Analysing Electric Vehicles Performance and Future sells by Python

Abstract

The integration of electric vehicles (EVs) into our transportation landscape necessitates a thorough understanding of their performance, efficiency, and impact. In this pursuit, we conducted an electric car analysis using Python, employing its versatile tools for data manipulation, visualization, and machine learning. Through comprehensive data collection, preprocessing, and exploratory analysis, we unearthed insights into EV attributes, charging behaviors, and driving patterns. Leveraging Python's predictive modeling capabilities, we developed algorithms to forecast EV range and evaluate efficiency factors. The results reveal not only the technical prowess of EVs but also their potential to revolutionize sustainable transportation. By harnessing Python's prowess, this analysis illuminates a path toward informed decision-making, optimized EV usage, and the realization of a cleaner and smarter mobility ecosystem

Table of content

[Table of Content 3](#_Toc143900150)

[INTRODUCTION 6](#_Toc143900151)

[Need for study 8](#_Toc143900152)

[Aim and objectives 9](#_Toc143900153)

[Objectives: 9](#_Toc143900154)

[Scope of the Study 10](#_Toc143900155)

[Research Methodology 11](#_Toc143900156)

[Data Collection and Preparation: 11](#_Toc143900157)

[Python Setup and Libraries 11](#_Toc143900158)

[Exploratory Data Analysis (EDA): 11](#_Toc143900159)

[Data Visualization: 11](#_Toc143900160)

[Temporal Analysis: 11](#_Toc143900161)

[Battery Technology Insights 12](#_Toc143900162)

[Charging Infrastructure Mapping 12](#_Toc143900163)

[Policy Impact Visualization 12](#_Toc143900164)

[Insights and Interpretation 12](#_Toc143900165)

[Reporting and Presentation 12](#_Toc143900166)

[Data collection 13](#_Toc143900167)

[Review of literature 14](#_Toc143900168)

[Limitations of the electric cars 16](#_Toc143900169)

[Industry & Company Profile 17](#_Toc143900170)

[Industry Overview 17](#_Toc143900171)

[Key Segments: 17](#_Toc143900172)

[Trends and Challenges: 17](#_Toc143900173)

[Key Players: 18](#_Toc143900174)

[Regulation and Emissions: 18](#_Toc143900175)

[Conclusion: 18](#_Toc143900176)

[Company profile 19](#_Toc143900177)

[History of the company 19](#_Toc143900178)

[Key Focus Areas: 20](#_Toc143900179)

[Challenges and Criticisms: 22](#_Toc143900180)

[Conclusion: 22](#_Toc143900181)

[Conceptual framework/ Theoretical framework 23](#_Toc143900182)

[Meaning 23](#_Toc143900183)

[Definition 24](#_Toc143900184)

[Conceptual framework 24](#_Toc143900185)

[1. Data Collection and Preprocessing: 24](#_Toc143900186)

[2. Exploratory Data Analysis (EDA): 25](#_Toc143900187)

[3. Feature Engineering: 25](#_Toc143900188)

[4. Predictive Modeling: 25](#_Toc143900189)

[5. Model Training: 25](#_Toc143900190)

[6. Model Evaluation: 25](#_Toc143900191)

[7. Interpretability and Visualization: 25](#_Toc143900192)

[8. Time Series Analysis (Optional): 25](#_Toc143900193)

[9. Incorporating External Factors (Optional): 25](#_Toc143900194)

[10. Future Trends and Insights: 26](#_Toc143900195)

[11. Visualization and Reporting: 26](#_Toc143900196)

[12. Continuous Improvement: 26](#_Toc143900197)

[Data Description 26](#_Toc143900198)

[Result and Analysis 28](#_Toc143900199)

[Overview 28](#_Toc143900200)

[Library Import 28](#_Toc143900201)

[Basic information about the data 28](#_Toc143900202)

[Data Visualizations 30](#_Toc143900203)

[Question no. 1: whether the target column is balanced or not? 31](#_Toc143900204)

[Question no.2: Distribution of postal code whether it is normal distribution or not? 32](#_Toc143900205)

[Question no.4: Distribution of 2020 Census Tract whether it is normal distribution or not? 34](#_Toc143900206)

[Question no.5: Checking in each attribute any outlier exists or not? 35](#_Toc143900207)

[Question no.6: Checking correlations to reduce the dimension if possible? 36](#_Toc143900208)

[Question no.7: Checking correlations to reduce the dimension if possible? 37](#_Toc143900209)

[Question no.8: Comparison between Legislative District and Electric Vehicle Type? 38](#_Toc143900210)

[Question no.9: Finding which County has highest and lowest Base MSRP. 39](#_Toc143900211)

[Question no.10: County wise electric bike type demand. 40](#_Toc143900212)

[Question no.11: Finding which company produce more electric vehicle? 41](#_Toc143900213)

[Question no.12: Finding which type of electric vehicle type are more in CAFV? 42](#_Toc143900214)

[Question no.12: Multivariate plot to find the distribution of all attributes. 43](#_Toc143900215)

[Conclusion 44](#_Toc143900216)

[Reference 45](#_Toc143900217)

# INTRODUCTION

Over the course of the past three decades, the automotive industry has undergone a remarkable evolution, primarily steered by the emergence and progression of electric vehicles (EVs). Amidst growing concerns related to climate change, energy security, and deteriorating urban air quality, electric vehicles have emerged as a promising solution, poised to transform personal transportation and mitigate the environmental impact of the automotive sector. This extensive data analysis project embarks on a thorough exploration of the EV landscape spanning from 1997 to 2024, employing the potent tools of data visualization to uncover trends, patterns, and shifts that have guided the trajectory of this dynamic industry. The automotive domain has experienced a profound upheaval with the rise of electric vehicles, necessitating a comprehensive understanding of their growth trajectory, technological advancements, and market integration. This data-centric analysis aims to deeply delve into the realm of electric vehicles over a span of almost thirty years, stretching from 1997 to 2024. By harnessing the capabilities of data visualization, this initiative seeks to unlock valuable insights from historical data. Through a meticulous analysis of these datasets, the objective is to unveil the intricate narrative of electric vehicles, tracing their journey from novelty to becoming formidable contenders in the pursuit of sustainable mobility. The cornerstone of any robust data analysis lies in the meticulous gathering and preparation of data. A comprehensive dataset encompassing electric vehicle sales, battery technology advancements, deployment of charging infrastructure, and governmental policies spanning the years from 1997 to 2024 has been meticulously compiled. Rigorous data cleaning and formatting procedures have been applied to ensure the dataset's reliability and consistency, thereby laying a solid foundation for accurate and meaningful analysis.

The preliminary phase of the project involved a thorough examination of the dataset to extract initial insights. Initial observations highlighted a rapid surge in electric vehicle sales, particularly evident around the mid-2010s, signifying a significant shift towards heightened adoption. Furthermore, a closer look was taken at the relationship between battery capacity and vehicle range, shedding light on the technological leaps that have propelled electric vehicles into the mainstream market. A fundamental goal of this analysis was to capture the evolutionary trajectory of electric vehicle market penetration over time. To achieve this, data visualization techniques were adeptly utilized to construct compelling line charts, effectively showcasing the consistent upward trajectory of EV sales. These visualizations provide a comprehensive overview of the growth pattern, uncovering pivotal milestones such as the introduction of mass-market electric vehicles and their subsequent acceptance. Undoubtedly, the evolution of battery technology has played a pivotal role in determining the viability of electric vehicles. Employing scatter plots and trend lines, this project effectively demonstrates the remarkable strides made in battery capacity, directly correlating with the enhanced range of vehicles. This visual representation underscores the pivotal role of innovation in alleviating concerns surrounding range anxiety and fostering consumer confidence in EVs.

As electric vehicle adoption depends significantly on charging infrastructure availability, this project extensively examines its impact. Utilizing heatmaps and geographical visualizations, the expansion of charging stations across different regions and periods is vividly portrayed. The analysis highlights the concentrated efforts aimed at establishing a robust charging network, ultimately augmenting the practicability of electric vehicle ownership. In shaping the landscape of electric vehicles, governmental policies and incentives hold immense significance. To elucidate this aspect, stacked area charts were employed to visually represent the influence of diverse policies, tax incentives, and subsidies on EV sales. These visualizations not only underscore the correlation between policy interventions and spikes in sales but also emphasize the necessity of sustained governmental support to sustain momentum. Traversing the electric vehicle landscape from 1997 to 2024, illuminated through the lens of data visualization, reveals a captivating narrative of transformation. By amalgamating historical sales data, technological advancements, infrastructure expansion, and policy dynamics, a comprehensive picture of an industry poised on the brink of a revolution emerges. Electric vehicles, once a niche concept, have transcended into symbols of sustainable transportation, primed to redefine the future of mobility. Through the lens of data visualization, this data science project underscores the capacity to decipher intricate trends, ultimately providing stakeholders with indispensable insights to navigate a path toward a cleaner and more sustainable automotive future.

## Need for study

The study titled "Electric Vehicle Data (1997-2024) Analysis through Data Visualization: A Journey into Sustainable Mobility" is a critical endeavor driven by the imperative to unravel the multifaceted evolution of electric vehicles (EVs) within the automotive landscape across nearly three decades. Amidst intensifying concerns surrounding environmental degradation, climate change, and energy security, the emergence of EVs presents a promising avenue toward sustainable mobility. This study, through meticulous data analysis and visualization techniques, aims to decode the intricate dynamics shaping the EV industry's journey. By scrutinizing historical sales data, tracking technological advancements in battery technology and charging infrastructure, and assessing the impact of government policies and market trends, this study seeks to illuminate the underlying patterns, trends, and transformations that have steered the course of the EV revolution. The insights garnered from this analysis are not only indispensable for policymakers, industry stakeholders, and manufacturers but also hold the potential to guide consumers and society at large toward an environmentally conscious and forward-looking automotive paradigm. In essence, this study serves as a compass, navigating the realm of sustainable mobility through the data-driven exploration of EVs, encapsulating their profound significance in shaping a more ecologically harmonious future.

