Impact Analysis of Covid-19 on the Housing Market of Shelby County, TN

Introduction

For this project, I am analyzing the impact of covid-19 on the housing market of Shelby county, Tennessee. Housing market was one of the prime areas that faced the severe brunt of the pandemic combined with anxious buying, remote work, relocation to suburbs, piling up of recession, etc. Pandemic created uncertainty in the overall housing market by creating short periods of bumps and falls in the overall trend- in short, it was hard to understand and analyze where the housing market was going next!

This project is human-centered in two-fold ways- in terms of the causes, a sudden change in the lifestyle due to remote work combined with other factors like anxious buying from people, fluctuations in the mortgage interest rate, etc. created an environment of uncertainty that we all witnessed during the pandemic. And in terms of the effects, we still see how this has directly impacted us in all the aspects- be it emotionally, financially or even physically. With such fluctuations in the housing market, a vast majority of Americans who are/were looking to purchase homes are affected and with the housing prices just soaring up, it would also mean that a lot of people from lower-income brackets will suffer as they will not be able to afford a house, or will be pushed to the suburbs thus widening the wealth gaps between the rich and the poor. And as homeownership is an important tool for building long-term wealth and children of homeowners are likely to become homeowners, this trend can further exacerbate wealth inequality for future generations.

By performing this analysis I hope to learn if there was any correlation between the number of covid cases and housing prices. This analysis is of particular scientific interest as it will help others see trends between housing prices and a fluctuating economy during a pandemic.

Background/Related Work

As the pandemic progressed, the cases started increasing nationwide. The state and national government issued several lockdowns to control the spread of the virus, a lot of the businesses were shut and there was a widespread fear of a major recession coming up. Tracking changes in the housing market will provide us insights into the economy of Shelby county of Tennessee state, as it will help us to contextualize growth and decline, thereby giving an insight into the market. As part of this project, I plan to explore the following questions to help understand the trend in the housing market since the start of the pandemic:

- Did Covid-19 impact the housing market for Shelby county, TN with respect to housing inventory, sale price, number of listings from February 2020 to October, 2021? If yes, then how?
- How did Covid-19 impact the mortgage interest rates and how did that in turn impact the housing market in terms of homes sold and median sale price?
- How was the overall housing market trend impacted during the covid-19 from February 2020 to October, 2021? Did it move away from its original trend?

There are several interesting articles published about the impact of COVID-19 on the housing market. This Bloomberg article helped me understand how there are several factors that play a role when it comes to housing prices. Trends such as work from home (WFH) has resulted in people, especially young millennials relocating to the suburbs for cheaper housing. However, it has also led to a new term **'Rise of the Rest'** signifying increase in the cost of housing in the suburban areas during covid, thus, resulting in increased housing prices.

This Federal Reserve article on the link between covid-19 and the housing market also lists certain factors that influenced the housing market. During the start of the pandemic several people were reluctant to sell homes to prevent risk of getting covid-10 thereby creating a shortage in inventory. Similarly stay-at-home orders and lockdowns also impacted how many homes were sold during the pandemic. Due to these fluctuations and strong linkage to the number of covid-19 cases, it is important to analyze this market to see how the overall trends diverted from the projected housing market trends.

Methodology

In terms of my methodology for implementing this project, I have addressed a variety of human-centered data science considerations including :

1. **Licenses**: I have ensured that the research is licensed under a MIT license so as to allow free use of it for further research, analysis or expansion.

2. **Copyright**: I have appropriately cited all sources of data, tools, and inspiration (blogs, articles, readings, etc.)

3. **Interpretability**: I have followed a literate programming style to explain decisions made along the way that impact the final results of my analysis.

4. **Repeatability and Reproducibility:** I have taken extensive efforts to make my research reproducible by providing information about my environment, how I collected and used the data, including my data in my repository so that anyone could run my notebooks to replicate my results

5. **Interpretability**: I have used methods like Linear regression that are very transparent and easy to understand for the audience of the research. I have also used visualizations to convey the information in a simple and easier manner to promote this goal. I have also included enough information even in my notebooks to make everything interpretable to the users.

