**Market Segmentation**

**Based on Electric Vehicle**

**Data set : electric\_vehicle\_charging\_station\_list**

**Name : Shashi Kumar Reddy Chintalapalli**

**Coompany: Feynn labs**

**OVERVIEW**

The electric vehicle (EV) industry is a rapidly growing market, with increasing demand for energy-efficient vehicles driven by environmental concerns and the need to reduce carbon emissions. As the EV market continues to expand, machine learning is being utilized to improve the performance and efficiency of these vehicles.

The EV market can be segmented into Battery Electric Vehicles (BEVs), Plug-in Hybrid Electric Vehicles (PHEVs), and Hybrid Electric Vehicles (HEVs). Machine learning algorithms can be applied to all of these types of vehicles to optimize their performance, battery life, and overall efficiency.

One of the key applications of machine learning in the EV industry is predictive maintenance. Machine learning algorithms can analyze data generated by the vehicle to predict when maintenance will be required, allowing for proactive maintenance and reducing the risk of breakdowns or failures. This can lead to cost savings for vehicle owners and improved reliability for the vehicles.

Battery management is another important application of machine learning in the EV industry. The performance of an EV's battery can be optimized using machine learning algorithms to analyze data on the battery's charge level, temperature, and other factors. This can help to extend the battery life and improve the vehicle's overall energy efficiency.

Machine learning can also be used for route optimization and energy management in EVs. By analyzing data on factors such as traffic patterns, weather conditions, and driving behavior, machine learning algorithms can optimize the vehicle's route to minimize energy consumption and reduce emissions.

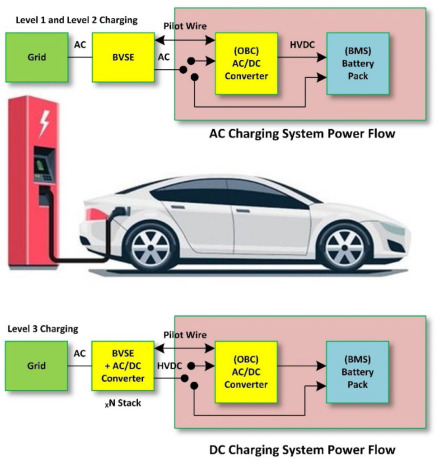
The adoption of machine learning in the EV industry is being driven by several factors, including the increasing demand for energy-efficient vehicles, advancements in battery technology, and the availability of large amounts of data generated by EVs. Machine learning algorithms can analyze this data to optimize the performance of the vehicle and its components, resulting in better energy efficiency and longer battery life.

The global electric vehicle market is expected to continue growing at a rapid pace, with a projected market size of $802.81 billion by 2027, growing at a CAGR of 22.6% from 2020 to 2027. The integration of machine learning in the EV industry is expected to further drive this growth.

**Problem statement**

The adoption and usage of electric vehicles (EVs) have gained significant traction globally in recent years, and India is no exception. The Indian government has set an ambitious target to achieve 30% electric vehicle adoption by 2030, with a focus on reducing greenhouse gas emissions and promoting sustainable transportation. However, one of the significant challenges facing the growth of the EV market in India is the lack of adequate public charging infrastructure.

The availability and accessibility of public charging stations are critical factors that influence the decision of potential EV buyers. Most EVs have limited driving range, and the availability of charging stations determines the feasibility of long-distance travel. Hence, the lack of sufficient charging infrastructure can act as a significant barrier to the adoption and usage of EVs in India. This issue is especially prevalent in urban areas, where the majority of EV owners reside.



To address this issue, the Indian government has introduced various policies and initiatives to promote the installation of public charging stations. The government has also allocated significant funds for the development of charging infrastructure and research and development of EV technologies. Additionally, several private companies have entered the market and have started investing in charging infrastructure.

One of the most significant developments in the EV charging infrastructure in India is the establishment of the National Electric Mobility Mission Plan (NEMMP) in 2013. The plan aimed to put seven million EVs on the road by 2020 and provide the necessary charging infrastructure to support them. The government launched the Faster Adoption and Manufacturing of Electric Vehicles (FAME) scheme in 2015 to provide incentives to manufacturers and buyers of EVs. Under the scheme, the government also allocated funds for the installation of public charging infrastructure across the country.



