

Electric Vehicles for India

A Market Segmentation Analysis

Arvind Haridas



Problem Statement

The task is to analyse customer purchases in the automobile industry, in order to obtain insights and develop a market strategy for electric vehicle development. This report will study a database of known passenger vehicle purchases, along with the characteristics of the purchaser.

Fermi Estimation

These are the sales figures for passenger vehicles in India for the past four financial years.

Previous Sales			
2020-21	2021-22	2022-23	2023-24
30,62,280	36,50,698	45,87,116	49,01,844

From this, we can estimate that there is roughly a 6 lakh increase year-on-year for passenger vehicle sales.

In European markets, approximately 1 in 5 vehicles sold are EVs. Given the rapid progress in EV technologies and the associated reduction in barriers for purchase, it stands to reason that Indian manufacturers can aim to reach those numbers as well. EV 4 wheeler sales are currently low in India at approximately 1 lakh per annum.

This gives us a target of supplying 9 lakh EVs per annum at the very least. This can also increase by 1.2 lakh per year if the projected sales growth is correct. In addition, climate change will eventually necessitate that non-EV vehicles need to be replaced with EV variants, with passenger vehicles being prime targets for legislation.

Data Collection

The data used here was obtained from Feynn Labs. The data descriptions are:

1. "Age" tells us the age of the person purchasing the vehicle.
2. "Profession" tells us whether the purchaser is a business owner or a salaried worker.
3. "Marital Status" tells us whether the purchaser is married.
4. "Education" tells us whether the person has completed their graduate or postgraduate degree.
5. "No of Dependents" tells us the number of people depending on the purchaser and their income.
6. "Personal Loan" and "House Loan" tell us whether the purchaser has those respective loans.
7. "Wife Working" tells us whether the purchasers spouse (if they have one) is also employed.
8. "Salary", "Wife Salary" and "Total Salary" tells us the wages of the purchaser, their spouse (if employed) and the sum of the two wages respectively.
9. "Make" tells us the model of the purchased vehicle.
10. "Price" tells us the cost of the purchased vehicle.

Github link

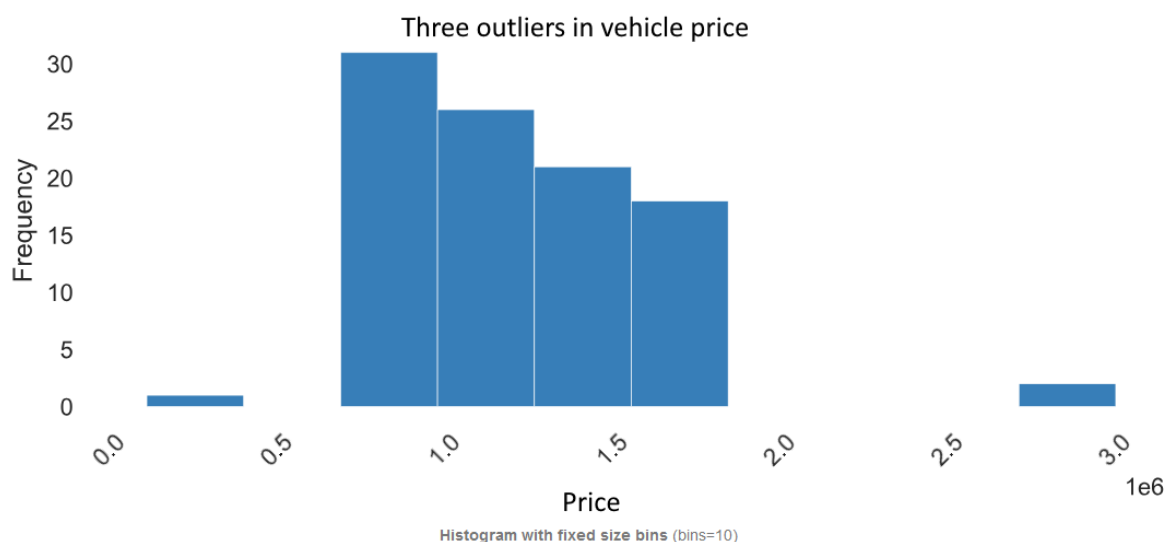
<https://github.com/Drakus98/Feynn-Labs-EV-Market-Segmentation>

