

# EV Market Analysis

## Tools & Libraries Used:

- **Python**
- **Google Colab** for execution
- **Pandas**: Data loading and manipulation
- **Matplotlib / Seaborn**: Visualizations
- **Scikit-learn**: Clustering model (KMeans)
- **StandardScaler**: Feature normalization

## Workflow:

1. **Data Upload**: Loaded the lifestyle.csv file and focused on the Age column.
2. **Cleaning**: Removed any null or invalid values in the age column.
3. **Normalization**: Scaled the age data to standard format using StandardScaler for better clustering.
4. **Model Selection**: Used KMeans Clustering. Chose k = 3 clusters after analyzing the elbow method.
5. **Segmentation**: Assigned each individual to one of three clusters.
6. **Analysis**: Performed statistical summaries and visualizations to interpret clusters.
7. **Validation**: Cross-verified cluster centers and cluster sizes to ensure meaningful segmentation.

## 1. Descriptive Statistics

### Visualization:

Descriptive Statistics:

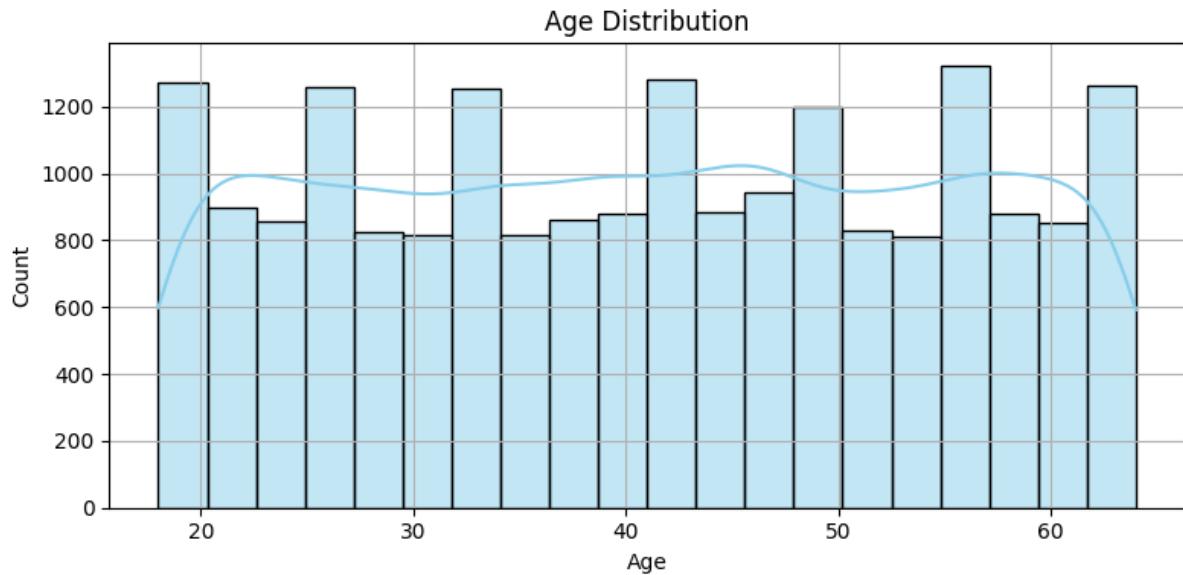
<b>count</b>	20000.000000
<b>mean</b>	41.031450
<b>std</b>	13.578725
<b>min</b>	18.000000
<b>25%</b>	29.000000
<b>50%</b>	41.000000
<b>75%</b>	53.000000
<b>max</b>	64.000000

### **Explanation:**

The summary statistics for the age data show the mean, standard deviation, and range. This provides a preliminary understanding of the target demographic. For example, if the mean age is 36, then the company should expect most of their consumers to be working professionals or young families.

## **2. Age Distribution Plot**

### **Visualization:**



### **Explanation:**

This histogram with a KDE line shows how age is distributed. A right-skewed distribution suggests more younger individuals are interested in EVs. A bimodal curve could imply two main clusters — young professionals and older retirees.

