Dheeraj Nair

**Electric Vehicle Market in India**

**Market Segment**

https://github.com/Dheeraj1002/Fynn-labs-code-conversion

****

We will not stop until every car on the road is electric

**-Elon Musk**

**Conceptual -**

Market segmentation becomes a crucial tool for evolving transportation technology such as electric vehicles (EVs) in emerging markets to explore and implement for extensive adoption. EVs adoption is expected to grow phenomenally in near future as low emission and low operating cost vehicle, and thus, it drives a considerable amount of forthcoming academic research curiosity.

**Market Analysis Using Geographic Datasets: Vehicles and Charging Stations**

This analysis aims to understand market trends through geographic datasets that encompass the number of vehicles per state and the distribution of charging stations. By examining these data points, we can identify patterns in vehicle ownership, assess the availability and accessibility of charging infrastructure, and evaluate regional disparities in electric vehicle adoption.

**Key Objectives:**

1. **Vehicle Distribution Analysis**:
   * Assess the number of vehicles registered in each state to understand regional demand and preferences for different vehicle types, including electric and hybrid models.
   * Assess the total number of vehicles registered in each state, distinguishing between electric and non-electric vehicles. This will help in understanding regional demand and preferences for different vehicle types.
2. **Charging Station Mapping**:
   * Analize the location and density of charging stations across states to evaluate the adequacy of infrastructure supporting electric vehicle usage.
3. **Correlation Studies**:
   * Investigate the relationship between the number of vehicles and the availability of charging stations, identifying areas that may require infrastructure investment or incentives to promote electric vehicle adoption.

**Data Collection**

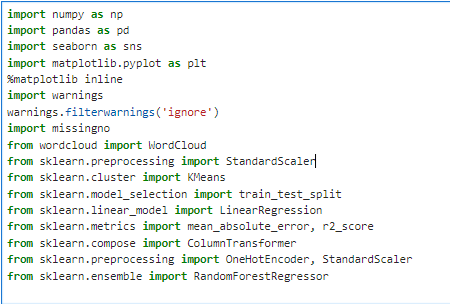
Data Collection The data has been collected manually, and the sources used for this process are listed below:

• <https://www.kaggle.com/datasets>

• <https://data.gov.in/>

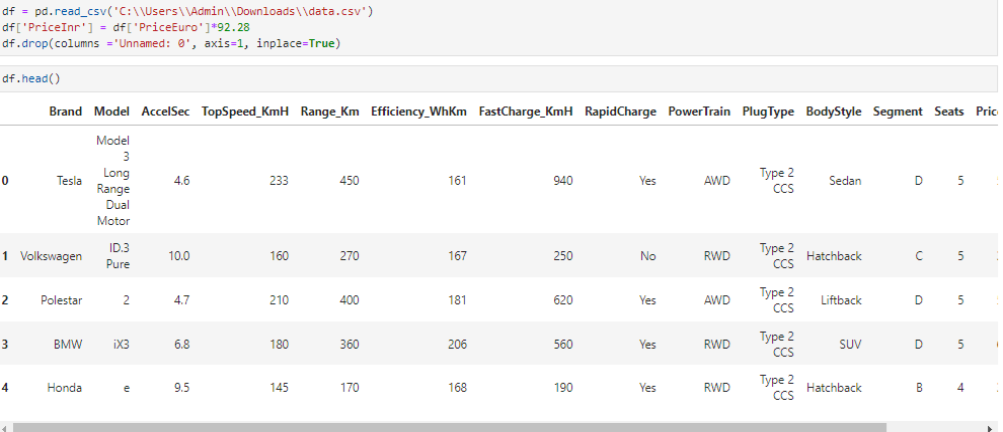
**Implementation**

**Packages/Tools used**

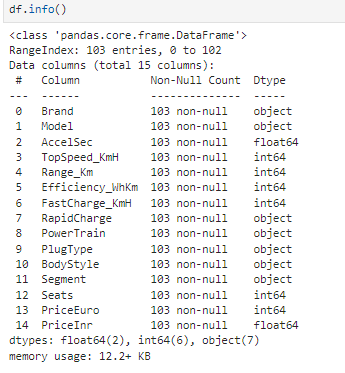
****

* **NumPy**: A library for numerical computations in Python, providing support for arrays, matrices, and a collection of mathematical functions.
* **Pandas**: A data manipulation and analysis library, offering data structures like Data Frames for handling structured data.
* **Seaborn**: A statistical data visualization library built on top of Matplotlib, providing a high-level interface for drawing attractive statistical graphics.
* **Matplotlib**: A plotting library for Python, used for creating static, animated, and interactive visualizations in Python.
* **Warnings**: A module to manage warnings in Python. It's used to suppress warnings that may clutter the output.
* **Filter Warnings**: This line tells Python to ignore all warnings, which can be useful in notebooks or scripts where warnings are not relevant.
* **Missingno**: A library for visualizing missing data in Python, providing an easy way to visualize and understand data completeness.
* **WordCloud**: A library to generate word clouds from text data, visually representing word frequency.
* **StandardScaler**: A preprocessing tool from Scikit-learn that standardizes features by removing the mean and scaling to unit variance.
* **KMeans**: An unsupervised machine learning algorithm in Scikit-learn for clustering data into k distinct groups based on feature similarity.
* **train\_test\_split**: A utility function from Scikit-learn to split arrays or matrices into random train and test subsets.
* **LinearRegression**: A class in Scikit-learn for performing linear regression, modeling the relationship between dependent and independent variables.
* **mean\_absolute\_error**: A function to calculate the mean absolute error of a regression model.
* **r2\_score**: A function to compute the coefficient of determination (R²) regression score function
* **ColumnTransformer**: A class that allows you to apply different preprocessing steps to different columns of your dataset.
* **OneHotEncoder**: A preprocessing technique for converting categorical variables into a format that can be provided to ML algorithms.
* **RandomForestRegressor**: An ensemble learning method in Scikit-learn for regression that fits a number of decision tree regressors on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control overfitting.

**Data-Preprocessing**



Added price Inr column which shows the Inr price of the vehicle



**>> Columns details**

**Brand, Model: Identifiers for the car manufacturer and the specific model.**

**AccelSec: Acceleration time from 0 to 100 km/h, a performance metric.**

**TopSpeed\_KmH: The maximum speed the vehicle can achieve.**

**Range\_Km: Distance the vehicle can cover on a full charge.**

**Efficiency\_WhKm: Energy efficiency, measured in watt-hours per kilometer.**

**FastCharge\_KmH: Charging speed in terms of kilometers per hour.**

**RapidCharge: A binary feature indicating whether rapid charging is supported.**

**PowerTrain, PlugType: Technical specifications related to the car’s drivetrain and charging type.**

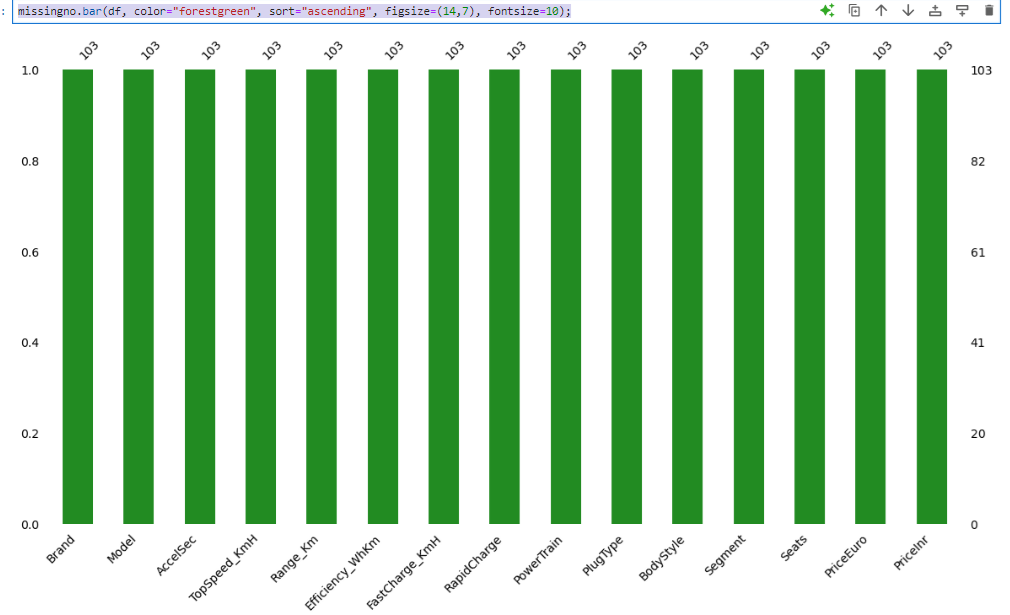
**BodyStyle, Segment: The car’s design and market segment (e.g., sedan, SUV).**

