MARKET SEGMENTATION FOR ELECTRIC VEHICLE STARTUP IN INDIA

ASWANTH K

11 APRIL 2024

ABSTRACT

This project uses machine learning to divide up the electric vehicle (EV) market for a startup. We're using techniques like clustering and predictive modeling to analyze different types of data, such as information about people's age, where they live, and how they behave. This helps us find groups of customers who are similar to each other. The cool thing is, we're making sure our results are easy to understand. This means we can use them to make smart decisions about how to advertise our EVs and what features they should have. By doing this, we're helping the startup target and keep customers in a tough market where lots of companies are competing.

1. PROBLEM STATEMENT

This project is about helping an electric vehicle (EV) startup in India figure out which group of people they should focus on. India has lots of different types of people, so it's important for the startup to pick the right group to target. They'll do this by studying different types of customers based on where they live, their age, income, and other factors. By doing this, they can understand what these different groups of people want and what other companies are offering them. Then, they'll find the best group of people who are most likely to be interested in buying the startup's EVs. With this information, they'll make a plan to enter the market in a way that appeals to this chosen group, considering any limitations in the data they have and how the market is changing. The goal is to set the startup up for long-term success in the growing electric vehicle market in India.

2. DATA COLLECTION

To kickstart our market segmentation analysis for launching our EV startup in India, I began by gathering data from a variety of sources available online. This involved carefully searching through different websites and databases to find relevant information for our project. By doing this, we're laying the foundation for the next important step: figuring out which customer segment holds the most potential for our startup to thrive in the Indian electric vehicle market. Essentially, we're collecting all the necessary facts and figures that will help us understand the market better and make informed decisions about where to focus our efforts. This thorough data collection process is crucial for guiding our strategy and ensuring that we target the right group of customers for a successful entry into the dynamic and growing Indian EV market.

Websites used for researching:

- https://www.kaggle.com/
- https://datasetsearch.research.google.com/
- https://data.gov.in/
- https://trends.google.com/trends/explore

Datasets I used on the project

1. https://www.kaggle.com/datasets/prasenjitsharma/fuel-type-wise-vehicle-registration-india/data

The dataset provides information on the total number of vehicles registered in India from January 2014 to July 2023, categorized by the type of fuel used by the vehicles. By analyzing this data, we can understand the buying patterns in the Indian market and how they've evolved over the years. Essentially, we'll be looking at whether people are buying more petrol, diesel, electric, or other types of vehicles during different periods. This information can give us insights into the preferences of Indian consumers when it comes to choosing vehicles and how those preferences might have changed over time.

2. https://pib.gov.in/Pressreleaseshare.aspx?PRID=1808115

This dataset specifies the sanctioned EV Charging Stations in India:

a. State-wise sanctioned EV Charging Stations

b. City-wise sanctioned EV Charging Stations

c. Sanctioned EV Charging Stations on Expressways and Highways

By analyzing this data, we can figure out which states or cities in India already have or are planning infrastructure to support electric vehicles (EVs). This information is crucial for making decisions about where to target our EV startup's efforts. If certain areas already have charging stations or other facilities for EVs, it might be easier for us to introduce our electric vehicles there. Essentially, we're looking for places where the groundwork has been laid for EV adoption, making it more likely for people to buy and use our EVs. This way, we can focus our resources and efforts on areas that are ready for electric vehicles, maximizing our chances of success.

