EV MARKET ANALYSIS

(INDIA)

SAUMAN SARKAR

1. **Problem Statement:**

- You are a team working under an Electric Vehicle Startup. The Startup is still deciding in which vehicle/customer space it will be develop its EVs.
- You must analyse the Electric Vehicle market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles.

Solution:

1. **Fermi Estimaton:**

Q1. What EV vehicle should the startup enter the market with?

Ans: we can break down the problem into smaller components using rough assumptions to estimate key factors. The goal is to assess whether this space is worth pursuing. We'll consider the following aspects: market size, growth, and the number of units that could potentially be sold.

- **Percentage of EV sales in the four-wheeler market**:
  - As of 2023, EVs made up about 8% of new car sales. Let's estimate the current percentage of EVs is **10%** (accounting for growth), which means about:
    EV sales=15 million×10%=1.5 million EVs per year\text{EV sales} = 15 \, \text{million} \times 10\% = 1.5 \, \text{million EVs per year}EV sales=15million×10%=1.5million EVs per year
- **Startup's market penetration**:
  - Let's assume the startup aims for a modest 0.1% market share within its first few years of production. This would mean annual sales of:
    Startup EV sales=5.6 million×0.1%=5,600 units per year\text {Startup EV sales} = 5.6 \, \text{million} \times 0.1\% = 5,600 \, \text {units per year}
    Startup EV sales=5.6million×0.1%=5,600units per year
- **Average price of an EV**:
  - If the average price of a four-wheeler EV is around $40,000, then the potential revenue for the startup would be:
    Revenue=5,600 units×$40,000=$224 million per year\text {Revenue} = 5,600 \, \text{units} \times \$40,000 = \$224 \, \text {million per year}
    Revenue=5,600units×$40,000=$224million per year

Summary of Fermi Estimation for EV Four-Wheeler Market:

This estimation shows a clear opportunity for a startup to capture a growing segment of the market, with the potential for hundreds of millions in annual revenue by targeting the four-wheeler EV space. Of course, these numbers could vary widely depending on market conditions, competition, and the startup's capabilities.

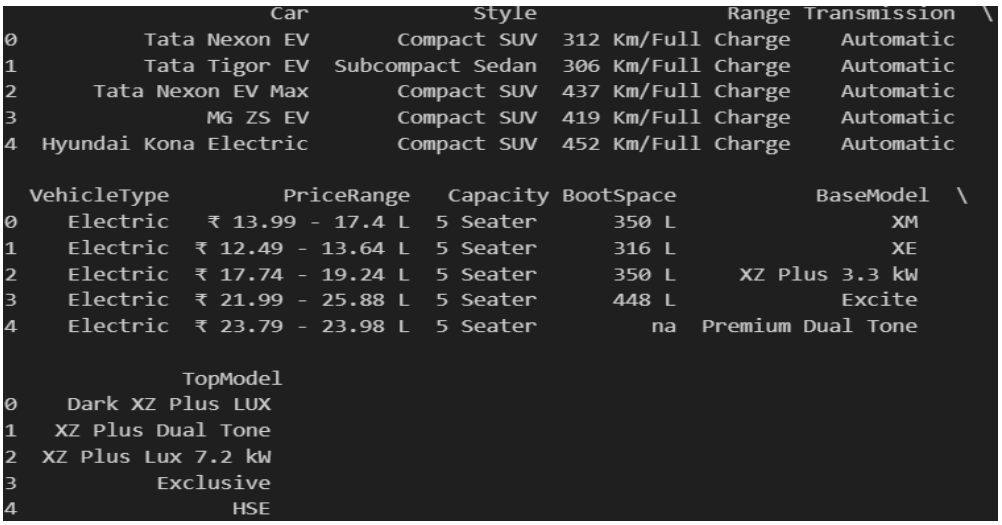
2. **Dataset Collected:**

   - EV_India vehicles.csv

- https://github.com/FROSTYOO/Feynn-Labs-2nd-Project-EV_market-segmentation_India/blob/main/EV_India%20vehicles.csv

3. **Data Preprocessing:**

- The code starts by importing necessary libraries, including pandas, NumPy, and matplotlib.
- It reads a CSV file containing the EV market dataset.
- The dataset is pre-processed by handling missing values, converting categorical variables to numeric, and encoding categorical variables using one-hot encoding.
- The pre-processed dataset is stored in a new data frame called **df_cleaned**.

