MARKET SEGMENTATION ANALYSIS

OF

ELECTRIC VEHICLES IN INDIA

SOUBHAGYA RANJAN BEHERA

Problem Statement :

Electric Vehicle (EV) market in India to determine the most viable customer and vehicle segments for a new EV startup. Given the complexity of the market and the diverse factors influencing EV adoption, the task requires a comprehensive segmentation analysis.

This includes geographic, demographic, psychographic, and behavioral segments, among others. Due to the scarcity of readily available data, a Fermi estimation approach is proposed to break down the problem into manageable parts, make educated assumptions, and estimate the size of potential target markets.

This will help the startup identify the most promising segments to focus on and develop a feasible market entry strategy.

Fermi Estimation:

Fermi estimation is a method used to tackle complex problems by breaking them down into smaller components and making reasonable assumptions.

For analyzing the EV market in India, we start with available data: India's population is approximately 1.4 billion.

Assume 20% can afford a car, and of these, 5% are early adopters of EVs. With 35% of the population living in urban areas and the current EV market penetration at about 1%,

we can calculate the potential market size: 1.4 billion × 0.2 × 0.05 × 0.35 = 4.9 million potential customers.

To refine these estimates, we consider additional factors like government policies, incentives, and the availability of charging infrastructure. This approach helps in estimating the target market size, aiding in strategic decision-making for market entry.

Data Sources :

Data was extracted from Various Websites. Extracted data are given below,

Dataset-1

https://drive.google.com/file/d/1qQ1AudJAFVuqnRI_9loysYqp8n8iFloU/view?usp=drive_link

Dataset-2

https://drive.google.com/file/d/1s1HQ_y60oaSs0Lqo8W86_CkGlRZM_MQu/view?usp=drive_link

Column Explanation :

- State column specifies about the state name .
- Two Wheeler ,Three Wheeler , Four Wheeler , Goods Vehicle , Public Service Vehicle , Special Category Vehicles ,Ambulance , Construction Equipment Vehicle describes about the type and number of vehicles from various state.
- Grand Total says about total number of vehicle from each state .
- Total charging station column says about number of charging station available in each state.
- Region column describes about the vehicle belongs to which region.
- Address and Aux Address describes about the detailed address of vehicle.
- Type also says which type of vehicle it is.
- Power column says the power of the vehicle in KW.
- Also about the service of vehicle i.e. whether it is self or not is described in service column.
- And Latitude , Longitude describes the region of vehicle.

Data Preprocessing:

These steps includes Pre Processing The data.

- Ordinal encoded “Power”
- Used Standard Scaler for Pre processing the data.

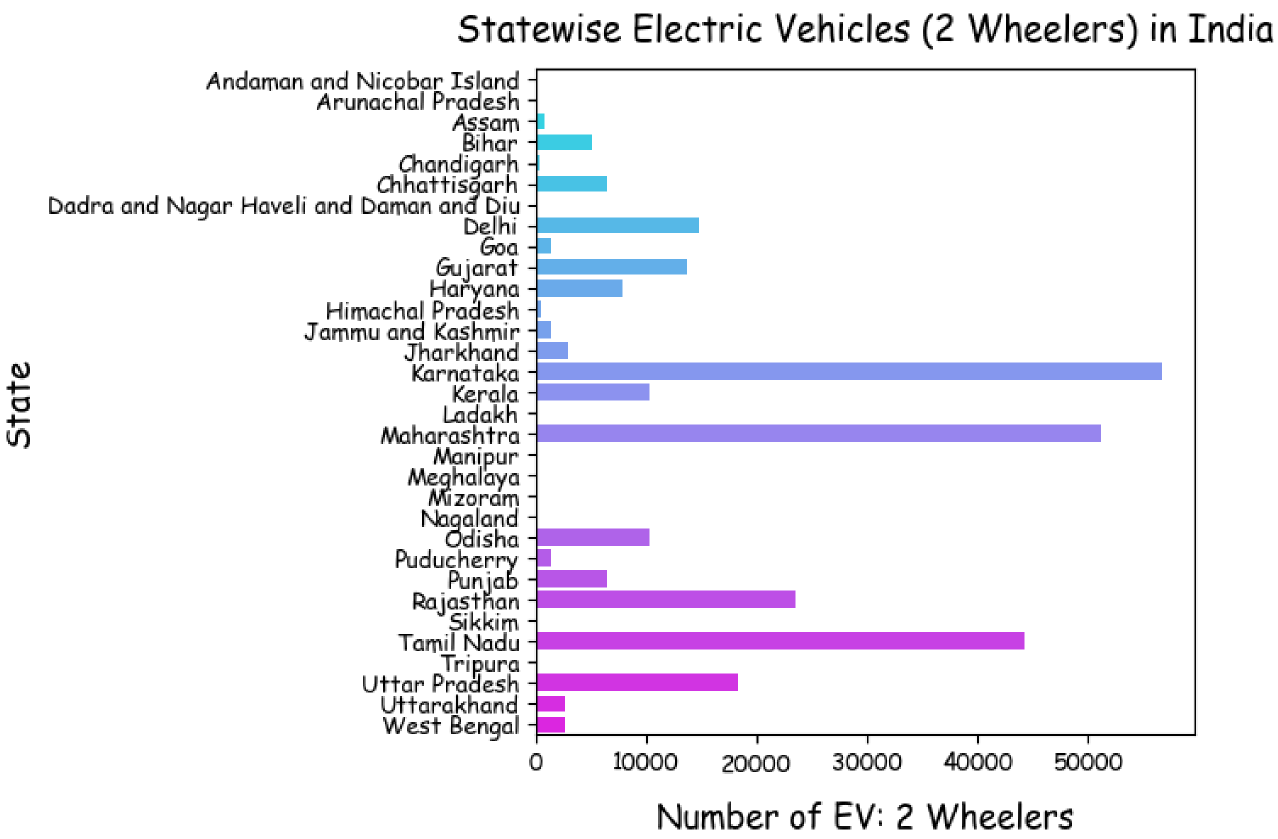
Exploratory Data Analysis:

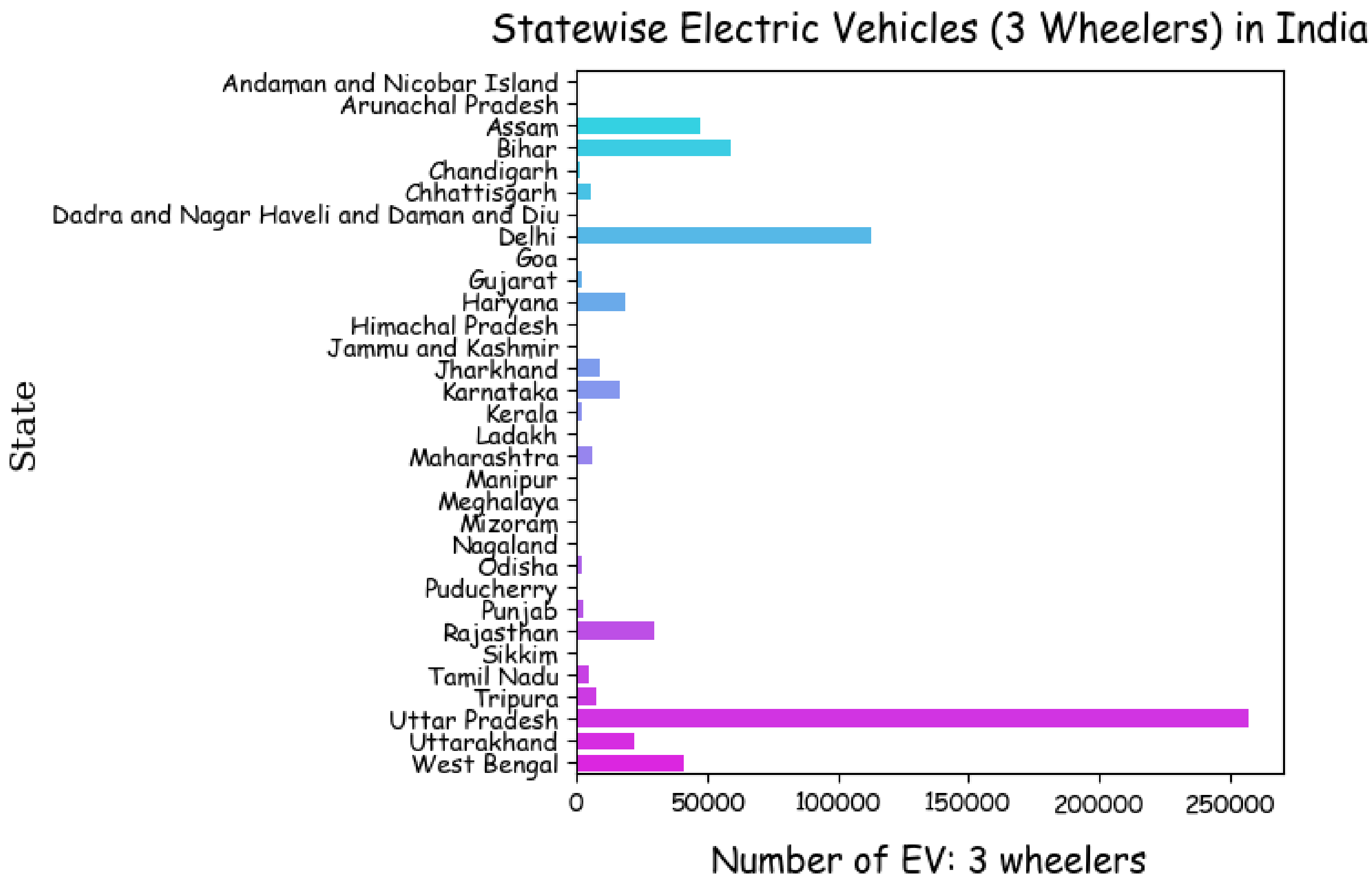
An Exploratory Data Analysis or (EDA) is a critical step in understanding and interpreting the underlying patterns and relationships within a dataset.

