Table of Contents

PROBLEM STATEMENT

Question

Based on Market Analysis, the entire segmentation problem boils down to Two main Questions-

What type of EV the company will produce?

Who are the target customers?

In this case I am specifically looking into the 2wheeler EV market as it is much bigger and more dynamic market comparatively. Thus, I look for what type of EV 2 Wheelers a EV 2Wheeler company should focus it's resources on.

Approach

Task is to analyse the Electric Vehicles Market in India using *Segmentation* analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use their product in terms of Geographic, Demographic, Psychographic, and Behavioural.

In this report I have analysed the 2Wheeler Electric Vehicles Market in India using segmentation analysis and tried to answer some of the crucial questions. Along with that, I have also performed Segmentation on customers and their Income Data.

The Segmentation is Done using PCA and K-Means clustering and Hierarchical Clustering is also shown.

At the end the possible segments to choose for such an EV company, are evaluated based of selected features.

The Git Hub Link to the Analysis- https://github.com/mrinmoymondal898/EV-MARKET-SEGMENTATION

DATA COLLECTION

Both the Demographic dataset and EV bikes dataset are collected from https://www.kaggle.com/
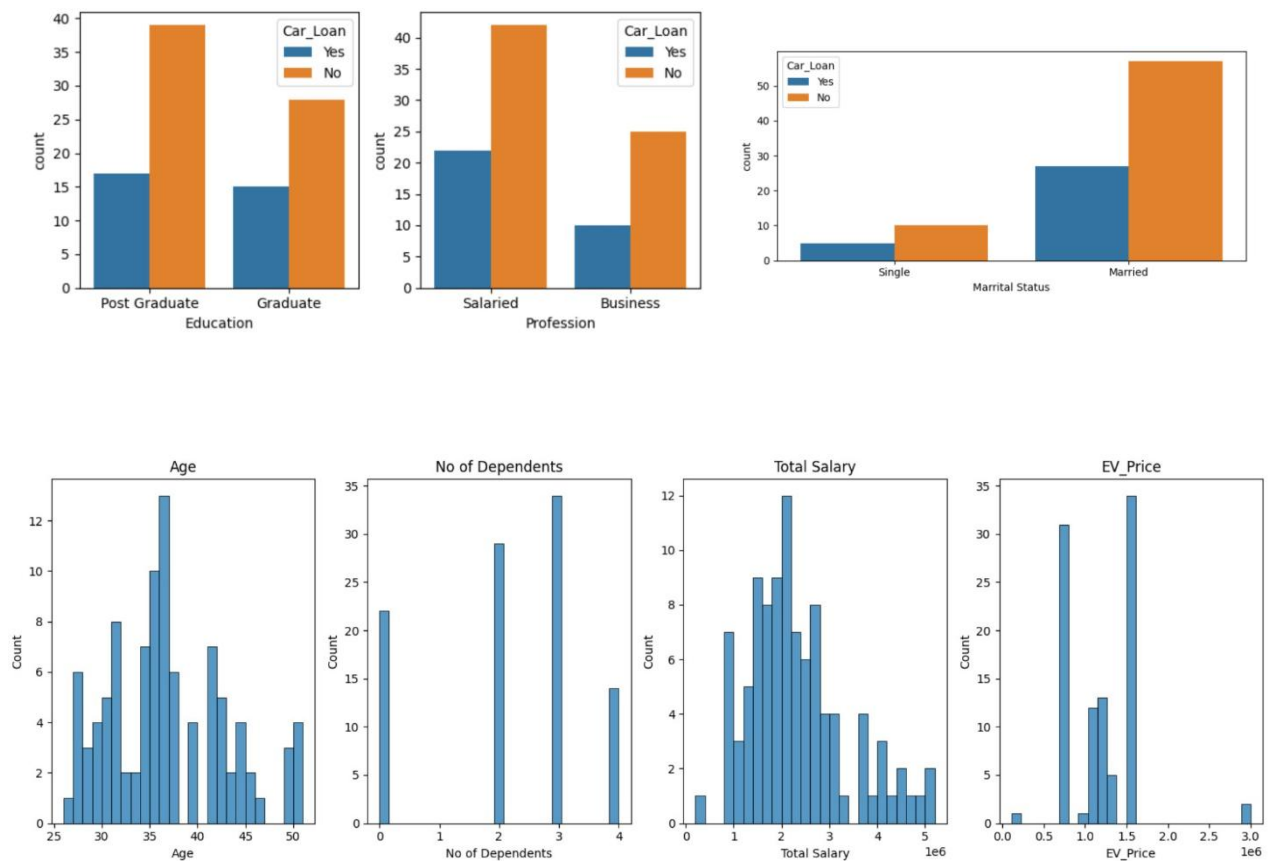
BEHAVIOURAL SEGMENTATION

In this part I show the methods taken to do a demographic segmentation on the Dataset. But first some Exploratory Data Analysis is performed. An Exploratory Data

Analysis or EDA is a thorough examination meant to uncover the underlying structure of a data set and is important for a company because it exposes trends, patterns, and relationships that are not readily apparent.
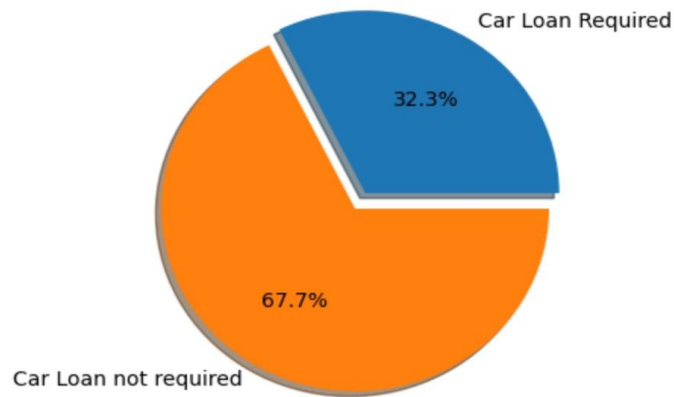
Exploratory Data Analysis