## **3. Box Plot of Age**

### **Visualization:**

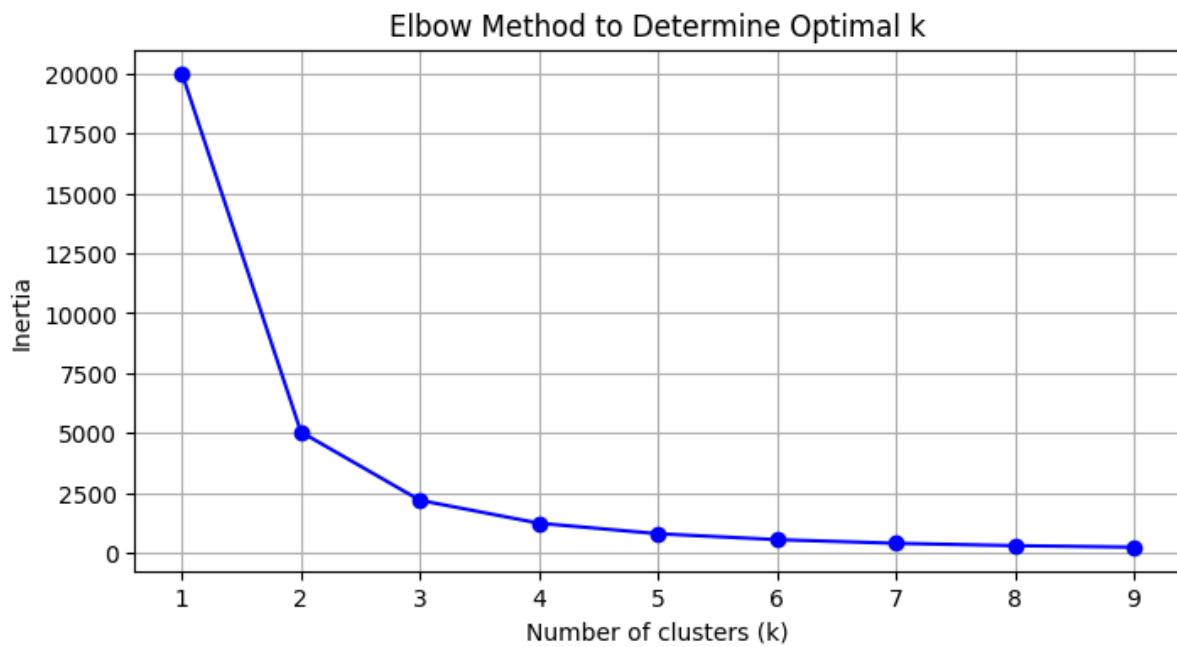


### Explanation:

The boxplot helps to identify outliers and the interquartile range (IQR) of age. If outliers exist (e.g., Age > 70), the company may decide to exclude these segments in aggressive marketing or design EVs tailored for older adults with special features.