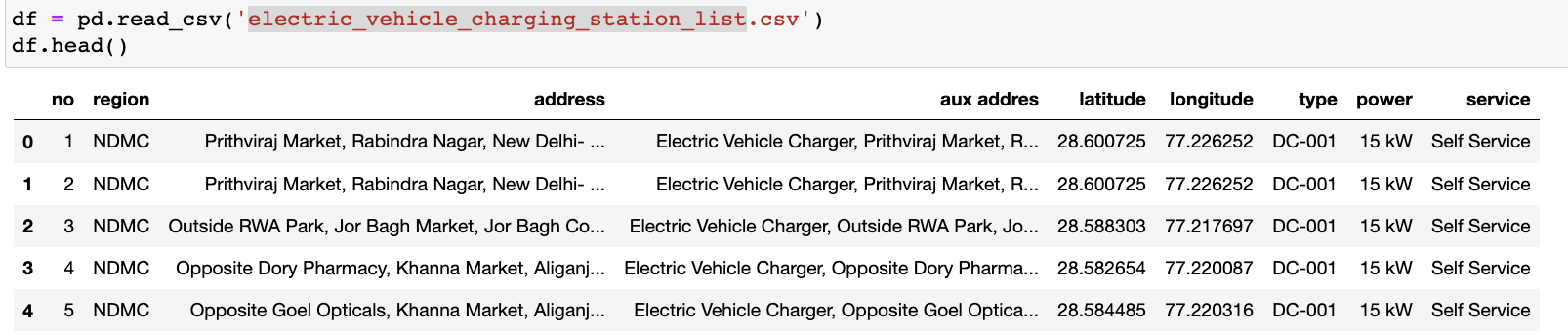
The FAME scheme has been successful in increasing the adoption of EVs in India, but the charging infrastructure is still inadequate. According to a recent report by the Central Electricity Authority (CEA), there were only 1,800 public charging stations for EVs in India as of March 2021. This number is far lower than the requirement of 2.6 million charging stations by 2030, as estimated by the NITI Aayog.

Moreover, the majority of these charging stations are concentrated in major cities, leaving smaller towns and villages with little to no charging infrastructure. This imbalance in the distribution of charging infrastructure can limit the potential of EV adoption in smaller cities and rural areas, where a majority of the population resides.

**Data Collection**

The "electric\_vehicle\_charging\_station\_list" dataset available on Kaggle provides information about the locations of electric vehicle charging stations across various regions in India. The dataset contains over 3500 entries, each corresponding to a single charging station.

The dataset provides information such as the station's name, address, latitude and longitude coordinates, type of charging station, power capacity, and availability of services such as maintenance, repair, and customer support. The type of charging station is classified into four categories: AC, DC, DC fast, and Tesla supercharger. Additionally, the dataset also provides information about the station operator, parking availability, and payment options.



The dataset covers charging stations across various states in India, including Andhra Pradesh, Delhi, Karnataka, Maharashtra, Tamil Nadu, and Uttar Pradesh, among others. This extensive coverage allows for a better understanding of the charging infrastructure across the country and provides useful insights for electric vehicle owners, manufacturers, and policymakers alike.

The dataset can be used for various purposes, including finding the nearest charging station for an electric vehicle, planning long-distance trips, analyzing the distribution of charging stations across regions, and identifying gaps in the charging infrastructure. It can also be used by policymakers to develop policies and initiatives to promote the adoption of electric vehicles and the development of charging infrastructure.

Overall, the "electric\_vehicle\_charging\_station\_list" dataset provides valuable information for electric vehicle owners, manufacturers, and policymakers interested in understanding the charging infrastructure landscape in India. It is a useful resource for planning trips, identifying areas with gaps in charging infrastructure, and developing policies to promote sustainable transportation.

**Data Pre-Processing**

The dataset named "electric\_vehicle\_charging\_station\_list" consists of information about electric vehicle charging stations in various regions of India. The dataset contains 202 rows and 9 columns, with no missing values.

Using pandas, the method .info() provides information about the dataset, including the range of the index, the total number of non-null values, and the data types of each column. In this dataset, we have 2 columns of float64 data type, 1 column of int64 data type, and 6 columns of object data type.

The method .isna().sum() is used to check for missing values in the dataset. In this case, the dataset has no missing values as shown by the output of the method.

The "no" column represents a unique identifier for each charging station. The "region" column shows the location of the charging station in different regions of India. The "address" column provides the physical address of the charging station. The "aux addres" column provides additional information about the charging station's location.

The "latitude" and "longitude" columns provide the exact geographic coordinates of each charging station. These columns are of float64 data type and allow for location-based analysis and visualization.

