PROJECT 1-R REPORT

Analyzing Market Segmentation for Electric Vehicle Market Entry in India & Online Vehicle Booking Analysis in India

By ~ Vijay Patil

Table of Contents ~

1.0 Abstract

The present paper offers a thorough examination of market segmentation, which is a crucial element in devising a strategic plan for an Electric Vehicle (EV) startup operating in the ever-changing and dynamic Indian market. Finding the most viable market niches for the effective launch and uptake of electric vehicles in India is the main goal of this research.

Geographic, demographic, psychographic, and behavioral aspects are all included in the segmentation study, with a particular emphasis on early adopters of new technologies and areas where sustainable transportation innovation is most likely to be welcomed. Important insights into the market environment can be gained by carefully examining data sources, such as EV market data, data on general vehicle types, data on charging stations, and statistics on vehicle usage.

In order to identify regions in India that correspond with the traits of early adopters of technology, the report goes deeper into the life cycle of technology adoption. This facilitates the establishment of an early market. When complete datasets are not easily accessible, data collection techniques are carefully thought out to guarantee impartial and accurate decision-making.

This research presents a strategic pricing range for EV products based on the psychographics of the early market in addition to market segmentation. The startup leverages its emphasis on environmental conscience, innovation, and sustainability to project a positive image in the eyes of prospective clients.

The report also emphasizes the significance of marketing and promotion, which place a strong emphasis on convenience and environmental advantages. In order to boost the usage of electric vehicles, it encourages partnerships with local organizations and government support.

The recommended approach's essential element is continuous data analysis, which guarantees that the company is adaptable and sensitive to client feedback, market conditions, and the shifting Indian EV industry landscape.

Serving as a cornerstone for the startup's entry strategy into the Indian electric vehicle business, this project report offers vital insights into market segmentation, early market discovery, and the execution of a successful market entry plan.

2.0 Market Segmentation ~

We will be segmenting the market for our strategies according to the available data with different segmentation constraints ~

2.1 Geographic Segmentation ~

At the initiation we will be segmenting or analyzing the data about the EV market in India, for knowing the active use of electric vehicles in different states we will be using one dataset with relevant information about the active EV used in different states and union territories.

States and UT's actively using Electric Vehicle ~

```
import pandas as pd
import matplotlib.pyplot as plt
data = {
      "State/UT": ["Andaman and Nicobar Island", "Arunachal Pradesh",
"Assam", "Bihar", "Chandigarh", "Chhattisgarh", "Delhi", "Goa", "Gujarat",
"Haryana", "Himachal Pradesh", "Jammu and Kashmir", "Jharkhand",
"Karnataka", "Kerala", "Ladakh", "Maharashtra", "Manipur", "Meghalaya",
"Mizoram", "Nagaland", "Odisha", "Puducherry", "Punjab", "Rajasthan",
"Sikkim", "Tamil Nadu", "Tripura", "Dadra and Nagar Haveli and Daman and
Diu", "Uttarakhand", "Uttar Pradesh", "West Bengal"],
    "Total EVs (08.12.2021)": [157, 20, 43707, 58655, 1791, 11998, 126111,
1315, 13270, 24379, 623, 1334, 11060, 72018, 12109, 6, 52159, 519, 31, 18,
53, 10001, 1393, 8190, 47480, 22, 45368, 7146, 123, 23524, 258105, 43432]
}
df = pd.DataFrame(data)
```

```
plt.figure(figsize=(10, 6))
plt.bar(df["State/UT"], df["Total EVs (08.12.2021)"])
plt.xlabel("State/UT")
plt.ylabel("Total EVs as of 08.12.2021")
plt.title("Geographic Segmentation of Active Electric Vehicles")
plt.xticks(rotation=45, ha="right")
plt.tight_layout()

plt.show()
```

The code above helps us to visualize the max number of electric vehicle using states and union territories in india as a result we get a bar graph as shown below ~

