Market Segmentation
EV Market
Mohd Shahvez
22-8-2024

Dataset Description:

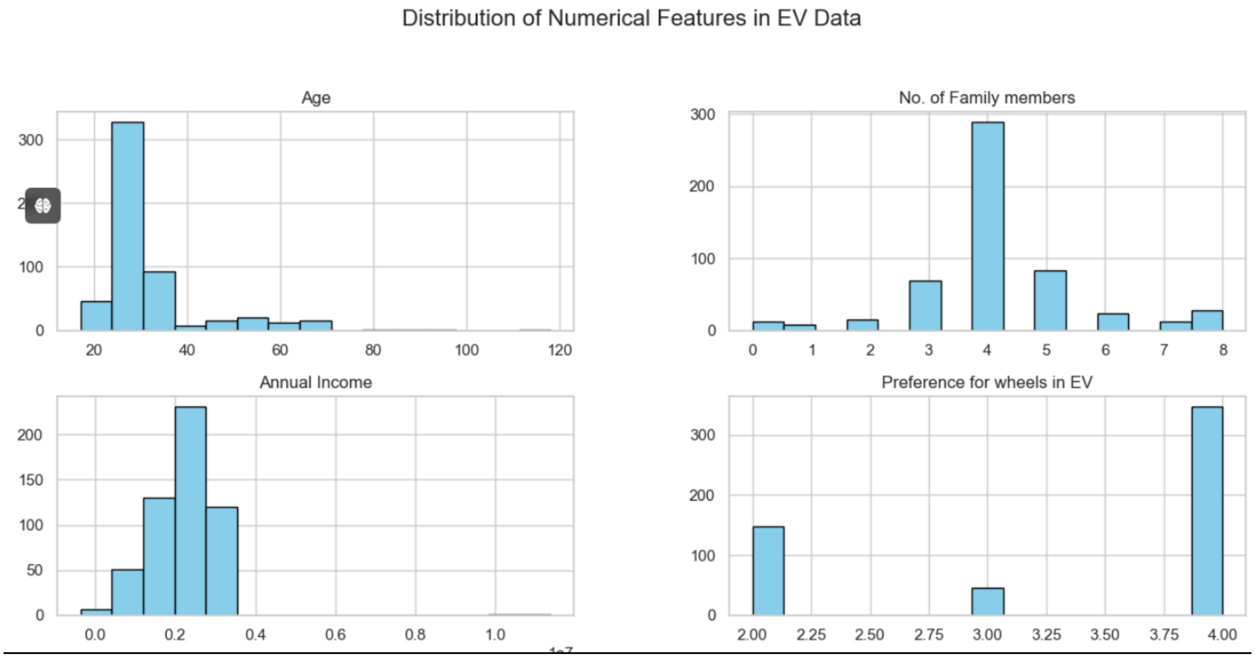
The dataset, `df1`, contains responses from 540 individuals regarding their demographic, psychographic, and behavioral preferences, particularly focused on their views about adopting Electric Vehicles (EVs). It includes a mix of numerical and categorical variables, representing a wide array of factors that could influence EV adoption. Here's a breakdown of the key columns:

1. **Age**: Numerical data indicating the age of the respondents.
2. **City**: Categorical data indicating the city of residence of the respondents.
3. **Profession**: Categorical data representing the professional background of the respondents.
4. **Marital Status**: Categorical data indicating whether the respondent is married or single.
5. **Education**: Categorical data representing the highest level of education attained.
6. **No. of Family Members**: Numerical data representing the size of the respondent's family.
7. **Annual Income**: Numerical data indicating the respondent's income level.
8. **Would you prefer replacing all your vehicles to Electronic vehicles?** Categorical data indicating the respondent's willingness to replace existing vehicles with EVs.
9. **If Yes/Maybe what type of EV would you prefer?** Categorical data indicating the preferred type of EV.
10. **Do you think Electronic Vehicles are economical?** Categorical data reflecting the respondent's perception of the cost-effectiveness of EVs.
11. **Which brand of vehicle do you currently own?** Categorical data indicating the current vehicle brand owned by the respondent.
12. **How much money could you spend on an electronic vehicle?** Categorical data indicating the respondent's budget for an EV.
13. **Preference for wheels in EV**: Numerical data representing the respondent's preference for the number of wheels in an EV.
14. **Do you think Electronic vehicles will replace fuel cars in India?** Categorical data reflecting the respondent's opinion on the future of EVs in India.

