

Diwakar Kumar and Doordarshita Purohit

Electric Vehicle (EV) Market Analysis Report

Summary

This report explores used car data to uncover trends, segments, and market insights using data science techniques. Using clustering algorithms (KMeans), dimensionality reduction (PCA), and extensive visualizations, we identified market patterns that could inform customer targeting, pricing strategy, and inventory decisions. Key variables analysed include selling price, year, kilometres driven, fuel type, ownership, and other features.

Introduction

With the rapid expansion of the electric and used vehicle market in India, data-driven insights have become critical for understanding consumer preferences and optimizing business decisions. This project analyses a dataset of over 8,000 + used car listings to identify underlying market segments, trends, and patterns.

Methodology & Data Analysis

1. Data Collection

The dataset (car_details.csv) contains **8128** entries and **13** features including:

- **Name, Year, Selling Price, KM Driven, Fuel Type, Transmission, Owner, Mileage, Engine, Max Power, Torque, Seats**

Dataset link:

<https://drive.google.com/file/d/1yeTKNvAxCALz4QIKluGZqDFc6GbHt9dV/view?usp=sharing>

2. Data Cleaning

- Missing values found in mileage, engine, max_power, torque, and seats.
- Null values were dropped to maintain model integrity.

```
# Check for missing values
print(car_details.isnull().sum())

# Drop or fill missing values if necessary
car_details.dropna(inplace=True)
```

