

EV Market Segmentation
-Raima Joseph
-Feynn Labs Internship Project 2.2



Abstract

The following report is a detailed analysis of the principles demonstrated through the study task given previously. It contains a practical application of the 10 steps studied for the problem of Electric Vehicle Market Segmentation in India. Leveraging real-world data, the analysis encompasses various facets such as the state of Electric Vehicles (EV), type of charge, car and style of EV, and the behaviour of non-EV customers.

Background

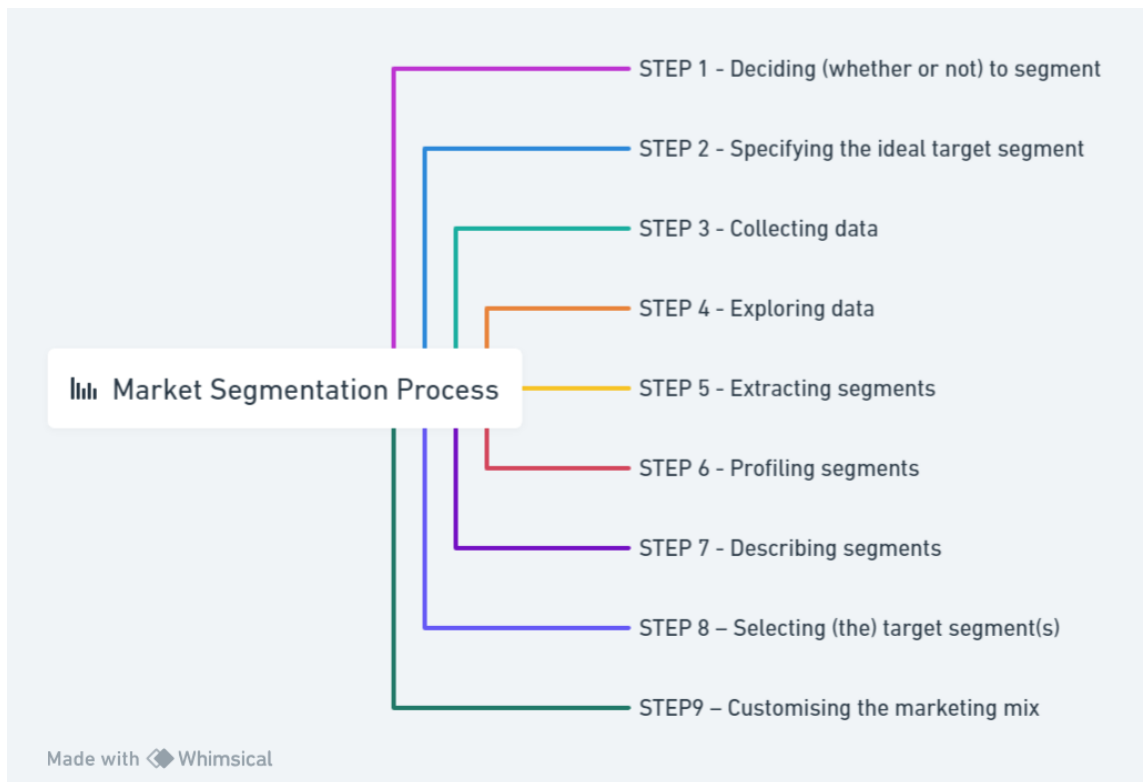
To match the needs of the consumers of a product or service, it is crucial to understand the audience that gets attracted by the product or service. This match is what makes the marketing planning process of an organisation. The simplest definition for the same is that market segmentation is the process of creating a decision-making tool for managers to divide a large audience into smaller targets for designing and deploying appropriate services. This study goes beyond theoretical exploration by delving into real-life implementation, where the principles of MS are applied to a practical case study. The intent is to bridge the gap between theory and application, providing a comprehensive guide for professionals, researchers, and decision-makers seeking to harness the potential of MS in their organizational contexts.

When it comes to the Electric Vehicle Market in India, it is but a newfound cultivation. The Innovation Adoption Life Cycle is like a roadmap, guiding us through the stages of how people and businesses in India are adopting and adapting to EV technology. As the popularity of EVs in India grows, businesses need to strategically price and analyse EVs and understand the early adopters' behaviour.

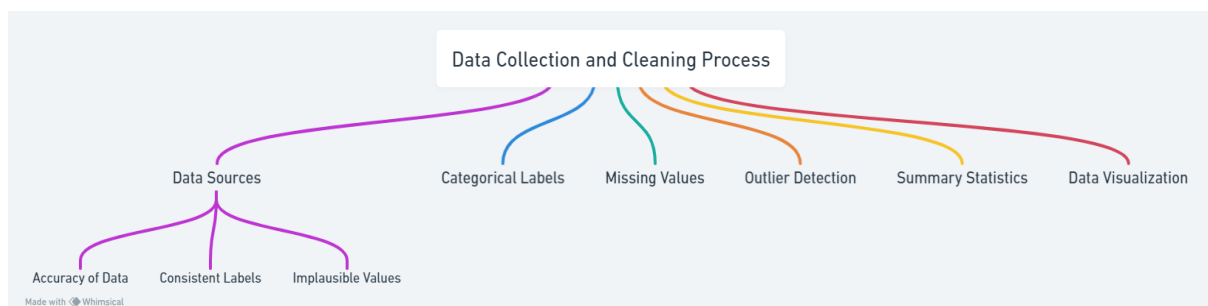
Based on the available datasets, this report tries to analyse various segmentations of Electric as well as Non-Electric vehicles in India based on geographical, popular and behavioural data.



Summarising the study task conducted, the market segmentation analysis involves several crucial steps to transform raw data into meaningful insights, especially when applied to the Electric Vehicle (EV) market in India. This pictorial representation explains the steps well.



Key preprocessing steps include merging levels and converting categorical variables into numerical values, often employing binary classification. Standardizing variables ensures consistency, with Python tools like StandardScaler simplifying this process.



Principal Component Analysis (PCA) further helps in dimensionality reduction, facilitating clearer visualization of complex multivariate datasets. These preprocessing steps collectively prepare the data for subsequent machine-learning models.

For extracting market segments the study task proved K-Means clustering to be most useful.. The 'elbow method' and a combination of k-means help determine the optimal number of clusters which leads the way for analysing the segments.

Profiling involves identifying essential features within each market segment based on segmentation variables. Visualizations, such as charts and graphs, play a key role in analysing segmentation breakdowns.

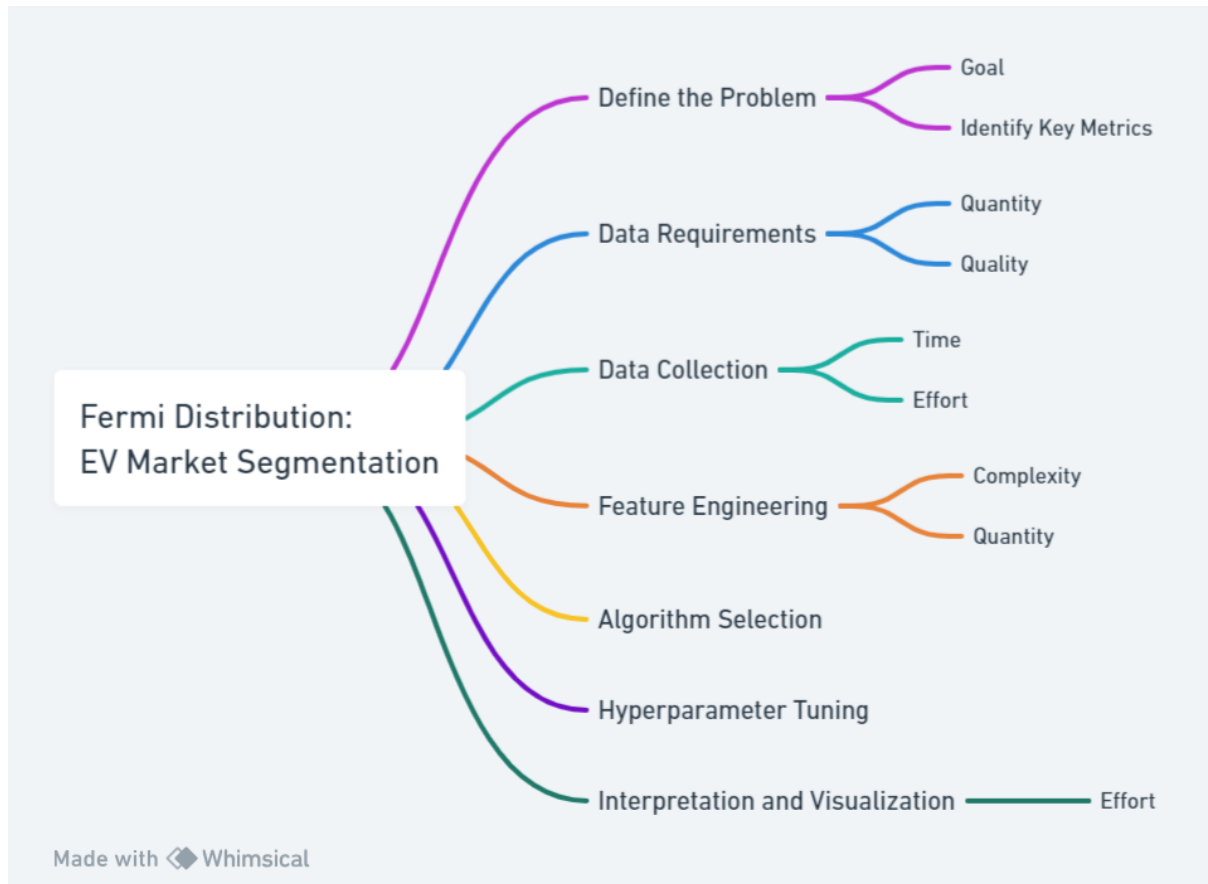
A Summary of the Case Study

In the case study conducted on the Fast Food (McDonald's) Dataset, steps 1-3 covered data import, preprocessing, and exploratory data analysis (EDA). Barplots, Histograms, Correlation heat maps, and a Biplot were utilized for feature relevance assessment, along with PCA to identify key features. Aligning with the report's methodology steps 4-8 included cluster extraction using the elbow method, cluster visualization, segment profiling using bar and violin plots, and the strategic selection of target segments. This comprehensive approach mirrors real-world market segmentation scenarios and highlights the practical application of the discussed methodologies.

Similar to the case study conducted, this report is an implementation of the same steps on EV-related datasets for the purpose of EV Market Segmentation in India.

Fermi Estimation Breakdown Based on Analysis Steps

Fermi estimation breakdown provides a structured approach to understanding the various steps required for analysing a market and its segmentation. Here is a pictorial representation of the same regarding the EV Market.



1) Define the Problem:

Goal: Segment the EV market geographical, behavioural and popularity data.

2) Identify Key Metrics:

Metrics: EV Cars in India, Type of charge, Body and Performance details of EV, Behavior of non-EV Customers, Statewise Registered EV in India.

3) Data Requirements:

Quantity: Based on the availability of data, I was able to gather and combine various datasets for the purpose of achieving the criteria of segmentation.