Data Pre-processing

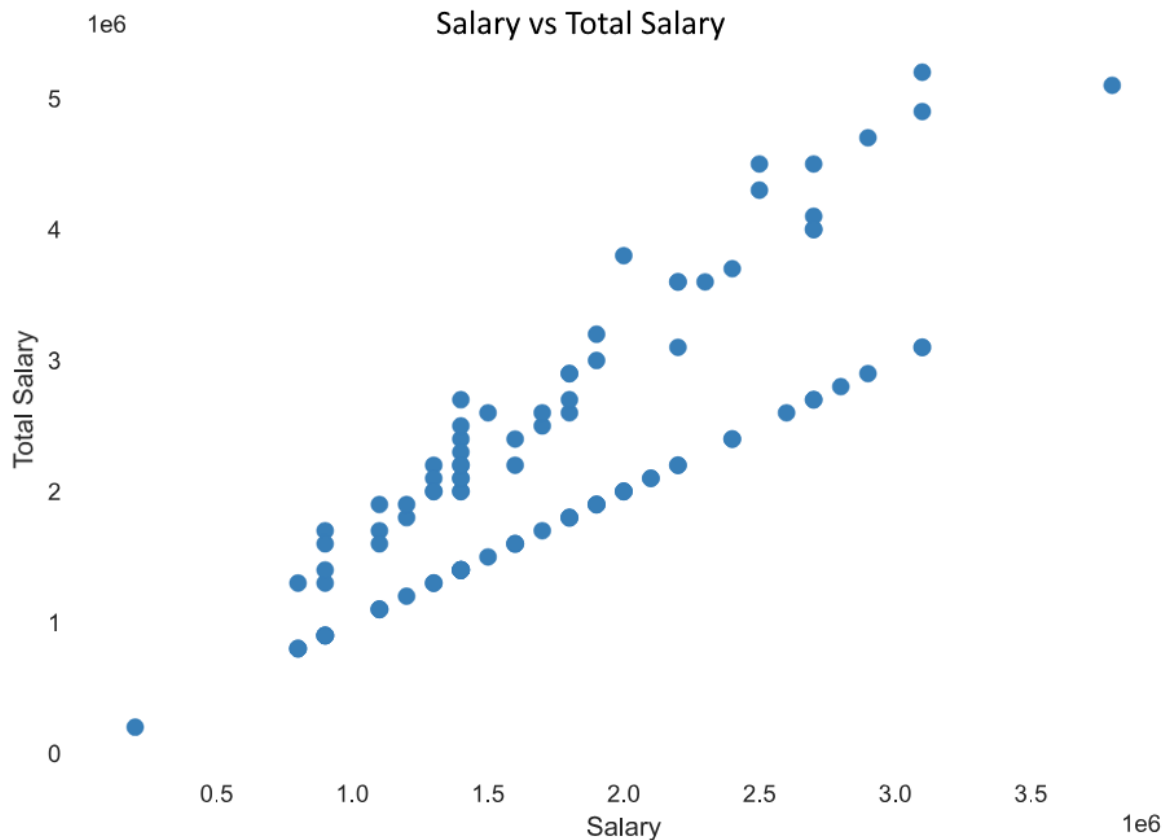
1. Yes/No was re-encoded as 1s and 0s.
2. "Make" was dropped, as it was almost entirely correlated with "Price".
3. "Salary" and "Wife Salary" were dropped, as both of those together combine into "Total Salary". There is naturally also significant correlation between the spouse having employment and having a salary at all.
4. Three records were dropped due to being extreme outliers, namely indices 12, 15 and 54.
5. "Total Salary" and "Price" were scaled using MinMaxScaler.

