Report on EV Market Segmentation in India

By Mohammed Ahmed Hussain (*Batch-SB-23-9-5 MLI*)

Problem Statement

To analyze the Electric Vehicle market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles.

(CUSTOMER/VEHICLE/B2B) SEGMENTS: Apart from Geographic, Demographic, Psychographic, Behavioral segments, teams can consider different CATEGORY of Segments for the Segmentation Tasks, based on AVAILABILITY OF DATA. Market Segmentation comes with a wide scope of possibility and Segments created can change based on different datasets collected.

Objective

Analyzing the Electric Vehicle market in India using Segmentation analysis and coming up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles.

Introduction

The electric vehicle (EV) market in India stands at the forefront of transformative change, marked by a surge in innovation, infrastructure development, and a growing awareness of sustainable mobility. This report embarks on a journey to dissect the market through the lens of segmentation analysis. We aim to unravel the intricacies of consumer behavior, preferences, and expectations, coupled with a discerning look at the optimal product profiles that will resonate with the Indian audience.

In an era of data abundance, our approach is underpinned by the rich tapestry of information available. This not only allows us to explore traditional segmentation parameters such as geography, demographics, psychographics, and behavior but also affords the flexibility to unearth novel insights from an array of datasets. As we navigate through this segmentation odyssey, the ultimate goal is to derive a

strategy that aligns with the core values and needs of our target audience, ensuring a seamless integration of electric vehicles into their lives.

This report sets the stage for a comprehensive analysis, revealing the mosaic of opportunities and challenges that characterize the Indian EV landscape. Join us as we decipher the code to not just understand the market but strategically shaping our approach to become a key player in the electric vehicle revolution.

Datasets Used

- EV STATS.csv
- Indian automoble buying behavour study 1.0.csv

Exploratory Data Analysis

1. Indian automoble buying behavour study 1.0

It has 13 columns namely, 'Age', 'Profession', 'Marrital Status', 'Education', 'No of Dependents', 'Personal loan', 'House Loan', 'Wife Working', 'Salary', 'Wife Salary', 'Total Salary', 'Make' and 'Price' with 99 rows.

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
0	27	Salaried	Single	Post Graduate	0	Yes	No	No	800000	0	800000	i20	800000
1	35	Salaried	Married	Post Graduate	2	Yes	Yes	Yes	1400000	600000	2000000	Ciaz	1000000
2	45	Business	Married	Graduate	4	Yes	Yes	No	1800000	0	1800000	Duster	1200000
3	41	Business	Married	Post Graduate	3	No	No	Yes	1600000	600000	2200000	City	1200000
4	31	Salaried	Married	Post Graduate	2	Yes	No	Yes	1800000	800000	2600000	SUV	1600000

We aim to find out what is the age group, income group, professionality, geography etc. of the customer.

The following are the unique values under each column.

```
Profession : ['Salaried' 'Business']
Marrital Status : ['Single' 'Married']
Education : ['Post Graduate' 'Graduate']
Personal loan : ['Yes' 'No']
House Loan : ['No' 'Yes']
Wife Working : ['No' 'Yes' 'm']
Make : ['i20' 'Ciaz' 'Duster' 'City' 'SUV' 'Baleno' 'Verna' 'Luxuray' 'Creata']
```

