Housing Market Sale-price Predictor

The housing market has been quite volatile in recent years. As the economy waxes and wanes, interest rates inch their way up, then come crashing down, for instance with the recent pandemic, or the 2008 recession, it can be difficult to know what a home is worth. Buzz words abound in real estate media, terms like buyers market, sellers market, features like pools, large garages, decks or patios. What adds value and what doesn't? And who can say? With home prices quickly trending up and down, can realtors accurately price homes in the rapidly changing market conditions that exist today?

Enter HoMS(Housing Market Sale-price) Predictor. With a few details about a particular home, anyone can determine a fair market sale price for that home. However, this model doesn't replace a realtor, and should not be taken in place of the opinions of a qualified realtor. There are always exceptional homes that can't be accurately predicted by this model. Primarily, this model is excellent for "average" homes, with typical features. It does, for instance account for the presence of a pool, but the pool itself could either be in disrepair, or be an exceptional pool, say an infinite pool or a pool with exceptional rock features or masonry work that would warrant a higher price than could be predicted by this model.

In addition to an inability to accurately value an exceptional(non-typical) home, this model does have a few other limitations. For instance, this model cannot take into account the cost of renovations. A home seller may expect to have a return on a renovation, and such a return might be possible with the common understanding between a buyer and seller, however, these types of renovations can't really be determined. Whatsmore, some data used in the model is open to some level of interpretation. For instance the quality of a home isn't something that is precisely quantifiable, and as a qualitative judgment, a buyer would need to verify all details of the particular home, to determine whether such values are appropriate.

The HoMS Predictor could be improved in the future by allowing photos of key features to drive out some of these qualitative judgments. For instance, a photo of a fireplace, or kitchen cabinetry inside a home may be a good indication of its quality, and offer a more reliable means of predicting a fair sale price. Additionally, receipts for renovations could also play an integral part in assessing both quality and expected condition. For instance a recently renovated home with a renovation with a moderate price point could likely be more accurately defined for the purposes of our model. Perhaps details such as particular features or tile, carpet or wall color could be incorporated to determine what may drive inflated valuations, or what buyers may perceive to warrant a higher sale price. Such an example of this might be that green reminds us of money, and tile with a green tint may be determined to be perceived by buyers as more money, or ceiling mounted pot racks may give a kitchen a more distinguished look that buyers would be willing to pay more money for.

The models that were considered for these purposes includes Support Vector Machine Classifier, Gaussian Naive Bayes, and Random Forest Classifier models. Each of these models was selected for it's classifier power. As most of our variables are discrete, and there is a high number of variables, it would make sense for these models to work well, as they are classifiers. In fact, as classifiers, they do roughly what a realtor would do, which is to use sales of the most similar homes to determine an appropriate sale price. SVM and RandomForest models are very similar, while the Naive Bayes model takes variables to be independent of one another. One difference between the RandomForest model and the SVM is that the variables in the SVM must be scaled, but this step is unnecessary for the random forest model, however the data was already scaled when it was consumed by the Random Forest model, which had no effect on the result.

Some of the key findings, as one may expect, is that square footage is most highly correlated with sale price. Larger homes sell for more money, but what may or may not surprise users is that ground floor square footage is also strongly correlated, so a home with two floors of equal area would generally sell for less than a home of the same size on a single level. While one might assume that this could be due to a larger lot size, which would be a requirement for a home with a larger footprint, this model does not take into account lot size, as it was not very strongly correlated with sale price. This is likely because a very small home in disrepair on a large plot of land may sell for the price of the land, which may be quite low, as the price per acre generally decreases as the plot grows, whilst also a very large expensive home on a very small plot of land also sells for quite a high price.

The primary use case for this model is for a home buyer or seller, or a realtor working with buyers and sellers, to be able to find an accurate value for homes based on recent sales data. One of the primary jobs of a realtor is to find ‘comps’ in the area. Such a task, is somewhat similar to what our model does here, however it is largely up to the realtors judgment. This model may point out comps that a realtor would have otherwise dismissed for one reason or another, or it may not include comps that a realtor would include in a report.

A secondary use case would be for a buyer to make the most cost efficient upgrades to their home prior to sale. For instance, if I’m looking to sell in the next year, perhaps a $10,000 kitchen renovation would warrant an upgrade in overall quality and will increase my sale price by $20,000 increase in sale price. Or adding an $18,000 detached garage might warrant a meager $5,000 increase in sale price. This is not something that should be done without the expertise of a realtor however, as mentioned before, such qualitative judgments are not able to be made by this model, and particular features play no part in it’s predictive power, with the exception of a pool feature.

Another strong use case would be to help buyers compare homes to determine the best value home out of a few homes that they are looking at putting offers in on. If a buyer is otherwise indifferent, they may as well choose the home that is the best deal, or if a buyer is looking at paying more for their favorite home, perhaps this model could reveal that the higher price is in fact justified, or if it really isn’t worth the extra cheddah.