PROJECT REPORT

ON

AIR QUALITY IN HYDERABAD: POLLUTION ANALYSIS.

By

Rohith Kamal Kumar Yenduri
AP21110010303

Under Guidance of

K. Meenakshi Mam, APSSDC
K. Narmada Mam, APSSDC

Prepared in the partial fulfillment of the
Summer Internship Course

AT

Andhra Pradesh State Skill Development Corporation
(APSSDC)
3rd floor,Infosight Building, survey No. 78/2,Near Pathuru Road Junction,NH-16
Service Road, Tadepalli - 522501,Guntur District,Andhra Pradesh

ACKNOWLEDGEMENTS

We would like to express my special thanks to our trainers Meenakshi and Narmada Mam for their time and efforts that they provided throughout the internship. Your useful advice and suggestions were really helpful to us during the project's completion. In this aspect, We are eternally grateful to you.

We are highly indebted to APSSDC management for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

ABSTRACT

The project facilitates the observation of seasonal variations and potential correlations between pollutant levels, leading to a deeper understanding of air quality dynamics in the city. By leveraging this dataset, Many researchers, stakeholders and organizations can make well-informed decisions, implement targeted pollution reduction strategies, and contribute to a greener and healthier future for Hyderabad City. The project offers an invaluable resource for analyzing air quality patterns and trends, fostering a sustainable approach towards urban environmental management.

The dataset is from trusted government sources to analyze the air quality landscape of Hyderabad City.This enables a detailed exploration of air quality patterns and trends. The dataset's historical data spanning different months allows for the observation of seasonal variations and potential correlations between pollutant levels, providing insights into the dynamics of air quality in the city.This project's involves rigorous data processing, exploratory data analysis, and statistical techniques to extract meaningful patterns and trends from the dataset. The results reveal significant variations in pollutant concentrations over time, shedding light on the impact of various factors such as pollutants PM10, SO2, NOx, PM2.5, Ammonia, O3, CO, Benzene.The analysis which was made adheres to standard terminology, ensuring accessibility to a wide range of readers, and serves as a crucial reference point for further research and decision-making in urban environmental management.

System Requirements

- PC(Minimum 4 GB Ram and 256 SSD)
- OS (Windows , Mac)
- Python(Version 3.10)
- Jupyter Notebook/JupyterLab (IDE)

INTRODUCTION

The project aims to address the pressing issue of air quality in Hyderabad City, which has been adversely affected by rapid urbanization and industrialization. The rising pollution levels pose significant health and environmental risks to the residents. The primary goal is to conduct a comprehensive analysis of historical air quality data sourced from reliable government sources.This project leverages a dataset consisting of 72 rows and 11 columns, containing essential parameters related to air quality, such as PM10, PM2.5, SO2, NOx, Ammonia, O3, CO, Benzene, and the Air Quality Index (AQI). The data spans multiple months, allowing for a comprehensive understanding of seasonal variations and long term trends.

The deteriorating air quality in Hyderabad City is a major concern, with pollutants such as PM10, PM2.5, SO2, NOx, and Benzene exceeding safe levels. This project is driven by the urgent need to improve air quality through evidence-based decision-making and targeted pollution reduction strategies. By analyzing the extensive dataset, we aim to identify pollution hotspots, explore seasonal variations, and establish potential correlations among different pollutants.

Uses of Data Analysis library

Data analysis in libraries has various uses that can significantly improve library services, operations, and user experiences. Some of the key uses of data analysis in libraries include User Behavior Analysis, Collection Development, Resource.Utilization, Service Improvement, Predictive Analysis, Space Management, Cost Management, Decision Making, Marketing and Outreach.Here are some libraries used in this project

Numpy : Python library for numerical computing and array manipulation.

Pandas: A data manipulation and analysis library that offers easy-to-use data structures and data analysis tools, making it efficient for data preprocessing and cleaning tasks.

Matplotlib: A plotting library that produces publication-quality visualizations, allowing users to create a wide variety of plots and charts.

Seaborn: A statistical data visualization library that builds on top of Matplotlib, providing additional functionality and more visually appealing plots.

SQL : A database library used for storing AQI values

Requests: Used for making HTTP requests to a specified URL.

bs4 BeautifulSoup: To scrape information from web pages. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree.

