

CONTEXT:

Market Segmentation Analysis

EV Market

Team Shreya:

1.	Shreya Sharma
2.	Charan
3.	Soumabrata Manna
4.	Sahildeep Buttar

INTRODUCTION -

An electric vehicle (EV) is one that operates on an electric motor, instead of an internal-combustion engine that generates power by burning a mix of fuel and gasses. Therefore, such a vehicle is seen as a possible replacement for current-generation automobiles, in order to address the issue of rising pollution, global warming, depleting natural resources, etc. Though the concept of electric vehicles has been around for a long time, it has drawn a considerable amount of interest in the past decade amid a rising carbon footprint and other environmental impacts of fuel-based vehicles.

Electric vehicles have low running costs as they have fewer moving parts for maintenance and also are very environmentally friendly as they use little or no fossil fuels (petrol or diesel). While some EVs used lead acid or nickel metal hydride batteries, the standard for modern battery electric vehicles is now considered to be lithium-ion batteries as they have greater longevity and are excellent at retaining energy, with a self-discharge rate of just 5% per month. Despite this improved efficiency, there are still challenges with these batteries as they can experience thermal runaway, which has, for example, caused fires or explosions in the Tesla Model S, although efforts have been made to improve the safety of these batteries.

ACTUAL MARKET SITUATION FOR THE LAND OF EVs -

- India has a target for EV sales penetration - 30% for private cars, 70% for commercial vehicles, 40% for buses, and 80% for 2-wheelers.

CONTEXT:

- For this Central Government has reduced the GST price for EVs, from 12 % to 5 %, providing subsidies and incentives to the manufacturers which helped the 2 wheelers models to reduce the price to almost similar to the petrol models
- Market says that last year OLA Electric sold 85,000 models in just 2 days worth 11cr.
- EV two-wheeler sales have gone up to 145%
- Nitin Gadkari (Union Transport Minister) said that the prices of the Ev vehicles will be the same as the price of the petrol vehicles. This point will be a game changer since the prices are the same and with some extra features, there is a high possibility of EV sales in the market.
- Market says that last year between Jan - Jul there was an investment of 2,50,00,00,00,000 INR.
- In the upcoming future there is a new idea going to launch I.e battery swapping from the charging stations, in which one can change the battery for charging the EV models. There is no need to wait for charging at the stations in case of any emergency. But for this, the market needs a lot of investment
- Private players are also installing charging stations in malls, restaurants, sports grounds, and near offices.
- Indian Oil - 10k Charging stations in the next 3 years, Bharat Petroleum - 7k charging stations
- Even the petrol pumps are also installing powerful charging stations
- Real Market says that NHAI will install 700 charging stations by the year 2023 each at a distance of 40 - 60 km with solar powered technology
- Work is going on at a higher rate to install around 4k charging stations in around 25 states in INDIA with super fast charging technology, which can charge up to 40kms in just 10 mins.

Fermi Estimation

Wild Guess : Around 8-10% people will have electric vehicles by the end of 2023 in India.

Educated Guess :

Employment rate = it is the ratio of number of available labor force to the population of people in the working age.

CONTEXT:

I think there are about 1.5 billion Indians in the world. Let's assume the only people over 18 and under 60 work, assuming that they account for around 60% of the population then that would make 0.9 billion Indians in the working class.

Out of the 0.9 billion people not all are employed, assuming only 2022 had 45% employment rate that would bring the number around 405 million.

Since, not everyone can afford an electric vehicle, let's assume only people above middle class can afford an electric vehicles, that would be 40 million. Not everyone buys an electric vehicle. Let's assume out of these 40 million only 10 million are willing to buy an electric vehicle.

Variables and Formulas :

Let $E(x)$ be the employment rate of the year x (in %).

Let $P(x)$ be the population of the year x .

Let $A(x)$ be the number of available Labor in the year x .

Let r be the ratio of Indians between the age of 18 and 60 to the total population of India.

$$E(x) = (A(x) * 100) / (P(x) * r)$$

This formula will formulate the Employment ratio for the year x .

Gathering More Information :

Estimation for the population of the year 2022 can be obtained by the increase in population each year

$$P(2019) = 1.3676 \text{ billion}$$

$$P(2020) = 1.3786 \text{ billion}$$

$$P(2021) = 1.39199 \text{ billion}$$

$$P(2020) - P(2019) = 11 \text{ million}$$

$$P(2021) - P(2020) = 13.39 \text{ million}$$

the mean would be 12.195 million

$$\text{thus } P(2022) = 1.44185 \text{ billion}$$

$$\text{assuming } A(x) \text{ is constant every year} = 471,688,990$$

$$r = 0.6$$

$$C = 0.75$$

$$E(2022) = (471,688,990 / (1,441,850,000 * 0.6)) * 0.75$$

$$E(2022) = 42\%$$

CONTEXT:

Conclusion :

By this analysis, I conclude that by the end of the year 2023 there would a Employment rate of 42%. That would make 42% of 405 million i.e. 170 million. Out of these 170 million only 10% afford EV'S. So around 17 million people will have EV's by the end of 2023"

Data Sources

This is a dataset of electric vehicles.

One of the more popular data science datasets is the mtcars dataset. It is known for its simplicity when running analysis and visualizations.

When looking for simple datasets on EVs, there don't seem to be any. Also, given the growth in this market, this is something many would be curious about. Hence, the reason for creating this dataset.

For more information, please visit the data source below.

<https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset>

TASKS:

Some basic tasks would include

1. Which car has the fastest 0-100 acceleration?
2. Which has the highest efficiency?
3. Does a difference in power train affect the range, top speed, and efficiency?
4. Which manufacturer has the most number of vehicles?
5. How does price relate to rapid charging?

CONTENT:

I've included two datasets below:

1. 'ElectricCarData_Clean.csv'
-- original pulled data.
2. 'ElectricCarData_Norm.csv'
-- units removed from each of the rows
-- rapid charge has a binary yes/no value

CONTEXT:

Column Name | Description
Brand | Manufacturer of the vehicle
Model | Model name
Accel | Acceleration as 0-100 km/h
TopSpeed | The top speed in km/h
Range | Range in km
Efficiency | Efficiency Wh/km
FastCharge | Charge km/h
RapidCharge | Yes / No
PowerTrain | Front, rear, or all wheel drive
PlugType | Plug type
BodyStyle | Basic size or style
Segment | Market segment
Seats | Number of seats
PriceEuro | Price in Germany before tax incentives

The point of both is to have users practice some data cleaning.

CREDITS:

There are two credits and sourcing that need to be mentioned:

1. *Datasource*: ev-database.org/
2. *Banner image*: freepik - author - 'macrovector'

UPDATES:

There will be future updates when we can attain additional data.

Data Pre-processing

For data pre-processing we mostly used the pandas library which is **Pandas is an open-source library that is made mainly for working with relational or labeled data both easily and intuitively**. It provides various data structures and operations for manipulating numerical data and time series. This library is built on top of the NumPy library.

All the dataset features were already perfectly scaled.

