Second Hand Car Price Prediction

Dataset:

The dataset for used car price was obtained from Kaggle and imported to the notebook through csv file.

Problem:

In today's economy it is nearly impossible to know what the cost of things will be in 5 year or even in 5 months. We hear on the news of used car prices skyrocketing more than brand new cars, which gives us some sense that something in the economy is taking an unexpected shift. Yet with the use of training and modeling data, we are able to have some sense of what the prices of cars may be based on trends we have seen.

Approach:

The Approach to take in order to determine this is to first obtain data on used cars for sale, taking down their prices, year of the car, make and model of the car, and other things like mileage and type of fuel consumption. Once the data is obtained, it can be trained and compared through different models (such as RandomForest and Linear Regression models). Then we can compare the accuracy of these models and use them to predict data for the future.

Findings:

Using both models, it was determined that they both product about the same level of accuracy (about 75%) in determining car prices.

Some problems during data cleaning were found that this data was based only in Indian Rupees, and the data did not inform us if this car market is exclusive only to India or also the US. But if we applied the same approach to US data we could use the same models to determine used car prices as well.

Recommendations for Client:

1. First: I would recommend that the client obtain US driven car price data and that this data is obtained from different sources and regions, as prices can fluctuate based on different factors such as states and economy.
2. Second: I would recommend that the data be of all car makes and models as this would give us more diversity in the data and/or we could obtain further information where we know which are the most common used car makes and models and only obtain data from those.