AIR QUALITY PROJECT

Project Lead: STACY WAWERU
Date: 1/09/2024

TABLE OF CONTENTS

PROJECT OVERVIEW

Project Title: Predictive Modeling for Air Quality Management in an Italian Industrial City

Project Duration: March 2004 – February 2005

Objective:
The primary objective of this project is to analyze air quality data from an industrial city in Italy and develop predictive models to assist city planners and health officials in monitoring and managing air pollution effectively. The focus is on understanding the relationships between key pollutants—Carbon Monoxide (CO), Nitrogen Oxides (NO_x), and Ozone (O_3)—and environmental factors such as temperature, relative humidity, and time of day.

BUSINESS UNDERSTANDING

Air pollution poses a significant threat to public health, particularly in industrial areas where pollutant levels often exceed safe thresholds.

For city planners and health officials in urbanized regions, the ability to monitor and predict air quality is crucial for making informed decisions that protect the population and improve environmental conditions.

The need for effective air quality management has driven the development of this project.

DATA UNDERSTANDING

Data Sources: Air Quality

The dataset contains 9358 instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device.

Initial analysis revealed strong correlations between pollutants, and skewed data distributions necessitated log transformations to manage outliers. This data forms the foundation for developing predictive models to monitor and manage air quality effectively.

DATA ANALYSIS

Methodology:

- Data Cleaning and preprocessing
- Exploratory Data Analysis
- Visualize key insights
- Prediction Models

Technology used:

Python

DATA ANALYSIS

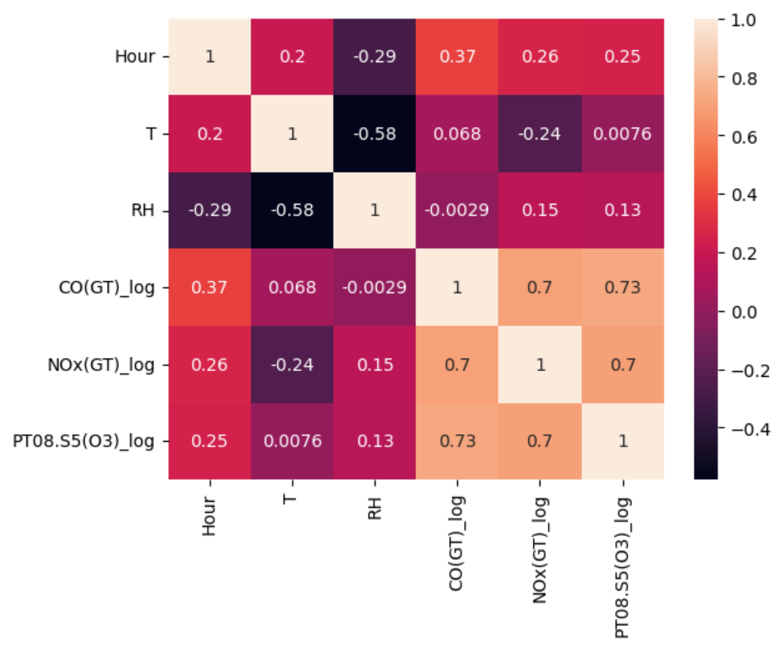
Python is a powerful and versatile programming language used for data analysis among other applications.
Python is used because it has strong and powerful tools and libraries for handling and analyzing data.
Tools and Libraries within python that were used:

- Pandas
- Matplotlib and Seaborn
- Numpy
- Scipy
- Sklearn

MAIN POLLUTING GASES

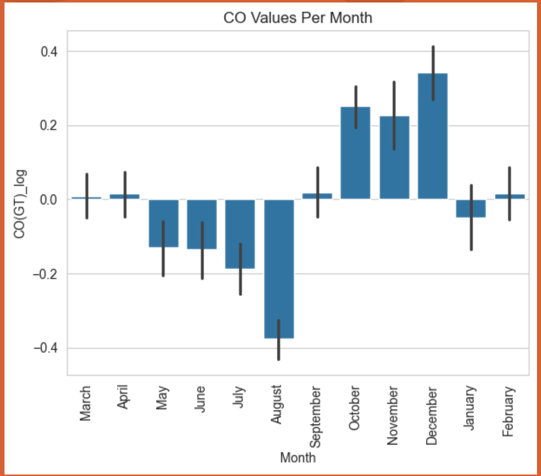
1. CARBON MONOXIDE
2. CARBON DIOXIDE
3. NITROGEN DIOXIDE
4. NITROGEN OXIDE

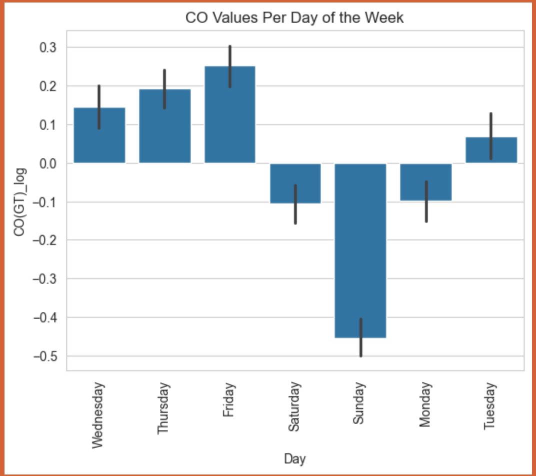
CORRELATIONS

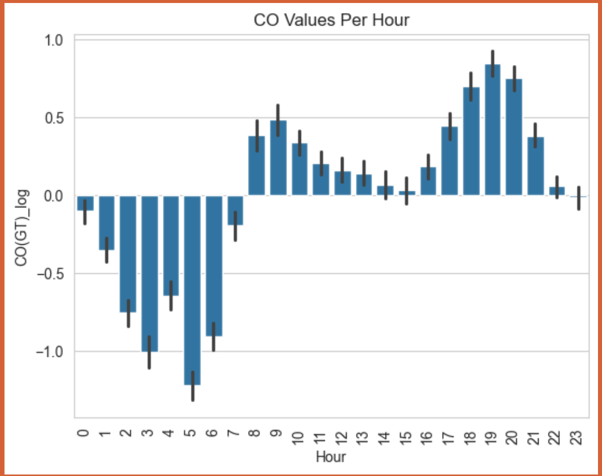


Pollutant Interrelations: The strong positive correlations between the log-transformed pollutant variables (`CO(GT)_log`, `NOx(GT)_log`, and `PT08.S5(O3)_log`) suggest that they are often elevated together, potentially due to common sources or atmospheric conditions.

- Temperature and Humidity: The strong negative correlation between temperature (`T`) and relative humidity (`RH`) aligns with general atmospheric behavior, where warmer air can hold more moisture, reducing relative humidity.







Months

Higher CO Levels: Months like January, February, and March have higher average CO levels.

Lower CO Levels: June, July, and August exhibit lower average CO levels.

Day

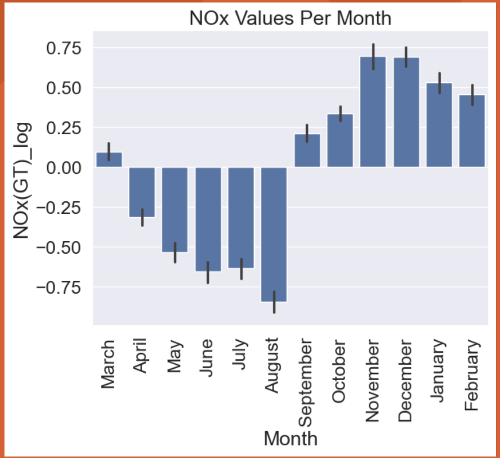
CO levels are lowest on Sundays and highest on Friday.

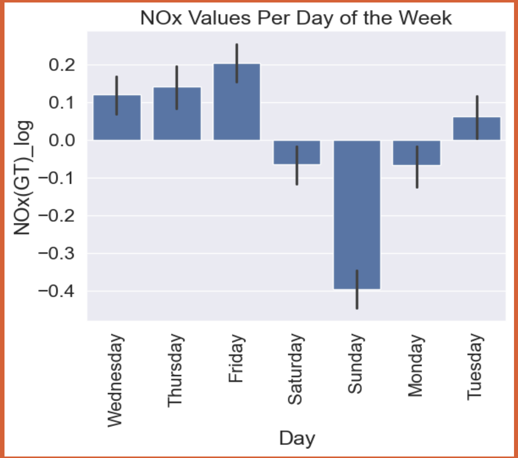
Hour

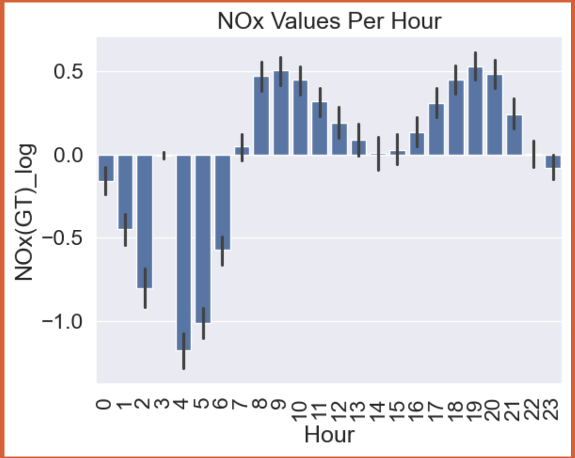
CO Levels: Certain hours (e.g., early morning or late evening) may have higher average CO levels.

