The COVID-19 Impact On The New York City Metropolitan Housing Market

Tarmah Abbas

Ramiro Baldeon

Nicholas Zakelj

Zahyra Ceballos

Lilliane Liang

Executive Summary

This project aims to look at the impact that the Covid-19 pandemic had on the housing market in the New York Metropolitan Area and use data mining techniques to gain insights into how the housing market would behave in the future. The scope of the project was based on public real estate and COVID-19 data from the City and State of New York and a Python Crawler and API were used to crawl and scrape price and sales data from Zillow.

Our group then split up the housing data into two sections, before the pandemic and during the pandemic. We then were able to split the New York City data by the five boroughs and were able to look at price changes over time for each of the boroughs as well as data for Westchester County. By further cleaning the data to only consider residential properties we were able to get a better understanding of how residents of the city reacted to being in a densely populated area during the pandemic. To examine the price changes and visualize the contrast between the different boroughs the python packages Seaborn and Matplotlib were used.

After looking at the initial results of our analysis we decided to further group our data by neighborhood as well as property type. We were able to see which neighborhoods in all the boroughs and which property types were affected the most. Using this data we were able to see that the number of sales for homes in New York City dropped significantly in 2020 and the borough of Manhattan experienced the greatest drop in sales. This drop in sales wasn't permanent as we were able to see a recovery in the market beginning in 2021 in which the numbers returned to pre-pandemic levels. During this time we were also able to see an increase in sales as well as prices for homes in Westchester County, showing that many people were looking to leave the city and find a place of residence in a place less densely populated.

This project is highly relevant in the context of housing shortages and a general increase in home prices. The COVID-19 pandemic is an event that has affected multiple lifestyles and its ramifications continue to have a significant impact on the real estate sphere. Faced with this panorama, this project seeks to find solutions and a future direction to address this problem in New York City.

Business Goal

In 2020, the economy of the United States was disrupted by the outbreak of COVID-19, which led to the shut down of the vast majority of businesses and disrupted the lifestyles of thousands of people around the world. As a direct result of this pandemic, metropolitan areas, specifically New York City's housing market was affected to a great extent. According to data made available by the Federal Reserve, the supply of residences has dropped to an all-time historical low whilst the price increase has heightened, thus affecting the chain of supply and demand.

In order to further investigate the housing marking during COVID, the use of data mining and analytical techniques can be implemented. Thus, the major goal of this project is to explore the housing market and do such. This goal was achieved through a comparison of the housing market with data extracted from Zillow for the metropolitan area of New York City. Indicators such as home price history, date sold, building class amongst other factors were taken into consideration when examining the COVID-affected housing market. In order to address the goal of this project and to compare the real estate data associated with the selected area for analysis, we used web crawling techniques with the aid of Python. Furthermore, we performed a descriptive analysis as we were reliant on the historical data and also normal calculations. Through the use of the data obtained with web crawling techniques such as the HTTP requests to crawl data from Zillow can demonstrate how COVID affected demand and supply and whether it increased, decreased, or steadied the prices of property.

As the data was openly available in public domains we used various techniques which were introduced in web analytics to mine the required data. With the data of the sales from the years 2019, 2020, and 2021 we were able to provide charts for visual understanding. With the results, both buyers and sellers can gain an insight into the market pre-COVID, when COVID first was discovered and now. The quantity of available properties can be a result of the owner hoarding the properties waiting for the seller market to boom so they could gain the most profit.

Additionally, with insights gained from this analysis landlords can estimate when the market will decline thus taking off their listings and waiting until the market improves to increase their profits. With the landlord benefitting from such data, buyers can also buy knowing when would be the best time to purchase real estate. Overall, the data provides the buyers and sellers with insight into sale trends surrounding an economy-altering situation such as a pandemic.

Dataset Description

As we are trying to investigate the data from two periods: pre-COVID and post-COVID, this need presented a challenge for the project – both the timeframe and the overall amount of

data would be enormous. In this case, we chose to go with both the available off-the-shelf database and web crawling from scratch.

After some research, we went with the official NYC property sales database from NYC OpenData. This database, originally, contains 29 columns describing a wide variety of aspects on the properties in NYC. Furthermore, it contains over 400k rows of data, ranging from 2018 to 2021, which gives us a long enough history to better understand this market.

