

Data Analysis

Real Estate Market in NYC



Katz School of Science and Health
AIM 5001 Data Acquisition & Management

Radek Jan Holik
Sheng-Han Yueh
Chi-Kuang Chen

Dec 15, 2022



Table of Contents

Research Purposes and Motivations	- 3
Data Sources and Data Management Strategy	- 4
Special Findings in EDA	- 7
Explore the Relationships between the Following Factors and Price	- 8
• The Characteristics of the House	
• Household income	
• The influence of the pandemic	
• Safety (crime rate)	
Conclusions	- 12

Research Purposes and Motivations

As an International Student... we care about...

- How do the characteristics of the house affect the sale price?
- How likely is it that residents who live in zip codes with higher housing prices will earn more money?
- How has the Covid-19 pandemic affected the sale price or the rental price?
- Do safety or the crime rate have a relationship with housing prices?



Data Sources and Data Management Strategy

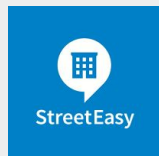
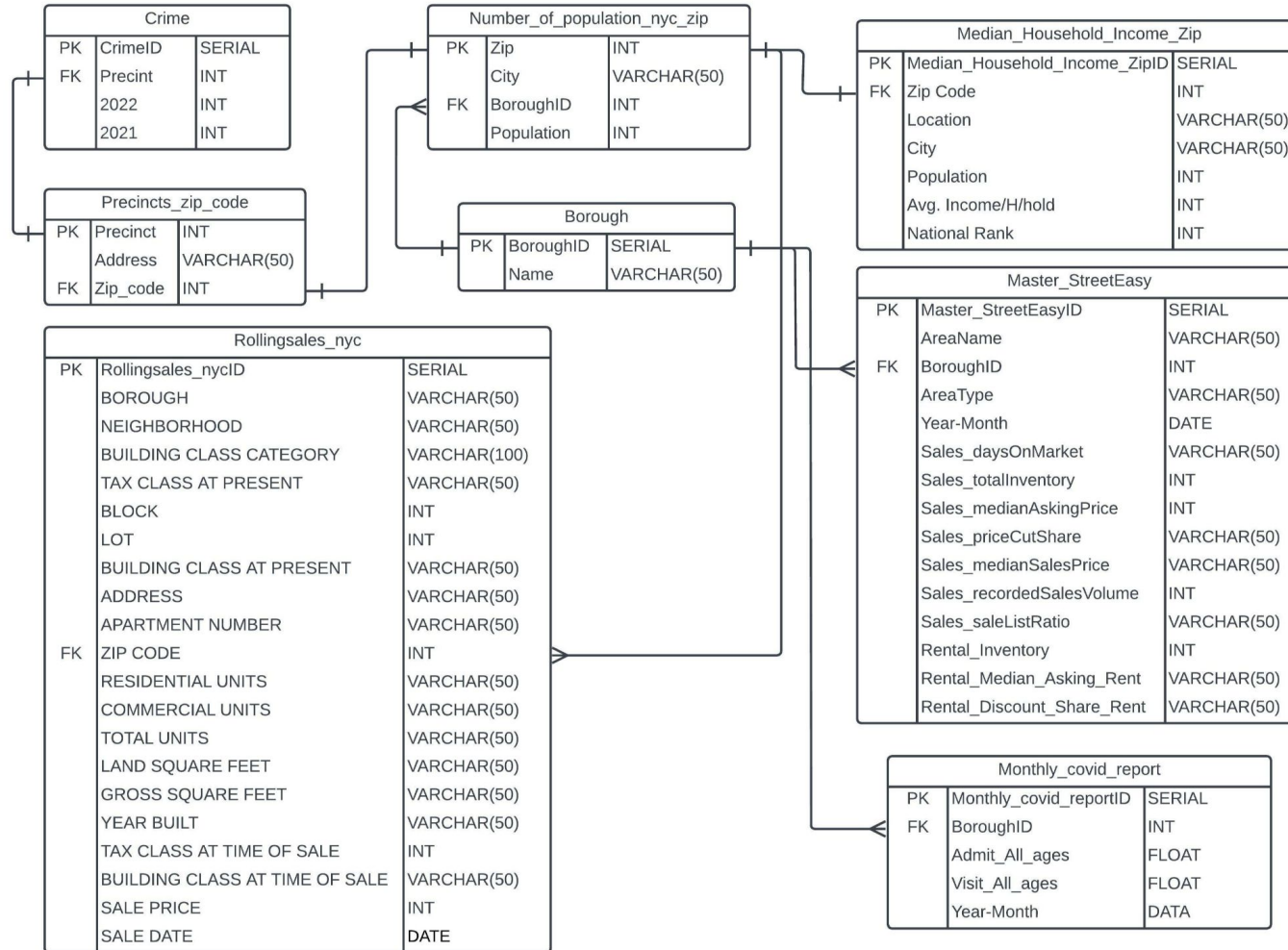
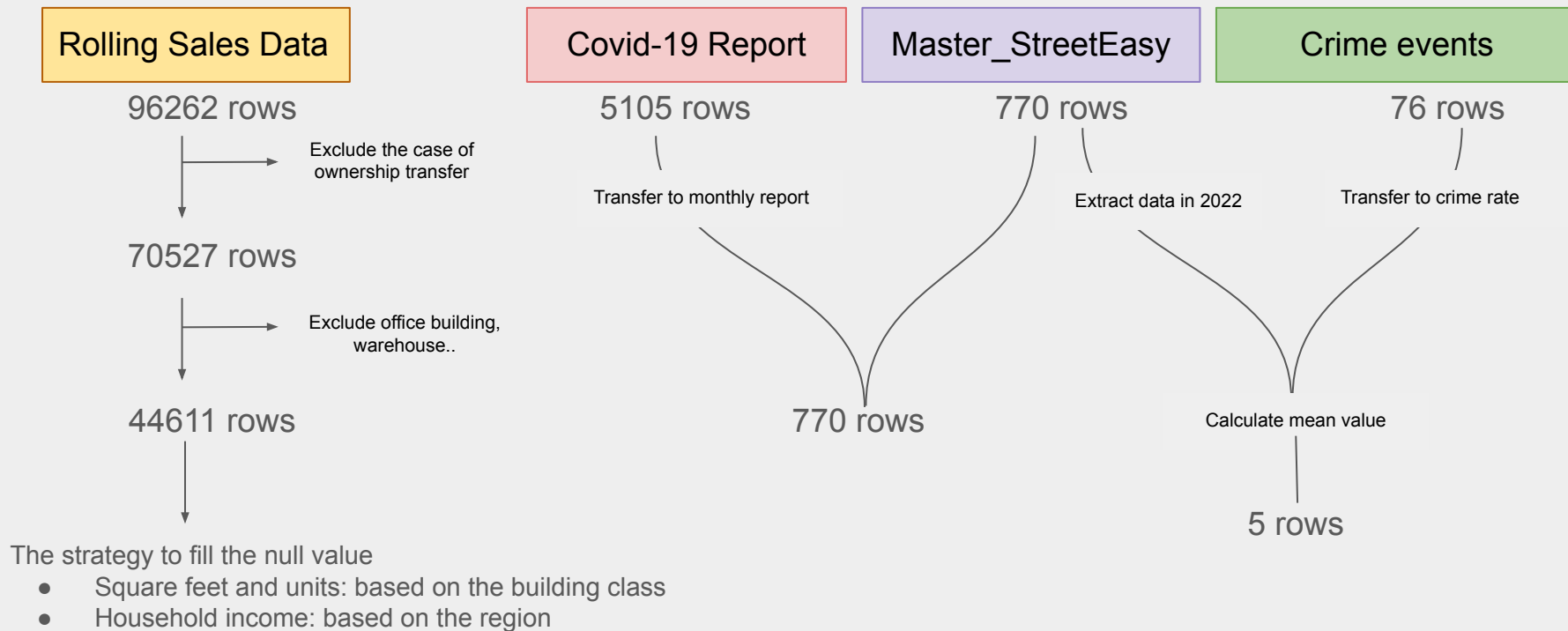