Descriptive Statistics: To summarize and describe the main characteristics of a dataset, such as measures of central tendency (mean, median, mode) and measures of dispersion (standard deviation, range).

Data Visualization: Data visualization techniques, such as histograms, bar plots, scatter plots, and box plots, are used to visually represent the distribution, relationships, and patterns within the data.

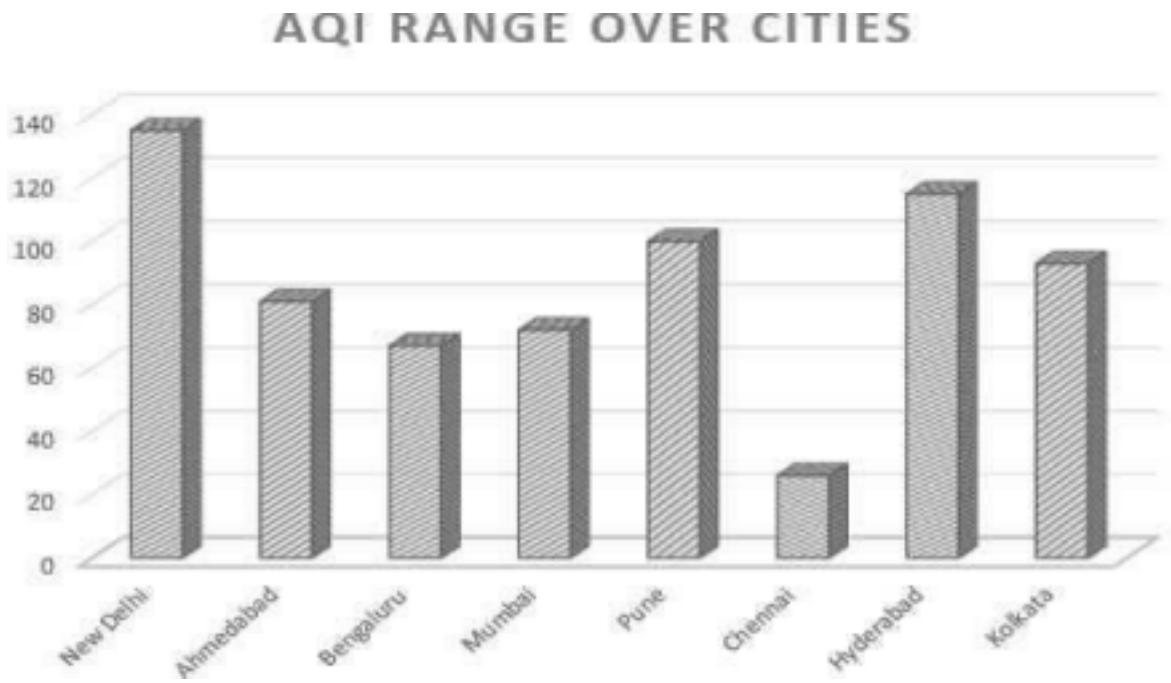
Correlation Analysis: Correlation analysis examines the relationship between variables in a dataset. It helps identify the strength and direction of the relationship using correlation coefficients, such as Pearson's correlation coefficient.

Missing Data Handling Data Transformation and Filtering: Concepts used in exploratory data analysis to address missing values, standardize data, and extract specific subsets, respectively.

PURPOSE OF THE PROJECT

The main purpose of this project is to examine how air pollution refers to the presence of harmful substances in the air, primarily resulting from human activities and natural processes. The main causes of air pollution, includes industrial emissions, vehicular exhaust, burning of fossil fuels, agricultural activities, and dust from construction sites. These activities release various pollutants into the atmosphere.

Pollution levels in Metros of India are rising to alarming levels in recent decades. This issue needs to be addressed immediately because it is hazardous to people's health. The present work is focused to highlight the major air pollutants in various areas of Hyderabad using publicly available data at Kaggle.com. By consolidating more air pollutants into fewer factors, this study's key objective is to reduce the complexity of air pollution. This helps to understand the interdependency of air pollutants.



