MARKET SEGMENTATION ANALYSIS OF ELECTRIC VEHICLES MARKET IN INDIA

Sharath Pai

1 Problem Statement

Our goal is to analyze market segment for an Electric Vehicle based startup to decide which vehicle/customer space is suitable for developing its EVs..

In this report we analyze the Electric Vehicles Market in India using segments such as region, price, charging facility, type of vehicles (e.g., 2 wheelers, 3 wheelers, 4 wheelers etc.), body type (e.g., Hatchback, Sedan, SUV, Autorickshaw etc.), safety, plug types and much more.

2 Fermi Estimation

Around 4.52 million electric vehicles are estimated to be sold in India in the year 2024 out of which 75% of the market will be occupied by public transport. gives us a rough idea of how many people in India could use EVs by 2024. The Indian EV market in 2024 could reach between Rs. 63,700 crores and Rs. 161,500 crores, depending on the pace of EV adoption. Moderate estimates suggest a 3% adoption rate and a slight price decrease, leading to a market size of Rs. 102,300 crores. However, optimistic scenarios with 5% adoption and continued government support could push the market towards Rs. 161,500 crores. These are the rough estimations which gives us a rough idea of how many people in India could use EVs by 2024.

3 Data Collection

Data was collected from one of the most popular data science platforms which is Kaggle and public datasets to extract meaningful insights. We'll be using these datasets for our market segmentation analysis.

https://www.kaggle.com/datasets/saketpradhan/electric-vehicle-charging-stations-in-india

https://www.india-briefing.com/news/indias-ev-manufacturing-capacity-and-market-preferences-progress-25840.html/

4 Data Preprocessing

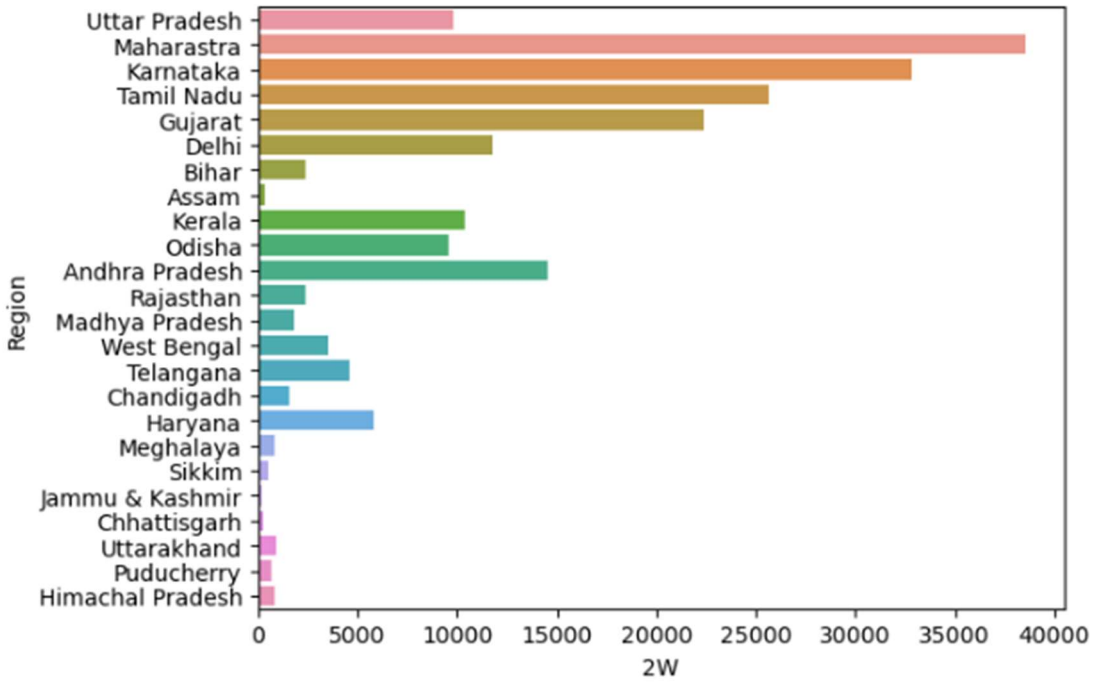
Using Python libraries such as NumPy and Pandas, the data was loaded into the notebook and preprocessed. The preprocessing involved encoding of categorical variables, extraction of

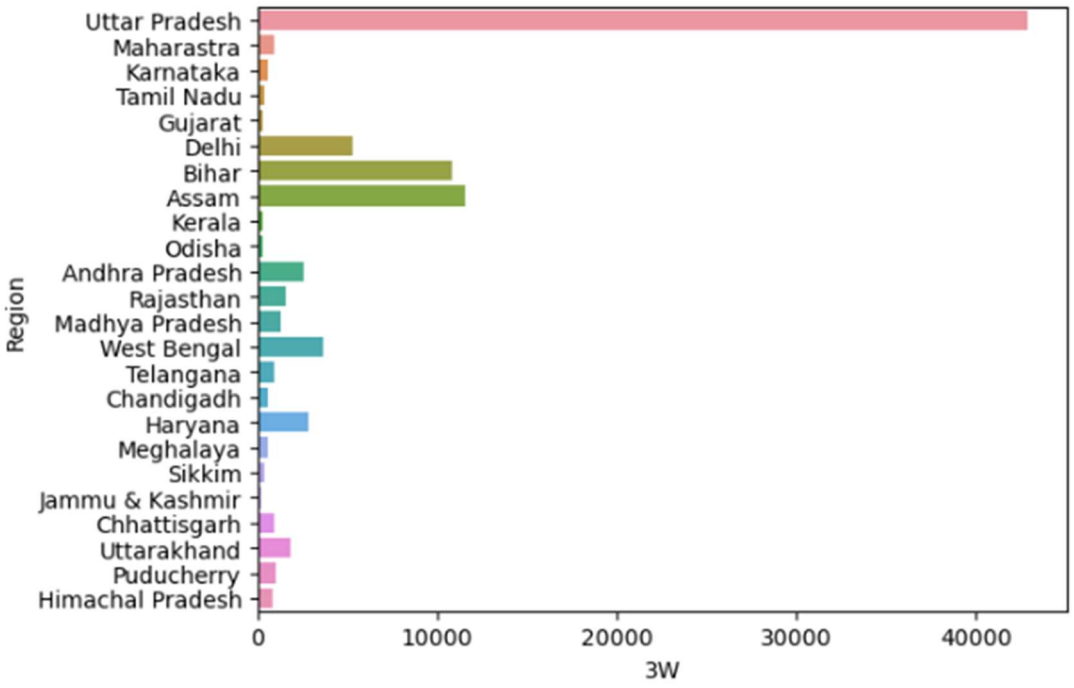
important features, deletion of unnecessary variables. This part ensured that our datasets are for further analysis of features.