Quality: Assess the data quality, especially for features related to state, type of charge, car, style, and customer behaviour.

4) Data Collection:

Time: The overall time taken for collection of data and researching through datasets.

It took more than 3-4 days to collect viable data which was up-to-date.

Effort: Cleaning, preprocessing and combining datasets took ample effort.

5) Feature Engineering:

Complexity: Assess the complexity of feature engineering for the selected features. For the features which included metrics or addresses, it increased complexities.

Quantity: Estimate the number of features relevant to the EV Market; state, type of charge, car, style, behaviour, range, customer status and charging stations available.

6) Algorithm Selection and Model Training:

Time and Resources: Reflect on the time and computational resources needed for implementing K-means clustering and the elbow method.

Status achieved after completion:

Python 3 Google Compute Engine backend

Showing resources from 8:44 PM to 10:41 PM

System RAM

1.9 / 12.7 GB

Disk

26.7 / 107.7 GB

7) Hyperparameter Tuning:

Time: Estimate the time needed for implementing and interpreting the elbow method for determining the optimal number of clusters. For the purpose of selecting which features must be dropped and which must be used, tuning and estimation were required which was time-consuming.

8) Model Evaluation:

Effort: Assess the effort needed to evaluate and interpret the results of K-means clustering.

9) Interpretation and Visualization:

Effort: Estimate the time and effort needed for interpreting and visualizing the segmentation results.

Communication: Plan for effective communication of segmentation findings to stakeholders.

By leveraging this breakdown and incorporating the steps analysed for the case study conducted, we can estimate the time and resources needed.

Dataset Sources:

Websites with Links
https://www.smev.in/charging-infrastructure
https://dataspace.mobi/dataset/electric-vehicle-charging-station-list/resource/f39bb18a-bf5b-4e93-a22e-91f13b2ad9a7?view_id=e1be1b81-cd77-43b2-9102-2f3e85559a75
https://pib.gov.in/PressReleasePage.aspx?PRID=1947389 (statewise)
https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1944594
(statewise-registered under vahan)
https://www.iea.org/data-and-statistics/charts/electric-vehicle-share-of-vehicle-sales-by-mode-and-scenario-in-india-2030
https://dataspace.mobi/dataset/electric-vehicle-charging-station-list (outdated charging stations)
https://tn.data.gov.in/resource/city/highway-wise-electric-vehicle-ev-charging-stations-out-which-479-charging-stations
https://www.iea.org/data-and-statistics/charts/electric-vehicle-share-of-vehicle-sales-by-mode-and-scenario-in-india-2030
https://ev-database.org/#sort:path~type~order=.rank~number~desc range-slider-range:prev~next=0~1200 range-slider-acceleration:prev~next=2~23 range-slider-topspeed:prev~next=110~350 range-slider-battery:prev~next=10~200 range-slider-towweight:prev~next=0~2500 range-slider-fastcharge:prev~next=0~1500 paging:currentPage=0 paging:number=9
https://www.kaggle.com/datasets/karivedha/indian-consumers-cars-purchasing-behaviour/data

Datasets and their description:

Using the Pandas DataFrame, I have imported CSV files of the following datasets and have also used Info and Describe for the same.

EV_DATA_INDIA

Data columns (total 18 columns):

#	Column	Non-Null Count	Dtype
0	Brand	15 non-null	object
1	Car	15 non-null	object
2	Style	15 non-null	object
3	Plug_Type	15 non-null	object
4	Rapid_Charge	15 non-null	object
5	Top_Speed	15 non-null	int64
6	Range	15 non-null	int64
7	Transmission	15 non-null	object
8	VehicleType	15 non-null	object
9	StartPrice	15 non-null	int64
10	MaxPrice	15 non-null	int64
11	Capacity	15 non-null	int64
12	BootSpace	15 non-null	int64
13	BaseModel	15 non-null	object
14	Reviews_Carwale	15 non-null	float64
15	Ratings_Carwale	15 non-null	int64
16	Reviews_Cardekho	15 non-null	float64
17	Ratings_Cardekho	15 non-null	int64

EV_GEOGRAPHICAL_POPULATION

Data columns (total 2 columns):

#	Column	Non-Null Count	Dtype
0	State	34 non-null	object
1	EV	34 non-null	int64

EV_CHARGING_STATIONS

Data columns (total 7 columns):

#	Column	Non-Null Count	Dtype
	---	-----	----
0	region	202 non-null	object
1	state	202 non-null	object
2	address	202 non-null	object
3	latitude	202 non-null	float64
4	longitude	202 non-null	float64
5	type	202 non-null	object
6	power	202 non-null	int64

NON-EV_BEHAVIOUR

Data columns (total 13 columns):

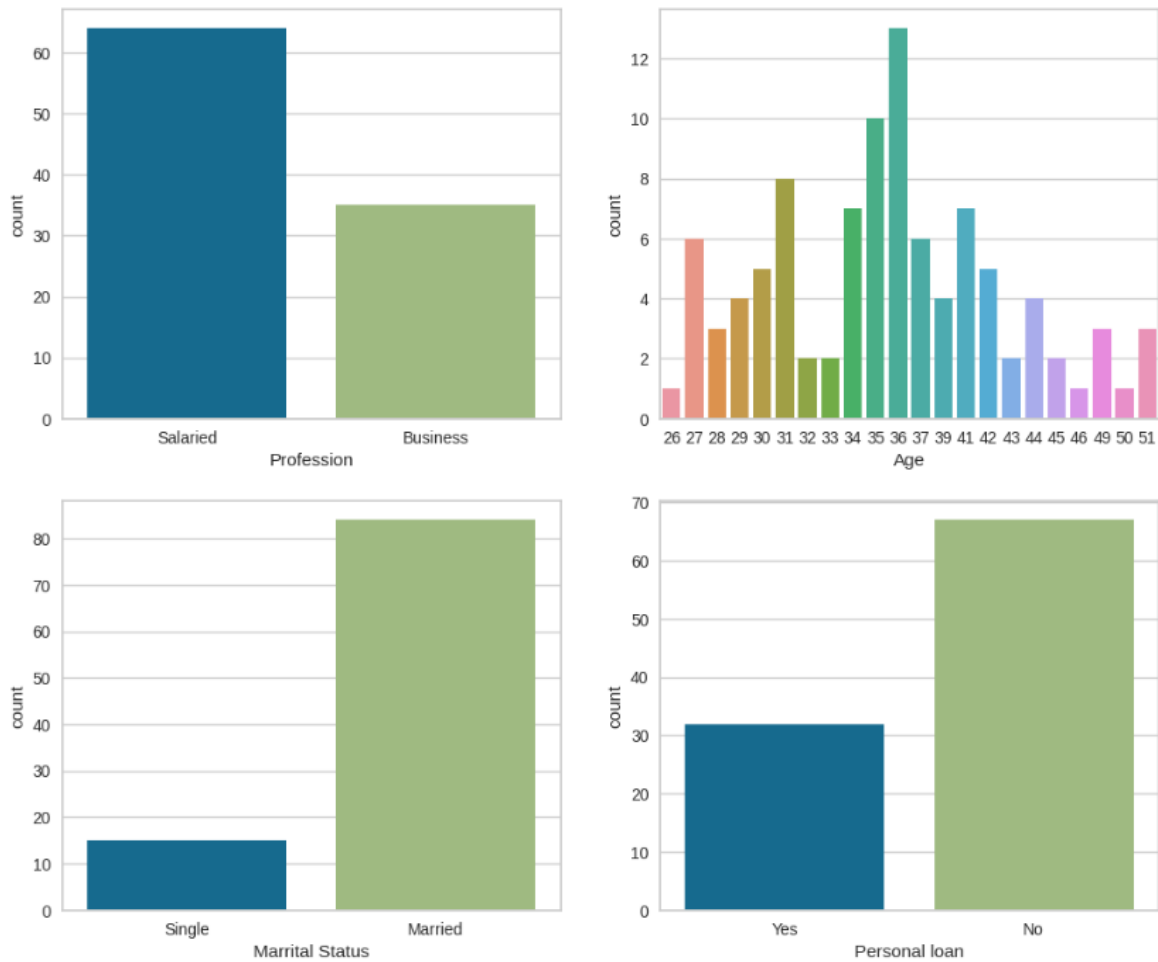
#	Column	Non-Null Count	Dtype
	---	-----	----
0	Age	99 non-null	int64
1	Profession	99 non-null	object
2	Marital Status	99 non-null	object
3	Education	99 non-null	object
4	No of Dependents	99 non-null	int64
5	Personal loan	99 non-null	object
6	House Loan	99 non-null	object
7	Wife Working	99 non-null	object
8	Salary	99 non-null	int64
9	Wife Salary	99 non-null	int64
10	Total Salary	99 non-null	int64
11	Make	99 non-null	object
12	Price	99 non-null	int64

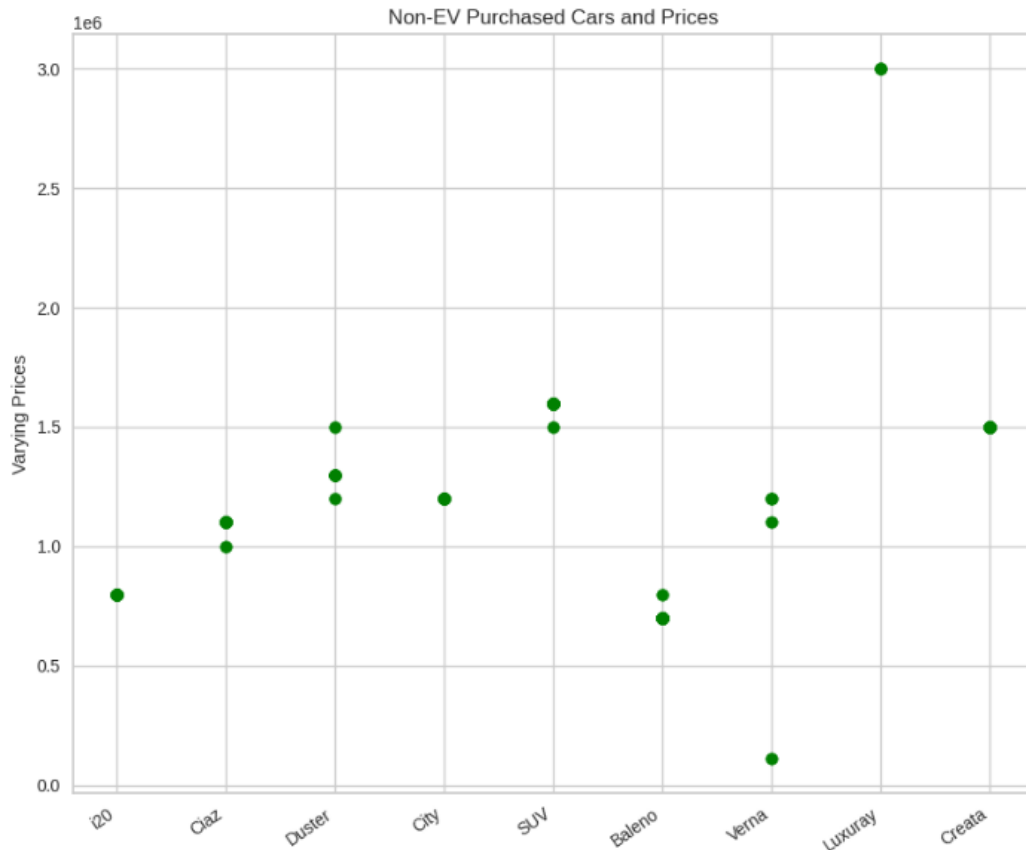
DATASET 1:

The behaviour of Customers Purchasing Non-EV in India.

