

Proposal: Air quality analysis system

Problem: Creating a system to analyze the air quality based on location, its content and history.

Dataset: Pollution in US open dataset: <https://www.kaggle.com/sogun3/uspollution>.

The dataset presents information about the state, the county, the city, the date of the measurements (from 2000 to 2016), and measurements of the pollutants level in the air. The dataset format is csv (size: 382 MB).

Proposal solution and application:

Carbon Monoxide, Nitrogen Dioxide, Ozone and Sulfur Dioxide are the four main gases that cause air pollution and burning of fuels is the source that creates large amount of carbon dioxide and other three gases into the atmosphere. Our proposal solution is to use the U.S pollution data from 2000 to 2016, to identify when, where and how much of these gases had been created. Also, after extracting and cleaning the data, we will visualize the information of air content in a given date and location. For instance, each of the pollutants included has its own five columns of metrics. For NO2:

- NO2 Units : The units measured for NO2
- NO2 Mean : The arithmetic mean of concentration of NO2 within a given day
- NO2 AQI : The calculated air quality index of NO2 within a given day
- NO2 1st Max Value : The maximum value obtained for NO2 concentration in a given day
- NO2 1st Max Hour : The hour when the maximum NO2 concentration was recorded in a given day

The application of this system is to help visualizing how is the pollution demographics in US and how it evolved over time. Knowing the pollution trends can help to predict future scenarios. In addition, solutions can be taken to avoid catastrophic pollution scenarios. Another application is the comparison of air quality in different locations. This can help to brainstorm possible reasons for the differences.

Project steps:

Step	Estimate Completion Time	Person in charge
Extract and clean up data	One week	Isabel & Haocheng
Data analysis 1. Quality of the air over time 2. Research on the cause of the problem	Two weeks	Haocheng, Isabel, Anh
Data visualization	One week	Anh, Haocheng