**AIR QUALITY MONITIORING**

**A SUMMER INTERNSHIP REPORT**

***Submitted by***

**VARUN.S**

**71762332054**

**Coimbatore Institute Of Technology**

***in partial fulfillment for the Completion of***

**Summer Internship 2024**

***In***

**PROJECT DEVELOPMENT CELL (PDC)**

**COMPUTER SCIENCE AND ENGINEERING**

****

**COIMBATORE INSTITUTE OF TECHNOLOGY**

***(Government-Aided Autonomous Institution Affiliated to Anna University)* COIMBATORE-641014**

**ANNA UNIVERSITY: CHENNAI 600 025**

**July 2024**

**COIMBATORE INSTITUTE OF TECHNOLOGY**

***(A Govt. Aided Autonomous Institution Affiliated to Anna University)* COIMBATORE – 641014**

**BONAFIDE CERTIFICATE**

Certified that this summer internship’2024 project **“AIR QUALITY MONITORING”** is the bonafide work of **Varun S, 71762332054, MSc Data Science, Coimbatore Institute Of Technology** under my mentorship during the period **28th June to 13th July 2024.**

Certified that the candidates were examined continuously by us during the summer internship held at our premises through PDC.

**Mentor name Dr.A.Kunthavai Designation Convener – PDC** Department of CSE, Department of CSE, Coimbatore Institute of Technology, Coimbatore Institute of Technology, Coimbatore – 641014. Coimbatore – 641014.

Place:Coimbatore

Date:15/07/2024

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO** | **TITLE** | **PAGE NO** |
|  | ABSTRACT | I |
|  | LIST OF FIGURES | III |
|  | LIST OF ABBREVIATIONS | IV |
| 1 | INTRODUCTION | 1 |
| 2 | SYSTEM METHODOLOGY | 3 |
| 2.1 | SYSTEM DESIGN | 3 |
| 2.2 | SYSTEM ARCITECTURE | 4 |
| 3 | RESULTS AND DISCUSSION | 9 |
| 4 | CONCLUSION AND FUTURE WORK | 14 |
|  | REFERENCES | 14 |

**ABSTRACT**

#### Background:

Air pollution is a critical global issue, impacting both human health and the environment. Poor air quality contributes to respiratory diseases, cardiovascular conditions, and overall reduced quality of life. As urbanization and industrialization continue to grow, monitoring and improving air quality has become more important than ever. This project focuses on real-time monitoring of air quality to provide insights and recommendations for pollution control, ensuring a healthier living environment.

#### Scope of the Project:

The scope of this project encompasses the development of an interactive dashboard using Tableau, designed to monitor air quality in real-time across various cities. The project utilizes a dataset that includes metrics such as the number of days with good, moderate, and unhealthy air quality, as well as specific pollutant levels (CO, NO2, Ozone, PM2.5, PM10). The dashboard aims to visualize this data through various charts and graphs, offering a comprehensive view of air quality trends and potential areas for improvement.

#### Algorithm/Idea and Results:

The core idea of this project involves leveraging data visualization techniques to enhance the understanding of air quality metrics. The dataset is pre-processed to ensure consistency and accuracy, followed by the creation of calculated fields and pivoting of relevant data columns. Several visualization techniques are employed, including heat maps, bubble charts, dual-axis charts, stacked bar charts, and pie charts, to present the distribution and trends of air pollutants.

The results of the proposed project include an easy-to-navigate dashboard that provides detailed insights into air quality patterns. Users can interact with the dashboard to filter data by city and time period, identify high-pollution areas, and track the effectiveness of pollution control measures over time. This tool not only aids policymakers in making informed decisions but also raises public awareness about air quality issues.

In conclusion, this project leverages Tableau's powerful data visualization capabilities to create an effective air quality monitoring system, contributing to healthier communities and a cleaner environment.

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO** | **TITLE** | **PAGE NO** |
| 1.1 | RAW DATASET | 2 |
| 2.1 | INFORMATION DATASET | 3 |
| 2.2 | DIMENSION OF DATASET | 4 |
| 2.3 | SUMMARY STATISTICS | 4 |
| 2.4 | FEW ROWS OF DATASET | 5 |
| 2.5 | DISPLAY OF NULL VALUES | 6 |
| 3.1 | HISTOGRAM OF AIR QUALITY DAYS | 9 |
| 3.2 | BELL CURVE OF MAX AQI | 10 |
| 3.3 | HEAT MAP OF AQI MEASURES | 10 |
| 3.4 | SCATTERPLOT OF AQI COMPARISON | 11 |
| 3.5 | BAR CHART OF POLLUTANT | 12 |
| 3.6 | STACK BARCHART OF POLLUTANTS | 13 |
| 3.7 | GEOSPATIAL ANALYSIS | 13 |
| 3.8 | DASHBOARD OF AQI | 14 |

