Task 3
Report
EV Segmentation
Ashudeep Dubey

Geographic Segmentation

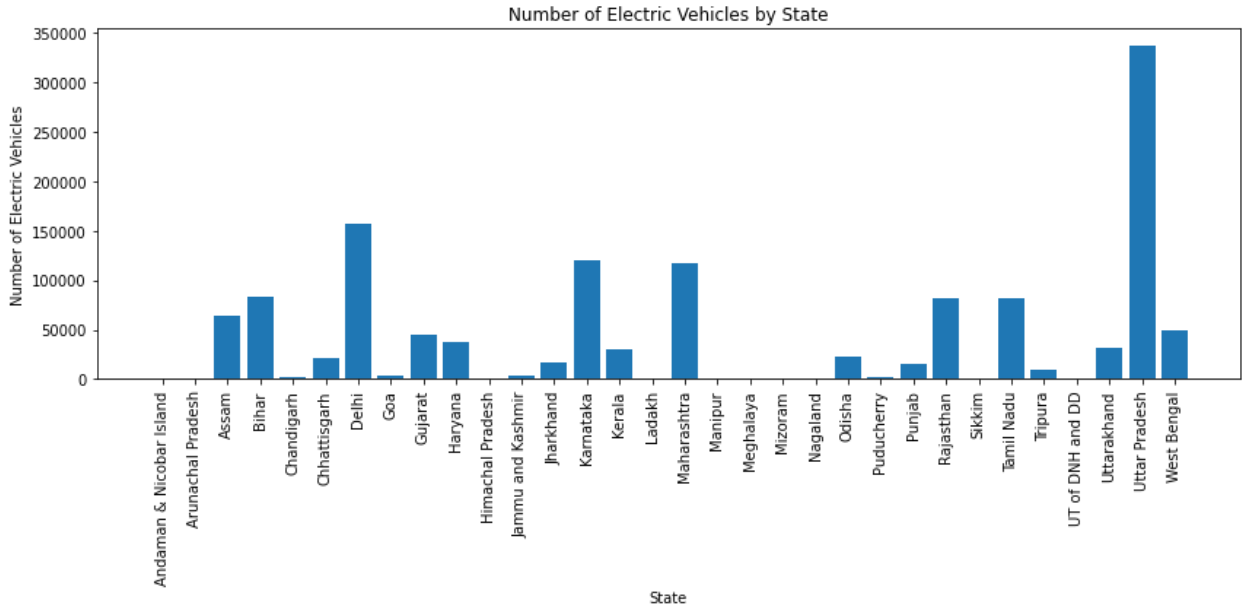
Pre-processing

The data contained NaN values for some states and union territories, which had to be removed before segmentation.

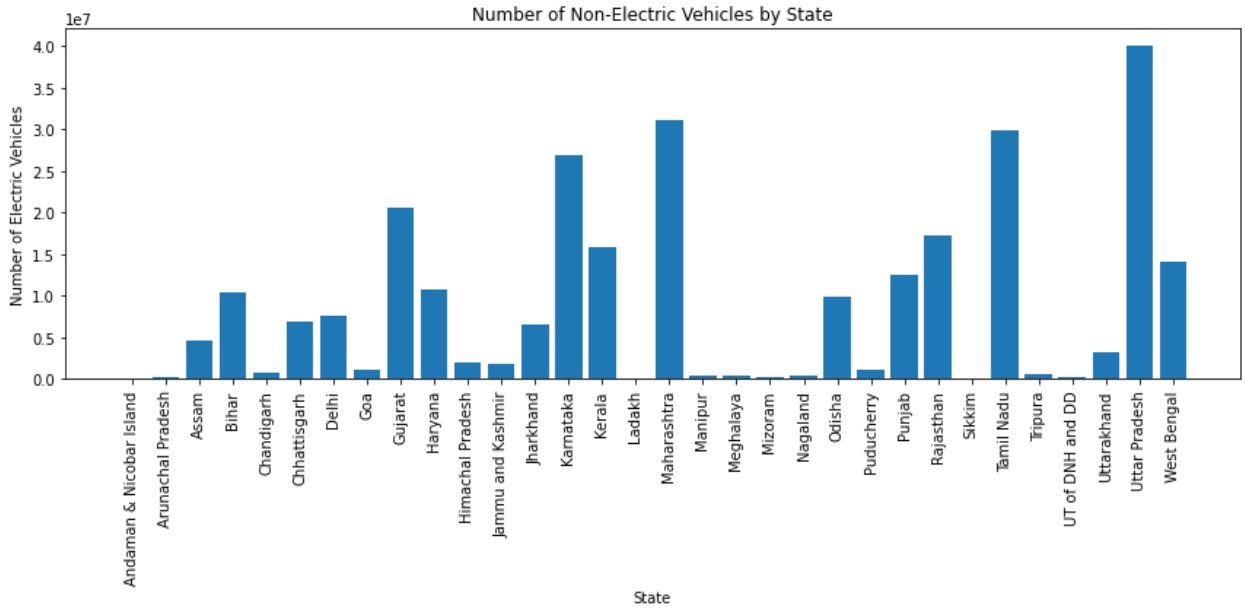
	State Name	Total Electric Vehicle	Total Non-Electric Vehicle	Total	Total Vehicles	Proportion Electric
0	Andaman & Nicobar Island	162	146945	147107	147107	0.001101
2	Arunachal Pradesh	20	252965	252985	252985	0.000079
3	Assam	64766	4677053	4741819	4741819	0.013658
4	Bihar	83335	10407078	10490413	10490413	0.007944
5	Chandigarh	2812	746881	749693	749693	0.003751
6	Chhattisgarh	20966	6836200	6857166	6857166	0.003058
7	Delhi	156393	7685600	7841993	7841993	0.019943
8	Goa	3870	1071570	1075440	1075440	0.003599
9	Gujarat	45272	20605484	20650756	20650756	0.002192
10	Haryana	37035	10778270	10815305	10815305	0.003424
11	Himachal Pradesh	1175	1964754	1965929	1965929	0.000598
12	Jammu and Kashmir	2941	1869962	1872903	1872903	0.001570
13	Jharkhand	16811	6486937	6503748	6503748	0.002585
14	Karnataka	120532	26870303	26990835	26990835	0.004466
15	Kerala	30775	15774078	15804853	15804853	0.001947
16	Ladakh	26	38302	38328	38328	0.000678
19	Maharashtra	116646	31058990	31175636	31175636	0.003742
20	Manipur	586	499324	499910	499910	0.001172
21	Meghalaya	49	459001	459050	459050	0.000107
22	Mizoram	21	315626	315647	315647	0.000067

