

Zillow Zuperstars

An analysis of Zillow house price changes due to the pandemic



Evani Dalal, Morgan Lo, Preeti Nagalamadaka and Vandana Ramesh

Background

The pandemic greatly altered housing patterns for aspects like sale prices, the number of listed properties, and locations. We sought to examine the degree of the change in sale prices over time, how many homes were listed, and how property prices were affected in small and larger cities. More specifically, we looked at whether there was a correlation between the number of properties available and the profit the seller would receive, whether the sale price of houses increased at a faster rate before or after the pandemic, and how property prices changed depending on the size of the city.

Data Collection

We got all of our house price data from Zillow’s publicly available datasets. It consists of inventories, sale prices and list prices of houses. Our final dataset had roughly 900 data points (cities) for median listing prices, and roughly 500 data points for the median sale prices. It also contained a size rank value for all the cities. The data spans multiple years from 2020 to 2022. We know that the data is reputable because it comes directly from Zillow.

Methodology

We had three main hypotheses we wanted to test
1: There is a correlation between the number of houses available for sale and the profit that the seller would receive (sale price - list price). **regression analysis and one-sided t test**
2: There is not a relationship to suggest that sale prices of houses over time increased more rapidly before the pandemic than after the pandemic. **regression analysis**
3: During the pandemic, many people migrated out of hubs like NYC and LA to smaller cities. There were roughly four groups - large cities that saw property prices fall, small cities that saw property prices increase, somewhat large cities that saw property prices increase, and small cities that saw property prices decrease. **Clustering analysis**

Results

Claim 1



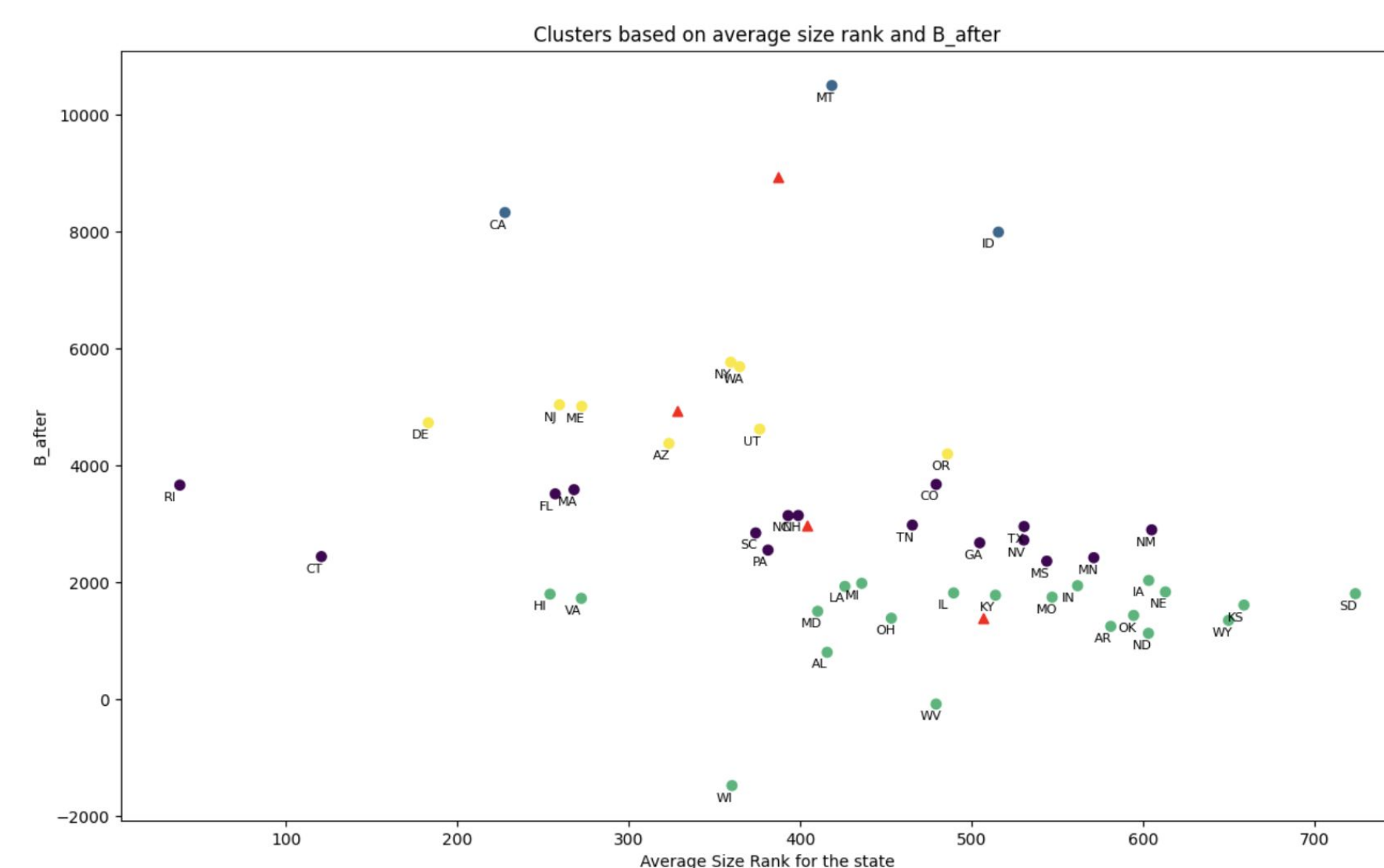
The first graph shows a correlation between the number of houses for sale affects the profit that one receives from selling the house. We then ran a one sided t-test on $b < 0$, which would imply that when there are more houses for sale in a particular city the profit of the sale decreases (i.e. the difference between sale price and list price goes down). When we ran this hypothesis test, we found that a t-value of -3.9137 and a p-value of 0.0001030. Because our p-value is less than our significance level of 0.05, we can conclude that our results are significant and that there is a correlation between the number of houses per sale and the amount of profit received.