3. CODE IMPLEMENTATION:

Importing all the neccesarry libraries

In [5]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
import warnings
warnings.filterwarnings('ignore')
```

- **Fuel Type wise vehicle registration in India**

I. Fuel Type wise vehicle registration in India

The dataset specifies the total no. of vehicles registered in India from January 2014 to July 2023.

The data has been categorized into fuel variant of the vehicle registered into the following.

- Only CNG
- Diesel
- Diesel/Hybrid
- Dual Diesel/CNG
- Electric(BOV)
- Ethanol
- LPG Only
- Petrol
- Petrol/CNG
- Petrol/Ethanol
- Petrol/Hybrid
- Petrol/LPG
- Solar
- Fuel cell Hydrogen
- LNG
- Methanol
- Dual Diesel/LNG

Data has been scrapped from Government of India Vehicle Registration website MPARIAVAHAN.

➢ Data Loading and Preprocessing

[13]:
```
print(pd.isnull(datasetFuelType).sum())
```

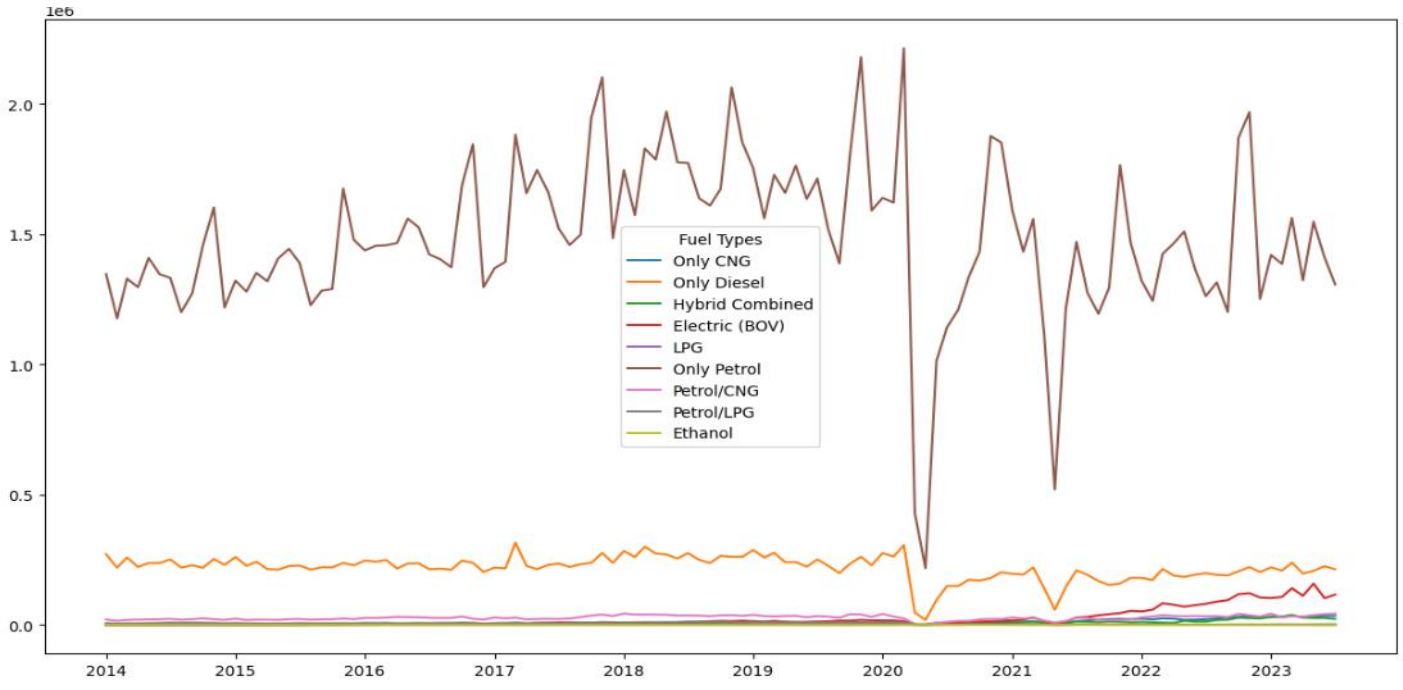
```
Month                0
CNG ONLY             0
DIESEL               0
DIESEL/HYBRID        0
DUAL DIESEL/CNG      0
ELECTRIC(BOV)        0
ETHANOL              0
LPG ONLY             0
NOT APPLICABLE       0
PETROL               0
PETROL/CNG           0
PETROL/ETHANOL       0
PETROL/HYBRID        0
PETROL/LPG           0
SOLAR                0
FUEL CELL HYDROGEN   0
LNG                  0
METHANOL             0
DUAL DIESEL/LNG      0
dtype: int64
```

[14]:
```
datasetFuelType["Month"]=pd.to_datetime(datasetFuelType["Month"], format='%b-%y')
datasetFuelType["Month"].head()
```

[14]:
```
0   2014-01-01
1   2014-02-01
2   2014-03-01
3   2014-04-01
4   2014-05-01
Name: Month, dtype: datetime64[ns]
```

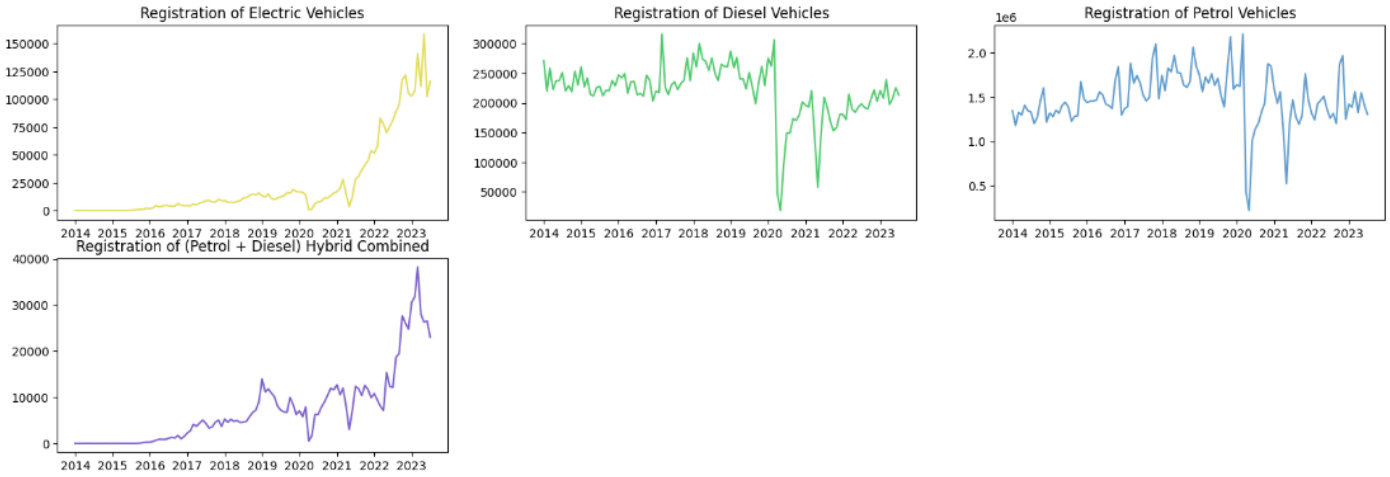
[15]:
```
datasetFuelType['HYBRID COMBINED'] = datasetFuelType['PETROL/HYBRID'] + datasetFuelType['DIESEL/HYBRID']