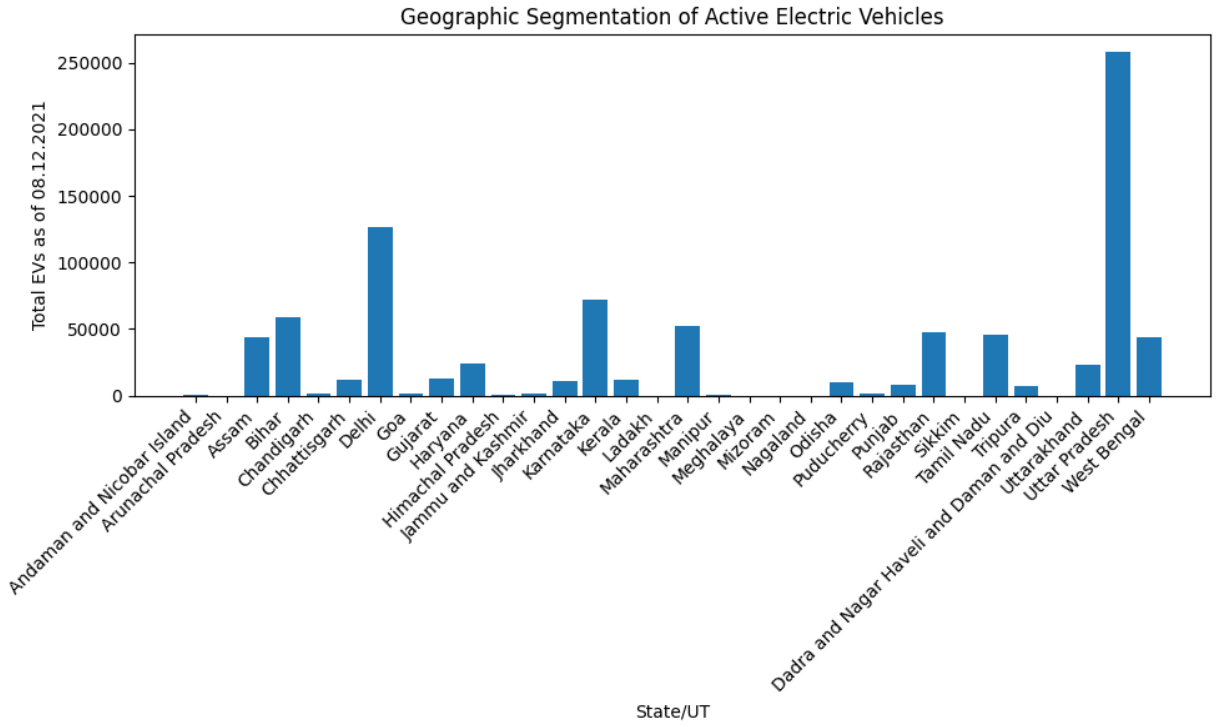


Figure 1 - Geographic Segmentation

The above visualization shows some of the clear picture about the EV market and active EV uses in the different states and union territories in India where we can have insightful information for our strategy planning where Uttar Pradesh has the max number of active electric vehicles in use.

In accordance to the Technology Adoption Life Cycle we can use certain constraints in the code to sort the data out for the best location for the EV startup ~

```
import pandas as pd
import matplotlib.pyplot as plt

data = {
       "State/UT": ["Andaman and Nicobar Island", "Arunachal Pradesh",
"Assam", "Bihar", "Chandigarh", "Chhattisgarh", "Delhi", "Goa", "Gujarat",
"Haryana", "Himachal Pradesh", "Jammu and Kashmir", "Jharkhand",
"Karnataka", "Kerala", "Ladakh", "Maharashtra", "Manipur", "Meghalaya",
"Mizoram", "Nagaland", "Odisha", "Puducherry", "Punjab", "Rajasthan",
"Sikkim", "Tamil Nadu", "Tripura", "Dadra and Nagar Haveli and Daman and
Diu", "Uttarakhand", "Uttar Pradesh", "West Bengal"],
    "Total EVs (08.12.2021)": [157, 20, 43707, 58655, 1791, 11998, 126111,
1315, 13270, 24379, 623, 1334, 11060, 72018, 12109, 6, 52159, 519, 31, 18,
53, 10001, 1393, 8190, 47480, 22, 45368, 7146, 123, 23524, 258105, 43432]
}

df = pd.DataFrame(data)
df = df.sort_values(by="Total EVs (08.12.2021)", ascending=False)

plt.figure(figsize=(10, 6))
plt.bar(df["State/UT"], df["Total EVs (08.12.2021)"])
plt.xlabel("State/UT")
plt.ylabel("Total EVs as of 08.12.2021")
plt.title("Geographic Segmentation of Active Electric Vehicles")
plt.xticks(rotation=45, ha="right")
plt.tight_layout()

plt.show()
most_suitable_location = df.iloc[0]["State/UT"]
print("Most Suitable Location for Early Market: ", most_suitable_location)
```

And as according to the code we get the result which is displayed below and we also get a clear visualization of which location is best to start a startup.

