MARKET SEGMENTATION ON ELECTRIC VEHICLES

TEAM: ISHITA

Date: 29/07/2022

Members:

- Ishita A Kumar
- Rudraksh Kansal
- Aditi Deshpande
- Amar Parameswaran

GitHub link:

https://github.com/Aditi271/Electric-Vehicle-Market-Segmentation-Team-Ishita

PROBLEM STATEMENT:

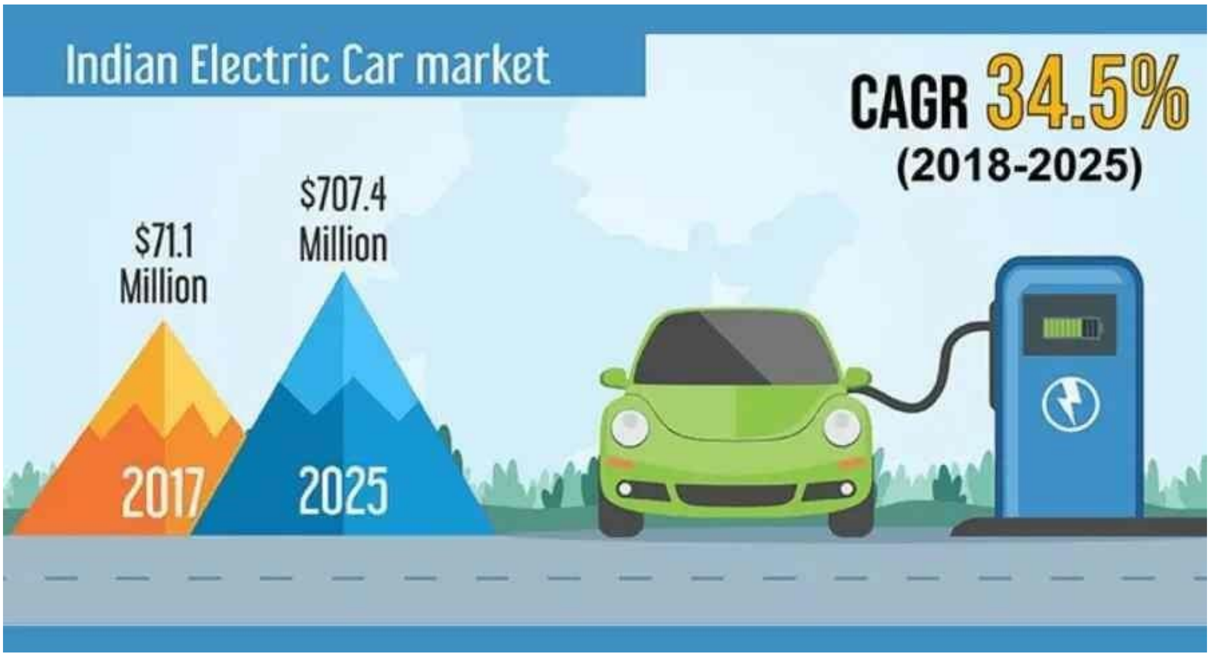
Using market segmentation, analyse the electric vehicles in India and come up with a feasible strategy to enter the market and targeting the segments that are most likely to buy the Electric Vehicles.

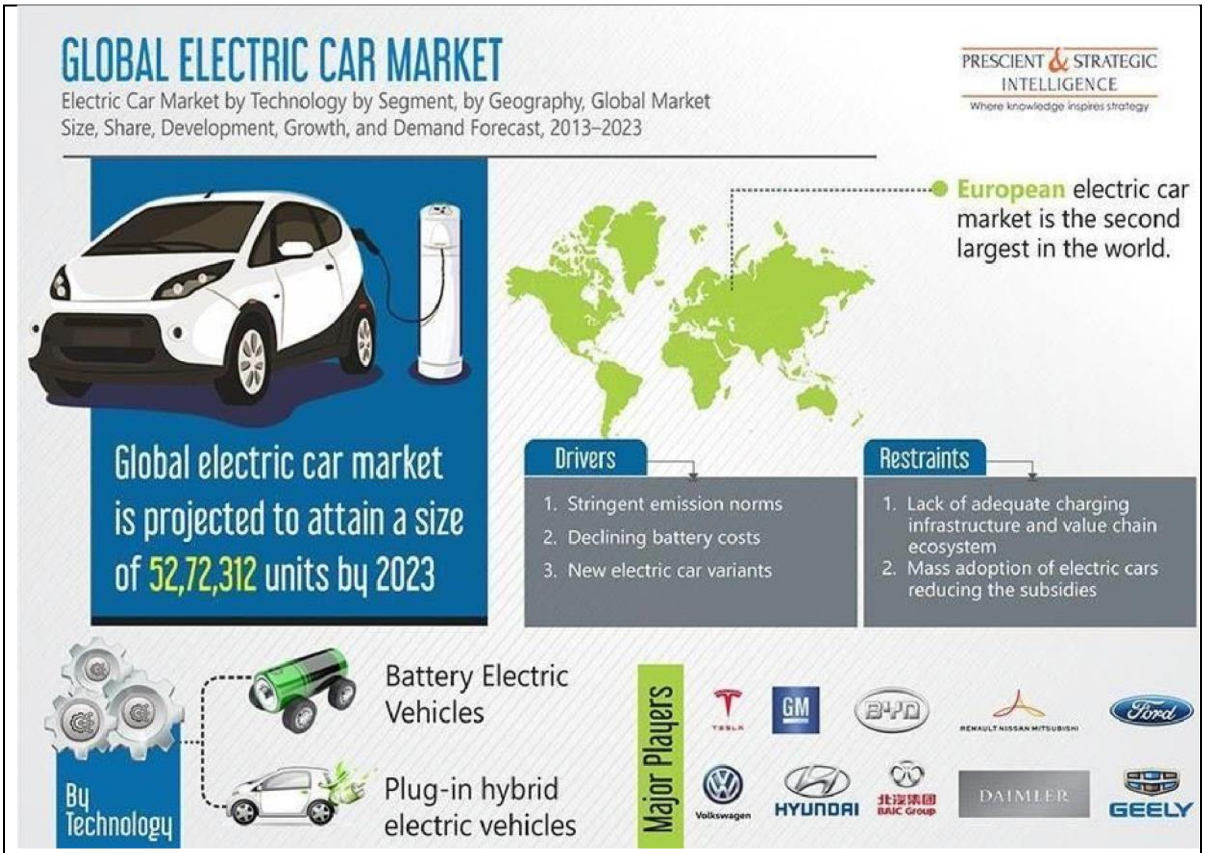
Overview:

What is Electric Vehicle?

