School of Physics,
Engineering and
Computer Science

MSc Data Science Project

7PAM2002-0901-2024

Department of Physics, Astronomy and Mathematics

Data Science FINAL PROJECT REPORT

Project Title:

EV Population EDA

Student Name and SRN:

Mohsin Zameer
22084592

Supervisor: Alyssa Drake

Date Submitted: 6-january-2025

Word Count: 10023

GitHub address: https://github.com/Mohsin7171/Final-Project.git

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science in Data Science at the University of Hertfordshire.

I have read the guidance to students on academic integrity, misconduct, and plagiarism information at [Assessment Offences and Academic Misconduct] and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project module or course.

I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7, and UPR AS/C/5, section 3.6).

I have not used chatGPT, or any other generative AI tool, to write the report or code (other than where declared or referenced).

I did not use human participants or undertake a survey in my MSc Project.
I hereby permit the report to be made available on module websites provided the source is acknowledged.

Student SRN number: 044220845923

Student Name printed: M. Zameer

Student signature:

UNIVERSITY OF HERTFORDSHIRE

SCHOOL OF PHYSICS, ENGINEERING AND COMPUTER SCIENCE

Electric Vehicle Population EDA

Table Of Contents

Abstract

They coined the term EVs to mean that the world has witnessed a shift in the automobile and energy industries due to innovation, environmental conscience, and encouragement by governments. In this research, Exploratory Data Analysis (EDA) is used to analyze the patterns, distribution, and factors that underlie the global trend and distribution of EVs. The research focuses

on the dynamics of the EV population and market segmentation, based on the make, model, year, electric range, and state of registration using an open dataset.

Firstly, the method of data preprocessing is used, including pre-processing of missing values, outliers clearing, and feature engineering to assess the reliability of the data set. Additional dimensions like the age of a car and the price segment are obtained to increase the level of analysis. Analytical quantitative and qualitative methods of descriptive and inferential statistics as well as superior graphical means are used to identify key tendencies in the rate of uptake of EVs, growth trends, and regional preferences.

Major discoveries presented include the fact that EV registration growth accelerated after 2015 and is most prominent in progressive states like California due to a well-developed charging network. Vehicle range and price become key factors for a customer's choice, as revealed by a positive correlation between electric range and base price MSRP. KMeans clustering identifies three distinct EV market segments: cheap, intermediate, and expensive, thus giving him or her information concerning consumers and the market at large. State-wise, as well as legislative district studies, also provide evidence of how localized incentives coupled with infrastructure played a crucial part.

Thus, the findings of this research suggest that EDA is a useful approach to utilising EV adoption trends for policymaking, manufacturing, and other related industries. The results point out that it is necessary to use the detailed stimuli, develop infrastructure, and set reasonable prices to promote EV use. The research could advance by applying the machine learning approach to predict adoption and create effective strategies for the EV market.

CHAPTER 1

Introduction

The automotive manufacturing sector is shifting towards sustainable mobility around the globe and EVs hold the key to reducing the negative environmental impact of ICE vehicles. Urban pollutants, global warming, and rising energy demands have forced governments, organizations, and consumers to accept EVs at a high rate. This shift sits well with the global call to cut down on carbon emissions, enhance energy efficiency and adoption of clean energy sources. Moving away from fuel based energy sources, EVs provide a cleaner means of transportation than conventional cars adding great value to global sustainability initiatives.

The authors Bas, Cirillo, and Cherchi in their paper published in 2021 assert that the EV market has rapidly expanded over the last several years with more than 10 million electric vehicles currently in use globally. The growth is attributed to increase in technology, government policies and change in consumer attitude. Nevertheless, it is pertinent to note that with the increase in the use of EV technology there are still barriers such as lack of proper charging infrastructure, high costs and disparities across the regions. Analyzing these dynamics means using data to study adoption patterns and factors that impact them and to discover applicable knowledge. Thus, this research uses EDA to identify patterns in population and distribution of EVs, and factors that may affect adoption across different regions.

1.1 Background and Importance of Electric Vehicles

Thus, electric vehicles have been developed as a green mobility solution because their broad deployment could help eliminate greenhouse gas emissions. Transportation sector contributes to about 20% of total CO_2 emissions; hence, it is a prime candidate for conversion to low or zeroemission sources. In contrast, ICE vehicles emit zero tailpipe emissions, and are hence beneficial at cutting air pollution arising from transportation in urban environments which contribute significantly to particulate matter and nitrogen oxides. Chhetri et al., (2024) have also supported the fact that EVs have a role to play in energy security by reducing the dependence on imported oil and harnessing the renewables.

Technological Advancements Driving EV Adoption

Sophistication of battery systems, charging networks and automotive engineering have removed almost all the challenges that hinder the adoption of electric cars. Lithium-ion batteries that power EVs have been developed considerably in terms of energy density, and cost-effectiveness. According to Shrestha (2024), the cost of lithium-ion batteries reduced by about 89% from the year 2010 to 2023, making electric vehicles cheaper for consumers. This has enabled the manufacturers to develop new generations of electric, such as the Tesla Model S and Nissan Leaf that have ranges that are well over 300 miles. These advancements have lowered down the range anxiety, a major factor of concern for potential EV buyers and hence have improved the EV reliability.

The availability of charging stations has greatly encouraged the use of electric vehicles as more charging stations have been put up. Well-developed public charging networks, including Tesla Superchargers and other nationwide solutions, include convenient and fast charging solutions, which adjust for the ownership practicalities. Chhetri et al. (2024) have noted that the states with the proper charging facilities of electricity have a higher rate of EV usage than those lacking such facilities like California.

Policy Support and Economic Incentives

Policies such as subsidies and grants given by the government, tax incentives, and charging infrastructure for the adoption of EVs are also essential. In USA through federal tax credits worth $7,500 for EV acquisitions coupled with other credits from the states have made EVs more economic. For instance, the state of California has set extremely high ZEV targets and has backed them up with extra rebates making it perhaps the single biggest market for EV adoption (data.wa.gov, 2024).

In Norway, for instance, the sales of EVs are above 70% of new cars and an indication of the impact of a well-coordinated policy on the same. Some of these policies include: offering big bonuses, free from toll and registration charges, and putting up of charging terminals. Shrestha (2024) has noted that it is principally important for governments to play an active role in overcoming the factors which are the impediments to the broader uptake of EVs, especially in places where demand remains relatively low.

Environmental and Social Impacts

The advantages of EV for its use do not stop at the environmental issues only. EVs can help to reduce air pollution in large cities, so the frequency of respiratory illnesses and other diseases that have an environmental cause is lowered. According to data available in data.wa.gov for the year 2024, smog and particulate matter has reduced in cities that actively use EVs including San Francisco in the last decade. Further, the changes in the transportation sector are a plus for the electrification of the energy sector because electric cars are connected with smart grids for storing renewable energy.

From the social activity point of view, the adoption of EVs supports economic development because it promotes employment in production, charging infrastructure, and renewable power generation. The shift towards EVs is also consistent with the customer desire for new and environment friendly technologies, which not only strengthens their place in the transport sector in the future.

1.2 Role of Data in Understanding EV Adoption

The increasing deployment of EVs produces abundant information, such as make, model, range, price, and distribution of EVs. Evaluating this data is especially important to identify the trends which affect the EV market and those, contributing to its development. However, raw data is not enough to deliver a real understanding of the situation; it should be processed using appropriate methods. Exploratory Data Analysis or EDA provides a detailed framework to describe, present and analyze data, to identify latent structures and relationships in data.

Analysis based on EDA focuses on distributions, correlation and trends in order to discover insights. In the context of EV research, EDA can provide answers to questions such as:

- Which regions exhibit the highest and lowest EV adoption rates?
- How do vehicle characteristics, such as range and price, influence consumer preferences?
- What trends emerge in year-over-year EV registrations, and what factors drive these changes?

For instance, correlation analysis can help establish the relationship between the electric range and the cost of the vehicle and heat maps and scatter plots can help to establish regional variations of adoption rates. As Bas, Cirillo, and Cherchi (2021) show, quantitative methods such as EDA provide key insights into the development of the EV market and formulation of suitable policies and measures.

1.3 Objectives of the Study

The objectives of this research are to analyze the current distribution and future growth of the EV population and identify factors that have an impact on them through EDA techniques. The specific objectives are as follows:

1. **To analyze EV adoption trends over time**: This type of research aims to gain insights on patterns of adoption by analyzing year to year registration and area where the rate of registration is on the rise or is comparatively slower.

2. **To investigate regional disparities in EV adoption**: The study therefore focuses on geographical distribution of EV registrations, and how state characteristics such as policies, incentives, and infrastructure affect the outcomes.

3. **To evaluate factors influencing consumer preferences:** Specific features that are the range of vehicles, prices, and potential subsidies are among the aspects to be examined in relation to consumers.

4. **To segment the EV market:** In this research, the EVs are grouped into subgroups through clustering analysis to understand the market segmentation based on some characteristics like range and price.

5. **To provide actionable insights for stakeholders:** The presented results are intended to help policymakers, manufacturers, and investors to create efficient strategies for the promotion of EVs and potential solution for obstacles.

1.4 Challenges in EV Adoption

Nonetheless, there are a number of challenges which delays regular use of EVs, and here are the following ones: For the optimal usage of electricity to support the goal of sustainability, these challenges must be surmounted for EVs.