Segment Extraction -

- **LINEAR REGRESSION -**

CONTEXT:

Linear regression shows the linear relationship between the independent(predictor) variable i.e. X-axis and the dependent(output) variable i.e. Y-axis, called linear regression.

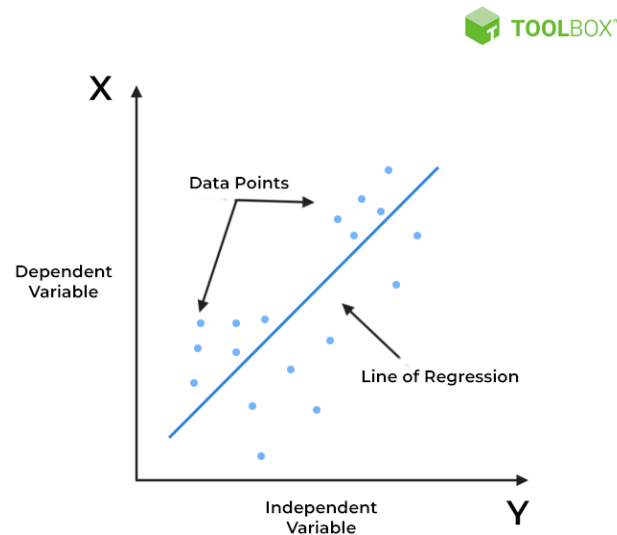


Fig - Basic understanding of Linear Regression in graphical format

The above graph presents the linear relationship between the output(y) variable and predictor(X) variables. The blue line is referred to as the *best fit* straight line. Based on the given data points, we attempt to plot a line that fits the points the best.

To calculate best-fit line linear regression uses a traditional slope-intercept form which is given below,

$$Y_i = \beta_0 + \beta_1 X_i$$

where Y_i = Dependent variable, β_0 = constant/Intercept, β_1 = Slope/Intercept, X_i = Independent variable.

The goal of the linear regression algorithm is to get the best values for B_0 and B_1 to find the best fit line. The best fit line is a line that has the least error which means the error between predicted values and actual values should be minimum.

CONTEXT:

The best fit line is a line that fits the given scatter plot in the best way. Mathematically, the best fit line is obtained by minimizing the Residual Sum of Squares(RSS).

CODE IMPLEMENTATIONS -

Brand	Model	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH	RapidCharge	PowerTrain	PlugType	BodyStyle	Segment	Seats	PriceEuro	lnr(10e3)
Tesla	Model 3 Long Range Dual Motor	4.6000	233	450	161	940	1	AWD	Type 2 CCS	Sedan	D	5	55480	4615.936
	wagen ID.3 Pure	10.0000	160	270	167	250	0	RWD	Type 2 CCS	Hatchback	C	5	30000	2496.000
	olestar 2	4.7000	210	400	181	620	1	AWD	Type 2 CCS	Liftback	D	5	56440	4695.808
	BMW iX3	6.8000	180	360	206	560	1	RWD	Type 2 CCS	SUV	D	5	68040	5660.928
	Honda e	9.5000	145	170	168	190	1	RWD	Type 2 CCS	Hatchback	B	4	32997	2745.350

Fig - The Original Dataset

	AccelSec	Range_Km	TopSpeed_KmH	Efficiency_WhKm	RapidCharge	PowerTrain
0	4.6000	450	233	161	1	2
1	10.0000	270	160	167	0	0
2	4.7000	400	210	181	1	2
3	6.8000	360	180	206	1	0
4	9.5000	170	145	168	1	0

Fig - The final dataset used for linear regression after data preprocessing

STEPS USED IN LINEAR REGRESSION ALGORITHMS -

- The data is being divided between independent and dependent variables
- The target variable is the price (**PriceEuro**)i.e the dependent variable.

CONTEXT:

- After checking the dataset, its shape, and parameters, it is being split between the training and testing data using the `train_test_split` method with a test size of 30% and training size of 70%.
- The Linear Regression model is being fitted and the accuracy and r^2 score of the training data is being calculated
- Following the above step the accuracy and r^2 score of the testing data is being calculated
- Finally it is being compared and it is concluded that the accuracy and the r^2 score of testing data are being better than training data.

- **K MEANS CLUSTERING -**

The K-means clustering algorithm computes centroids and repeats until the optimal centroid is found. It is presumptively known how many clusters there are. It is also known as the flat clustering algorithm. The number of clusters found from data by the method is denoted by the letter 'K' in K-means.

In this method, data points are assigned to clusters in such a way that the sum of the squared distances between the data points and the centroid is as small as possible. It is essential to note that reduced diversity within clusters leads to more identical data points within the same cluster.

CONTEXT:

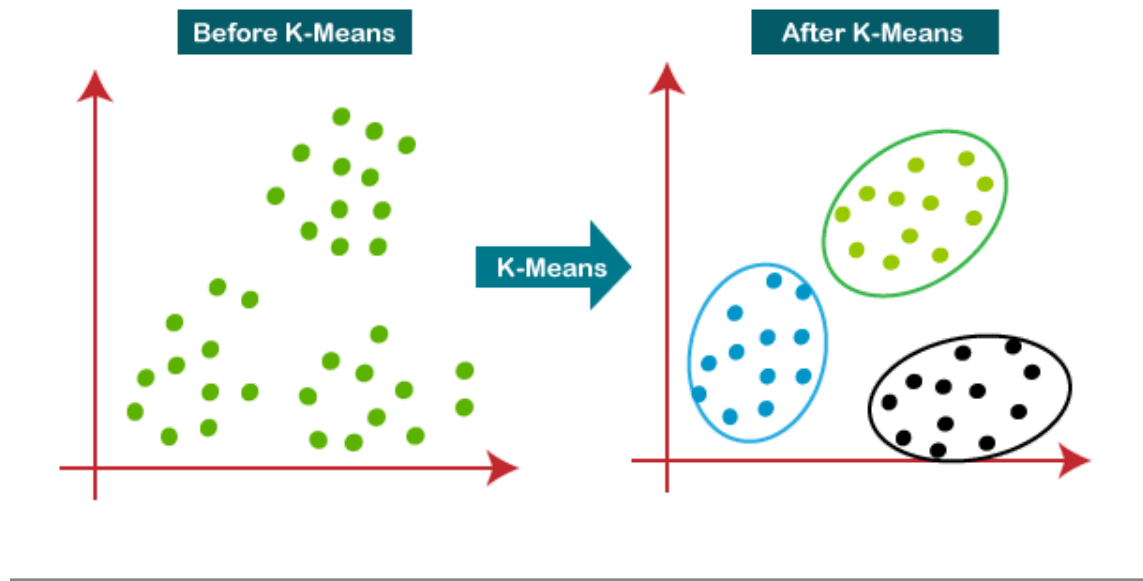


Fig - Basic understanding of K means clustering

- **ELBOW METHOD -**

In the Elbow method, we are actually varying the number of clusters (K) from 1 – 10. For each value of K , we are calculating WCSS (Within-Cluster Sum of Square). WCSS is the sum of the squared distance between each point and the centroid in a cluster. When we plot the WCSS with the K value, the plot looks like an Elbow. As the number of clusters increases, the WCSS value will start to decrease. WCSS value is largest when $K = 1$. When we analyze the graph we can see that the graph will rapidly change at a point and thus creating an elbow shape. From this point, the graph starts to move almost parallel to the X-axis. The K value corresponding to this point is the optimal K value or an optimal number of clusters.

