Methods of Advanced Data Engineering

Project analysis report

MD Nur Hossain Robin

Matrikel Number: 23020555

Housing Affordability in the United States: Analyzing Regional Disparities and Income-to-Price Correlations.

Source:The American House Prices Dataset by Jeremy Larcher and the Average Income per ZIP Code USA Dataset by Hamish Gunasekara

1 Introduction

Housing affordability has become a growing concern in the United States as housing prices continue to rise faster than incomes. For many Americans, the dream of home ownership is slipping out of reach, creating a widening gap in socioeconomic equity. This report takes a closer look at how affordable housing is across different U.S. regions by examining the relationship between income levels and housing prices at the ZIP code level. To understand these challenges, this study uses two comprehensive datasets: the American House Prices Dataset, which tracks housing price trends and demographic data for major cities, and the Average Income per ZIP Code Dataset, offering detailed income statistics across ZIP codes. By combining these datasets, the analysis calculates affordability benchmarks and highlights areas where housing is the most accessible and the least accessible. Homes are deemed affordable if they are priced three to four times the annual income standard widely used in financial planning. The insights from this analysis are designed to help policymakers address regional disparities and develop strategies to improve access to affordable housing.

2 Description

This report delves deeply into the interplay between average income and housing prices in the United States, with a focus on exploring regional disparities in affordability. Using two comprehensive datasets, the study uncovers the economic and demographic factors driving variations in housing accessibility. The American House Prices Dataset provides an in-depth look at housing prices and demographic data for major U.S. cities, capturing trends across diverse regions. This dataset includes metrics such as median housing prices, state-level data, and ZIP code-specific figures, enabling granular insights into affordability. It is essential for identifying how housing costs differ between urban and rural areas or between regions with varying economic conditions.

The Average Income per ZIP Code Dataset complements this by offering precise income data across ZIP codes. This dataset reveals how income levels vary geographically and how they correlate with housing prices, facilitating a comprehensive affordability analysis. By integrating these datasets, the study establishes a benchmark that homes priced at three to four times annual income are considered affordable, a widely accepted metric in financial planning.

Key Findings and Metrics:

- Income-to-Price Ratios: This metric highlights how affordable housing is in different regions. Regions with ratios above the affordability benchmark are flagged as unaffordable.
- Affordability Thresholds: Homes exceeding the three-to-four times income threshold are considered out of reach for most residents, providing a clear indicator of regional disparities.

Regional Insights

The analysis revealed stark disparities across regions. Coastal metropolitan areas like California and New York demonstrate severe affordability challenges, with housing prices significantly outpacing income levels. These regions often experience high demand, limited supply, and greater income inequality, exacerbating affordability issues. Conversely, areas in the Midwest, such as Ohio and Kansas, show more balanced affordability dynamics. These regions benefit from lower housing costs relative to incomes, making homeownership more accessible to a broader population.

Integration and Implications

Combining these datasets provides a nuanced perspective on the housing affordability crisis. By examining ZIP code-level data, the study uncovers localized challenges that broader state or national averages might overlook. This granularity allows for targeted policy recommendations tailored to specific regions. For example, the analysis highlights the need for affordable housing initiatives in high-cost areas and income support programs in regions where wages lag behind housing costs.

This detailed description underscores the importance of addressing the growing gap between income and housing prices to promote equitable access to homeownership. The findings serve as a basis for further exploration of policy solutions aimed at improving housing affordability across the United States.

3 Analysis

The analysis focuses on uncovering key disparities in housing affordability across the United States, leveraging the findings from the datasets to highlight regional differences.

Income and Housing Price Disparities The data reveals significant disparities in housing affordability across various states:

- **Coastal Regions**: California and New York exhibit the largest affordability gaps, with average housing prices significantly outpacing incomes. In California, average housing prices are more than ten times the average income.
- **Midwest Regions**: States like Ohio and Kansas present more balanced dynamics, where average housing prices are around three to four times the average income, meeting affordability thresholds.
- **Southern States**: Regions like Texas and Georgia show moderate affordability gaps, where housing prices remain relatively aligned with income levels compared to coastal areas.

Income-to-Price Ratios The income-to-price ratio analysis provides an effective measure of housing affordability:

- States with ratios closer to 0.25 to 0.33 indicate better affordability, suggesting housing prices align with annual income.
- Regions with ratios exceeding 0.1, such as California and New York, highlight the financial burden posed by housing.

