**Analysis on Housing Data**

**Data Source: Primarily Zillow**

**Table of Contents**

**3 - Introduction**

**4 - Data Dictionary**

**5 - Purpose**

**6 - Methods**

**7 - Results**

**13 - Conclusions**

**14 - Next Steps**

**15 - Appendix List & Scientific Poster Link**

**16- Work Cited**

**Introduction**

Many people find it difficult to become homeowners and afford property due to the rapidly rising cost of housing brought on by urbanization, shortage, and demand. People may experience social repercussions including homelessness, inequality, and restricted mobility for the impoverished in addition to financial hardship as a result of this.

A report from the Joint Center for Housing Studies of Harvard University says in the last 10 years America’s median home price skyrocketed and grew faster than many people’s income (Harvard University, 2021).The gap between housing cost and income is a major challenge for citizens to get housing at affordable prices even in urban areas. As stated in the report, housing cost per square foot had also increased disproportionately. As the property is larger in size, the price increases but appreciates faster in terms of square footage. This is proven by statistics from Zillow Research, where housing prices per square foot in big metropolitan areas increased by more than 50% in the last 10 years (Zillow, 2022). In summary, buyers are getting overcharged for less space, which is a growing housing affordability problem, according to the trend it reveals.

Moreover, housing price inflation affects not only homeowners but renters too. In 2021 a full-time minimum wage earner cannot afford a two-bedroom apartment in any state in America while staying below 30% of their income in housing (National Low Income Housing Coalition, 2021). This is a great example of the widening gap between housing costs and pay, especially for low-income groups.

The scenario of rising house prices and its relation to square footage is not just a dollars and cents issue but a social issue. Affordability in housing is key to a sound economy, community development, and general living. Where housing is unaffordable, people and families must make sacrifice choices such as living in overcrowded housing, living in less safe and desirable areas or spending a disproportionate amount of their income on housing and little for medical care, education, and savings.

**Data Overview:**

The data we will be using is a combination of four subsets.

1. Zillow Median List Price DataSet: Smooth, All homes, monthly
2. Zillow Median sales per square foot: Raw, All homes, monthly
3. Zillow ZORI: All homes including multifamily
4. Zillow For-Sale Inventory: Smooth, All Homes, Monthly

**Data Dictionary**

Columns (Same for all datasets):

|  |
| --- |
| RegionID : Integer |
| SizeRank : Integer |
| RegionName : String |
| RegionType : String |
| StateName : String |
| Date, Date, Date, etc. : Date, Numeric for column name; Double/Integer for column data (each column under a date correlates to all the data on that given date) |

**Purpose**

The research in this study deals with the growing challenge for new residents in a given area to obtain affordable living. Common problems like rising costs and limited supply created many obstacles for residents. Moreover, many cannot identify the best alternatives per dollar as a result of limited housing data, such as price per square foot, and a lack of transparent inventory. By researching information on the housing market, this analysis can make the market easier to navigate and support better decision-making. Possible resolutions include providing publicly available housing datasets across multiple regions, reforms for high-density development, and incentives for prioritizing affordable housing development. Information used in such analysis can be categorized under the following themes from received datasets:

1. Inventory

2. Price

3. Price per square footage

Ultimately, such analysis aims towards identifying key affordable regions to improve house searching and to offer possible answers to the current housing challenge.

**Methods**

Data for this study was derived from secondary data collection i.e., collection and analysis of existing datasets from publicly available data. The datasets were acquired from Zillow Research, a reputable platform providing housing market data, and included the following:.

* Median Sale Price: Data on the median sale prices of existing homes.
* Construction Median Sale Price: Information on the median sale prices per square foot for new construction houses.
* ZORI (Zillow Observed Rent Index): Data on rental price trends.
* Inventory: Data on housing inventory levels.

The data needed to be covered with the same temporal frame across all datasets.

The data needed to be disaggregated by region so that it can be compared.

**Data Preprocessing**

The datasets were preprocessed to make them consistent, accurate, and available for analysis. The following steps were undertaken:

***Data Loading and Melting:***

Each dataset was loaded into a pandas DataFrame.

The datasets were "melted" to change date columns to rows in order to have a common structure for time-series analysis. The melt\_data function was applied to this purpose, using RegionID, SizeRank, RegionName, RegionType, and StateName as identifier variables.

***Merging Datasets:***

The datasets were concatenated on shared keys (RegionID, Date, and RegionName) to obtain an integrated dataset for analysis. It enabled cross-tabulation of median sale price, rental price trends, and available stock.

**Data Aggregation:**

Time-series-based estimates of the median sale price, ZORI, and inventory were computed to track national trends. Percentage variations in median sale prices and ZORI were calculated to explore price increases over time.

