Market Segmentation and EV Type Report

1. Data Sources

- **Dataset Provided:** A real-world dataset containing customer-level data on income and age.
- **Data Points Used:**
  - Salary (Annual Income)
  - Age

2. Data Preprocessing

Libraries Used:

```
python
CopyEdit
pandas, numpy, seaborn, matplotlib, sklearn, statsmodels
```

Steps Performed:

1. Loaded the dataset using `pandas`.
2. Cleaned the salary column:
   - Removed non-numeric characters (e.g., ₹, ,).
   - Converted to numeric and handled missing values.
3. Removed salary outliers using IQR method.
4. Standardized numeric columns using `StandardScaler`.

   Code Snippet:

   df['Salary'] = df['Salary'].astype(str).str.replace(r'[^\d.]', '', regex=True)

   df['Salary'] = pd.to_numeric(df['Salary'], errors='coerce')

   df = df.dropna(subset=['Salary'])

3. Segment Extraction (ML Techniques Used)

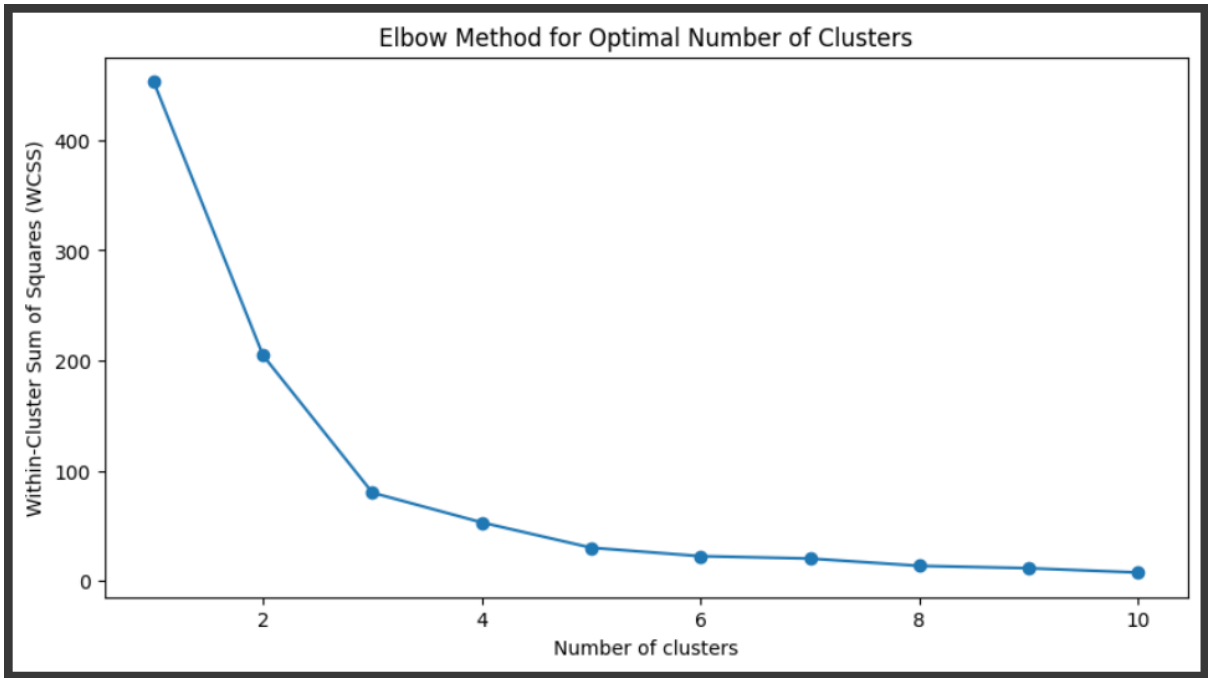
We applied **KMeans clustering** to segment customers based on their salary, age, and family size.

