The Rise of Electric Vehicles and Its Impact on Charging Infrastructure in the United States

Ritik Jain
MSc Data Analytics
(School of Computing)
National College of Ireland
Dublin, Ireland
x22119469@student.ncirl.ie

Taniya Bagh
MSc Data Analytics
(School of Computing)
National College of Ireland
Dublin, Ireland
x22120831@student.ncirl.ie

Arnab Hati
MSc Data Analytics
(School of Computing)
National College of Ireland
Dublin, Ireland
x22107321@student.ncirl.ie

***Abstract*— In today's world use of Electric Vehicles is much more crucial than the conventional types as pollution levels are increasing day by day. This report "Analysis of Electric Vehicle Population, and Electric Charging Stations" examines the current state of the electric vehicle market and charging infrastructure for the last 10 years. The analysis includes a review of the current electric vehicle population, registration trends, and the availability and distribution of charging stations. The report highlights the importance of the electric vehicle market for reducing greenhouse gas emissions and improving air quality of various cities. The analysis also identifies the challenges and opportunities facing the electric vehicle market, including the need for increased investment in charging infrastructure and public awareness campaigns. A data pipeline was created to fetch the raw data from web API and loaded itinto the MongoDB database. Preprocessing is done in all the datasets', cleaned data is loaded into the PostgreSQL database. Various visualizations have been created to analyze the yearly sales of various Electric Vehicles also the total vehicles per city was plotted. It is also observed that in 2020, there is a sudden drop in sales of EVs due to covid-19 restrictions. Furthermore, it is also noticed that after 2021, sales increased rapidly, and the charging infrastructure was also built more over the years. The analysis shows that the number of charging stations is increasing, but there are still gaps in the charging infrastructure, particularly in rural and remote areas. The report identifies the need for further investment in charging infrastructure to support the growing electric vehicle market.**

***Keywords*— *Data pipeline,* Preprocessing, *Electric vehicles (EVs), Charging Infrastructure,* MongoDB, PostgreSQL**

I. Introduction

Electric vehicles have emerged as an environmentally friendly alternative to gasoline-powered vehicles. The increasing awareness of the need to reduce greenhouse gas emissions and the limited supply of fossil fuels has led to a global shift towards electric vehicles. As the popularity of electric vehicles continues to rise, it is important to understand the trends in electric vehicle population and registration, as well as the availability and distribution of charging stations.

To this end, the objective of this project is to gather meaningful insights into the current state of the electric vehicle market and charging infrastructure. We analyzed the availability of electric vehicle charging stations across various cities in the United States and the correlation between the number of charging stations and the number of electric vehicles registered in each city.

From this analysis report, several companies and organizations can find it helpful by understanding the conclusions and they can further improvise, eliminate, or prepare themselves in the future. We have collected the data, preprocessed it, and used two different databases so as to store the datasets in various phases of the project. From the individual dataset so many visualizations were plotted with the help of graphs and at the end three datasets were also merged to draw different conclusions which can show some amazing trends and insights about the electric vehicle's usage and adoption in the United States.

The datasets used for this analysis was sourced from publicly available datasets from the official US websites. It has the data of around 10 years. Dataset names are 'Electric Vehicle (EV) Registration Data', 'Electric Vehicle (EV) Population Data'. 'Fuel Station Location Data'.

Overall, in the II-section Related work is given from which we have taken help to study and understand similar work which was already done in past. In section III- methodology is explained in a clear and concise manner. This explains from the very first step which was Data fetching and importing then preprocessing, data management, data exploration, and finally, data storing. Section IV – gives us a detail understanding of data visuals which we plotted and with each and every visual you get the insights and trends of the data. This is done on both individual as well as merged data, so this gives a very easy understanding of the insights just with the help of simple readable visuals. In section V we have finally discussed the conclusions which are drawn from all the analysis. Also, in this section we have talked about the future work which can be done in addition to our project.

This highlights the need for targeted policies and initiatives to promote the adoption of electric vehicles in areas with lower charging infrastructure availability.

Furthermore, the analysis of the trends in electric vehicle registration and population over the years shows a steady increase, indicating a growing interest in electric vehicles. This highlights the need for continued investment in charging infrastructure to meet the increasing demand.

