The Impact of Electric Vehicles on Air Quality

Rohan Kanagal Sathyanarayana
National College of Ireland
School of Computing
Dublin, Ireland
x19203829@student.ncirl.ie

Aravind Hallimysore Kalegowda
National College of Ireland
School of Computing
Dublin, Ireland
x22104275@student.ncirl.ie

Abstract— It is generally presumed that increasing the use of electric cars (EVs) will decrease air pollution and benefit public health. The quantity of EVs on the road and the underlying sources of air pollution in a specific location are two factors that affect how much air pollution EVs can truly help reduce, thus in this assessment, our objective with this data visualization is to learn more about how EV use affects air quality. Henceforth for the analysis we have used two datasets, where the air pollution dataset is unstructured (XML) and electric vehicle population dataset is structured [source: https://catalog.data.gov/]. We fetched the data from API stored in MongoDB and then transformed the data and loaded it into PostgreSQL for visualization. The relationship between EV's and Air pollution along with its pattern of distribution will be illustrated.

Keywords— Air quality, Electric Vehicle, ETL

Introduction

Due to the detrimental effects that transportation has had in recent years on both the environment and public health, the shift towards more environmentally friendly modes of transportation has emerged as an issue of growing significance **[1].** Because electric cars (EVs) have the potential to cut emissions and improve air quality, they have emerged as a viable alternative to traditional gasoline-powered vehicles in recent years. Currently, to direct policy choices and help assist the transition towards sustainable mobility, it is vital to conduct an accurate analysis of the effect electric vehicles have on air pollution **[2].** Using methods of machine learning which involves data visualization, the evaluation will attempt to investigate the effect that EVs have on levels of air pollution. This research underlines the necessity for precise projections of the effect that electric vehicles will have on air pollution to inform policy choices and help the transition towards more sustainable modes of transportation.

According to the findings of the study, the effect that electric vehicles have on the amount of pollution released into the atmosphere is likely to change depending on several factors. These include the type of electricity that is used to charge the vehicles, how efficient the vehicles are, and the road conditions. In this evaluation, we will be utilizing various types of datasets, one of which will be structured and the other of which will be unstructured **[3].** As a result, we will extract the data and carry out the analysis to determine which nation has the lowest levels of air pollution as a direct result of the growing popularity of electric vehicles in that region. This assessment presents an innovative method for measuring the effect of electric vehicles (EVs) on air pollution by making use of machine learning strategies. This research underlines the necessity for precise projections of the effect that electric vehicles will have on air pollution to inform policy choices and help the transition towards more sustainable modes of transportation. Datasets linked to Air quality and Electric vehicle population, an unstructured XML dataset which contained air pollution statistics retrieved through an API and structured CSV dataset which contained EV's statistics respectively were sourced from the US government data site, The chosen datasets are stored in unstructured and structured formats, respectively, in the NoSQL databases Mongo DB and PostgreSQL and further used for visualization using the python libraries.

.

I. Related Work

In this research **[4],** the synergy as well as co-benefits of lowering CO2 or even air pollutant levels in Shanghai through employing electric individual automobiles, taxis, as well as buses are studied. The research was carried out by the University of Hong Kong. To detect and assess the produced co-benefits, the co-control coordinates systems as well as pollutants reductions cross-elasticity (Elsa/b) have been used in this paper. The unit's pollutants equivalents (Apeq) reduction costs of reaching synergies among three different kinds of cars are evaluated based on the complete operational life - cycle cost. When potential benefits as well as expenditure studies are combined, as they are here, they allow for discussion of such co-benefits provided by employing the three different kinds of electric cars. According to the findings (Alimujiang & Jiang, 2020), the greatest number of co-benefits are generated by public transport. Hence, switching from conventional fuel cars to electric buses is one way to cut down on both air quality as well as CO2 pollutants at the same time. If we consider the subsidies that the governments provide to electric cars, we can see that the implementation of public transport in Shanghai would have significant positive effects on both the surroundings as well as the economy. In contrast, electric automobiles used for personal transportation as well as taxis have the additional advantage of lowering pollutants of CO, PM10, NMHC, and NOx. Though without considering government subsidies, the use of electric cabs in Shanghai has a significantly positive impact on the city's economy. This research **[5],** seeks to use advanced ML concepts, extreme gradient boosting, as well as to make a comparison with conventional machine learning modelling techniques, multiple linear regression, light gradient boosting machine, as well as artificial neural network to address the problem of accurately predicting the amount of energy consumed by electric vehicles. Japan's Aichi Prefecture was the location where the data used in the investigation about the energy usage of electric cars were gathered. The effectiveness of the estimation methods was evaluated using three different metrics: and the mean absolute error, the root means square error, the coefficient of determination (R2). The conclusion of the analysis shows that the machines that use extreme gradient boosting as well as mild gradient boosting delivered more accurate and reliable outcomes compared to those that use multiple linear regression models as well as artificial neural networks**[6].**The models that are built on stochastic gradient boosting as well as light gradient boosting machines have been shown to have greater significance of R2,higher values of root mean square error and lower values of mean absolute error. These results indicate that these predictions are much more efficient. Yet, the findings showed that the moderate gradient boosting machine executed much better than the extreme gradient boosting models. To illustrate the effect but also comparative significance of various input factors on the forecast of the amount of energy used by electric cars, a comprehensive functionality evaluation was performed. Based on the findings, it appears that such enhanced machine learning approach has the potential to improve the forecast accuracy of energy usage by electric cars.