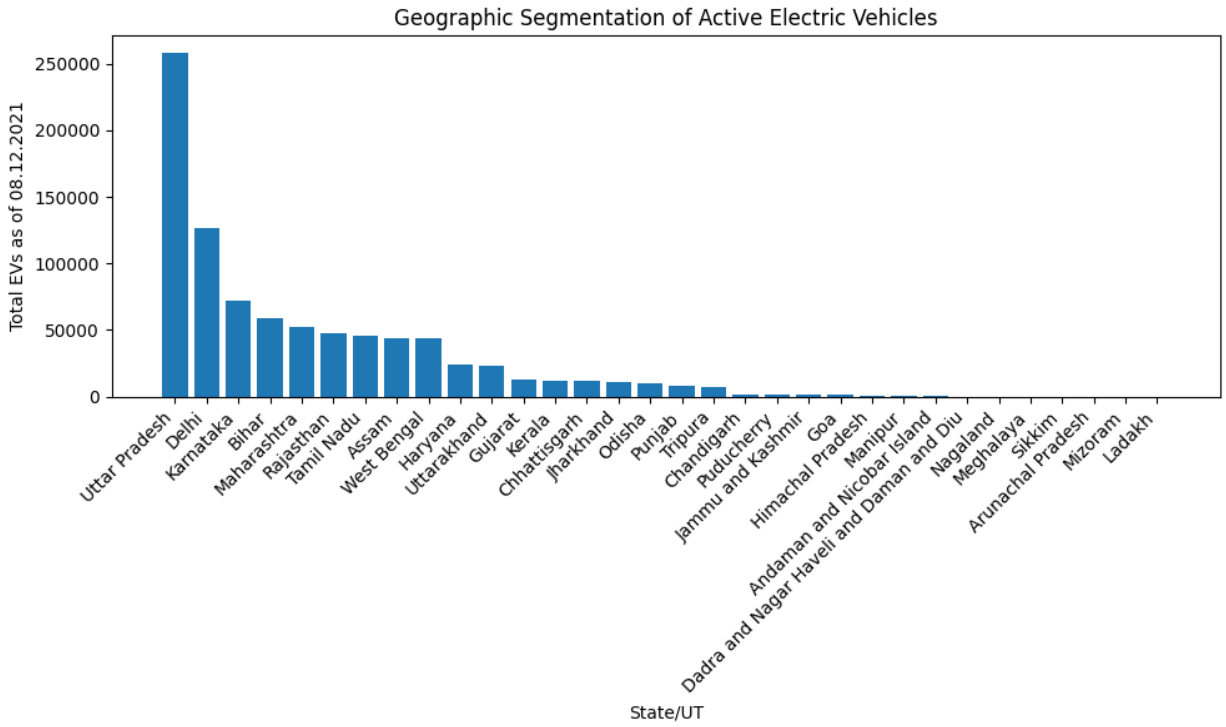


Figure 2 - Geographic segmentation using the Innovation Adoption Life Cycle concept.

2.2 Psychographic Segmentation ~

The EV market in India also depends on the type or the category of area on the basis of the psychographic analysis where we can also use this data to analyze the market of EV in India in terms of which type of vehicle does the Indian market have and what's their price range.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('evindia.csv')
data.head()
data.describe()
plt.figure(figsize=(10, 6))
plt.bar(data['Car'], data['PriceRange'])
plt.xlabel('Car')
plt.ylabel('Price Range')
plt.title('Price Range of Different Electric Vehicles')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

