MARKET SEGMENTATION OF ELECTRIC VEHICLE FOR STARTUP

YUKTHA T D
14 APRIL 2024

ABSTRACT

In this study, we looked at information about electric cars in India. We wanted to understand which places in India might be best for selling electric cars. We also wanted to know who might want to buy them. We found data about where people live, how much money they make, and how they use cars. This helped us figure out which cities might like electric cars the most. We also looked at how much pollution cars make and how the government helps with electric cars. Our goal is to create a plan for selling electric cars to people who want them.

1.PROBLEM STATEMENT

We're figuring out where in India people might want electric vehicles. We want to know who could be interested and where they live. We're looking at data on city living, income, and how people use transport means. This helps us see which cities might like EVs more. We're also checking pollution from vehicles and what the government does to help electric cars. Our aim is to make a plan to sell electric cars to those who want them. We're studying electric cars in India to see where they might sell best. We want to know which cities and people are interested in buying electric cars. By looking at where people live, their incomes, and how they use cars, we'll find the best places. We're also checking how much pollution cars make and how the government helps with electric cars. Our goal is to make a plan for selling electric cars to people who want them in India. We would love to know where in India people might like electric cars. We also want to understand who might want to buy them. Our goal is to make a plan for selling electric cars to those who want them.

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 relevantinformation 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 allthe 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 ofcustomers 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/

Datasets I used on the project
1.charging_station

Information based on number of charging stations in each state of India is provided here. It helped out my research since fuel is ultimate food of these electric vehicles and without electricity they are of no us. Suppose a customer is travelling a long distance and isn't able to find a charging port for the vehicle, then the EV is basically of no use. Hence, availability of charging points in every frequent check point is necessary and based on this information we can sell out more EVs in the state. Here, there is more number of charging points. And in states like Uttarakhand, Puducherry, Himachal Pradesh where the charging points are less can be changed for the better by installing more charging points.

1. combine_rating_all_vehicle

This dataset helped us with discrimination vehicles' good and bad features using rating system. All the customers who got to use the vehicle rated its its different features. So a lot of vehicles rating wasn't as expected. Two wheelers came out with flying colours where as the four wheelers' rating went down the drain. Although few four wheelers performed very well in some aspects while few of them were still not negligible enough to be discarded. Three wheelers performance was moderate and did very well on capacity holding and speed,.four wheelers did well on safety and two wheelers on speed and agility.

3. CODE IMPLEMENTATION:

- **Downloading all the libraries needed to analyze the datasets**

```
In [126...
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

- **Compiling the dataset and importing it to the notebook**

```
In [127… EV_user_rating = pd.read_csv("combine_rating_all_vehicle.csv")
```

```
In [128… EV_user_rating
```

Out[128…

	Rating	Model Name	Type
0	1.0	TVS iQube	2-wheeler
1	1.0	TVS iQube	2-wheeler
2	3.0	TVS iQube	2-wheeler
3	1.0	TVS iQube	2-wheeler
4	1.0	TVS iQube	2-wheeler
...	...	...	...
1108	4.0	tata tigor ev	4-wheeler
1109	5.0	tata tigor ev	4-wheeler
1110	5.0	tata tigor ev	4-wheeler
1111	5.0	tata tigor ev	4-wheeler
1112	5.0	tata tigor ev	4-wheeler

1113 rows × 3 columns

- Calculating number of ratings received for different types of vehicles