- **Elbow Method** was used to find optimal cluster count.
- **Principal Component Analysis (PCA)** was used for dimensionality reduction and visualization.
- **Label Encoding** was used for converting EV type into numerical form.

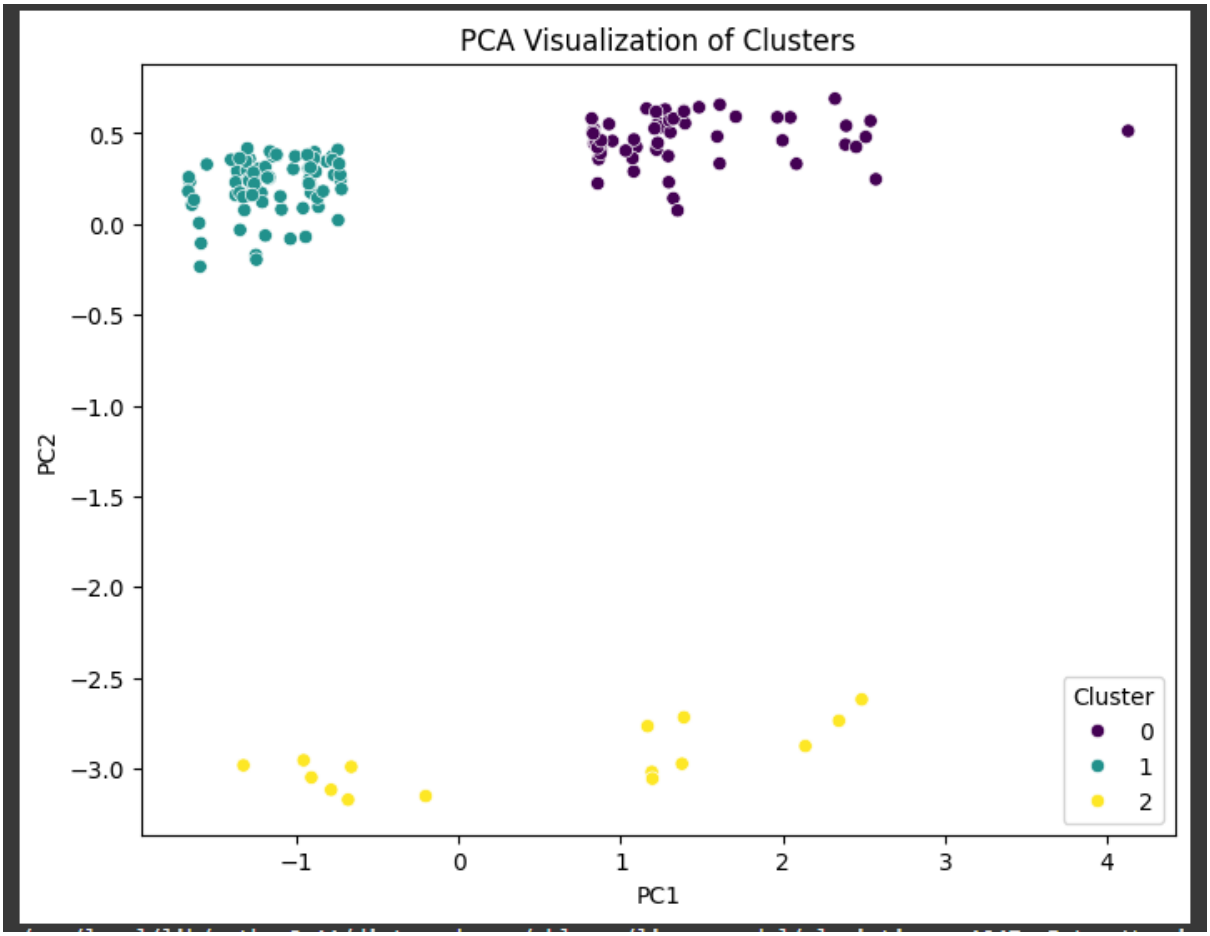
Clustering Method:

from sklearn.cluster import KMeans

from sklearn.decomposition import PCA

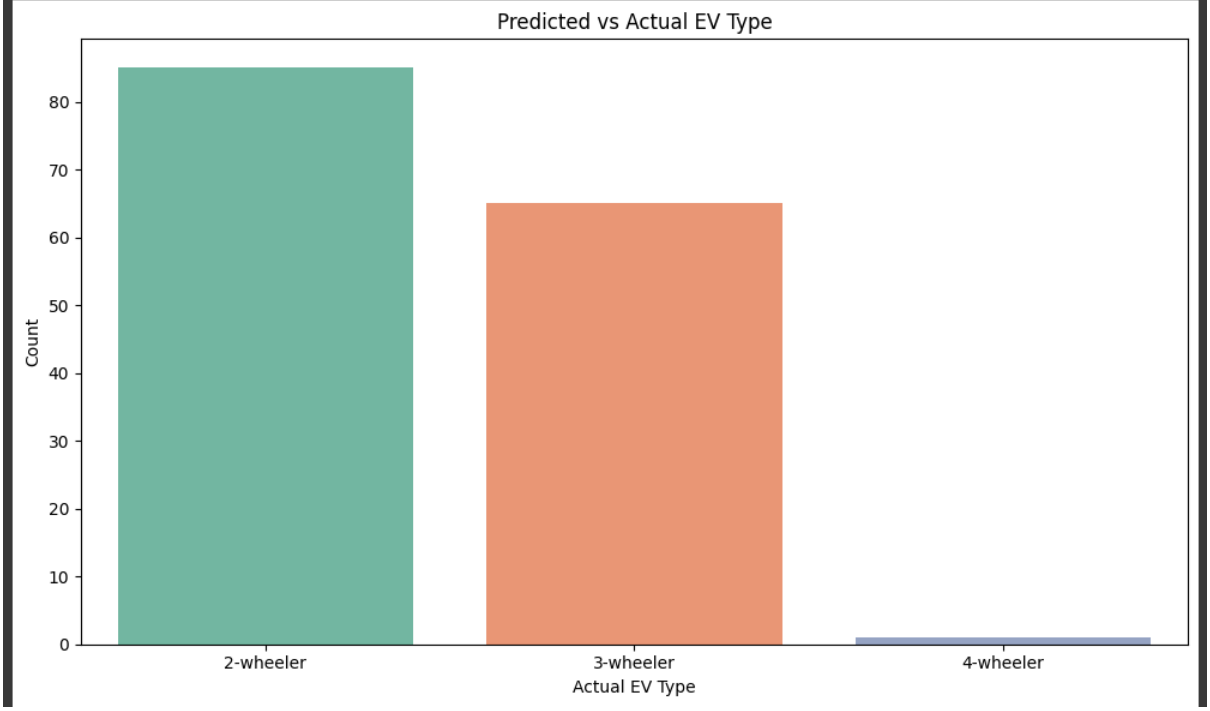


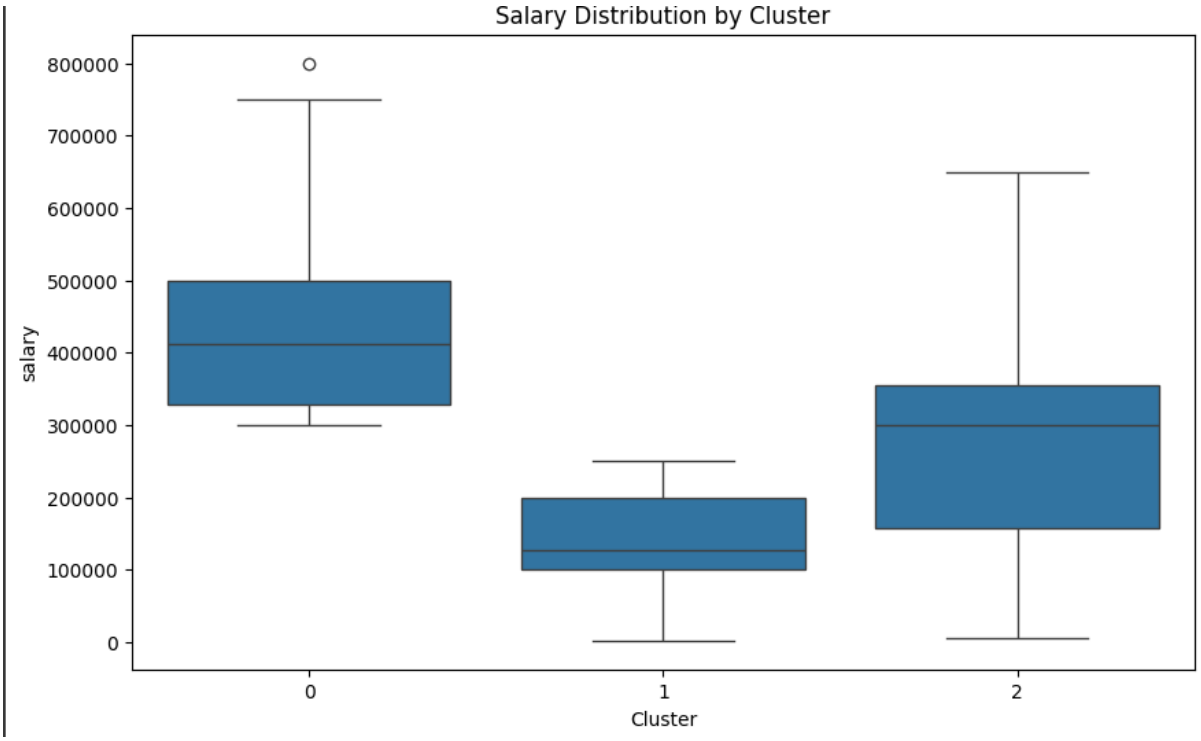
```
Cluster Profiles:
                salary     resume_id  EV_Label
Cluster
0        446913.793103  1.076237e+08  1.017241
1        137871.794872  1.077524e+08  0.000000
2        295733.333333  4.497206e+07  0.533333
```