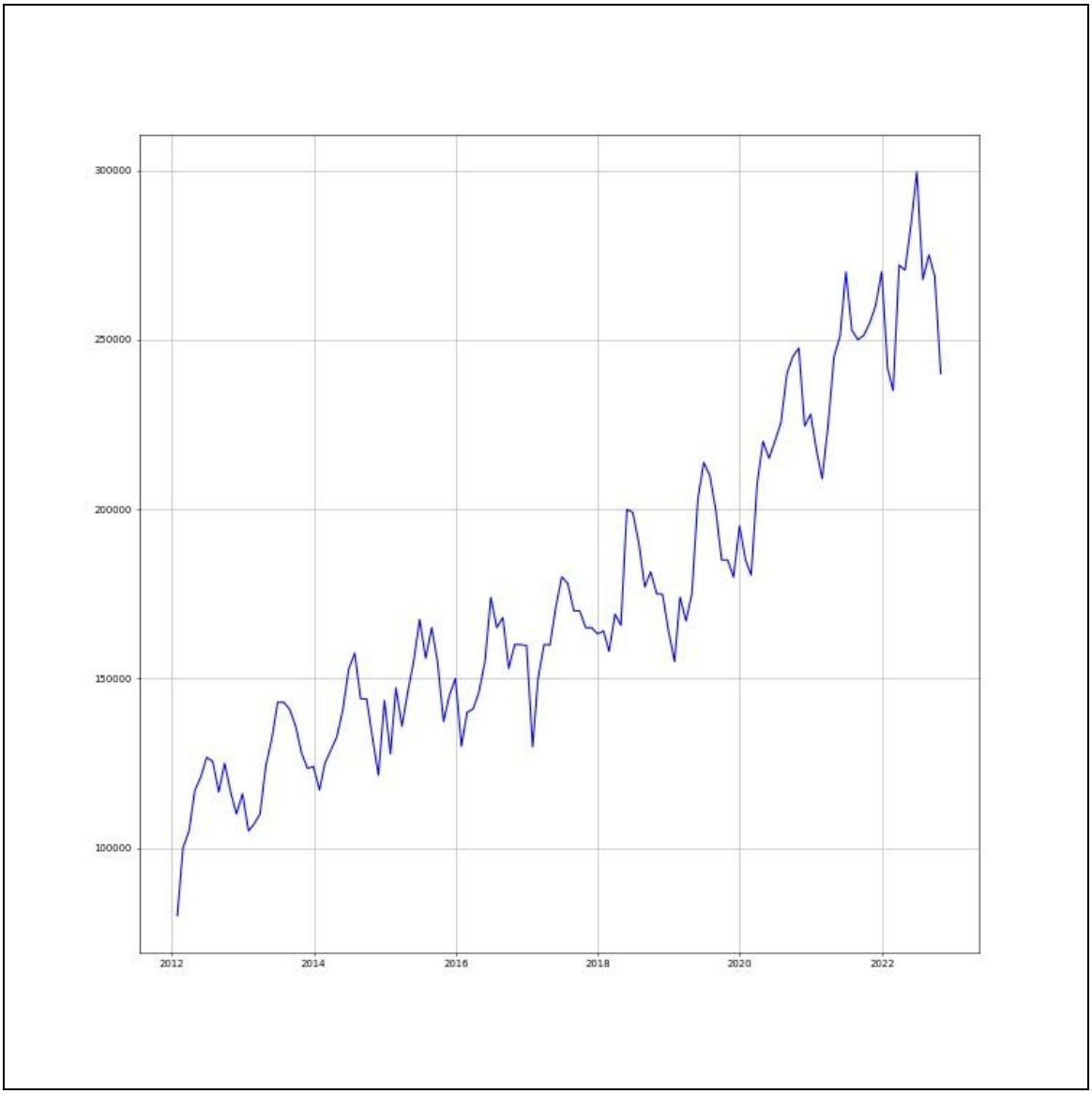
To further help in easy interpretation, I have broken down my overall analysis into the following 4 steps:

1. Data Acquisition - The data is downloaded from the below data sources (included in Data section) and saved to the raw_data folder for easier access in future to anyone reproducing the analysis. The data includes the covid data from Kaggle, the Weekly and Monthly housing market data from Redfin and the 30-year mortgage interest rate data.

2. Data Processing and Cleaning - Here I performed data processing to filter data for Shelby county, Tennessee and also filtered the data for data range (January 2020 to October 2021). This is also where I compute the 7-day moving averages to account for variation caused by weekends and public holidays. After performing the cleaning, I merge the housing data and covid cases data into a combined data set and save the results to the data-processed folder such that someone else can leverage this csv file directly for continuing the analysis.

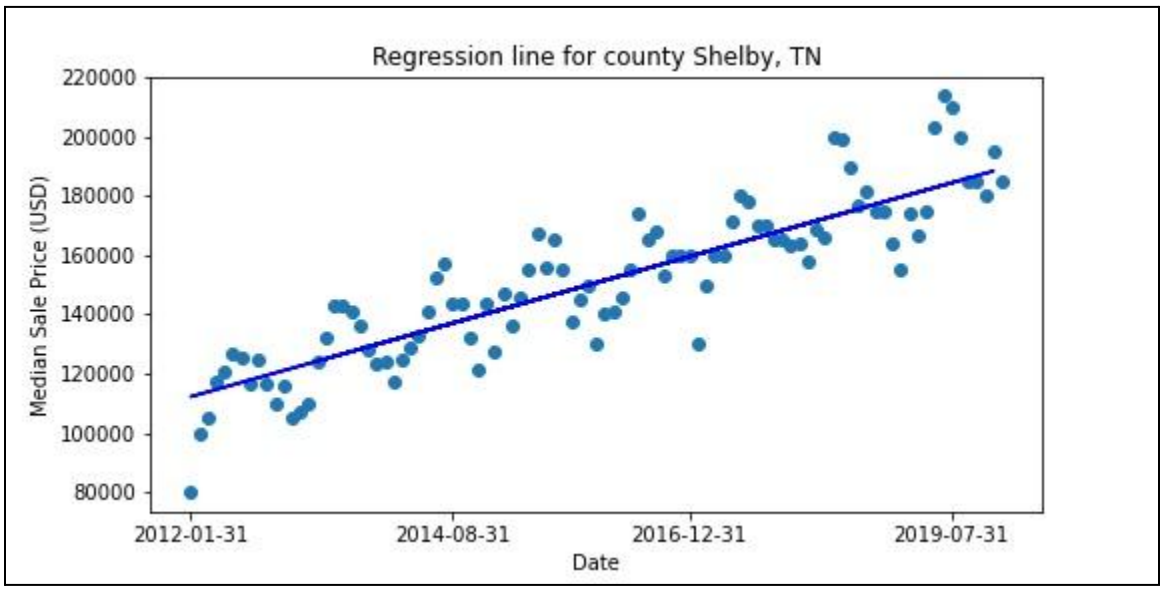
3. Data Analysis / Linear Regression: This is where I compare my KPIs which are housing inventory, median list price, number of new listings to the covid-19 cases. Comparing their behavior, I see how they are impacted. I also perform simple linear regression to predict housing prices for 2020 and 2021 and compare it with actual housing prices to see if there is a difference between predicted and actual housing prices for 2020 and 2021.

