EV Market Segmentation

Mahesh Tiria

Abstract

Electric vehicles (EVs) represent a significant shift in transportation technology, aiming to reduce reliance on fossil fuels and minimize carbon emissions. They are powered by electricity stored in batteries, which propel an electric motor, eliminating the need for internal combustion engines.

India's electric vehicle (EV) market has witnessed significant growth and interest in recent years, driven by both government initiatives and market dynamics. The Indian government has been actively promoting EV adoption through various schemes, incentives, and policy measures aimed at reducing pollution and dependence on fossil fuels.

Data Source

We are using EV Sales data(State and Union Territory wise) provided by Indian Government .

https://data.gov.in/resource/stateuts-wise-current-sales-electric-vehicles-ev-country-various-segments-reply-unstarred.

Data Overview

Data contains the sales amount of electric vehicle type sold for each possible vehicle type. Each row represent the data for a state or union territory.

This data will allow us to segment the market based on the current market status of EV market across the states

State Name	Two Wheeler	Three Wheeler	Four Wheeler	Goods Vehicles	Public Service Vehicle	Special Category Vehicles	Ambulance/Hearses	Construction Equipment Vehicle	Other	Grand Total
Andaman and Nicobar Island	1	30.0	81	NaN	40.0	NaN	NaN	NaN	7.0	159
Arunachal Pradesh	14	NaN	5	NaN	NaN	NaN	NaN	NaN	1.0	20
Assam	721	47041.0	161	7.0	15.0	NaN	NaN	NaN	2.0	47947
Bihar	5003	59079.0	114	11.0	26.0	NaN	NaN	NaN	8.0	64241
Chandigarh	298	1410.0	182	NaN	40.0	NaN	NaN	NaN	1.0	1931

Data Pre-processing

Data contains NaN value in some places , to handle this it has been replaced with 0.

```
[5]: #replacing nan with 0
     data.fillna(0,inplace=True)
     data.head()
```

Segmentation Extraction

To extract segments various state market can be clustered together to form strategy according to the target cluster. Some market segments already have existing EV market, some might be in growing phase and some might have non existent EV market.

There exists a hierarchy of market segments based on the amount of sales in various states. So we are going to use **Agglomerative Clustering** to segment the markets.

Benefits of agglomerative clustering:

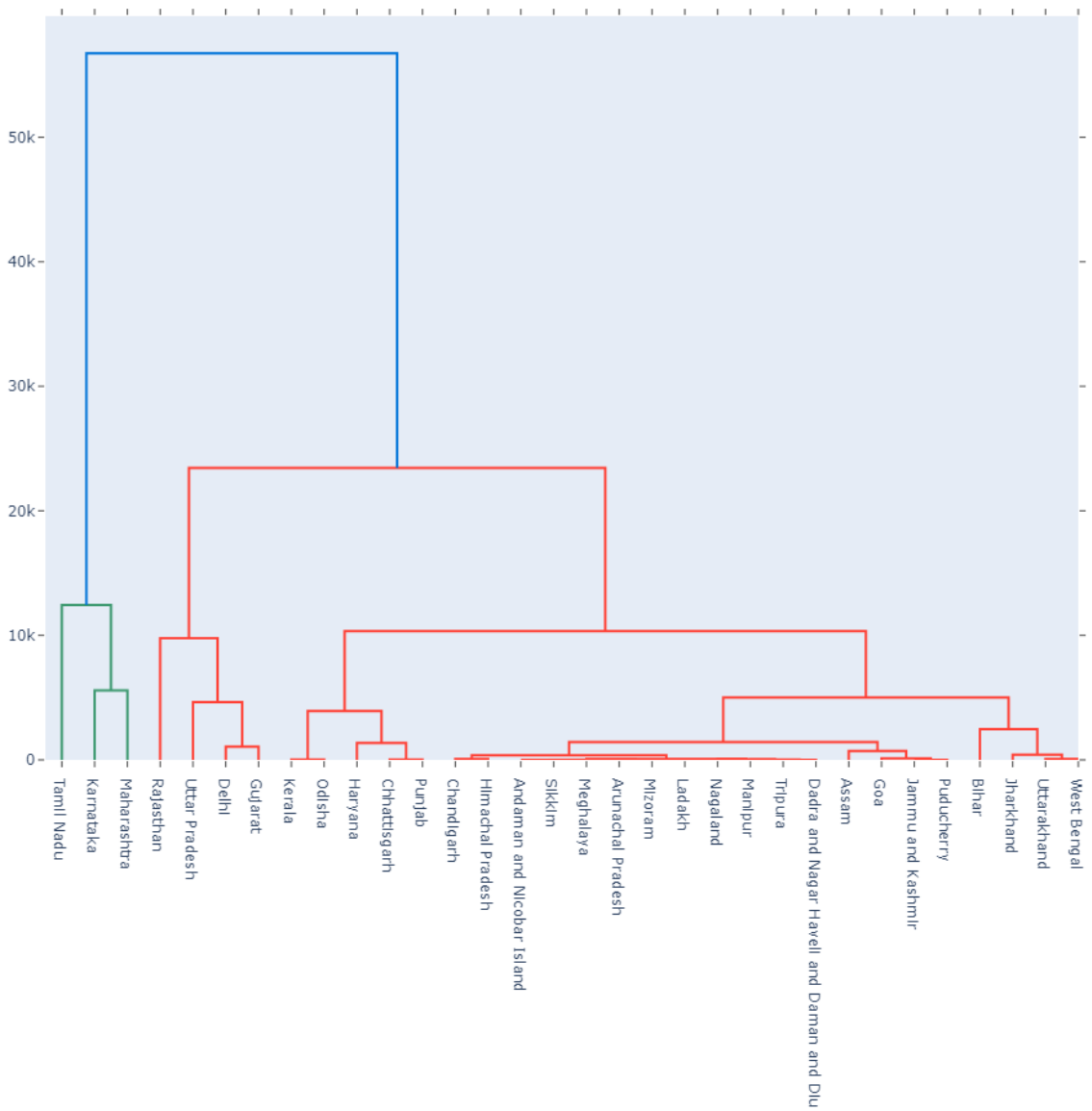
- Hierarchical Representation: One significant advantage is its hierarchical structure, which allows for the creation of dendrograms. These visualizations display the relationships between clusters at different levels of granularity, aiding in understanding the data's natural grouping.
- No Preset Cluster Number: Unlike some other clustering algorithms, agglomerative clustering doesn’t require specifying the number of clusters beforehand. It iteratively combines data points based on similarity, making it adaptable to various datasets without a priori knowledge of the ideal number of clusters.

- Flexibility in Linkage Criteria: Agglomerative clustering offers different linkage methods (e.g., single, complete, average linkage) to measure the similarity or dissimilarity between clusters. This flexibility allows for choosing the appropriate linkage criterion that best fits the dataset's structure.
- Scalability: While the computational complexity can increase with larger datasets due to its hierarchical nature, agglomerative clustering can still be scalable, particularly with efficient implementations and optimizations.
- Interpretability and Understanding: The hierarchical nature of the clusters provides a clear representation of how data points are grouped together, aiding in interpretation and understanding of the relationships within the dataset.
- Merging Strategy Insights: It allows insights into the merging strategy by tracking how clusters are formed step by step, which can be useful in understanding the data's structure and relationships between different clusters.
- Handling Different Data Types: Agglomerative clustering can handle various types of data (numerical, categorical, or mixed) and different distance metrics to calculate similarity or dissimilarity, making it versatile across different types of datasets.