Behavioral Analysis
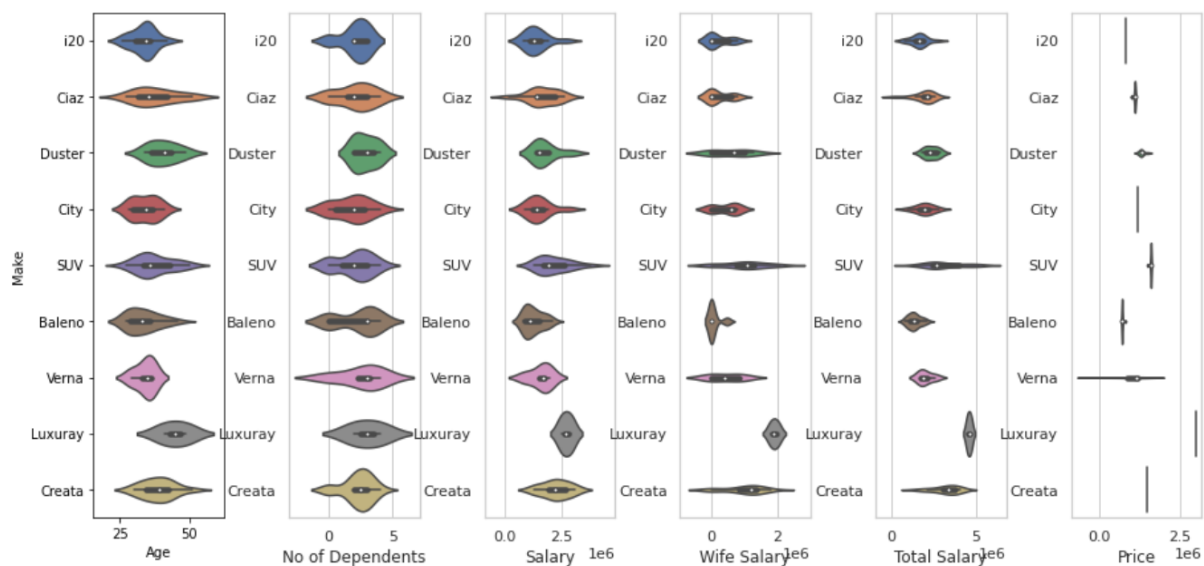
- Age: Younger consumers purchase less expensive vehicles.
- Number of Dependents: Greater number of dependents makes the consumer buy a vehicle with more seats and so they prefer SUVs.
- Salary: If you overlap the normalized salary plots with price plot, you will observe the median of salary violin plot matches that of the price of the vehicle indicating a very direct relationship.
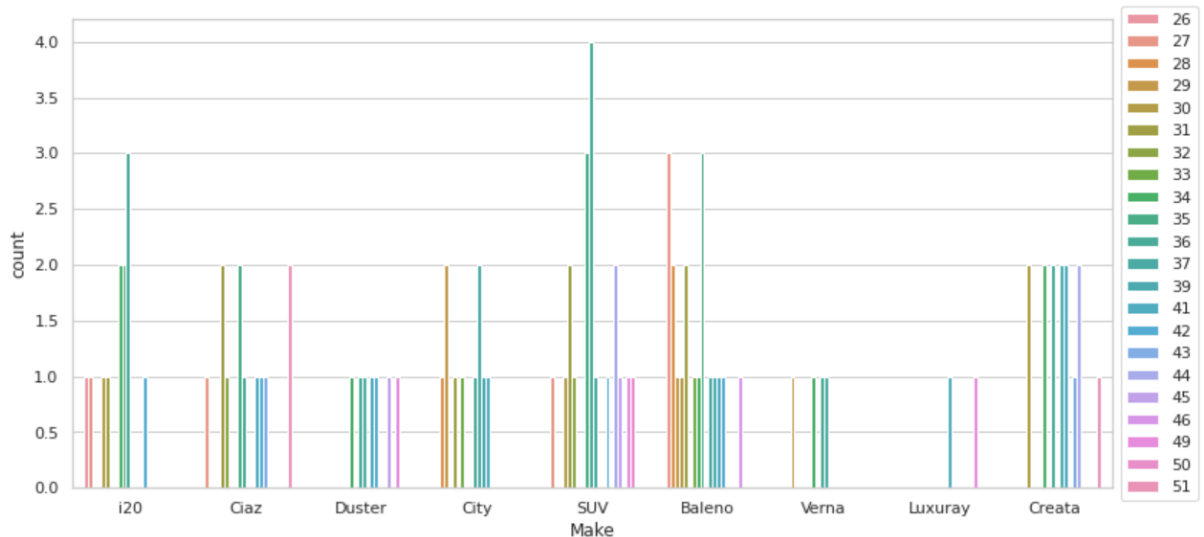