Clean Dataset

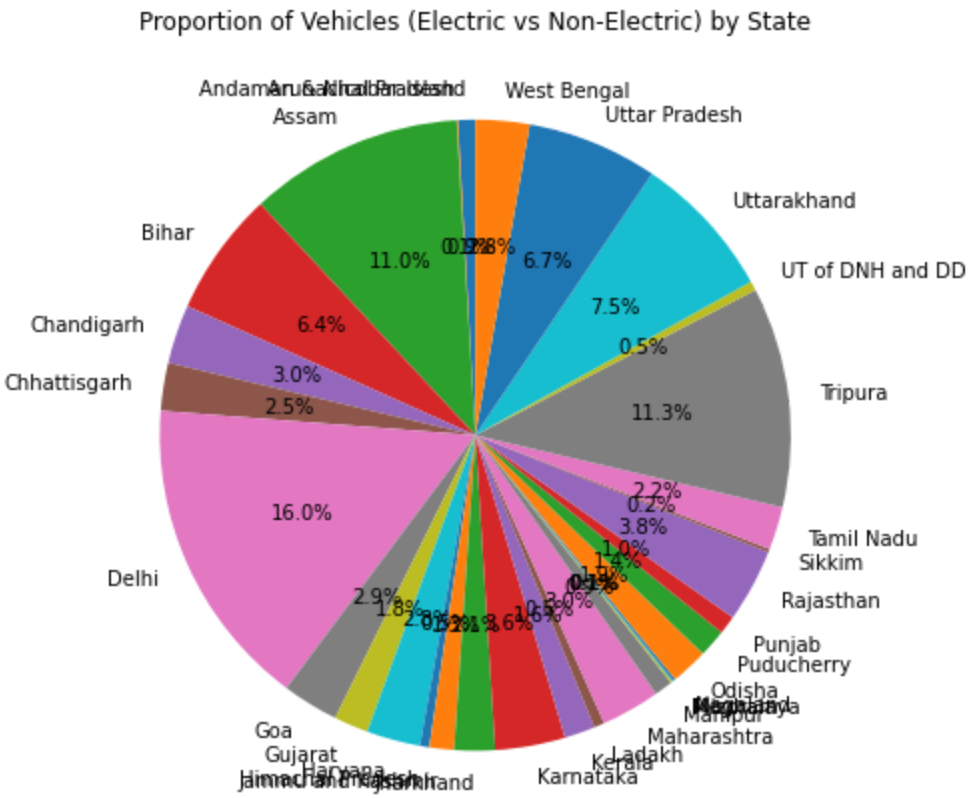
Segmentation



Number of EVs by State



Number of Non-Electric Vehicles by State



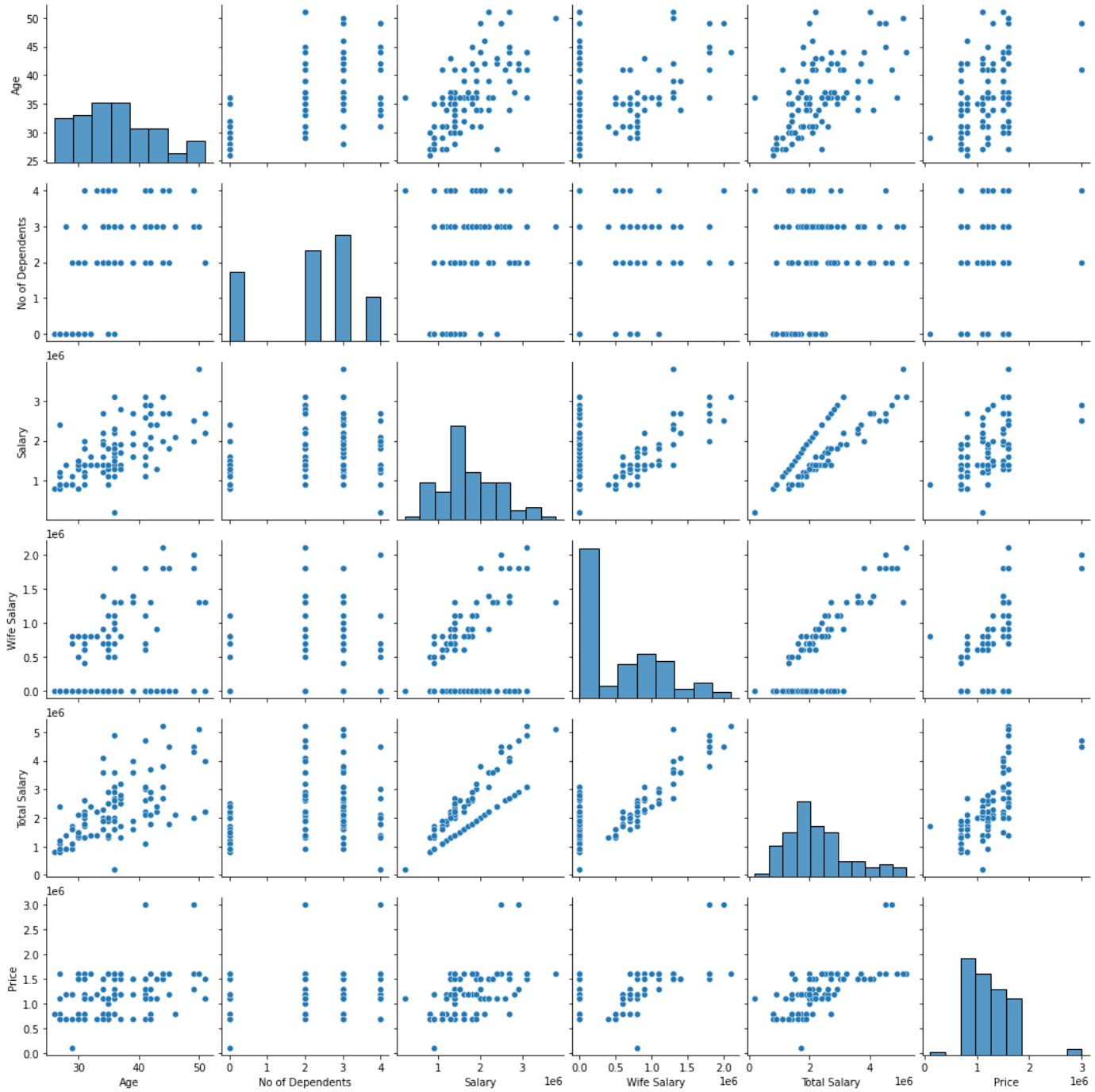
It is clear from the above statistics that many regions namely-
Andaman and Nicobar, Tamil Nadu, Sikkim, Rajasthan, Punjab, Puducherry, Odisha, Manipur, Maharashtra, Jammu and Kashmir, Jharkhand, and UT of DNH and DD have a deficient proportion of Electric Vehicles as compared to Non-Electric ones.

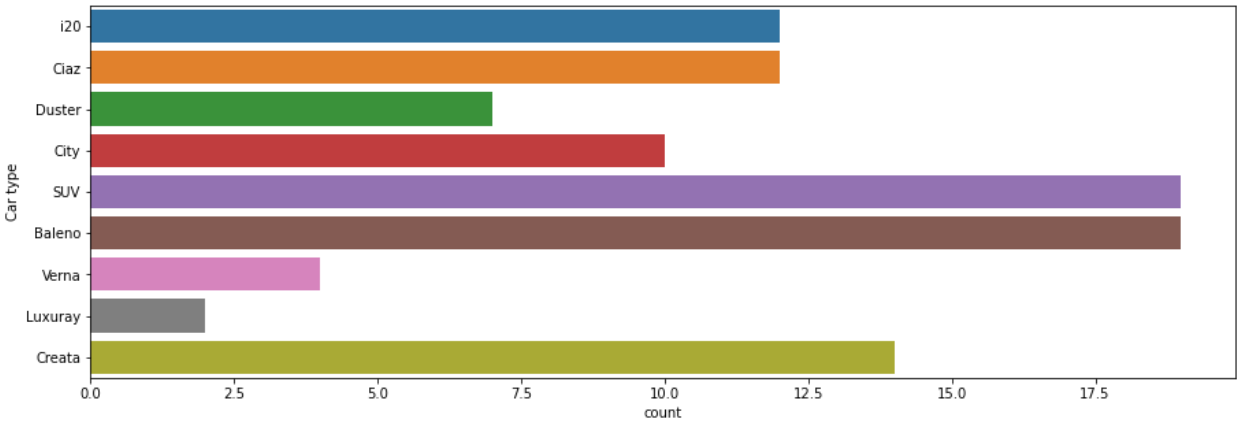
These states can be identified as an early market to set up EV factories and bring about their shift towards EV vehicles.