Performing EDA on the Behavioural Dataset and Visualising the features.

MatPlotLib and Seaborn plotting methods were utilised for this step.





The average price for each car ‘Make’ in the dataset was also found and plotted as a grouped bar graph.

This step is a crucial method in order to analyse and visualise the data we have at hand.

Performing PCA on the copied dataset.

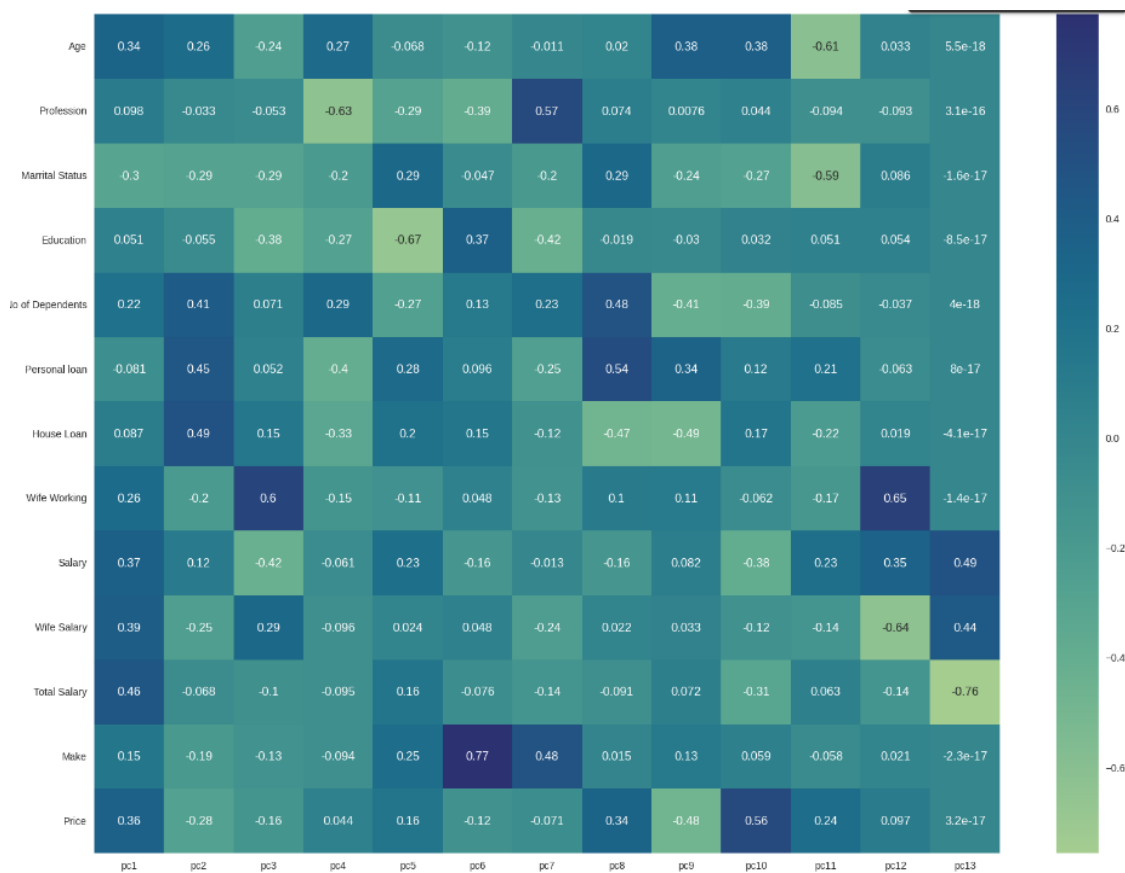
Simplifying complex multivariate datasets.

This step involves handling categorical and numerical features and variables in the dataset. By shifting the values of the n_components, we can find the values of each component.

After applying PCA							
	pc1	pc2	pc3	pc4	pc5	pc6	pc7 \
0	-3.508803	-1.242395	-1.070874	-2.108843	0.479000	1.505867	0.381268
1	-0.188394	1.284799	1.034237	-1.752116	-0.765243	-0.108937	-0.908496
2	-0.062948	2.877911	-0.258709	1.291994	1.331130	0.534794	0.057796
3	0.506741	-0.238137	0.291381	1.546050	-1.184148	0.327658	-1.059920
4	0.793199	-0.670607	0.232101	-1.472065	-0.216105	0.756490	-0.053368

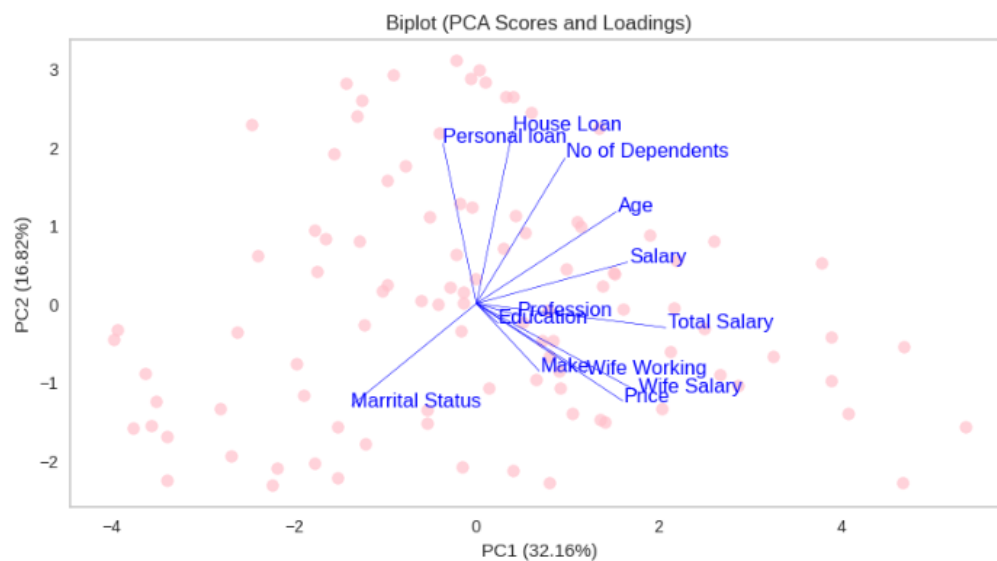
	pc8	pc9	pc10	pc11	pc12	pc13
0	1.001599	0.652114	0.258463	-0.320854	-0.365426	1.572707e-16
1	0.055397	0.053369	0.434174	0.032615	0.172405	2.072090e-16
2	0.560843	-0.193009	0.697789	-0.342153	-0.057021	-3.109129e-16
3	0.191520	0.250963	-0.071230	-0.240159	0.594051	-5.075310e-16
4	1.371336	0.502788	0.275563	1.227557	0.198080	3.428311e-16

The loadings indicate how the original variables are combined to form principal components. Here is a snippet of the resulting `loadings_matrix`.



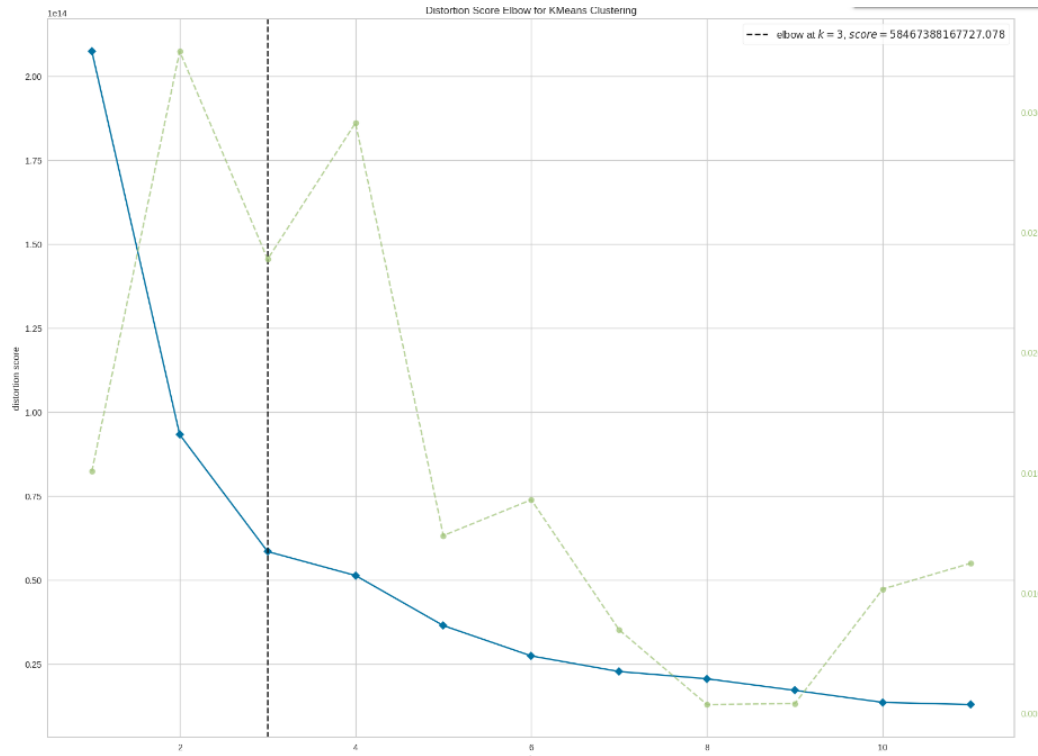
Making a Biplot Graph for the features with the Standardised Data.

Here, as I conducted in the case study, I have used the Scatter Plot with the PCA_scores and also added the Arrow plot.

