



- Home

- About

- Contact

Get Started ➞

USED CAR MARKET ANALYSIS

From Raw Data to Pricing Strategy
Comprehensive Exploratory Data Analysis (EDA)



MEET THE TEAM

- **Mohamed Younis**
- **Shahd Alhindawy**
- **Mariam Fouad**



PROJECT GOALS & OBJECTIVES

The Problem: High volatility and inconsistent pricing make fair valuation difficult for buyers and sellers.

Main Goal: Identify and quantify the key factors that drive the Used Car Selling Price (Selling_Price)



- **Key Objectives:**

- Perform rigorous data cleaning and feature engineering.
- Extract clear, data-driven business insights.
- Lay the foundation for a predictive Machine Learning model.

DATA OVERVIEW & INITIAL CHALLENGES

	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner
0	Maruti 800 AC	2007.0	NaN	70000.0	Petrrol	NaN	Manual	First Owner
1	Maruti Wagon R LXI Minor	Nan	135000.0	50000.0	Petrol	Individual	Manual	First Owner
2	Hyundai Verna 1.6 SX	2012.0	600000.0	100000.0	Deisel	Individual	Manual	First Owner
3	Datsun RediGO T Option	2017.0	250000.0	46000.0	Petrol	Individul	Manual	First Owner
4	Honda Amaze VX i-DTEC	2014.0	450000.0	141000.0	Diesl	Individul	Manual	Second Owner

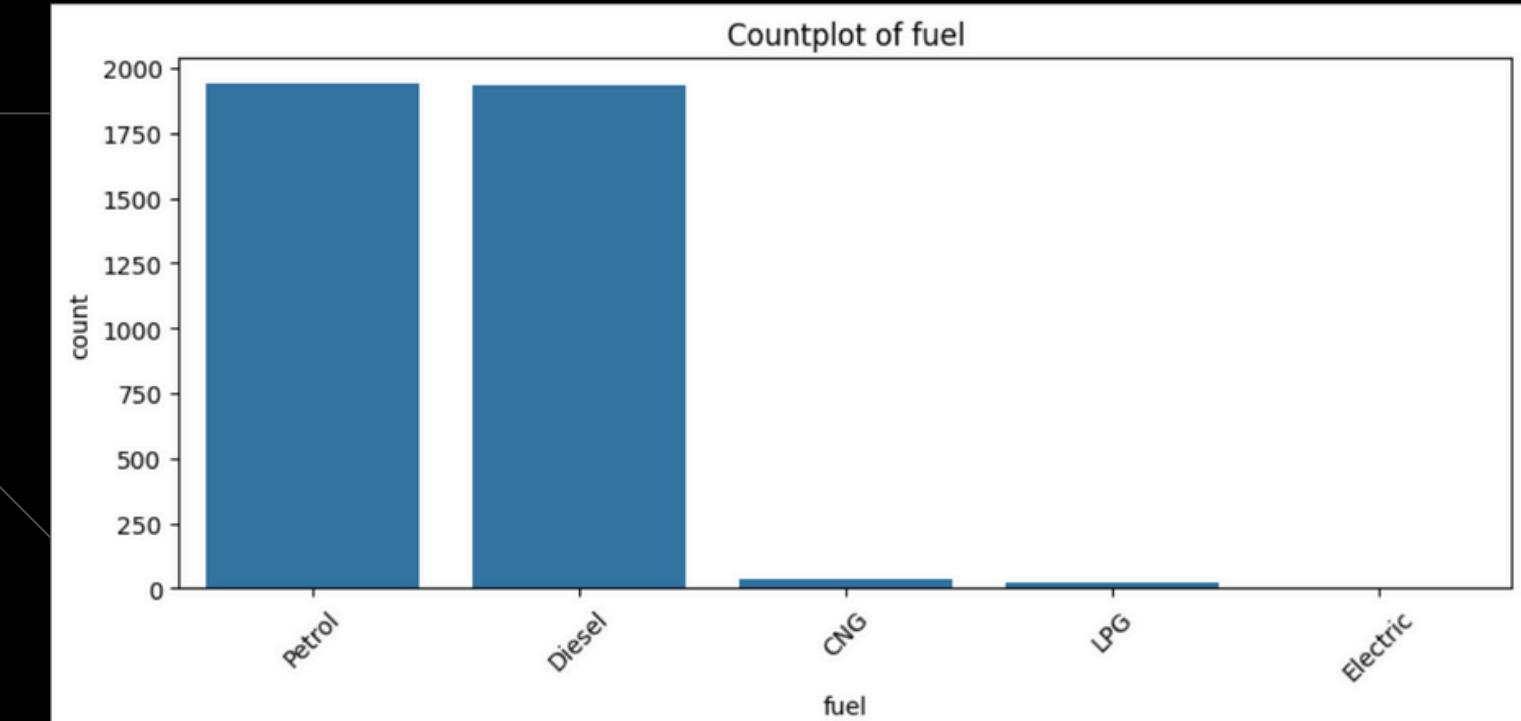
- **Source:** Used Car Resale Data
- **Volume:** 4345 Rows (Cars), 8 Columns (Features).
- **Key Columns:** Name, Year, Selling_Price, Km_Driven, Fuel_Type, Transmission.