II. Related Work

"Electric Vehicle Adoption: Insights from the US and Norway" by Deepak Krishnamurthy and Uday Rajan. This paper analyzes the factors that influence the adoption of electric vehicles in the US and Norway, including government policies, charging infrastructure, and consumer attitudes towards electric vehicles.[1]
"Electric Vehicles in the United States: A New Model with Forecasts to 2030" by Colin McKerracher and Pedro Haas. This paper provides a comprehensive analysis of the electric vehicle market in the US, including the current state of the market, trends in consumer behavior, and forecasts for the future. [2]
"A Comprehensive Review on Electric Vehicle Charging Infrastructures: State-of-the-Art, Challenges, and Research Directions" by Ahmed Abdelhady, Mohamed Hussein, and Tarek Sobh. This paper reviews the current state of electric vehicle charging infrastructure, including the different types of charging stations, the challenges associated with building and maintaining charging infrastructure, and future research directions. [3]
"Smart Charging of Electric Vehicles: A Review of the State-of-the-Art and Future Research Directions" by Luis Martinez-Ramos and Luis Martinez-Garcia. This paper reviews the current state of smart charging for electric vehicles, including the different types of smart charging systems, their benefits and drawbacks, and future research directions. [4]
"An overview of electric vehicle charging infrastructure deployment in smart cities" by Y. Liu, et al. (2019) - This paper presents an overview of the current state of electric vehicle charging infrastructure deployment in smart cities. It discusses various challenges and opportunities for electric vehicle adoption, including the need for interoperability, accessibility, and scalability of charging infrastructure.[5]
"Electric vehicle charging infrastructure planning: A review of models and methods" by G. Gökaydin, et al. (2017) - This paper provides an overview of the methods and models used for electric vehicle charging infrastructure planning. It discusses various factors that need to be considered for optimal planning, such as location, power capacity, and cost-effectiveness.[6]
"Electric Vehicle Charging Infrastructure Planning and Implementation: A Review of Best Practices" by R. Tiwari, et al. (2020) - This paper provides a comprehensive review of best practices for electric vehicle charging infrastructure planning and implementation. It discusses various factors that need to be considered, such as site selection, infrastructure design, and regulatory policies. The pap er also presents case studies of successful implementation of charging infrastructure in different regions.[7]

III. METHODOLOGY

A. DATA SOURCE AND DESCRIPTION:

The datasets used for this analysis was sourced from various publicly available datasets which is explained below:

i) Dataset 1- Electric Vehicle (EV) Registration:

This dataset was obtained from the California Department of Motor Vehicles (DMV) website. The DMV collects data on all registered vehicles in the state, including EVs. The dataset includes information such as the make, model, year, and zip code of registration for each EV. The data covers the period from 2011.

Dataset was fetched through API in JSON format and saved into jupyter notebook through Pandas library.

Dataset link: [Electric Vehicle Title and Registration Activity - Catalog (data.gov)]

ii) Dataset 2- Electric Vehicle (EV) Population Data:

This dataset was obtained from the California Department of Motor Vehicles (DMV) website. The DMV collects data on all registered vehicles in the state, including EVs. The dataset includes information such as the make, model, year, and zip code of registration for each EV. The data covers the period from 2011 to 2021.

Dataset was fetched through API in JSON format and saved into jupyter notebook through Pandas library.

Dataset link: [Electric Vehicle Population Data - Catalog]

iii) Dataset 2- Electric Vehicle (EV) Population Data:

This dataset was obtained from the United States Department of Energy (DOE) Alternative Fuels Data Center (AFDC) website. The AFDC maintains a database of fueling stations for alternative fuel vehicles, including EV charging stations. The dataset includes information such as the name, address, latitude, and longitude of each charging station, as well as the types of charging available and the number of charging ports. The data covers the period from 2011 to 2021.

Dataset link: [Alternative Fueling Stations | Alternative Fueling Stations | Geospatial at the Bureau of Transportation Statistics (arcgis.com)]

This dataset was downloaded in GEOJSON format and imported into jupyter notebook through Pandas library.