In the end, we chose to adopt the columns listed in the following chart:

Column Name	Column Description
Borough	The name of the borough in which the property is located
Neighborhood	DOF assessors determine the neighborhood name in the course of valuing properties.
Building Class Category	This identifies properties with broad usage (e.g. One Family Home).
Building Class As Of Final Roll	The building classification is used to describe a property's constructive use
Address	The street address of the property as listed on the Sales File. Coop sales include the apartment in the address field
Zip Code	The property's postal code
Residential Units	The number of residential units at the listed property
Commercial Units	The number of commercial units at the listed property
Gross Square Feet	The total area of all e floors of a building as measured from the exterior surfaces of the outside walls of the building , including the land area and space within any building structure on the property
Sale Price	The price paid for the property
Sale Date	Date the property sold

System Design
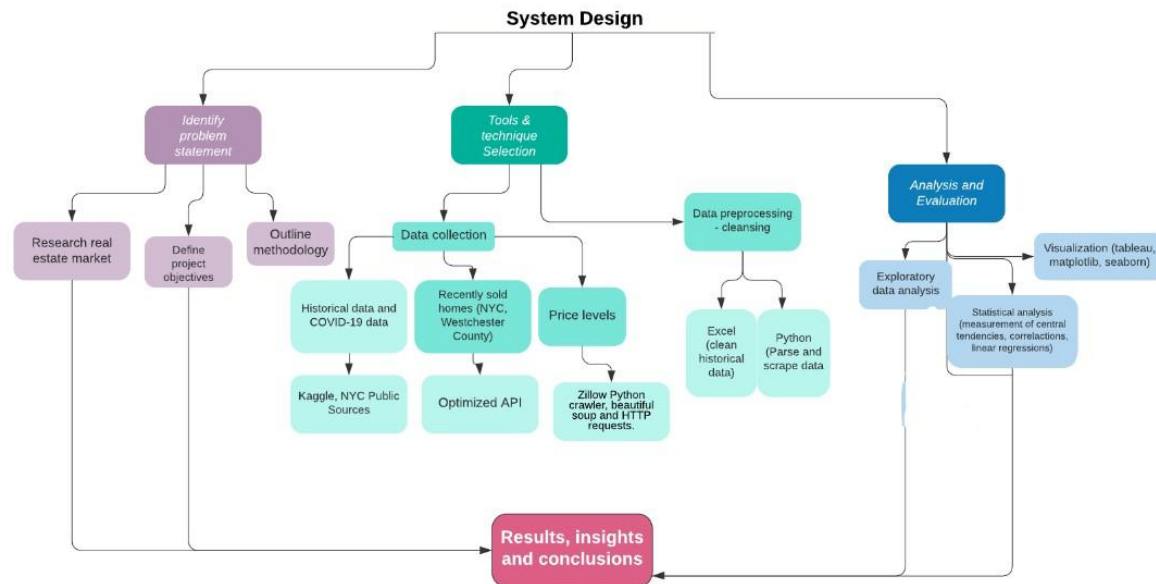
System Implementation

In the tools and techniques selection stage we divided it into two steps: data collection and data preprocessing. To collect the data we referred to different sources and tools. For historical data and data on COVID-19, we extracted datasets from Kaggle and also public data from the City and State of New York. To identify recently sold properties we use an API optimized to extract as many listings as possible from Zillow (500 for New York City and 540 for Westchester County) and to obtain the property price levels we designed a crawler that was implemented with the help of Python's BeautifulSoup and HTTP requests libraries.

Regarding the cleaning of the data, we use excel to clean the historical data and reduce the number of attributes to be considered, thus, we only consider residential properties, eliminating commercial properties, and only one, two, and three-family homes, condos, walkup and elevator apartments were brought into the analysis. Afterward, we used Python to parse and scrape data and so that the data would remain in the same format and with the same columns to facilitate analysis.

```
In [163]:  ▶  col_names = ["borough","neighborhood","building_class", "address", "zip",
                           "residential_units","total_units","squareft","year_built", "price", "sale_date"]

               data2021.columns = col_names
               data2021.head()
```

Out[163]:

	borough	neighborhood	building_class	address	zip	residential_units	total_units	squareft	year_built	price	sale_date
0	BRONX	BATHGATE	01 ONE FAMILY DWELLINGS	410 EAST 179TH STREET	10457	1	1	2048	1901	600000	15/01/2021
1	BRONX	BATHGATE	01 ONE FAMILY DWELLINGS	2052 BATHGATE AVENUE	10457	1	1	1516	1910	580000	22/07/2021
2	BRONX	BATHGATE	01 ONE FAMILY DWELLINGS	2317 BASSFORD AVENUE	10458	1	1	1188	1901	251600	17/08/2021

