AIR QUALITY ANALYSIS IN MAJOR NORTH AMERICAN CITIES WITH THE RISE OF ELECTRIC VEHICLES (EVS).

PRESENTED BY

YEASIN ARAFAT SHAMPOD

MASTER'S IN DATA SCIENCE

FAU ERLANGEN-NÜRNBERG

Module: Methods of Advanced Data Engineering (MADE)

INTRODUCTION

Background

• Rapid adoption of electric vehicles (EVs) promoted as a solution to reduce air pollution in urban areas.
• The analysis in this study focuses on environmental improvements through trends in…

- PM2.5
- NO2
- CO

• Focus: Major North American cities from 2016–2021.

Objective

• Quantify the environmental benefits of transitioning to electric vehicles by studying the relationship between EV adoption and air quality improvements.

Key Research Question

How has the rise of electric vehicles (EVs) affected air quality in major cities in North America?

DATA SOURCES

DataSource-1: Air Quality

- Data Type: CSV
- Licence: CC BY 4.0
- Source:
  - Collected from the U.S. Environmental Protection Agency (EPA).
  - Includes pollutant readings (PM2.5, NO_2, and CO).

Description:

- Timeframe: 2016 to 2021
- Coverage: Major cities in the U.S. and Canada.
- Details: Daily measurements with metadata on monitoring stations.

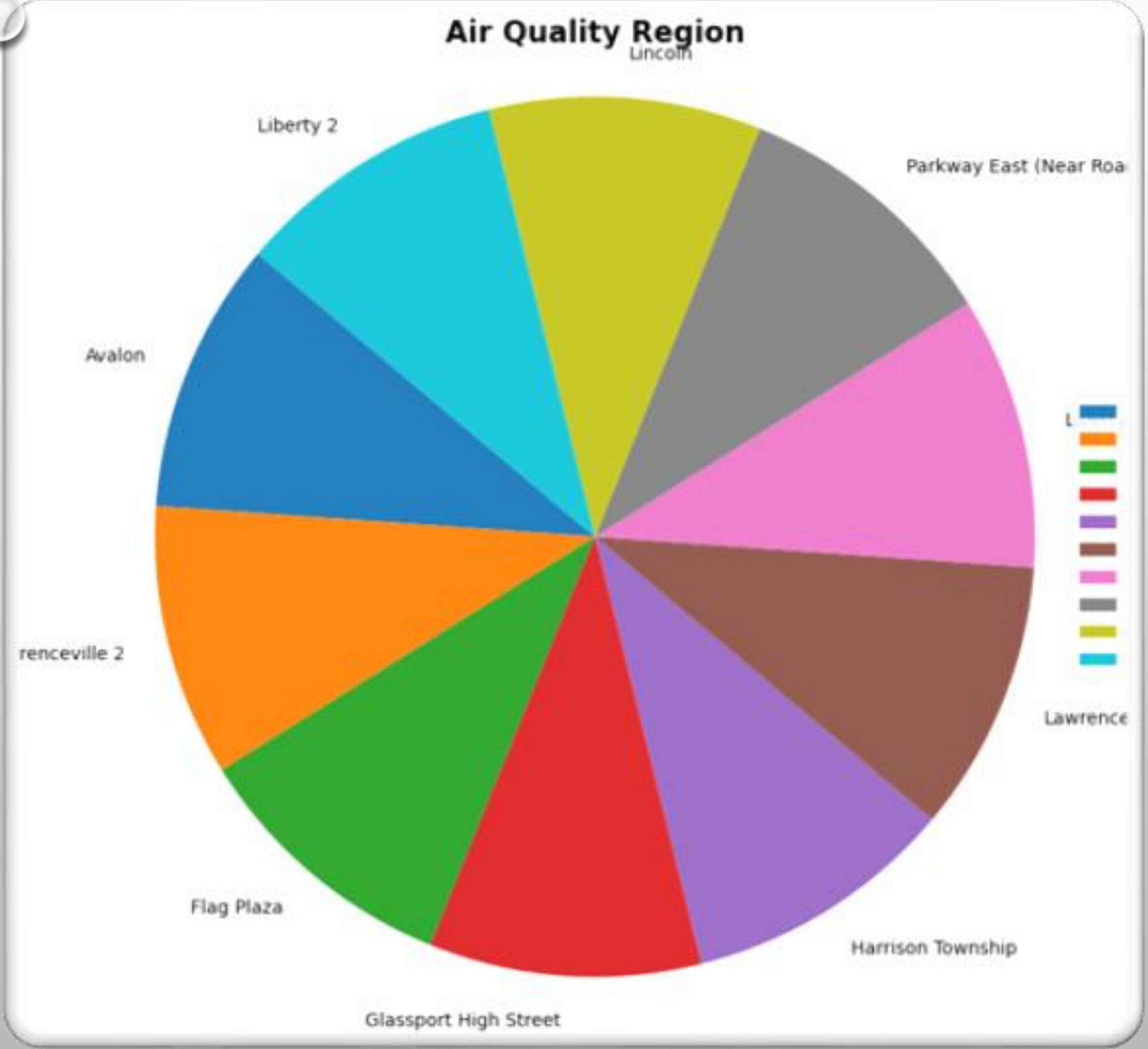
DataSource-2: Electric Vehicle Adoption

- Data Type: CSV
- Licence: Public Domain (Open Source)
- Source:
  - Provided by the U.S. Department of Energy.
  - Contains yearly EV registration trends by city and state.

