

Correlation Investigation on House Prices and Covid Confirm Cases in Polk, Florida

Matthew Chau
DATA512
University of Washington

Introduction

During the pandemic, the housing prices surged in the United States and presently it is still showing the sign of further uprise, despite the fact that a lot of rents were reduced in some regions. It has always been rumored that the house prices are related to the pandemic, and understanding the relationship behind could be beneficial for various groups of people. For homeowners, this could help them gain insights on their personal wealth and better manage their cost of livings. For investors, the relationship between the pandemic and house prices could be a significant factor when making investment decisions. Moreover, if there does exist a relationship between pandemic and house prices, it could say a lot of things about the economy as the house prices both reflect the ongoing economic/political policies, and signal potential future policies. Thus in this analysis, I would be investigating the relationship of the pandemic and the housing prices in Polk, Florida.

Background/Related Work

There have been analysis on the pandemic and the house prices and how pandemic affects the houses. Duca (2021) stated that the pandemic has had an impact on the supply and demand of houses, as well as the momentum. The pandemic led to an uprising unemployment, low mortgage rates and accomodative monetary policies, which eventually raised the demand for housing. Moreover, the supplies had been constrained, such as disruptions of building supply chains and restrictions on work practices, leading to a decreased supply. Both factors would result in a uprising price, according to the fundamental economic theoroies.

To further justify it, I would be working on this project with the following research question: does the pandemic have an impact on the housing prices of Polk, Florida? If so, how was the local housing prices related to the pandemic?

Methodology

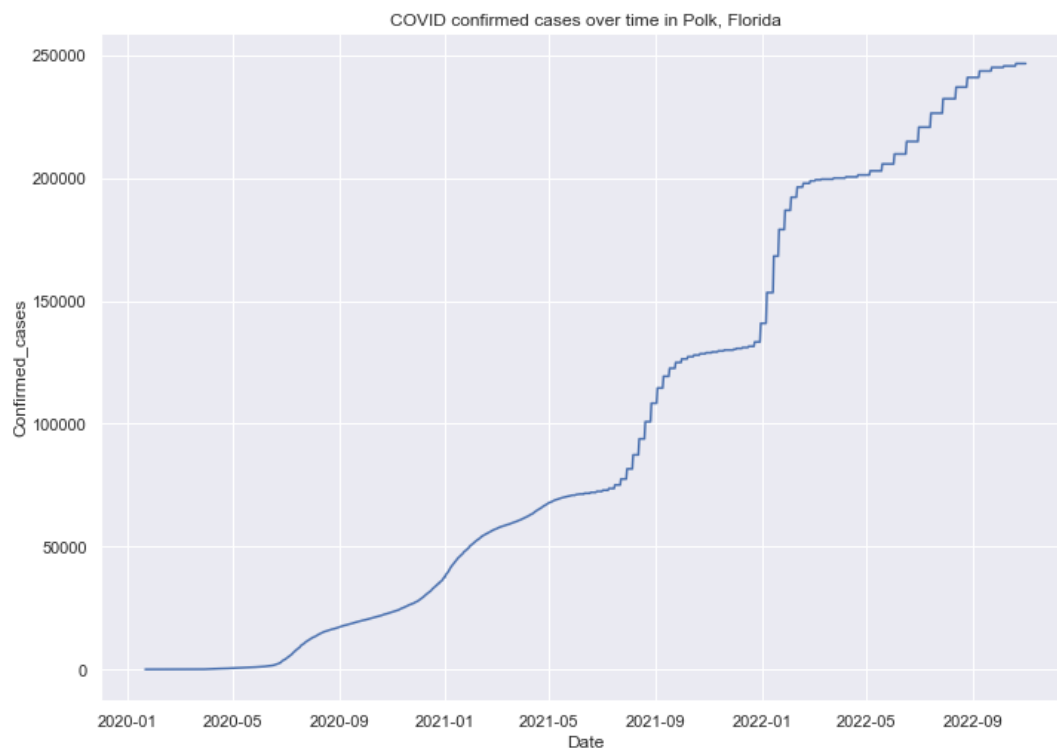
Three methods have been used in this study, including an autocorrelation analysis of the Polk Confirmed cases, linear regression analysis on the time and House prices, and a cross correlation analysis for the Covid confirmed cases in Polk and the house prices. The autocorrelation analysis was conducted as masked policy data was not available for Polk, Florida, and it'd be reasonable that the pandemic could be influencing itself as the higher number it gets, the more awareness it raises in the public. The linear regression was selected because the housing prices have a strong linear relationship, and it'd be useful to use the slopes for a comparison before and after covid. The cross correlation was done as I'd like to find out if there is a relation between Covid confirmed cases and house prices, and how does the correlation varies across different lags of days.

