

Name: Shashikumar P Kulkarni

## Problem Statement: Consumer Segmentation for Electric Vehicle Preferences

In launching a new startup in the electric vehicle (EV) market, understanding the preferred types of vehicles by end-users is critical. This analysis focuses on the psychographic segmentation of consumer data to inform strategic marketing decisions and drive revenue growth.

### 1. Data Cleaning and Preprocessing

- **Data Loading and Initial Cleaning:**

- Loaded the dataset `ev_data.csv`.

	age	city	profession	marital_stat	education	family_count	annual_income	replace_with_ev	preferred_ev_type	think_ev_are_economical
0	30	New Delhi	Working Professional	Single	Post Graduate	4	1.770466e+06	Yes	SUV	Yes
1	27	Ahmedabad	None	Single	Post Graduate	4	7.600731e+05	Yes	SUV	Yes
2	28	Pune	Working Professional	Single	Post Graduate	4	2.779701e+06	Maybe	Sedan	Yes
3	26	Pune	Salaried	Single	Post Graduate	4	1.473698e+06	Yes	Sedan	Yes
4	25	Pune	Working Professional	Single	Post Graduate	4	2.430643e+06	Yes	Sedan	Yes

- Standardized city names (Pune, Mumbai, Bengaluru, New Delhi, Haldwani, Nashik).
- Corrected and standardized values in categorical columns (city, preferred\_price\_range, preferred\_ev\_type, etc.)

### 2. Exploratory Data Analysis

- **Descriptive Statistics:**

- Utilized `df.describe()` to summarize numerical features (annual\_income)

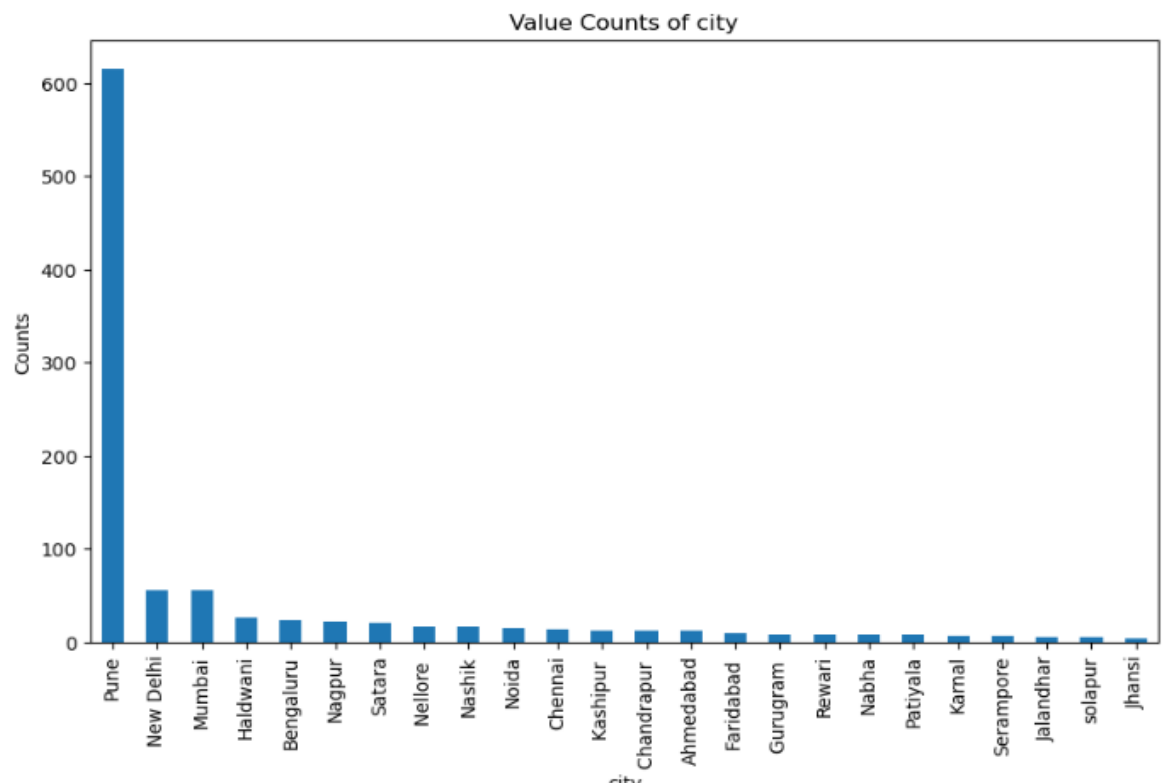
annual_income	
count	9.900000e+02
mean	2.261180e+06
std	1.001444e+06
min	-3.761509e+05
25%	1.794900e+06
50%	2.329246e+06
75%	2.758737e+06
max	1.282128e+07