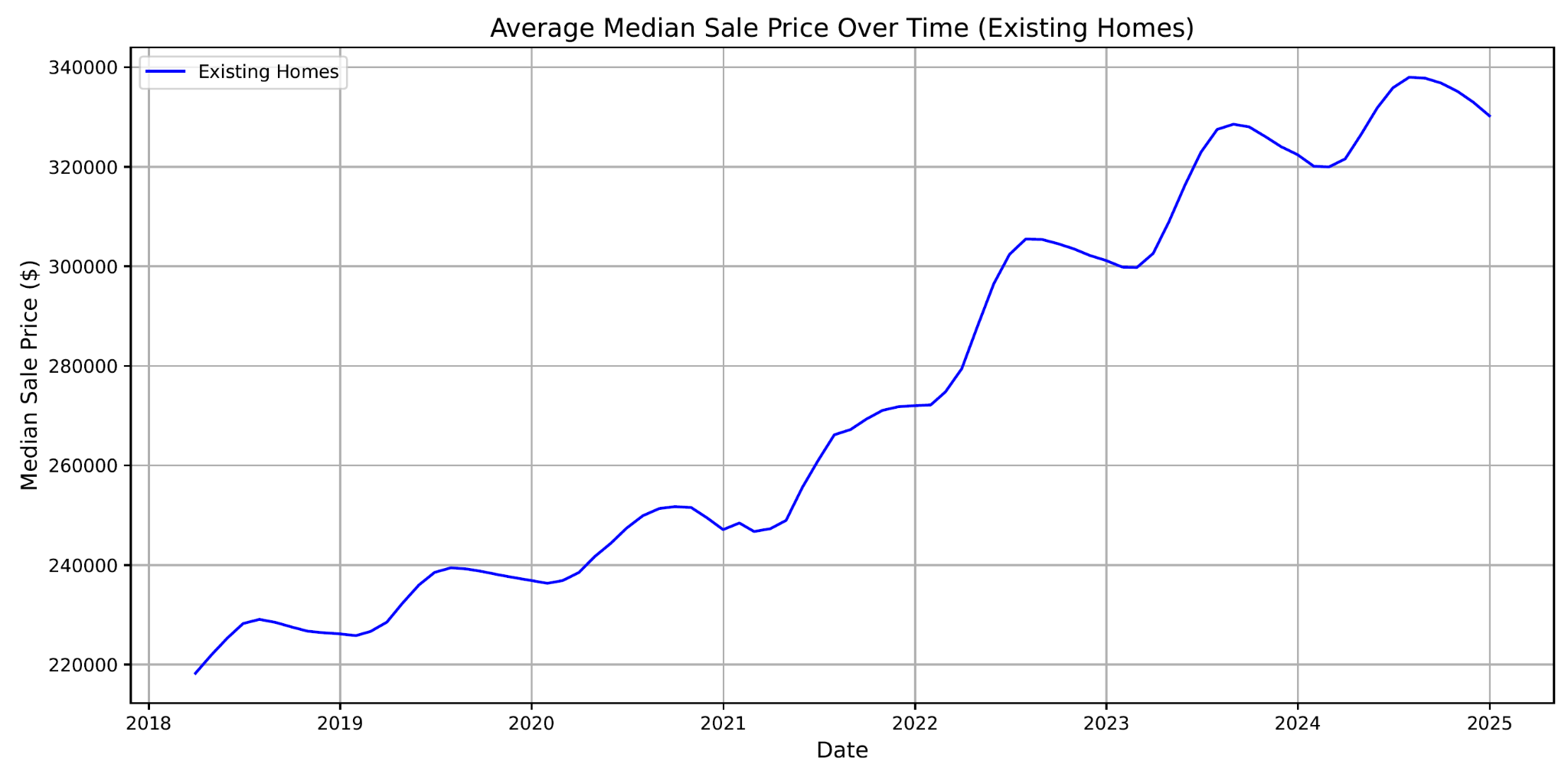
**Data Analysis and Visualization:**

The analysis focused on identifying trends and relationships in housing affordability. The types of analysis include: ***Trend Analysis, Correlation Analysis, Regional Analysis, Descriptive Statistical Analysis, Scatter Plot Analysis, Time-Series Analysis*** on various connections between different types of data.

