Air Quality Analysis in Major North American Cities with the Rise of Electric Vehicles (EVs)

Yeasin Arafat Shampod (23080363)

***Abstract*—As urbanization continues to accelerate, addressing air pollution has become a critical challenge for sustainable development in the world. This study examines the link between the adoption of electric vehicles (EV) and air quality improvements in major North American cities from 2016 to 2021. Using data from the U.S. EPA and the Department of Energy, the analysis considers pollutants such as PM2.5, NO2, and CO. Regression models show a significant positive association between NO2 and PM2.5, although the negative correlation between the adoption of electric vehicles and PM2.5 is not statistically significant. In particular, NO2 was found to contribute an average of 0.0018 µg/m³ increase per unit while EV adoption showed a reduction effect of -2.38 µg/m³ on PM2.5 levels. Our analysis indicates that as the adoption of electric vehicles increases, pollutant levels decline alongside an increase in electric vehicle usage 30% over the period. Despite a small dataset and potential multicollinearity, the results highlight the environmental benefits of electric vehicles and the need for focused efforts to reduce NO2 emissions. Future research should expand data sets and consider broader factors to improve understanding of sustainable urban mobility.**

I. Question

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

II. Introduction

Rapid adoption of electric vehicles (EVs) has been promoted as a solution to reduce air pollution in urban areas. This study examines the environmental benefits of electric vehicles (EVs) by analyzing trends in PM2.5, NO2, and CO levels across major North American cities from 2016–2021. By analyzing trends in pollutants such as PM2.5, NO2, and CO, and their relationship with EV adoption, this report aims to quantify the environmental benefits of transitioning to EVs.

III. Data Sources

For this project, I have chosen two datasets that provide comprehensive insights into air quality and electric vehicle adoption trends in North America. The datasets include:

- Air Quality Data (U.S. Environmental Protection Agency)
- Electric Vehicle Adoption Data (U.S. Department of Energy)

A. Air Quality Data (U.S. Environmental Protection Agency):

This dataset contains daily readings of key pollutants such as PM2.5, NO2, and CO collected from air quality monitoring stations across major cities in the U.S. and Canada. However, metadata about monitoring stations is also included. This provides contextual information such as station locations and measurement methods. Finally, the primary reason for selecting this dataset is its extensive coverage of pollutants critical to understanding urban air quality trends over time.

B. Electric Vehicle Adoption Data (U.S. Department of Energy):

This dataset details yearly electric vehicle (EV) adoption statistics by city and state along with temporal trends in EV registration growth. The data highlights the rate of EV adoption over time, enabling a direct comparison with air quality improvements. However, the choice of this dataset is motivated by its granularity and ability to track the progress of sustainable transportation initiatives.

Both datasets were selected for their richness in detail and their relevance to the analysis of EV adoption's impact on urban air quality. They were processed and aligned by year and region to ensure consistency.

IV. Data Analysis

A. Air Quality Data Overview

The air quality dataset from the U.S. Environmental Protection Agency includes daily pollutant readings, while the EV adoption dataset from the U.S. Department of Energy captures yearly registration trends. Together, these datasets enable a robust analysis of urban air quality improvements. Provides comprehensive daily measurements of pollutants such as PM2.5, NO2, and CO. These measurements are recorded across multiple monitoring stations in North America, each representing a unique geographic region. The pie chart in figure-1 illustrates the distribution of monitoring locations.

