MARKET SEGMENTATION ANALYSIS
OF ELECTRIC VEHICLE (EV) MARKET IN INDIA

Project 02
Data Analysis Report

Sona Teresa Jossy

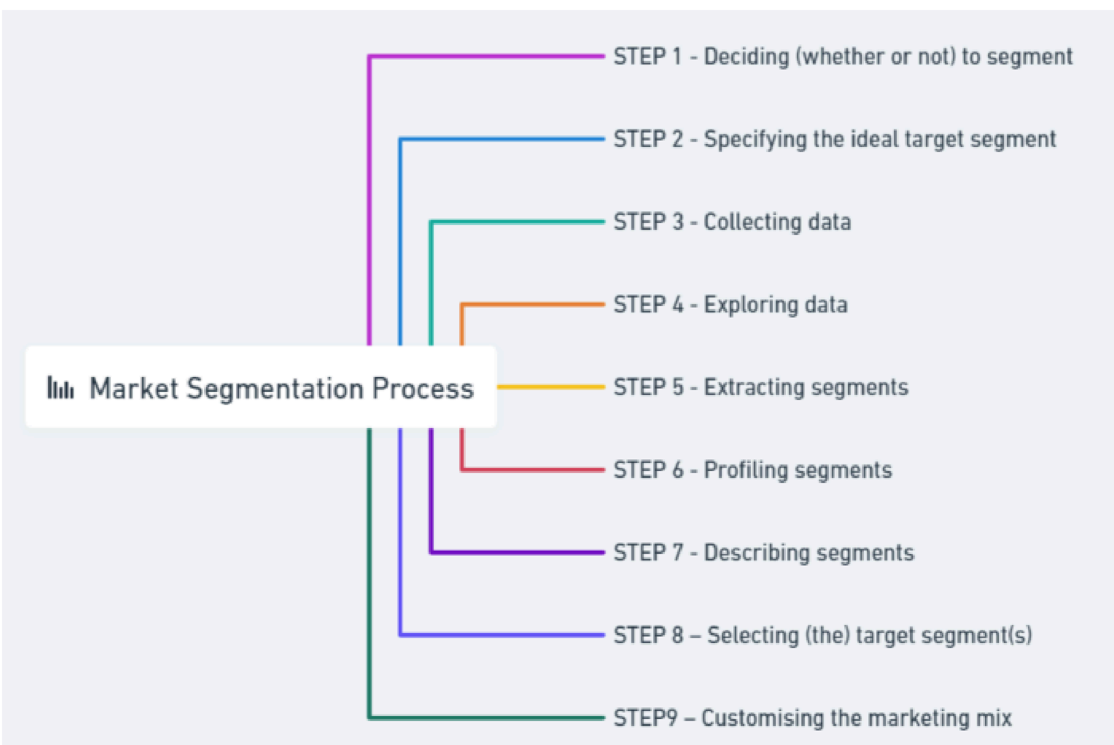
January 2024

Machine Learning Internship Program
Feynn Labs Services

BACKGROUND

The current Electric Vehicle (EV) scenario in India is experiencing a notable transformation, fueled by a heightened focus on sustainability, technological advancements, and government initiatives. The shift towards EVs is gaining momentum due to environmental consciousness, escalating fuel prices, and a drive towards cleaner mobility. Government policies, particularly the National Electric Mobility Mission Plan (NEMMP) and Faster Adoption and Manufacturing of Hybrid and Electric Vehicles (FAME) initiatives, play a crucial role in incentivizing EV adoption and developing charging infrastructure.

Despite these advancements, challenges like limited charging infrastructure, battery concerns, and affordability persist. To thrive in this evolving market, our startup aims to employ segmentation analysis, considering geographic, demographic, psychographic, and behavioral factors. Urban centers exhibit distinct adoption patterns, influenced by population density and economic development. The younger, tech-savvy generation is increasingly open to EVs, driven by shifting consumer attitudes and environmental awareness. Navigating this complex landscape strategically positions our startup to seize early market opportunities and contribute to India's electric mobility transformation.



PROBLEM STATEMENT

Electric Vehicle (EV) Market Entry Strategy for an Indian Startup

The EV startup faces a critical decision in determining the optimal vehicle/customer space for its EV development in the Indian market. The primary challenge lies in conducting a comprehensive segmentation analysis of the Indian EV market, with a specific focus on identifying and targeting segments most likely to adopt electric vehicles. This entails traditional parameters like geographic, demographic, psychographic, and behavioral factors, as well as specific categories based on available data.

Our primary objective is to strategically establish an early market presence by pinpointing the most suitable location in India, aligned with the Innovation Adoption Life Cycle. This requires a nuanced understanding of technology adoption variations across regions and identification of relevant demographic, psychographic, and behavioral factors using datasets like EV market data and vehicle usage statistics.

Acknowledging potential data challenges, our strategy must enable decision-making in the face of incomplete datasets, ensuring accuracy and impartiality. The report will encompass Fermi Estimation, data sources identification, preprocessing steps, machine learning techniques for segment extraction, and profiling of potential segments. It will conclude with the selection of the target segment, customized marketing mix, estimated customer base and profit, offering clear insights into optimal market segments based on robust segmentation analysis. The accompanying GitHub profile will provide well-documented codes and datasets.

DATA COLLECTION

In order to craft a robust entry strategy into the Indian EV market, a comprehensive and data-driven approach is essential. The following datasets serve as critical sources for gathering insights into distinct market variables:

1. EV Market Data: EVData_01.csv

2. Charging Stations Data: EVData_02.csv

3. EV Customer Data: EVData_03.csv

These datasets are pivotal in shaping the Electric Vehicle (EV) market entry strategy in India. EV Market Data offers insights into vehicle performance, trends, and competition, addressing queries on EV popularity, pricing, and regional preferences. Simultaneously, Charging Stations Data illuminates infrastructure accessibility, distribution, and collaboration opportunities. The EV Customer dataset delves into diverse factors influencing adoption, facilitating targeted marketing, product development, and market segmentation. This comprehensive approach guides decisions on vehicle offerings, market segments, and infrastructure development, ensuring a strategic and successful entry into the burgeoning Indian EV market.

The **EV Market dataset** provides valuable insights into the performance, features, and specifications of different electric vehicle models. Here's breakdown of the data variables:

- *Brand*: Represents the manufacturer or brand of the electric vehicle.
- *Model*: Specifies the model name or identifier of the electric vehicle.
- *Accel (Acceleration)*: Indicates the acceleration time from 0 to 100 km/h (or 0 to 60 mph) in seconds.
- *TopSpeed*: Represents the maximum speed the electric vehicle can achieve in kilometers per hour (km/h).
- *Range*: Denotes the driving range of the electric vehicle on a single charge in kilometers.
- *Efficiency*: Measures the energy efficiency of the electric vehicle, in watt-hours per kilometer (Wh/km).
- *FastCharge*: Indicates the fast-charging speed in kilometers per hour.
- *RapidCharge*: Describes whether rapid charging is possible for the electric vehicle.
- *PowerTrain*: Specifies the type of powertrain used in the electric vehicle (e.g., All Wheel Drive, Rear Wheel Drive, Front Wheel Drive).
- *PlugType*: Represents the type of plug used for charging (e.g., Type 2 CCS).
- *BodyStyle*: Describes the body style or configuration of the electric vehicle (e.g., Sedan, Hatchback, SUV).
- *Segment*: Categorizes the electric vehicle into a specific market segment (e.g., A, B, C, D).
- *Seats*: Indicates the number of seats available in the electric vehicle.
- *PriceEuro*: Represents the price of the electric vehicle in Euros.

The **Charging Stations dataset** will be instrumental in assessing the coverage and accessibility of charging infrastructure. Here's an explanation of the data variables:

- *Name*: Represents the name or identifier of the charging station.
- *State*: Specifies the state in which the charging station is located.
- *City*: Indicates the city where the charging station is situated.
- *Address*: Provides the detailed address of the charging station, including street information.
- *Latitude*: Specifies the latitude coordinates of the charging station's location.
- *Longitude*: Specifies the longitude coordinates of the charging station's location.
- *Type*: Describes the type of charging station, potentially indicating the charging speed or capabilities.

The **EV Customer dataset** captures valuable insights related to individual preferences and perceptions towards Electric Vehicles (EVs) in India. Here's a breakdown of the key variables:

- *Age:* Represents the age of respondents.
- *City:* Indicates the city of residence.
- *Profession:* Specifies the respondents' professions.
- *Marital Status:* Provides information on marital status.
- *Education:* Captures the educational background of respondents.
- *No. of Family Members:* Indicates family size.
- *Annual Income:* Provides an understanding of respondents' income levels.
- *Preference for EV Replacement:* Captures respondents' willingness to replace all vehicles with EVs.
- *Type of EV Preferred:* Specifies the preferred type of EV, such as SUV, Hatchback, or Sedan.
- *Perception of EVs' Economical Aspect:* Captures respondents' perceptions regarding the economic viability of EVs.
- *Current Vehicle Brand:* Indicates the brand of the vehicle currently owned.
- *Budget for EV:* Specifies the amount respondents are willing to spend on an EV.
- *Wheel Preference in EV:* Captures the preference for the number of wheels in EVs.
- *Expectation of EVs Replacing Fuel Cars in India:* Gauges respondents' expectations regarding the timeline for EVs replacing traditional fuel cars in India.

DATA PRE-PROCESSING

EV Market Data:

The initial step involves cleaning and preparing the EV Market Data. This includes handling missing values, ensuring consistency in units, and addressing outliers.

Variables like "PriceEuro" may need currency conversion for uniformity. Categorical variables like "BodyStyle" may undergo encoding for machine learning models. Feature scaling may be applied to numeric variables for better model performance. Additionally, exploratory data analysis (EDA) techniques can identify patterns and correlations, providing valuable insights for subsequent analysis.

Charging Stations Data:

Data preprocessing for Charging Stations involves validating geographical coordinates, handling missing values in the "Type" variable, and standardizing station names. Geospatial data might require validation to ensure accuracy in location-based analyses. Additionally, clustering algorithms could be employed to categorize stations based on characteristics such as charging speed. Cleaning the dataset ensures accurate representation and facilitates effective utilization for market insights.

EV Customer Dataset:

Cleaning the EV Customer dataset involves addressing missing values in various columns, ensuring uniformity in categorical variables like "Profession" and "Marital Status," and handling outliers in numerical features like "Annual Income." Transforming categorical variables into numerical representations may be necessary for machine learning models. Imputation techniques can be applied judiciously, preserving the integrity of the dataset. EDA helps uncover patterns, aiding in crafting targeted strategies for different customer segments.

Effective data preprocessing ensures that the datasets are cleaned, standardized, and ready for in-depth analysis. It establishes a solid foundation for subsequent tasks such as segmentation analysis, machine learning model training, and the formulation of an informed market entry strategy in the dynamic landscape of the Indian Electric Vehicle market.

SEGMENT EXTRACTION

The initial Python script focuses on the pivotal stage of "Segment Extraction" within the electric vehicle (EV) dataset. This process is crucial for identifying distinct clusters using unsupervised learning techniques, particularly KMeans clustering, to reveal patterns and groupings that provide valuable insights into the diverse characteristics of EVs.

The script begins by loading and exploring the EV dataset, analyzing key features and statistics to comprehend the data's structure. Feature engineering is then employed to create or modify attributes, capturing essential aspects such as acceleration patterns, charging speeds, and vehicle types. This enrichment facilitates deeper analysis.

The primary emphasis is on clustering analysis, applying KMeans to group similar EVs based on selected features. This segmentation reveals distinct clusters representing segments of EVs with similar characteristics. Visualizations, including scatter plots, enhance the understanding of these segments.

Segment extraction is fundamental for uncovering nuanced insights within the EV dataset, allowing stakeholders to tailor strategies, marketing approaches, or product enhancements to specific segments. This approach aligns with the broader goal of understanding the diverse landscape of EVs, optimizing strategies for development, marketing, and adoption, and contributing to informed decision-making in the electric vehicle domain.

ML Techniques Used:

The machine learning techniques used in the provided Python scripts are:

1. KMeans Clustering:
   - → Purpose: Identifying distinct clusters or segments within the electric vehicle dataset based on specific features.
   - → Implementation: Utilizing the KMeans algorithm from the scikit-learn library to group similar electric vehicles together.

2. Principal Component Analysis (PCA):
   - → Purpose: Reducing the dimensionality of the dataset and transforming features into principal components.
   - → Implementation: Employing PCA from scikit-learn to visualize and analyze the dataset in a lower-dimensional space.

3. Logistic Regression:
   - → Purpose: Implementing a logistic regression model for classification tasks, and predicting cluster labels.
   - → Implementation: Using scikit-learn's LogisticRegression for classifying and analyzing the clusters.

Steps and Libraries Used:

1st Dataset:

Libraries Used:

- pandas
- matplotlib.pyplot
- seaborn
- sklearn.cluster.KMeans
- -sklearn.preprocessing.StandardScaler

Steps:

1. Data Loading and Exploration
2. Feature Engineering
3. Clustering Analysis
4. Visualization and Analysis

2nd Dataset:

Libraries Used:

- pandas
- matplotlib.pyplot
- seaborn
- sklearn.cluster.KMeans

Steps:

1. Data Loading and Exploration
2. Geospatial Visualization
3. Clustering Analysis

3rd Dataset:

Libraries Used:

- pandas
- matplotlib.pyplot
- seaborn
- -sklearn.preprocessing.LabelEncoder
- -sklearn.model_selection.train_test_split
- -sklearn.decomposition.PCA
- -sklearn.linear_model.LogisticRegression

Steps:

1. Data Loading and Renaming
2. Exploratory Data Analysis (EDA)
3. Variable Transformation
4. Handling Missing Values
5. Data Transformation and Encoding
6. Handling Outliers
7. Unsupervised Learning (KMeans)
8. Principal Component Analysis (PCA)
9. Logistic Regression
10. Visualization of Results

PROFILING AND DESCRIBING POTENTIAL SEGMENTS

In the context of electric vehicle (EV) market analysis and segmentation, the focus shifts beyond clustering towards a detailed exploration of potential segments. Following techniques like KMeans clustering, the next step involves profiling and describing these segments to understand their distinct characteristics. Through the utilization of descriptive statistics, visualizations, and analytics, each segment is scrutinized based on features such as acceleration, top speed, range, efficiency, and pricing. This process provides a nuanced understanding of consumer behaviors within each segment.

Profiling extends to demographic, psychographic, and behavioral aspects, offering a comprehensive view of potential buyers. Visual representations aid in depicting feature distributions, and external factors such as geography and regulations are considered. The willingness of consumers to adopt EVs within each segment is assessed, informing market penetration strategies. Profiling ensures that marketing, policy, and manufacturing decisions align with the diverse preferences and priorities within the dynamic landscape of electric mobility. By adopting a customer-centric approach, stakeholders can tailor their strategies, fostering sustainable growth and adoption of electric vehicles in the evolving market.

SELECTION OF TARGET SEGMENT

The selection of a target segment in electric vehicle (EV) market analysis is a critical strategic step that involves a meticulous evaluation of various factors. Following the segmentation process, businesses must identify a specific consumer group that aligns with their product offerings and organizational objectives. This decision encompasses a thorough analysis of each segment's size, growth potential, and profitability. Understanding the distinct needs, preferences, and behaviors of potential customers within each segment aids in selecting a target audience that is most receptive to the unique value proposition of electric vehicles.

Market trends, consumer demographics, and psychographic profiles play a pivotal role in this selection process. Additionally, factors such as regulatory support, infrastructure development, and technological advancements influence the viability of a target segment. Businesses need to assess their own capabilities and resources to effectively cater to the identified segment. This targeted approach enhances marketing efficiency, resource allocation, and overall competitiveness in the dynamic landscape of the electric vehicle industry.

DATASET 1:

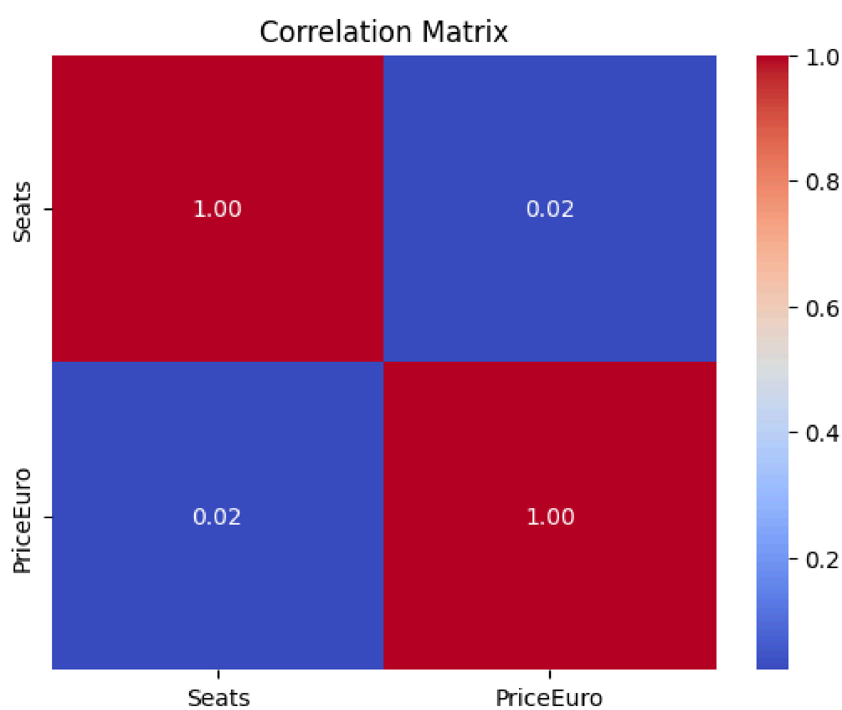
Columns in the dataset include:

```
ev_data.head()
```

	Brand	Model	Accel	TopSpeed	Range	Efficiency	FastCharge	RapidCharge	PowerTrain	PlugType	BodyStyle	Segment	Seats	PriceEuro
0	Tesla	Model 3 Long Range Dual Motor	4.6 sec	233 km/h	450 km	161 Wh/km	940 km/h	Rapid charging possible	All Wheel Drive	Type 2 CCS	Sedan	D	5	55480
1	Volkswagen	ID.3 Pure	10.0 sec	160 km/h	270 km	167 Wh/km	250 km/h	Rapid charging possible	Rear Wheel Drive	Type 2 CCS	Hatchback	C	5	30000
2	Polestar	2	4.7 sec	210 km/h	400 km	181 Wh/km	620 km/h	Rapid charging possible	All Wheel Drive	Type 2 CCS	Liftback	D	5	56440
3	BMW	iX3	6.8 sec	180 km/h	360 km	206 Wh/km	560 km/h	Rapid charging possible	Rear Wheel Drive	Type 2 CCS	SUV	D	5	68040
4	Honda	e	9.5 sec	145 km/h	170 km	168 Wh/km	190 km/h	Rapid charging possible	Rear Wheel Drive	Type 2 CCS	Hatchback	B	4	32997

Correlation matrix for the numerical columns in the dataframe and then creating a heatmap to visually represent the correlation coefficients.

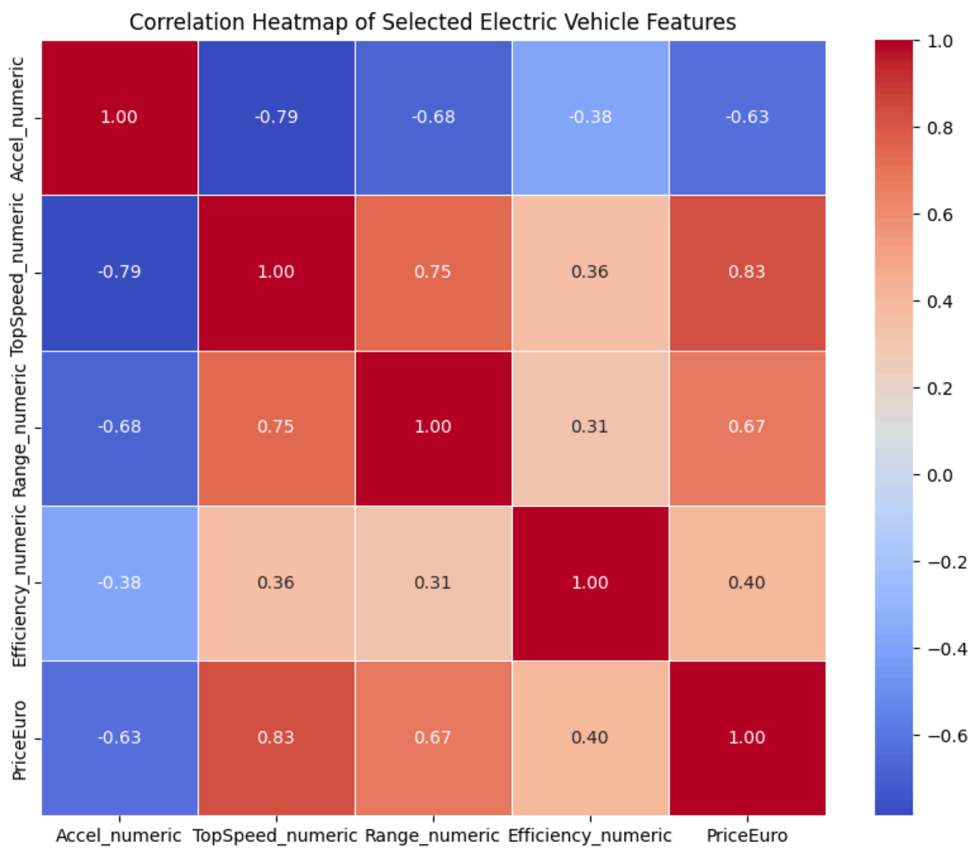
The visualization shows how strongly each pair of numerical variables in the dataset is correlated. Positive values indicate a positive correlation, negative values indicate a negative correlation, and values close to zero suggest a weak or no correlation between the variables. The color intensity in the heatmap helps visualize the strength of the correlation, with warmer colors (reds) indicating stronger positive correlations and cooler colors (blues) indicating stronger negative correlations.