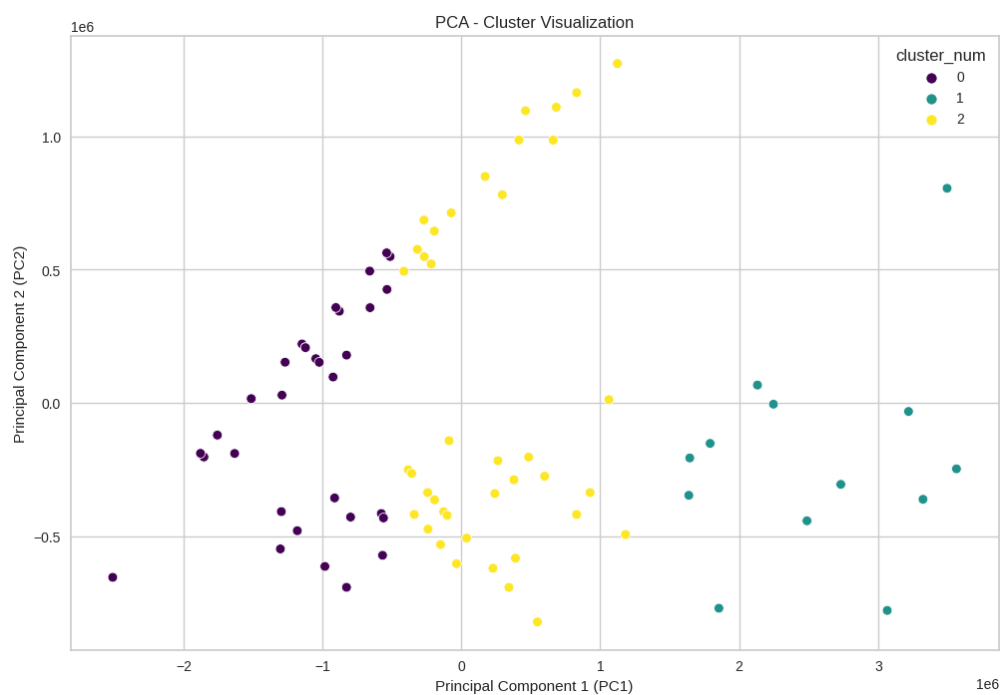


Using the Elbow method in order to find the Optimal Number of Clusters

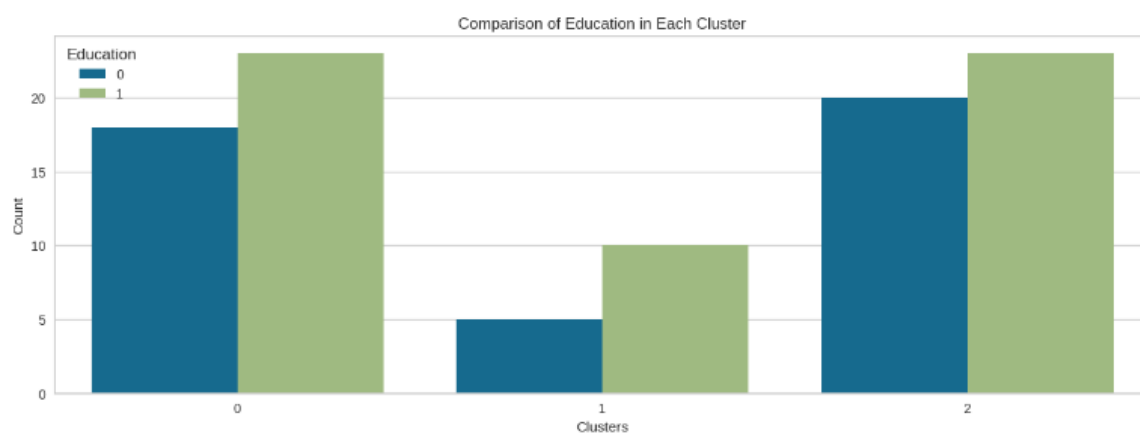
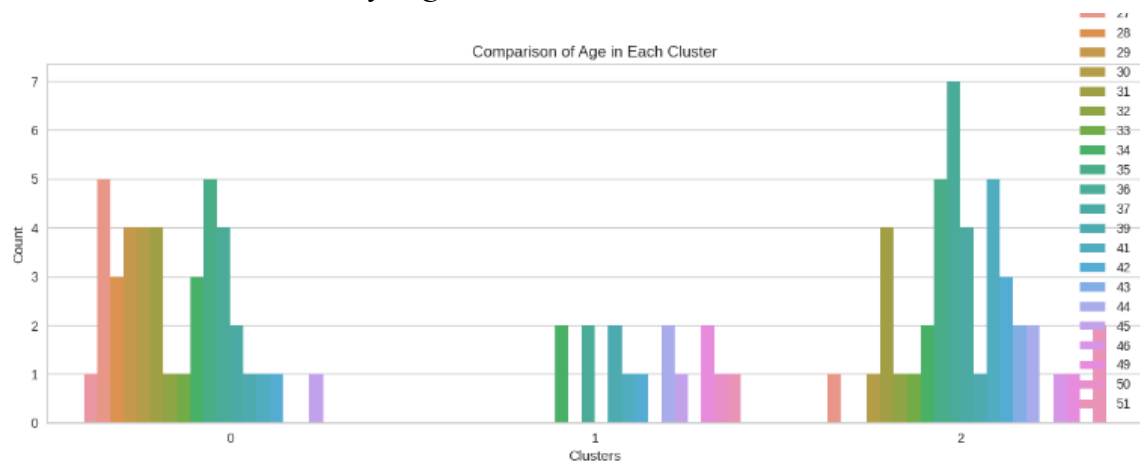
KElbowVisualizer from the YellowBrick library helps in achieving this purpose. The main aim is to find the point of decline in the graph.



