

Car Price Prediction Using Machine Learning

Presented By -
Daksh Dureja



Problem Statement

TITLE: WHY PREDICTING CAR PRICES MATTERS

In today's fast-paced automobile market, the resale value of used cars varies wildly—even between similar models. This inconsistency often leaves buyers overpaying and sellers underselling due to a lack of clarity and guidance.

Observing this gap, we asked ourselves: Can data help bring transparency to car pricing?

→ OBJECTIVE:

We decided to leverage machine learning to build a car price prediction model that considers various influencing factors like car age, fuel type, seller type, and more.

Objectives:

- Predict car resale value using data-driven models.
- Perform in-depth EDA to understand pricing trends.
- Compare algorithms to identify the most accurate predictor.

Data Preprocessing, Curation, and Exploratory Data Analysis (EDA)

DATASET OVERVIEW:

- Size: 10,000 rows × 8 columns
- Columns: Car_Name, Year, Selling_Price, Present_Price, Kms_Driven, Fuel_Type, Seller_Type, Transmission
- Dataset from cardekho.com

DATA PREPROCESSING STEPS:

- Removed duplicates and checked for null values (none found)
- Engineered feature Car_Age = 2025 - Year for model relevance
- Dropped Car_Name as it provided no predictive power
- Applied Label Encoding to:
 - Fuel_Type, Seller_Type, and Transmission

EXPLORATORY DATA ANALYSIS (EDA):

- Selling_Price distribution → automatic cars have high price as compare to manual ones
- Car Age vs. Selling Price:
 - Clear negative correlation – older cars depreciate significantly
- Fuel Type Impact:
 - Diesel cars generally have a broader price distribution compared to petrol cars



Slide 4: Literature Review & Methodology

LITERATURE REVIEW-

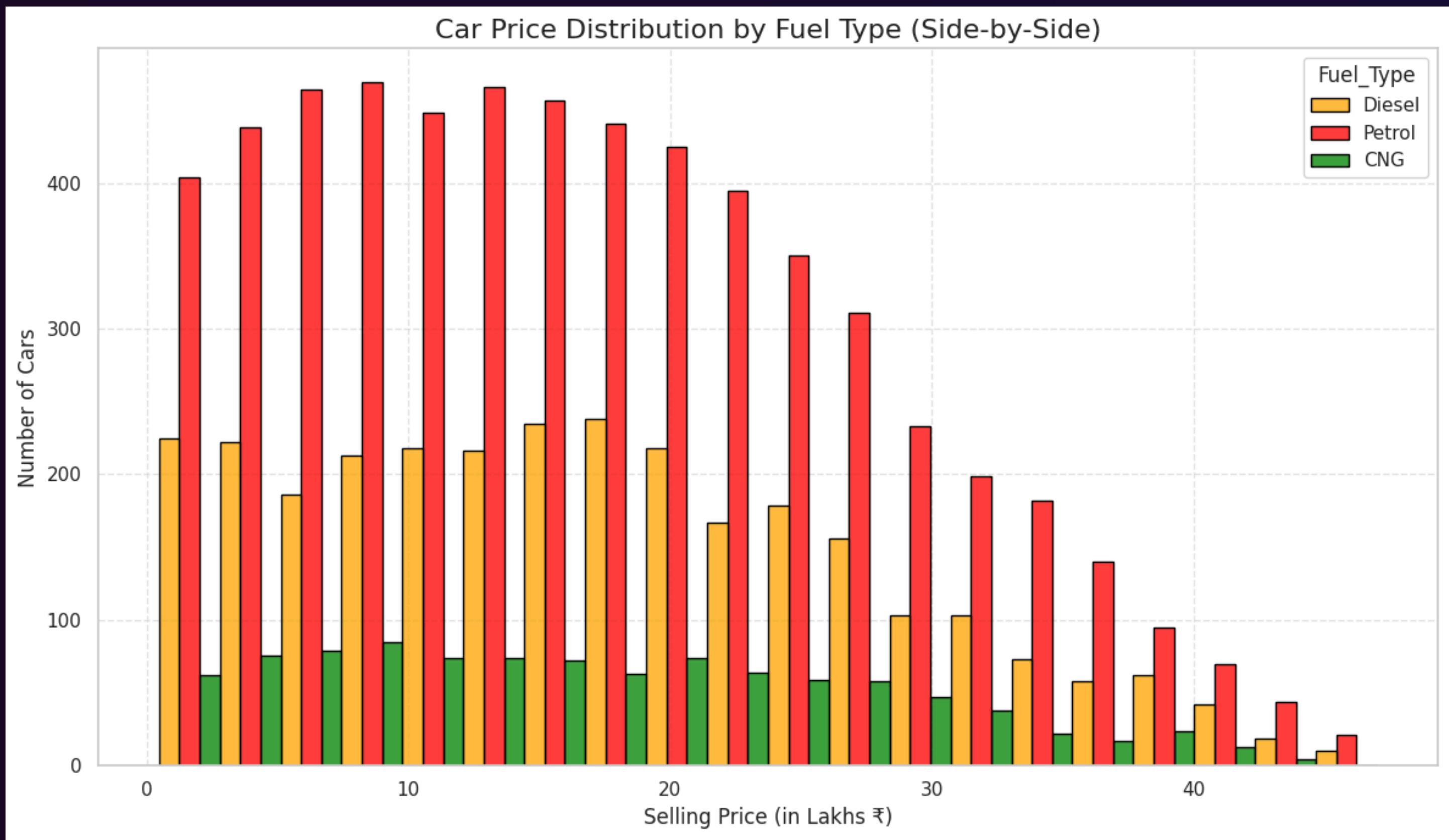
- ML models like RF, XGBoost outperform linear models in price prediction
- GBR & XGB give better accuracy in real-world data (Chenguang Li, 2024)
- Feature selection and data cleaning are crucial (Zhou et al., 2019)