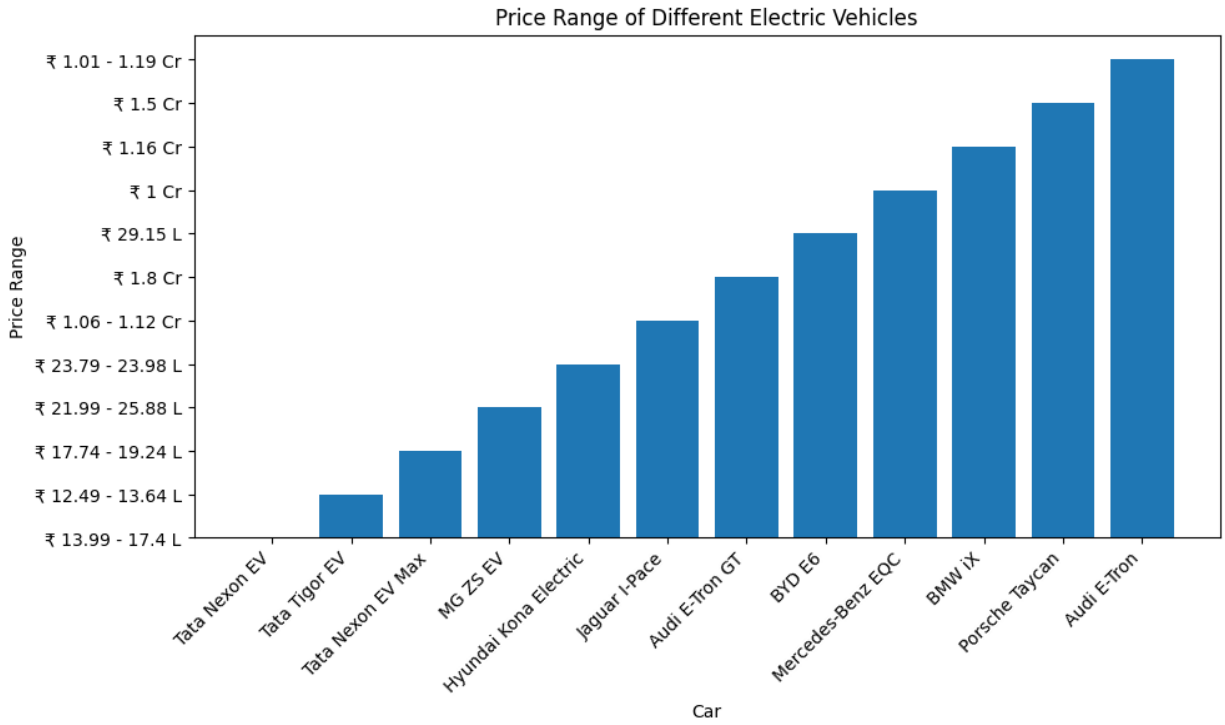


Figure 3 - Psychographic data of Indian EV cars market with EV type & their prices.

Also due to some unavailability of the data specifically orienting for these constraints we have or can create some synthetic data where we can analyze psychographic patterns and constraints for the project and analysis part.

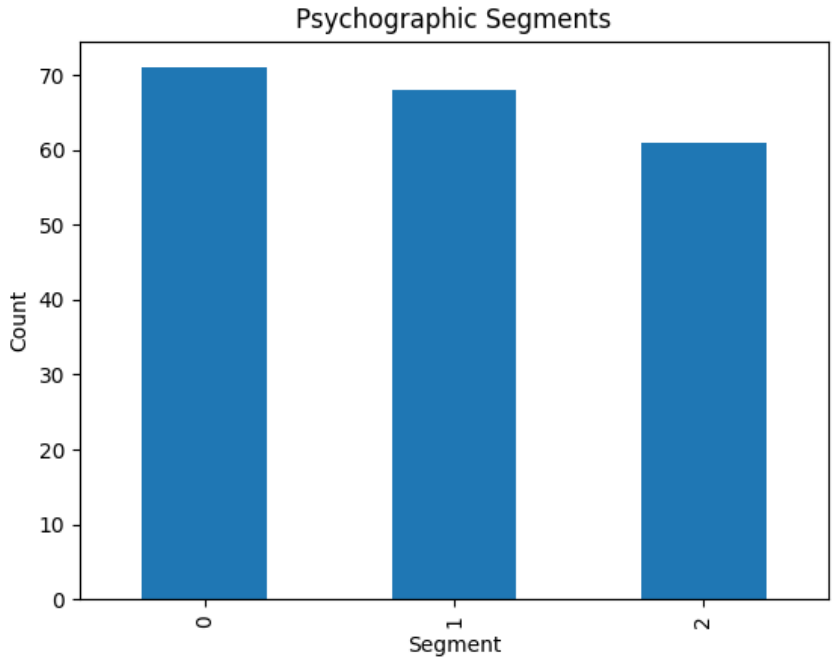
```
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
import random
random.seed(0)
np.random.seed(0)
n_samples = 200
data = {
    'Income': np.random.randint(20000, 100000, n_samples),
    'Age': np.random.randint(18, 75, n_samples),
    'Environment_Friendly': np.random.rand(n_samples),
    'Tech_Savvy': np.random.rand(n_samples),
}
df = pd.DataFrame(data)
```

```
n_clusters = 3
kmeans = KMeans(n_clusters=n_clusters, random_state=0).fit(df)
df['Segment'] = kmeans.labels_
print(df.head())
df.to_csv('psychographic_data.csv', index=False)
```

	Income	Age	Environment_Friendly	Tech_Savvy	Segment
0	88268	19	0.004048	0.019301	2
1	63567	35	0.269479	0.979573	0
2	62613	66	0.410492	0.148478	0
3	65891	53	0.428224	0.258702	0
4	41243	74	0.297842	0.215529	1
..	...	...	...	...	...
195	83413	69	0.696077	0.399660	2
196	68208	40	0.378326	0.530860	0
197	50142	72	0.100032	0.085233	0
198	52904	25	0.960257	0.665678	0
199	46813	59	0.072434	0.029635	1

Figure 3 - Synthetic Data Creation

Now we have created synthetic data based on the research done for the project to create or generate a dataset for our problem statement. Now let's visualize the above dataset and also lets take this dataset accordingly.



Now based on the above visualization lets take this in a note for further analysis purposes.

3.0 Strategy Analysis

With the above available data and visualization of the data we can have some strategic results where we can select the market space for our startup and and could apply for the strategic start of our startup we can also get some insights from the data available. We can use a variety of machine learning algorithms to assess the projections and plan for the project of joining the Indian electric vehicle (EV) industry. The particular kind of research you wish to perform will determine the model you should choose.

