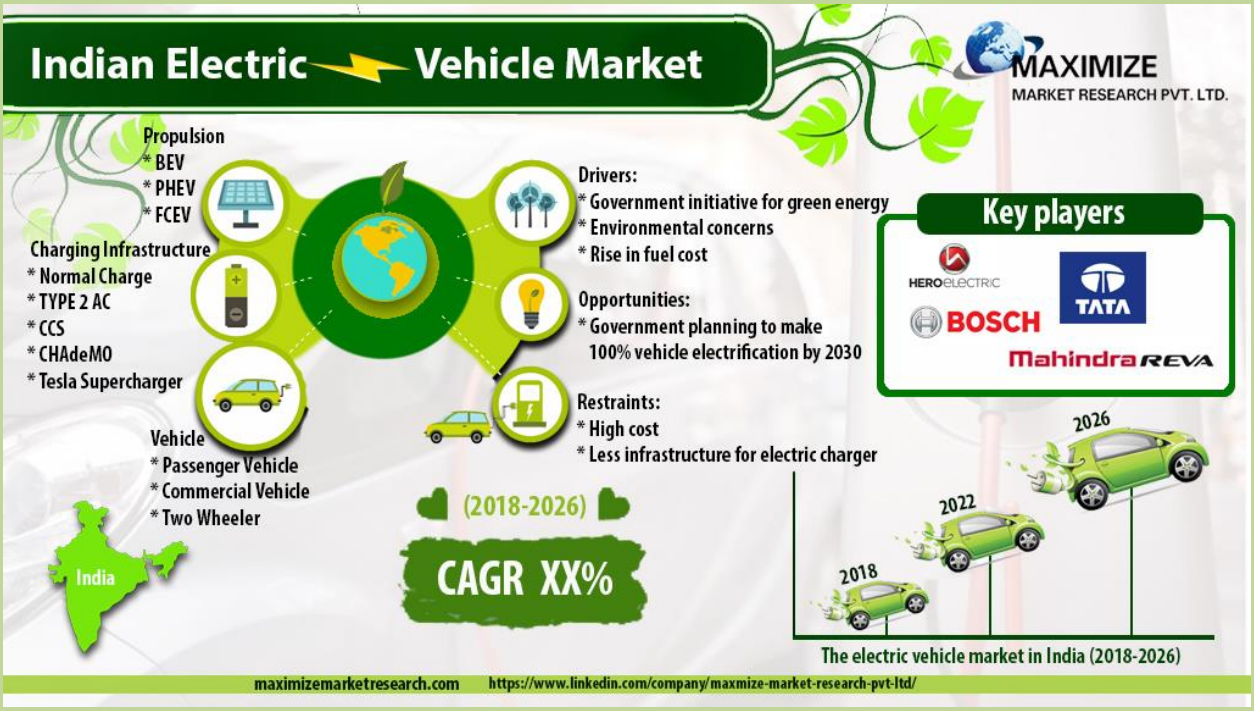
Market Segmentation of Electric Vehicles in India

By Sandeep kumar Pradhan

Date : 14/07/2024



Objective:

To perform clustering analysis on the Indian automobile dataset to identify groups of vehicles with similar fuel efficiency and related characteristics. This analysis aims to uncover patterns and insights that can inform manufacturers, policymakers, and consumers about vehicle performance and efficiency.

Problem Statement:

1 . **Data Exploration and Cleaning:**

- Perform an initial inspection of the dataset to understand its structure and contents.
- Clean the data by handling missing values and ensuring that all columns have appropriate data types.

