Teju Kandula

Professor McDonald

DATA512 Human Centered Data Science

December 12, 2022

DATA 512 Project Part 4 - Final Report

Introduction

Covid-19 is something that has affected many different aspects of our lives over the past three years. Over the course of the pandemic there was a lot of datafication, with data being an integral tool in telling the story and predicting the course of the pandemic. There have been many integral decisions being made from these analyses and this has offered a different perspective in terms of pandemic control. When performing this analysis it is important to make sure we take a human-centered approach. The implications of these data driven projects have vast impacts that will affect humans both immediately and in the long term especially in terms of social health, trade, global relations and the economy. Over the course of this project we will continue to use data to gain insights into the pandemic, but will put an emphasis in making sure our exploration is human-centered. I will be looking at the pandemic through a human-centered, data driven lens for Philadelphia County, Pennsylvania.

My analysis for this project is two-fold. First I will be looking at how masking policies changed the progression of confirmed COVID-19 cases in Philadelphia County. Mask mandates were highly controversial in America and the efficacy of wearing a mask was debated. Showing how mask mandates affect the rate of covid over a period of time can be very beneficial in figuring out how to slow the spread of covid and may end up saving human lives. My second question draws upon the initial analysis by looking at how the rate of covid and masking policies

affected the housing market in Philadelphia. This analysis is important because housing is something that affects everyone and is crucial to human survival. There has been much speculation during the pandemic about housing. Personally, I was living in the Chicago suburbs during the pandemic. In the city many people received ‘covid deals’ with very low rent. In the suburbs the housing rates were soaring with houses being sold in days. This experiment will be interesting as we will be able to use actual data to see if this speculation is real as well as analyze the rental and real-estate market from 2018 to 2022 in Philadelphia, PA and give us an understanding how the real estate and rental market fluctuated during unprecedented times.

Background/Related Work

The impact of the pandemic on the housing market is something that has been discussed, speculated and written about multiple times. There have been many news articles written about the impact of Covid-19 on housing in Philadelphia with titles including “3 Ways the Coronavirus Crisis Is Expected to Shape Housing Prices in and Around Philly”, “As COVID Drives Home Prices Higher, Crunching the Numbers on Housing Affordability” and “WHAT HAPPENS NEXT? COVID-19’S CONTINUOUS IMPACT ON PHILADELPHIA’S REAL ESTATE MARKET”. Although these are all very informative articles about how COVID-19 has impacted the housing market, they are neither human-centered or data driven. Most of these articles do not mention the sources of their data, are not interdisciplinary and rely on a few testimonials. They also focus on a single aspect of the housing market. This research informs my hypothesis by making me believe that the pandemic has had an impact on the housing market, but I would like to expand on this in my project to see to what extent there was an impact as well as explore both the rental and sale of houses.

Over the course of my project I aim to create a data driven analysis that utilizes human-centered design techniques and is reproducible. This will allow peer-review to be simple and convenient. The research question I am aiming to answer is how has the pandemic affected the real estate and rental market in Philadelphia, PA? This is a very broad question so I will be dividing this question into multiple sub questions. I will be analyzing the effect of the pandemic on rent in the city of Philadelphia, list and sale prices of houses for sale in Philadelphia as well as the inventory of houses for sale in Philadelphia,PA.

As I have three sub questions listed, I also have three hypotheses that I will be listing in the same order as the sub questions above.

1) As the change in confirmed covid cases increases, the average rent decreases in the city of Philadelphia
2) As the change in confirmed covid cases increases, the average difference between list and sale price decreases in the city of Philadelphia
3) As the change in confirmed covid cases increases, the number of houses for sale decreases in Philadelphia, PA.

Along with the above hypothesis that I will be analyzing using data visualizations I also hope to test whether the mean number of listings and the mean average sale to list ratio changed before and after the pandemic (January 1, 2020). The hypothesis are listed below:

Test 1

$H_0: \mu_{num\ listings\ before\ pandemic} = \mu_{num\ listings\ during\ pandemic}$

$H_A: \mu_{num\ listings\ before\ pandemic} \neq \mu_{num\ listings\ during\ pandemic}$

Test 2

$H_0: \mu_{avg\ sale\ to\ list\ ratio\ before\ pandemic} = \mu_{avg\ sale\ to\ list\ ratio\ during\ pandemic}$

