Housing Affordability Analysis: Rent-to-Price Trends in the United States

(COMP3125 Individual Project)

Nicholas Souza
Data Science Fundamentals

***Abstract*— This project analyzes housing affordability trends in the United States from 2015 to 2024 using rent and housing price datasets for single-family homes. The rent-to-price ratio was employed as a key affordability metric. The analysis highlights a decline in affordability from 2015 to 2020, followed by a sharp dip during the COVID-19 pandemic, and a gradual recovery in the subsequent years. Regional disparities were also observed, with some markets showing significantly higher rent-to-price ratios than others. This research provides valuable insights into the dynamics of the U.S. housing market.**

Keywords*— *housing affordability, rent-to-price ratio, single-family homes, U.S. housing market, COVID-19 impact

I. INTRODUCTION (*HEADING 1*)

Housing affordability has become a critical issue in the United States, affecting the financial well-being and quality of life for millions of individuals and families. The topic is particularly important as it reflects the balance between income levels, rent prices, and housing costs, which directly influence housing accessibility. This project focuses on analyzing housing affordability trends over the past decade (2015–2024), specifically by exploring the relationship between rent prices and housing prices for single-family homes.

The rent-to-price ratio is used as a key indicator to assess affordability, as it reflects the cost of renting relative to owning a home. Current research shows that the U.S. housing market has experienced significant changes in recent years, driven by factors such as the COVID-19 pandemic, fluctuating interest rates, and shifts in demand for housing. These events have impacted affordability, with many markets seeing surging housing prices and rents, particularly during the pandemic years.

This analysis contributes to the understanding of housing market dynamics by highlighting national trends and regional disparities in affordability. By leveraging datasets from Zillow's Housing Value Index (ZHVI) and Observed Rent Index (ZORI), this project uncovers critical insights that can inform future policy discussions and market strategies.

II. DATASETS

A. *Source of dataset (Heading 2)*

The datasets used in this project were sourced from Zillow's publicly available research data. The two datasets are:

Zillow Home Value Index (ZHVI):

This dataset provides monthly median home prices for single-family homes across various metro areas in the United States. It includes historical data from 2000 to 2024.

Zillow Observed Rent Index (ZORI):

This dataset contains monthly median rent prices for single-family homes in the same metro areas. It spans from 2015 to 2024.

Zillow is a widely recognized and credible source for housing market data, regularly used by researchers, policymakers, and industry professionals. The datasets are generated by Zillow using a combination of property listing data, public records, and machine learning models to estimate market trends accurately.

Both datasets were downloaded directly from the Zillow Research website. The inclusion of extensive historical data makes them ideal for analyzing long-term trends in housing affordability.

B. *Character of the datasets*

The datasets used in this project consist of time-series data with specific formats, sizes, and characteristics. The Zillow Home Value Index (ZHVI) dataset, which provides housing prices, is in CSV format and has a size of approximately 4.1 MB. It includes data for 920 regions (e.g., metro areas) and contains 303 columns, including metadata and monthly home price values from 2000 to 2024. The Zillow Observed Rent Index (ZORI) dataset, which provides rent prices, is also in CSV format with a size of about 0.6 MB. This dataset includes data for the same 920 regions and has 123 columns, including metadata and monthly rent price values from 2015 to 2024.

Both datasets share common parameters. Key columns include RegionID, which serves as a unique identifier for each metro area or region; RegionName, which specifies the name of the metro area (e.g., "Los Angeles, CA"); and StateName, which indicates the corresponding U.S. state. The monthly columns in both datasets provide time-series data with values for each month (e.g., "2024-01") representing housing or rent prices in U.S. dollars.

To prepare the data for analysis, missing data was handled by excluding rows with insufficient values for key time periods. The two datasets were combined by merging on the RegionName and StateName columns, ensuring alignment between housing and rent prices for the same regions. Additionally, a new metric called RentToPriceRatio was created to analyze housing affordability trends. This metric was calculated by dividing the rent price by the housing price for each region and time period.

III. Methodology

A. Method A. Rent-to-Price Ratio Calculation

Example: The primary method used in this project was the calculation of the Rent-to-Price Ratio, a metric designed to measure housing affordability. This ratio is defined as RentToPriceRatio which is equal to rent price divided by housing price.

The assumption of this method is that the rent-to-price ratio reflects the relative affordability of renting compared to owning a home. A higher ratio indicates that renting is relatively more expensive compared to purchasing a home, while a lower ratio suggests the opposite.

