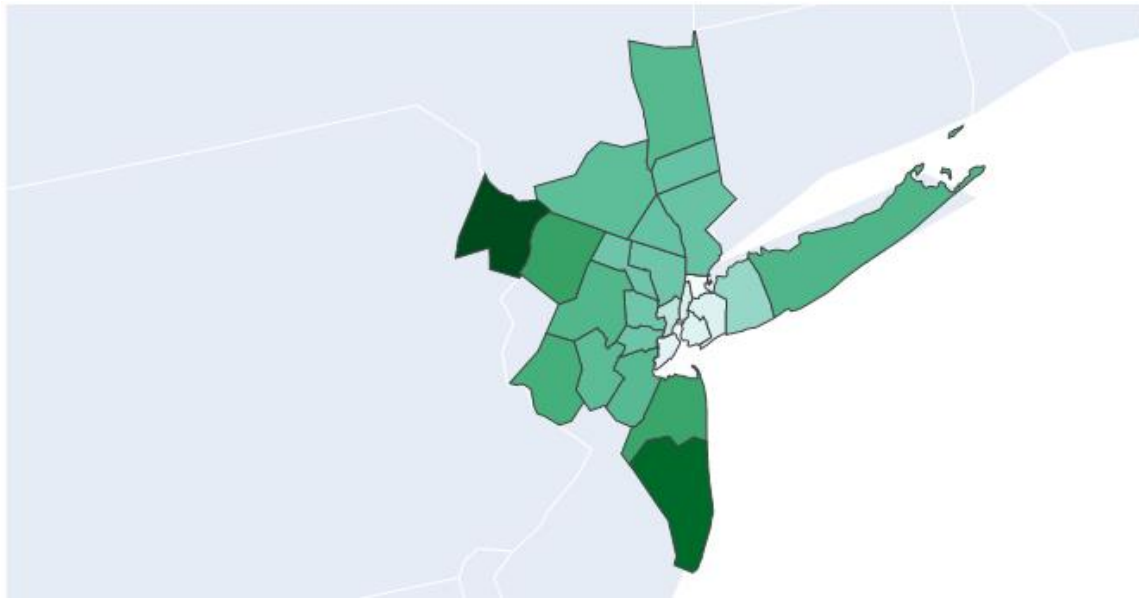


# The Impact of Covid-19 on the Housing Prices in The United States

An exploratory analysis

---



Kisha Kim, Sho Tsuyuki , Karthikeyan Ragunathan (Karthik)

August 2021

# 1. Introduction

## < Overview >

Due to the recent pandemic, there has been big shifts in the way in which people live and work. Change in lifestyle, where people no longer have to commute and spend most of their time at home, has led many people making decisions like moving to a different city or buying a house. This then led to the housing market behaving in a new pattern that we have not seen in the past. Our research will leverage the Covid-19 data, and the housing prices data in various cities to understand this change in the U.S Real Estate market.

We will start by exploring both the housing market data and covid data from the past 18 to 24 months at the State level to understand the trend of housing prices in each State, and if we can find any correlation between number of Covid cases and the change in housing prices. Next, we will deep dive into housing and covid data for a few select US metropolitan areas to understand the regional trends, if any, and the relationship of covid cases and the housing market. Finally, we will analyze the change in housing prices in downtown neighborhoods and suburban neighborhoods to see if the pandemic accelerated the change in housing prices in certain areas.

## < Research Question/Hypothesis >

Through our research, we plan on answering the following questions:

- Did the Covid-19 impact the housing price change in the past 2 years?
- Did the Covid-19 and the subsequent shift in working styles increase housing prices in suburbs of large cities?

The hypothesis for our research is that due to the recent pandemic, housing prices were impacted in a unique pattern compared to historical changes. Additionally, more people have been moving out to the suburban area, where the population and the covid rates are low. Due to the increased demand in these areas, the housing price has been fluctuating more in specific neighborhoods. Conversely, urban areas where people mostly resided for work/school, the population has decreased, resulting in a negative/less impact on the housing price.

## < Data >

Below are primary data sources used in the analysis:

- Covid Data
  - U.S. Statewise Latest Covid-19 Data: This details the number of Covid cases, deaths and recoveries against the population of each State  
(<https://www.kaggle.com/anandhuh/usa-statewise-latest-covid19-data>)
    - Columns['State', 'Death/1 mil population', 'Total Cases/1 mil population', 'Total Tests/1 mil population']
  - County level Covid Data:
    - NY State county covid cases: cases-by-day.csv  
<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/state/new-york>
      - Columns['date of interest', 'case count 7day avg']
    - Washington State county covid cases: wa covid19 cases.csv

<https://www.doh.wa.gov/Emergencies/COVID19/DataDashboard#downloads>

- Columns['County', 'WeekStartDate', 'ConfirmedCases']
- Housing Price Data
  - Housing Prices in the U.S.: This is the data for housing prices across US States and metropolitan areas, including some divide between 'City' and 'Suburb', which is part of what we will explore related to covid. (<https://www.zillow.com/research/data/>)
    - Zillow\_price\_data\_by\_state.csv
      - Columns ['RegionID', 'SizeRank', 'RegionName', 'RegionType', 'StateName', '2019/07/31', '2019/08/31', ... '2021/06/30']
      - Drop: Regional ID
    - Zip\_zhvi\_uc\_sfrcondo\_tier\_0.33\_0.67\_sm\_sa\_month.csv
      - Columns ['RegionID', 'SizeRank', 'RegionName', 'RegionType', 'StateName', '1996-01-31', ... '2021/06/30']
      - Drop: Regional ID, SizeRank, '1996-01-31'... '2019/06/30'
    - County\_zhvi\_uc\_sfrcondo\_tier\_0.33\_0.67\_sm\_sa\_month.csv
      - Columns ['RegionID', 'SizeRank', 'RegionName', 'RegionType', 'StateName', '1996-01-31', ... '2021/06/30']
      - Drop: Regional ID, SizeRank, '1996-01-31'... '2019/06/30'
- FIPS Data
  - This data was used to merge with housing prices data and plot on a geographic map using plotly. The FIPS data was taken from this source (<https://github.com/kjhealy/us-county/blob/master/data/census/fips-by-state.csv>) but is commonly available from the US census bureau.

### **<Assumptions>**

Below are assumptions that were made during this research:

- Assumption was made that people are moving within the state between urban/suburban. We acknowledge that there are cases where people move out of state. This research is limited non-out of state moves and the price impact due to this.
- Metropolitan area: when looking at the city level housing price change and covid cases, we used the Metropolitan Statistical Area defined by The U.S. Office of Management and Budget. The conventional metropolitan area usually means an urban area around major cities that include satellite cities, towns and rural areas that are tied to the urban core, measured typically by commuting patterns.

