CarDekho Used Car Listings Analysis

Project Category: Exploratory Data Analysis

Tools & Technologies Used: Python, NumPy, Pandas, Jupyter Notebook, Tableau

🌐 Introduction

This project explores a dataset of used car listings from a car-buy/sell platform similar to CarDekho. The objective is to uncover patterns and insights that influence car pricing and demand in the second-hand market. The goal is to build a Tableau dashboard that presents these findings in a clean, actionable, and interactive way for business stakeholders, analysts, and product teams.

Key Business Questions:

- Which features (e.g., mileage, power, brand) influence selling price?
- Which price ranges and car types dominate the used car market?
- Which brands and configurations are most in demand?

📜 Project Overview

Dataset Summary:

- **Source:** Provided internally, resembling CarDekho listings
- **Dataset:**

> Car details v3.csv

- **Size:** 8,128 records, 13 features
- **Features include:**
  - Car name, year, selling price, km driven
  - Fuel type, seller type, transmission, ownership
  - Mileage, engine size, power, torque, and seats

Main Analysis Steps:

1. Data Loading & Inspection (via Pandas)
2. Data Cleaning & Transformation
3. Feature Extraction (brand & model)
4. Exploratory Data Analysis (EDA)
5. Export to Tableau
6. Dashboard design and storytelling

🔍 Data Exploration

Cleaning & Preprocessing:

- Removed units from `mileage`, `engine`, `max_power` columns
- Converted them to numeric (float)
- Imputed missing values (mode for categorical, median for numeric)
- Extracted `brand_name` and `model_name` from the `name` column
- Cleaned inconsistent text entries in the `owner` column

Null Values (Before Cleaning):

- `mileage`, `engine`, `max_power`, `torque`, and `seats` had missing values
- Most entries were imputed successfully or dropped if unresolvable

---

⚖️ Analysis Methods

Techniques Used:

- Summary statistics & distributions (mean, median, etc.)
- Aggregations & group-by analysis (brand-level, fuel-level, etc.)
- Correlation plots (scatterplots with trendlines)
- Visual analysis via Tableau

Why These Methods?

These techniques were chosen to:

- Understand demand patterns (count, average prices)
- Explore relationships (mileage vs. price, power vs. price)
- Simplify insights for business presentation

📊 Results and Interpretation

Key Findings:

- **Maruti Suzuki** is the most listed brand, indicating strong resale presence
- **₹3L–6L** is the most common price range
- **Diesel** is the most preferred fuel type in the dataset
- **Manual transmission** dominates listings but **automatic** variants command higher average prices
- **Power (bhp)** shows a stronger correlation with price than **mileage**

Feature-Price Trends:

- Scatterplot shows that more powerful cars generally have higher resale prices
- Mileage shows a weaker correlation, suggesting other factors dominate

