Stock Market Prediction Using AI & EV Market Segmentation

Business Model and
Financial Forecasting

A report submitted as part of the requirements for the internship
at the Feynn Labs India

Report Submitted by: Sayantan Mandal

11th august 2024

Step 1: Prototype Selection

The stock market is known for its volatility and complexity, making it a challenging area for investors.This project aims to develop a machine learning-based prototype that predicts stock prices and trends using historical data. The objective is to assist investors in making informed decisions by providing accurate predictions, thus reducing investment risks.

Problem Statement:

The unpredictability of the stock market often leads to significant financial losses. By leveraging AI and machine learning, we aim to create a model that can predict stock prices and market trends, providing investors with actionable insights to maximize their returns.

Market/Customer/Business Need Assessment:

The financial market is highly competitive, with investors constantly seeking tools to gain an edge. A reliable stock prediction model can attract a wide range of users, from individual investors to financial institutions, offering them a powerful tool for making data-driven investment decisions.

Target Specifications and Characterization:

1. **Accuracy**: The model should provide highly accurate predictions to be trusted by investors.
2. **User-Friendly Interface**: An easy-to-use interface that can be accessed by both novice and experienced investors.
3. **Real-Time Data Processing**: The ability to process and analyze real-time market data for up-to-date predictions.

External Search (Information and Data Analysis): Extensive research on stock market prediction techniques, including time series analysis, regression models, and AI-based approaches, will be conducted. The project will utilize publicly available stock market datasets from sources like Yahoo Finance or Kaggle.

Benchmarking: Comparing the performance of our prototype with existing stock prediction models to ensure competitiveness and identify areas for improvement.

Applicable Constraints:

- **Data Availability**: Reliable and comprehensive data is crucial for accurate predictions.
- **Market Fluctuations**: Unpredictable events like economic crises can impact the model's accuracy.

Step 2: Prototype Development

Model Building:

The prototype will be built using Python, leveraging libraries like `pandas`, `numpy`, and `scikit-learn` for data preprocessing, and `tensorflow` or `keras` for implementing machine learning models. A time series forecasting model like Long Short-Term Memory (LSTM) will be developed for predicting stock prices

Step 3: Business Modeling

Business Model:

The service can be offered as a subscription-based platform where users pay a monthly fee to access stock predictions. The platform can also provide premium features, such as real-time alerts and detailed market analysis, to enhance user experience and generate additional revenue.

Market Analysis: The product will be targeted at individual investors, financial advisors, and investment firms. A detailed market analysis will be conducted to identify potential user segments and forecast adoption rates.

Financial Equation: The financial model will be based on linear regression, with the following equation

Step 4: Financial Modeling

Market Analysis: The product will be targeted at individual investors, financial advisors, and investment firms. A detailed market analysis will be conducted to identify potential user segments and forecast adoption rates.

Financial Equation: The financial model will be based on linear regression, with the following equation: y=mx (t) +cy = mx (t) + cy=mx (t) +c

Where:

- y = Total profit
- mmm = Subscription fee
- x(t)x(t)x(t) = Number of subscribers over time
- ccc = Operational costs

If the market grows exponentially, the financial model will be adjusted accordingly to reflect an exponential growth trend.

Code Implementation

Below is a simplified version of the Python code for stock market prediction using LSTM:

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM

# Load the dataset
data = pd.read_csv('stock_data.csv')
data = data['Close'].values.reshape(-1, 1)

# Data preprocessing
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data)

# Split data into training and testing sets
train_size = int(len(scaled_data) * 0.8)
train_data, test_data = scaled_data[:train_size], scaled_data[train_size:]

# Create a dataset with 60 timesteps
def create_dataset(data, time_step=60):
    X, y = [], []
    for i in range(len(data) - time_step - 1):
        X.append(data[i:(i + time_step), 0])
        y.append(data[i + time_step, 0])
    return np.array(X), np.array(y)

X_train, y_train = create_dataset(train_data)
X_test, y_test = create_dataset(test_data)

# Reshape the input to be [samples, time steps, features]
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], 1)
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], 1)

# Build the LSTM model
model = Sequential()
model.add(LSTM(50, return_sequences=True, input_shape=(X_train.shape[1], 1)))
model.add(LSTM(50, return_sequences=False))
model.add(Dense(25))
model.add(Dense(1))

# Compile the model
```

```
model.compile(optimizer='adam', loss='mean_squared_error')