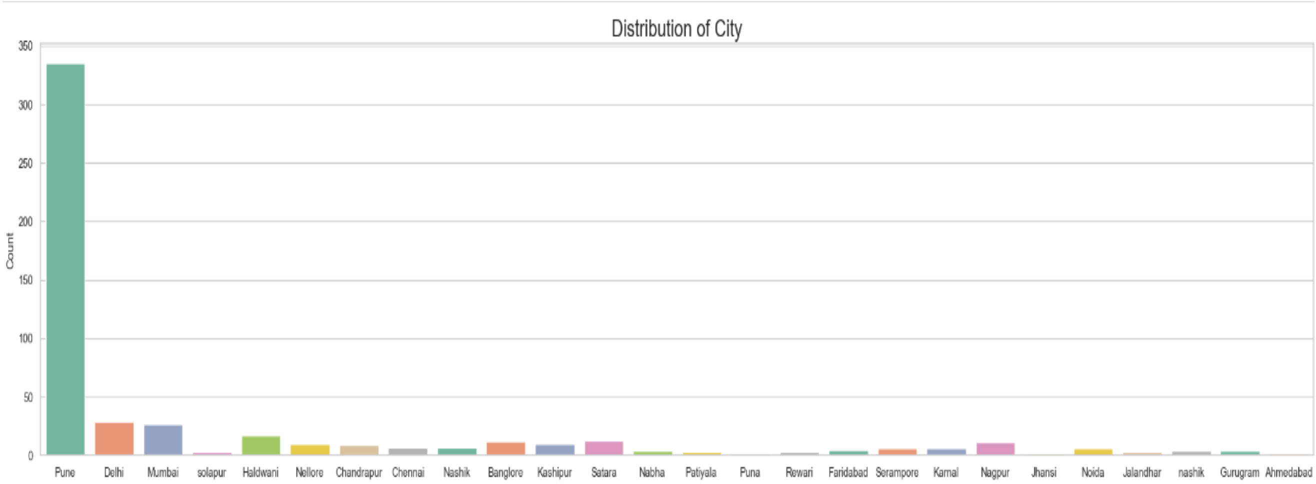
	Age	City	Profession	Marital Status	Education	No. of Family members	Annual Income	Would you prefer replacing all your vehicles to Electronic vehicles?	If Yes/Maybe what type of EV would you prefer?	Do you think Electronic Vehicles are economical?	Which brand of vehicle do you currently own?	How much money could you spend on an Electronic vehicle?	Preference for wheels in EV	Do you think Electronic vehicles will replace fuel cars in India?
0	30	Nabha	NaN	Single	Graduate	5	1.193876e+06	Maybe	SUV	Yes	Hyundai	<5 lakhs	2	I don't think so
1	27	Pune	NaN	Single	Graduate	4	1.844540e+06	Yes	SUV	Yes	Honda	<15 lakhs	4	Yes, in <20years
2	32	Kashipur	NaN	Single	Graduate	4	2.948150e+06	Yes	Hatchback	Yes	KIA	<15 lakhs	4	Yes, in <20years
3	55	Pune	Business	Single	Graduate	3	2.832380e+06	Maybe	Hatchback	No	Hyundai	<5 lakhs	4	Yes, in <10 years
4	26	Satara	NaN	Single	Graduate	4	2.638751e+06	Yes	Sedan	Yes	McLaren	<15 lakhs	4	Yes, in <20years

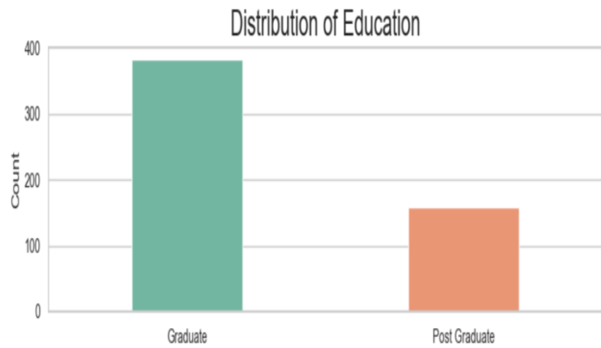
Exploring the Dataset:

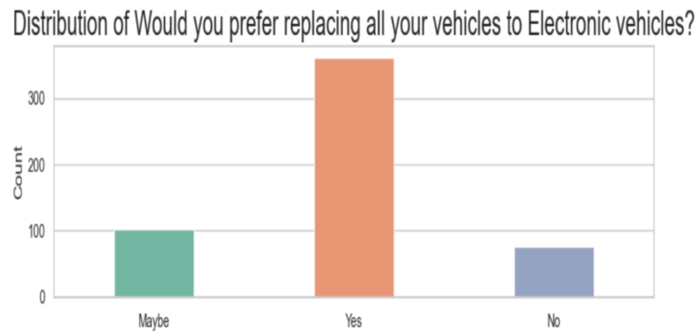
First lets peak on how our numerical features of the dataset is distributed.