```
In [131… sns.countplot(x = 'Rating',data = EV_user_rating,color = 'blue')
```
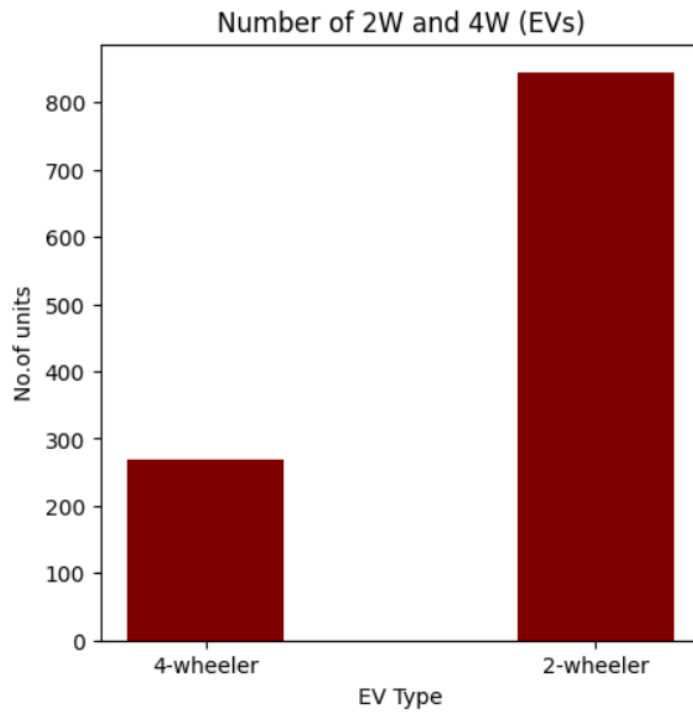
```
Out[131… <Axes: xlabel='Rating', ylabel='count'>
```
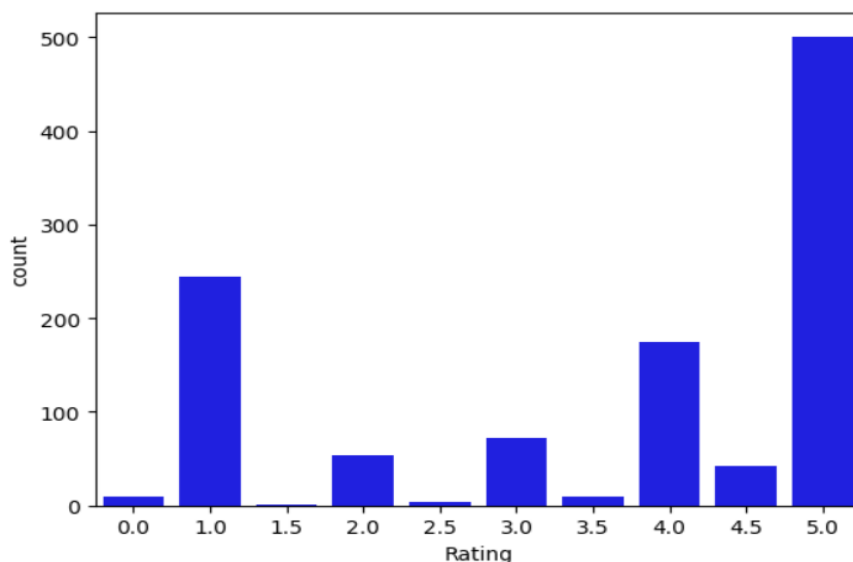
- Figuring out which type of EV is the best

```python
fig = plt.figure(figsize = (5, 5))
plt.bar(labels, units, color ='maroon', width = 0.4)
plt.xlabel("EV Type")
plt.ylabel("No.of units")
plt.title("Number of 2W and 4W (EVs)")
plt.show()
```

- Compiling the second dataset and importing it

In [135…
```python
charging_stations_India = pd.read_csv("charging_station.csv")
charging_stations_India.head(15)
```

Out[135…

	State/ UTs	No. of charging stations
0	Maharashtra	317
1	Andhra Pradesh	266
2	Tamil Nadu	281
3	Gujarat	278
4	Uttar Pradesh	207
5	Rajasthan	205
6	Karnataka	172
7	Madhya Pradesh	235
8	West Bengal	141
9	Telangana	138
10	Kerala	211
11	Delhi	72
12	Chandigarh	70
13	Haryana	50
14	Meghalaya	40

- Analyzing the data and collecting the number of the charging points

In [136…   `charging_stations_India.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 2 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   State/ UTs               25 non-null     object
 1   No. of charging stations 25 non-null     int64
dtypes: int64(1), object(1)
memory usage: 532.0+ bytes
```

