ELECTRIC VEHICLE MARKET SEGMENTATION ANALYSIS

Sibsankar Maity

Github Link: https://github.com/sibsankar-4/EV-MARKET-SEGMENTATION

1 Abstract

This project provides a detailed look at India's electric vehicle (EV) market, focusing on electric two-wheelers and income data. It highlights the strong growth of India's two-wheeler market, driven by rising environmental concerns, government incentives, and improvements in battery technology. Using behavioral data, we performed a market segmentation analysis with the k-means algorithm. This analysis identifies different market segments within the electric two-wheeler sector, offering insights into consumer preferences, purchasing power, and market opportunities. The study aims to help stakeholders, such as manufacturers and investors, with targeted marketing strategies and to support the overall development of India's EV ecosystem.

2 Introduction

The global electric vehicle (EV) market has grown rapidly in recent years, driven by advancements in technology, environmental concerns, and government initiatives for sustainable transportation. India, too, is becoming an important player in the EV sector, with many companies adopting this new technology. This article compares the global EV market with India's market, focusing on key trends, challenges, and the leading Indian companies in the EV space.

Around the world, EV adoption is accelerating. Major car manufacturers are investing heavily in electric vehicles. Countries like China, the United States, and several European nations have seen significant increases in EV sales. This growth is supported by government policies, better charging infrastructure, and a variety of electric models. The global push for EVs has led to advancements in battery technology and innovations in autonomous driving and vehicle connectivity.

In India, the EV market is just beginning but shows great promise. The Indian government has introduced several initiatives to promote EVs, including tax incentives, subsidies, and nationwide charging infrastructure. Many Indian companies, from large automotive manufacturers to startups, are producing and promoting electric vehicles, contributing to the sector's growth.

In the next sections, we will look more closely at the global and Indian EV markets, exploring key players, market dynamics, and future prospects for electric mobility in both areas.

3 Problem Statement

The rapid growth of the Indian electric vehicle (EV) market necessitates a comprehensive analysis of EV segmentation, customer segmentation, and regional growth patterns. Key challenges to address:

- **Market Segmentation:** Refine and validate the current EV market segmentation based on K-means clustering. Identify additional influential factors such as driving range, charging infrastructure, vehicle types (sedans, SUVs, etc.), and specific features.

- **Customer Segmentation:**

  Conduct further analysis to understand the demographics, psychographics, and needs of potential EV buyers. Examine factors like income levels, environmental awareness, commuting patterns, and consumer preferences to develop targeted marketing strategies.

- **Regional Growth Patterns:**

  Analyze the variation in EV growth across Indian states. Identify factors driving or hindering EV adoption, such as state-level policies, infrastructure development, incentives, and consumer awareness.

By addressing these challenges, we can provide valuable insights and recommendations for automotive manufacturers, policymakers, and infrastructure providers. This will help accelerate EV adoption and growth in India, leading to a more sustainable and eco-friendly transportation landscape.

4 Data Collection and Preprocessing

Both the Demographic dataset and EV bikes dataset are collected from https://www.kaggle.com/. The data collected is compact and is used for visualization purposes and for clustering.

Features: The EV bikes dataset used in this project contain 53 rows and 11 features. The features are as follows: 'Unnamed: 0', 'Model', 'Manufacturer', 'Vehicle Type','Battery Capacity (kWh)', 'Range per Charge (km)', 'Charging Time','Price', 'Power (HP or kW)', 'Top Speed (km/h)', 'Year of Manufacture'.
The demographic dataset contain 99 rows and 8 features. The features are 'Age', 'Profession', 'Marrital Status', 'Education', 'No of Dependents', 'Personal loan', 'Total Salary', 'Price'.

5 Behavioural Segmentation

In this section, we will outline the methods used to perform demographic segmentation on the dataset. Before diving into segmentation, a thorough Exploratory Data Analysis (EDA) is conducted. EDA is crucial as it helps reveal the underlying structure of the data, uncovering trends, patterns, and relationships that might not be immediately visible. This step is essential for companies as it provides insights that can inform better decision-making and strategy development.

5.1 Exploratory Data Analysis

The sample of the data set is
Before conducting the analysis, the datasets were preprocessed to ensure data quality and accuracy. This involved data cleaning, handling missing values, resolving inconsistencies, and transforming categorical variables into numerical representations if required. The summary of the dataset is it contain 4 numerical data columns and 4 categorical data columns.

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	Total Salary	Price
10	35	Salaried	Married	Post Graduate	4	No	2000000	1600000
98	51	Salaried	Married	Post Graduate	2	Yes	2200000	1100000
72	37	Salaried	Married	Graduate	2	No	1300000	700000
91	36	Salaried	Married	Post Graduate	3	Yes	4900000	1600000
1	35	Salaried	Married	Post Graduate	2	Yes	2000000	1000000

Figure 1: Sample of the Data

```
1 df.describe()
```

	Age	No of Dependents	Total Salary	Price
count	99.000000	99.000000	9.900000e+01	9.900000e+01
mean	36.313131	2.181818	2.270707e+06	1.194040e+06
std	6.246054	1.335265	1.050777e+06	4.376955e+05
min	26.000000	0.000000	2.000000e+05	1.100000e+05
25%	31.000000	2.000000	1.550000e+06	8.000000e+05
50%	36.000000	2.000000	2.100000e+06	1.200000e+06
75%	41.000000	3.000000	2.700000e+06	1.500000e+06
max	51.000000	4.000000	5.200000e+06	3.000000e+06