Exploratory Data Analysis

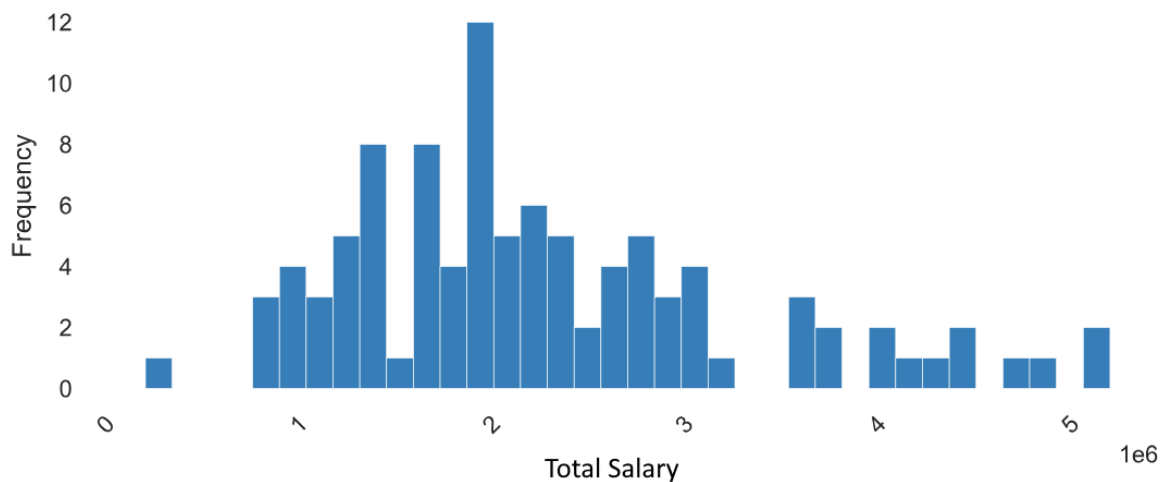
Due to the database being extremely sparse (99 records in total), EDA was primarily conducted through ProfileReports from pandas. This gives us an interactive dashboard for analysing the data as an html document. The highlights of the reports for EDA are detailed below.



Comparing the frequency of vehicle sales with the price of said vehicles gives us a snapshot of the overall market. In the above graph, we can see that there are three significant outliers in the dataset. Given the impact of outliers on segmentation and the unlikelihood of them being a good marketing target, the three outliers were removed.



Comparing “Total Salary” with the “Salary” shows significant issues with using this data for segmentation. Two correlation trends are clearly visible here. The bottom one represents those without spouses or with unemployed spouses, where the “Total Salary” is exactly equal to “Salary”. The one on top arises from “Total Salary” being a sum of the other two salary values.



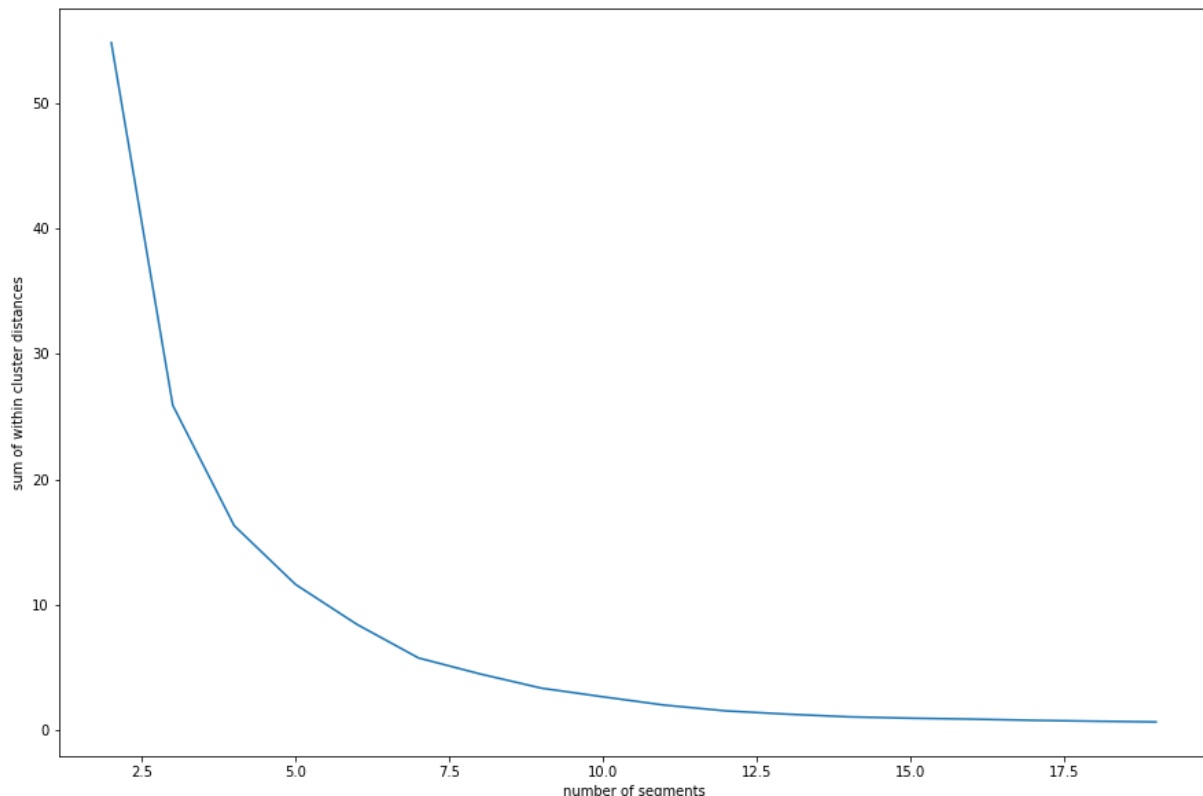
Since we are studying potential buyers, their income is of paramount importance. The above graph shows us the “Total Salary” distribution.