# Train the model
model.fit(X_train, y_train, batch_size=1, epochs=1)

# Make predictions
predictions = model.predict(X_test)
predictions = scaler.inverse_transform(predictions)

# Save the model
model.save('stock_prediction_mode
```

EV MARKET SEGMENTATION

“In order to have clean air in cities, you have to go electric “

- Elon Musk

Introduction

Electronic vehicles, also known as electric vehicles, are automobiles that are powered by one or more electric motors, using electrical energy stored in batteries. They are an alternative to traditional internal combustion engine (ICE) vehicles that rely on gasoline or diesel.

Types of Electric Vehicles:

1. Battery Electric Vehicles (BEVs):

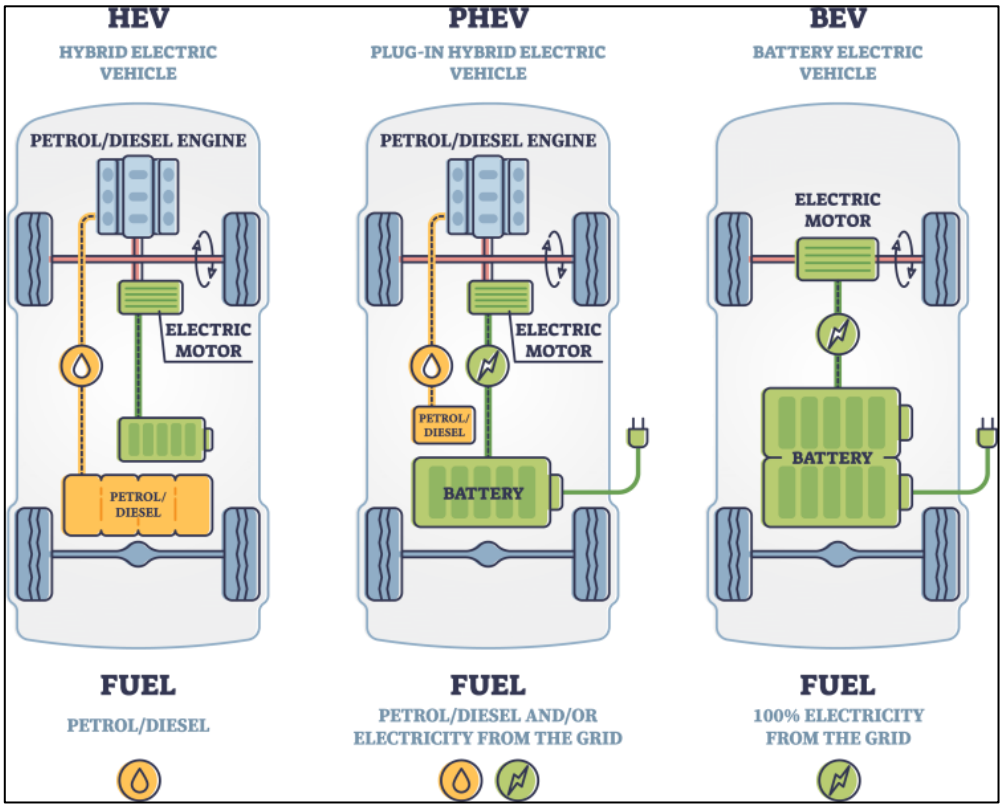
These vehicles run entirely on electricity stored in onboard batteries. They do not have an internal combustion engine and produce zero tailpipe emissions.

2. Plug-in Hybrid Electric Vehicles (PHEVs):

PHEVs have both an electric motor and an internal combustion engine. They can be plugged in to charge the battery and also use gasoline or another fuel source. They can run on electric power alone for shorter distances.

3. Hybrid Electric Vehicles (HEVs):

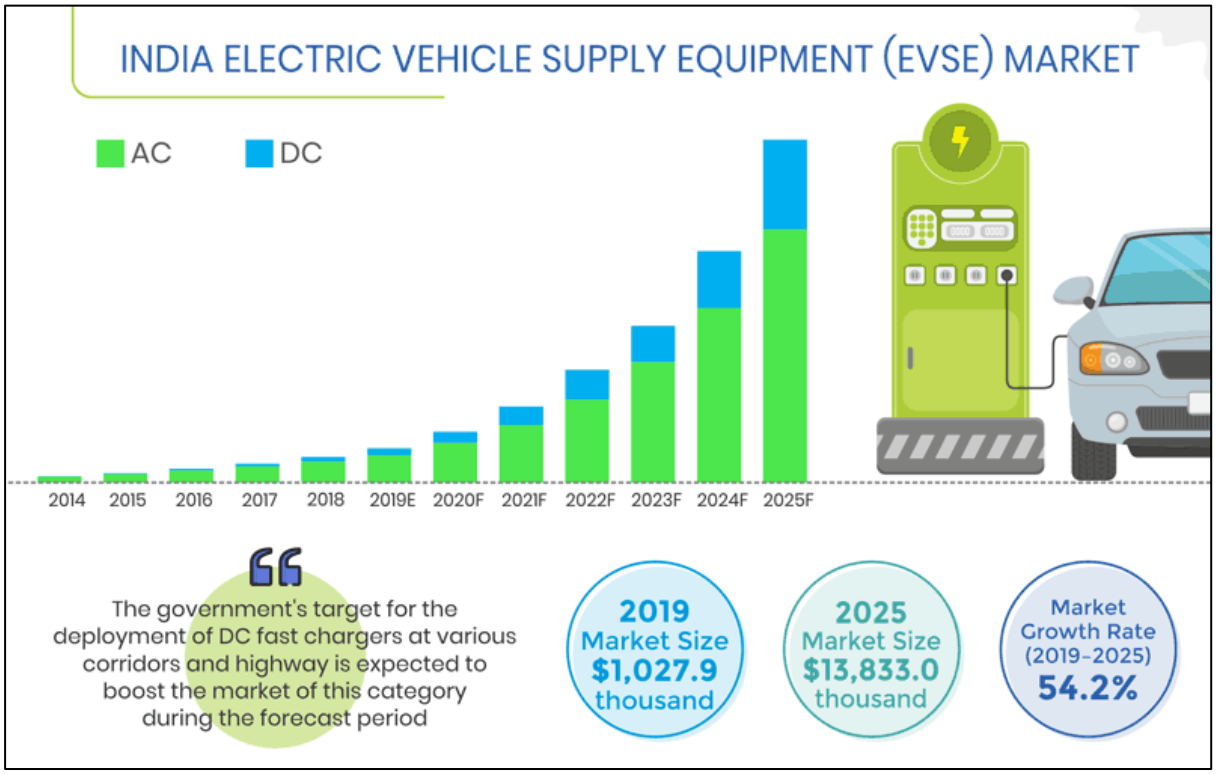
These vehicles combine an internal combustion engine with an electric motor. They do not need to be plugged in as the battery is charged through regenerative braking and the internal combustion engine.



2. Market of Electronic Vehicles in India

According to a recent report, India's electric vehicle (EV) market, including EV two-wheelers and three-wheelers, is expected to grow at a compounded annual growth rate (CAGR) of 90% to touch
$150 billion by 2030. The Indian electric vehicle industry is estimated to grow at a CAGR of 90%. The sales of electric vehicles accounted for barely 1.3% of total vehicle sales in India during the year 2020-21. The EV two-wheeler and three-wheeler market is growing rapidly. It is estimated that by the year 2030, the country's shift to high-speed electric mobility will help save nearly one gigaton of carbon dioxide emissions from vehicles. The global electric vehicle (EV) market is rapidly expanding, with the International Energy Agency (IEA) predicting that there could be as many as 145 million electric vehicles on the road by 2030. India is one of the key players in this market, and its EV industry is starting to gain traction both domestically and internationally. In this article, we will explore how the Indian EV industry is raising the temperature of the global market with well-researched data.

India has been a latecomer to the electric vehicle market, with only around 0.1% of total vehicles being electric. However, in recent years, the Indian government has implemented a series of policies and initiatives to boost the adoption of EVs. For instance, the government has launched the Faster Adoption and Manufacturing of Hybrid and Electric Vehicles (FAME) scheme, which provides financial incentives for the purchase of electric vehicles. Under FAME II, the government has allocated INR 10,000 crore ($1.4 billion) for electric vehicle adoption and infrastructure development over the next three years.



3. Market Segmentation

The global battery electric vehicles market size was $267.1 billion in 2022. The electric vehicles sector has registered exponential growth during the last decade due to the depleting fossil fuel reserves and growing awareness about the effect of global warming. Electric vehicles are a viable option to replace ICE vehicles as they do not cause environmental pollution, unlike ICE vehicles.

The EV market research report provides data and analysis on electric vehicle production and technology; sales volume, market size, policy, charging points, charging infrastructure, and market drivers and challenges for eleven key countries – The United States, Canada, China, India, Japan, South Korea, Germany, France, the Netherlands, Norway, and the United Kingdom.

Market Size (2022)	$267.1 billion
Key Regions	· Asia Pacific · Europe · Americas · Rest of the World
Key Body Type (Passenger)	· Conventional · MAV · Sporty · SUV · Van · Others
Key Category (Passenger)	· Non-Premium · Premium · Economy · Low-Cost · Super-Premium · Others
Leading Players	· Tesla Motors · BYD Auto · SAIC Group · Volkswagen Group · Hyundai Group

India is the second most populated country in the world after China, and just like China, which has the largest electric bus fleet in the world. India is also pushing hard for the electrification of buses. Many state governments have already started procuring electric buses from Chinese and local electric bus manufacturers.

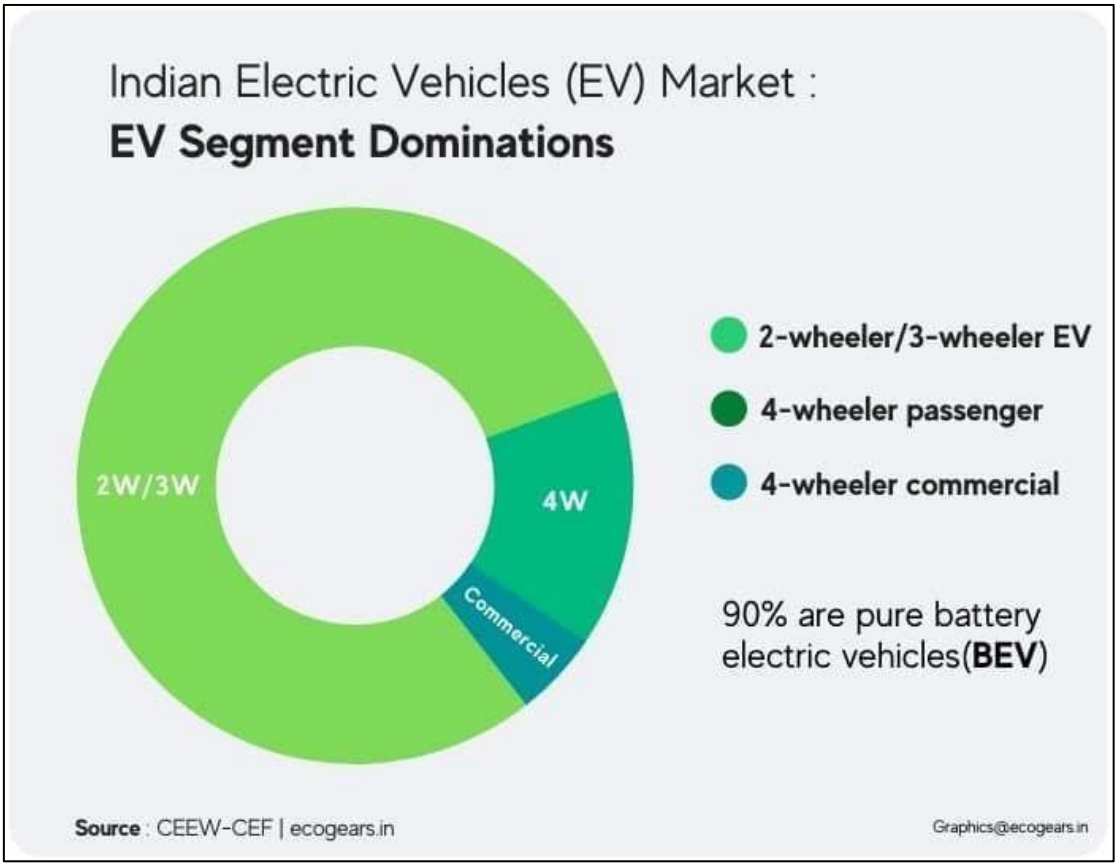
With the growing need for controlling GHG (Greenhouse gases) emissions emitted by vehicles, the government is encouraging the use of electric-powered vehicles across various states, boosting the demand for electric buses in India. The market is driven by factors such as the increase in domestic manufacturing, rapid urbanization, and a rise in environmental awareness.

For instance,

- In February 2020, the Union Transport Minister inaugurated India's first intercity electric bus service. These buses were manufactured by Mitra Mobility Solution, with a range of 300 km on a full charge.

Many local bus manufacturers who are in collaboration with some Chinese manufacturers are trying to cater to the rising demand for electric buses in India.

- In 2019, Foton PMI was planning to invest around INR 500 crore in a joint venture with Beiqi Foton Motor Co. of China to manufacture electric buses in India. The company has already given five electric buses to one of the airlines for internal operations.



4. Fermi Estimation

A market segmentation analysis for the electric vehicle (EV) market in India, keeping in mind the available data and the feasibility of targeting different customer segments. Based on the categories of segments mentioned, we'll analyze the Indian EV market and propose a feasible strategy for your startup to enter the market.

1. Geographic Segmentation:

India is a diverse country with varying geographic characteristics. You can consider segments based on regions, cities, and urban vs. rural areas. For example:

a) Urban Centers: Major cities like Delhi, Mumbai, and Bangalore, where urban congestion and pollution are significant concerns.
b) Tier 2 and Tier 3 Cities: Smaller cities with emerging markets and growing infrastructure.
c) Rural Areas: Areas with potential for last-mile connectivity solutions, agricultural use, etc.

2. Customer/Usage Segmentation:

Understanding how potential customers will use electric vehicles can be crucial. Consider segments based on customer needs and vehicle usage patterns:

(a) Daily Commuters: Individuals who need EVs for daily commuting within city limits.
(b) Commercial Fleets: Businesses operating delivery services, ride-sharing, and logistics.
(c) Tourism and Rentals: EVs for short-term rentals and tourism purposes.
(d) Government and Public Sector: EVs for government fleets, public transportation, etc.

3. Vehicle Type Segmentation:

Segmenting based on the type of EVs can help you focus on the most relevant offerings:

(a) Compact Urban EVs: Small-sized EVs designed for city commuting and short trips.
(b) Electric Sedans/Hatchbacks: EVs targeting family use and longer distances.
(c) Electric SUVs/Crossovers: Addressing the demand for larger vehicles.
(d) Commercial EVs: Vans and trucks for logistics and delivery purposes.

4. Demographic Segmentation:

While demographic data might be more challenging to obtain, you can still consider some broad categories:

(a) Age Groups: Younger individuals might be more open to adopting new technologies.
(b) Income Levels: Higher income groups might afford premium EVs.

5. SWOT Analysis for Electronic Vehicle Segmentation:

A SWOT analysis examines the Strengths, Weaknesses, Opportunities, and Threats of a particular business or industry. In this case, we'll focus on electronic vehicle segmentation within the automotive industry:

(A) Strengths:

a) Environmental Benefits: Electric vehicles are considered environmentally friendly due to zero tailpipe emissions, helping to reduce air pollution and greenhouse gas emissions.
b) Lower Operating Costs: EVs have lower fuel and maintenance costs compared to traditional internal combustion engine vehicles.
c) Innovative Technology: EVs showcase cutting-edge technology, including advanced batteries, regenerative braking, and connectivity features.
d) Government Incentives: Many governments offer incentives such as tax credits and rebates to encourage the adoption of EVs.

(B) Weaknesses:

a) Limited Range: Some electric vehicles still have limitations in terms of driving range compared to traditional gasoline-powered vehicles.
b) Charging Infrastructure Gaps: While charging infrastructure is growing, there are still areas with limited access to charging stations, which could deter potential buyers.
c) Higher Upfront Costs: EVs often have a higher upfront cost than traditional vehicles, primarily due to the cost of the battery.
d) Battery Degradation: Over time, the battery's capacity and performance may degrade, leading to reduced driving range.

(C) Opportunities:

a) Increasing Demand: As environmental concerns grow and technology advances, the demand for electric vehicles is expected to rise.
b) Advancements in Battery Technology: Improvements in battery technology could lead to longer ranges, shorter charging times, and overall better performance.
c) Diversification of Vehicle Types: EV technology can be applied to various vehicle segments, including passenger cars, SUVs, trucks, and even public transportation.
d) Smart Grid Integration: EVs can serve as a storage solution for renewable energy and be integrated into smart grids, enhancing energy efficiency.

(D) Threats:

a) Competition: Traditional automakers entering the EV market could increase competition and potentially challenge the market share of current EV manufacturers.
b) Dependency on Raw Materials: EVs rely on minerals like lithium, cobalt, and rare earth elements, which could face supply chain challenges and price fluctuations.
c) Infrastructure Strain: A rapid increase in EV adoption could strain the existing electricity grid and charging infrastructure.

6. Detailed Analysis

Collecting relevant data for electric vehicle (EV) segmentation is crucial for making informed business decisions. Data collection might involve a combination of primary data (collected directly) and secondary data (existing sources). Collaborating with research organizations, academic institutions, and industry associations can also provide access to valuable data. Ensure that any data you collect or use adheres to privacy regulations and ethical considerations.

It's important to conduct thorough research to gather accurate and up-to-date data that reflects the specific market conditions and trends in the Indian EV market. This data will serve as the foundation for your segmentation analysis and subsequent business strategy.

(A) Dataset 1: Electronic Car Data

Link for the dataset:

https://github.com/SayantanIITBombay/EV-market-segmentation/blob/main/Electric%20Car%20Data.csv

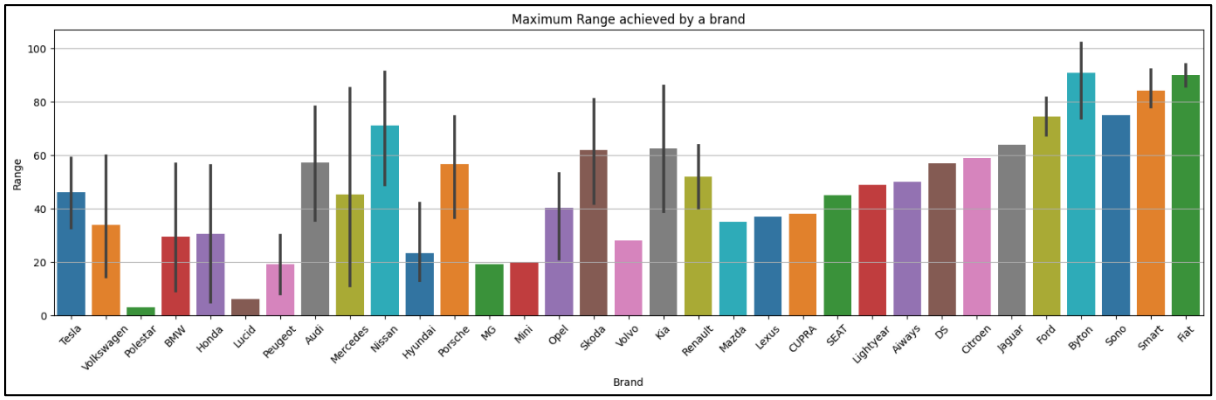
Description:

The dataset consists of various car models and on the basis of model, the speed, range, efficiency, etc. details achieved by the car is provided. It also tells us about the body style, charger type, number of seats, and price of the car.

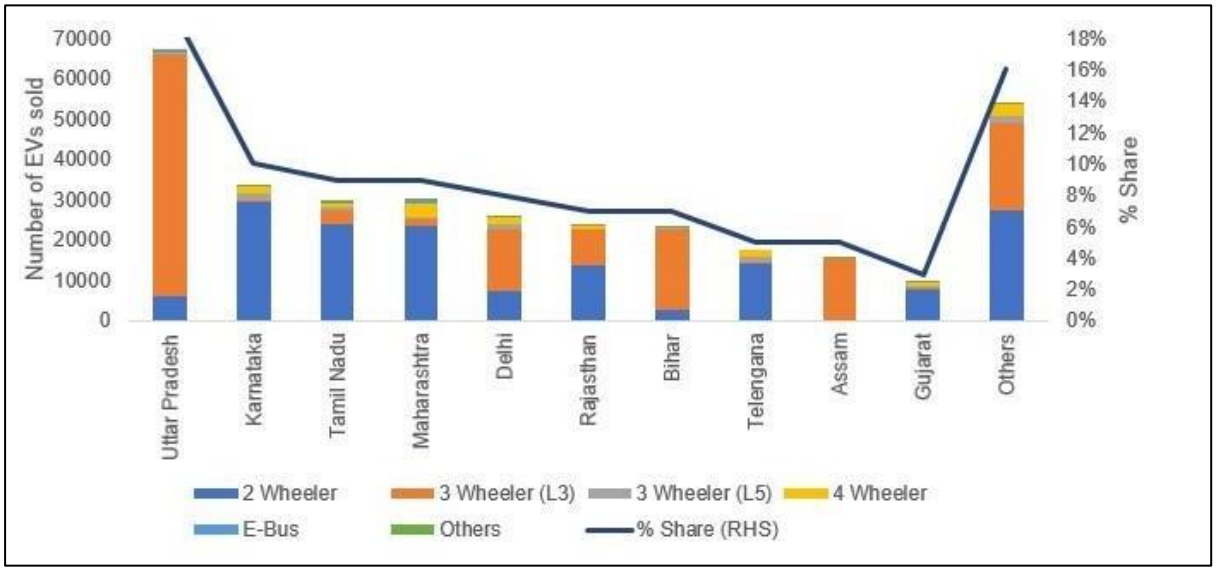
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 102 entries, 0 to 101
Data columns (total 16 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   Brand             102 non-null    object 
 1   Model             102 non-null    object 
 2   Accelsec          102 non-null    float64
 3   TopSpeed_KmH      102 non-null    int64  
 4   Range_Km          102 non-null    int64  
 5   Battery_Pack Kwh  102 non-null    float64
 6   Efficiency_WhKm   102 non-null    int64  
 7   FastCharge_KmH    102 non-null    int64  
 8   RapidCharge       102 non-null    object 
 9   PowerTrain        102 non-null    object 
 10  Plug Type         102 non-null    object 
 11  Body Style        102 non-null    object 
 12  Segment           102 non-null    object 
 13  Seats             102 non-null    int64  
 14  Price Euro        102 non-null    int64  
 15  INR               102 non-null    float64
dtypes: float64(3), int64(6), object(7)
memory usage: 12.9+ KB
```