This study **[7]** emphasizes the need to make the switch to electric cars as soon as possible to cut down on emissions and enhance the quality of the air we breathe. According to the findings of the research, it is critical to have a solid grasp of the elements that determine how electric cars affect levels of air pollution. According to the findings of the research, switching from traditional gasoline-powered cars to electric vehicles may drastically cut down on the amount of air pollution produced in metropolitan areas **[8].** According to the findings of the research, electric cars have zero emissions coming out of their tailpipes when they are in operation. This leads to a decrease in the releases of pollutant concentrations such as PM2.5 as well as NOx. The research also emphasized the fact that the influence of electric cars on air pollution is dependent on various variables such as the source of the power used to charge the vehicles, the efficiency of the vehicles themselves, and the circumstances under which they are driven. In addition, the report **[9]** places an emphasis on the need for the development of charging infrastructure as well as the accessibility of other energy sources to facilitate an even greater uptake of electric cars. According to the findings of the research, electric cars could greatly cut down on air pollution and enhance public health. Nevertheless, the study also found that the effect of electric vehicles might vary depending on the local circumstances and legislation.

II. METHODOLOGY

A. DATASET DESCRIPTION

The current outlook of electric vehicle adoption in the state may be better understood with the help of the electric vehicle dataset, which displays the number of Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) registered with the United States Department of Licensing (DOL). The "Air Pollution Data" collection represents information on the data collected during the surveillance of the air quality in New York City. Air quality readings for pollutants were collected from the Department of Health and Mental Hygiene (DOHMH). These datasets were chosen because they help us better understand how broad adoption of EVs would affect air quality. Both the EV dataset and the air pollution dataset are useful resources for understanding the current state of the EV market and pollution levels in the United States, respectively. By merging these data sets, we can identify the areas where the widespread use of electric cars would have the greatest impact on lowering air pollution levels.

B. DESCRIPTIONS AND JUSTIFICATIONS OF DATABASE

While working on this project, we extracted, transformed, and loaded data from a variety of sources using a variety of different data processing packages and methods. We extracted XML data from API endpoints by using the requests library to send GET queries to those endpoints. After that, we made use of the Panda's library to construct data frames and carry out various operations on the data, such as filtering and merging. In addition to that, we made use of matplotlib to generate graphical representations of the data, such as bar graphs, that assisted us in comprehending the connections that exist between the numerous data variables. We made use of both MongoDB and PostgreSQL for the storage of and management of our databases. We connected to the MongoDB database by utilizing the Pymongo module, and then we inserted data into the collections that were relevant to those connections. When working with PostgreSQL, we established connections to the database, created tables, and inserted data into them by utilizing the psycopg2 package. All these actions involving processing data were carried out to derive insights and links between the many variables contained in the data, as well as to carry out a variety of studies to investigate the effect that electric vehicles have on air pollution.

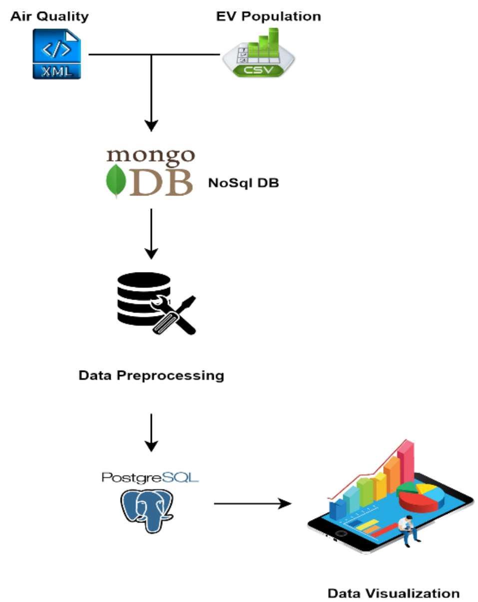

Fig.1 Dataflow

C. Descriptions and Justifications of Programming Language and Library Used

Python: Python is a widely used programming language for data science, and it contains a significant number of libraries for data analysis, data visualization, and machine learning.