Linear regression suits best to find the relationship between a dependent continuous variable (Median Sale Price) and one or more explanatory independent variables (Month/Year). Linear regression suits best here because we can see a linear trend in the dataset for housing prices and housing prices are normally distributed.



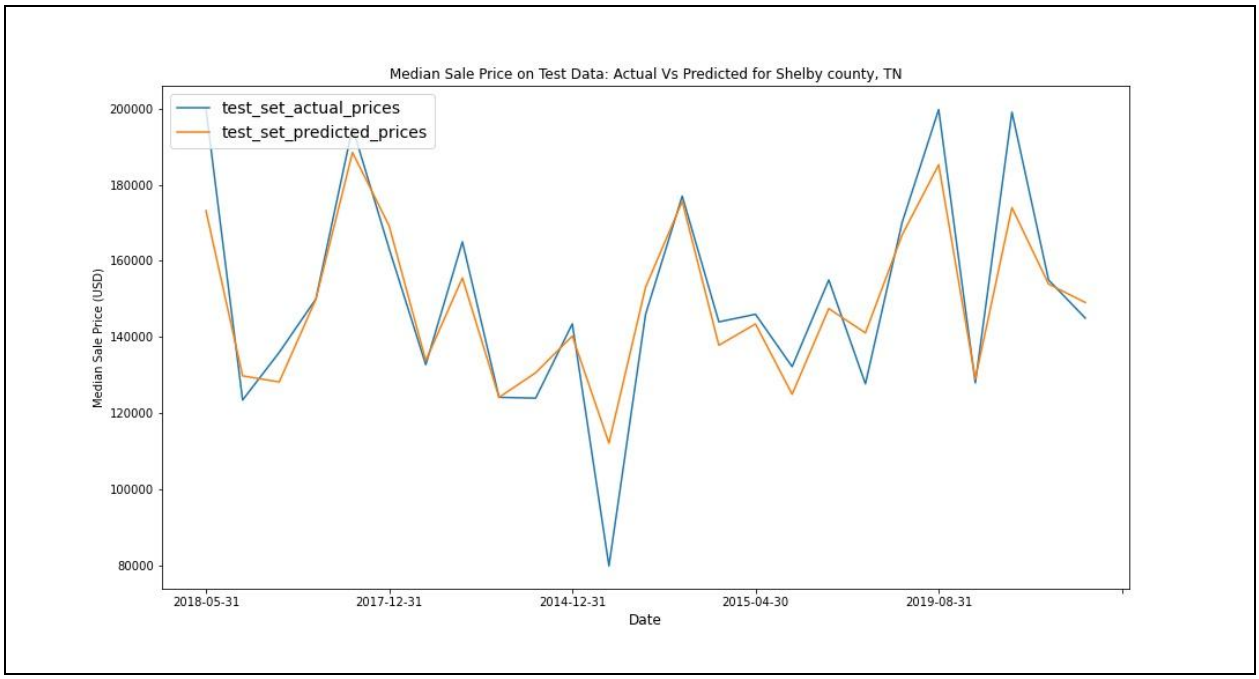
Median Sale Price in Shelby County from 2012-2022

I fit a univariate linear regression model using historical data from (2012- 2019) where the feature is the weekly dates and the target is the median housing price. To avoid the outlier years from the housing market crash in 2008, I have taken the data post that. I have split the data into 75:25 train-test split and fit a linear regression model using python scikit learn. I have also computed the RMSE (root mean square error) as a measure of model performance and used the model to predict housing prices for the years of analysis (2020 and 2021) to compare if the prediction is higher or similar to actual prices.