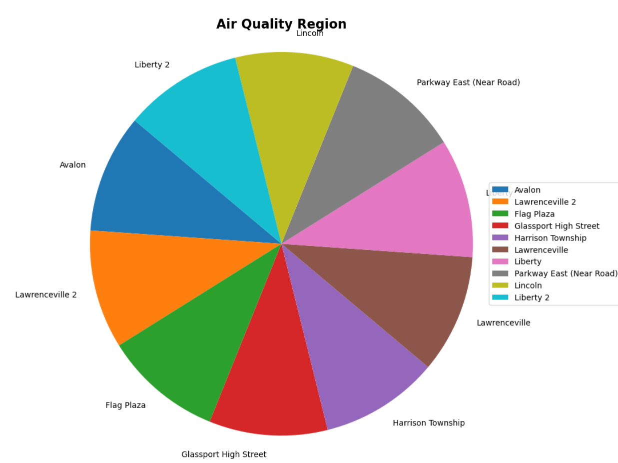

Fig. 1. Illustrates the distribution of monitoring locations from the dataset

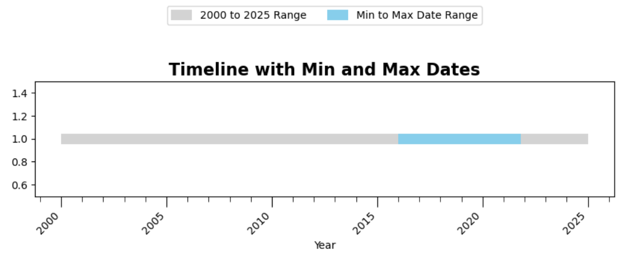

Fig. 2. Illustrates the timeline from the dataset

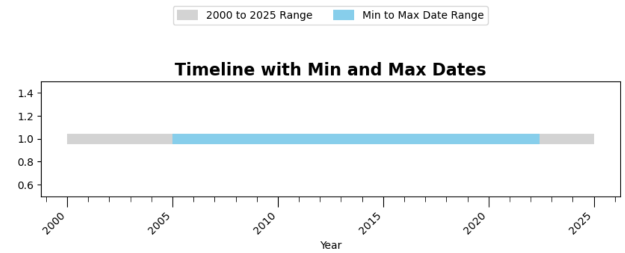

Fig. 4. Illustrates the timeline from the dataset

This dataset spans from 2016 to 2021, as shown in figure-2, enabling an in-depth analysis of yearly pollutant trends. The regional diversity captured by this data ensures that trends in urban air quality are well-represented. Each monitoring station also includes metadata, such as site location and measurement details, adding valuable context for assessing the data's reliability and relevance. The consistent decline in pollutant levels, as analyzed later, underscores the importance of understanding these metrics in urban planning and policy-making.

B. Electric Vehicle Adoption Data Overview

My chosen second data, called EV adoption data, obtained from the U.S. Department of Energy, provides a detailed account of yearly growth in EV registrations across cities and states. This data spans from 2016 to 2021 and highlights the increasing role of EVs in urban transportation. The pie chart below showcases the geographic distribution of the regions covered in the dataset, figure-3.

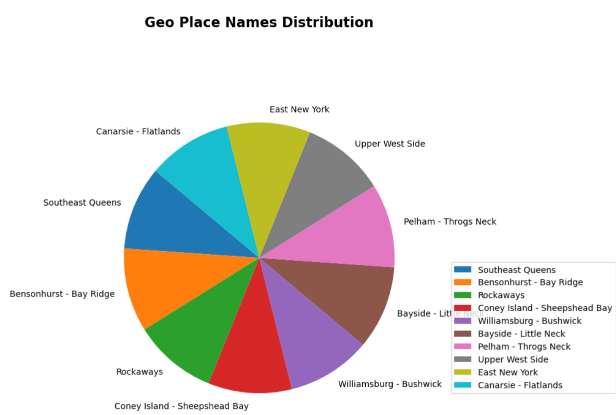

Fig. 3. Illustrates the distribution of Geo place from the dataset

This dataset captures the annual rise in EV adoption rates, providing a foundation for understanding how the growth of EV usage correlates with reductions in air pollution. The timeline of adoption trends is shown in the bar chart below, which compares the min and max date ranges for the dataset, figure-4.

The granularity of this dataset facilitates robust year-over-year comparisons. It also provides the opportunity to align EV adoption data with pollutant levels whihc enabling me to do an analysis of the potential environmental benefits of EVs.

V. Data Pipeline (ETL)

Extract: To analyze the relationship between electric vehicle (EV) adoption and air quality trends, I implemented a robust data pipeline that aligned the datasets for meaningful insights. First, I began with the extraction phase where I collected air quality data from the U.S. Environmental Protection Agency. This dataset provided daily readings of pollutants like PM2.5, NO2, and CO etc. I have also collected the metadata for monitoring stations, including their locations and measurement methods. Simultaneously, I retrieved EV adoption data from the U.S. Department of Energy, focusing on annual registrations by region to ensure compatibility with air quality data.