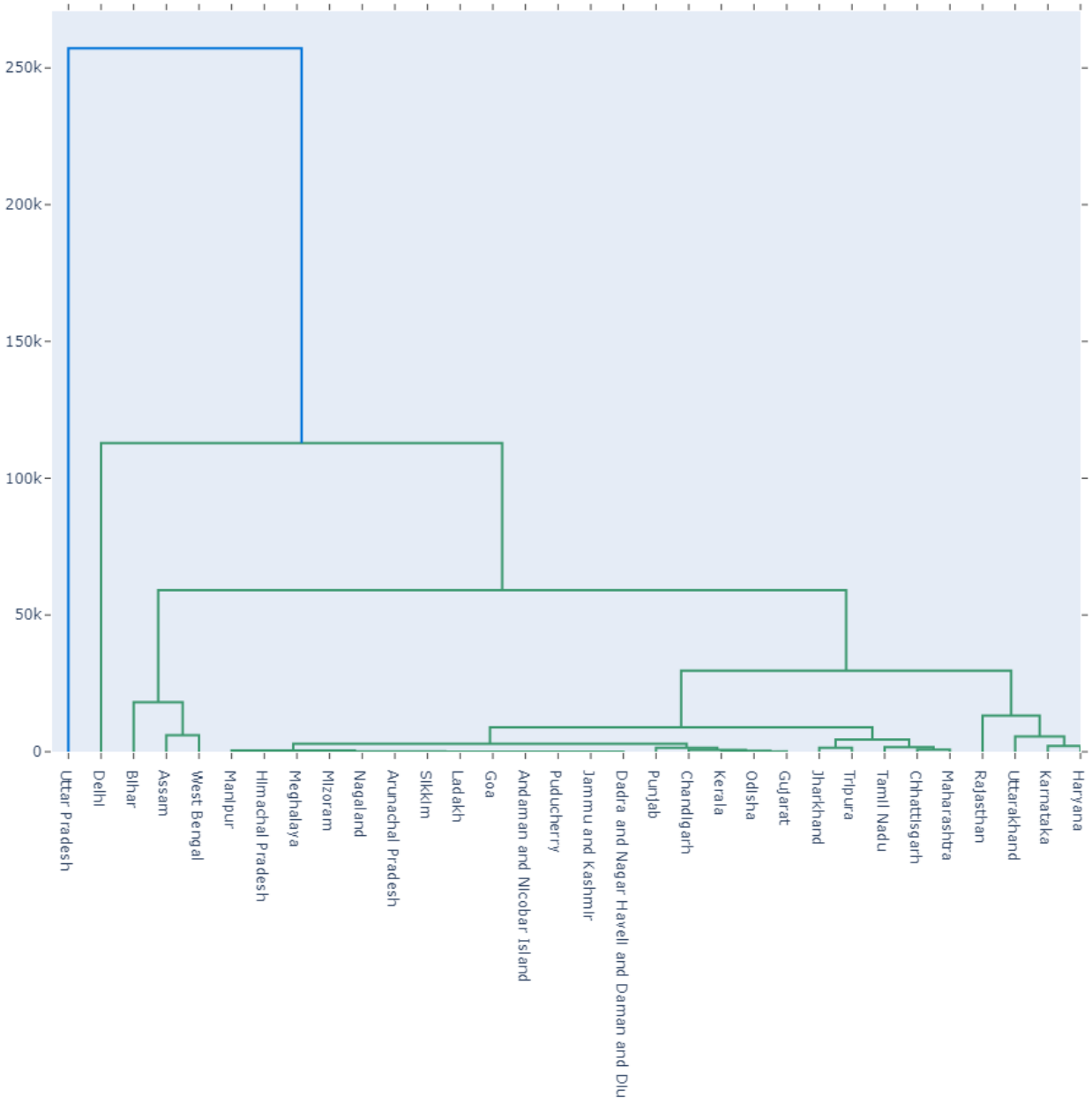
Analysis

Applying Agglomerative Clustering on various vehicle types

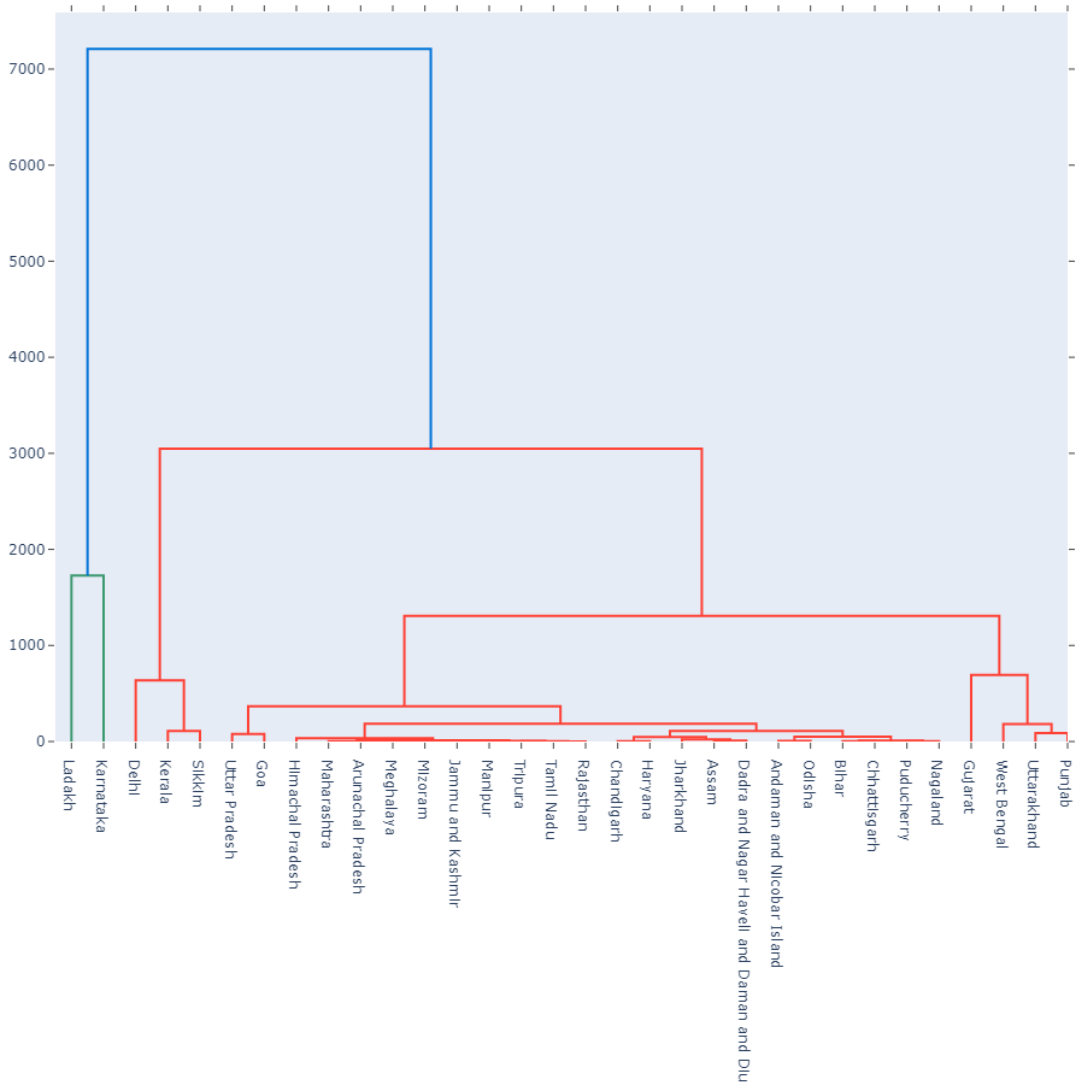
Two Wheeler

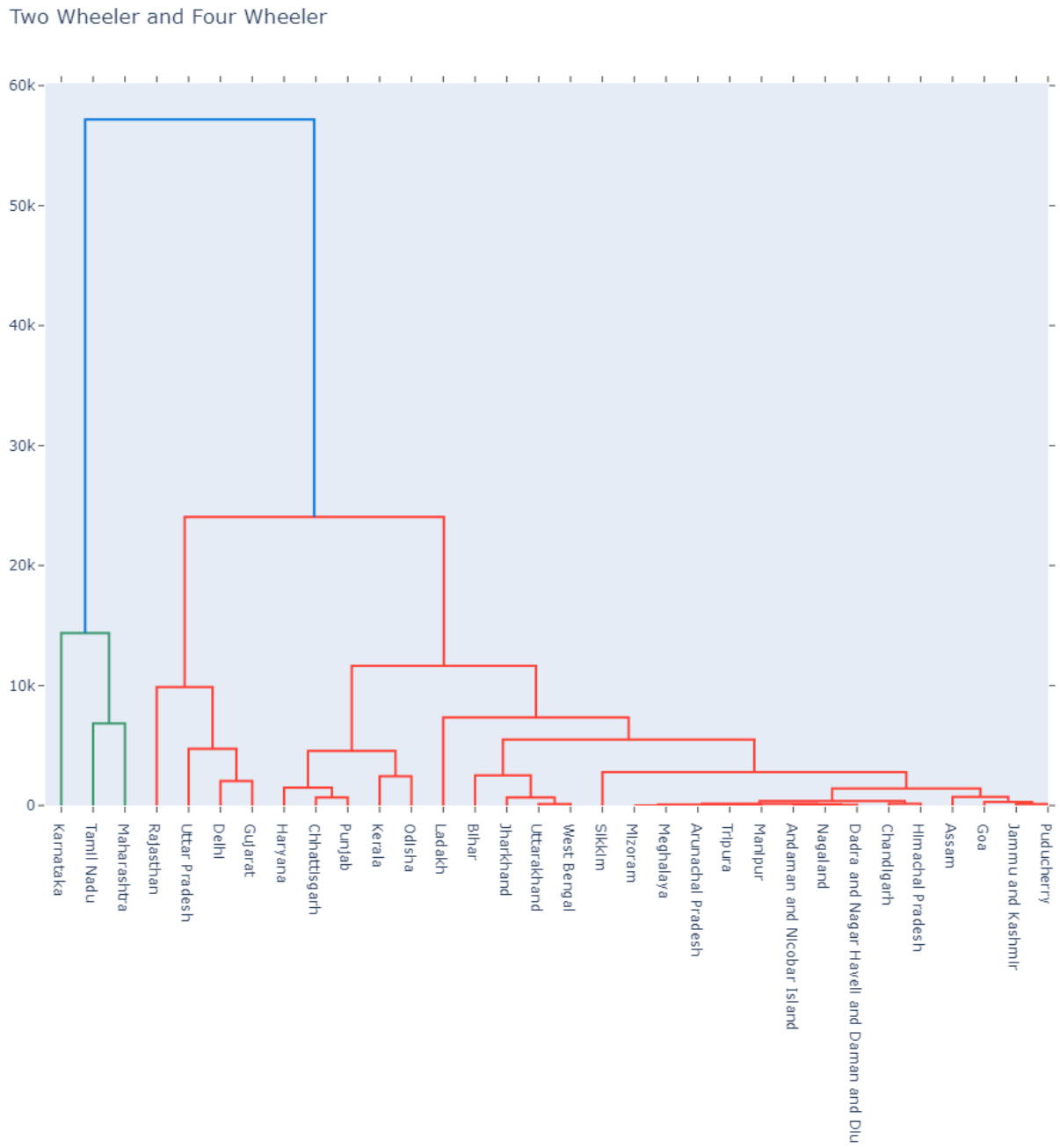

50k
40k
30k
20k
10k
0
Tamil Nadu
Karnataka
Maharashtra
Rajasthan
Uttar Pradesh
Delhi
Gujarat
Kerala
Odisha
Haryana
Chhattisgarh
Punjab
Chandigarh
Himachal Pradesh
Andaman and Nicobar Island
Sikkim
Meghalaya
Arunachal Pradesh
Mizoram
Ladakh
Nagaland
Manipur
Tripura
Dadra and Nagar Haveli and Daman and Diu
Assam
Goa
Jammu and Kashmir
Puducherry
Bihar
Jharkhand
Uttarakhand
West Bengal

Three Wheeler



Four Wheeler





Conclusion

In India EV market is at different stages in different states, so we can segment the states according to the sales data. By segmenting similar states together , we can deploy more strategic market plans suiting the target EV market.

Here three market segments can be observed

- One ,where there is high sales number