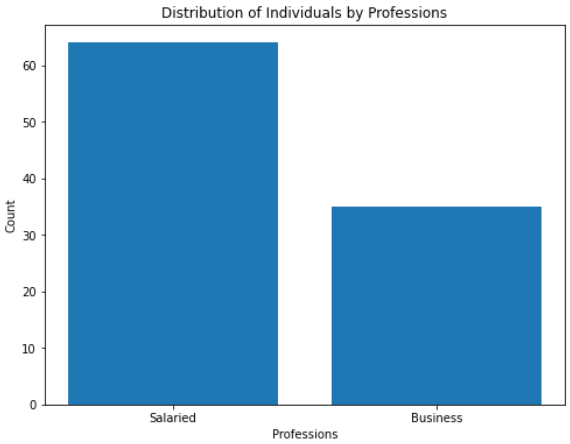
Consumer Market Segmentation

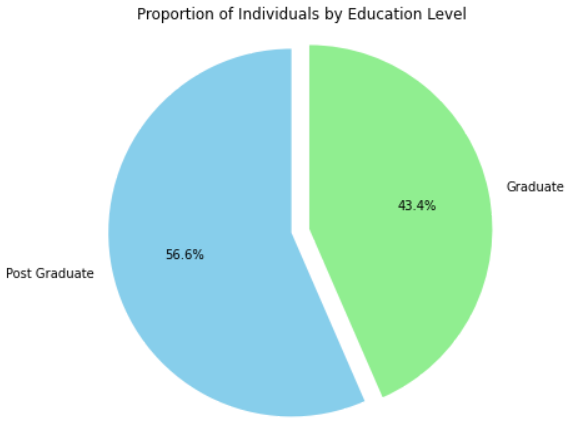
No pre-processing was required, and the data was clean.

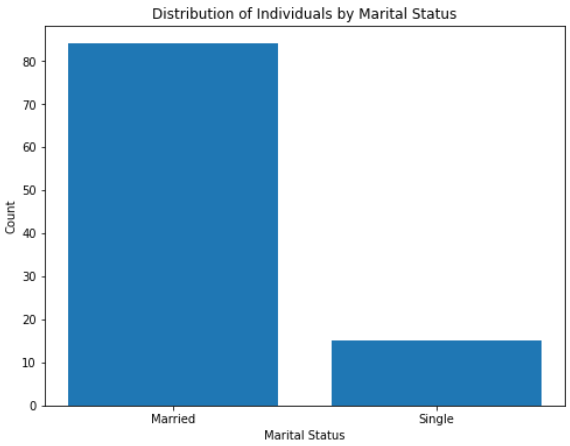
Stats



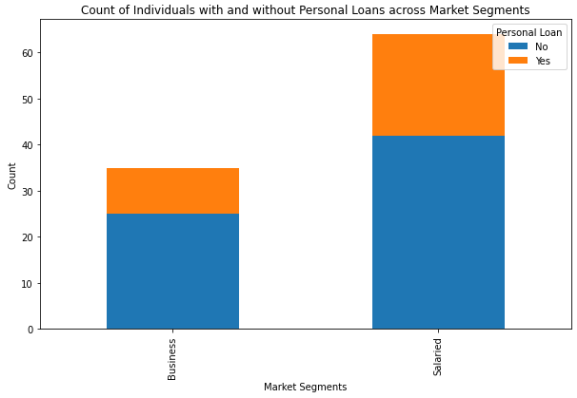

i20
Ciaz
Duster
City
SUV
Baleno
Verna
Luxuray
Creata
Car type
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
count

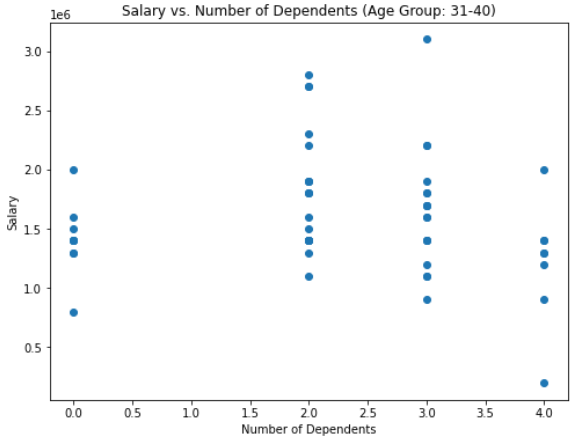

Distribution of Individuals by Professions
Count
Salaried
Business
Professions

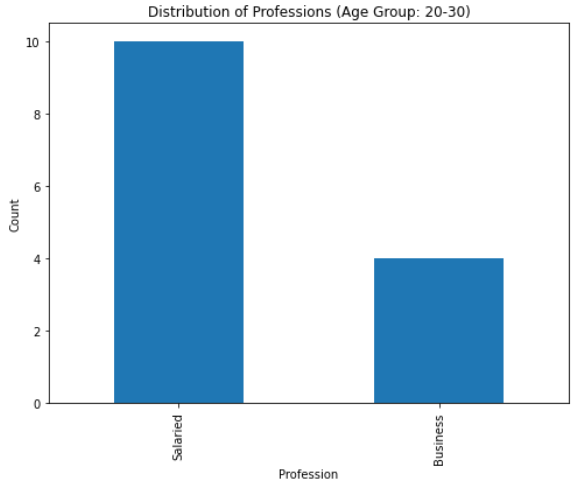

Proportion of Individuals by Education Level
Graduate
43.4%
56.6%
Post Graduate

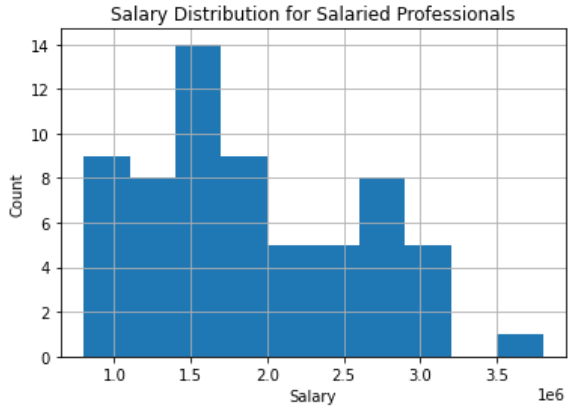

Distribution of Individuals by Marital Status
Count
Married
Single
Marital Status

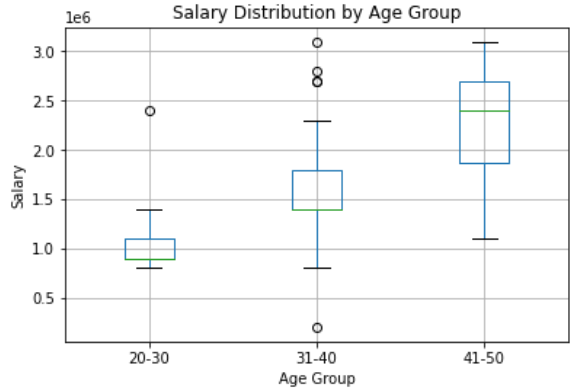

Relationship between Salary and Age
1e6
Salary
Age


Count of Individuals with and without Personal Loans across Market Segments
Personal Loan
No
Yes
Count
Business
Salaried
Market Segments