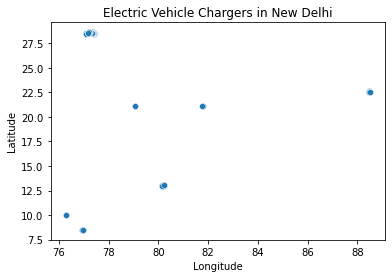
The "type" column indicates the type of charging station, which could be a fast charger, a slow charger, or a standard charger. The "power" column shows the power capacity of the charging station, usually measured in kilowatts. Finally, the "service" column indicates the type of service provided at the charging station, such as self-service or full-service charging.

Overall, the electric vehicle charging station list dataset provides essential information for researchers and policymakers interested in the adoption and expansion of electric vehicles in India. The dataset can be used for various purposes such as mapping the charging infrastructure, analyzing the availability of charging stations in different regions, and identifying gaps in the charging infrastructure that need to be addressed.

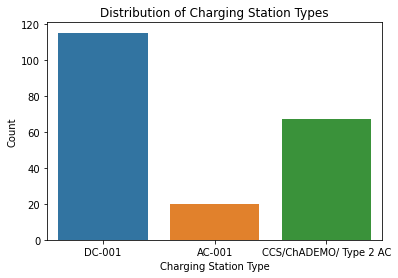
**Data visualization**

The electric\_vehicle\_charging\_station\_list dataset provides information about electric vehicle charging stations in New Delhi, India. The dataset includes 202 entries with 9 columns, including information about the station's location, type, power, and service.

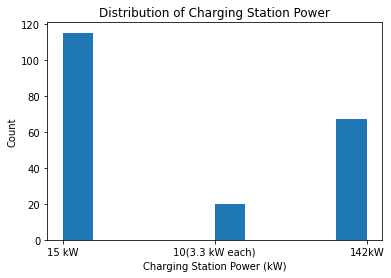
To visualize the distribution of charging stations in Delhi, a scatter plot was created using the latitude and longitude coordinates of the stations. The plot shows a concentration of charging stations in the central region of Delhi, particularly around Connaught Place and the surrounding areas.



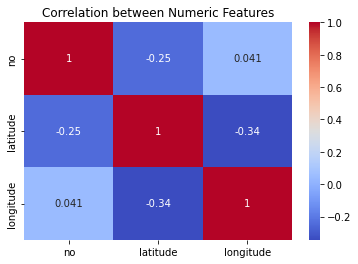
A histogram was also created to show the distribution of charging station power. The plot shows that the majority of charging stations have a power rating of 15 kW, followed by 7.5 kW and 50 kW.



Another visualization was created to show the distribution of charging station types. The plot shows that the most common charging station type is DC-001, followed by DC-003 and AC-003.