$H_A: \mu_{avg\ sale\ to\ list\ ratio\ before\ pandemic} \neq \mu_{avg\ sale\ to\ list\ ratio\ during\ pandemic}$

Methodology

For part one of the project I decided to create a data visualization showing how the course of the disease was changed by masking policies. I created a time series Figure showing the changes in the derivative function of the rate of infection (*Figure 1*) as well as the total number of cases over time (*Figure 2*), using color to indicate when the masking policies were in effect. Along with this I also added vertical lines indicating the change points showing when there was a significant change in the derivative function. I also created a pie chart of mask compliance in Philadelphia County (*Figure 3*) so we can further expand on the above graphs by seeing if the mask mandate resulted in compliance. This piece was important because if a mask mandate was implemented and the people were not complying graphs 1 and 2 would not be an accurate indication of whether masks were effective in changing the progression of the confirmed Covid-19 Cases.

For the extension part of the project I used two different analytical methods: data visualizations and t-tests. For the data visualizations we will create a time series visualization for all three data sets: rental prices, sale to list ratio and number of listings. Both the covid cases data and the Zillow data are time series so the visualization will show the change in rent (*Figure 4*), number of listings (*Figure 5*), sale to list ratio (*Figure 6*) over time with the number of covid cases overlaid. I also choose to include data for the housing market prior to the pandemic so we

can also visualize trends prior to the start of the pandemic and see how Covid-19 affected those trends. I chose this technique because I believe it is extremely human-centered. The access and interpretability of data analysis should not only be meant for people educated in statistical techniques. With these data visualizations the analysis of the pandemic and the housing market in Philadelphia can help a vast majority of people in the area understand how their housing works. I also believe data visualization is the best way to show all our data, without having to look at raw tables. Everything is upfront unlike creating models and statistical tests where a lot of the data is hidden in the background.

For the statistical analysis we will be using the two-sample t-test. The null hypothesis is that the means of each value before the start of the pandemic and after the start of the pandemic are equal. The alternative hypothesis is that the means are not equal. I chose this approach because we are investigating the difference of two means where the population standard deviation is unknown. It will be interesting to see whether the trends we see in the data visualizations are significant or not and not just viewing the trends in a data visualization. I chose not to include the ZORI in the statistical tests as I believe it is biased since there was an increasing trend prior to the pandemic.

Along with the methods mentioned above I would also like to look at news articles about housing and the pandemic. This will allow two more aspects of human-centered data science. The first being interdisciplinary research, allowing me to learn from experts in fields such as economics and social work. The news articles will also allow the use of ‘thick’ data, giving more context to the data being used in the above statistical techniques. I believe this final piece of the methodology will help round out my project and tie together data driven and ethnocentric research.

Findings

Over the course of this project we were able to find many insights into human actions during the pandemic. First we will discuss the findings of part 1 and explore how mask mandates affected the rate of Covid-19 over time.

