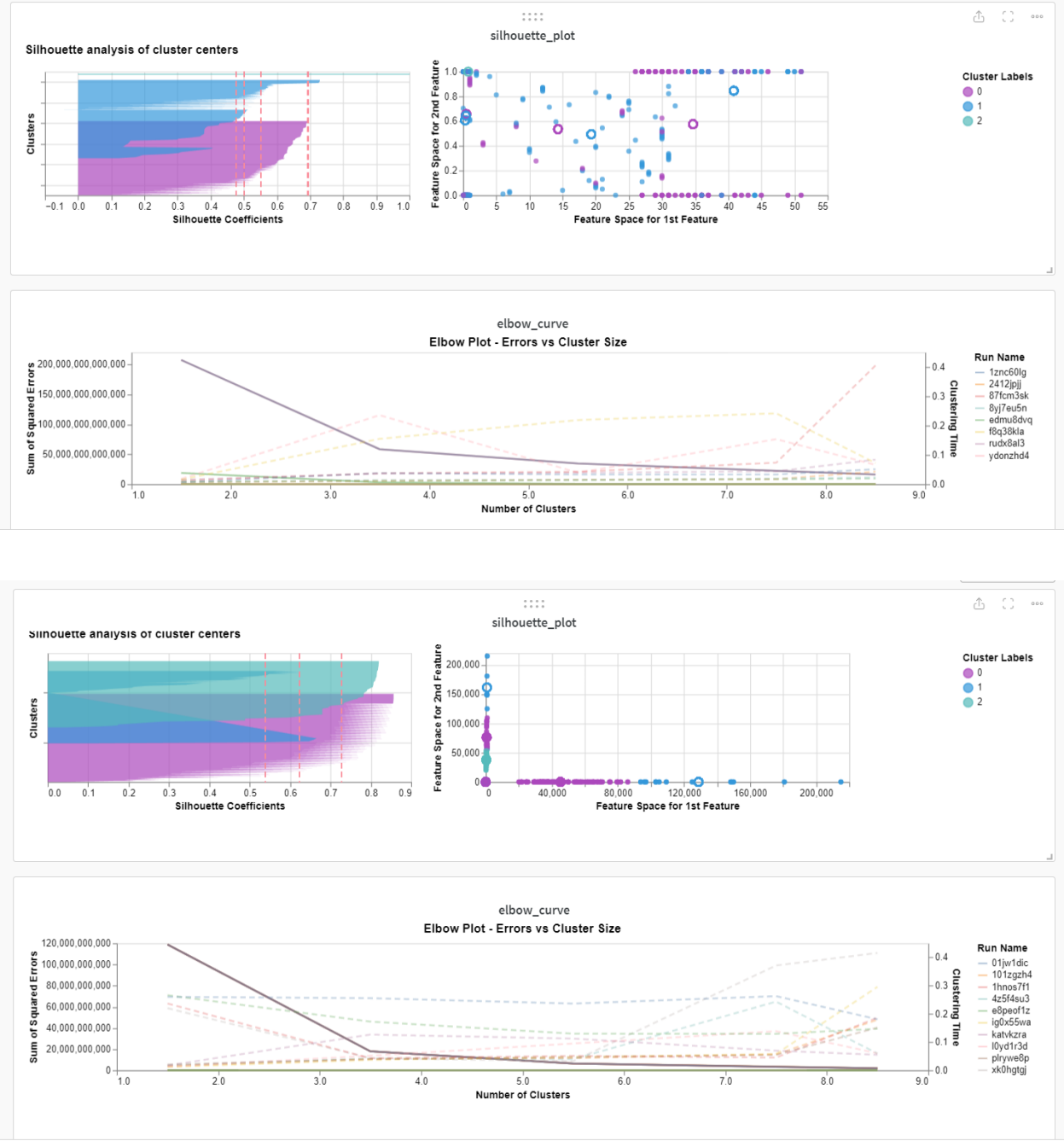
Market Segmentation

Market Segmentation Analysis of Electric Vehicles Market in India

a. Explain how and which ML model (algorithm) helped you in this Project?

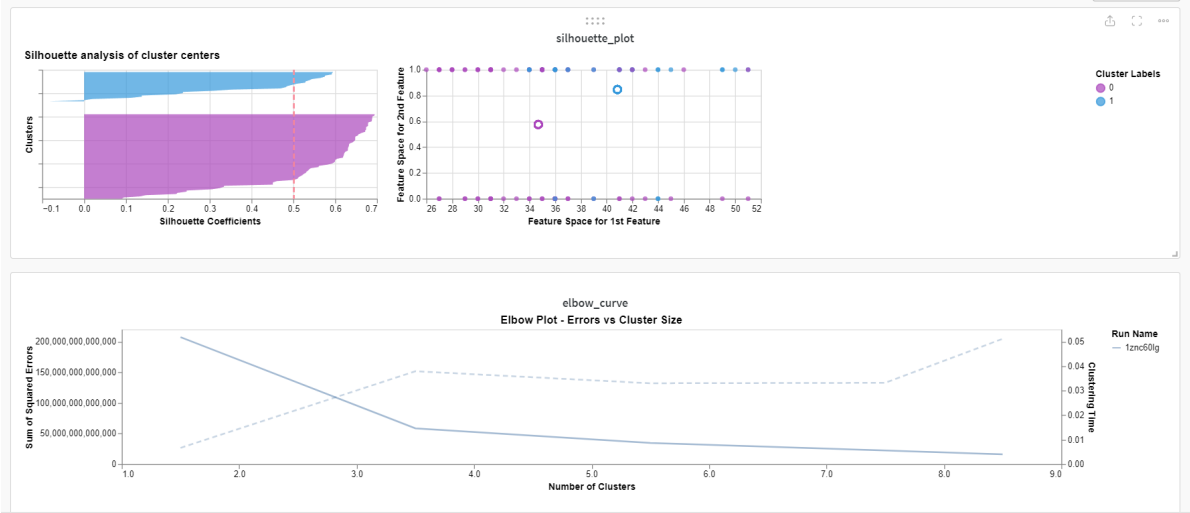
1. **PCA (Principal Component Analysis)**:
    - Role: PCA is used for **dimensionality reduction**, transforming correlated variables into uncorrelated principal components.
    - Market Segmentation: It helps identify the most influential features for segmentation by reducing data dimensionality and highlighting variables contributing to **cluster differences.**
2. **K-Means Clustering**:
    - Role: K-Means is an unsupervised clustering algorithm that partitions data into K clusters based on **feature similarity.**
    - Market Segmentation: It **groups customers** with similar characteristics to create distinct market segments for targeted marketing strategies.
3. **Silhouette Score**:
    - Role: The silhouette score **assesses the quality of clustering** by measuring data point cohesion within clusters and separation between clusters.
    - Market Segmentation: It quantifies the effectiveness of segmentation, ensuring well-defined and separate customer segments for precise targeting.
4. **Elbow Method (Kelbow Visualizer)**:
    - Role: The elbow method identifies the optimal number of clusters (K) by finding the **"elbow" point in the within-cluster sum of squares (WCSS) plot.**

