EV Market Segmentation Report

Manish Walurkar

12/9/2024

Git hub Link: https://github.com/Manish-128/Market-Segmentation/tree/4570972624bb8f378979461f288ef8ba167002ee/Feynn%20Laboratories

Abstract:

The Automobile market is very huge and despite the niche products there are many competitors in the market. The Electronic Vehicle's market in India is starting to grow, with some settled spaces across the map. The aim of this market segmentation is to understand the EV market and various factors that affects the market growth in a certain area. The research was conducted on segmentation variables for demographic, geographic, behavioural, infrastructure, and many mixed variables that supported the process. At the end of the research we succeeded in finding various customer profiles and examined them in details to understand the EV market.

Understanding the Market:

India's automotive market is one of the largest and fastest-growing globally, fueled by a rising middle class, increasing incomes, and rapid urbanization. Traditionally, this market has been dominated by vehicles with internal combustion engines, but we're seeing a clear shift toward electric vehicles (EVs). This shift is spurred by growing environmental awareness, government incentives, and advancements in EV technology. While EVs currently represent a small portion of total vehicle sales, the segment is steadily expanding, especially in states where policies and infrastructure are more favourable.

Leading states for EV adoption, such as Uttar Pradesh, Maharashtra, and Tamil Nadu, are benefiting from both their large populations and better economic conditions. These states account for a significant share of EV sales, although challenges like charging infrastructure still persist. Smaller states like Goa and Delhi, while contributing less to overall sales, show high EV readiness with better-developed infrastructure, including a higher density of Customer Service Points (CSPs) and surfaced roads that make EV adoption easier. In contrast, states like Tripura and Jharkhand are lagging behind in both sales and infrastructure, pointing to the need for foundational improvements to prepare for future EV growth.

1. Planning and data collection:

1.1 Dividing the Problem:

The segmentation task is itself is a huge task and requires a lot of variable and analysis technique. The first and the most important task is to break the problem into small pieces.

- Objectives to achieve: A clear, distinguishable and profiled segment needs to be delivered.
- Variables: Thinking in right direction is very necessary. Before starting with anything we will be needed to understand various variables that may affect the EV sales in a particular region.
  Here is a table of example variables for different segmentation that may affect EV sales:

	Demographic	Psychological	Geographical	Behavioural
Variable 1	Income	Environmental Awareness	Roads Quality	Yearly Registered Vehicles
Variable 2	Population	Government Facilities	Charging Station	EV Sales
Variable 3	Urban/Rural	Running Cost		

Along with the more common variables, more variable can be created of these, for example based on "Income" we can create "High class, Low class and Middle class" to help understanding the demographic pattern more accurately.
More will be discussed in the upcoming points.

- Data Collection: In order to bring these segmentation variables to life we will need to collect the data for these variables.
  The data collection must contain a variable or referred to a quality of variable that can be retrieved later.
  We are using Power BI for visualization and data pre-processing, following is an image that shows the collected data.

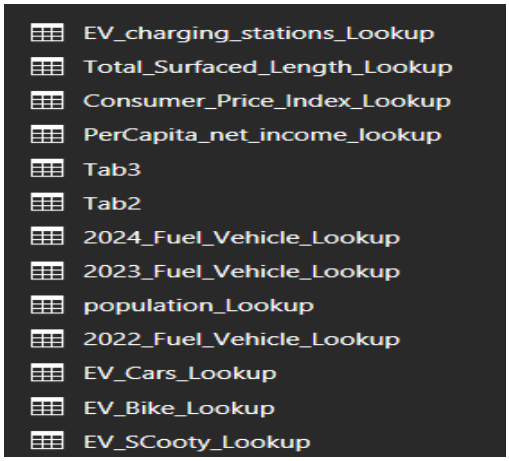


Fig. Data Collected

1.2 Approach:

The approach of our team is not to complex the relations between the data. We will follow the 10 steps rule for market segmentation.

2. Knock-Out Criteria

2.1 Per Capita Income: If the per capita income of the segment is above 100,000rs then the segment will pass, else the segment will fail and be knocked out.

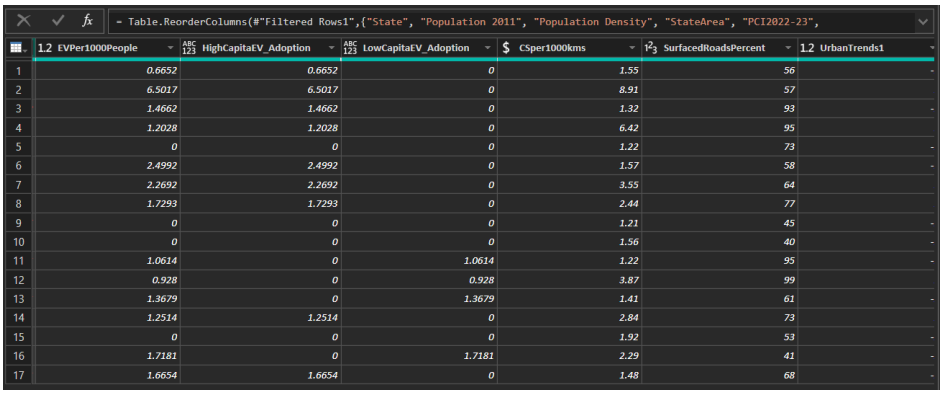
2.2 Charging Station per 100km^2: The second criteria is that there must be 1 charging station per 100 km2 of area within the region/state. If the segment fails to qualify then the segment will be rejected.

2.3 Surfaced Roads: The percent of Surfaced roads in a state when dropped below 40 then the segment will get rejected. If the percent of surfaced roads is more than 40, it shows an urban sign and better road quality thus the segment is passed.

3. Data Pre-processing:

3.1 Imputation: If the variable is necessary to segment and the data is not present then, either the data is imputed with StandardImputer from sklearn, or it is removed.

3.2 Creation of sub data: As mentioned above the data is divided further to answer questions in binary or numeric. For example the creation of EV sold per 1000 people for this the total (EV sales x 1000) / population of state. Creation of these sub data helps in forming better segmentation.