Table Name	Master_StreetEasy	Rolling Sales Data	Covid Daily Report	Crime Events / Precinct
Data Retrieval	CSV	CSV	CSV	Web Scraping
Data Type	Time Series Data (From 2010 - Present)	Cross Sectional Data (From Nov 21 - Oct 22)	Time Series Data (From Feb 20 - Present)	Real-time Data (in 2022)
Important Attributes	<ul style="list-style-type: none"> Borough Period (Year-Month) Sales: Day on Market, Total Inventory, Sale Price Rentals: Median Rent 	<ul style="list-style-type: none"> Physical Location Property Information: Square footage, Sale Price, Building Type, Tax Class, # of Units 	<ul style="list-style-type: none"> Borough Period (Year-Month-Day) Number of Infections per 100,000 people by age groups 	<ul style="list-style-type: none"> Borough / Precinct Crime events based on the Region



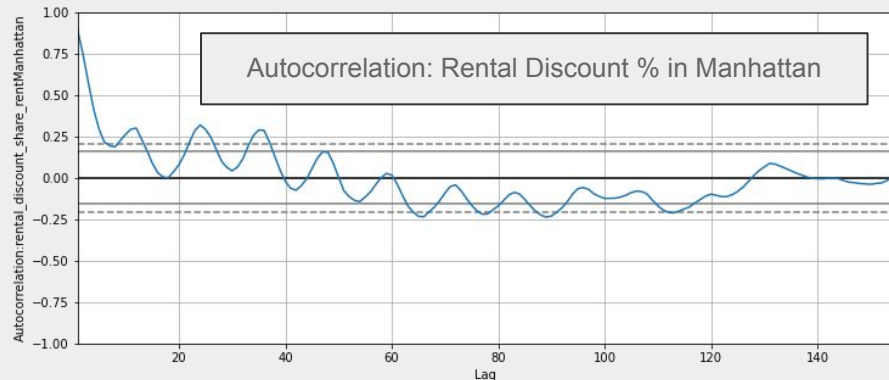
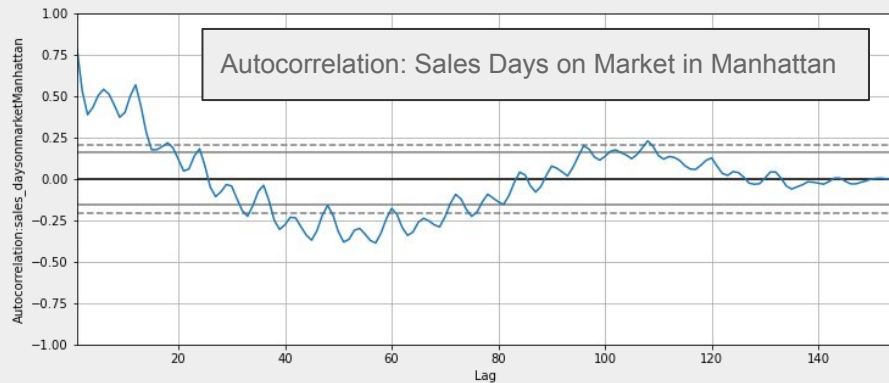
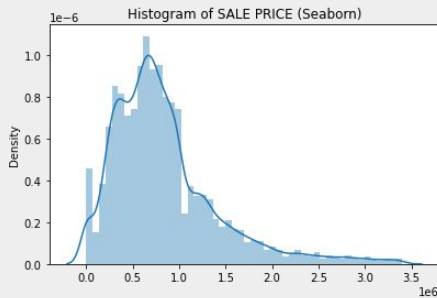
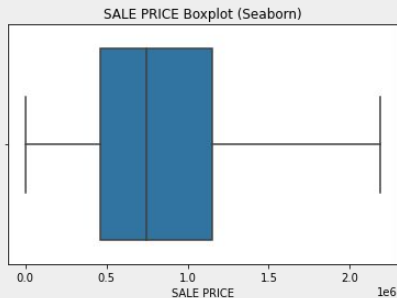
Data Preprocessing



Exploratory Data Analysis (EDA): Special Findings

Findings are counter-intuitive / interesting....

