

# EV Vehicle Segmentation Analysis of Electric Vehicles Market in India

By

**Darshan Sunil Shewale**

**Date: 22-09-2023**



# 1. Introduction

This project aims to perform market segmentation analysis in the Electric Vehicle (EV) market in India. The objective is to provide strategic insights for an Electric Vehicle Startup to target specific customer segments effectively. The segmentation analysis is conducted using K-means clustering in combination with Principal Component Analysis (PCA) for dimensionality reduction.

The project aims to provide insights into potential customer segments for an Electric Vehicle Startup to target. The segmentation analysis utilizes K-means clustering in conjunction with Principal Component Analysis (PCA) to identify distinct customer segments.

## 2. Data Exploration and Preprocessing

Data Collection research reference links

Dataset research:

<https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset>

<https://data.world/data-ny-gov/thd2-fu8y>

<https://data.world/smart-columbus/2f0e672b-f6dd-4718-8557-d484b2f2f14e>

<https://data.world/smart-columbus/19fb49cb-0875-484d-9b3f-5acd2d57cfa2/workspace/file?filename=connected-electric-autonomous-vehicle-shift-segments-1.csv>

<https://www.kaggle.com/code/sjsumanth/ev-segmentation/input>

<https://www.kaggle.com/datasets/utkarshx27/electric-vehicle-population-data>

<https://www.kaggle.com/datasets/gunapro/electric-vehicle-population-data>

<https://www.kaggle.com/datasets/kkhandekar/cheapest-electric-cars>

<https://www.kaggle.com/datasets/iottech/electric-vehicle-data-1997-2024-update-version>

<https://www.kaggle.com/datasets/edsonmarin/historic-sales-of-electric-vehicles>

<https://www.kaggle.com/datasets/kkhandekar/electric-vehicles-india>

<https://www.kaggle.com/datasets/atechnohazard/battery-and-heating-data-in-real-driving-cycles?select=TripA03.csv>

<https://www.kaggle.com/datasets/varpit94/tesla-stock-data-updated-till-28jun2021>

<https://www.kaggle.com/datasets/sjsumanth/electric-data-analysis>

<https://github.com/chargeprice/open-ev-data/tree/master/data>

<https://s3.amazonaws.com/capitalbikeshare-data/index.html>

<https://github.com/BairagiSaurabh/Electric-Vehicle-Market-Segmentation/tree/main/Vehicle%20Specification%20Segmentation>

## Data Collection

The project starts by fetching data from the 'Car details v3.csv' dataset.

This dataset contains information about used cars.

The columns in the given dataset are as follows:

- 0 name
- 1 year
- 2 selling\_price
- 3 km\_driven
- 4 fuel
- 5 seller\_type
- 6 transmission
- 7 owner
- 8 mileage
- 9 engine
- 10 max\_power
- 11 torque
- 12 seats

## Data Cleaning and Missing Value Handling

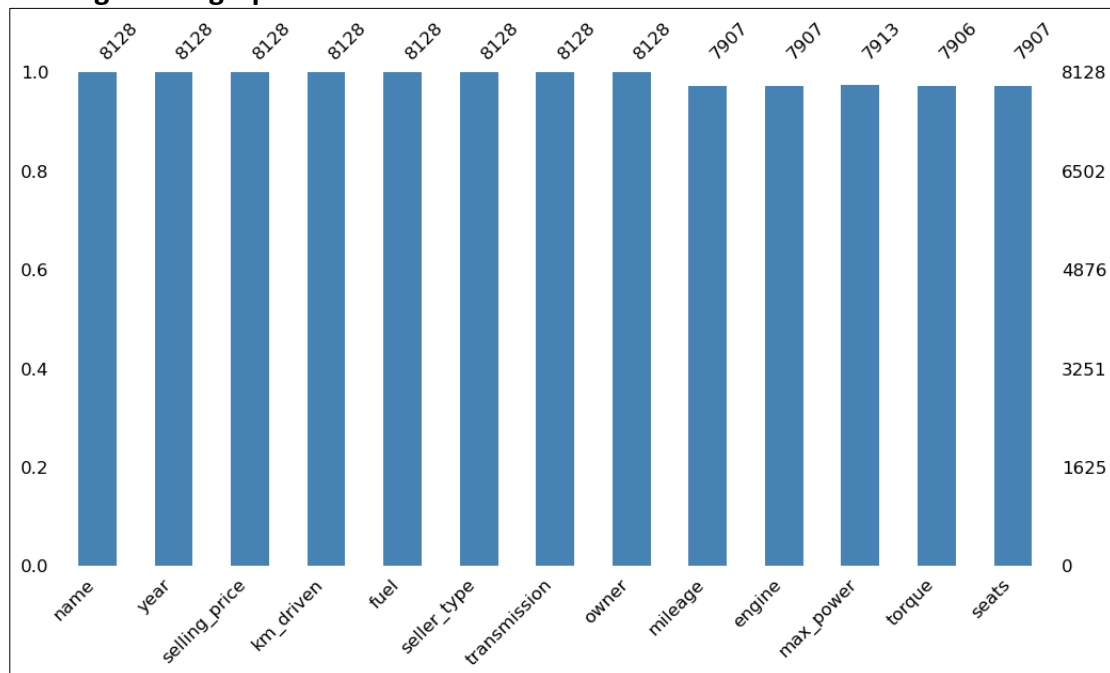
- Missing values are identified and visualized using 'missingno' library.
- Rows with missing values are dropped from the dataset.
- Several columns are cleaned and transformed for analysis.

