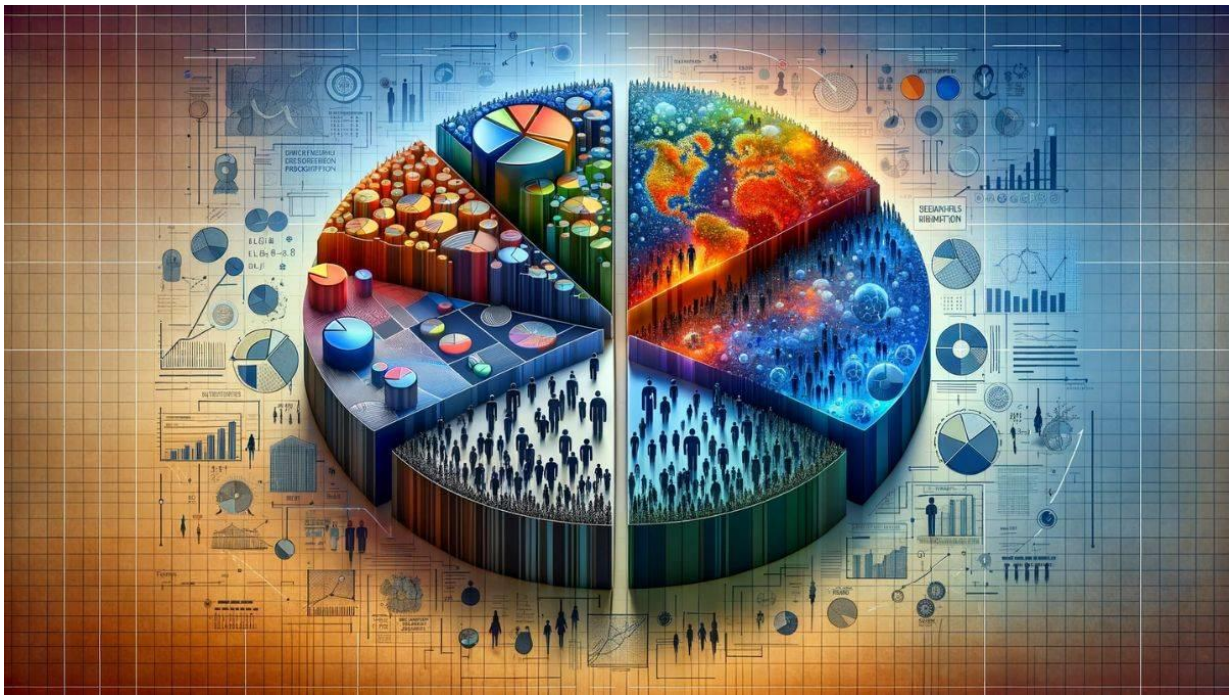
Vehicle Market Segmentation Analysis



Prepared By:

Tanzila Naz

Date: 5 Sep, 2024

https://github.com/Naztanzila/FenyyLabs_MarketSegementation

Contents

1. Introduction

Objective:

The idea of the project is to conduct market segmentation on the car market so that it is defined in terms of customer or product groups. Market segmentation is important for implementation in the car market because it provides an effective way of understanding the market and hence offering a way of meeting the various customer needs. When such segments are identified, firms can align their promotional campaigns, decide 'which products to offer to these segments, and improve on satisfying their customers to open new sales opportunities and enjoy increased market share. Another advantage of segmentation is that it assists in identifying the correct market for the correct product hence making marketing effective.

Scope:

This project therefore aims at segmenting the car market based on car data collected and analysed. Features of the dataset contain the car including name, year of manufacture, price, kilo-meter meter reading, fuel type, type of seller, transmission, type of owner, mileage, engine capacity, max power, torque, and seating capacity. In the analysis where market segmentation is central to the study, clustering techniques, specifically the K-Means clustering technique, are used to partition this market into different groupings. The analysis is envisaged to give a perception of the various customers' categories and propose how the category can be marketed uniquely. Unlike many academic papers, this project focuses on implementing the ML models to draw business insights from data.