B. DATA MANAGEMENT:

All the three dataset files were semi structured data and is fetched in JSON and GEOJSON format from the web API, after fetching it is converted into data frame, and the raw data is then stored in MongoDB. Further, the raw data is again retrieved from the MongoDB database for pre-processing and cleaning. Once the data is cleaned and preprocessed this cleaned data is finally stored in PostgreSQL database. Lastly, cleaned data is again retrieved from the PostgreSQL database to finally analyze the data through various visualizations using seaborn matplotlib libraries of python. To better understand the overall flow of the process, Flow charts is plotted. This flow chart displays all the steps from starting till the end in a very systematic manner.

C. SYSTEM FLOW DIAGRAM:

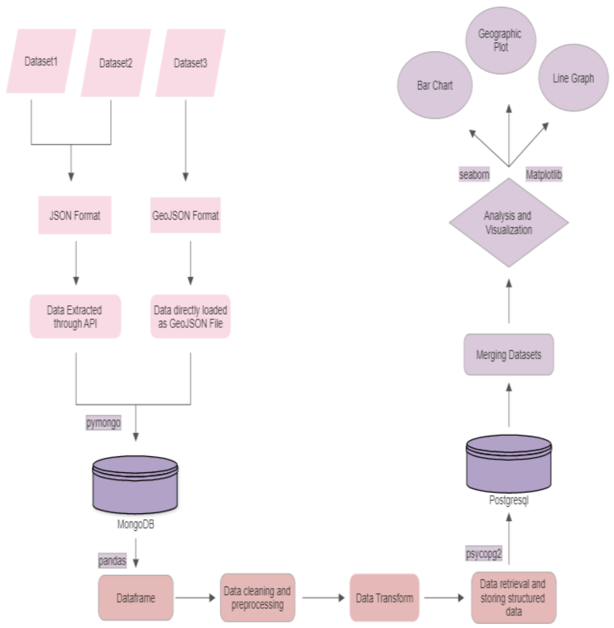

Figure.1. Workflow diagram of all Datasets.

Figure.1 shows the clear and concise outline of the project. This illustrates the work done on three datasets. Dataset 1 and Dataset 2 are taken in JSON Format whereas Dataset 3 is taken in GeoJSON format. The third dataset is taken in GeoJSON format since it contains fields like longitude and latitude fields which were later used for carrying out the analysis that was done individually and later on were used in drawing the conclusion. The first two datasets were fetched through API whereas the third dataset was directly loaded in GeoJSON format. These data were loaded into MongoDB then the data were retrieved and dataframes were created. After this data cleaning and preprocessing were done on these datasets using different approaches such as feature selection, removal of outliers, etc in order to clean data and fetch the data which was required for the analysis. Furthermore, these data were transformed into structured data and loaded into PostgreSQL for further analysis. Data retrieval for all three datasets was done from PostgreSQL and then they were merged for analysis of the final research question. Finally, analysis was done, and for showcasing the results various visualizations were used.

IV. DATA PREPROCESSING AND DATA STORAGE:

A. DATA PRE-PROCESSING:

Data preprocessing is a crucial step in data analysis that ensures the accuracy, completeness, and reliability of the data. In this process, the raw data is formatted and cleaned to prepare it for analysis.

Data cleaning is the process of locating and dealing with missing, incorrect, or inconsistent data in order to enhance its quality and dependability. The process of pre-processing entails formatting the raw data.

i) Feature Selection:

A data analysis approach called feature selection involves choosing a subset of the most important features or variables from a dataset. Reduce the dimensionality of the data while maintaining as much pertinent information as you can is the aim of feature selection. Out of every feature, we chose the ones that were most important to our analysis.

ii) Feature engineering:

The process of developing new variables or features from ones already present in a dataset is known as feature engineering. Create new features from the current columns that can be beneficial for the analysis. For instance, a "year" or "month" column may be retrieved from the "open_date" column to allow for time-based analysis.

iii) Handling missing data:

We looked to see whether there were any missing values in the dataset. We replace any incorrect values with zero if any exist. Numerous records included zero, thus to distinguish it and alter it to null, the null values are ultimately removed from the dataset.

iv) Handling Duplicate data:

Duplicate record detection and management is a crucial stage in data cleaning since it may raise the data's correctness and quality. When there are several records in a dataset with identical values in most or all of the columns, the records are considered duplicates.

v) Remove the Outliers:

Data points known as outliers diverge greatly from the rest of the dataset's data points. On the other hand, noisy data includes flaws or inconsistencies that might skew the data's general distribution or pattern. Because outliers and noisy data can significantly affect the outcomes of data analysis, identifying and addressing them is a crucial stage in data cleaning and pre-processing. To exclude outliers from the dataset, we employed z-score.

vi) Datatype conversion:

The process of altering the data type of one or more variables in a dataset is known as data type conversion. When a variable's data type is wrong or it has to be changed to a better type for more research or modelling, this procedure is utilized. Changing a text variable into a date variable, for instance, or changing a category variable into a numerical variable.