## Aim and objectives

**Aim:**

This study aims to comprehensively analyze electric vehicle (EV) data spanning the years 1997 to 2024 using data visualization techniques, with the overarching goal of unraveling the transformative journey of EVs within the automotive industry and their contribution to sustainable mobility.

## Objectives:

* To chronicle the evolution of EVs by analyzing sales data, technological advancements, and market dynamics over the specified period.
* To employ data visualization tools to depict trends and patterns in EV adoption, battery technology advancements, charging infrastructure growth, and market share shifts.
* To assess the impact of governmental policies and incentives on EV sales and market penetration, offering insights into the role of regulatory frameworks in shaping the industry.
* To extrapolate from historical data and trends, providing informed projections for the future growth and influence of EVs in the context of sustainable mobility.
* To provide stakeholders, including policymakers, manufacturers, investors, and consumers, with data-driven insights that facilitate informed decision-making, contributing to a more sustainable and environmentally conscious automotive landscape.

## Scope of the Study

encompasses a comprehensive exploration of the electric vehicle (EV) landscape within the automotive industry over a span of almost three decades. The scope of the study includes:

* The study delves into historical EV data from 1997 to 2024, examining the evolution of electric vehicles from their early inception to their current prominence, providing insights into the industry's transformative journey.
* The study utilizes advanced data visualization techniques to effectively present and analyze trends, patterns, and shifts in electric vehicle sales, battery technology advancements, charging infrastructure growth, and market dynamics over time.
* The scope includes evaluating the progress in battery technology and its impact on EVs, including improvements in battery capacity, range, and charging capabilities, elucidating the pivotal role of innovation in driving EV adoption.
* The study assesses the expansion of charging infrastructure networks across different regions and its correlation with the growth of electric vehicle adoption, offering insights into the practical feasibility of EV ownership.
* The study examines the impact of governmental policies, incentives, and regulations on electric vehicle sales and market penetration, providing a comprehensive view of the interplay between policy frameworks and industry growth.
* The scope encompasses the analysis of market share trends between electric vehicles and conventional internal combustion engine vehicles, shedding light on the changing landscape of the automotive industry.
* The study extrapolates insights from historical data to offer informed projections regarding the future trajectory of electric vehicles, contributing to a forward-looking understanding of their potential role in sustainable mobility.
* The study's insights aim to guide decision-making for various stakeholders, including policymakers, automotive manufacturers, investors, and consumers, by providing valuable information on industry trends and implications.

## Research Methodology

The research methodology for analyzing electric vehicle (EV) data from 1997 to 2024 using data visualization involves a systematic approach that integrates Python programming with various data visualization libraries to extract insights and trends. The primary goal is to uncover the evolution of the EV industry, technological advancements, policy impacts, and their role in fostering sustainable mobility. The following steps outline the research methodology:

### Data Collection and Preparation:

* Gather comprehensive datasets encompassing EV sales, battery technology, charging infrastructure, and policy data for the years 1997 to 2024.
* Preprocess and clean the data to ensure consistency and accuracy. Handle missing values, outliers, and format inconsistencies.

### Python Setup and Libraries

* Set up a Python environment with necessary libraries such as Pandas for data manipulation, Matplotlib, and Seaborn for data visualization, and NumPy for numerical computations.

### Exploratory Data Analysis (EDA):

* Conduct initial data exploration using Python to gain insights into the datasets.
* Generate summary statistics, histograms, and correlation matrices to understand the data distribution and relationships.

### Data Visualization:

* Utilize Matplotlib and Seaborn to create a variety of visualizations, such as line charts, bar charts, scatter plots, heatmaps, and stacked area charts.
* Visualize EV sales trends over time, market share comparisons, battery capacity improvements, charging infrastructure growth, and policy impact.

### Temporal Analysis:

* Create line charts to illustrate the temporal growth of EV sales, highlighting key milestones and inflection points.
* Overlay market share data to analyze the shift between EVs and traditional vehicles over time.

### Battery Technology Insights

* + Generate scatter plots with trend lines to showcase the correlation between battery capacity and vehicle range.
  + Highlight notable advancements in battery technology by visualizing the increase in capacity over the years.

### Charging Infrastructure Mapping

* + Utilize geographical visualization libraries (e.g., Plotly) to plot charging stations on maps.
  + Create heatmaps to depict the distribution of charging stations across different regions and observe the growth trends.

### Policy Impact Visualization

* + Employ stacked area charts to visualize the impact of various policies and incentives on EV sales.
  + Show how policy interventions have influenced adoption spikes and sustained growth.

### Insights and Interpretation

* + Analyze the generated visualizations to extract meaningful insights.
  + Conclude the evolution of the EV industry, the role of policies, and the potential for sustainable mobility.

### Reporting and Presentation

* Summarize findings in a comprehensive report or presentation.
* Use the visualizations to effectively communicate the research outcomes to various stakeholders.

## Data collection

Electric Vehicle Data 1997-2024 from the website Kaggle is a valuable archive of secondary data that summarizes the multifaceted evolution of electric vehicles. This data has 134474, 17 rows and columns. This multi-source and carefully curated dataset provide a comprehensive overview of electric vehicle sales, technology developments, charging infrastructure growth, and policy implications over nearly three decades. Using this secondary data, the analysis aims to reveal hidden trends, patterns, and changes in the electric vehicle industry, shedding light on its transformation from a niche innovation to a formidable challenge in the pursuit of sustainable mobility.

## Review of literature

According to Ullah et al., 2023 This study applies a new interpretive machine learning (ML) framework to predict electric vehicle charging choice behavior. The test was based on actual normal and rapid charging of 500 electric cars in Japan for two years. The results showed that the XGBoost model achieved the highest accuracy in predicting charging station selection behavior compared to other ML classifiers. In addition, this study used the newly developed SHAP approach to identify the importance of features and the complex nonlinear and interactive effects of different features on charging station choice behavior. This study suggests that combining ML models with SHAP can develop an interpretable ML model to predict EV charging station choice behavior. According to the vision of Zahed et al., 2022 This work evaluates the Holy in Canada. Energy supply/demand efficiency of a residential cluster in a village located in the Albert climate region. First, all the buildings in the community are modeled and the energy needed to meet demand and electric vehicles is calculated. The potential of each building to provide electricity through solar panels is then calculated. Finally, the municipality's energy supply/demand management is evaluated using a machine learning tool. The results show that the heating and cooling of the community are 420 kW and 121 kW respectively, and the total annual energy production of the solar energy system is 14,203 kWh per individual house. In terms of EV load, the solar energy system can generate 29.23% of the total community load per year. Finally, comparing the modeled pattern and the predicted pattern, the prediction accuracy is 88.6%. Ragonen et al. According to the State of Charge (SOC) estimate of 2021, the battery management system of lithium-ion battery (LIB) and #40; BMSand#41; in battery electric vehicle (BEV) applications. In this work, we propose a modeling framework for SOC estimation using different machine learning (ML) methods, namely support vector regressor (SVR), artificial neural network (ANN), and long-term memory (LSTM) networks. The required training data is developed using Matlab/Simulink auto simulations of the BEV integrated with the electrochemical COMSOL Multiphysics model of the LIBs. The developed Multiphysics operating model of BEVs and LIBs allows to study the influence of driving conditions on the electrochemical and degradation processes (i.e. formation and degradation of the solid electrolyte intermediate phase - SEI -) occurring in different chemical batteries that have been deployed. Tesla S. and Nissan Leaf BEVs.

Secinaro et al., 2020 Automotive Business Model Architectures are receiving increasing attention from researchers and decision-makers. Although research has been conducted to address the pressures of future business model change, no studies have examined bibliometric variables in this area. This study aims to fill the gap through a bibliometric analysis of 104 articles on electric vehicle business models. The analysis showed that the literature on EV business models is extensive and focuses on charging technologies, driver services, electricity management, commercial contracts and service business models. China, the United States and Germany have done maximum research on the above-mentioned subject. The topic dendrogram identified two emerging threads of discussion: innovative technologies and resource optimization and electrical management systems and product life cycles. These findings could guide environmental policies for electric car production and help automakers reform their models.

Duraisamy, T. and Kaliyaperumal, D., 2021 Cell balancing is a key feature of a battery management system and#40; BMSand#41; an important feature implemented to extend battery life and service life. Due to the increasing demand for bigger and better batteries, there is a focus on different cell-balancing techniques. The passive balancing method is the most popular due to its low cost and ease of use. Since the balanced resistors dissipate balanced energy as heat, a proper thermal diagram of the balanced system is necessary to keep the BMS board temperature within tolerable limits. This paper proposes optimal balancing resistor selection according to cell unbalance degree, balancing time, C index, and temperature rise using a machine learning (ML) based balancing control algorithm to improve balancing time and optimal power loss control. The passive balanced system uses variable resistors to optimize power dissipation and achieve optimal thermal characterization. The performance of the proposed system is evaluated using Back Propagation Neural Network (BPNN), Radial Basis Neural Network (RBNN), and Long Short Term Memory (LSTM). To optimize the balance parameters, an error analysis of the balance system is carried out and the proposed algorithms are compared using performance indices such as root mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) to validate the balance. model presentation In the Matlab-Simscape environment, a possible optimization area for the implementation of passive balance with machine learning algorithms is tested.

## Limitations of the electric cars

1. Data Quality: The availability of accurate and comprehensive data on electric cars can be limited, leading to potential inaccuracies in the analysis.
2. Data Bias: Analysis results might be skewed due to biased data sources that focus on specific manufacturers, models, or regions.
3. Limited Historical Data: Electric vehicle trends and comparisons with conventional cars may be constrained by the relatively short history of electric car adoption.
4. Charging Infrastructure: Variations in charging options and availability across regions can impact usage patterns and analysis outcomes.
5. Battery Complexity: Modeling battery performance accurately requires considering intricate factors like temperature, charging habits, and degradation, adding complexity to the analysis.
6. Model Diversity: Diverse electric car models with varying features and efficiencies might lead to oversimplified generalizations in the analysis.
7. External Influences: External factors such as policy changes and technological advancements may not be fully integrated into the analysis, affecting its comprehensiveness.