Relationship between age and type of car consumers are likely to buy based on the make of vehicles. The age group between 33 to 39 are more likely to buy any kind of vehicle.
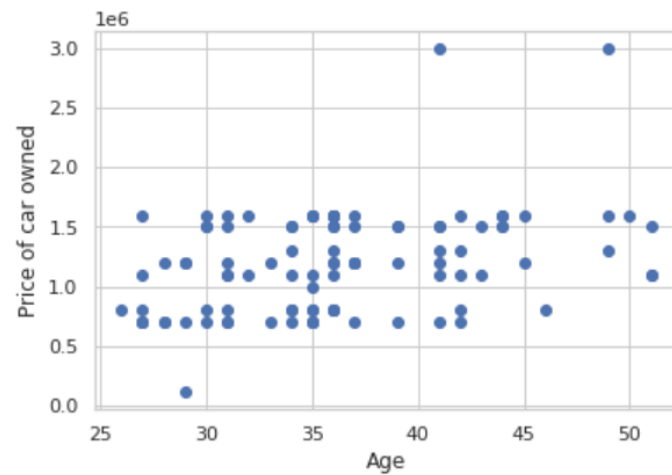
Older people are more likely to buy an expensive vehicle. This is probably the result of collected wealth over the years.
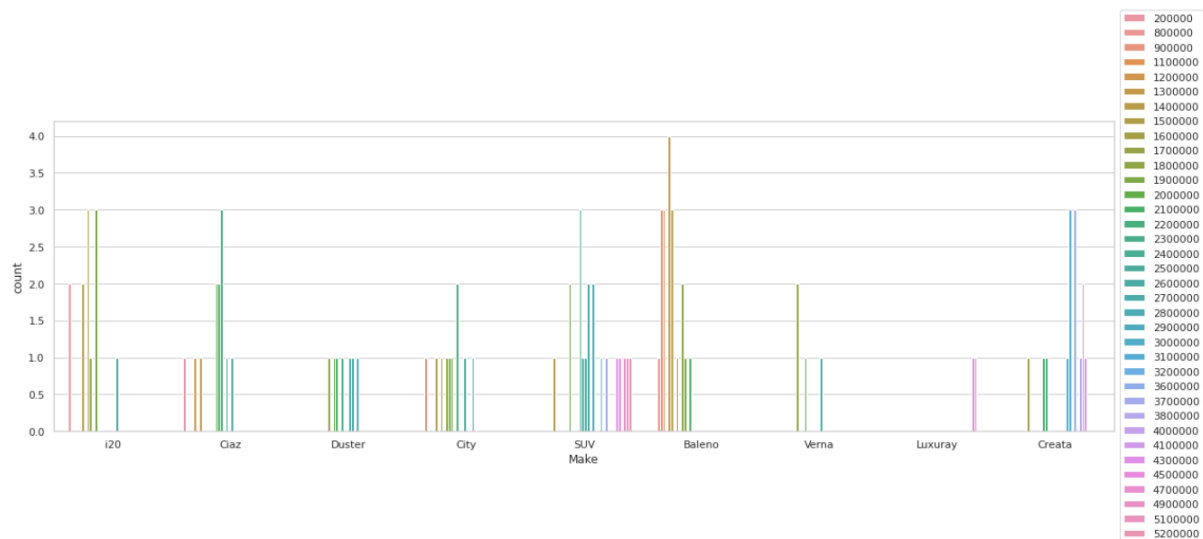