In conclusion, data preprocessing is important. By employing various techniques such as feature selection, feature engineering, and data type conversion, it is possible to prepare the data for analysis and derive meaningful insights.

B. DATA STORAGE:

To store the raw data, a Docker instance was used to set up the MongoDB application, which was a document-oriented NoSQL database designed for high performance, flexibility, and scalability. For storage of data, Mongo database and PostgreSQL were used at different instances. Mongo DB was used because of its features such as it is a document-oriented NoSQL database that is built for providing high performance, flexibility, and scalability also it provides a platform that is easy to use and provides easy handling and storage of large amounts of unstructured or semi-structured data. It allows developers to select the ideal data model for their application from a variety of supported data models, including document, key-value, graph, and column-family. Here we have used a docker image instead of setting up the MongoDB application on the individual local machines and the database and Python were connected using pymongo. To store the data collections are created and then these unstructured data are stored in it. In this report we have directly fetched Dataset 1 and Dataset 2 through API and Dataset 3 was directly loaded through GEOJSON formatted file. A schema named EV_database was created in MongoDB and three collections were made to store the unstructured/semi-structured data. Collections namely EV_Registration, EV_population and Fuel_Station111 were created to store Dataset 1, Dataset 2, and Dataset 3 respectively. After this, the data were retrieved from MongoDB and then loaded into dataframes namely EV_Registration_DF, EV_Population_DF and Fuel_Station_location_DF were created with the help of the function “create_dataframe_MongoDB” which was created while setting up the connection with MongoDB. The further analysis was done on these data and then structured data were stored in PostgreSQL which was also created through the docker instance. The connection between the database and python was created by using the psycopg2 library.

To store the structured data, a Docker instance was also used to set up the PostgreSQL database. PostgreSQL is a powerful, open-source relational database management system that provides robustness and scalability. The connection between the database and Python was created using the psycopg2 library. After establishing the connection, a database named "evpopulation" was created, and four tables were made, namely table 1, table 2, table 3, and table 4. Table 4 was used for the merged data. The data was then loaded into these tables, and data was retrieved again into dataframes for final analysis and visualization purposes.

This approach allowed for efficient and scalable storage and retrieval of data, enabling the team to easily manipulate and analyze large volumes of data while ensuring consistency and reliability. The use of Docker also made it easier to manage the deployment of the database instances and ensured consistency across different environments.

V. DATA VISUALIZATION

Data Visualization are graphs and plots which displays the highlights, trends, insight information through which to the reader can very easily understand and will be able to interpret the data. In this project, we plotted several visualizations, and these were created using Seaborn and matplotlib libraries.

These libraries provide an extensive selection of interesting and beautiful statistical plots to display the data. In order to assess the data and derive important conclusions, several plots, including line charts, bar plots, and stacked bar plots, were employed. Few of them are displayed below:

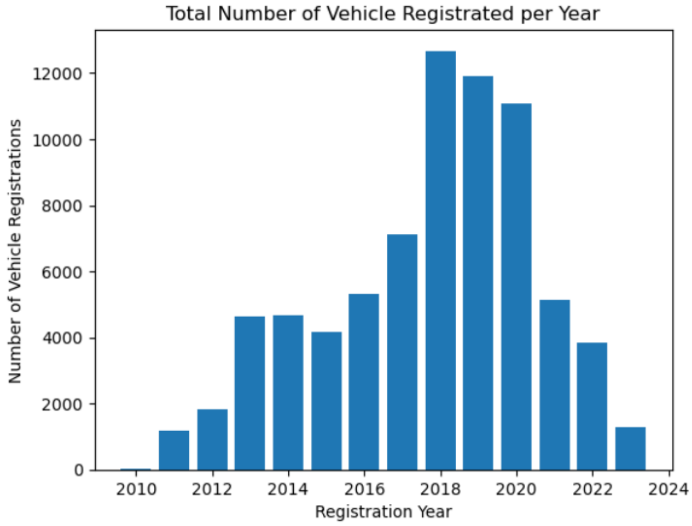

Figure. 2

Bar Diagram in Figure. 2 shows the total number of vehicles registered from 2010 till 2023. The graph shows that till 2010, there are not many registrations of Electric Vehicles but in the last decade starting from the 2011 we can see a clear rise in the registration count of the vehicle.

