

DS-GA 1007 PROGRAMMING FOR DATA SCIENCE COURSE PROJECT

Project Title: Analyzing NYC Property Sales for 2003-2022



Group "pip install Grade A"
(Chloe Zheng & Rodrigo Kreis de Paula)

December 2022

Today's Agenda

1

Project Recap

2

.....

What we did

3

.....

Results &
Conclusions

[Project Recap]

Subject: Property Sales in **NYC** over the last **2 decades** - 2003-2022(oct)

Why? Many **implications**: economic, social, and personal

Main goals:

- Observe and analyse trends over the years
- Find the most affordable locations in NYC
- Find which features have a higher/lower influence on prices

Data: Annualized Sales Update by the NYC Department of Finance ([link](#))



NYC 5 boroughs



20 independent variables
1 dependent variable (Sale Price)



1.85 million registers
30% "Sale Price" = 0 registers

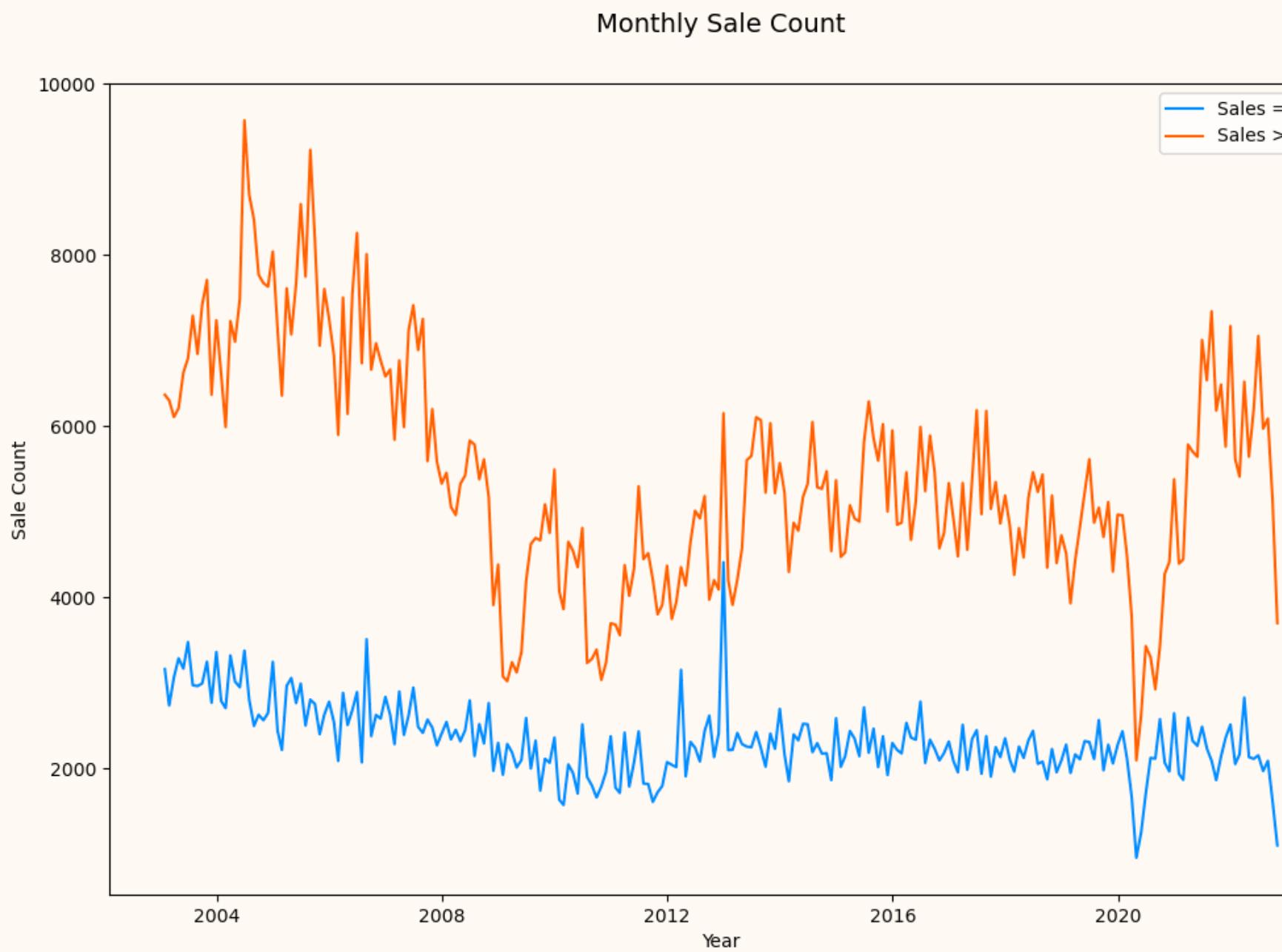
[What we did] - DS-GA 1007 in practice



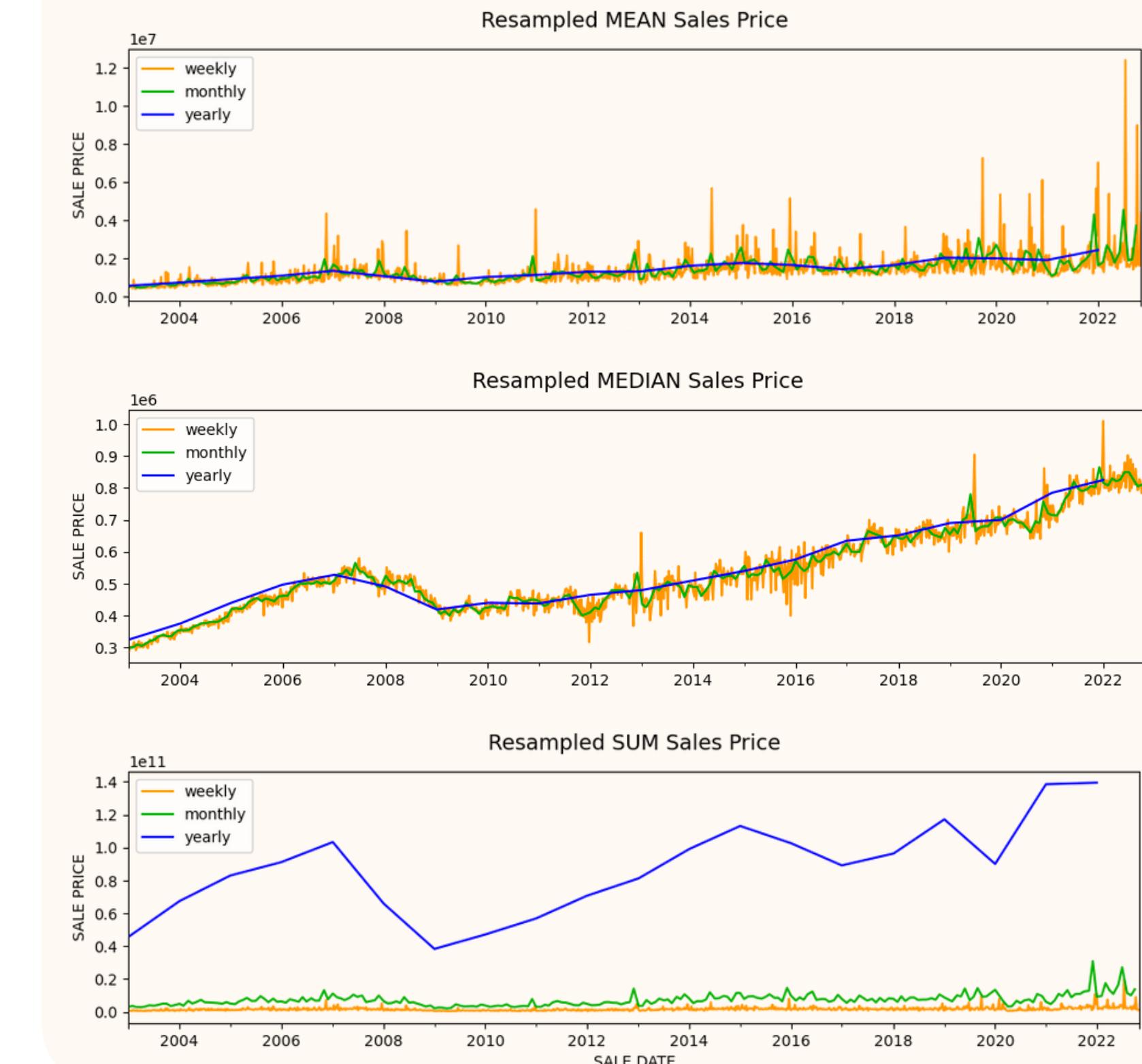
- **Control Flow** via Branching and Iteration
- Defining **functions**
- **Operations on strings**, such as splitting
- **Data Manipulation, Operation, Analysis and Statistical Analysis** with Pandas
- **Data Visualization** with Pandas and Matplotlib Pyplot
- **Time Series Manipulation** with Pandas

[Results & Conclusions] - Overall Sales

Comparing Count of Sales Price > 0
vs. Sale Price = 0



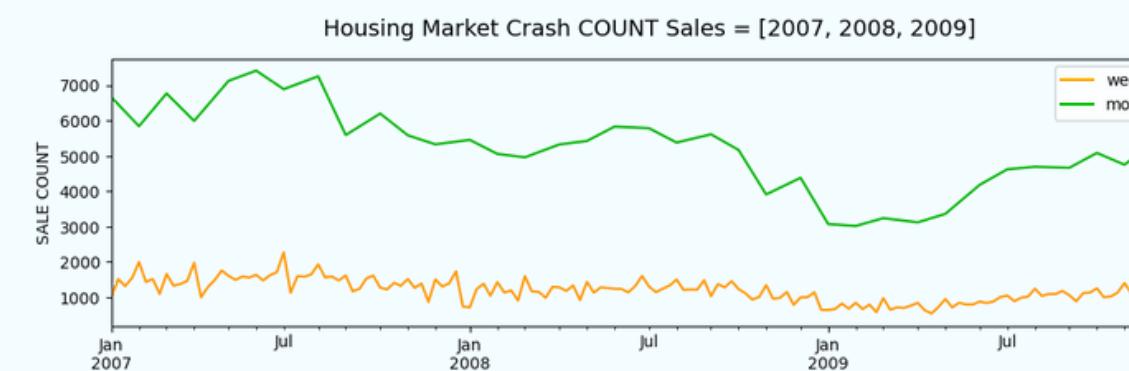
Evolution of Sales Price



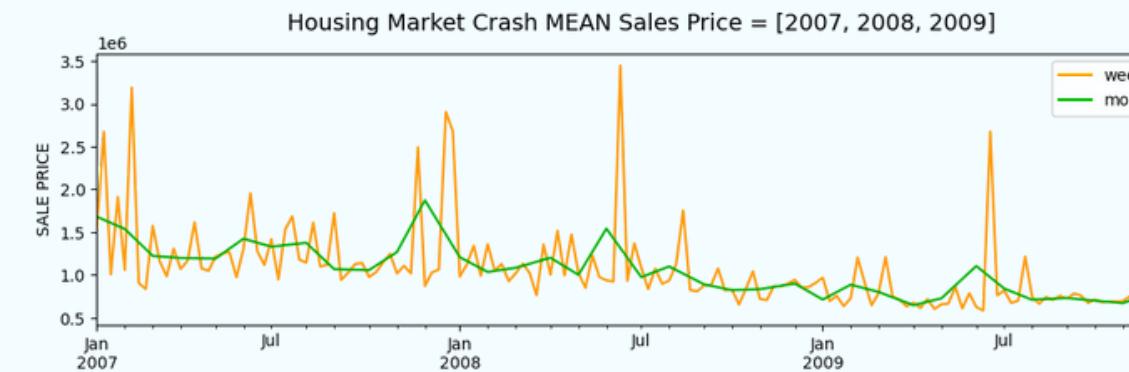
[Results & Conclusions] - Historical Events