Pandas is a library for the Python programming language that allows users to execute sophisticated operations for manipulating and analyzing data.

Matplotlib: Data visualization is a common use of the Python package known as Matplotlib, which can be found on many computers. It gives users the ability to build high-quality visualizations by giving a variety of plot types and customizing options to choose from.

Pymongo: Pymongo is the official MongoDB library for Python. It offers a user interface that is straightforward and uncomplicated, and it may be used to perform operations on MongoDB databases **[6]**. For dealing with unstructured data such as XML documents, which are often used in NoSQL databases such as MongoDB, this is a smart option to go with. Working with massive datasets is made simpler and more flexible because of this feature's availability.

Requests: The requests library is a Python library that is frequently used for sending HTTP requests to application programming interfaces (APIs) **[2].** It has a user interface that is intuitive and uncomplicated to use with, making it possible to send GET, POST, and other HTTP queries.

Psycopg2: psycopg2 is the most popular PostgreSQL database adapter for Python. It provides a robust and efficient way to connect to PostgreSQL databases and execute SQL queries.

In the analysis we used mongo dB and PostgreSQL to upload all the necessary dataset, for doing so we used Pymongo, psycopg2, SQL alchemy, request. First, we need to connect our python to mongo dB server, therefore we have used Pymongo library in the below image Fig.2 we can see that we have connected to our mongo dB serve by using connection URL and we can also see that we have successfully connected to the server.

```
#takes an air quality API url as an input parameter.
client=pymongo.MongoClient("mongodb://localhost:27017")
#establishing a connection to a local MongoDB server using default port(27017)
db = client['dbairpollution']
collection = db['airpoll_xml']
```

Fig. 2

the next step is to fetch all the required data using API for that we need to import request library and for fetching the data we have used request. Get method and pass the URL to the data which is in XML format now we have fetched the data and successfully stored it in a variable.Fig.3 depicts it.

```
import requests
response = requests.get("https://data.cityofnewyork.us/api/views/c3uy-2p5r/rows.xml?accessType=DOWNLOAD")
    #requesting the air quality data and then saving it in the variable
    #data is returned as xml file and then saved in 'Air_quality.xml' file
with open('Air_quality.xml', 'wb') as file:
    file.write(response.content)
air_quality_df = ET.parse("Air_quality.xml").getroot() #parsing the xml file and extracting the data from
```

Fig.3

As seen in Fig.3, the 'with open' function is used to create a new file named "Air_quality.xml," which is then written to using the contents of the response variable; the ‘response' object most likely contains data retrieved from the API. Finally, the root element of the XML document is extracted using the "ET. parse" function, and the extracted value is then stored in the "air_quality_df" variable for further processing.

