



Used Cars Data Analysis – EDA Project

Exploratory Data Analysis on Used Cars Dataset

Fatma shehata ewas
Abdelaziz elbana



Project Objective

Dataset Overview

Data Cleaning Process

Individual insight

Key Relationships

Conclusion

Project Objective



Objective of This Project:

Perform complete Exploratory Data Analysis (EDA) on a used cars dataset

Understand:

- Price behavior
- Usage patterns
- Market structure
- Factors affecting car prices

Prepare the data for:

- Future Machine Learning models
- Business decision making
- Interactive dashboard visualization

Dataset Overview



Dataset Features:

- Selling Price
- Km Driven
- Year
- Fuel Type
- Seller Type
- Transmission
- Owner
- Brand

Engineered Features:

- Car Age
- Km Per Year
- Name Length
- Brand Average Price

Final Dataset Size:

2,927 Clean Records

Data Cleaning Process



Main Cleaning Steps

Converted all numeric columns from object to numeric



Removed:

1. Negative & zero prices
2. Unrealistic kilometers
3. Invalid manufacturing years
4. Car age above 50 years
5. Handled missing values
6. Removed extreme unrealistic outliers
7. Recalculated Brand Average Price

Result:

Clean, reliable, and analysis-ready dataset

Selling Price Analysis



Average Selling Price \approx 465,000
Most cars fall between 200,000 – 600,000
High-priced cars exist but represent premium brands

After cleaning:

Price distribution became balanced and realistic



Insight:

The used car market mainly targets mid-range buyers

Km Driven & Km Per Year



- Average Km Driven \approx 520,000
- Average Km per Year \approx 46,000
- Extreme yearly usage values were removed



Insight:

Most cars show moderate and realistic usage behavior.

Car Age Analysis



- Average Car Age \approx 11 years
- Most cars range between 8 – 13 years
- Very old vehicles were removed



Insight:

The dataset mainly represents well-used market-active cars

Fuel Type Distribution

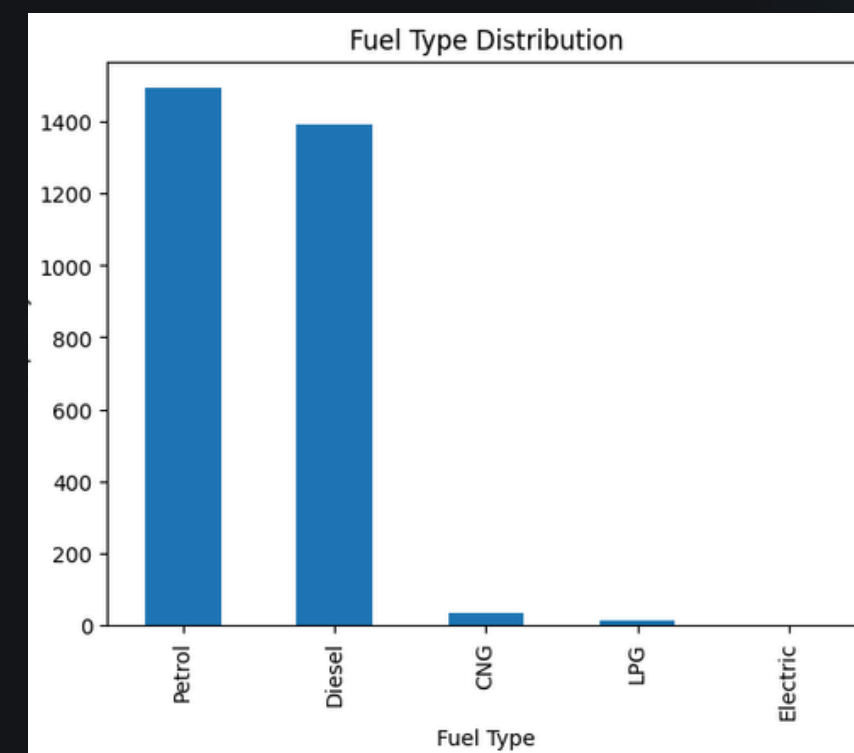


- Petrol is the dominant fuel type
- Diesel is second
- CNG, LPG, Electric → very limited presence



Insight:

Petrol remains the most preferred fuel in the used car market.



