**Introduction**

The electric vehicle market in India is still at a nascent stage but has tremendous potential for growth. The government has announced several initiatives and incentives to promote electric mobility, including subsidies, tax benefits, and the FAME (Faster Adoption and Manufacturing of Hybrid and Electric Vehicles) scheme. As a result, several domestic and international companies have entered the market with a range of electric vehicles, including two-wheelers, three-wheelers, and four-wheelers.

However, the market faces several challenges, including high upfront costs, limited charging infrastructure, and range anxiety. These challenges highlight the importance of market segmentation analysis to understand the needs and preferences of different consumer segments and develop strategies to address them.

**Deciding to Segment**

In the context of the electric vehicle market in India, market segmentation analysis is crucial for several reasons:

**Understanding Consumer Needs:** India is a diverse country with varying consumer preferences and needs. Market segmentation analysis helps to identify these diverse needs and tailor products and services accordingly. For example, urban consumers may prioritize convenience and charging infrastructure, while rural consumers may prioritize affordability and range.

**Targeting Specific Segments:** By identifying and targeting specific segments, companies can optimize their marketing efforts and resources. For instance, targeting segments with high potential for adoption can lead to faster market penetration and higher sales.

**Product Development:** Market segmentation analysis can provide insights into the features and specifications that are most important to different segments. This information can be used to develop products that are tailored to the needs of specific segments, increasing the likelihood of success in the market.

**Customized Marketing Strategies:** Different segments may require different marketing strategies to effectively reach and engage them. Segmentation analysis helps to identify the most effective channels, messages, and promotions for each segment.

**Competitive Advantage:** Understanding the market segments and their needs can provide a competitive advantage by enabling companies to differentiate their offerings and position themselves effectively in the market.

**Specifying the Ideal Target Segment (Fermi Estimation)**

*Electric vehicle (EV) configurations and specifications for 4-wheelers are a crucial aspect of the EV market, shaping consumer preferences and influencing purchasing decisions.*

Analysing this segment can provide valuable insights into the market dynamics, help companies tailor their products and marketing strategies to meet consumer needs effectively and capitalize on the growing demand for electric vehicles.

One key aspect of EV configurations for 4-wheelers is the vehicle segment. Each type of vehicle caters to different consumer segments based on factors like affordability, size, and usage patterns such as personal use, commuting or commercial purposes.

Another important aspect is the specifications of the EVs, including battery capacity, charging infrastructure, range, and performance metrics. Battery capacity and range are critical factors influencing consumer adoption, as higher capacities and longer ranges offer greater convenience and flexibility. For example, the trend towards higher battery capacities and longer ranges reflects advancements in battery technology and efforts to address range anxiety among consumers.

Moreover, understanding consumer preferences for EV configurations and specifications for 4-wheelers can help companies develop targeted marketing strategies and product offerings. For instance, highlighting the environmental benefits of EVs, such as reduced emissions and lower operating costs, can appeal to environmentally conscious consumers. Similarly, offering customizable options for battery capacity or range can cater to different consumer needs and preferences.

**Collecting Data**

The data set contains configuration of 104 4-wheeler vehicles in the market. They provide insights with respect to the following attributes: Brand, Model, Accelerating Sec, Top Speed, Range, Efficiency, Fast Charge, Rapid Charge, Power Train, Plug Type, Body Style, Vehicle Segments, Seats, Price.

**Brand:** The brand refers to the maker of the vehicle.

**Model:** The specific model name of the vehicle.

**Accelerating Sec:** The time it takes for the vehicle to accelerate from 0 to a certain speed in seconds. This is an indicator of the vehicle's performance.

**Top Speed:** The maximum speed (Km/h) that the vehicle can achieve under optimal conditions.

**Range:** The distance (Km) the vehicle can travel on a single charge. This is a key factor influencing the practicality and usability of the vehicle.

**Efficiency:** The efficiency of the vehicle's electric drivetrain, measured in km per kilowatt-hour (km/kWh). Higher efficiency indicates that the vehicle can travel farther on a single charge.

**Fast Charge:** Indicates whether the vehicle supports fast charging, which allows for quicker charging times compared to standard charging. Its values here correspond to Km/h.

**Rapid Charge:** “*Yes*” or “*No*” indicates whether the vehicle supports rapid charging, which is similar but faster than fast charging.

**Power Train:** The type of powertrain used in the vehicle, such as AWD, FWD or RWD.

**Plug Type:** The type of plug used for charging the vehicle. Here the vehicles fall into four categories – Type1 CHAdeMO, Type2, Type2 CHAdeMO, Type2 CCS.

**Body Style:** The body style of the vehicle, such as sedan, SUV, hatchback, etc.

**Vehicle Segments:** The segment or category of the vehicle based on size, price, and features, such as compact, mid-size, or luxury represented as A, B, C, D, E, F, N or S.

**Seats:** The number of seats in the vehicle, indicating its capacity for passengers.

**Price:** The price of the vehicle (in Lakhs), which vary based on the model, features, and market conditions.

**Segmentation Variables**

Segmentation variables are characteristics or attributes used to divide a market into segments. These variables help identify groups of customers with similar needs, preferences, or behaviours, allowing companies to tailor their marketing strategies to each segment. The following are considered segmentation variables:

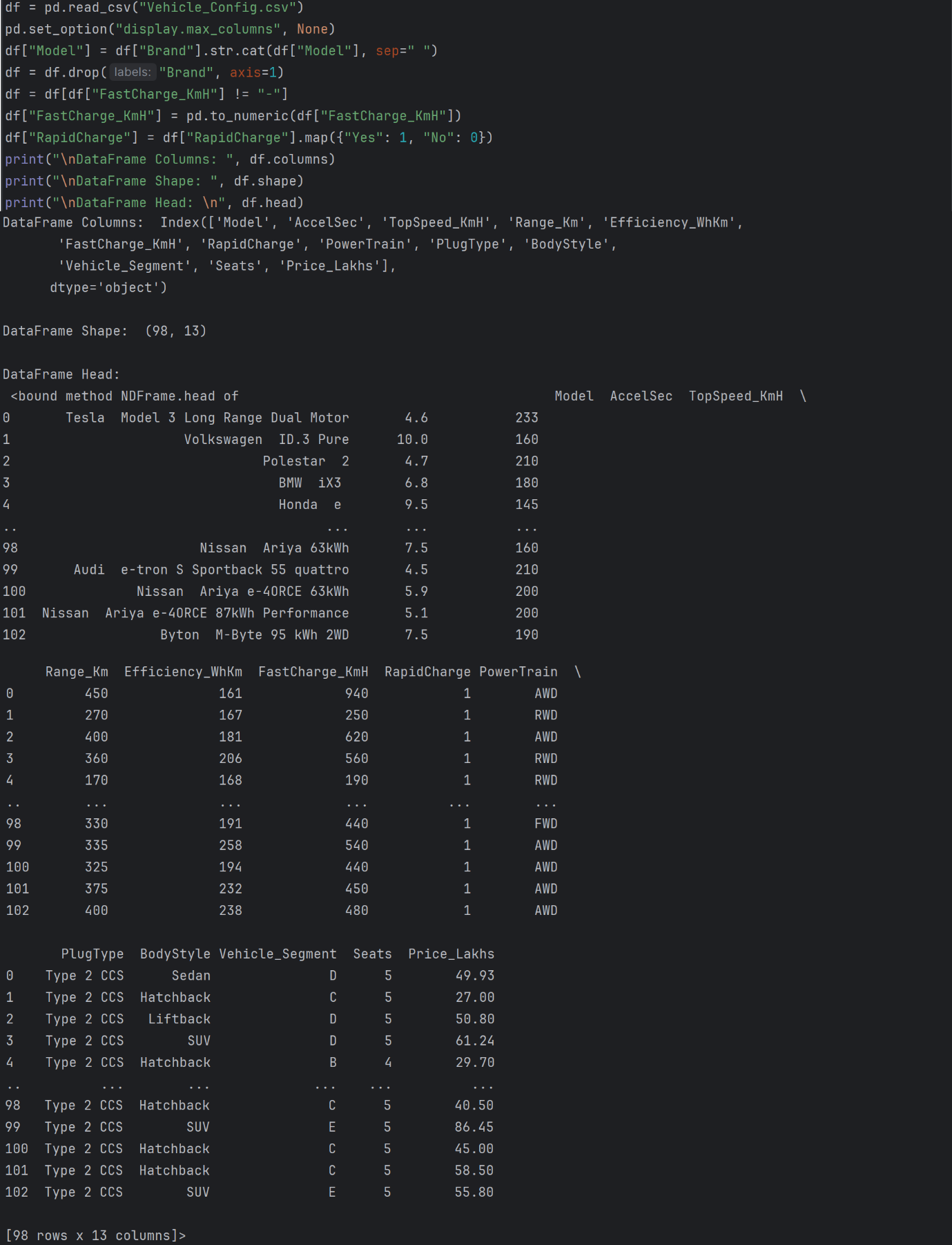
* Accelerating Sec
* Top Speed
* Range
* Efficiency
* Fast Charge
* Rapid Charge
* Seats

