Sale Price Prediction and Analysis of New York City Property

Padma Parameshwaran

The Problem

New York industry that has been impacted by the pandemic is real estate. The outlook changed significantly.

Brokers, buyers, and sellers are all navigating uncertainty amid the pandemic, as home hunters are forced to tour listings virtually—or not at all.
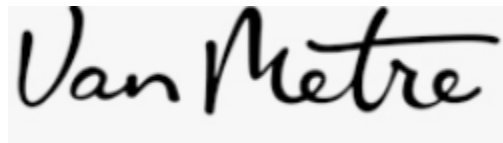
Brokers began working remotely and how was the sales during this year and the expectation is the same at least for another few months.

What is the trend in Property sales in New York city in 2020?

Can we predict the sale price of a property?

Who might care?

Real Estate Business

Home Builders

Commercial Builders

Home Buyers

Rental Agencies

Property Management Services

…and many more …

Data Information

New York Department of Finance has list of properties sold during the last 12 months sorted based on Tax classes.
https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page

The data has properties that were sold during 2020 in New York city with a separate listing for each borough namely Manhattan, Bronx, Brooklyn, Queens, and Staten Island.

Metrics collected: Borough, Neighborhood, Building Class Category, Tax class at present, Block, Lot, Easement, building class at present, Address, Apartment Number, ZIP code, Residential units, Commercial units, Total units, Land square feet, Gross square feet, Year built, Tax class at time of sale, Building class at time of sale, Sale price and Sale date

Class 1: Includes most residential property of up to three units (such as one-, two-, and three-family homes and small stores or offices with one or two attached apartments), vacant land that is zoned for residential use, and most condominiums that are not more than three stories.

Class 2: Includes all other property that is primarily residential, such as cooperatives and condominiums.

Class 3: Includes property with equipment owned by a gas, telephone, or electric company.

Class 4: Includes all other properties not included in class 1,2, and 3, such as offices, factories, warehouses, garage buildings, etc.

Building Class Category : similar properties grouped by broad usage

| BOROUGH	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AT PRESENT	BLOCK	LOT	EASE-MENT	BUILDING CLASS AT PRESENT	ADDRESS	APARTMENT NUMBER	ZIP CODE	RESIDENTIAL UNITS	COMMERCIAL UNITS	TOTAL UNITS	LAND SQUARE FEET	GROSS SQUARE FEET	YEAR BUILT	TAX CLASS AT TIME OF SALE	BUILDING CLASS AT TIME OF SALE	SALE PRICE	SALE DATE
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3029	36	NaN	A5	418 EAST 180 STREET	NaN	10457.0	1.0	0.0	1.0	1568.0	1526.0	1899.0	1	A5	0	2020-07-09
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3030	65	NaN	A1	4455 PARK AVENUE	NaN	10457.0	1.0	0.0	1.0	1622.0	1587.0	1899.0	1	A1	458000	2020-01-31
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3030	67	NaN	A1	4451 PARK AVENUE	NaN	10457.0	1.0	0.0	1.0	1646.0	1495.0	1899.0	1	A1	275000	2020-03-16
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3030	67	NaN	A1	4451 PARK AVENUE	NaN	10457.0	1.0	0.0	1.0	1646.0	1495.0	1899.0	1	A1	475000	2020-07-23
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3039	29	NaN	A9	2327 WASHINGTON AVE	NaN	10458.0	1.0	0.0	1.0	1095.0	1260.0	1910.0	1	A9	165000	2020-03-02

