MARKET SEGMENTATION ANALYSIS OF ELECTRIC VEHICLES

G L Katyayini Koneti

Date: 2^{nd} May 2024

Abstract

Electric vehicles are becoming an increasingly popular choice for consumers. However, the market for electric vehicles is not monolithic. There are a variety of different consumer segments that are interested in electric vehicles. This report identifies these segments and explores their characteristics. By understanding these different segments, businesses can develop more targeted marketing campaigns and products for electric vehicles.

1.0 Introduction

The electric vehicle (EV) market is experiencing rapid growth globally, but competition is fierce. To ensure success, our newly established EV start-up needs to identify a specific vehicle segment with high potential for early adoption and long-term profitability in Indian market.

2.0 Objective

Develop a market entry strategy by conducting a comprehensive historical data and segmentation analysis of the EV market in India. This analysis will identify the most promising segment for our initial EV offering.

3.0 About Data

The data used for analysis is the historical data of EV vehicles in the Indian market from 2016-2023.

It contains the following **features**:

- region: India
- category: Historical
- parameter: 'EV sales share', 'EV stock share', 'EV sales', 'EV stock', 'Oil displacement Mbd', 'Oil displacement, million lge', 'Electricity demand', 'EV charging points'
- mode: 'Vans', 'Cars', 'Buses', 'EV', 'Trucks'
- powertrain: 'EV', 'BEV', 'PHEV', 'FCEV', 'Publicly available fast', 'Publicly available slow'
- year: 2016-2023
- unit: Depends on parameter unit like 'percent', 'Vehicles', 'Million barrels per day', 'Oil displacement, million lge', 'GWh', 'charging points'
- value: Numerical value of parameter in the above units.

Column Description:

- powetrain:

  BEVs are battery electric vehicles.
  PHEVs are plug-in hybrid electric vehicles.
  FCEVs are fuel cell electric vehicles.
  EVs refers to all electric vehicles (BEVs + PHEVs).

- parameter:

  Sales and Stock:

  EV Sales: This refers to the number of new electric vehicles sold within a specific timeframe, typically a month, quarter, or year. It's a key metric for gauging the growth and adoption rate of EVs in the market.

EV Stock:

This represents the total number of electric vehicles currently in operation within a particular region or globally. It considers all previously sold EVs that haven't been scrapped or permanently removed from use.

Market Share:

EV Sales Share: This metric indicates the percentage of total vehicle sales that are attributed to electric vehicles within a specific timeframe. It helps understand the penetration level of EVs compared to gasoline-powered vehicles.

EV Stock Share:

This reflects the portion of all vehicles in operation that are electric vehicles, expressed as a percentage. It provides insights into the long-term trend of EV adoption and the gradual shift toward electrification.

Oil Displacement:

Oil Displacement Mbd (Million barrels per day): This metric estimates the volume of oil that is no longer needed due to the increased adoption of EVs. It highlights the potential impact of EVs on reducing dependence on fossil fuels and greenhouse gas emissions.

Oil Displacement, million lge (million liters gasoline equivalent):

Similar to Mbd, this metric expresses the volume of gasoline that is displaced by EVs, again emphasizing the environmental benefits of electric transportation.

Electricity Demand:

This refers to the amount of electricity needed to power the growing fleet of EVs. As more EVs are adopted, the demand for electricity is expected to rise, requiring grid upgrades and investments in renewable energy sources to meet this demand sustainably.

EV Charging Points:

This refers to the number of locations where EVs can be plugged in to recharge their batteries. This includes public charging stations, private charging points at homes or workplaces, and fast-charging stations for rapid recharging. The availability and accessibility of charging infrastructure are crucial factors influencing EV adoption.

External link : https://www.iea.org/data-and-statistics/data-tools/global-ev-data-explorer

4.0 Libraries Used

- Pandas and Numpy for loading and analysing data
- Seaborn and Matplotlib for visualization of data
- ScikitLearn for preprocessing
- Bioinfokit for visualizing cluster
- Scikit.Cluster for KMeans clustering
- Yellowbrick which extends the Scikit-Learn API to make model selection and hyperparameter tuning easier

5.0 Loading Data

The data set is loaded using pandas library:

1. Importing and Exploring the Data Set

```
df = pd.read_csv('EV_data_latest_till_2023.csv')
```

```
df
```

	region	category	parameter	mode	powertrain	year	unit	value
0	India	Historical	EV sales share	Vans	EV	2016	percent	0.056000
1	India	Historical	EV stock share	Vans	EV	2016	percent	0.003200
2	India	Historical	EV sales	Vans	BEV	2016	Vehicles	180.000000
3	India	Historical	EV stock	Vans	BEV	2016	Vehicles	180.000000
4	India	Historical	Oil displacement Mbd	Vans	EV	2016	Milion barrels per day	0.000009
...	...	...	...	...	...	...	...	...
223	India	Historical	EV charging points	EV	Publicly available fast	2022	charging points	4100.000000
224	India	Historical	EV charging points	EV	Publicly available slow	2022	charging points	6800.000000
225	India	Historical	Oil displacement Mbd	Trucks	EV	2022	Milion barrels per day	0.000001
226	India	Historical	Oil displacement, million lge	Trucks	EV	2022	Oil displacement, million lge	0.064000
227	India	Historical	Electricity demand	Trucks	EV	2022	GWh	0.240000

