TITLE: GLOBAL AIR POLLUTION ANALYSIS

Project Guide:

K.Ramakrishna(professor)

Presented By:

A.Deepshika(160122737004)
G.vyshnavi(160122737005)
M.Kavyasri(160122737013)

ABSTRACT

- *This data analysis project focuses on assessing global air quality using a comprehensive dataset that includes Air Quality Index (AQI) values from various countries and cities worldwide. The dataset encompasses AQI readings and corresponding categories, representing the level of air pollution from different regions. The primary objectives of the analysis are to identify geographical hotspots of air pollution and to understand seasonal variations in air quality. Through this analysis, we aim to pinpoint regions with consistently poor air quality and to uncover patterns in AQI fluctuations throughout the year. The outcomes will provide valuable insights into global air pollution trends, enabling stakeholders to prioritize interventions and develop targeted strategies for improving air quality* and public health.

Problem Statement,Objectives,Outcomes

- **Problem statement**: *This project aims to analyse the global air pollution dataset, focusing on AQI values across different countries and cities.*

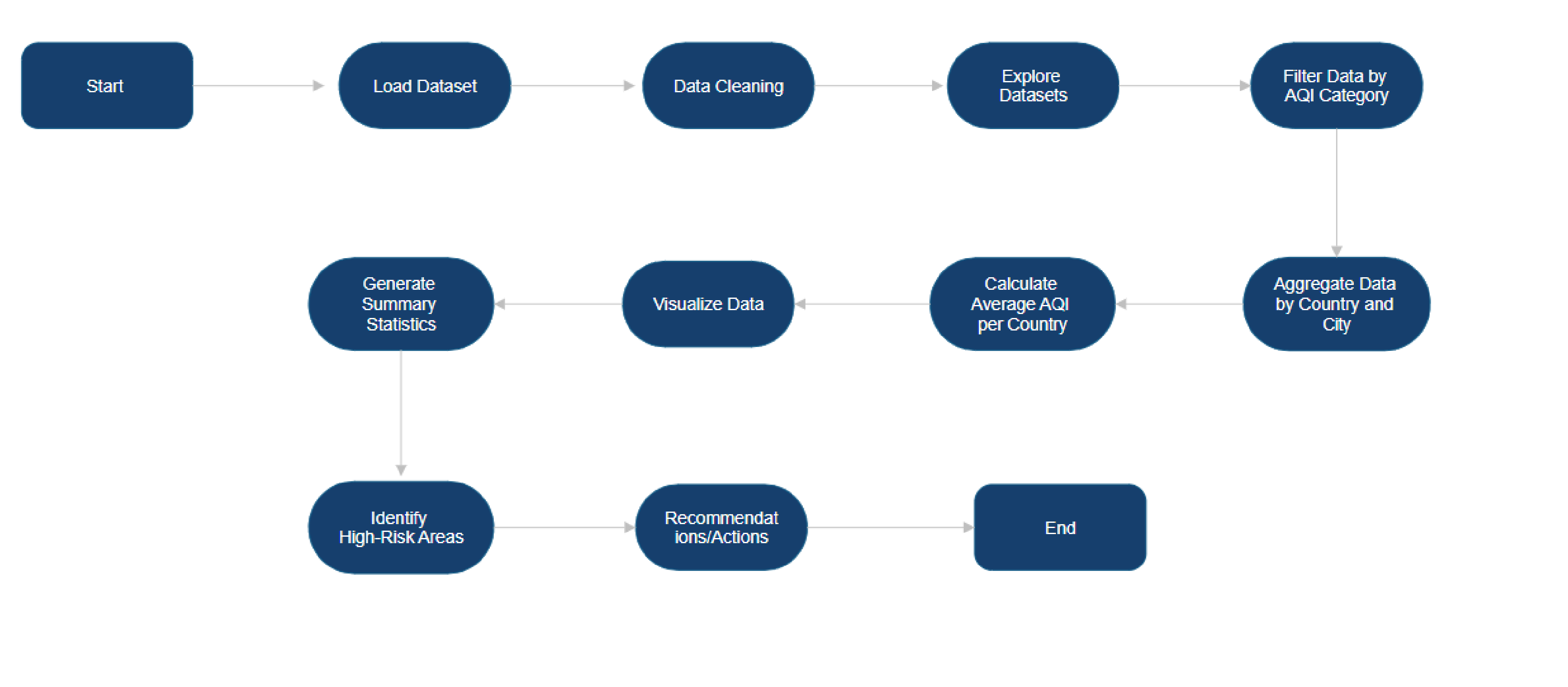
- **Objectives:**
- ***1. Regional Analysis of AQI Levels:***
- *• To identify and compare the countries and cities with the highest and lowest average AQI values.*
- *• To determine the geographical regions that consistently experience poor air quality based on AQI categories (e.g., Good, Moderate, Unhealthy).*
- ***2 . Temporal Trends and Seasonal Variations:***
- *To analyse the temporal trends of AQI values over a specified time period to identify any patterns or fluctuations.*
- *To investigate seasonal variations in air quality, understanding how different seasons impact AQI levels across various countries and cities.*

