EV Market Segmentation

November 22nd , 2024

Sharan P

Abstract

India's electric vehicle market presents a compelling opportunity for new entrants, driven by increasing environmental consciousness, rising fuel costs, and supportive government policies. As a startup planning to enter this space, the core challenge lies in identifying and targeting the most promising market segments through data-driven analysis. The task requires examining multiple segmentation dimensions - geographic (considering factors like charging infrastructure and urban density), demographic (income levels, age groups), psychographic (environmental consciousness, early technology adoption tendencies), and behavioral (daily commute patterns, vehicle usage). The goal is to identify early adopters who align with the Innovation Adoption Life Cycle, particularly focusing on innovators and early adopters who will drive initial market traction. The Indian automotive landscape is undergoing a significant transformation, with major cities experiencing severe air quality concerns and growing environmental awareness among urban populations. This shift in consciousness, coupled with government initiatives like FAME II and state-level EV policies, creates a fertile ground for innovative EV solutions. The rising middle class, with its increasing purchasing power and tech-savvy nature, represents a particularly promising demographic for EV adoption.

Key strategic decisions include:
- Optimal geographic entry point based on infrastructure readiness and target demographic concentration
- Product positioning and pricing strategy aligned with identified segment needs
- Marketing approach that resonates with early adopter psychographics

The challenge is compounded by potential data limitations, requiring creative approaches to market research and decision-making. Success will depend on balancing ambitious innovation with practical market realities to establish a sustainable presence in India's evolving EV landscape. Furthermore, the competitive landscape is rapidly evolving, with both domestic and international players vying for market share. Understanding the competitive dynamics and identifying underserved segments will be crucial for carving out a defensible market position. The startup must also consider the entire ecosystem, including charging infrastructure partners, battery suppliers, and service networks, to build a comprehensive go-to-market strategy.

The timing of market entry is particularly crucial, as India stands at an inflection point in EV adoption. Early movers have the opportunity to shape consumer preferences and establish brand loyalty, while benefiting from various government incentives and subsidies designed to accelerate EV adoption. However, this must be balanced against the risks of entering a market that is still in its nascent stages of development.

1. Problem Statement

Our electric vehicle startup needs to identify and target optimal market segments in India through comprehensive data analysis. The challenge involves analyzing geographic factors (like charging infrastructure), demographics (income, age), psychographics (environmental awareness, tech adoption), and behavioral patterns (commuting habits) to determine the most viable entry strategy. The team must develop a data-driven approach to select target locations aligned with the Innovation Adoption Life Cycle, define strategic pricing, and create effective marketing strategies despite potential data limitations in the Indian market.

2. Data Collection

Data is collected from these three major sources.
2.1 http://www.yocharge.com - A site for EV Charging Management Software that offers a way to launch, operate EV Charging Stations. The site has been scraped using libraries such as Beautiful Soup.
2.2 data.gov.in - Data.gov.in is India's national open data portal, providing government datasets across sectors like agriculture, health, transportation, and infrastructure
2.3 https://www.kaggle.com/datasets - A site for open datasets.

Each column are explained in the following datasets:

Dataset 1 - Number of charging stations sanctioned in respective states

1. **State/UT**
2. **Number of Electric vehicle charging stations sanctioned**

This dataset shows the distribution of electric vehicle charging infrastructure across different states and union territories in India. The first column lists all states/UTs, while the second column indicates the number of EV chargers that have been officially approved/sanctioned for installation in each region. This data is crucial for:

1. Understanding charging infrastructure readiness
2. Identifying potential market entry points
3. Assessing government support by region
4. Correlating with EV adoption potential

We can use this to determine which states have stronger EV ecosystem development and higher potential for early market success.

Dataset 2 - Number of electric vehicles in each state

2WN: Two-Wheeler Non-Transport
2WT: Two-Wheeler Transport
2WIC: Two-Wheeler Invalid Carriage (Special Use)
3WN: Three-Wheeler Non-Transport
3WT: Three-Wheeler Transport
LMV: Light Motor Vehicle
LPV: Light Passenger Vehicle
LGV: Light Goods Vehicle
4WIC: Four-Wheeler Invalid Carriage (Special Use)
MMV: Medium Motor Vehicle
MPV: Medium Passenger Vehicle
MGV: Medium Goods Vehicle

This table categorizes different types of vehicles based on their purpose and capacity. The distinctions include personal or commercial use (e.g., "Non-Transport" vs. "Transport"), size (e.g., "Light," "Medium," or "Heavy"), and specific use cases like "Invalid Carriage" for specialized vehicles. This segmentation helps in analyzing transportation trends, vehicle usage patterns, and market demands.

Dataset 3 - Electric Vehicles brand Info

CompanyName: Name of the vehicle manufacturer
Brand: Sub-brand under the company
EV ModelName: Specific model name of the electric vehicle
VehicleType: Type of vehicle (e.g., two-wheeler, car)
VehicleCMVRCategory: Category based on Central Motor Vehicle Rules (CMVR)
IncentiveAmountInINR: Government incentive offered in INR
Range: Distance the EV can travel on a single charge (in km)
Speed: Top speed of the EV (in km/h)
Acceleration: Time taken to reach a certain speed (e.g., 0–60 km/h)
Warranty: Warranty period offered by the manufacturer
EnergyConsumption: Energy usage efficiency (kWh/km or similar)
BatteryTechnology: Type of battery used (e.g., Lithium-ion)
BatteryCapacity: Capacity of the battery (in kWh)
BatteryCycles: Number of charge-discharge cycles the battery can sustain

The dataset provides a detailed overview of electric vehicle (EV) specifications and manufacturer details, which are crucial for market segmentation. **CompanyName** and **Brand** help identify key players in the industry and their positioning. **EV ModelName** and **VehicleType** offer insights into the diversity of vehicles being produced, catering to different customer needs. **VehicleCMVRCategory** further classifies these vehicles based on regulatory standards, making it easier to align offerings with specific market segments. Attributes like **Range**, **Speed**, and **Acceleration** reflect performance metrics, helping to target customers with preferences for high efficiency, speed, or practicality. Financial factors, including **IncentiveAmountInINR**, are critical for understanding pricing dynamics and affordability for different demographics. **Warranty** enhances the appeal of EVs to cautious adopters, while **EnergyConsumption**, **BatteryTechnology**, **BatteryCapacity**, and **BatteryCycles** provide a glimpse into operational costs, technological advancement, and durability. Together, these parameters enable segmentation by geographic, demographic, and psychographic factors, allowing the identification of target markets and crafting tailored strategies for effective market penetration.

Data Preprocessing

1. **Removed Null Values**: Cleaned the dataset by eliminating rows or columns containing null values to ensure consistency and avoid errors during analysis.
2. **Standardized Numerical Values**: Adjusted numerical attributes to their closest standard values (e.g., rounded ranges, speeds) for uniformity across records.
3. **Encoded Categorical Variables**: Converted textual data like VehicleType or Brand into numerical representations using techniques such as one-hot encoding or label encoding for model compatibility.
4. **Handled Outliers**: Identified and addressed outliers in columns like IncentiveAmountInINR, Range, or BatteryCapacity to prevent skewed results during analysis.
5. **Normalized Data**: Scaled numerical features (e.g., Speed, Acceleration) to a consistent range (e.g., 0 to 1) to eliminate bias from varying units of measurement.
6. **Removed Duplicates**: Identified and eliminated duplicate records to avoid redundancy and ensure accurate analysis.
7. **Feature Selection**: Retained only relevant columns (e.g., VehicleCMVRCategory, EnergyConsumption) to reduce dimensionality and focus on key aspects of the segmentation analysis.