In [137…   `charging_stations_India.describe()`
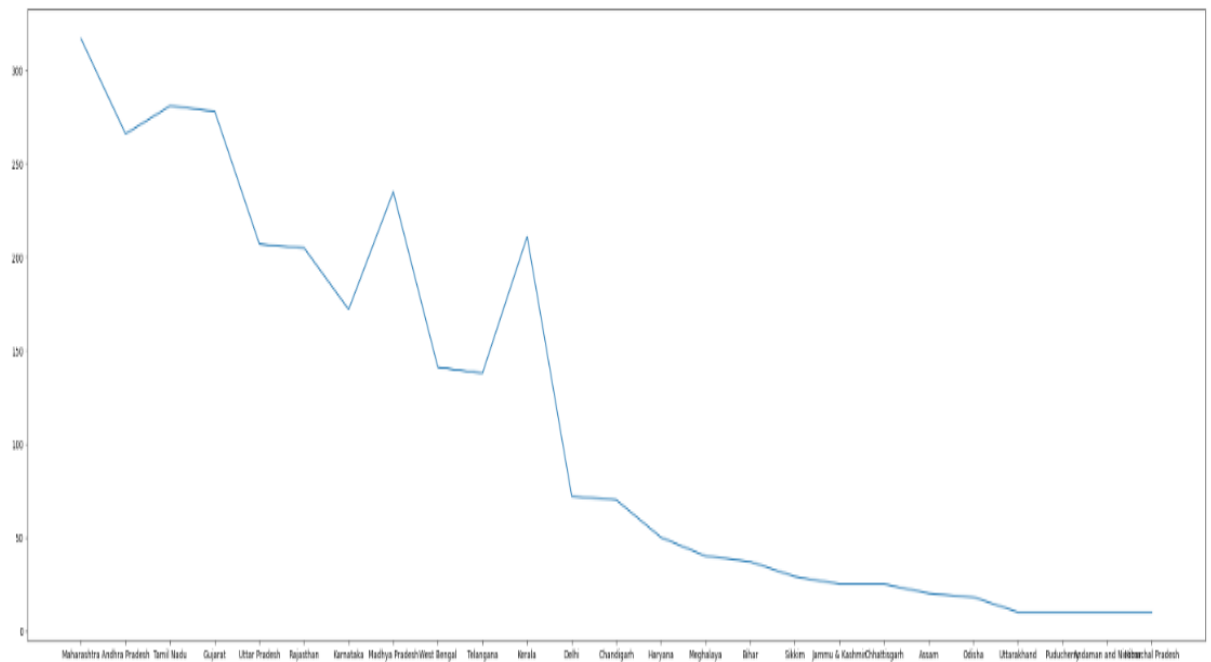
Out[137…

	No. of charging stations
count	25.000000
mean	115.080000
std	105.731941
min	10.000000
25%	25.000000
50%	70.000000
75%	207.000000
max	317.000000

- Number of stations vs all the states

```
x = charging_stations_India['State/ UTs']
y = charging_stations_India['No. of charging stations']
tickvalues = range(0,len(x))
plt.figure(figsize = (35,10))
plt.plot(x,y)
```

[<matplotlib.lines.Line2D at 0x21645f56e10>]

