Data 512: Human-Centered Data Science
Juhi Choubey

Impact of Covid-19 on the Housing Market in Multnomah County, OR

Final Project Report

1. Introduction

The objective of this project is to assess the impact of covid-19 cases on the real estate and housing market in Multnomah County, OR. The duration of this analysis will include a study between 2019-2021 to accommodate for the effect of time before and after a covid-19 surge.

The real estate and housing market is subject to many ups and downs - and largely depends on the standing of customers (humans) either through political amendments, the economy, the spread of diseases (in this case covid-19), or many other human-driven factors. Not just that, the relative fluctuations in real estate are in turn a cause of stir among the general public - and not just people of one socio-economic class, but from every class. Through this study, I aim to assess any connection between the rise/fall of confirmed covid-19 cases and the subsequent variation in the real estate market (housing prices in particular). While a general trend is a base observation for housing prices, I hope to detect any specific/unusual fluctuations that can be connected to the pandemic in a certain way.

This problem is particularly human-centered due to the direct impact on human society and the large population. Though, in this particular analysis, we will be narrowing the scope to Multnomah County, OR only, and for a fixed duration too (2019-2021). With this analysis, I hope to attempt to provide some insight into the impact of rising covid-19 cases on the prices of houses to the general public. I hope to make this analysis intuitive and reproducible to the people interested in understanding the correlation between the covid-19 surge and the housing market, so they can make better decisions.

2. Background/Related Work

An article was published in Forbes magazine on April 27, 2021 (link in the References section), titled ‘The Covid-19 Pandemic has fueled a crisis in the housing market’ that explored the various disastrous effects the pandemic has had on housing prices throughout the world. The claims were backed with some major key statistics including rising median house prices, the decline in total inventory (supply) of the houses, and a slight increase in the availability of new listings. This article is the base groundwork of this project and I hope to test these claims using the procured datasets (links in the Data Sources section).

Moreover, an economic journal published by the federal reserve (link in the References section) on July 08, 2021, reviewed the demand-supply dynamics of the housing market during covid-19 and provides an extremely detailed view of the housing market and proposes a model to determine the fluctuation in the housing prices using monthly log change of median housing prices data from Redfin. For the sake of simplicity and to avoid diving deep into the economic

intricacies of the subject, I have refrained from the usage of heavy economic factors. Considering this, the above journal can be utilized as a future work or build up on the current analysis.

After reviewing the above articles, the preliminary research questions I plan to assess from the dataset and answer are as follows:

- ➔ Was there a statistically significant difference between housing prices in 2020 and 2021 (during the pandemic) as opposed to 2019 (prior pandemic) in Multnomah County, OR?
- ➔ Did the number of confirmed cases have a statistically significant impact on the housing prices between 2020-2021 in Multnomah County, OR?

For a thorough and complete inference, there are many considerations that must be looked at, including political and economic effects, covid-19 severity (not just confirmed cases, but deaths and hospitalizations as well), etc. Due to the lack of availability of such detailed data sources, appropriate assumptions have been made and highlighted throughout the analysis (including the limitations section).

3. Methodology

The approach to investigating the answers to research questions has been broken down into specific details and drafted into a summary table below. Any other nuances and human-centered considerations for this study have also been included in this section.

I. Data Cleaning:

For all required data files:

a. Filter for the county: Multnomah, OR, and Time: 2019-2021
b. Discarding records with null values
c. Rollup #confirmed cases/deaths at the weekly level
d. Merging the datasets to prepare 1 master file containing covid-19 metrics and housing prices.
e. Feature Extraction - Infection rate, Change in #cases (weekly)

Human-Centered Considerations: These data cleaning steps have been performed in Python Jupyter notebook which has in turn been uploaded on GitHub. It is a public repository accessible on the internet as the data is also publicly available from Redfin and does not violate any privacy ethics. Additionally, the data cleaning steps are documented in a user-readable format and are completely reproducible ensuring

II. EDA (Visualizations):

Review trends for Multnomah County, OR between 2019-2021:

a. The weekly trend of Median Housing Prices vs Housing Inventory
Through this graph, we aim to review if there were any detectable fluctuations in supply-demand dynamics, especially before and after a covid-19 surge.

b. The weekly trend of #Confirmed Cases and Median Housing Prices

In case there were any detectable fluctuations observed from (a), it would be interesting to see how these fluctuations behave during 2020-21. While from (a), we hope to check if covid-19 surge had any impact on supply-demand dynamics, from (b), we hope to observe 'how' these dynamics fluctuate with the course of covid-19 spread.

c. Weekly progression of #Confirmed Cases and Interest Rates
While this analysis attempts to observe any correlation between rising covid-cases and median housing prices, we also hope to address any possible influencers that partake in the fluctuations of supply-demand dynamics. Changes in Interest Rates are a direct player in this behavior and it would be interesting to see how this influencing variable thrives between 2020-21 during the surge of covid-19.

