Name : Pradnya Devendra Ukey

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 have to analysis 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.

To better understand the market segmentation of the Electric Vehicle (EV) market in India, let's perform a Fermi estimation based on the given problem statement.

When opening a new startup based on EV market, the analysis of the preferred types of vehicles by the end-users/ customers is of utmost importance. So, the problem statement that I am covering here is to analyze the age segment of the data available. This will aid us in preparing new strategies to deploy in marketing and increasing revenue or sales

ELECTRIC VEHICLE MARKET

The electric vehicle (EV) market is growing rapidly, driven by several factors, including increasing environmental consciousness, government incentives, and technological advancements. In 2022, global sales of electric cars reached a record 10 million, and sales are expected to continue to grow in the coming years. The largest EV market is currently China, followed by the United States and Europe. China has been a major driver of the EV market, due to government policies that have supported the development of the industry. The United States and Europe are also seeing strong growth in EV sales, as consumers become more aware of the benefits of electric vehicles. There are a few factors that are expected to continue to drive the growth of the EV market in the coming years. These include:

• **Increased environmental consciousness:** As consumers become more aware of the environmental impact of gasoline-powered vehicles, they are increasingly looking for alternative transportation options. Electric vehicles are zero-emission vehicles, which makes them a more sustainable choice.

• **Government incentives:** Many governments around the world are offering incentives to consumers who purchase electric vehicles. These incentives can take the form of tax breaks, rebates, or free parking.

• **Technological advancements:** Technological advancements in battery technology are making electric vehicles more affordable and practical. Batteries are becoming more energy-dense, which means that electric vehicles can travel longer distances on a single charge. The growth of the EV market is having a significant impact on the automotive industry. Traditional automakers are investing heavily in electric vehicle development, and new startups are entering the market. The EV market is still in its early stages, but it is poised for significant growth in the coming years. Here are some of the key trends in the EV market:

• **The rise of battery electric vehicles (BEVs):** BEVs are the most common type of electric vehicle, and they are expected to continue to dominate the market in the coming years. BEVs are powered by batteries, and they do not have an internal combustion engine.

• **The growth of plug-in hybrid electric vehicles (PHEVs):** PHEVs are a type of electric vehicle that has both an electric motor and an internal combustion engine. PHEVs can be driven on electricity for a limited distance, and then the internal combustion engine kicks in.

• **The development of hydrogen fuel cell vehicles (FCEVs):** FCEVs are a type of electric vehicle that uses hydrogen fuel cells to generate electricity. FCEVs are still in the early stages of development, but they have the potential to offer longer range and faster refuelling than BEVs. The EV market is a rapidly growing and evolving market. The trends mentioned above are just a few of the factors that are shaping the future of the EV market. It will be interesting to see how the market develops in the coming

Data Source

Data view


```python
data = pd.read_csv('ev_data.csv')
data.head()
```

	age	city	profession	marital_stat	education	family_count	annual_income	replace_with_ev	preferred_ev_type	think_ev_are_economical	preferred_brand	preferred_price_range	preferred_wheel_count
0	30	New Delhi	Working Professional	Single	Post Graduate	4	1.770466e+06	Yes	SUV	Yes	MG	<5 lakhs	2
1	27	Ahmedabad	NaN	Single	Post Graduate	4	7.600731e+05	Yes	SUV	Yes	Nissan	<15 lakhs	4
2	28	Pune	Working Professional	Single	Post Graduate	4	2.779701e+06	Maybe	Sedan	Yes	Honda	<5 lakhs	4
3	26	Pune	Salaried	Single	Post Graduate	4	1.473698e+06	Yes	Sedan	Yes	Tata	<25 lakhs	2
4	25	Pune	Working Professional	Single	Post Graduate	4	2.430643e+06	Yes	Sedan	Yes	Hyundai	<15 lakhs	4

Steps used:

- Data preprocessing
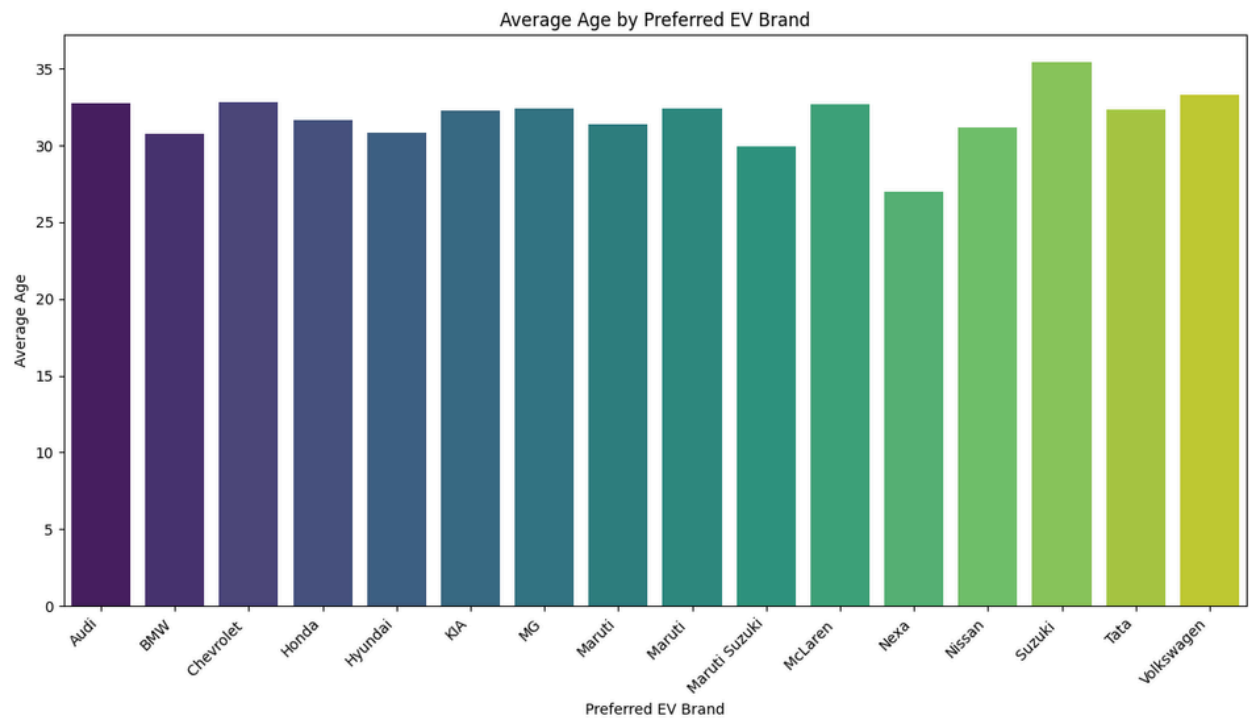- Finding optimum number of clusters:

Scree/ Elbow Plot
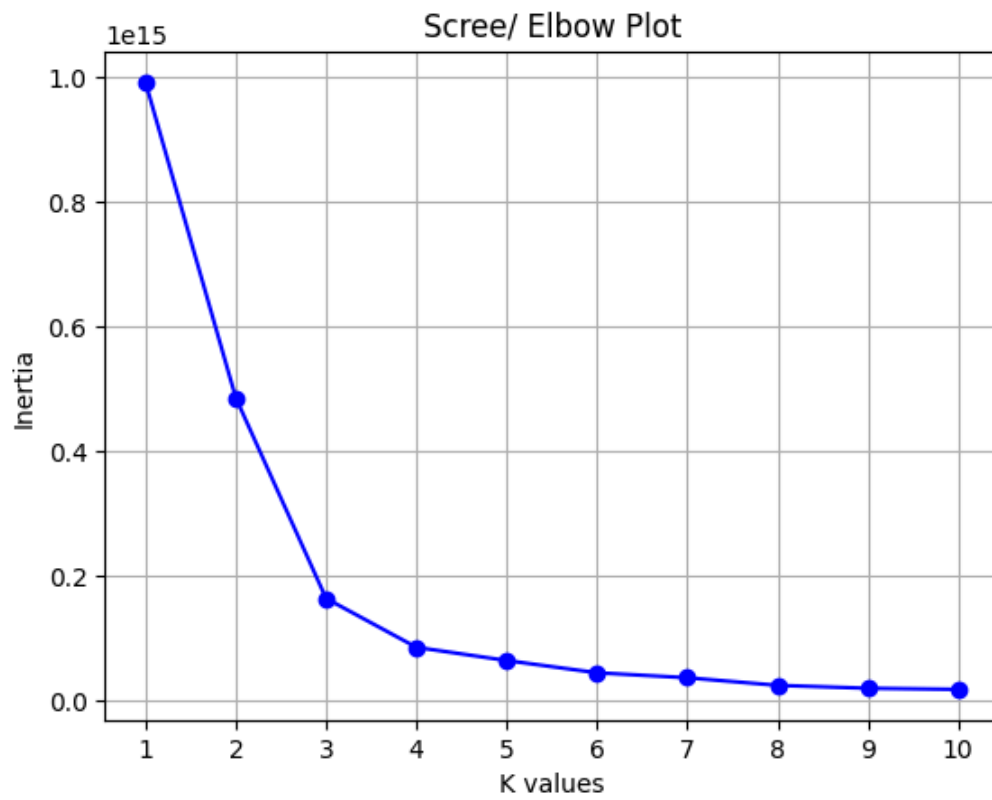
- we fit and predict the model using n_clusters = 4, and obtain the following plot:
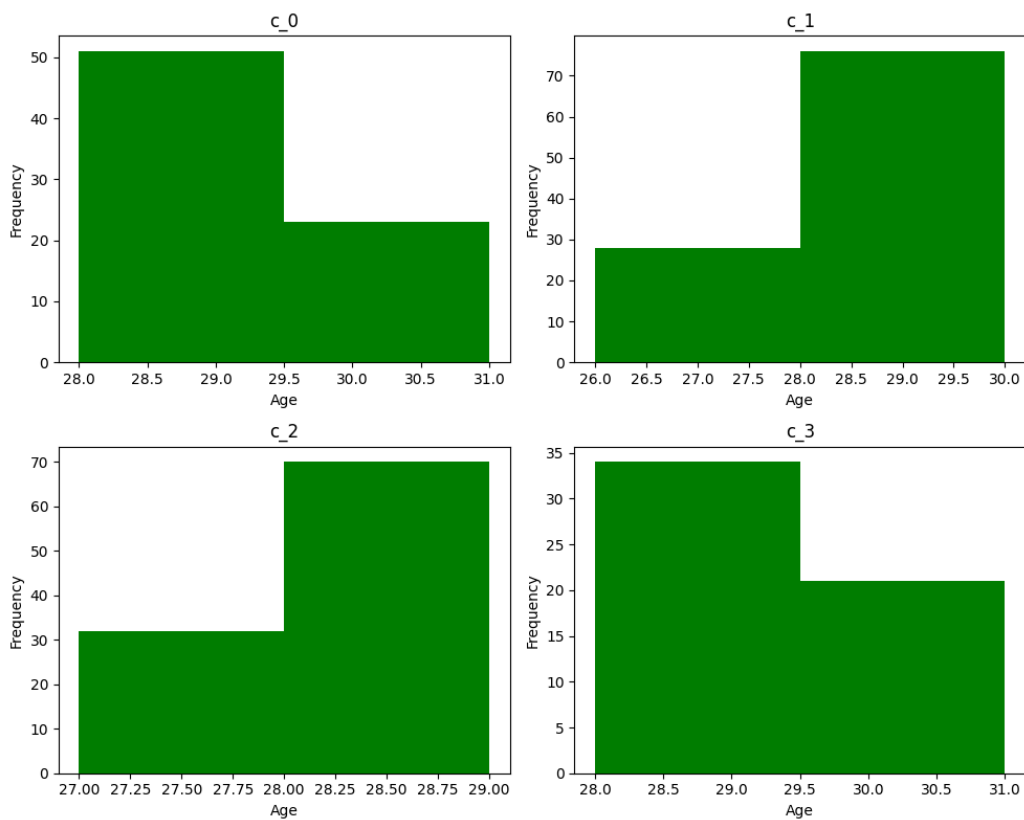