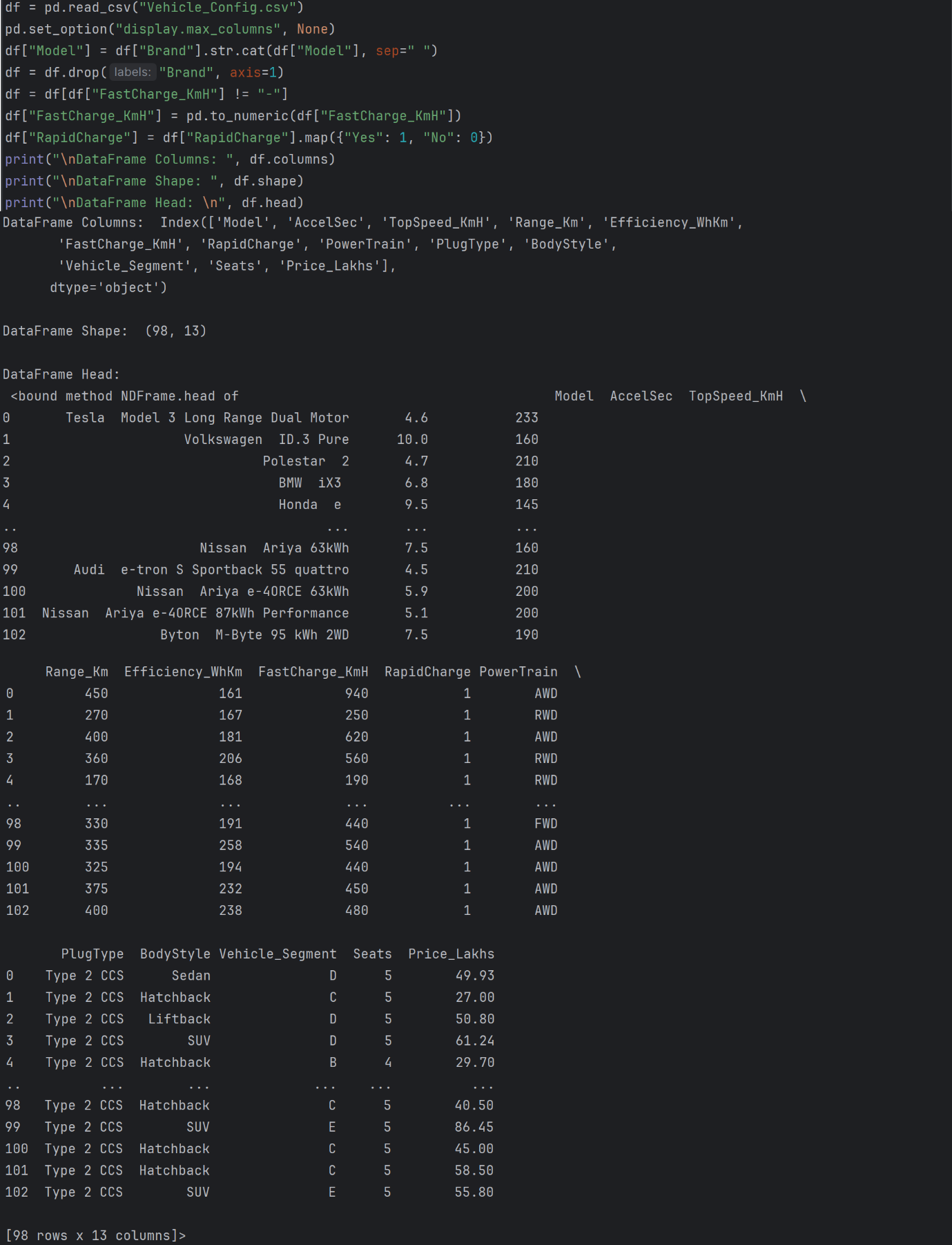
**Exploring Data**

***Binary Numeric Segmentation***

First we load the dataset from a CSV file named "*Vehicle\_Config.csv*" into a pandas DataFrame called **df** and set the option to *pd.set\_option("display.max\_columns", None)* to display all columns in the DataFrame without truncation.

We now inspect basic features such as variable names, the sample size and the first 5 rows of the data.





To Make the DataFrame suitable for segment extraction:

1. We concatenate the "Brand" and "Model" columns in the DataFrame with a space separator, assign the result to a new column called "Model" and drop the "Brand" column from the DataFrame.
2. Map the values in the "RapidCharge" column from strings ("Yes" and "No") to integers (1 and 0).
3. Replace the unknown values in “FastCharge\_KmH” column to 0 and convert it into numeric data.

We retrieve only the segment variables in a new DataFrame called **frame** and calculate the mean for each column in the frame.



The average values of the transformed binary numeric segmentation variables indicate that on an average the vehicles in the market have *Rapid Charge* and 4-5 *Seats*. The average *Top Speed* is 182 Km/h, can travel a *Range* of 350 Km with an *Acceleration* of 7 sec and *Efficiency* of ~190.

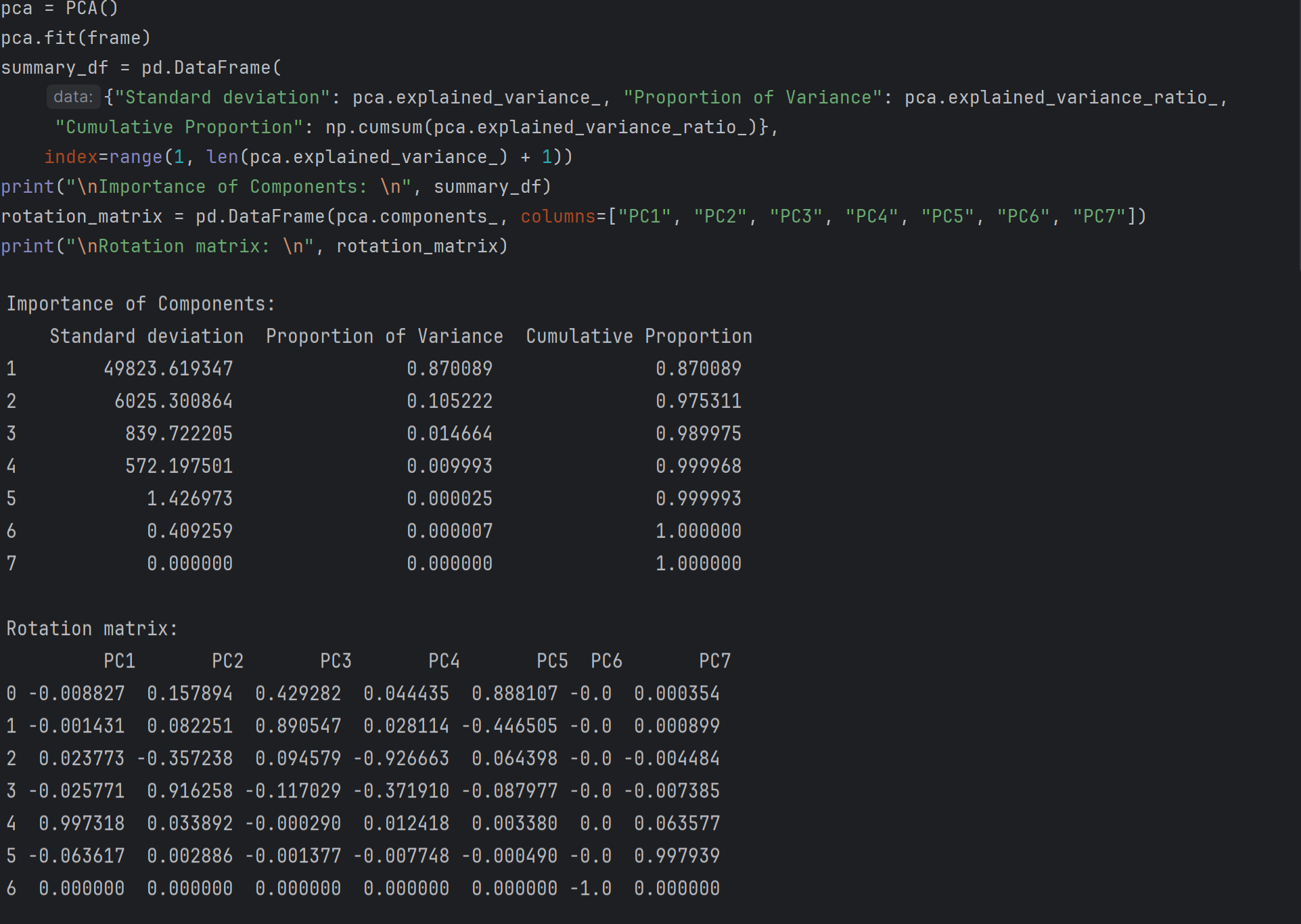
***Principal Component Analysis***

The following code performs **Principal Component Analysis (PCA)** on frame using the *PCA* class from *scikit-learn*.

*summary\_df = pd.DataFrame(...)* creates a DataFrame **summary\_df** containing the following columns:

* **Standard deviation:** The eigenvalues of the covariance matrix, which represent the variance explained by each principal component.
* **Proportion of Variance:** The proportion of the total variance explained by each principal component.
* **Cumulative Proportion:** The cumulative proportion of total variance explained up to each principal component.