Correlation heatmap for a subset of features from the dataframe. The selected features include:

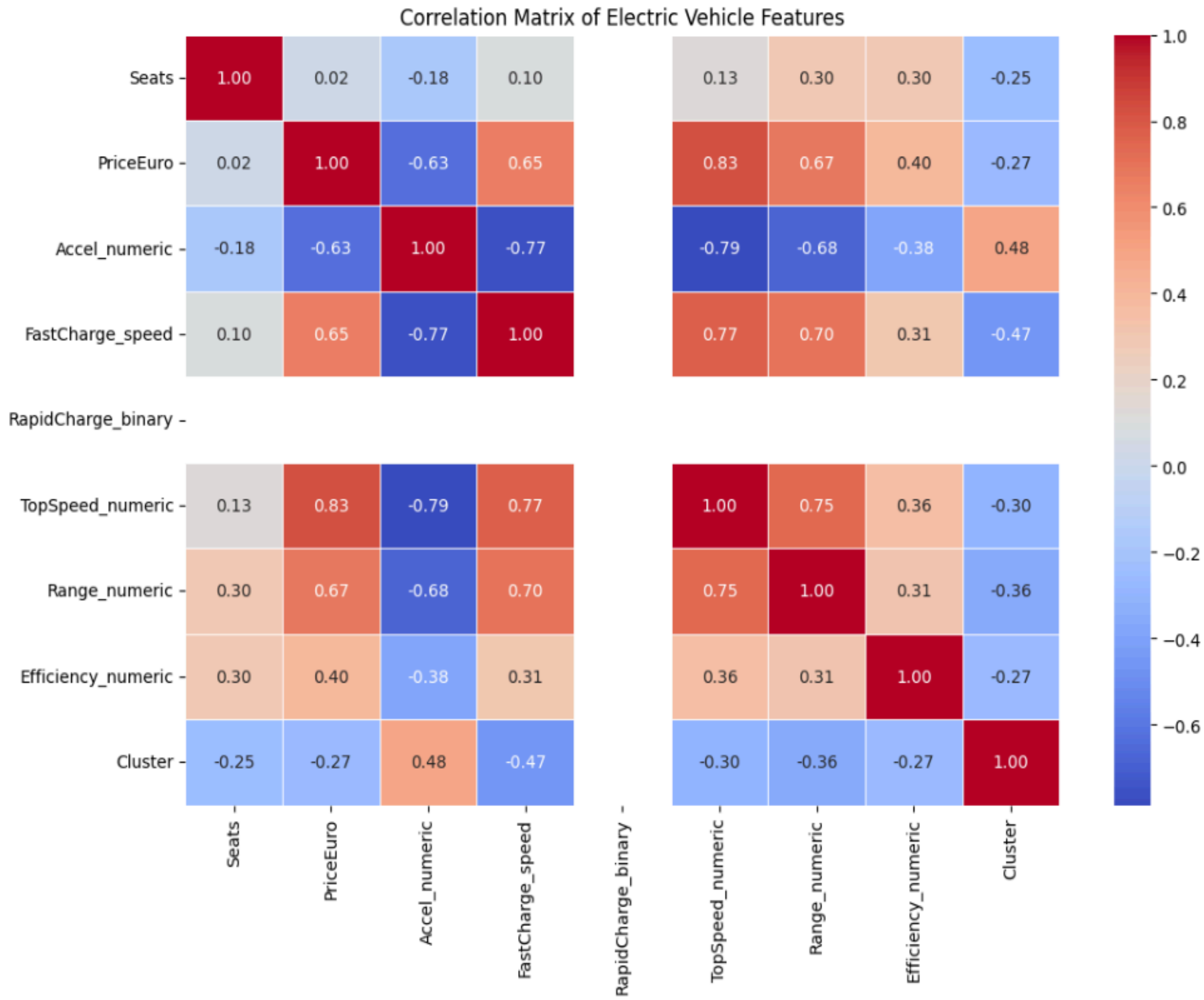
1. ***'Accel_numeric'***
2. ***'TopSpeed_numeric'***
3. ***'Range_numeric'***
4. ***'Efficiency_numeric'***
5. ***'PriceEuro'***

The visualization gives insights into how strongly the selected features are correlated with each other.



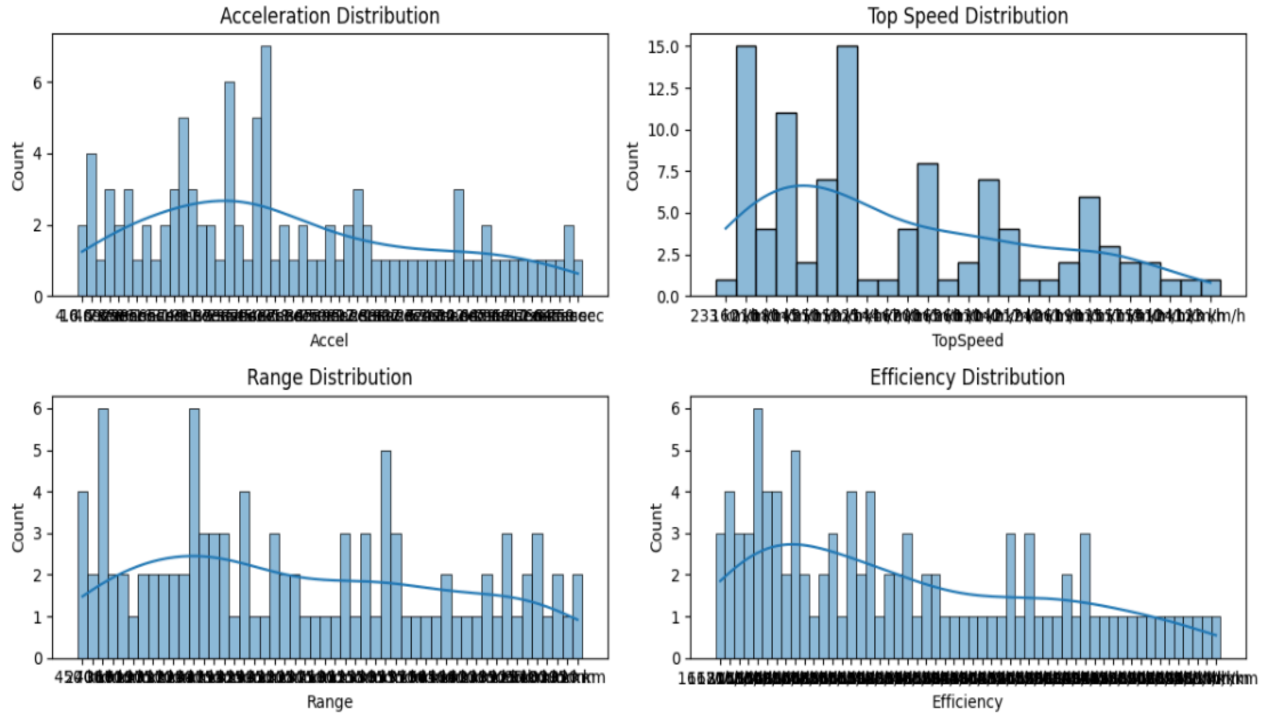
Correlation heatmap for all numerical features in the dataframe.

It provides an overview of how strongly different numerical features in the dataset are correlated with each other. The color intensity in the heatmap helps visualize the strength of these correlations.



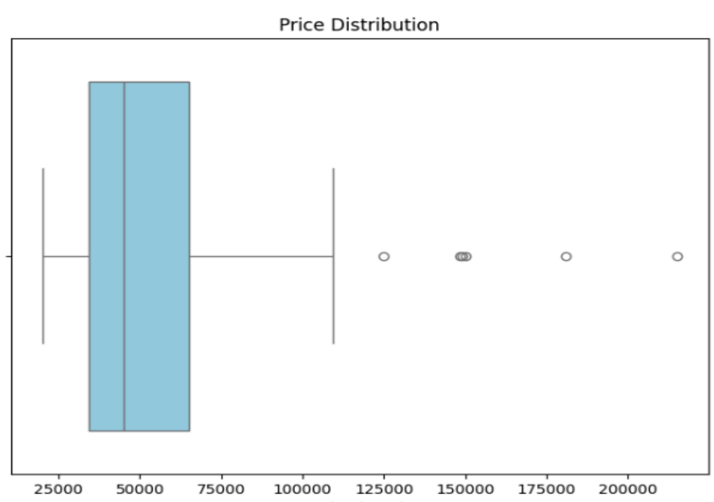
Subplot with four histograms each representing the distribution of a specific electric vehicle feature created using the `matplotlib` and `seaborn` libraries.

The resulting visualization shows the distributions of acceleration, top speed, range, and efficiency for the electric vehicles in the dataset. The histograms provide insights into the spread and frequency of values for each feature. The optional kernel density estimate (kde=True) adds a smoothed line to the histograms, offering a continuous representation of the distribution.



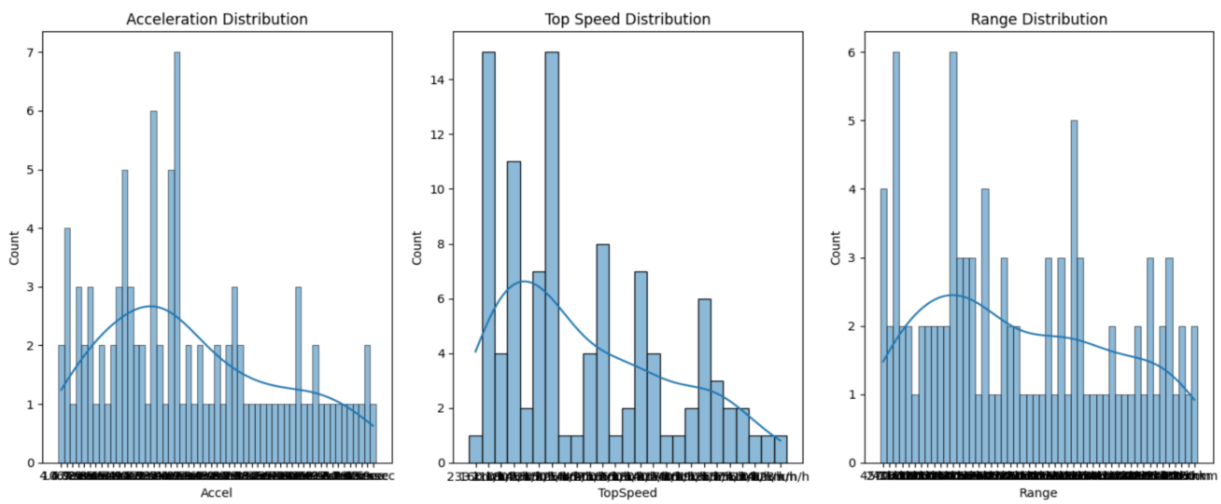
Boxplot to visualize the distribution of prices (in Euros) for electric vehicles in the dataframe.

It provides a summary of the distribution of prices for electric vehicles. The boxplot allows you to see the central tendency, spread, and presence of outliers in the price data. The median is represented by the line inside the box, while the box itself represents the interquartile range (IQR). Whiskers extend to show the range of most of the data, and individual points beyond the whiskers may indicate potential outliers.



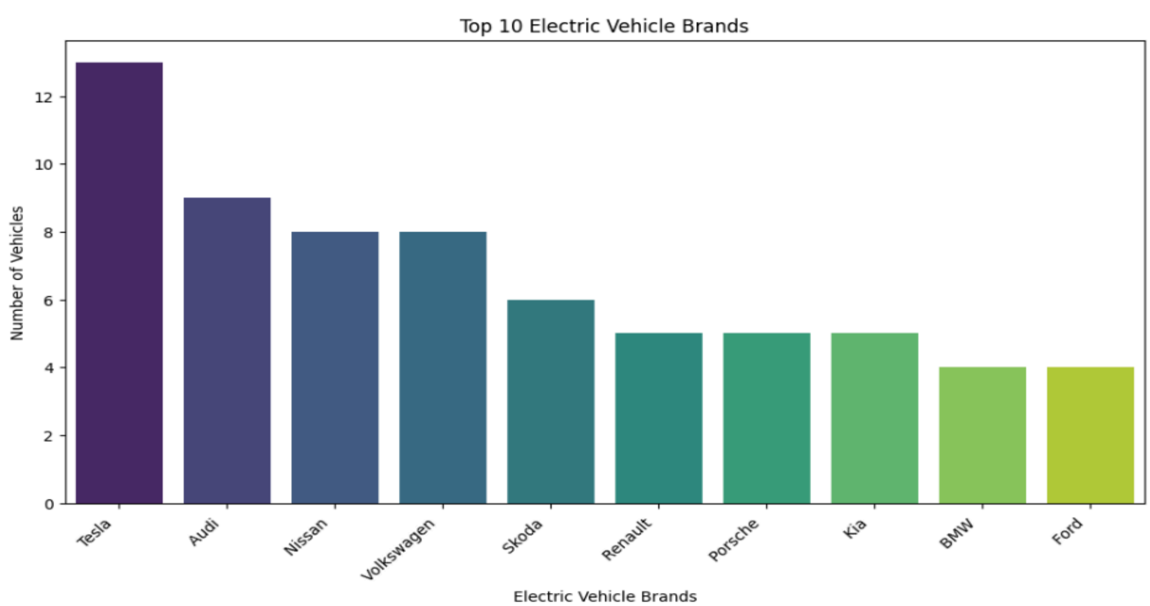
Subplot with three histograms each with a kernel density estimate for a specific electric vehicle feature.

The resulting visualization shows side-by-side histograms for the acceleration, top speed, and range distributions of the electric vehicles in the dataset. Each subplot provides insights into the spread and frequency of values for a specific feature.



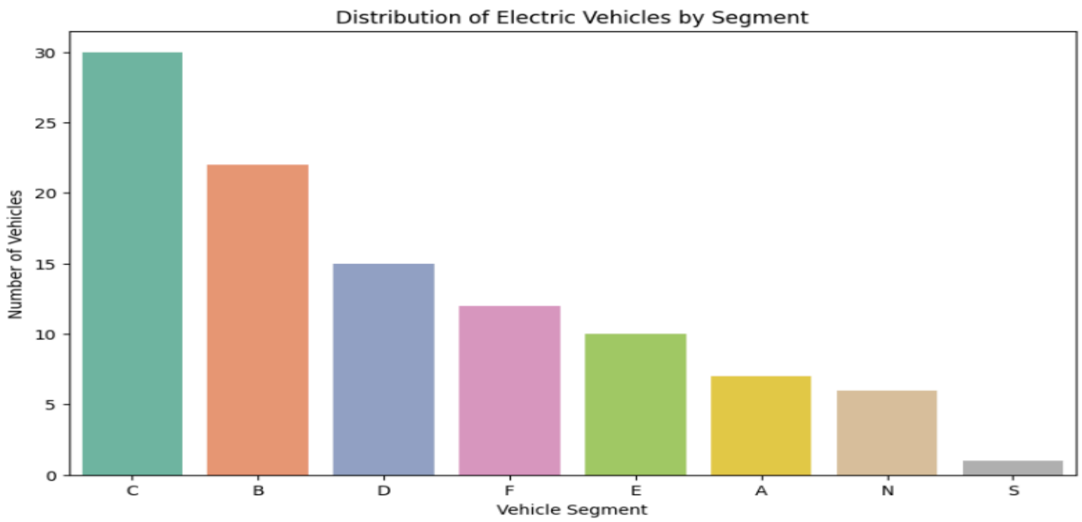
Bar plot to visualize the top 10 electric vehicle brands based on the number of vehicles in the dataframe.

The visualization provides a clear representation of the top 10 electric vehicle brands and their respective counts in the dataset.



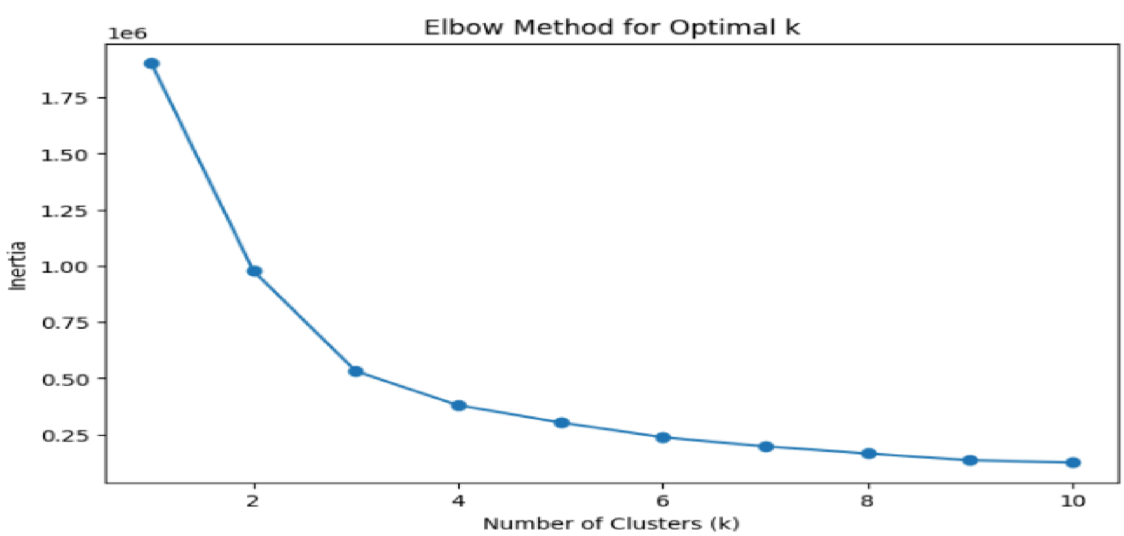
Count plot to visualize the distribution of electric vehicles by segment in the dataframe.

The visualization provides a bar chart showing the distribution of electric vehicles across different segments. The colors differentiate between segments, and the ordering of segments is based on their counts in the dataset.



Elbow Method Plot to help determine the optimal number of clusters (k) in a k-means clustering algorithm.

In the Elbow Method, the inertia is plotted for different values of k, and the "elbow" point is where the inertia starts to decrease at a slower rate. This point is often considered the optimal number of clusters for the given dataset.



Clusters and their characteristics:

The table represents characteristics of the clusters generated by a clustering algorithm. Here's an interpretation of the key characteristics for each cluster:

Cluster 0:

Average Seats: Approximately 5.06 seats.
Average PriceEuro: Approximately €66,761.82.
Average Acceleration: Approximately 5.96 seconds (0-100 km/h).
Average FastCharge Speed: Approximately 546.13 km/h.
RapidCharge Availability: Available (Binary representation: 1.0).
Average Top Speed: Approximately 195.05 km/h.
Average Range: Approximately 394.92 km.
Average Efficiency: Approximately 197.61 Wh/km.

Cluster 1:

Average Seats: Approximately 4.59 seats.
Average PriceEuro: Approximately €33,829.69.
Average Acceleration: Approximately 9.92 seconds (0-100 km/h).
Average FastCharge Speed: Approximately 272.65 km/h.
RapidCharge Availability: Available (Binary representation: 1.0).
Average Top Speed: Approximately 147.28 km/h.
Average Range: Approximately 222.82 km.
Average Efficiency: Approximately 173.31 Wh/km.