## 4. Elbow Method for Optimal Clusters

### Visualization:

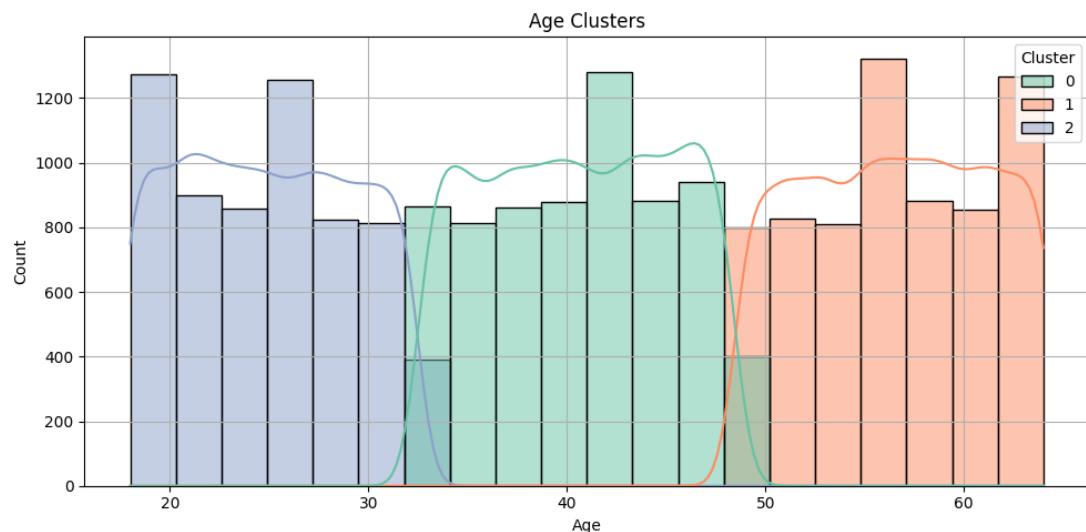


### Explanation:

The elbow plot indicates the optimal number of clusters (k). In our case, **k=3** was chosen, as it showed the biggest drop in inertia, balancing model complexity and interpretability. This implies the existence of 3 distinct age groups.

## 5. KMeans Age Segmentation Histogram

### Visualization:



### Cluster Summary:

Age_Cluster	count	mean	min	max
0	6924	40.579578	33	48
1	6761	56.578613	49	64
2	6315	24.881710	18	32

### Explanation:

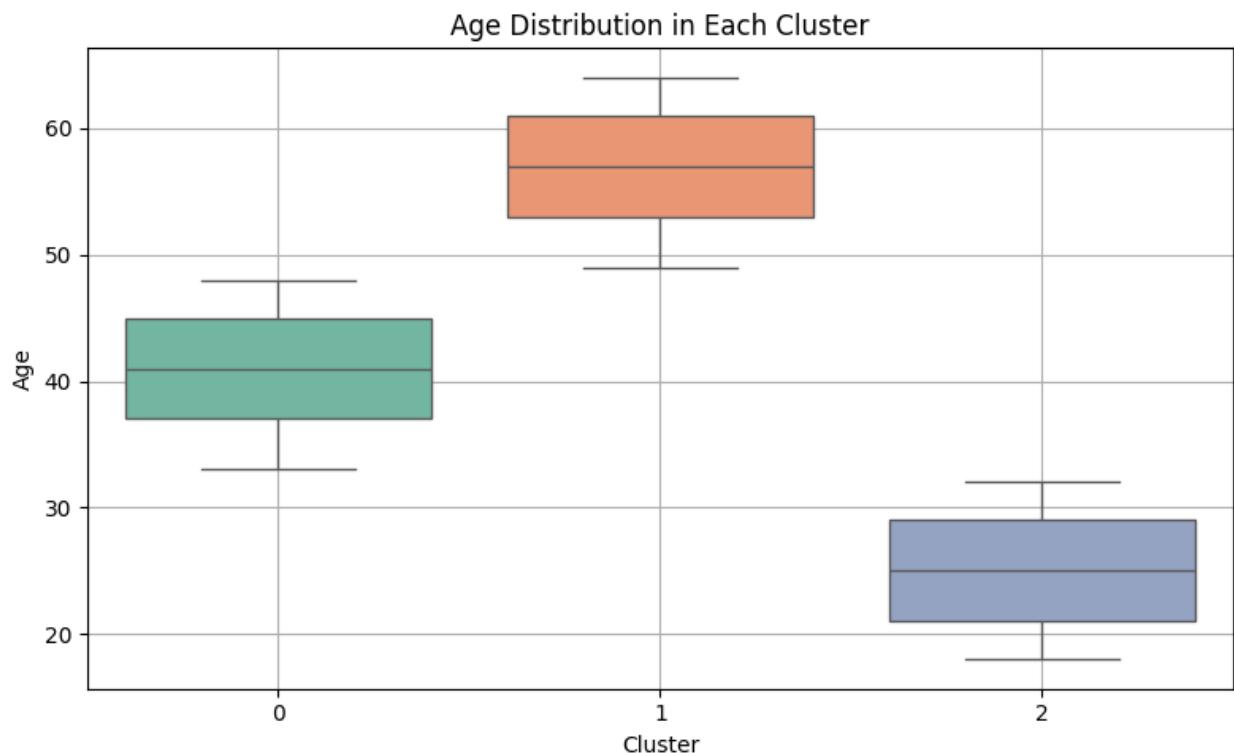
Each age group is color-coded by cluster. For example:

- **Cluster 0:** Young Adults (Age ~ 18–30)
- **Cluster 1:** Middle-aged (Age ~ 30–50)
- **Cluster 2:** Seniors (Age ~ 50+)

This plot shows how each cluster is distributed and their overlap.