The advantages of this method include its simplicity and its ability to provide an intuitive measure of affordability. However, a limitation is that it does not account for external factors such as household income, mortgage rates, or property taxes, which can also influence affordability.

This method was chosen because it directly addresses the research question of how housing affordability has changed over time and across regions. The Python libraries Pandas and NumPy were used to calculate the rent-to-price ratio for each metro area and month.

B. Method B. Time-Series Analysis

To identify affordability trends over time, the rent-to-price ratio was calculated for each month from 2015 to 2024. The **assumption** in this method is that the time-series data reflects real market trends over the specified period.

The **advantages** of this approach include the ability to observe long-term patterns and the impact of specific events (e.g., the COVID-19 pandemic). A disadvantage is that missing data for some regions or months can slightly limit the completeness of the trend analysis.

The Python libraries **Matplotlib** and **Pandas** were used to visualize the trends. Optional input adjustments included aligning the datasets to ensure that only common time periods were included in the analysis.

C. Method C. Regional Analysis

To analyze regional disparities, scatter plots were created to compare rent and housing prices for a specific time period (e.g., October 2024). The assumption of this method is that the scatter plot accurately represents the relationship between rent prices and housing prices across metro areas.

The advantages of this method include its ability to highlight outliers and its simplicity in visualizing the relationship between the two variables. The main disadvantage is that it provides only a snapshot in time and does not reflect dynamic changes.

The scatter plots were generated using the Matplotlib library in Python. This visualization revealed a positive correlation between rent prices and housing prices, with notable regional outliers.

IV. Results

The average rent-to-price ratio was calculated for each month from 2015 to 2024 to analyze national affordability trends. The results, visualized in a line chart, show three distinct phases:

2015–2020: A gradual decline in the rent-to-price ratio, indicating that housing prices were increasing faster than rent prices during this period.

2020–2022: A sharp dip in affordability during the COVID-19 pandemic, which coincides with a housing price surge driven by low interest rates and increased demand.

2022–2024: A gradual recovery in affordability as rent prices started to catch up with housing prices.

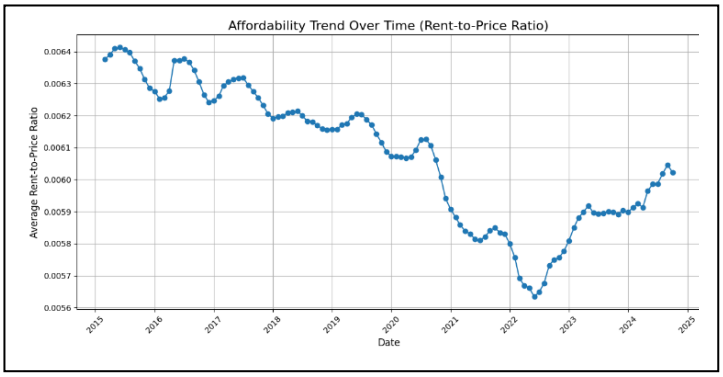


Figure 1 shows the national affordability trend from 2015 to 2024.

The results confirm the impact of significant economic events, such as the pandemic, on housing market dynamics. The steady increase in affordability after 2022 suggests a rebalancing of the market.

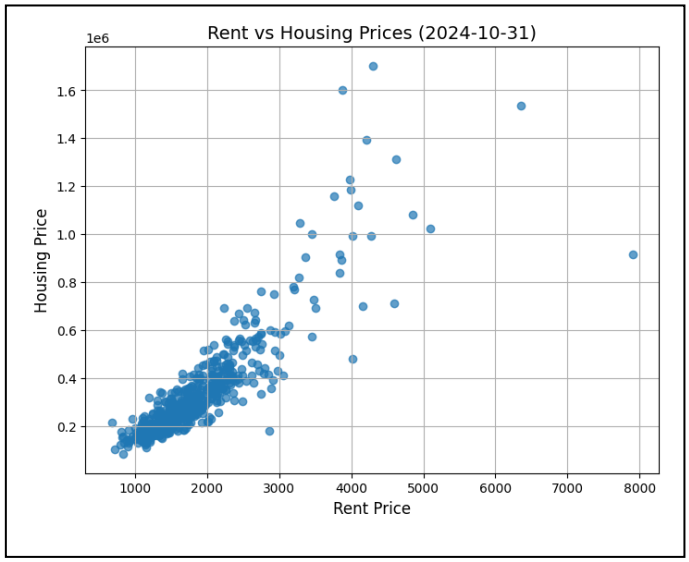
A. Result A. Regional Disparities

The scatter plot of rent prices versus housing prices for October 2024 highlights significant regional disparities. A positive correlation is observed, where regions with higher housing prices also tend to have higher rent prices. However, there are notable outliers:

Regions with High Rent-to-Price Ratios: Binghamton, NY, and Decatur, IL, show relatively high ratios, indicating that renting is relatively expensive compared to owning a home.