Description:

- Timeframe: 2016 to 2021
- Details:
  - Yearly EV growth rates across North America.
  - Highlights sustainable transportation progress.

DATA ANALYSIS



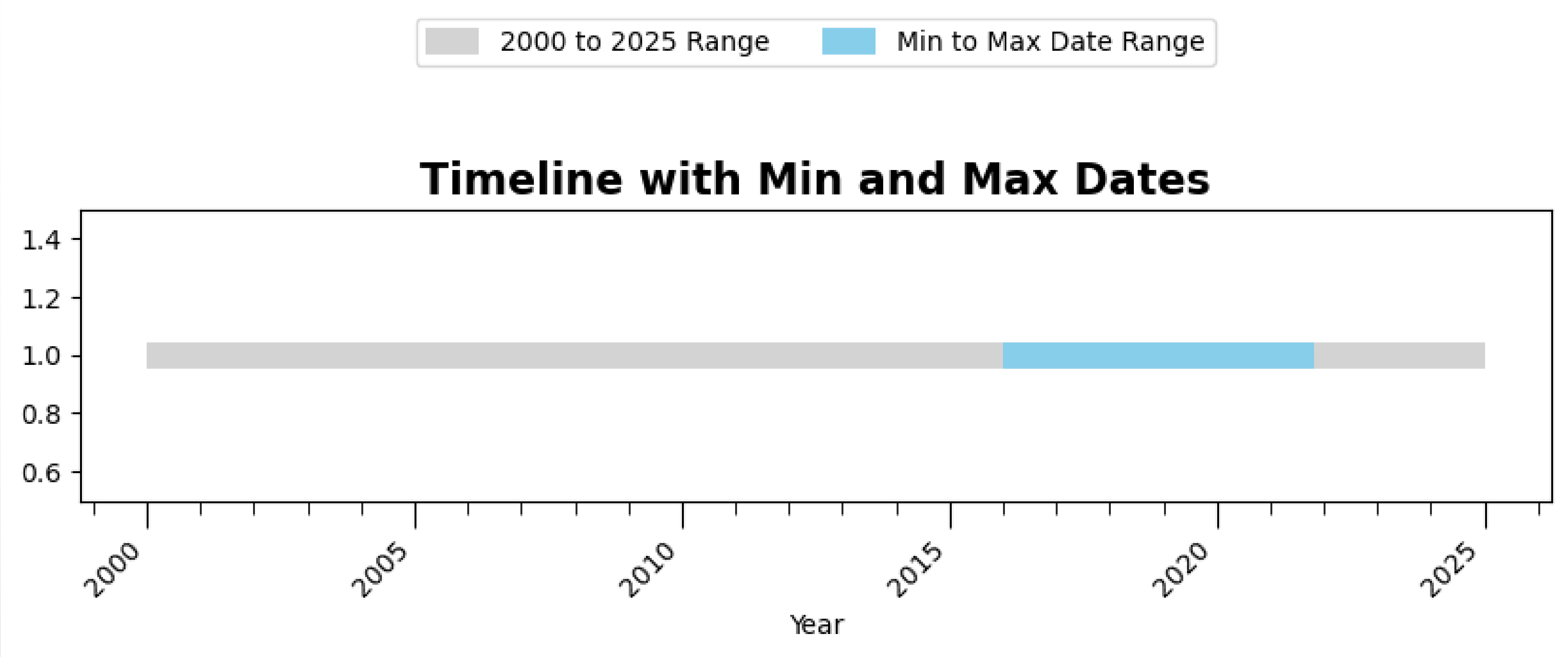
Illustrates the distribution of monitoring locations from the dataset