# Dropping 'PETROL/HYBRID' and 'DIESEL/HYBRID' columns
datasetFuelType.drop(['PETROL/HYBRID', 'DIESEL/HYBRID'], axis=1, inplace=True)
```

➢ Exploring the Data

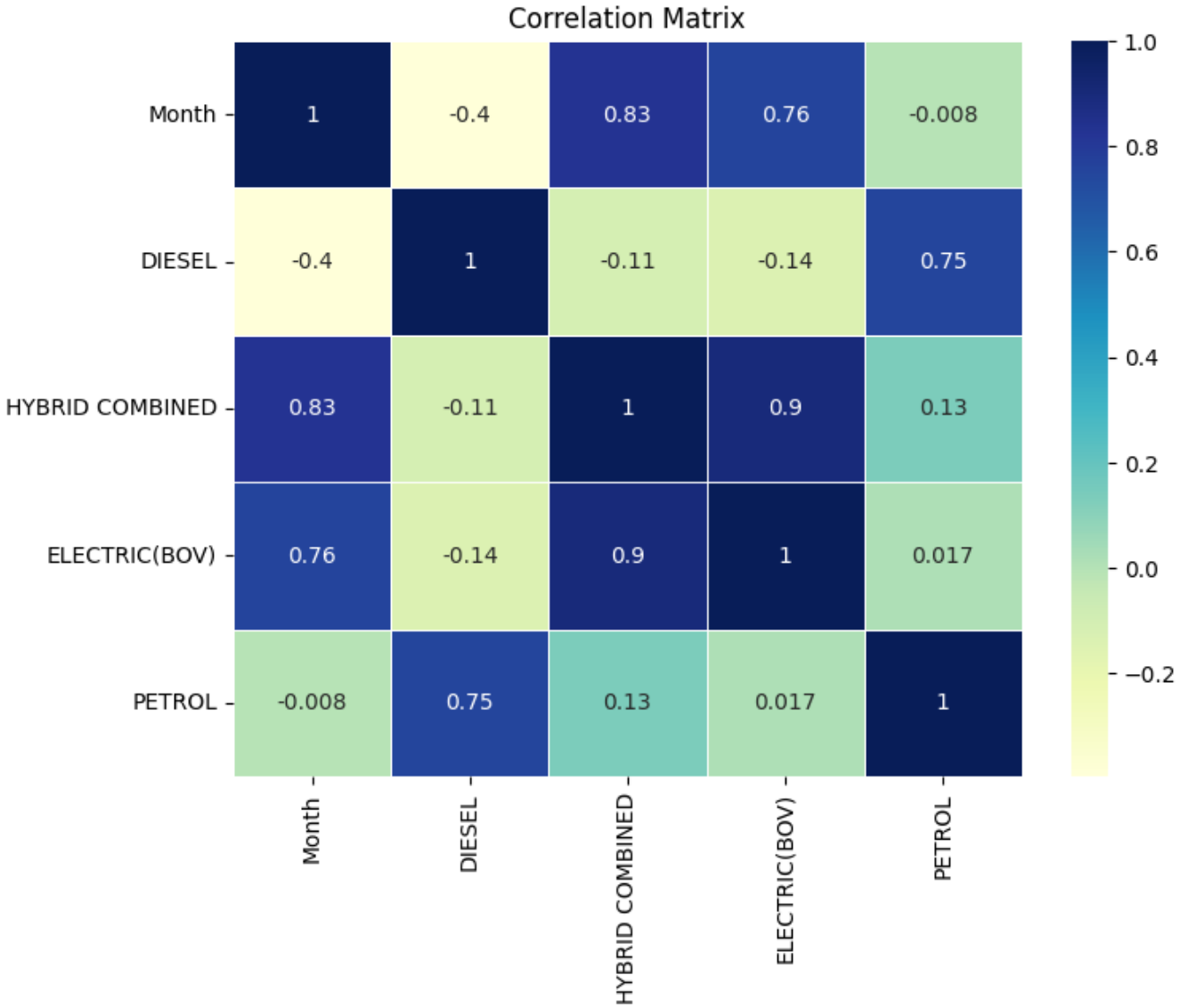


we can see that:

- Petrol vehicles consistently have the highest registrations throughout the time period, indicating their dominance in the market.
- Diesel vehicle registrations remain relatively stable over time, suggesting a consistent demand for diesel-powered vehicles.
- Electric vehicles show an increasing trend in registrations over time, indicating a growing interest or adoption of electric mobility, from late 2021 onwards.
- The increasing trend in EV registrations suggests a potential shift in the market toward cleaner and more sustainable transportation options.

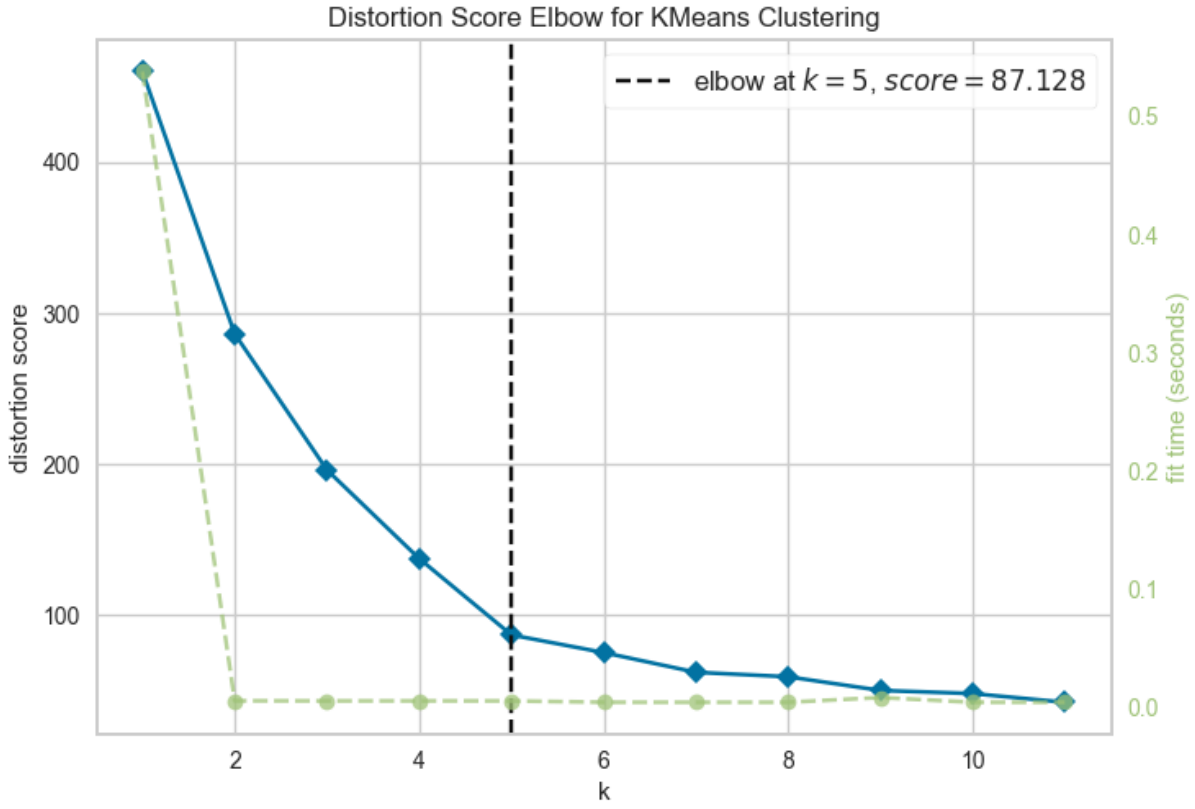


- Out of all the four variants of fuel type, the registrations of Electric Vehicles and Hybrid Vehicles show a noticeable increasing trend over the time period and seems to suggest a growing market in the near future.
- The upward trend may indicate a shift in consumer preferences towards more sustainable and environmentally friendly transportation options.
- The consistent demand for the remaining vehicles suggests a stable market presence, possibly driven by specific industry requirements or consumer preferences.



- Strong positive correlation exists between "DIESEL" and "PETROL" (0.87), suggesting that these fuel types tend to move together in terms of their values.
- Moderate positive correlation exists between "HYBRID COMBINED" and "PETROL" (0.54) and "HYBRID COMBINED" and "ELECTRIC(BOV)" (0.44), indicating some degree of co-occurrence.
- A weak positive correlation exists between "Year" and "ELECTRIC(BOV)" (0.23), suggesting a possible increase in electric vehicles over time.
- A strong negative correlation exists between "ELECTRIC(BOV)" and "DIESEL" (-0.84), indicating that an increase in electric vehicles is associated with a decrease in diesel usage.

[43]:
```
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from yellowbrick.cluster import KElbowVisualizer