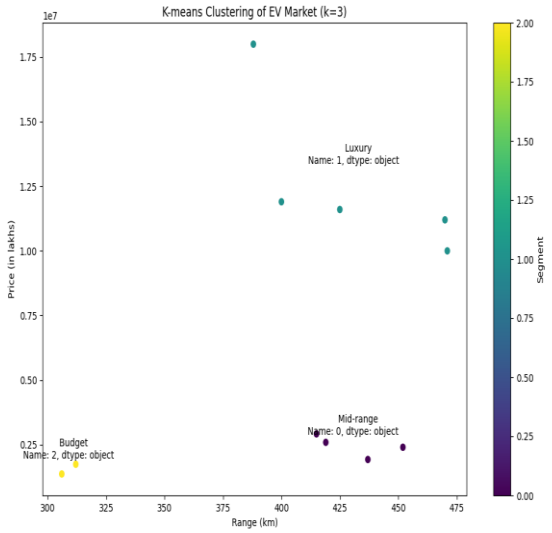


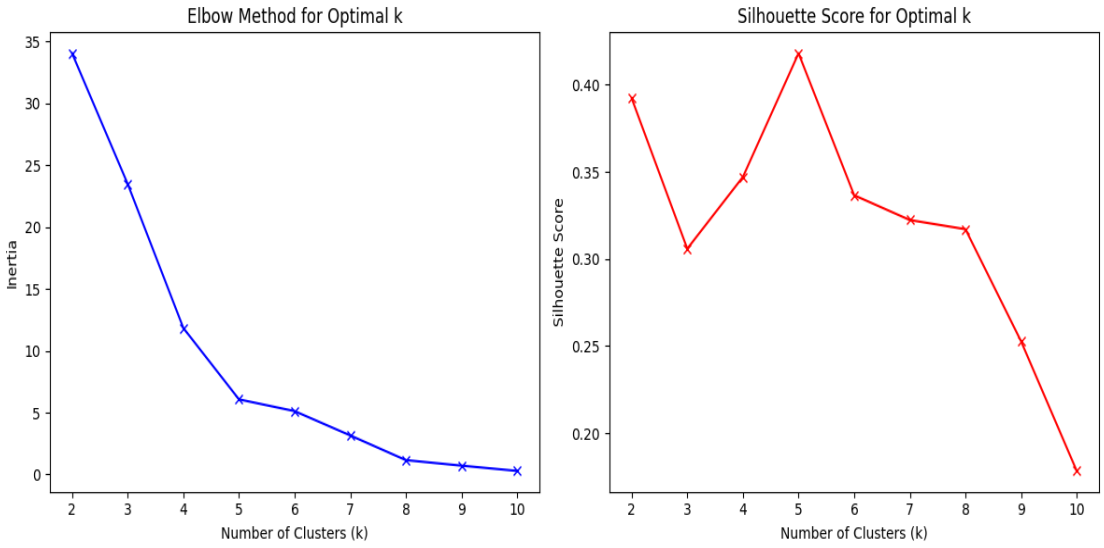
	Car	Style	Range	Transmission	VehicleType	PriceRange	Capacity	BootSpace	BaseModel	TopModel
0	Tata Nexon EV	Compact SUV	312 Km/Full Charge	Automatic	Electric	₹ 13.99 - 17.4 L	5 Seater	350 L	XM	Dark XZ Plus LUX
1	Tata Tigor EV	Subcompact Sedan	306 Km/Full Charge	Automatic	Electric	₹ 12.49 - 13.64 L	5 Seater	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

Step 2: Segmentation Using ML techniques and libraries used

- The code selects relevant features for segmentation, including price, range, boot space, and capacity.
- It normalizes the features using StandardScaler.
- K-means clustering is performed to segment the data into three clusters.
- The segments are labelled as 'Budget', 'Mid-range', and 'Luxury' based on their price means.
- The code performs K-means clustering to segment the EV market. It creates three segments (Budget, Mid-range, and Luxury) based on price, range, boot space, and capacity. The output includes segment analysis, distribution, and characteristics of each segment.
- Machine Learning Techniques and Libraries Used:
- 1. Clustering:
- - Technique: K-means clustering
- - Library: scikit-learn (sklearn)
- - Purpose: Market segmentation
- K-means clustering was used to segment the EV market into distinct groups based on features like price, range, boot space, and capacity. This unsupervised learning technique helped identify natural groupings in the data.
- Pandas: A library for data manipulation and analysis.
- NumPy: A library for numerical computing.
- Matplotlib: A library for creating static, animated, and interactive visualizations.
- Seaborn: A library for creating informative and attractive statistical graphics.
- Scikit-learn: A library for machine learning.

- SciPy: A library for scientific computing.
- Re: A library for regular expressions.





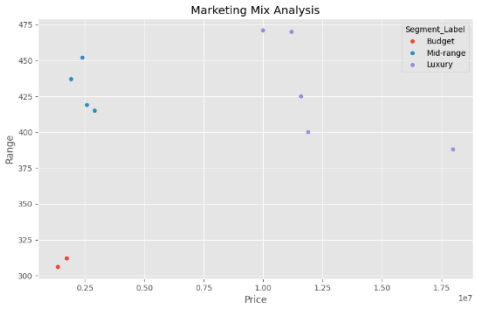
2. **Profiling and Describing Potential Segments**
   Profiling and describing potential segments is a crucial step in market research and customer analysis. It involves identifying and characterizing specific groups of customers or potential customers based on their demographics, needs, behaviours, and preferences.

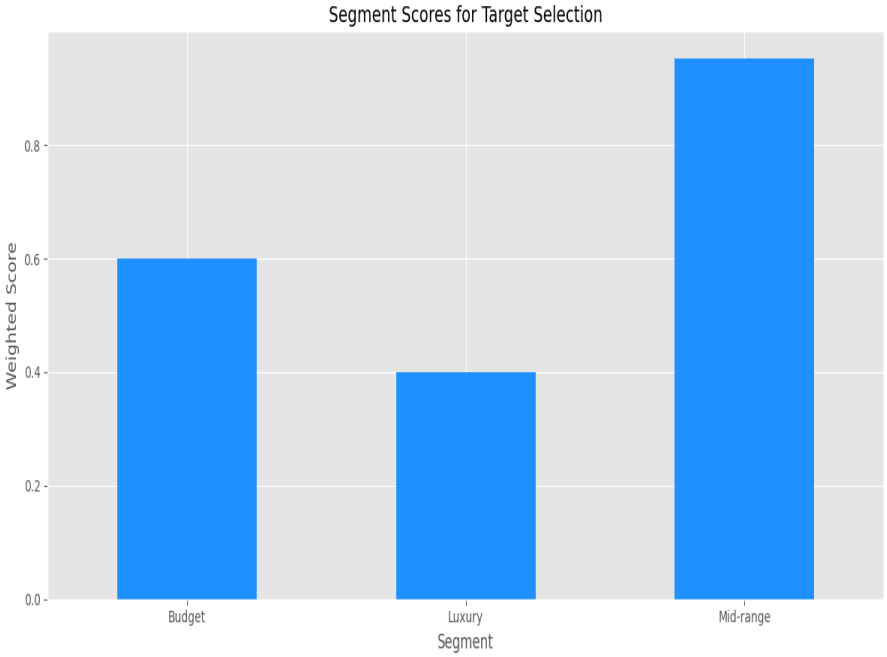
```
--- Budget Segment Profile ---
Count: 2
Most Common Type: Electric
Price Range: ₹1364000.00L - ₹1740000.00L
Avg Price: 1552000.0
Avg Range: 309.0
Avg Boot Space: 333.0
Avg Capacity: 5.0

Detailed Statistics:
            Range  Capacity   BootSpace         Price  Segment
count    2.000000       2.0    2.000000  2.000000e+00      2.0
mean   309.000000       5.0  333.000000  1.552000e+06      2.0
std      4.242641       0.0   24.041631  2.658721e+05      0.0
```

Selection of Target Segment

Once you have profiled and described potential segments, the next step is to select the target segment(s) that your business will focus on. This involves evaluating each segment's attractiveness and feasibility, and choosing the segment(s) that best align with your business goals and resources.