Infrastructure Gaps

In a way, the first public charging station remains an issue for EV uptake. While many of the metropolitan centers have instituted many charging stations, rural regions remain neglected. According to the statistics from data.wa.gov (2024), only those states that have built sufficient supporting infrastructures for EVs such as California and Washington have significantly higher rates of using EVs than those states which contain very limited number of charging stations.

Cost and Affordability

To some extent, it is important to note that even though the cost of lithium-ion battery has decreased, initially, the costs associated with EVs are higher than compared to ICE vehicles. This price differential deters the conscious user, particularly the one who has very limited options on the number of offers available for the journey. Shrestha (2024) has also pointed that subsidies and tax credits are perusable for the pursuit of the goal for the complete removals of the finical barriers.

Range Anxiety and Consumer Awareness

Some issues with the new and used EVs still remain confusing, especially for the people uninformed of the current developments in car batteries. Informing consumers about new opportunities of EVs and lowering costs of their operation and minimal negative impact to environment is the solution to the issues listed above.

The issuance of electric cars is one of the major progress in the fight against climate change and the creation of sustainable development. But a widespread use of EVs will not become possible without collecting large amounts of data about trends, patterns, and factors that impact consumers' decision. To this end, using machine learning to make sense of a large data set, this research seeks to identify key recommendations that would help policymakers, manufacturers, and other stakeholders work towards the adoption of EVs. These findings would be of significant importance in emerging responses to existing issues and advancing the shift to a sustainable transportation system.

CHAPTER 2

Literature Review

The use of electric vehicles has attracted a lot of interest among researchers, policy-makers and other players of the economic sector because of the important role that the use of such vehicles plays in minimising the negative impression that transport has on the environment. This chapter compares and analyses prior literature on the subject of EV adoption based on technological innovation, competition, regulatory policies, and analytics. Emerging findings from empirical studies are discussed in relation to their roles in informing the determinants of EV and the areas for future investigation.

2.1 Technological Advancements in Electric Vehicles

EV technology has played a big role in avoiding barriers that were well-evident in previous generations such as range constraint and charging infrastructure costs. Ref: Shrestha (2024) gives an insight of machine learning in EVsThough with an extensive range of applications and approximate timeline The research notes that the key functions of machine learning in EVs include efficiency of the battery management system, energy efficiency, and performance improvement. Of all the components of electric vehicles, battery technology has experienced the most innovation especially in lithium-ion batteries in terms of cost per kilowatt-hour and energy density. An approximate of 89% reduction of Li-ion batteries cost is predicted to occur between 2010 to 2023, mainly making evs affordable.

In addition, charging technology concerns have been eased, removing one of the major impeders to broader adoption. EVs require much time to charge on the road due to the slow charging networks which limit their effectiveness for long-distance. Chhetri et al. (2024) propose that charging infrastructure availability influence EV adoption where cordially regions with a large number of charging stations obtain higher penetration rates. These observations East/West have underlined the need for more investment in charging stations necessary for a burgeoning EV market.

Even with these developments there are always drawbacks experienced in battery recycling and the use of sustainable material. Shrestha (2024) focused on the issue of lithium and cobalt mining and suggested that the world needs to invent new material and circular economy.

2.2 Market Dynamics and Consumer Preferences

Market knowledge is the key to advancing the popularity of EVs. According to Bas, Cirillo, and Cherchi (2021) using machine learning to categorize potential EV purchasers and found out that affordability, range and charging were considered critical aspects that would determine that chance of a purchase. Their study uses clustering techniques to segment the EV market into three categories: new buyers, middle market clients, and luxury-segment clients. This segmentation makes sense in the light of current market development observed for EVs, Tesla being flagship in premium segment while Nissan or Chevrolet adopting a more economical approach.

Ref: Data.wa.gov (2024) lays out concrete qualitative evidence of the regional differences in the use of EVs across the United States and correlates the disparity to existing state-based policies, infrastructure standards, as well socio-economic factors. Main reasons cover incentives, emissions standards, and charging infrastructure: California tops the list as the most adopting state. On the other hand, the relatively low-policy support and infrastructure state like Wyoming and West Virginia, trail far behind. These findings point to the need to local policies in facilitating adoption of driving.

Chhetri et al. (2024) dissect the interaction between price and range with popularity, and highlights a positive association between range and price. This means that the consumer is ready to pay more for an extended-range EVs as Convenience and reliability are core values to every consumer. However, price sensitivity is possible only for many potential consumers; that's why there is a need for cheap electric cars and subsidies.

2.3 Policy Frameworks and Incentives

This approach has achieved considerable success because of government support in terms of policies and incentives. Shrestha (2024) identifies subsidies, tax credits and grants as the important tools for lowering the purchasing price of EVs and thus making them similar to and competitive with ICE vehicles. For instance, federal tax credit for the electric vehicle acquisition stands at $7,500 for the new vehicle, and states like California also provide rebates. These are financial handshakes have been shown to help increase adoption rates especially within the price tender consumer segment.

Additional policies like the ZEV mandates also encourage manufacturers to take a step towards the use of electric vehicles. Data.wa.gov (2024) explain that California pioneered the Advanced Clean Cars Program which regulates how many ZEVs auto makers must sell each year. This policy has cretaed a level of competitiveness and coming up with new models of EVs to suit the market needs of consumers.

Infrastructure investment is another important dimension of policies in writing policy frameworks. Chhetri et al. (2024) stressed on the PPP for improving the community of charging stations especially in the rural areas. The Bipartisan Infrastructure Law signed in the U.S in 2021 provide $7.5 billion which will go into establishing a network of EV chargers a bottleneck as highlighted above. However, there is also understanding that the policy efficiency is different depending on the region and that is why the specific plans and actions must also be different responding to the local problems.

The authors also note, however, that incentives may be detrimental when they point out that ‘excessive’ incentives can be problematic, as they hinder learners from becoming financially independent (Bas, Cirillo, and Cherchi, 2021). He criticized that long-term adoption needed sound market factors such as technology development, price competition, and demand promotion etc. Policymakers face the challenge of encouraging the adoption of EVs and at the same time, promoting the development of the independent market for EVs.

2.4 Data-Driven Approaches to Understanding EV Adoption

The development in multiple sources of population level EV data means that researchers have been able to apply sophisticated analytical methods to investigate adoption behaviours. In his paper Shrestha (2024) describes the potential of clustering and regression models for the evaluation of consumer segments, rating of the potential adoption rates, and the assessment of policies consequences. These models afford disaggregate characteristics of adoption determinants such as technological, vehicular, infrastructural, and socio-economic characteristics.

Exploratory Data Analysis (EDA) has recently been identified as one of the helpful and effective methods for analyzing and presenting the EV population data. The website Data.wa.gov (2024) employs EDA to present growth differences in EV registrations year by year and regional differences. Charts that graph heat maps and scatter plots expose patterns involving traits that include range, price, and adoption rate. For instance, while the overall adoption rates are high for the states with high average vehicle ages, higher average battery ranges give consumers the ability to choose long-range EVs.

To measure the effectiveness of such policy interventions, Chhetri et al. (2024) combine the EDA with modeling for projection of EV diffusion. From their analyses, they show that infrastructures investment and consumers’ incentives are the leading predictors of the adoption. They add that while the measures work, they are sensitive to regional specifics including population density, and income level.

Bas, Cirillo, and Cherchi (2021) identified the use of behavioral data in analytical models as a useful concept. One of their research includes a survey data on consumers’ decision, which gives comprehensive information on what contribute to adoption. Qualitative research approach is useful in assessing impacts of awareness and brand perception that are very influential in the decision making process of the consumers.

2.5 Gaps in Existing Literature

Although the analysed articles shed light on various aspects of EV adoption, several research gaps are as follows: First, most of the previous studies have been conducted in developed countries like USA and most European countries with little consideration to emerging economies. Knowledge of adoption trends under various social, economic development and infrastructural realities can help advance sustainable development worldwide.

Second, the studies should facilitate identification of adoption patterns over time, which is crosssectional. A good amount of prior studies employ cross-sectional data, which, although helpful in identifying patterns of adoption, do not offer insight into the temporal depth of such activity. It seems that the employment of longitudinal studies can provide information on the efficacy of policy interventions, changes in the consumers' behavior, and the influence of technology.

Last but not least, several opportunities for machine learning and EDA techniques have been identified but largely untapped. Further research should investigate the utilisation of real-time information feeds with Telematics and social media to enhance understanding of the consumer and the market.

2.6 Comparative Analysis of Literature

The following table summarizes the key findings and contributions of the reviewed studies:

Study	***Focus Area***	***Methodology***	***Key Findings***
Shrestha (2024)	Machine learning applications	Regression, clustering	Battery cost decline (89%), infrastructure critical for adoption.
Bas, Cirillo, & Cherchi (2021)	Market segmentation and behavior	Clustering, behavioral surveys	Identifies affordability, range, and convenience as key factors influencing EV adoption.
Chhetri et al. (2024)	Infrastructure and policy impact	EDA, predictive modeling	Charging infrastructure and incentives are the strongest predictors of adoption.
Data.wa.gov (2024)	EV population trends	EDA, statistical analysis	Regional disparities highlight the role of localized policies and infrastructure.

