Data Analytics Project 1 Written Analysis
August 12, 2024
Laura Magallanes,Rob Huibregtse,
Jane Hedges, James Pelletier

Project Title: COVID-19 and Real Estate: A Zillow Data Analysis

Hypothesis 1:

The Hypothesis 1 analysis evaluates whether larger cities experienced a significant increase in inventory levels in comparison to smaller cities due to individuals moving from larger cities to more remote locations during the pandemic period of 2020 through 2021. The inventory data used in the analysis was derived from Zillow Housing Data and, for Hypothesis 1, segmented into two groups, the 'Top 100' large city segment and the remaining size segments as smaller cities. Referencing the Comparison of Inventory Index and Interest Rates (Dual Y-Axes For Better Readability) graph, it displayed consistent fluctuation between the 2018 through 2022 regardless of Federal Interest Rate, showing relatively low correlation between Inventory Index and Federal Interest rates. Referencing the Annual Average Inventory Index graph which utilizes a dual Y-axis for better visualization, revealed clear differences in inventory levels between large and small cities. Large cities demonstrated significantly higher inventory levels in comparison to smaller cities, thus signifying significantly more available housing in larger cities. This indicates there is a strong correlation between city size and inventory levels. This graph revealed a significant drop in both large city and small city inventory levels around 2020, likely influenced by the COVID-19 pandemic. According to the graph, Yearly Percent Change in Inventory Levels, the pandemic years of 2019 and 2020, displayed a low yearly percent change changing from -0.769309 to-23.053959 in just one year, a drastic difference from its following three years.

The summary statistic analysis further supports the graph visualizations. Large cities continue to display much higher levels of inventory in comparison to smaller cities. A high standard deviation of 4811.225 suggests a wide range of values and significant variability in inventory index among large cities. In addition, The variance at 23147887.48, is consistent with the high standard deviation, further reflecting the wide dispersion in the data. For small cities, a low standard deviation of 121.477 suggests less variability in inventory index values among small cities. In addition, the variance at 14756.675 is much lower, consistent with the lower standard deviation, showing less spread in the data compared to large cities. Referencing the Large Cities Inventory Index Statistics by Year graph, it reveals a significant drop amongst mean, median, and mode between the years of 2019-2020 when the COVID-19 pandemic was at its peak. This signifies there were far fewer available housing options during these years suggesting a potential economic shock or event such as the COVID-19 pandemic that impacted inventory levels. The following years suggest a recovery in the inventory index as a response to adjustments after the peak of the pandemic. Referencing the Small Cities Inventory Index

Statistics by Year graph, it indicates that smaller cities experienced the same trend as the Larger Cities with the drastic drop in all three measures. In addition, the smaller cities also experienced a similar recovery period. Although there is a difference in inventory levels between the Large Cities and the Small Cities, they follow relatively the same trends along the pandemic years.

The statistical test that was run for inventory levels consisted of an ANOVA test which resulted in a F-Statistic: 246.7993546042154 and a P-Value: 5.0385975458100496e-55. The very high F-Statistic indicates the difference in the mean between large and small cities is much larger than the variability. This suggests that the differences in inventory levels between large cities and small cities are significant and city size does have an impact on inventory levels. The very small P-Value indicates the likelihood that the difference between larger and smaller cities is not random. The ANOVA results support the conclusion that there is a statistical difference in inventory levels between large and small cities.

Hypothesis 2:

This analysis evaluates the changes in rent index in larger cities and smaller cities during the 2020-2021 pandemic period. The hypothesis I am answering is whether larger cities experienced a decrease in rent value index compared to smaller cities during the pandemic period. The graph displaying the monthly percentage change in rent value index during the pandemic shows the percentage changes between large and small cities. The most significant drops in large cities occurred around July 2020 and November 2020, with the most pronounced decrease being in October 2020 where the percentage change drops below -0.02. When comparing this to smaller cities, the rent value index generally follows a more stable upward trend after mid-2020. The increases in rent value percentage change are particularly strong from October 2020 to mid-2021, peaking around January 2021. This graphical evidence initially supports the alternative hypothesis, suggesting that larger cities did experience a more substantial decrease in rent value index during the pandemic. However, the ANOVA test yielded a p-value of 0.86, indicating that the difference is not statistically significant. Therefore, I fail to reject the null hypothesis, as the p-value is greater than the 0.05 threshold. The F statistic was very low (0.0313), which suggests there is little variation in rent index percent changes between large and small cities.