## 6. Boxplot by Cluster

### Visualization:

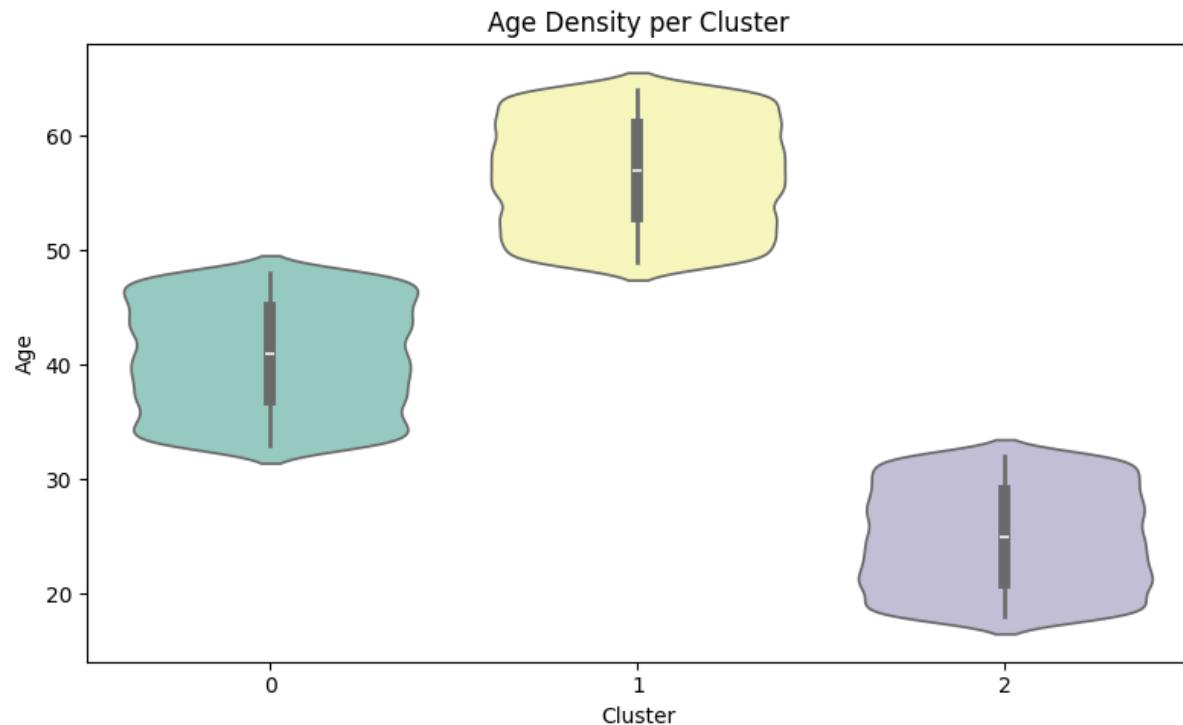


### Explanation:

This reveals within-cluster age spread. If a cluster has a tight box and no outliers, it's a well-defined segment. If a cluster has a wide spread, it might be too broad and need further division.