# Industry & Company Profile

## Industry Overview

The automotive sector plays a crucial role in the global economy, contributing significantly to employment, technological advancements, and international trade. It is characterized by rapid innovation, evolving consumer preferences, and increasing emphasis on sustainability and safety.

## Key Segments:

1. Automobile Manufacturers: These companies design, engineer, and manufacture vehicles, ranging from passenger cars and trucks to commercial vehicles.
2. Auto Parts and Components: This segment focuses on producing various components, systems, and parts used in vehicle manufacturing, including engines, transmissions, electronics, and more.
3. Aftermarket Services: Aftermarket businesses provide maintenance, repair, and replacement parts for vehicles after they are sold. This includes independent repair shops, parts retailers, and maintenance service providers.
4. Electric and Autonomous Vehicles: The industry is undergoing a transformation with a growing emphasis on electric vehicles (EVs) and autonomous vehicles (AVs). Companies are investing in EV technology, battery development, and AV software.
5. Global Markets: Major automotive markets include North America, Europe, Asia-Pacific, and emerging markets in Latin America, Africa, and the Middle East. Each market has unique consumer preferences, regulations, and competitive dynamics.

## Trends and Challenges:

* Electric Revolution: The shift towards electric vehicles is driven by environmental concerns and government regulations. Automakers are investing heavily in EV development and charging infrastructure.
* Autonomous Driving: Companies are working on self-driving technologies that promise safer and more efficient transportation. However, regulatory, ethical, and technological challenges remain.
* Connectivity and IoT: Vehicles are becoming more connected, enabling features like infotainment, real-time diagnostics, and remote control. This trend is creating opportunities for tech companies and cybersecurity challenges.
* Sustainability: The industry is addressing environmental concerns through initiatives like fuel efficiency improvements, lightweight materials, and recycling programs.
* Shared Mobility: Ride-sharing services and car-sharing platforms are changing the way people use and own vehicles, impacting traditional car ownership models.
* Supply Chain Disruptions: Events like semiconductor shortages and trade disputes highlight the sector's vulnerability to supply chain disruptions.



## Key Players:

Prominent automotive companies include Toyota, Volkswagen Group, General Motors, Ford, BMW, Honda, and many more. Additionally, newer entrants like Tesla are driving innovation in electric and autonomous vehicles.

## Regulation and Emissions:

Stringent emissions standards are pushing manufacturers to reduce their carbon footprint. Governments around the world are implementing regulations to promote cleaner and more efficient vehicles.

## Conclusion:

The global automotive sector is undergoing significant transformation due to technological advancements, changing consumer behaviors, and environmental concerns. This presents both challenges and opportunities for companies across the industry, driving them to innovate and adapt to a rapidly evolving landscape.

## Company profile

Tesla, Inc. is an American electric vehicle (EV) and clean energy company founded in 2003 by Martin Eberhard and Marc Tarpenning. Elon Musk, JB Straubel, and Ian Wright joined the company shortly after its founding. Tesla is renowned for its innovation in electric vehicle technology, sustainable energy solutions, and commitment to accelerating the transition to sustainable transportation.

## History of the company

* 2003: Tesla Motors (now Tesla, Inc.) is founded in Palo Alto, California, by Martin Eberhard and Marc Tarpenning, with the goal of producing electric vehicles with high performance and long electric range.
* 2004: Elon Musk becomes the chairman of Tesla's board of directors and leads the company's initial round of investment funding.
* 2006: Tesla unveils its first electric car prototype, the Tesla Roadster, based on the Lotus Elise chassis. The Roadster became the first highway-legal all-electric vehicle to use lithium-ion battery cells.
* 2008: Tesla delivers the first Tesla Roadster to customers. The Roadster gains attention for its acceleration, range, and sleek design.
* 2010: Tesla goes public with its initial public offering (IPO) on the NASDAQ stock exchange under the symbol "TSLA." This makes Tesla the first American car company to go public since Ford in 1956.
* 2012: Tesla introduces the Model S, a luxury all-electric sedan. The Model S receives widespread acclaim for its performance, range, and innovative features.
* 2013: Tesla announces its Supercharger network, a system of fast-charging stations designed to provide long-distance travel capability for Tesla vehicles.
* 2015: The company unveils the Model X, an electric SUV known for its distinctive falcon-wing doors and advanced safety features.
* 2016: Tesla announces its intention to acquire SolarCity, a solar energy services company, as part of its vision to create a vertically integrated sustainable energy company.
* 2017: The highly anticipated Tesla Model 3, a more affordable electric sedan, begins production. The Model 3's launch attracts significant attention and demand.
* 2018: Tesla introduces the Model Y, a compact electric SUV, expanding its vehicle lineup further.
* 2020: Tesla becomes the most valuable automaker by market capitalization, surpassing long-established companies like Toyota and Volkswagen.
* 2021: Tesla continues to deliver new vehicle models, including updates to existing ones. The company's Autopilot and Full Self-Driving features remain a topic of discussion and debate regarding their safety and regulatory implications.
* 2021 (Ongoing): Tesla's Gigafactories continue to be built around the world to increase production capacity for both vehicles and batteries. The company's impact on the electric vehicle market and the broader automotive industry remains significant



## Key Focus Areas:

Electric Vehicles (EVs):

Tesla is primarily known for its electric cars, which include various models such as the Model S (luxury sedan), Model 3 (mass-market sedan), Model X (SUV), and Model Y (compact SUV). These vehicles have gained popularity for their impressive performance, long-range capabilities, and advanced software features.

Autonomous Driving:

Tesla has been at the forefront of autonomous driving technology. Its vehicles are equipped with advanced driver-assistance features through the "Autopilot" system. Tesla's long-term vision involves achieving full self-driving capability, although this goal has faced regulatory and technological challenges.

Energy Storage and Solar Solutions:

Beyond vehicles, Tesla is engaged in providing energy storage solutions through products like the Powerwall (home battery) and Powerpack (commercial and utility-scale battery). The company also acquired SolarCity to integrate solar energy generation into its clean energy ecosystem.

Charging Infrastructure:

Tesla has developed a global network of Supercharger stations that provide fast charging for its vehicles, reducing charging times and increasing the convenience of long-distance travel for EV owners.

Electric Range and Performance:

Tesla's EVs are known for their impressive electric range, often outperforming competitors. The Model S, for instance, set records for being one of the first EVs to achieve over 300 miles of range on a single charge.

Over-the-Air (OTA) Updates:

Tesla has demonstrated the ability to remotely update its vehicles' software, introducing new features, improvements, and even performance upgrades through OTA updates.

Gigafactories:

Tesla's Gigafactories are large-scale manufacturing facilities designed to produce batteries, electric drivetrains, and complete vehicles. These factories are strategically located around the world to meet increasing demand.

Market Capitalization and Impact:

Tesla's market capitalization has surged significantly, making it one of the most valuable automakers globally. Its success has influenced other automakers to accelerate their EV plans and invest in electrification.

## Challenges and Criticisms:

* Production Challenges: Tesla has faced production challenges, including delays in meeting production targets and quality control issues.
* Autonomous Driving Concerns: The safety and regulatory implications of Tesla's Autopilot and Full Self-Driving features have raised debates about the readiness of self-driving technology for public roads.
* Financial Sustainability: Despite its market capitalization, Tesla has faced periods of financial uncertainty and cash flow challenges in the past.

## Conclusion:

Tesla has fundamentally reshaped the automotive industry with its electric vehicles, energy storage solutions, and advancements in autonomy. The company's ambitious goals, commitment to sustainability, and innovative approach have positioned it as a leading force in the transition to a more sustainable transportation and energy landscape.

# Conceptual framework/ Theoretical framework

## Meaning

Analyzing electric cars with Python requires applying the power of programming and computing to the complex world of electric vehicles (EV). This multifaceted process involves the systematic collection, cleaning, exploration, visualization and interpretation of EV-related data using Python's rich libraries and tools. Diving into this analysis, the aim is to open a deep understanding of the different dimensions of electric cars, including their technical performance, energy efficiency, charging dynamics and ecological relevance. Using state-of-the-art machine learning algorithms, statistical techniques and data visualization techniques, this analysis aims to unravel the complex patterns embedded in EV data streams. Python acts as an organizer that transforms raw data into actionable intelligence, whether it's estimating battery capacity, evaluating driving habits or figuring out the interaction between environmental factors and the efficiency of electric cars. In the middle of this analytical journey, Python data processing libraries like pandas make it easy to combine and clean raw data by forming it into a coherent structure. Libraries such as matplotlib and seaborn, which visualize data distribution, trends and correlations, contribute to a comprehensive understanding of electric cars. In addition, Python's scikit-learn library, a set of machine learning algorithms, shows the potential to predict the performance of electric vehicles - whether it predicts mileage, predicts charging times or simulates energy consumption in different scenarios. Each algorithm acts as a mathematical engine that decodes the complex relationships embedded in the EV data. But this analysis is not limited to mathematics and algorithms. This is research that examines real-world consequences. As algorithms refine their predictions, insights extrapolated from data models can shape EV usage practices. This includes creating optimal charging schedules that reduce the load on the electricity grid during peak hours and offering ecological driving strategies that expand the range of electric vehicles. These knowledge-based insights spread in society and influence both individual drivers and political decision-makers and direct the development direction towards a more sustainable transport future. Basically, electric car analysis with Python is a complete journey that includes data dissection, statistical exploration, algorithmic prediction and actionable interpretation. It's an exploration that begins with raw data and culminates in illuminating revelations, all driven by a symphony of Python code. This is not just an analysis of numbers, but an exploration of the synergy between technology, environment and human behavior, harmonized by Python's digital prowess. Ultimately, this analytical effort leads us to an era where electric vehicles are more than just cars; they become intricate pieces in a shifting puzzle that envisions a cleaner, greener and more sustainable future.