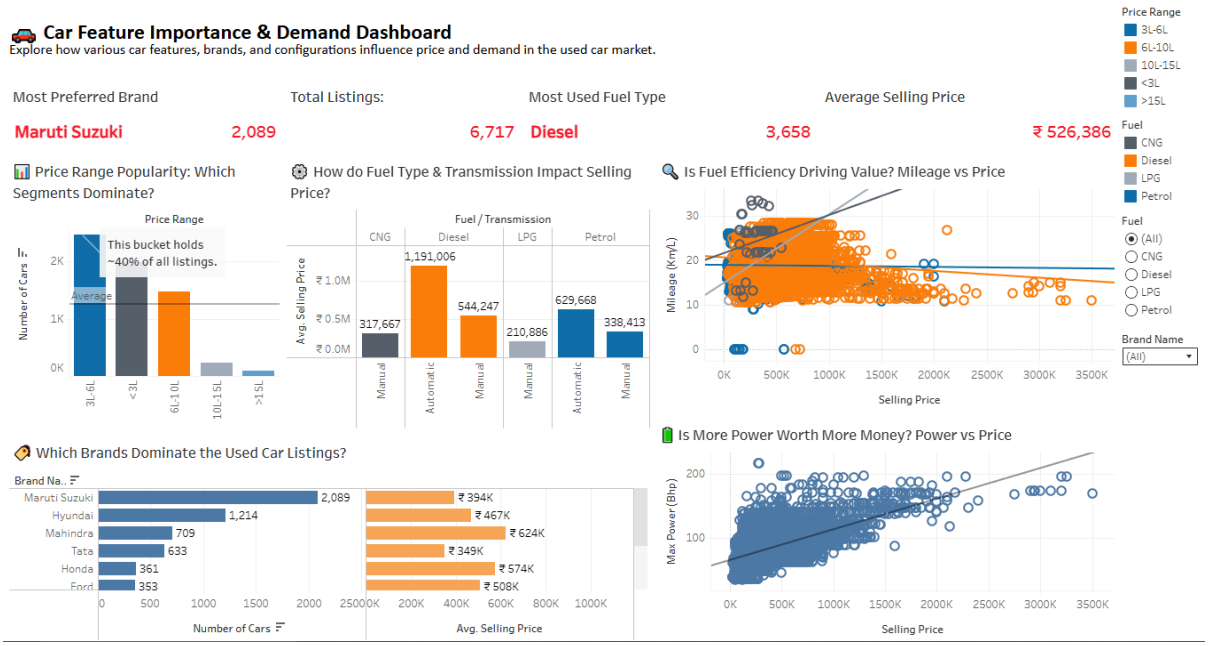
---

🚀 Conclusion

This analysis reveals:

- Price clusters between ₹3–10L and among brands like Maruti and Hyundai
- Feature-driven insights that can improve pricing algorithms and buyer targeting
- Fuel and transmission preferences that impact car value perception

The final Tableau dashboard provides a comprehensive, visual tool to explore these dynamics interactively.



Recommendations

Based on the analysis and dashboard insights, the following actions are recommended for business teams:

1. **Inventory Focus:**
   - Prioritize stock procurement from top-performing brands like Maruti Suzuki and Hyundai.
   - Concentrate listings in the ₹3L–6L and ₹6L–10L price brackets.
2. **Targeted Marketing:**
   - Promote automatic variants for higher margins, especially in urban areas.
   - Highlight engine power in premium listings to attract value-conscious buyers.
3. **User Experience & Filtering:**
   - Enhance filtering features on platforms based on brand, power, and fuel type.
   - Introduce recommendation engines prioritizing price-power combinations.
4. **Pricing Strategy:**
   - Refine price prediction models to include max power, brand, ownership history, and transmission.
   - Create tiered pricing based on power ratings and feature combinations.
5. **Expansion Considerations:**

- Explore strategic promotion of less common but emerging segments (e.g., automatic diesel).
- Track future demand shifts based on power and transmission trends.

⚠️ Limitations and Future Work

Limitations:

- No geographical data to analyze regional trends
- Torque values were too inconsistent to clean
- No timestamps to track price changes over time

Future Work:

- Add `location`, `listing date`, and `service history` to enrich analysis
- Build a predictive model for price estimation
- Add demand-side behavior data (e.g., buyer clicks or inquiries)

📓 References

- Dataset: Internal (based on public listings like CarDekho)
- Tools Used: Python (Pandas, NumPy, Matplotlib), Tableau Public
- Inspiration: Real-world used car marketplaces and analytics practices

---

💻 Code Documentation

- The code is structured in Jupyter Notebooks, step-by-step:
  - `data_cleaning.ipynb` → for cleaning, conversions, and feature extraction
- Key libraries: `pandas`, `numpy`, `matplotlib`, `seaborn`
- Run requirements: Python 3.9+, Pandas 1.3+, Tableau Desktop or Public

---

Materials & Resources

Interactive Python Notebook:
https://colab.research.google.com/drive/1pkZTtMhEzhZ3_XBIuHilCx1B7rW5RAoW?usp=sharing

GitHub Repository: https://github.com/Rudrajit12/Used-Car-Listings-Analysis

Tableau Dashboard:
https://public.tableau.com/views/CarDekhoListingsDashboard/CarFeatureInsights?:language=en-

US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Credits

Author: Rudrajit Bhattacharyya

Email: mailto:rudrajitb24@gmail.com

LinkedIn: https://www.linkedin.com/in/rudrajitb/

GitHub: https://github.com/Rudrajit12