FEYNN LABS

EV Market Segmentation

Team Prashant | Ayush | Dhruvin | Harshada | Prem |

EV Market Segmentation

-PRASHANT BHARTI

Market Segmentation:
Market segmentation is a marketing term that refers to aggregating prospective buyers into groups or segments with common needs and who respond similarly to a marketing action.

2. Types of Market Segmentation

#1. **Demographic Segmentation**: It refers to splitting up audiences based on observable, people-based differences. Demographic information is relatively easily accessible and low-cost to obtain.
(1) Ask your customers directly
(2) Use second-party and third-party data providers

#2. **Geographic Segmentation**: A customer's location can help you better understand their needs and enable you to send out location-specific ads.
(1) country, state, county, zip code
(2) climate, population density
(3) urban, suburban, or rural

#3. **Behavioral Segmentation:** It studies the behavioral traits of consumers, which include their knowledge of, attitude towards, use of, or response to a product, service, promotion, or brand.
(1) Purchasing and Usage Behavior
(2) Occasion Purchasing
(3) Benefits Sought
(4) Customer Loyalty

#3. **Psychographic Segmentation:** Personality traits include- Interests, Beliefs, Values, Attitudes, Lifestyles.

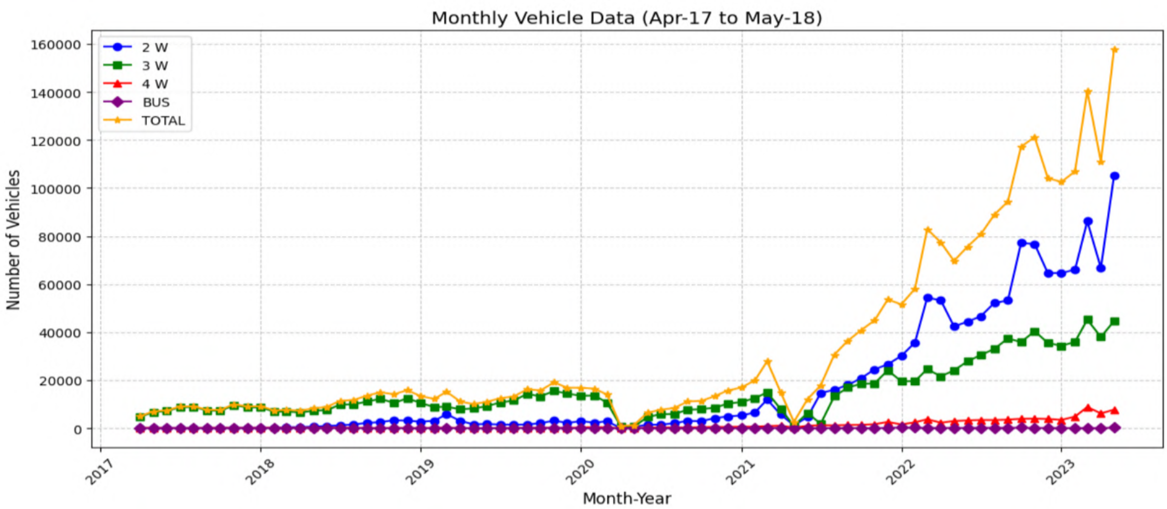
Motivation

We see steady increase in the sales of all types of Electric Vehicles be it 2-wheeler, 3-wheeler or 4-wheeler.

Code:

```python
import matplotlib.pyplot as plt
import pandas as pd

# Data

# Create DataFrame
df = pd.read_excel("sales_years.xlsx")

# Plotting
plt.figure(figsize=(12, 6))

# Plot each column with different colors
plt.plot(df['YEAR'], df['2 W'], label='2 W', color='blue', marker='o')
plt.plot(df['YEAR'], df['3 W'], label='3 W', color='green', marker='s')
plt.plot(df['YEAR'], df['4 W'], label='4 W', color='red', marker='^')
plt.plot(df['YEAR'], df['BUS'], label='BUS', color='purple', marker='D')
plt.plot(df['YEAR'], df['TOTAL'], label='TOTAL', color='orange', marker='*')

# Customize the plot
plt.title('Monthly Vehicle Data (Apr-17 to May-18)', fontsize=14)
plt.xlabel('Month-Year', fontsize=12)
plt.ylabel('Number of Vehicles', fontsize=12)
plt.xticks(rotation=45)
plt.grid(True, linestyle='--', alpha=0.7)
plt.legend()

# Adjust layout to prevent label cutoff
plt.tight_layout()

# Show plot
plt.show()
```

Plot:



Dataset:

YEAR	2 W	3 W	4 W	BUS	TOTAL
Apr-17	96	4748	198	0	5042
May-17	91	6720	215	2	7028
Jun-17	137	7178	149	1	7465

….

Jan-23	64649	34308	3490	98	102545
Feb-23	66033	35995	4850	99	106977
Mar-23	86194	45225	8852	89	140360
Apr-23	66755	38016	6193	84	111048
May-23	105154	44615	7736	283	157788

Inference:

1. Understanding EV Market Growth

1. The data shows a rapid surge in EV adoption, particularly in 2-wheelers (2W) and 3-wheelers (3W).
2. Example:
    a) April 2017: Only 96 electric 2W sold.
    b) March 2023: 86,194 electric 2W sold (~900x growth in 6 years).
3. Helps policymakers and manufacturers forecast demand and scale infrastructure.

2. Impact of Government Policies (FAME I & II)

1. The Faster Adoption and Manufacturing of Electric Vehicles (FAME) scheme boosted EV sales.
2. Post-2019, EV sales surged, especially in 2021-2023.
3. Example:
    a) March 2021: 11,963 electric 2W.
    b) March 2023: 86,194 electric 2W (7x growth in 2 years).
4. Helps assess policy effectiveness and guide future incentives.

3. Segment-Wise Adoption Patterns

1. 2-Wheelers (2W): Dominates EV sales (e.g., 105,154 in May 2023).
2. 3-Wheelers (3W): Strong in commercial use (e.g., 44,615 in May 2023).
3. 4-Wheelers (4W) & Buses: Slower but steady growth (e.g., 7,736 4W in May 2023).
4. Helps automakers prioritize R&D investments (e.g., battery tech for 2W vs. 4W).

4. COVID-19 Impact & Recovery

1. April 2020: Sales dropped sharply (only 85 electric 2W).
    a) Post-lockdown recovery: June 2020: 1,511 electric 2W (gradual rebound).
    b) 2021-2023: Exponential growth (e.g., 54,402 in March 2022).
2. Helps analyze resilience of EV demand during economic shocks.

5. Future Market Potential

1. The data suggests EVs are entering mass adoption phase (especially 2W & 3W).
2. Example: 2021-2023: 2W sales grew from 11,963 (Mar-21) → 86,194 (Mar-23).
3. Helps investors and startups identify high-growth segments.

Business & Policy Applications

1. Automakers: Optimize production for high-demand segments (e.g., affordable 2W EVs).
2. Battery & Charging Cos.: Plan infrastructure based on regional sales trends.
3. Government: Adjust FAME subsidies to accelerate 4W and bus electrification.
4. Investors: Identify fast-growing EV companies in 2W/3W space.

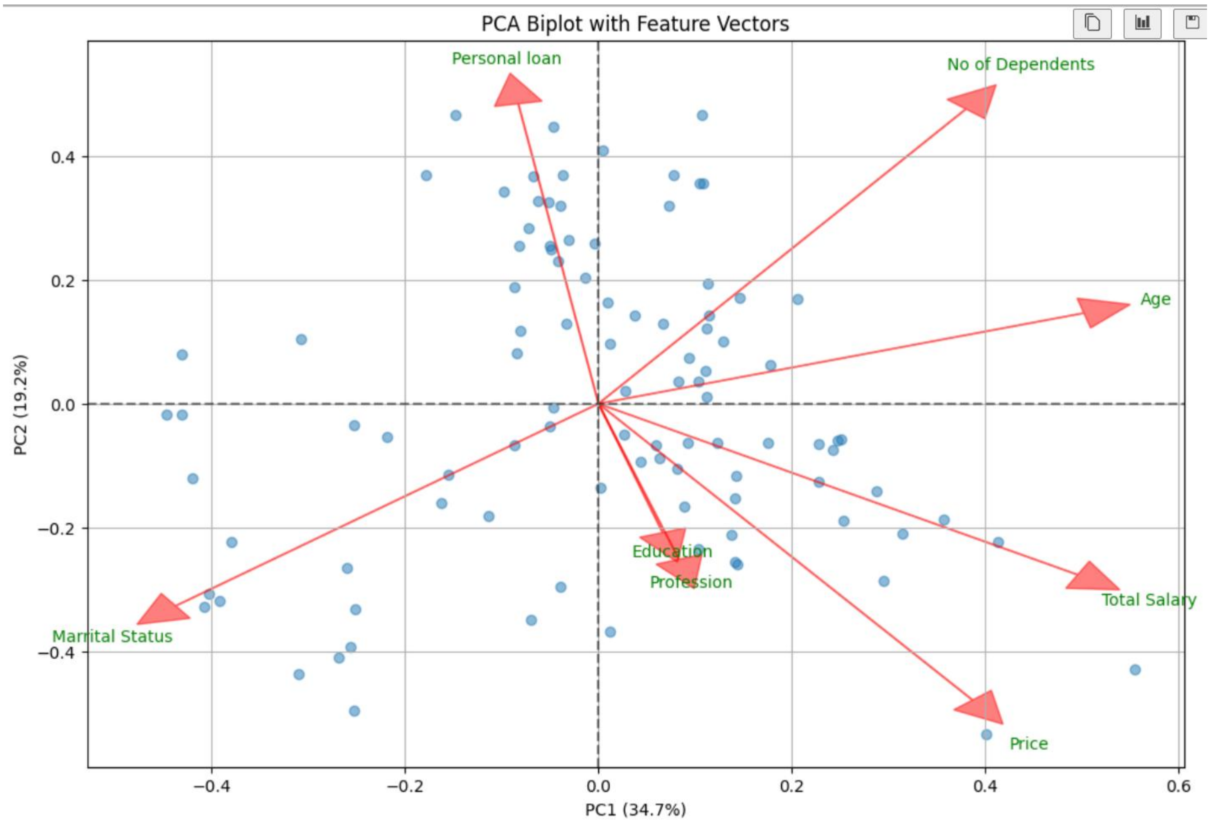
Analysis

DATASET 1. Demographic Dataset

Age	Profession	Marrital St	Education	No of Dep	Personal loan	Total Salary	Price
27	Salaried	Single	Post Graduate	0	Yes	800000	800000
35	Salaried	Married	Post Graduate	2	Yes	2000000	1000000
45	Business	Married	Graduate	4	Yes	1800000	1200000
41	Business	Married	Post Graduate	3	No	2200000	1200000
31	Salaried	Married	Post Graduate	2	Yes	2600000	1600000
28	Salaried	Married	Graduate	3	Yes	900000	700000
31	Salaried	Married	Graduate	4	No	1800000	1200000
33	Business	Married	Post Graduate	4	No	1400000	700000
34	Business	Married	Post Graduate	4	No	2000000	1100000
34	Salaried	Married	Graduate	3	Yes	1900000	800000
35	Salaried	Married	Post Graduate	4	No	2000000	1600000
35	Salaried	Married	Graduate	4	Yes	1400000	700000
29	Salaried	Married	Post Graduate	0	No	1700000	110000
30	Business	Single	Post Graduate	2	Yes	1400000	800000

i. Loading the data set and Pre processing:

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt

# Load the data
df = pd.read_csv('age_buynotbuy.csv')

# Display original data
print("Original Data:")
print(df.head())
print("\nData Types:")
print(df.dtypes)

# Convert categorical columns to numerical values
label_encoders = {}
for column in df.select_dtypes(include=['object']).columns:
    le = LabelEncoder()
    df[column] = le.fit_transform(df[column])
    label_encoders[column] = le

# Display encoded data
print("\nEncoded Data:")
print(df.head())
```



<u>Description</u>:
The dataset includes information about individuals, including their age, profession, marital status, education level, number of dependents, whether they have a personal loan, total salary, and price (possibly of a product or service).

<u>Insights</u>:
1. Education and Profession Influence Salary: People with higher education levels (like Post Graduate degrees) and those in professions tend to have higher salaries.
2. Marital Status and Dependents: Marital status and the number of dependents also play a role in financial decisions, such as taking personal loans.

<u>Implications for the EV Company</u>: For an Electric Vehicle (EV) company looking to market and sell their products, here are potential benefits from this dataset:

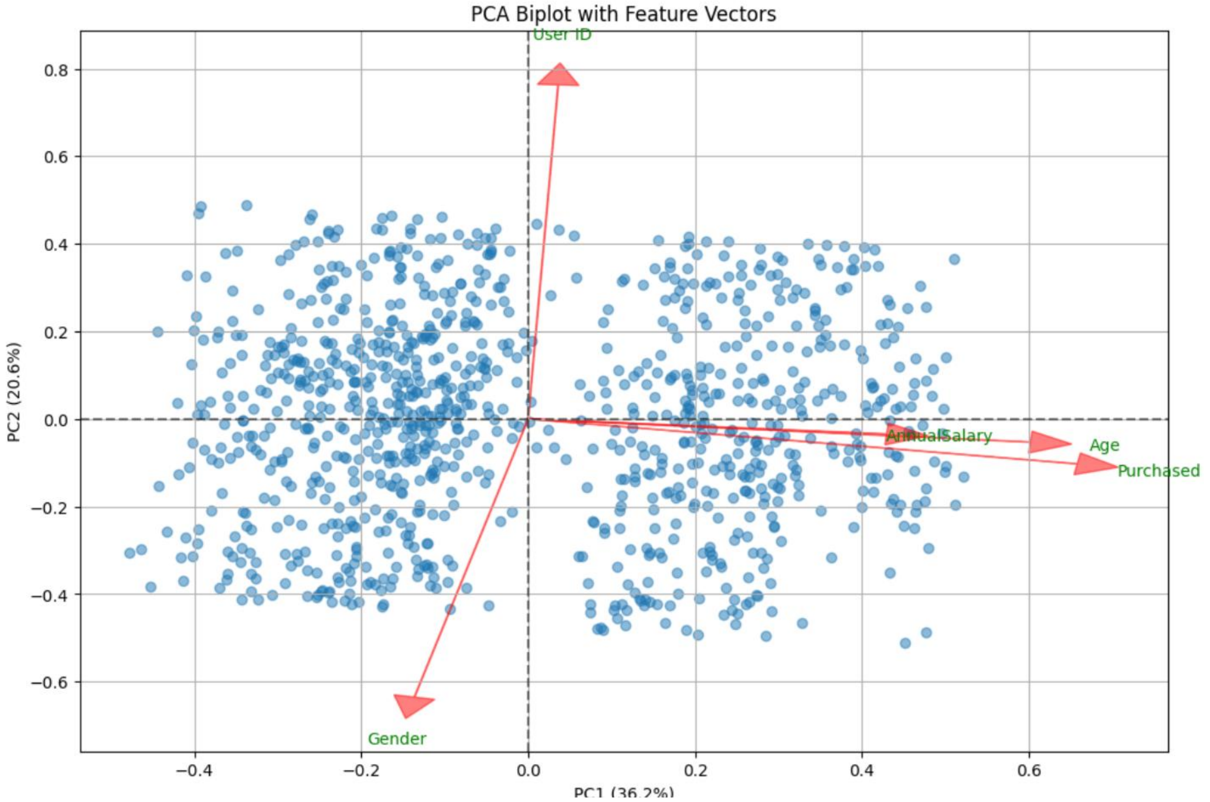
1. **Targeting High-Earning Groups**: Focus marketing efforts on individuals with higher salaries, especially those in professions and with advanced education levels. These groups are more likely to afford higher-priced products like electric vehicles.
2. **Understanding Financial Commitments**: Individuals who have taken personal loans might be more cautious about additional financial commitments. Tailor financing options or incentives to address these concerns.
3. **Segmentation Strategy**: Use demographic insights to segment the market effectively. For instance, target married individuals with higher incomes and educate them about the long-term financial benefits of EV ownership, such as lower operational costs and environmental impact.
4. **Education and Awareness Campaigns**: Leverage the correlation between education and income to design educational campaigns about the benefits of EVs, targeting educated professionals who are likely to appreciate the technological advancements and environmental benefits of EVs.

DATASET 2. Behavioural dataset:

User ID	Gender	Age	TotalSalary	Purchased
385	Male	35	20000	0
681	Male	40	43500	0
353	Male	49	74000	0
895	Male	40	107500	1
661	Male	25	79000	0
846	Female	47	33500	1
219	Female	46	132500	1

……………

PCA Analysis:



Inference: People who have "More Salary" and have "More Age" have more tendency to "**purchase**".

Dataset Description: Helps to analyze how Gender, Age, and Annual Salary influence the purchasing behavior of users.

Insight:
1. Older, more financially stable customers are more inclined to consider high-investment products like EVs.
2. EVs are premium products, and hence attract higher-income segments.

Implications for the EV Company:

1. **Target Demographics**: Focus on **urban professionals aged 40+ with income > 90K**. These are your prime EV buyers.

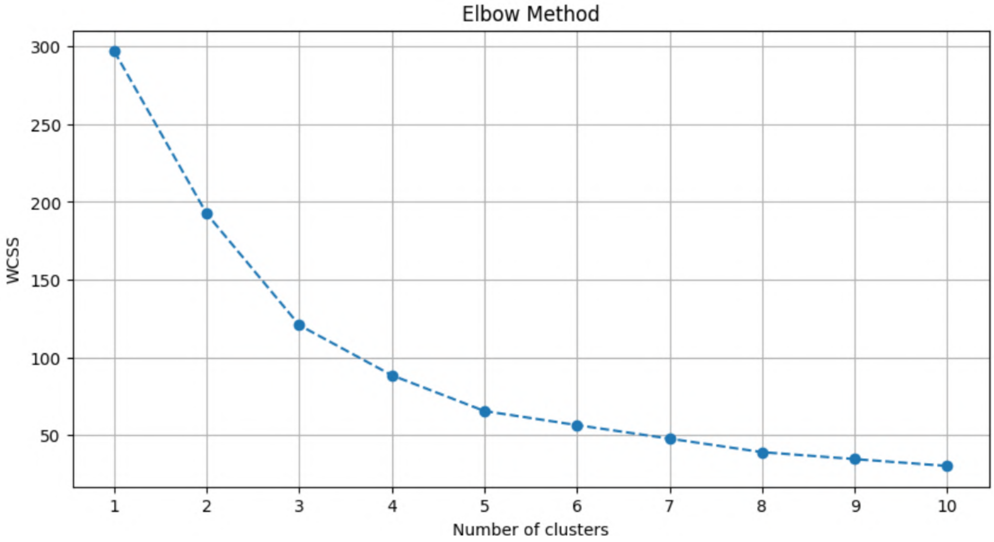
2. **Customer Segmentation**: Use Age and Income to segment and prioritize leads:
   -High-income, older buyers → Immediate targets
   -Young, mid-income users → Potential for **future targeting** via EMI schemes or smaller EV models
3. **Marketing Strategy**:
   -Highlight **eco-conscious luxury** and **long-term savings** to appeal to high-income groups.
   -Tailor campaigns for **working women** as a rising buyer segment.
4. **Product Bundling & Offers**:
   -Provide flexible payment plans for mid-income groups aged 30–40.
   -Offer **test-drive experiences** and **incentives** (e.g., charging vouchers) to convert hesitant mid-range customers.
5. **Expansion Strategy**:
   -Launch campaigns in metros and tier-1 cities where **average income is higher**.
   -Partner with corporates for **fleet EV adoption** targeting high-salary employees.

Segmentation and visualization:

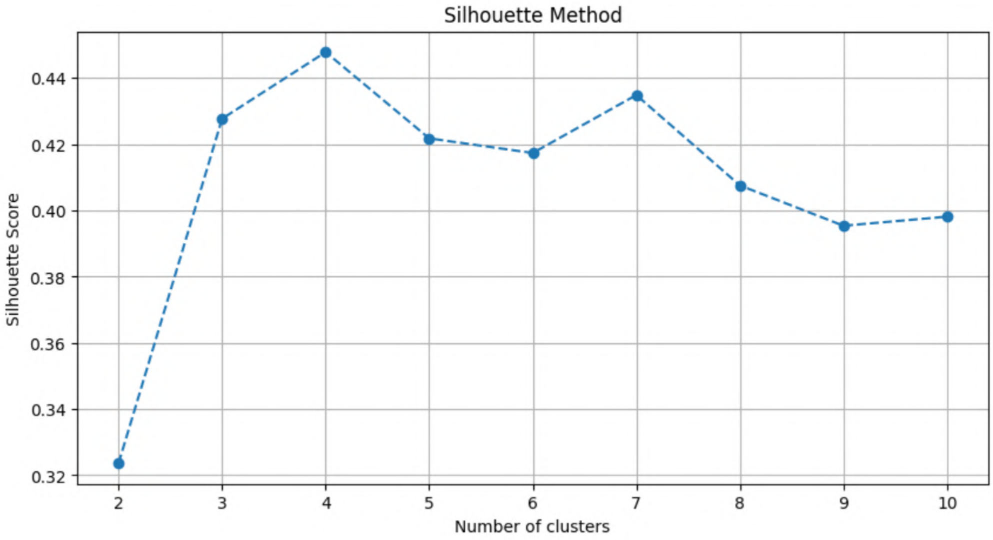
1. Segmentation for demographic dataset:

a) Clustering among: **Age, Personal Loan** and **Salary**

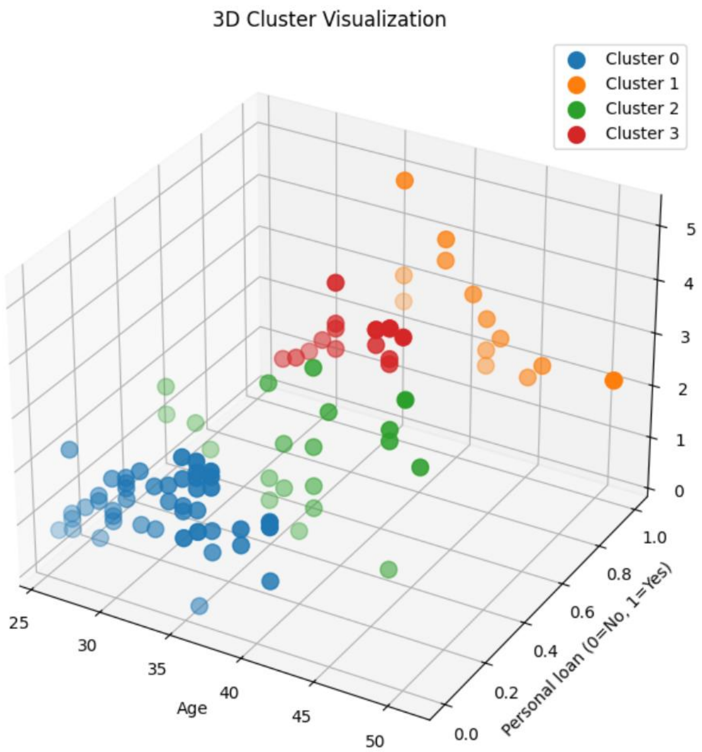
```
# Plot Elbow Method
plt.figure(figsize=(10, 5))
plt.plot(cluster_range, wcss, marker='o', linestyle='--')
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.xticks(cluster_range)
plt.grid()
plt.show()
```