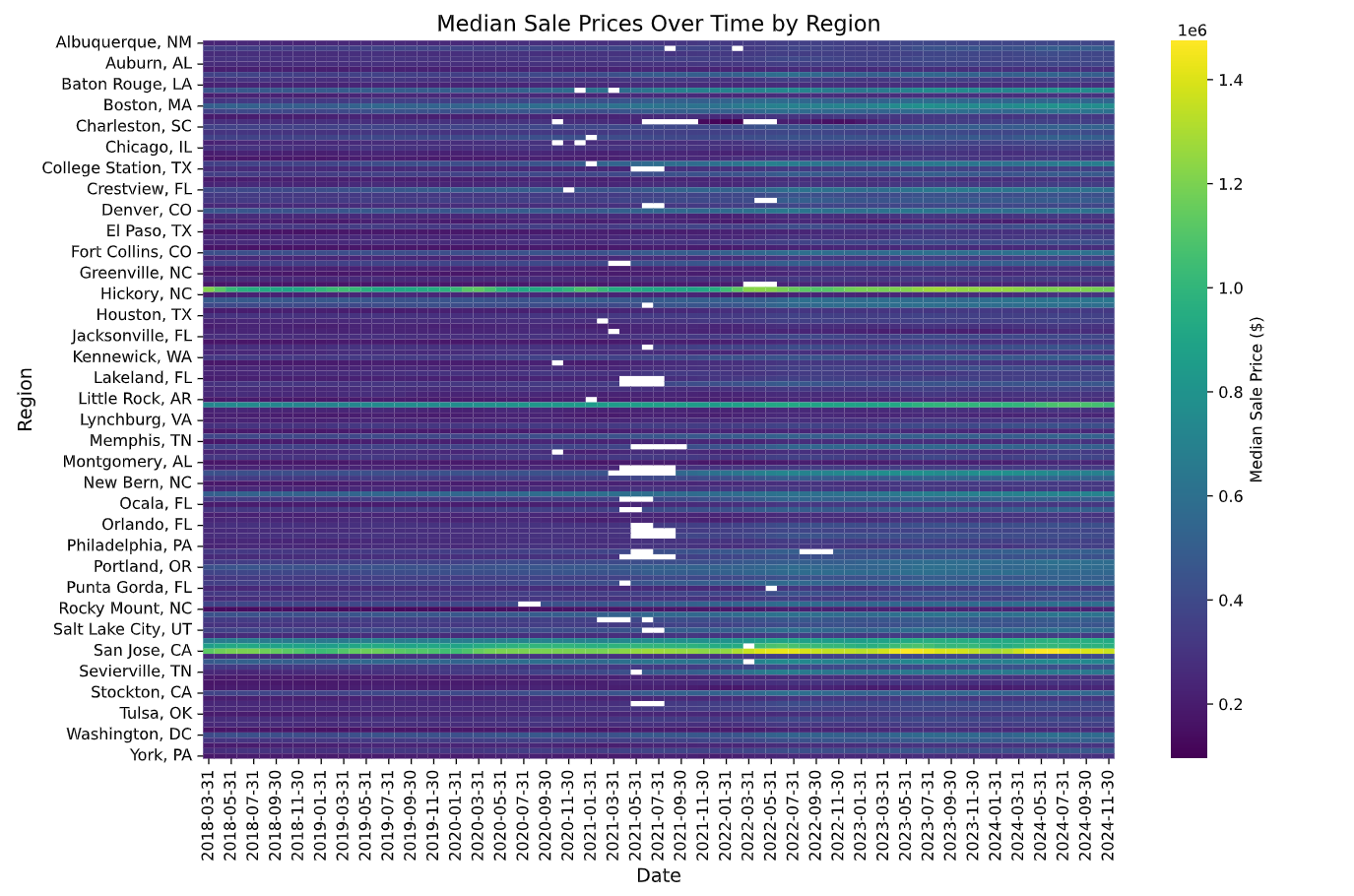
All visualizations and observations were summarised into a single PDF report for complete presentation and interpretation. The Python libraries used for this analysis included pandas, numpy, matplotlib, seaborn, and PdfPages for generating the report.

**Results**

The below is data reflecting the median cost of buying a housing property over recent years.