= Table.ReorderColumns(#"Filtered Rows1",{"State", "Population 2011", "Population Density", "StateArea", "PCI2022-23",

	1.2 EVPer1000People	HighCapitaEV_Adoption	LowCapitaEV_Adoption	$ CSper1000kms	SurfacedRoadsPercent	1.2 UrbanTrends1
1	0.6652	0.6652	0	1.55	56	-
2	6.5017	6.5017	0	8.91	57	
3	1.4662	1.4662	0	1.32	93	-
4	1.2028	1.2028	0	6.42	95	
5	0	0	0	1.22	73	-
6	2.4992	2.4992	0	1.57	58	-
7	2.2692	2.2692	0	3.55	64	
8	1.7293	1.7293	0	2.44	77	
9	0	0	0	1.21	45	-
10	0	0	0	1.56	40	-
11	1.0614	0	1.0614	1.22	95	-
12	0.928	0	0.928	3.87	99	
13	1.3679	0	1.3679	1.41	61	-
14	1.2514	1.2514	0	2.84	73	
15	0	0	0	1.92	53	-
16	1.7181	0	1.7181	2.29	41	-
17	1.6654	1.6654	0	1.48	68	-

Fig. Sub-data fields

CSPer100kms: This variable is formed by dividing the charging stations by state area per 100 km. This data variable shows the spread of EV charging stations across the state. Lowered Charging station per 100 km means that the state don't have distributed charging stations across the states which may lead to selective area where EVs are used.

SurfacedRoadsPercent: This variable is the percent of the surfaced road area by total road area. This shows how much percent of surfaced roads are there in state. This is also a crucial data for considering urban trends, and EV sales.

3.3 Data Model:

Fig. Data Model

The data model has population lookup table at its centre. All the tables are linked somehow with the population tables. The population table contains population data for all the states and UTs of 2011 census. Thus the states column is the foreign key for all states. The relationships were used to either merge the data, or use them to show visualizations in the report panel.

4. Understanding The Data:

4.1 Understanding the Data variables:

1. State: The data revolves around the state, and thus it is the primary key and the foreign key for most of the datasets.
2. Charging Station: The charging station is the number of total charging station available in a state.
3. CSper100kms: This variable shows the number of charging stations available in per 100 km of area in the state.
4. EVPer1000People: This shows how many Electronic vehicles are owned per 1000 people in a state.
5. EVSales2023: This is the summation of registered Electronic Vehicles in a state. Thus it is the summation for all the 2 wheelers, 3wheelers, and 4wheelers vehicles registered in the year 2023.
6. HighCapitaVS_EVAdoption: This data variable shows the spread of registered EV vehicles for higher per capita income in a state.
7. LowCapitaVS_EVAdoption: This data variable shows the spread of registered EV vehicles for lower per capita income in a state.
8. PCI2022-23: The Per Capita Income for the financial year 2022 and 23 is within this variable.
9. Population 2011: The variable contains the population per state and union Territories.

10. SurfacedRoads: The roads are mainly divided in two part, Surfaced and Unsurfaced. The surfaced roads shows the numeric value for km of roads in a state which is surfaced.
11. SurfacedRoadsPercent: This shows the percentage of surfaced roads there are in a state.
12. Total Rural Trends: This data variable consists of 3 total rural trends variable (Fewer Population density, More unsurfaced roads and low per capita income) thus the summation of all these groups became the value. If the condition satisfies then the value is 1 otherwise 0.
13. Total Urban Trends: It is the same as Rural Trends consisting of Urban Trends Variable as (Higher Per Capita Income, Higher Population Density and More Surfaced Roads). The criteria for summation and points is same as total rural trends.

4.2 Exploratory Data Analysis:

We decided to do the EDA in order to understand the influence of the data variables across the dataset.

As the state becomes the main criteria for merging datasets, most of the EDA will be focused on comparison of different data variables in accordance with the state.

