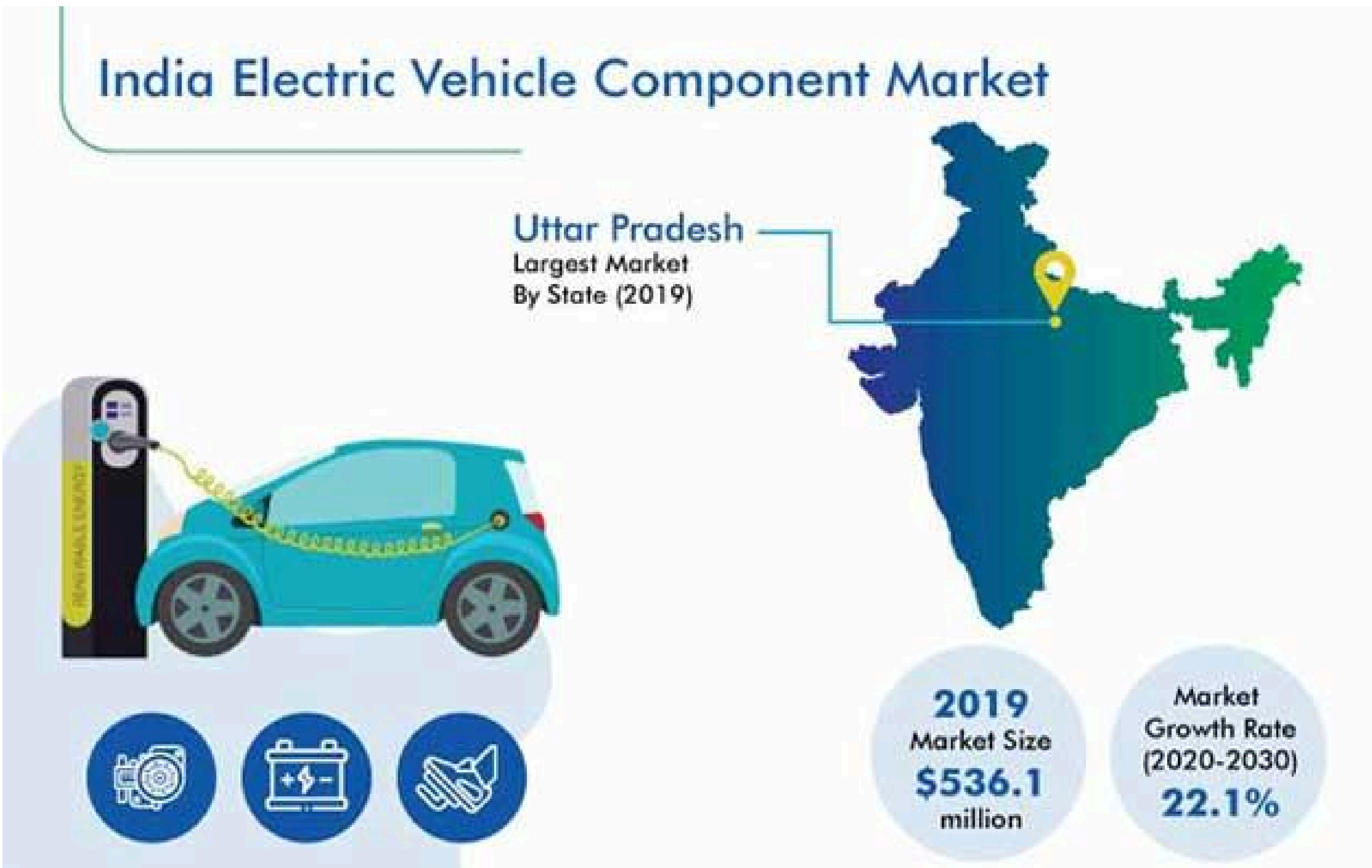
Market Analysis EV market in India

GitHub Link : https://github.com/nidhi-158/FeynnLabs_EV-market

Report by: Nidhi Patel



Report by: Nidhi Patel

Overview:
Electric vehicles (EVs) could account for more than 40% of India's automotive market and generate over $100 billion of revenue by 2030.

> Achieving this will require concerted strategies across five areas: new product development, go-to-market/distribution, customer segment prioritization, software development, and charging infrastructure.
>
> Several of these interventions will require category-specific stakeholder action.

Executive summary:

India's electric vehicle (EV) sector is growing rapidly, driven by government incentives, environmental concerns, and technological advancements. The FAME scheme aims to boost EV adoption, pushing India towards sustainable transportation.

By 2030, India targets 30% EV sales in private cars, 70% in commercial vehicles, 40% in buses, and 80% in two- and three-wheelers, aiming for 80 million EVs. The 'Make in India' initiative supports domestic production.

In 2023, global EV market value was $255.54 billion, projected to reach $2,108.80 billion by 2033, with a 23.42% CAGR. In India, EV sales grew by 49.25% in 2023 to 1.52 million units. The market is expected to grow from $3.21 billion in 2022 to $113.99 billion by 2029, with a 66.52% CAGR.

India's EV battery market is set to grow from $16.77 billion in 2023 to $27.70 billion by 2028. As of February 2024, there are 12,146 public EV charging stations, with Maharashtra leading. A CII report calls for 1.32 million charging stations by 2030, requiring 400,000 installations annually.

In this paper, I analyze electric vehicle (EV) sales based on individual demographics such as age and salary, as well as specific vehicle features including price (in Euros), number of seats, plug type, rapid charge capability, segment, model, and more. Additionally, the paper examines the distribution of charging stations across states, categorized by different types of vehicles.

Title: Analysis of Electric Vehicle (EV) Sales: A Comprehensive Study of Four-Wheeler EVs and Comparative Analysis with Other EV Categories

Abstract: This research paper presents a detailed analysis of the sales dynamics of four-wheeler electric vehicles (EVs), alongside other categories of EVs, including two-wheelers, three-wheelers, and buses. The study explores the influence of various demographic and geographic factors on EV sales, focusing on individual characteristics such as age, salary, and geographic location.

Introduction: The transition to electric vehicles represents a significant shift in the automotive industry, driven by environmental concerns and technological advancements. This research aims to provide insights into the sales trends of four-wheeler EVs compared to other EV categories and to understand the underlying factors affecting these sales.

Methodology: The analysis utilizes a comprehensive dataset encompassing sales data for different categories of EVs. Factors such as individual age, salary, and geographic location are examined to determine their impact on EV sales. Advanced statistical and data analysis techniques are employed to uncover patterns and trends.

Results: The study reveals notable trends in the sales of four-wheeler EVs, highlighting differences compared to two-wheelers, three-wheelers, and buses. The analysis indicates how demographic factors like age and salary, as well as geographic location, influence purchasing decisions and sales performance.

Discussion: The findings offer valuable insights into consumer behavior and market dynamics. The impact of age, salary, and geographic factors on EV sales is discussed, providing implications for policymakers, manufacturers, and marketers aiming to optimize their strategies in the growing EV market.

Conclusion: This research underscores the importance of understanding demographic and geographic influences on EV sales. The comparative analysis of four-wheeler EVs with other EV categories offers a comprehensive view of market trends and consumer preferences.