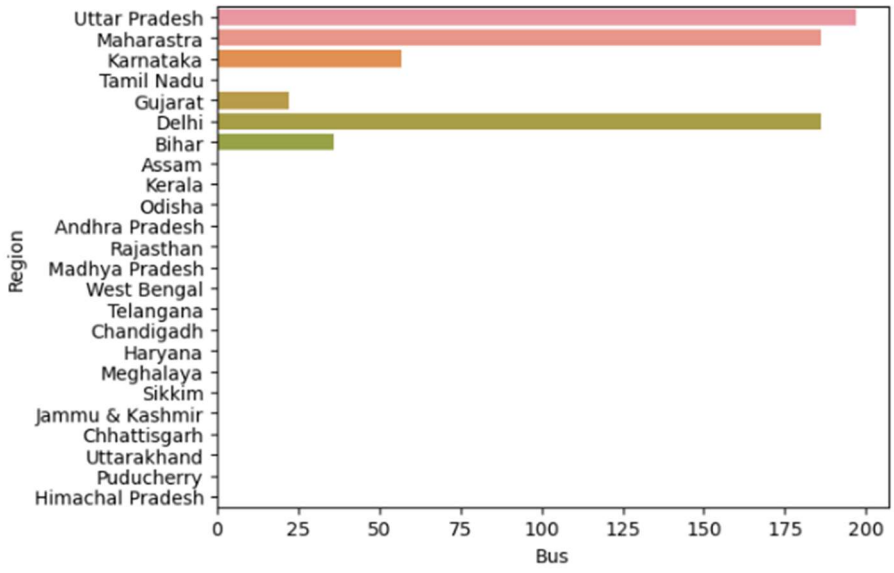
5 Exploratory Data Analysis

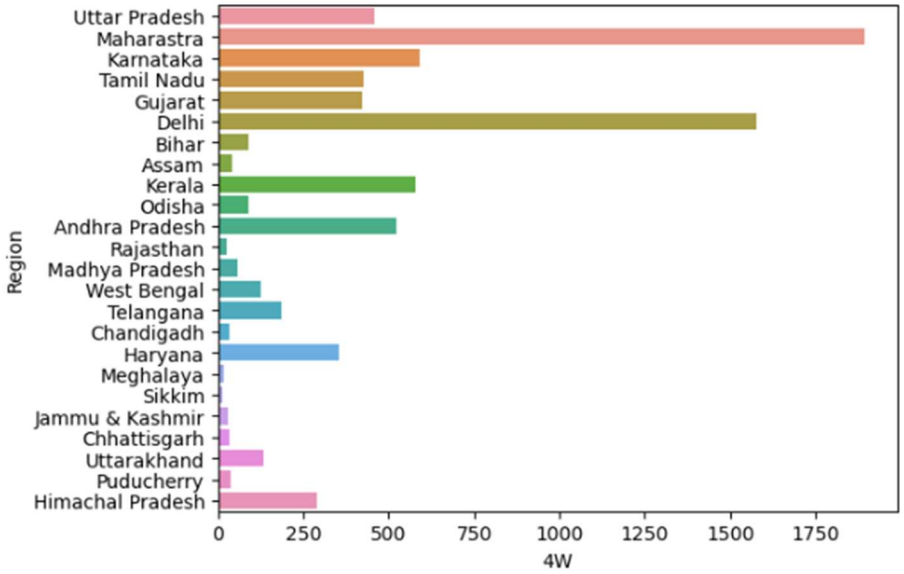
An Exploratory Data Analysis or EDA is a thorough examination meant to uncover the underlying structure of a data set and is important for a company because it exposes trends, patterns, and relationships that are not readily apparent.

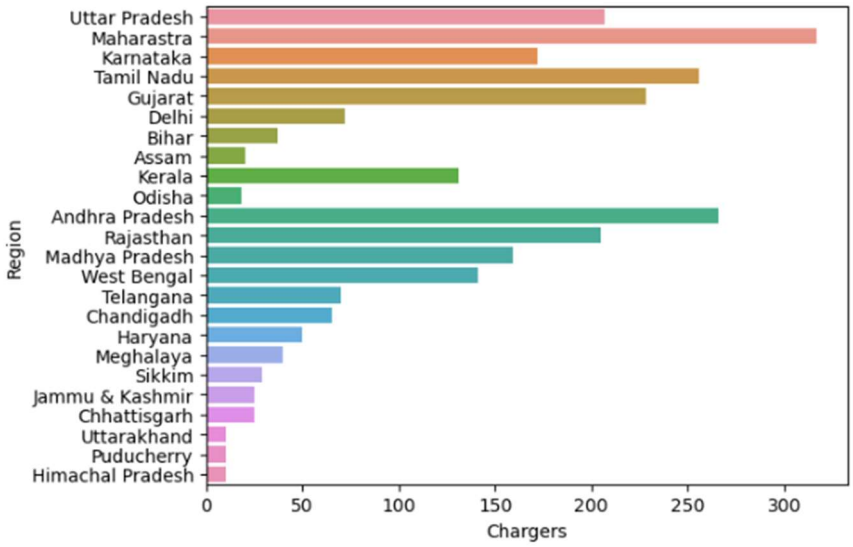
We performed EDA on two datasets. The first dataset determines the number of electric vehicles manufactured per state and number of charges in each state.