- Second, where there is absolute beginning of ev market
- Third, there is medium levels of growth

Indian Electric vehicle report based on the study of current vehicle

- Jaswant Singh Caldwe

Link for Dataset:

https://drive.google.com/file/d/16c3Z6GPBpKHDQ4Xl0sy-Yih-KjtiXkhv/view?usp=drivesdk

Features used in the dataset

- Brand
- Model
- AccelSec
- TopSpeed_KmH
- Range_Km
- Efficiency_WhKm
- FastCharge_KmH
- RapidCharge
- PowerTrain
- PlugType
- BodyStyle
- Segment
- Seats
- PriceEuro
- inr(10e3)

Exploratory Data Analysis:

Unique values:

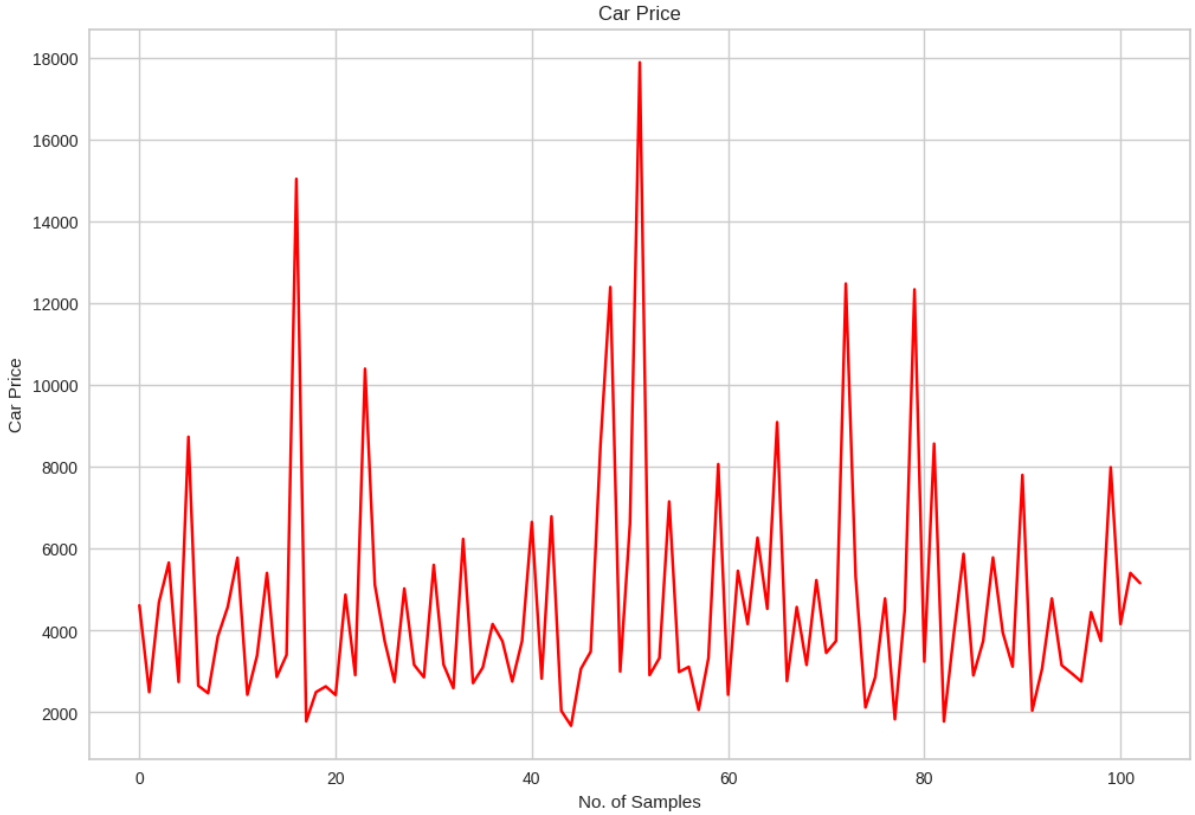
PowerTrain unique values: ['AWD' 'RWD' 'FWD']

PlugType unique values: ['Type 2 CCS' 'Type 2 CHAdeMO' 'Type 2' 'Type 1 CHAdeMO']

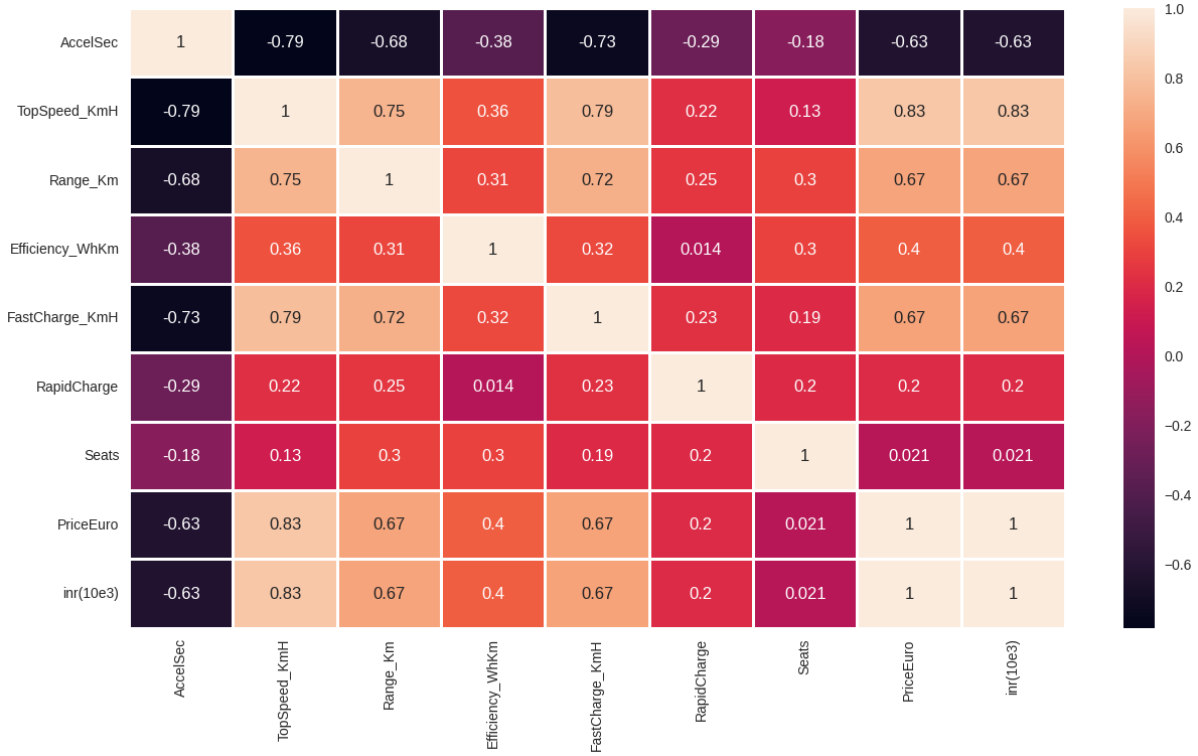
BodyStyle unique values: ['Sedan' 'Hatchback' 'Liftback' 'SUV' 'Pickup' 'MPV' 'Cabrio' 'SPV' 'Station']

Segment unique values: ['D' 'C' 'B' 'F' 'A' 'E' 'N' 'S']