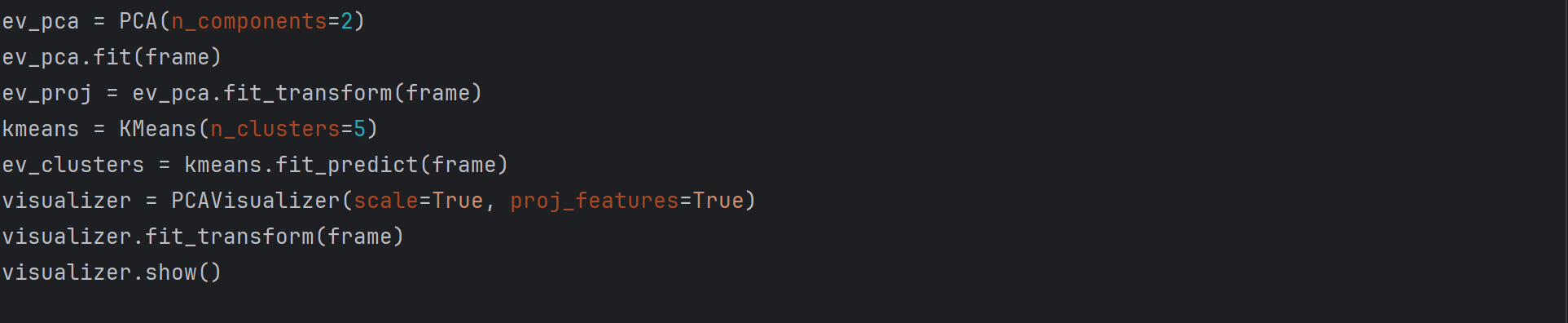
A DataFrame **rotation\_matrix** is created containing the loadings of each original variable on each principal component. Each row represents a principal component, and each column represents an original variable.

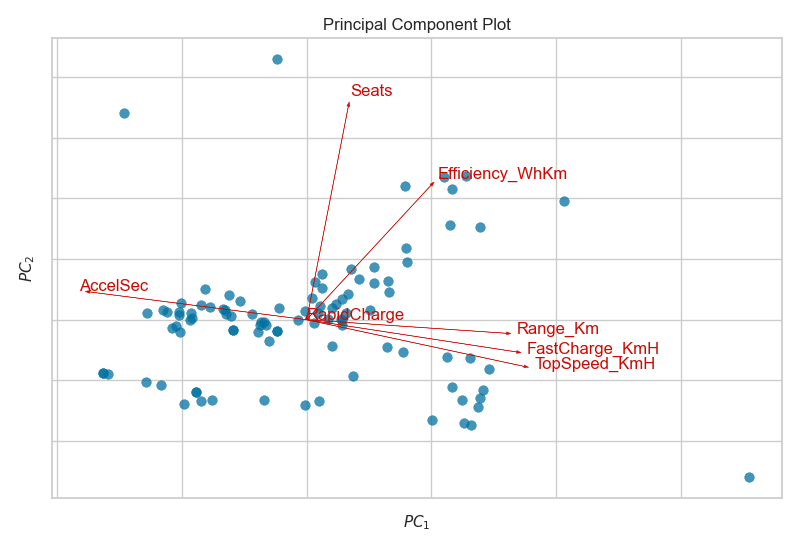


The loadings indicate how the original variables are combined to form principal components. Loadings guide the interpretation of principal components. In our example, the two segmentation variables with the highest loadings (in absolute terms) for principal component 1 are *Fast Charge* and *Rapid Charge* indicating that this principal component captures the charging and battery dimension. Similarly for principal component 2, highest loadings are *Range* and *Efficiency*, indicating that this principal component captures the performance dimension.

***Factor-Cluster Analysis***

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. Here, we calculate principal components because we use the resulting components to rotate and project the data for the perceptual map. We use unstandardised data because our segmentation variables are all binary.





The output shows the resulting perceptual map. The attributes *Range*, *Fast Charge*, *Top Speed* align in the same direction while *Acceleration* aligns opposite to it, indicating that vehicles with great *Range* and *Fast* *Charge* also tend to provide appreciable *Top* *Speed*. In the upper part of the map, we have *Seats* and *Efficiency* pointing away from each other indicating that market considers *Efficiency* and *Seats* as non-relatable attributes. Majority of the vehicles tend to have *Rapid Charge* as it is neutral.

The observations cluster at five positions:

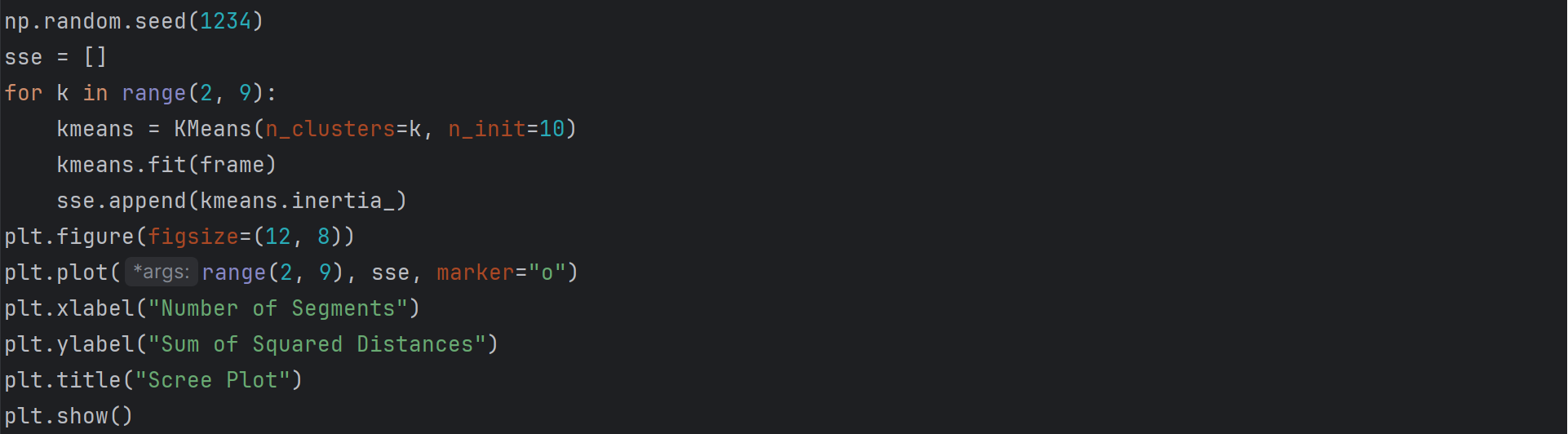
1. A group of vehicles at the top around the arrow pointing to Efficiency.
2. A group of vehicles at the bottom that don’t possess any considerable attributes.
3. A group of vehicles in the middle (rising upwards) with Seats attributes.
4. A group of vehicles at the left having Acceleration.
5. A group of vehicles at the right pointing towards Range, Fast Charge, Top Speed.

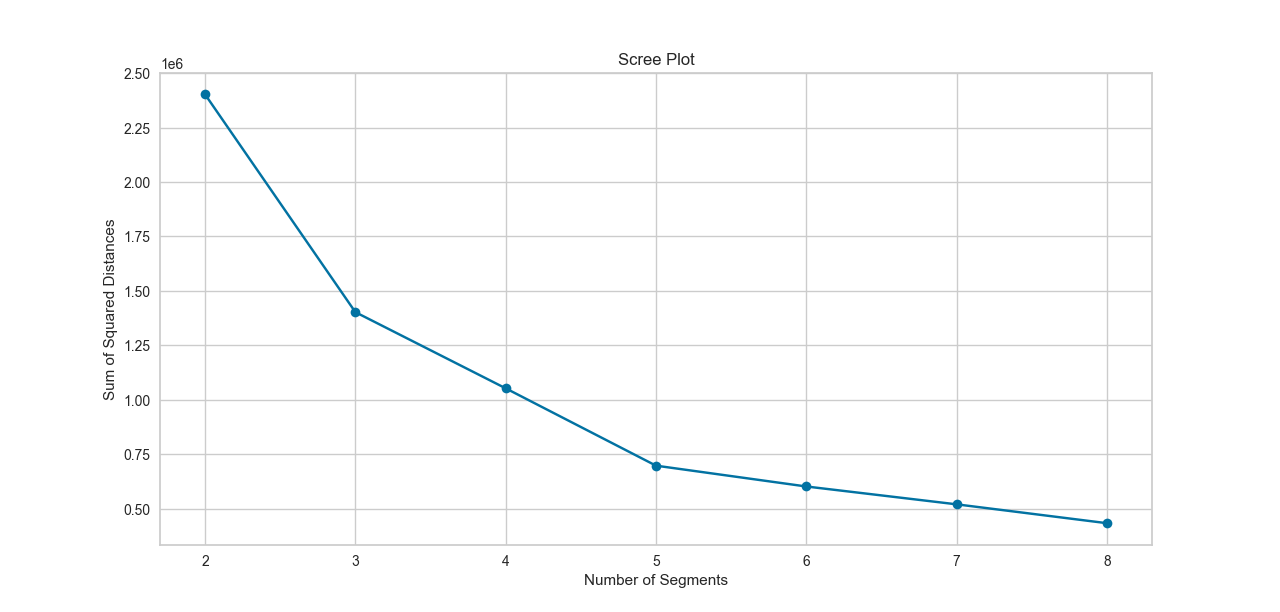
These initial exploratory insights represent valuable information for segment extraction. Results indicate that some attributes are strongly related to one another, and that the price dimension may be critical in differentiating between groups of consumers.

**Extracting Segments Using k-Means**

***Scree Plot***

We calculate solutions for two to eight market segments using standard *k*-means analysis with ten random restarts (*n\_init*). We then relabel segment numbers such that they are consistent across segmentations. We extract between two and eight segments because we do not know in advance what the best number of market segments is. If we calculate a range of solutions, we can compare them and choose the one which extracts segments containing similar consumers which are distinctly different from members of other segments. We compare different solutions using a **scree plot**.





