Used Car Price Prediction

Alan T Akkara
Springboard Data Science Bootcamp
10/05/2025

Introduction

- The used car market is vast and dynamic.

- Prices vary by mileage, brand, year, fuel type, and other attributes.

- Buyers and sellers struggle with transparency in pricing.

- **Goal:** Build a predictive model to estimate car prices more accurately.

Problem Statement

- Used car buyers risk **overpaying**.

- Seller's risk **undervaluing** their vehicles.

- Businesses need **data-driven tools** for fair and competitive pricing.
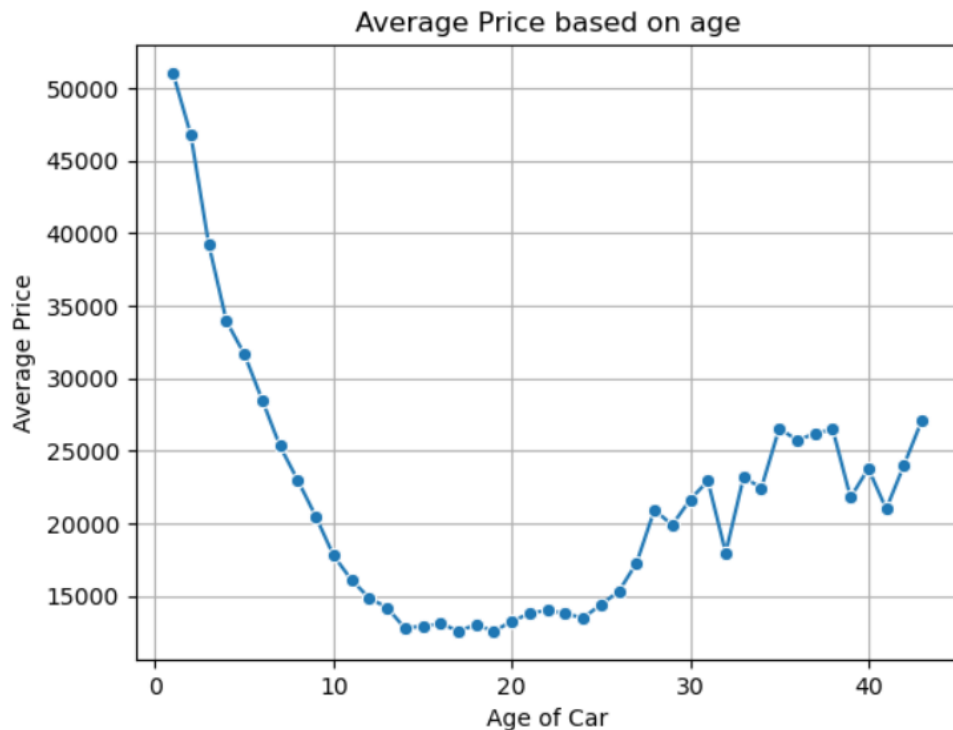
Dataset

	manufacturer	model	year	mileage	engine	transmission	drivetrain	fuel_type	mpg	exterior_color	interior_color	accidents_or_damage	one_owner	personal_us
0	Acura	ILX Hybrid 1.5L	2013	92945.0	1.5L I-4 i-VTEC variable valve control, engine...	Automatic	Front-wheel Drive	Gasoline	39-38	Black	Parchment	0.0	0.0	
1	Acura	ILX Hybrid 1.5L	2013	47645.0	1.5L I4 8V MPFI SOHC Hybrid	Automatic CVT	Front-wheel Drive	Hybrid	39-38	Gray	Ebony	1.0	1.0	
2	Acura	ILX Hybrid 1.5L	2013	53422.0	1.5L I4 8V MPFI SOHC Hybrid	Automatic CVT	Front-wheel Drive	Hybrid	39-38	Bellanova White Pearl	Ebony	0.0	1.0	
3	Acura	ILX Hybrid 1.5L	2013	117598.0	1.5L I4 8V MPFI SOHC Hybrid	Automatic CVT	Front-wheel Drive	Hybrid	39-38	Polished Metal Metallic	NaN	0.0	1.0	
4	Acura	ILX Hybrid 1.5L	2013	114865.0	1.5L I4 8V MPFI SOHC Hybrid	Automatic CVT	Front-wheel Drive	Hybrid	39-38	NaN	Ebony	1.0	0.0	

