New York City Real Estate

Anthony Rondos, Erick Sanmartin, Peter Morris, Nate Williamson

Dev10 Data Science and Engineering Capstone

October 7, 2022

I. INTRODUCTION

The New York City real estate market is especially competitive, with housing costs well above expected compared to other metropolitan areas. Individuals and families alike living in the city are finding it harder and harder to keep up with the rising rent and property prices. Many New Yorkers are slowly getting pushed out from the neighborhoods they call home, and the people of the city need action by lawmakers to ease these inflated housing costs and make living in the city more sustainable for the average New Yorker.

To make the necessary changes to New York's housing situation, lawmakers and government officials must have a clear understanding of New York's historic property costs, and how that compares to today. This project seeks to use New York rental and property cost information to perform informational analysis on the real estate market, and construct a predictive model that calculates property costs in the future based on 2021 to 2022 rolling property sales data. In order to provide utilizable information to lawmakers for the purpose of creating actionable policy to alleviate the skyrocketing housing costs in New York City, this capstone will answer the following questions:

1. How has the median cost of purchasing a home varied monthly in the last year?
2. Which neighborhoods' median rent increased the most for each borough in the last 12 years?
3. What are the most and least expensive neighborhoods for renting?
4. How long until renting costs outweighs owning costs?
5. How do sales in the past 12 months compare with one another by both zip code and borough?
6. How does employee pay compare to sales price in different zip codes of New York City?
7. How do median rent prices for each borough during the pandemic compare to the present?
8. Do property details have a relationship to changing prices in specific neighborhoods?
9. How has public policy influenced real estate prices?
10. Can we predict property prices based on property features and location?

II. DATA SOURCES AND ETL

We used a total of nine datasets in this capstone project with either information about housing costs or economic information by region. We extracted five datasets from the New York City Department of Finance's open data. These datasets were in the form of comma-separated value text files held rolling property sales information in New York City in the past year, with each dataset pertaining to a different borough. By creating a dataframe for each borough via url, we concatenated all five boroughs into one large New York City Rolling Property Sales dataset. We called two datasets from StreetEasy, a website that provides property sale and rental information for New York City, via API. These two datasets respectively contained property sales information and rental information by neighborhood. We took the final two datasets from the U.S Census website via API to provide economic background and context to the other real estate information extracted. We loaded the NYC Rolling Property Sales into a databrick and used Kafka to automatically stream the data into a consumer and producer. We created another static databrick for the StreetEasy and Census data, and then inputted both databricks into the same normalized SQL database. One can visualize the process as follows:
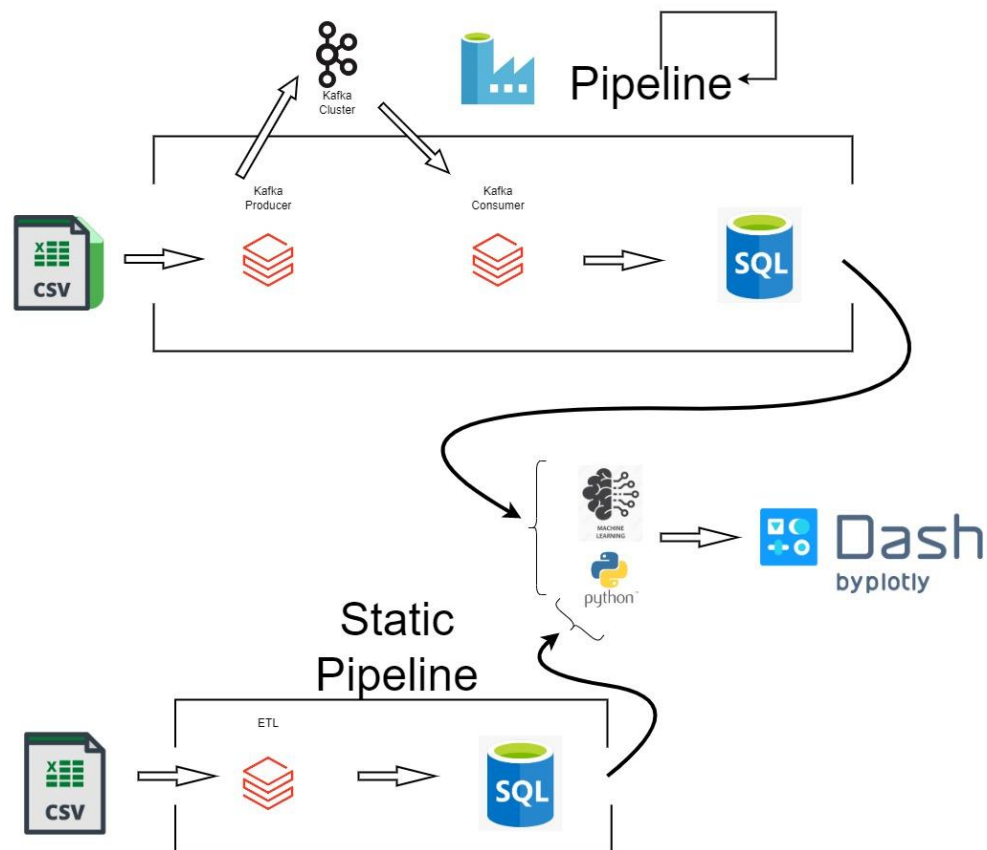
III. VISUALIZATIONS