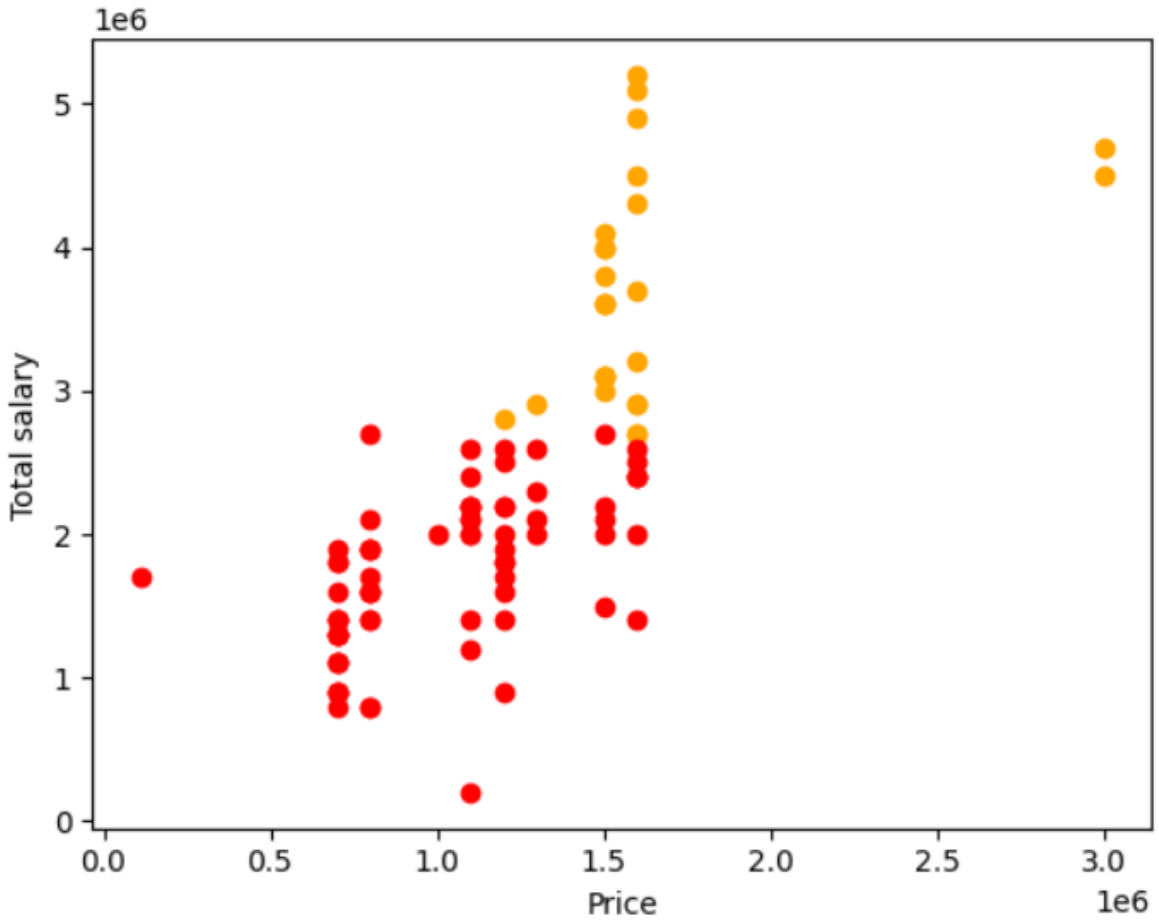
- Market Segmentation: It helps determine the most suitable number of segments for market analysis, ensuring that clusters capture meaningful distinctions among customers.



b. Elaborate on the final conclusion & insights gained from the research/analysis work.

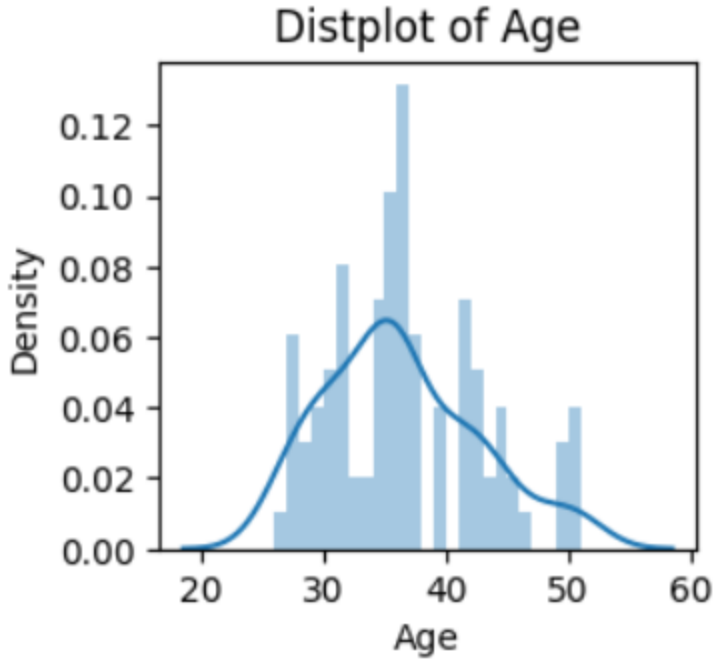
i. Most of the individuals in the dataset were in the range of 30 to 40 years and were married
ii. Number of dependents per family were between 2 to 4
iii. Individuals had a total salary between 2.3M to 2.5M out which their salaries were 1.75L to 2L
iv. Around 38 of the total individuals had taken a house loan and rest 60 (~) didn't
v. Most of the cars were either SUV or Baleno (~20) which were followed by Creata (~14)
vi. Most of the individuals were salaried (~62) and rest had their business (~34)
vii. Most of the individuals were married (~80) and only a few (~17) were single
viii. Few of the individuals received postgraduation (~57) and rest were graduates (~45)
ix. Around 33 of the individuals had 3 dependents and 27 of the individuals had 2 dependents
x. A few of the individuals had personal loans (~31) while most didn't (~69) and most didn't have any home loans (~61) while some did (~38)
xi. Next, we have performed K-Means clustering using Min-Max scaling and clustered data points based on the 'Price' and 'Total Salary' features.
    i. Checking out the silhoutte we found that it is 0.5 when we take 2 clusters into consideration