Regions with Low Rent-to-Price Ratios: Jackson, WY, and San Jose, CA, have low ratios, suggesting that homeownership is significantly more expensive relative to renting.

These findings highlight that affordability varies widely depending on the region, influenced by local market conditions and economic factors.



B. Results B. Unexpected Results

While most regions followed the expected trend of a positive correlation between rent and housing prices, a few anomalies were noted:

Flat or Negative Rent Growth: Some regions experienced stagnant rent prices despite rising housing prices, likely due to specific local factors such as declining demand or economic shifts.

Regions with Missing Data: A few metro areas had incomplete data for specific time periods, which may have slightly impacted the completeness of the analysis.

These anomalies warrant further investigation to understand the underlying causes.

V. Discussion

A. Limitations of the Analysis

While this project provides valuable insights into housing affordability trends in the United States, several limitations were identified:

Exclusion of Additional Factors:

The analysis focuses solely on rent and housing prices, without considering other factors such as household income, mortgage rates, or property taxes. These factors play a significant role in determining overall affordability.

Missing or Incomplete Data:

Some regions had missing or incomplete data for specific time periods. This may have slightly impacted on the accuracy of the affordability trend analysis and regional comparisons.

Static Snapshot for Regional Analysis:

The regional comparison relies on a single time point (October 2024). While this highlights disparities at a specific moment, it does not account for changes over time in individual regions.

Rent-to-Price Ratio Simplification:

The rent-to-price ratio provides a straightforward measure of affordability but does not account for the benefits or costs associated with homeownership, such as equity-building, maintenance costs, or rental market volatility.

B. Suggestions for Future Work

To address these limitations and enhance the analysis, the following improvements are proposed:

Incorporate Additional Variables:

Future studies should integrate factors such as income levels, interest rates, and demographic data to provide a more comprehensive view of affordability.

Improve Data Completeness:

Efforts should be made to fill gaps in the datasets by supplementing missing data with alternative sources or predictive modeling techniques.

Expand Regional Analysis:

Conduct time-series analyses for individual regions to identify trends and anomalies in affordability over time. This would help uncover unique regional dynamics.

Apply Advanced Statistical Methods:

Using machine learning or econometric models could provide deeper insights into the relationships between rent, housing prices, and external factors. For example, regression analysis could quantify the impact of specific variables on affordability.

Policy Implications:

The results of this analysis could be linked to potential policy recommendations, such as increasing affordable housing initiatives or revising zoning laws to address regional disparities.

By addressing these limitations and incorporating the suggested improvements, future research can provide a more nuanced understanding of housing affordability trends and inform effective policymaking.

VI. Conclusion

This project analyzed housing affordability trends in the United States from 2015 to 2024 using rent and housing price data for single-family homes. The rent-to-price ratio was employed as a key metric to measure affordability. The analysis revealed three distinct phases in national affordability trends: a gradual decline from 2015 to 2020, a sharp dip during the COVID-19 pandemic (2020–2022), and a steady recovery from 2022 onward.

The regional analysis highlighted significant disparities in affordability, with regions like Binghamton, NY, exhibiting high rent-to-price ratios and markets like Jackson, WY, showing low ratios. These findings underscore the varying economic conditions and market dynamics across the United States.

The results of this project have real-world implications for renters, homeowners, and policymakers. Understanding these trends can inform housing policies, guide market strategies, and provide individuals with insights into the affordability of renting versus owning a home in different regions. Future research incorporating additional economic variables and advanced methods could further enhance the understanding of housing affordability.

ACKNOWLEDGMENT *(Heading 5)*

The author thanks Zillow Research for providing the publicly available data used in this analysis. Gratitude is also extended to the instructor of the data science course for their guidance and support throughout this project.

REFERENCES

[1] Zillow Research, "Zillow Home Value Index (ZHVI)," Available: https://www.zillow.com/research/data/. [Accessed: Dec. 9, 2024].

[2] Zillow Research, "Zillow Observed Rent Index (ZORI)," Available: https://www.zillow.com/research/data/. [Accessed: Dec. 9, 2024].

[3] G. VanderPlas, Python Data Science Handbook. Sebastopol, CA: O'Reilly Media, 2016.

[4] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[5] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference*, Austin, TX, USA, 2010, pp. 51–56.