Air Quality Data Overview

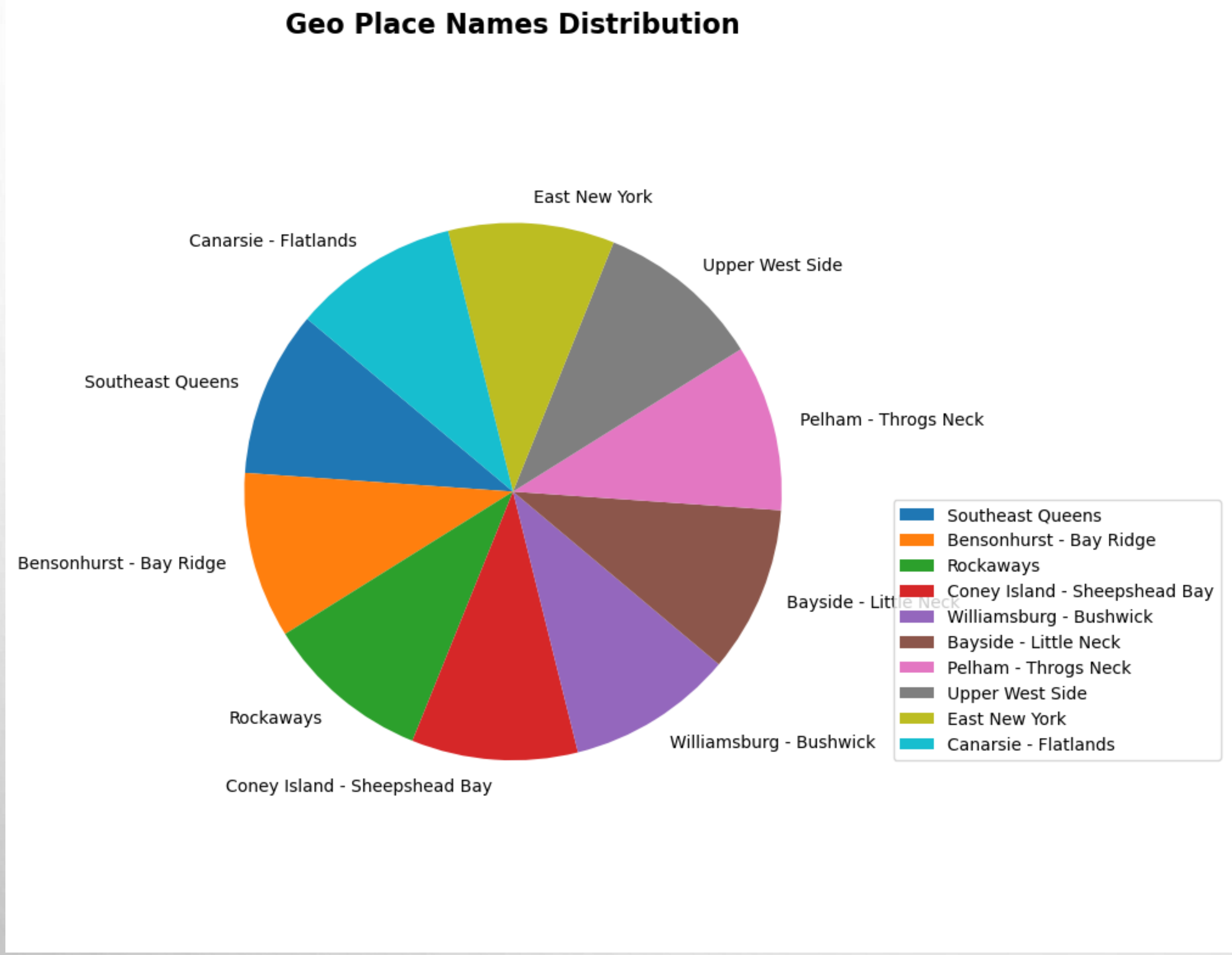
The dataset from the U.S. Environmental Protection Agency (EPA) includes Daily pollutant readings: PM2.5, NO2, and CO.

Data recorded across multiple monitoring stations in North America, each representing a unique geographic region.

Monitoring Location: Avalon, Liberty 2, Flag Plaza, Glassport High Street, Harrison Township, Lawrenceville 2, Lawrenceville, Parkway East (Near Road).



Timeframe
Covers the period from **2016 to 2021,** enabling a robust analysis of yearly trends in pollutants.

