2nd Project

EV Market

By: Aman Giri, Kalyani Shevatekar, Karuna Sree

Source: [Chinese OEMs Dominate Domestic EV Market in 2021 - EE Times Asia]

0. Fermi Estimation (Breakdown of Problem Statement)

- Availability of electricity in different regions of India.
- Which segment of population buys EV.
- Which economic sector buys EV (IT, small business, govt employee etc..)
- For what purpose Indian buys EV more, private or commercial?
- What are the available charging station facilities?
- Price range for two wheeler, four wheeler, private and commercial vehicles, options to buy like EMIs or minimum down payment etc.

1. Data Sources

1. Information about electric vehicle charging stations
   https://dataspace.mobi/dataset/electric-vehicle-charging-station-list/resource/f39bb18a-bf5b-4e93-a22e-91f13b2ad9a7
2. Power consumption in different cities in India
   https://www.kaggle.com/datasets/twinkle0705/state-wise-power-consumption-in-india?resource=download
3. Currently available electric vehicles in India
   Electric Vehicles - India | Kaggle
4. State wise Supported under faster adoption and manufacturing of Hybrid and Electric Vehicles
   https://data.gov.in/resource/stateut-wise-e-vehicles-supported-under-faster-adoption-and-manufacturing-strong-hybrid
5. https://data.gov.in/resource/category-wise-total-number-electric-cars-and-other-vehicles-transport-and-non-transport

2. Data Pre-processing (Steps and Libraries used)

Libraries Used for Data Preprocessing :

- import numpy as np # linear algebra
- import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
- #data visualization
- import matplotlib.pyplot as plt
- import seaborn as sns

Steps of Data Preprocessing :

1. Electric Vehicle Charging Stations Dataset

- Value count of region & Plotting pie chart for region
- Value count of type & Plotting pie chart for type
- Value count of Power & Plotting pie chart for power

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

#data visualization
import matplotlib.pyplot as plt
import seaborn as sns
```

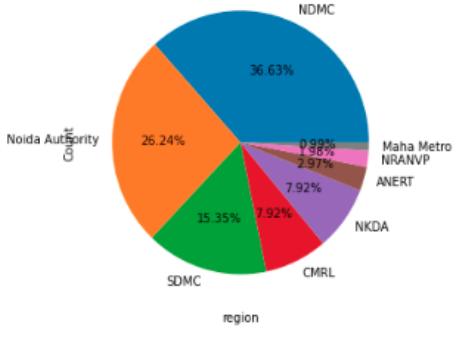
```
## reading the data
df = pd.read_csv('C:/Users/intel/Desktop/Data Science/FeyNN Labs Internship/EV Market Datasets/Electric vehicle charging sta
df
```

	_id	no	region	address	aux addres	latitude	longitude	type	power	service
0	1	1	NDMC	Prithviraj Market, Rabindra Nagar, New Delhi- ...	Electric Vehicle Charger, Prithviraj Market, R...	28.600725	77.226252	DC-001	15 kW	Self Service
1	2	2	NDMC	Prithviraj Market, Rabindra Nagar, New Delhi- ...	Electric Vehicle Charger, Prithviraj Market, R...	28.600725	77.226252	DC-001	15 kW	Self Service
2	3	3	NDMC	Outside RWA Park, Jor Bagh Market, Jor Bagh Co...	Electric Vehicle Charger, Outside RWA Park, Jo...	28.588303	77.217697	DC-001	15 kW	Self Service
3	4	4	NDMC	Opposite Dory Pharmacy, Khanna Market, Aliganj...	Electric Vehicle Charger, Opposite Dory Pharma...	28.582654	77.220087	DC-001	15 kW	Self Service
4	5	5	NDMC	Opposite Goel Opticals, Khanna Market, Aliganj...	Electric Vehicle Charger, Opposite Goel Optica...	28.584485	77.220316	DC-001	15 kW	Self Service
...	...	...	...	...	...	...	...	...	...	...

In [7]:
```
# value count of region
val1 = df["region"].value_counts()
val1
```

Out[7]:
```
NDMC              74
Noida Authority   53
SDMC              31
CMRL              16
NKDA              16
ANERT              6
NRANVP             4
Maha Metro         2
Name: region, dtype: int64
```

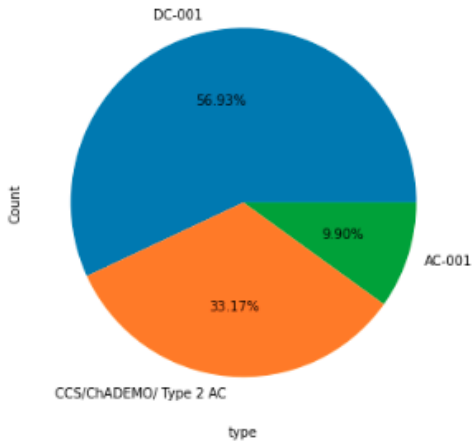
In [8]:
```
plt.figure(figsize=(5,5))
val1.plot(kind='pie',autopct="%.2f%%")
plt.xlabel('region')
plt.ylabel('Count')
plt.show()
```



