Market Segmentation Analysis of Electric Vehicles Market in India

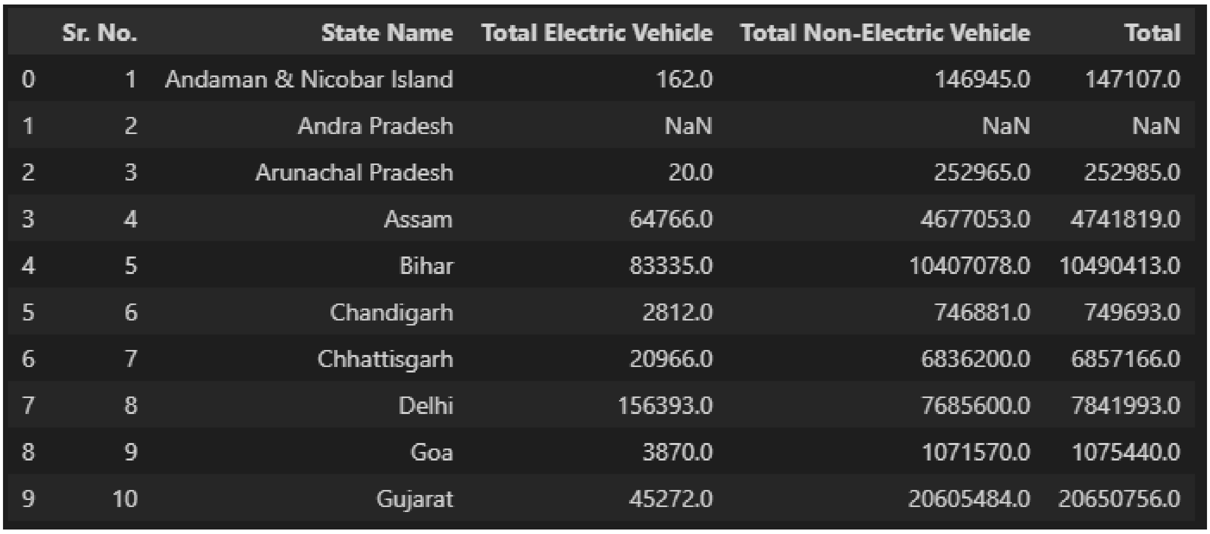
Vincent Isaac Jeyaraj J

Problem Statement:

Git-hub link: vincent-isaac/Market-Segmentation-Analysis-of-Electric-Vehicles-Market-in-India (github.com)

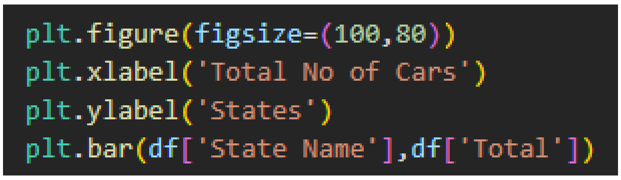
The task is to analyze the Electric Vehicle market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles.

Dataset:

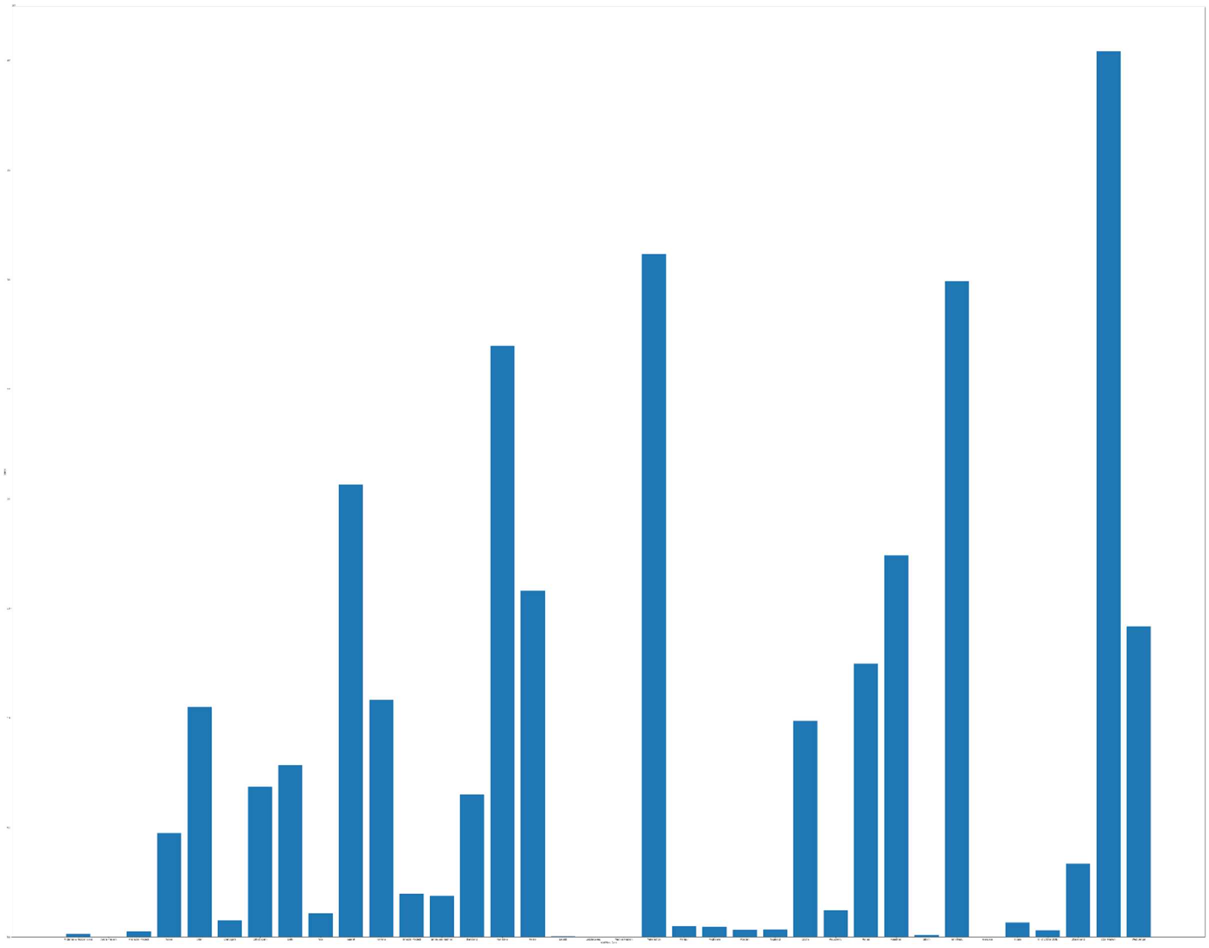


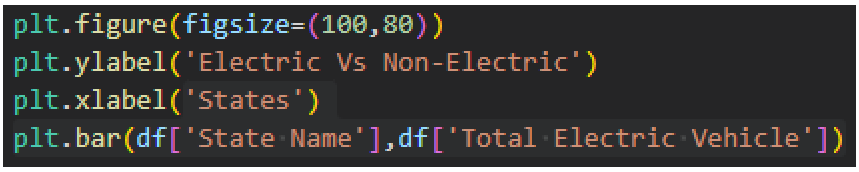
	Sr. No.	State Name	Total Electric Vehicle	Total Non-Electric Vehicle	Total
0	1	Andaman & Nicobar Island	162.0	146945.0	147107.0
1	2	Andra Pradesh	NaN	NaN	NaN
2	3	Arunachal Pradesh	20.0	252965.0	252985.0
3	4	Assam	64766.0	4677053.0	4741819.0
4	5	Bihar	83335.0	10407078.0	10490413.0
5	6	Chandigarh	2812.0	746881.0	749693.0
6	7	Chhattisgarh	20966.0	6836200.0	6857166.0
7	8	Delhi	156393.0	7685600.0	7841993.0
8	9	Goa	3870.0	1071570.0	1075440.0
9	10	Gujarat	45272.0	20605484.0	20650756.0

Visualization:

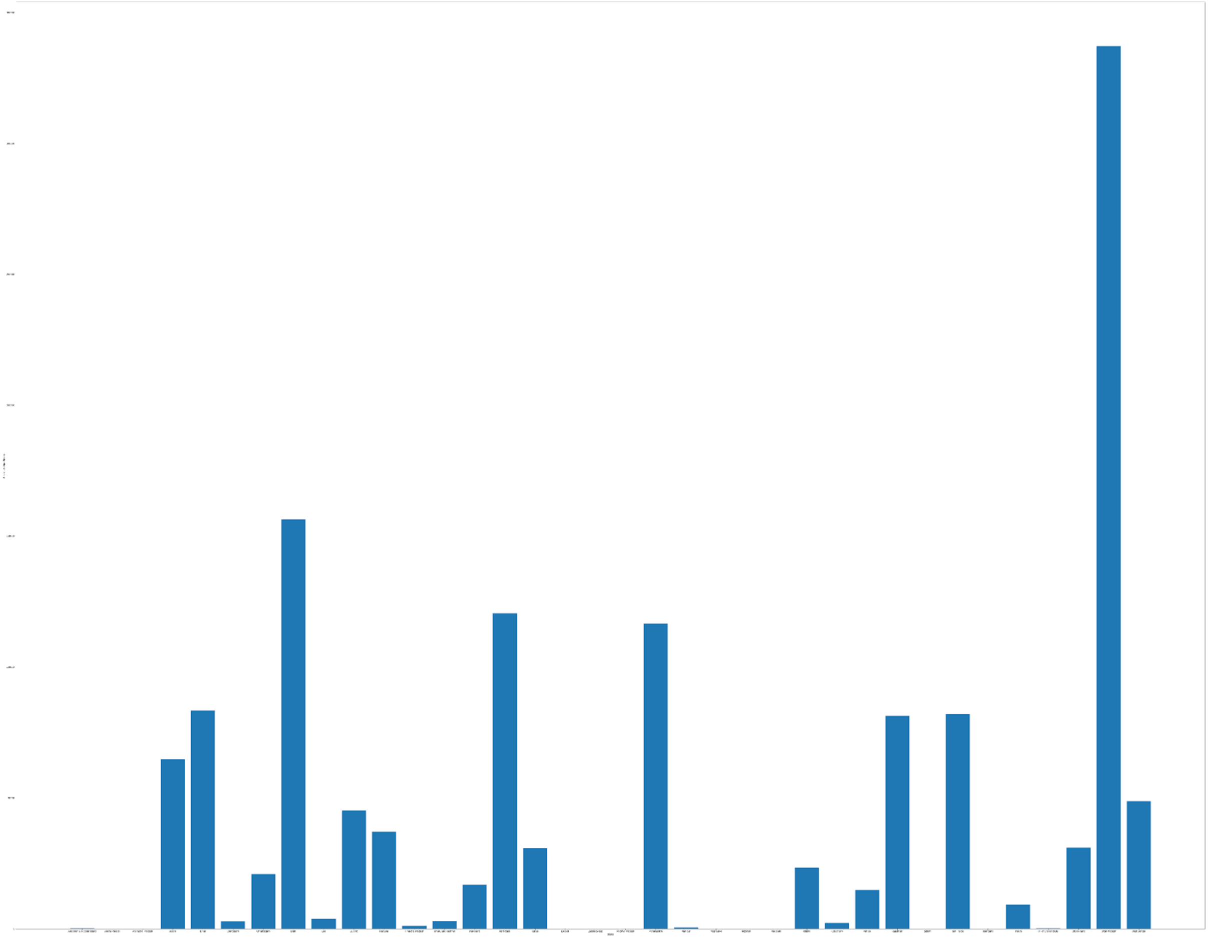


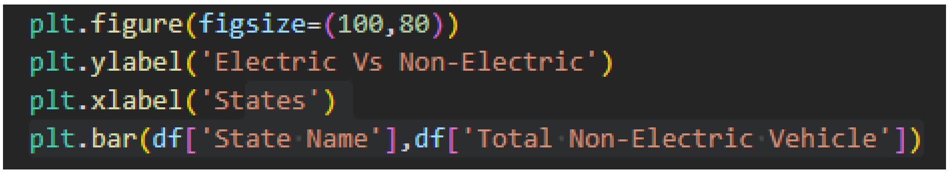
```
plt.figure(figsize=(100,80))
plt.xlabel('Total No of Cars')
plt.ylabel('States')
plt.bar(df['State Name'],df['Total'])
```