**LIST OF ABBREVIATIONS**

KPI - Key Performance Indicator

AQ - Air Quality

AQI – Air Quality Index

EDA – Exploratory Data Analysis

CO – Carbon Di Oxide

NO2 – Nitrogen Di Oxide

PM 2.5 – Particulate Matter 2.5

PM 10 - Particulate Matter 10

ETL – Extract Transform Load

GIS – Geographical Information System

CSV - Comma-Separated Values

**CHAPTER 1**

**INTRODUCTION**

* 1. **Background Information:**

Air pollution poses a significant threat to both human health and the environment. It is associated with respiratory diseases, cardiovascular conditions, and overall reduced quality of life. With increasing urbanization and industrial activities, monitoring air quality has become essential to mitigate these adverse effects. This project leverages Exploratory Data Analysis (EDA) and Tableau for real-time visualization of air quality, aiming to provide actionable insights for pollution control and public awareness.

* 1. **Problem Statement:**

Despite the critical importance of air quality, many urban areas lack effective tools for real-time monitoring and analysis. Traditional air quality monitoring systems often fall short in providing easily interpretable data for immediate action. This project seeks to fill this gap by creating an interactive, user-friendly dashboard that offers comprehensive and real-time insights into air quality metrics.

* 1. **Scope:**

The project encompasses the entire lifecycle of air quality data analysis, from data preparation to visualization. It involves collecting air quality data, performing EDA, creating various charts and key performance indicators (KPIs), and integrating these elements into a cohesive dashboard. The scope also includes making the dashboard interactive and easy to navigate for diverse users, from policymakers to the general public.

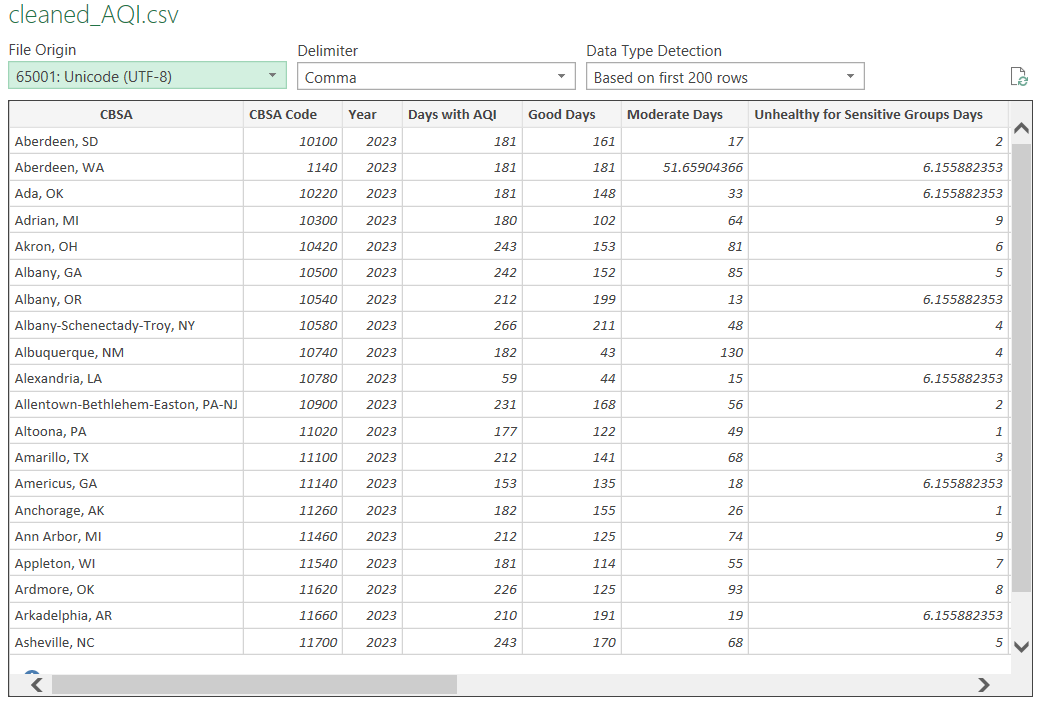
**1.4 Objectives:**

* To collect and preprocess air quality data from reliable sources.
* To perform exploratory data analysis to uncover trends and patterns in air quality.
* To create informative visualizations using Tableau, including heat maps, bubble charts, and pie charts.
* To integrate KPIs and interactive elements into the dashboard.
* To provide actionable insights and recommendations for air quality improvement.

**1.5 Applications**

* Public health agencies can use the dashboard to monitor air quality and issue warnings during high pollution periods.
* Policymakers can utilize the insights to implement effective pollution control measures.
* Environmental researchers can analyze trends and identify sources of pollution.
* The general public can stay informed about air quality in their area and take necessary precautions.