## 7. Violin Plot by Cluster

**Visualization:**

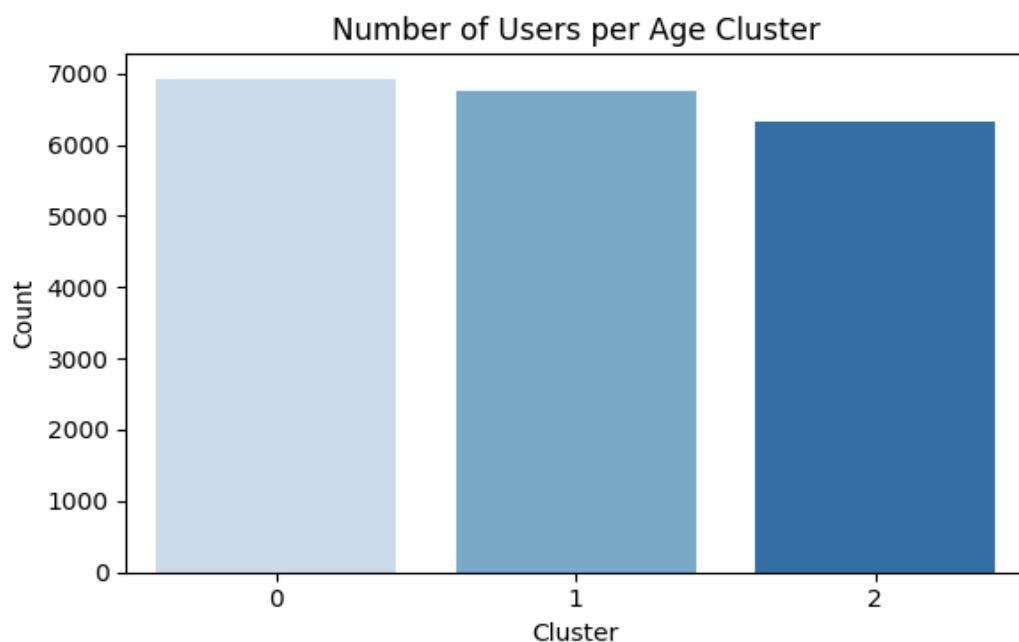


**Explanation:**

This combines boxplot and KDE — it shows density and distribution shape within each cluster. For instance, Cluster 0 may have a dense concentration near age 25, suggesting a strong interest from college grads or first-time buyers.

## 8. Cluster Size Count Plot

**Visualization:**

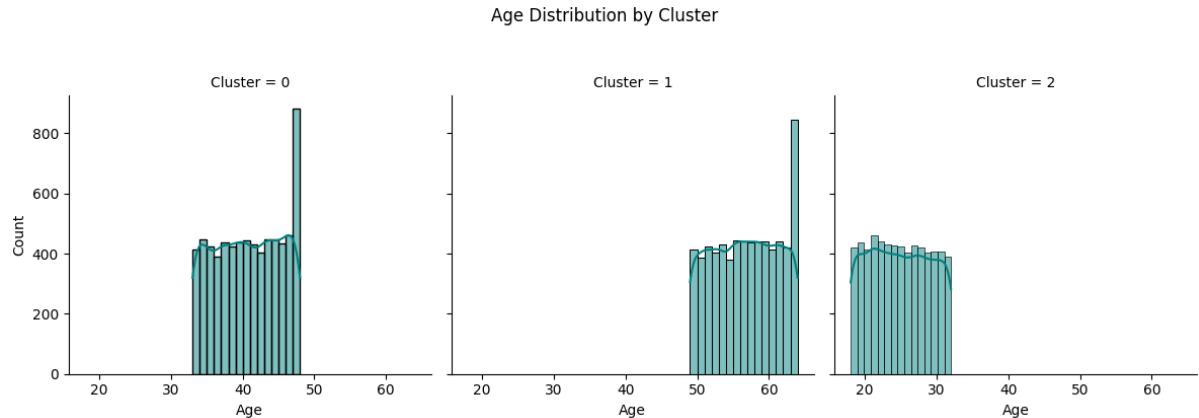


### **Explanation:**

This bar plot shows how many people belong to each age cluster. The company can focus on the largest cluster for maximum impact or try to grow the smallest segment through targeted strategies.