## Definition

Electric Vehicle Analysis with Python refers to an in-depth and data-driven study of various aspects related to Electric Vehicles (EVs) using the Python programming language and related libraries. This complex process involves the systematic collection, organization, processing, and interpretation of data on electric cars, including their technical data, performance measurements, charging behavior and environmental impact. Using the synergy of Python's programming capabilities and its specialized tools for data processing, visualization and machine learning, this analysis aims to uncover complex patterns, correlations and insights embedded in the data shedding light on the complex dynamics of electric vehicles. From predicting driving distance and energy consumption to interpreting the effects of external factors such as weather and terrain, the aim is to gain practical insights that can inform electric vehicle users, manufacturers, and decision makers. Using the versatility of Python, electric car analysis goes beyond numbers and algorithms and becomes a conduit for informed decision-making and sustainable traffic solutions in an increasingly electrified automotive world.

## Conceptual framework

A conceptual framework for analyzing electric cars using Python involves several key components and steps. Here's a high-level outline of the framework:

### 1. Data Collection and Preprocessing:

* Gather data on electric cars, including specifications, performance metrics, charging data, and environmental factors.
* Clean and preprocess the data to handle missing values, outliers, and inconsistencies.

### 2. Exploratory Data Analysis (EDA):

* Visualize data to understand distributions, correlations, and trends.
* Identify patterns and insights related to electric car attributes, usage patterns, and performance.

### 3. Feature Engineering:

Create relevant features from the collected data, such as battery capacity, charging times, efficiency metrics, and driving conditions.

### 4. Predictive Modeling:

* Choose a suitable predictive task, such as range prediction, charging time estimation, or efficiency analysis.
* Split the data into training and testing sets.
* Select appropriate machine learning algorithms (regression, classification, etc.) for the chosen task.

### 5. Model Training:

* Train the selected machine learning models using the training dataset.
* Tune hyperparameters to optimize model performance.

### 6. Model Evaluation:

* Evaluate model performance on the testing dataset using appropriate evaluation metrics (e.g., Mean Absolute Error for regression tasks).
* Compare different models to identify the best-performing one.

### 7. Interpretability and Visualization:

* Interpret model results to understand the factors influencing predictions.
* Visualize predictions, residuals, and model insights for better comprehension.

### 8. Time Series Analysis (Optional):

If analyzing temporal data (e.g., charging patterns), apply time series analysis techniques to uncover trends and seasonality.

### 9. Incorporating External Factors (Optional):

Integrate external data sources, such as weather data or charging station locations, to enhance analysis accuracy.

### 10. Future Trends and Insights:

* Analyze the implications of the analysis results for the electric vehicle market, charging infrastructure, and consumer behavior.
* Provide recommendations for improving electric car performance, optimizing charging strategies, or addressing limitations.

### 11. Visualization and Reporting:

* Create clear and informative visualizations and reports to communicate findings effectively.
* Present insights to stakeholders or decision-makers.

### 12. Continuous Improvement:

Continuously refine and update the analysis framework as new data becomes available or as the electric vehicle landscape evolves.

Throughout each step, Python libraries such as pandas for data manipulation, matplotlib and seaborn for visualization, sci-kit-learn for machine learning, and potentially time series libraries like statsmodels can be utilized. The framework should be adapted and expanded based on the specific analysis goals and data availability.

## Data Description

Here the author collected the dataset from an open-source Kaggle because it is very difficult to find or collect a primary dataset. That’s why the author chose a secondary dataset. The dataset link is “https://www.kaggle.com/datasets/iottech/electric-vehicle-data-1997-2024-update-version”. In this dataset, there are 17 attributes present in the dataset and 134474 rows exist in the dataset. All the data attributes represent the electric vehicle data. Among them few are the most important attributes that represent or give some good information about an electric vehicle, those are:

* **Model year** – In which year the car is launched it is represented by this column.
* **Make** – the car made by which company is represented by this attribute.
* **Electric vehicle type** - Battery- or plug-in-hybrid-powered electric vehicles are two different types of electric vehicles.
* **CAFV** - Refers to vehicles that have to meet specified criteria in order to participate in programs, get incentives, or have advantages related to using clean or alternative fuels.
* **Electric Range** - The term "electric range" refers to the maximum distance that an electric vehicle (EV) may go before its battery needs to be recharged. It represents the distance an electric vehicle (EV) can travel in a single trip using solely the electricity stored in its battery pack.
* **Base MSRP** - It speaks about an automobile model's manufacturer-suggested retail price (MSRP), which excludes any options, upgrades, or accessories.
* **(DOL)The Department of Licensing** - The DOL is responsible for managing the issuance of driver's licenses, identification cards, and vehicle registrations.
* **Electric Utility** - Electric utilities are required to provide the infrastructure, such as charging stations, grid connections, and power distribution networks, needed to supply electricity to EV owners.

# Result and Analysis

## Overview

The result plays an important role in any research activity because after applying different techniques and effort the obtained results it is a crucial part of the research or study work. That’s why in this chapter the author explains different results obtained by this study work.

## Library Import

Here, after connecting Google Drive to Colab Notebook the author gets access to loading the data from the drive. To load the data the author required some library.

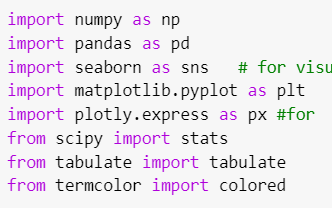


Figure 1.1 Importing necessary libraries

Here in Fig. 1.1, it can be seen that the author imported some of the necessary libraries such as NumPy to deal with matrix and numeric calculations, pandas to deal with some basic visualizations, seaborn, matplotlib, and plotly express used to deal with some basic and advance visualizations to represents the data.

## Basic information about the data

Here basic information represents the shape of the data, whether is there any null value or not, how many unique values are there in each attribute, and many more. All these things are shown in this section in the visual form of images.

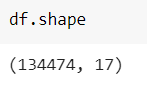


Figure 1.2 Shape of the dataset

Fig. 1.2 represents the shape of the dataset and it can be seen that the dataset contains 134474 rows and 17 columns. The volume of the dataset is very good for creating or analyzing the data.

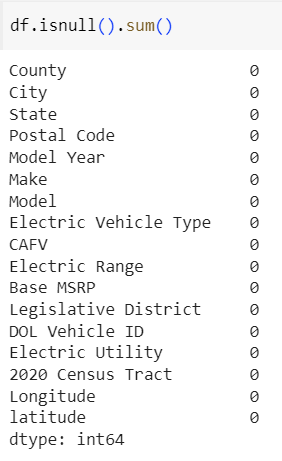


Figure 1.3 Representing null values existence

Checking null values’ existence is important for any ML project because if any null values exist in the data at that time model can’t manipulate or handle the data to predict the feature values. That’s why before creating or processing the model it is necessary to check null values. If there are any null values exist then either drop or treat them with the appropriate techniques. Fig. 1.3 represents that there are no null values in the dataset so there is no requirement for any techniques to treat them which means the data is good quality.

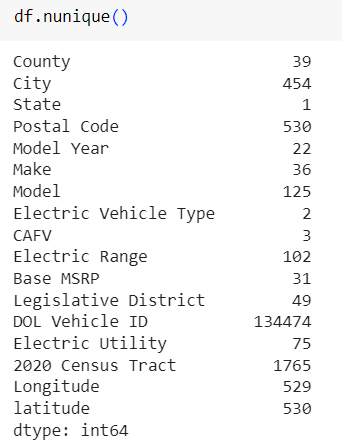


Figure 1.4 Showing unique values in each attribute

It can be seen in Fig. 1.4, there are so many unique values present in each attribute. Here in Fig. 1.4, it can be seen that the Electric Vehicle Type has 2 unique values which one considered the target column for this problem statement.

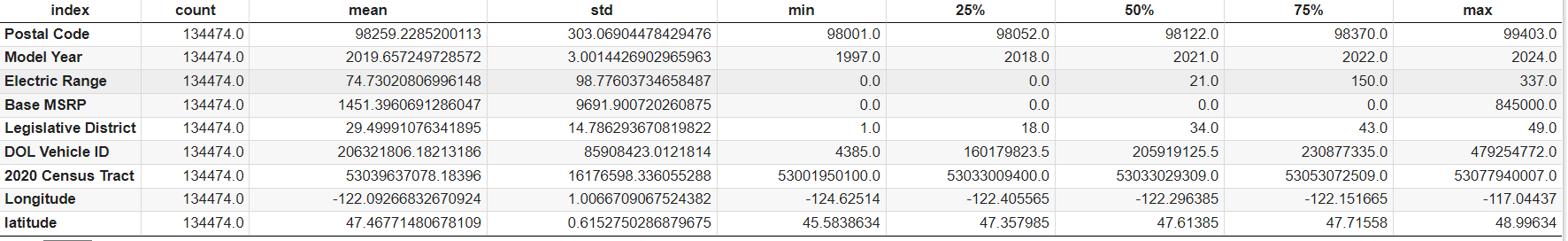


Figure 1.5 Descriptive statistics

Descriptive statistics gives different statistical information such as mean, median, mode, standard deviation, and percentile. Fig. 1.5 represents descriptive statistics of all numerical values.

## Data Visualizations

Data visualizations are essential tools for unlocking the insights hidden within complex datasets. They provide a concise and intuitive way to present information, enabling individuals to grasp patterns, trends, and relationships that might otherwise go unnoticed in raw data. By transforming data into visual formats such as graphs, charts, and maps, data visualizations enhance comprehension, aid in decision-making, and facilitate effective communication. They bridge the gap between data analysis and human cognition, allowing analysts, researchers, and decision-makers to quickly extract meaningful conclusions, identify outliers, and convey findings to both technical and non-technical audiences. Ultimately, data visualizations play a pivotal role in making data-driven insights more accessible, actionable, and impactful. Below the author provides all the visualizations to represent the data information.