3. Exploratory Data Analysis

3.1 Importing all the libraries

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sb
import plotly.express as px
from sklearn.preprocessing import StandardScaler,PowerTransformer
from sklearn.decomposition import PCA
from scipy.cluster.hierarchy import dendrogram, linkage
from sklearn.cluster import KMeans, MeanShift, estimate_bandwidth
from sklearn.datasets import make_blobs
from yellowbrick.cluster import KElbowVisualizer, SilhouetteVisualizer, InterclusterDistance
from collections import Counter
from sklearn.model_selection import cross_validate,train_test_split
from sklearn import metrics
import plotly.io as pio
import io
```

3.2 Loading the datasets

```
from google.colab import files
uploaded = files.upload()
```

```
Choose Files No file chosen          Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving EV_ChargingStations.csv to EV_ChargingStations (1).csv
Saving EV_India.csv to EV_India (1).csv
Saving EV_VehiclesData.csv to EV_VehiclesData (1).csv
```

```
df1 = pd.read_csv(io.BytesIO(uploaded['EV_ChargingStations (1).csv']))
df2 = pd.read_csv(io.BytesIO(uploaded['EV_India (1).csv']))
df3 = pd.read_csv(io.BytesIO(uploaded['EV_VehiclesData (1).csv']))
```

3.3 Retrieving the head

```
[ ] df1.head()
```

	State/UT	No. of EV Chargers Sanctioned
0	Maharashtra	317
1	Andhra Pradesh	266
2	Tamil Nadu	281
3	Gujarat	278
4	Uttar Pradesh	207

	CompanyName	Brand	EV ModelName	Vehicle Type& Segment	VehicleCMVRCategory	IncentiveAmountInINR	Range (Km)	Max. SpeedKm/Hr)	Acceleration(m/s2	WarrantyInYears)	ElectricEnergyconsumptionKWh per100KM	BatteryTechnology
0	AltigreenPropulsionLabs Pvt Ltd	Altigreen	NEEV	ThreeWheeler (e-3W	L5N	75000	117.0	53.4	0.65	3.0	8.10	Lithium ionLiFeP04(Lithium Ironphosphate)
1	AltigreenPropulsionLabs Pvt Ltd	Altigreen	NEEV HD	ThreeWheeler (e-3W	L5N	90000	151.0	53.7	0.99	3.0	8.70	Li ion batterybased onLiFePO4(Lithium IonPhosp...
2	AltigreenPropulsionLabs Pvt Ltd	Altigreen	NEEV LR	ThreeWheeler (e-3W	L5N	85000	151.0	53.7	0.99	3.0	8.70	Li ion batterybased onLiFePO4(Lithium IonPhosp...
3	AltigreenPropulsionLabs Pvt Ltd	Altigreen	NEEV HDx	ThreeWheeler (e-3W	L5N	92000	160.0	54.0	0.95	3.0	8.00	Li ion batterybased onLiFePO4
4	AmpereVehiclesPrivateLimited	Ampere	ZEAL	Two Wheeler(e-2W	L1	18000	108.0	41.6	0.65	3.0	2.26	LI ion NCM
5	AmpereVehiclesPrivateLimited	Ampere	Magnus	Two Wheeler(e-2W	L1	18000	90.0	48.0	0.65	3.0	2.50	Lithium ion
6	AmpereVehiclesPrivateLimited	Ampere	Zeal VX1	Two Wheeler(e-2W	L1	19600	84.0	41.6	0.65	3.0	2.47	Lithium ion
7	AmpereVehiclesPrivateLimited	Ampere	ZEAL CA	Two Wheeler(e-2W	L1	18000	90.0	42.0	0.65	1.0	2.52	Lithium NickelManganeseCobalt Oxide
8	AmpereVehiclesPrivateLimited	Ampere	ZEAL EX	Two Wheeler(e-2W	L1	34500	124.0	44.3	0.99	3.0	2.60	Lithium NickelManganeseCobalt Oxide
9	AmpereVehiclesPrivateLimited	Ampere	MAGNUS EX	Two Wheeler(e-2W	L1	34500	120.0	46.4	0.86	3.0	2.70	Lithium NickelManganeseCobalt oxide
10	AmpereVehiclesPrivateLimited	Ampere	ZEAL CA EX	Two Wheeler(e-2W	L1	34500	94.0	44.3	0.99	3.0	3.00	Lithium NickelManganeseCobalt oxide

```
df3.head()
```

	S.No.	State Name	2WN	2WT	2WIC	3WN	3WT	LMV	LPV	LGV	4WIC	MMV	MPV	MGV	HPV	HGV	OTH	Grand Total
0	1	Andaman and Nicobar Island	2	5.0	NaN	NaN	30.0	86	6.0	NaN	NaN	NaN	NaN	NaN	40.0	NaN	NaN	169
1	2	Andhra Pradesh	27629	NaN	2.0	374.0	108.0	1050	3.0	166.0	NaN	NaN	NaN	NaN	NaN	NaN	1117.0	30449
2	3	Arunachal Pradesh	14	NaN	NaN	NaN	NaN	6	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	21
3	4	Assam	2287	NaN	NaN	NaN	79891.0	233	5.0	15.0	NaN	NaN	NaN	NaN	15.0	NaN	NaN	82216
4	5	Bihar	13472	NaN	NaN	2.0	96560.0	231	8.0	21.0	1.0	NaN	NaN	1.0	27.0	2.0	NaN	110325

3.4 Describing the datasets

```
[ ] df1.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26 entries, 0 to 25
Data columns (total 2 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   State/UT                     26 non-null     object
 1   No. of EV Chargers Sanctioned 26 non-null    int64
dtypes: int64(1), object(1)
memory usage: 544.0+ bytes
```

```
df2.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 160 entries, 0 to 159
Data columns (total 15 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   CompanyName                             160 non-null    object
 1   Brand                                   160 non-null    object
 2   EV ModelName                            160 non-null    object
 3   Vehicle Type& Segment                   160 non-null    object
 4   VehicleCMVRCategory                     160 non-null    object
 5   IncentiveAmountInINR                    160 non-null    int64
 6   Range (Km)                              158 non-null    float64
 7   Max. SpeedKm/Hr)                        158 non-null    float64
 8   Acceleration(m/s2                       158 non-null    float64
 9   WarrantyInYears)                        158 non-null    float64
 10  ElectricEnergyconsumptionKWh per100KM   158 non-null    float64
 11  BatteryTechnology                       158 non-null    object
 12  BatteryCapacity(kWh)                    158 non-null    float64
 13  BatteryDensityWh/Kg)                    158 non-null    float64
 14  Battery cycleNo. ofCycles)              158 non-null    float64
dtypes: float64(8), int64(1), object(6)
memory usage: 18.9+ KB
```

```
df3.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 18 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   S.No.        35 non-null     object
 1   State Name   35 non-null     object
 2   2WN          35 non-null     int64
 3   2WT          11 non-null     float64
 4   2WIC         11 non-null     float64
 5   3WN          16 non-null     float64
 6   3WT          30 non-null     float64
 7   LMV          35 non-null     int64
 8   LPV          32 non-null     float64
 9   LGV          26 non-null     float64
 10  4WIC         8 non-null      float64
 11  MMV          3 non-null      float64
 12  MPV          14 non-null     float64
 13  MGV          11 non-null     float64
 14  HPV          25 non-null     float64
 15  HGV          20 non-null     float64
 16  OTH          14 non-null     float64
 17  Grand Total  35 non-null     int64
dtypes: float64(13), int64(3), object(2)
memory usage: 5.0+ KB
```

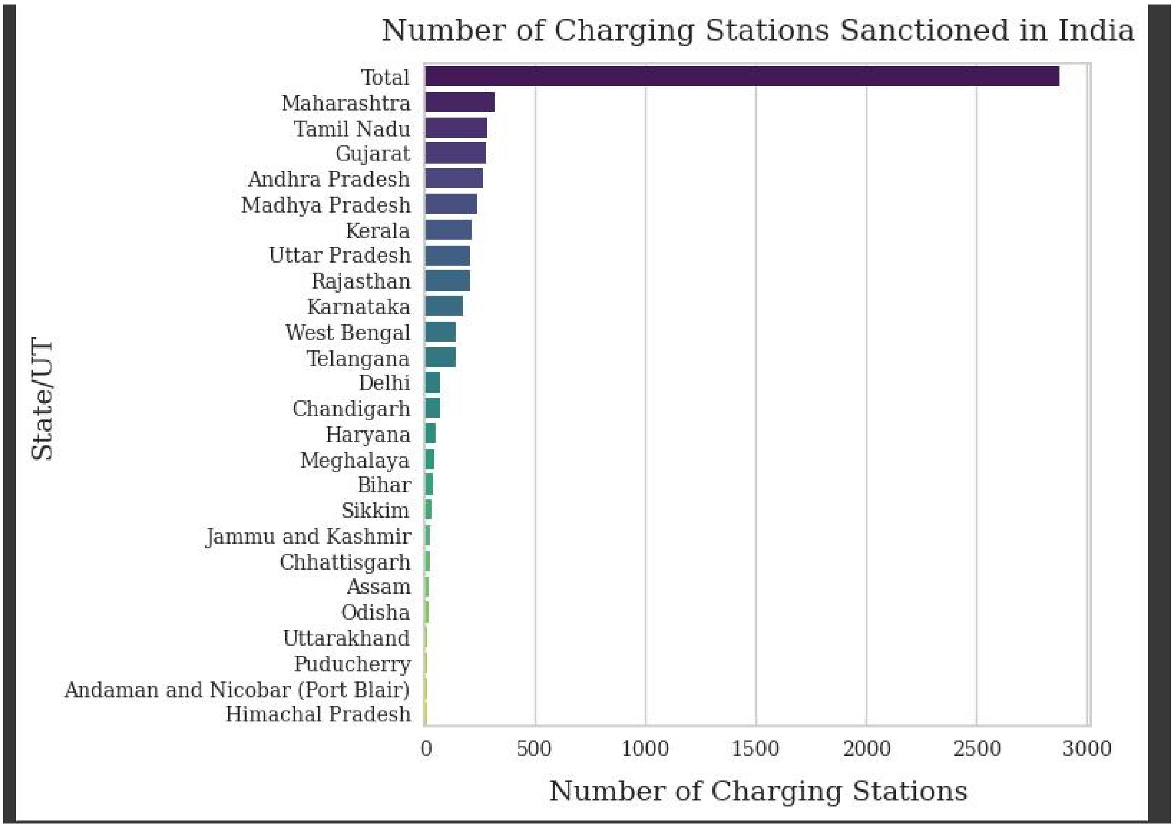
```
[ ] df2.describe()
```

	IncentiveAmountInINR	Range (Km)	Max. SpeedKm/Hr)	Acceleration(m/s2	WarrantyInYears)	ElectricEnergyconsumptionKWh per100KM	BatteryCapacity(kWh)	BatteryDensityWh/Kg)	Battery cycleNo. ofCycles)
count	160.000000	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000
mean	69009.550000	129.048797	38.361076	0.861582	2.987342	6.173418	7.110696	153.116266	1675.329114
std	63889.509182	48.356604	28.648991	2.249225	0.159111	3.377130	6.667228	47.738918	664.145612
min	17000.000000	80.000000	0.000000	0.000000	1.000000	2.260000	1.540000	76.050000	1000.000000
25%	33456.250000	99.000000	0.000000	0.000000	3.000000	3.410000	2.900000	121.000000	1000.000000
50%	41000.000000	115.800000	42.000000	0.650000	3.000000	5.045000	4.400000	141.660000	2000.000000
75%	76000.000000	142.750000	51.982500	1.040000	3.000000	7.975000	7.700000	186.750000	2000.000000
max	302000.000000	314.000000	116.500000	28.000000	3.000000	14.900000	30.200000	269.000000	5000.000000