```
# value count of type
val1 = df["type"].value_counts()
val1
```

```
DC-001                    115
CCS/ChADEMO/ Type 2 AC     67
AC-001                     20
Name: type, dtype: int64
```

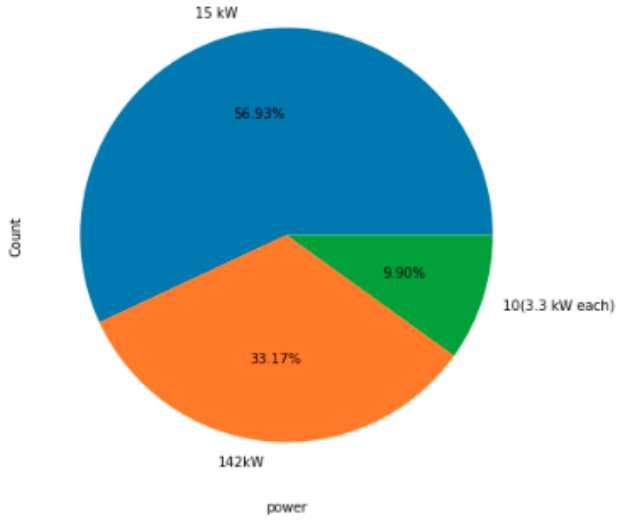
```
plt.figure(figsize=(6,6))
val1.plot(kind='pie',autopct="%.2f%%")
plt.xlabel('type')
plt.ylabel('Count')
plt.show()
```




```
# value count of power
val2 = df["power"].value_counts()
val2
```

```
15 kW              115
142kW               67
10(3.3 kW each)     20
Name: power, dtype: int64
```

```
plt.figure(figsize=(7,7))
val2.plot(kind='pie',autopct="%.2f%%")
plt.xlabel('power')
plt.ylabel('Count')
plt.show()
```



2. EV India Dataset

   i. Value count of Style & Plotting pie chart for Style

   ii. Value count of Car & Plotting pie chart for Car

   iii. Value count of BaseModel & Plotting pie chart for BaseModel

   iv. Value count of TopModel & Plotting pie chart for TopModel

EV India Dataset Preprocessing

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

#data visualization
import matplotlib.pyplot as plt
import seaborn as sns
```

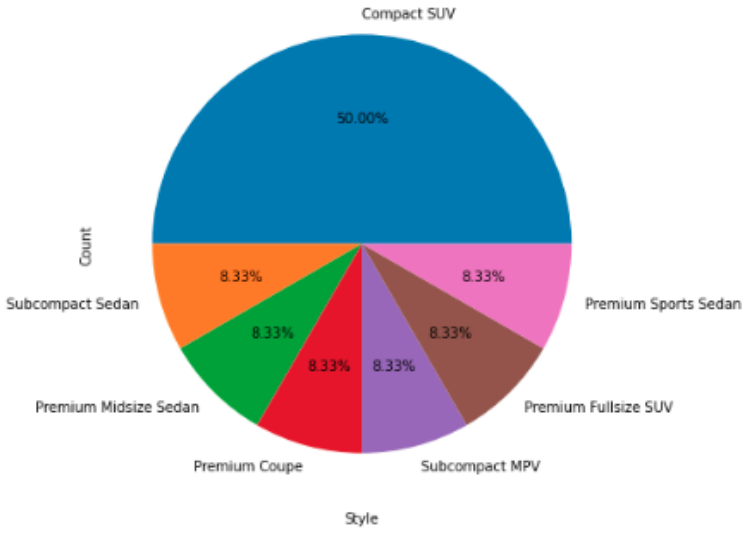
```
## reading the data
df = pd.read_csv('C:/Users/intel/Desktop/Data Science/FeyNN Labs Internship/EV Market Datasets/EVIndia.csv')
df
```