The decline in environmental quality as a result of pollution is evidenced by a decline in agricultural yield, a loss of biological diversity and excessive amounts of toxic and recalcitrant contaminants in the atmosphere, water reservoirs and in food supplies.Pollution becomes problematic when the use or contamination of natural resources occurs at a higher rate than nature’s capacity to restore itself.

OBJECTIVE AND SCOPE OF THE PROJECT

The primary objective of this project is to conduct a detailed analysis of air pollution data in Hyderabad. The specific objectives include

- Analyzing the concentrations of major air pollutants such as particulate matter (PM10, PM2.5), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), and carbon monoxide (CO).
- Identifying temporal patterns and trends in air pollution levels over months, seasons, and years. Exploring spatial distribution and identifying pollution hotspots across different regions within the city.
- Investigating the potential correlation between meteorological factors (temperature, humidity, wind speed) and air pollution levels. ● Providing data-driven insights and recommendations for policymakers and stakeholders to implement effective air quality management strategies.

The scope of the air pollution analysis project on Hyderabad's dataset is extensive and multifaceted because it includes various steps through the process like Data Acquisition, Data Preprocessing and Cleaning, Data Exploration and Visualization, Data Analysis and Modeling, Interpretation of Results Documentation and Reporting.

This project also has some limitations that may arise during the analysis. Some limitations could include data gaps, inconsistencies, or the inability to account for all potential contributing factors to air pollution.

The project may identify areas for future research or additional data collection to gain deeper insights into specific aspects of air pollution in Hyderabad, paving the way for continuous improvement and more targeted interventions.

LITERATURE REVIEW

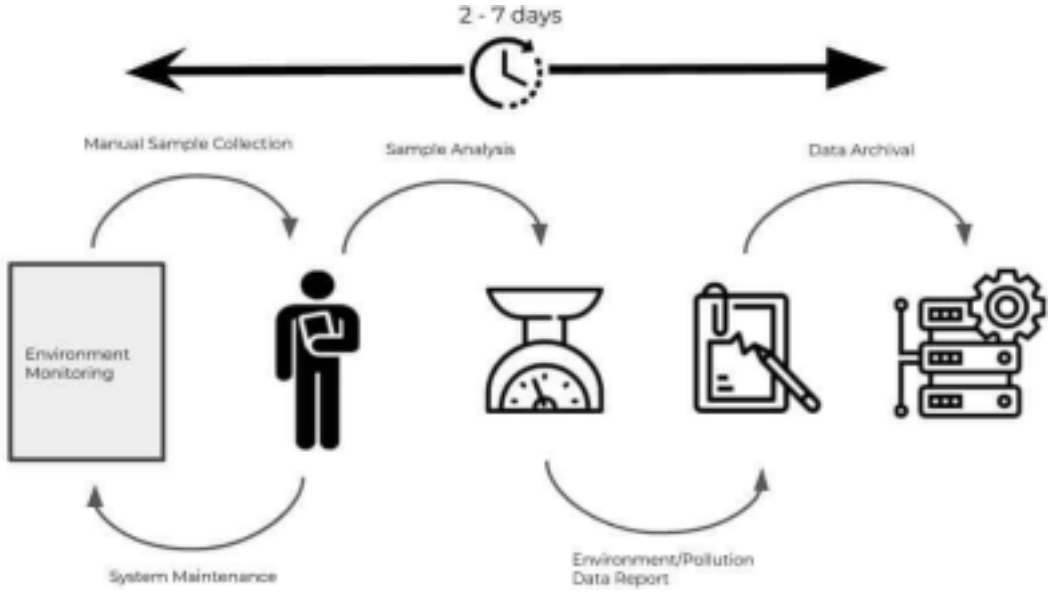
Several studies have explored the trends and patterns of air pollution in urban centers. A study showed analysis of air quality data in Hyderabad over a decade, highlighting the increasing concentrations of particulate matter (PM10 and PM2.5) and nitrogen dioxide (NO2) due to growing vehicular emissions and industrial activities.

The study emphasized the need for targeted pollution control measures to curb air pollution levels in the city.Understanding the seasonal variations of air pollutants is crucial for designing effective pollution control strategies. Air pollution has adverse effects on public health, leading to various respiratory and cardiovascular diseasesVarious data analysis techniques and machine learning algorithms have been applied to air pollution datasets for prediction and forecasting.