We then created filters to divide the data frames into housing information from before the pandemic (prior to March 13th, 2020), "during" the pandemic as the entirety of the year 2020 and the crawled data was used as a reference point for the current market prices. By slicing our data by dates we were able to perform a statistical analysis of measurement of central tendencies according to different time periods. And even at this initial exploratory stage, we were immediately able to identify a property price spike by looking at the two date benchmarks.

```
In [174]:  ▶  # Average price of sold homes before the pandemic
              filter1 = data2020.sale_date < "2020-03-13"
              beforeCovid = data2020[filter1]
              round(beforeCovid.price.mean())

Out[174]:  938280
```

```
In [175]:  ▶  # Average price of sold homes during the pandemic
              filter2 = (data2020.sale_date > "2020-03-13") & (data2020.sale_date < "2020-12-31")
              duringCovid = data2020[filter2]
              round(duringCovid.price.mean())

Out[175]:  942500
```

We decided to analyze how prices had changed according to different factors and also to analyze how certain places in New York had been affected differently. At this stage we refer to the Python packages seaborn and matplotlib to visualize the comparisons and facilitate the process of price change analysis as well as the identification of trends over time.

First, we did a comparison between the average property prices before the pandemic and grouped it by borough. When grouped together we could observe slight increases in the average, especially in Manhattan and Staten Island.

```python
# Average price of sold homes before the pandemic by borough
beforeCovid.groupby("borough").price.mean().sort_values(ascending=False)
```

```
borough
MANHATTAN        1704389
BROOKLYN         1070698
QUEENS            761315
BRONX             611801
STATEN ISLAND     588222
Name: price, dtype: float64
```

```python
# Average price of sold homes during the pandemic by borough
duringCovid.groupby("borough").price.mean().sort_values(ascending=False)
```

```
borough
MANHATTAN        1783595
BROOKLYN         1079425
QUEENS            781091
BRONX             618277
STATEN ISLAND     600559
Name: price, dtype: float64
```

However, a bigger difference can be identified when looking at the average price of sold homes before the pandemic compared to the average of recently sold homes that we crawled from Zillow.
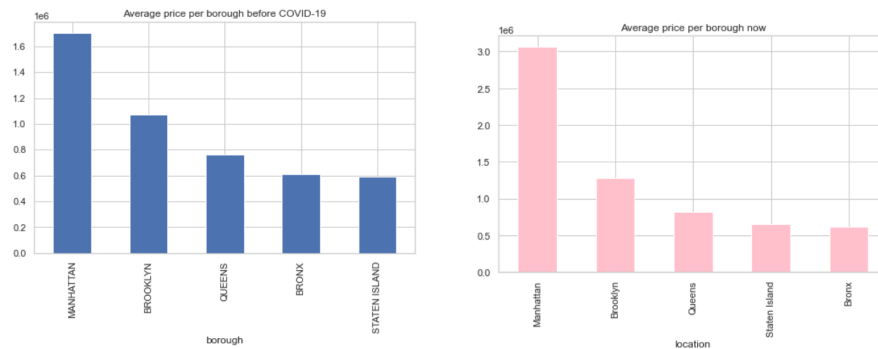
Furthermore, we performed the same analysis except it was grouped by neighborhood and extracted the top and bottom five cheapest and most expensive neighborhoods by looking at median property price. After analyzing this we found a reconfiguration of the top more expensive and cheapest neighborhoods of New York City.
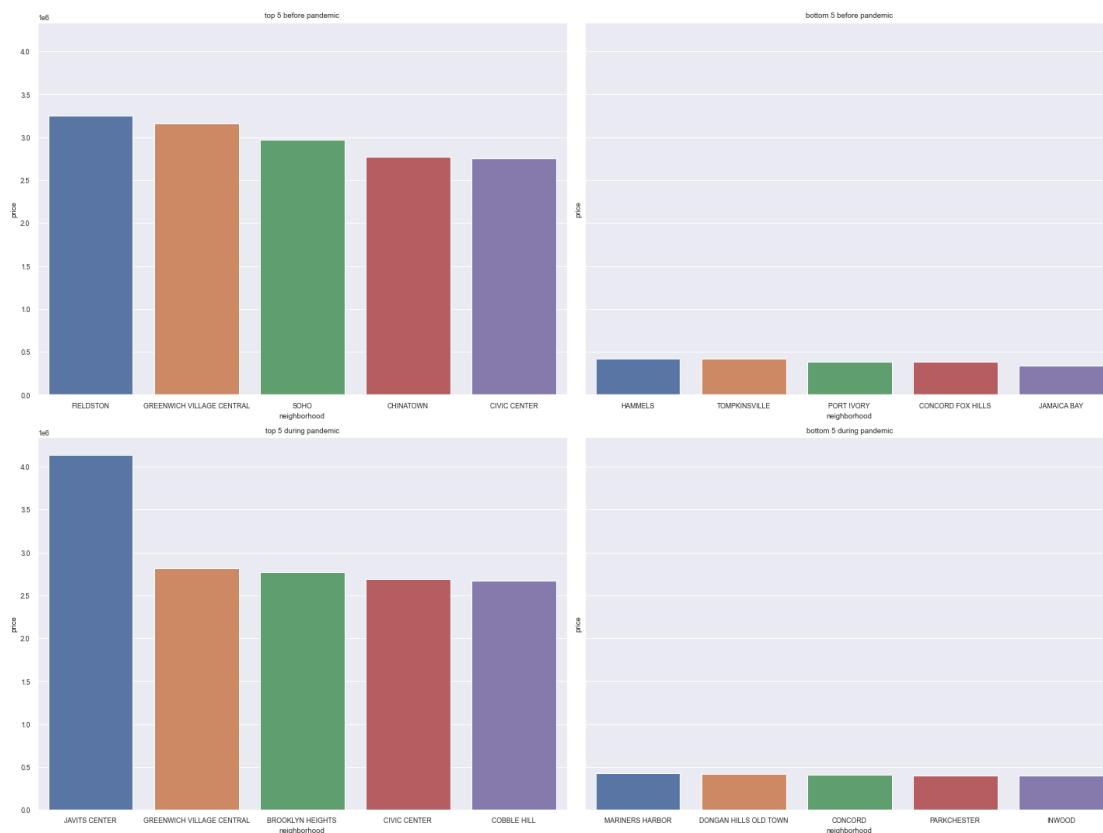
By double grouping the neighborhood and borough simultaneously, we were able to observe a general rise in the price of homes in Brooklyn, which went from having none to having two of its neighborhoods among the most expensive during the pandemic. On the other hand, Staten Island experienced an opposite effect, during the pandemic three of its neighborhoods came to occupy the first positions within the 5 cheapest neighborhoods.

