# MARKET SEGMENT ANALYSIS FOR ELECTRIC VEHICLE

Contributed by

Vuppu Uditha

**Github link:**

https://github.com/123uditha/Market-segmentation-analysis-on-Electric-vehicles

**Problem Statement**

Problem is to analyse the Electric Vehicle market in India so a startup can use and come up with a feasible strategy to enter in the market. Target segments are Geographic, Behavioural, Profession and so on. In this report, I analyse Electric_vechile market in India using 2wheelers, Passengers cars, Buses, 3 wheelers, age, profession, total salary, number of dependents of the family and so on.

**Fermi Technique**

Fermi is a famous technique for solving a problem. Steps to solve a problem using this technique are:

- Identify the core problem
- Break it down: divide the problem into smaller more manageable subproblems.
- Make educated guesses: for each subproblem make reasonable assumptions and estimations.
- Multiply the estimations.
- Consider the range of possible values.
- Refine and iterate.

**Identifying the core problem**

Analyse the electric vehicle market and come up with a feasible strategy to enter in the market.

**Break it down**

- What type of Electric Vehicle will the company produce?
- Who are target customers?

**Population of the India over 5 years**

- **2023:** 1.438 B
- **2022:** 1.425 B

- ➢ **2021:** 1.414 B
- ➢ **2020:** 1.402 B
- ➢ **2019:** 1.389 B

**Make assumptions**

- ➢ Let's assume that bikes are a more popular electric vehicle in India and are increasing with an average of 8-10 % every year.
- ➢ Uttar Pradesh is the most populated state in India, most industrialised state is Tamil nadu, Maharashtra is a place for commercial capital, IT capital is Benguluru and so on.
- ➢ Based on these assumptions Benguluru is one of the options to start a startup.
- ➢ Since we are targeting Benguluru our target customers would be people around 25-47.

**Data Sources**

Data was extracted from various sources

https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset

https://www.kaggle.com/datasets/adarshde/electric-vehicle-population-dataset

https://www.kaggle.com/datasets/gunapro/electric-vehicle-population-data

https://www.kaggle.com/datasets/fatihilhan/electric-vehicle-specifications-and-prices

Databases, Tables & Calculators by Subject

**Data Preprocessing**

Data preprocessing is a critical step in the data analysis and machine learning pipeline. It involves preparing raw data into a clean and usable format to improve the performance of models and ensure accurate analysis. Raw data collected from various sources is often incomplete, noisy, inconsistent, and contains missing or duplicate values. Preprocessing helps address these issues by transforming the data into a structured, standardized, and meaningful form.

- ➢ Ensures that models learn from relevant, high-quality data.
- ➢ Reduces bias and variance in the model.
- ➢ Increases the accuracy, efficiency, and reliability of results.
- ➢ Makes the dataset interpretable for humans and machines alike.

Libraries used are pandas, numpy.

```
print("Dataset 1 ------------------------->")
df1.info()
```

```
print("Dataset 2 ------------------------->")
df2.info()
```

Finding missing values

```
d1 = df1.isnull().sum()
d2 = df2.isnull().sum()
print("DF1:----------------------> ", d1)
print("DF2:------------------------------> ", d2)
```

Finding duplicate values

```
d1 = df1.duplicated().sum()
d2 = df2.duplicated().sum()
print("Dataset1 ---------------------->", d1)
print("Dataset2 ---------------------->", d2)
```

**Exploratory data analysis**

Exploratory Data Analysis (EDA) involves summarizing, visualizing, and understanding the main characteristics of a dataset to discover patterns, spot anomalies, and check assumptions before applying any machine learning models.

➢ Import necessary libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
```

➢ Understand the dataset

```
#Dataset1
DF1 = "C:\\Users\\dell\\Downloads\\EVStats.csv"
df1 = pd.read_csv(DF1)
df1.head()
```

```
#Dataset2
DF2 = "C:\\Users\\dell\\Downloads\\Buying.csv"
df2 = pd.read_csv(DF2)
df2.head()
```

```
#Shape of datasets
```

```
print("DF1 shape: ", df1.shape)
print("DF2 shape: ", df2.shape)
```

```
d1 = df1.describe().T
d2 = df2.describe().T
display("Dataset1 ---------------->", d1, "Dataset2
----------------------->", d2)
```

**Segment Extraction**

Segment extraction is the process of identifying, isolating, and extracting distinct groups (segments) or meaningful subsets from a dataset. These segments are typically formed based on specific characteristics, attributes, or behaviors of the data points. Segment extraction is commonly used in clustering, image processing, and data analysis.

➢ Segment extraction for dataset 1

```
data = df1
kmeans = KMeans(n_clusters=4, random_state=0)
kmeans.fit(data)
cluster_1 = data[kmeans.labels_ == 0]
cluster_2 = data[kmeans.labels_ == 1]
cluster_3 = data[kmeans.labels_ == 2]
cluster_4 = data[kmeans.labels_ == 3]

print("Cluster 1:", cluster_1)
print("Cluster 2:", cluster_2)
print("Cluster 3:", cluster_3)
print("Cluster 4:", cluster_4)
```

➢ Segment extraction for dataset 2

```
data = df2
kmeans = KMeans(n_clusters=4, random_state=0)
kmeans.fit(data)
cluster_1 = data[kmeans.labels_ == 0]
cluster_2 = data[kmeans.labels_ == 1]
cluster_3 = data[kmeans.labels_ == 2]
cluster_4 = data[kmeans.labels_ == 3]
```

```
print("Cluster 1:", cluster_1)
print("Cluster 2:", cluster_2)
print("Cluster 3:", cluster_3)
print("Cluster 4:", cluster_4)
```

**Conclusion**

In conclusion, we observe that

- ➢ There are more salaried persons than businessmen.
- ➢ Gujarat has more two wheelers than any other state.

**The startup has to focus on building two wheelers more and target customers of age 23 - 27.**