Our investigation began with a look at pricing trends on the broadest scale. A look at the Street Easy data of median prices since 2010 paints a clear picture: prices are rising. Condo sale prices exceed other types of residential properties even when outlier luxury condo sales exceeding $100 million were removed. Although one can discern no causal link from the graph, it is worth noting that important New York City government policies, or the looming possibility of their passing in the near future, coincides with changes in observed average median prices. In March of 2016, the City Council approved the Zoning for Quality and Affordability Initiative, an amendment that made constructing larger affordable housing units easier for developers. Around the same time in NYC, average median condo prices peaked to historic highs before trending back down. The passing of the Housing Stability and Tenant Protection Act in 2019, which afforded tenants more rights while aiming to make rents more affordable, came after a spike in average price across all residential property types.
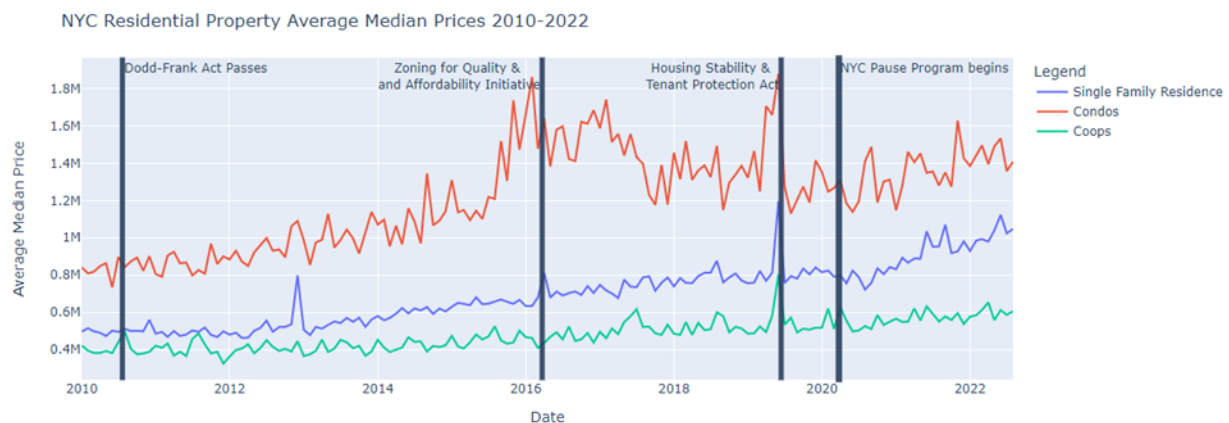

The NYC Pause Program marks the beginning of policy shifts in response to the COVID-19 pandemic. While the graph above gives a sense that prices began to increase sharply at that time, we needed to investigate this time period further, occasionally breaking prices down geographically and across longer time periods, to fully understand the trend we observed.

Median Sale Price by Month
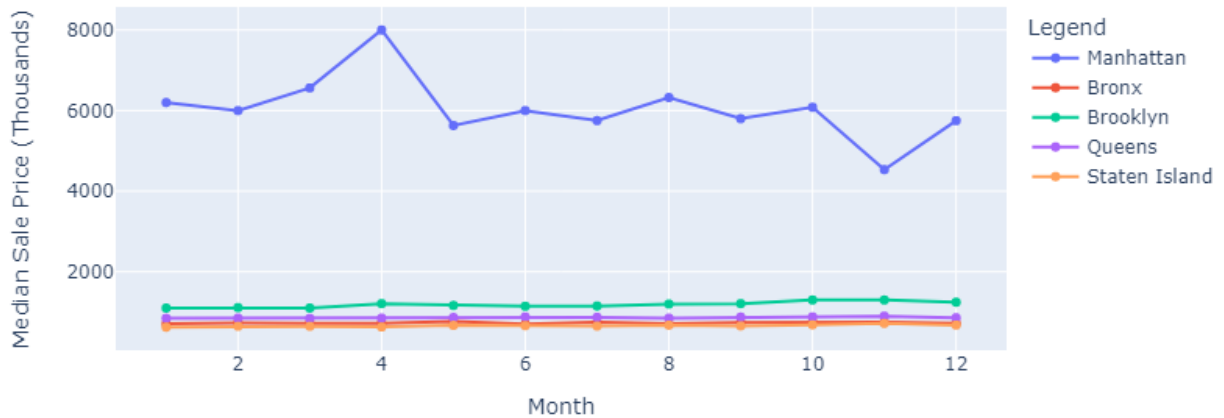
This graph measures the change in median sale price of homes sold in the past twelve months in the five boroughs of New York. Four of the five boroughs have similar sale prices while Manhattan is a clear outlier that makes the trends in the other boroughs difficult to see. To get a clearer view, we can remove Manhattan to create the following graph:

Median Sale Price by Month
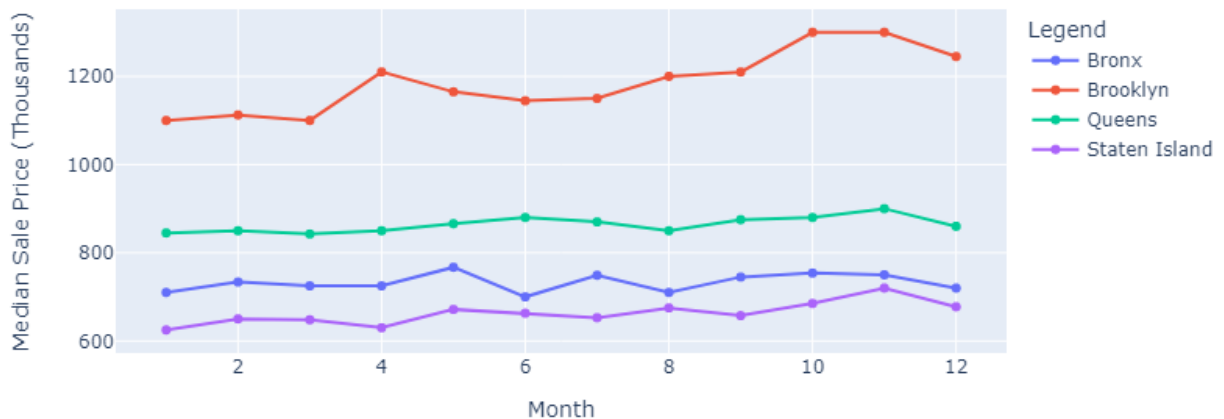
Home prices have been fairly constant, but generally are more expensive today than twelve months ago. This could possibly be because of high inflation rates, as well as people returning to the city and lower rates of remote work due to the COVID-19 pandemic dying down.