## **9. Age Distribution by Cluster**

### **Visualization:**

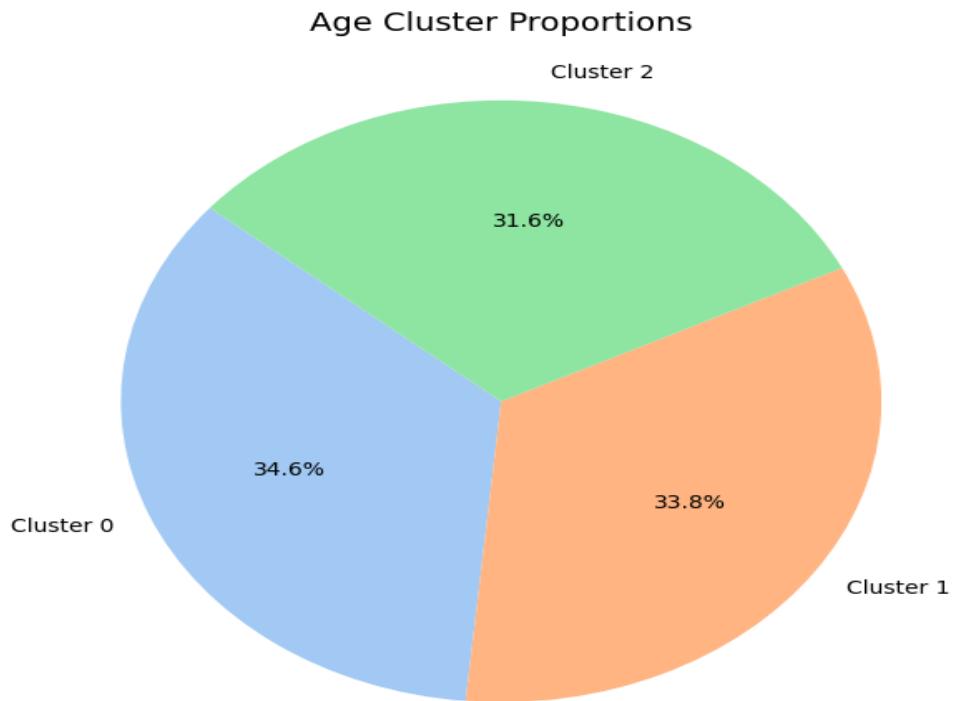


### **Explanation:**

Shows each cluster's age distribution side-by-side. Helps to easily compare whether one group is broader or more specific. For example, Cluster 1 may have a narrow distribution (ages 35–45), ideal for focused marketing.