Transmission Analysis



- Manual transmission dominates the market
- Automatic cars form a small proportion



Insight:

Manual transmission is still the most common choice.

Seller Type Analysis



Most cars are sold by:

- Individual sellers
- Dealers represent a smaller portion



Insight:

The market depends heavily on direct individual selling

Brand Analysis

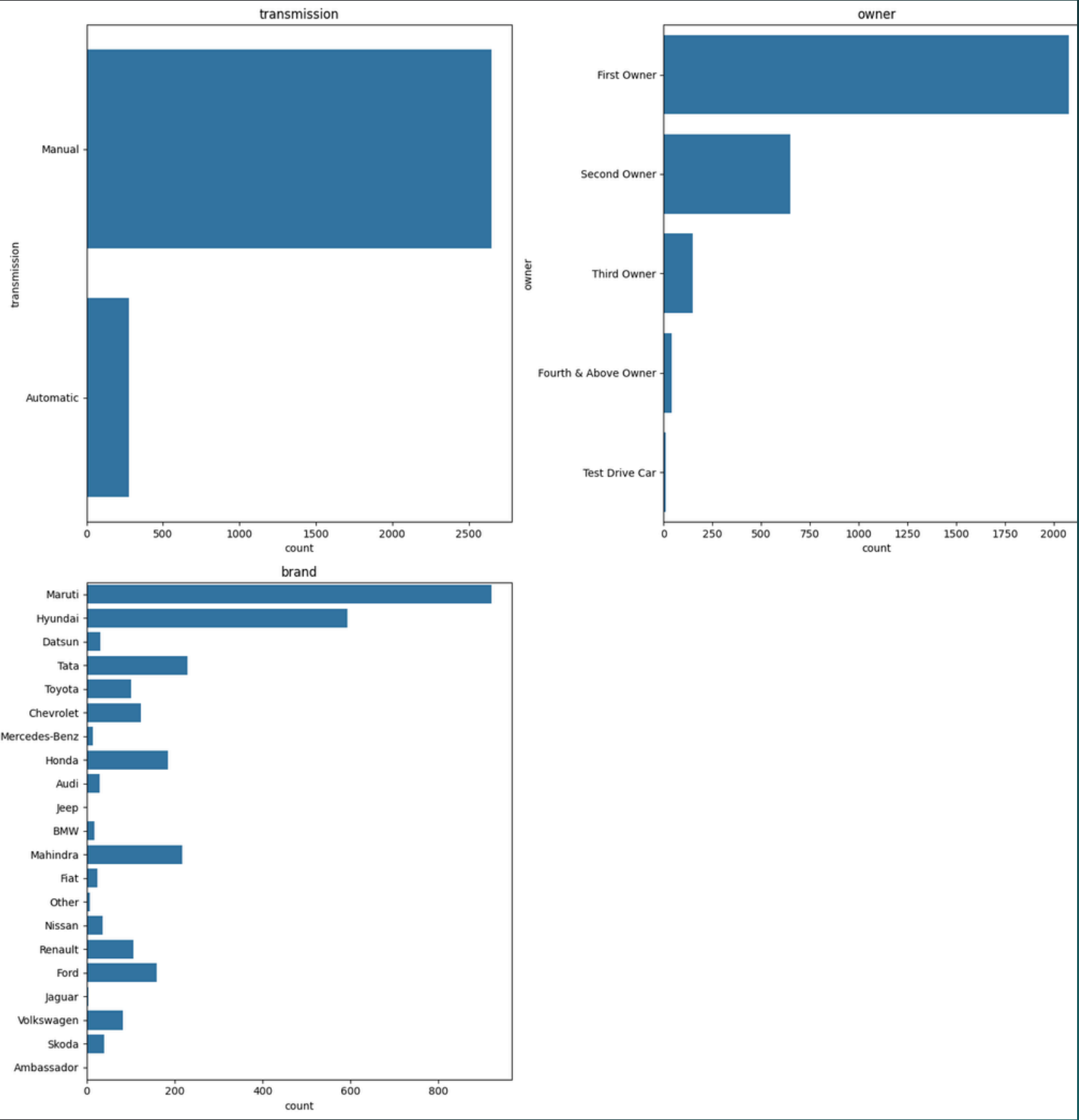


- The dataset includes 19 brands
- Maruti is the most dominant brand
- Brand average prices vary significantly



Insight:

Economy brands dominate the used car market.



