Homework 6

Devashish C. Thakar

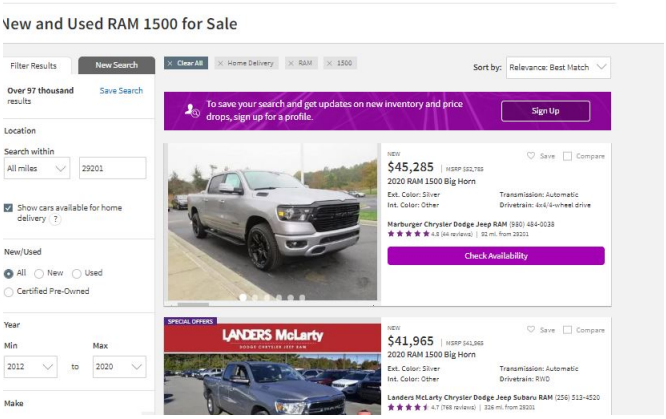
11 Nov 2019

Part I: Description of Data Collection

This is a part of my effort to collect the details of prices of car listings from cars.com, and collect data over a long period of time to build a model to determine the price changes on the listings over time and the possible antecedents. I am still far away from that, but I have started collecting this data, and I describe the process here.

The first step is to collect the basic details of the listings. I filter for the 'most relevant' listings in the zip code 29201. They look like this:

FIGURE 1: **cars.com listings**



I extract the details such as price, dealer rating, dealer reviews, brand etc. These details are stored in a json type format. Therefore, collecting these details was a simple case of looping over all pages and storing the details in a list.

In the preceeding step, I also collected the listing URL that provides me access to more details of the listing. I use these URLs to collect the details of features such as Safety, Convenience, Entertainment, Exterior,and seating. The data collection for these features was a little twisted, since each feature was located within a 'class', and the actual features were lists within each class (As shown in the figure):

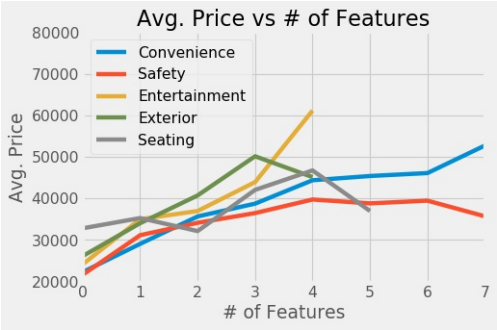
FIGURE 2: **cars.com listings**

```
<div class="page-section__row details-section__normalizedFeatures" ng-non-bindable>
        <div class="details-feature-list details-feature-list--normalized-features">
            <h2 class="page-section__title--sub cui-heading-2">Convenience</h2>
            <ul class="details-feature-list__list">
                <li class="details-feature-list__item">Keyless Start</li>
                <li class="details-feature-list__item">USB Port</li>
            </ul>
        </div>
        <div class="details-feature-list details-feature-list--normalized-features">
            <h2 class="page-section__title--sub cui-heading-2">Entertainment</h2>
            <ul class="details-feature-list__list">
                <li class="details-feature-list__item">Bluetooth</li>
            </ul>
        </div>
        <div class="details-feature-list details-feature-list--normalized-features">
            <h2 class="page-section__title--sub cui-heading-2">Safety</h2>
            <ul class="details-feature-list__list">
                <li class="details-feature-list__item">Backup Camera</li>
                <li class="details-feature-list__item">Brake Assist</li>
                <li class="details-feature-list__item">Stability Control</li>
            </ul>
        </div>
        <div class="details-feature-list details-feature-list--normalized-features">
            <h2 class="page-section__title--sub cui-heading-2">Exterior</h2>
            <ul class="details-feature-list__list">
                <li class="details-feature-list__item">Alloy Wheels</li>
                <li class="details-feature-list__item">Tow Hitch</li>
```

It involved findings all tags for features, then storing each class within those tags as the 'key'(feature name), and the actual feature as the 'value' (all this done using a dictionary). Thus, I had to iterate over each feature class, and within each feature listing within the class.