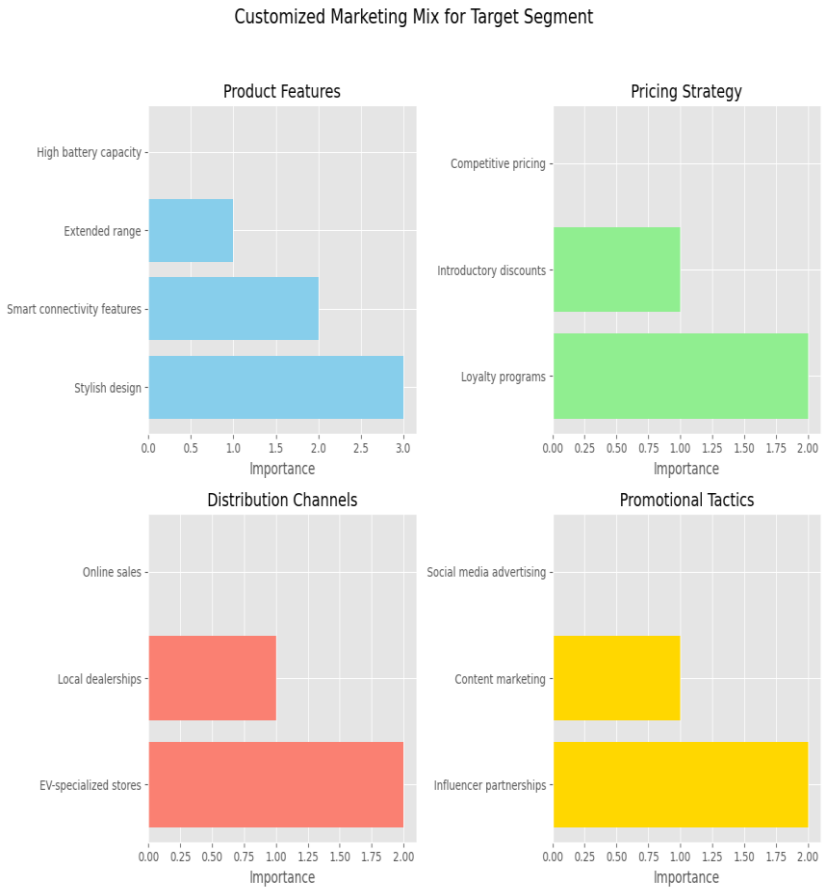
```
Summary Statistics for Target Segment:
       Range  Capacity  BootSpace  Price
count    0.0       0.0        0.0    0.0
mean     NaN       NaN        NaN    NaN
std      NaN       NaN        NaN    NaN
min      NaN       NaN        NaN    NaN
25%      NaN       NaN        NaN    NaN
50%      NaN       NaN        NaN    NaN
75%      NaN       NaN        NaN    NaN
max      NaN       NaN        NaN    NaN

Top 5 Vehicles in Target Segment:
Empty DataFrame
Columns: [Range, Capacity, BootSpace, Price, Segment_Label]
Index: []

Segment size comparison plot has been saved as 'segment_size_comparison.

Percentage of missing values in each column:
Range                          8.333333
Capacity                       0.000000
```





Customizing the marketing mix involves tailoring the 4Ps—Product, Price, Place, and Promotion—to effectively engage a specific target segment.

- **Product**: Adapt features, quality, variety, and services to meet the unique needs and preferences of the segment.
- **Price**: Choose a pricing strategy that aligns with the segment's purchasing power, offering discounts or flexible payment options as needed.
- **Place**: Select distribution channels that are accessible to the target audience, focusing on geographical areas where the segment is concentrated.
- **Promotion**: Develop targeted advertising and promotional strategies, using channels and messages that resonate with the segment's values and interests.



Conclusion For Q.1:

The provided code is a comprehensive analysis of the electric vehicle (EV) market in India, focusing on segmenting the market based on various features such as price, range, boot space, and capacity. The code uses K-means clustering to identify three distinct segments: Budget, Mid-range, and Luxury.

The analysis reveals that the Budget segment has the highest average range, while the Luxury segment has the highest average price. The Mid-range segment has a balance of price and range, making it an attractive option for customers.

The code also provides a detailed profile of each segment, including statistical summaries of numeric columns and vehicle type distribution. Additionally, it visualizes the segment comparison using bar plots and heatmaps.

The selection of the target segment is based on a weighted score that balances price and range. The target segment is selected as the one with the highest weighted score.

Finally, the code provides a customized marketing mix for the target segment, including product features, pricing strategy, distribution channels, and promotional tactics.

Key Takeaways

- The EV market in India can be segmented into three distinct segments: Budget, Mid-range, and Luxury.
- The Budget segment has the highest average range, while the Luxury segment has the highest average price.
- The Mid-range segment has a balance of price and range, making it an attractive option for customers.
- The selection of the target segment should be based on a weighted score that balances price and range.

- A customized marketing mix can be developed for the target segment, including product features, pricing strategy, distribution channels, and promotional tactics.
    - Recommendations
- EV manufacturers should focus on developing products that cater to the Mid-range segment, which has a balance of price and range.
- The marketing mix should be customized for the target segment, taking into account their specific needs and preferences.
- The analysis should be regularly updated to reflect changes in the market and customer preferences.
    - Future Work
- The analysis can be extended to include additional features such as charging time, battery size, and safety features.
- The marketing mix can be further customized based on customer feedback and market trends.
- The analysis can be applied to other markets and regions to identify similarities and differences in customer preferences.

GitHub Link1: https://github.com/FROSTYOO/Feynn-Labs-2nd-Project-EV_market-segmentation_India/blob/main/Ev_cars_India.py

Q2. What should me targeted segment for the ev market.

Fermi Estimation:

- Customer Space: Middle-class segment in India.
- Target Customer Segment: 60 million people who are interested in buying EVs.
- Specific Needs and Preferences: Affordable EVs with a range of at least 200 km, battery life of at least 5 years, and advanced safety features.
- Competition: 5 players in the EV market in India that are targeting the middle-class segment.
- Revenue Potential: ₹3,000 crores
- It will be based on market scenarios which contains the factors such as behavioural, geographical, demographic etc.