The complexity and multidimensional nature of air pollution analysis and management. The reviewed studies underscore the urgent need for effective policies, comprehensive monitoring, and innovative approaches to mitigate air pollution and safeguard public health and the environment. This project seeks to contribute to the growing body of knowledge on air pollution analysis and provide data-driven insights to support evidence-based air quality management strategies.

AIR QUALITY MONITORING AND DATA COLLECTION

Air quality monitoring and data collection are fundamental steps that provide the foundation for the subsequent data analysis and interpretation. The project team must collaborate with relevant authorities, such as environmental agencies or research institutions, to access existing air quality monitoring data and establish data-sharing agreements.It involves assessing the level of pollutants in the atmosphere and understanding the spatial and temporal variations of air pollution. It helps in identifying pollution hotspots, determining the sources of pollutants, and evaluating the effectiveness of pollution control measures. Accurate and up-to-date air quality data are indispensable for evidence-based decision-making and formulating effective air quality management strategies.



Data can be collected through Monitoring stations, Meteorological stations, Remote Sensing or Real time monitoring.data accuracy and consistency is of utmost importance to ensure the validity of the project's findings and the reliability of the recommendations made for air quality management.

In conclusion, air quality monitoring and data collection form the backbone of this project. They are vital for obtaining reliable data on air pollution levels in Hyderabad, enabling comprehensive analysis, and offering valuable insights for policymakers and stakeholders to implement effective air quality management strategies.

HEALTH IMPACTS OF AIR POLLUTION IN URBAN CITIES

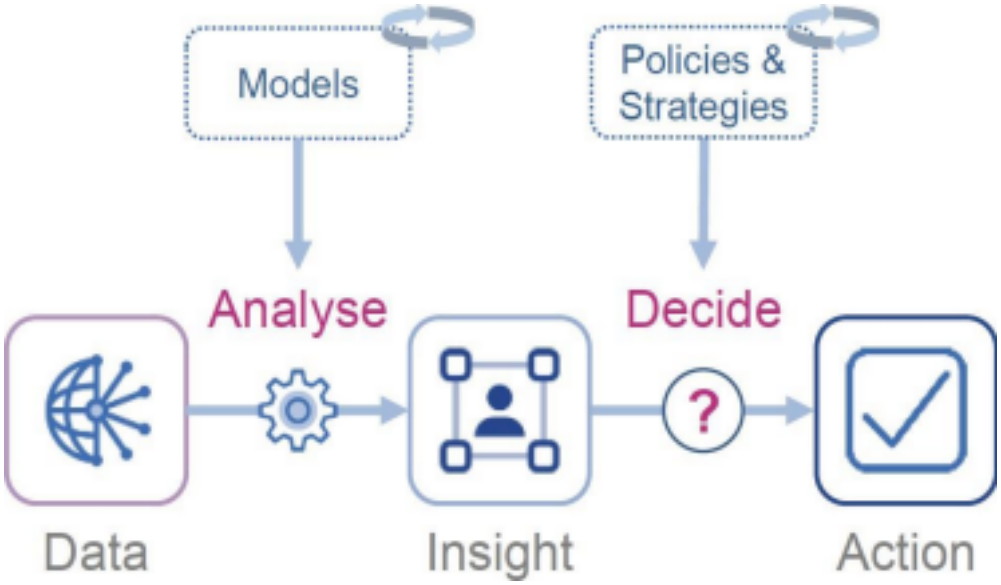
Air pollution is a persistent and pervasive environmental challenge that poses significant health risks to populations living in urban cities worldwide. With rapid industrialization, increasing vehicular emissions, and growing urbanization, urban areas have become hotspots for air pollution, resulting in severe health implications for millions of people.Urban outdoor air pollution is estimated to cause 1.3 million deaths worldwide per year. Those living in middle-income countries disproportionately experience this burden.