Here is an analysis regarding the top speed achieved by each car brand:



State wise sales in each state in India of electric vehicle:

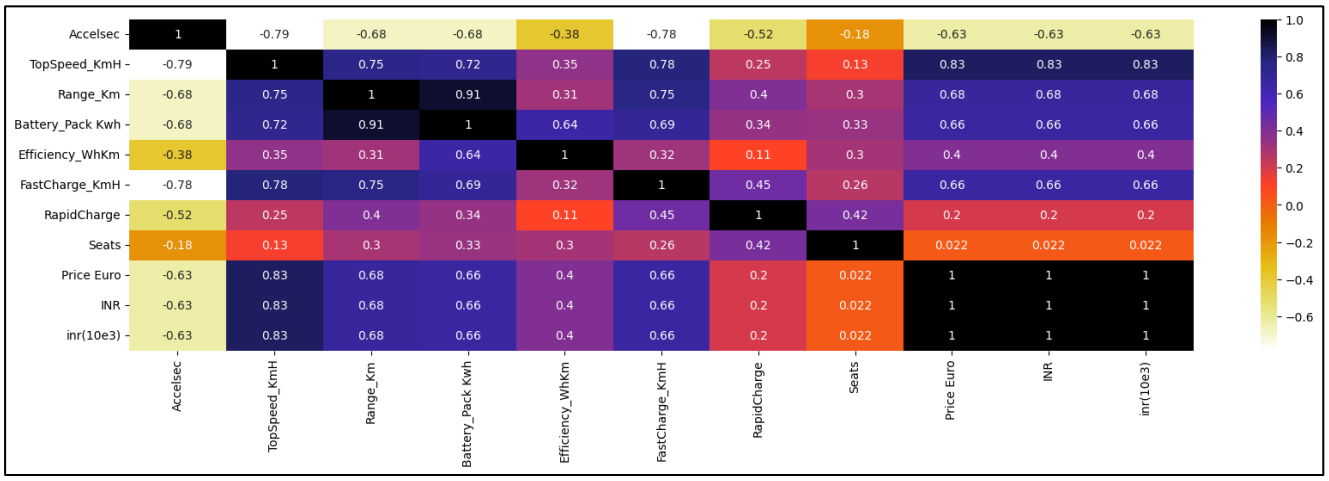


The Indian automobile industry is the fifth largest in the world and is expected to become the third largest by 2030. As per India Energy Storage Alliance (IESA), the Indian EV industry is expected to expand at a CAGR of 36%. As population rises and demand for vehicles grow, dependence on conventional energy resources is not a sustainable option as India imports close to 80% of its crude oil requirements. NITI Aayog aims to achieve EV sales penetration of 70% for all commercial cars, 30% for private cars, 40% for buses and 80% for two and three- wheelers by 2030. This is in line with the goal to achieve net zero carbon emission by 2070. Over the last three years, 0.52 million EVs were registered in India, according to the Ministry of Heavy Industries. EVs recorded robust growth in 2021, supported by the implementation of favourable policies and programmes by the government.

In India, Uttar Pradesh held the highest share in EV sales in 2021, with the number of units sold across all segments reaching 66,704, followed by Karnataka with 33,302 units and Tamil Nadu with 30,036 units. Uttar Pradesh dominated the three-wheeler segment, while Karnataka and Maharashtra led the two-wheeler segment and four-wheeler segment, respectively.

Heatmap:

We also plotted a heatmap to find correlation between the datapoints of the dataset.

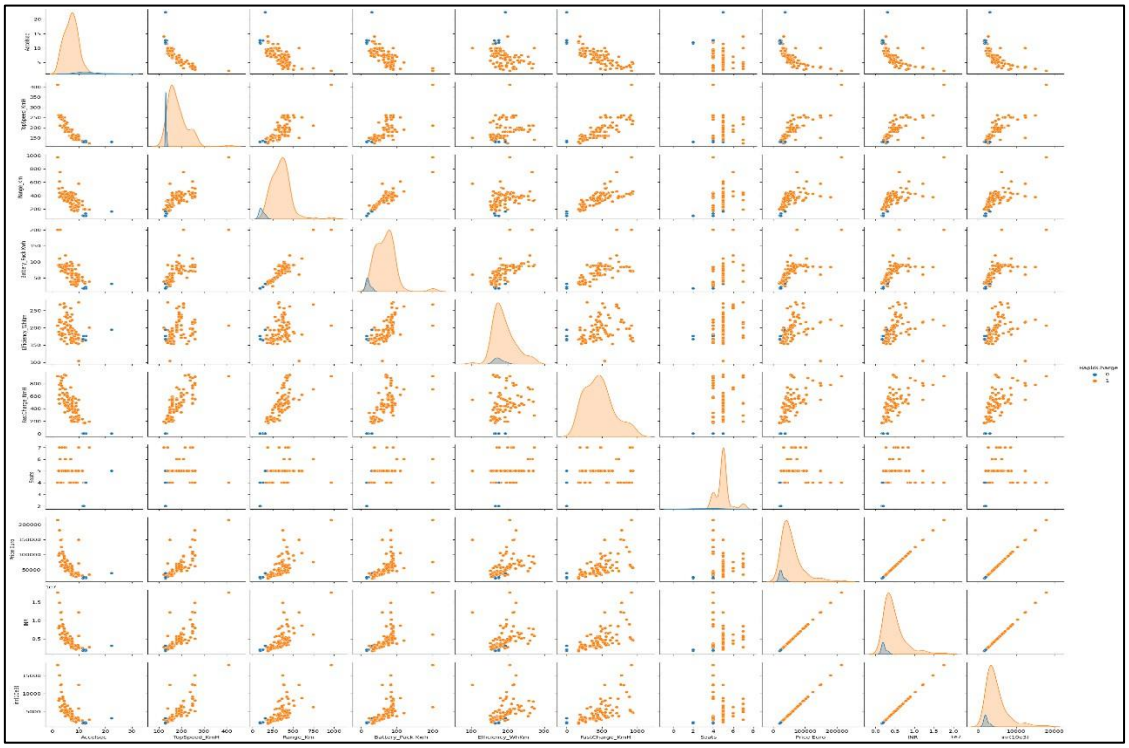


It can be clearly stated that there is a high correlation between the top speed, range, and efficiency of the car. It is also visible that the range and battery pack of the car are also related.

Pair plot:

To plot multiple pairwise bivariate distributions in a dataset, we have used the
pairplot() function.

The diagonal plots are the univariate plots, and this displays the relationship for the (n, 2) combination of variables in a DataFrame as a matrix of plots. Through this pairplot we can depict the relation between the correlation between the datasets.



Regression Analysis:

Through regression, we can find the relationship between the independent variables such as top speed, range, and efficiency with that of the dependent variable price.

```
df['PowerTrain'].replace(to_replace=['RWD','AWD','FWD'],value=[0, 2,1],inplace=True)
x=df[['Accelsec','Range_Km','TopSpeed_KmH','Efficiency_WhKm', 'RapidCharge','PowerTrain']]
y=df['Price Euro']
```

(A) Linear Regression:

```
from sklearn.linear_model import LinearRegression
lr= LinearRegression()
lr.fit(X_train, y_train)
pred = lr.predict(X_test)
pred
```

```
array([ 3744.    ,  2496.    ,  5233.28  ,  3243.7184,  2043.808 ,
        5459.584 ,  5876.4992,  3328.    ,  4451.2   ,  2594.5088,
        2654.08  ,  3744.    ,  2041.2288, 15040.9792,  6609.824 ,
        3170.336 ,  3744.    ,  2866.9888,  3744.    , 17888.    ,
        4877.184 ,  5660.928 ,  3902.08  ,  2062.528 , 12396.8   ,
        8565.024 , 12338.6432,  3328.    ,  4695.808 ,  5408.    ,
        3064.8384,  3952.    ,  2826.3872, 10400.    ,  2745.6   ,
        2912.    ,  3153.28  ,  5028.3584,  2424.9472,  2766.0672,
        5604.1856])
```

Here, we can see the predicted prices for the given data values. The prices predicted are the dependent variable whose values depends on x.

The accuracy of the model is calculated using r2_score that depicts how close the estimated value is to the actual data values. The coefficient of determination, or R2 , is a measure that provides information about the goodness of fit of a model. In the context of regression it is a statistical measure of how well the regression line approximates the actual data.

```
from sklearn.metrics import r2_score
r2=(r2_score(y_test,pred))
print(r2*100)
```

```
79.66245261701704
```

With this, we can clearly say that our model is giving an accuracy of 80% (approximately).