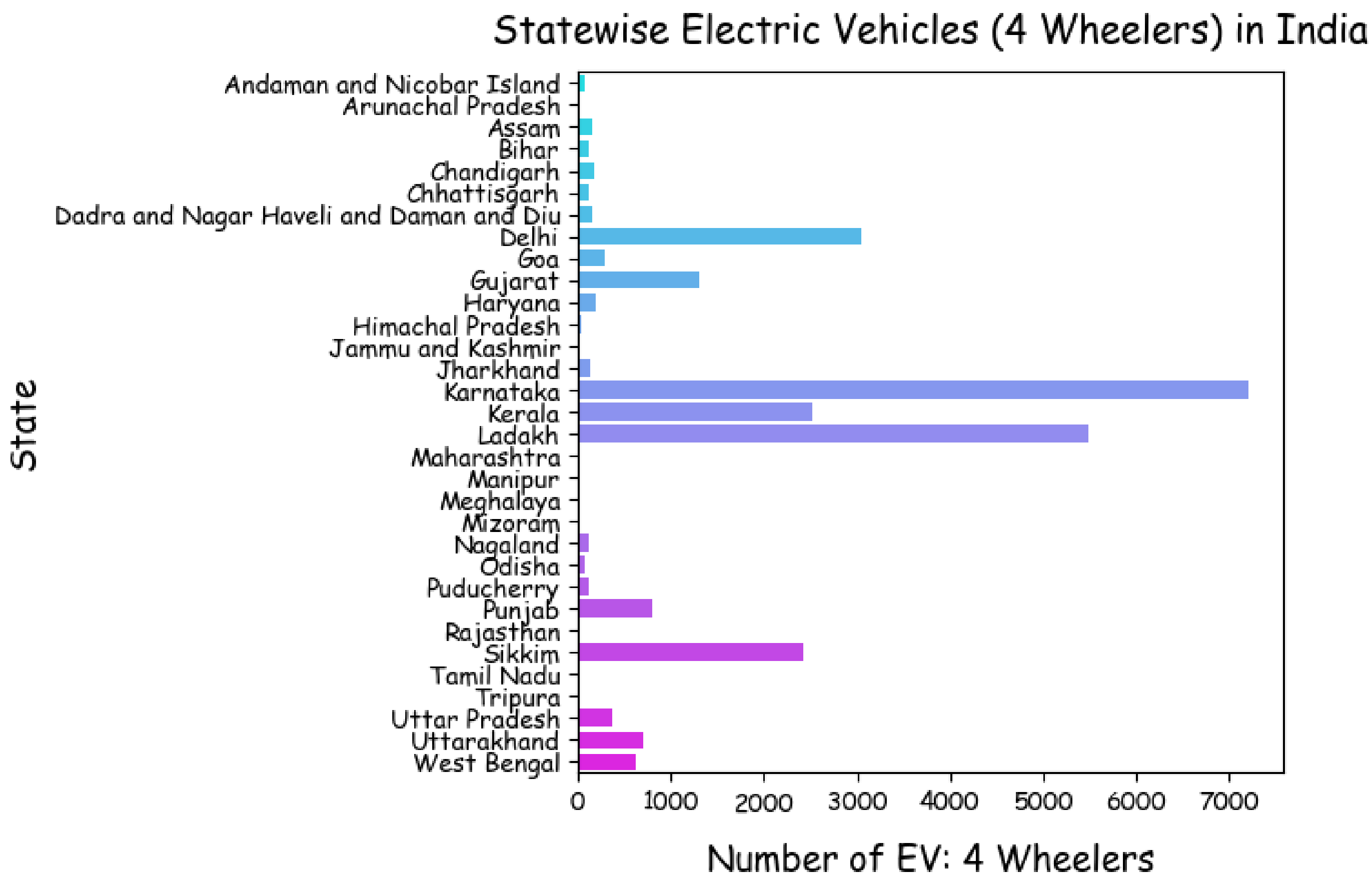
When analyzing the Electric Vehicle (EV) market in India, EDA helps in identifying key insights that can inform strategic decisions for a new EV startup.

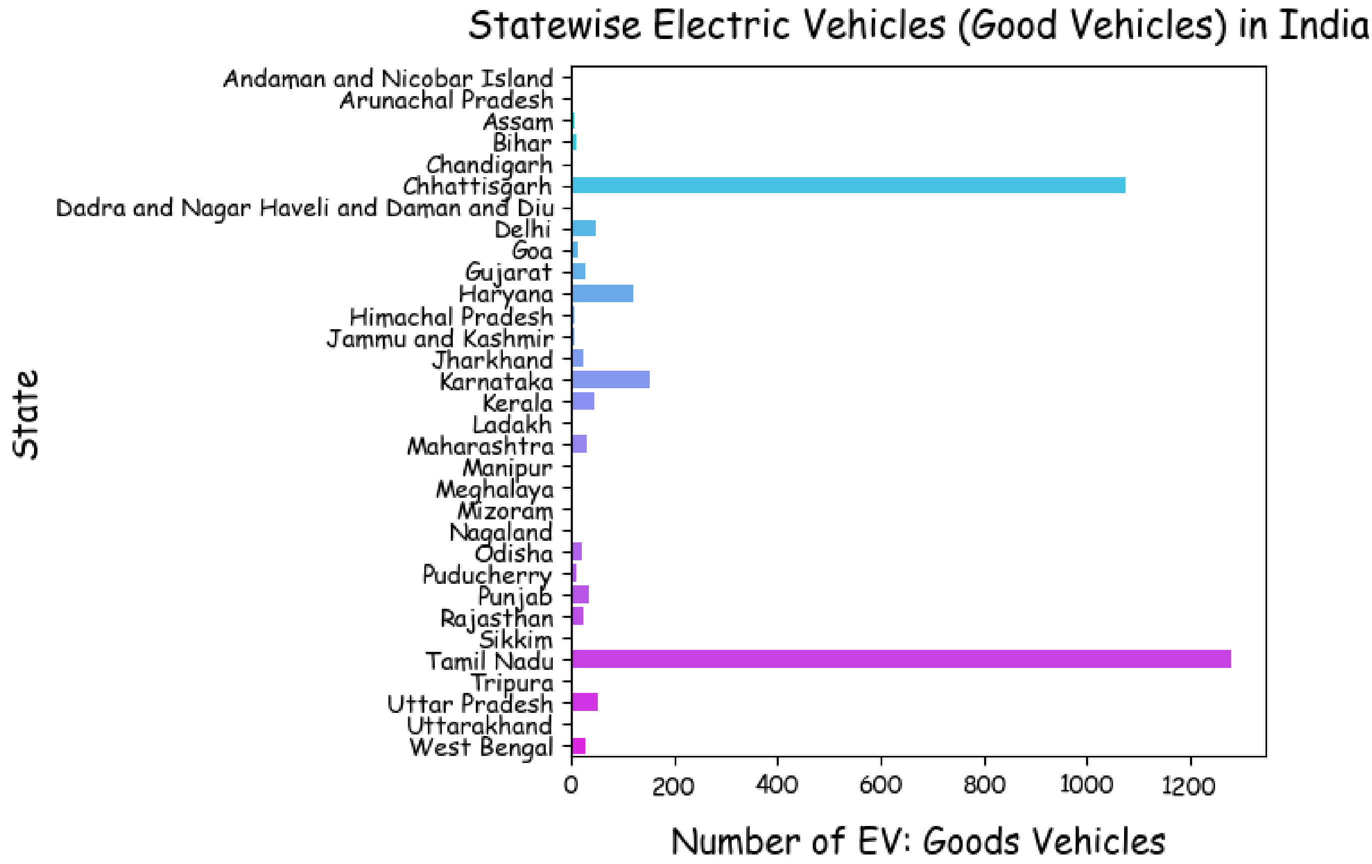
We analyzed our dataset using UNIVARIATE (analyze data over a single variable/column from a dataset) and MULTIVARIATE (analyze data by taking two variable/column from a dataset)

