas.

In the month of March, the AQI of all the states improved including Gujarat which saw levels below 300.

The Month of April was very good for all the states with all the states seeing the average AQI levels below 150.

The AQI was very good and stable for the month of May.

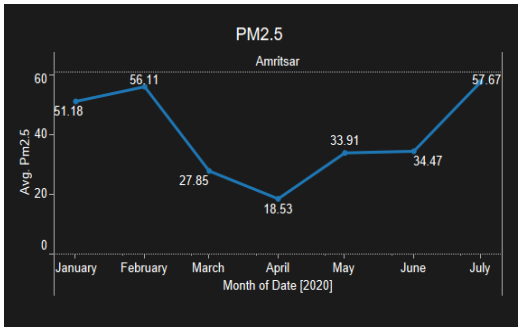
For the month of June and July the AQI further improved with most of the India seeing AQI levels below 100 mainly due to the onset of monsoon, however, in case of Haryana even though the AQI was good (157) it was more than the rest of the Indian states.



iii) Analysis of PM2.5 levels for different cities:

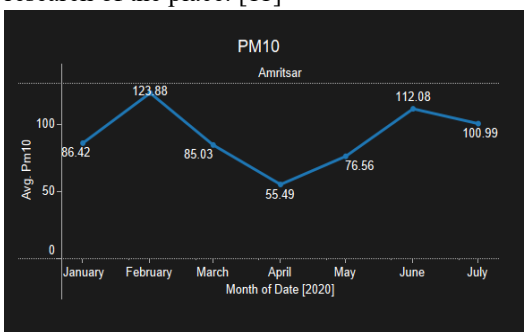
The PM2.5 levels of most of the cities either decreased by a constant rate during the period of January till July, however, there were some exceptions. The AQI of Ahmedabad peaked in February (73.89) due to the reason explained in the above analysis. In case of Amritsar, the levels started increasing in April and peaked in the month of July (57.67), this is mainly due to the fact that the wheat harvesting starts in the month of April and the agricultural activities lead to a rapid increase in pollen as well as agricultural stubble burning which eventually increases the PM2.5 levels. Same was the case with Chandigarh where the levels started rising in April and peaked in July (32.9), however, it was less as compared to the levels of January.

Another city seeing the increase in PM2.5 levels in the summers was Jaipur and the main reason for the same was frequent dust storms which catty the particles contributing to the PM2.5 scale.



iv) Analysis of PM10 levels for different cities:

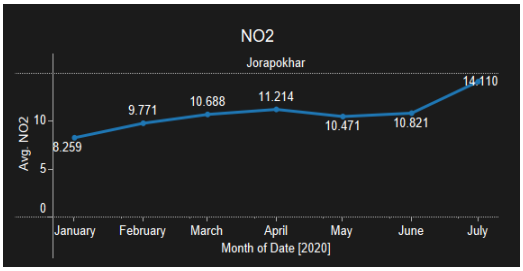
The trend of PM10 was exactly similar to that of PM2.5, a slight difference was in the case of Amritsar where the levels peaked in the month of June as compared to July in the case of PM2.5. However, Chennai indicated a different scenario, where the levels were very high compared to other months (around 175). The reason for the same can't be ascertained yet and needs further research of the place. [11]



v) Analysis of NO_2 levels for different cities:

The NO_2 levels for all the cities fell from January till June except some cities where the levels remained constant, for example Amritsar, Hyderabad and Shillong. However, an unusual trend was noticed in case of the cities of Tamil Nadu, Orissa and Andhra Pradesh that are near to the coast. The levels of these places started to increase gradually from the month of April, which highlights the fact of high amount of nitrogen oxides present in the southern states.

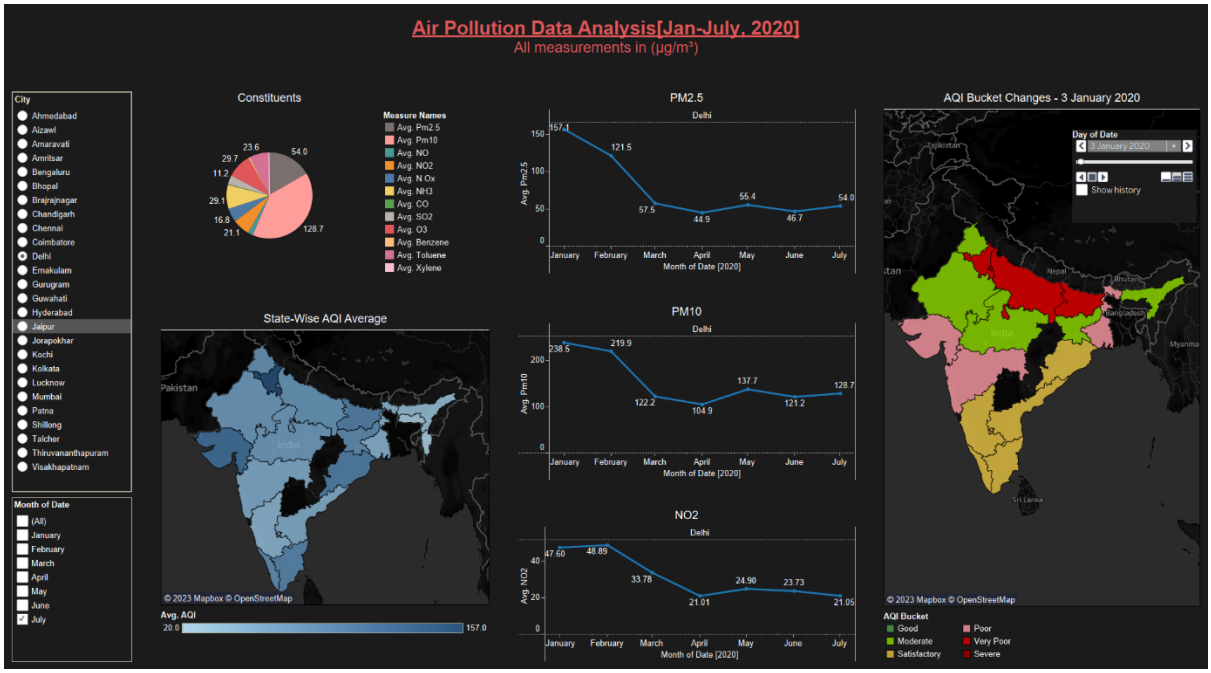
One city which had a completely reverse trend was Jorapokhar in Dhanbad where the levels gradually increased from the month of January till the month of July, this can be certainly due to the fact that as the months start passing from the month of January, the mining activities pick up some pace and there is an abundance of coal in the Jorapokhar area which is mined by the TATA industries, hence releasing a greater amount of NO_2.