- **Initial Challenges:**
 - Missing and unrealistic values (e.g., car year 1800).
 - Inconsistent spelling in categorical features (e.g., "Petrrol").
 - Extreme outliers in Selling_Price and Km_Driven.

DATA CLEANING: CATEGORICAL VARIABLES

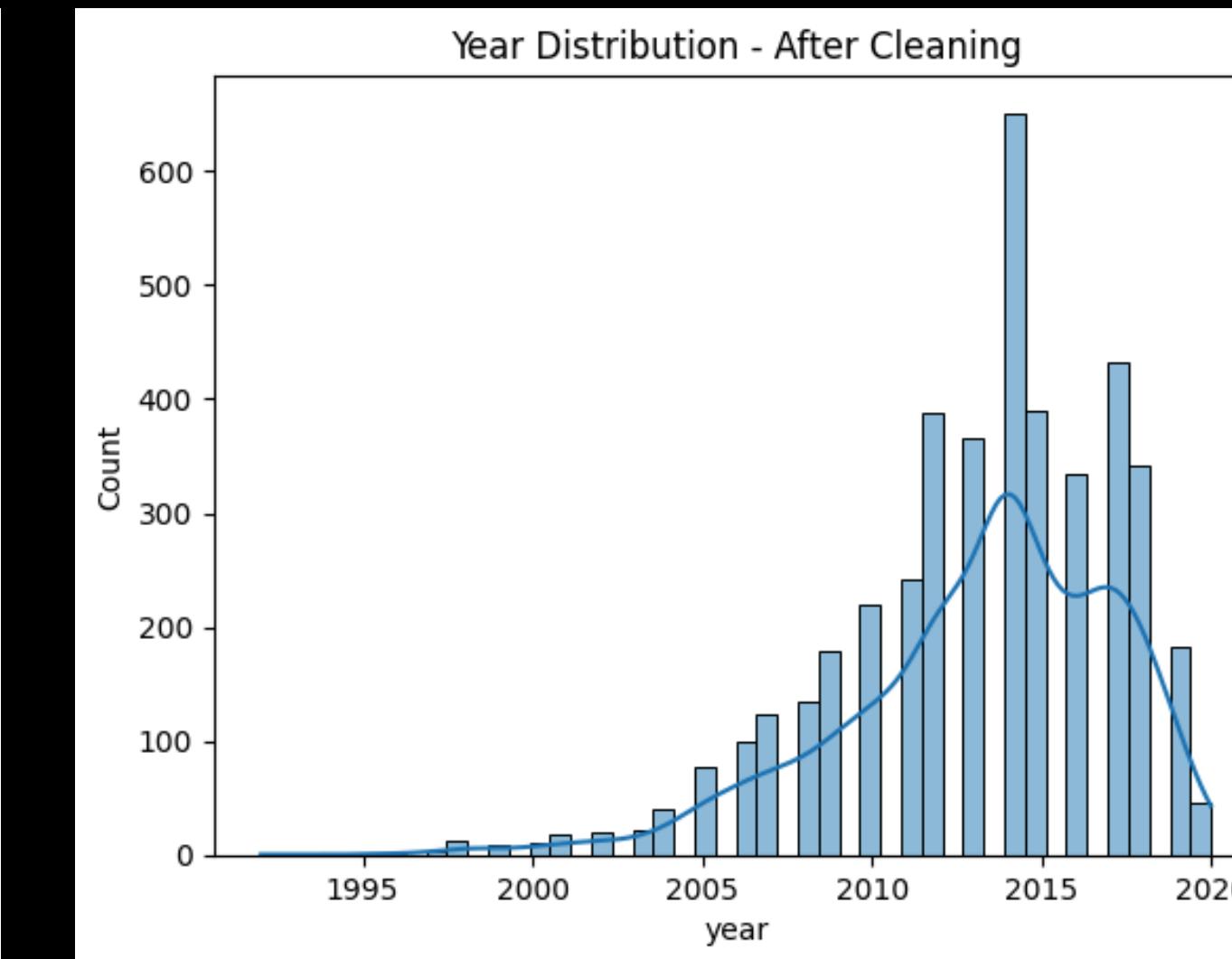
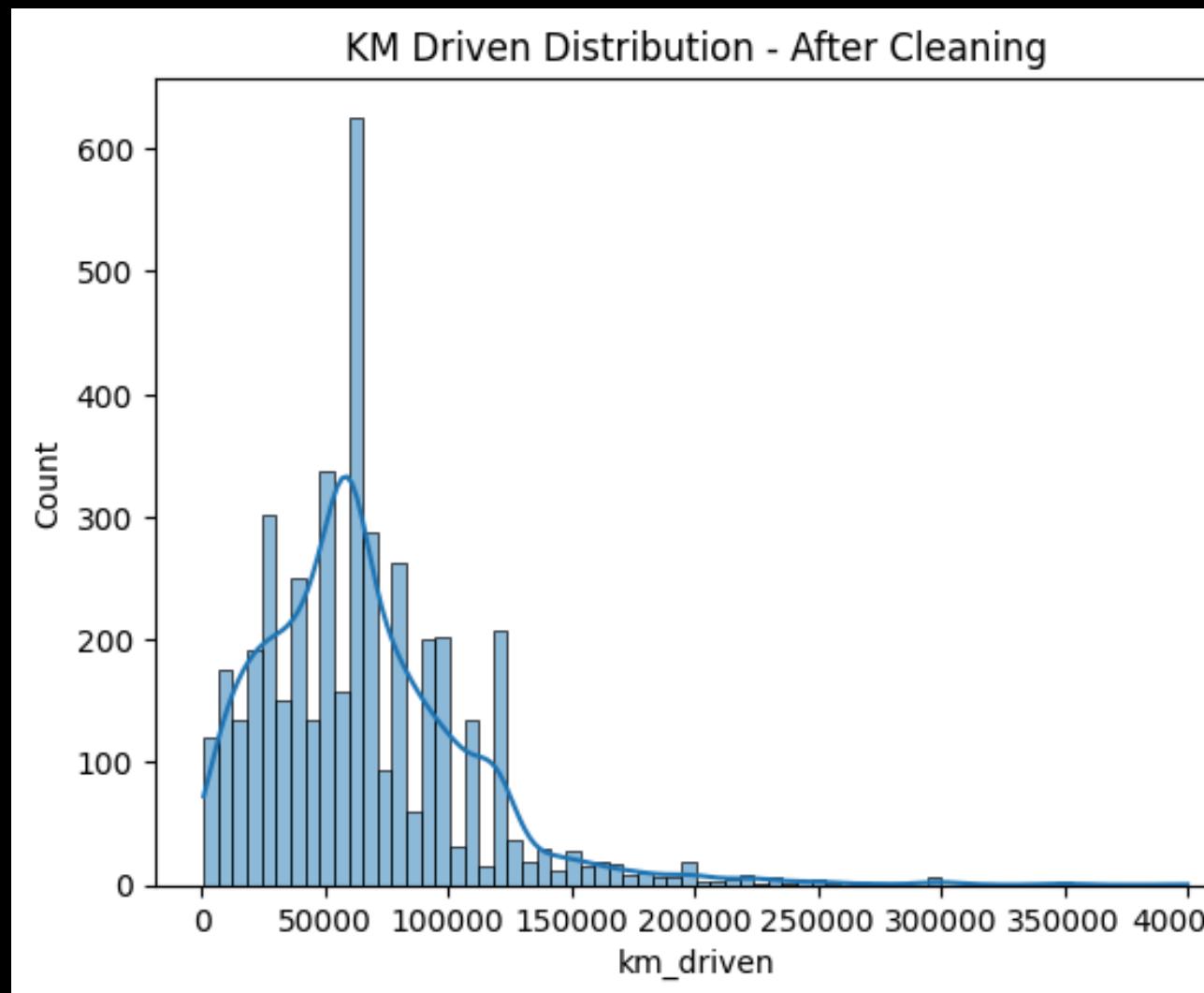
- **Focus:** Fuel_Type, Seller_Type, Transmission.
- **Action Taken:** Uniform formatting (Lowercasing and stripping whitespace).
- Correcting spelling errors and consolidating similar categories.
- **Result:** Consistent and clean categories, ready for accurate grouping and analysis.

```
Unique values in fuel BEFORE cleaning:  
['Petrrol' 'Petrol' 'Deisel' 'Diesl' 'Petrl' 'CNG' nan 'LPG' 'Electric']  
  
Unique values in seller_type BEFORE cleaning:  
[nan 'Indvidual' 'Individul' 'Deelar' 'Dealerr' 'Trustmark Dealer']  
  
Unique values in transmission BEFORE cleaning:  
['Manual' nan 'Automatic']  
  
Unique values in owner BEFORE cleaning:  
['First Owner' 'Second Owner' nan 'Fourth & Above Owner' 'Third Owner'  
'Test Drive Car']
```



