

Housing Data Analysis in the US

By Sangram Sanjiv Malekar, Cecilia Chen, and Minyan Gao

Background and Motivation

Owning a home offers the benefits of having a secure place to reside in the long term, the freedom to make renovations, and potential for rental income. Buying a property is an important decision and understanding the real estate market conditions can greatly impact the success of the purchase. The market experiences fluctuations between buyer's and seller's market, each with advantages and challenges for homebuyers. In articles, such as *Buyer's Market Vs. Seller's Market: What Does Each Mean For You, And What Is The Current Market State*, the importance of understanding when these markets occur and how they impact potential homebuyers is discussed. Detecting and predicting when a buyer's market is likely to occur helps buyer's optimize the purchasing power and secure their dream home.

We will be using data analysis techniques to analyze Zillow's housing data, mortgage rates from FreddieMac, and immigration data from the Department of Homeland Security (DHS) to gain and provide insights into the market conditions and identify indicators of buyers' market. This will help buyers navigate the market fluctuations and avoid badly timed investments as it could lead to severe consequences. The Housing Market Crash 2008 happened when many homeowners who had initially taken out adjustable-rate mortgages (ARMs) with introductory interest rates, only to face issues when those rates were later adjusted to higher rates. This made it difficult for them to afford their monthly mortgage payments which contributed to the 2008 crash. The article *Price-to-Rent Ratio: Determining if It's Better To Buy or Rent* explores how a notable indicator that acts as a warning sign of the 2008 crash was the surge in the price-to-rent ratio. This ratio is calculated by dividing the median home price by the median annual rent and assesses if it's cheaper to rent or buy a property in a given market.

Questions:

1. Can proxies for buyer's/seller's market be used to predict home prices?
2. Are there specific mortgage rates thresholds or trends that indicate a shift towards a buyer's or seller's market?
3. Can mortgage rates be used to predict home prices?
4. Is it worth competing with multiple offers when purchasing a property for rental purposes?
5. How might immigration impact the real estate market?
6. Which indicators in Zillow's housing data are most useful for determining whether the current market favors buyers or sellers?



Data Sources - Primary

Name	Description	Estimated Size	Access
Zillow's Housing Data	Zillow's housing data is a vast dataset that includes dozens of metrics related to the housing market, available at both the national and neighborhood levels. For our project, we will be utilizing several datasets, including those pertaining to home values, rentals, inventory, list and sale prices, as well as sales count and price cuts.	<p>9.35 MB in total size, made up of 15 csv that varies in time from early 2000s to April 2023.</p> <ul style="list-style-type: none">● Home price sales<ul style="list-style-type: none">○ Median List Price, Median Sale Price (Raw, All Homes, Monthly)○ Home Value Index (ZHVI) All Homes (SFR, Condo/Co-op) Time Series, Smoothed, Seasonally Adjusted(\$)● Price-to-Rent Ratio<ul style="list-style-type: none">○ Real estate inventory -- For-Sale Inventory, New Listings, Newly Pending Listings, Mean/Median Days to Pending (Using Raw, All Homes, Monthly)○ Rentals -- ZORI (Smoothed): All Homes Plus Multifamily Time Series (\$)● Sales Count and Price Cuts<ul style="list-style-type: none">○ Sales Count Nowcast (Raw, All Homes)○ Share of Listings with a Price Cut (Raw, All Homes, Monthly)○ Median Price Cut (\$, Raw, All Homes, Monthly)● List<ul style="list-style-type: none">○ Percent of Homes Sold Above/Below List (Raw, All Homes, Monthly View)○ Median Days to Close (Raw, All Homes, Monthly View)	<p>Zillow Housing Data (CSV Download)</p>