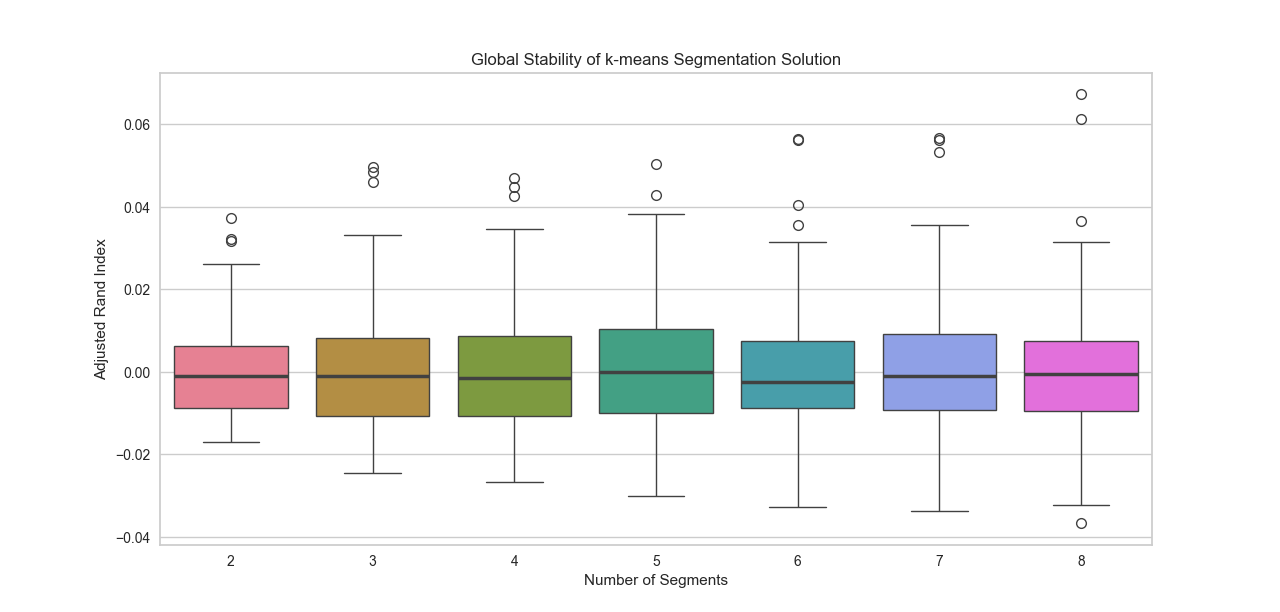
The above scree plot has an elbow at 5corresponding to *Number of Segments*. This is the point where the curve starts to flatten out. The *Sum of Squared Distances* within market segments drops slowly as the number of market segments increases. We expect the values to decrease because more market segments automatically mean that the segments are smaller and, as a consequence, that segment members are more similar to one another. *Number of Segments* beyond this point does not significantly increase the explained variance.

***Stability Based Data Structure Analysis***

Another approach to determine a good number of segments is to use stability-based data structure analysis. **Stability-based data structure analysis** also indicates whether market segments occur naturally in the data, or if they have to be artificially constructed. Stability-based data structure analysis uses stability across replications as criterion to offer this guidance. Such a solution would give management a little confidence in terms of investing substantial resources into a market segmentation strategy. Assessing the stability of segmentation solutions across repeated calculations ensures that unstable, random solutions are not used.

**Global stability** is the extent to which the same segmentation solution emerges if the analysis is repeated many times using *bootstrap samples* (randomly drawn subsets) of the data. Global stability is calculated using the following code, which conducts the analysis for each number of segments using 2 × 100 bootstrap samples (*nboot*) and ten random initialisations (*n\_init*) of *k*-means for each sample and number of segments.

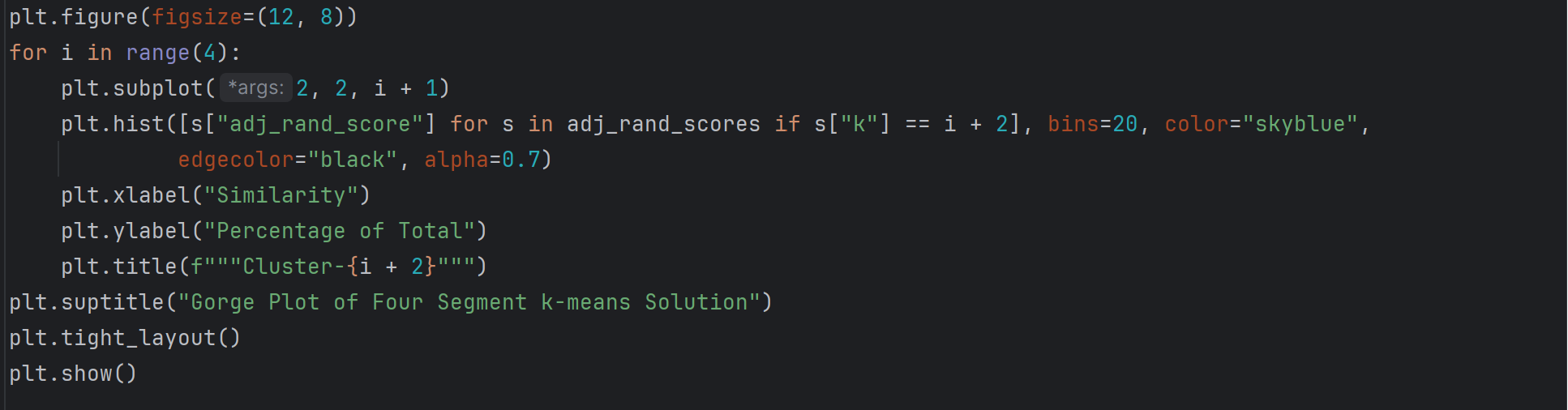


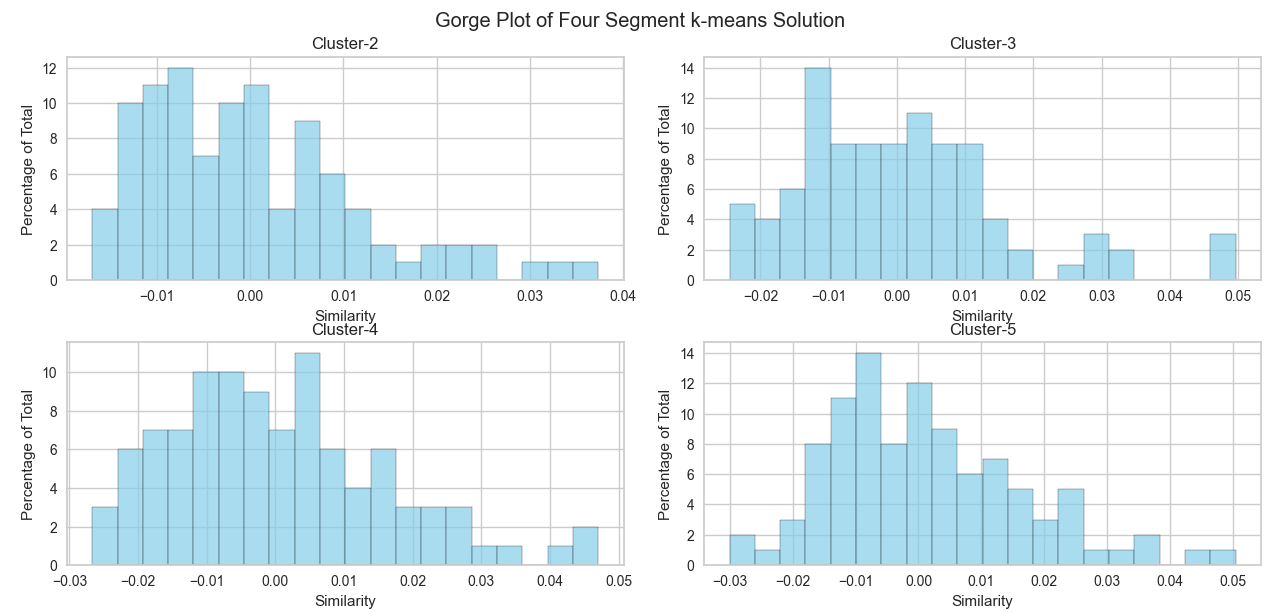


The vertical **boxplots** show the distribution of stability for each number of segments. The median is indicated by the thick black horizontal line in the middle of the box. Higher stability is better. Figure points that almost all the segment solutions appear to be quite stable. However, the two-, six- and eight- segment solutions do not offer a broad view of the market. Solutions containing a small number of segments typically lack the market insights managers are interested in. While three-, four- and seven- segment solutions offer similar market view. The five-segment solution thus emerges as the solution containing the most market segments.

***Gorge Plot***

We gain further insights into the structure of the four-segment solution with a gorge plot.