- Sale Price Distribution for Sale Properties
- Autocorrelation of StreetEasy Market Data
 - Sales Market vs. Rental Market
 - Business Cycle vs. Seasonality



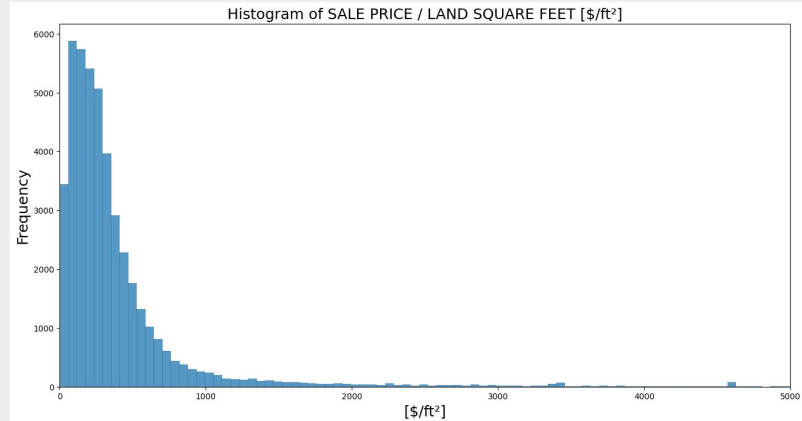
Influential Characteristics of Sale Price

1. Location

- Borough: Manhattan
- Zip Code 10069
- Neighborhood: Financial District

2. Property Size

Prices tend to increase as the square footage increases, the average price per square foot is \$445

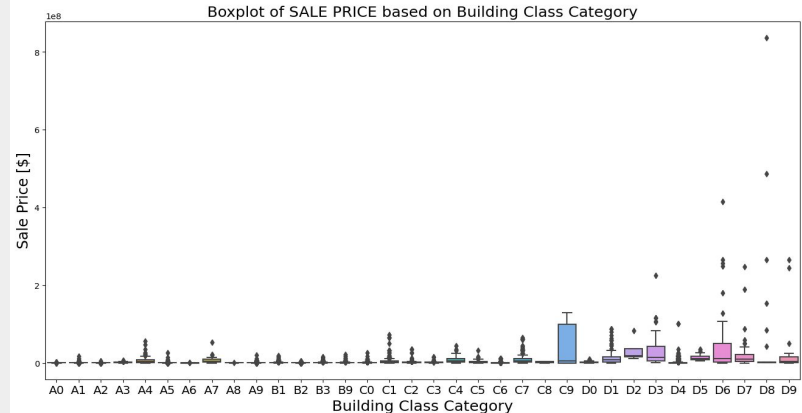


3. Age of the Property

Costliest houses are those below 20 years of age (new-built), and the least expensive houses are those between 50 and 80 years of age

4. Building Class Category

The price of a unit in a large complex is typically higher than in a smaller complex)



Region Comparison: Median Household Income vs. Property Sale Price

The neighborhood/zip code/borough with the highest median income per household

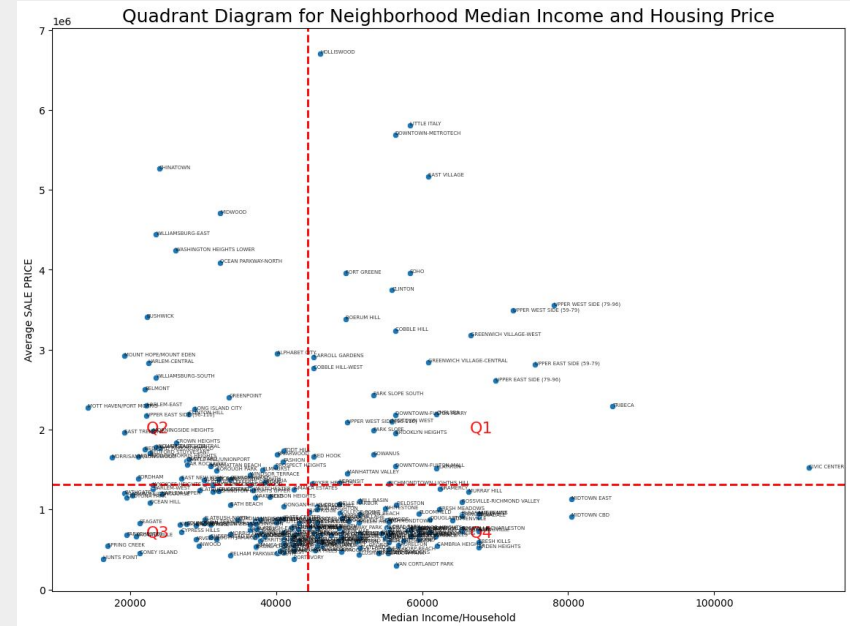
- Neighborhood: Civic Center with \$113,000
- Zip Code: 10007 with \$113,000 (overlapping with Civic Center)
- Borough: Manhattan with \$66,600

