EV MARKET IN INDIA

Driving The Future: An Segmentation Analysis Of India's Electric Vehicle Market

By Divya Gazinkar

Table Of Contents

Problem Statement

To help Electric Vehicle Startup to decide in which vehicle/customer space it will develop its EVs. To analyze the Electric Vehicle market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles.

Fermi Estimation (Breakdown of Problem Statement)

Fermi estimation is a technique used to make rough yet educated guesses about complex problems by breaking them down into simpler, more manageable components.

Breakdown Using Fermi Estimation:

- Market Size Estimation:
- Electric Vehicle Market Share Estimation:
- Segmentation Analysis:
    - Demographic Segmentation: Breakdown of potential customers by age, income, occupation, etc.
    - Geographic Segmentation: Analyzing which regions in India have higher EV adoption rates.
    - Psychographic Segmentation: Understanding consumer values and attitudes toward green technology.
    - Behavioral Segmentation: Identifying early adopters and their preferences.
- Feasibility Strategy:
    - Competitor Analysis: Study of existing EV manufacturers in India.
    - Cost Estimation: Calculating the cost of EV production, charging infrastructure, and marketing.
    - Regulatory and Government Policies: Assessment of incentives, subsidies, and regulations favoring EVs.
    - Partnerships and Alliances: Identifying potential collaborations for manufacturing, distribution, and charging infrastructure.
- Targeting the Most Likely Segments:
    - Using the results from the segmentation analysis, prioritize segments that align with the startup's strengths and resources.
    - Develop marketing and product strategies tailored to the selected segments.

By breaking down the problem statement in this manner, the Electric Vehicle Startup can gain a clearer understanding of the market's size, potential growth, and which segments are most promising for their entry strategy.

Data Collection

1. Electric Car Data
2. India Air Quality Index (AQI)
3. State wise operational Public EV Charging Stations
4. Vehicle count State wise

First, the study explores AQI data state-wise, shedding light on the environmental context in which the Indian EV market operates. It examines how air quality levels influence consumer choices and government policies favoring cleaner mobility solutions.

Second, the distribution of power charging stations across states is a crucial factor in determining the viability of EV adoption. We analyze the dataset on charging infrastructure, highlighting regional disparities and opportunities for expansion to support the growing EV market.

Third, the analysis of the count of EVs and non-EVs at the state level offers a comprehensive view of market penetration. This segmentation helps identify regions with untapped potential and those where EVs have gained significant traction.

Lastly , we analyze the dataset on EV cars in India, examining factors like make, model, price, and sales trends. This analysis aids in understanding the evolving preferences of Indian consumers when it comes to EVs and provides insights into the competitive landscape within the EV industry.

Data Pre-processing

Libraries and their imported modules and classes:

NumPy (np): NumPy is a fundamental package for scientific computing in Python, providing support for multi-dimensional arrays and mathematical functions.

Pandas (pd): Pandas is a powerful library for data manipulation and analysis, offering data structures like DataFrames and Series to handle structured data efficiently.

Matplotlib.**pyplot** (plt): Matplotlib is a popular plotting library that provides tools for creating static, animated, and interactive visualizations in Python.

Seaborn (sns): Seaborn is a data visualization library built on top of Matplotlib, designed for making statistical graphics more informative and attractive.

Warnings: The warnings library allows you to control how warning messages are displayed in your Python environment.

scikit-learn (sklearn): Scikit-learn is a popular machine learning library in Python that provides tools for data preprocessing, modeling, evaluation, and more.
Imported Modules and Classes:

- StandardScaler: Used for standardizing (scaling) numerical features to have mean 0 and standard deviation 1.
- OneHotEncoder: Used for one-hot encoding categorical features, converting them into binary columns.
- ColumnTransformer: Allows for applying different transformers to different subsets of the columns in a dataset.
- Pipeline: Enables the creation of a processing pipeline that combines multiple preprocessing steps and a machine learning model.
- PCA (Principal Component Analysis): A dimensionality reduction technique used for feature extraction and reducing the number of features.
- KMeans: A clustering algorithm for unsupervised learning, used to group data into clusters based on similarity.
- train_test_split: Used to split a dataset into training and testing subsets for model evaluation.
- LinearRegression: A linear regression model for predicting a target variable based on input features.
- metrics: Provides various metrics for evaluating model performance, such as mean squared error, R-squared, etc.

scipy.cluster.hierarchy: Part of SciPy, this library provides hierarchical clustering methods and tools.
Imported Modules and Classes:

- dendrogram: Allows for creating dendrogram plots, which visualize hierarchical clustering results.

EDA

1. Loaded the data using df=pd.read_csv().
2. Displayed the rows of the dataset using df.head().
3. Checked for the missing values in the columns using df.isnull().sum(). There were no missing values in the datasets.
4. Printed information about a DataFrame including the index dtype and columns, non-null values and memory usage using df.info

5. Generate descriptive statistics of DataFrame columns using df.describe()
6. Dimension of dataframe using df.shape
7. Data visualization:

India Air Quality Index - AQI

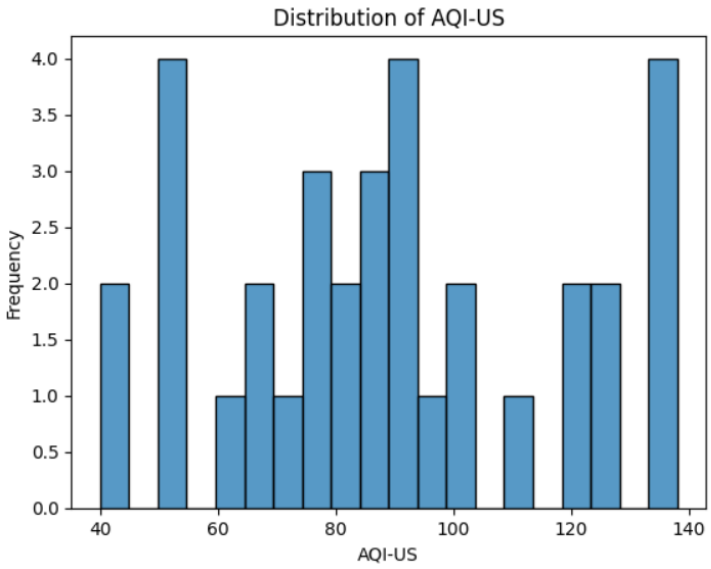

Fig. 1 Distribution of AQI-US

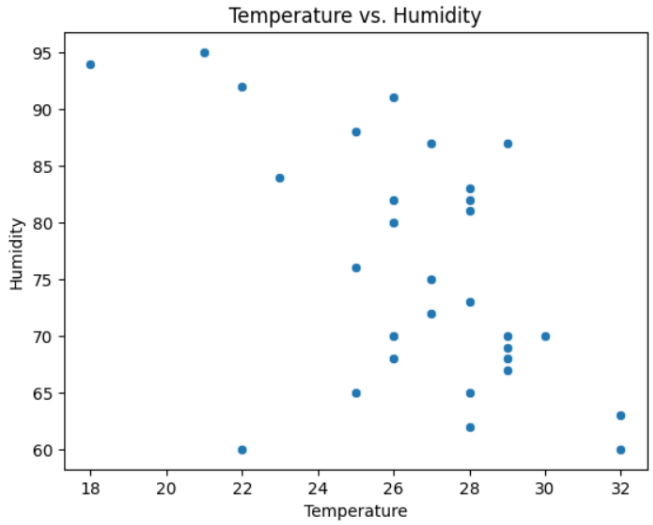

Fig. 2 Temperature vs Humidity

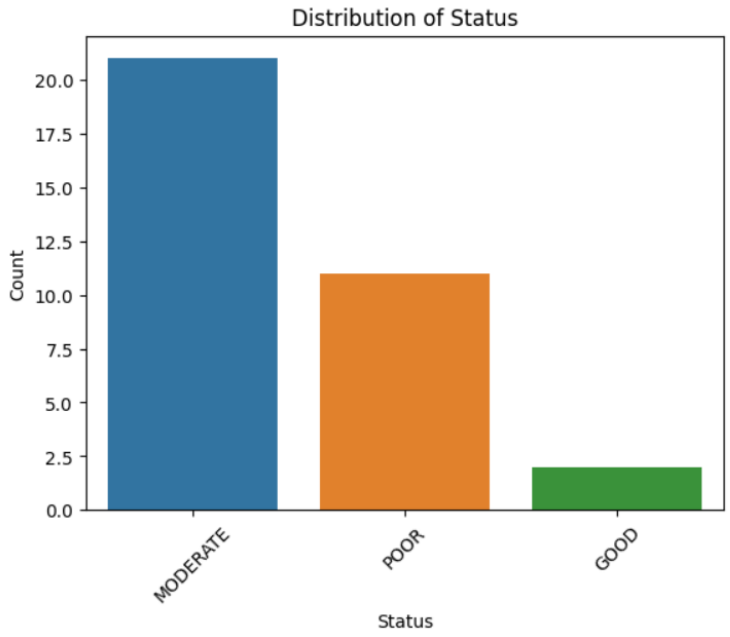

Fig 3. Distribution of Status

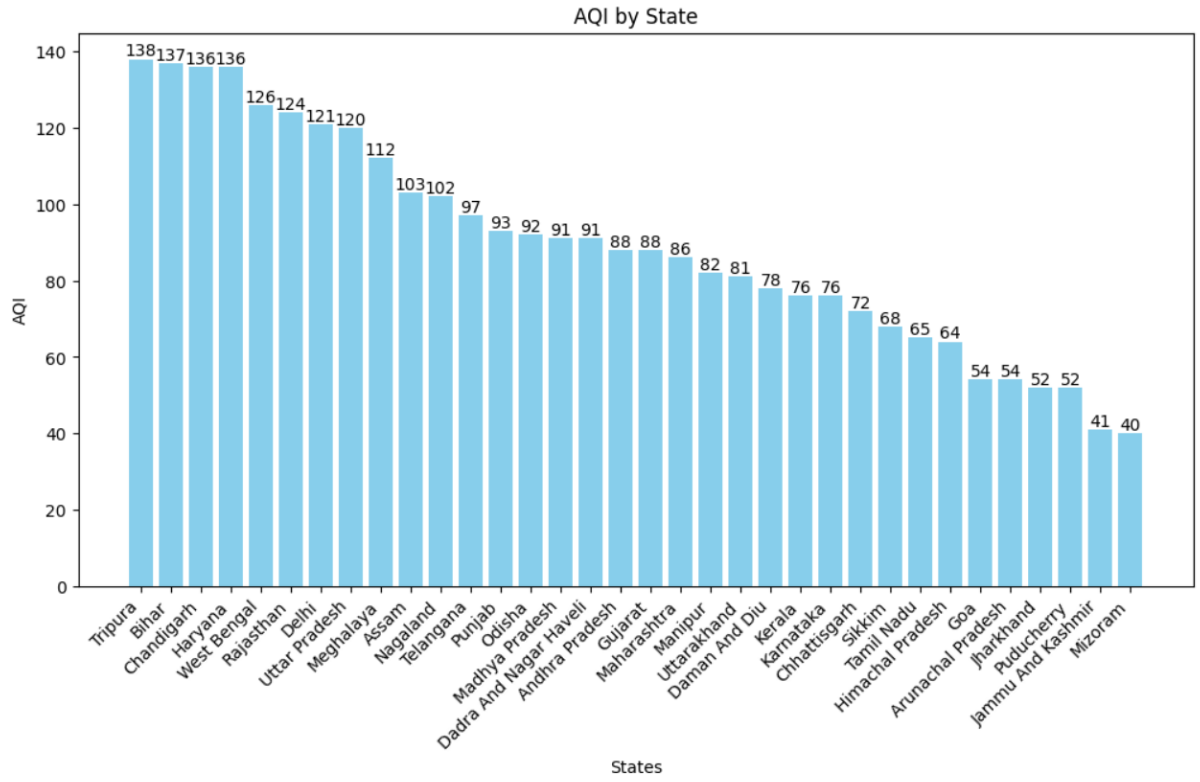

Fig 4 AQI by state. Tripura has the highest AQI whereas Mizoram has least. When AQI values are above 100, air quality is unhealthy. EVs can reduce the AQI.