Data Wrangling

| BOROUGH	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AT PRESENT	BLOCK	LOT	EASE-MENT	BUILDING CLASS AT PRESENT	ADDRESS	APARTMENT NUMBER	ZIP CODE	RESIDENTIAL UNITS	COMMERCIAL UNITS	TOTAL UNITS	LAND SQUARE FEET	GROSS SQUARE FEET	YEAR BUILT	TAX CLASS AT TIME OF SALE	BUILDING CLASS AT TIME OF SALE	SALE PRICE	SALE DATE
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3029	36	NaN	A5	418 EAST 180 STREET	NaN	10457.0	1.0	0.0	1.0	1568.0	1526.0	1899.0	1	A5	0	2020-07-09
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3030	85	NaN	A1	4455 PARK AVENUE	NaN	10457.0	1.0	0.0	1.0	1622.0	1587.0	1899.0	1	A1	458000	2020-01-31
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3030	67	NaN	A1	4451 PARK AVENUE	NaN	10457.0	1.0	0.0	1.0	1646.0	1495.0	1899.0	1	A1	275000	2020-03-16
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3030	67	NaN	A1	4451 PARK AVENUE	NaN	10457.0	1.0	0.0	1.0	1646.0	1495.0	1899.0	1	A1	475000	2020-07-23
2	BATHGATE	01 ONE FAMILY DWELLINGS	1	3038	29	NaN	A9	2327 WASHINGTON	NaN	10458.0	1.0	0.0	1.0	1095.0	1260.0	1910.0	1	A9	165000	2020-03-02

- **Combined data from all boroughs**

- **Sale price filtered to contain 100,000 to 3 million**

- **Coop homes shares sold treated as selling price**

- **Missing Total units and Residential units were populated based on other entries for the same building type.**

- **KNN Imputing to fill the missing entries**

- **One hot encoding and Standard scaling**
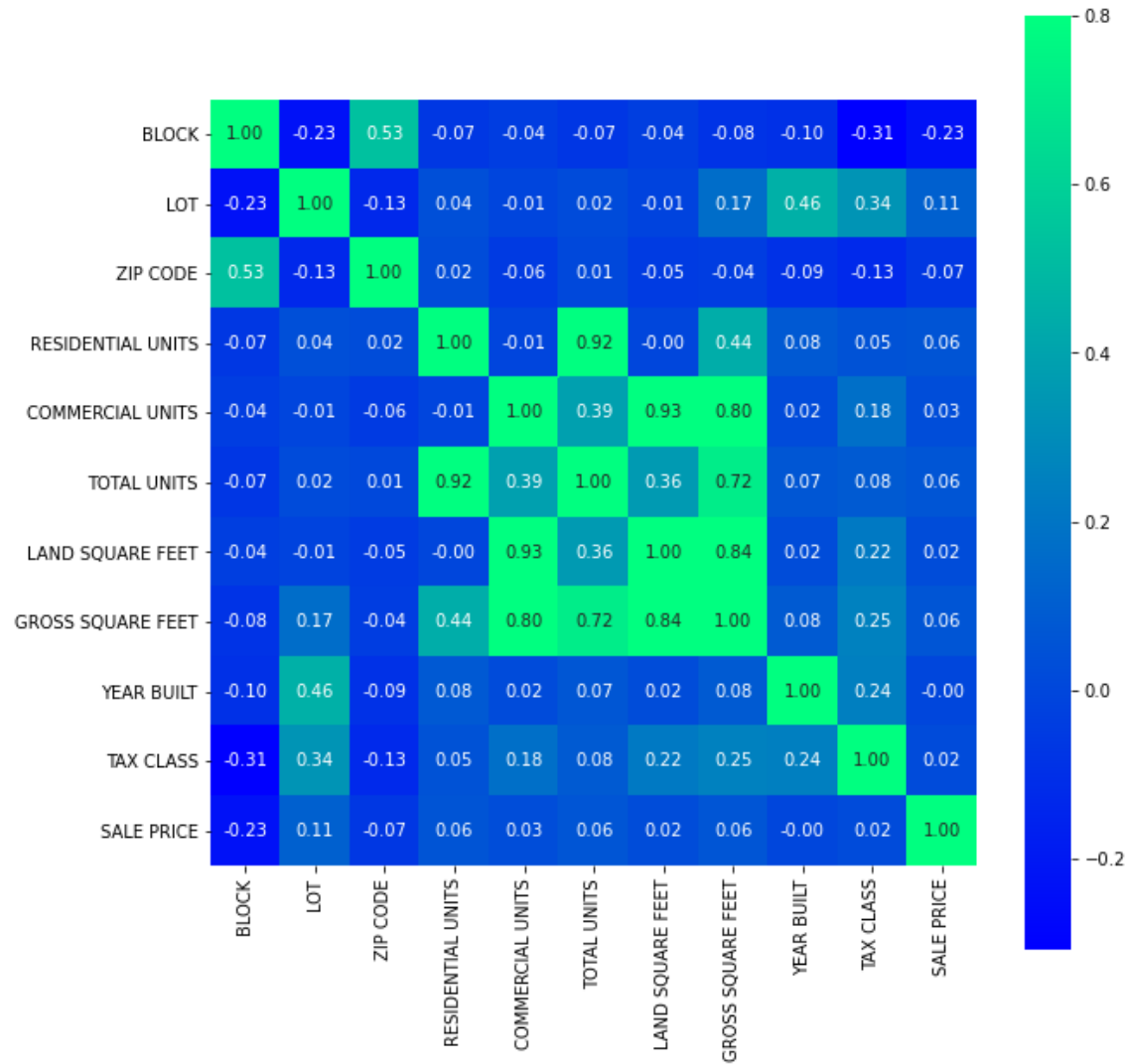
- **Residential Units and Total Units are highly correlated.**

