Electric Vehicle Market Segmentation

Analysing the respective market in India using Segmentation analysis for ELECTRIC VEHICLE

by

Murari Muniprathap
Archana S
Padala Akhil Reddy
Ishmeet Kaur
Shashi Kumar Reddy

Abstract:

The electric vehicle market is rapidly growing, and with the increasing number of electric vehicle models available, it has become essential for companies to identify and target specific customer segments. Market segmentation involves dividing a larger market into smaller groups of consumers with similar needs and characteristics. This approach helps companies to develop targeted marketing strategies and tailor their product offerings to meet the specific needs of each customer group. In the electric vehicle market, segmentation is typically based on factors such as customer demographics, driving habits, and charging infrastructure availability. By understanding the different segments within the electric vehicle market, companies can better position their products and services to meet the unique needs and preferences of each group, ultimately driving sales and market share.

Problem Statement:

(EV Market) You are a team working under an Electric Vehicle Startup. The Startup is still deciding in which vehicle/customer space it will be develop its EVs. You have to analyse the Electric Vehicle market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles.

Fermi Estimation

Assumptions:

- India's population is approximately 1.4 billion people.
- The percentage of people who own a vehicle in India is around 10%, or 140 million people.
- The electric vehicle market in India is currently at a nascent stage, with an estimated 1% market share.
- The electric vehicle market is expected to grow at a compound annual growth rate of 30% over the next five years.

Estimations:

- The potential market size for electric vehicles in India is around 1.4 million vehicles.
- Assuming a segmentation approach, the team could consider the following categories of segments based on data availability:
    - Geographic: Segmenting customers by their location and the availability of charging infrastructure.
    - Demographic: Segmenting customers by age, gender, income, education, and occupation.
    - Psychographic: Segmenting customers by their lifestyle, values, and attitudes towards electric vehicles.
    - Behavioral: Segmenting customers by their purchase behavior, usage patterns, and attitudes towards electric vehicles.
    - B2B: Segmenting customers by their industry, fleet size, and usage patterns.

Based on the above assumptions and estimations, the startup should focus on targeting the following segments:

- Urban consumers in large cities with a high density of vehicles and a growing network of charging infrastructure.
- Early adopters who are environmentally conscious and have a higher income level.
- Fleet operators in the B2B segment who have a high usage rate and a need for cost savings.
- Customers who frequently commute short distances and can benefit from the lower running cost of electric vehicles.

1. Data Sources:

The data used in the report are obtained from the following sources:

1. Kaggle:

https://www.kaggle.com/code/fabiendaniel/customer-segmentation/input?select=data.csv
https://www.kaggle.com/datasets/kkhandekar/electric-vehicles-india
https://www.kaggle.com/datasets/fathimaibrahimkunju/electric-vehicle-in-india-2022
https://pib.gov.in/PressReleasePage.aspx?PRID=1842704
https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset

We gather all this information by browsing through internet. Most of the datasets used in this project are from Kaggle, some other data from government sites and one was collected by survey.

2.**Data Preprocessing:**

Data preprocessing is an important step in the data mining process. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data mining task.
Data preprocessing transforms the data into a format that is more easily and effectively processed in data mining, machine learning and other data science tasks. The techniques are generally used at the earliest stages of the machine learning and AI development pipeline to ensure accurate results.
There are several different tools and methods used for preprocessing data,including the following:
• sampling, which selects a representative subset from a large population of data;
• transformation, which manipulates raw data to produce a single input;
• denoising, which removes noise from data;
• imputation, which synthesizes statistically relevant data formissing values;
• normalization, which organizes data for more efficient access; and
• feature extraction, which pulls out a relevant feature subset that is significant in a particular context.

Libraries That are used in this market segmentation are pandas, numpy, matplotlib, seaborn, sklearning models etc.

there are the imported libraries

```
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import Kmeans
from statsmodels.graphics.mosaicplot import mosaic
```

State-wise EV and non EV dataset

	Sr. No.	State Name	Total Electric Vehicle	Total Non-Electric Vehicle	Total
0	1	Andaman & Nicobar Island	162	1,46,945	1,47,107
1	2	Andra Pradesh	NaN	NaN	NaN
2	3	Arunachal Pradesh	20	2,52,965	2,52,985
3	4	Assam	64766	46,77,053	47,41,819
4	5	Bihar	83335	1,04,07,078	1,04,90,413

Dataset of 2W 3W 4W and bus and chargers

```
df1.head()
```

	Region	2W	3W	4W	Bus	Chargers
0	Uttar Pradesh	9852	42881	458	197	207
1	Maharastra	38558	893	1895	186	317
2	Karnataka	32844	568	589	57	172
3	Tamil Nadu	25642	396	426	0	256
4	Gujarat	22359	254	423	22	228

Customer based Dataset

	Unnamed:	Age	City	Profession	Marital Status	Education	No. of Family members	Annual Income	Would you prefer replacing all your vehicles to Electronic vehicles?	If Yes/Maybe what type of EV would you prefer?	Do you think Electronic Vehicles are economical?	Which brand of vehicle do you currently own?	How much money could you spend on an Electronic vehicle?	Preference for wheels in EV	Do you think Electronic vehicles will replace fuel cars in India?
0	0	30	Nabha	None	Single	Graduate	5	1193875.647	Maybe	SUV	Yes	Hyundai	<5 lakhs	2	I don't think so
1	1	27	Pune	None	Single	Graduate	4	1844540.398	Yes	SUV	Yes	Honda	<15 lakhs	4	Yes, in <20years
2	2	32	Kashipur	None	Single	Graduate	4	2948150.113	Yes	Hatchback	Yes	KIA	<15 lakhs	4	Yes, in <20years
3	3	55	Pune	Business	Single	Graduate	3	2832379.739	Maybe	Hatchback	No	Hyundai	<5 lakhs	4	Yes, in <10 years
4	4	26	Satara	None	Single	Graduate	4	2638750.576	Yes	Sedan	Yes	McLaren	<15 lakhs	4	Yes, in <20years