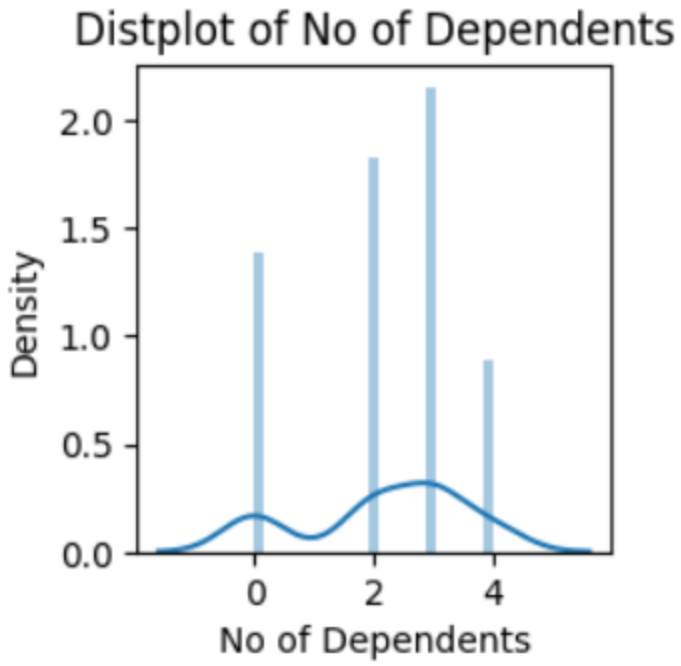


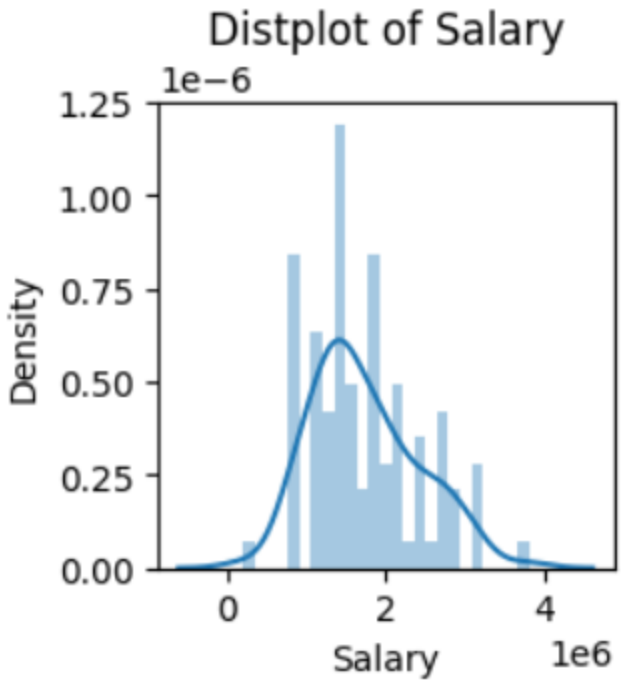


ii. Then we also clustered data points based on age and price

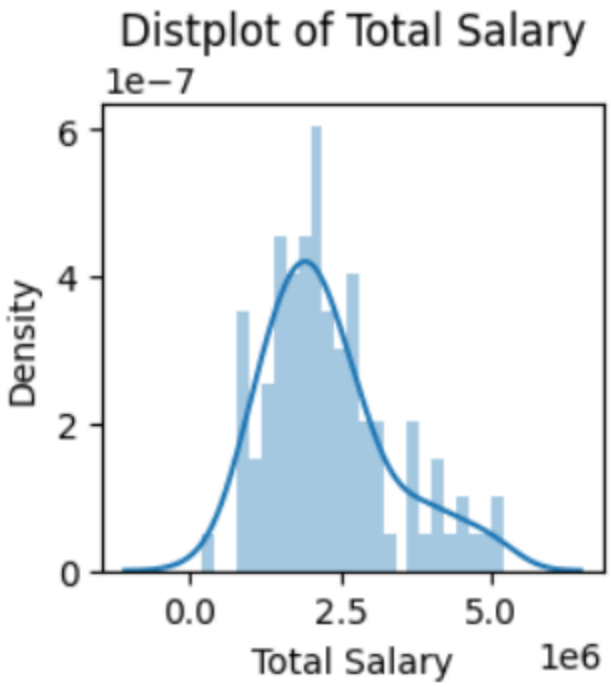
Some Supporting Visualizations:

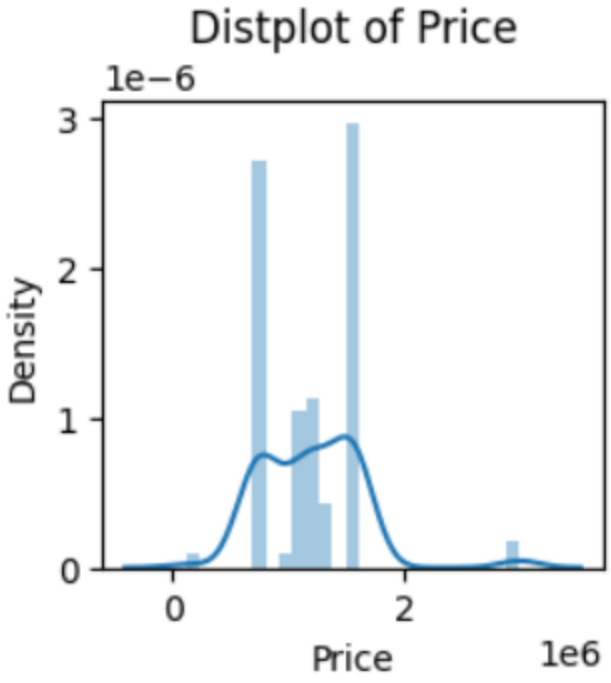

Distplot of Age
Density
0.12
0.10
0.08
0.06
0.04
0.02
0.00
20
30
40
50
60
Age

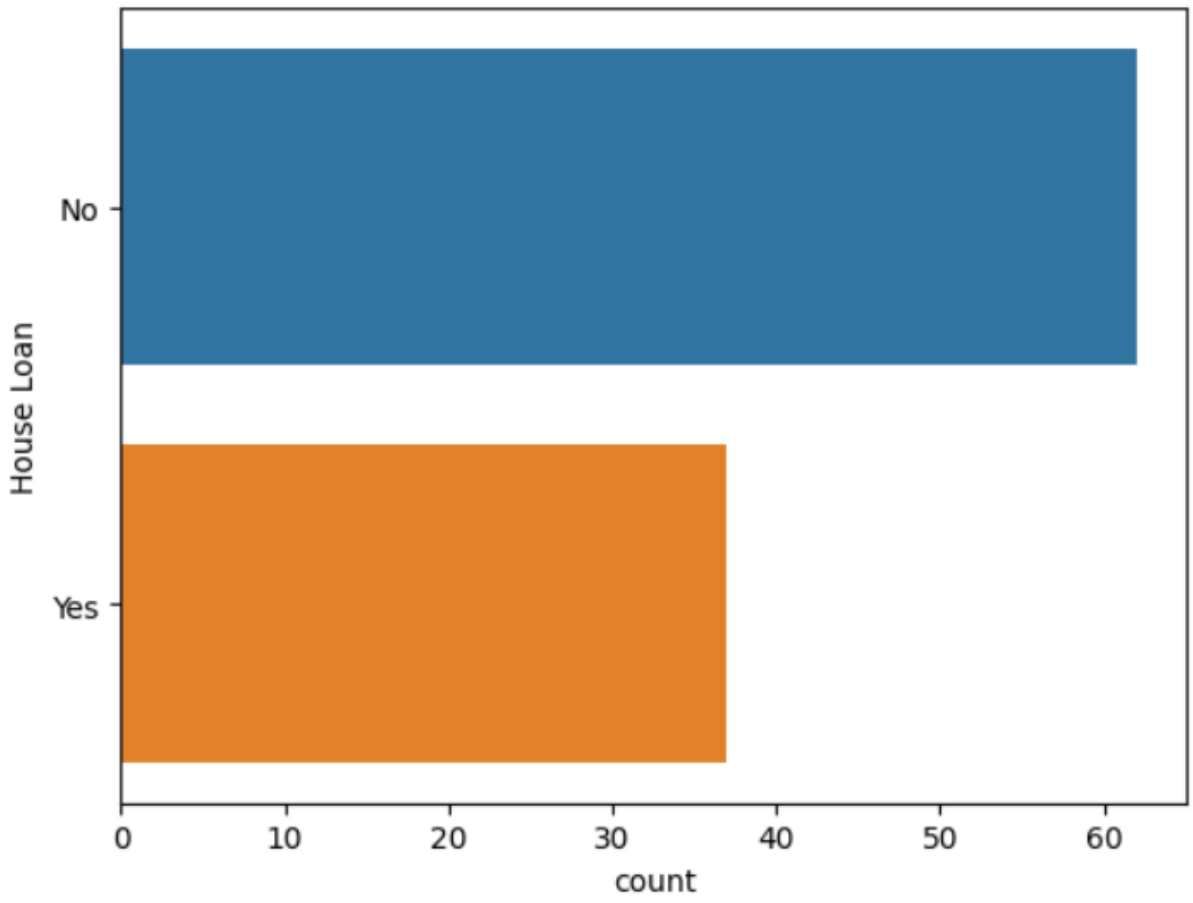

Distplot of No of Dependents
Density
2.0
1.5
1.0
0.5
0.0
0
2
4
No of Dependents

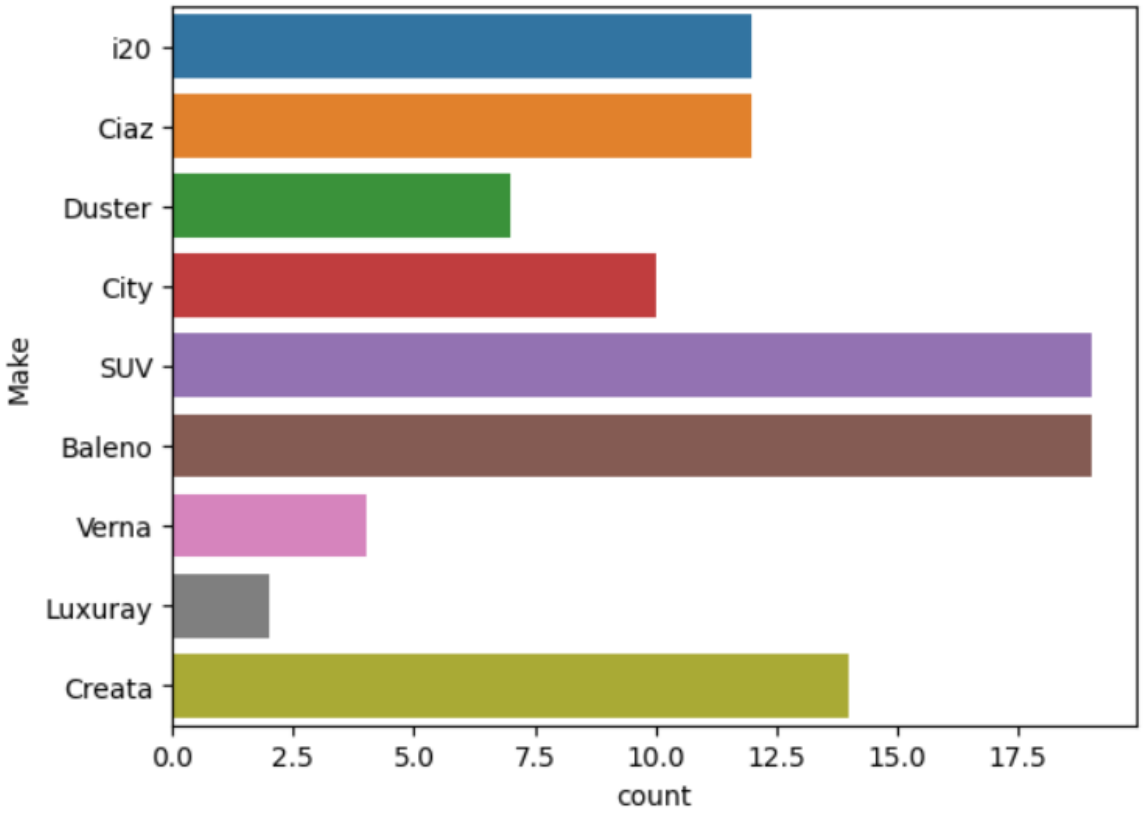

Distplot of Salary
1e−6
1.25
1.00
0.75
0.50
0.25
0.00
Density
0
2
4
Salary
1e6

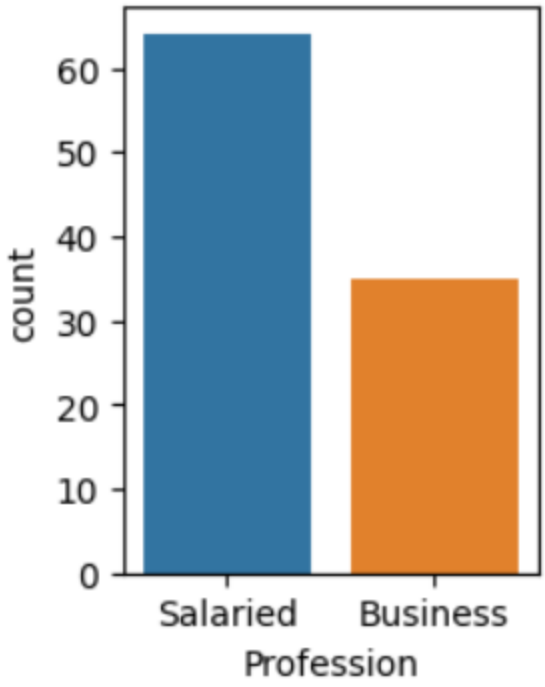

Distplot of Wife Salary
1e−6
6
4
2
0
Density
0
1
2
3
Wife Salary