- **Outcomes:**
- ***1. Geographical Hotspots of Air Pollution:***
- *Identification of countries and cities that consistently report high AQI values, pinpointing geographical areas as "hotspots" for air pollution.*
- *Visualization maps highlighting these hotspots to provide a clear picture of areas most affected by poor air quality.*
- ***2. Seasonal Air Quality Profiles:***
- *Creation of seasonal AQI profiles for different regions, illustrating how air quality varies throughout the year.*
- *Insights into the seasons where certain regions experience heightened air pollution levels, potentially correlating with climatic conditions or seasonal activities.*

Technology Stack

- Python
- Libraries: NumPy, Pandas, Matplotlib
- IDE: Jupyter Notebook, Google Colab
- Data Source: Global Air Pollution Analysis Dataset
- Version Control: Git, GitHub
- Documentation: Jupyter Notebook or report

System Design


Start
Load Dataset
Data Cleaning
Explore Datasets
Filter Data by AQI Category
Aggregate Data by Country and City
Calculate Average AQI per Country
Visualize Data
Generate Summary Statistics
Identify High-Risk Areas
Recommendations/Actions
End

DATASET,MATHEMATICAL MODEL FOR DATA ANALYSIS

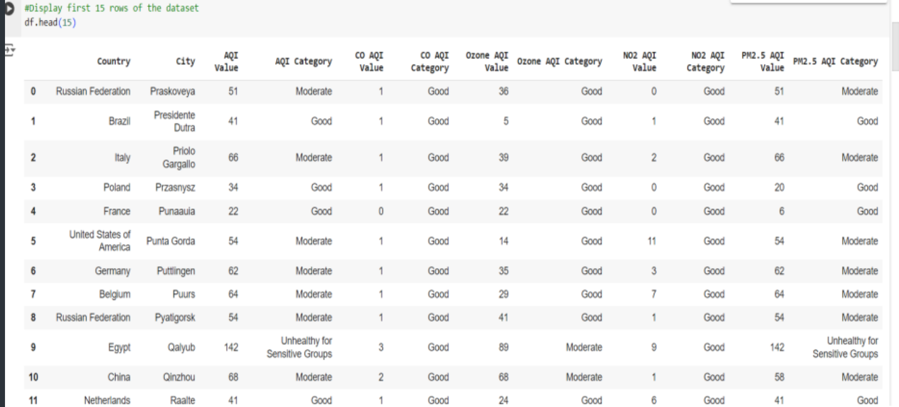
Dataset:

The global air pollution dataset is an invaluable resource for understanding the spatial and temporal distribution of air quality across the world. Compiled from various sources including satellite observations, ground-based monitoring stations, and atmospheric models, this dataset provides comprehensive data on pollutants such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and carbon monoxide (CO). The dataset enables researchers, policymakers, and public health officials to analyze trends, identify pollution hotspots, and assess the effectiveness of air quality regulations. By offering insights into the correlation between air pollution and health outcomes, economic impacts, and environmental degradation, the global air pollution dataset plays a crucial role in driving international efforts to mitigate air pollution and improve public health and environmental quality.