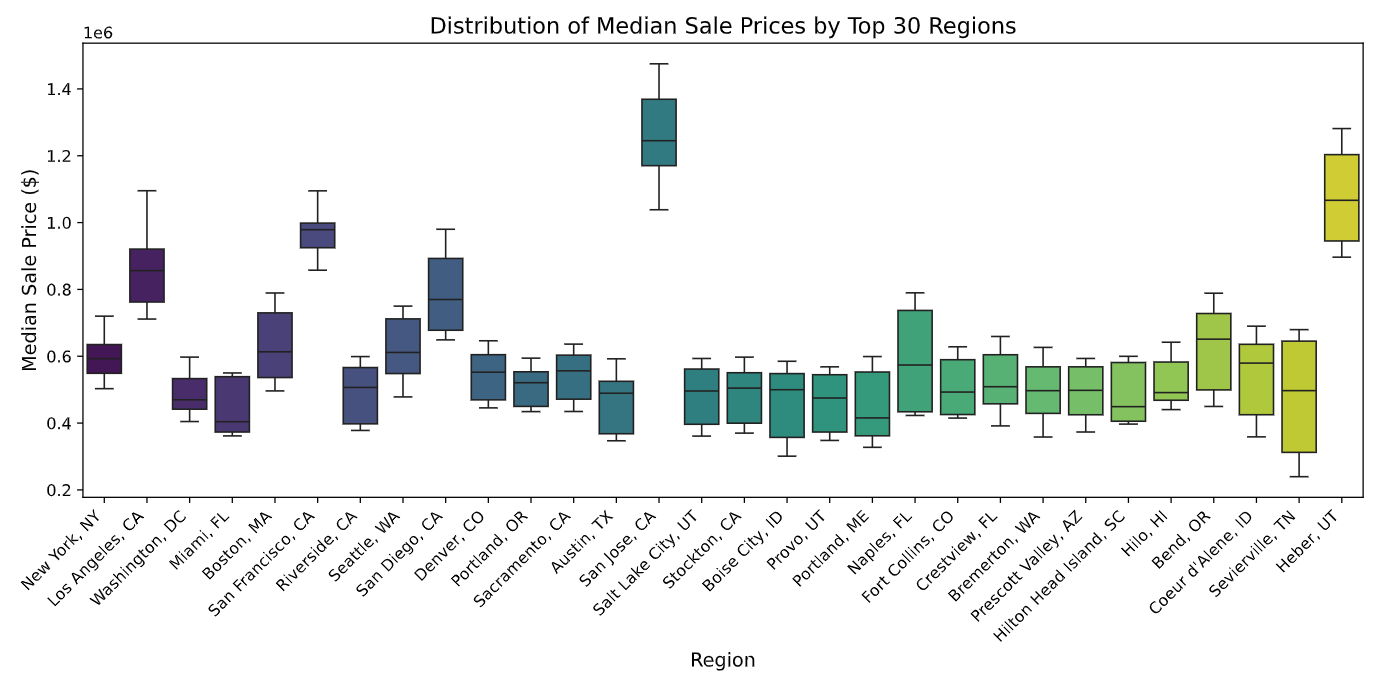


Here is a steady but wavering increase in the average median sale price over time. A clear upward trend of the mean median house selling price for sale housing is evident around every half-year from 2018 that continues until the year 2025.