The resulting object are then extracted from the hierarchy of the XML file by indicating the path that contains specific entity and then empty list is passed further looping through all the elements in the data that matches the path specified in Fig.4.The XML data is then processed and converted into a panda's Data Frame using the panda's library where we carry transform process after extracting the data. The function then uses the Pymongo library to connect to a MongoDB database and inserts the data from the Data Frame into a specific collection within the database. Finally, the function reads the data from the MongoDB collection and saves it to a CSV file named data_air.csv. Overall, this function streamlines the process of collecting and storing data from an API endpoint, making it easier to work with large datasets and perform more advanced analysis.

```python

#defining the paths for easy extraction of the data from the hierarchy of the XML file
ID_PATH          = ".//row/row/unique_id"
indicator_id_path = ".//row/row/indicator_id" #indicates an XML file path that contains specific entity,such as indic
name_path        = ".//row/row/name"
measure_path     = ".//row/row/measure"
measure_info_path = ".//row/row/measure_info"
geo_type_path    = ".//row/row/geo_type_name"
geo_id_path      = ".//row/row/geo_join_id"
geo_place_path   = ".//row/row/geo_place_name"
time_path        = ".//row/row/time_period"
start_date_path  = ".//row/row/start_date"
data_value_path  = ".//row/row/data_value"


Unique_ID = []
Indicator_ID = []
name = []
measure = []
measure_info = []
geo_type = []
geo_id = []
geo_place = []
time = []
start_date = []
data_value = []

#Looping through all the elements in the data that matches the path specified above:

for elements in air_quality_df.iterfind(ID_PATH):
    Unique_ID.append(elements.text)
for elements1 in air_quality_df.iterfind(indicator_id_path):
    Indicator_ID.append(elements1.text)
for elements2 in air_quality_df.iterfind(name_path):
    name.append(elements2.text)
for elements3 in air_quality_df.iterfind(measure_path):
    measure.append(elements3.text)
for elements4 in air_quality_df.iterfind(measure_info_path):
    measure_info.append(elements4.text)
for elements5 in air_quality_df.iterfind(geo_type_path):
    geo_type.append(elements5.text)
for elements6 in air_quality_df.iterfind(geo_id_path):
    geo_id.append(elements6.text)
for elements7 in air_quality_df.iterfind(geo_place_path):
    geo_place.append(elements7.text)
for elements8 in air_quality_df.iterfind(time_path):
    time.append(elements8.text)
for elements9 in air_quality_df.iterfind(start_date_path):
    start_date.append(elements9.text)
for elements10 in air_quality_df.iterfind(data_value_path):
    data_value.append(elements10.text)

airquality_data_frame = pd.DataFrame(
{'unique_id': Unique_ID,
    'indicator_id': Indicator_ID,
    'name': name,
    'measure': measure,
    'measure_info': measure_info,
    'geo_type_name': geo_type,
    'geo_join_id': geo_id,
    'geo_place_name': geo_place,
    'time_period': time,
    'start_date': start_date,
    'data_value': data_value
})

airquality_data_frame = airquality_data_frame.dropna()
airquality_data_frame = airquality_data_frame.drop_duplicates(subset=['unique_id'])

air_df_dict = airquality_data_frame.to_dict("records")
db.air_quality.insert_many(air_df_dict)
airquality_data_frame.to_csv("data_air.csv",index=False)
```

Fig. 4

Now that we have extracted, transformed the unstructured data of XML into structured format, we will be further loading it into PostgreSQL same has been depicted in Fig.5

```python
from sqlalchemy import create_engine as ce
import pandas as pd
```

```python
data = data_air_csv
#data.head(100)
```

```python
from sqlalchemy import create_engine

# Set up database connection parameters
DB_USER = "postgres"
DB_PASSWORD = "dap"
DB_HOST = "localhost"
DB_PORT = "5432"
DB_NAME = "postgres"

conn_str = f"postgresql://{DB_USER}:{DB_PASSWORD}@{DB_HOST}:{DB_PORT}/{DB_NAME}"

engine = create_engine(conn_str)

conn = engine.connect()
conn.close()
data.to_sql('air_pollution',engine,if_exists='replace', index=False)
```

Fig. 5

First, the necessary modules are imported. Then, the connection parameters are defined with the username, password, host, port, and database name. The converted data is subsequently stored in PostgreSQL (RDBMS), tables are created as per formatted dataset and further used for visualization.

III. Results and Evaluation

AQI is the measure of the air quality determined on the levels of pollutants such as Ozone, particle pollution and nitrogen dioxide. The higher the air quality index number, the more pollution there is, and the more health problems people are worried about. Figure 6 shows a depiction of the Air Quality Index (AQI) for the 15 most polluted cities in the United States.

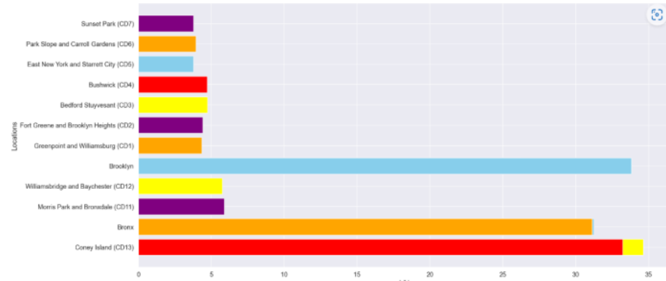

Fig.6

Ozone (O3) is a pollutant which is formed by the chemical reaction and combination of nitrogen oxides and highly volatile compounds with the presence of sunlight, Environmental Protection Agency (EPA) in US has setup an AQI standards and been monitoring by locations, below pie chart (Fig.7) shows the percentage of O3 for 5 locations.