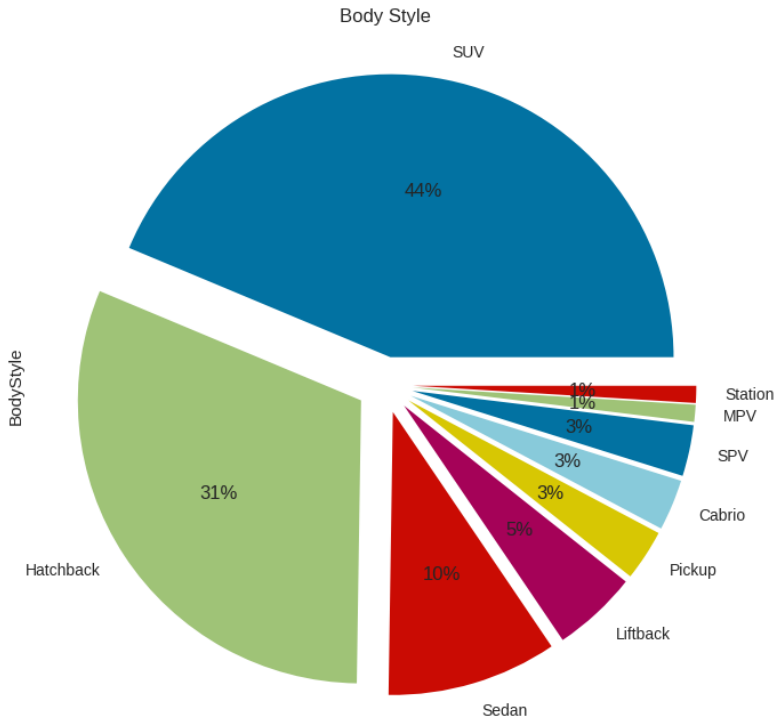
Price range distribution:

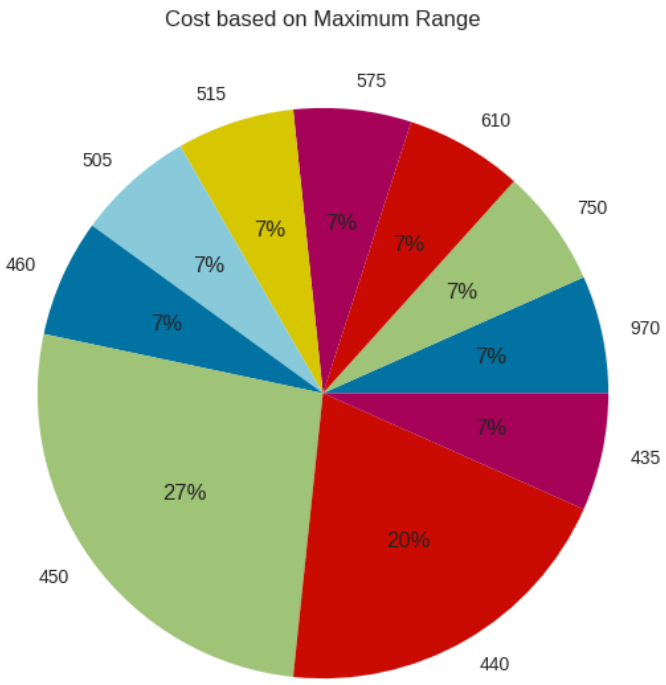


Correlation coefficient:



Segment Body type:





Ols regression values:

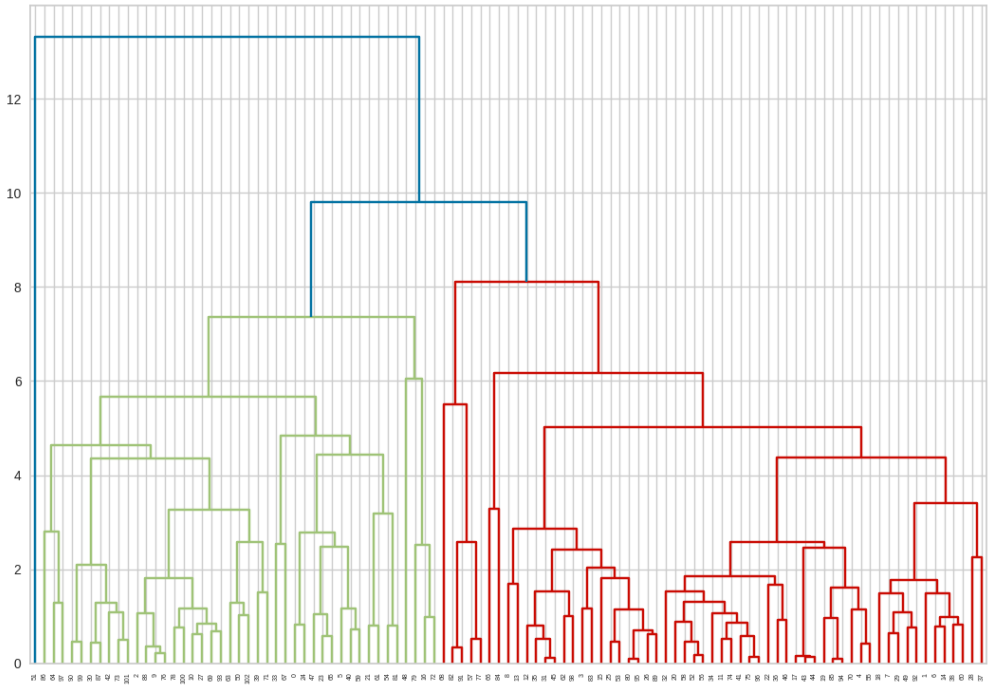
OLS Regression Results

Dep. Variable:	PriceEuro	**R-squared:**	0.721
Model:	OLS	**Adj. R-squared:**	0.704
Method:	Least Squares	**F-statistic:**	41.36
Date:	Sun, 26 Nov 2023	**Prob (F-statistic):**	1.57e-24
Time:	16:04:32	**Log-Likelihood:**	-1155.0
No. Observations:	103	**AIC:**	2324.
Df Residuals:	96	**BIC:**	2342.
Df Model:	6		
Covariance Type:	nonrobust		

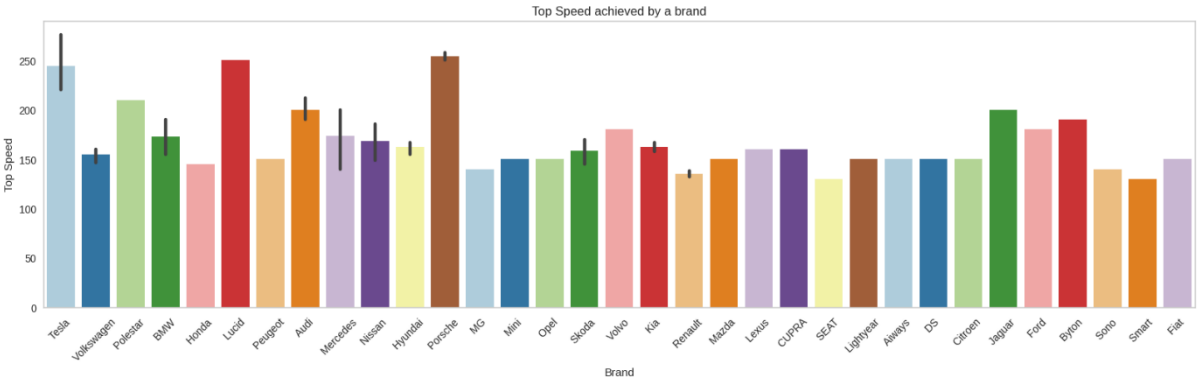
	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.031e+05	2.38e+04	-4.324	0.000	-1.5e+05	-5.58e+04
TopSpeed_KmH	581.7484	80.158	7.257	0.000	422.636	740.861
Efficiency_WhKm	117.6685	70.307	1.674	0.097	-21.890	257.227
PowerTrain	5235.8309	2956.235	1.771	0.080	-632.248	1.11e+04
Range_Km	36.3000	22.629	1.604	0.112	-8.618	81.218
AccelSec	1753.0004	1048.759	1.672	0.098	-328.769	3834.770
RapidCharge	1465.5687	4496.958	0.326	0.745	-7460.822	1.04e+04

Omnibus:	84.867	**Durbin-Watson:**	2.060
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	741.645
Skew:	2.644	**Prob(JB):**	8.99e-162
Kurtosis:	15.036	**Cond. No.**	5.79e+03

Dendogram:



Top speed achieved by brands:



Conclusion:

The electric vehicle market in India holds immense potential to revolutionize the way we commute, offering a greener and moresustainable future. As the infrastructure for EVs continues to develop, along with advancements in battery technology and cost-effectiveness, we can anticipate a significant surge in the popularity and accessibility of electric vehicles across the country. The journey towards a cleaner and more environmentally conscious transportation system is well underway, andthe electric vehicle sector in India is poised to play a pivotal role in shaping this future.