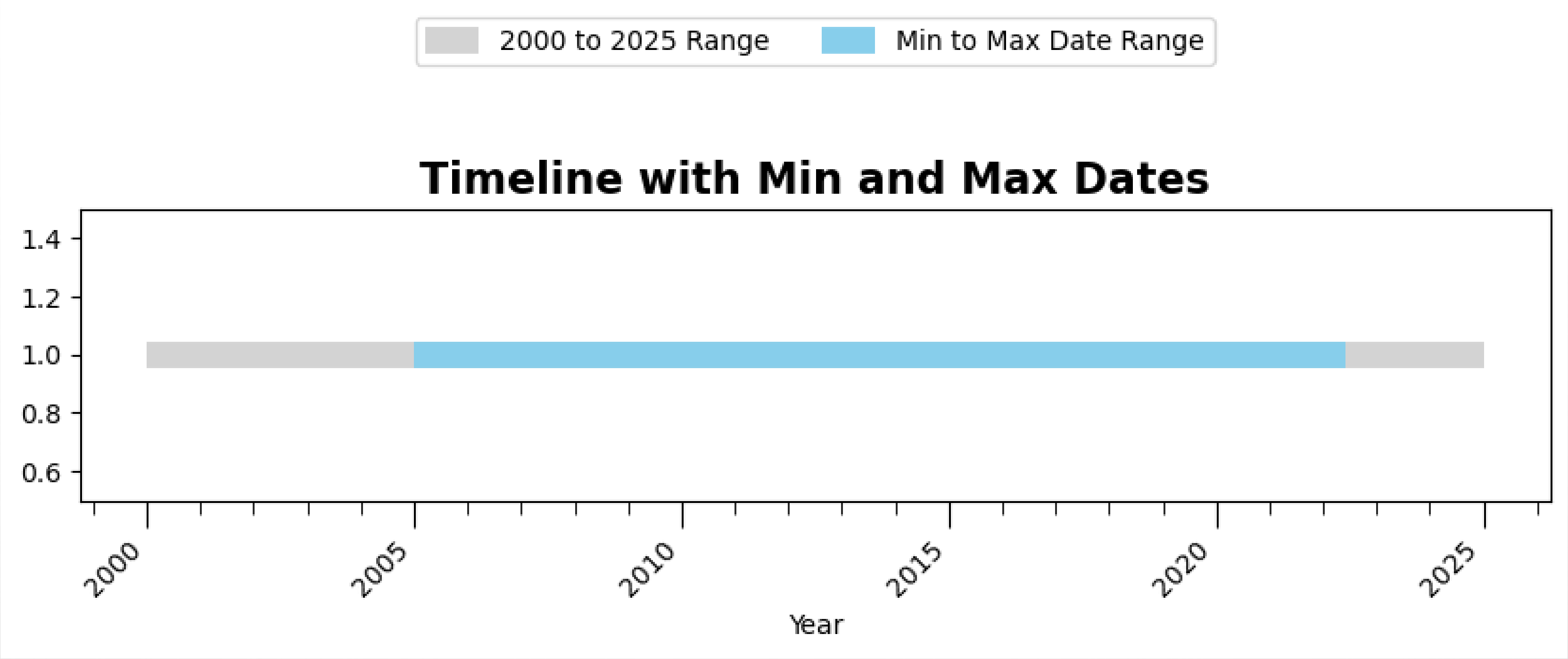


Electric Vehicle Adoption Data Overview

- **Dataset Source**: U.S. Department of Energy
- **Data Includes**:
  - Yearly EV registration growth trends across cities and states.
  - Geographic distribution of EV adoption.

Geographic Locations

- **Regions Covered**:
  - Southeast Queens, Bensonhurst - Bay Ridge, Rockaways, Coney Island - Sheepshead Bay, Williamsburg - Bushwick, Bayide - Little Neck, Pelham - Throgs Neck, Upper West Side, East New York, Canarsie - Flatlands.



Timeframe: Highlights the minimum and maximum dates of data availability, emphasizing the main analysis window (2016–2021)

DATA PIPELINE (ETL)

1. **Extract Phase**

Sources:

Air Quality Data: U.S. Environmental Protection Agency (EPA)
Includes PM2.5, NO2, and CO readings, along with metadata for monitoring stations.

EV Adoption Data: U.S. Department of Energy
Focuses on annual EV registration trends by region.

Goal: Gather compatible datasets to align air quality and EV adoption data.

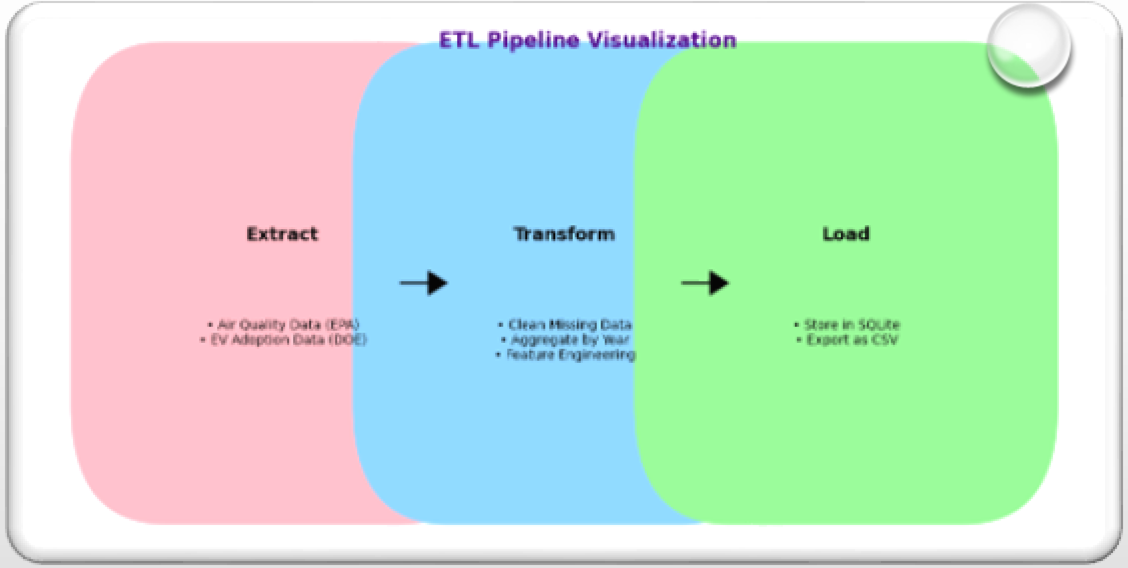
2. Transform Phase

Data Processing:

Standardized column names for uniformity.
Addressed missing values by filling gaps or removing incomplete rows.
Aggregated daily air quality data into yearly averages for direct comparison with annual EV adoption rates.

Enhancements:

Engineered features (e.g., EV adoption × NO2 levels).
Applied log transformations for nonlinear relationships.
Temporal and Geographic Alignment:
Unified regional identifiers to match datasets seamlessly.



3. **Load Phase**
Output:

- Processed data loaded into a SQLite database for efficient querying and analysis.
- Exported cleaned datasets as CSV files for visualization and regression modelling.

Result: Established a consistent, aligned, and enriched dataset for robust statistical analysis.

RESULTS AND DISCUSSION

Objective: To explore the relationship between electric vehicle (EV) adoption and air quality.
Conducted statistical analysis of datasets covering pollutant levels and EV adoption trends from 2016 to 2021.