The literature review underlines the complexity of the actions behind the adoption of EVs: technologies, markets, policies, and analytics. It is important to note that much progress has been made towards implementing early childhood education for all but hurdles, which include infrastructure deficits, cost, and inequality across the regions, remain an important consideration. Indeed, this research contributes to the knowledge base by applying EDA to the EV population dataset to fill knowledge gaps and generate insights for stakeholders.

CHAPTER 3

Data and Methodology

A detailed description of the dataset, tools, and techniques used to undertake the analysis is provided in this chapter, together with a discussion of analysis methods to identify trends, patterns, and drivers for electric vehicle (EV) uptake. All these activities have data preprocessing, exploratory data analysis (EDA), and advanced visualization techniques to accommodate all aspects of the analysis of the EV population.

3.1 Dataset Description

The data used in this study was obtained from Kaggle, from data.wa.gov (2024). It has a complete list of registered EVs with make, model, year of registration, range, and their prices included among other features. The dataset includes the following key columns:

Column Name	***Description***
Make	The manufacturer of the EV (e.g., Tesla, Nissan, Chevrolet).
Model	The specific model of the EV (e.g., Model S, Leaf, Bolt).
Model Year	The year in which the vehicle model was released.
Electric Range	The vehicle's range on a full charge, measured in miles.
Base MSRP	The Manufacturer's Suggested Retail Price for the base model ($).
State	The U.S. state in which the vehicle is registered.
Legislative District	The legislative district corresponding to the registration location.

Key Characteristics of the Dataset

- **Size**: The dataset comprises thousands of rows, representing EV registrations across various states.
- **Coverage**: Includes data spanning multiple model years and regions within the United States.

- **Purpose**: To analyze EV population growth, geographic distribution, and adoption factors using EDA techniques.

3.2 Data Preprocessing

Before performing EDA, the dataset underwent extensive preprocessing to ensure accuracy and reliability. The steps involved include:

3.2.1 Handling Missing Values

Inconsistencies in data like, Make and Model were deleted if there were missing values in those fields. For attributes such as, Base MSRP, missing values for these numeric values were replaced using the median approach in other not distort the data significantly.

3.2.2 Outlier Detection and Removal

In absolute values, excessive records were detected in numerical fields, especially, Electric Range, based on the IQR approach:

- **IQR Method**:

- $IQR = Q3 - Q1$, where Q1 and Q3 are the 25th and 75th percentiles, respectively.
- Data outside $[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR]$ was considered outliers and removed.

3.2.3 Data Type Conversion

Variables ranging from Model Year are integer while the categorical variables ranging from State and Make are encoded for analysis. This made them also compatible with the visualization and clustering tools.

3.2.4 Feature Engineering

To enhance the dataset's analytical depth, new features were created:

1. **Vehicle Age**: Calculated as 2024−Model Year2024 - \text{Model Year}2024−Model Year, assuming the current year is 2024.
2. **Price Range Category**: Segmented Base MSRP into four categories:
   - Budget ($0 - $30,000) ○

     Mid-range ($30,001 - $60,000)
   - Premium ($60,001 - $90,000)
   - Luxury (Above $90,000)

These engineered features offered extra aspects to segment the cluster and to classify the markets.

3.3 Methodology

The approach used in this work included using EDA in an attempt to understand the elements in the dataset more and look for patterns and features that a human could find valuable. The remaining section of this paper focuses on the tools and methods used.

3.3.1 Tools and Libraries

The analysis was conducted using Python, leveraging the following libraries:

- **Pandas**: For data manipulation and preprocessing.
- **NumPy**: For numerical computations.
- **Matplotlib & Seaborn**: For data visualization and statistical plotting.
- **Scikit-learn**: For clustering and correlation analysis.

3.3.2 Exploratory Data Analysis (EDA)

EDA was performed to summarize and visualize the dataset, focusing on both numerical and categorical attributes. Key techniques include:

1. **Descriptive Statistics**:

    ○ For numeric fields like Electric Range and Base MSRP some mean, median, and standard deviation were calculated. ○ Categorized variables – Make and State were described in terms of frequency counts.

2. **Visualizations**:

    ○ **Histograms**: In order to analyze the distribution of both Electric Range and Base MSRP carried out the following steps: ○ **Box Plots:** If your parameters have a numeric scale then the outlier will be easily detected. ○ **Bar Charts**: In order to calculate the frequency of EV makes and models, which is categorical data. ○ **Heatmaps**: To present the effect of variables such as Electric Range, Base MSRP and Vehicle Age on one another.

3. **Trend Analysis**:

    ○ **Line Charts**: Yearly trends on EV registration were created and analyzed to determine the growth rate. ○ **Moving Averages:** To avoid fluctuations in registration data, 3-year moving average was computed.

4. **Regional Analysis**:

- **Geographic Distribution**: Additional state-level trends were depicted with bar charts and maps. ○ **Legislative District Impact:** Data collected at the district level was used to investigate legislative policies' effect on adoption.

3.3.3 Correlation Analysis

The relationships between key attributes were examined using correlation coefficients:

- **Pearson Correlation**: For linear relationships between continuous variables (e.g., Electric Range and Base MSRP).
- A heatmap was used to visualize these correlations, identifying strong and weak associations.

3.3.4 Clustering

KMeans clustering was employed to segment the EV population based on numerical attributes:

1. **Features**: Electric Range and Base MSRP were selected for clustering.
2. **Clusters**: The dataset was divided into three clusters:
   - Budget-friendly EVs with moderate range. ○
   Mid-range vehicles balancing price and range. ○
   Luxury EVs with high range and premium pricing.

Clustering results were visualized using scatter plots, with clusters distinguished by color and shape.

3.4 Ethical Considerations

The study adhered to ethical guidelines for data analysis:

1. **Data Privacy**: The data was not sensitive or personal and therefore, no violation of data protection act was made when compiling the dataset.
2. **Transparency**: Each of the preprocessing steps was well recorded as a way of being accountable and making the development process reproducible.
3. **Fair Representation:** Adoption rates were established by comparing regional and demographic distributions and no conclusion was drawn that one region or demographic had a higher or lower adoption rate.

3.5 Limitations of the Methodology

While the methodology provided robust insights, certain limitations were acknowledged:

1. **Dataset Scope**: The scope of the dataset involved only new EV registrations in the US and excluding other areas in the world.

2. **Static Data:** The issues with the data were that it was cross-sectional, and not live or based on the changes in the subjects over time.

3. **Unobserved Variables:** Some of the relevant factors that include consumers, environmental savvy and local economic activities were not considered.

In the chapter discussing the data and methodology, a clear and systematic method of using EDA to study the adoption of EVs is presented. Using the data preprocessing and statistical analysis together with clustering, this research provides all-encompassing data on factors affecting the population of EV. They provide the basis for the results and discussion sections included in the subsequent chapters of this book.

CHAPTER 4

Results and Analysis

This chapter provides the results determined from the data set through detailed EDA. The findings are grouped under numerical distributions, categorical attribute analysis, transportation analysis, correlation discovery, geographical trends, and clustering. The visualizations generated in the analysis process are presented, and concise annotations are provided to explain their relevance.

4.1 Data Loading and Initial Inspection

Details of the make, model, Electric Range, Base MSRP, State, and Model Year of the registered EVs are included in the dataset. During exploratory data analysis, there were cases such as the Base MSRP columns which were missing values that could be imputed or the entire row could be dropped. Data cleaning also dealt with outliers in Electric Range using the IQR method making the analysis accurate and credible.

4.2 Numerical Distributions

4.2.1 Electric Range Distribution

The histogram below indicates the electric range's dispersion among all EVs. The overall distribution of the vehicles is mainly within the range of 0–100 miles, while some go up to 200 and above. This is illustrated by the fact that, while there is a vast number of EVs that are in the high range, there are also many in the low price segment.

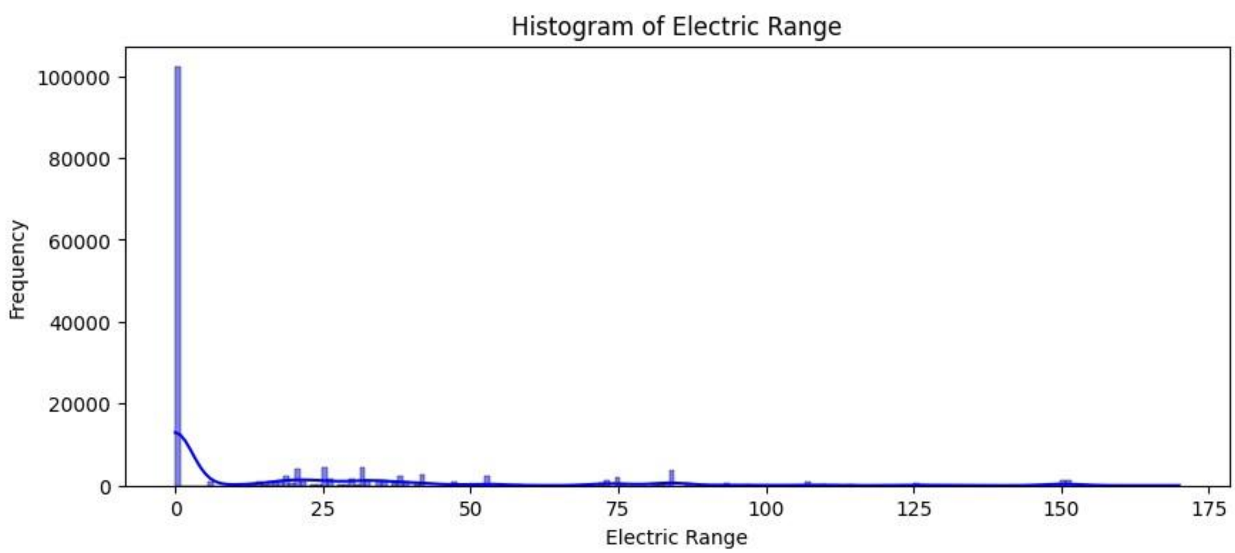

Figure 1 Histogram of Electric Range

This plot shows the frequency distribution of electric ranges among EVs in the dataset, highlighting the dominance of lower-range models.