- One of the most evident and well-documented health effects of air pollution is its impact on respiratory health. Fine particulate matter (PM2.5), nitrogen dioxide (NO2), sulfur dioxide (SO2), and other pollutants are known to penetrate deep into the respiratory system, causing inflammation and irritation of the airways
- Air pollution is not limited to affecting the respiratory system; it also has detrimental effects on the cardiovascular system.
- Particulate matter and other pollutants can enter the bloodstream, triggering systemic inflammation and oxidative stress. This, in turn, contributes to the development of cardiovascular diseases such as heart attacks, strokes, and hypertension
- The particulate matter has been found in brain tissues, and evidence suggests that air pollution may contribute to cognitive decline and an increased risk of neurodegenerative diseases, such as Alzheimer's and Parkinson's disease. As urban cities continue to grapple with air pollution, the emerging concern for neurological health adds a new dimension to the already complex public health challenge.
- Long-term exposure to pollutants has been associated with increased stress, anxiety, and depression. As urban dwellers contend with both physical and mental health challenges arising from air pollution, the comprehensive impact on overall well-being becomes evident.

To tackle these health risks, comprehensive air quality management strategies are imperative. Implementing cleaner transportation systems, transitioning to sustainable energy sources, promoting green urban planning, and raising public awareness are vital steps towards ensuring a healthier, more sustainable urban environment for present and future generations.

DATA ANALYSIS OVERVIEW

Data analysis is the process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.The project used data analysis techniques to achieve the final outcomes with the following steps



1.Data Collection:

The first step involves collecting air quality data from various monitoring stations in Hyderabad. Hourly measurements of major air pollutants, including particulate matter (PM10, PM2.5), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), and carbon monoxide (CO), will be obtained. Additionally, meteorological data such as temperature, humidity, wind speed, and wind direction will be collected to analyze their correlation with air pollution levels.

2. Data Preprocessing:

To ensure data quality, data preprocessing will be performed. This step involves handling missing values, removing duplicates, and addressing any inconsistencies in the dataset. Outliers, if present, will be carefully analyzed to determine their validity and potential impact on the analysis. Data cleaning will be conducted using Python programming and libraries like Pandas and NumPy.

3. Exploratory Data Analysis (EDA):

EDA aims to gain insights into the dataset and identify patterns, trends, and correlations. Data visualization techniques, including histograms, box plots, scatter plots, and heatmaps, will be employed to visualize the distribution of pollutants, identify seasonal variations, and explore the relationships between

pollutants and meteorological factors. EDA will be conducted using Python libraries like Matplotlib and Seaborn.

4. Statistical Analysis:

Statistical analysis will be conducted to examine the relationships between air pollutants and meteorological variables. Correlation analysis will help determine if certain meteorological conditions influence air pollution levels.

5. Data Visualization:

Data-driven insights will be presented using clear and informative visualizations, making the analysis easily understandable to various stakeholders. Data visualizations will be included in the form of graphs and charts.

6. Data Interpretation and Recommendations:

The results of the analysis will be carefully interpreted and translated into actionable insights. Based on the findings, data-driven recommendations will be formulated for policymakers and stakeholders to implement effective air quality management strategies and pollution control measures in Hyderabad.

Using the above steps we can provide valuable insights into air pollution levels and their implications, contributing to evidence-based decision-making and efforts to improve air quality in the city.

DESIGN OF THE SOLUTION

The dataset which was taken consists of 72 rows and 11 columns of data. Columns include pollutant parameters such as PM2.5, PM10, NO2, SO2, CO, O3, etc., along with the Air Quality Index (AQI) for each observation.The dataset spans different months,enabling the exploration of seasonal variations and potential correlations between pollutant levels.The solution includes data preprocessing,exploratory data analysis, and correlation analysis to understand air quality patterns. Seasonal analysis will examine variations in air quality across different seasons in Hyderabad.

The initial step involved data preprocessing to ensure data quality. Missing values were handled by filling them with appropriate measures, and duplicates or outliers were eliminated. The data was formatted to ensure consistency in data types. EDA was conducted to gain insights into the air quality dataset. The distribution of pollutant parameters was visualized through histograms and box plots. The analysis revealed the central tendency, spread, and skewness of the data.A heatmap was generated to visualize the correlation matrix, identifying strong positive or negative correlations.Seasonal trends in pollutant levels and AQI were analyzed and compared.

This analysis provided insights into the variation in air quality during different times of the year.Visualizations, including line plots, bar charts, and heatmaps, were used to present the findings clearly. The visual representations facilitated a better understanding of air quality patterns and seasonal variations.

