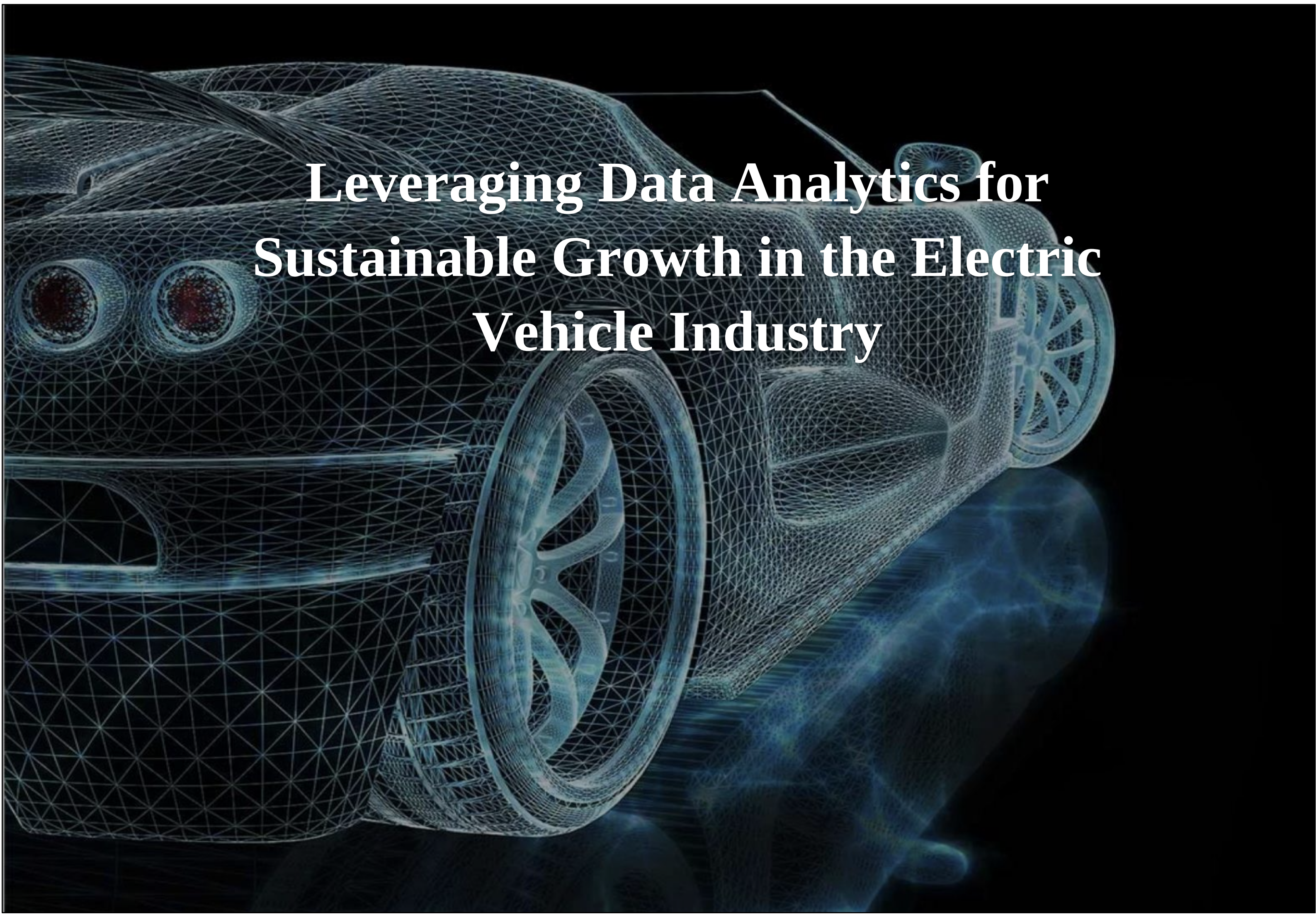
Northeastern
University



Leveraging Data Analytics for Sustainable Growth in the Electric Vehicle Industry

Submitted By

I. Amadin Cynthia Aigbe
II. Ebenezer Opoku-Manu
III. Md. Shahriar Bin Mohsin
IV. Suraju Olanrewaju Ibiyemi

Submitted To

Instructor:	Andy Chen
Couse Title:	Foundations of Data Analytics
CRN:	19846
Period:	Fall 2023

1. Introduction

1.1. Overview of the Electric Vehicle (EV) Industry

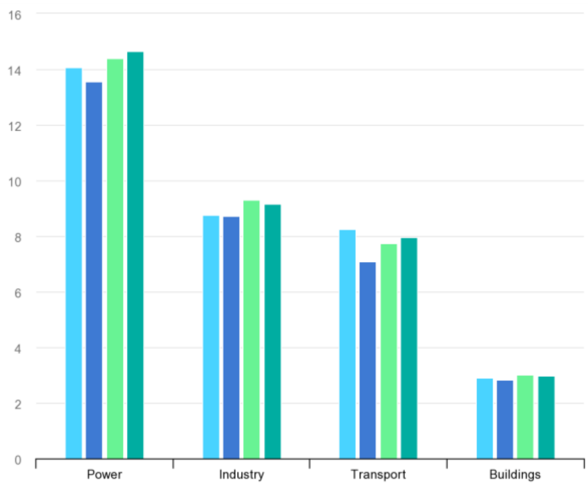
The Electric Vehicle (EV) industry has undergone a transformative shift in recent years, marked by an increasing global focus on sustainable transportation. EVs have emerged as a pivotal solution in the fight against climate change, aiming to reduce carbon dioxide (CO2) emissions attributed to the automotive sector. As the world embraces cleaner and more environmentally friendly alternatives, the EV industry has witnessed exponential growth and technological advancements.

Electric vehicles, once seen as a niche market, have evolved from conceptual prototypes to viable alternatives to traditional internal combustion engine vehicles. The inception of EVs dates back to the early 19th century, but their commercialization gained momentum in the late 20th and early 21st centuries. Advancements in battery technology, cost efficiency, and government initiatives supporting clean energy have accelerated the adoption of EVs.

1.1.1. Environmental Concerns Driving EV Adoption

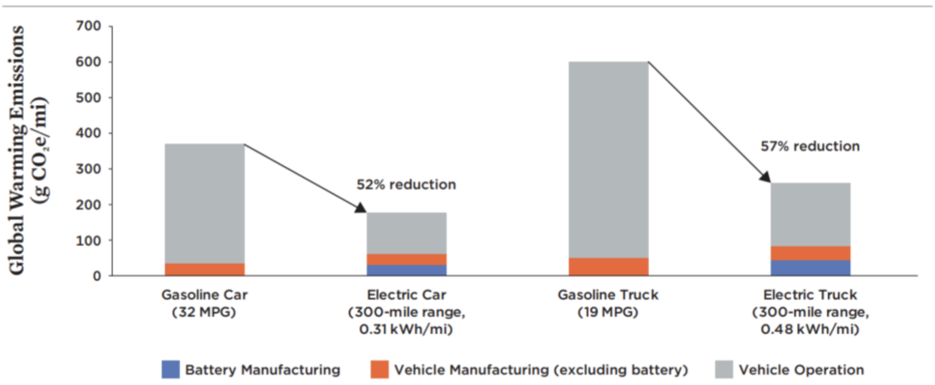
The transportation sector is a significant contributor to global CO2 emissions, accounting for approximately 17% of the total emissions. The surge in automotive emissions by 2.1% in 2022 highlights the urgency to address this environmental concern.

Source: IEA, Global CO2 emissions by sector, 2019-2022, IEA, Paris



Electric vehicle production includes battery manufacturing, leading to initial emissions. However, when comparing lifetime emissions, EVs fare better. Despite their emissions during production, electric cars produce 52% fewer greenhouse gas emissions than gas cars, while electric trucks emit 57% less than gas trucks over their lifespan due to cleaner operation.

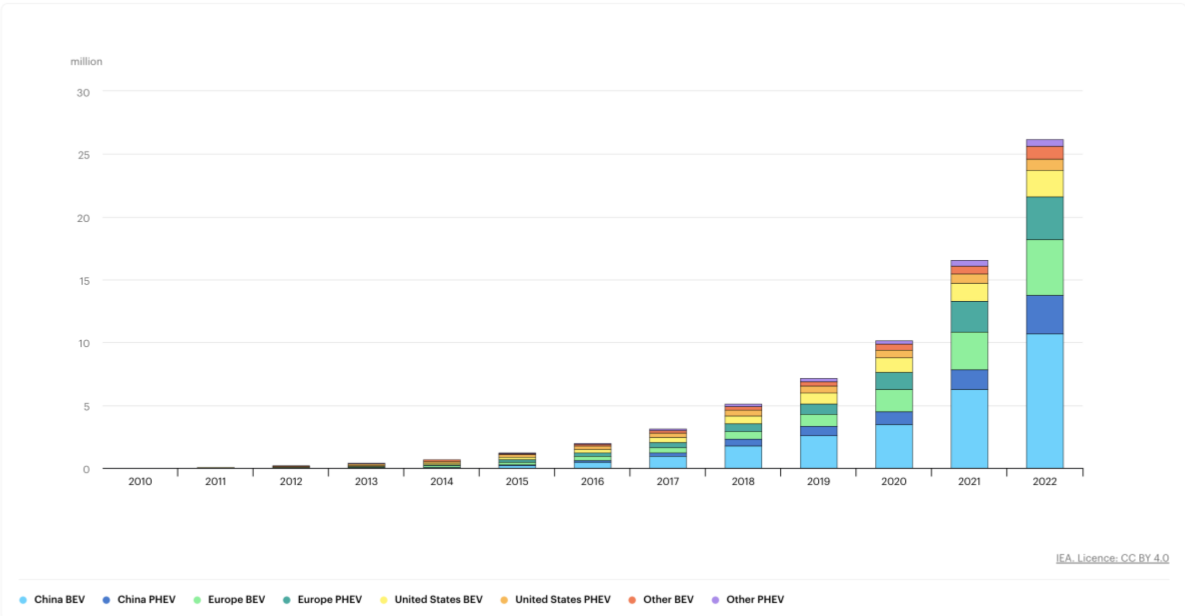
Source: driving-cleaner-report PDF (www.ucsusa.org)



1.1.2. Global Shift towards Electric Mobility

In response to the growing environmental crisis, the global shift towards electric mobility has gained traction. The impact of EV adoption is profound; if all new electric cars in 2022 were traditional oil-based vehicles, global emissions would have risen significantly. The surge in electric car sales, surpassing 10 million and accounting for over 14% of global sales in 2022, demonstrates a substantial shift towards sustainable transportation.

Sources: IEA analysis based on country submissions, ACEA, EAFO, EV Volumes and Marklines.



Technological innovations and decreasing costs have significantly contributed to the widespread acceptance of EVs. Over the last decade, electric vehicle battery prices have plummeted by 89%, making EVs more affordable and competitive in the automotive market. These advancements have propelled the EV industry into a pivotal position in the global automotive landscape.

1.2. Business Requirement: Sustainable Growth to Reduce CO2 Emissions

The imperative to achieve sustainable growth within the Electric Vehicle (EV) industry stems from the pressing need to mitigate carbon dioxide (CO2) emissions and curb the environmental impact of the automotive sector. This business requirement is not just a market necessity but a global imperative driven by environmental concerns and the need for responsible, eco-friendly transportation solutions.

1.2.1. Addressing Environmental Challenges

The automotive sector's contribution to global CO2 emissions cannot be overstated, accounting for approximately 17% of the world's total emissions. This sector's emissions surged by 2.1% (equivalent to 137 million metric tons) in 2022 alone, underscoring the urgency to address these environmental challenges.

Electric Vehicles (EVs) play a pivotal role in reducing greenhouse gas emissions. On average, EVs produce approximately 54% fewer greenhouse gas emissions than conventional internal combustion engine vehicles. If all new electric cars sold in 2022 were traditional oil-based vehicles, global emissions would have soared an additional 13 million metric tons*(Source: IEA Reports/ CO2 Emissions in 2022).*

Beyond environmental considerations, embracing sustainable growth in the EV industry underscores a sense of social responsibility. It not only contributes to reducing emissions but also drives economic viability and innovation, fostering a healthier and more sustainable future for generations to come.

1.2.2. Impact of Electric Car Sales

The shift towards embracing electric mobility is evident in the substantial increase in electric car sales. Surpassing 10 million units globally in 2022, electric cars accounted for over 14% of total vehicle sales. This surge highlights a promising trend towards sustainable transportation.

1.2.3. Affordability and Market Penetration

Advancements in technology have significantly contributed to making EVs more affordable. The drastic 89% reduction in electric vehicle battery prices over the last decade has played a pivotal role in enhancing their market penetration, making them increasingly competitive against conventional vehicles.

1.2.4. Business Imperative: Leveraging Data Analytics

Achieving sustainable growth in the EV industry not only aligns with environmental goals but also presents a significant business opportunity. Leveraging data analytics to forecast market needs, guide optimal actions, and derive insights into historical performance is crucial for fostering this sustainable growth trajectory.