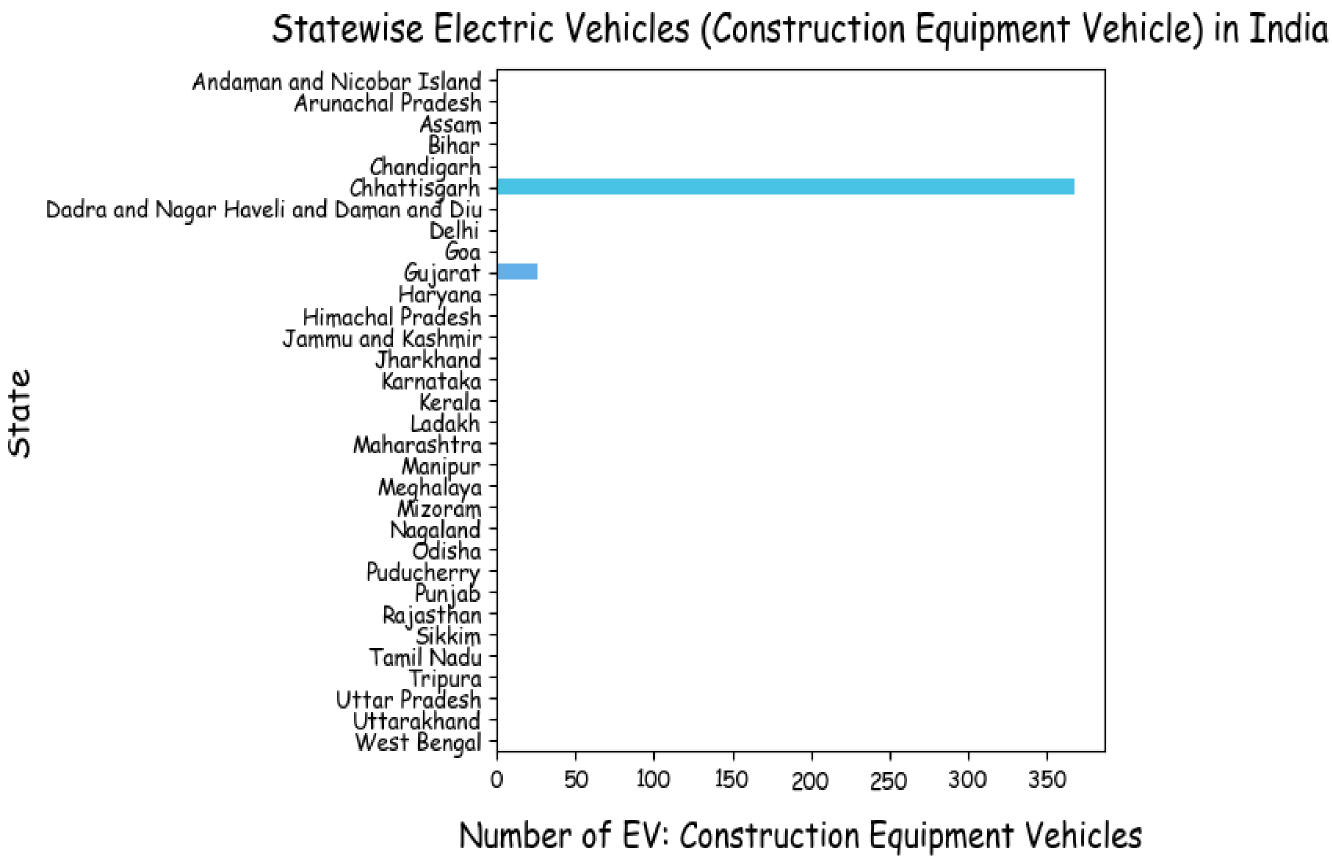
The Bar Graph shows the diversity of data geographically. We can see that we have maximum amount of vehicles(2 wheelers) of states are Karnataka, Maharashtra and Tamil Nadu and minimum amount of data from Arunachal Pradesh, Manipur ,Meghalaya , Nagaland and Tripura State.

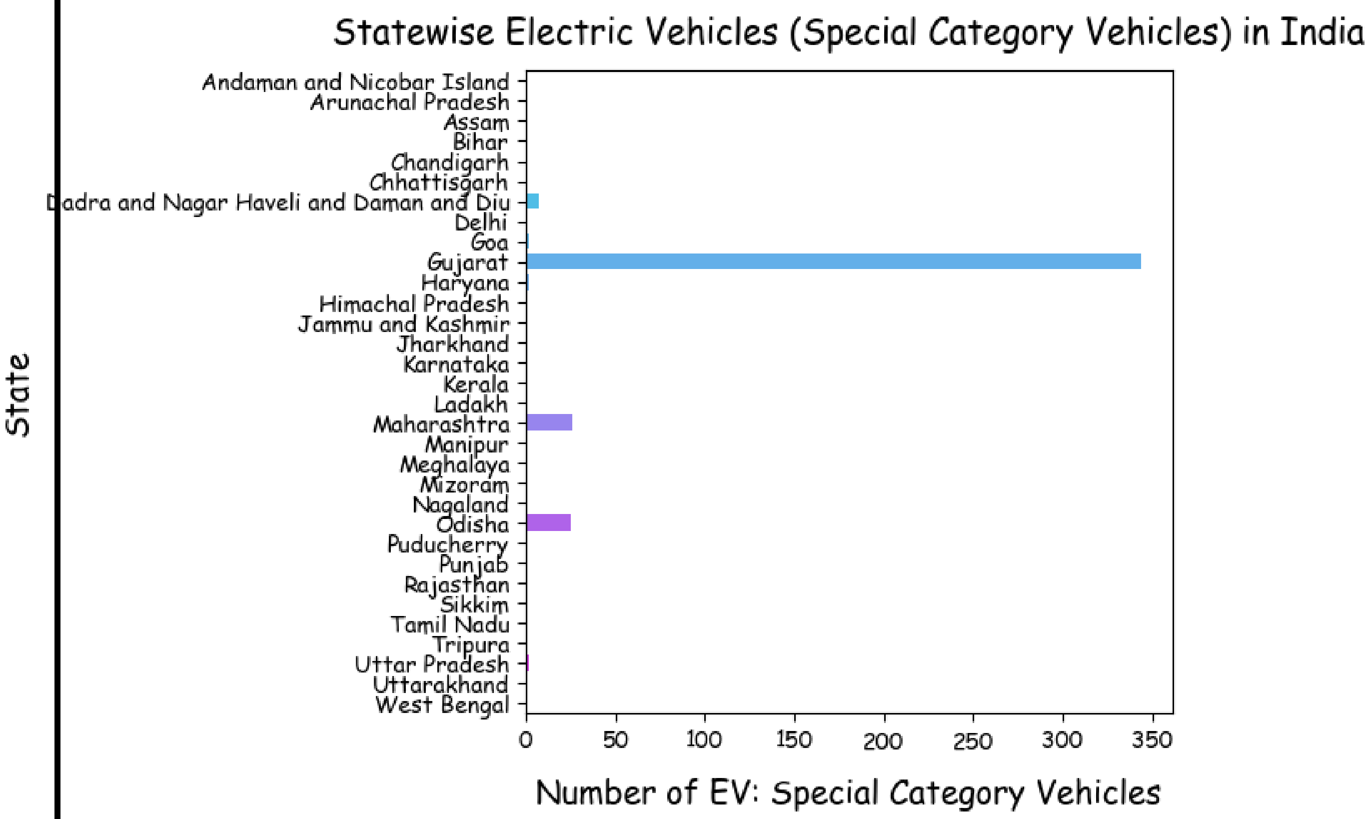



Statewise Electric Vehicles (3 Wheelers) in India
State
Andaman and Nicobar Island
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Dadra and Nagar Haveli and Daman and Diu
Delhi
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Ladakh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Tripura
Uttar Pradesh
Uttarakhand
West Bengal
0
50000
100000
150000
200000
250000
Number of EV: 3 wheelers