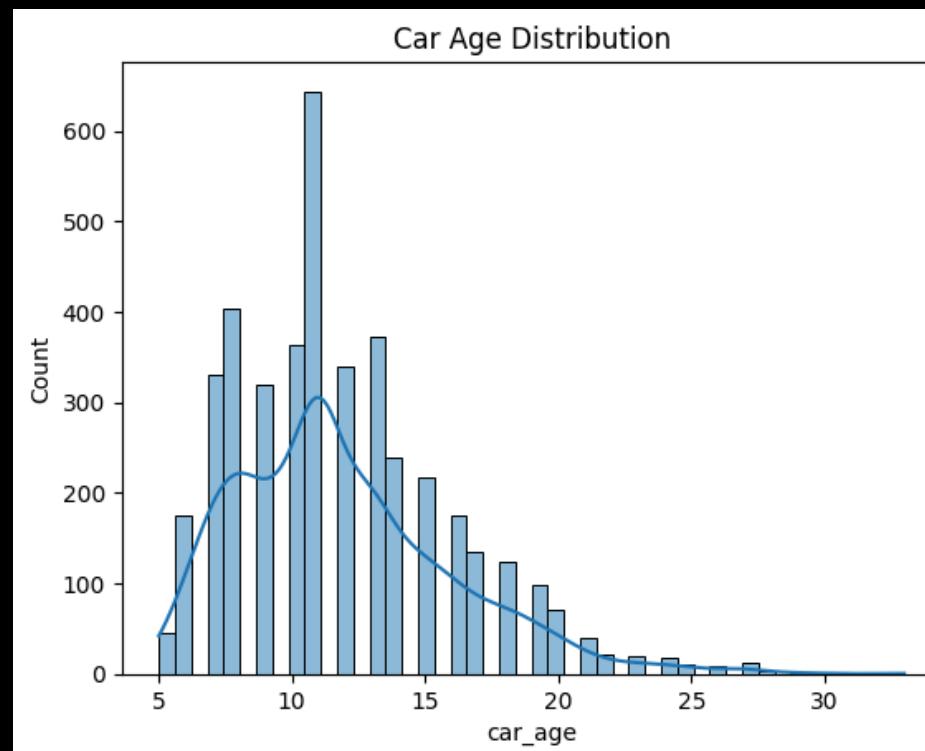
DATA CLEANING: YEAR, KM, AND MISSING VALUES

- **Year:** Filtered out logically impossible years (e.g., < 1980 or > 2025).
- **Km_Driven:** Cleaned text (removed units/commas) and addressed non-realistic mileage values.
- **Missing Values (Imputation):** Filled numeric Nulls with the Median, Filled categorical Nulls with the Mode.