```
# Silhouette Method to find optimal number of clusters
silhouette_scores = []
for k in range(2, 11):
    kmeans = KMeans(n_clusters=k, init='k-means++', random_state=42)
    kmeans.fit(scaled_data)
    score = silhouette_score(scaled_data, kmeans.labels_)
    silhouette_scores.append(score)

# Plot Silhouette Method
plt.figure(figsize=(10, 5))
plt.plot(range(2, 11), silhouette_scores, marker='o', linestyle='--')
plt.title('Silhouette Method')
plt.xlabel('Number of clusters')
plt.ylabel('Silhouette Score')
plt.xticks(range(2, 11))
plt.grid()
plt.show()
```




```python
# Based on the plots, choose optimal number of clusters
optimal_clusters = 4

# Perform K-Means clustering
kmeans = KMeans(n_clusters=optimal_clusters, init='k-means++', random_state=42)
clusters = kmeans.fit_predict(scaled_data)

# Add cluster labels to original dataframe
df['Cluster'] = clusters

# 3D Visualization
fig = plt.figure(figsize=(12, 8))
ax = fig.add_subplot(111, projection='3d')

# Plot each cluster
for cluster in range(optimal_clusters):
    cluster_data = df[df['Cluster'] == cluster]
    ax.scatter(cluster_data['Age'],
               cluster_data['Personal loan'],
               cluster_data['Total Salary'],
               label=f'Cluster {cluster}',
               s=100)

ax.set_title('3D Cluster Visualization')
ax.set_xlabel('Age')
ax.set_ylabel('Personal loan (0=No, 1=Yes)')
ax.set_zlabel('Total Salary')
ax.legend()
plt.show()

# Print cluster centers in original scale
cluster_centers = scaler.inverse_transform(kmeans.cluster_centers_)
centers_df = pd.DataFrame(cluster_centers, columns=['Age', 'Personal loan', 'Total Salary'])
print("Cluster Centers:")
print(centers_df)
```



<u>Inference</u>:

Cluster 0 (Blue): People with **Less age** and **No Loan** tend to have **Lesser Salary.**

Cluster 1 (Green): People with **Average age** and **No Loan** tend to have **Average Salary.**

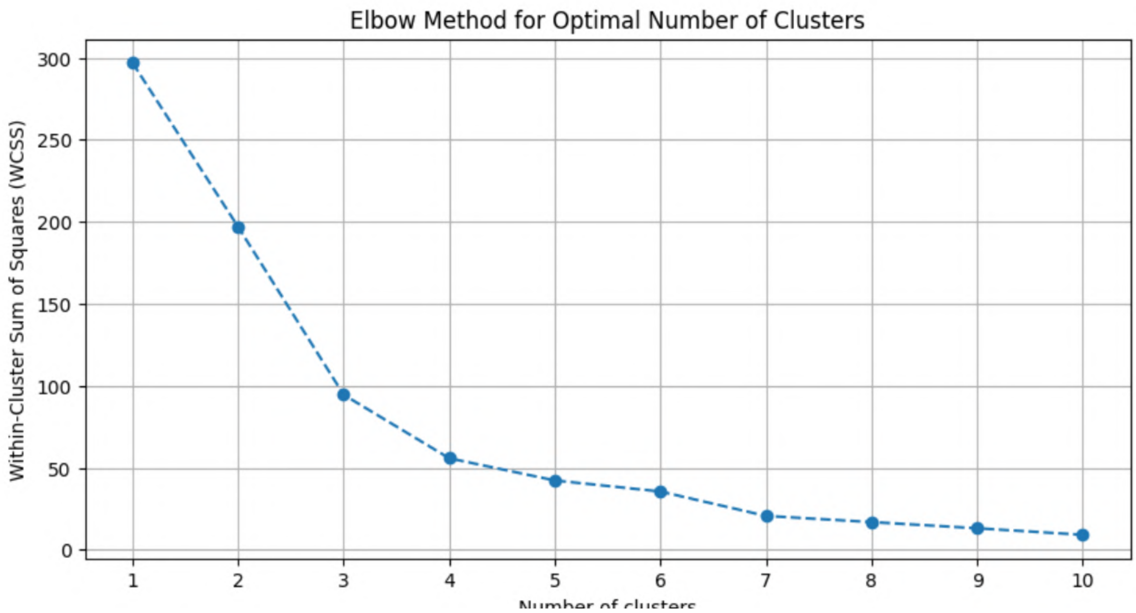
Cluster 2 (Red): People with **Less age** and **Loan** tend to have **Lesser Salary.**

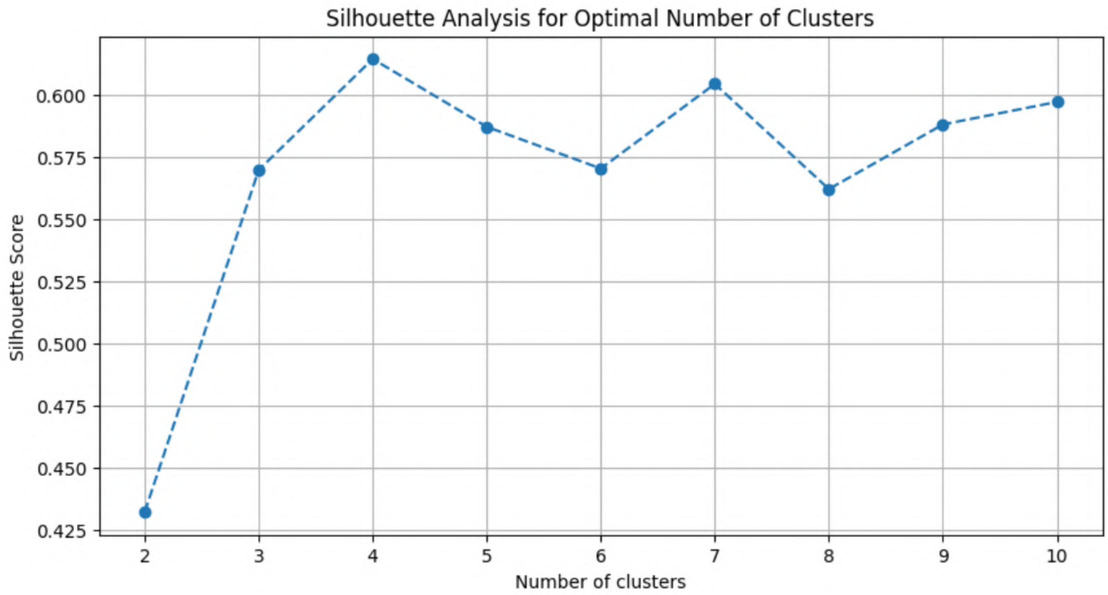
Cluster 3 (Orange): People with **Higher age** and **Loans** tend to have **Higher Salary.**

Preferred Clusters: Cluster 1 and Cluster 3.

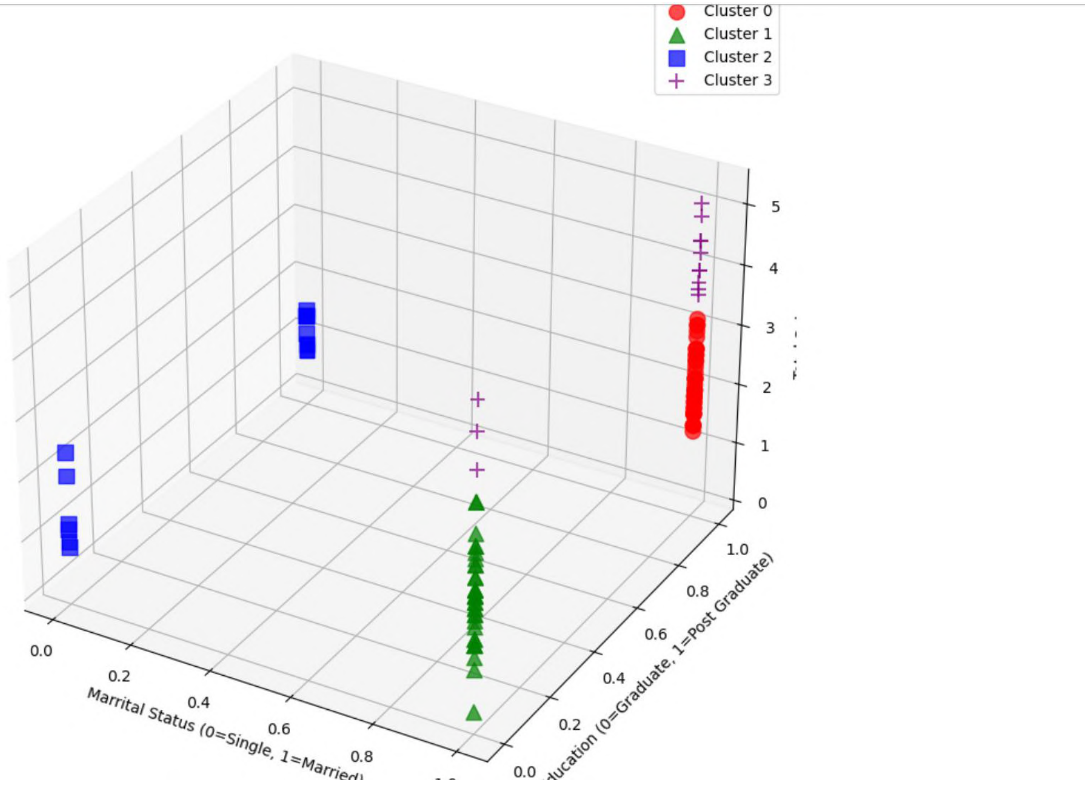
b) Clustering among: **Marital Status, Education** and **Total Salary**

```python
# Plot Elbow Method
plt.figure(figsize=(10, 5))
plt.plot(cluster_range, wcss, marker='o', linestyle='--')
plt.title('Elbow Method for Optimal Number of Clusters')
plt.xlabel('Number of clusters')
plt.ylabel('Within-Cluster Sum of Squares (WCSS)')
plt.xticks(cluster_range)
plt.grid()
plt.show()

# Silhouette Method to find optimal number of clusters
silhouette_scores = []
for k in range(2, 11):
    kmeans = KMeans(n_clusters=k, init='k-means++', random_state=42)
    kmeans.fit(scaled_data)
    score = silhouette_score(scaled_data, kmeans.labels_)
    silhouette_scores.append(score)

# Plot Silhouette Method
plt.figure(figsize=(10, 5))
plt.plot(range(2, 11), silhouette_scores, marker='o', linestyle='--')
plt.title('Silhouette Analysis for Optimal Number of Clusters')
plt.xlabel('Number of clusters')
plt.ylabel('Silhouette Score')
plt.xticks(range(2, 11))
plt.grid()
plt.show()
```






```python
# Based on the plots, choose optimal number of clusters
optimal_clusters = 4

# Perform K-Means clustering
kmeans = KMeans(n_clusters=optimal_clusters, init='k-means++', random_state=42)
clusters = kmeans.fit_predict(scaled_data)

# Add cluster labels to original dataframe
df['Cluster'] = clusters

# 3D Visualization
fig = plt.figure(figsize=(12, 8))
ax = fig.add_subplot(111, projection='3d')

# Define colors and markers for each cluster
colors = ['r', 'g', 'b','purple']
markers = ['o', '^', 's','+']

# Plot each cluster
for cluster in range(optimal_clusters):
    cluster_data = df[df['Cluster'] == cluster]
    ax.scatter(cluster_data['Marrital Status'],
               cluster_data['Education'],
               cluster_data['Total Salary'],
               c=colors[cluster],
               marker=markers[cluster],
               label=f'Cluster {cluster}',
               s=100,
               alpha=0.7)

ax.set_title('3D Cluster Visualization')
ax.set_xlabel('Marrital Status (0=Single, 1=Married)')
ax.set_ylabel('Education (0=Graduate, 1=Post Graduate)')
ax.set_zlabel('Total Salary')
ax.legend()
plt.tight_layout()
plt.show()
```



Inference:

Cluster 0 (Red): People **Not Married** and **Educated** tend to have **Lesser Salary.**

Cluster 1 (Green): People **Married** and having **No Education** tend to have **Average Salary.**

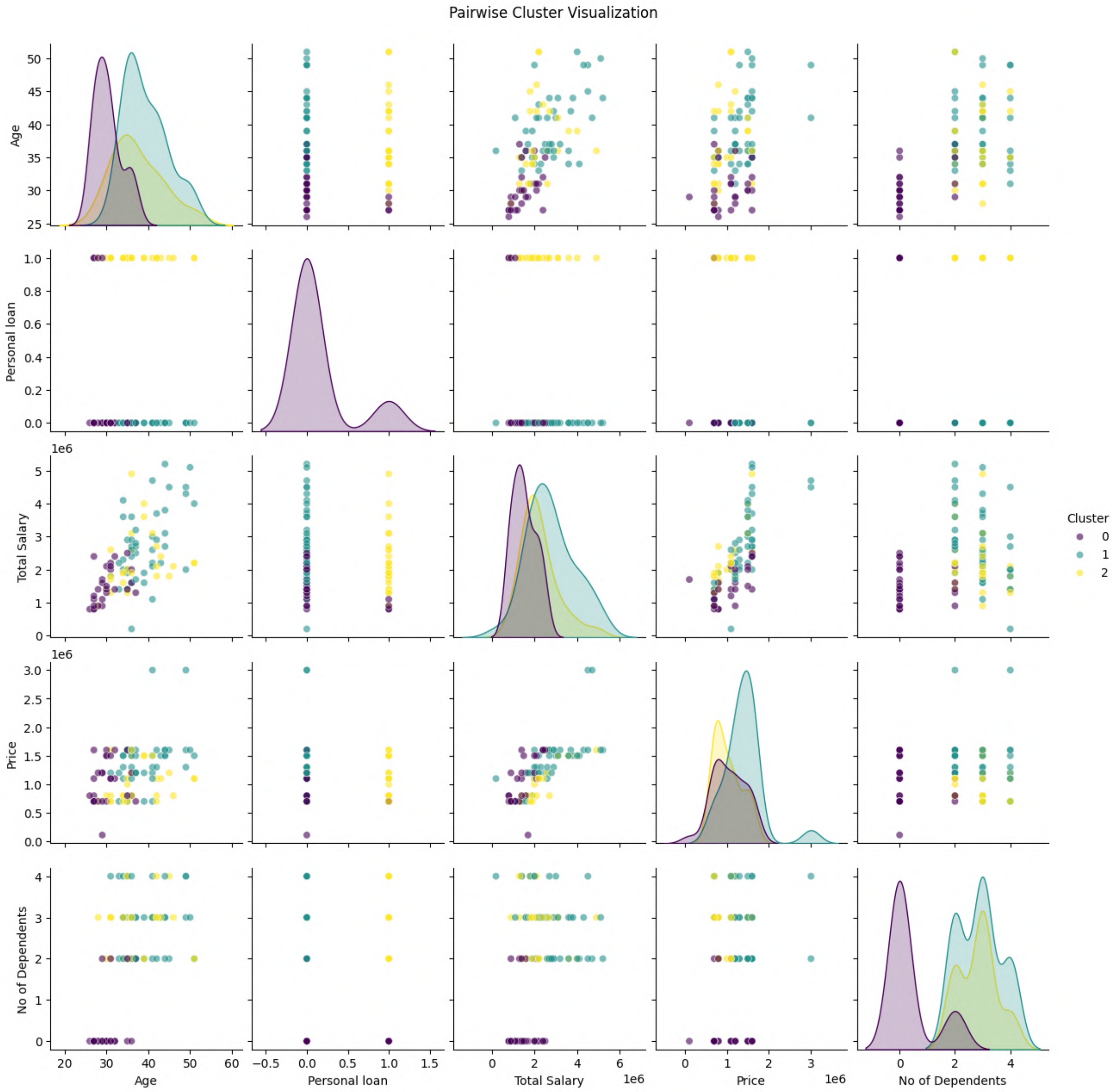
Cluster 2 (Blue): People **Not Married** and having **Education** tend to have **Average Salary.**

Cluster 3 (Plus): People **Married** and **Educated** tend to have **Higher Salary.**

Preferred Clusters: Cluster 1 and Cluster 3.

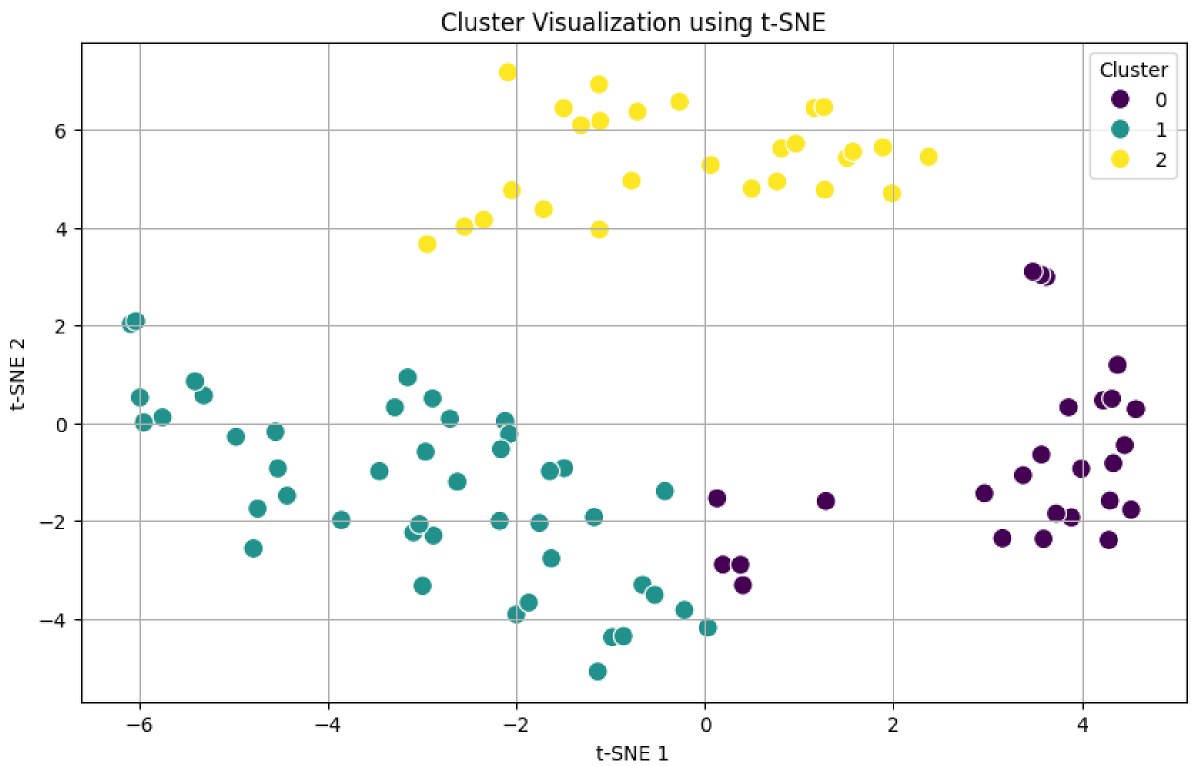
c) Clustering among multiple variables: **<u>Age', 'Personal loan', 'Total Salary', 'Price'</u>** and **<u>'No of Dependents':</u>**

```
# Preprocessing
df['Personal loan'] = df['Personal loan'].map({'Yes': 1, 'No': 0})
df['Marrital Status'] = df['Marrital Status'].map({'Single': 0, 'Married': 1})
df['Profession'] = df['Profession'].map({'Salaried': 0, 'Business': 1})
df['Education'] = df['Education'].map({'Graduate': 0, 'Post Graduate': 1})