1e6

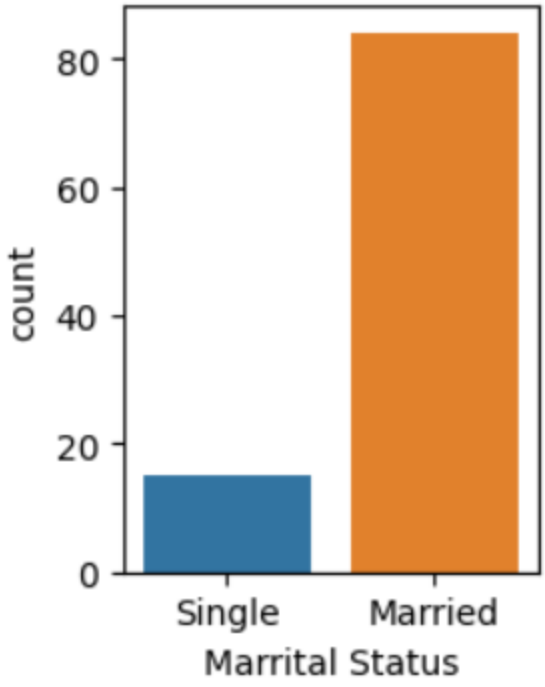

Distplot of Total Salary
1e−7
6
4
2
0
Density
0.0
2.5
5.0
Total Salary
1e6

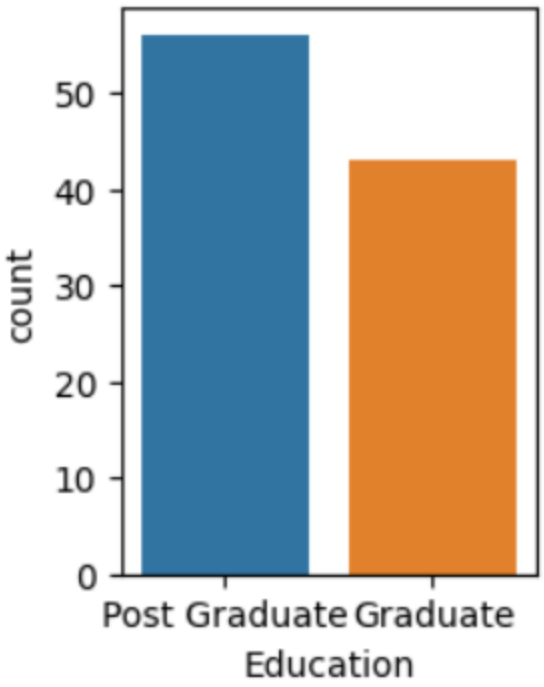

Distplot of Price
1e−6
3
2
1
0
Density
0
2
Price
1e6

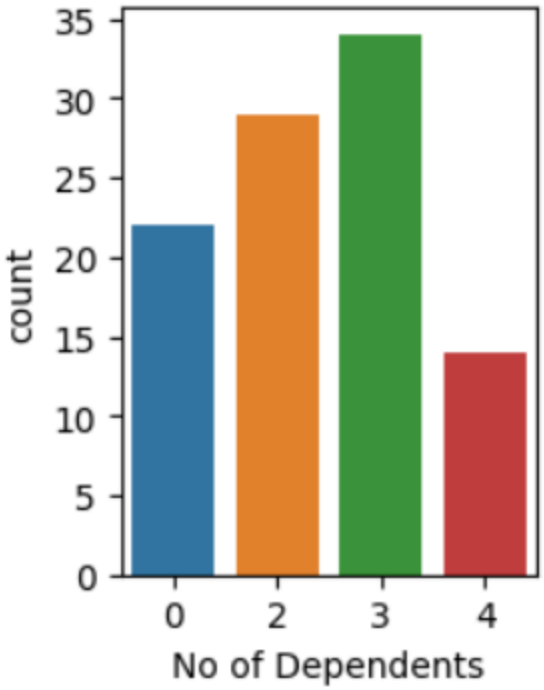

House Loan
No
Yes
0
10
20
30
40
50
60
count

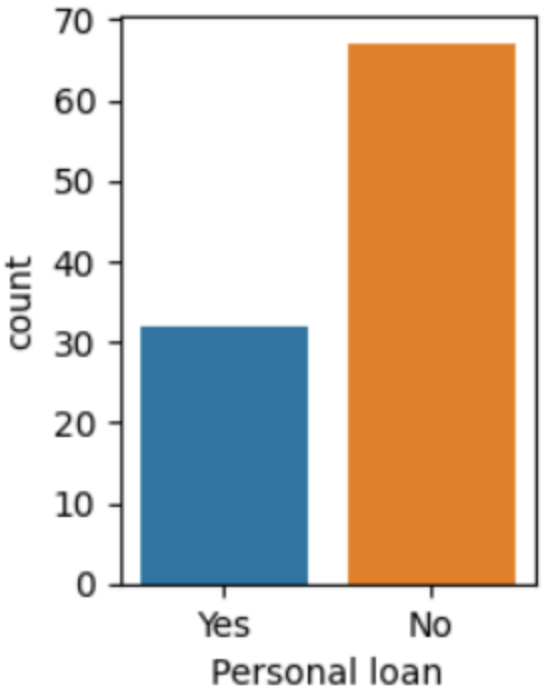

i20
Ciaz
Duster
City
SUV
Baleno
Verna
Luxuray
Creata
Make
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
count