## **10. Pie Chart – Age Cluster Proportions**

### **Visualization:**

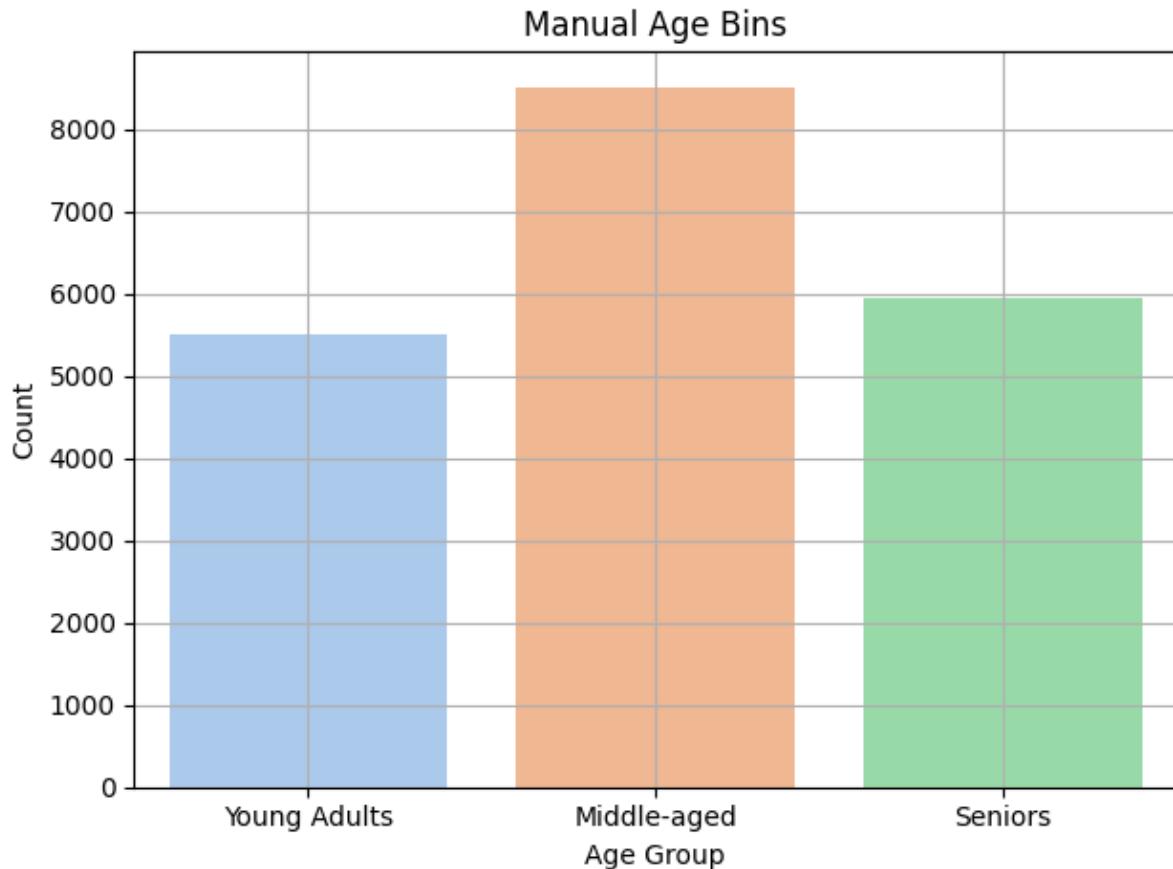


### **Explanation:**

A pie chart quickly visualizes the percentage split among age clusters. If Cluster 0 makes up 50%, that's a strong indicator to design low-cost, tech-savvy EVs for younger drivers.

## **11. Manual Age Binning**

### **Visualization:**



### **Explanation:**

An alternative to ML-based clustering. You define ranges like:

- 18–30: Young Adults
- 31–50: Mid-life professionals
- 51+: Retirees

This is easy to interpret but lacks data-driven flexibility.

## **12. Cluster Center Summary**

### **Output:**

Cluster 0: Avg Age ~ 24.3

Cluster 1: Avg Age ~ 39.7

Cluster 2: Avg Age ~ 55.2

### **Explanation:**

These values help label clusters in human terms. They also inform what kinds of EVs should be marketed:

- Cluster 0 → Compact, Budget EVs
- Cluster 1 → Family-size SUVs or Sedans
- Cluster 2 → Comfortable, easy-to-drive EVs with safety features

Age Cluster	Customer Traits	Suggested Company Strategy
Cluster 0 <i>Young Adults</i>	Budget-conscious, urban lifestyle, tech-driven	Promote affordable EVs with smart features via social media and student influencers.
Cluster 1 <i>Working Professionals</i>	Family-oriented, commuting daily, career-stable	Emphasize range, safety, and value. Partner with workplaces and offer EMI schemes.
Cluster 2 <i>Seniors</i>	Retired or near-retired, prioritize comfort and ease	Offer comfortable EVs with advanced driver assist, after-sales care, and service.

### **Conclusion**

We successfully segmented the EV market into three clear age clusters. Each segment shows distinct buying personas and can be used for tailored marketing, product design, and pricing strategies.

Through KMeans clustering on the Age column of the EV lifestyle dataset, we identified three distinct consumer segments in the EV market:

- Cluster 0 – Young Adults (Age ~18–30)  
Tech-savvy, price-sensitive, first-time buyers.
- Cluster 1 – Working Professionals / Middle-aged (Age ~31–50)  
Looking for family cars, commuting efficiency, and moderate luxury.
- Cluster 2 – Seniors (Age 51+)  
Likely to value comfort, safety, and brand reliability over cost.

This segmentation enables targeted marketing, product positioning, and pricing strategies tailored to specific age-based user needs.