Human-Centered Considerations: Similar to I, the EDA and visualization code steps are documented in a user-readable format and are completely reproducible ensuring the key principles of human-centered data science.

III. Research Methodology:

- Research Question I: No difference b/w Median Housing Prices between 2019-21

  Research Method: Trend Chart Visualization

  Rationale: Viewing a trend chart is the first and foremost step to detect any fluctuations in a metric. Through this chart, we aim to review if there were any detectable fluctuations in supply-demand dynamics, especially before and after a covid-19 surge. In case the change is not clear via visualization, further attempts at any statistical tests are suggested.

  Human-Centered Consideration: An important consideration here has been viewing trends in Housing Inventory & Interest Rates along with Median Housing Prices. While the above research method alone is sufficient to detect any fluctuations in Median Housing Prices, the inference will be incomplete if the supply & interest rates in the housing market are not considered. This will in turn affect the overall understanding of the Housing Market and provide an incorrect inference of the evidence to the general public, jeopardizing the intent of the analysis.

- Research Question II: No impact of #confirmed cases on Median Housing Prices

  Research Method: ARIMA model

  Rationale: Median Housing Prices across the years are not just dependent on implicit economic factors, but also on their own value in history. Using this literary claim, we develop an auto-correlation model of median housing prices. But this time, we condition this mathematical equation on #confirmed cases. To study this model, we would be interested to review the ARIMA coefficient & standard error associated with this coefficient of #confirmed cases. This coefficient will enable us to argue for the research equation and the p-value will be most indicative of the rejection of the research

hypothesis.

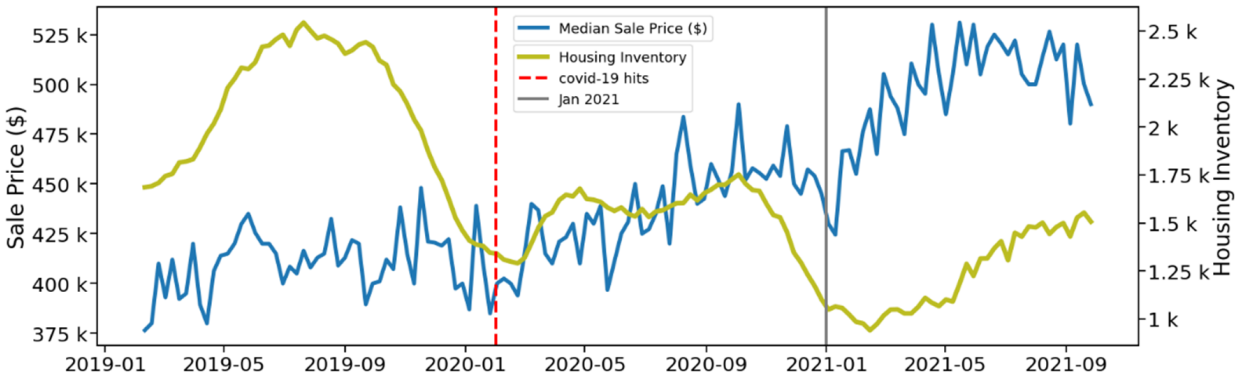
Human-Centered & Other Considerations: It is important to consider all the aspects whether included or not within the proposed ARIMA model. A couple of metrics such as interest rates have not been considered in the model as they are already indirectly flowing in through the auto-correlation lag component of Median Housing Prices. More has been explored in the limitations section and the inference from this model will be built upon in a further section while taking into such nuances and considerations.