The model's performance on the test set is evaluated using the common evaluation metrics for regression problems

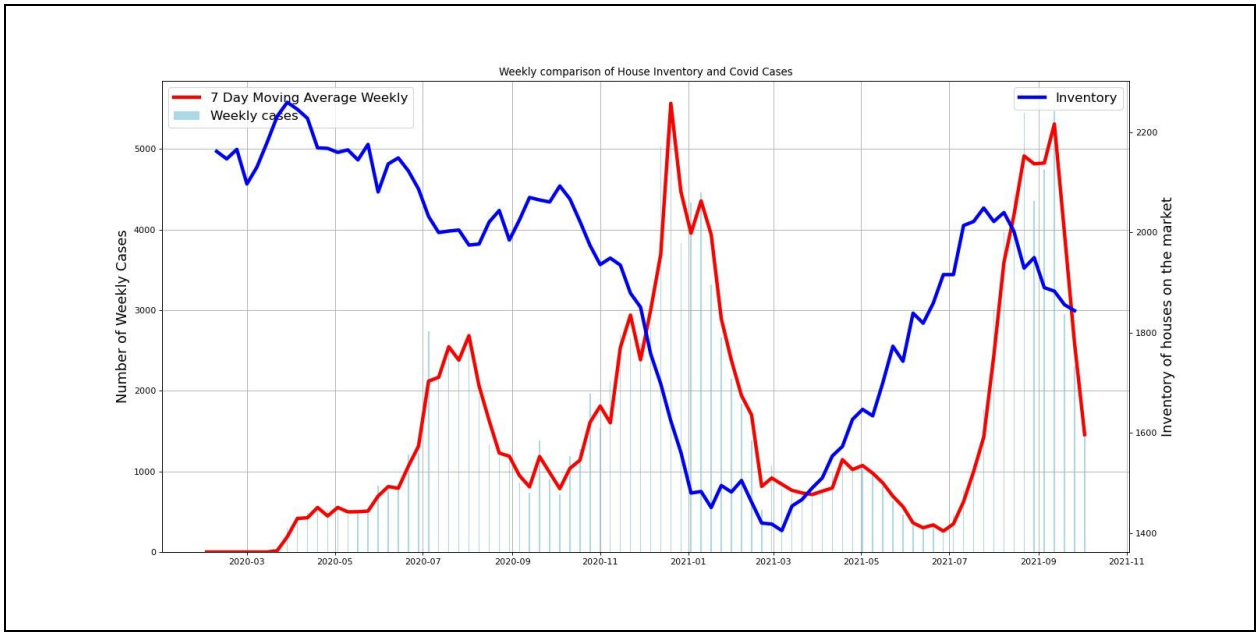
```
MAE, MSE and RMSE on Test Data for Shelby
Mean Absolute Error: 8012.865700163395
Mean Squared Error: 133009722.47508258
Root Mean Squared Error: 11532.984109721237
R2 Square 0.8322598010063076
```



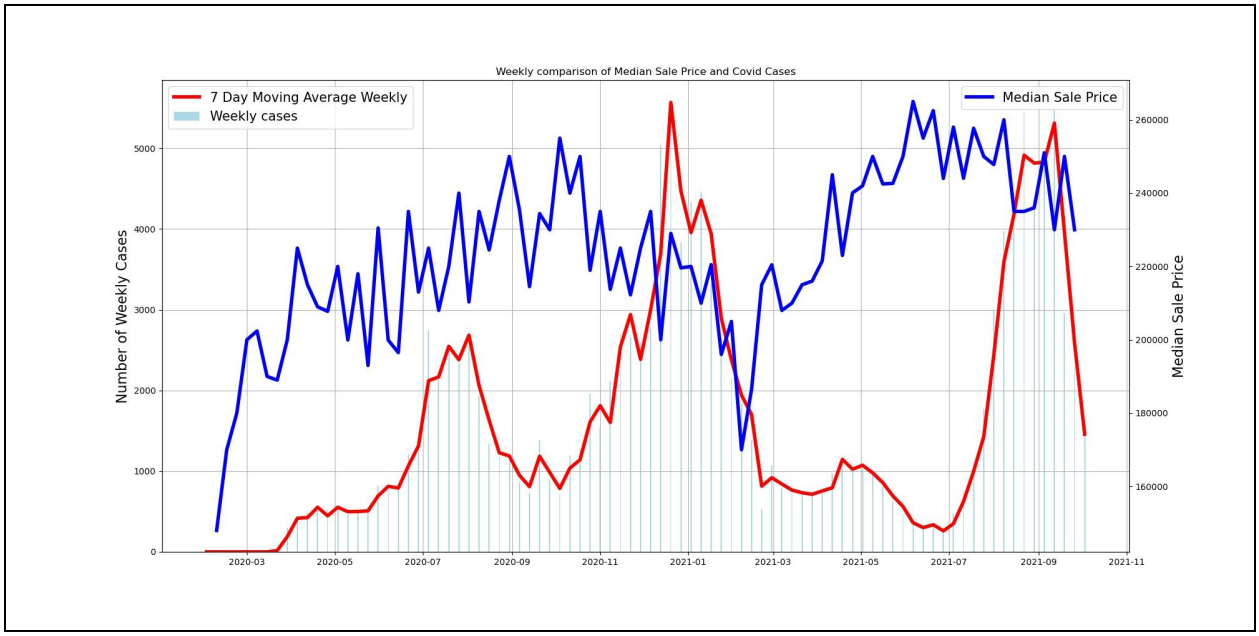
Findings

For our findings, we consider and compare our analysis based on the following KPIs:

1. **Housing Inventory:** During the pandemic in 2020 and 2021, we can see a sharp decline in housing inventory (blue line) in the first quarter of the year 2021 where we also observe a peak in covid cases as seen on the red line. A decline in inventory was caused by multitude of reasons such as decline in new constructions due to supply chain disruptions, people not wanting to sell homes due to economic uncertainty, people not wanting to sell homes to avoid home visits by strangers to avoid contracting the virus, and also because of strict stay at home orders and lockdowns in place by the government.



