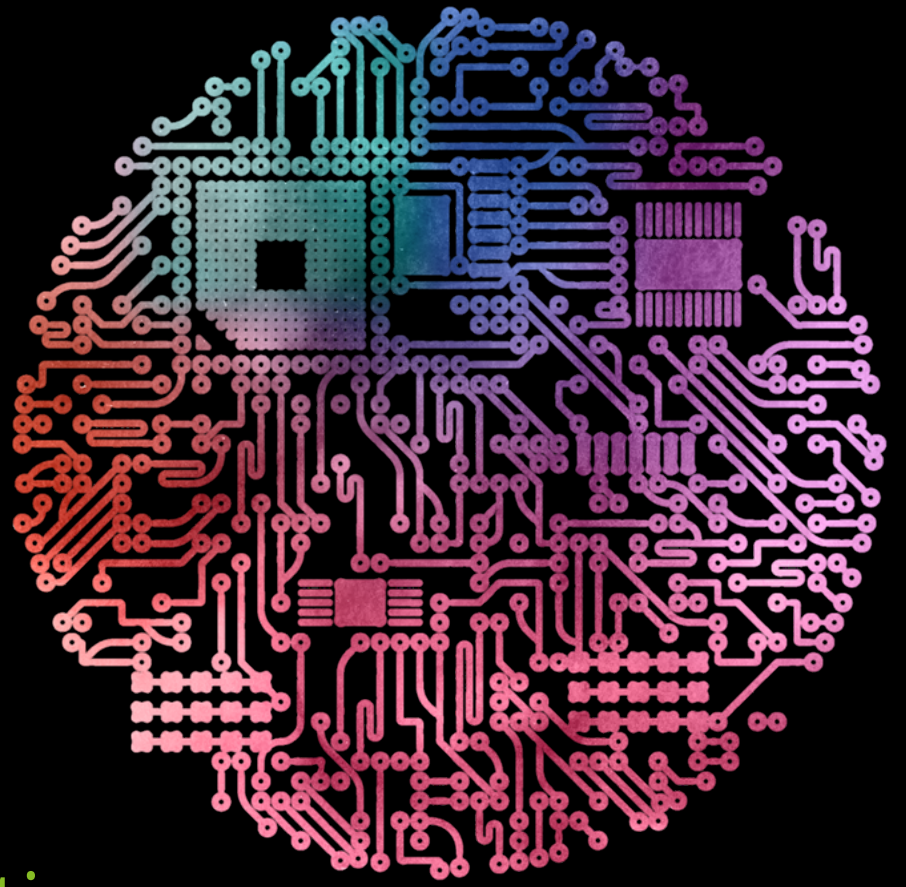
Deloitte.



Empower Automotive Used Car Price Analysis

August 2023

Group 3 Capstone

MAKING AN IMPACT THAT MATTERS since 1845

Team Members

Hannah LoChiatto
Consulting Analyst

Sasha Shahidi
Advisory Analyst

Cristian González Ramírez
ARDC Consulting Analyst

Oscar Muñoz Sanchez
AI & Data Analyst

Jordan Mai
Advisory Analyst

Contents

Introduction

- Project Overview / Business Understanding
- What We Heard From Stakeholders

Data Understanding

- Data Understanding
- Model Selection Reasoning

Data Analysis

- Data Analysis & Visualizations – Linear Regression
- Data Analysis & Visualizations – Random Forest
- Data Analysis & Visualizations – XGBoost with LASSO

Conclusion

- Our Recommendation
- Next Steps

Thank you! Questions?

Introduction

Project Overview / Business Understanding

Empower Automotive is an app that users can use to help set a price on their used car. They are trying to become a major player in the automotive industry.

Key Business Questions

What features are the most important when accurately pricing a car?

How can Empower Automotive become a trusted source for used car prices?

Stakeholders

Empower Automotive Executives

Car Sellers & Buyers

Our goal is to provide a recommendation based on data analysis of linear regression, random forest, and XGBoost with LASSO to answer these business questions while considering stakeholder input.

