BubbleWarn: Housing Affordability & Housing Market Abnormalities Using a Multi-Data Source Approach

Derek Cheng *Mahaveer Jain* *Micah Jeng* *Peter Hernandez* *Victor Cerabone*

Summary

BubbleWarn is a visual analytics system that combines predictive model algorithms with fundamental housing market factors to provide insights into housing market volatilities, risks, and rewards. The system combines user interaction, a predictive model, and Tableau visualization. The user can interactively explore various affordable housing markets and determine the market condition using two very intuitive ratios.

Practical Insights

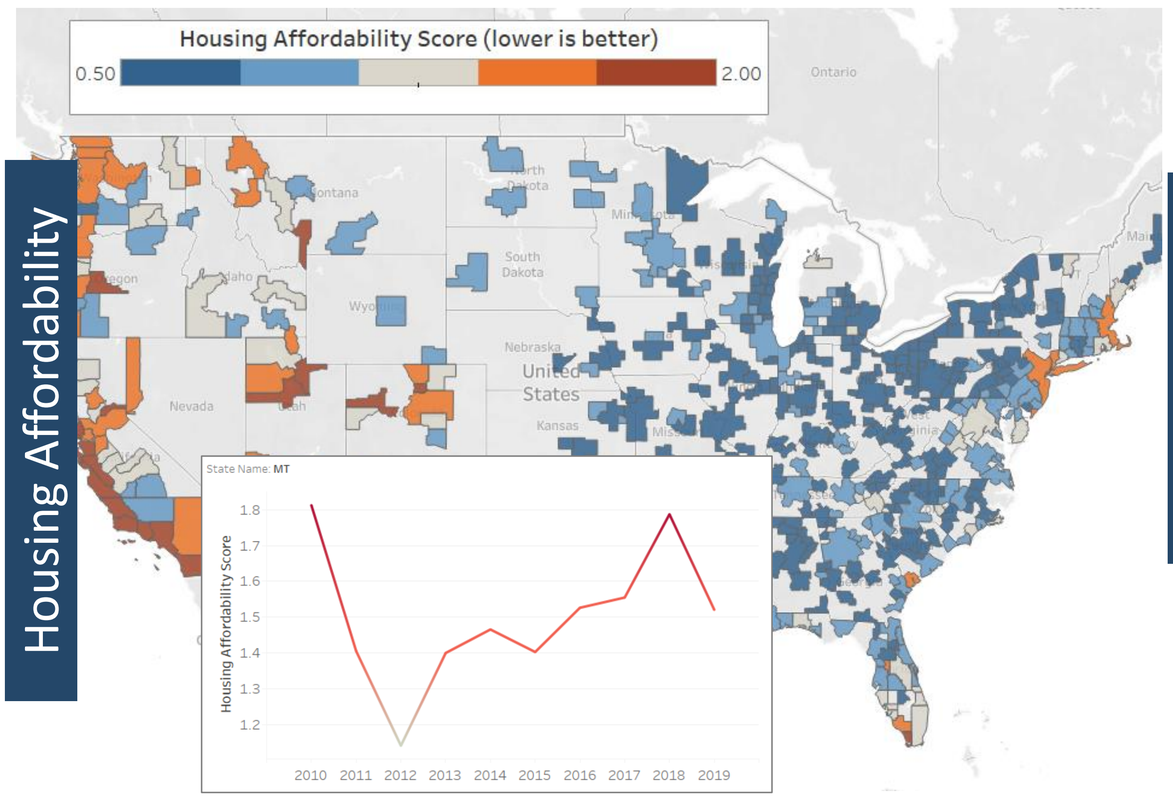
BubbleWarn provides data driven insight into the current housing market to potential home buyers, investors, and policy makers by combining various predictive variables and public sentiment through Google search index. BubbleWarn also provides housing affordability scores by occupation for the same audience. The user can select different region, time, debt to income ratio and other relevant attributes of interest to visualize the trend.