model = KMeans()
visualizer = KElbowVisualizer(model, k=(1, 12))
visualizer.fit(scaledData)

# Display the elbow p
visualizer.show()
```

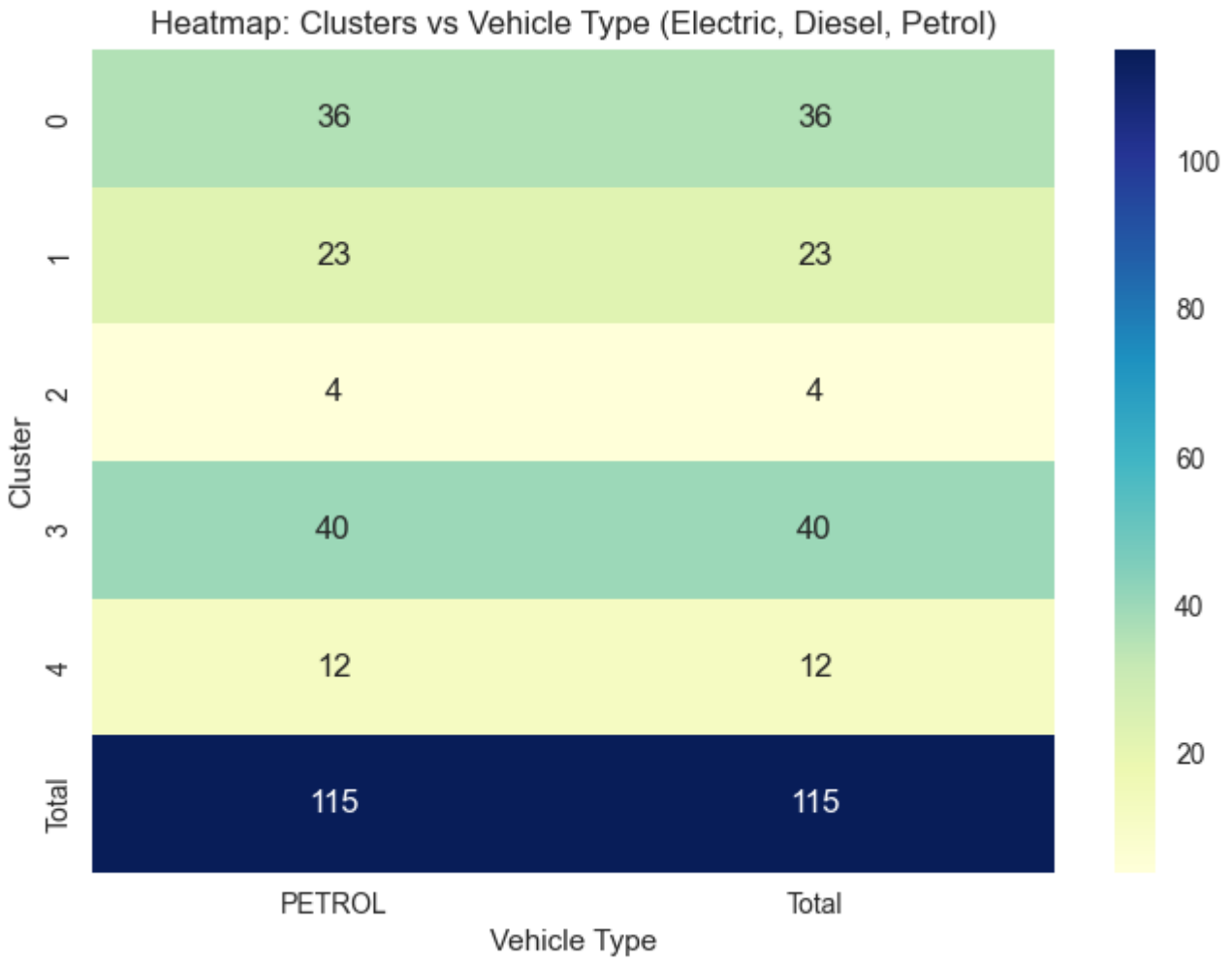


There is a clear and sharp elbow around k=5, where the decrease in distortion score starts to slow down significantly. This suggests that adding more clusters beyond k=5 does not provide a substantial improvement in clustering quality.

Based on the elbow method, the optimal number of clusters for the given data is likely k=5. This means that the KMeans algorithm can effectively group the data points into five distinct clusters with minimal distortion.

[44]:
```
# Applying K-means clustering
kmeans = KMeans(n_clusters=5)
finalDatasetFuelType['Cluster'] = kmeans.fit_predict(scaledData)