Count

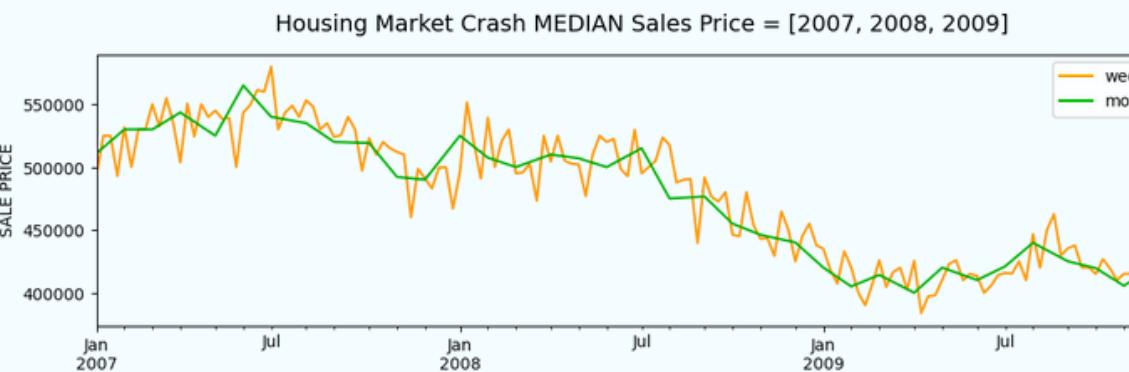
2008 Housing Market Crash



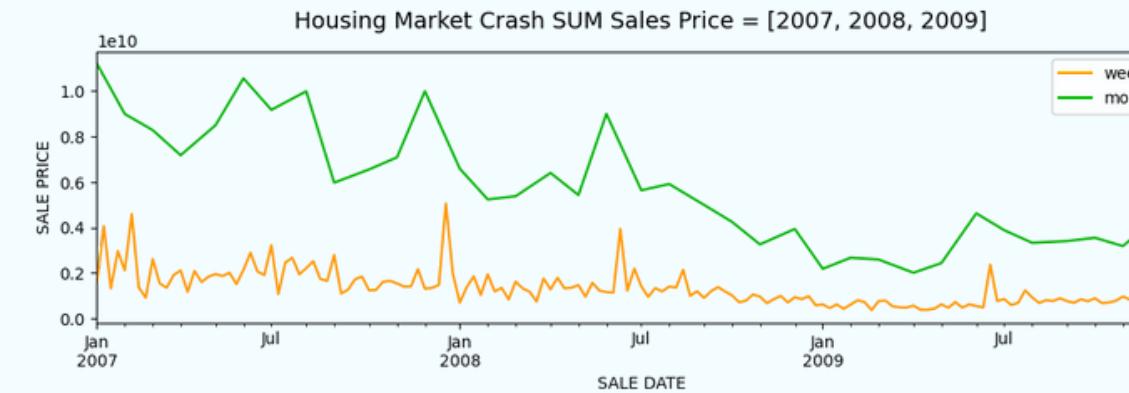
Mean



Median

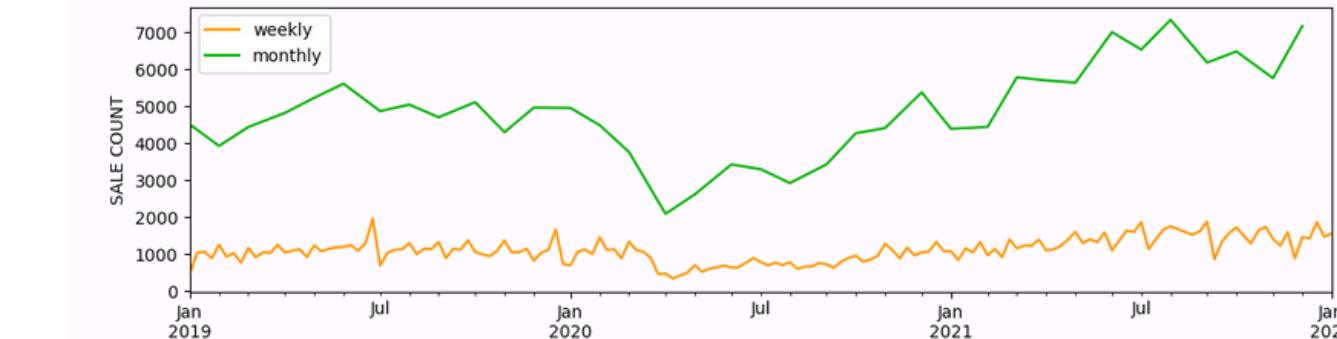


Sum

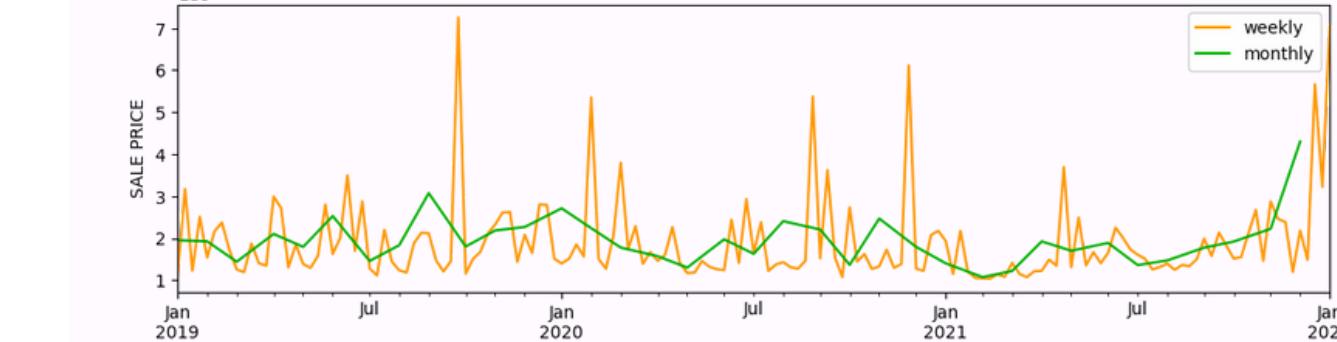


2020 Coronavirus Pandemic

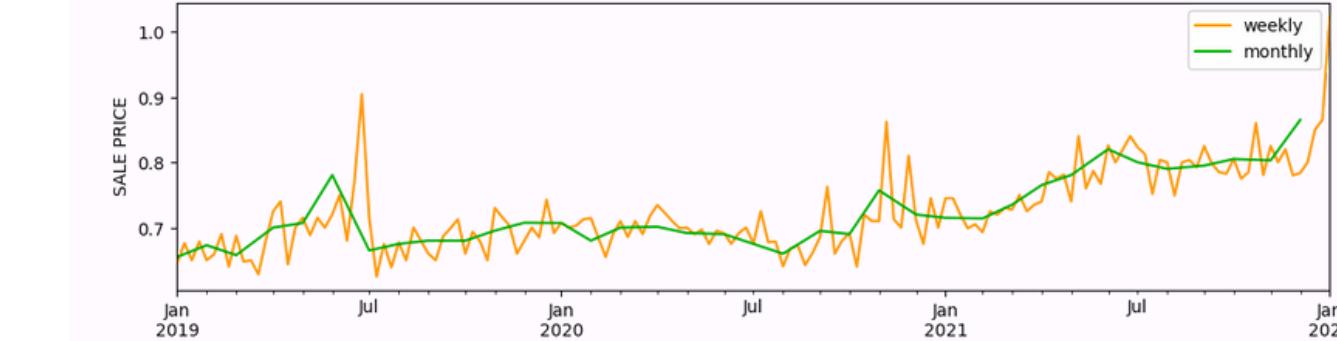
Coronavirus Pandemic COUNT Sales = [2019, 2020, 2021]



