

Ryan Neumann
Artificial Intelligence
Project Milestone
8 November 2016

My project is going to be based around housing prices, and how the features of the home influence the price. Currently, there are hundreds of data points considered when looking to buy a house. With an almost infinite amount of options, determining the price of a house can be quite a difficult thing to accomplish.

My idea is to narrow it down to a handful of factors that are able to calculate an accurate price without going through all the hassle of considering every single aspect. Primarily, I would calculate my estimate around the location of the house. Considering factors such as distance to schools, type of neighborhood, and general environment are essential when looking to buy a home. Therefore, it would have the greatest influence on the estimated price.

After calculating the price due to location, I would then focus on the details of the house. This would also help people who are trying to sell their homes. As a seller, an accurate appraisal of the home is extremely important, especially with the size of the investment on their end. Determining price points for what a home does and doesn't have will aid in calculating an estimate that will be worth considering. Each of these would then be added into my algorithm, either raising or lowering the value of the home. Keeping it simple, I would like to have it either be a 1 or a 0. Either the home has it, or it doesn't. This way, it will be a simple algorithm, described to buyers/sellers in a simple way. The transparency of what is determining the value of the home allows for it to be trusted, having confidence that you know it's worth it's determined value.

Currently, the main method of appraisal for real estate agents is with the automated valuation model (AVM), which calculates a ranking and rating for the property through various different inputs. AVM looks at any previous sales for the home, as well as the homes around it, along with the amount of search hits for that specific location. With this data collected, it provides an estimated valuation along with how it ranks among other properties.

However, much of this information that is entered into the AVM is useless in today's time. What the house sold for in 1950 is completely irrelevant to what it would sell for currently. With inflation and the value of the dollar constantly decreasing, these variables would best be left out of the equation. This model serves as an advantage to real estate agents, while simultaneously confusing the investor involved. With the properties being listed in descending order of popularity, the investor is most likely to click the first couple listings.

This, in turn, would cause the top listings to skyrocket in price, while lowering the value of equivalent homes. Essentially, the AVM is an algorithm set up to make real estate agents the most profit.

This is unfair to the buyers of a home, as they are directed to the most popular estates rather than one that best suites their needs. In hopes to repair this hindered algorithm, I will recreate my own version which suites the needs of all individuals rather than just the needs of the real estate agent. As described before, my algorithm will consist of two factors: the location and the details of the house. Compared to the AVM, I will disregard the hit rate of a house as well as any previous sales.

Using the Boston house-price data by Harrison, D. and Rubinfeld, D.L., I factored in the proportion of residential land zoned for lots over 25,000 square feet, the average number of

rooms per dwelling, as well as the pupil to teacher ratio by town. Out of the fourteen different factors given for the dataset, I found only these three have a significant impact on the evaluation of an estates worth. While the provided information specifically looks at the Boston area, I plan to use my algorithm at a much bigger scale, allowing anyone around the world to get an accurate appraisal.

With this information, I can calculate the population as well as the number of homes owned in the region of the house for sale. These detailed statistics are key when attempting to sell a home, as they are the largest contributors to the price point. For families with young children, being close to an educational facility is an aspired benefit for the whole family. For individuals, knowing how populated the area surrounding the home can significantly impact their decision to purchase.

Having the buyer's input when they are purchasing a home not only benefits the real estate agent, as they are making an informed decision on what houses to offer, but also benefits the buyer, as they are only given options that suite their needs. Associating a price value to all of the buyer inputs will add transparency to the determined price of a house, putting both the buyer and seller at ease throughout the transaction. Currently, many individuals use a real estate agent which acts as a middleman during this transaction. Often times, the real estate agent creates complications for both ends, making the process long and dreadful. Creating an algorithm that would essential remove the middleman will dispel much of the hassle of purchasing a home.

Ultimately, a person selling a home would like to be given a fair price for their investment. The buyer ideally would like to know exactly what it is they're paying for, as well as what factors lead up to that price. With those involved in the current real estate market, they

tend to create a barrier between the buyer and seller, creating an unneeded discrepancy.

Removing this barrier, making it a one to one transaction, would primarily lower the cost of purchasing a home while simultaneously increasing the profit for the seller.

References

<http://lib.stat.cmu.edu/datasets/boston>

<https://ww2.amstat.org/publications/jse/v19n3/decock.pdf>

<https://www.google.com/patents/US20070033122>