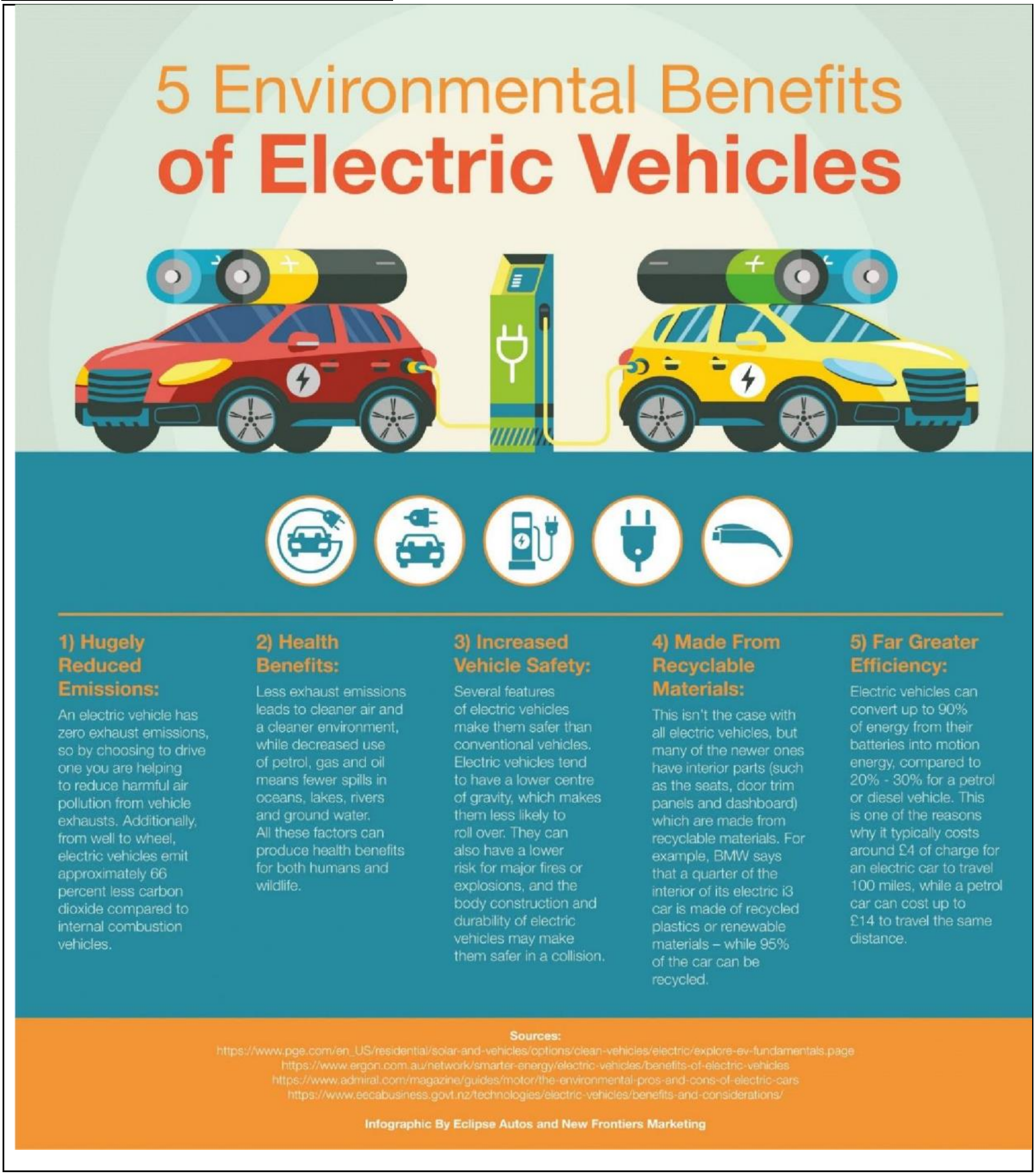
The electric vehicle is a vehicle that runs on electricity alone. Such a vehicle does not contain an internal combustion engine like the other conventional vehicles. Instead, it employs an electric motor to run the wheels. These vehicles are becoming very popular nowadays. They are considered to be a promising solution for the future transportation. The most common example is Tesla.

Business Opportunity:





Benefits of Electric Vehicles:



Data Sources:

Data was taken from the Kaggle website

[Indian Consumers Cars purchasing behaviour | Kaggle](#) **Market Segmentation:**

1. Behavioural Segmentation:

Behavioural segmentation is a form of marketing segmentation that divides people into different groups who have a specific behavioural pattern in common. Users may share the same lifecycle stage, previously purchase particular products, or have similar reactions to your messages.

Benefits of Behavioural Segmentation

- Improves targeting accuracy
- Helps provide better-personalized experience
- Sifts engaged users from uninterested
- Saves money
- Makes it easier to track success
- Helps build loyalty to your brand

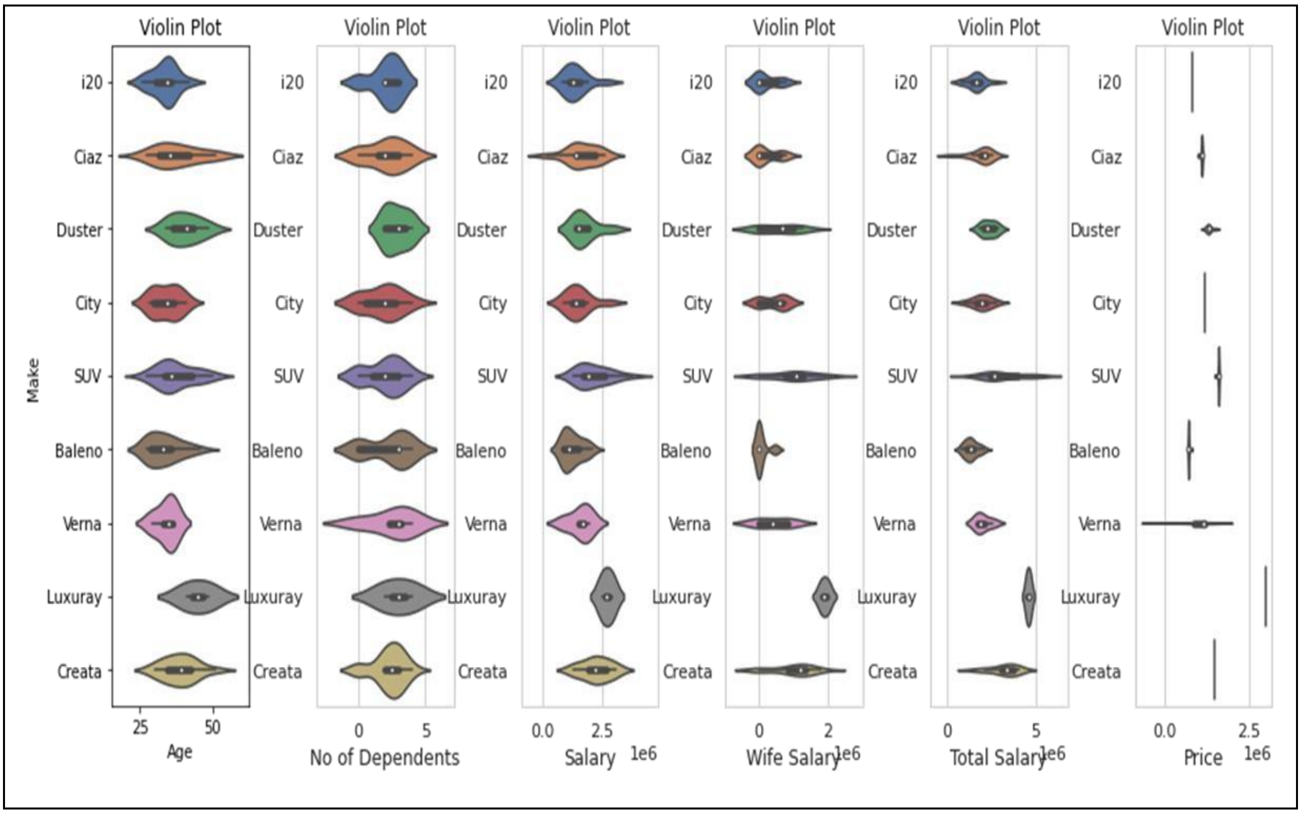
2. Psychographic Segmentation:

Psychographic segmentation's emphasis on characteristics like personality and values differs from demographic segmentation, which uses a specific trait (like gender, age, income, etc.) to categorize potential audiences.