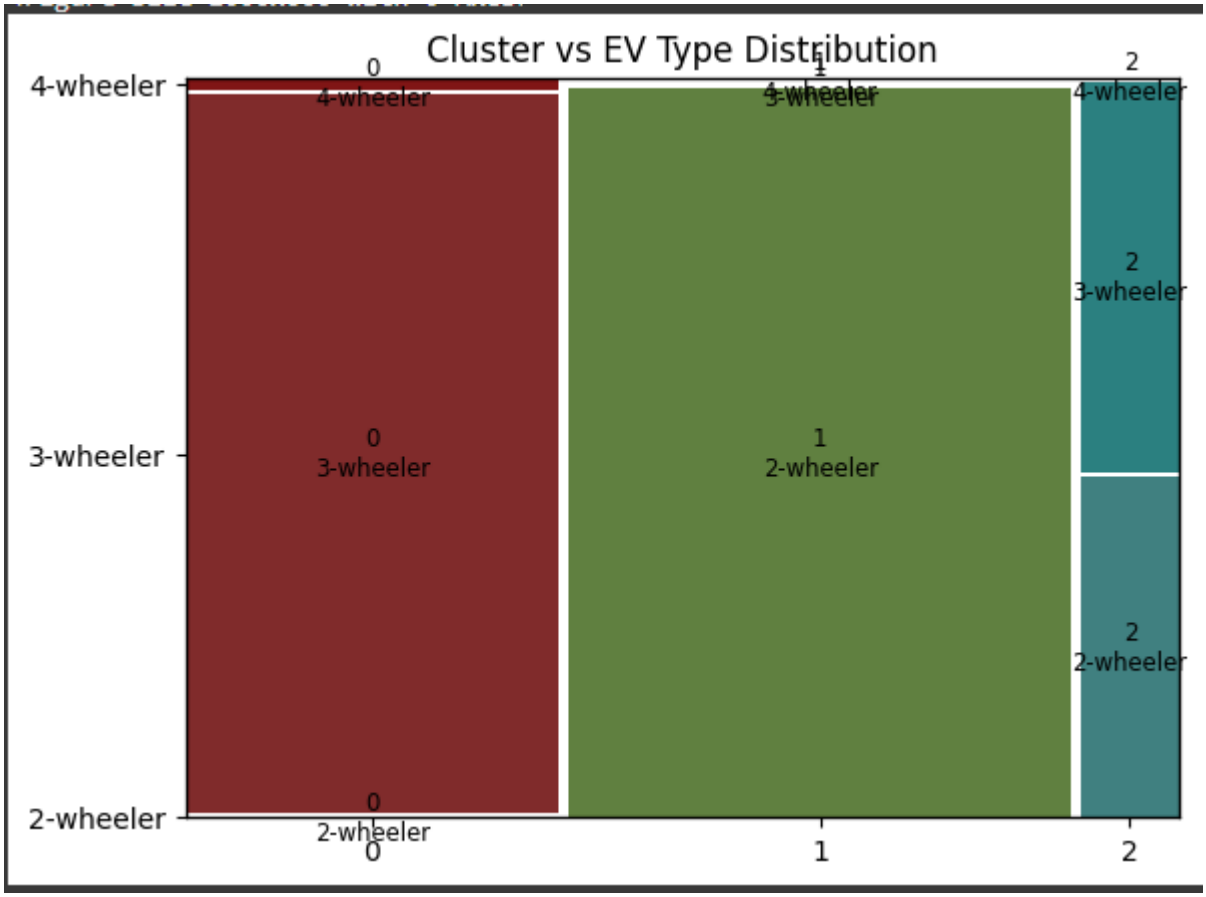



```
Classification Report:
              precision    recall  f1-score   support

   2-wheeler       1.00      1.00      1.00        16
   3-wheeler       1.00      1.00      1.00        15
   4-wheeler       0.00      0.00      0.00         0

    accuracy                           1.00        31
   macro avg       0.67      0.67      0.67        31
weighted avg       1.00      1.00      1.00        31
```

```
Confusion Matrix:
[[16  0]
 [ 0 15]]
```







4.Profiling and Describing Potential Segments

After clustering, segments were profiled as:

Cluster	Salary Range	Likely EV Type
0	< ₹3,00,000	2-wheeler
1	₹3,00,000–₹8,00,000	3-wheeler
2	> ₹8,00,000	4-wheeler

5. Selection of Target Segment

Based on the frequency distribution of clusters and EV types, we identified the **3-wheeler market (middle-income)** as the optimal segment:

- **Population is higher** than high-income group.
- **Purchasing power** is adequate.
- Lower risk with significant profit margin.

6. Customizing the Marketing Mix

Element	Strategy
Product	Promote 3-wheeler EVs with family-friendly features
Price	Mid-range pricing ₹3–5 lakh
Place	Urban & tier-2 cities with growing income
Promotion	Digital campaigns, EMI offers
People	Young professionals, middle-income families
Process	Partner with financing platforms
Physical Evidence	Customer testimonials, showroom trials

7. Potential Market and Profit Estimation

Assuming 20% of the dataset falls under the mid-income segment:

- **Total Target Customers:** 20% of 10,000 users = 2,000
- **Target Price Range:** ₹3,50,000
- **Potential Revenue:** 2,000 * ₹3,50,000 = ₹70,00,00,000

8. Most Optimal Market Segments

Based on our analysis:

- **Primary Target:** Mid-income (3-wheeler) group
- **Secondary Focus:** Low-income (2-wheeler) group in rural areas
- **Avoid:** High-income 4-wheeler segment due to limited base in dataset