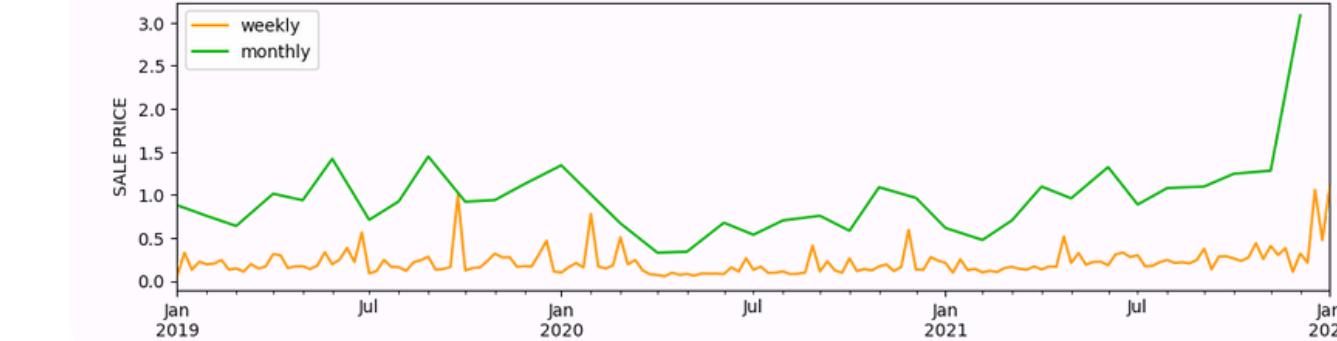
Coronavirus Pandemic MEAN Sales Price = [2019, 2020, 2021]



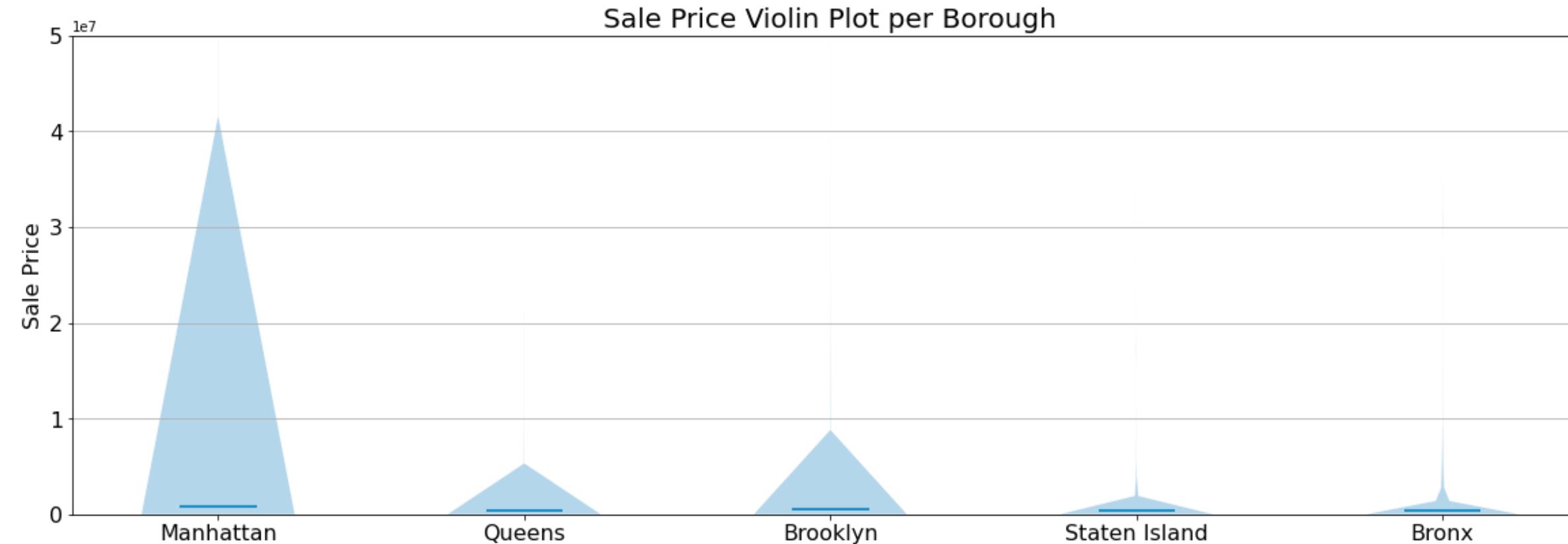
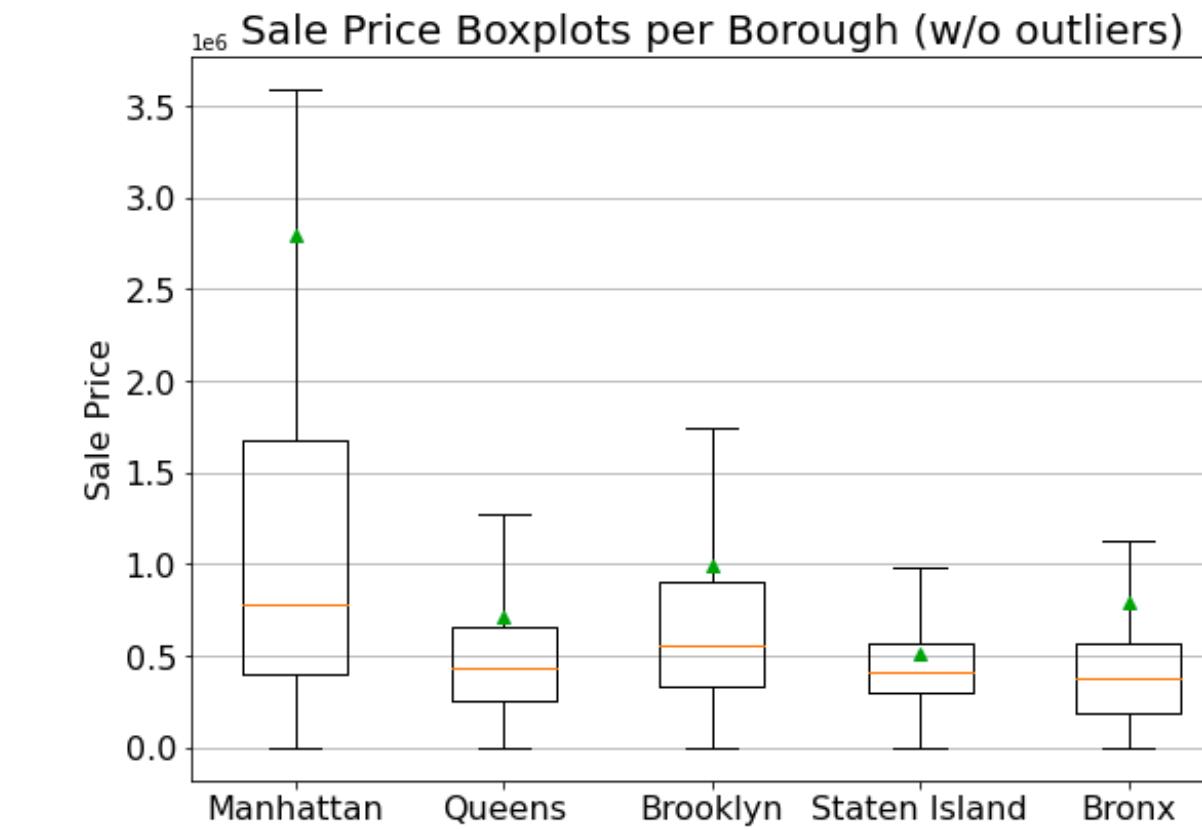
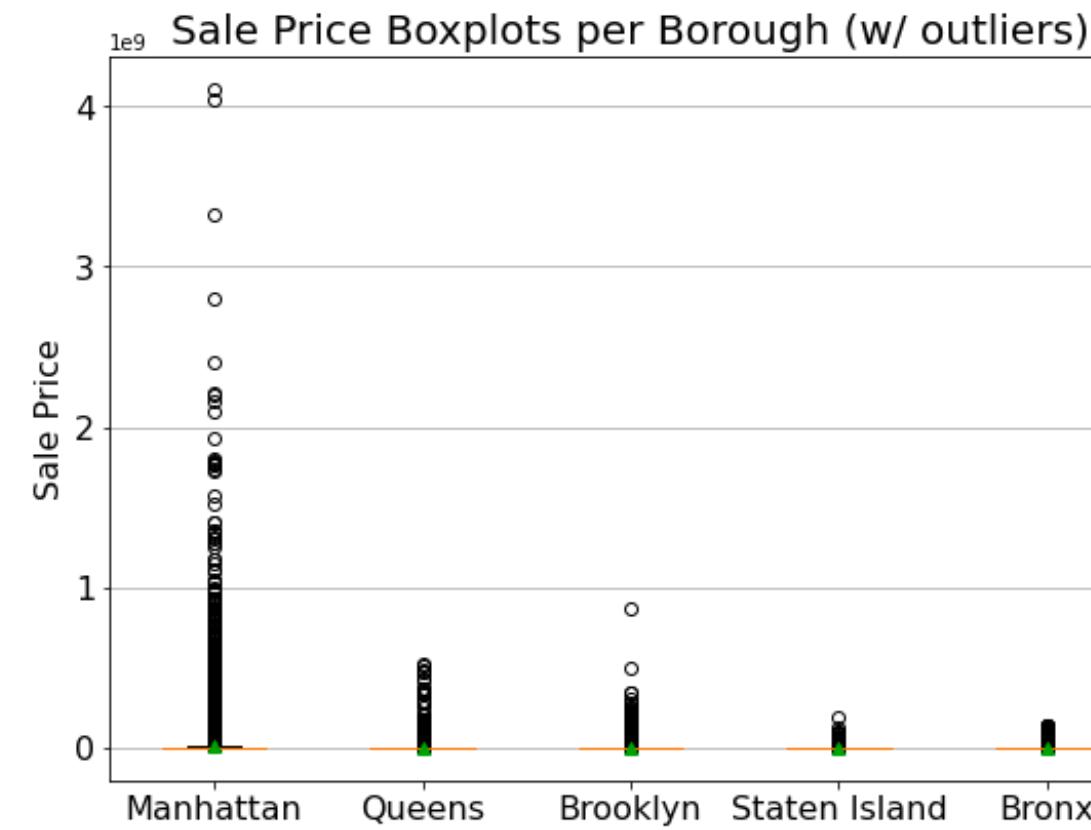
Coronavirus Pandemic MEDIAN Sales Price = [2019, 2020, 2021]