A REPORT ON EV POPULATION DATASET ANALYSIS

- Samarth R Joshi

Link to the dataset - https://www.kaggle.com/datasets/willianoliveiragibin/electric-vehicle-population/data

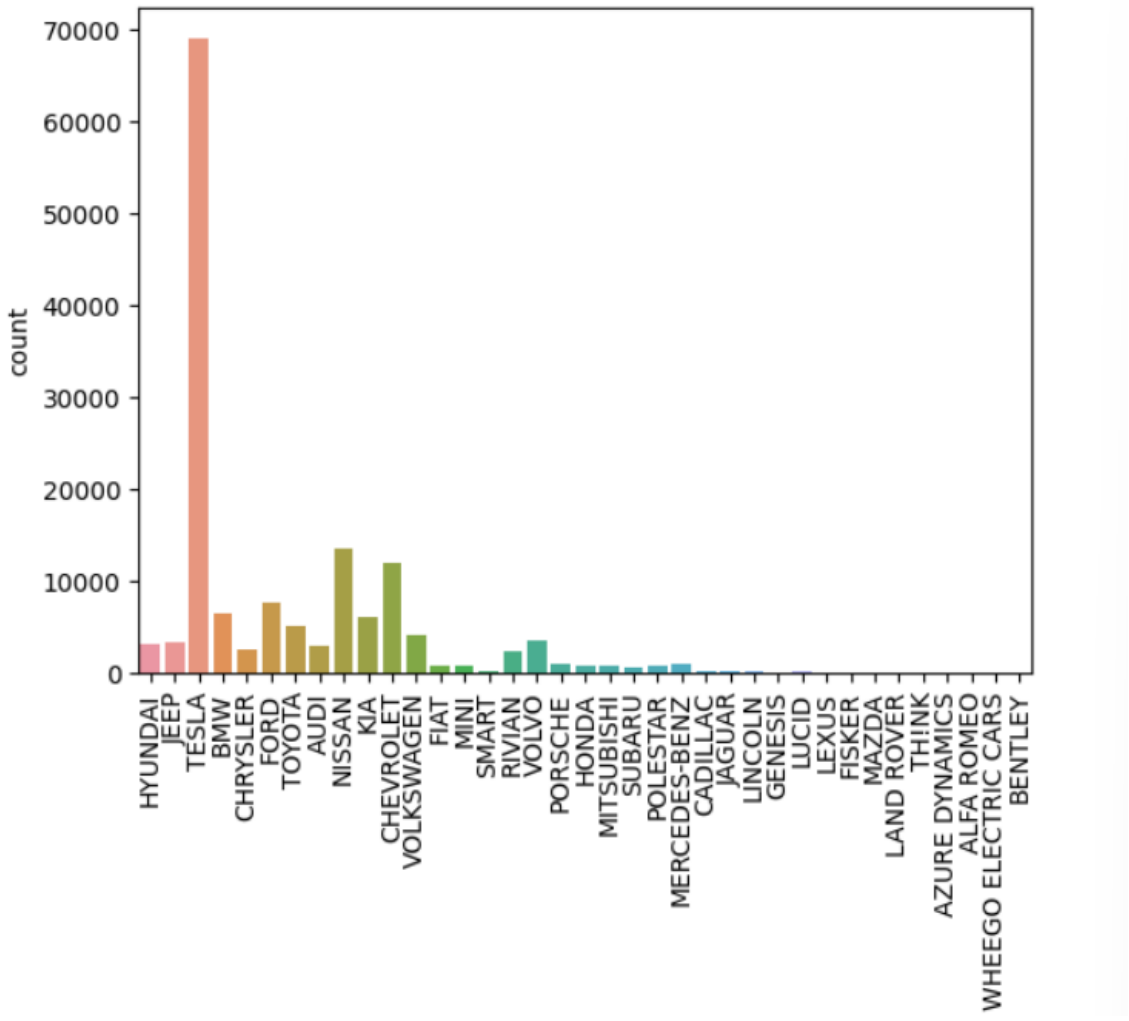
Features in the data set:

- VIN (1-10)
- County
- City
- State
- Postal Code
- Model Year
- Make
- Model
- Electric Vehicle type
- Clean alternative fuel vehicle [CAFV] Eligiblity
- Electric Range
- Base Msrp
- Legislative District
- DOL Vehicle ID
- Vehicle Location
- Electric Utility'
- 2020 Census Tract

Exploratory data analysis has been performed on the following data and following conclusions has been drawn:

1) Count Plot

   The count plot provides a quick visual summary of the distribution of vehicles among different makes, offering insights into the dataset's composition. The plot allows you to identify popular and less common makes, helping you understand the overall landscape of vehicle makes in the dataset.



- X-axis (Vehicle Makes):

The x-axis of the plot represents the various vehicle makes present in the dataset. Each unique make is displayed along the x-axis.

- Y-axis (Count of Vehicles):

The y-axis shows the count or frequency of each vehicle make. Each bar's height corresponds to the number of vehicles belonging to a specific make.

- Bar Heights:

The height of each bar indicates how many vehicles of a particular make are present in the dataset. Taller bars represent makes with a higher count, while shorter bars represent makes with a lower count.

- Rotation of X-axis Labels:

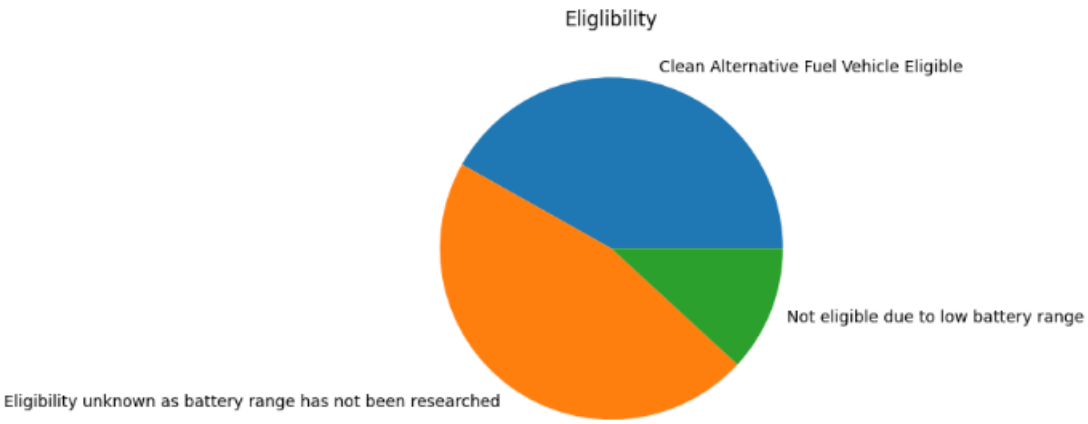
The x-axis labels, which are the names of the vehicle makes, are rotated by 90 degrees. This rotation is applied for better visibility, especially when there are many unique makes, preventing overlap and improving readability.

- Figure Size:

The figure size is set to 15 inches in width and 10 inches in height. This adjustment is made to ensure that the plot is presented in an appropriately sized and clear format within the report.

2) Pie Chart

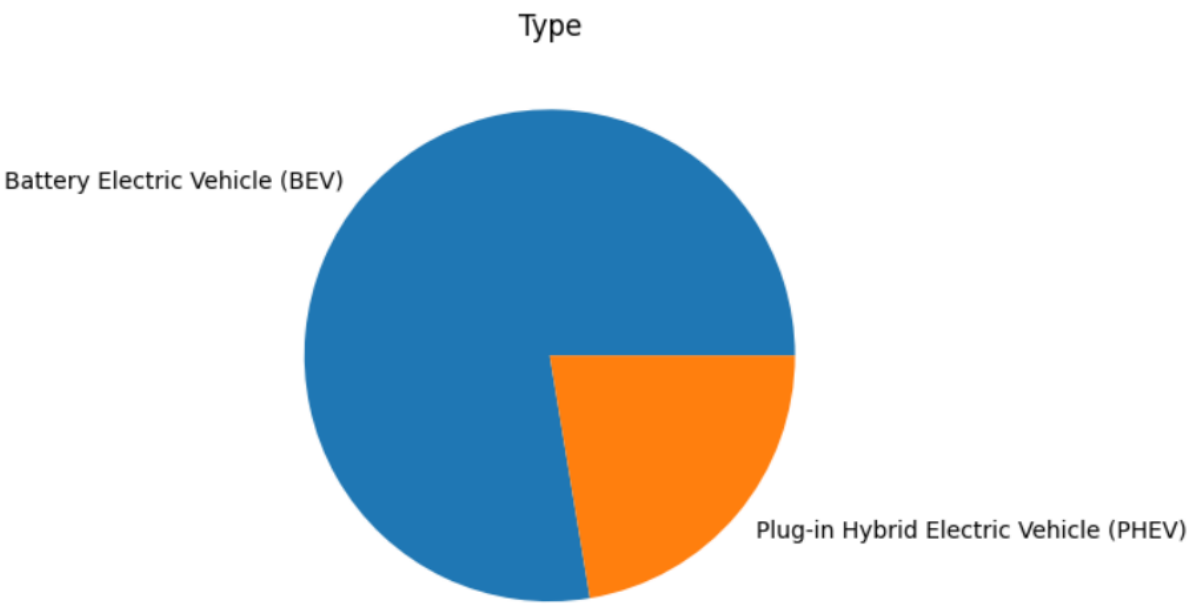
```
Text(0.5, 1.0, 'Eliglibility')
```



The pie chart above illustrates the distribution of vehicles based on their 'Clean Alternative Fuel Vehicle (CAFV) Eligibility.' Each slice represents a category of eligibility, and the size of each slice corresponds to the proportion of vehicles with that eligibility status. The chart provides a visual overview of the distribution of eligibility categories within the dataset.