- Compiling the third dataset and importing the packages along with dataset

```
In [139_   !pip install plotly
```

```
Requirement already satisfied: plotly in c:\users\yuktha t d\appdata\local\programs\python\python312\lib\site-packages (5.20.
0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\yuktha t d\appdata\local\programs\python\python312\lib\site-packag
es (from plotly) (8.2.3)
Requirement already satisfied: packaging in c:\users\yuktha t d\appdata\local\programs\python\python312\lib\site-packages (fr
om plotly) (24.0)
```

```
In [140_   EVs_India = EV_user_rating['Model Name'].value_counts()
           print(EVs_India)
```

```
Model Name
Hero Electric Flash      102
Okinawa Praise            95
Hero Electric Optima      82
tata nexon ev             75
Tata Nexon EV             74
hyundai kona              56
Revolt RV 400             52
PURE EV EPluto 7G         50
Tork Kratos               38
Hero Electric Photon      38
Ather 450X                30
OLA S1                    29
Ampere Magnus EX          28
OLA S1 Pro                28
Ampere REO                24
Benling Aura              22
Ampere Magnus Pro         22
tata tigor ev             20
PURE EV ETrance Neo       19
TVS iQube                 17
Bajaj Chetak              15
Okinawa Ridge Plus        15
Tata Tigor EV             15
Ampere Zeal               13
MG ZS EV                  12
Kia EV6                   12
Bounce Infinity E1        12
Okinawa i-Praise          11
Revolt RV 300             10
Techo Electra Emerge      10
Hero Electric Optima CX    9
Techo Electra Raptor       9
Hero Electric NYX          9
Okinawa R30                8
Okinawa Lite               8
Gemopai Ryder              5
BYD E6                     5
Joy e-bike Wolf            5
e-bike Gen Nxt             5
Yo Drift                   5
BGauss B8                  4
Gemopai Astrid Lite        3
Hero Electric Atria        3
Odysse Evoqis              3
Hero Electric NYX HX       2
Evolet Polo                2
Joy e-bike Monster         2
Name: count, dtype: int64
```

- Data types

```
In [141_   import plotly as px
```

```
In [142_   Indian_purchase_behaviour = pd.read_csv("behavioural_dataset.csv")
           Indian_purchase_behaviour.rename(columns={'Personal loan':'Car_Loan'},inplace=True)
           Indian_purchase_behaviour.rename(columns={'Price':'EV_Price'},inplace=True)
           Indian_purchase_behaviour.head()
```

Out[142_

|	Age	Profession	Marrital Status	Education	No of Dependents	Car_Loan	Total Salary	EV_Price
0	27	Salaried	Single	Post Graduate	0	Yes	800000	800000
1	35	Salaried	Married	Post Graduate	2	Yes	2000000	1000000
2	45	Business	Married	Graduate	4	Yes	1800000	1200000
3	41	Business	Married	Post Graduate	3	No	2200000	1200000
4	31	Salaried	Married	Post Graduate	2	Yes	2600000	1600000