After 2020, we all know there was a covid-19 outbreak worldwide and lockdown was imposed in many countries therefore there was a sudden drop in the registrations of new vehicles. After 2021 and 2022, there can be a rise in the registrations since 2023 in not yet completed therefore we don’t have the full data. In future maybe we can conclude.

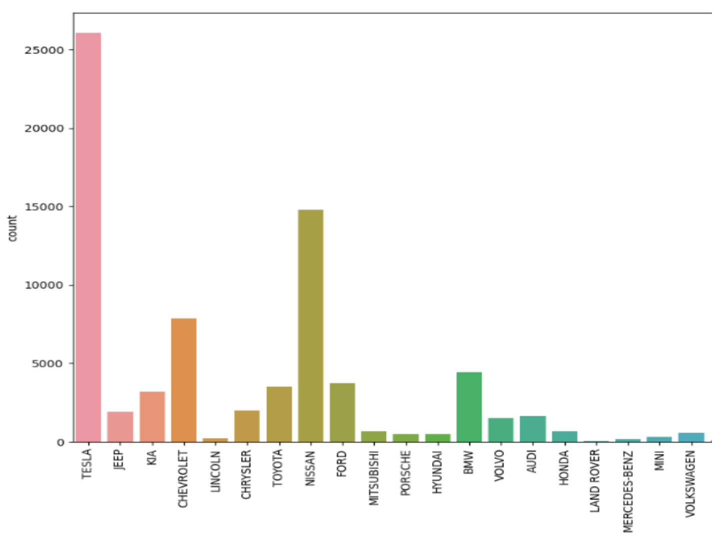

Figure. 3

Figure. 3 shows the vehicle brand type total registrations.

We can clearly see that Tesla is the most registered brand out of all brands having more than 25000 registrations. Second is the Nissan brand having near about 15000 registrations. Third is Chevrolet with nearly 9000 total registrations. Rest all other brands have few counts.

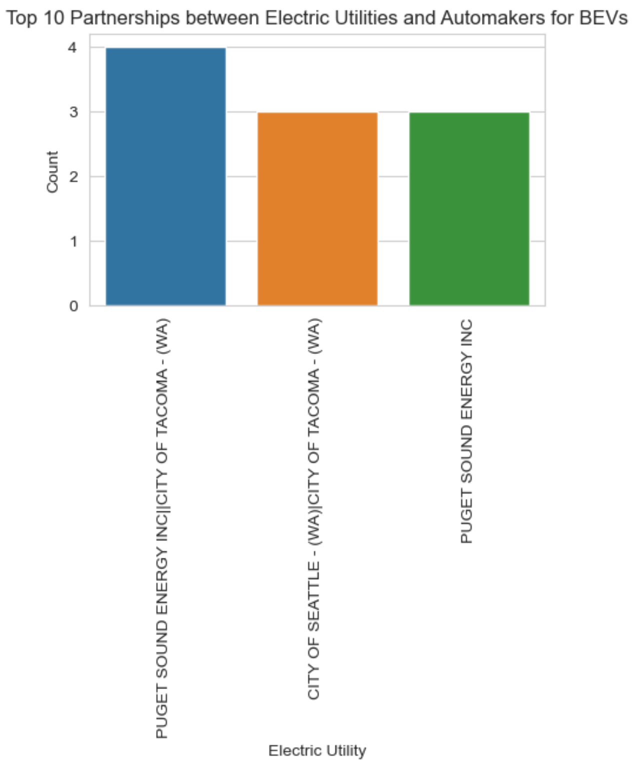

Figure. 6

Figure. 6 shows the Top Partnerships between Electric Utility Providers and Tie-up with the Automakers. We can clearly see that Puget Sound Energy Inc provides the most number of charging infrastructure.