3)

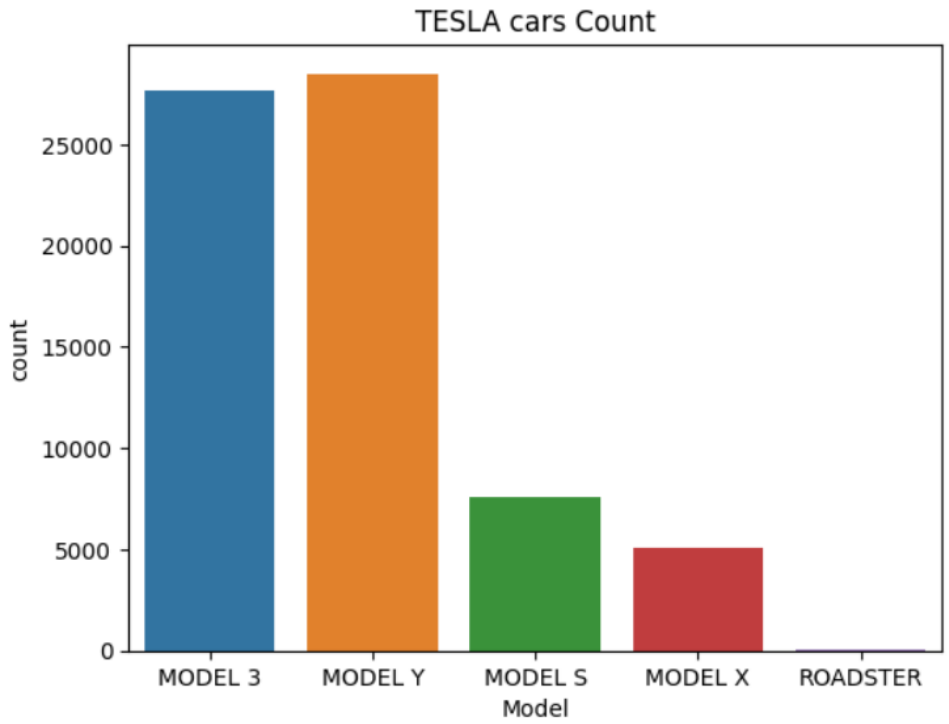
```
Text(0.5, 1.0, 'Type')
```



The pie chart above illustrates the distribution of electric vehicle types within the dataset. Each slice represents a specific electric vehicle type, and the size of each slice corresponds to the proportion of vehicles belonging to that type. The chart provides a visual overview of the composition of electric vehicle types present in the dataset.

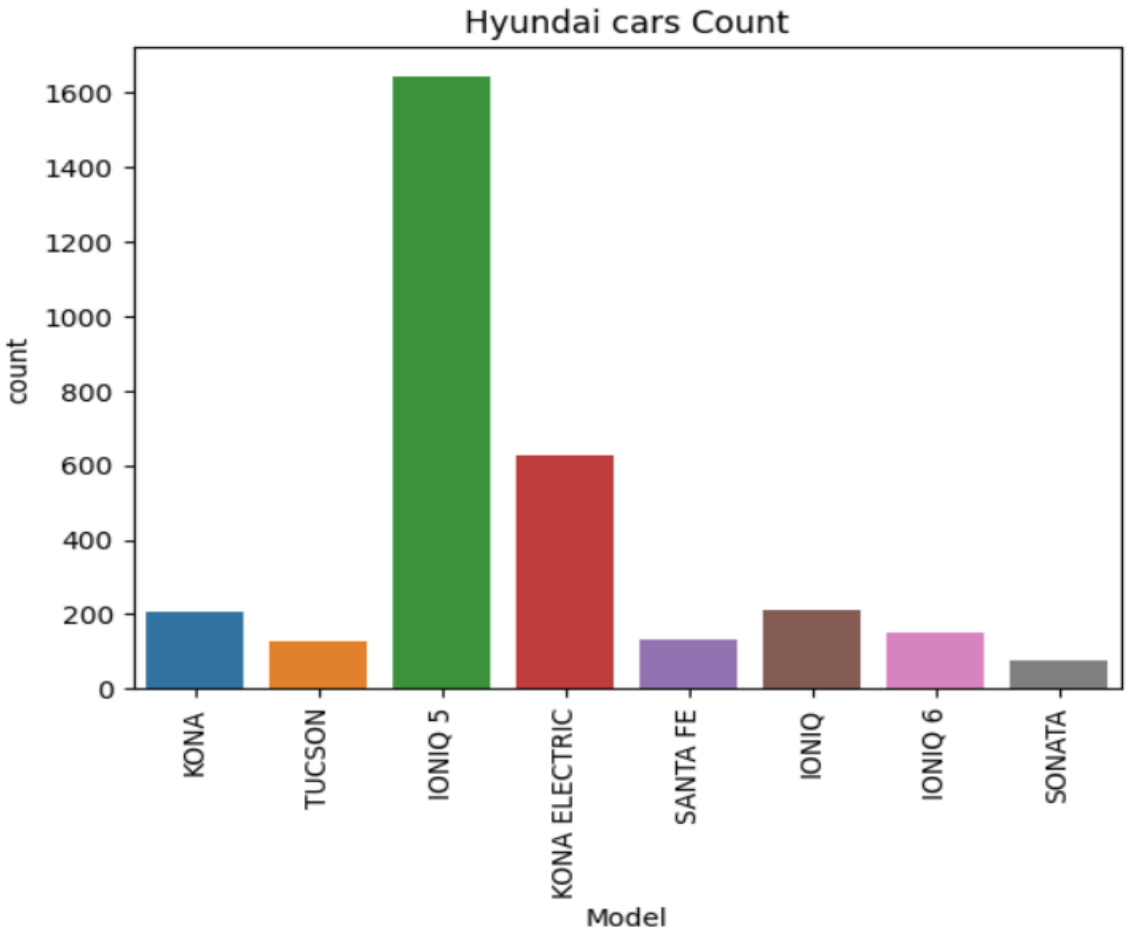
3) BAR PLOTS:

```
Text(0.5, 1.0, 'TESLA cars Count')
```



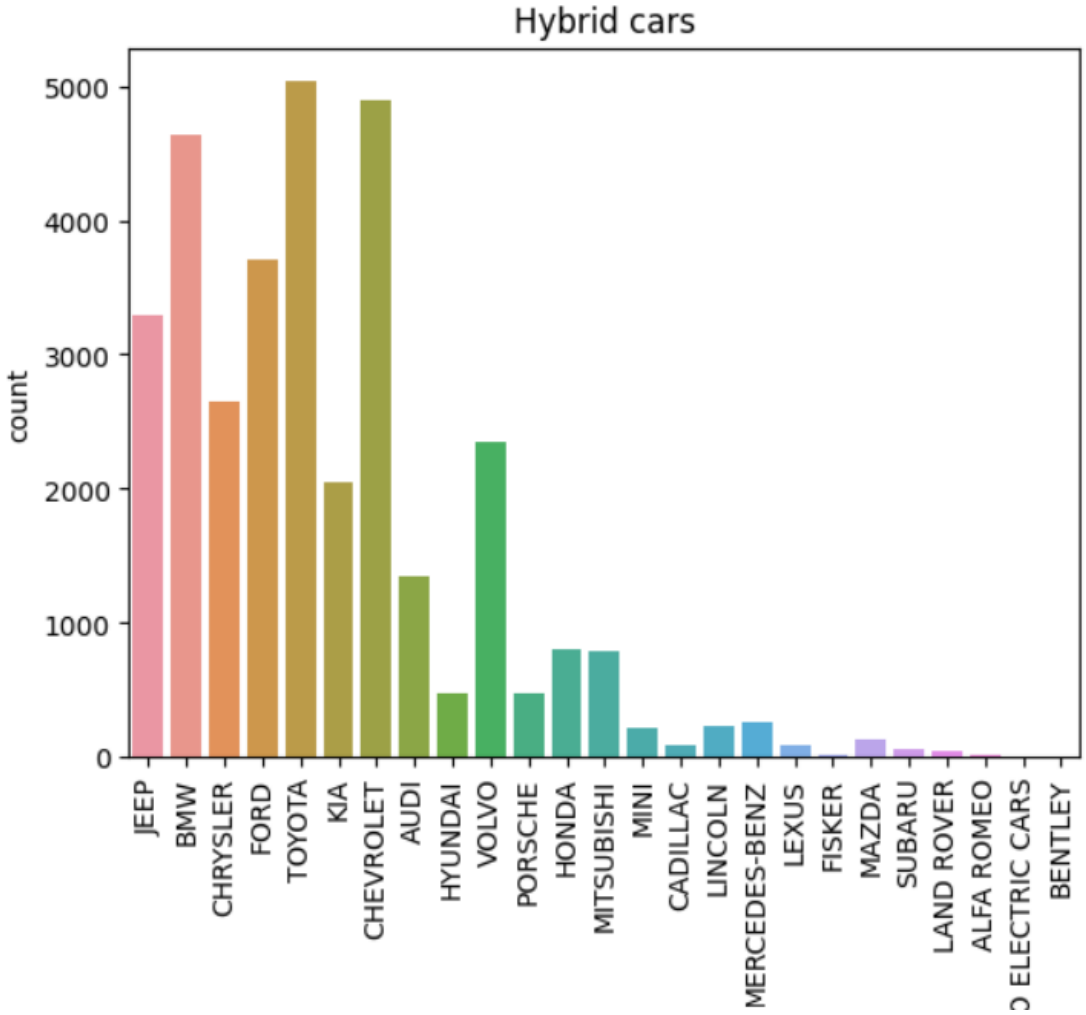
The count plot above visualizes the distribution of Tesla car models within the dataset. Each bar represents the count of a specific Tesla model, providing insights into the popularity of different models among Tesla vehicles in the dataset.