Uttar Pradesh
Maharastra
Karnataka
Tamil Nadu
Gujarat
Delhi
Bihar
Assam
Kerala
Odisha
Andhra Pradesh
Rajasthan
Madhya Pradesh
West Bengal
Telangana
Chandigadh
Haryana
Meghalaya
Sikkim
Jammu & Kashmir
Chhattisgarh
Uttarakhand
Puducherry
Himachal Pradesh
Region
0
10000
20000
30000
40000
3W

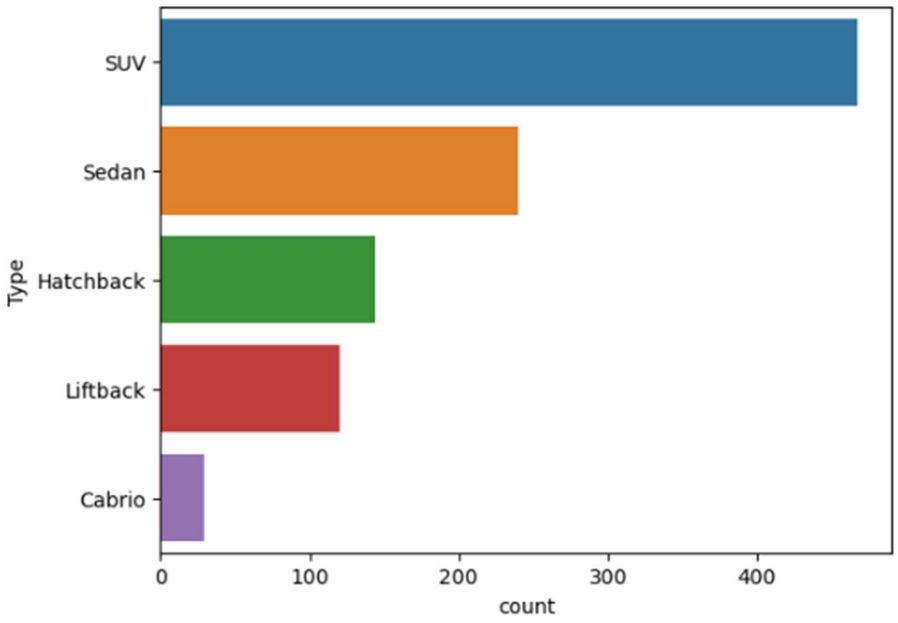

Uttar Pradesh
Maharastra
Karnataka
Tamil Nadu
Gujarat
Delhi
Bihar
Assam
Kerala
Odisha
Andhra Pradesh
Rajasthan
Madhya Pradesh
West Bengal
Telangana
Chandigadh
Haryana
Meghalaya
Sikkim
Jammu & Kashmir
Chhattisgarh
Uttarakhand
Puducherry
Himachal Pradesh
Region
0
25
50
75
100
125
150
175
200
Bus

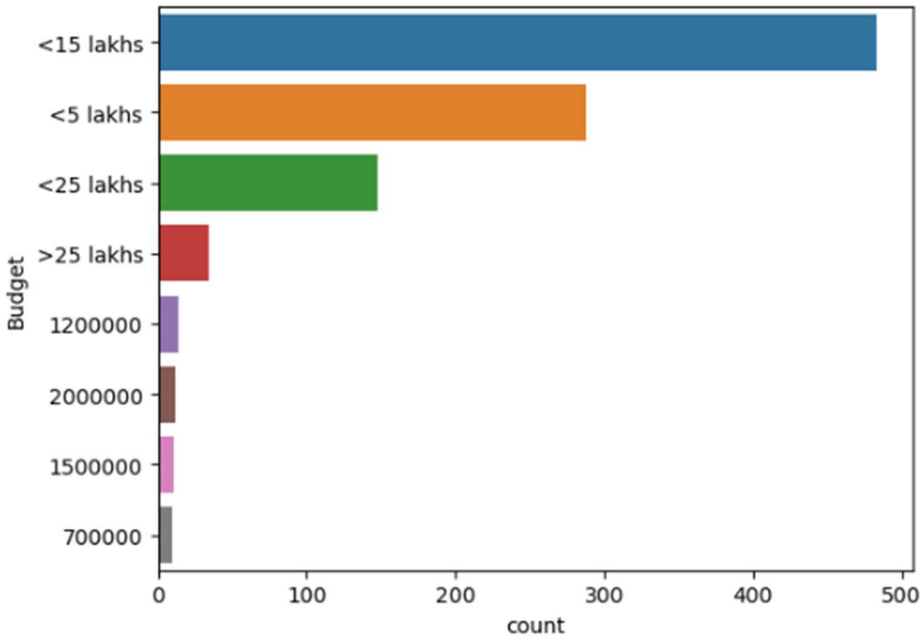

Uttar Pradesh
Maharastra
Karnataka
Tamil Nadu
Gujarat
Delhi
Bihar
Assam
Kerala
Odisha
Andhra Pradesh
Rajasthan
Madhya Pradesh
West Bengal
Telangana
Chandigadh
Haryana
Meghalaya
Sikkim
Jammu & Kashmir
Chhattisgarh
Uttarakhand
Puducherry
Himachal Pradesh
Region
0
250
500
750
1000
1250
1500
1750
4W