#### Question no. 1: whether the target column is balanced or not?

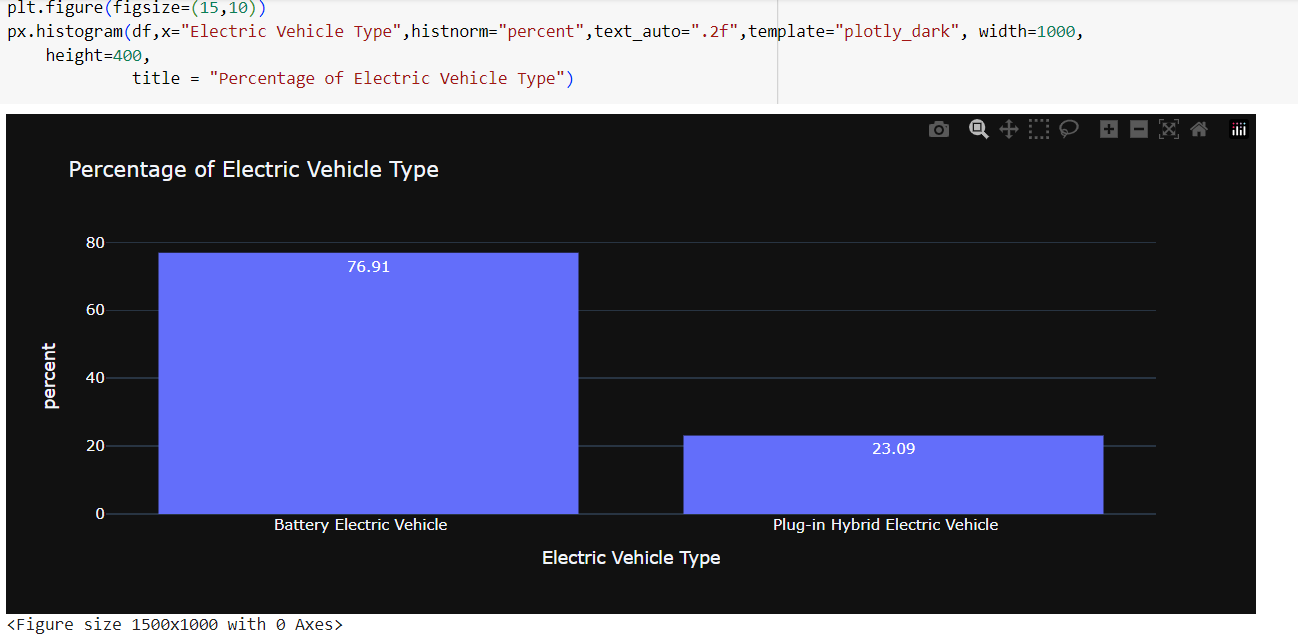


Figure 1.6 Target column distribution in percentage

**Interpretations:**

The Fig. 1.6 and Fig. 1.7 represents that the target column means the electrical vehicle type is not balanced because the Battery’s electrical is present at almost 76.91 and plug I hybrid vehicle presents only 23.09 percent.



Figure 1.7 Target column distribution in total count

**Interpretations:**

Fig. 1.7 represents the imbalanced of target column which is distributed in high imbalanced. One is almost 103424 times and anther is 31050.

#### Question no.2: Distribution of postal code whether it is normal distribution or not?

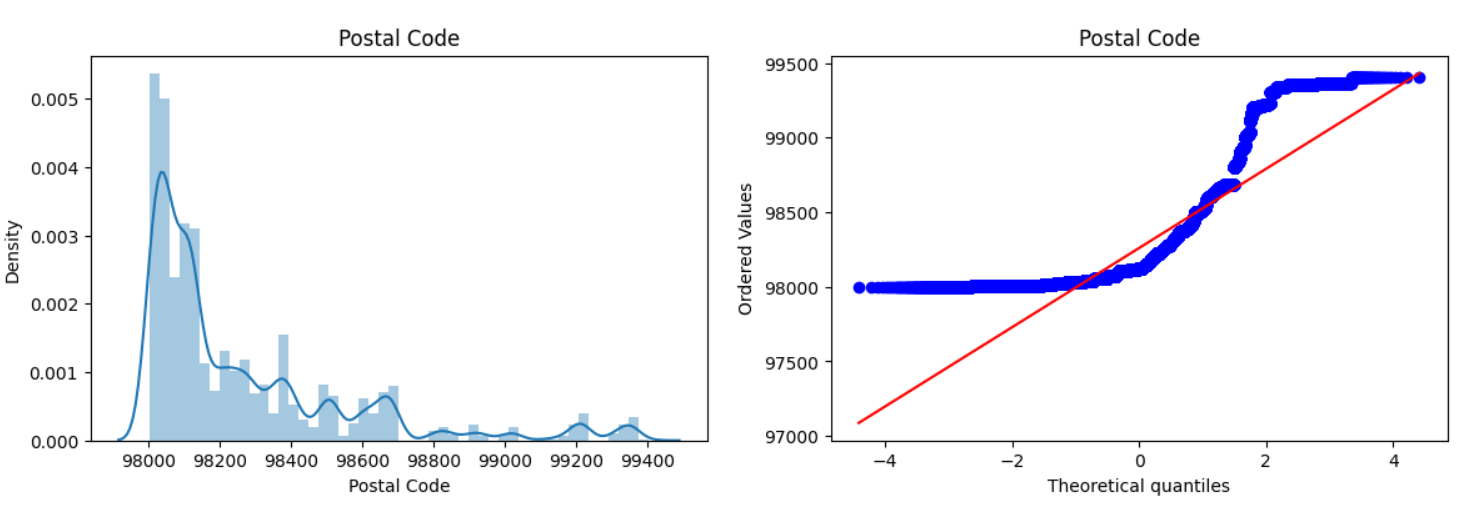


Figure 1.8 Distribution of postal code

**Interpretations:**

The data is not normally distributed and it can be seen in Fig. 1.8 because all the data does not come or falls down in the red line. Question no.3: Distribution of Model Year whether it is normal distribution or not?

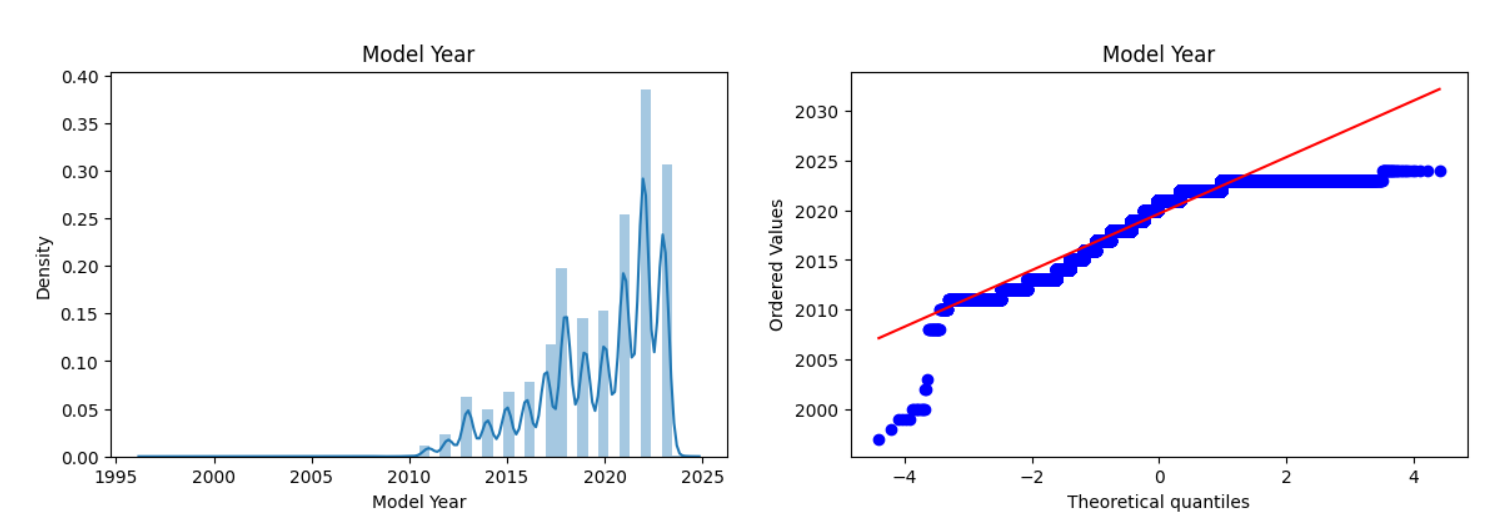


Figure 1.9 Distribution of Model Year

**Interpretations:**

Model Year is quite balanced but does not properly follow normal distribution and it can be seen in Fig. 1.9.

#### Question no.4: Distribution of 2020 Census Tract whether it is normal distribution or not?

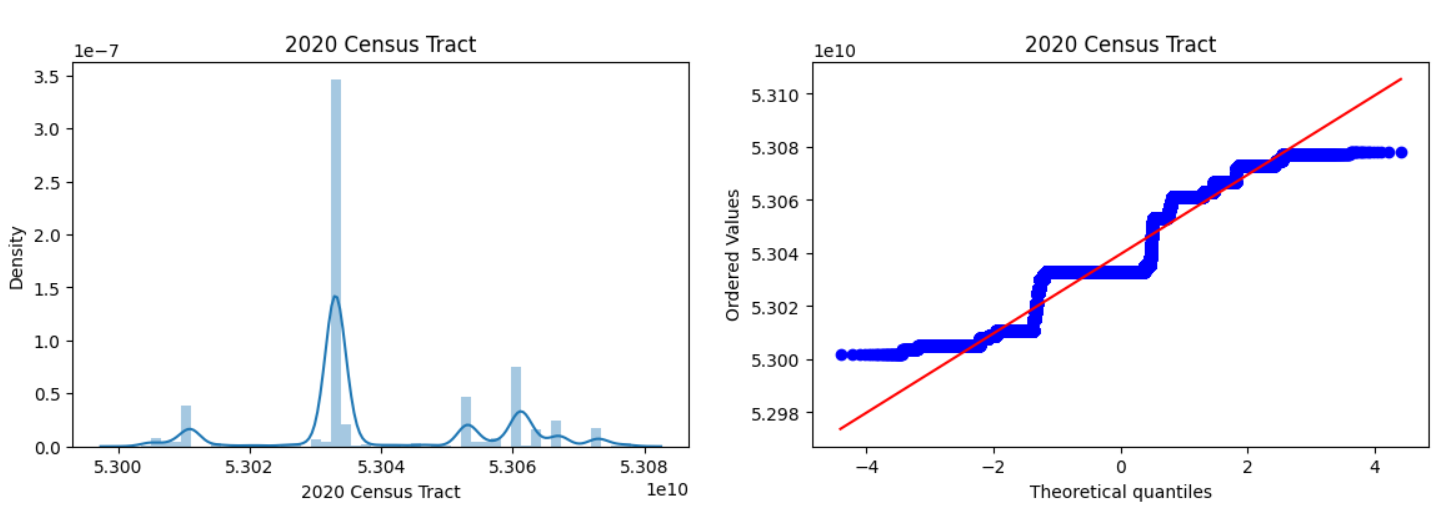


Figure 1.10 Distribution of Model Year

**Interpretations:**

Fig. 1.9 shows that the model year almost follows the normal distributions which means the distribution of the data is well to create the model.