Keywords: Electric Vehicles, Four-Wheeler EVs, Sales Analysis, Demographic Factors, Geographic Location, Market Trends

Data

Behavioral Data: (https://www.kaggle.com/datasets/srinrealyf/india-ev-market-data)
Geographical Data:
https://www.kaggle.com/datasets/sanhitasaxena/indian-electric-vehicle-dataset

Vehicle Data
(https://www.kaggle.com/datasets/praveenchoudhary1217/electric-vehicle-sales-in-india)
EV Sales Data
https://www.kaggle.com/datasets/praveenchoudhary1217/electric-vehicle-sales-in-india

EDA

In the data preparation phase, we perform the following steps to ensure that our dataset is optimized for analysis:

1. **Column Selection**:
   - We conduct a feature selection process to retain only the relevant columns necessary for our analysis. This involves dropping columns that do not contribute to our analysis or that might introduce noise into the dataset. In this case, we focus on the `State Name` and `Four Wheeler` columns as they are pertinent to our analysis.
2. **Encoding Categorical Variables**:
   - To prepare categorical data for analysis, we employ Label Encoding. This technique transforms categorical variables into numerical values. Specifically, the `State Name` column is encoded into a format suitable for computational models. Label Encoding assigns a unique integer to each category, facilitating the integration of categorical variables into numerical computations.
3. **Feature Scaling**:
   - To ensure that all features contribute equally to the analysis, we apply Standard Scaling to the numerical variables. Standard Scaling, achieved using `StandardScaler`, normalizes the data by removing the mean and scaling to unit variance. This preprocessing step standardizes the `Four Wheeler` values, which is essential for algorithms that are sensitive to the scale of input features.

Linear Regression:

This project aims to create a predictive tool for estimating car prices in Euros based on various vehicle attributes. The tool will support manufacturers, dealers, and consumers in making informed pricing and purchasing decisions.

Key Features

The dataset includes the following attributes:

- **Brand**: Vehicle manufacturer.
- **Model**: Specific vehicle model.
- **AccelSec**: Acceleration time from 0 to 100 km/h.
- **TopSpeed_KmH**: Maximum speed.
- **Range_Km**: Distance per charge (for electric vehicles).
- **Efficiency_WhKm**: Energy efficiency.
- **FastCharge_KmH**: Fast charging rate.
- **RapidCharge**: Rapid charging capability.
- **PowerTrain**: Type of powertrain (e.g., electric, hybrid).
- **PlugType**: Charging plug type.
- **BodyStyle**: Vehicle body design (e.g., sedan, SUV).
- **Segment**: Market segment (e.g., luxury, economy).
- **Seats**: Number of seats.
- **PriceEuro**: Vehicle price in Euros (target variable).

Benefits and Applications

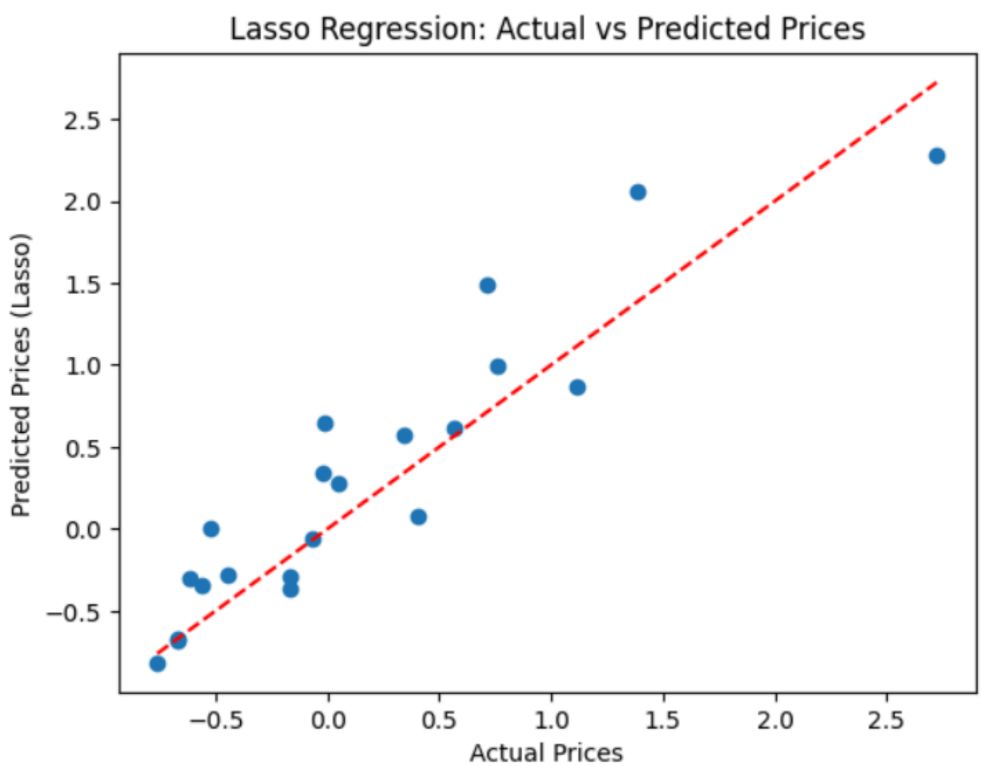
- **Manufacturers**:
  - **Pricing Strategy**: Optimize pricing for new models and align with market expectations.
  - **Product Development**: Tailor features and specifications based on price predictions.
- **Dealers**:
  - **Competitive Pricing**: Set competitive prices and adjust based on predicted vehicle values.
  - **Marketing Strategies**: Design targeted marketing campaigns based on pricing insights.
- **Consumers**:
  - **Informed Decisions**: Make better purchasing decisions with clear price predictions relative to vehicle features.

Applications

1. **Pricing Optimization**:
    - Adjust vehicle prices dynamically based on predicted values to maximize sales and profitability.
2. **Market Analysis**:
    - Analyze trends and pricing patterns across different vehicle segments and attributes.
3. **Strategic Marketing**:
    - Develop marketing strategies and promotions informed by predictive pricing data.
4. **Product Positioning**:
    - Align vehicle features and pricing with consumer preferences and market demand.

By implementing this predictive tool, stakeholders will gain valuable insights into vehicle pricing, enhancing decision-making and market strategy effectiveness.

Here we have plotted the regression that shows the predicted price for the ev vehicle to purchase based on various features.



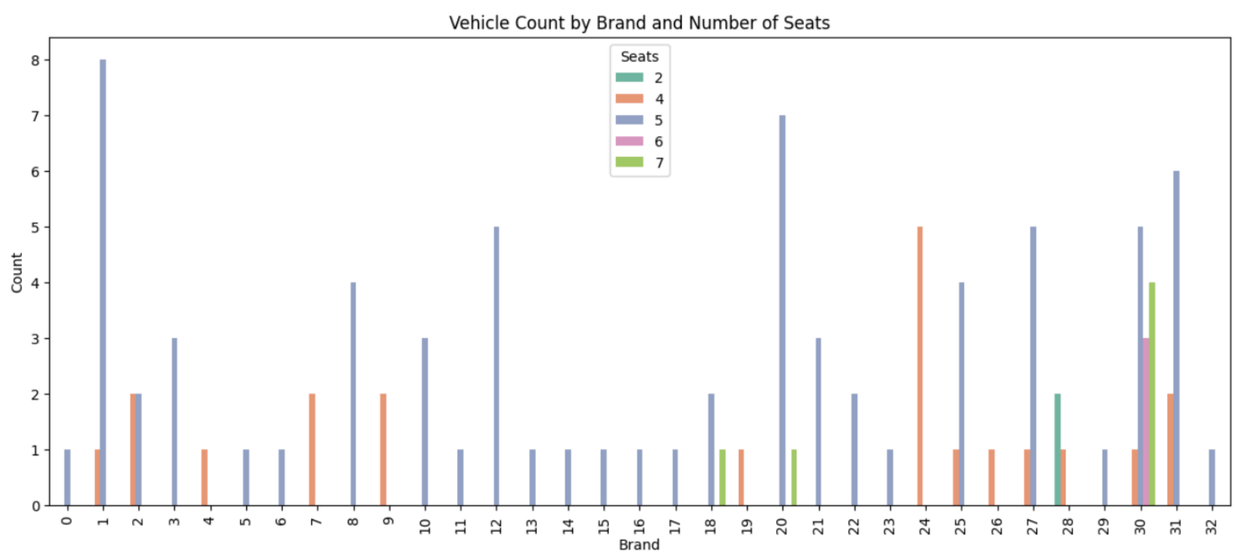
Visualization:

Analysis of Vehicle Distribution by Seats and Brand

In the following analysis, we present a graphical representation of the relationship between vehicle brands and the number of seats for four-wheelers. The graph illustrates the distribution of vehicles across different seat configurations, categorized by their respective brands.