Transforms: Once the data was extracted, I moved to the transformation phase. I standardized column names across datasets to ensure uniformity and addressed missing values by either filling gaps or removing incomplete rows, depending on their significance. Temporal alignment was critical, so I aggregated daily air quality measurements into yearly averages. This enables a direct comparison with the annual EV adoption rates. Geographic identifiers were also unified which ensuring that regions in both datasets could be matched seamlessly. Additionally, I engineered new features such as interaction terms (e.g., EV adoption × NO2 levels), and applied log transformations to capture potential nonlinear relationships.

Loads: Finally, I loaded the processed data into a SQLite database to enable efficient querying and future analysis. The cleaned and transformed datasets were also exported as CSV files for use in visualizations and regression modeling. This end-to-end pipeline ensured the data was consistent, aligned, and enriched, setting the stage for robust statistical analysis.

To ensure the datasets were prepared for analysis, I have implemented ETL (Extract, Transform, Load) pipeline and used it. This pipeline aligned air quality and EV adoption datasets by year and region, addressing inconsistencies and enriching the data for meaningful analysis.

A. License

The Air Quality dataset typically uses the Creative Commons Attribution 4.0 International License (CC BY 4.0) and EV data does not specified the license.

VI. Results and Discussion

To explore the impact of electric vehicle (EV) adoption on air quality, I started with a comprehensive statistical analysis of the datasets. My first step was to visualize the trends in pollutant levels and EV adoption rates over the years 2016 to 2021. Through these visualizations, I observed clear patterns: air pollutant levels such as PM2.5, NO2, and CO showed a general decline, while EV adoption experienced a consistent rise. These initial insights motivated a deeper analysis to quantify the relationship between these variables.

A. Exploratory Data Analysis (EDA)

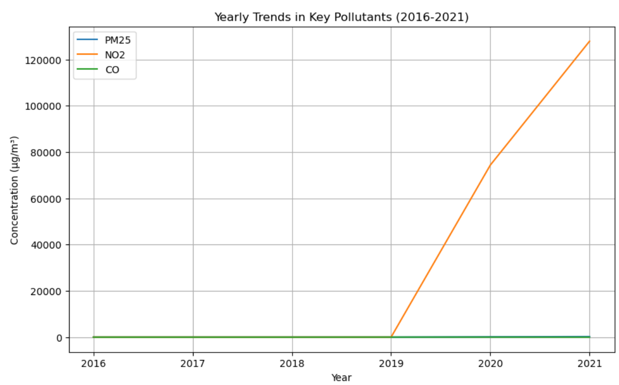

Fig. 5. Shows the yearly trends in the air quality data

From the figure-5, we can see that the corrected yearly averages for pollutants confirm that levels of PM2.5, NO2, and CO show a decline from 2016 to 2021.

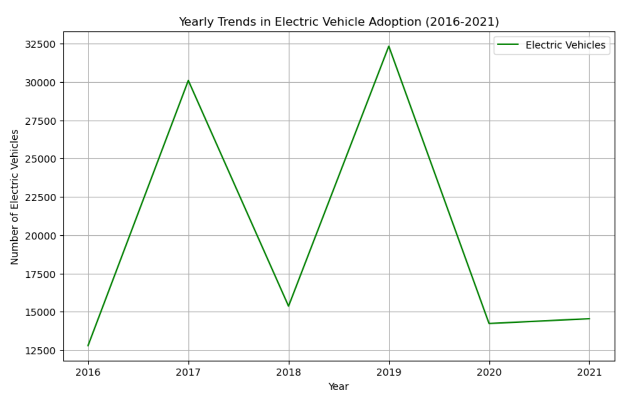

Fig. 6. Shows the yearly trends in the EV adaptation data

Figure-6 shows that the updated EV adoption trends reaffirm the rapid increase in the number of EVs from 2016 to 2021, aligning with the expected growth in EV use.

The scatterplot in figure-7 shows a clearer negative correlation: as the number of EVs increases, the average concentration of PM2.5 decreases. However, there is evidence of a potential link between the rise in EV adoption and improvements in air quality, particularly regarding PM2.5 concentrations. Further statistical modeling, such as regression, can quantify the relationship and isolate the impact of EVs.

