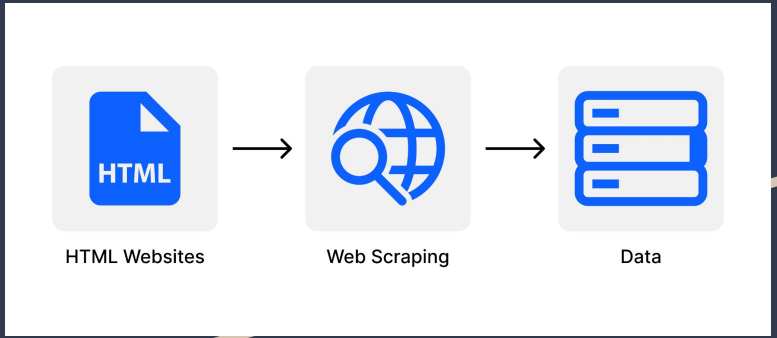
How to Predict the Value of Trucks

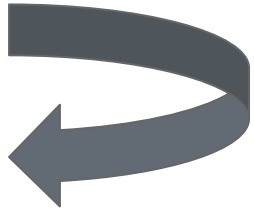
By Erik Paulson

First: Utilize Web Scraping



Web scraping is a method used to gather usable information from web pages, instead of using already gathered data.

- A typical flow looks like this



- KBB.com
  - Individual Truck Listing
    - Scrape HTML for information
      - SAVE YOUR DATA

Kelley Blue Book

- KBB has millions of listings from all over the country!

- This is a great website to find data on all types of cars!

Some typical Information found in a page listing is:

1. Color (Interior/exterior)
2. Year
3. Mileage
4. Make
5. Engine information
6. Drivetrain Information
7. Transmission Information

My Data

From scraped KBB pages,

- 1216 Trucks
- 11 Features
  - 6 Categorical
  - 4 Numeric
  - 1 Target (continuous)

Cleaning the Data

For many trucks, especially certain brands. Many colors have unique names.

EX: Charcoal Gray vs Metallic Gray vs Gray

Are these colors more or less equal?

For some cars, probably no. (EX: sports cars)

For trucks, probably yes.

Any numeric column has to be cleaned before converting from a string to a numeric!

Bucketing Features

Because of the vast, unique values of color. I decided to bucket the feature.

All shades of colors would be bucketed to a base color.

New examples of colors are:

- Green
- Gray
- Blue
- Red
- White

Don’t forget to validate your data!

To validate the data I split the data into chunks.

- 80% Train Data
- 10% Validation Data
- 10% Test Data

The Selling Points

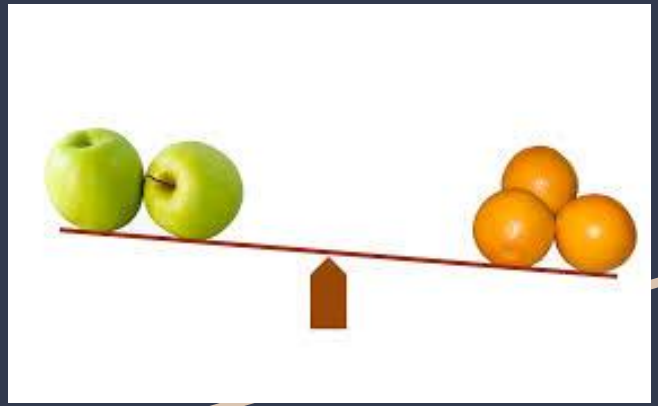
Average Truck Price in my sample: $43,993.44

The Root Mean Square Error: $6925.76

Percentage Prediction Error: 15.74%

This is an ok model given that when pricing a car in KBB, you receive a range and not a singular value.

Compared to an Existing Model



To try and compare my model to the KBB model, I priced my own car to check the price.

20,155 - 22,402 (approx a 10% value range)

However, they have some additional features not included on their car listings.

My model's 15.74% error vs KBB's 10% pricing range is ok considering they have additional features and mine is predicting an exact price.

Explanation of Price!