- Utilized `df.describe(include='object')` for categorical features (city, preferred\_brand, etc.)

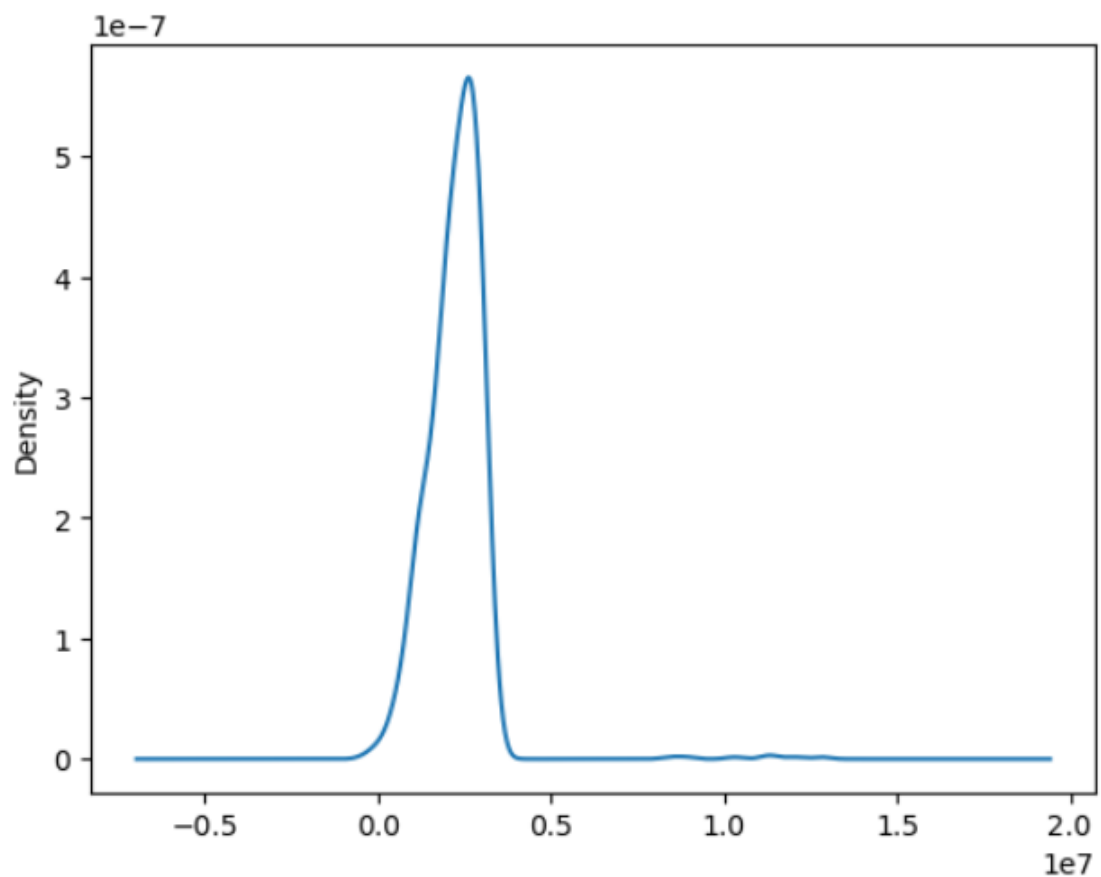
	city	profession	marital_stat	education	replace_with_ev	preferred_ev_type	think_ev_are_economical	preferred_brand	preferred_price_range
count	990	990	990	990	990	990	990	990	990
unique	35	4	2	2	3	5	3	16	8
top	Pune	None	Single	Graduate	Yes	SUV	Yes	Tata	<15 lakhs
freq	565	455	631	685	655	462	760	184	479

- **Univariate Analysis:**

- Plotted bar charts for categorical variables (city, preferred\_brand, preferred\_ev\_type, etc.)