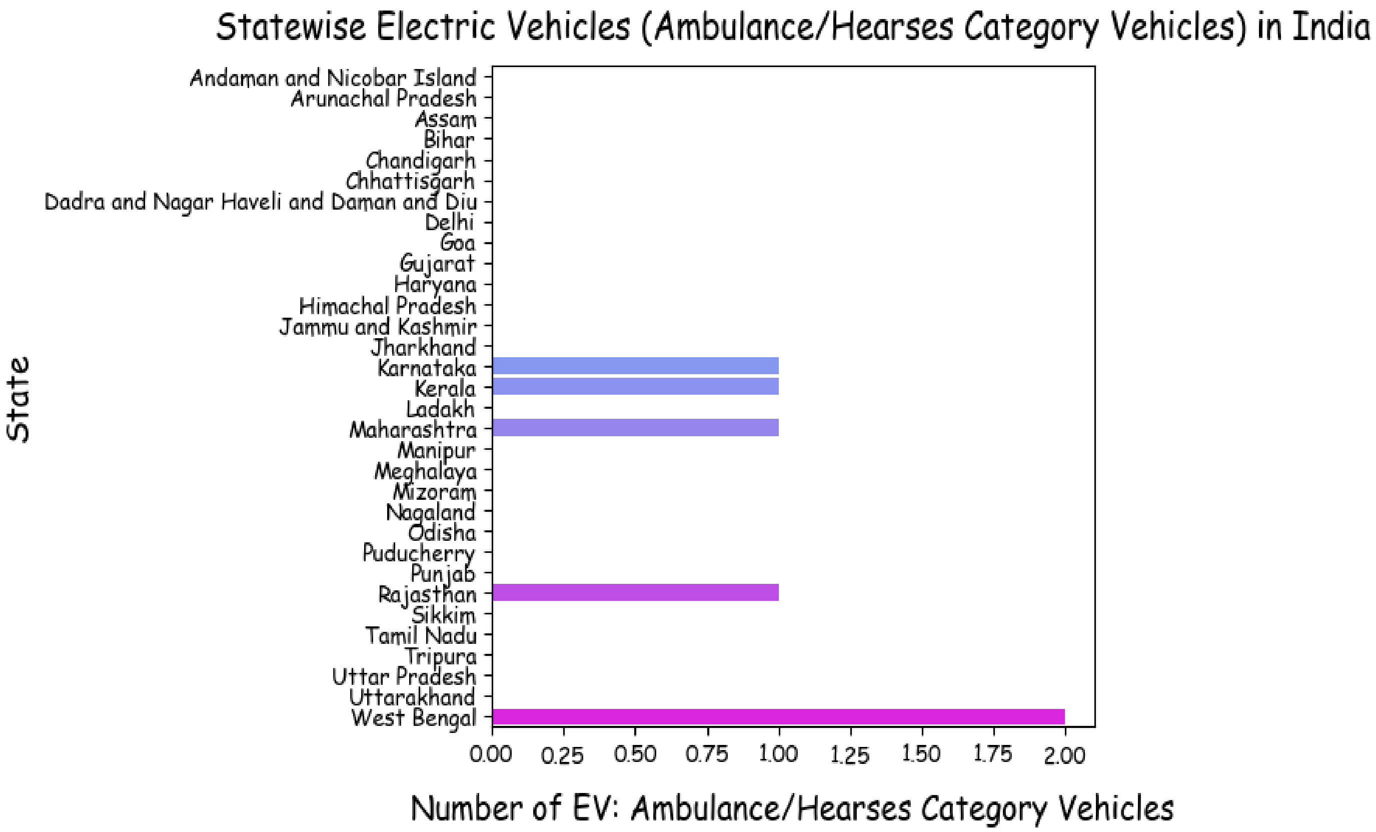

Statewise Electric Vehicles (4 Wheelers) in India
State
Andaman and Nicobar Island
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Dadra and Nagar Haveli and Daman and Diu
Delhi
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Ladakh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Tripura
Uttar Pradesh
Uttarakhand
West Bengal
0
1000
2000
3000
4000
5000
6000
7000
Number of EV: 4 Wheelers


Statewise Electric Vehicles (Good Vehicles) in India
State
Andaman and Nicobar Island
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Dadra and Nagar Haveli and Daman and Diu
Delhi
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Ladakh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Tripura
Uttar Pradesh
Uttarakhand
West Bengal
0
200
400
600
800
1000
1200
Number of EV: Goods Vehicles

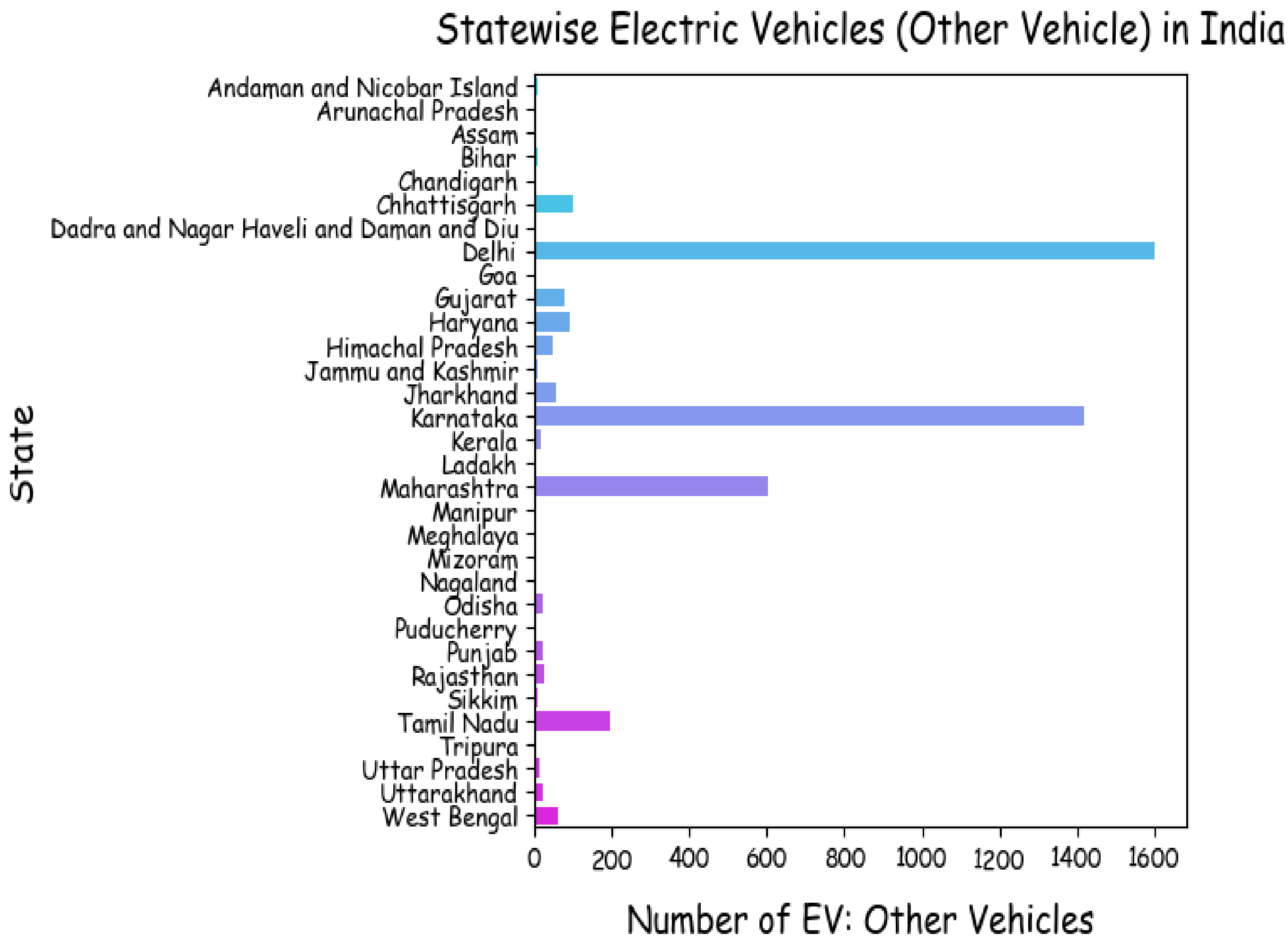

Statewise Electric Vehicles (Construction Equipment Vehicle) in India
State
Andaman and Nicobar Island
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Dadra and Nagar Haveli and Daman and Diu
Delhi
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Ladakh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Tripura
Uttar Pradesh
Uttarakhand
West Bengal
0
50
100
150
200
250
300
350
Number of EV: Construction Equipment Vehicles



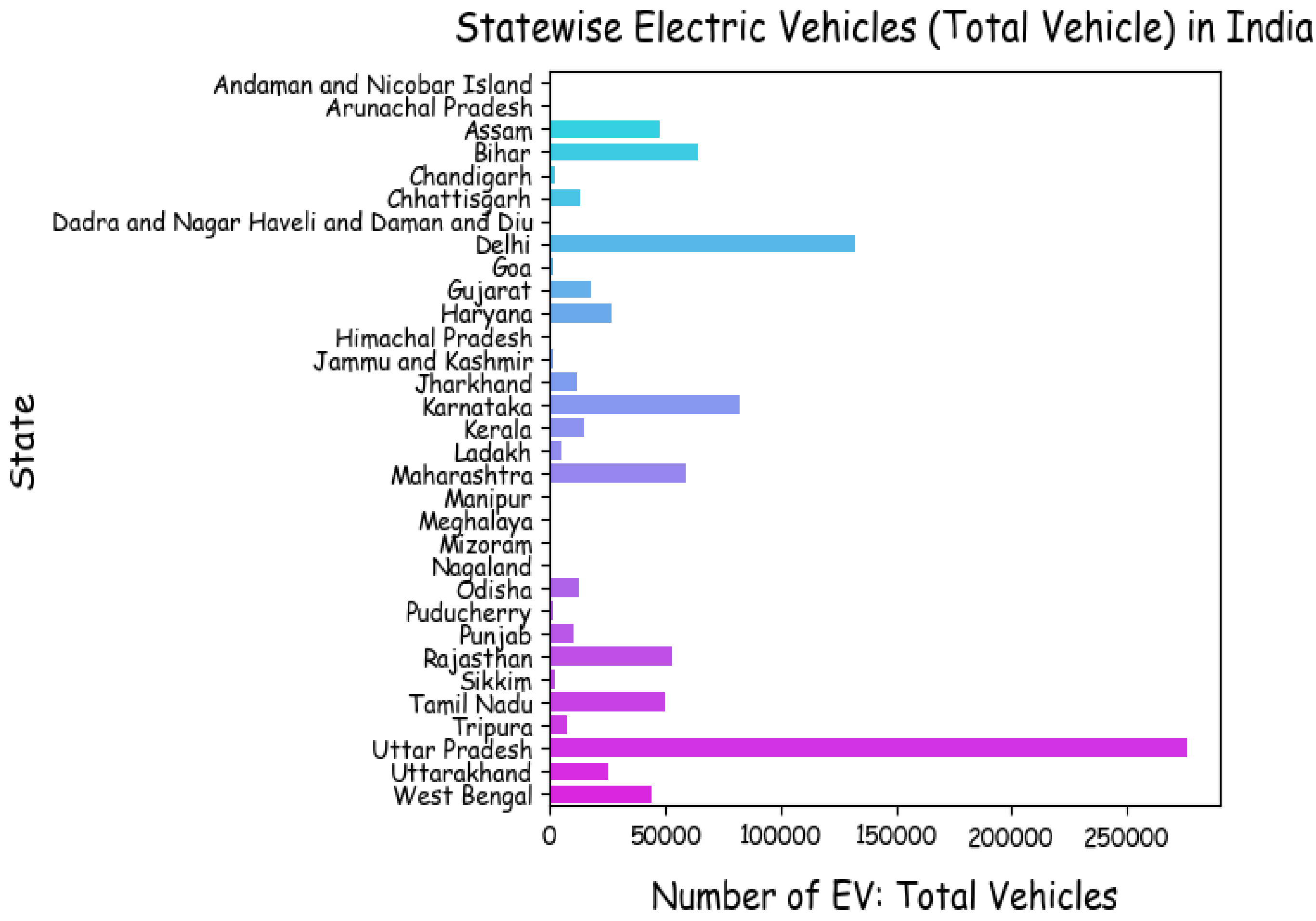
Statewise Electric Vehicles (Special Category Vehicles) in India
State
Andaman and Nicobar Island
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Dadra and Nagar Haveli and Daman and Diu
Delhi
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Ladakh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Tripura
Uttar Pradesh
Uttarakhand
West Bengal
0 50 100 150 200 250 300 350
Number of EV: Special Category Vehicles