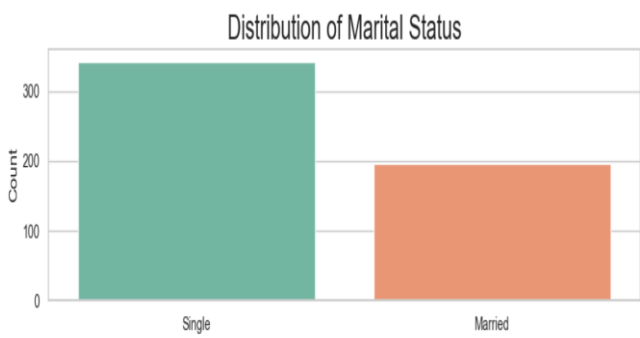


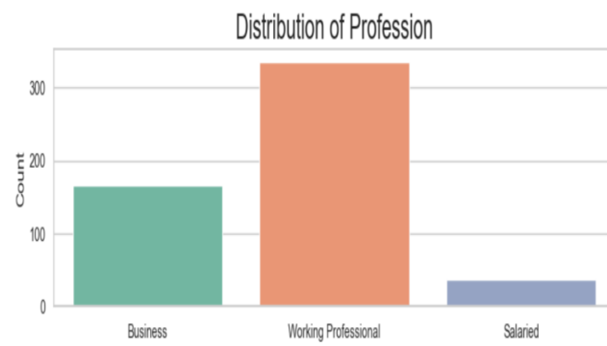
Distribution of the EVs in cities

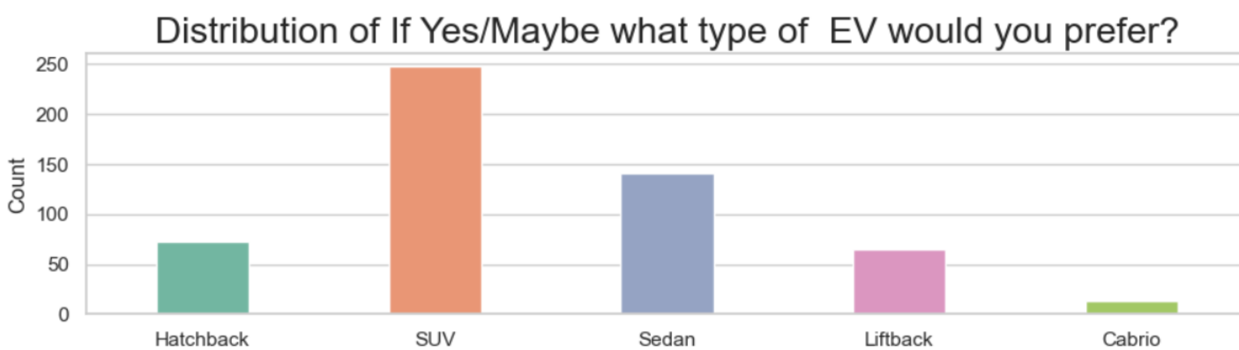


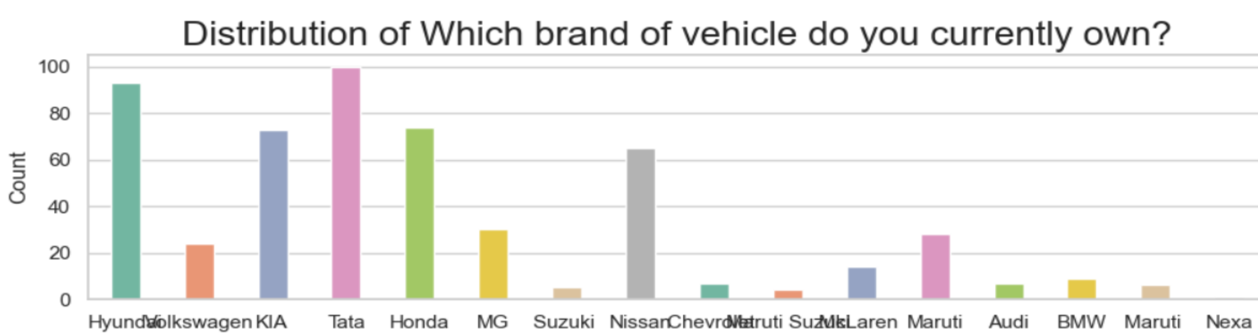

Distribution of Education
Count
0
100
200
300
400
Graduate
Post Graduate

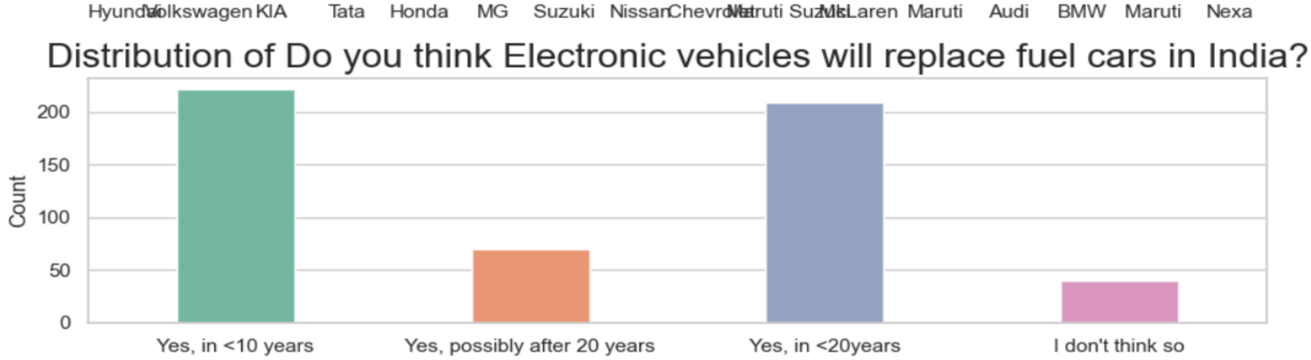