Clustering using K-Means
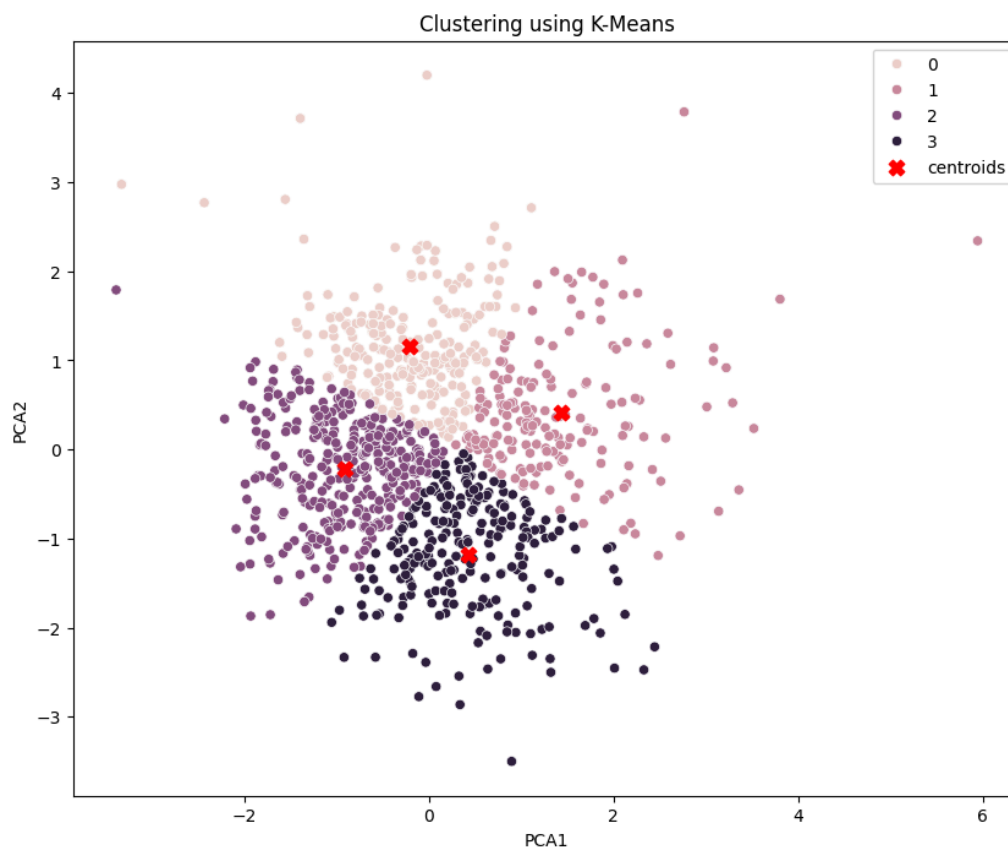
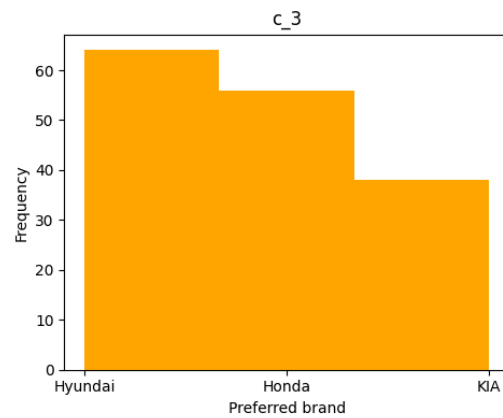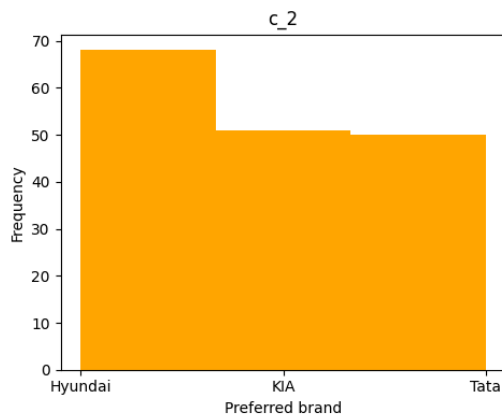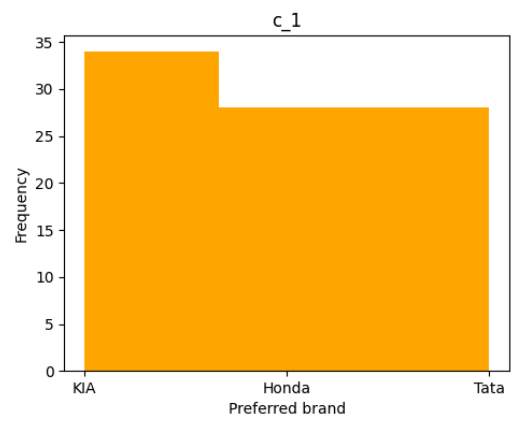