# Introduction

The goal of this project is to analyze how total housing listing count was affected by the COVID-19 pandemic with a scope limited to Fort Bend County, Texas. While the project does not solve a problem it does seek to incite discourse regarding the nature of housing proprietors, their motivations, and their concern for the greater population health. Ideally, this project will allow reviewers to draw inferences from the results such to formulate conclusions to be tested against others in a pragmatic ethical fashion. While seeking to understand the housing market and its influences is not a novel concept, it is one that dynamically changes with the environment around it; therefore, it should be continuously under analysis.

# Background/Related Work

## Related Research

According to the [US Census Bureau](https://www.census.gov/library/stories/2021/10/zillow-and-census-bureau-data-show-pandemics-impact-on-housing-market.html), the housing market came to an abrupt halt followed shortly after by a resounding rebound with “hyper competitive buying conditions” [1]. However, it is worth noting that this conclusion comes from data not only on the national level but also from a third-party source – Zillow. As a result, it would not be surprising if Fort Bend County, Texas diverges from the national average.

Bankrate.com’s [article](https://www.bankrate.com/real-estate/covid-impact-on-the-housing-market/#process) also supports the claims made by the Census Bureau through stating that housing market sales stalled when the pandemic first hit followed by them picking back up along with the stock market [2]. In reference to the housing boom, this article cites home sellers “acerating their plans to sell at premium prices.” [2]

As a result of this research, it is likely to see a sharp decline in the housing market early in the pandemic with a rise in total listings shortly afterwards; however, it may not be the case that Fort Bend County, Texas aligns with the housing market shifts for the rest of the county.

The low presence of research finding correlation between covid-19 cases and total housing market listings provides an incentive to look for Pearson and Spearman correlation between the two with both a static and time-lagged approach.

## Research Question

How were total housing market listings influenced by the COVID-19 pandemic?

## Hypotheses

**Hypothesis 1:** I predict that a time lag of two months will occur between the slope of confirmed COVID-19 cases and total market listings present in Fort Bend County, Texas

**Hypothesis 2:** I predict every 25% change in the slope of confirmed COVID-19 cases will result in a 5% decrease in total market listings in Fort Bend County, Texas.

# Methodology

## Data Transformation

**RAW\_us\_confirmed\_cases.csv Data Cleaning**

In order to filter results to those including Fort Bend County, I indexed the original dataframe based on Federal Information Processing System codes that matched Fort Bend’s. From there, I dropped all columns not relevant to date or confirmed Covid-19 cases such to produce time series data.

**COVID-19 Confirmed Cases Derivation**While total confirmed cases are a good gauge as to the severity of the epidemic limited to the scope of an entire timeframe, the slope of confirmed cases would yield a better gauge of the pandemic severity limited to the scope of each individual day. In order to produce this better gauge of pandemic severity, I employed numpy’s gradient function on the total confirmed cases dataframe.

**Housing Inventory: Total Listing Count in Fort Bend County, TX Discretization**

In order to produce market listings data that is discretized by day instead of month, I duplicated market listing count for each months corresponding day. As I result, I was able to plot and compare the two dataframes directly.

**Curve Smoothing**

In an attempt to reduce the noise of each plotted dataframe and better understand their tendencies, I decided to implement curve smoothing to allow for better visualization inference. The three methodologies being used for curve smoothing are as follows: piecewise polynomials, natural cubic splines, and moving average.

## Correlation

**Pearson Correlation**

To understand the *overall* correlation between Covid-19 slope cases and total market listings, I calculated Pearson Correlation to give an understanding of the overall linear relationship

**Spearman’s Correlation**

To supplement Pearson Correlation, I decided to also calculate Spearman’s Correlation to gauge the validity of a monotonic relationship between market listings and the pandemic severity.

**Time Lagged Cross Correlation (TLCC)**

Implementing time lagged cross correlation aids in discovering a leader-follower relationship and therefore is vital in understanding response time between two time-series dataframes as well as their peak synchrony.

**Windowed Time Lagged Cross Correlation (WTLCC)**

To understand how correlation changes over time, I implemented windowed time lagged cross correlation such to gain a more fine-tuned assessment of any leader-follower relationships over a given timeframe.

## Visualizations

**Dual-Axis Line Graph**

I will be producing dual-axis line graphs to visually understand correlation between two dataframes with corresponding timeframes. To assess the difference between confirmed Covid-19 cases and their derivation, I plotted them together along with vertical lines indicating the release of the vaccine as well as breakpoints calculated through the pelt search algorithm. To view the success of calculated smoothed curves, I plotted them against their original “noisy” counterparts such to see how much noise was captured or removed. Finally, the more vital deliverable of this project is a line graph with both the derivation of confirmed Covid-19 cases and total market listings plotted against time.