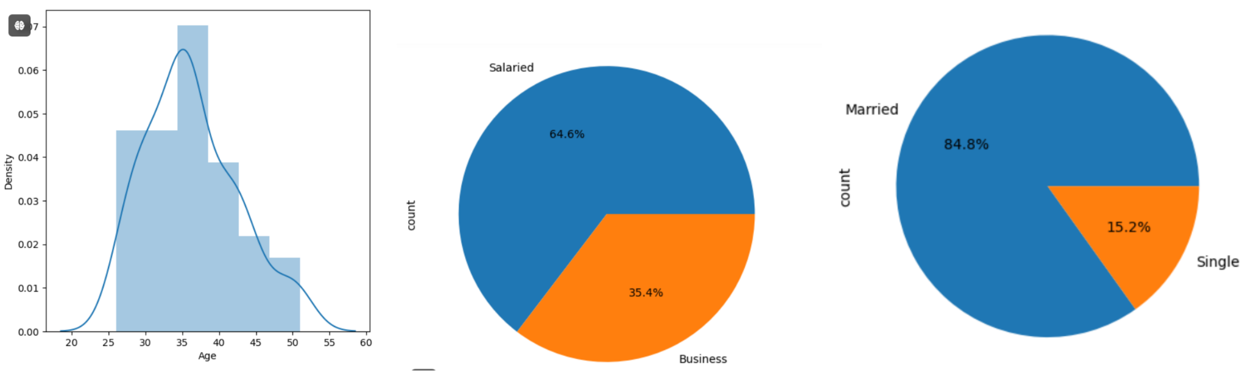
(a)

```
1 df.describe(include="object")
```

	Profession	Marrital Status	Education	Personal loan
count	99	99	99	99
unique	2	2	2	2
top	Salaried	Married	Post Graduate	No
freq	64	84	56	67

(b)

Figure 2: Summary of the data

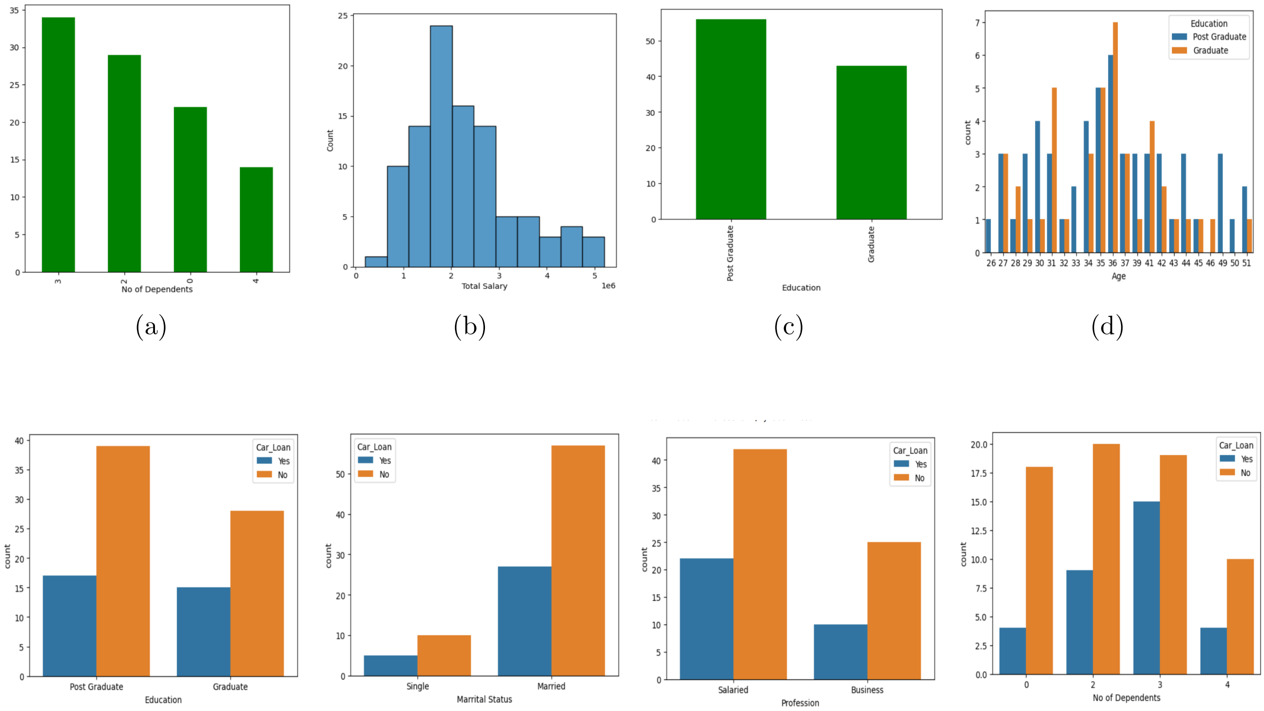


(a) (b) (c)

Figure 3: Sample Plot

5.1.1 Data Visualisation

By using various data visualization methods like pie-plot, distplot, countplot, histplot we visualize four important characteristics, namely age, profession and martial status, of the dataset.