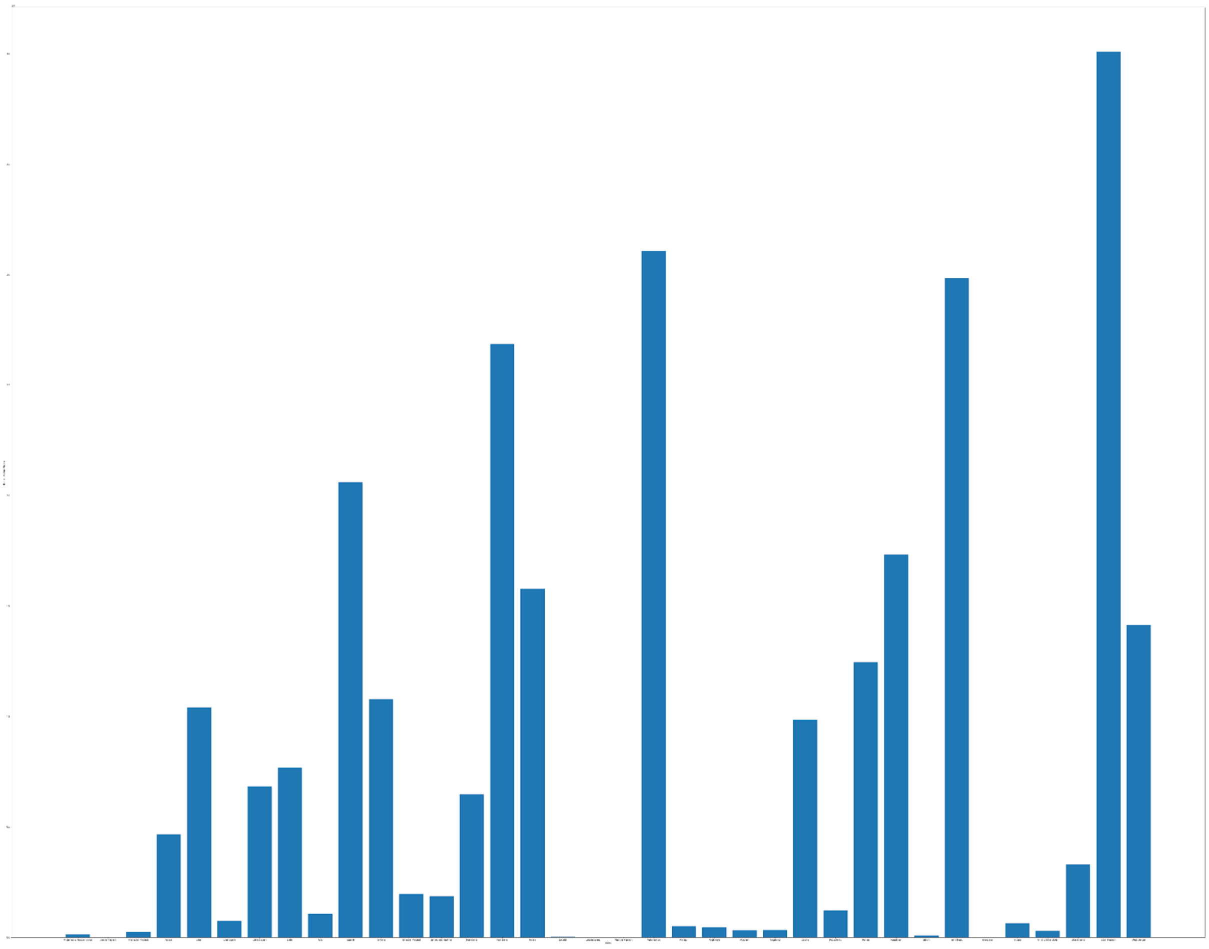


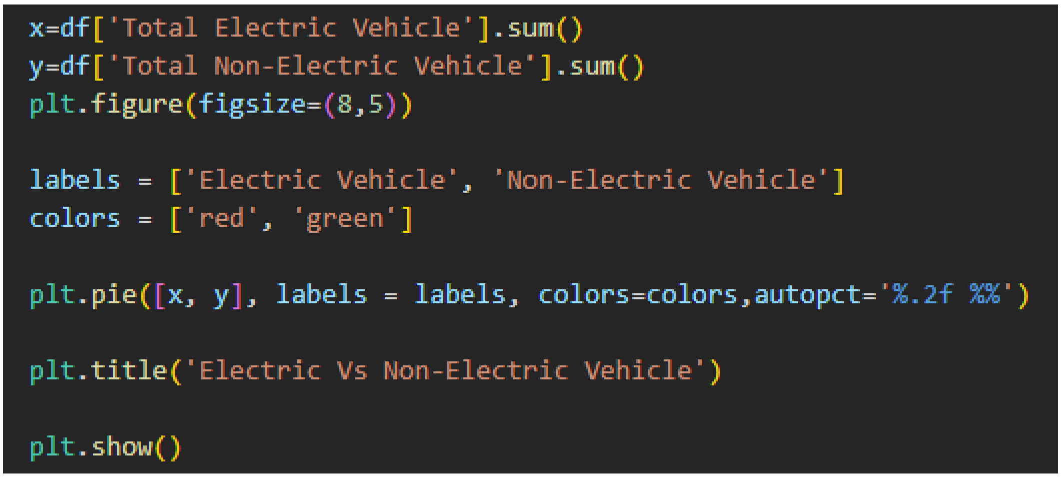
```
plt.figure(figsize=(100,80))
plt.ylabel('Electric Vs Non-Electric')
plt.xlabel('States')
plt.bar(df['State Name'],df['Total Electric Vehicle'])
```





```
plt.figure(figsize=(100,80))
plt.ylabel('Electric Vs Non-Electric')
plt.xlabel('States')
plt.bar(df['State Name'],df['Total Non-Electric Vehicle'])
```





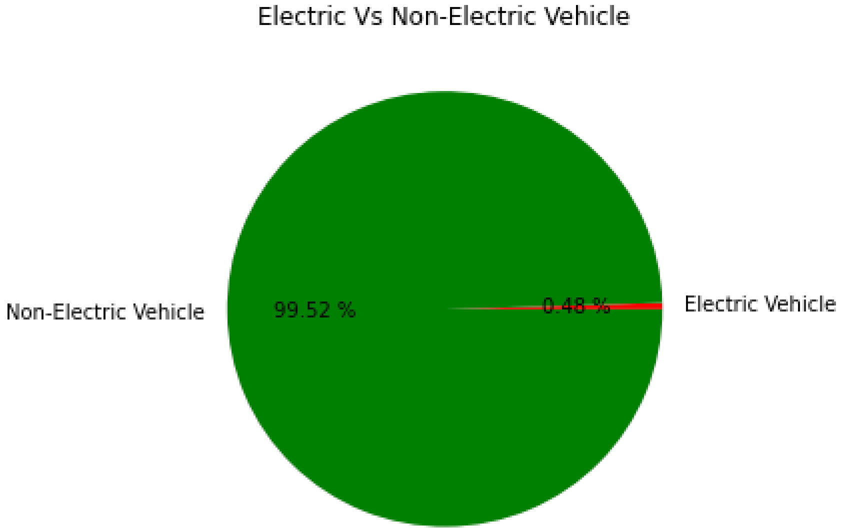
```
x=df['Total Electric Vehicle'].sum()
y=df['Total Non-Electric Vehicle'].sum()
plt.figure(figsize=(8,5))

labels = ['Electric Vehicle', 'Non-Electric Vehicle']
colors = ['red', 'green']

plt.pie([x, y], labels = labels, colors=colors,autopct='%.2f %%')

plt.title('Electric Vs Non-Electric Vehicle')

plt.show()
```



ML Algorithm used:

K-Means Clustering:

K-means clustering is a popular algorithm for partitioning data points into clusters based on similarity. The algorithm aims to minimize the within-cluster variance by iteratively assigning data points to clusters and updating the cluster centroids.

