Feynn Labs Final Assignment

Market Analysis of Electric Vehicles In India

Contributors :- Pratik Datey, Shruthy Radhakrishnan

Project Link :- https://github.com/pratikdatey/Electric-Vehicle-market-Analysis

Background

The electric vehicle industry in India is a growing industry. Electric Vehicles include a large range of vehicles from electric two - wheelers, three - wheelers (rickshaws), cars and electric buses. An electric vehicle can be classified on the basis of their attributes such as charging time, driving range, and the maximum load it can carry. Of these attributes, the two most important characteristics of an electric vehicle of concern to the consumer are: Driving range(i.e. Maximum distance an EV can run when fully charged) Charging time of batteries(i.e. the time required to fully charge the battery) and Charging time depends on the input power characteristics (i.e. input voltage and current), battery type and battery capacity. 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.

In this paper, our motto is to obtain information on the present Electrical Vehicle market in India using Business Analytics and Market Segmentation techniques. A detailed analysis of the various aspects of Indian EV's and its customer base is carried out through machine learning techniques such as Clustering. The case study focuses on the clustering pipeline using limited amounts of data obtained from several trusted platforms, including government open source.

Problem Statement

Task is to analyse the Electric Vehicles Market in India using Segmentation analysis and come up with a feasible strategy to enter the market.

In this report we analyse the Electric Vehicles Market in India using segments such as price, top speed, range, safety, battery capacity, fuel types, fast charging, boot space and much more.

Data Collection

Data was scraped from the website https://e-amrit.niti.gov.in/home. This website is the e-AMRIT (**Accelerated e-Mobility Revolution for India's Transportation**) portal for creating awareness about electric mobility in India. Also for some specification of Electrical Vehicle we gathered from https://www.cardekho.com/.

From the above links we gathered a datestes and merged them to create the final data. The data is partly used for visualization purpose and partly for clustering.

Final data generated:

https://github.com/pratikdatey/Electric-Vehicle-market-Analysis/blob/master/Final EV data.xlsx

```
data=pd.read_excel('Final EV data.xlsx')
data.head()
```

	Vehicle full name	Manufacturing	Model	Top speed (km/hr)	Price (INR)	Fuel Type	Wheelers type	Battery capacity [kWh]	Full charging time (HR)	Kerb weight (KG)	Range (km/hr)	Fast Charging	Drive Type	Number of Seats	boot space (L)	Number of Airbags	Type of brakes	Max Torque (N-M)	Type of Vehicle
0	Revolt RV400	Revolt Motors	RV400	85.0	134000.0	Electric	Two wheeler	4.0	4.5	108.0	150.0	YES	Belt Drive	2	NaN	NaN	Disc	170.0	Motor cycles
1	Revolt RV300	Revolt Motors	RV300	65.0	94999.0	Electric	Two wheeler	2.7	4.2	101.0	180.0	YES	Hub Drive	2	NaN	NaN	Disc	NaN	Motor cycles
2	Tork Motors(Kratos)	Tork Motors	Kratos	100.0	192499.0	Electric	Two wheeler	4.0	5.0	NaN	180.0	NO	NaN	2	NaN	NaN	Disc	28.0	Motor cycles
3	Tork Motors(Kratos R)	Tork Motors	Kratos R	105.0	207499.0	Electric	Two wheeler	4.0	5.0	NaN	180.0	YES	NaN	2	NaN	NaN	Disc	38.0	Motor cycles
4	Oben Rorr	Kabira Mobility	Oben Rorr	100.0	102999.0	Electric	Two wheeler	4.4	2.0	110.0	200.0	YES	Belt Drive	2	NaN	NaN	Disc	NaN	Motor cycles

Fig no.1. Final dataset

Each column explained below:

- Vehicle full name - Name of vehicle

- Manufacturing - Manufacturing company of vehicle

- Model - Model of vehicle

- Top speed (km/hr) - Maximum speed of vehicle in (km/hr)

- Price (INR) - Price of vehicle

- Fuel Type - Type of fuel (Electrical, Hybrid)

- Wheelers type - Type of wheelers(Two,Three,Four wheelers)

- Battery capacity [kWh] - Capacity of battery in (kwh)

- Full charging time (HR) - Total charging time 100% in (hr)

- Kerb weight (KG) - Total weight of vehicle in (kg)

- Range (km/hr) - Maximum kilometer covered per charging in (km/hr)

- Fast Charging - Vehicle have fast charging or not

- Drive Type - Type of Drive

- Number of Seats - Number of Seats in vehicle

- boot space (L) - Space for luggages in (Liter)

- Number of Airbags - Airbags for safety

- Type of brakes - Type of brakes

- Max Torque (N-M) - Max torque (n-m)

- Type of Vehicle - Vehicle types (Scooter, Cars,etc.)

- Income - Price range of vehicle (Thousands, Lakhs, Crore)

Data Preprocessing

Steps taken to preprocess the raw data scraped:

1. Dealing with different variables names but having the same information in columns, so we replace it.

```
In [5]: data['Wheelers type']=data['Wheelers type'].replace('four wheeler','Four Wheeler')
        data['Wheelers type']=data['Wheelers type'].replace('Four Wheeler','Four wheeler')
        data['Fast Charging']=data['Fast Charging'].replace('NO','No')
        data['Fast Charging']=data['Fast Charging'].replace('YES','Yes')
        data['Fuel Type']=data['Fuel Type'].replace('electric','Electric')
```

Fig no.2. Deals with dummy variable in columns

2. Deals Null values in the dataset by using filling mean values.

```
data['Top speed (km/hr)']=data['Top speed (km/hr)'].fillna(data['Top speed (km/hr)'].mean())
data['Price (INR)']=data['Price (INR)'].fillna(data['Price (INR)'].mean())
data['Battery capacity [kWh]']=data['Battery capacity [kWh]'].fillna(data['Battery capacity [kWh]'].mean())
data['Kerb weight (KG)']=data['Kerb weight (KG)'].fillna(data['Kerb weight (KG)'].mean())
data['Max Torque (N-M)']=data['Max Torque (N-M)'].fillna(data['Max Torque (N-M)'].mean())
```

Fig no.3. Deals with Null values

3. Create Income feature for range between Low(Thousands), Medium(Lakhs), High(Crore).

```
In [6]: def income(price):
            if price <= 100000:
                return 'Low (Thousands)'
            elif price>100000 and price<10000000:
                return 'medium (Lakhs)'
            else:
                return 'High(Crore)'
```

```
In [8]: data['Income'] = data['Price (INR)'].apply(income)
```

Fig no.3. Income Feature

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a process of describing the data by means of statistical and visualization techniques in order to bring important aspects of that data into focus for further analysis.