# Select all numerical columns (or specific ones you want)
num_cols = ['Age', 'Personal loan', 'Total Salary', 'Price', 'No of Dependents']
X = df[num_cols]

# Scale data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# K-Means clustering
kmeans = KMeans(n_clusters=4, random_state=42)
clusters = kmeans.fit_predict(X_scaled)
df['Cluster'] = clusters

# Pairplot visualization
sns.pairplot(df, vars=num_cols, hue='Cluster', palette='viridis', plot_kws={'alpha': 0.6})
plt.suptitle('Pairwise Cluster Visualization', y=1.02)
plt.show()
```



d) **t-SNE** Visualisation for multidimensional data:

```
# t-SNE visualization
tsne = TSNE(n_components=2, random_state=42)
X_tsne = tsne.fit_transform(X_scaled)

plt.figure(figsize=(10, 6))
sns.scatterplot(x=X_tsne[:, 0], y=X_tsne[:, 1], hue=clusters, palette='viridis', s=100)
plt.title('Cluster Visualization using t-SNE')
plt.xlabel('t-SNE 1')
plt.ylabel('t-SNE 2')
plt.legend(title='Cluster')
plt.grid()
plt.show()
```



Cluster Visualization using t-SNE
Cluster
0
1
2
t-SNE 1
t-SNE 2
−6
−4
−2
0
2
4
6

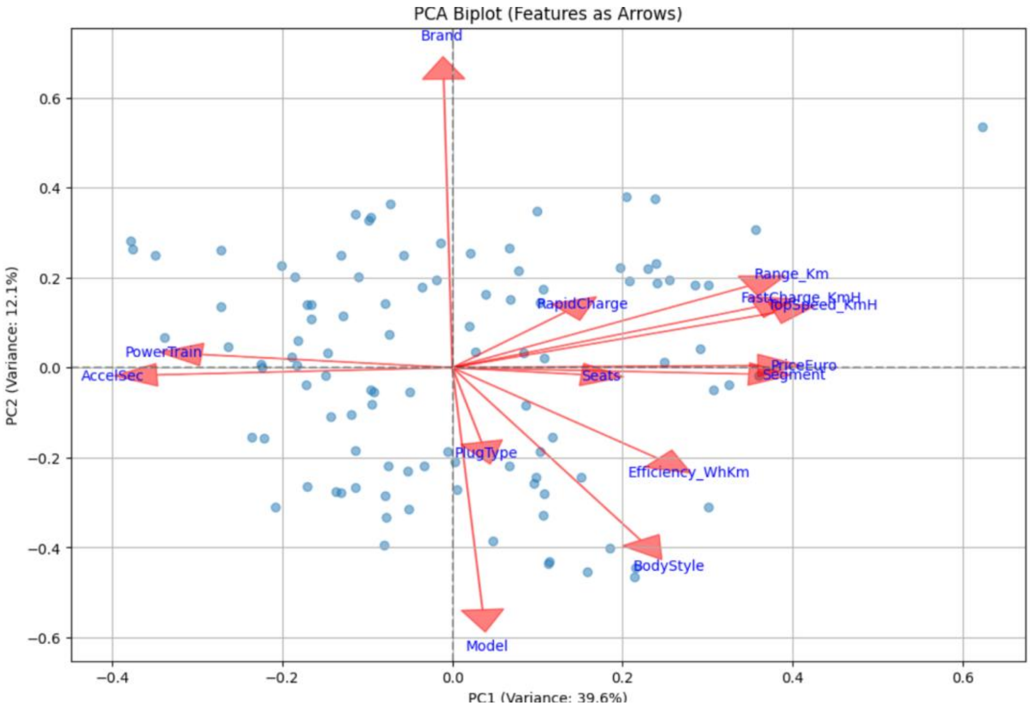
ANALYSIS

DATASET 3: Vehicle Info DATASET(Parametric Dataset) (speedefficiencyPrice.csv)

```python
# 1. Load the Excel file
file_path = 'speedefficiencyPrice.csv'  # Replace with your file path
df = pd.read_csv(file_path)

# 2. Convert categorical columns to numerical (if any)
label_encoders = {}
for column in df.select_dtypes(include=['object']).columns:
    le = LabelEncoder()
    df[column] = le.fit_transform(df[column])
    label_encoders[column] = le

# 3. Standardize the data (important for PCA)
scaler = StandardScaler()
scaled_data = scaler.fit_transform(df)
```

Brand	Model	AccelSec	TopSpeed_	Range_Km	Efficiency_	FastCharg	RapidChar	PowerTrai	PlugType	BodyStyle	Segment	Seats	PriceEuro
Tesla	Model 3 Lo	4.6	233	450	161	940	Yes	AWD	Type 2 CCS	Sedan	D	5	55480
Volkswage	ID.3 Pure	10	160	270	167	250	No	RWD	Type 2 CCS	Hatchback	C	5	30000
Polestar	2	4.7	210	400	181	620	Yes	AWD	Type 2 CCS	Liftback	D	5	56440
BMW	iX3	6.8	180	360	206	560	Yes	RWD	Type 2 CCS	SUV	D	5	68040
Honda	e	9.5	145	170	168	190	Yes	RWD	Type 2 CCS	Hatchback	B	4	32997
Lucid	Air	2.8	250	610	180	620	Yes	AWD	Type 2 CCS	Sedan	F	5	105000
Volkswage	e-Golf	9.6	150	190	168	220	No	FWD	Type 2 CCS	Hatchback	C	5	31900
Peugeot	e-208	8.1	150	275	164	420	No	FWD	Type 2 CCS	Hatchback	B	5	29682
Tesla	Model 3 St	5.6	225	310	153	650	Yes	RWD	Type 2 CCS	Sedan	D	5	46380
Audi	Q4 e-tron	6.3	180	400	193	540	Yes	AWD	Type 2 CCS	SUV	D	5	55000
Mercedes	EQC 400 4	5.1	180	370	216	440	Yes	AWD	Type 2 CCS	SUV	D	5	69484
Nissan	Leaf	7.9	144	220	164	230	Yes	FWD	Type 2 CHA	Hatchback	C	5	29234
Hyundai	Kona Elect	7.9	167	400	160	380	Yes	FWD	Type 2 CCS	SUV	B	5	40795
BMW	i4	4	200	450	178	650	Yes	RWD	Type 2 CCS	Sedan	D	5	65000

```python
# 4. Perform PCA
pca = PCA()
pca_result = pca.fit_transform(scaled_data)

# 5. Define a function to create a biplot (scatter + arrows)
def biplot(score, coeff, feature_names):
    plt.figure(figsize=(12, 8))

    # Scatter plot of PCA results
    xs = score[:, 0]
    ys = score[:, 1]
    scalex = 1.0/(xs.max() - xs.min())
    scaley = 1.0/(ys.max() - ys.min())
    plt.scatter(xs*scalex, ys*scaley, alpha=0.5)

    # Add arrows for each feature
    for i, feature in enumerate(feature_names):
        plt.arrow(0, 0, coeff[i, 0], coeff[i, 1], color='r', alpha=0.5,
                  head_width=0.05, head_length=0.05)
        plt.text(coeff[i, 0] * 1.15, coeff[i, 1] * 1.15, feature,
                 color='b', ha='center', va='center', fontsize=10)

    # Add labels and grid
    plt.xlabel(f"PC1 (Variance: {pca.explained_variance_ratio_[0] * 100:.1f}%)")
    plt.ylabel(f"PC2 (Variance: {pca.explained_variance_ratio_[1] * 100:.1f}%)")
    plt.title("PCA Biplot (Features as Arrows)")
    plt.grid(True)
    plt.axhline(0, color='k', linestyle='--', alpha=0.3)
    plt.axvline(0, color='k', linestyle='--', alpha=0.3)
    plt.show()
```



Inference:

1. Attribute **Brand** is independent of most of the important attributes like **Price**, **Segment** and **Charging Rate**.
   This means Each Brand has almost all classes of vehicle and expands itself in every other use case.
2. More the rate of charging, that is **Rapid Charge** and **Fast Charge, More Expensive** the EV is.

DATASET 4: Vehicle Fuel Type vs Sales DATASET(Parametric Dataset)

(vehicle_fuel2.csv)

Vehicle Category	CNG ONLY	DIESEL	DIESEL/HY	DI-METHY	DUAL DIES	DUAL DIES	DUAL DIES	ELECTRIC(I	ETHANOL	FUEL CELL	LNG	LPG ONLY
HEAVY GOODS VEH	46,939	58,55,610	92	0	0	40	52	809	1	1	749	51
HEAVY MOTOR VEH	17	96,792	115	0	0	0	0	154	0	0	0	33
HEAVY PASSENGER	49,672	7,55,697	49	0	0	13	0	9,498	25	16	8	4
LIGHT GOODS VEHI	4,37,612	90,66,552	221	0	2	11	0	16,437	13	1	0	6,043
LIGHT MOTOR VEH	19,562	2,79,14,500	1,67,676	1	0	6	0	2,12,288	59	2	7	13,639
LIGHT PASSENGER	52,822	30,70,265	4,657	0	0	4	0	25,348	2	2	1	1,185
MEDIUM GOODS V	50,446	8,16,824	19	0	0	6	0	44	1	0	2	11
MEDIUM MOTOR V	34	1,61,865	20	0	0	1	0	61	1	0	0	3
MEDIUM PASSENG	18,161	3,21,928	9	0	0	5	0	835	0	1	4	2
THREE WHEELER (I	0	7	0	0	0	0	0	1	0	0	0	0
THREE WHEELER(N	6,822	5,30,699	0	0	0	0	0	1,674	0	0	0	630
THREE WHEELER(T	12,44,494	43,65,633	25	0	0	0	0	23,75,611	4	0	3	1,63,791
TWO WHEELER (IN	0	140	0	0	0	0	0	141	0	0	0	0
TWO WHEELER(NT)	54,577	1,39,157	45	0	1	0	1	25,47,087	816	11	4	1,001
TWO WHEELER(T)	46	98	0	0	0	0	0	14,297	1	0	0	1

Vehicle Category	METHANC	NOT APPLI	PETROL	PETROL/CI	PETROL/ET	PETROL/HY	PETROL/LF	PETROL/M	PLUG-IN H	PURE EV	SOLAR	STRONG H	Total
HEAVY GOODS VEH	0	60,547	13,562	263	0	1	320	0	0	72	56	0	59,79,165
HEAVY MOTOR VEH	0	4,695	748	0	0	1	23	0	0	0	4	0	1,02,582
HEAVY PASSENGER	0	9,665	3,948	136	0	0	68	0	0	1,137	11	0	8,29,947
LIGHT GOODS VEHI	4	6,22,585	3,51,373	51,671	36	2	8,536	0	0	937	93	0	1,05,62,129
LIGHT MOTOR VEH	7	12,80,609	3,18,15,614	42,46,223	52,276	11,73,650	12,08,102	1	74	49,714	268	86,849	6,82,41,127
LIGHT PASSENGER	0	37,237	4,85,966	7,24,336	265	28,395	53,274	0	0	4,618	63	2,224	44,90,664
MEDIUM GOODS V	0	23,701	3,833	106	0	0	31	0	0	0	25	0	8,95,049
MEDIUM MOTOR V	0	37,604	1,045	7	0	10	43	0	0	0	4	0	2,00,698
MEDIUM PASSENG	0	1,809	943	78	0	0	27	0	0	39	4	0	3,43,845
THREE WHEELER (I	0	0	13	1	0	0	0	0	0	0	0	0	22
THREE WHEELER(N	0	3,729	1,05,614	8,667	0	0	11,872	0	0	31	5	0	6,69,743
THREE WHEELER(T	0	52,586	12,64,524	12,66,458	1	8	7,19,838	0	0	86,689	92	0	1,15,39,757
TWO WHEELER (IN	0	58	1,26,422	8	55	0	33	0	0	14	0	0	1,26,871
TWO WHEELER(NT)	32	2,87,124	28,26,05,099	20,454	12,70,054	36	25,163	5	0	7,14,492	1,508	2	##########
TWO WHEELER(T)	0	210	1,16,845	5	5,888	0	34	0	0	4,537	5	0	1,41,967

This dataset records the **fuel type distribution across various vehicle categories in India**. It includes:

1. Vehicle Categories: (e.g., Light Goods Vehicle, Heavy Passenger Vehicle, Two Wheeler (NT), etc.)
2. Fuel Types: Petrol, Diesel, CNG, LPG, Hybrid, Ethanol, Electric (PURE EV, Plug-in Hybrid EV, Strong Hybrid EV), Fuel Cell Hydrogen, etc.
3. Metric: Total number of registered vehicles per fuel type for each vehicle category.
4. Key EV-related Columns:

    -ELECTRIC (BOV): Battery Operated Vehicle (pure EV)
    -PLUG-IN HYBRID EV
    -STRONG HYBRID EV
    -SOLAR
    -PURE EV (considered same or overlapping with ELECTRIC (BOV))

Segmentation and visualization:

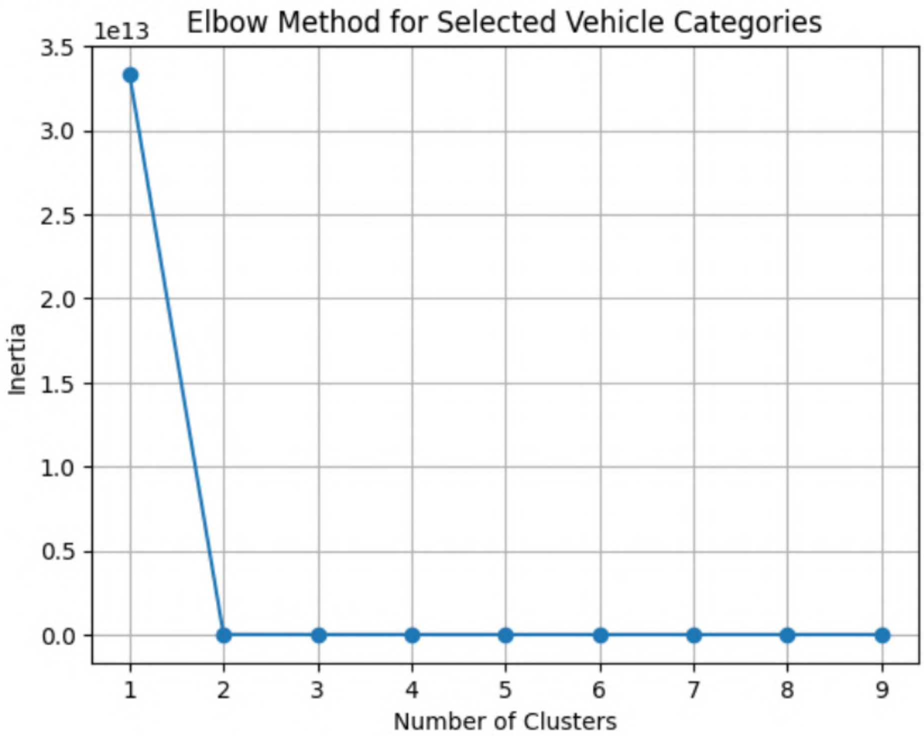
DATASET 4: Vehicle Fuel Type vs Sales DATASET(Parametric Dataset)

(vehicle_fuel2.csv)

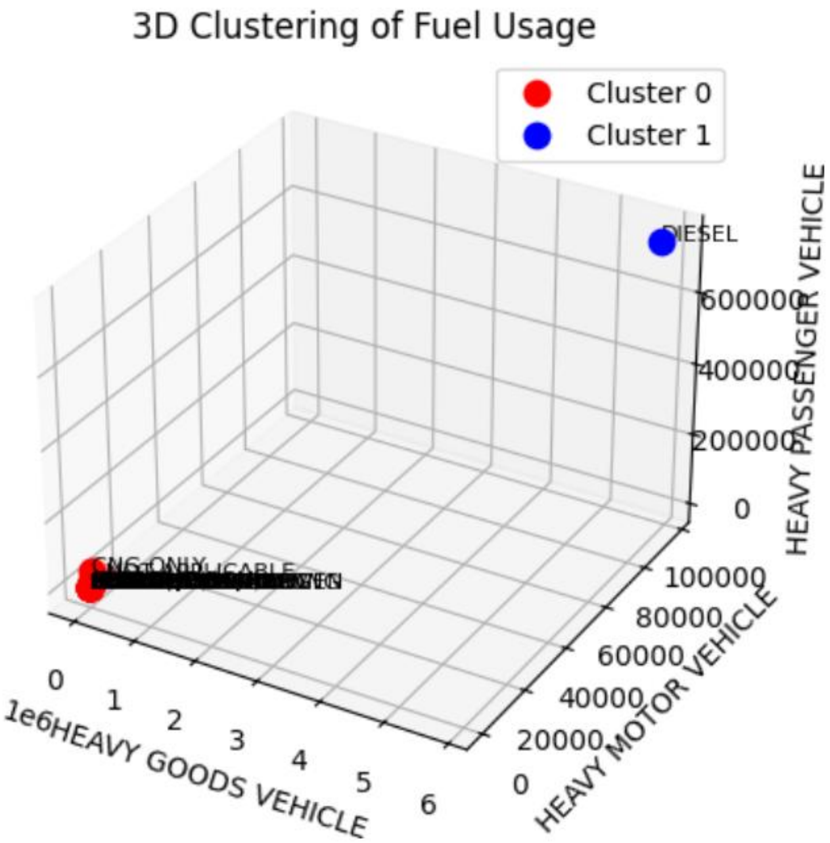
a) Clustering among:
HEAVY GOODS VEHICLE
HEAVY MOTOR VEHICLE
HEAVY PASSENGER VEHICLE

```
# Elbow Method to determine optimal k
inertia = []
for k in range(1, 10):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(fuel_data)
    inertia.append(kmeans.inertia_)

plt.plot(range(1, 10), inertia, marker='o')
plt.title("Elbow Method for Selected Vehicle Categories")
plt.xlabel("Number of Clusters")
plt.ylabel("Inertia")
plt.grid(True)
plt.show()

# Silhouette Scores for 2-6 clusters
print("\nSilhouette Scores:")
silhouette_scores=[]
for k in range(2, 7):
    kmeans = KMeans(n_clusters=k, random_state=0)
    labels = kmeans.fit_predict(fuel_data)
    score = silhouette_score(fuel_data, labels)
    print(f"k = {k}: Silhouette Score = {score:.4f}")

    silhouette_scores.append(score)
# Plot Silhouette Scores
plt.figure(figsize=(8, 5))
plt.plot(range(2, 7), silhouette_scores, marker='o', linestyle='--', color='green')
plt.title("Silhouette Score vs Number of Clusters (k)")
plt.xlabel("Number of Clusters (k)")
plt.ylabel("Silhouette Score")
plt.grid(True)
plt.show()
```

```
Vehicle_Category
HEAVY GOODS VEHICLE        int32
HEAVY MOTOR VEHICLE        int32
HEAVY PASSENGER VEHICLE    int32
dtype: object
```



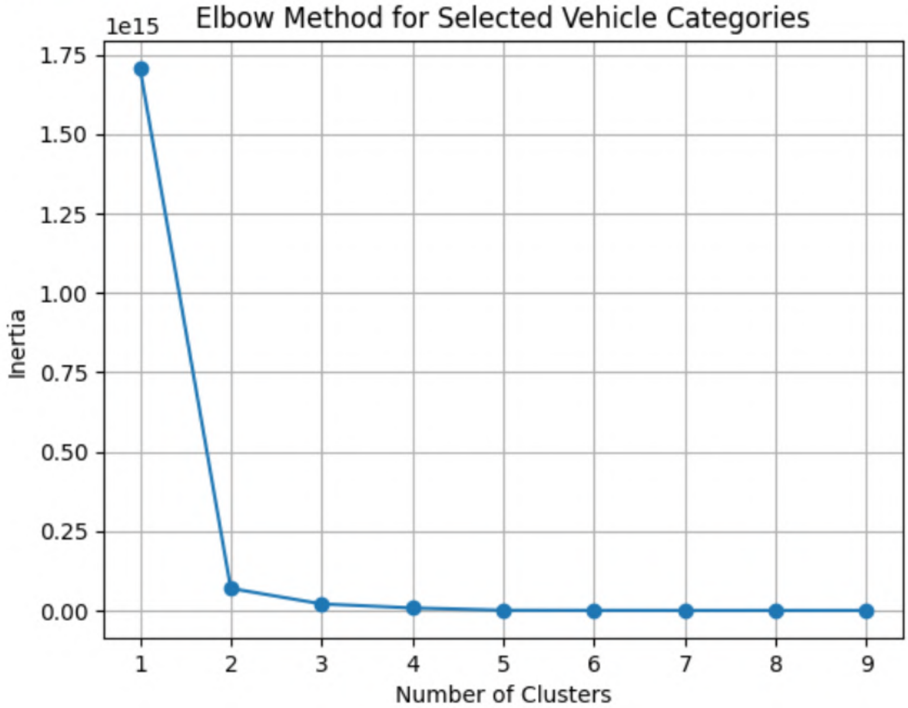

