Report

Knowledge Representation and Insights Generation from Structured Datasets

1. Introduction

Air quality is a crucial aspect of environmental health, impacting human health and the ecosystem. This project focuses on analyzing an Air Quality Index (AQI) dataset to uncover patterns and insights, develop interactive dashboards, and create a user-friendly interface for a wider audience. The methodology, steps, results, and conclusions are detailed below.

2. Methodology

Objectives:

1. **Uncover patterns and insights:** Utilize Exploratory Data Analysis (EDA) and machine learning algorithms to understand relationships and trends within the data.
2. **Develop interactive dashboards:** Create compelling visualizations using Streamlit and Power BI to effectively communicate the findings.
3. **Design a user-friendly interface:** Implement front-end technologies (HTML, CSS, JavaScript) to build an interactive AQI dashboard accessible to a wider audience.

Tools and Technologies:

- **EDA and Machine Learning:** Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn)
- **Interactive Dashboards:** Streamlit, Power BI
- **Front-End Development:** HTML, CSS, JavaScript

3. Steps

1. Loading the Dataset

- The dataset is loaded into the environment for analysis.

2. Mean and Median Calculation

- Calculating mean and median values for various pollutants to understand the central tendency and spread of the data.

3. EDA Analysis

- **Distribution of AQI Across Categories:** Visualize how AQI values are distributed across different categories.
- **Correlation Heatmap:** Analyze correlations between different pollutants.
- **Residual Plot:** Evaluate the residuals of the AQI predictions.

4. Analysis Based on Pollutants

- **AQI Value Analysis:** Identify patterns and insights from the overall AQI values.
- **CO AQI Value Analysis:** Focus on Carbon Monoxide AQI values.
- **Ozone AQI Value Analysis:** Focus on Ozone AQI values.
- **NO2 AQI Value Analysis:** Focus on Nitrogen Dioxide AQI values.
- **PM2.5 AQI Value Analysis:** Focus on Particulate Matter (PM2.5) AQI values.

5. Model Training and Accuracy

- Training machine learning models to predict AQI and selecting the best-performing model based on accuracy.

4. Results

Insights from AQI Model Representation:

Sum of CO AQI Value by City Analysis:

- **Highest CO AQI Value:** Durango
- **Other High CO AQI Cities:** Miri, Kasongo Lunda, Boksburg, Alberton
- **Lower CO AQI Cities:** Seoul, Caxito

Sum of Ozone AQI Value by City Analysis:

- **Highest Ozone AQI Value:** Eslamshahr
- **Other High Ozone AQI Cities:** Karaj, Qom, Jinan, Robat Karim
- **Lower Ozone AQI Cities:** Hainan, Shaping

Ozone AQI Value by City (Line Chart):

- **Trend:** Descending trend from Eslamshahr to Zhaoqing

CO AQI Value by AQI Category (Pie Chart):

- **Moderate Category:** 37.16%
- **Good Category:** 23.91%
- **Unhealthy for Sensitive Groups:** 21.44%
- **Unhealthy:** 10.59%

Sum of CO AQI Value by Country Analysis:

- **Highest CO AQI Value:** India
- **Other High CO AQI Countries:** United States, China, Brazil, Russian Federation
- **Lower CO AQI Countries:** Japan, France

Sum of Ozone AQI Value by Country Analysis:

- **Highest Ozone AQI Value:** India
- **Other High Ozone AQI Countries:** United States, China, Germany, Italy
- **Lower Ozone AQI Countries:** Brazil, Indonesia

PM2.5 AQI Value by Country Analysis (Pie Chart):

- **Highest PM2.5 AQI Value:** India (23.13%)
- **Other High PM2.5 AQI Countries:** United States, China, Brazil, Germany

Overall AQI Value by Country (Pie Chart):

- **Highest AQI Value:** India (22.52%)
- **Other High AQI Countries:** United States (10.15%), China (5.97%)
- **Smaller AQI Contributions:** Many other countries with smaller shares

Overall Analysis (Pie Chart):

- **Majority CO AQI Category:** Moderate (37.16%)
- **Other Significant Categories:** Good (23.91%), Unhealthy for Sensitive Groups (21.44%), Unhealthy (10.59%)

5. Conclusion

From the above insights and patterns, we concluded a comparative study on AQI based on city, category, location, and pollutants. The analysis reveals significant variations in AQI values across different regions and pollutants, highlighting areas with high pollution levels and those with better air quality. These insights can inform policymakers, environmental agencies, and the public about the state of air quality and help in devising strategies for air quality management and improvement.

The interactive dashboards and user-friendly interface further enhance the accessibility and understanding of AQI data, making it easier for a wider audience to engage with the findings and take informed actions towards improving air quality.