- Differentiation of data types

```
In [143_   labels = ['Car Loan Required','Car Loan not required']
           Loan_status = [Indian_purchase_behaviour.query('Car_Loan == "Yes"').Car_Loan.count(),Indian_purchase_behaviour.query('Car_
           print(Loan_status)

           import matplotlib.pyplot as plt

           explode = (0, 0.1)
           fig1, ax1 = plt.subplots()
           ax1.pie(Loan_status, explode=explode, labels=labels, autopct='%1.1f%%',shadow=True, startangle=90)
           ax1.axis('equal')

           plt.show()
```
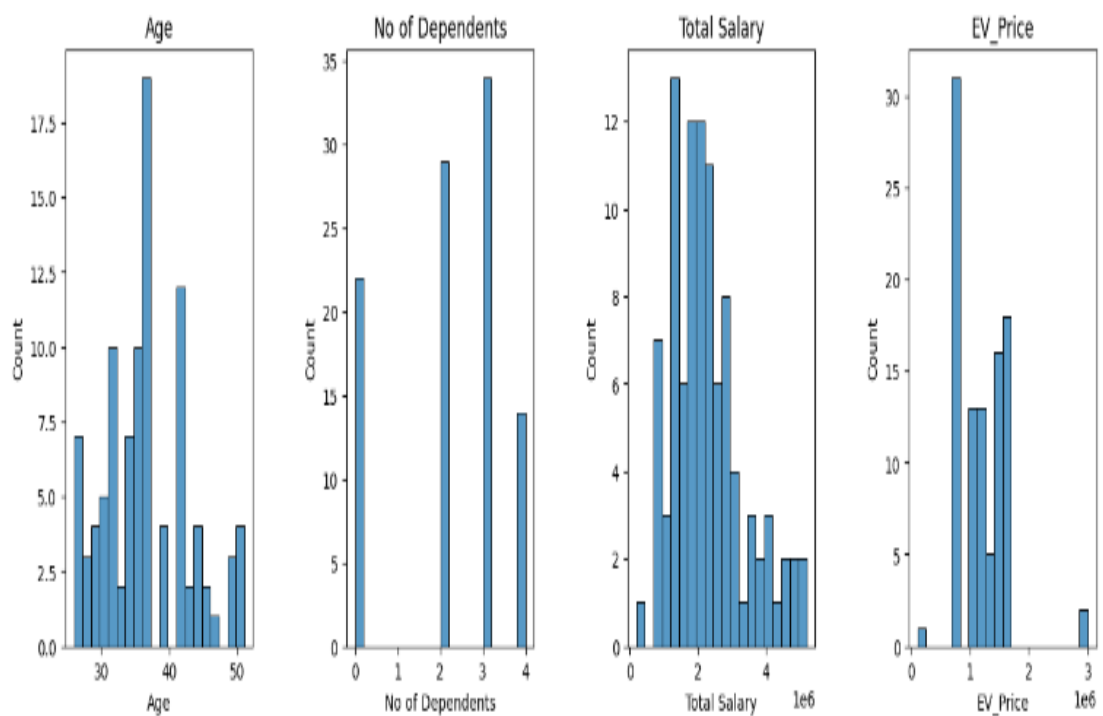
```
[32, 67]
```
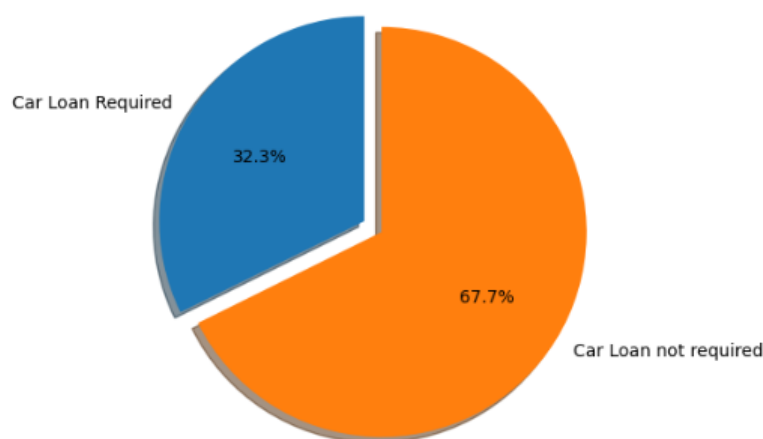
- Further more data types

```python
plt.figure(1, figsize=(15,5))
n=0
group_labels = ['Group 1', 'Group 2', 'Group 3','Group 4']
for x in ['Age', 'No of Dependents' ,'Total Salary'  ,'EV_Price']:
  n += 1
  plt.subplot(1,4,n)
  plt.subplots_adjust(hspace=0.5, wspace=0.5)
  sns.histplot(Indian_purchase_behaviour[x], bins= 20)
  plt.title('{}'.format(x))
plt.show()
```

```python
features = list(Indian_purchase_behaviour.columns)
features
```

```
['Age',
 'Profession',
 'Marrital Status',
 'Education',
 'No of Dependents',
 'Car_Loan',
 'Total Salary',
 'EV_Price']
```