Cluster 2:

Average Seats: 5 seats.
Average PriceEuro: €145,000.
Average Acceleration: Approximately 2.55 seconds (0-100 km/h).
Average FastCharge Speed: Approximately 815 km/h.
RapidCharge Availability: Available (Binary representation: 1.0).
Average Top Speed: Approximately 310 km/h.
Average Range: Approximately 860 km.
Average Efficiency: Approximately 236.50 Wh/km.

Each cluster represents a group of electric vehicles with similar characteristics based on the features used in the clustering algorithm.

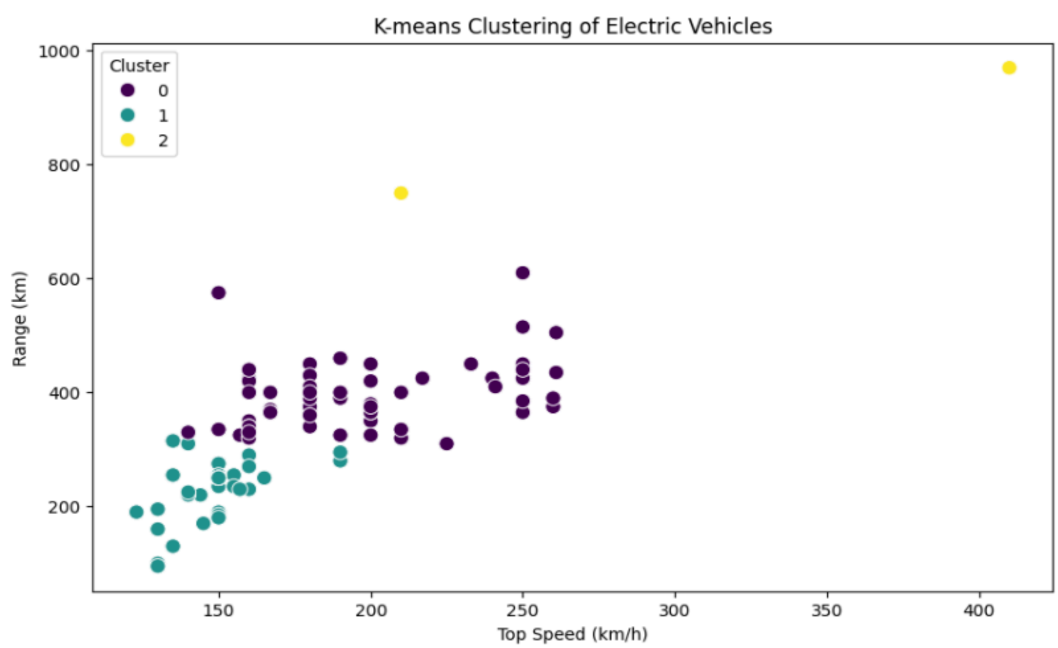
```
Clusters and Their Characteristics:
            Seats      PriceEuro  Accel_numeric  FastCharge_speed
Cluster
0        5.064516   66761.822581       5.962903        546.129032
1        4.589744   33829.692308       9.923077        272.647059
2        5.000000  145000.000000       2.550000        815.000000

         RapidCharge_binary  TopSpeed_numeric  Range_numeric  \
Cluster
0                       1.0        195.048387     394.919355
1                       1.0        147.282051     222.820513
2                       1.0        310.000000     860.000000

         Efficiency_numeric
Cluster
0                197.612903
1                173.307692
2                236.500000
```

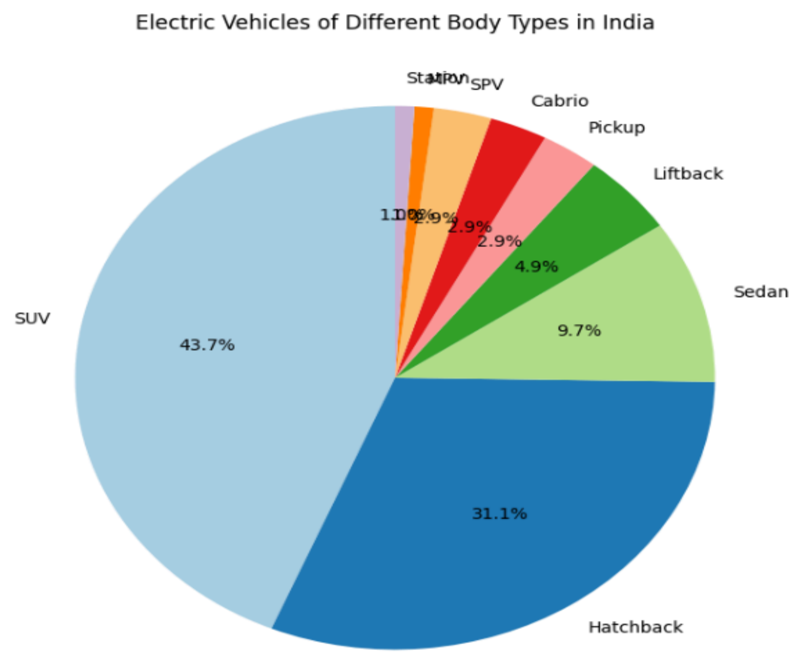
Scatter plot to visualize the k-means clustering results for electric vehicles based on the 'TopSpeed_numeric' and 'Range_numeric' features.

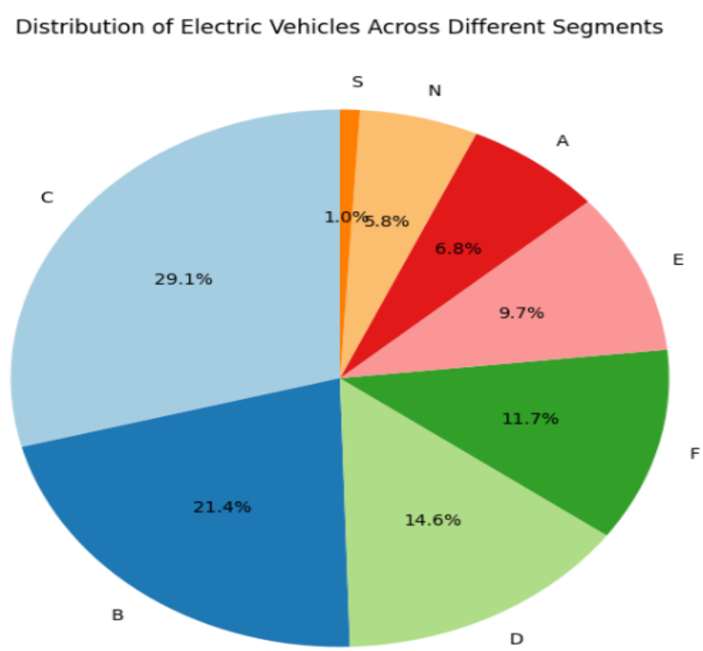
The resulting visualization allows you to see how electric vehicles are distributed in the 'Top Speed' and 'Range' feature space, with different colors representing different clusters assigned by the k-means algorithm. It helps in understanding the separation or grouping of vehicles based on these two features.



Pie charts to visualize the distribution of electric vehicles in India based on their body types and the distribution of electric vehicles across different segments in the dataframe.

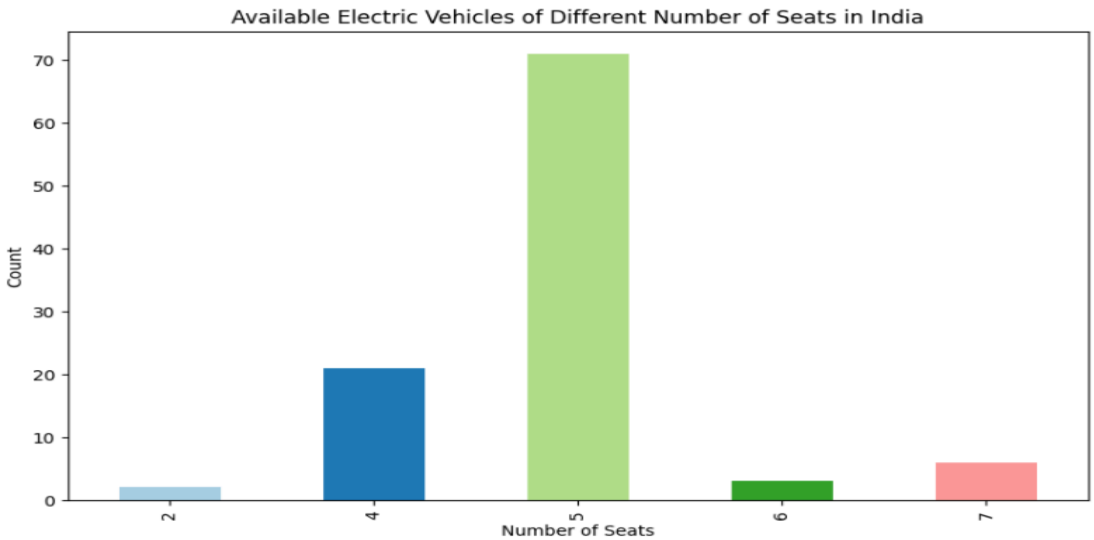
The size of each slice corresponds to the percentage of vehicles in that category. The percentage values are displayed on the chart as well. The color palette helps differentiate between the different body types.





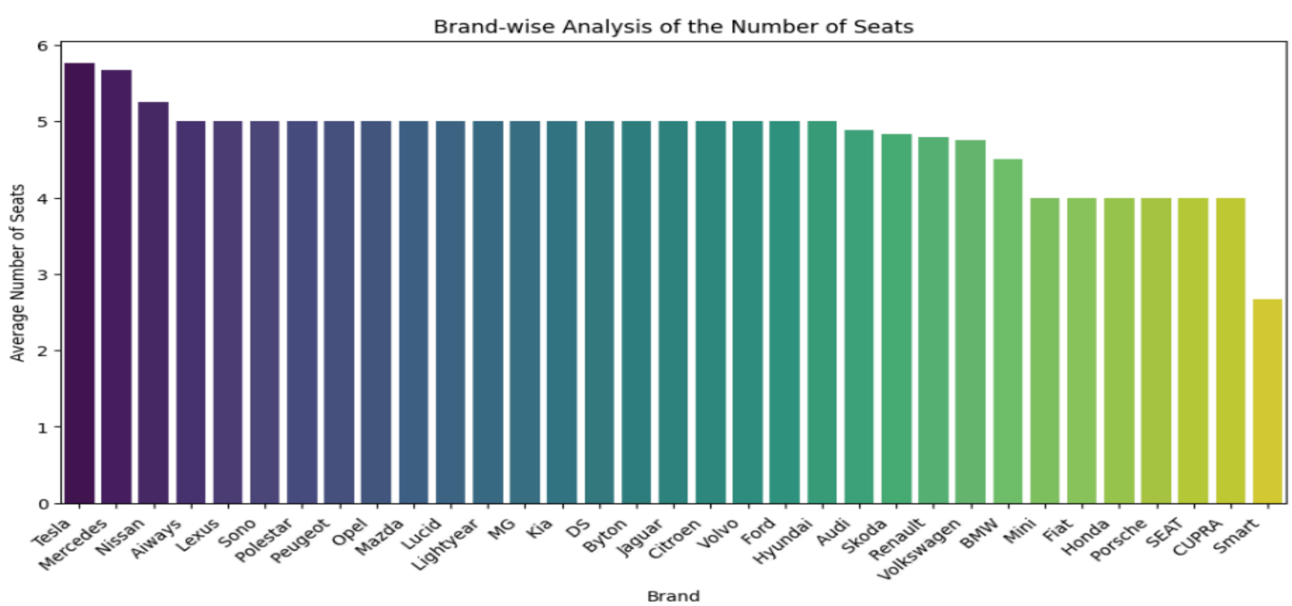
Bar chart to visualize the distribution of available electric vehicles in India based on the number of seats.

Each bar represents a different number of seats, and the height of the bar corresponds to the count of electric vehicles with that specific number of seats.



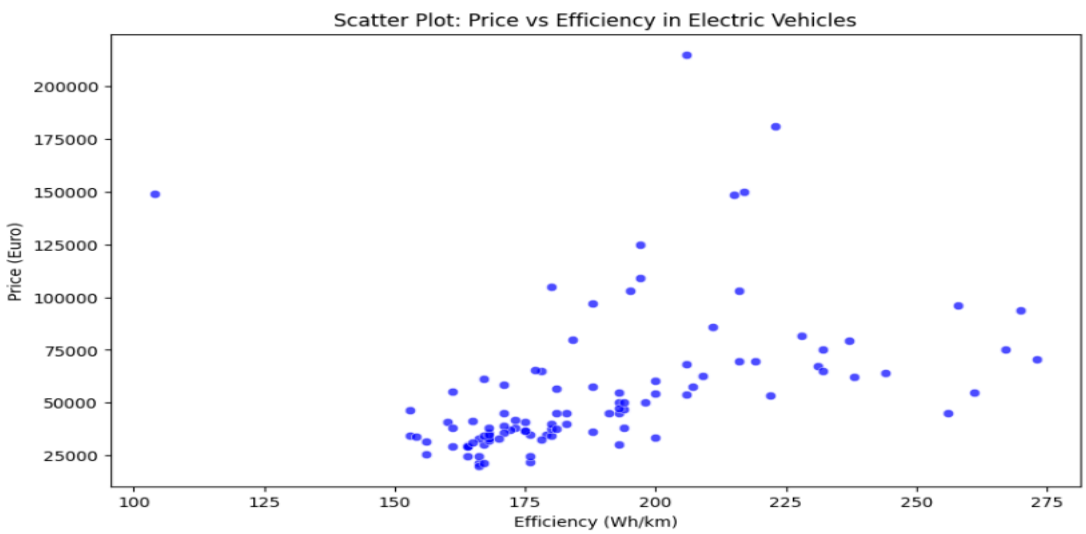
Bar plot to visualize the average number of seats for different electric vehicle brands in the dataframe.

Each bar represents a brand, and the height of the bar corresponds to the average number of seats for that brand.



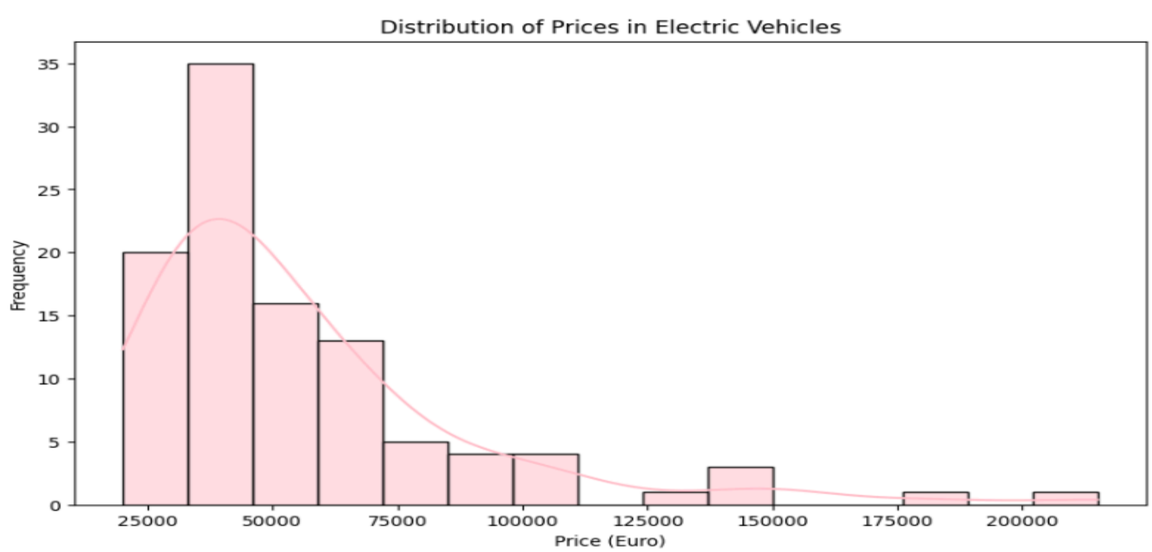
Scatter plot to visualize the relationship between the efficiency (measured in Wh/km) and the price (in Euros) of electric vehicles.

Each point on the plot represents a specific electric vehicle, and the position of the points provides insights into how the efficiency and price are distributed across the dataset.



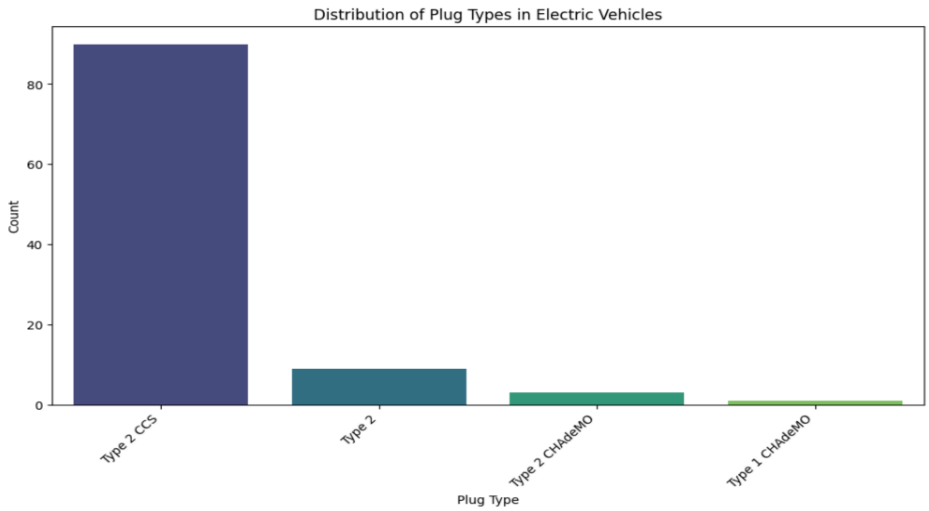
Histogram to visualize the distribution of prices for electric vehicles in Euros.

The histogram displays the frequency of different price ranges, and the kernel density estimate provides a smoothed curve showing the overall distribution trend.



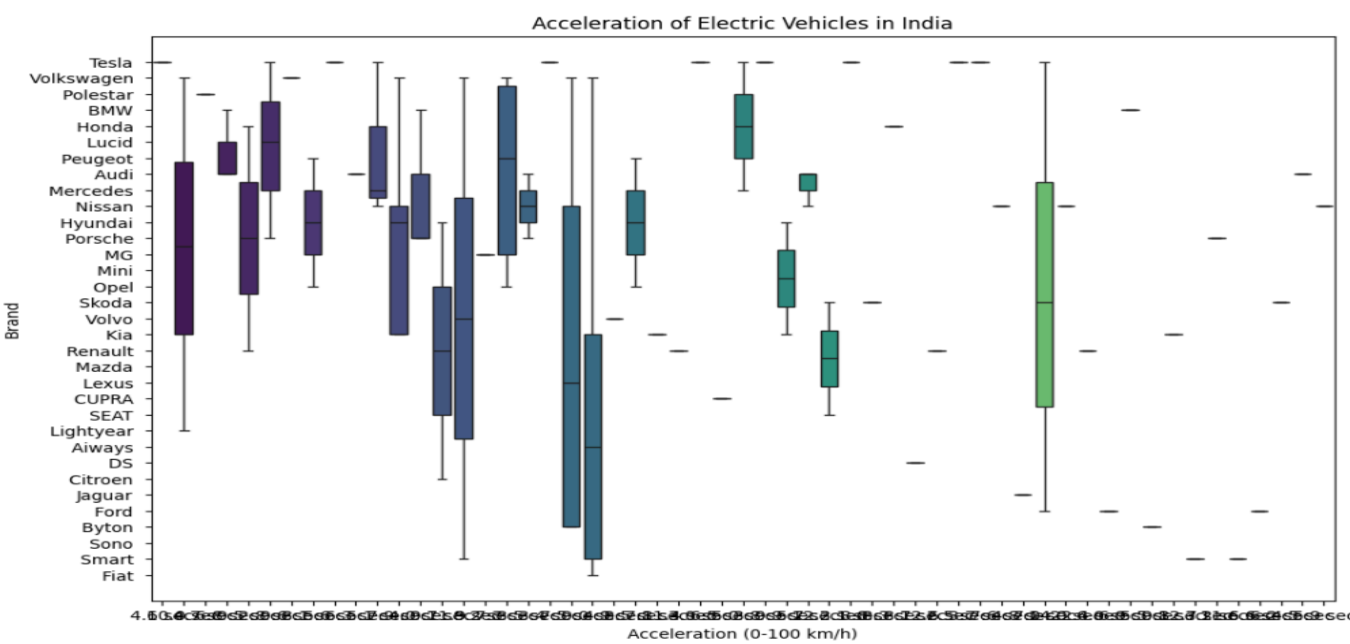
Bar plot to visualize the distribution of plug types in electric vehicles based on the 'PlugType' column in the dataframe.

