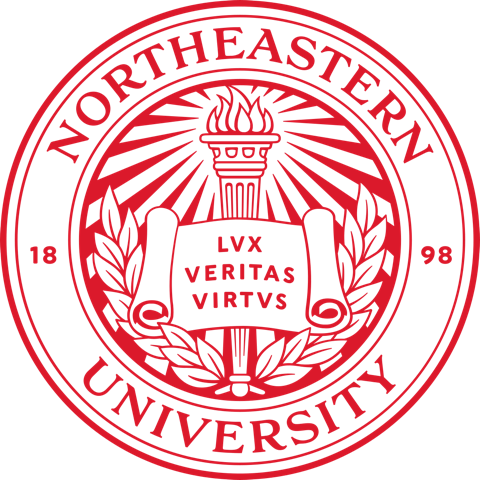
**Module 2 Midweek Project**

**Car Prices**

Sunil Raj Thota



College of Professional Studies, Northeastern University ALY 6020 – Predictive Analytics

Prof. Justin Grosz

November 09, 2021

**Introduction:**

An automobile manufacturer intends to enter the US market by establishing a production facility there and producing automobiles locally to compete with US and European competitors. They have hired an automobile consultancy firm to help them understand the elements that influence car pricing. They are particularly interested in learning about the factors that influence car pricing in the United States, as these may differ significantly from those in other markets. It wants to determine which factors are important in determining a car's pricing. Based on several market surveys, the consulting firm has amassed a big data set of different types of automobiles across the American market to see how well those variables explain the price of a car. It will be used by management to figure out how the pricing differs. They can then adjust the car's design, commercial strategy, and other factors to fulfill specified price targets. Additionally, this study can assist management in comprehending the price dynamics of a new market.

**Analysis:**

The goal of this report is to use the Car Prices dataset and predict the prices of the cars. By fitting a linear equation to observed data, linear regression seeks to model the relationship between two variables. One variable is regarded as an explanatory variable, while the other is regarded as a dependent variable. When we wish to anticipate the value of a variable based on the value of another variable, we apply this formula. The collection contains statistics about various automobiles. I'll try to anticipate the price using a Linear Regression model. I'm going to use the OLS method. This is accomplished by re-fitting the model using the supplied machine learning method, ranking features by relevance, deleting the least important features, and fitting the model again.

**Data Modelling:**

Multiple linear regression (MLR), often known as multiple regression, is a statistical technique that predicts the result of a response variable by combining numerous explanatory variables. Multiple regression is a variant of linear (OLS) regression in which just one explanatory variable is used.

1. **What were the three most significant variables?**

Car width, engine size, and rear are the three most important elements. The significant variable's coefficient is as follows: (See Appendix for Figure 1) p-value of car width is 0.002. engine size p-value as 0.000, The price is defined by the engine size. The P-value of the rear is 0.000, which are the significant elements in overall performance.

1. **Of those three, which had the greatest positive influence on car prices?**

Engine size has the largest beneficial impact on car costs of the three. The cost of producing a larger engine is usually higher for the manufacturer. Increasing the price toward the customer frequently recovers this expense (plus a good profit). With the larger engine, many cars will come with more standard equipment that would be costly extras if the car were chosen with a smaller engine. The price of this item is added to the car's base price.

1. **How accurate was the model?**

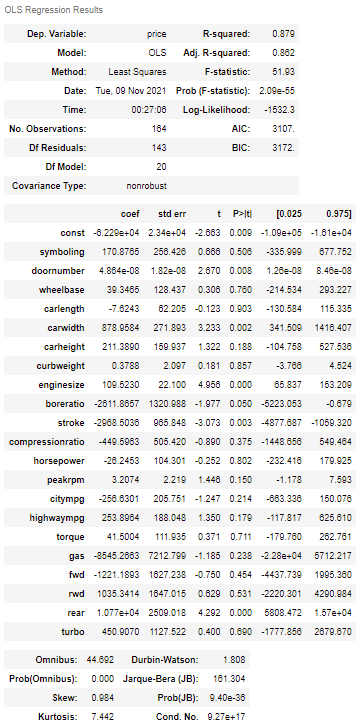
I would rate the algorithm as good because we were able to get an 88% accuracy, which is a great outcome. (see Figure 2 in Appendix).

**Conclusion:**

The model predicted with an accuracy of 88% that the corporation could increase sales by focusing on car width, engine size, and rear. The higher the R-squared, the better the model fits your data in general. An MAE of 2175.7 will be a variation in the price predictions. A good MAE is dependent on the dataset we are working with. It's a good idea to use a naive predictive model to create a baseline MAE for our dataset, such as predicting the mean target value from the training dataset.

**Appendix:**

**Figure 1: OLS Regression Results**

****

**Figure 2: Metrics Results**