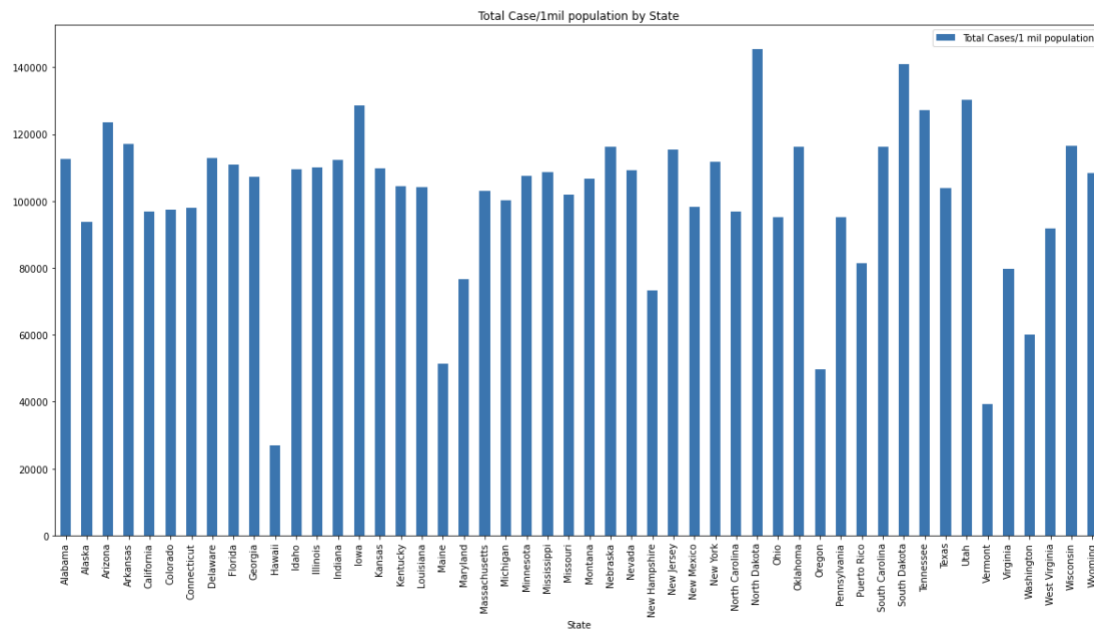
## **2. First level analysis at the State level**

We started looking at both covid data and housing price change data at the state level.

### **< Covid Cases >**

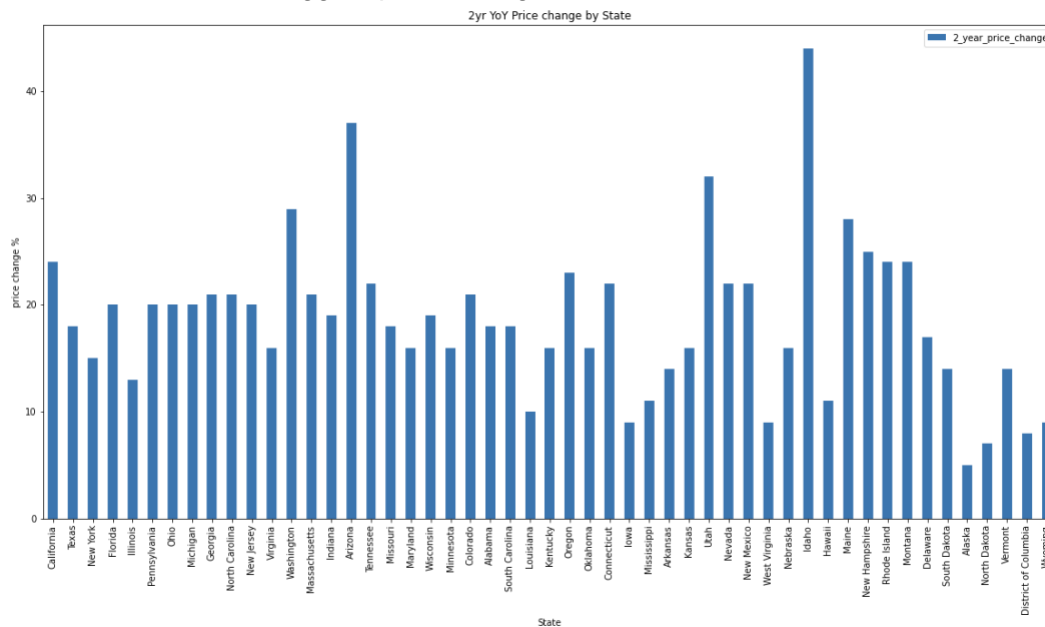
There are three different variables that we are focusing on in the Covid data - Total Cases/1 mil population, Death/1 mil population, Tests/1 mil population. We've focused on the Total Cases/1mil population variable by each State. See Appendix A for additional covid measures.

Below chart identifies the total cases per million population by State. For the most part, it's an even spread in total number of cases in most of the state.