2. Role of Data Analytics in Achieving Sustainable Growth

2.1. Descriptive Analytics: Offering Insights into Historical Performance

Descriptive Analytics constitutes the foundational stage of data analytics, focusing on summarizing and interpreting historical data to gain valuable insights. Within the Electric Vehicle (EV) industry, Descriptive Analytics plays a crucial role in understanding past trends, market behavior, and performance metrics to inform strategic decision-making.

2.1.1. Summarizing Historical Data

Descriptive Analytics involves summarizing vast amounts of historical EV data. It encompasses techniques like summary statistics, data aggregation, and visualization to present a clear and concise overview of past EV sales, market trends, consumer demographics, and regional preferences.

2.1.2. Understanding Market Trends

By analyzing historical EV sales data, Descriptive Analytics allows stakeholders to understand market trends and patterns. It helps identify peak sales periods, seasonal variations, regional preferences, and shifts in consumer behavior over time.

2.1.3. Customer Segmentation and Demographics

Descriptive Analytics facilitates customer segmentation based on historical data, enabling organizations to categorize consumers by preferences, behaviors, and demographics. This segmentation aids in targeted marketing strategies and product development tailored to specific consumer segments.

2.1.4. Performance Evaluation and Benchmarking

Within the EV industry, Descriptive Analytics enables performance evaluation by benchmarking key metrics such as sales volumes, market shares, and growth rates against historical data. This analysis helps in assessing progress, identifying areas of improvement, and setting realistic goals for sustainable growth.

2.1.5. Informed Decision-Making

Descriptive Analytics serves as a fundamental tool for informed decision-making within the EV industry. By providing insights into historical performance, market trends, and consumer behavior, it equips stakeholders with the necessary information to formulate effective strategies and make data-backed decisions

2.2. Predictive Analytics: Anticipating Market Needs

Predictive Analytics serves as a pivotal tool in the dominion of data analytics, enabling organizations to forecast and anticipate future trends, behaviors, and market demands. In the context of the Electric Vehicle (EV) industry, Predictive Analytics plays a significant role in shaping sustainable growth strategies by forecasting market needs and consumer preferences.

2.2.1. Forecasting Market Demand

Leveraging historical EV sales data, market trends, and various external factors, Predictive Analytics enables industry stakeholders to forecast future market demand for Electric Vehicles. By analyzing patterns and correlations, predictive models can anticipate shifts in consumer preferences, thereby facilitating strategic planning for production, inventory management, and marketing initiatives.

2.2.2. Enhancing Product Development

Predictive Analytics aids in enhancing product development strategies within the EV industry. By analyzing consumer feedback, preferences, and technological advancements, organizations can tailor their offerings to meet evolving consumer needs. This proactive approach ensures the development of innovative EV models aligned with market demands, ultimately driving sustainable growth.

2.2.3. Optimizing Inventory and Supply Chain

Anticipating market needs allows for better inventory management and optimization of the supply chain. Predictive Analytics assists in forecasting demand fluctuations, enabling companies to streamline production schedules, manage inventory levels efficiently, and mitigate unnecessary stockpiling or shortages.

2.2.4. Personalizing Marketing Strategies

Predictive Analytics empowers organizations to personalize marketing strategies. By analyzing consumer behavior, demographic trends, and purchase patterns, companies can create targeted marketing campaigns. These campaigns not only maximize reach but also foster a more sustainable approach by directing marketing efforts towards potential EV consumers.

2.2.5. Future-Focused Insights

Predictive Analytics serves as a cornerstone in steering the Electric Vehicle industry towards sustainable growth. By providing future-focused insights into market trends, consumer preferences, and product development, Predictive Analytics empowers industry stakeholders to make informed decisions, optimize operations, and contribute to the proliferation of environmentally friendly mobility solutions.

2.3. Prescriptive Analytics: Guiding Optimal Actions

Prescriptive Analytics represents the pinnacle of data analytics, offering actionable insights that guide optimal decision-making processes within the Electric Vehicle (EV) industry. This sophisticated form of analytics not only predicts future outcomes but also prescribes the best course of action to achieve desired objectives and foster sustainable growth.

2.3.1. Data-Driven Decision Making

Prescriptive Analytics harnesses advanced algorithms and modeling techniques to simulate various scenarios based on historical data and predictive models. By assessing these simulations, it provides decision-makers with actionable insights, empowering them to make informed and data-driven choices.

2.3.2. Optimizing Operational Efficiencies

Within the EV industry, Prescriptive Analytics plays a crucial role in optimizing operational efficiencies. It identifies inefficiencies in production, distribution, and resource allocation, offering recommendations to streamline processes, reduce waste, and enhance overall productivity.

2.3.3. Pricing and Revenue Optimization

Prescriptive Analytics aids in pricing strategies and revenue optimization for EV manufacturers and service providers. By analyzing market demand, competitor pricing, and consumer behavior, it suggests optimal pricing structures that maximize revenue while ensuring competitive positioning.

2.3.4. Resource Allocation and Investment Decisions

In the realm of sustainable growth, Prescriptive Analytics guides resource allocation and investment decisions. It assists in identifying strategic areas for investment, such as research and development for innovative EV technologies, charging infrastructure expansion, or sustainability initiatives aligned with market trends.

2.3.5. Strategic Decision Support

Prescriptive Analytics acts as a strategic decision support system within the EV industry. By providing actionable recommendations, it empowers stakeholders to make informed decisions, optimize operations, allocate resources effectively, and steer towards sustainable growth pathways.

3. Stakeholders in the EV Industry

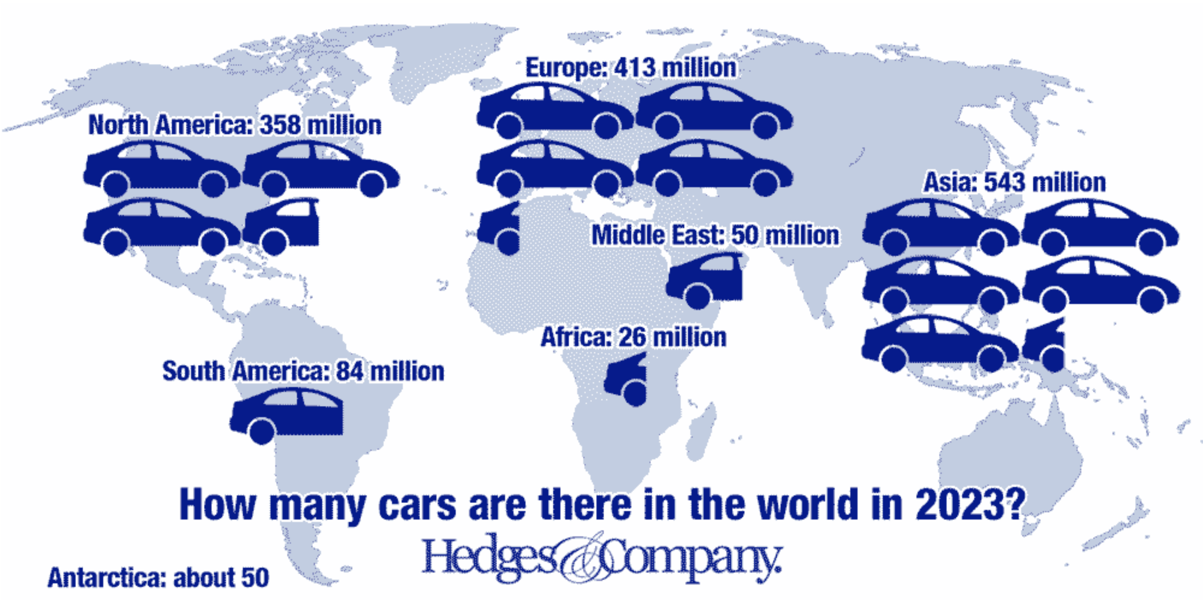
3.1. Consumers: Global Vehicle Distribution Statistics

Consumers are pivotal stakeholders within the Electric Vehicle (EV) industry, shaping market dynamics and the adoption of environmentally friendly mobility solutions. Understanding global vehicle distribution among different regions offers insights into consumer behaviors, preferences, and the potential for EV market penetration.

3.1.1. Global Vehicle Distribution Statistics

As of 2023, there are approximately 1.474 billion vehicles worldwide, with the United States accounting for about 19% of the total. Interestingly, electric vehicles represent approximately 1 in every 250 cars on the road, constituting a global market share of around 2.2%.

Source: HOW MANY CARS ARE THERE IN THE WORLD IN 2023?, How Many Electric Cars in the World?



3.1.2. Overview of Regional Vehicle Ownership

North America: With an ownership ratio of approximately 710 vehicles per thousand people, North America exhibits a relatively high vehicle ownership rate.

Europe: Boasting around 520 vehicles per thousand people, Europe stands as another significant automotive market.

Other Continents: South America, Middle East, Asia/Oceania, Africa, and Antarctica exhibit varying vehicle ownership ratios, with Asia/Oceania hosting over a third of all vehicles globally.

3.1.3. Insights into Consumer Behavior

Contrary to popular belief, North America doesn't possess the highest number of vehicles globally; Europe holds a 28% share, indicating a robust automotive presence. Asia/Oceania showcases a considerable market potential with its diverse ownership patterns, presenting opportunities for EV adoption.

3.1.4. Tailored Strategies for EV Penetration

Understanding these regional disparities in vehicle ownership guides the industry toward tailored strategies. This includes targeted EV promotion, marketing campaigns, and infrastructure development, catering to specific regional preferences and behaviors. A consumer-centric approach is vital for effective EV adoption, considering these variations in vehicle ownership worldwide.

3.1.5. Guiding the Consumer-Centric Approach

The global distribution statistics of vehicles serve as a compass for the Electric Vehicle industry. They underscore the significance of adopting tailored strategies to promote EV adoption, considering regional disparities in vehicle ownership and consumer behaviors. This insight aids in developing effective approaches to drive sustainable growth in the EV market.

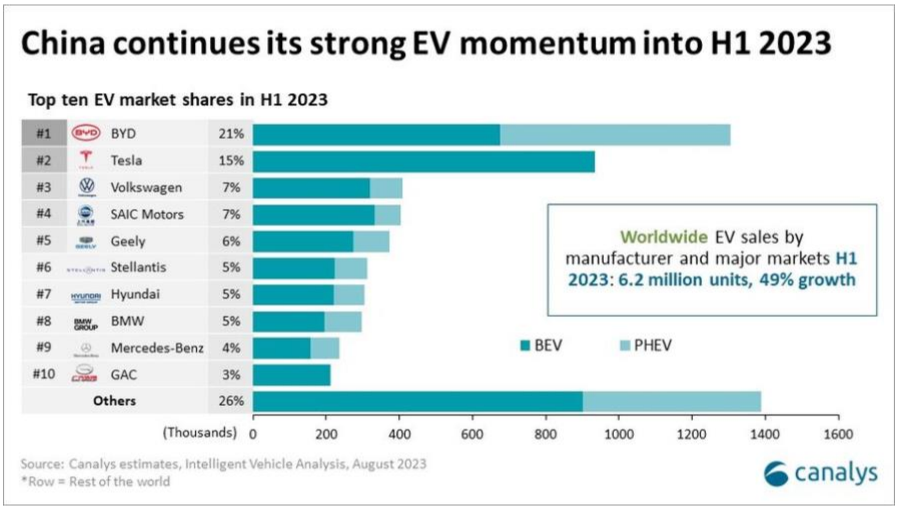
3.2. Automakers: Market Share Analysis and Leading Brands

Automakers play a pivotal role in driving innovation, technological advancements, and market competitiveness within the Electric Vehicle (EV) industry. Analyzing market shares and identifying leading brands offers critical insights into the landscape of EV manufacturers and their contributions to sustainable mobility solutions.

3.2.1. Market Share Analysis

In the first half of 2023, the EV market experienced a remarkable 49% growth, reaching 6.2 million units, constituting 16% of the global light vehicle market. This substantial increase of 12.4% from the first half of 2022 signifies a significant surge in EV adoption. Insights from comprehensive market analysis highlight the dominance and influence of key automakers:

(Source: worldwide sales of electric vehicles)



BYD: Leading the market share, BYD emerged as a frontrunner with an impressive sales volume of 287,454 new energy vehicles (NEVs) in September 2023, marking a 43.0% increase from the previous year and a 4.8% rise from August 2023. Cumulative sales in 2023 reached 2,079,638 units, reflecting a substantial year-on-year increase of 75.5%.

(Source: BYD delivered a record-breaking 287,454 EVs in September)

Tesla: Renowned for its innovative technologies, Tesla continues to maintain a significant market share. Setting a quarterly record in Q2 2023 with 466,140 vehicles delivered and aiming to reach 2 million deliveries by the end of 2023, Tesla remains a leader in the EV sector.

(Source: Tesla Sales, Revenue & Production for 2023)

Nissan, Chevrolet, and Other Key Players: Noteworthy contributions from Nissan, Chevrolet, and several other automakers have contributed to the diversification of the EV market, albeit with varying market shares.

3.2.2. Leading Brands and Their Impact

The market dominance of BYD, Tesla, and other leading brands significantly influences the pace and direction of the EV industry's growth trajectory.

Tesla's innovative EV models, sustainable approach, and extensive charging infrastructure have propelled the company to the forefront, setting benchmarks for technological advancements and consumer expectations.