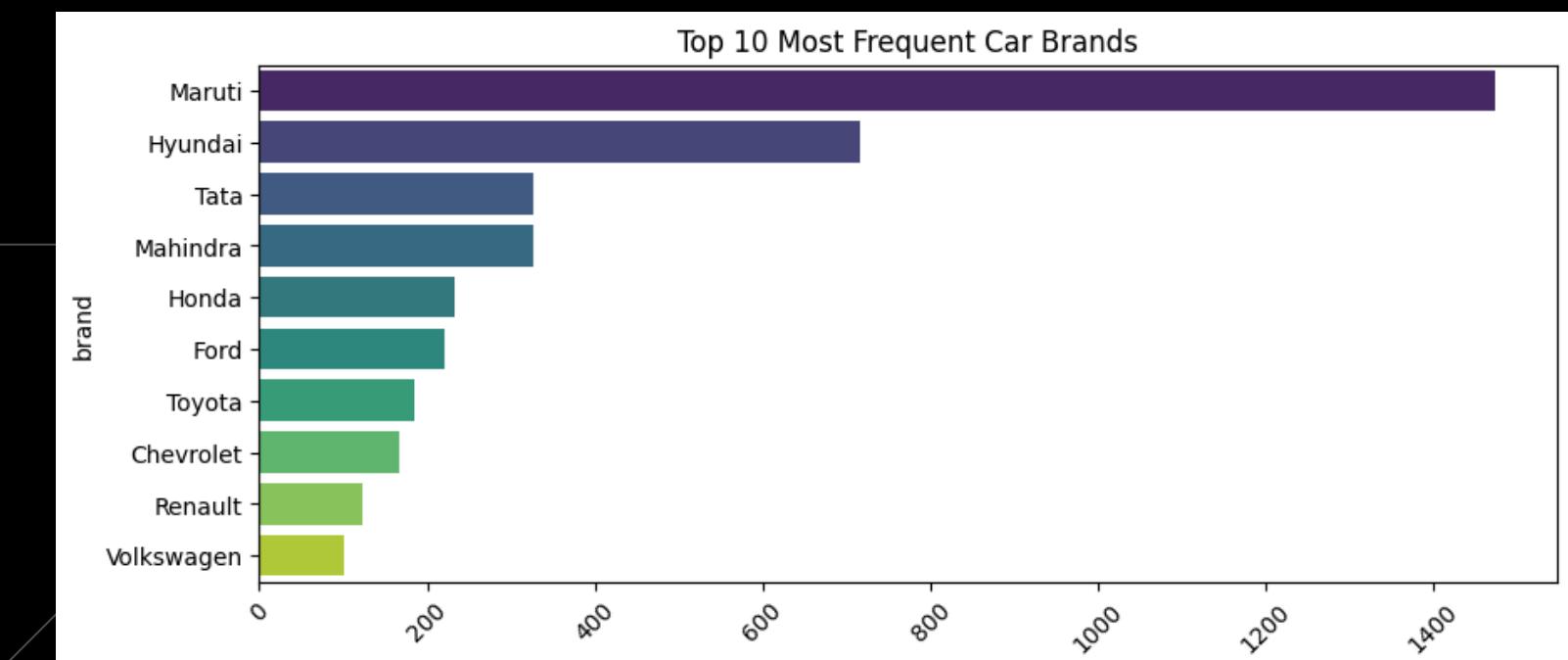


FEATURE ENGINEERING

- **Car_Age:** Calculated as (Current Year - Year).
This is a more intuitive measure of depreciation.