In K-means clustering, the following steps are typically followed:

1. Specify the number of clusters (K) to create.
2. Initialize K centroids randomly or based on some heuristics.
3. Assign each data point to the nearest centroid, forming K clusters.
4. Recalculate the centroids by taking the mean of the data points within each cluster.
5. Repeat steps 3 and 4 until convergence (when the centroids no longer change significantly) or until a maximum number of iterations is reached.

K-means clustering is an iterative and computationally efficient algorithm, but it is sensitive to the initial random centroids and can converge to suboptimal solutions. Techniques like multiple random initializations or K-means++ initialization can be employed to mitigate this.

The resulting clusters can be analyzed to gain insights into patterns, group similar data points together, or assist in data exploration and visualization. K-means clustering is widely used in various fields such as machine learning, data mining, image processing, and customer segmentation, among others.

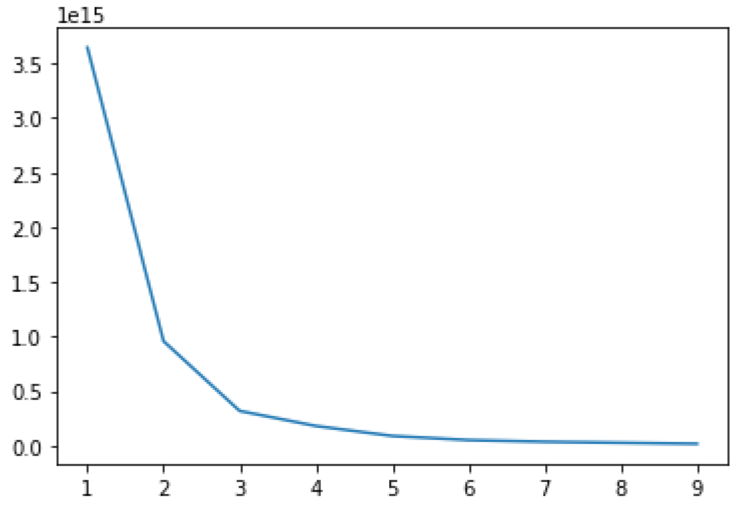
Let us assume we have N unlabeled multivariate datasets with different attributes from our dataset, such as water availability, price, city, etc. Clustering is a method for categorizing datasets into different groups based on shared traits and characteristics. Clusters are the entities that are forming these groups.

Unsupervised learning algorithms in machine learning use clustering because they may divide multivariate data into different groups without the need for a supervisor based on a common pattern concealed within the datasets.

The Elbow Method:

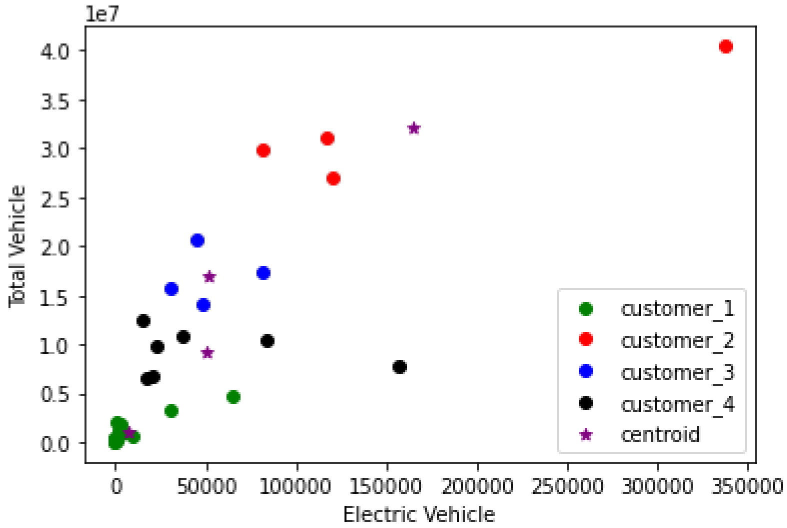
In the Elbow approach, the number of clusters (K) is truly variable and ranges from 1 to 10.
We are calculating WCSS (Within-Cluster Sum of Square) for each value of K.
The sum of the squared distances between each point and the cluster's centroid is known as WCSS. The plot of the WCSS with the K value resembles an elbow.

Below is the graph obtained for the electric vehicles in India dataset:



From the WCSS graph, we are getting 4 different clusters. Hence, our K value is 4.

