

APPLIED BUSINESS ANALYTICS

CAPSTONE PROJECT

TITLE: Purify- Air-Purifying Solutions

Air Quality Analysis for effective Purifier Deployment

By- Riya Singh

BACKGROUND

In today's fast-paced life where people try to find easy and convenient solutions to save up their time, such developments are being made which are helpful to the human race but is adversely affecting the environment and causing various pollutions. One of them being, air quality. It is a critical environmental issue that affects public health, ecosystem stability, and climate change. Urban areas are particularly vulnerable to high levels of pollutants, which can cause respiratory problems, cardiovascular diseases, and other health issues. Monitoring and analyzing air quality data is essential for developing strategies to mitigate these risks and protect public health. And NYC being a huge example for this.

Goals of the Project

This project aims to analyze air quality data to identify patterns and trends in pollutant concentrations across different regions. The goals are:

- To determine the regions most affected by specific pollutants.*
- To provide data-driven recommendations for the placement of air purifiers and other mitigation strategies.*

Data Extraction/Collection/Scraping

The dataset used for this analysis has been extract by using the web-scraping technique using the API key and the API endpoint from NYC Open Data.

URL: 'https://data.cityofnewyork.us/resource/c3uy-2p5r.json'

The dataset has over 18000 rows and 11 columns with the important fields/columns being the name which indicates the air pollutant, the data value indicating the pollutant concentration and the region which has the different regions in NYC affected by these pollutants.

Data Exploration/Data Visualization

I have first cleaned and prepared the data for use by:

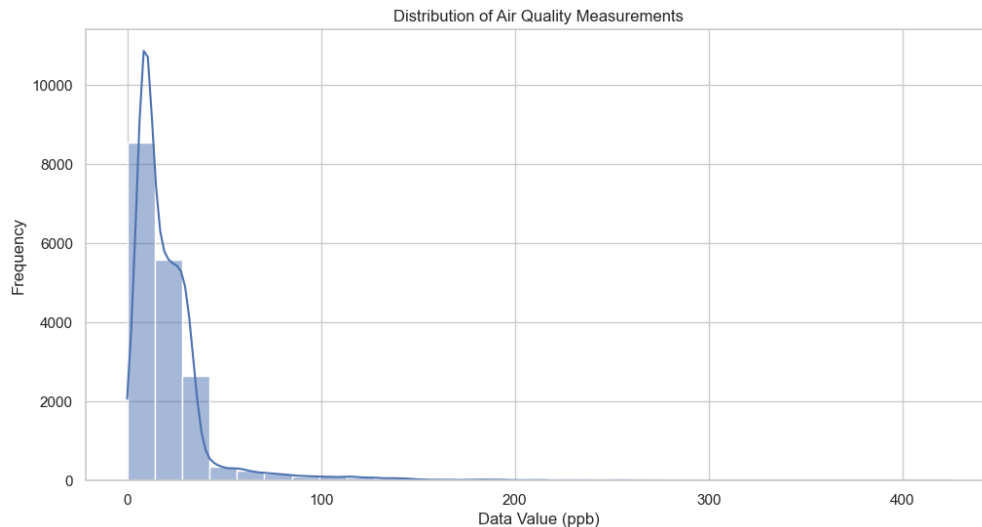
- loading the dataset into the pandas.*
- Dropped a few irrelevant columns. Irrelevant in the sense that these columns necessarily did not add any weightage to the analysis I performed.*
- Handled the missing values in the dataset.*

Then, performed analysis on the fields which included data about the pollutants present, the concentration of these pollutants and the regions in NYC here the presence of these particles were found.