STEPS INVOLVED IN K-MEANS CLUSTERING ALGORITHMS -

CONTEXT:

- The features parameters are being chosen for implementing k means algorithms and are being separated.
 - The feature parameter uses StandardScaler to resize the distribution of values so that the mean of the observed values is 0 and the standard deviation is 1 following fit_transform used on the training data so that we can scale the training data and also learn the scaling parameters of that data.
 - Elbow method graph is being plotted for selecting the number of features
-

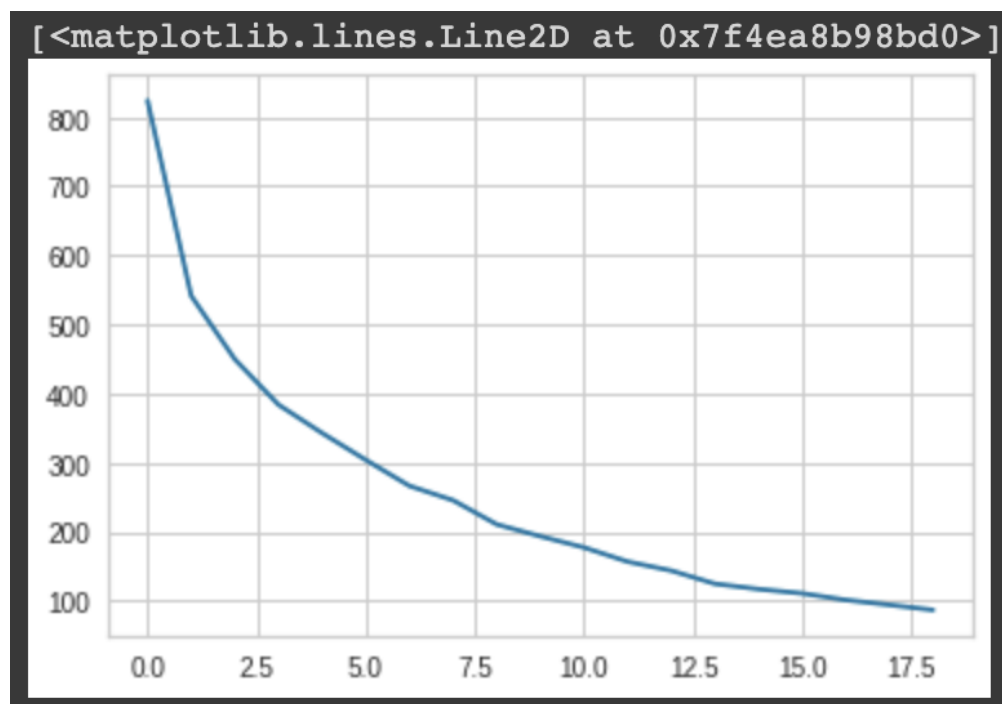


Fig - Elbow method graph

- After the elbow method graph visualization the k means clustering is being fitted with a number of clusters equal to 3 .

CONTEXT:

- Then the labels are assigned to each data point and the k-means inertia is being calculated.
 - After that the number of iterations that k means algorithms run to get a minimum within-cluster sum of squares is calculated.
 - Following the above step the location of the centroids on each cluster is being printed in the form of an array.
 - Finally the data and the clusters are plotted in a graphical format.
-

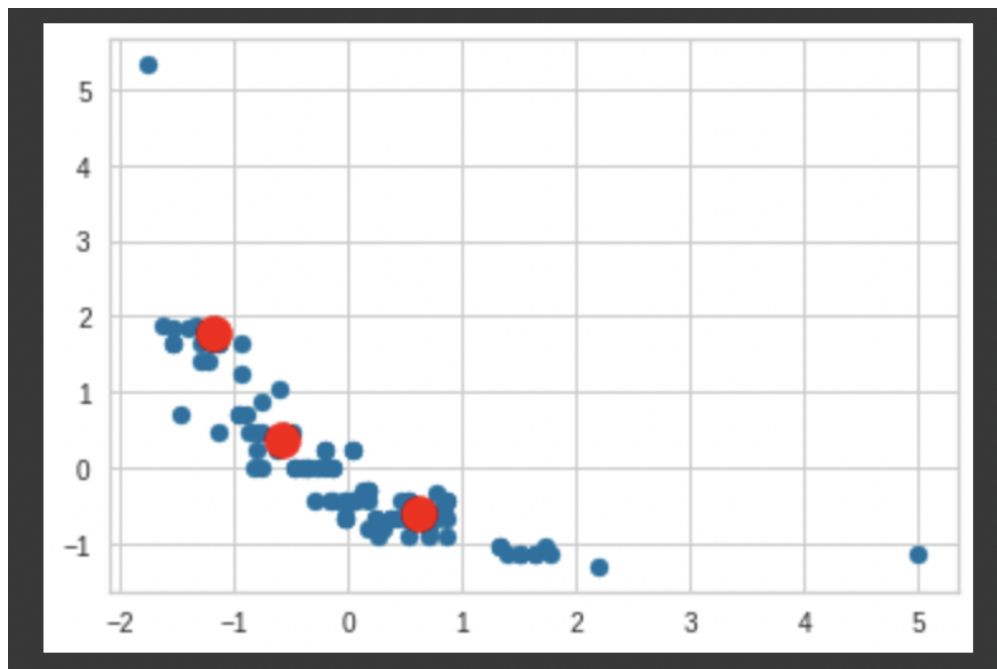


Fig - Unique clusters are being plotted

- **PRINCIPAL COMPONENT ANALYSIS (PCA) -**

CONTEXT:

Principal component analysis (PCA) is the process of computing and analyzing the essential features of data and utilizing them to modify the foundation of the data, often just using the first few critical features and not using the rest.

In the modern machine learning processes, PCA is one of the most widely used multivariate statistical approaches. It is a statistical approach with the broader field of feature analysis that has been commonly employed in pattern recognition and signal processing.

