NITTE (Deemed to be University) | NMAM INSTITUTE OF TECHNOLOGY

Nitte (DU) established under Section 3 of UGC Act 1956 | Accredited with 'A+' Grade by NAAC

Department of Computer Science & Engineering

Report on Mini Project

Air Quality Analysis

Course Code: CS1602-1
Course Name: Data Analysis using R Programmig

Semester: IIISEM Section: A

Submitted To:

Ms. Keerthana B C
Assistant Professor [GD-III]
Department of Computer Science and Engineering

Submitted By:

Ananya Dash (NNM22CS025)
Ansh Vashisht (NNM22CS029)
Arkal Varun Hegde (NNM22CS033)

Date of submission:
17/11/2023

Signature of Course Instructor

ABSTRACT

This study thoroughly analyses air quality parameters in various states of India. The focus is on Sulphur Dioxide (SO2), Nitrous Dioxide (NO2), Respirable Suspended Particulate Matter (RSPM), and Suspended Particulate Matter (SPM). Extensive use of graphs has been done so that we can compare the situation what was before and what was now so that we can know where we went wrong and how will the future look if we don't stop exploiting.

In analysing our project, we employed bar charts to contrast the different types of pollutants across each state. To evaluate the pollution levels over the past few years, specifically from 1990 to 2010, a heat map was utilized. For a comparison of pollution levels in Delhi, we employed a line chart to determine the highest sources of pollution, including industry, residential areas, and rural regions.

Lastly, we have used Faceted graphs to check the condition between various states having 1 graph consisting of all 4 pollutants and 1 graph having solely SO2.

The use of graphs exemplifies the integration of various libraries within the R environment, utilizing the necessary packages.

Rephrase Graphs indicates where the pollution was highest and lowest and what all measures can be taken in the future to curb them. The main significance of the project is to know that it's time that we start thinking about nature and starts taking all the necessary measures which are needed to be taken not only by an individual but also as a group.

TABLE OF CONTENTS

INTRODUCTION

Air pollution stands as a critical global concern with far-reaching implications for both public health and environmental well-being. As our world undergoes rapid industrialization and urbanization, understanding and monitoring air quality data become paramount for informed decision-making and strategic planning.

This study delves into the sources of comprehensive air quality data provided by various agencies of India. This dataset encompasses pollutants like Sulphur dioxide (SO2), Nitrous dioxide (NO2), Respirable Suspended Particulate Matter (RSPM), and Suspended Particulate Matter (SPM). These pollutants, recognized as major determinants of air quality, are associated with various health issues such as respiratory diseases and cardiovascular ailments.

On a global scale, air pollution is not confined to specific regions but is a pervasive issue affecting nations worldwide. The historical evolution of air quality monitoring has witnessed significant technological advancements, enhancing our capacity to collect and analyze data with precision.

The direct impact on public health cannot be overstated, with poor air quality linked to a spectrum of health problems. Respiratory diseases and cardiovascular issues are among the notable consequences, affecting communities at large. Additionally, air pollution extends its consequences to the broader environment, posing threats to ecosystems, biodiversity, and contributing to the global challenge of climate change.

Within the Indian context, government initiatives have been pivotal in addressing air pollution. Various policies, programs, and regulatory measures have been implemented to mitigate the impact of pollutants. The data utilized in this study, sourced from esteemed organizations like the Ministry of Environment, Forests, and the Central Pollution Control Board of India, ensures reliability and credibility.

Air Quality Analysis

As we explore the nuances of Air Quality data, it is crucial to acknowledge emerging challenges and trends. Beyond the conventional factors of industrialization and urbanization, recent developments or events may significantly influence air quality. This study aims to shed light on these intricacies, contributing to a deeper understanding of the dynamic relationship between human activities, air quality, and overall well-being.

PROBLEM STATEMENT

Navigating the world of environmental science and public health, we encounter a puzzling challenge: how do we address the impact of air pollution across different parts of India? To solve this complex puzzle, we need to explore the extensive air quality data. Imagine data about Sulphur Dioxide (SO_2), Nitrous Dioxide (NO_2), Respirable Suspended Particulate Matter (RSPM), and Suspended Particulate Matter (SPM) all mixed up in a web of information. But fear not, we have the key to unlock these secrets. Our mission is to analyze the data carefully, leaving no stone unturned. We'll focus on the patterns of air quality over space and time, guided by the details. We face a dual challenge: the fast growth of industries and cities. How do these busy forces influence the levels of pollutants? And where do we find areas consistently experiencing high or low pollution levels? An adventure awaits as we embark on a quest to decode the mysteries of India's air quality. Are you ready to join the journey?

OBJECTIVES

The Objectives of our project is divided in 3 main parts which are sub-divided in 2 subparts are each which are as follows

A. Local air quality differences:

1: Analyze and compare the concentrations of major air pollutants (SO2, NO2, RSPM, SPM) in different parts of country, assess spatial distribution, identify areas of consistently high or low pollution levels. 2: To investigate the relationship between air quality and geographical factors such as urbanization and industrialization, and to understand the factors that may contribute to regional differences.

B. Time trends in Delhi:

1: To analyze the quality of air in Delhi during the analyzed years, focusing on the identification of significant changes and factors influencing these trends.

2: To investigate the role of various pollution sources (e.g., industrial, residential) in contributing to the observed temporal trends in Delhi.

C. Comprehensive National and Annual Surveys:

1: Conduct a comprehensive analysis of the state over the years, providing insight into how air quality varies by state and identifying any emerging patterns.

2: Use statistical models to identify changes and stability in air quality parameters across countries and years to provide a solid basis for comparative analysis.