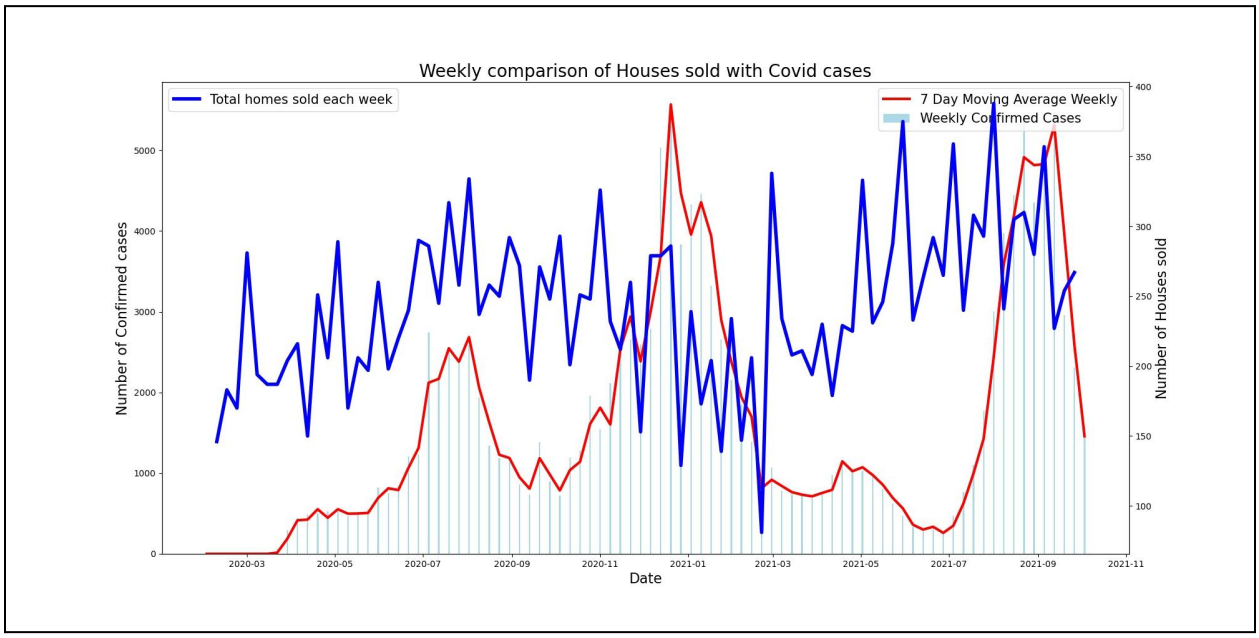
2. **Median Sale Price:** We can see an overall upward trend in median housing prices of Shelby county, TN. However, this is still followed by a lot of quick falls and ups- this hints at the overall chaos in the housing market combined with other factors such as anxious buying among people, falling interest rates, limited inventory, etc. The housing price dipped briefly again in the first quarter of 2021 when the cases peaked, but it steadily increased after that.