Input:

```
fig, axes = plt.subplots(2,2, sharey=True)
sns.barplot(x='neighborhood', y='price',hue="borough", dodge=False, data=b_top5,  ax=axes[0,0]).set(title='top 5 before pande
sns.barplot(x='neighborhood', y='price',hue="borough", dodge=False, data=b_bottom5, ax=axes[0,1]).set(title='bottom 5 before
sns.barplot(x='neighborhood', y='price',hue="borough",dodge=False, data=d_top5,  ax=axes[1,0]).set(title='top 5 during pandem
sns.barplot(x='neighborhood', y='price',hue="borough",dodge=False, data=d_bottom5, ax=axes[1,1]).set(title='bottom 5 during p

sns.set(font_scale = 0.8)
fig.set_size_inches(20, 15)
fig.tight_layout()
plt.show()
```

Finally, we compared the average price of homes according to the type of property and building, before the pandemic and after the pandemic, and at the same time we grouped it by the borough to facilitate its visualization.
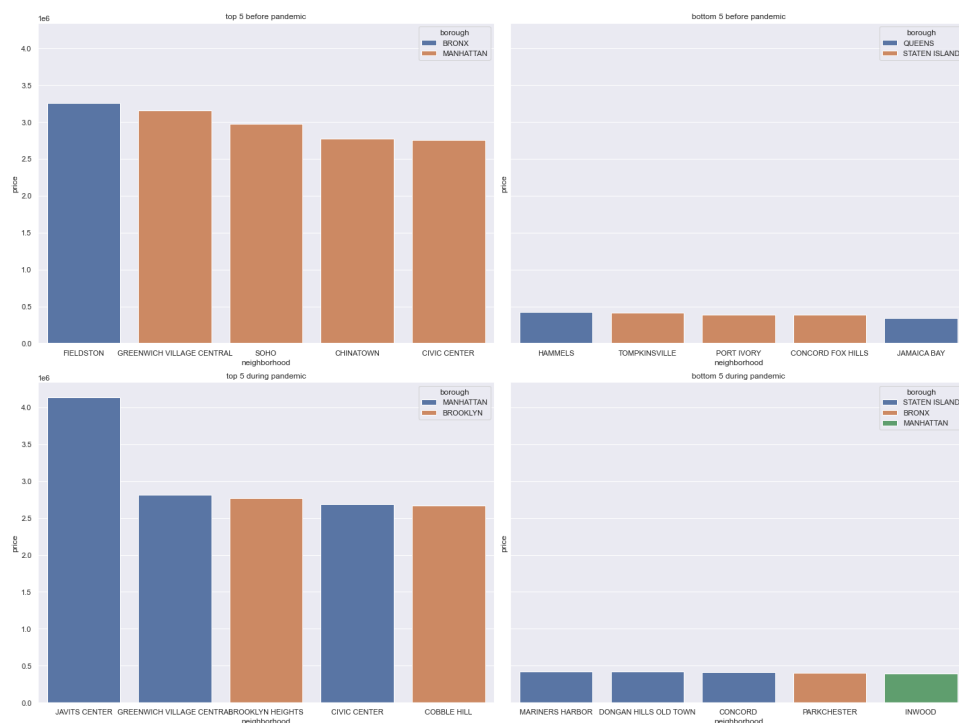
Manhattan experienced the most aggressive changes by property type. Here we can see that the