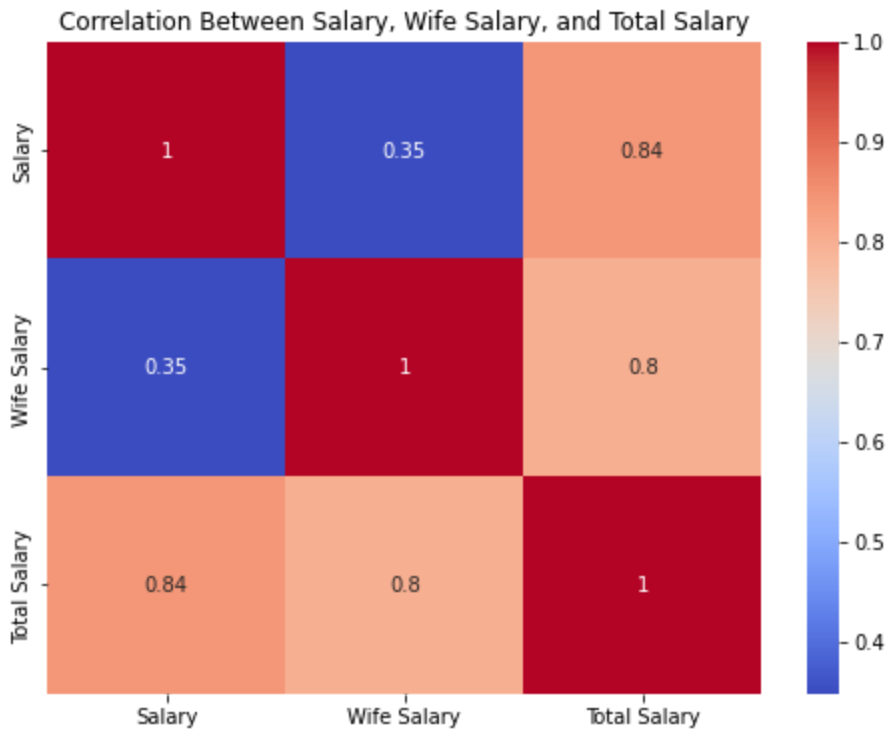

Salary vs. Number of Dependents (Age Group: 31-40)
1e6
Salary
Number of Dependents


Distribution of Professions (Age Group: 20-30)
Count
Salaried
Business
Profession


Salary Distribution for Salaried Professionals
Count
Salary
1e6


Salary Distribution by Age Group
1e6
Salary
20-30
31-40
41-50
Age Group

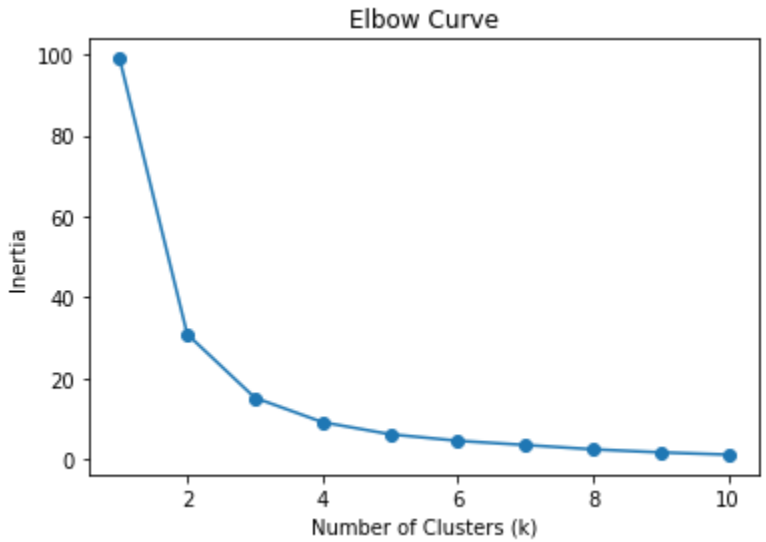

Total Salary Comparison: Salaried vs. Business Professionals
1e6
Total Salary
Salaried
Business
Profession



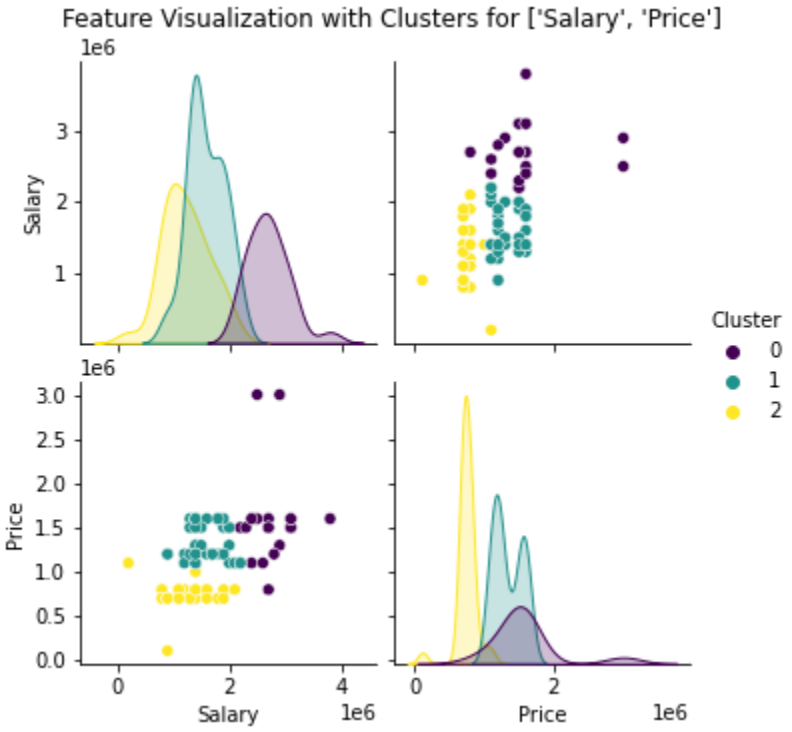
Most car owners are-
Married, Aged - 41-50, with most of them **Not** taking **Personal Loans** and having a decent **Wife Salary** also.

Segmentation

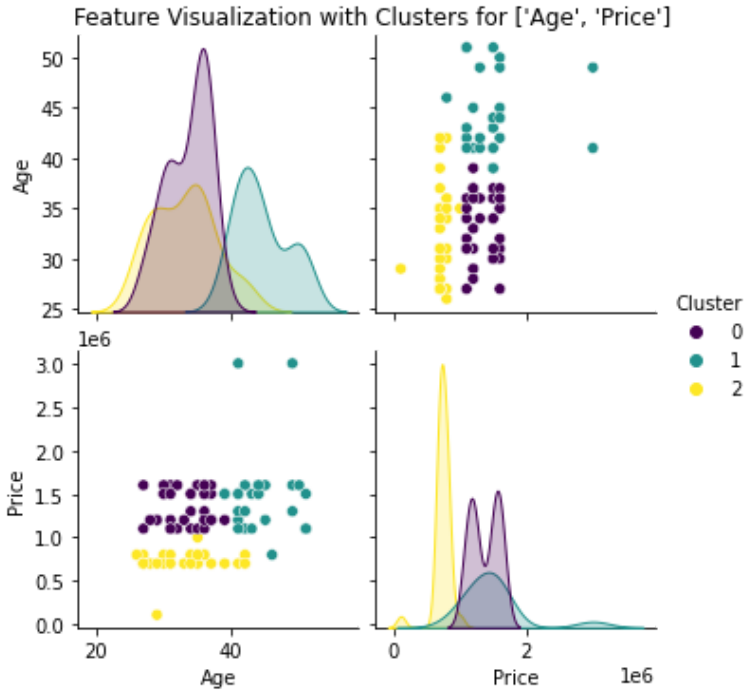
We perform K-Means on the given dataset and identify the best cluster number