State wise operational Public EV Charging Stations (PCS) - dataset

```
Sum_PCS=df2['No. of Operational PCS'].sum()
print(Sum_PCS)
```

```
6586
```

```
State with the Highest Number of Operational PCS:
  State Name  No. of Operational PCS
7      Delhi                    1845

State with the Lowest Number of Operational PCS:
                          State Name  No. of Operational PCS
16                        Lakshadweep                       1
26                             Sikkim                       1
30  Dadar &Nagar Haveliand Daman &Diu                       1
```

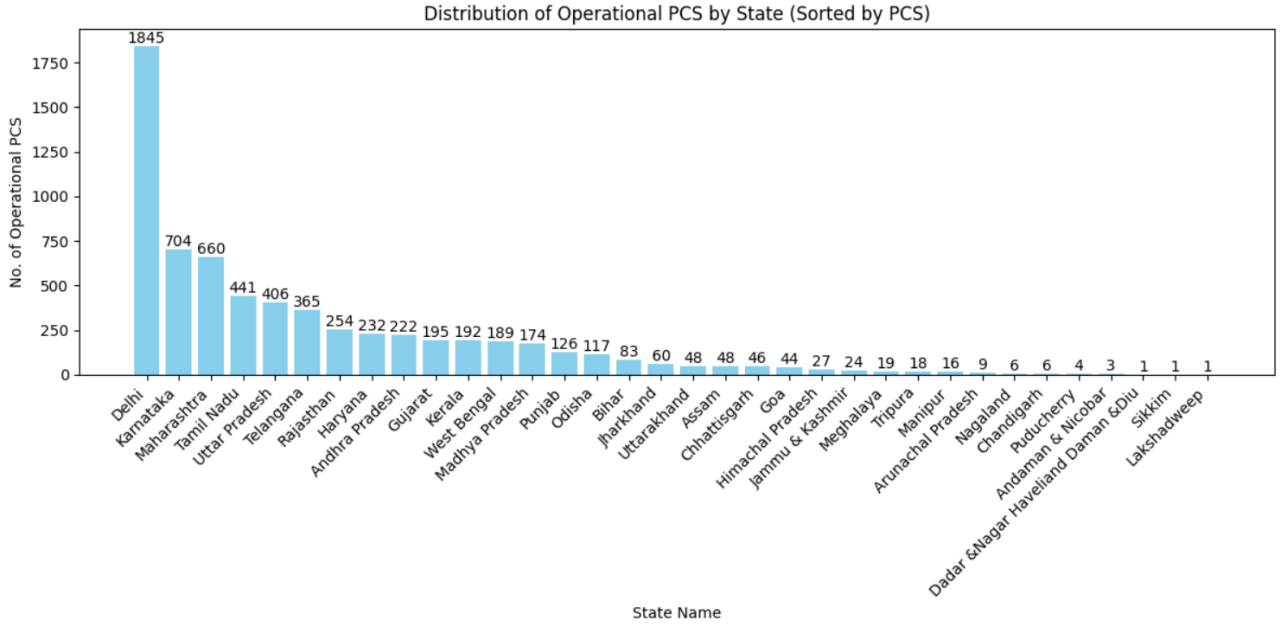

Fig. 5 distribution of Operational PCS by State. Delhi contains the highest no. of operational PCS in India. Karnataka and Maharashtra are 2nd and 3rd largest.