	Age	Profession	Marrital Status	Education	No of Dependents	Car_Loan	Total Salary	EV_Price
0	27	Salaried	Single	Post Graduate	0	Yes	800000	800000
1	35	Salaried	Married	Post Graduate	2	Yes	2000000	1000000
2	45	Business	Married	Graduate	4	Yes	1800000	1200000
3	41	Business	Married	Post Graduate	3	No	2200000	1200000
4	31	Salaried	Married	Post Graduate	2	Yes	2600000	1600000

Count plots are used to show categorical features like Profession, Education, etc. and how many of them take Loan to buy EV.
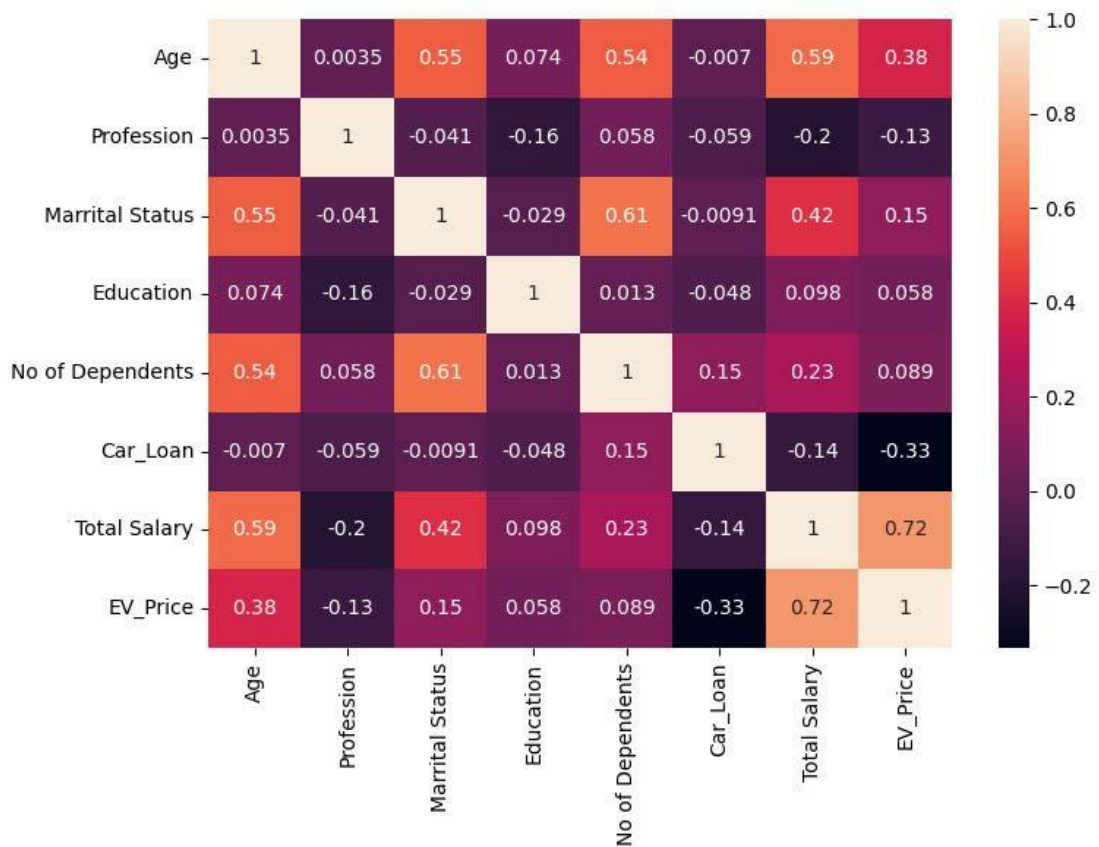
A data to show percentage of candidates taking loan, which might be useful for purchase options using EMI and other options given by the company.
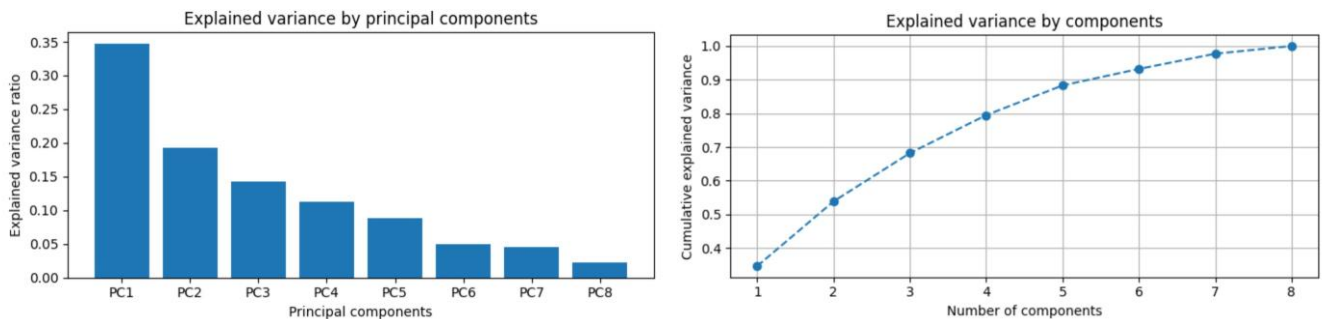

Principal Component Analysis

Before clustering the datapoints I have shown the correlation matrix where the dataset showing interdependency between features.

The data are pre-processed using Standard Scalar class in ScikitLearn and I proceed for PCA to extract the independent components and less than the number of features for which most of the information is intact i.e the explained variance.