Customer based dataset

```
data.head()
```

	Names	gender	age	education	income	employment	state
0	Aadhya	F	25	graguate	650000	yes	Maharashtra
1	Aaliyah	F	32	10th	950000	yes	Uttar Pradesh
2	Aaradhya	F	28	postgraguate	1100000	yes	Rajasthan
3	Aaryan	M	25	12th	300000	yes	Harayana
4	Abhinav	M	32	10th	360000	yes	Uttar Pradesh

preprocessing for customer related data

```
df.isnull().sum()
```

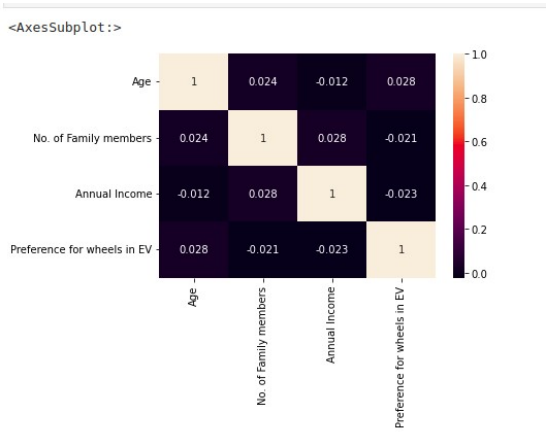
```
Unnamed:                                                          0
Age                                                               0
City                                                              0
Profession                                                        0
Marital Status                                                    0
Education                                                         0
No. of Family members                                             0
Annual Income                                                     0
Would you prefer replacing all your vehicles to Electronic vehicles?    0
If Yes/Maybe what type of EV would you prefer?                    0
Do you think Electronic Vehicles are economical?                  0
Which brand of vehicle do you currently own?                      0
How much money could you spend on an Electronic vehicle?          0
Preference for wheels in EV                                       0
Do you think Electronic vehicles will replace fuel cars in India?    0
dtype: int64
```

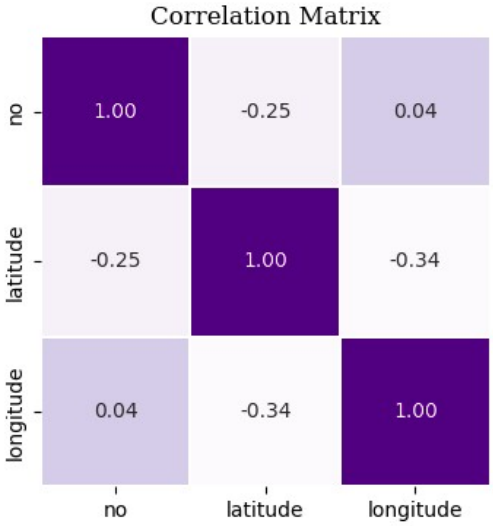
preprocessing for wheels based dataset

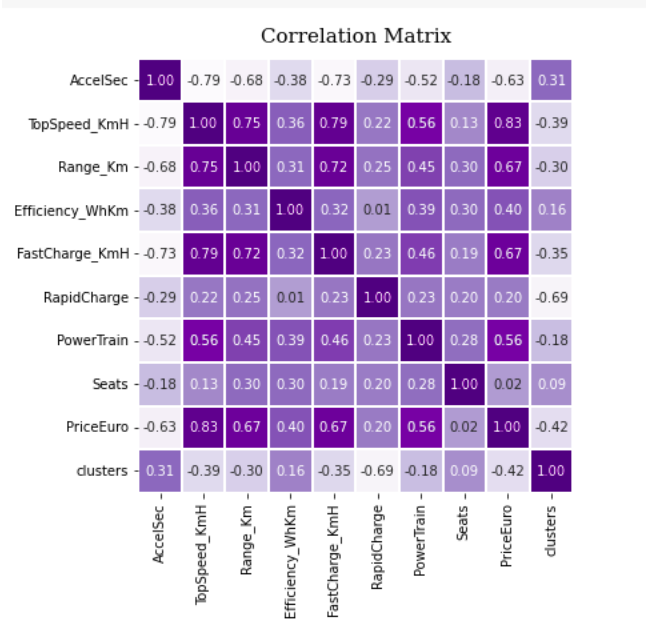
```
df1.info()
```

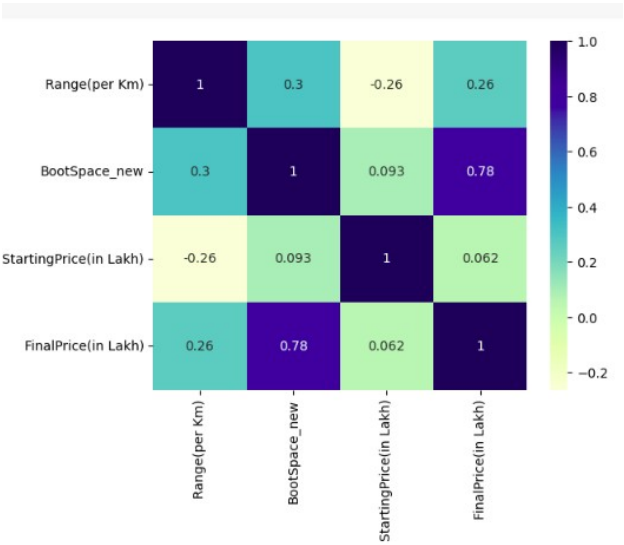
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24 entries, 0 to 23
Data columns (total 6 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   Region    24 non-null     object
 1   2W        24 non-null     int64 
 2   3W        24 non-null     int64 
 3   4W        24 non-null     int64 
 4   Bus       24 non-null     int64 
 5   Chargers  24 non-null     int64 
dtypes: int64(5), object(1)
memory usage: 1.2+ KB
```

correlations of datasets used









3: Segment Extraction

Segmentation with K-means Clustering We are going to use K-means algorithm from scikit-learn. Let’s first understand how the algorithm will form customer groups:

Initialize k=n centroids=number-of-clusters randomly or smartly

Assign each data point to the closest centroid based on Euclidian distance, thus forming the groups

Move centres to the average of all points in the cluster

While running the steps through, the algorithm checks the sum of squared distances between clustered-point and centre for each cluster.

Mathematically speaking, it tries to minimize — optimize the within-cluster sum-of-squared-distances or inertia of each cluster.

$$\sum_{i=0}^{n} \min_{\mu_j \in C} (||x_i - \mu_j||^2)$$

Mathematical expression of within-cluster sum-of-squared-distances or inertia where X is the points in the cluster and μ is the current centroid

When inertia value does not minimize further, algorithm converges. Thus,iteration stops.