(a) (b) (c) (d)

(e) (f) (g) (h)

5.2 PCA

Before clustering the data points, we present the correlation matrix to illustrate the interdependencies between features within the dataset. This step is essential for understanding the relationships among variables and ensuring an effective clustering process.

The data are preprocessed using the 'StandardScaler' class in Scikit-Learn. Following this, Principal Component Analysis (PCA) is applied to extract the independent components. The aim is to reduce the number of features while retaining most of the information, as indicated by the explained variance.

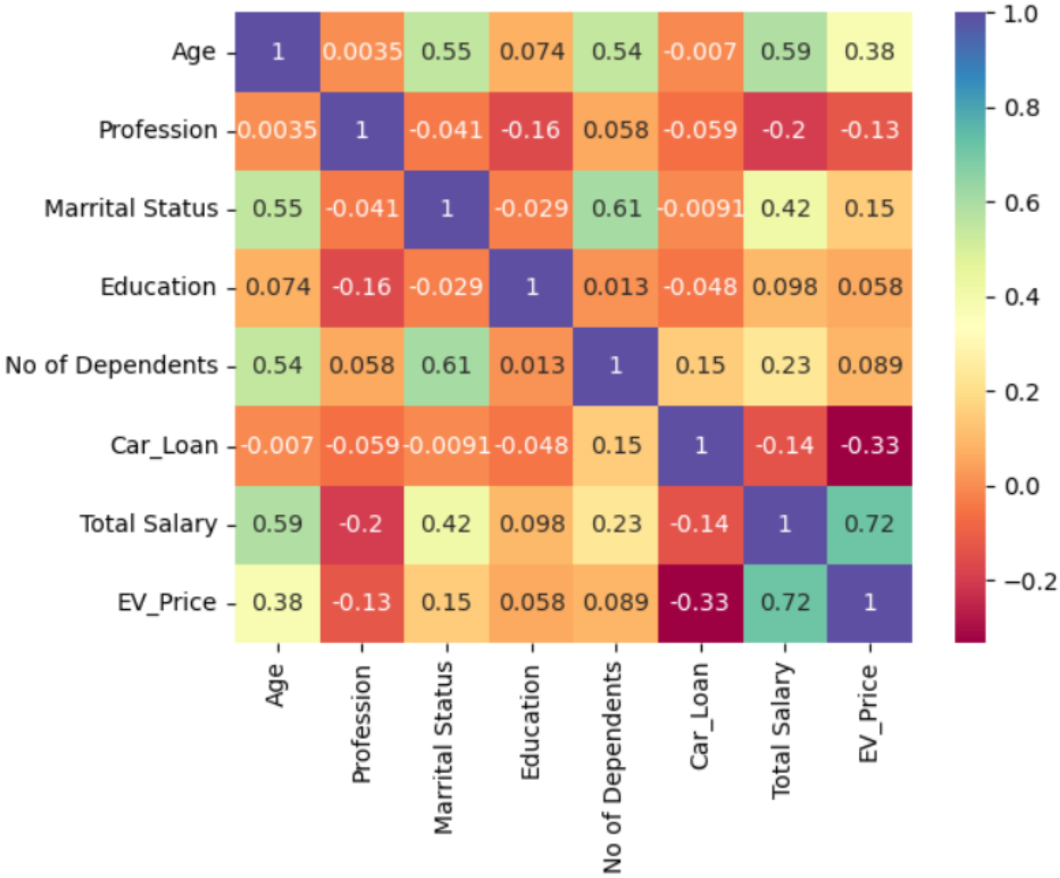


Figure 4: Caption

The PCA analysis reveals that only five components explain more than 85% of the variance when all features are considered.

5.3 K-Means Clustering

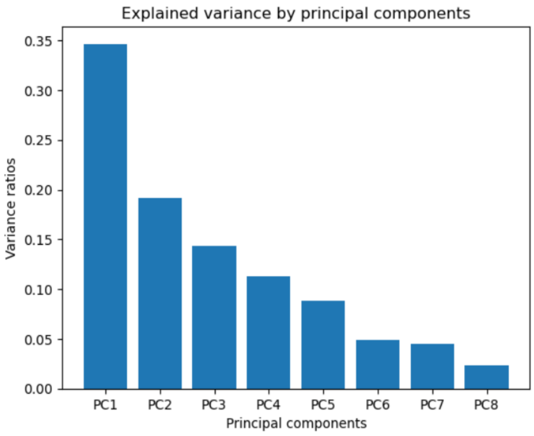
Next, I performed K-means clustering with different numbers of clusters and plotted the Elbow curve to determine the optimal number of clusters, as the algorithm requires this input.