## Exploratory Data Analysis (EDA)

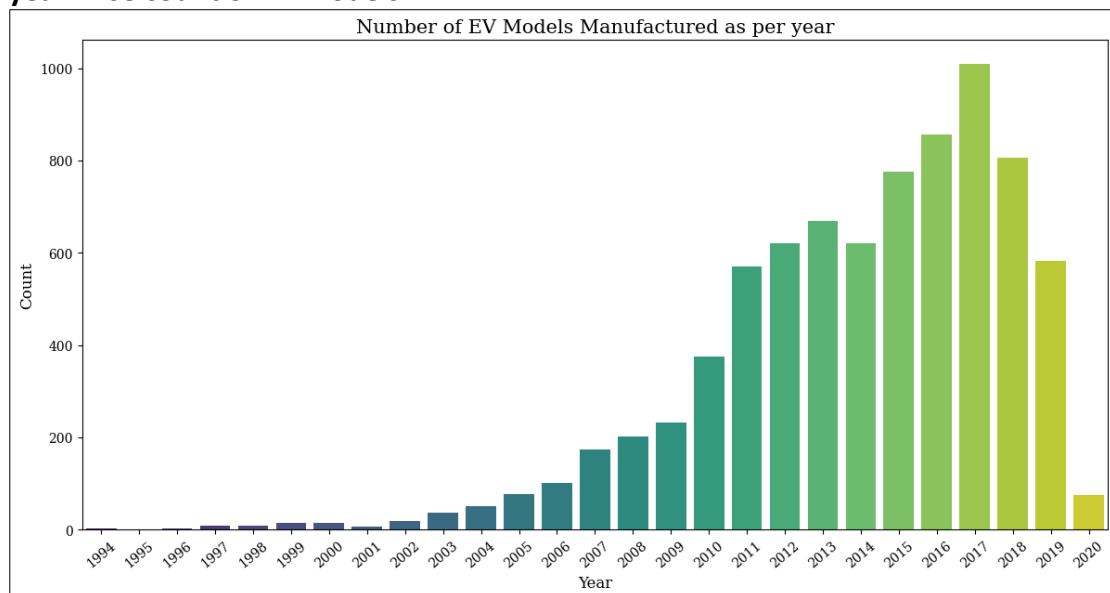
An Exploratory Data Analysis or EDA is a thorough examination meant to uncover the underlying structure of a data set and is important for a company because it exposes trends, patterns, and relationships that are not readily apparent.

- Key statistics, counts, and visualizations are generated to understand the dataset.
- Features like 'year', 'mileage', 'engine', 'max\_power', 'seats', and 'torque' are explored to gain insights into EV market trends.