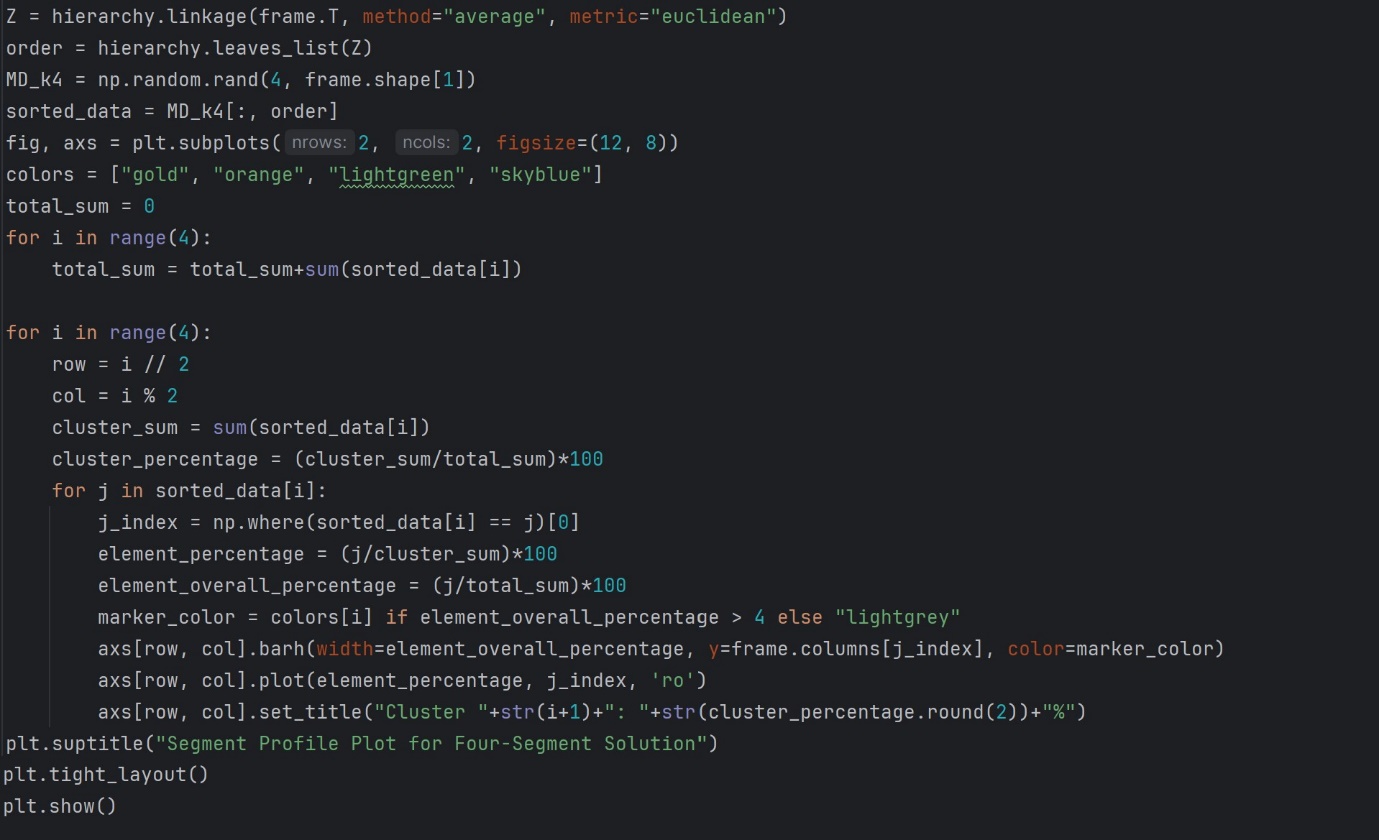


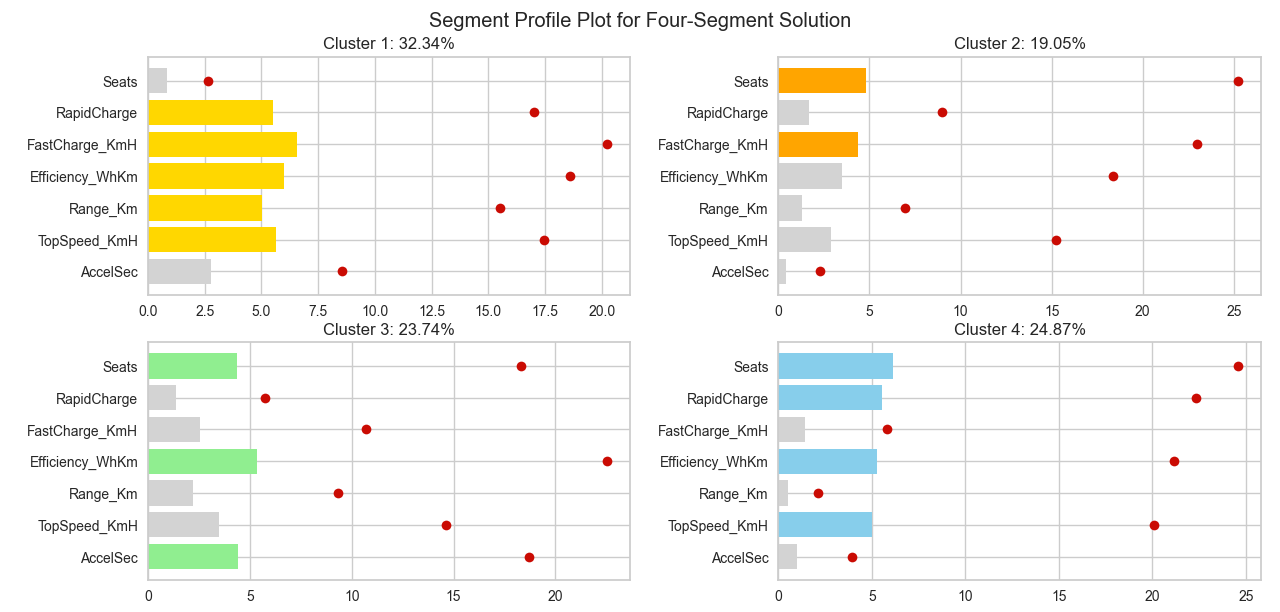
None of the segments shown in **histogram** is well separated from the other segments, and proximity to at least one other segment is present as indicated by the similarity values all being between -0.01 and 0.02.

**Profiling Segments**

***Hierarchical Cluster Analysis***

The core of the segmentation analysis is complete: market segments have been extracted. Now we need to understand what the five-segment *k*-means solution means. The first step in this direction is to create a **segment profile plot**. The segment profile plot makes it easy to see key characteristics of each market segment. It also highlights differences between segments. To ensure the plot is easy to interpret, similar attributes should be positioned close to one another. We achieve this by calculating a hierarchical cluster analysis. **Hierarchical cluster analysis** used on attributes (rather than consumers) identifies – attribute by attribute – the most similar ones.





The above figure is easy for companies to interpret. They can see that there are four market segments. They can also see the size of each market segment. The smallest segment (Cluster 2) contains 19% of vehicles, the largest (Cluster 1) 32%. The names of the segmentation variables (attributes) are written on the left side of the plot. The horizontal lines with the dot at the end indicate the percentage of vehicles in the entire sample that associate with segmentation variable. The bars plot the percentage of vehicles *within each segment* that associate with segmentation variable.

Marker variables are coloured differently for each segment. All other variables are greyed out. Marker variables differ from the overall sample percentage either by more than 4% points in absolute terms.

To understand the market segments, companies need to do two things:

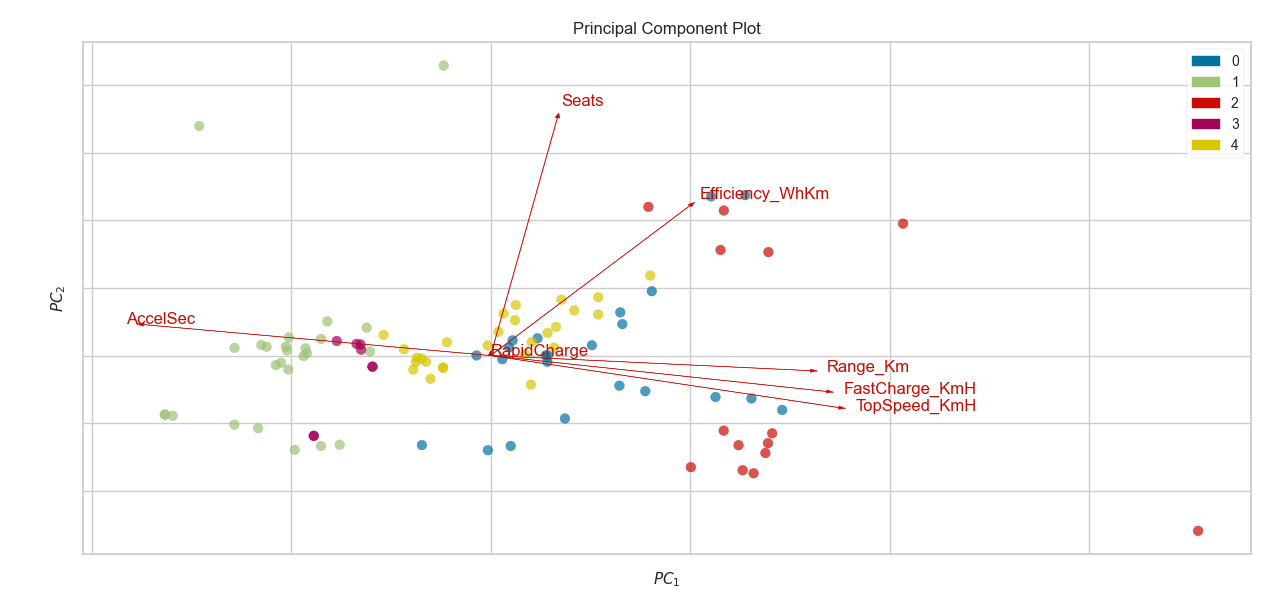
1. Compare the bars for each segment with the horizontal lines to see what makes each segment distinct from all vehicles in the market.
2. Compare bars across segments to identify differences between segments.

Segment 2 shows that vehicles have *Fast Charge* and *Seats*. Segment 3 shows that vehicles have *Seats*, *Efficiency* and *Acceleration* *Time*. While vehicles in Segment 4 have *Rapid* *Charge* and *Top* *Speed* in addition to *Seats*, *Efficiency*. Finally Segment 1 has a very distinct perception, setting apart vehicles of this segment from others. Apart from *Seats* and *Acceleration* *Time*, they have every other attributes.

