Task-1-R

Problem Statement-1. EV Market

Team 1

Adarsh Kumar

23rd January, 2024

Market Segmentation Report for Electric Vehicle Startup

Introduction

The objective of this report is to perform a segmentation analysis of the electric vehicle (EV) market in India using two different datasets and to come up with a feasible strategy to enter the market, targeting the segments most likely to use the EVs.

Data Sources and Preprocessing

Dataset 1: EV Data (India) Collected (Web) - 18MI31001.xlsx: The dataset contains 114 entries, each representing a unique EV model and contains information such as price, range, charging time, top speed, battery capacity, and vehicle type of various EVs available in India. I have collected this data through web scraping from zigwheels.com, a popular online platform for buying and selling vehicles in India. I have also manually searched for the missing values for some of the EVs in India from other sources, such as official websites, news articles, and blogs. The vehicle types included bikes, cars, and scooters.
I dropped the 'Name' column as it was not relevant for clustering, scaled the numerical features and encoded the categorical feature 'Vehicle Type' using One-Hot Encoding.

Dataset 2: electric_vehicle_charging_station_list.csv: With 202 entries, this dataset focuses on the infrastructure aspect of EV adoption, providing details about charging stations across different regions in India. The columns include 'no', 'region', 'address', 'aux address', 'latitude', 'longitude', 'type', 'power', and 'service'. Regions are classified into categories such as 'NDMC,' 'CMRL', 'Maha Metro', 'Noida Authority', 'SDMC', 'NKDA', 'NRANVP', and 'ANERT.'
I converted the latitude and longitude columns to radians to prepare them for distance calculations using the Haversine formula and also created a new column called "lat_rad" and "lon_rad" for this purpose.

Segmentation Analysis

Dataset 1: EV Data Analysis:

Clustering

I have used the K-Means clustering algorithm to segment the EVs based on their features. K-Means is a popular unsupervised machine learning technique that partitions the data into k clusters, where each data point belongs to the cluster with the nearest mean. To find the optimal number of clusters, I have used the elbow method, which plots the within-cluster sum of squares (WCSS) against the number of clusters and looks for the point where the WCSS curve bends sharply.
Using the optimal number of clusters as 2 obtained from the elbow method, I have fitted the K-Means model on the data and assigned the cluster labels to each EV. I have then visualized the clusters using a pair plot (as shown in Fig.1), which shows the pairwise relationships between the features for each cluster.

Insights from EV Clustering

From the pair plot, I can observe that the two clusters have distinct characteristics. Cluster 0 consists of EVs that have lower price, lower range, lower charging time, lower top speed, and lower battery capacity than cluster 1. Cluster 0 also has more bikes and scooters than cars, while cluster 1 has more cars than bikes and scooters. I can infer that cluster 0 represents the low-end segment of the EV market and represents more budget-friendly EVs with shorter ranges, suitable for cost-conscious consumers, while cluster 1 represents the high-end segment of the EV market encompassing higher-end models with longer ranges, appealing to consumers seeking premium performance.
To further understand the differences between the two clusters, I have calculated the average values of each feature for each cluster, as shown in the table below:

Cluster	Price(INR)	Range (km)	Charging Time at Home(hr)	Top Speed (kmph)	Battery Capacity(kwh)	Vehicle Type
0	1.510711e+05	133.731707	4.918659	68.97561	3.653415	Bike/Scooter
1	1.119006e+07	505.531250	8.938750	190.37500	76.793750	Car

The table confirms the observations from the pair plot and shows the significant differences between the two clusters in terms of price, range, charging time, top speed, battery capacity, and vehicle type.