Mathematical models for data analysis:

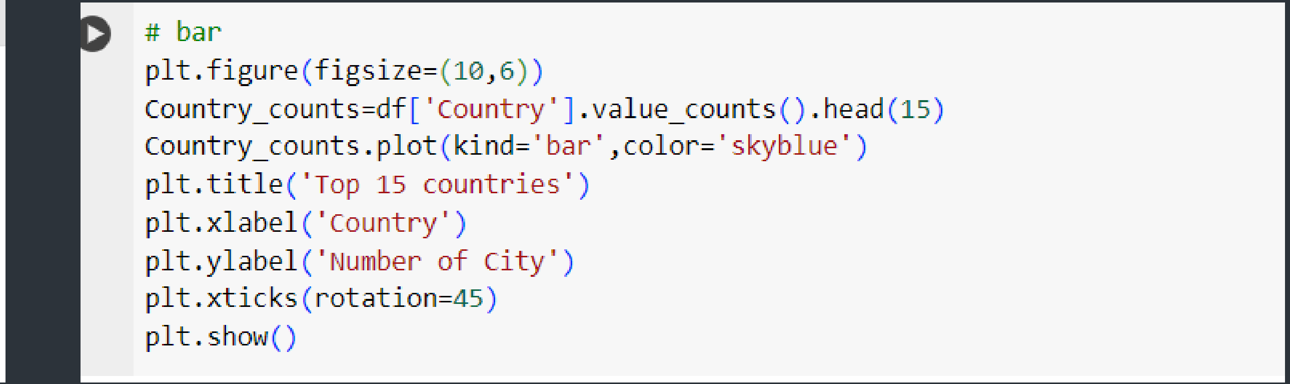
Mathematical models are pivotal in analyzing global air pollution data, providing crucial insights into the complex dynamics of pollutant distribution and transformation. Key models include dispersion models, which simulate the spread of pollutants from various sources based on meteorological data; chemical transport models (CTMs), like CMAQ and GEOSChem, which integrate atmospheric chemistry with transport processes to predict pollutant levels and interactions; and statistical models that utilize regression analysis, machine learning, and time-series analysis to identify trends and make predictions from large datasets. Inversion models help estimate emission sources by using observed pollutant concentrations, while climate models study the interactions between air quality and climate change. These models collectively enable researchers to assess pollution sources, predict future trends, and evaluate the effectiveness of mitigation strategies, ultimately guiding informed policy decisions to improve global air quality.

IMPLEMENTATION

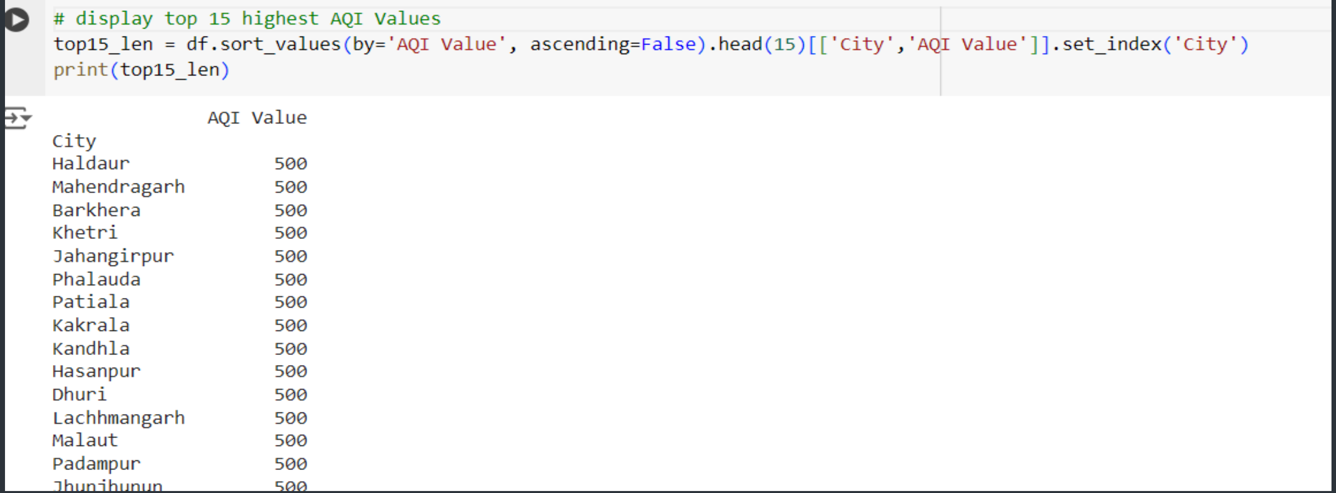


```python
#Display first 15 rows of the dataset
df.head(15)
```