(B) Principal Component Analysis (PCA):

Principal Component Analysis (PCA) is the analysis of principal features of the data. The analysis is done by reducing the dimensionality of the feature space. In other words, it is a tool to reduce the features from the data to get only the required features or principal components for the learner. PCA has three major components which help to reduce dimensionality:

- The covariance matrix is the measure of how much the variables are associated with each other.
- The eigenvectors are the directors in which the data is dispersed.
- The eigenvalues are the relative importance of the directions.

```
from sklearn.decomposition import PCA
pca = PCA(n_components=9)
t = pca.fit_transform(x)
data2 = pd.DataFrame(t, columns=['PC1', 'PC2','PC3','PC4','Pc5','PC6', 'PC7', 'PC8','PC9'])
data2
```

	PC1	PC2	PC3	PC4	Pc5	PC6	PC7	PC8	PC9
0	2.342036	0.254881	-1.297001	-0.569622	0.770965	-1.304321	0.401787	-0.762618	-0.354884
1	-1.914561	-0.569047	-0.709043	0.593886	0.206999	0.352992	-0.397134	0.238896	-0.325262
2	1.515695	0.024369	-0.382073	-0.679665	0.107532	-0.729073	0.266925	0.042262	-0.087392
3	0.252431	-0.104398	-0.257001	1.601049	-0.142045	0.190743	-0.412853	-0.301661	0.185571
4	-2.561396	0.254023	-0.791450	0.322643	-0.893104	0.348270	-0.495068	0.109750	-0.165978
...	...	...	...	...	...	...	...	...	...
97	-0.305650	-0.460376	-0.075838	0.164252	-0.163668	-0.041783	0.151686	-0.126566	0.132994
98	2.231842	0.177451	1.792758	0.158852	-1.120332	0.077190	-0.342025	-0.091941	0.101495
99	0.739027	-0.181728	0.277008	-0.811593	-0.287049	-0.448344	0.056445	0.286306	-0.265459
100	1.476350	-0.129236	1.151317	-0.179274	-0.719638	-0.118880	0.269224	0.310137	-0.101881
101	1.212530	-0.258307	1.390068	-0.094184	-0.482151	0.262949	0.589068	-0.206575	-0.367528

PCA is been imported from the sklearn library and stored in a variable for easier applications. The PCA is fitted with the independent variables for dimensionality reduction. The percentage of variance in the dependent variable is explained by adding each principal component to the model.

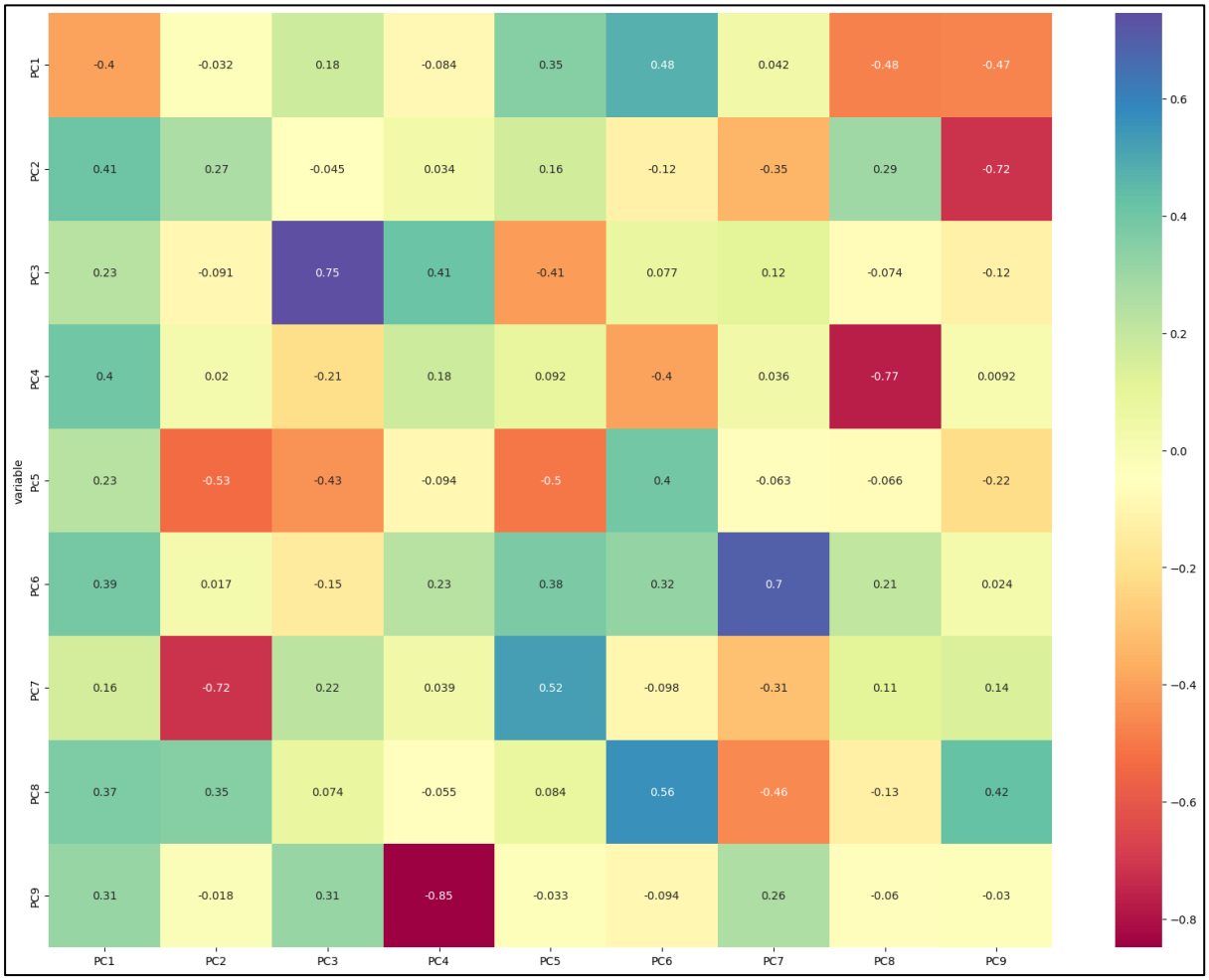
```
np.cumsum(np.round(pca.explained_variance_ratio_, decimals = 4)*100)
```

```
array([ 53.78,  68.  ,  78.72,  84.98,  90.83,  94.23,  96.8 ,  98.94,
       100.  ])
```

The above output is explained as:

- By using the first principal component, we can explain 53.78% of the variation in the dependent variable.
- By using the second principal component, we can explain 68% of the variation in the dependent variable.
- Similarly, by using others we can explain 78.72%, 84.98%, 90.83%, 94.23%, 96.8%,
  98.94% and 100%.

(C) Correlation Matrix:



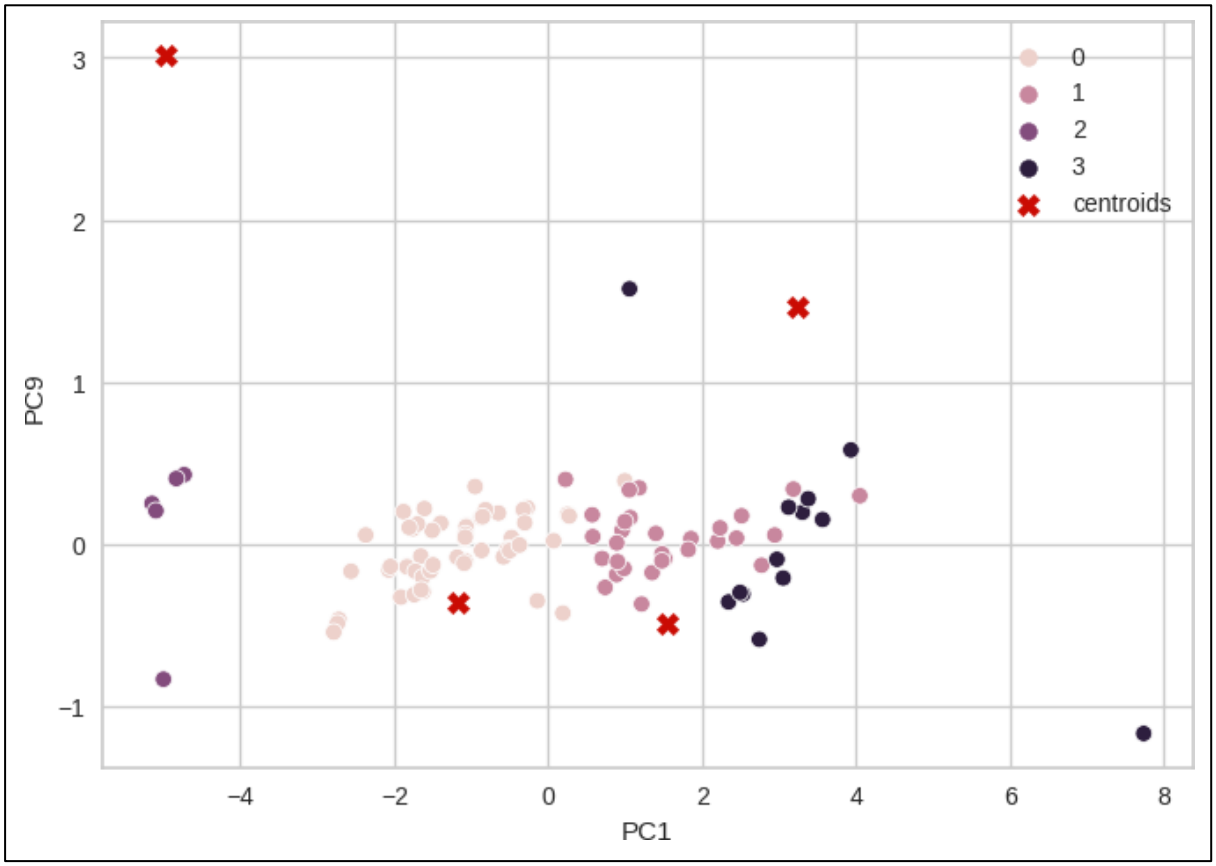
A correlation matrix is a square matrix showing the correlation coefficients between two variables. Correlation coefficients measure how strong and in which direction two variables are linked in a straight line. A correlation matrix often examines how different variables relate in multivariate analysis and statistics.

```
#Correlation matrix plot for loadings
plt.rcParams['figure.figsize'] = (20,15)
ax = sn.heatmap(loadings_df, annot=True, cmap='Spectral')
plt.show()
```