Owner Distribution



Majority are:

- First Owner cars
- Fewer second and third owner vehicles



Insight:

Most cars have low ownership turnover.

Selling Price decreases with

- Higher Car Age
- Higher Km Driven

Higher prices are linked to

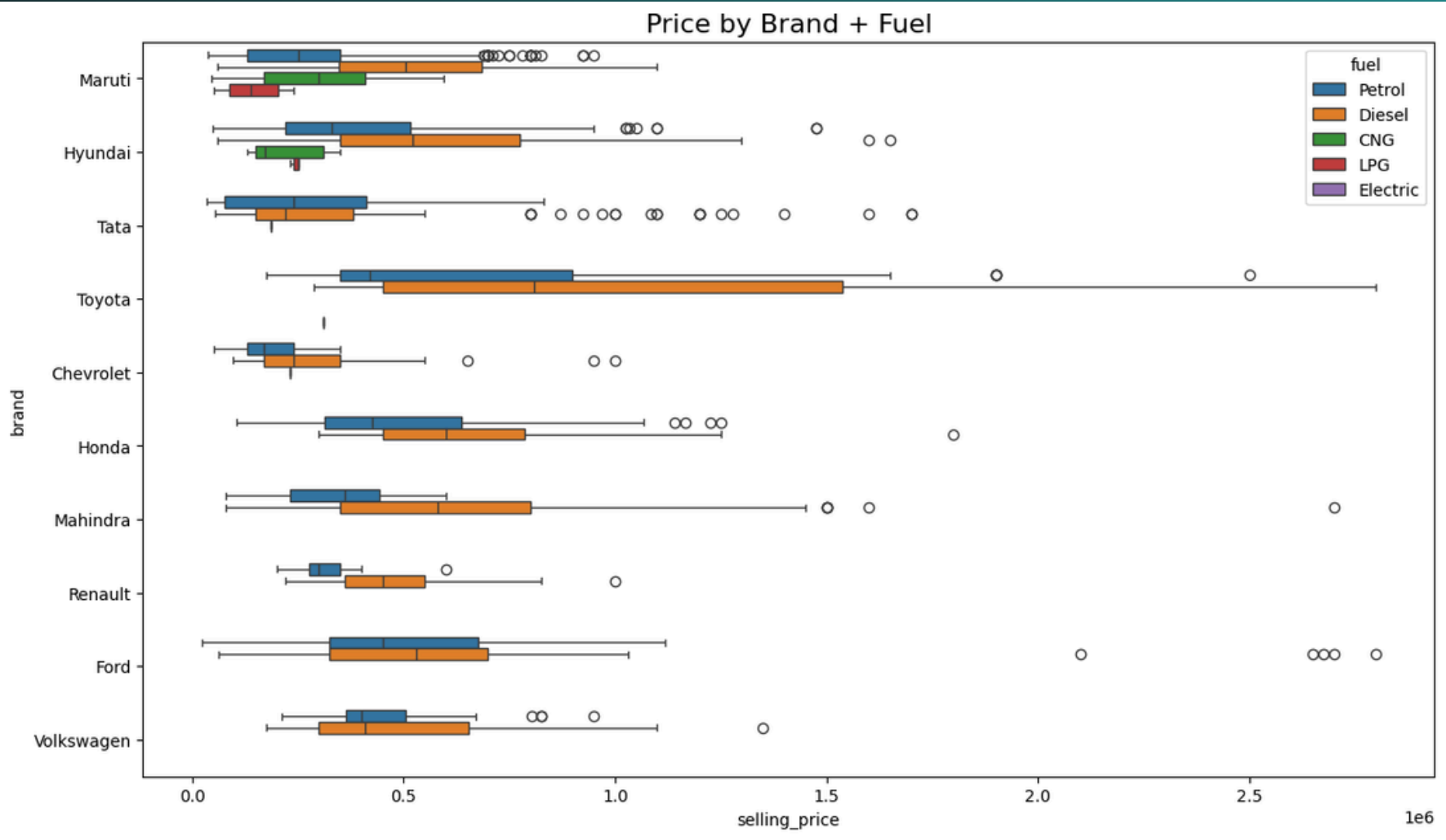
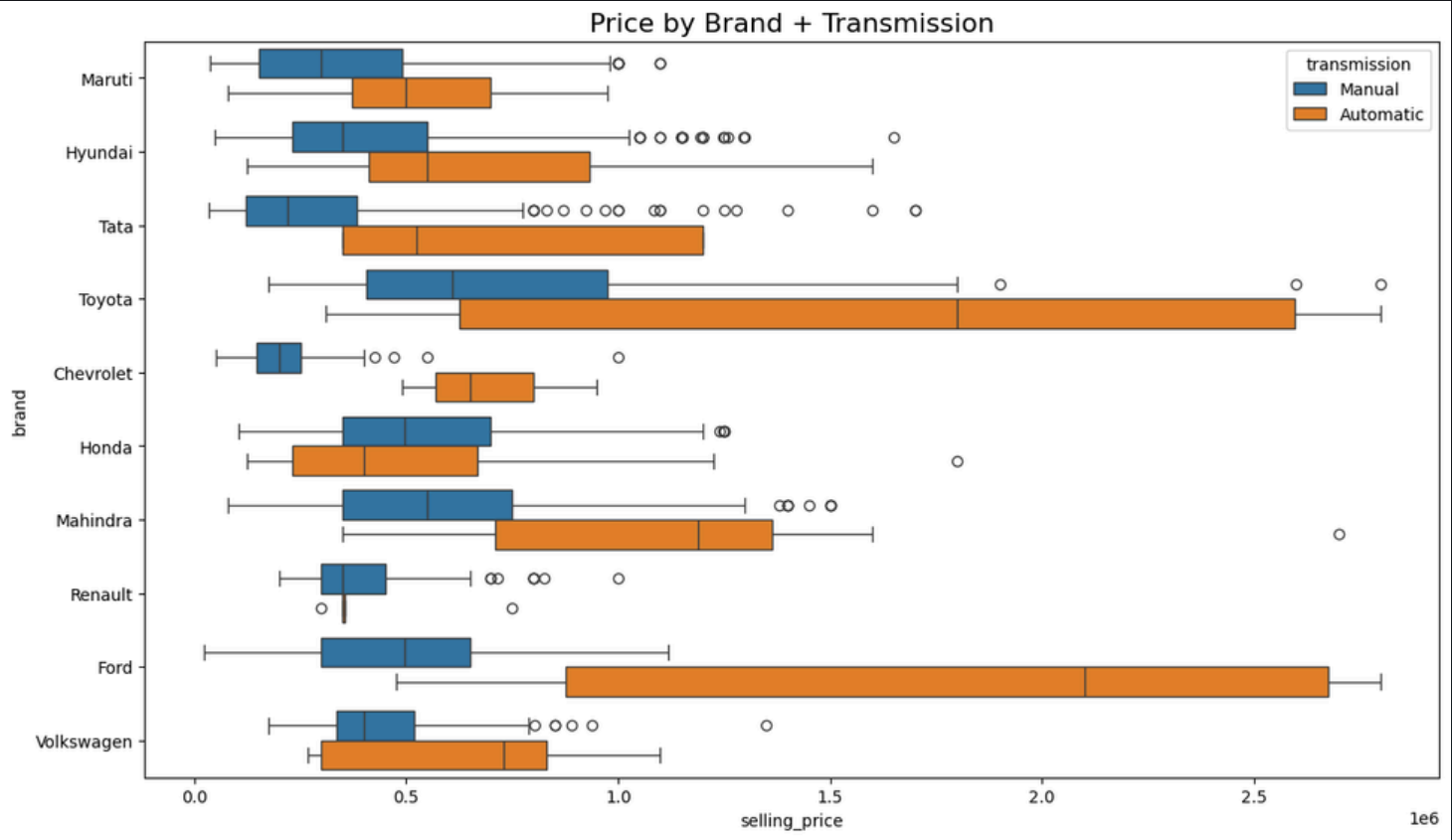
- Automatic transmission
- First owner
- Low mileage

Key Relationships



Insight:

Mileage, age, and ownership are the strongest price drivers.



conclusion

The used car market represented in this dataset is mainly driven by affordable petrol-powered manual cars sold by individuals. Car price is strongly affected by age, mileage, ownership history, and transmission type. After applying strong data cleaning and outlier handling, the dataset became reliable for professional data analysis and dashboard visualization.

thank you