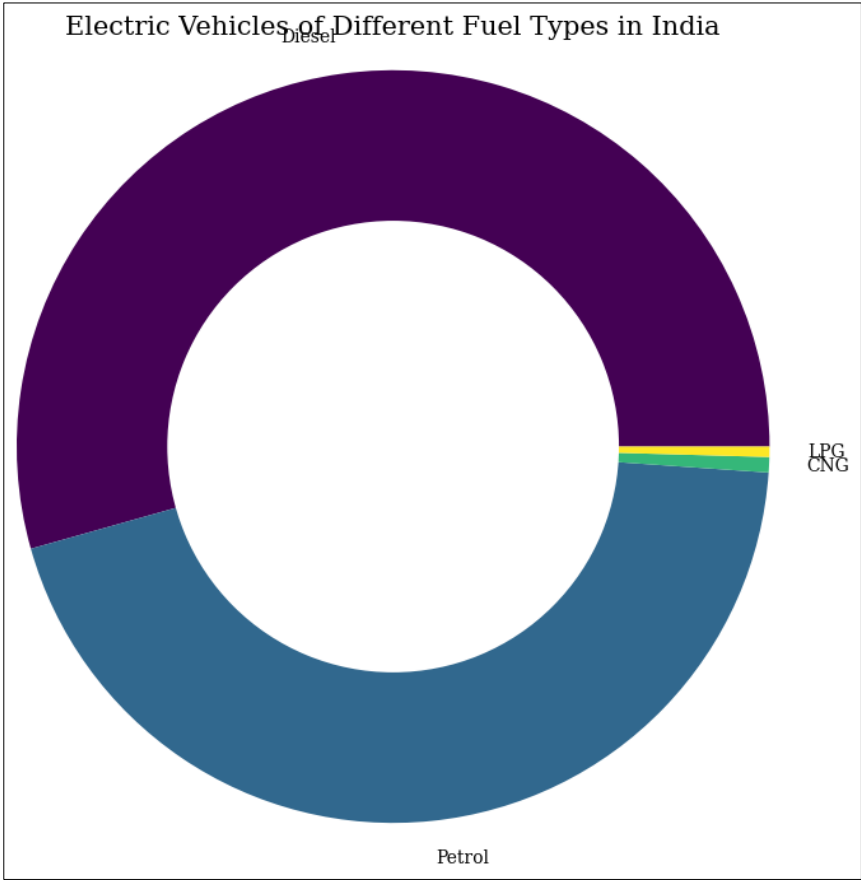
Missing values graph



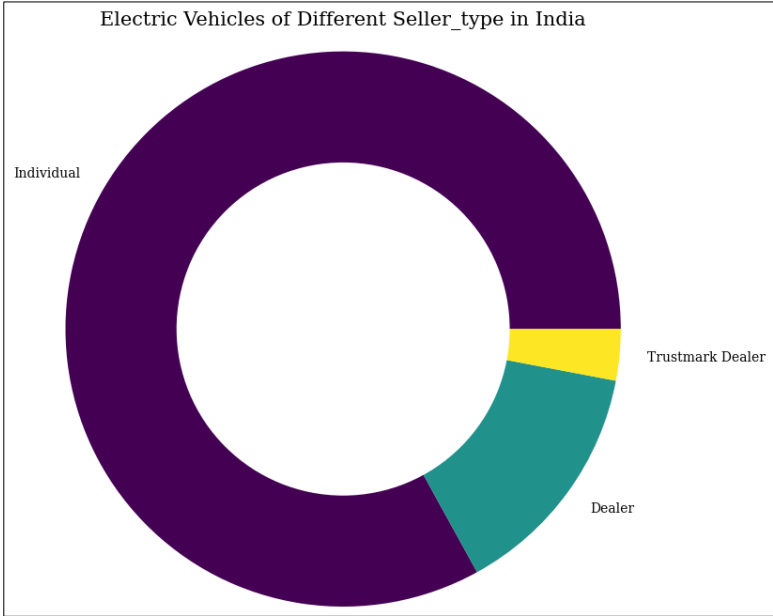
year-wise count of EV models



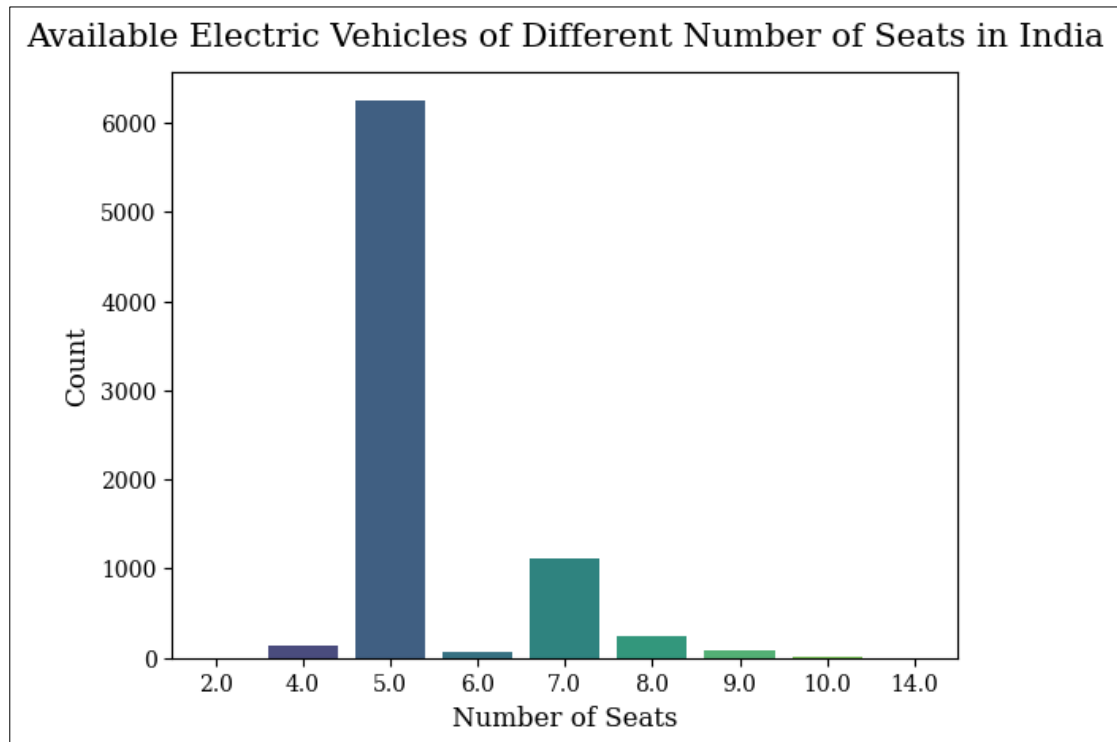
different body types EVs visualization from dataset



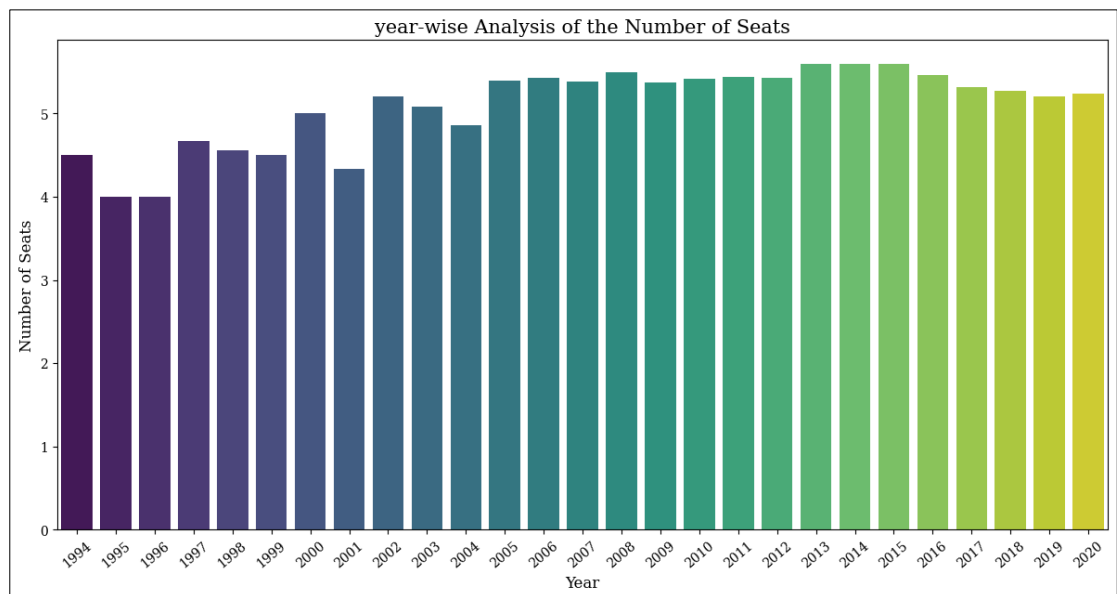
Electric Vehicles of Different Seller\_type in India