Top 3 and Bottom 3 Brooklyn Neighborhoods for Rate of Rent Increase 2010(11)-2022
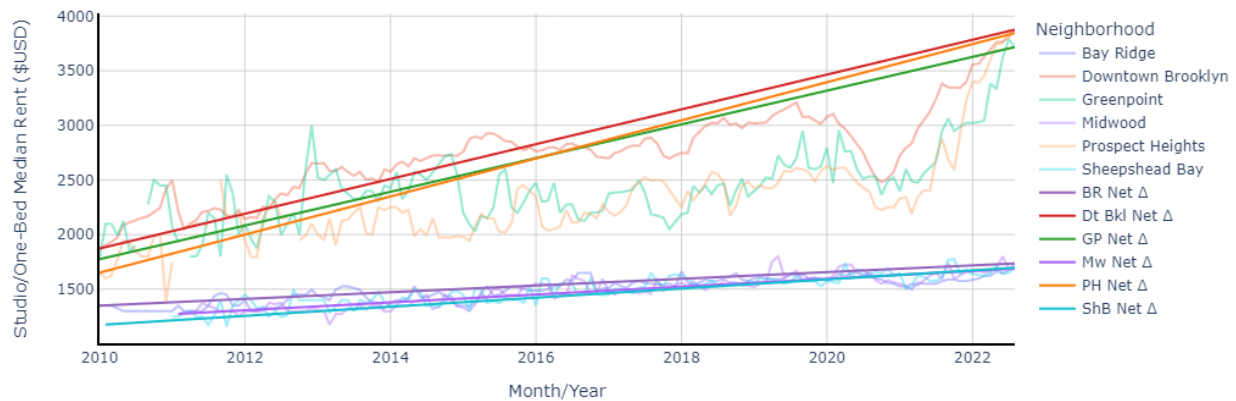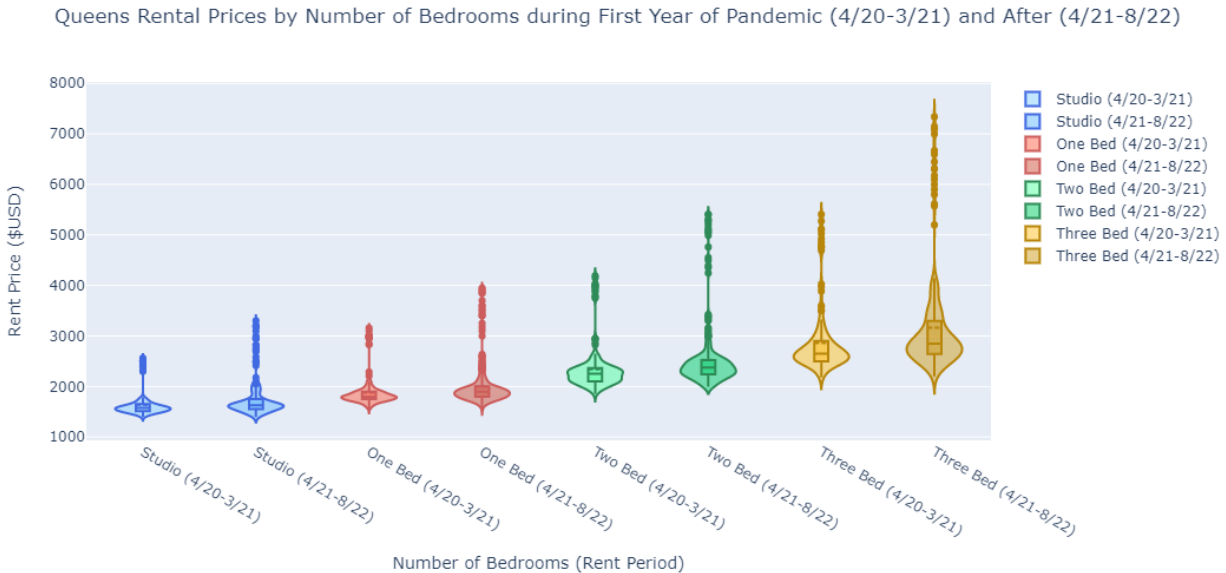
The rental market forces in Brooklyn, as with other boroughs, are spatially determined in that the closer the neighborhood is to Manhattan and north of Prospect Gardens (more or less at the geographic center of Brooklyn), the greater the increase in rent reported from 2010 (or 2012) to 2022. Here, the slope of the bolded line is what matters most. If the data were missing for the 2010 season, we included the latest date that was present as long as that date was before 2016 and there was not much discontinuity. As of 2022, the difference between the largest increase and smallest increase in asking-for price since 2010 crept above $1,600 between Downtown Brooklyn and Bay Ridge, respectively. Also of note, there is no crossover of neighborhoods that were more expensive in 2010 with other less expensive neighborhoods over the years, meaning the former are becoming concentrators of wealth, possibly pushing away people who may have lived there to the outskirts.