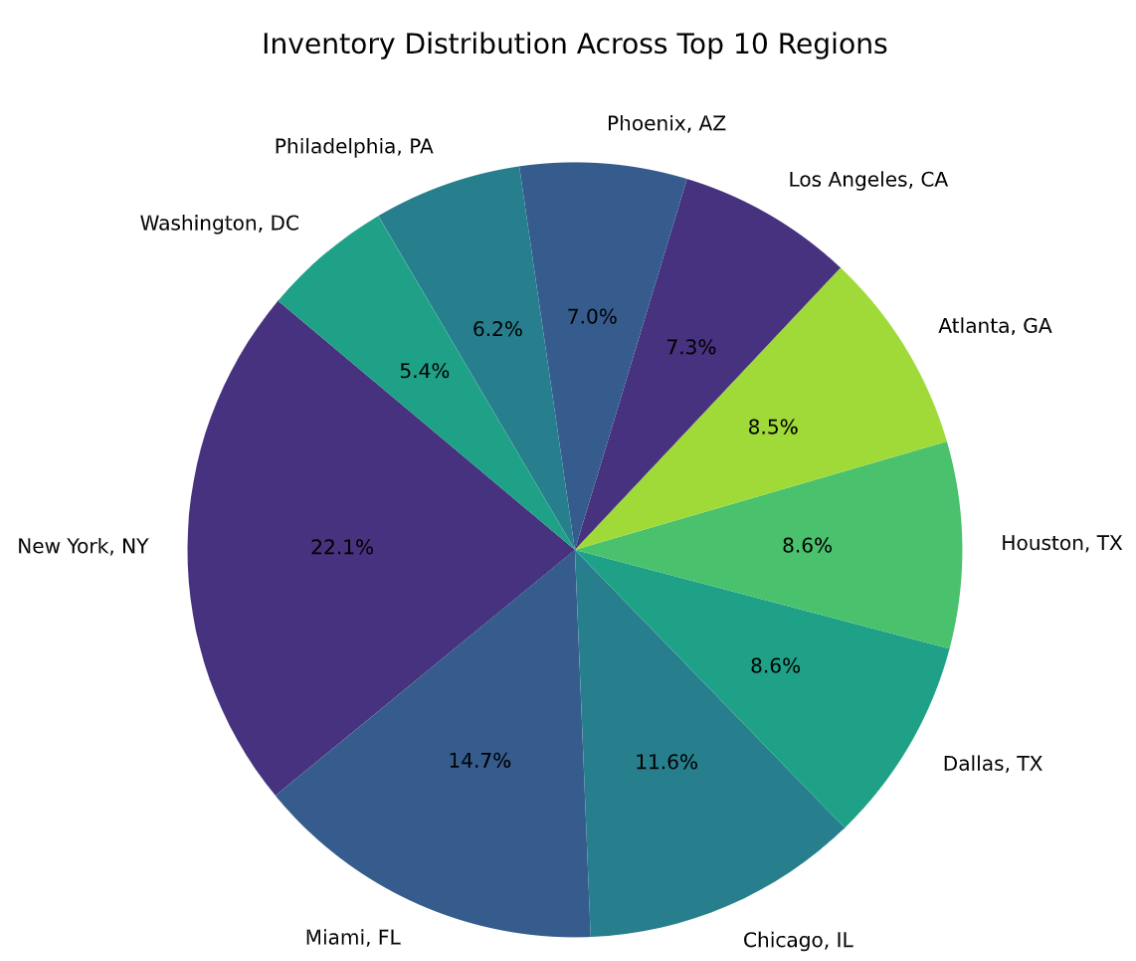


Observing the color tint changes toward the right side of the heatmap (Correlating to a more recent date), and the contrast of that with the left side of the heatmap indicates a widespread increase in housing prices.