**Single-Axis Line Graph**

A single-axis line graph is produced when calculating time lagged cross correlation with the x-axis encoded to number of days shifted and the y-axis encoded to the Pearson correlation coefficient. Within this line graph, a vertical line is added to point to the time-lag that corresponds to the highest correlation within the time frame.

**Heatmap**

A heatmap with bucketized time frames encoded to the y-axis, bucketized offset encoded to the x-axis, and color encoded to Pearson correlation is produced to understand windowed time-lagged cross correlation.

# Findings

## Analysis of Confirmed Cases in Fort Bend County, TX

Chart, histogram

Description automatically generated

Figure 1: Confirmed Cases of COVID-19 and Derivation in Fort Bend County, Texas

This is the deliverable from part one of this research project. This visualization allows the viewer to not only gain an understanding of how COVID-19 cases changed over time, but also how their slope or instantaneous severity changed over time as well. Furthermore, the viewer can see the effects of the vaccine rollout with the red vertical line and the outlier breakpoints with the yellow vertical lines.

## Discretized Total Listing in Fort Bend, County, TX

Chart, histogram

Description automatically generated

Figure 2: Discretized Total Housing Market Listings by Date in Fort Bend County, Texas

## Curve Smoothing

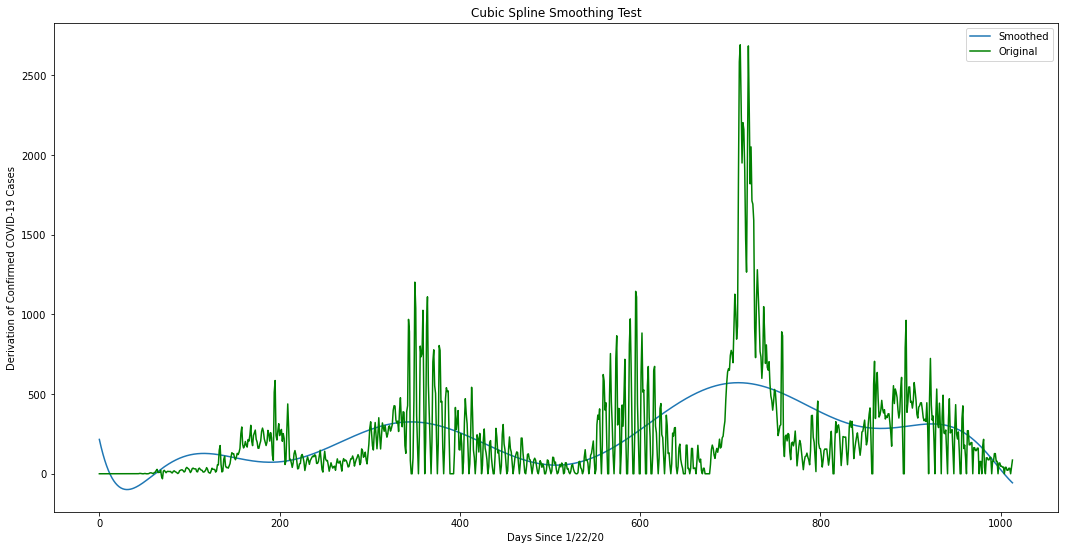


Figure 3: Cubic Spline Smoothing Test

Chart, histogram

Description automatically generated

Figure 4: Piecewise Polynomial Smoothing Test

Chart, histogram

Description automatically generated

Figure 5: Rolling Average of Derivation of Confirmed COVID-19 Cases in Fort Bend County, Texas

Chart, histogram

Description automatically generated

Figure 6: Rolling Average of Total Housing Market Listings in Fort Bend County, Texas

As can be seen, moving average with a window size of 50 yielded far better results than the products of both natural cubic splining and piecewise polynomial splining. Due to these findings, moving average smoothed curves will be utilized in the following visualizations.

## Confirmed Cases vs. Total Market Listings in Fort Bend County

Chart, line chart

Description automatically generated

Figure 7: Smoothed Derivation of COVID-19 Cases vs. Smoothed Total Market Listings in Fort Bend County, Texas

This visualization allows the viewer to directly relate COVID-19 case derivation against total housing market listings for Fort Bend County, Texas. Furthermore, the viewer can view the self-reported masking survey results as well as the date of the vaccination release.