Datasets Used:

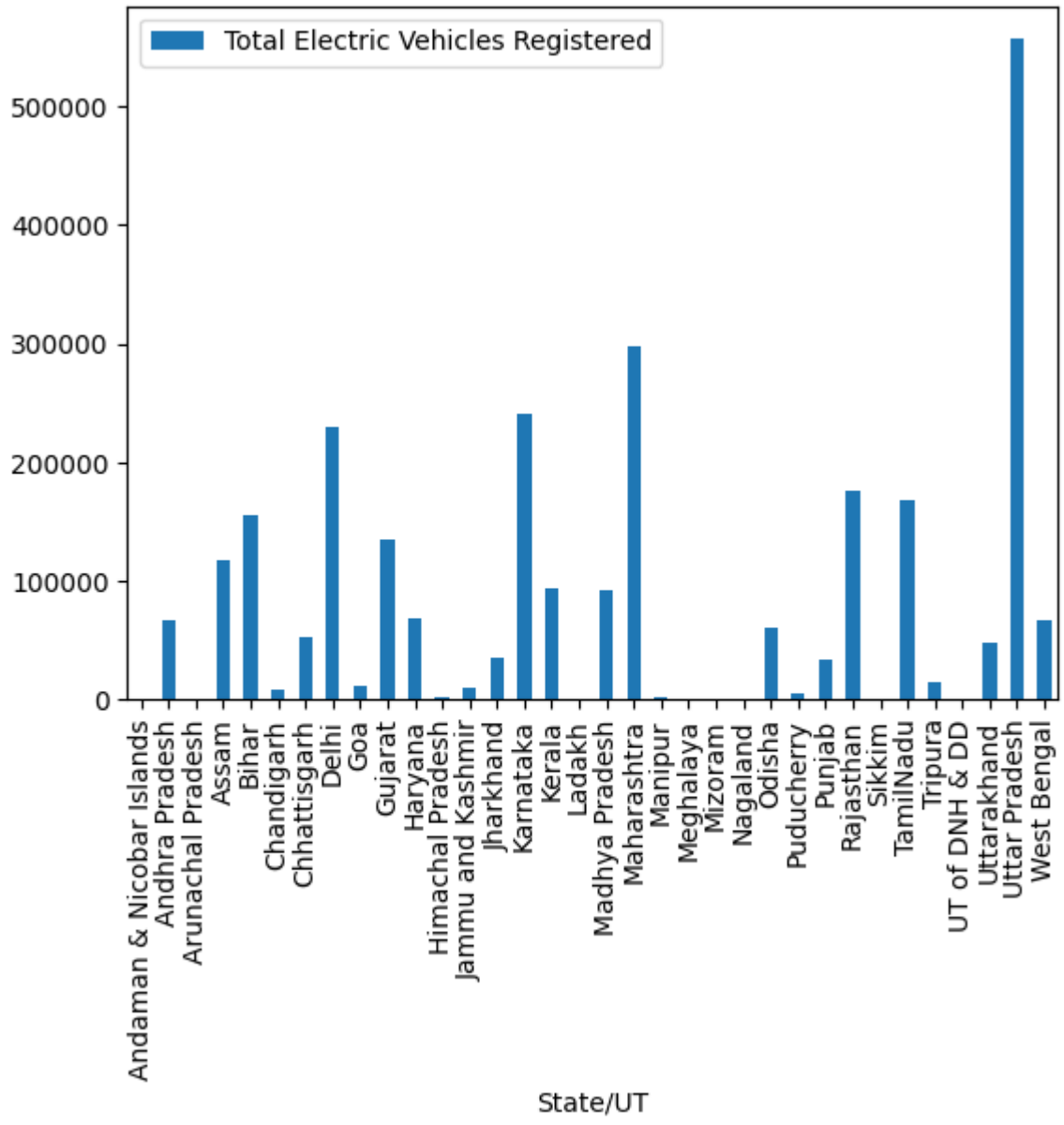
- https://github.com/FROSTYOO/Feynn-Labs-2nd-Project-EV_market-segmentation_India/blob/main/RS_Session_258_AU_1241_2.i_data_gov_in.csv
- https://github.com/FROSTYOO/Feynn-Labs-2nd-Project-EV_market-segmentation_India/blob/main/data1.xlsx
- https://github.com/FROSTYOO/Feynn-Labs-2nd-Project-EV_market-segmentation_India/blob/main/demographicData.csv
- https://github.com/FROSTYOO/Feynn-Labs-2nd-Project-EV_market-segmentation_India/blob/main/evcount.xlsx

Based on Geographic Conditions:

```
#%%
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#read the excel
df1 = pd.read_excel(r'C:\Users\sauma\OneDrive\Desktop\EV_Market Analysis India\Dataset\data1.xlsx')

print(df1.head())
```

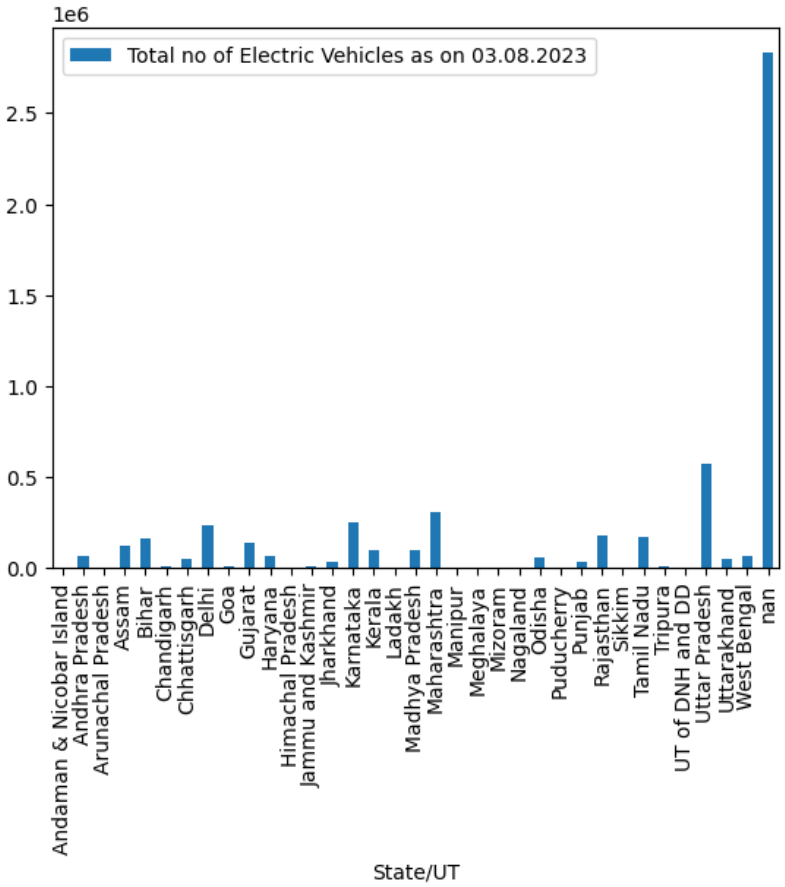
Data Preprocessing:

```
# %%
print(df1.info())
print(df1.shape)
print(df1.isnull().sum())
print(df1.describe())

df1.plot.bar(x='State/UT',y='Total Electric Vehicles Registered')
plt.show()

#sort the values
df1.sort_values(['Total Electric Vehicles Registered'], ascending=False).head(10)
```