2022 Most and Least Expensive Neighborhoods by Borough in terms of Studio/One-Bedroom Rent
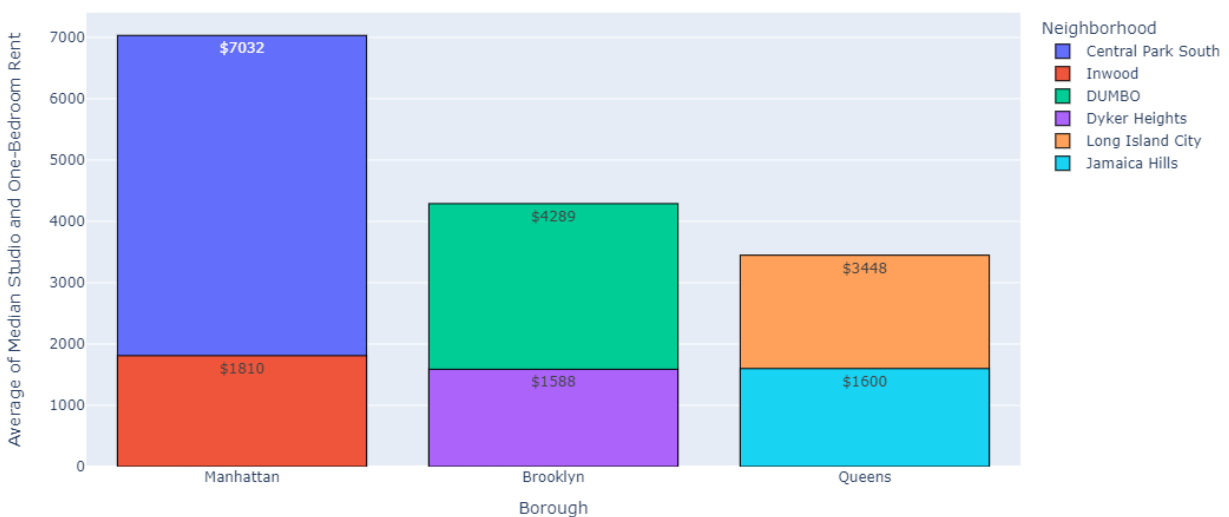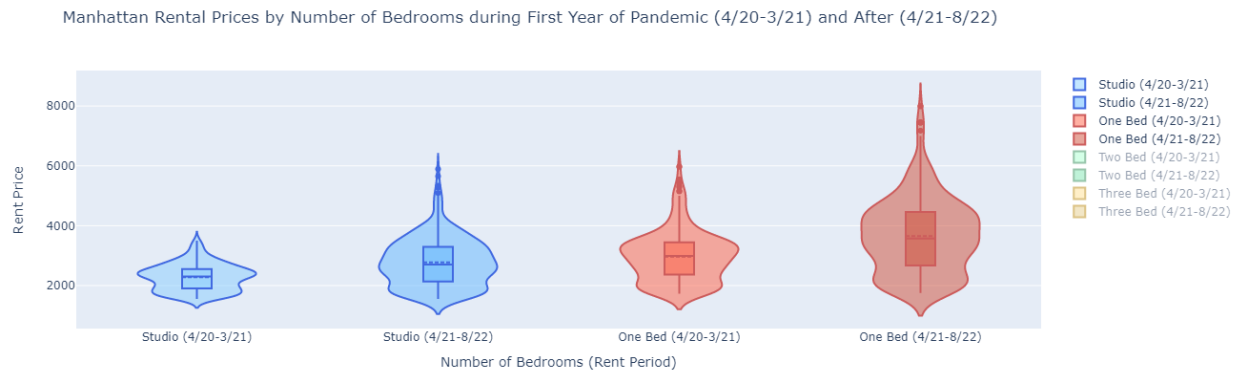
Another point of comparison worth drawing is the current state of price inequality across neighborhoods within the same borough. Unfortunately, the Bronx and Staten Island lacked sufficient data for this aim. The ratio of the most to least expensive neighborhood in Manhattan this year for a studio/one-bed apartment was 3.88 between Central Park South and Inwood, far higher than that of Brooklyn at 2.70 and Queens at 2.16.

Queens Rental Prices by Number of Bedrooms during First Year of Pandemic (4/20-3/21) and After (4/21-8/22)

In all neighborhoods, one perceives a price hike in Queens monthly rent for each bedroom count between the average of April 2020 to March 2021 and the average of April 2021 to August 2022. While the global COVID-19 pandemic began in March 2020, we decided to investigate prices in the month following and up to a year post in representing its potential consequences on the real estate market. As shown previously in the rate of rent increase line plot, rents reach a local minimum in the top three Brooklyn neighborhoods and rise rapidly thereafter, signaling a new, if unwelcome, normal for tenants. We performed a Games-Howell post-hoc test (Pingouin library) to characterize these pairwise differences in Queens and other boroughs. This is similar to a post-ANOVA Tukey's HSD but more appropriate for unequal variances and sample sizes and nonparametric nature of the data.

	A	B	pval
0	One Bed (4/20-3/21)	One Bed (4/21-8/22)	0.000792
1	Studio (4/20-3/21)	Studio (4/21-8/22)	0.006859
2	Three Bed (4/20-3/21)	Three Bed (4/21-8/22)	0.002645
3	Two Bed (4/20-3/21)	Two Bed (4/21-8/22)	0.001513

Manhattan Rental Prices by Number of Bedrooms during First Year of Pandemic (4/20-3/21) and After (4/21-8/22)


Nowhere is this year-over-year jump in aggregate prices clearer than in the studio and one-bedroom apartments of Manhattan. Both the average and the median of the price data have trended upward since April 2021, with more outliers accounted for beyond the upper quartile. One might hypothesize that landlords were motivated to recoup the revenue they lost during the first year of the pandemic, pricing people out and making it a seller's market. The results of the same test as above for Manhattan indicate a stronger statistical difference than Queens between pre- and post-pandemic rent prices.

|	A	B	pval
0	One Bed (4/20-3/21)	One Bed (4/21-8/22)	0.000000e+00
1	Studio (4/20-3/21)	Studio (4/21-8/22)	0.000000e+00
2	Three Bed (4/20-3/21)	Three Bed (4/21-8/22)	5.029240e-03
3	Two Bed (4/20-3/21)	Two Bed (4/21-8/22)	3.710598e-08

Looking further into the question of renting versus buying, we calculated how long it took for renting costs to overtake buying costs based on year of purchase, and created shaded line charts for each borough.