Data Sources - Secondary

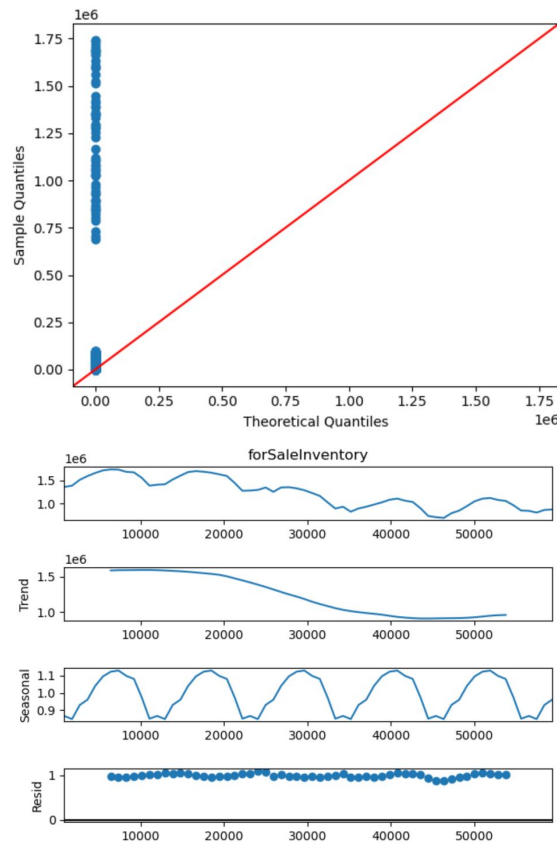
Name	Description	Estimated Size	Access
FreddieMac's Mortgage Rates Data	Weekly mortgage rates data is provided by FreddieMac from 1971 to 2023, who surveys lenders on the rate and points for their most popular 30-year fixed-rate, 15-year fixed-rate and 5/1 hybrid amortizing adjustable-rate mortgage products since April 1971. As of November 17, 2022, the survey results are based on actual applications from lenders across the country that are submitted to Freddie Mac when a borrower applies for a mortgage.	167 KB	FreddieMac Mortgage Rates Inch Down (xlsx. Download)
The 2021 Yearbook of Immigration Statistics	The 2021 Yearbook of Immigration Statistics provided by the Department of Homeland Security contains a wide range of tables that offer detailed data on various aspects of immigration. For our analysis, we will be focusing on Table 4, which provides a count of individuals with Lawful Permanent Resident Status by State or Territory of Residence for Fiscal Years 2012 to 2021.	19 KB	Yearbook of Immigration Statistics 2021 (dhs.gov) (xlsx. Download)
Dates of U.S. recessions as inferred by GDP-based recession indicator	The series assigns dates to U.S. recessions based on a mathematical model of the way that recessions differ from expansions. The data span is from 1967 to 2022.	15.5 KB	Recession Index (FRED) (xls. Download)

Data Manipulation Methods

Given that the datasets are based on time series, the code leverages the temporal information present in the data to estimate values that are missing between observed data points. The `interpolate(method='time')` method in Pandas is used for filling missing values, allowing for a more precise representation of the underlying patterns and ensuring that the filled values align with the temporal characteristics of the data. The datasets are structured in a wide format, with dates represented as columns. To convert this wide format to a long format, Pandas `melt()` function is employed. The resulting unpivoted data has long format where the dates are values in a single column.

The housing metrics analyzed exhibit a strong trend and seasonality, indicating significant patterns over time. However, these metrics are not normally distributed, as revealed by the QQ Plots. For example, while analyzing For Sale Inventory metric, statsmodels seasonal_decompose API is used to analyze the trend and seasonality of the national-level data. The analysis reveals a decreasing trend and significant seasonality. Further, the QQ plot does not exhibit the expected alignment of quantiles with the theoretical quantiles when utilizing the statsmodels qqplot API.