```
from sklearn.cluster import Kmeans
kmeans_model = KMeans(init='k-means++', max_iter=500, random_state=42)
```

init parameter with the k-means++ allows the algorithm to place initialcenters smartly, rather than random.

max_iter is the maximum number of iterations of the algorithm in a single run, default value is 300.

random_state guarantees the reproducibility of the model results.

This algorithm is easy to understand, fits well to large datasets in terms of computing times and guarantees convergence. However, when centroids are initialized randomly, algorithm may not assign the points to the groups in the most optimal way.

One important consideration is the selection of k. In other words, how many groups should be formed? For example, K-means applied above uses k=8 as a default value.

Hyperparameter Tuning

While selecting k, we are going to decide against the optimization criteria of the K-means, inertia, using elbow method. We are going to build different K-means models with k values 1 to 15, and save the corresponding inertia values.

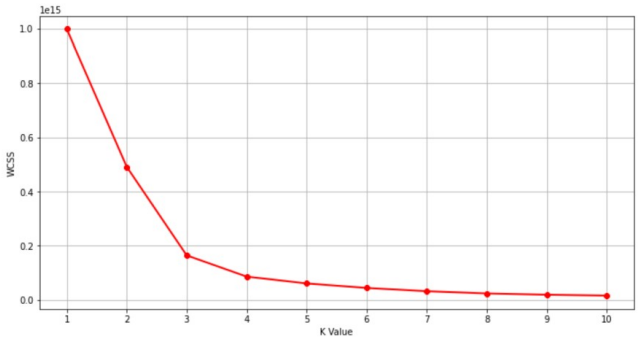
With the elbow method, we are going to select the k value where the decrease in the inertia stabilizes.

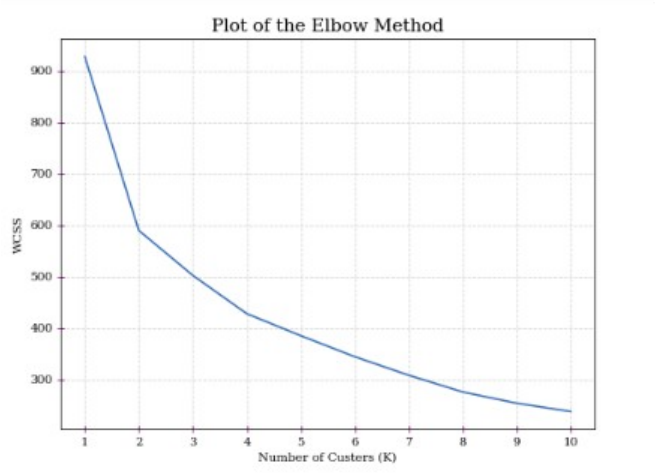
When k=1 inertia is at the highest, meaning data is not grouped yet. Inertia decreases steeply until k=2. Between k=2 and 4, the curve continues to decrease fast.

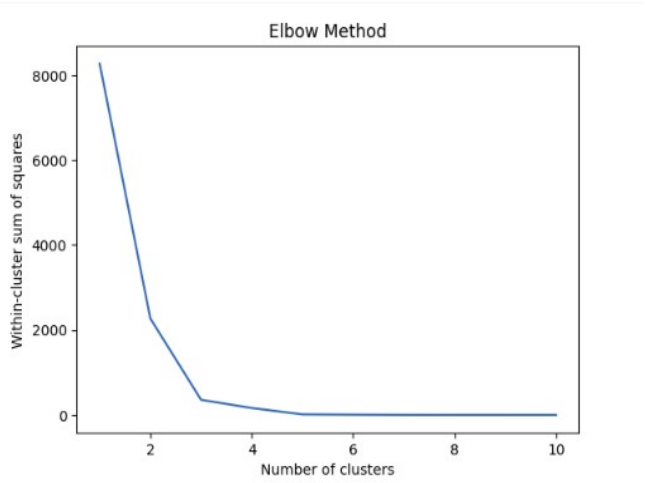
At k=4, the descent stabilizes and continues linearly afterwards, forming an elbow at k=4. This points out the optimal number of customer group is 4.

K MEANS MODEL

```
In [142]: from sklearn.cluster import KMeans
          wcss = []
          for k in range(1,11):
              kmeans = KMeans(n_clusters=k, init="k-means++",random_state=28)
              kmeans.fit(X)
              wcss.append(kmeans.inertia_)
          plt.figure(figsize=(12,6))
          plt.grid()
          plt.plot(range(1,11),wcss, linewidth=2, color="red", marker ="8")
          plt.xlabel("K Value")
          plt.xticks(np.arange(1,11,1))
          plt.ylabel("WCSS")
          plt.show()
```

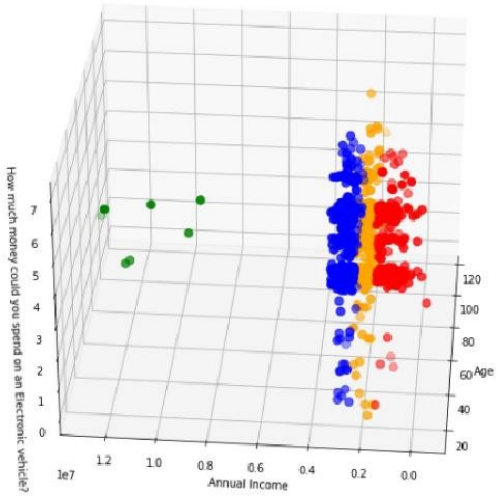






K-Mean Model for Customer data:

```
In [143]:
km = KMeans(n_clusters=4, random_state=28)
clusters = km.fit_predict(df)
df["Cluster"] = clusters

df_Original["Cluster"] = clusters

from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

fig = plt.figure(figsize=(20,10))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(df.Age[df.Cluster == 0], df["Annual Income"][df.Cluster == 0], df["How much money could you spend on an Electronic vehicle?"][
ax.scatter(df.Age[df.Cluster == 1], df["Annual Income"][df.Cluster == 1], df["How much money could you spend on an Electronic vehicle?"][
ax.scatter(df.Age[df.Cluster == 2], df["Annual Income"][df.Cluster == 2], df["How much money could you spend on an Electronic vehicle?"][
ax.scatter(df.Age[df.Cluster == 3], df["Annual Income"][df.Cluster == 3], df["How much money could you spend on an Electronic vehicle?"][

