What are we trying to do?

We seek to build an interactive visualization that will allow home buyers to understand **why** a house is valued the way it is. For example, we seek to determine what amenities, such as restaurant proximity or school rankings, are driving the value of specific houses. Our tool will allow a consumer to look at a house or neighborhood and understand what are the drivers of its relative cheapness or expensiveness. We will develop a test case for this visualization using one city.

How is it done today?

Most products and research do not focus on the perspective of the home buyer. Many websites look at prices from the perspective of a real estate agent or appraiser. Real estate agents and appraisers are primarily focused on determining the right price to sell a house, not determining why the right price is the right price. Websites like Zillow or Trulia are commonly used by consumers to visualize what a house is worth. These websites tend to use models that focus heavily on neighborhood peer sales and qualitative adjusts to determine housing values. Research in this area has focus on the decomposition of house and land value using neighbor peers(Xu 2018). Pagourtzi 2003). The data outputs result from these research models typical takes the form of a house price estimate. Although we will not use these estimation methods, we will use estimated house price values as the independent variable in our analysis.

Significant amounts of research has also been done from the perspective of an asset owner or government policy maker. This research tends to focus on what variable might change a houses price in the future. In particular, this research looks at the impact of macro-economic factors on housing prices (Diewert 2015). This type of information is important to entities, such as real estate investment trusts or central banks, who are interested in how housing prices might evolve with respect to macro economic factors. For example, researchers have attempted to use statistical learning technique to identify time series risk and trends in housing prices (buck 2008, clap 2002). Others have examined the housing market from an efficient markets perspective (case 1989). This work has very limited application in our project because we are not interested in house prices over time, but may serve to help estimate the variability of estimates produced by our project.

Research on the “why” factor of housing prices has tended to be single variable with a focus on time series analysis instead of cross sectional with a focus on house price differentiation. These papers tend to be more recent than other research. Several researchers have touched up topics relevant for this project. For example, Evans and Malin determined the relationship between magnet school status and neighborhood home values. Glaeser, Lim and Luca have measured the effect so gentrification using YELP data. The types of data sources used in these research papers will form the basis of our project.

What’s new in your approach?

The information provided by Zillow or the macroeconomic approach falls short in providing home buyers information about “why” a house is priced the way it is. Answer like “because its neighbors are expensive” or “because GDP is expanding” don’t help anyone figure out why a house is worth what is worth. For instance, 2 homes with identical zestimates might derive that value through completely different amenities. For example house A is in a top school district, but it is far from public transportation and highly rated restaurants whereas house B is in a poor school district but is within walking distance to multiple 5-star restaurants and other entertainment. The consumer is likely more interested in why a price is priced where it is.

Our approach is different because it explains what drives house prices using the factors considered important by ordinary consumers. Our approach will use statistical techniques that cross reference multiple data sources to build a model of a houses value based on amenity. This cross-sectional approach will provide a different set of information compared to other tools. The cross-sectional approach we plan to use is inspired by factor risk attribution in the asset management industry (Meucci 2010). It will be successful because this is the information consumers actual want and need when making a house decision.

Who cares?

Consumers should be more focused on what they are getting for their money, and what tradeoffs they might need to make if they don’t have enough purchasing power to buy their perfect house. Price sensitive home buyers who only want to pay for the amenities that are most important to them personally will be most impacted by this tool’s capabilities.

What impact will it make? How do you measure them?

The impact of this project will be a shift in the way homebuyers budget for a house from a single, opaque home value to a more detailed plan of how much they are willing to spend per amenity.

What are the risks and payoffs?

The main payoff is increased transparency in the home buying process. The main risk is that we will not be able to distinguish a big enough difference in the component price of homes in various neighborhoods to make it worth the user switching tools.

How much will it cost? How long will it take? (grouping these two)

We plan to use free data sources and open source analysis tools. The monetary cost of this project will be small. The real cost will primarily be in time spend developing the project. In the scope of this class, we expect the construction of a prototype of this model will take about 400 person hours (5 people, 10 hours per week, 8 more weeks). A full implementation of this model to cover all US cities could take a year or longer and is beyond the scope of this class.

PUT in detailed plan of action here

How will progress be measured?

USE that plan of action to dictate progress measurement

Example

10/17/18 Finishing pulling from APIs

10/24/18 begin statistical analysis

10/31/18 determine visualization tool and begin work on visualization using preliminary data

[Discuss with group. I think we need to more clearly define a goal. Is it to reduce the time it takes to buy a home? Reduce home turnover rate (meaning how many years until the same home is sold again) or is it something else]

1. Edward L. Glaeser, Hyunjin Kim, Michael Luca. "Measuring Gentrification:Using YELP Data to Quantify Neighborhood Change", 2018
2. Angela Buck, Takisha Harrison, Linden Johnson, Holly Sontag. "Using Real Figures to Invest in Real Estate: A Multivariate Statistical Analysis of the US Housing Market", 2008
3. Chad Evans and Joel R. Malin. "The Relationship Between Magnet Status and and Neighborhood Home Values in Chicago"
4. Yangfei Xu, Qinghua Zhang, Siqi Zheng and Guozhong Zhu. "House Age, Price and Rent: Implications from Land-Structure Decomposition", 2018
5. W. Erwin Diewert, Jan de Haan, and Rens Hendriks, "Hedonic Regressions and the Decomposition of a House Price Index into Land and Structure Components", 2015
6. Elli Pagourtzi, Vassilis Assimakopoulos, Thomas Hatzichristos, Nick French, "Real estate appraisal:a review of valuation methods, Journal of Property Investment & Finance", 2003
7. Stephen Sheppard and Andrew Udell, "Do Airbnb properties affect house prices?", 2018
8. John M. Clapp and Carmelo Giaccotto. "Evaluating House Price Forecasts", 2002
9. Case, Karl E. and Shiller, Robert J. “The Efficiency of the Market for Single-Family Homes”, 1989
10. Meucci, Attilio, Factors on Demand: Building a Platform for Portfolio Managers, Risk Managers and Traders, 2010

This paper contains a summary of methods used for generalized asset price factor contribution estimation. These methods vary from simple to state of the art. The paper empshises point in time estimation, which is critical for our cross sectional analysis house prices.