Predicting the Price of Trucks

Abstract:

The goal of my project was to scrape data and build a model to predict prices of trucks using the website Kelley Blue Book. I used this website to gather numeric and categorical features of trucks as well as their prices for the model. The first step was to scrape the data, then build the linear model, and then create a powerpoint to display my findings.

Design:

My project stemmed from the growing car price phenomenon in the United States right now. For my current job, a truck would be hugely useful for carrying different supplies. However, current prices seem insanely high and so I was driven to create a model and try to determine which truck features matter most with respect to price.

Data:

My dataset consists of 1216 trucks each with 6 categorical and 5 numeric features including the target variable price. Some examples of features include, mileage, exterior color, interior, make, age, and condition. In some of these categories, cleaning had to be done and I was also able to bucket some of the features to create more meaningful categorical variables.

Algorithms:

Feature Engineering - Creating Dummy Variables for Categorical Features
Linear Regression Modeling
Data was split into 80% train 10% validation 10% test using the train_test_split function in sklearn, different models with more/less features and more/less data points were compared to each other using adjusted r^2 and RMSE. After changing and improving the model through EDA and cleaning, various techniques like SelectKBest features and minimizing less important coefficients using Lasso, a base model with an adjusted r^2 of .2875 was improved to .6513. RMSE “rose” from 3401 to 6925 but was effectively lowered from 26.6% to 15.7% because I had also added additional data points to my dataset.

Tools:

BeautifulSoup for scraping HTML websites
Python Pandas for building and modifying dataframes
Scikit-learn for models
Numpy
Matplotlib and Seaborn for graphing/plotting

Communication:
Google Docs/Slides for presenting