We construct a K-Means clustering graph or a multivariate cluster graph based on our segmentation analysis.



In the above graph, we can see that there are 4 different segments. Customers in India according to states are divided into segments called customer_1, customer_2, customer_3, and customer_4.

Segment 1 (customer_1) has a very small number of electrical vehicles compared to the total number of vehicles present in their state, which states that customers mostly don’t prefer electric cars and it may not be suitable with their terrain or location.

Segment 2 (customer_2) has the most number of electric cars present in their respective states, stating that EVs are trending, and suitable with their place which means these states have appropriate facilities to accommodate electric cars and are affordable for this segment of customers.

Segment 3 (customer_3) has the second most electric cars in their states which means these states are developing fast and customers have started to change and prefer electric cars. This segment can soon become a target segment since its market is easy to capture and it has a faster development rate.

Segment 4 (customer_4) has the second least number of electric cars in their state and customers from this segment are still doubt full to changing to electric cars and their state might not have appropriate facilities to accommodate the change.

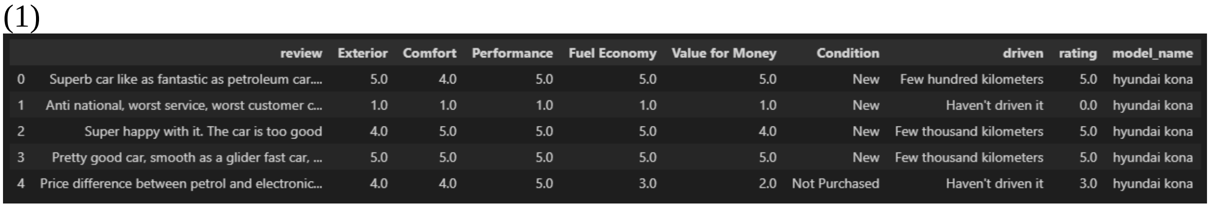
To find more insight into why customers are segmented and the reason behind it can be extracted from other datasets.

Dataset:

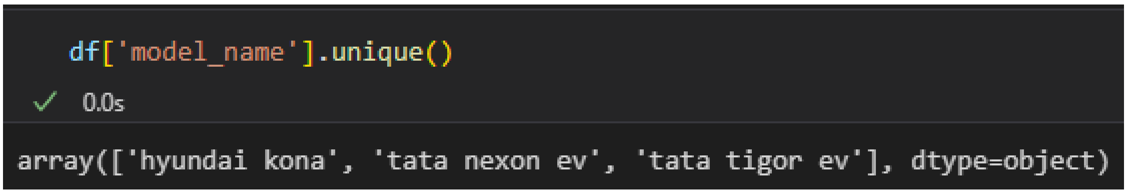
Dataset (1) and (2) contains a review of the electric car in India, and dataset (3) has the details of every purchasable electric car in India.

These datasets let us observe some key details to say which type of cars customers prefer and don’t.

(1)



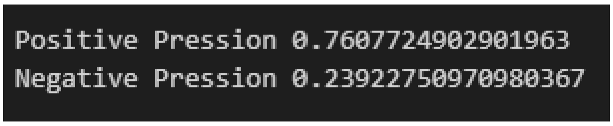
	review	Exterior	Comfort	Performance	Fuel Economy	Value for Money	Condition	driven	rating	model_name
0	Superb car like as fantastic as petroleum car....	5.0	4.0	5.0	5.0	5.0	New	Few hundred kilometers	5.0	hyundai kona
1	Anti national, worst service, worst customer c...	1.0	1.0	1.0	1.0	1.0	New	Haven't driven it	0.0	hyundai kona
2	Super happy with it. The car is too good	4.0	5.0	5.0	5.0	4.0	New	Few thousand kilometers	5.0	hyundai kona
3	Pretty good car, smooth as a glider fast car, ...	5.0	5.0	5.0	5.0	5.0	New	Few thousand kilometers	5.0	hyundai kona
4	Price difference between petrol and electronic...	4.0	4.0	5.0	3.0	2.0	Not Purchased	Haven't driven it	3.0	hyundai kona



```
df['model_name'].unique()
```

✓ 0.0s

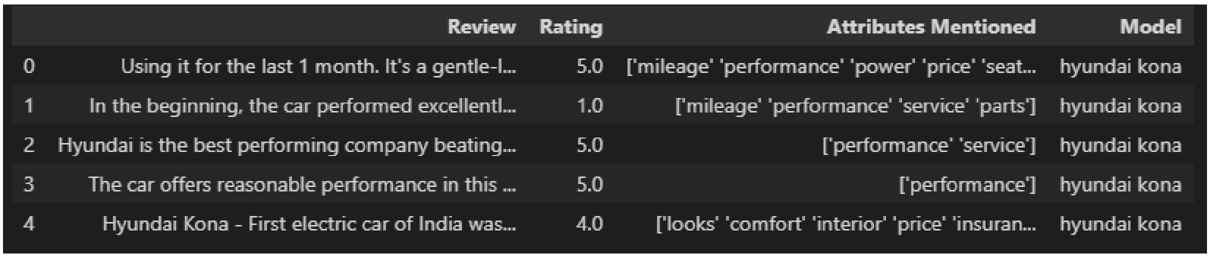
```
array(['hyundai kona', 'tata nexon ev', 'tata tigor ev'], dtype=object)
```