Source : https://www.kaggle.com/datasets/andreinovikov/used-cars-dataset
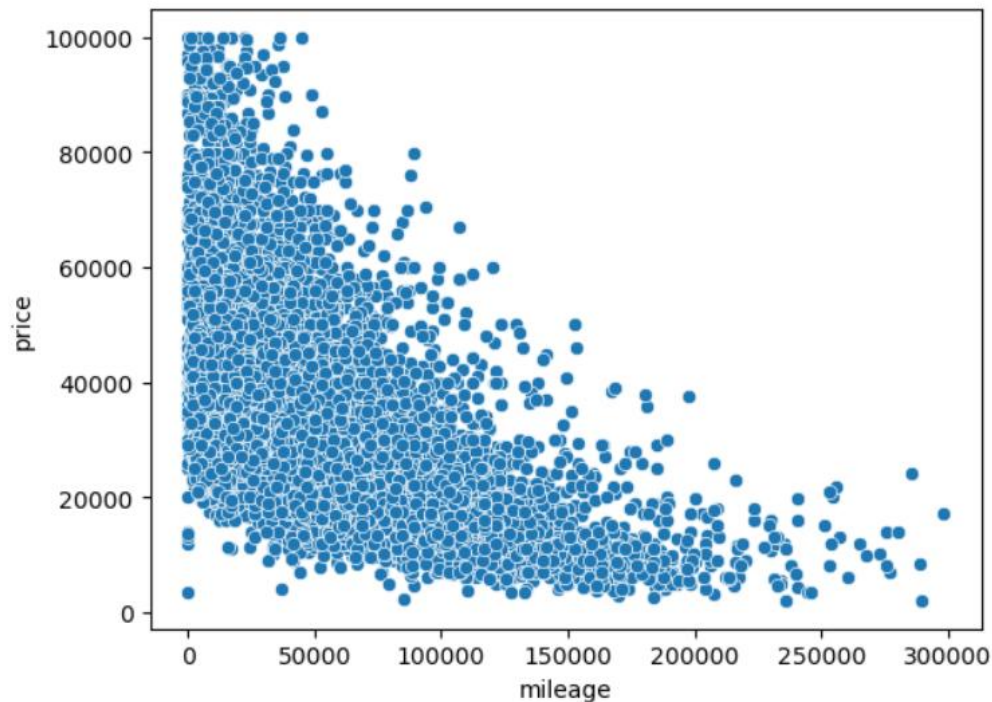
Data Wrangling

- Removed duplicate rows to ensure data integrity.

- Limited extreme values in **year**, **mileage**, and **price**.

- Restricted categorical features to top 10 values.

- Converted **mpg** ranges into their mean values for consistency.

- After preprocessing: **690,370 rows × 17 columns**

Exploratory Data Analysis

Average Price based on age

- Price decreases as car age increases.