It is one such technique by which dimensionality reduction(linear transformation of existing attributes) and multivariate analysis are possible. It has several advantages, which include reduction of data size(hence faster execution), better visualizations with fewer dimensions, maximizes variance, reduces overfitting, etc

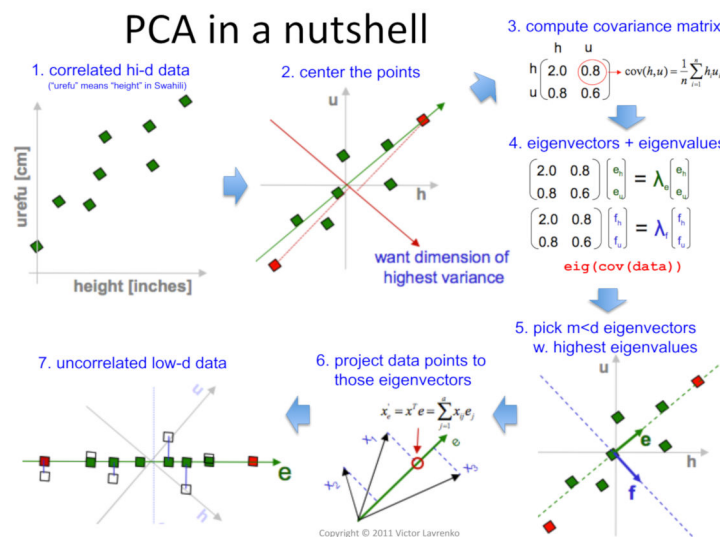


Fig- Basic understanding of PCA

PRINCIPAL COMPONENT ANALYSIS STEPS -

CONTEXT:

- **STANDARDIZATION -**

Standardizing is a method of scaling the data. If the different features in a given data set are on different scales, the corresponding variances will be different, and all features' contributions will not be the same. Variables with high variance will dominate the variables with low variance resulting in biased output.

- **COVARIANCE MATRIX -**

Computation of covariance matrix will help us understand how variables are varying with each other. Also, the covariance matrix is used to calculate eigenvectors and eigenvalues.

- **COMPUTE EIGEN VECTORS AND EIGEN VALUES -**

Eigenvectors are principal components and Eigen Values tell us about the % of information (variance) explained.

The formula for Eigenvector and Eigenvalues calculation

$$Av = \lambda v$$

where

A = Matrix

v = Eigen vector

λ = Eigen Value

CONTEXT:

The above formula is used to compute Eigenvectors and Eigen Values. Here A is any matrix, in our case, we will put our covariance matrix instead of A. Eigenvalue is a *scalar*.

- **FEATURE VECTOR -**

A feature vector is a matrix of all the Eigenvectors which we have decided to keep column-wise. If we choose to keep *m* Eigenvectors out of *n*, the final data will have m dimensions.

- **ALIGN / RECAST DATA ALONG PRINCIPAL COMPONENT -**

The feature vector with Eigen Vectors(Directions) is used to align/reorient the data from original axes to new principal component axes.

$$\text{Final Data} = \text{Feature Vector}^T * \text{Standardized Original Data}^T$$

CONTEXT:

STEPS INVOLVED IN CODING PCA ALGORITHMS -

- The first step is started with the number of components i.e equal to 2
- PCA algorithm is being fitted with fit_transform.
- Following the above 2 steps PCA data frame is being created.
- Number of unique clusters is being presented with value_counts method and is being added to the dataframe

	pca1	pca2	cluster
0	2.4292	-0.5546	2
1	-2.3225	-0.3454	1
2	1.5879	0.0089	0
3	0.2910	-0.0001	0
4	-2.6027	-0.6265	1
...
98	-0.2972	0.4467	1
99	2.3350	0.6307	0
100	0.7806	0.4268	0
101	1.5409	0.6988	0
102	0.9151	0.2615	0
103 rows x 3 columns			

Fig - Final Dataframe with Clusters

CONTEXT:

- Seaborn python library is being imported and finally the complete graph for the pca_df data frame is being plotted.