- We can see from the graph that the states Uttar Pradesh, Maharashtra, Delhi, Karnataka
- are the major states with highest number of electric vehicles.

```
#%%
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df3 =pd.read_excel(r'C:\Users\sauma\OneDrive\Desktop\EV_Market Analysis India\Dataset\evcount.xlsx')
print(df3.head())
print(df3.info())
print(df3.shape)
print(df3.isnull().sum())
print(df3.describe())
#plot a bar graph for df3 states on x label vs electric vehicle count on y label
df3.plot.bar(x='State/UT',y='Total no of Electric Vehicles as on 03.08.2023')
plt.show()
```



- From the graph we can conclude that for geographic conditions we can choose states like Maharashtra, Delhi, Karnataka, UP as the entry market foe our EV startup.

Demographic Condition:

Dataset Used:

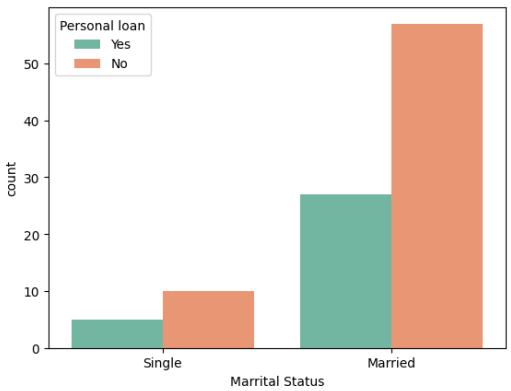
Data Loading and Preprocessing:

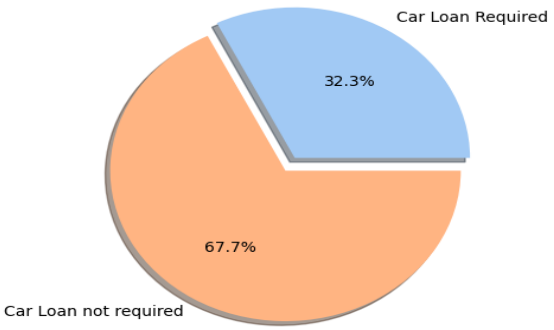
```python
# %%
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import preprocessing
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
df4 =pd.read_csv(r'C:\Users\sauma\OneDrive\Desktop\EV_Market Analysis India\Dataset\demographicData.csv')
print(df4.head())
print(df4.info())
print(df4.shape)
print(df4.isnull().sum())
print(df4.describe())

# Plotting the Car loan status with respect to Marrital Status
sns.countplot(x ='Marrital Status', hue = 'Personal loan',  data = df4, palette = 'Set2')
plt.show()
```

```
   Age Profession Marrital Status      Education  No of Dependents  \
0   27   Salaried          Single  Post Graduate                 0
1   35   Salaried         Married  Post Graduate                 2
2   45   Business         Married       Graduate                 4
3   41   Business         Married  Post Graduate                 3
4   31   Salaried         Married  Post Graduate                 2

  Personal loan House Loan Wife Working   Salary  Wife Salary  Total Salary  \
0           Yes         No           No   800000            0        800000
1           Yes        Yes          Yes  1400000       600000       2000000
2           Yes        Yes           No  1800000            0       1800000
3            No         No          Yes  1600000       600000       2200000
4           Yes         No          Yes  1800000       800000       2600000

     Make    Price
0     i20   800000
1    Ciaz  1000000
2  Duster  1200000
3    City  1200000
4     SUV  1600000
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 13 columns):
```

Visualizations:

Based on marriage and loan:

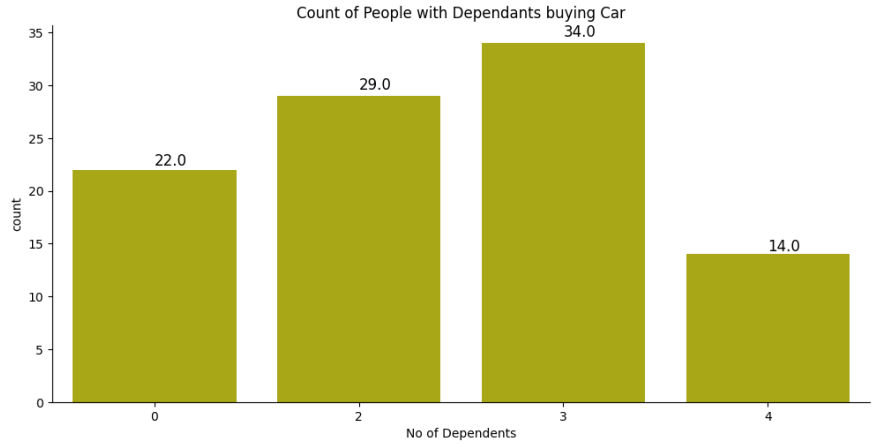




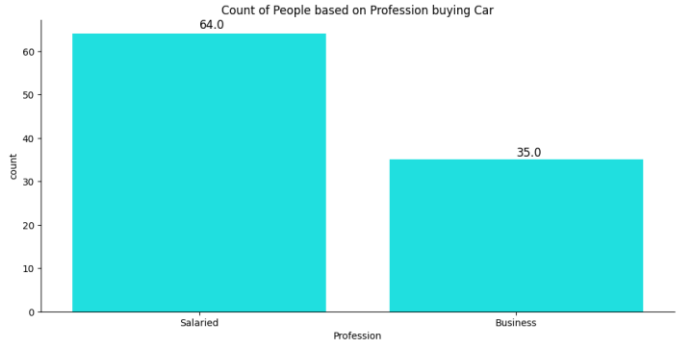
- From the above graphs we can see that the persons who are married takes high percentage of personal loan.
- From the pe chart we can conclude that maximum percentage does not require car loans.