Our research is based on the geographic dataset more priorly than the psychographic or demographic data because we have a synthetic dataset for our project which we have collected with our research from different platforms and sources mentioned in the reference section of our project, so as per the available data and its feature we have to use a machine learning model which is compatible and will suit the dataset, so for this we will be going with Random Forest Classifier as Geographical and psychographic data are both excellent candidates for the flexible and potent Random Forest model classification algorithm, which can handle a variety of data types. Several decision trees are combined in this ensemble learning technique to generate predictions.

For the above we will be needing to prepare the data for our project so we will be using our most firm data which is the geographical data to analyze the situation for our result to be accurate.

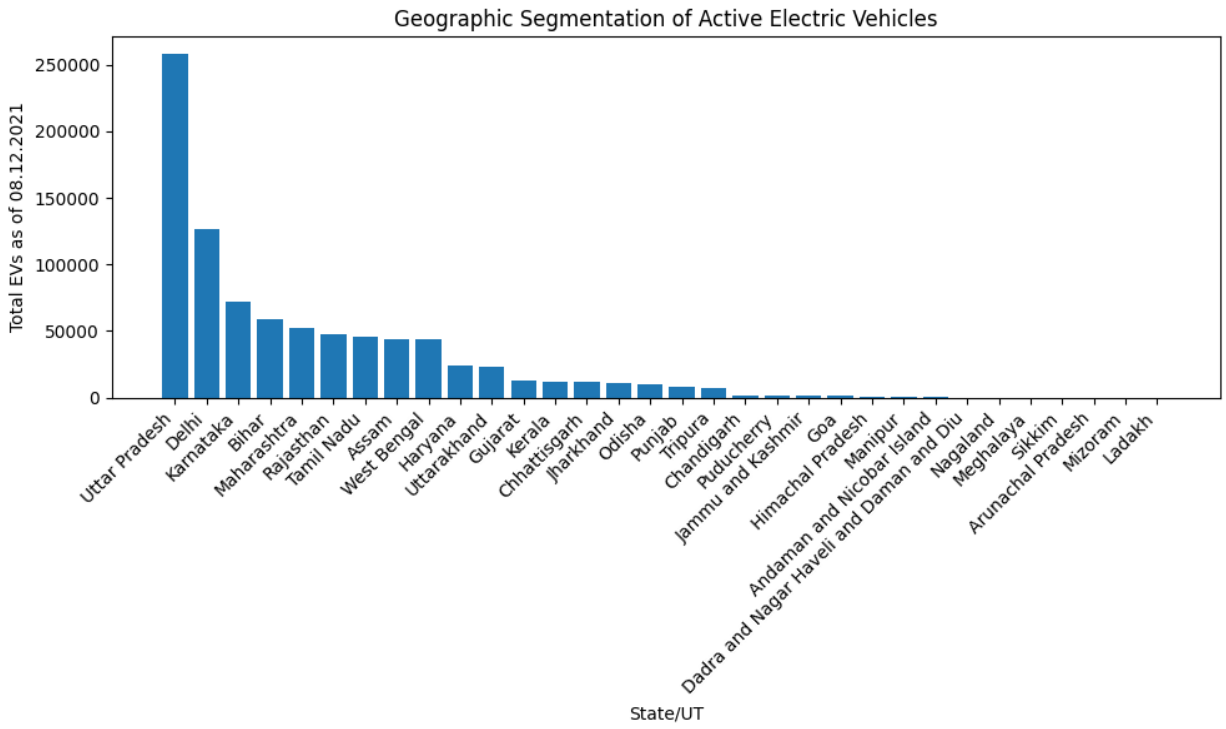
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
data = pd.read_csv('evmarketdata.csv')
X = data.drop('EV_Sales', axis=1)
y = data['EV_Sales']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
```

```
print(f"Accuracy: {accuracy}")
predicted_sales = clf.predict(X_new)
```

Above is a basic code for the prediction of best geographical data analysis for our data which wil be providing an accurate result for the problem.

4.0 Conclusion

Here we can conclude that as per the data web visualization and segmentation we can evaluate that we can have a market set up at Uttar Pradesh as the state has more number of ev users and demand and as per the data we can start plotting the branches at several places according to the visualization.



Online Vehicle Booking Market Analysis

5.0 Abstract

The route for a new player into India's fiercely competitive car rental sector is laid out in this strategy plan. This firm wants to carve out a niche by focusing on segmentation analysis and a data-driven approach to get an early market foothold and create revenue, despite the formidable presence of industry giants like Ola and Uber. The strategy encompasses a wide range of segmentations, including B2B and geographic, demographic, psychographic, and behavioral segments. When there are no complete datasets available, methods for reaching objective decisions are talked about. These methods include primary research, data analytics, and collaborations with research organizations.

The goal of the price plan is to remain competitive while luring early adopters. It consists of subscription plans, surge pricing, tiered pricing structures, dynamic pricing, and a referral scheme.

6.0 Importing Libraires

For segmentation these are some of the important libraries which we could be using for our project for making an accurate prediction ~

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

import warnings
warnings.filterwarnings('ignore')
```

The above libraries will be helping us for our project to evaluate an accurate result of prediction for our dataset.