Relation between consumer's total salary and the vehicles they purchase based on the make of vehicles. It is clear that SUVs and Luxury vehicles are only purchased by people with more salary/income.
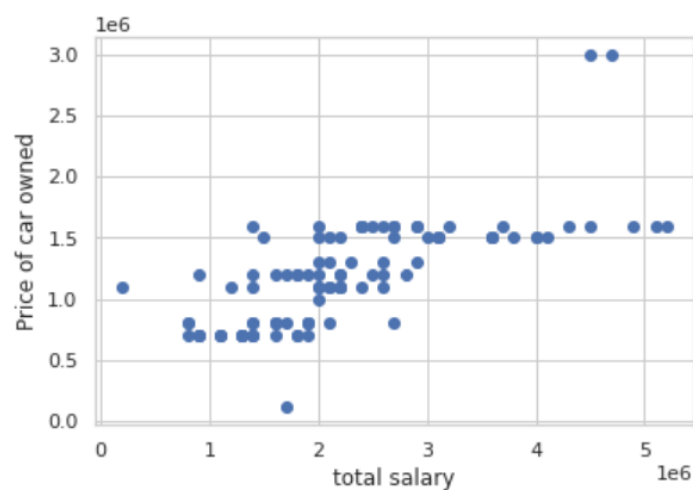
The relationship between the number of dependents on a consumer and the vehicles they tend to purchase, based on the make of vehicles they tend to purchase.
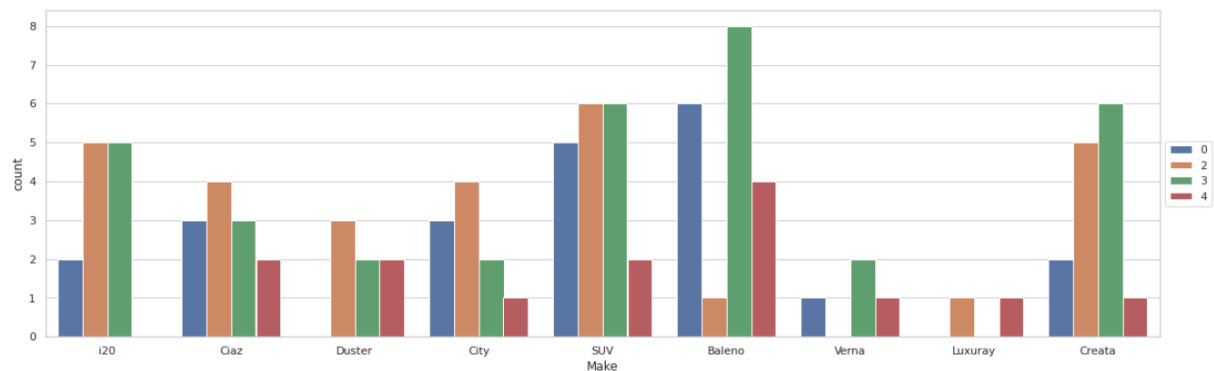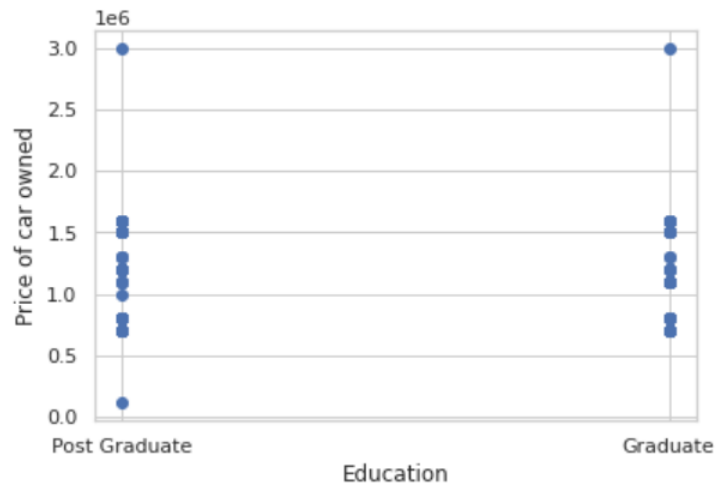