BYD's achievements, especially in the Chinese market, have demonstrated the scalability and potential for widespread EV adoption, showcasing the brand's commitment to producing and selling millions of new energy vehicles (NEVs).

3.2.3. Impact on Sustainable Mobility

The market share analysis and prominence of leading brands underscore the pivotal role automakers play in advancing sustainable mobility solutions. Their contributions drive technological innovations, economies of scale, and widespread adoption of environmentally friendly transportation alternatives.

Collaborations, investments in research and development, and strategic initiatives by these leading brands contribute significantly to shaping the future of the EV industry, emphasizing the importance of competition and collaboration in driving sustainable growth.

3.2.4. Driving Innovation and Market Growth

Automakers' market share analysis highlights the competitive landscape and impact of key players in steering the Electric Vehicle industry towards sustainable growth. Their innovations, market strategies, and commitment to advancing EV technology underpin the industry's transformation and global transition towards cleaner and greener mobility solutions.

3.3. Government Agencies: Regulatory Role and Incentivizing Sustainable Mobility

Government agencies wield significant influence and responsibility in shaping policies, regulations, and incentives that drive the adoption of Electric Vehicles (EVs) and facilitate the transition towards sustainable transportation solutions.

3.3.1. Regulatory Role

Governments worldwide play a pivotal role in crafting regulations that influence the development, manufacturing, and adoption of EVs. Emission standards, vehicle efficiency regulations, and mandates on clean energy adoption contribute to shaping the EV landscape

3.3.2. Incentivizing Sustainable Mobility

Various agencies implement incentives to stimulate EV adoption, including tax credits, rebates, grants, and subsidies for purchasing EVs, installing charging infrastructure, or developing renewable energy sources. These incentives aim to reduce the barrier to entry and encourage consumers, automakers, and other stakeholders to transition towards cleaner mobility options.

3.3.3. Funding and Infrastructure Developmen

Government agencies allocate funds for research, development, and infrastructure projects essential for expanding EV adoption. Investments in charging networks, grid modernization, and renewable energy deployment are crucial to support the growing EV market and enhance its accessibility.

3.3.4. International Collaboration and Policy Advocacy

Collaboration between governments, international organizations, and advocacy groups is instrumental in fostering a supportive global environment for EV adoption. Joint agreements, policy frameworks, and shared standards facilitate a harmonized approach towards sustainable mobility solutions on a global scale

3.3.5. Impacts on Industry Growth and Consumer Perception

The regulatory framework and incentives crafted by government agencies significantly influence industry growth, market competitiveness, and consumer perceptions. Clear policies and supportive measures create an environment conducive to innovation, investment, and consumer confidence in EV technologies.

3.3.6. Government as an Enabler for Sustainable Mobility

Government agencies serve as key enablers in the evolution of the EV industry towards sustainability. Their regulatory actions, incentives, investments, and advocacy efforts foster an ecosystem conducive to the widespread adoption of EVs, driving the global transition towards cleaner and more environmentally friendly transportation solutions.

3.4. Investors: Driving Forward Momentum with Substantial Investments

Investors play a pivotal role in propelling the Electric Vehicle (EV) industry forward, channeling substantial investments that are set to redefine the landscape of sustainable transportation.

3.4.1. Significant Capital Infusion

Between August 2022 and March 2023, major EV and battery manufacturers made groundbreaking announcements, pledging a cumulative post-IRA investment of USD 52 billion in North American EV supply chains. These investments are strategically distributed, with 50% earmarked for battery manufacturing, and approximately 20% each allocated to battery components and EV manufacturing.

(Source: Trends in electric light-duty vehicles)

3.4.2. Comprehensive Industry Commitments

Overall, company announcements, coupled with tentative commitments for future battery and EV production in the United States, estimate a staggering investment range of USD 75-108 billion. These commitments exemplify a robust dedication to revolutionizing the EV industry's infrastructure and production capacities.

3.4.3. Transformative Ventures

Leading the charge, Tesla's strategic move to relocate its Berlin-based lithium-ion battery gigafactory to Texas marks a significant step. Collaborating with China's CATL, Tesla aims to produce next-generation EVs in Mexico, fueling innovation and expanding manufacturing capabilities.

3.4.4. Collaborative Industry Advancements

Ford's collaboration with CATL for a battery plant in Michigan signifies a pivotal shift towards scaling up electric car manufacturing. Their plan includes a sixfold increase in electric car production by the end of 2023, reaching 600,000 vehicles annually and setting sights on scaling up to 2 million by 2026.

3.4.5. Global Expansion Initiatives

BMW's initiative to expand EV manufacturing at its South Carolina plant post-IRA demonstrates a concerted effort to bolster production capabilities and cater to the growing demand for electric vehicles. Meanwhile, Volkswagen's investments in Canada for its inaugural battery plant outside Europe and substantial investments in its South Carolina plant indicate a commitment to global expansion strategies.

3.4.6. Anticipating Growth and Impact

These monumental investments are expected to catalyze high growth potential in the EV industry. However, their full impact might be realized post-2024 as these strategic plants come online, marking a transformative phase towards a sustainable, electrified future of mobility.

4. Dataset Overview and Preprocessing

4.1. Original Dataset Information and Features

The original dataset, "Electric Vehicle Title and Registration Activity," for State of Washington sourced from data.wa.gov, serves as a foundational resource for understanding the Electric Vehicle (EV) landscape. The dataset encapsulates a plethora of information crucial for analysis and decision-making within the EV industry.

4.1.1. Dataset Source and Scale

Source: The dataset originates from the State of Washington, providing a comprehensive overview of EV title and registration activity.

Scale: Featuring an extensive repository, the dataset encompasses 832,178 entries and 35 columns, presenting a substantial pool of EV-related data for analysis.

```
RangeIndex: 832178 entries, 0 to 832177
Data columns (total 35 columns):
 #   Column                                                 Non-Null Count   Dtype
---  ------                                                 --------------   -----
 0   Clean Alternative Fuel Vehicle Type                    832178 non-null  object
 1   VIN (1-10)                                             832178 non-null  object
 2   DOL Vehicle ID                                         832178 non-null  int64
 3   Model Year                                             832178 non-null  int64
 4   Make                                                   832178 non-null  object
 5   Model                                                  832178 non-null  object
 6   Vehicle Primary Use                                    832178 non-null  object
 7   Electric Range                                         832178 non-null  int64
 8   Odometer Reading                                       832178 non-null  int64
 9   Odometer Code                                          832178 non-null  object
 10  New or Used Vehicle                                    832178 non-null  object
 11  Sale Price                                             832178 non-null  int64
 12  Sale Date                                              248278 non-null  object
 13  Base MSRP                                              832178 non-null  int64
 14  Transaction Type                                       832178 non-null  object
 15  DOL Transaction Date                                   832178 non-null  object
 16  Transaction Year                                       832178 non-null  int64
 17  County                                                 832142 non-null  object
 18  City                                                   832106 non-null  object
 19  State of Residence                                     832177 non-null  object
 20  Postal Code                                            832130 non-null  float64
 21  2015 HB 2778 Exemption Eligibility                     832178 non-null  object
 22  2019 HB 2042 Clean Alternative Fuel Vehicle (CAFV) Eligibility 832178 non-null  object
 23  Meets 2019 HB 2042 Electric Range Requirement          832178 non-null  bool
 24  Meets 2019 HB 2042 Sale Date Requirement               832178 non-null  bool
 25  Meets 2019 HB 2042 Sale Price/Value Requirement        832178 non-null  bool
 26  2019 HB 2042: Battery Range Requirement                323028 non-null  object
 27  2019 HB 2042: Purchase Date Requirement                832178 non-null  object
 28  2019 HB 2042: Sale Price/Value Requirement             832178 non-null  object
 29  Electric Vehicle Fee Paid                              832178 non-null  object
 30  Transportation Electrification Fee Paid                741147 non-null  object
 31  Hybrid Vehicle Electrification Fee Paid                741147 non-null  object
 32  2020 Census Tract                                      832142 non-null  float64
 33  Legislative District                                   829268 non-null  float64
 34  Electric Utility                                       832142 non-null  object
dtypes: bool(3), float64(3), int64(7), object(22)
memory usage: 205.5+ MB
```

4.1.2. Data Format

The Electric Vehicle (EV) dataset is organized and kept in a machine-readable format. The dataset is structured in the Comma-Separated Values (CSV) format. Each row in the CSV file corresponds to a single record, and the columns include different variables pertaining to different models of electric vehicles.

4.1.3. Key Features and Field.

The dataset comprises a diverse range of features that encapsulate crucial information regarding Electric Vehicles and their registration activities. Some key fields include:

- Clean Alternative Fuel Vehicle Type: Categorization of vehicle types based on alternative fuel usage.
- VIN (1-10): Vehicle Identification Number capturing unique vehicle identities.
- Model Year, Make, Model: Specifications detailing the year, make, and model of EVs.
- Electric Range: Information on the electric range provided by EVs.
- Odometer Reading, Sale Price: Details encompassing odometer readings and sale prices.
- Transaction Details: Information regarding transaction types, dates, and associated codes
- Geographical Details: Fields such as county, city, state of residence, and postal code provide geographic context.

4.1.4. Purpose and Importance

The dataset holds paramount importance in dissecting trends, understanding market dynamics, and formulating strategies within the EV industry. It serves as a foundation for uncovering insights related to EV sales, consumer preferences, market penetration, and geographical patterns.

Its breadth and depth offer an opportunity to conduct diverse analyses, predictive modeling, and prescriptive recommendations, contributing significantly to sustainable growth initiatives within the EV sector.

4.1.5. Start and End Date

For this extensive Electric Vehicle (EV) dataset, data gathering began(Sale Date) in 2008 and continue until September, 2023 . This long period of time which spans one and half decades, from 2008 to 2023 was carefully selected to capture the development and evolution of electric vehicle types over time.

4.1.6. Dataset Significance

The original dataset's richness in EV-related information and comprehensive nature positions it as a cornerstone for deriving actionable insights, predictive modeling, and making informed decisions. Its utilization in preprocessing and subsequent analysis holds immense potential for uncovering trends and patterns crucial for fostering sustainable growth in the Electric Vehicle industry.

4.2. Data Cleaning: Handling Missing Values and Refining Columns

Data cleaning stands as a critical step in the preprocessing phase, ensuring data accuracy and reliability for subsequent analysis within the Electric Vehicle (EV) dataset. Addressing missing values and refining columns involved meticulous strategies to enhance the dataset's quality and usability.

4.2.1. Handling Missing Values

Backward and Forward Filling: Initially, missing values in specific columns, particularly 'Sale Date,' were handled using backward and forward filling techniques. This approach aimed to impute missing date entries based on adjacent valid date values, enhancing temporal data accuracy.

Dictionary-based Imputation for 'Sale Price': To tackle the challenge of missing 'Sale Price' entries, a strategic approach involved the creation of distinct dictionaries for new and used cars. These dictionaries meticulously stored the yearly average sale prices corresponding to each make's specific model cars.

```
average_sale_prices_used = {
    2019: {
        'Toyota': {
            'Corolla': 15000,
            'Camry': 20000,
            # Other Toyota models and their average prices for 2019
        },
        # Other makes and their models and prices for 2019
    },
    # Other years and their respective makes, models, and average prices
}
```

Fig: Structure of the dictionary

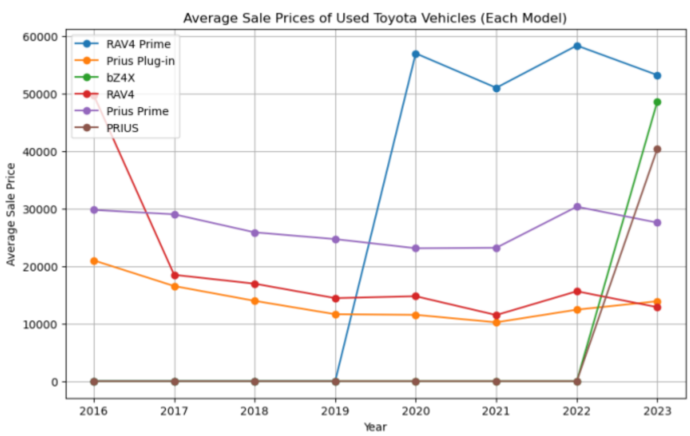


Fig: Contents in the dictionary

Leveraging these dictionaries, the missing 'Sale Price' values were expertly imputed by cross-referencing the available data. This process notably reduced the overall percentage of missing sale prices from a staggering 77.5% down to a significantly improved 15%.