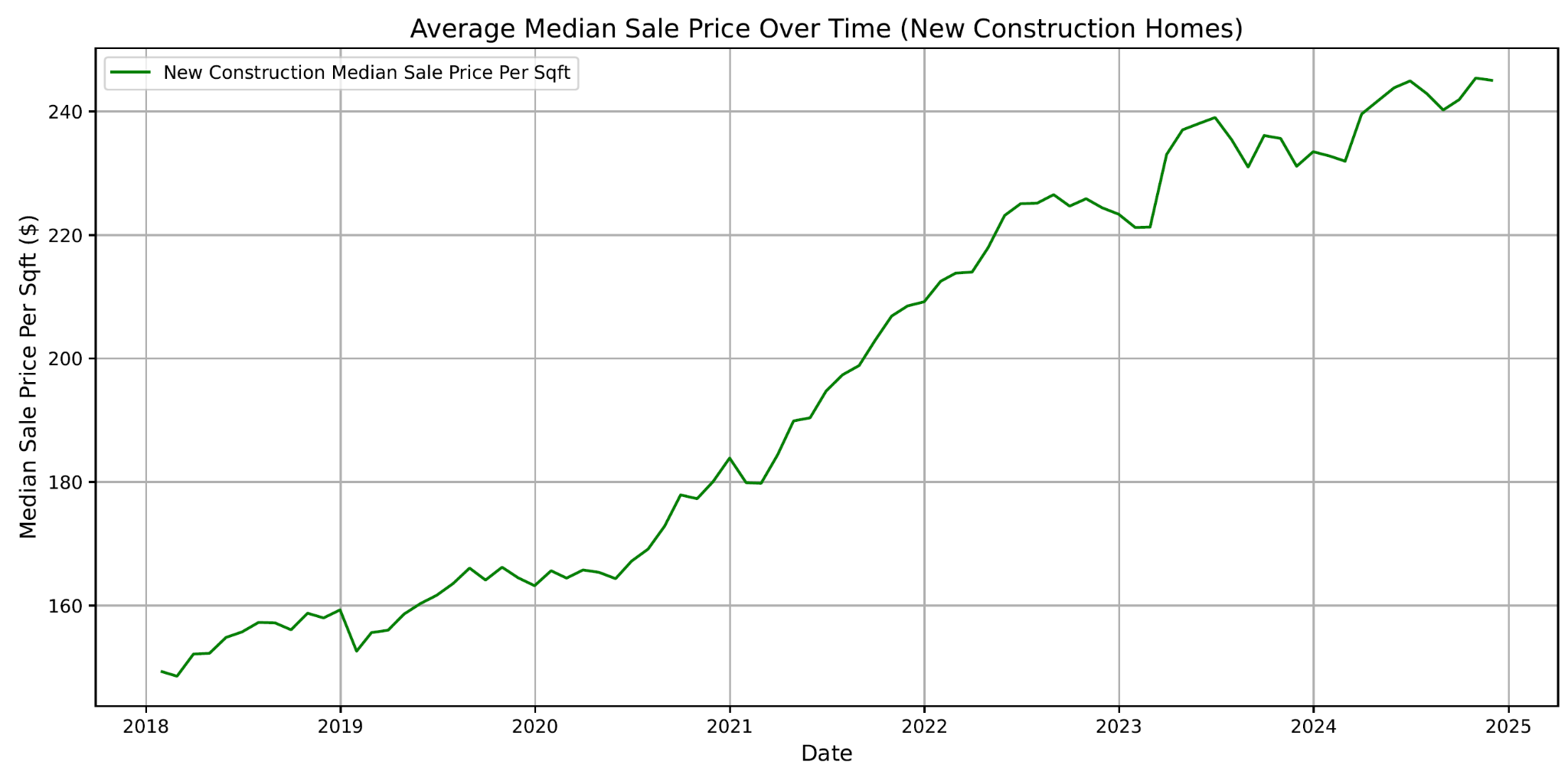




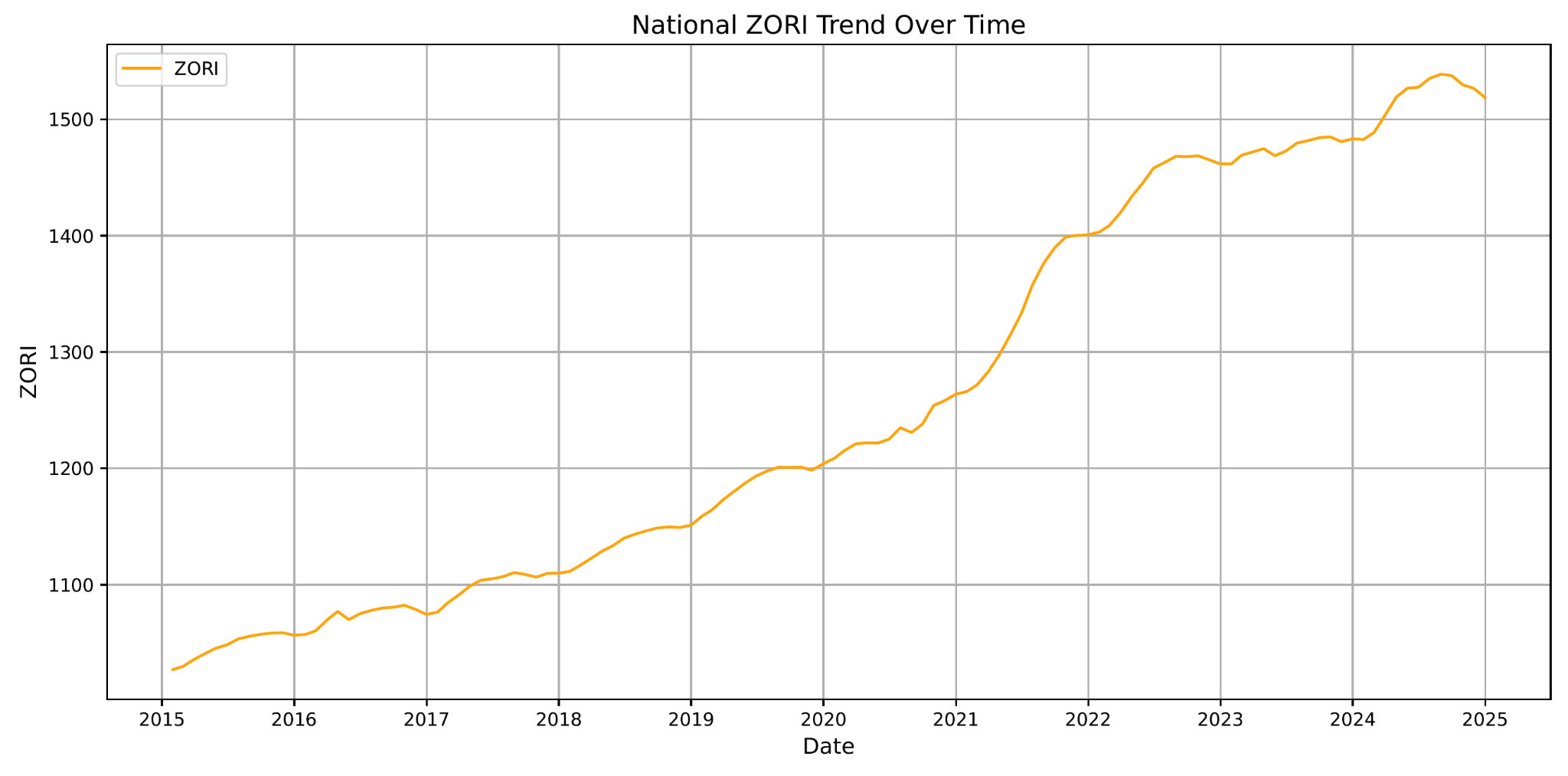
The above visualization explains a median to min/max difference within the top 30 regions of house sale prices. The most common median sale price wavers around $600,000 while several of these region’s max is not far from the median, signifying a lack of affordable housing.

