Car Resale Value: Prediction & Key Influencing Factors

Rosa A. Estrada

June – July 2025

Objective

This project explores whether the resale value of a car can be accurately predicted using features like brand, age, mileage, fuel type, and more, while uncovering which attributes contribute most to price depreciation.

Questions Answered:

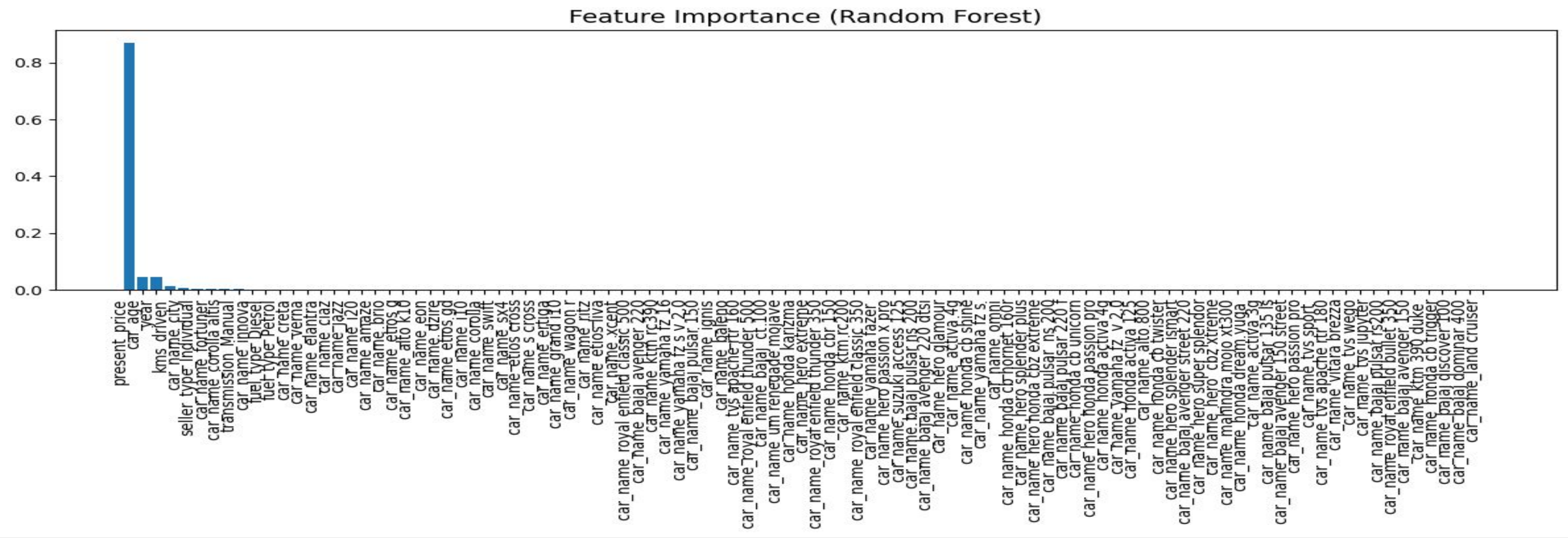
- Can we predict the optimal selling price of a car?
- What factors influence resale value?
- Which brands hold their value best?
- Do automatic cars depreciate faster than manual ones?
- Does high mileage equal lower price?

Dataset/Tools Description

Source	• Kaggle – Vehicle Dataset • From 1992 – 2020
(Observations, Features)	• (301, 9)
Features	• Year • Car Age • Care Name • Selling Price • Present Price • KMS Driven • Fuel Type • Seller Type • Transmission • Price Difference
Tools/Libraries	• Python = 3.12.3 • Pandas = 2.2.2 • NumPy = 1.26.4 • Matplotlib = 3.9.2 • Seaborn = 0.13.2 • Scikit=learn = 1.5.1 • XGBoost = 3.0.2

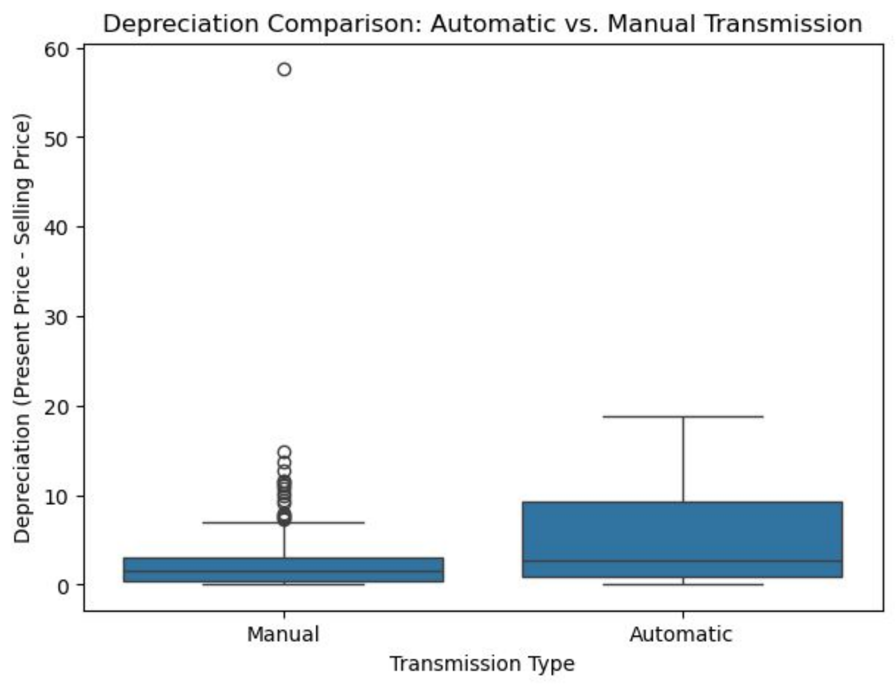
Key Insights:

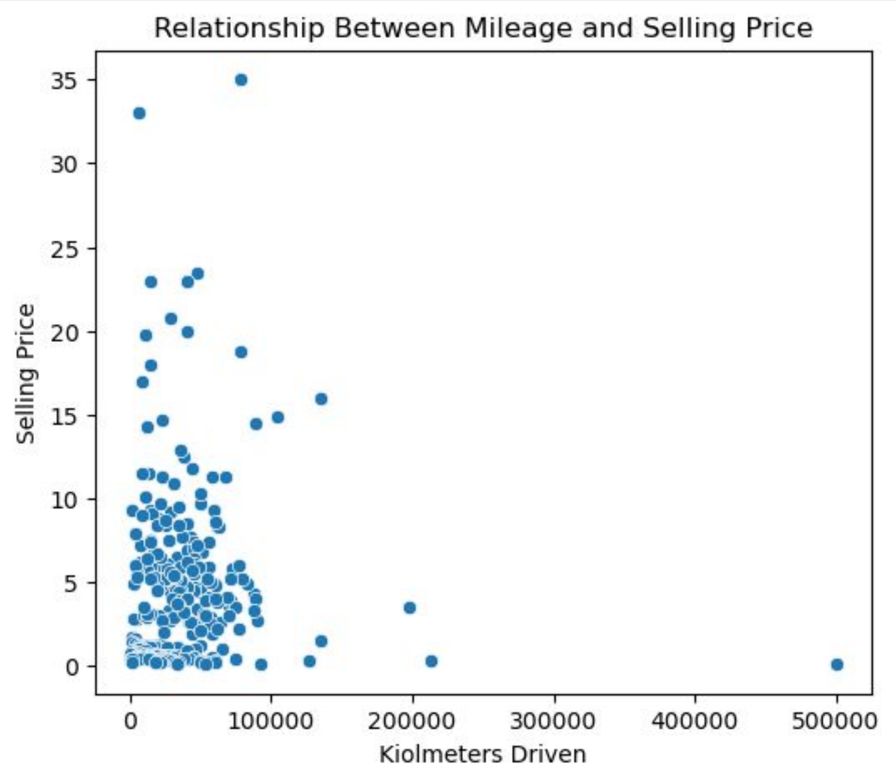
Present price, car age, year, and kms driven are the strongest predictors of resale value



Key Insights:

Automatic cars depreciate faster than manual ones, losing 5.50 more value on average





Key Insights:

High mileage has little predictive power on resale price in this dataset (Correlation ≈ 0.03)

Key Insights:

Random Forest outperformed Linear Regression and XGBoost, achieving the highest accuracy ($R^2 = 0.72$)

Car Name	Present Price	Selling Price	Value Retained %
Vitara brezza	9.83	9.250	94.099695
Bajaj avenger 150	0.80	0.750	93.750000
Um renegade mojave	1.82	1.700	93.406593
Tvs sport	0.52	0.480	92.307692
Yamaha fz 16	0.82	0.750	91.463415
Honda activa 4g	0.51	0.465	91.176471
Hero passion x pro	0.55	0.500	90.909091
Bajaj dominar 400	1.60	1.450	90.625000
Royal enfield bullet 350	1.17	1.050	89.743590
Honda dream yuga	0.54	0.480	88.888889

Key Insights:

Brands like VItara Brezza and Bajaj Avenger retained over 90% of their original price, signaling strong demand and low depreciation

Tools + Skills Demonstrated

Data Preprocessing & Cleaning

- Standardized column names
- Converted data types
- Removed missing values
- Verified duplicates to ensure data integrity

Outlier Detection & Handling

Used the Interquartile Range (IQR) method to identify and drop outliers in:

- Selling Price
- Present Price
- KMS Driven
- And Ownership fields

Feature Engineering

Created new features like:

- Car_age
- Price difference
- Depreciation

to enhance model input and analytical depth

Exploratory Data Analysis (EDA)

Visualized car counts, fuel types, transmission types, depreciation patterns, and explored relationships

- e.g., mileage vs. selling price

Model Building & Evaluation

Trained and compared Linear Regression, Random Forest, and XGBoost models.

- Random Forest performed best (R^2 = 0.72, MAE ≈ $1,100)

GitHub + Kaggle links

- github.com/rosaaestrada/Car-Resale-Value-Prediction
- kaggle.com/code/rosaaestrada/car-resale-value-prediction

Connect with me!

- linkedin.com/in/rosa-a-estrada-ms/
- rae.estrada03@gmail.com