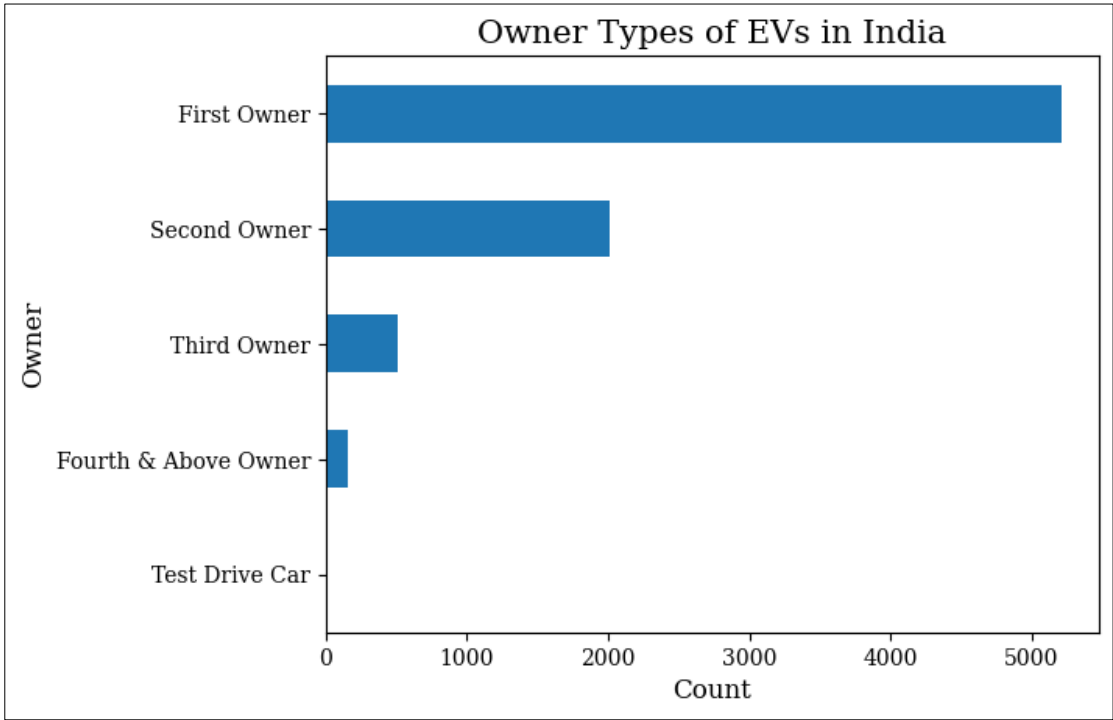
### EVs of different number of seats from dataset



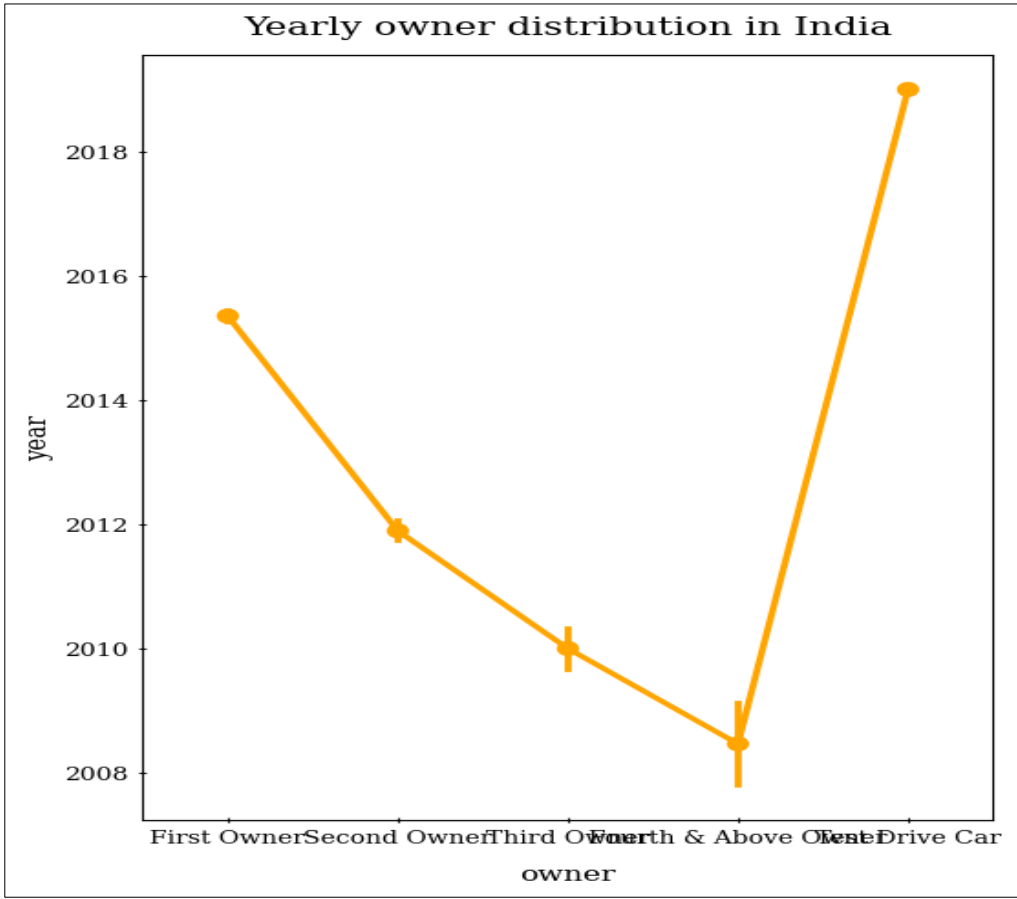
### year-wise analysis of the number of seats