(D) KMeans Clustering:

```
#Visulazing clusters
sn.scatterplot(data=data2, x="PC1", y="PC9", hue=kmeans.labels_)
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1],
            marker="X", c="r", s=80, label="centroids")
plt.legend()
plt.show()
```

The above code is used to portray the formed clusters in the given dataset.



It can be explained that the nearest clusters formed for corresponding 0 and 1 to be close and 2 and 3 to be quite far from the actual data.

Dataset 2: Indian automobile purchasing behaviour.

Link for the dataset:

https://github.com/SayantanIITBombay/EV-market-segmentation/blob/main/Indian%20automoble%20buying%20behavour%20study%201.0.csv

Description:

This dataset describes the buying behaviour of Indian customers on the basis of various factors such as age, marital status, education, loan, and salary. On this basis, it is predicted whether the consumer will be able to purchase the vehicle or not.

```
data.info()
```

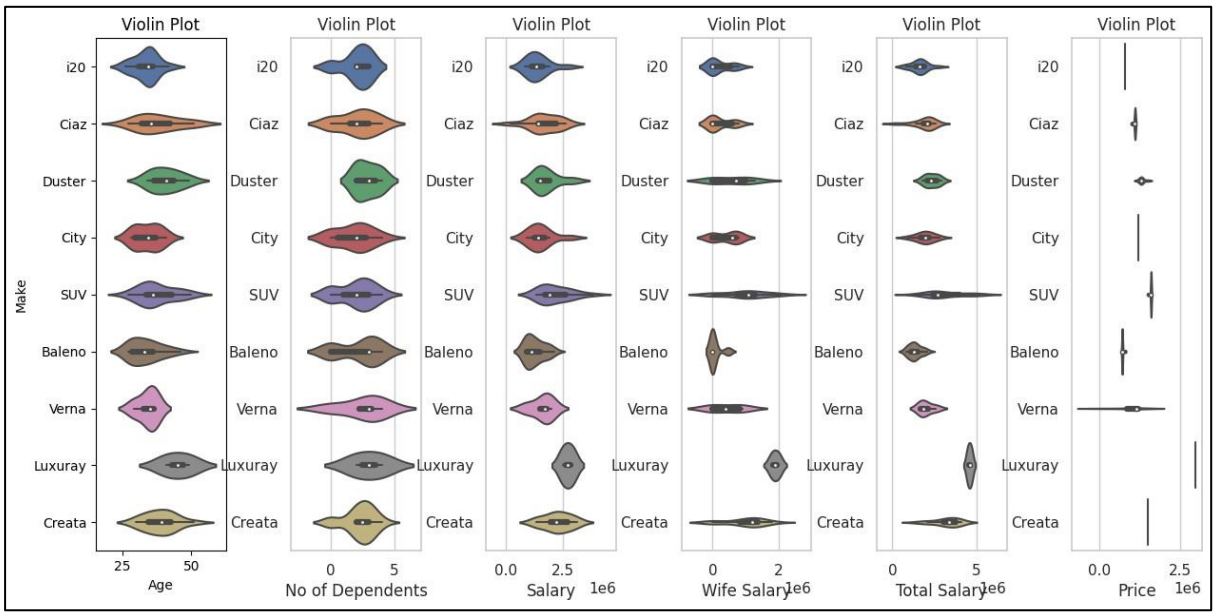
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Age               99 non-null     int64
 1   Profession        99 non-null     object
 2   Marrital Status   99 non-null     object
 3   Education         99 non-null     object
 4   No of Dependents  99 non-null     int64
 5   Personal loan     99 non-null     object
 6   House Loan        99 non-null     object
 7   Wife Working      99 non-null     object
 8   Salary            99 non-null     int64
 9   Wife Salary       99 non-null     int64
 10  Total Salary      99 non-null     int64
 11  Make              99 non-null     object
 12  Price             99 non-null     int64
dtypes: int64(6), object(7)
memory usage: 10.2+ KB
```

The Indian EV Industry is slowly gathering momentum, supported by government initiatives and rise in crude oil prices, as people look for alternative sources to reduce their monthly bills. However, a mass shift from internal combustion engine (ICE) vehicles to EVs requires expansion of infrastructure facilities, including charging stations, and vehicles that could provide a higher range (KM range with a single charge). Several initiatives taken by the government to support the manufacturing and adoption of electric vehicles in the country should help in achieving the target of 100% EV adoption by 2030.