	Country	City	AQI Value	AQI Category	CO AQI Value	CO AQI Category	Ozone AQI Value	Ozone AQI Category	NO2 AQI Value	NO2 AQI Category	PM2.5 AQI Value	PM2.5 AQI Category
0	Russian Federation	Praskoveya	51	Moderate	1	Good	36	Good	0	Good	51	Moderate
1	Brazil	Presidente Dutra	41	Good	1	Good	5	Good	1	Good	41	Good
2	Italy	Priolo Gargallo	66	Moderate	1	Good	39	Good	2	Good	66	Moderate
3	Poland	Przasnysz	34	Good	1	Good	34	Good	0	Good	20	Good
4	France	Punaauia	22	Good	0	Good	22	Good	0	Good	6	Good
5	United States of America	Punta Gorda	54	Moderate	1	Good	14	Good	11	Good	54	Moderate
6	Germany	Puttlingen	62	Moderate	1	Good	35	Good	3	Good	62	Moderate
7	Belgium	Puurs	64	Moderate	1	Good	29	Good	7	Good	64	Moderate
8	Russian Federation	Pyatigorsk	54	Moderate	1	Good	41	Good	1	Good	54	Moderate
9	Egypt	Qalyub	142	Unhealthy for Sensitive Groups	3	Good	89	Moderate	9	Good	142	Unhealthy for Sensitive Groups
10	China	Qinzhou	68	Moderate	2	Good	68	Moderate	1	Good	58	Moderate
11	Netherlands	Raalte	41	Good	1	Good	24	Good	6	Good	41	Good



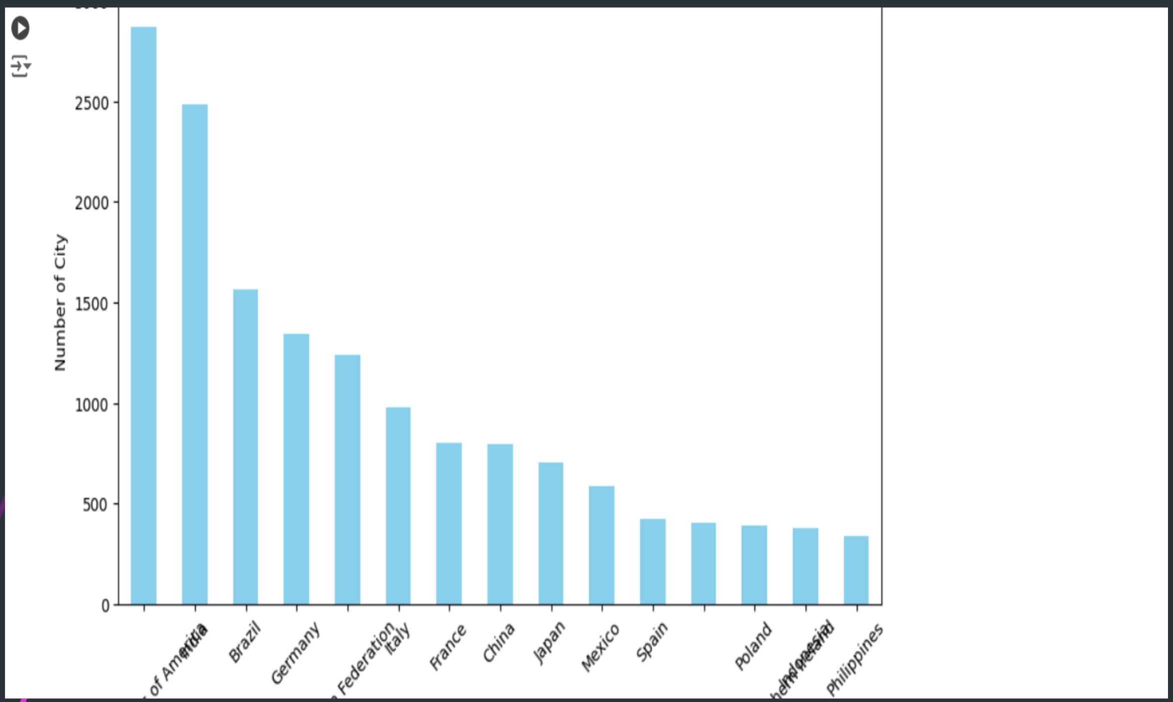
```python
# bar
plt.figure(figsize=(10,6))
Country_counts=df['Country'].value_counts().head(15)
Country_counts.plot(kind='bar',color='skyblue')
plt.title('Top 15 countries')
plt.xlabel('Country')
plt.ylabel('Number of City')
plt.xticks(rotation=45)
plt.show()
```