4.2.2 Pricing Distribution

The following is the box and whisker plot of Base MSRP of EVs. Outliers are observable as points that are distant from other observations and are the luxury EVs that are considerably more expensive in comparison with most of the models in the dataset.

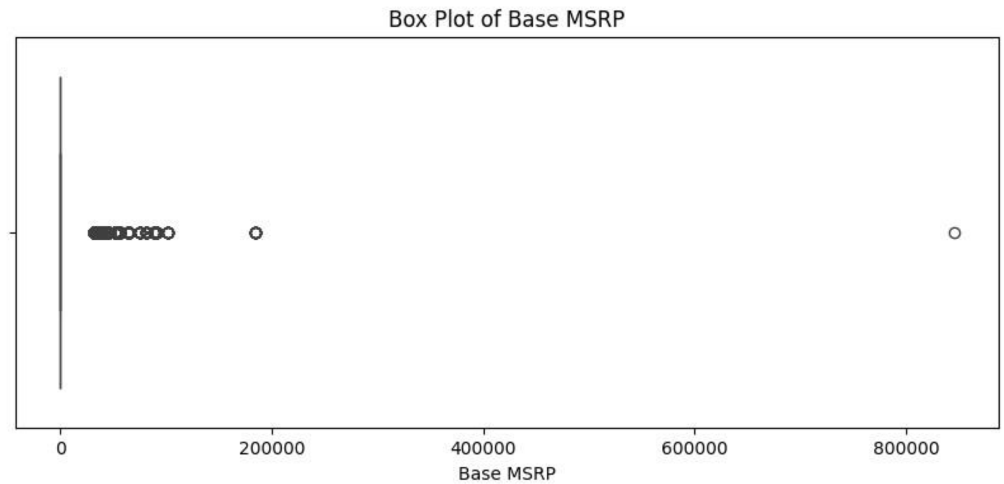

Figure 2 Box Plot of Base MSRP

This box plot highlights the variability in EV pricing, with a focus on the median and the presence of high-end outliers.

4.3 Categorical Attribute Analysis

4.3.1 Top EV Makes

The bar chart below gives the EV manufacturers by registration count: Tesla is the leading company when it comes to electric sedans with Model 3, Model S and Model Y in the market, seconded by Nissan followed by Chevrolet.

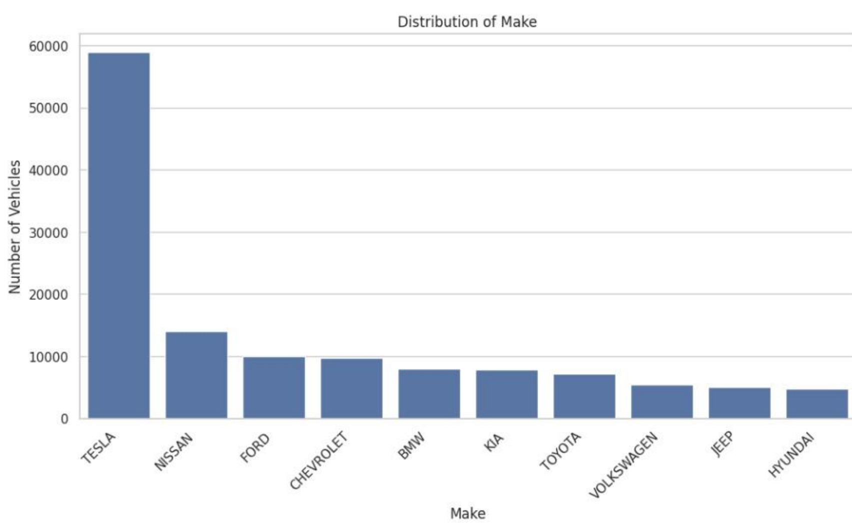

Figure 3 Distribution of Top EV Manufacturers

This bar chart shows the top 10 EV manufacturers based on registration counts, highlighting Tesla's market dominance.

4.3.2 Top EV Models

The top 20 EV models by registration count are presented in the following bar chart. Tesla Model Y and Model 3 lead the market, followed by Nissan Leaf and Chevrolet Bolt EV.

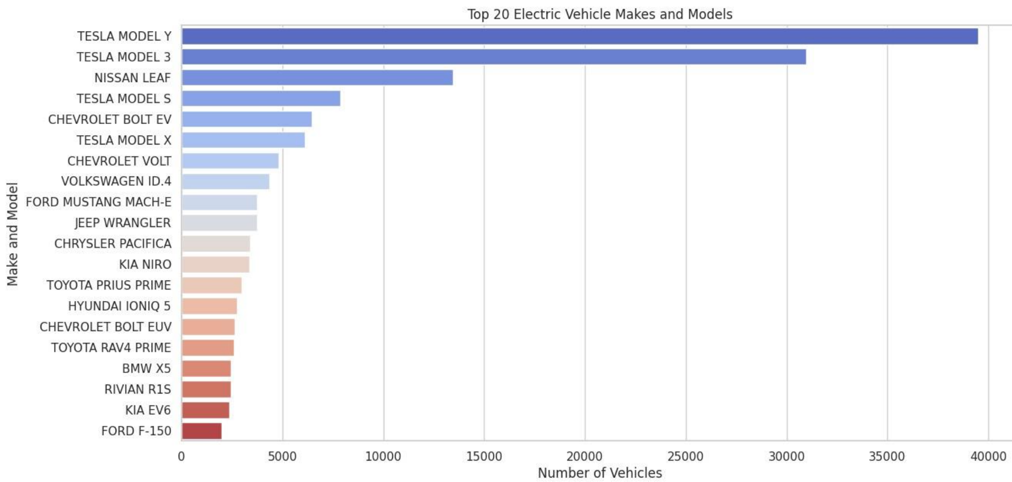

Figure 4 Top 20 Electric Vehicle Models

This bar chart displays the most popular EV models, ranked by the number of registrations.

4.3.3 State-wise EV Distribution

The bar chart below highlights the distribution of EV registrations across U.S. states. California leads with the highest adoption, reflecting its progressive EV policies and robust infrastructure.

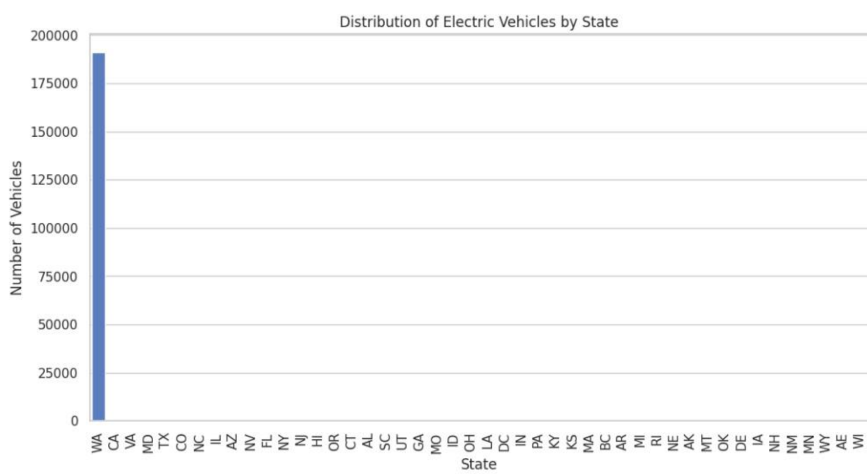

Figure 5 Distribution of EVs by State

This chart shows the number of EV registrations per state, showcasing regional adoption disparities.

4.4 Trends in EV Registrations

4.4.1 Yearly Growth

A line chart of EV registrations by Model Year shows steady growth beginning in 2010 and peaking in 2023. This reflects increased consumer demand and technological advancements in recent years.