ax.view_init(30, 185)
plt.xlabel("Age")
plt.ylabel("Annual Income")
ax.set_zlabel('How much money could you spend on an Electronic vehicle?')
plt.show()
```



4:Target Market

A target market is a group of people that have been identified as the most likely potential customers for a product because of their shared characteristics such as age, income, and lifestyle.

Identifying the target market is a key part of the decision-making process when a company designs, packages, and advertises its product.

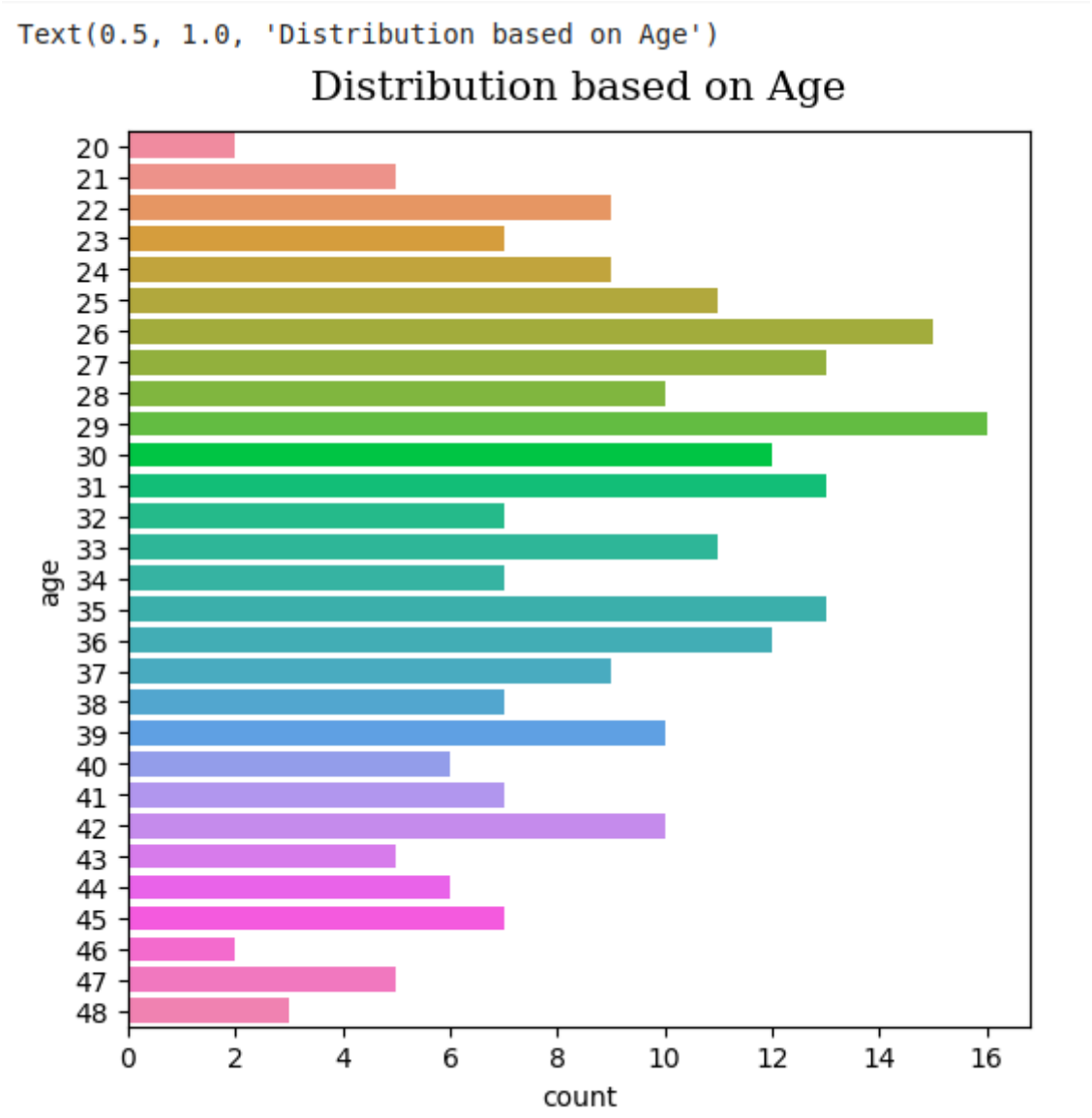
The 4 Target Markets are:
Marketing professionals divide consumers into four major segments:

Demographic: These are the main characteristics that define your target market. Everyone can be identified as belonging to a specific age group, income level, gender, occupation, and education level.

Distribution based on Age

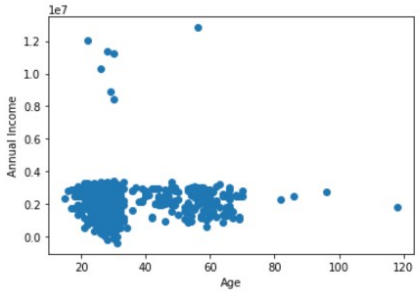
```
fig = plt.figure(figsize=(6,6))
sns.countplot(y="age", data=data)
plt.title(label="Distribution based on Age",weight=200, family='serif',
size=15, pad=12)
```