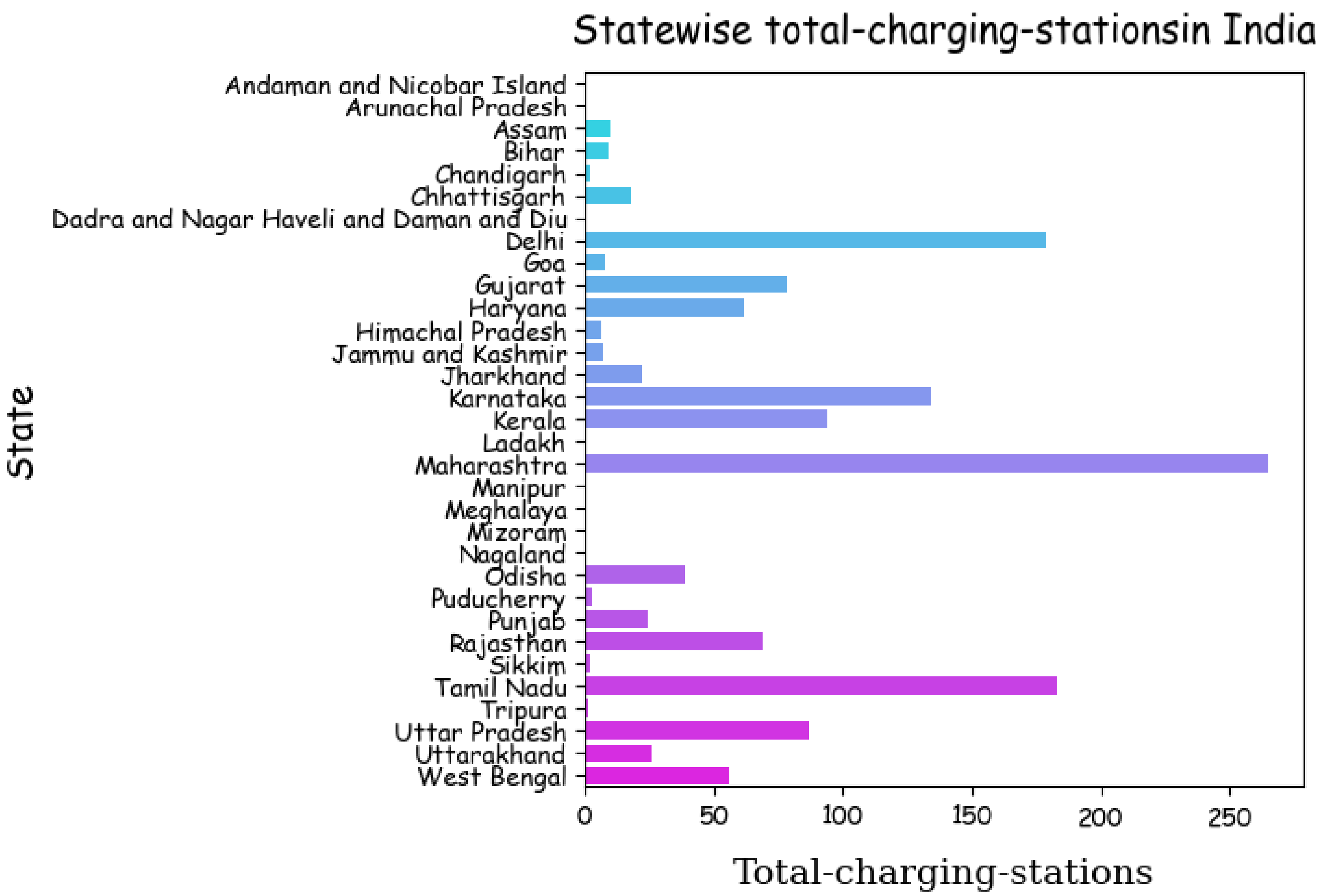
Statewise Electric Vehicles (Ambulance/Hearses Category Vehicles) in India
State
Andaman and Nicobar Island
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Dadra and Nagar Haveli and Daman and Diu
Delhi
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Ladakh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Tripura
Uttar Pradesh
Uttarakhand
West Bengal
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75 2.00
Number of EV: Ambulance/Hearses Category Vehicles



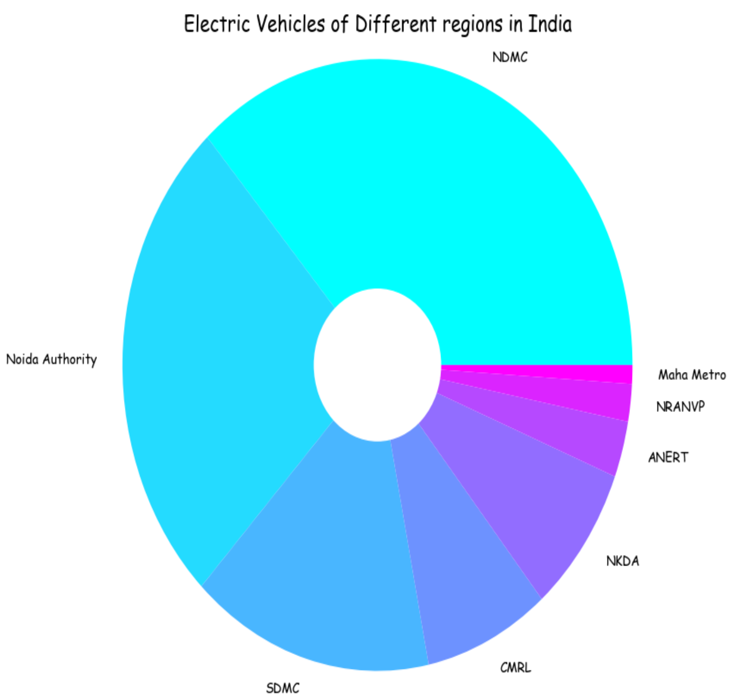
The above Bar Graph shows the number of vehicles i.e. Three Wheeler , Four Wheeler , Goods Vehicle , Public Service Vehicle , Special Category Vehicles ,Ambulance , Construction Equipment Vehicle and other vehicle state-wise.



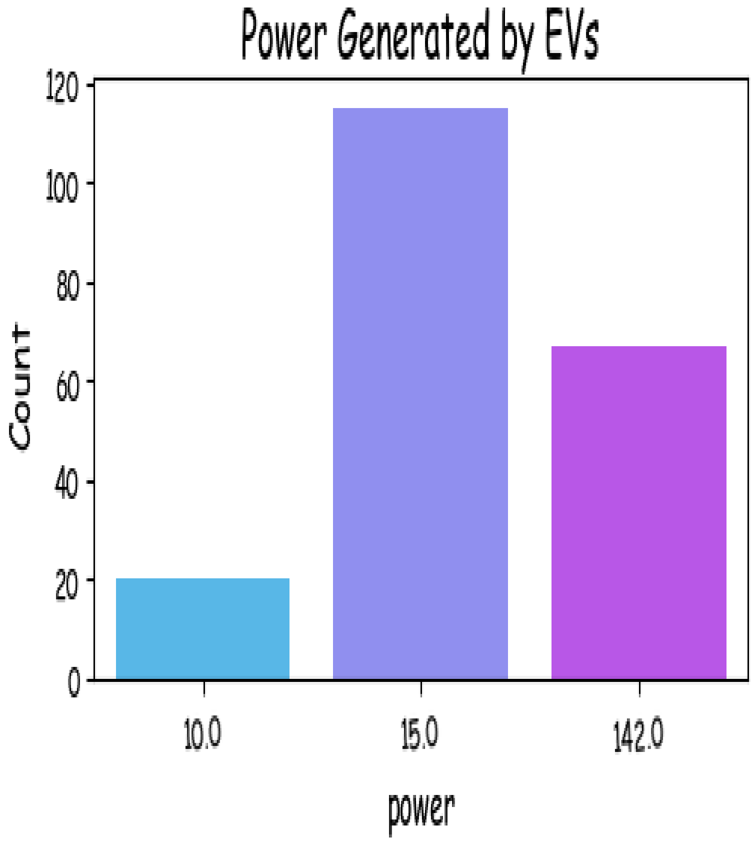
The above Bar Graph shows the Total number of vehicles State-wise where Maximum number of Electric Vehicle are in Uttar Pradesh , Delhi and Karnataka state and minimum in Andaman and Nicobar Island ,Himanchal Pradesh ,Manipur , Meghalaya ,Mizoram states



