**Market Segmentation Analysis Report for Electric Vehicle Market in India**

By Adarsh Herle

4/7/2024

**Introduction**

This report presents the analysis of the Electric Vehicle (EV) market in India using segmentation analysis. The goal is to identify key market segments and develop a feasible strategy for an EV startup to target the segments most likely to use electric vehicles.

**Problem Statement**

As an EV startup, we need to determine the appropriate vehicle/customer space to develop our EVs. This analysis aims to segment the EV market in India and identify the most promising segments for market entry.

**Methodology**

The analysis involved several steps:

1. Data Collection and Preprocessing
2. Feature Engineering
3. Segmentation Analysis
4. Strategy Formulation

**Data Collection and Preprocessing**

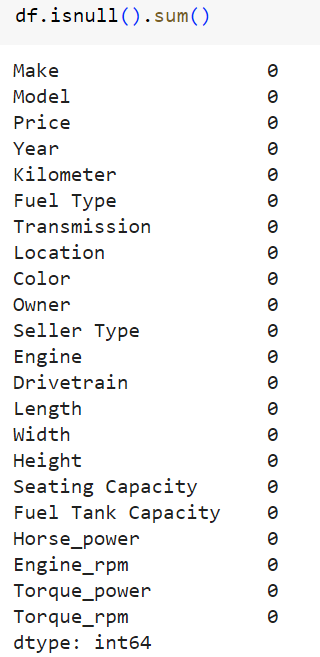
The dataset used for this analysis, car details v4.csv, contains various attributes of vehicles as well as locations. The preprocessing steps included:

1. **Handling Missing Values:**
   * Filling missing values in the "Max Power" and "Max Torque" columns with default values.
   * Splitting the "Max Power" and "Max Torque" columns into horsepower, engine RPM, and torque RPM.
   * Filling missing values in the "Engine", "Length", "Width", "Height", "Seating Capacity", and "Fuel Tank Capacity" columns with mean values.
2. **Data Cleaning:**
   * Removing unnecessary characters (e.g., "cc" from the "Engine" column).
   * Converting data types to appropriate formats for analysis.

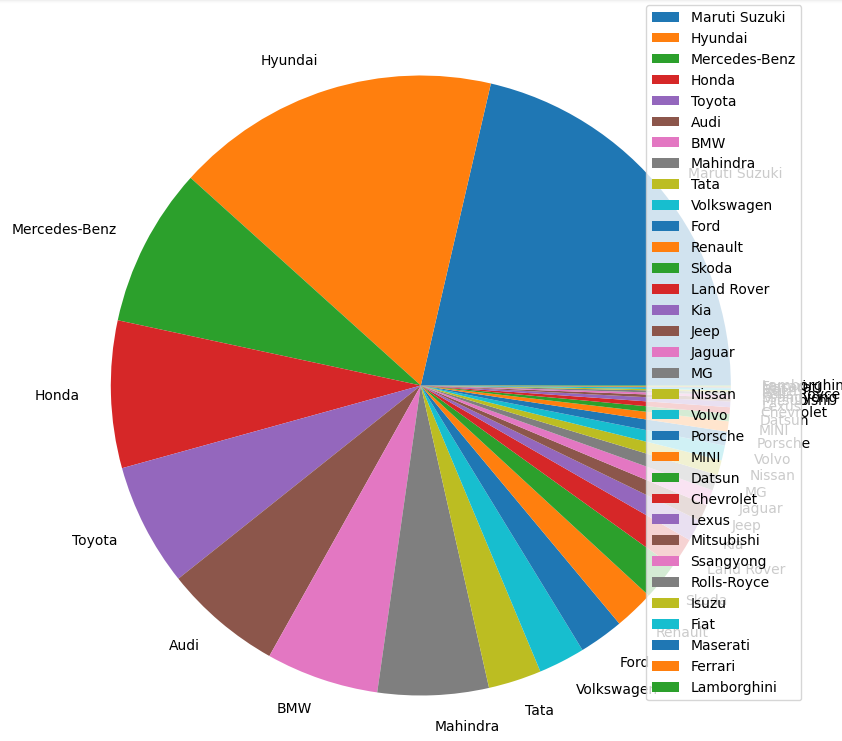
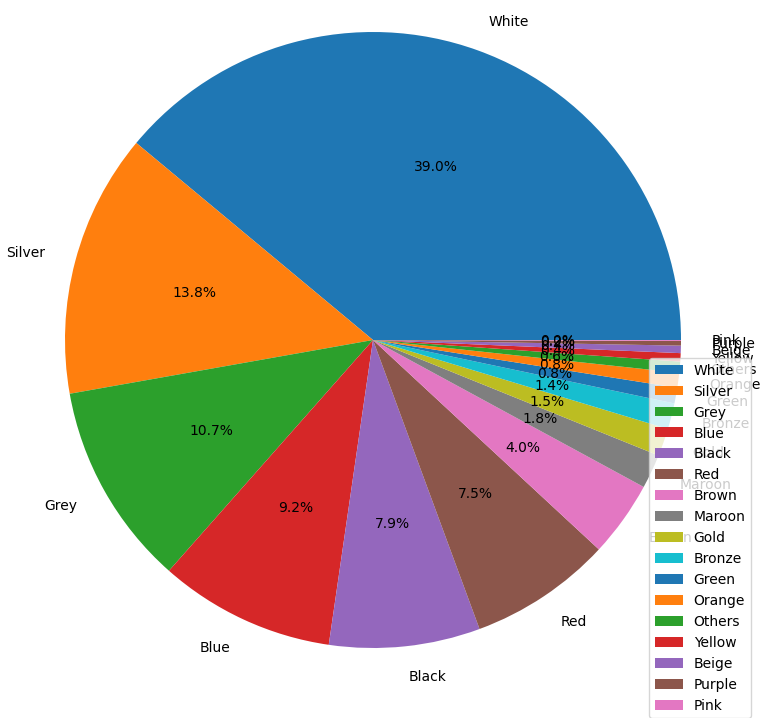
**Feature Engineering**