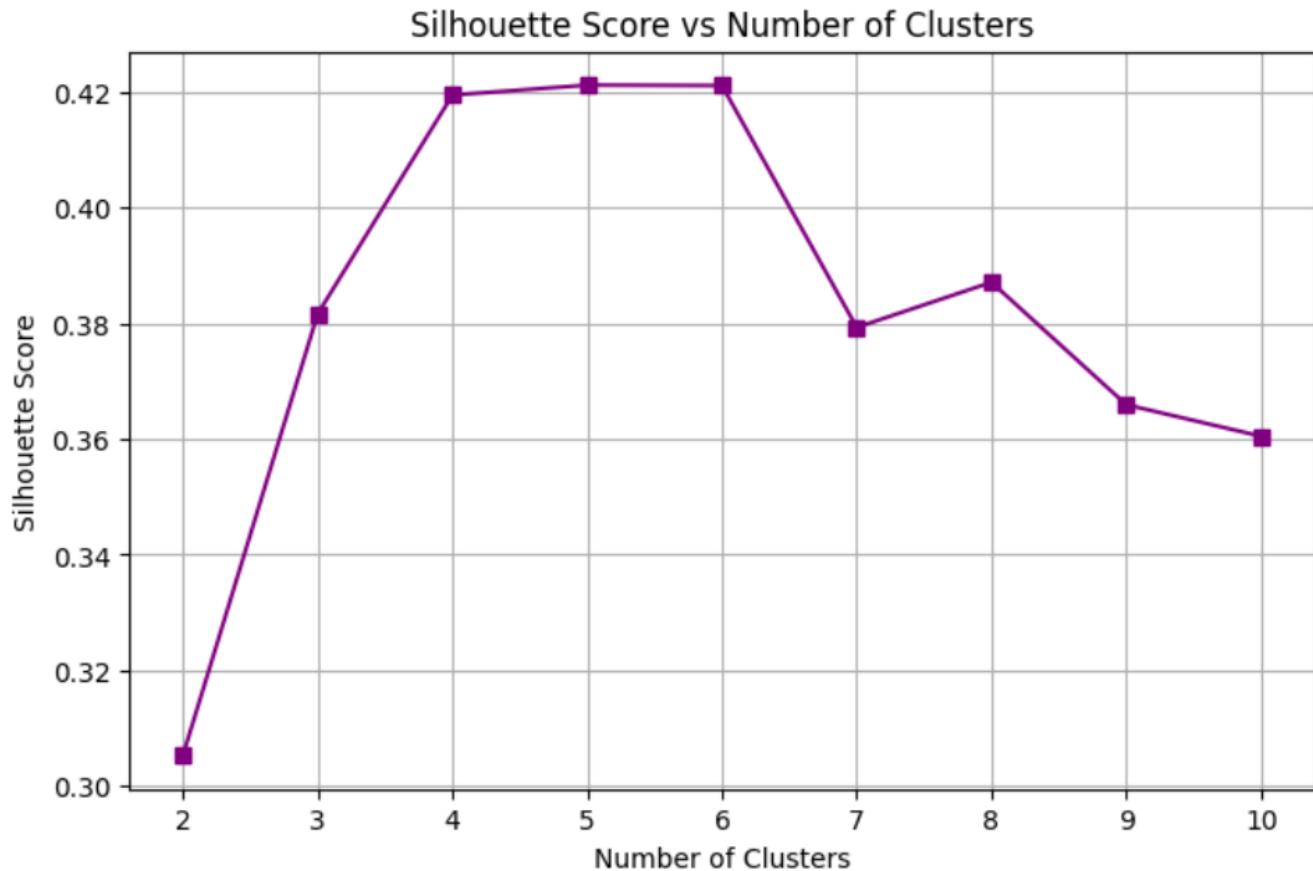
3. Feature Selection

- For clustering: year, selling_price, km_driven, and seats were used.
- These were scaled using StandardScaler to normalize the range.

4. Clustering with KMeans

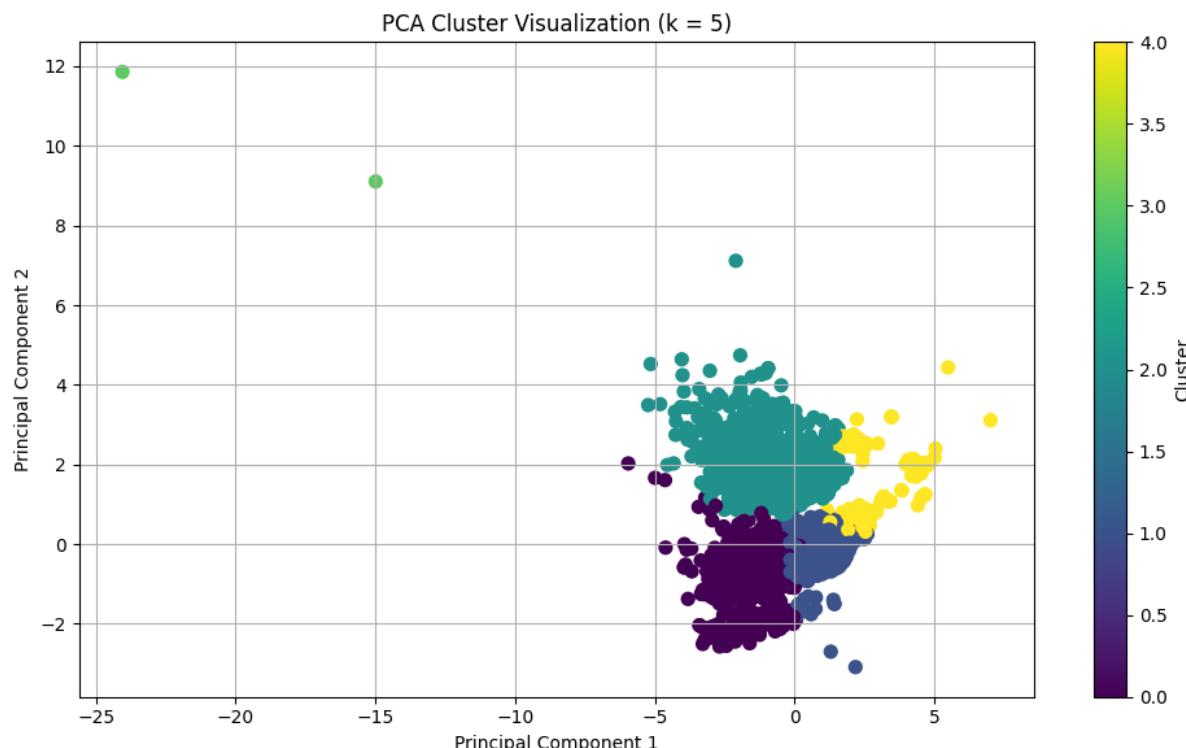
- The best number of clusters (k) was determined using the **Silhouette Score** method.