Using the Cluster Num with K-Means Analysis

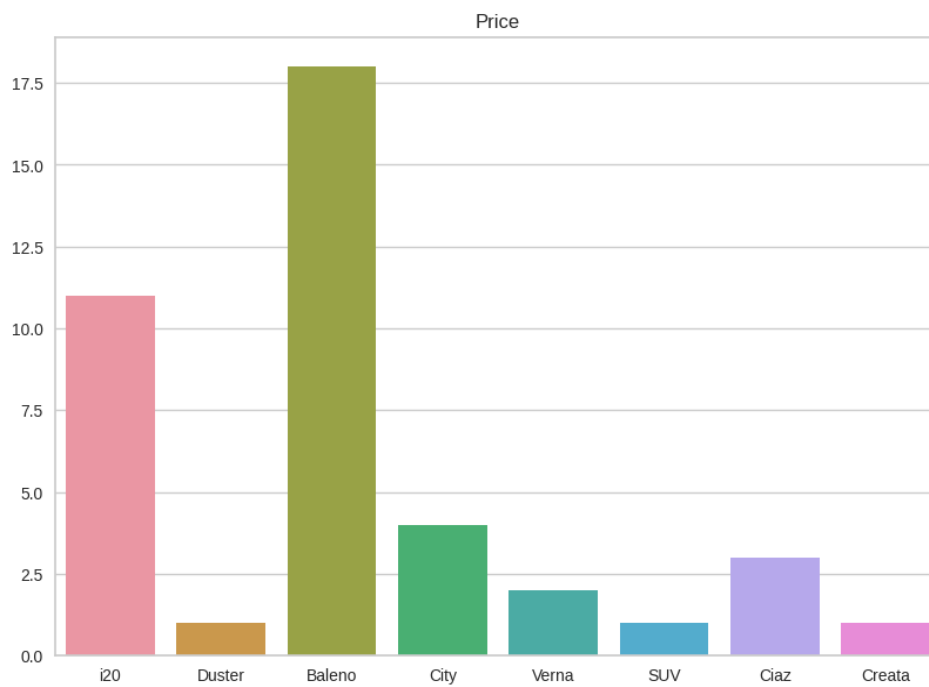


Analysing the Clusters Based on Features

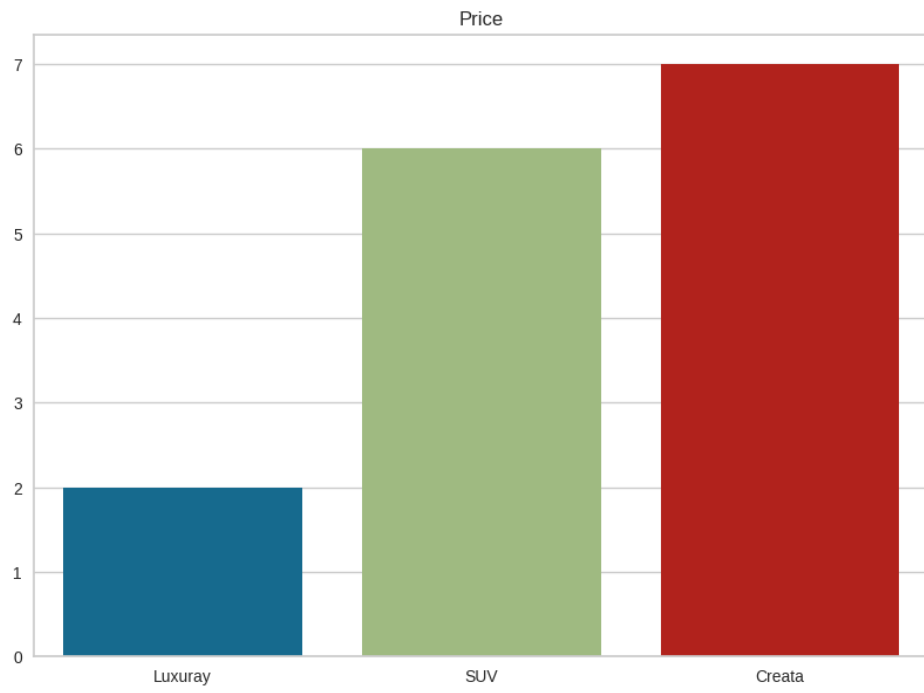


Comparison of No of Dependents in Each Cluster

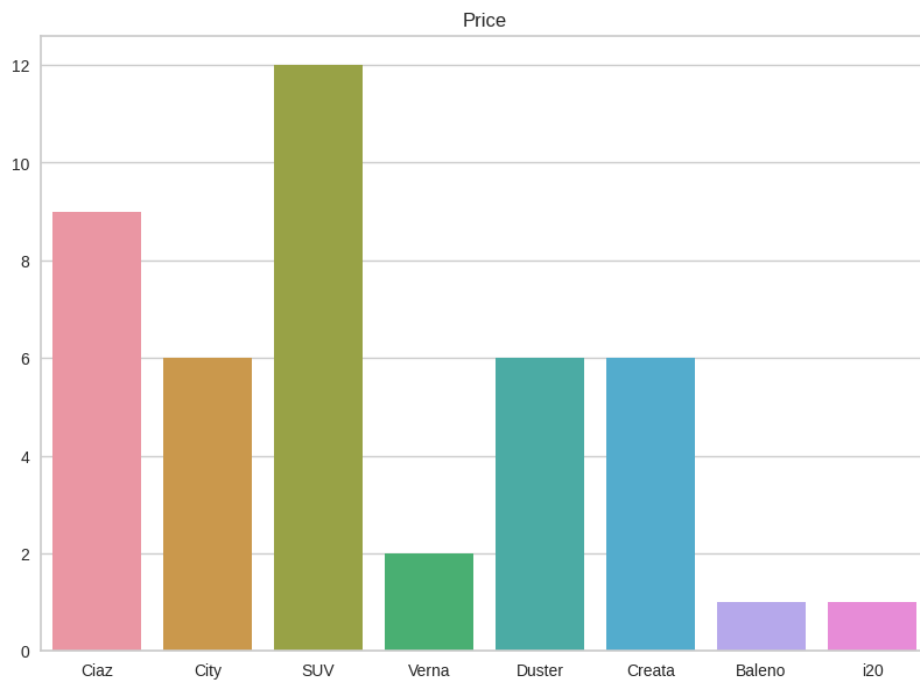
Cluster 0 Feature Distributions



Cluster 1 Feature Distributions

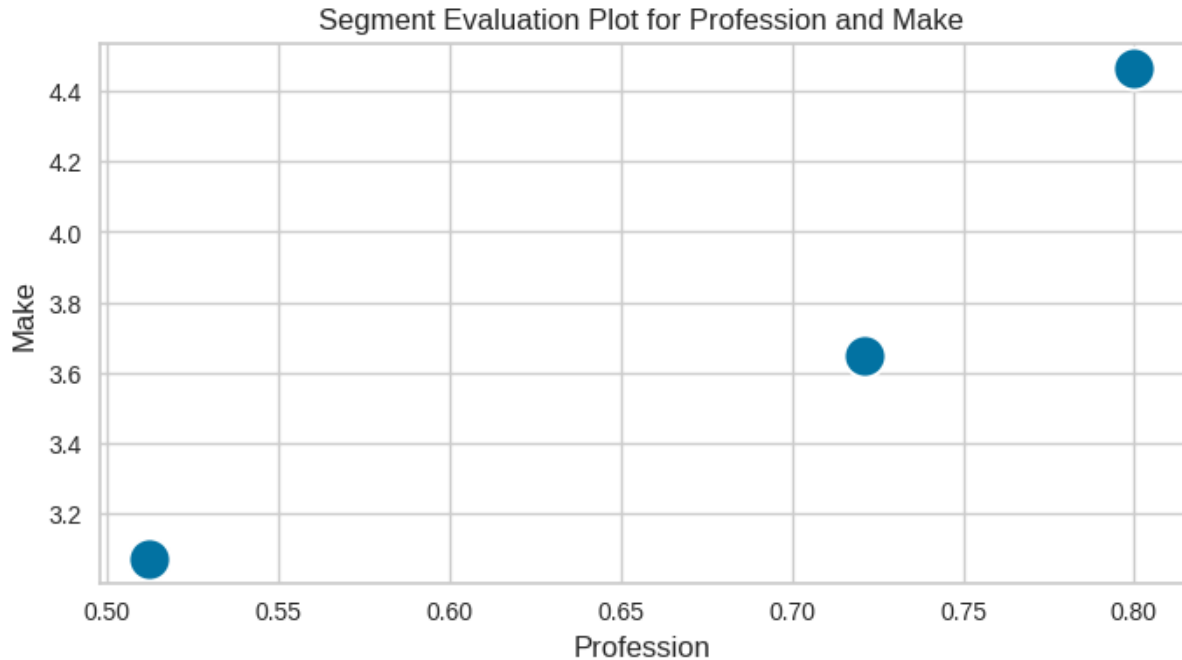


Cluster 2 Feature Distributions



Making A Simple Segmentation Evaluation Plot

This step involves making data frames of various categorical features to analyse the segments.

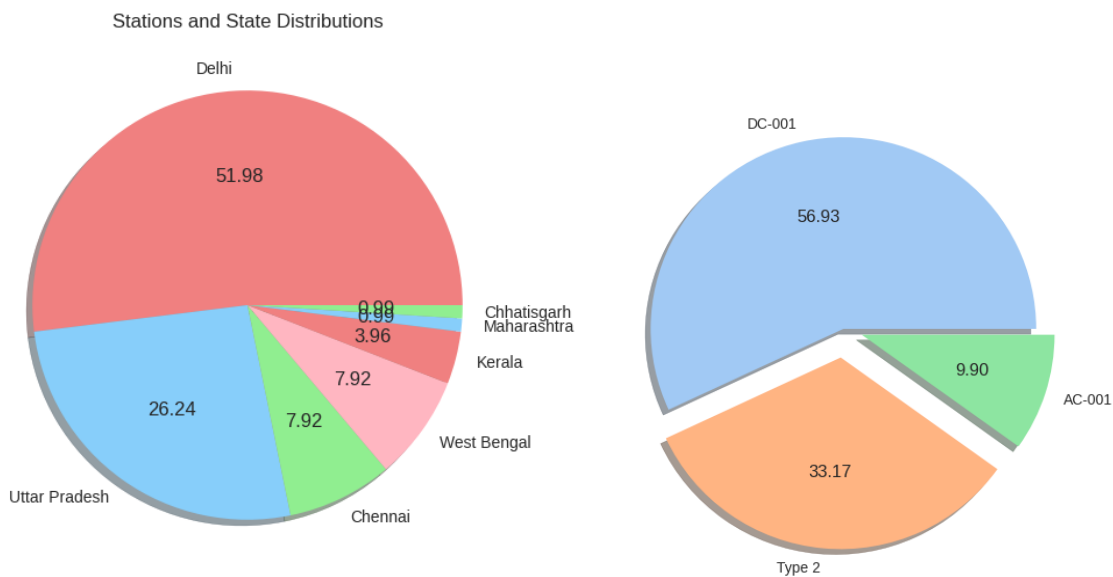
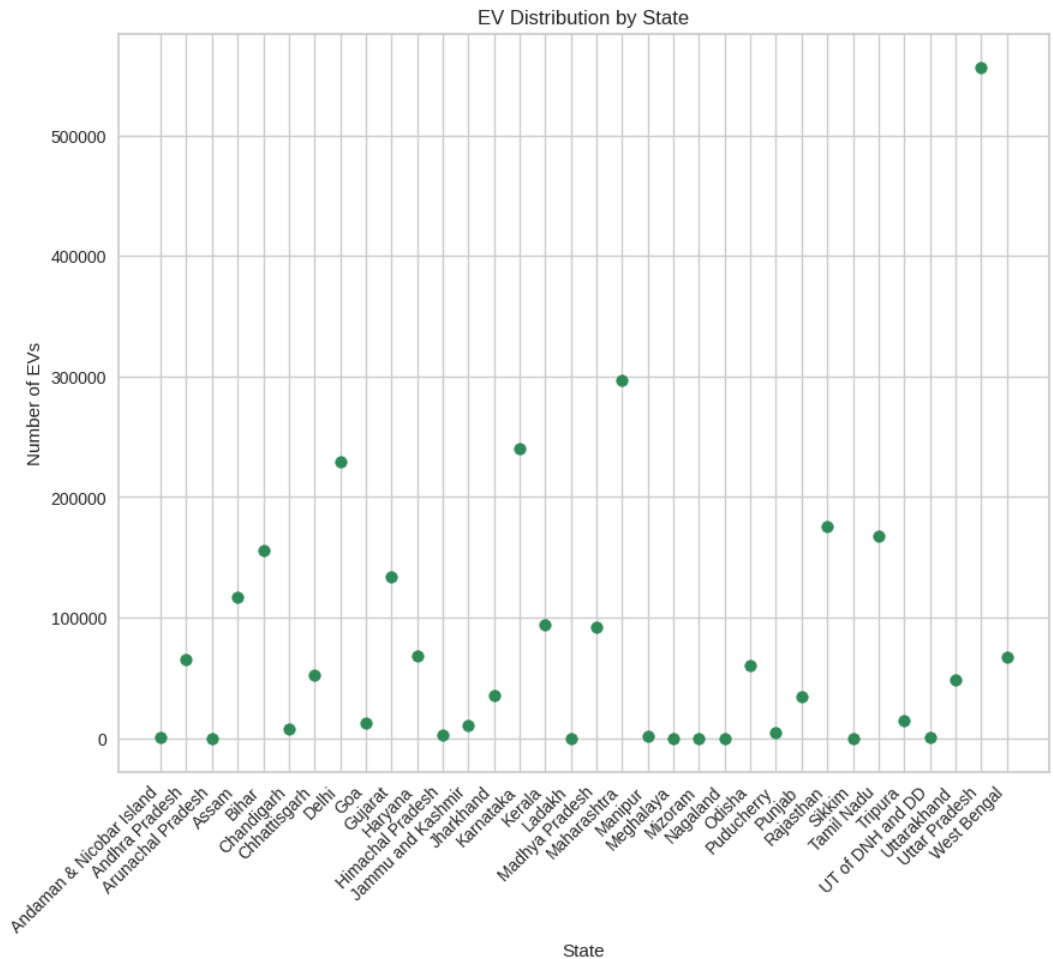


Summary of Segments in Dataset 1 (Behavioural)

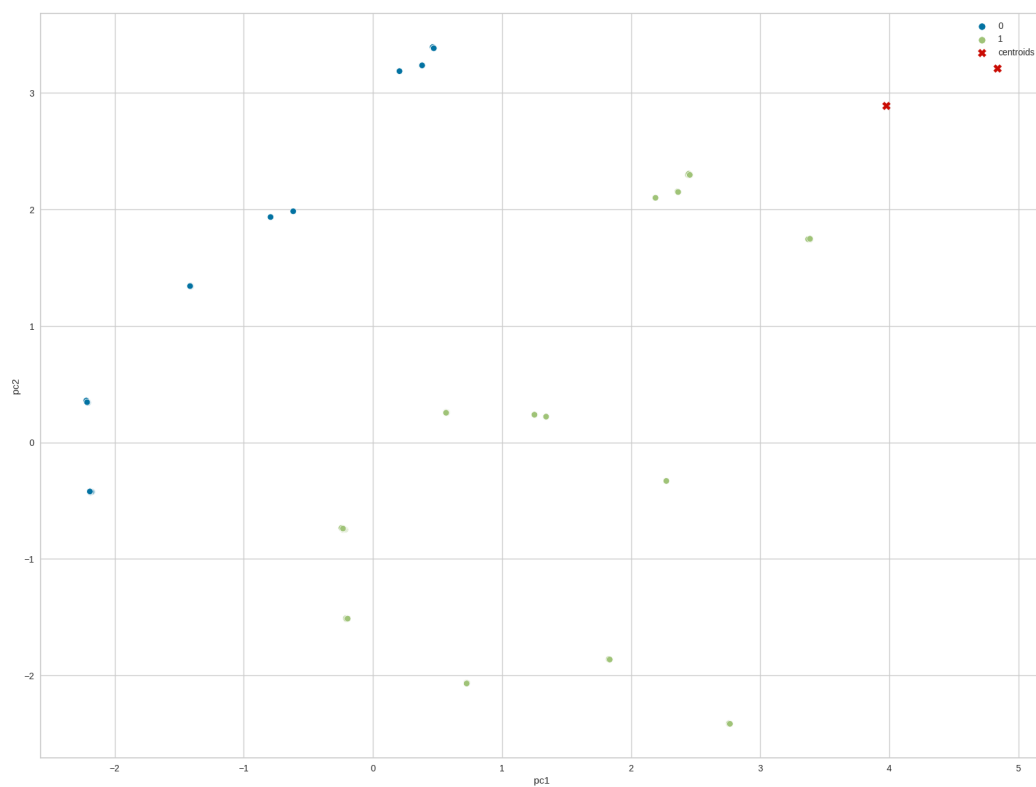
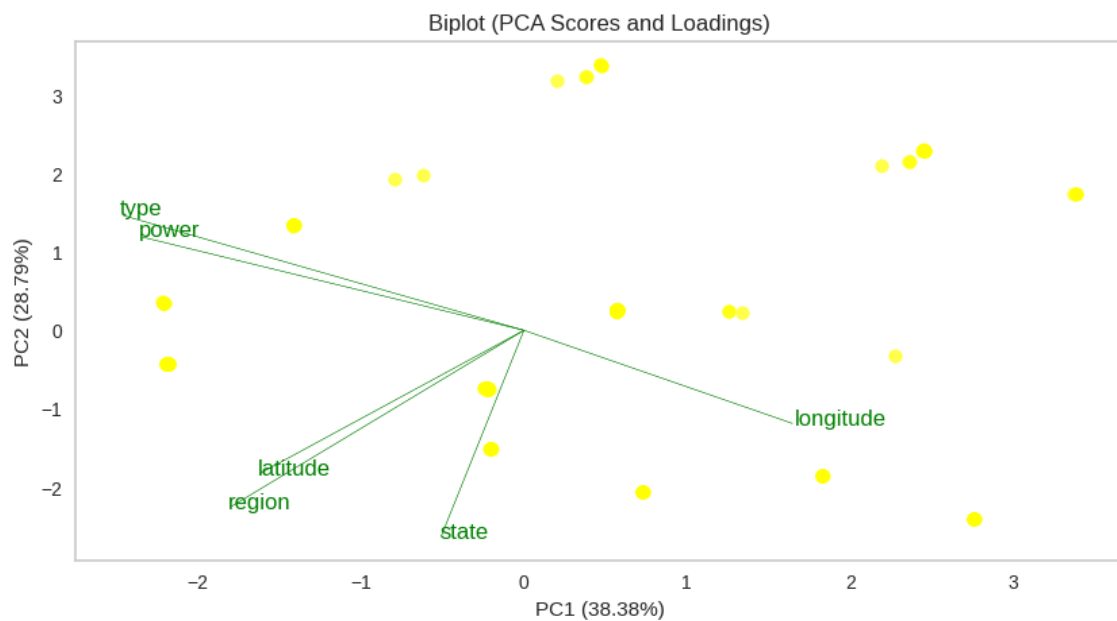
The steps taken to create segments of Non-EV customers prove that the segmentation may happen concerning the Car Make and Price. The three clusters have a vast difference in Marital status as well as Salary, which is also reflected in the Car they choose to purchase.