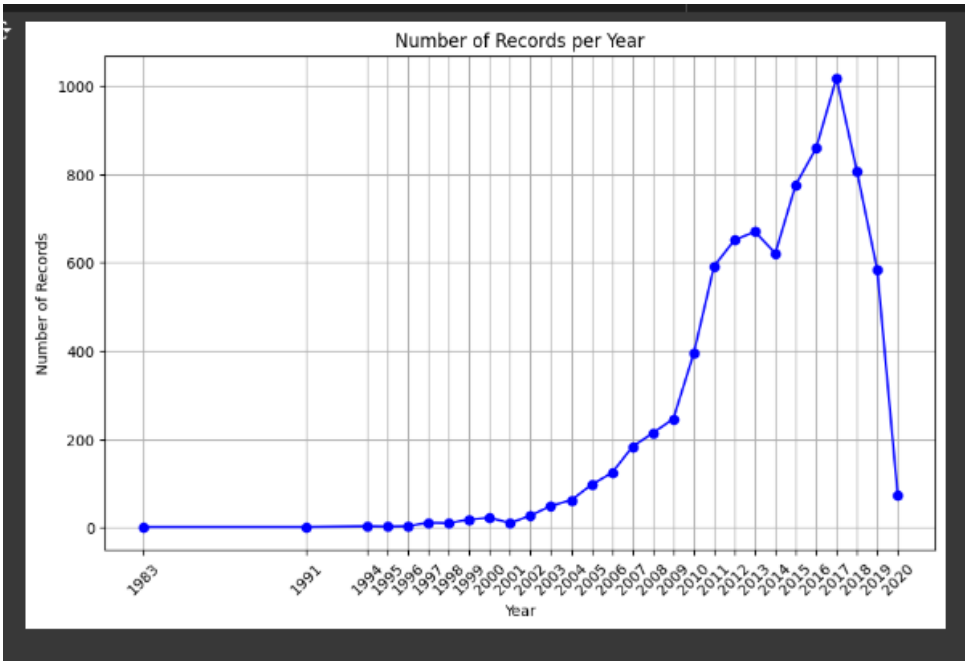
2. Data Description

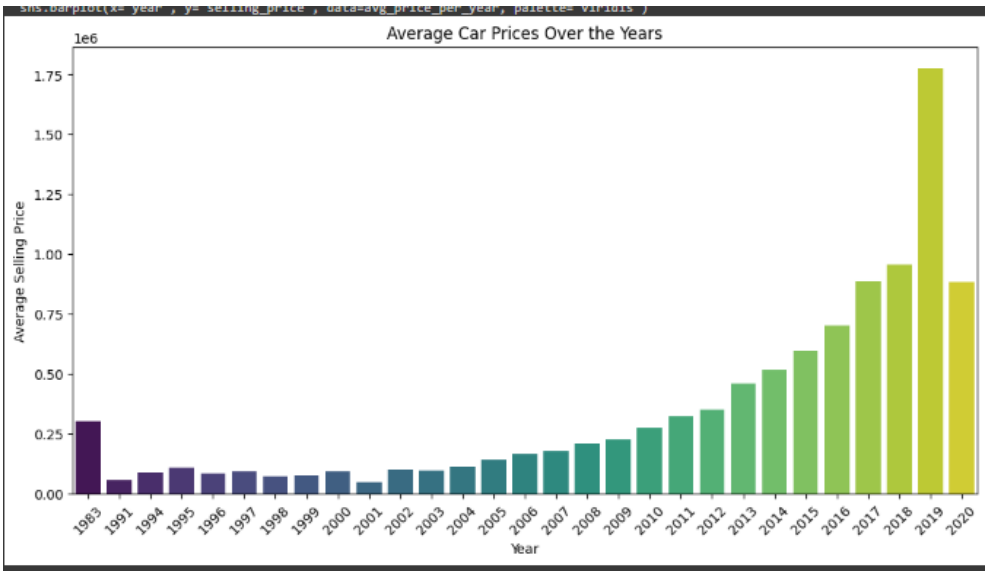
- **Data Features:** The dataset used in this project includes the following columns:
  - *name*: Car model name
  - *year*: Year of manufacture
  - *selling_price*: The selling price of the car
  - *km_driven*: Kilo-meters driven
  - *fuel*: Fuel Type (Diesel, Petrol, LPG, CNG)
  - *seller_type*: Seller type (Individual, Dealer, Trustmark Dealer)
  - *transmission*: Transmission type (Manual, Automatic)
  - *owner:* Number of previous owners
  - *mileage*: Mileage of the car
  - *engine*: Engine capacity in CC
  - *max_power:* Maximum power in BHP
  - *torque:* Torque of the car
  - *seats:* Number of seats
- **Data Cleaning:**
  - **Handling Missing Values**: Missing values were identified in crucial columns like *'mileage', 'engine', 'max_power', 'torque', and 'seats'.*