1. State VS Population

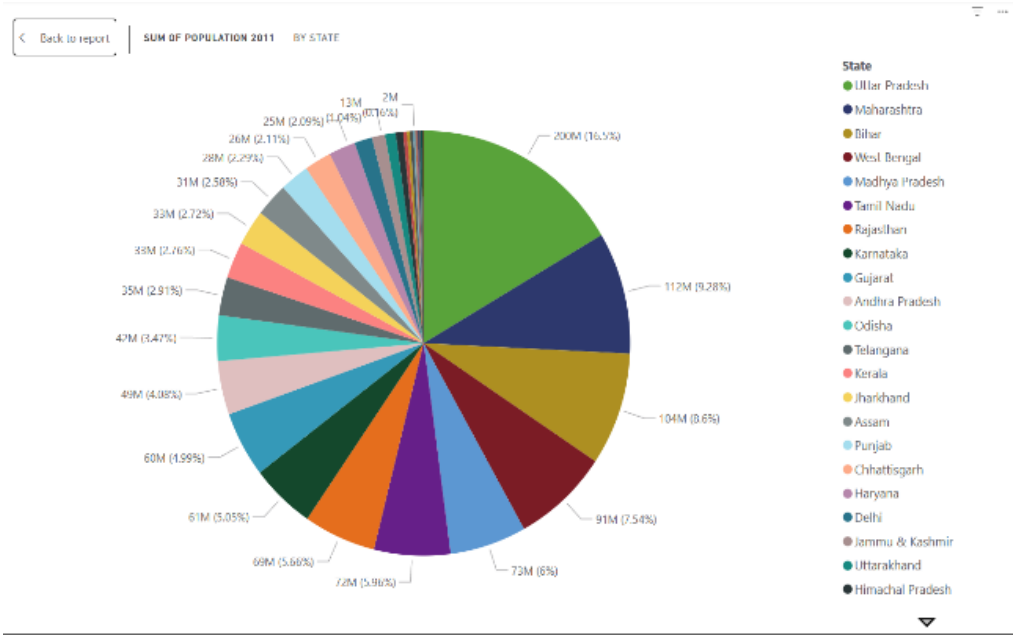


Fig. Pie Chart for Population by States

2. State VS EV Charging Station

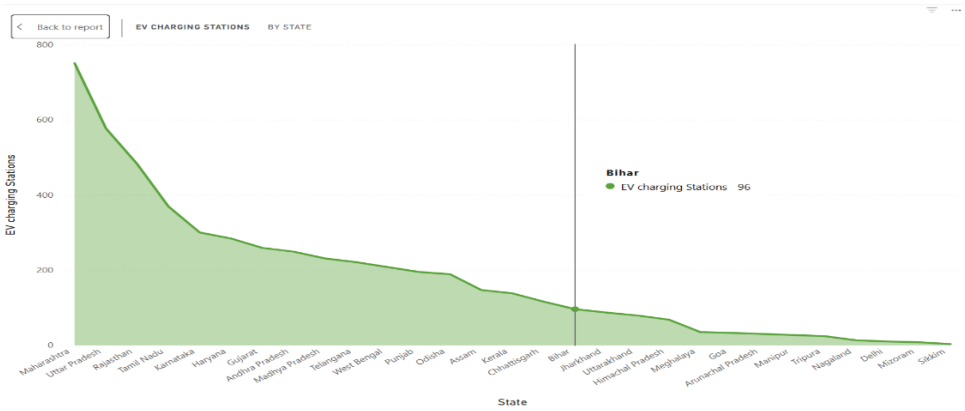


Fig. Number of EV charging stations by State

3. State VS EV Sales

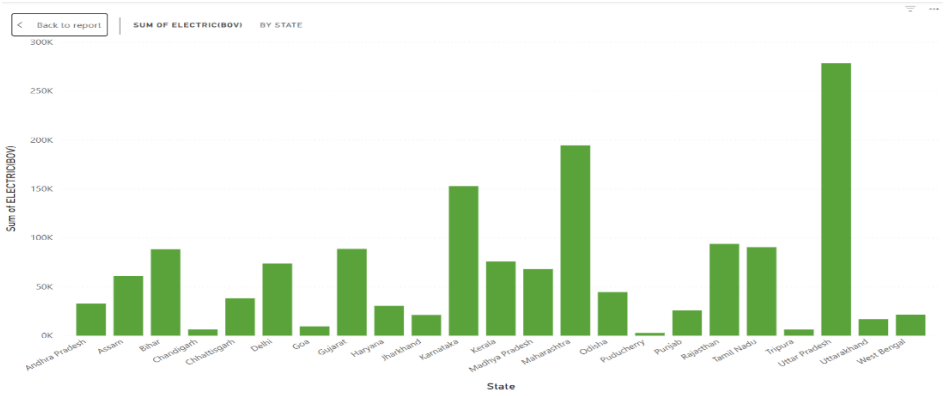


Fig. Bar Graph for EV sales by State

4. State VS EV per 1000 Peoples

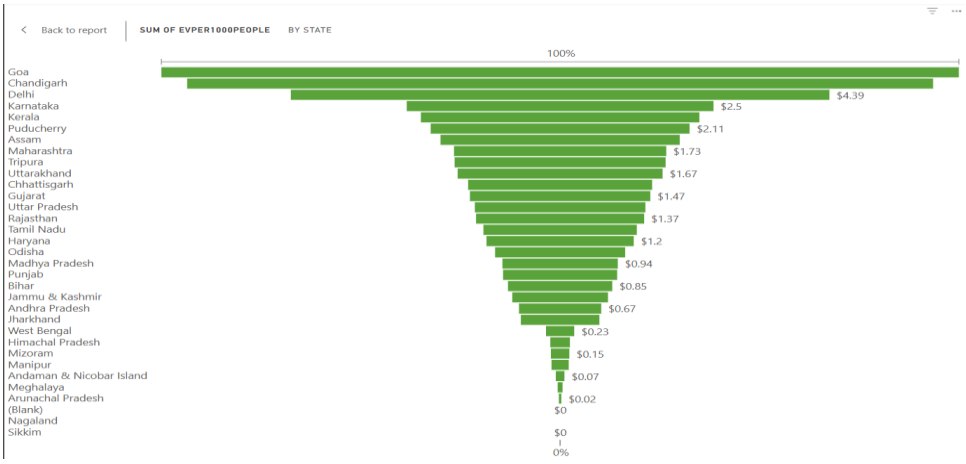


Fig. Funnel Chart for EV owned per 1000 people by State

5. State VS Per Capita Income

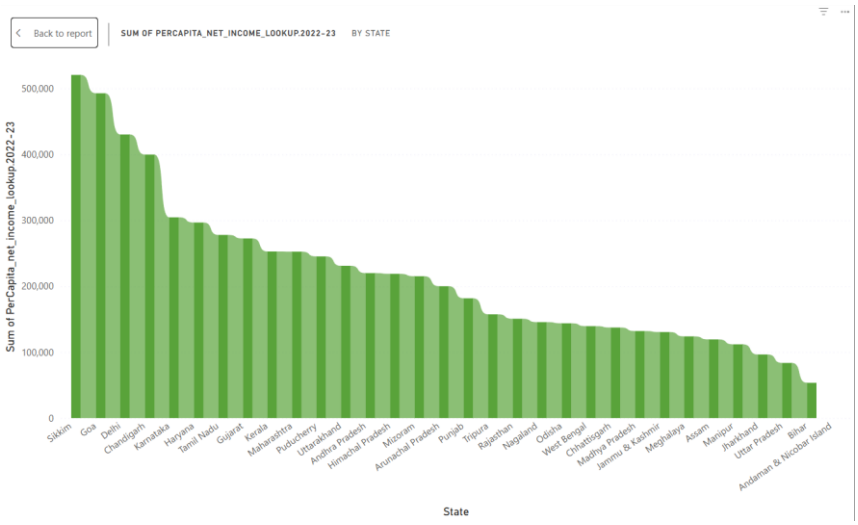


Fig. Curved Bar Graph for per capita income by State

4.3 Principal Component Analysis:

The PCA is done to reduce the dimensionality of the data. We considered a total of 10 Principal components for the PCA analysis. The PC1 had good spread, across the demographic segmentation variable and geographic segmentation variables. The PC2 focused more on EV sales, population, and charging station.

We focused more on PC1 vs PC2 as both of the Principal components showed good spread of points on variables. The PCA has its most of the variance on first 3-4 PCs thus, we decided to use the first and second PCs for the clustering.

State, ChargingStations, CSper1000kms, EVPer1000People, EVSales2023, HighCapitaEV_Adoption, LowCapitaEV_Adoption, PCI2022-23, Population 2011, SurfacedRoads, SurfacedRoadsPercent, TotalRuralTrends and TotalurbanTrend

Fig. Data Variables used in PCA

By using the above mentioned data variables, the cold map was created for all the PCA components describing their spread on all the data variables. This map will be helpful in determining the best PCA to use in the analysis.