After getting this data in a columnar format, with each column representing the feature, and each cell containing a list of the actual features, I appended this data to the listing details. I check if the average price is somehow related to the number of features in each category. Here is how that relationship looks like:

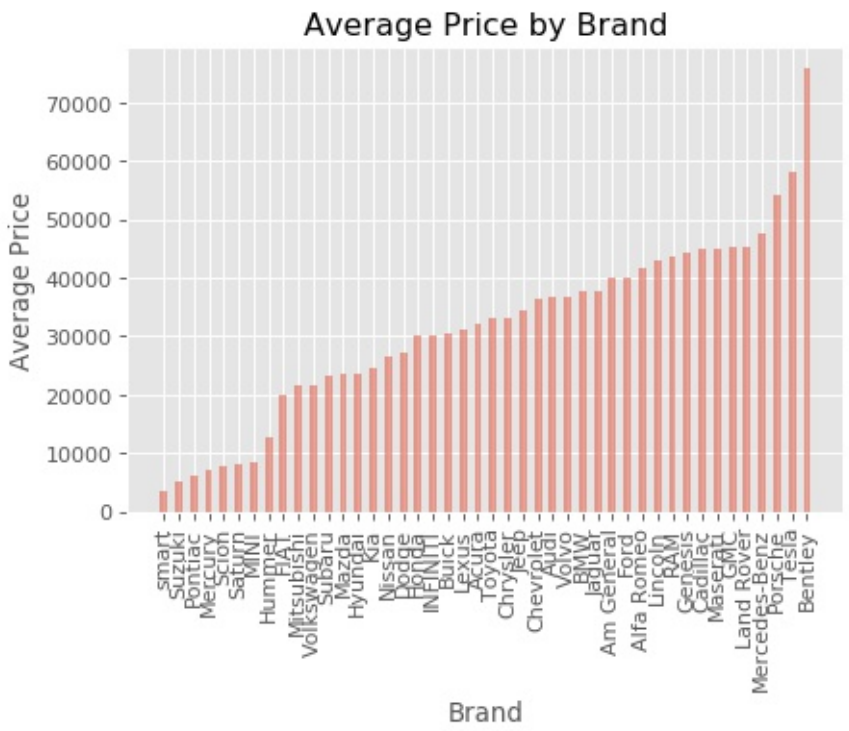
FIGURE 3: **Price vs Number of Features**



Generally, this kind of relationship is expected, but the relationship with number of safety and number of seating features is interesting. With number of safety features, the average price flattens at 4 and declines at 6. With the number of seating features, it declines at 4. I test these relationships econmetrically later.

Next, I look at the average prices by the car brand. I leave out Bentley and Rolls-Royce, since their prices were outliers (avg. prices of about $2,50,000). The average price by the brand is as follows:

FIGURE 4: **Price vs Brand**



Part II: Econometric Model

I test two models to check the econometric reltionship between the features and price: 1) Brand FE 2) Standard Errors clustered on the brand. The results are shown in table 1. It can be seen that for every unit increase in the number of safety features, the price reduces by approximately $ 1100 (FE model) or $1800 (Robust SE model) (from the regression coefficient for Num Safety Features). It seems like the number of exterior features provide the highest per unit increase in price. Specifically, with every extra exterior feature, the price increases by approximately $4200 (FE) or $ 5300 (Robust SE model). Interestingly, the number of entertainment features does not have a significant impact on the price. On a different note, dealer ratings does seem to positively impact the price. This could be highly endogeneous, since good dealers might be selling only high end cars. But I do not have enought variables to test the endogeneity. Maybe in the future, I will collect some more data and test it.

TABLE 1: Regression Results for price

	Dependent variable: Price	
	Brand FE	*Brand Robust Errors*
	(1)	(2)
Dealer Rating	408.554**	652.908***
	(202.793)	(214.128)
Dealer Num of Review	−0.360*	−0.442
	(0.201)	(0.343)
Num Conveneince Features	1,344.907***	1,547.592***
	(147.104)	(581.811)
Num Safety Features	−1,127.663***	−1,809.265***
	(132.993)	(412.821)
Num Exterior Features	4,154.664***	5,330.887***
	(243.840)	(866.693)
Num Seating Features	2,131.789***	1,803.976***
	(149.352)	(398.568)
Num Entertainment Features	390.123	−22.167
	(266.724)	(980.450)
Constant	826.250	2,946.785***
	(10,285.630)	(1,035.595)
Year FE	Yes	Yes
Brand FE	Yes	No
Observations	4,493	4,493
R^2	0.654	0.35951
Adjusted R^2	0.648	0.35477
Residual Std. Error	10,081.400 (df = 4414)	
F Statistic	107.023*** (df = 78; 4414)	75.8451*** (df = 33; 4459)
Note:	*p<0.1; **p<0.05; ***p<0.01	