Key Findings:

- **Vehicle Distribution by Seats**: The graph provides a detailed view of how the number of seats varies across different vehicle brands. It highlights the count of vehicles available for each seat configuration, allowing us to understand the variety and distribution of vehicle types based on seating capacity.
- **Brand Segmentation**: Each vehicle brand is represented, showcasing how it contributes to the overall distribution of vehicles with various seat configurations. This segmentation helps in identifying which brands offer more diversity in seating options.



Plug Type Distribution Analysis

The graph presents the distribution of various plug types used in electric vehicles, including:

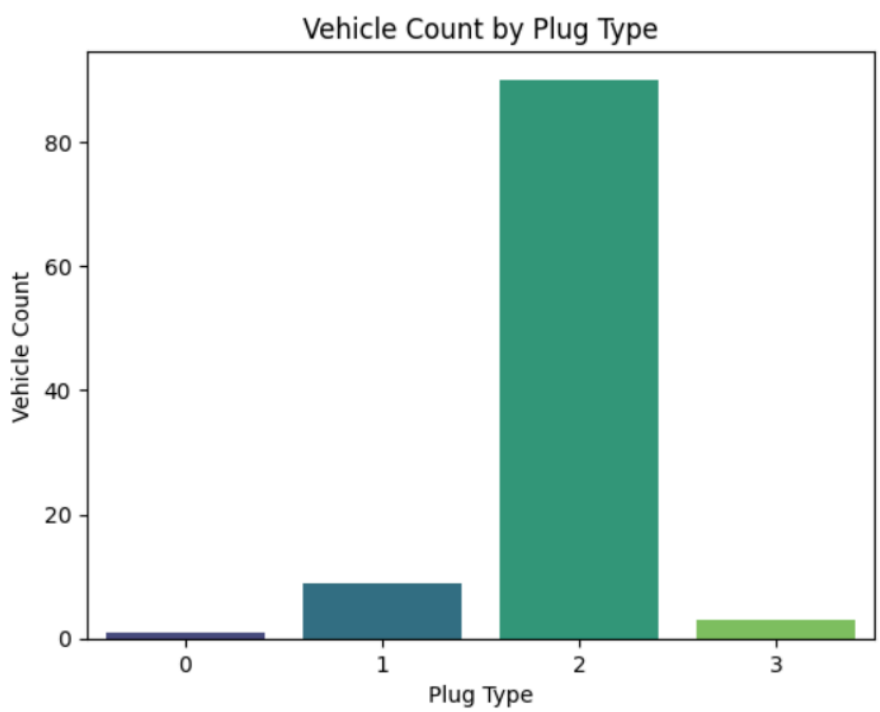
- **Type 2 CCS**
- **Type 2 CHAdeMO**
- **Type 2**
- **Type 1 CHAdeMO**

The values have been standardized using Standard Scaling, which normalizes the data for accurate comparison.

Key Points:

- **Normalized Data**: Standard Scaling has been applied to ensure all plug types are on a comparable scale.
- **Distribution Insight**: The graph shows how frequently each plug type appears in the dataset.

Most of the EV 4 wheelers are of PlugType 2 that is (Type 2).



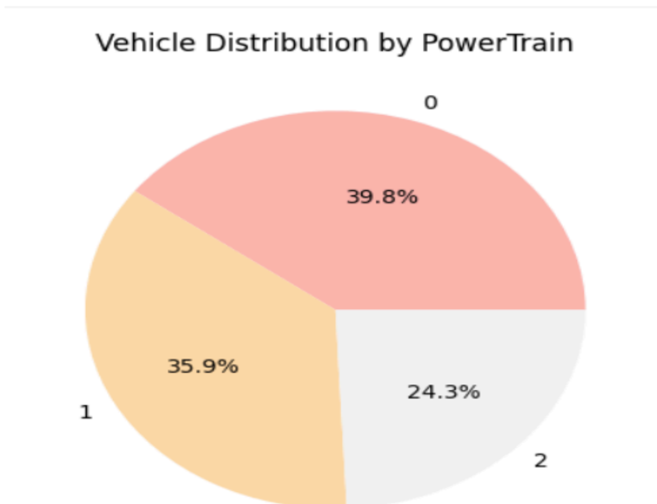
Analysis of PowerTrain Distribution

The pie chart below illustrates the distribution of four-wheel vehicles by their powertrain types, categorized into:

- **AWD (All-Wheel Drive)**
- **FWD (Front-Wheel Drive)**
- **RWD (Rear-Wheel Drive)**

Summary:

The distribution clearly shows that AWD is the predominant powertrain type for four-wheel vehicles, followed by FWD. RWD is the least common among the three powertrain configurations. This distribution provides insights into market preferences and the prevalence of different powertrain systems in the dataset.

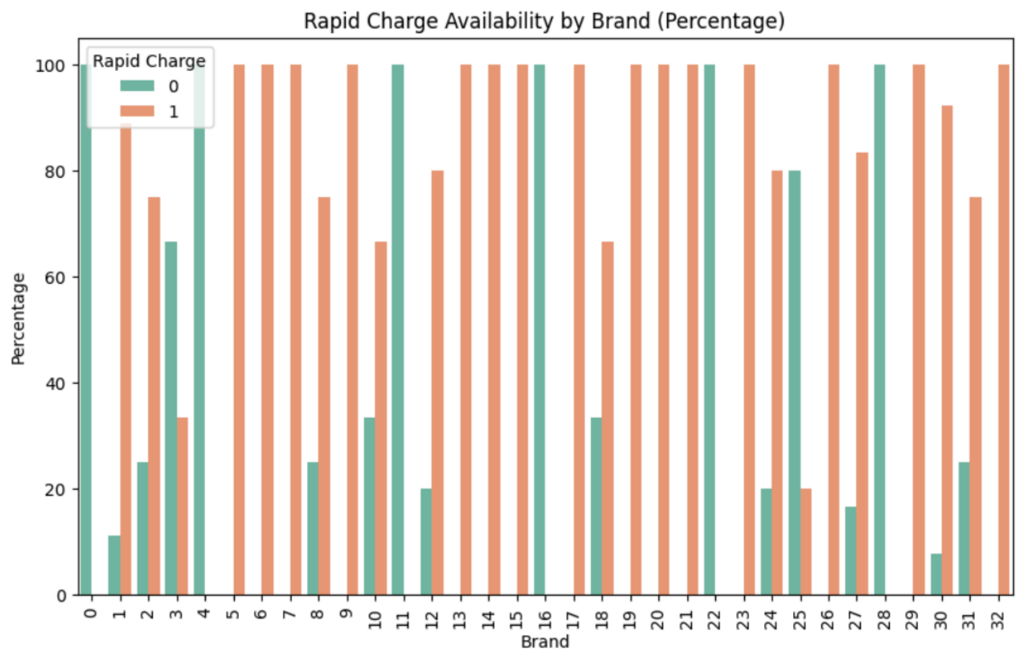


Rapid Charging Facility Availability by Brand

Analysis of rapid charging facilities across various brands indicates that:

- **75%** of the brands provide rapid charging capabilities.
- **25%** of the brands do not offer rapid charging facilities.