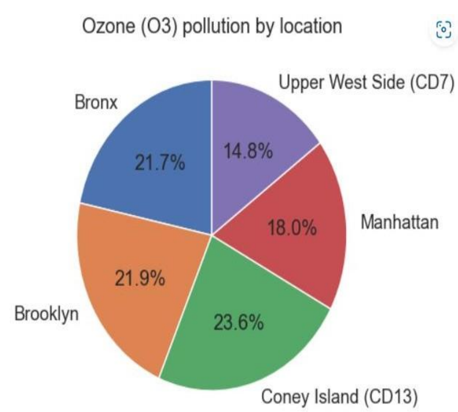

Fig.7

As there are different aspects for air pollution large population, heavy traffic and warm climate are one of the responsible factors and one of the serious problem that affects the health and wellbeing of the environment , Fig.8 displays the visualization of air pollution for top 5 location ["Central_Harlem", "East_Harlem", "Morrisania", "East_Harlem", and "Harlem"] and bar chart (Fig.8) has been generated to display the Volumetric ratio of pollution in air for top 5 locations.

Fig. 8

While Electric vehicles playing a major role on air pollution Fig 9 displays the top 10 counties in the state based on the number of electric vehicles per county, which may be used to infer how much of impact electric vehicles have on air pollution in that region. These allow us to determine the overall connection between EV usage and pollutant levels.

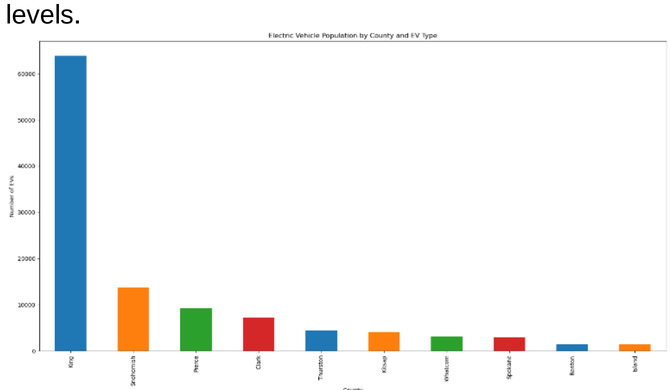

Fig. 9

Due to the thriving market for EVs and strong competition among manufacturers, consumers can now choose from a wide variety of models that promise improved range, technology, safety features, and dependability. Figure 10 displays the top 12 percentage of Electric vehicles produced by various models, and Tesla has become a market leader in this sector.

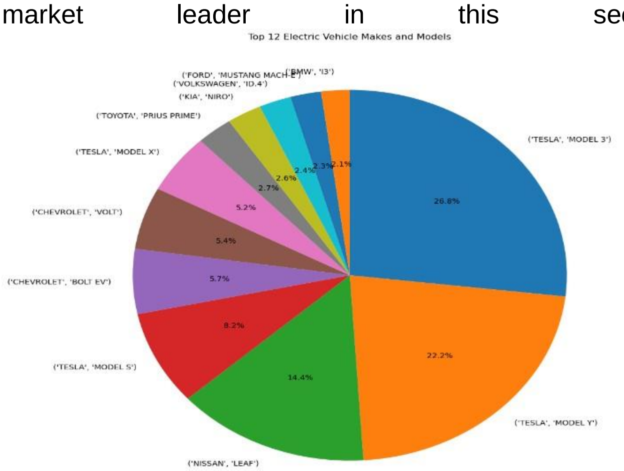

Fig. 10

Government is providing subsidies and encouraging people to buy electric vehicles for various reasons and environmental objective is one among the reasons. In Fig.11 we can see that the number of EV's sold in the years keep on raising.

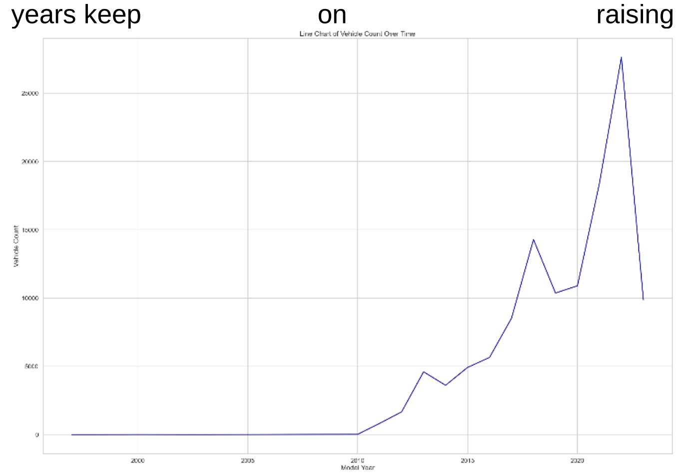

Fig. 11

Electric vehicles have a much smaller impact on air pollution compared to gasoline powered vehicles as EV's exhaust pipes produce no pollutants. Fig.12 depicts the comparative analysis of the AQI and number of EV's in relation to the year.