Coronavirus Pandemic SUM Sales Price = [2019, 2020, 2021]



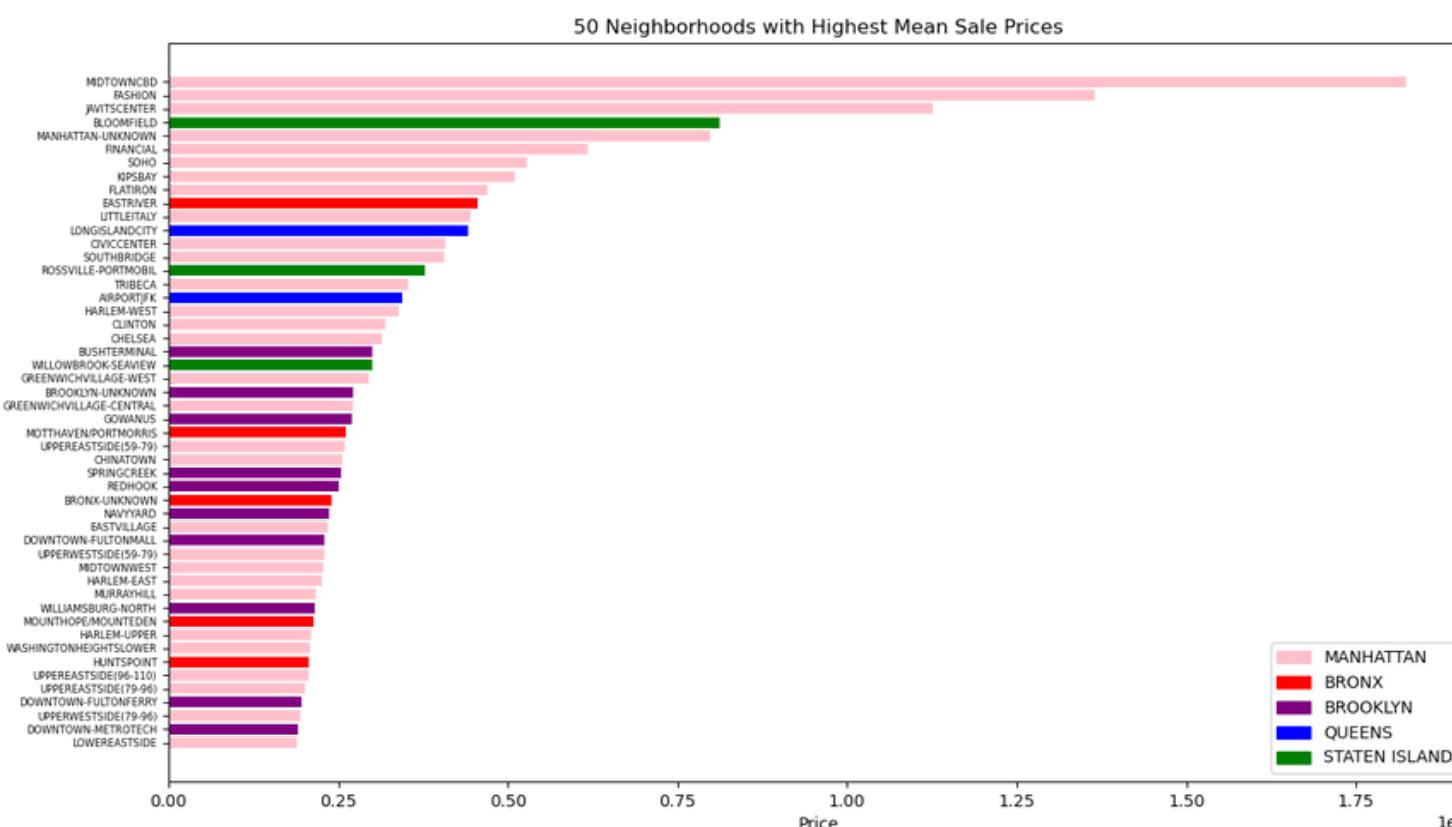
[Results & Conclusions]