Data Analysis

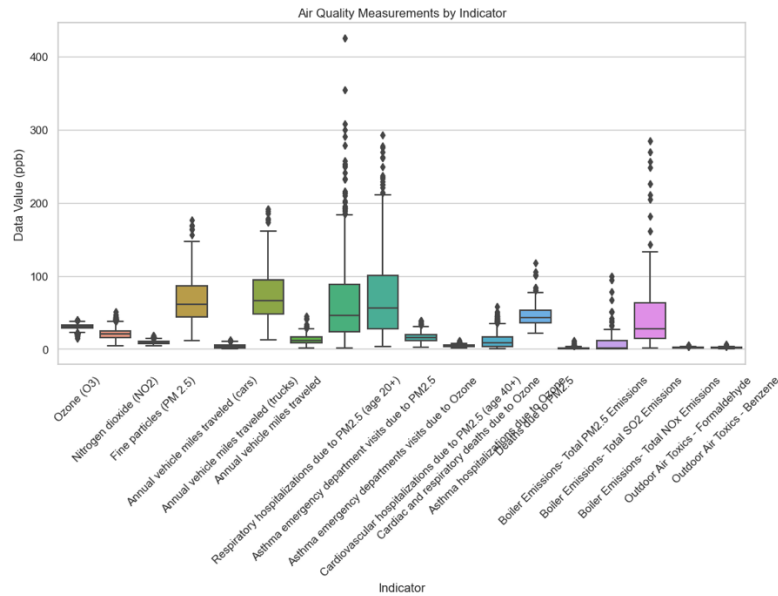
Did the below analysis for the dataset:

First is a histogram that shows the distribution of air quality measurements in parts per billion (ppb).

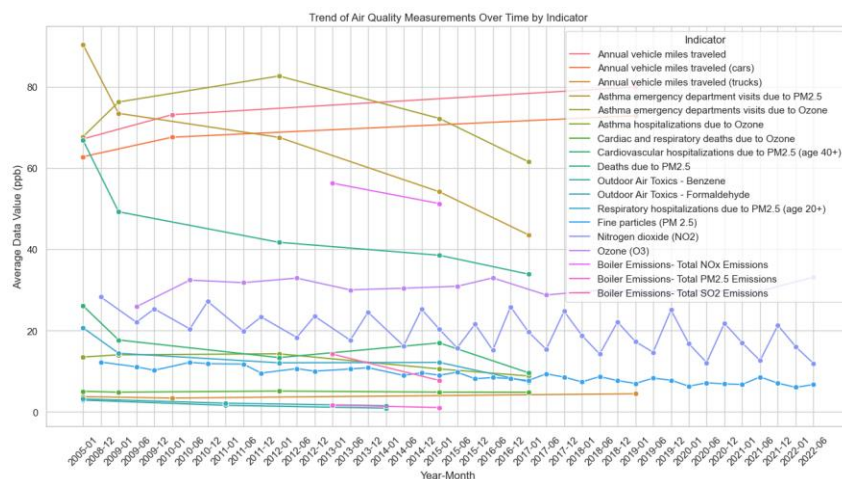


Most of the time, the air quality measurements are low (which is good), but there are a few instances where the air quality gets much worse, and those instances might require investigation or action.