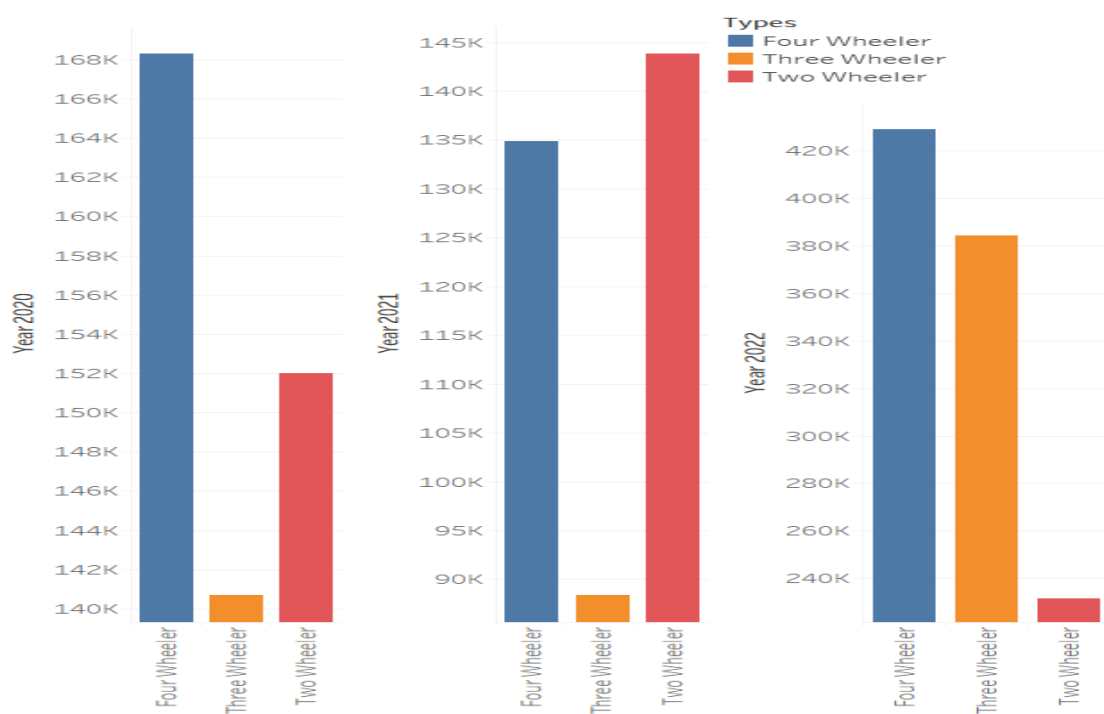For analyse purpose we took some features for visualization from our dataset as shown below:

Fig no.4. EV sales by years

This bar graph shows sales of Two ,Three, Four wheeler of Electrical Vehicles in years of 2020, 2021 & 2022. We can see the maximum average sales in all three years is of Four wheelers and it is still increasing year by year.

```
plt.figure(figsize=(30,20))
sns.barplot(charging_station['State wise'],x=charging_station['State wise'],
    y=charging_station['Number of Electric Vehicle Charging Sanctioned'],color='skyblue')
plt.title('Charging Sation by State wise ',fontsize = 40)
```
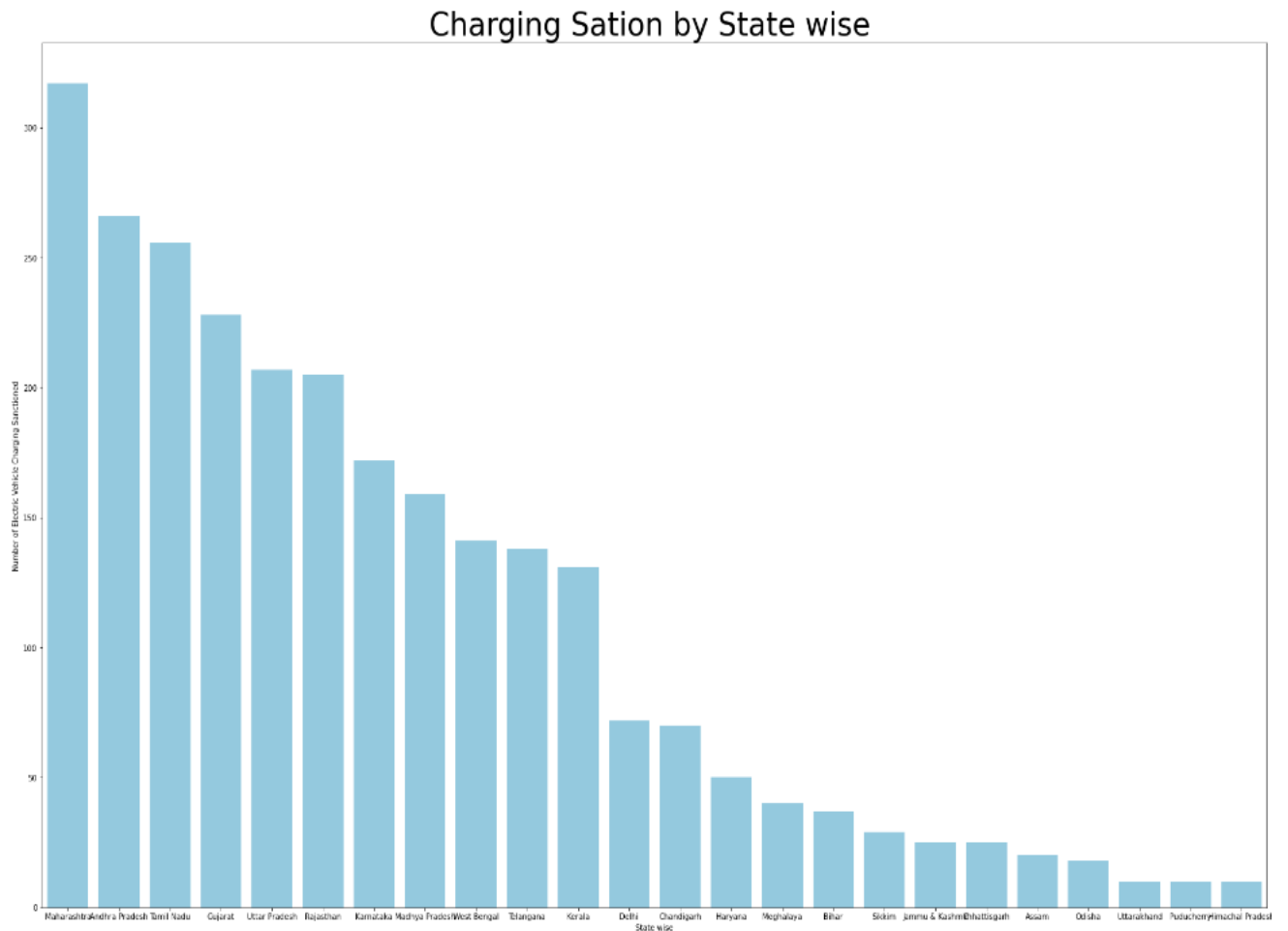
Charging Sation by State wise

