1. Introduction

In this project, I analyze the impact of covid on the housing market. One specific area that specifically experienced the brunt of this unprecedented crisis is the housing market. As a home buyer during the pandemic, my husband and I were going through the process of searching for a home. But the housing market at the start of 2021 gave us immense anxiety and worry as the demand for the houses was very high, prices of the houses kept escalating at a very fast rate, the inventory of houses kept declining and the bidding war to top it off made things worse. We spent countless hours every day trying to read the news, watching youtube videos, and talking to other buyers to understand the housing market to help with the confusion.

A lot of the home-buyers across the country are frustrated with this increasing trend in housing prices. This topic is particularly human-centered as it affects a vast majority of Americans who are looking to purchase homes. If the prices keep increasing at this rate, a lot of the people from lower-income brackets will suffer as they will not be able to afford to house, or will be pushed to outskirts or suburbs thus causing a widening of the wealth gap between the rich and the poor. Because 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 and hope to see how the market fluctuated during the pandemic. I also want to see if the housing prices have gone down recently or if they are still on the increasing trend. This analysis is of particular scientific interest as it will help others see trends between housing prices, inventory, number of listings and a fluctuating economy during a pandemic and help them in making a decision if they are looking to buy or sell a home in Mecklenburg County, North Carolina.

2. 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 widespread fear of another recession. Tracking changes in the housing market will provide us insights into the economy of North Carolina state, help us contextualize growth and decline in this county, and give us 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.

1. During the COVID-19 pandemic did the housing prices go up or down from January 2020 through October 2021?
2. Did the number of COVID-19 cases have an impact on the housing prices, inventory, median sale price, number of new listings in the market from January 2020 through October 2021?

3. Did changes in 30-year fixed Interest Rates cause any fluctuations in Median Sale price and number of homes sold in the market?
4. Were there any other trends in the data related to covid cases and the housing market?

There are several interesting articles published about the impact of COVID-19 on the housing market. This [Forbes article](#) helped me understand how there are several factors that play a role when it comes to housing prices. Trends such as WFH have fueled an increase in lumber prices due to increased interest in home renovation and remodeling in turn affecting new constructions and home prices. Also other factors such as corporate relocations are creating a surge in demand for housing prices. Apple announced its plan to invest 1B in North Carolina to build a new campus and engineering hub in the research triangle. This news is also fueling the housing prices in NC as demand will soar in the future.

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.

3. Methodology

In my methodology I have made an effort to address a variety of human-centered data science considerations including

1. **Licenses**: I ensured that the research is licensed under a MIT license so others can use it.
2. **Copyright**: Appropriately citing sources of data, tools, and inspiration
3. **Interpretability**: Following a literate programming style to explain decisions made along the way that impact the final results of my analysis
4. **Repeatability and Reproducibility:** Making my research reproducible by providing information about my environment, how I used the data, and where I received my data in addition to 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. Also I have used visualisation to convey the information in a simple manner to promote this goal.

To further help in easy interpretation, I broke down the steps into the below 4 steps -

1. **Data Acquisition -** The data is downloaded from the below data sources and saved to the data-raw folder for easier access in future for anyone reproducing the analysis. The data includes the covid data from Kaggle and the Weekly and Monthly housing market data from Redfin.
2. **Data Processing and Cleaning -** Here I performed Data processing to filter data for Mecklenburg county and North Carolina State and also filter data for data range (January 2020 to October 2021). This is also where I compute the 7-day moving average 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 perform 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.
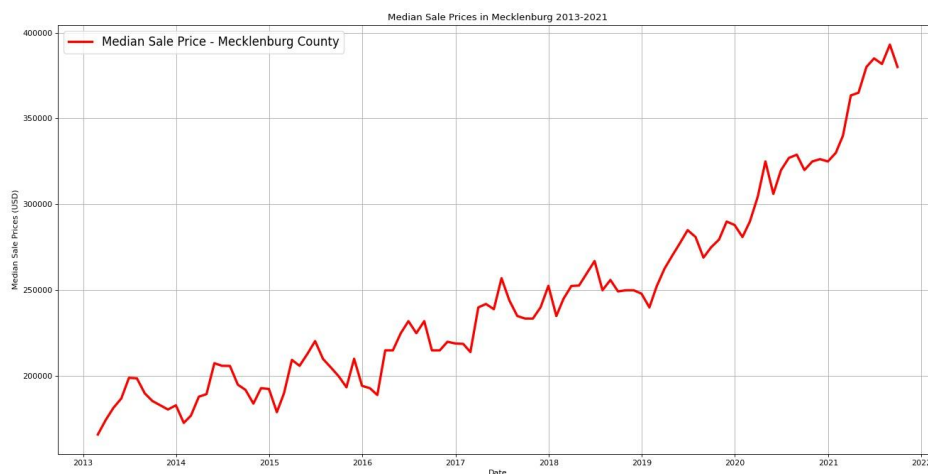
I fit a univariate linear regression model using historical data from (2013- 2019) where the feature is the weekly dates and the target is the median housing price. We are all aware of the housing market crash in 2007-2008 and it took the market some time to stabilize after the crash, hence I have decided to train the model with data post-2013. I will split the data into (80-20 train test split) and fit a linear regression model using python scikit learn. I will also compute the RMSE (root mean square error) as a measure of model performance and use the model to predict housing prices for 2020 and 2021 and compare if the prediction is higher or similar to actual prices.