METHODOLOGY

Our project had 5 stages which were thoroughly done so that complete information can be provided

1. **Data Collection and Preparation**: The first phase of our analysis involved the acquisition and preparation of the air quality dataset. The dataset, sourced from the various agencies of India, includes information on key pollutants such as Sulphur Dioxide (SO2), Nitrous Dioxide (NO2), Respirable Suspended Particulate Matter (RSPM), and Suspended Particulate Matter (SPM). We utilized the read.csv function to import the data and subsequently examined the information.

2. **Analysis of the Data**: During the initial steps of checking data, we conducted descriptive analysis to gain insights into the distribution and summary statistics of all 4 pollutants. This included generating summary tables, assessing the overall trends, and identifying any errors in the data.

3. **State-Wise Analysis:** We grouped the data by state and calculated the average concentrations of SO2, NO2, RSPM, and SPM. The results were visualized using the ggplot2 package, creating bar plots to highlight the variations in air quality across different states and for comparison which states are highly polluted and which are least polluted.

4. **Temporal Trends in Delhi:** To investigate temporal trends in air quality, particularly in Delhi as Delhi being one of the highest polluted city, not only in India but also in world, we employed time series analysis. We facilitated the creation of line plots to depict the changes in pollutant levels over the years, categorized by different pollution source types.

5. **Comprehensive National and Annual Survey:** The analysis was expanded to have a national perspective. State and year-wise summaries were generated using statistical models, allowing for a better understanding of air quality variations across different regions and over time. The results were presented through bar plots and heat maps to enhance

visual interpretation and to see how the trend is going.

IMPLEMENTATION

The implementation of the air quality analysis was executed in a systematic manner, involving several key steps to derive meaningful insights from the dataset:

1. Data Import and Exploration: The analysis commenced with the importation of the air quality dataset, obtained from the Ministry of Environment, Forests, and the Central Pollution Control Board of India. The `glimpse()` and `summary()` functions were employed to gain a preliminary understanding of the dataset's structure and content.

2. Data Cleaning: A meticulous data cleaning process was undertaken to ensure data integrity. This encompassed handling missing values, standardizing pollutant type categories, and refining the dataset for subsequent analysis. This step aimed to enhance the dataset's consistency and accuracy.

3. Aggregation by State: Utilizing the powerful `dplyr` package, the data was aggregated at the state level. Average values for key pollutants (SO2, NO2, RSPM, SPM) were calculated, providing a concise overview of air quality metrics across different states.

4. Visualization - Bar Plots: The `ggplot2` package facilitated the creation of informative bar plots, visually depicting the average concentrations of each pollutant by state. These visualizations offer an intuitive comparison of pollution levels among different regions.

5. Time-Trend Analysis - Delhi: A specific focus was placed on the capital city, Delhi. The dataset was filtered to isolate Delhi's air quality data, and line plots were generated to illustrate how pollutant levels evolved annually. This temporal analysis aimed to uncover trends and patterns in Delhi's air quality.
6. Visualization - Heat Maps: Leveraging the `reshape2` package, the data underwent transformation for the creation of heat maps. These visualizations showcased the spatial distribution of average pollutant concentrations across states and years, providing a comprehensive understanding of regional variations.

7. Exploratory Data Analysis: Further exploratory data analyses included bar plots and heat maps highlighting pollutant levels over time for specific pollutants. These visualizations offered nuanced insights into the dynamics of air quality parameters.

This step-by-step implementation process ensures the reliability and validity of the analysis, leading to a robust foundation for the subsequent interpretation of air quality trends and patterns in India.

RESULTS AND DISCUSSIONS

In our analysis, we did a series of graphs and visual representation for better understanding of air quality index(AQI) and how Sulphur dioxide (SO2), Nitrous dioxide (NO2), Respirable Suspended Particulate Matter (RSPM), and Suspended Particulate Matter (SPM) are affecting it.

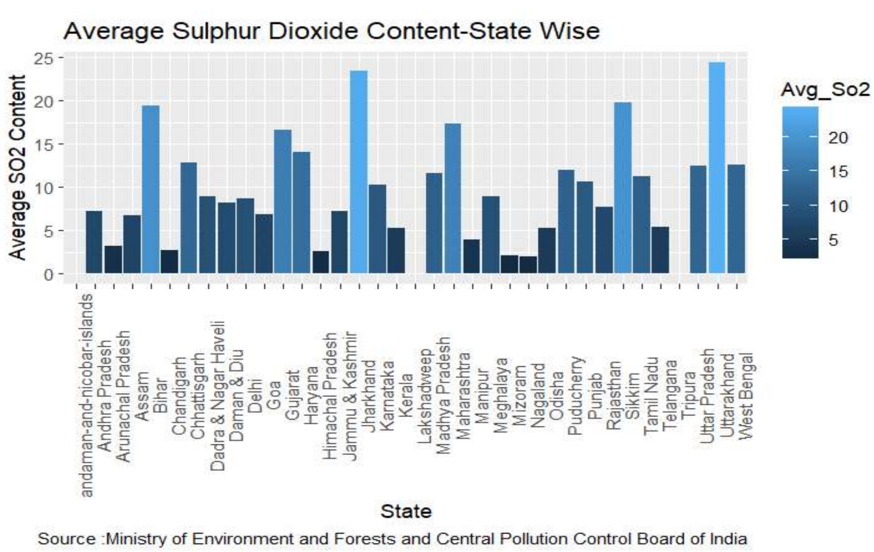
A. State-wise Analysis of Air Quality:

1. Sulphur dioxide content: The state-wise analysis uncovered substantial variations in average Sulphur Dioxide (SO2) content across different regions.

Code:

```
ggplot(by_state_wise,aes(x=state,y=Avg_So2,fill=Avg_So2)) +
  geom_bar(stat="identity") +
  theme(axis.text.x =element_text(angle=90)) +
  ggtitle("Average Sulphur Dioxide Content-State Wise") +
  xlab(label="State") +
  ylab(label="Average SO2 Content") +
  labs(caption="Source :Ministry of Environment and Forests and Central Pollution Control Board of India")
```

Output:



Outcomes:

- Uttarakhand and Jammu and Kashmir displays the highest average SO2 content

- Meghalaya and Mizoram consistently exhibited lower SO2 levels, indicating comparatively cleaner air.