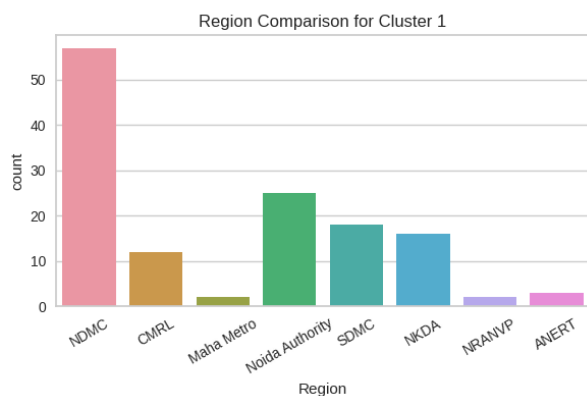
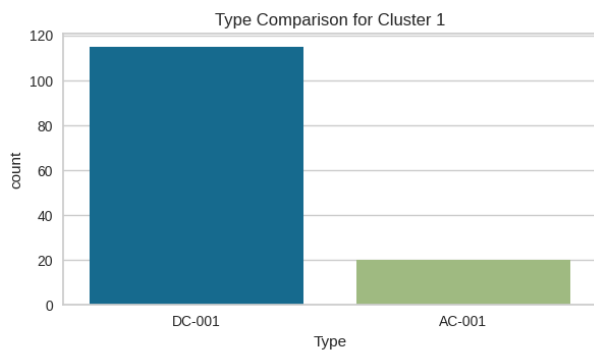
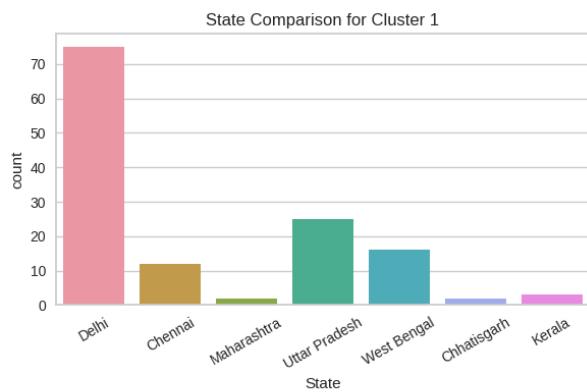
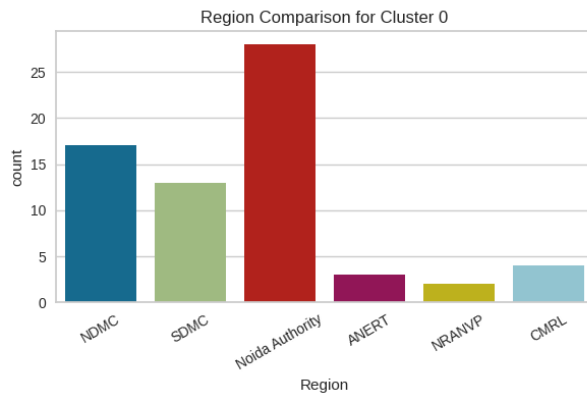
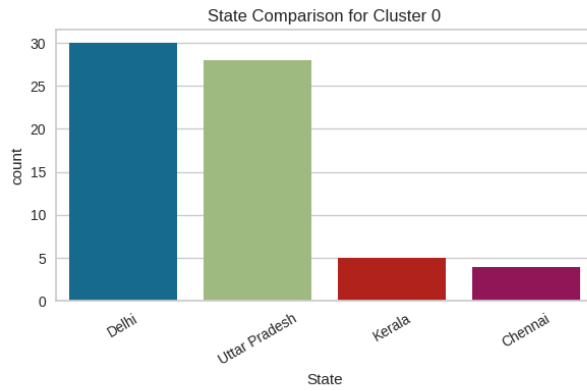
This point is crucial when it comes to EV Cars as these are generally at a higher range of price. Based on surveying and calculating the census of the popularity of electric cars, we can analyse the behaviour of customers and similarly segment them.

DATASETS 2 AND 3:
Statewise Electric Vehicles registered under the Govt.
Charging Stations Available in Each State and Their Distributions



By performing similar steps in these Geographical Datasets, we can segment them based on varying features. The results for PCA and Segmentation for this dataset are as follows.





Summary of Analysing Segments of Geographical Charging Stations Dataset

The two clusters prove that the segments may be formed based on the type of charge and the major cities connecting in between. This is a piece of pivotal information needed as EVs can in general be categorised based on the availability of their Plug-Type. Producing

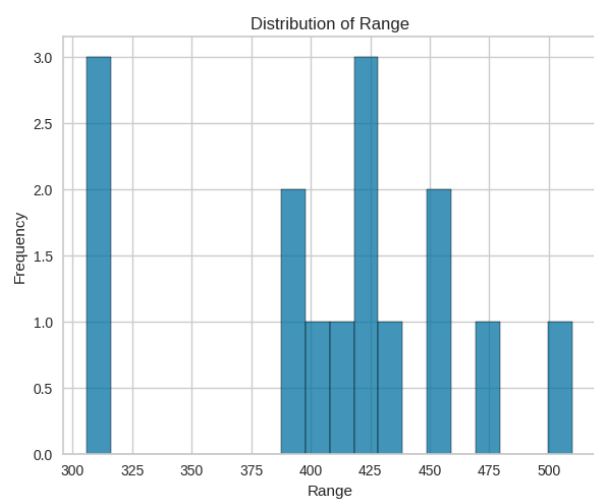
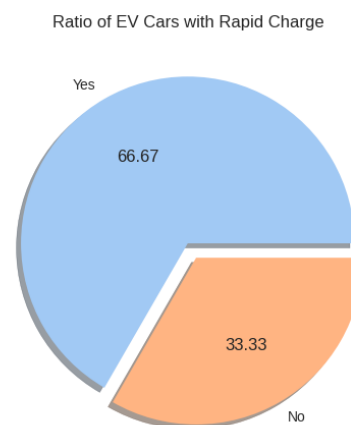
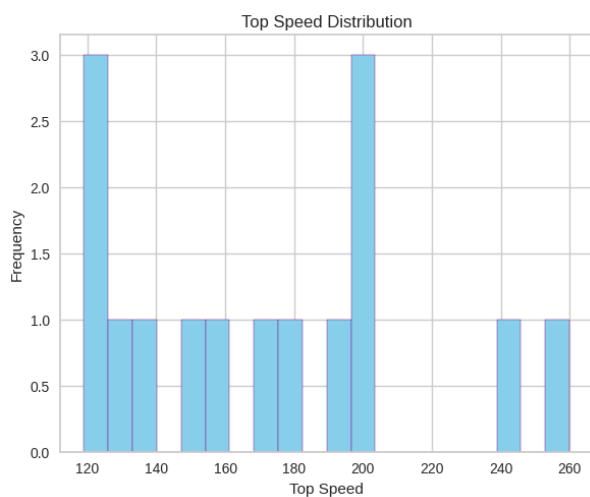
and marketing EVs based on the more important or popularly available Type of Charge is vital.

DATASET 4:

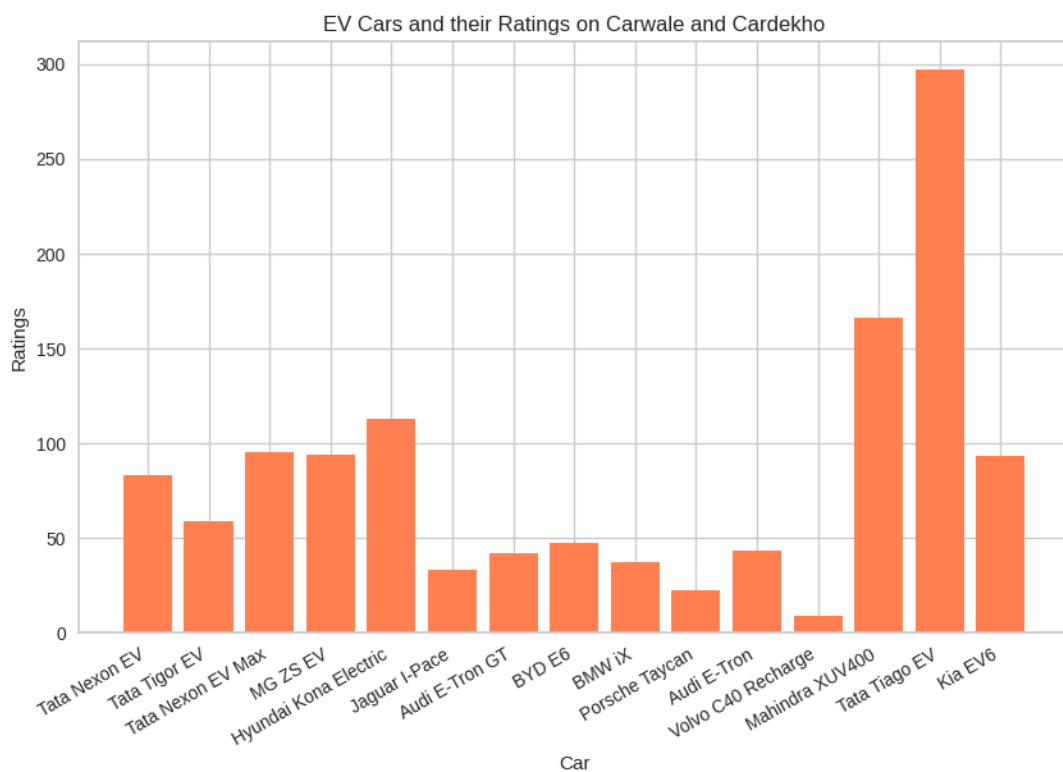
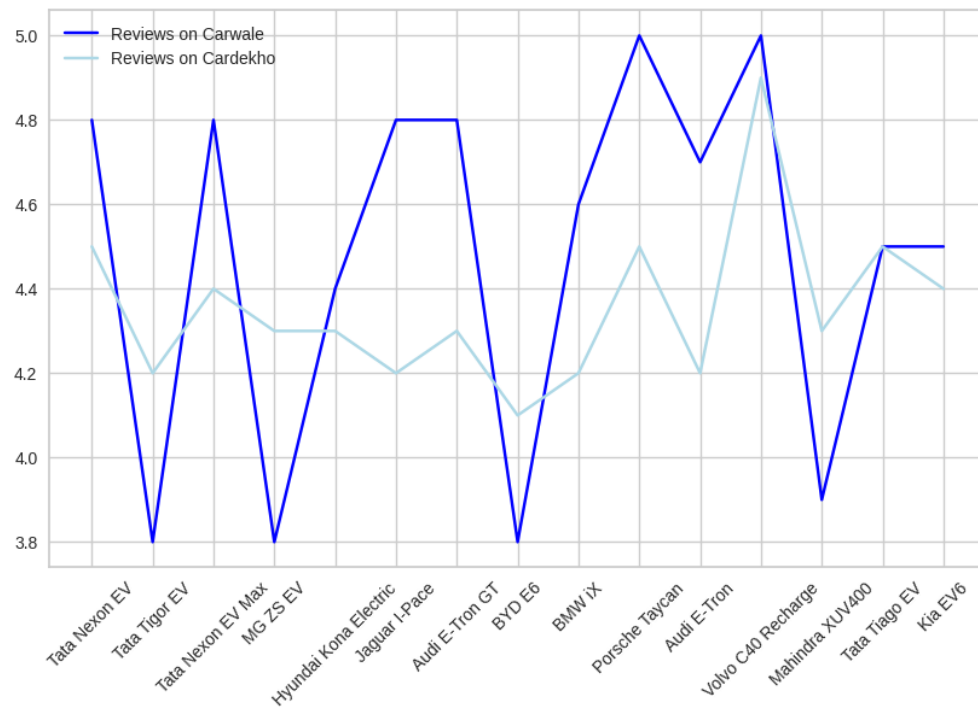
Electric Cars that are available and on-road in India along with their details.

This dataset is essential for segmentation as the above-performed analyses can also be used to study the segments created by this dataset.

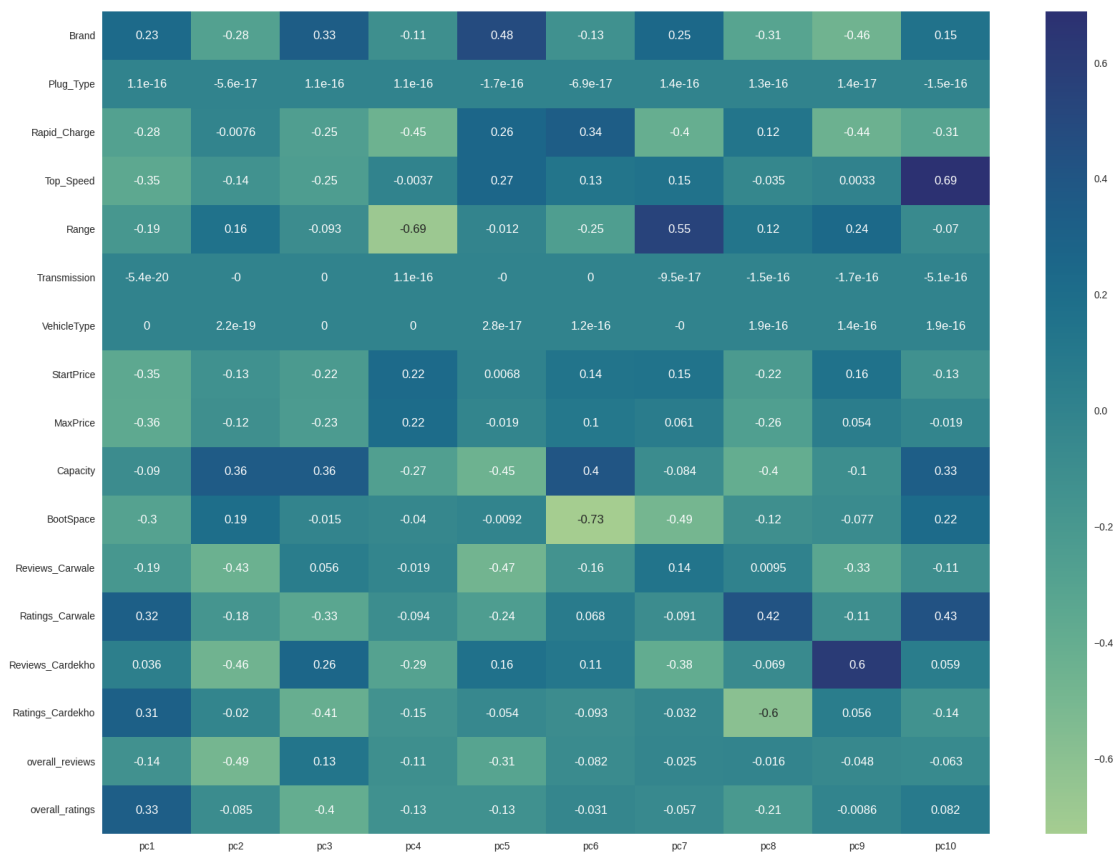
For the EDA on this dataset, various plots have been used to better visualise the features and their value.



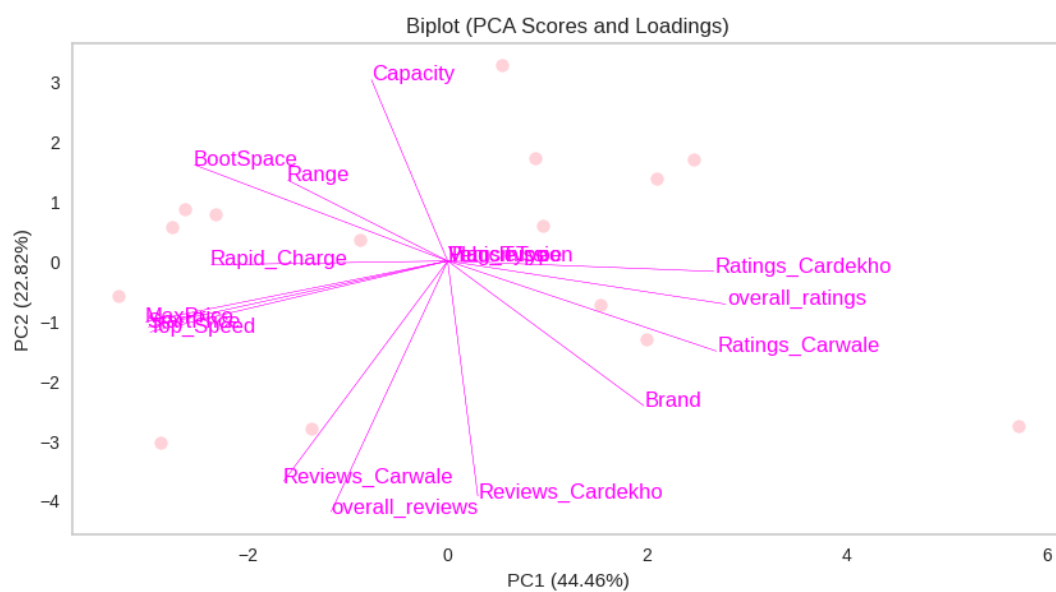
For the feature engineering and dataset enhancement for this data, I have also manually gathered data of Reviews and Ratings from websites such as Carwale and Cardekho to better understand the segmentation process.



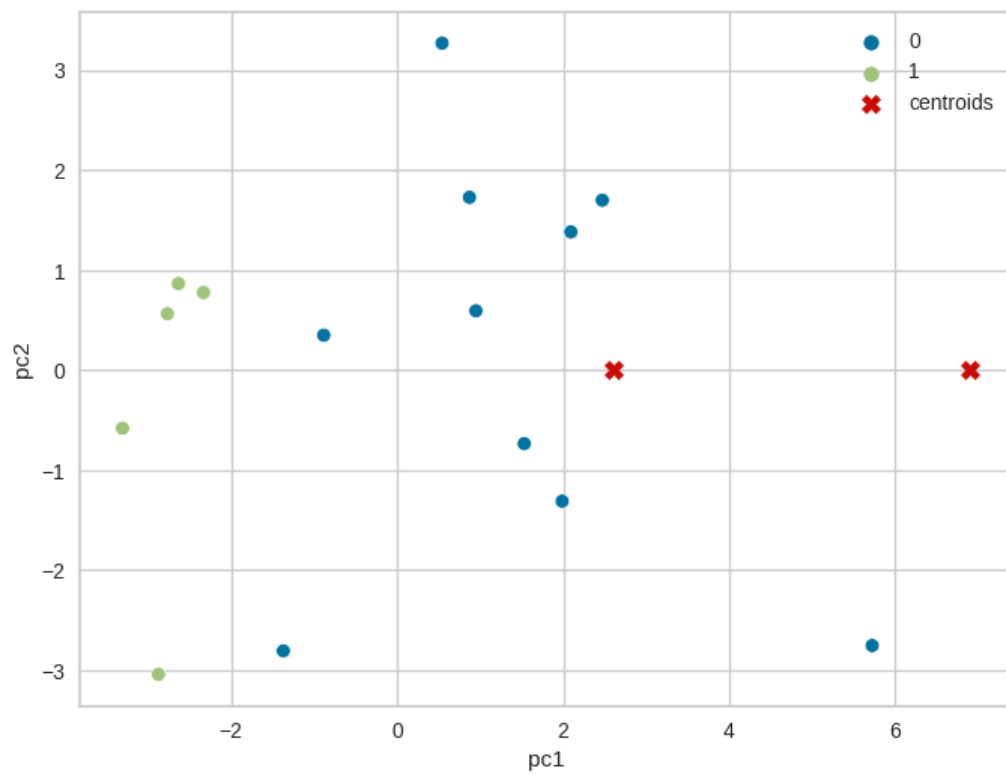
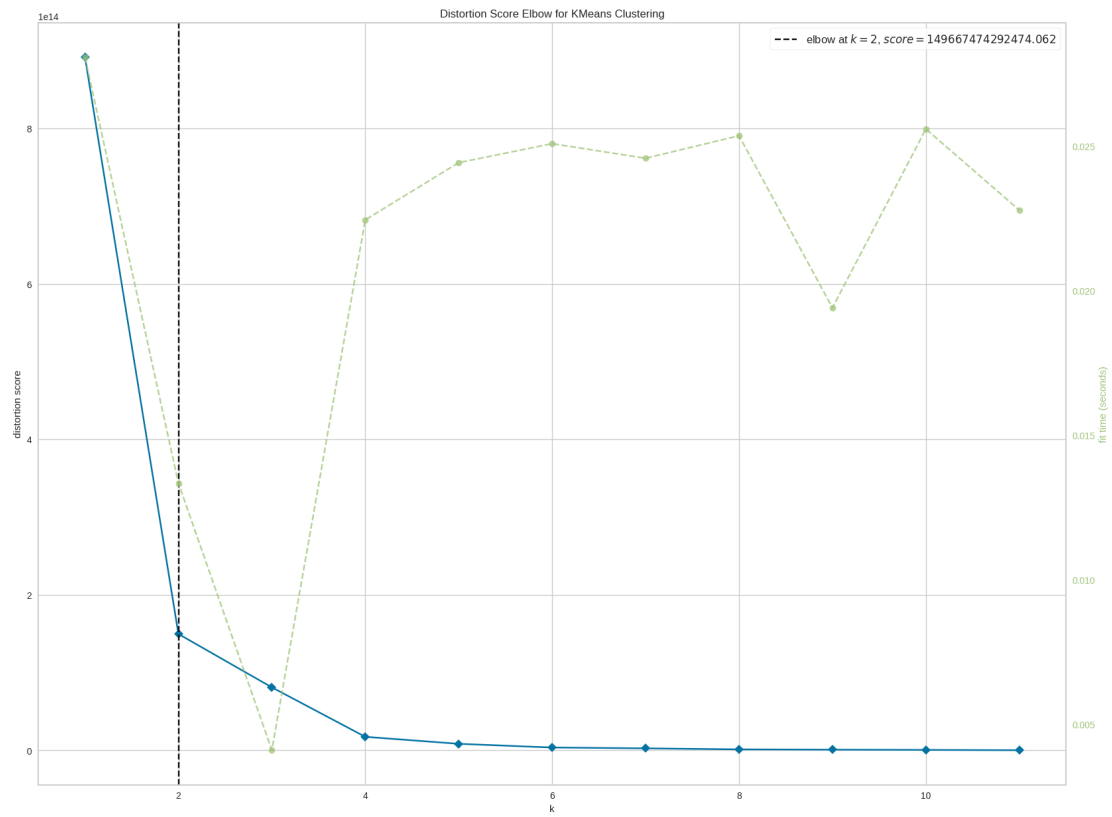
After applying PCA on categorical and numerical values and dropping complex features, the loadings_matrix was as follows.