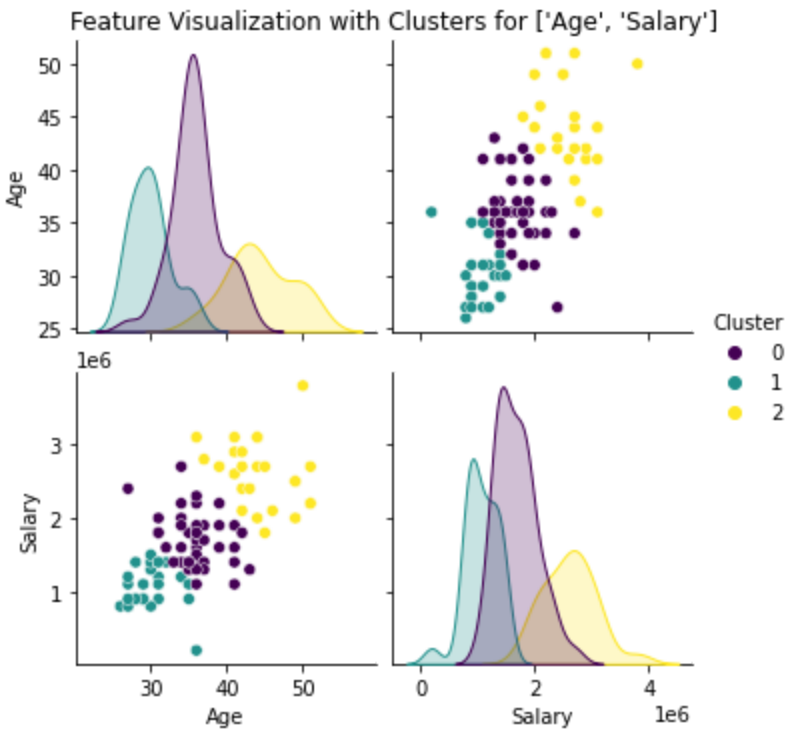
As apparent best segmentation would have 3 groups of customers (clusters)



We can observe that as the salary of individuals increases so does the Price of their cars which is obvious

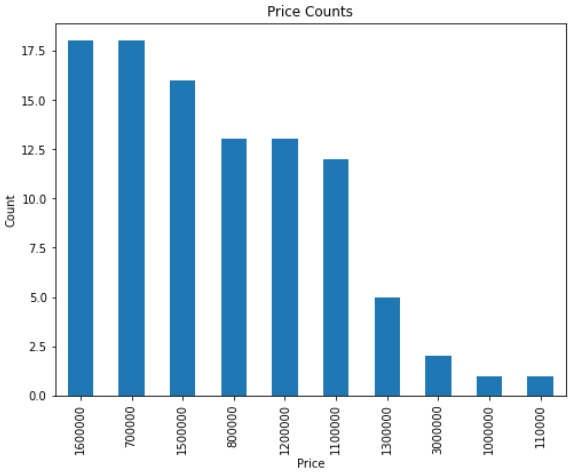


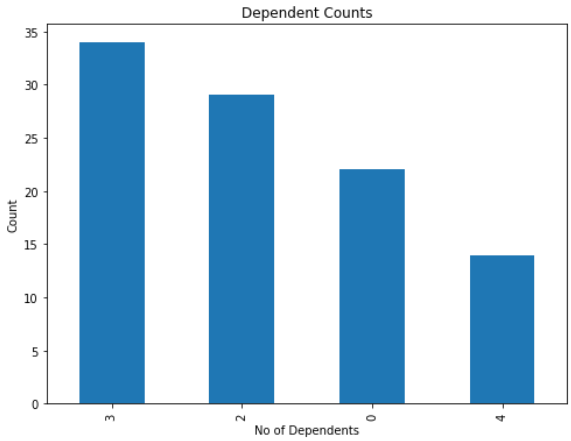
The Price of cars owned shows little or no difference with Age



The Salary of each segment also increases with Age

Therefore it can be observed that a majority of the population (in the dataset) have an affordable income to buy a vehicle





People are not hesitant about the price when it comes to vehicles as many have a vehicle costing 1600000.
The best range to price the new EVs would be 700000-1300000 as apparent from the limited info of the dataset.
Also many have No of Dependents at least 2 which implies family.

The best Age group to target will be 31-40 as these are the people with the most stable income according to the stats and most of them are married therefore the needs of a family must be kept in mind.

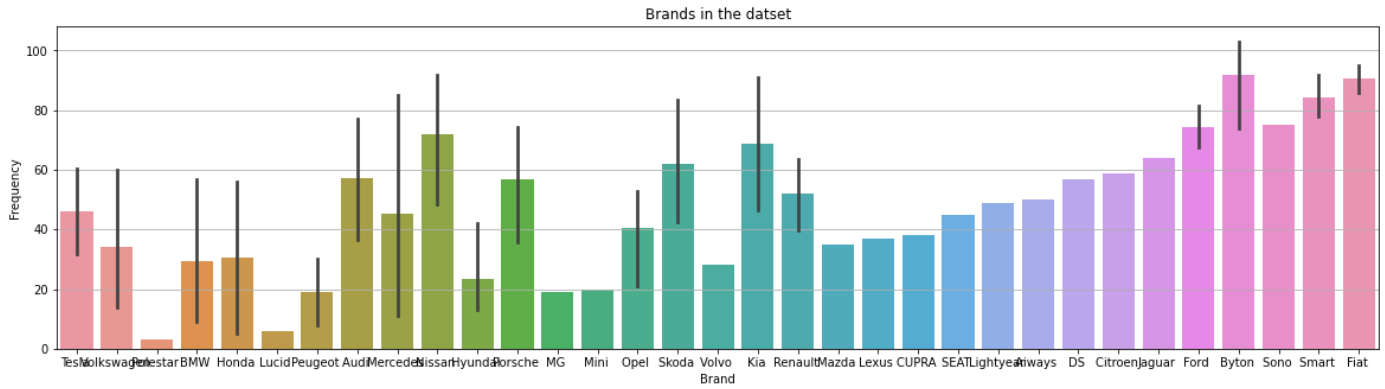
Based on their needs EVs can be:
Mid-sized and compact
Versatile and Practical EVs

These are the broad categories of EVs a company can focus on.