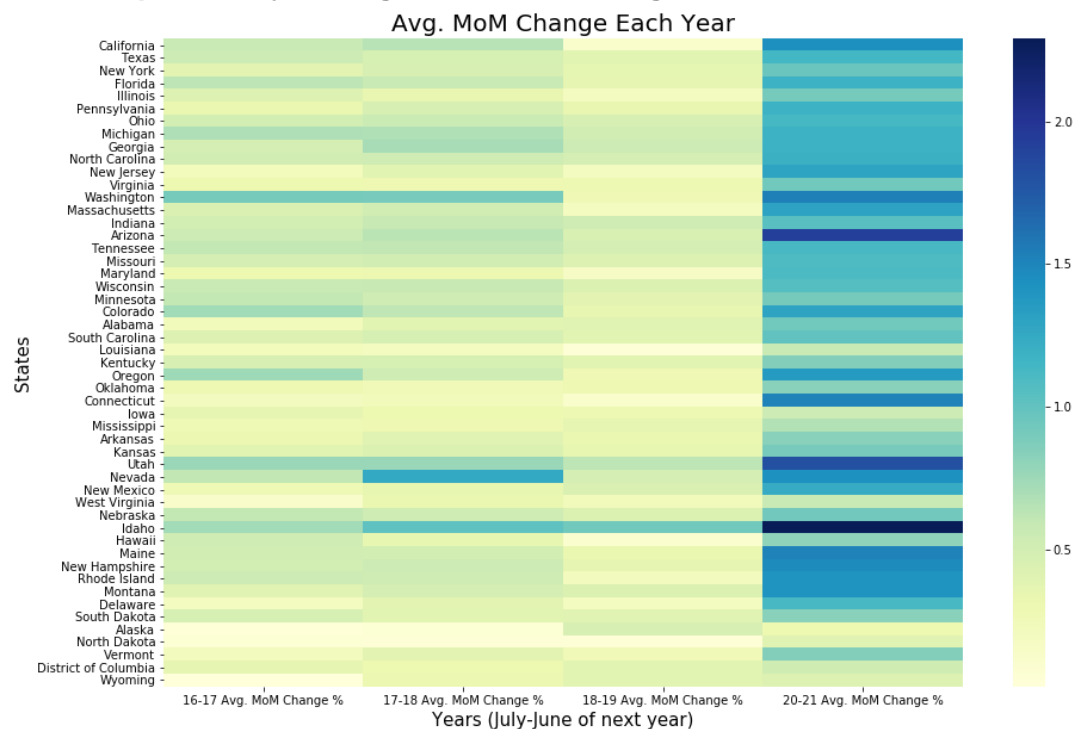
## < Housing Prices >

We've analyzed the year over year housing price change for the past 2 years by State. Idaho and Arizona had the biggest price change.



We've also analyzed the month to month housing price change by State.

### < Heatmap of Yearly Average MoM Price Change >

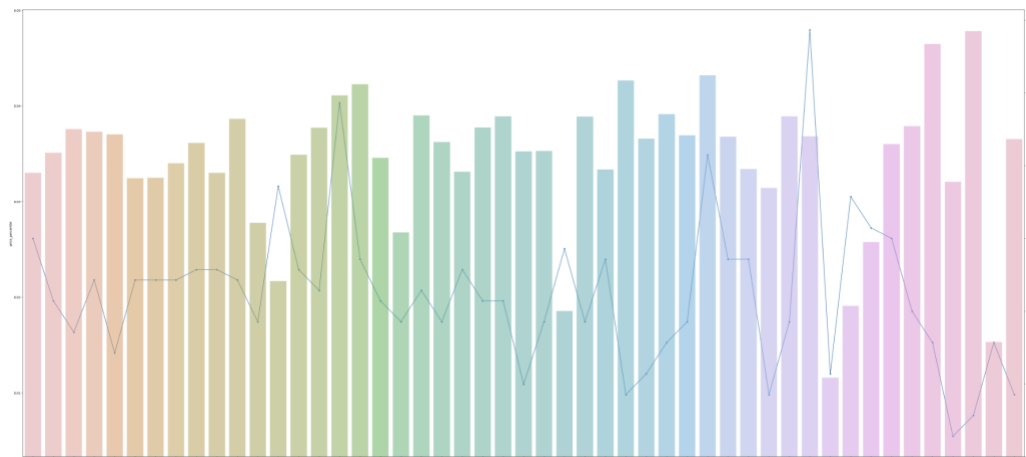


By using a heatmap to show the average MoM change of the 3 years prior to the Covid-19 Outbreak, and the year after the Covid-19 Outbreak, we can clearly see that the US has seen a dramatically higher MoM increase in housing prices compared to before. This helps us validate that the change in housing prices in the past year is greater than the organic increase in housing prices that we may have experienced if the Pandemic had not happened.

### < Correlation between Covid data and housing price change >

We wanted to understand the correlation between covid cases and housing price fluctuation in the past 2 years at the state level. We used seaborn to depict the price change in a line graph, on top of total case per million population in a bar graph. Here, out of all 3 variables related to covid, we focused on total case per million to see the correlation between price change impact

and the covid cases.



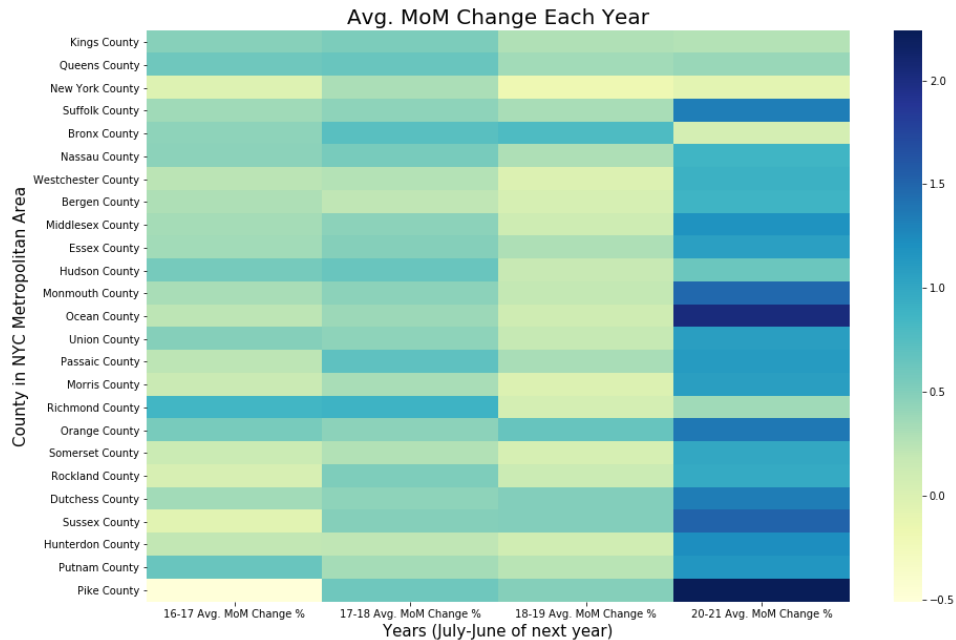
See Appendix B for additional visualization on covid and housing price change correlation.

**Takeaway:** In the correlation analysis, we've noticed that Idaho and Arizona may be states that their housing price got impacted due to Covid. Before looking at the data, we've had assumed that New York and California would have had the biggest impact, which was a bit different.

We wanted to deep dive into a few states to analyze the data at the county level to continue with our hypothesis/research question. As there wasn't a clear state that was indicated by the state level research, we've picked New York and Washington, as our teammates reside in and are familiar with the suburban/urban areas of the state.

### 3. County Level Deep Dive - Case: New York Metropolitan

The same operation of State level MoM housing data change can be looked at at a county level. We will be looking at the New York City Metropolitan area to understand the difference in price behavior between city centers and suburbs. For the definition of Metropolitan area, we will be using the New York-Newark-Jersey City, NY-NJ-PA Metropolitan Statistical Area defined by The U.S. Office of Management and Budget (see appendix for more information). A similar method of data cleaning as the State level can be carried out to arrive at a county level price change heatmap to understand the price change in the New York Metropolitan area, and validate that the increase in housing prices in 2020 to 2021 is more than an organic, normal market price increase.