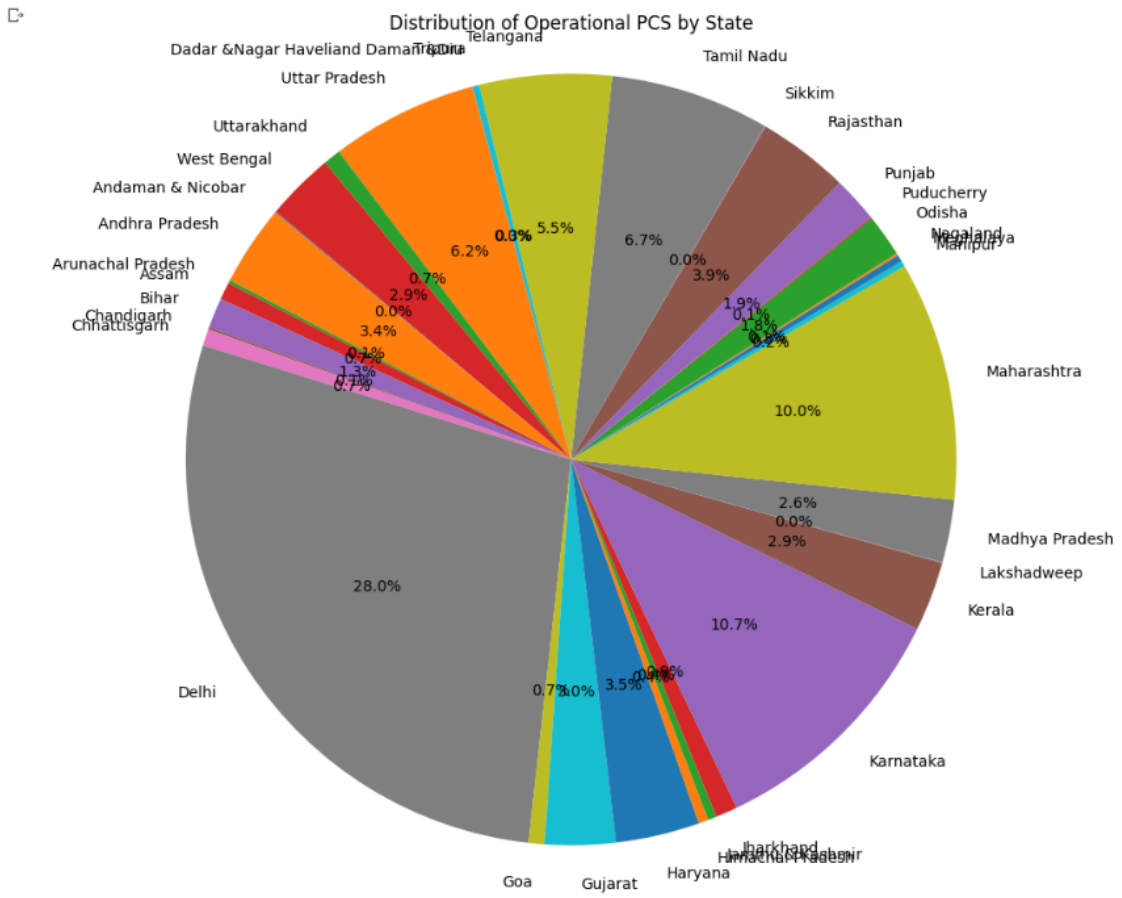

Fig. 6 Distribution of operational PCS by state

Vehicle count State wise - dataset

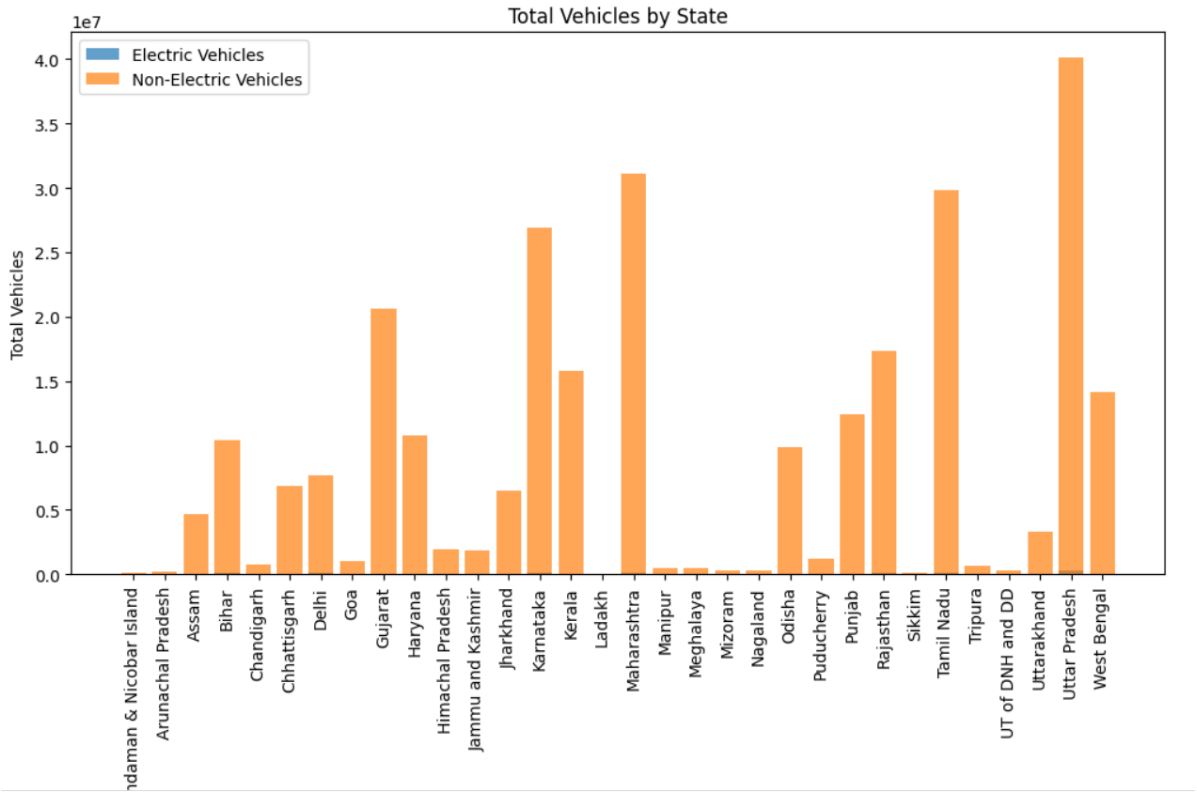

Fig. 7 Total vehicles by state

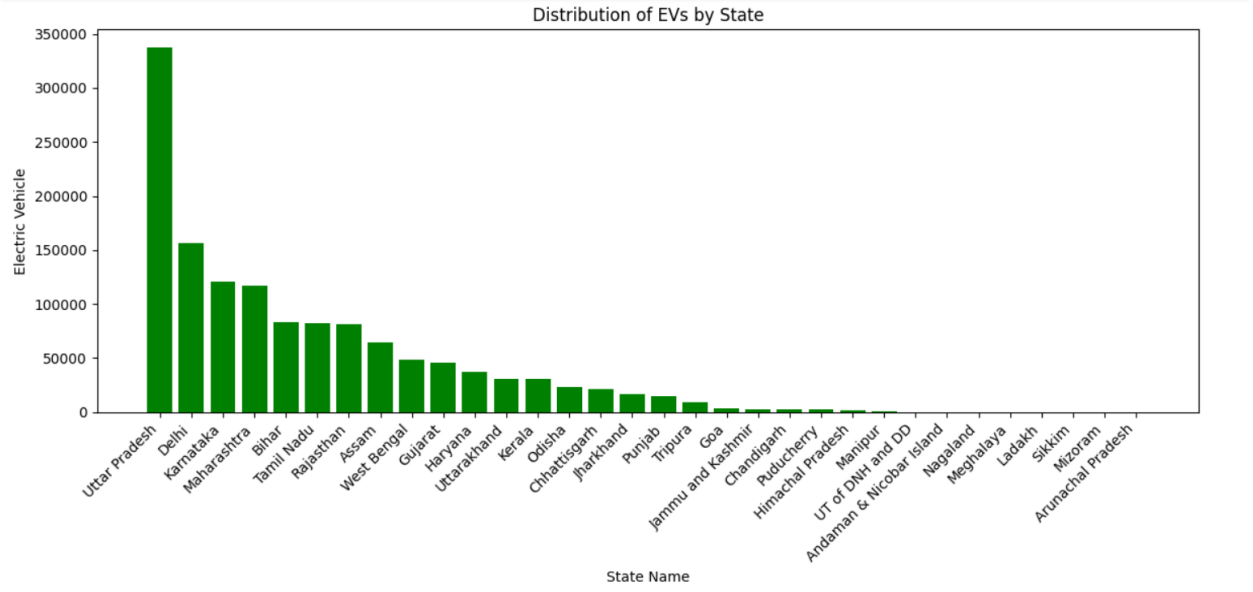

Fig 8 Distribution of EVs by state. Uttar Pradesh shows high no. of EVs.