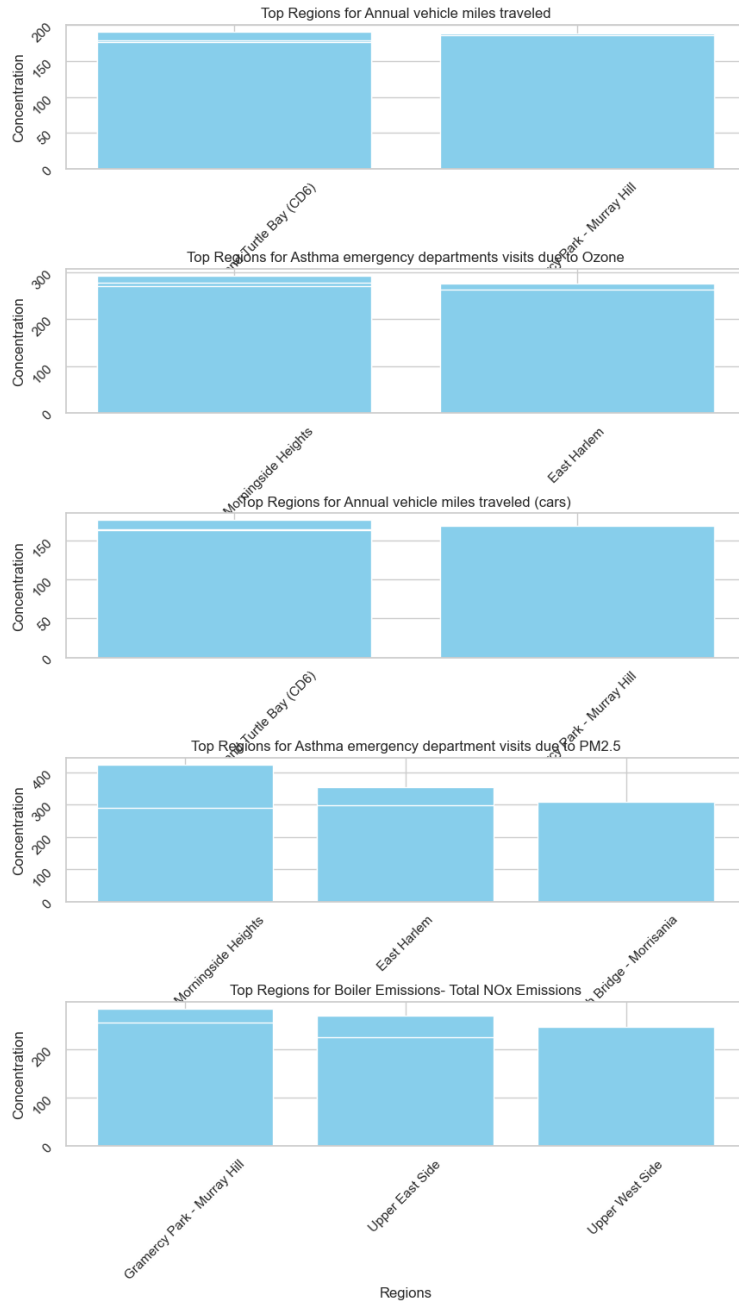
Second is boxplot. This graph gives a snapshot comparison of different air pollutants, showing how common certain levels of pollution are, and highlights the ones with unusually high or low readings that might require special attention.



Third is a line graph showing that the efforts to reduce air pollution may be having a positive effect, as indicated by the general downward trend in pollutant levels and related health issues over the time frame shown. However, it's also clear that pollution from certain sources, like boilers, remains a concern and may require ongoing attention.



I also have identified the top 5 pollutants present and the regions where we can see emission of these pollutants in higher concentration.



Conclusion

Identified the top pollutants,

1. Annual Vehicle Miles Traveled: The region labeled 'East Bay (CD6)' has the highest recorded vehicle miles, suggesting heavy traffic and potential for vehicular pollution.

2. Asthma Emergency Department Visits due to Ozone: A particular region, which could be associated with high ozone levels, has seen a significant number of asthma-related emergency department visits.

3. Annual Vehicle Miles Traveled (Cars): Just as with total vehicle miles, 'East Bay (CD6)' shows a notably high figure for miles traveled by cars, reinforcing the area's link to high traffic and pollution levels.

4. Asthma Emergency Department Visits due to PM2.5: is known for its adverse health effects, and one region stands out for a high number of asthma emergency visits due to this pollutant, indicating serious air quality concerns.

5. Boiler Emissions - Total NOx Emissions: There are notable emissions of nitrogen oxides (NOx) from boilers in certain areas, highlighting spots that might benefit from cleaner energy sources or updated equipment.

This will help us design our products in such a way which filters out these pollutants, increasing the efficiency of our products.

And, the regions help us understand such areas in New York, which could be a growing and potential market to sell and market our products more.

Bibliography

- <https://www.nyc.gov>
- <https://opendata.cityofnewyork.us>
- https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r/about_data
- <https://a816-dohbesp.nyc.gov/IndicatorPublic/L>
- GitHub Link:
https://github.com/Riyasingh27/BAProject/blob/main/Air_Quality_final.ipynb