This distribution highlights that a significant majority of brands are equipped to meet rapid charging needs, whereas a smaller fraction does not currently support this feature.



Clustering Analysis

Data Preparation

- **Loading and Cleaning**: Import and clean the dataset.
- **Categorical Encoding**: Convert categorical variables to numeric using `LabelEncoder`.
- **Feature Scaling**: Standardize numeric features with `StandardScaler`.

PCA (Principal Component Analysis)

- **Dimension Reduction**: Reduce features to 2 principal components to facilitate clustering.
- **Explained Variance**: Plot cumulative explained variance to determine how much information is retained.

Clustering

- **Elbow Method**: Determine optimal clusters by plotting Within-Cluster Sum of Squares (WCSS) against the number of clusters.

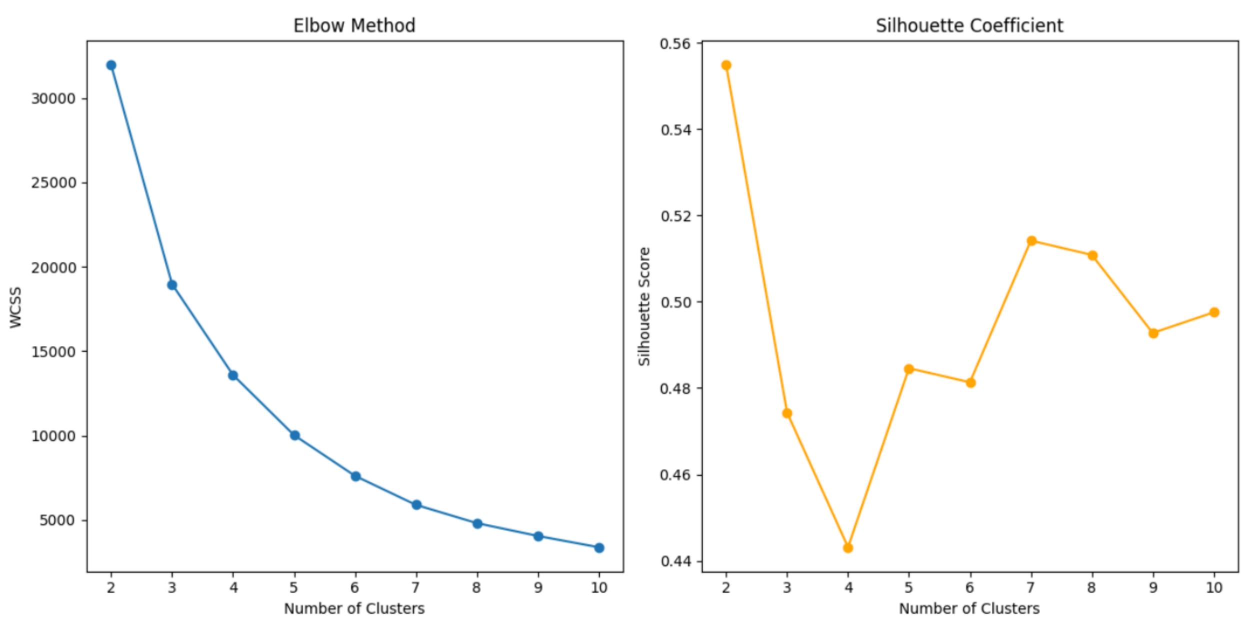
- **Silhouette Score**: Evaluate clustering quality by measuring how well-separated clusters are.

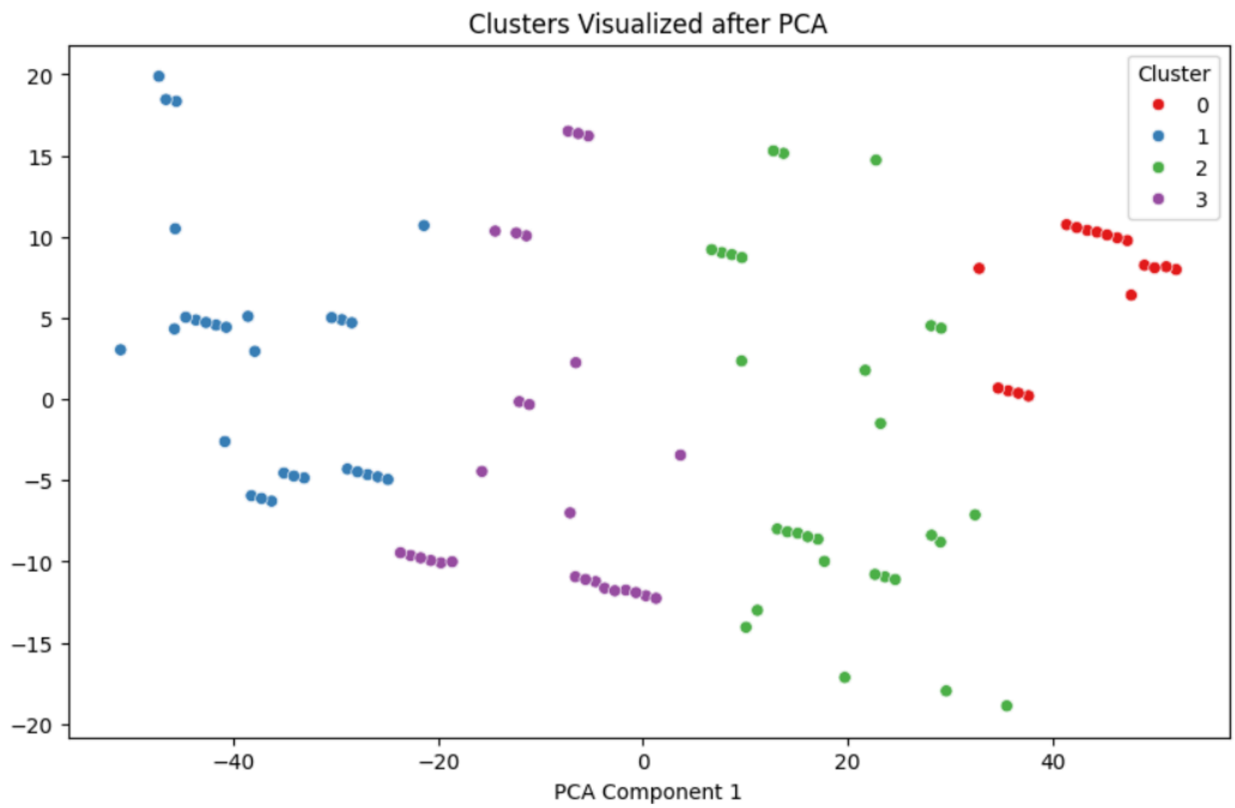
Visualization

- **PCA Visualization**: Scatter plot of clusters based on PCA components.

