Dataset Overview

- Source: CarDekho Open Dataset
- 4340 entries, 8 initial features
- Key columns: Selling Price, Year, Km Driven, Fuel, Transmission

Data Inspection

- Loaded CSV using pandas
- Checked for missing values
- Inspected data types and sample records

Data Cleaning

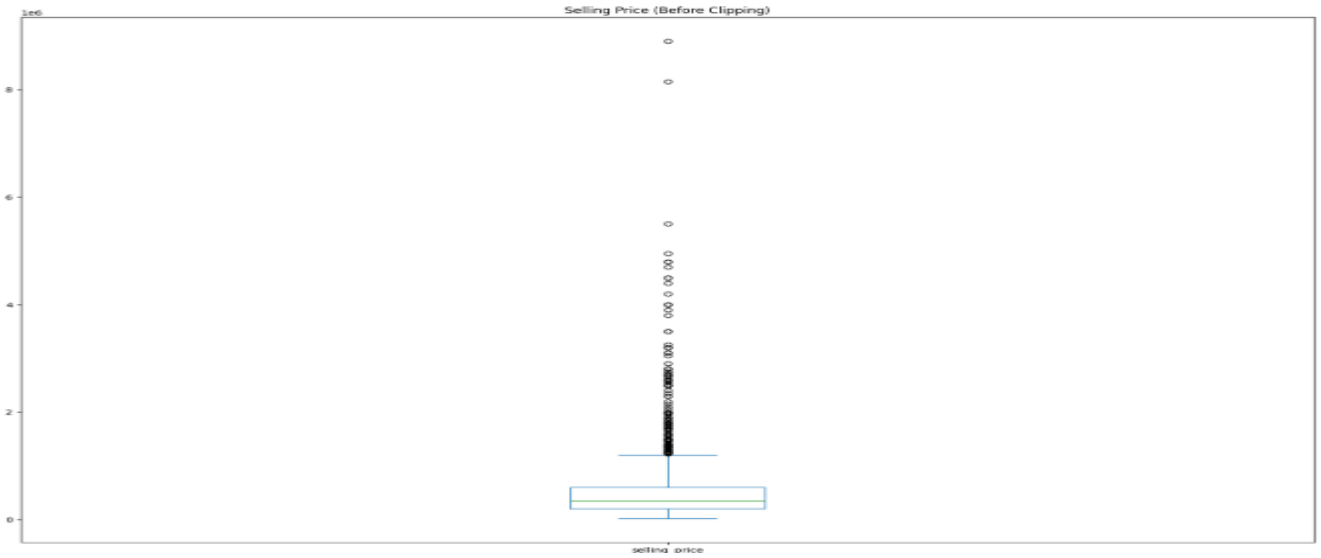
- Removed 767 duplicate records
- Dropped irrelevant 'Owner' column
- Confirmed no missing values

Feature Engineering

- Created 'Car Age' = 2025 - Year
- Calculated 'Price per KM' = Selling Price / Kms Driven
- Grouped prices into categories