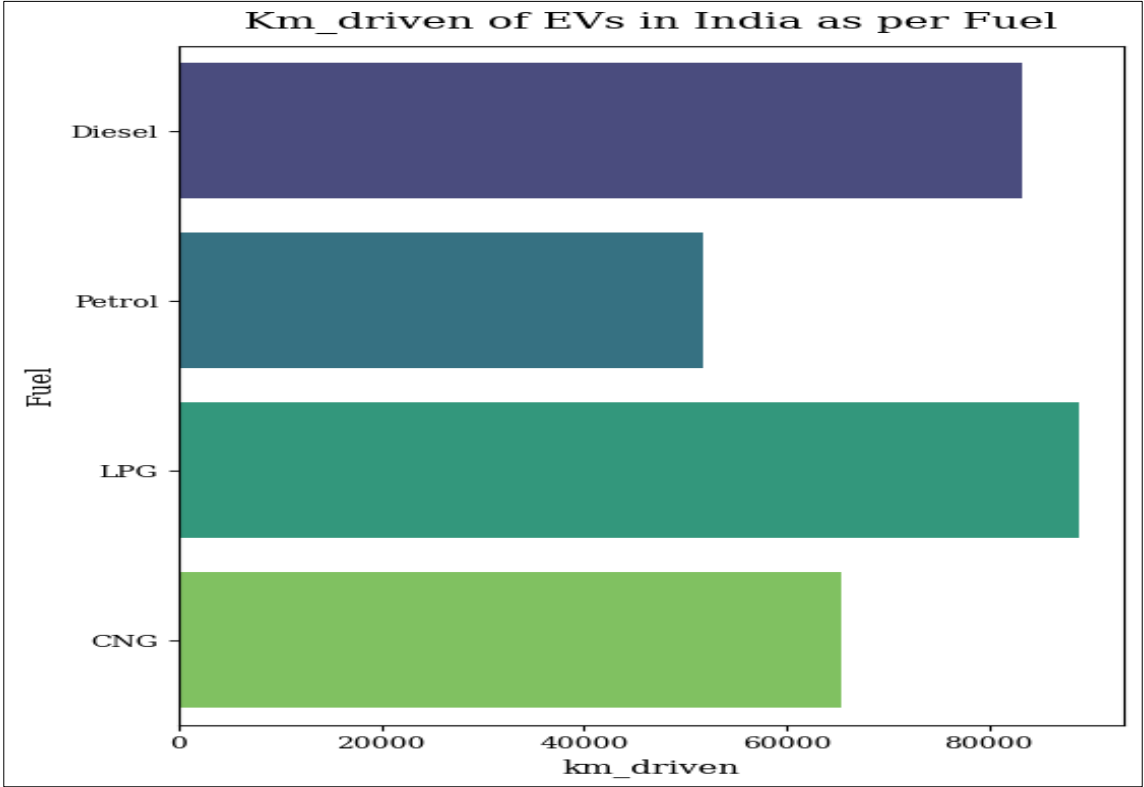
owner types visualization from dataset



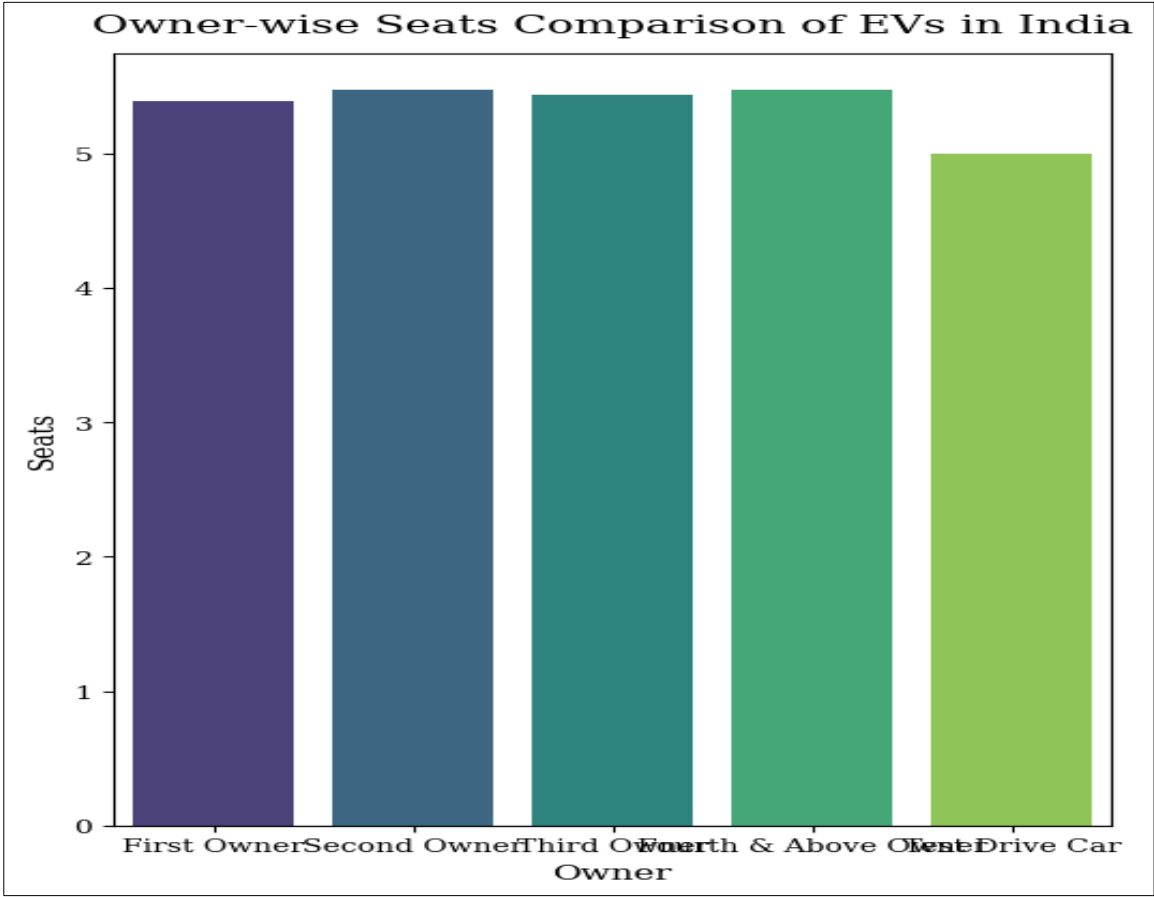
retail outlets visualization from dataset