	Car	Style	Range	Transmission	VehicleType	PriceRange	Capacity	BootSpace	BaseModel	TopModel
0	Tata Nexon EV	Compact SUV	312 Km/Full Charge	Automatic	Electric	₹ 13.99 - 17.4 L	5 Seater	350 L	XM	Dark XZ Plus LUX
1	Tata Tigor EV	Subcompact Sedan	306 Km/Full Charge	Automatic	Electric	₹ 12.49 - 13.64 L	5 Seater	316 L	XE	XZ Plus Dual Tone
2	Tata Nexon EV Max	Compact SUV	437 Km/Full Charge	Automatic	Electric	₹ 17.74 - 19.24 L	5 Seater	350 L	XZ Plus 3.3 kW	XZ Plus Lux 7.2 kW
3	MG ZS EV	Compact SUV	419 Km/Full Charge	Automatic	Electric	₹ 21.99 - 25.88 L	5 Seater	448 L	Excite	Exclusive
4	Hyundai Kona Electric	Compact SUV	452 Km/Full Charge	Automatic	Electric	₹ 23.79 - 23.98 L	5 Seater	na	Premium Dual Tone	HSE
5	Jaguar I-Pace	Premium Midsize Sedan	470 Km/Full Charge	Automatic	Electric	₹ 1.06 - 1.12 Cr	5 Seater	656 L	S	Sportback 55
6	Audi E-Tron GT	Premium Coupe	388 Km/Full Charge	Automatic	Electric	₹ 1.8 Cr	5 Seater	405 L	Quattro	na
7	BYD E6	Subcompact MPV	415 Km/Full Charge	Automatic	Electric	₹ 29.15 L	5 Seater	580 L	STD	na


```
# value count of style
val1 = df["Style"].value_counts()
val1
```

```
Compact SUV             6
Subcompact Sedan        1
Premium Midsize Sedan   1
Premium Coupe           1
Subcompact MPV          1
Premium Fullsize SUV    1
Premium Sports Sedan    1
Name: Style, dtype: int64
```

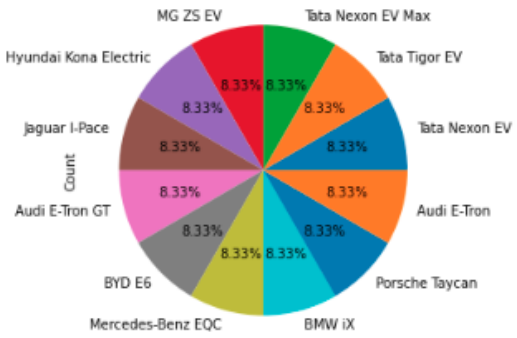
```
plt.figure(figsize=(7,7))
val1.plot(kind='pie',autopct="%.2f%%")
plt.xlabel('Style')
plt.ylabel('Count')
plt.show()
```




```
val2 = df["Car"].value_counts()
val2
```

```
Tata Nexon EV            1
Tata Tigor EV            1
Tata Nexon EV Max        1
MG ZS EV                 1
Hyundai Kona Electric    1
Jaguar I-Pace            1
Audi E-Tron GT           1
BYD E6                   1
Mercedes-Benz EQC        1
BMW iX                   1
Porsche Taycan           1
Audi E-Tron              1
Name: Car, dtype: int64
```