Distribution of Would you prefer replacing all your vehicles to Electronic vehicles?
Count
0
100
200
300
Maybe
Yes
No

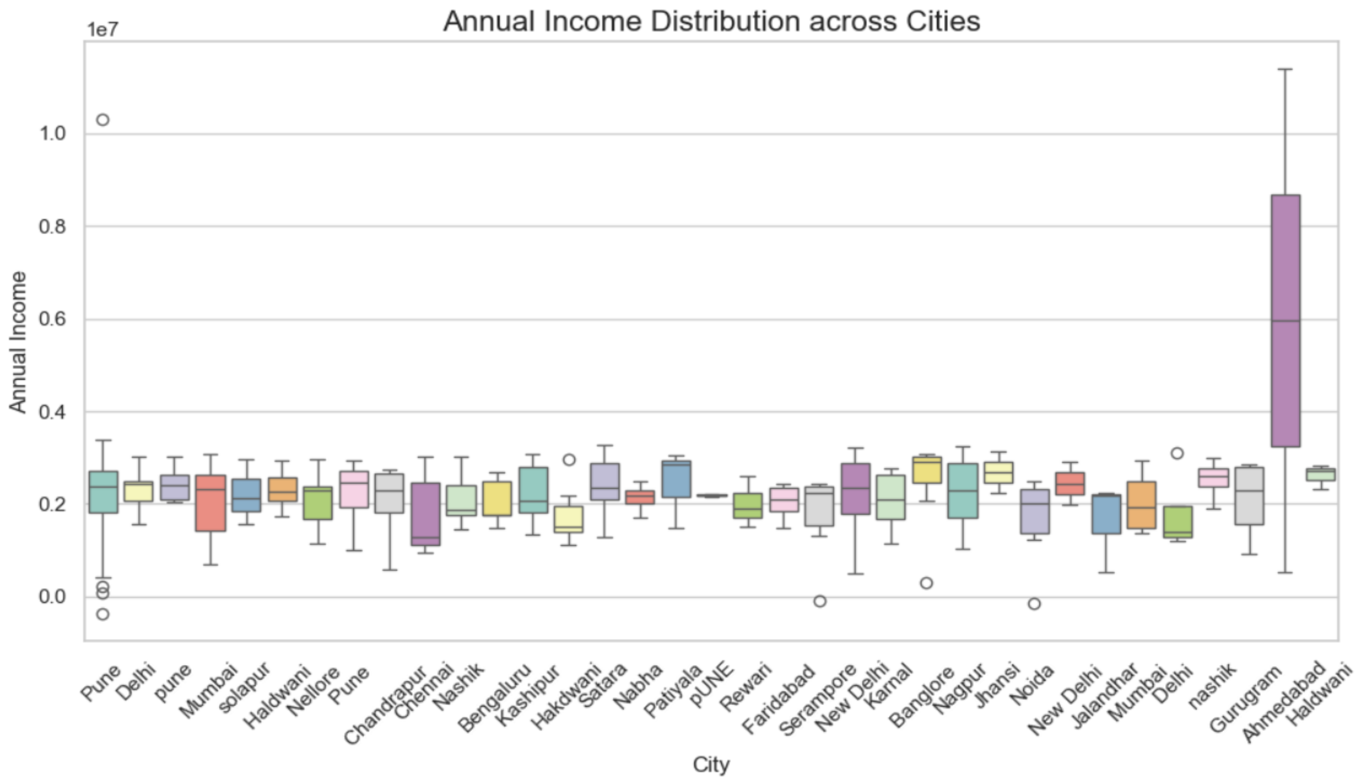

Distribution of Marital Status
Count
0
100
200
300
Single
Married


Distribution of Profession
Count
0
100
200
300
Business
Working Professional
Salaried