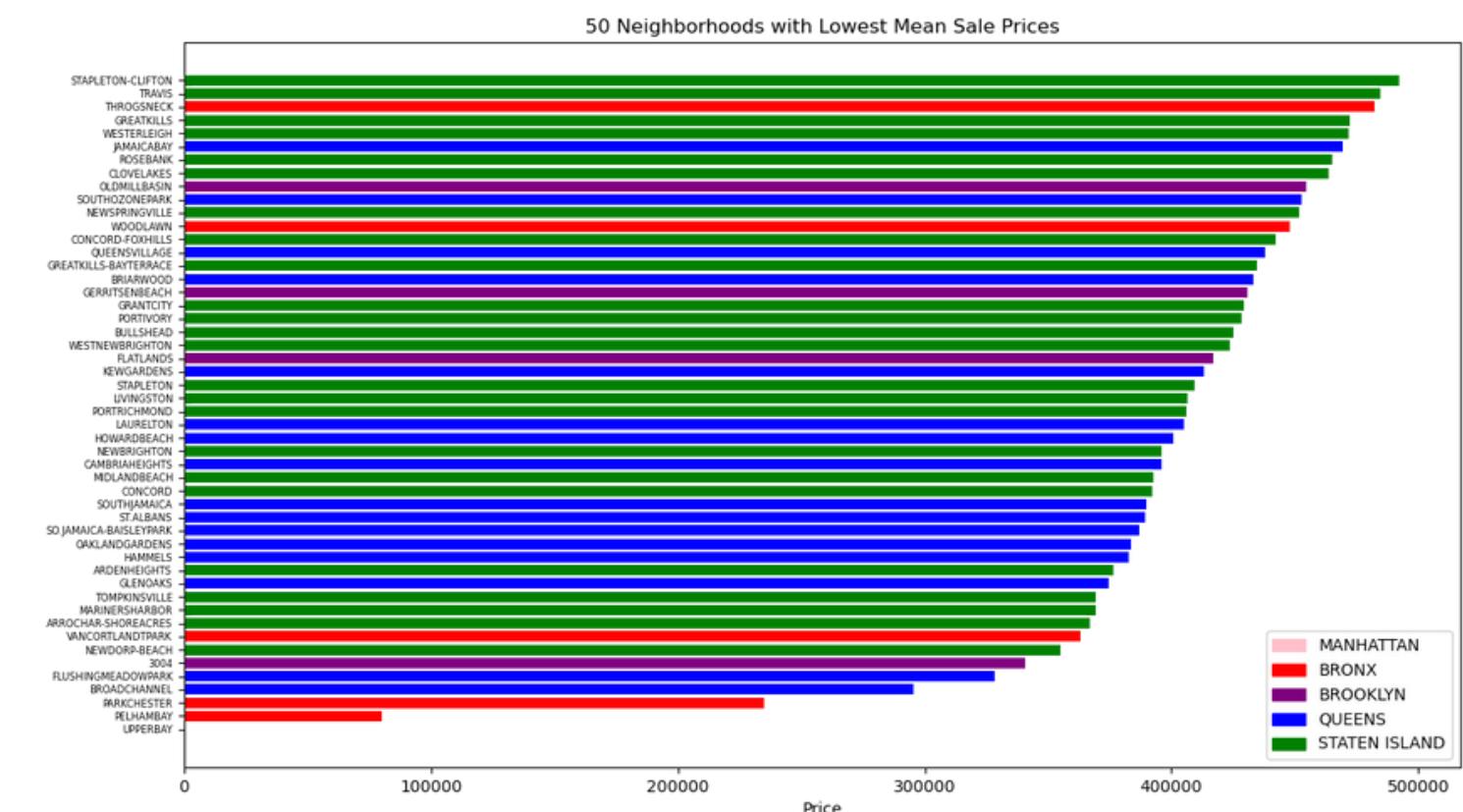
[Results & Conclusions] - Explore Top/Bottom 50 neighborhoods