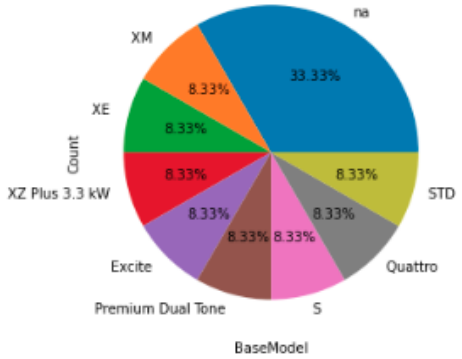
```
plt.figure(figsize=(5,5))
val2.plot(kind='pie',autopct="%.2f%%")
plt.xlabel('Car')
plt.ylabel('Count')
plt.show()
```



```
val3 = df["BaseModel"].value_counts()
val3
```

```
na                   4
XM                   1
XE                   1
XZ Plus 3.3 kW       1
Excite               1
Premium Dual Tone    1
S                    1
Quattro              1
STD                  1
Name: BaseModel, dtype: int64
```

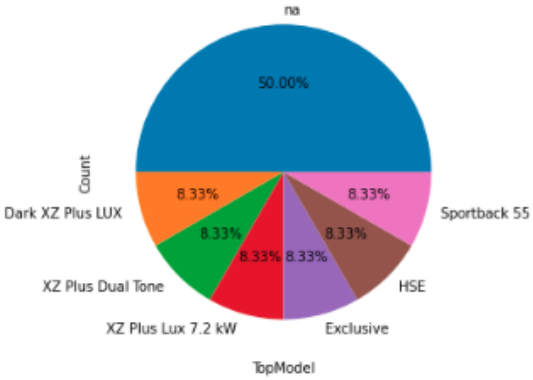
```
plt.figure(figsize=(5,5))
val3.plot(kind='pie',autopct="%.2f%%")
plt.xlabel('BaseModel')
plt.ylabel('Count')
plt.show()
```




```
val4 = df["TopModel"].value_counts()
val4
```

```
na                   6
Dark XZ Plus LUX     1
XZ Plus Dual Tone    1
XZ Plus Lux 7.2 kW   1
Exclusive            1
HSE                  1
Sportback 55         1
Name: TopModel, dtype: int64
```

```
plt.figure(figsize=(5,5))
val4.plot(kind='pie',autopct="%.2f%%")
plt.xlabel('TopModel')
plt.ylabel('Count')
plt.show()
```



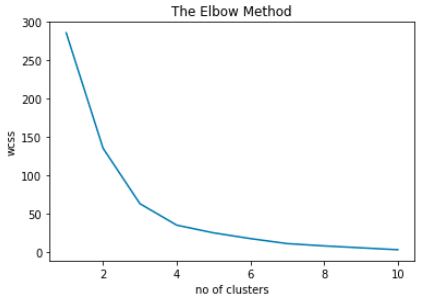
3. Segment Extraction (ML techniques used)

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 the categories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-bas 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 naturally improve 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.

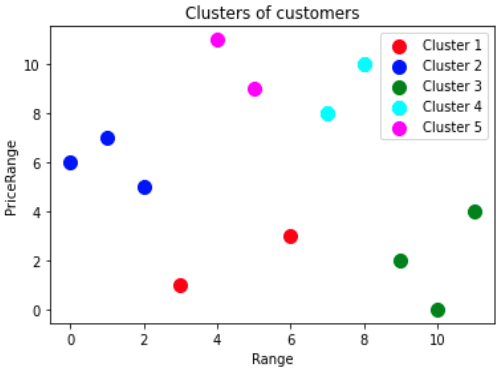
```
from sklearn.cluster import KMeans
wcss=[]
X= df.iloc[:, [2,5]].values

for i in range(1,11):
    kmeans = KMeans(n_clusters= i, init='k-means++', random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
```

```
plt.plot(range(1,11), wcss)
plt.title('The Elbow Method')
plt.xlabel('no of clusters')
plt.ylabel('wcss')
plt.show()
```