Fig no.5. Charging station by state wise in India

As from Fig.no.5, We can see numbers of charging stations present in India as per states. The maximum number of charging stations present in Maharashtra and lowest in Himachal Pradesh.

```
#EXPLORING DATA
plt.figure(figsize=(25,8))
explode = [0.01,0.04,0.04,0.04,0.04,0.04]
labels=['Cars','Scooters','Motor cycles','Bus','Rickshaws','Cargo three wheeler']
plt.pie(data['Type of Vehicle'].value_counts(),
        labels=labels,autopct = '%.2f%%',explode=explode)
plt.title('Type of Vehicle', fontsize = 30)
```
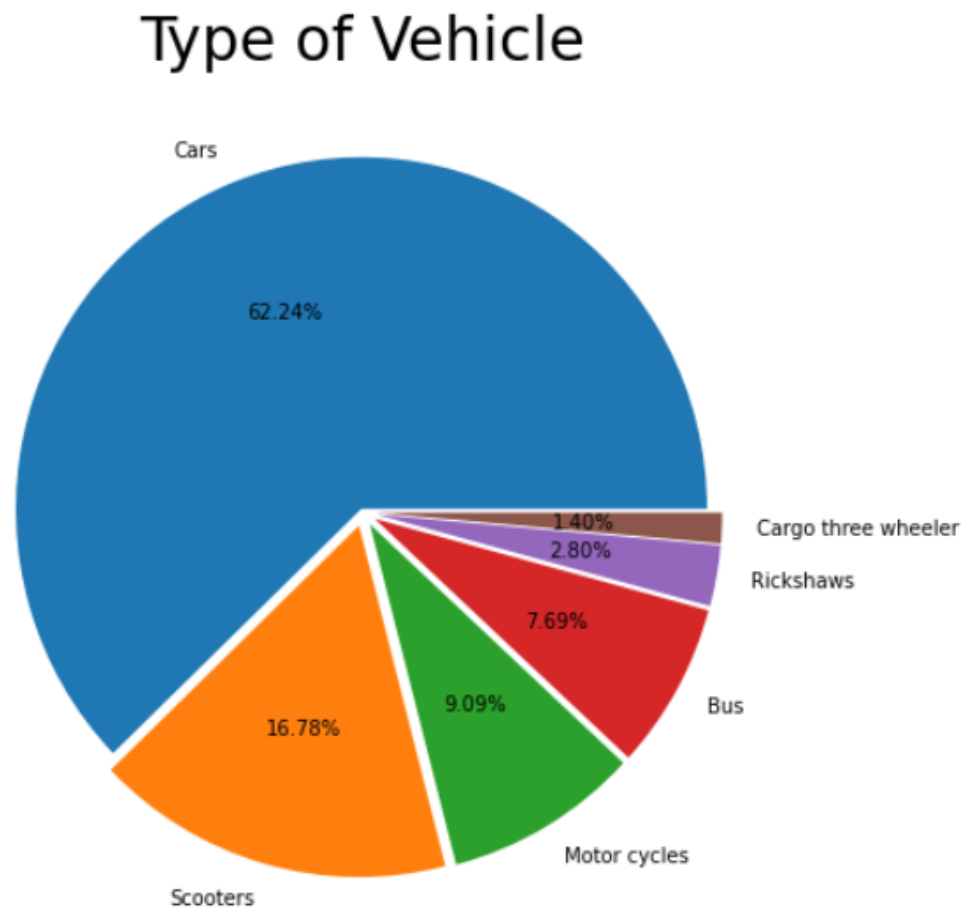

Fig no.6. Types of Electrical Vehicles

Fig.no.6 shows percentages of Electrical vehicles types in india. Basically it shows a manufacturing market percentage of every type of vehicle. In that we can see that the market of Cars is high. A lot of EV startup companies are manufacturing or focused on only Electricals Cars. Also there is less market for Cargo and Rickshaws. A very less number of companies are focusing on Cargo and Rickshaws.

```
#Scatter plot between Price and Top speed
plt.figure(figsize=(12,6))
sns.scatterplot(x='Price (INR)',y='Top speed (km/hr)',data=new_data)
```
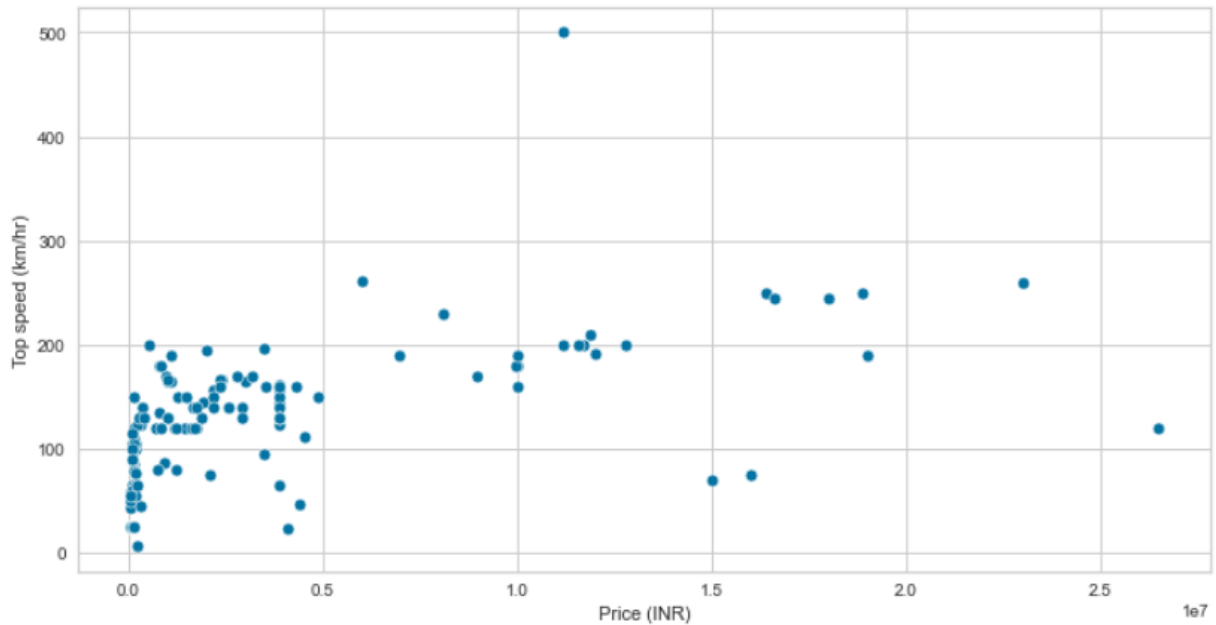

Fig no.7. Top speed VS Price

Above Fig.no.7 shows a scatter plot between Top speed vs Price to see the relation between them. As from this scatter plot ,we can conclude that if the Top Speed of EV is increasing then the Price of EV is also increasing. They both are directly proportional to each other.

```
#Income Feature
plt.figure(figsize=(12,8))
sns.countplot(new_data['Income'])
plt.title('Income')
plt.ylabel('Counts')
```
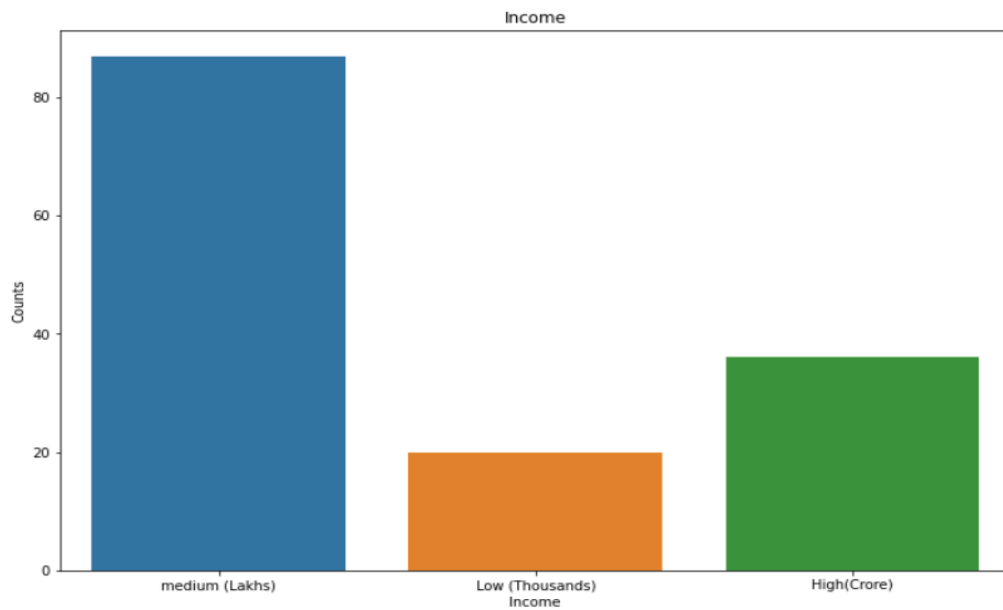