Based on dependencies and professional categories:

```
# %%
g = sns.catplot(x = "No of Dependents", data=df4, aspect=2, kind="count", color='y')

ax = g.facet_axis(0,0)
for p in ax.patches:
    ax.text(p.get_x() + p.get_width() / 2,
            p.get_height() * 1.02,
            format(p.get_height()),
            color='black', rotation='horizontal', size='large')
plt.title('Count of People with Dependants buying Car', color='black')
plt.show()
```



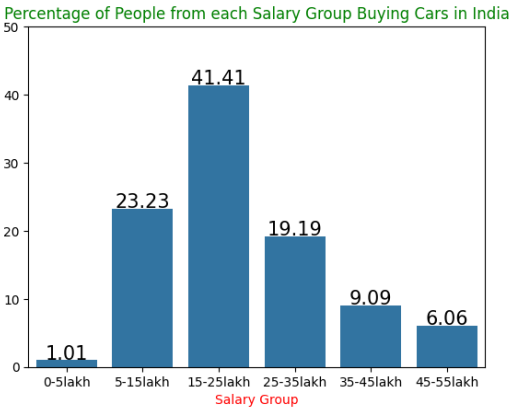

```
# %%
g = sns.catplot(x = "Profession", data=df4, aspect=2, kind="count", color='cyan')

ax = g.facet_axis(0,0)
for p in ax.patches:
    ax.text(p.get_x() + p.get_width() / 2,
            p.get_height() * 1.02,
            format(p.get_height()),
            color='black', rotation='horizontal', size='large')
plt.title('Count of People based on Profession buying Car', color='black')
plt.show()
```



- Here from the above graphs we can conclude the number of people buys the car on the basis of family dependents and professional status.

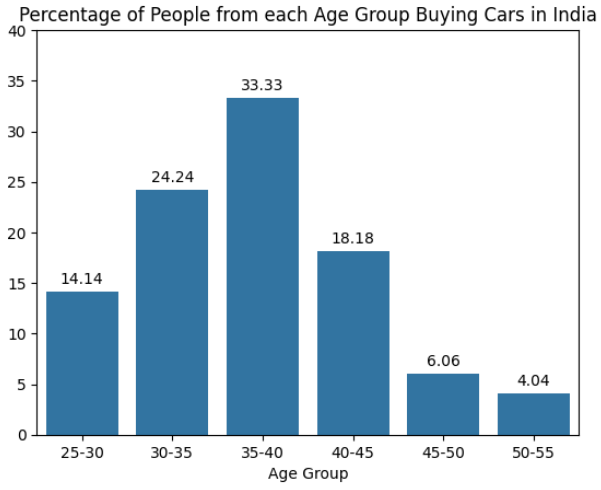
Based on salary group and age group:

```
SalaryGroup={}
length=df4.shape[0]
SalaryGroup['0-5lakh']=round(df4[(df4['Total Salary']>=100000)&(df4['Total Salary']<500000)].s
SalaryGroup['5-15lakh']=round(df4[(df4['Total Salary']>=500000)&(df4['Total Salary']<1500000)]
SalaryGroup['15-25lakh']=round(df4[(df4['Total Salary']>=1500000)&(df4['Total Salary']<2500000
SalaryGroup['25-35lakh']=round(df4[(df4['Total Salary']>=2500000)&(df4['Total Salary']<3500000
SalaryGroup['35-45lakh']=round(df4[(df4['Total Salary']>=3500000)&(df4['Total Salary']<4500000
SalaryGroup['45-55lakh']=round(df4[(df4['Total Salary']>=4500000)&(df4['Total Salary']<5500000
fig,ax=plt.subplots()
plots=sns.barplot(x = list(SalaryGroup.keys()), y = list(SalaryGroup.values()))
for bar in plots.patches:
    plots.annotate(format(bar.get_height(), '.2f'),
                   (bar.get_x() + bar.get_width() / 2,
                    bar.get_height()), ha='center', va='center',
                   size=15, xytext=(0, 5),
                   textcoords='offset points')
ax.set_ylim(0,50)
plt.title('Percentage of People from each Salary Group Buying Cars in India', color='green')
plt.xlabel('Salary Group', color='r',)
plt.show()
```



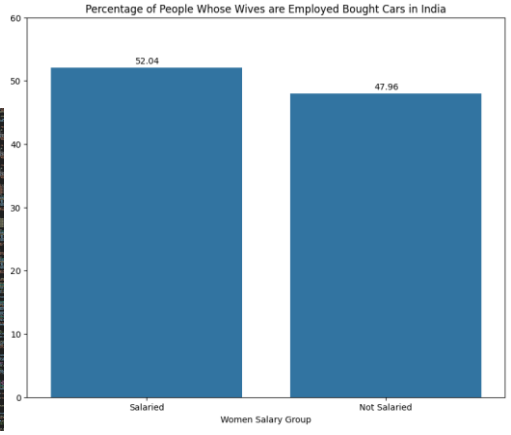
- Here we can conclude that the person with a salary range from 15-25 lakhs buys cars in India.

```
# %%
AgeGroup={}
length=df4.shape[0]
AgeGroup['25-30']=round(df4[(df4['Age']>=25)&(df4['Age']<30)].shape[0]*100/length,2)
AgeGroup['30-35']=round(df4[(df4['Age']>=30)&(df4['Age']<35)].shape[0]*100/length,2)
AgeGroup['35-40']=round(df4[(df4['Age']>=35)&(df4['Age']<40)].shape[0]*100/length,2)
AgeGroup['40-45']=round(df4[(df4['Age']>=40)&(df4['Age']<45)].shape[0]*100/length,2)
AgeGroup['45-50']=round(df4[(df4['Age']>=45)&(df4['Age']<50)].shape[0]*100/length,2)
AgeGroup['50-55']=round(df4[(df4['Age']>=50)&(df4['Age']<55)].shape[0]*100/length,2)

fig,ax=plt.subplots()
plots=sns.barplot(x=list(AgeGroup.keys()),y=list(AgeGroup.values()))
for bar in plots.patches:
    plots.annotate(format(bar.get_height(), '.2f'),
                   (bar.get_x() + bar.get_width() / 2,
                    bar.get_height()), ha='center', va='center',
                   size=10, xytext=(0, 8),
                   textcoords='offset points')
ax.set_ylim(0,40)
plt.title('Percentage of People from each Age Group Buying Cars in India')
plt.xlabel('Age Group')
plt.show()
```