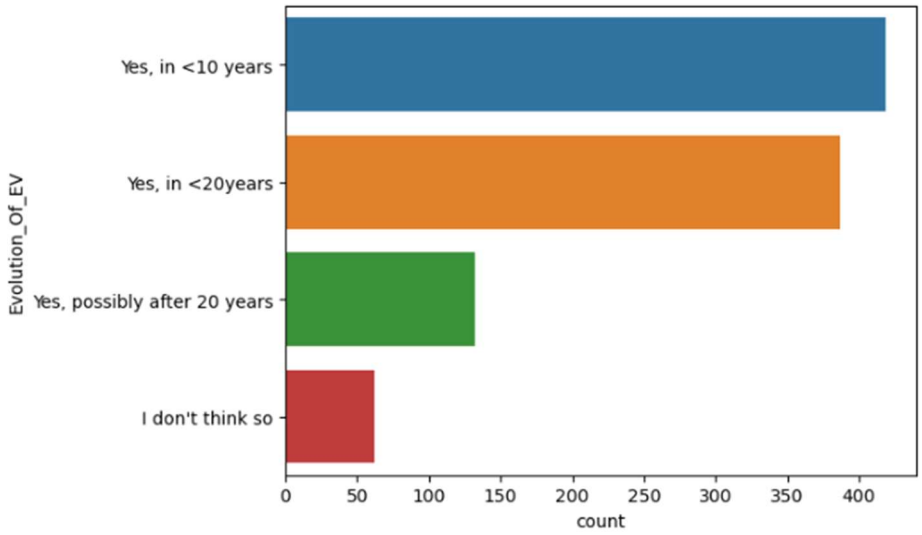

Uttar Pradesh
Maharastra
Karnataka
Tamil Nadu
Gujarat
Delhi
Bihar
Assam
Kerala
Odisha
Andhra Pradesh
Rajasthan
Madhya Pradesh
West Bengal
Telangana
Chandigadh
Haryana
Meghalaya
Sikkim
Jammu & Kashmir
Chhattisgarh
Uttarakhand
Puducherry
Himachal Pradesh
Region
0
50
100
150
200
250
300
Chargers

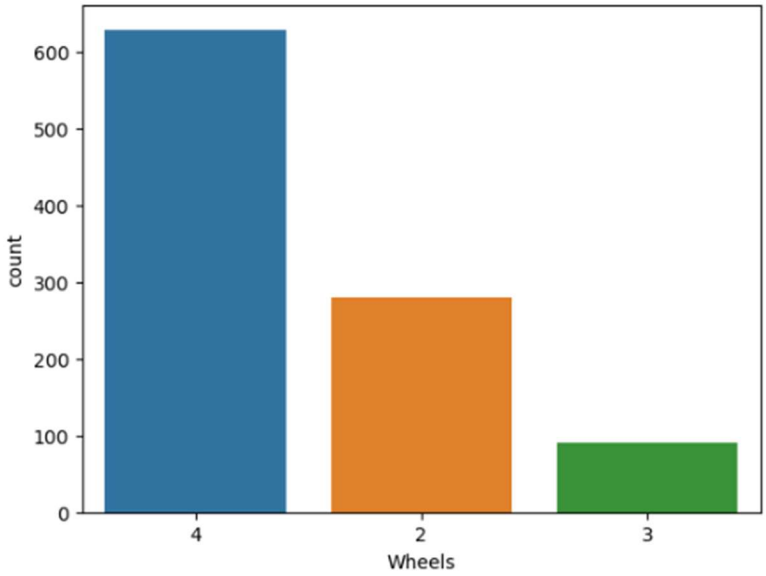
Maharashtra uses the highest number of electric vehicles in entire India with the highest usage of 2 wheeler and 4 wheeler vehicles. Maharashtra and Karnataka have the highest number of 2 wheeler EVs in India. Uttar Pradesh has the highest number of 3 wheeler EVs in India. Maharashtra and Delhi have the highest number of 4 wheeler EVs in India. Only 6 states have electric buses throughout India. Maharashtra, Andhra Pradesh and Tamil Nadu are leading producers of EV chargers.

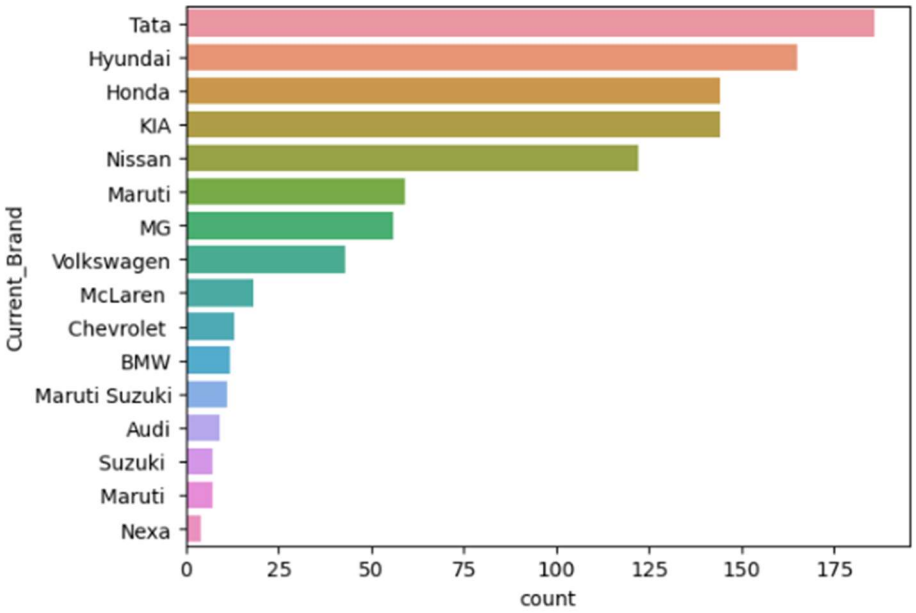
The second dataset was a survey conducted regarding the preferences and views of different people regarding electric vehicles. The variables consist of profession, annual income, vehicle choice, budget, their views regarding EV evolution.






Yes, in <10 years
Yes, in <20years
Yes, possibly after 20 years
I don't think so
Evolution_Of_EV
0
50
100
150
200
250
300
350
400
count


600
500
400
300
200
100
0
count
4
2
3
Wheels


Tata
Hyundai
Honda
KIA
Nissan
Maruti
MG
Volkswagen
McLaren
Chevrolet
BMW
Maruti Suzuki
Audi
Suzuki
Maruti
Nexa
Current_Brand
0
25
50
75
100
125
150
175
count