228 rows × 8 columns

Insight:

```
The data contains 228 rows and 8 columns.
```

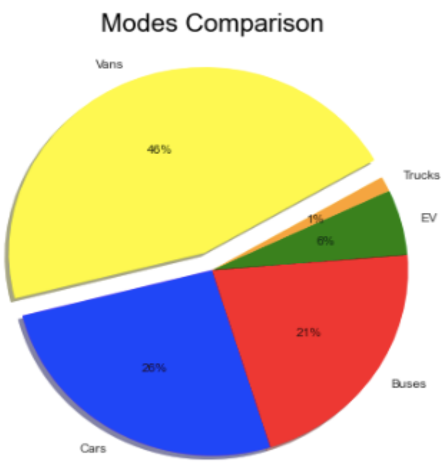
6.0 Exploratory Data Analysis

6.1 Univariate analysis

Below is the univariate analysis of mode and prowetrain of the EVs.

Univariate Analysis

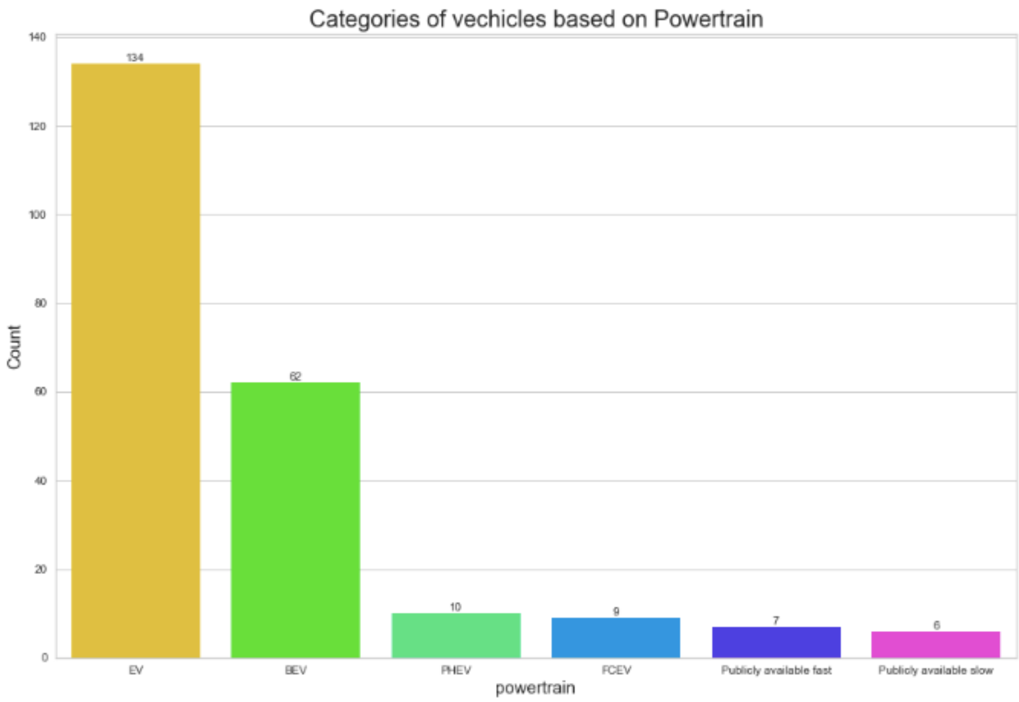
[53]:
```
plt.figure(figsize=(12,7))
plt.pie(df['mode'].value_counts(), labels= df['mode'].unique(),
        colors=['yellow','blue','red','green','orange'], autopct='%.0f%%', explode = [0.1, 0,0,0,0], sh
plt.title('Modes Comparison',fontsize=20, color = 'Black')
plt.show()
```



[ ]:
```
Insight:

    This data contains 46% information about vans , 26% about cars, 21% buses , 6% EV and 1% trucks.
```

```
plt.figure(figsize=(15,10))
f = sns.countplot(x=df['powertrain'],palette = 'hsv')
f.bar_label(f.containers[0])
plt.xlabel('powertrain',fontsize='15')
plt.ylabel('Count',fontsize='15')
plt.title('Categories of vechicles based on Powertrain', fontsize='20')
plt.show()
```



Insights:

```
From the above plot it is known that BEVs dominate the other categories in the Indian market.
```

6.2 Multi-Variate Analysis

Below are the insights from the multivariate analysis of yearly EV category data based on parameter

Vans:

'**EV sales share**' of Vans in the year peaked in 2021 with 42% which says out of all the vans sold 42% are EV vans.

'**EV Stock Share**' of vans is highest in 2021 22 and 23 saying that out of all vans on road currently used 8.8% are EV vans.

'**EV Sales**' of vans is highest in 2021, there after gradually decreased in 2022 and 2023.But overall 'EV Stock' of vans is rapidly increasing from 2020-23 which are all BEV(Battery Electric Vehicles).

'**Oil displacement**' is highest in the year 2022 which represents battery power is replaced by gasoline and oils products which is 0.00026 million barrels per day.

'**Electricity Demand**' of vans in the year 2023 is 41GWh(the amount of electricity needed to power the growing fleet of EVs) in the year 2023 which will be increasing as the number of EVs increase.