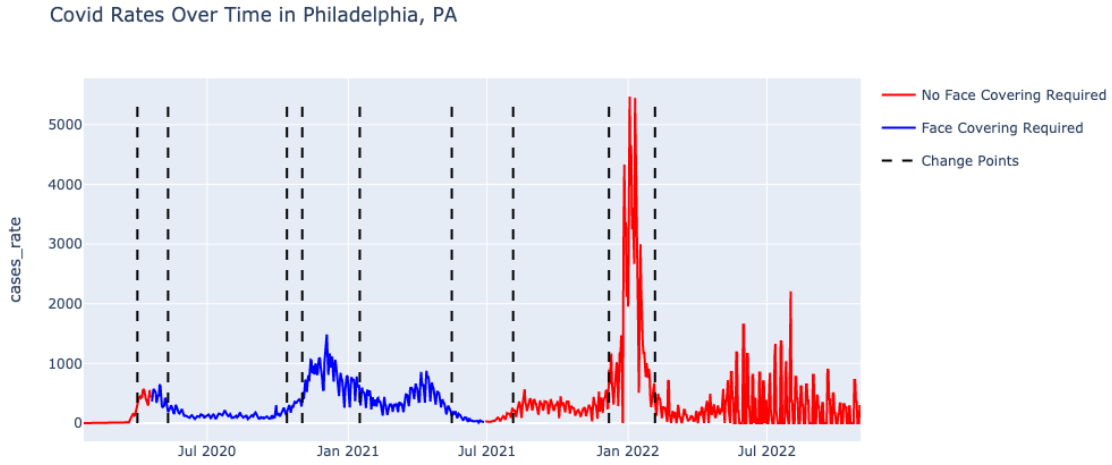

Figure 1

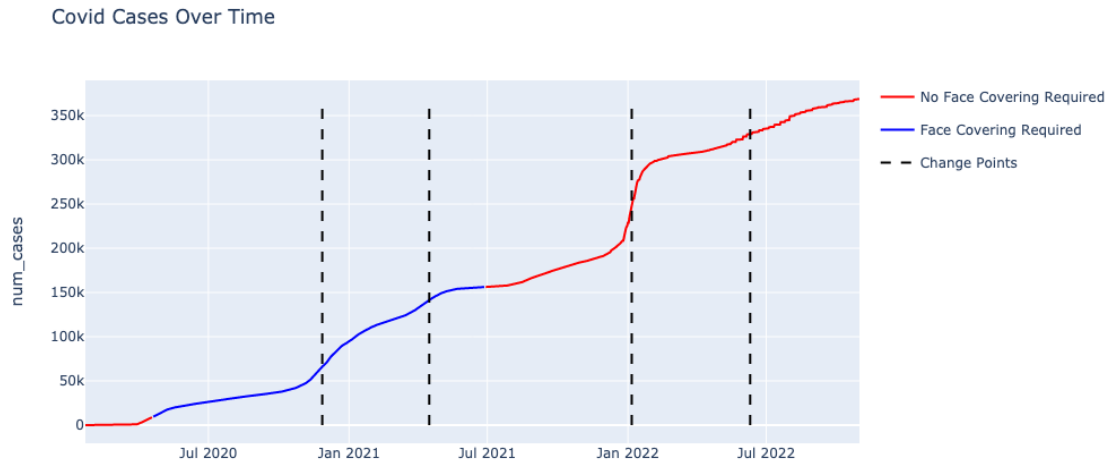

Figure 2

From the graph we can see that when the mask mandate was first introduced there was not much change in the rate of Covid-19. Once the mask mandate was removed it remained steady for a little while but then increased drastically. This may be due to the time it takes for

infection to occur. After the peak, there was still a decrease in the rate of Covid-19 even with no mask mandate. Overall I think we can get a little bit of information from these visualizations but nothing to make a solid scientific conclusion about, we will continue to discuss this in the *Discussion* and *Limitations* section of the paper. It is also important to see whether the Mask Mandates were actually followed in Philadelphia County.

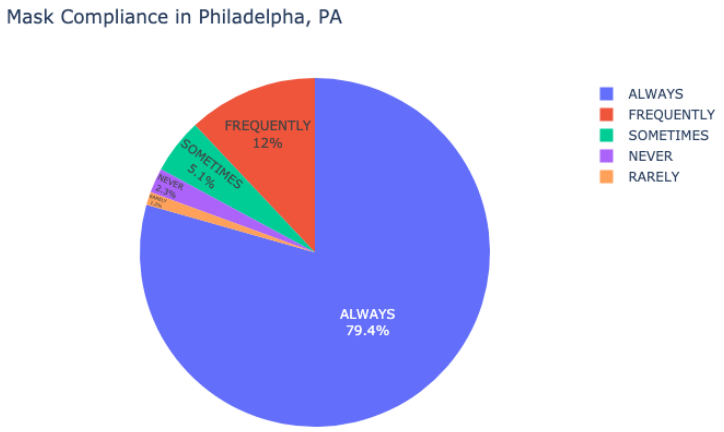

Figure 3

I simply modeled the masking compliance using a pie chart. I know pie charts are not usually favorable, but in this case we have very simple data which made a simple visualization the best choice. We are able to quickly see the overall mask compliance in Philadelphia. Here we can see that most people do comply with the local mask mandates which make the other visualizations more reliable. The first two graphs would make no sense looking at the difference between the rates of Covid-19 during mask mandates and no mask mandates if no one was following the mask mandates.

Next we will expand on the above visualizations and see how the rate of Covid-19 affected the housing market in Philadelphia County. We will begin by looking at the rental prices

in Philadelphia using the ZORI score which is a smoothed measure of the typical observed market rate rent across a given region.