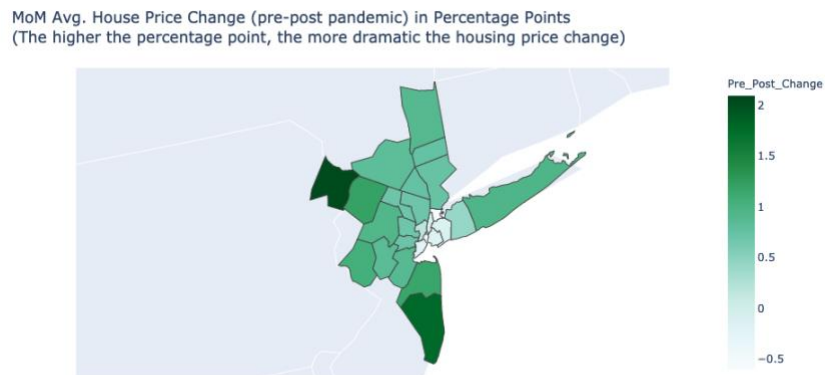


We then looked at a 3 year average of the MoM price change in housing in the area, then compared that to this past year's (July 2020 to June 2021) price change. The delta in percentage points in the housing market post covid can show us how dramatic the price change is since the covid pandemic, compared to the average pre-pandemic.

RegionName	3 Year Avg. MoM Change %	20-21 Avg. MoM Change %	Pre_Post_Change
Kings County	0.44	0.27	-0.17
Queens County	0.54	0.40	-0.14
New York County	0.04	-0.07	-0.11
Suffolk County	0.38	1.33	0.95
Bronx County	0.66	0.06	-0.60
Nassau County	0.44	0.87	0.43
Westchester County	0.17	0.91	0.74

A glimpse into the numbers show that certain counties in the NYC metropolitan area actually saw a decrease in the MoM average change. Someone familiar with the area will be able to tell that all the counties listed here that saw a decrease in the MoM price change was in New York City, whereas the counties that saw an increase in MoM price change were the suburbs. For example, New York County refers to the borough of Manhattan, and Kings country refers to the borough of Brooklyn, which both saw a decline in average MoM price change, but Nassau county in Long Island, NY, and Westchester county, situated in the north of New York City are considered the typical suburbs of New York City, both saw a jump in housing prices.

Finally, by merging this data with FIPS (Federal Information Processing System) code and using Plotly, we can plot these, we are able to show the difference between the pre-covid average monthly change in price and post covid average monthly change in price even for those that are not familiar with the county names in the region.



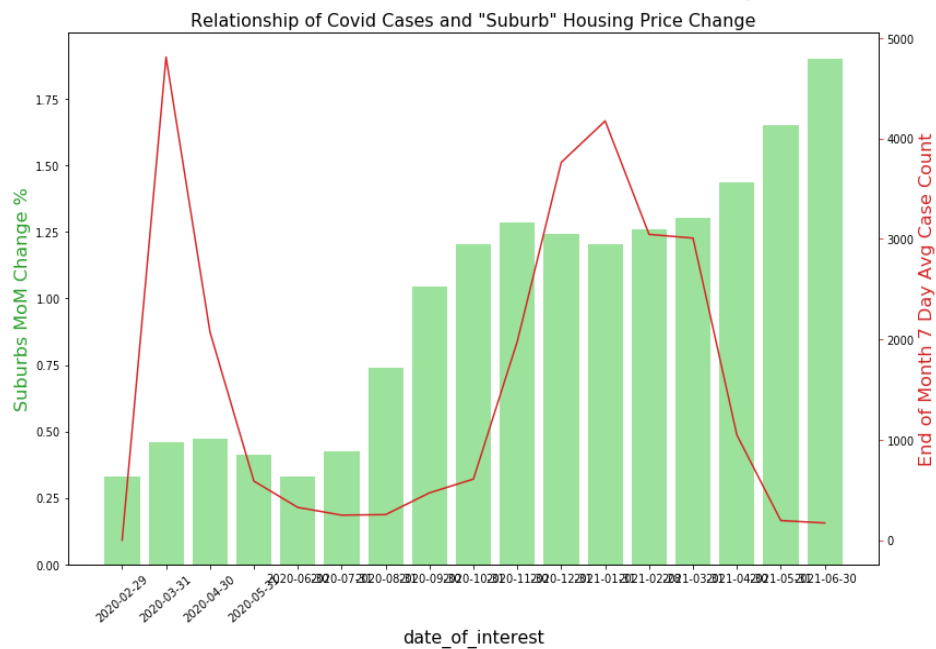
This map clearly shows that the area with the light colors is where New York City is (Manhattan, Queens, Brooklyn), housing prices MoM declined on average compared to normal times, whereas the suburbs of the Metropolitan area saw a sharp rise in MoM price change compared to normal times.

This helps us partially answer our questions about the impact of Covid-19 on housing in cities. It looks like New York City is seeing a price decline in the city center most likely driven by an outflow of people from the city center to the neighboring counties, perhaps in search of more space, which is causing the prices in the suburbs to increase. It also looks like the farther you go from the city center the more dramatic this change is post covid.

Lastly, if we grouped the housing data by 'City' counties (New York, King, Queens, Richmond and Bronx)(New York County, for example, is Manhattan), we are able to see not just the pre-post covid price change, but also how the housing prices behaved MoM during the covid pandemic. The first chart, which shows the price change in the city, indicates that as covid hit back in March of 2020, prices in the city were negative MoM for a few months. After that, as the summer months came and cases started to decrease (red line chart) the housing prices MoM started to rise again, but starting in November of 2020 when New York saw its second big wave of cases, MoM prices started to shrink. On the other hand, housing prices in the suburbs increased throughout the pandemic, and started to see a dramatic spike especially 6 months in to the pandemic, as people may have realized that this was going to be a long term issues, and