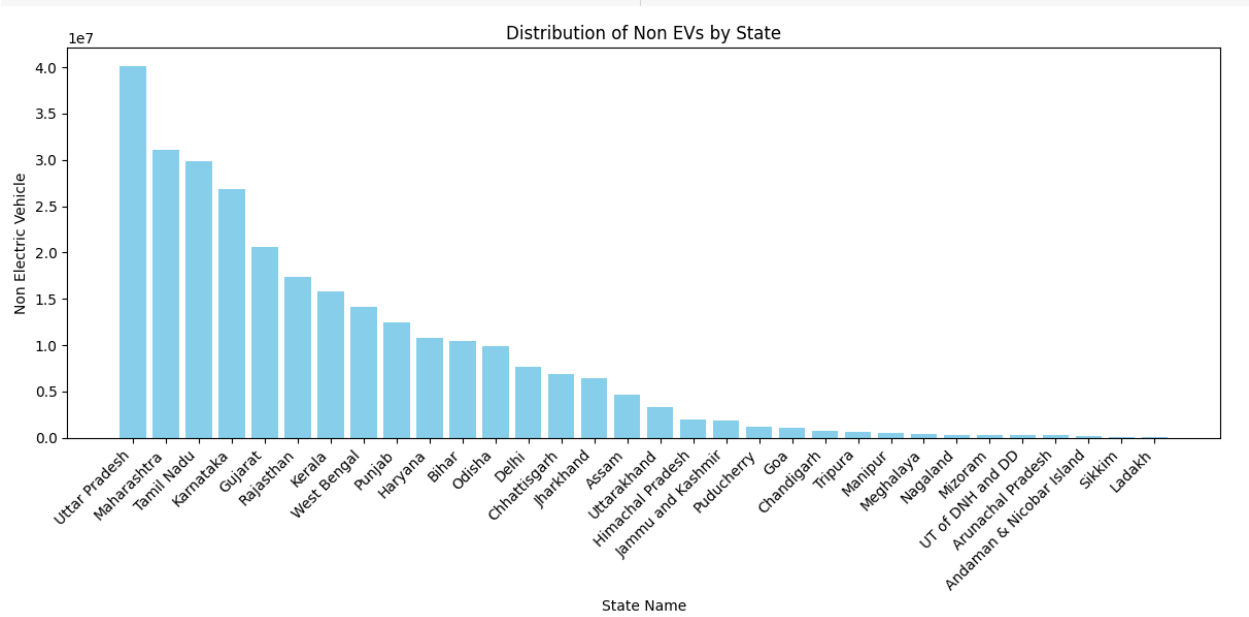

Fig 9 Distribution of Non EVs by state

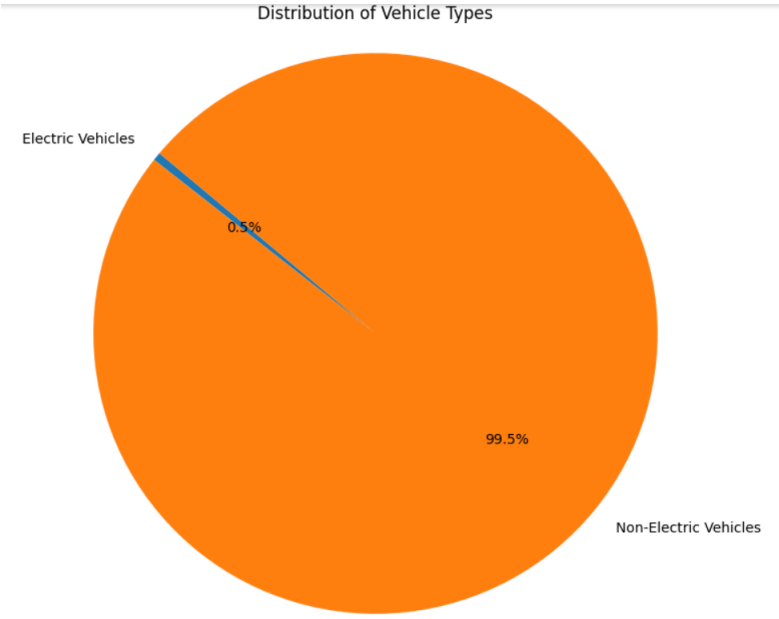

Fig 10. Distribution of Vehicle Types. In India there is a majority of Non Electric Vehicles. EVs only contribute to 0.5%

Electric Car Data - dataset

I have created a column ‘INR’ by taking the product of 'PriceEuro’ and 89.
Dropped the 'PriceEuro’ column as it won't be used in analysis.