**Seats: Number of seats in the vehicle.**

**PriceEuro: Price of the vehicle in Euros.**

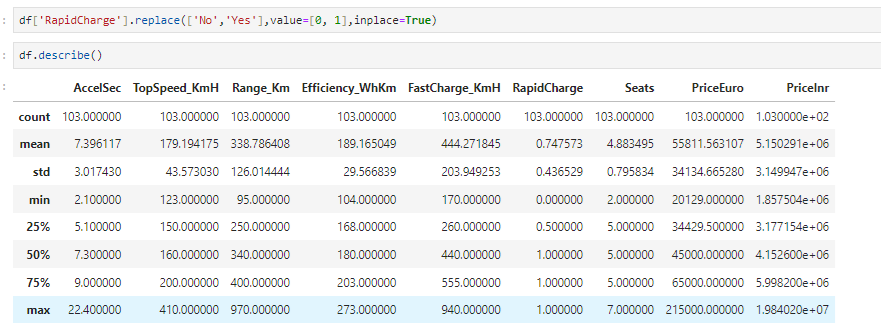
**PriceInr : Price of the vehicle in Rupees**[**¶**](http://localhost:8888/notebooks/Documents/Python%20class/EV%20Market%20in%20India.ipynb?#PriceInr-:--Price-of-the-vehicle-in-Rupees)