New Cars are more Expensive than Used Cars

Jeep>Ford>Toyota>RAM>Nissan>Honda>Hyundai in Terms of Price

AWD adds to Truck Value

As age increases value goes down by 1036

For every mile added, value decreases by 0.144

A Larger Engine in Cylinders or Liters increases Value by (1860 per cylinder) or (280 per liter)

How to Improve

Additional features that might be interesting to explore:

- Premium Sound System
- Leather/Heated Seats
- Number of Doors/Cab Size
- Bed Size
- Premium Wheels
- Owner Modified/Aftermarket Parts
- Overall Condition (KBBs metric)

Conclusion

Scraping data and using regression can lead to a great linear model. Sometimes adding additional features can help improve the accuracy of a model as well!

Using data collected exclusively from KBB.com, a decently accurate model that attempts to price a truck based on different features is possible!

Many people can utilize a model such as this one when they are searching for a truck to buy/sell and want to understand the fair price of the product!

Appendix

	0	coefficient	intercept
0	-0.144062	mileage	37192.624159
1	279.899848	engine_size	37192.624159
2	1859.730746	cylinders	37192.624159
3	-1035.584180	age	37192.624159
4	-833.384239	ext_color_Black Exterior	37192.624159
5	-1829.423808	ext_color_Bright White Clearcoat Exterior	37192.624159
6	1818.636989	ext_color_Diamond Black Crystal Pearlcoat Exte...	37192.624159
7	2011.192358	ext_color_Granite Crystal Metallic Clearcoat E...	37192.624159
8	-2500.895190	ext_color_Gray Exterior	37192.624159
9	-2643.427497	ext_color_Onyx Black Exterior	37192.624159
10	-4508.939932	ext_color_Satin Steel Metallic Exterior	37192.624159
11	-2051.652874	ext_color_Silver Exterior	37192.624159
12	-892.922956	ext_color_Silver Ice Metallic Exterior	37192.624159
13	-3600.395907	ext_color_White Exterior	37192.624159
14	-1705.124790	ext_color_other	37192.624159
15	-2163.228659	int_color_Gray Interior	37192.624159
16	-1719.623056	int_color_other	37192.624159
17	3952.214906	transmission_Manual	37192.624159
18	7419.987218	awd_yes	37192.624159
19	2830.287360	make_Ford	37192.624159
20	208.763686	make_GMC	37192.624159
21	-6802.825887	make_Honda	37192.624159
22	-10597.710799	make_Hyundai	37192.624159
23	9600.914186	make_Jeep	37192.624159
24	-6425.062085	make_Nissan	37192.624159
25	918.782259	make_RAM	37192.624159
26	1014.107411	make_Toyota	37192.624159
27	-1122.810937	condition_New	37192.624159
28	-2540.434404	condition_Used	37192.624159

```
big_df
```

	mileage	ext_color	int_color	engine_size	cylinders	transmission	awd	price	make	age	condition
0	105410	White Exterior	Gray Interior	5.7	8	Automatic	no	18588	RAM	8	Used
1	115000	White Exterior	other	5.3	8	Automatic	no	18788	Chevrolet	11	Used
3	128976	other	Gray Interior	5.7	8	Automatic	no	18990	Toyota	14	Used
4	88150	Silver Exterior	other	3.4	6	Automatic	no	18995	Toyota	20	Used
5	112241	White Exterior	other	2.5	4	Automatic	yes	18995	Chevrolet	6	Used
...	...	...	...	...	...	...	...	...	...	...	...
1263	13157	other	Black Interior	3.5	6	Automatic	no	43641	Toyota	2	Used
1264	68078	other	Gray Interior	3.5	6	Automatic	yes	43599	Ford	2	Used
1265	49850	Silver Ice Metallic Exterior	Black Interior	5.3	8	Automatic	no	43549	Chevrolet	3	Used
1266	25856	White Exterior	Gray Interior	3.5	6	Automatic	yes	43494	Toyota	2	Used
1267	31335	Black Exterior	Black Interior	5.3	8	Automatic	no	43289	Chevrolet	3	Certified

1216 rows × 11 columns