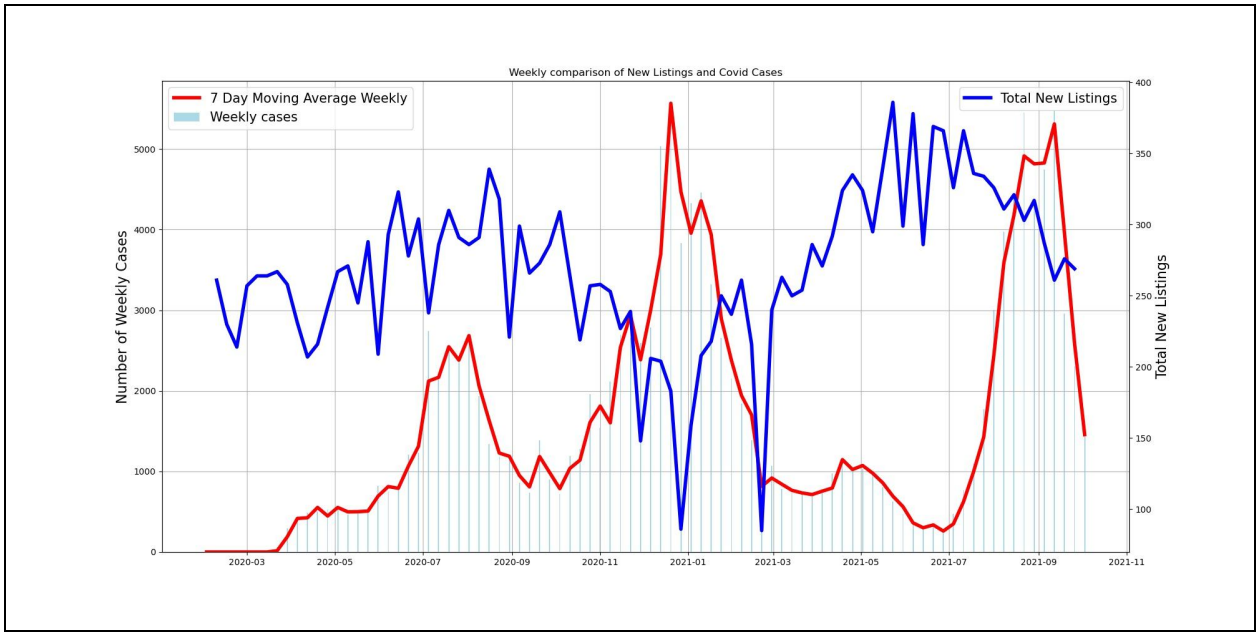


3. **Total Number of Homes Sold and Total Number of New Listings:**

The first graph shows the fluctuations in the total number of homes sold. Overall we see that the number of homes sold over the covid period fluctuates a lot. There are again short periods of increases and decreases in this number. With the covid cases peaking up in the first quarter of 2021, we observe that the number of homes sold also goes down but then continues to go up with the fall in the number of covid-19 cases.

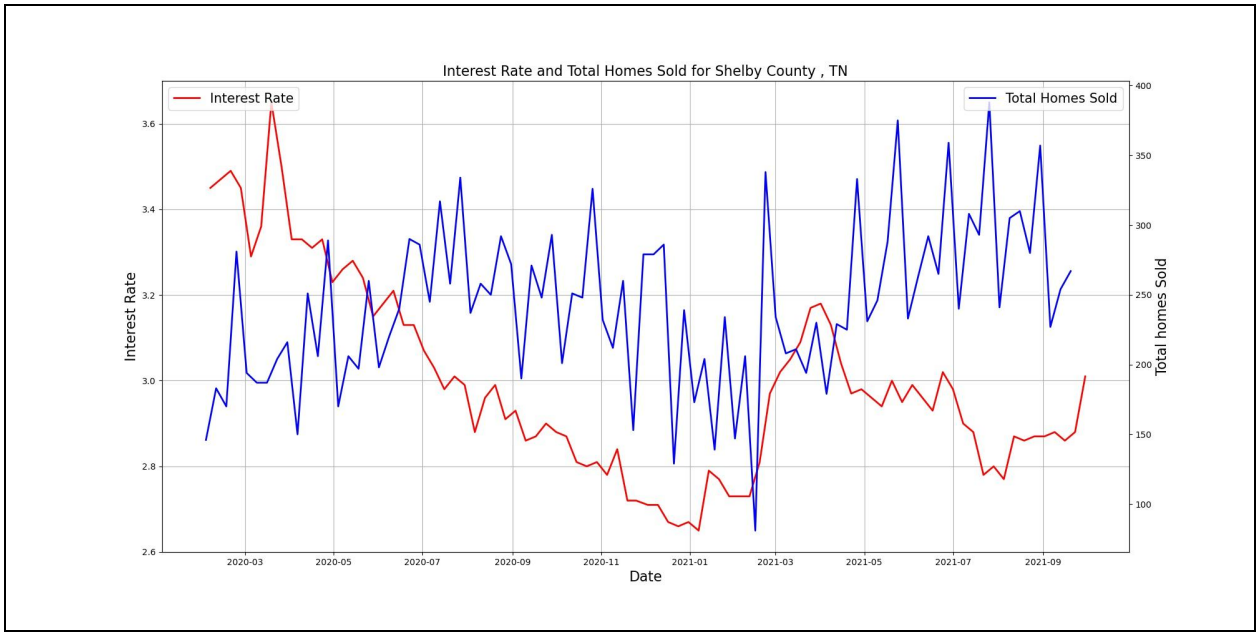
Similarly, we can see a sharp drop in the number of new listings around January 2021 when the cases peak. From these two graphs, we can conclude that everytime the covid-19 cases increase, the number of homes sold and number of new listings in the market decrease. This could be caused by a number of factors like lockdown, fear among people of getting covid by doing home tours, sellers thinking they might not be able to sell at a high price if there are not enough buyers, real-estate agents not wanting to do in-person home tours etc.