For the segmentation, the EDA narrowed down three ideal segmentation variables: Total Salary, Price and No. of Dependents. These represent the variables that are most likely to be critical to the purchaser’s decision-making process when selecting their ideal vehicle. The remaining variables are to be used to describe the segments later on.

Segment Extraction

The segmentation was conducted using K-means clustering, an unsupervised learning algorithm that groups the various data points into clusters. It is an iterative process that divides the dataset into N number of clusters, where this number can be dictated by us and altered to obtain the ideal segments. It functions by associating each segment with a centroid, such that the distance between each cluster and its constituent data points are minimized. In the process, each data point is assigned to one and only one cluster, so we end up with N completely distinct clusters.

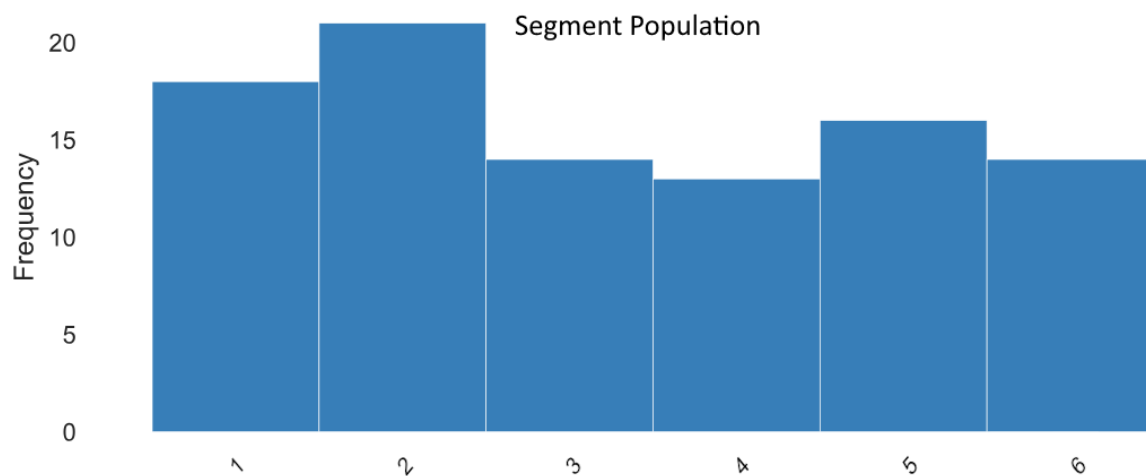
Identifying the ideal number of clusters is one of the main challenges. The first and simplest technique used is the “Elbow” method. Since K-means works by minimizing distances within each cluster, increasing the number of clusters results naturally results in an overall decrease in the sum of such distances from all clusters. This decrease is typically most significant when you have very few segments, with the magnitude of the decrease going down as you increase your cluster count. The “Elbow” refers to the point at which the magnitude of this decrease reduces dramatically, i.e. increasing the number of clusters is no longer giving a significant improvement in the cluster quality.