**Missing Data**

****

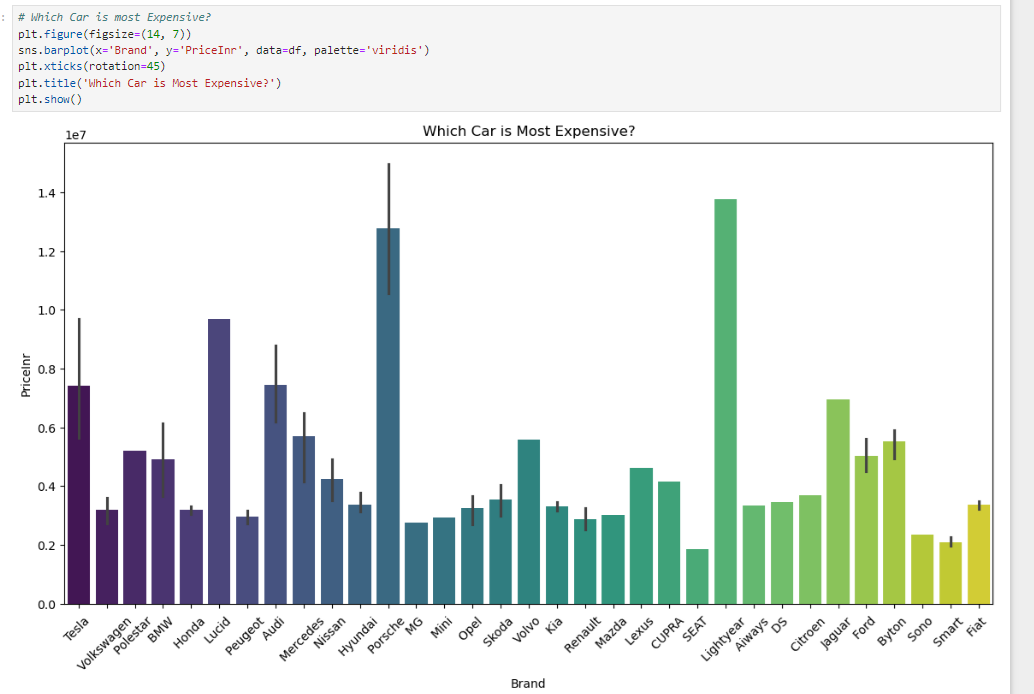
There are no missing Data

**Descriptive Analysis**

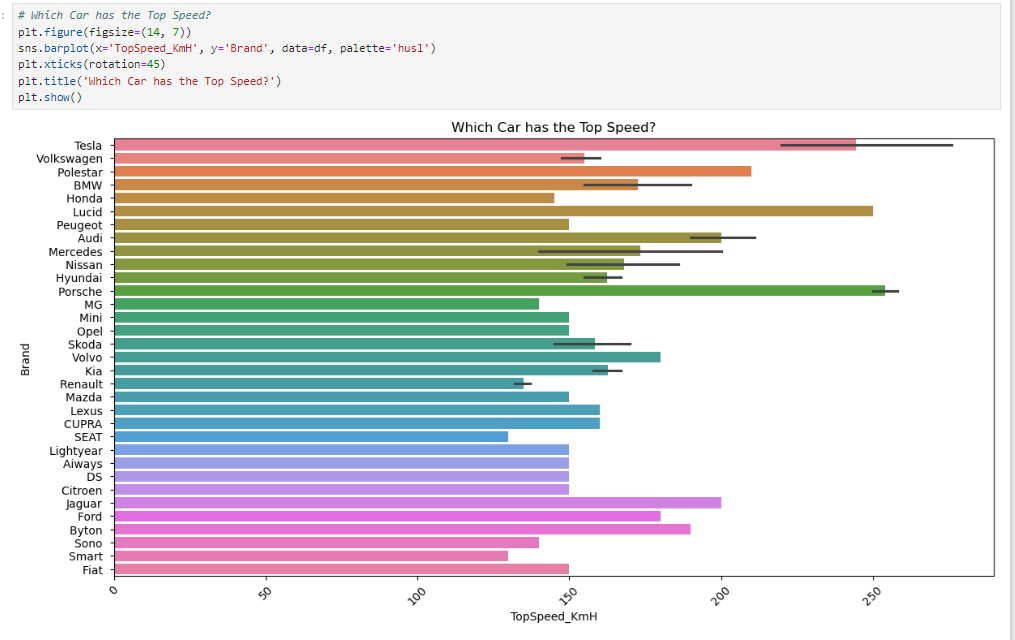


Descriptive data shows the mean, Standard deviation etc

**Most Expensive car**

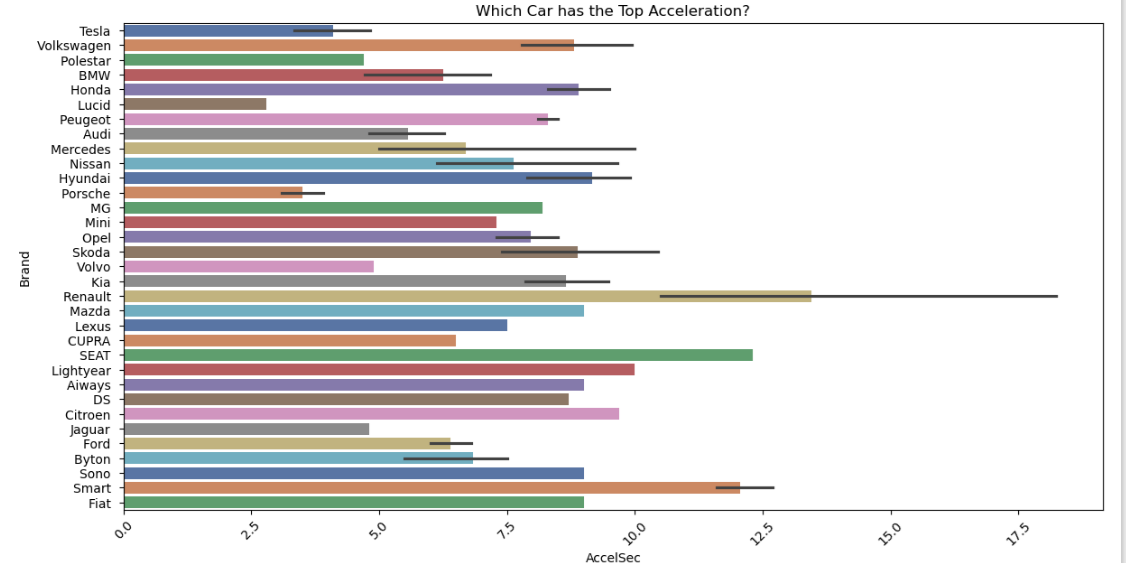
****

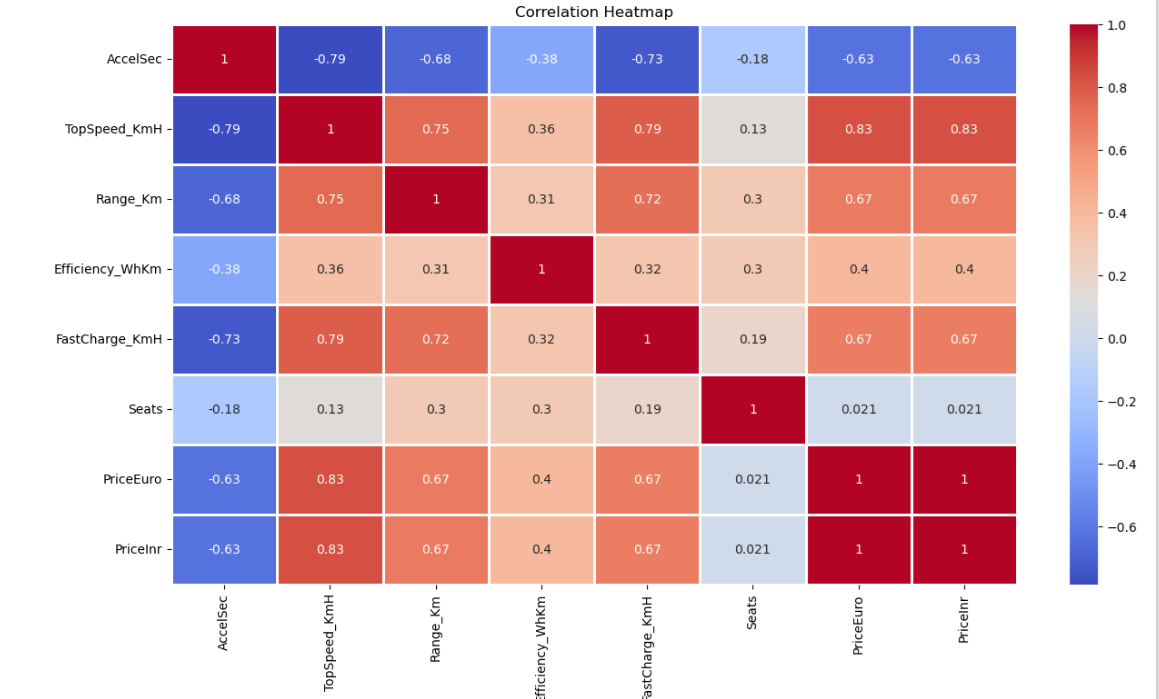
Porsche is the most expensive and seat in the cheapest.