4. Findings

The observations and findings from the above-proposed methodology are discussed below. In this section, an attempt has been made to answer the primary research hypothesis. Considerations and other nuances associated with the inference of these results will be highlighted in the discussions section.

a. The weekly trend of Median Housing Prices vs Housing Inventory in Multnomah County, OR between 2019-21:

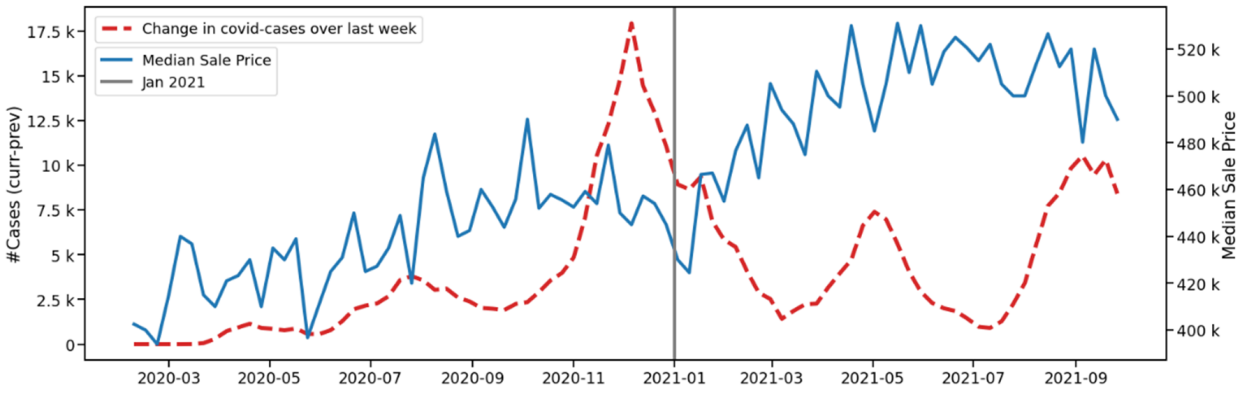
The graph below presents the week-over-week trend of Median Housing Price ($) and Housing Inventory between 2019-21. The onset of covid-19 in Feb 2020 has been marked in red while another important time - the onset of Omnicron in Jan 2021 has been marked in gray. These timelines coincide with highly noticeable fluctuations in both Median Housing Prices & Housing Inventory - showing that in both these times, the median housing price sharply increases, while the housing inventory plummets. This is highly indicative of a connection between covid-19 surge times with fluctuations in the housing market.



b. The weekly trend of Median Housing Prices ($) and weekly change in #confirmed cases in Multnomah County, OR between 2020-21:

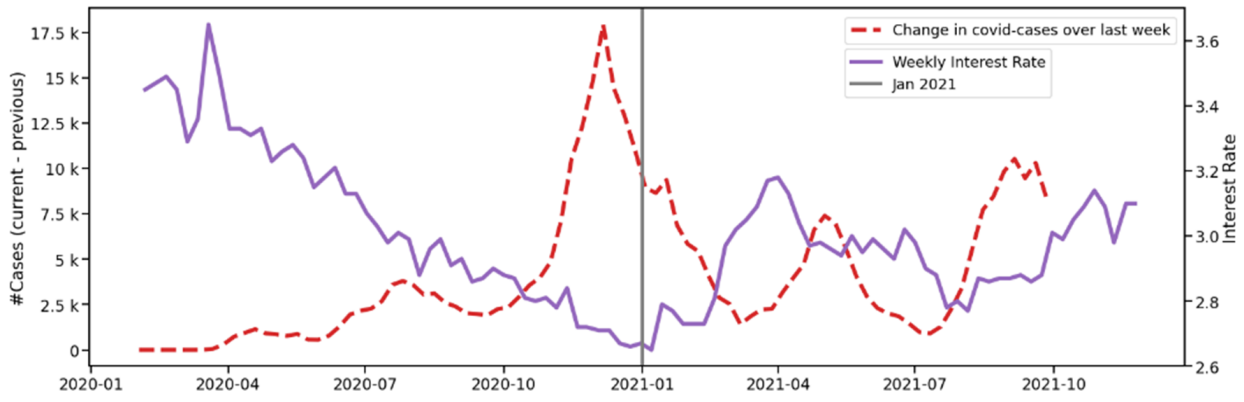
The below graph builds up on the previous graph to highlight ‘how’ the covid-19 surge has affected the median housing prices. Instead of using the raw #confirmed cases, we proceed to pick an improved metric - week-over-week change in #confirmed cases - to observe any hidden patterns associated with the size of the covid-19 surge. The surge of Omnicron variant around Jan 2021 sees the most increase in weekly covid-19 cases.