Highest Price

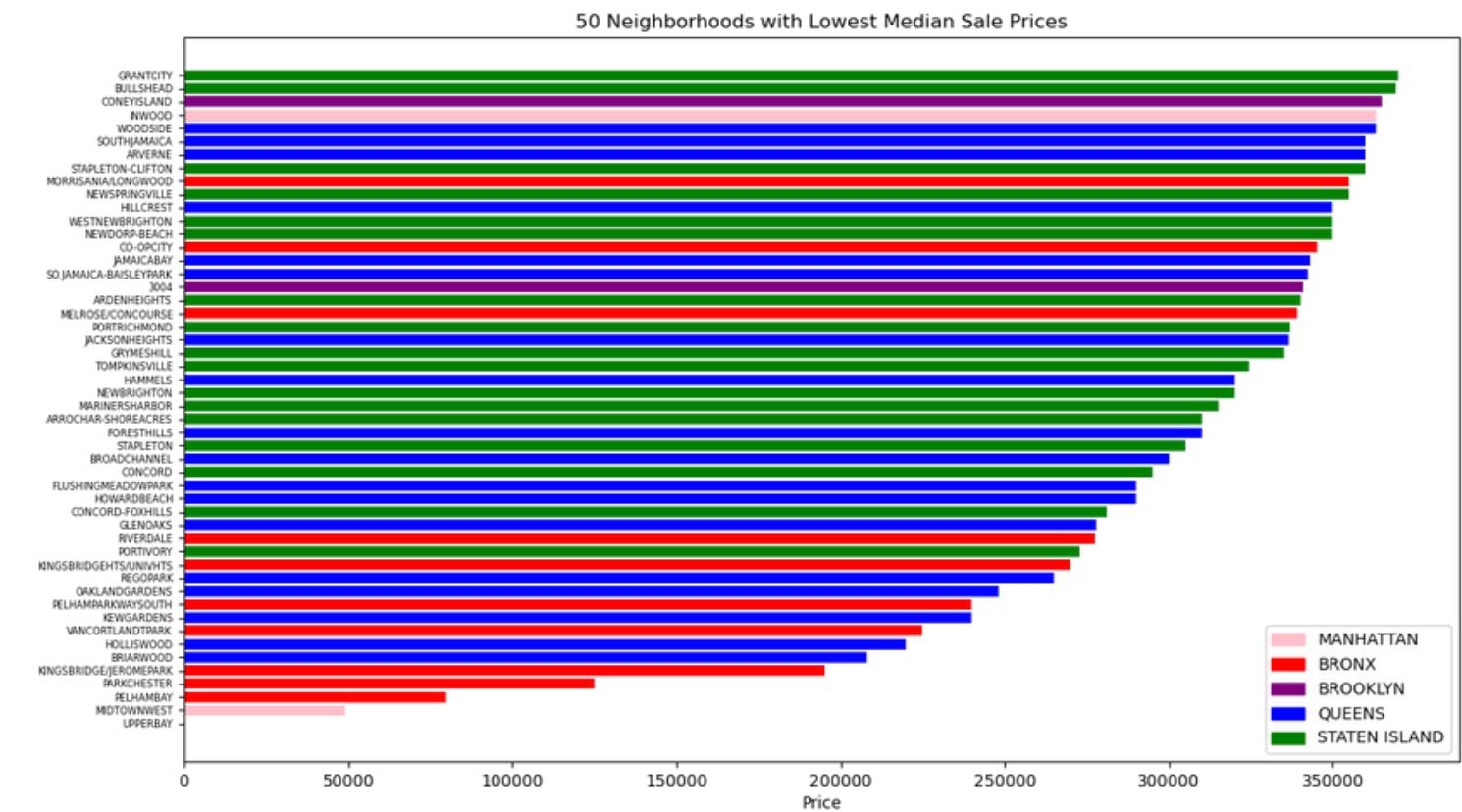
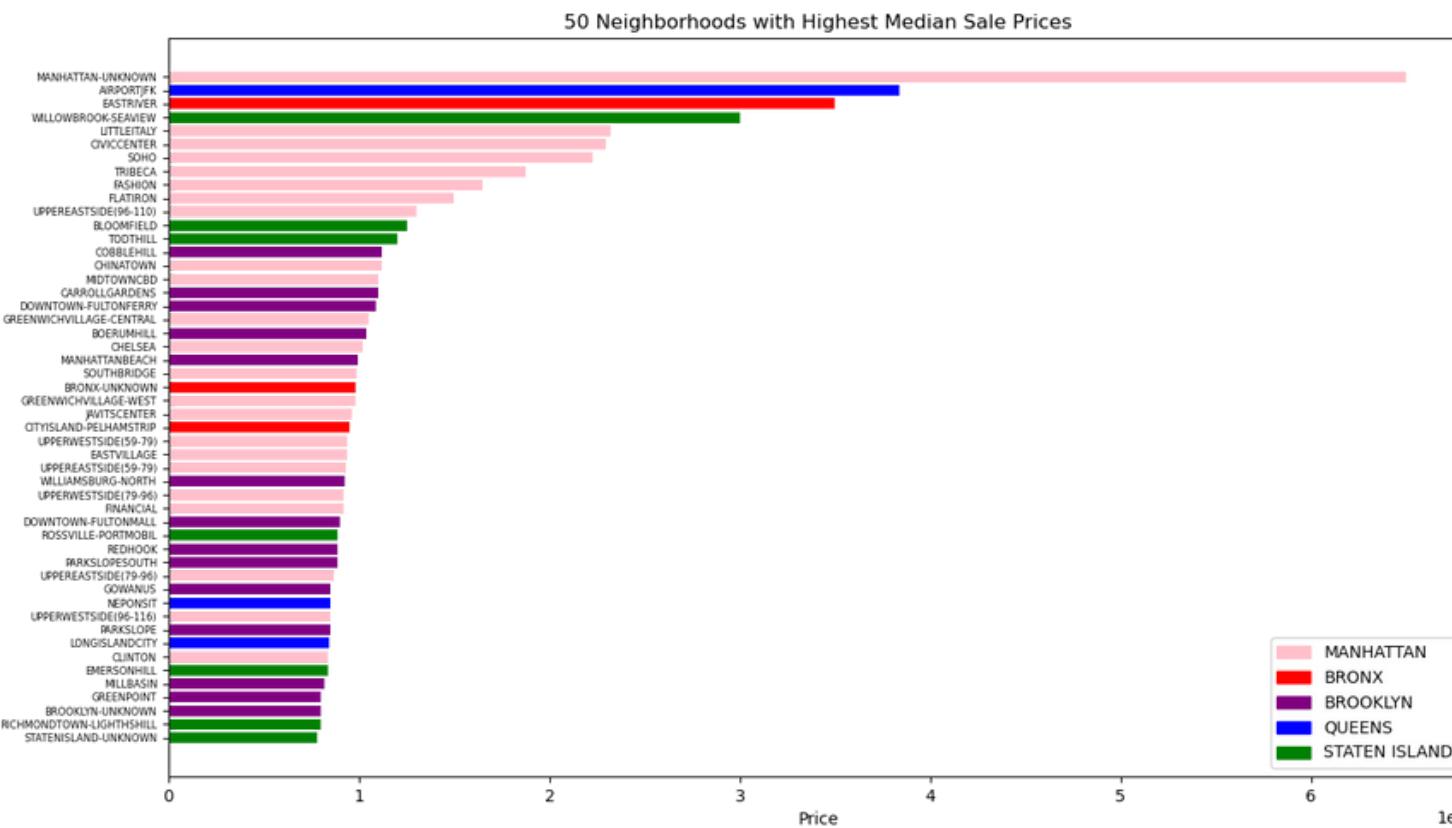
Mean



