What Factors Influence the Price of a Car

In our analysis we examine the “vehicles dataset” which looks at four hundred twenty-six thousand observations of car sales and the attributes of the car at time of sale. The intent of our analysis is to determine which attributes have the greatest impact on the final sales price of a car. In our analysis we looked at the correlation between sale price and 17 other attributes of which the following are present: id, region, price, year, manufacturer, model, condition, cylinders, fuel, odometer, title status, transmission, VIN, drive, size, type, paint color, and state license plate.

We approach this data using the CRISP DM process, beginning with understanding the business need for the analysis, understanding the data used for the analysis, preparing the data for analysis, visualizing data patterns, creating our data model, and evaluating the performance of our model. In data preparation we normalize and scale our numerical data attributes: price, year, and odometer and convert to determine data skew. In this dataset, price outliers greatly skewed our data so vehicles with a price above $150,000.00 were excluded for our analysis.

After preparing our data we dedicated 25% of the dataset as testing data and ran a regression model where our predicted y variable is the price of the car. Based on the outputs from running the train test split the model was evaluated on Lasso, Ridge, and LinReg. The outputs of the model are then run through GridSearch to determine which variables are the best predictors for determining the sale price of a vehicle.

After running a GridSearchCV pipeline we arrived at the following results: Having a lower odometer, clean title, and newer car are the best comprehensive features for the sale price of a car. There is a penalty on price where cars are 4 cylinder and front wheel drive, it is possible that the approach has led to a misrepresentation of these attributes and a different model would not arrive at this conclusion. On the “type” column the values SUV, Sedan, and Pickup are the most beneficial features. Further model development could refine the features selected to identify additional features that drive vehicle prices. However, in the final results of the model, a low odometer, clean title, and newer car are the most impactful elements to predict a cars price followed by whether or not the car is a SUV, Sedan, or a Pickup Truck.