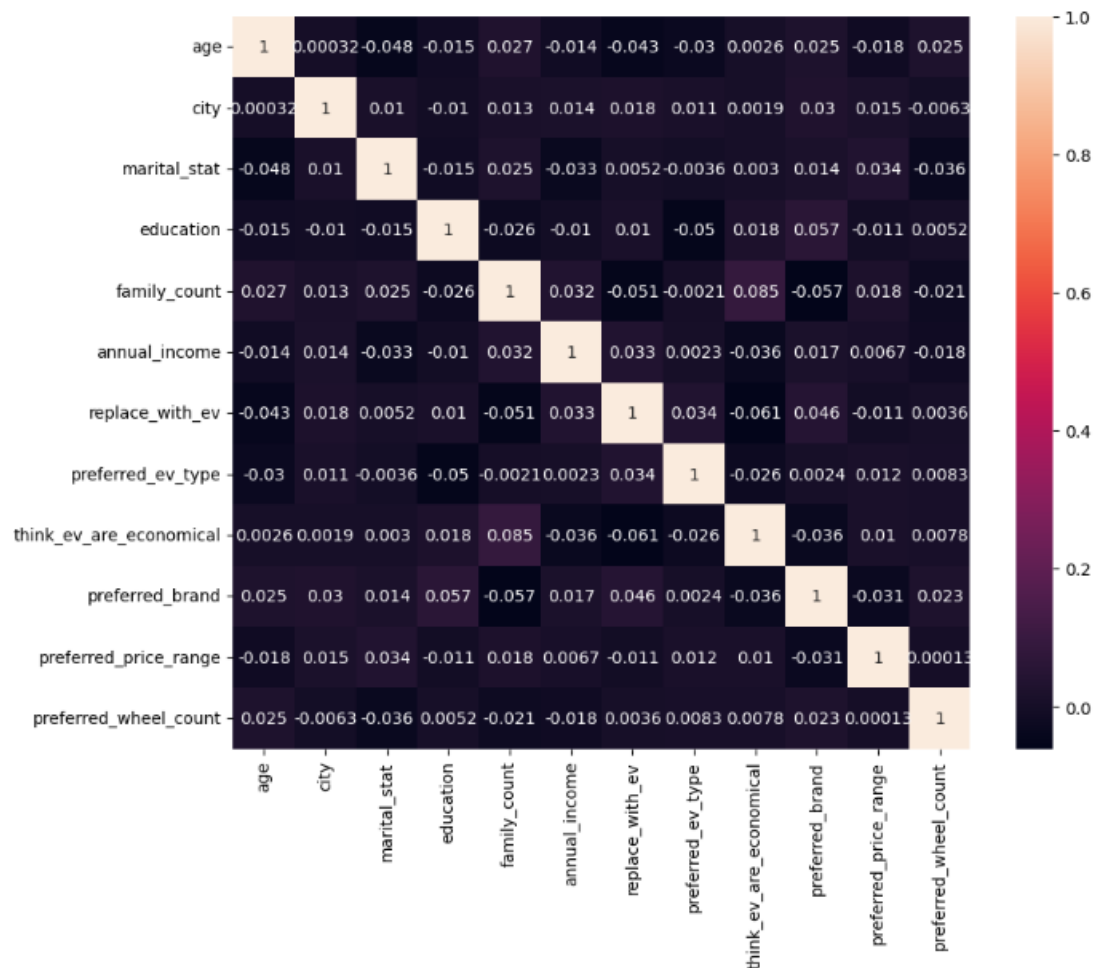


- Used KDE plot for annual\_income to understand its distribution.



- **Correlation Analysis:**

- Employed heatmap visualization (`sns.heatmap`) to explore correlations among numerical features after scaling



- **Encoding:**
  - Converted categorical variables into numerical using LabelEncoder for modeling.