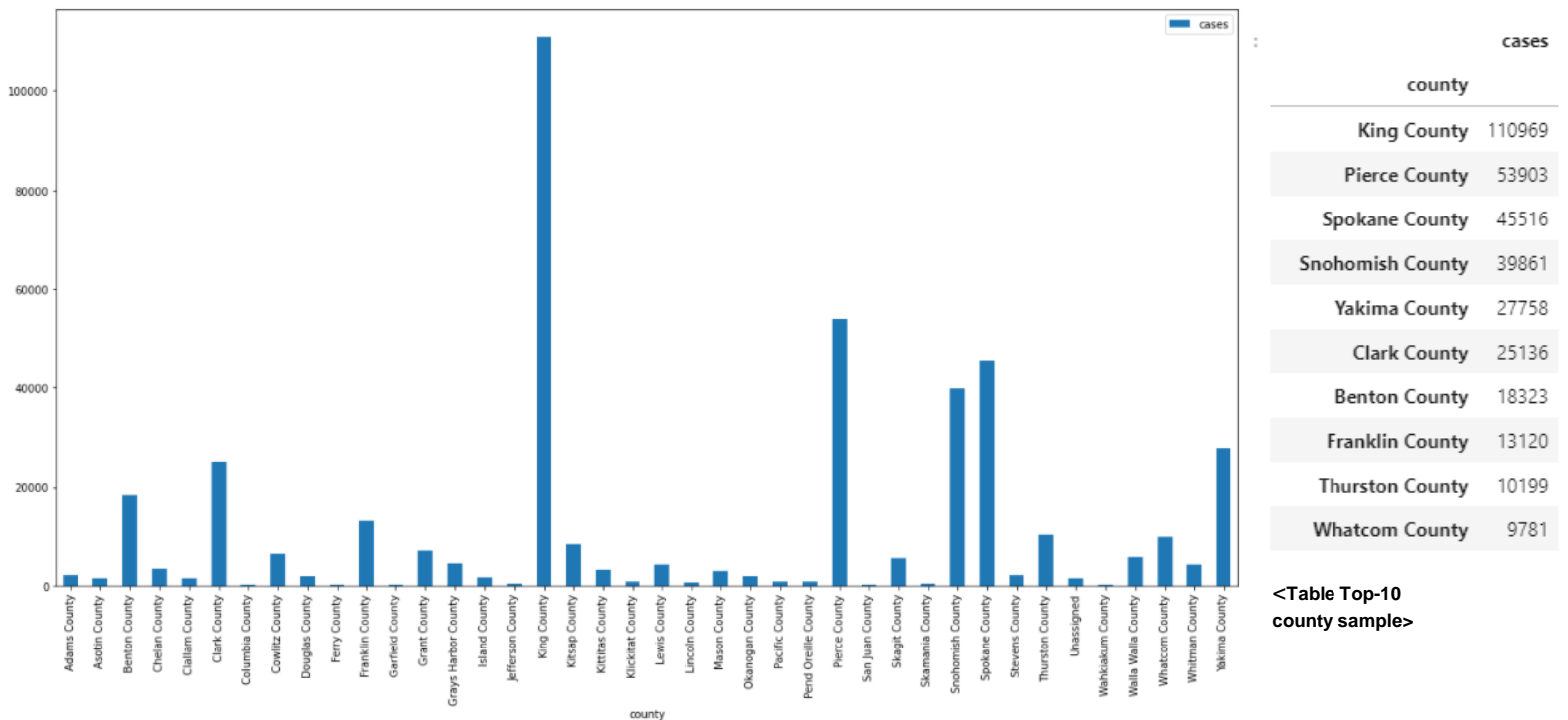
started to move more to the suburbs in search for more space



**Takeaway:** By looking at the data from multiple lenses, we are able to see that the New York Metropolitan Area's housing price was significantly affected by the Covid-19 pandemic, and the trends of city dwellers in New York City moving out to the suburbs is reflected in the housing price data.

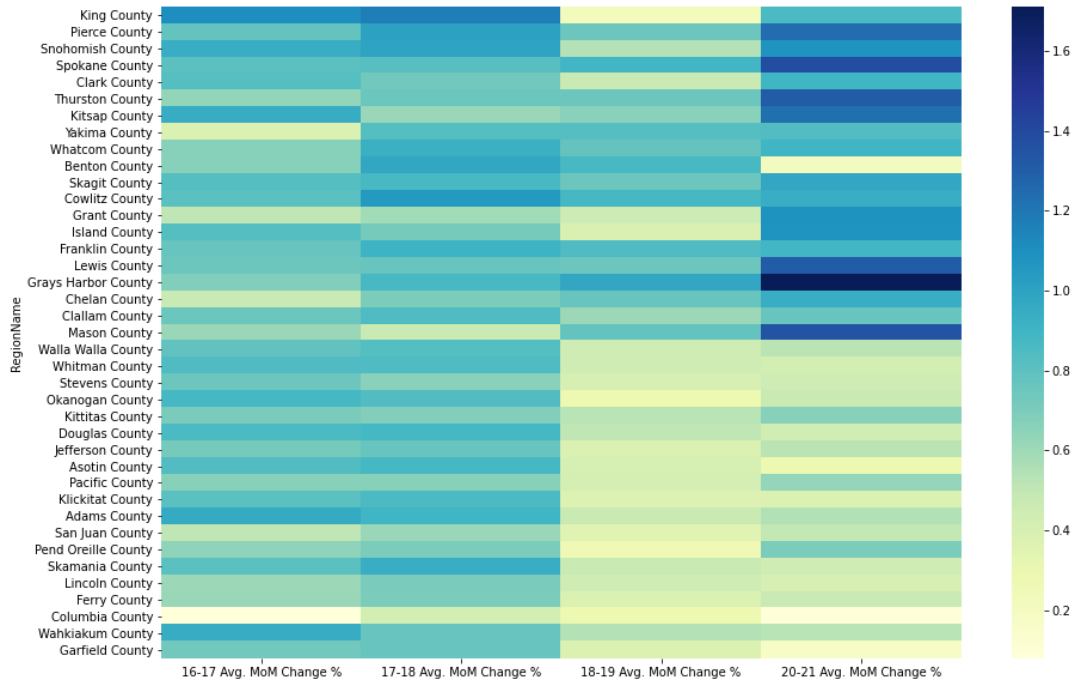
## 4. County Level Deep Dive - Case: Washington State

In case-2 we look at the state of Washington at the county level.

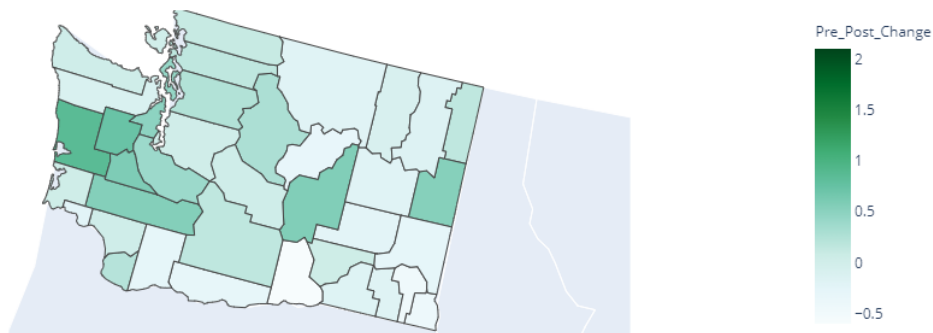


We can see that King county, which includes cities such as Seattle, has the highest number of Covid cases. Unlike the metropolitan cities of NY (NY-NJ-PA), King county is a combination of metro city and suburban region. We follow the similar method of data cleaning as we did with State and NY county levels to arrive at the county level price change heatmap to understand the price changes at each county in Washington state.