7.0 Data Analyzing

Now at this stage lets first analyze our available data for the future reference and computation for our project ~

```
df = pd.read_csv('/kaggle/input/rideshare/rideshare_kaggle.csv')
```

```
df.shape
```

```
~ (693071, 57)
```

We can analyze the shape of our data which also describes how big our data is .

```
df.head()
```

Here we have analyzed some of our initial data for checking if the dataset is responsive or not.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 693071 entries, 0 to 693070
Data columns (total 57 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   id                      693071 non-null  object
 1   timestamp               693071 non-null  float64
 2   hour                    693071 non-null  int64
 3   day                     693071 non-null  int64
 4   month                   693071 non-null  int64
 5   datetime                693071 non-null  object
 6   timezone                693071 non-null  object
 7   source                  693071 non-null  object
 8   destination             693071 non-null  object
 9   cab_type                693071 non-null  object
 10  product_id              693071 non-null  object
 11  name                    693071 non-null  object
 12  price                   637976 non-null  float64
 13  distance                693071 non-null  float64
 14  surge_multiplier        693071 non-null  float64
 15  latitude                693071 non-null  float64
 16  longitude               693071 non-null  float64
 17  temperature             693071 non-null  float64
 18  apparentTemperature     693071 non-null  float64
 19  short_summary           693071 non-null  object
 20  long_summary            693071 non-null  object
 21  precipIntensity         693071 non-null  float64
```

Here we have the whole info of our data for the prediction purposes and uses .

8.0 Exploratory Data Analysis

Now at this stage we will be doing some of the EDA on our dataset (Exploratory Data Analysis)

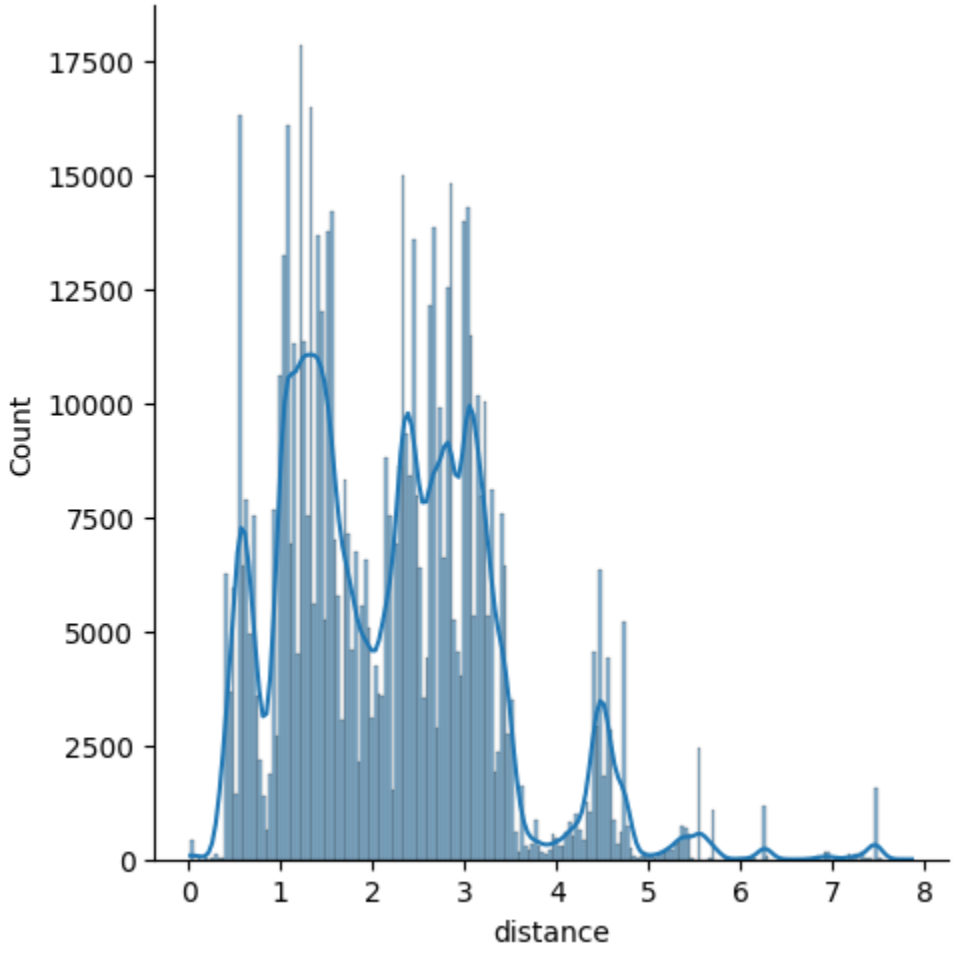
```
latest_df = df[['hour', 'day', 'month', 'datetime', 'timezone', 'source', 'cab_type', 'destination', 'latitude', 'name', 'distance', 'surge_multiplier', 'product_id', 'longitude', 'temperature', 'long_summary', 'short_summary']]
```