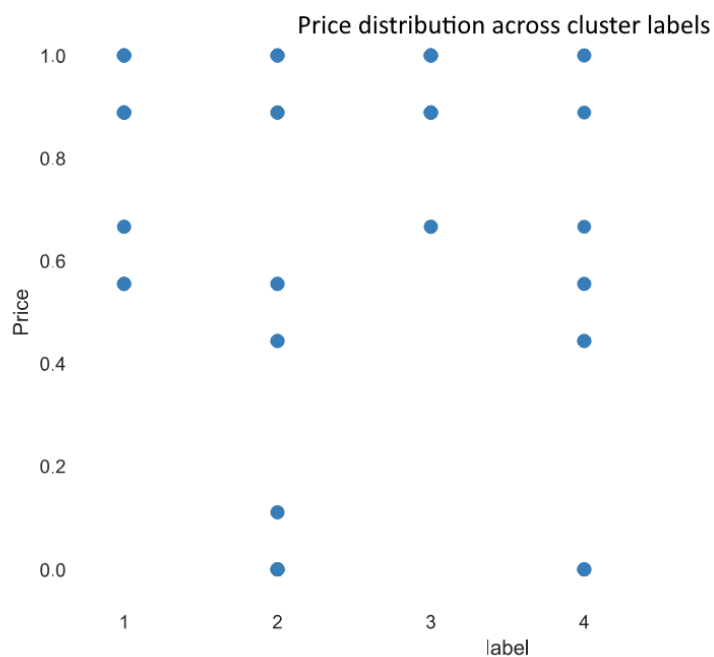
The graph above shows this elbow pattern. We can see that the curve starts flattening out when you approach 5-7 segments. We can confirm this using the silhouette coefficient, which is a metric for calculating the goodness of the clustering output. A higher value for the coefficient indicates better clusters. In this case, we have a peak of 0.616 in this coefficient at 6 segments, with the preceding peak being at 2 segments (too few) and the following peak being at 9 segments (too many). Therefore, we are conducting the segmentation with 6 segments.

Profiling Potential Segments

Given the nature of the data at hand, we are limited to studying demographic factors here. Certain trends arise when profiling and describing the segments.



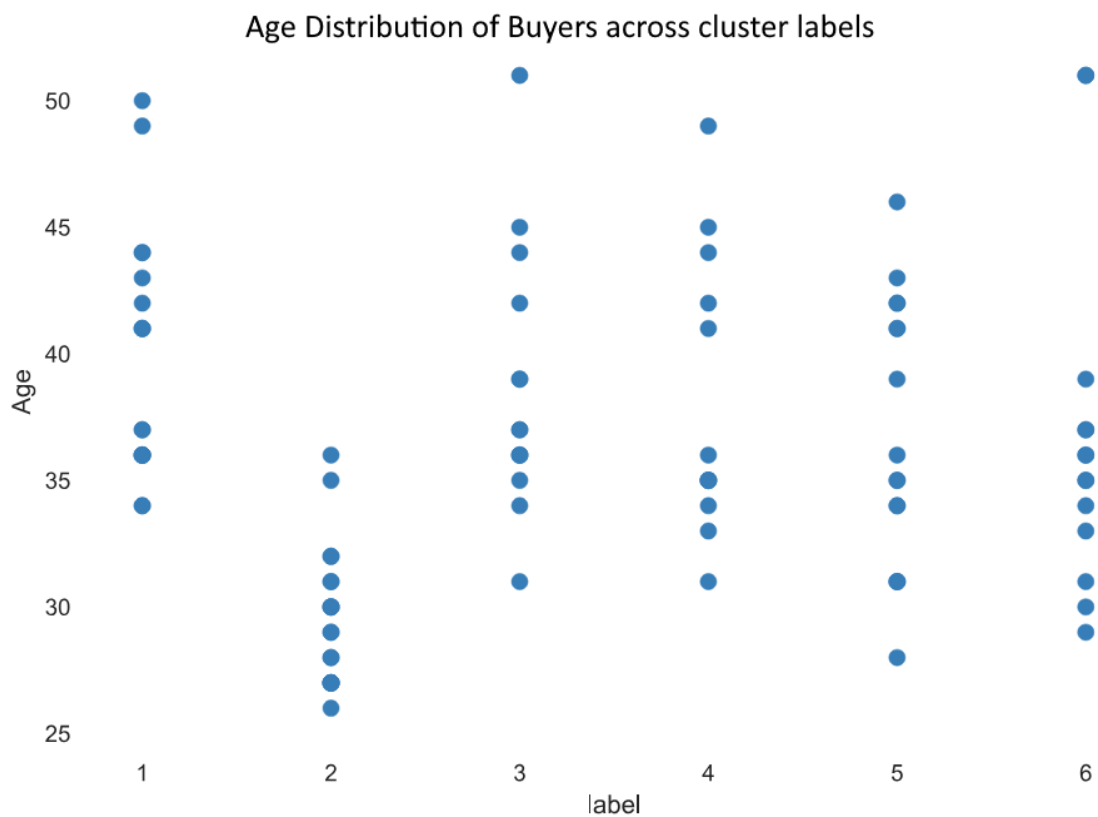
When looking at the size of each segment, segment 2 pops out more than the others. A larger segment size is of massive benefit, since the returns on the time and effort investment are potentially much larger if the rest of the segment's characteristics are desirable.