## Correlations

Table 1: Pearson and Spearman Correlation Coefficient between Derivation of COVID-19 Cases and Total Market Listings in Fort Bend County, Texas

|  |  |
| --- | --- |
| **Pearson** | **Spearman** |
| -0.409 | -0.387 |

These results are indicative of there being a significant relationship between COVID-19 cases and total market listings. Furthermore, they indicate that there is a somewhat linear relationship between the two and that for every 25% change in COVID-19 slope there is a 10% change in total housing market listings.

Chart, line chart

Description automatically generated

Figure 8: Time Lagged Cross Correlation between Derivation of COVID-19 Cases and Total Market Listings in Fort Bend County, Texas

This visualization indicates that the housing market listings lead confirmed COVID-19 case derivation by 259 days. Furthermore, the viewer can see the change in Pearson correlation over time to understand the synchrony of varying offset.

Background pattern

Description automatically generated

Figure 9: Windowed Time Lagged Cross Correlation between Derivation of COVID-19 Cases and Total Market Listings in Fort Bend County, Texas

This visualization supports the results from the previous figure with the overwhelming presence of red indicating housing market listings lead confirmed case derivation. However, it is interesting to note that the presence of blue in certain rows reveals that the established leader-follower relationship does in fact change between the two curves.

# Discussion/Implications

My findings concluded that, with a correlation coefficient of 0.4, for every 25% change in confirmed case derivation there is a 10% change in total market listings in Fort Bend County, Texas. Not only does this imply a stronger relationship than my hypothesis suggested, but it provides evidence for the connection between population health and the housing market. Furthermore, these results bolster claims from my initial research in a more fine-tuned fashion. Housing market listings do drop as the pandemic emerges and there is a resurgence shortly after; however, there is a drop following the resurgence as cases exponentially increase. Finally, while my results provide evidence for association and a leader-follower relationship, they are contrary to intuition. A time-lag of much less than 259 days can be viewed in figure 7 between 10/15/21 and 6/26/22. This is significant as it indicates a complexity when attempting to predict housing market response to similar events.

Future research could build on my confounding results by limiting the timeframe in question to filter out the initial and ending responses to the pandemic to understand the direct impact of public health rather than the anticipation of going into a crisis or leaving one.

# Limitations

The primary ethical consideration regarding the total housing market dataset is the level of completeness that realtor.com provides such that my analysis is a true comparison of total market listings and not just those mentioned within the scope of realtor.com. Another ethical consideration relates to the Belmont Report and whether this dataset violates the requirement of informed consent regarding the property proprietors whose estates comprise the bulk of this dataset.

As a part of the Realtor.com Real Estate Data Library, I will be providing the requested attribution for partial use of this data with the following [link.](https://www.realtor.com/research/data/)

This dataset does not seem to have an associated license nor terms of service aside from the previous attribution.

Furthermore, a limitation present in my research is the assumption of linearity when calculating the Pearson correlation. If a correlation technique that better accounts for non-linearity is implemented, then the correlation could potentially be stronger.

# Conclusion

In my analysis of how the COVID-19 pandemic affected total housing market listings in Fort Bend County, Texas, I found that there is a much stronger correlation than I initially thought. For every 25% change in cases there is a 10% change in total housing market listings. This doubles my initial estimate of a correlation coefficient of 0.2. Furthermore, my results yield a much longer response time-lag than two weeks. Instead, I calculated a time-lag of 259 days.

It is my hope that this research can allow its viewers to draw conclusions about the nature of housing market proprietors in Fort Bend County, Texas. There exists a fairly strong correlation between surrounding population health and total number of houses on the market. How do these results affect your predefined notions? What do the results imply in terms of housing proprietor concern for the local population? Is this the appropriate response? I urge you to use this research as a springboard for discussion regarding the nature of housing proprietors and test your conclusions against others in a pragmatic ethical fashion.

# References

* Zillow and Census Bureau Data Show Pandemics Impact on Housing Market [1]
  + https://www.census.gov/library/stories/2021/10/zillow-and-census-bureau-data-show-pandemics-impact-on-housing-market.html
* COVID’s impact on the housing market [2]
  + https://www.bankrate.com/real-estate/covid-impact-on-the-housing-market/

# Data Sources

* Housing Inventory: Total Listing Count in Fort Bend County, TX
  + Provided by [Realtor.com](https://fred.stlouisfed.org/series/TOTLISCOU48157)
* RAW\_us\_confirmed\_cases
  + Provided by the Kaggle repository of [John Hopkins University COVID-19 data](https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university)
* Mask Compliance Survey
  + Provided by [New York Times github](https://github.com/nytimes/covid-19-data/tree/master/mask-use) repository