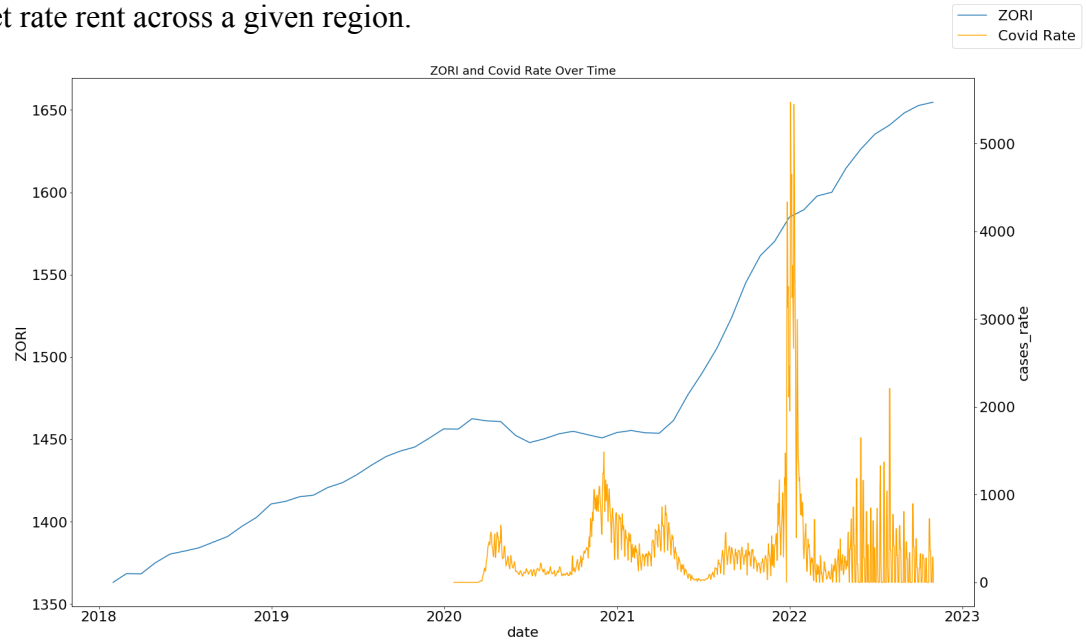

Figure 4

From the above graph we can see that the average rental prices in Philadelphia county have been increasing over time. During the first couple peaks of covid rates the ZORI seemed to steady and even slightly decrease at points. Later in the pandemic the ZORI continued to increase even with a big peak in covid rates in January 2022. Early in the pandemic caused a huge stall in a rental market that had been increasing for years before. Even with the stall in rental prices they continued to rise after the first significant dip in covid cases. One thing to note about this graph is that the data had been smoothes to get rid of any seasonality. Next we will look at the number of listings in Philadelphia County.