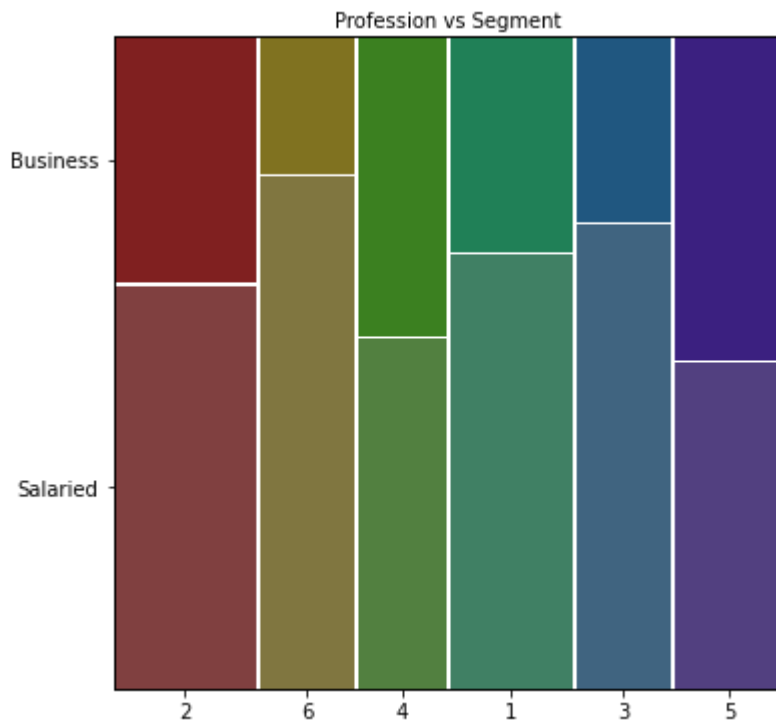
Looking at the price distribution of the purchases across our six segments, it is clear that there isn't any neat clustering of price points. Segments 1, 3, 5 and 6 cover half the total price range, slightly narrowing down the target price point for any EV model development. Segments 2 and 4 cover the entire range.