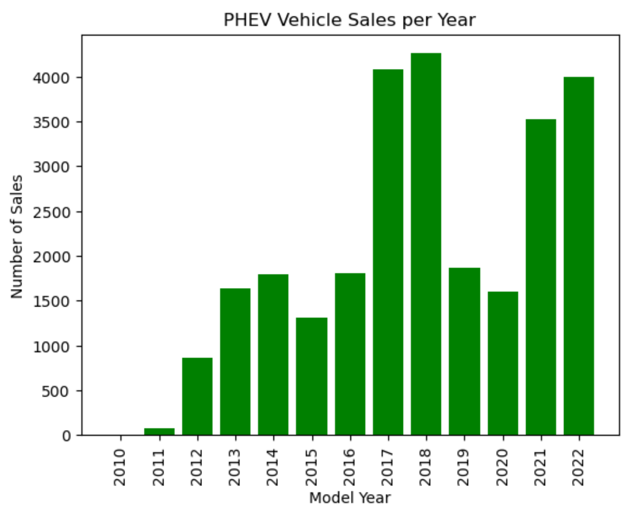

Figure. 4

Bar graph in Figure. 4 shows the overall sales data of a decade in the Washington state of the United States. This graph depicts that there were many sales in 2017 and 2018. Then in 2020 again because of Covid-19 outbreak the sales went down and from 2021, 2022 sales again got increased as covid-19 restrictions got reduced.

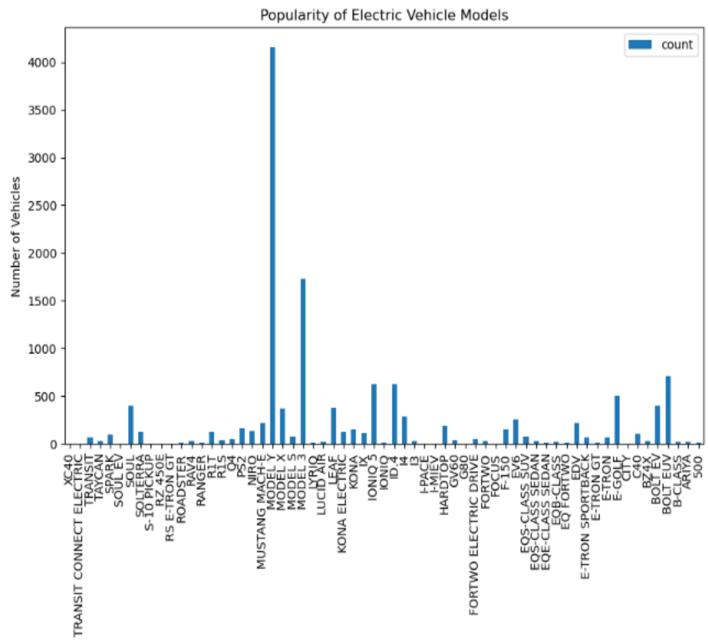

Figure. 5

The bar graph shown in Figure. 5 shows the popularity of Electric Vehicle model out of all the models available on the streets of the United States.

We can see that Tesla's model-Y are the most sold model of Electric Vehicles. Also, the model-3 of Tesla is the second most popular in the streets. The rest of the other Spikes which are seen are the model of Nissan brand. A few Chevrolet Brands like Bolt EUV and Bolt EV are also somewhat popular out of all vehicles.