Each bar represents a different plug type, and the height of the bar corresponds to the count of electric vehicles with that specific plug type.



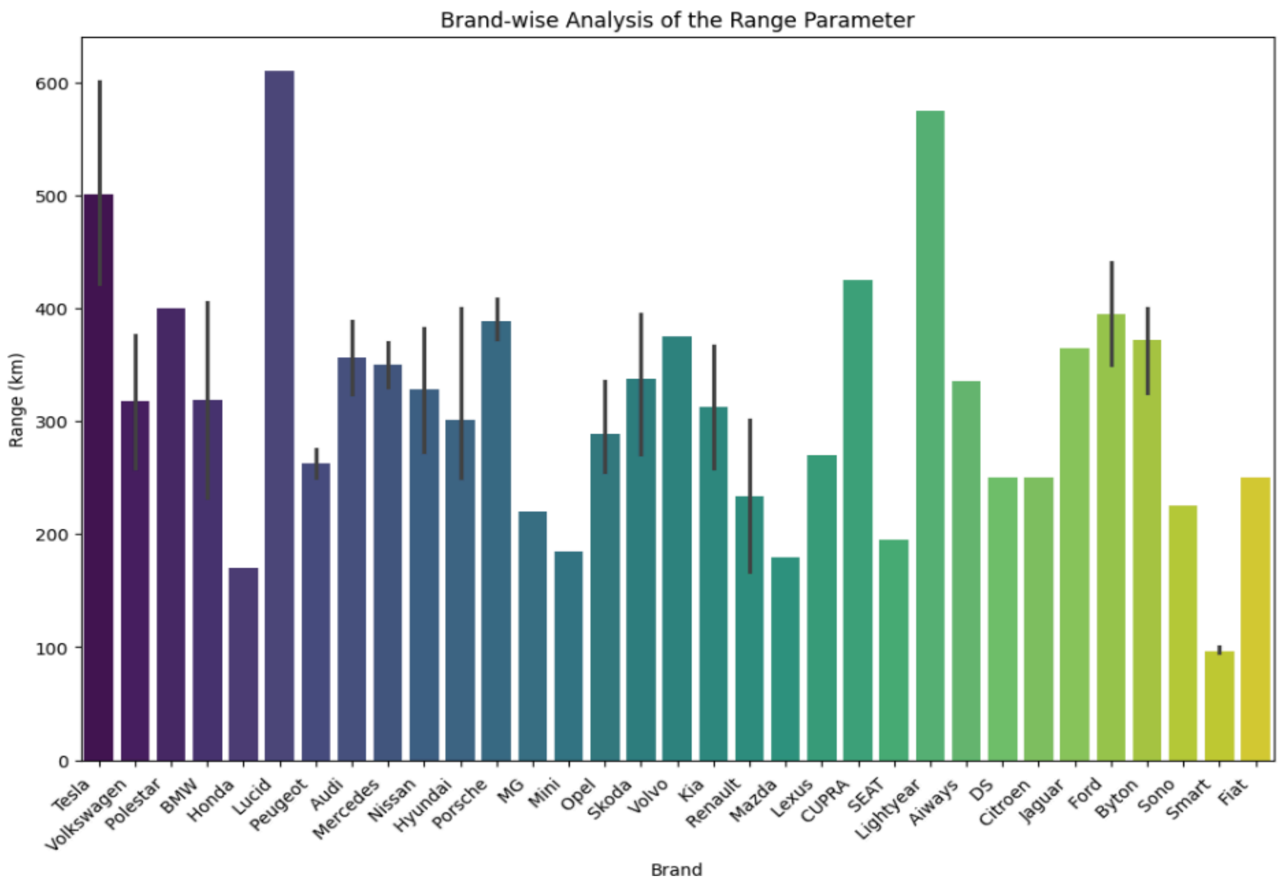
Horizontal boxplot to visualize the distribution of acceleration times (0-100 km/h) for different electric vehicle brands in India.

Each box represents the interquartile range (IQR) of acceleration times for a specific brand, and the whiskers show the range of the data.



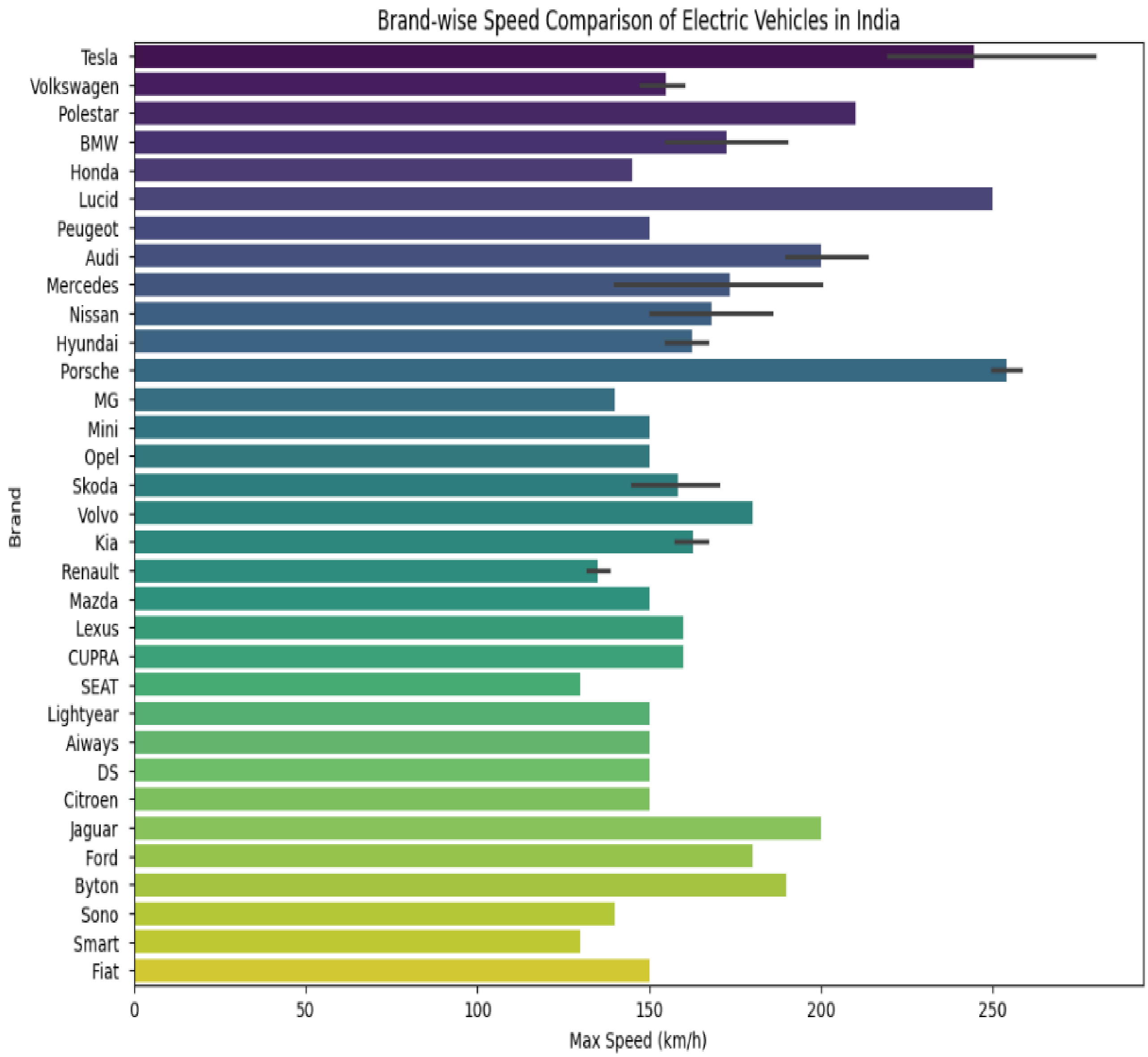
Bar plot to visualize the range (in kilometers) of electric vehicles for different brands.

The resulting visualization provides a comparison of the average range of electric vehicles for different brands. Each bar represents a brand, and the height of the bar corresponds to the average range for that brand.



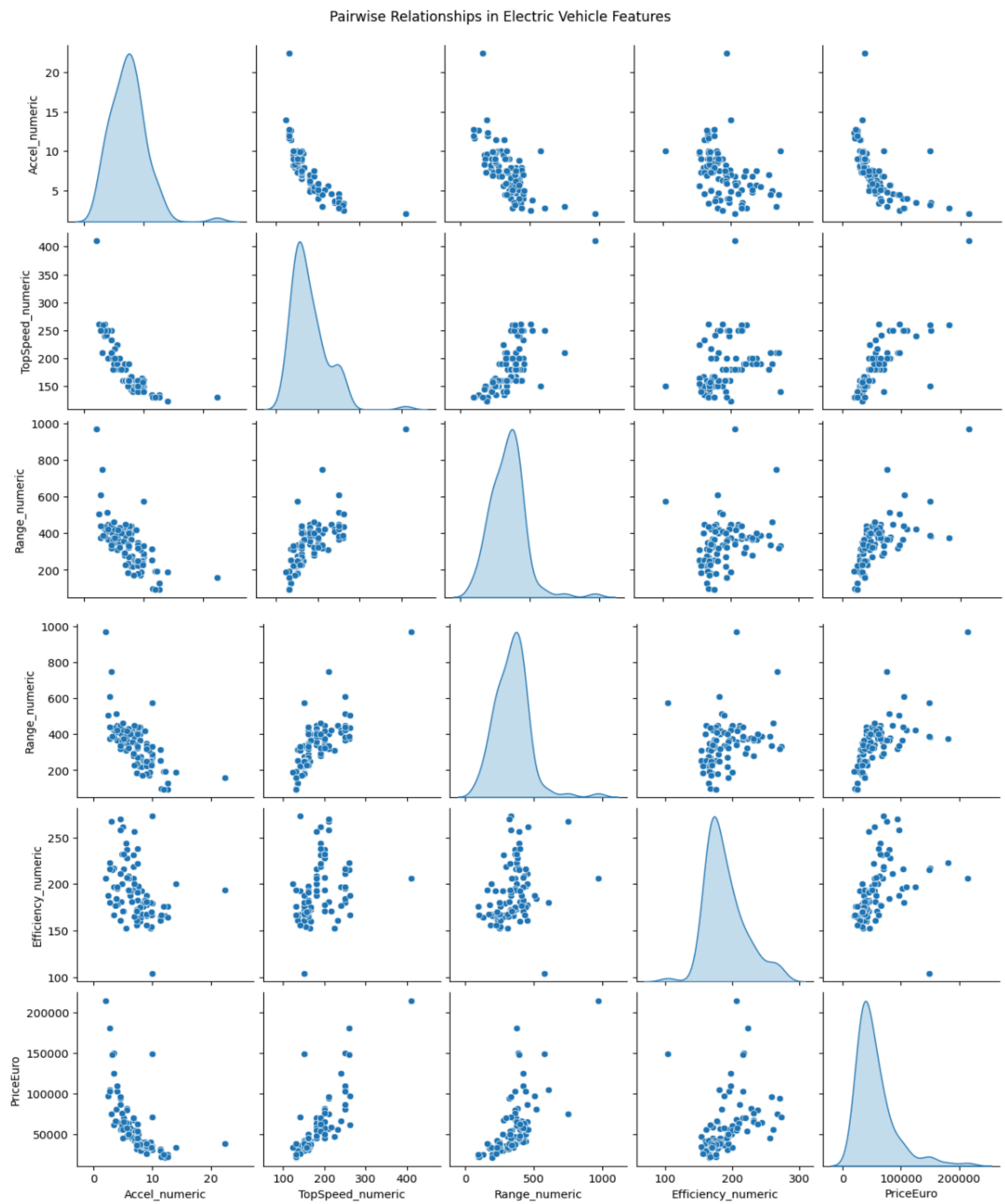
Bar plot to compare the maximum speed (in kilometers per hour) of electric vehicles for different brands in India.

The resulting visualization provides a comparison of the average maximum speed of electric vehicles for different brands. Each bar represents a brand, and the length of the bar corresponds to the average maximum speed for that brand. This analysis can be valuable for understanding how the speed parameter varies across different electric vehicle brands in the Indian market.



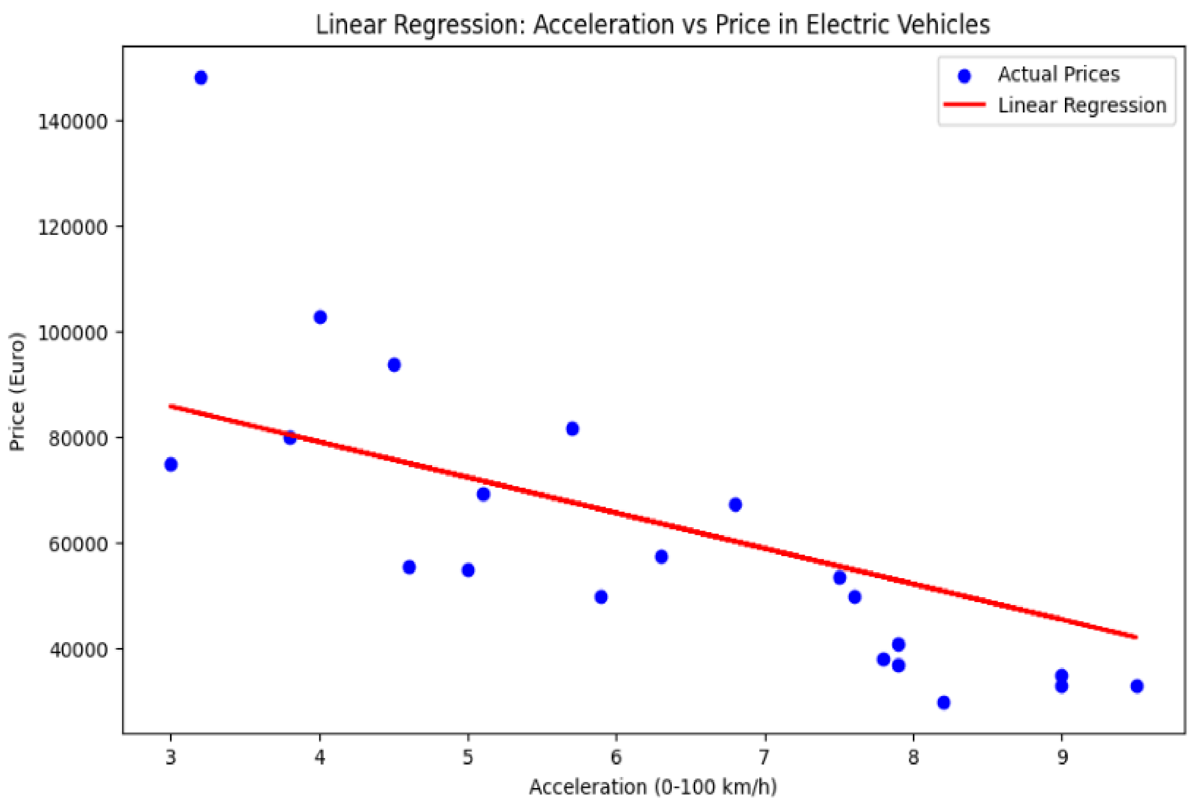
Pairplot to visualize pairwise relationships among selected features of electric vehicles.

The resulting pairplot provides a matrix of scatter plots for the pairwise relationships between the selected features, histograms along the diagonal (with kernel density estimates), and correlation coefficients. This visualization helps in understanding how different features are correlated and distributed with respect to each other in the dataset.


Pairwise Relationships in Electric Vehicle Features
Accel_numeric
TopSpeed_numeric
Range_numeric
Range_numeric
Efficiency_numeric
PriceEuro
Accel_numeric
TopSpeed_numeric
Range_numeric
Efficiency_numeric
PriceEuro

Scatter plot with a linear regression line to visualize the relationship between acceleration (0-100 km/h) and price (in Euros) for electric vehicles.

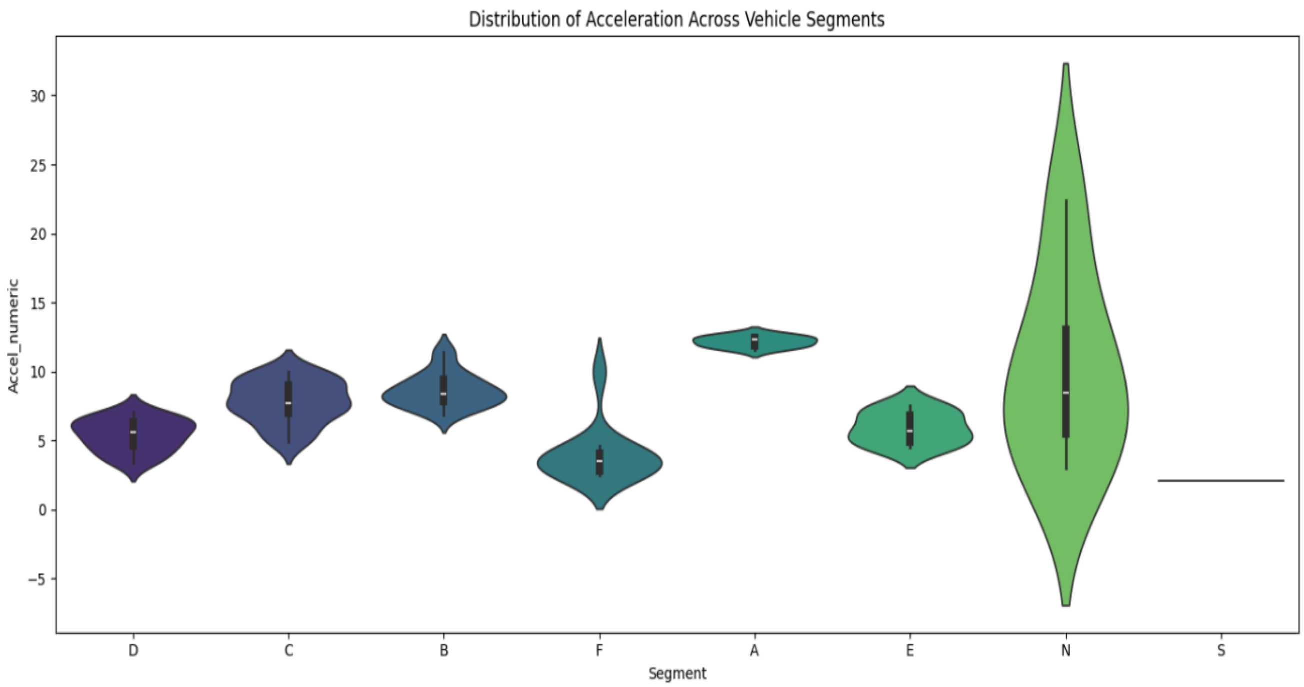
The resulting visualization allows you to compare the actual prices (blue points) with the predicted prices based on the linear regression model (red line). This is useful for assessing how well the linear regression model fits the relationship between acceleration and price in electric vehicles.



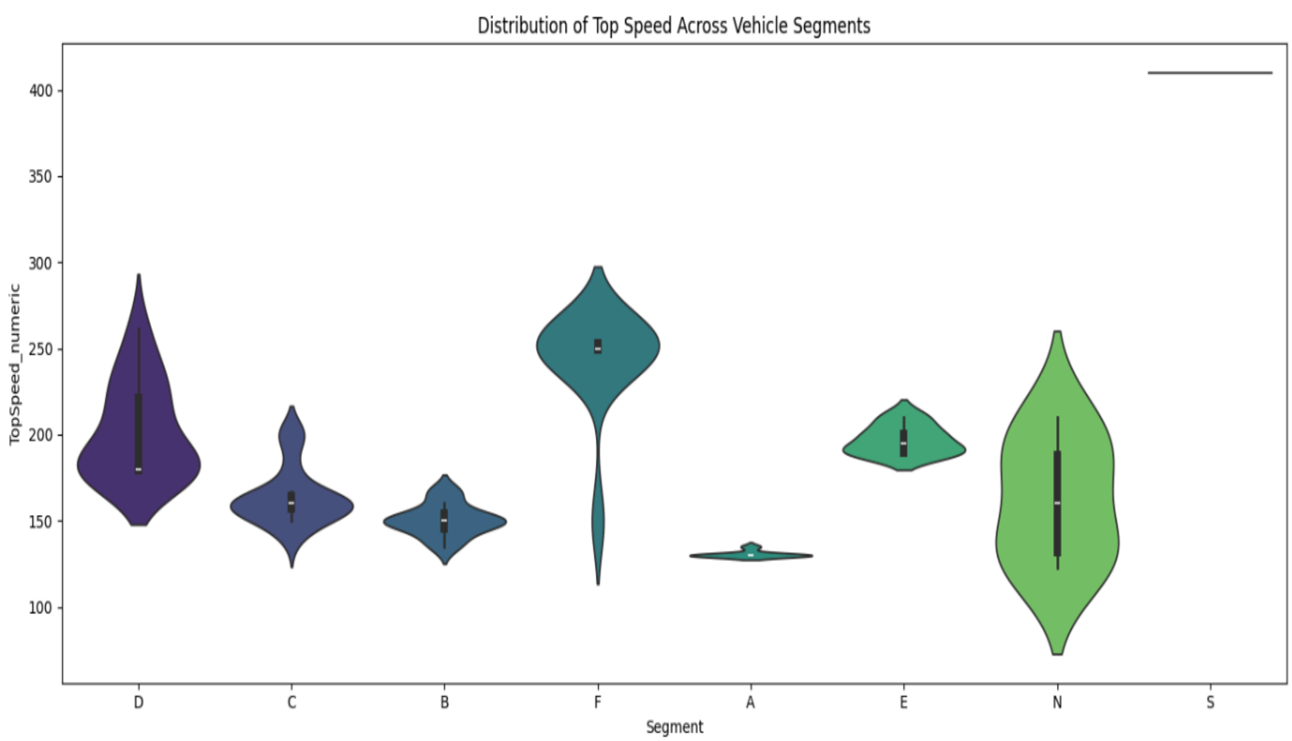
Set of three vertically stacked violin plots to visualize the distribution of acceleration, top speed, and efficiency across different vehicle segments.