New features were created from the existing data:

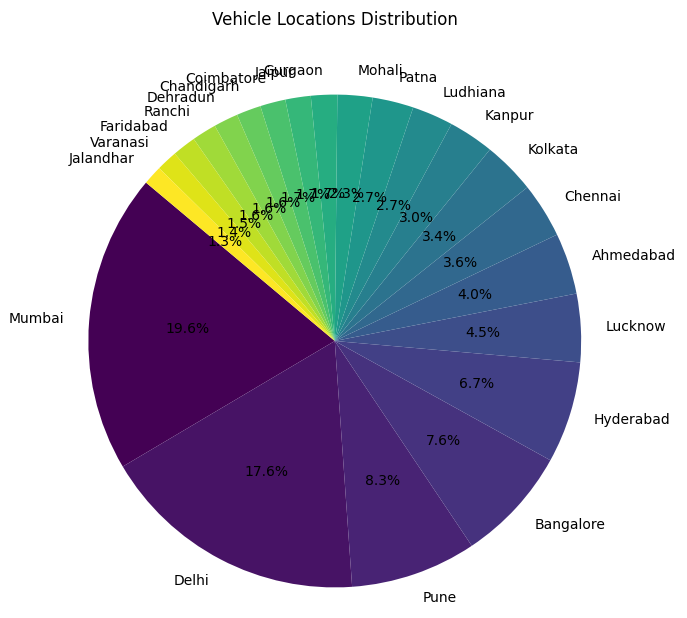
* Horsepower and engine RPM from the "Max Power" column.
* Torque power and torque RPM from the "Max Torque" column.



After doing the visualizations for some features such as colour, make of car, locations, we found white to be the most common colour, Maruti and Hyundai being the most preferred brands, Mumbai and Delhi being the most common locations in terms of car sales.



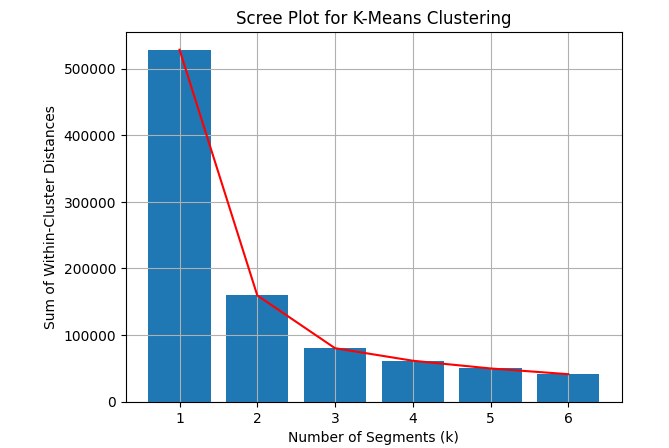
I decided to remove locations that occurred less than 20 times in the dataset so that it becomes easier to interpret.