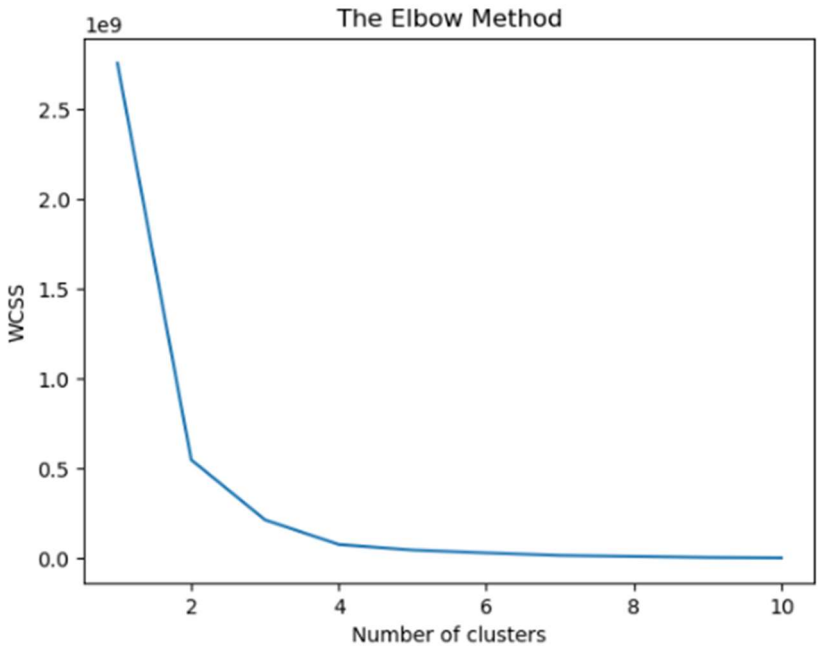
A large number of the survey takers owned Tata, Hyundai, Honda and Kia. They prefer to own a SUV, Sedan as their preferred electric vehicle. They prefer to buy such electric vehicles having a budget less than 15 lakh rupees. Very less people want to buy a two wheeler EV. Majority of the people are positive about evolution of electric vehicles and believe that the EV market will evolve within the next 10 years.

6 K-Means Clustering

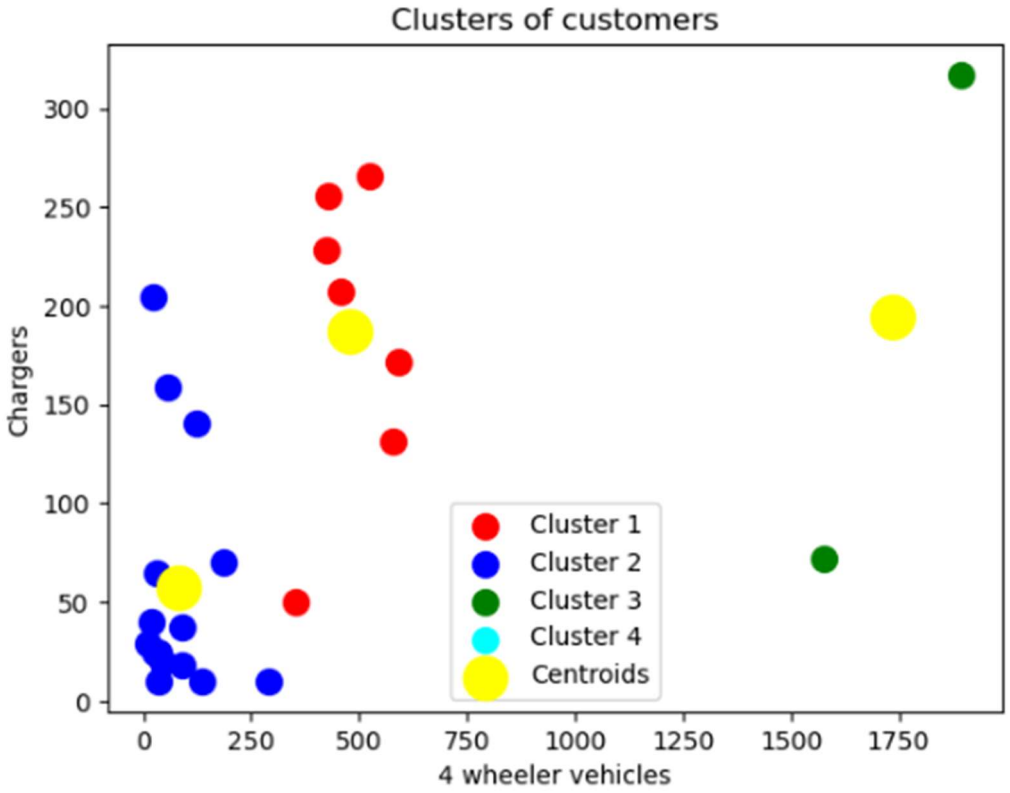
K-Means Clustering is one of the most popular Unsupervised Machine Learning Algorithms Used for Solving Classification Problems. K Means segregates the unlabeled data into various groups, called clusters, based on having similar features, common patterns. Suppose we have N number of Unlabeled Multivariate Datasets of various features like water-availability, price, city etc. from our dataset. The technique to segregate Datasets into various groups, on the basis of having similar features and characteristics, is called Clustering. The groups being Formed are known as Clusters. Clustering is being used in Unsupervised Learning Algorithms in Machine Learning as it can segregate multivariate data into various groups, without any supervisor, on the basis of a common pattern hidden inside the datasets. In the Elbow method, we are actually varying the number of clusters (K) from 1 – 10. For each value of K, we are calculating WCSS (Within-Cluster Sum of Square). WCSS is the sum of squared distance between each point and the centroid in a cluster. When we plot the WCSS with the K value, the plot looks like an Elbow.

1st dataset

For the survey of the 2nd dataset, we observe that more people intend to purchase 4 wheeler vehicles, so in the 1st dataset we considered the features: 4 wheeler, bus and number of chargers.