**Which car has the top speed **

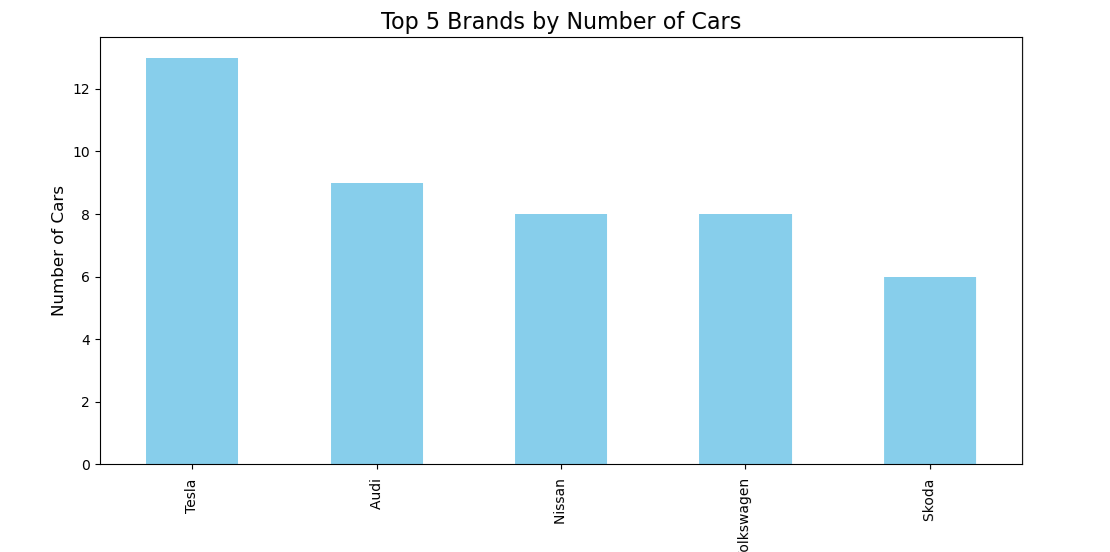
Tesla being the top seat in the low side in speed.