Datasets :

a) Name | Manufacturer | Location | Year | Kilometers Driven | Fuel Type | Transmission | Owner Type | Engine CC | Power | Seats | Mileage Km/L | Price

b) Brand | Model | AccelSec | TopSpeed _KmH | Range_Km | Efficiency_WhKm | FastCharge_KmHRapidCharge | PowerTrain | PlugType |BodyStyle | Segment | Seats | PriceEuro

Statistical summary of Data 1

	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH	Seats	PriceEuro
count	103.000000	103.000000	103.000000	103.000000	103.000000	103.000000	103.000000
mean	7.396117	179.194175	338.786408	189.165049	444.271845	4.883495	55811.563107
std	3.017430	43.573030	126.014444	29.566839	203.949253	0.795834	34134.665280
min	2.100000	123.000000	95.000000	104.000000	170.000000	2.000000	20129.000000
25%	5.100000	150.000000	250.000000	168.000000	260.000000	5.000000	34429.500000
50%	7.300000	160.000000	340.000000	180.000000	440.000000	5.000000	45000.000000
75%	9.000000	200.000000	400.000000	203.000000	555.000000	5.000000	65000.000000
max	22.400000	410.000000	970.000000	273.000000	940.000000	7.000000	215000.000000

Statistical summary of Data 2

	Unnamed: 0	Year	Kilometers_Driven	Engine CC	Power	Seats	Mileage Km/L	Price
count	5975.00000	5975.000000	5.975000e+03	5975.000000	5975.000000	5975.000000	5975.000000	5975.000000