The nature of this research determines there would be human centered considerations associated, since this research looks into the economy and social phenomenon of the US. First of all, there would be social biases related to data retrieval. We did not consider the attributes of the houses, nor the newly built houses into account. Moreover, the covid data could be falsely reported due to political reasons. Some countries report a higher number to raise awareness among the public, and some countries report a lower number to suppress public panic.

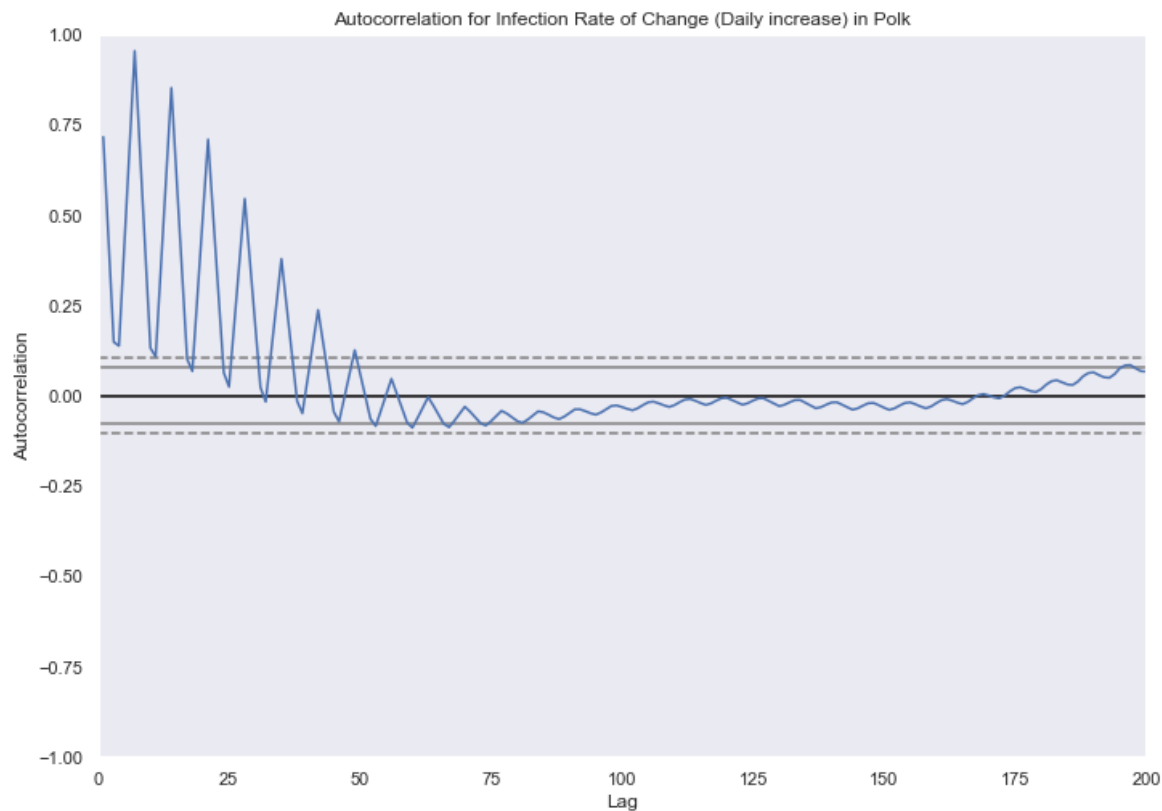
There might also be ethical concerns regarding this topic, as of how the pandemic could bring profits to people. If there is indeed a relationship between the confirmed cases with the house prices, it'd be dangerous if people would like to manipulate the house prices with Covid. It's been reported in some countries that medicine and vaccination companies profited from falsely reporting Covid cases so that they could sell more of their products. That would put people's healths at risk if people with capabilities to spread Covid for potential profits. As a result, we would only be looking at the correlations between Covid and house prices, but not the causational relationships.