# Which Car has the Top Acceleration? # Acceleration time from 0 to 100 km/h

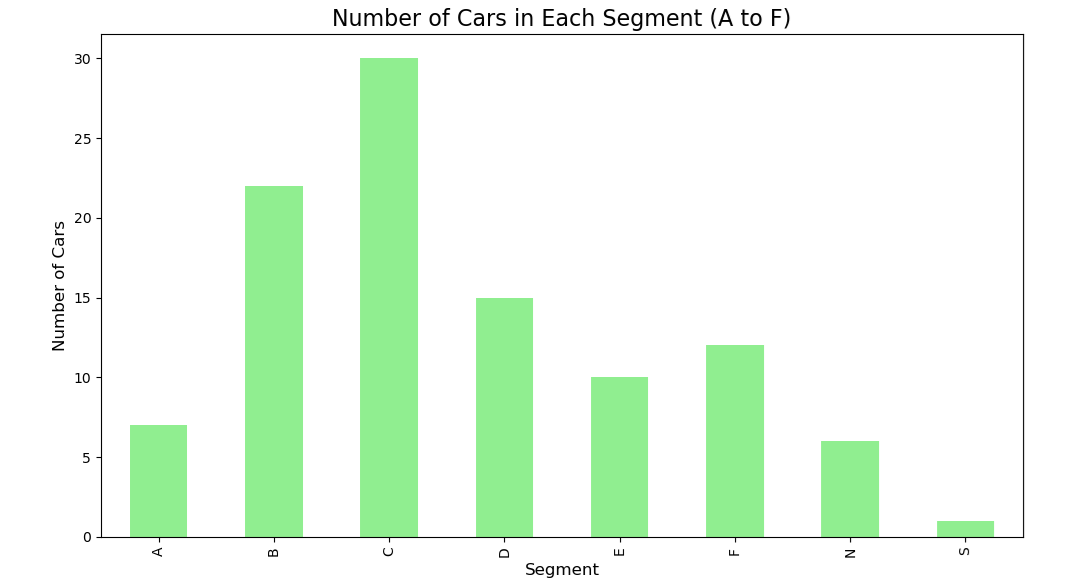
****

**Correlation of the data using Heatmap**

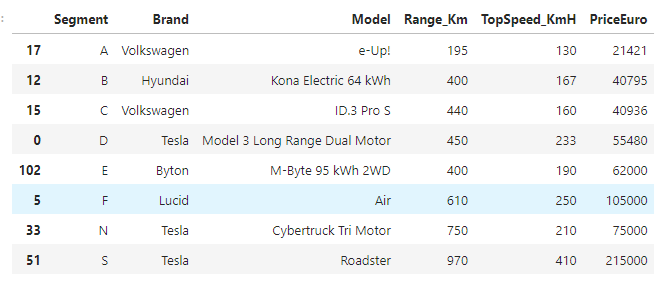
**# Plot the top 5 brands in a bar chart**

****

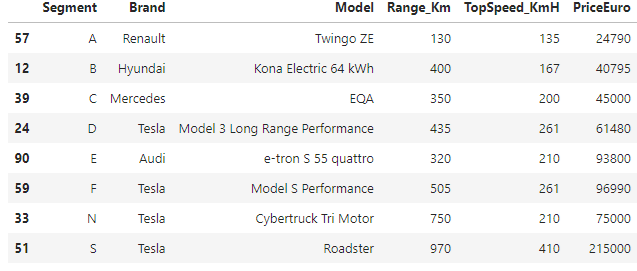
**Segment of the cars**

****

**# Finding the best car in each segment through Range\_Km**

****

**# Finding the best car in each segment through TopSpeed\_KmH**

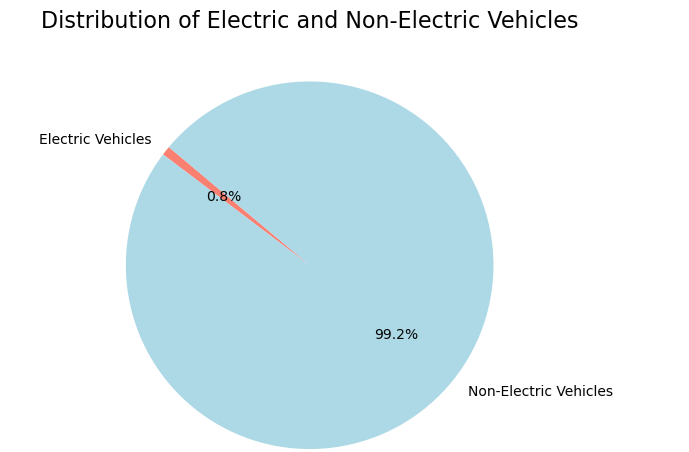
****

**Here is a summary of the market segment analysis:**

1. **Segment A (Low-end vehicles):**
   * Average Price: €22,693
   * Average Range: 143 km
   * Average Acceleration: 12.2 seconds (0-100 km/h)
   * Average Top Speed: 131 km/h
   * Efficiency: 169 Wh/km
   * Seats: 3.4
2. **Segment B (Entry-level compact vehicles):**
   * Average Price: €34,799
   * Average Range: 266 km
   * Average Acceleration: 8.7 seconds
   * Top Speed: 151 km/h
   * Efficiency: 169 Wh/km
   * Seats: 4.7
3. **Segment C (Mid-range compact vehicles):**
   * Average Price: €41,199
   * Average Range: 329 km
   * Acceleration: 7.8 seconds
   * Top Speed: 164 km/h
   * Efficiency: 181 Wh/km
   * Seats: 4.9
4. **Segment D (Upper-mid-range vehicles):**
   * Average Price: €58,488
   * Average Range: 400 km
   * Acceleration: 5.4 seconds
   * Top Speed: 202 km/h
   * Efficiency: 187 Wh/km
   * Seats: 5.3
5. **Segment E (High-end vehicles):**
   * Average Price: €74,269
   * Average Range: 346 km
   * Acceleration: 5.9 seconds
   * Top Speed: 197 km/h
   * Efficiency: 238 Wh/km
   * Seats: 5
6. **Segment F (Luxury vehicles):**
   * Average Price: €119,691
   * Average Range: 455 km
   * Acceleration: 4.0 seconds
   * Top Speed: 243 km/h
   * Efficiency: 194 Wh/km
   * Seats: 4.8
7. **Segment S (Supercars):**
   * Average Price: €215,000
   * Range: 970 km
   * Acceleration: 2.1 seconds
   * **Top Speed: 410 km/h**
   * Efficiency: 206 Wh/km
   * Seats: 4