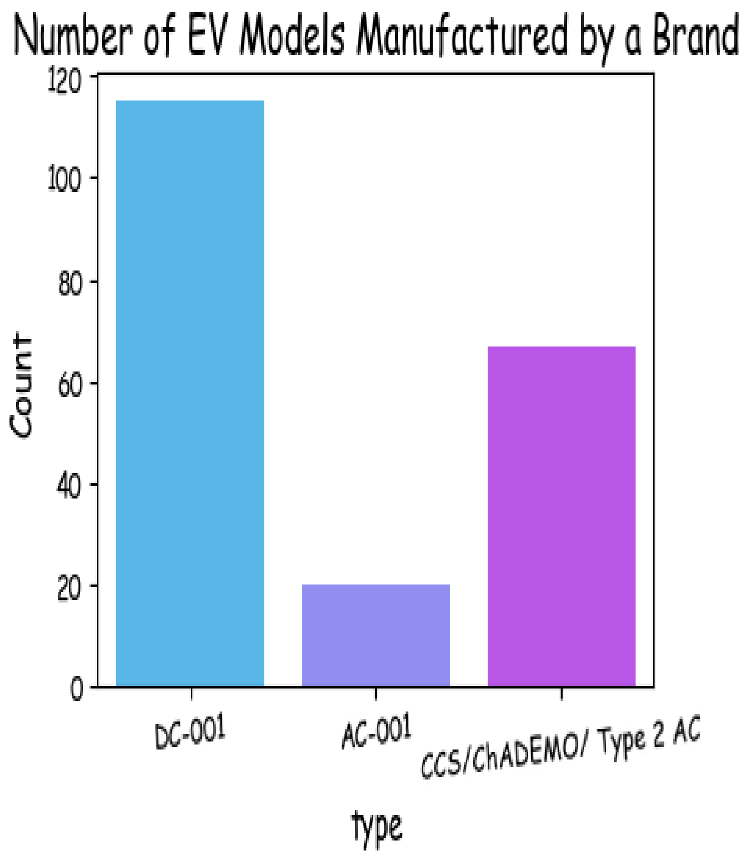
The above Bar Graph shows the Total number of Charging station State-wise where Maximum number of Electric Charging station are in Maharashtra , Delhi ,Tamil Nadu and Karnataka state and minimum in Andaman and Nicobar Island ,Himanchal Pradesh ,Manipur , Meghalaya ,Mizoram states.



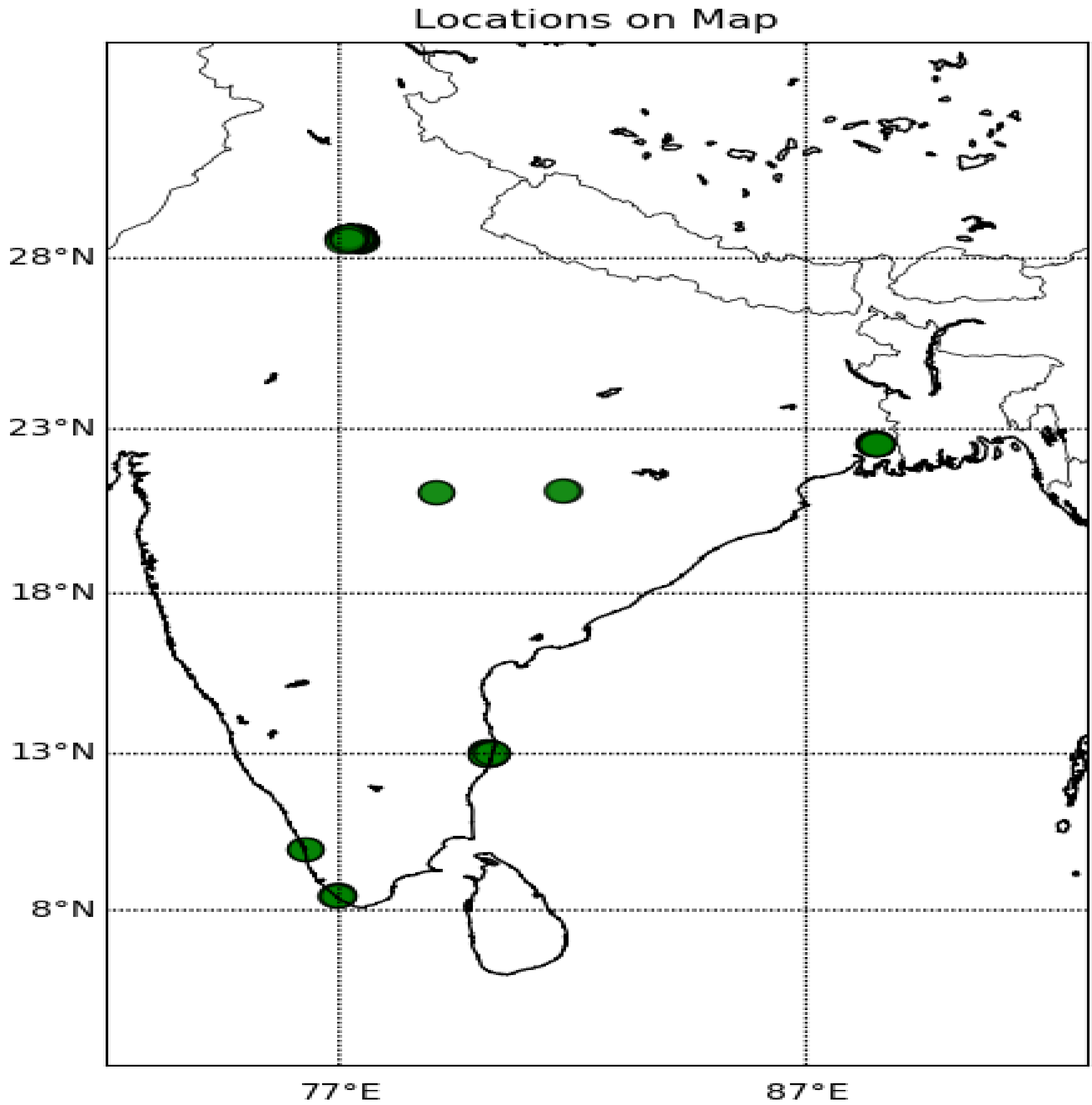
The Pie-Chart shows the Different Regions of Electric vehicle State-wise where Maximum number of region occupied by NDMC and Noida Authority and minimum in ANERT , NRANVP and Metro.



The Bar Graph shows the power generated by the Electric Vehicle where as maximum power generated by EVs is 15kW followed by 142kW and 10kW.