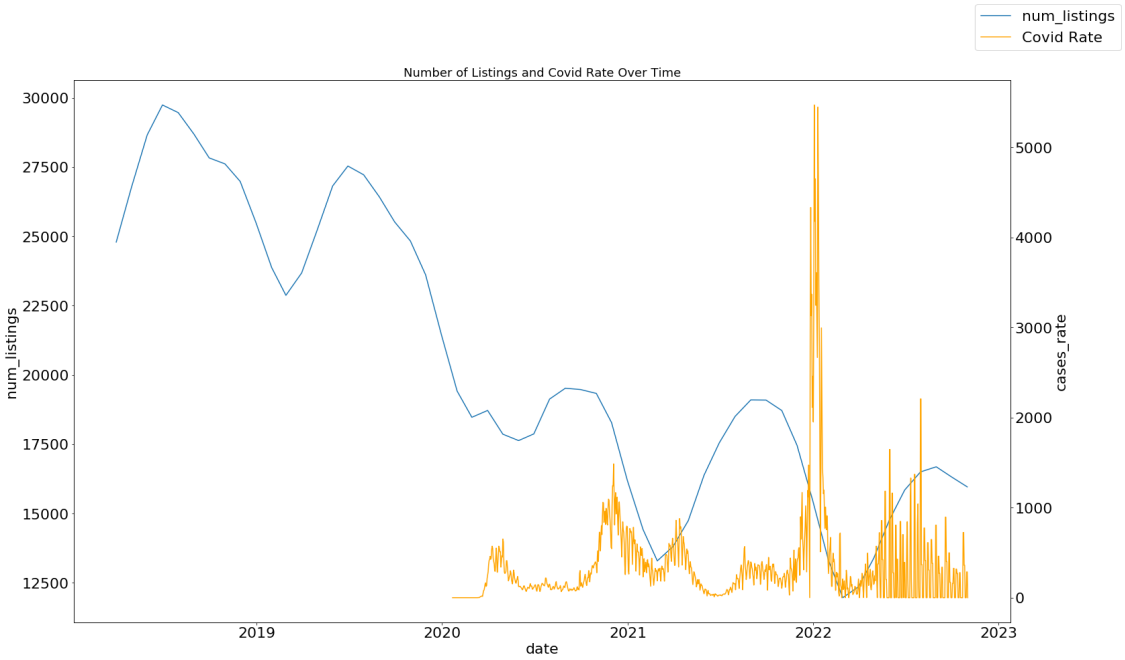

Figure 5

From the graph above we can see that prior to the pandemic there was already a seasonal trend with more listings during the middle of the year (the summer months). Even with the seasonality the overall number of listings remained relatively steady. After the first peak in the covid rates, the number of listings dropped significantly. During the pandemic the number of listings followed the same seasonal trend, but overall there were much less listings during the pandemic than prior. Next, we will be looking at the sale to list ratio of homes in Philadelphia County.

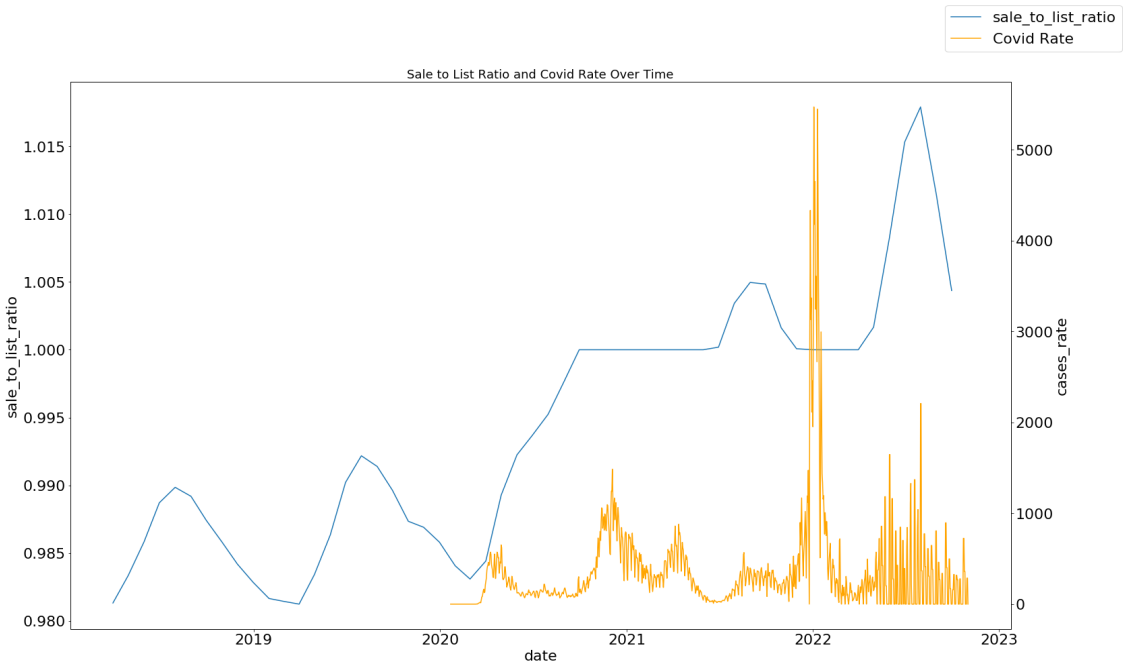

Figure 6

Prior to the pandemic the sale to list ratio looked pretty steady with some seasonality. Just from the visualization it seems as if the rate of spread of Covid-19 has some sort of relation with the sale to list ratio. Prior to the pandemic the average sale to list ratio remained under 1, meaning homes were usually sold for less than they were listed. After the start of the pandemic this value went up significantly with homes being bought for more than they were listed and this value continued to increase steadily over the course of the pandemic with stalls after peaks in the covid rate.

Finally I performed a t-test to see if the mean of the number of listings and the sale to list ratio changed before and after the pandemic. The hypotheses are listed above in the *Background* section. The results are shown below.