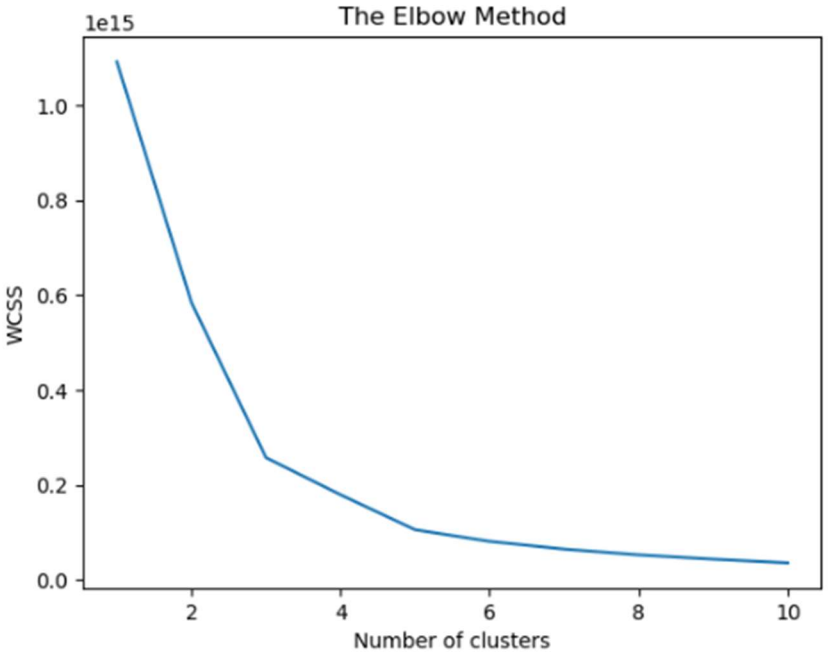
As we saw, there is a clink on 4th cluster. So we'll be selecting 4 clusters



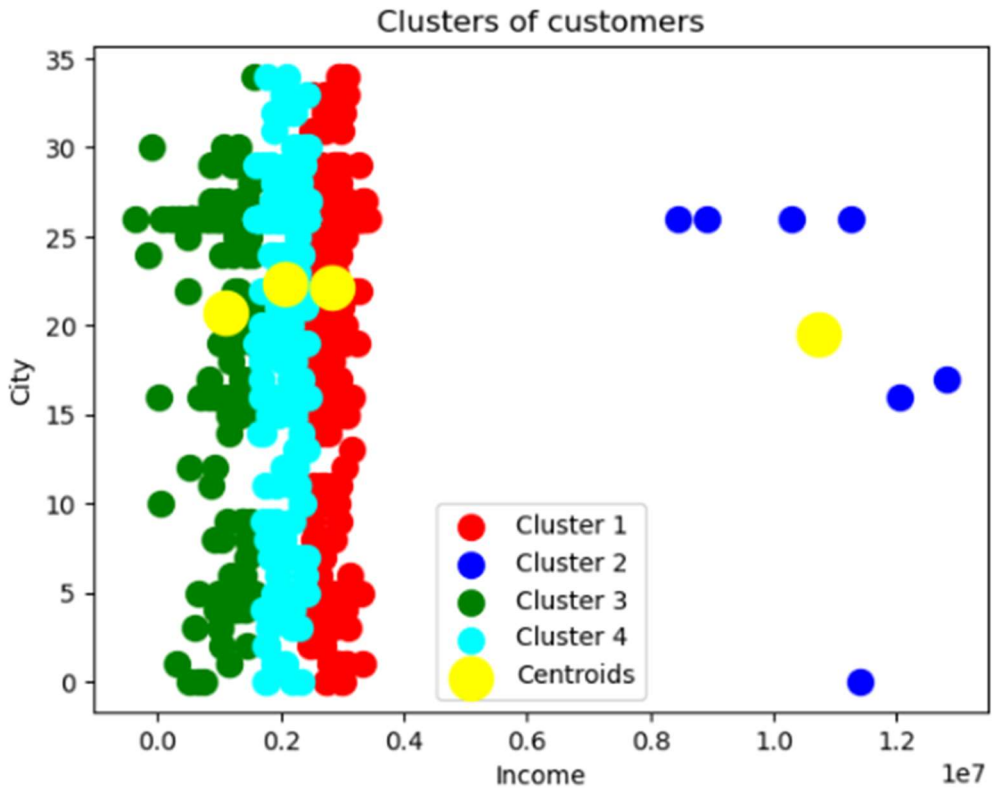
The K-Means cluster plot for the 1st dataset is as shown above.

2nd dataset

For the 2nd dataset, we considered Age, Income and Budget as our essential features for the model



As we saw, there is a clink on 4th cluster. So we'll be selecting 4 clusters



The K-Means cluster plot for the 2nd dataset is as shown above.

7 Principal Component Analysis

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into values of linearly uncorrelated variables called principal components. This transformation is defined so that the first principal component has the largest possible variance (that is, it accounts for as much of the variability in the data as possible). Each succeeding component has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components.

PCA is a widely used dimensionality reduction technique for data analysis. It can reduce the number of features in a dataset while preserving as much information as possible. This can help make the data easier to visualize and analyze and improve the performance of machine learning algorithms.

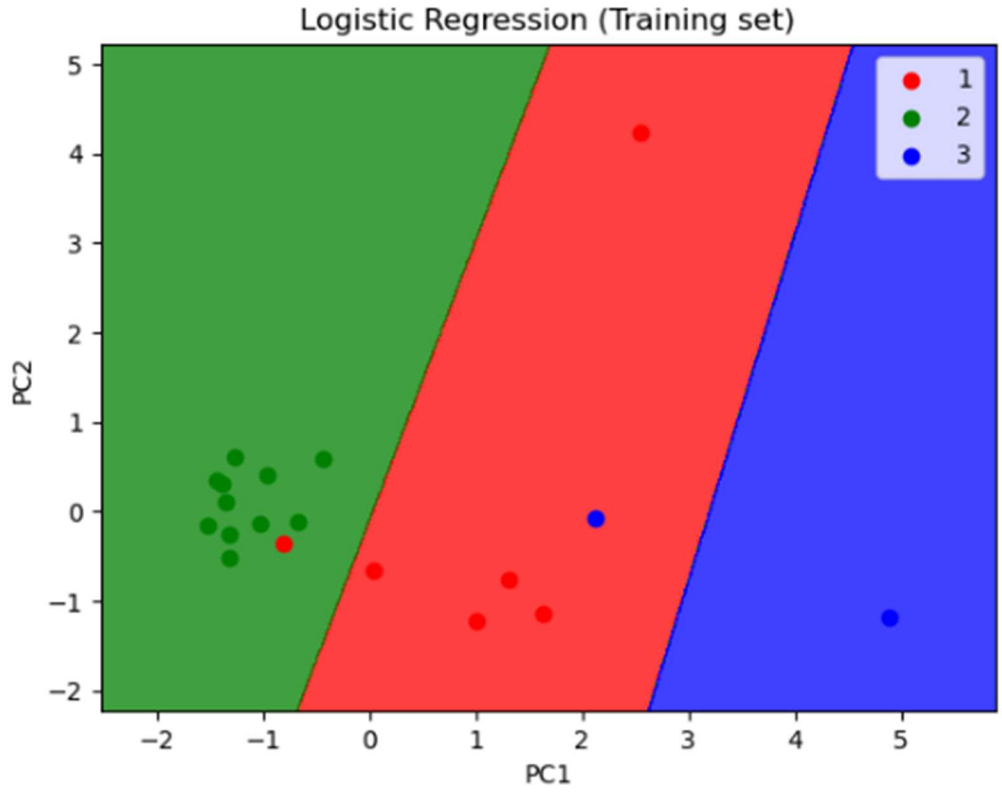
The clusters of each observation obtained from the K-Means model were added to the dataset and were used in the PCA model to check the accurate prediction of each cluster point.