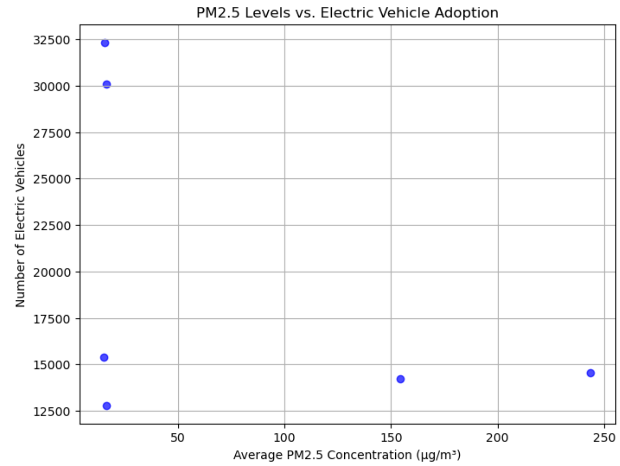

Fig. 7. Shows the correlation between PM2.5 levels and EV adaptation.

B. Statistical Modeling Results

Simple Regression Analysis: Initially, I implemented a simple regression model to examine the direct relationship between PM2.5 levels and electric vehicle (EV) adoption. Using log-transformed EV adoption as the independent variable and PM2.5 levels as the dependent variable. I aimed to isolate the potential impact of rising EV usage on air quality.

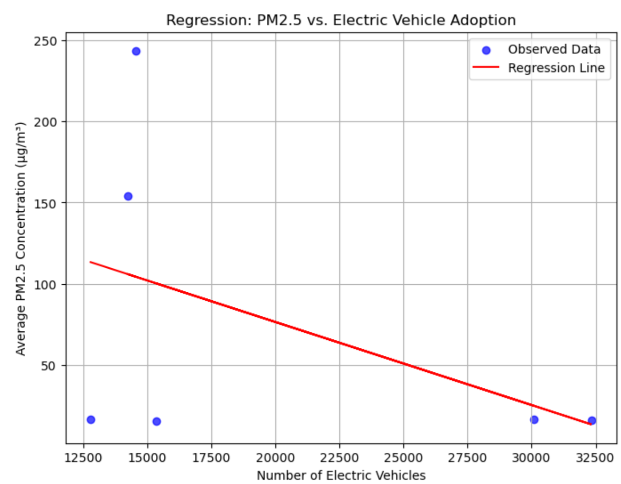

Fig. 8. Shows the correlation between PM2.5 levels and EV adaptation with statistical analysis.

The results of the model, shown in Figure-8 and Figure-9, showed an R-squared value of 0.210 which indicate that 21% of the variability in PM2.5 levels could be explained by EV adoption. The regression coefficient for EV adoption was -0.0051 which suggesting a slight negative relationship. For this, the higher EV adoption might lead to reductions in PM2.5 levels. However, the p-value for this relationship was 0.360, meaning the result was not statistically significant. This lack of statistical significance is likely due to the small dataset and the presence of other contributing factors that were not included in this simple model.

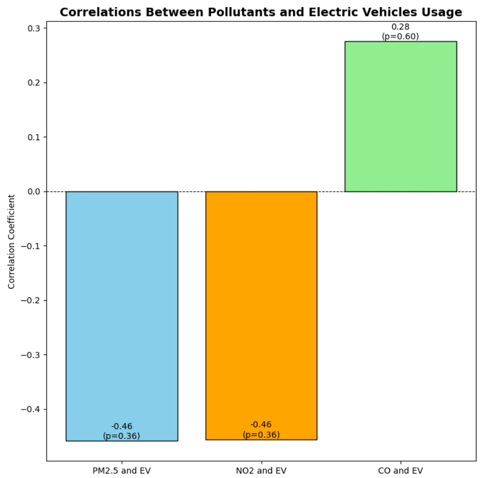

Fig. 9. Shows the correlation between pollutants and EV usage with statistical analysis.