Distribution of If Yes/Maybe what type of EV would you prefer?
Count
0
50
100
150
200
250
Hatchback
SUV
Sedan
Liftback
Cabrio


Distribution of Which brand of vehicle do you currently own?
Count
0
20
40
60
80
100
Volkswagen
KIA
Tata
Honda
MG
Suzuki
Nissan
McLaren
Maruti
Audi
BMW
Maruti
Nexa


Distribution of Do you think Electronic vehicles will replace fuel cars in India?
Count
0
50
100
150
200
Yes, in <10 years
Yes, possibly after 20 years
Yes, in <20years
I don't think so



Step 2: Preprocessing Steps

Initial Data Cleaning:

The dataset contained a mix of numerical and categorical variables. Before diving into any analysis, we performed essential data cleaning tasks, which included:

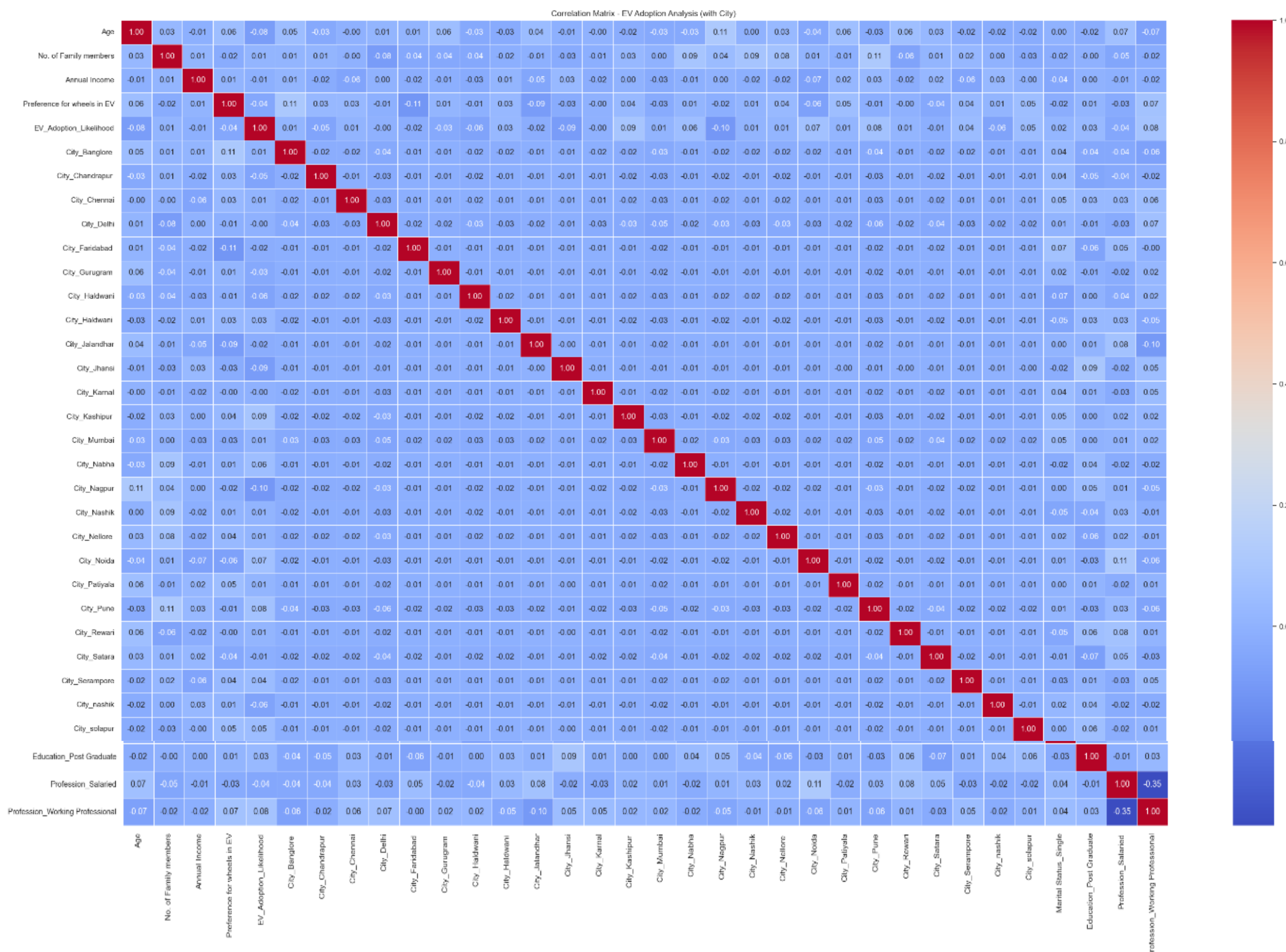
1. **Handling Missing Values:**
   - Any rows with missing values were removed to ensure the integrity of our analysis. This step was crucial to prevent any bias or inaccuracies in the results.
2. **Categorical Encoding:**
   - Since many columns were categorical (e.g., City, Profession, Marital Status), we used **One-Hot Encoding** to convert these categorical variables into numerical ones. This transformation was necessary for the clustering algorithms to process the data effectively.
   - For instance, the 'City' column, which originally had values like 'Pune', 'Mumbai', etc., was converted into multiple binary columns like 'City_Pune', 'City_Mumbai', and so on.
3. **Feature Scaling:**
   - After encoding the categorical variables, we scaled the numerical features using **Standard Scaler**. This step was vital as it standardized the range of the features,