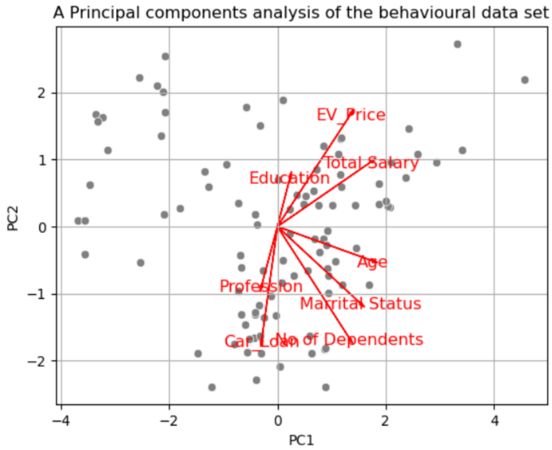
Based on the Elbow curve, I chose to use 5 clusters for the K-Means clustering. The resulting clusters are visualized in the plot, which displays the first principal component against the second principal component and also i give the first five rows.

5.4 Prediction

In this subsection, I address the initial questions by selecting target clusters that the two-wheeler EV company aims to cater to. The top five variables for targeting these segments are "Age," "Total Salary," "Profession," "Education," and "Car_Loan." The selection of target variables and clusters can vary based on the company's policies and goals. Here, we identify some possible target clusters, with their feature values displayed for different clusters-

Note: (Car Loan: 1->Yes, 0->No), (Profession: 1->Business, 0->Salaried), (Education: 0->Graduate, 1->Postgraduate)





	Standard Deviation	Proportion of Variance	Cumultive Proportion
PC1	1.673480	0.346531	0.346531
PC2	1.245957	0.192091	0.538622
PC3	1.076239	0.143324	0.681946
PC4	0.954883	0.112824	0.794770
PC5	0.844415	0.088229	0.882999
PC6	0.628961	0.048950	0.931948
PC7	0.603903	0.045127	0.977075
PC8	0.430429	0.022925	1.000000

From the cluster data, an EV company may choose the first cluster due to its members having:

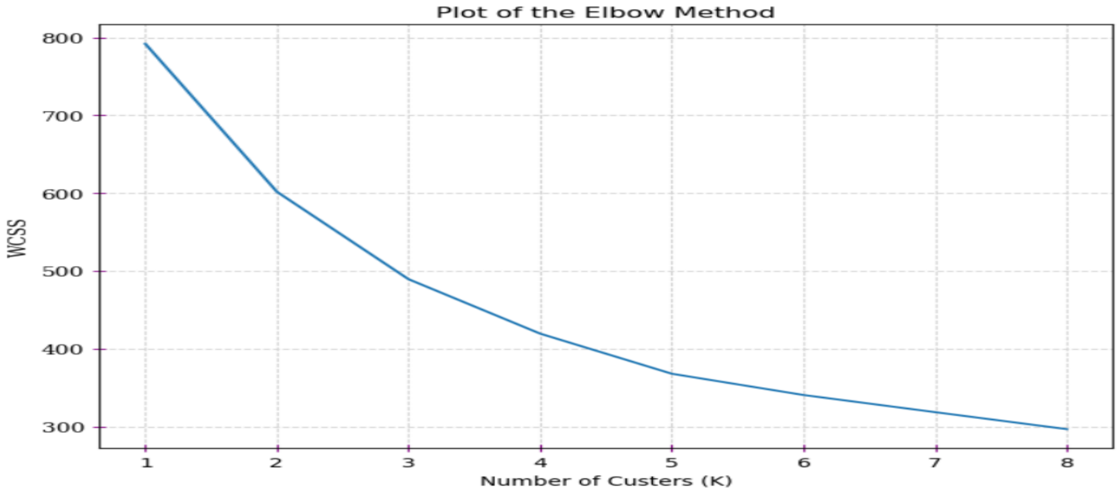
- A decent salary in the range of 13-29 LPA,
- An age range of 31-37,
- Not Having a Car Loan,
- No conclusion in education
- Salaried Employee.

Another preferable segment to target can be the 5th cluster, which includes members with:

- A decent salary in the range of 14-22 LPA,
- An age range of 33-40,
- Not Having a Car Loan,
- No conclusion in education
- Business Man.

6 Vehicle Segmentation

Here, I perform a segmentation analysis on the potential target segments based on the product (2-wheelers) features during product development. Let's explore the data through Exploratory Data Analysis (EDA).