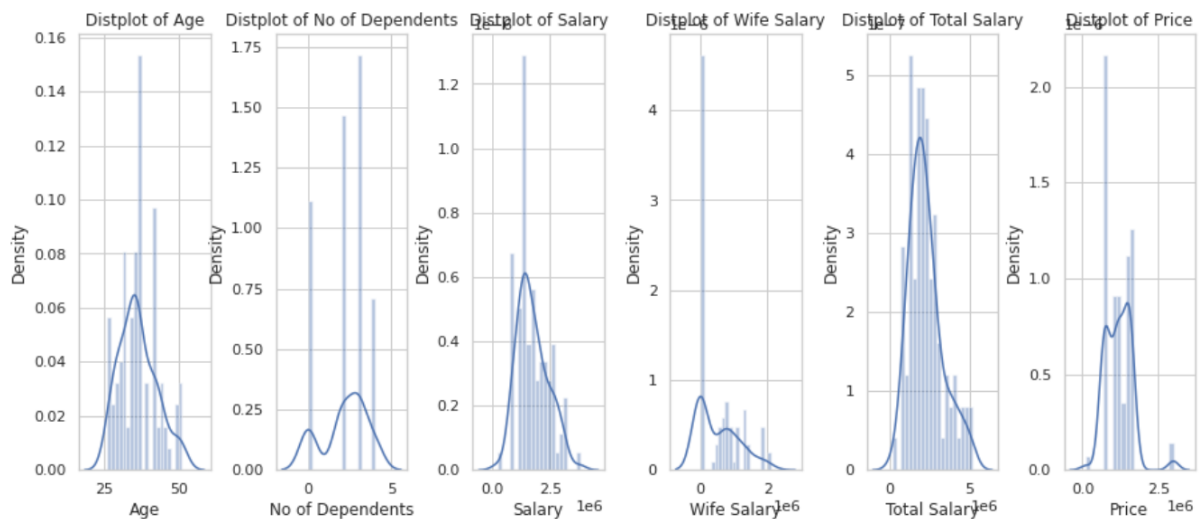The people with a family tend to buy SUVs and Sedans.
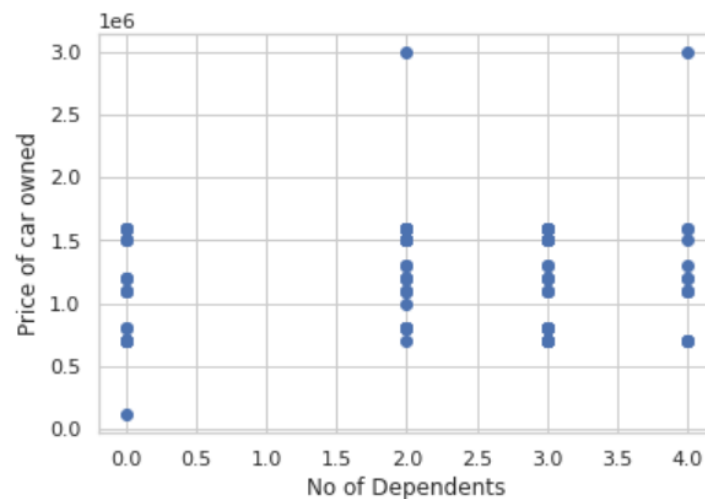
The relationship between consumers education and the vehicles purchased based on the make of vehicle owned. The people who are post graduate are more likely to buy a bigger car.
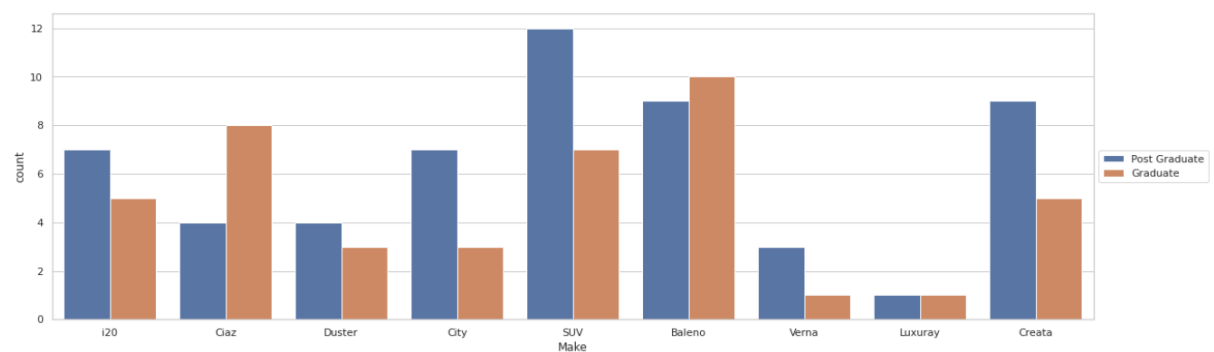
But it does not affect based on price of the vehicle. As both the groups are not affected by it.

Demographic Analysis

- People belonging to the age group of 25 to 50 make up most of the consumer market.

- Most people having an average total salary of around 30 lakhs tend to purchase more.

- Most people spent 10 to 20 lakh on buying vehicles.

2. EV STATS.csv

The dataset contains the state wise count on the variables such as

- Two Wheelers (Category L1 & L2 as per Central Motor Vehicles Rules
- Two Wheelers (Category L2 (CMVR))
- Two Wheelers (Max power not exceeding 250 Watts)
- Three Wheelers (Category L5 slow speed as per CMVR)

- Three Wheelers (Category L5 as per CMVR)
- Passenger Cars (Category M1 as per CMVR)
- Buses
- Total in state