OUTCOMES

- Through data analysis, the project will help identify the primary sources and contributors to air pollution in Hyderabad. This information will be crucial for authorities and policymakers to develop targeted measures aimed at controlling emissions from these sources.
- Armed with insights from the data analysis, authorities can implement specific and effective measures to reduce pollution levels. This could include stricter regulations for industries, promoting cleaner technologies, and incentivizing sustainable practices.
- By implementing targeted measures based on the data analysis, the project aims to lead to a significant improvement in air quality in Hyderabad. This, in turn, will contribute to a healthier living environment for the city's inhabitants.
- The project will create accessible and comprehensible data visualizations that will help raise public awareness about air quality conditions in different localities. This knowledge will empower citizens to make informed decisions about their outdoor activities and take necessary precautions to protect their health during periods of high pollution.
- A data-driven approach to AQI management will provide valuable inputs for sustainable urban planning. By integrating air quality considerations into city development plans, Hyderabad can foster a more environmentally friendly and sustainable urban environment. ● The project's successful implementation is expected to have a positive impact on the environment, reducing greenhouse gas emissions and other harmful pollutants that contribute to climate change and environmental degradation
- As air quality improves over time, the long-term health of Hyderabad's residents is expected to see positive effects. Reduced exposure to pollutants can lead to lower rates of respiratory and cardiovascular diseases, improving the overall well-being of the population.
- The project will establish a precedent for data-driven policy making, demonstrating the importance of evidence-based decision-making in addressing complex environmental issues.
- The approach and methodologies used in this project can serve as a model for other cities facing air pollution challenges, facilitating the

Advantages

This project plays a pivotal role in understanding and mitigating the impact of air pollution on the city's inhabitants and environment. Firstly, data analysis provides valuable insights into the trends and patterns of air pollution, enabling authorities to identify the main sources and contributors to poor air quality. By pinpointing these sources, targeted measures can be implemented to control emissions and reduce pollution levels, fostering a healthier living environment for residents. By analyzing the data we can develop more precautions for improving air quality in the city. Building insightful dashboards can help us to get a glimpse about trends that have been happening from the past few years.By pinpointing these sources, targeted measures can be implemented to control emissions and reduce pollution levels, fostering a healthier living environment for residents.

In addition to aiding authorities and policymakers, data analysis of AQI in Hyderabad also benefits the general public. With accessible and comprehensible data visualizations, citizens can gain a clear understanding of air quality conditions in their localities. Armed with this information, they can take proactive steps to safeguard their health, such as modifying outdoor activities during high pollution days or using face masks to reduce exposure to harmful pollutants

In the long run, a data-driven approach to AQI management can contribute to sustainable urban planning, promoting the adoption of cleaner technologies, and enhancing Hyderabad's overall environmental quality.

CONCLUSIONS

The project's ultimate objective is to utilize data-driven insights to improve air quality and promote sustainable practices in Hyderabad City. By addressing air pollution, we aim to enhance the overall well-being of residents, foster environmental stewardship, and set an example for other cities to follow towards achieving cleaner and greener urban environments.

This approach involves collaborating with government agencies, research institutions, local communities, and private sectors to foster a culture of environmental stewardship and sustainable practices. Through the dissemination of information, awareness campaigns, and education initiatives, we seek to empower residents and businesses to actively participate in the city's transformation towards cleaner air and a greener future. In addition to improving public health, our efforts aim to enhance the overall well-being of the city's inhabitants, creating a healthier and more livable urban environment. By setting a benchmark for other cities to follow, we hope to inspire a domino effect where environmental consciousness becomes an integral part of urban planning and development across the nation and beyond. Ultimately, the success of this project relies on collaboration, innovation, and an unwavering commitment to a sustainable future. By combining our collective efforts, we can realize the vision of a cleaner, healthier, and more vibrant Hyderabad City, showcasing the power of data-driven solutions in shaping a brighter tomorrow for generations to come.

REFERENCES

https://aqicn.org/city/hyderabad/
https://www.iqair.com/in-en/india/telangana/hyderabad https://www.news18.com/aqi-india/hyderabad/