Fig no.8. Income

Fig no.8. Shows a plot of information about Income feature. we categorized Income features in three different types as, first in Low means the price of EV is in thousands rupees (Less than 1 lakhs), second in Medium means the price of EV is in lakhs (Between 1 lakh to 1 crore) and Third in High means the price of EV is in crore (Greater than 1 crore). As from countplot we can conclude that the maximum EV's price is in lakhs (Medium).

```
#Scatter plot between Income and Top speed
plt.figure(figsize=(12,6))
sns.scatterplot(x='Income',y='Top speed (km/hr)',data=new_data)
plt.title('Income VS Top Speed')
```
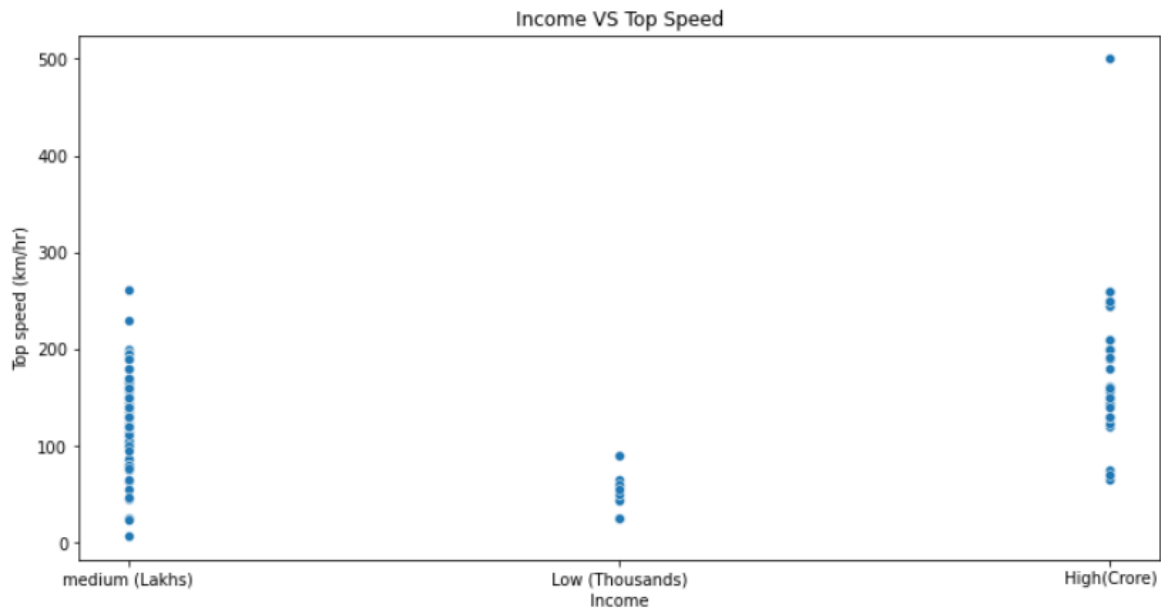

Fig no.9. Income vs Top speed

This figure shows a relationship between Income and Top speed. We can see that if the price of EV in Low (thousands) then your top speed lies within 0-110 km/hr. As the price increases your vehicle's top speed also increases.
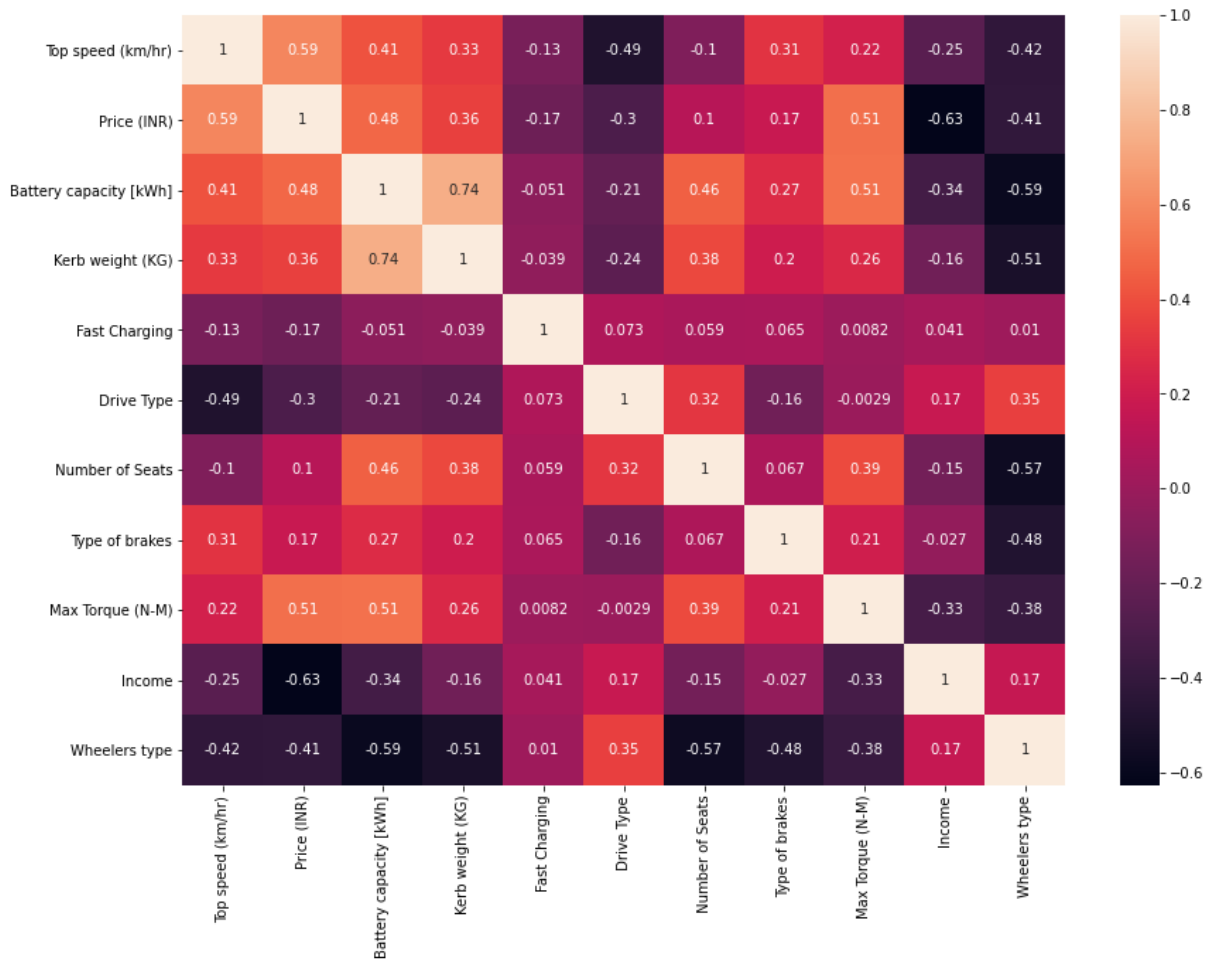
Fig no.10. Heat map