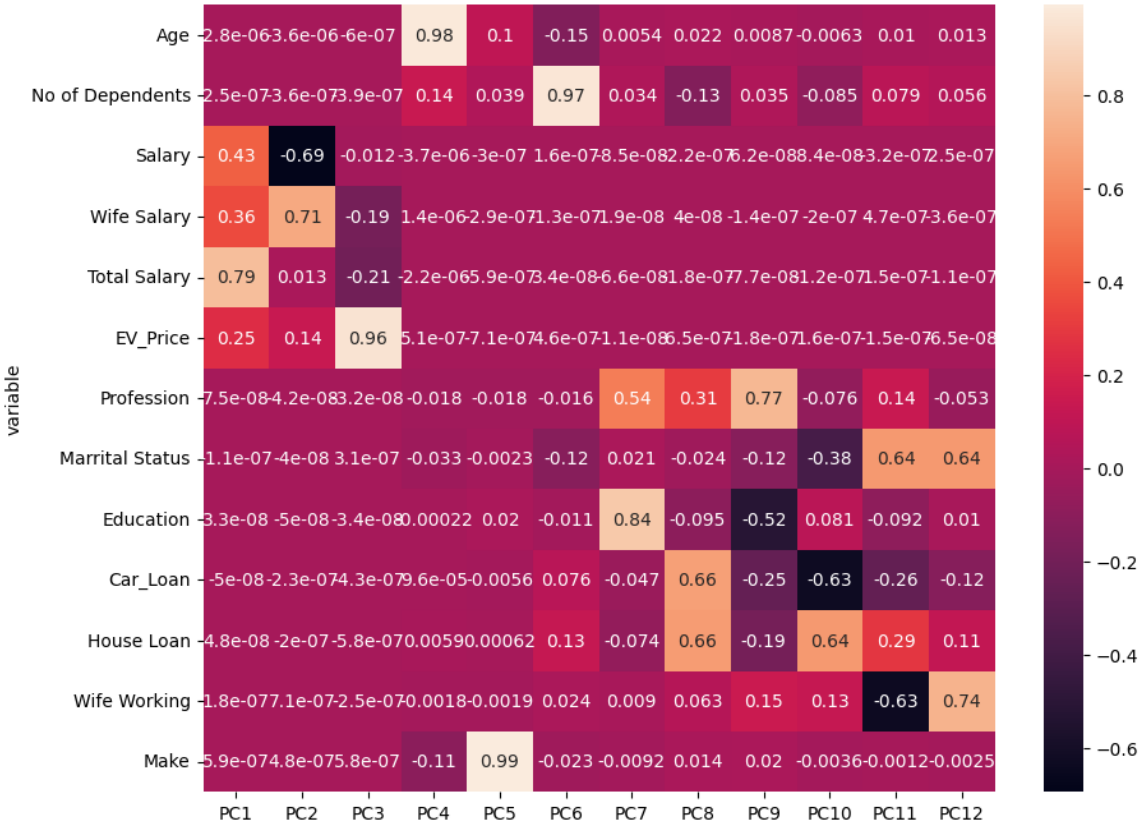
- From the plot we can conclude that the persons at the age group of 35-40 buys most cars in India.

```
WomenSalaryGroup={}
length=df4.shape[0]
WomenSalaryGroup['Salaried'] = round(df4[(df4['Wife Salary']>0)].shape[0]*100/length,2)
WomenSalaryGroup['Not Salaried'] = round(df4[(df4['Wife Salary'] == 0)].shape[0]*100/length,2)

fig,ax=plt.subplots()
plots=sns.barplot(x=list(WomenSalaryGroup.keys()),y=list(WomenSalaryGroup.values()))
for bar in plots.patches:
    plots.annotate(format(bar.get_height(), '.2f'),
                   (bar.get_x() + bar.get_width() / 2,
                    bar.get_height()), ha='center', va='center',
                   size=10, xytext=(1, 8),
                   textcoords='offset points')
ax.set_ylim(0,60)
plt.title('Percentage of People Whose Wives are Employed Bought Cars in India')
plt.xlabel('Women Salary Group')
plt.show()
```



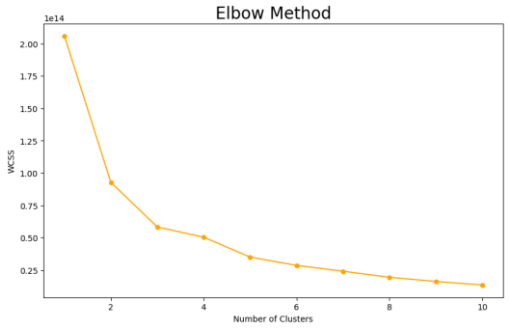
- Here the graphs show info about the women's salary status and who buys cars or not.

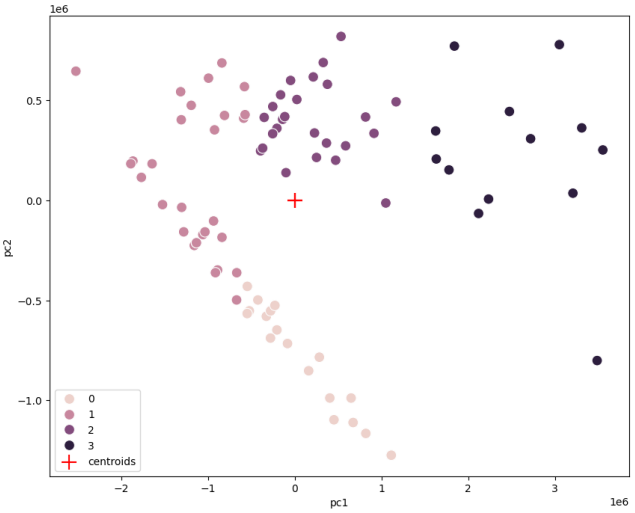
```
# %%
plt.rcParams['figure.figsize'] = (10,8)
ax = sns.heatmap(loadings_df, annot=True)
plt.show()
```




```
# %%
wcss=[]
for i in range(1,11):
    #preventing random initialization: 'init=k-means++'
    kmeans= KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
```

