DS740 Midterm Project Executive Summary

Motor Vehicle Retail Price Prediction

In 2019, 284.5 million vehicles were registered in the United States. That number was expected to increase to 287.3 million in 2020 (Stasha, 2021). The automobile industry possesses an enormous amount of opportunity, and all auto manufacturers should pay close attention to the following information regarding motor vehicle price predictions. Accurately identifying which characteristics drive the price of a vehicle can be pivotal in determining where to invest capital and how to target customers.

In this analysis, information about new cars and trucks from 2004 was analyzed to identify an accurate means of predicting the retail price of a vehicle based on certain characteristics. The class of the vehicle (Sport, Pickup, Wagon, Minivan, SUV, and other) were amongst the categories evaluated. In addition, engine size, the number of cylinders, horsepower, city and highway gas mileage, and general size parameters (height, weight, length, and wheelbase) were analyzed. Another parameter that was evaluated pertains to if the vehicle possesses all-wheel drive, rear-wheel drive, or some other drive function.

The first step of data cleansing was the removal of missing values. Most of the missing values were height measurements for Pickup trucks. The information was removed based on the relationship of vehicle height with the other size variables, such as weight and wheelbase. The native dataset provided height data in a characterized format, so the remaining rows were converted to numeric values for better analysis. Correlations between variable were present, which can require attention. In this case, the predictive modelling methods filtered out the risk of any collinearity.

Log transformations of highly skewed variables (Horsepower, HwyMPG, CityMPG) were considered. The transformed data reduced the overall accuracy of the model assessment, and the original, untransformed data was used in the final model selection.

Robust Regression and Elastic Net Regression were the predictive methods evaluated in this study. Elastic Net regression requires tuning parameters to drive model performance. In this case, a wide variety of alpha values (penalty parameter) and lambda values (shrinkage parameter) were used to cover all possible combinations of the two parameters. Alpha values ranged from 0.1 to 0.9 stepping by 0.1 and lambda values ranged from 1 to about 22000 stepping exponentially 100 times.

An initial, inner layer of model selection was completed and verified by a subsequent, outer layer of model assessment to determine the most accurate model. The best model, and the model recommend moving forward was the Elastic Net model with tuning parameters of α= 0.1 and λ = 90.017. The assessment of this model represents 70.5% of the variability in the response variable, which in this case is the retail price vehicles. Two of the evaluated predictors deserve specific attention. Engine Volume (Liters) and Engine Size (Number of Cylinders) were both positively correlated with the response variable. These two variables are the best predictors of the retail price of a new vehicle from the 2004 data as displayed in graphs A and B in Figure 1. Graphical representation of the relationships of two other variables are also displayed as graphs C and D in Figure 1. There does not seem to be any discernable relationship between Wheelbase or Highway gas mileage. These variables were examined based on their correlation the other size and gas mileage variables.

*Chart, scatter chart

Description automatically generatedFigure 1. (Selection of Quantitative Variables compared to Retail Price)*

The relationship between engine size and volume and the price of the vehicle is not surprising. It is commonly known that the size and volume of an engine have a direct correlation with the power of a vehicle and are usually found in larger, more expensive vehicles. For example, a full-size pickup truck with an eight-cylinder engine will cost much more than a compact, four-cylinder sedan.

There are direct and indirect impacts to price as related to engine size and volume. A direct example is that of material cost. Larger engines require more physical material to manufacture increasing the overall cost of the vehicle. Vehicles with greater power must also have other improvements, such as suspension, steering, and exhaust to support the increased engine size.

Some indirect impacts to vehicle price with respect to engine size include maintenance and fuel costs. There are appropriate applications for large and small engines. A larger engine is able to burn more fuel with each revolution it turns in a minute (rpm), and usually consumes more fuel than a smaller engine would during the same journey (Carbuyer, 2021).

Much can be done by auto manufacturers to optimize these parameters based on specific use cases as well as control price to drive bottom line profit. People generally buy larger vehicles for applications of hauling loads or increasing physical space and comfort. Optimizing engine performance and reducing cylinders and/or volume not only reduces the cost of manufacturing, but also the ownership costs for the consumer. Overall, the price of the vehicle will decrease, and more vehicles will be sold for the same purpose. Raw material optimization could also prove beneficial to overall profit. If a vehicle can be physically larger, yet the same weight as a counterpart it will likely be more profitable. This can be achieved by investigating various types of alloys and other light, yet sturdy building materials.

Based on the information above, and the performance of my model selection process (70.5%), the aforementioned Elastic Net model is sufficiently accurate for making new predictions on individual data points especially with lower cost vehicles as displayed by the tightness of the data points on the lower end of the scale in Figure 2. The model is best at predicting price when the vehicle is worth ≈$50,000 or less.

It is important to understand that the data is from 2004. There have been many technological and performance-based improvements by automobile manufacturers over the past 17 years. Predicting the retail price of vehicles build in the last three to five years is not recommended without further analysis and verification that the model can perform the same or better with newer data. There may also be variables that were not included in this analysis that may impact the performance of the model. Consideration around vehicle manufacturer and fuel type (electric, gasoline, diesel, and hybrid) should be included.

*Figure 2. (Predicted Values compared to Actual Values with trend line)*

Chart, scatter chart

Description automatically generated

REFERENCES

Carbuyer. (2021, July 26). *Engine size explained*. Carbuyer. Retrieved November 4, 2021, from https://www.carbuyer.co.uk/tips-and-advice/146778/engine-size-explained.

Stasha, S. (2021, September 15). *How many Americans own cars?: Policy advice*. PolicyAdvice. Retrieved November 4, 2021, from https://policyadvice.net/insurance/insights/how-many-americans-own-cars/.