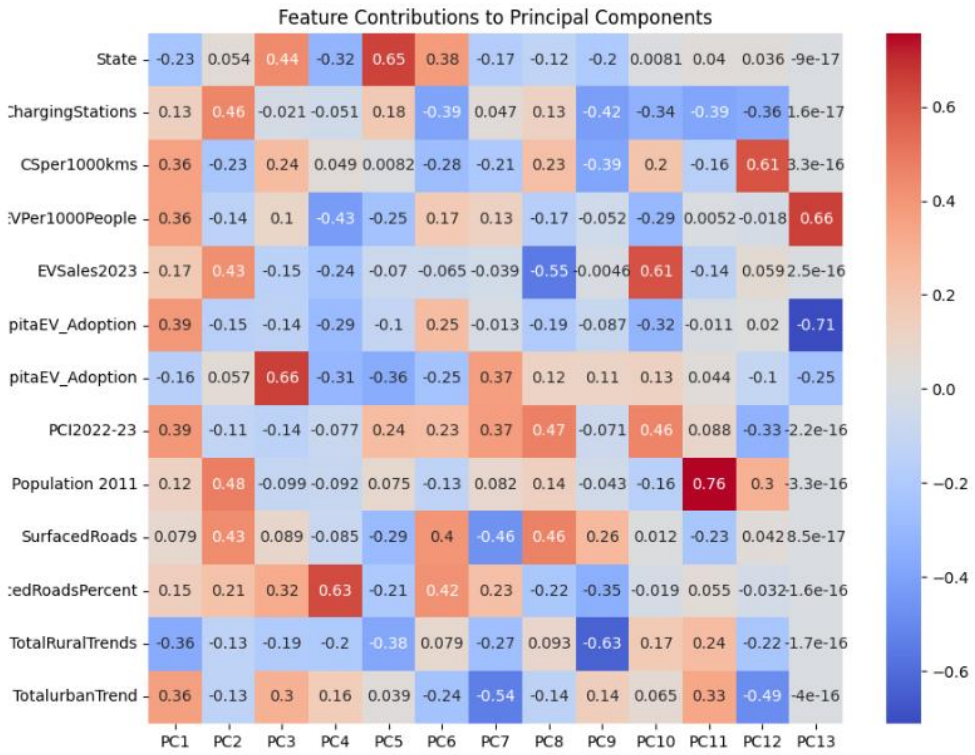


Fig. Feature Distribution for Principal Components

The spread of the points for data variables across all Principal components can be seen here. The reason for not selecting PC3 was its major focus on High Per capita vs EV adoption, as a good spread of data variables is considered better, thus PC3 was avoided. The PC4 was avoided for the same reason. After PC4 there was not much of the variance within PCs so we decided to continue with the PC1 and PC2.

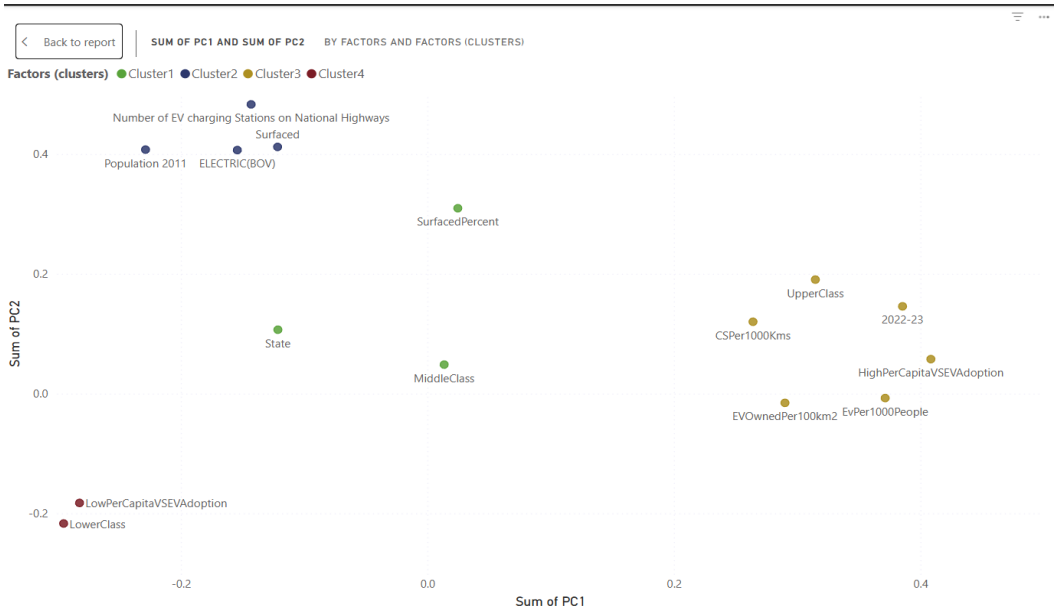


Fig. PC1 vs PC2

The distribution for the variables for PC1 and PC2 was visualized on a scatter plot to see which variables have similar spreads on both the principal components. The cluster below shows those data variables.