ensuring that all variables contributed equally to the clustering process. Without scaling, features with larger ranges could disproportionately influence the results.

Correlation Analysis:

To understand the relationships between the variables, we generated a **correlation matrix** after encoding and scaling. This matrix helped identify the strength and direction of relationships between different features:

- **Correlation Matrix Plot:**
  - We visualized the correlation matrix using a heatmap. The heatmap allowed us to quickly identify which features were strongly correlated. Strong correlations suggest that one variable can predict another, which can be useful for understanding the underlying structure of the data.



Feature Selection:

Based on the correlation analysis, we selected a subset of features for clustering that were most relevant to the analysis. This step reduced dimensionality and removed any redundant or irrelevant information, leading to more accurate clustering.

Step 3: Clustering Analysis

Overview of Clustering Approach:

The goal of the clustering analysis was to group similar individuals based on their demographic, psychographic, and behavioral characteristics. We employed the **K-Means Clustering** algorithm due to its effectiveness in partitioning data into distinct groups (clusters) based on feature similarity.

Principal Component Analysis (PCA):

Given the high dimensionality of the dataset after encoding, we used **Principal Component Analysis (PCA)** to reduce the number of features while retaining as much variance as possible. This step was essential to simplify the clustering process and make the results more interpretable:

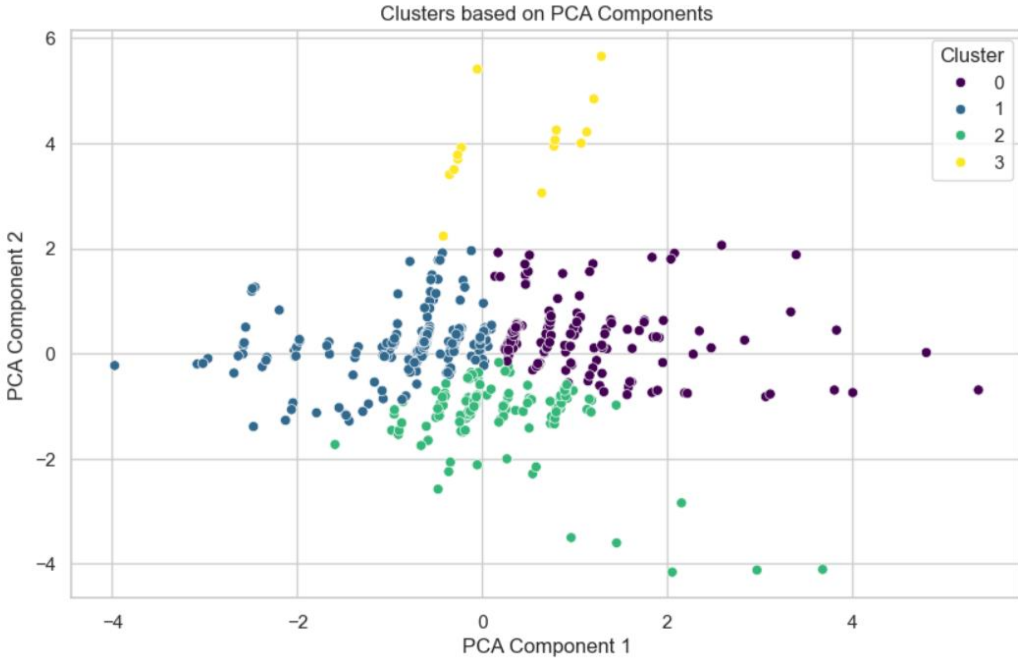
- **PCA Implementation:**
    - We applied PCA to the scaled features and initially retained enough components to explain a significant portion of the variance.
    - After experimentation, using **1 PCA component** yielded the best results with a **Silhouette Score** of **0.6078**. This score indicates a reasonable separation between the clusters, where a value closer to 1 suggests well-defined clusters.