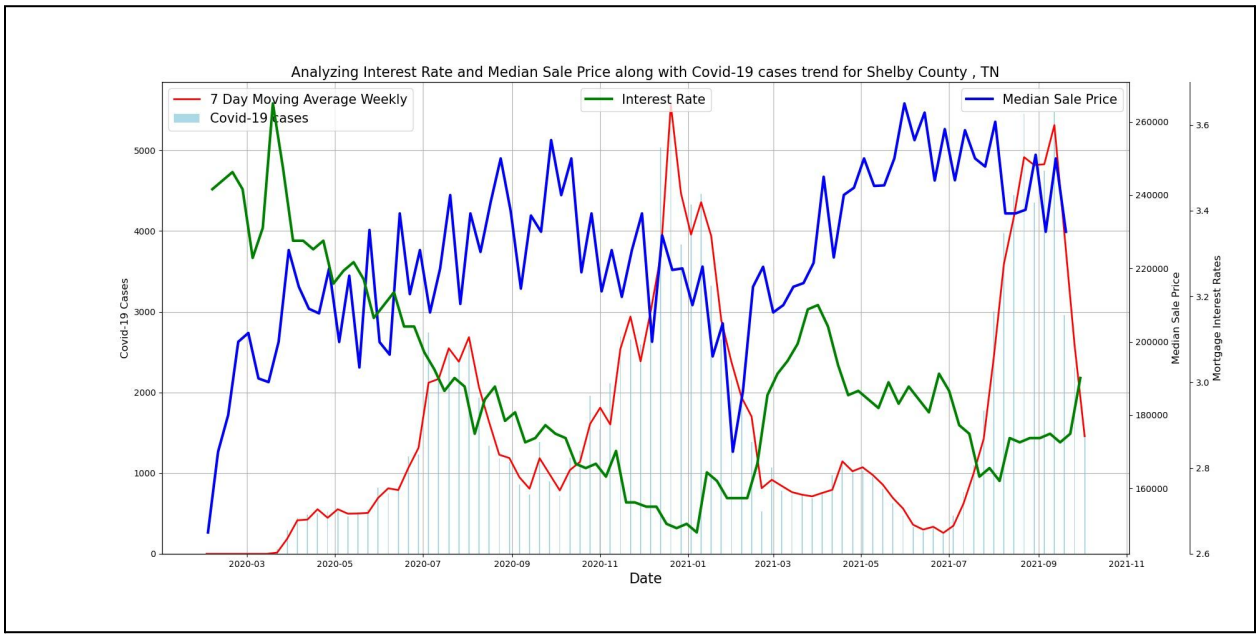




4. Interest Rate and Housing Trends:

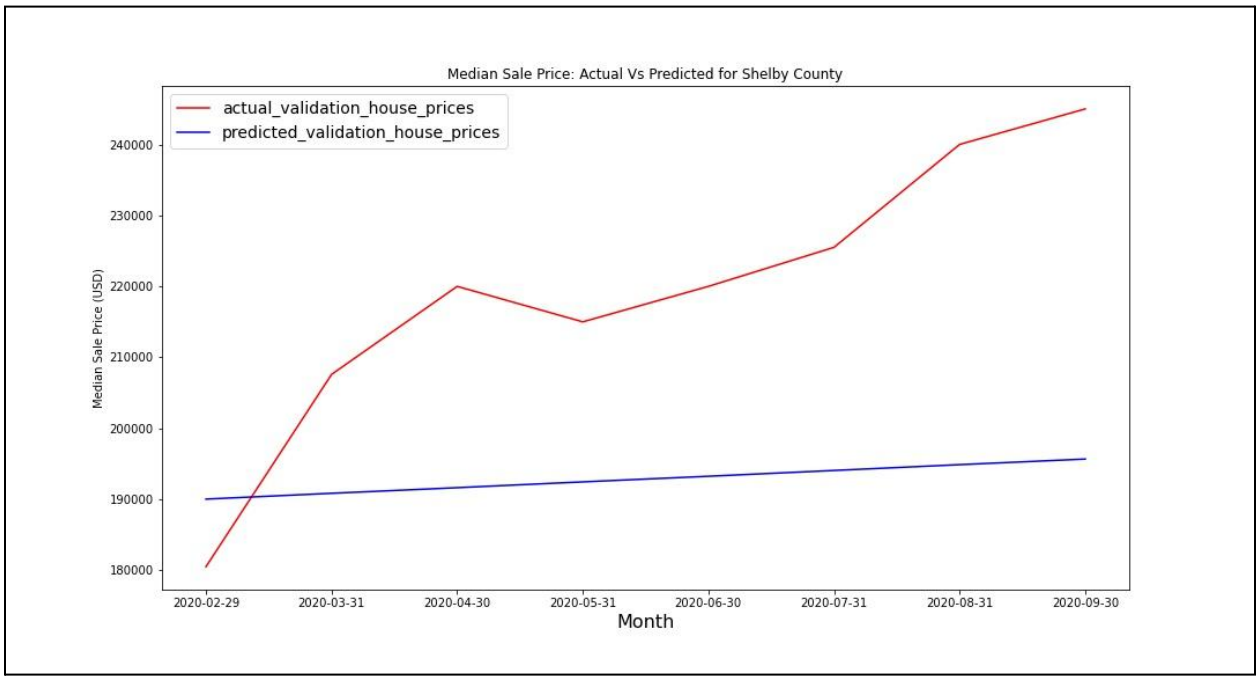
Another interesting aspect to look into was the 30-year mortgage interest rate. Overall we can see a declining interest rate trend in 2020 followed by interest rates dropping significantly to a record of less than 2.7% in January 2021. Although we cannot directly correlate interest rate to total homes sold, we can see that lower interest rates have fueled a higher demand for homes as people want to take advantage of lower mortgage rates. However due to a shortage in inventory, the total number of homes sold remained lower which caused the median home price to increase.





5. Linear Regression Output:

From the output of the linear regression model, we can see that there is a big difference between the actual housing and predicted housing price trends for 2020 and 2021. The model was trained with historic housing price data from (2012 to 2019) and based on that was asked to predict the housing price for 2020 and 2021. The model output shows that the housing price during the pandemic increased significantly more compared to previous years and that houses sold for a higher price than they should have been.