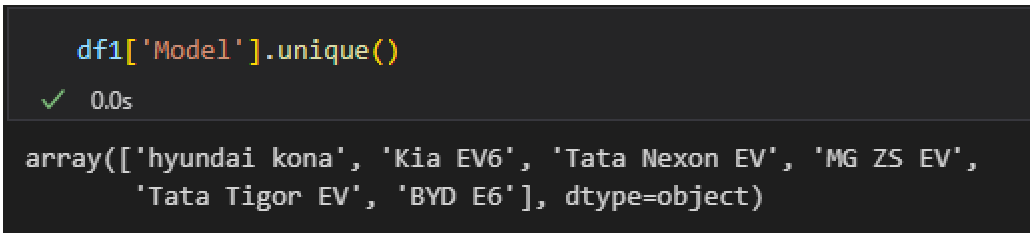
```
Positive Pression 0.7607724902901963
Negative Pression 0.23922750970980367
```

In dataset (1) we have the user reviews for the cars 'hyundai kona', 'tata nexon ev', and 'tata tiger ev'.

(2)



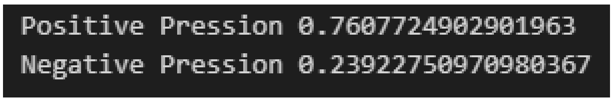
	Review	Rating	Attributes Mentioned	Model
0	Using it for the last 1 month. It's a gentle-l...	5.0	['mileage' 'performance' 'power' 'price' 'seat...	hyundai kona
1	In the beginning, the car performed excellentl...	1.0	['mileage' 'performance' 'service' 'parts']	hyundai kona
2	Hyundai is the best performing company beating...	5.0	['performance' 'service']	hyundai kona
3	The car offers reasonable performance in this ...	5.0	['performance']	hyundai kona
4	Hyundai Kona - First electric car of India was...	4.0	['looks' 'comfort' 'interior' 'price' 'insuran...	hyundai kona



```
df1['Model'].unique()
```

✓ 0.0s

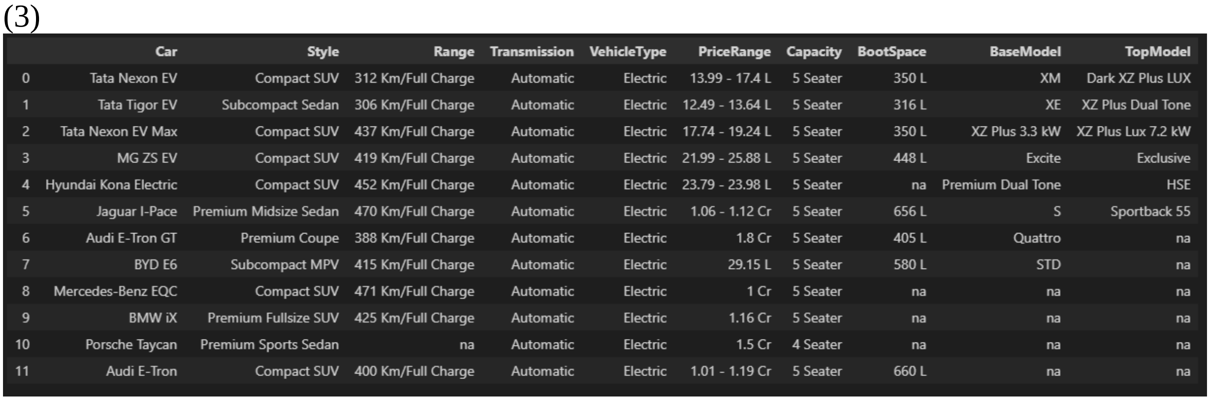
```
array(['hyundai kona', 'Kia EV6', 'Tata Nexon EV', 'MG ZS EV',
       'Tata Tigor EV', 'BYD E6'], dtype=object)
```



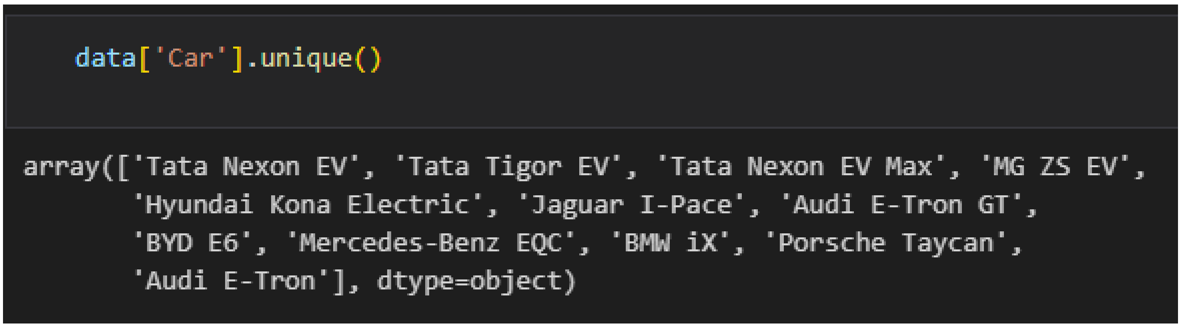
```
Positive Pression 0.7607724902901963
Negative Pression 0.23922750970980367
```