Findings

First, we visualized the number of confirmed cases in Polk, Florida and we could see how it grew from the beginning of 2020 till present. As we could notice, the curve went steep and flat sometimes, and it is still rising until now.



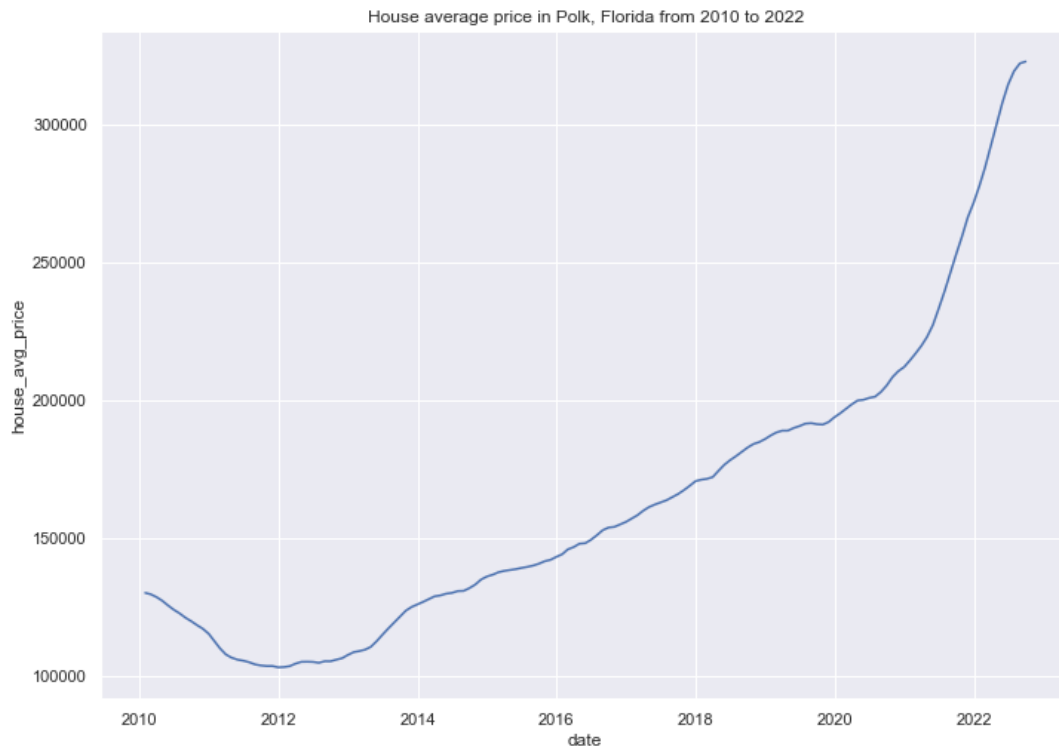
We created an autocorrelation graph below, showing the autocorrelation of the Infection Rate of Change in Polk, Florida, trying to find the impact of Infection Rate on itself.



Next, we investigated the relationship between the Covid Confirmed Cases in Polk, Florida with the local average house prices. The housing data is collected from the US Housing Market Dataset from Zillow: <https://www.zillow.com/research/data/>. The home type of “ZHVI all homes time series, smoothed, seasonally adjusted” is selected and geography location is specific to county. According to the description on the webpage, ZHVI is “a smoothed, seasonally adjusted measure of the typical home value and market changes across a given region and housing type”, and “reflects the typical value for homes in the 35th to 65th percentile range”.