Above fig.no.10 shows the correlation between every individual variable. We can see that Kerb weight and Battery capacity have the highest correlation. Meaning of this is if we want more battery capacity our EV weight will increase.

```
#Histogram
plt.rcParams['figure.figsize']=(20,17)
data.hist(['Price (INR)','Top speed (km/hr)','Battery capacity [kWh]','Kerb weight (KG)','Range (km/hr)',
          'boot space (L)','Max Torque (N-M)',' Number of Seats'])
```
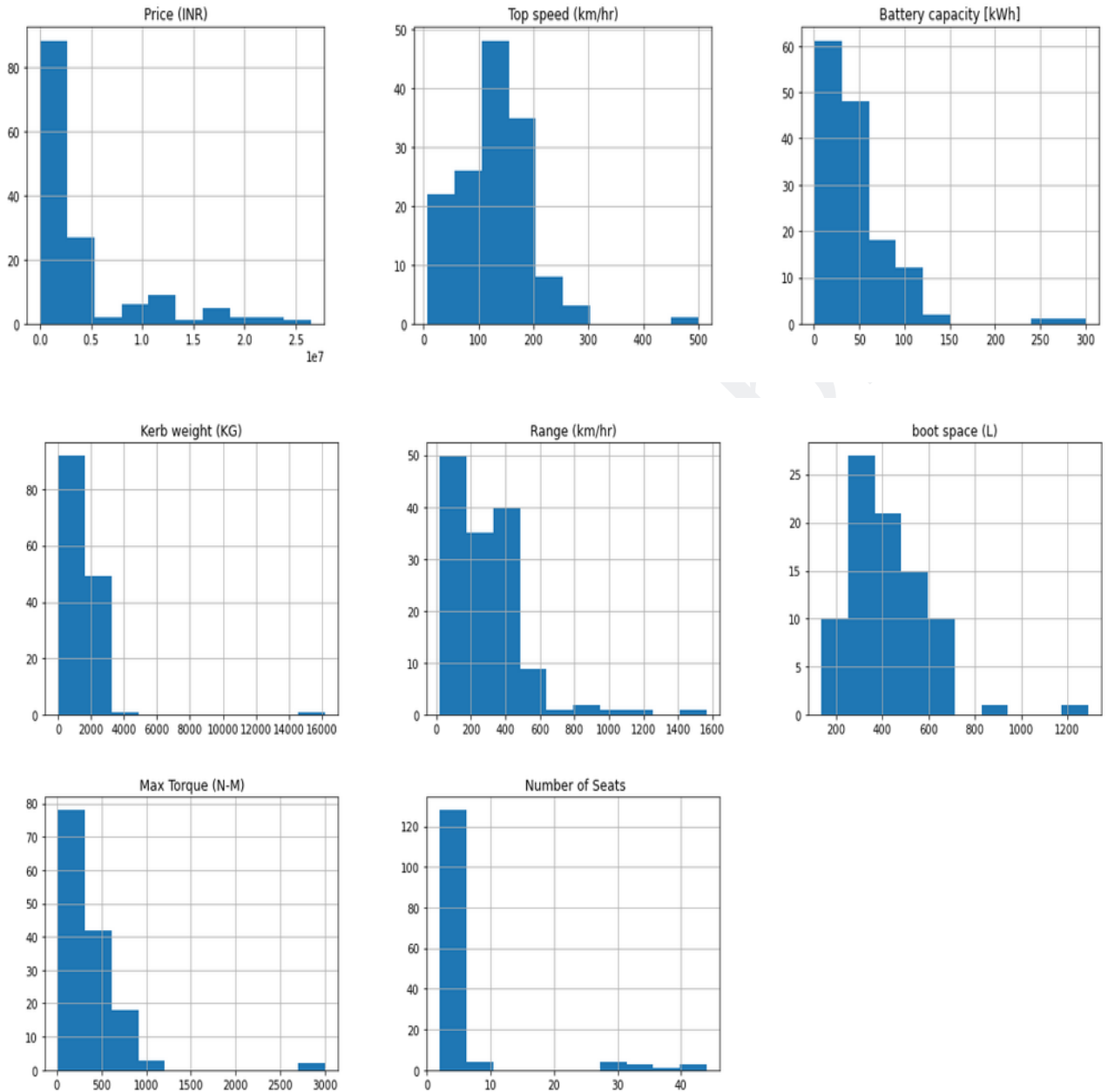

Fig no.11. Histogram

In the above figure we plot histograms of every single feature. As from that we can see that mostly Price ranges between thousands to lakhs. In Top speed maximum average value is around 150km/hr, same as for Battery capacity ranges around 0-50Kwh. As a Kerb weight it averages at 0-2000kg. Most

EVhas Range between 0-100km/hr. For boot space we can conclude that most EVs have 300 liter boot space. Also for maximum Torque and number of Seats, we can see torque lies between 0-400 and average EVs have 5 seats.

Market Segments

K-Means Clustering

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

We start by pre-processing the data and cleaning it. This essentially involves null-handling ,label encoding and dummies variables in the ordinal parameters of the data. The data is then passed into the Scikit-Learn K-Means Clustering model to obtain the elbow curve for the ideal number of clusters. Using the "elbow" or "knee of a curve" as a cutoff point is a common heuristic in mathematical optimization to choose a point where diminishing returns are no longer worth the additional cost.
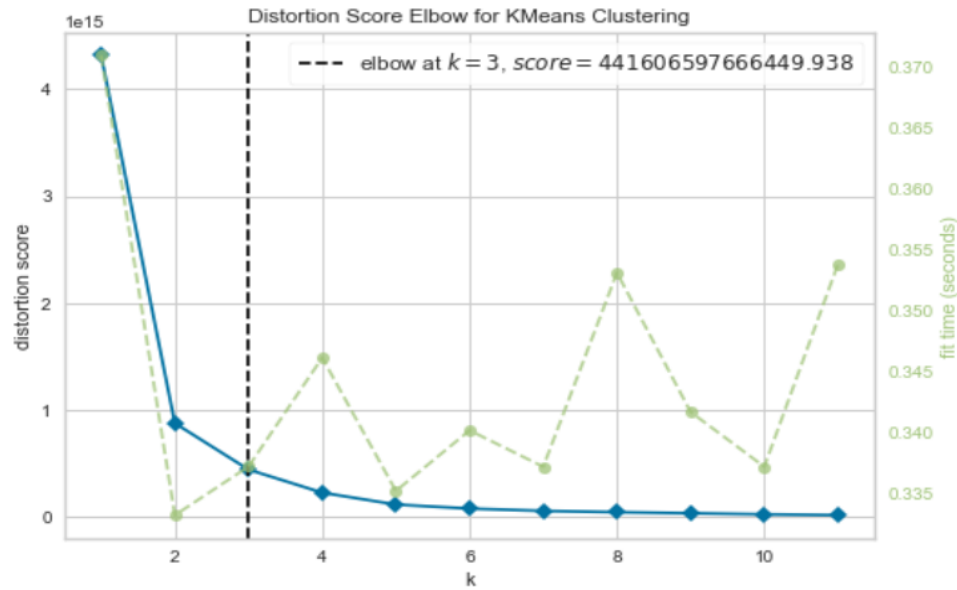
Fig no.12. Elbow Method