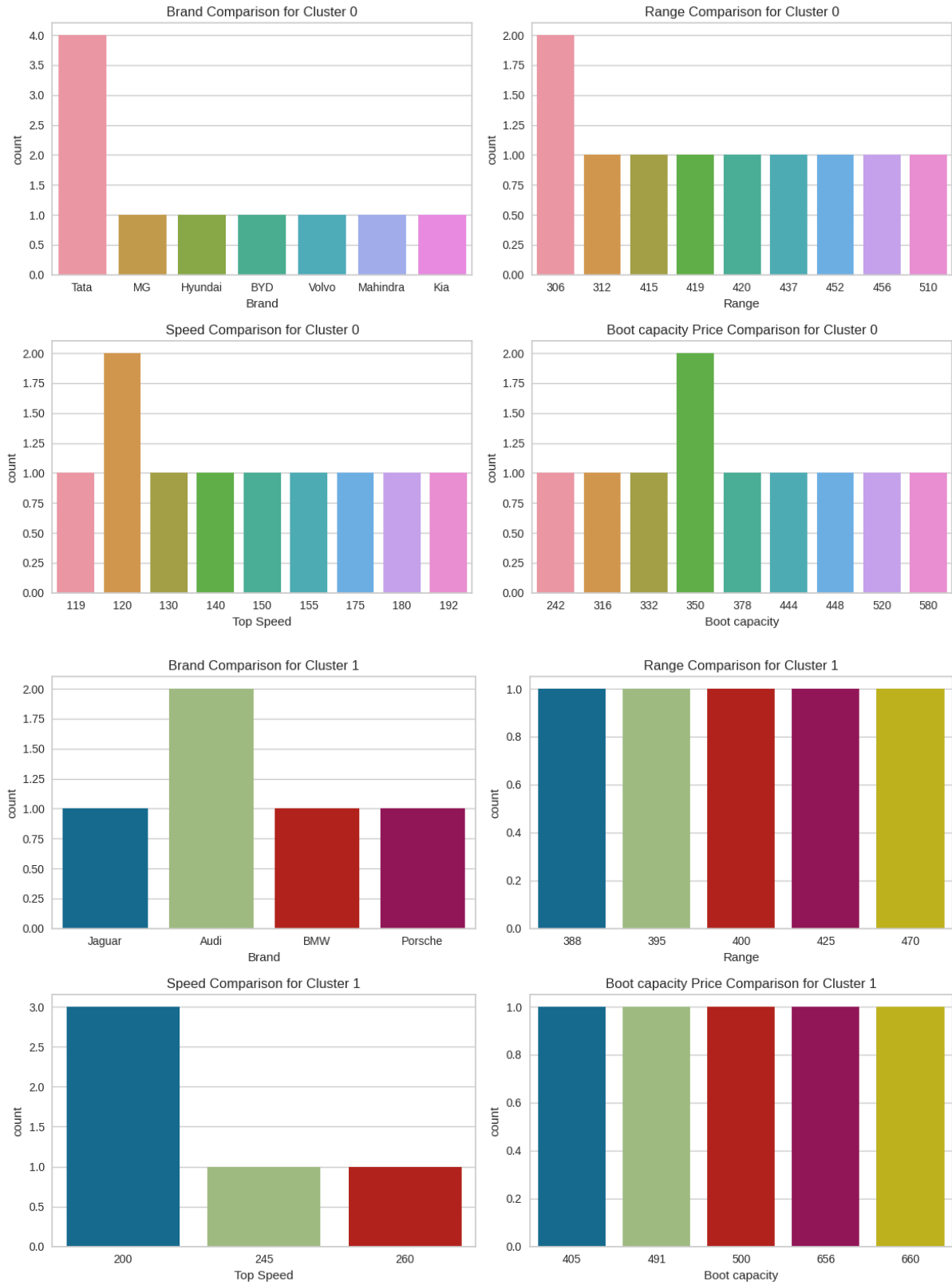
And the resulting biplot came as:

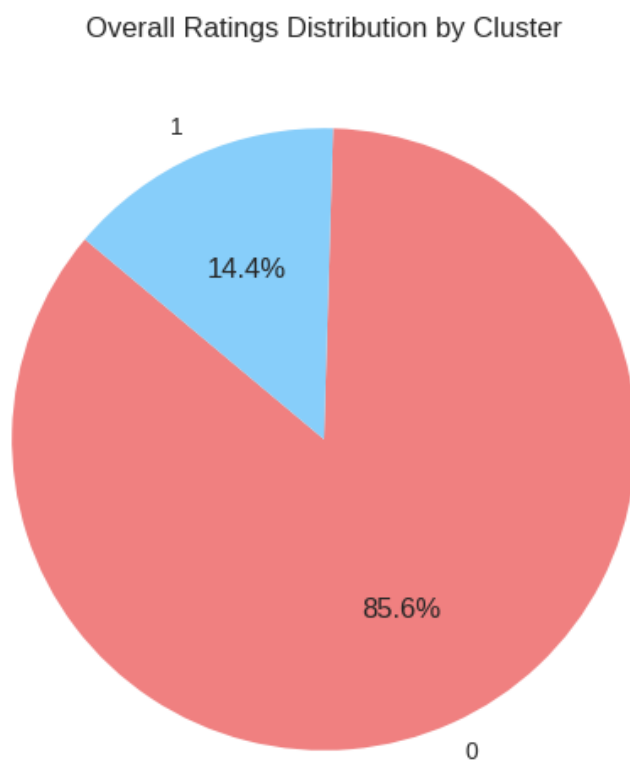
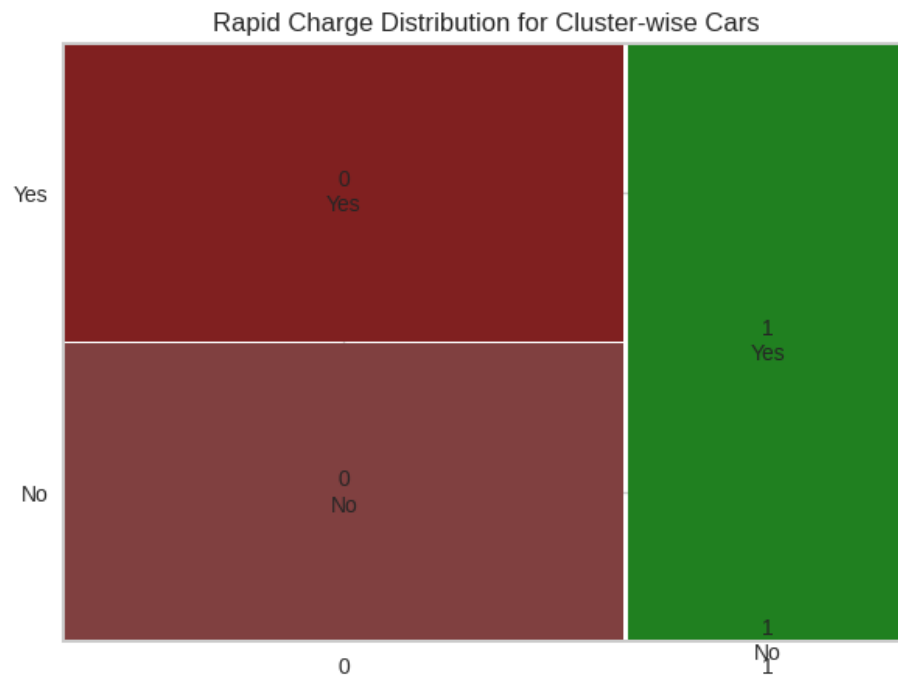


By the result of obtaining 2 clusters from the elbow method, the plot of the K-Means clustering is shown below.



Analysing and Visualising the Features in Each Segment.





Conclusion

The final segmentation of the EV dataset gives us two prominent segments; one segment contains EV cars priced at an affordable rate with a decent range and other features and the second segment with cars that are higher priced and deemed as luxury or expensive. The previous datasets' segmentation has segmented customers based on their salary and the price of the car purchased, the second dataset segmentation considered the state of the charging station and the two segments of Type of chargers, one of them being solely Type 2 Charger Segment. The segments also highlight the availability of Rapid Charge available for the more expensive cars in the segment.

The final dataset shows that all the passenger vehicles available as electric have a Type 2 Plug-type. These two segments may be merged based on the affordability of on-road electric cars for the general public of India.



By integrating these elements into the marketing mix, the company can create a holistic strategy that aligns with the characteristics and preferences of the identified market segments, ultimately enhancing the success of its EV offerings in the Indian market.

The link to the final Python Notebook is in this Github Permalink:

https://github.com/raima-j/Feynn-Labs-Internship/blob/d13e817edf5335591cbe603942b0a5ce415623b9/P2_2_EV.ipynb