How Long Untill Renting Cost Overtakes Buying Cost (Brooklyn)
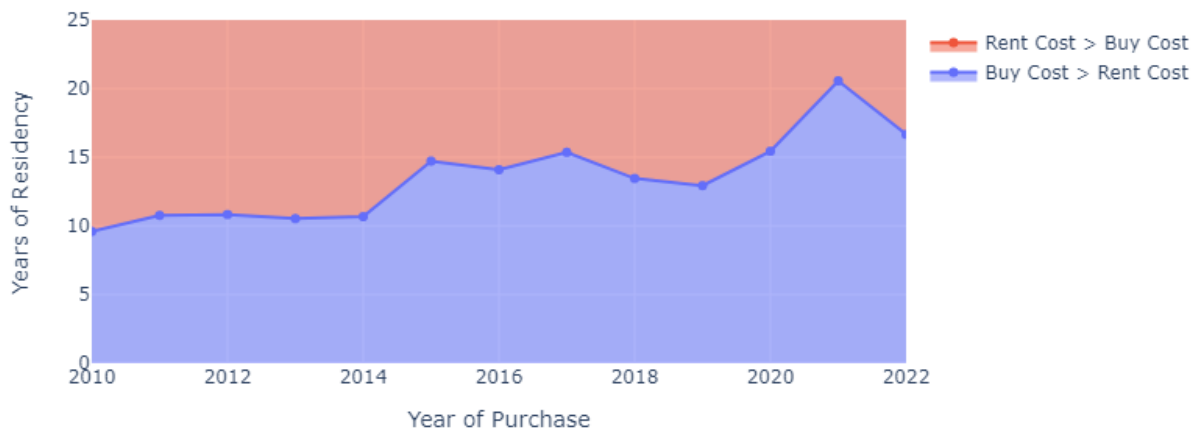
Above is the created line chart for Brooklyn. Most boroughs followed the same trend of buying becoming relatively more and more difficult over time with fluctuation either way based on current events. The spike in 2021 can likely be attributed to COVID-19, where rent prices plummeted due to a sudden decrease in demand as people ventured away from urban areas to avoid contracting the disease. Housing prices may have taken a dip as well, but due to the nature of real estate investments, property ownership holds value well and remains relatively expensive compared to renting. The only borough that does not seem to follow the same trend as Brooklyn is Staten Island.

How Long Untill Renting Cost Overtakes Buying Cost (Staten Island)
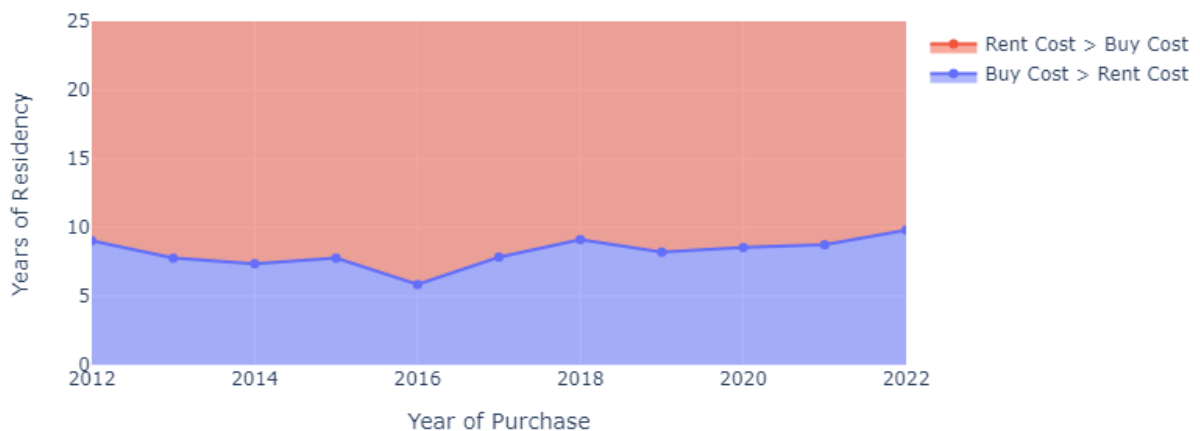
Staten Island is an especially suburban borough with many single family homes distinct from the large apartment complexes common across the city. Because of this unique landscape, owning a

house is often the norm and renting is overall less common with less options for residents interested in renting their home.

All five boroughs saw renting become relatively more financially sensible after 2019, when the housing stability and tenant protection act was passed. This law enacted limited security deposits, rent gouging, and evictions creating a much more hospitable housing market for renters in New York City.