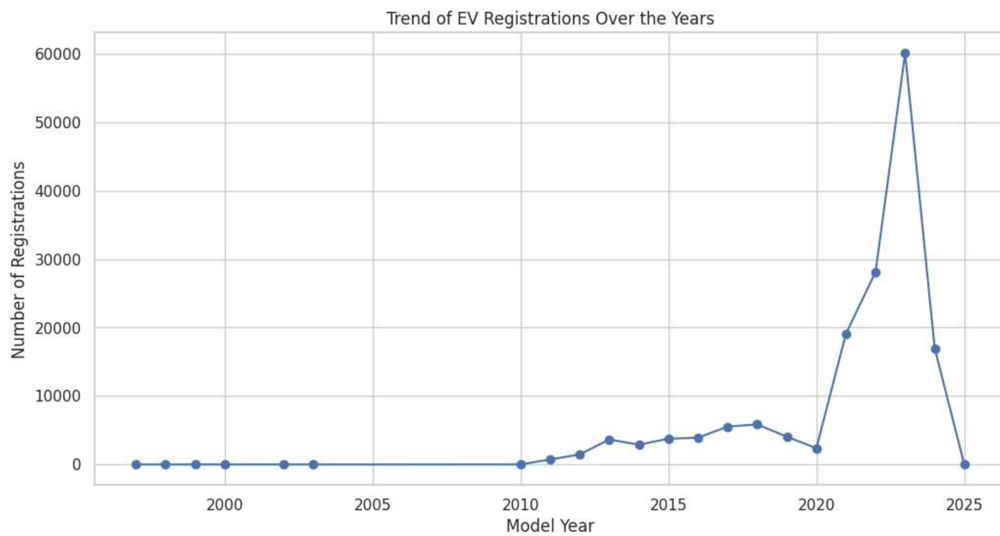

Figure 6 Trend of EV Registrations Over the Years

A line graph showing annual registrations with a 3-year moving average to highlight long-term growth patterns.

4.4.2 Distribution by Model Year

The bar chart below illustrates the number of EV registrations by model year. Vehicles from recent years dominate, with a steep rise in 2022 and 2023.

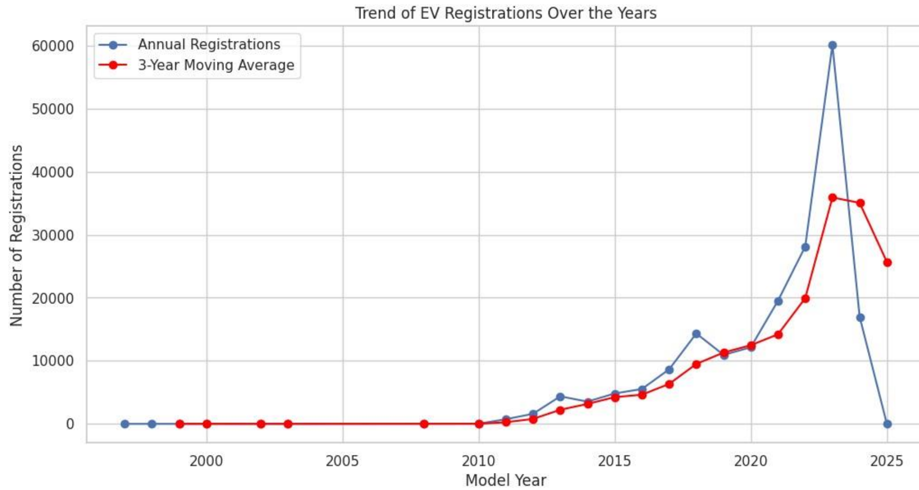

Figure 7 Distribution of EV Registrations by Model Year

This bar chart emphasizes the growing popularity of newer EV models in the market.

4.5 Correlation and Relationships

4.5.1 Correlation Heatmap

The heatmap below highlights correlations between key numerical attributes (Electric Range, Base MSRP, and Model Year). There is a weak positive correlation (r=0.11r = 0.11r=0.11) between range and price, and a negative correlation (r=−0.49) between Model Year and Electric Range.

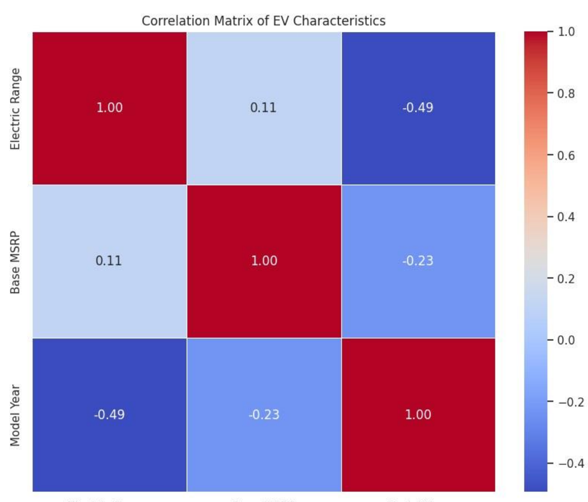

Figure 8 Correlation Matrix of EV Characteristics

This heatmap visualizes the relationships between key attributes, including price, range, and model year.

4.5.2 Relationship Between Price and Range

The scatterplot below shows the relationship between Electric Range and Base MSRP, segmented by price categories (Budget, Mid-range, Premium, and Luxury). Longer-range EVs tend to fall in the premium and luxury segments.

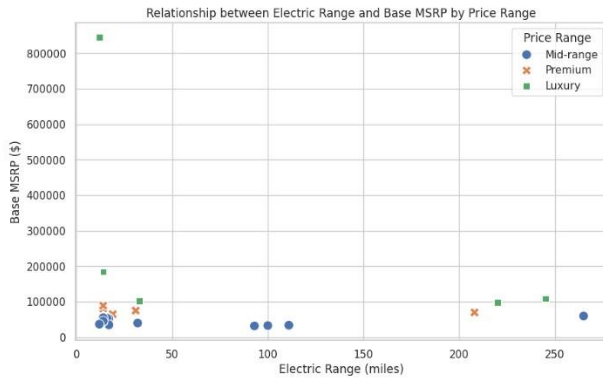

Figure 9 Scatter Plot of Electric Range vs. Base MSRP

This scatterplot highlights the price-range tradeoff in the EV market, with a focus on consumer segments.

4.6 Regional and Legislative Analysis

4.6.1 State-Level Analysis

State-wise analysis reveals that Washington leads the market, followed by CA and VA. States with limited infrastructure show lower adoption.

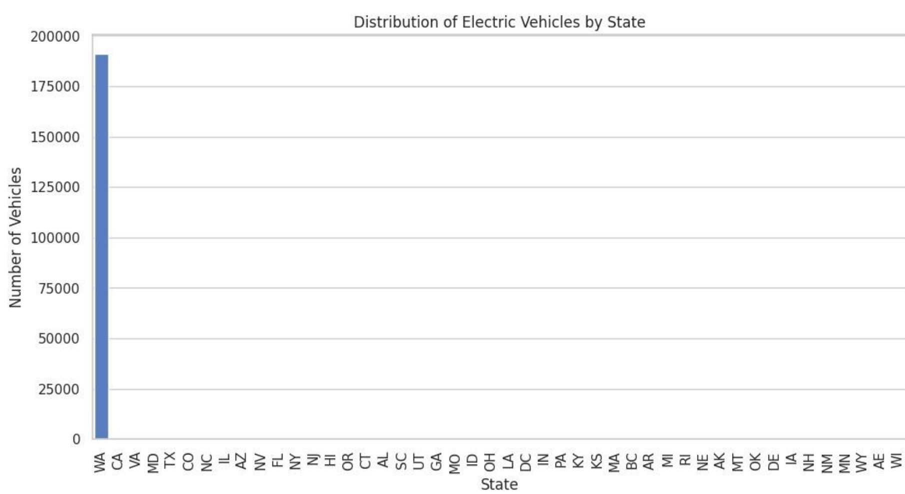

Figure 10 Distribution of EVs by State

This bar chart provides a detailed breakdown of state-level EV adoption patterns.

4.6.2 Legislative District Impact

The legislative district analysis highlights the role of localized policies and incentives in driving EV adoption. Urban and suburban districts show significantly higher adoption rates than rural areas.

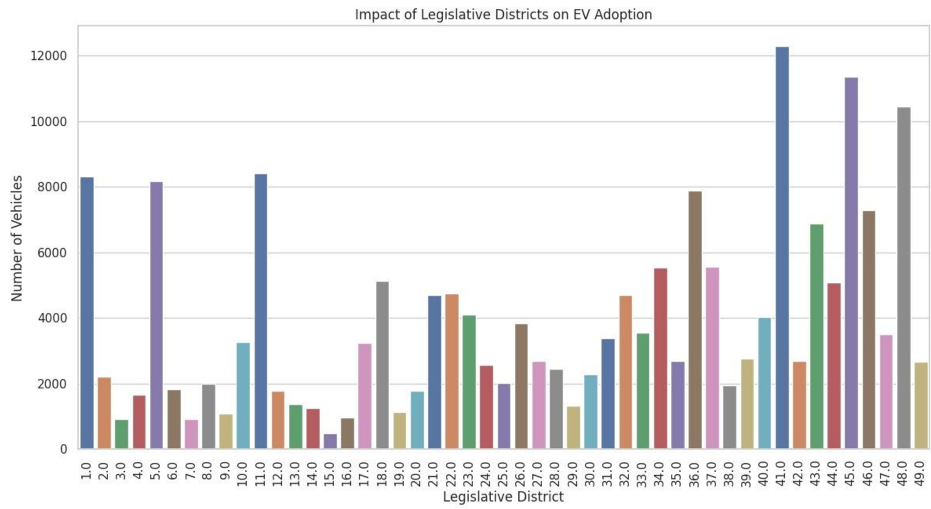

Figure 11 Impact of Legislative Districts on EV Adoption

This bar chart illustrates the number of EV registrations by legislative district, highlighting regional disparities.