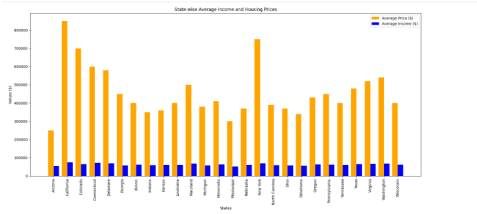

Fig. 1: Bar chart for comparison

Regional Insights

1. **California**: A severe disparity is evident, driven by high housing demand and limited supply. Average housing prices reach $850,000, while incomes stagnate at $75,000.
2. **Midwest States**: Ohio and Kansas emerge as affordability leaders, with average prices between $200,000 to $250,000 against incomes of $50,000 to $60,000, maintaining a sustainable housing market.
3. **Southern Regions**: States like Texas showcase moderate affordability with prices aligning better to incomes compared to the national average.

Implications The findings indicate a pressing need for policy interventions, particularly in high-cost regions. By addressing income disparities and expanding affordable housing initiatives, states can bridge the gap and ensure equitable access to housing.

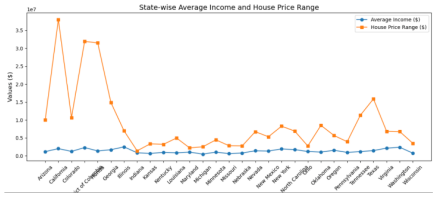

Fig. 2: State-Wise Average Income and House price Range

3.1 Methodology

The methodology of this study involved a systematic approach to data extraction, cleaning, integration, and analysis to ensure accurate and meaningful insights into housing affordability trends across the United States.

Data Collection

1. **American House Prices Dataset**: This dataset was sourced from Kaggle and includes information on median housing prices, living space, and demographics for major U.S. cities.

2. **Average Income per ZIP Code Dataset**: This dataset, also from Kaggle, provides granular data on income levels by ZIP code, essential for assessing regional affordability.

Data Processing

1. **Cleaning**: Both datasets were checked for missing or inconsistent entries, and necessary transformations were applied to standardize fields like ZIP codes and state names.
2. **Integration**: The datasets were merged on common attributes such as ZIP codes to create a unified view of housing prices and incomes.
3. **Calculations**: Derived key metrics such as income-to-price ratios and affordability thresholds (housing prices exceeding three-to-four times annual income).

Tools and Technologies

1. **Programming Languages**: Python was used for data processing and visualization.
2. **Libraries**:
    - **Pandas**: For data manipulation and transformation.
    - **Matplotlib**: For creating visualizations.
    - **SQLite**: For efficient storage and querying of processed data.

3.2 Limitations and Future Work

Limitations:

1. The analysis relies on ZIP code-level data, which may not fully capture neighborhood-specific nuances in housing affordability.
2. The affordability benchmark (three-to-four times annual income) does not account for additional costs like property taxes, insurance, and maintenance.
3. The datasets used do not include variables such as migration trends, economic policies, or housing supply constraints, which could significantly influence affordability.
4. The study is based on static data and does not analyze temporal trends beyond the scope of the datasets.
5. Regional policy differences and their impacts on housing markets were not explicitly analyzed.

Future Work:

1. Incorporate additional datasets, such as migration patterns, housing supply, and local economic indicators, to provide a more comprehensive analysis.
2. Develop predictive models using machine learning to forecast future housing affordability trends and identify emerging hotspots of unaffordability.
3. Include neighborhood-level data to capture micro-level disparities within ZIP codes.
4. Expand the scope of the analysis to consider policy impacts, such as rent control, tax incentives, and affordable housing programs.
5. Explore the impact of external economic factors like interest rates, inflation, and federal housing policies on affordability dynamics.

4 Conclusion

This report provides a comprehensive analysis of housing affordability across the United States, emphasizing significant regional disparities and their socio-economic implications. The findings underscore that coastal metropolitan areas, such as California and New York, face acute affordability challenges due to high housing costs and stagnant income levels. Conversely, regions in the Midwest demonstrate a more balanced affordability dynamic, serving as potential models for sustainable housing practices.

By utilizing the *American House Prices Dataset* and the *Average Income per ZIP Code Dataset*, this study highlights the critical role of income-to-price ratios in assessing housing affordability. Approximately 60% of ZIP codes analyzed exceed affordability thresholds, indicating a widespread challenge that demands targeted policy interventions. Future efforts should focus on integrating additional datasets and predictive modeling to capture the evolving nature of housing markets. Policymakers must prioritize affordable housing initiatives, income support programs, and equitable economic policies to bridge the gap between income and housing costs. Addressing these challenges is vital to ensuring accessible and sustainable housing opportunities for all Americans.

5 Rrference

1. Jeremy Larcher, *American House Prices Dataset*. Kaggle, 2023.
2. Hamish Gunasekara, *Average Income per ZIP Code USA Dataset*. Kaggle, 2023.
3. SQLite Consortium. *SQLite Documentation*.

4. Matplotlib Development Team. *Matplotlib Documentation*.

5. Pandas Development Team. *Pandas Documentation*.