- Test 1: Number of Listings
  - Test Statistic: 15.408
  - P-value: 2.037 x 10-21

- Test 2: Sale to List Ratio
    - Test Statistic: -7.845
    - 1.97 x 10-10

Both these statistical tests result in a significant p-value indicating that the means before and during the pandemic are not equal. Along with the statistical tests shown above I was also able to review some articles to get quotes from other fields as well as testimonials by residents of Philadelphia discussing how the pandemic has affected their lives in terms of real estate. The quotes are listed below.

- "Typically, in our previous research looking at property tax increases, the evidence shows that the landlords pass it on to the tenants" - Michael Shields, director of research for the Economy League of Greater Philadelphia
- "There's going to be a bunch of landlords selling their properties because we can't pay for this anymore," (In response to tenants being unable to pay rent) - Ebony Harris, Philadelphia Landlord
- "The pandemic was kind of a pause on the economy and now that things are reopening, inflation is picking up, rents are going up and people are realizing they don't have as much disposable income as they might have thought they had." - Daryl Fairweather, Redfin's chief economist

We are able to see insights from fields other than data science with the testimonial from economist Michael Shields explaining other extraneous factors as to why rent increased later in the pandemic even when covid rates were at an all time high. A first hand testimonial from a Philadelphia landlord gives context as to why rent increases had stopped early in the pandemic with people being unable to pay rent so landlords were struggling to even get their initial rent let alone an increased rent. Another quote from a housing economist shows how the pandemic may affect housing past the end of the pandemic giving predictions on data we do not have yet.

Discussion/Implications

From our above findings we are able to see many interesting insights into human actions during the pandemic. From part 1 we can see that most people in Philadelphia county followed masking guidelines set by the city but there were still no significant trends in the Covid-19 rate during this mandate. There are however other extraneous factors that we will discuss in the *Limitations* section. Overall, I think the first part of this project was human-centered as we were able to see how the actions of humans affected a global pandemic, but there is much more that could be done. The visualizations we created only tell part of the story and more ethnographic research could explain the effectiveness or lack thereof of masks on the spread of Covid-19.

For the second part of the project I had more control on the methodology and data, so I believe I was able to make this a very human-centered analysis. With the three graphs we created for housing data as well as testimonials for residents and economists in Philadelphia we were given many insights into human actions during the pandemic. This was the most human-centered aspect of my project as I was able to gain context around the data being used to explore how something so crucial to human survival was being affected. The stunted growth in rent seen in *Figure 4* could be due to people losing their jobs during the pandemic and being unable to pay rent. The subsequent rise in rent in *Figure 4* can be explained by inflation and the rise of property taxes after the pandemic.

There was also an increased sale to list ratio. More people were buying homes during the pandemic but the number of listings fell. *Figure 5* and *Figure 6* are related as they show how a decrease supply in listings made people want to spend more money on housing. I think this would be an interesting topic for future research. We can look at how the average sale price changed over the pandemic and get testimonials from sellers, buyers and real-estate agents to

explore why people were so desperate to buy homes that they were going much above the asking price.

Limitations

As many data science projects go there are limitations to this analysis both in terms of data and conclusions. For the visualizations in part 1 it seems as if the mask mandate had no effect on the rate of Covid-19. It is hard to see just the effect of mask mandates on the number of covid cases because there are so many other factors to be considered. Holidays, reopening of bars and restaurants and vaccines are just some of the many factors that could also contribute to the rate of Covid-19 in Philadelphia. There are also many different types of mask mandates with some being just for indoor activities, only at bars and restaurants, etc. The type of mask mandate also may have had an effect on the rate. The hardest part of this analysis was trying to figure out and control for extraneous factors.

Much like part 1, there were also similar limitations for Part 2. The housing market depends on multiple different factors and the pandemic is just one of them. By looking at testimonials by economists we can also see that rising interest rates, property taxes and inflation all have an effect on the housing market which are not represented in the visualizations we created. There were also not many testimonials from a variety of residents in Philadelphia, County. I would have liked to get perspectives from both landlords and tenants as well as homebuyers to get a more well rounded ‘thick’ data. Along with that, this is real world data so the assumptions (normal distribution, homogeneity) for the t-test were not perfect which may result in inaccurate results.

Conclusion

Overall we were able to get many insights into the pandemic both in regards to mask mandates as well as housing. We first studied how mask mandates may affect the rate of spread of Covid-19. Although the visualizations show no correlation between the two we also discussed limitations in our analysis that may have affected that. This informed my study of human centered data science by showing the importance of looking at the big picture and not just focusing on the data in front of us. Humans are complicated and cannot be reduced to a set of data points and we must look at much more complex values behind them.

