Data Report

November 28, 2024

1. Question How has the adoption of electric vehicles in Washington State, particularly in Seattle, influenced air quality over time?

2. Data Sources

2.1 Electric Vehicle Population Size History By County (WA State)

- **Metadata URL:** EV Population
- **Data URL:** Download EV Population
- **Data Type:** CSV
- **Description:** This dataset details the monthly vehicle registrations recorded by the Washington State Department of Licensing (DOL), which combines the data from the National Highway Traffic Safety Administration (NHTSA) and the Environmental Protection Agency (EPA) fuel efficiency ratings with its own titling and registration records.
- **Data Structure and Quality:** Data includes historical Electric Vehicle population counts from January 2017 through October 2024. It categorizes the data by county of Washington State, distinguishing between passenger vehicles and trucks.
- **License and Obligations:** Open Database License. It is stated in the Metadata. To comply with the obligation, appropriate credit to the Washington State Open Data Portal will be given and the link to the License will be provided.

2.2 Air Quality System (AQS) Data

- **Metadata URL:** Air Data Description
- **Data URL:** For each year and each substance a different zip file has to be downloaded from Air Data
- **Data Type:** zip folder containing csv file.
- **Description:** The Air Quality System (AQS) contains ambient air pollution data collected by the Environmental Protection Agency (EPA), state, local, and tribal air pollution control agencies from over thousands of monitors.
- **Data Structure and Quality:** For our study Daily Summary Files will be used. CSV files that aggregate the measurements by day. There are up to 29 different variables or columns, many of which might be redundant for our use, like the different ways to express the location

of the monitor. We will use 4 substances: Carbon Monoxide (CO), Nitrogen Dioxide (NO2), PM2.5 Speciation and PM10 Speciation

- **License and Obligations:** Public Domain License. The License statement can be found here. To comply with the obligation, citation will be included as mentioned here.

3. Data Pipeline The whole pipeline is implemented in Python.

- **3.1 EV Population Dataset:** A direct download link of the CSV file is available, allowing an easy extraction with Pandas. The data is then filtered by *state = 'WA'* before saving it to *../data/electric_car_population.csv*
- **3.2 AQS Data:** This data requires a more sophisticated extraction. The data is grouped by year and by substance, each in a different zip file, which in turn contains a CSV file. We will use 4 substances: Carbon Monoxide (CO), Nitrogen Dioxide (NO2), PM2.5 Speciation and PM10 Speciation. Since the EV Population Data goes from 2017 to 2024 (range of 7 years), we need to request at least 28 datasets. In the pipeline, a *../data/temp/* folder is created where the zip files are downloaded and extracted. For each of the 4 substances then yearly data frames are concatenated and saved in 4 different CSV files inside *../data/*.

4. Result and Limitations The output is 5 different CSV files. All of them are complete, consistent and relevant for the initial question, however, during the data exploration some doubts arose:

- **EV Population:** The percentage of electric vehicles in Washington State has been rising monotonically since 2017, but by 2024 we only reached around 10% of the total passenger vehicles. The truck population is even lower being around 1%.
- **AQS Data:** On certain days there are extreme spikes for certain substances, which makes it question the accuracy of the measurements, and also the contribution to a certain level of a certain substance might have many different causes, making this study more challenging than initially thought.