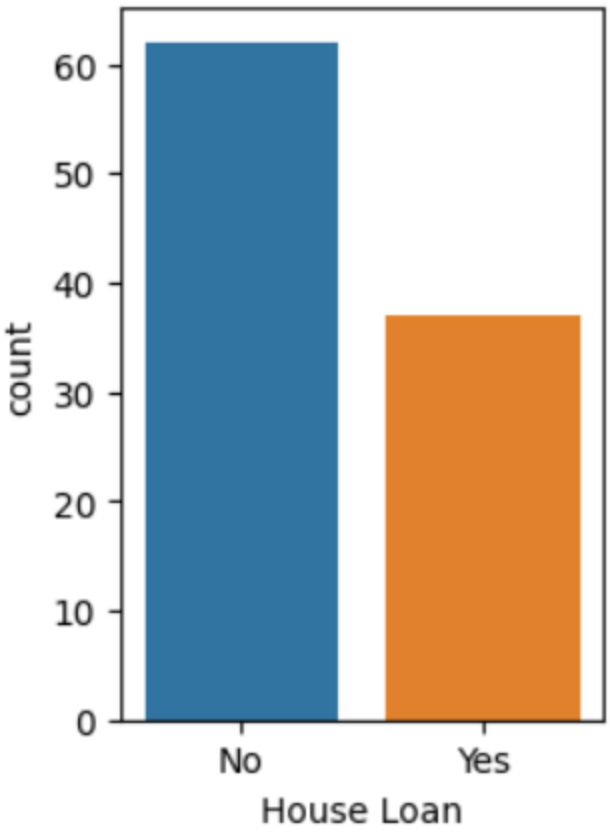

60
50
40
30
20
10
0
count
Salaried
Business
Profession

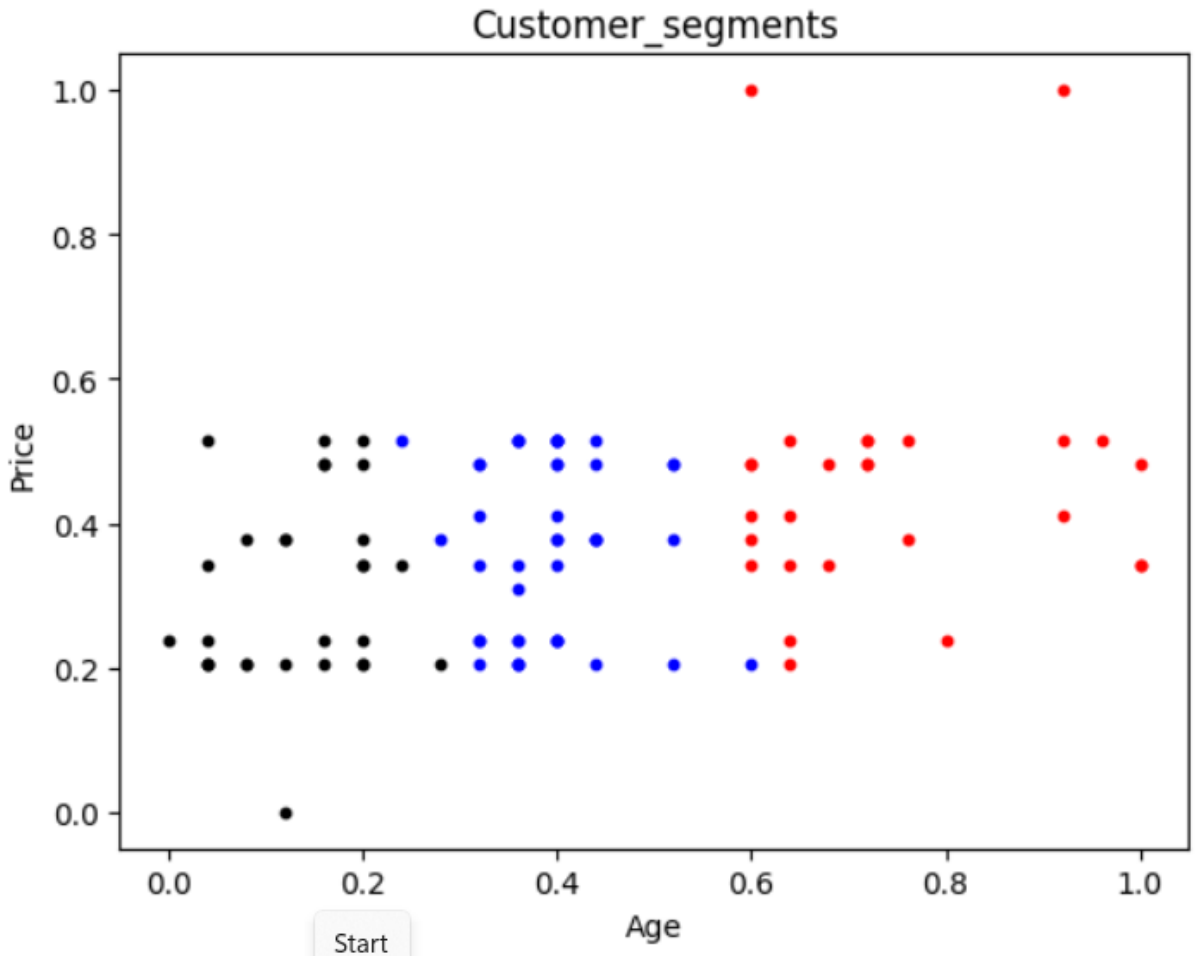

80
60
40
20
0
count
Single
Married
Marrital Status

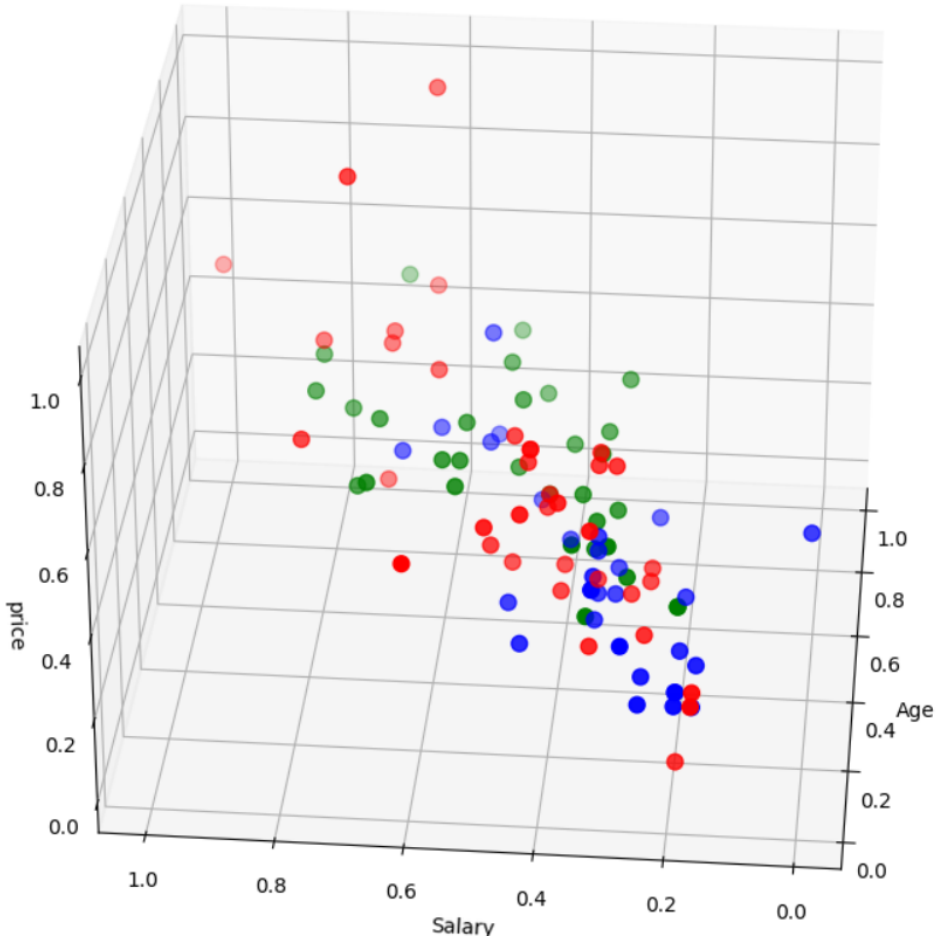

50
40
30
20
10
0
count
Post Graduate
Graduate
Education

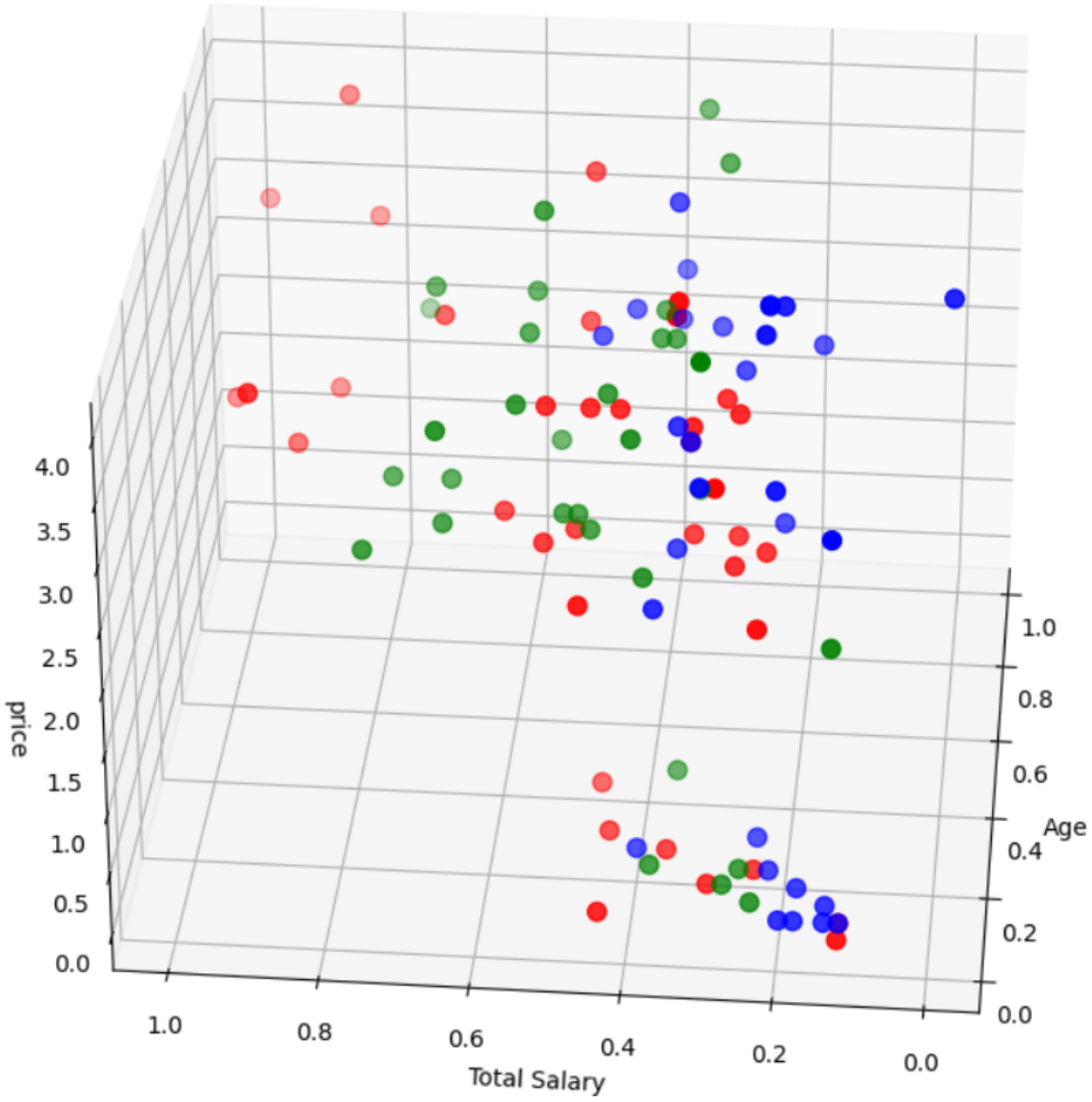

35
30
25
20
15
10
5
0
count
0
2
3
4
No of Dependents


70
60
50
40
30
20
10
0
count
Yes
No
Personal loan


60
50
40
30
20
10
0
count
No
Yes
House Loan


Customer_segments
Price
Age
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Start


price
1.0
0.8
0.6
0.4
0.2
0.0
Salary
1.0
0.8
0.6
0.4
0.2
0.0
Age
1.0
0.8
0.6
0.4
0.2
0.0



c. How will you improve upon the Market Segmentation Project given additional time & some budget to purchase data? (in terms of Datasets collection-name what columns points you will search for & what additional ML models you would like to try)

To improve upon the Market Segmentation Project with additional time and budget for data collection, you can consider the following steps:

Dataset - Indian Automobile Buying Behavior Study

1. **Data Collection**:

- **Demographic Data**: Collect additional demographic data, such as gender, location (city/region), and family size. This can provide more insights into the customer base.
- **Behavioral Data**: Gather data related to customer behavior, such as purchase history, frequency, and types of products bought.
- **Psychographic Data**: Include data on lifestyle, interests, and preferences to better understand customer segments.

2. **Machine Learning Models:**

- **Classification Models**: We can use classification models like Logistic Regression, Random Forest, or Neural Networks to predict customer behavior, such as loan acceptance or product purchase.
- **Recommendation Systems**: We can implement recommendation systems, such as collaborative filtering or content-based filtering, to provide personalized product recommendations.