Silhouette Score Summary

- **Purpose**: Measures how well each point is clustered.
- **Calculation**:
  - **Intra-cluster Distance**: Average distance to points in the same cluster.
  - **Nearest-cluster Distance**: Average distance to points in the nearest cluster.
- **Scores**:
  - **+1**: Well-clustered point.
  - **0**: Point on cluster boundary.
  - **-1**: Misclassified point.
- **Use**: Assesses cluster quality and helps choose the optimal number of clusters.





Behavioral data

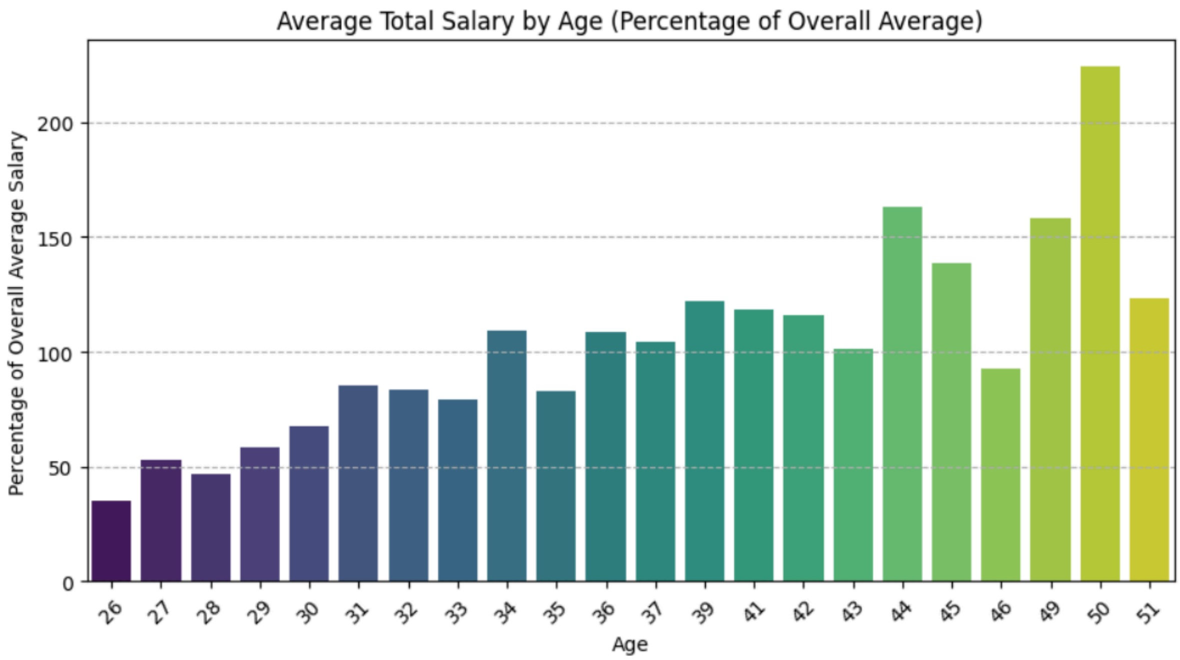
Dataset Columns

The dataset includes the following columns:

1. **Age**: The age of the individual.
2. **Profession**: The individual's current profession (job or business).
3. **Marital Status**: The individual's marital status (e.g., single, married).
4. **Education**: The highest level of education attained by the individual.
5. **Number of Dependents**: The total number of dependents the individual is responsible for.
6. **Personal Loan**: Indicator of whether the individual has a personal loan (e.g., Yes/No).
7. **Total Salary**: The individual's total annual salary.
8. **Price**: The price of the EV vehicle.

Age vs. Salary Analysis

The graph illustrates the relationship between age and salary. The observed trend indicates that salary generally increases with age. This suggests a positive correlation between an individual's age and their salary, where older individuals tend to have higher salaries.



1. Logistic Regression Results

- **Accuracy**: The Logistic Regression model achieved an accuracy of 65%. This indicates that 65% of the predictions made by the model were correct. In other words, the model was able to correctly predict whether an individual would obtain a personal loan 65% of the time based on their total salary.
- In 65% of cases a person gets a loan based on his/her salary to purchase an EV vehicle.

```python
# Import necessary libraries
from sklearn.metrics import mean_squared_error, mean_absolute_error, accuracy_score, r2_score

X=df[['Total Salary']]
y=df['Personal loan']

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=42)

model=LogisticRegression()
model.fit(X_train,y_train)

y_pred=model.predict(X_test)

# Calculate metrics using the imported functions, use a different variable name than the function name
mse  = mean_squared_error(y_test,y_pred)
mae = mean_absolute_error(y_test,y_pred)

print("Mean Squared Error:", mse)
print("R2 Score:", r2_score(y_test,y_pred))
print("Accuracy:", accuracy_score(y_test,y_pred))
```

```
Mean Squared Error: 0.35
R2 Score: -0.5384615384615388
Accuracy: 0.65
```

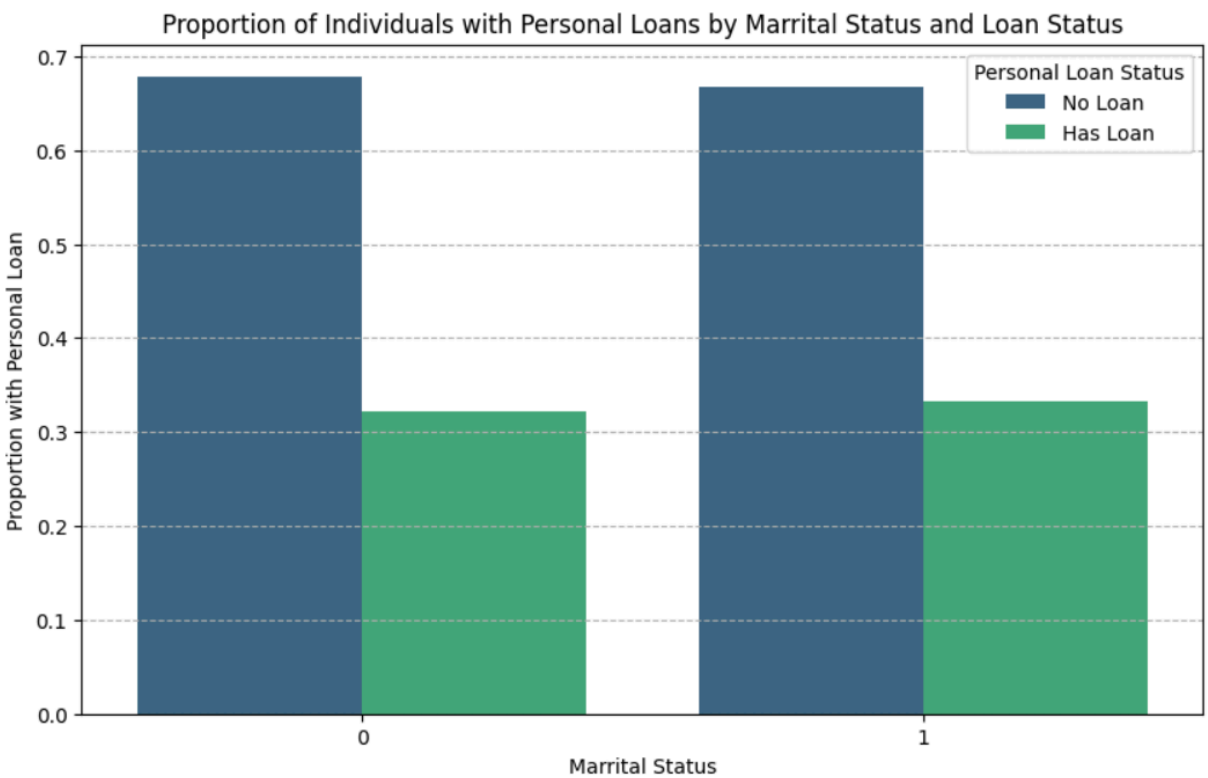
2. SVM Results

In SVM we have done the same thing that if the person is able to get the loan to purchase an EV vehicle but based on all the factor provided in dataset such as `'Age', 'Profession', 'Marrital Status', 'Education', 'No of Dependents', 'Total Salary', 'Price'.`

```
SVM Classification Report:
              precision    recall  f1-score   support

           0       0.68      1.00      0.81        13
           1       1.00      0.14      0.25         7

    accuracy                           0.70        20
   macro avg       0.84      0.57      0.53        20
weighted avg       0.79      0.70      0.62        20

SVM Confusion Matrix:
[[13  0]
 [ 6  1]]
```