The neighborhood/zip code/borough with the lower tier median income per household

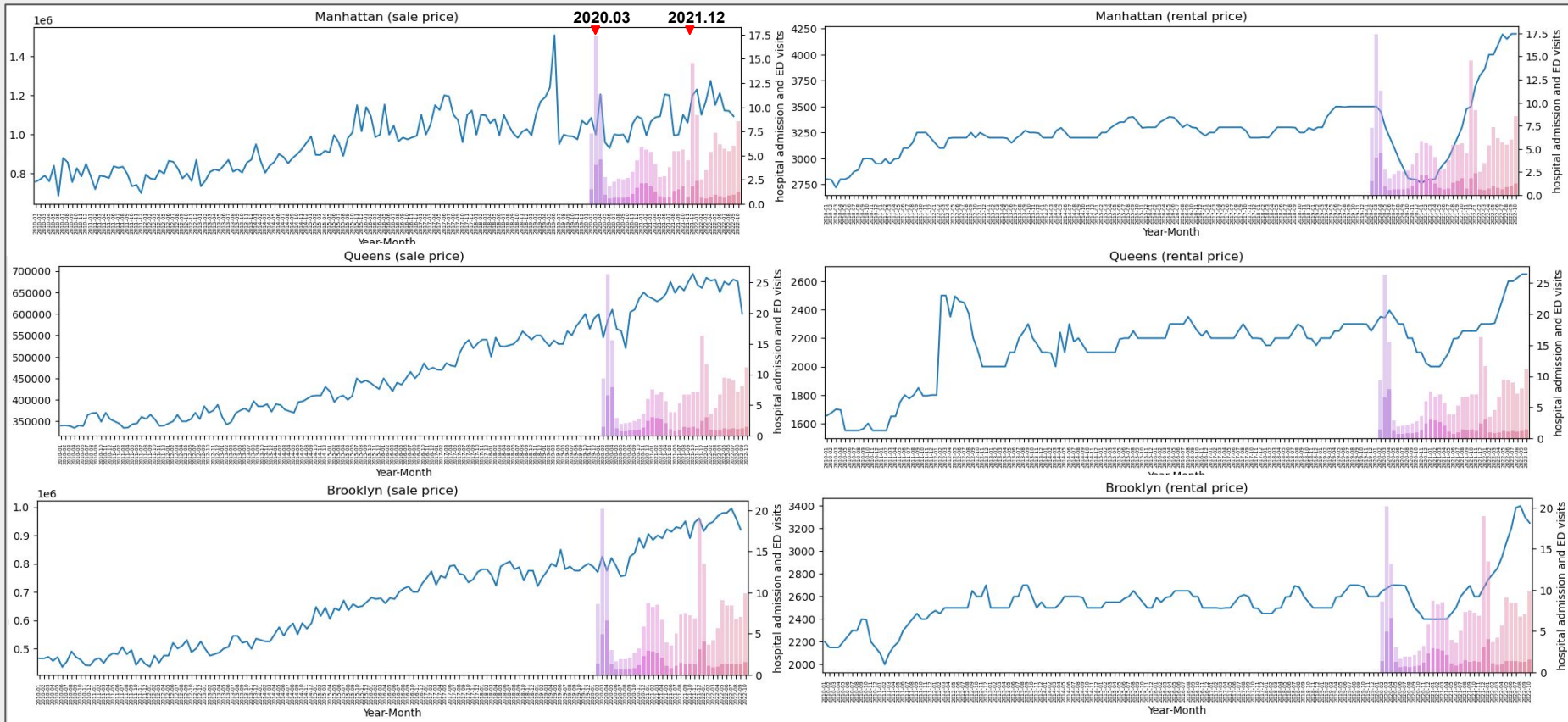
- Neighborhood: Mott Haven/Port Morris with \$14,300
- Zip Code: 10454 with \$14,300 (overlapping with Mott Haven/Port Morris)
- Borough: Brooklyn with \$33,765