**Fig.1.1 Raw Dataset**



**CHAPTER 2**

**SYSTEM METHODOLOGY**

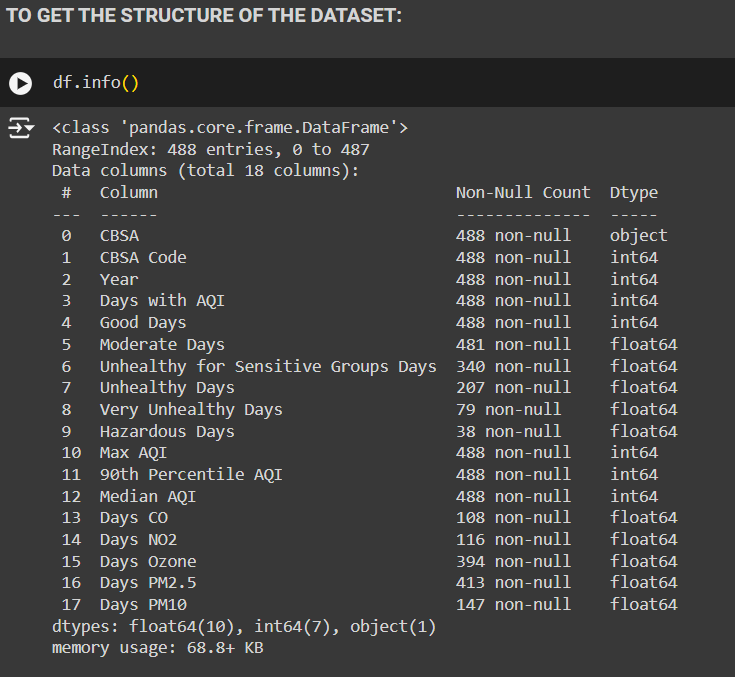
#### 2.1 System Design:

**System Design Overview** The system design for the Air Quality Monitoring Project involves several stages: data collection, data preprocessing, exploratory data analysis (EDA), visualization, geospatial analysis, and dashboard creation. Each stage is crucial for transforming raw air quality data into actionable insights presented in an interactive dashboard.

**2.1.1 Data Collection:**

* **Sources:** Data is collected from reputable sources such as government databases, environmental agencies, and sensor networks.
* **Data Types:** Includes metrics like days with AQI, good days, moderate days, unhealthy days, and specific pollutant levels (CO, NO2, Ozone, PM2.5, PM10).

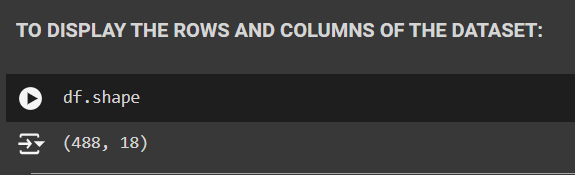
**Fig.2.1 INFORMATION DATASET**



**2.1.2 Data Preprocessing:**

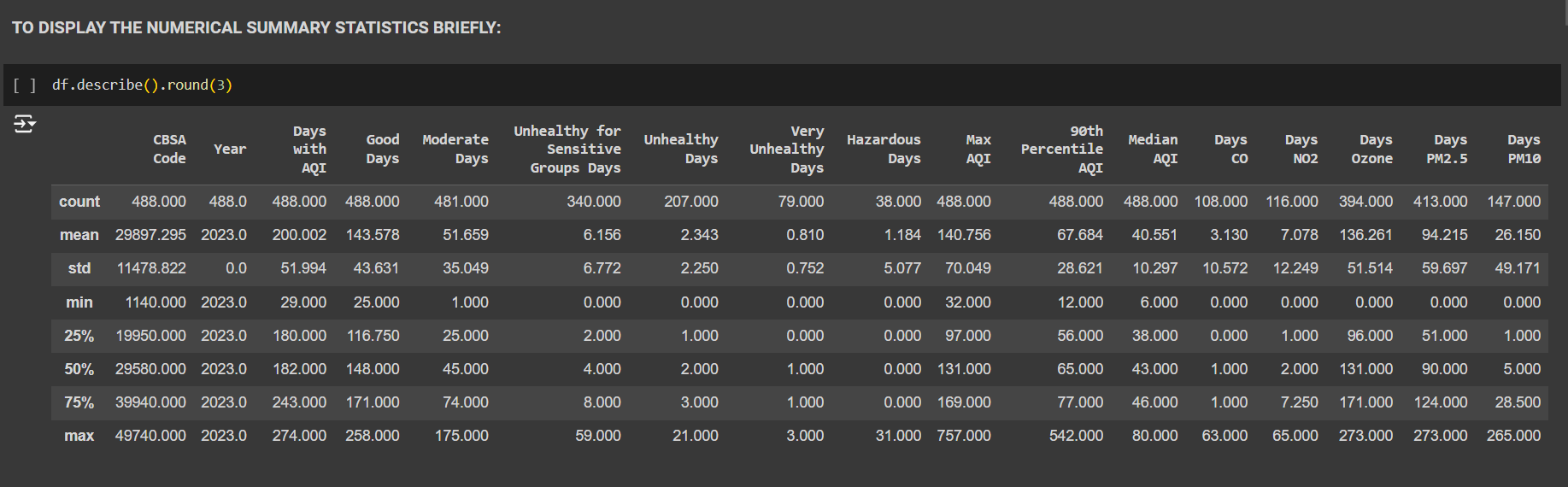
* **Data Cleaning:** Removing duplicates, handling missing values, and correcting inconsistencies.