Cars:

'**EV sales share**' of Cars in the year peaked in 2023 with 2% which says out of all the cars sold 2% are EV cars.

'**EV Stock Share**' of Cars is highest in 2023 with 30% saying that out of all cars on the road, 30% are EV cars.

'**EV Sales**' of Cars is highest in 2023 with 82k vans sold. The trend is almost quadrapuled from 21 to 22 and doubled from 22 to 23.The sameway overall 'EV Stock' of cars is rapidly increasing from 2020-23 which are all BEV(Battery Eletric Vehicles).

'**Oil displacement**' is highest in the year 2023 which represents battery power is replaced by gasoline and oils products which is 0.0025 million barrels per day.

'**Electricity Demand**' of cars in the year 2023 is 450GWh(the amount of electricity needed to power the growing fleet of EVs) which will be increasing as the number of EVs increase.

Buses:

'**EV sales share**' of Buses in the year peaked in 2021 with 7.5% which says out of all the buses sold 7.5% are EV buses and this trend slightly decreases in 2022 and 2023.

'**EV Stock Share**' of Buses is highest in 2023 with 44% saying that out of all buses on the road, 44% are EV buses.

'**EV Sales**' of Buses is highest in 2023 with 3000 BEV buses sold. There are 100 FCEV sold in 2020 and 2022. The same way overall 'EV Stock' of buses is rapidly increasing from 2020-23 which 8500 are BEV(Battery Electric Vehicles) and 150 are FCEV.

'**Oil displacement**' is highest in the year 2023 which represents battery power is replaced by gasoline and oils products which is 0.0033 million barrels per day.

'Electricity Demand' of Buses in the year 2023 is 790GWh(the amount of electricity needed to power the growing fleet of EVs) which will be increasing as the number of EVs increase.

‘Charging Points’

The number of publicly available charging points was been increasing from 2017 and reached 4100 fast charging points and nearly 7000 slow charging points over the country in the year 2023.

Trucks:

'Oil displacement' trucks in the year 2022 is 1.1e^-0.6 million barrels per day.

'Electricity Demand' of trucks in the year 2022 is 0.24GWh(the amount of electricity needed to power the growing fleet of EVs) which will be increasing as the number of EVs increase.

7.0 Data Pre-Processing

This step includes encoding categorical variables and standardization.

One hot encoding technique is used in this analysis as we want binary data for analysis.

Data after encoding is like:

```
#Extract categorical columns from the dataframe
#Here we extract the columns with object datatype as they are the categorical columns
categorical_columns = ['parameter','mode','powertrain']

#Initialize OneHotEncoder
encoder = OneHotEncoder(sparse_output=False)

# Apply one-hot encoding to the categorical columns
one_hot_encoded = encoder.fit_transform(df[categorical_columns])

#Create a DataFrame with the one-hot encoded columns
#We use get_feature_names_out() to get the column names for the encoded data
one_hot_df = pd.DataFrame(one_hot_encoded, columns=encoder.get_feature_names_out(categorical_columns))

# Concatenate the one-hot encoded dataframe with the original dataframe
df_encoded = pd.concat([df, one_hot_df], axis=1)

# Drop the original categorical columns
df_encoded = df_encoded.drop(categorical_columns, axis=1)

# Display the resulting dataframe
```

```
df_encoded
```

parameter_EV charging points	parameter_EV sales	parameter_EV sales share	parameter_EV stock	parameter_EV stock share	parameter_Electricity demand	parameter_Oil displacement Mbd	...	mode_Cars	mode_EV	mode_Trucks	mod
0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	
0.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	
1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	
1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	
0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	1.0	
0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	
0.0	0.0	0.0	0.0	0.0	1.0	0.0	...	0.0	0.0	1.0	

```
df_encoded.describe().round(2)
```

	parameter_EV charging points	parameter_EV sales	parameter_EV sales share	parameter_EV stock	parameter_EV stock share	parameter_Electricity demand	parameter_Oil displacement Mbd	parameter_Oil displacement, million lge	mode_Buses	mode_Car
count	228.00	228.00	228.00	228.00	228.00	228.00	228.00	228.00	228.00	228.0
mean	0.06	0.18	0.14	0.18	0.14	0.04	0.14	0.14	0.26	0.4
std	0.23	0.38	0.34	0.38	0.34	0.18	0.35	0.35	0.44	0.5
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
25%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
50%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
75%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.0
max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.0

From the above table we can say that EV, BEV, cars, buses are having highest standard deviation which says that these are high variance Features.

8.0 Principal Component Analysis

Principal component analysis (PCA) is a dimensionality reduction and machine learning method used to simplify a large data set into a smaller set while still maintaining signific ant patterns and trends.

It includes following steps

8.1 Standardization

As our data values are between 0 ,1 there is no need for standardization before Principal Component Analysis(PCA).