Geographic Analysis
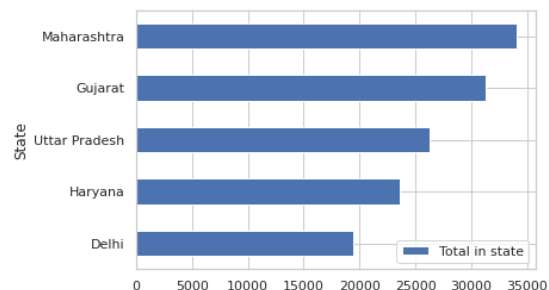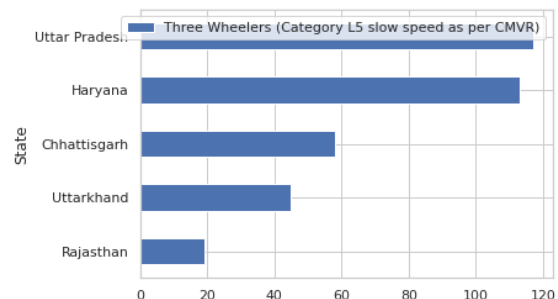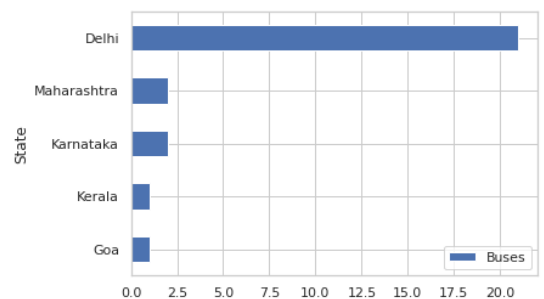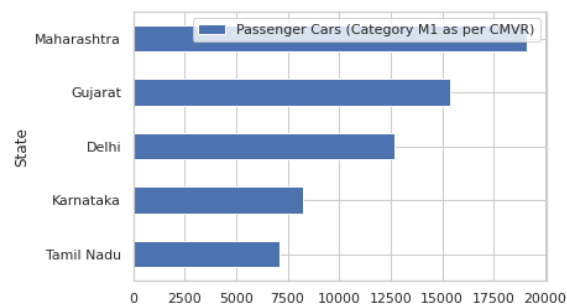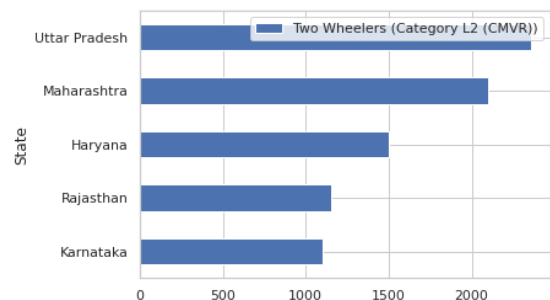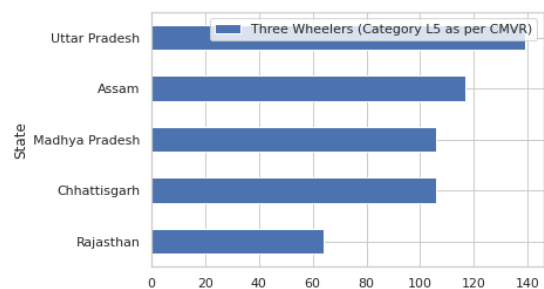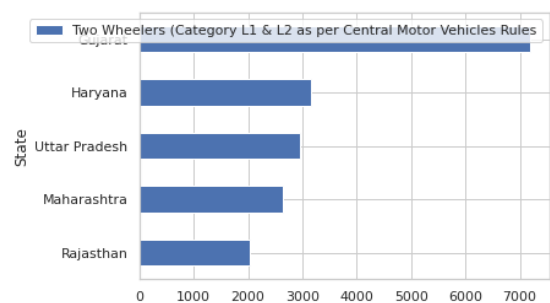
Based on the type of electric vehicle, specific states with a higher concentration of specific vehicles can be targeted for greater sales.

Model Building

The machine learning method that we are going to use is K-Means Clustering. It is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the K-means clustering algorithm are:

1. The centroids of the K clusters, which can be used to label new data.

2. Labels for the training data (each data point is assigned to a single

cluster) The 'means' in the K-means refers to averaging of the data; that is, finding the centroid.

But first we need to perform label encoding to convert text into numerical values to feed into our model.

|	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Price
0	27	0	0	1	0	1	0	0	800000	0	800000	800000
1	35	0	1	1	2	1	1	1	1400000	600000	2000000	1000000
2	45	1	1	0	4	1	1	0	1800000	0	1800000	1200000
3	41	1	1	1	3	0	0	1	1600000	600000	2200000	1200000
4	31	0	1	1	2	1	0	1	1800000	800000	2600000	1600000