The resulting visualization allows you to compare the distributions of acceleration, top speed, and efficiency across different vehicle segments using violin plots. The width of each violin represents the density of the data at different values, and the overall shape provides insights into the distribution characteristics within each segment.

First subplot: Distribution of Acceleration Across Vehicle Segments



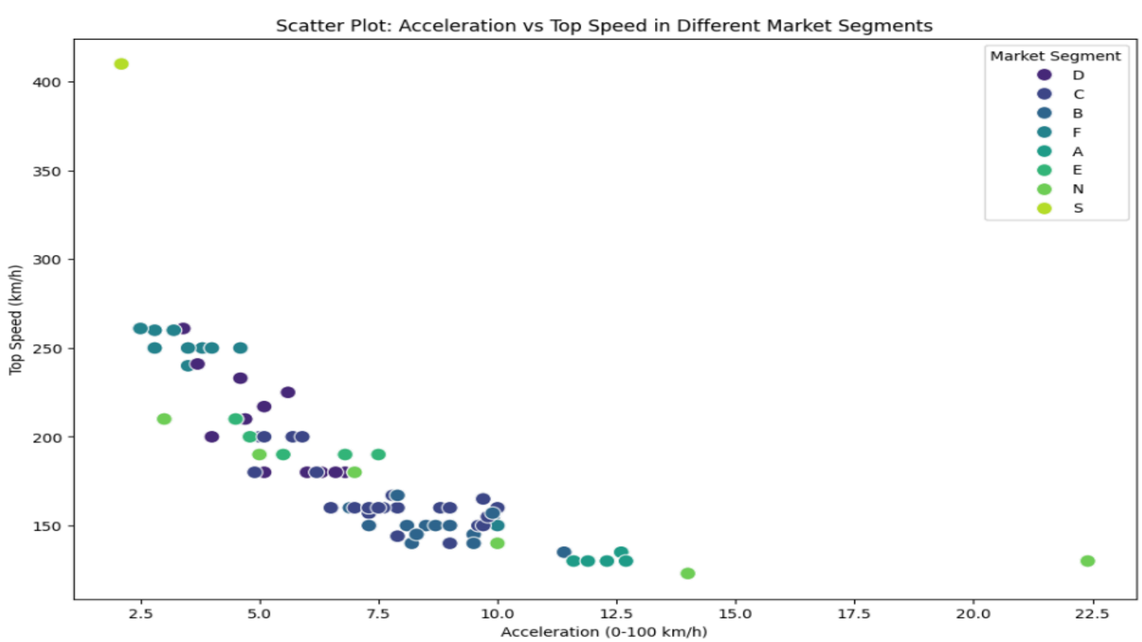
Second subplot: Distribution of Top Speed Across Vehicle Segments



Third subplot: Distribution of Efficiency Across Vehicle Segments

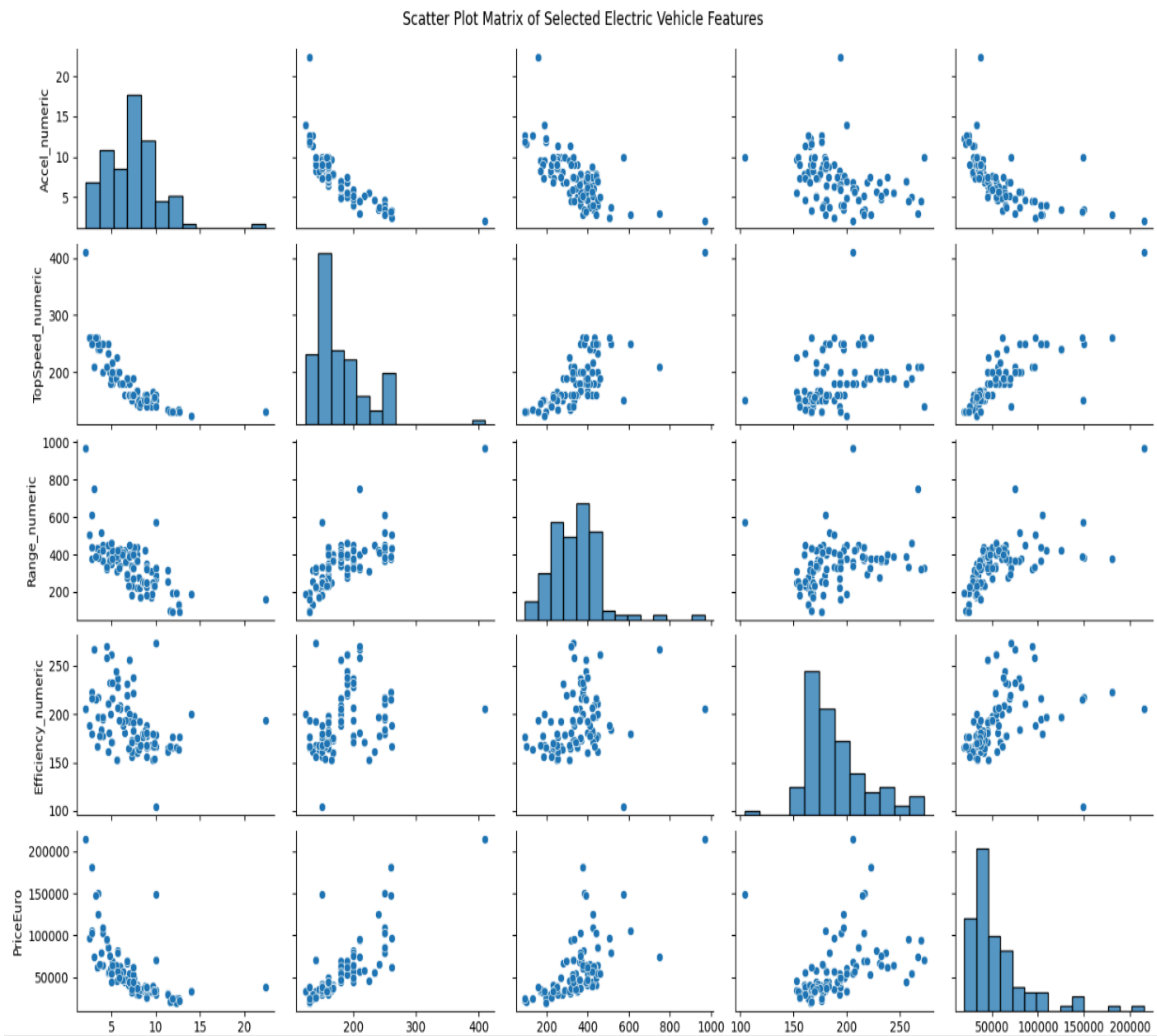
Scatter plot to visualize the relationship between acceleration (0-100 km/h) and top speed (km/h) for electric vehicles, with different market segments distinguished by color.

Each point on the plot represents a specific electric vehicle, and the color distinguishes the market segment to which it belongs.



Scatter plot matrix to visualize pairwise relationships among selected features of electric vehicles.

The scatter plot matrix provides a visual representation of the pairwise relationships and distributions among the selected features. Diagonal subplots typically display histograms or kernel density estimates for each individual feature.



Classification Report for evaluating classification performance:

Here's an interpretation of the key metrics:

Class 'B': The model has good precision (80%) and recall (80%).
Class 'C': The model has low precision (25%) and recall (25%).
Class 'D': The model has moderate precision (40%) and high recall (67%).
Class 'E': The model has high precision (80%) and recall (100%).
Class 'F': The model has perfect precision (100%) but lower recall (67%).
Class 'N': The model has low precision (0%) and recall (0%).

The overall accuracy of the model is 62%.

Classification Report:

	precision	recall	f1-score	support
B	0.80	0.80	0.80	5
C	0.25	0.25	0.25	4
D	0.40	0.67	0.50	3
E	0.80	1.00	0.89	4
F	1.00	0.67	0.80	3
N	0.00	0.00	0.00	2
accuracy			0.62	21
macro avg	0.54	0.56	0.54	21
weighted avg	0.59	0.62	0.59	21

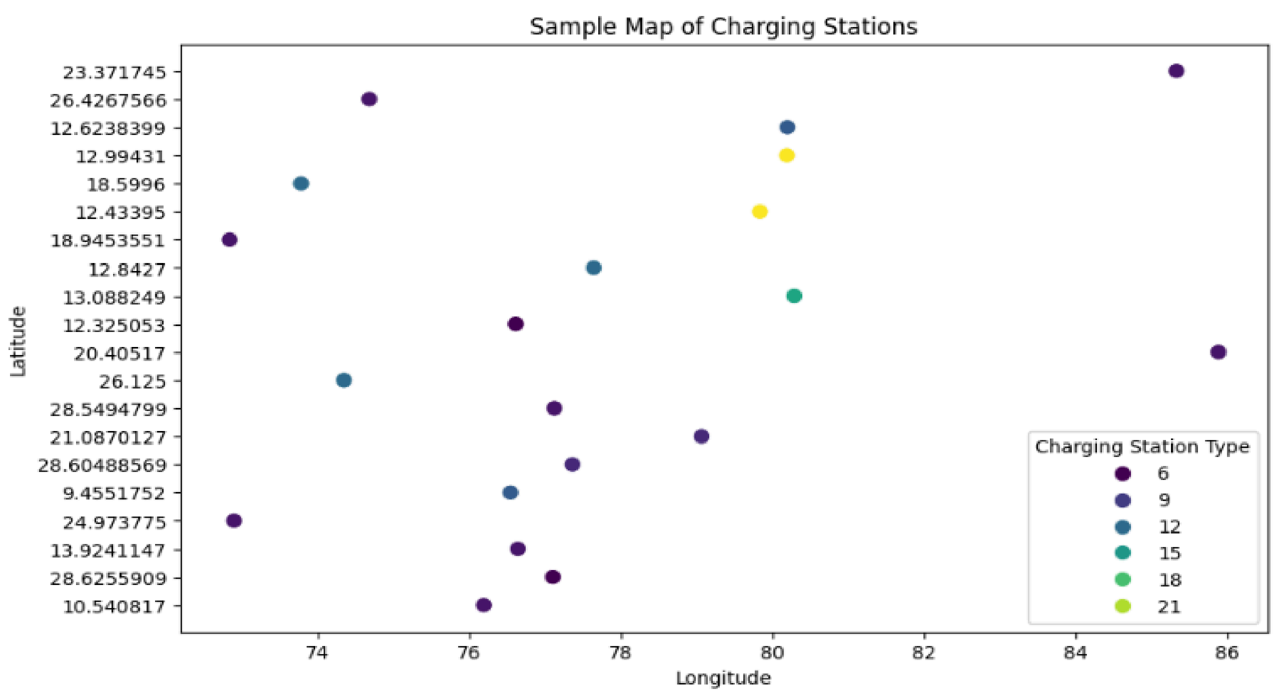
DATASET 2:

Columns in the dataset include:

	name	state	city	address	latitude	longitude	type
0	Neelkanth Star DC Charging Station	Haryana	Gurugram	Neelkanth Star Karnal, NH 44, Gharunda, Kutail...	29.6019	76.9803	12.0
1	Galleria DC Charging Station	Haryana	Gurugram	DLF Phase IV, Sector 28, Gurugram, Haryana 122022	28.4673	77.0818	12.0
2	Highway Xpress (Jaipur-Delhi) DC charging station	Rajasthan	Behror	Jaipur to Delhi Road, Behror Midway, Behror, R...	27.8751	76.2760	12.0
3	Food Carnival DC Charging Station	Uttar Pradesh	Khatauli	Fun and Food Carnival, NH 58, Khatauli Bypass,...	29.3105	77.7218	12.0
4	Food Carnival AC Charging Station	Uttar Pradesh	Khatauli	NH 58, Khatauli Bypass, Bhainsi, Uttar Pradesh...	29.3105	77.7218	12.0

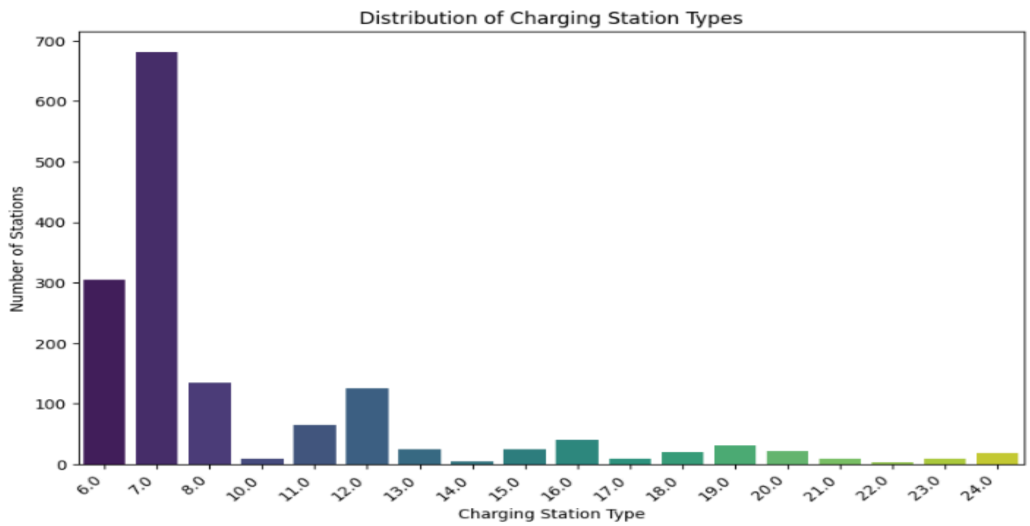
Scatter plot of a sample of 20 electric vehicle (EV) charging stations on a map.

The plot represents a sample of 20 charging stations on a map, with points colored based on their types. The x-axis represents longitude, the y-axis represents latitude, and the legend provides information about the types of charging stations.



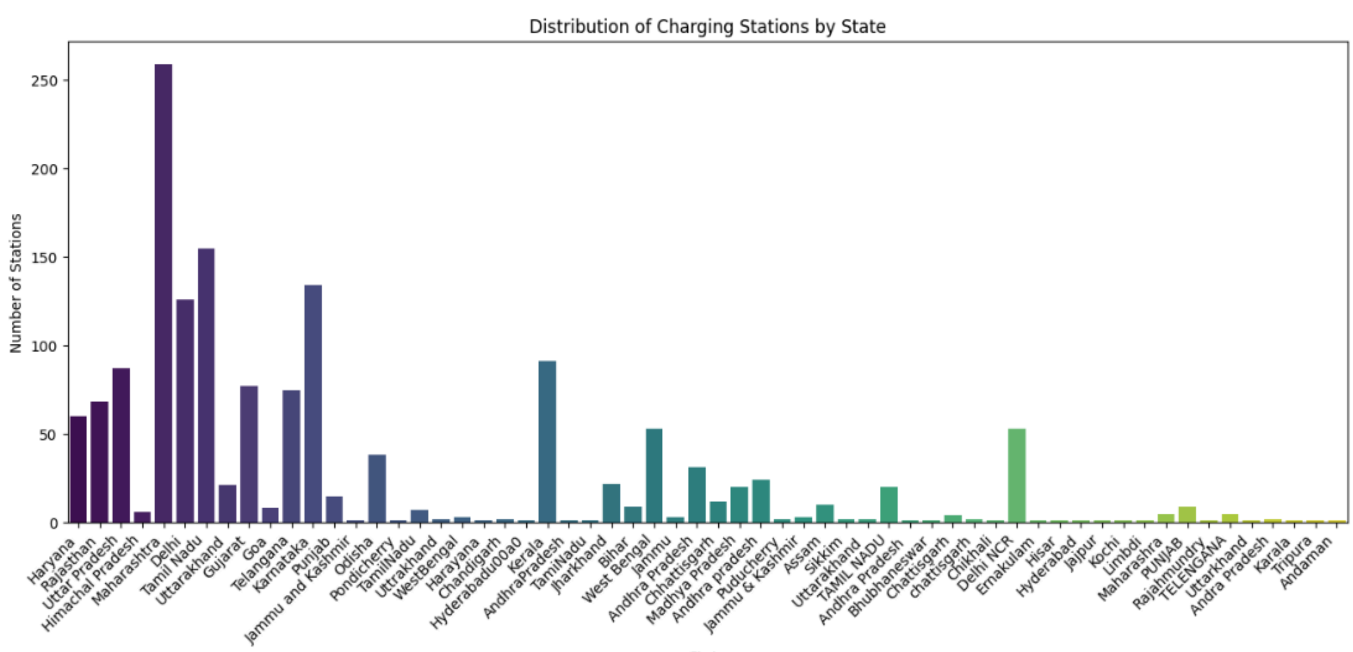
Count plot that displays the distribution of different types of charging stations in the dataframe.

The resulting visualization is a bar chart that shows the distribution of different charging station types, with each bar representing the count of charging stations for a particular type. This can help in understanding the prevalence of each type of charging station in the dataset.



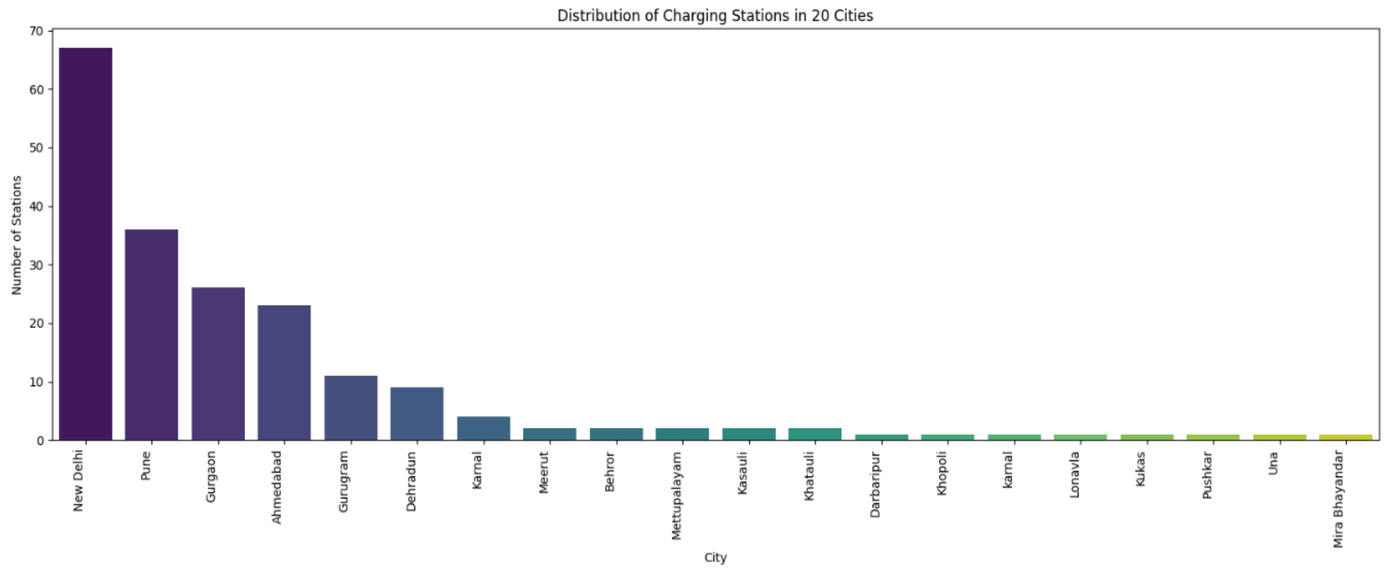
Count plot for displaying the distribution of charging stations across different states in the dataframe.