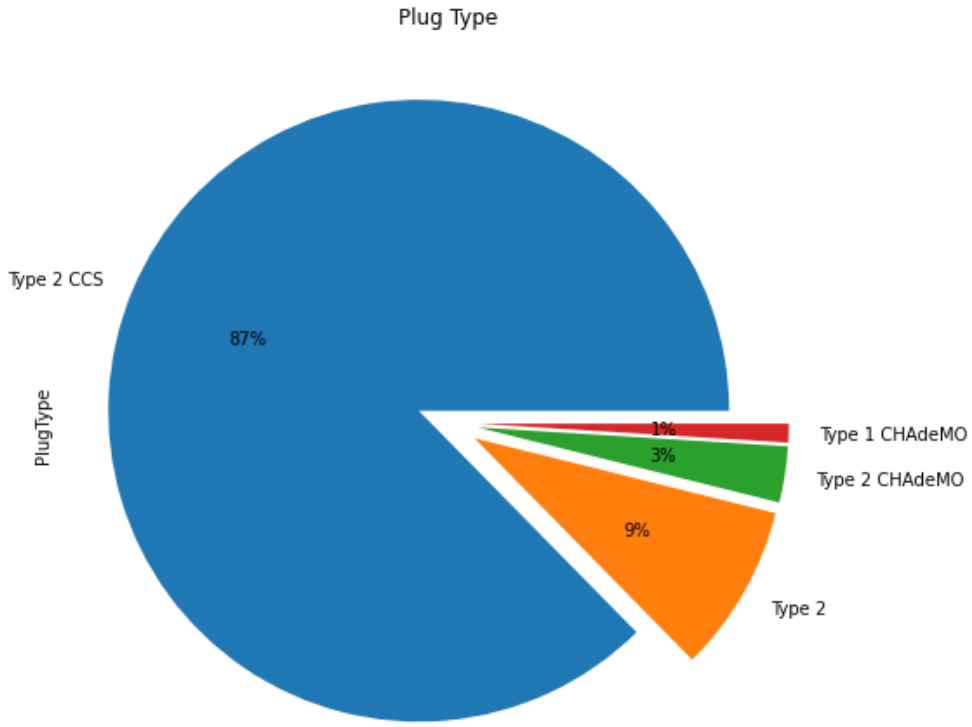
Vehicle Segmentation

No pre-processing was required

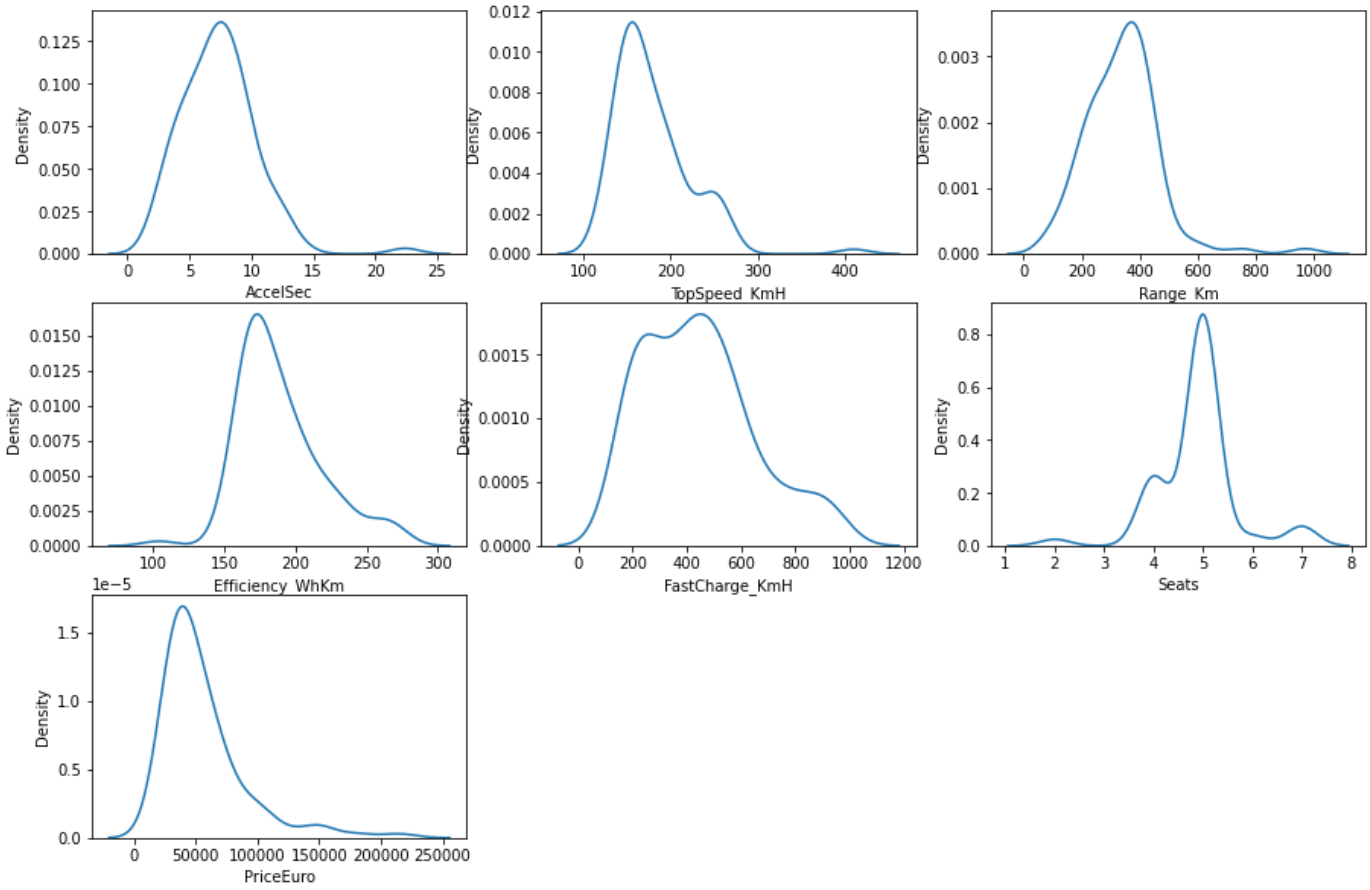
Stats



Most famous EV brands



Plug Type of EVs

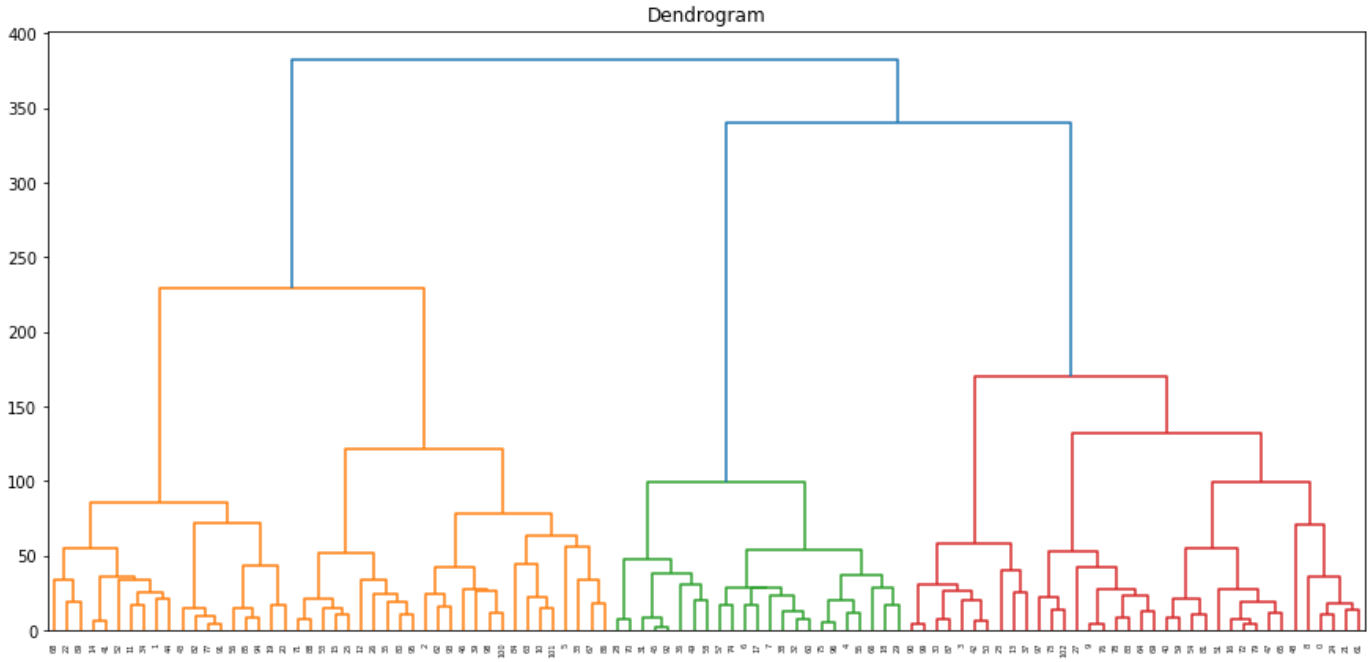


Some of the key observations are-

- **8-10 km/h^2** AccelSec has the most density meaning most of the cars have this acceleration range
- Most cars have a top speed of around **150-160 km/hr**
- **Plug Type - Type 2 CSS**
- Most of them are **5** seaters
- Most have a range of around **400km**
- The price of most EVs is around **50000 Euros**