km\_driven visualization from dataset

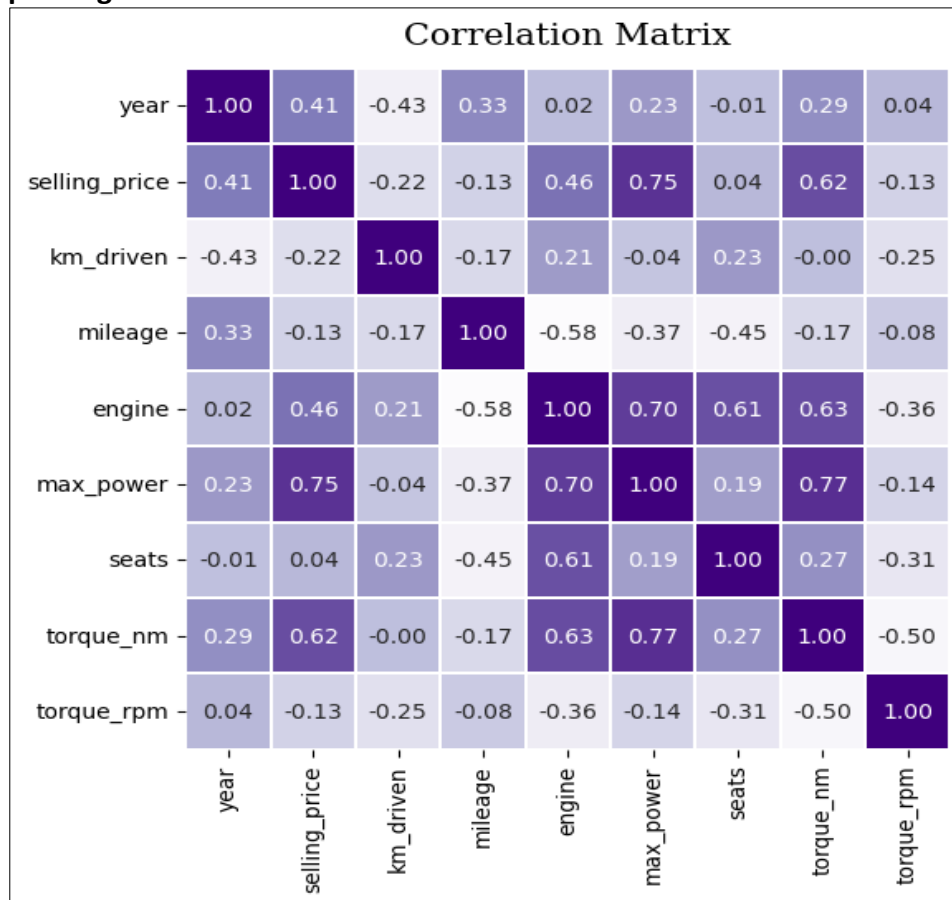


owner visualization from dataset





plotting the correlation matrix from dataset



### 3. Feature Engineering

Parsing and Engineering Torque Features

- The 'torque' column is parsed to extract relevant information, including torque value and engine speed.
- Units are standardized, and 'torque' is split into 'torque\_nm' and 'torque\_rpm' columns.
- Data types are converted for analysis.

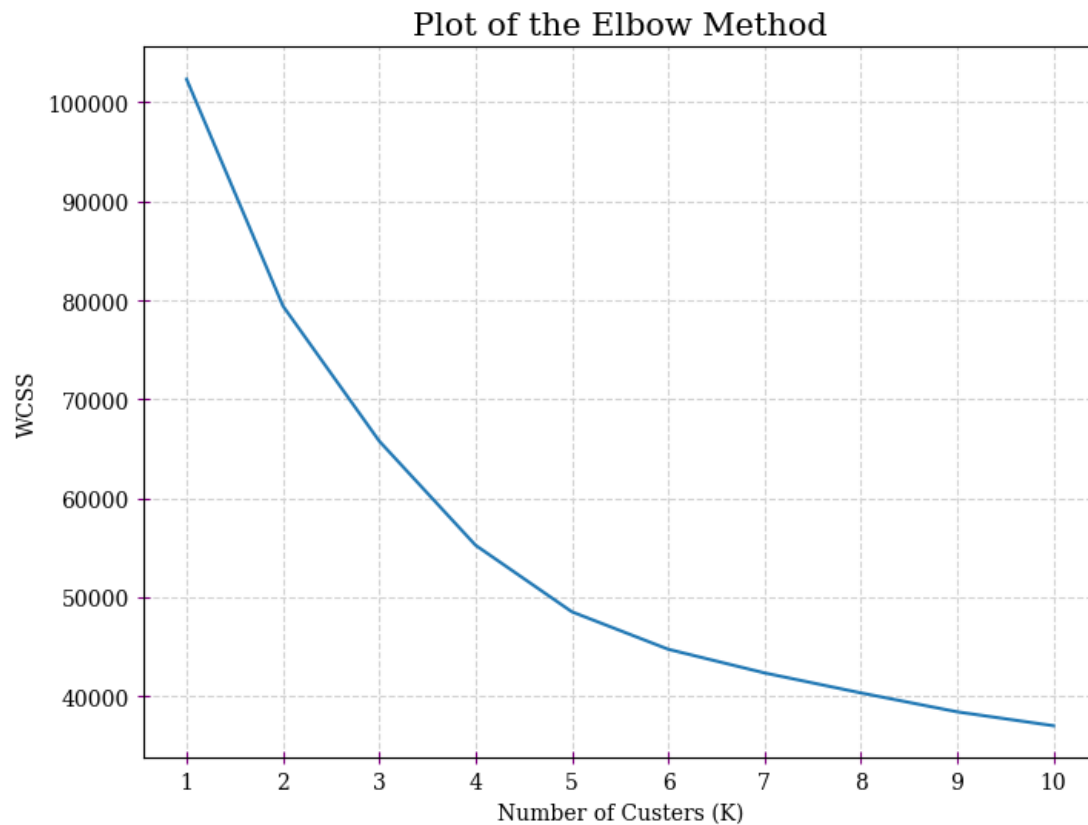
### 4. Dimensionality Reduction with PCA

- Principal Component Analysis (PCA) is applied to reduce feature dimensionality.
- The dataset is transformed into principal components for clustering analysis.

## 5. K-Means Clustering

### Optimal Number of Clusters

- The Elbow Method is used to determine the optimal number of clusters.
- The analysis suggests 'k=4' clusters as the best choice.