# Visualize or analyze the clusters
print(finalDatasetFuelType['Cluster'].value_counts())
```

```
Cluster
3    40
0    36
1    23
4    12
2     4
Name: count, dtype: int64
```



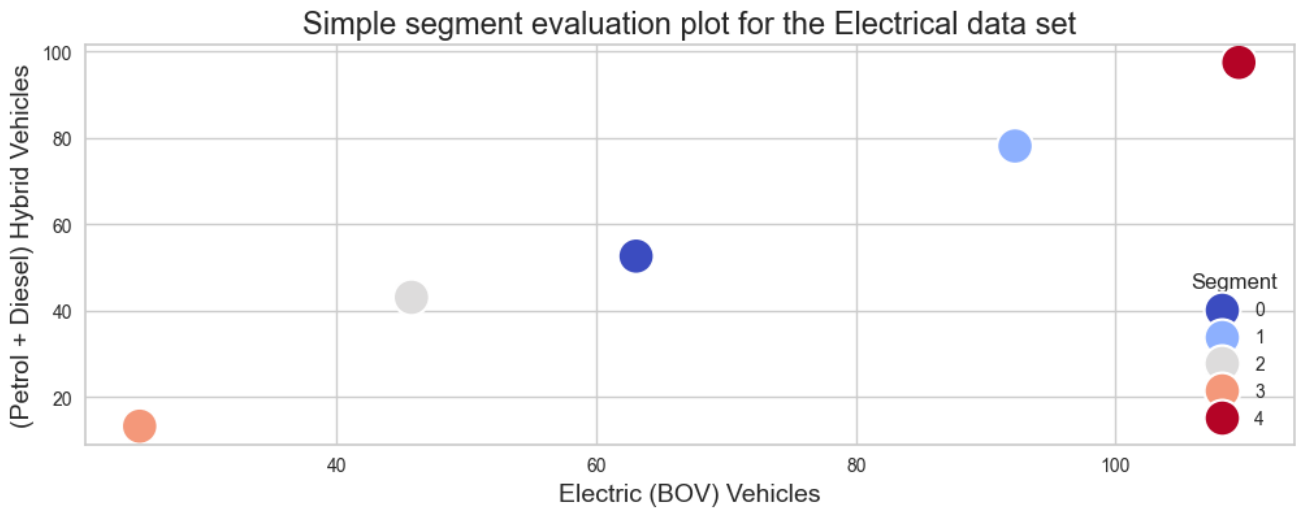
Cluster 0: This cluster has a more balanced distribution, with petrol vehicles being the most common, followed by electric and diesel.

Cluster 1: This cluster is heavily dominated by electric vehicles, significantly outnumbering diesel and petrol vehicles.

Cluster 2: This cluster has a moderate number of electric vehicles, followed by diesel and petrol.

Cluster 3: Diesel vehicles are the clear majority in this cluster, with very few electric or petrol vehicles.

[58]:
```
# Target segments