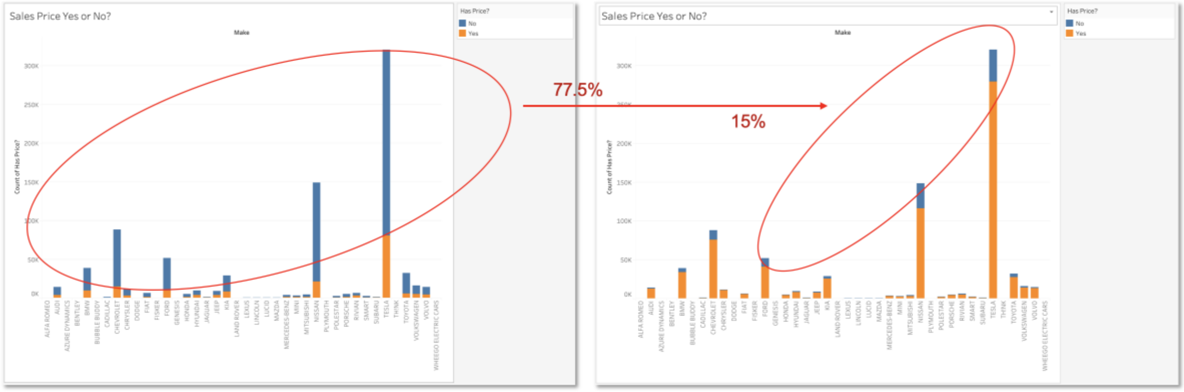


Fig: Comparison of before and after Imputation

However, it's important to note that the remaining 15% of missing sale prices were those for which no reference value was available within the dictionaries. These entries, lacking reference values, remained unchanged as part of the data integrity preservation process.

4.2.2. Data Format Validation

Carried out a careful and thorough examination of the Sale date variable, using stringent validation checks and parsing methods to ensure format consistency to make sure the Sale date values followed the dataset's standardized date format.

4.2.3. Refining Column

Drop Unnecessary Fields: To streamline the dataset for analysis, columns deemed irrelevant or redundant were dropped. This process eliminated unnecessary noise and reduced the dataset's complexity while focusing on pertinent features essential for EV analysis.

Final Dataset Configuration: Post-cleaning and refinement processes, the dataset was configured to contain 20 columns. This optimized dataset ensures improved usability and facilitates more effective analysis and modeling.

4.2.4. Importance of Data Cleaning

Data cleaning plays a pivotal role in enhancing the dataset's quality and reliability, ensuring that subsequent analyses and modeling are based on accurate, consistent, and complete data.

The refined dataset resulting from these cleaning techniques forms a robust foundation for comprehensive exploratory data analysis, predictive modeling, and prescriptive analytics within the EV industry. It facilitates more precise insights and informed decision-making.

4.2.5. Enhanced Data Integrity

By addressing missing values through specialized imputation techniques and refining columns to focus on relevant features, the data cleaning phase significantly bolstered the dataset's integrity and usability. This cleaned and refined dataset serves as a reliable basis for driving insightful analyses and deriving actionable insights in the Electric Vehicle domain.

4.3. Final Cleaned Dataset Structure

The culmination of rigorous data cleaning and refinement processes resulted in a refined dataset structure, setting the stage for comprehensive analysis and exploration within the Electric Vehicle (EV) industry.

4.3.1. Optimized Dataset Configuration

The dataset underwent meticulous cleaning methodologies, addressing missing values, refining columns, and dropping unnecessary fields to enhance its quality and usability.

Post-cleaning, the dataset now stands configured with 832,178 entries and 20 columns, ensuring a more streamlined and pertinent representation of EV-related information.

4.3.2. Enhanced Data Integrity and Usability

Through meticulous handling of missing values and refining columns, the cleaned dataset significantly improved data integrity, minimizing inconsistencies and inaccuracies that could hinder subsequent analyses.

The refined dataset's streamlined structure focuses on essential features crucial for insightful analyses, predictive modeling, and informed decision-making within the EV industry.

4.3.3. Key Features and Fields

The final cleaned dataset encapsulates pertinent information essential for in-depth exploratory data analysis (EDA), predictive modeling, and prescriptive analytics:

Features such as 'Clean Alternative Fuel Vehicle Type,' 'Model Year,' 'Make,' 'Model,' 'Electric Range,' 'Odometer Reading,' 'Sale Price,' and 'Transaction Details' remain pivotal for comprehensive analysis.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 832178 entries, 0 to 832177
Data columns (total 20 columns):
 #   Column                             Non-Null Count   Dtype
---  ------                             --------------   -----
 0   Clean Alternative Fuel Vehicle Type  832178 non-null  object
 1   VIN (1-10)                         832178 non-null  object
 2   DOL Vehicle ID                     832178 non-null  int64
 3   Model Year                         832178 non-null  int64
 4   Make                               832178 non-null  object
 5   Model                              832178 non-null  object
 6   Vehicle Primary Use                832178 non-null  object
 7   Electric Range                     832178 non-null  int64
 8   Odometer Reading                   832178 non-null  int64
 9   Odometer Code                      832178 non-null  object
 10  New or Used Vehicle                832178 non-null  object
 11  Sale Price                         832178 non-null  int64
 12  Sale Date                          832178 non-null  datetime64[ns]
 13  Base MSRP                          832178 non-null  int64
 14  Transaction Type                   832178 non-null  object
 15  DOL Transaction Date               832178 non-null  object
 16  Transaction Year                   832178 non-null  int64
 17  County                             832178 non-null  object
 18  City                               832178 non-null  object
 19  State of Residence                 832178 non-null  object
dtypes: datetime64[ns](1), int64(7), object(12)
memory usage: 127.0+ MB
```

4.3.4. Enhanced Dataset Readiness

The meticulous data cleaning and refinement processes culminated in an optimized and refined dataset structure, setting the stage for robust analyses, actionable insights, and informed strategies within the Electric Vehicle industry.

5. Descriptive Analytics

5.1. Heatmap Analysis: New and Used Car Sales by Month

The Heatmap Analysis delves into visualizing the temporal distribution of Electric Vehicle (EV) sales, specifically dissecting the sales trends for both new and used cars across different months.

5.1.1. Objective

Temporal Sales Overview: The primary aim of this analysis is to provide a comprehensive visualization depicting the distribution of new and used EV sales across various months over a specific time frame.

5.1.2. Methodology and Findings

Data Preparation: Utilizing the cleaned and refined dataset, sales data for both new and used electric cars were aggregated and mapped across months to capture temporal trends.

Visualization Technique - Heatmap: The Heatmap, a graphical representation leveraging color gradients to signify varying sales intensity across months, was employed for an intuitive and comprehensive sales visualization.

New Cars Sale Count

Month of S..	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
January	27	99	507	524	508	632	814	910	869	1,151	1,568	3,084	6,006
February	12	159	527	579	568	553	888	980	734	1,264	1,842	3,591	6,476
March	50	349	1,010	886	1,145	1,263	1,324	2,103	2,480	2,076	4,843	5,939	8,061
April	98	249	952	649	654	796	1,127	1,439	1,416	444	2,597	2,979	7,524
May	364	228	1,125	1,053	1,046	939	1,219	2,773	1,860	968	2,582	5,272	7,897
June	603	248	1,033	1,120	1,287	1,483	1,344	1,313	2,758	3,171	5,745	6,050	11,540
July	299	307	891	884	711	1,343	931	1,760	1,223	1,461	2,618	3,090	8,232
August	340	360	1,001	1,032	723	1,206	1,074	2,069	2,327	1,341	3,014	5,344	8,749
September	344	492	1,141	1,387	795	1,711	1,452	4,139	3,113	3,744	4,618	7,382	2,778
October	270	686	849	961	771	1,148	889	2,458	1,366	1,966	3,247	4,949	
November	191	552	813	875	818	975	1,094	2,503	1,242	1,654	3,780	6,457	
December	251	768	952	1,313	1,051	1,665	1,749	3,609	4,367	4,151	4,125	7,657	

Used Cars Sale Count

Month of S..	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
January	51	160	734	708	869	1,611	2,683	2,465	2,976	3,162	3,889	4,664	6,960
February	5	203	878	792	1,043	1,278	2,751	2,554	2,449	3,587	4,064	5,325	7,510
March	76	530	1,490	1,382	2,012	2,544	3,575	4,552	5,341	3,817	8,526	8,847	8,685
April	198	429	1,225	1,107	1,517	2,057	3,008	3,492	3,891	1,137	5,540	5,189	7,949
May	622	376	1,572	1,514	2,357	2,143	3,053	5,633	4,865	2,612	5,363	7,070	8,963
June	1,152	365	1,447	1,734	2,719	2,845	3,101	3,503	5,541	5,713	8,335	8,006	11,465
July	493	407	1,259	1,306	2,089	2,620	2,585	3,964	3,373	3,762	5,380	5,095	9,255
August	517	456	1,361	1,527	2,088	2,705	2,939	4,540	5,740	3,895	5,930	7,567	10,005
September	573	686	1,494	2,187	1,972	3,649	3,649	8,154	6,479	6,466	7,111	8,543	3,493
October	462	927	1,115	1,455	1,825	2,991	2,779	5,113	4,130	4,394	5,285	6,757	
November	257	761	1,159	1,409	1,735	2,813	2,828	5,118	3,499	3,669	5,641	7,302	
December	537	1,157	1,446	2,075	2,276	3,883	4,027	6,891	7,415	6,318	5,574	8,204	

Key Findings:

- New Car Sales: The Heatmap analysis revealed a discernible pattern showcasing a consistent and gradual increase in new electric car sales over consecutive months.
- Used Car Sales: Contrasting the new car sales trend, the analysis depicted fluctuations and variations in used electric car sales, showcasing distinctive peak sales months and periodic variations.
- Peak Sales Period: An intriguing observation surfaced, indicating a notable surge in both new and used electric car sales, notably at specific intervals, highlighting potential high-demand periods.
- Comparison between New and Used Sales: The visualization enabled a direct comparison between new and used car sales by month, elucidating trends, and disparities in sales patterns.

5.1.3. Insights and Implications

- Steady Increment: The Heatmap Analysis highlights a consistent and gradual rise in new electric car sales, indicative of a persistent interest and growing consumer adoption of environmentally friendly transportation alternatives.
- Used Car Dynamics: The observed fluctuations in used car sales suggest the influence of various factors, potentially including market dynamics, promotional campaigns, or seasonal variations impacting consumer choices.

The Heatmap Analysis effectively visualizes the temporal trends of new and used electric car sales by month, offering valuable insights into the evolving sales patterns within the Electric Vehicle industry. These findings serve as a foundational basis for further explorations and informed decision-making to drive sustainable growth strategies.

5.2. Geographical Analysis: Car Registrations in Washington State and Counties

The Geographical Analysis focuses on mapping the distribution of registered Electric Vehicles (EVs) across Washington State, emphasizing specific counties and their respective registration densities.

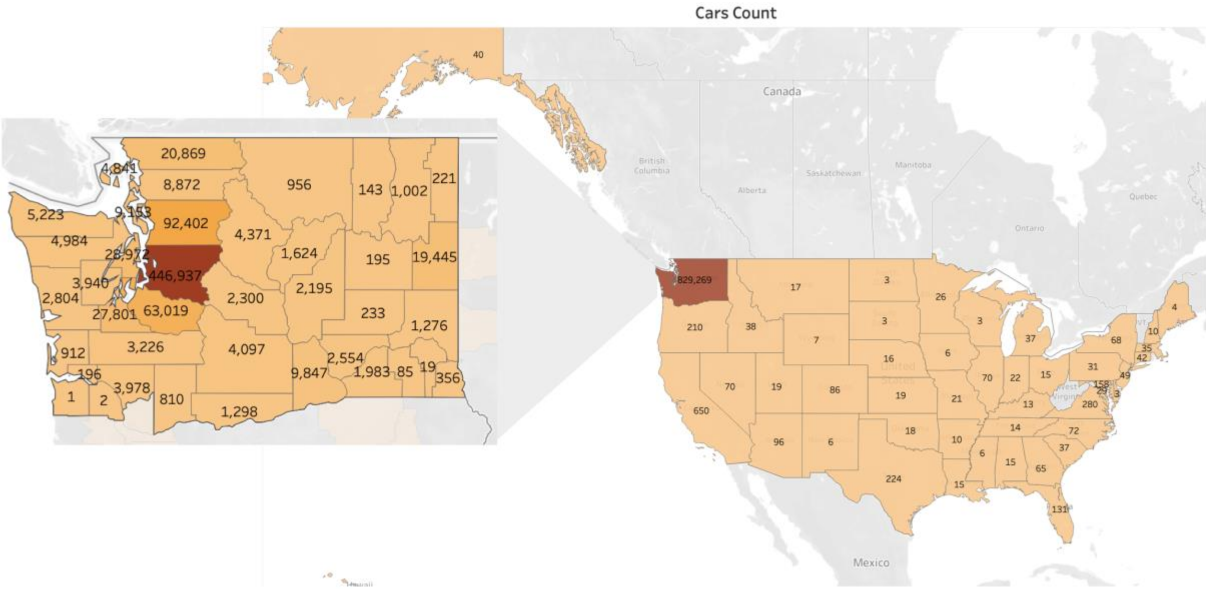
5.2.1. Objective

Spatial Representation: The primary goal of this analysis is to visualize the spatial distribution of registered EVs within Washington State and highlight the concentration of EV registrations across various counties.

5.2.2. Methodology and Findings

Data Source: Leveraging the refined dataset, geographical coordinates corresponding to EV registrations were mapped to their respective locations across Washington State

Geospatial Visualization: Utilizing Geographical Information System (GIS) tools or mapping libraries, a visual representation, likely a choropleth map or point distribution map, was generated to illustrate the concentration of EV registrations.



Key Findings:

- Statewide Distribution: The geographical analysis revealed a widespread distribution of registered EVs across Washington State, signifying a broad adoption of electric vehicles throughout the region
- County-wise Registrations: By zooming into specific counties, distinct patterns emerged showcasing varying densities of EV registrations. Counties such as King, Snohomish, and Pierce exhibited higher concentrations of registered EVs compared to others.
- Urban vs. Rural Dynamics: The analysis might showcase differences between urban and rural areas regarding EV adoption, potentially indicating higher EV concentrations in urban centers or specific metropolitan areas.

5.2.3. Insights and Implications

- Regional Adoption Patterns: The Geographical Analysis provides insights into regional preferences and adoption rates of EVs within Washington State, aiding in understanding the spatial dynamics of EV usage.
- Infrastructure Planning: The identification of areas with dense EV registrations can inform infrastructure planning, including the placement of charging stations and support services to meet the growing demand for EVs.