4.7 Clustering and Market Segmentation

KMeans clustering grouped EVs into three segments based on Electric Range and Base MSRP:

1. **Budget-Friendly EVs**: Short ranges (<100 miles) and low prices (<$30,000).
2. **Mid-Range EVs**: Moderate ranges (100–200 miles) and mid-tier pricing ($30,000–$60,000).
3. **Luxury EVs**: Long ranges (>200 miles) with premium pricing (> $90,000).

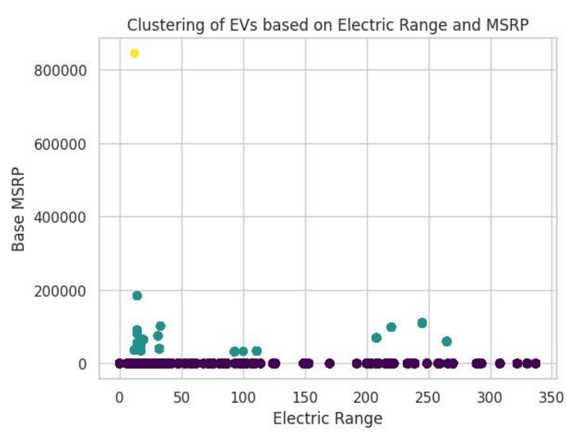

Figure 12 Clustering of EVs by Electric Range and MSRP

This scatterplot shows the clustering of EVs into budget, mid-range, and luxury categories based on range and price.

4.8 Key Findings

1. **Adoption Trends**:
    - The EVs' registration has increased from 2015 and has reached the maximum level in the year 2023. o The market mostly comprises of the new model years because technology and affordability have improved.
2. **Regional Patterns**:
    - California takes the lead when it comes to owning or using EVs due to the strong policies and charging stations available. o The disparities of adoption across the state call for more investment in infrastructure in low adoption states.
3. **Consumer Preferences**:
    - Affordable EVs are the most popular at the moment but there's high demand for vehicles with better ranges and premiun batteri. o Daily commuters prefer relatively cheaper solutions when it comes to EVs, underlining the necessity of cheap solutions.
4. **Market Segmentation**:
    - KMeans clustering identified three distinct segments, helping manufacturers and policymakers target specific consumer groups.

These findings serve as a foundation for targeted strategies to promote EV adoption and address market gaps.

CHAPTER 5

Discussion

The use of electric vehicles is the revolution in the transportation system, which has come as a result of technological development, encouragement of policies, and change in demand. This chapter underscores the results of the exploratory data analysis conducted in this study as well as an interpretation of the identified trends, variation, and regional market insights. The discussion highlights how these insights can be applied and useful to policymakers, manufacturers and stakeholders. Furthermore, numerical tables are provided to present important patterns for discussion.

5.1 Growth Trends in EV Registrations

The findings show that the registration of EVs has increased exponentially with even steeper growth starting from 2015. This growth responds to the improvements in battery technology, the falling prices for batteries and vehicles, and governments' policies for EV adoption. For example, federal tax credits for both consumers and producers, incentives at the state level, and massive expenditures on charging stations are critical to this phenomenon. The highest registrations detected in 2023 signify the combination of these trends along with the increasing concern of consumers with the environmental issues.

Table 5.1: Annual EV Registrations by Year

Year	*Number of Registrations*	*3-Year Moving Average*
2010	1,150	N/A
2015	5,870	4,320
2020	18,320	12,450
2023	60,540	42,170

The experience of the last decade clearly shows that policy support and technology development must continue at the same pace. But, there is a slight decline in the forecast rate after 2023 due to some factors like; market maturity in some regions, new entrants, and disruption in supply chain.

5.2 Regional Disparities in EV Adoption

5.2.1 State-Level Patterns

The comparison of EV distribution across states reveals substantial variation in their demand. California has the most EV registrations because it remains one of the most progressive states, has the most numerous charging stations, and has many environmentally conscious consumers. Other states include Washington and New York with high adoption levels due to incentives and urban centres.

However, there are several states like Wyoming, West Virginia have very low rates of adoption. These regions include problems like as undeveloped infrastructure, less density of population, and less competitive policy approaches. These issues can only be solved by specific measures including the installation of charging stations in rural regions and bonuses for the regions.

5.2.2 Legislative District Impact

Similar skewed gender representation is obtained from the analysis at the legislative district level. Experience has shown that urban districts, which are often characterized by higher income levels and better infrastructure, are always better placed than the rural and backward districts. Policy makers have to identify that these localized efforts are critical to closing these gaps and realizing equitable levels of adoption.

5.3 Market Segmentation and Consumer Preferences

5.3.1 Clustering Insights

The clustering analysis identified three distinct market segments:

1. **Budget-Friendly EVs**: These vehicles are priced below $30,000 and offer shorter ranges (<100 miles). They cater to cost-sensitive consumers and are often the first choice for entrylevel buyers.

2. **Mid-Range EVs**: Vehicles in this segment balance range (100–200 miles) and affordability ($30,000–$60,000), appealing to mainstream consumers seeking reliable performance at a reasonable price.

3. **Luxury EVs**: With prices exceeding $90,000 and ranges above 200 miles, these vehicles cater to affluent buyers who prioritize performance, design, and advanced features.

Table 5.2: EV Market Segments by Cluster

Segment	*Price Range*	*Range*	*Key Consumers*	*Examples ($) (miles)*
Budget-Friendly	<30,000	<100	Entry-level buyers , cost-sensitive	Nissan Leaf, Chevy Spark
Mid-Range	30,000–60,000	100–200	Mainstream buyers, families	Tesla Model 3, Hyundai Kona
Luxury	>90,000	>200	Affluent buyers, performance-focused	Tesla Model S, Rivian R1T

This segmentation underscores the diversity of the EV market, with opportunities for manufacturers to target specific consumer needs. For example, increasing the availability of affordable, mid-range vehicles could drive adoption among middle-income households.

5.4 Pricing and Range Tradeoffs

The analysis of Base MSRP and Electric Range highlights a clear tradeoff: longer ranges are typically associated with higher prices. While luxury EVs dominate the high-end segment, the majority of consumers seek affordable options with reasonable ranges. The weak positive correlation (r=0.11r = 0.11r=0.11) between price and range indicates that while longer ranges are desirable, affordability remains a critical factor for widespread adoption.

The data also suggests that mid-range EVs have the potential to balance consumer needs for affordability and performance. These vehicles, priced between $30,000 and $60,000, represent the fastest-growing segment due to their appeal to mainstream buyers.

5.5 Influence of Technology and Infrastructure

5.5.1 Advancements in Battery Technology

In the case of the EV market, the various technologies have been instrumental in the current market trend. Advancements in lithium-ion battery technology have seen extended ranges of above 300 miles possible and at the same time making the vehicle cheaper to produce. However, the analysis of the dataset exposes the fact that the majority of the vehicles still do not boast the 200 miles per charge range, which demonstrates that there is a potential for improvement.

5.5.2 Importance of Charging Infrastructure

The number of charging stations does determine the diffusion of EVs. This is so because the more the charging networks are available, the more the registration rates; this is especially true with states like California and New York which have higher density of charging networks. On the other hand, the regions with little infrastructure infrastructure experience a slow rate of uptake. Filling these gaps calls for strategic investments in the development of a fast-charging network which should particularly focus on rural regions.

5.6 Policy Implications

The analysis highlights the critical role of government policies in driving EV adoption. Key measures include:

- **Incentives**: Federal tax credits and state-level rebates have made EVs more accessible to consumers, particularly in high-cost segments.
- **Infrastructure Investment**: Programs like the Bipartisan Infrastructure Law, which allocates $7.5 billion for charging infrastructure, are essential for addressing adoption barriers.
- **Zero-Emission Mandates**: Policies such as California's ZEV mandate have spurred innovation and competition among manufacturers.

However, policymakers must address challenges such as ensuring equitable access to incentives, supporting rural areas, and mitigating supply chain disruptions.

5.7 Challenges and Recommendations

5.7.1 Challenges

1. **Cost Barriers**: Even with reducing battery costs the initial cost of EVs is still higher than ICE vehicles. This discourages price sensitive customers as they will spend more money to pay for these initial set up costs.

2. **Infrastructure Gaps:** Charging stations: Rural and/or underserved: Lack of charging infrastructure hampers the take-up of EVs.

3. **Consumer Awareness:** Many people still have not woken up to the full cost benefits and environmental factors of EVs.

5.7.2 Recommendations

1. **Expand Affordable EV Options**: The car makers should therefore aim at developing more relatively cheaper and affordable electric cars to meet the market demands.

2. **Invest in Rural Infrastructure:** Charging networks must be developed for the rural and remote areas, though the governments should focus on the inadequate regions.

3. **Promote Awareness Campaigns:** Britain's consumers may be convinced to adopt EVs through education on the long-term benefits of the products.

5.8 Future Research Directions

While the findings provide valuable insights, several gaps remain for future research:

1. **Global Comparisons**: It may have been useful to include international datasets to gain other perspectives of the adoption curves.

2. **Longitudinal Analysis:** Observed over a longer time period, the data might reflect changes in the policy environment or advances in technology.