Another graph showing rent index over time from 2018 to 2023 across size segments contradicts the hypothesis and suggests that larger cities did not experience a decrease in rent value index during the pandemic. Instead, larger cities experienced significant growth in rent values, especially after 2020. While smaller cities also saw a growth, it was more moderate compared to larger cities.

When comparing the summary statistics between large and small cities during the pandemic period, I observed that larger cities had a slightly higher average rent increase of 0.74%, compared to 0.69% in smaller cities. The range of rent changes was also broader in larger cities, with a maximum increase of 2.88% and a minimum decrease of -2.17%, while smaller cities saw a maximum increase of 2.80% and a smaller minimum decrease of -0.36%. The

maximum of smaller and larger cities was roughly the same which suggests that both cities experienced similar increases in rent value index during the pandemic period. This could entail that factors driving maximum rent increases were impactful across city size and could have external influences during the pandemic.

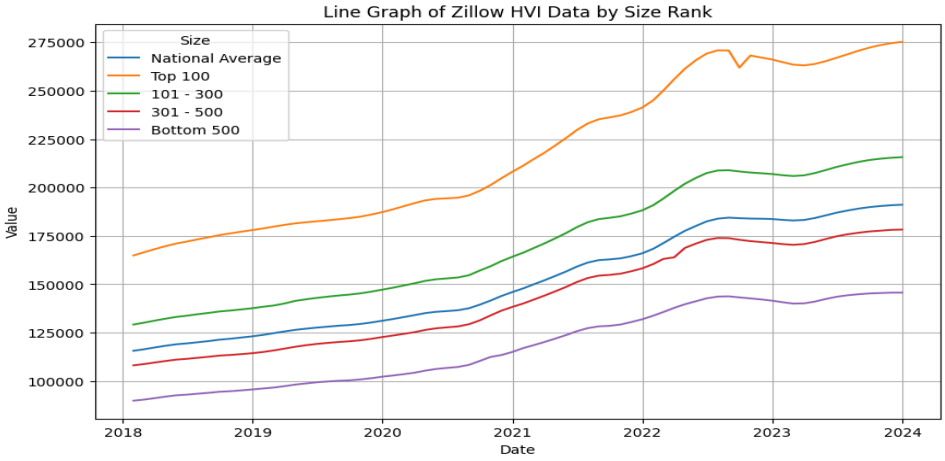
Additionally, I looked at the relationship between the rent index and the federal funds rate and discovered that the rent index rises when the federal funds rate is low. This finding may indicate that lower interest rates are a factor in rising rent prices. The data indicates a complex relationship that is probably influenced by more than simply the Federal Funds Rate, but it does show a correlation between low interest rates and rising rent indexes. The consistent rise in rent pricing over the past few years is indicative of broader developments in the housing and economic markets.

Hypothesis 3:

The hypothesis I attempted to answer for my group was if larger cities experienced a decrease in home value index (HVI) compared to smaller cities due to individuals moving from larger cities to more remote locations during the pandemic period (2020-2021). Using Zillow data, I looked at the average cost of homes, rent, and condos in the United States. The data was separated by city. Using the data and to further my hypothesis, I will answer the questions of if there has been a notable dip in the HVI for large cities between 2018 - 2023 and if home prices increased at a faster rate in smaller cities compared to large cities.