**Fig.2.2 DIMENSION OF DATASET**

****

* **Summary Statistics:** Calculating mean, median, standard deviation, and other descriptive statistics.

**Fig.2.3 SUMMARY STATISTICS**

****

* **Data Transformation:** Normalizing data, creating calculated fields, and pivoting relevant columns.
* **Null Value Removal:** Ensuring all fields are complete and ready for analysis.

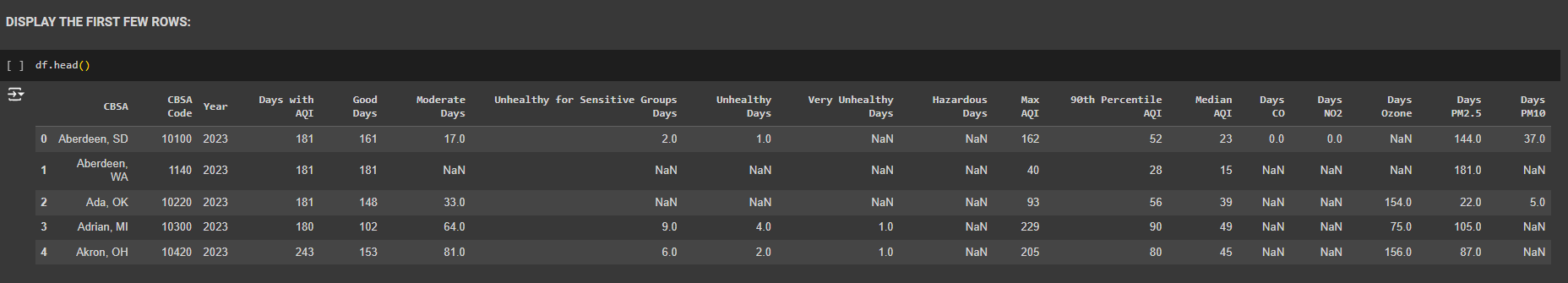
#### 2.2 System Architecture:

**2.2.1 System Architecture Overview:**

The system architecture is designed to support seamless data processing and visualization, comprising the following components:

1. **Data Ingestion Layer:**
   * Collects and stores raw air quality data.
   * Uses ETL (Extract, Transform, Load) processes to clean and transform data

**Fig.2.4 FEW ROWS OF DATASET**



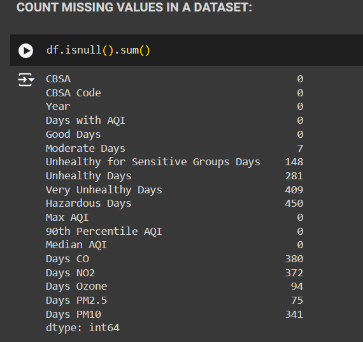
1. **Data Processing Layer:**
   * Implements EDA techniques to uncover patterns and trends.
   * Uses statistical analysis and data transformation methods.
2. **Visualization Layer:**
   * Utilizes Tableau for creating interactive charts and dashboards.
   * Integrates various visual elements for comprehensive data representation.
3. **Geospatial Analysis Layer:**
   * Uses geospatial data to map pollutant levels across cities.
   * Incorporates GIS tools for spatial analysis.
4. **Presentation Layer:**
   * Hosts the Tableau dashboard.
   * Provides an intuitive interface for users to explore data and insights.

#### 2.2.2 Module Description with Explanation

**1. Data Preparation Module:**

* **Data Cleaning:** Involves removing duplicates and correcting inconsistencies to ensure data quality.
* **Summary Statistics:** Calculation of basic statistical measures to understand data distribution.
* **Data Transformation:** Normalizes and pivots data for analysis.
* **Null Value Removal:** Ensures all fields are complete for accurate analysis.

**Fig.2.5 DISPLAY OF NULL VALUES**



**2. Exploratory Data Analysis (EDA) Module:**

* **Objective:** Uncover patterns and trends in the data.
* **Techniques Used:** Descriptive statistics, visualization of data distributions.
* **Outputs:** Insights into data quality, initial trends, and potential anomalies.