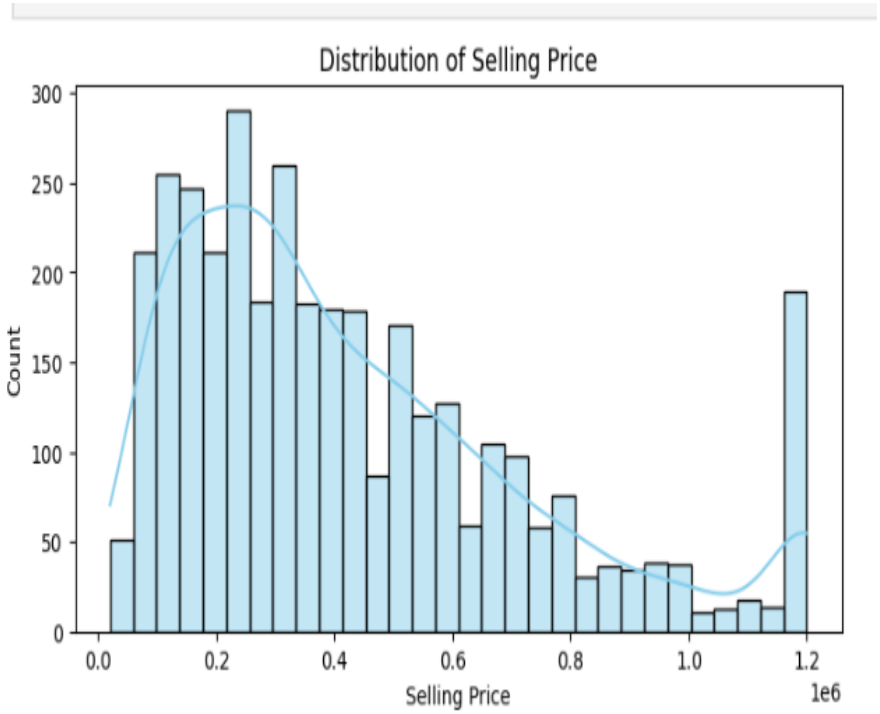
Outlier Handling

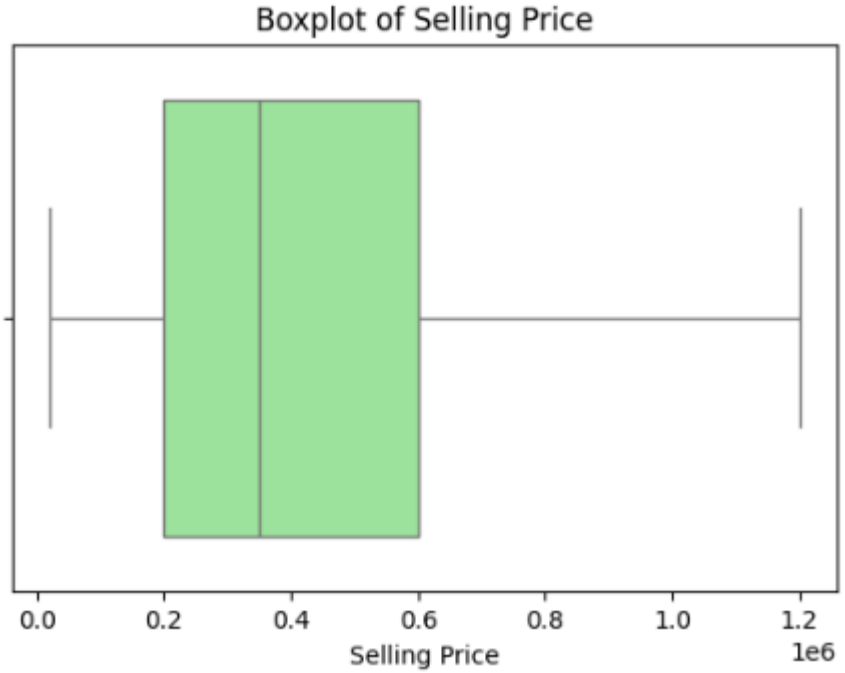
- Used boxplots to detect outliers
- Applied IQR clipping to Selling Price
- Stabilized price distribution



Univariate Analysis

- Selling Price: Right-skewed distribution
- The boxplot shows that most car prices lie between ₹2–5 lakhs, with a few high-value cars extending the range.