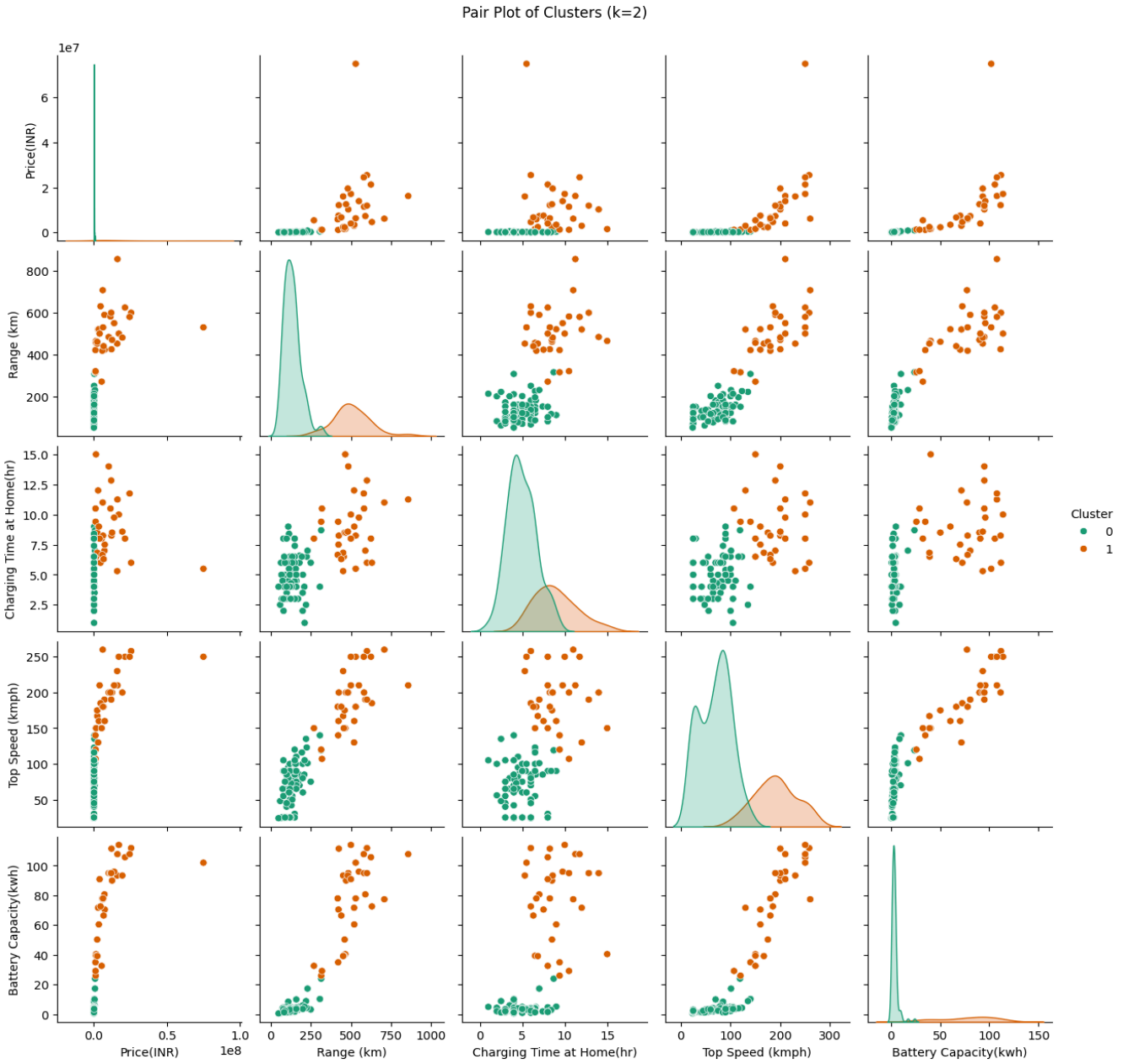

Fig.1: Pair Plot of the pairwise relationships between the features for each cluster

Dataset 2: EV Charging Station Analysis:

Clustering

For the second dataset, I have also used the K-Means clustering algorithm to segment the EV charging stations based on their locations that is latitude and longitude, and used the elbow method, which plots the WCSS against the number of clusters and looks for the point where the

WCSS curve bends sharply. I found that the optimal number of clusters for this dataset is 3, (as shown in Fig.2 below).