Age: Younger people have a smaller number of dependents, less salary, and are single so the they are not usually our target segment but they are the most likely to buy electric vehicles are they are informed about climate change and it effects and want to help the planet. The price range for younger target segment is below 10 Lakhs.

Number of Dependents: The more the number of dependents, the bigger the need of cars for transportation, here SUVs are preferred for higher target segments.

Salary: Different demographic has different people of different salaries. This is the main thing we should consider while segmenting the market based on 4-wheeler and 2-wheeler automobiles as higher salaried people are highly likely to purchase a 4-wheeler.



To enter a market an in-depth knowledge of the end user psychology, behavior is required. This market research is imperative for setting prices, studying spending habits, studying the product they use the most, like 4-wheel diesel/petrol automobiles, what is their price range, the requirement of the automobile, etc. The next series of visualizations is regarding this niche where we do a requirement analysis.

The above violin plot shows the top existing cars and their dependency on various variables such as age, marital status, price, salary, and number of dependents.

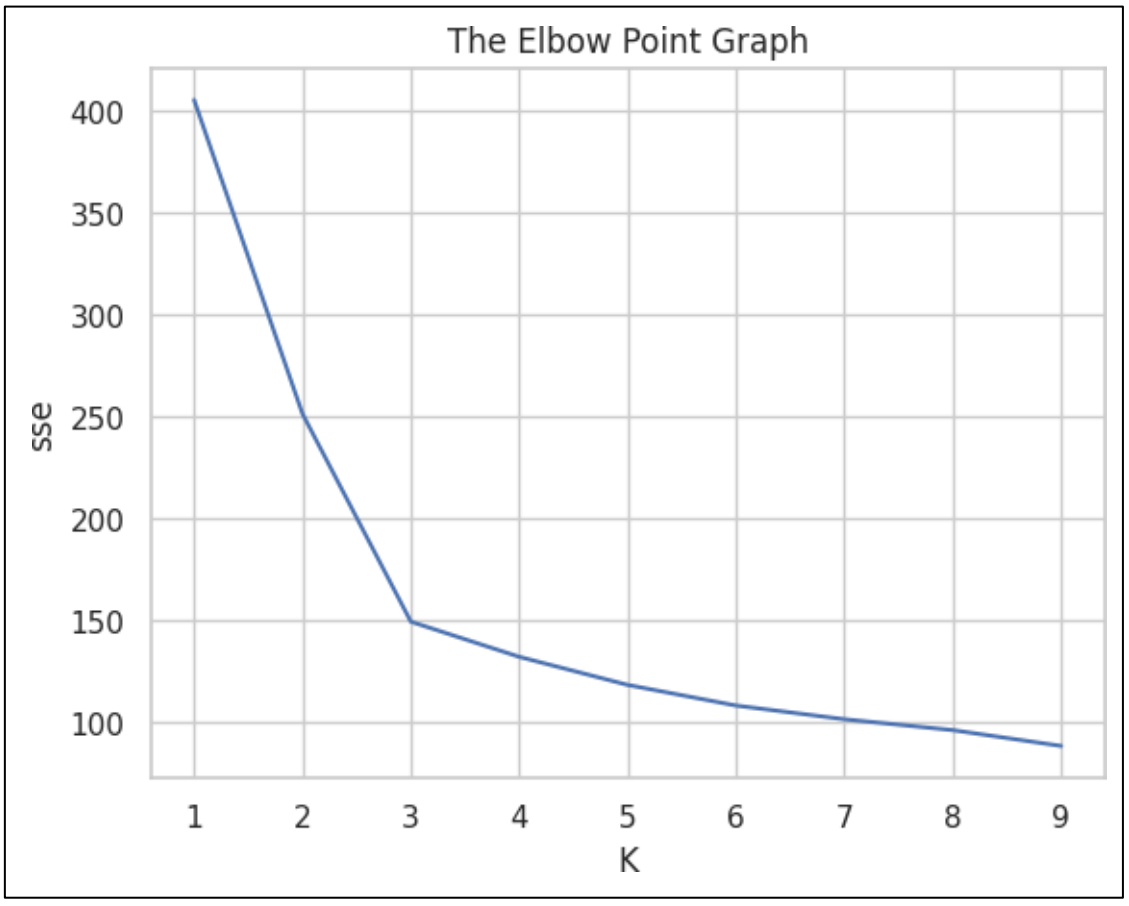
(A) KMeans Clustering:

Clustering is one of the most common exploratory data analysis techniques used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as Euclidean-based distance or correlation-based distance. The decision of which similarity measure to use is application-specific.

Firstly, the number of clusters are calculated using the Elbow method:

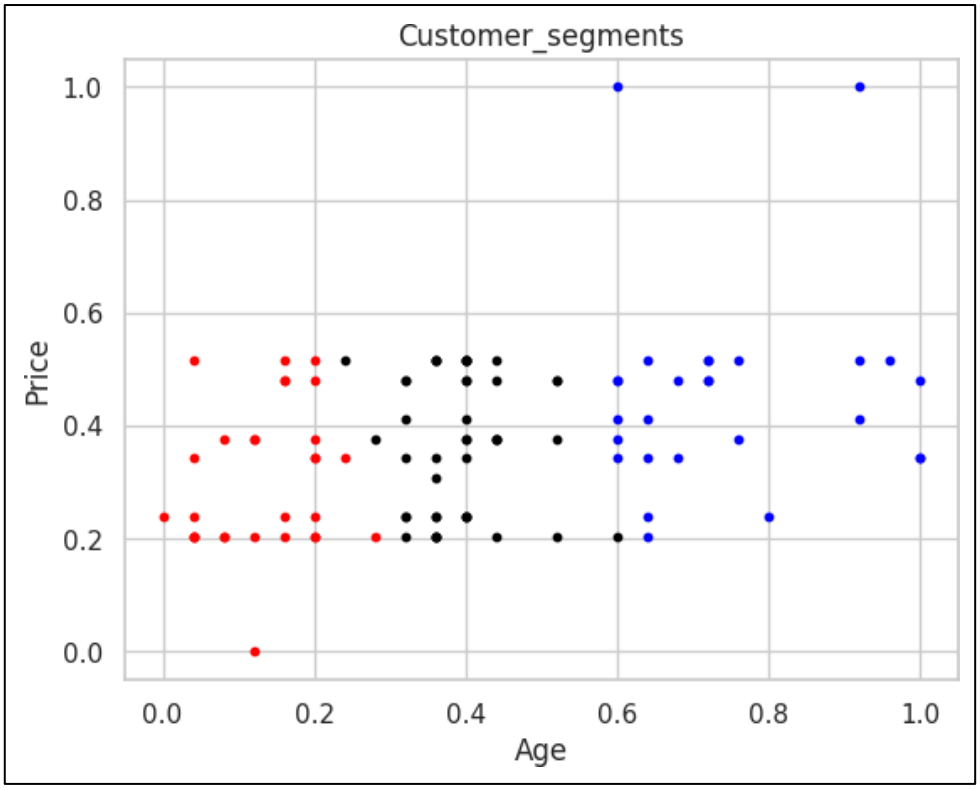
```
k_range = range(1,10)
#sum of squared error
sse = []
for k in k_range:
  km = KMeans(n_clusters=k)
  km.fit(df)
  sse.append(km.inertia_)
plt.title('The Elbow Point Graph')
plt.xlabel('K')
plt.ylabel('sse')
plt.plot(k_range,sse)
```

The elbow graph plotted as:



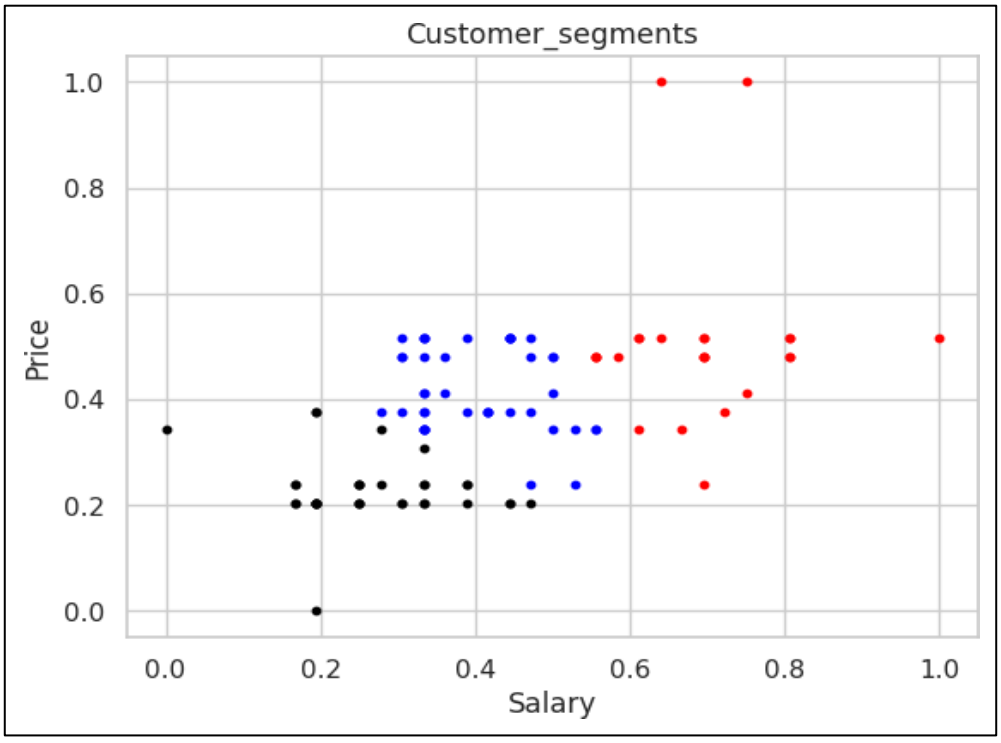
We get the number of clusters as 3. The clustering of age vs price is as follows. As we can see that as age increases the price of the car also increases. This can be justified as older people tend to be in high-paying jobs.