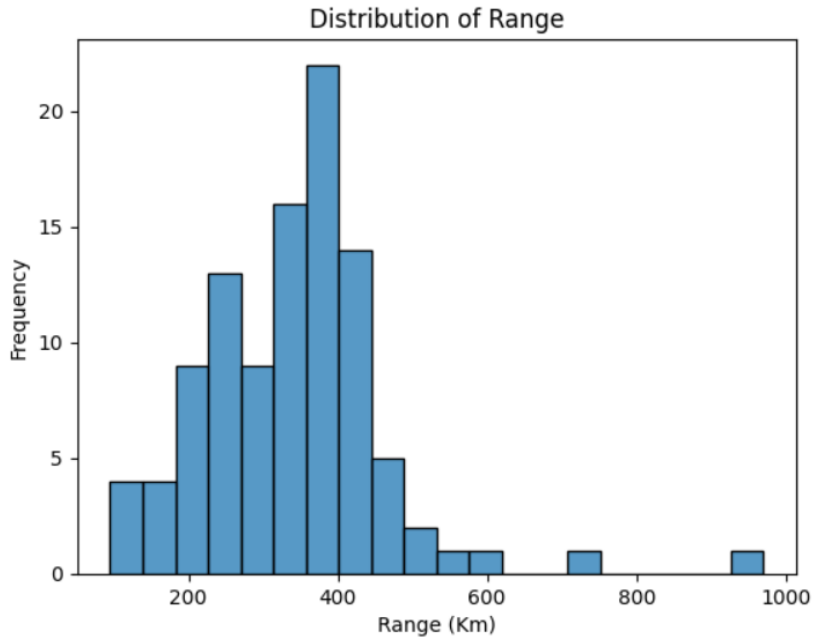

Fig 11. Distribution of Range

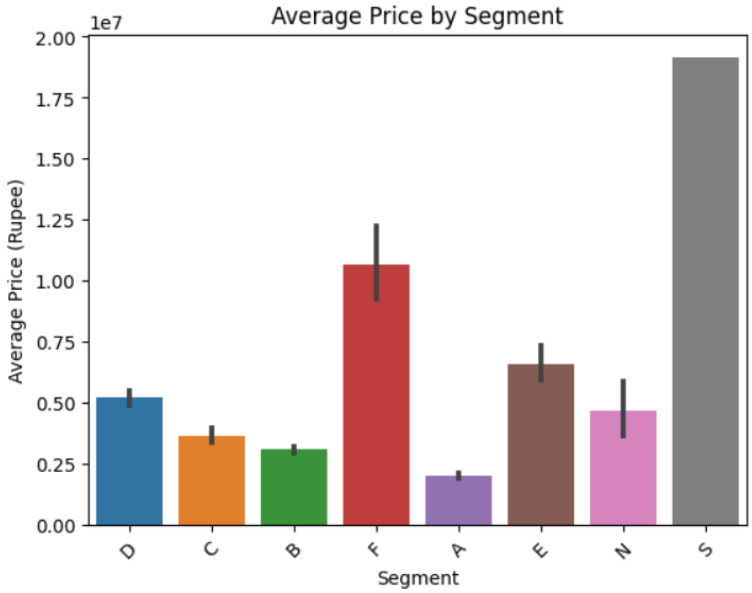

Fig 12. Average Price by Segment. Segment S shows highest Average price

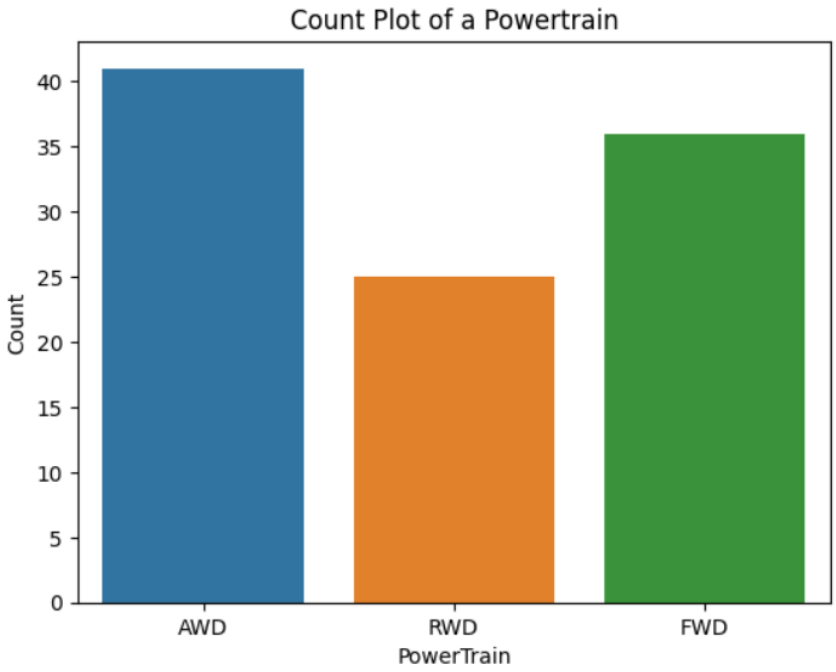

Fig 13. Count plot of powertrain

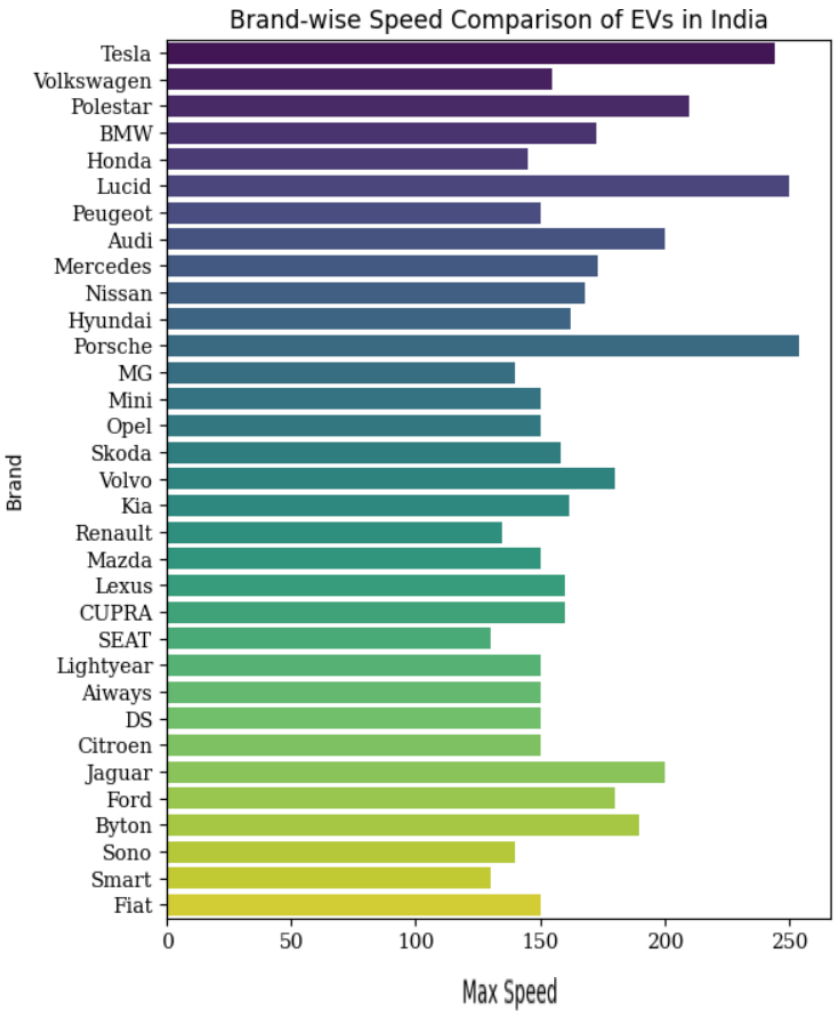

Fig. 14 Brand-wise Speed Comparison of EVs in India. Tesla, Lucid, and Porsche are among the highest max speed compared to others brands

```
Fastest 0-100 Acceleration Car:
    Brand     Model  AccelSec
51  Tesla  Roadster       2.1

Highest Efficiency Car:
        Brand Model  Efficiency_WhKm
48  Lightyear   One              104

Manufacturer with Most Vehicles: Tesla
```