plt.figure(figsize = (12,4))
sns.scatterplot(x = "ELECTRIC(BOV)", y = "HYBRID COMBINED",data=segment,s=400, hue='Cluster Number', palette='coolwarm')
plt.title("Simple segment evaluation plot for the Electrical data set",
          fontsize = 17)
plt.xlabel("Electric (BOV) Vehicles", fontsize = 14)
plt.ylabel("(Petrol + Diesel) Hybrid Vehicles", fontsize = 14)
plt.legend(title='Segment', loc='lower right')
plt.show()
```

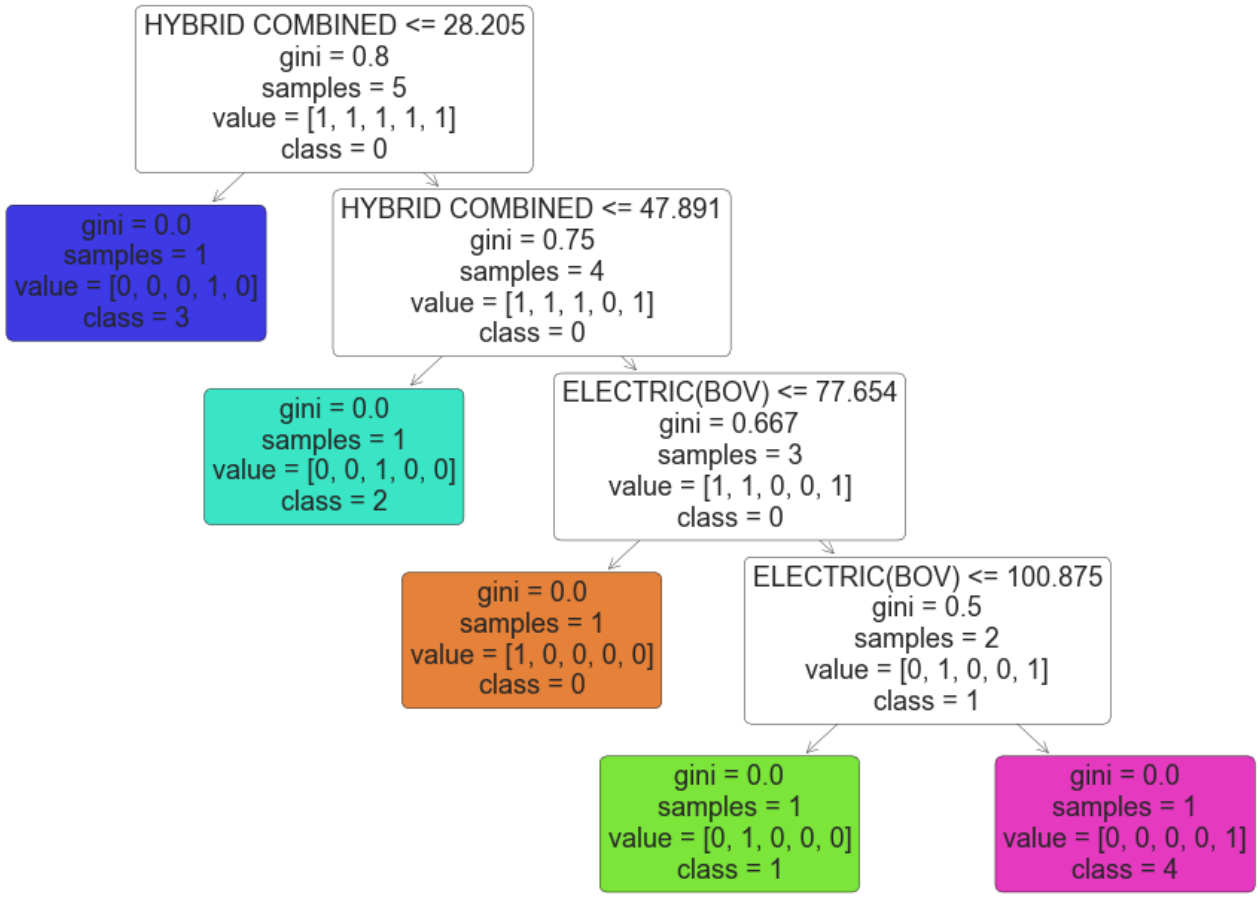


Electric (BOV) Vehicles: This refers to Battery Operated Vehicles, which encompass pure electric cars and electric two-wheelers.

(Petrol + Diesel) Hybrid Vehicles: These are vehicles that combine an electric motor with a traditional gasoline or diesel engine.

Data being divided into two types of electric vehicles: Battery Operated Vehicles (BOVs), which include electric cars and electric two-wheelers, and Hybrid Vehicles, which use both electricity and traditional fuels like petrol or diesel. Now, if you plot this data on a graph, you'll see points spread out in all four sections of the graph, showing a mix of adoption for both electric and hybrid vehicles across different groups. However, you might notice that there are more points clustered in the bottom left and

top right corners of the graph. This suggests that some groups prefer either electric or hybrid vehicles more, while others might like both types equally.