Silhouette Score vs Number of Clusters



- ✓ Optimal value found: $k = 3$

5. Dimensionality Reduction

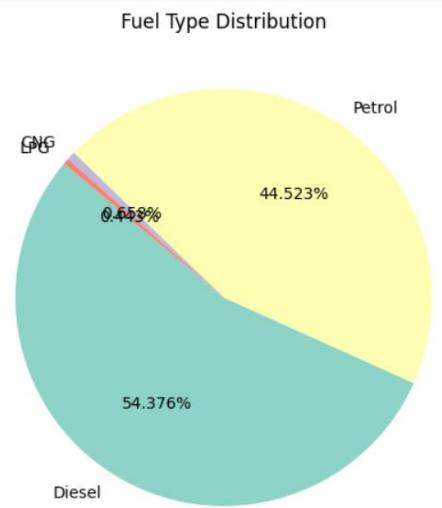


- Applied *PCA* (*Principal Component Analysis*) to reduce dimensions for visualization.

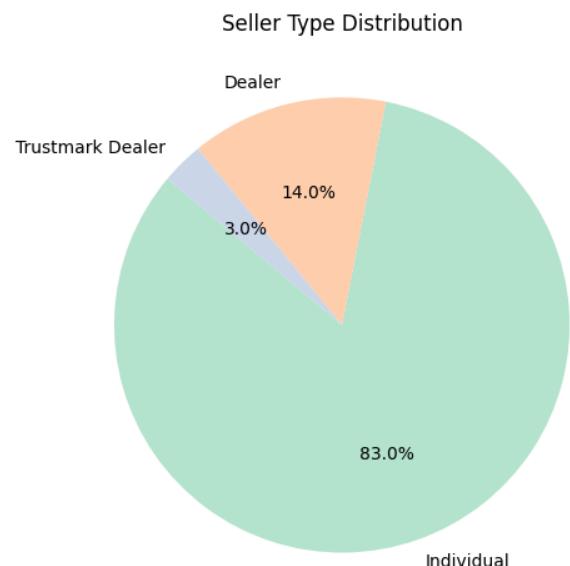
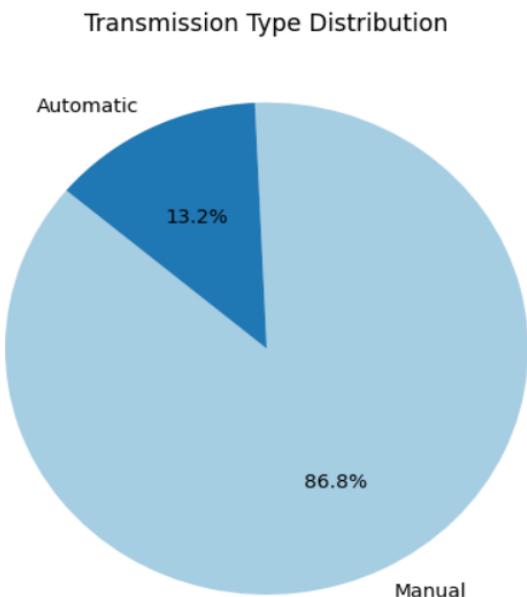
Market Overview & Trends

Fuel Distribution:

- Petrol and Diesel dominate the used car market.
- Petrol vehicles are especially popular among urban commuters due to their smoother ride and lower maintenance cost.
- Diesel vehicles, often used for longer trips and commercial purposes, are valued for better fuel efficiency and engine performance.
- LPG and CNG cars, though comprising a smaller portion, indicate a niche eco-conscious market.
- CNG vehicles are commonly found in metro areas where refuelling infrastructure is available.
- The growth of CNG/LPG reflects increasing environmental awareness and stricter emission norms.



Transmission & Seller Trends



Transmission Type:

- *Manual transmission vehicles dominate the listings, making up a large majority of second-hand cars.*
- Manual cars are often more affordable, and maintenance is simpler, contributing to their resale popularity.

- *Automatic transmissions*, while growing in popularity in metro areas, are still limited in the used car segment due to their higher initial cost and maintenance expense.
- Younger buyers and city dwellers are slowly pushing the demand for automatics in traffic-heavy regions.