Market researchers use psychographic characteristics to help develop and position their products and marketing messages for different target groups

Marketers use both demographics and psychographics in their market research to create their marketing strategy. So we will combine these both categories as well.

The violin plot below gives us some insight on the relation between the segmentation and descriptive variables in our data.



Observations:

<u>Age</u>: Younger consumers purchase less expensive vehicles. This can be explained simply as they have lesser dependents, lesser income and are single, and so they don't have both the option and the need to buy more expensive vehicles.

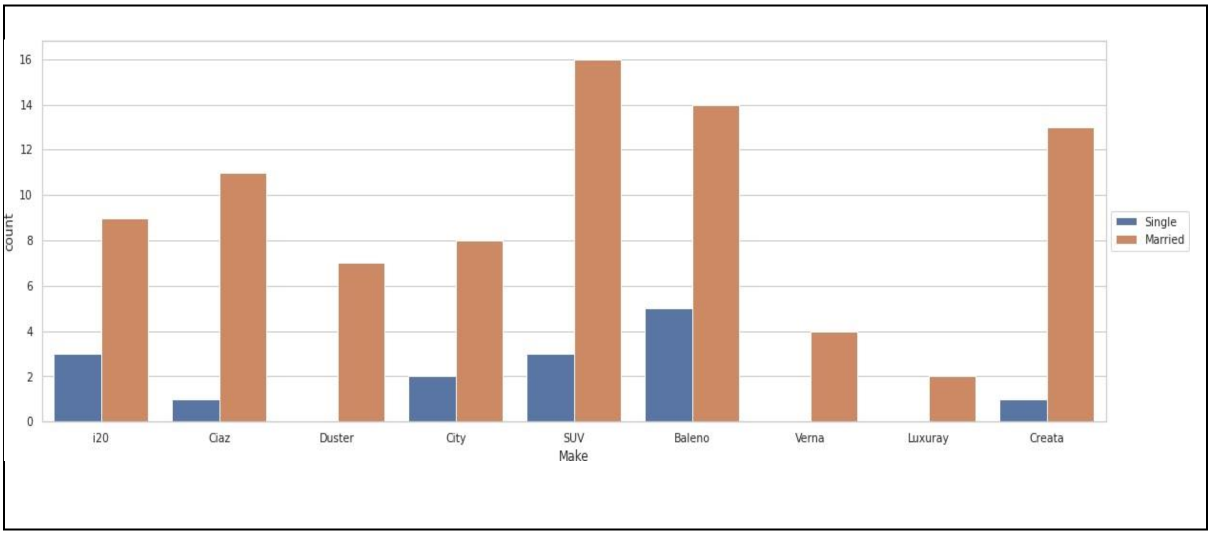
<u>Number of Dependents</u>: Greater number of dependents makes the consumer buy a vehicle with more seats and so they tend to prefer SUVs.

<u>Salary</u>: If you overlap the normalised salary plots with price plot, you would observe the median of salary violin plot matches that of the price of the vehicle indicating a very direct relationship, which makes sense as most people would buy vehicles they can afford.

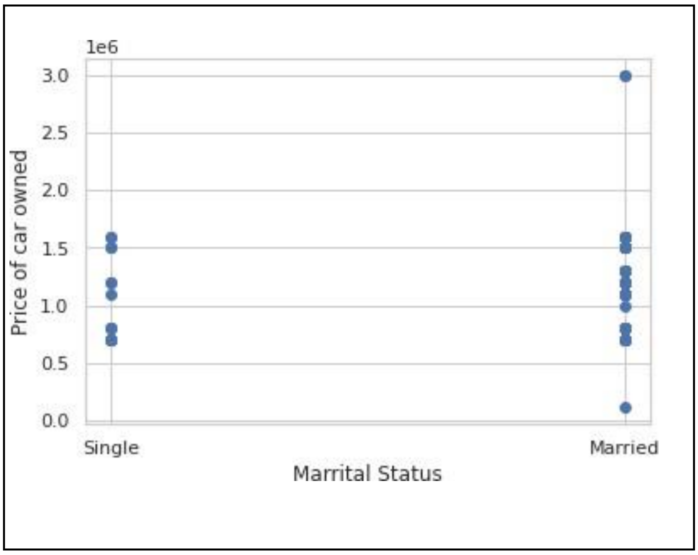
Dependency of make and price of vehicles on other descriptor variables

1) Marital Status:

Make of vehicle they tend to purchase

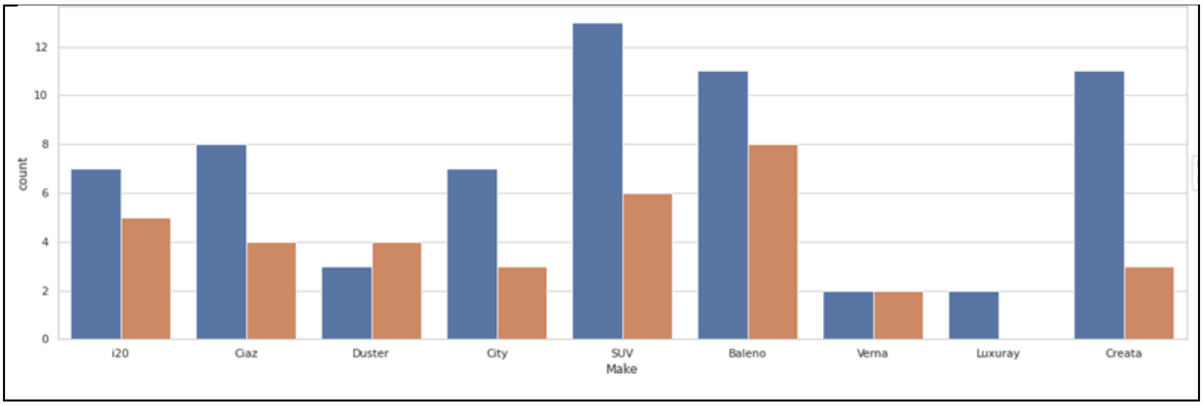


Price of vehicle they owned:

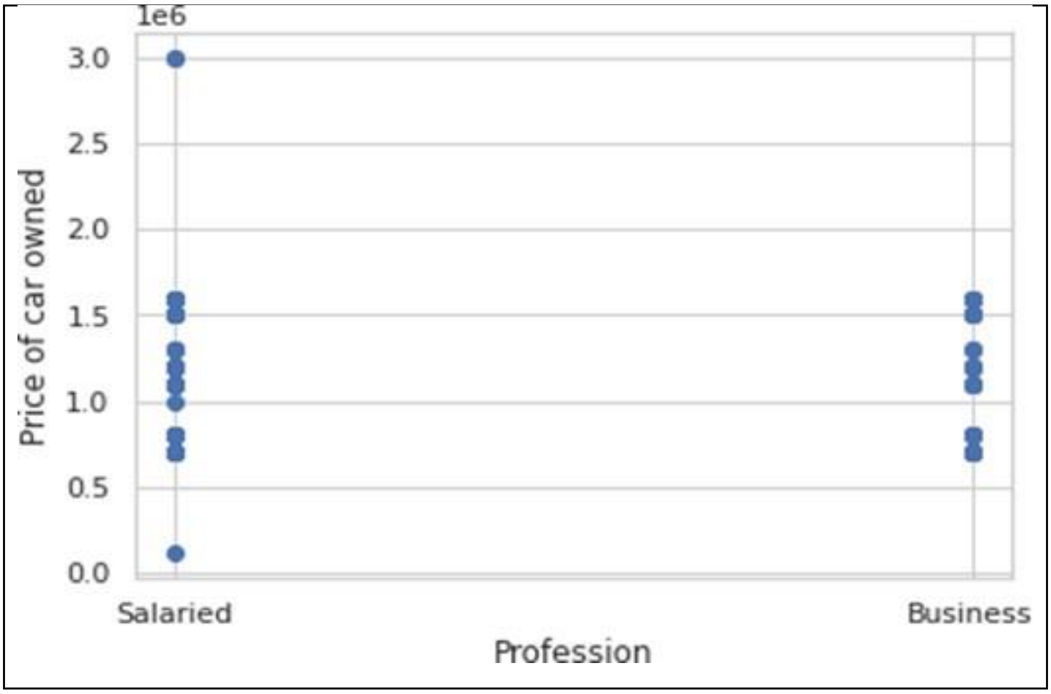


2) Profession:

Make of vehicle they tend to purchase:

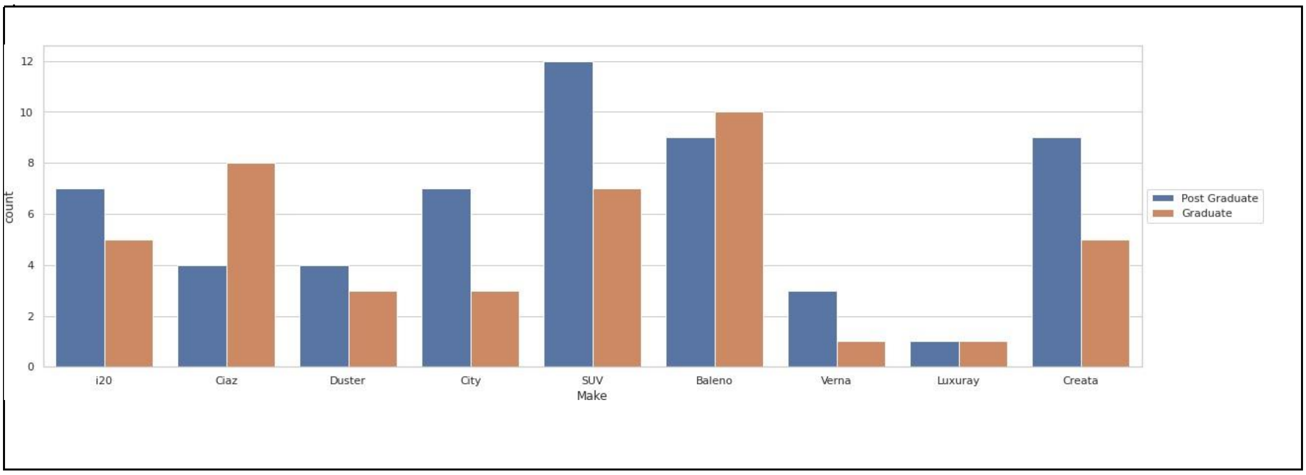


Price of vehicle owned:

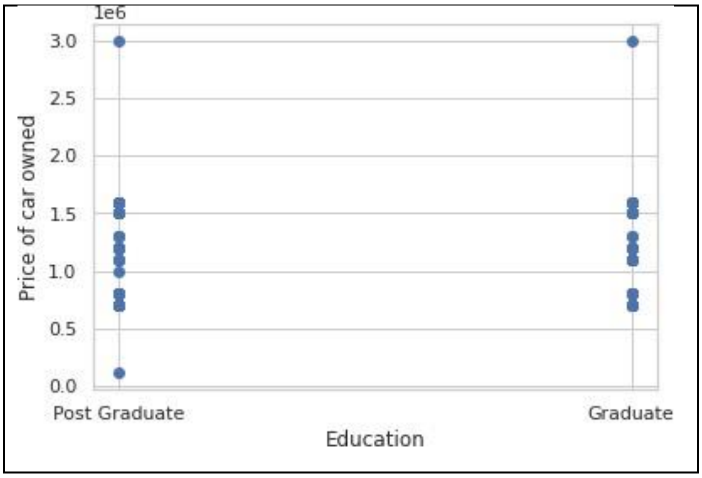


3) Education:

Make of vehicle they tend to purchase:

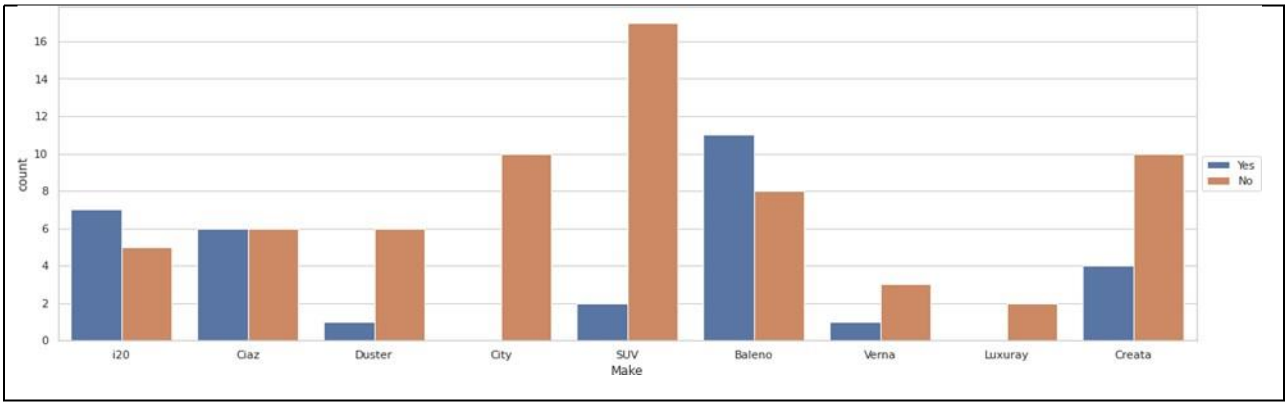


Price of vehicle owned:

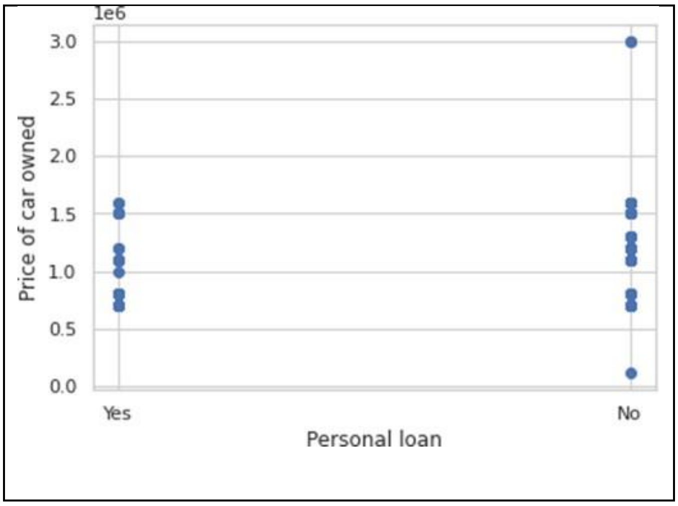


4) Personal Loan:

Make of vehicle they owned:



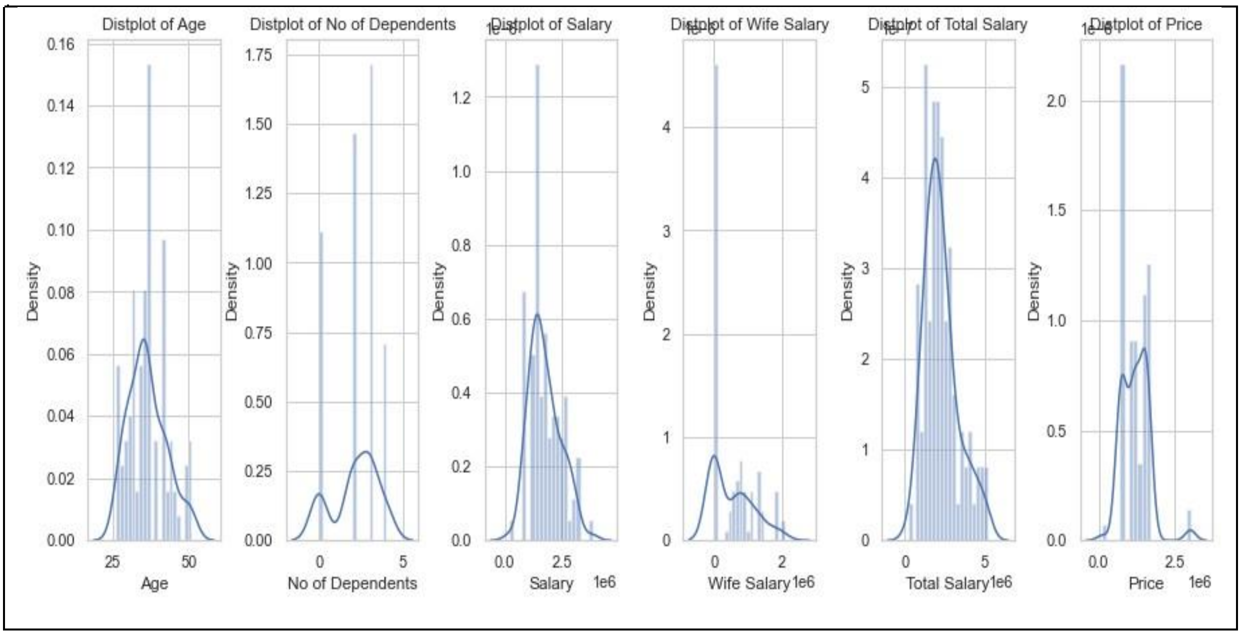
Price of Vehicle owned:



<u>Demographic Segmentation:</u>

Demographic segmentation is a market segmentation technique where an organization's target market is segmented based on demographic variables such as age, gender, education, income, etc. It helps organizations understand who their customers are so that their needs can be addressed more effectively. When an organization looks at the demographic segmentation, it focuses on the people who are most likely to buy a product. This helps in identifying the target market.

We have used the same dataset we used for behavioral and psychographic analysis and the following plots help us understand the socio-demographic structure of the market:



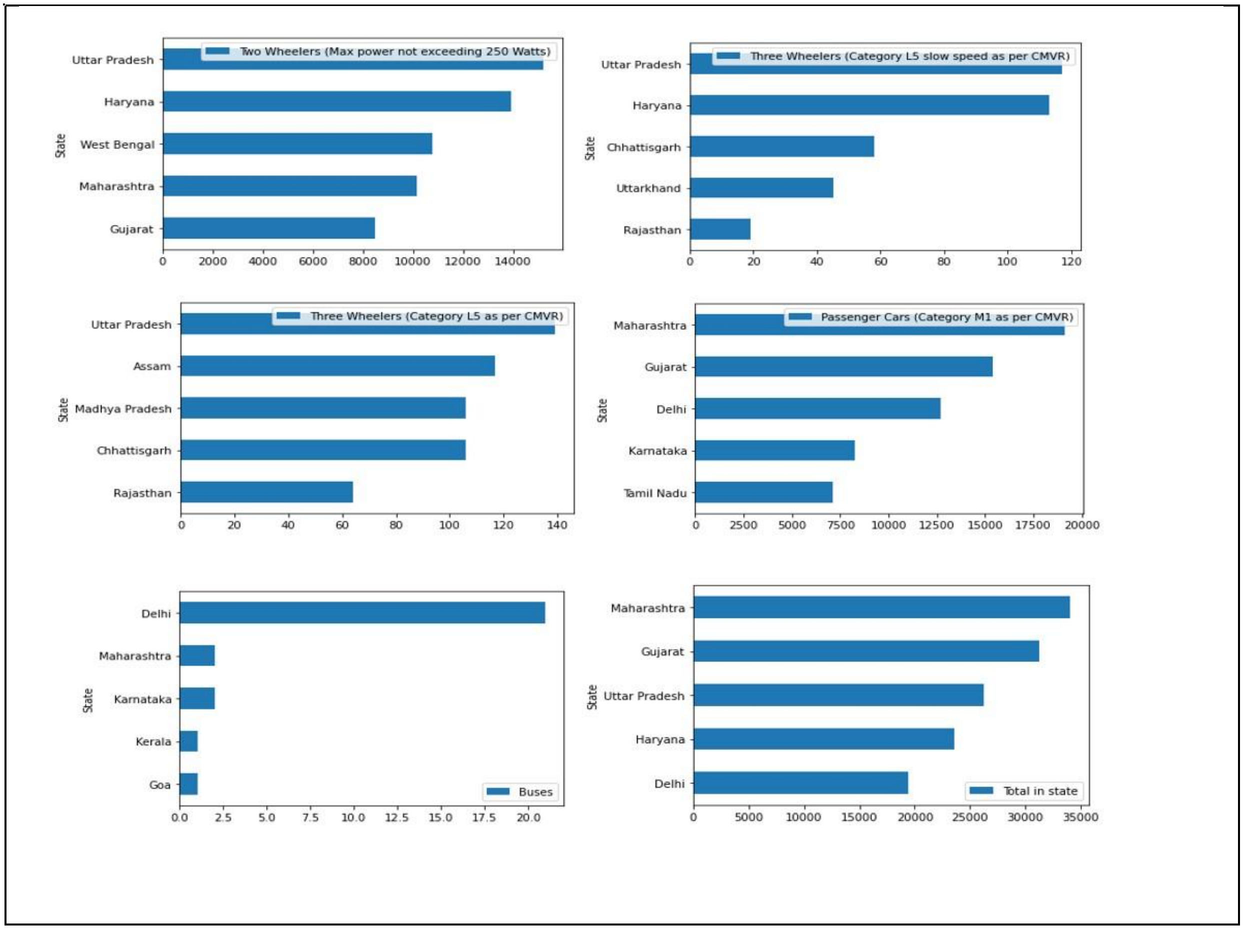
Geographic Segmentation:

Geographic segmentation is the process of dividing people into groups based on location, such as city, country, state, region, and even continent. It can help you tailor your approach during seasons customers may need your product.