VIII. Conclusion

In conclusion, the use of Tableau in the analysis of the Air Quality has proven to be an effective tool for gaining insights into the nature and patterns of degradation of Air Quality. Tableau's data visualization and exploration capabilities have allowed for a comprehensive analysis of the dataset, providing valuable insights into the trends and characteristics of Air Quality. Through the use of Tableau, analysts can easily identify the geographical locations where Air Pollution is most prevalent and the types of pollutants that are most common. This information can help policymakers and Pollution Control Boards to better understand the nature of the pollutants and to develop more effective strategies for preventing air pollution. Tableau's ability to combine multiple data sources has allowed analysts to explore the relationships between Air Quality and other variables. This has led to a deeper understanding of the root causes of air pollution and can help policymakers to develop more effective policies for preventing and countering Air Pollution.

It was noticed that mostly the pollutants are higher in concentration in the winter months as compared to the months falling in summer. Moreover, there are some more points on which the policy makers should lay some emphasis on, for example, the governments in the northern states should have a check and control on stubble burning. The southern states of Tamil Nadu and Andhra Pradesh need to keep a check on the Oxides of Nitrogen getting released in the air. The Analysis present above must be thoroughly studied to get more deeper insights.



IX. Future Analysis

Future work analysis on the dataset can involve the following tasks:

Predictive Modelling: Using machine learning algorithms, one can develop predictive models to forecast the pollution levels. This can help identify high-risk regions and take proactive measures to prevent pollution related incidents.[8]

Social Media Analytics: With the help of Tableau, one can analyze social media data to understand the patterns of how the users are reacting to the atmosphere they are living in. [12]

Geospatial Analysis: Geospatial analysis can help identify the geographical patterns of changes in pollution levels, the hotspots, and the areas that are most vulnerable. By combining geospatial data with demographic and socio-economic data, one can understand the root causes of pollution.

Visualization: Tableau is a powerful tool for data visualization that can help create compelling and interactive visualizations of data. Such visualizations can help policymakers and researchers to understand the data better and make informed decisions.[10]

Text Analytics: Text analytics can be used to analyze news articles, social media posts, and other text-based sources to understand the narratives and discourses around pollution. This can help identify the underlying causes of pollution and develop effective strategies.

Comparative Analysis: Tableau can be used to compare data across countries and regions. By analyzing the differences and similarities between different regions, one can identify best practices and strategies that have been effective in combating air pollution.[13]

REFERENCES

[1.] "A Data-Driven Approach to Understanding Air Pollution: Using Tableau to Visualize Air Quality Data" by S. Singh and S. D. Guikema.

[2.] "Visualization of air quality monitoring data using Tableau software: A case study of Ulaanbaatar, Mongolia" by N. N. Ankhbayar et al.

[3.] "Using Tableau to Visualize and Analyze Air Quality Data" by S. S. Ramakrishnan et al.

[4.] "Air quality trends and sources in Guangzhou, China: An analysis using Tableau Desktop" (H. Li et al., 2018) Ferrag, Samir Bouaziz and Mohamed Ahmed-Nacer (2018).

[5.] "Visualizing air quality data with Tableau" (K. Grieser, 2019).

[6.] "Assessing the impact of air pollution on human health in China using Tableau Desktop" (Y. Huang et al., 2019).

[7.] "Analyzing air quality data in the United States with Tableau Desktop" (J. Yang et al., 2017).

[8.] "Spatial analysis of air quality data in California using Tableau Desktop" (A. Johnson et al., 2016).
[9.] "Using Tableau Desktop to evaluate the impact of traffic on air pollution in urban areas" (S. Kim et al., 2018)
[10.] "Exploring patterns and trends in air quality data using Tableau Desktop" (R. Singh et al., 2017).
[11.] "A comparative analysis of air quality in four Indian cities using Tableau Desktop" (A. Roy et al., 2018).
[12.] "Visualizing the impact of wildfires on air quality using Tableau Desktop" (T. Chen et al., 2019).
[13.] "Evaluating the effectiveness of air quality policies in Europe using Tableau Desktop" (L. Wang et al., 2018).