**Segmentation Analysis**

It was performed using clustering techniques. The steps included:

1. **Choosing the segmentation variables:**
   * Price, Fuel Type, Location, Transmission, Colour, Seating Capacity.
2. **Mapping:**
   * Encoding categorical features in the ‘dfp’ dataframe into numerical values.
   * Dropping the original categorical columns.
3. **Standardization:**
   * Standardizing the data using StandardScaler to ensure all features contribute equally to the clustering process.
4. **K-means Clustering:**
   * Determined the optimal number of clusters using scree plot and decided to go with 3.

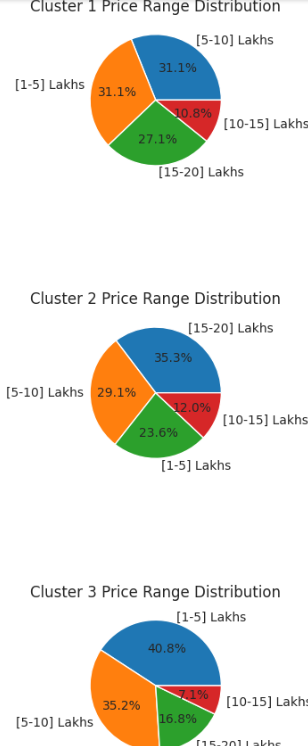
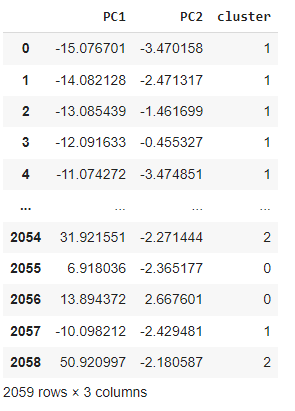


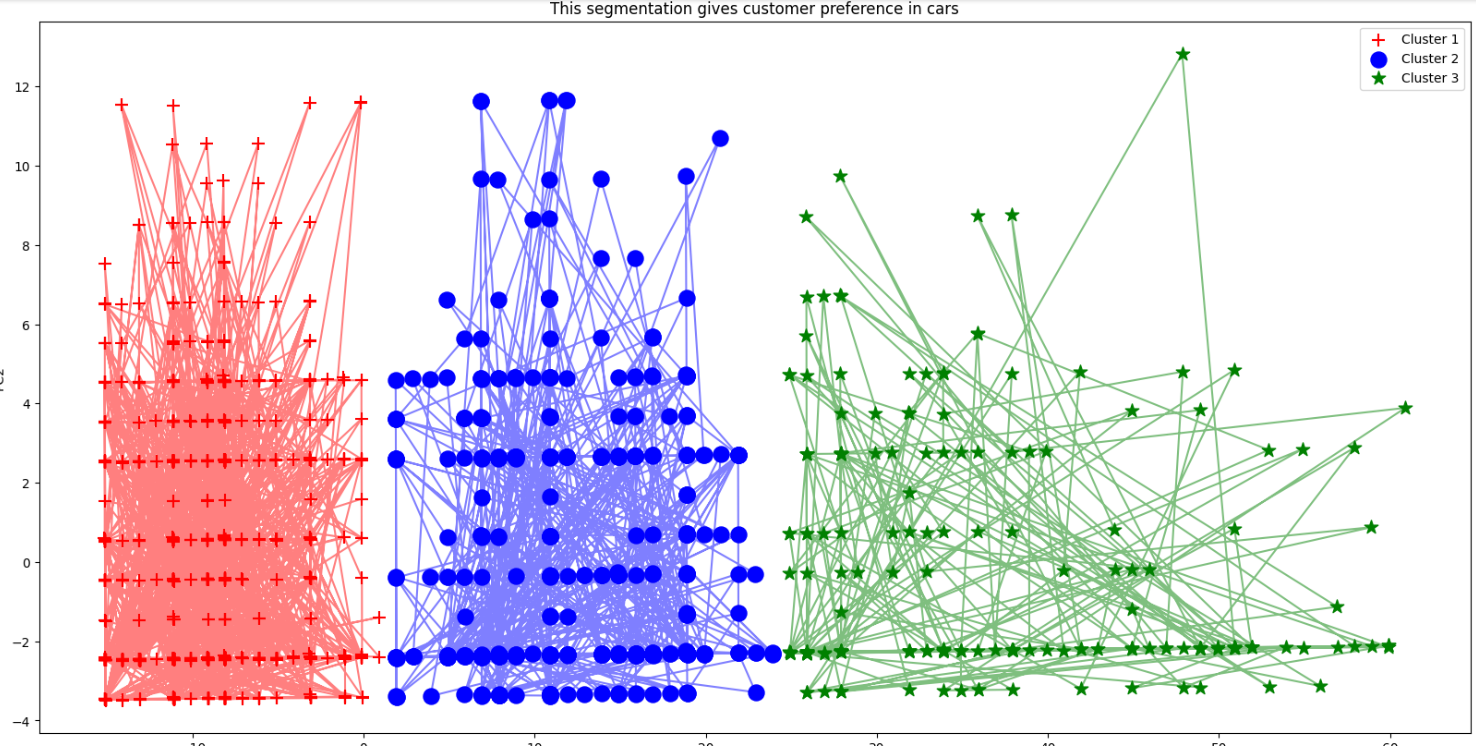
* + Selecting those relevant features and considering the numerical columns Price and Seating Capacity for clustering.
  + Applying K-means to identify distinct segments within the data.