The visualization is a bar chart that shows the distribution of charging stations across different states. Each bar represents the count of charging stations in a particular state, providing an overview of the geographical distribution of charging infrastructure in the dataset.



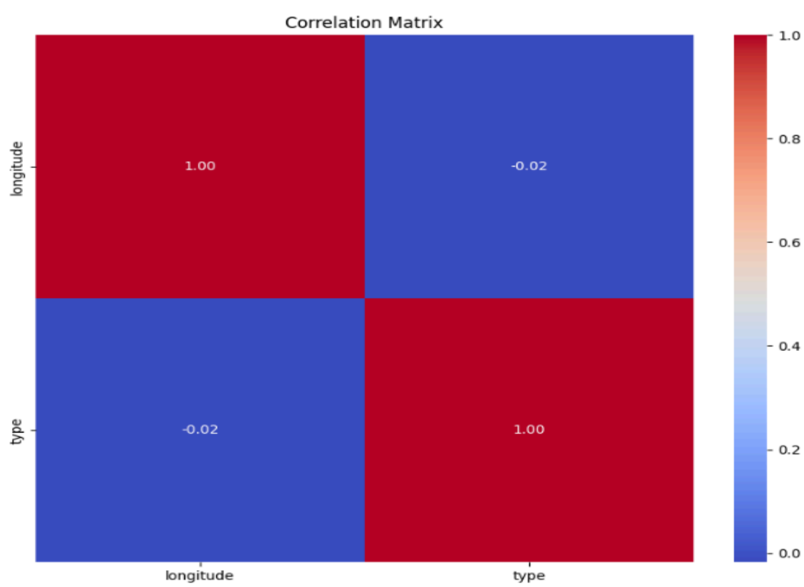
Count plot to visualize the distribution of charging stations in the top 20 cities.

The visualization is a bar chart that shows the distribution of charging stations in the top 20 cities, ordered by the number of stations in each city. This can provide insights into the concentration of charging infrastructure in specific urban areas.



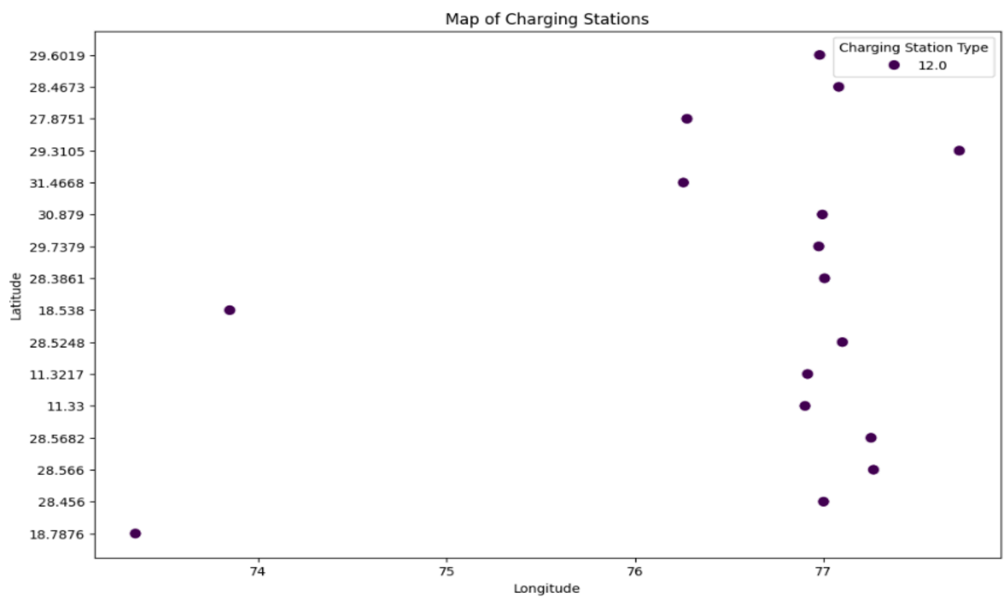
Heatmap of the correlation matrix for the numeric columns in the dataframe.

The heatmap visually represents the correlation matrix, where each cell corresponds to the correlation coefficient between two variables. The color intensity and direction indicate the strength and nature of the correlation (positive or negative).



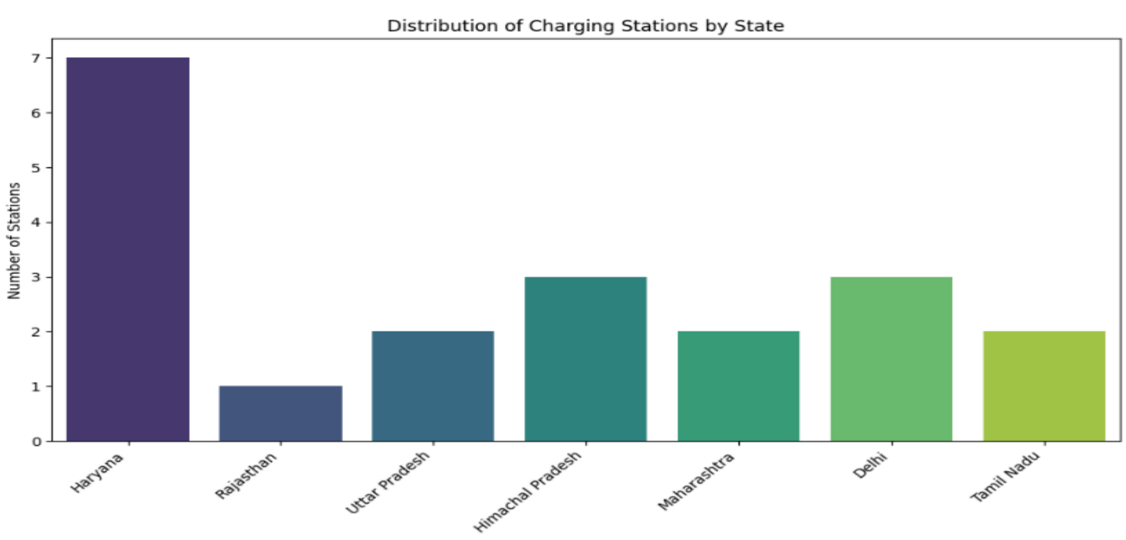
Scatter plot to visualize the geographic distribution of the first 20 charging stations in the dataframe.

The color of each point representing a charging station is determined by the 'type' column, providing information about the type of each charging station.



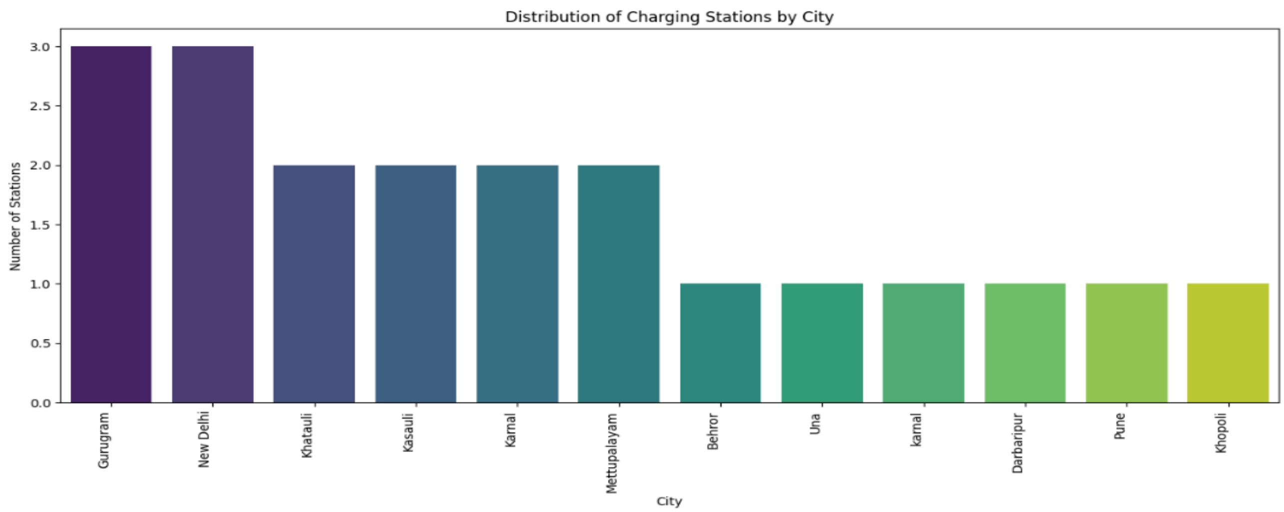
Count plot to visualize the distribution of charging stations by state for the first 20 charging stations in the dataframe.

Each bar represents the count of charging stations in a particular state, providing insights into the geographical distribution of the initial set of charging stations.



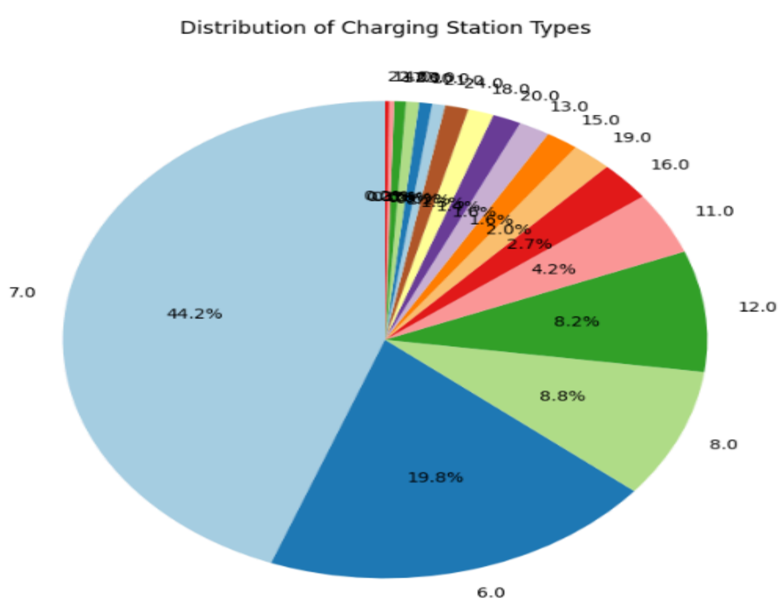
Count plot to visualize the distribution of charging stations by city for the first 20 charging stations in the dataframe.

Each bar represents the count of charging stations in a particular city, providing insights into the concentration of charging infrastructure in specific urban areas within the initial set of charging stations.



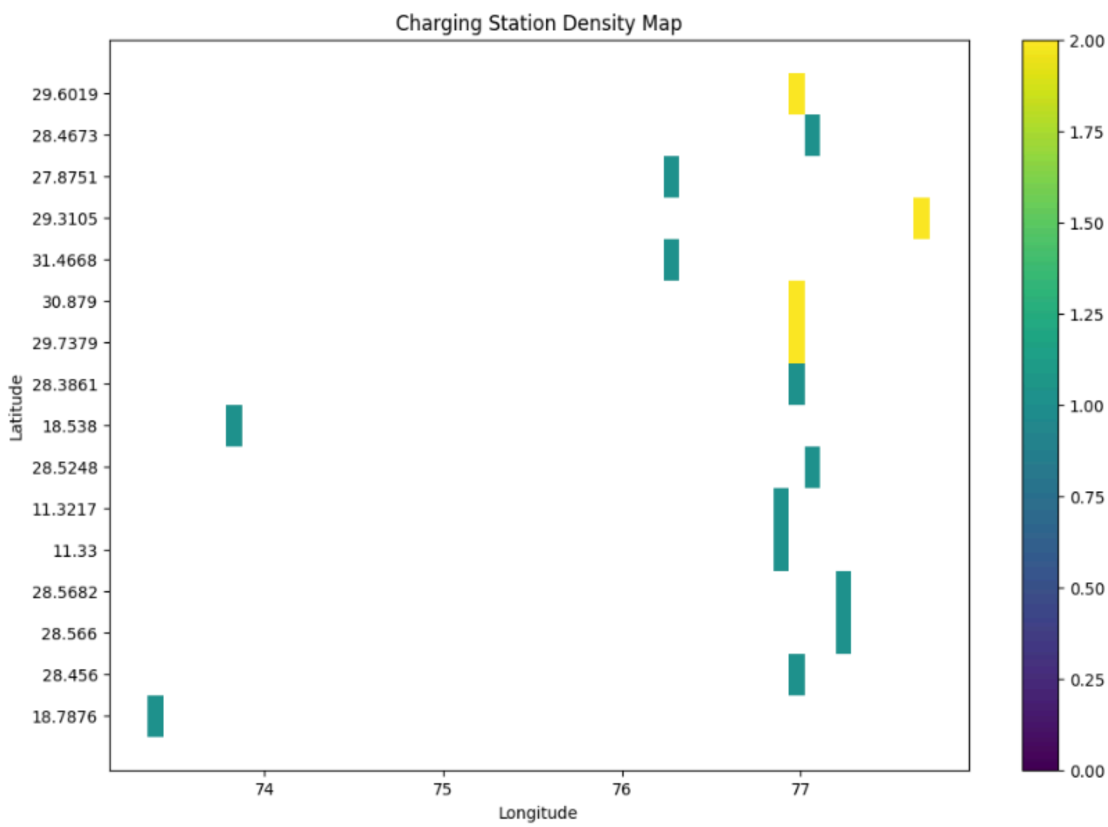
Pie chart to visualize the distribution of charging station types in the dataframe.

The pie chart illustrates the proportion of each charging station type in the dataset. Each slice represents a charging station type, and the size of the slice corresponds to the percentage of that type in the dataset.



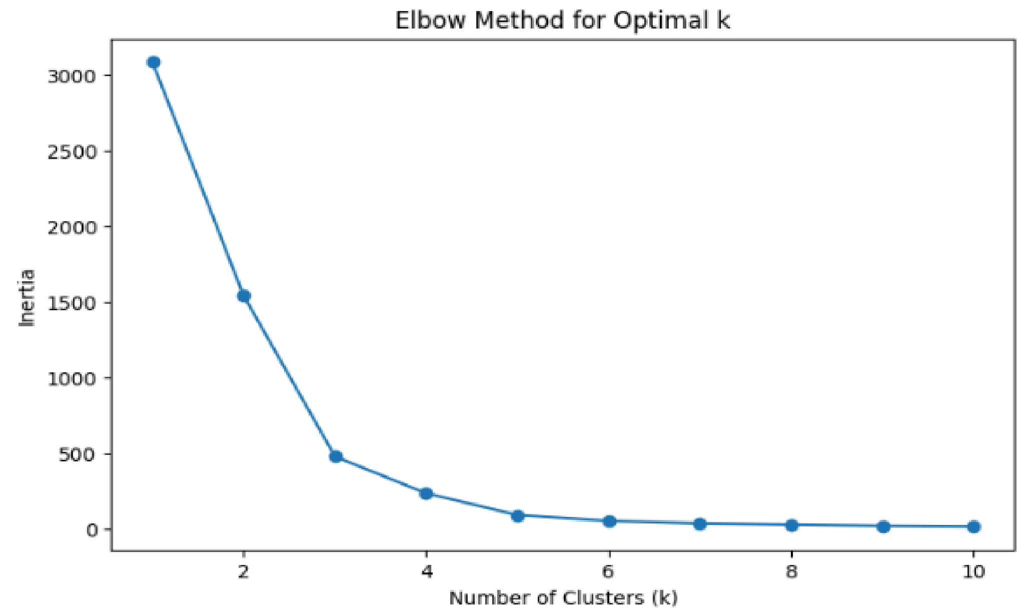
2D histogram (density map) to visualize the density of charging stations based on their geographical coordinates (longitude and latitude) for the first 20 charging stations in the dataframe.

The density map shows the concentration of charging stations in different geographical areas. Darker regions on the map indicate higher charging station density, while lighter regions indicate lower density. This type of plot can be useful for identifying hotspots or patterns in the distribution of charging infrastructure.



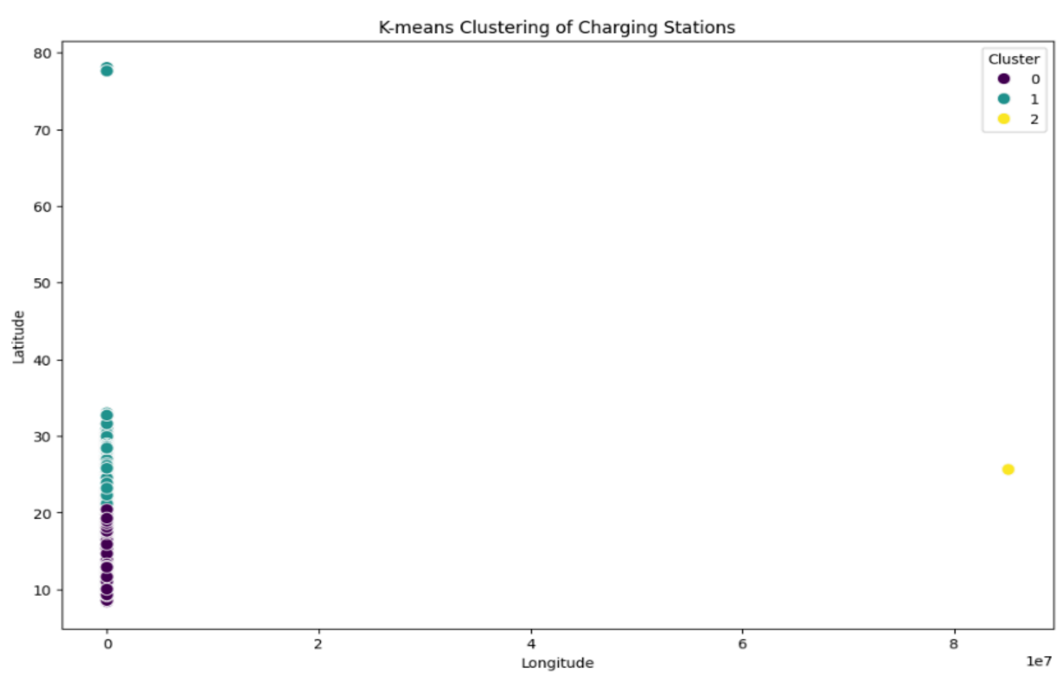
Elbow Method Plot to determine the optimal number of clusters (k) for a k-means clustering algorithm.

The Elbow Method is used to visually identify the "elbow" point on the plot where the inertia starts to decrease at a slower rate. This point is often considered the optimal number of clusters for the k-means algorithm. In the plot, you'll want to look for the point where adding more clusters doesn't significantly reduce the inertia. The idea is to choose a k value that provides a good balance between minimizing inertia and avoiding overfitting.



Scatter plot to visualize the results of K-means clustering applied to charging stations in the dataframe.

The scatter plot shows the geographical distribution of charging stations, where each point is colored based on the cluster to which it belongs according to the K-means clustering algorithm. This helps visualize how the algorithm has grouped the charging stations based on their geographical coordinates.



DATASET 3:

Columns in the dataset include:

	Age	City	Profession	Marital Status	Education	No. of Family members	Annual Income	Would you prefer replacing all your vehicles to Electronic vehicles?	If Yes/Maybe what type of EV would you prefer?
0	30	Nabha	None	Single	Graduate	5	1193875.647	Maybe	SUV
1	27	Pune	None	Single	Graduate	4	1844540.398	Yes	SUV
2	32	Kashipur	None	Single	Graduate	4	2948150.113	Yes	Hatchback
3	55	Pune	Business	Single	Graduate	3	2832379.739	Maybe	Hatchback
4	26	Satara	None	Single	Graduate	4	2638750.576	Yes	Sedan

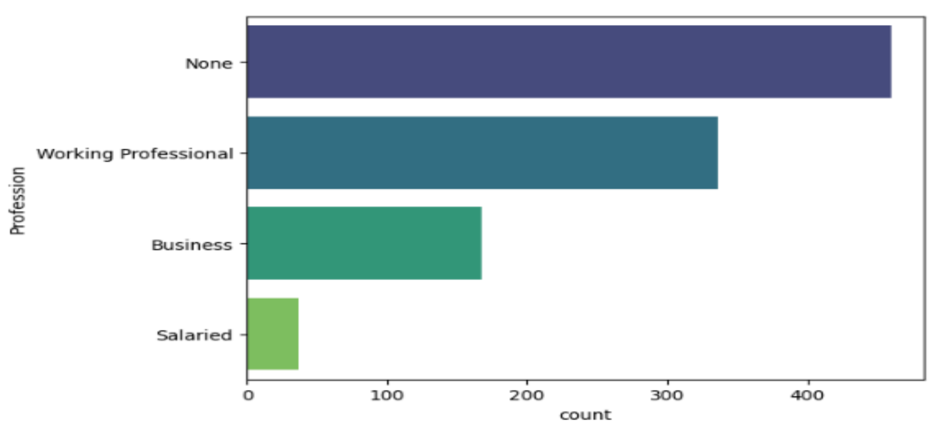
Do you think Electronic Vehicles are economical?	Which brand of vehicle do you currently own?	How much money could you spend on an Electronic vehicle?	Preference for wheels in EV	Do you think Electronic vehicles will replace fuel cars in India?
Yes	Hyundai	<5 lakhs	2	I don't think so
Yes	Honda	<15 lakhs	4	Yes, in <20years
Yes	KIA	<15 lakhs	4	Yes, in <20years
No	Hyundai	<5 lakhs	4	Yes, in <10 years
Yes	McLaren	<15 lakhs	4	Yes, in <20years

Renaming the columns of the dataframe to make them more concise or user-friendly.