This pattern could be related to factors like traffic and industrial activity.

Lower CO Levels: Other hours (e.g., midday or midnight) exhibit lower average CO levels.



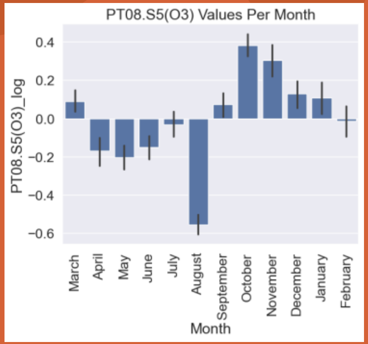


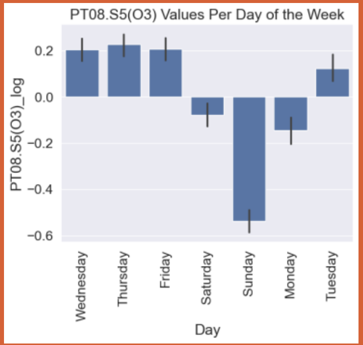


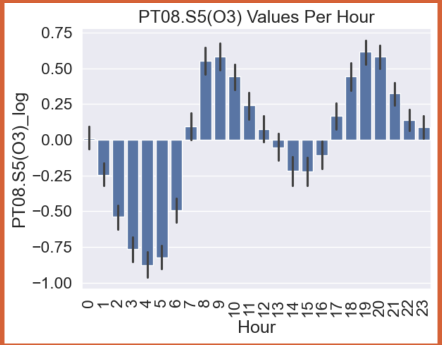
Month
Higher NOx Levels: Certain months (e.g., January, February, and March) have higher average NOx levels.
Lower NOx Levels: Other months (e.g., June, July, and August) exhibit lower average NOx levels.
This pattern could be related to factors like weather conditions, industrial emissions, or seasonal variations

Day
Higher NOx Levels: Monday, Tuesday, and Wednesday have higher NOx(GT) values. This might be due to increased traffic and industrial activity.
Lower NOx Levels: Thursday, Friday, Saturday, and Sunday exhibit lower NOx(GT) values. This could result from reduced traffic or different sources of pollution

Hour
Peak Hours: Certain hours (e.g., early morning and late evening) exhibit higher average NOx levels.
Lower Levels: Midday and midnight tend to have lower average NOx levels.
This pattern could be related to factors like traffic, industrial emissions, or atmospheric conditions







Month
Higher Levels: Months like January, February, and March have higher average PT08.S5(O3) values.
Lower Levels: July exhibits a particularly low average PT08.S5(O3) value compared to other months.
Seasonal patterns may be the reason for this

Day
Higher Levels: Some days (e.g., Monday, Tuesday, and Wednesday) have higher average PT08.S5(O3) values.
Lower Levels: Other days (e.g., Thursday, Friday, Saturday, and Sunday) exhibit lower average PT08.S5(O3) values.
This pattern could be related to factors like traffic, industrial emissions and working during the week.

Hour
Peak Hours: Certain hours (e.g., early morning and late evening) exhibit higher average PT08.S5(O3) values.
Lower Levels: Midday and midnight tend to have lower average PT08.S5(O3) values.
This pattern could be related to factors like traffic, weather, or industrial emissions

PREDICTION MODELLING

Linear Regression

Decent performance

Decision Trees

Poor performance

Gradient Boosting

Best performance

CONCLUSION

The study identified strong interdependencies among these pollutants, particularly in their correlations with time of day and seasonality. These relationships suggest that vehicular emissions, industrial activities, and atmospheric conditions play significant roles in driving pollution levels

These findings indicate that while the models provided some predictive capability, they were limited in their ability to fully capture the complex dynamics of air quality in this environment. The study highlights the need for more advanced modeling techniques and additional data to improve predictive accuracy

RECOMMENDATIONS

a) Enhanced Feature Enginereing – Introduce new features such as wind speed, solar radiation, or traffic density

b) Public health and policy Initiatives – to create awareness and promote behavior change

c) Seasonal emission controls – increasing industrial inspections during colder months

a) Ongoing model refinement and integration – refine the models as new data becomes available

WHAT CAN YOU DO?

REDUCE

RECYCLE

REPAIR

REUSE

"There's so much pollution in the air that if it weren't for our lungs ther'd be no place to put it all."

–ROBERT ORBEN

Air pollution from industries

Vehicles playing a role in increasing pollution levels

THANKS!

DO YOU HAVE ANY QUESTIONS?

waweru.stacy@student.moringaschool.com