price of three-family homes and walkup apartments in Manhattan rose dramatically during the pandemic. Prices in the Bronx and Brooklyn increased proportionally, not much different depending on the type of building. Queens experienced a significant increase in rentals - walk-up apartments And Staten Island saw an increase mostly in one family, two-family, and three-family homes.
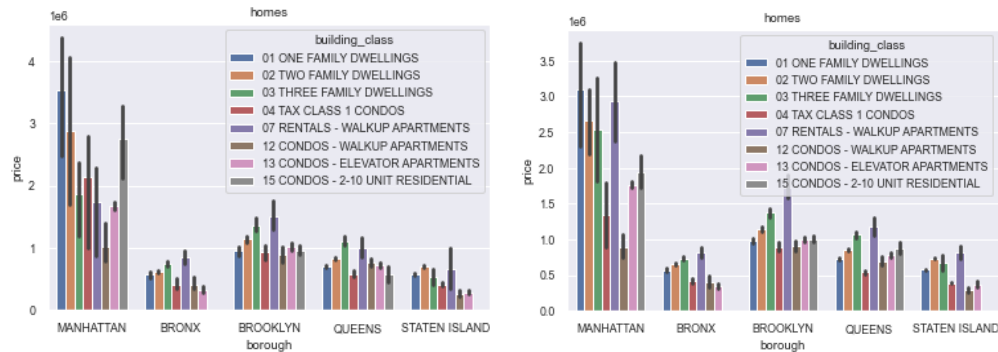
Results

Our results were divided into three parts, the first part of the results was the housing market from New York City, the second part was the housing market from Westchester County, and the third part was COVID-19 case counts from New York State. The data was viewed on a timeline starting from 2019 to 2021. The results were primarily focused on residential property sales for New York City and Westchester County.

Our results started with viewing the number of residential sales for New York City.
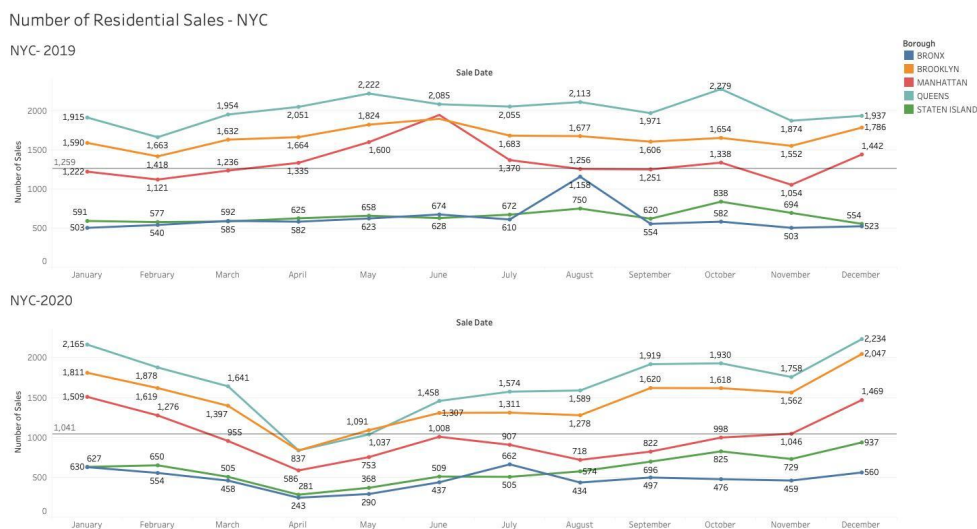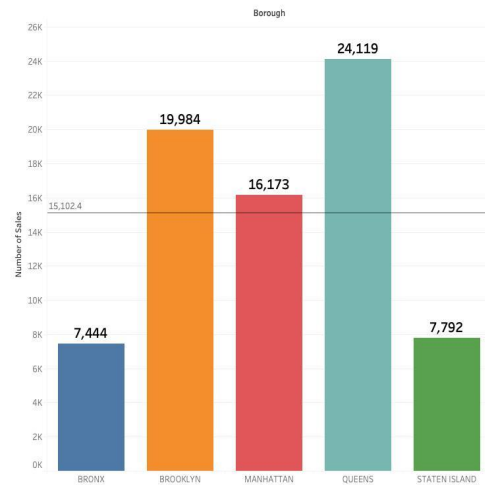