mean	3008.80887	2013.386778	5.867431e+04	1621.606695	112.599819	5.278828	18.179408	9.501647
std	1739.30056	3.247238	9.155851e+04	601.036987	53.659495	0.808959	4.521801	11.205736
min	0.00000	1998.000000	1.710000e+02	624.000000	34.200000	0.000000	0.000000	0.440000
25%	1502.50000	2012.000000	3.390800e+04	1198.000000	74.000000	5.000000	15.200000	3.500000
50%	3010.00000	2014.000000	5.300000e+04	1493.000000	92.700000	5.000000	18.160000	5.650000
75%	4514.50000	2016.000000	7.300000e+04	1984.000000	138.100000	5.000000	21.100000	9.950000
max	6018.00000	2019.000000	6.500000e+06	5998.000000	560.000000	10.000000	33.540000	160.000000

EDA :

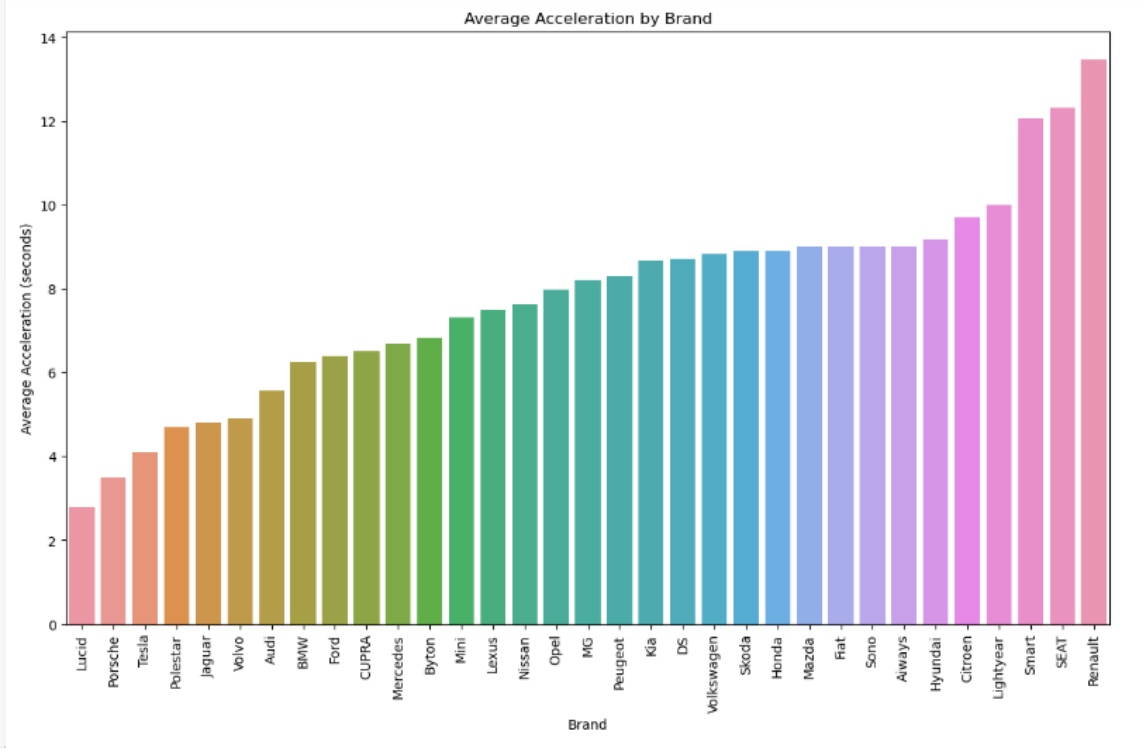
```
Model with the highest top speed:

================================

Brand              Tesla
Model           Roadster
TopSpeed_KmH         410
Name: 51, dtype: object

Model with the lowest top speed:

================================

Brand                  Nissan
Model           e-NV200 Evalia
TopSpeed_KmH               123
Name: 66, dtype: object
```

```
Model with the longest range:

================================

Brand          Tesla
Model       Roadster
Range_Km         970
Name: 51, dtype: object

Model with the shortest range:

================================

Brand            Smart
Model       EQ forfour
Range_Km            95
Name: 77, dtype: object
```