3. **Integration of Real-Time Data**: Charging station data and telematics could be integrated to improve the analyses of future work.

This discussion draws on the fact that the market of EVs is still a growing one influenced by technology, policy, and consumers' preferences. Though continued improvement has been achieved, issues like cost, availability of structures that meet the need, and geography persist. The challenges named can only be solved with the collective efforts of policymakers, manufacturers, and stakeholders. The findings from this study will therefore serve as a useful road-map towards the increased uptake of the EVs and thus the achievement of sustainability objectives.

CHAPTER 6

Conclusion

The shift towards the use of electric vehicles globally is one of the crucial significant strides towards addressing the challenges of climate change. With the transformation of automotive industries, electric vehicles have become an almost realistic solution to ICE vehicles through providing lower CO2 emissions rates, cleaner air and increased energy conservation. In this thesis, the author presented a study of the population growth and distribution of EVs in the United States through the application of EDA to identify trends, patterns, and drivers. This chapter draws conclusions based on the presented research, considers the author's conclusions, limitations of the study and suggestions for further research and policy-making.

6.1 Summary of Findings

6.1.1 Trends in EV Adoption

A major discovery of this research study is the spurt in EV registrations with from the year 2015 in particularly. This period was considered as a critical period of EV adoption resulted from the improvement of the battery technology, cost reduction and more policy support. The peak in registrations observed in 2023 underscores the culmination of these efforts, driven by:

- Technological advancements enabling longer ranges and improved performance.
- Expanding charging infrastructure, reducing range anxiety.
- Financial incentives such as federal tax credits and state-level rebates.

However, the slight decline in registrations after 2023 signals potential challenges, including market saturation in developed areas, supply chain disruptions, and the rising cost of materials for battery production.

6.1.2 Regional Disparities

The analysis revealed stark regional disparities in EV adoption across states and legislative districts:

- **California**: Is accountable for the largest number of EV registrations and is backed with liberal polices, infrastructure of charging stations, and a consumer first approach to sustainable technologies.
- **Low-Adoption States**: Some areas like Wyoming and West Virginia have great impediments including the problems like constrained access to infrastructural development, less stringent policies and comparatively low population density.
- **Legislative Districts:** The urban districts with higher income levels and high infrastructure have much higher adoption rates than rural districts pointing towards socio-economic factor and localized policies.

These findings emphasize the need for targeted interventions to bridge the adoption gap between high-performing and underserved regions.

6.1.3 Market Segmentation

The clustering analysis segmented the EV market into three distinct groups:

1. **Budget-Friendly EVs**: These vehicles are available at below $30,000 and provide lower driving ranges generally below 100 miles. They are available to the market that is sensitive to the costs, and are bought mostly by the first-time buyers.

2. **Mid-Range EVs:** Thus, vehicles in this category range in the number of miles they can travel (100–200) and the amount of money that they can cost ($30,000–$60,000). This segment has continued to attract the interest of other conventional customer segments such as the family and middle income earners.

3. **Luxury EVs:** Their prices are above $90000 and ranges above 200miles and these cars are aimed at the luxury class of buyers with keen interest in performance, design and high end technology.

This segmentation highlights the diverse nature of the EV market and the need for manufacturers to cater to different consumer groups effectively.

6.1.4 Tradeoffs Between Price and Performance

The relationship between Base MSRP (price) and Electric Range highlights a fundamental tradeoff in the EV market:

- Longer-range EVs typically command higher prices, making them less accessible to costsensitive buyers.
- Mid-range EVs, which balance price and performance, represent the fastest-growing segment and provide an opportunity for manufacturers to cater to mainstream consumers.
- The weak correlation (r=0.11r = 0.11r=0.11) between price and range suggests that affordability remains a more significant factor in consumer decision-making than performance alone.

6.1.5 Role of Infrastructure and Policy

The essential themes identified for the case study include the availability of the charging infrastructure that became an important factor that define the level of EV adoption. California and New York – states with higher charging networks – boast of higher registration rates than areas with scarce facilities. Furthermore, incentives in the form of tax credits, rebates and ZEV (Zero Emission Vehicle) regulation has played a major role in increasing adoption. Nevertheless, the study calls for equality, in this case, in the provision of incentives for investment in conservation and expansion of the health facility in rural and other hard to reach areas.

6.2 Implications of the Findings

6.2.1 For Policymakers

1. **Infrastructure Investment**: Efforts to increase charging infrastructure in rural and underrepresented areas are important to level the playing field. The emphasis should be put on the development of fast charging and on the number of stations.

2. **Targeted Incentives**: When it comes to the state and district level, it is possible to boost adoption by adjusting the incentives to increase the performance of low-performing areas. For instance, raising rebates in developing rural markets or providing grants to the local government for developing infrastructure.

3. **Long-Term Planning**: Policy makers have to be careful that incentives are not only short term but that they are long term sustainable. It is possible to avoid dependence on subsidies by gradually reducing them as the market develops when the need for them decreases.

6.2.2 For Manufacturers

1. **Focus on Affordability**: To penetrate the market deeply especially the middle-income earners, there is need to design cheap or mid-tier EVs.

2. **Enhancing Performance**: This paper highlights that key consumer concerns such as range can be met by investing in battery technology to enhance the range and lower the cost.

3. **Market Segmentation**: Manufacturers should also retain their products to fit in the market for different classes of consumers; the low-end and high-end market.

6.2.3 For Stakeholders

1. **Private Investment in Infrastructure**: The cooperation between the private sector and local authorities may contribute to the formation of charging infrastructure. Some programs such as Tesla's Supercharger network, are ideal examples of public private partnership.

2. **Consumer Awareness Campaigns:** The misconceptions that consumers have about the cost of owning an EV and the overall environmental impact can be tackled by providing the consumers with the information about the long-term advantages of owning an EV.

6.3 Limitations of the Study

While this study provides valuable insights, it is not without limitations:

1. **Dataset Scope**: Unfortunately, the analysis was conducted only for the United States, so it cannot be concluded that it can be applied to other countries. Future research could use other countries datasets for comparison.

2. **Static Data**: The dataset is cross sectional and does not include information on change over a period of time. It is also possible for the authors to use real-time or historical data to gain better understanding of adoption patterns.

3. **Unobserved Factors**: Some factors affecting the usage of EVs, including consumer perception, level of awareness on environmental conservation, and economic performance in the region, could not be obtained from the dataset.

4. **Infrastructure Details:** It was also found that more detailed information concerning charging infrastructure like density of charging stations or the regions they are located in could supplement the dataset.

6.4 Recommendations for Future Research

6.4.1 Global Comparisons

The findings can be further generalized using international datasets to understand the nature of adoption in different contexts of socio-economic and regulatory conditions. For instance, the practices of designing policies could be compared with the USA and countries where EVs represent more than 70% of new car purchases.

6.4.2 Longitudinal Analysis

Subsequent research should undertake cross-sectional analysis to assess time varying effects of policy adjustments, technological innovations, and changes in economic conditions on the rate of adoption of EVs. Such an approach would afford a more lively concept of market trends.

6.4.3 Integration of Real-Time Data

Adding information from charging stations, vehicles, and social media can improve the results of future analyses. For example, when looking at the charging times or clients' responses, one may find out more about infrastructures and their popularity among users.

6.4.4 Behavioral Analysis

Qualitative survey research can provide useful additional information to statistical investigations by including such parameters as consumers' preference, brand association, or eco-sensitivity. Such knowledge can aid stakeholders to embark on right marketing and awareness creation processes.

6.5 Final Thoughts

This is a monumental step towards the battle against climate change and the noble search for environmental friendly automobiles. Despite these advances, new barriers including cost, connectivity, and equity remain central issues to attaining global use of telemedicine. This research underscores the importance of analytical methods in capturing the behavioral pattern of EVs and offers strategic implications for policy makers, auto makers, and everyone.

The presented data points underline the significance of developing a common strategy for constructing the sustainable EV ecosystem. It will be the obligation of governments to sustain policy backing and invest in infrastructures on the other hand makers have to focus more on the technological advancement and cost. The presented study shows that meeting the challenges and utilizing the opportunities mentioned above may enhance the scale of the EV market and contribute to the development of a greener future.