Figure 1.

Number of Residential Sales - NYC
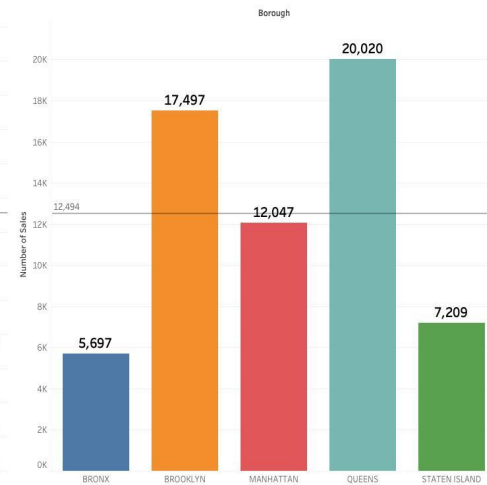
NYC-2019

NYC-2020

Figure 2.

Figure 1, indicates how the number of sales for homes dropped drastically specifically in the months from March to May in 2020. The average number of sales for homes sold in New York City dropped by 17.3% in 2020 over the 12 months. Manhattan experienced the greatest drop in residential sales more than any other borough in New York City, as seen in Figure 2. The next results indicated the condition of New York City for 2021. This data shows how NYC reacted to the pandemic of 2020. Below is Figure 3, which shows the number of sales for 2021.

Number of Residential Sales - NYC

2021 Sales Per Borough
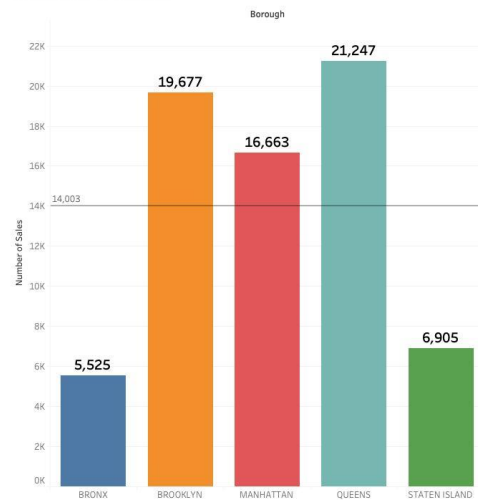
2021 Number of Sales - Residential

Figure 3.

In 2021, it was clear that New York City had begun to recover from the pandemic the year before. The overall residential market of New York City bounced back by an average of 12.1% in sales for all five boroughs combined. We also saw that in 2021 the number of sales started to rise and began to match the number of sales from 2019.

The second part of the data was from Westchester County. This particular data gave us a comparison of what is going on outside of New York City in the residential market. Below we see in Figure 4, how the residential market acted in recent months for Westchester County.
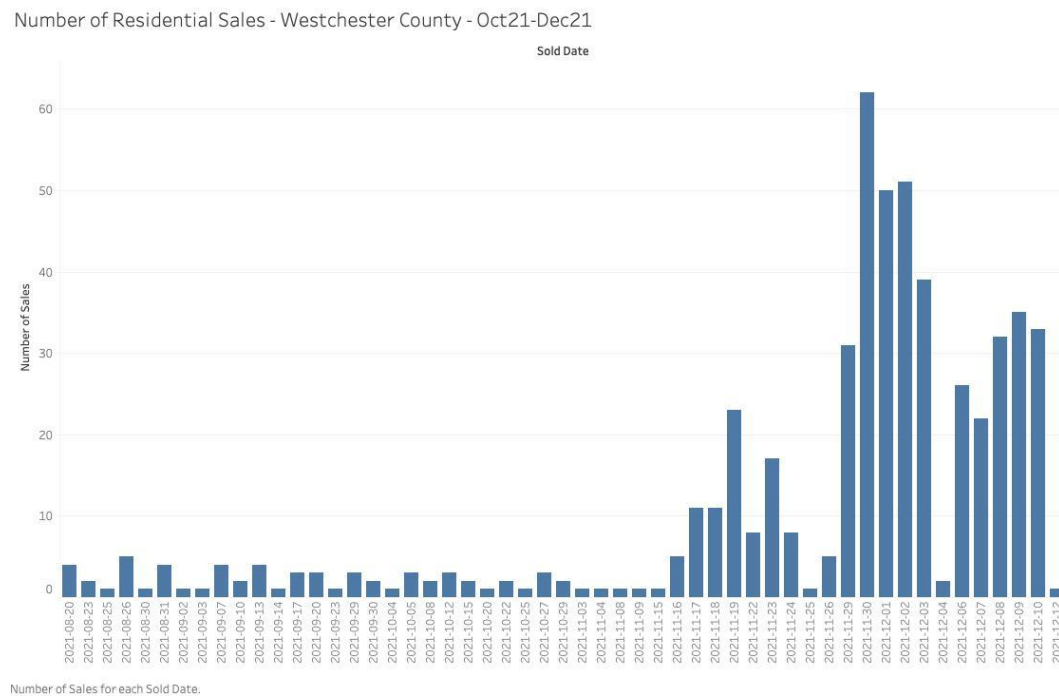