High Acceleration Brands: Brands with lower average acceleration times (shorter bars) indicate vehicles that accelerate faster. These brands may be known for their performance-oriented vehicles.

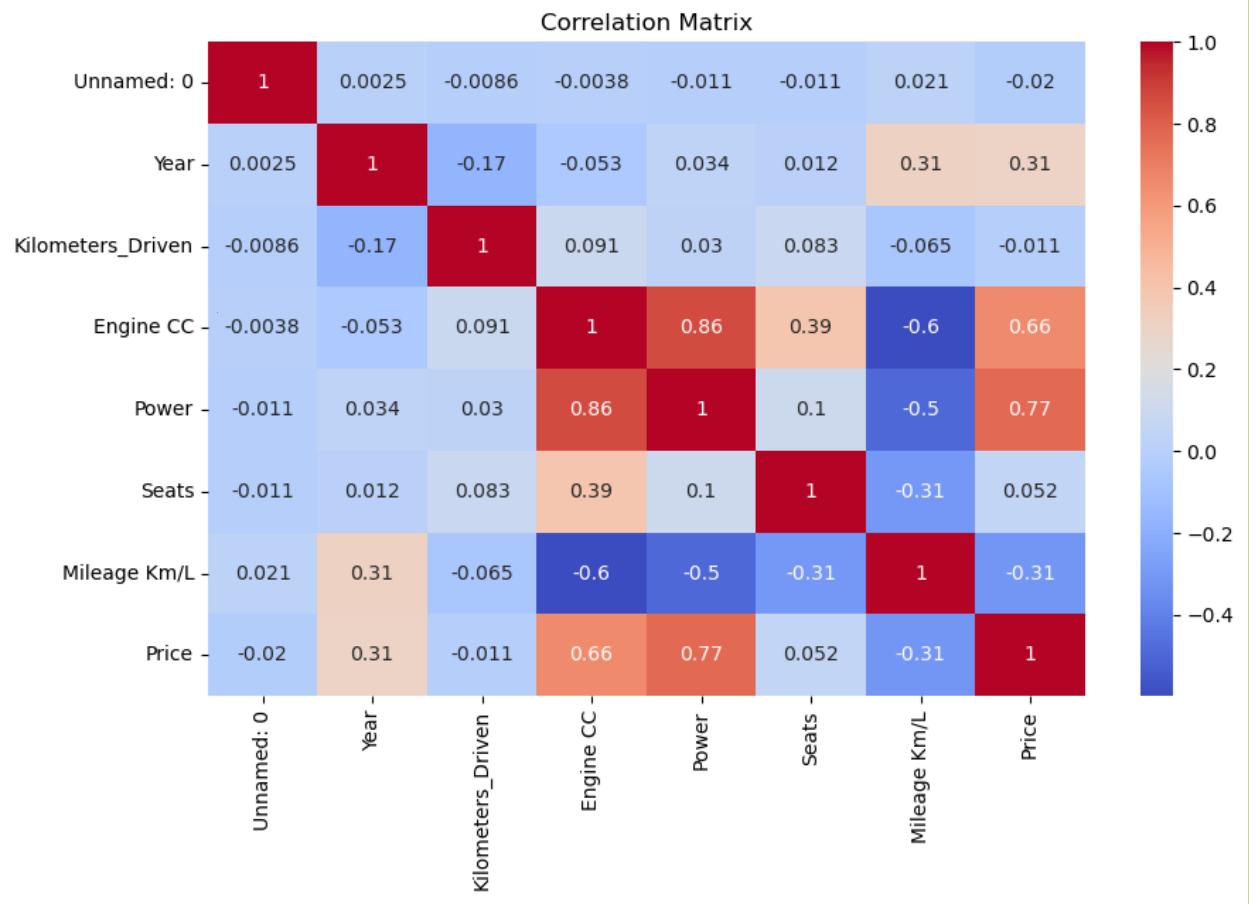
Low Acceleration Brands: Brands with higher average acceleration times (longer bars) indicate vehicles that accelerate slower. These brands may focus more on fuel efficiency or other attributes rather than performance.