4)



The count plot above visualizes the distribution of Hyundai car models within the dataset. Each bar represents the count of a specific Hyundai model, providing insights into the popularity of different models among Hyundai vehicles in the dataset. The x-axis labels are rotated for improved readability.

5)



The count plot above visualizes the distribution of Plug-in Hybrid Electric Vehicles (PHEV) by their make. Each bar represents the count of PHEV cars for a specific make, providing insights into the popularity of different makes among PHEV vehicles in the dataset. The x-axis labels are rotated for improved readability.

CONCLUSION:

Vehicle Distribution by Make:

The dataset exhibits a diverse range of vehicle makes, with some makes being more prevalent than others. Notable makes include Tesla and Nissan while Bentley and Jaguar are less represented.

Electric Vehicle Types:

The majority of vehicles in the dataset fall into the category of Battery Electric Vehicle (BEV). Other types such as Plug-in Hybrid Electric Vehicles (PHEV) and Hybrid Electric Vehicles (HEV) show varying degrees of representation.

Clean Alternative Fuel Vehicle (CAFV) Eligibility:

A significant portion of the vehicles in the dataset are eligible for Clean Alternative Fuel Vehicle (CAFV) incentives. This suggests a commitment to environmentally friendly vehicles within the dataset.

Customer Segmentation

Sameer

Dataset link: https://drive.google.com/file/d/1aFpqyF-2nr9BbFqDNPmB7Dq0-KCrzu3V/view?usp=sharing

[5]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
5	6	Female	22	17	76
6	7	Female	35	18	6
7	8	Female	23	18	94
8	9	Male	64	19	3
9	10	Female	30	19	72

This dataset gives us an idea about the customer gender, annual income of the customer and annual spending of the customer. Based on this data we can determine our target market and what pricing would be most ideal for launching our Electric Vehicle also what should be the pricing of the acceessories.

Data Preprocessing:

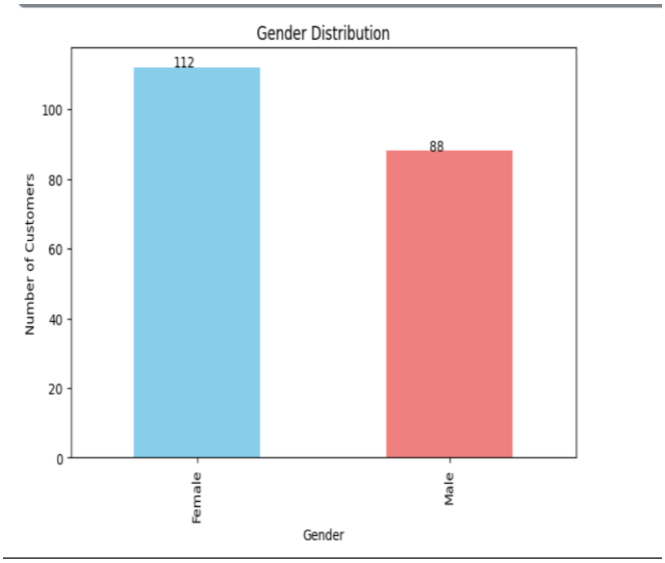
```
null_values = df.isnull().sum()


print("Null Values in Each Column:")
print(null_values)
```

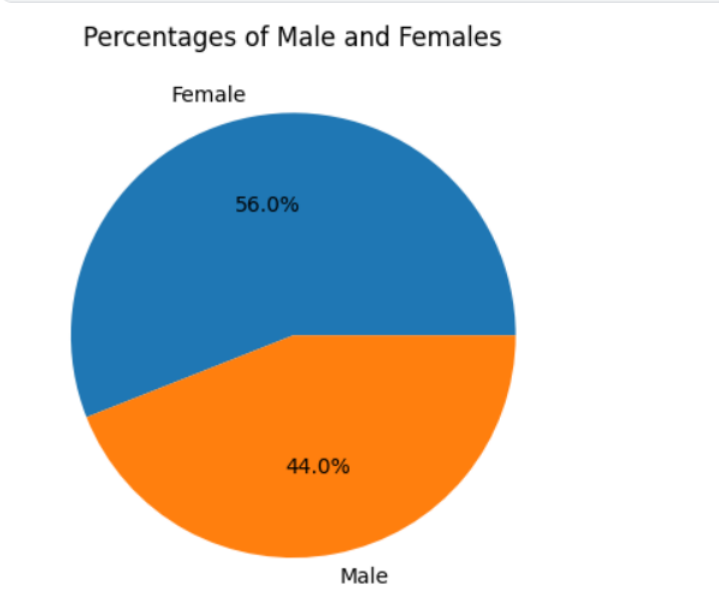
```
Null Values in Each Column:
CustomerID       0
Gender           0
Age              0
Annual_Income    0
Score            0
dtype: int64
```

The data contains none NULL values. Hence no data preprocessing is needed.

1.) Gender Distribution:

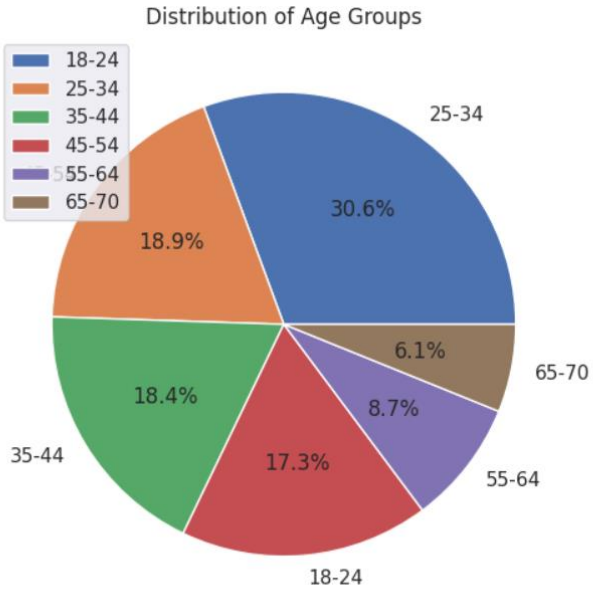


As can be seen from the bar plot our dataset contains 112 Female customers and 88 Male customers.

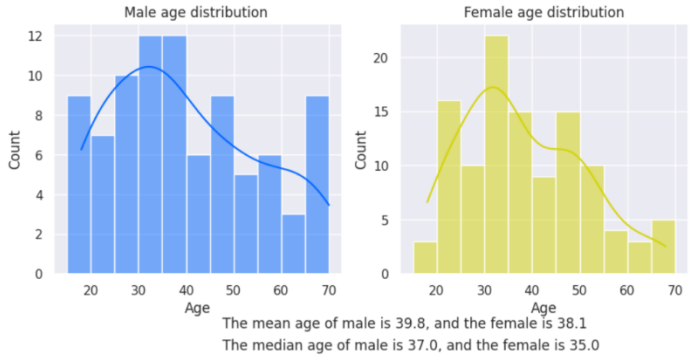


As can be seen from the bar plot our dataset contains 112 Female customers and 88 Male customers.