```
df3.describe()
```

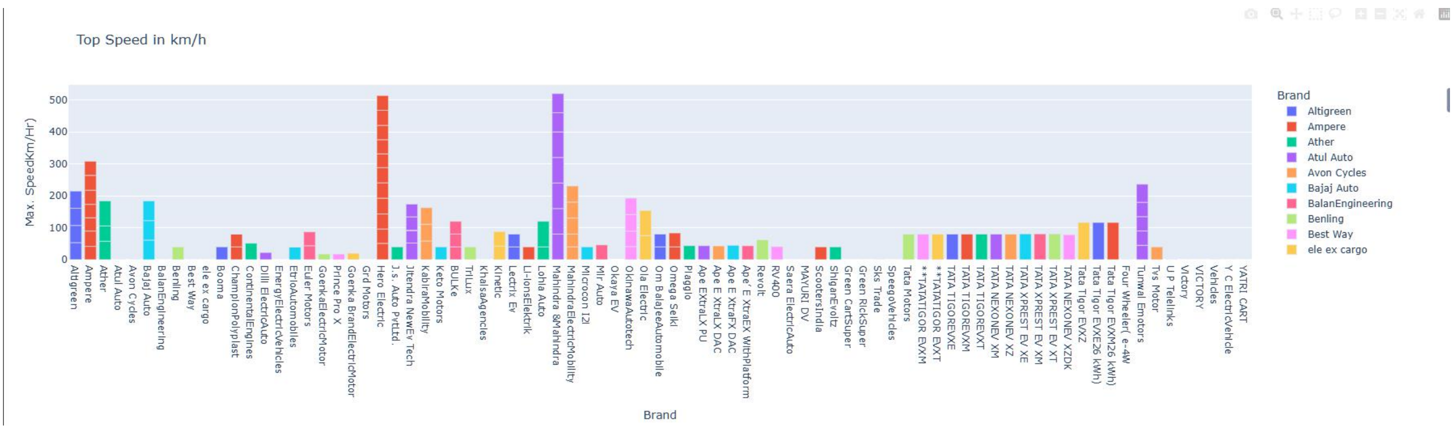
	2WN	2WT	2WIC	3WN	3WT	LMV	LPV	LGV	4WIC	MMV	MPV	MGV	HPV	HGV	OTH	Grand Total
count	35.00000	11.000000	11.000000	16.000000	30.000000	35.000000	32.000000	26.000000	8.000000	3.0	14.000000	11.00000	25.00000	20.000000	14.000000	3.500000e+01
mean	45930.00000	1745.090909	10.545455	93.000000	61226.400000	3070.114286	501.000000	256.846154	4.000000	4.0	60.142857	4.00000	255.76000	10.600000	237.285714	1.030267e+05
std	136815.20672	3832.732014	18.079622	197.159496	177269.244642	9382.845695	1542.957717	710.709150	5.182388	2.0	122.108311	6.21289	653.39385	22.792889	509.640058	3.070064e+05
min	1.00000	1.000000	1.000000	1.000000	1.000000	3.000000	1.000000	1.000000	1.000000	2.0	1.000000	1.00000	1.00000	1.000000	1.000000	2.100000e+01
25%	169.50000	6.000000	2.000000	2.000000	289.750000	43.000000	4.000000	9.500000	1.000000	3.0	1.250000	1.00000	9.00000	2.000000	1.000000	1.171000e+03
50%	8744.00000	20.000000	3.000000	18.500000	6995.000000	233.000000	35.000000	29.500000	1.500000	4.0	2.000000	1.00000	40.00000	5.500000	3.500000	2.348600e+04
75%	35121.00000	61.500000	5.500000	50.250000	30016.250000	1177.500000	93.000000	73.500000	4.500000	5.0	50.500000	4.00000	74.00000	7.750000	116.000000	7.893550e+04
max	803775.00000	9598.000000	58.000000	744.000000	918281.000000	53727.000000	8016.000000	3339.000000	16.000000	6.0	421.000000	22.00000	3197.00000	106.000000	1661.000000	1.802967e+06

3.5 Number of Charging Stations Sanctioned

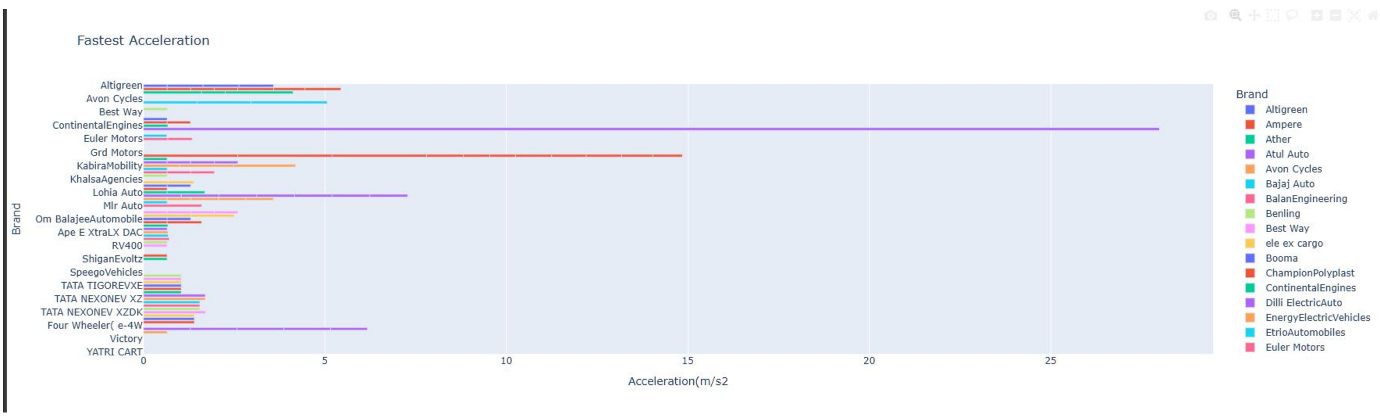


Maharashtra has the maximum number of Charging Stations followed by Tamilnadu and Gujarat.

3.6 Top speed in Km/h

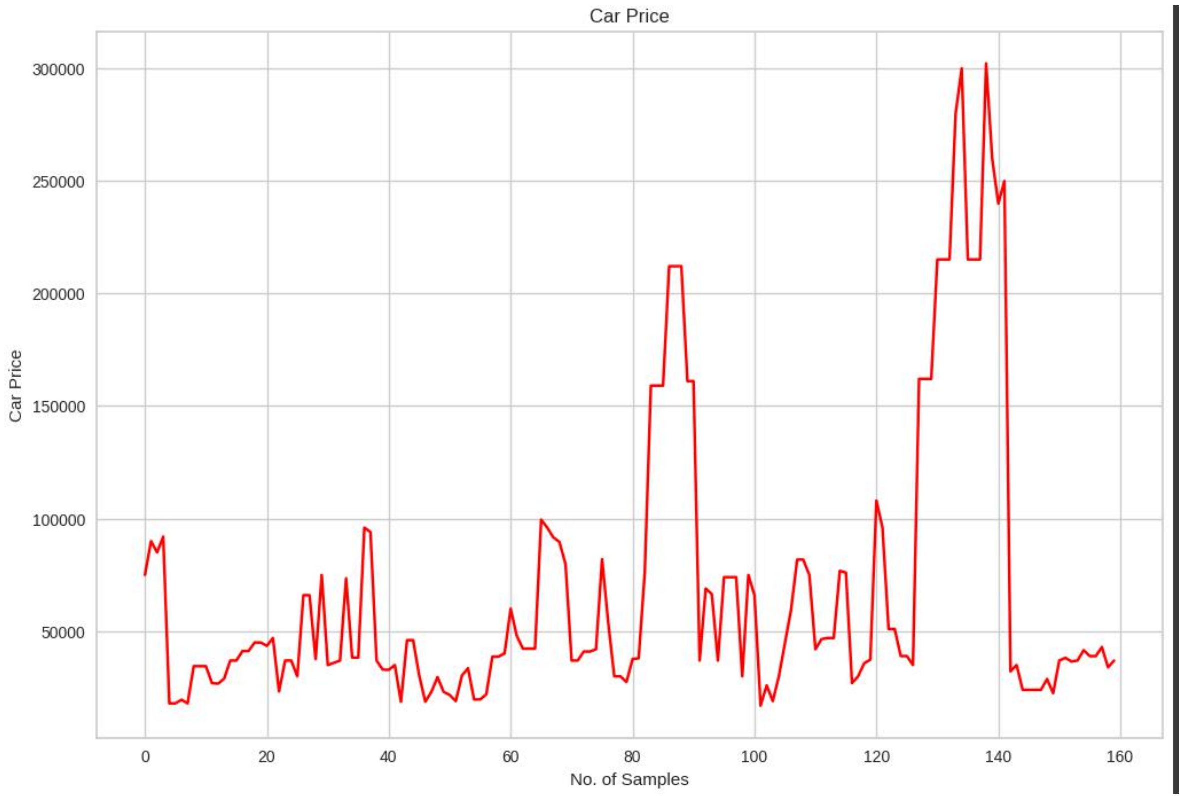


3.7 Top acceleration



3.8 Incentives provided

3.9

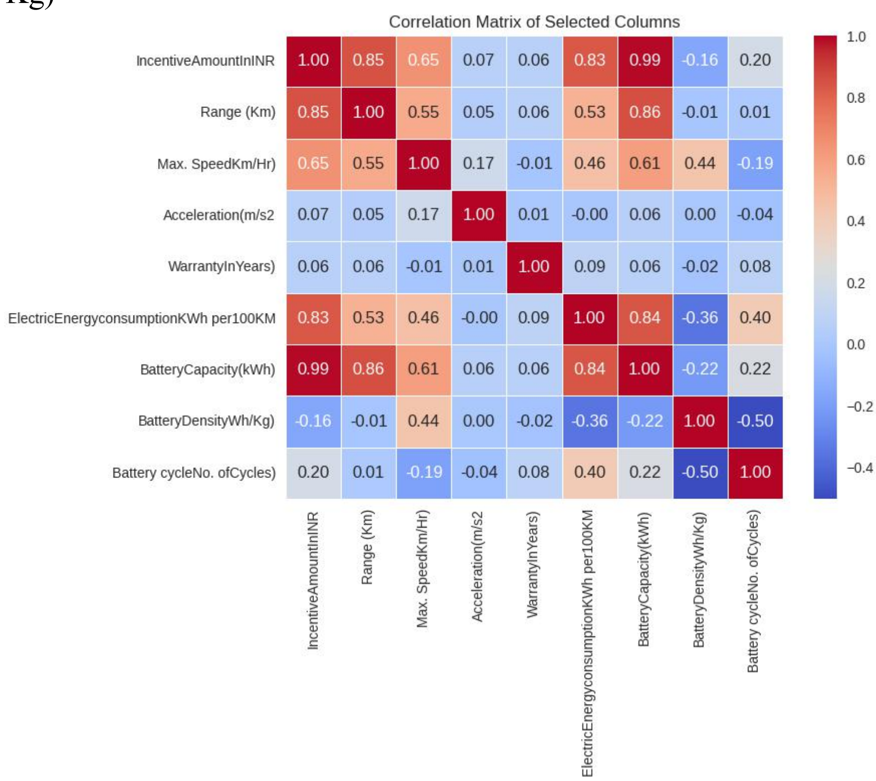


Correlation Matrix

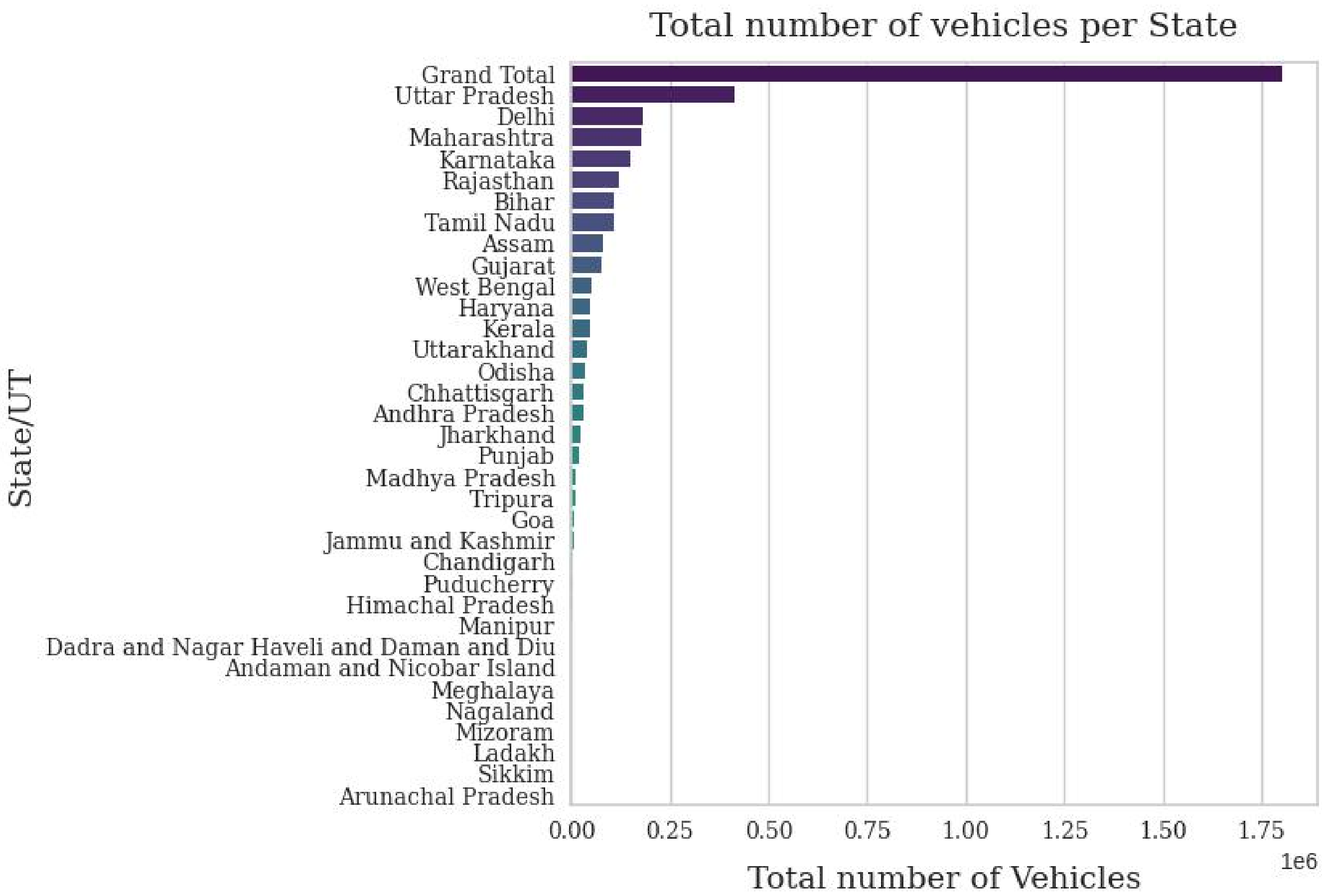
Following columns have been selected for the correlation matrix

IncentiveAmountInINR

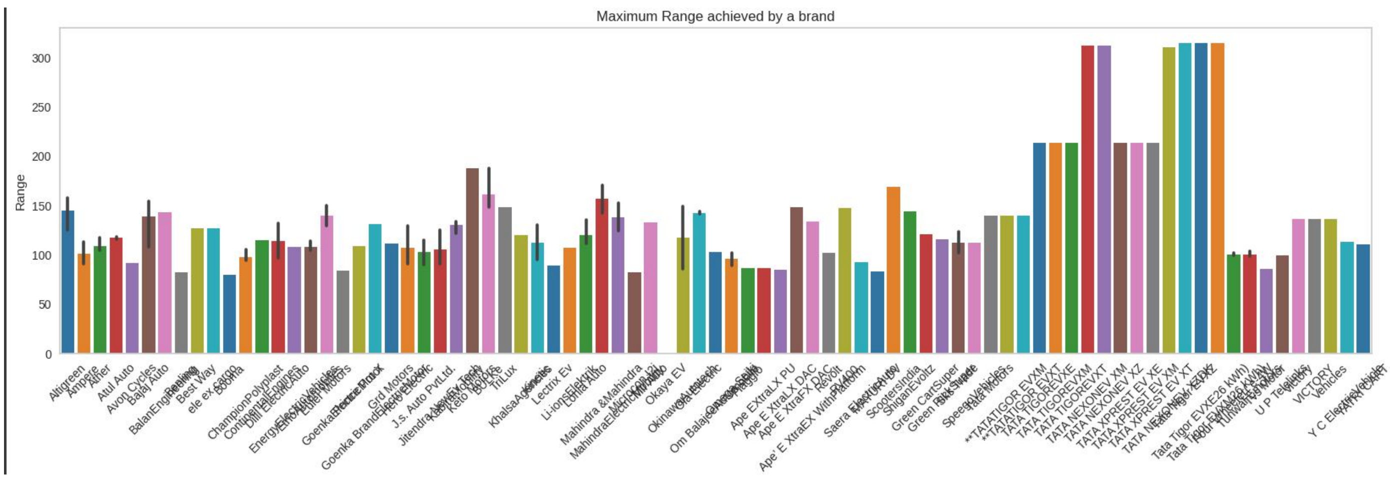
2. Range (Km)
3. Max. Speed (Km/Hr)
4. Acceleration (m/s²)
5. Warranty (In Years)
6. Electric Energy Consumption (KWh per 100 KM)
7. Battery Capacity (kWh)
8. Battery Density (Wh/Kg)
9. Battery Cycles.



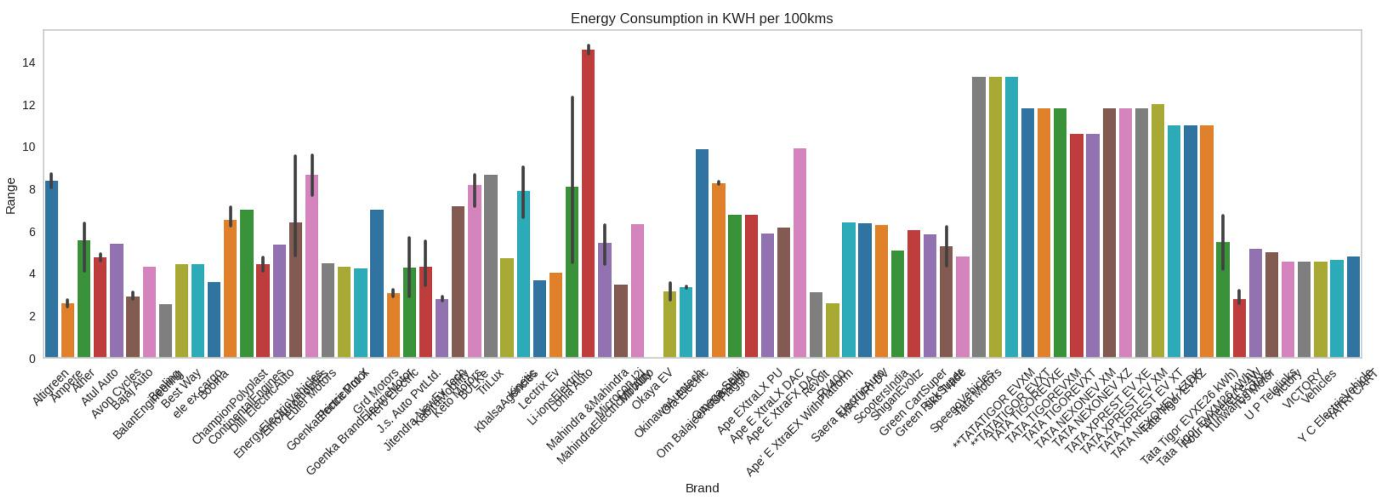
3.10 Total number of Vehicles Per State



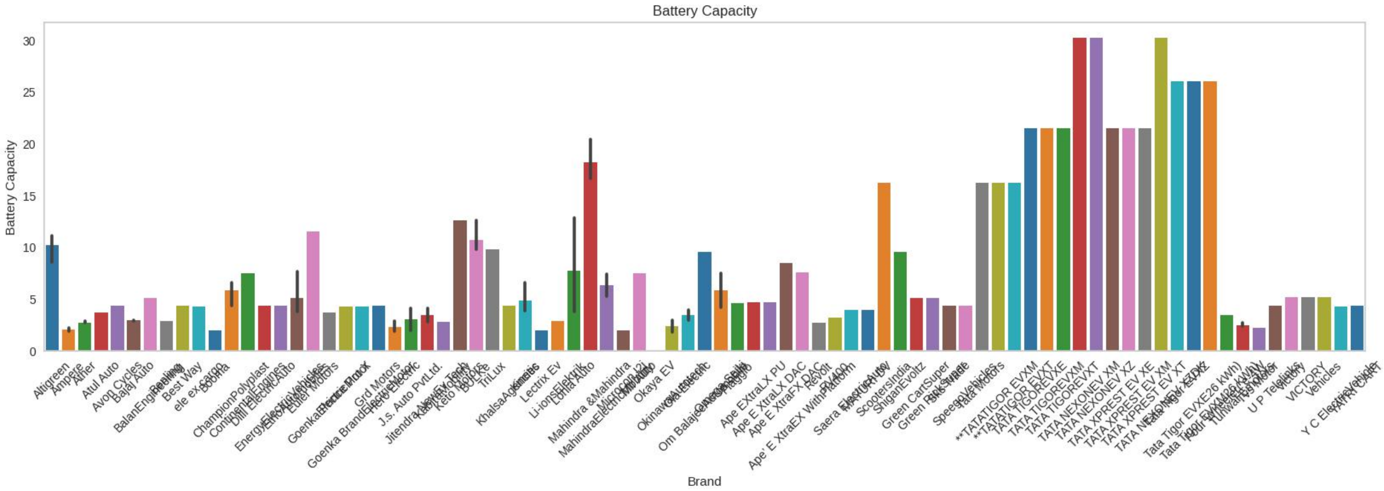
3.11 Maximum range achieved by a brand



3.12 Energy consumption in KWH per 100 kms



3.13 Battery Capacity



Key Takeaways from EDA

1. Maharashtra has the maximum number of charging stations followed by Tamilnadu and Gujarat.
2. Uttar Pradesh has the maximum number of electric vehicles followed by Delhi and Maharashtra.
3. Continental Engines had the fastest acceleration followed by GRD motors and Lohio Auto.
4. Tata had the most number of vehicles that would give the maximum range
5. Lohio Auto had consumd the most energy and almost all the vehicles in tata brand has almost same level of energy consumption.

Based on the exploratory data analysis (EDA) provided, we can draw the following conclusions:

Charging Station Infrastructure
Maharashtra has the highest number of charging stations, followed by Tamil Nadu and Gujarat. This indicates that Maharashtra has the most developed infrastructure for supporting electric vehicles (EVs), making it a promising region for EV adoption. Tamil Nadu and Gujarat also have substantial infrastructure, which positions them as key regions for EV growth.

Electric Vehicle Adoption
Uttar Pradesh leads in the number of electric vehicles, followed by Delhi and Maharashtra. This suggests that Uttar Pradesh is becoming an important market for EV adoption, potentially driven by regional policies, incentives, or growing awareness. Delhi and Maharashtra, being major urban centers, also show strong adoption, making them important markets for EV manufacturers.

Performance and Acceleration