### 3. Machine Learning Modeling

- **Feature Scaling:**
  - Standardized numerical features using StandardScaler.

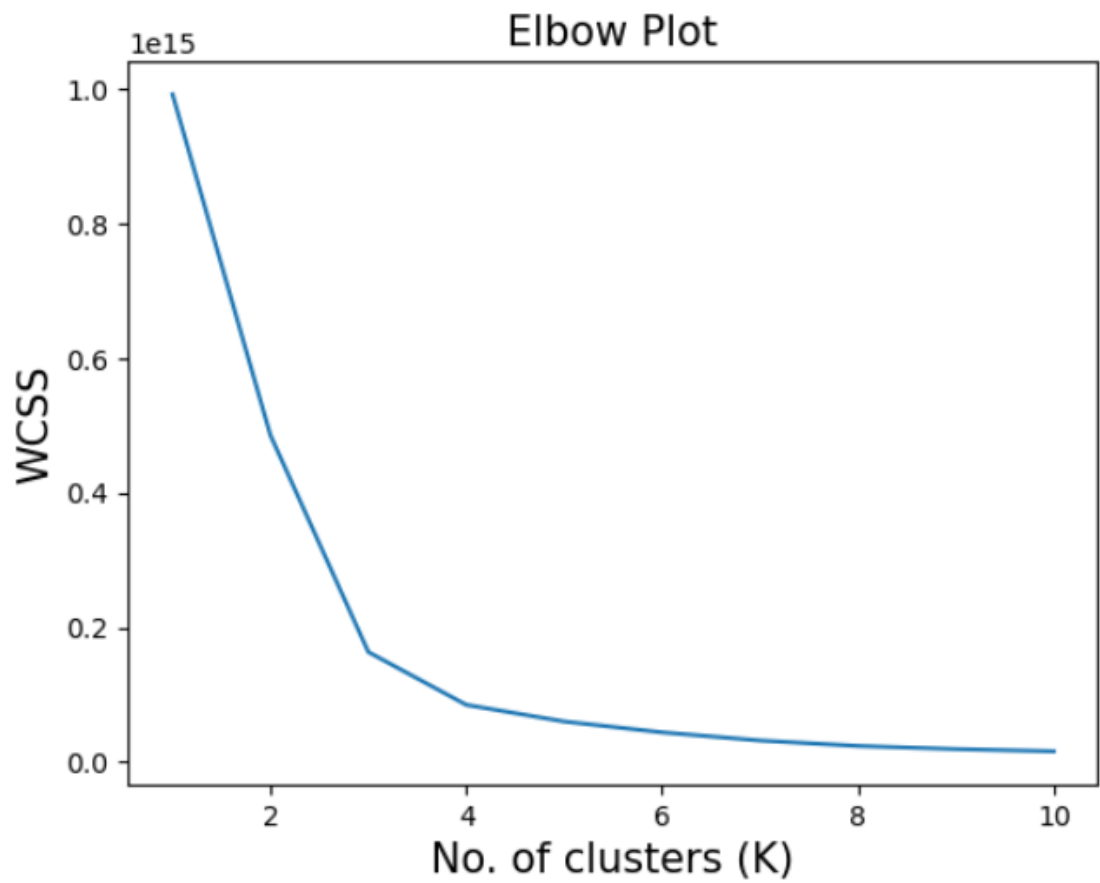
	age	city	marital_stat	annual_income	replace_with_ev	preferred_ev_type	preferred_brand	preferred_price_range
0	-0.159835	-0.018039	0.75428	-0.490254	0.669051	0.242489	-0.392423	1.337083
1	-0.425929	-3.025745	0.75428	-1.499700	0.669051	0.242489	0.934267	-0.887643
2	-0.337231	0.545906	0.75428	0.518035	-1.807062	1.180238	-1.055768	1.337083
3	-0.514627	0.545906	0.75428	-0.786744	0.669051	1.180238	1.376497	0.224720
4	-0.603325	0.545906	0.75428	0.169305	0.669051	1.180238	-0.834653	-0.887643

- **Principal Component Analysis (PCA):**
  - Applied PCA for dimensionality reduction (n\_components=4).