1.) Age Distribution:

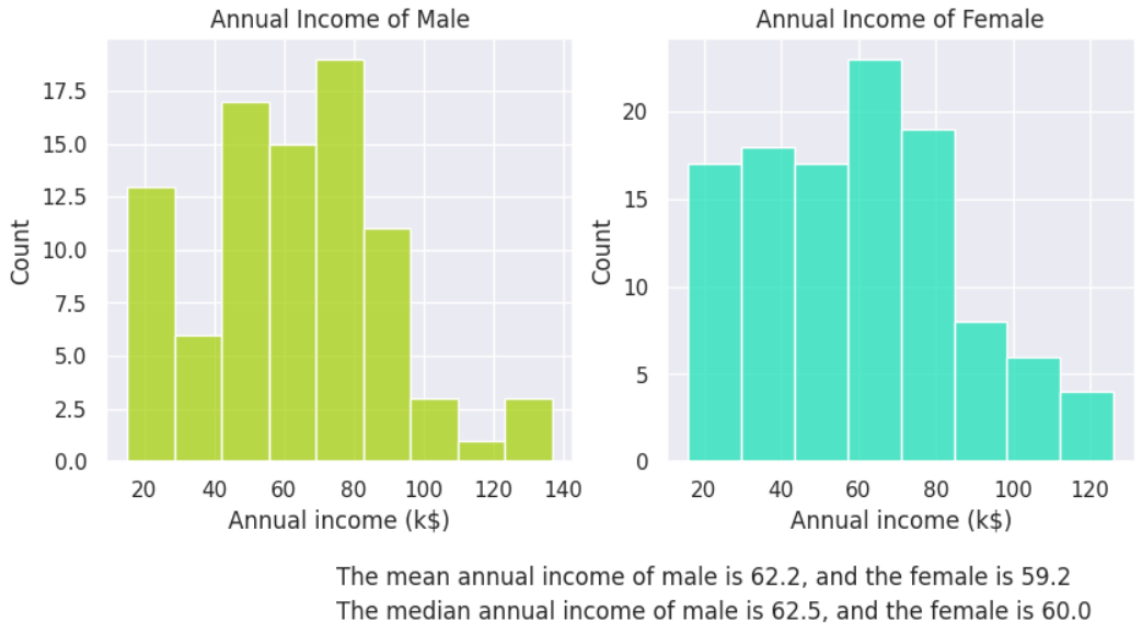


This pie chart shows the various age groups of the customers. Most of the customers are from the age group 18-24 which accounts for 30.6%. Age groups 25 -34 and 35-44 account for 18.9% and 18.4% respectively. 17.2% are the customers from the age group 45-54. Age group from 55-70 account for 14.8%.



The above histogram shows the count of customers in various age groups gender wise.
From the histogram we can conclude that the mean age of male is 39.8 years while mean age of female is 38.1 years. The median age of male is 37 years and median age of female is 35 years.

2.) Annual Income Distribution:



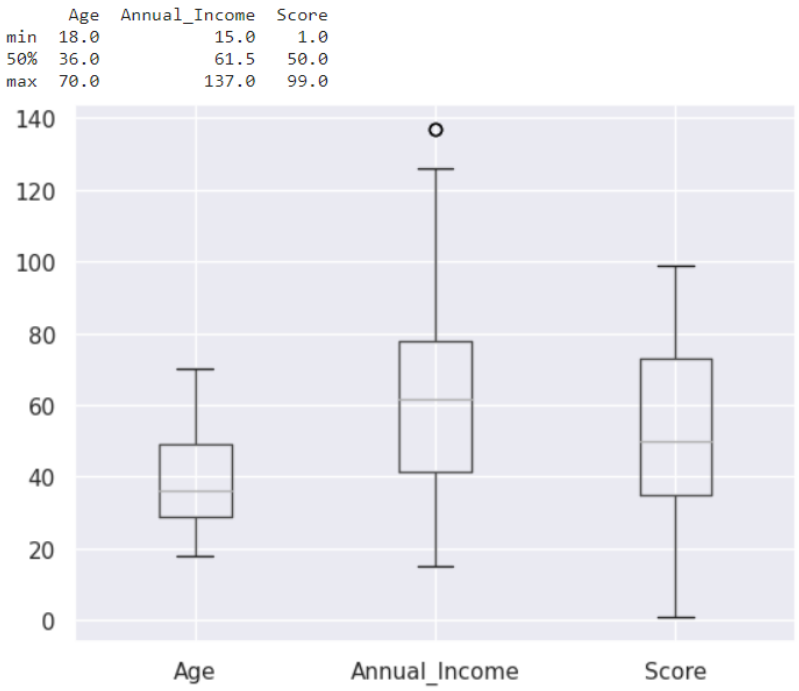
The mean annual income of male is 62.2, and the female is 59.2
The median annual income of male is 62.5, and the female is 60.0

The histogram represents the annual income of customers with respect to customer count gender wise. It can be observed that the mean income of male is $62.2K and mean income of female is $59.2k. The median income of male is $62.5k while the median income of female is $60k.

3.) Spending Score vs customer count:

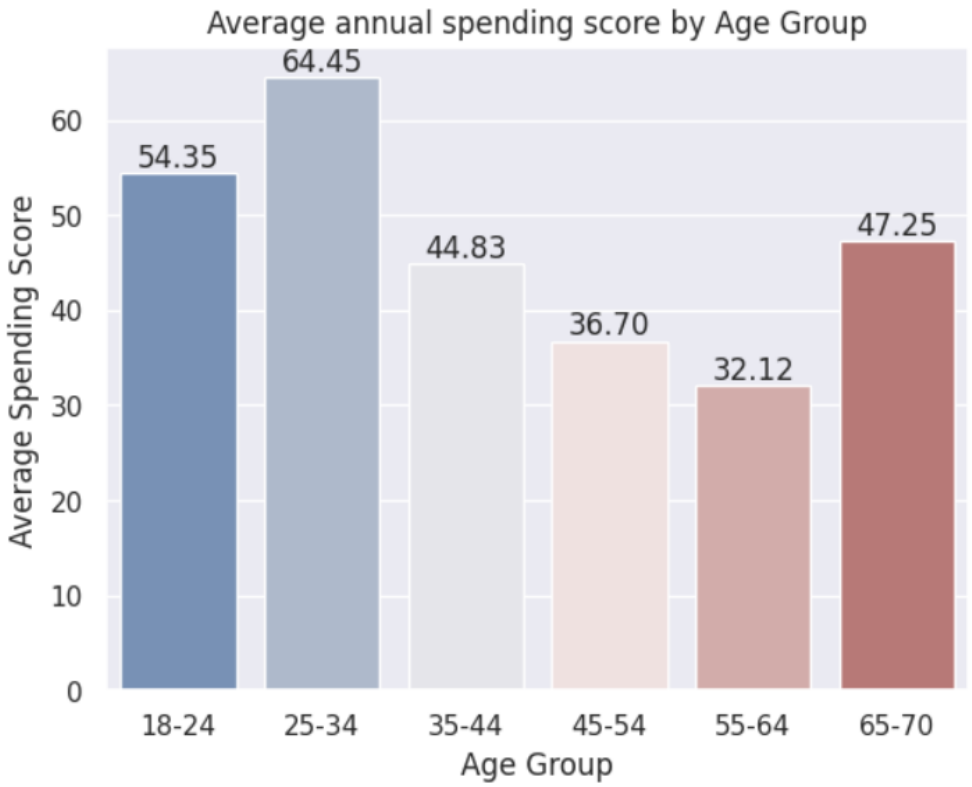
The above histogram shows the spending score of customers with respect to customer count gender wise. It can be observed that the mean spending score of male is 48.5 and mean income of female is 51.5. The median income of male is 50.0 and also the median income of female is 50.0.

4.) Box plot:



The box plot provides a concise summary of the distribution of three variables: Age, Annual Income, and Score. The Age distribution spans from a minimum of 18 years to a maximum of 70 years, with a median age of 36 years. Annual Income ranges from a minimum of $15,000 to a maximum of $137,000, with a median value of $61,500. The Score variable has a minimum of 1, a median of 50, and a maximum of 99.

5.) Bar plot Age group vs. average annual spending:



According to the above bar plot we can say, that the most average spending score comes from age 25 – 34 being 64.45 followed by age 18-24 with average spending score is 54.35. Age 35-44 comes at the 3rd place with the average score of 44.83 which is just little more than age 65-70 accounting for 47.25. The least average spending score comes from age 45-54 and 55-64 which is 36.70 and 35.12 respectively.

6.) Cluster scatter plot:

The cluster plot basically shows us the different clusters of customers based on their annual income and spending score. The blue cluster represents customers that have low annual income and low spending score. The green cluster represents high annual income customer with low spending score. Purple cluster represents customers with high spending score and low annual income. Yellow cluster representing high annual income and high spending score. The red cluster being the most ideal target market due to its decent annual income and spending score.

Conclusion:

In conclusion, the market segmentation analysis based on the provided data reveals several key insights. The dataset consists of 112 female customers and 88 male customers, with a predominant number of customers falling in the age group of 18-24 (30.6%). The mean age for males is 39.8 years, while for females, it is 38.1 years. The median age for males and females is 37 and 35 years, respectively. Regarding income, the mean income for males is $62.2K, and for females, it is $59.2K, with median incomes of $62.5K and $60K, respectively. In terms of spending score, the mean for males is 48.5, and for females, it is 51.5, with median scores of 50.0 for both genders. The cluster plot further emphasizes the distinct customer segments, with the red cluster identified as the most attractive market due to its decent annual income and spending score. Understanding these segmentation patterns allows for targeted marketing strategies tailored to the specific needs and behaviors of each customer group, enhancing overall marketing effectiveness and customer satisfaction.