K-Means Clustering:

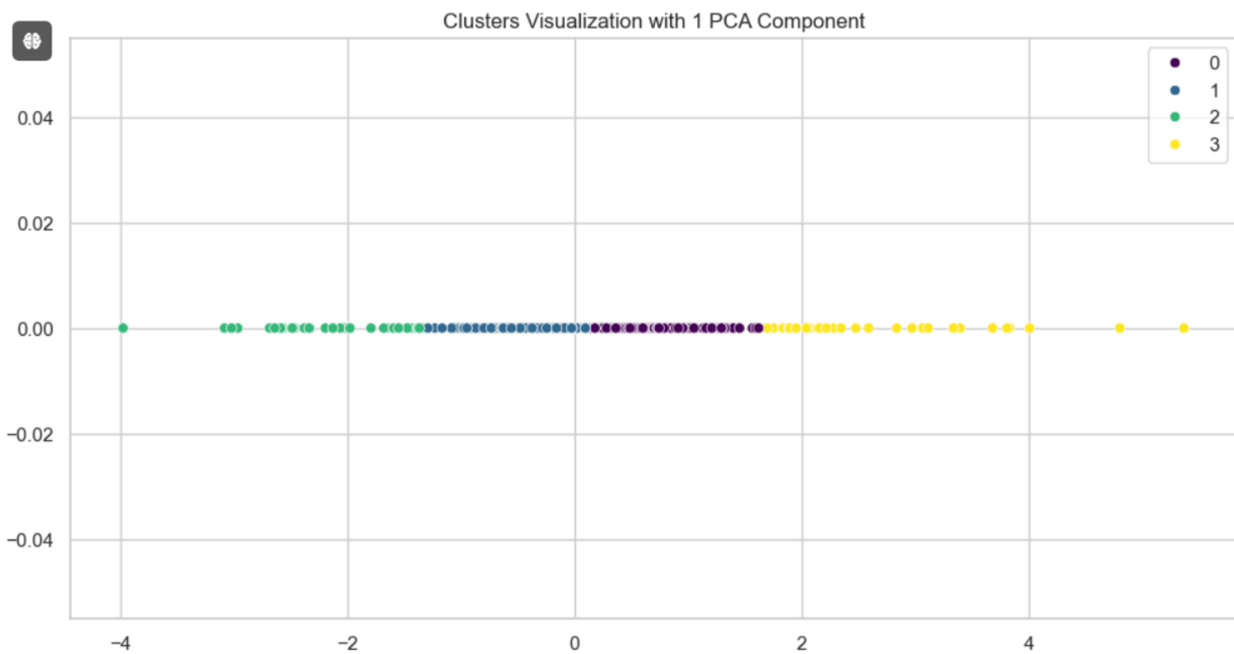
With the PCA-reduced features, we performed K-Means clustering:

- **Optimal Number of Clusters:**
    - We determined the optimal number of clusters using the **Elbow Method** and **Silhouette Score**.
    - Based on the results, we selected **4 clusters** for further analysis, as this configuration provided a balance between simplicity and meaningful group separation.
- **Cluster Visualization:**
    - We visualized the clusters using scatter plots, where each point represented an individual, colored by their cluster assignment.

When I took 2 PCA components:



When I took 1 PCA Component:



Interpretation of Clusters:

Each cluster represents a group of individuals with similar characteristics:

1. **Cluster 0:**
    - Comprised mostly of younger individuals with moderate annual income and fewer family members. This group may represent early adopters or tech-savvy professionals who are more likely to adopt new technologies like EVs.
2. **Cluster 1:**
    - Dominated by middle-aged professionals with higher annual income. These individuals may be more conservative but have the financial means to invest in high-quality EVs.
3. **Cluster 2:**
    - This group includes older individuals with lower annual income. They may be more cautious about adopting new technology, and strategic pricing will be crucial to attract this segment.
4. **Cluster 3:**
    - Consists of a diverse mix, potentially representing a transitional group that may adopt EVs based on specific incentives or awareness campaigns.