What We Heard From Stakeholders:

"I want to make sure I am getting a fair and accurate price for a used car based on what other similar used cars are selling for."
- Buyers

"My car is in great shape but is repeatedly getting appraised far lower than what I expected. I want to understand why and what features are important to customers."
- Sellers

"Empower Automotive is new to the Automotive Industry and have relevant data but are counting on you to extract what is absolutely necessary to find the most valuable features when pricing a car so that sellers and buyers can conduct business with confidence."
– Empower Automotive Executives

Data Understanding

Data Understanding

The data source we are using for analysis and to make our recommendation

Main Source of Data

- ***Car Details Dataset***

We used pre-processing and cleaned the data to remove outliers and ensure we were pulling the most relevant features for our models and predictions.

Kaggle Car Price Prediction Dataset

	Price	Levy	Manufacturer	Prod. year	Category	Leather interior	Fuel type	Engine volume	Aspiration	Mileage	Cylinders	Gear box type	Drive wheels	Doors	Wheel	Color	Airbags
0	13328	1399.0	LEXUS	2010	Jeep	Yes	Hybrid	3.5	Natural	186005.0	6	Automatic	4x4	4	Left wheel	Silver	More than 10
1	16621	1018.0	CHEVROLET	2011	Jeep	No	Petrol	3.0	Natural	192000.0	6	Tiptronic	4x4	4	Left wheel	Black	5-9
2	8467	0.0	HONDA	2006	Hatchback	No	Petrol	1.3	Natural	200000.0	4	Variator	Front	4	Right-hand drive	Black	0-4
3	3607	862.0	FORD	2011	Jeep	Yes	Hybrid	2.5	Natural	168966.0	4	Automatic	4x4	4	Left wheel	White	0-4
4	11726	446.0	HONDA	2014	Hatchback	Yes	Petrol	1.3	Natural	91901.0	4	Automatic	Front	4	Left wheel	Silver	0-4

Model Selection Explanation

The regression models we selected are linear regression, random forest, and XGBoost with LASSO

Explanations:

1. ***Linear Regression***

   - Base will always be linear

   - There was no significant difference between the base and linear

   Continued to follow up with more complex models that are more common and suitable for regression problems:

2. ***Random Forest***

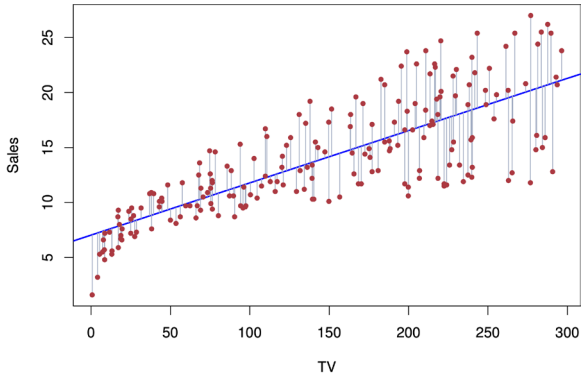
   - One of the most common models for regression

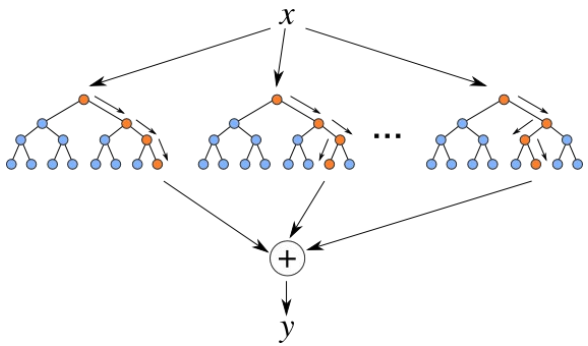
   - Simplicity and high accuracy

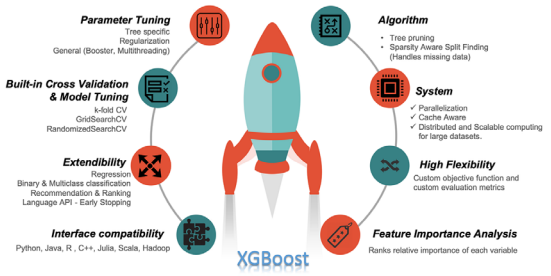
3. ***XGBoost with LASSO***

   - Execution speed and model performance

   - Simplicity and high accuracy







Data Analysis

Data Analysis & Visualizations – Linear Regression

What feature is the most important when accurately pricing a car?