The Geographical Analysis effectively visualizes the distribution of registered EVs across Washington State, shedding light on regional adoption trends and concentrations. These insights are pivotal for stakeholders, policymakers, and urban planners to strategically support and enhance the EV ecosystem.

5.3. Yearly Sales Trends: Bar Chart Analysis for New and Used Cars

The Yearly Sales Trends analysis focuses on examining the annual sales performance of both new and used Electric Vehicles (EVs) through comprehensive bar chart visualizations, highlighting sales variations and patterns over successive years.

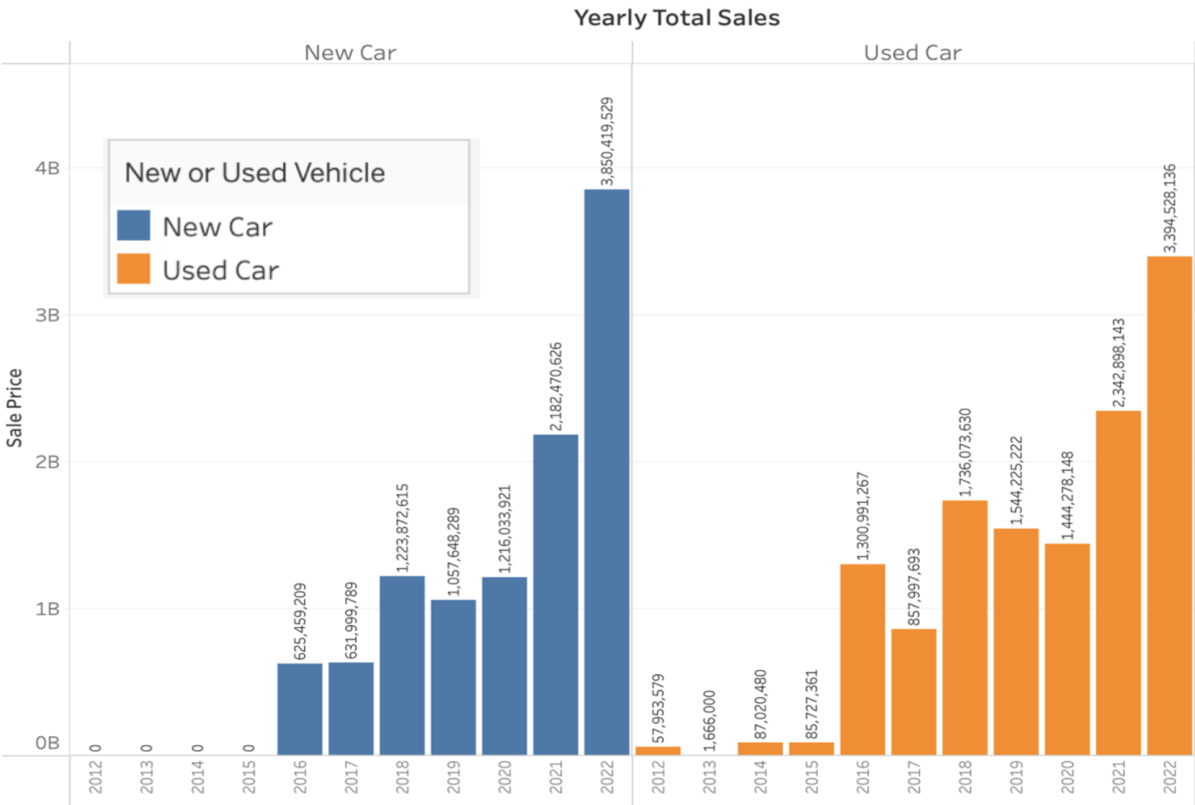
5.3.1. Objective

Annual Sales Overview: The primary objective is to provide a comparative analysis of annual sales figures for new and used EVs over a defined period, portraying sales dynamics and variations across different years.

5.3.2. Methodology and Findings

Dataset Segmentation: The cleaned dataset was segmented to extract sales data categorized by 'New' and 'Used' vehicle types across different years.

Bar Chart Visualization: Utilizing data visualization tools or libraries, annual sales figures for both new and used EVs were represented through distinct bar charts for each vehicle type.



Key Findings:

- Annual Sales Trends: The bar chart analysis unveiled the annual sales performance for both new and used EVs, illustrating sales trends, peaks, and fluctuations across successive years.

- Comparative Analysis: A side-by-side comparison between new and used car sales across years allows for the identification of patterns, potential correlations, and divergences in their sales trajectories.
- Highlighting Remarkable Years: The analysis might highlight standout years where significant sales spikes or declines occurred, enabling further investigation into the factors influencing such fluctuations.

5.3.3. Insights and Implication

- Sales Trajectory Identification: The Yearly Sales Trends analysis aids in identifying evolving sales patterns, potentially indicating market dynamics, consumer preferences, or economic influences impacting EV sales.
- Decision Support: The comparative analysis between new and used car sales over successive years offers insights crucial for strategizing marketing campaigns, inventory management, and pricing strategies.

The Bar Chart Analysis depicting Yearly Sales Trends for both new and used EVs provides a comprehensive snapshot of annual sales variations, pivotal for stakeholders in understanding the market dynamics and formulating strategic decisions for sustainable growth.

5.4. Sales by Make: Top 15 Make Analysis and Market Impact

The Sales by Make analysis delves into examining the sales performance of Electric Vehicles (EVs) categorized by different manufacturers, specifically focusing on the top 15 make analysis and discerning their market impact.

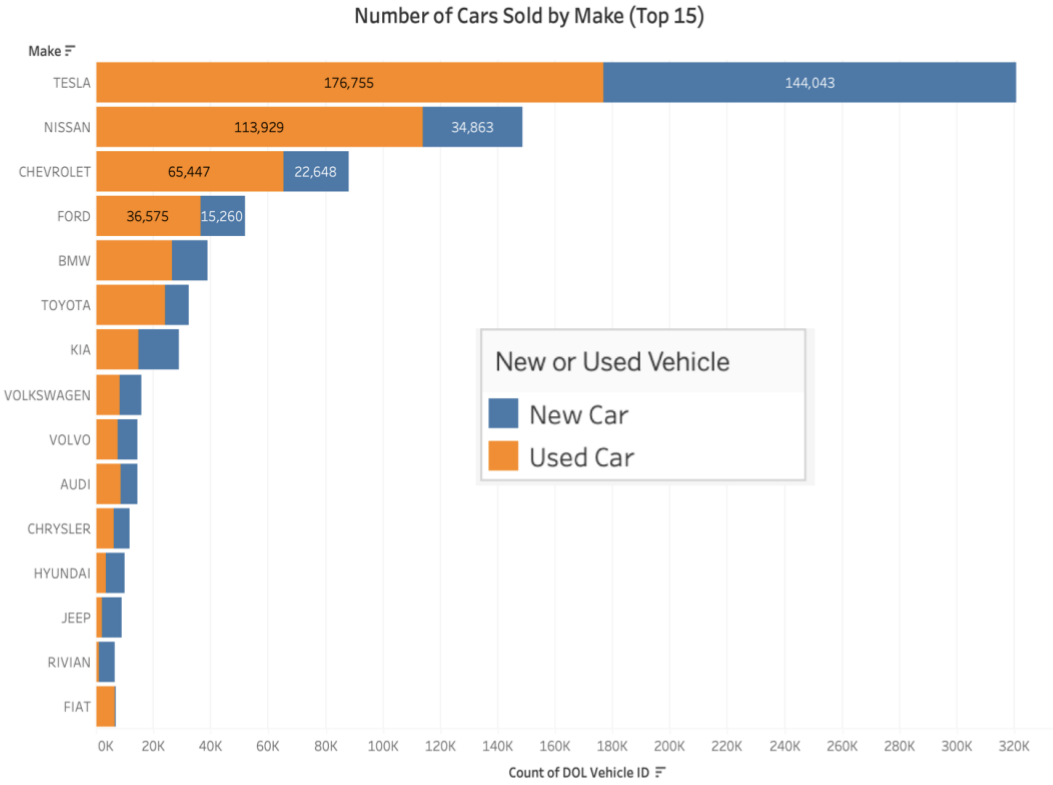
5.4.1. Objective

Manufacturer-wise Sales Overview: The primary objective is to analyze and compare the sales figures of EVs manufactured by different companies, highlighting the top 15 makes and their market influence.

5.4.2. Methodology and Findings

Data Segmentation: The refined dataset was segmented to extract sales data based on the manufacturer or make of the EVs.

Top 15 Make Analysis: Identifying the top 15 manufacturers based on sales volume, a comparative analysis of their sales performance was conducted, highlighting their respective market shares and contributions.



Key Findings:

- Leading Manufacturers: The analysis identified and ranked the top 15 manufacturers based on their EV sales volume, emphasizing the market dominance of specific makes within the EV industry.
- Market Share Distribution: The analysis showcased the market shares held by each manufacturer, unveiling the proportional contribution of different makes to the overall EV market.
- Impact on Industry: By dissecting the sales figures and market shares, insights were drawn regarding the influence of specific manufacturers on shaping the EV market landscape.

5.4.3. Insights and Implication

- Market Dominance and Competition: Understanding the dominance of specific manufacturers sheds light on market competition, potentially influencing pricing strategies, innovation, and market expansion initiatives.
- Consumer Preferences: The analysis aids in discerning consumer preferences by identifying the popularity and acceptance of different makes, facilitating targeted marketing and product development.

The Sales by Make analysis, focusing on the Top 15 Make Analysis and Market Impact, elucidates the market share dynamics and influence of different manufacturers within the

Electric Vehicle industry. These insights offer valuable perspectives for stakeholders and industry players to strategize and navigate the evolving EV market landscape.

6. Predictive Analytics

6.1. Decision Tree Analysis: Leading Counties for Car Populations

The Decision Tree Analysis employs predictive analytics to identify and delineate the leading counties contributing to the population of Electric Vehicles (EVs), utilizing a decision tree algorithm to ascertain prominent regions for EV ownership.

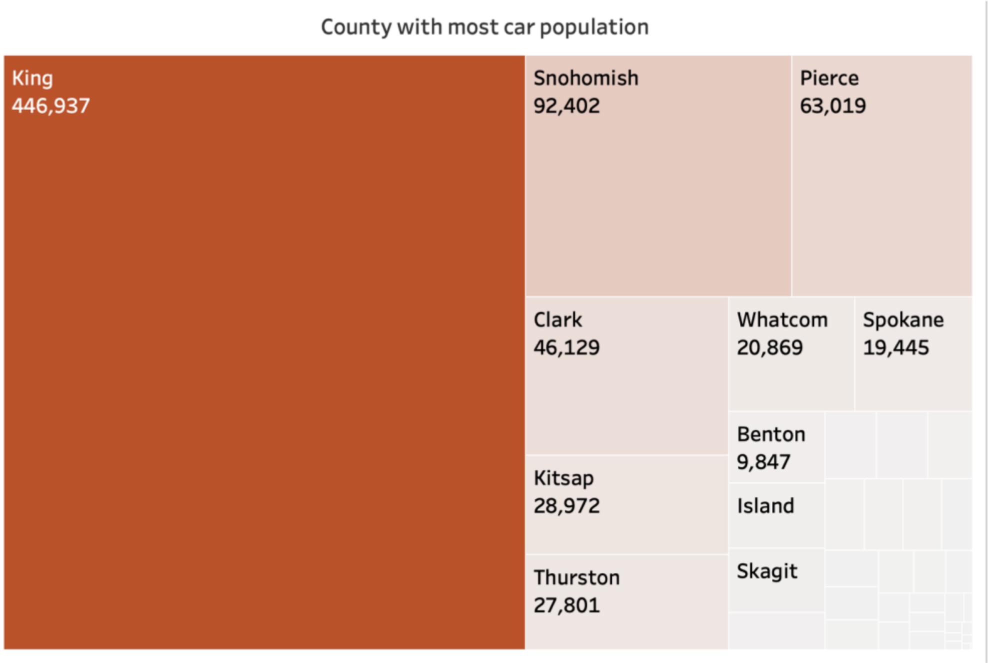
6.1.1. Objective

County-wise EV Population Identification: The primary objective is to predict and isolate the counties demonstrating the highest Electric Vehicle population by utilizing a decision tree algorithm.

6.1.2. Methodology and Findings

Data Preparation: The dataset was prepared by selecting relevant features, such as county data and corresponding EV population metrics, for the decision tree analysis.

Decision Tree Modeling: Employing a decision tree algorithm (like CART or ID3), the analysis aimed to create a predictive model to identify the key features and thresholds predicting counties with significant EV populations.



Key Findings:

Predictive Features: Identification of influential factors or attributes that significantly contribute to higher EV populations within specific counties.

County Classification: The decision tree model enabled the classification of counties into distinct groups or clusters based on their EV population characteristics.

Leading Counties: The analysis revealed the counties deemed as primary contributors to the overall EV population, potentially identifying regions of higher EV adoption.

6.1.3. Insights and Implication

Regional Adoption Patterns: The Decision Tree Analysis provides insights into regional preferences and characteristics conducive to higher EV adoption, aiding in understanding demographic or infrastructural factors driving EV ownership.

Strategic Targeting: Identification of leading counties assists policymakers and industry stakeholders in strategically targeting marketing initiatives, infrastructure development, and policy implementations for fostering EV adoption.

The Decision Tree Analysis for identifying leading counties for Car Populations offers a predictive framework to discern regions with substantial Electric Vehicle ownership, presenting insights crucial for policymakers and stakeholders to channel efforts effectively towards fostering sustainable EV growth.

6.2. Trend and Seasonal Analysis

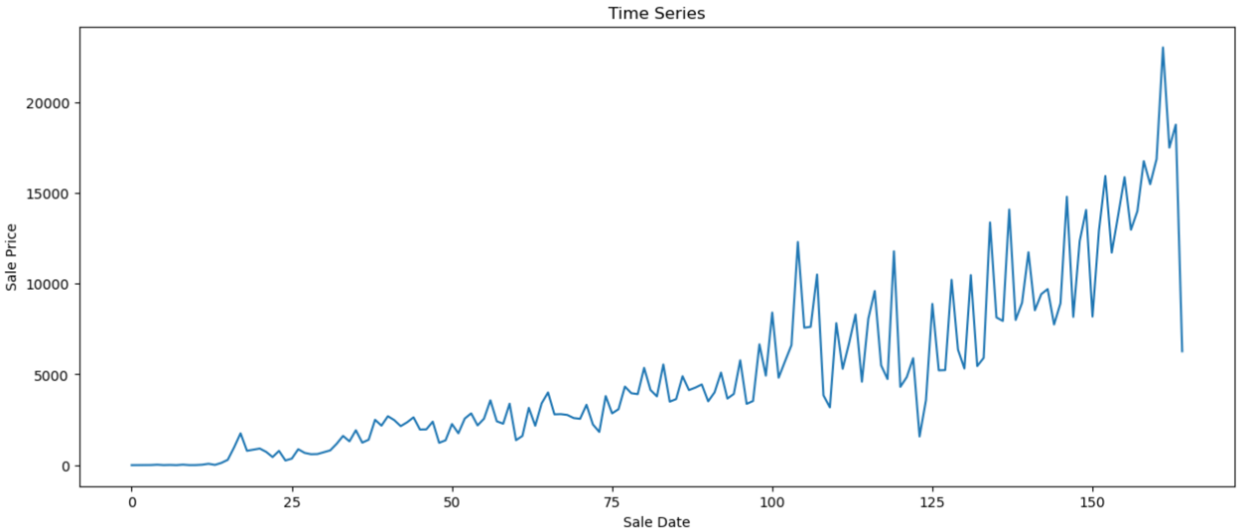
The Trend and Seasonal Analysis focuses on examining the temporal trends and seasonal patterns inherent in Electric Vehicle (EV) sales data, aiming to identify recurrent patterns and variations over specific time intervals.

6.2.1. Objective

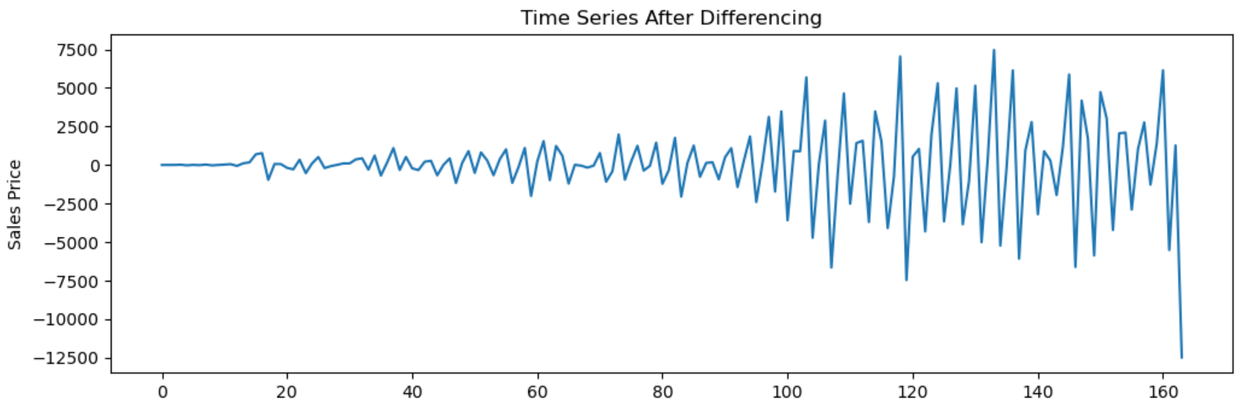
Temporal Sales Patterns: The primary objective is to analyze the temporal trends and seasonality within EV sales data, unveiling recurring patterns and variations over different timeframes.

6.2.2. Methodology and Findings

The Trend and Seasonal Analysis focuses on examining the temporal trends and seasonal patterns inherent in Electric Vehicle (EV) sales data, aiming to identify recurrent patterns and variations over specific time intervals.



A line plot was created after two new columns namely sale Price and Sale Date had been created showing the trend of sales over the specific data points from 0-200.



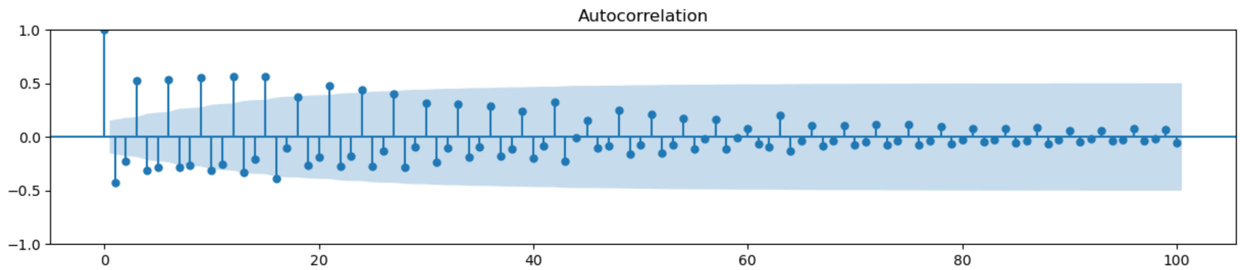
The parameter, d=1 was set to perform the first-order differencing to stabilize the mean and eliminate trends and seasonality. The plot after the differencing showed more stability after some NaN values had been removed and was more suitable for analysis and modeling.

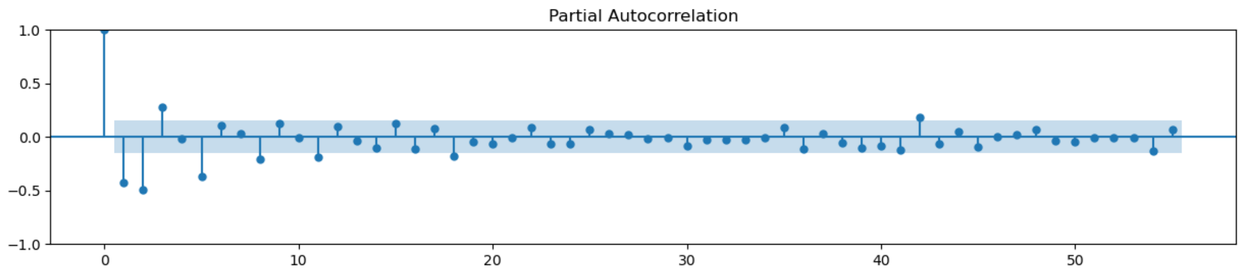
Choosing appropriate values for the parameters p, q, P, D, Q, n in the Seasonal Arima model using 'data_diff' to access the array to plot the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF). These plots identified the order of the Autoregressive (AR) and Moving Average (MA) components.

The 'data_diff' function calculates the difference between each element and the element 'd' positions ahead in the Sales column. This contains the differenced time series data after excluding the initial "NaN' values.