8.2 PCA

- Creating PCA object
- Fit and transform the encoded data in the pca object
- The pca takes all the features and creates 'Principal components' which are new variables that are **constructed as linear combinations or mixtures of the initial variables**. These combinations are done in such a way that the new variables (i.e., principal components) are uncorrelated and most of the information within the initial variables is squeezed or compressed into the first components.
- As the encoded data has 19 columns there will 19 principal components created with decreasing importance from PC1-PC19 which can be known from the following PCA summary:

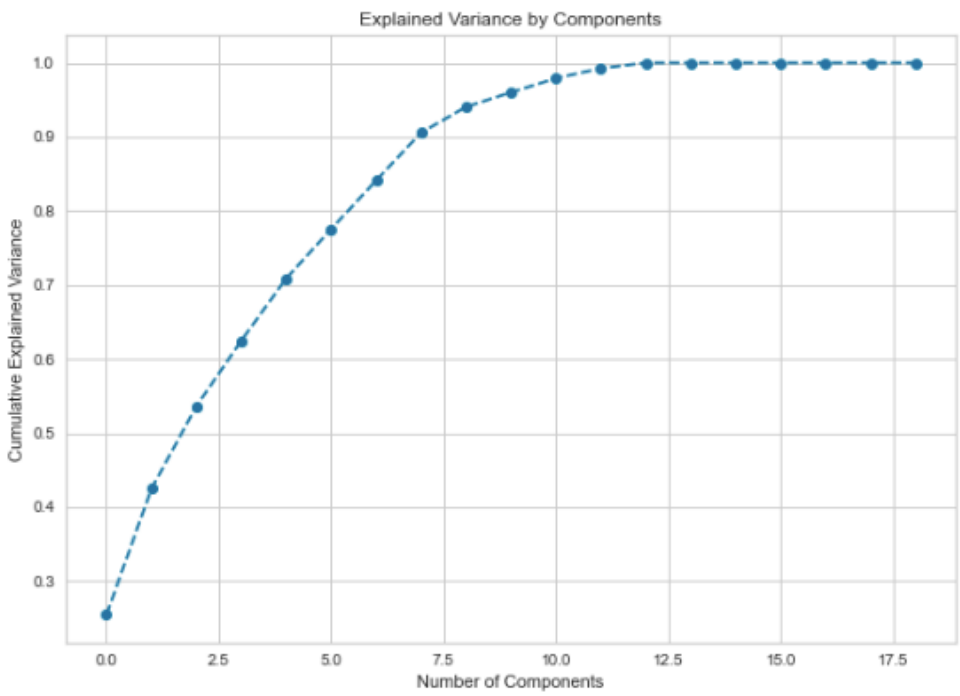
  From. The following picture 80% of the crucial data is covered from PC1 through PC6 which indicates these can be taken as the main components for segmentation of the data.

Principal Component Summary:

	Standard deviation	Proportion of Variance	Cumulative variance Ratio
PC1	0.54	0.25	0.25
PC2	0.36	0.17	0.42
PC3	0.24	0.11	0.54
PC4	0.19	0.09	0.62
PC5	0.18	0.08	0.71
PC6	0.14	0.07	0.78
PC7	0.14	0.07	0.84
PC8	0.14	0.06	0.91
PC9	0.07	0.03	0.94
PC10	0.04	0.02	0.96
PC11	0.04	0.02	0.98
PC12	0.03	0.01	0.99
PC13	0.02	0.01	1.00
PC14	0.00	0.00	1.00
PC15	0.00	0.00	1.00
PC16	0.00	0.00	1.00
PC17	0.00	0.00	1.00
PC18	0.00	0.00	1.00
PC19	0.00	0.00	1.00

- Deciding which components to keep for the analysis based on cumulative variance plot.

  The following is the cumulative variance plot for the components



The graph shows the amount of variance captured (on the y-axis) depending on the number of components we include (the x-axis).

A rule of thumb is to preserve around 80 % of the variance. So, in this instance, we decide to keep 6 components.

- Performing PCA with the chosen number of components, 6 in this case.

```
# We choose 6 components
pca = PCA(n_components=6)

# Fit the data with selected number of components
pca.fit(df_encoded)

# Transforming the data to the principal component space
scores_pca = pca.transform(df_encoded)
```

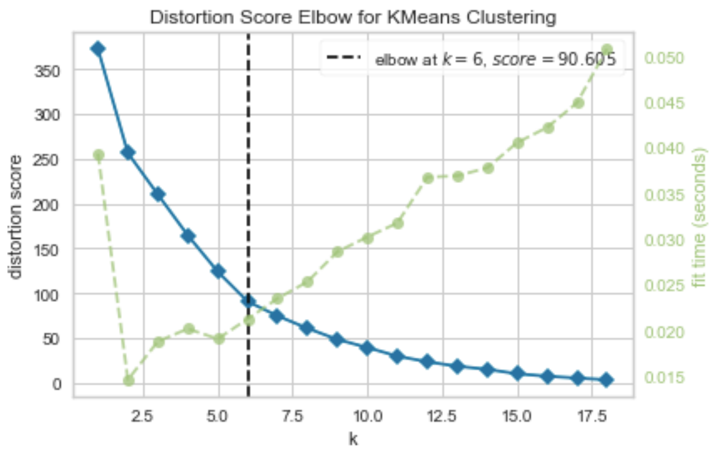
```
pd.DataFrame(scores_pca)
```

	0	1	2	3	4	5
0	-0.647358	0.423777	0.786291	-0.155544	-0.043884	-2.313166e-14
1	-0.647358	0.423777	0.786291	-0.155544	-0.043884	7.871313e-15
2	0.993061	0.568518	0.790250	-0.271742	-0.740787	5.212561e-16
3	0.987559	0.521078	0.882280	-0.310003	0.668354	-6.161104e-17
4	-0.654189	0.426815	0.788792	-0.148228	-0.043112	-7.071068e-01
...	...	...	...	...	...	...
223	0.270026	0.315399	0.164735	1.714337	0.056966	-5.240841e-16
224	0.267703	0.311236	0.161269	1.667466	0.055311	4.454403e-16
225	-0.623405	0.154184	0.056335	0.084688	0.007268	-7.071068e-01
226	-0.623405	0.154184	0.056335	0.084688	0.007268	7.071068e-01
227	-0.495317	0.126556	0.054687	0.149501	0.005981	1.067592e-15