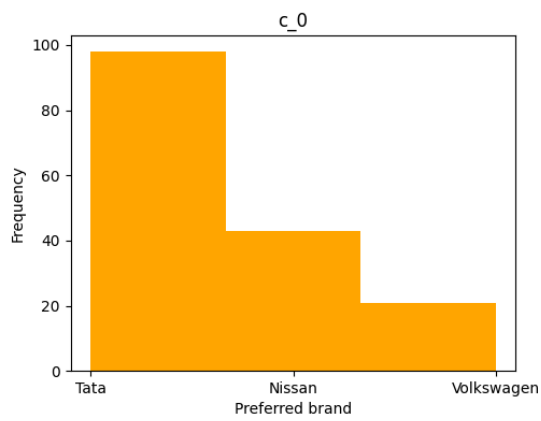I find the below target segments by performing K-Means Clustering on the dataset:
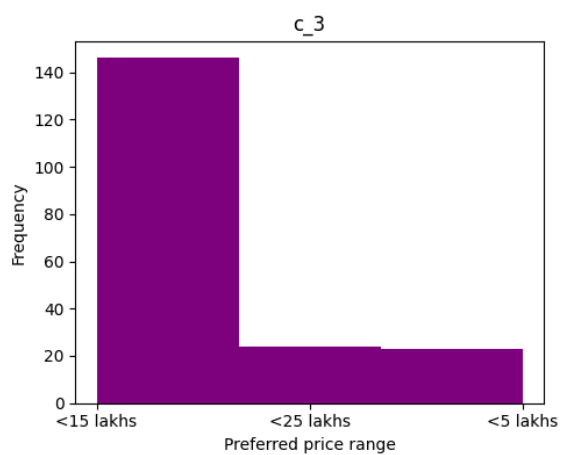
1. Average age By Preferred EV Type:

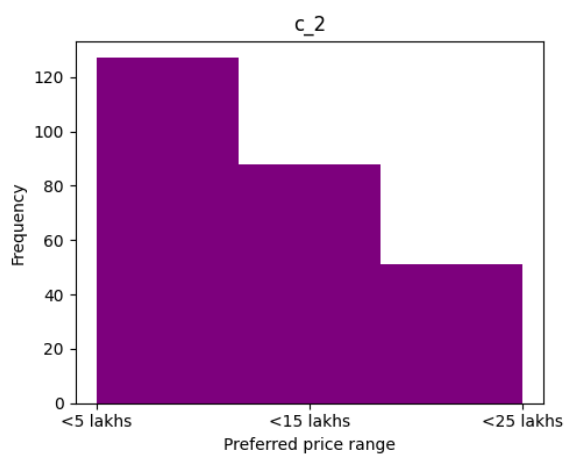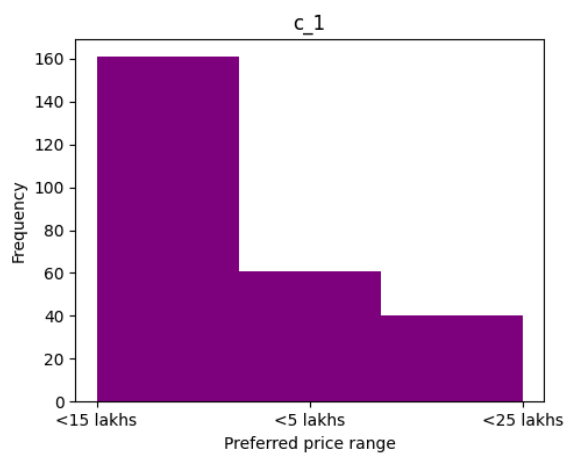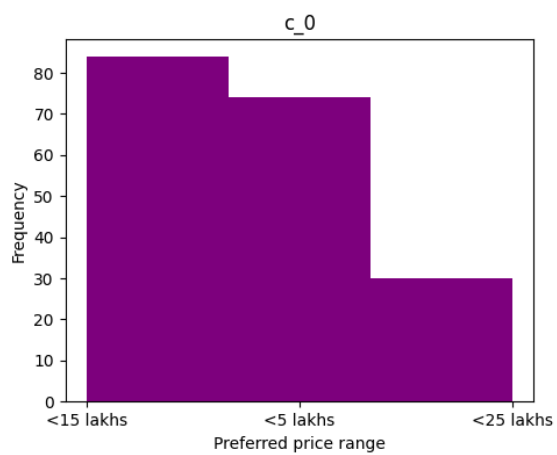Average Age by Preferred EV Brand

