**Team 012: BubbleWarn**

*Derek Cheng, Mahaveer Jain, Micah Jeng, Peter Hernandez, Tsz Kin Chan, Victor Cerabone*

## **1. The Problem**

Since the 2008 housing market crash, there have been economic debates over the potential of a similar bubble in the United States. According to Hans Lind (2008), a bubble occurs if the price of an asset increases and falls dramatically within a short term **[1]**. When housing demand increases, it attracts attention from speculative investors. Karl E. Case et al. (2004) conducted a survey of homebuyers in San Francisco in 2002 before the 2008 housing crash. On average, the results showed respondents were hopefully optimistic, expecting a 15.7% yearly average rate of return over the next decade **[2]**. Dieter Gerdesmeier et al. (2011) explains further that the eventual overvaluation of these assets causes these prices to be unsustainable, and a bust of this bubble that may lead to macroeconomic instability **[3]**.

Dean Baker (2006), Björn Sjöling (2012), and Paolo Gelain et al. (2018) explain that there are economic factors that can contribute to this cycle. In particular, Baker found that measuring the change in housing prices through the House Price Index (HPI) gave strong evidence of a bubble through the mid-1990s **[4]**. Sjöling concluded that, from 2000 to 2010, comparing housing prices vs. income had the predictive power needed for the Spanish and UK markets **[5]**. Gelain gathered that the relaxed lending standards that supported excessive borrowing contributed to the US housing boom **[6]**.

Of all these factors, we decided to look at the trend of housing affordability over the years. Aaron Hedlund (2019) found that the rent-to-income ratio has grown from 25% to over 28% in the last decade **[7]**. The disproportionality between the increase in income and property prices causes a decrease in affordability and some of the changes are without any fundamental support.

To resolve these challenges, we plan on building a tool that can provide insights on the state of housing affordability and abnormality. However, according to Quan Gan et al. (2009), it is important to not focus on just the average income households because the housing affordability problem may be worse for lower income households than is suggested by standard median measures **[8]**. We also plan to provide interactive options to select by region, profession, race, etc. in our data to accommodate this finding.

## **2. Limits of Current Practices**

The typical method of determining when and where to invest into real estate is opinion based recommendations from reputable sources as stated by M. Stone (2010) **[9]**. Sources such as Zillow **[10]** have gone as far as to compute a housing index (ZHVI) that measures home value and market changes in a given region. ZHVI is a viable indicator but it is derived from aggregating prices from similar houses and suffers from the cheerleader effect rather than actual support for the price. The index is further limited in that it alone does not provide sufficient information to estimate disruptions in the housing market.

There has been a lot of research done on predicting house prices using various machine learning approaches such as Artificial Neural Network (Limsombunchao) **[11]**, multiple regression or hedonic model (SC Bourassa et.al) **[12]** but suffered the same issue as ZHVI, where most models rely on only current public listings such as the regression model proposed by Alejandro Baldominos et. al (2018) **[13]**.

## **3. Project Approach and Implementation**

Data will be automatically collected through API and scraping and combined from various sources such as Zillow **[10]** and ACS **[14]**. We will explore various parameters to define affordability as it is examined by Zi Cai(2017) [**15**]. Paul Waddel **[16]** utilized data science to examine house affordability with different criteria, we will examine similar methods in our problem. Users will be able to make key decisions using a data-driven approach instead of opinions, an important foundation for informing audiences as backed by Betsy Gardner (2019) **[17]**. Decision-making time will also be greatly reduced by presenting results on one platform, instead of having the users switching between platforms.

We will compile an intuitive interface allowing the user to customize their search to the area(s) of interest and provide their preferences and mark as many results of interest such as mortgage rate (a large input for Rob Axtell et. al’s model (2014) for predicting housing market bubbles in Washington **[18]**) and DTI (Debt to Income), etc. The proposed interface allows users to drag and select a geographical region varying in size from a county to multiple/all states to populate results based on the filtered region.

## **4. Targeted Audience**

Potential buyers and investors: The tool can provide the risk associated with the current market and can help to identify the potential abnormality in the house market for a given area of interest.

Policy makers: To decide whether to build affordable houses or pass other measures.

## **5. Difference and Impact Upon Success**

Users will be able to avoid endless hours of market research to determine the housing market situation and will no longer need to swap between apps to make the decision. It can also be useful for policy makers to decide where to build more affordable houses and improve the housing market inequality.