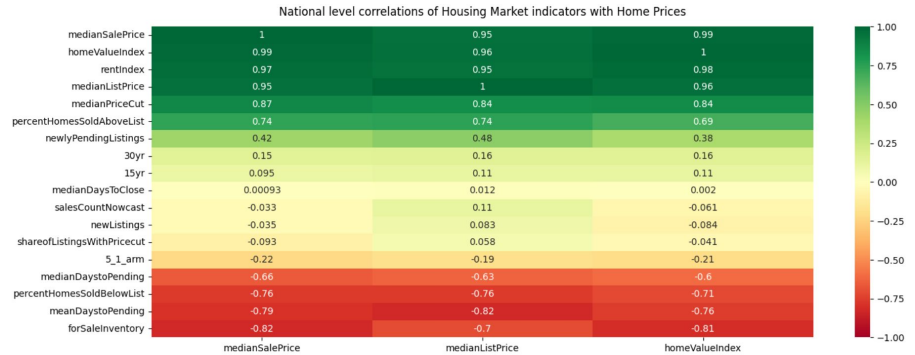
To create a comprehensive dataframe, multiple datasets are merged using inner joins. First, the Zillow datasets are merged based on the region and date columns. Next, the Mortgage rate dataset is aggregated by computing the mean on a monthly basis, and this aggregated dataset is joined with the previous dataframe using the month as the key. Finally, the resulting dataframe is further aggregated by computing the mean on a yearly and state basis, and it is joined with the immigration dataset based on the year and state columns.



Can proxies for buyer's/seller's market be used to predict home prices?

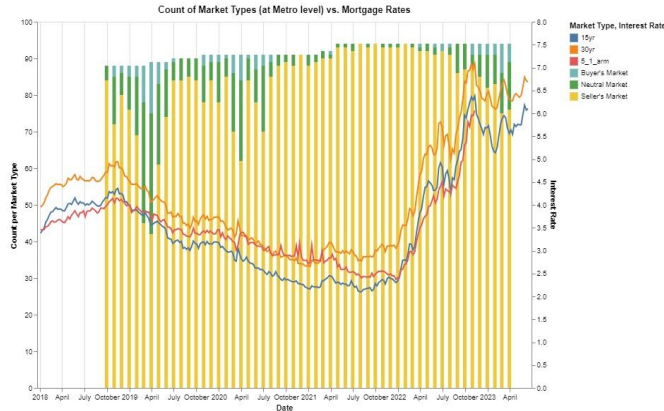
The correlation analysis uses the seaborn heatmap() and pandas corr() APIs. Spearman coefficient reveals relationships between home price metrics (MedianSalePrice, MedianListPrice, HomeValueIndex) and other variables. The spearman coefficient was chosen because all of the housing data used here are not normally distributed.

Positive correlations were observed with following:	Conversely, negative correlations were found with following:
<ol style="list-style-type: none"> 1. Zillow Observed Rent Index. <ul style="list-style-type: none"> - Higher rental prices tend to be associated with higher home prices. 2. Price cuts. <ul style="list-style-type: none"> - Larger price cuts are associated with higher home prices. 3. Percentage of homes sold above the list price. <ul style="list-style-type: none"> - A higher proportion of homes selling above the list price is associated with higher home prices. 	<ol style="list-style-type: none"> 1. Percent of Homes Sold Below List. <ul style="list-style-type: none"> - A higher proportion of homes selling below the list price is associated with lower home prices. 2. Days to Pending. <ul style="list-style-type: none"> - Longer time on the market is associated with lower home prices. 3. For Sale Inventory. <ul style="list-style-type: none"> - It suggests that a higher supply of homes on the market may be associated with lower home prices.



These proxies provide an indication of the overall supply and demand dynamics within the real estate market conditions, which in turn influences home prices. However, it's also important to note that while these proxies provide valuable insights into the market conditions, they are not the only determinants of home prices.

Are there specific mortgage rates thresholds or trends that indicate a shift towards a buyer's or seller's market?



On August 26, 2021, mortgage rates reached their lowest levels:

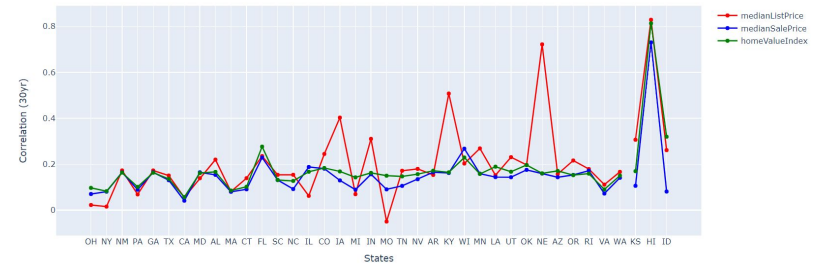
- 30-year Fixed-Rate Mortgage (FRM) was at 2.87%.
- 15-year FRM was at 2.17%.
- 5/1 Adjustable Rate Mortgage (ARM) was at 2.42%.