Seller Type:

- *Individual sellers form the majority*, indicating a peer-to-peer market is dominant in the used car ecosystem.
- *Dealers*, though fewer in number, may represent more curated inventories with warranties or service packages.
- *Trust factors and price negotiation flexibility* are stronger with individual sellers, but dealers offer convenience and faster transactions.

Ownership Trends

First-owner vehicles dominate, which usually implies:

- Better documentation and maintenance records.
- Higher resale value due to perceived quality.
- Less wear and tear compared to multi-owner cars.

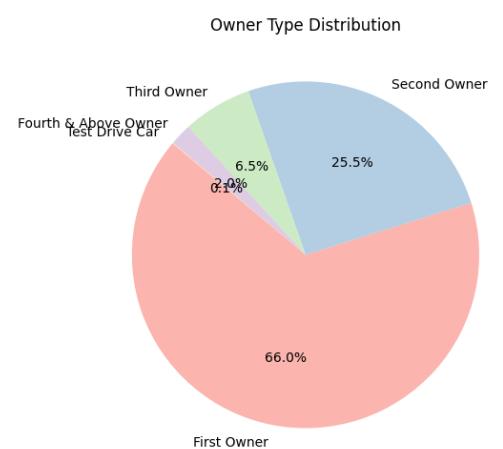
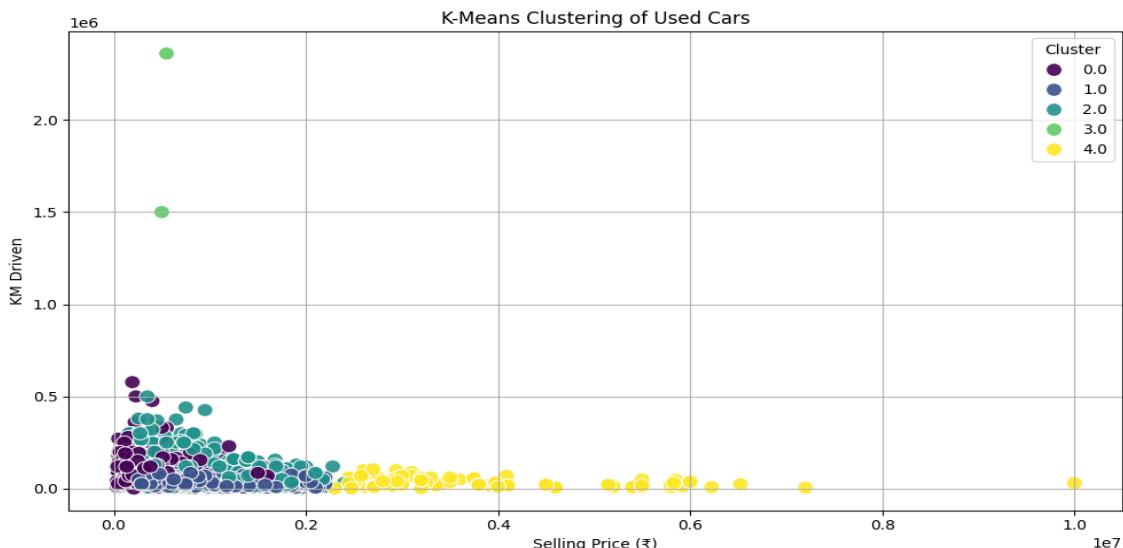
Second-owner cars are also common, often traded in the mid-life phase (around 5–7 years).

Third or fourth-and-above owners represent older or heavily used vehicles, typically priced lower but riskier in terms of reliability.

Buyers are increasingly prioritizing fewer past owners as a signal of trust and vehicle longevity.