Here is the Elbow method plot. It helps in identifying optimal clusters, identifying trade-off between Complexity and Fit and making data driven decisions.
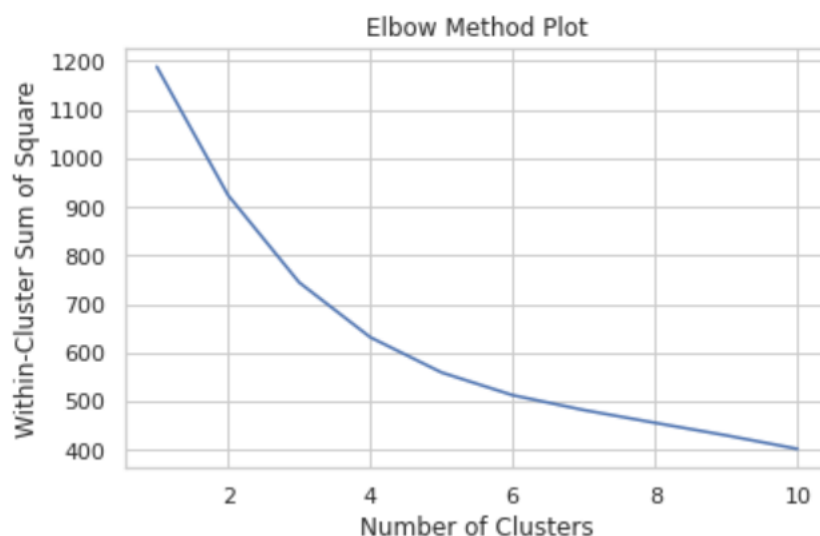
Elbow Method Plot

For K=3
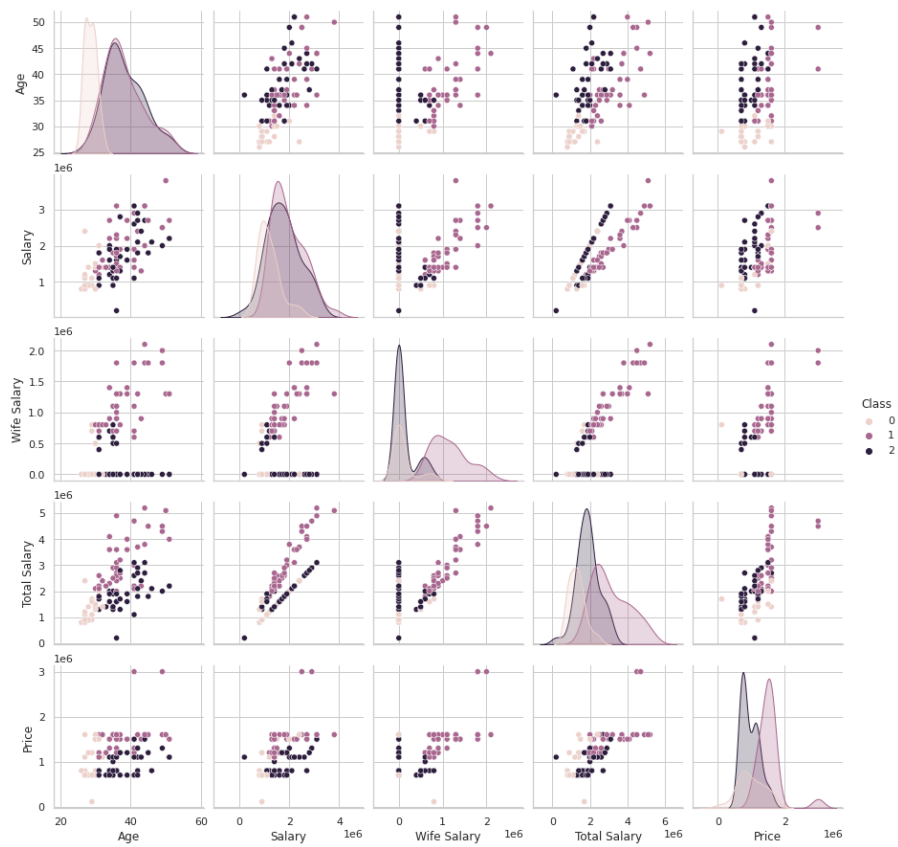
For K = 5
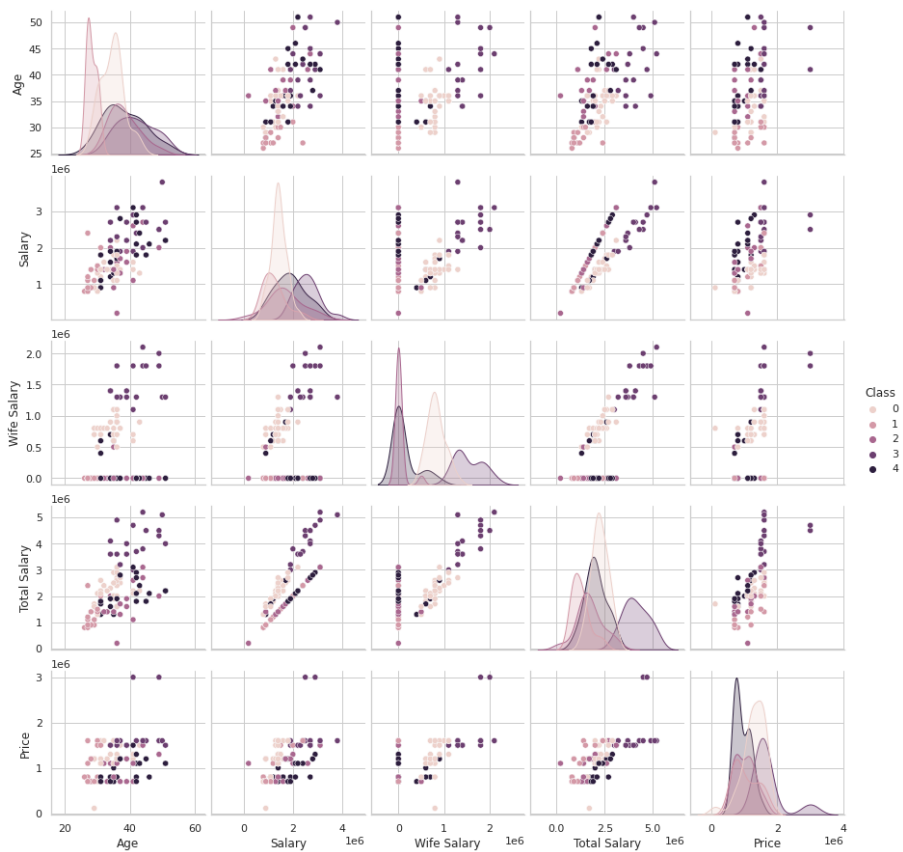
Firstly, we are trying to find the optimal K value using the Elbow Method wherein we will be finding the Within Cluster Sum of Square (WCSS) and try to find the point where it rapidly decreases which makes the graph look like an "elbow" there.