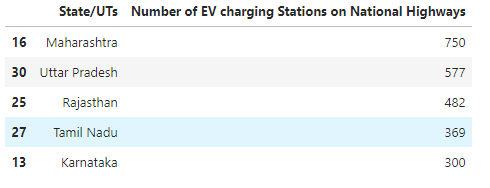
**Insights**

* + Supercars in Segment S have extreme performance and high range, while low-end vehicles in Segment A are affordable but offer modest performance.
  + Segment F (luxury cars) strikes a balance between range, acceleration, and top speed, at a premium price.
  + Segment D vehicles provide a good middle ground in terms of price, range, and performance for the average consumer.



Only 0.8 % of the Indian vehicles are Electric Vehicles

**Top 5 Charging station across India**



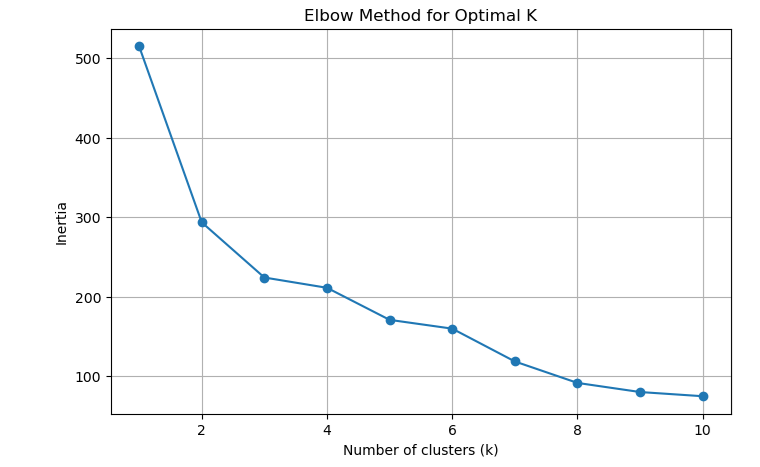
To perform **clustering** and **regression** tasks based on the dataset, we will break the analysis into two sections:

**1. Clustering Analysis**

We can use clustering to identify different groups of vehicles based on attributes like **Price**, **Range**, **Efficiency**, **Top Speed**, and **Acceleration**. Clustering will help uncover patterns in the vehicle market, revealing natural groupings that might correspond to budget cars, performance vehicles, or luxury cars.

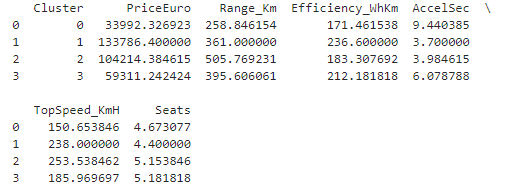
**2. Regression Analysis**

In a regression task, we will try to predict the vehicle's **Price** based on key features such as **Range**, **Acceleration**, **Top Speed**, and other relevant attributes. This will allow us to understand how various factors influence the price of a vehicle.



The **Elbow Method** plot shows the optimal number of clusters where the inertia (sum of squared distances to the nearest cluster center) begins to flatten, typically indicating the best balance between cluster quantity and quality.

Cluster Analysis



* Cluster 0

This cluster represents vehicles with a moderate price and range. The efficiency is relatively high, and the acceleration is slower compared to other clusters. This may indicate vehicles targeted at budget-conscious consumers who prioritize efficiency and moderate range over performance.

* Cluster 1

This cluster features high-end vehicles with a high price and good range. The acceleration is quick, indicating a focus on performance. These vehicles may cater to luxury consumers who value speed and range over cost.

* Cluster 2

This cluster showcases vehicles that balance performance, range, and price. With a substantial range and good acceleration, these vehicles are likely aimed at consumers looking for a mix of comfort, practicality, and performance.