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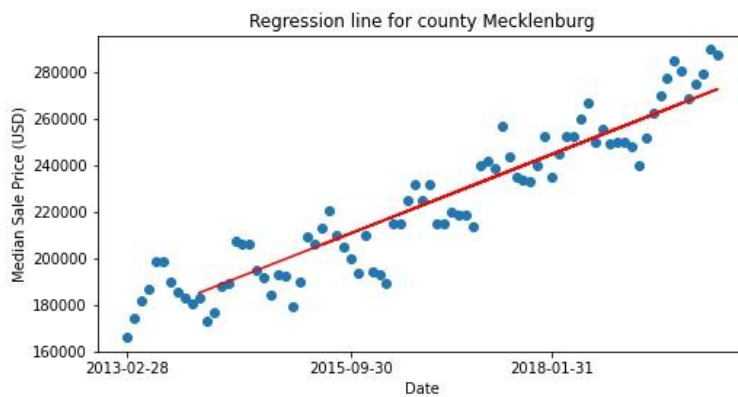
The model's performance on the test set is evaluated using the common evaluation metrics for regression problems

MAE, MSE and RMSE on Test Data for Mecklenburg

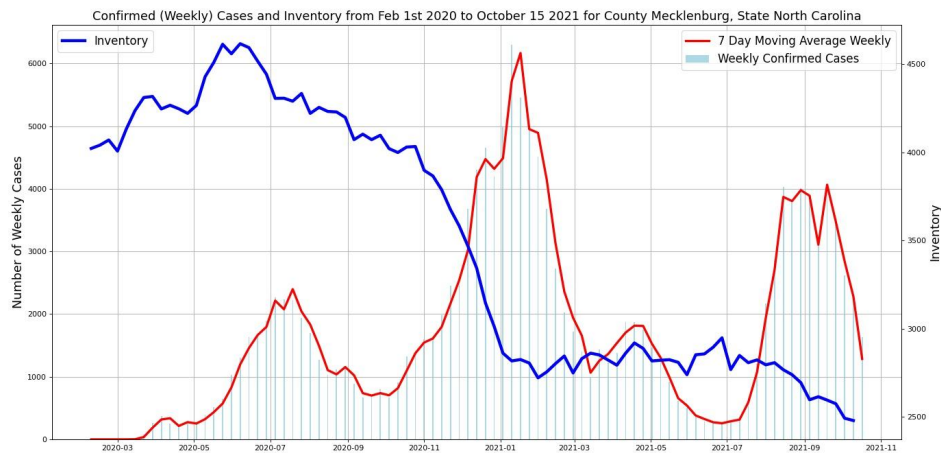**Mean Absolute Error:** 9957.248703801895