Row index	Date	House average price
0	2000-01-31	101224.0
1	2000-02-29	101585.0
2	2000-03-31	101858.0
3	2000-04-30	102091.0
4	2000-05-31	102364.0
...
268	2022-05-31	307742.0
270	2022-07-31	319160.0
271	2022-08-31	321991.0
272	2022-09-30	322653.0

The above table shows the average house prices for each month in Polk, Florida after some data cleaning, filtering and pivoting steps and it's been visualized in the line chart below.



As we could see there is slope difference between before the middle of 2020, and after the middle of 2020. We would like to know if there is a statistically significant difference between the before and after covid outbreak, to see if covid did have an impact on the rate of change in house prices.

We have filtered out two datasets from the Covid 19 Data: before the outbreak in Polk and after the outbreak in Polk. The first outbreak was occurred on 2020.03.17, and the data available for house prices only contains till 2022.09.30. Thus we have taken 2022-03.17 – 2022.09.30 as our data for after covid, and the data from 2017.09.02 – 2020.03.16 as before covid. The two datasets are ensured to have the same size for consistency.

Before testing slope difference before and after covid, linear regression was used to find out the slope of the two trends and the strength of the slope. The linear regression results are in the table and charts below:

data	slope	y-intercept	R2	std err	T value	P value
before covid	33.91	168003.41	0.96	31.136	10.107	< 0.001
after covid	151.56	177446.25	0.94	31.388	13.965	< 0.001



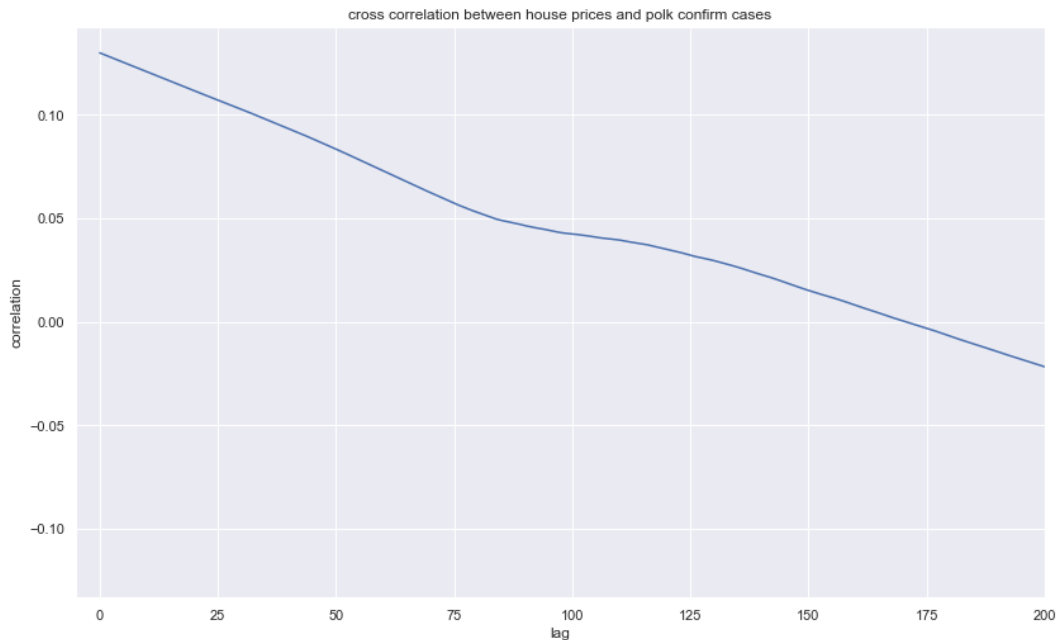
We compared the two slopes with a t-test and got the following results.

Test Statistic	Value
t-score	2.66
Degrees of Freedom	57
p-value	0.0101

Next, we have also looked into the correlation between house prices and the Covid-19 confirmed cases. However, since the house prices are recorded every month, we have estimated the daily house prices with interpolation. A cross correlation graph is then created to examine the correlation of house prices and Polk confirm case with different days of lag.