Patterns and Trends: Any notable patterns, such as luxury brands typically having higher acceleration compared to economy brands, can be highlighted.



Visualize Correlation Matrix:

- Use seaborn's `heatmap()` function to create a heatmap of the correlation matrix.
- The `annot=True` parameter adds the correlation coefficient values to each cell in the heatmap.
- The `cmap='coolwarm'` parameter sets the color palette for the heatmap.
- The `fmt='.2f'` parameter formats the correlation coefficient values to two decimal places.
- The `linewidths=0.5` parameter adds lines between the cells for better readability.



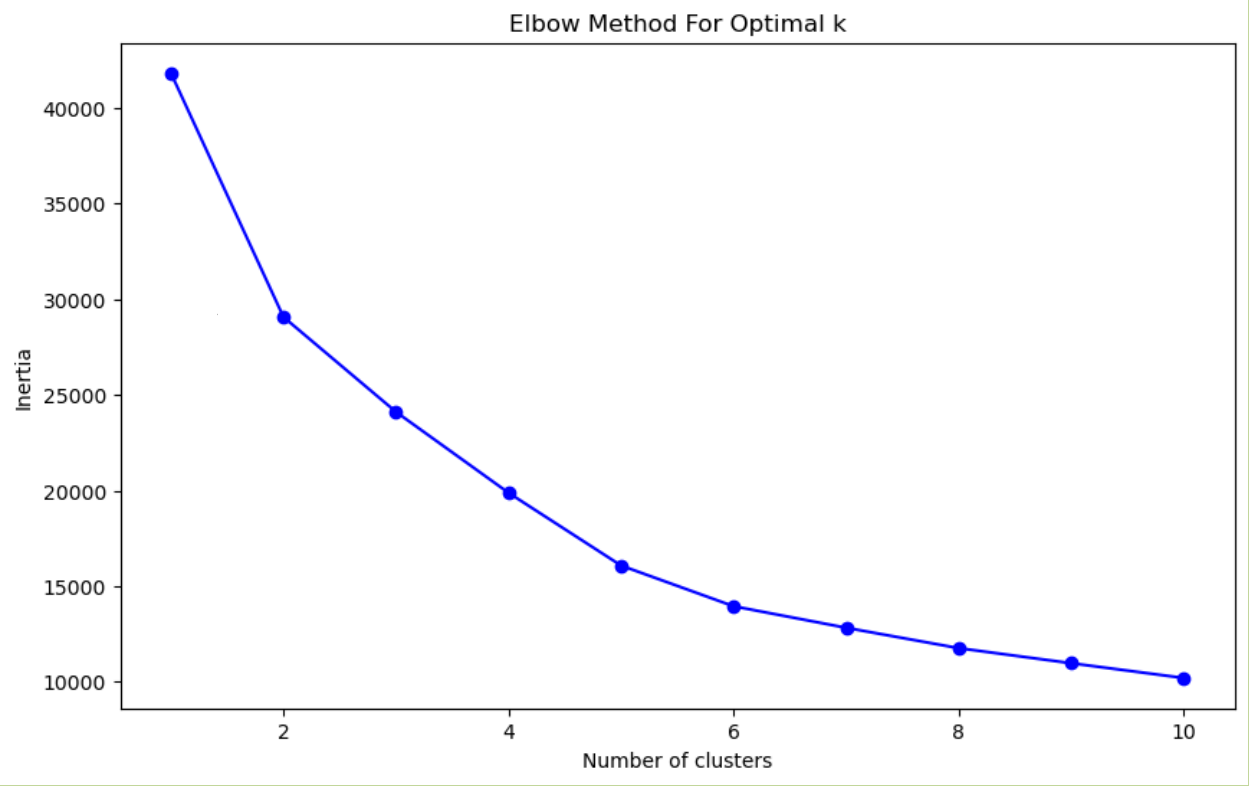
2. **Scale the Features:**

- Standardize the numerical features in the dataset to ensure fair comparison during clustering. Determine the optimal number of clusters using the elbow method to ensure meaningful segmentation of the data.
- Apply the K-Means clustering algorithm to segment the dataset into distinct groups based on fuel efficiency and related attributes.
- Assign cluster labels to each vehicle in the dataset.

	Year	Kilometers_Driven	Engine CC	Power	Seats	Mileage Km/L	Price
0	-1.043059	0.145555	-1.037638	-1.014627	-0.344705	1.862377	-0.691815
1	0.496840	-0.193055	-0.065903	0.253475	-0.344705	0.329673	0.267595
2	-0.735080	-0.138440	-0.703188	-0.445435	-0.344705	0.004554	-0.446384
3	-0.427100	0.309398	-0.621656	-0.444317	2.127816	0.572960	-0.312513
4	-0.119120	-0.196659	0.576374	0.525583	-0.344705	-0.658954	0.735252