* Cluster 3

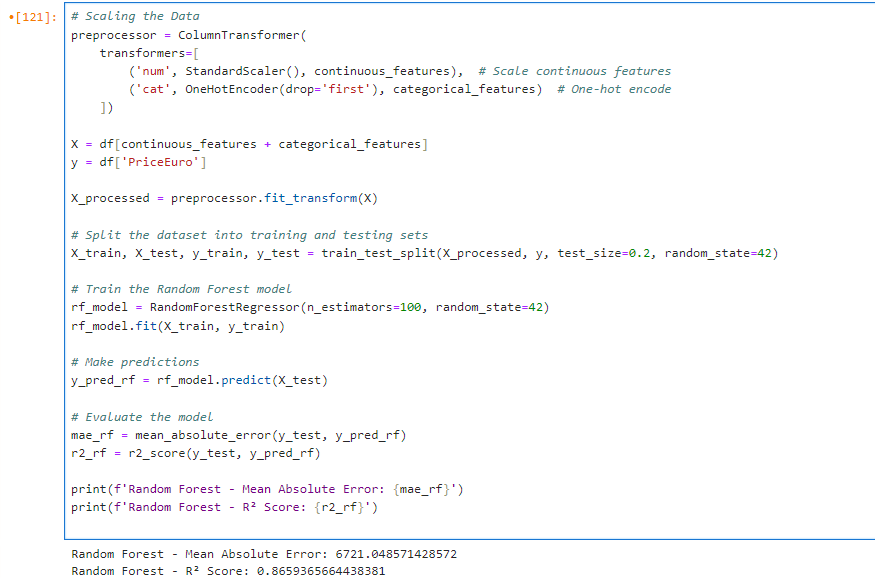
Vehicles in this cluster have a moderate price and good range, with decent acceleration. They might appeal to consumers looking for a reliable and comfortable vehicle without venturing into the luxury market.

**Random Forest**

**Definition: Random Forest is an ensemble machine learning algorithm that utilizes multiple decision trees to improve predictive accuracy and control overfitting. It is widely used for both classification and regression tasks.**

**How It Works:**

* **Ensemble Learning: Random Forest builds multiple decision trees during training and merges their outputs. Each tree is trained on a random subset of the data and features, which introduces diversity and improves generalization.**
* **Bootstrap Aggregating (Bagging): Each tree is trained on a random sample of the dataset (with replacement), helping to reduce variance.**
* **Feature Randomness: For each split in the decision trees, a random subset of features is considered, which enhances the model’s robustness and reduces overfitting.**

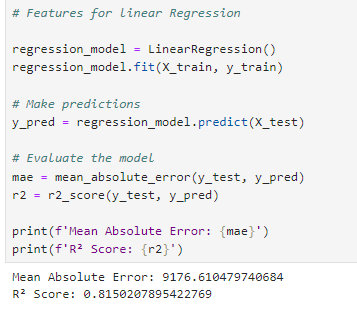
****

**Linear Regression**

**Definition**: Linear Regression is a statistical method used for modeling the relationship between a dependent variable and one or more independent variables. The relationship is assumed to be linear.

**How It Works**:

* **Modeling**: Linear Regression fits a straight line (or hyperplane in higher dimensions) to the data that minimizes the sum of the squared differences (errors) between the observed values and the values predicted by the model.
* **Equation**: The general form of the linear regression equation is: y=b0+b1x1+b2x2+...+bnxn+ϵy = b\_0 + b\_1x\_1 + b\_2x\_2 + ... + b\_nx\_n + \epsilony=b0​+b1​x1​+b2​x2​+...+bn​xn​+ϵ where yyy is the dependent variable, b0b\_0b0​ is the intercept, b1,b2,...,bnb\_1, b\_2, ..., b\_nb1​,b2​,...,bn​ are the coefficients for the independent variables x1,x2,...,xnx\_1, x\_2, ..., x\_nx1​,x2​,...,xn​, and ϵ\epsilonϵ is the error term.

****

**Conclustion**

Random Forest demonstrated superior predictive performance, evidenced by a Mean Absolute Error (MAE) of approximately 6271.04 and an R² score of about 0.86. This indicates that the Random Forest model captured a substantial portion of the variance in the data, making it effective for predicting the number of EV charging stations based on various features. The model's ensemble nature allowed it to account for non-linear relationships and interactions between variables, which is particularly beneficial given the complexities of real-world data.

Linear Regression, while straightforward, produced less accurate predictions in this context. Its reliance on the assumption of linear relationships may have limited its ability to effectively model the underlying data patterns, especially in a dataset with potentially complex interactions among predictors.

**Marketing mix**

• Price: refers to the value that is put for a product. It depends on segment targeted, ability of the companies to pay, ability of customers to pay supply - demand and a host of other direct and indirect factors.

• Product: refers to the product actually being sold – In this case, the service. The product must deliver a minimum level of performance; otherwise even the best work on the other elements of the marketing mix won’t do any good.

• Place: refers to the point of sale. In every industry, catching the eye of the consumer and making it easy for her to buy it is the main aim of a good distribution or ’place’ strategy. Retailers pay a premium for the right location. In fact, the mantra of a successful retail business is ’location, location, location’.

• Promotion: this refers to all the activities undertaken to make the product or service known to the user and trade. This can include advertising, word of mouth, press reports, incentives, commissions and awards to the trade. It can also include consumer schemes, direct marketing,contest and prizes

All the elements of the marketing mix influence each other. They make up the business plan for a company and handle it right, and can give it great success. The marketing mix needs a lot of understanding, market research and consultation with several people, from users to trade to manufacturing and several others

Thank you