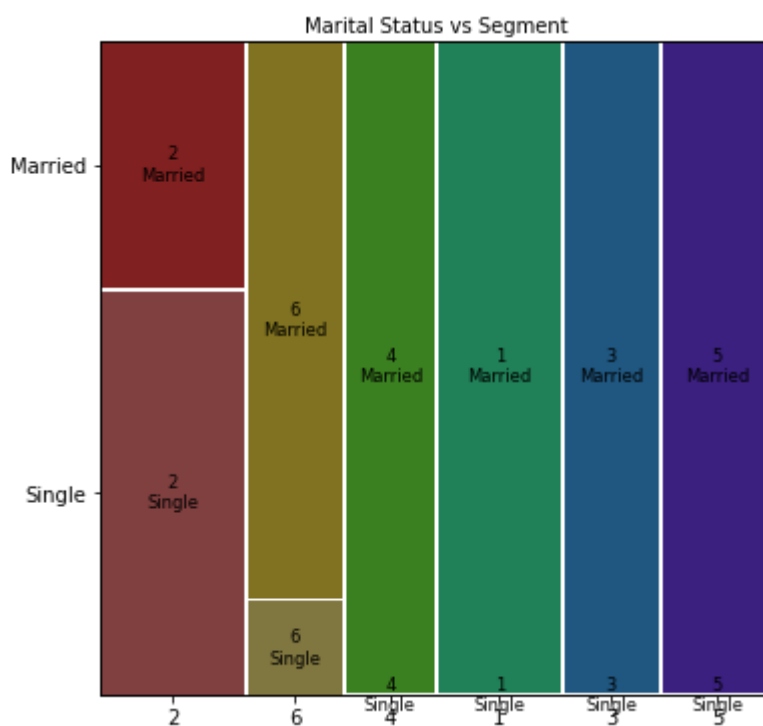
The distribution of salary is more marked. All six segments only cover approximately 30-60% of the total salary range. Segments 2, 5 and 6 are especially tightly clustered here, which could potentially help with marketing to them.



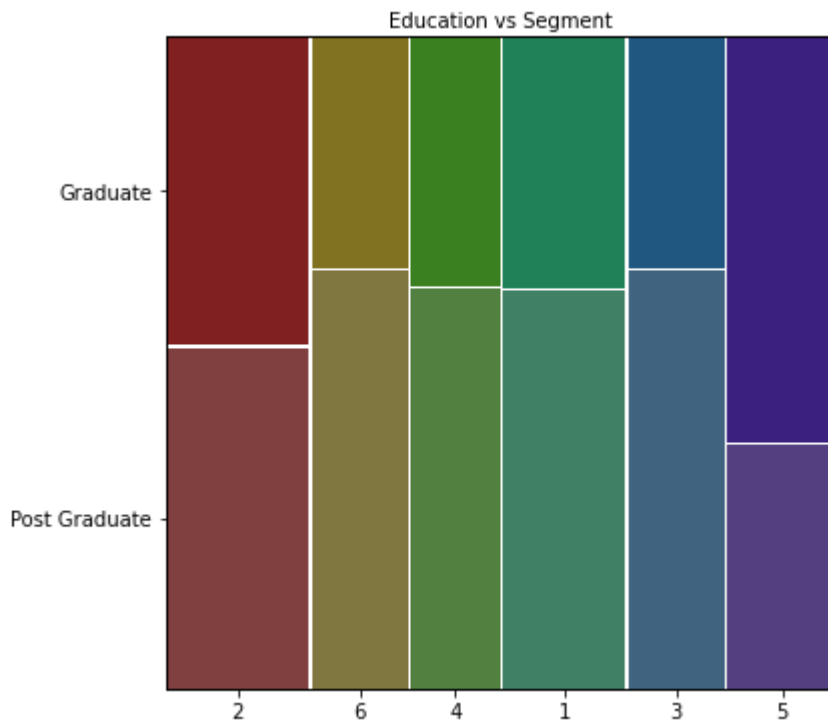
When looking at age, segments 2 and 6 are noteworthy, with the remainder showing a broad, even distribution across most of the overall age range. Segment 2 has the tightest clustering in age, with everyone below the age of 37. Segment 6 also shows similar clustering, except for one anomalous outlier.