	Age	City	Profession	Marital Status	Education	Family_Members	Income	Standard_To_EV	Type	Economical	Current_Brand	Budget	Wheels	Evolution_Of_EV
0	30	Nabha	None	Single	Graduate	5	1193875.647	Maybe	SUV	Yes	Hyundai	<5 lakhs	2	I don't think so
1	27	Pune	None	Single	Graduate	4	1844540.398	Yes	SUV	Yes	Honda	<15 lakhs	4	Yes, in <20years
2	32	Kashipur	None	Single	Graduate	4	2948150.113	Yes	Hatchback	Yes	KIA	<15 lakhs	4	Yes, in <20years
3	55	Pune	Business	Single	Graduate	3	2832379.739	Maybe	Hatchback	No	Hyundai	<5 lakhs	4	Yes, in <10 years
4	26	Satara	None	Single	Graduate	4	2638750.576	Yes	Sedan	Yes	McLaren	<15 lakhs	4	Yes, in <20years

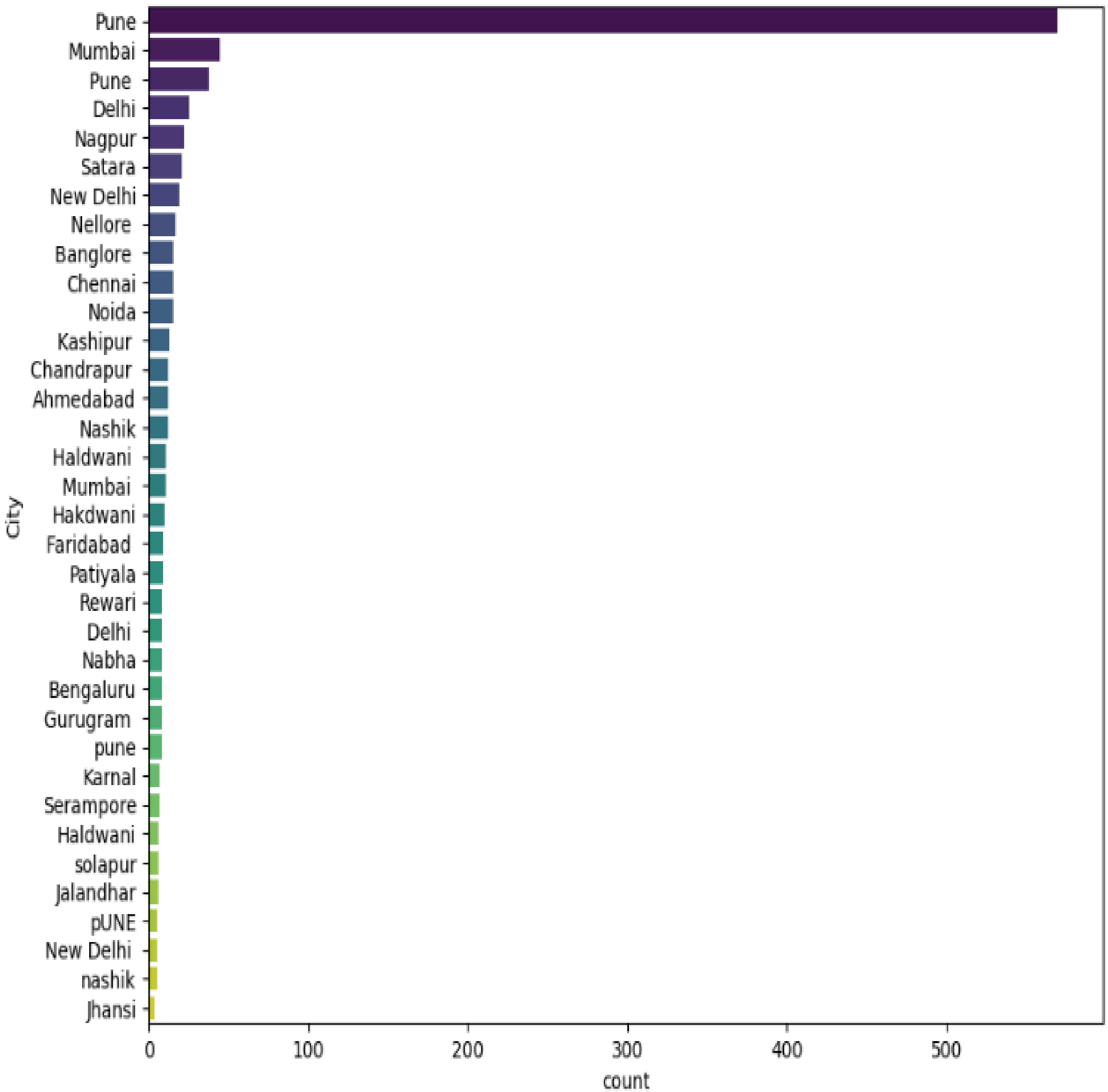
Count plot of the 'Profession' column in the dataframe.

The bars are ordered based on the frequency of each profession.

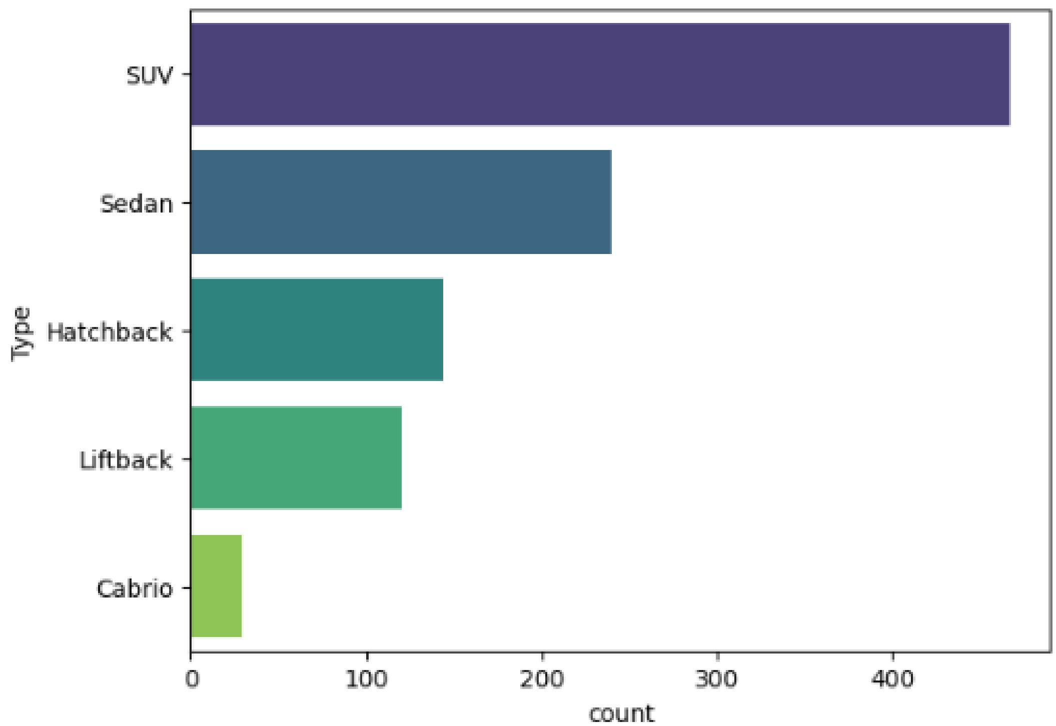


Count plot of the 'City' column in the dataframe.

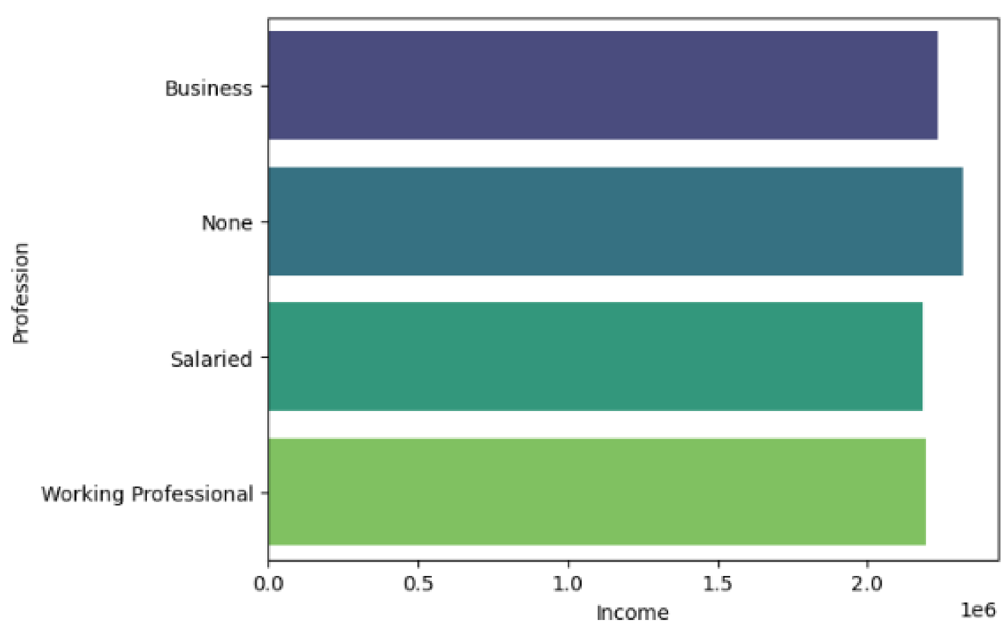
The bars are ordered based on the frequency of respondents from each city. The larger the bar, the more respondents from that particular city.



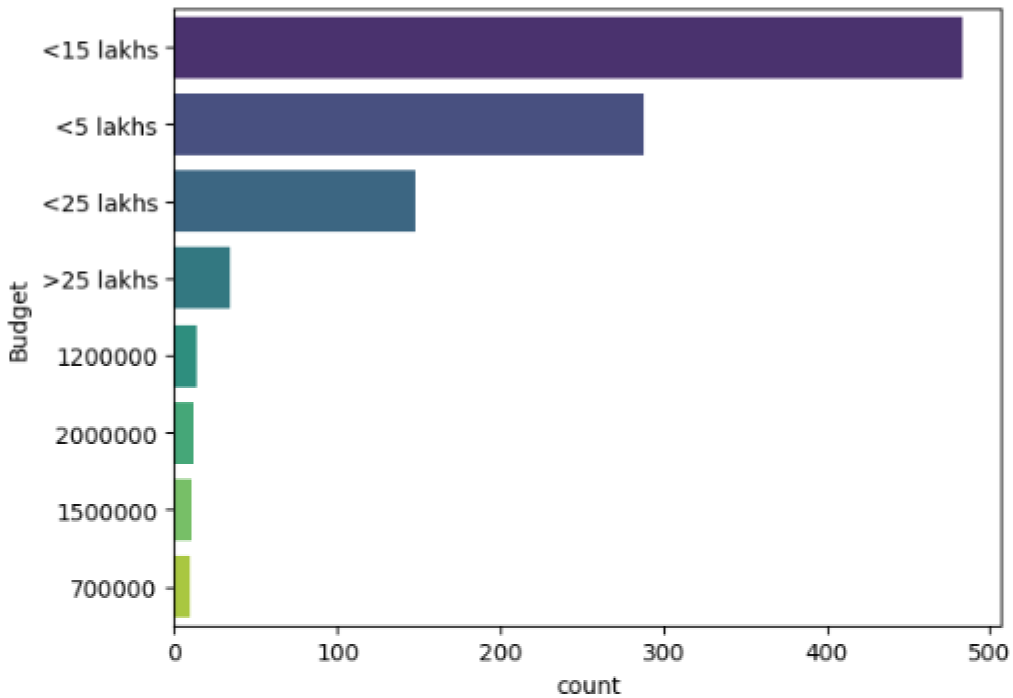
Count plot for the 'Type' column



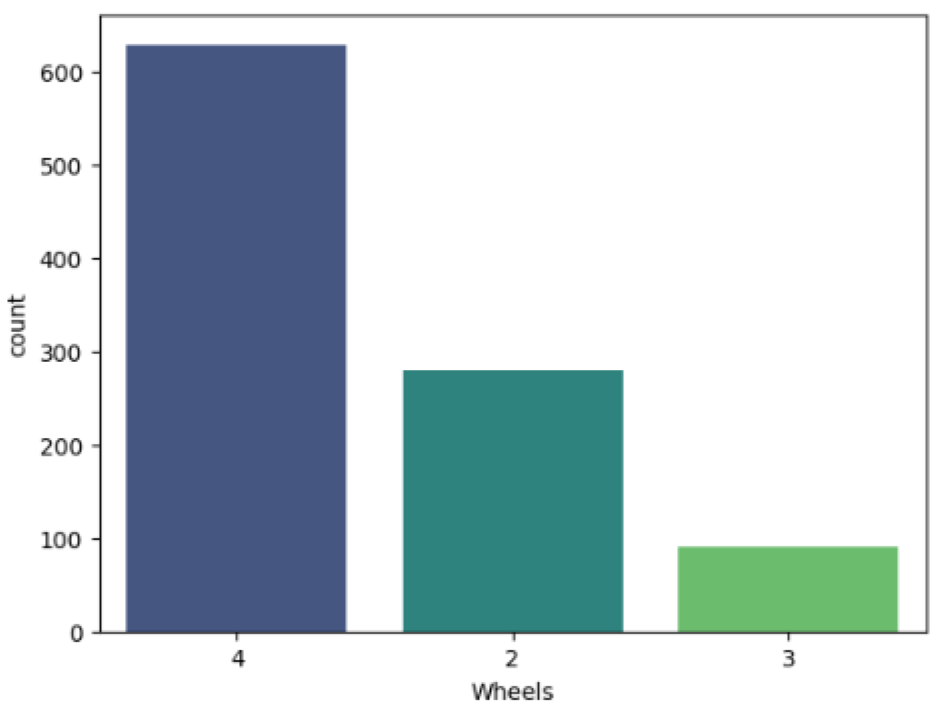
Horizontal bar plot that visualizes the average income (`Income`) for each profession (`Profession`) in the dataframe.



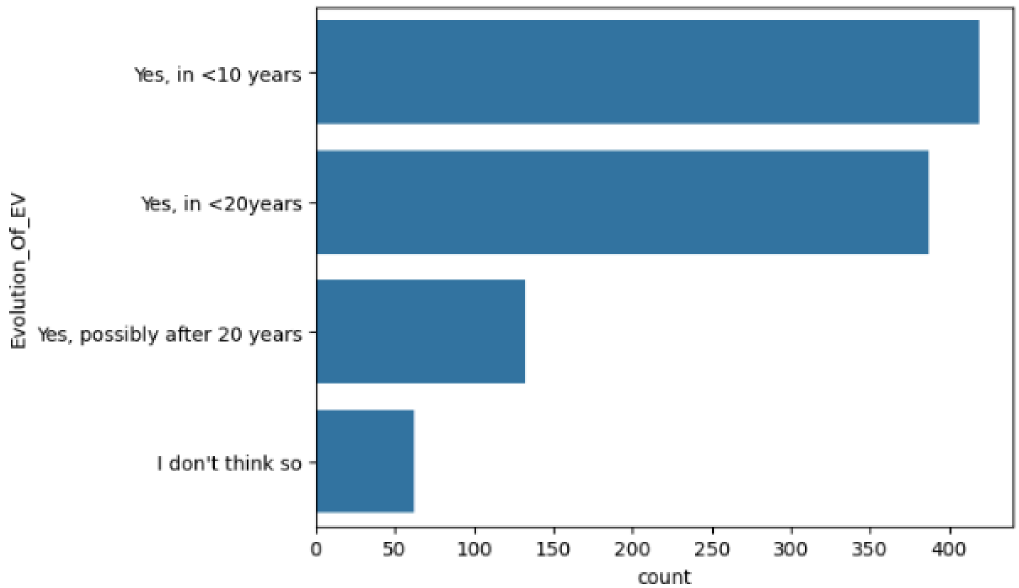
Count plot of the 'Budget' column in the dataframe.



Count plot of the 'Wheels' column in the dataframe.

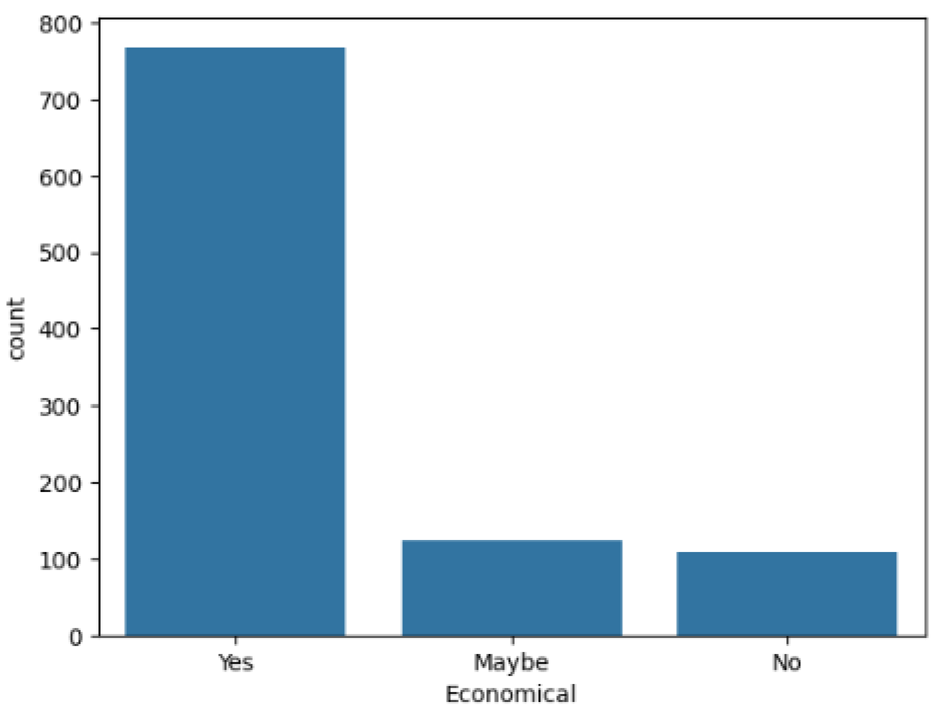


Count plot of the 'Evolution_Of_EV' column in the dataframe.



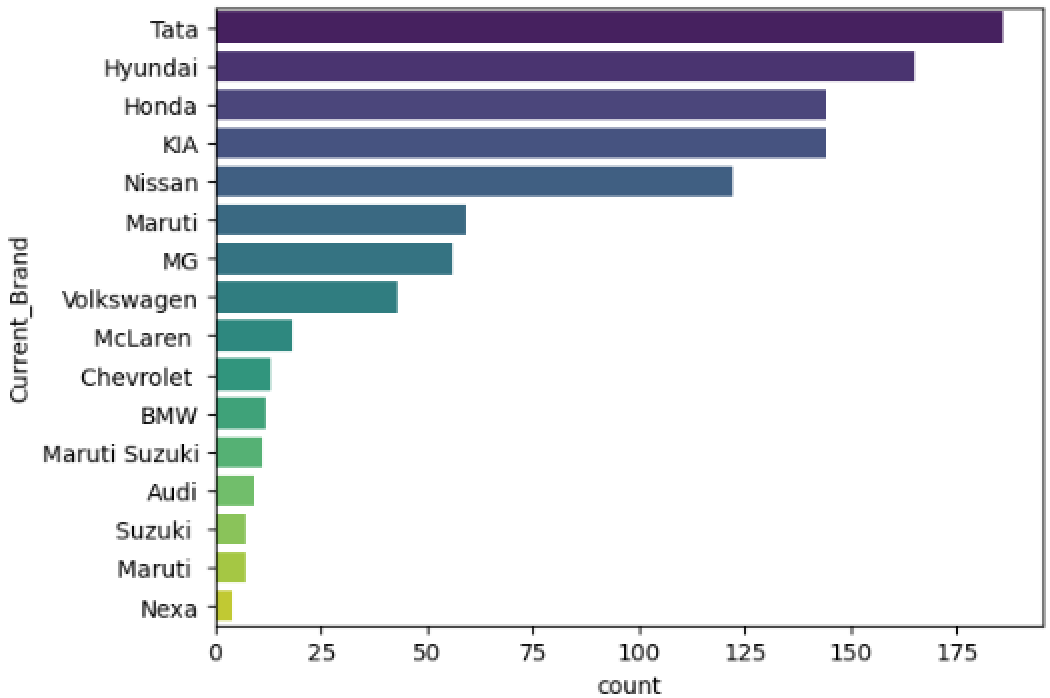
Count plot of the 'Economical' column in the dataframe.