OUR METHODOLOGY-

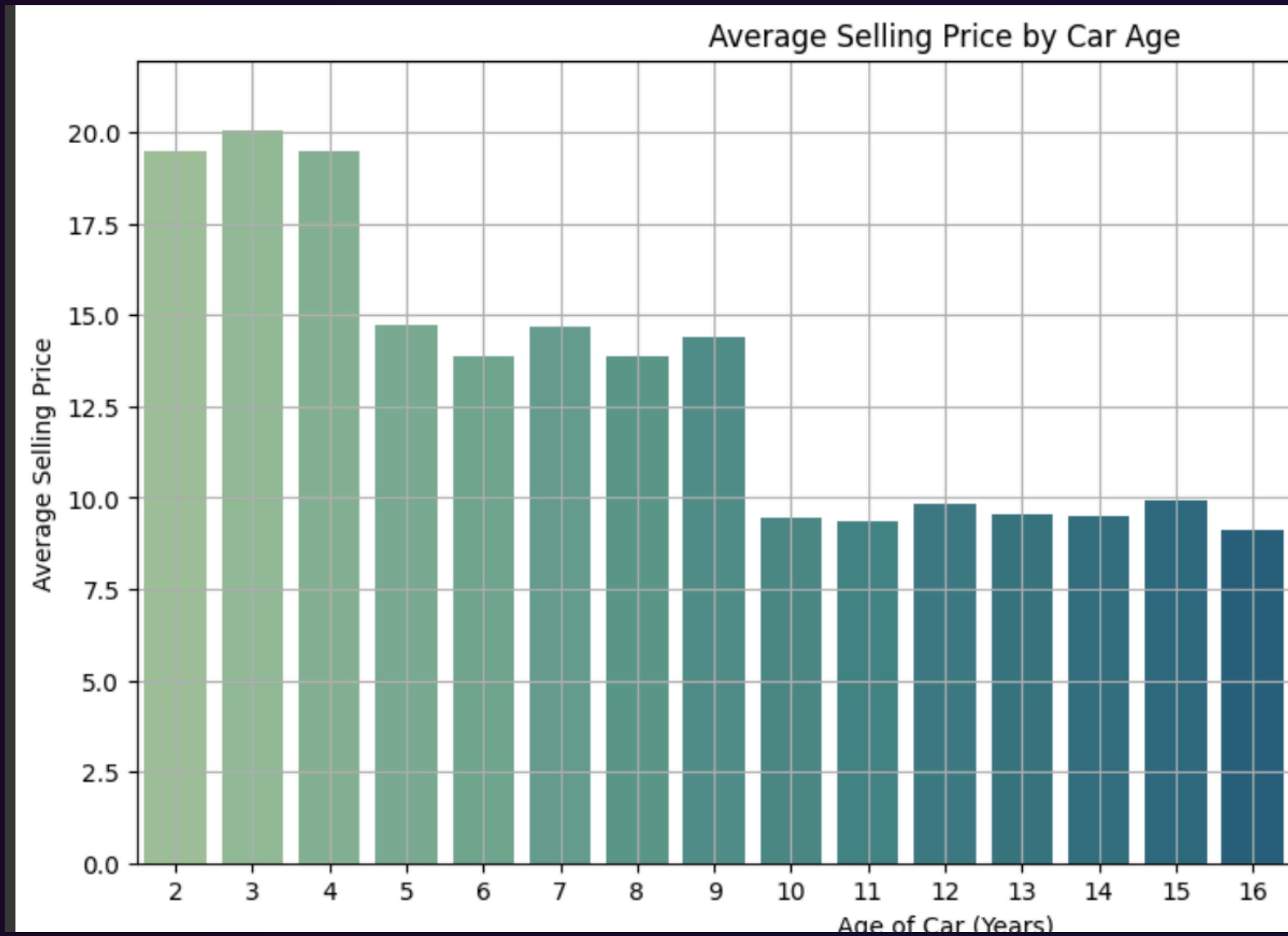
- 1.Cleaned & Preprocessed Data
- 2.Created Car_Age, encoded categories
- 3.Tested 4 Models – GBR gave best results
- 4.Saved Final Model using joblib



It shows manual cars have high selling range as compare to automatic transmission cars and most of the cars in Price bracket of (0 - 20 lakhs)