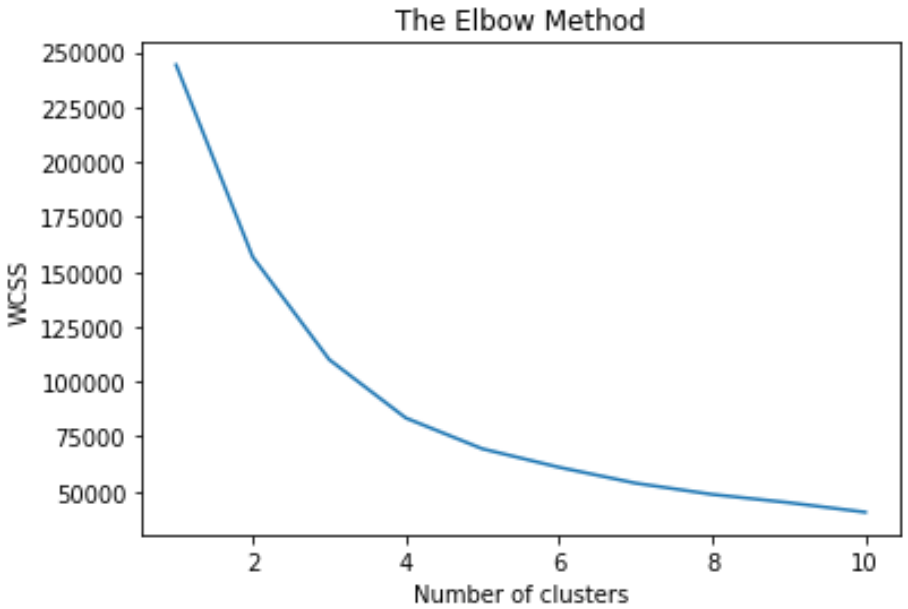
Segmentation

a) Hierarchical



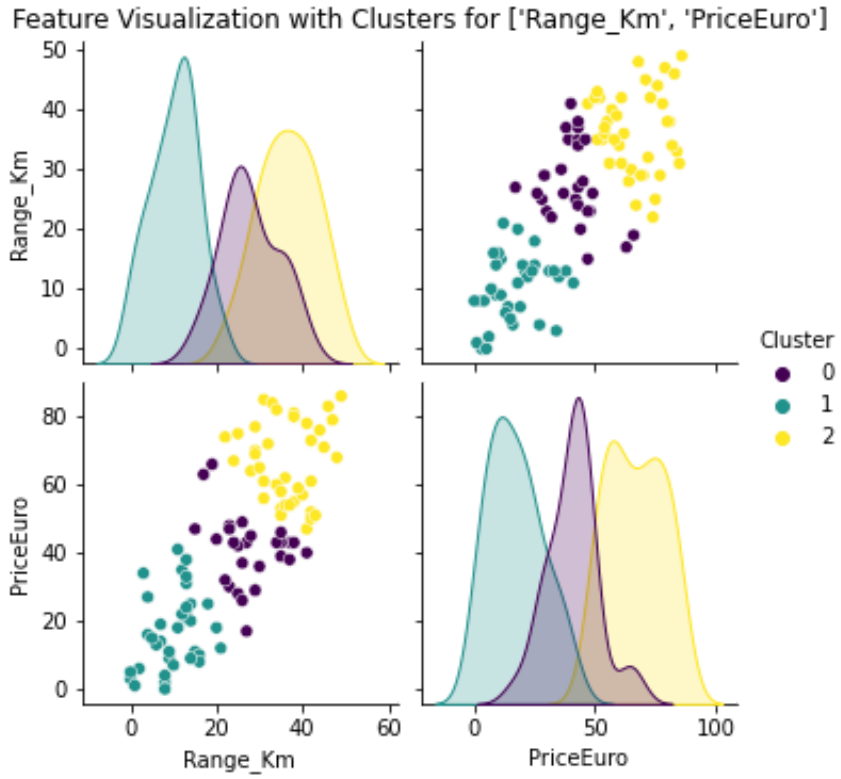
It is evident that the above graph suggests that best clusters = 3 for this dataset.

b) Elbow Method

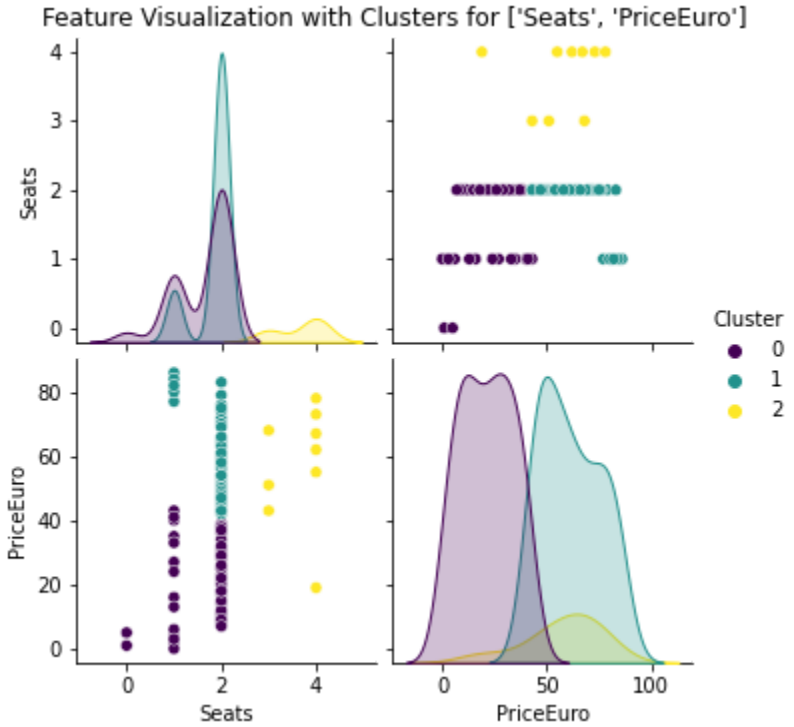


Elbow graph suggests that clusters =3 is the best segmentation

We perform K-Means again with clusters=3



There is a direct correlation between Price and Range



More Seats equals more Price

We perform a special K-Means clustering of k=7 for Segments vs Price as there are 7 segments namely A, B, C, D, E, F, N, and S

The significance is as follows-

Segment A: This segment typically includes the smallest and most compact EVs. These vehicles are designed for urban commuting and short-distance travel. They are often characterized by their small size, efficient performance, and agility.

Segment B: Vehicles in this segment are slightly larger than Segment A and offer more interior space. They are still compact and suitable for urban driving but may provide additional features and amenities compared to Segment A EVs.

Segment C: This segment represents mid-size EVs that offer a balance between compactness and interior space. Segment C EVs typically provide more room for passengers and cargo compared to Segments A and B. They are suitable for longer drives and may have a wider range of features and technologies.

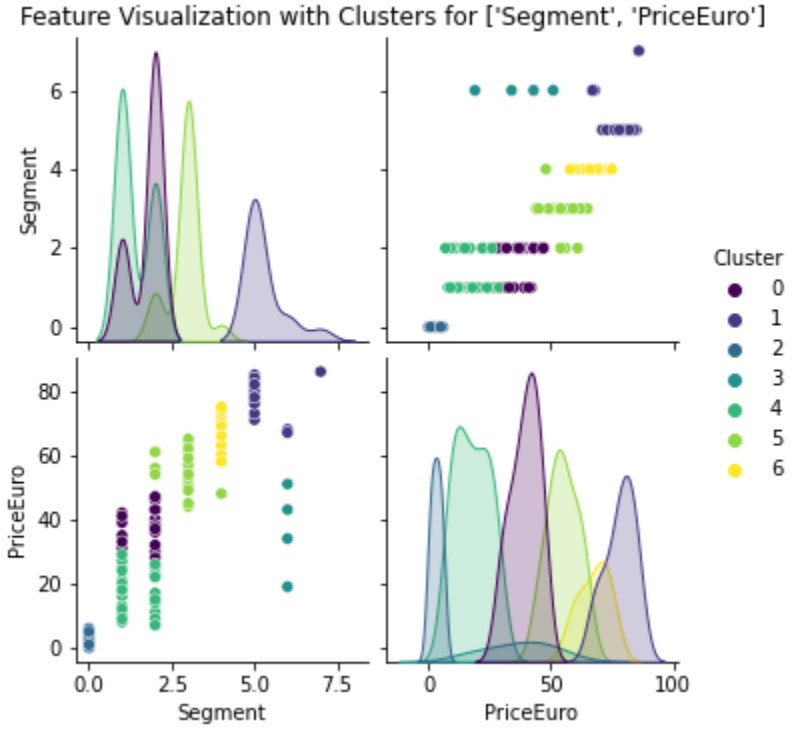
Segment D: Vehicles in this segment are larger and more spacious than those in the previous segments. They offer ample seating and cargo space, making them suitable for families or individuals requiring greater comfort and practicality. Segment D EVs often come with advanced features and technologies.