Detailed Analysis with Graphical Insights

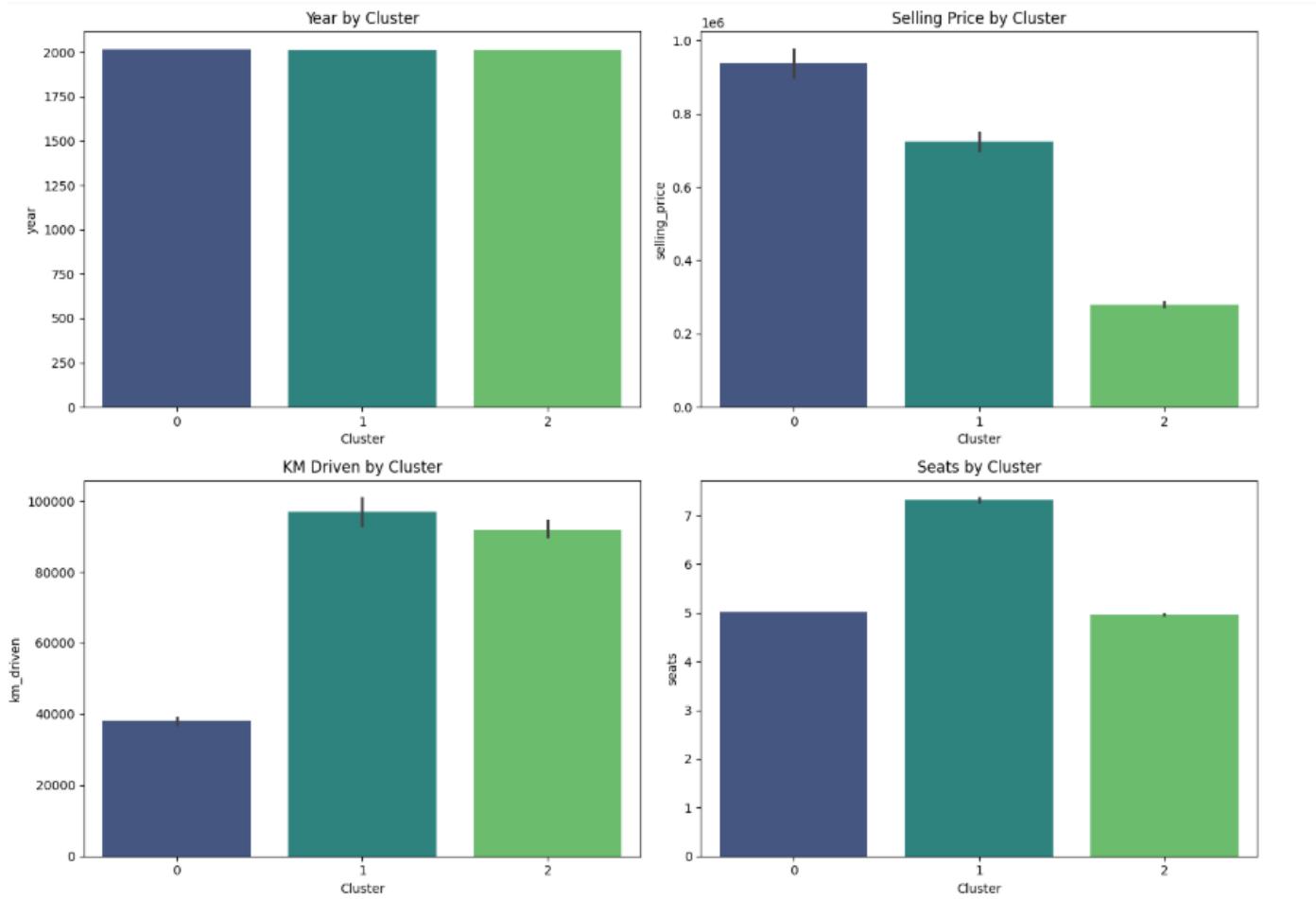
1. Cluster Analysis of Car Features



- *Cluster 0*: Low mileage, higher price – likely newer cars.
- *Cluster 1*: Older cars, mid-priced, balanced mileage.
- *Cluster 2*: High mileage, lower price – older cars or lower resale value.

2. Cluster-wise Bar Graphs

These bar plots help analyse how key car features vary across the clusters identified by KMeans. The clustering was based on standardized values of *year*, *selling_price*, *km_driven*, and *seats*.



Year by Cluster

- *Cluster 0* has the highest average year, indicating this group includes newer model vehicles.
- *Cluster 2* contains the oldest cars, suggesting budget vehicles with longer usage histories.
- *Cluster 1* sits between, representing moderately old cars, typically 6–10 years old.

Insight: This graph highlights the temporal segmentation of the market—clustered by how new or old the car is.

Price by Cluster

- Cluster 0 shows the highest average selling price, consistent with newer models and less usage.