Based on the elbow curve, we assume the number of clusters to be optimally around 3. In clustering, this means one should choose a few clusters so that adding another cluster doesn't give much better modeling of the data. The intuition is that increasing the number of clusters will naturally improve the fit (explain more of the variation), since there are more parameters (more clusters) to use, but that at some point this is over-fitting, and the elbow reflects this.
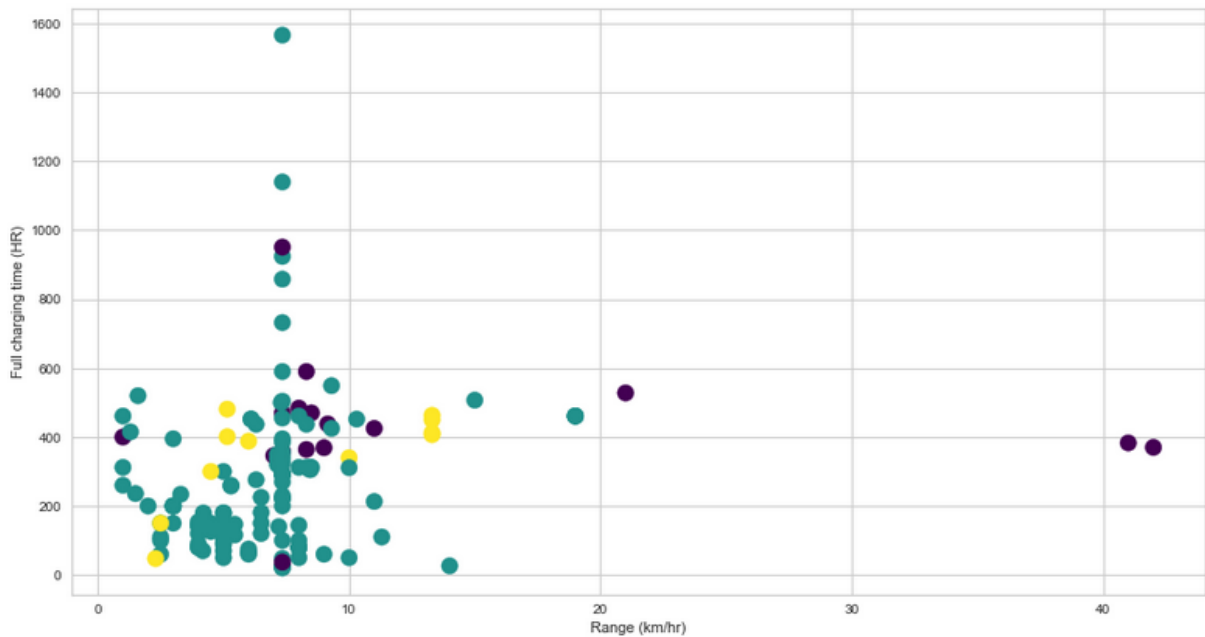

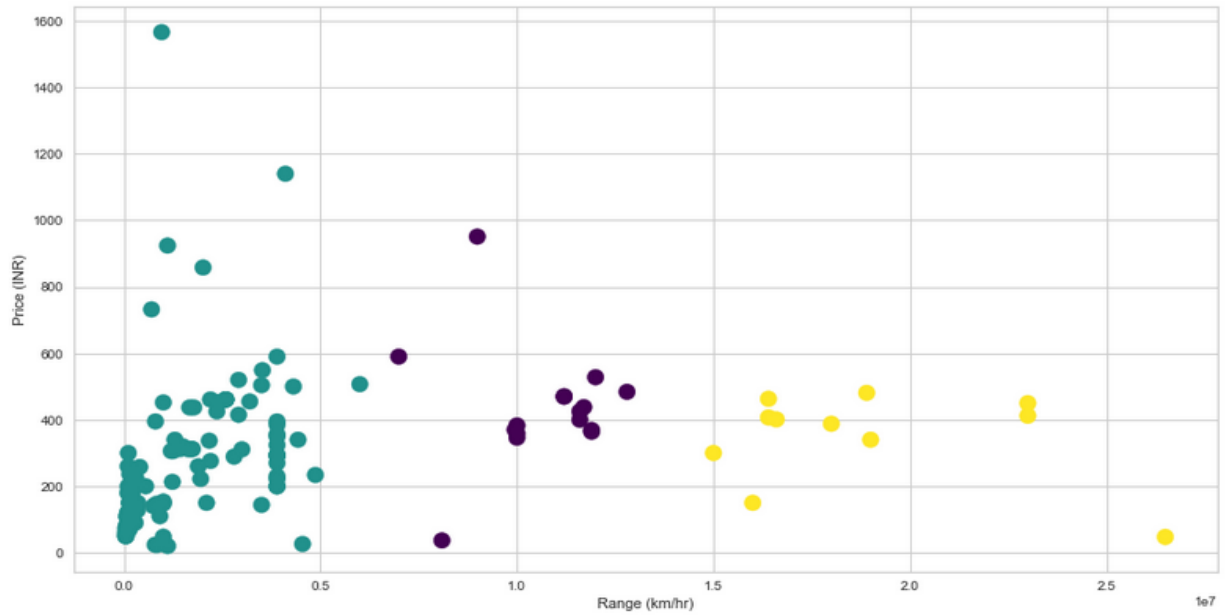Fig no.13. K-Means Clustering on Range vs charging time