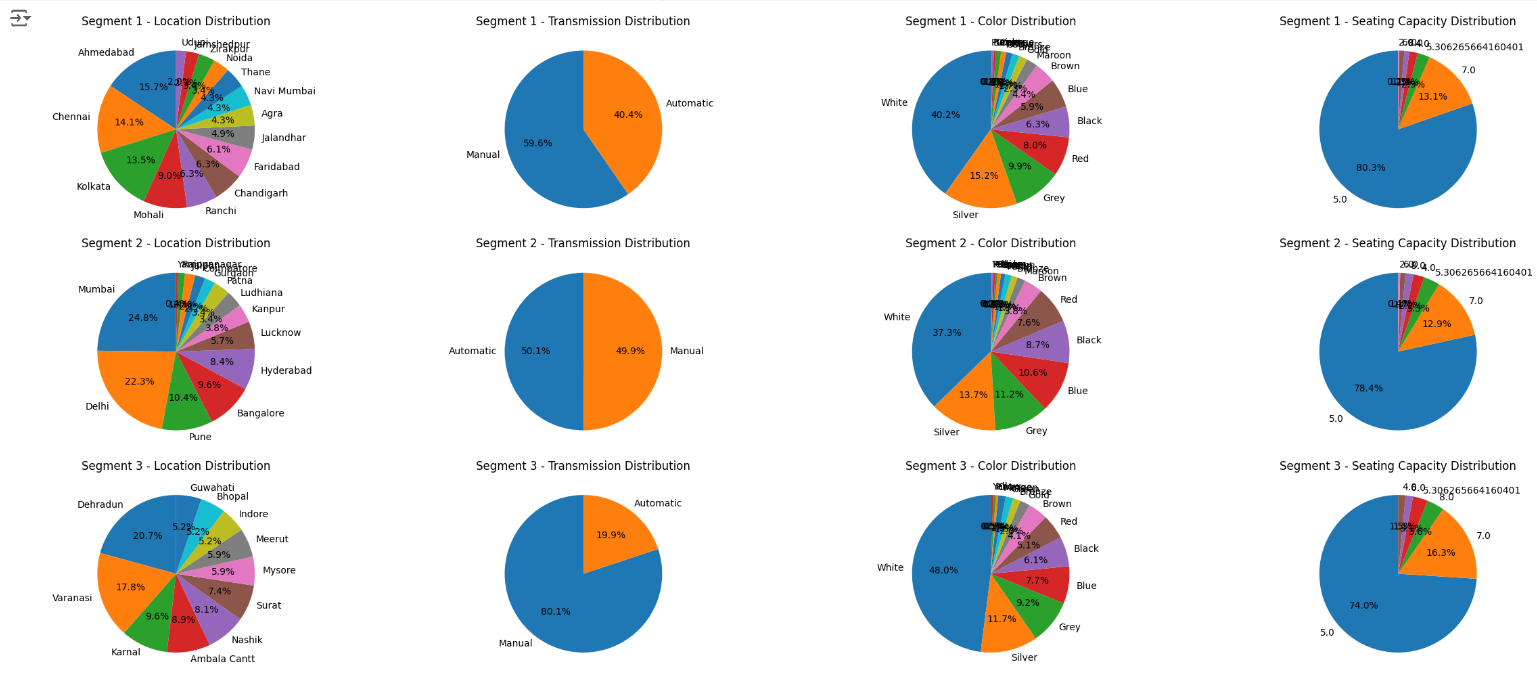
1. **Principal Component Analysis (PCA):**
   * Reducing the dimensionality of the data to visualize the clusters better.