The K value corresponding to that point is the optimal K value. After looking into the plot, we can find that there are 2 points at which elbows are formed (which can be seen with the slight bent at K=3 and K=5). After finding the possible optimal K value, we will try to find the K value which provides us with the right clustering. Therefore, we will try to train K-Means Clustering by taking K = 3 and K = 5.

From the behavioral, geographic and demographic analysis, we were able to see some attributes having an impact in the way of grouping consumers. However, while looking into the clustered dataset, we were able to find 5 attributes that contributed most to the clustering (i.e., Age, Salary, Wife Salary, Total Salary, and Price). This can be seen in the pair plot done for both K=3 and K=5 conditions. In this case, we are able to see the dataset being clustered naturally. From here we are also able to see that the model is trying to cluster people on the basis of their total income were:

● Class 1 are the group of people who have Total Salary close to Salary (husband's salary)

● Class 2 is the group of people who have Total Salary higher than Salary (husband's salary)

● Class 0 are the group of people who have Total Salary close to Salary (husband's salary) but their total salary is relatively less compared to other people.

In this case, we are able to see the dataset is being clustered into very small groups of people that the model recognizes as a trend which is however not the case. We don't want to lose the homogeneity between our segments, therefore, going with K=3 will give the best results in the clustering analysis that is being done.

Target Segment

Here are some insights from the above done market segmentation:

• The younger population is more likely to purchase products with new technology, especially Electric Vehicles as they are aware of the environmental benefits and would like to bring that change, but our report

showed that younger population buys less expensive vehicles and so Electric Vehicles not being affordable can be a downside.

- It is then suggested to target a segment which is still eager to try new technologies but financially well enough to be able to afford Electric Vehicles. These people are likely to be in an age-group of 30 to 40 years.
- People from urban cities with available infrastructure and education about technology and its benefits will tend to purchase electric vehicles more.
- People who are married and who have dependents are more likely to go ahead and purchase a vehicle and so they could be targeted.
- The average salary of people who buy vehicles is around 30 lakh and the most purchase for automobiles lies in the range 10-20 lakh and less for two-wheelers. These aspects need to be kept in mind too.

(*GitHub*)