Understanding Factors Influencing Used Car Prices

Andre Londono

11/20/24

Jupyter Notebook Link

The full analysis, including code and detailed outputs, can be accessed in the Jupyter Notebook available on GitHub: Price of Cars Jupyter Notebook.

Business Objective

The goal of this project is to identify the key factors that influence used car prices. By understanding these factors, I aim to help a used car dealership make informed decisions on pricing, inventory management, and marketing strategies, ultimately improving profitability and market competitiveness.

The CRISP-DM Framework

This analysis followed the **CRISP-DM (Cross-Industry Standard Process for Data Mining)** framework, ensuring a structured and systematic approach to solving the problem. The six phases of CRISP-DM were applied as follows:

1. Business Understanding

The project began with a clear business goal—to help the dealership optimize pricing and inventory acquisition decisions based on data backed insights into what drives used car prices. I focused on delivering actionable insights that could directly inform business strategies such as focusing on fuel efficient cars in great condition with low miles.

2. Data Understanding

I started by performing analyses on the data to understand key variables including car manufacturer, model, year, odometer mileage, type of fuel, condition, and drivetrain. I also plotted the data to get a sense of the overall characteristics of the data. **Data Cleaning**: I handled missing values, removed unrealistic

data (e.g., zero or single-digit prices and mileage), and standardized numerical values to ensure data integrity.

3. Data Preparation

The dataset was cleaned and prepared using techniques such as:

- Removing invalid records (e.g., unrealistic prices, odometer readings).
- Handling missing data by removing incomplete records.
- Dropping irrelevant columns like VIN, ID, and region, which didn't affect the price prediction.
- Selecting features with a known impact on pricing, such as year, condition, and mileage.

4. Modeling

Once the data was ready, I applied multiple machine learning models, including **Linear Regression**, **Ridge**, **Lasso**, and **Random Forest**, to predict car prices.

- **Feature Selection**: I used correlation analysis and feature importance to identify the most relevant factors for the model.
- **Model Evaluation**: The models were evaluated based on metrics like Root Mean Squared Error (RMSE) and R-squared to determine the best-performing model. The **Random Forest** model emerged as the most accurate, providing the best predictions.

5. Evaluation

The results showed that key factors influencing car prices include year, mileage, fuel type, drive train, size, model, and condition. By analyzing these factors, we were able to draw actionable insights that could guide pricing and inventory decisions.

- **Model Performance**: The **Random Forest** model had the best accuracy, with an RMSE of 6321.15, and an R-squared of 0.78, significantly outperforming the other models.

6. Deployment

- **Actionable Recommendations**: Based on the model findings, I recommended that the dealership focus on cars that are relatively new, in good condition, and have lower mileage. Additionally, marketing efforts should highlight key features like low mileage, newer model year, fuel efficiency, manufacturer and drive train to justify higher prices.

Insights and Recommendations

Key Findings

- **Age & Mileage**: Older cars with higher mileage tend to be less expensive.
- **Brand & Model**: Certain brands, such as Ferrari, retain their value better over time.
- **Condition & Features**: Cars in better condition command higher prices.

Model Performance

The **Random Forest** model outperformed others with the lowest RMSE and highest R-squared, indicating its strong ability to predict car prices accurately.

Feature Importance

Key factors affecting car prices include:

- **Year** and **Odometer** (mileage)
- **Fuel Type** and **Drive Type** (e.g., AWD, RWD)
- **Size** and **Condition**

Actionable Recommendations

- **Inventory Management**: Focus on sourcing cars that are newer, in good condition, and have lower mileage.
- **Pricing Strategy**: Adjust prices based on the car's condition and features, and highlight premium options in marketing materials.
- **Marketing Strategy**: Emphasize the unique selling points of each car (e.g., low mileage, advanced features) to attract customers.

Next Steps

- **Further Analysis**: Conduct a deeper dive into specific car categories (e.g., luxury cars, SUVs) to refine recommendations.
- **Customer Feedback**: Gather customer insights to better understand preferences and refine the model.
- **Model Updates**: Continuously update the model with fresh data to ensure it adapts to changing market conditions.

Conclusion

Using the **CRISP-DM** framework, I effectively identified the key factors that influence used car prices and provided actionable insights for the dealership. By focusing on the most valuable cars (in terms of year, condition, and features), the dealership can optimize pricing, improve inventory decisions, and better meet market demand.