Next we were also able to see how the rate of Covid-19 affected the housing market in Philadelphia. We hypothesized that as the rate of Covid-19 increases the average rent decreases, when in actuality it remained steady while stopping the trend of steadily increasing rental prices. Second we hypothesized that as the average rate of Covid-19 increases the average sale to list ratio decreases which ended up being completely wrong. The sale to list ratio increased drastically and this is something I would further like to study as an ethnographic research analysis. Finally we hypothesized that as the rate of Covid-19 increases the number of listings decreases which in actuality was true. We also were able to see that the mean number of listings and sale to list ratio prior to the pandemic were not equal to during the pandemic.

Overall, this study informed my understanding of human centered data science by making me apply techniques learned over the course in my own ethnographic research. It is important to understand the limitations of your research and how the data analysis being done is actually affecting people, especially when it comes to things needed for human survival. I was

able to perform a well rounded research study by giving context to my data and thinking about the ethical ramifications of my research.

References

- Allen, Taylor. “Philadelphia's Small Landlords Have Run out of Time.” *WHYY*, WHYY, 17 Mar. 2021, https://whyy.org/articles/will-philadelphias-small-landlords-be-another-covid-19-casualty/.
- “As COVID Drives Home Prices Higher, Crunching the Numbers on Housing Affordability.” *Drexel Home*, 10 Jan. 2022, https://drexel.edu/news/archive/2022/january/as-covid-drives-home-prices-higher-crunching-the-numbers-on-housing-affordability.
- Bhattarai, Abha. “Four Reasons Your Rent Is Going Up.” *The Washington Post*, WP Company, 10 Feb. 2022, https://www.washingtonpost.com/business/2022/02/10/rent-rising-inflation-housing/.
- Seymour, Kara. “Coronavirus' Impacts on the Suburban Philly Real Estate Market.” *Newtown, PA Patch*, Patch, 26 Mar. 2020, https://patch.com/pennsylvania/newtown-pa/coronavirus-impacts-suburban-philly-real-estate-market.
- “What Happens next? Covid-19's Continuous Impact on Philadelphia's Real Estate Market.” *KSS Architects*, 8 Apr. 2021, https://kssarchitects.com/intersections/blog/2021/04/what-happens-next-covid-19%E2%80%99s-continuous-impact-philadelphia%E2%80%99s-real-estate.
- Williams , Stephen. “Landlords Expected to Raise Rents Due to City Reassessment.” *The Philadelphia Tribune*, 12 July 2022, https://www.phillytrib.com/news/local_news/landlords-expected-to-raise-rents-due-to-city-reassessment/article_c640a6ab-22ce-5e21-a042-871142d4f1a9.html.

Data Sources

- US Confirmed Cases: Covid-19 Data from John Hopkins University covering confirmed cases and deaths on a country and US county level
- Mask Mandate Data: U.S. State and Territorial Public Mask Mandates From April 10, 2020 through August 15, 2021 by County by Day from the CDC
- Mask Compliance Survey: Data from the New York Times mask compliance survey
- Vaccine Data: Vaccine Administration data from the Philadelphia Health Department
- Rentals: Data about Zillow Observed Rent Index (ZORI), a smoothed measure of the typical observed market rate rent across a given region.The index is dollar-denominated by computing the mean of listed rents that fall into the 40th to 60th percentile range for all homes and apartments in a given region. This data is separated by counties all across America and has data monthly from March 2015 to September 2022. There is also a dataset available that adjusts for seasonality. This will help me answer the question by giving access to how the mean rent has changed before, during and after the pandemic which will allow us to compare that change to the rate of change in covid cases.
- For Sale Inventory: Data about current listings on Zillow. The data includes the count of unique listings that were active at any time in a given month. This includes data from March 2018 to September 2022 with the number of listings reported for each month. This will help me answer the question by giving access to how the number of listings has changed before, during and after the pandemic which will allow us to compare that change to the rate of change in covid cases.
- Sale-to-List Ratio (mean/median): This includes data from March 2018 to September 2022 with the ratio of sale vs. final price monthly. This will help me answer the question by giving access to how the sale to list ratio has changed before, during and after the pandemic which will allow us to compare that change to the rate of change in covid

cases. We will be able to examine if houses sold over or under the listing rate over the course of a couple years.