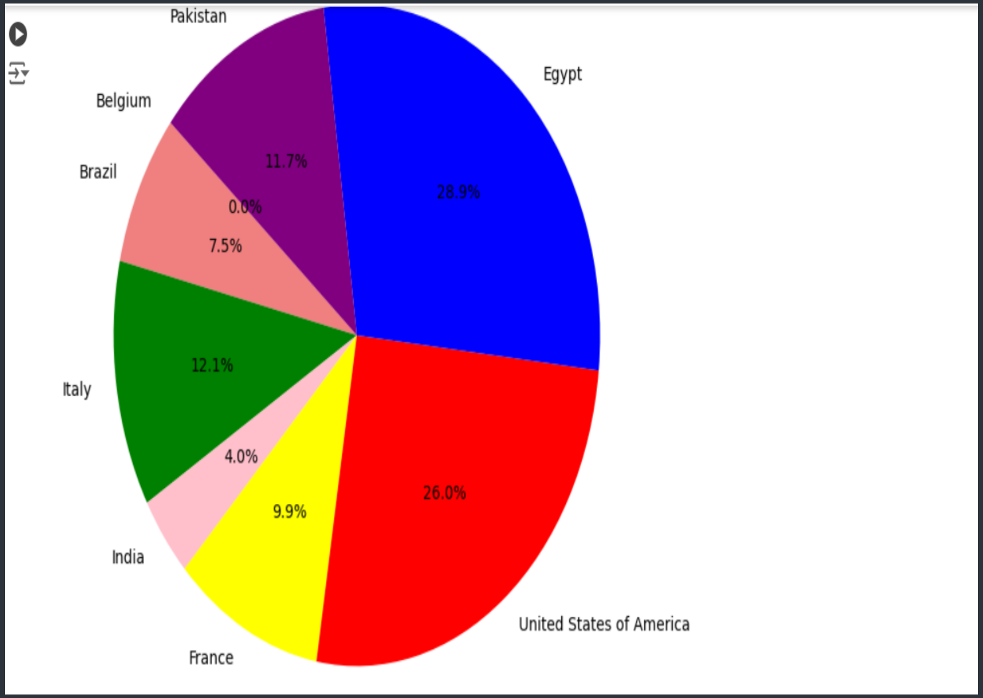


```python
# display top 15 highest AQI Values
top15_len = df.sort_values(by='AQI Value', ascending=False).head(15)[['City','AQI Value']].set_index('City')
print(top15_len)
```

```
              AQI Value
City
Haldaur             500
Mahendragarh        500
Barkhera            500
Khetri              500
Jahangirpur         500
Phalauda            500
Patiala             500
Kakrala             500
Kandhla             500
Hasanpur            500
Dhuri               500
Lachhmangarh        500
Malaut              500
Padampur            500
Jhunjhunun          500
```

RESULT ANALYSIS AND VISUALISATION PLOTTING





CONCLUSION AND FUTURE STUDY

Conclusion:

In conclusion, the global air pollution analysis dataset is an indispensable tool for comprehending the intricate patterns and impacts of air quality worldwide. It empowers researchers, policymakers, and health professionals to identify pollution trends, pinpoint sources, and evaluate the efficiency of regular measures. By integrating data from various monitoring technologies and models, this dataset enhances our ability to address the pervasive challenge of air pollution. Ongoing advancements and expanded access to this data are essential for fostering informed decision-making and driving global efforts to mitigate air pollution, ultimately leading to improved public health and environmental sustainability.

Future Study:

Future studies of global air pollution analysis will likely focus on enhancing the accuracy and granularity of pollution data through advanced monitoring technologies and machine learning algorithms. Researchers will aim to better understand the interactions between air pollution and climate change, as well as the socio-economic factors influencing pollution patterns. Additionally, there will be an increased emphasis on developing real-time monitoring and predictive models to support proactive policy interventions. Collaborative international research efforts will be crucial to address the disparities in air quality data availability and to develop comprehensive strategies for global air pollution mitigation.

REFERENCES

Matplotlib: A 2D Graphics Environment ,John D. Hunter
Computing in science & engineering (Print) 2007. 17993 Citations, 1 References.
Python Data Analytics: With Pandas, NumPy, and Matplotlib Fabio Nelli 2023

Research on Big Data Analysis Data Acquisition and Data Analysis
Hong Li
2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA)

Thank you