Claim 2

Type of Value	Value
Significance Level	0.05
t-value	-6.485771675905118
p-value	0.4999999989805289

There is not a relationship to suggest that sale prices of houses over time increased more rapidly before the pandemic than after the pandemic.

Claim 3



Cluster Pair	Centroid A	Centroid B	t-statistic	p-value
(0, 1)	[403.81952235 2974.83326049]	[387.24994613 8942.74825175]	-15.08984108	1.88896718e-51
(0, 2)	[403.81952235 2974.83326049]	[506.98787975 1396.21278721]	6.95376467	3.55664716e-12
(0, 3)	[403.81952235 2974.83326049]	[328.08642677 4930.37718531]	-9.2150282	3.11196488e-20
(1, 2)	[387.24994613 8942.74825175]	[506.98787975 1396.21278721]	13.87458329	9.03164512e-44
(1, 3)	[387.24994613 8942.74825175]	[328.08642677 4930.37718531]	7.2656593	3.71223161e-13
(2, 3)	[506.98787975 1396.21278721]	[328.08642677 4930.37718531]	-11.17252743	5.55766007e-29

The first graph exhibits k-means clustering to cluster states based on the values of B_after and the average of size_rank. B_after represents the relationship between the month and the sale price of a property after the pandemic and size_rank is a metric used to compare the sizes of the cities in the dataset. This was used to identify distinct groups of states based on their property price trends during the pandemic. The graph indicates 4 clusters in the data. To determine whether the clusters were statistically significant we ran pairwise Z-tests to compare the slopes of each cluster. Each of these p-values for the pairwise tests were statistically significant, so we can conclude that there are indeed 4 main clusters of property price trends during the pandemic. The second graph illustrates the p-values and t-stats for the Z-tests to determine whether the results were statistically significant.

Conclusions

From our analysis we found that there was a relationship between the number of houses for sale and the profit received. To further look into this relationship and how it may have changed during the pandemic, we found that the prices of houses increased at a greater rate before rather than after the pandemic. In terms of where people were migrating to, people moved to smaller cities. Property prices fell in large cities and some small cities, and they rose in other small cities and somewhat large cities.

Challenges

Some challenges we faced came from data collection. For some metrics such as sale price and list price of properties, there were not values available consistently for each of the months we performed an analysis and thus, those months were dropped or zeroed during analysis. Other challenges included adhering our analysis to the cities that were included in the dataset, and thus, not every major city is represented in the data. For example, the linear regression has values represented from certain cities that are not represented in the k means clustering, reducing the consistency of the analyses.

Ethics

The data, sourced from Zillow, may contain historical or societal biases. For example, the data may not represent different geographic regions equally, potentially overlooking trends in marginalized communities. This could affect the analysis of the data by leading to incomplete or skewed insights. Similarly, the data collected by Zillow may not capture the heterogeneity of the housing market, so groups, such as low income individuals, may be underrepresented in the dataset and limit generalizability.

Acknowledgements

We would like to thank our mentor TA Ben Shih for all his assistance through this project.