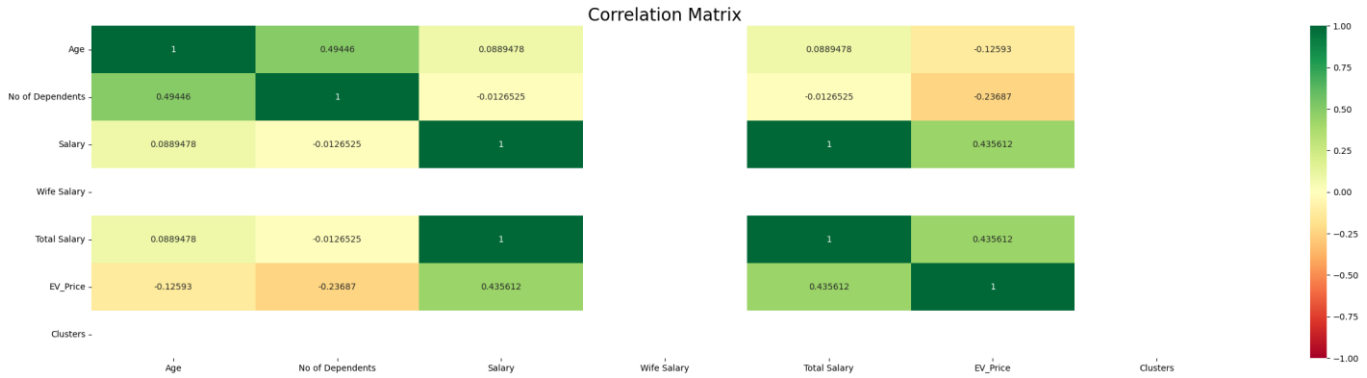
```
Run Cell | Run Above | Debug Cell | Go to [78]
# %%
plt.figure(figsize=(10,6))
plt.plot(range(1,11),wcss,color='orange', linestyle='solid', marker='o',
         markersize=5)
plt.title('Elbow Method', size=20)
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.show()
```






```
# %%
df4['Clusters'] = x['cluster']


newdf = df4.groupby('Clusters')


newdf.get_group(0)
```

	Age	Profession	Marrital Status	Education	No of Dependents	Car_Loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	EV_Price	Clu
2	45	Business	Married	Graduate	4	Yes	Yes	No	1800000	0	1800000	Duster	1200000	
8	34	Business	Married	Post Graduate	4	No	No	No	2000000	0	2000000	Verna	1100000	
34	42	Salaried	Married	Graduate	4	Yes	Yes	No	2100000	0	2100000	Ciaz	1100000	
50	49	Business	Married	Post Graduate	4	No	No	No	2000000	0	2000000	Duster	1300000	
52	44	Salaried	Married	Post Graduate	4	No	No	No	2700000	0	2700000	SUV	1600000	
55	41	Salaried	Married	Post Graduate	3	Yes	Yes	No	3100000	0	3100000	Creata	1500000	
56	41	Salaried	Married	Graduate	3	No	No	No	2600000	0	2600000	Ciaz	1100000	



- **Conclusion for Q2.:**

The code is a comprehensive analysis of the electric vehicle (EV) market in India, focusing on various aspects such as the number of EVs registered, charging stations, and demographic data of car buyers.

The code starts by reading in four different datasets:

- df1: Data on the number of EVs registered in different states in India.
- df2: Data on the number of charging stations sanctioned in different states in India.
- df3: Data on the total number of EVs in different states in India.
- df4: Demographic data of car buyers in India.

The code then performs various analyses on these datasets, including:

- Data visualization: The code creates bar plots to visualize the number of EVs registered, charging stations, and demographic data of car buyers.
- Data cleaning: The code cleans the data by removing missing values and encoding categorical variables.
- Principal component analysis (PCA): The code performs PCA on the demographic data to reduce the dimensionality.
- K-means clustering: The code performs K-means clustering on the demographic data to group car buyers into different clusters.
- Hierarchical clustering: The code performs hierarchical clustering on the demographic data to group car buyers into different clusters.
- Correlation analysis: The code performs correlation analysis on the demographic data to understand the relationships between different variables.

The code also creates various visualizations, including:

- Bar plots to visualize the number of EVs registered, charging stations, and demographic data of car buyers.
- Heatmaps to visualize the correlation matrix of the demographic data.
- Scatter plots to visualize the clusters of car buyers.

➢ **Key Findings**

- The number of EVs registered in India is increasing rapidly, with Maharashtra, Delhi, and Karnataka being the top three states.
- The number of charging stations sanctioned in India is also increasing, with Maharashtra, Delhi, and Karnataka being the top three states.
- The demographic data of car buyers in India shows that most car buyers are married, have a high income, and are between the ages of 25-40.
- The PCA analysis shows that the demographic data can be reduced to 12 principal components, which explain 95% of the variance.

- The K-means clustering analysis shows that car buyers can be grouped into four clusters, with each cluster having distinct characteristics.
- The hierarchical clustering analysis shows that car buyers can be grouped into four clusters, with each cluster having distinct characteristics.
- The correlation analysis shows that there are strong correlations between different variables in the demographic data.

➢ **Recommendations**

- The government should incentivize the adoption of EVs by providing subsidies and tax benefits.
- The government should invest in building more charging stations to support the growth of EVs.
- Car manufacturers should focus on producing EVs that are affordable and have a long range.
- Car manufacturers should also focus on producing EVs that are suitable for different types of car buyers, such as families and individuals.
- The government should also invest in educating car buyers about the benefits of EVs and promoting their adoption.
- Overall, the code provides a comprehensive analysis of the EV market in India and provides insights into the demographic data of car buyers. The findings and recommendations of the code can be used to inform policy decisions and business strategies in the EV industry.
- By this we can conclude that the target segment for EV startup will be the markets of UP, Maharashtra, Karnataka etc are geographically best areas to target market.
- Age group of 25-40 were most demanded one and hence most targeted as well. People from age group 25-31 with no of dependants > 2 and salary between 13 lakhs and 17 lakhs are buying cars.

GitHub Link 2: https://github.com/FROSTYOO/Feynn-Labs-2nd-Project-EV_market-segmentation_India/blob/main/Target%20Segment.py