Visualization:

The graph depicts the distribution of personal loan approvals for electric vehicle (EV) purchases categorized by marital status. It highlights how the likelihood of securing a personal loan varies among different marital status groups, such as Single, Married, and Divorced. The x-axis represents the various marital statuses, while the y-axis shows the proportion of individuals within each category who have obtained a personal loan for an EV. The bars are color-coded to differentiate between those who have taken a loan and those who have not. By analyzing this graph, we can discern patterns or trends regarding whether certain marital statuses are more likely to result in obtaining a personal loan for an electric vehicle. For example, if married individuals have a higher proportion of loans compared to single individuals, it might suggest different financial behaviors or lending preferences based on marital status.



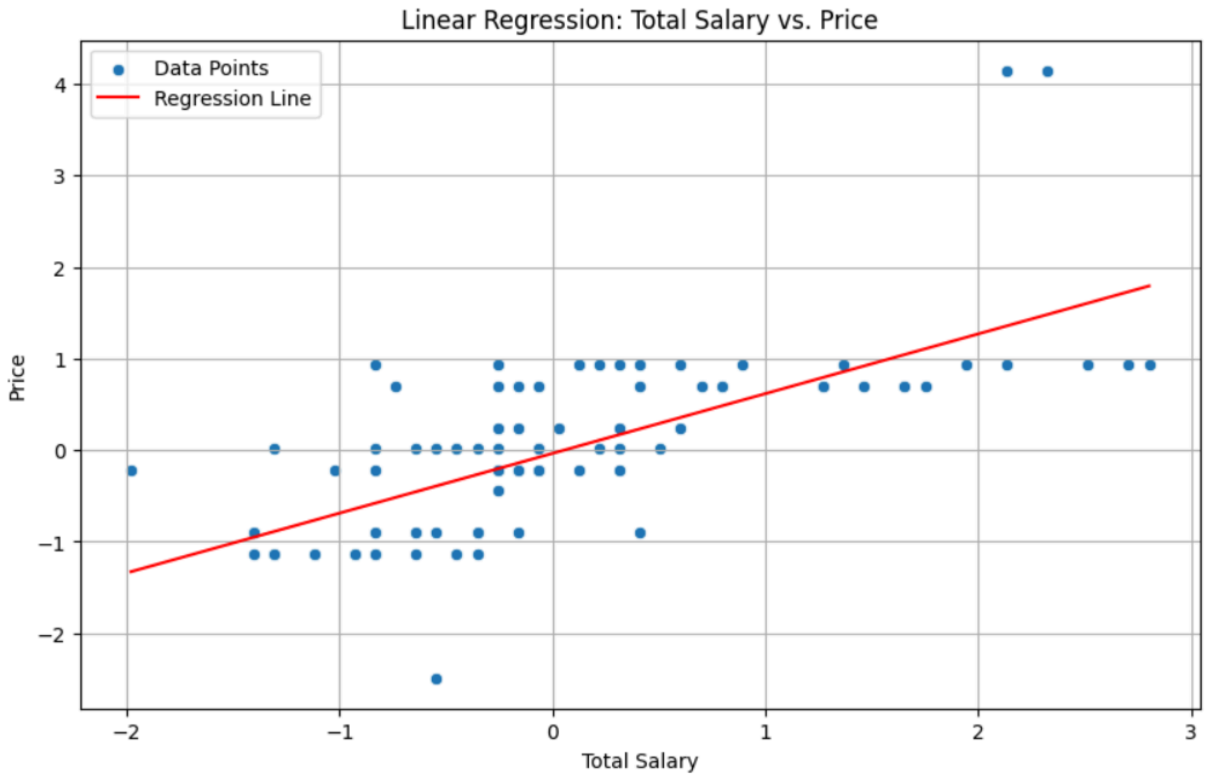
Relationship Between Individual's Salary and EV Vehicle Price

The graph demonstrates the relationship between an individual's salary and the price of the electric vehicle (EV) they purchase. Key observations include:

- **Price Proximity to Salary**: The price of the EV purchased is generally close to the individual's salary. However, there is variability, with some vehicles priced either slightly above or below the individual's salary.
- **Trend Analysis**: The graph reveals that while there is a tendency for higher salaries to correspond with higher-priced EVs, the actual price of the vehicle can vary significantly relative to the salary.

This variation suggests that while salary is an important factor in determining the price of an EV, other factors may also influence the final purchase decision.

```
R-squared: 0.60
```



<u>EV Sales Data:</u>

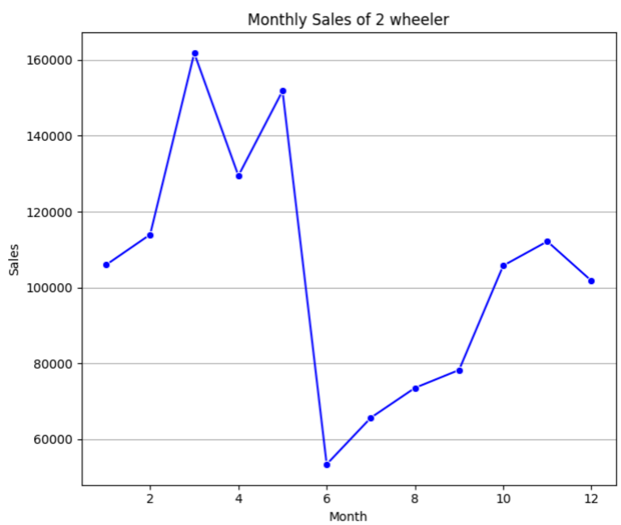
Dataset Columns

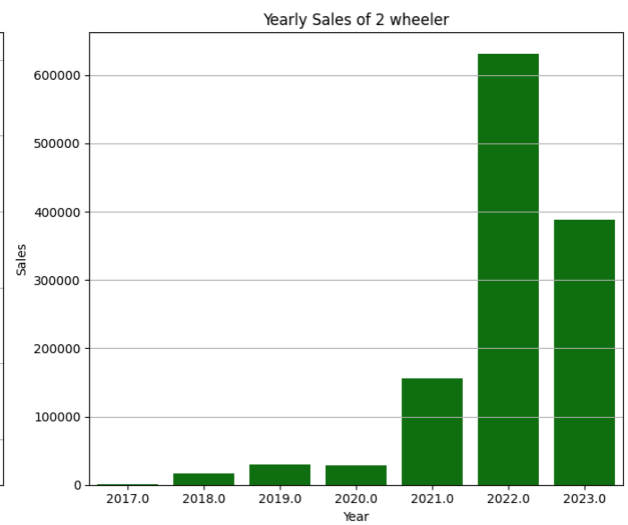
```
Index(['YEAR', '2 W', '3 W', '4 W', 'BUS', 'TOTAL'], dtype='object')
```

Monthly and Yearly Sales of 2-Wheelers

The graph presents an analysis of 2-wheeler sales both on a monthly and yearly basis.