Additionally, during the month of August 2021, out of the 94 Metropolitan statistical areas analyzed, all of them were categorized as seller's markets, with no regions classified as buyer's markets or neutral markets.

There is a loosely inverse relationship between mortgage rates and the prevalence of seller's markets.

Can mortgage rates be used to predict home prices?

Correlation for 30yr vs medianListPrice, medianSalePrice, homeValueIndex



There's typically an inverse relationship between interest rates and house prices. When mortgage rates are lower, buyers find it more affordable to purchase a home, which leads to increased demand and drives up home prices. However, this is more of a guideline than a hard rule which is evident from the analysis that that is not always the case.

The spearman correlation (since data is not normally distributed) between the 30-year fixed-rate mortgage (FRM) and home price metrics varies among states, but the correlations are mostly positive with only Missouri showing a negative correlation between 30-year FRM and the Median List Price. A positive spearman correlation here means that as the 30-year fixed-rate mortgage (FRM) increases, Median List Price, Median Sale Price, and Home Value Index also increases.

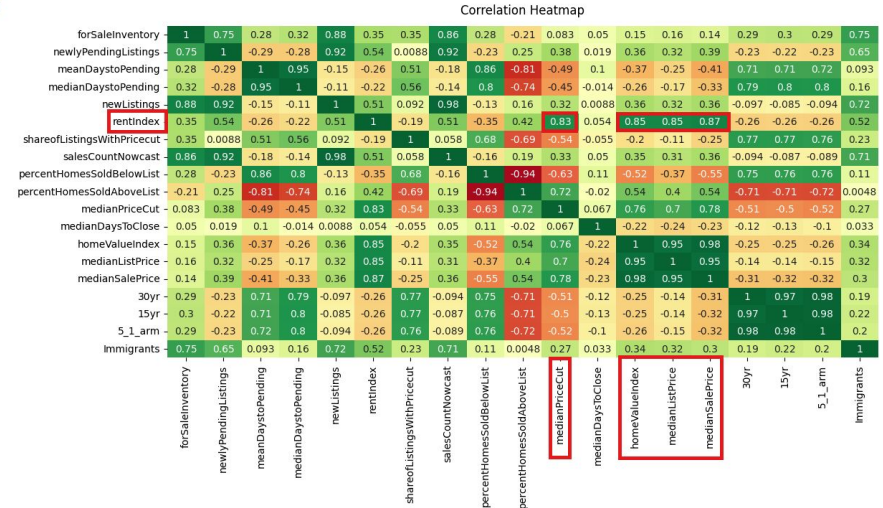
Is it worth competing with multiple offers when purchasing a property for rental purposes?

The rent index shows a positive correlation with median price cuts, home value index, median listing price, and median sale price. The correlation was calculated by using Spearman's correlation because the relationship is not expected to be strictly linear.

When the price-to-rent ratio is below 15, it is generally more advantageous to buy rather than rent. On the other hand, when the ratio exceeds 15, it is usually more favorable to rent. However, the analysis suggests that renting is often the recommended option based on the ratio.

When it comes to an investment, a higher demand for rentals can lead to a decrease in the price-to-rent ratio, which can potentially increase potential profits. In such cases, it may be worth competing with multiple offers by offering a higher purchase price, as long as the ratio remains reasonable.

When the ratio is very high, remains persistently high, it could potentially indicate that housing markets are in a bubble. In such situations, caution should be exercised before making investment decisions.