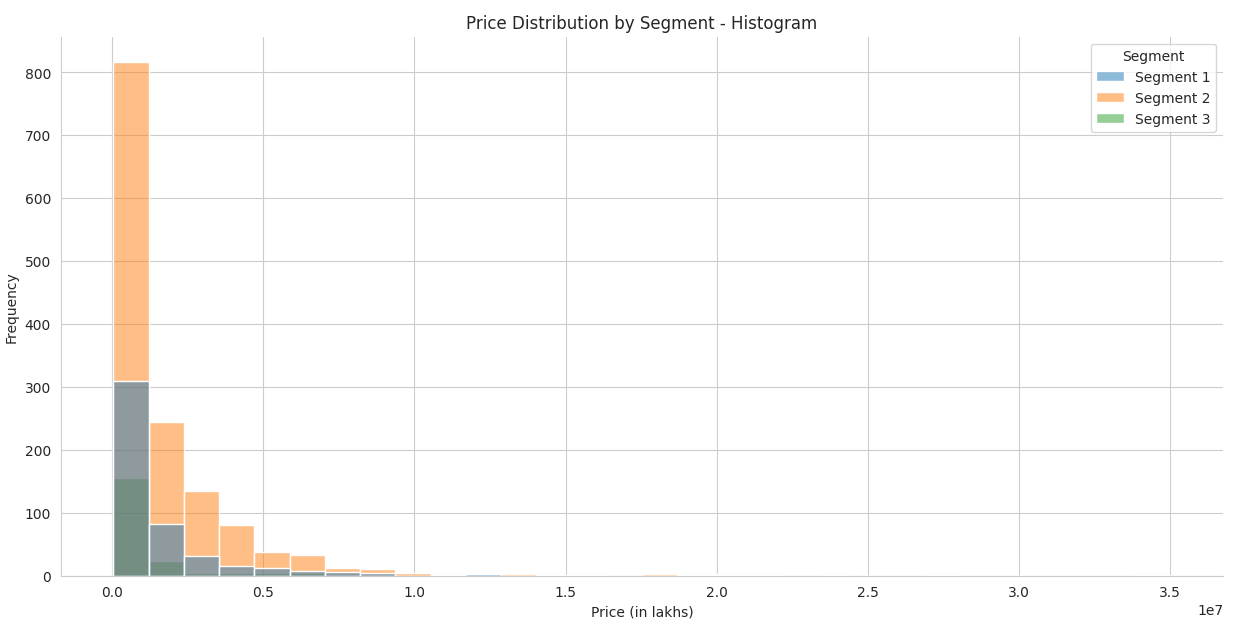
**Results**

The results of the clustering analysis were visualized using scatter plots and PCA components to illustrate the distribution of vehicles, locations, their features and the prices across different segments.









The histogram plot I plotted shows the distribution of price across three segments.

**Key Observations:**

1. **Concentration at Lower Prices**:
   * The majority of the prices are concentrated in the lower price ranges, specifically below 0.5 lakhs. This is indicated by the high bars on the left side of the plot.
   * Segment 2 (orange) has the highest frequency in the lowest price range.
2. **Overlap Between Segments**:
   * There is an overlap in the price distributions of the segments in the lower price ranges. However, Segment 2 seems to dominate the lower end of the price spectrum.
   * Segment 1 (blue) and Segment 3 (green) also have a presence in the lower price ranges but with lower frequencies compared to Segment 2.
3. **Higher Price Ranges**:
   * As the price increases, the frequency of items decreases sharply for all segments.
   * There are very few items in the higher price ranges (above 1 lakh), and these higher prices are mostly dominated by Segment 2, with some presence of Segment 1.

**Insights:**

1. **Market Focus**:
   * Segment 2 appears to focus on lower-price, as indicated by the highest frequency in the lowest price range.
   * Segment 1 and Segment 3 have a more diverse distribution but still lean towards lower prices.
2. **Price Distribution**:
   * The data shows a heavy skew towards lower prices, with very few high-priced items across all segments.
   * This could imply that the majority of the market is in the lower price range, which might be a target for competitive pricing or marketing strategies.

**Conclusion**

Given the price sensitivity of the Indian market, targeting the low-cost segment is crucial. Developing affordable EVs that provide best value for money can capture a significant market share.

Since people prefer a manual white, grey or a silver coloured 5-seater car, if we consider these points, our EV startup can offer vehicles that cater to all 3 types of consumer segments in a price range of 1-10 lakhs which is more affordable for most people and our visualizations also indicate the ideal places for the startup should be in Mumbai, Delhi, Ahmedabad, Chennai, Kolkata and Dehradun.

The Indian EV market presents significant opportunities for growth and expansion. By focusing on these key market segments, developing market-specific EVs, our EV startup can successfully enter and thrive in the market.

**Github link-** <https://github.com/adarsh1102/EV-market-segmentation>