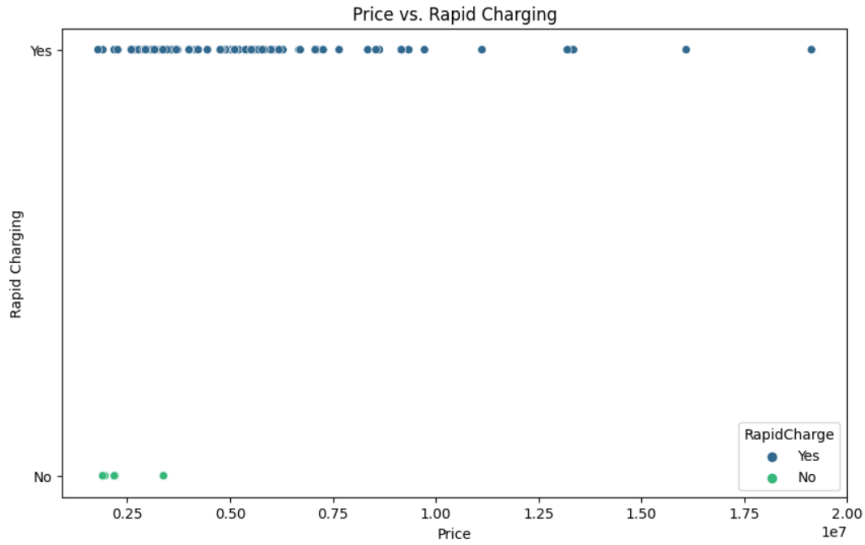

Fig 15 Price vs Rapid Charging

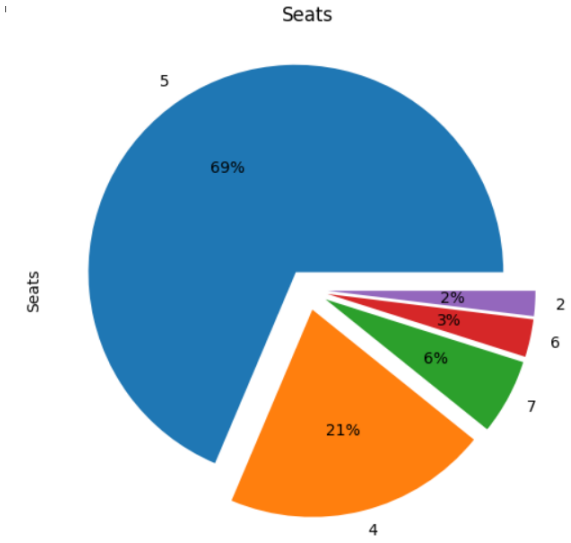

Fig 16. Distribution of seats. 5 seater cars are most preferred

Feature Engineering

- Encoded the categorical features - ‘PowerTrain’ and ‘RapidCharge’.
- Used feature scaling for 'AccelSec', 'TopSpeed_KmH', 'Range_Km', 'Efficiency_WhKm', 'FastCharge_KmH','RapidCharge', 'Seats','PowerTrain', 'INR' columns

Segment Extraction

Principal Component Analysis (PCA)

Performing PCA to reduce the dimensionality of the feature space. This will help in visualizing the data and potentially improve clustering results. Another way of exploring data initially is to compute a principal components analysis, and create a perceptual map. A perceptual map offers initial insights into how attributes are rated by respondents and, importantly, which attributes tend to be rated in the same way. Principal components analysis is not computed to reduce the number of variables. This approach – also referred to as factor-cluster analysis – is inferior to clustering raw data in most instances
Applied PCA, for the features.

Hierarchical Clustering and Dendrogram

After PCA to reduce the dimensionality of the data to nine principal components, it creates a dendrogram plot to visualize hierarchical clustering based on these components. The dendrogram can help identify cluster structures or relationships between data points based on their PCA-transformed features.

K-Means

K-Means is an unsupervised machine learning algorithm that partitions a dataset into K clusters based on the similarity of data points. To find the optimal K for K-Means clustering, "Elbow Method used. Determining the optimal number of clusters (K) based on the reduced feature space after PCA.From the Elbow method, the optimal K value is 4. And performed kmean clustering using k=4.

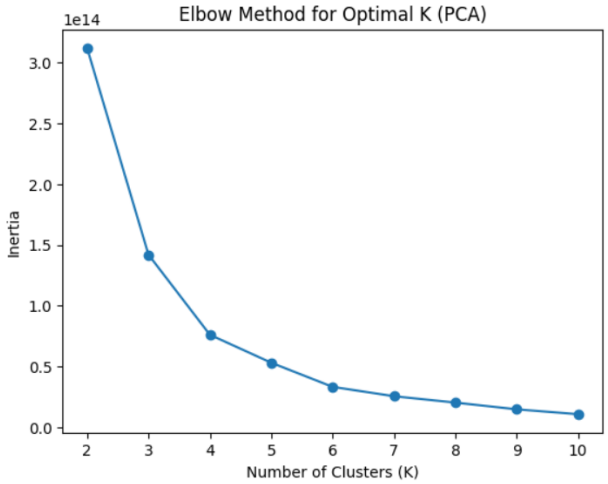

Fig 17. Elbow method

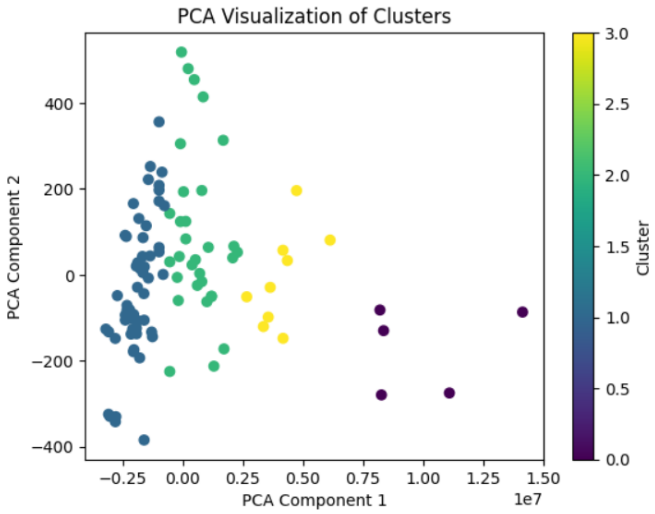

Fig 18 PCA visualization of cluster.