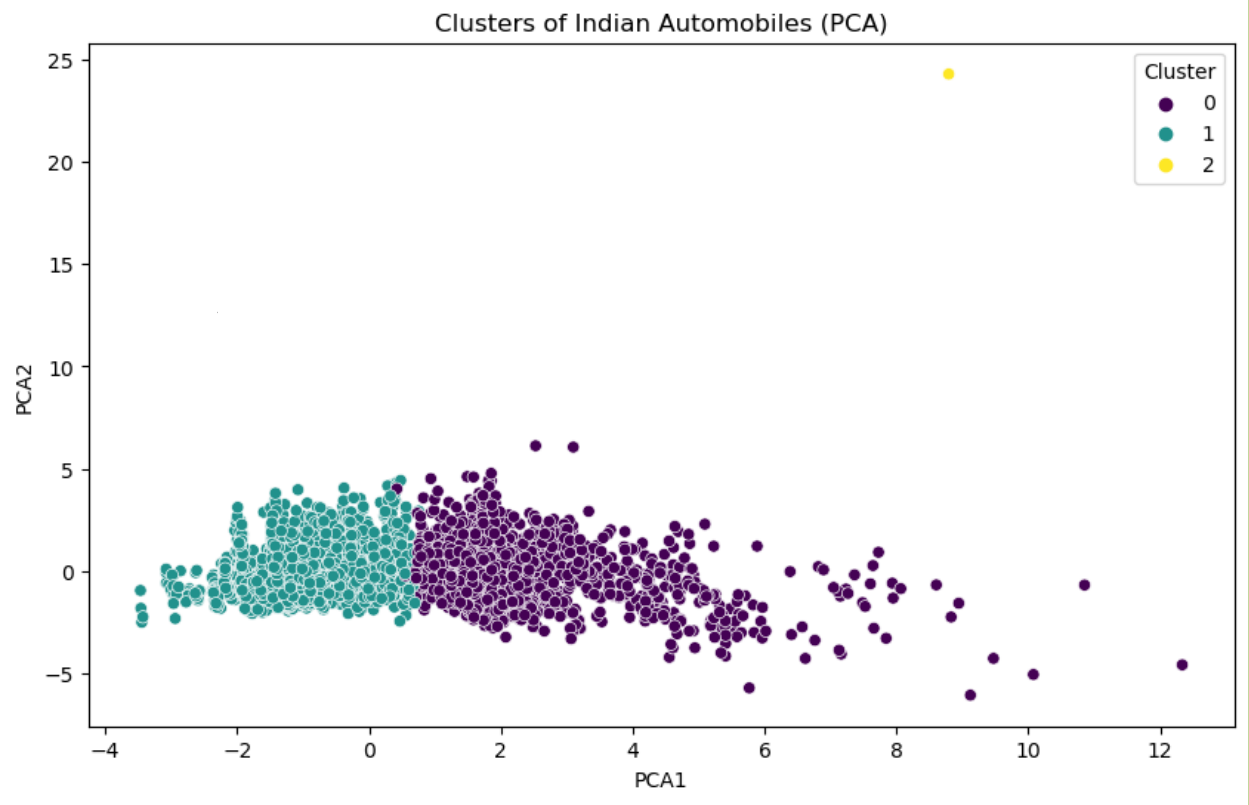
3. Cluster Evaluation and Visualization:

- Use Principal Component Analysis (PCA) to reduce the dimensionality of the dataset and visualize the clusters in a 2D space.
- Analyze the characteristics of each cluster to understand the common attributes of vehicles within the same cluster.



4. Insights and Recommendations:

- Provide insights based on the clustering results, such as identifying the most and least fuel-efficient clusters.
- Offer recommendations for manufacturers on vehicle design and for consumers on vehicle choices based on fuel efficiency.
- Suggest potential policy implications for promoting fuel-efficient vehicles.



Conclusion

The clustering analysis of the Indian automobile dataset has provided valuable insights into the segmentation of vehicles based on their fuel efficiency and related attributes. Through the application of data exploration, preprocessing, and K-Means clustering, we have identified distinct groups of vehicles that share similar characteristics. These findings can be used to inform various stakeholders, including manufacturers, policymakers, and consumers.

Recommendations:

1. **For Manufacturers**:
    - Focus on producing vehicles with characteristics similar to those in the high-efficiency cluster to meet consumer demand for fuel-efficient cars.
    - Invest in research and development to enhance fuel efficiency without compromising performance.
2. **For Policymakers**:
    - Promote policies and incentives that encourage the production and purchase of fuel-efficient vehicles.
    - Implement stricter fuel efficiency standards to reduce the environmental impact of automobiles.
3. **For Consumers**:
    - Consider the fuel efficiency and related attributes of vehicles when making purchasing decisions. Opt for vehicles in the high-efficiency cluster to save on fuel costs and reduce environmental impact.

Future Work:

- **Further Analysis**:
    - Conduct additional analyses to explore the impact of other factors, such as vehicle age and maintenance, on fuel efficiency.
    - Investigate the potential of other clustering algorithms and advanced machine learning techniques to improve segmentation accuracy.
- **Data Enrichment**:

- Enrich the dataset with additional features and more recent data to capture evolving trends in the automobile market.
- Collaborate with automobile manufacturers and industry experts to gain deeper insights and validate the clustering results.

By leveraging the insights gained from this clustering analysis, stakeholders can make informed decisions that promote fuel efficiency, enhance vehicle performance, and contribute to a more sustainable automobile market in India.