However, the Polk confirm cases might not be sufficient to demonstrate the relationship with the house prices. Therefore, we have also plotted a cross correlation graph of the rate of change of Covid cases and the house prices.



Discussion/Implications

From the findings above, we are able to analyze on the results. From the autocorrelation chart, it shows that the correlation of the daily increase data with itself with a time lag. The x-axis represents the size of the lag in terms of days. For example, a 25-day lag compares the data[0:n-25] with data[25:n], indicating whether there could potentially be an impact of the past data on the future. The y-axis, on the other hand, represents the strength of the correlation. The closer the values are to 1 and -1, the stronger the correlation, and the closer the values are to 0, the weaker the correlation is. The horizontal lines represent the 95% confidence interval and the dotted line shows the 99% confidence interval. That is, for lag 1 to 25 days, there is a strong evidence that there is some level of autocorrelation. This implies that the daily increase of the past 1-25 days are highly correlated to the daily increase of the present.

The t-test for the slope comparison shows a statistically significant difference between the slopes before and after Covid, with a t-value of 2.66, degree of freedom 57 and p-value 0.0101. That indicates that there is indeed a difference between the increase rate of house prices before covid and after covid. The cross correlation shows a very high correlation between covid confirm cases and house prices, and the decrease as lag goes up almost perfectly linearly. This shows a very strong correlation between the uprise of Covid confirmed cases in Polk and the house prices. However, the correlation between the increase rate of Covid confirmed cases in Polk and the house prices are very little, starting from above 0.1 to below, which indicate that the house price does not have a micro level relationship with Covid. Instead, it's more likely have a more higher level relationship – the existence of Covid could be the trigger, instead of the increase rate.

Limitations

This is a required section for your report. There are often many, many limitations for any study. If you honestly tried to list them all, this might end up being the longest section. You should prioritize and list the ones that are most likely to have a significant impact on your results. Specific license issues could be a limitation, depending on what data sources you used. Flaws in the data, data cleaning techniques, potential assumptions and/or how you handled missing values could be a limitation. Statistical techniques often have specific assumptions and preconditions; if you're not certain all of the data meets those requirements - this is a good place to make that clear.

This is a very preliminary investigation and suffers from several flaws, listing from the most significant to the least:

- The correlation does not mean anything causal, and is insufficient to state that the Covid-19 drives the house prices up even from the fact that they have an exceptionally strong correlation. The correlation could be a result from a common cause, which leads to both the Covid 19 and house prices surging, or it could be simply coincidental.
- As mentioned before, the Covid-19 data has a potential to be biased. The number of confirmed cases could be falsely reported, both higher or lower. The accuracy of the data cannot be validated, thus we could only analyze the results based on what we have in the data.
- The data used to calculate the correlation is an estimation based on the interpolation results, and could have a slight impact on the actual correlation values.

Conclusion

Restate your research questions/hypotheses and summarize your findings. Explain to the reader how this study informs their understanding of human centered data science.

From the above findings, we are able to make two conclusions. First, the increase rate of housing prices in Polk does have a statistically significant difference between Covid and after Covid. Second, although the change in housing prices have a very strong correlation with the Covid confirmed cases in Polk, Florida, the rate of change in confirmed cases does not have a strong correlation with the house prices. It's more likely that Covid itself affects the house prices, rather than the how fast it grows.

The results suggest that there is a high chance that COVID 19, although might not be the primary one, could be a trigger of the surge in the house prices. If further analysis show that this is the main cause, it could imply that similar pandemics could have similar impacts on the house prices as well. Also, the results bring insights to investors and house owners in Polk, since the strong correlation could imply that there's a high possibility that when confirmed cases increase rate decreases, the house prices would have a lower increase rate or even decrease.

References

Duca, J. V., & Murphy, A. (2021, December 28). Why House Prices Surged as the COVID-19 Pandemic Took Hold. Dallas Fed Economics. Retrieved December 12, 2022, from <https://www.dallasfed.org/research/economics/2021/1228.aspx>

Data Sources

The Covid-19 cases data: <https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university>

Zillow Home Value Index (ZHVI): <https://www.zillow.com/research/data/>