**3. Visualization Module:**

* **Charts Created:**
  + **Bar Chart:** Air quality days across cities.
  + **Trend Distribution Chart:** Max AQI trends.
  + **Heatmap:** Distribution of AQI measurements.
  + **Dual-Axis Chart:** Max AQI vs. Median AQI trends.
  + **Pie Chart:** Total pollutant levels.
  + **Bar Chart:** Days with specific pollutant levels.
* **Tools Used:** Tableau for creating interactive and informative visualizations.

**4. Geospatial Analysis Module:**

* **Objective:** Map pollutant levels across different cities.
* **Techniques Used:** GIS tools for spatial analysis, integrating geospatial data with pollutant metrics.
* **Outputs:** Visual maps showing pollutant distribution and identifying hotspots.

**5. Dashboard Creation Module:**

* **Objective:** Combine all visualizations into a cohesive, interactive dashboard.
* **Components:**
  + KPIs for quick insights.
  + Filters for user-driven data exploration.
  + Linked charts for comprehensive analysis.
* **Tool Used:** Tableau for integrating charts and ensuring user-friendly navigation.

#### 2.2.3 Algorithm for Data Processing and Visualization:

**Step-by-Step Process:**

1. **Data Ingestion:**
   * Extract data from various sources.
   * Load data into a central storage system.
2. **Data Cleaning:**
   * Identify and remove duplicates.
   * Handle missing values by imputation or removal.
3. **Summary Statistics:**
   * Calculate mean, median, standard deviation, etc.
4. **Data Transformation:**
   * Normalize data to ensure consistency.
   * Pivot relevant columns for analysis.
5. **EDA:**
   * Perform descriptive statistics.
   * Visualize initial data distributions and trends.
6. **Visualization:**
   * Create charts in Tableau:
     + Bar Chart for air quality days.
     + Trend Distribution Chart for Max AQI.
     + Heatmap for AQI distribution.
     + Dual-Axis Chart for Max vs. Median AQI.
     + Pie Chart for pollutant levels.
     + Bar Chart for specific pollutant days.
7. **Geospatial Analysis:**
   * Integrate geospatial data.
   * Create maps showing pollutant distribution.
8. **Dashboard Creation:**
   * Combine all visualizations in Tableau.
   * Add KPIs, filters, and interactivity for user navigation.

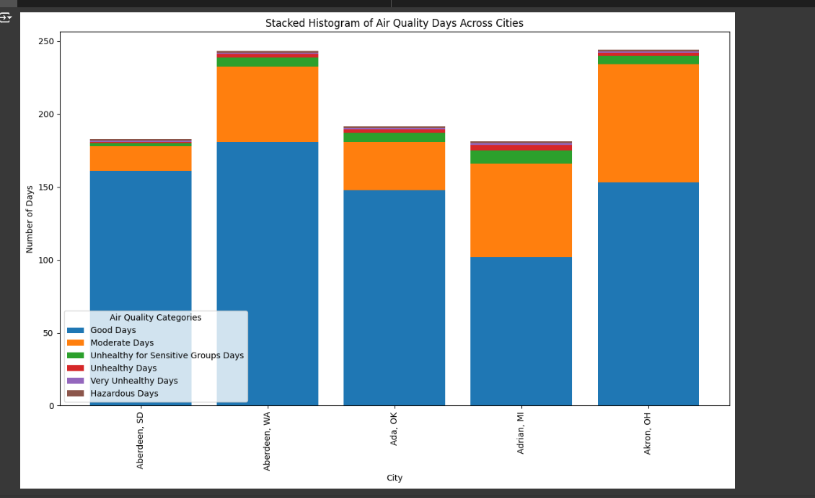
By following these steps, the Air Quality Monitoring Project successfully transforms raw data into meaningful insights, facilitating informed decision-making and public awareness.

**CHAPTER 3**

**RESULTS AND DISCUSSIONS**

**3.1 STACKED HISTOGRAM OF AQ DAYS :**

**Fig.3.1 HISTOGRAM OF AIR QUALITY DAYS**



**Title:** Stacked Histogram of Air Quality Days Across Cities

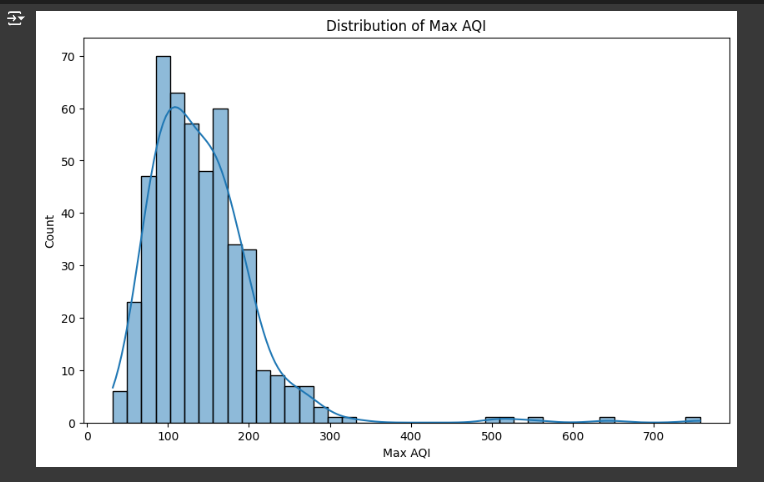
**X-Axis:** City

**Y-Axis:** Number of Days

The stacked histogram displays the number of days across various air quality categories for different cities. The categories include Good Days , Moderate Days , Unhealthy for Sensitive Groups Days, Unhealthy Days , Very Unhealthy Days, and Hazardous Days. Each city has a bar with different segments representing the number of days in each air quality category.