It is obvious from the PCA analysis by taking all the features that only 5 components explain more than 90% of the variance. The principal components-

|	PC1	PC2	PC3	PC4	PC5
0	-3.456779	-0.618798	-1.775729	-0.151357	1.047749
1	-0.274146	0.658946	-1.695786	-0.042108	0.291098
2	0.879912	2.388637	0.925667	-0.363319	1.116538
3	0.771649	0.376038	0.757106	1.685393	0.163439
4	0.219923	-0.252808	-1.513077	-0.399144	0.710610

K-Means Clustering

Now I performed K-means clustering for different number of clusters and plot the Elbow curve to determine the number of clusters, as the algorithm needs the number of clusters to be given as an input. One point to note that I have used "k-means++". Only difference from normal one is that it initialises the clusters smartly rather than randomly in normal k-means.
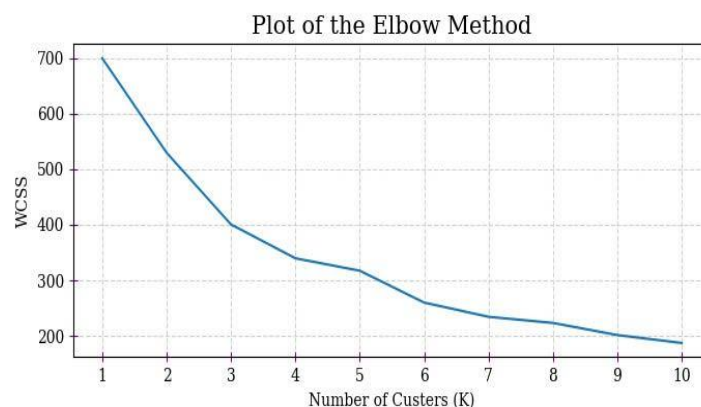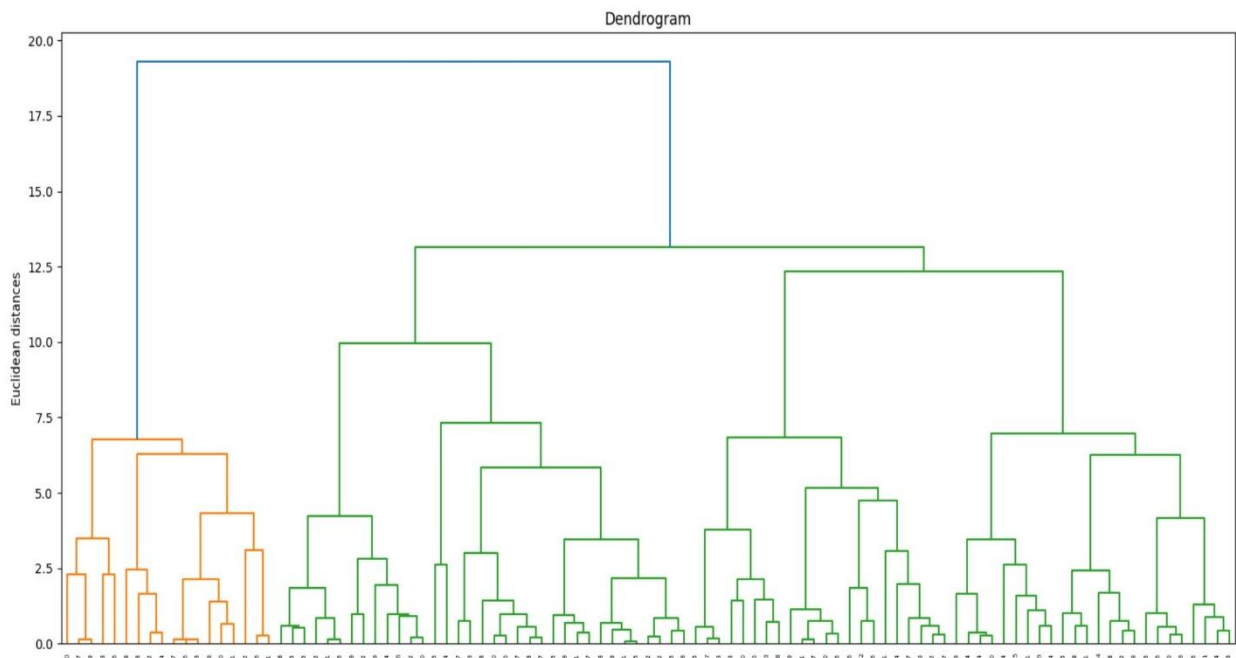
According to the Elbow curve I have chosen 4 clusters to preform K-Means. The clusters are shown in the plot(First vs Last principal components).
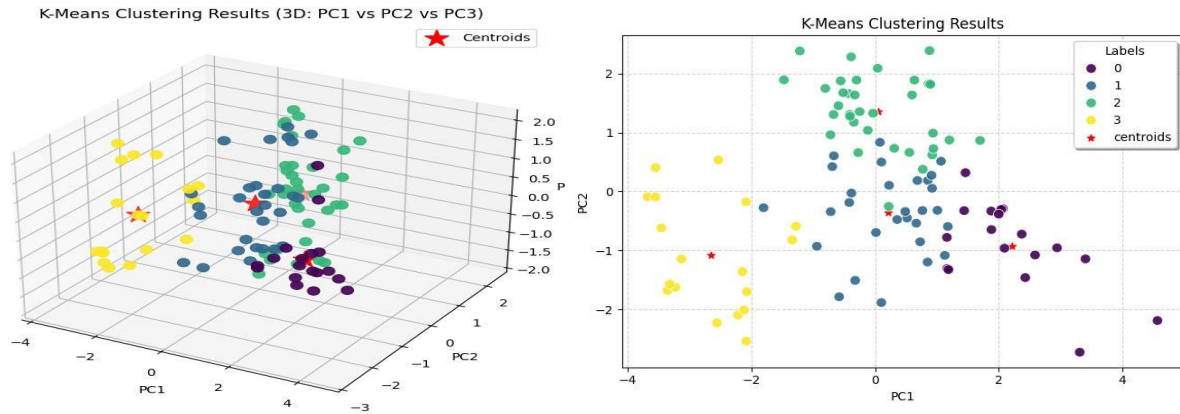
Hierarchical Clustering

A similar hierarchical clustering is also performed which shows 2 different components, shown in the Dendrogram.