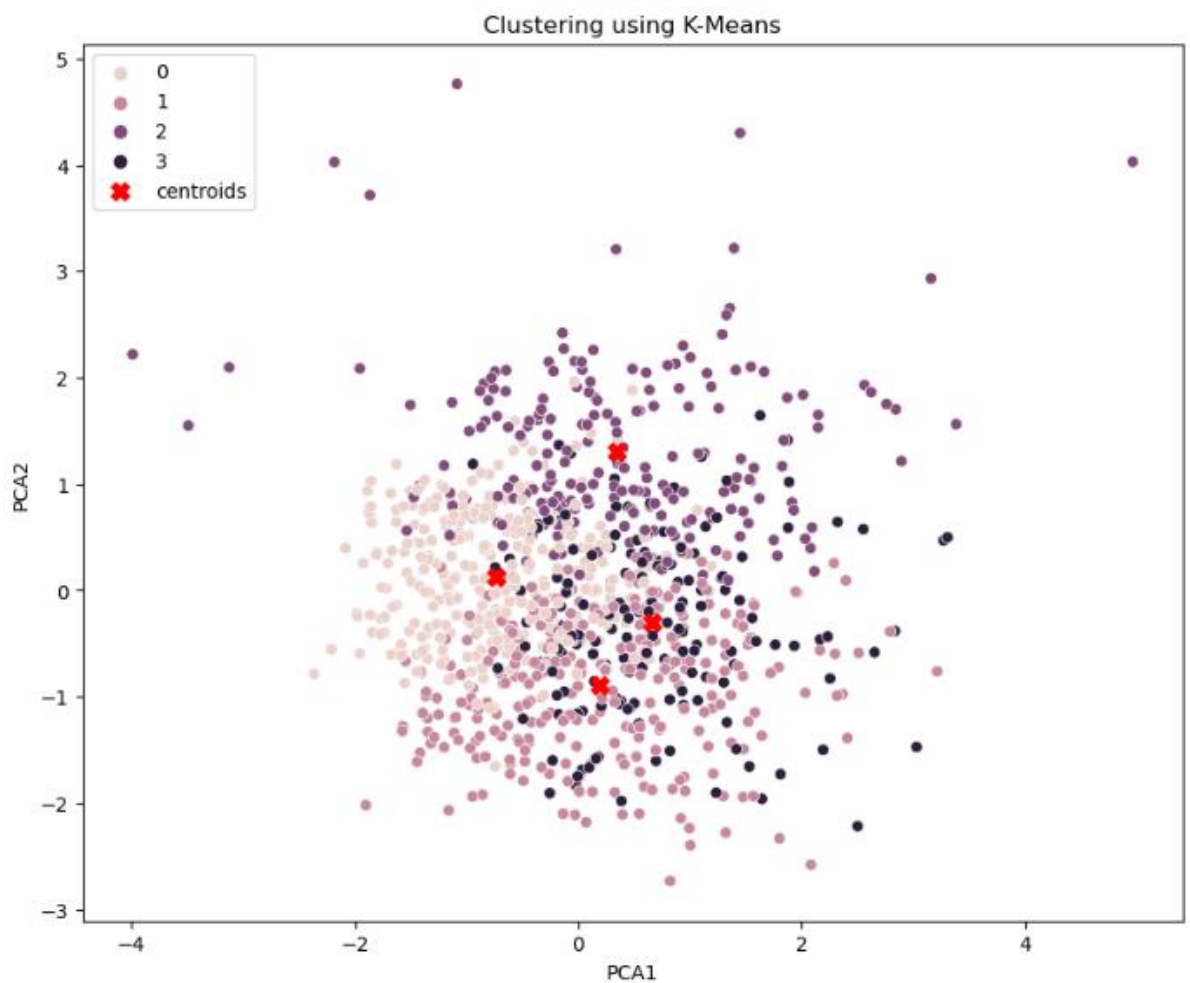
```
np.cumsum(percent_var)
```

```
array([ 13.78697092,  27.25743657,  40.08304372,  52.70219712,
        65.06202468,  77.02826126,  88.69527228, 100.          ])
```

- Analyzed variance explained by each principal component and cumulative explained variance.
- **Clustering (K-Means):**
  - Determined optimal number of clusters (K=4) using elbow method.