```
latest_df.head()
```

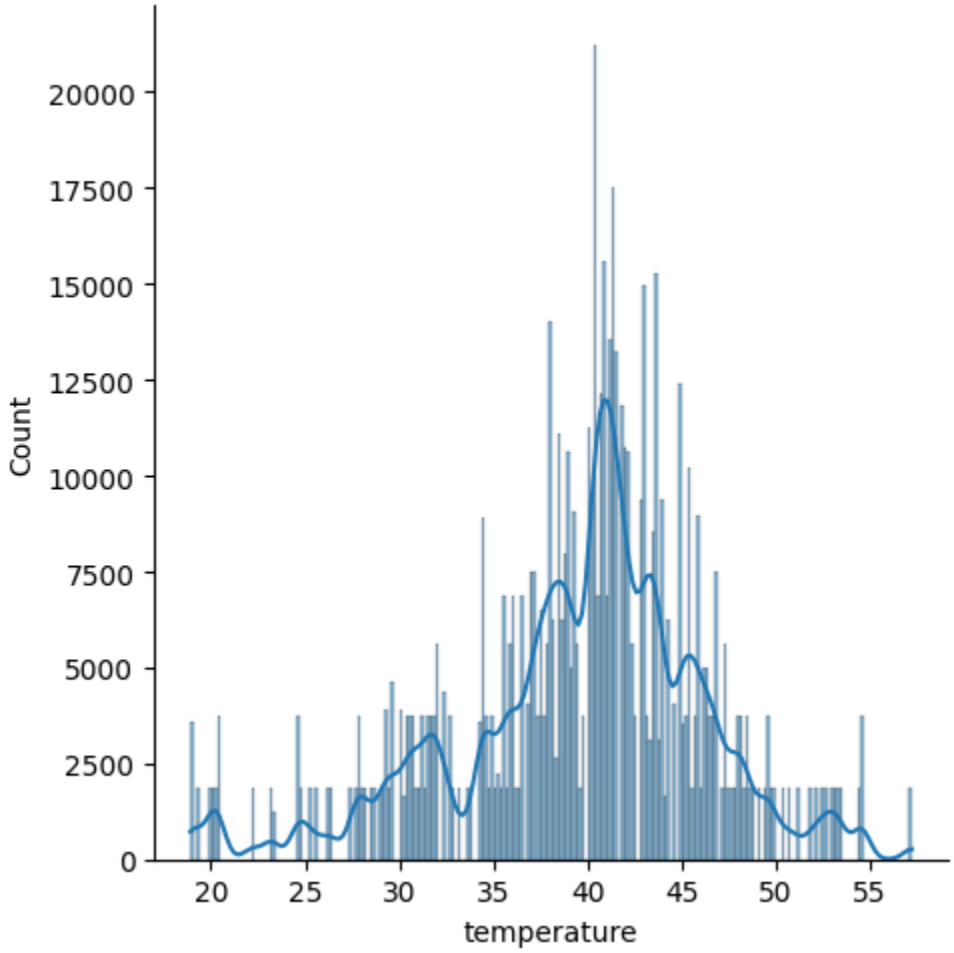
Here we have analyzed and checked our dataset. We have segregated some features which will be important for us to analyze.

Let's extract time and date features ~

```
new_df['datetime'] = pd.to_datetime(new_df['datetime'])
new_df['day_name'] = new_df['datetime'].dt.day_name()
```




```
sns.displot(new_df['temperature'], kde = True)
```



These were some of the data visualization lets do segmentation now

9.0 Segmentation

Let's do some of the segmentation for our data to predict the result.

```
df1 = pd.read_csv('/kaggle/input/sigma-cabs/sigma_cabs_kaggle.csv')
df1.head()
```

	Trip_ID	Trip_Distance	Type_of_Cab	Customer_Since_Months	Life_Style_Index	Confidence_Life_Style_Index
0	T0005689460	6.77	B	1.0	2.42769	A
1	T0005689461	29.47	B	10.0	2.78245	B
2	T0005689464	41.58	NaN	10.0	NaN	NaN
3	T0005689465	61.56	C	10.0	NaN	NaN
4	T0005689467	54.95	C	10.0	3.03453	B

Lets remove some non required features form the dataset

```
data1 = df1.drop(['Trip_ID', 'Var1', 'Var2', 'Var3'], axis = 1)
```

```
data1.head()
```

	Trip_Distance	Type_of_Cab	Customer_Since_Months	Life_Style_Index	Confidence_Life_Style_Index	Destination_Type	Customer_R
0	6.77	B	1.0	2.42769	A	A	3.90500
1	29.47	B	10.0	2.78245	B	A	3.45000
2	41.58	NaN	10.0	NaN	NaN	E	3.50125
3	61.56	C	10.0	NaN	NaN	A	3.45375
4	54.95	C	10.0	3.03453	B	A	3.40250

To drop the null values we will be doing ~