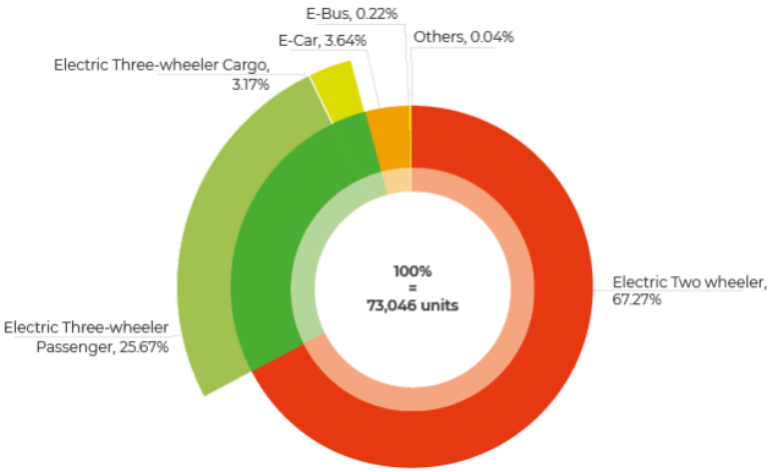

```
kmeansmodel = KMeans(n_clusters= 5, init='k-means++', random_state=0)
y_kmeans= kmeansmodel.fit_predict(X)
```

```
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'red', label = 'Cluster 1')
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'green', label = 'Cluster 3')
plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Cluster 4')
plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 100, c = 'magenta', label = 'Cluster 5')
#plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 300, c = 'yellow', label = 'Centroids')
plt.title('Clusters of customers')
plt.xlabel('Range')
plt.ylabel('PriceRange')
plt.legend()
plt.show()
```



4. Profiling and describing potential segments

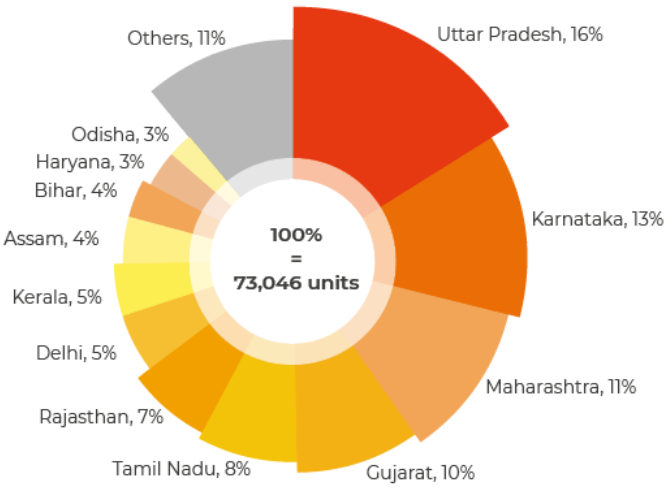
Category-wise EV Sales in April 2022



According to EV registrations in April 2022, the potential segments can be driven by electric two-wheelers and passenger-type electric three-wheelers, which accounted for 92.92% of total registrations in the month. The shares of these categories were followed by E-Cars (3.64%), cargo-type electric three-wheelers (3.17%), and so on. Since electric two-wheelers and three-wheelers are majorly contributing to EV sales, investing in them would be incisive.