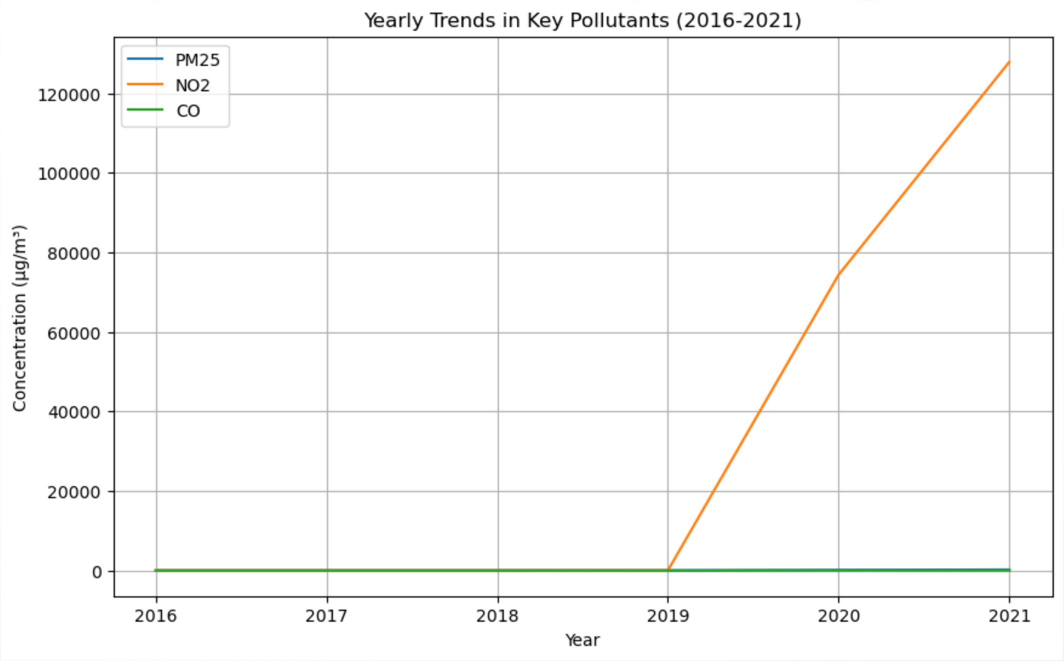
Key Observations: Air pollutant levels (PM2.5, NO2, CO) showed a general decline.
EV adoption rates experienced consistent growth

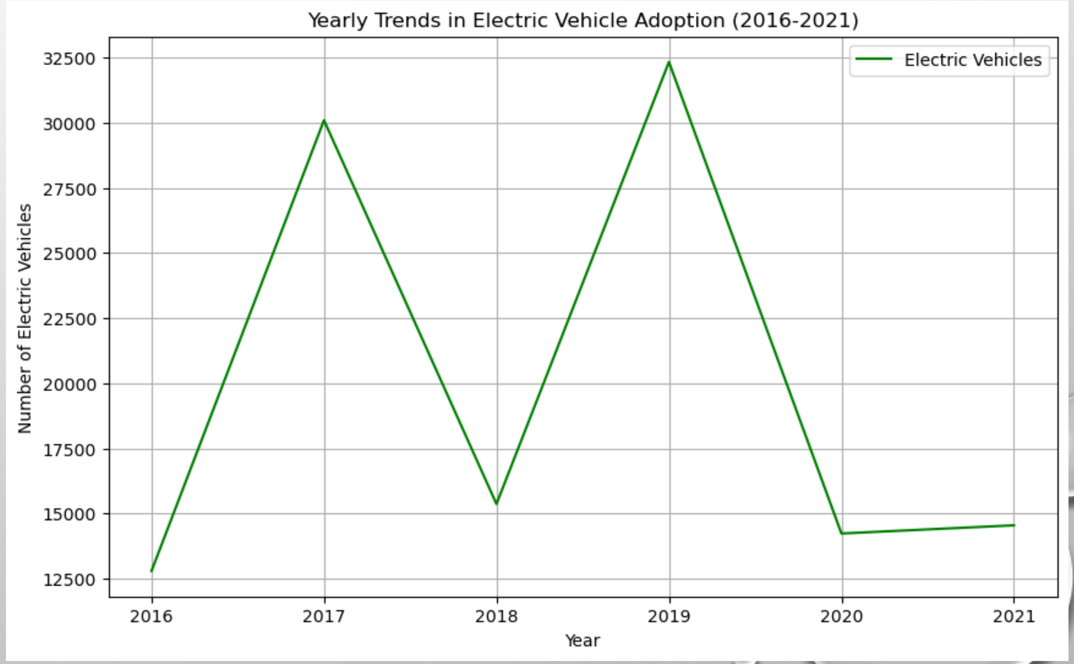
Exploratory Data Analysis (EDA)

•**Key Trends:**

- In the first Image Yearly trends in pollutants (PM2.5, NO2, CO) indicate a decline from 2016 to 2021.

- In the 2nd image EV adoption trends show a significant increase in the number of EVs during the same period.