- Cluster 2 represents the lowest price segment, indicating cars that are both old and heavily driven.
- Cluster 1 again falls in the middle range, indicating fairly priced, mid-aged vehicles.

Insight: Selling price is highly correlated with both car age and usage—clear economic stratification among clusters.

KM Driven by Cluster

- Cluster 2 cars have the highest average kilometers driven, often exceeding 100,000 km—likely ex-fleet or long-term private usage.
- Cluster 0 has the lowest KM, suggesting recently purchased or lightly used cars.
- Cluster 1 lies in the moderate range.

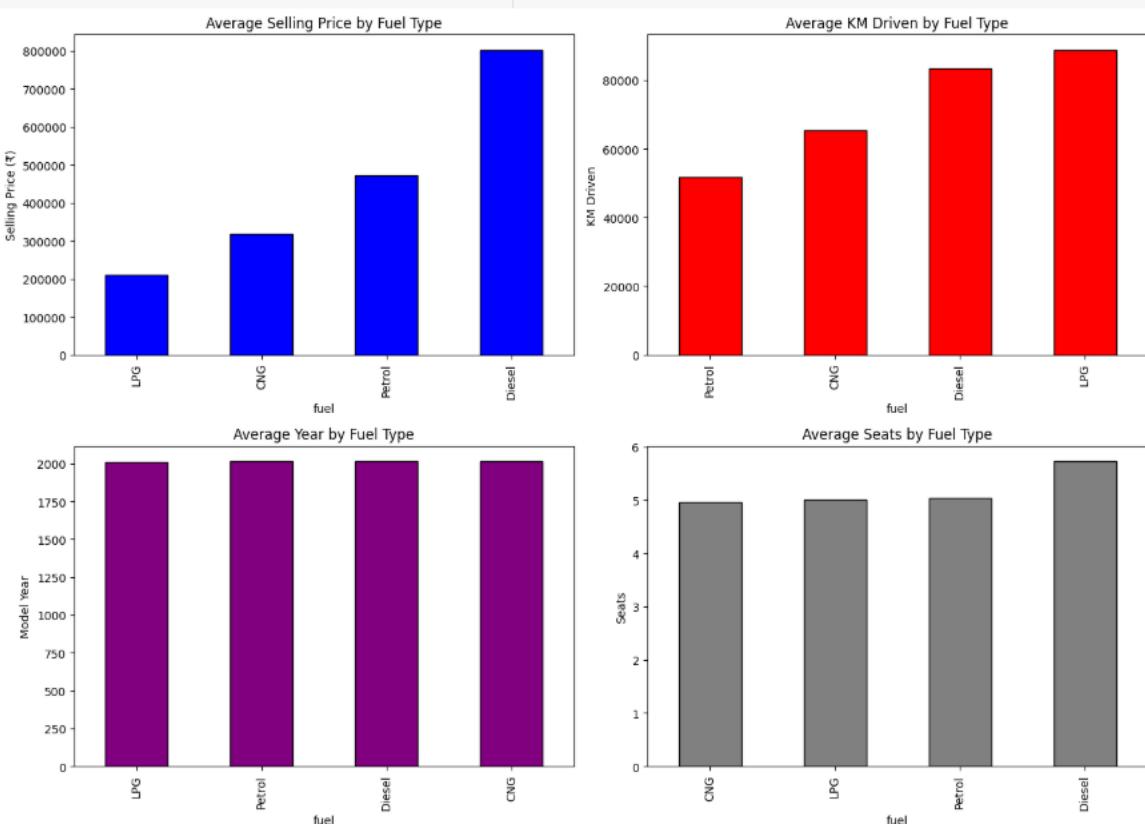
Insight: KM driven acts as a proxy for wear and tear. High mileage is a strong factor in downward price pressure and cluster assignment.

Seats by Cluster

- Most clusters average 5 seats, the standard for sedans and hatchbacks.
- Slight variation may indicate the presence of larger vehicles (7-seaters or SUVs), especially in Cluster 1, which shows a slightly higher average.

Insight: While seat count doesn't drastically affect pricing or clustering, it may reflect target buyer preferences (family vs. compact needs).