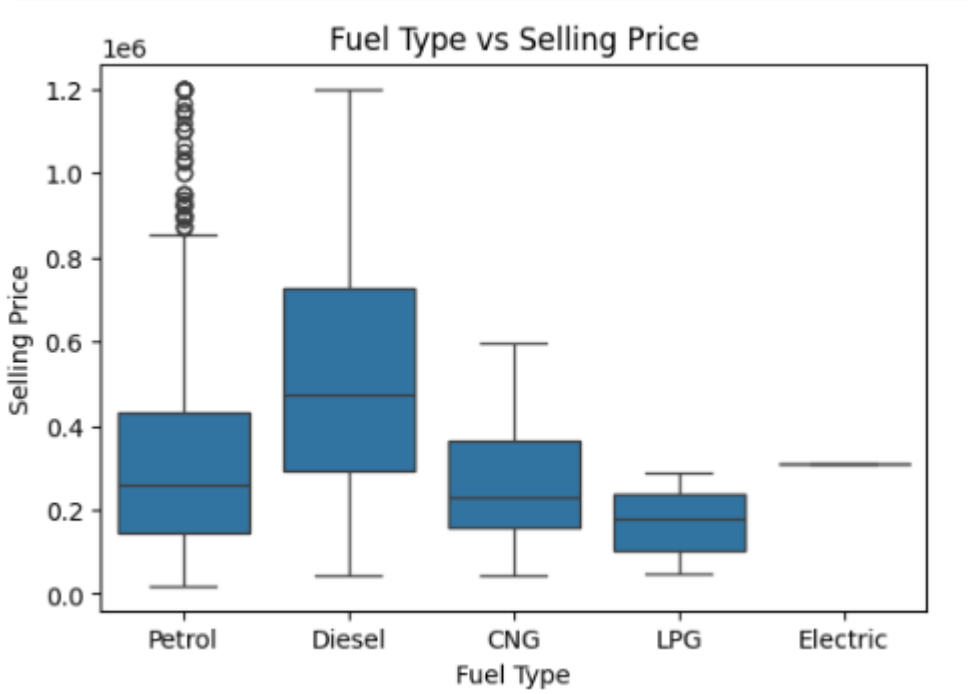


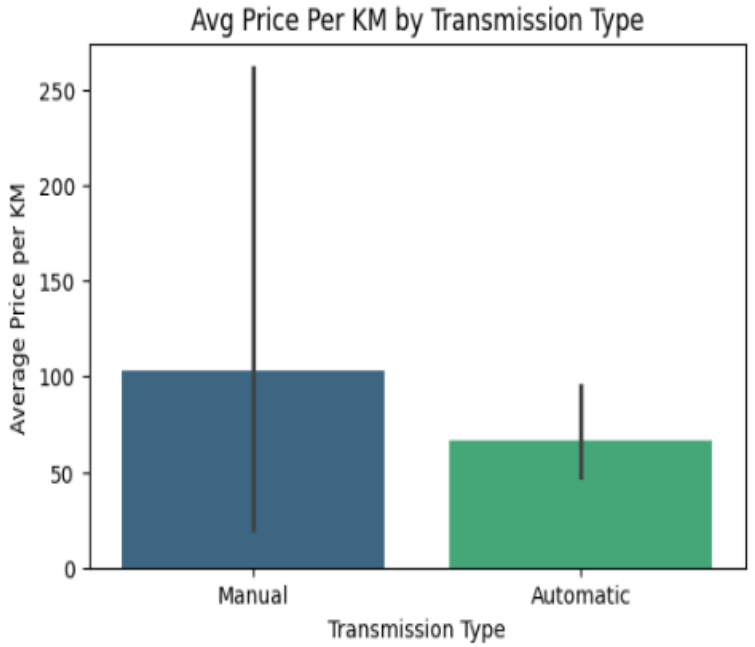
Bivariate Analysis

- Scatter Plot: Car Age vs Selling Price
- Correlation Matrix: Car Age and Price (-0.6)

Visualizations

- Boxplot: Selling Price vs Fuel Type
- Barplot: Transmission Type Count





Key Findings

- Older cars depreciate faster
- Diesel cars retain more value
- Manual transmission dominates market
- Lower Km Driven improves price

Conclusion

- Age, Fuel, Transmission major factors
- Data Cleaning improved quality
- Feature Engineering enabled better insights