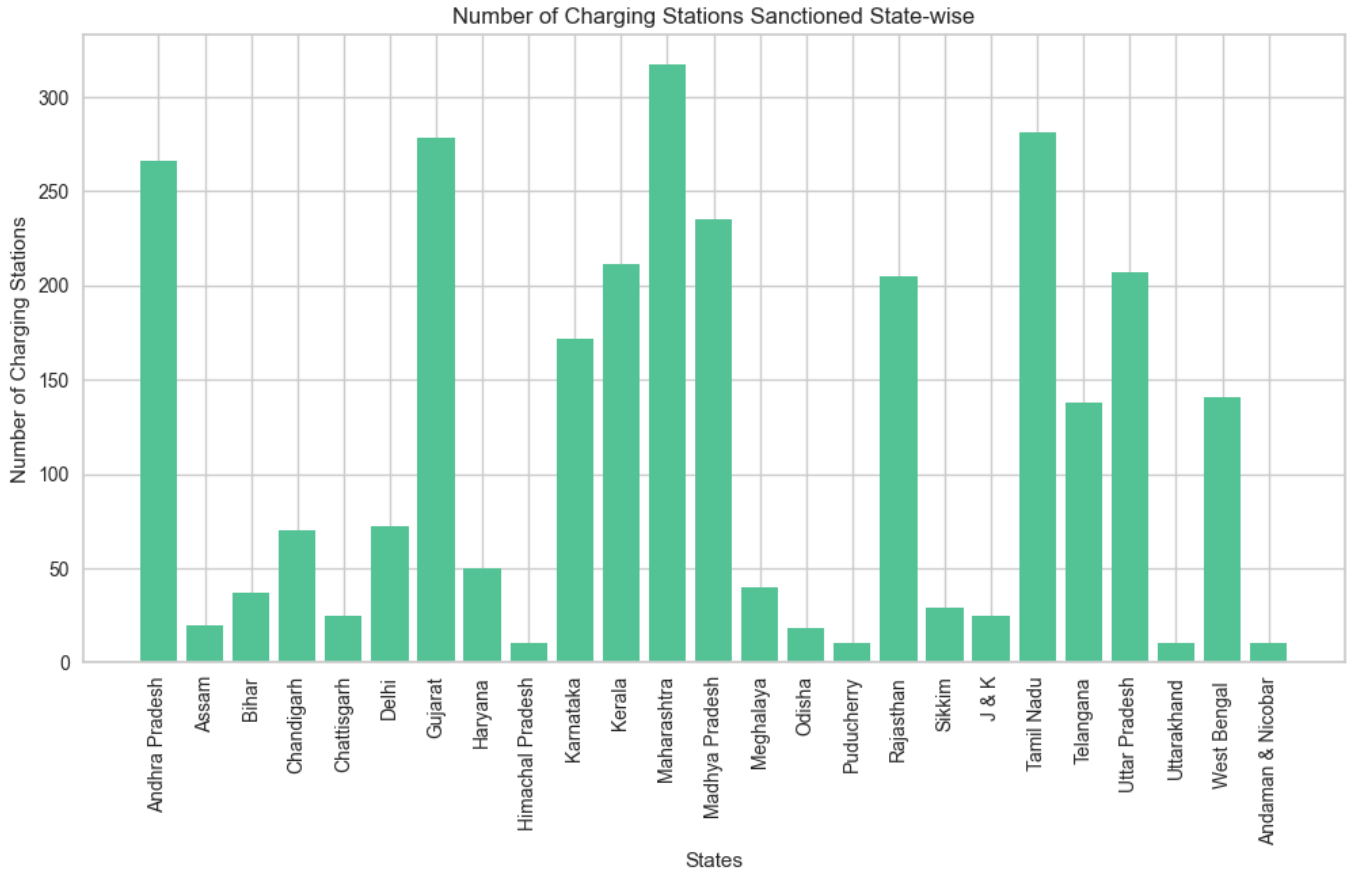
Target Segment 0: Prioritize marketing EVs' environmental benefits and lower running costs. Ensure easy access to charging infrastructure information.

Segment 4: Address price concerns and highlight the flexibility of having both hybrid and electric options.

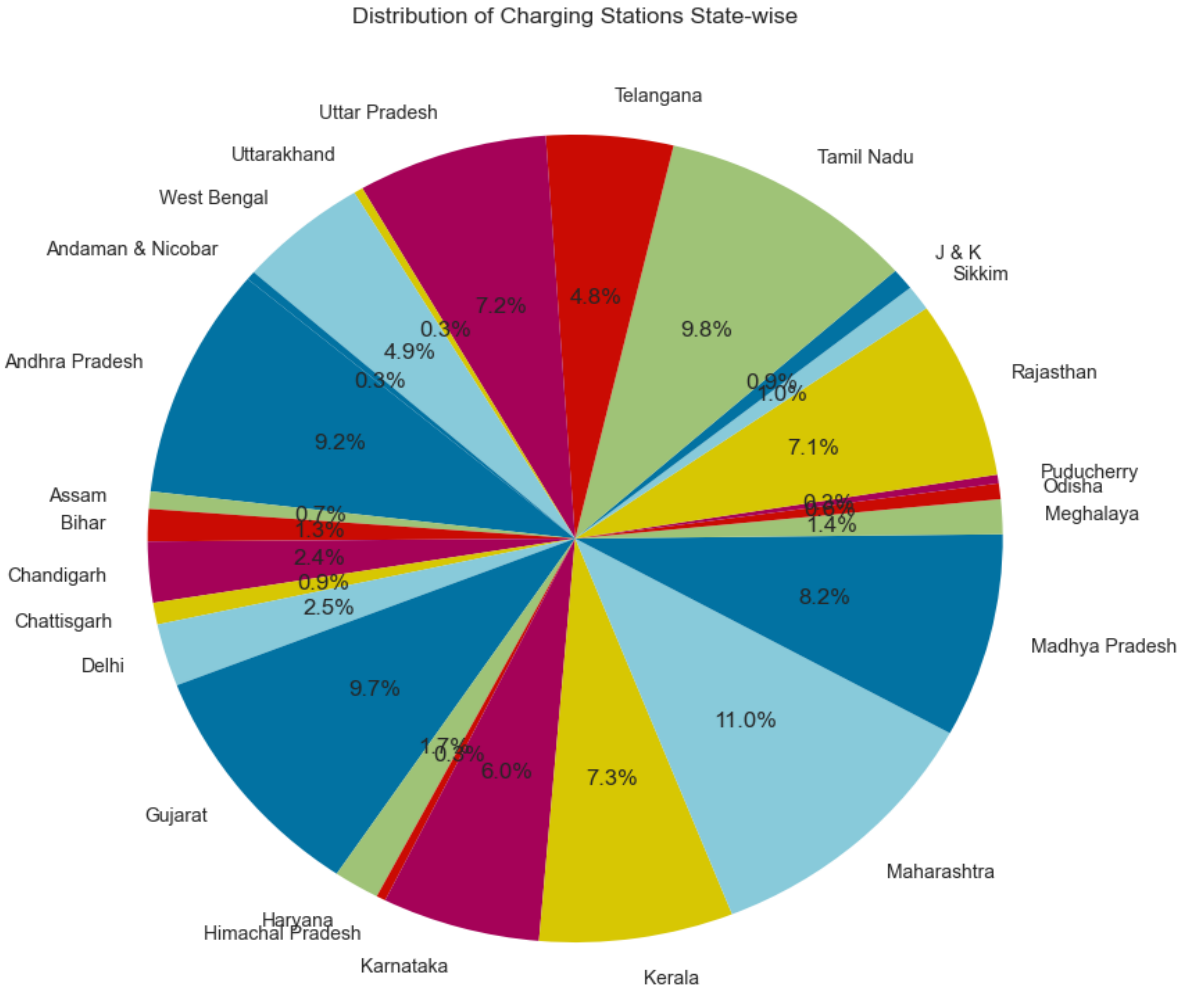
Segments 1 and 5: Offer a diverse range of EV and hybrid options cater to various needs and budgets. Emphasize fuel efficiency and the evolving charging infrastructure landscape.

Segments 2 and 3: Focus on the practicality and reliability of hybrids while acknowledging the growing appeal of EVs. Address range anxiety concerns.

State-wise Charging Station Sanctioned



PERCENTAGE WISE

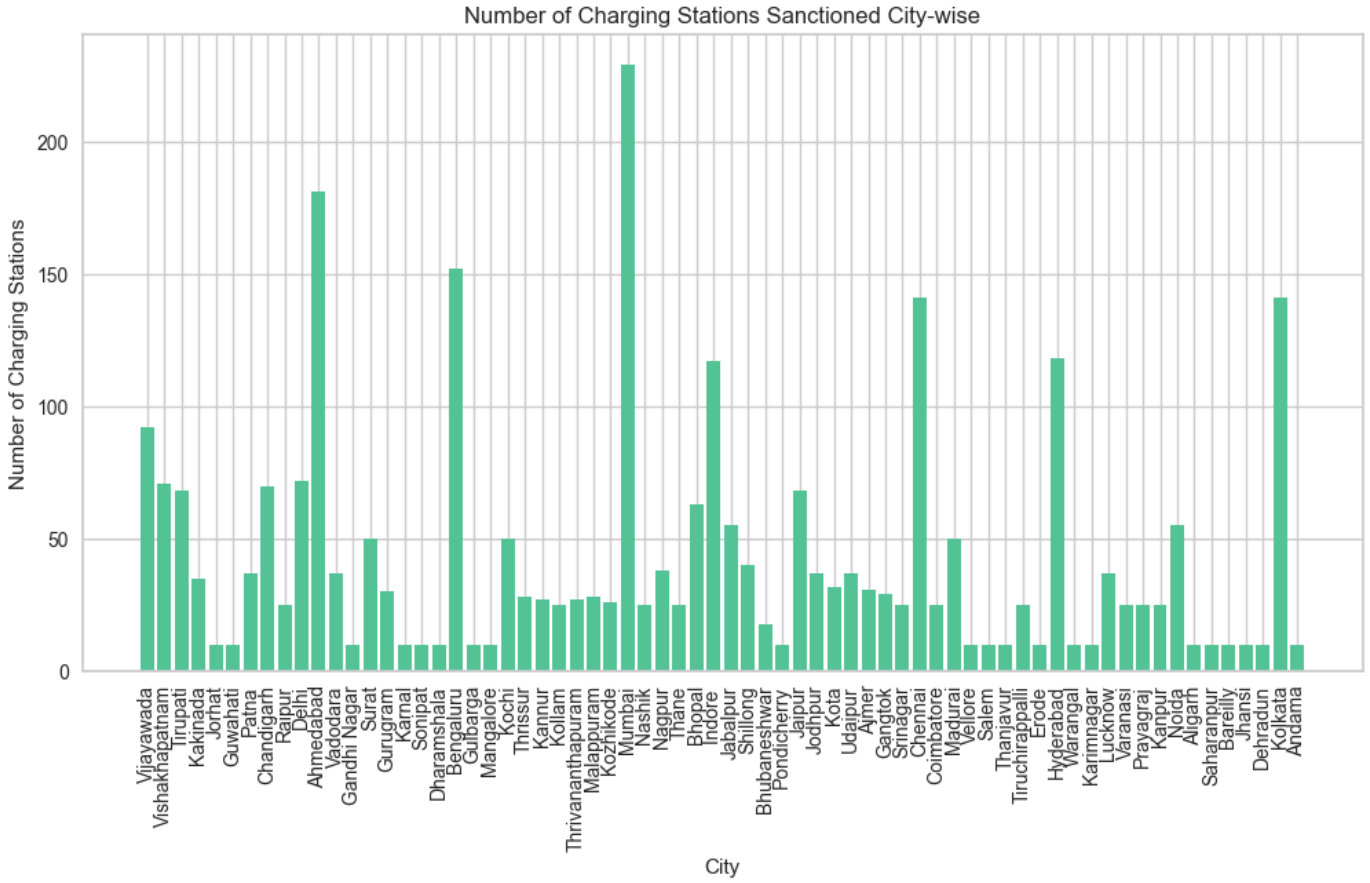


Based on the following graph we can see that the states with maximum number of sanctioned charging

stations are:

1. Maharashtra
2. Tamil Nadu
3. Gujarat
4. Andhra Pradesh
5. Madhya Pradesh

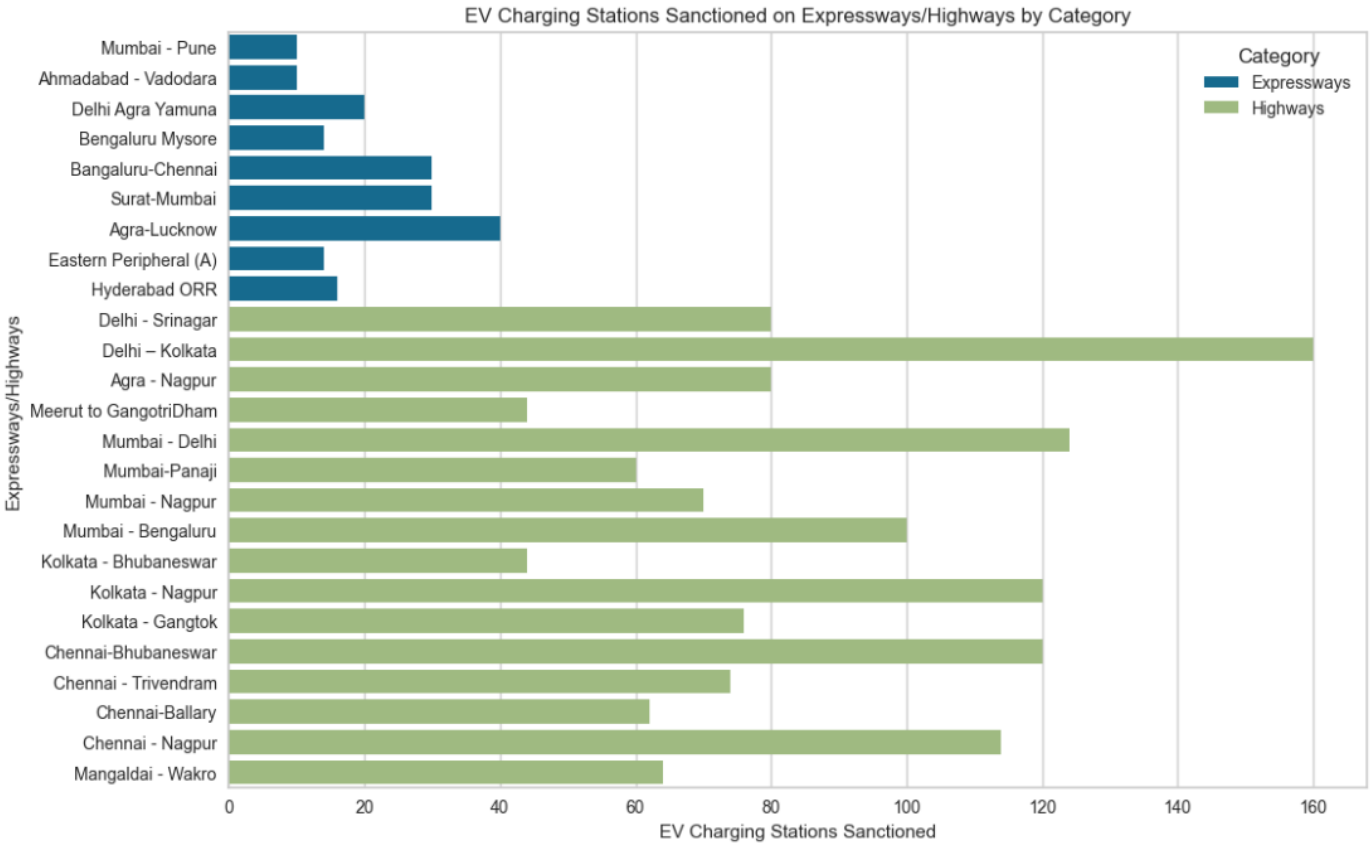
City-wise Charging Station Sanctioned



Based on the following graph we can see that the cities with maximum number of sanctioned charging

stations are:

1. Mumbai
2. Ahmedabad
3. Bengaluru
4. Kolkata
5. Chennai



Delhi-Agra Yamuna Expressway boasts the most stations, reflecting its high traffic volume, while several highways lack any infrastructure.

With a total of 1576 stations across the network, expansion seems crucial to support wider EV adoption on Indian highways.

CONCLUSION

The way people move around in India is changing a lot. More and more people are starting to care about the environment, so they're interested in electric vehicles (EVs) instead of just petrol cars. Even though petrol cars are still the most common, places like Maharashtra, Tamil Nadu, and Gujarat, as well as big cities like Mumbai, Ahmedabad, and Bengaluru, are seeing a big increase in people using electric vehicles. This is great news for our EV startup because it means there are two important groups of people we can focus on: those states and cities where EVs are really catching on. So, by targeting these areas, we have a good chance of success in the growing electric vehicle market in India.

Target Audiences for the Startup

Target Audience 1: People Who Care About the Environment and Like Trying New Things

- They care about the environment and like to use products that are good for it.
- They are interested in electric cars but worry about where to charge them.
- They like to see new and cool features in the cars they buy.

Target Audience 2: People Who Want Good Value for Money and Are Open to Different Types of Cars

- They want to spend their money wisely and look for good deals.
- They are open to buying either electric cars or hybrid cars.
- They care about saving money on fuel and want cars that are efficient.