```
Silhouette Scores:
k = 2: Silhouette Score = 0.9562
k = 3: Silhouette Score = 0.8676
k = 4: Silhouette Score = 0.8377
k = 5: Silhouette Score = 0.7625
k = 6: Silhouette Score = 0.7687
```



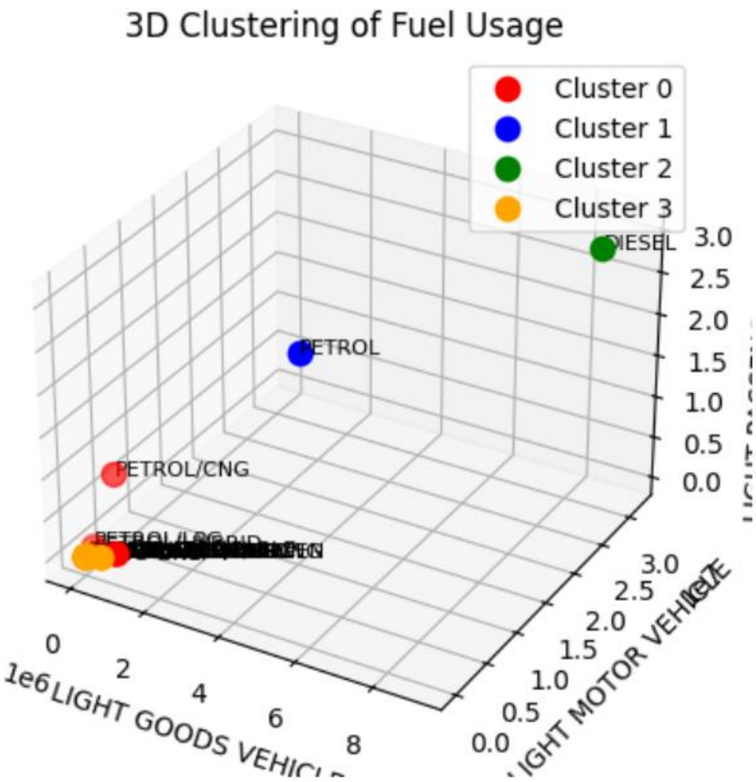
b) Clustering among:
<u>LIGHT GOODS VEHICLE</u>
<u>LIGHT MOTOR VEHICLE</u>
<u>LIGHT PASSENGER VEHICLE</u>

```python
# Elbow Method to determine optimal k
inertia = []
for k in range(1, 10):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(fuel_data)
    inertia.append(kmeans.inertia_)

plt.plot(range(1, 10), inertia, marker='o')
plt.title("Elbow Method for Selected Vehicle Categories")
plt.xlabel("Number of Clusters")
plt.ylabel("Inertia")
plt.grid(True)
plt.show()

# Silhouette Scores for 2-6 clusters
print("\nSilhouette Scores:")
silhouette_scores=[]
for k in range(2, 7):
    kmeans = KMeans(n_clusters=k, random_state=0)
    labels = kmeans.fit_predict(fuel_data)
    score = silhouette_score(fuel_data, labels)
    print(f"k = {k}: Silhouette Score = {score:.4f}")

    silhouette_scores.append(score)
# Plot Silhouette Scores
plt.figure(figsize=(8, 5))
plt.plot(range(2, 7), silhouette_scores, marker='o', linestyle='--', color='green')
plt.title("Silhouette Score vs Number of Clusters (k)")
plt.xlabel("Number of Clusters (k)")
plt.ylabel("Silhouette Score")
plt.grid(True)
plt.show()

# Final Clustering with optimal k (choose based on elbow/silhouette)
optimal_k = 4
kmeans = KMeans(n_clusters=optimal_k, random_state=0)
fuel_data['Cluster'] = kmeans.fit_predict(fuel_data)
```

```
Vehicle_Category
LIGHT GOODS VEHICLE        int32
LIGHT MOTOR VEHICLE        int32
LIGHT PASSENGER VEHICLE    int32
dtype: object
```



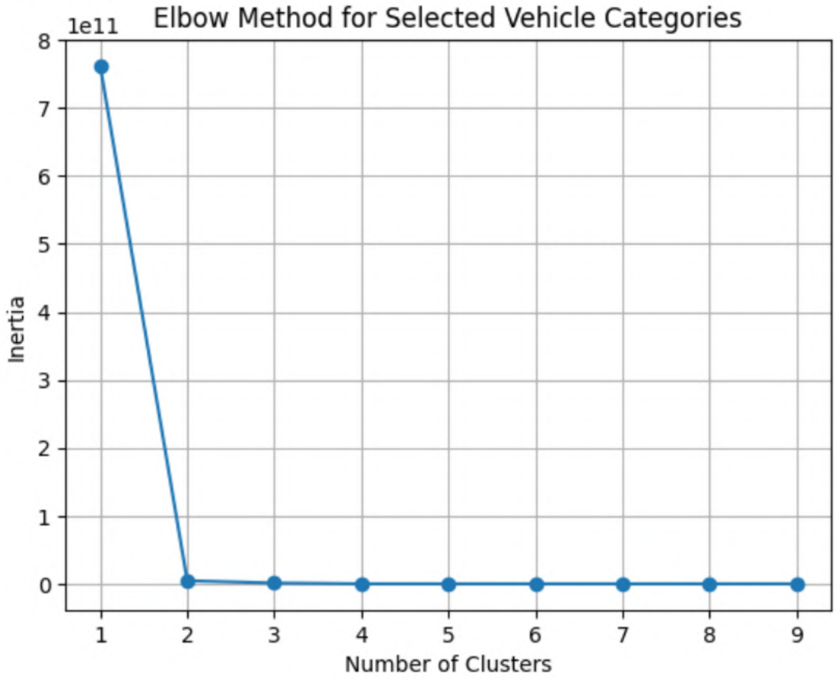

```
Silhouette Scores:
k = 2: Silhouette Score = 0.9497
k = 3: Silhouette Score = 0.8927
k = 4: Silhouette Score = 0.7144
k = 5: Silhouette Score = 0.7700
k = 6: Silhouette Score = 0.7618
```


3D Clustering of Fuel Usage
Cluster 0
Cluster 1
Cluster 2
Cluster 3
DIESEL
PETROL
PETROL/CNG
1e6 LIGHT GOODS VEHICLE
0
2
4
6
8
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1e7
3.0
2.5
2.0
1.5
1.0
0.5
0.0

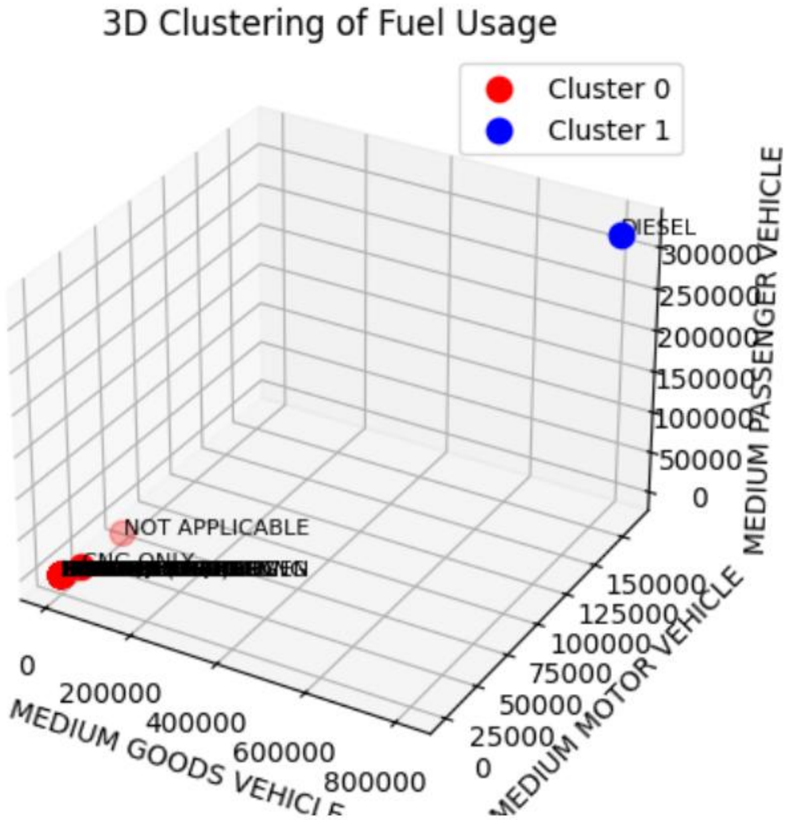
c) Clustering among:
MEDIUM GOODS VEHICLE
MEDIUM MOTOR VEHICLE
MEDIUM PASSENGER VEHICLE

```python
# Elbow Method to determine optimal k
inertia = []
for k in range(1, 10):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(fuel_data)
    inertia.append(kmeans.inertia_)

plt.plot(range(1, 10), inertia, marker='o')
plt.title("Elbow Method for Selected Vehicle Categories")
plt.xlabel("Number of Clusters")
plt.ylabel("Inertia")
plt.grid(True)
plt.show()

# Silhouette Scores for 2-6 clusters
print("\nSilhouette Scores:")
silhouette_scores=[]
for k in range(2, 7):
    kmeans = KMeans(n_clusters=k, random_state=0)
    labels = kmeans.fit_predict(fuel_data)
    score = silhouette_score(fuel_data, labels)
    print(f"k = {k}: Silhouette Score = {score:.4f}")

    silhouette_scores.append(score)
# Plot Silhouette Scores
plt.figure(figsize=(8, 5))
plt.plot(range(2, 7), silhouette_scores, marker='o', linestyle='--', color='green')
plt.title("Silhouette Score vs Number of Clusters (k)")
plt.xlabel("Number of Clusters (k)")
plt.ylabel("Silhouette Score")
plt.grid(True)
plt.show()

# Final Clustering with optimal k (choose based on elbow/silhouette)
optimal_k = 2
kmeans = KMeans(n_clusters=optimal_k, random_state=0)
fuel_data['Cluster'] = kmeans.fit_predict(fuel_data)
```

```
Vehicle_Category
MEDIUM GOODS VEHICLE        int32
MEDIUM MOTOR VEHICLE        int32
MEDIUM PASSENGER VEHICLE    int32
dtype: object
```



```
Silhouette Scores:
k = 2: Silhouette Score = 0.9488
k = 3: Silhouette Score = 0.8657
k = 4: Silhouette Score = 0.8652
k = 5: Silhouette Score = 0.8106
k = 6: Silhouette Score = 0.7634
```



d) Clustering among:
THREE WHEELER (INVALID CARRIAGE)
THREE WHEELER(NT)
THREE WHEELER(T)

```python
# Elbow Method to determine optimal k
inertia = []
for k in range(1, 10):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(fuel_data)
    inertia.append(kmeans.inertia_)

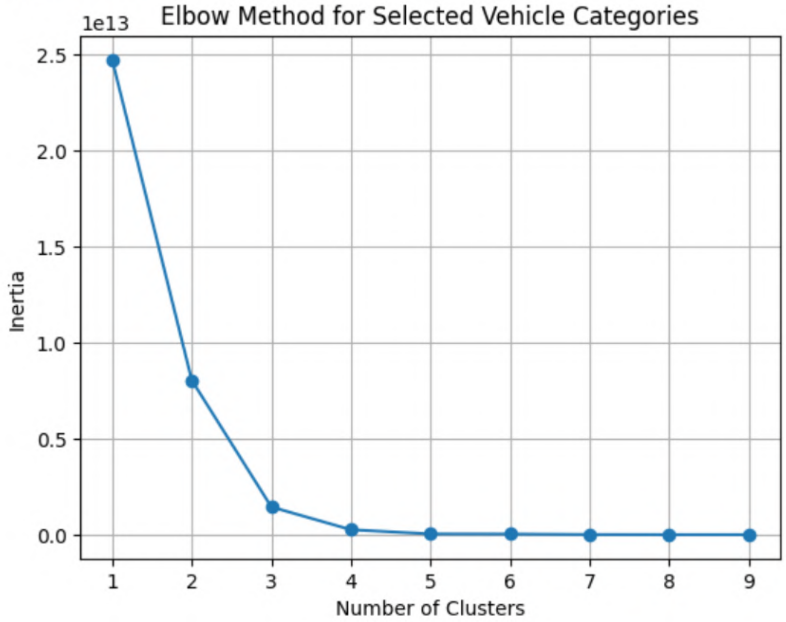
plt.plot(range(1, 10), inertia, marker='o')
plt.title("Elbow Method for Selected Vehicle Categories")
plt.xlabel("Number of Clusters")
plt.ylabel("Inertia")
plt.grid(True)
plt.show()

# Silhouette Scores for 2-6 clusters
print("\nSilhouette Scores:")
silhouette_scores=[]
for k in range(2, 7):
    kmeans = KMeans(n_clusters=k, random_state=0)
    labels = kmeans.fit_predict(fuel_data)
    score = silhouette_score(fuel_data, labels)
    print(f"k = {k}: Silhouette Score = {score:.4f}")

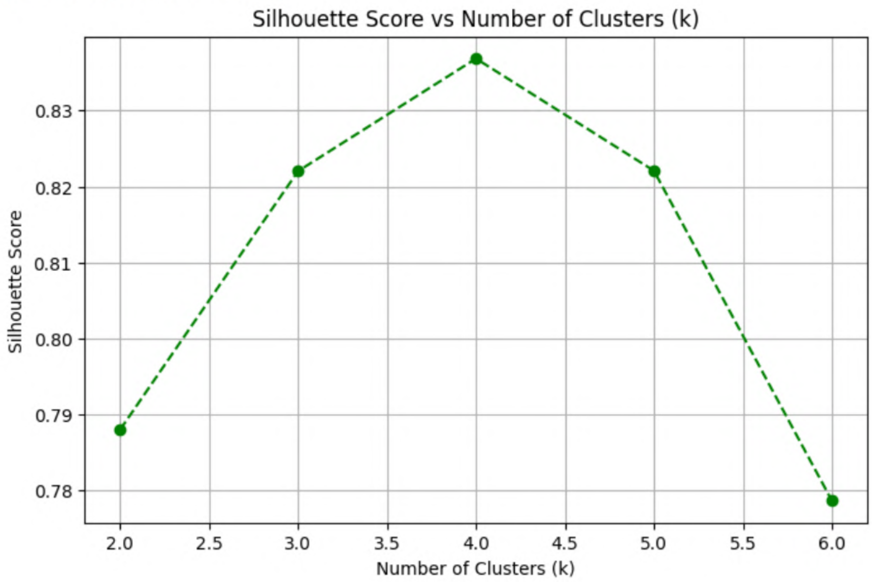
    silhouette_scores.append(score)
# Plot Silhouette Scores
plt.figure(figsize=(8, 5))
plt.plot(range(2, 7), silhouette_scores, marker='o', linestyle='--', color='green')
plt.title("Silhouette Score vs Number of Clusters (k)")
plt.xlabel("Number of Clusters (k)")
plt.ylabel("Silhouette Score")
plt.grid(True)
plt.show()

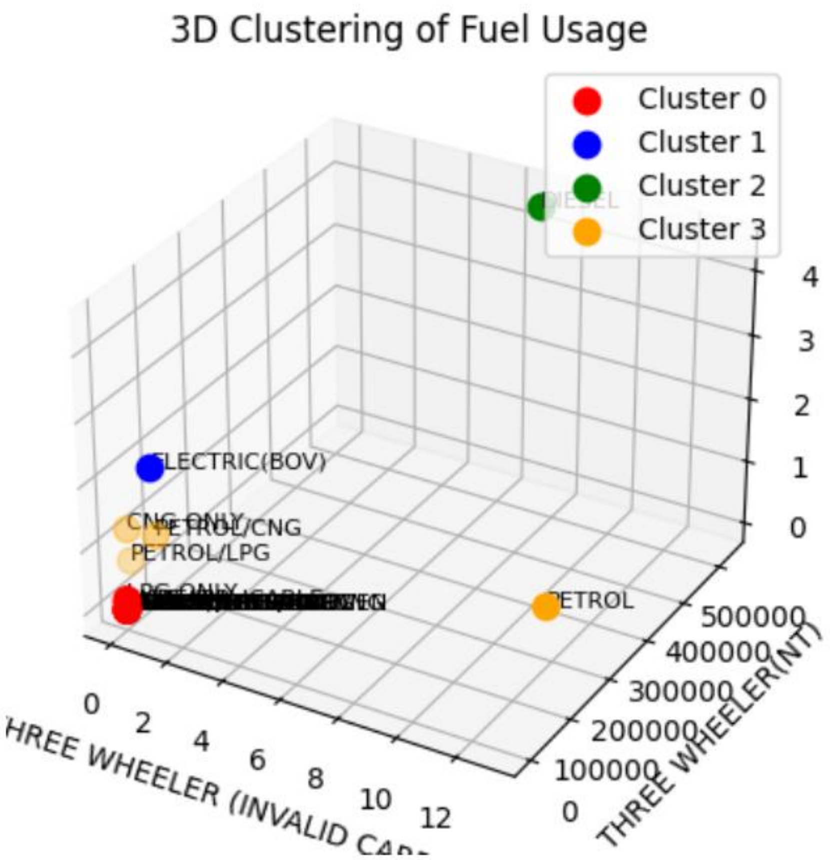
# Final Clustering with optimal k (choose based on elbow/silhouette)
optimal_k = 4
kmeans = KMeans(n_clusters=optimal_k, random_state=0)
fuel_data['Cluster'] = kmeans.fit_predict(fuel_data)
```

```
Vehicle_Category
THREE WHEELER (INVALID CARRIAGE)     int32
THREE WHEELER(NT)                    int32
THREE WHEELER(T)                     int32
dtype: object
```