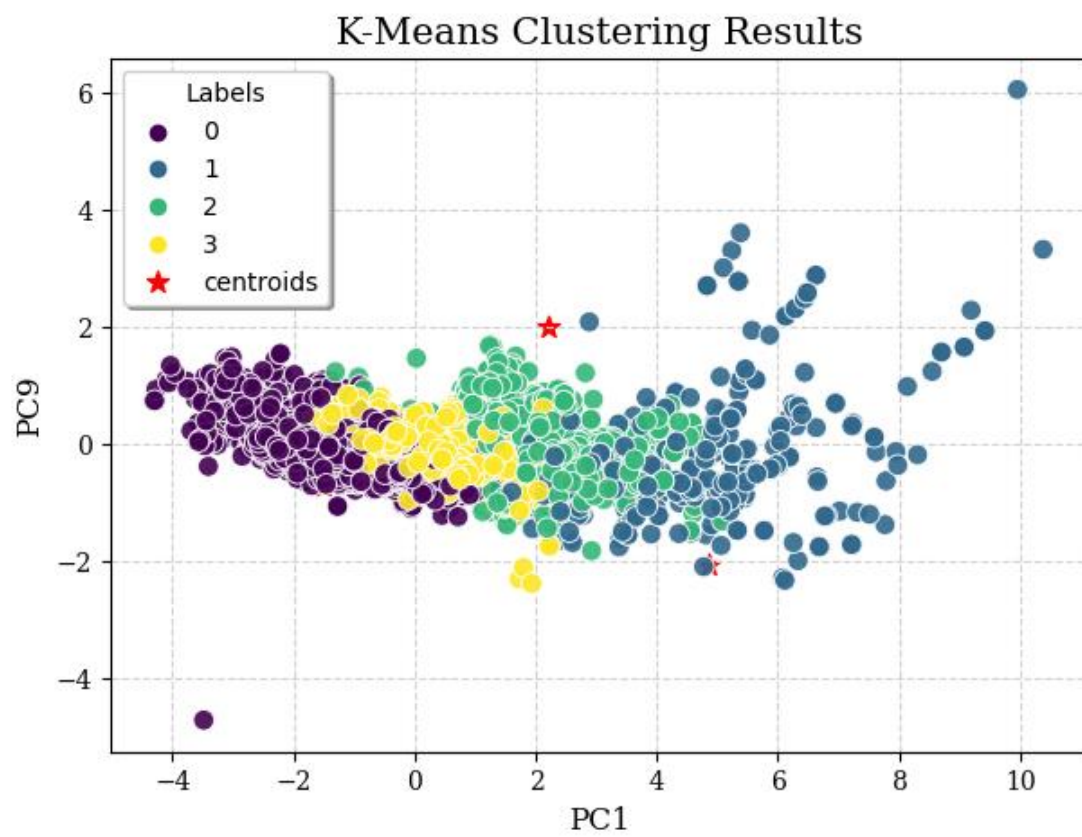
### Model Training and Cluster Assignment

- K-means clustering is performed with 'k=4' clusters.
- Data points are assigned to clusters based on their attributes.
- Cluster centers are visualized.

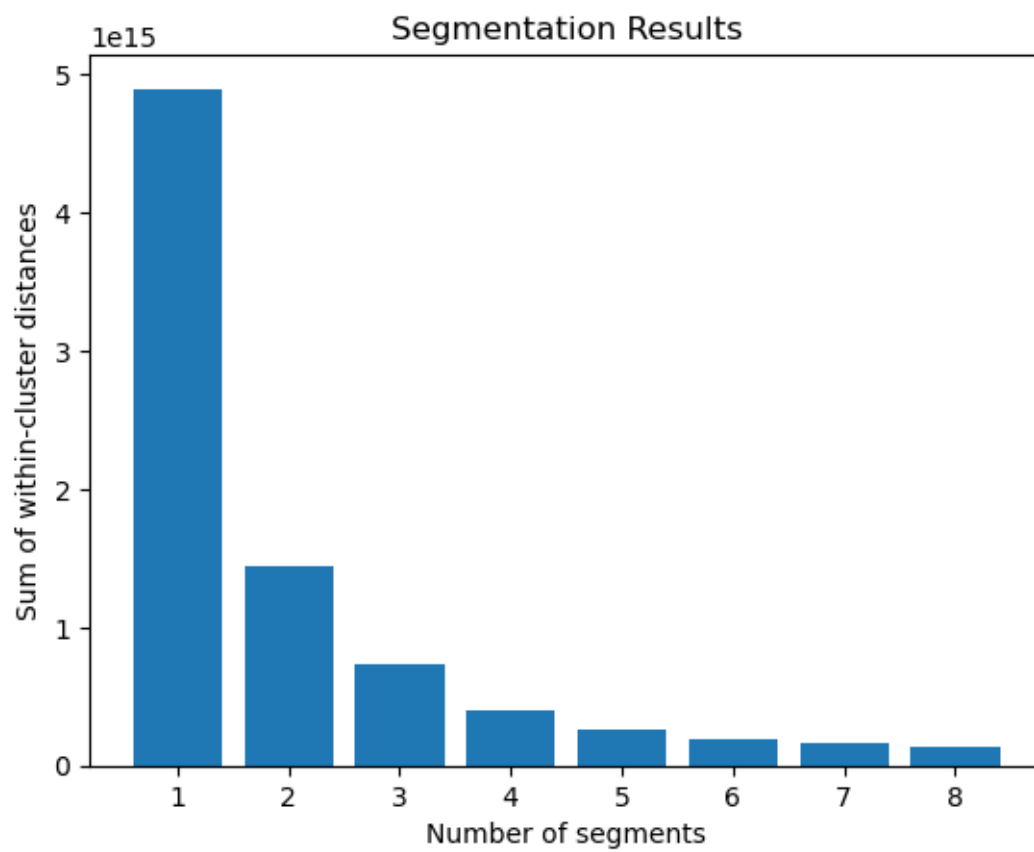
## 6. Cluster Visualization

- Clusters are visualized in a scatterplot, showing the distribution of data points in reduced dimensions.
- Centroids are marked for each cluster.