The relationship between median income and property price

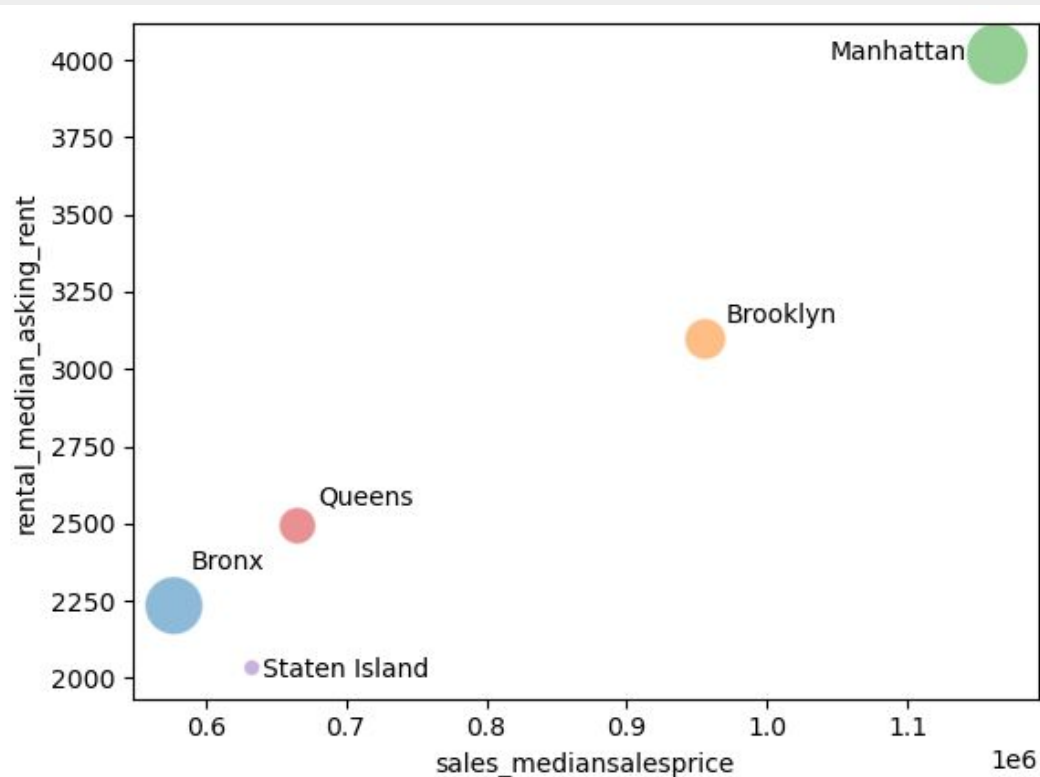
- There is no general connection between higher housing prices and higher income in New York City.
- There are neighborhoods where incomes are not very high, yet real estate prices are higher. (Holliswood, East Williamsburg, Midwood)
- On the other hand, there are exceptions in neighborhoods where a certain connection between incomes and real estate prices can be observed. (Upper West Side, Upper East Side, Greenwich Village)



The outbreak of the pandemic caused a severe perturbation in the house rental market



Relationship between Crime Rates vs. Sale Price



borough	crime rate/ 1000 people
Bronx	18.015627
Brooklyn	12.496215
Manhattan	19.692957
Queens	9.639718
Staten Island	5.311231

Our Observations...

Generally, **Staten Island** is the most affordable and safest county in NYC, and in terms of commute convenience, safety index, and affordability, **Queens** would be the best choice

Conclusions

Conclusions:

- The characteristics such as location, property size, age of the property, and building class affect the sale price
- The income is not necessarily related to the sale price or vice versa
- The pandemic severely influenced the rental price
- The house price does not relate to the crime rate and mostly depends on the convenience of transportation

Challenges:

- Web scraping is not stable
- A crime event cannot be matched with a zip code for each precinct