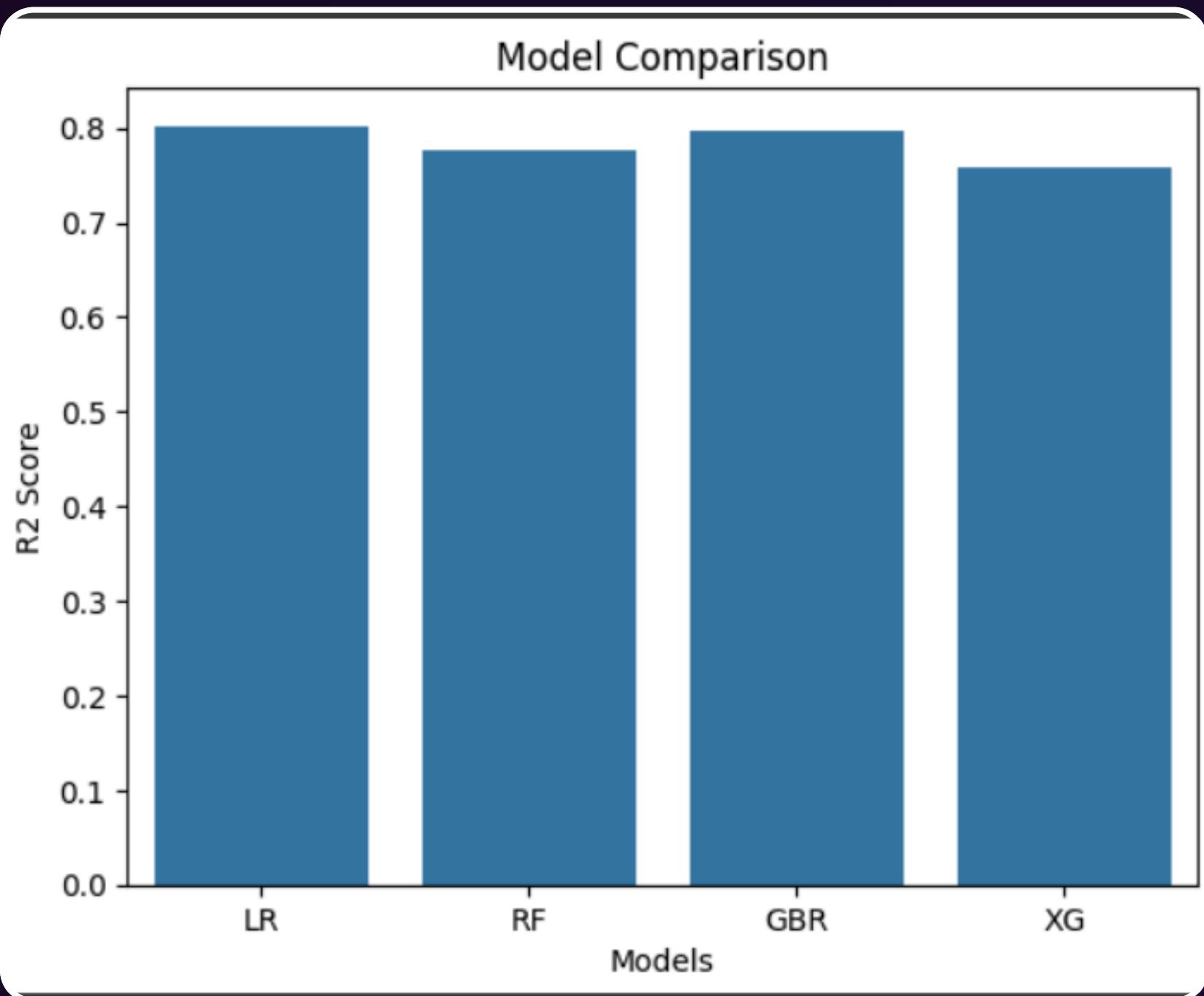
- Petrol cars dominate the used car market and have a wide price range.
- CNG cars are least common and priced lower.
- Most cars, regardless of fuel type, fall under the ₹10-25 price range.



- There is a clear inverse relationship between a car's age and its selling price.
- The trend aligns with real-world depreciation: as cars age, they typically lose market value.

Results-

Model Chosen: Gradient Boosting Regressor



| Model | R-squared (Train) | R-squared (Test) |
|-------|-------------------|------------------|
| 0 LR | 0.813036 | 0.802186 |
| 1 RF | 0.971168 | 0.779132 |
| 2 GBR | 0.826847 | 0.797012 |
| 3 XG | 0.936276 | 0.758879 |

we used gradient boosting regression as Random forest and XG boost is giving case of overfitting and according to this situation it is best it can also handle non linear value

Conclusion & future scope

Conclusion:

- The model successfully predicts car selling prices based on factors like age, mileage, fuel type, transmission, and owner type.
- Analysis shows that car price decreases as the car gets older, with sharper depreciation in early years.
- The model can help sellers set fair prices and buyers identify good deals in the used car market.
- Achieved good accuracy with reasonable error rates on test data.

Future Scope:

- we can also predict prices of ev segment by collecting ev car data over time
- Include more features like insurance status, accident history, and service records to improve accuracy.
- Deploy the model as a web app or mobile app for real-time price estimation.
- Regularly update the model with latest market data to stay relevant.

Thank You