Dataset - Automobile Dataset

1. **Data Collection:**

- **Gather More Data:** Acquire additional data related to electric vehicles (EVs) and their attributes. You can consider data sources like industry reports, EV manufacturers, and third-party data providers.
- **Additional Features:** Look for datasets that include more features related to electric vehicles, such as battery capacity, charging infrastructure, environmental impact, customer reviews, and maintenance costs. These additional features can provide richer insights.
- **Historical Data:** Collect historical data to analyze trends and changes in the EV market over time.

2. **Machine Learning Models:**

- **Regression Models:** We can use regression models to predict PriceEuro or other key metrics. Algorithms like Linear Regression, Random Forest

Regressor, or Gradient Boosting Regressor to help us understand the price determinants.

- **Classification Models:** We can consider using classification models to segment vehicles into different categories based on features like body style, segment, and powertrain. Models such as Decision Trees or Support Vector Machines (SVM) are the ones we thought of.
- **Time Series Analysis:** For historical data, explore time series analysis to understand how market dynamics and vehicle attributes change over time.

d. Name top 4 Variables/features which can be used to create most optimal Market Segments for your Market Domain.

The top 4 variables/features that can be used to create optimal market segments:

Dataset - Indian Automobile Buying Behavior Study

1. **Age**: Age is an important demographic feature that can be used to segment the market. Different age groups may have varying preferences, needs, and affordability when it comes to electric vehicles.
