Spring 2024

Project Proposal

**Team Member Names**: David Grice, Aaron Kwan, Kevin Rice

**Project Title**: Bid Smart: Navigating Real Estate Valuations

**Problem Statement:** Helping North Carolina home buyers successfully bid the right amount on a house listed for sale in a given market.

Real estate market conditions are constantly changing, and the landscape of supply and demand is turbulent. Though there may be a negotiable range on the valuation of a home from a buyer’s and/or seller’s perspective, a home’s value in any given market is relatively easy to measure in confidence. Despite any range of value that may exist, that range can only be so wide, as there are several industry resources that make up the valuation. The primary resource when it comes to home valuation is an appraisal by a licensed professional which takes into consideration several key factors such as square footage, structural improvements/additions/renovations, number of bedrooms, architectural style, HVAC system age and operational efficiency, foundation, and appliances. Generally, lenders will *not* loan any funds to a buyer more than the appraised value.

With that being said, a home’s value does not necessarily correlate with how much it will sell for, especially in a “hot” market where supply does not meet demand. This holds true for North Carolina in particular. In addition, the COVID-19 pandemic has drastically changed the norms of real estate markets across the country, especially in highly populated metropolitan areas. With a large portion of the work force working from home, over the past few years since the pandemic many people have moved their families *out of* traditionally expensive areas and *into* traditionally cheaper areas. This is where the correlation between home value and what a home actually sells for begins to break. Particularly in these “hot” markets, but around the country in general, it is commonplace for a home to sell for thousands of dollars *over* listing price with no concessions to the buyer at all. Prior to the pandemic this was virtually unheard of and has now created an “extreme” seller’s market. This makes buying a home a difficult situation to be in. Is the market cooling down and can buyers go back to making “normal” offers or do they need to offer over listing price if they want a competitive chance at beating competition? Additionally, *how much* (if at all) does a buyer need to offer over listing price to have a competitive bid? The answers to these questions are the ultimate goal for this project, as we aim to build a model that can accurately predict how much a home will actually sell for in a given North Carolina locality – helping buyers enter a negotiation with a realistic expectation of affordability and how to be competitive.

**Statement of Work:** Rather than just simply estimating the *value* of a home listed for sale, the team aims to focus on **regression**, **decision tree**, and **game theory** based models to predict what the optimal amount a home buyer would need to bid to get a fair, competitive offer in any given market in North Carolina. We will then apply our model to other regions in the United States to evaluate efficacy.

**Methodology:** The methodology to estimate a home’s *value* has largely been discussed in the problem statement. Having established the growing lack of correlation between a home’s value and what it will actually sell for, the methodology for our model needs to expand to new and/or alternative features and predictors to accurately estimate a home’s sell price (what a buyer should anticipate is needed to win the bidding war). We find it best and most useful to break the data down to the zip code level of locality to help reduce some level of noise and variance within the data. Realtor.com (*Realtor.com Real Estate Data and Market Trends for Download*, 2024) offers 56 unique metrics by zip code that offer insight into supply and demand, home features, and popularity. Zillow (*Housing Data - Zillow Research*, 2024) provides a Zillow Home Value Index (ZHVI) metric which is a smoothed and seasonally adjusted measure of typical home value in the 35th to 65th percentile range. This potentially provides a unique insight into the market, apart from medians and averages which are more frequently used in the Realtor.com metrics. We have also gathered monthly average interest rates in the U.S. (*30-Year Fixed Rate Mortgage Average in the United States*, 2024). Drawing inspiration for new perspectives from (*The hottest U.S. housing markets,* 2024) and (*Surveys of Consumers*, 2024), we have also added in our own calculated measure which is the ratio of active listings to total listings to help give a unique perspective on supply and demand. We plan to evaluate these sources more and potentially add additional metrics similar to this as we become more familiar with the data.

All data from these sources was combined by zip code and month to form a workable dataset to train our model on. Due to the year-by-year variability in housing markets, especially considering the effects of the COVID-19 pandemic, we wanted to limit our training data to as recent as possible to hopefully increase predictability. We plan to train our model on monthly North Carolina data from the entire year of 2022, and then validate the model on monthly North Carolina data from the entire year of 2023, to test the model to make predictions for the current year of 2024.

North Carolina real estate markets tend to be seasonal and much more active in the spring and early-summer months. Thus, we will have to account for this potential seasonality in the model and make a seasonal adjustment to provide a clearer signal.

**Evaluation:** As we develop and validate the model, we will evaluate the results with real sales that have already taken place in 2024 in North Carolina and see how closely the model is able to match the sale prices. If the results are less than optimal, we will evaluate the variables further to determine how results can be approved. Likewise, we will consider additional variables that can be added to further improve results, such as new calculated measures, sentiment analysis, or additional sources. The team will also look into dimensionality reduction techniques like Principal Component Analysis before utilizing our predictor model to try to simplify the inputs to our model and to reduce potential overfitting caused by one feature or another. We plan on using North Carolina as our initial set of data, splitting 80% of the data as a training set and evaluating using the remaining 20%. We will then use other states to further evaluate our model to see whether our model is generic enough to be applied to other markets or if it’s region specific. When buying a home, every dollar counts, and small percentages of inaccuracy can have a significant impact. Therefore, we would ultimately like our model to predict offer values within 2% of actual sale values during testing.

**Data Source(s):**

*Realtor.com real estate data and market trends for download*. (2024, March 5). Realtor.com Economic Research. <https://www.realtor.com/research/data/>

*Housing data - Zillow Research*. (2024, February 1). Zillow. <https://www.zillow.com/research/data/>

*30-Year fixed rate mortgage average in the United States*. (2024, March 5). <https://fred.stlouisfed.org/series/MORTGAGE30US>

*The hottest U.S. housing markets* (D. Bradbury & L. Rajala, Eds.). (2024, February 12). U.S. News. <https://realestate.usnews.com/real-estate/housing-market-index/articles/the-hottest-housing-markets-in-the-us>

*Surveys of Consumers - Demographic Subgroups*. (n.d.). <https://data.sca.isr.umich.edu/subset/subset.php>