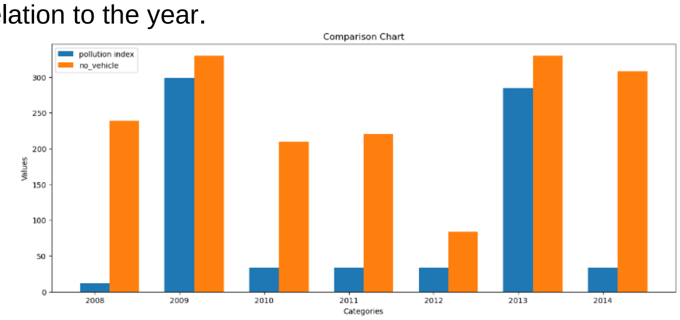

Fig. 12

IV. Conclusion & Future Works

Based on the analysis performed, from Fig.10 it was determined that Tesla was the most widely purchased model. NO2, O3, SO2 and other pollutants were found to be present in the atmosphere where Fig.7 shows the percentage of O3 present in different locations of US. In general, we concluded that there is a correlation between a rise in the number of electric cars and a decrease in the amount of pollution in the air. If we take Fig.12 as an example and look at the year 2014, for instance, we can see that the number of electrical cars was high. As a direct consequence of this, the air pollution index witnessed a considerable decline in the year 2014. In a similar way, if we consider the year 2010, we can see that not only are there a greater number of electric cars already on the road, but there was also a noticeably lower level of pollution in the air in that specific year.

In later study, we can widen the analysis to incorporate extensive range of electric vehicle models and further using machine learning techniques to predict their influence on air pollution reduction.

V. REFERENCES

[1] wa, “Electric Vehicle Population Data,” Catalog, 18-Mar-2023. [Online]. Available: https://catalog.data.gov/dataset/electric-vehicle-population-data.

[2] A. Alimujiang and P. Jiang, “Synergy and co-benefits of reducing CO2 and air pollutant emissions by promoting electric vehicles—a case of Shanghai,” *Energy for Sustainable Development*, vol. 55, pp. 181–189, 2020.

[3] A. S. Gillis and B. Botelho, “What is mongodb? features and how it works – techtarget definition,” *Data Management*, 07-Mar-2023. [Online]. Available: https://www.techtarget.com/searchdatamanagement/definition/MongoDB.

[4] I. Ullah, K. Liu, T. Yamamoto, R. E. Al Mamlook, and A. Jamal, “A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards Sustainability,” *Energy & Environment*, vol. 33, no. 8, pp. 1583–1612, 2021.

[5] J. Guo, X. Zhang, F. Gu, H. Zhang, and Y. Fan, “Does air pollution stimulate electric vehicle sales? empirical evidence from twenty major cities in China,” *Journal of Cleaner Production*, vol. 249, p. 119372, 2020.

[6] L. Wang, X. Chen, Y. Zhang, M. Li, P. Li, L. Jiang, Y. Xia, Z. Li, J. Li, L. Wang, T. Hou, W. Liu, **D.** Rosenfeld, T. Zhu, Y. Zhang, J. Chen, S. Wang, Y. Huang, J. H. Seinfeld, and S. Yu, “Switching to electric vehicles can lead to significant reductions of PM2.5 and no2 across China,” *One Earth*, vol. 4, no. 7, pp. 1037–1048, 2021.

[7] Nik, “Python requests: Get request explained • datagy,” *datagy*, 30-Dec-2022. [Online]. Available: https://datagy.io/python-requests-get-request/.

[8] S. Asadi, M. Nilashi, S. Samad, R. Abdullah, M. Mahmoud, M. H. Alkinani, and E. Yadegaridehkordi, “Factors impacting consumers’ intention toward adoption of electric vehicles in Malaysia,” *Journal of Cleaner Production*, vol. 282, p. 124474, 2021.

[9] X. Huang and J. Ge, “Electric Vehicle Development in Beijing: An analysis of consumer purchase intention,” *Journal of Cleaner Production*, vol. 216, pp. 361–372, 2019.