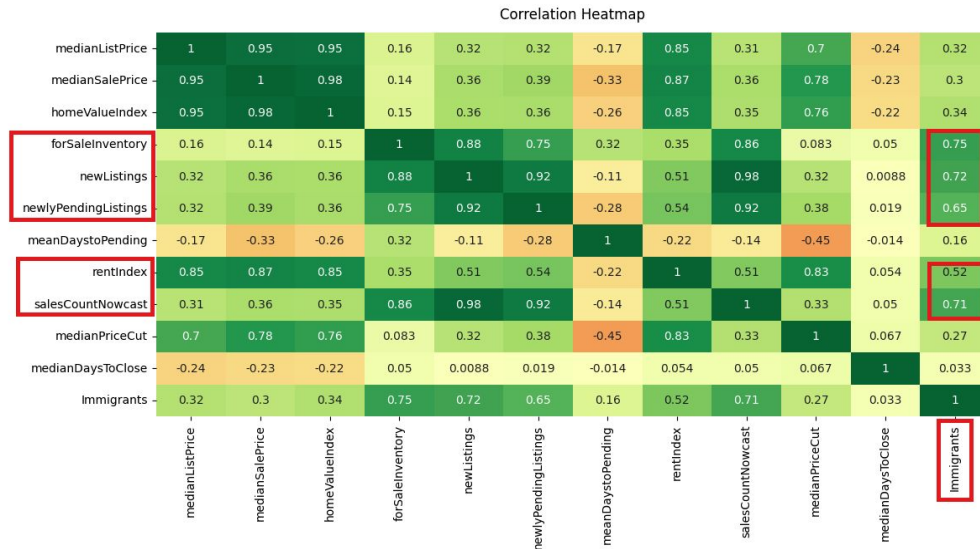
$$ratio = \frac{\{\text{Median Home Price}\}}{\{\text{Median Annual Rent}\}}$$

How might immigration impact the real estate market?

The correlation between immigrant counts with various factors in the real estate market was calculated by using Spearman's correlation because the relationship is not expected to be strictly linear. Immigrant counts has a robust positive relationship with various factors: sale inventory, new listings, newly pending listings, and sales count nowcast. This means the higher the immigrant counts, the more the demand for housing increases.

This indicates that an increase in the number of immigrants corresponds to a more active real estate market as evident from the positive correlations between immigrant counts and the various factors in the real estate market. Moreover, immigrants also have a positive impact on the rental market, as evidenced by a correlation coefficient of 0.52.

American real estate market is fuelled by immigration.

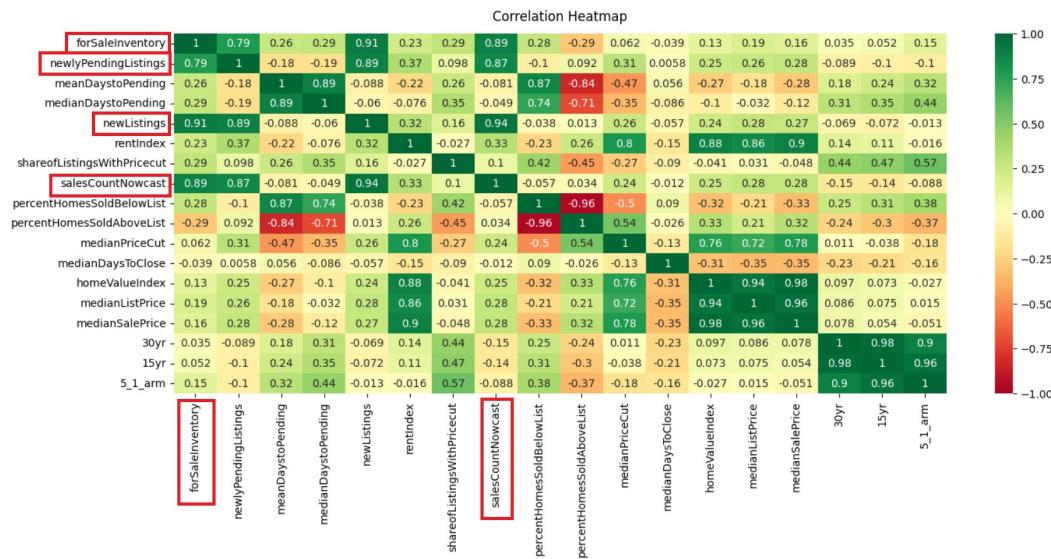


Which indicators in Zillow's housing data are most useful for determining whether the current market favors buyers or sellers?