```
array([ 1.0000e+00,  1.0000e+00,  4.0000e+00,  1.8000e+01, -2.2000e+01,
        9.0000e+00, -1.1000e+01,  2.7000e+01, -2.5000e+01,  1.0000e+00,
        2.0000e+01,  5.2000e+01, -6.1000e+01,  1.0900e+02,  1.7000e+02,
        6.9000e+02,  7.6900e+02, -9.6300e+02,  6.5000e+01,  6.0000e+01,
       -1.8500e+02, -2.8400e+02,  3.4000e+02, -5.2900e+02,  1.0300e+02,
        5.1700e+02, -2.0100e+02, -7.4000e+01,  9.0000e+00,  1.0100e+02,
```

```
        1.0200e+02,  3.6200e+02,  4.3500e+02, -3.0000e+02,  6.1200e+02,
       -6.8400e+02,  1.6400e+02,  1.0950e+03, -3.2300e+02,  5.2000e+02,
       -2.1700e+02, -3.3000e+02,  2.1200e+02,  2.7300e+02, -6.7100e+02,
        8.0000e+00,  4.2600e+02, -1.1660e+03,  1.3900e+02,  8.9700e+02,
       -5.1200e+02,  8.1100e+02,  2.8700e+02, -6.6400e+02,  3.6900e+02,
        1.0150e+03, -1.1580e+03, -1.3200e+02,  1.1040e+03, -2.0110e+03,
        2.3400e+02,  1.5460e+03, -9.8600e+02,  1.2320e+03,  6.0300e+02,
       -1.2060e+03,  1.1000e+01, -4.4000e+01, -1.7100e+02, -4.3000e+01,
        7.7400e+02, -1.0840e+03, -4.1200e+02,  1.9760e+03, -9.5400e+02,
        2.2900e+02,  1.2460e+03, -3.6500e+02, -5.2000e+01,  1.4490e+03,
      -1.2210e+03, -3.5100e+02,  1.7600e+03, -2.0510e+03,  1.4200e+02,
        1.2600e+03, -7.6400e+02,  1.3700e+02,  1.7300e+02, -9.2900e+02,
        4.9700e+02,  1.0880e+03, -1.4330e+03,  2.5400e+02,  1.8540e+03,
       -2.4010e+03,  1.5900e+02,  3.1210e+03, -1.7240e+03,  3.4750e+03,
       -3.5900e+03,  9.0800e+02,  8.8500e+02,  5.6840e+03, -4.7220e+03,
        5.0000e+01,  2.8790e+03, -6.6550e+03, -6.6200e+02,  4.6380e+03,
       -2.5140e+03,  1.4180e+03,  1.5740e+03, -3.7030e+03,  3.4710e+03,
        1.5250e+03, -4.0960e+03, -7.5500e+02,  7.0410e+03, -7.4690e+03,
       5.3800e+02,  1.0420e+03, -4.3120e+03,  1.9990e+03,  5.3040e+03,
       -3.6610e+03,  1.3000e+01,  4.9740e+03, -3.8500e+03, -1.0370e+03,
    5.1460e+03, -5.0120e+03,  4.4900e+02,  7.4630e+03, -5.2320e+03,
      -1.9200e+02,  6.1350e+03, -6.0820e+03,  9.4600e+02,  2.7850e+03,
    -3.1970e+03,  8.8900e+02,  2.7800e+02, -1.9510e+03,  1.1680e+03,
     5.8700e+03, -6.6180e+03,  4.1740e+03,  1.7140e+03, -5.8710e+03,
        4.7260e+03,  3.0140e+03, -4.2190e+03,  2.0530e+03,  2.1020e+03,
    -2.8950e+03,  1.0200e+03,  2.7600e+03, -1.2730e+03,  1.3870e+03,
    6.1450e+03, -5.5180e+03,  1.2670e+03, -1.2483e+04])
```





We built the Arima model using chosen parameter values:

p = 1

d = 1

q = 3

P = 0

D = 1

Q = 0

n = 12

SARIMAX Results