9.0 Segmentation

After performing the PCA on chosen components we use K-Means Clustering to know t he number clusters to be considered for the target selection by fitting the pca scores of th e chosen component data.

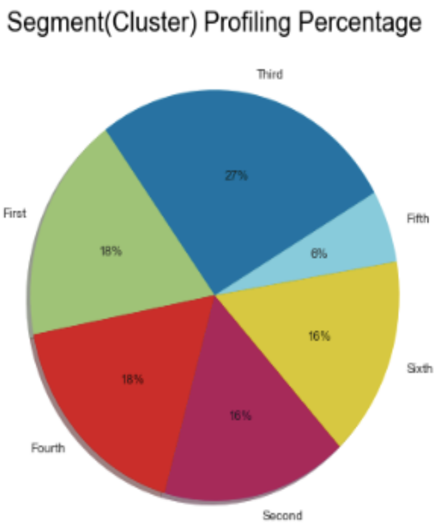
```
model1 = KMeans()
visualizer = KElbowVisualizer(model1, k=(1,19)).fit(scores_pca)
visualizer.show()
```



This plots says that the number of clusters to be considered is 6 as the graph gets stabilis es after a certain point which is called 'Elbow' in this particular graph, so got the name Elbow Visualizer.

9.1 Profiling Segments

```
]: plt.figure(figsize=(12,7))
   plt.pie(df_final['Segment'].value_counts(), labels= df_final['Segment'].value_counts().index,
           autopct='%.0f%%',  shadow = True, startangle=30)
   plt.title('Segment(Cluster) Profiling Percentage',fontsize=20, color = 'Black')
   plt.show()
```



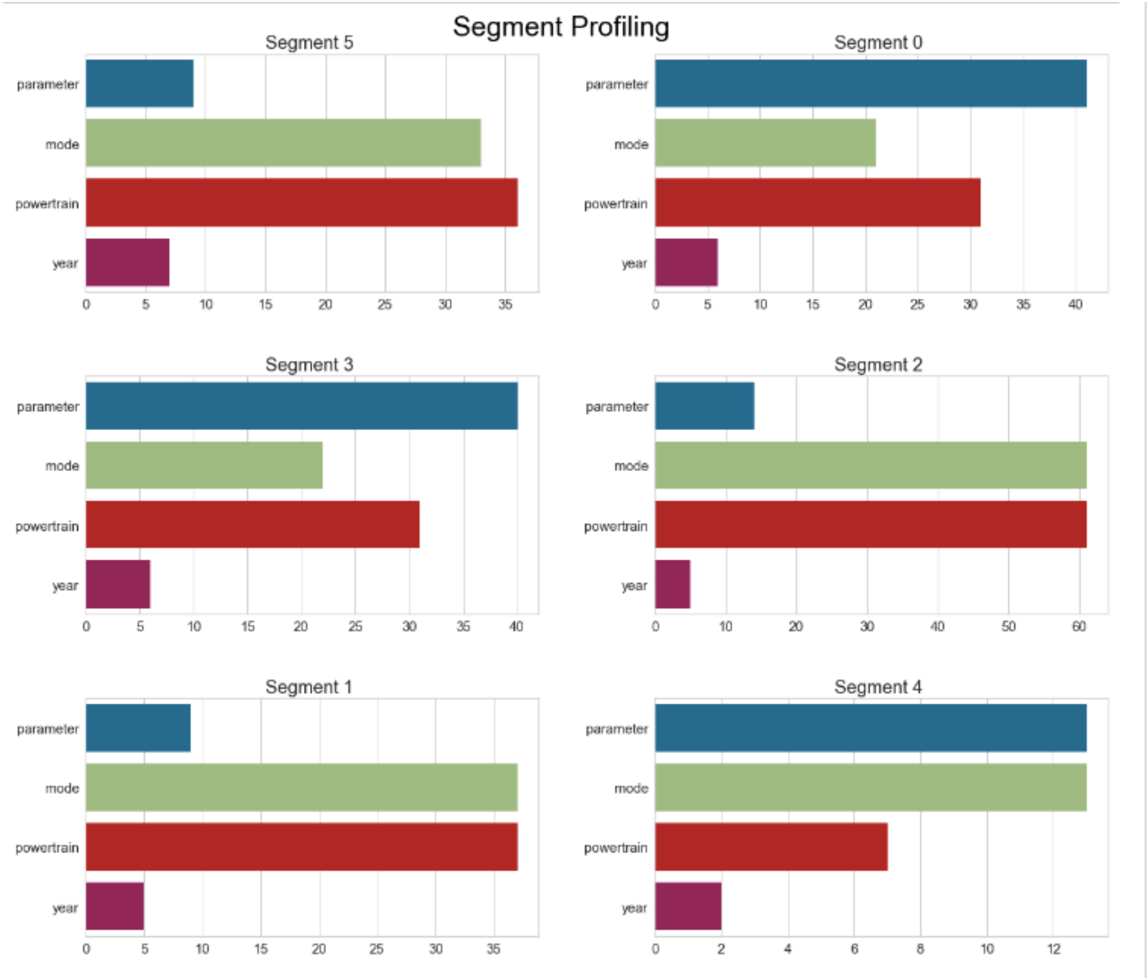
Insights:

```
Segment 0: First segments has 18% data.
Segment 1: Second segments has 16% data.
Segment 2: Third segments has 27% data.
Segment 3: Fourth segments has 18% data.
Segment 4: Fifth segments has 6% data.
Segment 5: Sixth segments has 16% data.

Segment 2 has the highest preference then segment 0,3.
```