	year	car_age
0	2007	18
1	2014	11
2	2012	13
3	2017	8
4	2014	11



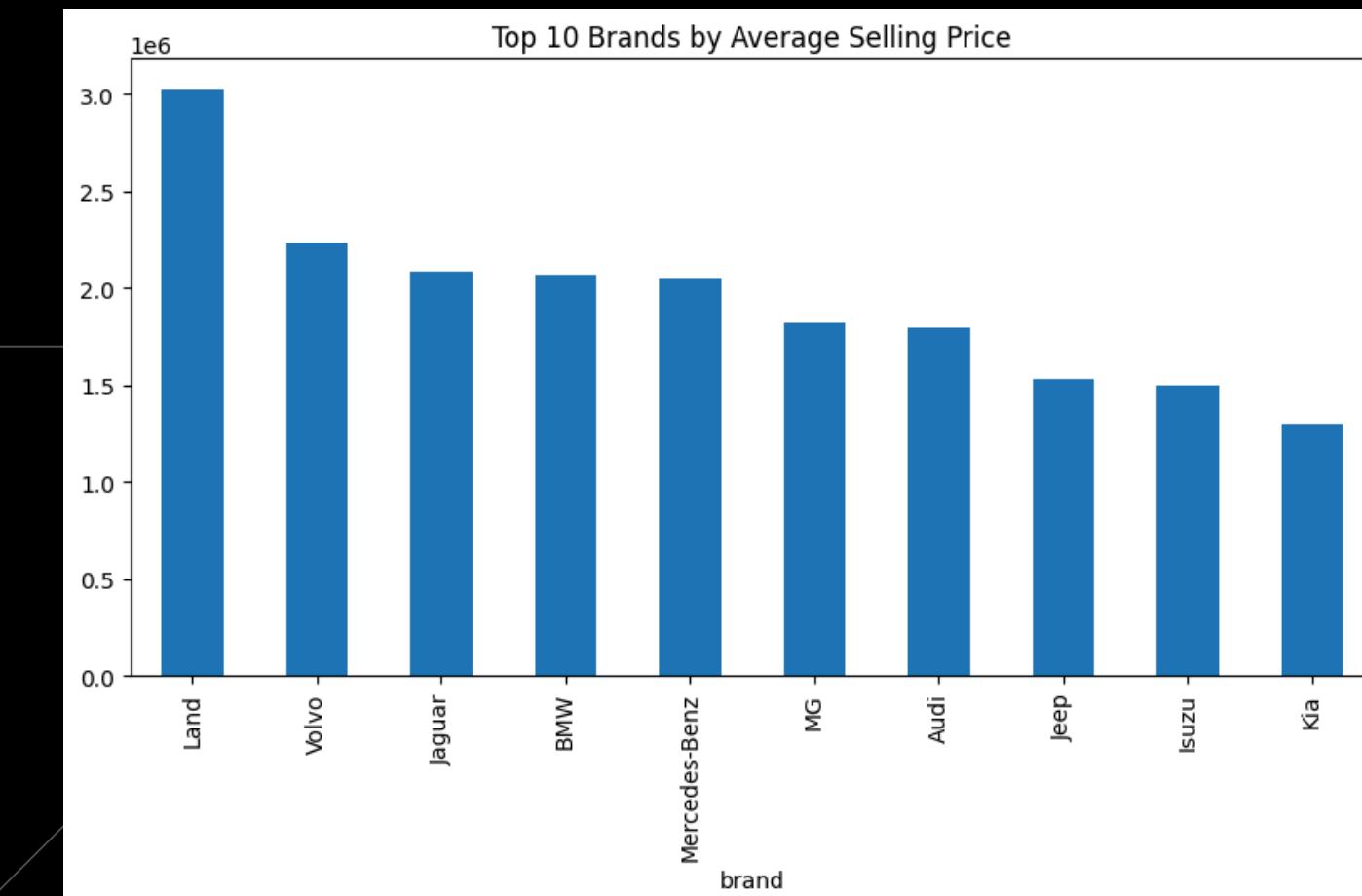
FEATURE ENGINEERING

- **Price_Per_Km:** Ratio for value assessment.

	selling_price	km_driven	price_per_km
0	350000.0	70000.0	5.000000
1	135000.0	50000.0	2.700000
2	600000.0	100000.0	6.000000
3	250000.0	46000.0	5.434783
4	450000.0	141000.0	3.191489

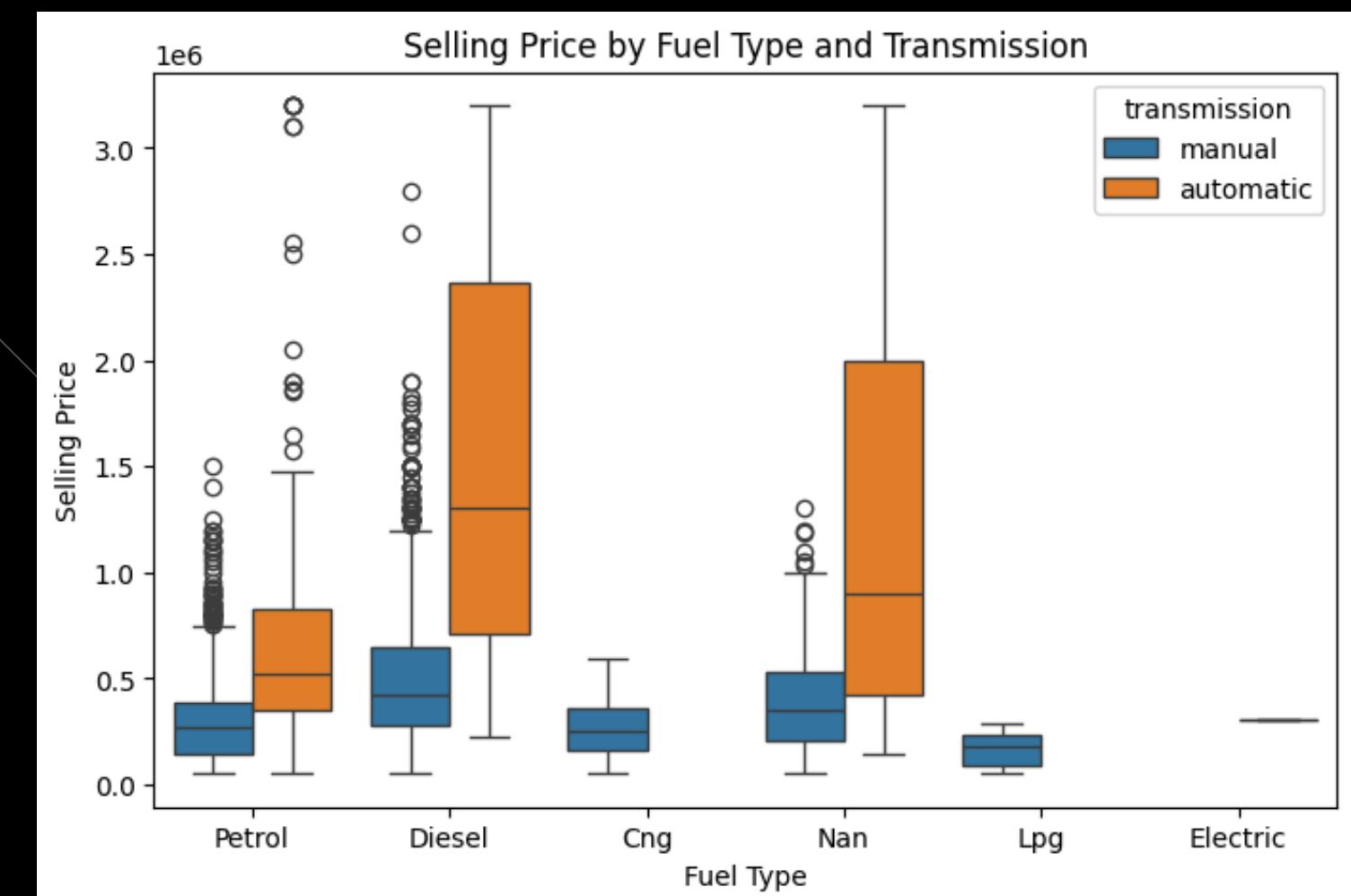
- **Insight 1: The Depreciation Engine**