- **Monthly Sales**: The data indicates that 2-wheelers are most frequently sold during the months of March, April, May, and June. This trend suggests a seasonal peak in sales during these months.
- **Yearly Sales**: Analysis of annual sales data reveals that the highest number of 2-wheelers were sold in 2022, while the lowest sales occurred in 2017. This upward trend in yearly sales implies a potential for continued growth in future sales.

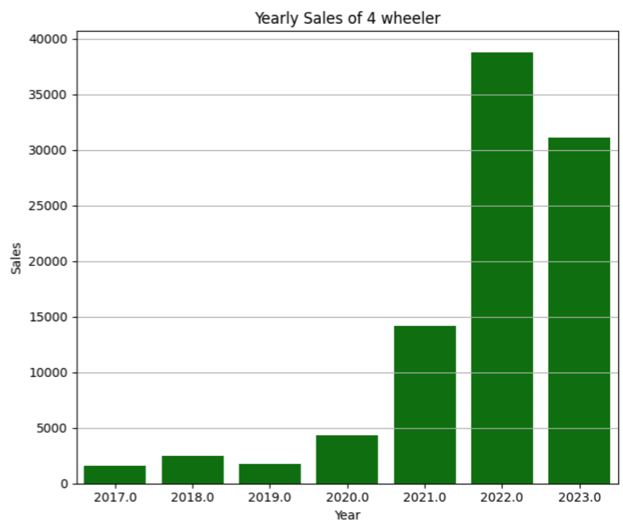




Monthly and Yearly Sales of 4-Wheelers

The graph provides insights into the sales patterns of 4-wheelers, broken down by month and year.

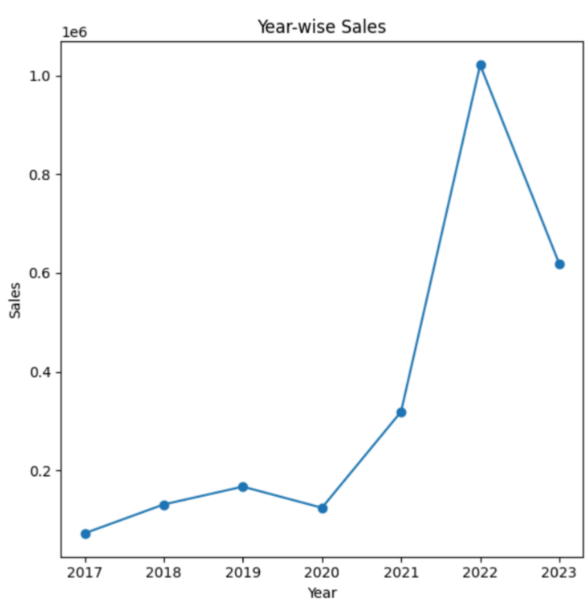
- **Monthly Sales**: The data shows that the highest number of 4-wheelers are sold during the months of March, April, and May. This indicates a peak in sales during these spring months.
- **Yearly Sales**: Analysis of annual sales data reveals that the greatest number of 4-wheelers were sold in 2022, while the lowest sales occurred in 2017. This trend suggests a general increase in sales over the years, indicating potential for further growth in the future.



Sales Analysis of Electric Vehicles (EVs)

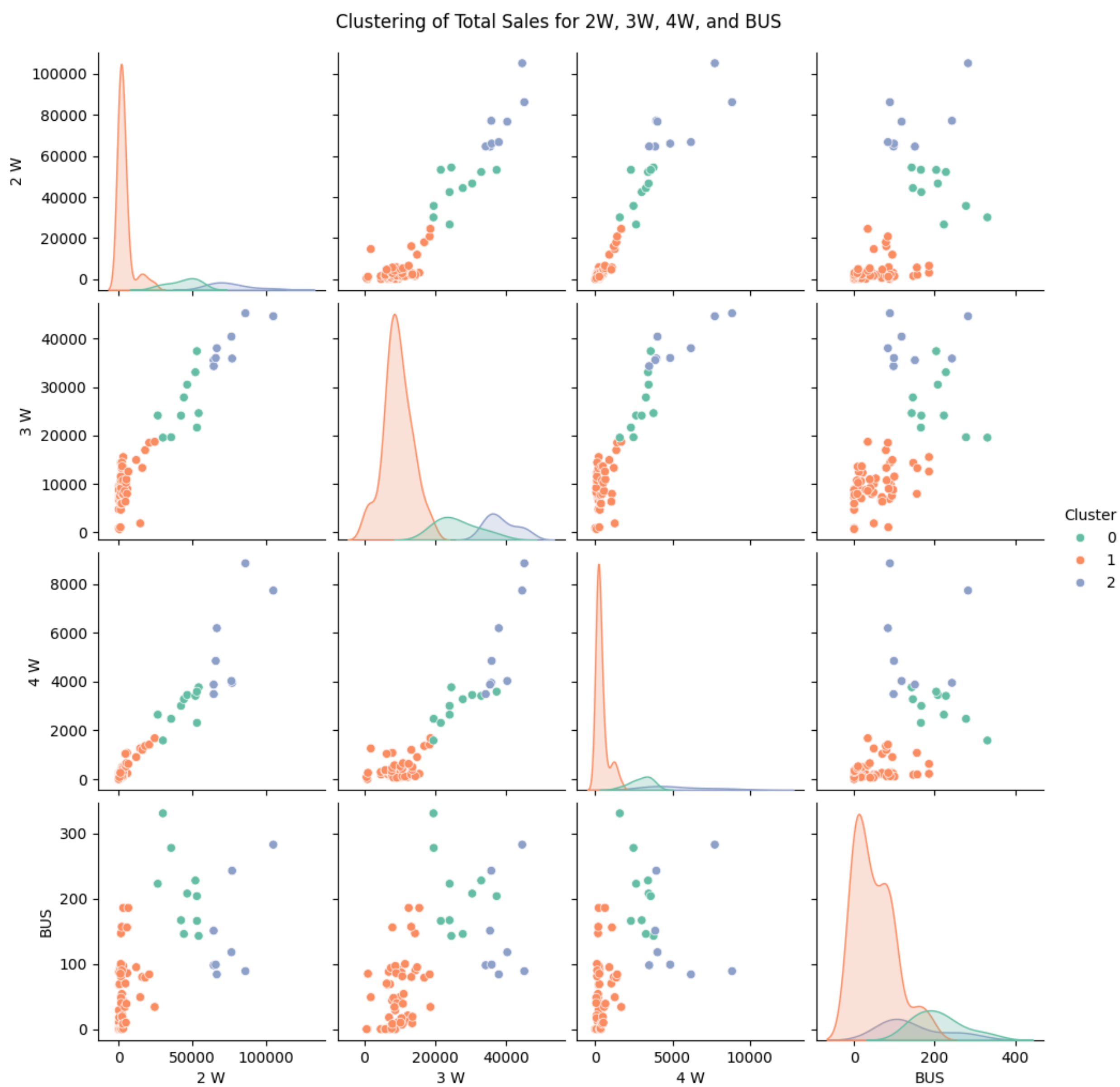
The graph provides a detailed view of the sales trends for all types of electric vehicles (EVs), segmented by year and month.