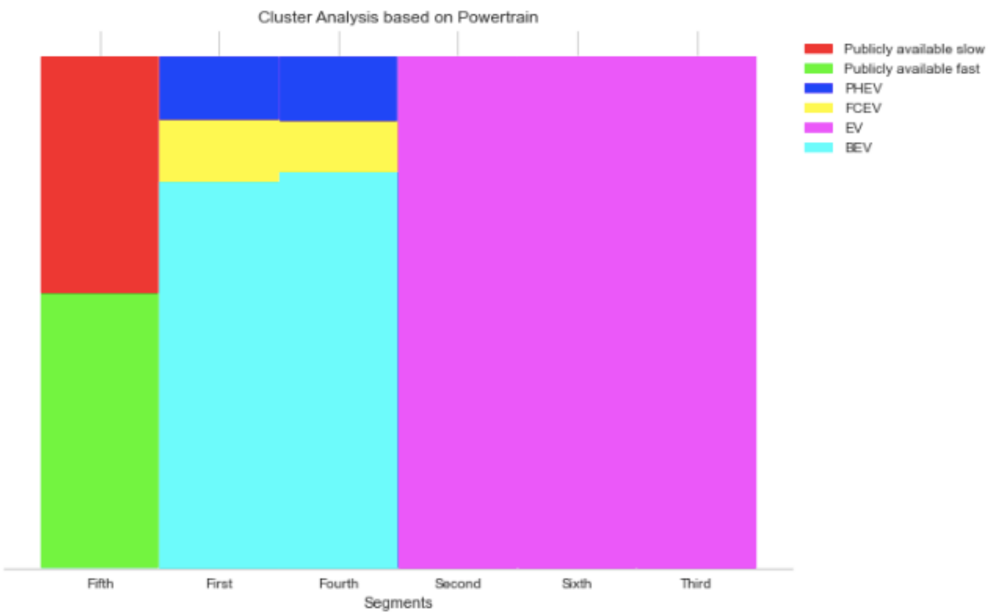
Segment 2 has 27% of data and is the dominating segment which says mode(Vans,Truck s,Cars,Buses,EV) and powertrain(EV,BEV,PHEV,FCEV) are the deciding variables.

Overall all the segments says **mode and powertrain will be Target Variables.**

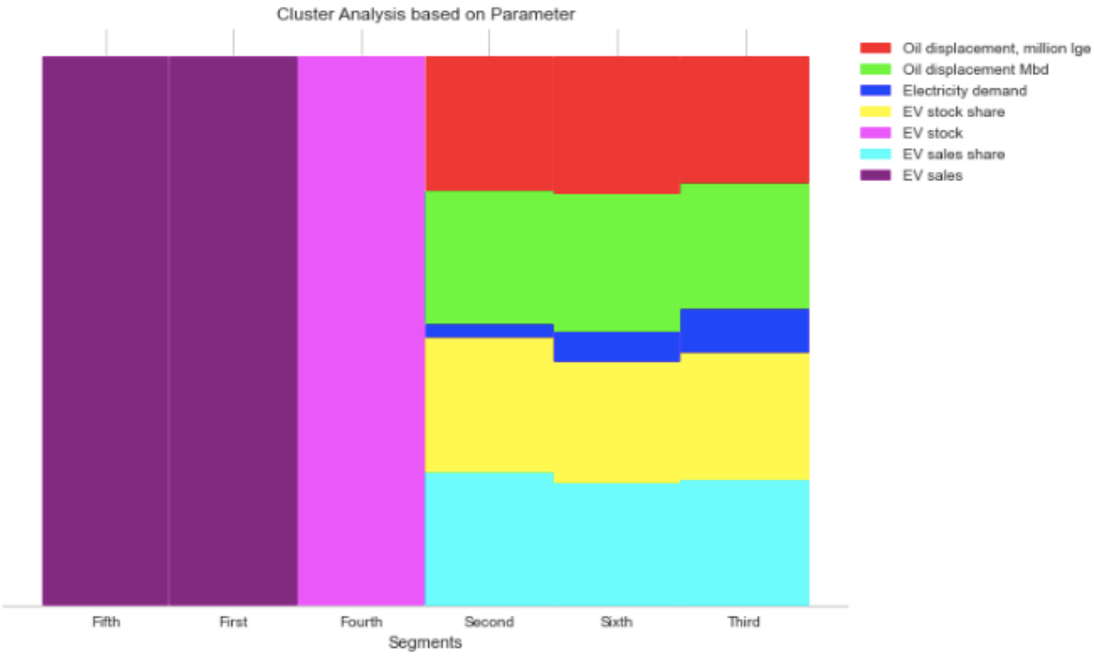


9.2 Segment Description: Detailing Segment Characteristics

This section offers a thorough overview based on insights drawn from various mosaic pl
ots and graphical representations.