  - **Handling Categorical Features:** Categorical variables such as *'fuel', 'seller_type', 'transmission'*, and '*owner*' were transformed into numerical representations using Label Encoding.
  - **Feature Engineering:**
    - ✓ The 'torque' column initially contained values as strings with varying formats (e.g., '190Nm@2000rpm').
    - ✓ The 'mileage' values were initially stored as strings with units (e.g., '23.4 kmpl').

3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to gain an initial understanding of the dataset and uncover key patterns and relationships among the variables. Several visualizations and statistical methods were used to explore the data, focusing on distributions, correlations, and other notable trends.

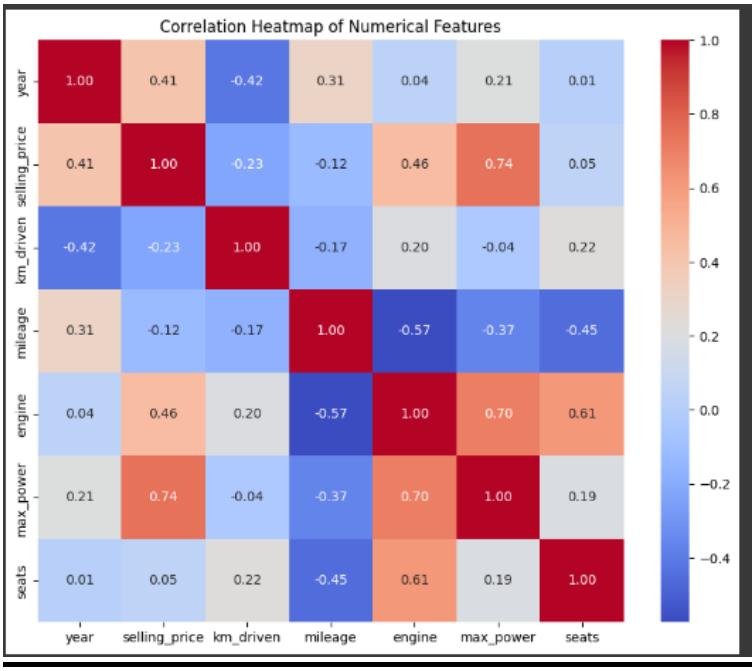
- *Distributions of Key Features*:





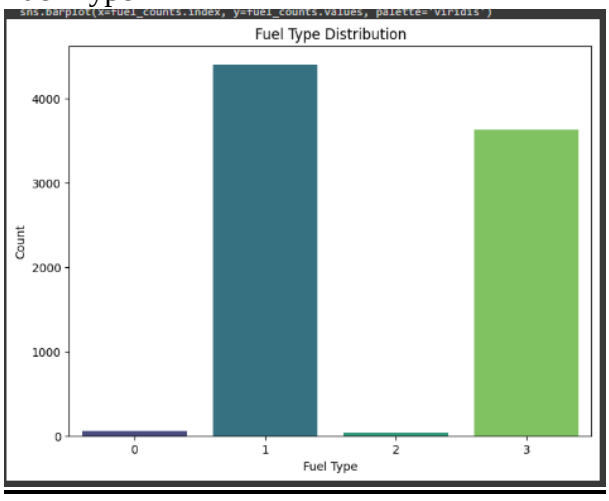
https://github.com/Naztanzila/FenyyLabs_MarketSegementation

➢ *Correlations*



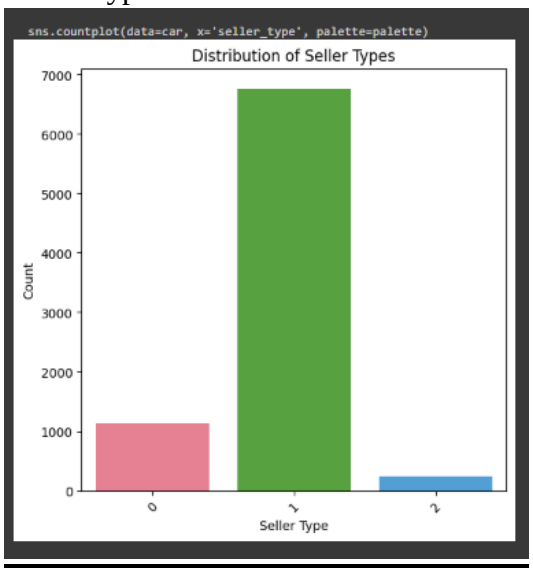
➢ *Categorical Feature Analysis:*

- ✓ Fuel Type

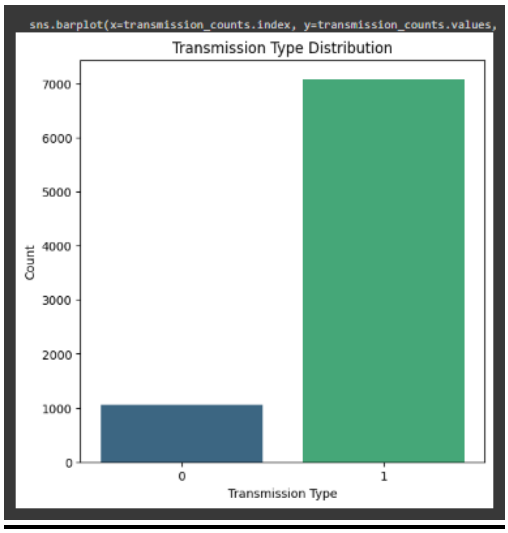


https://github.com/Naztanzila/FenyyLabs_MarketSegementation

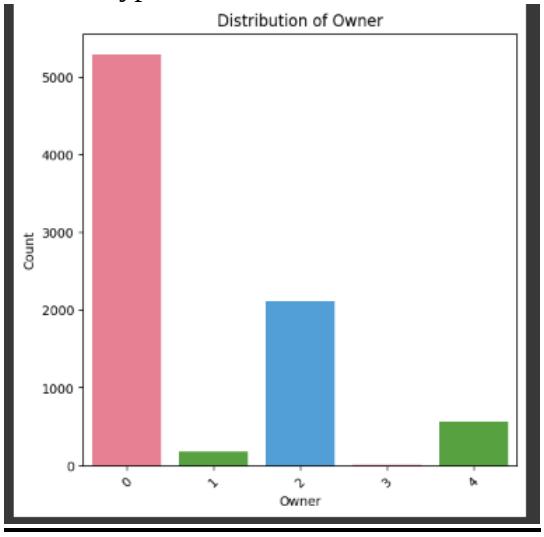
✓ Seller Type



✓ Transmission

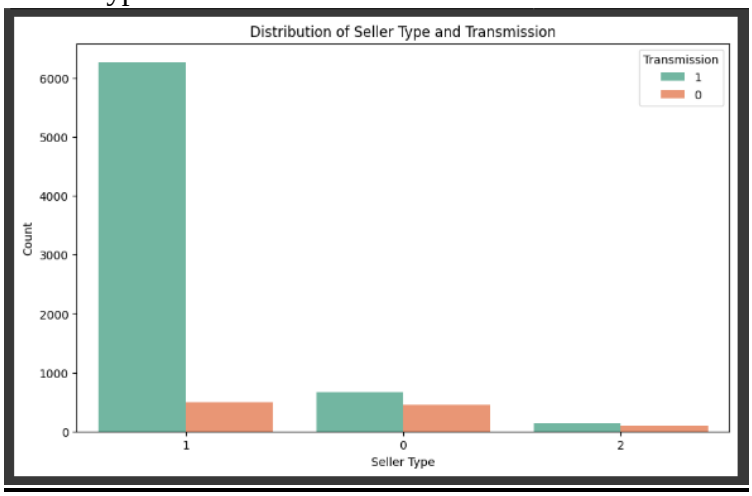


✓ Owner Type



https://github.com/Naztanzila/FenyyLabs_MarketSegementation

✓ Seller Type VS Transmission



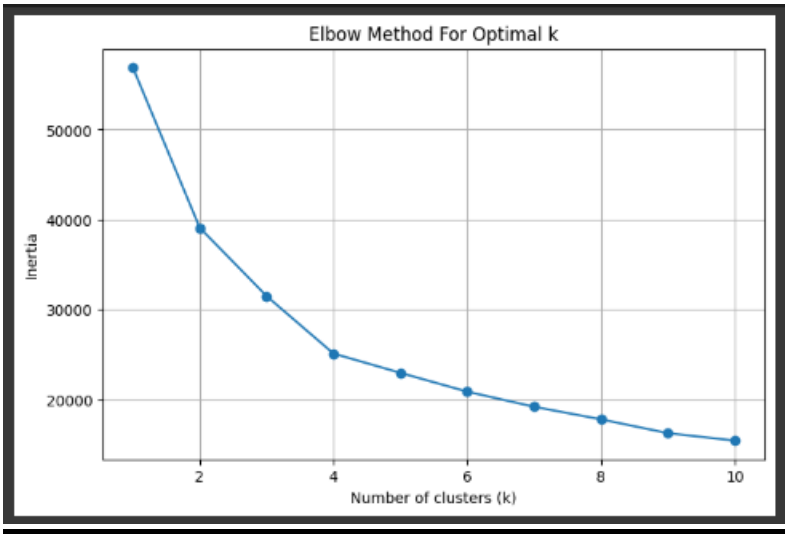
➢ *Segmentation Patterns*

```
                year  selling_price      km_driven      fuel  seller_type  \
Cluster
0        2013.706918   4.250050e+05   65362.741987  2.074738     0.914948
1        2017.354978   3.358634e+06   29733.268398  1.476190     0.426407
2        2013.049853   7.022016e+05  103988.426686  1.183284     0.934751

         transmission     owner    mileage       engine   max_power  torque  \
Cluster
0            0.927801  0.829737  20.681117  1229.039829   79.581419   200.0
1            0.069264  0.147186  16.927100  2182.833333  185.313182   200.0
2            0.879032  0.969208  14.268526  2239.939883  114.967801   200.0

            seats
Cluster
0        5.064424
1        5.123377
2        7.076246
```

Cluster 0: Represents the "Economy" segment, with older, more affordable cars that have been driven more and offer less power.

Cluster 1: Represents the "Mid-Range" segment, with cars that balance age, price, features, and performance.