- **Yearly Sales**: The total annual sales of EVs were highest in 2022, with a general increase in sales over the years accompanied by small fluctuations. This indicates a positive growth trend in the EV market.
- **Monthly Sales**: The total sales for all types of EVs peak during March, April, and May. Organizations aiming to maximize profits should consider focusing their sales strategies and promotions during these months.



Relation Between Total Sales of 2-Wheelers, 3-Wheelers, 4-Wheelers, and Buses

The plot illustrates the relationship between the total sales figures for different vehicle categories: 2-wheelers, 3-wheelers, 4-wheelers, and buses. The plot provides a comparative view of sales across these categories, allowing for an assessment of their relative performance and trends.

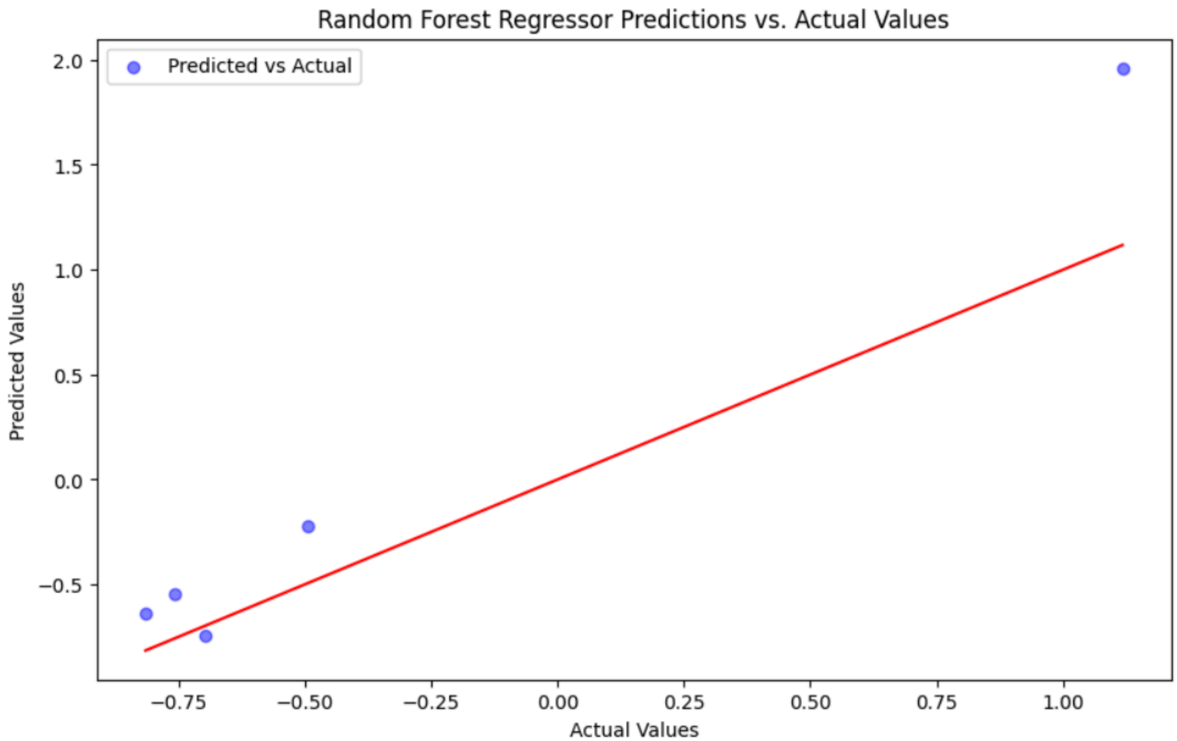


Estimation of Charging Stations Needed Using Random Forest Regression:

We applied a Random Forest Regressor to estimate the total number of charging stations required across various states, considering all types of EV vehicles, including 2-wheelers, 3-wheelers, 4-wheelers, and buses.

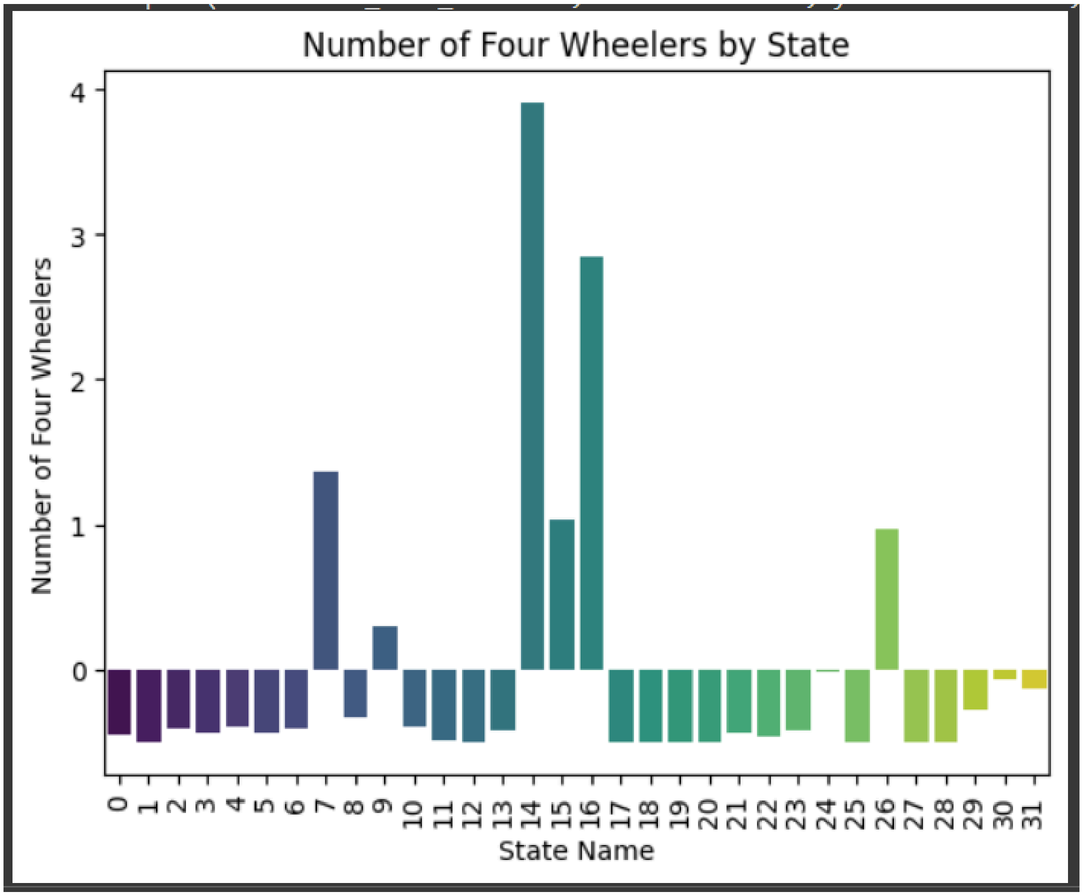
The graph clearly demonstrates that as the number of EV vehicles increases in a particular state, the demand for charging stations correspondingly rises. This highlights the direct relationship between EV adoption and the need for expanded charging infrastructure.

```
Mean Squared Error: 0.17
```



Number of EV 4-Wheelers by State

The graph displays the number of electric 4-wheelers registered in each state, providing a comparative overview of EV adoption across different regions.



Github link:
(https://github.com/nidhi-158/FeynnLabs_EV-market)