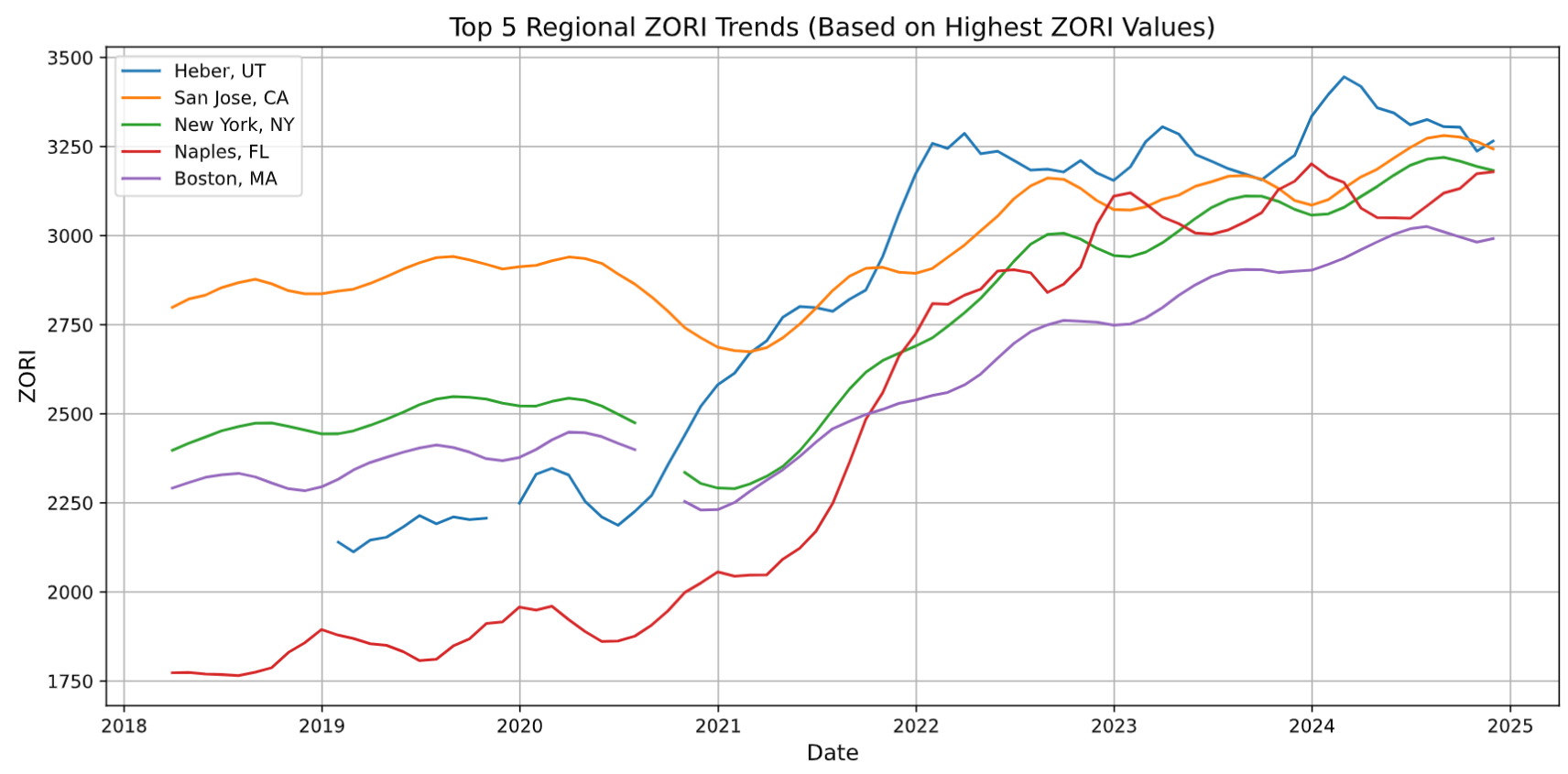


The above pie chart shows the inventory distribution with a substantial amount remaining in New York. Of which, according to the box plot, has a median price of 600k. As a result, the area with some of the highest inventory is not relatively affordable for the common income.