Interestingly enough, in both the previous and current graphs, we observe a considerable increase in Median Housing prices. Moreover, during this exact time, we also observe the total housing inventory plummeting so sharply that it completely shifts the overall trend.



c. The weekly trend of Interest Rates and weekly change in #confirmed cases in Multnomah County, OR between 2020-21:

We assessed the trend of interest rates as well during the set time to view if it fluctuated with rising and falling covid-19 cases. Until Jan 2021, while the covid-19 rose rapidly, there was a steady fall in the interest rates, however, it did begin recovering after Jan 2021 - after the highest increase in covid-19 cases was recorded. It can be seen that the interest rate sort of follows an inverse relationship with rising covid-19 cases through the graph below.



d. Results of ARIMA model of Median Housing Prices conditioned on #confirmed cases in Multnomah County, OR between 2020-21:

The summary of ARIMA model results is provided below. The model was built to explain Median Housing Prices ($) as a function of an auto-correlation component of Median Housing Prices ($) along with #confirmed covid-19 cases (as a covariate). The coefficient of the covariate in the proposed ARIMA model is 0.2665 with a standard error

of 0.029 and a p-value close to 0. The auto-correlation component given by lag 1 of the Median Housing Prices is 0.5070 with a standard error of 0.097 and a p-value close to 0.

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	87
Model:	ARIMA(1, 0, 0)	**Log Likelihood**	-970.323
Date:	Mon, 05 Dec 2022	**AIC**	1948.646
Time:	21:26:51	**BIC**	1958.510
Sample:	0	**HQIC**	1952.618
	- 87		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
const	4.273e+05	5464.009	78.200	0.000	4.17e+05	4.38e+05
x1	0.2665	0.029	9.324	0.000	0.210	0.323
ar.L1	0.5070	0.097	5.244	0.000	0.318	0.696
sigma2	2.812e+08	0.096	2.92e+09	0.000	2.81e+08	2.81e+08

Ljung-Box (L1) (Q):	0.43	**Jarque-Bera (JB):**	0.33
Prob(Q):	0.51	**Prob(JB):**	0.85
Heteroskedasticity (H):	1.05	**Skew:**	0.11
Prob(H) (two-sided):	0.91	**Kurtosis:**	2.80

Other factors such as masking mandate and interest rates have not been included in the model to avoid problems of confounding variables and multi-collinearity. However, there are other factors too at play that can affect the understanding of the results or even the results themselves. These considerations will be highlighted in the limitations so it it important to read the results keeping these considerations in mind.

5. Discussion/Implications

The sharp decline in total housing inventory during the two important timeframes namely, the strike of covid-19 in Feb 2020 and the surge in cases around Jan 2021 - suggest an important link between the supply of the housing market and covid-19 crisis. From graph (a), we observe that the decline is so steep that it completely moves the mean trend of the supply-demand cycle in the housing market. It is interesting because it does not simply lay concerns over rising Median Housing Prices, but actually goes above that to show fluctuations in total housing inventory aka

supply - prompting further investigation into how multiple facets of the housing market were shaken up due to covid-19. This helps provide better clarity on the real estate situation - to both home buyers and sellers - and provides a better picture of the situation so they may choose to wait or go right ahead with their decisions.

The analysis not only probes into the exact time frames (Feb 2020 & Jan 2021), but also maps these with the week-over-week changes in covid-19 cases. This is again super-interesting because it provides a solid background and rationale behind these two-time frames - informing home buyers/sellers of the fluctuations in the housing market immediately after the high surges in covid-19 cases. While the rise in Median Housing prices and decline in total Housing Inventory were consistent for certain durations between 2020-21, the major movements that led to displacing the overall trend occurred in these two exact time frames. This finding is very important to account for as it puts into perspective a lot of the fluctuations in the housing market during covid-19.

The summary statistics derived from the results of the ARIMA model, particularly the coefficient of #confirmed cases within the equation go on to show that there has been a positive association between Median Housing Prices & #Confirmed Covid-19 cases. However, the implications of this inference must be accounted for with a grain of salt. There are primarily two reasons for this: First, there are many other factors at play that can be considered in the model and may alter the results drastically. By considering an auto-correlation component of Median Housing prices, we have attempted to include the indirect effects of several of these factors and thereby make the inference more reliable. However, these indirect variables have not been explicitly studied, so it makes our inference malleable to external implications.

While the trend study is not sufficient alone to imply any causality, it does support the claim laid in Forbes magazine (as mentioned in the background section) with transparency in both data and analysis. This builds on the key principles of human-centered data science by not only providing analysis transparency and reproducibility but also aligning with ethical concerns and privacy by not utilizing any person-specific data or making it public.