For example, a fisherman in Alaska may only buy more equipment leading up to the salmon season. Whereas a fisherman in Orange Beach, Alabama, might purchase equipment all year round.

In contrast to other types of segmentation — demographic, psychographic, and behavioral — location-based segmentation analysis is easier to see results from. It doesn't take a lot of research to identify someone's location and the characteristics of a certain area, versus figuring out potential customers purchasing behaviors and psychographics. Here we have made divisions in terms of states and union territories in India.

For geographic analysis we used state-wise sales of different types of Electric Vehicles dataset which would help us understand our target region. Based on the type of electric vehicle, states with higher numbers of electric vehicles can be targeted as people in these states are more likely to purchase them. Given below are bar charts showing the top 5 states in sales of a particular EV type:



Depending on the type of Electric Vehicle the start-up comes with, it can target that particular state. What is important to consider is that for most of these electric vehicles that market would be a fairly developed city in that state, because consumers should be willing to purchase the electric vehicle and factors like cost versus average consumer income and the resources to charge the EV (e.g., Charging Stations) and being able to maintain it are important

Model fitting:

K-Means Clustering Algorithm

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. In this topic, we will learn what is K-means clustering algorithm, how the algorithm works, along with the Python implementation of k-means clustering.

What is K-Means Algorithm?

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of predefined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim oif this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

The algorithm takes the unlabeled dataset as input, divides the dataset into knumber of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

The k-means clustering algorithm mainly performs two tasks: ○ Determines the best value for K centre points or centroids by an iterative process.

○ Assigns each data point to its closest k-centre. Those data points which are near to the particular k-centre, create a cluster.

Data pre-processing: The

libraries used are:

- Numpy – For Computations
- Pandas – Manipulating the datasets
- Scikit-learn – For ML-based applications

```
In [1]: #Importing required libraries
        import numpy as np
        import pandas as pd
```

```
In [8]: #Label Encoding Categorical Variables
        from sklearn.preprocessing import LabelEncoder
```

With the help of this dataset, we can implement Behavioural, Psychographic and Demographic Segmentation of Indian Automobile Market. This helps us with understanding the various attributes leading to the consumer buying behaviour.

```
In [2]: #Loading the dataset
        df = pd.read_csv("Indian automoble buying behavour study 1.0.csv")
```

After uploading the 'Indian automobile buying behaviour study 1.0' dataset, we inspect the dataset. It has 99 rows and 13 columns.

```
In [4]: #Size of the dataset
        df.shape

Out[4]: (99, 13)
```

The various features include Age, Profession, Education, Number of dependents of the person buying the Vehicle and if the buyer is married or not, spouse is earning or not to ascertain the net income of the family. It further includes the Maker and the Price of the Vehicle of the buyer.

```
In [3]: #Inspecting the dataset
        df.head()
```

Out[3]:

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
0	27	Salaried	Single	Post Graduate	0	Yes	No	No	800000	0	800000	i20	800000
1	35	Salaried	Married	Post Graduate	2	Yes	Yes	Yes	1400000	600000	2000000	Ciaz	1000000
2	45	Business	Married	Graduate	4	Yes	Yes	No	1800000	0	1800000	Duster	1200000
3	41	Business	Married	Post Graduate	3	No	No	Yes	1600000	600000	2200000	City	1200000
4	31	Salaried	Married	Post Graduate	2	Yes	No	Yes	1800000	800000	2600000	SUV	1600000

We then use describe() function to gain in-depth insight about the mean, median and other statistical figures about the attributes of the dataset. This gives us an idea about the buyers' essential ranges.

In [5]:
```
#Statistics for Numerical Variables
df.describe()
```

Out[5]:

	Age	No of Dependents	Salary	Wife Salary	Total Salary	Price
count	99.000000	99.000000	9.900000e+01	9.900000e+01	9.900000e+01	9.900000e+01
mean	36.313131	2.181818	1.736364e+06	5.343434e+05	2.270707e+06	1.194040e+06
std	6.246054	1.335265	6.736217e+05	6.054450e+05	1.050777e+06	4.376955e+05
min	26.000000	0.000000	2.000000e+05	0.000000e+00	2.000000e+05	1.100000e+05
25%	31.000000	2.000000	1.300000e+06	0.000000e+00	1.550000e+06	8.000000e+05
50%	36.000000	2.000000	1.600000e+06	5.000000e+05	2.100000e+06	1.200000e+06
75%	41.000000	3.000000	2.200000e+06	9.000000e+05	2.700000e+06	1.500000e+06
max	51.000000	4.000000	3.800000e+06	2.100000e+06	5.200000e+06	3.000000e+06

There aren't any missing values or irrelevant attributes and thus needs no data handling.

In most of the algorithms, categorical values cannot be handled. Thus, there is a need to convert these categorical values to numerical levels. For this, we make use of the Label Encoder function.

In [8]:
```python
#Label Encoding Categorical Variables
from sklearn.preprocessing import LabelEncoder
labelen = LabelEncoder()
df['Profession'] = labelen.fit_transform(df['Profession'])
df['Marrital Status'] = labelen.fit_transform(df['Marrital Status'])
df['Education'] = labelen.fit_transform(df['Education'])
df['Personal loan'] = labelen.fit_transform(df['Personal loan'])
df['House Loan'] = labelen.fit_transform(df['House Loan'])
df['Wife Working'] = labelen.fit_transform(df['Wife Working'])
```

In [9]:
```python
#Encoded dataset
df.head()
```

Out[9]:

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
0	27	1	1	1	0	1	0	0	800000	0	800000	i20	800000
1	35	1	0	1	2	1	1	1	1400000	600000	2000000	Ciaz	1000000
2	45	0	0	0	4	1	1	0	1800000	0	1800000	Duster	1200000
3	41	0	0	1	3	0	0	1	1600000	600000	2200000	City	1200000
4	31	1	0	1	2	1	0	1	1800000	800000	2600000	SUV	1600000

ML algorithms work better when feature values are on relatively on a similar scale and close to normalized distribution. Using StandardScaler(), we scale the entire numerical portion of the dataset for an enhanced productive result.