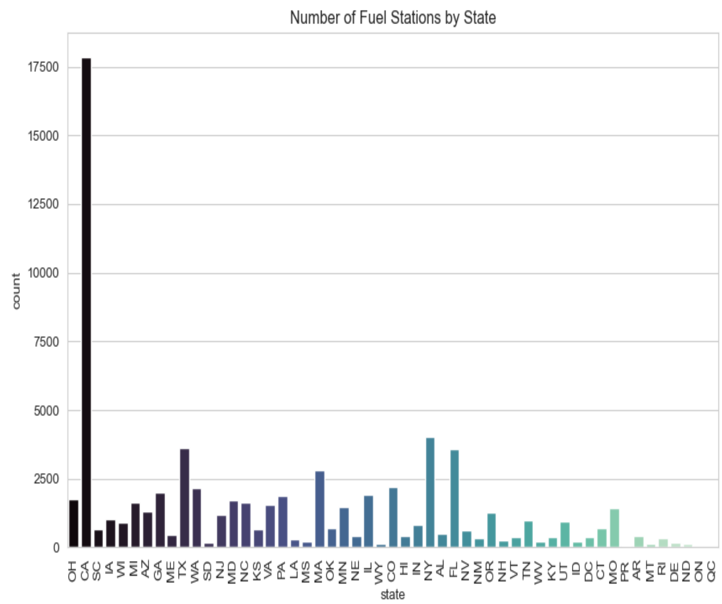

Figure. 5

The above graph shows the total count of the electric charging stations across all the various States of the United States of America. Seeing this graph it is clear that that California (CA) is having the most number of charging infrastructure more than 17500. New York is having the second most number of charging stations.

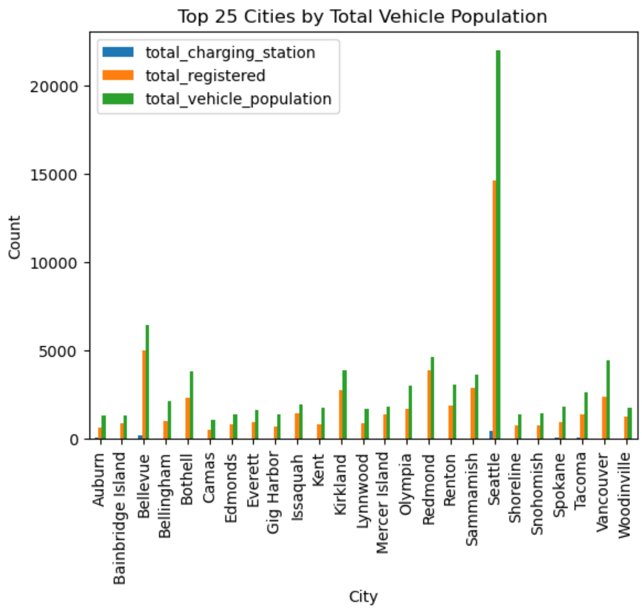

Figure. 6

Figure. 6 is a plot of top 25 cities which have the highest number of electric vehicles. Also out of these top 25 which have the highest number of vehicles. So, according to graph Seattle is have the most number of Electric Vehicles with a count of more than 20000. On a second spot Bellevue is the second highest city to have more number of electric vehicles. EV population in Bellevue is around 7000.

The other cities can be seen having less number of Electric Vehicles but this count will increase day by day as more number of vehicles are registering after sales.

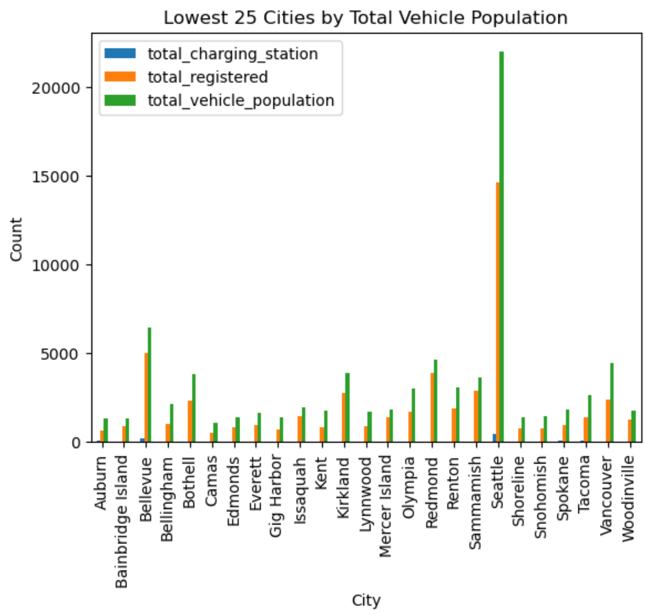

Figure. 7

Figure. 7 shows the 25 cities having the lowest Electric vehicle population.