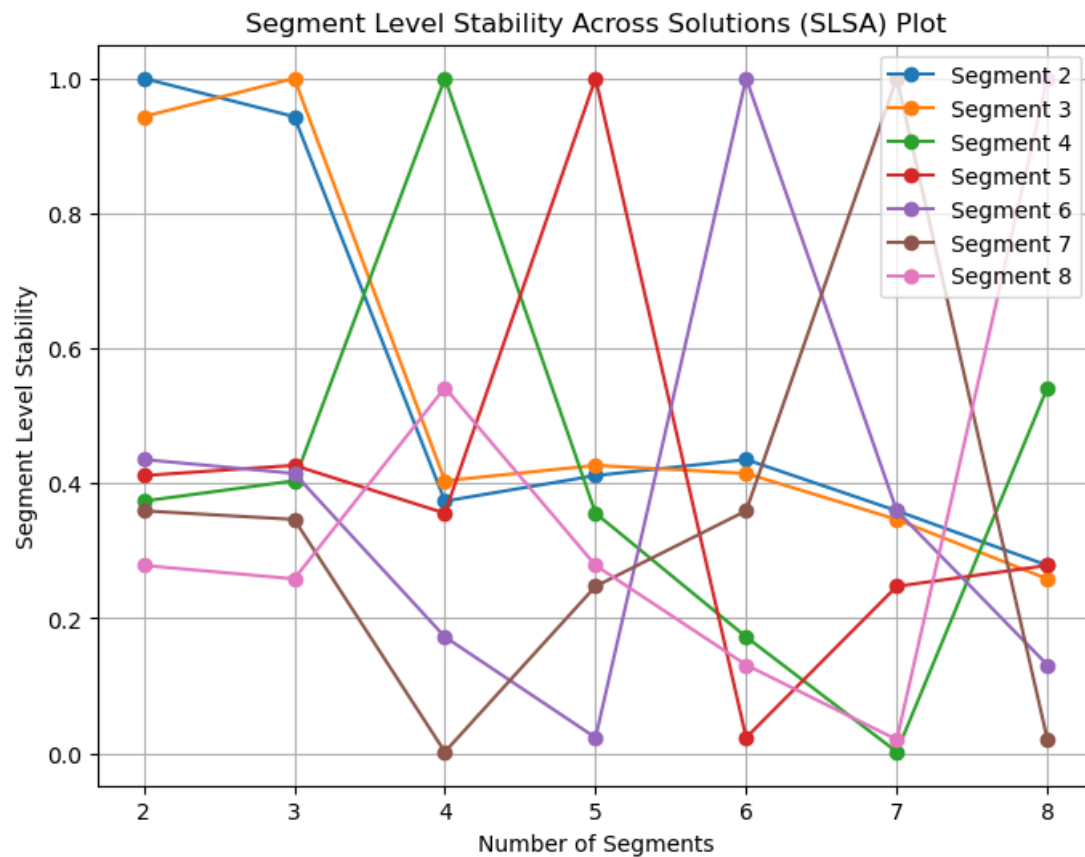
visualizing clusters



Segmentation Results



### Segment Level Stability Across Solutions (SLSA) Plot



## 7. Segment Stability Analysis

- Segment stability analysis is conducted to assess the stability of clusters.
- Stability results are visualized to determine the robustness of the clustering solution.

## 8. Conclusion and Recommendations

Based on our analysis, we provide the following recommendations for the Electric Vehicle Startup:

**Targeted Marketing:** Tailor marketing campaigns to the identified customer segments. Each segment exhibits distinct preferences and characteristics, so personalized marketing messages and offers can be more effective.

**Product Development:** Consider developing electric vehicle models that align with the preferences of specific segments. For example, if one segment values high torque and power, design EVs that cater to these requirements.

**Pricing Strategies:** Pricing strategies should vary based on the segments. Some segments may be more price-sensitive, while others may be willing to pay a premium for specific features.

**Geographic Focus:** Identify regions where specific segments are concentrated. This information can guide the geographic targeting of sales and marketing efforts.

**After-Sales Services:** Tailor after-sales services, warranties, and support based on the needs and preferences of each segment. Providing excellent post-purchase experiences can enhance customer loyalty.

**Data Collection:** Continue to collect data on customer behavior and preferences to refine the segmentation model and adapt strategies as the EV market evolves.

**Product Innovation:** Invest in research and development to innovate and stay ahead of the competition. Understanding customer segments can drive innovation in EV technology and features.

**Sustainability Initiatives:** Highlight the environmental benefits of electric vehicles to segments that prioritize sustainability and eco-friendliness.

## 9. Acknowledgments

**Dataset Providers:** We extend our appreciation to the providers of the dataset used in this project. Without access to this data, our analysis and segmentation would not have been possible.

**Open Source Community:** We acknowledge the open-source community for developing and maintaining the various libraries and tools used in this project, including Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn. These libraries played a crucial role in data manipulation, visualization, and machine learning.

**Educational Resources:** We are grateful for the wealth of online tutorials, articles, and educational resources available on data analysis, machine learning, and clustering. These resources provided valuable insights and guidance.

## 10. References

Dataset research:

<https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset>

<https://data.world/data-ny-gov/thd2-fu8y>

<https://data.world/smart-columbus/2f0e672b-f6dd-4718-8557-d484b2f2f14e>

<https://data.world/smart-columbus/19fb49cb-0875-484d-9b3f-5acd2d57cfa2/workspace/file?filename=connected-electric-autonomous-vehicle-shift-segments-1.csv>

<https://www.kaggle.com/code/sjsumanth/ev-segmentation/input>

<https://www.kaggle.com/datasets/utkarshx27/electric-vehicle-population-data>

<https://www.kaggle.com/datasets/gunapro/electric-vehicle-population-data>

<https://www.kaggle.com/datasets/kkhandekar/cheapest-electric-cars>

<https://www.kaggle.com/datasets/iottech/electric-vehicle-data-1997-2024-update-version>

<https://www.kaggle.com/datasets/edsonmarin/historic-sales-of-electric-vehicles>

<https://www.kaggle.com/datasets/kkhandekar/electric-vehicles-india>

<https://www.kaggle.com/datasets/atechnohazard/battery-and-heating-data-in-real-driving-cycles?select=TripA03.csv>

<https://www.kaggle.com/datasets/varpit94/tesla-stock-data-updated-till-28jun2021>

<https://www.kaggle.com/datasets/sjsumanth/electric-data-analysis>

<https://github.com/chargeprice/open-ev-data/tree/master/data>

<https://s3.amazonaws.com/capitalbikeshare-data/index.html>

<https://github.com/BairagiSaurabh/Electric-Vehicle-Market-Segmentation/tree/main/Vehicle%20Specification%20Segmentation>

[https://www.researchgate.net/figure/Electric-vehicle-market-segment-in-India\\_fig1\\_352994590](https://www.researchgate.net/figure/Electric-vehicle-market-segment-in-India_fig1_352994590)

<https://www.kaggle.com/code/tabrznagiyevev/car-details-v3>