The larger the bar, the more respondents share that particular opinion on the economic viability of electric vehicles.



Count plot of the 'Current_Brand' column in the dataframe.

The larger the bar, the more respondents own vehicles from that particular brand.



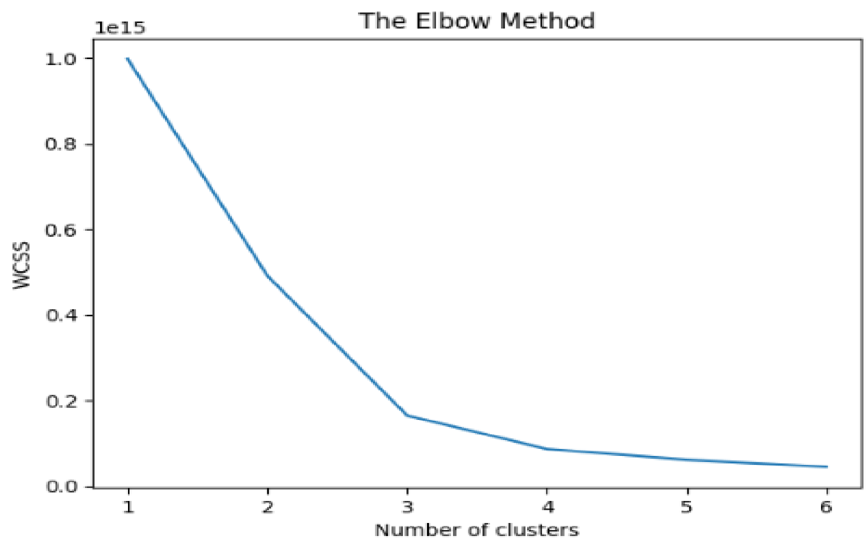
Transformation of Variables

	Age	City	Marital Status	Family_Members	Income	Type	Economical	Current_Brand	Budget	Wheels
0	30	18	0	5	1193875.65	3	1	4	5	2
1	27	26	0	4	1844540.40	3	1	3	15	4
2	32	15	0	4	2948150.11	1	1	5	15	4
3	55	26	0	3	2832379.74	1	0	4	5	4
4	26	29	0	4	2638750.58	4	1	10	15	4

Evolution_Of_EV	Post_Graduate	Standard_to_EV	Profession_None	Profession_Salaried	Profession_Working Professional	ClusterLabel
0	0	0	1	0	0	1
20	0	1	1	0	0	4
20	0	1	1	0	0	3
10	0	0	0	0	0	3
20	0	1	1	0	0	3

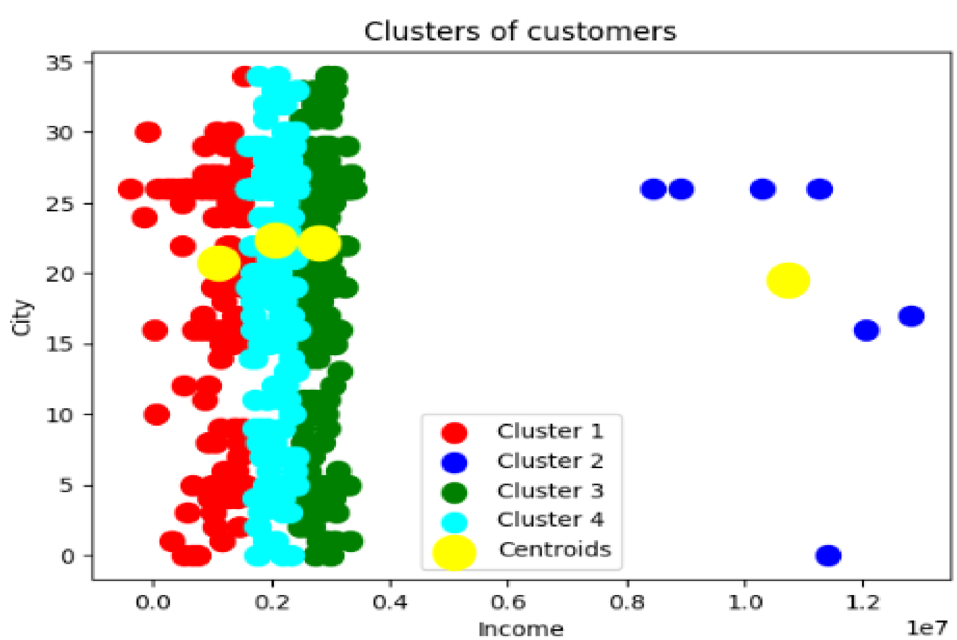
Elbow Method Plot using the k-means clustering algorithm to find the optimal number of clusters (k) for a given dataset.

This method helps in choosing the most appropriate value for the number of clusters in k-means clustering based on the trade-off between model simplicity and accuracy.



Scatter plot to visualize the clusters generated by the k-means algorithm.

It helps in understanding how the clusters are distributed in the feature space (Income and City in this case) and where the centroids of each cluster are located.



Applying a classifier to a test set and evaluating its performance using a confusion matrix and accuracy score.

1. Confusion Matrix:
    - Rows represent the actual classes (true labels).
    - Columns represent the predicted classes.
    - For example, in the first row, the model predicted 34 instances of class 3 when the true label was also class 3, and 14 instances of class 4 when the true label was class 3.

2. Accuracy Score: 0.365
    - The accuracy score is a measure of how many predictions the model got correct out of the total number of predictions.
    - In this case, the model achieved an accuracy of 36.5%, meaning that 36.5% of the predictions were correct.

```
[[ 0  0 34 14]
 [ 0  0  2  0]
 [ 0  0 63 18]
 [ 0  0 59 10]]
0.365
```

Applying classifier to the training set and evaluating its performance using a confusion matrix and accuracy score.

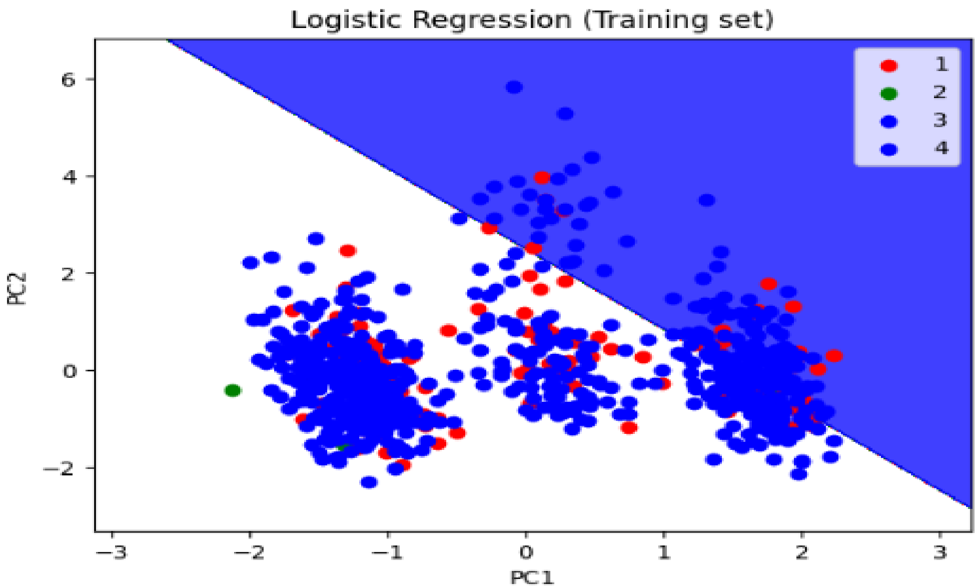
1. Confusion Matrix:
    - Rows represent the actual classes (true labels).
    - Columns represent the predicted classes.
    - For example, in the first row, the model predicted 114 instances of class 3 when the true label was also class 3, and 33 instances of class 4 when the true label was class 3.

2. Accuracy Score: 0.4530663329161452
    - In this case, the model achieved an accuracy of 45.3%, meaning that 45.3% of the predictions were correct on the training set.

```
[[  0   0 114  33]
 [  0   0   5   0]
 [  0   0 286  63]
 [  0   0 222  76]]
0.4530663329161452
```

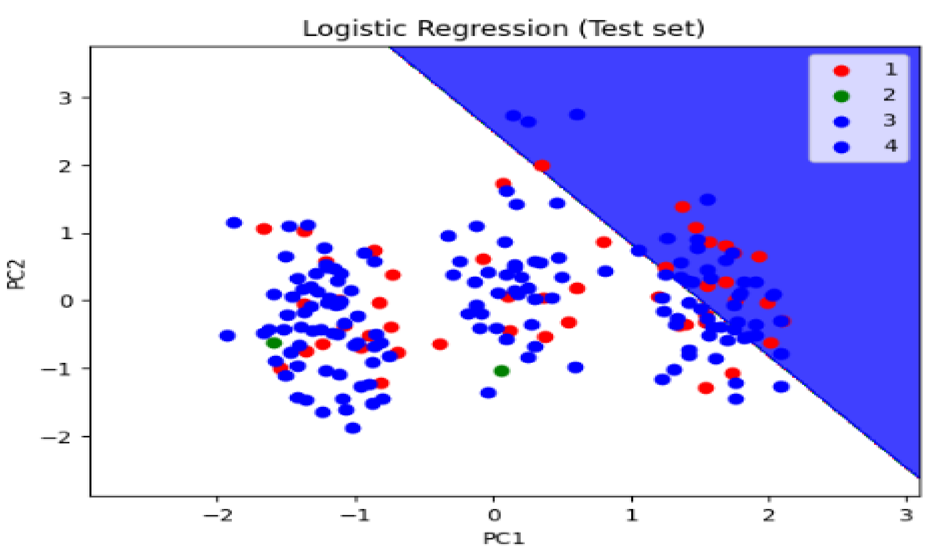
Contour plot to visualize the decision boundaries of a logistic regression classifier on the training set.

This visualization helps you understand how the logistic regression model has learned to separate the different classes in the feature space.



Contour plot to visualize the decision boundaries of a logistic regression classifier on the test set.

This visualization helps you understand how well the logistic regression model generalizes to the test set and how the decision boundaries perform on unseen data.



CUSTOMIZING THE MARKETING MIX

Customizing the marketing mix in the electric vehicle (EV) industry involves tailoring key elements to address the unique characteristics and preferences of the target segment. In the context of EVs, product customization involves developing models with diverse features and accommodating varying consumer needs.

Pricing Strategy:

Offer a diverse range of electric vehicle (EV) models to cater to various consumer preferences, including compact city cars, sedans, SUVs, and electric two-wheelers. Emphasize features such as extended range, fast-charging capabilities, advanced safety features, and smart connectivity options to enhance the appeal of the EV lineup.

Price Competitiveness:

Implement competitive pricing strategies to ensure affordability and attractiveness in the Indian market. Utilize government incentives, subsidies, and tax benefits to lower upfront costs, making EVs more cost-effective than traditional vehicles. Explore flexible financing options, affordable lease programs, and enticing warranty packages to incentivize adoption and ease financial considerations for potential customers.

Promotion and Awareness:

Launch targeted marketing campaigns to raise awareness about the benefits of EVs, emphasizing their positive environmental impact and long-term cost savings. Develop educational initiatives to dispel misconceptions and provide information about EV technology, charging infrastructure, and available government incentives. Collaborate with influencers, celebrities, and industry experts to enhance credibility and generate enthusiasm around EVs. Participate actively in EV expos, trade shows, and events to showcase the latest models and technologies, fostering a direct connection with potential consumers.

POTENTIAL CUSTOMER BASE (FOR BUSINESS MARKETS)

In the business markets of the electric vehicle (EV) industry, identifying the potential customer base is crucial for estimating early market sales and profitability. By understanding the potential customer base and setting an appropriate target price range, businesses can estimate the potential profit in the early market. This strategic approach allows companies to focus efforts on sectors most inclined towards EV integration, maximizing sales opportunities and establishing a strong foundation for profitability in the evolving landscape of sustainable transportation.

The regional landscape of electric vehicle (EV) adoption in India reflects distinctive patterns. Maharashtra emerges as a frontrunner, leading the nation in overall EV usage, particularly in the 2-wheeler and 4-wheeler categories. Karnataka closely follows with notable 2-wheeler EV adoption. Uttar Pradesh takes the lead in 3-wheeler EVs, showcasing a unique vehicular preference. In the 4-wheeler segment, Maharashtra and Delhi exhibit the highest numbers, emphasizing urban EV utilization. Electric buses, a nascent yet impactful sector, are currently operational in only 6 states. Maharashtra, Andhra Pradesh, and Tamil Nadu stand out as key contributors to the EV revolution by leading in the production of electric vehicle chargers, marking a pivotal stride towards sustainable mobility.

A substantial portion of survey participants currently own vehicles from reputable brands like Tata, Hyundai, Honda, and Kia. Their inclination towards SUVs and Sedans as their preferred electric vehicle aligns with a preference for larger and more versatile transportation options. Budget considerations also play a crucial role, with a significant preference for electric vehicles priced below 15 lakh rupees, indicating a demand for affordable options. Interestingly, a comparatively smaller percentage expresses interest in two-wheeler EVs. The overwhelmingly positive sentiment regarding the evolution of electric vehicles, with the majority anticipating market advancements within the next 10 years, signifies a promising outlook for the electric mobility landscape.

The strategic analysis outlines a comprehensive approach for the successful entry and growth of the electric vehicle (EV) company. Mumbai and Delhi, being economically prosperous states, are identified as key markets due to their higher purchasing power. While Delhi exhibits widespread electric vehicle adoption, the need for an increased number of chargers is recognized to bolster production and compete effectively with Mumbai. The target segment is strategically defined, focusing on individuals in Mumbai, Pune, and Ahmedabad with an annual income between 8 to 12 lakhs. This demographic is deemed more inclined to transition from conventional to electric vehicles, aligning with their income considerations. This approach ensures a methodical and adaptable trajectory for the company's early phases, setting the stage for sustainable growth in the dynamic EV market.

MOST OPTIMAL MARKET SEGMENTS

Target Market Strategy for EV Entry:

Demographic Segments

1. ***Age:*** The primary target age group is 28-31, aligning with the profile of young professionals who are likely to be more receptive to new technologies and environmentally conscious choices.

2. ***Income:*** The target income bracket is set between Rs. 20,64,995 to Rs. 28,12,149 annually, focusing on individuals with sufficient purchasing power to afford electric vehicles.

3. ***Education:*** The educational focus is on graduates, as they tend to be more open to adopting new technologies and are likely to have a higher awareness of the benefits of electric vehicles.

Geographic Segments

1. ***Location:*** Pune and Mumbai emerge as the primary markets for early market creation based on their suitability. New Delhi and Haldwani are identified as the next priority locations, followed by Satara and Bengaluru. The selection is based on a combination of market potential, infrastructure support, and demographic alignment.

2. ***Charging Infrastructure:*** Cities selected for early market creation should have robust charging station infrastructure, a parameter to be analyzed using the Sanctioned Charging Stations in India Dataset.

Psychographic Segments

1. ***Lifestyle:*** The psychographic focus is on individuals with a single marital status and families with 3-5 members. This aligns with the observed interest in EVs among this demographic, indicating a potential market segment with specific lifestyle preferences.

Brand Preferences

The target market should take into consideration the brand preferences of the identified demographic. Tata, Hyundai, Honda, and KIA are the most owned brands, each associated with specific attributes. This insight can guide marketing strategies and product positioning.

Vehicle Preferences:

Considering the observed preferences for SUVs, sedans, and hatchbacks, the company should focus on producing sedans, given the majority preference for this vehicle type. Additionally, understanding the willingness to spend on different vehicle types helps in strategic decision-making.
The most optimal market segments, as identified through comprehensive market research and segmentation analysis, are strategically chosen based on factors such as consumer preferences, purchasing behavior, and demographic trends. These segments exhibit the highest potential for growth, profitability, and alignment with the product offering. By targeting these specific market segments, businesses can optimize their marketing strategies, tailor product features to meet consumer needs, and efficiently allocate resources. This targeted approach enhances market penetration, fosters customer satisfaction, and positions the business for sustainable success in the competitive landscape.

FERMI ESTIMATION

Fermi Estimation Breakdown for EV Market in India:

1. Overall EV Market Size:
    - Estimate: Utilize recent reports and government data to estimate the total size of the electric vehicle market in India.
    - Estimated Value: ***Approximately 2 million EVs in the market.***

2. Geographic Segmentation:
    - Estimate: Consider infrastructure development and population density to estimate the regional distribution of EV adoption.
    - Estimated Value: ***Highest adoption in Maharashtra and Karnataka, followed by Delhi, Uttar Pradesh, and Tamil Nadu.***

3. Demographic Segmentation:
    - Estimate: Analyze age, gender, income, and profession data to estimate the distribution of potential customers.
    - Estimated Value: ***Majority in the 28-31 age group, with an annual income between Rs. 20,64,995 to Rs. 28,12,149.***

4. Psychographic Segmentation:
    - Estimate: Assess lifestyle preferences and environmental awareness to estimate psychographic factors influencing EV adoption.
    - Estimated Value: ***Targeting environmentally conscious individuals with a preference for sustainable living.***

5. Behavioral Segmentation:
    - Estimate: Analyze data related to vehicle usage, commuting habits, and preferences to understand behavioral patterns.
    - Estimated Value: ***Majority interested in SUVs, sedans, and hatchbacks; limited interest from the salaried population.***

6. Availability of Data:
    - Estimate: Evaluate the reliability and completeness of available datasets for each segmentation variable.
    - Estimated Value: ***Data availability is moderate, with some gaps in income distribution by geographic region.***

7. Machine Learning Techniques:
    - Estimate: Assess the complexity and feasibility of implementing clustering algorithms for segmentation.
    - Estimated Value: ***Feasible to implement clustering algorithms for demographic and behavioral segmentation.***

8. Potential Customer Base and Profit Calculation:
    - Estimate: Use demographic and psychographic data to estimate the potential customer base and calculate potential profit.
    - Estimated Value: ***Potential profit of Rs. 500 crores in the early market with a focus on sedan production for the 28-31 age group.***

This Fermi Estimation breakdown provides a structured approach to understanding and estimating key aspects of the Electric Vehicle market in India, laying the foundation for a data-driven market entry strategy.

CONCLUSION

In the pursuit of entering the burgeoning Electric Vehicle (EV) market in India, a meticulous and data-driven market segmentation process has been undertaken. By employing Fermi estimation, I delved into diverse facets, starting with estimating the overall EV market size, strategically breaking it down into geographical, demographic, psychographic, and behavioral segments. The process aimed to provide a nuanced understanding of the Indian market landscape.

Geographic segmentation revealed the dominance of Maharashtra and Karnataka, followed by key markets in Delhi, Uttar Pradesh, and Tamil Nadu. The demographic focus on the 28-31 age group, coupled with an income bracket of Rs. 20,64,995 to Rs. 28,12,149 and an emphasis on graduates, provides a clear profile for targeted marketing. Psychographic insights emphasized environmental consciousness and sustainable living preferences, shaping the positioning strategy for EVs.

Behavioral segmentation showcased preferences for SUVs, sedans, and hatchbacks, with a notable disinterest among salaried individuals. The evaluation of available data and the feasibility of machine learning techniques laid the groundwork for robust segmentation implementation.

In conclusion, this comprehensive market segmentation process is poised to guide the EV startup's entry into the Indian market strategically. The refined understanding of diverse consumer segments, coupled with an assessment of market dynamics, sets the stage for a well-informed, targeted, and adaptive market entry strategy in the dynamic landscape of the Indian EV market. Ongoing research, adaptability, and a responsiveness to evolving consumer preferences will be pivotal for sustained success and growth in this transformative industry.

LINK TO GITHUB PROFILE

GitHub Profile Link: https://github.com/sonateresa/feynnlabs-sona

Visit the link to access comprehensive documentation, code implementations, and datasets for a detailed overview of the projects.