- Older classic cars show a slight price rise
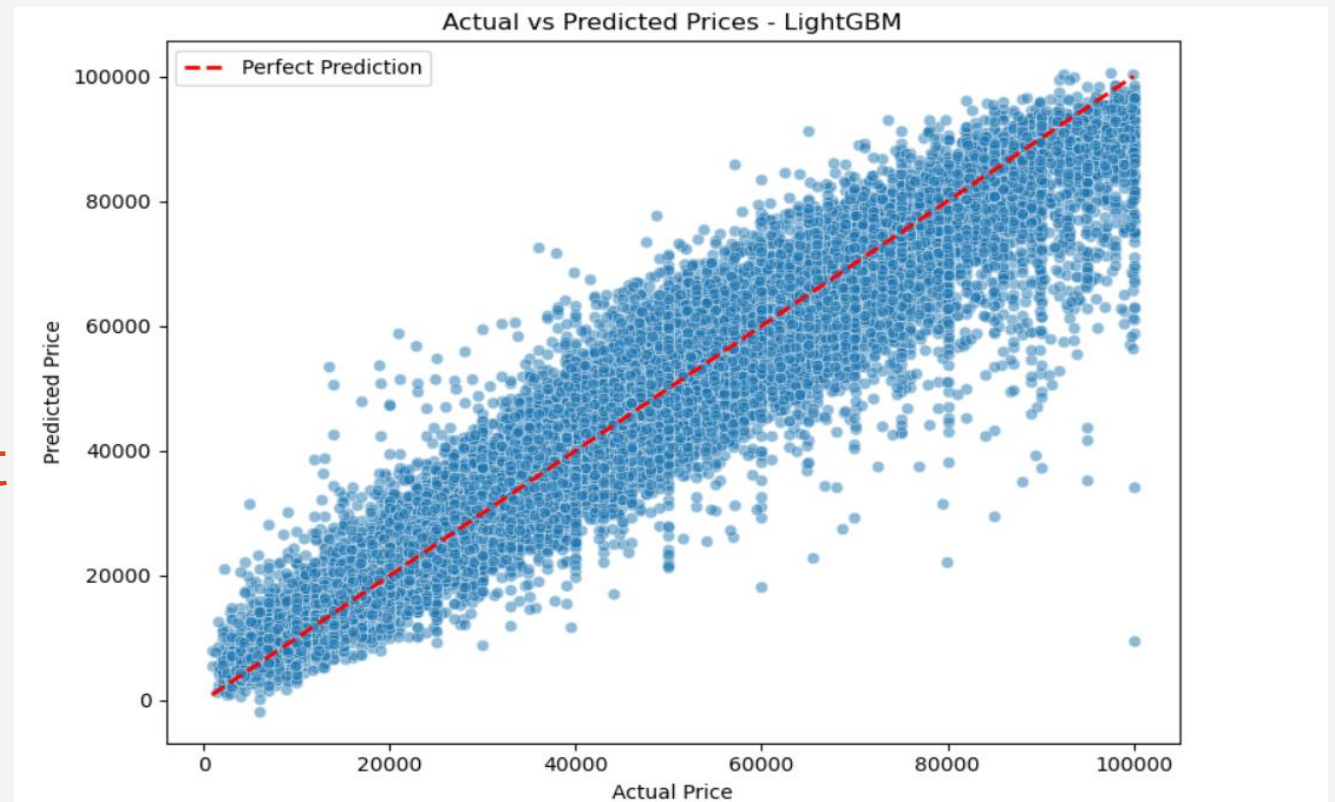
Exploratory Data Analysis

- Cars with higher mileage generally have lower prices

Modelling Approach

- Models tested:

    - Linear Regression

    - Random Forest

    - LightGBM

    - XGBoost

- Evaluation metric: **RMSE & R²**.
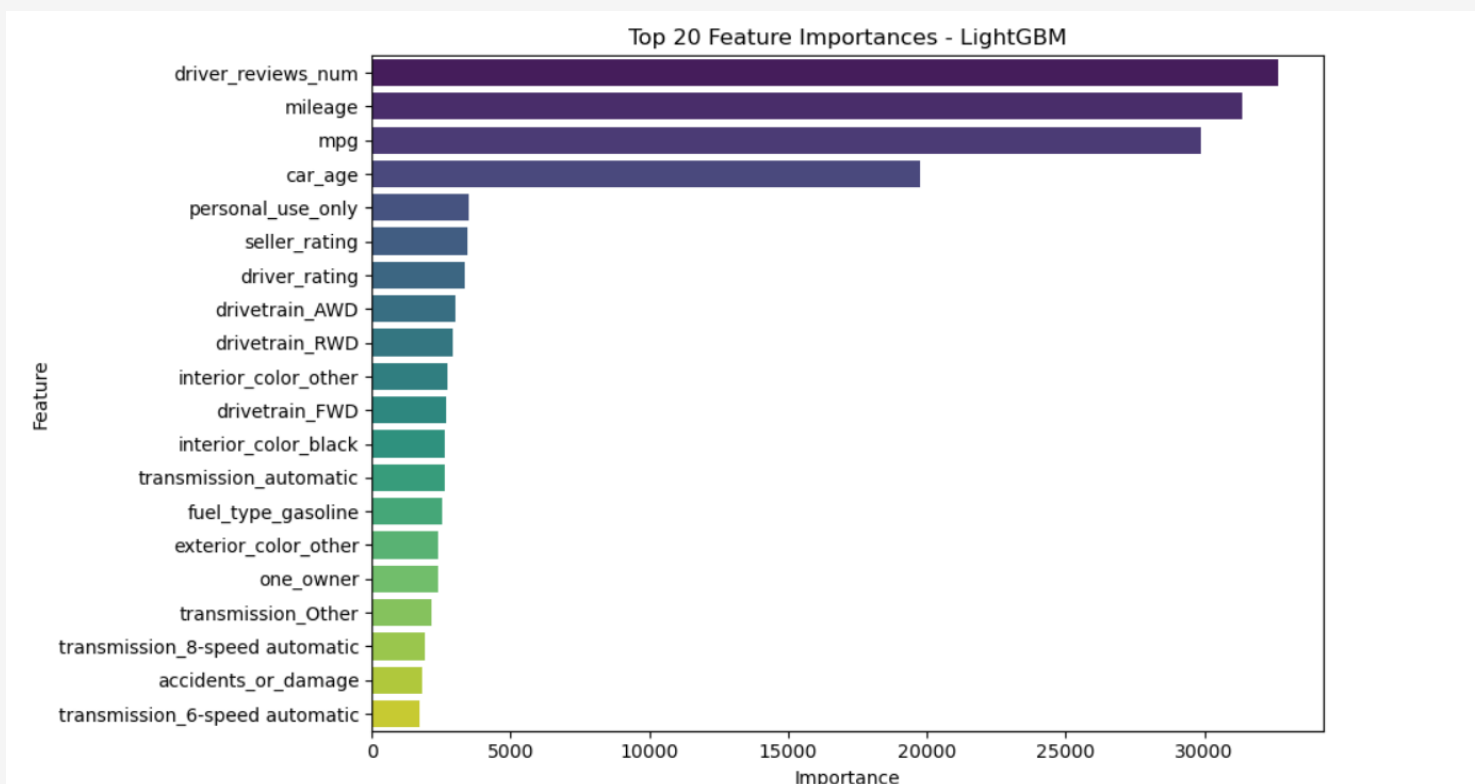
- Final model: **LightGBM** due to best performance.

Model Performance

- LightGBM Results:

  - RMSE: 3857.48

  - R²: 0.94

- Strong predictive power on test set.

Actual vs Predicted Prices - LightGBM

Feature Importance

- Top features :

  - Number of Driver Reviews

  - Mileage

  - Miles per Gallon

  - Age of the Car

Top 20 Feature Importances - LightGBM

Recommendations

- Businesses can use model outputs to:

  - **Set competitive prices** for listings.

  - **Provide transparency** to customers.

  - **Improve trust** in used car platforms.

- Buyers and sellers benefit from **fair pricing benchmarks**.

Future Research

- Develop a real-time car price prediction app.

- Add **features** like accident history and service records for better accuracy

- Analyze **time trends** to predict future price depreciation

Conclusion

- Built and evaluated multiple ML models.

- LightGBM achieved **94% accuracy**.

- Strong negative correlation between **car age**, **mileage**, and **price**

- The model can **enhance decision-making** in used car pricing.

THANK YOU