Most correlated feature to price is Production year, with a correlation value of 0.35

RMSE – Root Mean Squared Error

- Measures average difference between values predicted by the model and the actual values
- Lower values = better fit

***RMSE* = $14,038.19**

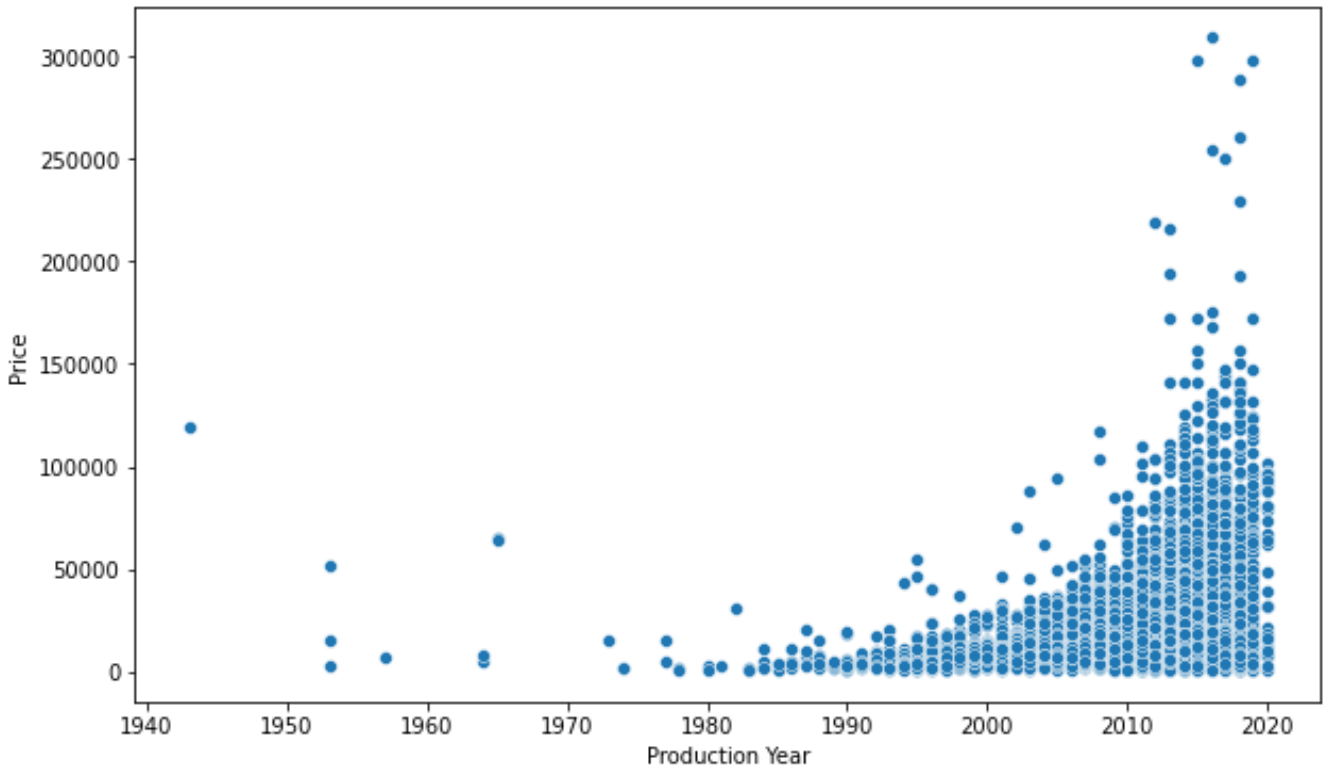
- Estimation is not very accurate
- Large error