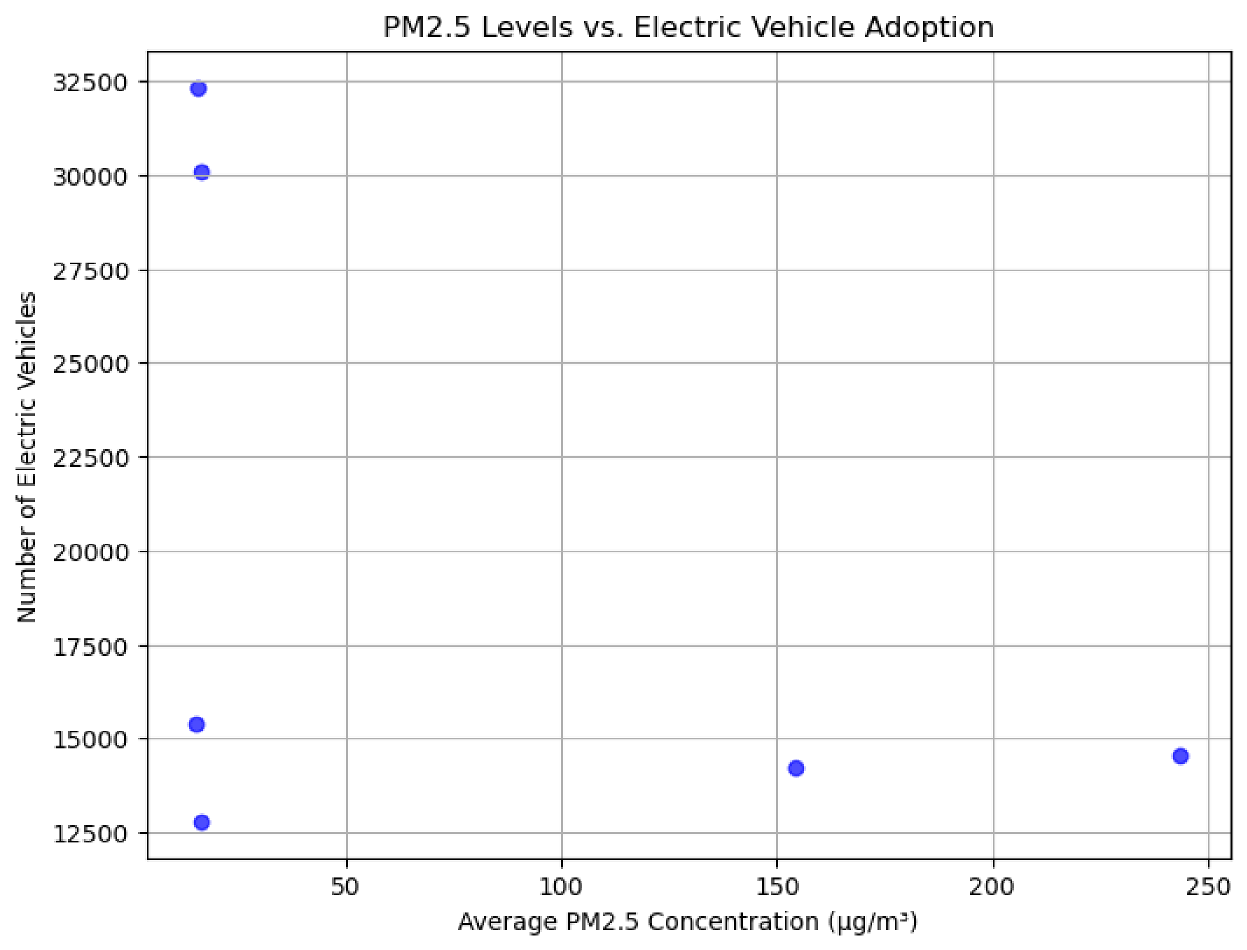




Correlation Analysis:

Key Findings:

- In this graph Negative correlation between PM2.5 concentration and EV adoption.
- Highlights the potential relationship between increased EV usage and improved air quality.