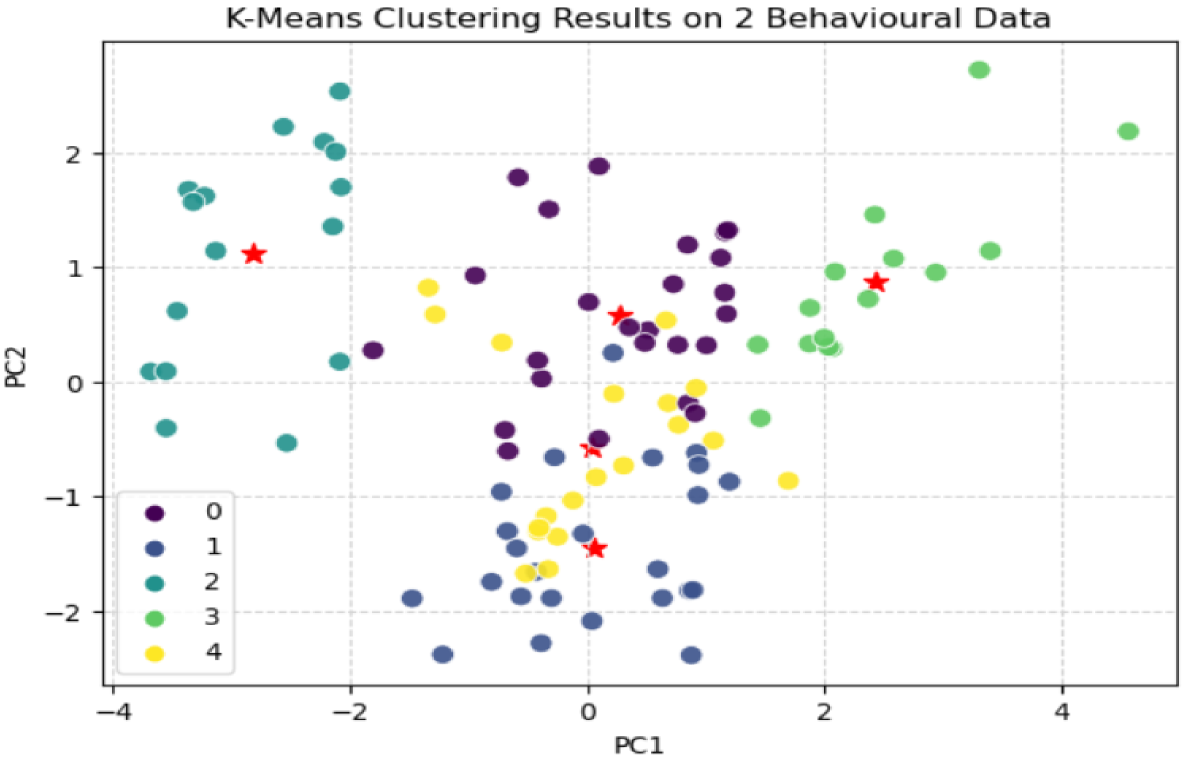


	Age	Profession	Marrital Status	Education	No of Dependents	Car_Loan	Total Salary	EV_Price	clusters
0	27	0	0	1	0	1	800000	800000	2
1	35	0	1	1	2	1	2000000	1000000	1
2	45	1	1	0	4	1	1800000	1200000	1
3	41	1	1	1	3	0	2200000	1200000	4
4	31	0	1	1	2	1	2600000	1600000	1

6.1 Exploratory Data Analysis

The sample of the data set is-

There exist two types of 2-wheelers: Bikes and Scooters. Below are their counts along with the year of manufacture of the 2-wheelers in the dataset:



The buying decisions also depend on the price of the product. Below are the price values extracted from the dataset:

6.2 PCA

Before clustering the data points, we present the correlation matrix to illustrate the interdependencies between features within the dataset. This step is essential for understanding the relationships among variables and ensuring an effective clustering process.

Therefore, it is necessary to perform PCA to extract independent principal components, focusing only on those that retain significant information from the entire dataset.

The data are preprocessed using the 'StandardScaler' class in Scikit-Learn. Following this, Principal Component Analysis (PCA) is applied to extract the independent components. The aim is to reduce the number of features while retaining most of the information, as indicated by the explained variance.

The PCA analysis reveals that only five components explain more than 95% of the variance when all features are considered.

```
[Education
 0     13
 1     12
 Name: count, dtype: int64,
 Education
 1     12
 0     12
 Name: count, dtype: int64,
 Education
 1     10
 0      6
 Name: count, dtype: int64,
 Education
 1     13
 0      2
 Name: count, dtype: int64,
 Education
 0     10
 1      9
 Name: count, dtype: int64]
```