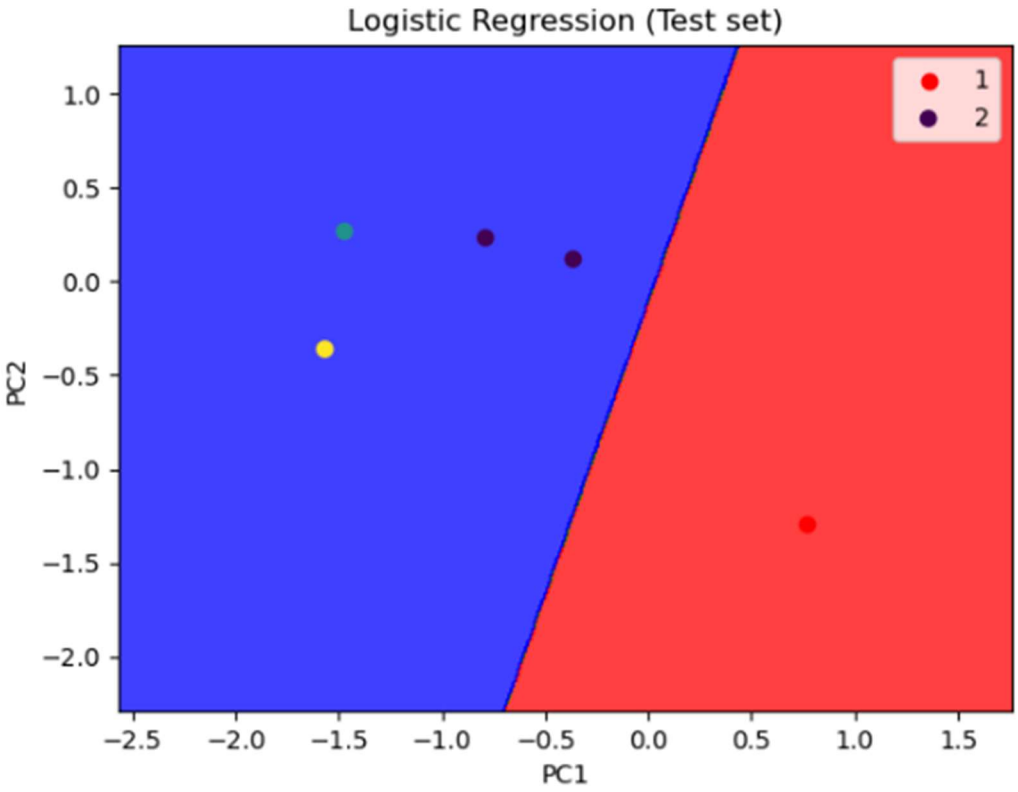
Like we do in supervised learning, we split the dataset into train and test data to see how our model works on unseen data. Then we used a logistic regressor to train the model and evaluate the principal components on the train and test data.

1st dataset

```
In [101]: from sklearn.metrics import confusion_matrix, accuracy_score
          y_pred = classifier.predict(X_test)
          print(confusion_matrix(y_test, y_pred))
          accuracy_score(y_test, y_pred)

          [[1 0]
           [0 4]]

Out[101]: 1.0
```

```
In [104]: from sklearn.metrics import confusion_matrix, accuracy_score
          y_pred = classifier.predict(X_train)
          print(confusion_matrix(y_train, y_pred))
          accuracy_score(y_train, y_pred)

          [[ 5  1  0]
           [ 0 11  0]
           [ 1  0  1]]

Out[104]: 0.8947368421052632
```