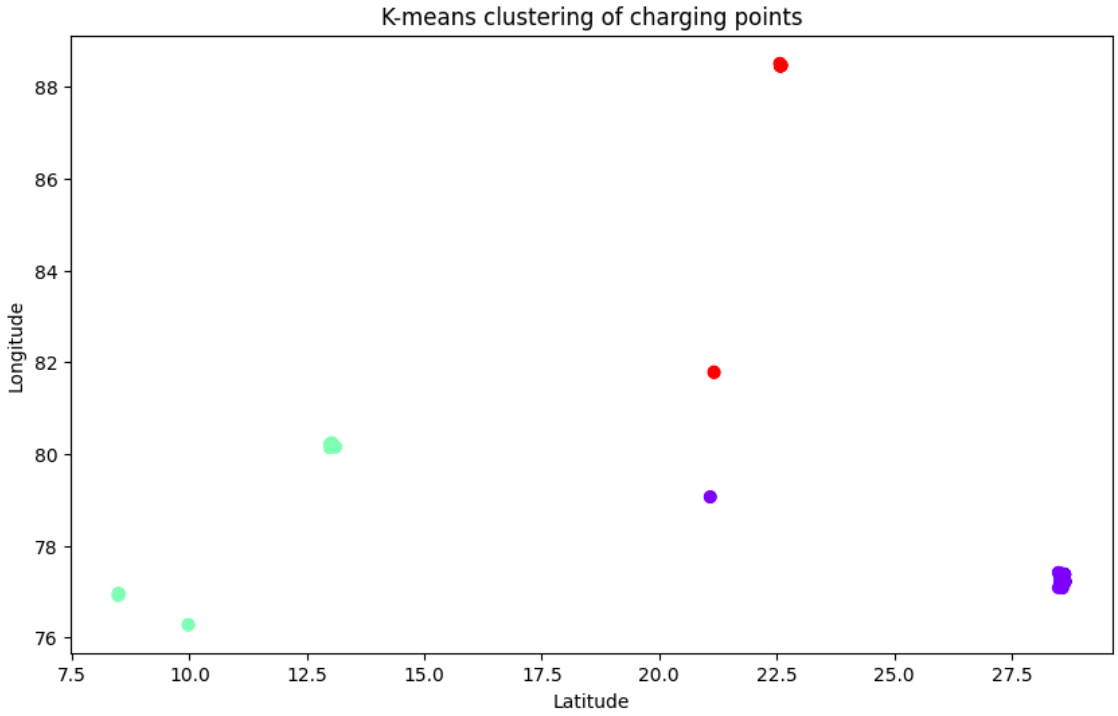

Fig.2: Optimal number of clusters for dataset is 3

However, upon viewing the cluster, I noticed that the two points in the center seem closer than their respective cluster. This may indicate that the elbow method did not capture the optimal number of clusters for this dataset, as there may be more than one elbow point or the elbow point may not be very clear. Therefore, I decided to increase the number of clusters to 4, using my domain knowledge and visual inspection. This may reflect the geographic proximity and similarity of those two points, as well as the business objectives of targeting different regions.

Using the number of clusters as 4, I have fitted the K-Means model on the data and assigned the cluster labels to each charging station. I've then visualized the clusters using a scatter plot, which shows the spatial distribution of the charging stations for each cluster. The scatter plot is (as shown in Fig.3 below):

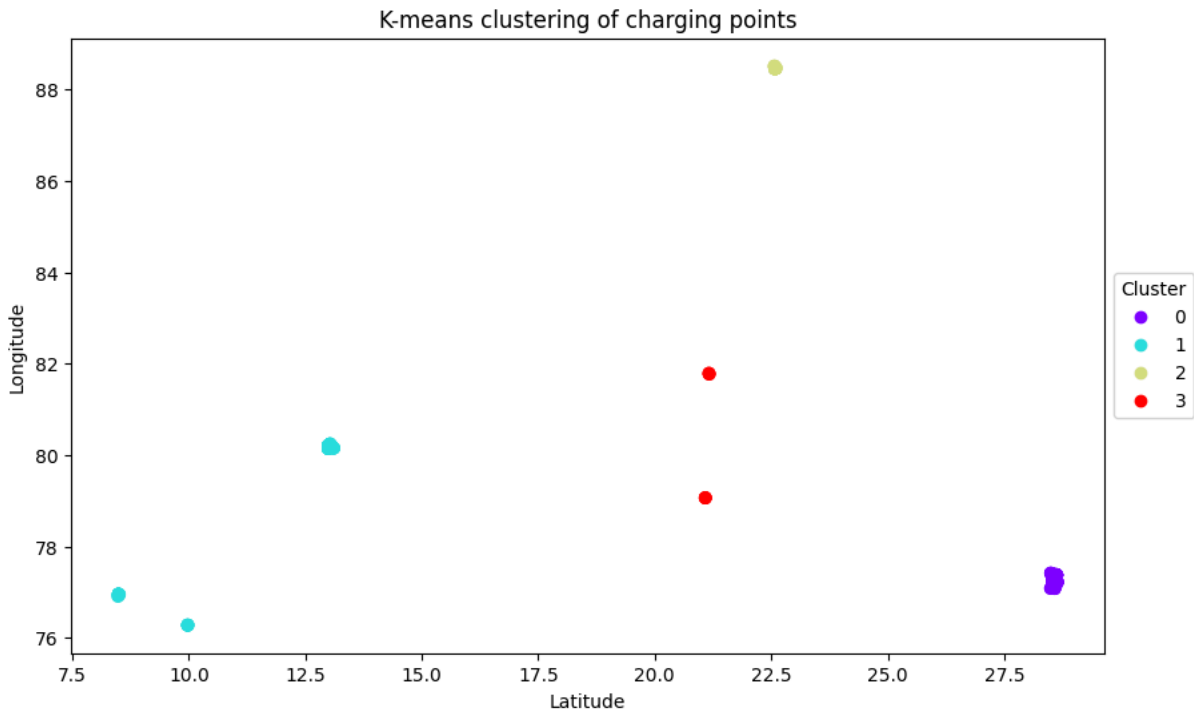

Fig.3: Optimal number of clusters for dataset is 4

Insights from Charging Station Clustering

The clustering of charging stations revealed geographical patterns, indicating areas with a higher density of charging infrastructure. This information is crucial for strategically placing our EVs to cater to regions with well-established charging networks.