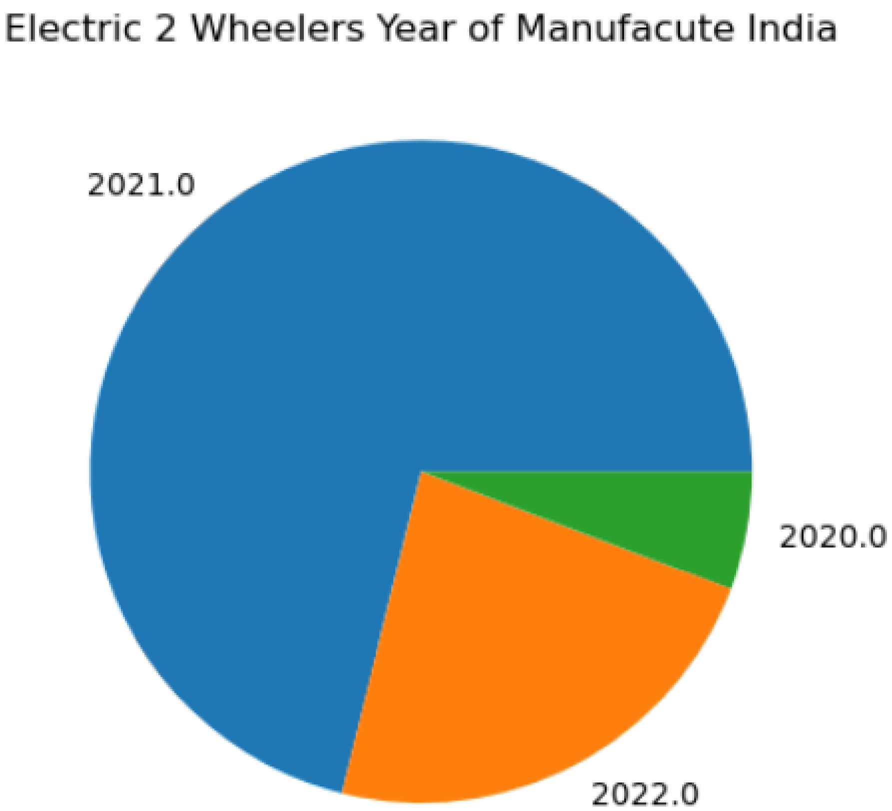
```
[Profession
 0     25
 Name: count, dtype: int64,
 Profession
 0     16
 1      8
 Name: count, dtype: int64,
 Profession
 0     10
 1      6
 Name: count, dtype: int64,
 Profession
 0     13
 1      2
 Name: count, dtype: int64,
 Profession
 1     19
 Name: count, dtype: int64]
```

```
[Car_Loan
 0     25
 Name: count, dtype: int64,
 Car_Loan
 1     24
 Name: count, dtype: int64,
 Car_Loan
 0     11
 1      5
 Name: count, dtype: int64,
 Car_Loan
 0     12
 1      3
 Name: count, dtype: int64,
 Car_Loan
 0     19
 Name: count, dtype: int64]
```

```
[Age
 35     4
 37     4
 31     3
 36     3
 34     3
 Name: count, dtype: int64,
 Age
 35     4
 31     4
 34     3
 36     3
 42     3
 Name: count, dtype: int64,
 Age
 27     6
 29     3
 30     3
 28     2
 26     1
 Name: count, dtype: int64,
 Age
 44     4
 41     3
 49     2
 42     1
 39     1
 Name: count, dtype: int64,
 Age
 36     6
 41     2
 35     2
 37     2
 33     1
 Name: count, dtype: int64]
```

```
[Total Salary
 2100000     3
 1300000     2
 2500000     2
 2900000     2
 2200000     2
 Name: count, dtype: int64,
 Total Salary
 1800000     3
 1900000     3
 2000000     2
 2600000     2
 1300000     2
 Name: count, dtype: int64,
 Total Salary
 800000      3
 1400000     3
 900000      3
 1100000     2
 1700000     1
 Name: count, dtype: int64,
 Total Salary
 4500000     2
 3100000     2
 4000000     2
 3000000     1
 3700000     1
 Name: count, dtype: int64,
 Total Salary
 1400000     3
 2000000     3
 1600000     3
 2200000     2
 1900000     1
 Name: count, dtype: int64]
```

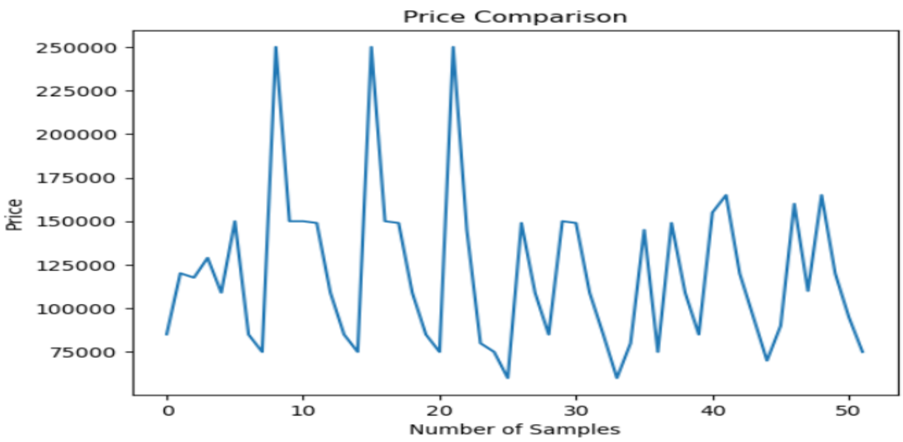
	Unnamed: 0	Model	Manufacturer	Vehicle Type	Battery Capacity (kWh)	Range per Charge (km)	Charging Time	Price	Power (HP or kW)	Top Speed (km/h)	Year of Manufacture
31	31	Pure EV Epluto 7G	Pure EV	Scooter	2.70	120	3.0	109000.0	5.0	80.0	2021.0
10	10	Bajaj Chetak Electric	Bajaj Auto	Scooter	4.00	95	5.0	150000.0	4.0	60.0	2020.0
33	33	Joy E-Bike Urbanite X1	Electric Vehicle Co.	Bike	2.20	80	2.5	60000.0	2.0	50.0	2021.0
0	0	Ola Electric S1	Ola	Scooter	2.98	181	5.0	85099.0	4.5	116.0	2021.0
11	11	Ather 450X	Ather Energy	Scooter	2.90	116	4.5	149000.0	6.0	80.0	2021.0



6.3 K-Mean Clustering

Next, I performed K-means clustering with different numbers of clusters and plotted the Elbow curve to determine the optimal number of clusters, as the algorithm requires this input.