Continental Engines stands out with the fastest acceleration, followed by GRD Motors and Lohio Auto. This highlights Continental Engines as a brand that offers high-performance vehicles, which could appeal to consumers seeking superior speed and acceleration. GRD Motors and Lohio Auto also provide competitive performance, though they may cater to different customer preferences based on other factors like pricing or features.

Range and Tata's Market Advantage
Tata leads with the highest number of vehicles offering the maximum range. This makes Tata a strong competitor in markets where longer-range EVs are in demand, especially in regions with limited charging infrastructure or where long-distance travel is common. Tata's focus on range could be an attractive feature for customers looking for convenience and sustainability in their EVs.

Energy Consumption Efficiency
Lohio Auto consumes the most energy, while almost all Tata vehicles have similar energy consumption levels. Lohio Auto's higher energy consumption could make its vehicles less efficient, potentially leading to

higher operating costs for consumers. In contrast, Tata's consistency in energy consumption across its vehicles suggests greater efficiency, which could appeal to cost-conscious and environmentally aware consumers.

Overall Conclusion
Maharashtra, Tamil Nadu, and Gujarat offer strong infrastructure for EV adoption, with Uttar Pradesh, Delhi, and Maharashtra showing the highest rates of EV adoption. Continental Engines stands out for performance, while Tata's emphasis on long-range and energy efficiency could make it a favorable choice for consumers. Lohio Auto may need to focus on improving energy efficiency to remain competitive in the market.

4. Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique used to simplify the complexity of high-dimensional data while retaining most of the variance (information) present in the data. PCA achieves this by transforming the original variables into a new set of uncorrelated variables called *principal components*.

What is PCA for?

PCA is primarily used for:

- **Dimensionality reduction**: It helps reduce the number of features (or dimensions) while maintaining the integrity of the original data, which is crucial for speeding up machine learning models, especially when the data is very large or has many features.
- **Data visualization**: It allows the reduction of high-dimensional data to 2D or 3D for visualization, which helps in understanding the structure and relationships in the data.
- **Noise reduction**: By focusing on the most important components, PCA can help reduce noise and irrelevant details, thereby improving the performance of machine learning algorithms.
- **Feature extraction**: PCA creates new features (principal components) that are linear combinations of the original features, which can sometimes help improve the performance of predictive models.

The steps involved in performing PCA are:

1. **Standardize the data**: If the features have different units or scales, standardization is done to bring all variables to a common scale (usually with mean = 0 and variance = 1). This step is crucial because PCA is sensitive to the scale of the data.
2. **Compute the covariance matrix**: The covariance matrix represents the relationship between the different features, indicating how much they vary together.
3. **Calculate the eigenvalues and eigenvectors**: Eigenvalues represent the amount of variance captured by each principal component, and eigenvectors represent the direction of these components in the feature space. The eigenvectors are essentially the axes of the new feature space.4.
4. **Sort the eigenvalues**: The eigenvectors are sorted by the corresponding eigenvalues in descending order, meaning that the principal component with the highest variance is selected first.
5. **Choose the top k principal components**: Depending on how much variance you want to retain, you select the top k eigenvectors (components). These k components form the new feature space.
6. **Transform the original data**: The original data is projected onto the selected eigenvectors (principal components), resulting in the reduced dataset.