## **6. Risks and Payoffs**

The risks are defining affordability, insufficient data, inability to join the data or to create meaningful results from an algorithm. The payoffs are to provide a single analytic tool to identify abnormality in house affordability and potential opportunities.

## **7. Projected Cost**

All of the data is free and provided by Zillow **[10]** and ACS **[14]**, but for practical application data costs, this will need to be assessed along with developer tool costs (i.e. website maintenance, API fees, ..).

## **8. Project Duration**

The scope of the project is paced to fit a 7 weeks. The first 2 weeks are allotted to dataset acquisition and cleaning using analytics tools such as OpenRefine and Python. Next 2.5 weeks are dedicated to model development in Python/R. The last 2.5 weeks are dedicated to data visualization and reporting duties.

## **9. Midterm and Final Checkpoints**

Midterm: Data set acquisition and preprocessing, and initial data analysis and modeling.

Final: Completed Visualization interface, project report and presentation.

## **Team members activities and contribution:**

The distribution of responsibilities is as follows:

1. Derek Cheng: Data cleaning and analytical modeling, report organization.
2. Mahaveer Jain: Data acquisition and preprocessing, analytical modeling and literature survey.
3. Micah Jeng: Data processing, visualization and report.
4. Peter Hernandez: Data collection and model development, visualization
5. Tsz Kin Chan: Tableau and data preprocessing.
6. Victor Cerabone: Tableau, Presentation and video.

Overall, all team members will contribute equally to the project.

**References**

1. Hans Lind. Price Bubbles on the Housing Market: Concept, theory and indicators. In International Journal of Housing Markets and Analysis 2(1):78-90
2. Karl E. Case and Robert J. Shiller. Is There a Bubble in the Housing Market? 2004. COWLES FOUNDATION PAPER NO. 1089.
3. Dieter Gerdesmeier, Hans-Eggert Reimers and Barbara Roffia. Early warning indicators for asset price booms. 2011. Review of Economics & Finance, Better Advances Press, Canada, vol. 1, pages 1-19, June.
4. Dean Baker, Is the Housing Bubble Collapsing? 10 Economic Indicators to Watch. 2006. In RePEc, January 2006.
5. Björn Sjöling, Indicators for Bubble Formation in Housing Markets. 2012. Master of Science Thesis.
6. Paolo Gelain, Kevin J. Lansing, Gisle J. Natvik. Explaining the Boom-Bust Cycle in the U.S. Housing Market: A Reverse-Engineering Approach. 2018. Federal Reserve Bank of San Francisco Working Paper 2015-02
7. Aaron Hedlund. [Housing Affordability: Trends, Consequences, and Policies](https://aaronhedlund.github.io/policy/CGO_housingaffordability.pdf).
8. Quan Gan and Robert J. Hill. Measuring Housing Affordability: Looking Beyond the Median. In Journal of Housing Economics, Volume 18, Issue 2, June 2009, Pages 115-125. 2009.
9. Michael E. Stone. What is housing affordability? The case for the residual income approach. 2010.
10. Zillow, Housing Data, <https://www.zillow.com/research/data/>
11. Limsombunchao, V. House price prediction: Hedonic price model vs.artificial neural network. In 2004 NZARES Conference.
12. SC Bourassa, E Cantoni, M Hoesli. Spatial dependence, housing submarkets, and house price prediction. In The Journal of Real Estate Finance, 2007 - Springer.
13. Alejandro Baldominos, Iván Blanco, Antonio José Moreno. Identifying Real Estate Opportunities using Machine Learning. In Appl. Sci. 2018, 8, 2321
14. [American Community Survey](http://www.census.gov/programs-surveys/acs/data/summary-file.html) (ACS) Survey summary file.
15. Zi Cai. [Analyzing measurements of housing affordability](https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/40314/Cai_washington_0250O_16966.pdf).
16. Paul Waddell, [Leveraging Data Science to Address Housing Affordability](https://bids.berkeley.edu/events/leveraging-data-science-address-housing-affordability)
17. Betsy Gardner. [How Data-Driven Decision Making Can Inform Affordable Housing | Data-Smart City Solutions (harvard.edu)](https://datasmart.ash.harvard.edu/news/article/how-data-driven-decision-making-can-inform-affordable-housing). 2019.
18. Rob Axtell, Doyne Farmer, John Geanakoplos, Peter Howitt. An Agent-Based Model of the Housing Market Bubble in Metropolitan Washington, D.C. 2014.