5. Extracting segments

5.1 Forming Clusters

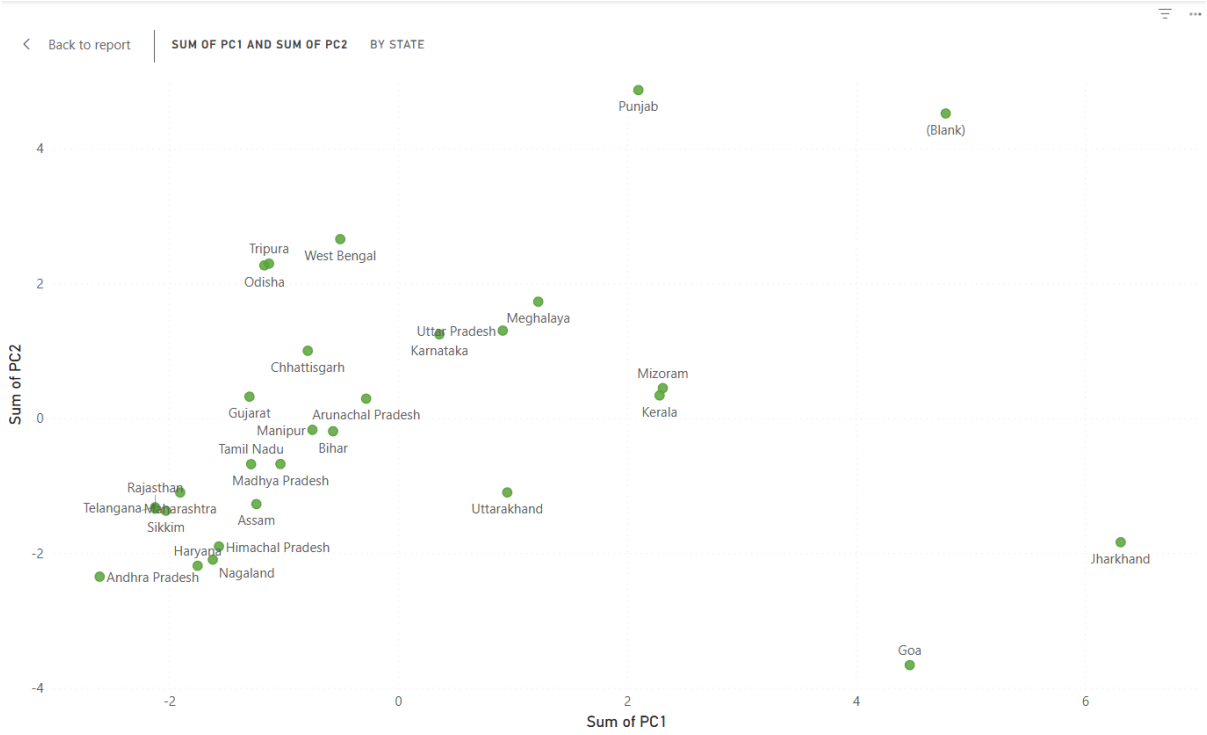


Fig. PC1 vs PC2 by states

For forming the clusters the scatter plot was used. The PC1 is used in the X axis and the Y axis has PC2. The states were used as the Values to show the category label. The data plotted was relatively complex and was based on all the data variables used in the PCA.

For selecting the number of clusters to form the K-Elbow method was used in accordance with the silhouette score to get the k values for clustering.

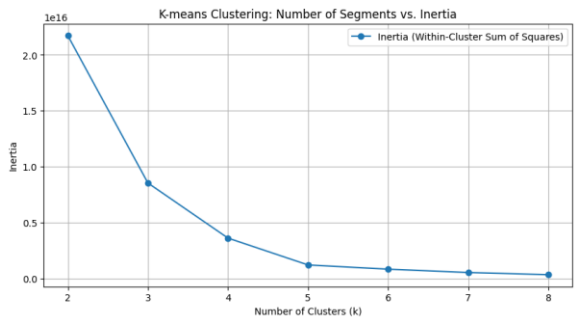


Fig. K-Elbow Method

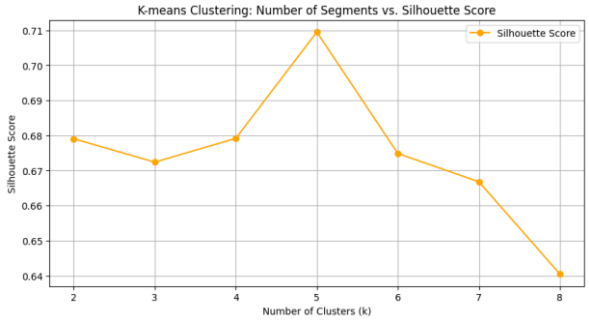


Fig. Segments vs Silhouette Score

The clusters were formed with the use of Power BI's inbuilt *automatically find clusters* technique. The clustered form were then visualized as

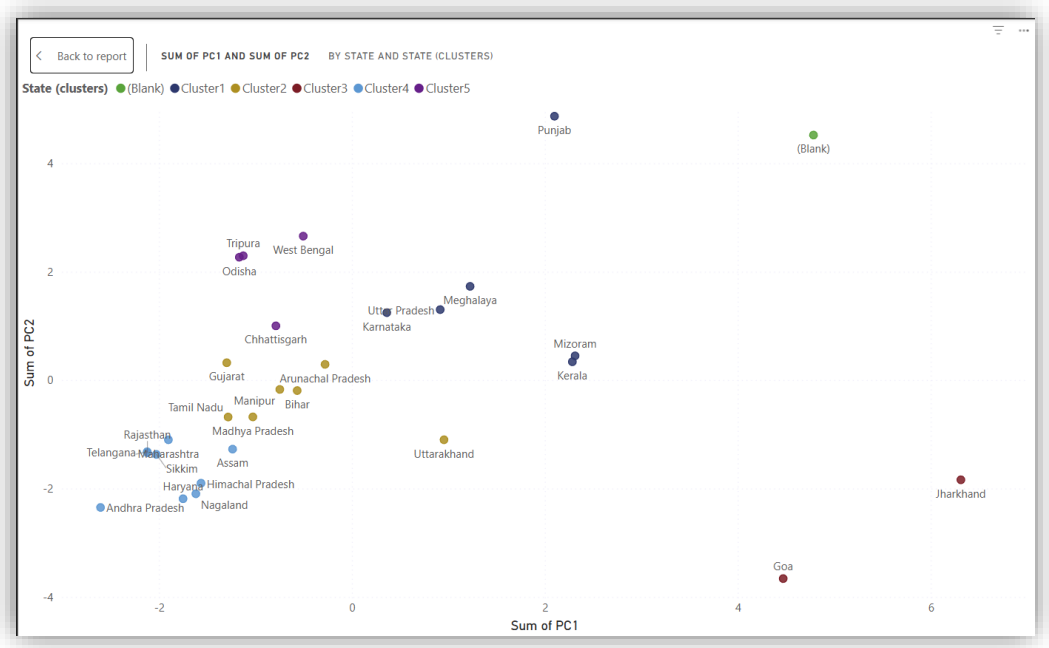


Fig. Clustered data based on PC1 and PC2

6. Profiling Segments

Before jumping right on to the clusters, we will need to define the analysis criteria for the clusters. The clusters below are shown with a dashboard with bunch of visuals which will assist us to profile the clusters.

