Group Members: Aamir Patel, Jay Patel, Parth Rana, Kavya Ramasharma

DSC 341- 601

Professor Hamilton

8 June 2021

**Non Technical Summary**

The dataset that we manipulated throughout this quarter contained information on cars sold in America. After further analyzing the many variables that were present, we decided to contain the data significantly by only assessing the value of used cars sold in the state of Wisconsin. All variables that were insignificant to finding car values were removed to determine the price at which the car was sold, year, make, model, color, condition, engine type, fuel type, mileage, title status, transmission type, drivetrain, as well as the type of car sold.

Using R Studio, we were able to find the influential points, or the outliers that greatly affected the slope, of the dataset. We removed 2,523 rows that contained non finite values by creating scatterplots. 87 outliers were also removed due to missing values that were significant to the models that we were creating. (add details about what was removed)

We created three histograms using the dataset to better understand the degree of variation and to show the distribution pattern. The histograms created were for price, odometer, and year. The histograms of price and odometer are skewed to the left, while the histogram of year is skewed to the right. Our histogram for price told us that the min value of cars sold in Wisconsin was 0, the max was $109,995, and the average of all cars sold was $13,543. The histogram for odometer told us the minimum value for odometer was 0 miles, the maximum was 1500000, and the mean was 112469 miles. The histogram for year told us that the oldest car sold in Wisconsin in this dataset was a 1954 model year and the newest car was a 2021 model year. The year histogram also told us that most of the cars sold in Wisconsin were from the 2010 model year.

Using R Studio, we were able to create a model that made it easier for us to calculate predictions for the value of a used car. Our first model made it possible to calculate the price of a car using variables such as odometer and year. We made a total of 3 predictions using our first model. In the first prediction, we used a 2014 car with the price of $15,000 and odometer reading of 20000. The second prediction was a car from 2016 with a $12,100 price and an odometer reading of 59000 miles. Our third prediction was a car from 2005, with a price of $9,000 and 130,000 miles on the odometer. The predictions that model 1 gave us using these variables were: $17216.014 for the first car, $18,382.075 for the second car, and $6,947.624 for the third car. For model 2, we used the same three predictions as model 1. However, model 2 only used the odometer variable to predict the value of a used car. The output for model 2 predictions were: $16,633.49 for the first car, $14,907.05 for the second car and $11,764.03 for the third car.

In conclusion, building these models helped us predict the price of used cars sold in Wisconsin. Using variables such as year and odometer we were able to predict the price that these cars would sell for.