- Applied K-Means clustering on PCA-transformed data and visualized clusters.



- **Supervised Learning (Classification):**

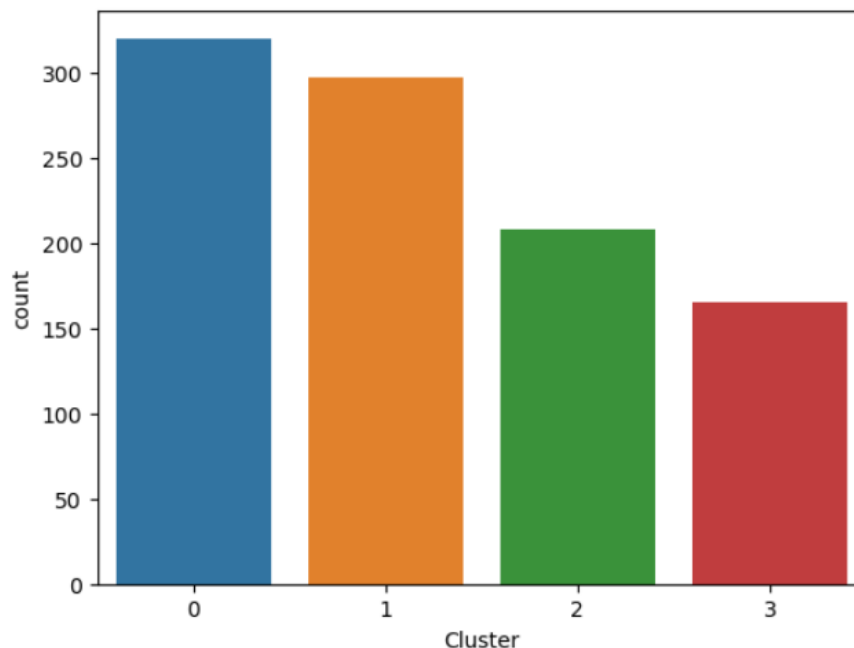
- Split data into training and test sets (`test_size=0.2`).
- Trained and evaluated models:
  - Decision Tree Classifier
  - Random Forest Classifier
  - AdaBoost Classifier
  - XGBoost Classifier
- Evaluated models using `classification_report` for both training and test sets

	precision	recall	f1-score	support
0	0.51	0.39	0.44	256
1	0.68	0.70	0.69	243
2	0.76	0.56	0.65	162
3	0.46	0.81	0.58	131
accuracy			0.59	792
macro avg	0.60	0.61	0.59	792
weighted avg	0.61	0.59	0.58	792

	precision	recall	f1-score	support
0	0.48	0.36	0.41	64
1	0.60	0.63	0.61	54
2	0.82	0.67	0.74	46
3	0.42	0.68	0.52	34
accuracy			0.56	198
macro avg	0.58	0.58	0.57	198
weighted avg	0.58	0.56	0.56	198