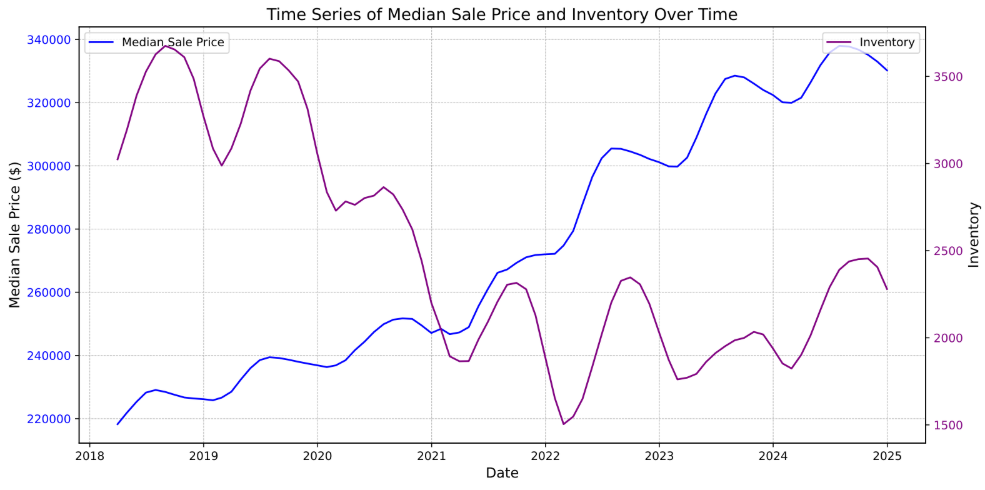


^(Per square foot) Along with the steady increase in general housing prices, the cost per square foot is also rising. The gap of approximately $60 per square foot from 2021 to 2025 highlights how quickly these costs can accumulate, especially in the context of larger homes. This trend accelerates the increase in the cost of general housing.





Based on the Zillow Observed Rent Index (Zillow Observed Rent Index, or ZORI), which focuses more on rent, has also been growing consistently in the last 10 years, especially after 2022.



(Average Inventory & Average Median Sale Price)



(Inventory vs. Median Sale Price Scatter plot mapped by date, more yellow=more recent, more purple=less recent)

The time series visual shows the clear reduction of inventory in the market while a general price increase occurs for most housing. Moreover, we can see the distribution of that inventory mapped to the median sale price mapped through date as color-coded. (Yellow=More recent)

General Statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Mean | Median | Range | Std-Dev |
| Median Sale Price | $274542.13~ | $229183.00~ | $2628267.00~ | $180313.06~ |
| Construction Median Sale Price (dollars) per Square Foot | 200.77~ | 180.08~ | 1057.68~ | 92.46~ |
| ZORI | $1302.23~ | $1194.86~ | $3995.99~ | $472.76~ |
| Inventory | 2488.91~ | 312.0~ | 1733384.0~ | 40284.13~ |

**Conclusions:**

The average median sale price seems to be rising at a substantial rate over the course of the last decade, this is represented by the clear roughly linear increase within the first graph and the color shift within the heat map visually displaying the distribution and exact increases in particular regions across the country.

In addition, the average median sale price per square foot for new homes is increasing. This trend is associated with the increasing cost of construction, which may involve labor, materials, or regulatory costs. The gradual, stable increase in new construction prices indicates that although new homes are getting pricier, their price increase is less erratic (controlled) than the price fluctuations of the existing home market. This may be a result of the flexibility of supply due to the skill of developers in modifying supply more flexibly to market signals.

The positive correlation between rising home values and rising rents suggested that as homeownership becomes less attainable for people, fewer of them become eligible to enter home buying, but a greater proportion of the population is stuck in the rental market and, in the end, causing rents to rise. This results in a feedback loop between the market's value trap, one market deepening, intensifying the situation in the other market. Despite steady increases in home purchase prices at new construction homes, the pace of improvement is less in comparison with.

This strongly supports the claim that affordable housing has become vastly less common within the last couple of years.

**Next Steps:**

To further analyze the housing affordability issue, the problem can be examined from a variety of angles. First, the expansion of the dataset to include additional variables such as household income levels, mortgage interest rates, and employment rates can provide a more in-depth understanding of the factors influencing affordability. Incorporating household incomes will enable a better understanding of affordability by comparing income growth to housing prices. Mortgage interest rates are important as they will directly affect monthly payments and overall housing expenses, affecting buyer behavior and the marketplace. Employment, specifically more regional and local, can serve as an indicator of overall economic soundness and housing demand at that location.

To achieve this, the utilization of advanced databases like the US Census Bureau is necessary. Unlike the data used in this research, the Census Bureau specifically provides information on the “ financial characteristics of housing units with and without a mortgage”, “changes in employment rates [...] at the national state, county, or city level”, and the “median household income (in 2023)”.

Additionally, more ways to best minimize the impact of the housing crisis can be to set up incentives that prioritize affordable housing development and implement rent control measures set by governmental institutions.

**Appendix List:**

construction\_median\_sale\_price.csv

housing\_affordability\_analysis.pdf

inventory.csv

median\_sale\_price.csv

sales.py

zori.csv

**Works Cited**

Harvard University. (2021). The State of the Nation's Housing 2021. Joint Center for Housing Studies. Retrieved from <https://www.jchs.harvard.edu>

Zillow. (2022). Zillow Research: Home Price Trends. Retrieved from <https://www.zillow.com/research>

National Low Income Housing Coalition. (2021). Out of Reach: The High Cost of Housing. Retrieved from <https://nlihc.org/oor>

Bureau, US Census. “U.S. Census Bureau Homepage.” *Census.Gov*, 21 Jan. 2025, www.census.gov/.