2) People in each age cluster/ group:

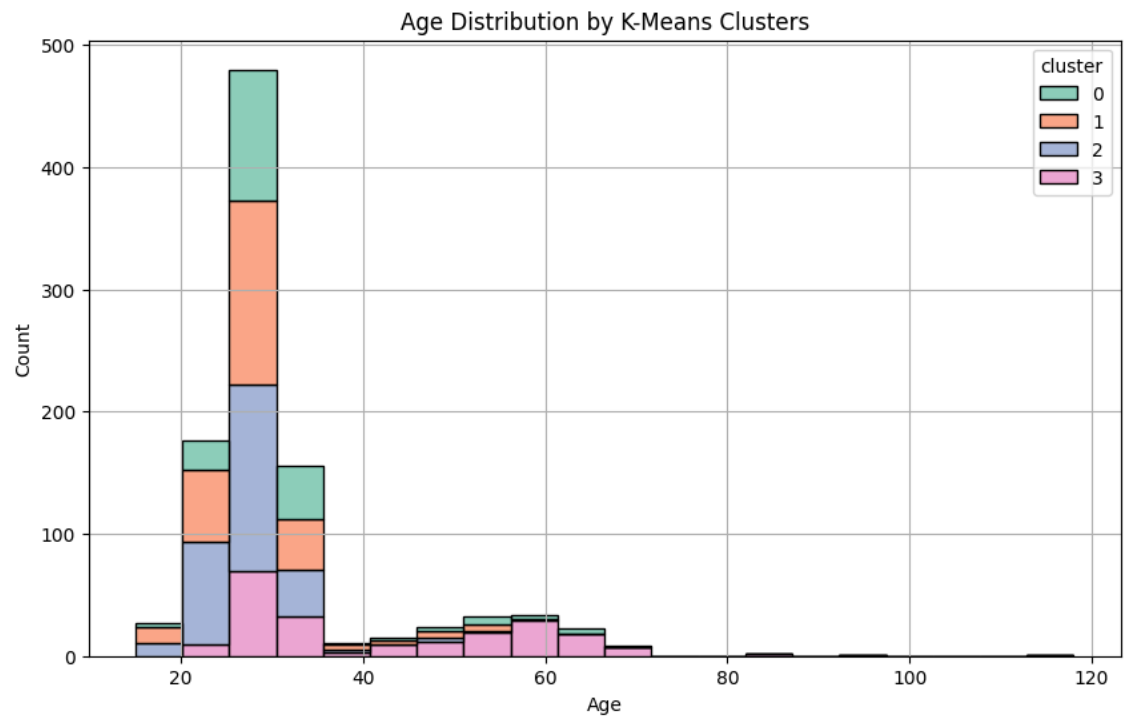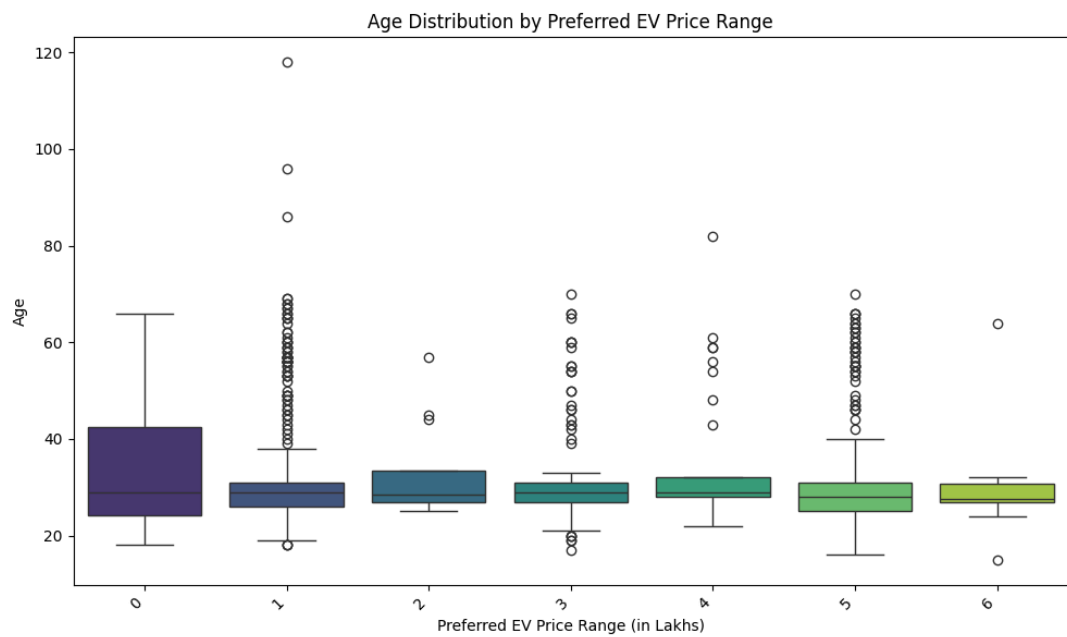## 3) Frequency. of People vs Preferred Brand:

Frequency of people with Prefered price range:-

AGE Distribution respondents

Age Distribution by K-Means Clusters

Age distribution by prefered price range:-

Age Distribution by Preferred EV Price Range

Libraries:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns


from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
import scipy.cluster.hierarchy as shc
from sklearn.metrics import silhouette_score, r2_score, mean_squared_error
from sklearn.model_selection import train_test_split, cross_val_score