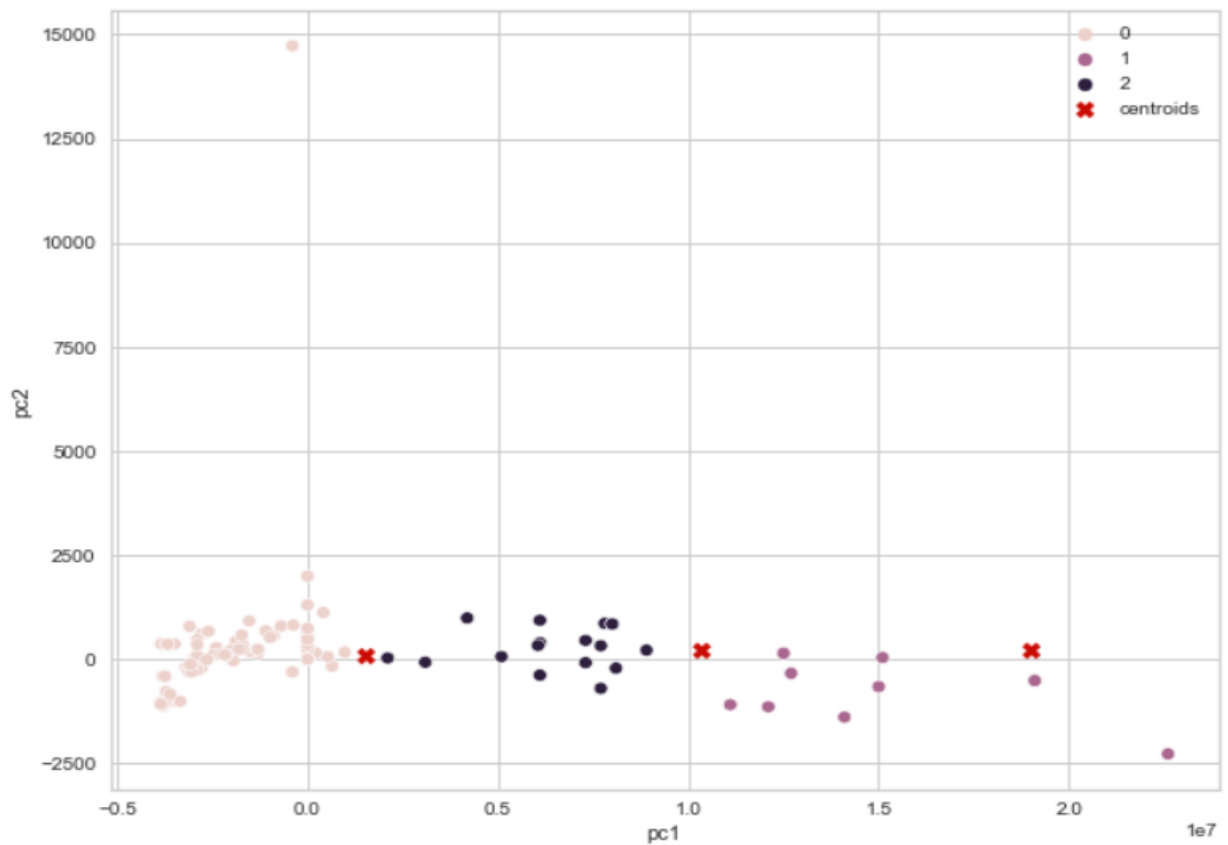Fig no.14. K-Means Clustering on Range vs Price

Fig no.15. K-Means Clustering on pc1 vs pc2 with Centroids

In the above Fig.no.15 we create 3 clusters by using K-Means Clustering and visualize for better understanding with Centroids.

Analysing Market Segments

There are several different variables by which segmentation is done:

Geographic segmentation

Geographic segmentation consists of creating different groups of customers based on geographic boundaries. The needs and interests of potential customers vary according to their geographic location, climate and region, and understanding this allows you to determine where to sell and advertise a brand, as well as where to expand a business.

- Charging station by State wise: State wise charging station will become a significant effect on consumer purchasing decisions. Those states with more charging stations may prefer to buy an EV and vice versa.

Demographic segmentation

Demographic segmentation consists of dividing the market through different variables such as age, gender, nationality, education level, family size, occupation, income, etc. This is one of the most widely used forms of market segmentation, since it is based on knowing how customers use your products and services and how much they are willing to pay for them.

- Income: Income levels have a significant effect on consumer purchasing decisions. Those with higher-income levels may prefer luxury vehicles. Conversely, individuals with lower income levels may prefer to get vehicles at the best deal and are likely to choose inexpensive products/services.

- Family size: Family size also determines consumers' purchase decisions. Those who have large family members may choose four wheelers and those who have less family members will choose two wheelers.

Psychographic segmentation

Psychographic segmentation consists of grouping the target audience based on their behavior, lifestyle, attitudes and interests. To understand the target audience, market research methods such as focus groups, surveys, interviews and case studies can be successful in compiling this type of conclusion.

- Lifestyle: A consumer whose profession is more time consuming than other average consumers , that consumer may select a vehicle who takes less time to charge a vehicle. This group of consumers only focus on the time required to charge an EV.

- Interests : Some consumers may have interest in particular manufacturing companies. Some consumers may like only vehicles made by the Tata company.

- Behavior : Behavior of consumers is the most important factor in the market segment. It shows what exactly consumers want from us?. Some consumers may want an EV who will cover far distance per a charging.

Customizing the Market Mix

The marketing mix refers to the set of actions, or tactics, that a company uses to promote its brand or product in the market. The 4Ps make up a typical marketing mix - Price, Product, Promotion and Place.

- **Price:** Refers to the value that is put for a product. It depends on costs of production, segment targeted, ability of the market to pay, supply - demand and a host of other direct and indirect factors. There can be several types of pricing strategies, each tied in with an overall business plan.

- **Product**: Refers to the item actually being sold. The product must deliver a minimum level of performance; otherwise even the best work on the other elements of the marketing mix won't do any good.

- **Place:** Refers to the point of sale. In every industry, catching the eye of the consumer and making it easy for her to buy it is the main aim of a good distribution or 'place' strategy. Retailers pay a premium for the right location. In fact, the mantra of a successful retail business is 'location, location, location'.

- **Promotion:** This refers to all the activities undertaken to make the product or service known to the user and trade. This can include advertising, word of mouth, press reports, incentives, commissions and awards to the trade. It can also include consumer schemes, direct marketing, contests and prizes.

All the elements of the marketing mix influence each other. They make up the business plan for a company and handle it right, and can give it great success. The marketing mix needs a lot of understanding, market research and consultation with several people, from users to trade to manufacturing and several others.

Target Segment

Target marketing involves breaking a market into segments and then concentrating your marketing efforts on one or a few key segments consisting of the customers whose needs and desires most closely match your product or service offerings. It can be the key to attracting new business, increasing sales, and making your business a success.

Below are the segments which we created from most important key factors from our dataset:

```
#Segment
segment = Income.merge(Type_of_Vehicle, on='cluster_num', how='left').merge(Fuel_Type, on='cluster_num',
how='left').merge(ranges,on='cluster_num', how='left').merge(Full_charging_time, on='cluster_num', how='left')
segment
```

	cluster_num	Income	Type of Vehicle	Fuel Type	Range (km/hr)	Full charging time (HR)
0	0	1.582609	2.634783	0.095652	265.971311	6.459936
1	1	0.000000	1.636364	0.090909	349.040909	8.072727
2	2	0.705882	2.000000	0.117647	440.647059	12.860567