Making Predictions

In this subsection I try to answer the questions that posed initially, by selecting target clusters which to the 2 wheeler EV Company wants to cater to. The top 4 variables on which the segment can be targeted are- "Age", "Total Salary", "Profession", and "Carload". The target variables and chosen cluster can depend on the company policies and their goal. Some possible target clusters are chosen here. The feature values for different clusters are shown here.

```
[Age
44    4
41    3
49    2
42    2
36    2
Name: count, dtype: int64,
Age
36    6
37    5
35    4
31    3
34    3
Name: count, dtype: int64,
Age
35    6
36    5
31    4
34    4
42    3
Name: count, dtype: int64,
Age
27    6
29    4
30    3
28    2
26    1
Name: count, dtype: int64]
```

```
[Car_Loan
0    15
1     3
Name: count, dtype: int64,
Car_Loan
0    29
Name: count, dtype: int64,
Car_Loan
1    24
0    10
Name: count, dtype: int64,
Car_Loan
0    13
1     5
Name: count, dtype: int64]
```

```
[Profession
0    16
1     2
Name: count, dtype: int64,
Profession
0    22
1     7
Name: count, dtype: int64,
Profession
1    18
0    16
Name: count, dtype: int64,
Profession
0    10
1     8
Name: count, dtype: int64]
```

```
[Education
1    16
0     2
Name: count, dtype: int64,
Education
0    19
1    10
Name: count, dtype: int64,
Education
1    18
0    16
Name: count, dtype: int64,
Education
1    12
0     6
Name: count, dtype: int64]
```

```
[Total Salary
4500000    2
3100000    2
4000000    2
3000000    1
3700000    1
Name: count, dtype: int64,
Total Salary
2100000    3
2200000    3
1300000    2
2500000    2
2400000    2
Name: count, dtype: int64,
Total Salary
2000000    4
1900000    4
1400000    3
1800000    3
1600000    3
Name: count, dtype: int64,
Total Salary
1400000    4
800000     3
900000     3
1100000    2
1700000    1
Name: count, dtype: int64]
```

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

From the Cluster data an EV company may choose the **4th Cluster** due to it having-

- *Decent Salary in range of 8-17 LPA,*

- *At an age of 27-30,*

- *Not Having a Car Loan,*

- *Salaried Employee.*

Another preferable Segment to target can be the **3rd Cluster** having-

- *Decent Salary in range of 14-20 LPA,*

- *At an age of 35-42,*

- *Having a Car Loan,*

- *Owning a Business.*

VEHICLE SEGMENTATION

Here I perform a segmentation analysis on the possible segment to target depending on the product (2 wheelers) features during product development. Let's explore the data through Exploratory Data Analysis

Exploratory Data Analysis

Let's look at the dataset-

|	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
0	Ola Electric S1	Ola	Scooter	2.98	181	5.0	85099.0	4.5	116.0	2021.0
1	Ola Electric S1 Pro	Ola	Scooter	4.00	181	6.5	120149.0	11.0	116.0	2021.0
2	TVS iQube	TVS	Scooter	3.40	75	6.5	117630.0	4.4	78.0	2020.0
3	Aura 300 Plus	Ather Energy	Scooter	2.90	116	4.5	129000.0	6.0	80.0	2021.0
4	Pure EV Epluto 7G	Pure EV	Scooter	2.70	120	3.0	109000.0	5.0	80.0	2021.0

There exists two Types of 2 wheelers: Bikes and Scooters, their counts along with the year of manufacture of the 2 wheelers in the dataset-
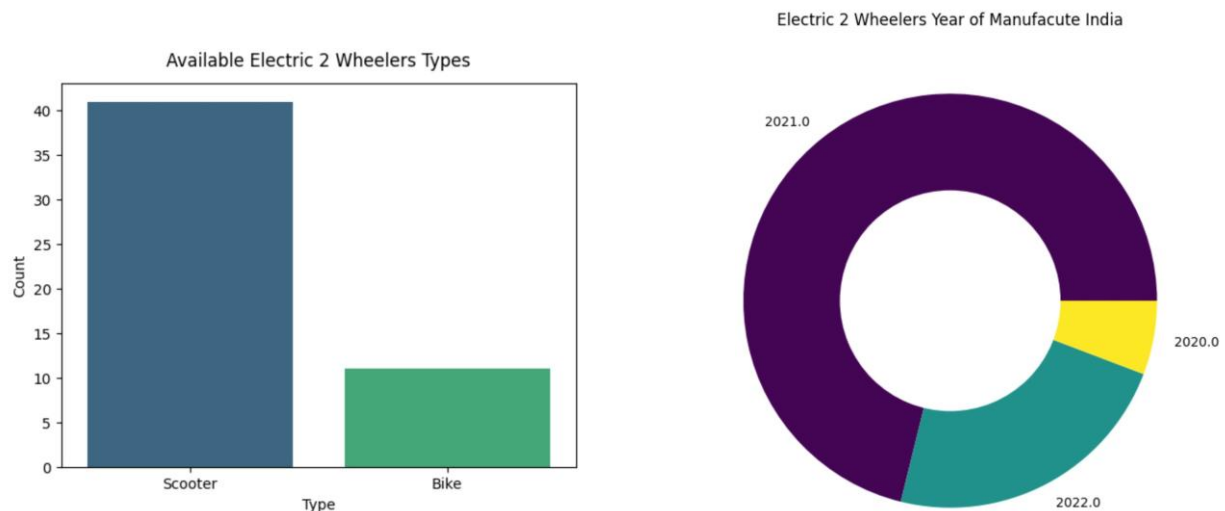

We only have 2 wheelers from 2020, 2021 and 2022 with most of them begin Electric Scooters.