#### Question no.5: Checking in each attribute any outlier exists or not?

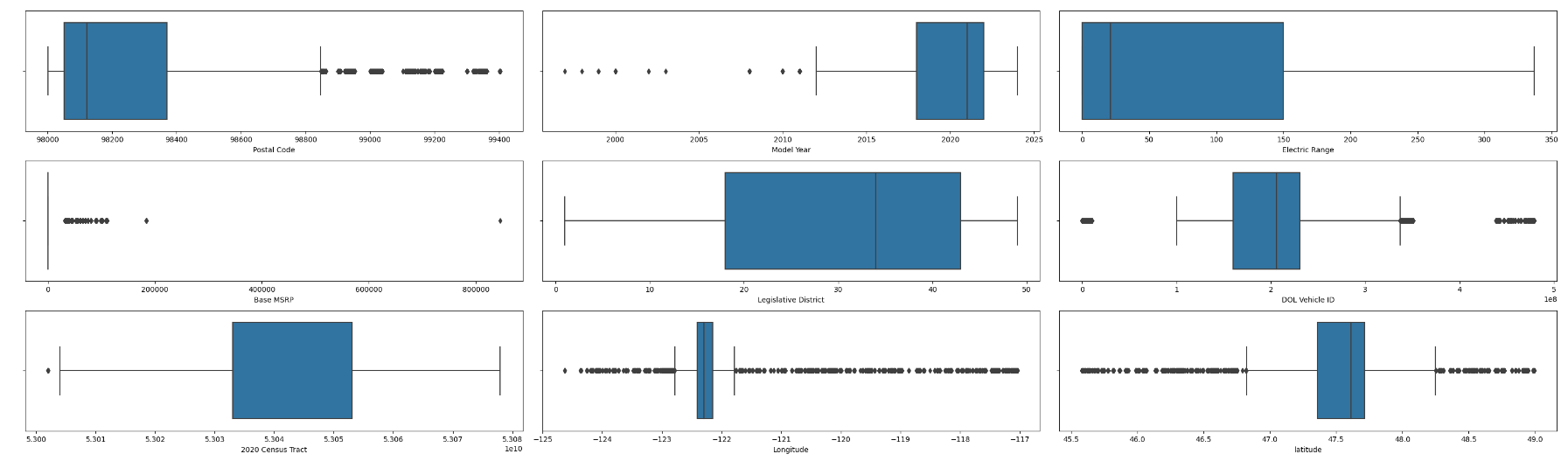


Figure 1.11 Showing outliers of the data

**Interpretations:**

Outliers are something called extreme values means in a group of data if any data is extreme or dissimilar from other values then it is called extreme values or outliers. It is necessary to identify outliers because outliers impact the model performance and take more computational time. That is why it is necessary to detect outliers and treat them with an appropriate approach. Fig. 1.11 represents that a few columns have no outliers and the remaining attribute has more outliers.

#### Question no.6: Checking correlations to reduce the dimension if possible?

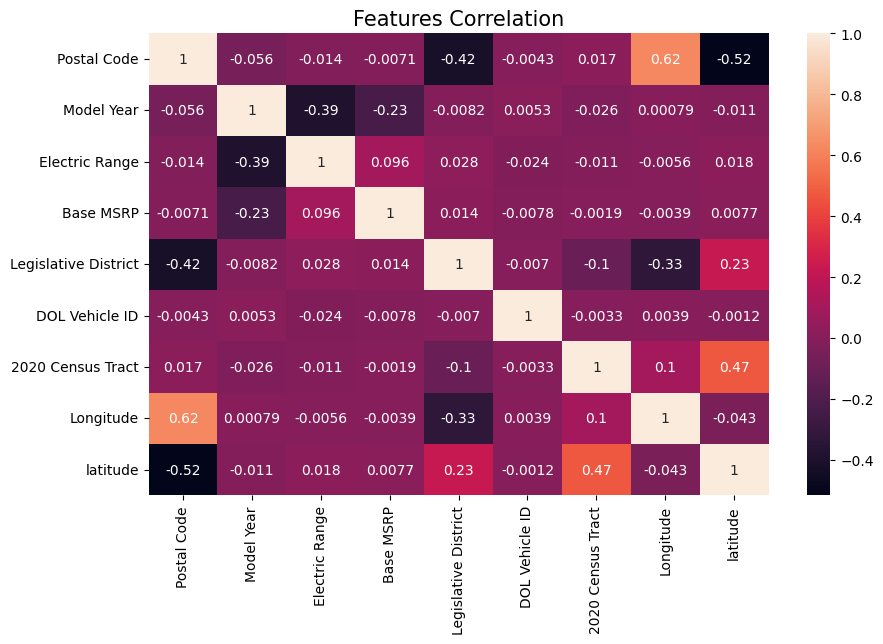


Figure 1.12 Correlations plots

**Interpretations:**

Correlation plays an important role while dealing with any dataset because there is a chance that any data may have a large or high dimension that’s why reducing the dimension correlation helps. If any independent columns have high correlations with another independent column that means both are representing the same thing that is why any column can be removed from the dataset to reduce the dimensionality. Generally, a high correlation is considered above 0.8. so, Fig. 1.12 represents that there is no such high correlation in the dataset. That means there is no chance to reduce the dimensionality.

#### Question no.7: Checking correlations to reduce the dimension if possible?

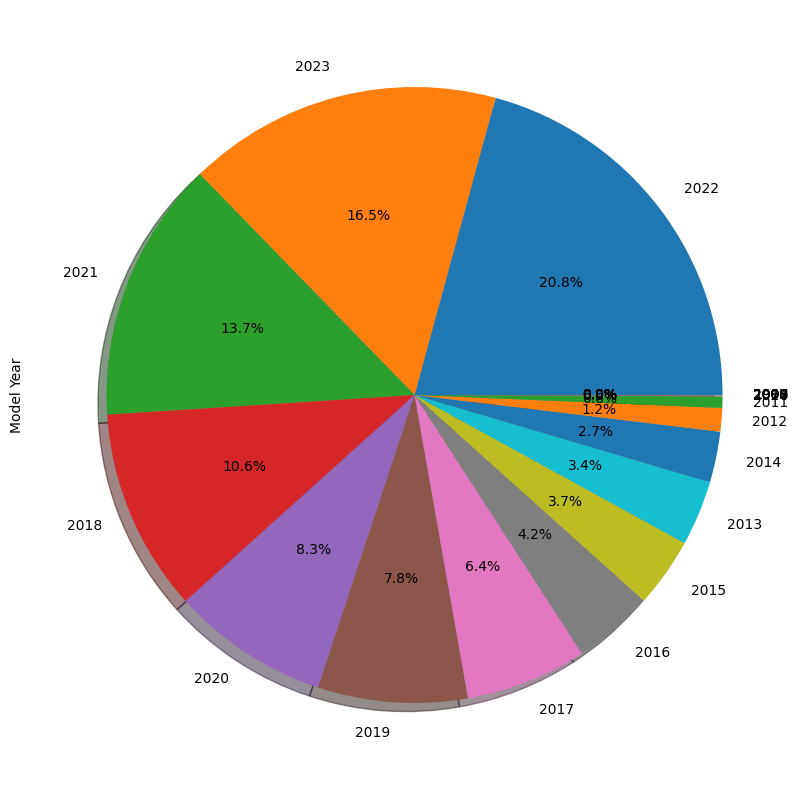


Figure 1.13 Year wise car model distributions

**Interpretations:**

It can be seen in Fig. 1.13 that increasing technology means developing the modern world People are also more attracted to buying an electric vehicle because it saves money and the biggest concern is reducing pollution. Here from 2010 to 2023 slowly people showed more interest in buying electric vehicles.