***Segment Separation Plot***

Another visualisation that can help managers grasp the essence of market segments is the **segment separation plot**.





Observations are coloured to reflect segment membership. Legend at the upper – right corner indicates the segment and its corresponding colour in the plot. As can be seen, Segment 1 focus on *Acceleration* *Time* while Segment 4, apart from *Rapid* *Charge* they don’t seem to possess any other features. Segment 0 is divided into two groups. One moving towards *Efficiency* while the other towards *Range*, *Fast* *Charge* and *Speed*. Segment 2 has all the attributes that Segment 4 is trying to achieve. Segment 3 holds a less flattering view with small attributes of *Acceleration* *Time*.

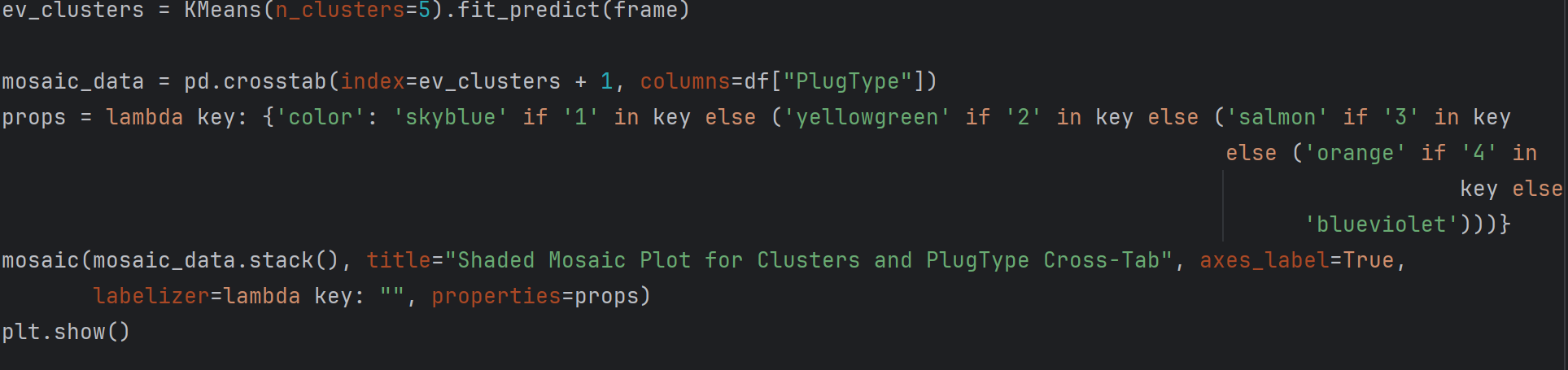
At the end of this step companies have a good understanding of the nature of the four market segments in view of the information that was used to create these segments.

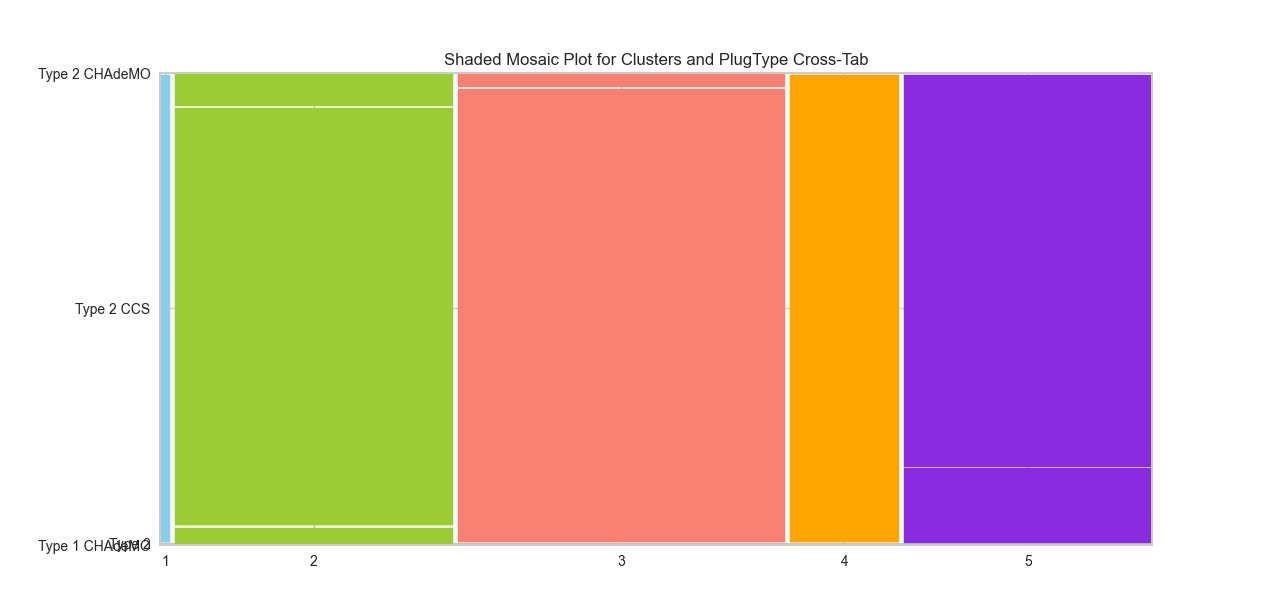
**Describing Segments**

Descriptor variables are additional pieces of information about consumers that are critically important to gaining a good understanding of market segments.

***Mosaic Plot***

One descriptor variable available in the vehicle data set is *Plug Type*. Using a simple mosaic plot, we can visualise the association between segment membership and *Plug Type*. To do this, we first extract the segment membership for each vehicle for the five-segment solution. Next, we cross-tabulate segment membership and the Plug\_Type variable. Finally, we generate the **mosaic plot** with cells colours indicating the observed frequencies.

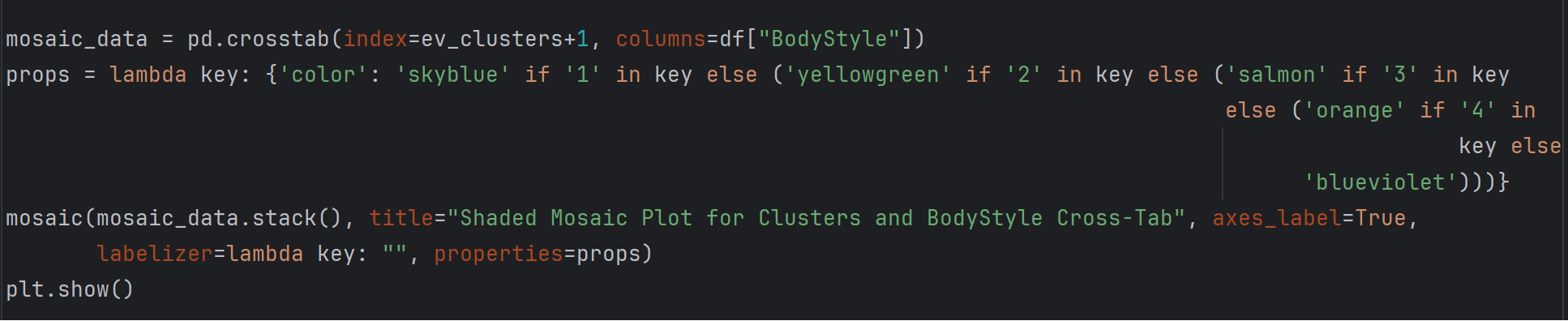


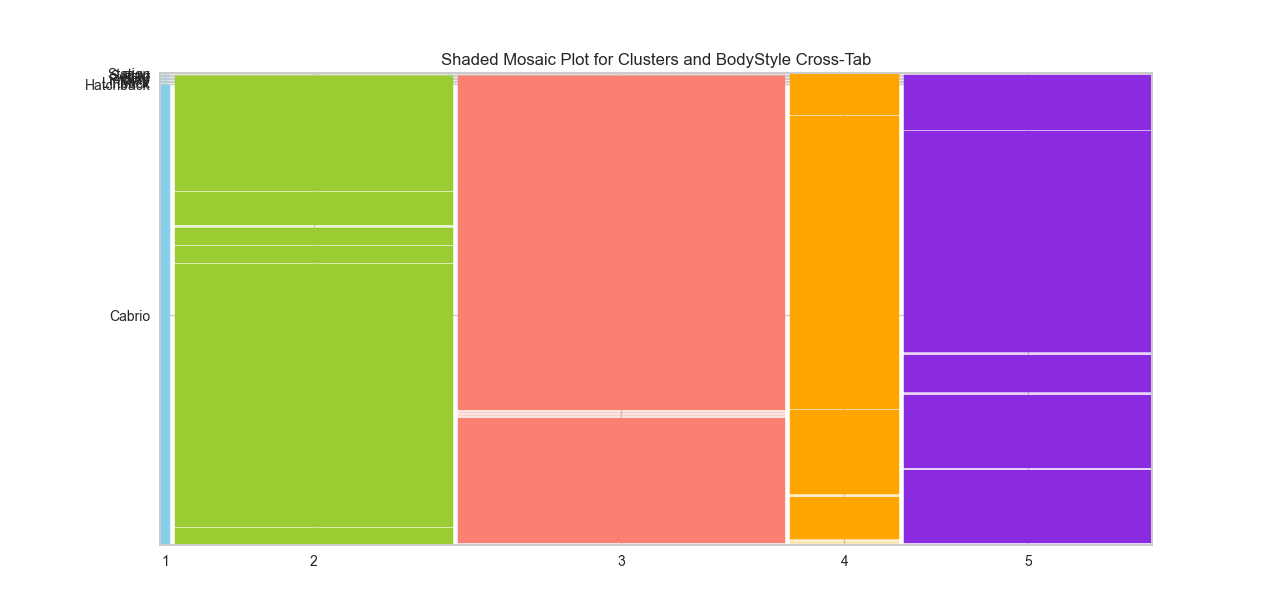


The mosaic plot, plots segment number along the *x*-axis, and *Plug Type* along the *y*-axis. The mosaic plot reveals a strong and significant association between those two variables. Based on the above plot, Segment 3 appears to be dominant. Segment 3 has 2 *Plug Type*: *Type2 CSS* and *Type2 CHAdeMO* with *Type2 CSS* being the largest shaded cell. Both these cells combined occupy greater area when compared to the rest.