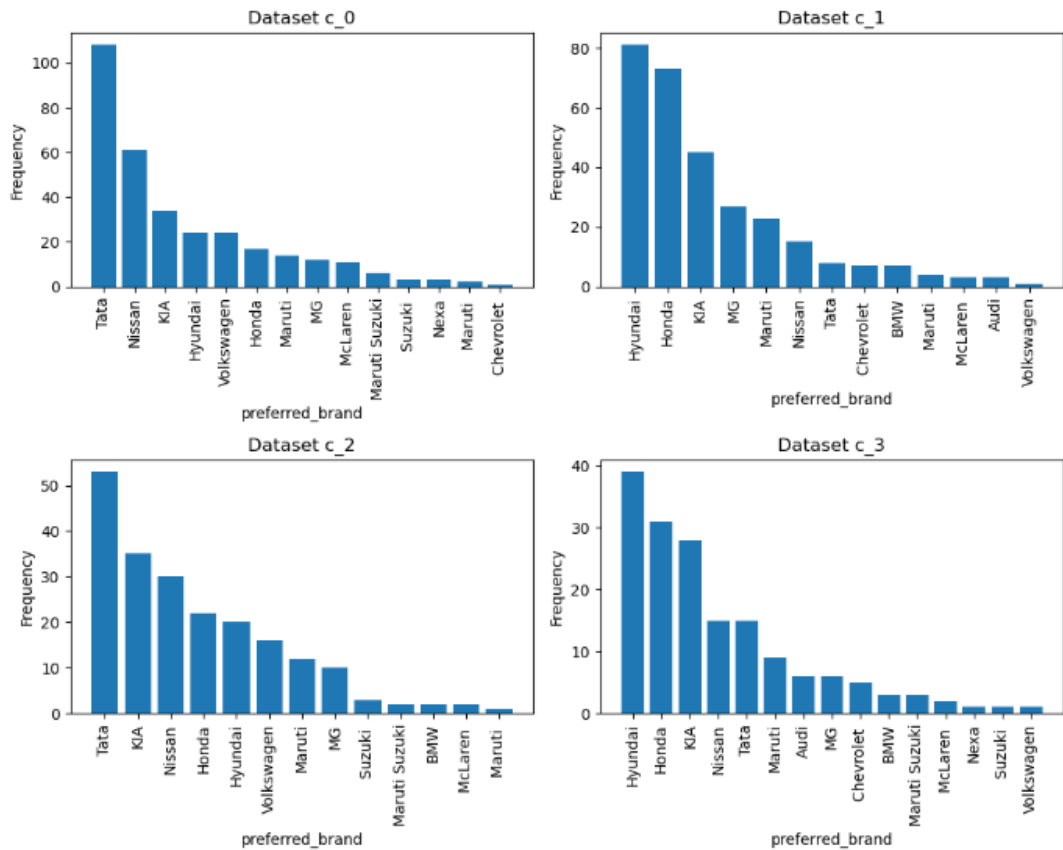
PCA Model works better with AdaBoost with accuracy of 88 and 86 on train and test data respectively



## 4. Insights and Visualizations

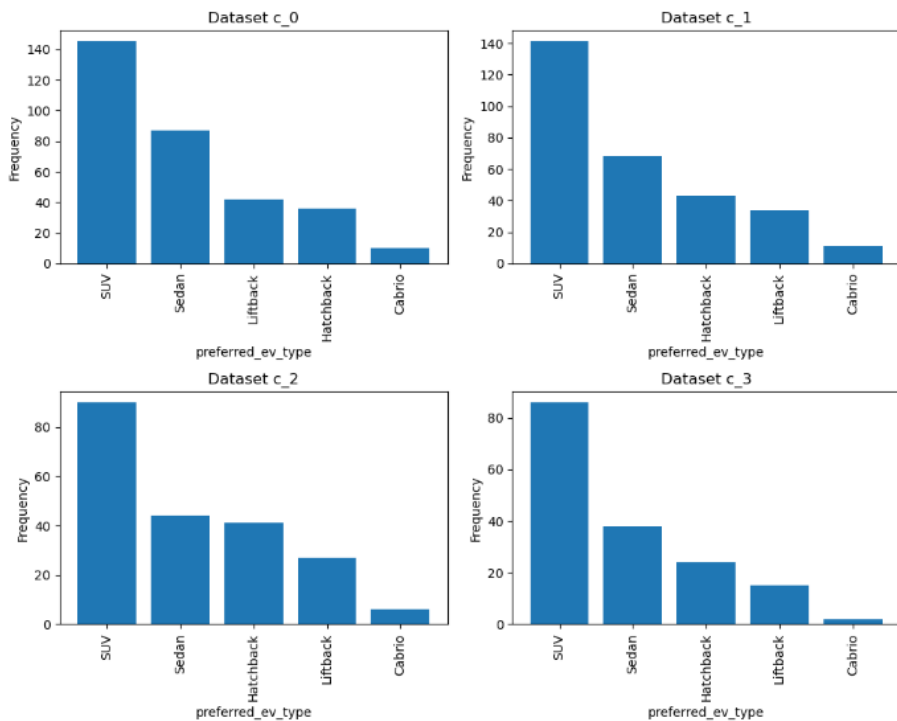
- **Demographic Insights:**

- Identified age groups and geographic locations (`city`) with highest interest in electric vehicles.