The most important features that the buyer might look for in an EV are Range, Battery Size and probably the Top Speed. These information help the buyer to choose a product. These features are plotted in bar plot-
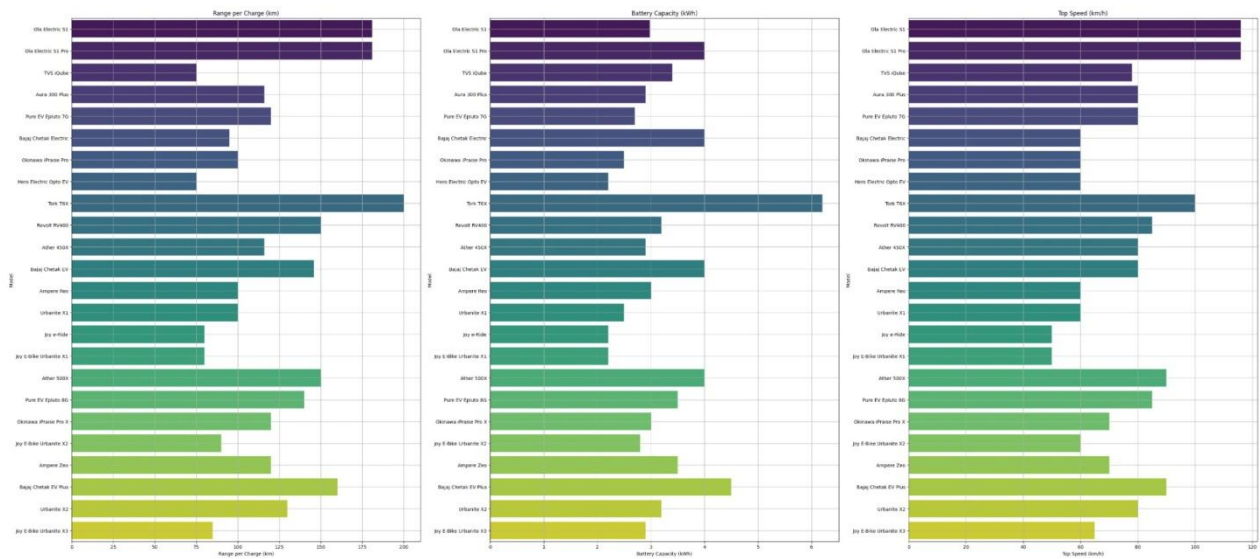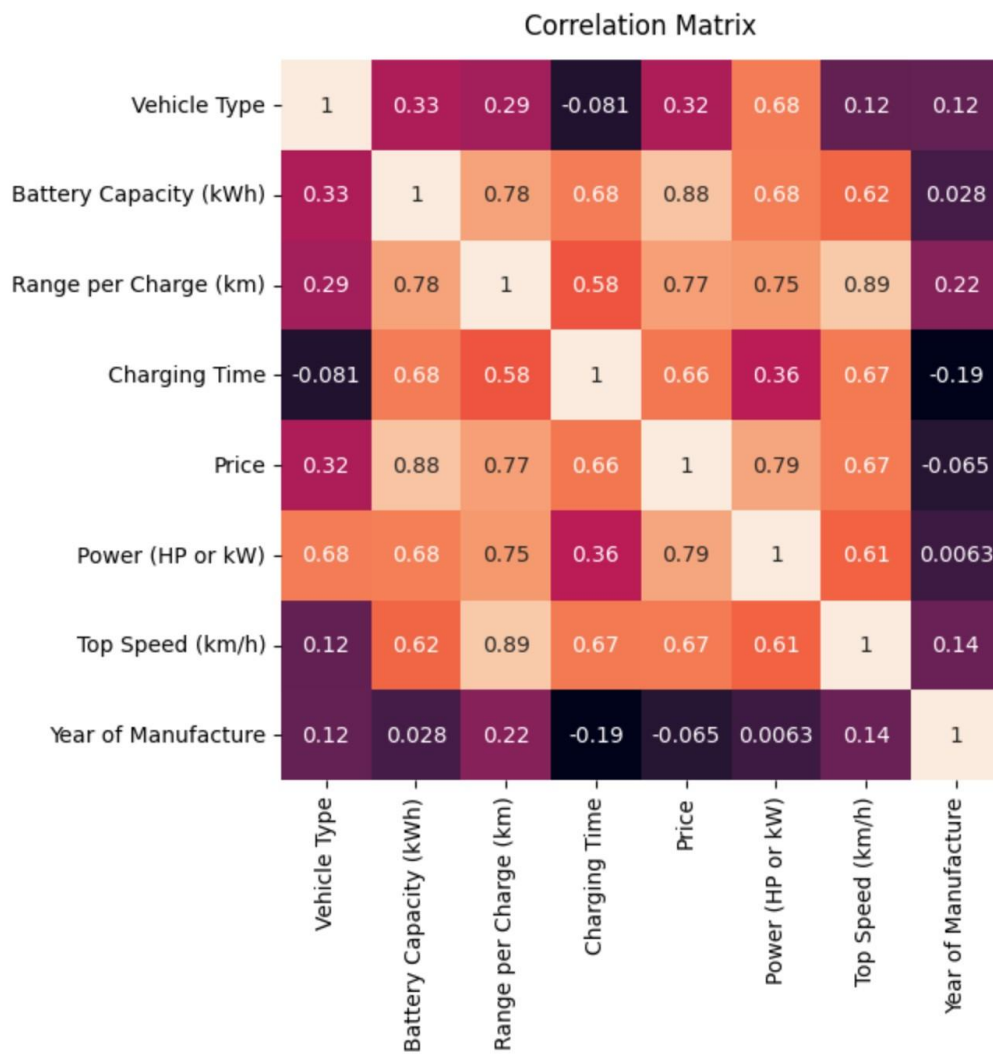
The buying decisions also depend on the price of the product. The price values from the dataset-
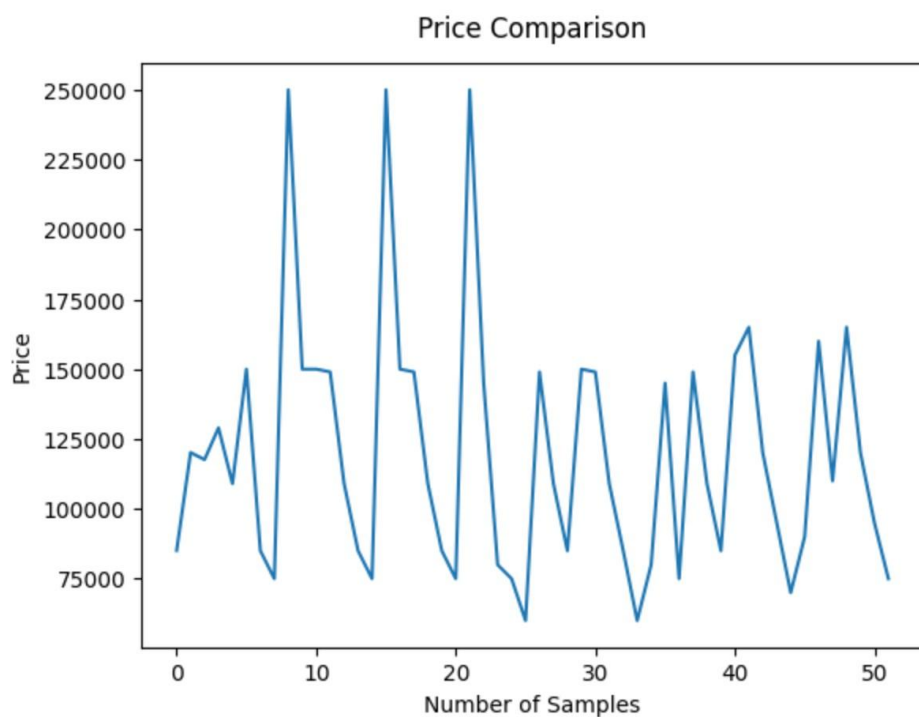
Principal Component Analysis