From the scatter plot, I can observe that the four clusters have different geographic regions:
Cluster 0 covers most of the northern regions of India, such as Delhi, Noida in Uttar Pradesh.
Cluster 1 covers the southern region of India, such as Tamil Nadu, Kerala, and Karnataka.
Cluster 2 covers the eastern region of India, such as Kolkata in West Bengal.
Cluster 3 covers the central region of India, such as Raipur in Chattisgarh and Nagpur in Maharashtra.

I can also infer that cluster 0 represents the most developed and urbanized region of India, while cluster 3 represents the least developed region of India as compared to Cluster 0. Cluster 1 and cluster 2 represent the intermediate regions of India, with varying degrees of development and urbanization.

Inferences and Insights

The clustering analysis of electric vehicles offers strategic insights into customer preferences and market segments. The segmentation reveals two distinct clusters – Cluster 0 representing budget-friendly EVs, mainly bikes and scooters, and Cluster 1 representing high-end models, predominantly cars. This information enables us to customize marketing strategies; for example, high-end electric bikes may cater to enthusiasts, while electric cars and scooters could attract a broader audience interested in eco-friendly commuting.

On the charging station front, the clustering provides geographical patterns, highlighting regions with varying degrees of charging infrastructure development. Cluster 0 encompasses northern urbanized regions like Delhi and Noida, while Cluster 3 covers less developed central regions like Raipur and Nagpur. This geographical understanding is crucial for planning the distribution and expansion of our electric vehicles, ensuring strategic placement to align with charging networks.

Improvements with Additional Time and Budget

To further enhance the market segmentation project, additional time and budget are crucial. For the EV dataset, collecting more detailed information such as customer reviews, preferences, and usage patterns could provide deeper insights into customer behavior. Exploring external datasets on consumer trends, economic indicators, and government policies could further refine the segmentation, ensuring a more comprehensive understanding of the market.

For the charging station dataset, incorporating real-time data on charging station usage, availability, and user feedback would enhance the accuracy of the clustering. Additionally, obtaining data on government initiatives and incentives related to EV charging infrastructure could be valuable for strategic decision-making.

Estimated Market Size

Determining the estimated market size for the entire EV market in India is a comprehensive task that requires further analysis. However, considering the growing interest in electric vehicles, it's evident that the market size is poised for substantial growth. Conducting additional market research and collecting more data would provide a more precise estimate, allowing us to gauge the potential size of our target market.

Top Variables for Optimal Market Segments

The top variables crucial for creating optimal market segments in the electric vehicle domain include:

- Price (Affordability): Understanding varied budget constraints within different customer segments is essential, influencing their preferences for electric vehicles.

- Range (Commute Distance): Tailoring products based on the typical commute distance of potential customers ensures our offerings meet their specific needs and expectations.

- Charging Time: Recognizing the importance of practicality, customers may prefer vehicles with shorter charging times, especially in regions with limited charging infrastructure.

- Vehicle Type: The type of vehicle (bike, car, or scooter) plays a pivotal role in shaping customer preferences and usage patterns. Our product offerings need to align with the diverse preferences within these vehicle types.

Conclusion

The ideal target segment for our EV startup in India is the low-end segment, characterized by affordability and practicality. By producing a diverse range of budget-friendly electric bikes and scooters, we can cater to a larger customer base and foster early adoption. Strategic distribution based on charging station clustering will ensure that our EVs are accessible and convenient for customers across different regions.

Entering the Indian EV market requires a nuanced approach that considers not only customer preferences but also the evolving landscape of charging infrastructure. With continuous refinement, data enrichment, and strategic planning, our startup can position itself as a key player in the dynamic and rapidly growing Indian Electric Vehicle market.

Google Colab/GitHub Link:
https://github.com/theadarshkr/Feynn-Labs-Assignment/blob/main/Adarsh_Kumar_T_1_R.ipynb

Adarsh_Kumar_T-1-R.ipynb