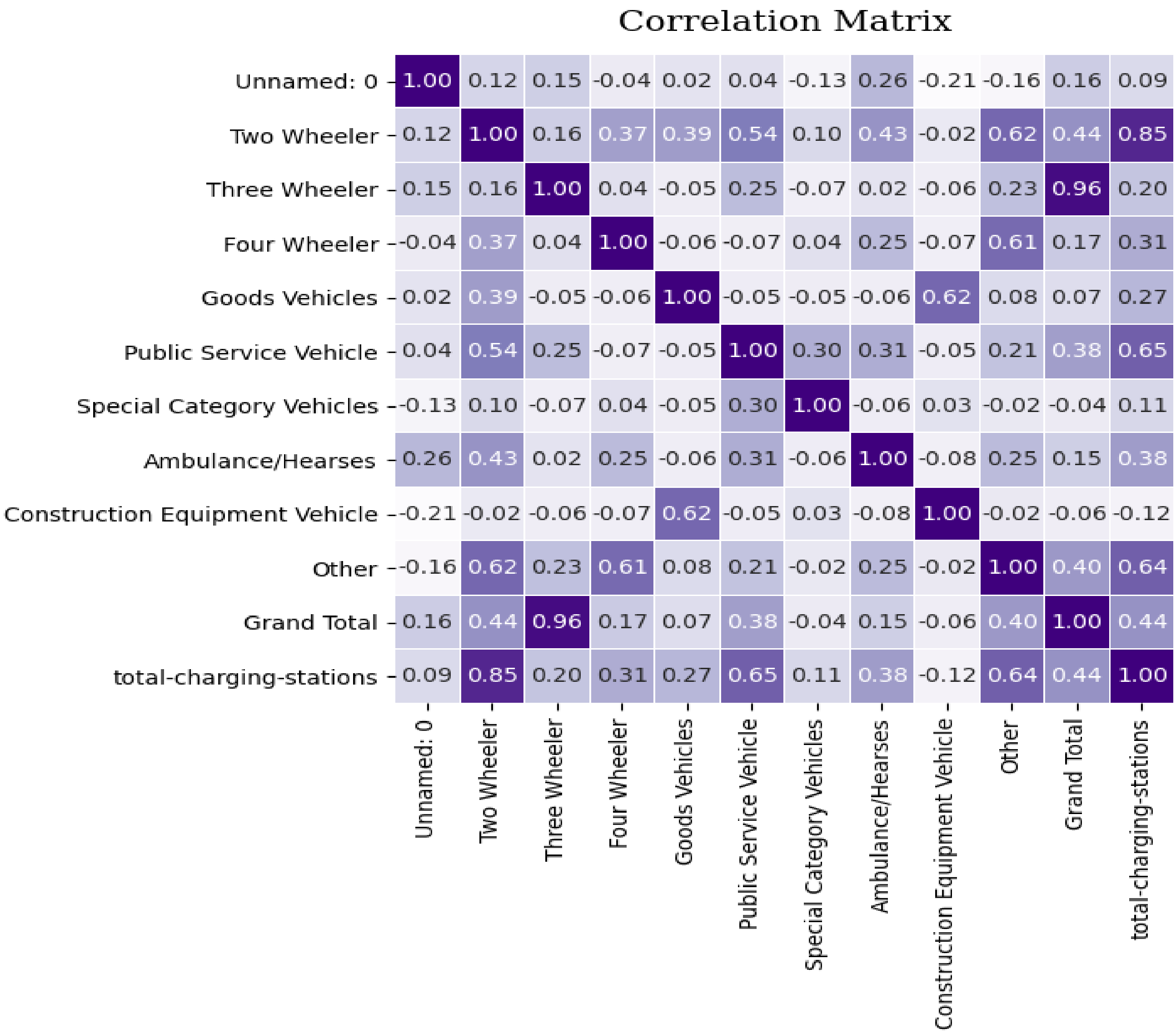


The Bar Graph shows the number of Electric Vehicle models manufactured by a brand. where as DC-001 produces maximum number of Electric Vehicles followed by CCS/ChADEMo/Type 2 AC and AC-001 Brand.

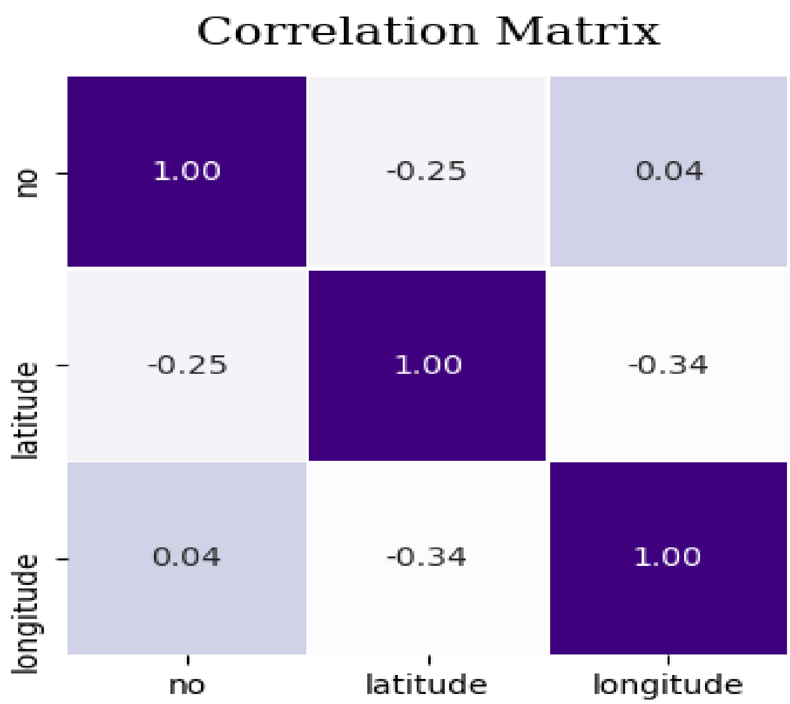


In the above figure I have plotted the latitude and longitude points on a map to visualize the geographic distribution of data.

This map provides a clear and intuitive representation of the locations, helping to identify spatial patterns and regional concentrations effectively. By using this visualization, we can better understand the geographic spread and focus areas for further analysis.



	Unnamed: 0	Two Wheeler	Three Wheeler	Four Wheeler	Goods Vehicles	Public Service Vehicle	Special Category Vehicles	Ambulance/Hearses	Construction Equipment Vehicle	Other	Grand Total	total-charging-stations
Unnamed: 0	1.00	0.12	0.15	-0.04	0.02	0.04	-0.13	0.26	-0.21	-0.16	0.16	0.09
Two Wheeler	0.12	1.00	0.16	0.37	0.39	0.54	0.10	0.43	-0.02	0.62	0.44	0.85
Three Wheeler	0.15	0.16	1.00	0.04	-0.05	0.25	-0.07	0.02	-0.06	0.23	0.96	0.20
Four Wheeler	-0.04	0.37	0.04	1.00	-0.06	-0.07	0.04	0.25	-0.07	0.61	0.17	0.31
Goods Vehicles	0.02	0.39	-0.05	-0.06	1.00	-0.05	-0.05	-0.06	0.62	0.08	0.07	0.27
Public Service Vehicle	0.04	0.54	0.25	-0.07	-0.05	1.00	0.30	0.31	-0.05	0.21	0.38	0.65
Special Category Vehicles	-0.13	0.10	-0.07	0.04	-0.05	0.30	1.00	-0.06	0.03	-0.02	-0.04	0.11
Ambulance/Hearses	0.26	0.43	0.02	0.25	-0.06	0.31	-0.06	1.00	-0.08	0.25	0.15	0.38
Construction Equipment Vehicle	-0.21	-0.02	-0.06	-0.07	0.62	-0.05	0.03	-0.08	1.00	-0.02	-0.06	-0.12
Other	-0.16	0.62	0.23	0.61	0.08	0.21	-0.02	0.25	-0.02	1.00	0.40	0.64
Grand Total	0.16	0.44	0.96	0.17	0.07	0.38	-0.04	0.15	-0.06	0.40	1.00	0.44
total-charging-stations	0.09	0.85	0.20	0.31	0.27	0.65	0.11	0.38	-0.12	0.64	0.44	1.00



	no	latitude	longitude
no	1.00	-0.25	0.04
latitude	-0.25	1.00	-0.34
longitude	0.04	-0.34	1.00

Correlation Matrix for the Features in Both of data set used.

Segmentation :

```
[98]: df2['power'] = df2['power'].apply(lambda x: float(x.split()[0]))
```

```
[99]: X = df2[['latitude', 'longitude', 'power']]
```

```
[100]: scaler = StandardScaler()
       X_scaled = scaler.fit_transform(X)
```

```
[101]: pca = PCA(n_components=3)
       X_pca = pca.fit_transform(X_scaled)
       df_pca = pd.DataFrame(X_pca, columns=['PC1', 'PC2', 'PC3'])
       df_pca.head()
```

The first step is converting the 'power' column in the data frame df2 to a float. It assumes that the values in the 'power' column are strings with numeric values followed by some text. The lambda function splits each string by spaces and takes the first part (the numeric value), then converts it to a float.

Standardization: Standardization transforms the data to have a mean of 0 and a standard deviation of 1.

Standard Scaler:

Standard Scaler is a class in scikit-learn that standardizes features by removing the mean and scaling to unit variance.

fit_transform(X) first fits the scaler on the data X (calculates the mean and standard deviation for each feature), and then transforms X by applying the standardization.