Based on the Elbow curve, I chose to use 5 clusters for the K-Means clustering. The resulting clusters are visualized in the plot, which displays the first principal component against the second principal component and also i give the first five rows.



	Vehicle Type	Battery Capacity (kWh)	Range per Charge (km)	Charging Time	Price	Power (HP or kW)	Top Speed (km/h)	Year of Manufacture
Vehicle Type	1.000000	0.320543	0.286497	-0.076636	0.318706	0.676764	0.108007	0.101519
Battery Capacity (kWh)	0.320543	1.000000	0.782845	0.671725	0.874510	0.679582	0.620106	0.037480
Range per Charge (km)	0.286497	0.782845	1.000000	0.569228	0.768037	0.745132	0.890465	0.228106
Charging Time	-0.076636	0.671725	0.569228	1.000000	0.663869	0.360537	0.658871	-0.194250
Price	0.318706	0.874510	0.768037	0.663869	1.000000	0.786668	0.669637	-0.063526
Power (HP or kW)	0.676764	0.679582	0.745132	0.360537	0.786668	1.000000	0.600609	-0.002850
Top Speed (km/h)	0.108007	0.620106	0.890465	0.658871	0.669637	0.600609	1.000000	0.156641
Year of Manufacture	0.101519	0.037480	0.228106	-0.194250	-0.063526	-0.002850	0.156641	1.000000

6.4 Prediction

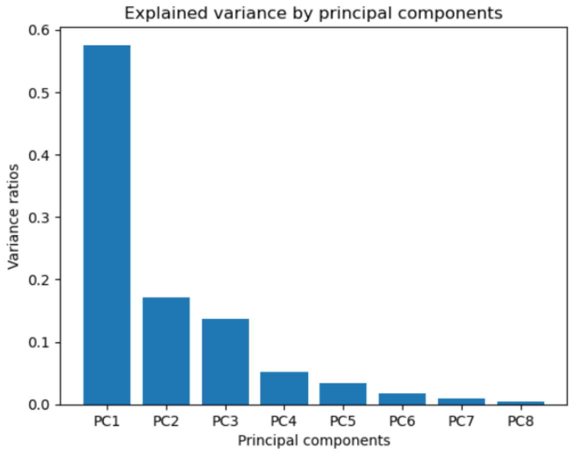
Now, I address the posed questions by selecting possible segments as targets. The choice of target variables and clusters can depend on the company policies and goals. Possible target clusters are identified here. The top features that a brand may focus on to select their segments are: Price, Vehicle Type, Top Speed, Range Per Charge, Battery Capacity, and Charging Time. The segments are:

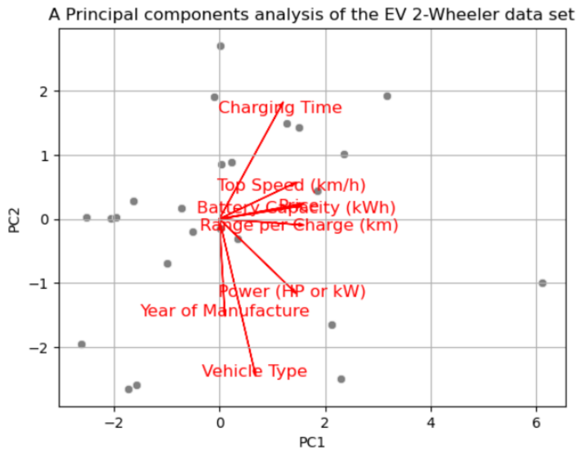
NOTE: Vehicle Type: 1->Bike, 0->Scooter The EV Manufacturer Company might choose the 1st Segment having feature values-

- Manufacture EV Bike
- Price Point of 150k-250k
- Top Speed having 85-100 km/h
- Range Having between 150-200 km,
- Battery Capacity of 3.2-6.2 kWh
- Charging Time of 4.0-5.0 Hrs

Another Possible Segment to target is 2nd Segment with features as-

- Manufacture EV Scooter
- Price Point of 109k-165k
- Top Speed having 80-90 km/h





	Standard Deviation	Proportion of Variance	Cumultive Proportion
PC1	2.166402	0.575380	0.575380
PC2	1.179781	0.170640	0.746020
PC3	1.057521	0.137105	0.883125
PC4	0.647789	0.051445	0.934570
PC5	0.524478	0.033723	0.968294
PC6	0.383907	0.018069	0.986362
PC7	0.268269	0.008823	0.995186
PC8	0.198169	0.004814	1.000000

- Range Having between 115-120 km,
- Battery Capacity of 2.7-4.0 kWh
- Charging Time of 3.0-5.5 Hrs