```
pca = PCA(n_components=9)
t = pca.fit_transform(df_scaled)
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	0.486021	-0.306660	0.003015	0.043200	0.645941	-0.406581	0.204295	-0.134171	0.008010
1	1.279161	-0.443066	0.057678	-0.072371	0.178417	-0.136725	0.233818	-0.274554	0.170793
2	1.240284	-0.444670	0.060755	-0.072183	0.183836	-0.137414	0.238908	-0.317502	0.223132
3	1.279168	-0.370073	0.041148	-0.050762	0.073030	0.086762	0.344584	-0.268892	0.134555
4	-1.607089	0.980899	0.044564	0.208277	-0.406088	-0.099600	0.515110	-0.004774	0.021014
...	...	...	...	...	...	...	...	...	...
155	-0.885761	-1.115223	-0.200433	0.224830	-0.483586	0.755617	-0.281577	-0.064319	0.060796
156	-0.885761	-1.115223	-0.200433	0.224830	-0.483586	0.755617	-0.281577	-0.064319	0.060796
157	-0.854660	-1.113940	-0.202895	0.224679	-0.487922	0.756168	-0.285648	-0.029960	0.018925
158	-1.129486	-1.609444	-0.142905	0.084850	-0.588577	0.272600	0.070909	0.025239	0.022659
159	-1.098123	-1.626347	-0.144605	0.081629	-0.565162	0.215405	0.041708	0.060935	0.003850