Age and Mileage are the two dominant factors determining resale value.



INSIGHT 2: TRANSMISSION AND FUEL TYPE IMPACT

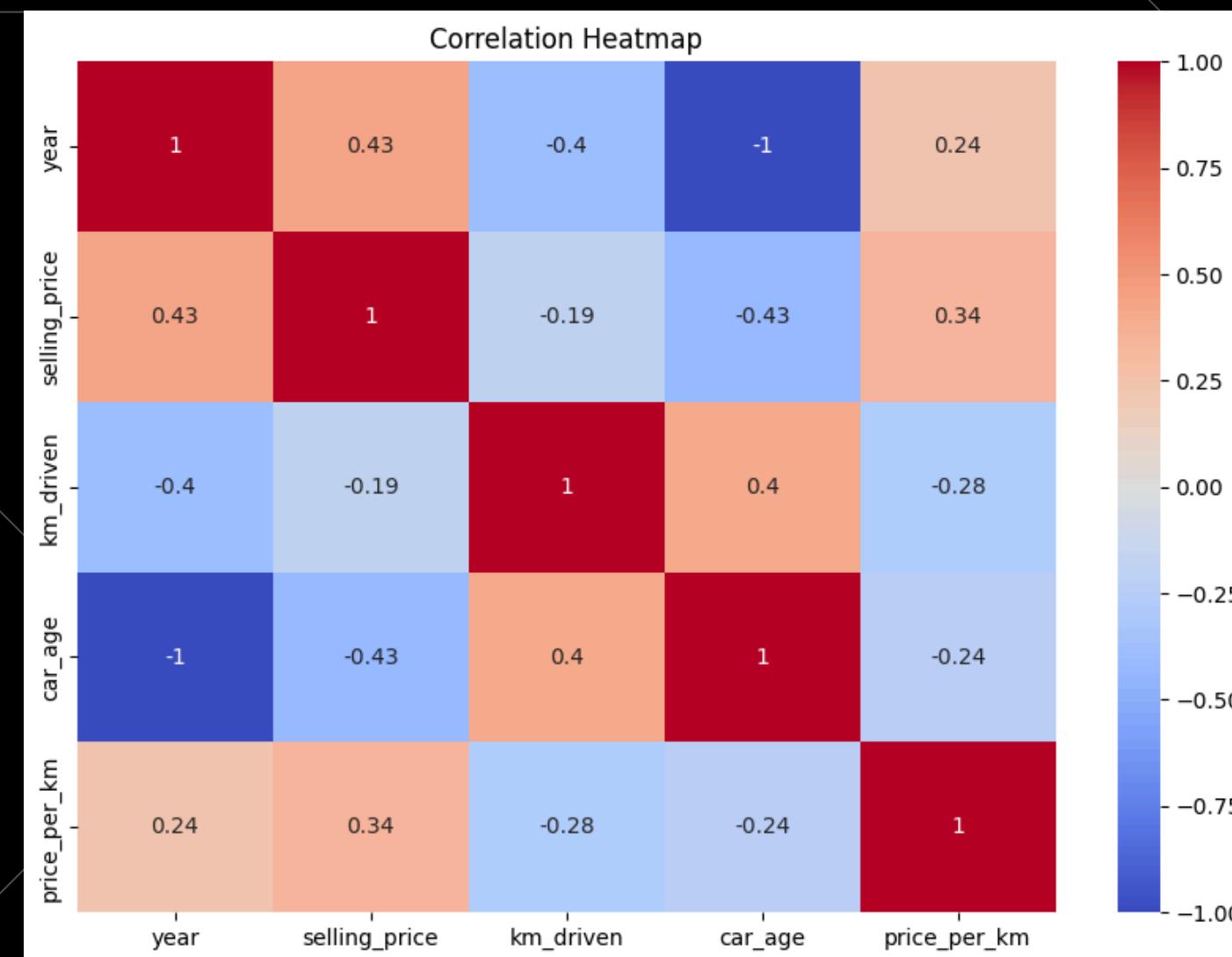
- **Visual Aid:** Box plots comparing Selling_Price across transmission and fuel types.
- **Transmission:** Automatic cars command a significantly higher average selling price compared to Manual.
- **Fuel Type:** Diesel cars generally maintain a higher resale value than Petrol cars, especially in higher price segments.
- **The Winner:** The Diesel + Automatic combination is the most valuable setup.



