

Topic

With the market in a current state of disarray, many homeowners are concerned about their homes and whether they will soon be underwater in their investment. We will be analyzing historical and current housing market data to determine whether a significant drop in pricing is potentially incoming.

Business Problem

Businesses and individuals alike invest in housing to provide for their family, or to rent out their property to gain residual income. According to the Internal Revenue Service, in 2018 approximately 10.3 million individual filers reported they were owners of rental properties while owning 1.72 properties on average (Pew Research Center, 2021). Throughout this project, we will attempt to discover whether these owners should be considering selling their investments, whether their homes will be massively underwater, and whether the housing market will remain stable for existing and new investors. We will also be attempting to gain insight for millions of homeowners of whom may be worried during these tumultuous times.

Datasets

The datasets used will be sourced from Kaggle — a massive dataset supplier — and the Federal Housing Finance Agency (FHFA) — an independent housing agency for the United States.

- HPI_master (FHFA): <https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx>
 - This data provides insights for the USA based on historical price indexes to understand the previous patterns of the housing market.

- Housing (Kaggle): <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>
 - This data provides information related to homes along with their area code to assess the current price for each area for each home.
- State_statistics_for_download (FHFA): <https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx>
 - This dataset contains additional information based on the state to show the average price per state per year/period.

Methods

Since the housing data provided by the FHFA is set in a time series, I intend to use a time series model consisting of ARIMA or LSTM for potential future forecasting. The main concept is to create an expected Home Price Index (HPI) for future performance by state to determine an approximation for what future quarters or months may experience. Alternatively, since the Kaggle housing data contains more features with categorical and continuous data, Random Forest Regression may be the optimal choice of model for this particular dataset. It is also likely that no machine learning model may be required with this data. With the understanding that both datasets are unique with their respective coverage, it's likely that the two model results will need to be compared against one another to determine significance. The state statistics dataset will be primarily used as a reference point for the HPI data to ensure physical amounts are identified, as opposed to percentage estimations.

Ethical Considerations

The main ethical concern regarding these datasets involves privacy. Based on the specifics that describe the home in the Kaggle dataset, it's possible that users may be able to identify the home based on the location. However, it's intended that we group the data by area to ensure that this provides no concern. The datasets used within this project provide no additional ethical considerations beyond the one stated.

Challenges/Issues

The data preparation may prove challenging to isolate the features, create dummies, and group them based on area for the Kaggle dataset; the main potential challenge here is that there may not be enough locations with similar features for a considerable average. Additionally, the comparison required at the end may not provide enough locations to get a full scope of the entire US and what areas may be affected the most.

References

- Federal Housing Finance Agency. (2022). House Price Index Datasets [Data set]. Retrieved from <https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx>
- Parker, K., & Stepler, R. (2021, August 2). As national eviction ban expires, a look at who rents and who owns in the U.S. Pew Research Center. <https://www.pewresearch.org/fact-tank/2021/08/02/as-national-eviction-ban-expires-a-look-at-who-rents-and-who-owns-in-the-u-s/>
- Yasser, H. (2020). Housing Prices Dataset. Kaggle. Retrieved February 14, 2023, from <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>