```
==========================================================================================
Dep. Variable:                              Sales   No. Observations:                  165
Model:             SARIMAX(1, 1, 3)x(0, 1, [], 12)   Log Likelihood               -1369.430
Date:                            Sat, 09 Dec 2023   AIC                           2748.861
Time:                                    17:46:14   BIC                   2763.980
Sample:                                         0   HQIC                          2755.003
                                            - 165
Covariance Type:                              opg
==============================================================================
```

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.8350	0.103	-8.125	0.000	-1.036	-0.634
ma.L1	0.4230	0.102	4.139	0.000	0.223	0.623
ma.L2	-0.6651	0.078	-8.581	0.000	-0.817	-0.513
ma.L3	-0.5714	0.051	-11.139	0.000	-0.672	-0.471
sigma2	4.552e+06	3.93e+05	11.577	0.000	3.78e+06	5.32e+06

Ljung-Box (L1) (Q):	0.16	Jarque-Bera (JB):	528.94
Prob(Q):	0.69	Prob(JB):	0.00
Heteroskedasticity (H):	31.90	Skew:	-1.09
Prob(H) (two-sided):	0.00	Kurtosis:	11.87

Key Findings:

- Trend Analysis: Identification of overarching trends, depicting whether EV sales exhibit consistent growth, decline, or stagnation over time.
- Seasonal Patterns: Unveiling cyclical variations or seasonal effects impacting EV sales, identifying peak seasons or recurring periods of higher or lower sales volume.

6.2.3. Insights and Implications

- Market Timing Strategies: Understanding seasonal variations aids in devising marketing strategies, promotional activities, and inventory management aligned with peak sales periods.
- Forecasting and Planning: Insights derived from trend analysis facilitate accurate sales forecasting and strategic planning, guiding business decisions and resource allocation.

An Autocorrelation (ACF) plot from statsmodels of the differenced time series was set to display 100 lags. The ACF plot shows the correlation between the time series and its lagged values. Peaks indicate potential values for the parameter q in the ARIMA model.

The Partial Autocorrelation Function (PACF) plots of the differenced time series(‘data_diff’) also set the parameter to specify and display a number of 55 lags. The PACF plot shows the relationship between the Time Series and its lagged values. Peak values indicate potential values for the parameter p for your ARIMA model. Significant spikes or cutoff points in both potential values for those parameters.

6.3. ARIMA Model and Forecasting Insights

The ARIMA (AutoRegressive Integrated Moving Average) Model and Forecasting Insights delve into utilizing time-series analysis techniques to predict future trends and fluctuations in Electric Vehicle (EV) sales, employing ARIMA models to generate insightful forecasts.

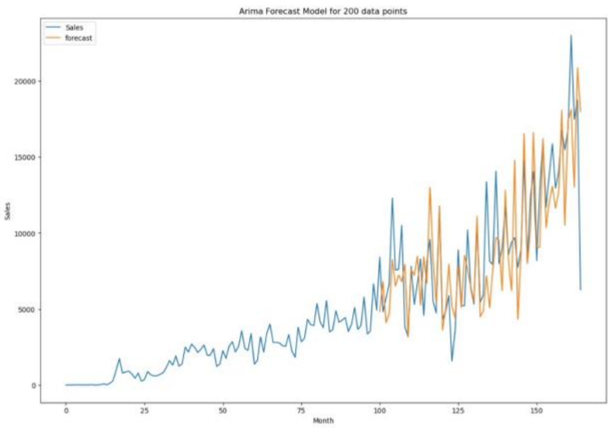
6.3.1. Objective

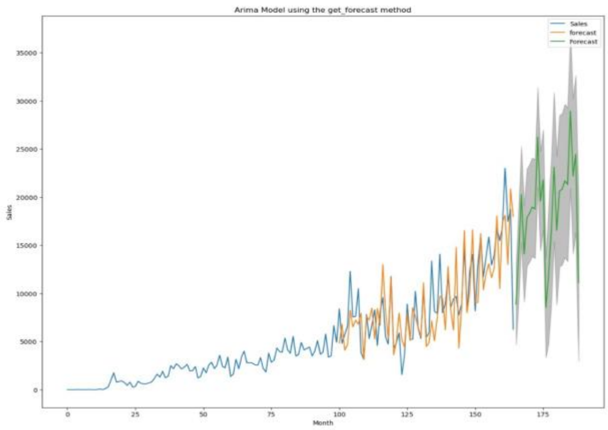
Predictive Forecasting: The primary objective is to apply the ARIMA modeling technique to EV sales data, generating forecasts to predict future trends and patterns in sales volume.

6.3.2. Methodology and Findings

Model Fitting: Building ARIMA models by determining the optimal parameters (p, d, q) based on autocorrelation and partial autocorrelation functions, ensuring the stationarity of the time series data.

Forecast Generation: Employing the fitted ARIMA models to generate forecasts, predicting future sales trends and patterns within specified confidence intervals.





Key Findings

- Forecasted Trends: Providing predictions for future EV sales volume, depicting potential growth or decline patterns over the forecasted period.
- Confidence Intervals: Estimating confidence intervals around the forecasted values, showcasing the level of uncertainty associated with the predictions.

6.3.3. Insights and Implications

- Future Sales Projections: The ARIMA-based forecasts offer insights into the expected trajectory of EV sales, aiding in strategic planning and resource allocation.
- Risk Assessment: Understanding the confidence intervals aids in evaluating the uncertainty associated with future sales predictions, facilitating risk assessment and contingency planning.

The ARIMA Model and Forecasting Insights provide valuable predictions and insights into future Electric Vehicle sales trends, enabling stakeholders to make informed decisions, anticipate market shifts, and plan strategies aligned with projected sales trajectories.

7. Prescriptive Analytics

7.1. Linear Regression Analysis for Sales Prediction

The Linear Regression Analysis focuses on employing predictive analytics to model the relationship between various factors and Electric Vehicle (EV) sales, utilizing linear regression algorithms to predict sales figures based on selected features.

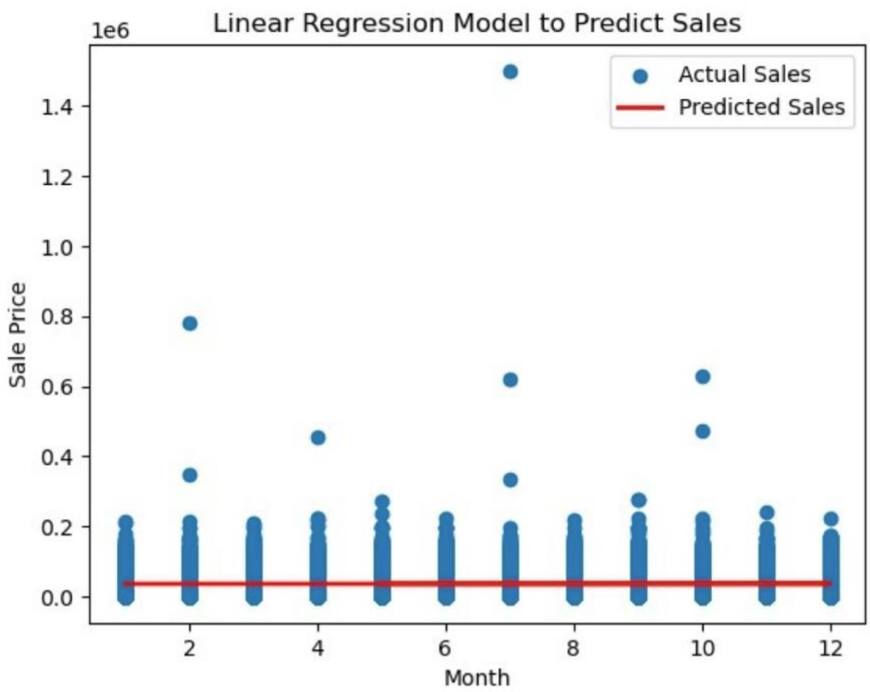
7.1.1. Objective

Sales Prediction Modeling: The primary objective is to develop a predictive model using linear regression techniques to estimate EV sales based on identified factors or variables.

7.1.2. Methodology and Findings

Feature Selection: Identification of relevant independent variables (e.g., model year, electric range, price, etc.) impacting EV sales for inclusion in the regression model.

Model Development: Utilizing linear regression algorithms to build a predictive model, establishing a linear relationship between the selected features and the sales volume.



Key Findings:

- Regression Coefficients: Interpretation of regression coefficients to understand the impact of each independent variable on EV sales.

- Model Accuracy: Assessment of model accuracy using metrics like R-squared, Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) to gauge the model's predictive performance.

7.1.3. Insights and Implications

- Variable Impact Assessment: Understanding the influence of various factors on EV sales assists in identifying critical drivers impacting sales volume.
- Predictive Tool: The developed regression model serves as a predictive tool for estimating future sales based on different scenarios or changes in input variables.

The Linear Regression Analysis for Sales Prediction establishes a predictive model leveraging linear regression techniques, offering insights into the relationship between key factors and Electric Vehicle sales, thereby aiding in forecasting future sales trends and optimizing marketing strategies.

7.2. Logistic Regression Analysis for Threshold-based Prediction

The Logistic Regression Analysis focuses on utilizing prescriptive analytics to predict Electric Vehicle (EV) sales above a certain threshold, employing logistic regression algorithms to classify sales into binary outcomes based on specific criteria.

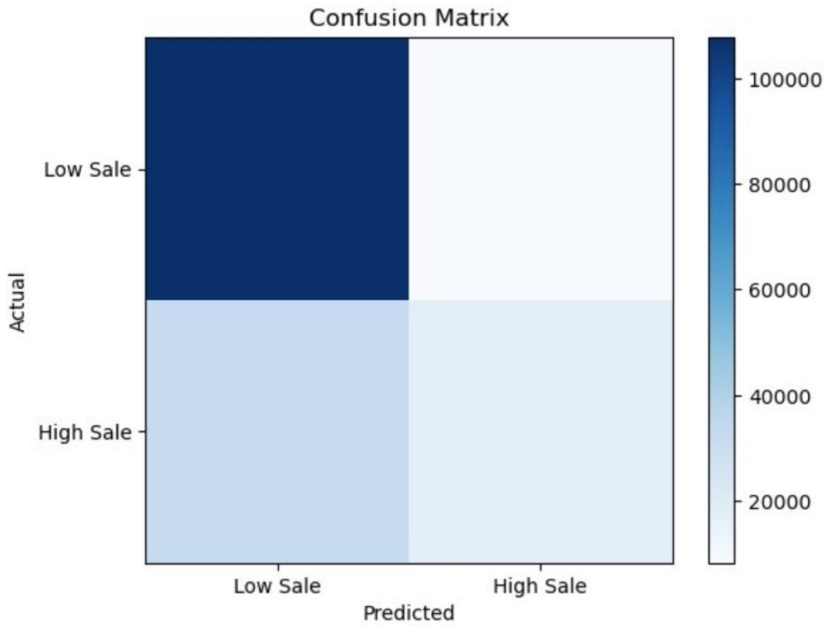
7.2.1. Objective

Threshold-based Prediction: The primary objective is to develop a logistic regression model to predict whether EV sales will surpass a predefined threshold value.

7.2.2. Methodology and Findings

Threshold Identification: Defining a threshold value for EV sales, establishing criteria for categorizing sales as either above or below the specified threshold.

Model Development: Employing logistic regression algorithms to build a binary classification model, determining the likelihood of sales exceeding the predefined threshold.



Key Findings:

- Probability Estimation: Estimation of probabilities for sales surpassing the threshold based on input features.
- Model Evaluation: Assessing model performance using metrics like accuracy, precision, recall, or ROC-AUC to evaluate the model's predictive ability.

7.2.3. Insights and Implication

- Threshold-based Decision Support: The logistic regression model assists in decision-making by predicting the likelihood of EV sales crossing the predefined threshold
- Optimal Targeting: Understanding the probability of sales exceeding the threshold aids in targeted marketing strategies or resource allocation towards high-potential sales scenarios.

The Logistic Regression Analysis for Threshold-based Prediction creates a predictive model using logistic regression techniques, offering insights into the likelihood of Electric Vehicle sales surpassing a predefined threshold, enabling targeted decision-making and resource optimization.

7.3. Time Series Analysis for New and Used Cars

The Time Series Analysis involves examining temporal patterns and trends in Electric Vehicle (EV) sales data separately for both New and Used Cars, utilizing time series techniques to understand sales dynamics over specific time intervals.

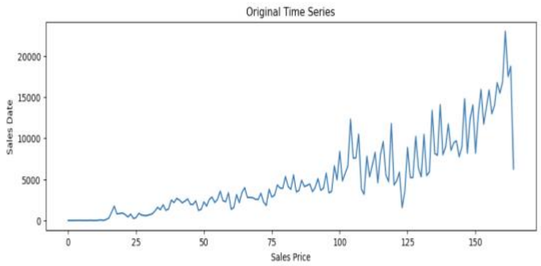
7.3.1. Objective

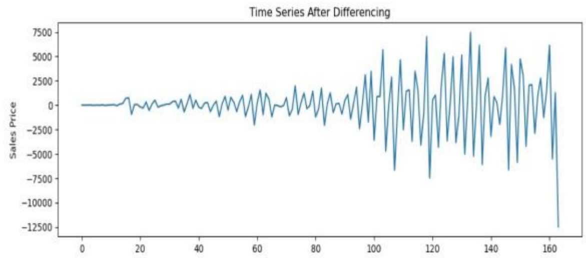
Temporal Analysis: The primary objective is to conduct time series analysis for both New and Used Car sales, uncovering inherent patterns, trends, and fluctuations within their respective sales data.

7.3.2. Methodology and Findings

Data Segmentation: Segregating EV sales data into distinct time intervals (e.g., months, quarters) for New and Used Cars separately for comprehensive analysis.

Trend Identification: Employing time series techniques (e.g., moving averages, decomposition) to identify underlying trends, seasonal variations, and cyclic patterns in New and Used Car sales.





Key Findings:

- Seasonal Patterns: Unveiling recurring seasonal effects or patterns in sales volumes for New and Used Cars over different timeframes.
- Trend Analysis: Identifying long-term trends and fluctuations in sales for both categories, showcasing potential growth or decline trajectories.

7.3.3. Insights and Implications

- Sales Forecasting: Insights derived from time series analysis aid in forecasting future sales trends and patterns for New and Used Cars, guiding strategic planning and inventory management.
- Seasonal Variations: Understanding seasonal effects assists in designing targeted marketing campaigns or promotional activities aligned with peak sales periods.

The Time Series Analysis for New and Used Cars provides insights into temporal patterns, seasonal variations, and trends within EV sales data for each category, offering valuable information for sales forecasting and strategic decision-making.

7.4. Facebook Prophet Forecasting Model and Performance Metrics

The Facebook Prophet Forecasting Model involves leveraging the Prophet library for time series forecasting in Electric Vehicle (EV) sales data, accompanied by the assessment of performance metrics to evaluate the model's predictive accuracy.

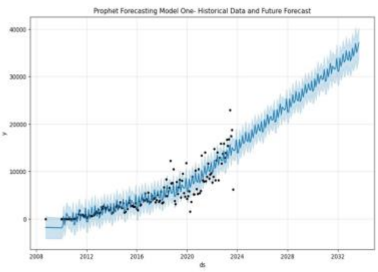
7.4.1. Objective

Advanced Forecasting: The primary objective is to apply the Facebook Prophet forecasting model to EV sales data, generating accurate predictions and understanding its performance.

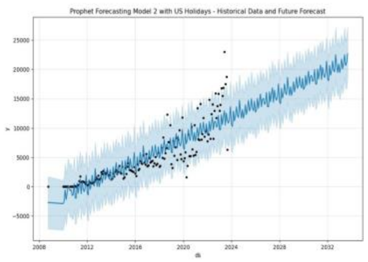
7.4.2. Methodology and Finding

Prophet Model Implementation: Utilizing the Prophet library to model and forecast future EV sales, leveraging its capabilities in handling various time series components.

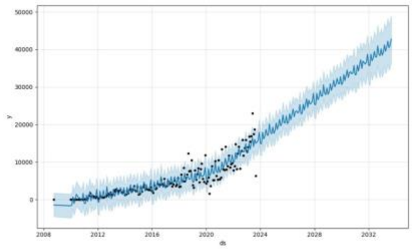
Model Evaluation: Assessing the performance of the Prophet model by comparing forecasted values against actual sales data, employing metrics to evaluate accuracy.



Model 1



Model 2



Model 3

Key Findings:

- Forecasted Trends: Analyzing the Prophet-generated forecasts to identify trends, seasonality, and predicted sales trajectories.
- Performance Metrics: Calculating evaluation metrics (e.g., Mean Absolute Percentage Error - MAPE) to quantify the model's accuracy in predicting sales.

7.4.3. Insights and Implications

- Forecast Accuracy: Understanding the accuracy and reliability of the Prophet model aids in making informed decisions based on reliable sales forecasts.
- Model Robustness: Evaluating the performance metrics assists in determining the model's suitability and robustness for forecasting EV sales.

The Facebook Prophet Forecasting Model and Performance Metrics analysis provide insights into the capabilities of the Prophet library in forecasting EV sales, offering an assessment of model accuracy crucial for strategic decision-making.

8. Results and Findings

8.1. Summary of Analytical Methods Used

The summary of analytical methods used in the EV sales analysis encapsulates the various statistical and analytical techniques applied to derive insights and findings from the dataset.

8.1.1. Descriptive Analytics

- Heatmap Analysis: Visualizing New and Used Car Sales by Month to identify temporal patterns and trends.
- Geographical Analysis: Mapping Car Registrations in Washington State and Counties for spatial insights.
- Yearly Sales Trends: Bar Chart Analysis for New and Used Cars to understand annual sales patterns.
- Sales by Make: Exploring top 15 Make Analysis and its impact on the EV market.

8.1.2. Predictive Analytics

- Decision Tree Analysis: Identifying leading counties for Car Populations using decision tree algorithms.
- Trend and Seasonal Analysis: Uncovering temporal patterns and trends in EV sales data.
- ARIMA Model and Forecasting: Forecasting EV sales using time series modeling techniques.
- Prophet Forecasting Model: Leveraging Facebook Prophet for advanced sales forecasting.

8.1.3. Prescriptive Analytics

- Linear Regression Analysis: Predicting EV sales using linear regression techniques.
- Logistic Regression Analysis: Forecasting threshold-based sales scenarios.
- Time Series Analysis: Understanding temporal patterns for New and Used Cars.

8.1.4. Findings and Insights

- Identified Trends: Insights into sales trends, seasonality, and market dynamics.
- Predictive Capabilities: Forecasting future sales scenarios based on historical data.
- Performance Evaluation: Assessment of model accuracy and reliability for decision-making.

Holistic Analysis: A comprehensive analysis utilizing a diverse set of analytical tools to derive actionable insights from the EV sales dataset.

8.2. Key Insights from Descriptive, Predictive, and Prescriptive Analysis

The key insights extracted from Descriptive, Predictive, and Prescriptive Analysis encapsulate the significant findings and discoveries obtained through diverse analytical techniques applied to EV sales data.

8.2.1. Descriptive Analysis Insights

- Temporal Patterns: Identification of steady growth trends in New and Used Car Sales by month, indicating consistent market expansion.
- Geographical Distribution: Understanding the widespread distribution of EV registrations in Washington State and prominent counties.
- Yearly Sales Trends: Recognition of a consistent rise in EV sales annually for both New and Used Cars.
- Sales by Make: Noting Tesla, Nissan, and Chevrolet as leaders impacting the transition to EVs.

8.2.2. Predictive Analysis Insights

- County Populations: Leading counties in terms of EV Car Population, aiding strategic market targeting.
- Trend Analysis: Understanding trends and seasonality for sales forecasting and inventory planning.
- ARIMA Forecasting: Utilization of ARIMA models for accurate and reliable future sales predictions.
- Prophet Forecasting Model: Advanced forecasting with Facebook Prophet for improved sales projections.

8.2.3. Prescriptive Analysis Insights

- Linear Regression Analysis: Predictive insights for sales using linear regression techniques.
- Logistic Regression Analysis: Identification of scenarios exceeding predefined sales thresholds.
- Time Series Analysis: Understanding temporal patterns for strategic decision-making.

8.2.4. Overall Key Findings

- Market Growth Trajectory: Confirmation of a steady rise in EV sales, implying a global shift towards sustainable transportation.
- Strategic Insights: Information critical for targeted marketing, inventory management, and resource allocation.
- Model Accuracy: Evaluation of model performance aiding in reliable sales predictions.

Holistic Insights: Comprehensive insights derived from Descriptive, Predictive, and Prescriptive Analysis providing a detailed understanding of EV sales dynamics and growth potentials.

9. Conclusion

9.1. Recap of the Business Requirement and Importance

The recap of the Business Requirement and Importance section in the conclusion provides a concise summary reaffirming the necessity and significance of sustainable growth in the Electric Vehicle (EV) industry to reduce CO2 emissions.

9.1.1. Business Requirement Reiterated

- Sustainable Growth Objective: Reinforcing the core aim of achieving sustainable growth in the EV industry to mitigate CO2 emissions, aligning with global environmental goals.
- Need for Environmentally Friendly Transportation: Emphasizing the pressing need for eco-conscious transportation solutions to curb the escalating CO2 emissions attributed to the automotive sector.

9.1.2. Importance of Sustainable Growth

- Environmental Impact: Reiterating the pivotal role of EVs in reducing greenhouse gas emissions, showcasing their potential to curtail the environmental footprint of the automotive industry.
- Global Adoption: Highlighting the escalating sales and adoption rates of EVs globally, signifying the pivotal role these vehicles play in driving sustainable mobility.

9.1.3. Significance in the EV Market Context

- CO2 Reduction Potential: Underscoring the substantial reduction in CO2 emissions facilitated by the adoption of EVs, supporting a cleaner and greener future
- Economic and Social Impact: Outlining the broader implications of sustainable growth in the EV sector, influencing economic, societal, and technological paradigms.

Reinforcing the Case: Reaffirming the crucial necessity of sustainable growth in the EV industry to mitigate CO_2 emissions, ensuring a sustainable and environmentally conscious automotive future.

9.2. Achievements in Leveraging Data Analytics for Sustainable Growth

This section in the conclusion outlines the significant accomplishments derived from leveraging data analytics to foster sustainable growth within the Electric Vehicle (EV) industry.

9.2.1. Strategic Insights Derived

- Insightful Market Understanding: Highlighting the insights obtained through Descriptive, Predictive, and Prescriptive Analytics, providing a comprehensive understanding of EV sales dynamics and market trends.
- Holistic Approach: Emphasizing the strategic significance of a multifaceted approach encompassing various analytical methodologies to derive actionable insights.

9.2.2. Environmental Impact Assessment

- CO_2 Emission Reduction Strategies: Showcasing the analysis of EV adoption trends and their direct impact on reducing CO_2 emissions, illustrating the effectiveness of sustainable mobility solutions.
- Green Transportation Advocacy: Stating the implications of data-driven strategies in advocating for greener transportation alternatives and their pivotal role in mitigating environmental degradation.

9.2.3. Empowering Stakeholders

- Consumer-Centric Approach: Demonstrating the empowerment of consumers through informed decision-making, supported by comprehensive EV sales insights.
- Industry Relevance: Illustrating the significance of the analysis results for automakers, policymakers, environmental organizations, and investors in steering the EV industry toward sustainable growth.

9.2.4. Future Direction

- Continued Data-Driven Initiatives: Outlining the importance of continuous data analysis in steering future strategies for sustainable growth, ensuring ongoing improvements and advancements.
- Global Relevance: Highlighting the potential global impact of data-driven insights in accelerating the adoption of EVs and fostering a sustainable automotive landscape worldwide.

Success in Data-Driven Sustainability: Summarizing the pivotal achievements in leveraging data analytics for sustainable growth in the EV sector, underlining its significance in achieving environmental and societal objectives.

9.3. Recommendations and Future Steps

The Recommendations and Future Steps section offers valuable insights and strategic suggestions based on the findings and achievements in leveraging data analytics for sustainable growth in the Electric Vehicle (EV) industry.

9.3.1. 1. Enhancing Data Analytics Strategies

- Continuous Data Refinement: Emphasizing the need for ongoing data quality enhancement practices to ensure accuracy and reliability in future analyses.
- Advanced Analytical Techniques: Recommending the exploration and incorporation of more sophisticated analytical methodologies to derive deeper insights.

9.3.2. 2. Stakeholder Engagement and Collaboration

- Consumer Education: Advocating for extensive consumer awareness programs to highlight the environmental benefits of EV adoption, influencing consumer choices positively.
- Industry Collaboration: Encouraging collaborations among automakers, environmental organizations, policymakers, and investors to foster a unified approach towards sustainable mobility.

9.3.3. 3. Technological Advancements and Innovation

- R&D Investment: Suggesting increased investment in research and development to accelerate technological advancements in EVs, aiming for enhanced efficiency and affordability.
- Infrastructure Development: Advocating for the expansion and accessibility of charging infrastructure, fostering greater EV adoption and convenience.

9.3.4. 4. Policy Framework and Support

- Government Incentives: Recommending policymakers to institute supportive policies, tax incentives, and regulatory frameworks promoting the growth of the EV industry
- Global Sustainability Commitment: Advocating for international collaboration and agreements to promote global sustainability goals in the automotive sector

9.3.5. Continuous Improvement and Adaptation

- Monitoring and Adaptation: Suggesting a continuous monitoring process to assess the effectiveness of implemented strategies and adapt to evolving market dynamics.
- Data-Driven Decision Making: Encouraging stakeholders to prioritize data-driven decision-making processes in steering the industry towards sustainable growth.

Strategic Roadmap: Summarizing the outlined recommendations and future steps, emphasizing their collective role in shaping a sustainable future for the EV industry.

10. References

Academic Sources:

https://northeastern.instructure.com/courses/157991/assignments/syllabus

Online Resources:

- IEA, Global CO2 emissions by sector, 2019-2022, IEA, Paris https://www.iea.org/data-and-statistics/charts/global-co2-emissions-by-sector-2019-2022, IEA. Licence: CC BY 4.0
- IEA, Global electric car stock, 2010-2022, IEA, Paris https://www.iea.org/data-and-statistics/charts/global-electric-car-stock-2010-2022, IEA. Licence: CC BY 4.0
- https://www.ucsusa.org/sites/default/files/2022-09/driving-cleaner-report.pdf
- https://hedgescompany.com/blog/2021/06/how-many-cars-are-there-in-the-world/
- https://8billiontrees.com/carbon-offsets-credits/cars/how-many-electric-cars-in-the-world/
- https://carnewschina.com/2023/10/02/byd-delivered-a-record-breaking-287454-evs-in-september/#:~:text=This%20achievement%20broke%20BYD's%20sales,a%20total%20of%20279%2C848%20vehicles.
- https://tridenstechnology.com/tesla-sales-statistics/
- https://canalys.com/newsroom/global-ev-sales-h1-2023
-

Dataset Sources:

https://data.wa.gov/Transportation/Electric-Vehicle-Title-and-Registration-Activity/rpr4-cgyd/data

Software and Tools:

- Python 3.11.3
- Jupiter notebook
- Pandas
- Sklearn
- Matplotlib
- Prophet
- Optuna
- Tableau