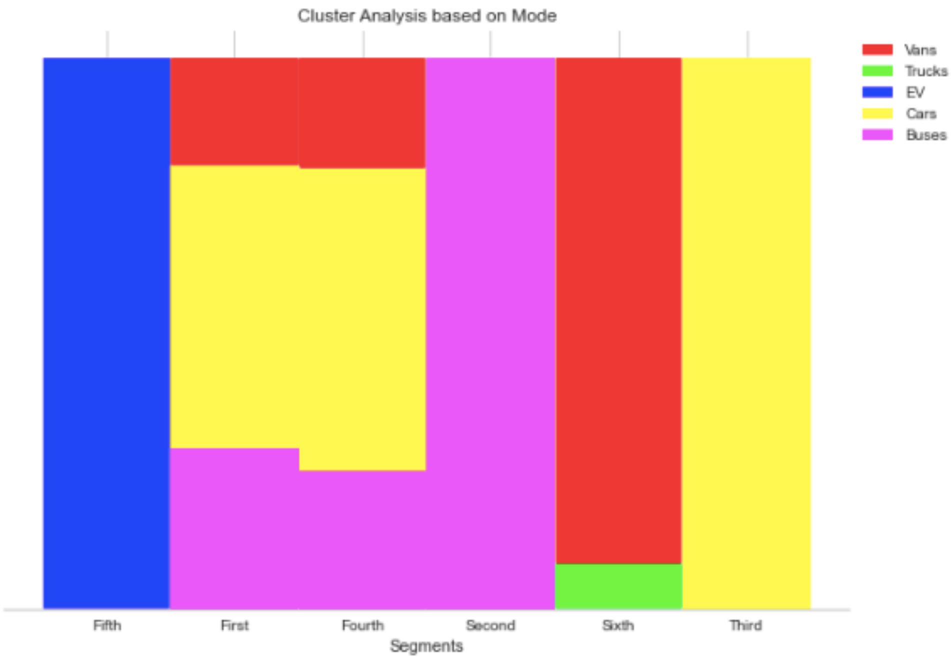


Concentrating on Segment 2 which has highest data covered whose cluster analysis reve als that it is dominated by all types of EV cars as the dominating mode.



The second dominating segments are segment0 and Segment 3 whose cluster analysis re veals that :

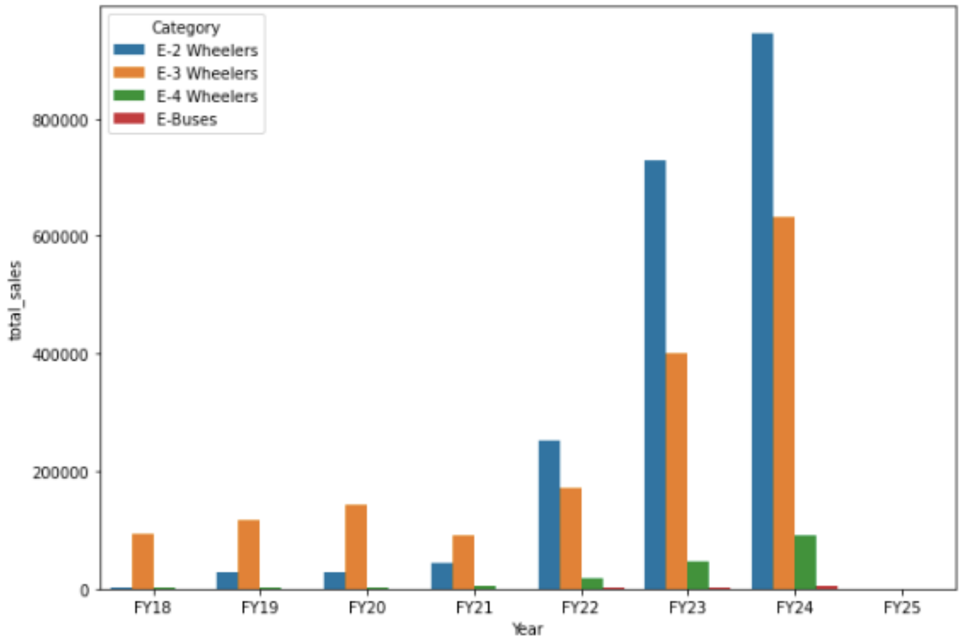
Segment 0, Segment 3 has more BEV vehicles especially cars, buses and vans.