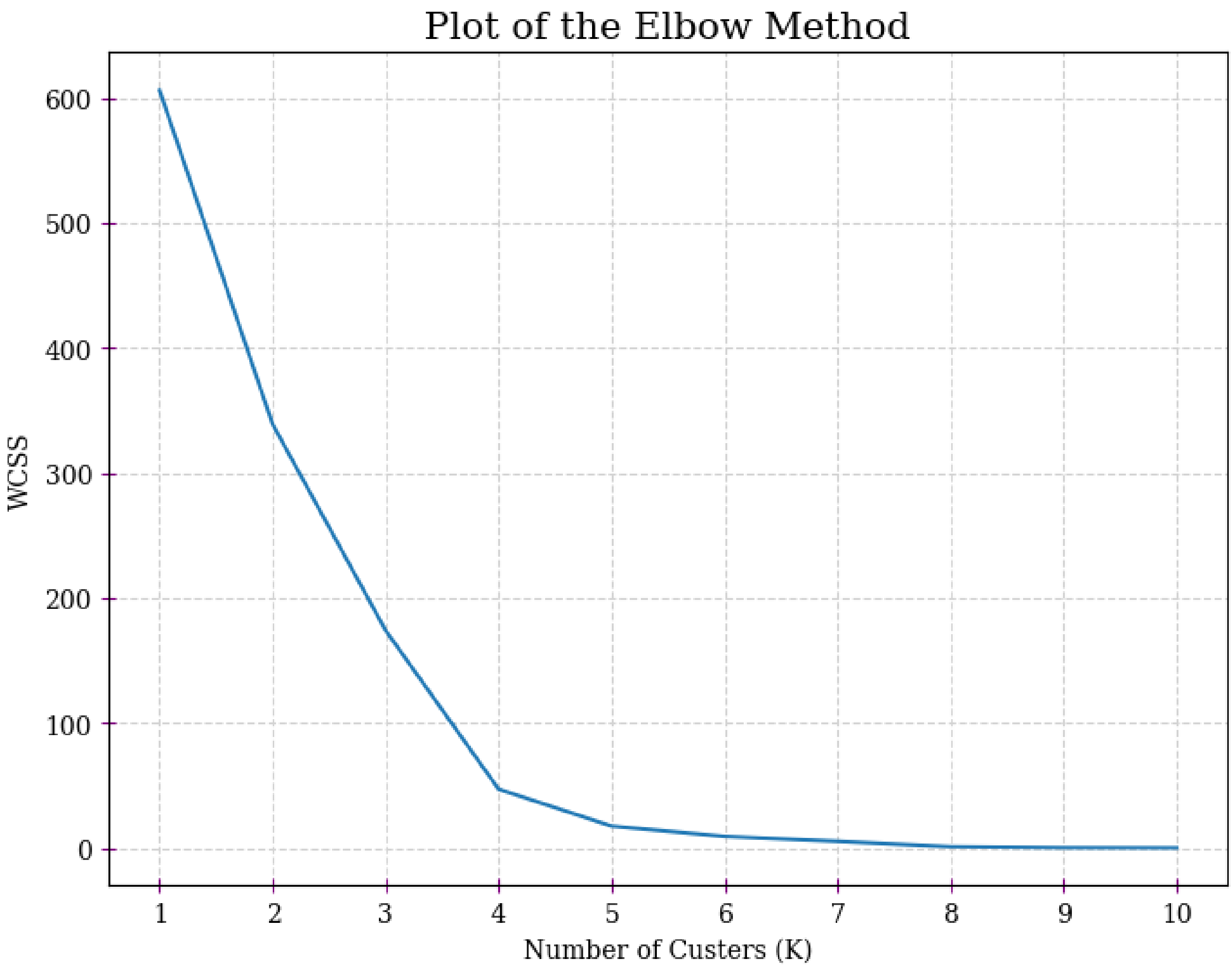
Many machine learning algorithms, especially those involving distance measurements (like PCA, k-means clustering, etc.), perform better when the data is standardized.

Principal Component Analysis (PCA):

- PCA is a statistical technique used to reduce the dimensionality of a dataset while preserving as much variance as possible.
- It identifies the directions (principal components) along which the variation in the data is maximized.

- The first principal component captures the most variance, the second captures the second most, and so on.
- Reducing the number of dimensions (features) can simplify models, reduce computational cost, and help mitigate the curse of dimensionality.
- n_components=3 specifies that we want to reduce the data to 3 principal components.
- PCA is a class in scikit-learn used to perform PCA.
- fit_transform(X_scaled) fits the PCA on the standardized data and transforms it to the new principal component space.
- PCA helps in reducing the complexity of the data and can be used to visualize high-dimensional data in a lower-dimensional space.

The Elbow Method used to determine the optimal number of clusters for K-Means clustering using the principal components obtained from PCA. It iterates through a range of cluster numbers (1 to 10), fitting the KMeans algorithm to the X_pca data for each cluster count, and computes the Within-Cluster Sum of



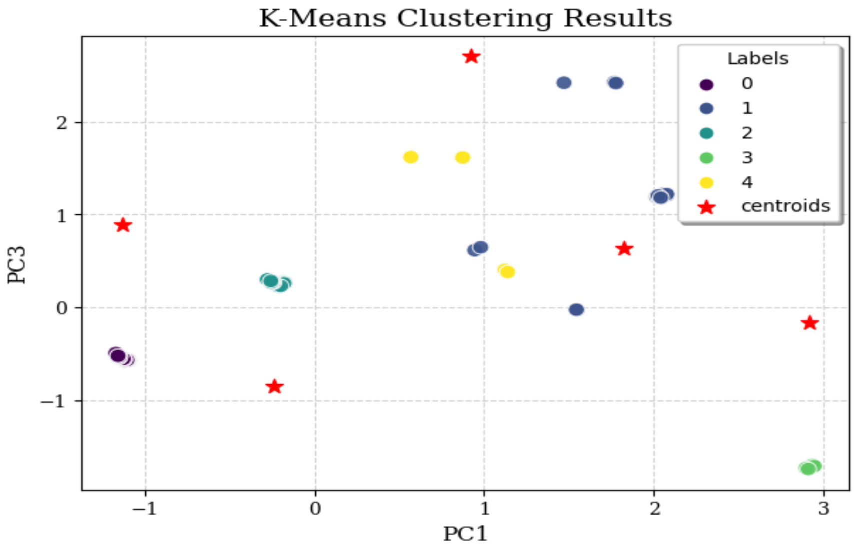
Squares (WCSS) for each model. The WCSS values are appended to a list and then plotted against the number of clusters . The 'elbow' in the plot, where the WCSS starts to decrease more slowly, indicates the ideal number of clusters.

K-Means Clustering :

K-Means Clustering is an unsupervised machine learning algorithm used to partition a dataset into a set of distinct, non-overlapping groups (clusters). The algorithm follows these steps:

- Select k initial cluster centroids (means). This can be done randomly or using methods like 'k-means++' which aims to select initial centroids that are far apart.
- Assign each data point to the nearest centroid based on a distance metric (usually Euclidean distance).
- Recalculate the centroids as the mean of all data points assigned to each cluster.
- Repeat the assignment and update steps until the centroids no longer change significantly or a predefined number of iterations is reached.
- The goal is to minimize the variance within each cluster (within-cluster sum of squares, WCSS).

K-Means clustering partitions data into k clusters by iteratively assigning data points to the nearest centroid and updating the centroids based on the mean of assigned points. The algorithm with number clusters, fits it to the PCA-transformed data, and prints and counts the cluster labels, which are then added to the original Data Frame for further analysis.



1. Machine Learning Model Used in the Second Project

In this project, K-Means clustering was the primary machine learning algorithm used for market segmentation.

This algorithm is particularly effective for partitioning a dataset into distinct groups or clusters based on similarity. K-Means clustering operates by iteratively assigning data points to the nearest centroid and then recalculating the centroids based on the current cluster memberships.

This process continues until the centroids stabilize. In our project, we applied K-Means clustering to segment electric vehicle data based on geographical location (latitude and longitude) and power usage.

The resulting clusters helped us identify distinct segments of vehicles with similar characteristics, facilitating targeted marketing strategies and resource allocation.

2. Final Conclusion & Insights

The final analysis provided valuable insights into the segmentation of electric vehicles based on their geographical distribution and power usage.

We identified five distinct clusters, each representing a unique segment of the market. Key insights included identifying high-power usage clusters in urban centers, suggesting a need for enhanced charging infrastructure in these areas.

Additionally, the segmentation revealed geographical patterns in electric vehicle distribution, allowing for more targeted marketing and operational strategies.

For example, the high-density clusters in city centers indicated a significant demand for quick and high-capacity charging solutions, whereas the more dispersed clusters highlighted regions where expansion of charging networks could capture new market opportunities.

3. Improvement on Market Segmentation Project

Given additional time and budget, I would enhance the market segmentation project by expanding the dataset and exploring additional machine learning models.

I would collect more comprehensive data, including columns such as vehicle type, age of the vehicle, driving patterns, charging habits, and customer demographics.

This enriched dataset would provide deeper insights into the behavior and preferences of different market segments. Additionally, I would experiment with other clustering algorithms like DBSCAN (Density-Based Spatial Clustering of Applications with Noise) for identifying clusters of varying densities and hierarchical clustering for more hierarchical segment relationships.

Incorporating supervised learning models, such as Decision Trees or Random Forests, could also help predict segment membership based on new data, further enhancing the robustness of our market segmentation.

4. Estimated Market Size

The estimated market size for the electric vehicle domain, considering it as a non-segmented market, is substantial.

As of recent industry reports, the global electric vehicle market size was valued at approximately USD 250 billion in 2023 and is expected to grow at a compound annual growth rate (CAGR) of over 20% in the coming years.

This growth is driven by increasing environmental concerns, government incentives, and advancements in battery technology.

The rising adoption of electric vehicles across various regions highlights the significant market potential and the importance of effective segmentation to capture and serve different customer segments effectively.

5. Key Variables for Optimal Market Segments

To create the most optimal market segments for the electric vehicle domain, the top four variables/features are:

Geographical Location (Latitude and Longitude): This feature helps in understanding the regional distribution of electric vehicles, which is crucial for infrastructure planning and targeted marketing.

Power Usage: Analyzing power consumption patterns allows for the identification of high-usage segments, informing the development of tailored charging solutions and services.

Vehicle Type: Differentiating between types of electric vehicles (e.g., sedans, SUVs, trucks) helps in understanding specific needs and preferences associated with each vehicle type, enabling more customized offerings.

Charging Habits: Data on how and where vehicles are charged (home, work, public stations) provides insights into user behavior and helps in designing better charging networks and services.

In conclusion, a comprehensive understanding of these variables will enable the creation of precise and actionable market segments, facilitating more effective business strategies and improved customer satisfaction.

The Data set and Codes are given below in Github link.

Github link