Mean Squared Error: 163812031.95049974

Root Mean Squared Error: 12798.90745143896

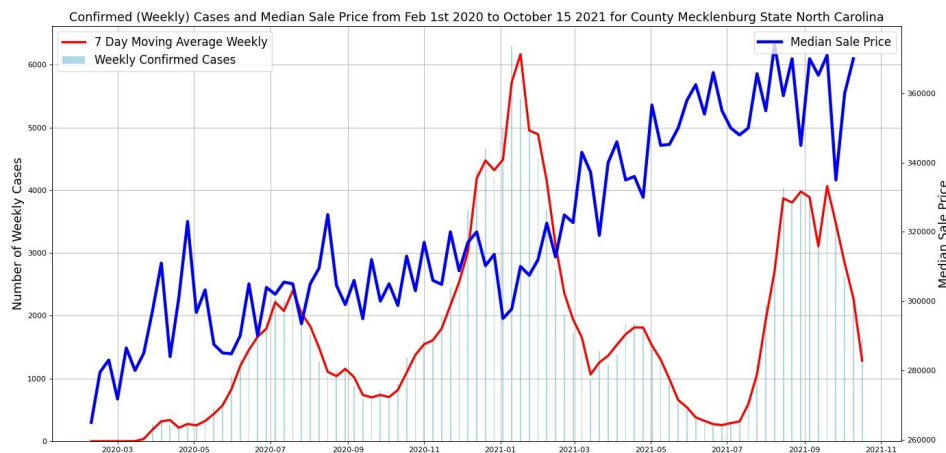Median Sale Price on Test Data: Actual Vs Predicted for city '{city}'
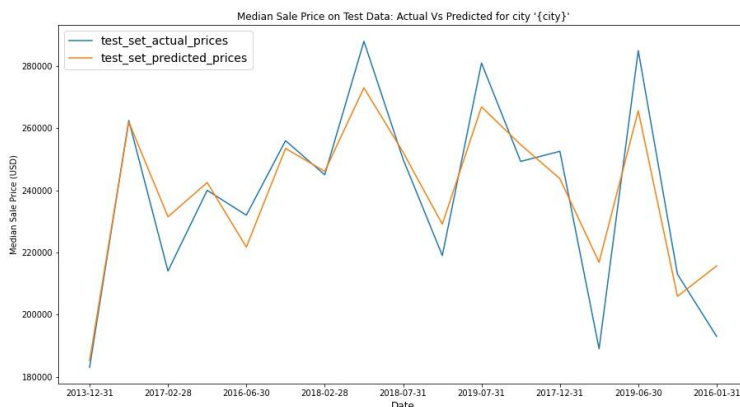
4. Findings

   Housing Inventory : During the pandemic in 2020 and 2021, we can see a sharp decline in housing inventory (blue line) since January 2021(peak of covid) 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.


Confirmed (Weekly) Cases and Inventory from Feb 1st 2020 to October 15 2021 for County Mecklenburg, State North Carolina

Median Sale Price: We can see an upward trend in median housing prices overall in NC. The housing price dipped briefly around January 2021 when the cases peaked, but it steadily increased after that. We can see a 100K increase in median sale price in a span of roughly two years from 260k to 360k. The increase in Median Sale Price can be correlated to the decline in inventory as can be seen from the above graph.


Confirmed (Weekly) Cases and Median Sale Price from Feb 1st 2020 to October 15 2021 for County Mecklenburg State North Carolina

Total Number of Homes Sold and Total Number of New Listings: The graph on the left shows the fluctuations in the total number of homes sold. We can see at the start of the pandemic the number of homes sold were lower as there was an overall uncertainty due to the pandemic. Around June 2020 we can see the number of homes sold start to increase and then it dips when the number of covid cases reaches its peak around January 2021. 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 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.
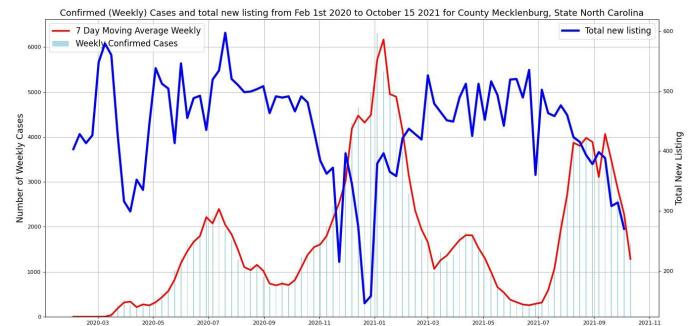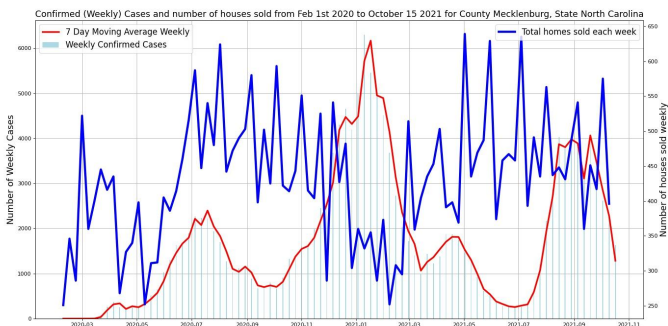
Interest Rate and Housing Trends: Another interesting aspect to look into was the interest rate. For my analysis I restricted to the 30 year fixed interest rate and how Single Family Homes fluctuated with regards to the 30 year fixed interest rate. Overall we can see that in 2020 the interest rate dropped significantly to a record low of 2.65% 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 ramain lower but causing the median home price to increase.
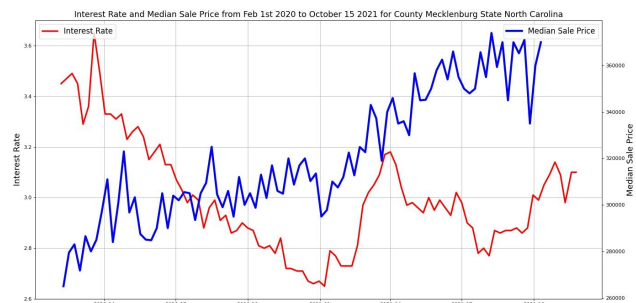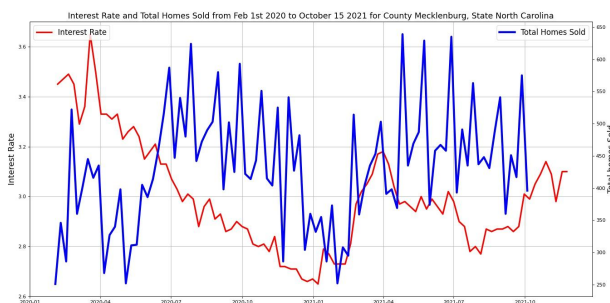
Linear Regression Output: From the output of the linear regression model, we can see that there is roughly 100K difference between the actual housing and predicted housing price for 2020 and 2021. The model was trained with historic housing price data from (2013 to 2019) and based on that was asked to 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.