Analysing the plot based on overall *Plug Type*, *Type2 CSS* is dominant of all *Plug Type*, followed by *Type2* (highest in Segment 5) and then *Type2 CHAdeMO* (highest in Segment 2).

The vehicle data contains a few other basic descriptor variables, such as *Body Style*, *Vehicle Segment* and *Price*. We cross-tabulate segment membership and the Body\_Style variable to generate a mosaic plot with cells colours indicating the observed frequencies.

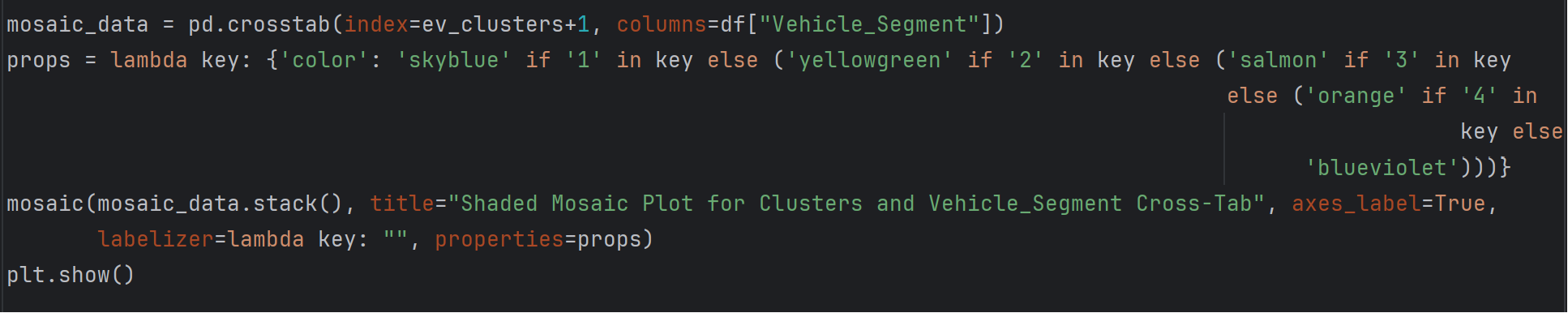


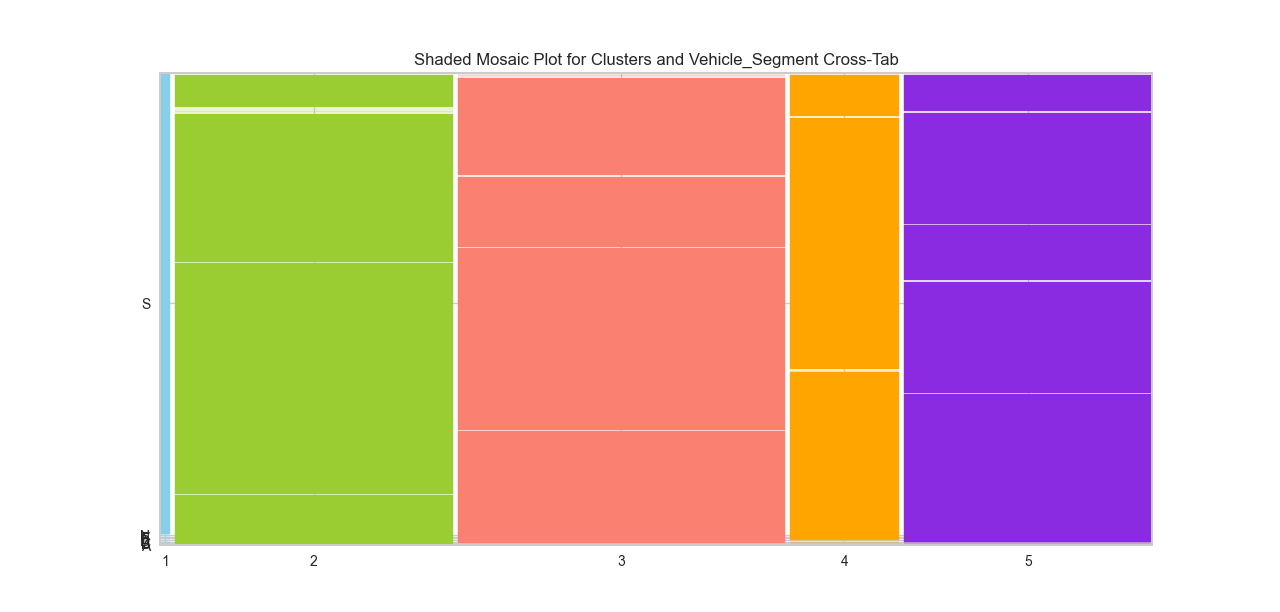


Based on the above plot, Segment 3 appears to be dominant. Segment 3 has 2 *Body* *Style*: *Hatchback* and *SUV* with *SUV* being the largest shaded cell. Both these cells combined occupy greater area when compared to the rest.

Analysing the plot based on overall *Body* *Style*, *SUV* is dominant of all *Body* *Style*, followed by *Hatchback* (highest in Segment 2) and then *Sedan* (highest in Segment 4).

We also generate a mosaic plot for Vehicle Segment by cross tabulating Vehicle\_Segment variable and segments.



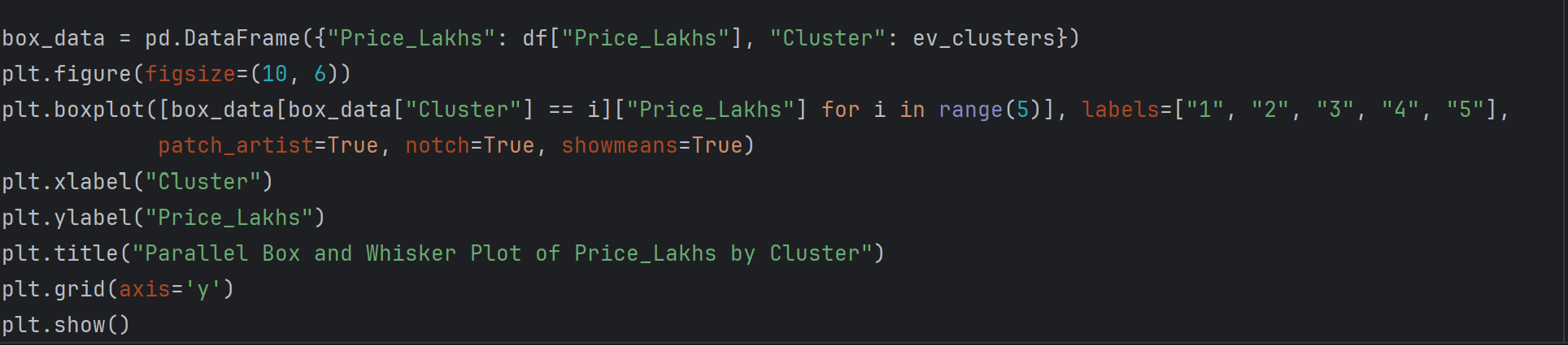


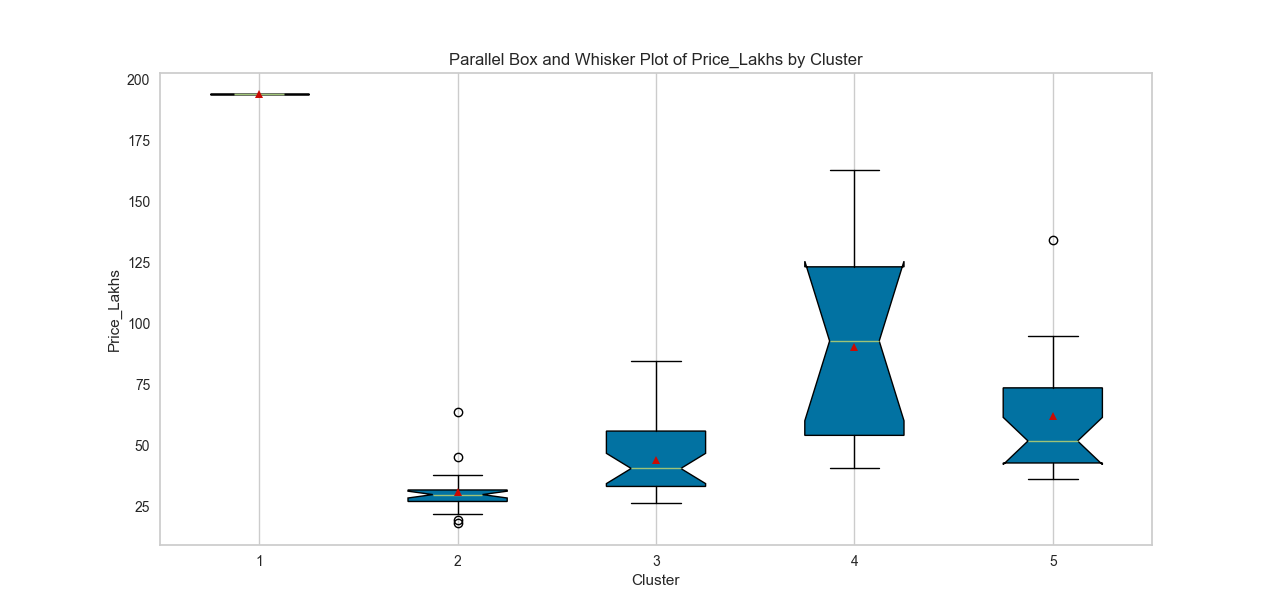
Based on the above plot, Segment 3 appears to be dominant. Segment 3 has 4 *Vehicle Segment*: *B, C, D, E* with *C* being the largest shaded cell. These cells combined occupy greater area when compared to the rest.