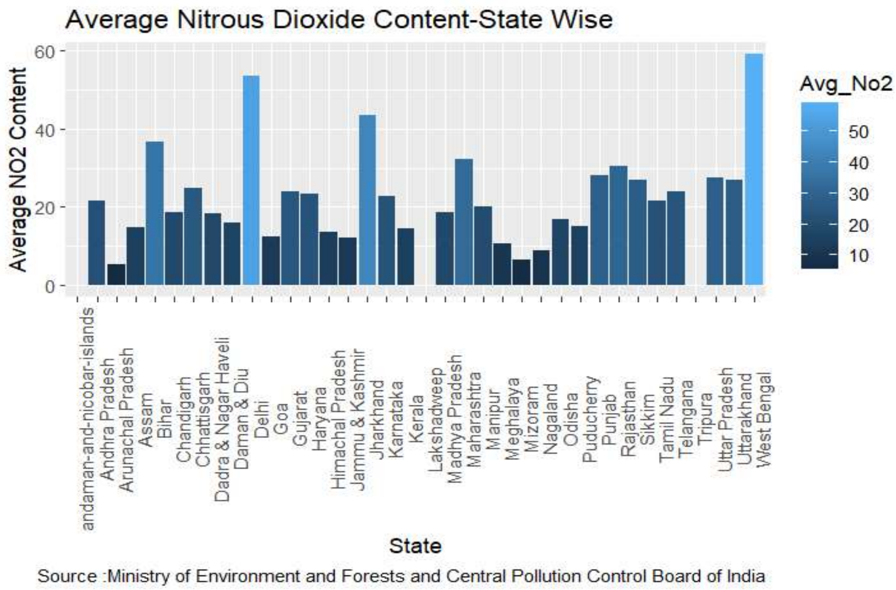
2. **Nitrous dioxide content:** The state-wise analysis uncovered substantial variations in average Nitrous Dioxide (NO2) content across different regions.

Code:

```
ggplot(by_state_wise,aes(x=state,y=Avg_No2,fill=Avg_No2)) +
  geom_bar(stat="identity") +
  theme(axis.text.x =element_text(angle=90)) +
  ggtitle("Average Nitrous Dioxide Content-State Wise") +
  xlab(label="State") +
  ylab(label="Average NO2 Content") +
  labs(caption="Source :Ministry of Environment and Forests and Central Pollution Control Board of India")
```

Output:

Air Quality Analysis



Outcomes:

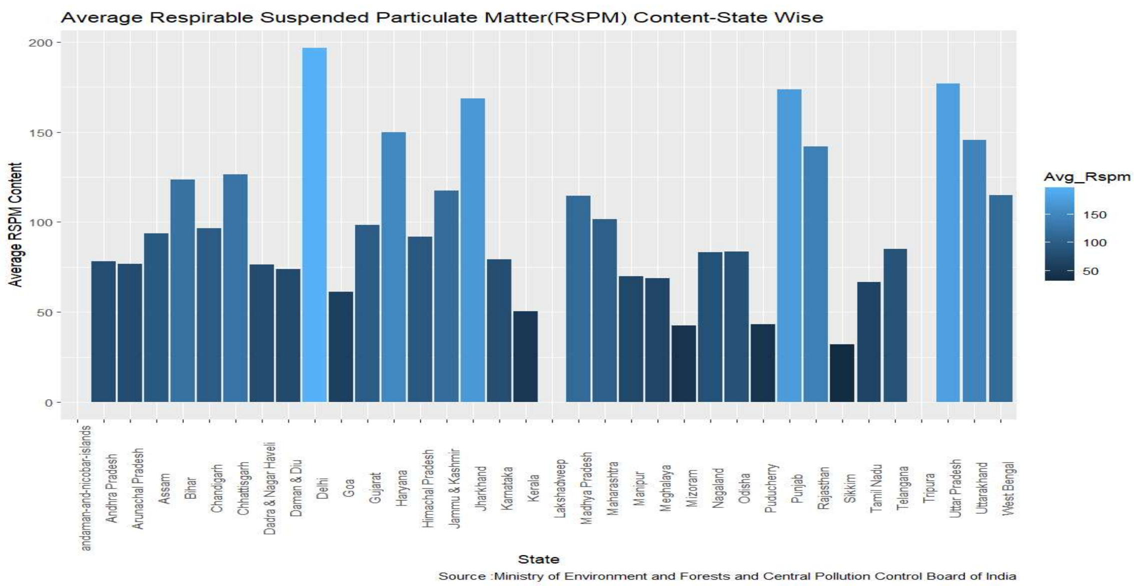
- West Bengal and Delhi displays the highest average SO2 content
- Meghalaya and Andhra Pradesh consistently exhibited lower SO2 levels, indicating comparatively cleaner air.

3. **Respirable Suspended Particulate Matter** content: The state-wise analysis uncovered substantial variations in average Respirable Suspended Particulate Matter (RSPM) content across different regions.

Code:

```r
ggplot(by_state_wise,aes(x=state,y=Avg_Rspm,fill=Avg_Rspm)) +
  geom_bar(stat="identity") +
  theme(axis.text.x =element_text(angle=90)) +
  ggtitle("Average Respirable Suspended Particulate Matter(RSPM) Content-State Wise") +
  xlab(label="State") +
  ylab(label="Average RSPM Content") +
  labs(caption="Source :Ministry of Environment and Forests and Central Pollution Control Board of India")
```

Output:



Outcomes:

- Delhi displays the highest average RSPM content

- Meghalaya and Mizoram consistently exhibited lower RSPM levels.