import xgboost as xg
from xgboost import XGBRegressor
```

ML Technique Used:

K-Means Clustering:

XGBOOST: XGBoost is used to **predict K-Means cluster labels** based on the PCA-reduced features (PCA1, PCA2). This turns the unsupervised clustering problem into a supervised classification task.
XGBoost learns patterns in the data and can **classify new data points into clusters**, offering better decision boundaries and feature importance insights.

XGBoost Decision Boundary over Clusters
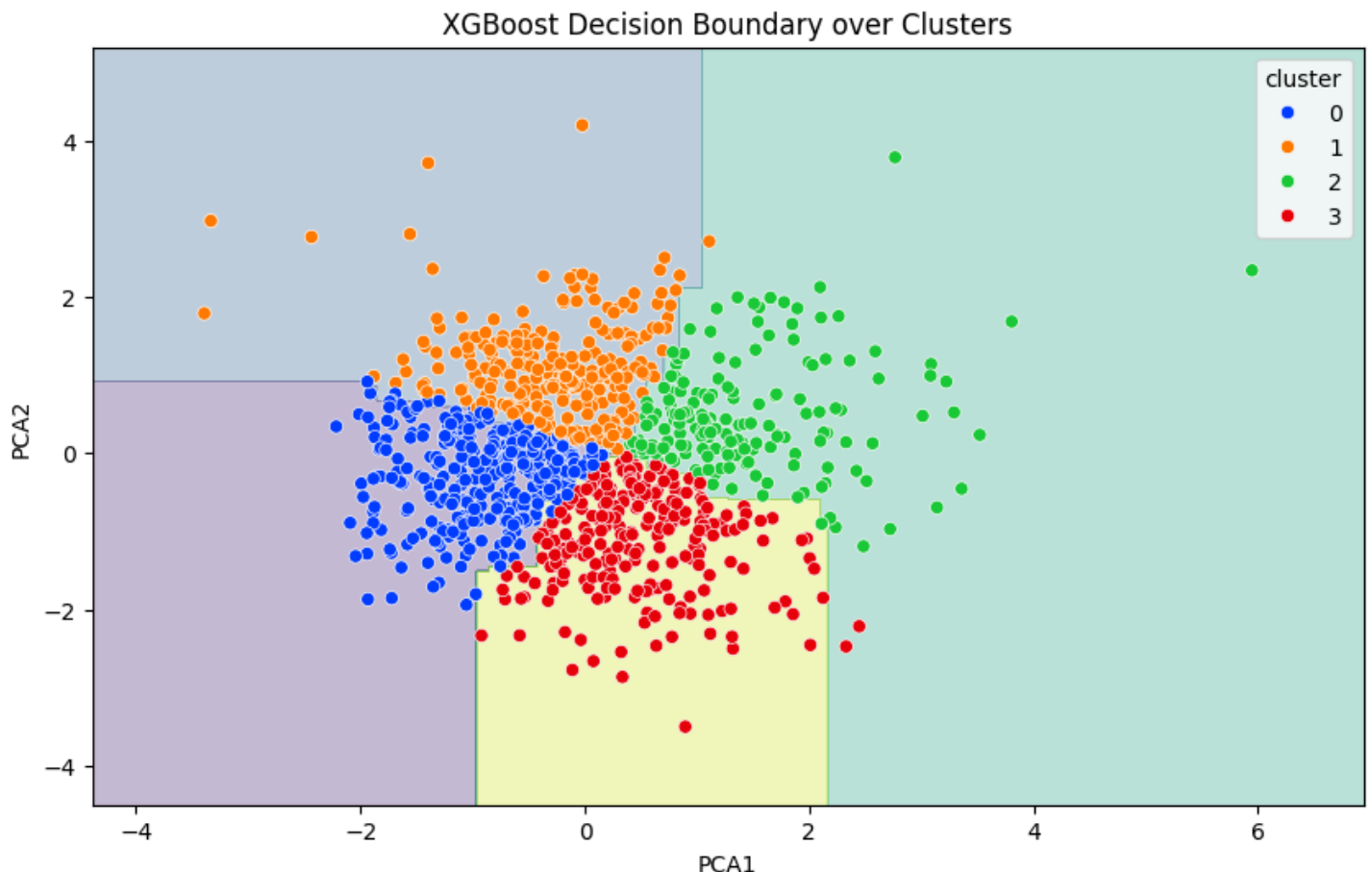
conclusion:

- Individuals aged between **25–35 years** are the most dominant in all clusters, indicating a primary target age group.
- Preferences among different age groups suggest that **younger customers (18–25)** lean toward affordability and compact models, while **older segments (35+)** prefer comfort and brand reliability.
- Marketing and product strategies should prioritize the **25–35 age group**, while also tailoring offerings for younger and older brackets based on their specific preferences.