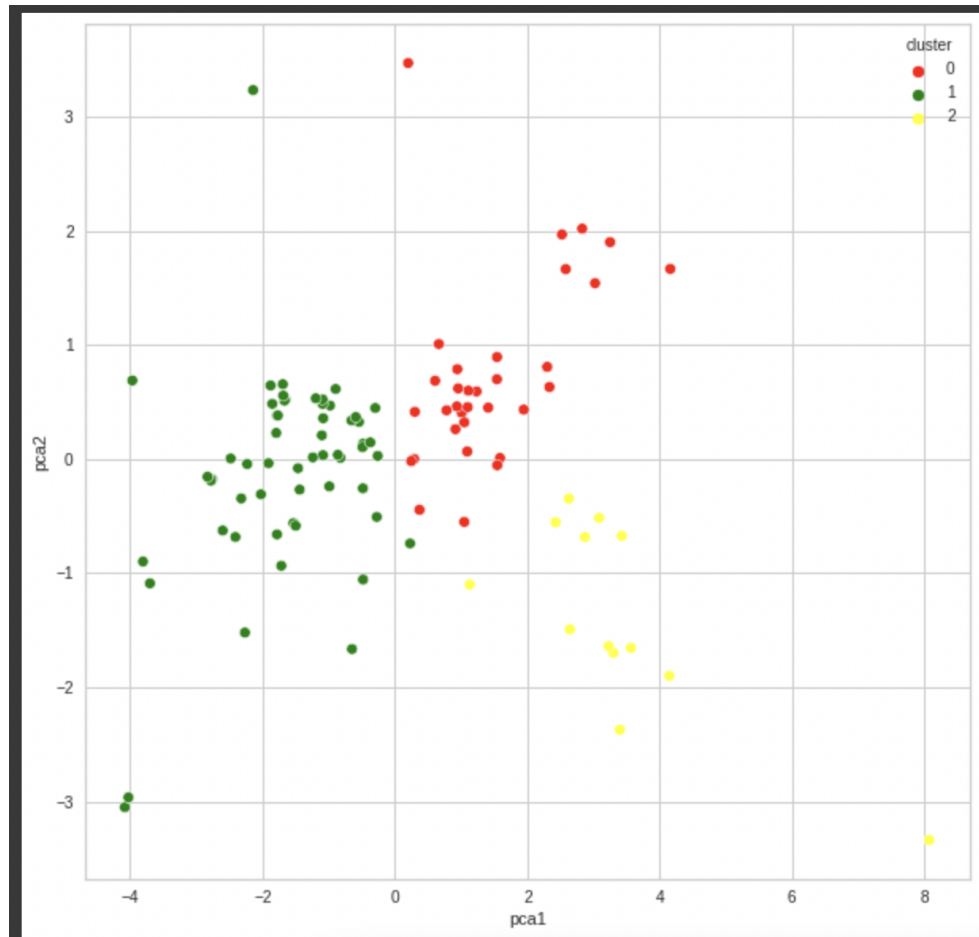


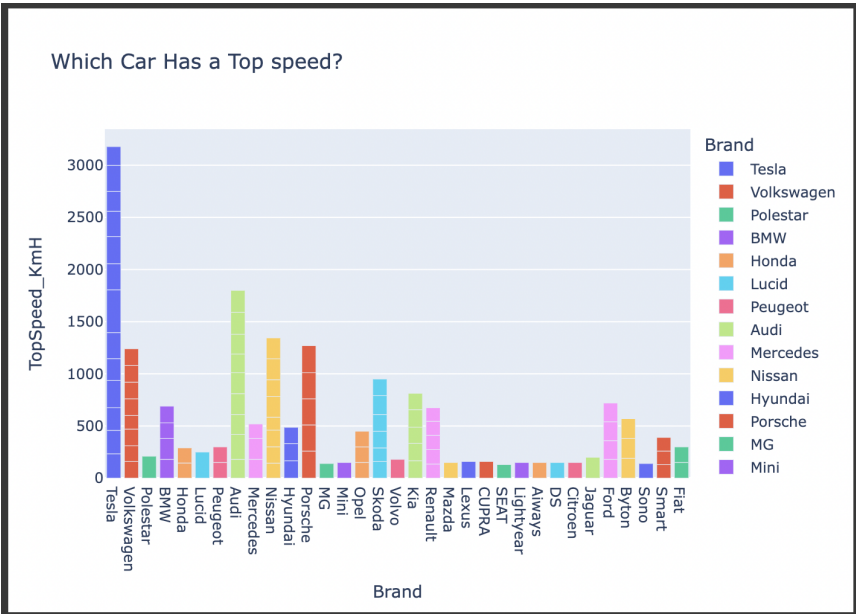
Fig - Final PCA graph

Profiling and Describing Potential Segments

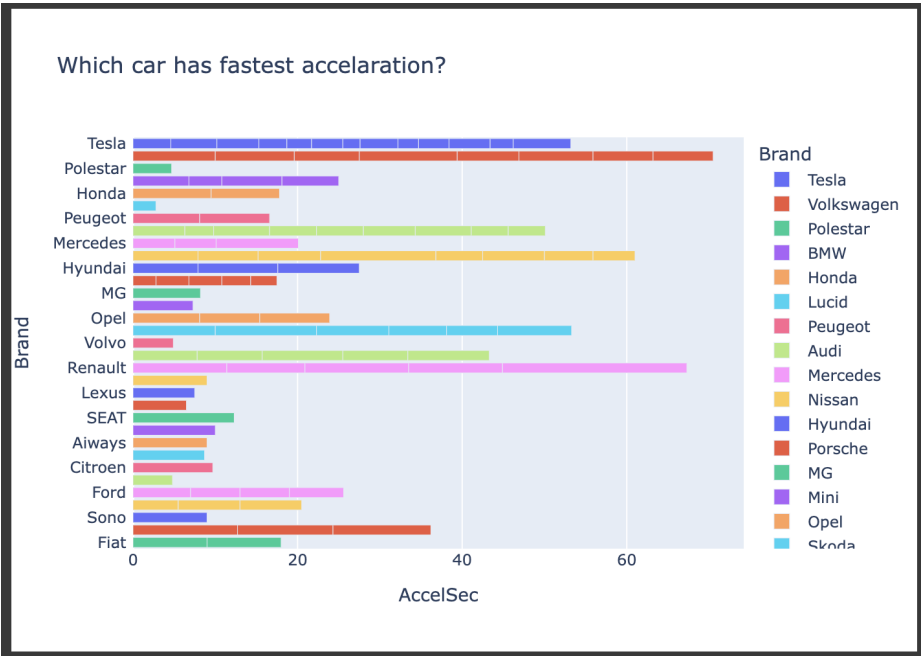
The Basic features and segments required for entering the electronic vehicle market are being described in a graphical and visualization format for easy and detailed understanding.

CONTEXT:

- TOP SPEED

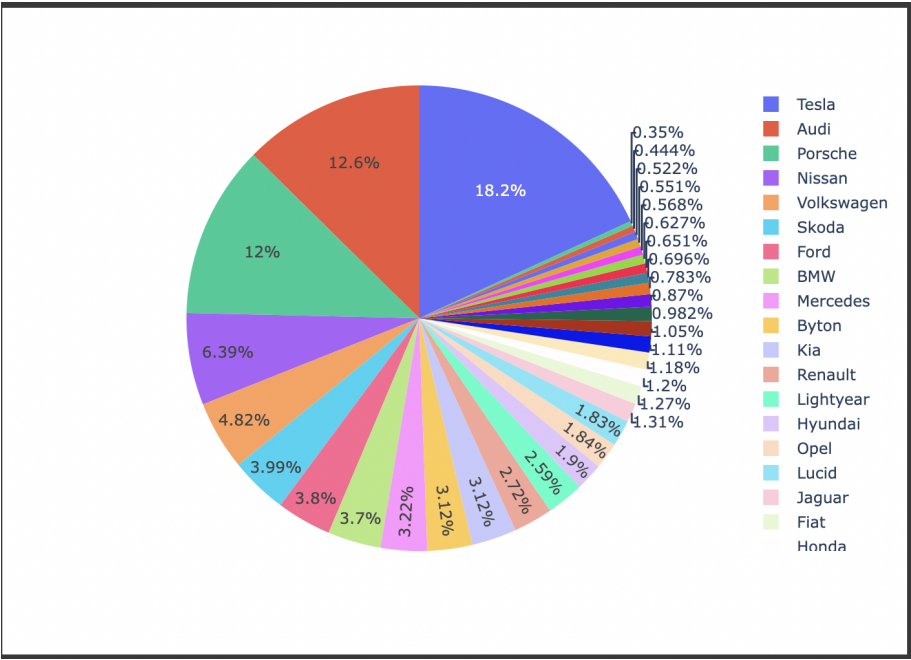


- FAST ACCELERATION -

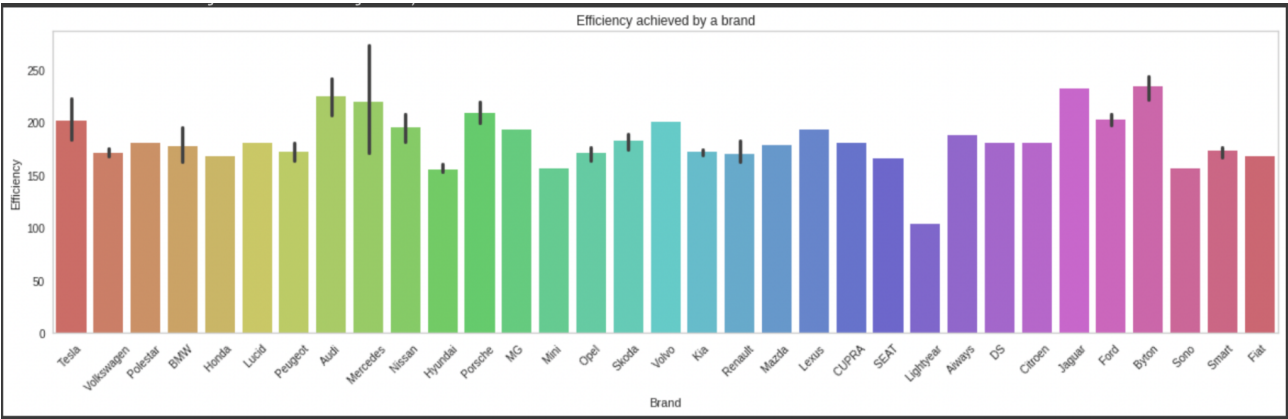


CONTEXT:

● BRAND VS PRICE PIE CHART VISUALIZATION -

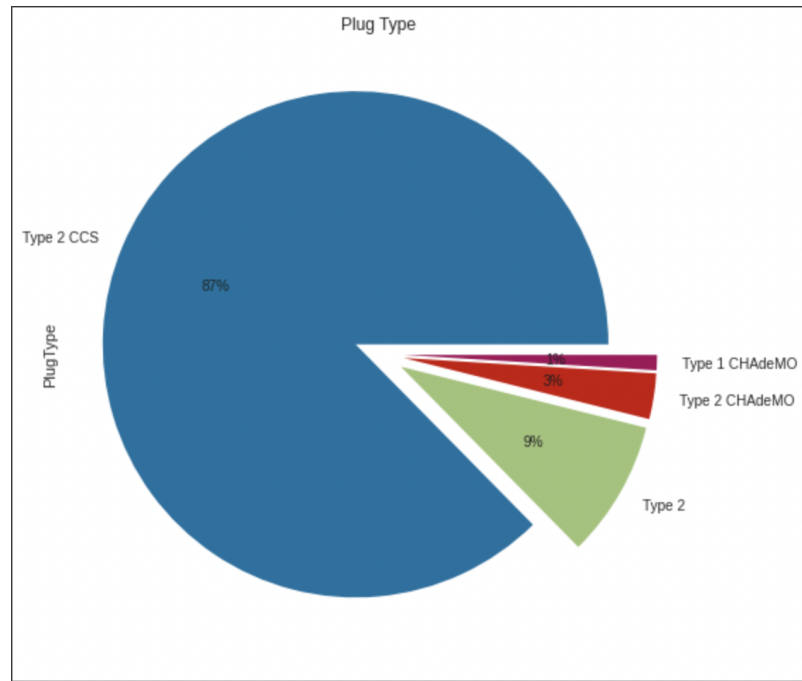


● EFFICIENCY ACHIEVED BY A BRAND -

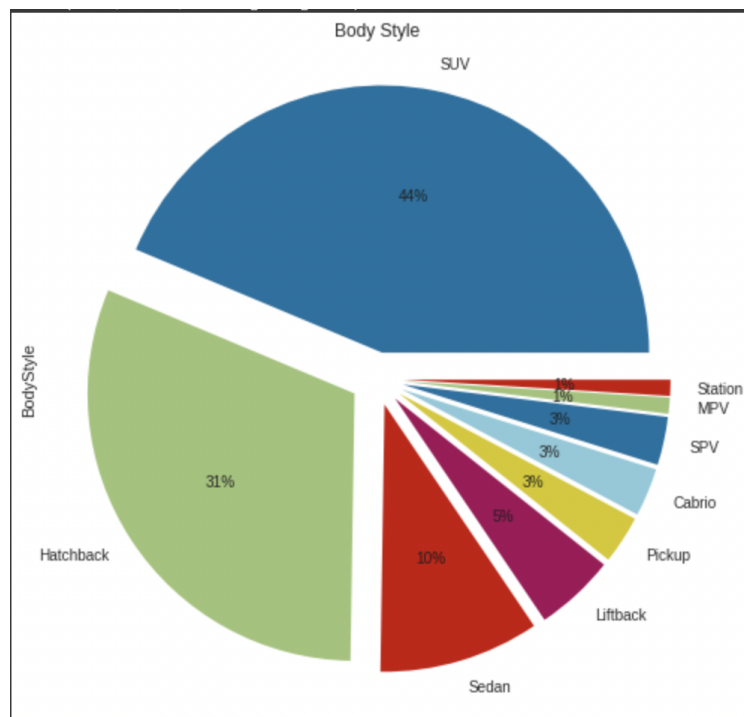


CONTEXT:

- PLUG TYPE FOR CHARGING -



- BODY STYLE -



CONTEXT:

Selection of Target Segment (Explanation about the Market Segmentation and the basic reason behind it)-

Market segmentation involves dividing a large homogenous market of potential customers into clearly identifiable segments. Customers are divided based on meeting certain criteria or having similar characteristics that lead to them having the same product needs. Segments are made up of customers who will respond similarly to marketing strategies. They share common interests, needs, wants, and demands.

The different types of market segmentation -

- **Demographic Segmentation** - This is the most common type of segmentation. A target audience is divided based on qualities such as age, gender, occupation, education, income, and nationality. Demographic segmentation is the easiest way to divide a market. Mixing demographic segmentation with another type of market segmentation can help to narrow your market down even further. The information required for demographic segmentation is easy to gather and doesn't cost a company too much to obtain.
- **Behavioral Segmentation** - A company can segment its market based on consumers' behaviors. Dividing your target audience based on their behaviors allows you to create specific messaging that will accommodate those behaviors.
 -
 - 1. What actions were taken on a website?
 - 2. What are their online shopping habits?
 - 3. How loyal are they to the brand/ product?
 - 4. What is the usage rate of your product?
 - 5. What need is a consumer trying to satisfy?

This information is relevant because it's directly related to how a consumer interacts with your products. Therefore, marketers can market more effectively to customers by knowing their behaviors.

- **Geographic Segmentation** - This involves splitting up a market based on location. Even though this is a basic form of segmentation it is highly effective. Knowing where a customer is located can help a company better understand the needs of their customers and companies can then target customers with location-specific ads. Dividing a segment based on the characteristics of their location, allows marketers to be even more specific with their targeting and messaging.

CONTEXT:

- **Psychographic Segmentation** - This form of segmentation is very similar to demographic segmentation however, it deals with characteristics that are related to mental and emotional attributes. Psychographic segmentation divides a group of customers based on their personality traits, values, interests, attitudes, and lifestyles. However, psychographics gives marketers valuable insights into customers' motives, preferences, and needs. It is psychographic information that informs you why people purchase or don't purchase a product or service.

The basic benefits of Market Segmentation -

- **Greater company focus -**

When a company has identified specific market segments, it helps them to focus on what segments they want to target with specific products/ services/ content/ blogs and campaigns. When a company focuses on specific segments, they ensure they are targeting the right segment with the right product which will see the greatest ROI.

- **better serve a customer's needs and wants -**

Having defined segments enables companies to satisfy a variety of customer needs by offering different bundles and incentives. Different forms and promotional activities will be used for different segments based on that segment's needs/ wants and characteristics.

- **Market competitiveness -**

When a company is focusing on a specific segment, its market competitiveness increases. Which in turn will lead to a higher ROI. The company is focused on specific segments and learns everything they need to know about that segment, to market their products to them.

- **Market expansion -**

With geographic segmentation as discussed earlier, the market expansion is possible immediately. When a company understands its segments and how to market to a segment in a particular location, it can expand immediately into another nearby location. If segmentation is based on demographics, then once the company knows its demographic segment it can expand in that segment with similar products.

- **Targeted communication -**

Even when product features and benefits are the same, it is important for companies to target segments with specific communication. For example, if your segment was senior engineers, they may respond better to technical information about a product in the form

CONTEXT:

of white papers or infographics, but a project manager might respond better to information regarding cost savings, efficiencies, etc in the form of a blog, case study or video. The key is to understand your segments and target communication relevant to them on the relevant platforms.

The main reason behind segmentation -

Segmentation helps marketers to be more efficient in terms of time, money, and other resources. Market segmentation allows companies to learn about their customers. They gain a better understanding of customers' needs and wants and therefore can tailor campaigns to customer segments most likely to purchase products.