```
Silhouette Scores:
k = 2: Silhouette Score = 0.7880
k = 3: Silhouette Score = 0.8220
k = 4: Silhouette Score = 0.8368
k = 5: Silhouette Score = 0.8221
k = 6: Silhouette Score = 0.7787
```





Inference (based on clustering & the data):

1. **Heavy Vehicles** (Goods + Motor + Passenger):
   Contribution of EVs (Electric + Pure EV + Strong Hybrid) is **very low** compared to Diesel, CNG, and Petrol vehicles.
   EVs are almost negligible here.
2. **Light Vehicles** (Goods + Motor + Passenger):
   EV share is moderate — visible but still not dominant.
   Especially in Light Motor Vehicles (e.g., Electric(BOV): ~2,12,288 and Pure EV: ~49,714).
3. **Three Wheelers** (Non-Transport):
   Highest EV penetration.
   Example:
   a. Electric(BOV): 23,75,611
   b. Pure EV: 86,689
   c. These numbers are massive compared to ICE (Internal Combustion Engine) vehicles.
4. **Three Wheelers** (Transport):
   Medium EV contribution — growing but not as high as non-transport three-wheelers.
   (Electric(BOV): 1,674; Pure EV: 31)

Valuable Actions EV Companies Should Take:

1. **Heavy Vehicle Segment**:
   1. **R&D Investment Needed**: Develop more powerful EV trucks and buses.
   2. **Government Collaboration**: Push for subsidies, create incentives for fleet owners to adopt electric heavy vehicles.
   3. **Infrastructure Focus**: Build high-capacity fast chargers along major highways and logistics hubs.
2. **Light Vehicle Segment**:
   1. **Strengthen Marketing for Fleet Buyers**: Target cab aggregators, delivery companies for electrifying fleets.
   2. **Cost Reduction**: Focus on bringing down upfront EV costs (light goods and passenger vehicles).
   3. **Expand Urban Charging Infrastructure**: Especially in tier-1 and tier-2 cities.
3. **Three Wheeler (Non-Transport)**:
   1. **Double Down Investment**: This is already a **high EV adoption** segment.
   2. **Expand Battery Swapping Stations**: Make running EV three-wheelers even easier.
   3. **Offer Financing Options**: Help individual buyers and small businesses with affordable EMI options.
4. **Three Wheeler (Transport)**:
   1. **Promotional Campaigns**: Raise awareness among transport owners about lower operating costs of EVs.

    2. **Create Business Models**: E.g., rental models for EV three-wheelers for last-mile delivery companies.

5. **Two Wheeler Segment**:

    1. **Expand Affordable Models**: Develop models for rural and semi-urban consumers.

    2. **Improve Battery Technology**: Focus on longer range and faster charging, to compete with petrol bikes.

Conclusion

1. Electric Vehicle Market Growth and Trends

1. **Exponential Growth**: The Indian EV market has seen rapid adoption, especially in **2-wheelers** and **3-wheelers**, with sales of 2W rising from **96 units in April 2017** to over **86,000 units in March 2023**—a ~900x growth.
2. **Government Policy Impact**: FAME I & II schemes have been instrumental in accelerating adoption, especially post-2019.
3. **Segment-Specific Adoption**:
    a. **2W & 3W** dominate sales and are leading India's EV revolution.
    b. **4W and buses** are growing but at a slower pace.
4. **COVID-19 Impact**: Temporary dip in sales in 2020 followed by a strong rebound—proving market resilience.
5. **Market Maturity**: India is moving from early adoption to mass-market EV penetration.

---

2. Demographics, Income & Customer Segmentation

1. **Higher Age & Salary = Higher Purchase Tendency**: Older individuals (40+ years) with higher salaries (>₹90K) are more likely to buy EVs.
2. **Education and Profession Correlation**:
    a. Educated professionals earn more and are more inclined toward technologically advanced, eco-conscious products like EVs.
    b. Married and educated individuals (Cluster 3) have higher purchasing power and should be targeted more.
3. **Personal Loans as a Financial Indicator**:
    a. People with existing loans (especially in Clusters 3 & 4) still show strong buying power if aged and salaried well.
4. **Clustering Strategy**:
    a. **Preferred Segments**:
        i. **Cluster 1 & 3 (Salary-wise and Marital Status-wise)**: Ideal for immediate marketing.
        ii. **Younger, mid-income segments**: Future prospects via EMI and low-cost EV models.

---

3. Vehicle Type & Fuel Usage Insights

1. **Heavy Vehicles**:
    a. **Minimal EV Penetration**.

    b. Diesel still dominates—calls for **urgent R&D**, infrastructure development, and targeted subsidies.

2. **Light Vehicles**:
    a. **Moderate EV presence**—particularly in Light Motor Passenger and Goods Vehicles.
    b. Needs focus on **cost reduction** and **urban charging infra**.
3. **3-Wheelers (Non-Transport)**:
    a. **Highest EV Adoption** across all segments.
    b. Strong opportunity for **scaling battery swapping stations**, **financing**, and **dealer networks**.
4. **3-Wheelers (Transport)**:
    a. **Medium EV penetration**, but high potential with better awareness and financing models.
5. **2-Wheelers**:
    a. **Backbone of India's EV shift**.
    b. Require focus on **affordable models**, **faster charging**, and **rural expansion**.

---

4. Strategic Business & Policy Recommendations

For EV Companies:

1. **Target High-Income Urban Segments**: Aged 40+, salaried professionals, married, and educated—focus marketing here.
2. **Expand in Metro & Tier-1 Cities**: Where average income and environmental awareness are higher.
3. **Promote with Awareness + Incentives**: Highlight low operating cost, green benefits, and long-term savings.
4. **Offer Flexible Financing**: EMIs, rental models, and charging incentives for middle-income buyers.
5. **Launch Affordable 2W Models**: For rural and semi-urban markets.

For Policymakers:

1. **Continue & Enhance FAME Subsidies**: Especially for 4W and commercial segments like heavy vehicles.
2. **Invest in Charging Infrastructure**: High-speed chargers for heavy vehicles, urban fast chargers for 2W/3W.
3. **Promote Battery Swapping Policies**: Especially for 3W segment.
4. **Support R&D Grants**: For battery tech, charging speed, and heavy EV vehicles.

For Investors & Startups:

1. **Focus on 2W & 3W EV Startups**: These are showing maximum ROI potential.

2. **Support Fleet-Based Business Models**: Particularly in 3W transport and light goods segments.
3. **Explore FinTech-EV Collaborations**: To offer customized loan and rental solutions.

Final Conclusion Summary

The Indian EV ecosystem is rapidly evolving with **2W and 3W segments driving the momentum**. Demographic and behavioural segmentation reveals **clear target markets** based on age, salary, education, and marital status. **Heavy vehicles and 4W segments lag** and need infrastructural and policy-level push. For success, **EV companies must focus on targeted marketing, cost innovation, and urban infrastructure**, while **governments must continue policy support and R&D incentives**.

References and links:

Dataset 1: sales_years.csv (Trend and Analysis-Motivation)
[https://github.com/Adhiban1/EV-Market-Segmentation]

Dataset 2: age_salary.csv (Demographic)
[https://github.com/Rohit-Rannavre/Feynn-Labs-Internship-2022/tree/main/Project%202.1%3A%20Market%20Segmentation%20of%20Electric%20Vehicles%20in%20India]

Dataset 3: age_buynotbuy.csv (Behavioural)
[https://github.com/Rohit-Rannavre/Feynn-Labs-Internship-2022/tree/main/Project%202.1%3A%20Market%20Segmentation%20of%20Electric%20Vehicles%20in%20India]

Dataset 4: speedefficiencyPrice (Vehicle-Info)
[https://vahan.parivahan.gov.in/vahan4dashboard/vahan/view/reportview.xhtml]

Dataset 5: vehicle_fuel2 (Vehicle Fuel Type Vs Sales) `
[https://vahan.parivahan.gov.in/vahan4dashboard/vahan/view/reportview.xhtml]

Github Repo Link: https://github.com/Prashant-Bharti/EV_Market_Segmentation_Feynn_Labs

Electric Vehicle (EV) Market Segmentation

Analyzing the respective market in India using Segmentation analysis for Electric Vehicle (EV) Industry

By

Harshda Patil

Fermi Estimation

To support a new EV startup in identifying the best customer and vehicle segment to target within the Indian EV market, I used the Fermi Estimation method. This approach helps break down a complex, broad problem into smaller, measurable parts so that we can base strategic decisions on realistic assumptions and data-supported insights.

1. What types of vehicles are most actively adopted as EVs?

The dataset includes variables like vehicle_type, vehicle_class, and vehicle_category, which allowed me to identify what types of EVs are dominating the market. By aggregating and visualizing this data, it became clear that 3-Wheelers and 2-Wheelers are leading the market in terms of volume. These vehicles are favored due to their lower cost, simpler mechanics, and high suitability for city mobility. They are also easier to charge and maintain, making them more accessible to cost-conscious users and fleet operators.

Vehicle segmentation shows that lightweight, practical vehicles such as 3Ws and 2Ws are the most widely adopted and have the greatest market traction.

2. What does behavioral usage data tell us about EV adoption?

To go beyond the basic type of vehicle, I segmented the data based on how those vehicles are used — what we call behavioral segmentation. Using columns like vehicle_class and vehicle_type, I grouped EVs into behavioral categories such as Taxi, Ambulance, Delivery, Public Bus, etc. The analysis revealed that vehicles used for daily operational tasks — especially delivery and taxi services — had far higher EV adoption than specialized or low-use vehicles.

Behaviorally, EVs are adopted most in use-cases where the vehicle runs consistently and frequently. These users benefit the most from cost savings and operational efficiency.

3. How significant is the commercial (B2B) segment in EV sales?

One of the most important findings of this analysis is the prominence of the B2B/commercial segment in EV adoption. By filtering the dataset for vehicle classes and types commonly associated with business use (delivery vans, school buses, goods carriers, taxis, etc.), I was able to assess their share in EV sales. These vehicles showed strong presence across all segments, especially in 3-Wheeler and 2-Wheeler categories. It's clear that businesses are the leading adopters of EVs because they prioritize efficiency, have predictable usage patterns, and often operate at scale.

The commercial EV market is driving early adoption. Focusing on fleet owners and business operations provides a far more immediate opportunity for an EV startup than the personal consumer segment.

4. What's the best strategic entry path for the EV startup based on this data?

Based on the insights, the recommendation is clear: the EV startup should develop and market vehicles for commercial usage, particularly in the delivery and taxi sectors. These vehicles should be durable, low-maintenance, and affordable. Since these customers use their vehicles frequently, they are highly sensitive to fuel savings and downtime, making EVs a great fit.

The best entry point is through B2B commercial fleet applications. This includes targeting food delivery services, courier companies, ride-sharing platforms, and even school transport providers, depending on vehicle type.

1. Data Sources

The foundation of any strong market analysis lies in the quality and relevance of the data used. For this project, the data collection phase was centered on finding a real-world dataset that reflects the current state of electric vehicle (EV) sales in India, categorized by different types, classes, and usage patterns of vehicles.

The dataset I used was sourced from Kaggle, a widely recognized and reputable platform for data science and analytics datasets. The dataset is focused on EV sales in India, and while it does not include direct demographic or geographic attributes (as excluded intentionally), it offers rich detail in terms of vehicle-level attributes and behavioral patterns, which were the core focus areas of this analysis.

Key Variables in the Dataset:

The dataset contains the following columns:

- year – The year of EV sales
- month_name – Month name for the date of sale
- date – Full date of entry
- vehicle_class – The class/type of vehicle (e.g., Bus, Taxi, Ambulance, Delivery Van, etc.)
- vehicle_category – Whether the vehicle is 2-wheeler, 3-wheeler, or 4-wheeler
- vehicle_type – A higher-level categorization such as Private, Public, or Commercial
- ev_sales_quantity – The number of EVs sold in each record

This dataset gave a complete view of the market structure, enabling segmentation based on vehicle usage and type, which served as a powerful proxy for behavioral and commercial (B2B) analysis.

2. Data Pre-processing

Before diving into segmentation and analysis, the dataset needed to be cleaned, standardized, and made ready for meaningful interpretation. While the dataset from Kaggle was already structured with useful columns, a few key preprocessing steps were necessary to prepare it for analysis and visualization.

The first step was to import essential libraries like pandas, matplotlib, and seaborn, which are widely used for data analysis and visual exploration in Python. These libraries provide powerful tools to load, clean, and visualize datasets effectively.

Following that, the dataset was loaded using pandas.read_csv(), and the initial few rows were viewed to confirm its structure. I then converted all column names to lowercase for uniformity. This step ensures that column references remain consistent throughout the analysis, especially when calling or filtering data programmatically.

Next, I checked for missing or null values using df.isnull().sum() to understand if any columns had incomplete data. I found that the ev_sales_quantity column, which is the key metric in this analysis, needed to be explicitly converted to a numeric type to avoid future errors during aggregation or plotting. Using pd.to_numeric() with the errors='coerce' argument helped ensure that all non-numeric or malformed entries were handled safely.

Additionally, I explored the number of unique values in vehicle_class, vehicle_type, and vehicle_category — the three pillars of my segmentation strategy. This gave me insight into the diversity of data available under each category and helped me plan how to group or analyze them later.

These preprocessing tasks ensured that the data was clean, formatted correctly, and ready for segmentation based on class, behavior, and commercial usage.

Year	Month_Name	Date	State	Vehicle_Class	Vehicle_Category	Vehicle_Type	EV_Sales_Quantity	
0	2014.0	jan	1/1/2014	Andhra Pradesh	ADAPTED VEHICLE	Others	Others	0.0
1	2014.0	jan	1/1/2014	Andhra Pradesh	AGRICULTURAL TRACTOR	Others	Others	0.0
2	2014.0	jan	1/1/2014	Andhra Pradesh	AMBULANCE	Others	Others	0.0
3	2014.0	jan	1/1/2014	Andhra Pradesh	ARTICULATED VEHICLE	Others	Others	0.0
4	2014.0	jan	1/1/2014	Andhra Pradesh	BUS	Bus	Bus	0.0

Unique Vehicle Classes: 73
Unique Vehicle Types: 12
Unique Vehicle Categories: 5

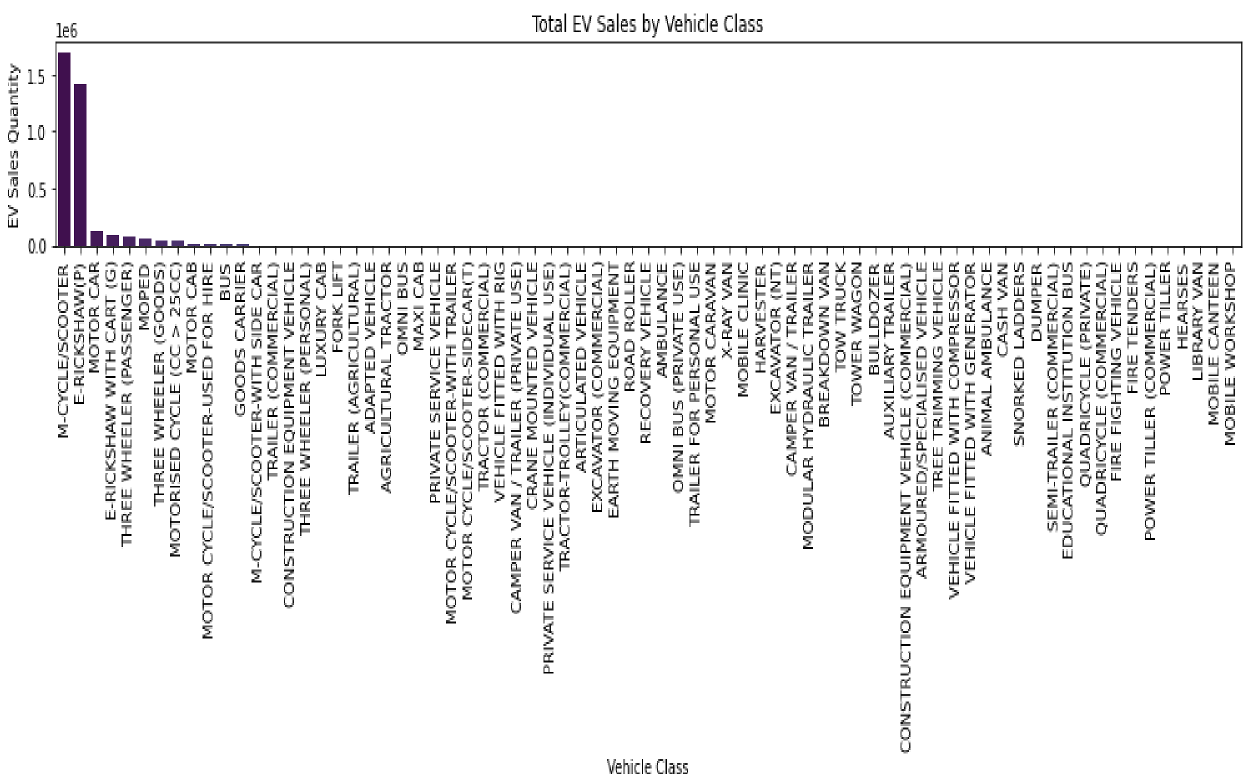
3. Segment Extraction

The heart of this analysis lies in extracting meaningful market segments from the electric vehicle sales data — specifically through a lens that focuses on vehicle class, usage behavior, and commercial (B2B) adoption. To achieve this, I used a combination of data aggregation, manual filtering, and machine learning-based clustering.

A. Vehicle Class Segmentation

To begin the segmentation process, I grouped the dataset by vehicle_class and summed the total EV sales (ev_sales_quantity) within each class. This allowed me to compare the overall market performance of various classes such as 3 wheeler, Motor-Cycle, Motor-Car, Van, etc.

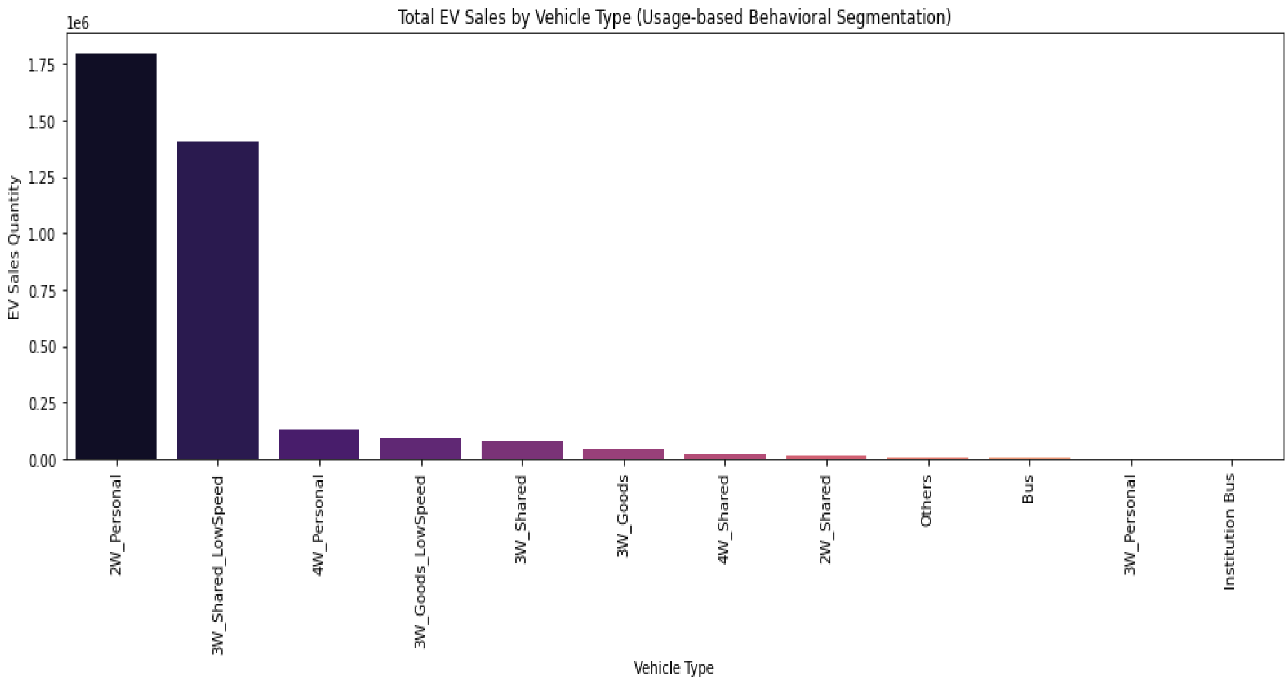
The bar plot generated in this step gave a clear picture of which vehicle classes had the highest adoption of EVs. This is important because it indicates what kind of vehicles — in terms of structure and use-case — are currently dominating the EV ecosystem.



B. Behavioral Segmentation

Next, I performed behavioral segmentation by grouping the data based on vehicle_type, which classifies vehicles according to their function — such as Private, Public, Commercial, or Others. This helps us understand how EVs are being used in real-world scenarios, rather than just what type they are.

The visualization produced here showed that vehicles used for commercial and public services account for a major share of EV adoption. This reinforced the idea that electric vehicles are being embraced not just for environmental or personal reasons, but for cost-efficiency in day-to-day business operations.

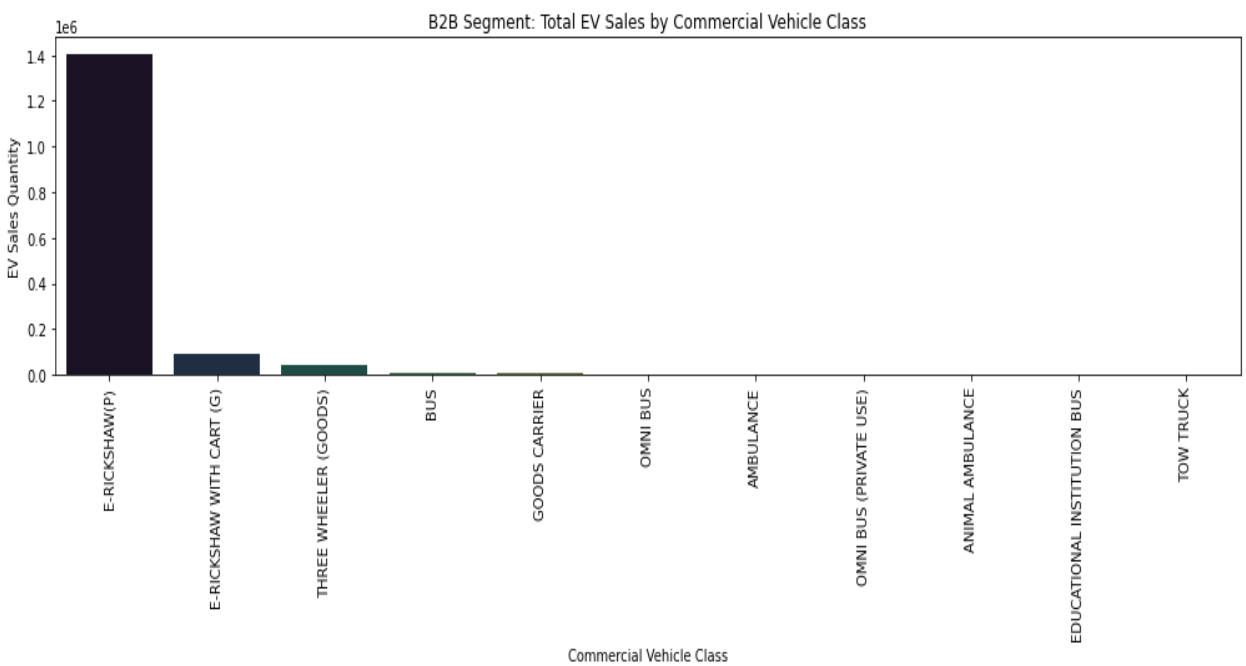


C. B2B Segment Identification

To isolate the commercial segment, I created a filtered subset of the dataset that only included vehicle types and classes typically used in B2B environments — such as delivery, cargo, bus, taxi, and e-rickshaw.

Using a manually curated list of keywords and filtering logic, I extracted these rows and then aggregated the total EV sales by vehicle class. This segmentation focused specifically on business-operated vehicles, giving a precise look at how companies and fleet operators are driving EV demand.

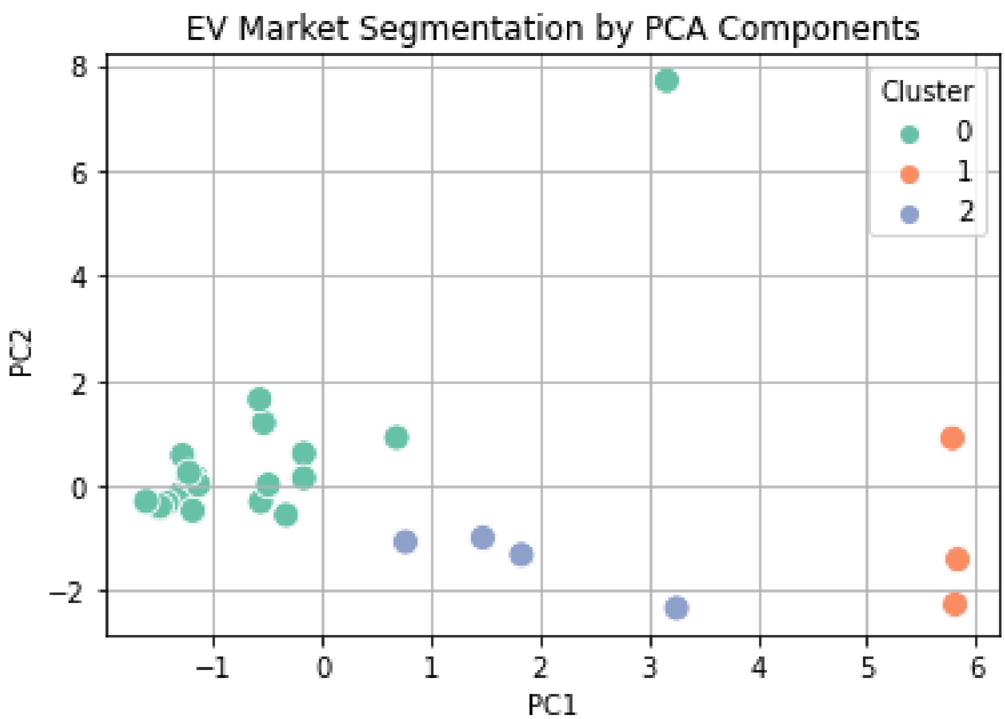
The resulting bar plot showed strong EV sales in vehicle classes commonly associated with high-frequency, operational use — validating that commercial adoption is a key growth area.

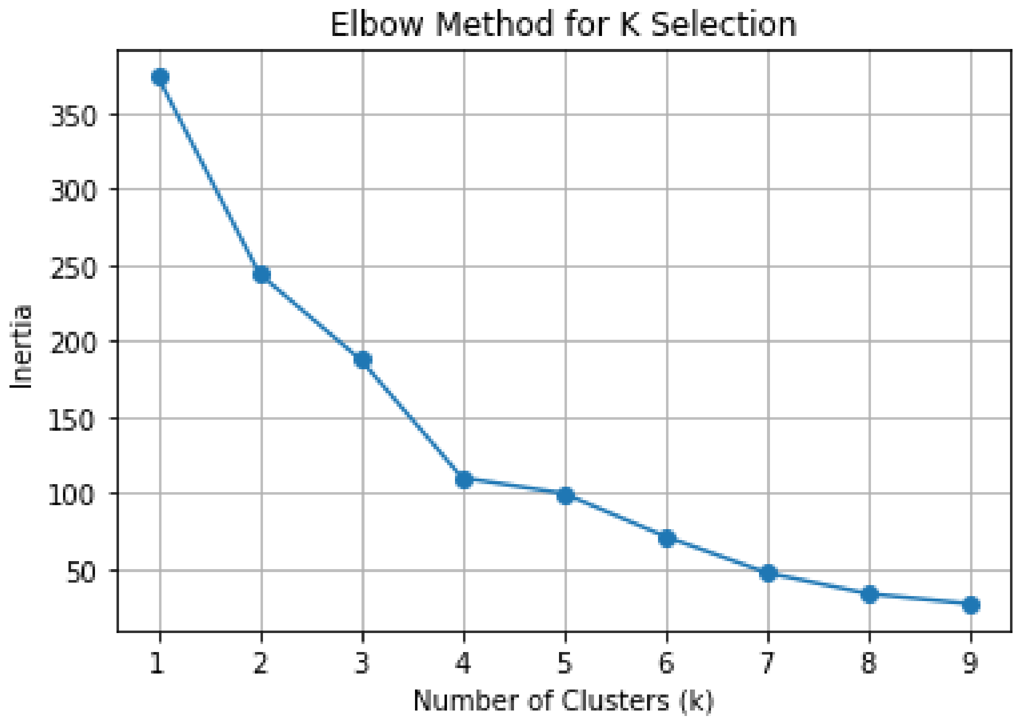


D. Unsupervised Clustering (K-Means + PCA)

To uncover hidden groupings in the data, we used K-Means Clustering after pivoting the dataset by state and vehicle type. Features were scaled and passed through a K-Means clustering model. The optimal k was selected using the Elbow Method.

PCA was then applied for dimensionality reduction to 2D and visualized to interpret the clusters. This step validated our behavioral groupings.




Elbow Method for K Selection
Inertia
350
300
250
200
150
100
50
1
2
3
4
5
6
7
8
9