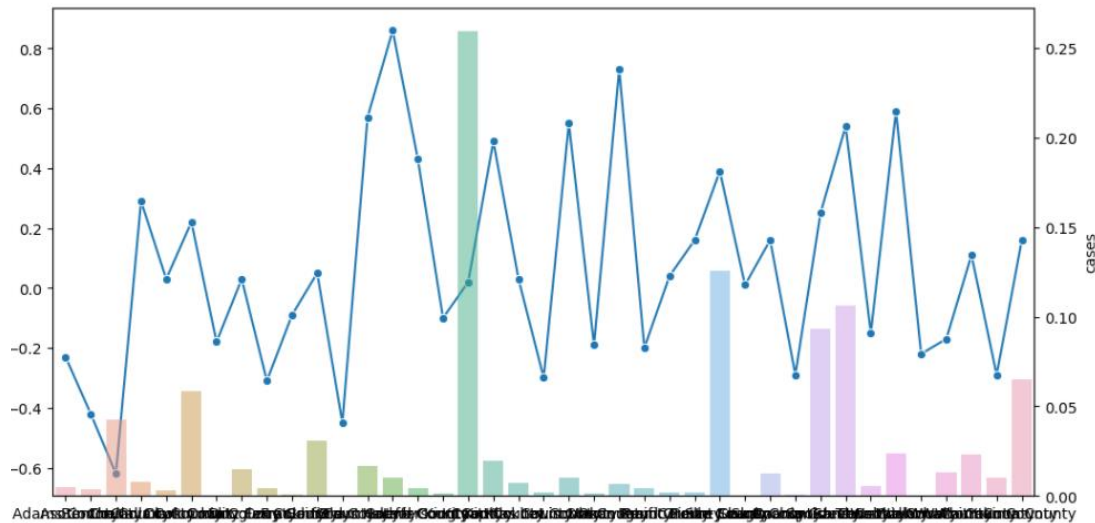
We generate the heatmap below using WA counties and month-on-month % housing price change from 2016-2019 (Pre-covid) to 2020-2021 (Post-covid)



Washington State Counties MoM Avg. House Price Change (pre-post pandemic) in Percentage Points  
(The higher the percentage point, the more dramatic the housing price change)



The heat map and the data show that the year 2020-2021 average month on month home price change has been higher than previous years pre-covid 2016-2019. But the interesting finding is that King County which has the metro city Seattle actually didn't have the highest rise in home prices post-covid compared to pre-covid. Instead we see suburban counties such as 'Gray's Harbor County' and 'Mason County'



### Takeaway:

The rise in Covid cases in metro areas in King County and subsequent rise of home prices in suburban counties other than King County might suggest that the home prices in metropolitan areas were affected by the covid pandemic, unlike home prices in suburban areas.

## 5. Conclusion

Through our research, we wanted to answer the following questions:

- Did the Covid-19 impact the housing price change in the past 2 years?
- Did the Covid-19 and the subsequent shift in working styles increase housing prices in suburbs of large cities?

For the first question, we can confidently say that Covid-19 impacted the housing price change in the past 2 years at the national level, and that last year's housing price change is significantly different from organic or regular market driven changes.

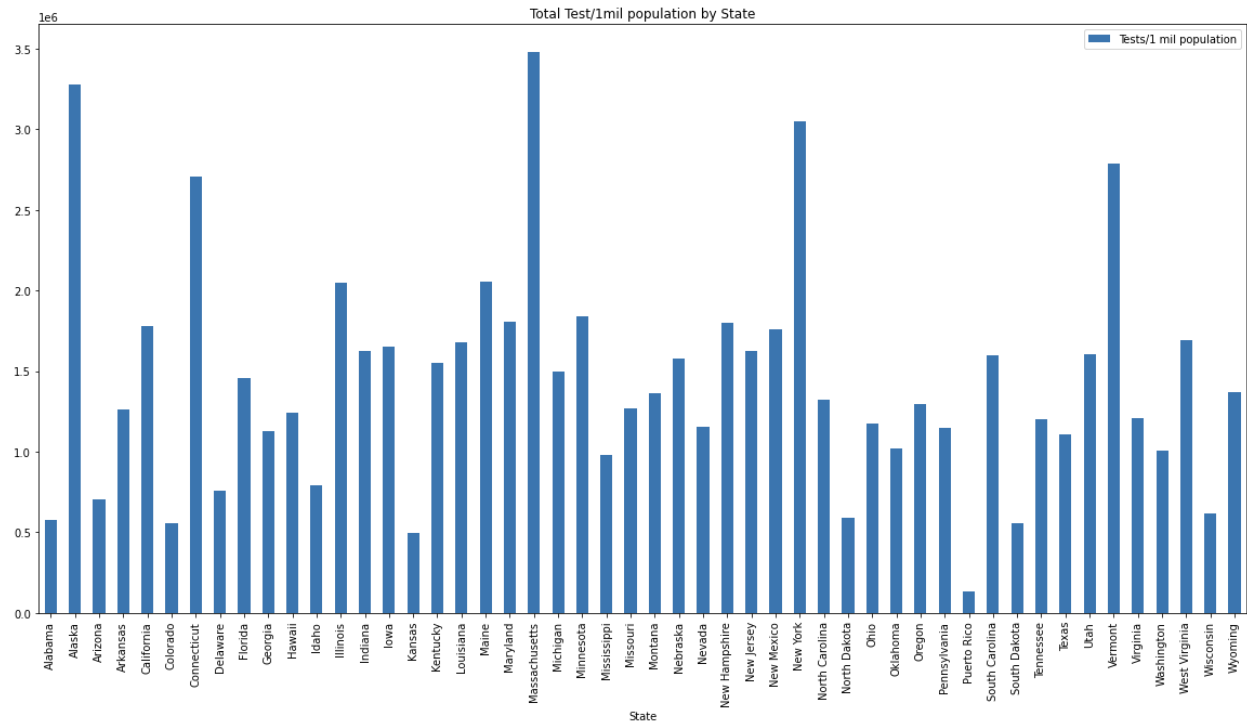
For the second question, we can also say that, yes, Covid-19 did cause a shift in housing price increasing more in the suburbs compared to the city.