**3.2 DISTRIBUTION OF MAX AQI:**

**Fig.3.2 BELL CURVE OF MAX AQI**



**Title:** Distribution of Max AQI

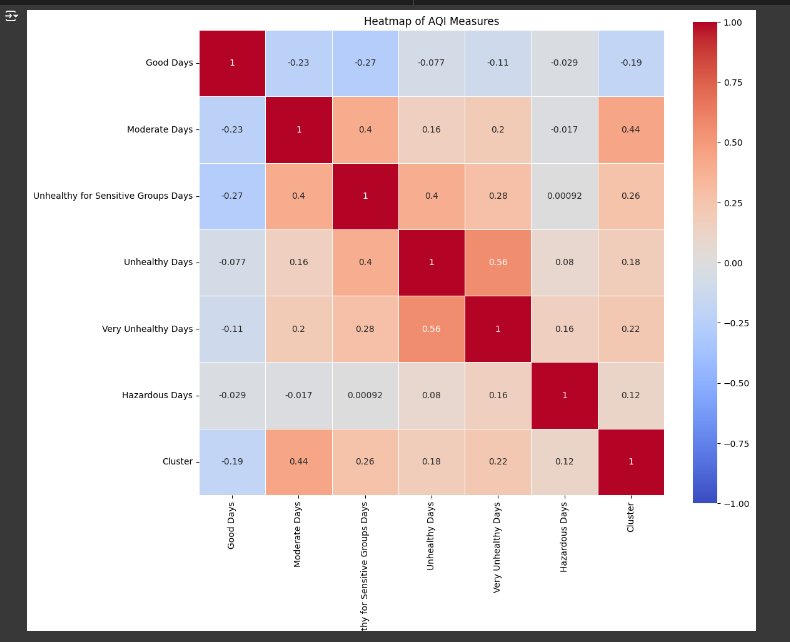
**X-Axis:** Max AQI

**Y-Axis:** Count

The histogram shows the distribution of maximum AQI values. The bars represent the frequency of days that fall into specific AQI ranges. A curve is overlaid on the histogram to indicate the distribution trend. The majority of days have a max AQI between 50 and 200, with a peak around 100-150. The frequency significantly decreases for higher AQI values, with very few days exceeding 300.

**3.3 HEATMAP OF AQI MEASURES:**

**Fig. 3.3 HEATMAP OF AQI MEASURES**



**Title:** Heatmap of AQI Measures

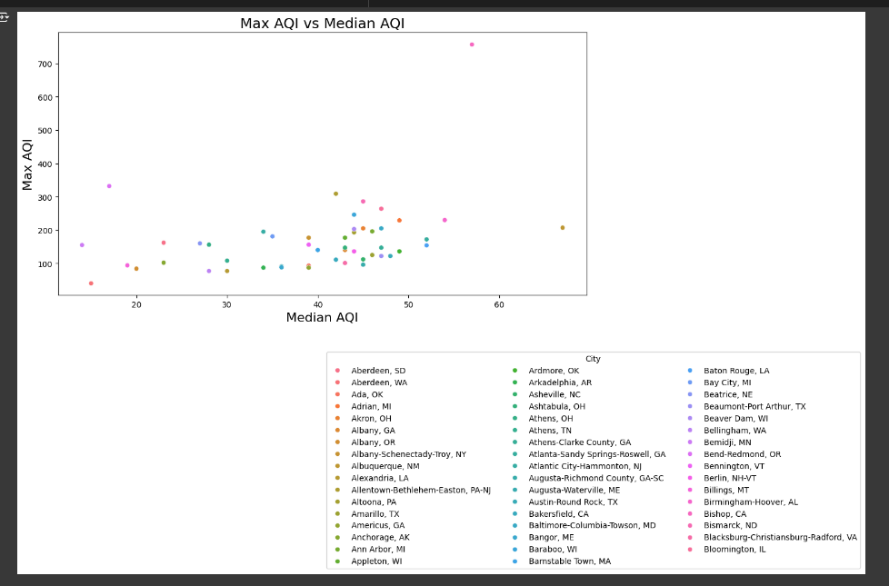
**X-Axis:** Air Quality Categories / Cluster

**Y-Axis:** Air Quality Categories

This heatmap illustrates the correlation between different air quality categories and a cluster variable. The color scale ranges from blue to red, where blue represents a strong negative correlation, red indicates a strong positive correlation, and shades in between show varying degrees of correlation.