Customizing The Market Mix:

Market Mix basically is a strategy that might be followed by a company to promote its product in a market. There are 4 P's of the marketing mix:

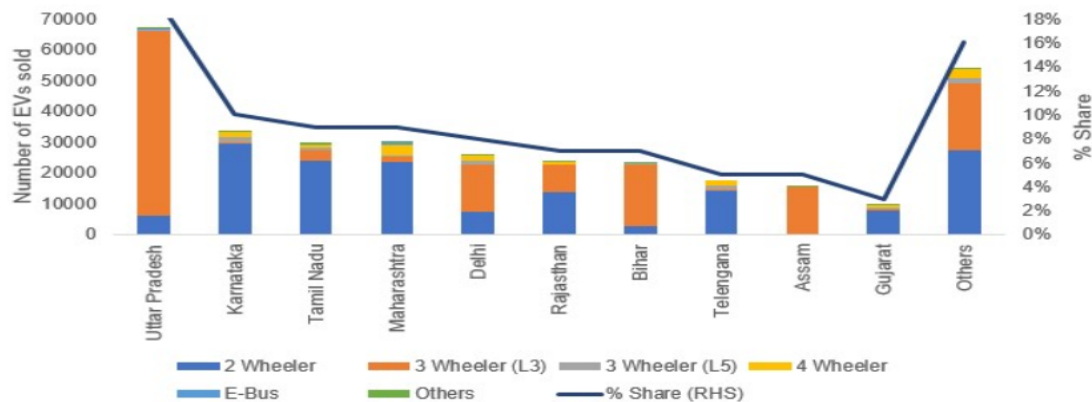
- Product
- Price
- Place
- Promotion

As we are aware of the fact that Electric vehicles run on an electrical motor instead of a combustion engine. These Mainly consist of rechargeable batteries and these features make them faster and a lot lighter to drive, but mainly electric vehicles are considered way better than the usual fuel-derived vehicles as they are zero-emission vehicles. They are considered to be the future of vehicles as a whole. They have been growing at a tremendous rate. Fuel-derived vehicles cause a lot of depletion of natural resources and are to be replaced by electric vehicles in the upcoming few years. As per India Energy Storage Alliance (IESA), the Indian EV industry is expected to expand at a CAGR of 36%.

CONTEXT:

wheeler segment and four-wheeler segment, respectively.

State -Wise-EV Sales Trend in 2021



Above is the state-wise distribution of the EV market.

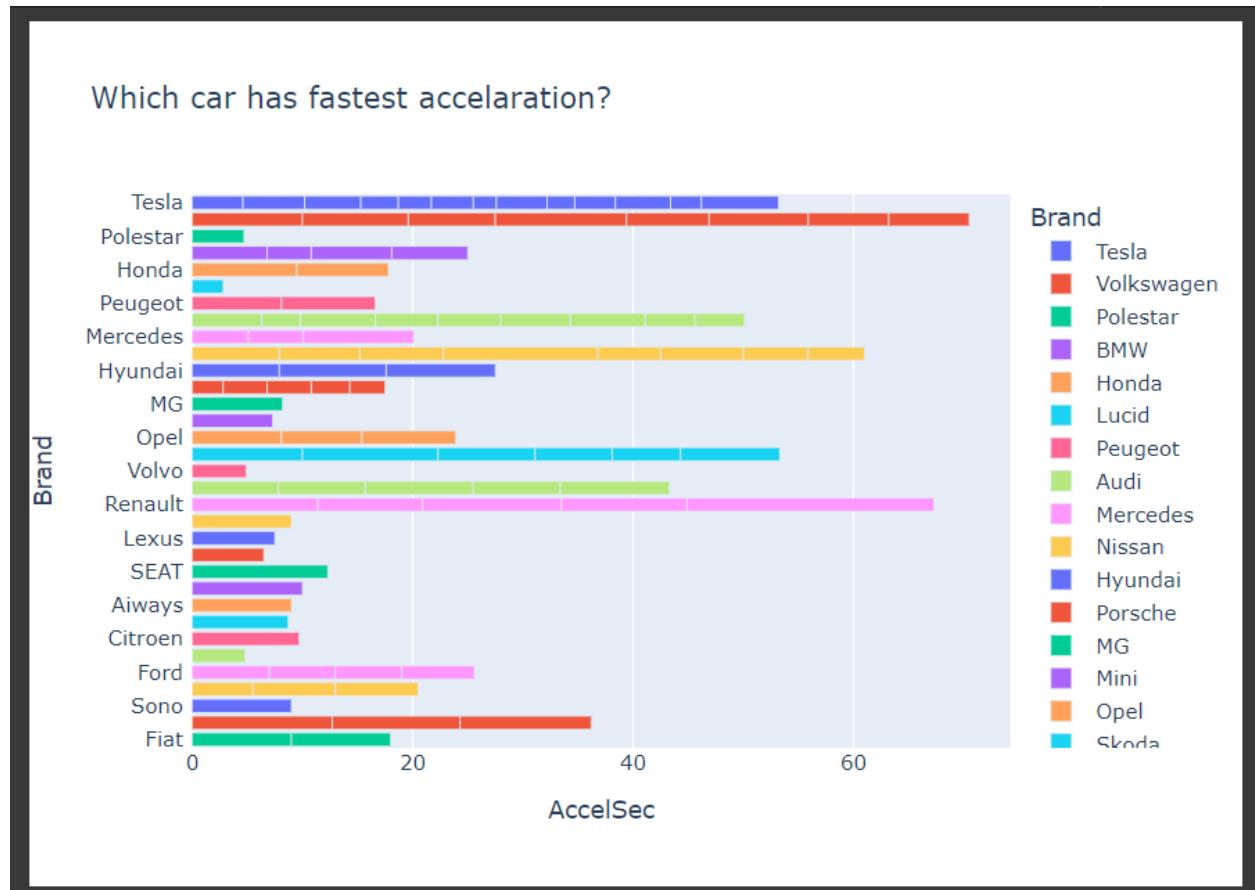
Despite growing at such a tremendous rate there are still certain shortcomings in this particular field some of which are listed below

1. Lack of recharge points at different locations.
2. Electric cars have been known for traveling a lesser distance as compared to fuel-derived vehicles.
3. These cars are currently pretty expensive for most people to afford.
4. Batteries need to be replaced every now and then which leads to higher maintenance costs.

There are certain areas we can still customize this particular market and we still have a certain scope of improvement.

We all are aware of the fact that Electric Vehicles have a higher acceleration as compared to other vehicles. So it is appropriate to target customers based on the acceleration of various types of EVs.

CONTEXT:



Above is a distribution of acceleration of various Electric Vehicle Brands all over the globe.