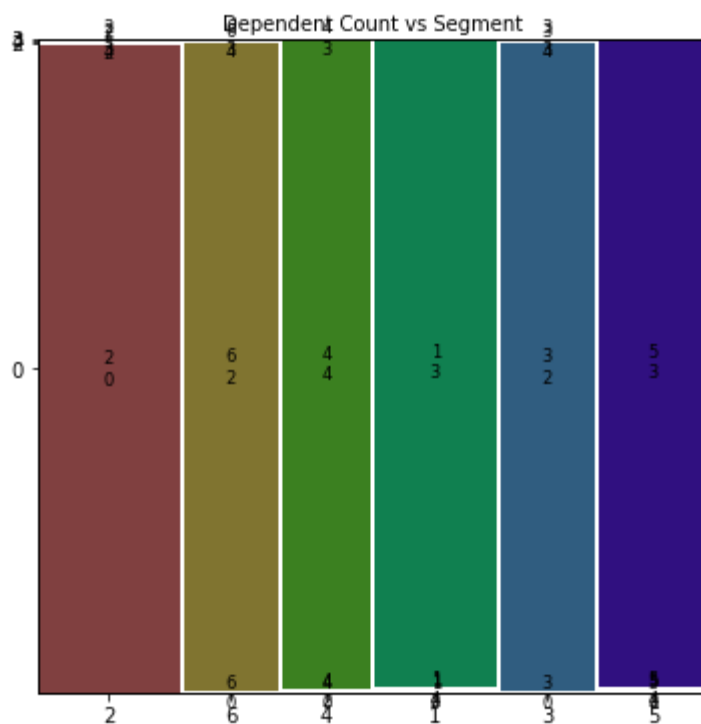
Studying the descriptive variables is best done via a mosaic plot, as shown above with profession distribution across the segments. Segment 6 stands out for having a large proportion of salaried individuals.



When it comes to marital status, Segments 1, 3, 4 and 5 are all filled with just married purchasers. Most single buyers are in Segment 2, with a much smaller number in Segment 6.



When looking at education, buyers who have completed only their graduate studies are most numerous in Segment 5 by far, followed by Segment 2. Segments 1, 3, 4, and 6 are dominated by postgraduate buyers.



The dependent count is clustered very effectively here. Every single segment is dominated by a single dependent count value.

Target Segments

Based on the segments created, the ideal target segments are:

- **Segment 2:** Age and salary are tightly clustered, making targeted marketing more effective. The majority of single buyers are in this segment too, so significant business opportunities would be lost if this segment isn't targeted. All buyers without dependents are in this segment.
- **Segment 5:** Salary is tightly clustered in this segment as well, but age covers a wide range here. There is a clear preference for cheaper vehicles. This group consists wholly of married buyers with multiple dependents. There is also a bias in favour of business owners with graduate qualifications. All this combines to cover the bases that Segment 2 doesn't cover.

Customizing the Marketing Mix

The conclusion of the market segmentation process is to identify the ideal target segments and develop a marketing mix that are tailored to those segments. This is done using the 4Ps, namely Price, Product, Place and Promotion.

Segment 2

- **Price:** This segment has no clear preference as far as price is considered. The organization may therefore choose a target price of their choosing based on their operational needs, desires and constraints.
- **Product:** There are a significant number of single buyers in this segment, and they tend to be younger as well. An EV targeting this segment should therefore cater to the lifestyle of an unmarried individual. There should be an emphasis on style and performance. The lack of dependents means that amenities catering to that aspect, such as numerous seating or luggage space, need not be a priority.
- **Place:** Given that we are dealing with demographic data, there aren't tight place constraints on the marketing. The prevalence of single, independent, young buyers in this segment suggests that places where they are found in larger numbers, such as technoparks or new urban developments, are ideal locations for selling the product.
- **Promotion:** The ideal way to promote the product to this segment is to utilize media that younger buyers frequent. A social media campaign would work well to that end. Another alternative is to use product placement in effective locations, such as movies or TV shows that also target this segment's demographics.

Segment 5

- **Price:** This segment prefers somewhat cheaper vehicles, so any product targeting this segment needs to be priced competitively,
- **Product:** Any EV targeting this segment needs to benefit married people with multiple dependents. This means that it needs to be larger, convenient to use and possess amenities that cater to a family unit rather than an individual. There are a noticeable number of business owners, so the product must possess the appearance and style to match an executive environment. A road-going SUV would be ideal here.
- **Place:** The dual presence of business and family related features suggests that the ideal place for selling EVs targeting this segment are business centres within major cities or the more developed suburbs where established families reside.

- **Promotion:** Family-oriented potential owners can be targeted through television advertisements or product placements in family-oriented shows. Business owners can also be targeted through professional networks, providing sales opportunities in bulk quantities.