Analysing the plot based on overall *Vehicle Segment*, *C* is dominant of all *Vehicle Segment*, followed by *B* (highest in Segment 2) and then *D* (highest in Segment 5).

***Parallel Box and Whisker Plot***

*Price* is metric rather than categorical; hence we use a **parallel box and whisker plot** to assess the association of *Price* with segment membership. A parallel box and whisker plot allows us to compare the distribution of prices across different vehicle segments. Each segment will have its own box and whisker, making it easy to see differences in median, quartiles, and variability between segments.



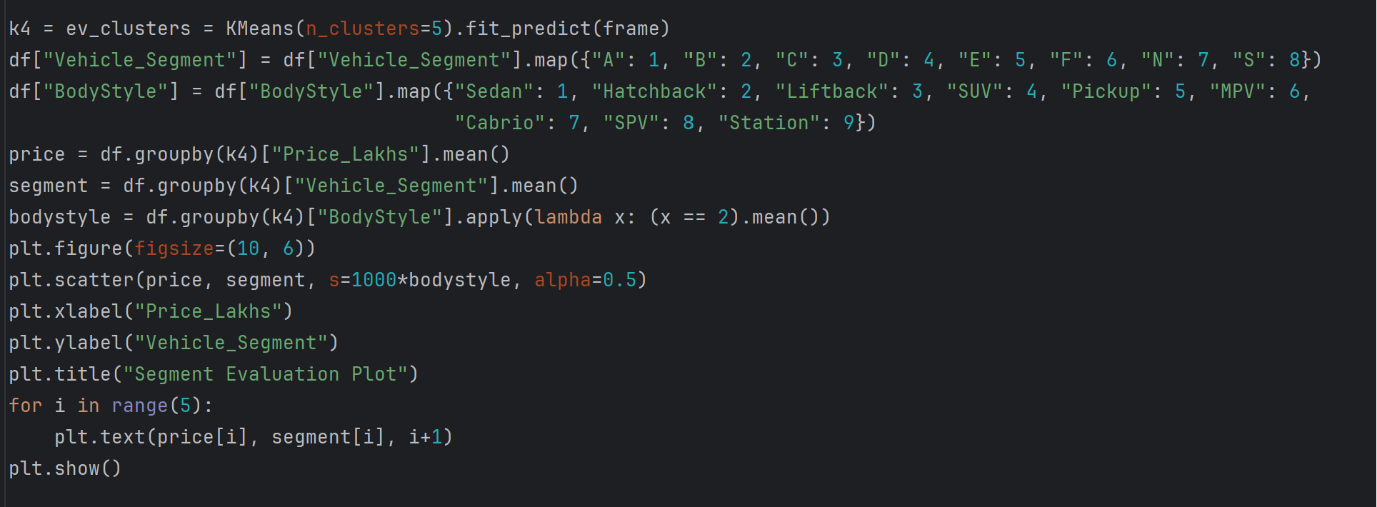


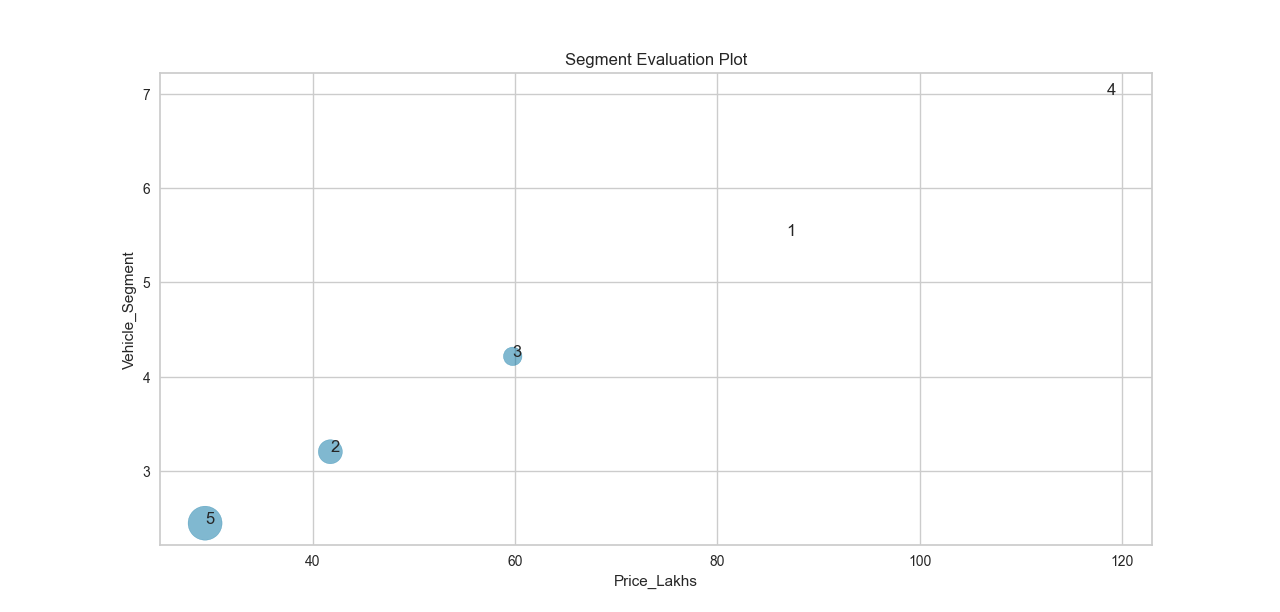
We plot Segments along the *x*-axis, and Price along the *y*-axis. By visually comparing the box and whisker plots for each segment, we can quickly identify that Segment 1 has the highest price range and Segment 2 has the lowest. We also see that Segment 1 and Segment 2 have the least variability in prices. Segment 4 offers wide range prices offering may varieties. This plot can provide insights for decision-making processes, such as pricing strategies or market positioning, by showing how prices are distributed across different segments and where adjustments may be needed.

**Selecting the Target Segment(s)**

Using the knock-out criteria and segment attractiveness criteria companies performing market segmentation can now proceed to develop a **segment evaluation plot**. The purpose of a segment evaluation plot is to help analysts and decision-makers understand how different segments are defined based on the available descriptors.

The *Price* of vehicles is plotted along the *x*-axis. The *Vehicle* *Segment* is plotted along the *y*-axis. The bubble size represents the *Body* *Style*.





Market segments 5 and 2 are located in the attractive quadrant of the segment evaluation plot. Members of these two segments offer affordable price and Segment C. These segments need to be retained, and their needs must be satisfied in the future. Market segment 4 is located at the far corner of the plot whose price is set very high and is Segment N. These vehicles are luxurious and exclusive. They have a different consumer base that differ from the usual commercial domain. Market segment 1 does not seem promising as it is Segment E or Segment F and expensive not in a positive way. Market segment 5 seems viable. Marketing action could attempt to address the negative factors of this segment, and re-in force positive factors. As a result, company may be able to broaden its customer base.

The segment evaluation plot serves as a useful decision support tool for company’s management to discuss which of the market segments should be targeted and, as such, become the focus of attention.

**Customising the Market Mix**

1. **Retain Segments 5 and 2:** These segments are located in the attractive quadrant of the segment evaluation plot, offering affordable prices and Segment C. The company should continue to meet the needs of these segments and ensure their satisfaction in the future. This could involve maintaining competitive pricing and offering products that align with Segment C preferences.
2. **Address Negative Factors of Segment 5:** While Segment 4 seems viable, there may be negative factors that need to be addressed. Marketing actions should focus on addressing these negative factors while reinforcing the positive factors of the segment. This could involve product improvements, targeted marketing campaigns, or pricing adjustments to make the segment more appealing.
3. **Reconsider Market Segment 1:** Market Segment 0 does not seem promising as it consists of Segment E or Segment F, which are expensive in a negative way. The company should reconsider its approach to this segment. This could involve repositioning the products to offer better value for the price or targeting a different consumer base that values luxury and exclusivity.
4. **Understand and Target Market Segment 4:** Market Segment 3 consists of luxurious and exclusive vehicles, appealing to a different consumer base. The company should understand the needs and preferences of this segment and tailor its marketing and product offerings accordingly. This could involve focusing on luxury and exclusivity in marketing campaigns and product features.
5. **Broaden Customer Base:** By addressing the needs and preferences of different market segments, the company can broaden its customer base. This could involve offering a diverse range of products that cater to different segments, implementing targeted marketing strategies, and ensuring competitive pricing and value proposition.

**Conclusion**

In conclusion, market segmentation analysis is essential for understanding the diverse and complex electric vehicle market in India. By identifying and targeting specific segments, companies can develop products and services that meet the needs of consumers, optimize their marketing efforts, and gain a competitive advantage in the market. As the electric vehicle market continues to grow and evolve, market segmentation analysis will play a crucial role in shaping the future of electric mobility in India.