- **Brand and Vehicle Preferences:**

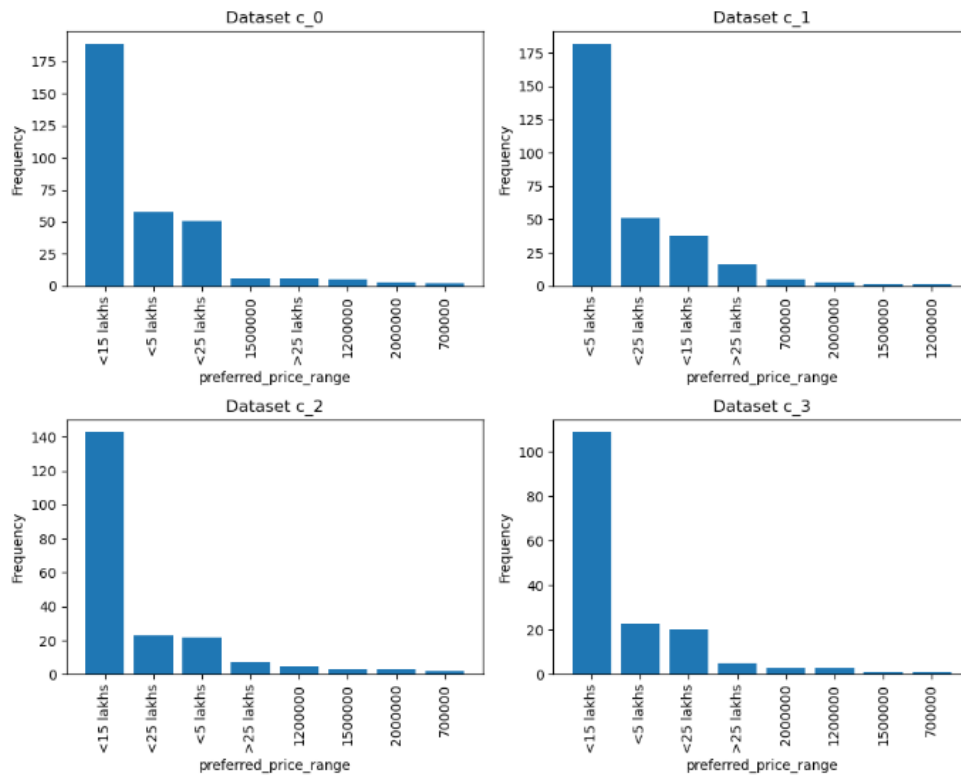
- Analyzed preferred brands (preferred\_brand) and types (preferred\_ev\_type) of electric vehicles.



- Visualized preferences using bar charts and pie charts.

- **Price Range and Economic Considerations:**

- Explored preferred price ranges (preferred\_price\_range) and perceptions on economic viability of electric vehicles (think\_ev\_are\_economical).



## 5. Conclusion

- **Market Segmentation:**
  - Electric vehicle buyers are predominantly located in urban areas such as Mumbai, Pune, Bengaluru, and New Delhi.
  - Buyers are typically aged between 25-31 years, with significant interest also observed among older age groups (~60 years).
- **Preferences:**
  - Hyundai, Honda, Kia, and Tata emerge as the top preferred brands for electric vehicles.
  - SUVs and sedans are the most favored types among buyers.
- **Economic Considerations:**
  - A majority of buyers perceive electric vehicles as economically viable, especially for models priced below 15 lakhs
- **Marketing Strategies:**
  - Focus marketing efforts on urban centers with high interest like Mumbai, Pune, and Bengaluru.
  - Highlight economic benefits and diverse model options (SUVs, sedans) to attract a wider demographic.
- **Product Development:**
  - Partner with popular brands like Hyundai, Honda, and Kia to offer attractive electric vehicle models.
  - Consider expanding offerings in SUV and sedan categories, priced competitively below 15 lakhs.

Github Link: [Shashi0170/EV\\_Market\\_Segmentation \(github.com\)](https://github.com/Shashi0170/EV_Market_Segmentation)