```
clus = df.loc[:,["Age","Price"]].values
kmeans = KMeans(n_clusters=3, init='k-means++')
Y = kmeans.fit_predict(clus)
plotseg(clus, Y, ["Age","Price"])
```



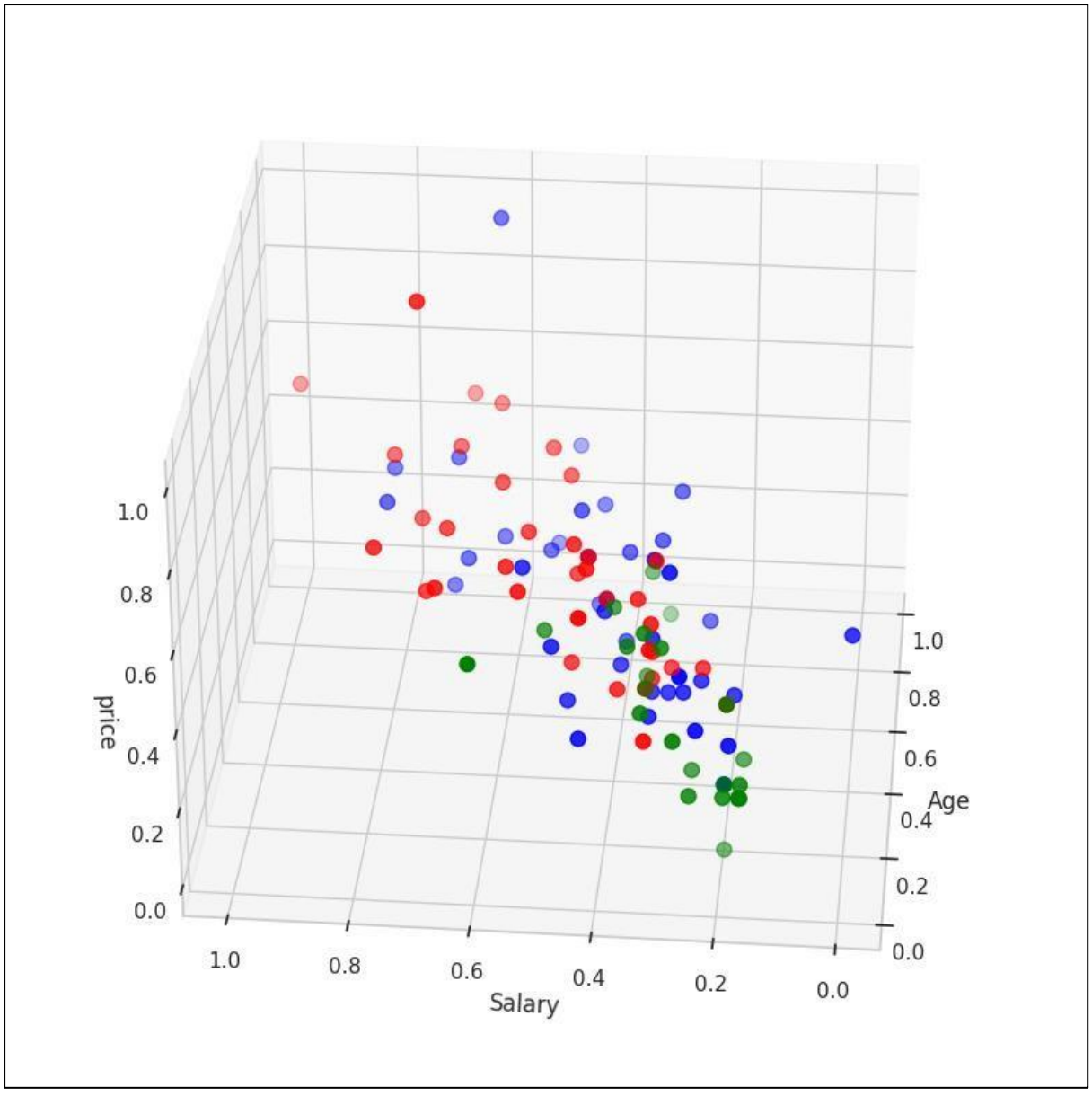
Next, we compare salary and price, as we can guess the higher the salary the higher the price of the car is.

```
clus = df.loc[:,["Salary","Price"]].values
kmeans = KMeans(n_clusters=3, init='k-means++')
Y = kmeans.fit_predict(clus)
plotseg(clus, Y, ["Salary","Price"])
```



Observations:

i. We found that 3 cluster groups can be formed from the data given (based on the price of the vehicle) using the K-Means algorithm and Elbow Point Method.
ii. While looking at the patterns, we find that as the Age increases the cost of the vehicle also rises.
iii. Also, the amount spent on the car goes up with the number of dependents. The same is true for the salary field too.
iv. The visualizations provided gives a clear idea about the patterns. Next, we have 3D plots to see how variables influence each other when they are not taken out of context. The 3D plot is between age, salary, and price.



7. Marketing Mix:

Designing the product is not successful without understanding the consumers need and putting the product in the right place in the consumer's mind. It is important to differentiate your product to ensure continuous growth. This approach will enable consumers to differentiate their products from those of their competitors. This approach will convince the consumers to save on their expenses, as well as ensure a clean and healthy environment. People start to think to purchase eco-friendly vehicles.

Objective:

(a) During first year we are aiming to capture a portion of 5% from the market.

(b) Attract new investor to increase the financial resources and decrease the threats.

(c) Introduce different lines with different feature to attract new customers.

(A) Product Strategy

- Our purpose is manufacturing premium eco-friendly & battery powered electric cars. To save the climate. In addition, their impact on the economy is incredible. Subsequently, a move to target them with this product will be useful to the economy as well as to the whole environment.
- EV would allow you to charge batteries either through plugging it directly into a wall outlet, which could take around 12 hours, or use one of the fast-charging stations that gets the job done in about 20-30 minutes.

(B) Pricing Strategy:

- Expansion of the guarantee period and the provision of lubrication oils can be competitive pricing strategies. In any case, the mental pricing approach remains suitable for middle-income workers since they always need to minimize cost in order to contribute in other areas. Such prices are within reach of the target consumers.
- Informing our interested customers with the latest offers & sales quote (send them Important information about our new updated features, interesting articles about what is going on with the amazing technology.

(C) Promotion and Advertising Strategy:

- **Educational Campaigns:** Raise awareness about EVs and dispel myths through informative campaigns that emphasize benefits such as environmental impact and cost savings.
- **Social Media:** Utilize platforms like Facebook, Instagram, and Twitter to engage with potential customers and share updates on new models and features.
- **Content Marketing:** Create blog posts, videos, and articles that address common questions about EVs, charging infrastructure, and technological advancements.

- **Partnerships:** Collaborate with influencers, environmental organizations, and technology companies to increase exposure and credibility.

(D) Place Strategy:

- **Charging Infrastructure:** Choose strategic locations for dealerships and showrooms near areas with well-developed charging networks.
- **Online Presence:** Enhance the online buying experience with virtual showrooms, configurators, and online purchase options.
- **Charging Stations:** Collaborate with partners to expand the charging infrastructure and establish charging points at key locations.

8. Potential Customer Base:

The potential customer base for electric vehicle (EV) segmentation is diverse and can include various consumer and business segments. As EV technology continues to evolve and becomes more mainstream, the following are some key customer groups that could be interested in adopting electric vehicles:

1. Environmentally Conscious Consumers:
    - Individuals who prioritize reducing their carbon footprint and want to contribute to environmental sustainability.
2. Early Adopters and Tech Enthusiasts:
    - Individuals who are excited about new technology and innovations.
    - Customers who want to experience the latest advancements in automotive technology, including autonomous driving features and connectivity options.
3. Urban Dwellers:
    - People living in urban areas with shorter commuting distances, making EVs a practical and cost-effective choice for their daily transportation needs.
4. Fleet Operators and Businesses:
    - Commercial and corporate fleets looking to reduce operating costs and demonstrate corporate social responsibility.
    - Businesses in industries such as delivery services, ride-sharing, and last-mile logistics, where EVs can offer operational savings and environmental benefits.
5. Cost-Conscious Consumers:
6. **Luxury Car Enthusiasts:**
    - Consumers interested in high-end electric luxury vehicles that offer cutting- edge technology, performance, and premium features.

7. Rural and Suburban Residents:
    - EVs can be a viable option for those with lower daily mileage and access to charging infrastructure.
8. Government and Municipalities:
    - Governments and local authorities aiming to reduce urban pollution and improve air quality through fleet electrification.
    - Municipalities looking to invest in EV charging infrastructure to promote sustainable transportation options.
9. Seniors and Retirees:
    - Seniors who may have shorter daily commutes and find EVs easier to handle due to simplified controls.
10. Second Car Owners:
    - Households with multiple vehicles that can benefit from having an electric car for short trips and daily commuting.
    - Customers who want to offset the emissions of their primary gasoline vehicle.

The potential customer base for electric vehicles is continuously expanding as technology improves, charging infrastructure grows, and consumer awareness increases. Effective marketing strategies should take into consideration the unique needs, preferences, and motivations of each of these customer segments.

9. Conclusion:

(A) Geographic:

With the analysis done on dataset 2, it can be concluded that **Uttar Pradesh and Maharashtra** are perfect locations to launch electronic vehicle startups considering the population and salaried employee ratio.

(B) Demographic:

From dataset 2, we can also conclude that people between **the age group 25 to 50** have a high density of purchasing electronic vehicles. Since **most of them are salaried earning** near about 30 lakhs, **they can easily spend anywhere between 10-20 lakhs on cars** considering their wives are also employed.

(C) Behavioural:

- **Social Media Analytics**: Analyze discussions, comments, and sentiment on social media platforms related to EVs.
- **Consumer Behavior Studies**: Research studies on consumer attitudes, behaviors, and perceptions toward EVs.

- **Early Adopter Communities:** Engage with communities of tech enthusiasts and early adopters to understand their preferences.

(D) Target Segment:

As the trend suggests, higher salaried, old people tend to buy cars in the range of 10-20L. However, the sharp rise in awareness in younger segments about climate change influences their decision to buy electric cars. So, the recommended segment is the mid-tier, with significant marketing about the pros of electric vehicles to the environment required.

10. Link to the project:

Dataset 1:

https://github.com/SayantanIITBombay/EV-market-segmentation/blob/7c60ded22ff993d80fe6c3859ef0f6f093477c80/Electric%20Car%20Data.csv

Dataset 2:

https://github.com/SayantanIITBombay/EV-market-segmentation/blob/7c60ded22ff993d80fe6c3859ef0f6f093477c80/Indian%20automoble%20buying%20behavour%20study%201.0.csv

GitHub Link:
https://github.com/SayantanIITBombay/EV-market-segmentation.git