```
Text(0.5, 1.0, 'Distribution based on Age')
```



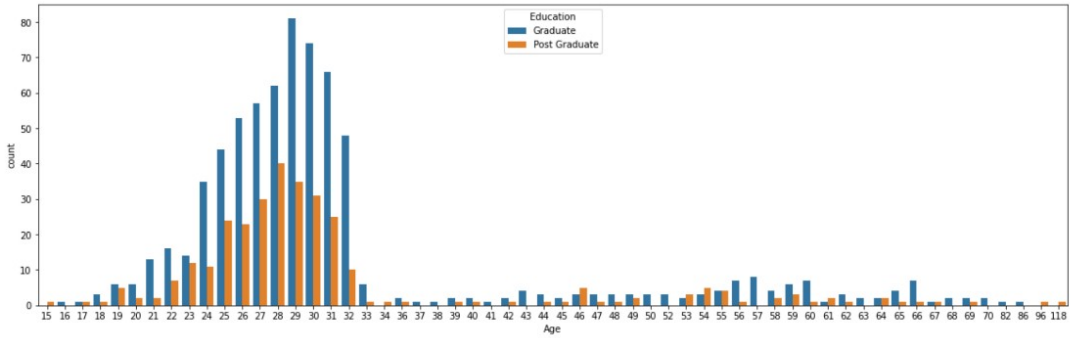
In [125]:
```
plt.xlabel('Age')
plt.ylabel('Annual Income')
plt.scatter(df['Age'],df['Annual Income'])
```

Out[125]:
```
<matplotlib.collections.PathCollection at 0x23f240774f0>
```



In [126]:
```
plt.figure(figsize=(20,6))
sns.countplot(x="Age", data=df, hue="Education")
```

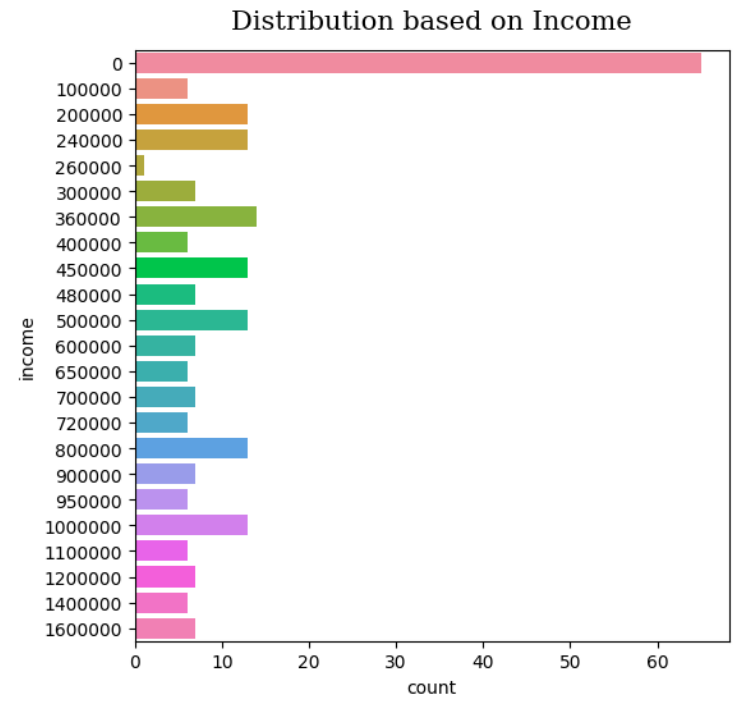
Out[126]:
```
<AxesSubplot:xlabel='Age', ylabel='count'>
```



Distribution based on Income

```
fig = plt.figure(figsize=(6,6))
sns.countplot(y="income", data=data)
plt.title(label="Distribution based on Income",weight=200, family='serif',
size=15, pad=12)
```

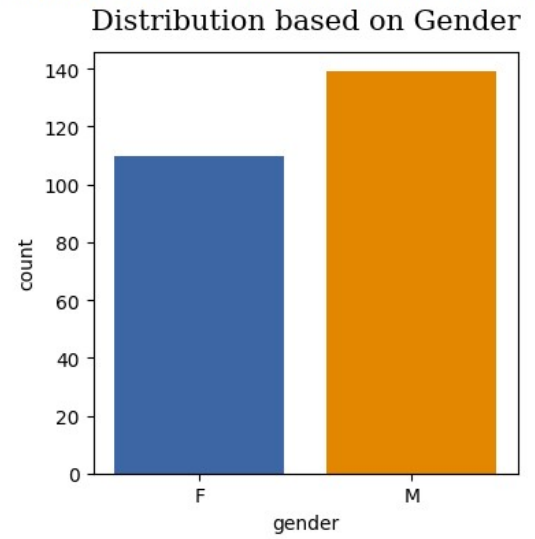
```
Text(0.5, 1.0, 'Distribution based on Income')
```



Distribution based on Gender

```
fig = plt.figure(figsize=(4,4))
sns.countplot(x="gender", data=data)
plt.title(label="Distribution based on
Gender",weight=200, family='serif',
size=15, pad=12)
```

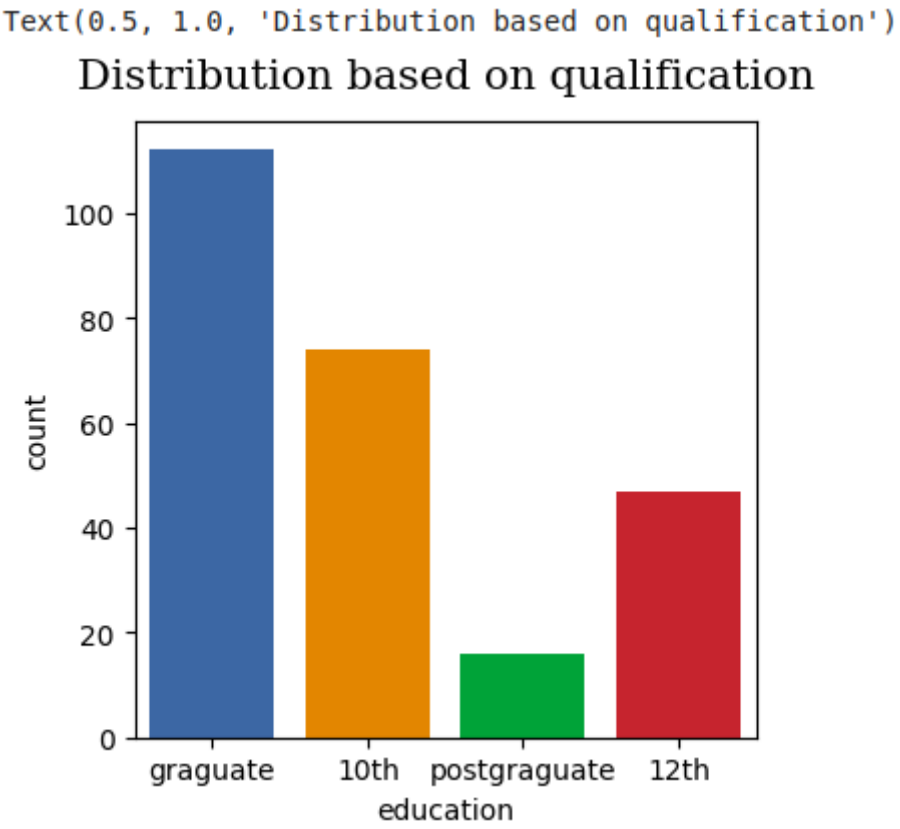
```
Text(0.5, 1.0, 'Distribution based on Gender')
```



Distribution based on qualification