Key Visuals Summary

1. Clusters by PC1 and PC2 (Scatter Plot)

   • This plot shows states grouped by clusters based on two principal components (PC1 and PC2), which are likely derived from various socioeconomic and demographic factors.

2. Population Distribution (2011 Census) (Pie Chart)

   • Displays population by state, giving an overview of population concentration.

3. Surface Roads (Donut Chart)

• Shows the distribution of surfaced roads across states, which could indicate infrastructure development levels.

4. Per-Capita Income (PCI) by State (2022-23) (Line Chart)

   • Provides PCI data for each state, which is a measure of economic prosperity.

5. EV Sales 2023 by State (Tree Map)

   • Displays Electric Vehicle (EV) sales distribution by state, reflecting market adoption levels and potential environmental considerations.

6. Charging Stations per 1000 km (CSP) (Bar Chart)

   • Indicates the number of charging stations relative to road length by state, which can be an indicator of readiness for EV adoption.

The clusters were found to share some common similarities in between them based on that the clusters were named. The following clusters were discussed in below.

1. Cluster 1: High Population, Low EV Readiness

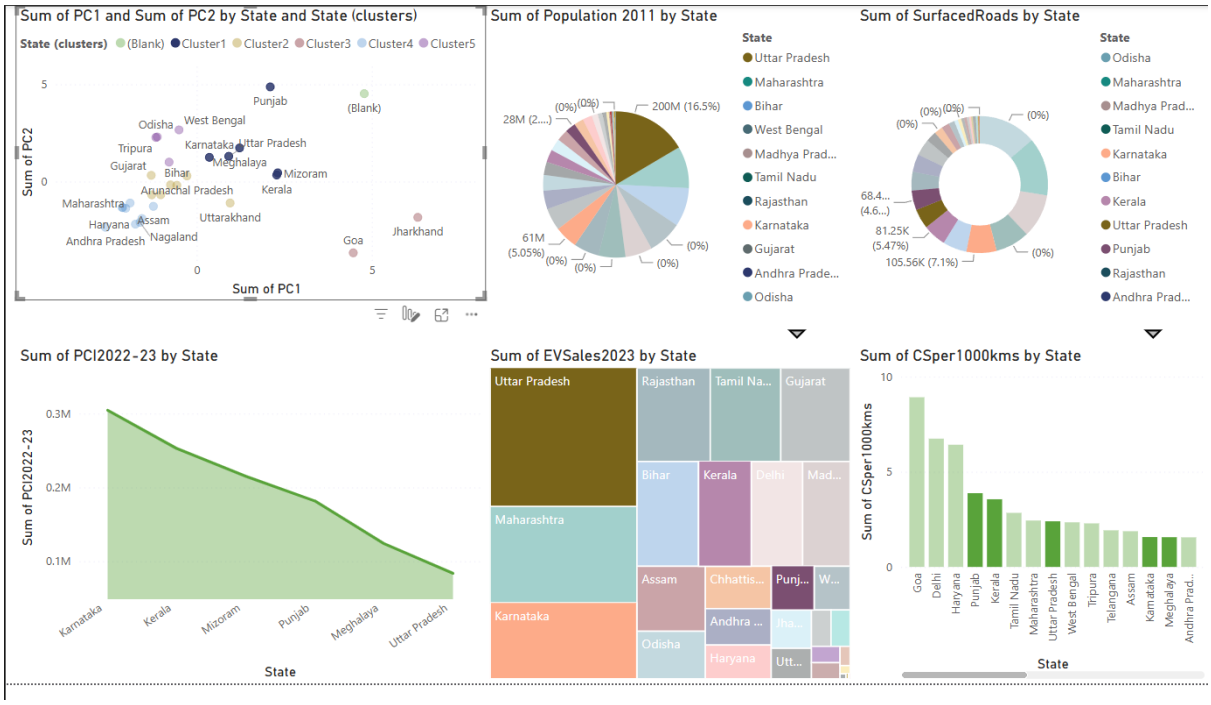


Fig. Cluster 1

1. States: Uttar Pradesh, Punjab, Kerala, Karnataka, Meghalaya
2. Characteristics: Large population, distributed per-capita income, limited EV infrastructure, moderate surfaced roads, moderate CSP per 1000 km

Uttar Pradesh is the biggest state in term of population and EV sales. Karnataka has the highest income in this segment, following with Kerala, Mizoram, Punjab, and Uttar Pradesh. The average of 4 charging stations have founded in this segment,

meaning there is already some infrastructure set up. This segment overall has “High population with Moderate Per Capita income”.

2. Cluster 2: Moderate EV Sales with Growing Infrastructure

   States: Tamil Nadu, Gujarat, Uttarakhand, Arunachal Pradesh, Madhya Pradesh, Bihar

   Characteristics:
   1. Lower PCI compared to other clusters, surfaced roads.
   2. Lower population density, relatively low EV sales, and fewer charging stations.
   3. Limited infrastructure and slower EV adoption rates.

   Cluster 2, excluding states like Gujrat and Tamil Nadu, other state shows lower levels of EV sales and infrastructure, with a relatively balanced distribution of surfaced roads. These states are actively moving towards EV readiness, with state-level policies and moderate consumer interest in EVs. The growing infrastructure in these states positions them as emerging EV markets where targeted investments in service points and marketing could yield substantial adoption rates. “This cluster demonstrates the potential for a near-term increase in EV sales with the right mix of support and development.”