I have performed correlation analysis among the features and they had significant correlation-

VEHICLE SEGMENTATION

Correlation Matrix

	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	0.33	0.29	-0.081	0.32	0.68	0.12	0.12
Battery Capacity (kWh)	0.33	1	0.78	0.68	0.88	0.68	0.62	0.028
Range per Charge (km)	0.29	0.78	1	0.58	0.77	0.75	0.89	0.22
Charging Time	-0.081	0.68	0.58	1	0.66	0.36	0.67	-0.19
Price	0.32	0.88	0.77	0.66	1	0.79	0.67	-0.065
Power (HP or kW)	0.68	0.68	0.75	0.36	0.79	1	0.61	0.0063
Top Speed (km/h)	0.12	0.62	0.89	0.67	0.67	0.61	1	0.14
Year of Manufacture	0.12	0.028	0.22	-0.19	-0.065	0.0063	0.14	1

Therefore, it is necessary to preform PCA to extract independent principal components and only those which have significant information of the entire dataset.

We first scale the data to perform PCA. Taking all the features to do PCA we get the explained variance as-

VEHICLE SEGMENTATION
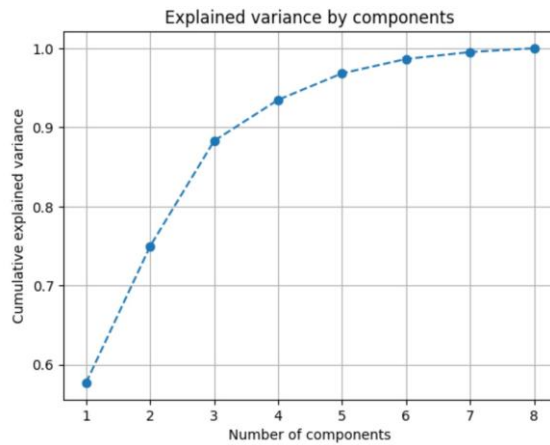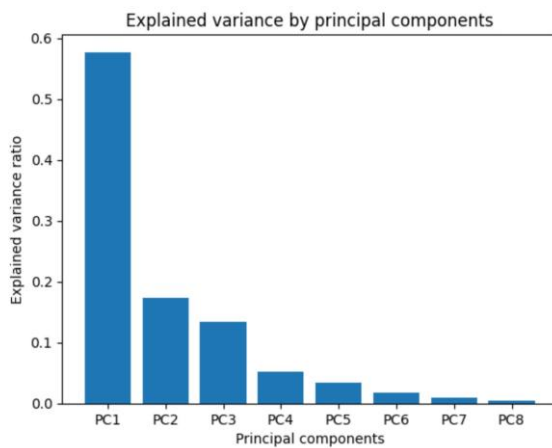
I have taken 4 Principal Components as it is enough for more than 90% of explained variance. The principal components along with the weightage of each feature with the components are-

	PC1	PC2	PC3	PC4
0	1.503078	-1.388905	1.384091	2.192278
1	3.154668	-1.876417	0.996621	1.296237
2	0.027291	-2.739152	-1.835036	-0.281784
3	0.038391	-0.851359	0.014584	0.209208
4	-0.712720	-0.165038	0.134792	0.754506

Now I proceed for the clustering task with the principal components. I have shown both K-means and Hierarchical clustering methods to extract the segments.

K-Means Clustering

K-means takes number of clusters as an input as it itself can't determine number of clusters. To do that I rum k-means for various number of clusters and choose the optimal one. One point to note that I have used "k-means++". Only difference from normal one is that it initialises the clusters smartly rather than randomly in normal k-means. The elbow plot-