- Forming a cluster

```
In [146…  cluster_data = Indian_purchase_behaviour[features]
          cluster_data.head()
```

	Age	Profession	Marrital Status	Education	No of Dependents	Car_Loan	Total Salary	EV_Price
0	27	Salaried	Single	Post Graduate	0	Yes	800000	800000
1	35	Salaried	Married	Post Graduate	2	Yes	2000000	1000000
2	45	Business	Married	Graduate	4	Yes	1800000	1200000
3	41	Business	Married	Post Graduate	3	No	2200000	1200000
4	31	Salaried	Married	Post Graduate	2	Yes	2600000	1600000

- Cluster data types

```
In [150…   cluster_data.dtypes
```

```
Out[150…   Age                 int64
           Profession          object
           Marrital Status     object
           Education           object
           No of Dependents    int64
           Car_Loan            object
           Total Salary        int64
           EV_Price            int64
           dtype: object
```
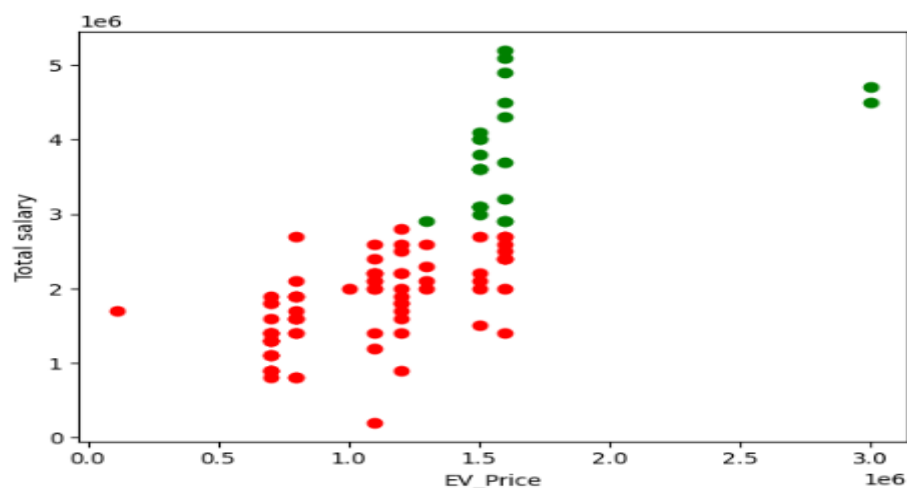
- Total salaries of target customers vs EV price

```
In [153…   Cluster_0 = cluster_data[cluster_data.clusters==0]
           Cluster_1 = cluster_data[cluster_data.clusters==1]

           plt.scatter(Cluster_0.EV_Price, Cluster_0['Total Salary'],color='red')
           plt.scatter(Cluster_1.EV_Price, Cluster_1['Total Salary'],color='green')

           plt.xlabel('EV_Price')
           plt.ylabel('Total salary')
           plt.show()
```