Multi-Factor Models

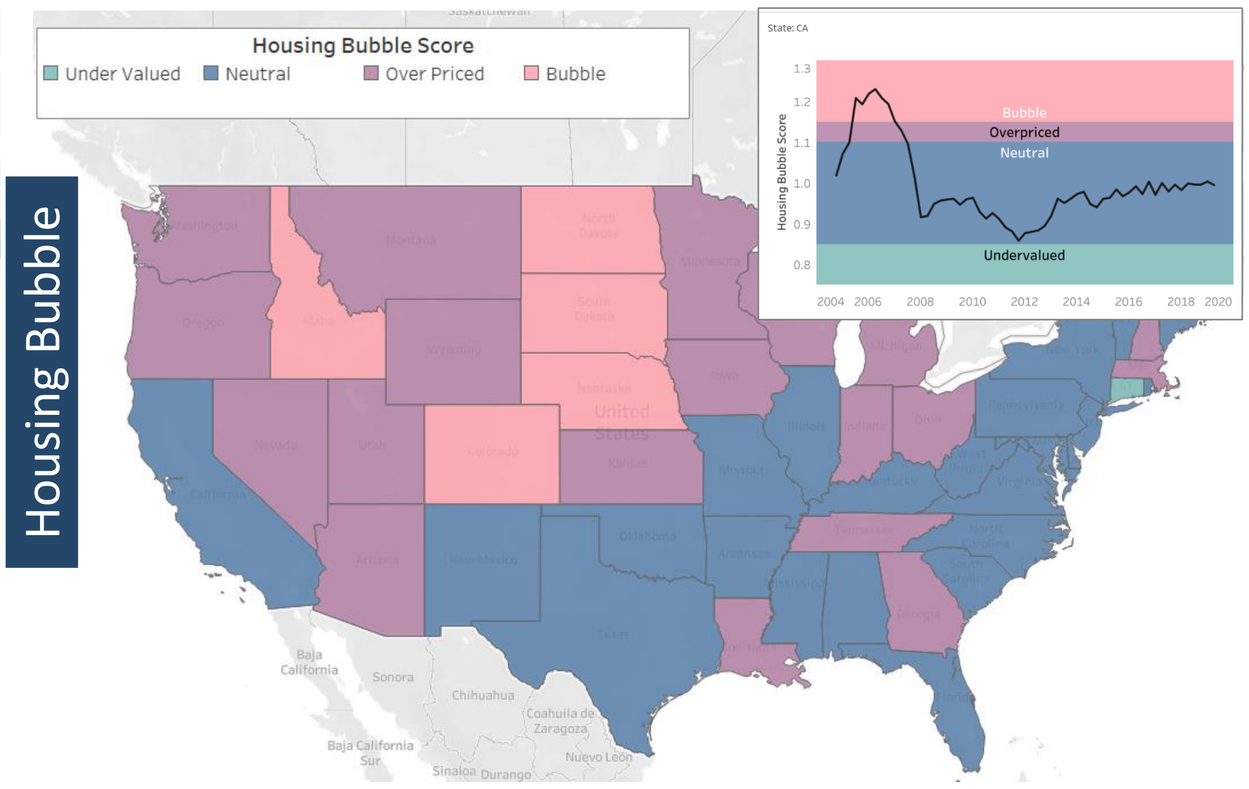
In our analysis, we used a multilinear regression model with 20+ factors in predicting housing price index and housing affordability ratio from 2005 to 2019. To capture the impact of public sentiment on the housing market, we integrated real estate related search terms from Google Trends and enhanced the model's predictive power. The predictive variables are carefully selected using ElasticNet regularization.

Interactive Visuals Provide Insights by Region: tinyurl.com/BubbleWarn

Housing Affordability



Housing Bubble



Housing Affordability Index & Data Sources

For a given down payment ratio β, debt to income ratio α, N year loan terms, Housing Affordability Index (HAI) for an occupation at time t is defined as:

$$HAI_t = HTI_t / AL_t$$

$$HTI_t = \left(\frac{ZHVI_t}{INC_t}\right), \qquad AL_t = \left(\frac{a}{1-\beta}\right)\left(\frac{1-(1+iR_t)^{-N}}{iR_t}\right)$$

Factor	Name	Access Method	File size
ZHVI	Zillow Housing value Index	Download	224KB
INC	Median income by occupation	Download	989KB
iR	30 years mortgage rate	Download	48KB

Housing Bubble Algorithm & Data Sources

The predicted Housing Price Index based on multi-predictive variables is represented by:

$$\hat{HPI}_{t,s} = \beta_0 + \beta_1 UR_{s,t} + \beta_2 PO_{s,t} + \beta_3 CCI_t + \beta_4 HPA_{s,t} + \beta_5 Factor(s) + \beta_{6\ldots N} GSVI(X)_{s,t}$$

GSVI(X) = Apartment, Real Estate, Zillow, Rent Query

Factor	Name	Access Method	File size
UR	Unemployment rate	API	207KB
PO	Population	API	229KB
CCI	Consumer Confidence Index	Download	654KB
HPA	# Housing permits authorized	Download	475KB
GSVI	Google search volume index	API	750KB

Experiments and Results

The BubbleWarn model is validated against 2006-07 U.S. housing market crisis data, where it has successfully identified the set of all benchmark states [9 out of 9 states] as being in a bubble period ahead of crisis by 4 quarters by combining public sentiments along with other predictive variables. The model performance further demonstrates it can estimate minimum support house price index and allows the user to make data driven decision. BubbleWarn's affordability metric provides further insight into the affordability across occupation & region and will allow the user to study the market at macro level and understand the affordability across regions. Combined, both metrics can allow the user to narrow down the time and region to invest in a housing market.