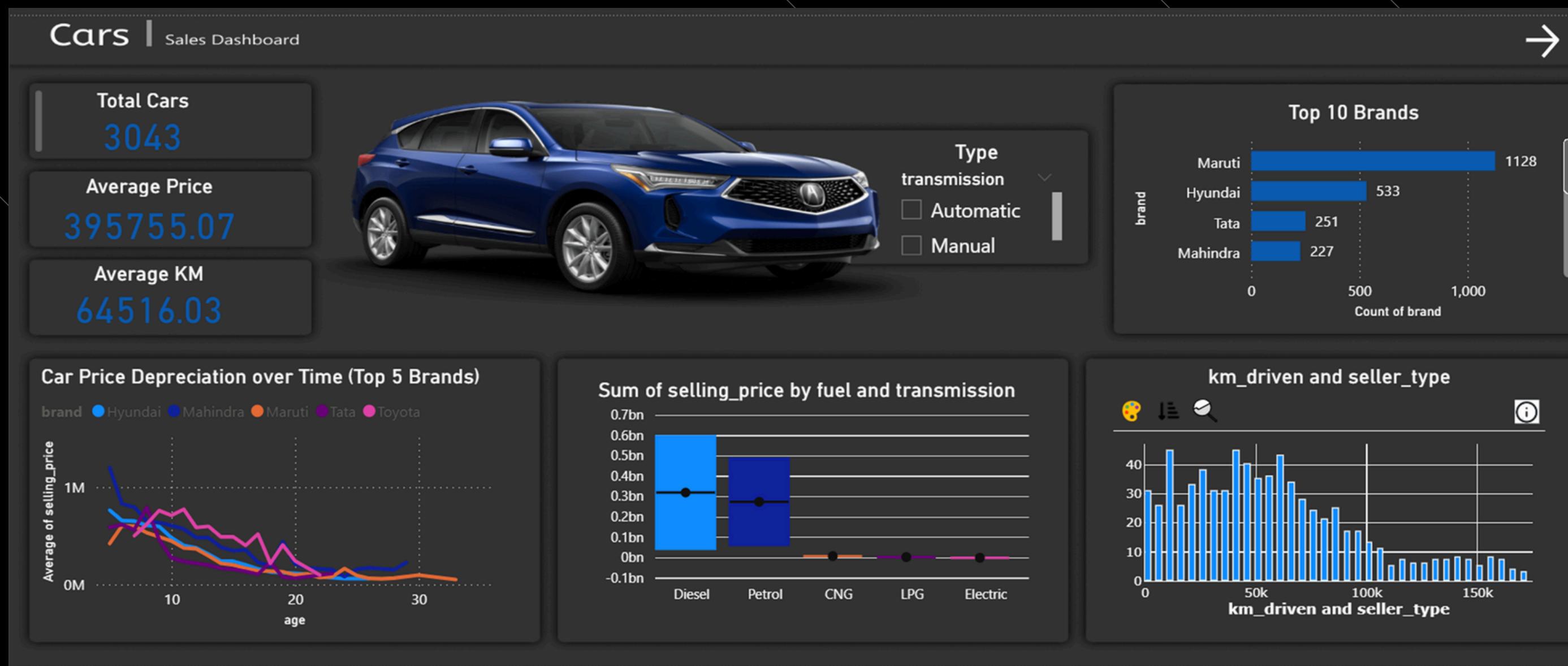


CORRELATION SUMMARY

- **Summary:** Confirms that Car_Age, Km_Driven, and Transmission are the most influential variables in determining Selling_Price.
- **Actionable Takeaway:** These variables must be prioritized in any pricing or valuation mode



INTERACTIVE DASHBOARD OVERVIEW



CONCLUSION & NEXT STEPS

- **Conclusion:** The analysis successfully identified the primary drivers of used car prices (Age, Km, Transmission, Owner).
- **Recommendations:**
 - Sellers should aim to sell before 5 years to maximize value.
 - Buyers seeking value should target well-maintained Second Owner, Manual cars.
- Next Steps: Develop and train a Regression Model (e.g., Random Forest) to predict price with high accuracy and Integrate the model into a live application for real-time pricing suggestions.



Home

About

Contact



THANK YOU



Slide

13