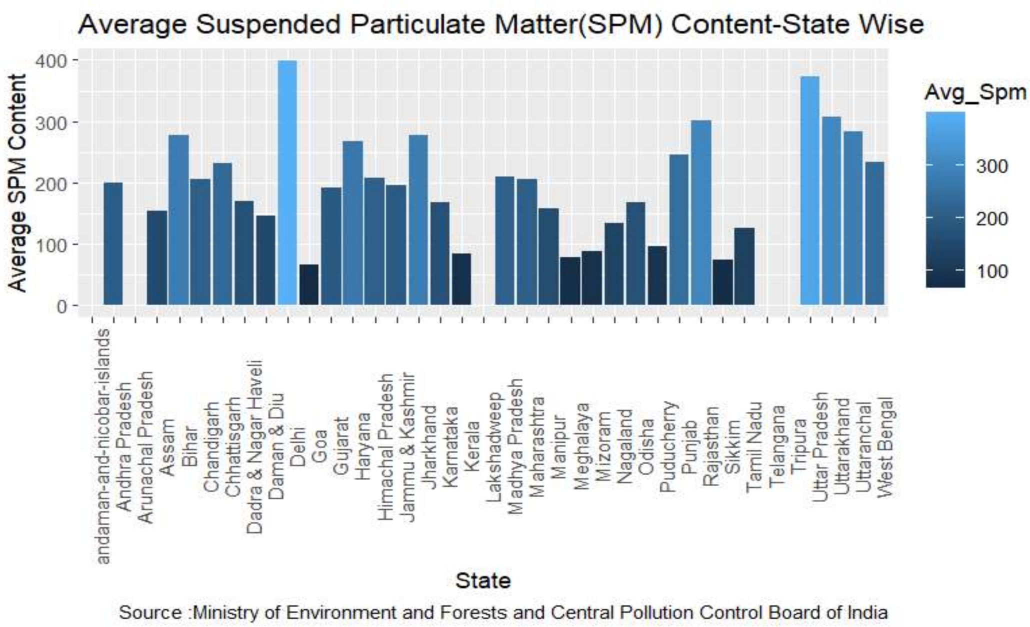
4. **Suspended Particulate Matter** content: The state-wise analysis uncovered substantial variations in average Suspended Particulate Matter (SPRM) content across different regions.

Code:

Air Quality Analysis

```
ggplot(by_state_wise,aes(x=state,y=Avg_Spm,fill=Avg_Spm)) +
  geom_bar(stat="identity") +
  theme(axis.text.x =element_text(angle=90)) +
  ggtitle("Average Suspended Particulate Matter(SPM) Content-State Wise") +
  xlab(label="State") +
  ylab(label="Average SPM Content") +
  labs(caption="Source :Ministry of Environment and Forests and Central Pollution Control Board of India")
```

Output:



Outcomes:

- Delhi displays the highest average SPRM content

- Meghalaya and Mizoram consistently exhibited lower SPRM levels, indicating comparatively cleaner air.

B. Temporal Trends in Delhi

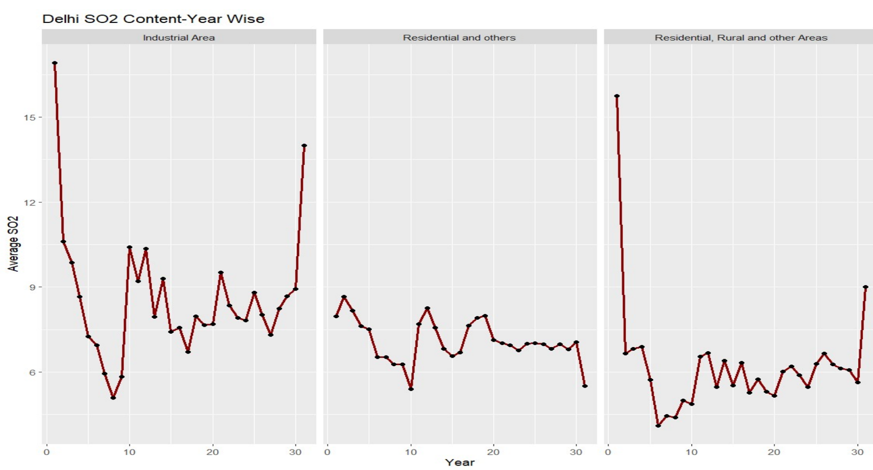
The temporal analysis focused on Delhi, revealing dynamic patterns over the years. Key observations are as follows:

1. Sulphur Dioxide content :

Code:

```
ggplot(Delhi,aes(x=year,y=Avg_So2)) +
  geom_line(size=1,color="darkred") +
  geom_point()+
  facet_wrap(~type) +
  ggtitle("Delhi SO2 Content-Year Wise")+
  xlab("Year") +
  ylab("Average SO2")
```

Output:



Outcomes:

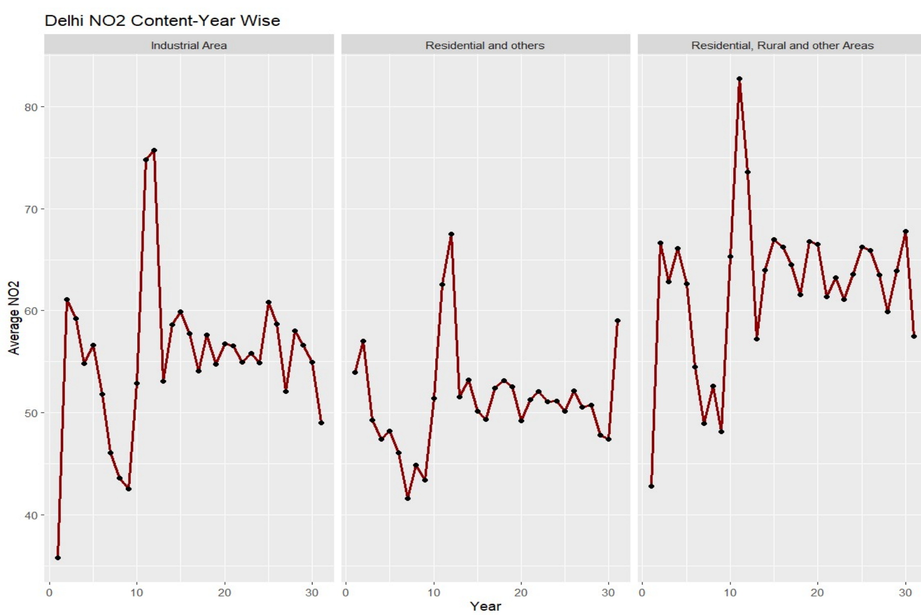
- SO2 level has shown fluctuations over the years.
- Residential areas exhibit less SO2 pollution than others