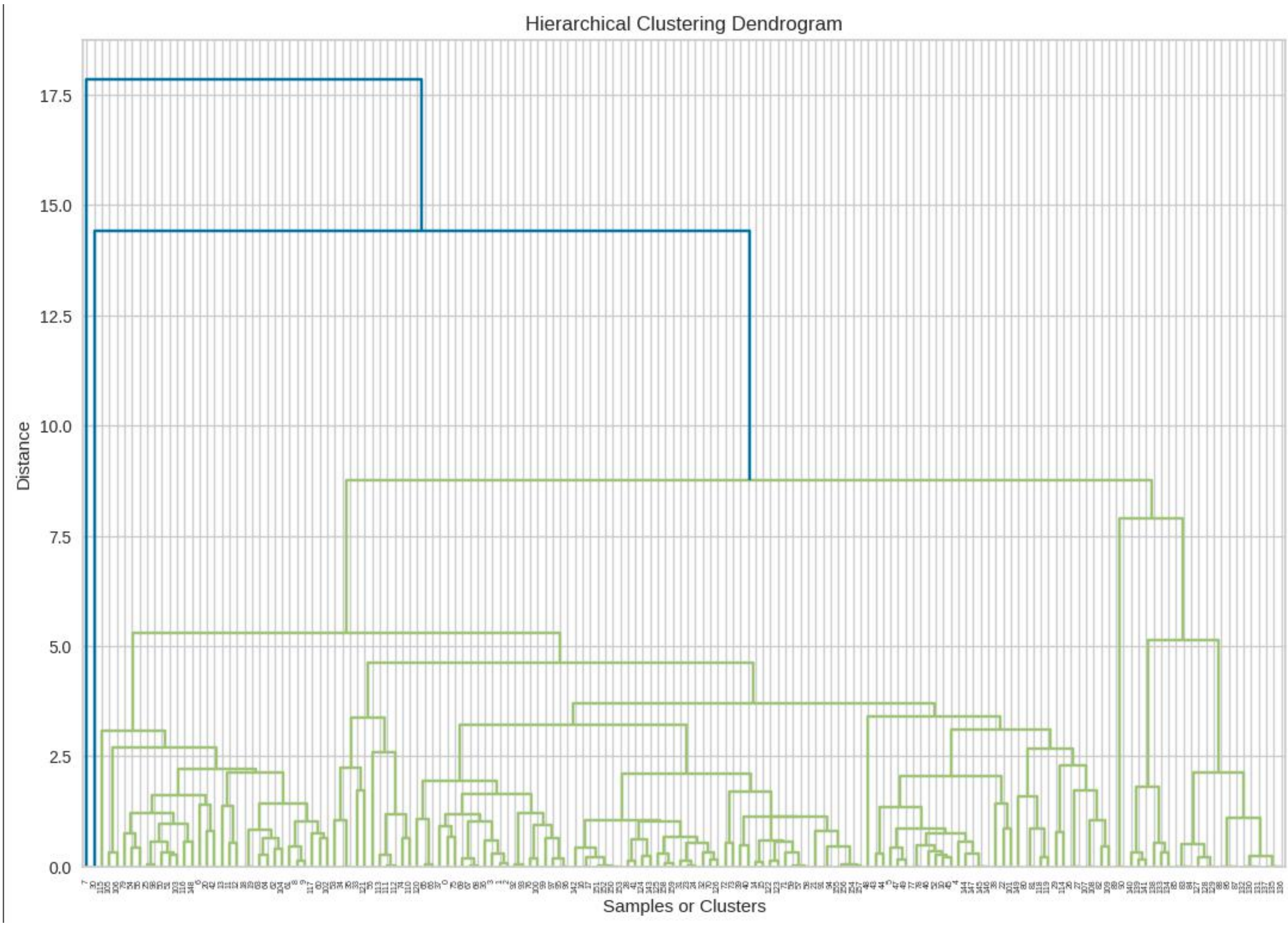
160 rows × 9 columns

The following code generates a DataFrame containing the loadings of each principal component (PC) from PCA. It labels the principal components and associates them with the original feature names, making it easier to interpret which features are most important in each principal component.

```
loadings = pca.components_
num_pc = pca.n_features_in_
pc_list = ["PC" + str(i) for i in list(range(1, num_pc + 1))]
loadings_df = pd.DataFrame.from_dict(dict(zip(pc_list, loadings.T)))
loadings_df['variable'] = df_9.columns.values
loadings_df = loadings_df.set_index('variable')
loadings_df
```

variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
PC1	0.495210	0.425483	0.336087	0.044725	0.046148	0.433653	0.495000	-0.090006	0.124365
PC2	0.020429	0.144745	0.447801	0.119802	-0.083876	-0.202112	-0.018613	0.631495	-0.562794
PC3	-0.039196	-0.055535	0.022850	0.784731	0.609433	-0.039909	-0.041635	-0.034019	0.056232
PC4	-0.002398	0.056998	-0.017397	-0.593177	0.782874	-0.001781	-0.009076	0.170676	-0.049765
PC5	-0.069027	-0.370417	0.382828	-0.028607	-0.046478	0.190816	-0.115209	0.465303	0.665057
PC6	0.008771	0.610156	-0.241574	0.039794	-0.047008	-0.536874	0.033045	0.255681	0.459110
PC7	-0.064825	0.004177	0.675061	-0.115873	0.044590	-0.485158	-0.084681	-0.522527	0.095258
PC8	0.547078	-0.527409	-0.156299	-0.012273	0.012843	-0.456284	0.427233	0.084119	-0.000556
PC9	-0.666685	-0.063274	0.025371	-0.002429	0.003460	-0.035796	0.740776	0.028904	-0.000531

This code is used to perform **Hierarchical Clustering**, a type of unsupervised machine learning algorithm that groups similar data points together into a hierarchy of clusters. The dendrogram provides a visual representation of how the clusters are formed. Hierarchical clustering can be useful in identifying patterns or groups in data when the number of clusters is not known in advance.The dendrogram helps to understand how clusters are merged or split at each step of the clustering process, and by analyzing the distances, you can determine the number of clusters that best fit your data.



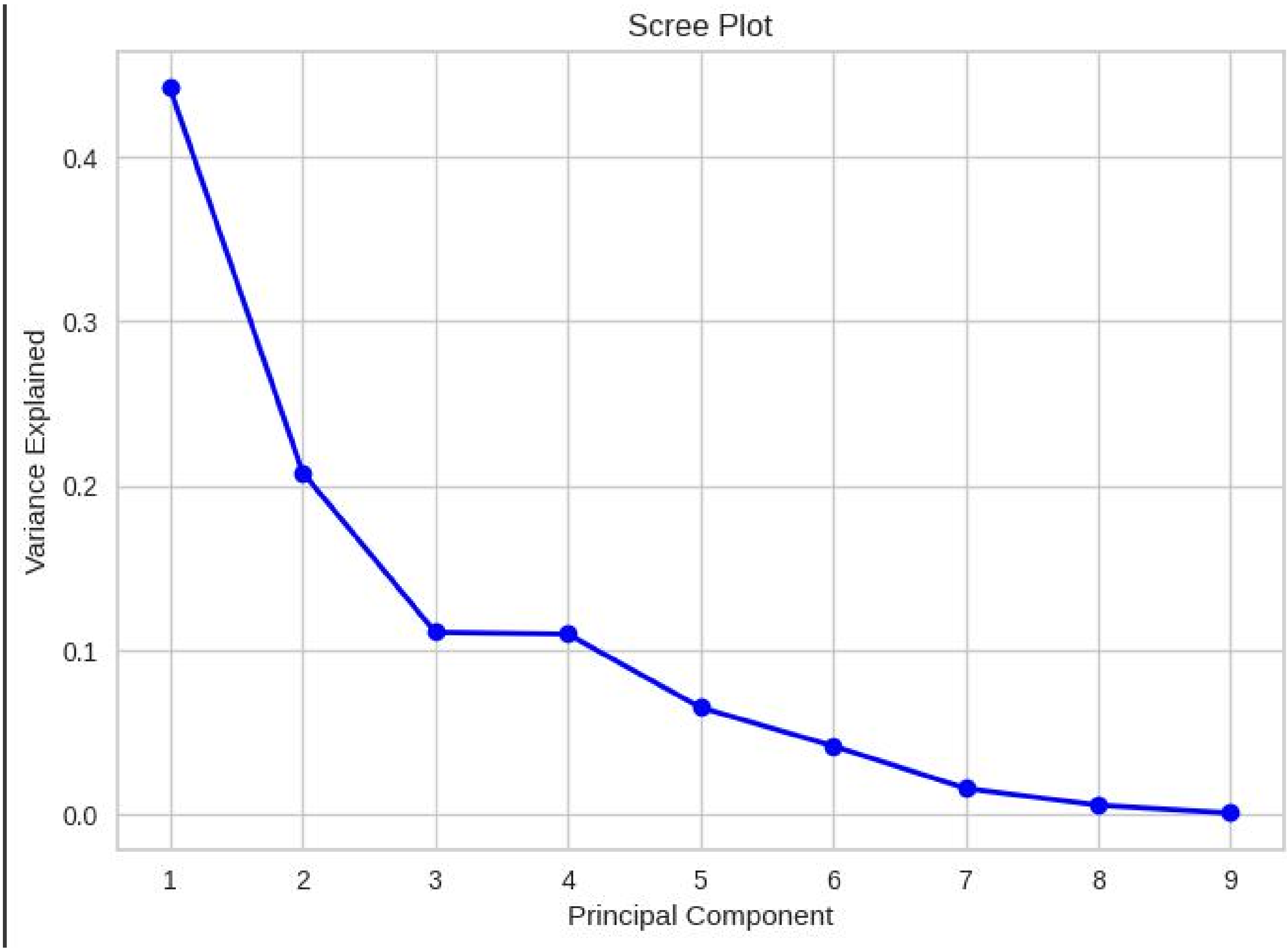
K means Clustering

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. It allows us to cluster the data into different groups and a convenient way to discover thec ategories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

We start by pre-processing the data and cleaning it. This essentially involves null-handling and label encoding the ordinal parameters of the data. The data is then passed into the Scikit-Learn K-Means Clustering model to obtain the elbow curve for the ideal number of clusters. Using the "elbow" or "knee of a curve" as a cutoff point is a common heuristic in mathematical optimization to choose a point where diminishing returns are no longer worth the additional cost. In clustering, this means one should choose a few clusters so that adding another cluster doesn't give much better modeling of the data. The intuition is that increasing the number of clusters will naturallyimprove the fit (explain more of the variation), since there are more parameters (more clusters) to use, but that at some point this is over-fitting, and the elbow reflects this.

Determining the number of Clusters

The Elbow Method involves plotting the **Within-Cluster Sum of Squares (WCSS)** (or **Inertia**) against the number of clusters (K). WCSS represents the total distance between data points and their respective cluster centroids. As the number of clusters increases, the WCSS value will decrease, because the clusters will become smaller and more tightly packed. However, after a certain number of clusters, the decrease in WCSS will start to slow down, forming an "elbow" in the graph. The idea is to choose the number of clusters at this "elbow," where the improvement in WCSS starts to level off.



Based on the Elbow Method, the plot of WCSS (Within-Cluster Sum of Squares) shows a clear flattening of the curve at 3 clusters. This indicates that increasing the number of clusters beyond 3 does not lead to a significant decrease in the WCSS. Therefore, we can conclude that 3 is the optimal number of clusters for the dataset. This choice of 3 clusters strikes a balance between reducing the variance within the clusters and avoiding overfitting, making it the most suitable number of clusters for the given data.