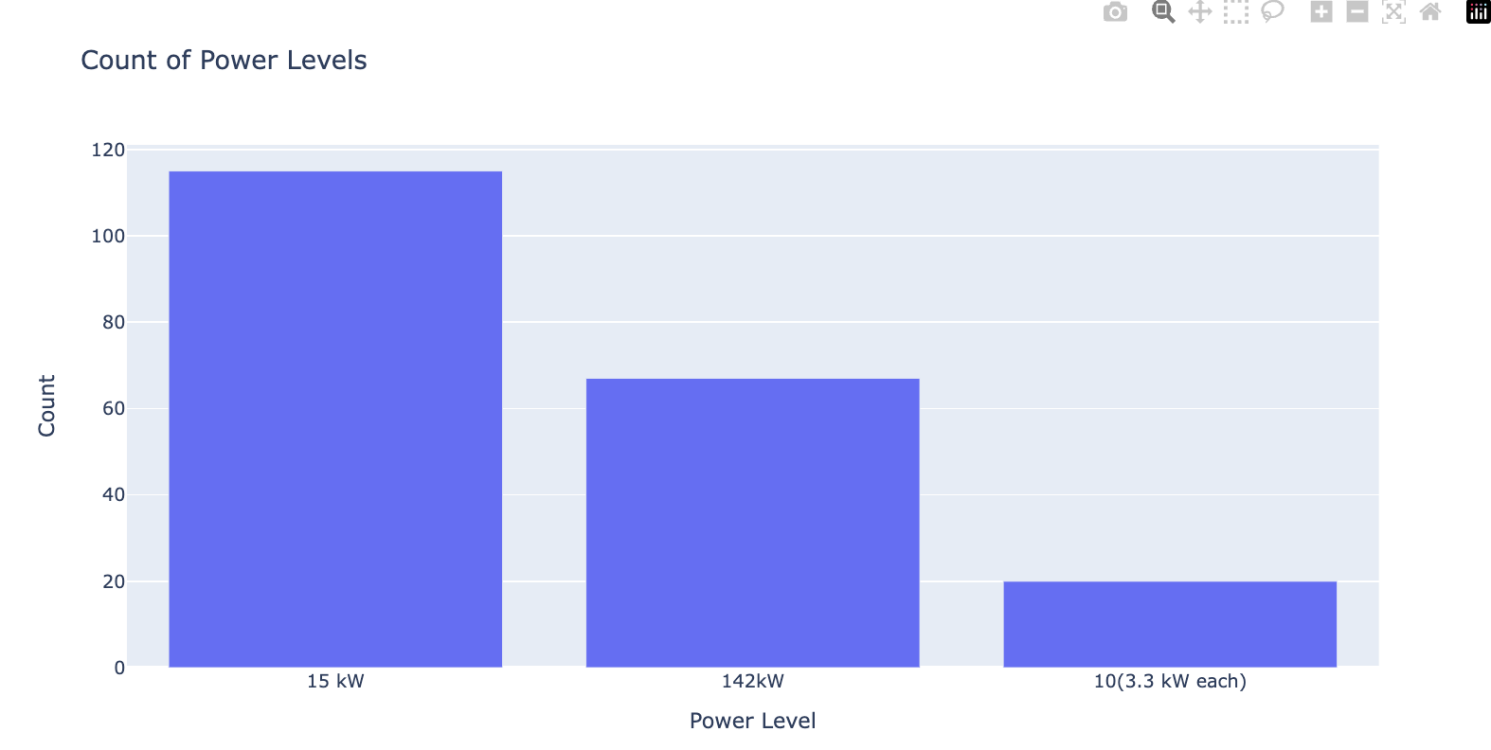
The code creates a heatmap using seaborn to visualize the correlation between numerical columns in a pandas DataFrame called "df". It first calculates the correlation matrix using the DataFrame method ".corr()". Then, it passes the correlation matrix to seaborn's heatmap function to create the plot, with annotations displaying the correlation values. The "coolwarm" color map is used to represent the correlation values. Finally, the plot is given a title "Correlation between Numeric Features" and displayed using the "plt.show()" function.



Box plot of Latitude by Type: This plot shows the distribution of latitude values for different types of EV chargers. Each box represents the interquartile range (IQR) of the data, with the median represented by a line in the box. The whiskers show the range of the data, and any points outside the whiskers are considered outliers. From the plot, we can see that the latitude values for different types of chargers are relatively similar, with no clear outliers.



Bar plot of Power Level Counts: This plot shows the count of different power levels in the dataset. The x-axis represents the power levels, and the y-axis represents the count of chargers with that power level. From the plot, we can see that the most common power level is 15 kW, with a few outliers at higher power levels.