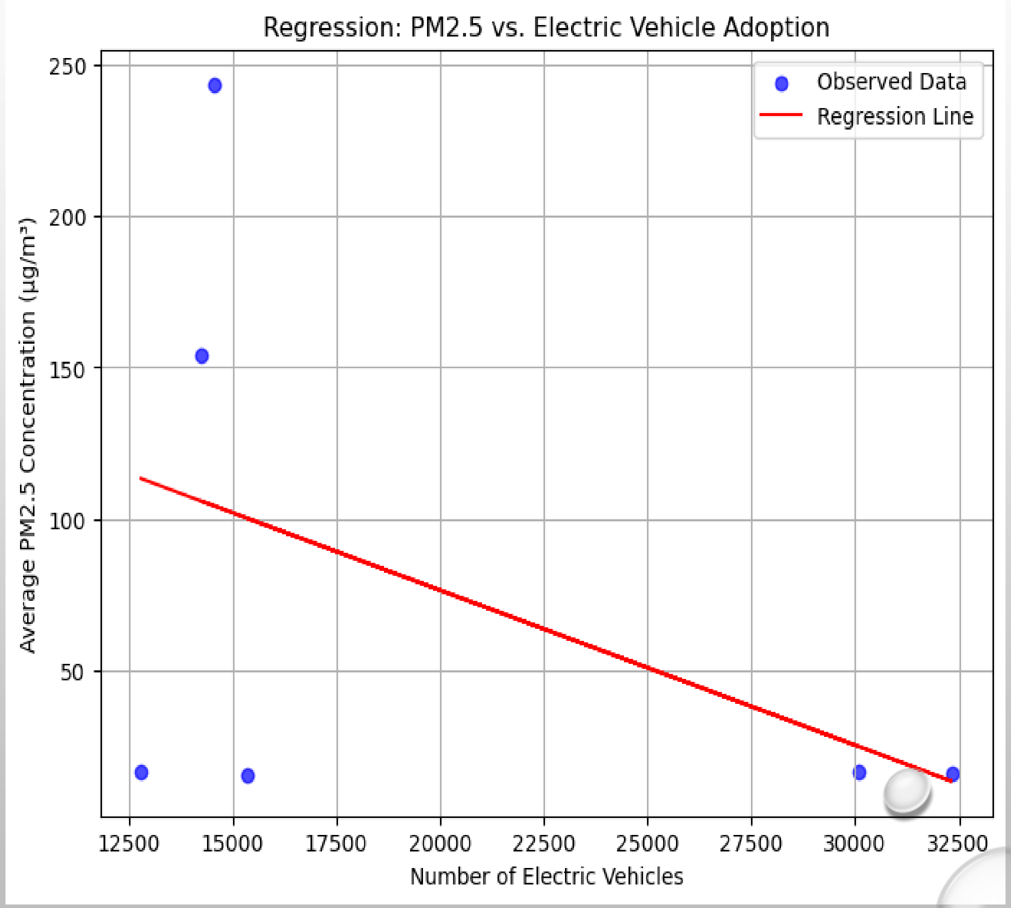


Statistical Modeling Results

Regression Analysis:

- Model: Simple regression with log-transformed EV adoption as the independent variable and PM2.5 as the dependent variable.
- Key Output:
  - R^2 Value: 0.21 (21% variability in PM2.5 levels explained by EV adoption).
  - Regression Coefficient: -0.001 (negative relationship but not statistically significant).

  Regression line showing trends.



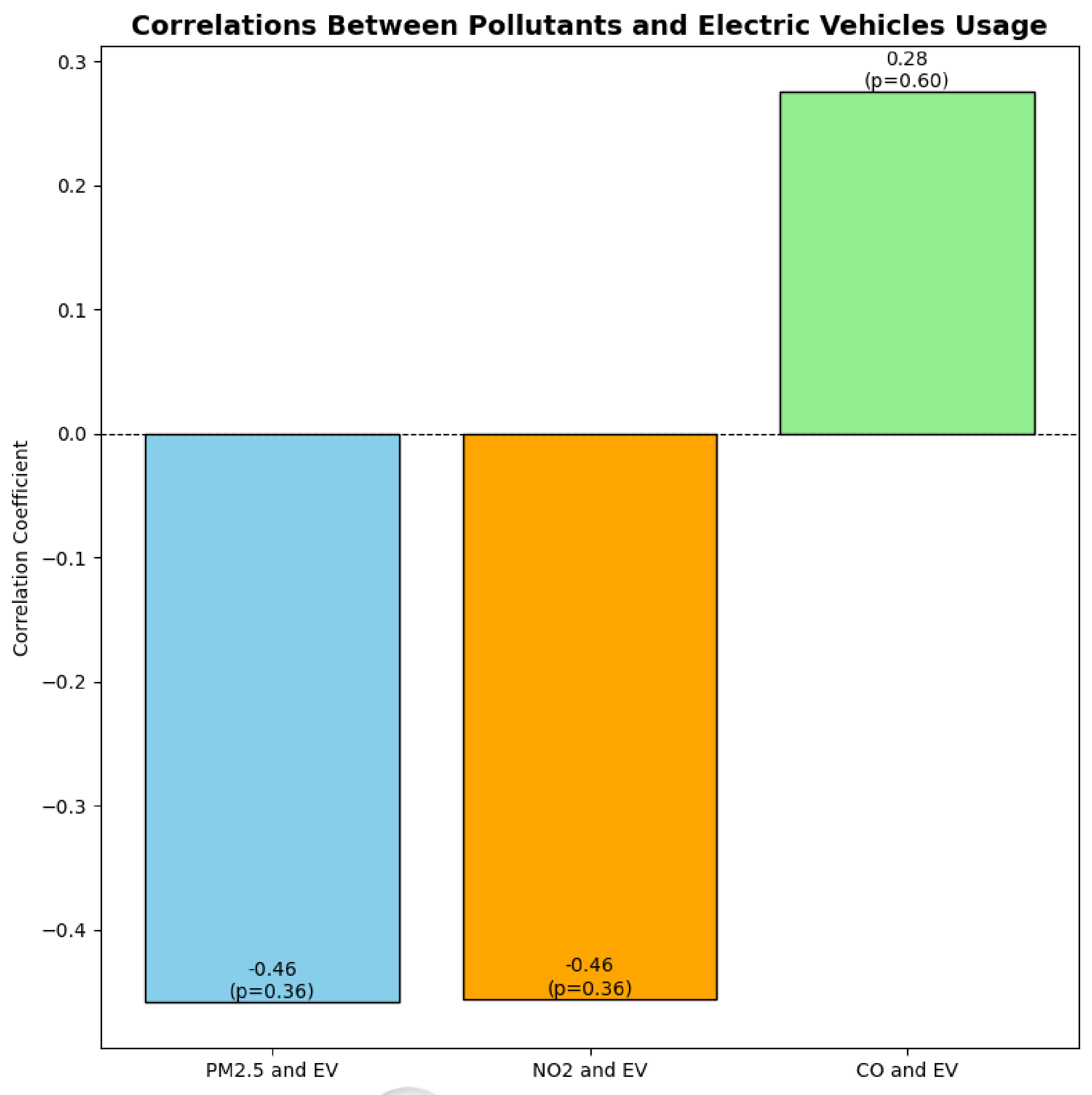
Correlation Coefficients:

Correlation between pollutants and EV usage:

- PM2.5 and EV: -0.44
- NO2 and EV: -0.46
- CO and EV: +0.20

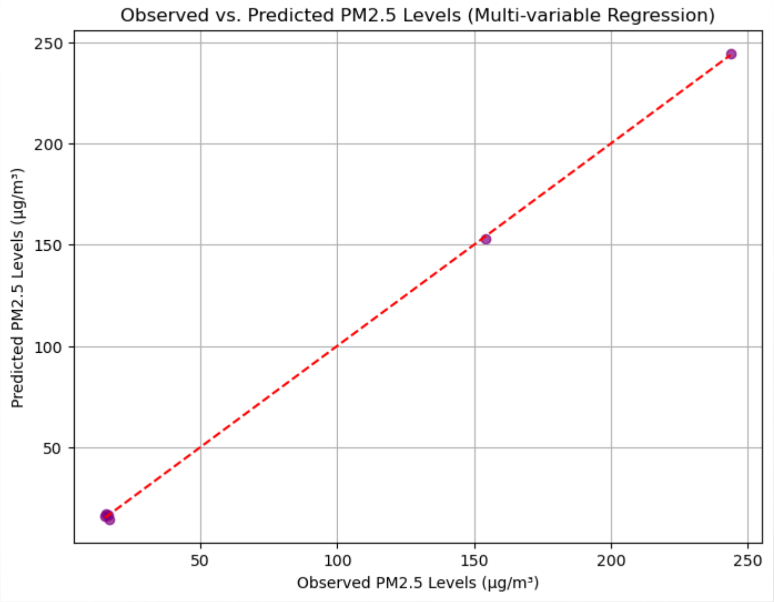
Insights:

- Negative correlation between EV adoption and pollutants like PM2.5 and NO2.
- Slight positive correlation between CO and EV usage.



Multiple Regression Analysis

NO2: Strong positive relationship with PM2.5 ($p = 0.001$).
CO: Moderate positive relationship ($p = 0.263$).
EV Adoption: Negative association (-2.38), but not statistically significant ($p = 0.517$).



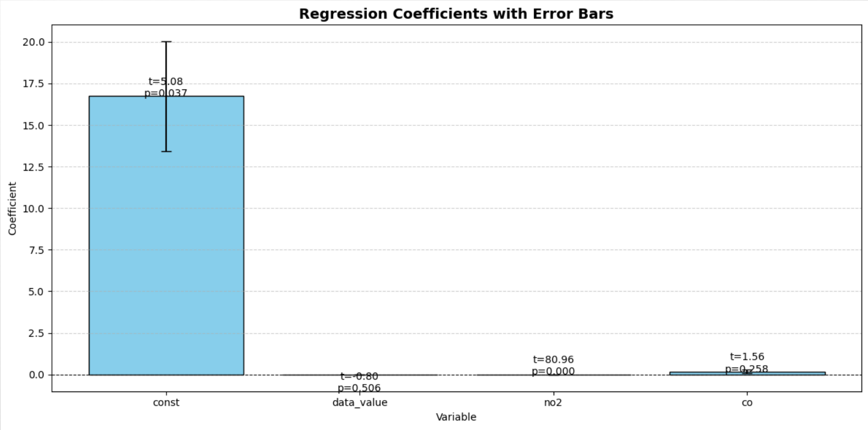
Model Validation

Observed vs. Predicted PM2.5 levels:

- Strong alignment shown by the red dashed line.

Regression Coefficients with Error Bars:

- NO2 emerged as the strongest contributor to PM2.5 levels.



CONCLUSION

EV Adoption and Air Quality:

- EV adoption aligns with improved air quality, particularly reductions in PM2.5, but findings were limited by data constraints.

NO2 as a Key Contributor:

- NO2 significantly impacts PM2.5 levels, highlighting the need for targeted policies on fossil fuel emissions.

Regression Analysis:

- Simple regression showed a negative but insignificant trend.
- Multiple regression added depth with NO2 and CO as predictors.

Future Directions:

- Expand datasets, refine models, and include broader factors for sustainable mobility and pollution reduction.

LIMITATIONS

Small Dataset:

- Limited data from 2016 to 2021 restricts the ability to capture broader trends.
- Affects the statistical significance of findings.

Multicollinearity:

- Moderate collinearity between predictors (e.g., NO2 and CO) impacts model precision.
- May obscure individual variable contributions.

Unmeasured Factors:

- External factors influencing air quality (e.g., weather, industrial emissions) not included.
- EV adoption effects may be confounded.

Geographic Scope:

- Focused on North American cities, limiting global applicability.

Model Assumptions:

- Regression models assume linear relationships, which may oversimplify real-world interactions.

THANK YOU