In dataset (2) we have the user reviews for the cars 'hyundai kona', 'Kia EV6', 'Tata Nexon EV', 'MG ZS EV', 'Tata Tigor EV', 'And BYD E6'.

(3)



	Car	Style	Range	Transmission	VehicleType	PriceRange	Capacity	BootSpace	BaseModel	TopModel
0	Tata Nexon EV	Compact SUV	312 Km/Full Charge	Automatic	Electric	13.99 - 17.4 L	5 Seater	350 L	XM	Dark XZ Plus LUX
1	Tata Tigor EV	Subcompact Sedan	306 Km/Full Charge	Automatic	Electric	12.49 - 13.64 L	5 Seater	316 L	XE	XZ Plus Dual Tone
2	Tata Nexon EV Max	Compact SUV	437 Km/Full Charge	Automatic	Electric	17.74 - 19.24 L	5 Seater	350 L	XZ Plus 3.3 kW	XZ Plus Lux 7.2 kW
3	MG ZS EV	Compact SUV	419 Km/Full Charge	Automatic	Electric	21.99 - 25.88 L	5 Seater	448 L	Excite	Exclusive
4	Hyundai Kona Electric	Compact SUV	452 Km/Full Charge	Automatic	Electric	23.79 - 23.98 L	5 Seater	na	Premium Dual Tone	HSE
5	Jaguar I-Pace	Premium Midsize Sedan	470 Km/Full Charge	Automatic	Electric	1.06 - 1.12 Cr	5 Seater	656 L	S	Sportback 55
6	Audi E-Tron GT	Premium Coupe	388 Km/Full Charge	Automatic	Electric	1.8 Cr	5 Seater	405 L	Quattro	na
7	BYD E6	Subcompact MPV	415 Km/Full Charge	Automatic	Electric	29.15 L	5 Seater	580 L	STD	na
8	Mercedes-Benz EQC	Compact SUV	471 Km/Full Charge	Automatic	Electric	1 Cr	5 Seater	na	na	na
9	BMW iX	Premium Fullsize SUV	425 Km/Full Charge	Automatic	Electric	1.16 Cr	5 Seater	na	na	na
10	Porsche Taycan	Premium Sports Sedan	na	Automatic	Electric	1.5 Cr	4 Seater	na	na	na
11	Audi E-Tron	Compact SUV	400 Km/Full Charge	Automatic	Electric	1.01 - 1.19 Cr	5 Seater	660 L	na	na



```
data['Car'].unique()
```

```
array(['Tata Nexon EV', 'Tata Tigor EV', 'Tata Nexon EV Max', 'MG ZS EV',
       'Hyundai Kona Electric', 'Jaguar I-Pace', 'Audi E-Tron GT',
       'BYD E6', 'Mercedes-Benz EQC', 'BMW iX', 'Porsche Taycan',
       'Audi E-Tron'], dtype=object)
```

In dataset (3) we have a total of 12 electric cars that are available in India for purchase. The cars include 'Tata Nexon EV', 'Tata Tigor EV', 'Tata Nexon EV Max', 'MG ZS EV', 'Hyundai Kona Electric', 'Jaguar I-Pace', 'Audi E-Tron GT', 'BYD E6', 'Mercedes-Benz EQC', 'BMW iX', 'Porsche Taycan', 'Audi E-Tron'.

In conclusion, dataset (3) contains 12 electric cars but dataset (1), (2) contains only 3, and 6 electric car data in it. We can either assume there is no available data for other cars or simply customers didn't prefer those cars.

Common electric cars from datasets (1), and (2) that are present in dataset (3) can be considered as the electric cars that customers use (According to the datasets). So, Electric car models 'Tata Nexon EV', 'Tata Tigor EV', 'Tata Nexon EV Max', 'MG ZS EV', 'Hyundai Kona Electric', 'BYD E6' are available in (3) from (1),(2).

To find the reason why customers do only prefer these cars, let's compare other details of these cars from the dataset (3). Let us take the major parameter that concerns customers buying an electric car, here if we do observe the cars that are common to dataset (3) compared to dataset (1), (2) the price of these cars only ranges a minimum of 12 Lakhs to a maximum of 30 Lakhs, and the other available electric cars cost more than 1Cr.

Therefore, we can conclude by saying the customer prefers to buy an electric car that has a price range of 12-30Lakhs.