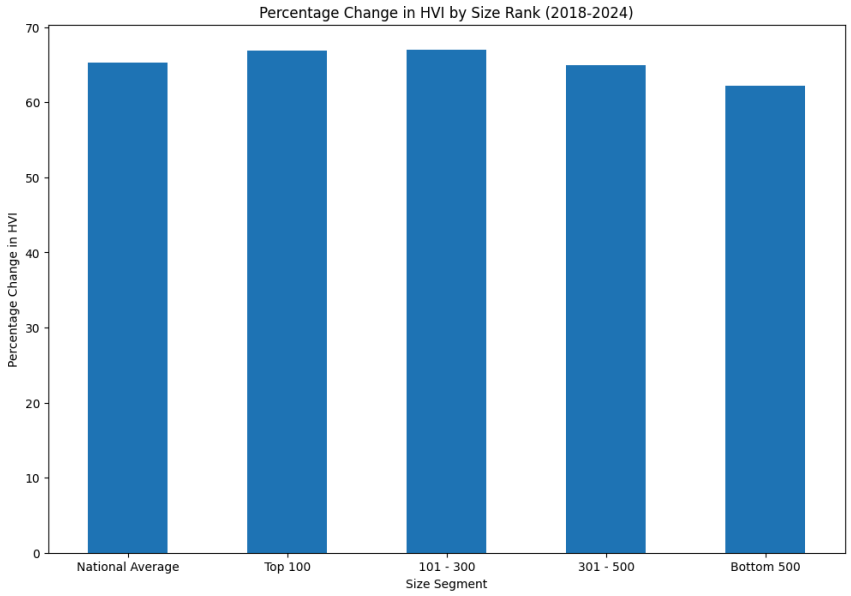
In order to answer these questions effectively, it was necessary to separate the data by city size. For this data set, I chose to divide the data into Top 100, 101-300, 301-500, and Bottom 500. This decision was made because the top 100 cities were the majority of the larger cities while every city below the top 100 was much smaller.

The graph below illustrates the Zillow HVI data from 2018 to 2024, separated into the aforementioned city size categories. Which will give us insight whether or not HVI did decrease in larger cities



Following the Top 100 cities which is the orange line, there is no statistically relevant dip. Similarly none of the other categories of cities had any dip during the pandemic. All HVI's are following an upward trend.

Before completely rejecting our hypothesis, it is also important to look at the percent change in HVI across all categories right before COVID-19 and right after. The bar graph below is a result of averaging each category between 2018 - 2024.



Once again there is no statistical relevance with the data to confirm that the pandemic decreased the HVI in larger cities. The percent change in the Home Value Index (HVI) across different city size categories from 2018 to 2024, as shown in the bar graph, further supports the conclusion that the pandemic did not cause a statistically significant decrease in the HVI for larger cities. The data indicates that all city size segments experienced a similar percentage increase in HVI during this period.

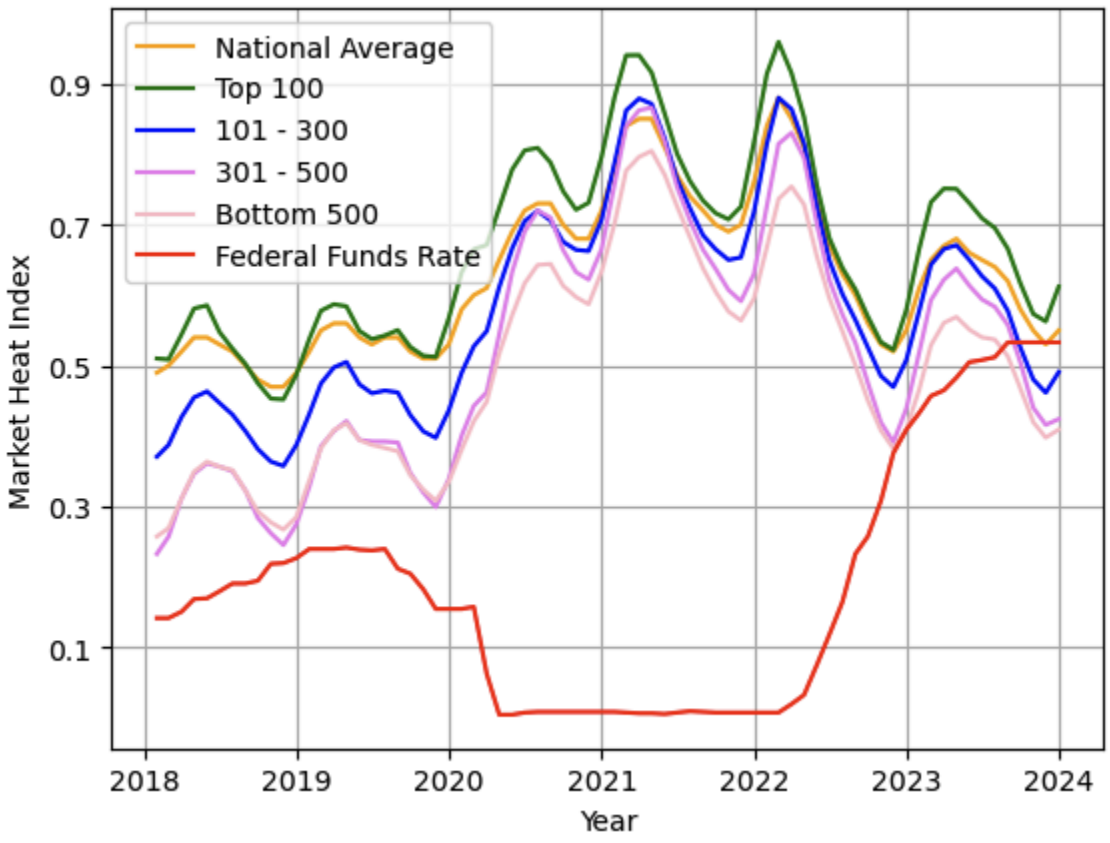
The summary statistics for HVI provided a view of the distribution of home values across the city size segments. Top 100 cities had the highest mean, averaging $218,552.83 with a standard deviation of $38,348.44, the highest of all categories. The bottom 500 cities had the lowest mean HVI at $118,202.88 with a standard deviation of $19,875.84, the lowest of all the categories. These values are expected. Larger cities will have more varying HVI's when compared to smaller cities. The p-value of 7.726e-67 is far above 0.05 which indicates that there is no statistical significance.