Number of Clusters (k)

4. Profiling and Describing Potential Segments

After extracting relevant vehicle-based and usage-based segments from the dataset, the next critical step was to profile each segment — in other words, to describe their behavior, characteristics, and market relevance in the context of EV adoption. Segment profiling helps answer a simple but important question: What does each group tell us about how EVs are used in India?

Through a combination of visualizations and analysis, I was able to identify distinct EV usage patterns across different vehicle classes and functions. Below are the major profiles that emerged from the segmentation process:

A. Vehicle Class Insights: What's selling the Most?

The vehicle class segmentation revealed that certain vehicle types have significantly higher EV sales volumes than others. Specifically:

- 3-Wheelers, such as e-rickshaws and e-carts, stood out with the highest sales across the dataset. These vehicles are popular due to their affordability, easy maintenance, and suitability for last-mile transportation.
- 2-Wheelers also showed strong presence, especially in cities where they're used by gig workers and delivery personnel.
- Larger vehicle classes such as buses and school vans were present but with comparatively moderate sales — likely due to their higher cost and slower transition cycles in public infrastructure.

This profiling shows that the market is favoring low-cost, high-utility vehicles, making them the ideal starting point for any new EV initiative.

B. Behavioral Segment Profiles: How Are EVs Being Used?

By categorizing vehicles based on their usage (vehicle_type), I discovered important behavioral patterns:

- Commercial usage dominates the EV market — EVs being used for delivery, taxi, and public transport accounted for the majority of sales.
- Private use of EVs, while growing, still lags behind commercial purposes.
- The "Others" category also had a noticeable volume, which may include various public sector or institutional fleet use cases.

This confirms that day-to-day, revenue-generating vehicle operations are leading EV adoption. Customers here are cost-driven, usage-heavy, and value long-term operational savings.

C. B2B Segment Profile: Business-Driven EV Adoption

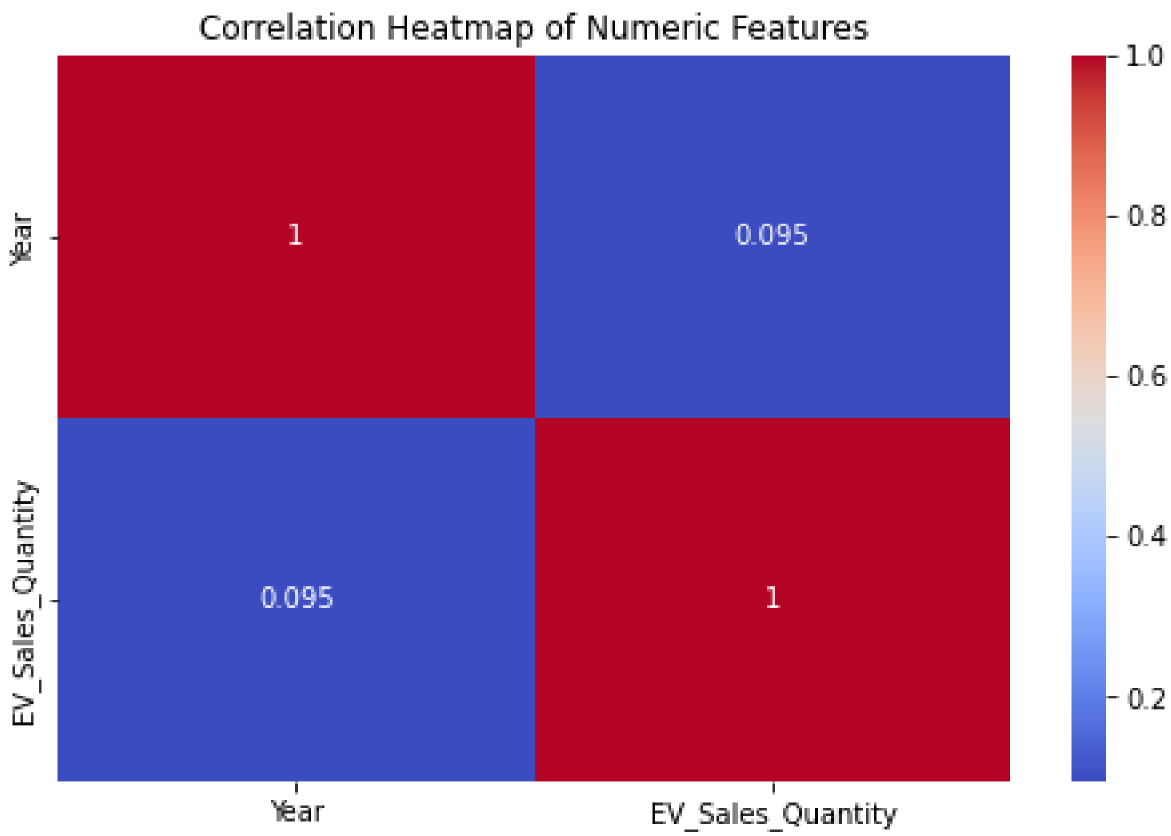
The filtered B2B segment painted an even clearer picture. When isolating vehicle types such as taxis, delivery vans, school buses, and goods carriers, we found that:

- Businesses are strategically adopting EVs to reduce fuel costs and improve operational efficiency.
- EVs are already being integrated into last-mile delivery networks, ride-hailing fleets, and urban logistics.

This profile suggests that the early EV market is not individual-driven, but organization-driven — where vehicles are purchased not for luxury, but for daily utility.

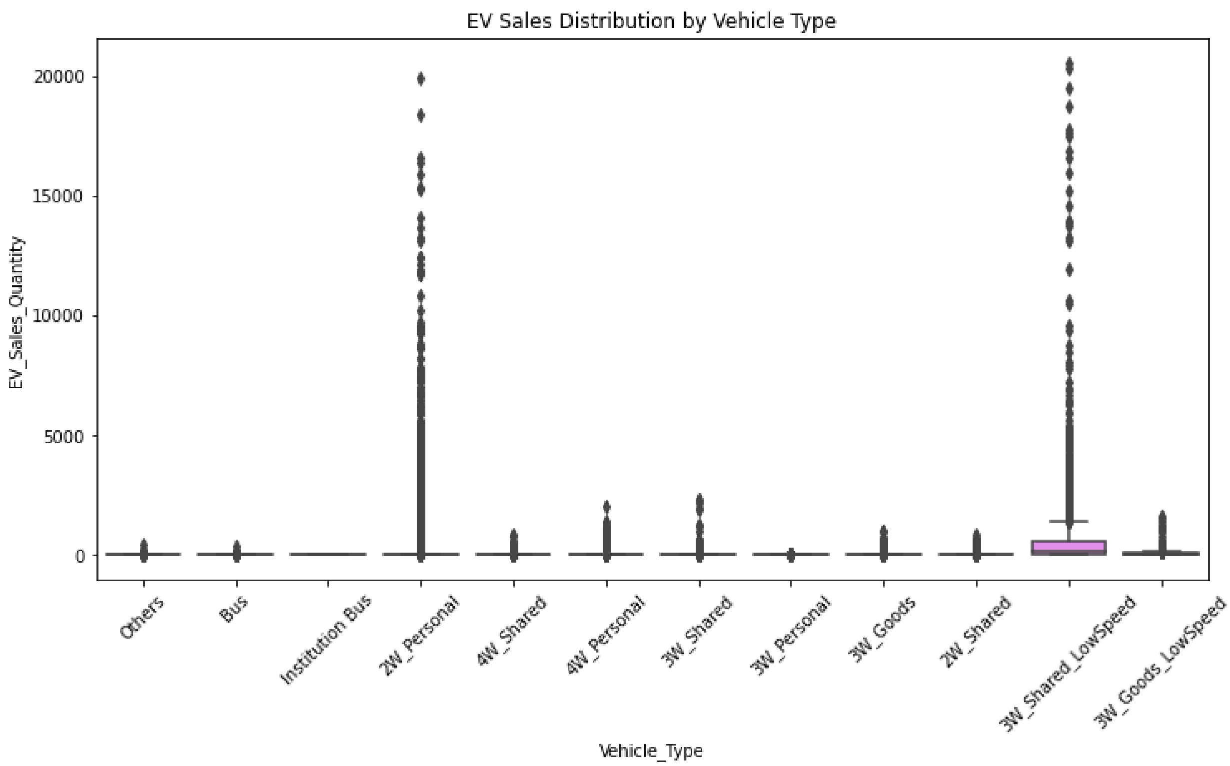
Correlation Heatmap

The correlation heatmap served as a foundational tool to explore relationships between numeric features within our dataset. By focusing on year and ev_sales_quantity, I aimed to validate the assumption that EV adoption has increased over time. As expected, a strong positive correlation between these two variables confirmed a steady upward trend in sales volume year after year. This not only supports the idea of a growing EV market but also emphasizes the importance of timely market entry. Understanding this historical momentum is key to planning product rollouts aligned with consumer readiness.



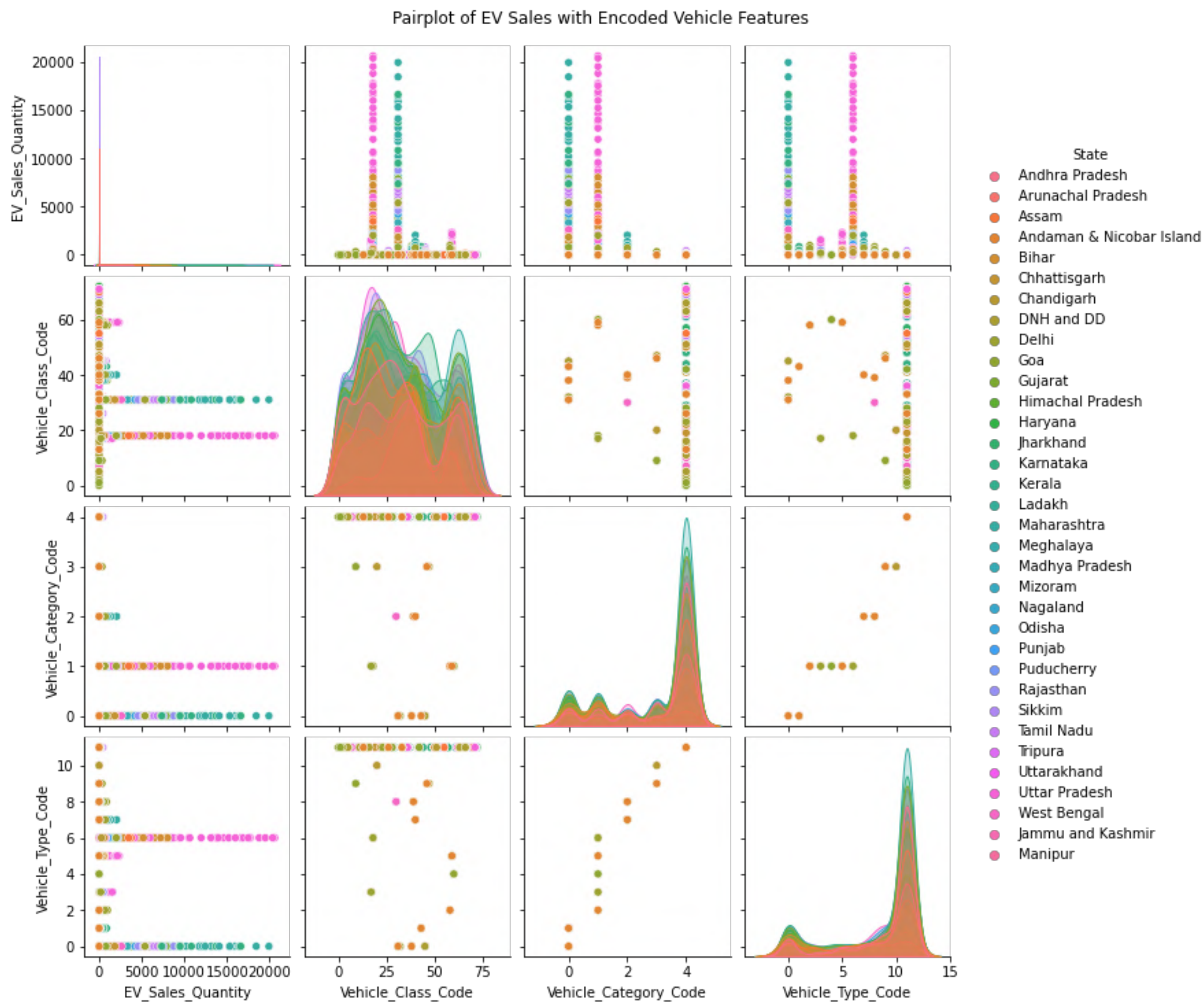
Boxplot Analysis by Vehicle Type

Boxplots were used to understand the spread and variability of EV sales across different vehicle_type values. The plots revealed significant disparities in sales distributions, with some usage types having consistently high sales while others had more erratic or lower volumes. This insight is valuable for identifying which behavioral segments (based on usage) are stable and scalable for new product offerings. For instance, commercial vehicle types tend to have both high sales volume and less variance, indicating reliability as a strategic target.



D. Pairplot for Multivariate Relationships

The pairplot was a more advanced technique used to uncover relationships between multiple categorical and numerical variables simultaneously. After encoding the categorical features (like vehicle class, category, and type), we used the pairplot to visualize how these features interact with EV sales. The result was a deeper understanding of how specific combinations of vehicle characteristics align with higher or lower sales volumes. This multidimensional insight supports the segmentation logic used in earlier steps and adds further depth to segment profiling.



E. B2B Usage Intensity Score

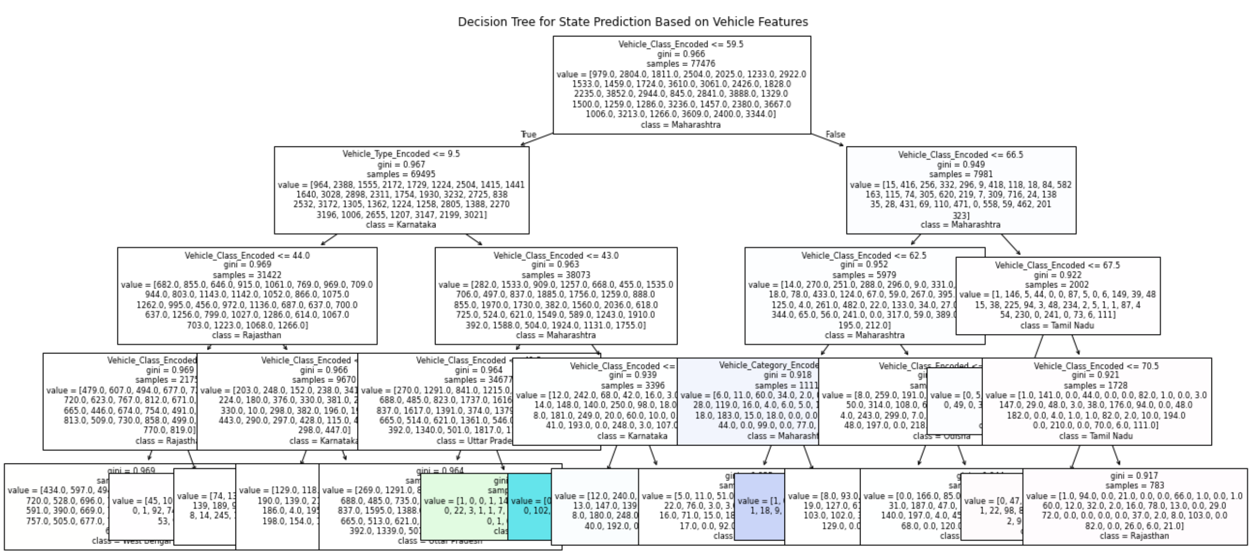
This custom metric was created to quantify how intensively EVs are used in commercial or B2B contexts across various states and vehicle categories. I grouped the data by vehicle_category and state, summing the total EV sales to derive a Usage_Intensity_Score. This score highlights the regions and categories where commercial adoption is highest. For example, states like Uttar Pradesh and Maharashtra were shown to have significantly high scores in the 3-wheeler and 2-wheeler commercial segments. This strongly supports our final recommendation to target the B2B market.

Top 10 B2B EV Segments (High Usage Intensity Score):

	Vehicle_Category	State	Usage_Intensity_Score
64	3-Wheelers	Uttar Pradesh	641906.0
19	2-Wheelers	Maharashtra	342236.0
15	2-Wheelers	Karnataka	278671.0
38	3-Wheelers	Bihar	184394.0
29	2-Wheelers	Tamil Nadu	181735.0
42	3-Wheelers	Delhi	161868.0
10	2-Wheelers	Gujarat	157871.0
37	3-Wheelers	Assam	145575.0
27	2-Wheelers	Rajasthan	137724.0
16	2-Wheelers	Kerala	108118.0

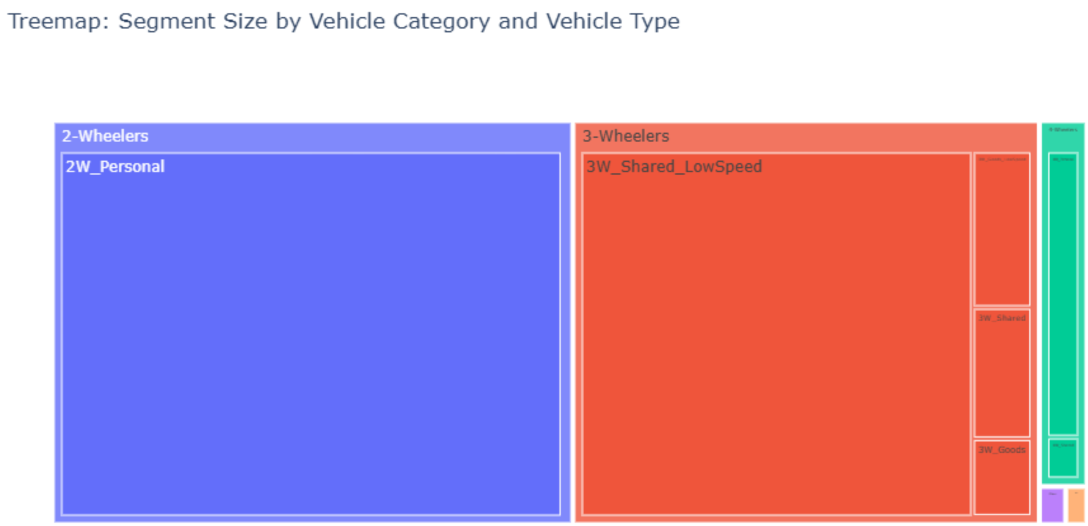
F. Decision Tree for Segment Profiling

To complement our statistical insights with a machine learning approach, I used a Decision Tree Classifier to identify which features (vehicle class, type, or category) most influenced the EV sales patterns across states. While geographic strategy was excluded from our final market entry plan, this analysis still provided crucial information about which vehicle attributes matter most. The decision tree confirmed that vehicle_type had a strong influence on sales behavior, reinforcing our behavioral segmentation approach.



G. Treemap for Segment Size Visualization

Finally, I used a treemap to visually represent the proportion of sales across different combinations of vehicle_category and vehicle_type. This powerful visualization offered an at-a-glance understanding of the biggest contributors to total EV sales. The chart clearly showed that segments such as delivery vehicles and commercial taxis dominate the landscape. This reinforced earlier conclusions and made the decision to focus on B2B and commercial vehicles even more convincing.



5. Selection of Target Segment

After profiling all the key segments extracted from the dataset, it became necessary to identify the most strategic and viable segment for a new EV startup to target. The selection is based on a combination of real sales trends, commercial potential, and ease of adoption — all interpreted from the data analysis.

While EV adoption is gradually expanding across various classes and types of vehicles, three segments clearly stood out as having the most promise for immediate and impactful market entry:

A. Target Segment Chosen: 3-Wheelers Used for Commercial Purposes

Among all the segments analyzed, 3-Wheelers used for delivery, cargo, and passenger transport emerged as the most dominant and promising segment. These vehicles are heavily used in both urban and semi-urban areas for last-mile logistics, small goods delivery, and public commuting.

They offer several advantages:

- Low upfront cost and low maintenance
- High usage frequency (ideal for EV ROI)
- Fast adoption rate in both Tier-1 and Tier-2 cities
- Already being embraced by fleet operators and individual drivers in commercial ecosystems

Their usage aligns perfectly with what EVs currently offer best: efficiency, cost-effectiveness, and utility. For a startup entering the EV space, this segment provides a low-risk, high-volume opportunity.

B. Why Not 4-Wheelers or Private Segments?

While 4-Wheelers and private users do show some level of EV interest, the sales data indicates a slower adoption curve. These segments tend to be more price-sensitive, dependent on infrastructure (like personal charging stations), and influenced by brand loyalty and aesthetics — areas where a new startup may struggle initially to compete.

Instead, the commercial and functional segments care primarily about:

- Total cost of ownership
- Battery range
- Charging time and availability
- Durability in frequent usage conditions

This makes them more rational buyers, focused on utility — and more willing to try emerging EV brands that meet their operational needs.

C. Final Justification for Segment Selection

Considering the insights from all three segment extractions — vehicle class, behavioral usage, and B2B orientation — the 3-Wheeler Commercial Vehicle Segment stands out as:

- The highest selling segment in the dataset
- Behaviorally aligned with cost-saving EV use-cases
- Least dependent on luxury features or brand recognition
- Actively scaling in urban delivery networks and public transport

These characteristics make it the most suitable target for an EV startup entering the Indian market.

6. Customizing the Marketing Mix

To build a successful entry strategy into the EV market, it’s not enough to just identify the right segment — we also need a complete Marketing Mix strategy (also known as the 4Ps) that aligns with the needs and behavior of the chosen audience.

Based on the insights from our data, we're targeting commercial users of 3-Wheelers, primarily operating in delivery, cargo, and public transport use-cases. These users are functional buyers who prioritize performance, cost-efficiency, and reliability.

Here's how the marketing mix would be designed to attract and serve this audience:

Product

The product should be a durable and high-efficiency 3-Wheeler EV, built for commercial usage. Key product features must include:

- Battery range of at least 100 km per charge to support delivery cycles
- Swappable battery technology (optional) for quick turnarounds
- Large cargo space or passenger comfort, depending on use-case
- Strong suspension and wheels to withstand Indian road conditions
- Real-time telemetry and GPS tracking for fleet managers
- Minimal maintenance design and fast after-sales support

This product isn't about luxury — it's about value, durability, and operational savings.

Price

Pricing should be competitive and clearly communicate long-term value. Since we're targeting small businesses and fleet operators, affordability is key:

- Base price should fall between ₹1.5 to ₹2.5 lakhs, depending on variant
- Offer financing options or lease-to-own schemes to lower entry barriers
- Highlight low running costs (₹0.80 per km vs ₹4+ for petrol/diesel)

By showcasing total cost of ownership (TCO) and return on investment, pricing becomes a strategic selling point rather than a barrier.

Place

The product should be made available through:

- Fleet channels and partnerships with delivery/logistics companies
- Micro-dealers and franchise outlets in urban hubs, especially around business centers
- Online presence for demo bookings, service scheduling, and financing queries
- Partner with fleet tech platforms (like ride-sharing or hyperlocal delivery apps)

Distribution should be focused on places where these vehicles operate, not just places where vehicles are typically sold.

Promotion

Promotional strategies should highlight practical benefits rather than emotional appeal:

- Testimonial marketing from early adopters (fleet drivers, delivery partners)
- Partnership campaigns with gig platforms like Zomato, Swiggy, Dunzo, etc.
- On-ground demos near commercial markets and transport hubs
- Incentives like free maintenance for 6 months, charging credits, or battery replacement discounts

Digital promotions can focus on ROI comparisons (EV vs petrol), operational efficiency, and government subsidies.

7. Potential Customer Base in the Early Market & Profit Estimation

After identifying the most promising segment — 3-Wheelers used for commercial purposes — the next logical step is to estimate the size of this opportunity and project the potential early-stage revenue or profit a startup might generate by entering this market.

This kind of rough, yet insightful estimation is essential for planning production, marketing budgets, and investment requirements.

Step 1: Estimating the Potential Customer Base

To estimate how many potential customers could exist in the early market, we rely on a combination of:

- Patterns from our EV sales data
- Practical assumptions about real-world adoption
- The business nature of the B2B customer segment

From the analysis:

- 3-Wheelers are the highest-selling EV category across the entire dataset.
- Commercial users (like delivery fleets, local transport operators, and cargo haulers) are early adopters.

Let's assume from the dataset that around 650,000+ EV units in this segment have already been sold across India cumulatively over the years (as per aggregation seen in the ev_sales_quantity for 3W commercial vehicles).

To define a reasonable early market opportunity, we'll assume the startup is targeting only a small but high-potential slice of this market — say 2% of active commercial 3W users in major cities.

Let's calculate that:

- If total active commercial 3W EVs = 650,000
- Targeting 2% in early stage = 13,000 potential customers

Step 2: Setting a Strategic Price Point

From our earlier Marketing Mix, we established that the pricing sweet spot for this product is ₹1.5 to ₹2.5 lakhs. For estimation, we'll go with an average price of ₹2,00,000 per unit.

Step 3: Estimating Revenue (Potential Profit)

Now let's estimate revenue generated if we sell to 13,000 early adopters:

 Estimated Early Sales Revenue
= 13,000 units × ₹2,00,000 per unit
= ₹260,00,00,000
= ₹260 crores

If an EV startup enters the market targeting commercial 3-Wheeler users (such as delivery operators and transport fleets) and captures just 2% of existing users, it can expect to generate ₹250–₹270 crores in early revenue. This makes the segment not just a practical target — but a high-value opportunity with serious scaling potential.

8. The Most Optimal Market Segments

After performing a comprehensive segmentation and profiling analysis of the Indian EV market using real-world sales data, a very clear picture emerges. The most optimal segment for a new EV startup to launch in is the commercial 3-Wheeler segment, primarily used for last-mile delivery, shared transport, and small-scale cargo operations.

This conclusion is drawn not from assumption or theory, but directly from the data:

- 3-Wheelers consistently ranked at the top in EV sales volume across all vehicle classes.
- Within behavioral segmentation, commercial and utility usage types dominated the dataset — showing that EVs are being adopted to meet business needs.
- The B2B filtering further confirmed that organizations, small businesses, and fleet operators are the key early adopters of electric vehicles — not private consumers.
- EV adoption is highest in use-cases where vehicles are used frequently and cost-efficiency matters, such as gig economy deliveries, public transport (rickshaws), and institutional fleets.

This market segment offers the following benefits to a startup:

- High demand and growth potential
- Practical product requirements (durability, range, charging support)
- A relatively lower barrier to entry compared to private or luxury EV markets
- A cost-driven customer base that responds to efficiency-focused messaging

Final Recommendation

A new EV startup should launch by building and marketing a commercial 3-Wheeler EV, optimized for businesses that need reliable, efficient, and affordable vehicles for everyday operations.

The focus should be:

- Functional product design
- Flexible pricing and financing
- Fleet-centric partnerships
- Marketing campaigns that emphasize cost savings and operational efficiency

This segment is not only the most active in terms of adoption, but also one of the most scalable due to the ongoing rise of e-commerce, logistics, and shared mobility in India.

Choosing this path gives the startup the best possible combination of market readiness, product-market fit, and revenue potential.

9. GitHub Link

https://github.com/HarshadaHPatil/Project-2.-EV-Market-Segmentation-in-India-Feynn-Labs

Dataset Link

https://github.com/HarshadaHPatil/Project-2.-EV-Market-Segmentation-in-India-Feynn-Labs./blob/main/EV_Dataset.csv

EV Performance Segmentation Using KMeans – Report

Name: Prem Modsing
GitHub link: EVMarketSegStatewise

EV Performance Segmentation:

Electric vehicles (EVs) vary widely in performance, driven by differences in acceleration and top speed. This project segments EVs into clusters based on these performance characteristics using the KMeans algorithm. By doing so, it enables better understanding of market tiers such as budget EVs, mid-range cars, and high-end performance models.

Market Analysis

1. Dataset Overview

The dataset contains detailed specs of electric cars with the following relevant features:

- AccelSec: Time (in seconds) to accelerate from 0 to 100 km/h
- TopSpeed_KmH: Vehicle's top speed in kilometres per hour
- Brand: Car manufacturer name

Initial Cleaning Steps:

- Removed rows with missing values
- Selected only numeric performance columns for clustering
- Final cleaned Data Frame: df_clean

2. Feature Scaling

Since acceleration and top speed are measured on different scales, we applied standardization:

```
scaler = StandardScaler()

scaled_data = scaler.fit_transform(df_clean[['AccelSec', 'TopSpeed_KmH']])
```

This step ensures fair clustering results.

3. Clustering Algorithm – KMeans

We used **KMeans** to segment the cars by performance.

Elbow Method (Fig 1)

Used to find the optimal number of clusters:

```
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k)
    kmeans.fit(scaled_data)
    inertia.append(kmeans.inertia_)
```

✅ **Optimal K chosen: 4**

```
kmeans = KMeans(n_clusters=4)
df_clean['Cluster'] = kmeans.fit_predict(scaled_data)
```

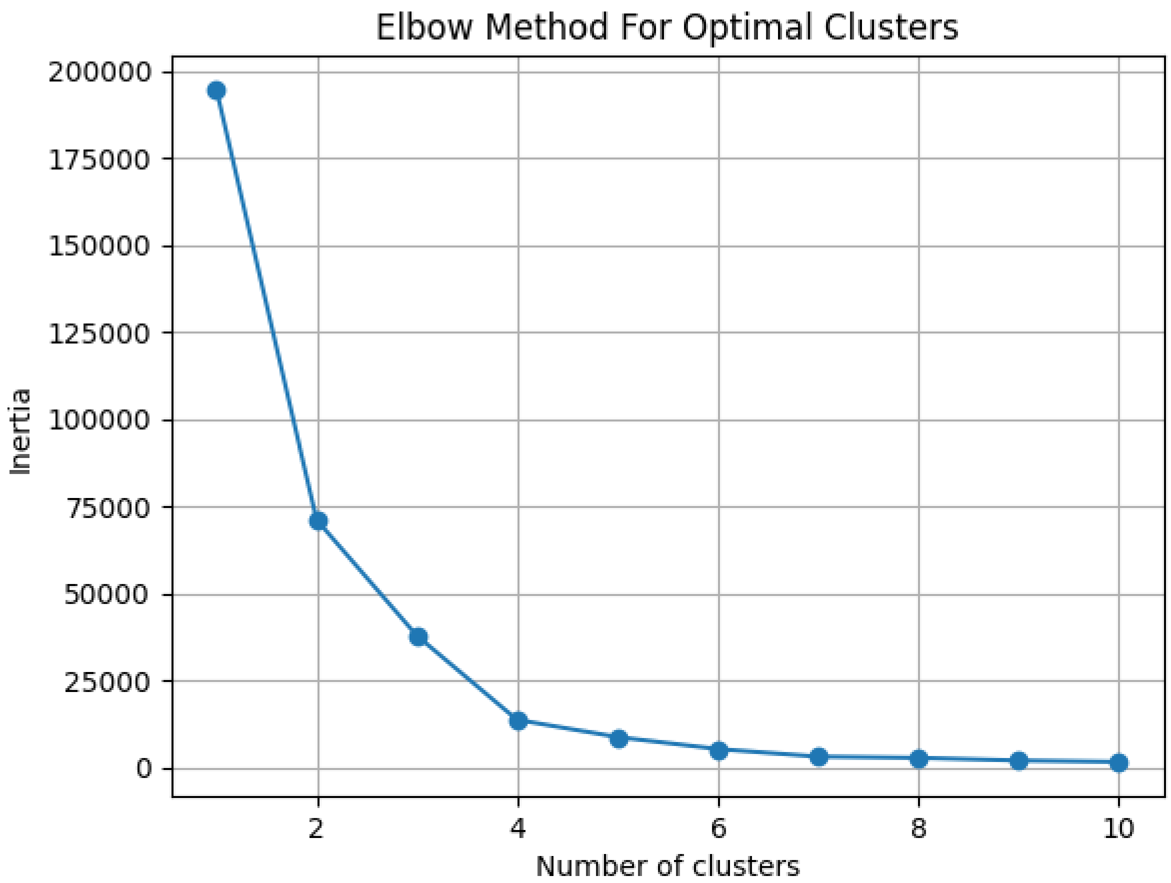


Fig 1: Elbow Curve to Find Optimal K

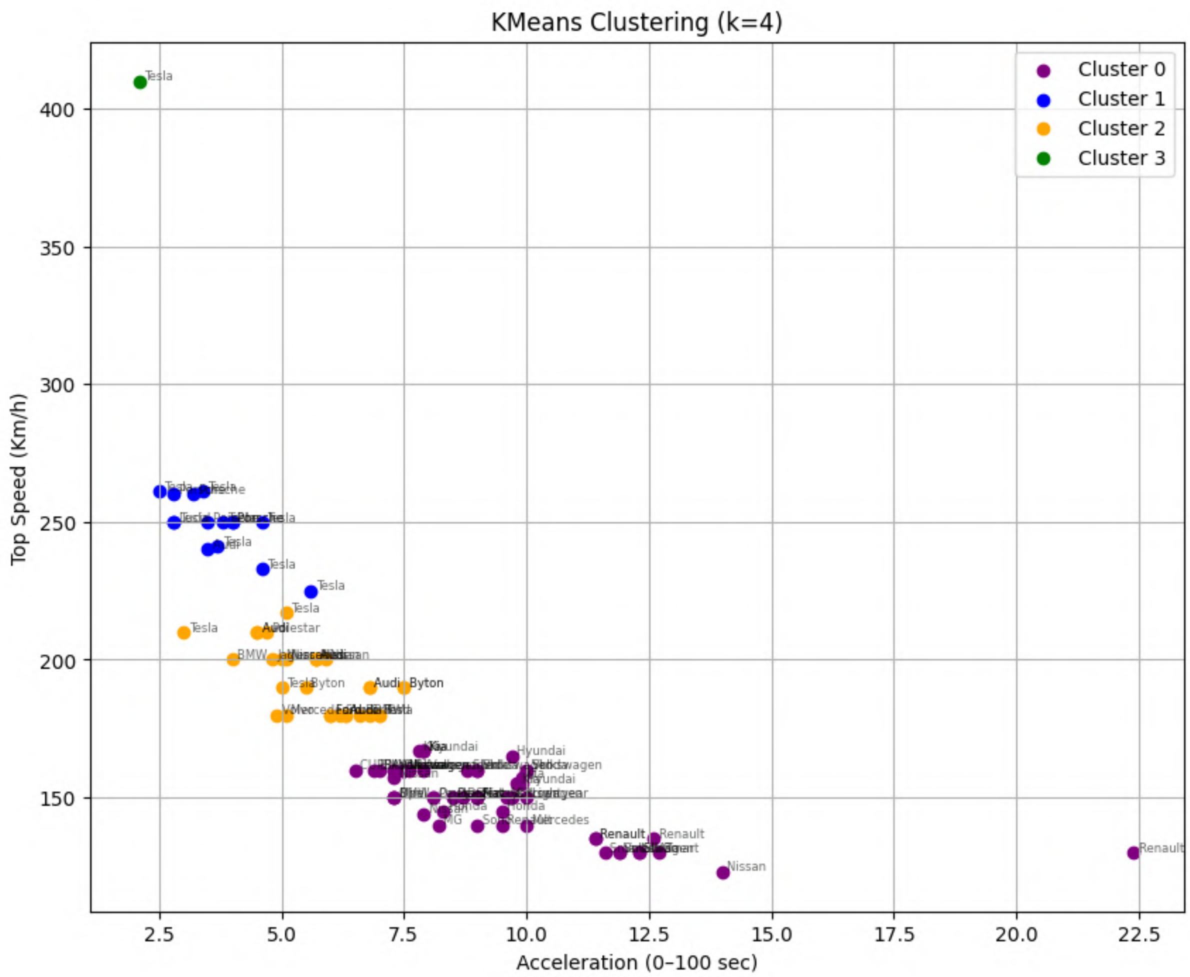


Fig 2: Cluster Visualization (Scatter Plot with Labels)

4. Cluster Interpretation

- Cluster 0 (🟣 Purple): Budget/city EVs – slower, lower top speed
- Cluster 1 (🔵 Blue): High-end performance EVs – quick & fast (e.g., Tesla, Porsche)
- Cluster 2 (🟠 Orange): Mid-range performers – balanced design (e.g., BMW, Audi)
- Cluster 3 (🟢 Green): Outlier/supercars – extremely fast (e.g., Tesla Roadster)

5. Hierarchical Clustering & Dendrogram

Hierarchical clustering was performed to complement the KMeans-based segmentation. This method provides a tree-like structure to visualize similarities between EVs based on their acceleration and top speed.

- The dendrogram was created using **Ward's linkage method**:

  ```
  linked = linkage(scaled_data, method='ward')
  ```

  ```
  dendrogram(linked, labels=df_clean['Brand'].values)
  ```

- Each branch of the dendrogram represents a **group of EVs with similar performance characteristics**.
- Useful for identifying **brand-level clusters** and **sub-segment relationships** not immediately obvious from flat clustering.

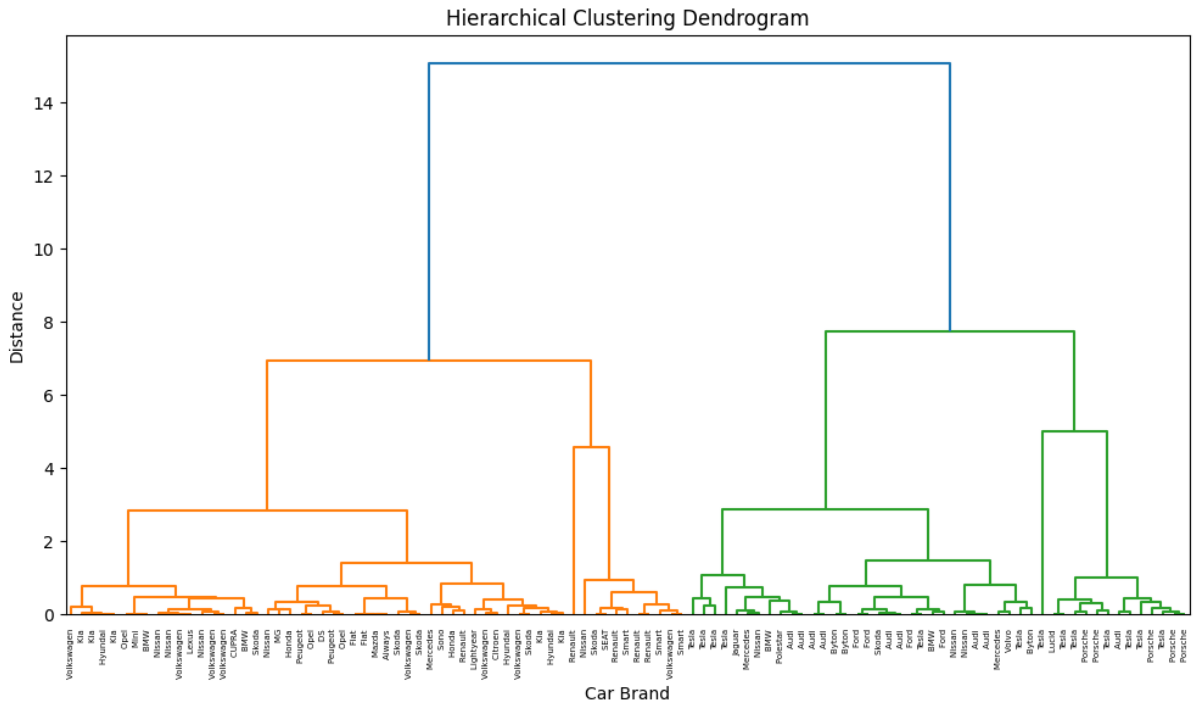


Fig 3: Dendrogram of Hierarchical Clustering

- Confirms KMeans segmentation with a visual grouping
- Highlights **performance similarities** among EV models
- Helps uncover **hierarchical relationships** (e.g., city cars vs. sports EVs)

Insights

- Performance clusters reveal natural segmentation within the EV market
- Helps manufacturers identify target markets and positioning strategies
- Outlier clusters point to niche or flagship vehicles
- Visualization simplifies explanation of complex specs to non-technical stakeholders
- Hierarchical clustering supports the KMeans-based segmentation
- Dendrogram reveals **multi-level similarity structure** among electric cars
- Useful for understanding both **macro** (cluster) and **micro** (brand-level) performance tiers

Conclusion:

- KMeans successfully segmented electric vehicles into 4 performance-based clusters.
- These clusters reflect real-world tiers such as city EVs, mid-range models, performance EVs, and outlier supercars.
- Strategic decisions in design, pricing, and marketing can be aligned with these performance segments.
- Hierarchical clustering provides additional validation by grouping similar EVs in a visual structure.
- EV performance varies significantly across brands, and such clustering helps identify key market differentiators.
- Combined, these methods offer a strong foundation for performance-driven EV market analysis.

EV Market Segmentation – Report

Name: Dhruvin Suthar
GitHub link: EVMarketSegStatewise

EV Market Segmentation:

India's electric vehicle (EV) market is rapidly evolving due to policy support and increasing environmental consciousness. The widespread adoption of EVs, backed by government subsidies and infrastructure development, is transforming the country's automotive landscape. This report segments the EV market in India using clustering algorithms based on state-wise EV sales data.

Market Analysis

1. Dataset Overview

The dataset included key columns such as:

- Year, Month_Name
- State, Vehicle_Category
- EV_Sales_Quantity

Initial cleaning steps:

- Removed rows with missing values in critical columns
- Resulted in a cleaned dataset: df_clean

2. State-Wise EV Sales Aggregation

- State-wise aggregation was done using:

```
state_sales = df.groupby('State')['EV_Sales_Quantity'].sum().reset_index()
```

- This allowed for understanding which states lead in EV adoption.

3. Feature Scaling

- Since sales numbers had large differences in scale, StandardScaler was applied:

```
scaled_sales = scaler.fit_transform(state_sales[['EV_Sales_Quantity']])
```

4. Clustering Algorithm – KMeans

- KMeans was used to segment Indian states based on total EV sales.
- Elbow Method (Fig 1) was used to determine the optimal number of clusters:

```
inertia = []
for k in range(1, 11):
    ...
```

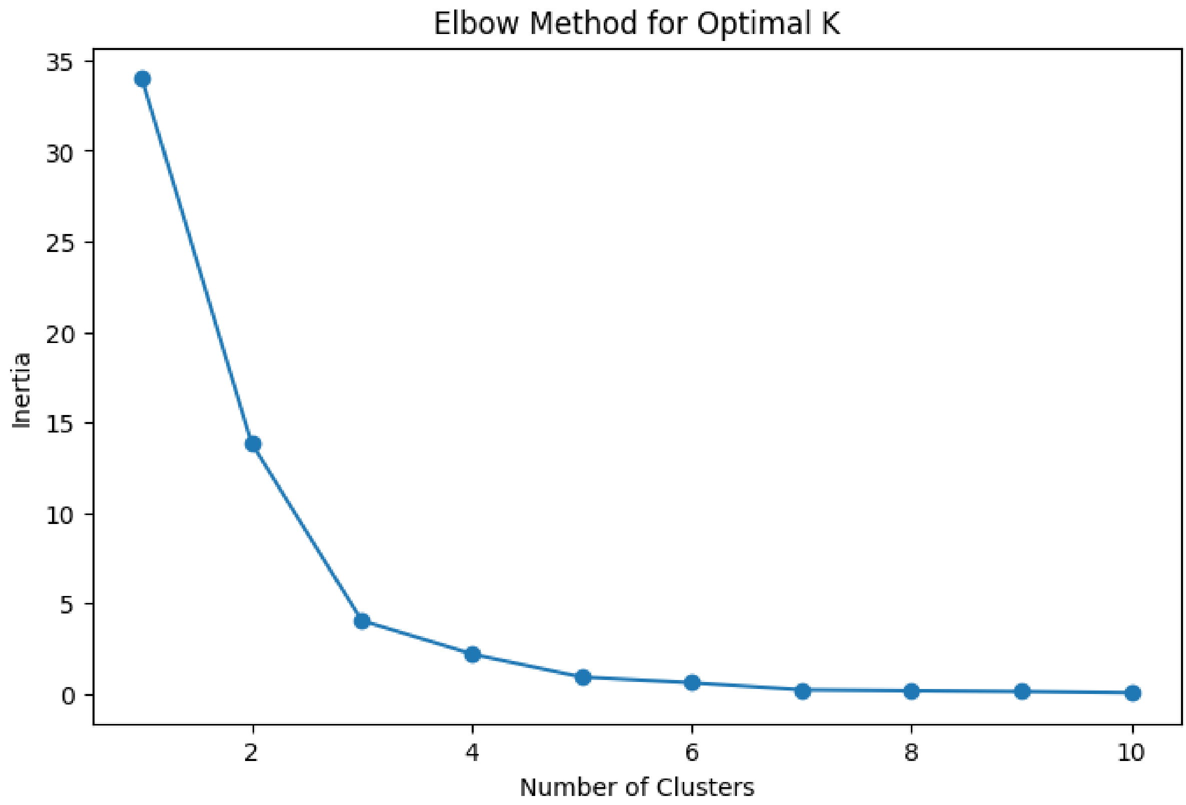


Fig 1: Elbow Method for Finding Optimal K

- Optimal K selected: **4**
- Final clustering applied:

```
kmeans = KMeans(n_clusters=4)
state_sales['Cluster'] = kmeans.fit_predict(scaled_sales)
```

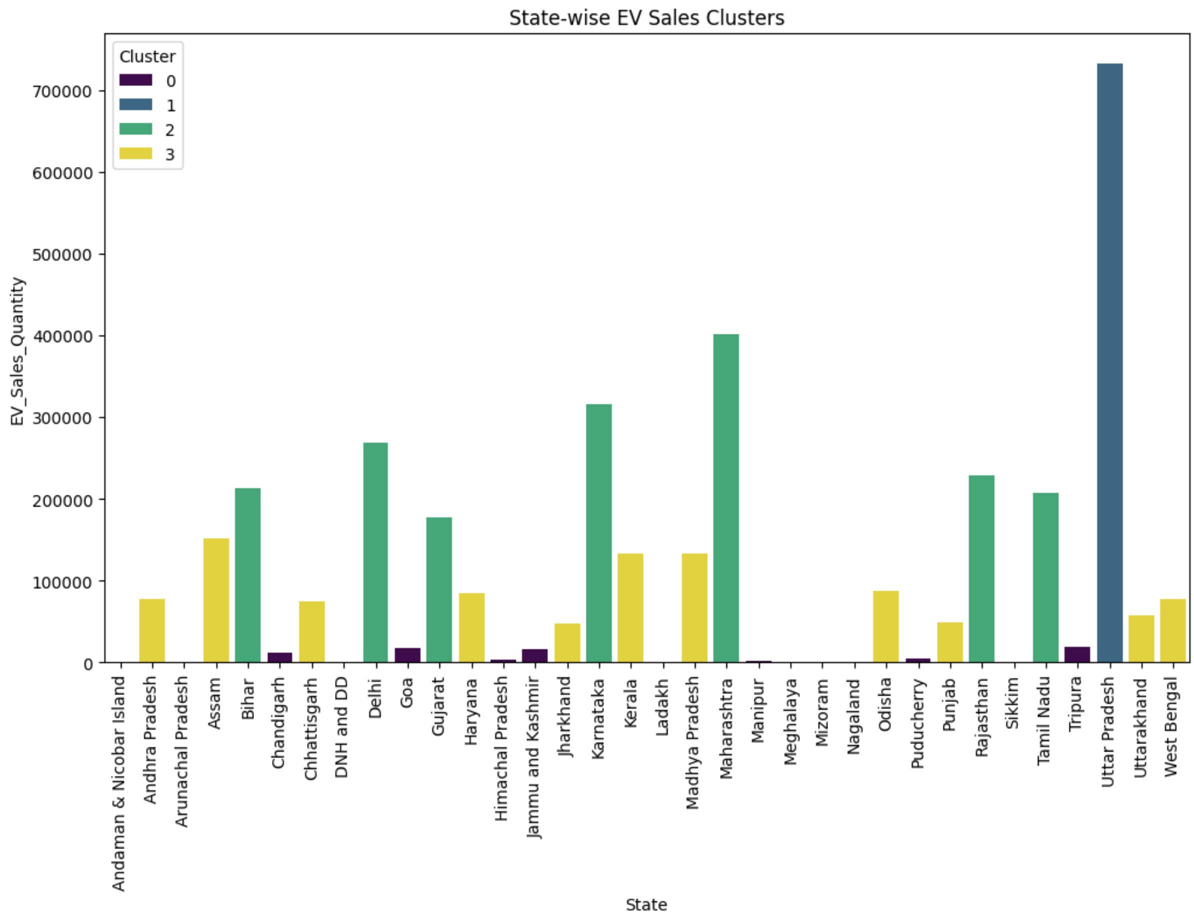


Fig 2: Bar Plot of EV Sales Clustered by State

- Clear segmentation shows which states have:
  - High EV sales (Cluster 0)
  - Moderate EV adoption (Cluster 1 & 2)
  - Emerging EV markets (Cluster 3)

5. Hierarchical Clustering & Dendrogram

- Dendrogram generated using scipy:

```linked = linkage(scaled_sales, method='ward')```

dendrogram(linked, ...)

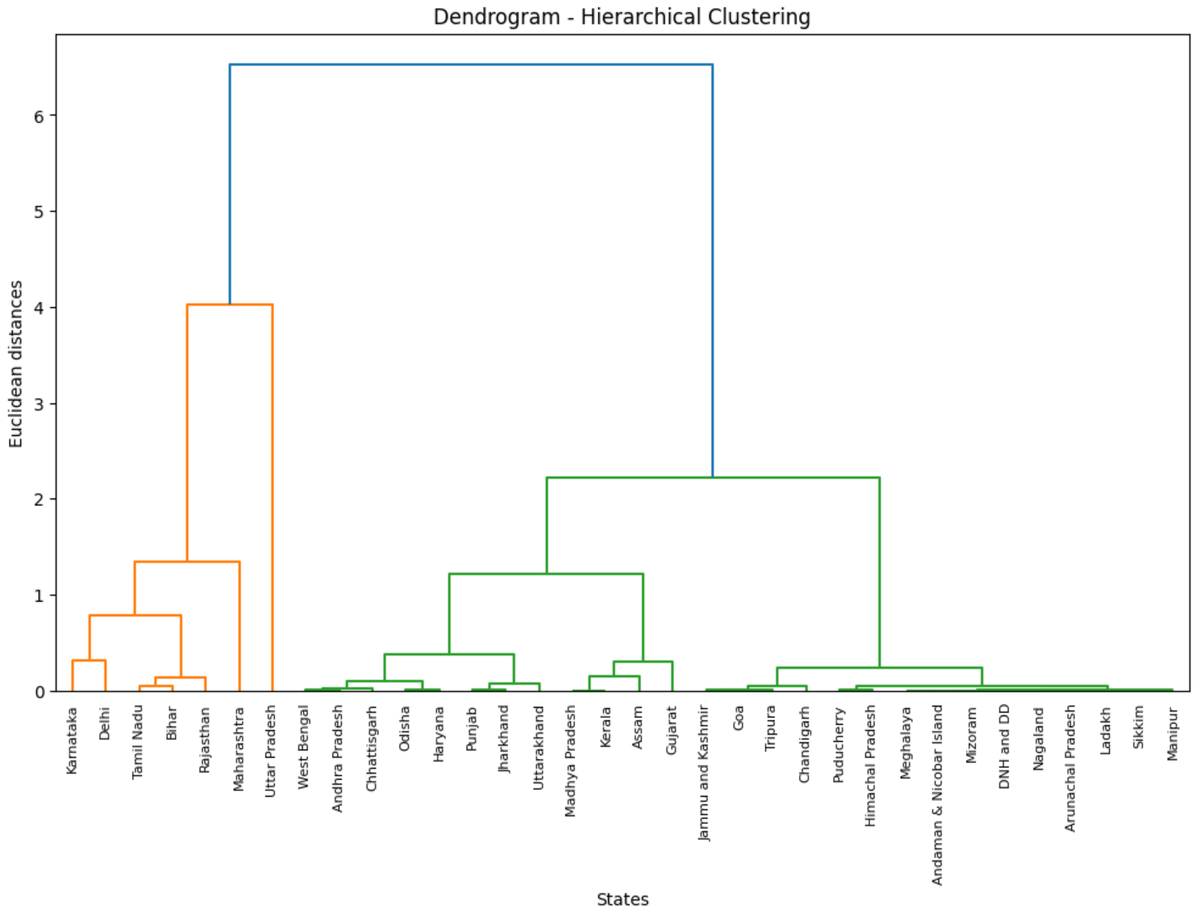


Fig 3: Dendrogram of Hierarchical Clustering

- States grouped by similarity in EV sales
- Offers insight for targeted infrastructure development
```

Insights

- **Top States by EV Sales:** Uttar Pradesh, Delhi, Maharashtra, Karnataka
- **Cluster 0 States:** Highly EV-ready states with developed infrastructure
- **Cluster 3 States:** Potential markets for EV investment
- **Hierarchical clustering confirms KMeans groupings with minor variations**

Conclusion:

- KMeans successfully segmented Indian states into 4 clusters based on EV sales.
- Strategic policy and infrastructure investment can be prioritized based on cluster results.
- Hierarchical clustering provides additional verification of segmentation.
- EV adoption is concentrated in a few regions targeted policies can help scale nationwide adoption.

Electric Vehicle (EV) Market Segmentation

Analysing the respective market in India using Segmentation analysis for Electric Vehicle (EV) Industry

By

Ayush Roy

Fermi Estimation

Goal: Determine which demographic or behavioral segments are most likely to purchase a product (binary classification), based on age, gender, and income.

To scope the market:

1. Assume a city of **1 million people**.
2. Approx. **20% (200,000)** are in the income bracket seen in the dataset (₹15K–₹1.5L/month).
3. Assume only **10%** are actively seeking to buy the type of product we're analyzing → **20,000 potential buyers**.
4. Out of these, data suggests ~**40% conversion rate → ~8,000 likely buyers.**
5. If our product costs ₹20,000:
   Estimated Early Profit = 8,000 × ₹20,000 = ₹16 crore

Data Source & Features

Dataset:

The dataset used for this analysis is a mock dataset named age_buynotbuy.csv, which includes the following columns:

- User ID: A unique identifier for each individual.
- Gender: The gender of the individual (Male/Female).
- Age: The individual's age (ranging from 18 to 63 years).
- Annual Salary: The salary of the individual in INR (₹15,000 to ₹1,52,500).
- Purchased: A binary flag (0/1) indicating whether the individual made a purchase (1 = Yes, 0 = No).

Feature Overview:

- **Demographic**: Age, Gender
- **Behavioral**: Purchase decision, Salary (influences purchase likelihood)

The dataset provides an excellent foundation to predict purchase behavior based on demographic and financial characteristics.

Summary Statistics

Feature	Mean	Std Dev	Min	25%	Median	75%	Max
Age	40.1	10.7	18	32	40	48	63
Annual Salary	72,689	34,488	15,000	46,375	72,000	90,000	152,500
Purchased	0.402	0.491	0	0	0	1	1

Correlation Analysis

- **Age vs. Purchase**
  - **Correlation coefficient**: **0.616**
  - **Interpretation**: There is a **moderately strong positive correlation** — older users tend to purchase more.
- **Salary vs. Purchase**
  - **Correlation coefficient**: **0.365**
  - **Interpretation**: There is a **moderate positive correlation** — users with higher income are more likely to purchase, though the effect is less than that of age.
- **Insights**
  - Age has a stronger influence on purchasing behavior compared to income.
  - Young users (under 30) and users with lower income (under 40,000) are less likely to purchase.

Data Preprocessing

Data preprocessing is a crucial step in ensuring that the data is clean, consistent, and ready for analysis. We employed the following preprocessing techniques:

- **Missing Values**: We performed an null check using df.isnull().sum() to ensure no missing data, as the dataset is well-structured and complete.
- **Label Encoding**: Since Gender is a categorical variable, we applied Label Encoding to convert it into a numerical format for machine learning algorithms. Female = 0, Male = 1.
- **Feature Scaling**: Features like Age and Annual Salary were scaled using StandardScaler. This ensures that these features are on the same scale, allowing the machine learning model to perform better and faster.
- **Splitting Data**: The data was split into training and testing sets (80% training, 20% testing) using train_test_split to ensure proper model validation. This step was critical in cleaning the data for analysis, ensuring no biases or inaccuracies from unscaled or missing data.

Segment Extraction & Visualization

Exploratory Data Analysis (EDA) was performed to visualize and understand the distribution of Age, Salary, and Purchase behavior across different segments.

1. Age Distribution

The Age Distribution plot shows the frequency of individuals across different age groups. The plot reveals that the majority of consumers fall within the age range of 30–50 years, which corresponds to a higher likelihood of purchase decisions.

2. Salary Distribution

The Annual Salary Distribution shows that most individuals fall within the ₹30,000 to ₹1,00,000 range, which corresponds to middle-income earners. Higher salary brackets, especially above ₹1,00,000, show a sharp decrease in the number of individuals. This suggests that the product has stronger appeal to mid-income consumers.

3. Gender vs Purchase Decision

This countplot shows the distribution of Gender against Purchase behavior. While the chart is nearly balanced, females show a slightly higher purchase rate compared to males.

However, the difference is not substantial, indicating gender may not be the strongest predictor of purchase in this context.

4. Age vs Salary by Purchase

This scatterplot shows how Age and Salary correlate with the purchase decision. We observe that the most frequent purchases happen among individuals aged 30–50 with salaries between ₹60,000–₹1,20,000. These individuals represent the most ideal target group for products aimed at mid-career professionals.

5. Salary by Purchase

The boxplot illustrates the spread of Annual Salary based on whether the individual made a purchase. The boxplot clearly shows that purchasers tend to have a higher salary compared to non-purchasers, with a noticeable income gap. This aligns with our assumption that individuals with higher disposable income are more likely to make purchases.

Logistic Regression Model

We trained a Logistic Regression model to predict whether a customer will purchase based on their Age, Salary, and Gender.

Key Findings:

- **Accuracy**: The model achieves 79.5% accuracy on the test set, indicating decent predictive performance.
- **Confusion Matrix**:

[[104, 8],

[33, 55]]

The model is more accurate in predicting non-purchasers (True Negative), with a higher recall for predicting non-buyers (92.9%).

Classification Report:

Class	Precision	Recall	F1-Score
Non-purchase (0)	75.9%	92.9%	83.5%
Purchase (1)	87.3%	62.5%	72.8%

Segment Profiling

Based on the visualizations and model output, we identified several key segments:

- **Age:** 30–50 years old are the primary buyers.
- **Salary**: Individuals with ₹60,000+ annual salary are more likely to make purchases.
- **Gender**: No significant gender disparity in purchasing.

The best segment for targeting includes middle-aged, mid-to-high income professionals.

Target Segment

The optimal target segment for the product is individuals who are:

- Aged 30–50 years.
- Earning between ₹60,000–₹1,20,000/month
- Likely to make purchase decisions based on value and necessity, especially if they are mid-career professionals.

Marketing Mix (4Ps)

To reach this segment, the marketing mix is tailored as follows:

- **Product**:
    - Focus on functional, affordable, and reliable offerings.
    - Products should offer value over luxury, emphasizing cost-effectiveness, durability, and ease of use.
- **Price**:
    - ₹15,000–₹25,000 range
    - Financing options like EMIs should be offered to ease purchase decisions for middle-income consumers.
- **Place**:
    - Focus distribution in urban centers with strong online presence and retail hubs.

- **Promotion**:
  - Use ROI-based messaging.
  - Digital ads on platforms like LinkedIn and Google, specifically targeting professionals.

Market Opportunity

- Target Market Size: ~8,000 active buyers.
- Price Point: ₹20,000
- Potential Revenue: ₹160,000,000 (~₹16 Crores)

Most Optimal Segment

The 30–50 years old, mid-to-high-income professionals (₹60,000–₹1,20,000/month) represent the optimal target market. They are ideal for reaching with value-driven, practical products that enhance their lifestyle.

Github:

https://github.com/ayushroy-117/EV-Market-Data-Analysis/blob/main/Notebook.ipynb