R^2 = 0.4143

- Around 41% of the variation of price is explained by linear model

Key Takeaways

- Overall, the graph demonstrates that the newer the car, the higher its price



Data Analysis & Visualizations – Random Forest

Production year and Engine volume, best predictors of price

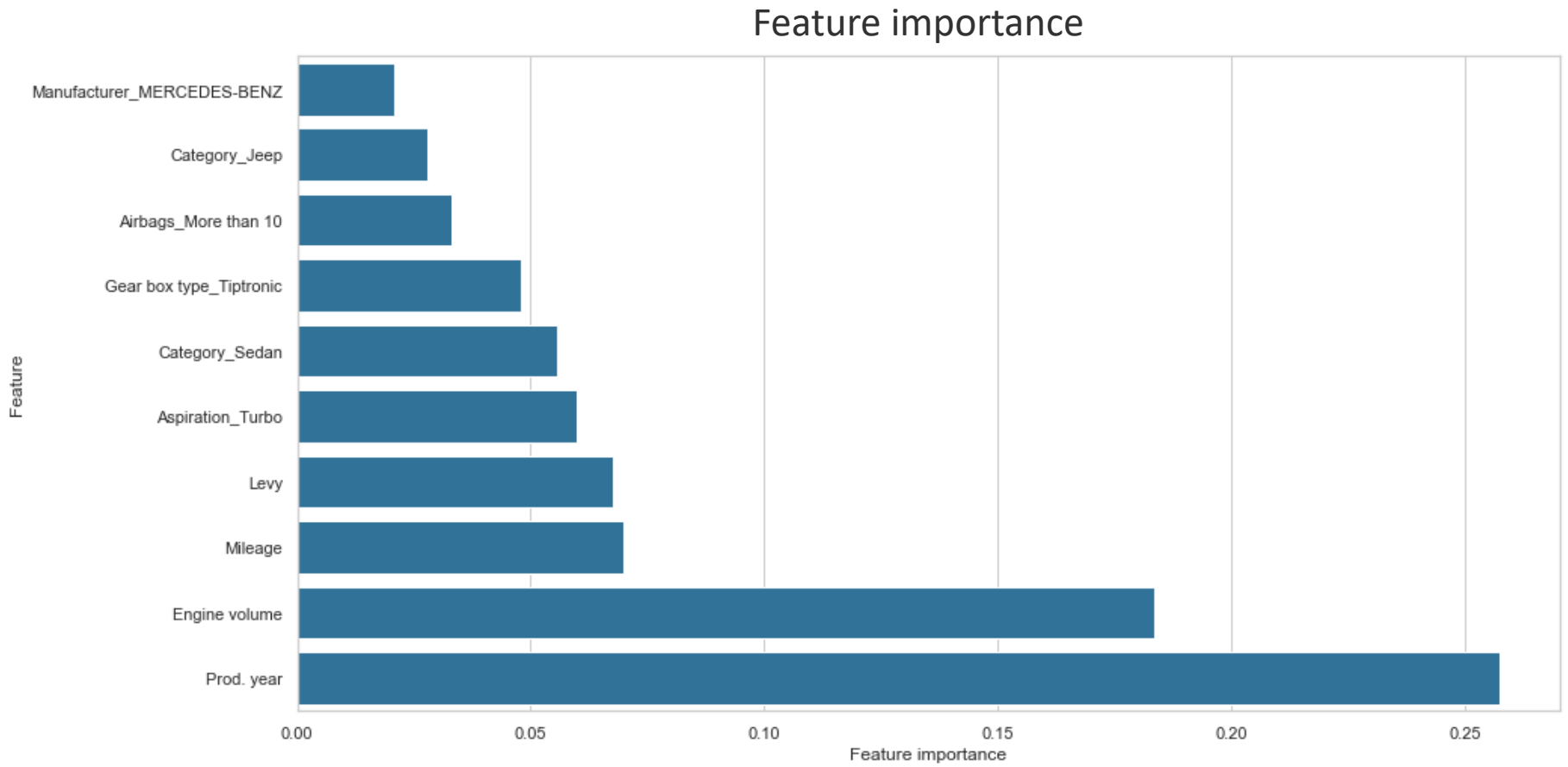
***R^2* = 0.7360**

- Almost 74% of the variation in the price is explained by this model

RMSE* = *$9,423.59

Key Takeaways

- Overall, the graph demonstrates newer models = higher price
- Engine volume and Mileage are also important influencers on the price for this model