2. **Profession**: The type of profession or occupation can provide insights into the lifestyle, income, and commuting patterns of potential EV customers. Professionals in urban areas might have different preferences compared to those in rural areas.
3. **Total Salary**: Total family income, which is the sum of "Salary" and "Wife Salary," can be a crucial factor. It reflects the financial capacity of a household to purchase an electric vehicle.
4. **Make**: The "Make" of the electric vehicle is an essential feature. Different EV brands and models may cater to diverse market segments. For example, some brands may focus on affordability, while others may emphasize luxury or performance.

Dataset - Automobile Dataset

1. **Range (Range_Km):** The range an EV can cover on a single charge is a critical factor for customers. Segmenting based on range can help identify different

customer groups, such as those who prioritize longer range for long-distance travel and those who are content with shorter ranges for daily commuting.

2. **Price (PriceEuro):** Price is a significant determinant in purchase decisions. Market segments can be created based on different price points, which may include premium, mid-range, and budget segments. This can help target customers with varying budgets.
3. **Body Style (BodyStyle):** The type of body style (e.g., sedan, SUV, hatchback) can significantly influence consumer preferences. Segmenting by body style can cater to customers looking for specific vehicle designs and functionalities.
4. **Segment (Segment):** This variable can be particularly important for market segmentation. Segments might include compact EVs, luxury EVs, electric SUVs, etc. Identifying which segment appeals most to different customer groups can be valuable for marketing and product development.

GitHub Code Link - https://github.com/Dishaa09/Market-Segmentation-Electronic-Vehicles-India