```
data1 = data1.dropna(axis = 0)
```

10.0 Data Pre-Processing

```
encoder = LabelEncoder()
for col in cat_cols:
    data1_copy[col] = encoder.fit_transform(data1[col])
```

```
data1_copy.head()
```

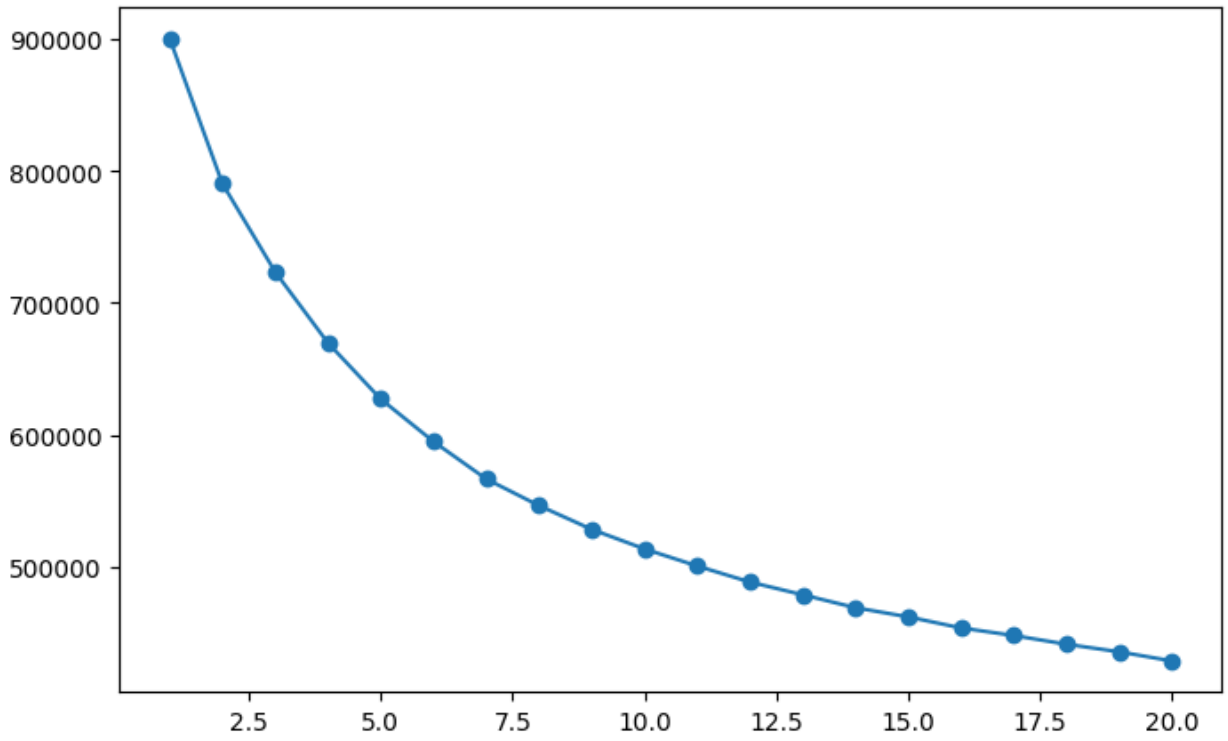
	Trip_Distance	Type_of_Cab	Customer_Since_Months	Life_Style_Index	Confidence_Life_Style_Index	Destination_Type	Customer_R
0	6.77	1	1.0	2.42769	0	0	3.9050
1	29.47	1	10.0	2.78245	1	0	3.4500
4	54.95	2	10.0	3.03453	1	0	3.4025
6	29.72	4	10.0	2.83958	2	1	2.9750
7	18.44	1	2.0	2.81871	1	0	3.5825

```
data1_scaled = pd.DataFrame()
scaler = StandardScaler()
data1_scaled[data1_copy.columns]                                    =
scaler.fit_transform(data1_copy[data1_copy.columns])


wcss = []
for cluster in range(1, 21):
    km = KMeans(n_clusters = cluster)
    km.fit_predict(data1_scaled)
    wcss.append(km.inertia_)
```

Let’s visualize and plot the data for analysis

```
plt.figure(figsize = (8, 5))
plt.plot(range(1, 21), wcss, marker = 'o')
```



11.0 Conclusion

We can conclude here that we have analyzed same data which we can use to predict our data set and to conclude our project with the result we have concluded that the geographic date segmentation is one of the most useful data to predict and conclude our market.

12.0 References

Kaggle

Government of India dataset

Github Link ~ https://www.github.com/vijaypatilarena