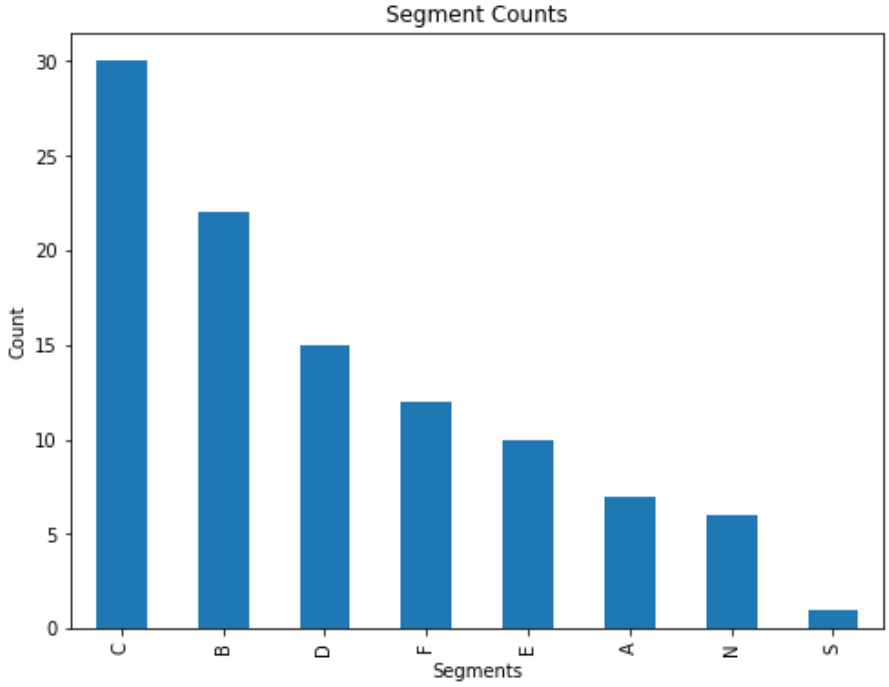
Segment E: This segment includes larger EVs that offer even more interior space and luxury features. They are designed for those seeking high levels of comfort, performance, and premium amenities. Segment E EVs are often associated with luxury electric vehicles.

Segment F: This segment represents electric SUVs (Sports Utility Vehicles) and crossovers. These EVs are known for their spaciousness, versatility, and off-road capabilities. Segment F EVs cater to individuals or families seeking utility and versatility in their electric vehicles.

Segment N: This segment typically refers to niche or specialty EVs that don't fit into the standard segment categorizations. It may include unique or specialized EV models designed for specific purposes or market niches.

Segment S: This segment is often used to represent high-end luxury electric vehicles or EVs with exceptional performance and cutting-edge technologies. These EVs often offer top-tier features, craftsmanship, and advanced driving capabilities.


Feature Visualization with Clusters for ['Segment', 'PriceEuro']
Segment
PriceEuro
Cluster
0
1
2
3
4
5
6
0.0
2.5
5.0
7.5
0
50
100
0
2
4
6
0
20
40
60
80


Segment Counts
Count
Segments
C
B
D
F
E
A
N
S
0
5
10
15
20
25
30

It can be inferred from the above analysis that the best market scope of EVs is-

- **Segment C or D**
- **Price around 50000-60000 Euros**
- **Minimum 4 seater**
- **Range around 400km**
- **Plug type preferred is Type 2 CSS**

These can be identified as the key features a new company willing to enter the EV market should keep in mind.

Report Analysis

ML Model and Algorithm Used

The most used model for this segmentation was K-Means as it helps to identify clusters from a given data which are key in market segmentation.

K-means clustering is a popular choice for market segmentation due to its simplicity, scalability, and interpretability. As an unsupervised learning algorithm, it can automatically group similar data points into distinct clusters based on their characteristics. With its flexibility in defining the number of clusters, K-means allows researchers to experiment and identify the optimal segmentation. The algorithm's efficiency makes it suitable for large datasets, and the resulting cluster centroids provide insights into segment characteristics. While K-means has advantages, its assumptions and sensitivity to initial centroids should be considered. Overall, K-means provides a practical and well-established approach for market segmentation analysis.

Final Conclusion

After conducting the research and analysis work on the EV market segmentation, several key insights and conclusions have been derived. These findings provide valuable information for understanding the market dynamics and consumer preferences within the EV industry.

Market Segmentation: Through the analysis, we have identified distinct market segments within the EV market based on various factors such as vehicle size, features, and target audience. The commonly observed segments include Segment A, Segment B, Segment C, Segment D, Segment E, Segment F, Segment N, and Segment S. Each segment caters to specific customer needs and preferences, offering different levels of compactness, luxury, performance, and utility.

Customer Preferences: By examining the characteristics of each segment, we gained insights into the preferences of EV consumers. It became apparent that customers have diverse requirements ranging from compact and efficient city commuting (Segment A and B) to larger, more spacious family vehicles (Segment D and E) and even specialized niche EVs (Segment F and N). Understanding these preferences helps automakers and marketers tailor their products and marketing strategies to target specific segments effectively.

Market Opportunities: The analysis has revealed potential market opportunities within the EV industry. For example, the growing demand for compact and affordable EVs (Segments A and B) indicates a need for cost-effective, urban-friendly electric vehicles with efficient battery ranges. On the other hand, the emergence of luxury EVs (Segment E and S) highlights the demand for high-end electric vehicles that offer advanced features, cutting-edge technology, and premium driving experiences.

The final conclusion for the EV market can be summarized in points as

- **Mid-sized and compact (Segment C), Versatile and Practical EVs(Segment D**
- **Price around 50000-60000 Euros**
- **Minimum 4 seater**
- **Range around 400km**
- **Plug type preferred is Type 2 CSS**

Targeted Age group should be 31-40
Targeted regions can be Andaman and Nicobar, Tamil Nadu, Sikkim, Rajasthan, Punjab, Puducherry, Odisha, Manipur, Maharashtra, Jammu and Kashmir, Jharkhand, and UT of DNH and DD

It's obvious that not all of these demands cannot be met by a single type of EV. The above segment is what the largest market is comprised of according to the given datasets and adjustments can be made accordingly.

Room for improvement
Given additional sources/budget a good dataset is required with columns

- Owners (Male/Female) (as these can affect the choice of vehicle)
- buying preferences,
- ownership duration
- distance traveled
- Charging patterns
- driving habits
- Data on EV market players, including manufacturers, dealerships, and service providers, can provide insights into market dynamics, competitive positioning, and market share.

Although the given datasets were good they lacked vastness and need to be more elaborative. Therefore a more elaborative dataset with more entries is required

Other ML models (given time) that can be used can surely be SVMs, Decision Trees, and Neural Networks. These can boost the segmentation process and provide a more detailed insight as well.

Market Size
Estimated Market from these given datasets can be taken to be around **42,000,000 according to the states that can be targeted** although it's a very rough estimate.

Best Variables
4 Variables/features which can be used to create the most optimal Market Segments can be:

- **Existing Price of EVs**
- **Segment Type**
- **Number of EVs by state**
- **Income of Buyers**

GitHub Link
https://github.com/AshDDftw/EV_MarketSegmentation