2. Nitrous Dioxide content :

Code:

```
ggplot(Delhi,aes(x=year,y=Avg_No2)) +
  geom_line(size=1,color="darkred") +
  geom_point()+
  facet_wrap(~type) +
  ggtitle("Delhi NO2 Content-Year Wise")+
  xlab("Year") +
  ylab("Average NO2")
```

Output:



Outcomes:

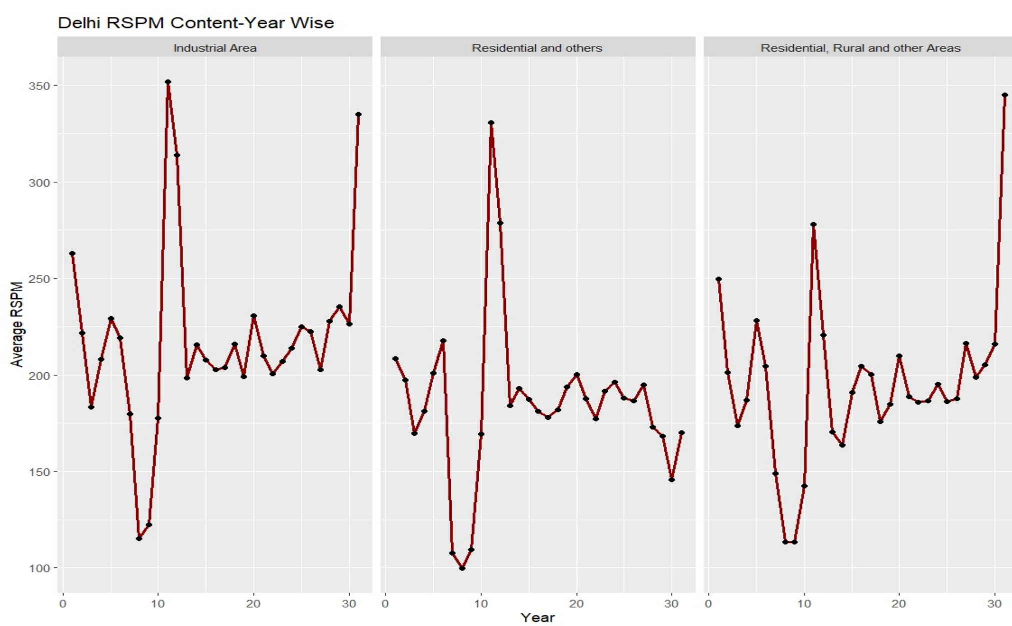
- NO2 level has increased rapidly especially after 2010.
- Industrial areas have maximum proportion of SO2 content.

3. **Respirable Suspended Particulate Matter** content :

Code:

```
ggplot(Delhi,aes(x=year,y=Avg_Rspm)) +
  geom_line(size=1,color="darkred") +
  geom_point()+
  facet_wrap(~type) +
  ggtitle("Delhi RSPM Content-Year Wise")+
  xlab("Year") +
  ylab("Average RSPM")
```

Output:



Outcomes:

- RSPM level is mostly high in Delhi, showcasing a prolonged problem which should be solves as fast as possible.
- Industrial areas have maximum proportion of RSPM content.

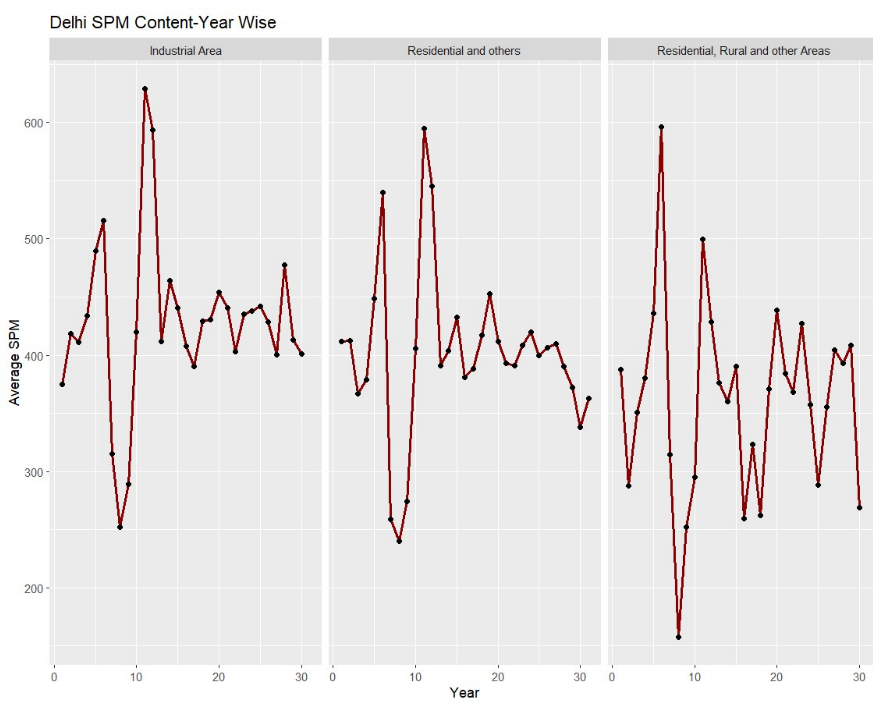
4. **Suspended Particulate Matter** content:

Air Quality Analysis

Code:

```
ggplot(Delhi,aes(x=year,y=Avg_Spm)) +
  geom_line(size=1,color="darkred") +
  geom_point()+
  facet_wrap(~type) +
  ggtitle("Delhi SPM Content-Year Wise")+
  xlab("Year") +
  ylab("Average SPM")
```

Output:



Outcomes:

- SO2 level has shown fluctuations over the years.
- Residential areas exhibit less SO2 pollution than others

C. Heat Map of Different States

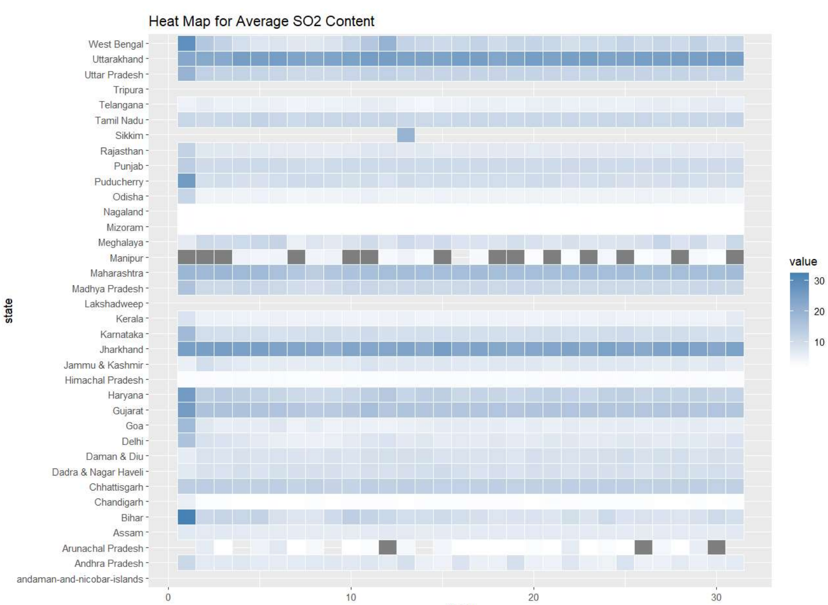
1. Sulphur Dioxide content : The heat map is used to demonstrate average SO2 in different states of India in a given time period.

   The grey darkened box represents the data which was not available at that specific period of time.

   Code:

```
State_Year_Wise.long %>%
  filter(variable=="So2") %>%
  ggplot(aes(x=year,y=state,fill=value)) +
  geom_tile(color="white") +
  scale_fill_gradient(low="white",high="steelblue") +
  theme(legend.position = "right") +
  labs(title="Heat Map for Average SO2 Content")
```

Output:



Outcomes:

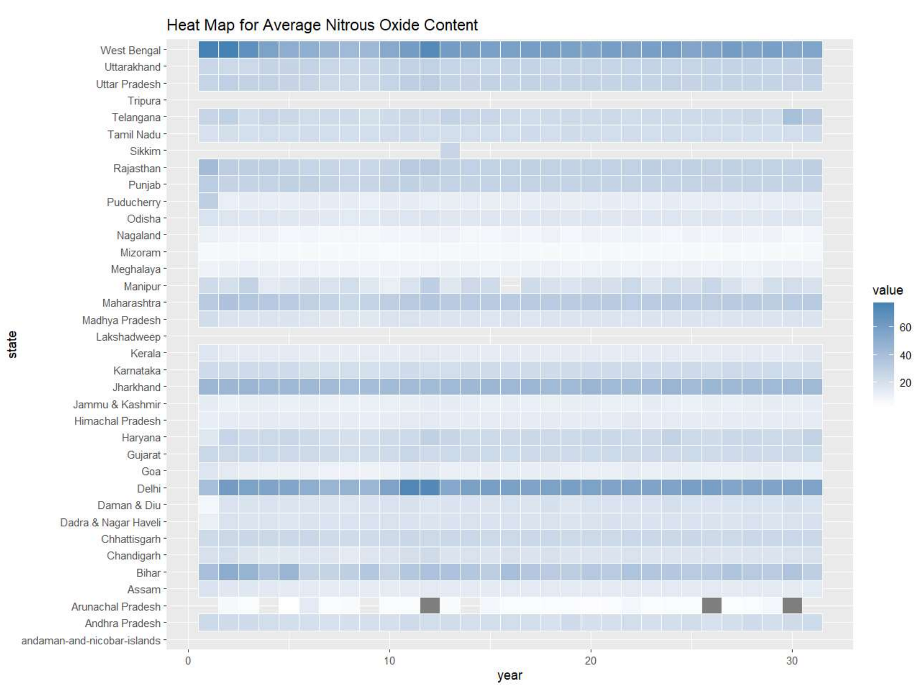
- Patterns in SO2 levels across states over the years are evident.
- Geographical disparities show various sources of pollution for SO2.