- Range of prices for target customers
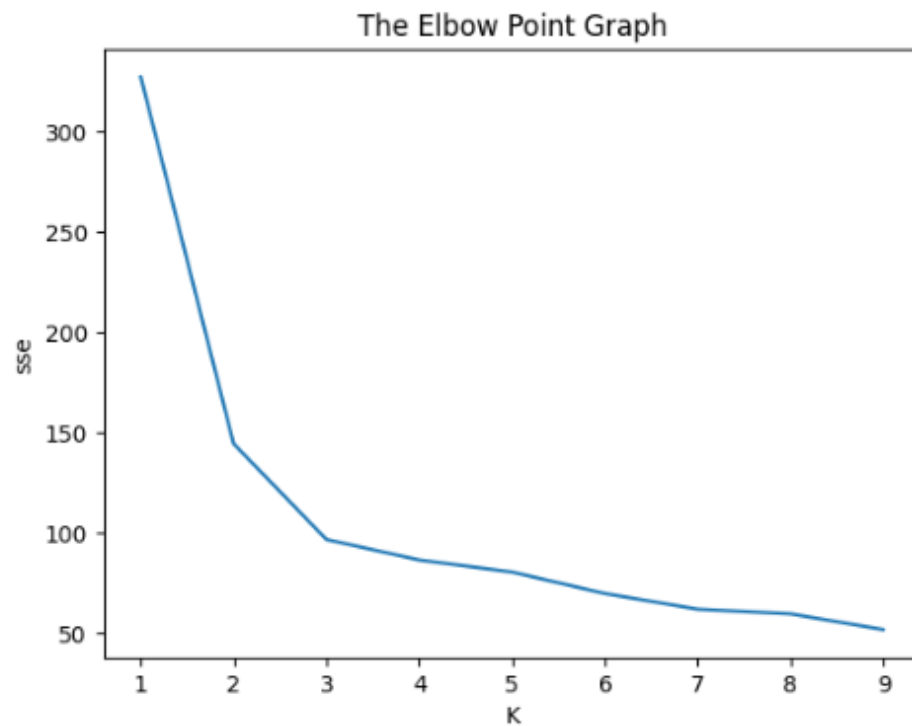
- 3D/spatial graph for better understanding

```
fig = plt.figure(figsize=(20,10))

ax = fig.add_subplot(111, projection='3d')
ax.scatter(cluster_data.Age[labels == 0], cluster_data["Total Salary"][labels == 0], cluster_data["EV_Price"][labels == 0],
ax.scatter(cluster_data.Age[labels == 1], cluster_data["Total Salary"][labels == 1], cluster_data["EV_Price"][labels == 1],
ax.scatter(cluster_data.Age[labels == 2], cluster_data["Total Salary"][labels == 2], cluster_data["EV_Price"][labels == 2],
ax.view_init(30, 185)

plt.xlabel("Age")
plt.ylabel("Total Salary")
ax.set_zlabel('EV_Price')
plt.show()
```
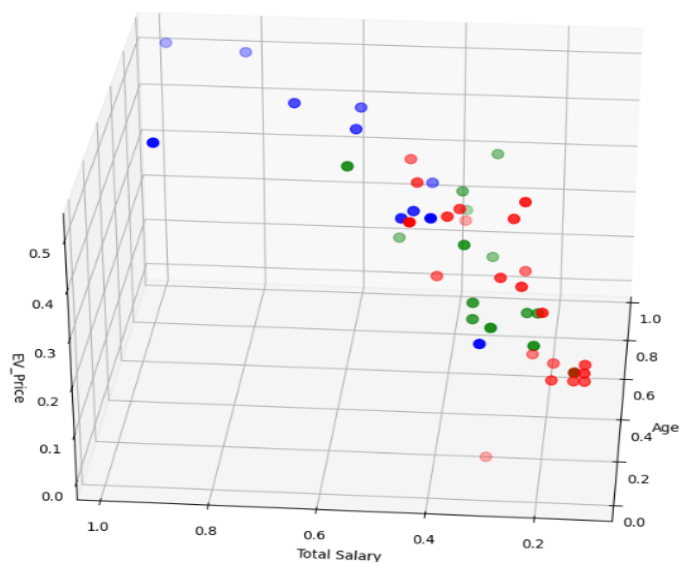
CONCLUSION

As the research came to an end, we were able to identify our target audience which is people with moderate salaries, post-graduate, married who also supposedly live in areas of high number charging points available, while the rating fluctuated between two wheelers and four wheelers, two wheelers were still high in demand due to its high movability and agility. This report has helped us to find people who would be potential buyer for our new EVs which would be produced later. Most of the information was taken from legit sites making it most authentic datasets. At the end, this will have huge part in start up's success and not in its downfall.