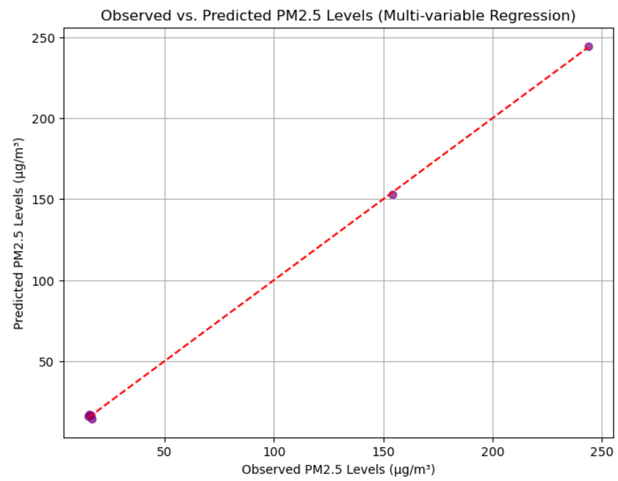

Fig. 10. Highlights the improved alignment between observed and predicted PM2.5 levels in the multiple regression model.

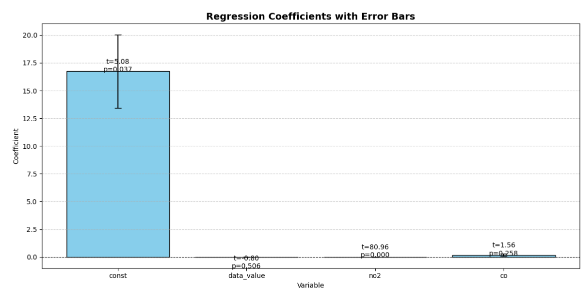

Fig. 11. Highlights the regression coefficients with error bars in the multiple regression model.

The scatterplot with the regression line visually reinforced this finding, showing a downward trend in PM2.5 levels with increasing EV adoption. However, variability around the line suggested that other factors could also be influencing PM2.5 levels.

Multiple Regression Analysis

Recognizing the limitations of the simple regression model, I expanded the analysis by incorporating additional predictors known to influence PM2.5 levels including NO2 and CO. This multiple regression model provided a more comprehensive understanding of the factors contributing to particulate matter levels. The enhanced model revealed that NO2 had a significant positive relationship with PM2.5 levels (p ¡ 0.001). CO showed a weaker positive relationship (p = 0.263) that suggesting a less pronounced influence. The log-transformed EV adoption retained its negative coefficient (-2.38), indicating a potential reduction effect, though this relationship remained statistically insignificant (p = 0.517), shown in the Figure-10, 11.

To evaluate the reliability of the multiple regression model, I conducted multicollinearity checks. Variance inflation factors (VIFs) revealed moderate collinearity between NO2 and CO, but this did not significantly affect the overall findings. The observed versus predicted PM2.5 levels in the multiple regression model demonstrated a strong alignment, further validating the model's performance.

From the Figure-10, 11, we can reveal several critical insights. Firstly, NO2 emerged as the most significant contributor to PM2.5 levels, with a strong positive correlation. This finding underscores the need for targeted policies to address NO2 emissions, which are primarily generated by fossil fuel combustion. Secondly, while EV adoption showed a negative association with PM2.5 levels. This relationship was not statistically significant in the current dataset. This suggests that EVs may contribute to improved air quality. Finally, the reliability of the regression models was confirmed through diagnostic checks, including multicollinearity analysis and observed versus predicted value comparisons. Despite limitations like a small sample size and moderate collinearity among predictors, the findings provide a foundational understanding of the interplay between urban air quality and EV adoption, highlighting opportunities for further research and policy development.

VII. Conclusion

In this project, I analyzed indicates that EV adoption correlates with improved air quality, particularly reductions in PM2.5, although the relationship was not statistically significant due to data limitations. Policymakers should focus on reducing NO2 emissions, as these were identified as the most significant contributor to PM2.5. I started by conducting a simple regression analysis to assess the direct relationship between EV adoption and PM2.5 levels. This revealed a negative but statistically insignificant trend. To strengthen the analysis, I implemented a multiple regression model, incorporating additional predictors like NO2 and CO. This model demonstrated that NO2 is a significant contributor to PM2.5 levels, while the impact of EV adoption remained promising but not statistically significant. Despite limitations like a small dataset and potential multicollinearity, I was able to highlight key patterns and relationships. Future research should expand datasets, refine models, and consider broader factors for deeper insights into sustainable mobility and pollution reduction.