In [10]:
```python
#Scaling the dataset
from sklearn.preprocessing import StandardScaler
scaled_df=df
scaled_df[["Age","Salary","Wife Salary","Total Salary","Price"]]=StandardScaler().fit_transform(df[["Age","Salary","Wife Salary",
scaled_df
```

Out[10]:

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price
0	-1.498630	1	1	1	0	1	0	0	-1.397118	-0.887055	-1.406760	i20	-0.904843
1	-0.211304	1	0	1	2	1	1	1	-0.501877	0.108995	-0.258937	Ciaz	-0.445579
2	1.397855	0	0	0	4	1	1	0	0.094950	-0.887055	-0.450240	Duster	0.013685
3	0.754191	0	0	1	3	0	0	1	-0.203464	0.108995	-0.067633	City	0.013685
4	-0.854967	1	0	1	2	1	0	1	0.094950	0.441012	0.314975	SUV	0.932213
...	...	...	...	...	...	...	...	...	...	...	...	...	...
94	-1.498630	0	1	0	0	0	0	0	0.990190	-0.887055	0.123671	SUV	0.932213
95	2.202434	1	0	1	3	0	0	1	3.079085	1.271054	2.706274	SUV	0.932213
96	2.363350	0	0	0	2	1	1	0	0.691777	-0.887055	-0.067633	Ciaz	-0.215947
97	2.363350	1	0	1	2	0	0	1	1.437811	1.271054	1.654102	Creata	0.702581
98	2.363350	1	0	1	2	1	1	0	0.691777	-0.887055	-0.067633	Ciaz	-0.215947

99 rows × 13 columns

Libraries used for the algorithm:

K - Means Algorithm

In [37]:
```python
# Importing Important Libraries
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
```

In [38]:
```python
X_scaled = StandardScaler().fit_transform(obj_df)
X_scaled = pd.DataFrame(X_scaled,columns=['Age', 'Profession', 'Marrital Status', 'Education', 'No of Dependents',
                                         'Personal loan', 'House Loan', 'Wife Working', 'Salary', 'Wife Salary',
                                         'Total Salary','Price'])
x = X_scaled.to_numpy()
X_scaled
```

Out[38]:

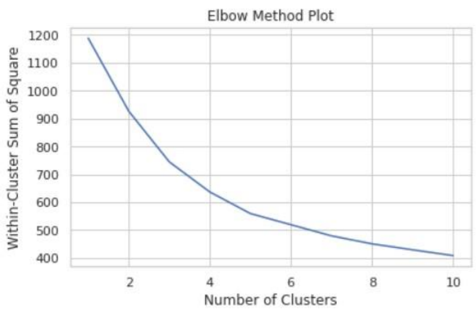
	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Price
0	-1.498630	-0.739510	-2.366432	0.876275	-1.642313	1.446980	-0.772512	-1.051847	-1.397118	-0.887055	-1.406760	-0.904843
1	-0.211304	-0.739510	0.422577	0.876275	-0.136859	1.446980	1.294479	0.950708	-0.501877	0.108995	-0.258937	-0.445579
2	1.397855	1.352247	0.422577	-1.141195	1.368594	1.446980	1.294479	-1.051847	0.094950	-0.887055	-0.450240	0.013685
3	0.754191	1.352247	0.422577	0.876275	0.615867	-0.691095	-0.772512	0.950708	-0.203464	0.108995	-0.067633	0.013685
4	-0.854967	-0.739510	0.422577	0.876275	-0.136859	1.446980	-0.772512	0.950708	0.094950	0.441012	0.314975	0.932213
...	...	...	...	...	...	...	...	...	...	...	...	...
94	-1.498630	1.352247	-2.366432	-1.141195	-1.642313	-0.691095	-0.772512	-1.051847	0.990190	-0.887055	0.123671	0.932213
95	2.202434	-0.739510	0.422577	0.876275	0.615867	-0.691095	-0.772512	0.950708	3.079085	1.271054	2.706274	0.932213
96	2.363350	1.352247	0.422577	-1.141195	-0.136859	1.446980	1.294479	-1.051847	0.691777	-0.887055	-0.067633	-0.215947
97	2.363350	-0.739510	0.422577	0.876275	-0.136859	-0.691095	-0.772512	0.950708	1.437811	1.271054	1.654102	0.702581
98	2.363350	-0.739510	0.422577	0.876275	-0.136859	1.446980	1.294479	-1.051847	0.691777	-0.887055	-0.067633	-0.215947

99 rows × 12 columns

In [39]:
```python
wcss = []

for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++',
                    max_iter = 300, n_init = 10,random_state = 0)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)
```

In [40]:
```python
plt.plot(range(1, 11), wcss)
plt.title('Elbow Method Plot')
plt.xlabel('Number of Clusters')
plt.ylabel('Within-Cluster Sum of Square') # Within cluster sum of squares
plt.tight_layout()
plt.show()
```



Either take K=3 or K= 5

1) when k= 3

k = 3

In [43]:
```
kmeans = KMeans(n_clusters = 3, init = 'k-means++',
                    max_iter = 300, n_init = 10,random_state = 42)
kmeans.fit(X_scaled)
```

```
C:\Users\dhruv\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:1332: UserWarning: KMeans is known to have a memory leak
on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment variable OM
P_NUM_THREADS=1.
  warnings.warn(
```

Out[43]:
```
KMeans
KMeans(n_clusters=3, random_state=42)
```

In [44]:
```
y = kmeans.predict(X_scaled)
y_df = pd.DataFrame(y,columns=['Class'])
```

In [45]:
```
final_data = pd.concat([df,y_df],axis=1)
final_data
```

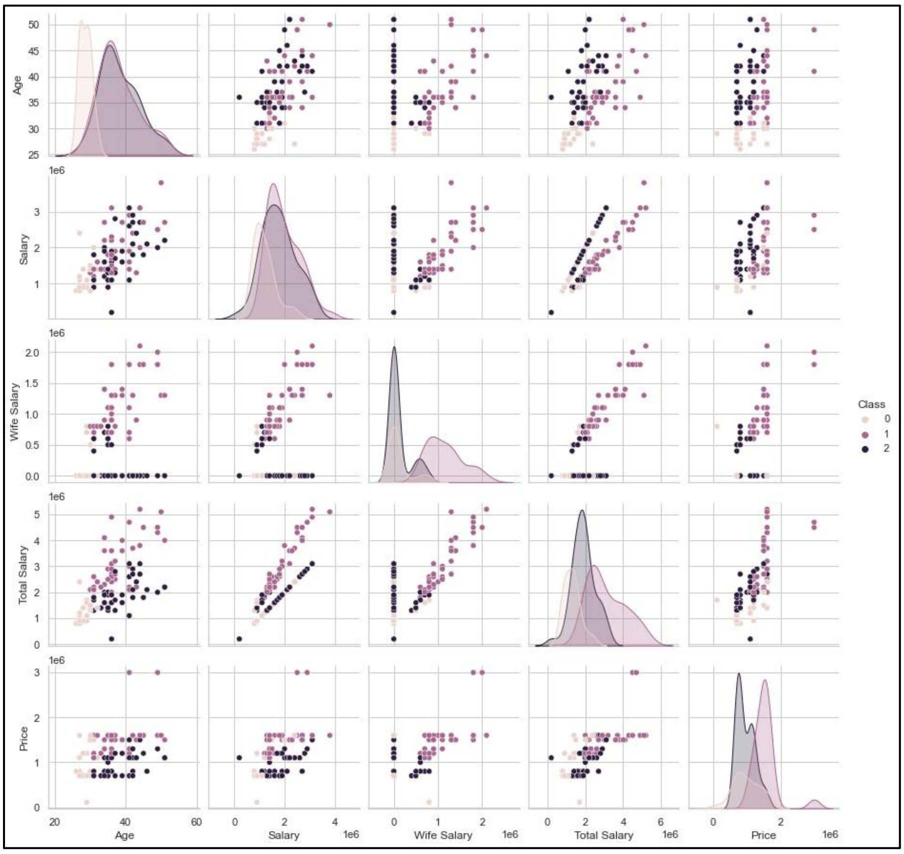
Out[45]:

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price	Class
0	27	Salaried	Single	Post Graduate	0	Yes	No	No	800000	0	800000	i20	800000	0
1	35	Salaried	Married	Post Graduate	2	Yes	Yes	Yes	1400000	600000	2000000	Ciaz	1000000	2
2	45	Business	Married	Graduate	4	Yes	Yes	No	1800000	0	1800000	Duster	1200000	2
3	41	Business	Married	Post	3	No	No	Yes	1600000	600000	2200000	City	1200000	1

Class 0: total salary equals to husband salary

Class 1: total salary is greater than husband salary

Class 2: total salary is nearly equal and greater than husband salary



2) when k= 5

k = 5

In [48]:
```
kmeans1 = KMeans(n_clusters = 5, init = 'k-means++',
                  max_iter = 300, n_init = 10,random_state = 42)
kmeans1.fit(X_scaled)
```

```
C:\Users\dhruv\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:1332: UserWarning: KMeans is known to have a memory leak
on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment variable OM
P_NUM_THREADS=1.
  warnings.warn(
```

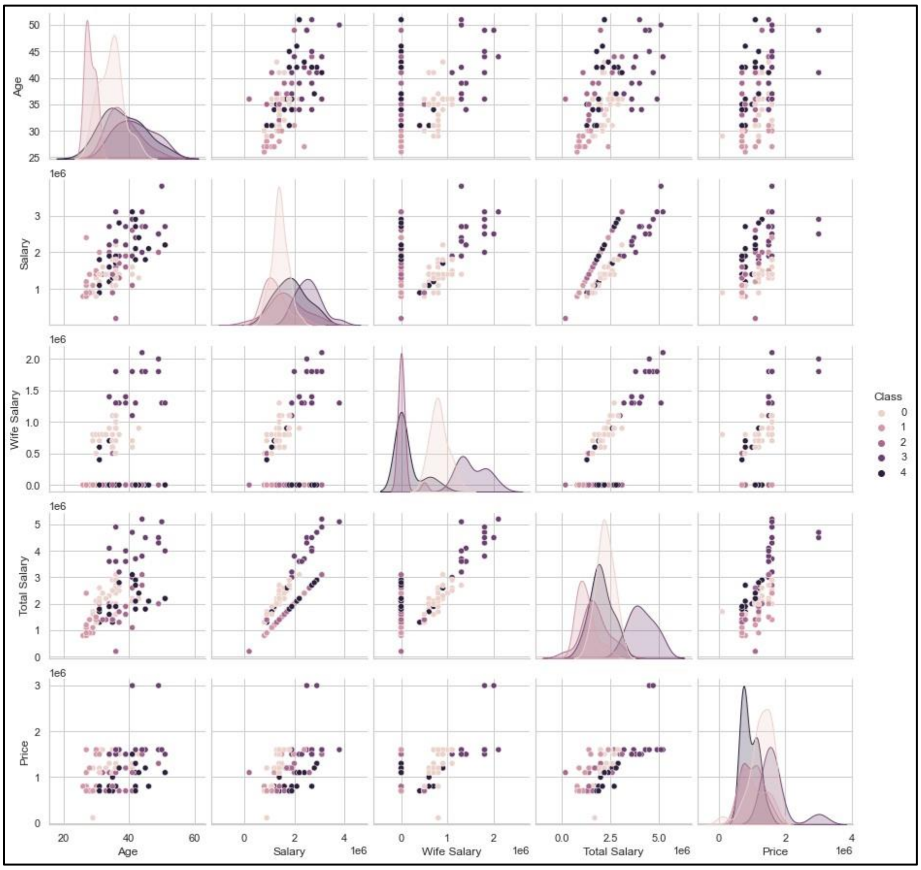
Out[48]:
```
KMeans
KMeans(n_clusters=5, random_state=42)
```

In [49]:
```
y1 = kmeans1.predict(X_scaled)
y1_df = pd.DataFrame(y1,columns=['Class'])
```

In [50]:
```
final_data1 = pd.concat([df,y1_df],axis=1)
final_data1
```

Out[50]:

	Age	Profession	Marrital Status	Education	No of Dependents	Personal loan	House Loan	Wife Working	Salary	Wife Salary	Total Salary	Make	Price	Class
0	27	Salaried	Single	Post Graduate	0	Yes	No	No	800000	0	800000	i20	800000	1
1	35	Salaried	Married	Post Graduate	2	Yes	Yes	Yes	1400000	600000	2000000	Ciaz	1000000	4
2	45	Business	Married	Graduate	4	Yes	Yes	No	1800000	0	1800000	Duster	1200000	4



To reduce the homogeneity, we chose k= 3

Potential Customer Base * Your Target Price Range = Potential Profit

For Electric 2-Wheeler

The per unit average price will be 1 lakh and the number of units sold will be around 5,00,000.

Potential Profit = 500000 *100000 = 50 billion

For Electric 3-Wheeler

The per unit Avg. price will be 2 lakhs and the number of units sold will be around 90000

Potential Profit = 90000*200000 = 18 billion

Optimal market segments to open in the market

From the above report, we conclude that to create an Electric Vehicle startup in India, the most optimal market segment for us will be based on Geographic and Demographic segments which would be the most amount of EVS sold in particular states and the type of electric vehicle respectively.

After analysing the EV market using Market Segmentation Analysis, the feasible strategy which we have come up with is that we will be focussing on the states that have more demand for EVs like Maharashtra and Gujrat. Also, one more reason to set the startup in these 2 states is that the infrastructure required for the EVS including the charging station is available which would ease our setting up of the startup process.

Marketing Mix with regard to EV:

PRICE

Affordability is the number one issue for any vehicle, more so in the case of EVs. The more cost efficient a product is the more it's sale. We can see from the above analysis that the product's price should ideally range between 10 to 20 lakh, as most people would make a purchase in this range.

PRODUCT

Product totally depends on the start-up, it's design, it's mechanics. Having said that, in general, if an EV start-up has to get successful in India, it's key would be to get into 2-wheeler EV business.

Another type of product EV Start up can look into is public transport vehicles, because the current government policies are supportive for revamping public transport to electric-based engines

PLACE

Major cities of the country (Especially metropolitan cities) should be targeted as these are the places where infrastructure would support. Another reason for targeting urban cities is that here it is more likely to have an aware population willing to buy Electric Vehicles.

For different types of vehicles, the list of top states which will promise a good market have been given in our geographical analysis.

PROMOTION

In EV business, awareness is the key, with its edge over fuel using vehicles, more and more people should be made aware of its advantages.