IV. DASHBOARD

We utilized the graphing library Plotly to create detailed and palatable visualizations from the data loaded into the SQL database to aid policy makers in making informed data driven decisions pertaining to the New York City real estate market. We then produced a plotly dashboard to provide users an easy to use interface that will allow for easy navigation to render varying graphs and interactive models about the city's housing market. We separated information by labeled pages, each of which we further divided by more specific criteria such as borough or neighborhood.

The Property Sale Trends page provides a general overhead view of the New York housing market as a whole, using line graphs to detail how median home price has changed for each borough over the last twelve months and how public policy may have impacted median home price for various types of housing over the last decade. The Maps page contains two maps, the first of which details the median home price by zip code, and the second the average employee pay by zip code. The first map allows the user to input their borough of interest to provide more precise and relevant information.

The Drilling Down page contains a wide array of close-up perspectives of property rental and purchasing cost trends in New York City. Rental graphics within the page include a line chart of the three neighborhoods where rent increased the most by borough and a filled line chart depicting when the time spent renting overtakes the cost of buying a home by year of purchase. Purchase cost graphics within the Drilling Down page include a layered bar chart that depicts the most and least expensive neighborhood for various housing types for Manhattan, Brooklyn, and Queens, as well as a violin chart that compares change in median home price for various types of housing for each of the five boroughs.

Lastly, the Machine Learning page contains information pertaining to a machine learning model constructed from the New York Rolling Property Sales dataset and predicts real estate prices based on property and locational information. Included here is a feature importance bar graph that measures how influential property details such as square footage are for a property's total sale price. Furthermore, we made scatter plots measuring actual versus predicted sale price and residuals for our Random Forest model. The Machine Learning page offers users the ability to perform a live sale price prediction by inputting certain property information and outputs a sale-price forecast based on the given parameters.

V. MACHINE LEARNING

We made our model using property sales data from the past year (September 2021 - August 2022) in New York City. The data has numeric and categorical features describing a property's physical characteristics, location, property classifications and price. Our model predicts the price of a residential property based on these features.


The density plot for sale price is shown above for sales greater than $100 and less than $5,000,000. Most sales occurred between $250,000 and $2,000,000. The dataset has numeric features of gross square feet, land square feet, residential units, commercial units and total units. Many rows had outliers several standard deviations from the mean in numeric columns. We removed rows with values greater than two standard deviations from the mean, which reduced sale price standard deviation from ~17M to ~2.7M. After removing outliers and setting a minimum sale price of $100,000, the range of sale price was (100K - 36.3M).

Zip code, neighborhood, building class at sale and tax class at sale were included as categorical features. Building class category was filtered for residential categories for best results. Land square feet was removed as it had a correlation with sales price of less than 0.35. The Total units column was removed as it had a correlation with residential units of 0.99, and we determined it was a redundant feature.

We tested the cleaned data with popular regression algorithms. The two best performing baseline models were random forest regression and XGBoost regression. The random forest model was tuned with grid search, giving optimal parameters of (max_depth=100, max_features=14, n_estimators=300, random_state=42). This was the best performing model with a test score of 0.71. The most important features were gross square feet, residential units,

land square feet and commercial units. As seen in the actual vs. predicted values plot, the model tended to underprice properties over $10M.
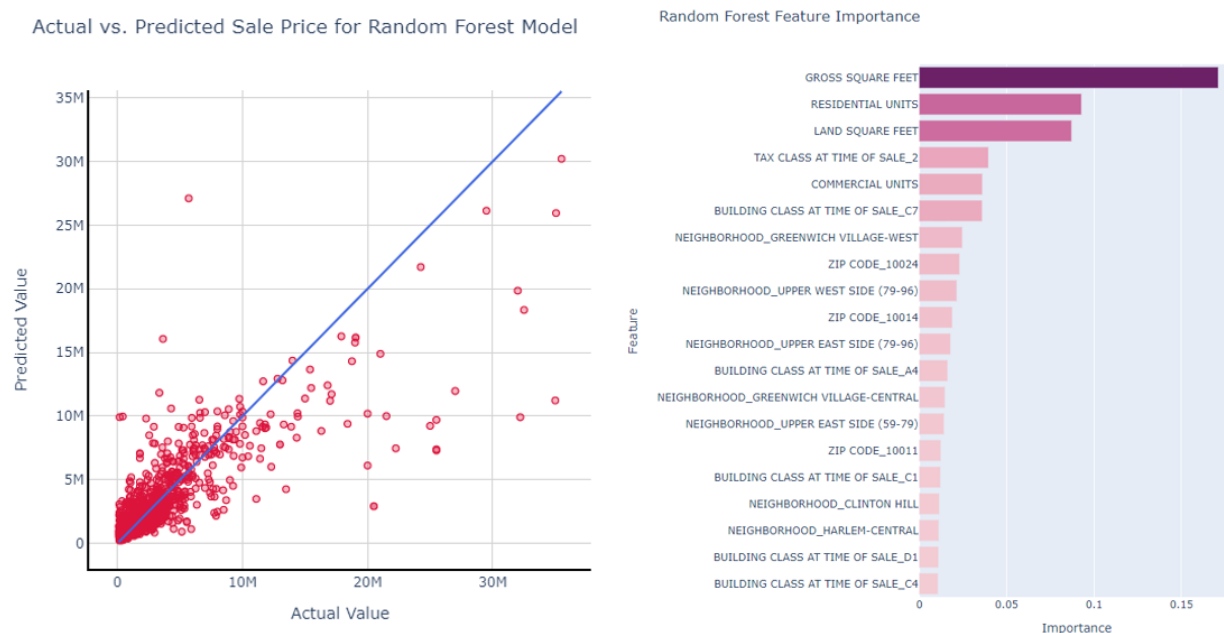
Random forest performs well with outliers and features that do not have strong linear relationships with the independent variable. The distribution of sale price in our data had a significant proportion of expensive properties that were outliers. Also, many features had a correlation with sale price of less than 0.5. Using an algorithm with decision trees gave us best results considering these limitations in our data.