Analysing Market Segments

Geographic Segmentation: Geographic Segmentation: Geographic segmentation divides a target market by location so marketers can better serve customers in a particular area. This type of market segmentation is based on the geographic units themselves (countries, states, cities, etc.), but also on various geographic factors, such as climate, cultural preferences, populations, and more. Geographic segmentation involves segmenting your audience based on the region they live or work in. This can be done in any number of ways: grouping customers by the country they live in, or smaller geographical divisions, from region to city, and right down to postal code. Geographic segmentation might be the simplest form of market segmentation to get your head around, but there are still plenty of ways it can be used that companies never think about. The size of the area you target should change depending on your needs as a business. Generally speaking, the larger the business the bigger the areas you'll be targeting. After all, with a wider potential audience, targeting each postcode individually simply won't be cost-effective.

Based on the insights derived from the previous exploratory data analysis (EDA), the following conclusions can be drawn about geographic segmentation in the Electric Vehicle (EV) market:

Charging Station Density: Maharashtra has the highest number of charging stations, followed by Tamil Nadu and Gujarat. This suggests that these states have better infrastructure for supporting electric vehicles, making them prime candidates for targeting EV customers. The availability of charging stations is a key factor in EV adoption, as it reduces range anxiety and increases the convenience of using EVs.

EV Adoption: Uttar Pradesh leads in the number of electric vehicles, followed by Delhi and Maharashtra. This indicates a higher level of EV adoption in these states, possibly driven by local government initiatives, environmental awareness, and the presence of EV-friendly policies. As these states have the highest EV penetration, they represent critical markets for further expansion, sales, and services related to electric vehicles.

Regional Preferences: The insights about vehicle performance and energy consumption indicate that different regions might have distinct preferences or needs based on their infrastructure and driving conditions. For example, regions with high numbers of charging stations and EVs like Maharashtra, Tamil Nadu, and Gujarat may prioritize models with longer range and lower energy consumption, whereas regions with fewer charging stations may focus more on affordable models that can support charging infrastructure development.

Targeting Early Adopters: States with better EV infrastructure (Maharashtra, Tamil Nadu, Gujarat) are likely to be early adopters of new EV technology. These regions can be targeted with premium EV models offering advanced features such as longer range, faster acceleration, and greater energy efficiency, appealing to tech-savvy consumers who are more willing to invest in new technologies.

EV Market Penetration Strategy: The insights from the geographic segmentation suggest that for states with fewer charging stations and lower EV adoption (e.g., Uttar Pradesh), an educational push about the benefits of EVs, alongside government incentives for setting up charging stations, may be essential to boost adoption. Similarly, these regions could benefit from more affordable EV options or models with shorter ranges suited for urban commuting, where charging stations are gradually being set up.

Behavioral Segmentation:

Usage Frequency: Different regions and consumers exhibit varying usage patterns for electric vehicles. Consumers who are likely to use their vehicles for short commutes (e.g., in urban areas like Delhi and Mumbai) may prefer EVs with lower costs and smaller ranges. On the other hand, consumers in suburban or rural areas may prefer EVs with longer ranges, especially in regions where charging stations are fewer and farther apart. This suggests that understanding the usage frequency and range requirements of customers is vital for targeting.

Price Sensitivity: Behavioral segmentation based on price sensitivity is important in identifying segments that are more likely to adopt EVs. Consumers who prioritize affordability may gravitate toward budget-friendly electric vehicle models, while those more concerned with advanced technology may be inclined toward premium EVs with better features such as enhanced range, faster acceleration, and more luxurious interiors. Pricing strategies need to be tailored to the sensitivity levels of each target market segment.

Loyalty to Brands: Behavioral segmentation can also take into account brand loyalty. Some consumers may be brand-conscious and prefer to purchase electric vehicles from well-established brands like Tata or Mahindra, who have a history in the Indian automotive market. Other consumers might be open to trying out new, less established brands like Lohio Auto, especially if they offer advanced features or competitive pricing.

Environmental Consciousness: Consumers who are environmentally conscious may be more likely to purchase EVs as part of their commitment to reducing their carbon footprint. Behavioral segmentation based on environmental awareness can help target the eco-conscious demographic, who are more likely to choose electric vehicles over traditional fuel-powered ones. This segment could be concentrated in urban centers where awareness about climate change is higher.

Psychographic Segmentation:

Lifestyle Preferences: Psychographics play an essential role in understanding EV adoption. Consumers who lead a tech-savvy and modern lifestyle might prefer advanced, high-tech electric vehicles, particularly those offering autonomous driving features or advanced infotainment systems. On the other hand, those with a more traditional lifestyle might prefer simpler, more functional EVs with just the basic features and a focus on reliability and cost-effectiveness.

Attitudes Towards Innovation: Consumers in the EV market can be classified based on their attitude towards innovation. Early adopters, who are excited by new technologies and are willing to experiment with new types of vehicles, represent a key psychographic segment for EVs. These consumers are more likely to invest in the latest EV models, including high-end electric cars that push the boundaries of automotive innovation.

Environmental Values: People who have strong environmental values may see owning an electric vehicle as a way to reduce their carbon footprint. This segment is particularly attracted to EVs because they offer a greener alternative to traditional internal combustion engine vehicles. In targeting this segment, marketing strategies could highlight the positive environmental impacts of switching to an EV.

Social Influence: Many consumers may purchase electric vehicles based on social trends, including peer influence or recommendations from influencers in their social circles. Targeting people who are highly influenced by societal trends or group behaviors can be an effective approach to increase EV adoption, especially in urban areas where trends are more rapidly embraced.

Demographic Segmentation:

Age Group: Age plays a significant role in the type of EV consumers are likely to purchase. Younger consumers (millennials and Gen Z) are more likely to embrace new technologies like electric vehicles due to their environmental concerns, tech-savvy nature, and interest in innovation. On the other hand, older generations may be more hesitant to switch to electric vehicles, preferring traditional cars with proven reliability.

Income Level: Income is another critical demographic factor in determining EV adoption. Electric vehicles are often seen as premium products, and higher-income consumers are more likely to afford them. Targeting affluent urban areas, where disposable incomes are higher, would likely result in higher adoption rates of premium EV models. However, budget-friendly models can cater to middle-income and lower-income segments, who may be more price-sensitive but still interested in transitioning to electric vehicles.

Geographic Location: Although geographic segmentation is already discussed, it also ties into demographics such as population density and urbanization. In more densely populated areas like metro cities (Delhi, Mumbai, Bengaluru), younger, higher-income individuals may be more inclined toward EV adoption, while suburban and rural areas may require more accessible EV models suited for family use, with an emphasis on durability and range.

Occupation and Profession: Occupation and professional background can also impact the adoption of electric vehicles. Professionals working in tech, government, and environmental sectors may show higher interest in electric vehicles due to their sustainability initiatives or affinity for innovative technologies. Similarly, corporate organizations looking to meet sustainability goals may promote the adoption of EVs for their employees.

Conclusion:

Focus on High-Demand Regions with Established Infrastructure: According to the EDA, regions like Maharashtra, Tamil Nadu, and Gujarat have a higher number of charging stations, making them ideal candidates for the initial market launch. Additionally, Uttar Pradesh and Delhi are the states with the highest number of electric vehicles. The strategy should focus on these regions where there is already a developing EV ecosystem, and infrastructure support will aid in consumer adoption. Launching in these locations can help capture early adopters and accelerate market penetration.

Diversify Vehicle Offerings for Multiple Segments: Based on the insights from the vehicle data, there is a need to cater to different customer needs, from budget-conscious city commuters to high-end vehicle seekers. Offering a variety of electric vehicles, ranging from affordable 2-wheelers for daily commuting to high-performance SUVs with longer ranges, will allow the startup to cater to different demographics and use cases. Tailoring the product lineup to match the purchasing power and preferences of different consumer segments will be key to ensuring market coverage.

Leverage Incentives and Pricing Strategy for Early Adoption: The analysis shows that vehicles with greater battery capacity and energy efficiency tend to attract more attention. Therefore, pricing vehicles strategically while also leveraging government incentives and subsidies can make the EV offerings more attractive, especially for the early adopters in the market. Offering competitive pricing for entry-level models and incentives for early buyers can create strong initial demand and generate buzz in the market.

Maximize Range and Energy Efficiency: Insights from the analysis suggest that Tata's vehicles offer the maximum range, which is a key factor for consumer adoption. To capitalize on this, the startup should focus on providing vehicles with long ranges and low energy consumption, addressing one of the major concerns of EV users. Promoting these features as key selling points will help in gaining consumer trust and overcoming the range anxiety that often comes with electric vehicles.

By focusing on these strategic areas, the EV startup can effectively launch and position itself in the market, aligning with both the current consumer needs and the developing infrastructure for electric vehicles in India.

Repo link: https://github.com/sharaaan/EV-Market-Segmentation-Sharan-P.git