2. Nitrous Dioxide content :

Code:

```
State_Year_Wise.long %>%
  filter(variable=="No2") %>%
  ggplot(aes(x=year,y=state,fill=value)) +
  geom_tile(color="white") +
  scale_fill_gradient(low="white",high="steelblue") +
  theme(legend.position = "right") +
  labs(title="Heat Map for Average Nitrous Oxide Content")
```

Output:



Outcomes:

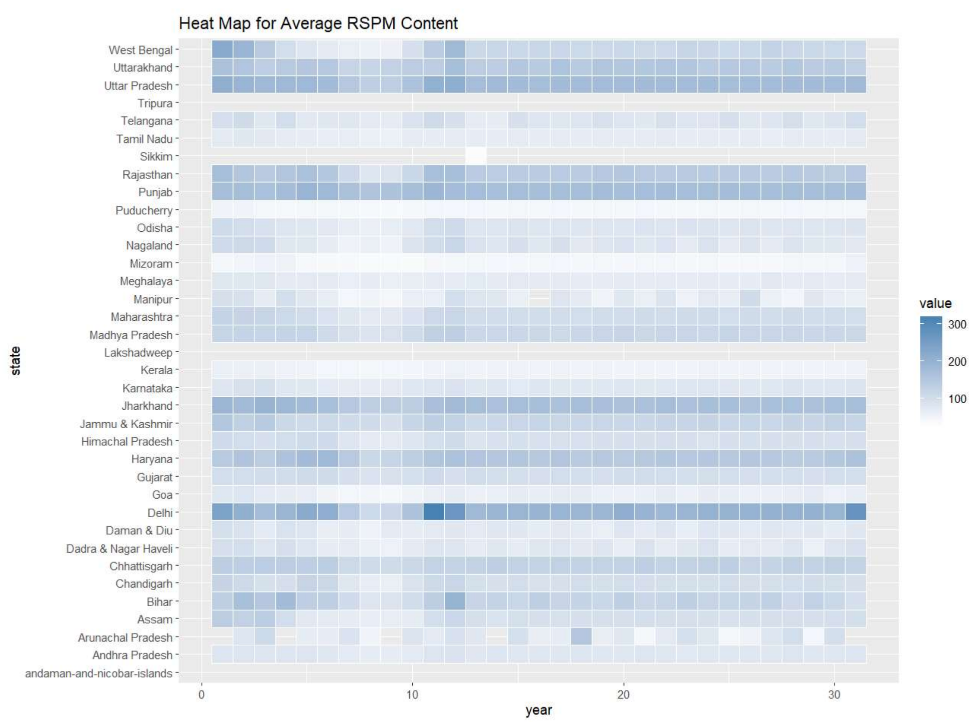
- The heat map reveals temporal patterns in NO2 levels across different states.
- Certain states consistently exhibit higher NO2 levels, highlighting potential geographical pollution.

3. Respirable Suspended Particulate Matter content :

Code:

```
State_Year_Wise.long %>%
  filter(variable=="Rspm") %>%
  ggplot(aes(x=year,y=state,fill=value)) +
  geom_tile(color="white") +
  scale_fill_gradient(low="white",high="steelblue") +
  theme(legend.position = "right") +
  labs(title="Heat Map for Average RSPM Content")
```

Output:



Outcomes:

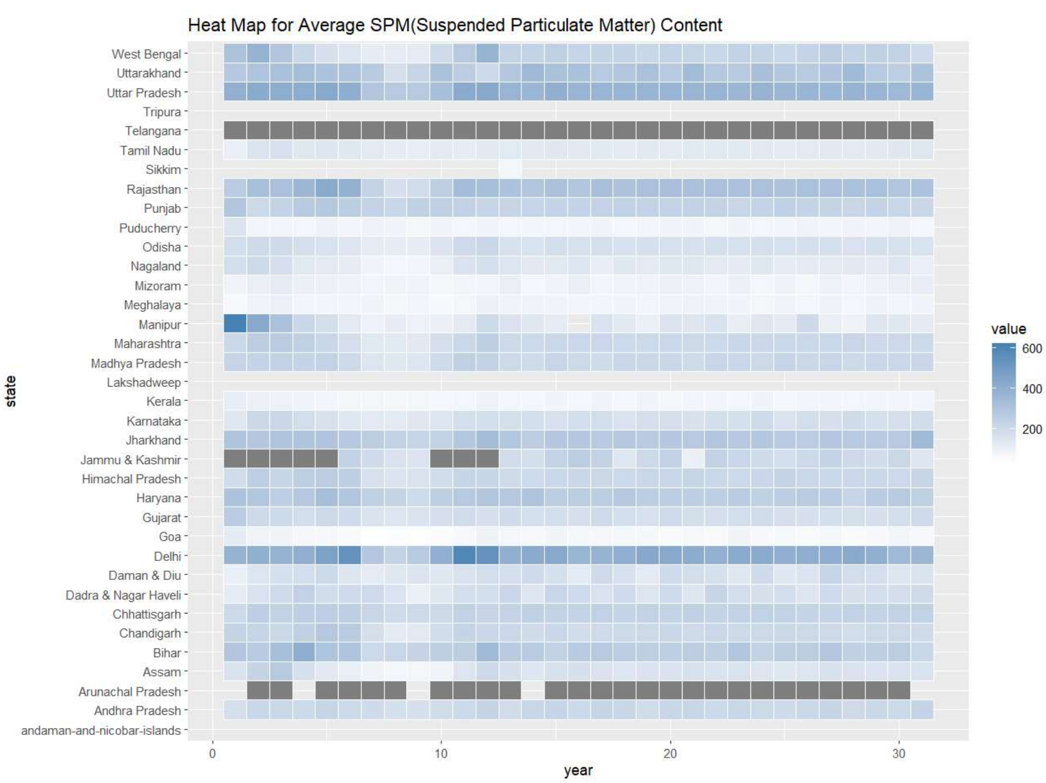
- Some states consistently face higher RSPM levels, indicating persistent air quality challenges.
- Certain regions contribute significantly to elevated RSPM levels.

4. Suspended Particulate Matter content :

Code:

```
State_Year_Wise.long %>%
  filter(variable=="spm") %>%
  ggplot(aes(x=year,y=state,fill=value)) +
  geom_tile(color="white") +
  scale_fill_gradient(low="white",high="steelblue") +
  theme(legend.position = "right") +
  labs(title="Heat Map for Average SPM(Suspended Particulate Matter) Content")
```

Output:



Outcomes:

- States with consistently high SPM levels necessitate targeted interventions.
- Certain states consistently exhibit higher SPM levels, highlighting potential pollution.