```
fig = plt.figure(figsize=(4,4))
sns.countplot(x="education", data=data)
plt.title(label="Distribution based on qualification",weight=200,
family='serif', size=15, pad=12)
```

```
Text(0.5, 1.0, 'Distribution based on qualification')
```

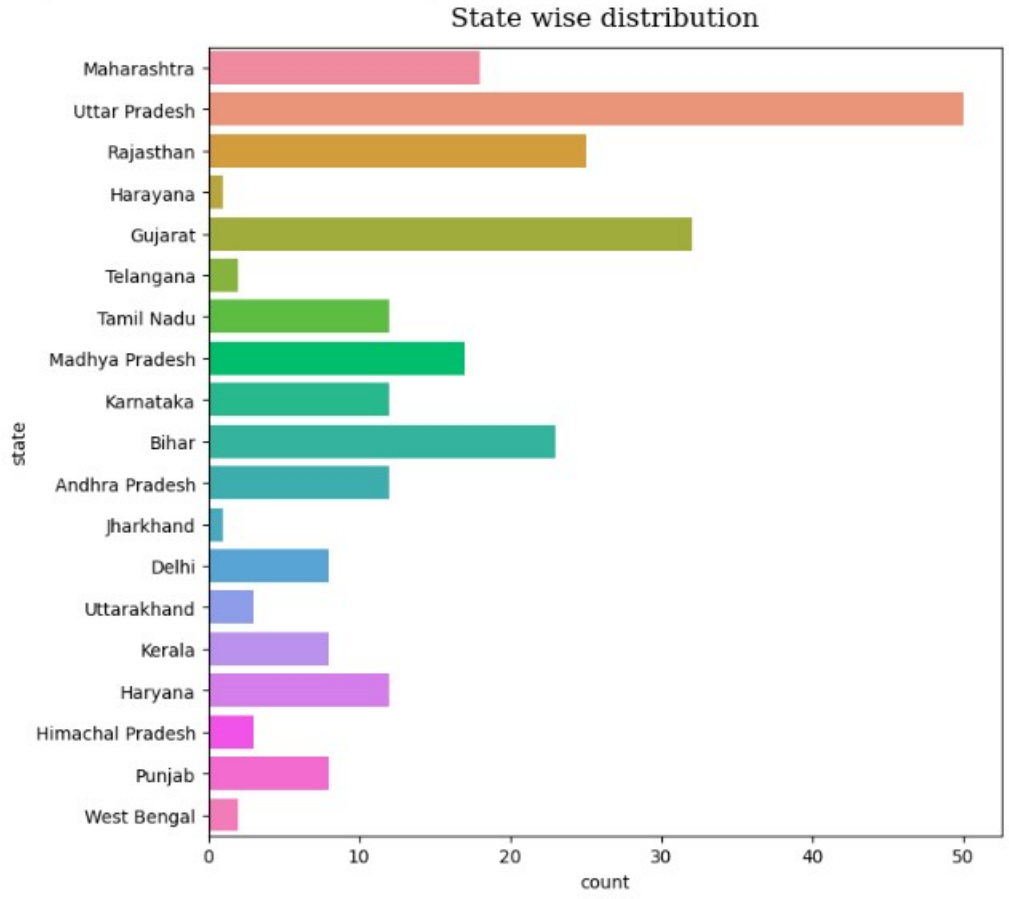


Geographic: This segment is increasingly relevant in the era of globalization. Regional preferences need to be taken into account.

State Wise Distribution

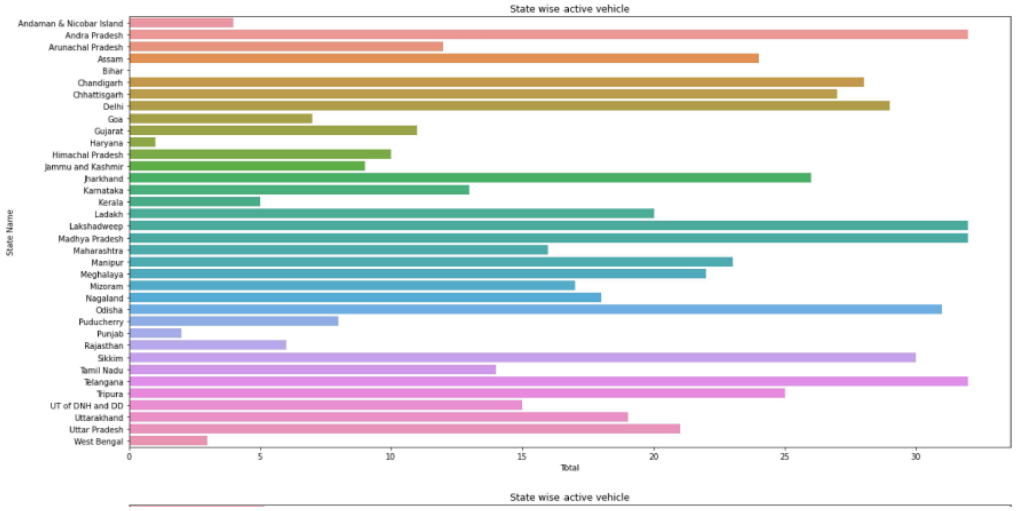
```
fig = plt.figure(figsize=(8,8))
sns.countplot(y="state", data=data)
plt.title(label="State wise distribution",weight=200, family='serif',
size=15, pad=12)
```

```
Text(0.5, 1.0, 'State wise distribution')
```



In [254]:

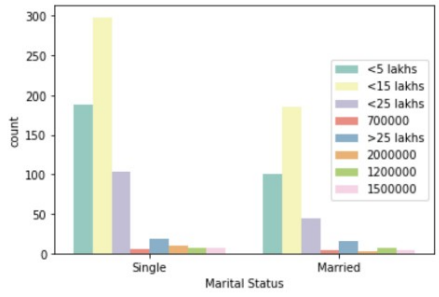
```
plt.rcParams['figure.figsize'] = (20, 10)
f = sns.barplot(y=Active_EVs['State Name'], x=Active_EV['Total'])
plt.title('State wise active vehicle')
plt.show()


plt.rcParams['figure.figsize'] = (20, 10)
f = sns.barplot(y=Active_EV['State Name'], x=Active_EV['Total Electric Vehicle'])
plt.title('State wise active vehicle')
plt.show()
```



Psychographic: This segment goes beyond the basics of demographics to consider lifestyle, attitudes, interests, and values.

In [123]:

```
sns.countplot(x ='Marital Status', hue = 'How much money could you spend on an Electronic vehicle?',  data = df, palette = 'Set3'
plt.legend(loc='center right')
plt.show()
```



5:Describing Segments:

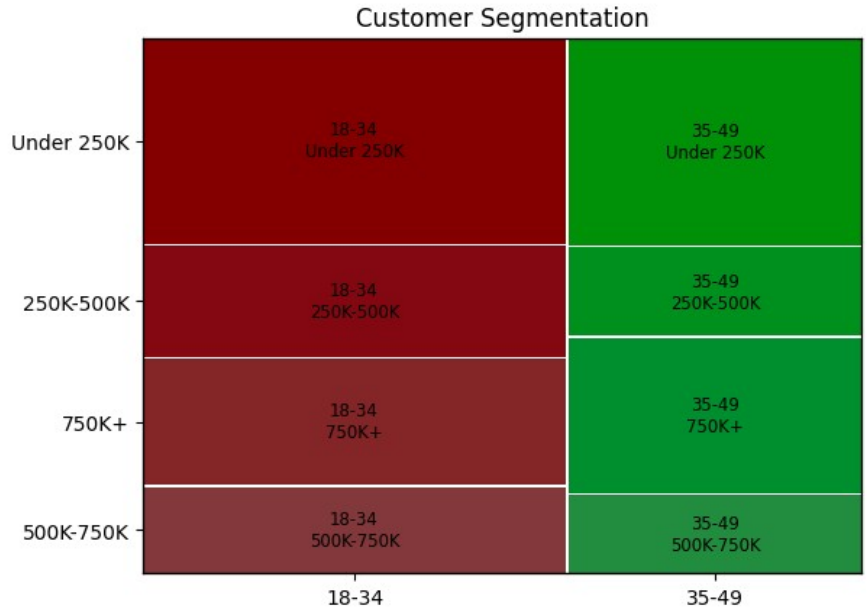
```
# segment customers based on age
age_groups = []
for age in data['age']:
      if age < 18:
            age_groups.append('Under 18')
      elif age >= 18 and age < 35:
            age_groups.append('18-34')
```

```
        elif age >= 35 and age < 50:
                age_groups.append('35-49')
        else:
                age_groups.append('50+')

# add age groups to customer data
data['age_group'] = age_groups

# segment customers based on income
income_groups = []
for income in data['income']:
        if income < 250000:
                income_groups.append('Under 250K')
        elif income >= 250000 and income < 500000:
                income_groups.append('250K-500K')
        elif income >= 500000 and income < 750000:
                income_groups.append('500K-750K')
        else:
                income_groups.append('750K+')

# add income groups to customer data
data['income_group'] = income_groups
```

```
[126] data.head()
```

	Names	gender	age	education	income	employment	state	age_group	income_group
0	Aadhya	F	25	graguate	650000	yes	Maharashtra	18-34	500K-750K
1	Aaliyah	F	32	10th	950000	yes	Uttar Pradesh	18-34	750K+
2	Aaradhya	F	28	postgraguate	1100000	yes	Rajasthan	18-34	750K+
3	Aaryan	M	25	12th	300000	yes	Harayana	18-34	250K-500K
4	Abhinav	M	32	10th	360000	yes	Uttar Pradesh	18-34	250K-500K

```
mosaic(data, ['age_group', 'income_group'], title='Customer Segmentation')
plt.show()
```




```
# segment customers based on gender
gender_groups = []
for gender in data['gender']:
      if gender == 'M':
            gender_groups.append('Male')
      elif gender == 'F':
            gender_groups.append('Female')
      else:
            gender_groups.append('Other')

# add gender groups to customer data
data['gender_group'] = gender_groups

# segment customers based on employment status
employment_groups = []
for employment in data['employment']:
      if employment == 'yes':
            employment_groups.append('employed')
      else:
            employment_groups.append('Unemployed')

# add employment groups to customer data
data['employment_group'] = employment_groups
```

```
data.tail()
```

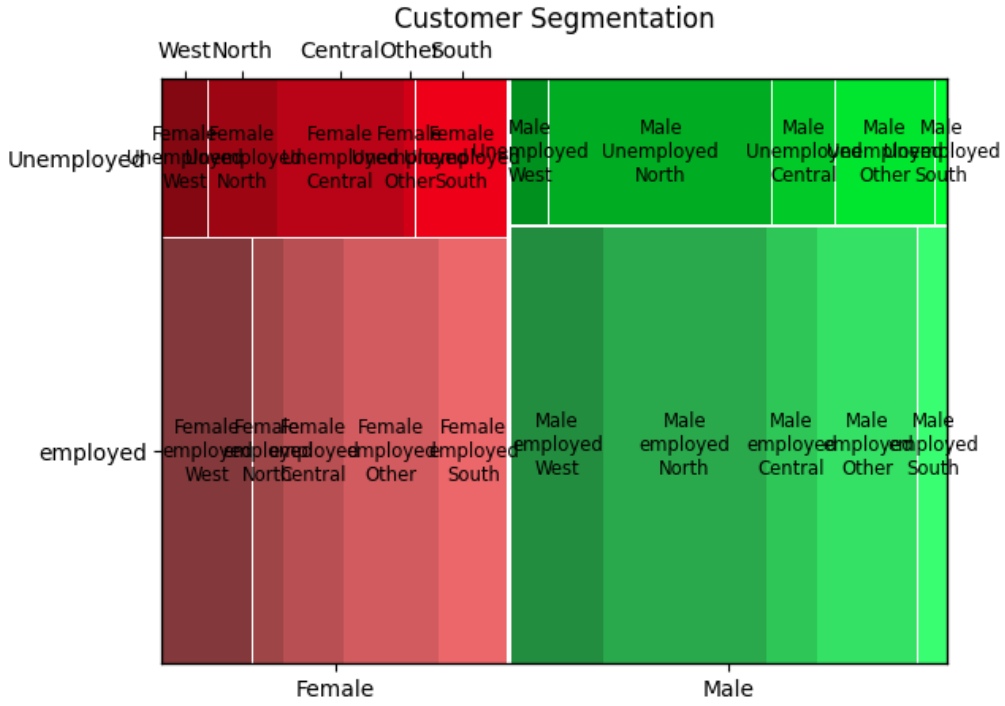
	Names	gender	age	education	income	employment	state	age_group	income_group	gender_group	employment_group
244	Sudhanshu	M	48	12th	1400000	yes	Uttar Pradesh	35-49	750K+	Male	employed
245	Sumit	M	23	graguate	500000	yes	Gujarat	18-34	500K-750K	Male	employed
246	Sunil	M	42	graguate	400000	yes	Bihar	35-49	250K-500K	Male	employed
247	Suresh	M	29	graguate	360000	yes	Haryana	18-34	250K-500K	Male	employed
248	Uday	M	36	12th	450000	yes	Madhya Pradesh	35-49	250K-500K	Male	employed