VI. CONCLUSION

The goal of this investigation was to elucidate underlying market trends within the New York City real estate market. Equipped with this information, policy makers would be able to make informed and effective decisions to protect New Yorkers from unaffordable rents and property prices. Our approach for analysis ensured that recent circumstantial dips in prices were only a small part of the story, and that takeaways from the information we provided would be actionable in the future.

We observed a general increase in prices for property sales over the past twelve years, and an alarming increase in prices within the last few years. Rent cost similarly trended upwards, especially in Brooklyn. An analysis of intra-borough prices revealed that some neighborhoods in every borough continued to trend upwards despite recent dips during the COVID-19 pandemic. Indeed, places like Corona, Queens have seen a 20% increase in median rent prices for one-bedroom apartments since December 2019: median rent has reached $3500 in a neighborhood with a reported $45,000 average payroll salary according to the US Census. We highlighted the most and least costly neighborhoods in each borough for renting and buying

property to shed light on the gap in affordability depending on location. Our plots and analyses of changes in prices between the first year of the pandemic and this year evinced significant differences in Queens rent prices. Furthermore, the threshold for rent costs superseding the cost of purchasing a property increased by over a few years in Brooklyn indicating a trend of decreased affordability within the borough.

       The information we have gathered and our model for price predictions provide a strong direction for policy makers who want to target key areas exhibiting troubling pricing trends. Even neighborhoods that do not immediately show rising costs may reach unaffordable levels for the people who live in them. Although Manhattan retains its characteristic high prices, Brooklyn has clear signs of runaway costs as of this year, and Queens shows signs of following suit. To adequately address these potentialities, policy makers can look to old government edicts and policies, discern their effectiveness, and address the trends observed in neighborhoods across New York City.

VII. REFERENCES

New York City Department of Finance. (2022). Rolling Sales Data.
    https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page

Street Easy. (2010-2022). Sales in NYC, All Home Type, Median Sale Price.
    https://streeteasy.com/blog/data-dashboard/?agg=Median&metric=Recorded%20Price&type=Sales

Street Easy. (2010-2022). Rentals in NYC, All Bedroom Count, Median Asking Price.
    https://streeteasy.com/blog/data-dashboard/?agg=Median&metric=Asking%20Rent&type=Rentals

United States Census Bureau. (2017). Economy-Wide Key Statistics 2017.
    https://www.census.gov/data/developers/data-sets/economic-census/2017.html

United States Census Bureau. (2020). County Business Patterns: 2020.
    https://www.census.gov/data/datasets/2020/econ/cbp/2020-cbp.html

United States Census Bureau. (2020). ZIP Code Tabulation Areas.
    https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2020&layergroup=ZIP+Code+Tabulation+Areas

White Karp, Jennifer. "Hit with 30 Percent (or Higher) Rent Increases, Many NYC Tenants Are in 'Shock.'" Brick Underground, 2 Mar. 2022,
    https://www.brickunderground.com/rent/why-are-landlords-raising-rents-extreme-increase-lease-renewal-nyc.

Zaveri, Mihir. "Why It's So Hard to Find an Affordable Apartment in New York." The New York Times, 1 Aug. 2022,
    https://www.nytimes.com/2022/08/01/nyregion/nyc-affordable-apartment-rent.html.