Data Analysis & Visualizations – XGBoost with LASSO

Type of aspiration, the predominant feature

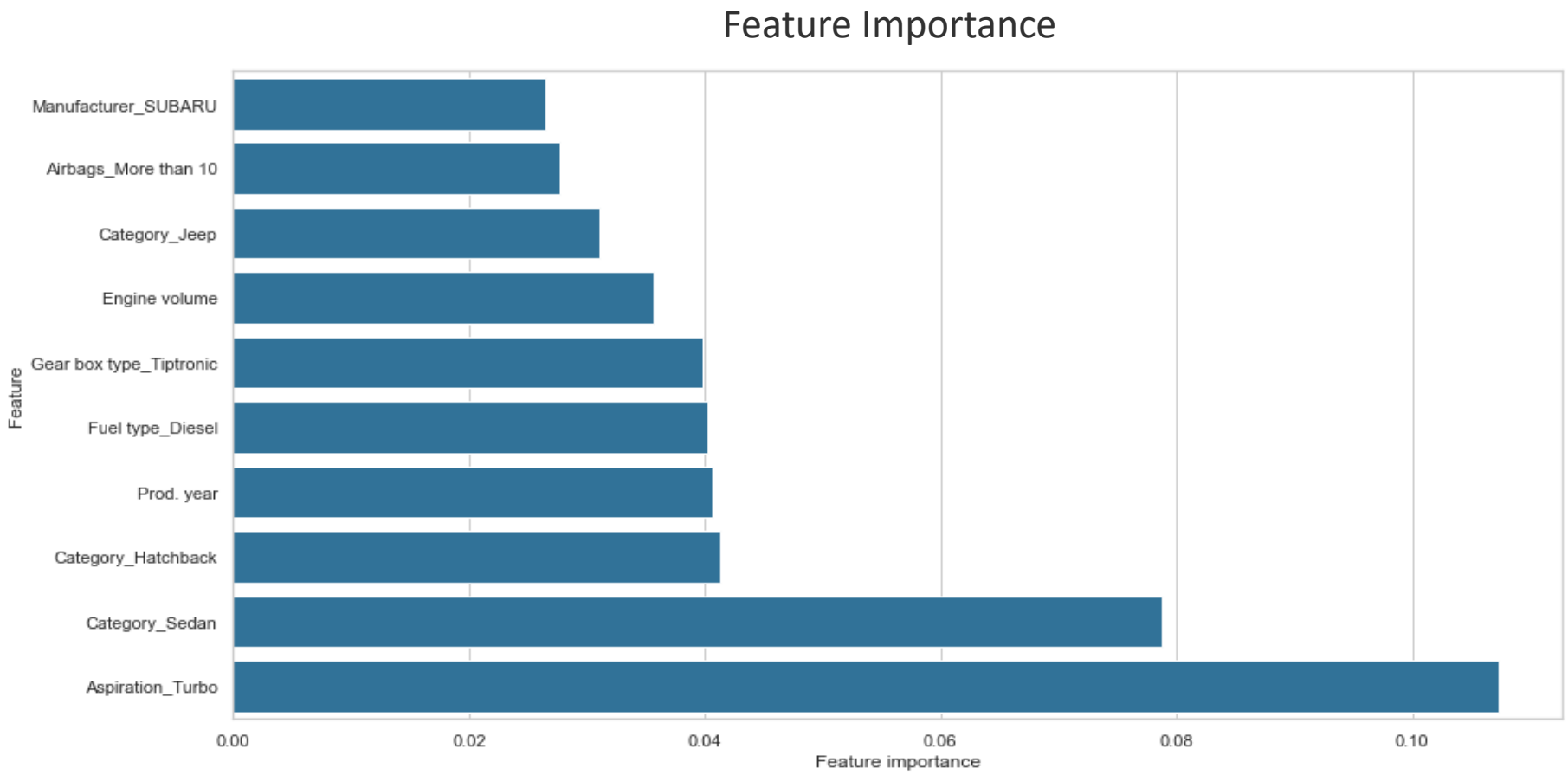
R^2 = 0.7727

- 77% of price variation is explained by our model

RMSE = $8,743.25

Key Takeaways

- Better performance than all previous models
- Different features account the most for this model



Conclusion

Our Recommendation

Based on our findings and comparing the data collected, we recommend these action plans for Empower Automotive

Key Business Questions	Solutions
What features are the most important when accurately pricing a car?	The most important features for determining car prices are Year of Production, Engine Volume and Aspiration Type.
How can Empower Automotive become a trusted source for used car prices?	Empower can become a trusted source by using the most up-to-date car data and continuing to add features that are useful for sellers and buyers, especially features related with the condition of the car.

While comparing linear regression, random forest, and XGBoost with LASSO, our recommendation is for Empower Automotive to use XGBoost with LASSO. This model proved to be the most accurate, with the smallest RMSE and a strong R^2.

Next Steps

To further improve our data analysis and model output, our next steps should be to:

Tune current model to include more correlated features to price and condition of the car

Implement more complex models:

Neural Network

Thank You!

Questions?