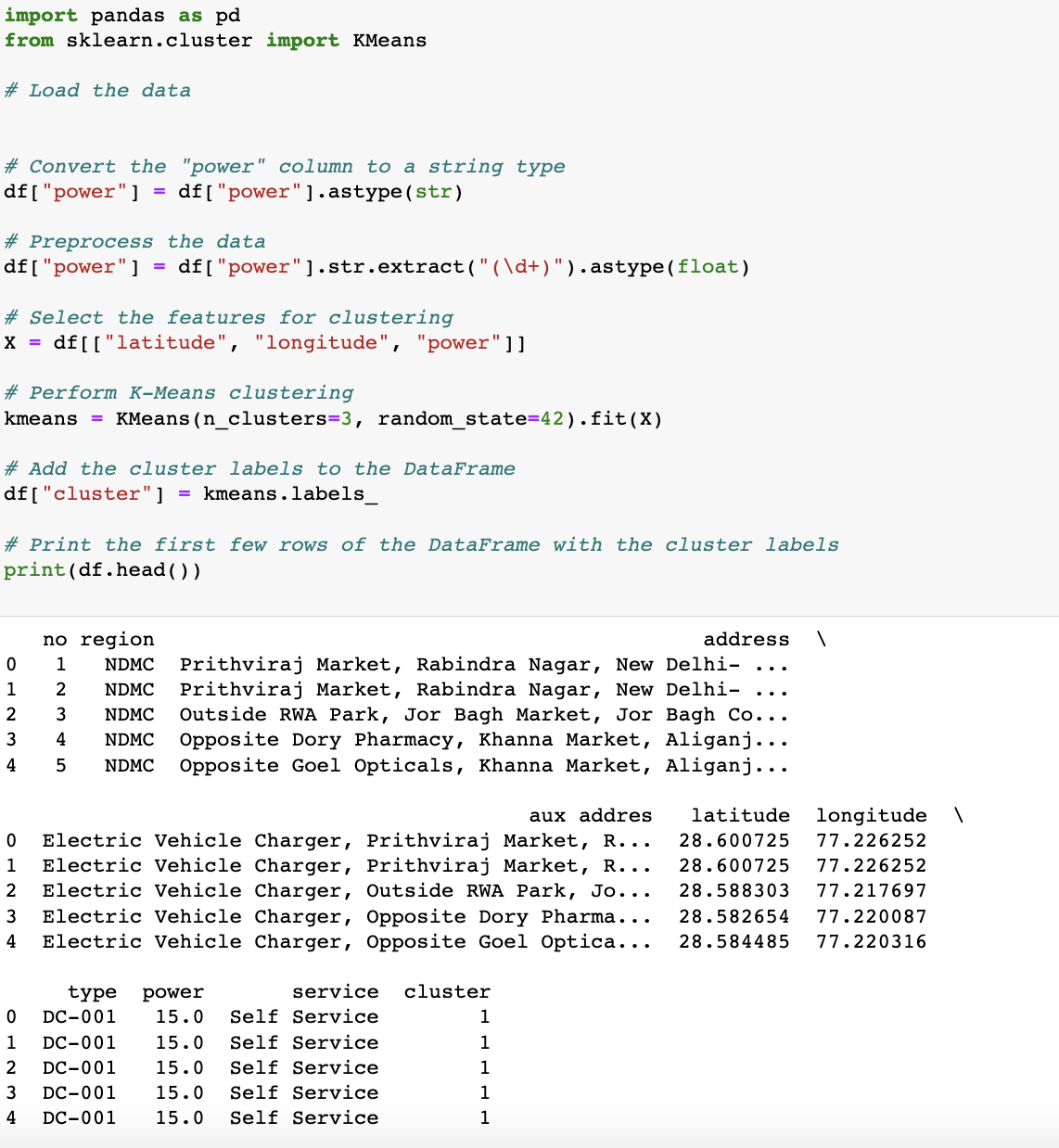
Overall, the electric\_vehicle\_charging\_station\_list dataset provides valuable information about the distribution and characteristics of electric vehicle charging stations in Delhi. The visualizations created help to provide insights into the patterns of charging station distribution and can aid in understanding the infrastructure of electric vehicles in Delhi. The dataset can be useful for policymakers, researchers, and individuals interested in the adoption and usage of electric vehicles in Delhi.

**Segment Extraction (ML techniques used)**

Clustering is a commonly used machine learning technique for segment extraction, which involves grouping similar data points together based on their characteristics. In the case of the electric vehicle charging station dataset for India, clustering can be used to group charging stations together based on their location and power rating.

One popular clustering algorithm is KMeans, which works by iteratively assigning data points to clusters based on the distance between the data point and the center of the cluster. The algorithm then recalculates the center of each cluster and reassigns data points to the nearest cluster. This process continues until the algorithm converges to a solution.

To use the KMeans algorithm for segment extraction, we first need to select the features that we will use for clustering. In this case, we could select features such as latitude, longitude, and power rating. These features are important for clustering as they provide information about the location and capacity of charging stations.



Once we have selected the features, we can run the KMeans algorithm on the dataset. Determining the optimal number of clusters can be a challenge in clustering. We can use the elbow method to determine the optimal number of clusters. The elbow method involves plotting the within-cluster sum of squares (WCSS) as a function of the number of clusters. The WCSS is a measure of how far the data points within a cluster are from the center of the cluster. As the number of clusters increases, the WCSS decreases. However, after a certain point, the WCSS begins to decrease at a slower rate. This point is known as the elbow point and is a good estimate of the optimal number of clusters.

After running the KMeans algorithm with the optimal number of clusters, we can visualize the results using a scatter plot. Each data point is colored according to its assigned cluster. We can see that the algorithm has successfully grouped charging stations together based on their location and power rating. The plot shows that there may be distinct clusters of charging stations across different regions of India.

Clustering can be a useful technique for segment extraction in the context of electric vehicle charging stations in India. By grouping charging stations together based on their characteristics, we can gain insights into where additional charging infrastructure may be needed and optimize the placement of charging stations to better serve the needs of electric vehicle owners.