Indicators for determining whether the current market favors buyers or sellers:

- For Sale Inventory
- Newly Pending Listings
- New Listings
- Sale Count Nowcast (estimated number of unique properties that sold during the month)

The market condition can be assessed based on the ratio of current home inventory to the number of homes sold in the previous month. Various factors, such as new listings, newly pending listings, sales count nowcast, and sale inventory, exhibit a strong correlation with each other based on the correlation heatmap. These data can serve as valuable indicators for predicting whether the current market is favorable for sellers or buyers.



Market Condition Formula

$$ratio = \frac{\{\text{Current Home Inventory}\}}{\{\text{Amount of Homes Sold in the Last Month}\}}$$

If ratio > 7, buyer's market;

If ratio < 5, seller's market;

If 5 <= ratio <= 7, neutral market

Summary and Next Steps

- Zillow's datasets, including the rent index, price cuts, and the percentage of homes sold above the list price, exhibit strong positive correlations with home prices. Conversely, other datasets such as the percent of homes sold below the list price, days to pending, and for sale inventory display significantly low negative correlations. These datasets can be effectively utilized for predicting home prices.
- For Sale Inventory, Newly Pending Listings, New Listings, and Sale Count Nowcast can be utilized for detecting real estate market type.
- In general, the current market conditions tend to favor sellers, with limited opportunities for buyers. The prevailing trend suggests that renting is often a more viable option than purchasing a property.
- However, a higher rental demand can lead to a reduction in the price-to-rent ratio, thereby increasing potential profits if purchasing a property is for investment purposes.

- The prevalence of seller's markets generally exhibits a loose inverse relationship with mortgage rates. Mortgage rates can serve as a predictive factor for home prices in certain states.

Future enhancements:

- Incorporate p-values for assessing statistical significance and explore alternative interpolation methods for handling missing values.
- Consider incorporating other macroeconomic datasets, such as inflation rates, recessions, and pandemic data, as independent variables in the analysis.
- Include housing data from additional sources like Redfin to expand the dataset and capture a longer time period, reducing biases in the analysis.
- Perform PCA (Principal Component Analysis) to reduce dimensionality and utilize it for developing housing market forecasting models.



Statement of Work and References

References

- [Price-to-Rent Ratio: Determining if It's Better To Buy or Rent](#)
- [Buyer's Market Vs. Seller's Market: What Does Each Mean For You. And What Is The Current Market State?](#)
- [Housing Market Crash 2008 Explained: Causes & Effects](#)
- [Interest Rate Vs. Home Price: Which Matters More?](#)
- [Real Estate Marketing Plan Presentation \(Slide Template\)](#)

Statement of Work:

We collaborated on the multivariate analysis by utilizing Google Colab and its comment features to keep notes. For future team projects, adopting Python script files for the development of Python modules would enhance efficiency. This approach enables team members to easily import and utilize the public methods within their own Python projects, promoting collaboration and code reusability.

All: Report editing, peer reviews, project goal formulation, dataset consolidation

Cecilia Chen: Univariate analysis on the datasets of immigration, mortgage, home value index, and median sale price. Multivariate analysis was conducted to examine the real estate market type, the relationship between rental price and property sale price, as well as the influence of immigration policies on the real estate market.

Minyan Gao: Univariate analysis on the sales count and price cuts data sets and list data sets from Zillow. Multivariate analysis was conducted to explore the relationships between mortgage rates and home prices to see if those rates can be used to predict home prices, as well as exploring how immigration impacts the real estate market.

Sangram Sanjiv Malekar: Univariate analysis on the Zillow Real estate inventory, Rentals data sets. Multivariate analysis to explore Market Conditions Across Years, Impact of market proxies to predict home prices, Mortgage rates impact on market conditions.