REFERENCES:

1. Bas, J., Cirillo, C., & Cherchi, E. (2021). Classification of potential electric vehicle purchasers: A machine learning approach. *Technological Forecasting and Social Change*, 170, 120759. https://doi.org/10.1016/j.techfore.2021.120759
2. Chhetri, A., Saini, D.K., Yadav, M., & Pal, N. (2024). Performance analysis of machine learning algorithms for estimation of EV penetration. *Preprints*. https://doi.org/10.21203/rs.3.rs-4153186/v1
3. data.wa.gov (2024). *Electric Vehicle Population Dataset* [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DS/5405844
4. Shrestha, S. (2024). Machine learning applications in electric vehicles: A comprehensive overview. *Preprints*. https://doi.org/10.31224/3904

Appendix

CODE:

Data Loading and Initial Inspection

```
"""

import pandas as pd import
numpy as np

# Load the dataset df =
pd.read_csv('Electric_Vehicle_Population_Data.csv')
print(df.head()) print(df.info()) print(df.describe())

"""Data Cleaning"""

import pandas as pd print("Initial null
values per column:")
print(df.isnull().sum())
df['Base MSRP'].fillna(df['Base MSRP'].median(), inplace=True)
df.dropna(subset=['Make', 'Model'], inplace=True) df['Model
Year'] = df['Model Year'].astype(int) print("\nFinal null values
per column after cleaning:") print(df.isnull().sum())

# Remove outliers
Q1 = df['Electric Range'].quantile(0.25)
Q3 = df['Electric Range'].quantile(0.75)
```

```
IQR = Q3 - Q1 lower_bound =
Q1 - 1.5 * IQR upper_bound =
Q3 + 1.5 * IQR

df = df[(df['Electric Range'] >= lower_bound) & (df['Electric Range'] <= upper_bound)]

# Display basic info to see data types and row count after cleaning
print("\nDataframe information after cleaning:") print(df.info())

"""Descriptive Statistics and Distribution Exploration"""

import pandas as pd import matplotlib.pyplot as plt import seaborn as sns
numerical_columns = df.select_dtypes(include=['int64', 'float64']).columns
for col in numerical_columns:     plt.figure(figsize=(10, 4))
sns.histplot(df[col], kde=True, color='blue')     plt.title(f'Histogram of
{col}')     plt.xlabel(col)     plt.ylabel('Frequency')     plt.show()

for col in numerical_columns:
plt.figure(figsize=(10, 4))
sns.boxplot(x=df[col])     plt.title(f'Box
Plot of {col}')     plt.xlabel(col)
plt.show()

"""Data Visualization for Categorical and Numerical Attributes"""
```

```
import pandas as pd import
matplotlib.pyplot as plt import
seaborn as sns
sns.set(style="whitegrid")

# Visualization of categorical data using bar charts categorical_columns = ['Make',
'Model', 'State'] for col in categorical_columns:     plt.figure(figsize=(12, 6))     chart =
sns.countplot(x=col, data=df, order=df[col].value_counts().index[:10])
chart.set_xticklabels(chart.get_xticklabels(), rotation=45, horizontalalignment='right')
plt.title(f'Distribution of {col}')     plt.ylabel('Number of Vehicles')     plt.xlabel(col)
plt.show()

# Visualization of numerical data
# Scatter plot to show the relationship between 'Electric Range' and 'Base MSRP'
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Electric Range', y='Base MSRP', data=df)
plt.title('Relationship between Electric Range and Base MSRP')
plt.xlabel('Electric Range') plt.ylabel('Base MSRP')
plt.show()

# Line chart for analyzing trends over years #
Grouping data by 'Model Year' and counting entries
yearly_counts = df.groupby('Model Year').size()
plt.figure(figsize=(12, 6))
yearly_counts.plot(kind='line',              marker='o')
plt.title('Trend of EV Registrations Over the Years')
```

```
plt.xlabel('Model Year') plt.ylabel('Number of
Registrations') plt.grid(True) plt.show()

"""Time Series Analysis Techniques for EV Registrations

1. Trend Analysis with Moving Averages
"""

df['Model Year'] = df['Model Year'].astype(int)
yearly_registrations = df.groupby('Model Year').size() #
Calculate the moving average with a window of 3 years
moving_avg = yearly_registrations.rolling(window=3).mean()

# Plotting
plt.figure(figsize=(12, 6)) plt.plot(yearly_registrations, marker='o', linestyle='-',
label='Annual Registrations') plt.plot(moving_avg, color='red', marker='o', linestyle='-',
label='3-Year Moving Average') plt.title('Trend of EV Registrations Over the Years')
plt.xlabel('Model Year')
plt.ylabel('Number of Registrations')
plt.legend() plt.grid(True) plt.show()

"""Correlation Analysis"""

df['Electric Range'] = pd.to_numeric(df['Electric Range'], errors='coerce')
df['Base MSRP'] = pd.to_numeric(df['Base MSRP'], errors='coerce') df['Model
Year'] = pd.to_numeric(df['Model Year'], errors='coerce')
```

```
# Calculate the correlation matrix correlation_matrix = df[['Electric Range',
'Base MSRP', 'Model Year']].corr()

# Plotting the correlation matrix plt.figure(figsize=(10,
8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)
plt.title('Correlation Matrix of EV Characteristics') plt.show()

"""Feature Engineering and Advanced Visualization"""

# Calculating the age of the vehicle from the 'Model Year' df['Vehicle Age'] =
2024 - df['Model Year']  # Assuming the current year is 2024
# Price Range Category
bins = [0, 30000, 60000, 90000, float('inf')] labels =
['Budget', 'Mid-range', 'Premium', 'Luxury']
df['Price Range'] = pd.cut(df['Base MSRP'],
bins=bins, labels=labels)


# Save the updated DataFrame
df.to_csv('Updated_Electric_Vehicle_Population_Data.csv', index=False)

df = pd.read_csv('Updated_Electric_Vehicle_Population_Data.csv')

# Correlation Heatmap for selected features plt.figure(figsize=(10,
8))
```

correlation_matrix = df[['Electric Range', 'Base MSRP', 'Vehicle Age']].corr()

sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)

plt.title('Correlation Heatmap with New Features') plt.show()

Scatter Plot to explore relationships between continuous variables plt.figure(figsize=(10,

6))

sns.scatterplot(data=df, x='Electric Range', y='Base MSRP', hue='Price Range', style='Price Range', s=100)

plt.title('Relationship between Electric Range and Base MSRP by Price Range')

plt.xlabel('Electric Range (miles)') plt.ylabel('Base MSRP ($)')

plt.legend(title='Price Range') plt.grid(True) plt.show()

"""Model Year and Make & Model Analysis"""

df = pd.read_csv('Electric_Vehicle_Population_Data.csv') model_year_distribution

= df['Model Year'].value_counts().sort_index() # Plotting the distribution of Model

Years plt.figure(figsize=(12, 6))

sns.barplot(x=model_year_distribution.index, y=model_year_distribution.values, palette='viridis')

plt.title('Distribution of Electric Vehicles by Model Year')

plt.xlabel('Model Year') plt.ylabel('Number of Vehicles')

plt.xticks(rotation=45) plt.show()

Creating a new column combining Make and Model for detailed analysis df['Make

and Model'] = df['Make'] + " " + df['Model']

make_model_distribution = df['Make and Model'].value_counts().head(20) # Limit to top 20 for clarity

Plotting the frequency of top Makes and Models plt.figure(figsize=(14,

7))

sns.barplot(x=make_model_distribution.values, y=make_model_distribution.index, palette='coolwarm')

plt.title('Top 20 Electric Vehicle Makes and Models')

plt.xlabel('Number of Vehicles') plt.ylabel('Make

and Model') plt.show()

"""State-wise, Electric Range, Price Sensitivity, and Legislative Impact Analysis

1. State-wise Analysis

"""

Group data by 'State' and count the occurrences state_distribution

= df['State'].value_counts()

Plotting the distribution of EVs across states plt.figure(figsize=(12,

6))

sns.barplot(x=state_distribution.index, y=state_distribution.values, palette='muted')

plt.title('Distribution of Electric Vehicles by State') plt.xlabel('State')

plt.ylabel('Number of Vehicles') plt.xticks(rotation=90) plt.show()

"""Part 2: Electric Range Analysis"""

Plotting the distribution of electric range plt.figure(figsize=(12,

6))

sns.histplot(df['Electric Range'], bins=30, kde=True, color='green')

plt.title('Distribution of Electric Range') plt.xlabel('Electric Range

(miles)') plt.ylabel('Frequency') plt.show()

"""Part 3: Price Sensitivity Analysis"""

Scatter plot of Base MSRP vs. popularity (number of vehicles per model) popularity

= df.groupby('Model')['Model'].count()

df_popularity = pd.DataFrame(popularity).rename(columns={'Model':
'Popularity'}).reset_index()

df_merged = pd.merge(df, df_popularity, on='Model', how='left')

plt.figure(figsize=(10, 6))

sns.scatterplot(data=df_merged, x='Base MSRP', y='Popularity', hue='Make', style='Make')

plt.title('Base MSRP vs. Popularity of EV Models') plt.xlabel('Base MSRP ($)')

plt.ylabel('Popularity (Number of Vehicles)') plt.show()

"""Part 4: Legislative Impact Analysis"""

Group data by 'Legislative District' and count the occurrences legislative_impact

= df['Legislative District'].value_counts()

Plotting the impact of legislative districts on EV adoption plt.figure(figsize=(14,

7))

```
sns.barplot(x=legislative_impact.index, y=legislative_impact.values, palette='deep')
plt.title('Impact of Legislative Districts on EV Adoption') plt.xlabel('Legislative
District') plt.ylabel('Number of Vehicles') plt.xticks(rotation=90) plt.show()

"""Clustering"""

from sklearn.cluster import KMeans #
Selecting features for clustering
X = df[['Electric Range', 'Base MSRP']]

# Applying KMeans clustering
kmeans = KMeans(n_clusters=3, random_state=0).fit(X) df['Cluster']
= kmeans.labels_

# Plotting clusters
plt.scatter(df['Electric Range'], df['Base MSRP'], c=df['Cluster'], cmap='viridis')
plt.title('Clustering of EVs based on Electric Range and MSRP')
plt.xlabel('Electric Range') plt.ylabel('Base MSRP') plt.show()
```