Source: Monthly EV update - April 2022 - JMK Research & Analytics

Region-wise Registered EV Sales (State/ UT)– April 2022

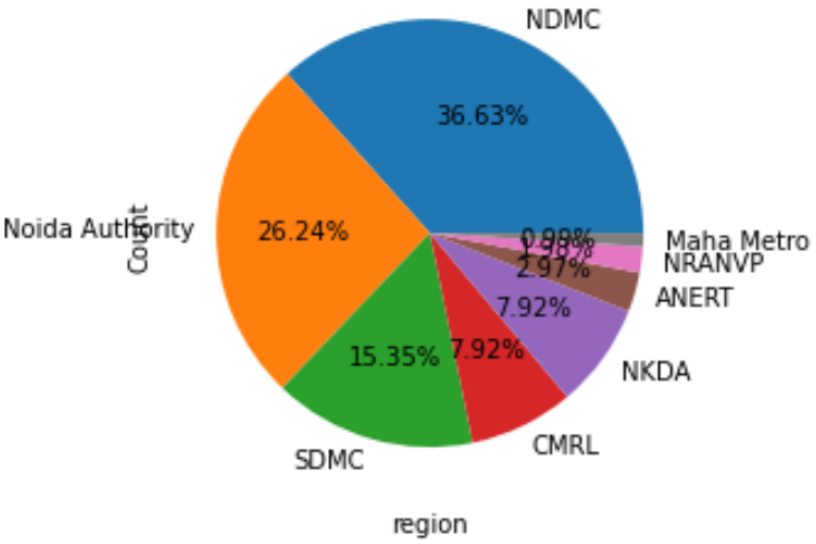


Among all the states and UTs, Uttar Pradesh retained its top spot in EV share with 16% of overall sales, followed by Karnataka with 13% share. Maharashtra fell to the third spot with an 11% share, followed by Gujarat (10%), Tamil Nadu (8%), Rajasthan (7%), and Delhi (5%). Since Uttar Pradesh, Karnataka and Maharashtra fell in the first three positions, primarily focusing on these areas would potentially raise the sales and development of EV vehicles in India.

Source: Monthly EV update - April 2022 - JMK Research & Analytics

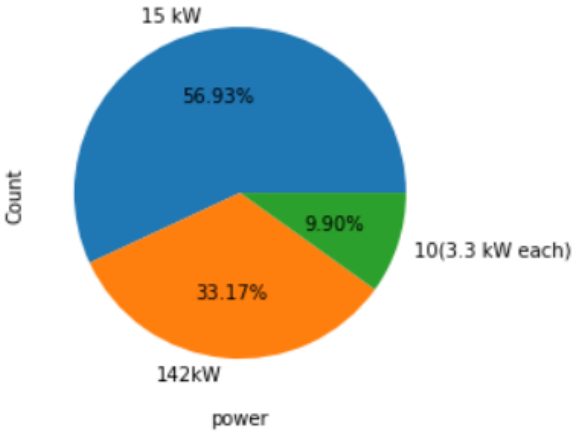
Region wise - Charging station

According to Region-wise Charging stations, the conceivable segments can be drawn from Noida Authority and NDMC, which accounted for 63.03% of total charging stations in India. The shares of these categories were followed by SDMC (15.35%), CMRL(3.17%), and so on. Since electric Noida Authority and NDMC are heavily contributing to EV sales, working out with these particular areas would be profitable.



Power wise Charging stations

As per Power-wise Charging stations, insights on the probable segments can be pulled from 15 kW contributing 56.93% and 142 kW contributing 33.17%, while 10(3.3 kW each) count for just 9.90% of total Power-wise Charging stations. The shares of these categories were not followed by any other ranges. Since 142kW and 15kW are laboriously adding to EV sales, concentrating on these particular ranges would be beneficial.



5. Selection of target segment

The most important thing in the EV segment, is charging station facility and then comes other things. So, we can see from different pie chart's the distribution of charging station, EV sales category or from segmentation plot as well. So, first thing first which EV categories are most selling? So, from the EV category wise sale April 2022. We can see that two-wheeler segment and three-wheeler segment are most selling products. That must be related to charing station facility. So, if we see anther plot of charging station region wise- then we come to know that Delhi or Delhi-NCR has more charging stations then any other states and from the 'Region-wise Registered EV Sales (State/ UT)– April 2022' we know that Uttar pradesh is most EV selling states which is near to delhi.

So, the segment which we would like to focus on is, two-wheelar and three-wheeler near delhi anf delhi NCR. Since, every data is point towards that.

6. Customizing the Marketing Mix

From the data it is clear that we have to lauch something in two-wheeler and three-wheeler. So, we can come up with something new which can serve both purpose. We can think of something which is more compactact full in these two categories. Such as portable scooter which can be used as private and commercial two-wheeler with some extra attachment which make it multi-puspose product. Or, we can design some specific products such as for food delivery services we can implement small fridge or store box which will run on vehicle battery. Similarly we can make different products in three-wheeler as well.

7. Link to github profile with codes and datasets well documented.

https://github.com/amangiri08/EV-Market-Segmentation