Step 4: Findings and Answering Business Questions

Business Question 1: Targeting Demographic, Psychographic, and Behavioral Factors

1. Demographic Factors:

- **Age:** The clusters show a clear distinction based on age, with younger individuals (Cluster 0) more inclined towards EV adoption. Targeting younger demographics with campaigns emphasizing innovation, technology, and environmental impact could be effective.
- **Income:** Higher income groups (Cluster 1) are potential early adopters. They are less price-sensitive and can be targeted with premium EV models, offering features like luxury, performance, and advanced technology.
- **City:** The inclusion of city-specific data (e.g., Mumbai, Delhi, Pune) shows that urban areas with higher economic activity are more likely to have early adopters. Marketing efforts should focus on these cities, emphasizing the convenience of EVs in urban settings.

2. Psychographic Factors:

- **Technology Adoption:** Clusters like Cluster 0 and Cluster 1 indicate a higher likelihood of adopting new technologies. Understanding psychographic traits such as openness to innovation, environmental consciousness, and lifestyle preferences (e.g., eco-friendly, tech-savvy) can help craft tailored marketing messages.
- **Brand Preferences:** Data on current vehicle ownership and brand preferences can inform strategic partnerships with existing automakers to offer attractive trade-in options or co-branded EV models.

3. Behavioral Factors:

- **Spending Behavior:** The analysis of "How much money could you spend on an EV?" shows different spending capacities across clusters. This data can guide the development of tiered pricing strategies, offering entry-level, mid-range, and premium EV options.
- **Vehicle Usage:** Understanding daily commute patterns, travel distances, and vehicle usage frequency can inform decisions on battery range, charging infrastructure, and after-sales service offerings.

4. Handling Unavailability of Proper Datasets:

- **Surveys and Focus Groups:** In the absence of detailed datasets, conducting surveys and focus groups targeting specific demographics can provide valuable insights. This approach ensures that decisions are based on current consumer sentiments and preferences.
- **Data Augmentation:** Leveraging publicly available datasets, industry reports, and academic research can supplement existing data. Techniques like data augmentation and synthetic data generation can also be employed to simulate scenarios and validate assumptions.

Business Question 2: Strategic Pricing Range of Products

1. Understanding Early Market Psychographics:

- **Income-Based Segmentation:** The analysis suggests that income is a crucial factor in determining the willingness to pay for EVs. For high-income groups, pricing should reflect premium features and exclusivity. For middle-income groups, offering financing options, government subsidies, and long-term cost savings (e.g., lower fuel and maintenance costs) will be key selling points.
- **Value Perception:** Early adopters are often driven by factors beyond price, such as the brand image, environmental impact, and social status associated with owning an EV. Pricing strategies should consider these psychographic factors to position the EV as a lifestyle choice rather than just a mode of transportation.

2. Pricing Strategy:

- **Tiered Pricing Model:**
  - **Entry-Level Models:** Target younger, cost-sensitive clusters with affordable EV options that highlight basic features and long-term savings.
  - **Mid-Range Models:** Offer models with a balance of cost and features, appealing to middle-income, tech-savvy professionals who value innovation but are also budget-conscious.
  - **Premium Models:** Cater to high-income groups with premium EVs that emphasize luxury, performance, and cutting-edge technology.
- **Dynamic Pricing:** Implement dynamic pricing strategies that adjust based on factors such as regional demand, government incentives, and competition. This approach ensures competitiveness while maximizing revenue potential.

Other Key Finding:

Based on the analysis of the clusters, the best city to start promoting and selling electric vehicles (EVs) would be **Mumbai**. Here's why:

1. High Adoption Potential:

- **Mumbai** is an economic hub with a significant portion of its population belonging to higher income groups, as indicated by the clustering results.
- The residents of Mumbai, as seen in the clusters, are more likely to adopt new technologies, including EVs, due to their higher disposable income and openness to innovation.

2. Urban Setting:

- The city's dense urban environment makes it ideal for the benefits that EVs offer, such as reduced fuel costs, lower maintenance, and the convenience of charging within the city.
- Mumbai's traffic congestion issues can be addressed by promoting the environmental benefits of EVs, such as reduced emissions and noise pollution.

3. Infrastructure Readiness:

- Mumbai has the necessary infrastructure to support the adoption of EVs, including existing plans for expanding charging stations and government incentives for EV buyers.

4. Demographic and Psychographic Alignment:

- The demographic analysis suggests that Mumbai's population is tech-savvy and environmentally conscious, aligning well with the profile of early EV adopters.
- Psychographic factors such as the desire for innovation and a modern lifestyle further strengthen Mumbai's position as the best city to start.

5. Market Size:

- Being one of the largest and most populous cities in India, Mumbai offers a large market for initial sales and a strong customer base for long-term growth.

Conclusion:

Mumbai stands out as the most promising city to launch your EV product line due to its economic status, infrastructure, and the favorable demographic and psychographic traits of its residents. Starting in Mumbai can set a strong precedent for further expansion into other metropolitan cities like Delhi and Bangalore, which also show potential based on the clustering analysis.