In case we consider demographic segments following are some ways to target appropriate customers:

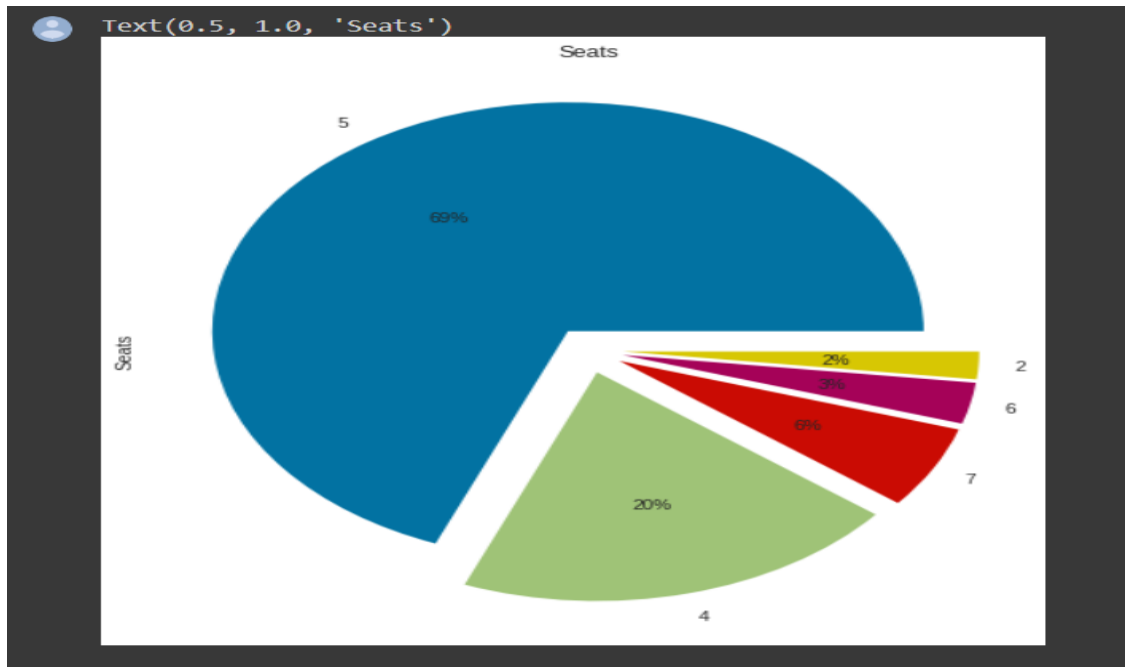
1. Studies have shown that most younger people tend to drive faster than elder age groups, who eventually have a family traveling with them.
2. Based on gender, studies have shown that most women tend to prefer a mildly accelerated car, which they consider to be safer as compared to men.
3. Yearly Income plays a huge role when it comes to buying an electric vehicle. If we take the example of Tesla (An automotive and energy company), it has an average price of 70 Lakh Indian rupees for the electric vehicles sold. So, only a well-settled person with a good income can afford a vehicle of this kind.

Combining all these points we can infer that a middle-aged, well-settled person can only afford an Electric Vehicle at this point in time.

To expand this market even further we must devise a strategy where we know which people are capable of making.

CONTEXT:

Similarly, if we consider a person with a family to care after, they would most probably also take the number of seats in the car into consideration. Below is a chart distribution of seats among various Electric vehicles as per our analysis:

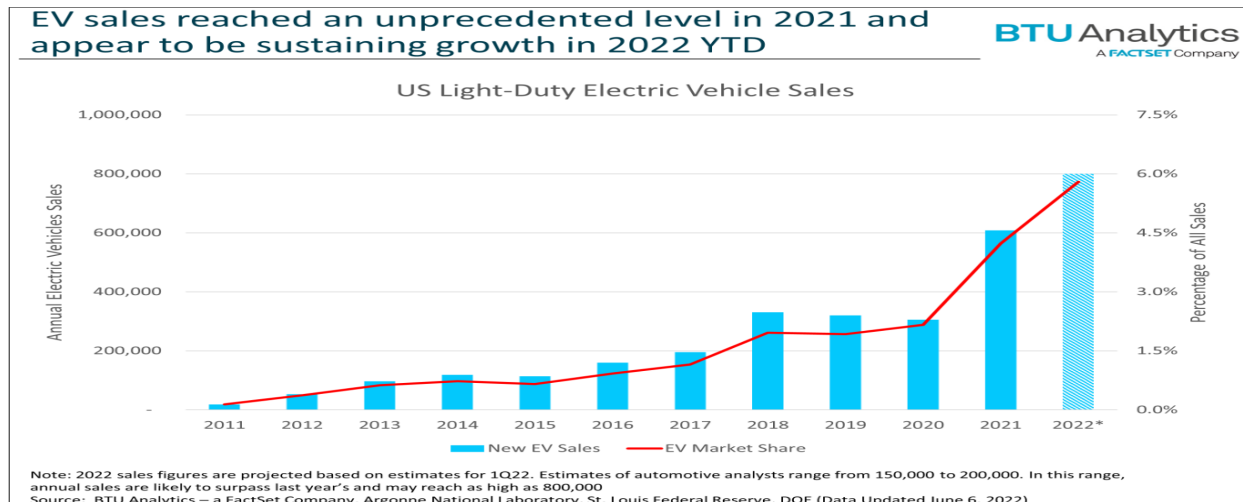


We can infer from this that 89% of Electric Vehicles in this dataset have more than 4 seats which means this can be family oriented.

Potential Customer Base In Early Market

EV market has been growing at a tremendous rate and is expected to be reached all over the world within a few years. The growth trends have not only been maintaining themselves but also have been increasing at a great speed as depicted in the figure below:

CONTEXT:



We are aware of the fact that electric vehicles are not very cheap at this point in time as given in our analysis of the price of a vehicle. We must target people who have an average income of at least 100,000 dollars annually. In India, we must target middle-aged millionaires or billionaires. There are around 700,000 dollar millionaires in India. Most males are interested in the EV market as per studies and also not all males are interested in buying an electric vehicle, which leaves us with an estimated 400,000 millionaires. The target price range can be taken as 4,500,000. It gives a total profit of 1800 billion.

The Most Optimal Market Segments

There are many possible market segments as discussed. There are many aspects in which we can define these market segments. Some of them are discussed below:

- Demographically segregated segments: As discussed earlier demographic segmentation is segregated based on various qualities like age, gender, occupation, etc., In this case, we have seen a pattern where males are more likely to buy electric vehicles as compared to females. Also, well-settled, middle-aged men mostly are able to afford this kind of vehicle.
- Behavioral segregated segments: In this case, we define our segments based on different actions taken by customers. Even if a customer has more than sufficient resources to buy an electric vehicle, still if he/she doesn't hold any kind of interest in electric vehicles there would be no point in targeting such a customer. Therefore, we must detect some behavior patterns as to if a person holds any kind of interest in this area or not.
- Geographically segregated segments: geographical segments are those which are segregated based on various locations. If we consider a particular area which

CONTEXT:

has wealthier aspects as compared to another country, the former would offer a wider market segment. For example, the USA is one of the richest countries so it would offer a wider market for electric vehicles.

- Psychographically segregated segments: This segment involves the emotions and traits of people. Some people who are inclined more towards environmental preservation are most likely to buy an electric vehicle as compared to a person who still has problems with the shortcomings an electric vehicle still offers.

Github Link To The Code Implementation

<https://github.com/Charanpoojary/Project-14-EV-Market-Segmentation>