3.4 MAX AQI VS MEDIAN AQI:

**Fig.3.4 SCATTERPLOT OF AQI COMPARISON**



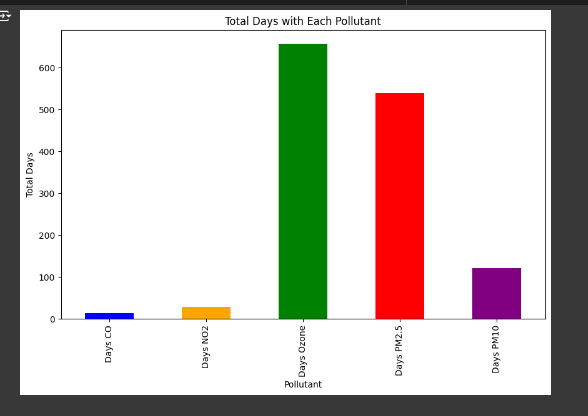
Title: Max AQI vs Median AQI.

* **X-axis**: Median AQI
* **Y-axis**: Max AQI

Each point represents a city, with its color corresponding to a specific city listed in the legend below the plot. The plot compares the maximum Air Quality Index (AQI) values with the median AQI values for various cities.

**3.5 TOTAL DAYS OF EACH POLLUTANT:**

**Fig.3.5 BAR CHART OF POLLUTANT**



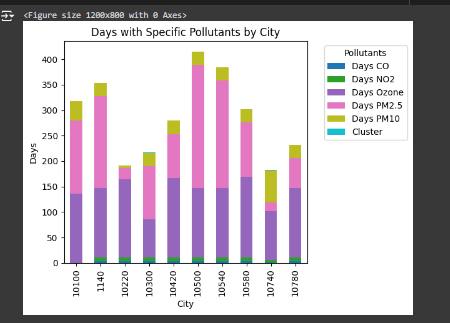
Title: Total Days with Each Pollutant

* **X-axis**: Pollutant (categorized as Days CO, Days NO2, Days Ozone, Days PM2.5, Days PM10)
* **Y-axis**: Total Days

Each bar represents the total number of days a specific pollutant was recorded. The chart shows the number of days for each pollutant, with the pollutants being CO, NO2, Ozone, PM2.5, and PM10. The height of each bar indicates the total days for each pollutant.

**3.6 DAYS WITH SPECIFIC POLLUTANTS BY CITY:**

**Fig.3.6 STACK BARCHART OF POLLUTANTS**



Title: Days with Specific Pollutants by City

* **X-axis**: City (represented by numerical codes such as 10100, 11400, 10220, etc.)
* **Y-axis**: Days

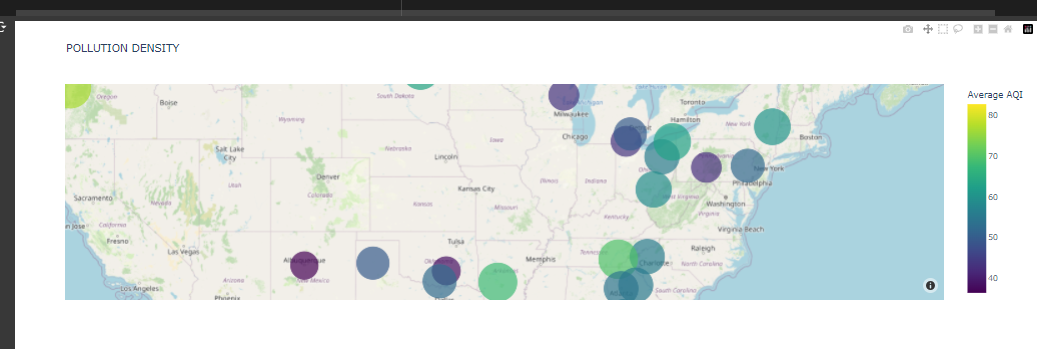
Each bar is divided into segments that represent the number of days each specific pollutant was recorded in each city. The colors in each bar correspond to different pollutants:

* Dark blue: Days CO
* Orange: Days NO2
* Green: Days Ozone
* Pink: Days PM2.5
* Purple: Days PM10

The legend on the right indicates which color corresponds to which pollutant. The height of each segment within a bar shows the number of days that particular pollutant was recorded in the corresponding city.

**3.7 GEOSPATIAL ANALYSIS:**

**Fig.3.7 GEOSPATIAL ANALYSIS**



Title Pollution Density: The map shows pollution levels.