Fig no.16. Segments

```
#Target segments

plt.figure(figsize = (12,4))
sns.scatterplot(x = "Range (km/hr)", y = "Income",data=segment,s=400, color="r")
plt.title("Simple segment evaluation plot for the Electrical data set",
          fontsize = 17)
plt.xlabel("Income", fontsize = 14)
plt.ylabel("Range (km/hr)", fontsize = 14)
plt.show()
```
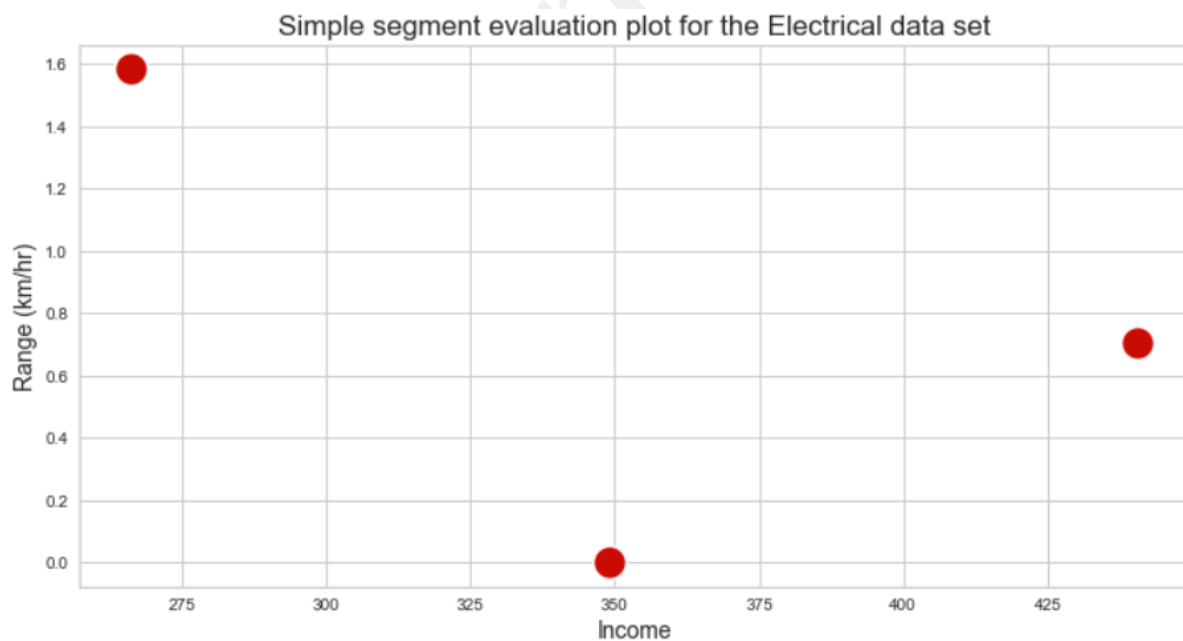

Fig no.17. Target Segment on Income vs Range

```
#Target segments
plt.figure(figsize = (12,8))
sns.scatterplot(x ='Full charging time (HR)' , y ="Range (km/hr)",data=segment,s=400, color="r")
plt.title("Simple segment evaluation plot for the Electrical data set",
          fontsize = 17)
plt.xlabel("Full charging time (HR)", fontsize = 14)
plt.ylabel("Range (km/hr)", fontsize = 14)
plt.show()
```
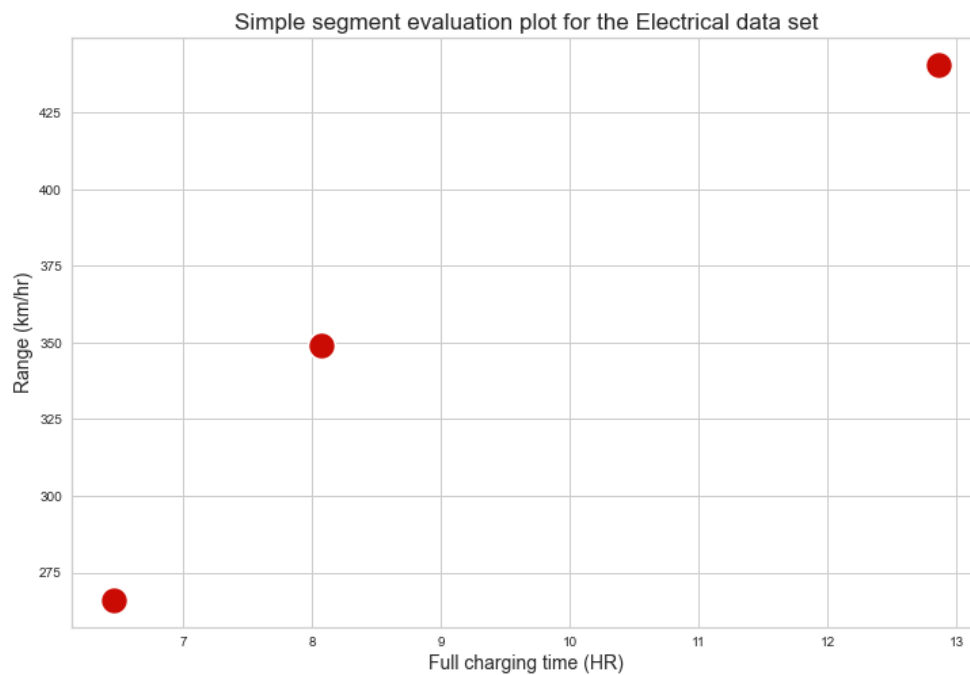

Fig no.18. Target Segment on Full charging time vs Range

It can be concluded from above figures that Range, Top Speed, Full charging time, Income and Types of Vehicles can be the most important segment categories for consumer purchasing decisions. These are the key factors who make markets different and similar at the same time. This segments have formed with distinct features which may indicate that their preferences for EVs are motivated by different factors.

Insight and Recommendation

The penetration of EV in India has Increased Significantly in the last five years as they are more efficient. In addition, growing fuel prices are further helping to boost substantial growth in the product adoption, mainly due to their extended range and efficiency. The global Electric Vehicle Market size is projected to grow from 8,151 thousand units in 2022 to 39,208 thousand units by 2030, at a CAGR of 21.7%. Factors such as growing demand for low emission commuting and governments supporting long range, zero emission vehicles through subsidies & tax rebates have compelled the manufacturers to provide electric vehicles around the world. Increasing investments by governments across the globe to develop EV charging stations and Hydrogen fueling stations along with incentives offered to buyers will create opportunities for OEMs to expand their revenue stream and geographical presence.

From this analysis we create different types of segments to affect consumers' purchasing decisions. **Geographic** segmentation is about places, cities, states that where consumers live will affect market sales. Like if a consumer lives in a rural area there may be less possibility of having charging stations and vice versa in urban areas. Now in 2022 yet we have only 1742 public charging stations available. So if a consumer is from those states who have more available charging stations ,the probability of buying is more as compared to others who have less charging stations in their states. **Demographic** segmentation focuses on education level, family size, occupation, income, etc. since it is based on knowing how customers use your products and services and how much they are willing to pay for them. That depends on consumers' education, Financial status and purpose of buying EV's. If a customer's purpose is to buy an EV for transporting goods in different cities or states, that customer will focus on the boot space and maximum range of a vehicle. On a **psychological** segment some customers may go for a product which gives them satisfaction and others may go with a product who is cheaper in cost and their other factors are average.

Acknowledgement

We would like to express our special thanks of gratitude to Feynn Labs who gave us the golden opportunity to do this wonderful project on the Electric Vehicle Market Segmentation Analysis, which also helped us in doing a lot of research and we also came to know about so many new things.

We perceive this opportunity as a big Milestone in my career development. We will strive to use gained skills and knowledge in the best possible way, and will continue to work on improvement, in order to attain desired career objectives.

Thank You.