6. Limitations

There are definitely other factors at work in explaining the variability within housing prices in Multnomah County, OR between 2019-2021, such as political and economic amendments, increased rental prices, lockdown restrictions leading to remote work, covid-19 severity (not just a rise in confirmed cases, but deaths and hospitalizations as well), fluctuating interest rates, etc. Due to limitations on both time and data availability, this project will narrow the scope to assessing the impact of rising/falling covid-19 confirmed cases on housing rates, making necessary assumptions for other factors.

Additionally, there had been a consideration to include available factors such as masking mandate and the proportion of the masked population. However, including these co-variates with the number of confirmed cases would have led to the bias of confounding variables within the ARIMA model, and hence this consideration was discarded. A further deep-dive into confounding variables was needed for this to work - and it can be implemented as a future scope in this work.

Moreover, the housing data was available at a monthly level as well. Grouping the confirmed cases dataset at a monthly level and then performing this entire analysis at a monthly level instead of a weekly was another consideration. However, the interest rates data was only available at the weekly level. Performing any aggregation on the interest rates dataset to arrive at a monthly granularity would have involved a lot many assumptions. To avoid including further assumptions, the analysis has been carried out at weekly granularity only. A monthly analysis of the impact of covid-19 on housing data can be considered a scope for future work.

7. Conclusion

In this project, we have reviewed the popular claims laid by certain articles around the impact of covid-19 on the housing market. This study has been performed only for Multnomah County, Oregon. We have determined that there has been a significant difference between housing prices in 2020 and 2021 (during the pandemic) as opposed to 2019 (prior pandemic) in Multnomah County, OR with the help of trend chart visualization using data from Kaggle & Redfin. We have also marked significant time frames during this view to characterize the exact points in time that were associated with unusually large increase and decrease. These timelines were specifically the ones with the beginning of covid-19 (in Mar 2020) and the influx of the omicron variant in Jan 2021. Moreover, we built an ARIMA model (AR0) to explain median housing prices conditioned on lag 1 of median housing prices and #confirmed covid-19 cases. With this, we were able to confirm that the number of confirmed cases have a statistically significant positive impact on the housing prices between 2020-2021 in Multnomah County, OR. This conclusion has also been accredited with multiple caveats that have been outlined in the limitations section so as to prevent unjust inferences and conclusions.

This project lays a special emphasis on human-centered data science by invoking and implementing multiple key principles of human-centeredness that govern a data science project. Starting from the very basics, care has been taken to prepare this study in the most explainable and reproducible manner for the general audience. The data sourced and shared for this study is ensured to not leak any personal/private information of any individual - privacy ethics is followed not just with data but also with the analysis. Moreover, it is important to present observations and inferences in the most wholesome manner that takes into consideration various caveats and nuances. In this study, we have outlined many such considerations that hold importance and potential to skew the results.

8. References

- https://www.forbes.com/sites/saibala/2021/04/27/the-covid-19-pandemic-has-fueled-a-crisis-in-the-housing-market/?sh=3de63f635928

- Anenberg, Elliot, and Daniel Ringo (2021). "Housing Market Tightness During COVID-19: Increased Demand or Reduced Supply?," FEDS Notes. Washington: Board of Governors of the Federal Reserve System, July 08, 2021, https://doi.org/10.17016/2380-7172.2942.

9. Data Sources

There are multiple datasets to be used during the analysis covering both covid-19 cases and the real estate market in the specified duration (2019-2021). Below lists the source of the datasets and the link to download them. We are currently planning to use all the below datasets, but this is subject to change as the analysis progresses.

- ➔ The RAW_us_confirmed_cases.csv file from the Kaggle repository of John Hopkins University COVID-19 data: https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv

- ➔ The RAW_us_deaths.csv file from the Kaggle repository of John Hopkins University COVID-19 data: https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_deaths.csv

- ➔ The weekly housing market data from Redfin: https://redfin-public-data.s3-us-west-2.amazonaws.com/redfin_covid19/weekly_housing_market_data_most_recent.tsv

- ➔ Interest Rate (fixed for 30 years): https://fred.stlouisfed.org/series/MORTGAGE30US

The redfin weekly housing market data has data for each county on a weekly basis. Furthermore, the data is categorized by property type (All Residential, Single Family, Condo, Multi-Family, Townhouse, etc.) Redfin has published a data dictionary to understand every column and interpret them. This data set is licensed under Redfin's Terms of Use. The guidelines for using the data state to cite the data source appropriately and provide a link to Redfin.