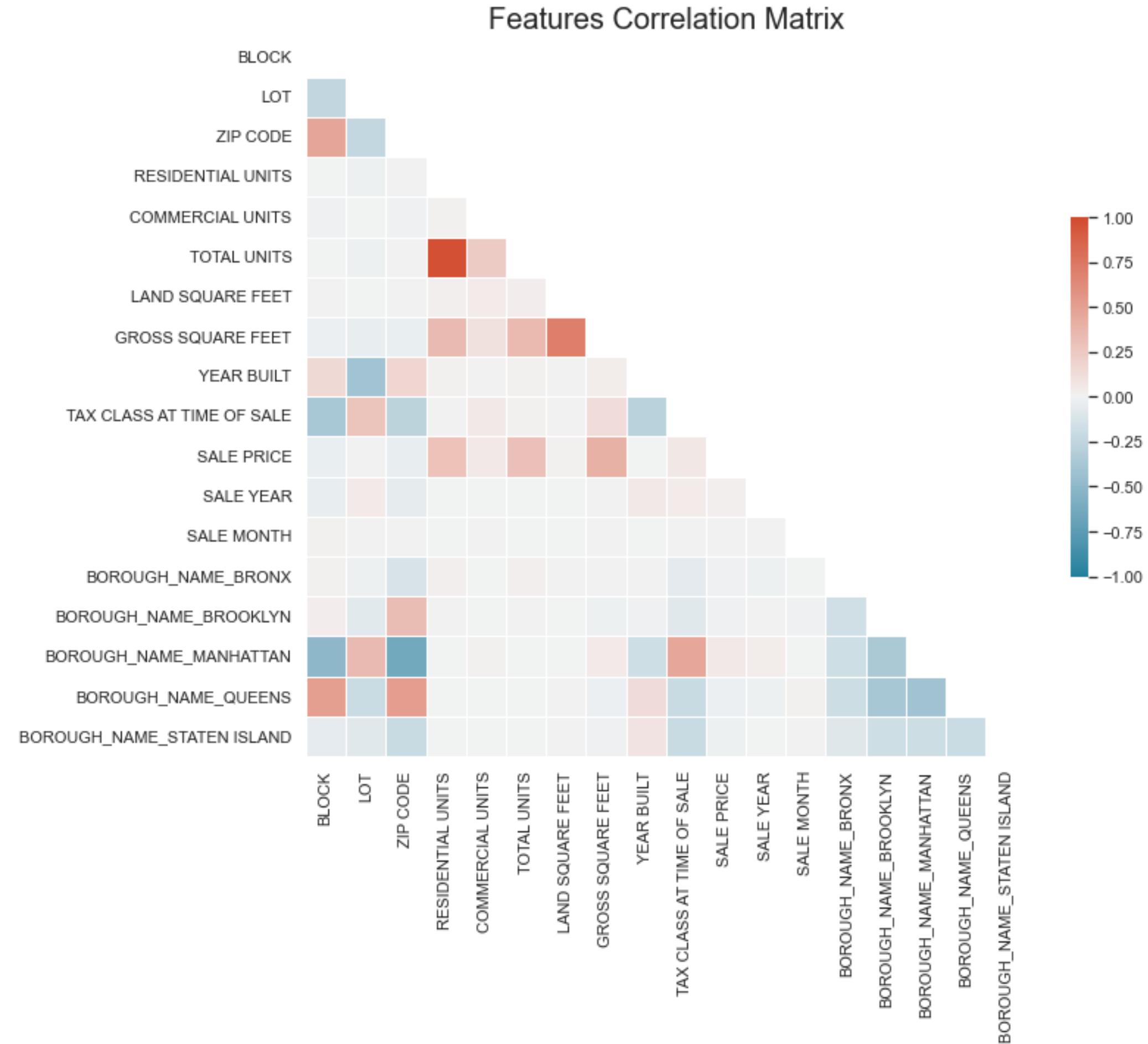
Lowest Price



Median



[Results & Conclusions] - Low correlations with Sale Price



[Summary]

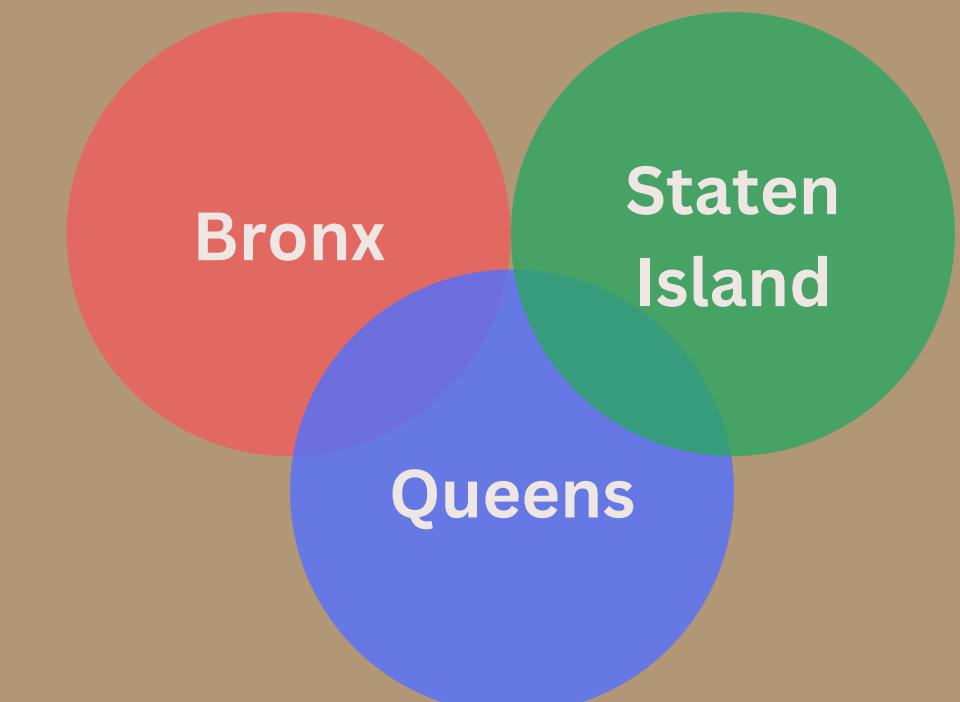
Trends over the years

Overall count of sales slightly decline, but sales prices continue to trend upward

The 2008 Housing Market Crisis caused a decline in prices (more intense and extended than the 2020 Coronavirus Pandemic)

The 2020 Coronavirus caused a brief disruption, followed by renewed buying

Most affordable locations in NYC



Features with a higher influence on prices

Gross Square Feet (bigger)

Total Units (buildings)

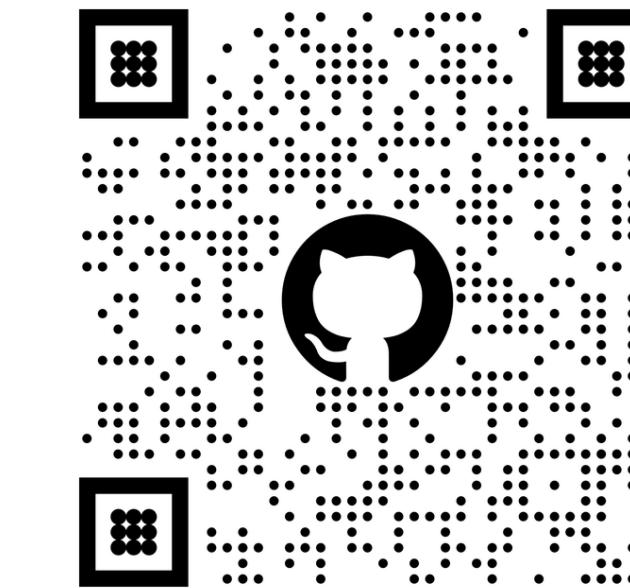
Residential Units (condos)

Final Conclusions and Future Directions

- Annualized Sales Update by the NYC Department of Finance had **useful information to analyze trends, gauge affordability of NYC, and identify important features.**
- More information about the seller and buyer may be useful to understand **who is or is no participating and benefiting** from the housing market
- This data can be used for **predictive modeling** about future trends.



Check the full project
on [GitHub](#)



Thanks!