3. Fuel Type Insights



Average Selling Price by Fuel Type

- Diesel vehicles have the highest average selling price, likely due to better fuel efficiency and higher torque output.
- Petrol cars are moderately priced and dominate the market, especially in urban areas.
- CNG and LPG vehicles are priced lower, making them attractive for cost-conscious and eco-conscious buyers.

Insight: The resale value of diesel cars remains strong, while CNG/LPG options represent a budget-friendly, fuel-efficient segment.

Average KM Driven by Fuel Type

- Diesel cars show the highest average KM driven, supporting their common use in long-distance travel or commercial applications.
- CNG vehicles also have high usage, often employed by cab services or fleet operators.
- Petrol and LPG cars show lower average KM, suggesting more private, short-distance use.

Insight: KM driven reveals usage intent—diesel for endurance, petrol for daily urban drives, and CNG/LPG for economical city travel.

Average Year by Fuel Type

- Petrol and Diesel vehicles are newer on average, indicating sustained demand and availability in recent years.
- LPG vehicles tend to be older, possibly because the fuel system is often retrofitted on aging vehicles.
- CNG vehicles show a moderately recent average year, reflecting their increasing popularity in urban centers.

Insight: Petrol and diesel dominate newer models, while LPG is declining, and CNG is gaining steady adoption.

Average Seats by Fuel Type

- All fuel types average around 5 seats, showing standard sedan/hatchback configurations.
- Diesel vehicles may show slightly higher average seats, possibly reflecting a higher share of SUVs or utility vehicles.
- No significant variation across fuel types in terms of seating.

Insight: Seating capacity remains largely uniform across fuel types, indicating that body style and segment are more critical to seating than fuel type.

Discussion & Key Findings

- Cluster 0 represents premium or newer listings.
- Diesel cars tend to be driven more than petrol ones on average.
- Manual transmission dominates the second-hand market.

- Selling price has a strong inverse relation with kilometers driven.
- Most used cars are 5-seaters and models between 2011–2017.

These insights can assist used car platforms and dealers in:

- *Pricing strategy*
- *Inventory categorization*
- *Targeted advertisements based on cluster profiles*

Future Outlook & Recommendations

1. *Integrate Electric Vehicle (EV) Labels*

Add specific EV tags to compare EV market growth and pricing strategies.

2. *Use NLP on Car Names*

Model names can be tokenized to extract manufacturer trends and performance groupings.

3. *Geo-tagging & Region-wise Segmentation*

Adding location data would unlock regional price trends and buyer preferences.

4. *Customer Sentiment Analysis*

User reviews or seller notes could offer qualitative insights.

5. *Prediction Models*

Implement regression models to predict selling price or classify quality tiers.

Conclusion

This analysis of the used car_deatails market reveals important patterns and trends through a combination of clustering algorithms, visualization techniques, and statistical summaries. By examining key features such as *fuel type, year of manufacture, selling price, kilometers driven, transmission type, and ownership*, we were able to identify distinct clusters of vehicles with varying characteristics and resale values. The clustering process, using KMeans and PCA, segmented the cars into three meaningful groups: *newer, high-priced, low-mileage vehicles; mid-aged, moderately priced cars; and older, high-mileage vehicles with lower resale value*. These clusters reflect the diversity of buyer preferences and vehicle conditions in the second-hand market.

Fuel type analysis showed that petrol and diesel cars still dominate, but there is a small yet growing presence of CNG and LPG vehicles, suggesting increased awareness toward cost-efficiency and environmental impact. Diesel vehicles tend to have higher resale values and usage, reflecting their continued appeal for long-distance and commercial applications.

Transmission trends revealed a strong preference for manual vehicles, while the majority of sales come from individual sellers rather than dealerships, indicating that peer-to-peer transactions remain the dominant mode in this market. The ownership trend emphasized that first-owner vehicles are highly preferred, with better price retention and perceived reliability.

Overall, the insights derived from this project provide valuable direction for used car dealers, online platforms, and buyers. These findings can support strategic decisions in pricing, inventory management, and customer targeting. Moreover, as India's automotive sector moves toward greater adoption of electric vehicles and stricter environmental policies, incorporating EV-specific data, regional segmentation, and customer sentiment could greatly enhance future analyses. This project underscores the value of data science in uncovering actionable insights and preparing businesses for a more informed and sustainable automotive future.