## 6. Appendix

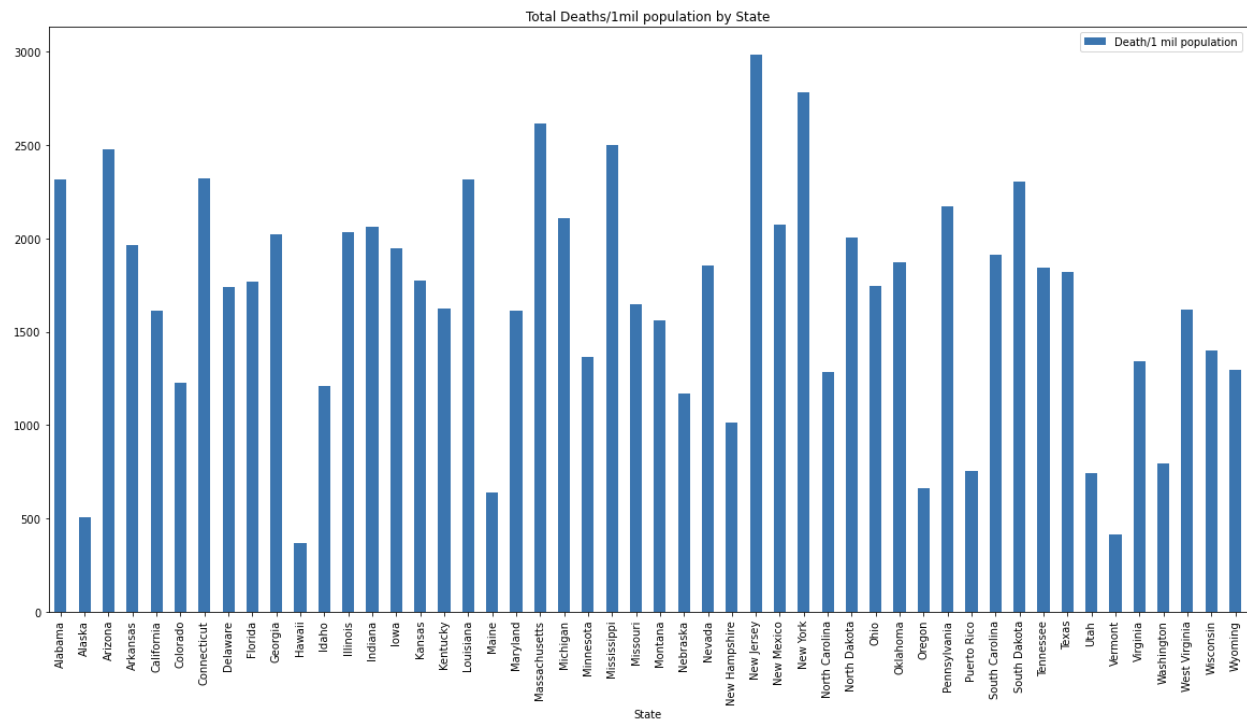
### A. Additional Covid Charts by State

Below are additional charts analyzed for Covid data by State. We wanted to look at variables outside of Total Cases, as this might not holistically reflect of number of tests, which shows how people are cautious of Covid, and number of deaths in each state.

Below chart identifies the total tests per million population by State. Massachusetts and Alaska had the highest number of total covid tests per million population.



Below chart identifies the total deaths per million population by State. There isn't a small number of states that have way higher deaths over any other state. New Jersey, New York came in as two top states with covid deaths per million population.



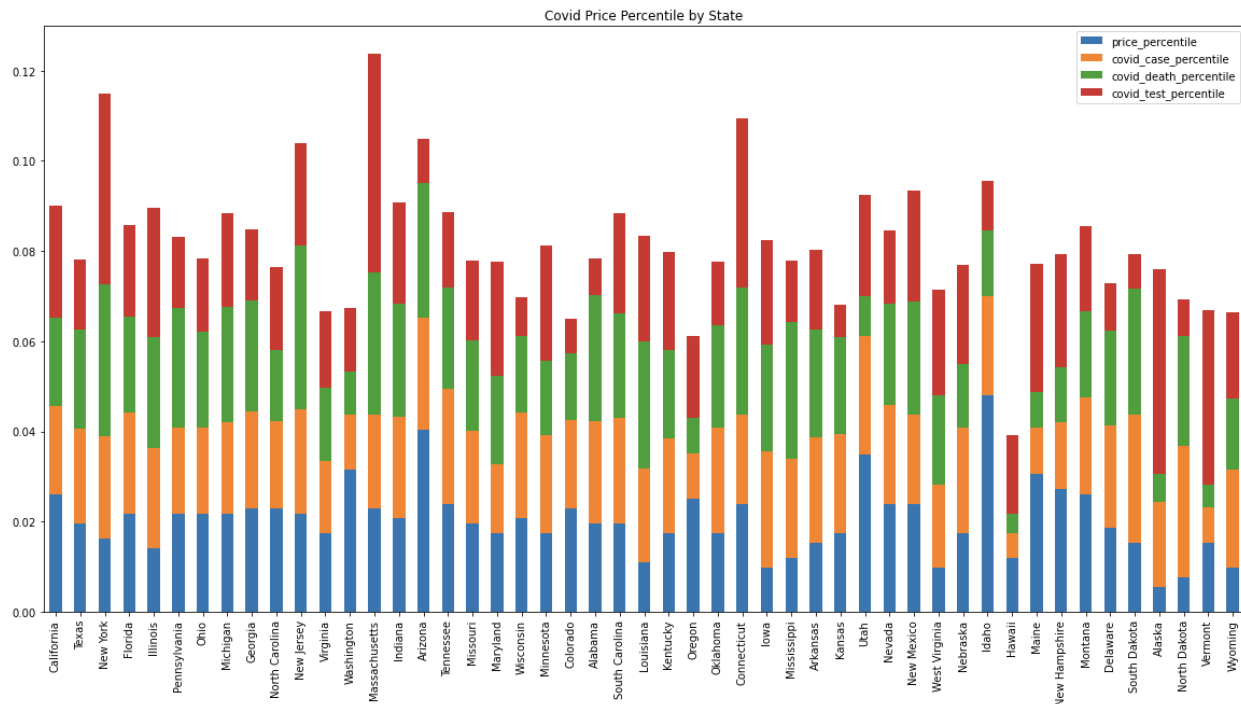
B. Additional Covid and housing price correlation analysis by State

In Covid data by state, we focused on three variables - Total Cases/1 mil population, Death/1 mil population, Tests/1 mil population. In the zillow home price data by state, we focused on the 2 year YoY price change by state.