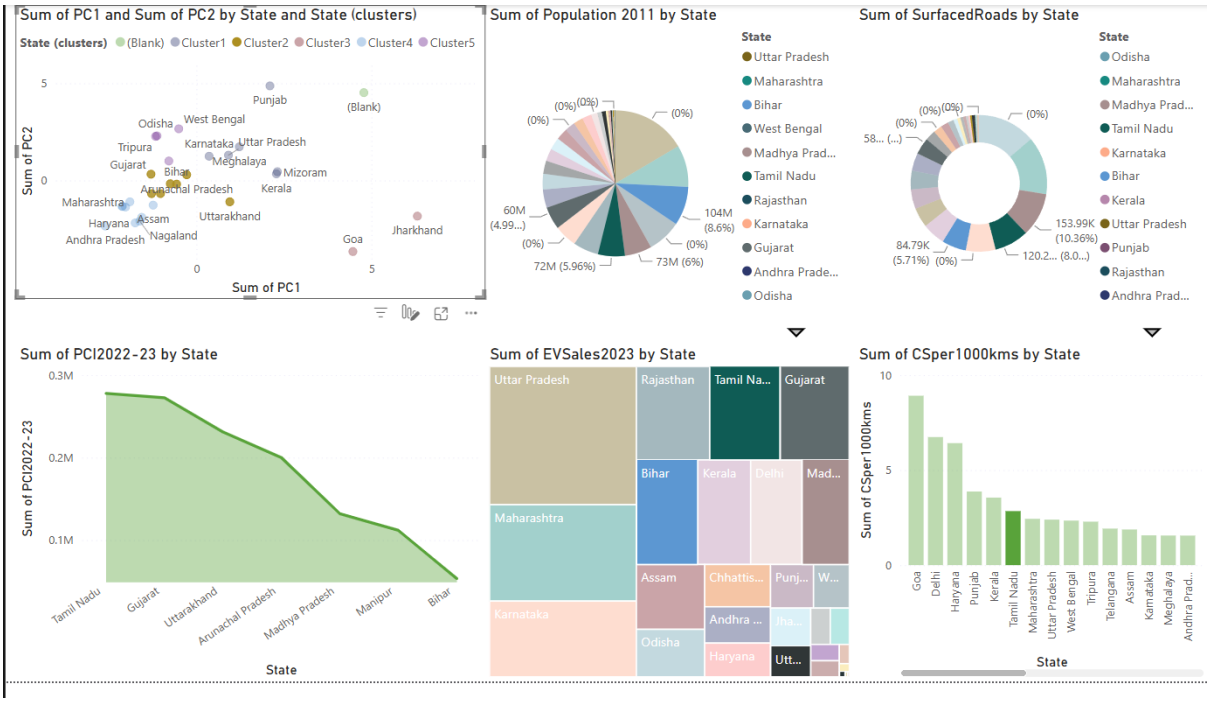


Fig. Cluster 2

3. Cluster 3: High Per Capita Income, High EV Readiness

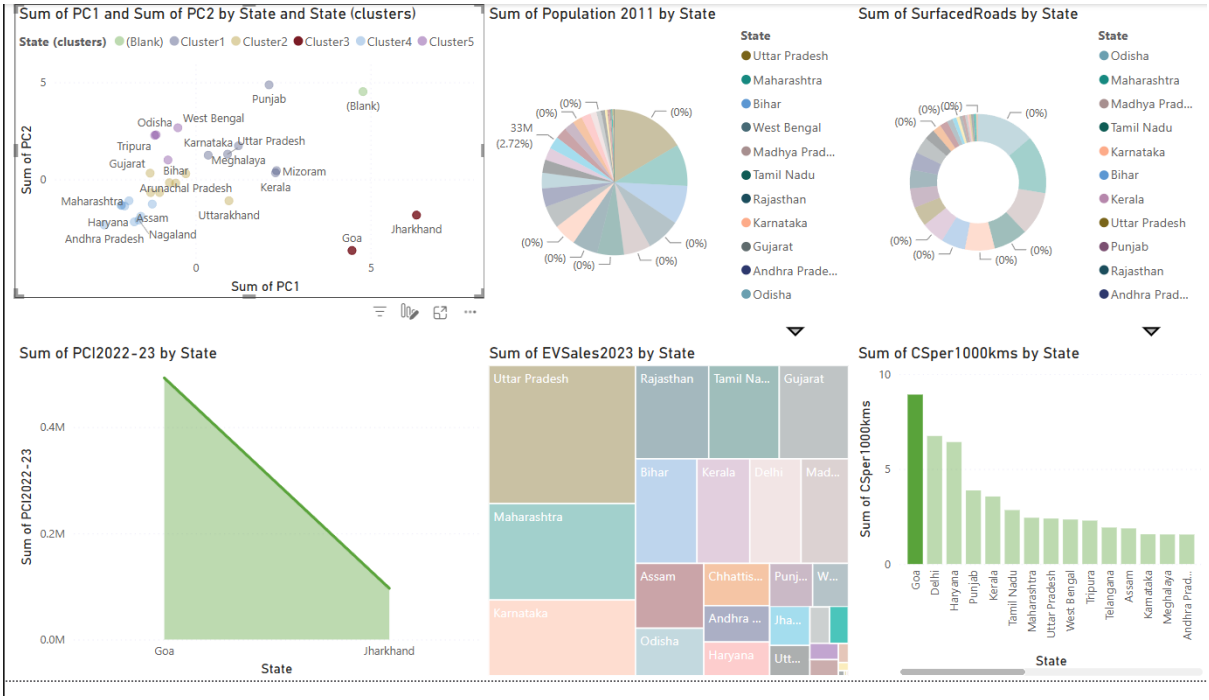


Fig. Cluster 3

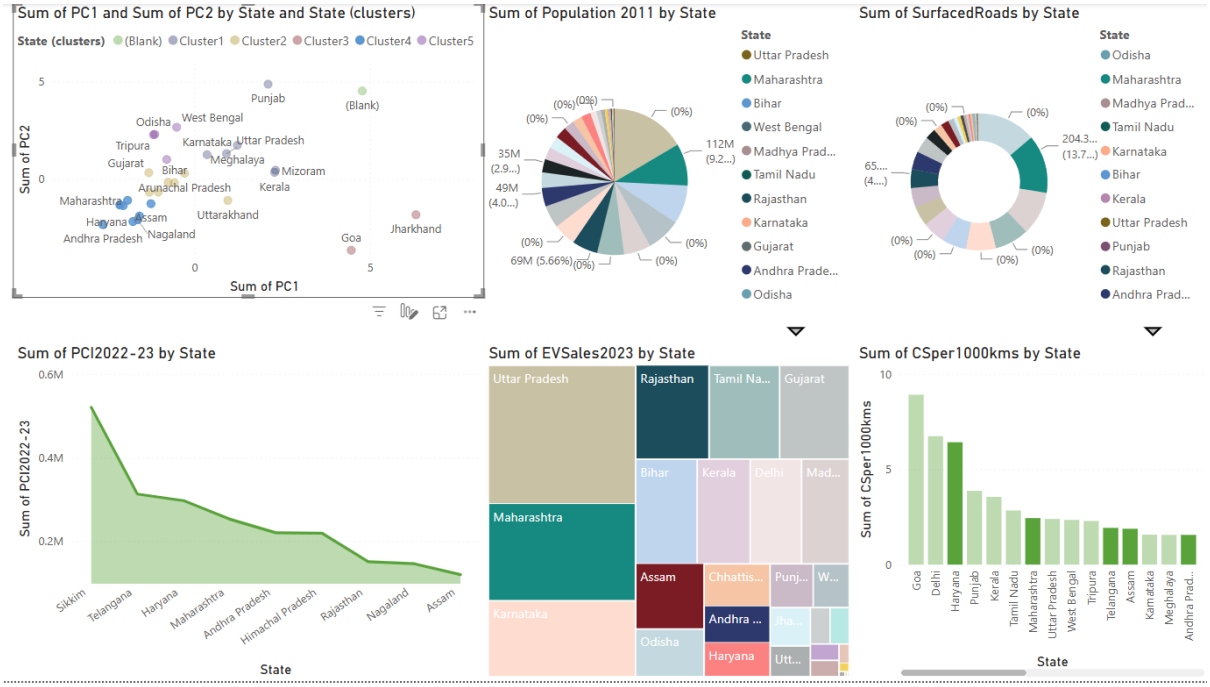
States: Delhi, Goa, Jharkhand

Characteristics:

1. High PCI, moderate surfaced roads, and leading in CSP per 1000 km.
2. Higher EV sales in comparison with its area, reflecting a strong interest and support for EVs.

States in Cluster 3, including Goa and Delhi, are characterized by relatively small populations but advanced EV infrastructure. Goa, for example, has a high concentration of CSPs and fair EV sales relative to its population and size. These states serve as EV adoption models, showcasing how strong infrastructure and accessibility can drive adoption despite limited population. Cluster 3 is well-positioned for further EV growth, potentially serving as a testing ground for new EV models and charging technologies due to its readiness and receptive market.

4. Cluster 4: High EV sales, Distributed Population and Per Capita Income



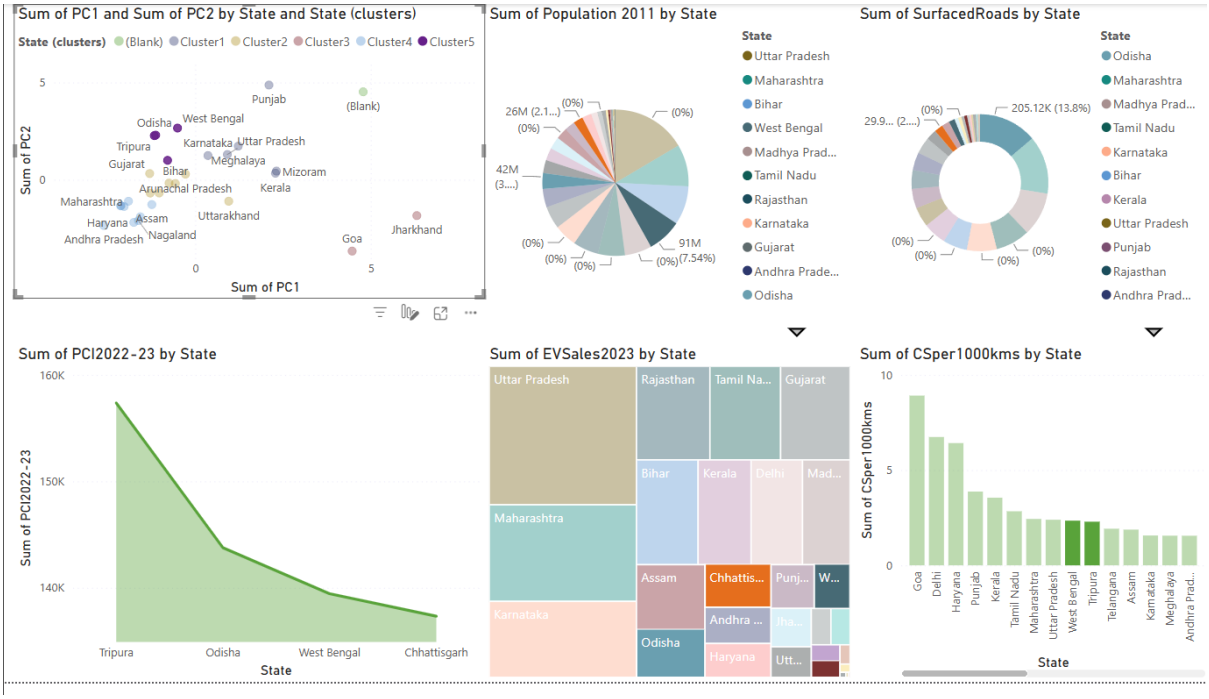
States: Sikkim, Telangana, Haryana, Maharashtra, Andhra Pradesh, Himachal Pradesh, Rajasthan and others.

Characteristics:

1. Moderate PCI, high population, and higher road infrastructure.
2. High EV sales, indicating a large potential market due to population but lagging CSP.

Cluster 4, containing states like Maharashtra and Haryana, has a higher population base but noticeable gaps in EV infrastructure. Although EV sales have started to gain traction in these states, there is still a lack of sufficient CSPs to support widespread EV adoption. Infrastructure development, including additional charging points and improved road networks, could significantly boost the EV market potential in these states. Cluster 4 represents regions where government support and infrastructure investment could bridge gaps to foster a more conducive environment for EV growth.

5. Cluster 5: Low Readiness, Emerging Interest



States: Tripura, Odisha, West Bengal, Chhattisgarh

Cluster 5 includes states like Tripura and Odisha, where both population density and EV infrastructure are relatively low. EV adoption here faces challenges due to limited CSPs and lower surfaced road coverage, creating accessibility issues. However, there is emerging interest in these states as awareness of EV benefits grows. Investment in basic infrastructure and targeted marketing efforts could gradually improve readiness and interest. These states represent long-term opportunities, where foundational steps towards EV infrastructure could eventually yield a viable market for future EV adoption.