```
state_groups = []
for state in data['state']:
      if state in ['Maharashtra', 'Gujarat', 'Goa']:
            state_groups.append('West')
      elif state in ['Delhi', 'Uttar Pradesh', 'Haryana']:
            state_groups.append('North')
      elif state in ['Tamil Nadu', 'Karnataka', 'Kerala']:
            state_groups.append('South')
      elif state in ['Rajasthan', 'Madhya Pradesh', 'Chhattisgarh']:
            state_groups.append('Central')
      else:
            state_groups.append('Other')


# add state groups to customer data
data['state_group'] = state_groups
# select market segment
market_segment = data.loc[data['state_group'] == 'West']

market_segment.head()
```

	Names	gender	age	education	income	employment	state	age_group	income_group	gender_group	employment_group	state_group
0	Aadhya	F	25	graguate	650000	yes	Maharashtra	18-34	500K-750K	Female	employed	West
5	Abhishek	M	28	12th	450000	yes	Gujarat	18-34	250K-500K	Male	employed	West
15	Akshara	F	26	graguate	240000	yes	Gujarat	18-34	Under 250K	Female	employed	West
18	Amaira	F	33	10th	0	no	Maharashtra	18-34	Under 250K	Female	Unemployed	West
23	Amit	M	29	10th	240000	yes	Gujarat	18-34	Under 250K	Male	employed	West

```
plt.figure(figsize=(12,12))
from statsmodels.graphics.mosaicplot import mosaic
mosaic(data, ['gender_group', 'employment_group','state_group'], title='Customer Segmentation')
plt.show()
```

```
<Figure size 1200x1200 with 0 Axes>
```



6. Customizing the Marketing Mix

Marketing mix refers to the combination of product, price, promotion, and place (distribution) strategies that a company uses to meet its marketing objectives. To customize the marketing mix for the electric vehicle (EV) market segmentation, you need to consider the specific needs and preferences of different customer segments. Here are some strategies that you can use for each element of the marketing mix:

1. Product: The product strategy for EVs should focus on meeting the unique needs and preferences of different market segments. For example, some customers may be more interested in EVs that have a longer range, while others may be more interested in EVs that are more affordable. To appeal to different segments, you can offer a range of EV models with different features and price points. You can also emphasize the environmental benefits of EVs and their advanced technology to appeal to customers who value sustainability and innovation.
2. Price: Pricing strategies for EVs should also be tailored to different market segments. To appeal to budget-conscious customers, you can offer lower-priced EV models with fewer features. To appeal to more affluent customers, you can offer higher-end models with more

advanced features and luxury amenities. You can also offer incentives, such as tax credits and rebates, to encourage customers to buy EVs.

3. Promotion: Promotion strategies for EVs should focus on educating customers about the benefits of EVs and addressing common misconceptions. You can use social media, email marketing, and content marketing to reach out to different market segments and share information about the benefits of EVs. You can also partner with influencers and other organizations to increase awareness and credibility.
4. Place (Distribution): Distribution strategies for EVs should be designed to make them accessible to different market segments. You can partner with dealerships and other retailers to make EVs available in a range of locations. You can also offer online sales and home delivery options to appeal to customers who prefer to shop from home.

Overall, customizing the marketing mix for the EV market segmentation requires a deep understanding of your target customers and their needs and preferences. By tailoring your product, price, promotion, and distribution strategies to different segments, you can increase the appeal of EVs and drive adoption among a wider range of customers.

7.The MOST OPTIMAL MARKET SEGMENTS to open in the market as per your Market Research and Segmentation

India is one of the world's fastest-growing electric vehicle markets, with the government's push towards electrification of transportation and growing awareness of the benefits of electric vehicles. As electric vehicle adoption continues to grow, the need for charging infrastructure becomes increasingly important. Electric vehicle charging stations are essential for EV owners who need to recharge their vehicles on the go. In this regard, the optimal market segments for an electric vehicle start-up to enter in India would be in regions where electric vehicles are being adopted at a high rate.

According to recent market research, regions such as Delhi-NCR, Mumbai, Pune, Hyderabad, and Bangalore are experiencing a significant increase in electric vehicle adoption. For instance, Delhi-NCR and Bangalore are the leading electric vehicle markets in India, accounting for more than 60% of electric vehicle sales in 2020. Other regions such as Mumbai, Pune, and Hyderabad are also witnessing growth in electric vehicle sales.

By targeting these regions, an electric vehicle start-up can capitalize on the growing demand for electric vehicles and provide charging infrastructure to support this adoption. The optimal market segments for such a start-up would be urban commuters, ride-sharing services, and environmentally conscious consumers. These segments are more likely to purchase electric vehicles and require charging infrastructure to support their daily use.

Moreover, fleet management companies are also a viable market segment, especially in urban areas where delivery and logistics companies operate. These companies often have large fleets and can benefit from the reduced operating costs and carbon footprint of electric vehicles. In addition, government agencies and public transportation providers are also potential market segments that can benefit from electric vehicle charging infrastructure.

In summary, electric vehicle charging stations are crucial for the continued growth of electric vehicle adoption in India. By targeting regions where electric vehicles are more widely adopted and focusing on optimal market segments such as urban commuters, ride-sharing services, environmentally conscious consumers, fleet management companies, and government agencies, an electric vehicle start-up can position itself for success in the rapidly growing electric vehicle industry in India.

GitHUb Links:

Murari Muniprathap- https://github.com/muni-prathap-goud/Market-segmentation_EV
Archana S - https://github.com/ArchanaSooraj
Padala Akhil Reddy - https://github.com/lonewolff03/Feynn-Project2
Ishmeet Kaur - https://github.com/Ishmeet7/EV-Market-segmentation
Shashi Kumar Reddy - https://github.com/Shashi960/Ev-market-segment