Cluster 2: Represents the "Luxury/Performance" segment, with newer, more expensive cars that offer better performance and potentially more space.



https://github.com/Naztanzila/FenyyLabs_MarketSegementation

4. Methodology

Clustering Algorithm: For the market segmentation in this project, we used the **K-Means Clustering** algorithm. K-Means is an unsupervised machine learning algorithm that is highly effective for partitioning data into distinct groups based on feature similarity. The algorithm works by initializing several centroids and iteratively adjusting their positions to minimize the distance between the data points and the nearest centroid, ultimately creating clusters. The choice of K-Means was driven by its simplicity and efficiency for segmenting a dataset with numerical features such as mileage, engine size, and selling price. The optimal number of clusters was determined using the elbow method, which involves plotting the sum of squared distances (inertia) against the number of clusters and selecting the point where the decrease in inertia slows down significantly.

5. Insights and Recommendations

Insights:

- The used car market is dominated by cars with a single previous owner and manual transmissions.
- Diesel is the most common fuel type, followed by Petrol.
- Maruti Suzuki models are highly popular in the used car market.
- There's a clear preference for 5-seater cars, followed by 7-seater cars.
- Newer cars generally command higher prices, with a significant price jump for cars manufactured after 2010.
- Smaller engines tend to offer better mileage.
- Five distinct customer segments were identified: Economy, Mid-Range, Luxury/Performance, and two additional segments with specific characteristic combinations

Recommendations:

- Target marketing efforts towards first-time car buyers and those seeking manual transmission vehicles.
- Optimize inventory to include a variety of Maruti Suzuki models and cars with 5 or 7 seats.
- Adjust pricing strategies based on the year of manufacture, mileage, and identified customer segments.
- Highlight fuel efficiency for cars with smaller engines.

- Consider offering specialized services or promotions for each identified customer segment to maximize sales and customer satisfaction.

6. Future Work

Datasets Collection: Acquiring additional data points would enrich the analysis.

- Customer Demographics: Data on buyer age, income, and location would help in understanding market segments better.
- Brand Sentiment: Customer reviews and sentiment analysis on brands could further differentiate segments.
- Hierarchical Clustering: To explore different cluster shapes and structures, hierarchical clustering could provide a more nuanced understanding of relationships between data points.

7. Market Analysis

The estimated market size for the car market domain in India (non-segmented) is vast, with millions of vehicles sold annually. For the year 2023, the total car market size is estimated to be approximately 3.7 million units, which includes all segments from economical to luxury cars. This number reflects a growing demand driven by urbanization, increasing disposable incomes, and a preference for personal mobility.

8. Cluster Analysis

Based on the analysis, the top four variables that can be used to create the most optimal market segments for the car market are:

- Engine Size: Strongly correlates with car performance and price.
- Mileage: A key determinant for cost-conscious buyers, influencing the perceived value of a car.
- Selling Price: Directly linked to customer affordability and market positioning.
- Fuel Type: Differentiates cars based on running costs and environmental impact preferences, such as petrol, diesel, or electric.

9. GitHub Repository

The full project code, including all data cleaning, analysis, and modeling scripts, can be found in the GitHub repository: https://github.com/Naztanzila/FenyyLabs_MarketSegementation

10. Conclusion

In this project, I used K-Means clustering to segment the car market based on various attributes such as engine size, mileage, and selling price, aiming to uncover distinct customer groups. Analysis revealed that the car market is characterized by diverse segments, each defined by unique combinations of features. Key attributes like engine size, mileage, fuel type, and seller type played a significant role in distinguishing these segments. For instance, larger engines were generally associated with higher power but lower mileage, catering to performance-oriented customers, while smaller engines offered better fuel efficiency, appealing to cost-conscious buyers. The use of visual tools like scatter plots and heatmaps further illustrated the relationships between these variables, highlighting how different factors influence market segments and providing valuable insights for targeted marketing and product development strategies.