#### Question no.8: Comparison between Legislative District and Electric Vehicle Type?

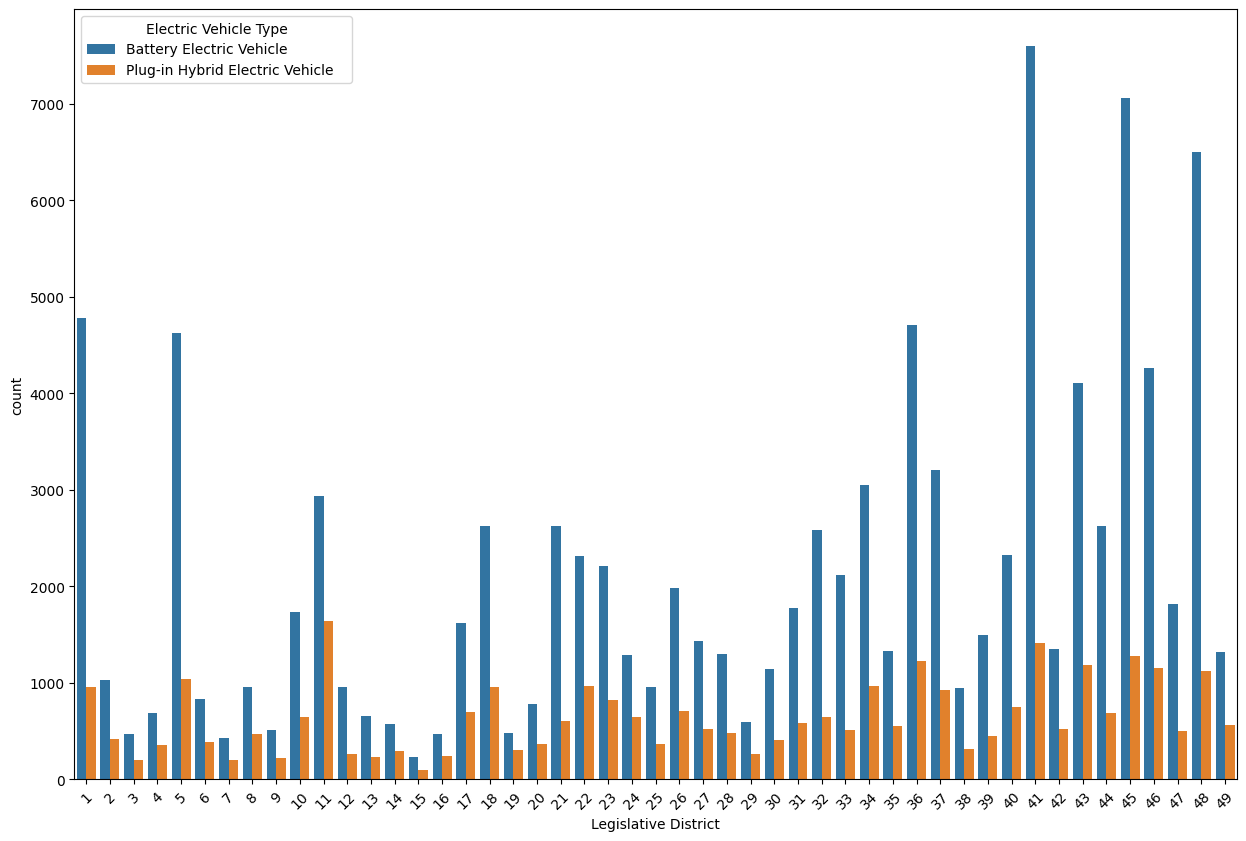


Figure 1.14 Comparison plot between Legislative District and Electric Vehicle Type

**Interpretations:**

It can be seen in Fig. 1.14 that the author represents a comparison plot between the Legislative district and electric vehicle type. From the figure it is understood that 41 Legislative districts have high-battery electric vehicles and it can be also seen that compared to plug-in hybrid electric vehicles battery electric vehicles is more in every legislative district.

#### Question no.9: Finding which County has highest and lowest Base MSRP.

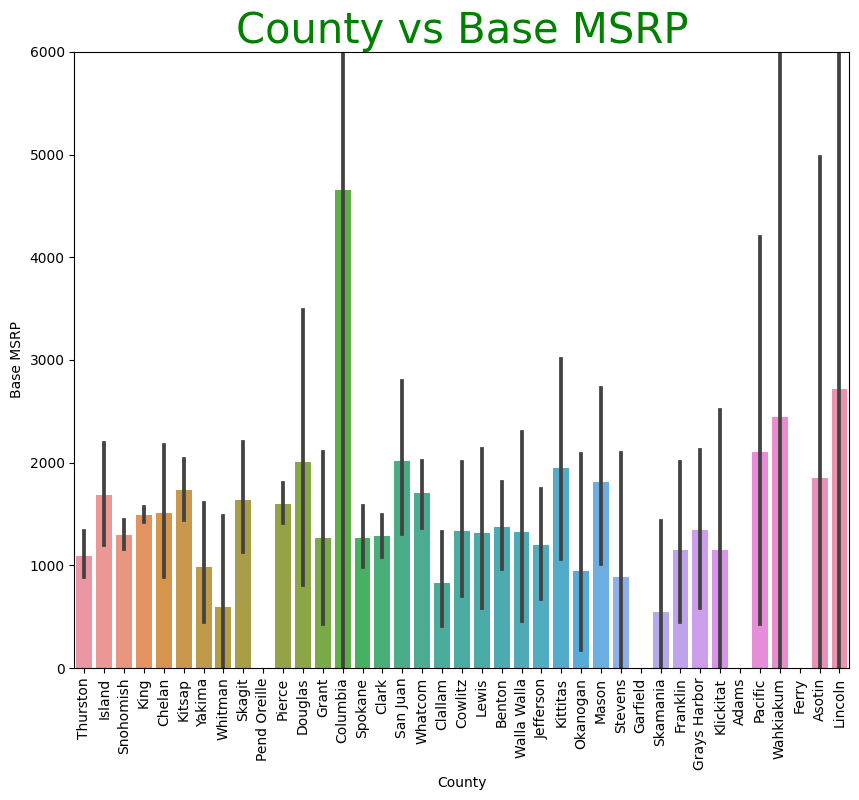


Figure 1.15 County vs Base MSRP distribution

**Interpretations:**

From the Fig. 1.15 it can be clearly observed that compared to all county the Columbia is the highest Base MSRP and the least MSRP county holder are Pend Oreille, Garfield, Adams and Ferry County.

#### Question no.10: County wise electric bike type demand.

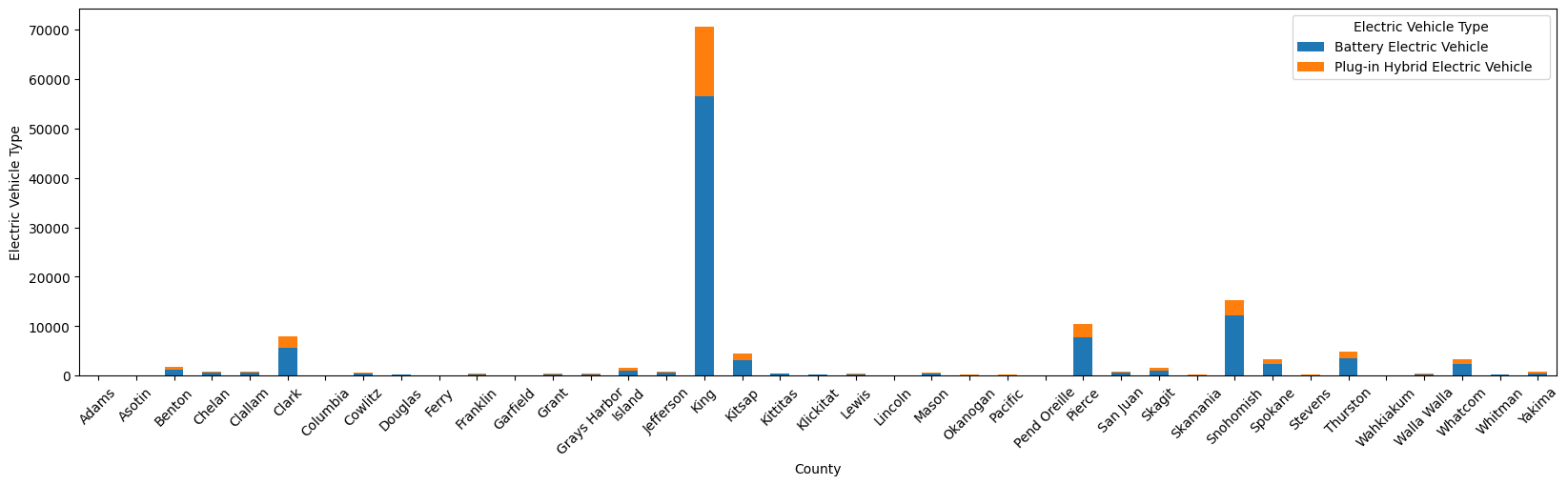


Figure 1.16 County-wise electric bike type demand.

**Interpretations:**

It is visible that King County is the most highly demanded county among all counties for electrical vehicles. It is also found that battery electric vehicle is more in demand compared to plug-in hybrid electric demand. And all these things are represented in Fig. 1.16.

#### Question no.11: Finding which company produce more electric vehicle?

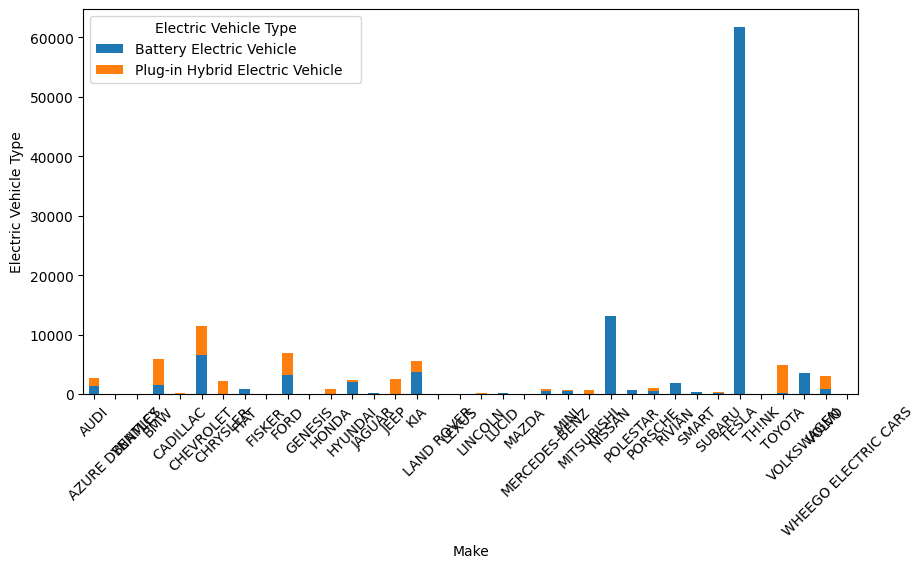


Figure 1.17 Highest electric vehicle producer

**Interpretations:**

From Fig. 1.17 it is seen that compared to all-electric vehicle producer companies SABARU company makes the largest produced electric vehicle and it is also observed they only make battery electric vehicles.