Then, we've cleansed the data and created a percentile score for each variable by state :

- Price Percentile : 2 yoy price change by state / Sum(2yoy price change for all states)
- Covid Case Percentile : Total Cases/1mil population by state / Sum(Total Cases/1mil population for all states)
- Covid Death Percentile : Total Death/1mil population by state / Sum(Total Death/1mil population for all states)
- Covid Test Percentile : Total Test/1mil population by state / Sum(Total Test/1mil population for all states)

Below is the stacked bar plot using four percentiles calculated. It indicates that Massachusetts and New York were the top states that had the top score per our analysis.



### C. Data Cleaning Process for Zillow House Data

Zillow provides a robust set of housing prices data for the public called Zillow Home Value Index (ZHVI). These data sets are "A smoothed, seasonally adjusted measure of the typical home value and market changes across a given region and housing type"<sup>1</sup>

There are numerous ways to segment and download these data sets, and if we download the State level data and open it in Python, we can see a glimpse of a fairly clean time series data for housing prices by State.

<sup>1</sup> <https://www.zillow.com/research/data/>

Out [26]:

	RegionID	SizeRank	RegionName	RegionType	StateName	1996-01-31	1996-02-29	1996-03-31	1996-04-30	1996-05-31	...	2020-09-30	2020-10-31	2020-11-30
0	9	0	California	State	CA	160569.0	160342.0	160195.0	159908.0	159714.0	...	589875.0	596686.0	6031
1	54	1	Texas	State	TX	98663.0	98688.0	98717.0	98844.0	98983.0	...	219097.0	221039.0	223:
2	43	2	New York	State	NY	138610.0	138296.0	138111.0	137831.0	137729.0	...	336932.0	339991.0	343:
3	14	3	Florida	State	FL	101472.0	101569.0	101682.0	101894.0	102097.0	...	262308.0	264696.0	267:
4	21	4	Illinois	State	IL	127993.0	127980.0	127893.0	127957.0	127876.0	...	212960.0	214764.0	216:

5 rows x 311 columns

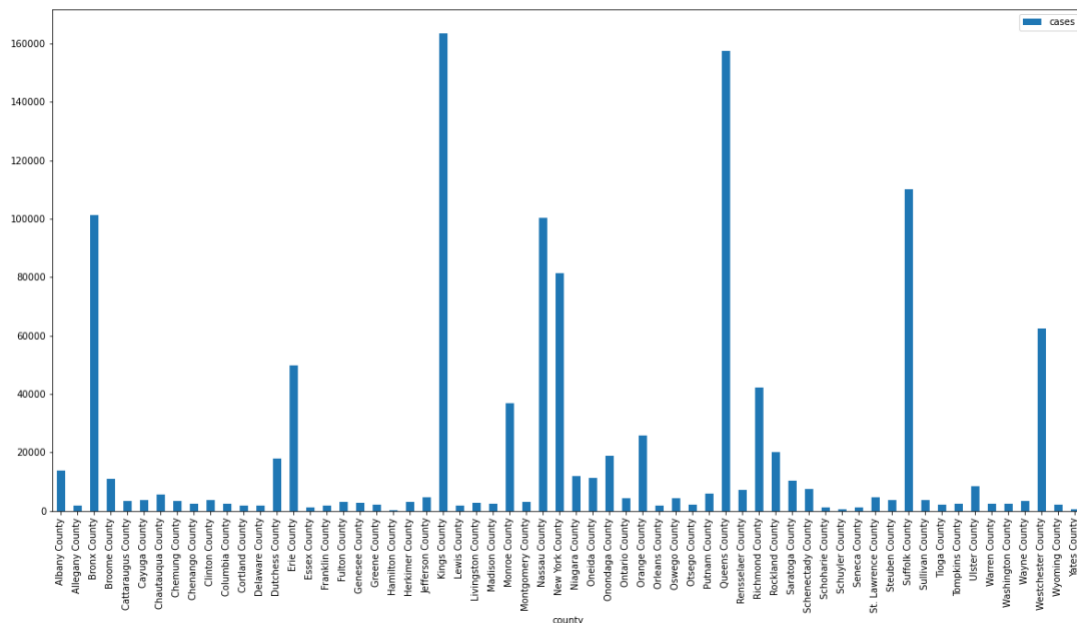
Using the 'shape' method, we are able to see the size of this data set is (51, 311), and the columns represent each month, starting from 1996-01-31 to 2021-06-30. For the initial purpose of our research, we dropped most of the data prior to 2016 since for the purpose of this research we only plan to look at data dating back 3 years before covid.

Once irrelevant columns are dropped, we can create a data frame that looks at the month over month (MoM) price change for each state over the past 4 years, and aggregate that data into a yearly average of MoM price change by each state.

#### D. DataFame of MoM Price Change

	2016-06-30	2016-07-31	2016-08-31	2016-09-30	2016-10-31	2016-11-30	2016-12-31	2017-01-31	2017-02-28	2017-03-31	...
RegionName											
California	NaN	0.004868	0.006204	0.005686	0.006196	0.005216	0.004494	0.004938	0.005304	0.006321	...
Texas	NaN	0.006071	0.005261	0.005593	0.005037	0.006256	0.006565	0.005870	0.005792	0.005043	...
New York	NaN	0.002481	0.002874	0.003051	0.003528	0.003749	0.004623	0.004378	0.003360	0.002948	...
Florida	NaN	0.007231	0.006817	0.006026	0.005537	0.006361	0.006099	0.006306	0.005614	0.005916	...
Illinois	NaN	0.004229	0.004006	0.003406	0.003768	0.004601	0.004951	0.005290	0.004050	0.004781	...

#### E. NewYork County Total Covid cases (Year 2021)>



## F. WA State counties month-to-month housing price change data

RegionName	16-17 Avg. MoM Change %	17-18 Avg. MoM Change %	18-19 Avg. MoM Change %	20-21 Avg. MoM Change %
King County	1.11	1.17	0.23	0.86
Pierce County	0.78	1.01	0.75	1.24
Snohomish County	0.94	1.00	0.54	1.08
Spokane County	0.80	0.82	0.89	1.38
Clark County	0.83	0.73	0.47	0.90
Thurston County	0.64	0.76	0.75	1.31
Kitsap County	0.95	0.62	0.66	1.23
Yakima County	0.38	0.83	0.83	0.84
Whatcom County	0.67	0.93	0.78	0.90