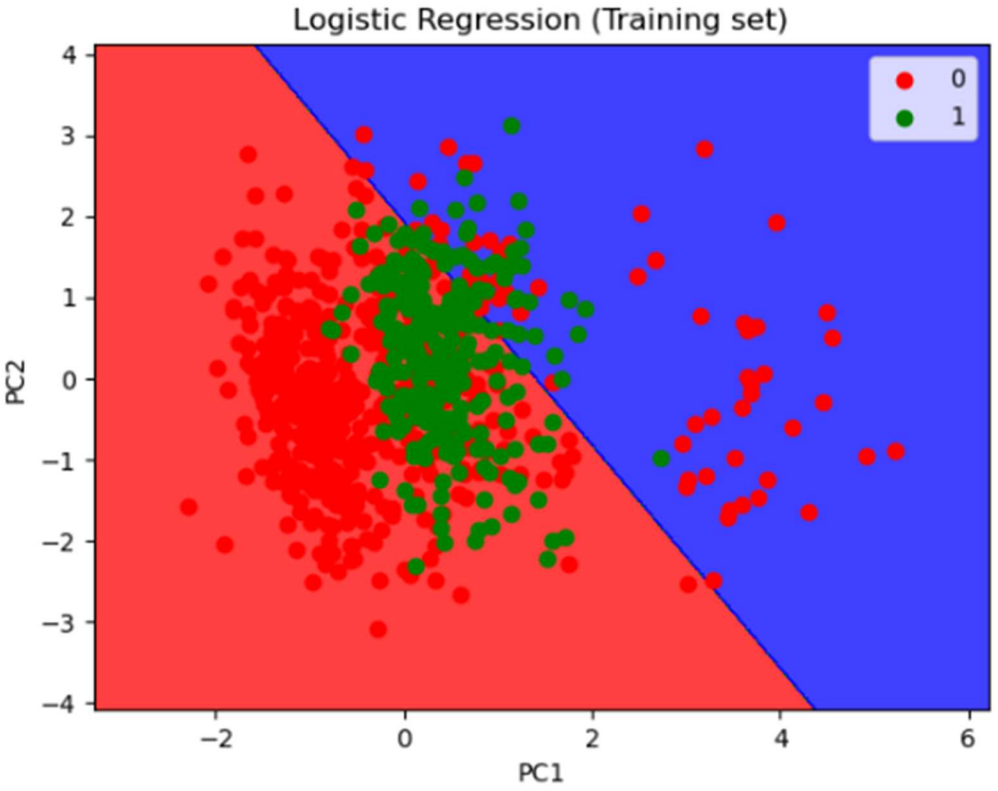


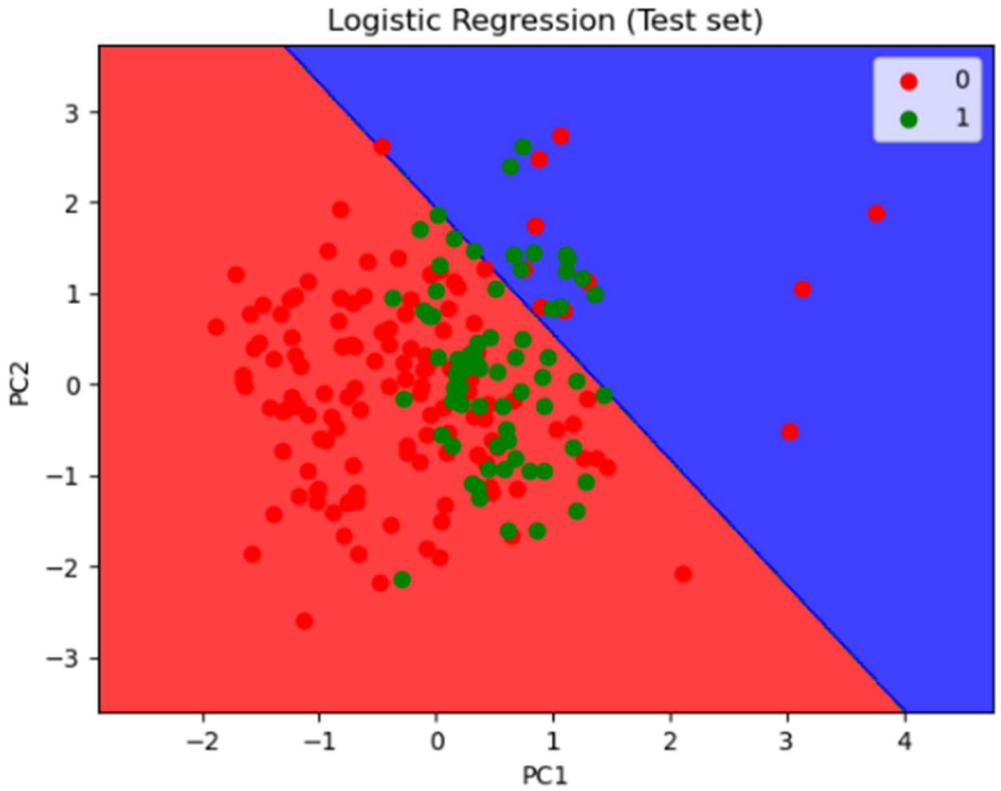


We were able to achieve an accuracy of 89.47% on the training set and 100% accuracy on the test set.

2nd dataset

```
In [115]: from sklearn.metrics import confusion_matrix, accuracy_score
          y_pred = classifier.predict(X_test)
          cm = confusion_matrix(y_test, y_pred)
          print(cm)
          accuracy_score(y_test, y_pred)

          [[123   7]
           [ 63   7]]

Out[115]: 0.65

In [116]: from sklearn.metrics import confusion_matrix, accuracy_score
          y_pred = classifier.predict(X_train)
          print(confusion_matrix(y_train, y_pred))
          accuracy_score(y_train, y_pred)

          [[488  46]
           [242  23]]

Out[116]: 0.639549436795995
```





We were able to achieve an accuracy of 63.95% on the training set and 65% accuracy on the test set.

8 Final Strategy

The first dataset identifies cluster 3 as its potential target segment to set up more 4 wheeler vehicles in the state of Maharashtra and Delhi. This is because Mumbai and Delhi are economic states, so they have the purchasing power to automate more vehicles and chargers. Delhi has widely adopted electric vehicles, but it needs to introduce more chargers to be able to produce more vehicles and compete with Mumbai. The customers in this cluster are mostly working class and are expected to be aware of technology, traffic problems, and the environment's cleanliness. The states belonging to cluster 2 such as Bihar, Assam, Rajasthan,, Madhya Pradesh, etc has more enough number of chargers but the import rate of vehicles is very low.

In the second dataset, the cluster 2 is the target segment with people belonging from Mumbai, Pune and Ahmedabad having an annual income between 8 to 12 lacs. They are more likely ready to convert their standard cars to electric vehicles and the vehicles should be priced considering their income.

Following these plans ensures a well-organized path to the launch and growth of the company in the early business phases. However, extensive research and an active feedback loop are always necessary to quickly address issues, enhance product quality and be aware of customer experiences.

9 Code Implementation

The directory containing the code can be found here on my GitHub repository: https://github.com/Sharath1036/feynn-labs-projects/tree/main/Task%202.1:%20Electric%20Vehicles%20Segmentation