Number of Residential Sales - Westchester County - Oct21-Dec21

Figure 4.

We started to see the number of sales increase and match the sales from before the pandemic. The data indicated that once again by November the number of sales began to increase. The demand in the residential market in Westchester County was quickly rising and prices continued to rise as well.

The third part of the data was the COVID-19 case count for New York State. The data was focused on two areas of New York State: New York City and Westchester County. Below in Figure 5, one may see how the case count changed over time for Westchester County, and for New York City in Figure 6 for 2020. It is clear from Figures 5 and 6, that a spike in cases occurred during the months of March through May. Case counts started to level out by the

summer, and then once again, we saw the case count starting to rise in October and continuing to increase by December.
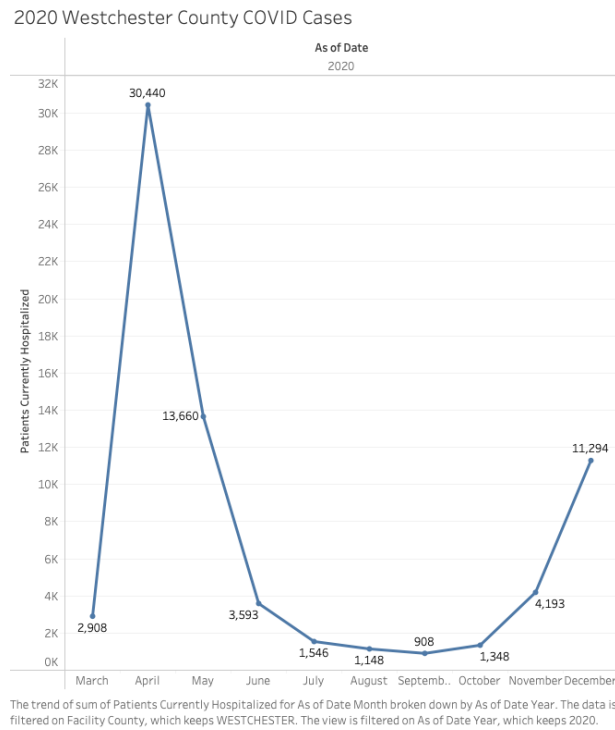

2020 Westchester County COVID Cases

The trend of sum of Patients Currently Hospitalized for As of Date Month broken down by As of Date Year. The data is filtered on Facility County, which keeps WESTCHESTER. The view is filtered on As of Date Year, which keeps 2020.

Figure 5.