Linear Regression

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

x=df2[['PC1', 'PC2','PC3','PC4','Pc5','PC6', 'PC7','PC8','PC9']]
y=df1['INR']

Splitted the data using a test size of 0.3.

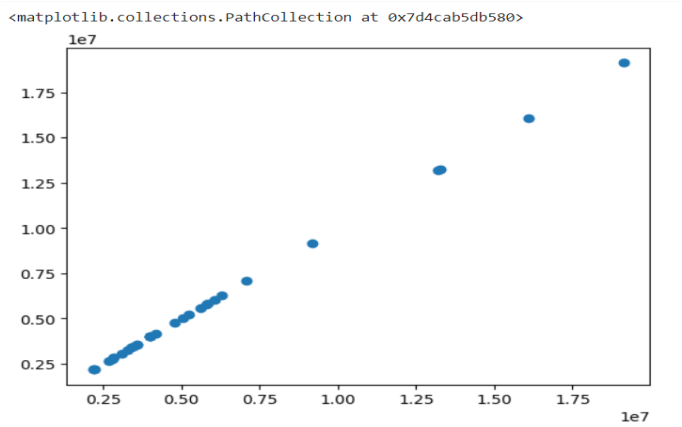

Fig 19.scatter plot of y_testand predictions

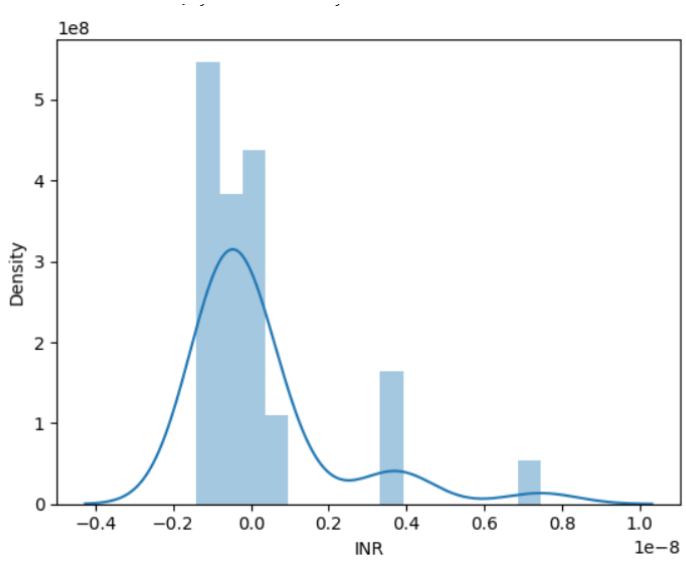

Fig. 20 distplot((y_test-predictions)

```
print('MAE:',metrics.mean_absolute_error(y_test,predictions))
print('MSE:',metrics.mean_squared_error(y_test,predictions))
print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,predictions)))
```

```
MAE: 1.081535893101846e-09
MSE: 3.553385184662169e-18
RMSE: 1.885042488821451e-09
```

```
metrics.mean_absolute_error(y_test,predictions)
```

```
1.081535893101846e-09
```

```
metrics.mean_squared_error(y_test,predictions)
```

```
3.553385184662169e-18
```

```
np.sqrt(metrics.mean_squared_error(y_test,predictions))
```

```
1.885042488821451e-09
```

Selection of target segment

Geographic Segmentation: Analyzing which regions in India have higher EV adoption rates.

- Uttar Pradesh, Delhi, Karnataka and Maharashtra have already started using EVs in a good number
- Delhi contains the highest no. of operational PCS in India. Karnataka and Maharashtra are 2nd and 3rd largest.

Psychographic Segmentation: Understanding consumer values and attitudes toward green technology.

- With use of EVs the air quality will also improve, as it can be seen customers are moving towards green technology.

Behavioral Segmentation: Identifying early adopters and their preferences.

- 5 seaters cars are mostly preferred.
- Manufacturer with Most Vehicles is Tesla
- Rapid charging cars are mostly used.
- Price range is 10L to 60 L

The startup should sell cars which are 5 seaters, cars should have rapid charging feature and price range of 10L-60L, and stores should be located in states like Karnataka, Maharashtra, Delhi and Uttar Pradesh as they have high no. of operational power stations.

Customizing the Marketing Mix

The marketing mix, often referred to as the 4Ps (Product, Price, Place, and Promotion), plays a crucial role in shaping the success of electric vehicles (EVs) in the market. To effectively market EVs, manufacturers and stakeholders must tailor these elements to meet the unique demands of the EV industry. Here's a breakdown of how to customize the marketing mix in the EV sector:

- Product: Tailoring electric vehicle offerings to meet specific consumer needs, whether it's affordable models for the mass market or high-performance options for enthusiasts.
- Price: Strategically pricing EVs to compete with conventional vehicles, considering government incentives, subsidies, and total cost of ownership to make them more attractive.
- Place: Expanding distribution channels to ensure easy access to EVs, including online sales, dealerships, and partnerships with charging infrastructure providers.
- Promotion: Implementing marketing campaigns that highlight the environmental benefits, cost savings, and advanced technology of EVs, targeting both eco-conscious consumers and those seeking innovation.

The MOST OPTIMAL MARKET SEGMENTS to open in the market as per my Market Research and Segmentation is Rapid Charging feature of cars. This will help in reducing the long hour wait for charging. Also cost will also increase wrt to rapid charging feature and also there are more operation power stations being installed in the states.

Conclusion

Demand for EVs has grown markedly over the past decade thanks to heightened environmental concerns, greater availability of models, increased cost competitiveness with conventional gas vehicles, and improved vehicle ranges. These factors are expected to continue to drive increased adoption.

GitHub Link:

https://github.com/DivyaGazinkar/Machine-Learning-Internship-2023/tree/main/EV%20Market