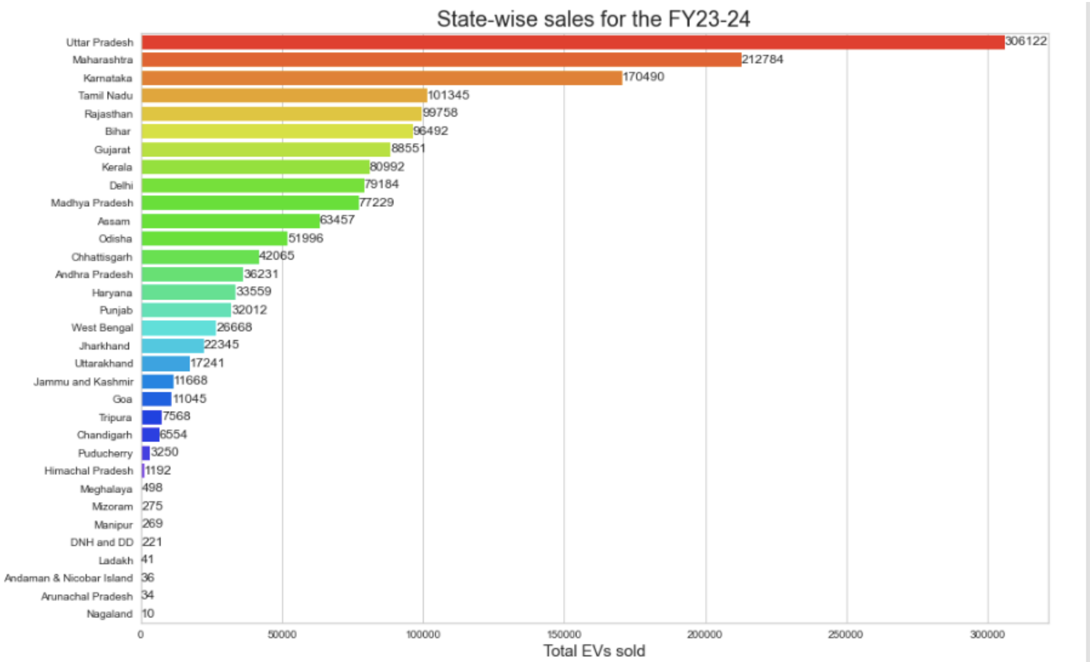
The over all segment profiling says us to concentrate on segments 2,0,3 which in turn says **all types of EV cars and BEV Cars, Buses and Vans are dominating the Indian Market.**

From other analysis like state wise sales and yearly sales in India:

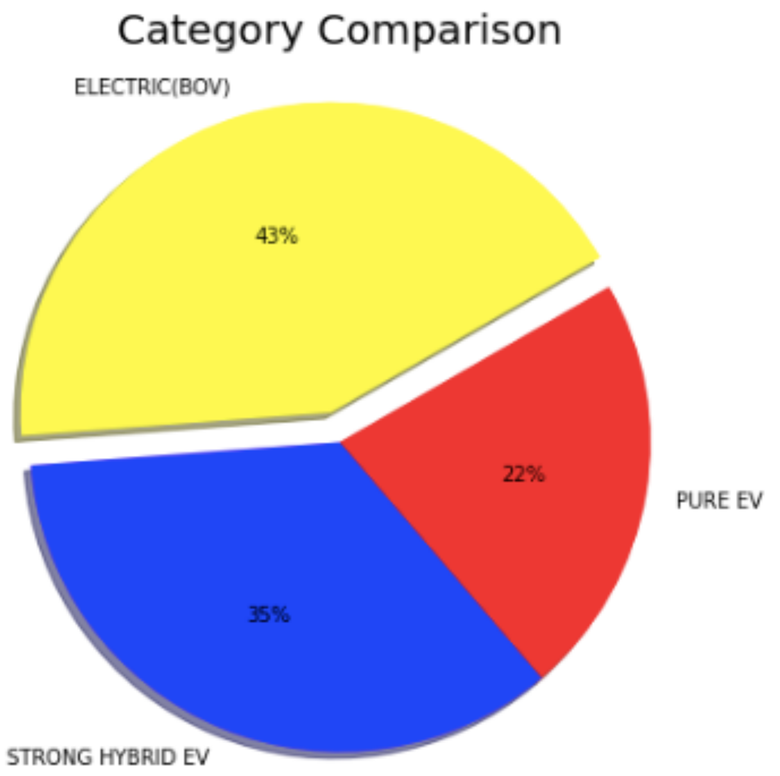
We can say that the sales of all types of Electric vehicles are increasing yearly out of which the 2-wheeler sales is high.



Uttar Pradesh, Maharashtra ,Karnataka are the top three states with highest sales.



Battery operated vehicles are dominating the market next comes StrongHybridVehicles.



10.0 Conclusion

In-depth segmentation analysis reveals that segment 2 is the best segment fo r a start-up to launch.

Which says launching all types of EV cars and 2-Wheelers will best for a st art-up in the dominating Indian states like Uttar Pradesh , Maharashtra and Karnataka .

GitLinks for Code:

https://github.com/katyayini0583/Market_Segmentation_Analysis_ElectricVehicles_India/blob/main/EV_Data_Analysis_India.ipynb

https://github.com/katyayini0583/Market_Segmentation_Analysis_ElectricVehicles_India/blob/main/State_Wise_EV_Sales_FY23-24_India.ipynb

https://github.com/katyayini0583/Market_Segmentation_Analysis_ElectricVehicles_India/blob/main/Yearly_EV_Sales_Analysis_India.ipynb