Median Sale Price: Actual Vs Predicted for County Mecklenburg
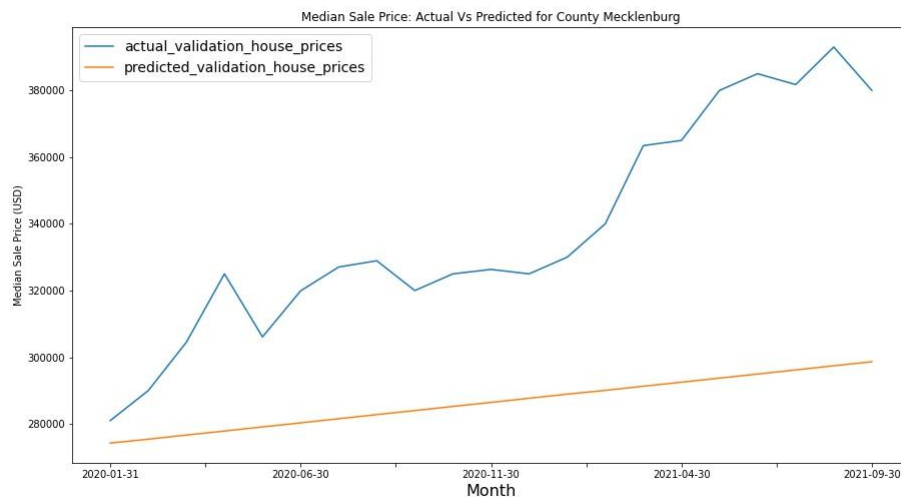
We can also confirm this by looking at the percentage change in median sale price from year to year to see that the increase in sale price is indeed higher in 2020 and 2021.

YEAR	% CHANGE
2014	0.026586
2015	0.051282
2016	0.131707
2017	0.012931
2018	0.089362
2019	0.050781
2020	**0.222770**
2021	0.194801

5. Discussion/Implications

It is eye opening to see the trends in Covid 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 analyse 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.

6. Limitations
1. All my analysis and conclusions are based on Single Family Homes (SFH). This could mean that there could be other patterns and trends for different property types and my results do not extend to other residential properties such as condos, townhomes, etc
2. 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.
3. The analysis uses a 7 day rolling average number of covid cases to see correlation between the number of covid cases and housing market and does not include deaths.
4. The Redfin weekly and monthly data is not seasonally adjusted. Doing so could yield slightly different results.

7. Conclusion

Through my analysis, I can conclude that the COVID-19 pandemic had an impact on the housing market in Mecklenburg 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 caused the demand for housing to go up.

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. 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.

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.

8. References

1. 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 11, 2021, from https://www.stlouisfed.org/publications/regional-economist/fourth-quarter-2020/impact-covid-residential-real-estate-market.

2. Demsas, J. (2021, February 5). *Covid-19 caused a recession. so why did the housing market boom?* Vox. Retrieved December 11, 2021, from https://www.vox.com/22264268/covid-19-housing-insecurity-housing-prices-mortgage-rates-pandemic-zoning-supply-demand.

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

4. Valkov, V. (2019, July 5). *Predicting house prices with linear regression: Machine learning from scratch (part II)*. Medium. Retrieved December 11, 2021, from https://towardsdatascience.com/predicting-house-prices-with-linear-regression-machine-learning-from-scratch-part-ii-47a0238aeac1.

5. Faressayah. (2021, August 5). *Linear regression 📈 house 🏡 price 💵 prediction*. Kaggle. Retrieved December 11, 2021, from https://www.kaggle.com/faressayah/linear-regression-house-price-prediction.

6. *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.

9. 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