D. Melted Data Representation:

1. Faceted Bar Plots: SO2, NO2, RSPM, and SPM by State:

Code:

```
State_Year_Wise.long <-melt(State_Year_Wise,id=c("year","state"),measure=c("So2","No2","Rspm","spm"))
State_Year_Wise.long

ggplot(State_Year_Wise.long,aes(x=year,y=value,color=variable)) +
  geom_bar(stat="identity",position="dodge") +
  facet_wrap(~state)
```

Output:

Outcomes:

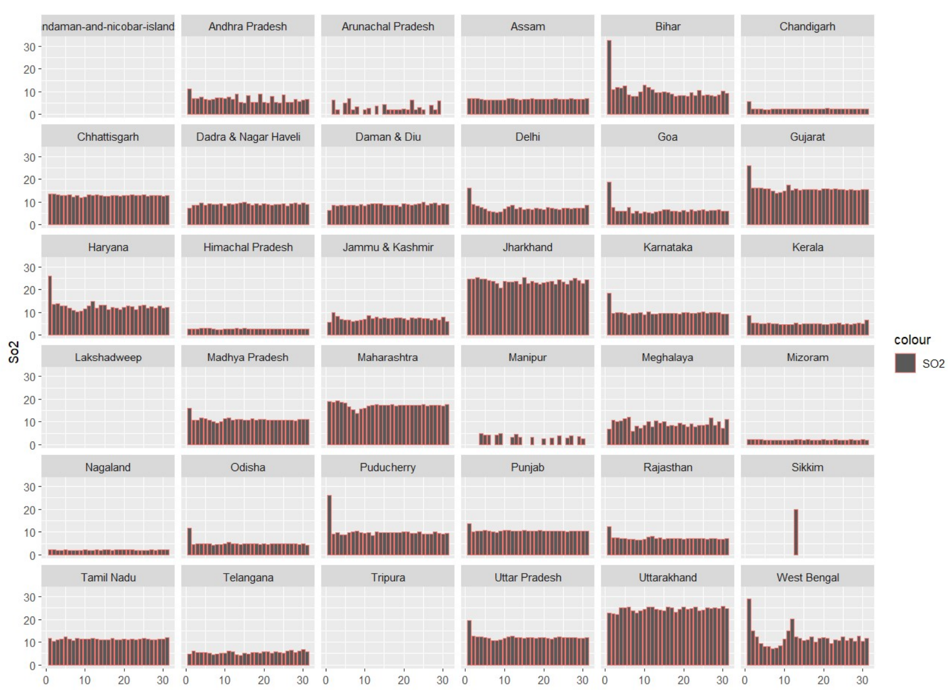
- The heat map reveals temporal patterns in NO2 levels across different states.
- Certain states consistently exhibit higher NO2 levels, highlighting potential geographical pollution.

2. Faceted Bar Plot: SO2 by State:

Code:

```
ggplot(State_Year_Wise,aes(x=year,y=So2,color= "SO2")) +
  geom_bar(stat="identity",position="dodge") +
  facet_wrap(~state)
```

Output:



Outcomes:

- The heat map reveals temporal patterns in NO2 levels across different states.
- Certain states consistently exhibit higher NO2 levels, highlighting potential geographical pollution.

CONCLUSION AND FUTURE SCOPE

In conclusion, this air quality analysis provides valuable insights into the patterns and trends of major pollutants (SO2, NO2, RSPM, SPM) across different states in India. The study navigated through the complexities of a comprehensive dataset from authoritative sources, allowing for a nuanced understanding of air quality dynamics.

The key takeaways include notable variations in pollution levels among states, with Delhi standing out for elevated RSPM and SPM concentrations. Meghalaya and Mizoram emerge as regions with comparatively lower pollution levels. Furthermore, the study unveiled the temporal trends in Delhi's air quality, shedding light on the city's evolving pollution scenario over the analyzed years.

The exploratory data analyses, visualizations, and trend assessments contribute to a comprehensive narrative, highlighting the multifaceted nature of air quality challenges in India. These findings are crucial for informed decision-making, policy formulation, and targeted interventions aimed at mitigating the adverse effects of air pollution.

FUTURE SCOPE

The findings of this study lay the groundwork for a broader scope in understanding and managing air quality in India. The report's final scope extends to:

1. Policy Implications: Delving deeply into the realm of air quality, policymakers face a challenging task: formulating robust strategies and regulations that tackle the pressing issue of regional variations. The air is heavy with the need for change, and it's essential to equip our policymakers with actionable insights that pave the way for effective interve-ntions.
2. Public Health Interventions: In the realm of public health, where lives hang in the balance, it becomes imperative to take swift action to mitigate the deleterious effects of air pollution-related health issues. Let us delve into the depths of this analysis, which serves as a guiding light, illuminating the areas where healthcare resources and awareness campaigns may hold the key to salvation.
3. Long-Term Monitoring: When it comes to ensuring the quality of our air, long-term monitoring plays a crucial role. It's not just about recommending a basic system, but rather urging the adoption of an enhanced and comprehensive monitoring framework. Why is this important? Well, it's simple. Long-term observations offer a window into the effectiveness of our existing regulations and interventions.
4. Interdisciplinary Research: Diving headfirst into the realm of interdisciplinary research, we embark on an exhilarating quest to unravel the intricate web of connections between air quality and a myriad of geographical factors, urbanization, and industrialization. Picture a harmonious symphony of minds coming together, with environmental scientists, health professionals, and policymakers joining forces in a grand collaboration.

REFERENCES

[1] Dataset- Kaggle

[2] https://www.w3schools.com/ (W3Schools)

[3] https://www.geeksforgeeks.org/ (Geeksforgeeks)

[4] Various YouTube Channels were referred