#### Question no.12: Finding which type of electric vehicle type are more in CAFV?

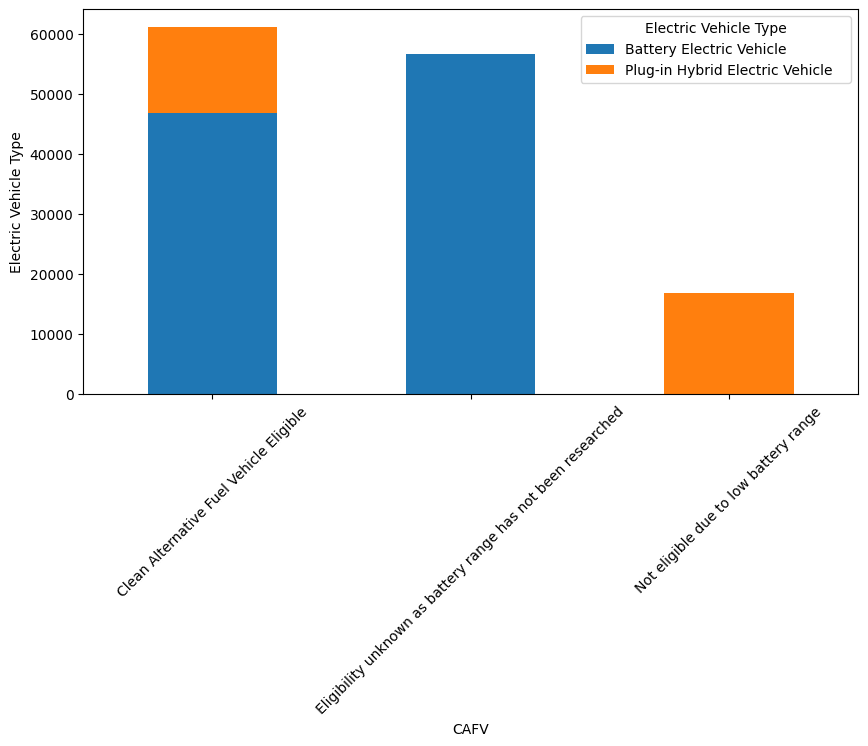


Figure 1.18 Compare between electric vehicle type and CAFV

**Interpretations:**

It is finding or observed from the Fig. 1.18 that eligibility unknown as battery range has not been researched is more compare to other two CAFV types in electric vehicle type.

#### Question no.12: Multivariate plot to find the distribution of all attributes.

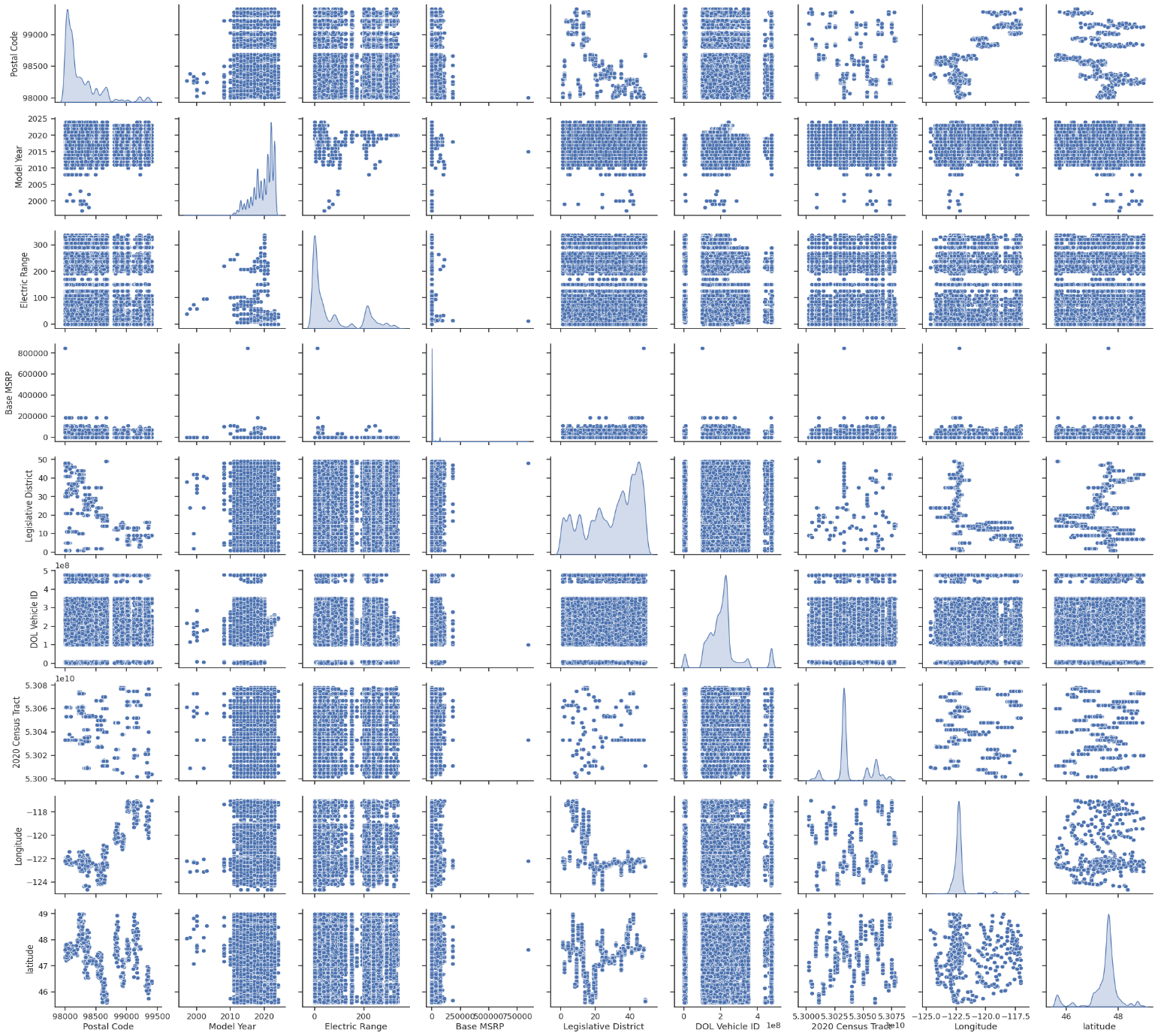


Figure 1.19 Distribution of plots through Pair plot

**Interpretations:**

From Fig. 1.19 it is understood that none of the attributes follows the normal distributions.

# Conclusion

In conclusion, performing EV analysis using Python opens the door to insights that can guide the future of transportation. Through data collection, preprocessing, visualization and predictive modeling, this analysis reveals patterns that provide deeper insights into EV behavior. Python tools allow us to predict driving distance, estimate efficiency and simulate charging scenarios, allowing us to optimize the use of electric cars and promote sustainable mobility. In this analytical journey, Python transforms raw data into actionable intelligence, bridging the gap between technological innovation and environmental management. The ability to leverage Python's capabilities puts us at the forefront of an evolving landscape where electric cars are not just a means of transportation, but a catalyst for change in a world that demands cleaner, smarter, and more efficient mobility solutions. By developing our methods and embracing the power of Python, we are paving the way to a future where electric vehicles are seamlessly integrated into our lives, redefining the way we move and shaping a more sustainable tomorrow.

## Findings

* the type of electric vehicle is not balanced because battery electricity is almost 76.91 percent and plug-in hybrid car is only 23.09 percent. Correlation plays an important role when dealing with any data set because it is possible that some data may have high or high dimensionality, so reducing dimensional correlation helps.
* If an independent column has a high correlation with another independent column, it means that both represent the same thing, so each column can be removed from the dataset to reduce dimensionality. In general, a correlation greater than 0.8 is considered high.
* King County is the most in-demand county for electric cars out of all the counties. It has also been observed that the demand for a battery electric vehicle is higher than for a plug-in hybrid.
* battery range not studied is comparable to the other two CAFV types in EV types. modern world People are also more interested in buying an electric car because it saves money and the biggest concern is to reduce pollution. Here, from 2010 to 2023, people slowly showed more interest in buying electric cars.
* 41 legislative districts have a high rate of battery electric vehicles, and it can also be seen that each legislative district has more battery electric vehicles than plug-in hybrids.

## Suggestions

* Data Collection: Gather relevant data on electric car models, including specifications such as battery capacity, charging speed, range, and price.
* Data Cleaning: Preprocess the collected data by handling missing values, removing duplicates, and ensuring consistent formatting.
* Visualization: Create visualizations using libraries like Matplotlib or Seaborn to compare key features like range vs. price, charging speed distribution, and battery capacity trends.
* Statistical Analysis: Utilize libraries such as Pandas and NumPy to calculate descriptive statistics, correlation coefficients between variables, and distribution insights.
* Machine Learning: Implement regression models to predict factors like range based on battery capacity and charging speed, using libraries like Scikit-Learn or TensorFlow.
* Cost Analysis: Develop a cost of ownership model by factoring in initial purchase price, charging costs, and potential savings compared to traditional gasoline vehicles.
* User Interface: Build a simple user interface using frameworks like Tkinter or web frameworks like Flask to allow users to input car specifications and receive analysis results

# Reference

Ullah, I., Liu, K., Yamamoto, T., Zahid, M. and Jamal, A., 2023. Modeling of machine learning with SHAP approach for electric vehicle charging station choice behavior prediction. Travel Behaviour and Society, 31, pp.78-92.

Zahedi, R., hasan Ghodusinejad, M., Aslani, A. and Hachem-Vermette, C., 2022. Modelling community-scale renewable energy and electric vehicle management for cold-climate regions using machine learning. Energy Strategy Reviews, 43, p.100930.

Ragone, M., Yurkiv, V., Ramasubramanian, A., Kashir, B. and Mashayek, F., 2021. Data driven estimation of electric vehicle battery state-of-charge informed by automotive simulations and multi-physics modeling. Journal of Power Sources, 483, p.229108.

Secinaro, S., Brescia, V., Calandra, D. and Biancone, P., 2020. Employing bibliometric analysis to identify suitable business models for electric cars. Journal of cleaner production, 264, p.121503.

Duraisamy, T. and Kaliyaperumal, D., 2021. Machine Learning-Based Optimal Cell Balancing Mechanism for Electric Vehicle Battery Management System. IEEE Access, 9, pp.132846-132861.