Data Analysis and Visualization Project

Project Title: Exploring Air Quality in Tamil Nadu

Project by: Prashant Kumar

1.Introduction

The objective of this project is to analyze and visualize air quality data in Tamil Nadu. The dataset, obtained from [data source], includes information on various air pollutants measured across different locations in the state. This document outlines the steps taken for loading, preprocessing, and analyzing the data using IBM Cognos.

2. Data Loading and Preprocessing

2.1 Data Source

The dataset was sourced from [provide details of the data source]. It comprises [describe the key attributes and variables].

2.2 Data Cleaning

Handling Missing Values: Missing values were addressed using forward fill method for key variables like SO2, NO2, and RSPM/PM10.

Column Removal: Unnecessary columns, such as Stn Code, Sampling Date, Agency, and Location of Monitoring Station, were removed.

Handling PM 2.5 Data: Due to a high percentage of missing values, the PM 2.5 column was dropped.

3. Data Exploration and Analysis

3.1 Descriptive Statistics

Descriptive statistics, including count, mean, and percentiles, were calculated to provide an overview of the dataset.

4. Data Visualization:

- **Create Visualizations:** Utilize various chart types (bar charts, line charts, scatter plots, etc.) to represent data insights.
- **Dashboard Creation:** Develop interactive dashboards to showcase key metrics.
- **Geospatial Analysis:** If applicable, perform geospatial analysis and visualize results

5. Conclusion

In conclusion, this analysis provides valuable insights into air quality in Tamil Nadu. The use of IBM Cognos facilitated efficient data exploration and visualization, enabling the identification of patterns and trends.