The data does not support the hypothesis that the pandemic caused a decrease in the Home Value Index (HVI) in larger cities compared to smaller cities. Both larger and smaller cities experienced steady growth in home values during the period studied, with no notable dip

during the pandemic. This suggests that the expected migration from larger cities to more remote locations did not result in a significant decline in home values in urban areas.

Hypothesis 4:

In analyzing the Market Heat Index, which we see as a good proxy for overall housing demand, much like the other metrics analyzed in this project, we looked for statistical evidence that there was an increase in the market heat index in smaller metro areas relative to large metro areas. Visually, we could find non such evidence; the trends in market heat index throughout the time period of 2018 - 2023 seemed to move in unison, and we did not see any specific size segment in our data break away from the general trendlines. When the market heat index rose among large metro areas, it also rose in smaller metro areas. The gap between market heat index at the beginning of our analysis period between the largest and smallest metro area segments shrank from 25.24 to 20.27 over the course of the analysis, but we don't believe this to be a significant enough change to support the idea that market heat index was much higher in the small areas than in large metro areas. This is backed up by our inability once again to reject the null hypothesis that there was no significant change in market heat index in larger metro areas (top 100) versus smaller metro areas. An ANOVA test on the rate of change of market heat index during the pandemic years of 2020-2021 yielded a 0.200 F-value and a P-value of 0.656, indicating that there was no statistical evidence to indicate that there as a significant difference between the different size segments. Simply put, a decrease in demand in large metro areas does not appear to have increased demand in smaller metro areas.



We did uncover an interesting trend when comparing the market heat index to the Federal Funds Rate (interest rates) by analyzing the correlation between the two. In the pre-pandemic period (2018-2019), there was a slight positive correlation coefficient of 0.279 between these two variables. This goes against basic economic theory, which tells us that as interest rates go up, market heat index would go down, holding all else constant. Of course, the US economy holds very few actual constants; I suspect this is due to a long-standing housing market trend of increasing prices and demand that was fueled by buyers who simply did not care about interest rates as much. Fear of missing out on a home purchase in the face of rising home prices likely drove buyers to simply accept not give interest rates as much weight in their home-buying decisions as before. However, when the Federal Reserve lowered interest rates to a near-zero value in response to the COVID pandemic, this factored into a significant shock to the housing market, and now interest rates and market heat index were observed to have a significant negative correlation (as is expected) at -0.629 during the pandemic (2020-2021) and even stronger in the post-pandemic period (2022-2023) at -0.715. The positive correlation from before the pandemic has been “corrected”. I’m sure the Federal Reserve are happy about this development.