Discussion/Implications

It is eye opening to see the trends in covid-19 cases and the housing market. The most interesting observation is to see how the median sale price of homes is increasing sharply due to a surge in demand for homes since the start of the pandemic. The lower interest rates have also fueled the increase in demand for homes. However, to match the demand, there isn't enough housing on the market as can be seen from the decline in housing inventory.

Future Work

1. It will be interesting to build on top of my analysis to see patterns in the housing market related to other factors such as work from home, commuting patterns, supply chain disruptions, wood prices, lockdown restrictions that have also had some impact on the housing market indirectly.

2. Another interesting future work is to see how the housing market performed in different states and counties and if there was any commonality that can be extended from such an analysis to generalize the results with similar size states and counties.

3. Another potential analysis involves how rental prices fluctuated throughout the pandemic and if there was any correlation between rental price and housing prices.

4. Another analysis that can be built on top of this includes looking at the median household income of the people purchasing the homes to further analyze how the government can create schemes to help racial and generational inequality.

5. Going forward it would also be useful to put this visualization in a dashboard that gets refreshed with data on a day-to-day basis so we can see how the housing market evolved as COVID and its variants and cases change on a timely basis.

Limitations

There were certain limitations to my analysis which are listed as follows:

1. I limited my analysis to a fixed period of time which is from February 2020 to October 2021. If looked at longer periods of time, this analysis could infer different conclusions.

2. I limited my analysis to Single Family Homes (SFH), that is, I did not include other home types like townhomes, condos, etc. This could mean that there could be other patterns and trends for different property types and my results do not extend to those other residential properties.

3. In my interest rate analysis, I focus on the 30 year fixed interest rate. This could mean that other interest rates could have influenced the housing market differently which is not explored in my analysis.

4. The Redfin weekly and monthly data is not seasonally adjusted. Doing so could yield slightly different results.

Conclusion

Through my analysis, I can conclude that the COVID-19 pandemic had an impact on the housing market in Shelby County. We can see that COVID-19 pandemic has impacted both supply and demand in the housing market. People wanted to take advantage of lower mortgage rates which in turn caused the demand for housing to go up- we see a lot of high and low bumps in the median sale price that indicates the anxiety and uncertainty in the housing market.

However the pandemic also fueled a shortage in supply of homes - both newly built and those sold by existing owners. We can see a decline in housing inventory, number of new listings on the market as covid cases increased during the first quarter (Q1) of 2021. We see that with a sharp increase in the number of covid cases, our KPIs except median sale price fall sharply. Thus we can conclude that an increased demand and a shortage in supply fueled the median sale price of the homes to go up throughout the pandemic.

In terms of mortgage interest rate, we see that they very closely follow the covid-19 trend, that is, the mortgage interest rates keep falling down and reach their lowest point during the Q1 of 2021. With the number of covid-19 cases going down, the mortgage interest rate starts coming up again.

To conclude, this study informs the reader of their understanding of human centered data science as it is important to pay attention to this trend and for the government to take action. It is important to fix this gap between supply and demand by building more homes where people need it, otherwise this inequality will continue to skyrocket and a growing number of Americans will be shut out of the housing market altogether.

References

1. The impact of coronavirus on the U.S Housing Market - Redfin. (n.d.). Retrieved December 11, 2021, from https://www.redfin.com/guides/coronavirus-housing-market-impact.

2. Florida, R. (2022, September 8). *Why did housing costs explode during the pandemic?* Bloomberg.com. Retrieved December 11, 2022, from https://www.bloomberg.com/news/features/2022-09-08/why-did-housing-costs-explode-during-the-pandemic

3. Federal Reserve Bank of St. Louis. (2021, December 9). The impact of covid-19 on the residential real estate market: St. Louis Fed. Saint Louis Fed Eagle. Retrieved December 10, 2022, from https://www.stlouisfed.org/publications/regional-economist/fourth-quarter-2020/impact-covid-residential-real-estate-market.

4. Demsas, J. (2021, February 5). Covid-19 caused a recession. So why did the housing market boom? Vox. Retrieved December 11, 2022, from

https://www.vox.com/22264268/covid-19-housing-insecurity-housing-prices-mortgage-rates-pandemic-zoning-supply-demand.

5. Sai Balasubramanian, M. D. (2021, December 10). The COVID-19 pandemic has fueled a crisis in the housing market. Forbes. Retrieved December 11, 2022, from https://www.forbes.com/sites/saibala/2021/04/27/the-covid-19-pandemic-has-fueled-a-crisis-in-the-housing-market/?sh=49f3de175928.

6. Valkov, V. (2019, July 5). Predicting house prices with linear regression: Machine learning from scratch (part II). Medium. Retrieved December 11, 2022, from https://towardsdatascience.com/predicting-house-prices-with-linear-regression-machine-learning-from-scratch-part-ii-47a0238aeac1.

Data Sources

To perform the analysis I will use the following different datasets.

1. 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

2. 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

3. The monthly housing market data from Redfin - https://redfin-public-data.s3.us-west-2.amazonaws.com/redfin_market_tracker/county_market_tracker.tsv000.gz

4. 30 year fixed Interest Rate - https://fred.stlouisfed.org/series/MORTGAGE30US