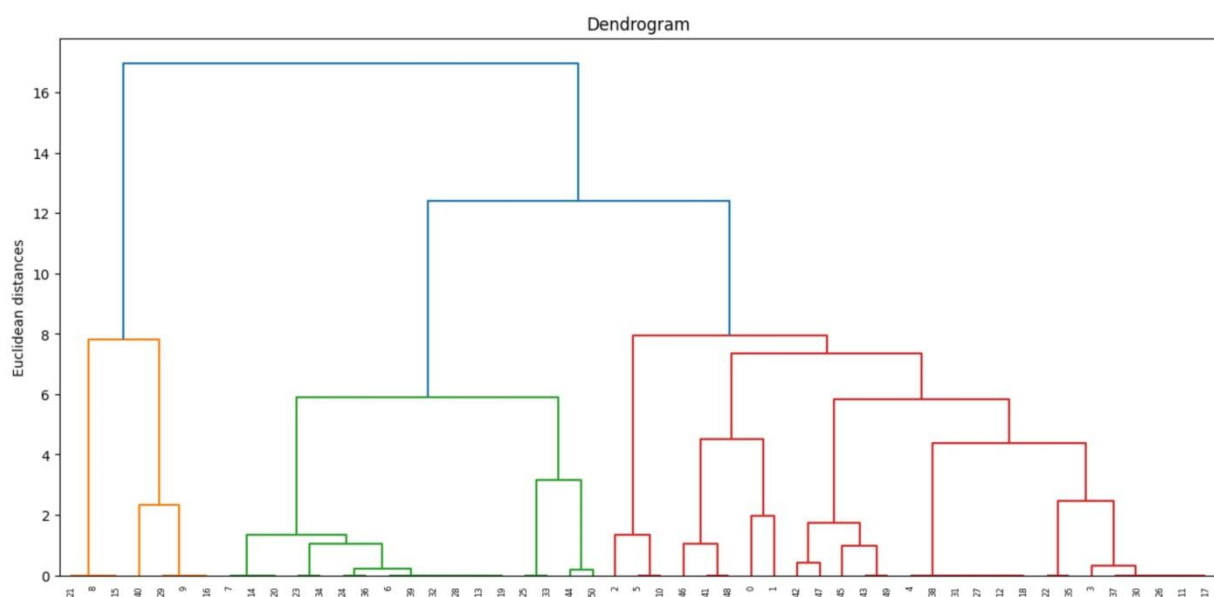

Plot of the Elbow Method
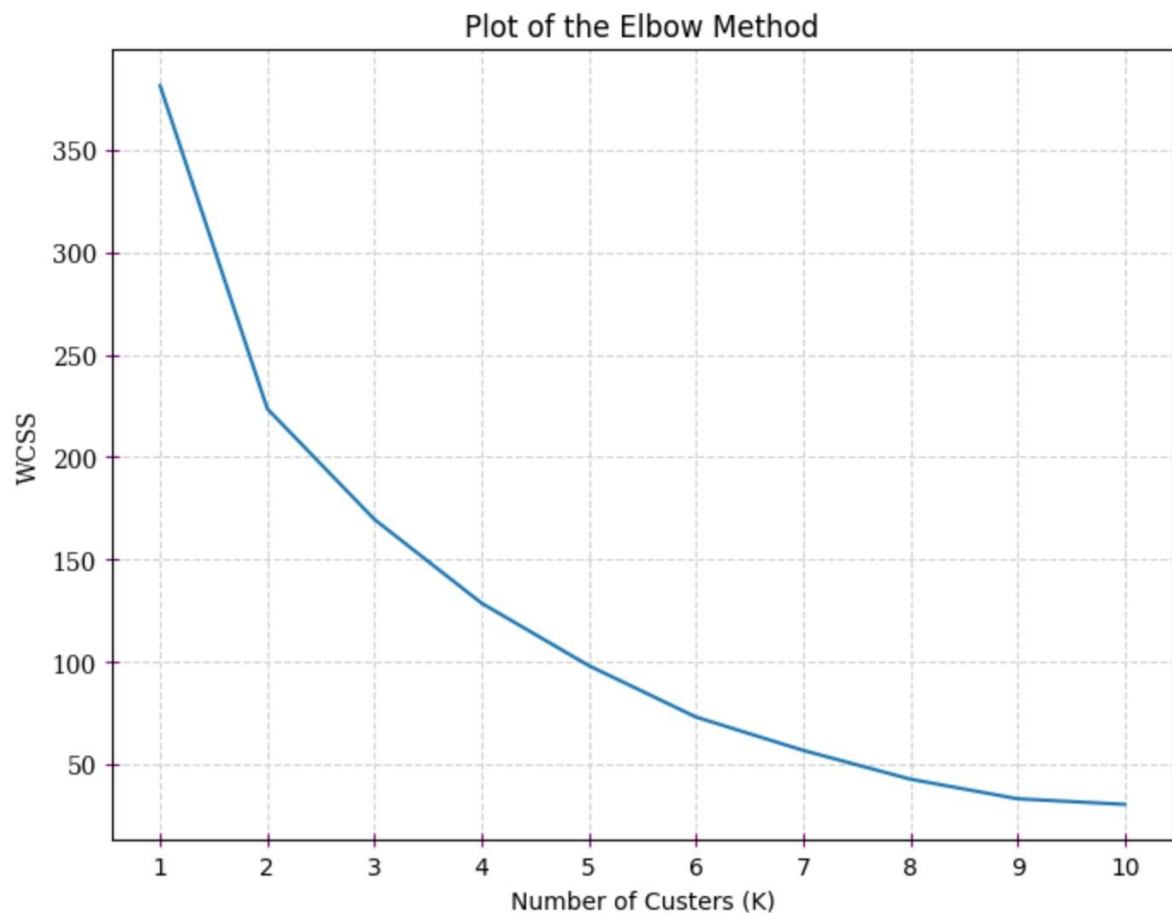
I have selected 3 clusters for the clustering process.

The clustering result is plotted with highest explained variance component vs lowest-
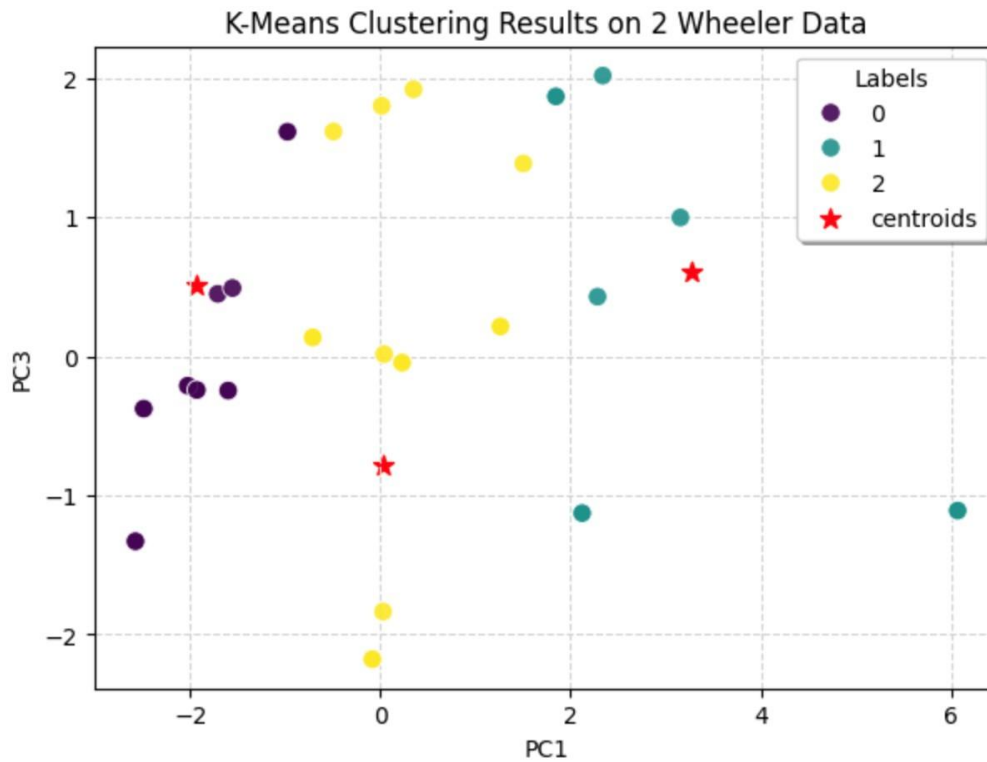
Hierarchical Clustering

