
ISyE 2028 – Basic Statistical Methods - Fall 2015

Bonus Project: “Big Data” Analytics

Final Report

Methods to Evaluate the Market Price of Used Cars

John Barbour

Background:

Determining the price at which a product should be listed for sale is an important part of business strategy. This also holds true in the used car market. Resources like NADA and Kelly Blue Book help in determining a list price, but knowledge of the local market is also essential. Car buyers have access to a deep source of inventory through sites like auto-trader and craigslist, putting retailers at a disadvantage if they are not considering the pricing found in these channels.

Method:

We attempt to solve this problem by applying statistics within a python script. Each paragraph below describes the application of a statistical method to what is a ‘Big-Data’ problem. The end result untangles the ‘last mile’ of data, helping businesses maximize their list prices without losing sales.

Filtering out bad data using interquartile range and outliers-Craigslist in particular is a data structure nightmare. Individual ads can be posted by anyone and contain anything. Instead of posting the actual list price, dealers will routinely list the ‘monthly payment’ or ‘minimum down-payment’ for a vehicle. This information shows up in the listing title and shouldn’t be considered as part of the distribution. The image below shows a sample ad of \$315 for a 2013 Honda Accord. This filtering occurs after putting the data into the format shown below.

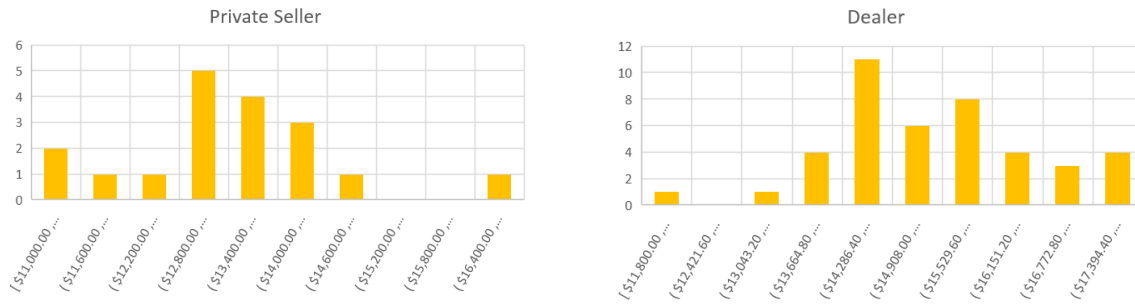
★ Nov 21 2014 Honda Accord LX - \$17988 (Canton) [pic](#) [map](#) [x]

★ Nov 20 2013 Honda Accord EX-L SKU:DA020313 Honda Accord EX-L Sedan - \$20456 (AUTONATION - Financing Available OAC) [pic](#) [x]

★ Nov 20 2013 Honda Accord EX-I V6 \$315 Mo.- All Credit Situations - \$315 (Atlanta) [pic](#) [map](#) [x]

```
['16500', '2013', 'Honda', 'Accord', '2015-11-26 13:30']
['20456', '2013', 'Honda', 'Accord', '2015-11-26 01:12']
['20456', '2013', 'Honda', 'Accord', '2015-11-20 22:43']
['315', '2013', 'Honda', 'Accord', '2015-11-20 20:11']
['315', '2013', 'Honda', 'Accord', '2015-11-13 13:32']
['315', '2013', 'Honda', 'Accord', '2015-11-06 13:18']
['315', '2013', 'Honda', 'Accord', '2015-11-06 09:11']
['20456', '2013', 'Honda', 'Accord', '2015-10-31 11:45']
['16999', '2013', 'Honda', 'Accord', '2015-11-28 11:01']
['16500', '2013', 'Honda', 'Accord', '2015-11-02 21:53']
['18500', '2013', 'Honda', 'Accord', '2015-11-02 17:10']
```

Using histograms to communicate the pricing distribution-Since the pricing data comes from two separate populations, dealer owned vehicles and privately owned vehicles, we compute a histogram for each and compare results. These are shown below along with a summary of our data for each population.



Results			
	Private	Dealer	Joint
n	18	42	60
mean	\$13,387.33	\$15,400.62	\$14,796.63
st. dev	\$ 1,361.33	\$ 1,407.20	\$ 1,666.04

Using a c.d.f. complement to recommend a sale price-Although advertising a price is subject to things like quality of exterior, cleanliness of interior, and specific trim items like navigation and heated seats, getting into the search results is the first step to getting seen. While it is impossible to determine the probability distribution for the price that a customer will inquire about a particular vehicle, we can use statistics to indicate the extent that a car is over or under priced relative to the market.

To do this we consider the entire scale of vehicle prices and the supply and demand relationship.