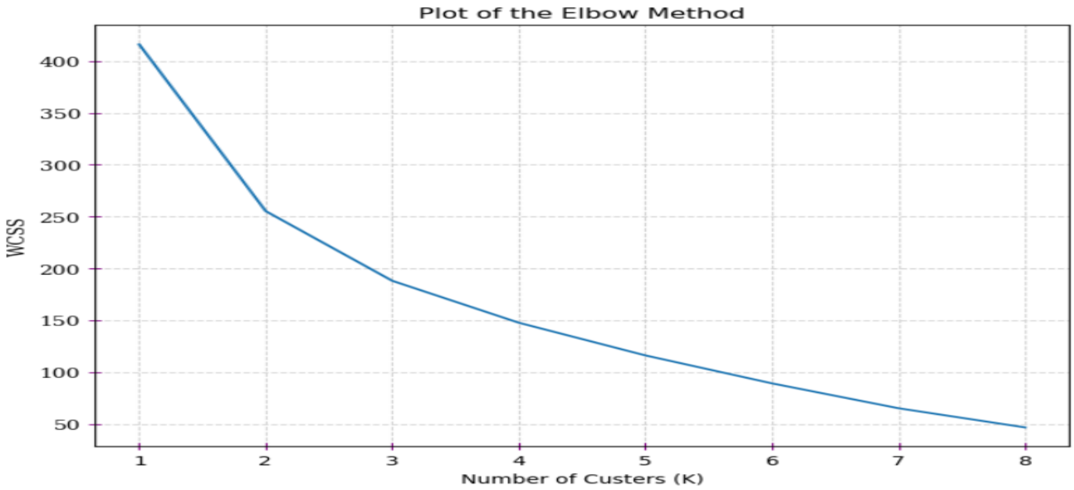
Another Possible Segment to target is 3rd Segment with features as-

- Manufacture EV Scooter
- Price Point of 75k-85k
- Top Speed having 50-60 km/h
- Range Having between 75-100 km,
- Battery Capacity of 2.2-3.0 kWh
- Charging Time of 3.0-5.5 Hrs

7 Improvement Possibilities

Even though analysis has been performed and results have been obtained, there is always room for improvement. In this case, different algorithms like DBSCAN or Gaussian Models could also be employed to gain further insights.

Having clearer data and more features would enhance segmentation precision. Such data could be generated through web scraping, manual data collection, surveys, etc.



8 Conclusion

There are many two-wheeler EV manufacturing companies in the country, including Ola Electric, Hero Electric, Ather Energy, TVS, Bajaj, and Okinawa, among others. The demand for electric vehicles is expected to grow, given the automotive nature of the market. Although the investments and policies surrounding this sector are significant, it will take some time for the market to fully settle in India, even though we are already witnessing the spread of EVs across the country.

Such segmentation analysis on EV products and demographic segmentation proves to be an invaluable resource for both new EV brands and established ones. It helps them better understand the market and predict the dynamics of this early-stage market, particularly where investments are happening.


```
[Manufacturer
 Revolt Motors     4
 Tork Motors       3
 Name: count, dtype: int64,
 Manufacturer
 Ather Energy      8
 Pure EV           8
 Bajaj Auto        5
 Ola               2
 TVS               1
 Name: count, dtype: int64,
 Manufacturer
 Okinawa Autotech          8
 Electric Vehicle Co.      5
 Hero Motocorp             3
 Ampere Vehicles           2
 Joy E-Bike                1
 Name: count, dtype: int64]
```

```
[Charging Time
 4.0     4
 5.0     3
 Name: count, dtype: int64,
 Charging Time
 4.5     6
 3.0     6
 5.0     5
 6.5     2
 5.5     2
 Name: count, dtype: int64,
 Charging Time
 3.0     13
 2.5      4
 3.5      2
 Name: count, dtype: int64]
```

```
[Vehicle Type
 1    7
 Name: count, dtype: int64,
 Vehicle Type
 0    26
 Name: count, dtype: int64,
 Vehicle Type
 0    15
 1     4
 Name: count, dtype: int64]
```

```
[Battery Capacity (kWh)
 3.2     4
 6.2     3
 Name: count, dtype: int64,
 Battery Capacity (kWh)
 4.00     7
 2.90     6
 2.70     6
 3.50     3
 2.98     1
 Name: count, dtype: int64,
 Battery Capacity (kWh)
 2.5     8
 2.2     5
 3.0     4
 2.8     1
 2.9     1
 Name: count, dtype: int64]
```

```
[Range per Charge (km)
 150     4
 200     3
 Name: count, dtype: int64,
 Range per Charge (km)
 120     7
 116     6
 181     2
 95      2
 146     2
 Name: count, dtype: int64,
 Range per Charge (km)
 100     10
 75       3
 80       2
 120      2
 90       1
 Name: count, dtype: int64]
```

```
[Price
 250000.0     3
 150000.0     3
 155000.0     1
 Name: count, dtype: int64,
 Price
 109000.0     6
 149000.0     5
 150000.0     2
 145000.0     2
 165000.0     2
 Name: count, dtype: int64,
 Price
 85000.0     6
 75000.0     6
 80000.0     2
 60000.0     2
 95000.0     2
 Name: count, dtype: int64]
```

```
[Top Speed (km/h)
 85.0      4
 100.0     3
 Name: count, dtype: int64,
 Top Speed (km/h)
 80.0      15
 90.0       3
 116.0      2
 60.0       2
 85.0       2
 Name: count, dtype: int64,
 Top Speed (km/h)
 60.0     14
 50.0      2
 70.0      2
 65.0      1
 Name: count, dtype: int64]
```