The Dendrogram showing the clusters performed using hierarchical clustering which also gives 3 clusters-

Making Predictions

Now I try to answer the posed questions by selecting possible segments as targets. The target variables and chosen cluster can depend on the company policies and their goal. Possible target clusters are chosen 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**

```
[Vehicle Type
0    15
1     4
Name: count, dtype: int64,
Vehicle Type
1     7
0     4
Name: count, dtype: int64,
Vehicle Type
0    21
Name: count, dtype: int64]
```

```
[Price
85000.0     6
75000.0     6
80000.0     2
60000.0     2
95000.0     2
Name: count, dtype: int64,
Price
250000.0    3
150000.0    3
165000.0    2
120149.0    1
155000.0    1
Name: count, dtype: int64,
Price
109000.0    6
149000.0    5
150000.0    2
145000.0    2
85099.0     1
Name: count, dtype: int64]
```

```
[Top Speed (km/h)
60.0    14
50.0     2
70.0     2
65.0     1
Name: count, dtype: int64,
Top Speed (km/h)
85.0     4
100.0    3
90.0     3
116.0    1
Name: count, dtype: int64,
Top Speed (km/h)
80.0    15
60.0     2
116.0    1
78.0     1
85.0     1
Name: count, dtype: int64]
```

```
[Range per Charge (km)
100    10
75      3
80      2
120     2
90      1
Name: count, dtype: int64,
Range per Charge (km)
150     6
200     3
181     1
160     1
Name: count, dtype: int64,
Range per Charge (km)
120     7
116     6
95      2
146     2
181     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,
Battery Capacity (kWh)
3.2     4
4.0     3
6.2     3
4.5     1
Name: count, dtype: int64,
Battery Capacity (kWh)
2.90    6
2.70    6
4.00    4
3.50    2
2.98    1
Name: count, dtype: int64]
```

The EV Manufacturer Company might choose the **1st Segment** having feature values-

- Manufacture EV Scooter
- *Price Point of 85k-95k*
- *Top Speed having 50-70 km/h*
- *Range Having between 80-120 km,*
- *Battery Capacity of 2.2-3 kWh*
- *Charging Time of 2.5-2.5 Hrs*
- *Already* *Present Manufacturers in the Segment are Hero, Okinawa, Ampere, etc.*

```
[Charging Time
3.0    13
2.5     4
3.5     2
Name: count, dtype: int64,
Charging Time
5.0     5
4.0     4
6.5     1
6.0     1
Name: count, dtype: int64,
Charging Time
4.5     6
3.0     6
5.0     3
5.5     2
4.0     2
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,
Manufacturer
Revolt Motors    4
Tork Motors      3
Ather Energy     2
Ola              1
Bajaj Auto       1
Name: count, dtype: int64,
Manufacturer
Pure EV         7
Ather Energy    6
Bajaj Auto      4
Ola             1
TVS             1
Name: count, dtype: int64]
```

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

- Manufacture EV Scooter
- *Price Point of 85k-149k*
- *Top Speed having 80-116 km/h*
- *Range Having between 95-181 km,*

- *Battery Capacity of 2.7-4 kWh*
- *Charging Time of 3-5.5 Hrs*
- *Already* **Present Manufacturers in the Segment are Ola, Aether, Bajaj , TVS, etc.**

POSSIBLE IMPROVEMENTS

Even though through Analysis have been performed and Results have been obtained, improvements are always possible. In this case, different algorithms like DBSCAN or Gaussian Models might have also been used for further insights.

A clearer data and more features would have helped to segment having more precision. Such data could also be generated Scraping, Manual Data Collection, Surveys, etc.

MARKET MIX

The four main points of a marketing mix called 4P's- Product, Place, Promotion, and Price. They specify the approaches and strategies that address the target market, based on the details of the marketing plan.

Product

Each product line represents a group of outputs or products. The set of all the product lines is called the product mix. The company can have different Kind of Products, Models of EV. They might offer- • Automobiles

- Automobile parts
- Commercial vehicles
- Financial services

Pricing Strategies

The Company should have a strategy for the pricing of it's products depending on the segments the company targeting. Both Premium and Affordable strategies are necessary e.g-

- Market-oriented pricing strategy
- Premium pricing strategy

Promotion

One of the most important task for the Company is to tell the customer about the product they have produced in a convincing way. The activities are-

- Advertising (Primary Method)
- Direct marketing
- Personal selling
- Sales promotion
- Public relations

Place

In this aspect, the virtual or physical locations of transactions are considered. Such locations are significant because they enable the company to reach target customers in specific markets. The places can be-

- Official websites
- Dealerships
- Automotive shows and exhibits

CONCLUSION

There are many 2 wheeler EV manufacturing companies in the country like Ola Electric, Hero Electric, Aether Energy, TVS, Bajaj, Okinawa, etc. The demand will get higher & higher since it is automotive so the investments and policies and all that would be bigger but it will take some time to perfectly settle in India even though we have already started to see them spreading throughout the country.

Such a segmentation analysis on EV Product, Demographic Segmentation, etc. proves to be invaluable resource to New EV Brands as well as already established brands to better understand the market and predict the dynamics of such and early market especially where investments are happening.

Report Prepared by: *Mrinmoy Mondal* (Machine Learning Intern, Feynn Labs)