Map Features

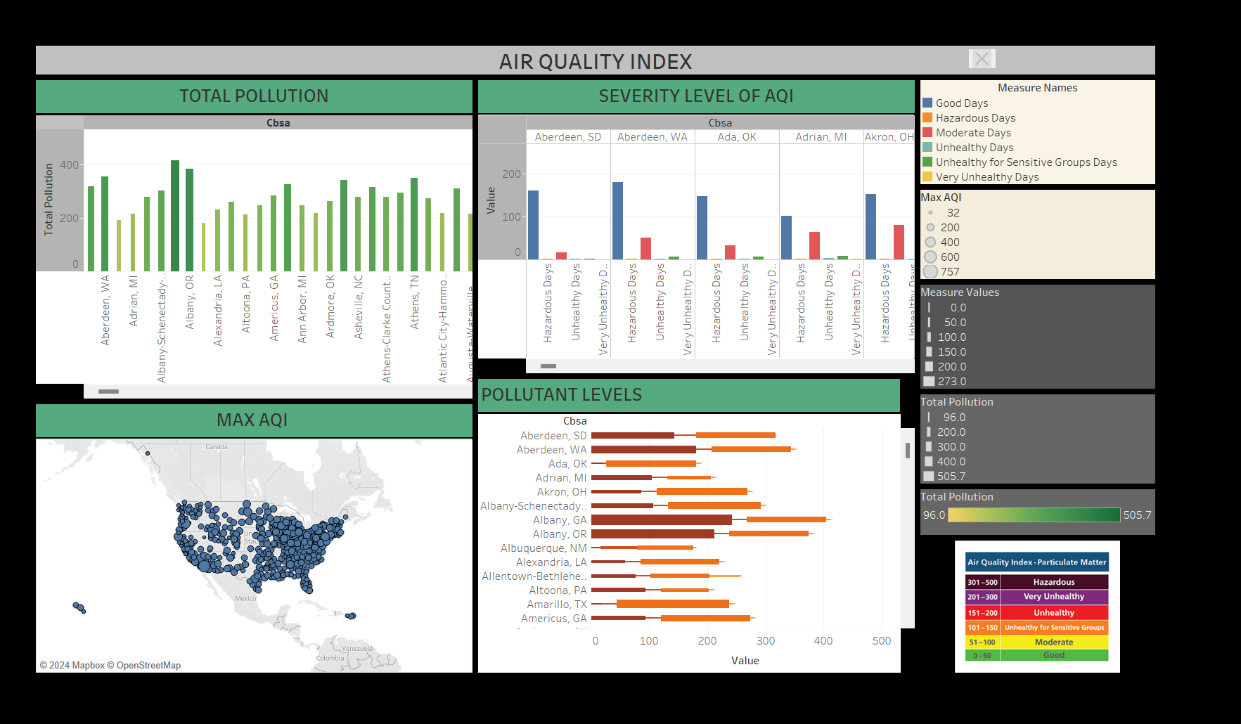
Geographical Coordinates: Locations are plotted by latitude (y-axis) and longitude (x-axis).

Circles: Represent different locations.

Color Scale: Indicates Average AQI (Air Quality Index), ranging from green (low AQI) to purple (high AQI).

**3.8 AQI DASHBOARD:**

**Fig.3.8 DASHBOARD OF AQI:**



This image displays a dashboard for the following sectors:

Total Pollution: Bar graph showing total pollution levels for different cities.

Severity Level of AQI: Bar graph categorizing days by AQI severity levels (Good, Moderate, Unhealthy, etc.) for different cities.

Max AQI: Map indicating the maximum AQI recorded across various locations in the US.

Pollutant Levels: Horizontal bar graph displaying the levels of different pollutants for specific cities.

AQI Legend: Color-coded legend explaining the AQI values.

**CHAPTER 4**

**CONCLUSION AND FUTURE WORK**

### Conclusion:

The Air Quality Monitoring Project effectively uses data visualization and EDA to provide real-time insights into air quality across cities. By utilizing Tableau, the project offers an interactive dashboard that helps policymakers, researchers, and the public make informed decisions about air quality management. Key trends and patterns identified through this project underscore the need for continuous monitoring and data-driven strategies to improve air quality and public health.

### Future Work:

Future enhancements include:

* **Real-Time Data Integration:** Adding more real-time data sources.
* **Predictive Analytics:** Developing forecasting capabilities.
* **Geographic Expansion:** Covering additional areas.
* **User Experience:** Improving dashboard usability.
* **Advanced Visualizations:** Introducing sophisticated visualization techniques.

These upgrades will enhance the tool's accuracy, usability, and effectiveness in air quality management.

**REFERENCES:**

1. <https://colab.research.google.com/drive/1dwE3VOidAbm8F69D__qlT1NYsFhtmLkK?usp=drive_link> - google collab
2. "C:\Users\Varun S\OneDrive\Desktop\tableau\AQI\_Dashboard.twb" - tableau
3. <https://aqs.epa.gov/> - Air Quality data set