2020 Covid Case Count Per Borough

The trends of Bk Case Count, Bx Case Count, Mn Case Count, Qn Case Count and Si Case Count for Date Of Interest Month. Color shows details about Bk Case Count, Bx Case Count, Mn Case Count, Qn Case Count and Si Case Count. The data is filtered on Date Of Interest Year, which keeps 2020.
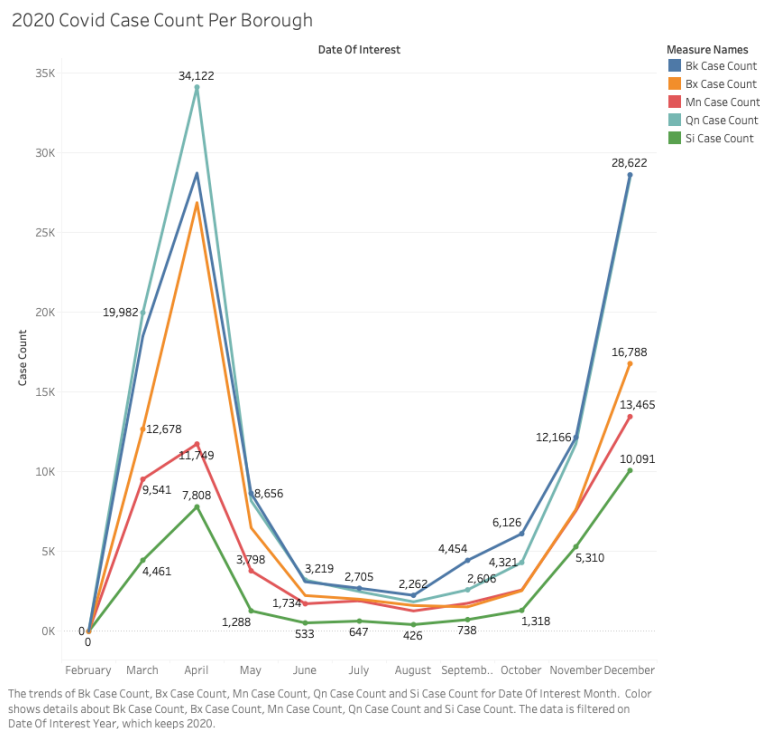
Figure 6.

Conclusion and Future Direction

During early 2020, a clear correlation was found between the first peak of COVID-19 cases and the number of homes sold throughout the state of New York. Between February through May of 2020, as the case count increased the number of sold homes decreased in New York State. The residential market was heavily affected by the pandemic.

As 2020 came to an end, the housing market started to increase in a number of sales, and very quickly demand for homes began to skyrocket again. Similar to New York City, Westchester County experienced a 12.5% increase in homes sold for quarter 3 of 2020 and continued to rise. This increase in demand is creating a peak for the residential market, and it relates differently than the first peak did when compared to the COVID-19 data. This rise in demand led to a rise in prices all across the state of New York. In New York City, many up-and-coming neighborhoods began to rise in property prices. Westchester County also experienced a rise in residential prices. The majority of residential sales were focused on single-family homes and caused our efforts to provide more insights on single-family homes.

Initially, home price changes and amount of homes sold could be directly related to the pandemic. But now that connection is no longer so evident. The market is no longer behaving according to the severity of COVID and after analyzing price trends and the behavior of recently sold homes across all boroughs and neighborhoods before and after the pandemic, we are inclined to believe that the current low levels of home sales may actually be a consequence of a lack of supply of homes and that this lack of supply is going to lead to a rise in the cost of residential real estate in New York.

The housing market experienced a quick and drastic demand in supply and rising prices. One method to help the residential housing market and future homeowners is to allocate our efforts and resources in building more affordable housing. Primarily focusing on building single-family homes followed by multi-family homes. Redirecting our resources to the residential market and reducing resources put into the commercial real estate market to ultimately provide housing for future generations.

References

Anenberg, E. (2021). Housing Market Tightness During COVID-19: Increased Demand or Reduced Supply?. Retrieved from https://www.federalreserve.gov/econres/notes/feds-notes/housing-market-tightness-during-covid-19-increased-demand-or-reduced-supply-20210708.htm

Campisi, N. (2021). New York After Covid-19: How Are Rent And Real Estate Prices Doing? Retrieved from https://www.forbes.com/advisor/mortgages/new-york-real-estate-after-covid/

Chung, S. (2021). COVID Takes a Toll on New York City's Real Estate Market - NY City Lens. Retrieved from https://nycitylens.com/covid-takes-toll-new-york-citys-real-estate-market/

Gascon, C. (2020). The Impact of COVID-19 on the Residential Real Estate Market. Retrieved from https://www.stlouisfed.org/publications/regional-economist/fourth-quarter-2020/impact-covid-residential-real-estate-market

Manhattan Apartment Sales Surge to Three-Decade High. (2021). Retrieved from https://www.nytimes.com/2021/10/05/realestate/new-york-real-estate-sales.html

Theis, E. (2021). NYC Median Sale Price Hits Yearly Low, Sales Drop Below 2,000 Once Again | PropertyShark. Retrieved from https://www.propertyshark.com/Real-Estate-Reports/nyc-real-estate-covid19/

Wade, T. (2021). Understanding the National Increase in House Prices. Retrieved from https://www.americanactionforum.org/insight/understanding-the-national-increase-in-house-prices/