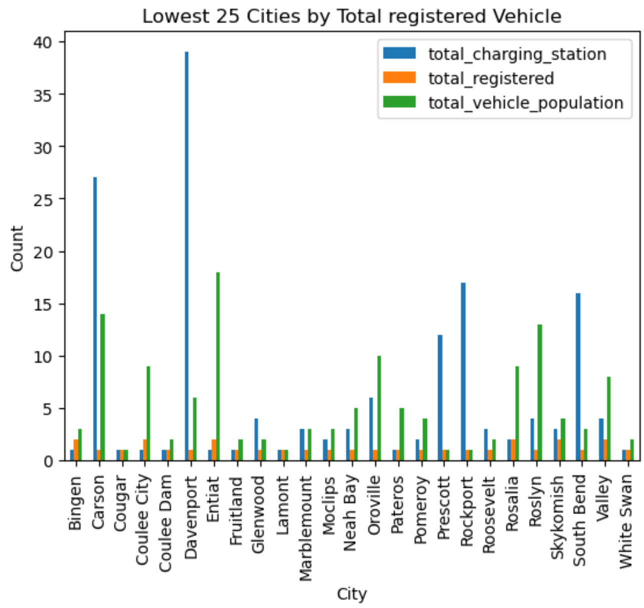

Figure. 8

The above graph shows the insights of the cities having lowest number of Electric Vehicle registered. Rockport is the city where the lowest number of Electric vehicles were registered. Also, there are many other cities which do not have a good sales record of EV's. This concludes that there are very less number of charging stations and the people of these cities do not prefer to buy EV's. May be From these insights EV manufacturers can do some research and promotion to motivate peoples to buy the EV's in these cities.

VI. Conclusion

In conclusion, this project provides insights into the current state of the electric vehicle market and charging infrastructure in the United States. The analysis of the datasets shows a positive trend in electric vehicle adoption, with an increase in the number of electric vehicle registrations and population over the years.

- It was observed that before 2010 there were very less number of electric vehicles registered. This is because, in earlier years there were very less number of charging stations which leads to think customers about the long run of the vehicle. Between 2011 and 2018 there is a sudden increase in electric vehicle sales. The number of EV users increased more and so as the EV manufacturers started manufacturing more new variants of the EV's.
- It is noted that after 2019 there was a covid-19 outbreak in all over the world. Due to this pandemic there is a sudden drop in sales and registration of EV's. But soon after mid 2021 the sales of EV's were again increased and it seems that it will linearly increase over the years.

about charging stations and its infrastructure California state have the most number of charging stations.

However, there is still a need for targeted policies and initiatives to promote the adoption of electric vehicles in areas with lower charging infrastructure availability. Continued investment in charging infrastructure is also necessary to meet the increasing demand for electric vehicles.

This project can be used by companies and organizations to understand the trends in the electric vehicle market and charging infrastructure, and to make informed decisions regarding future investments and initiatives in the field. Furthermore, the methodology used in this project can serve as a guide for future research on electric vehicles and charging infrastructure.

VII. References

[1] Krishnamurthy, D., & Rajan, U. (2019). Electric vehicle adoption: Insights from the US and Norway. Energy Policy, 127, 356-362.

[2] McKerracher, C., & Haas, P. (2015). Electric vehicles in the United States: A new model with forecasts to 2030. Energy Policy, 86, 273-285.

[3] Abdelhady, A., Hussein, M., & Sobh, T. (2018). A comprehensive review on electric vehicle charging infrastructures: State-of-the-art, challenges, and research directions. Renewable and Sustainable Energy Reviews, 82, 3000-3013.

[4] Martinez-Ramos, L., & Martinez-Garcia, L. (2018). Smart charging of electric vehicles: A review of the state-of-the-art and future research directions. Renewable and Sustainable Energy Reviews, 82, 536-544.

[5] Gökaydin, G., Akkaya, G., & Uysal, M. (2017). Electric vehicle charging infrastructure planning: A review of models and methods. Renewable and Sustainable Energy Reviews, 67, 515-535. doi: 10.1016/j.rser.2016.09.039

[6] Yi He, Ziqi Song, Zhaocai Liu, Fast-charging station deployment for battery electric bus systems considering electricity demand charges, Sustainable Cities and Society, Volume 48,2019,101530,ISSN 2210-6707, https://doi.org/10.101/j.scs.2019.101530.

[7] R. Tiwari, M. Alam, M. K. Jha, and P. K. Jha, "Electric Vehicle Charging Infrastructure Planning and Implementation: A Review of Best Practices," Journal of Advanced Transportation, vol. 2020, p. 8839897, Dec. 2020, doi: 10.1155/2020/8839897.