- **Commercial units and Land Square Feet is correlated.**

- **Very few entries did not have year built.**

Number of Sales per Borough in 2020
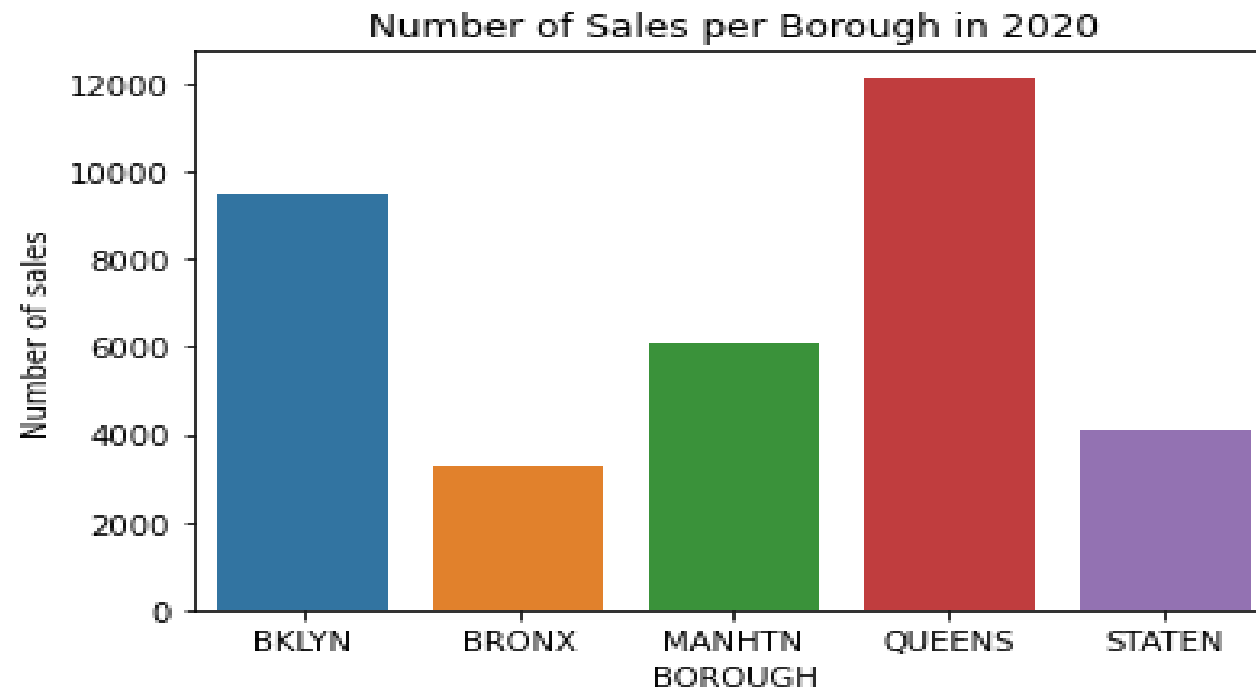
More sales were in Queens than other boroughs last year and Bronx had the least property sales

Sale Price by Building Class Category
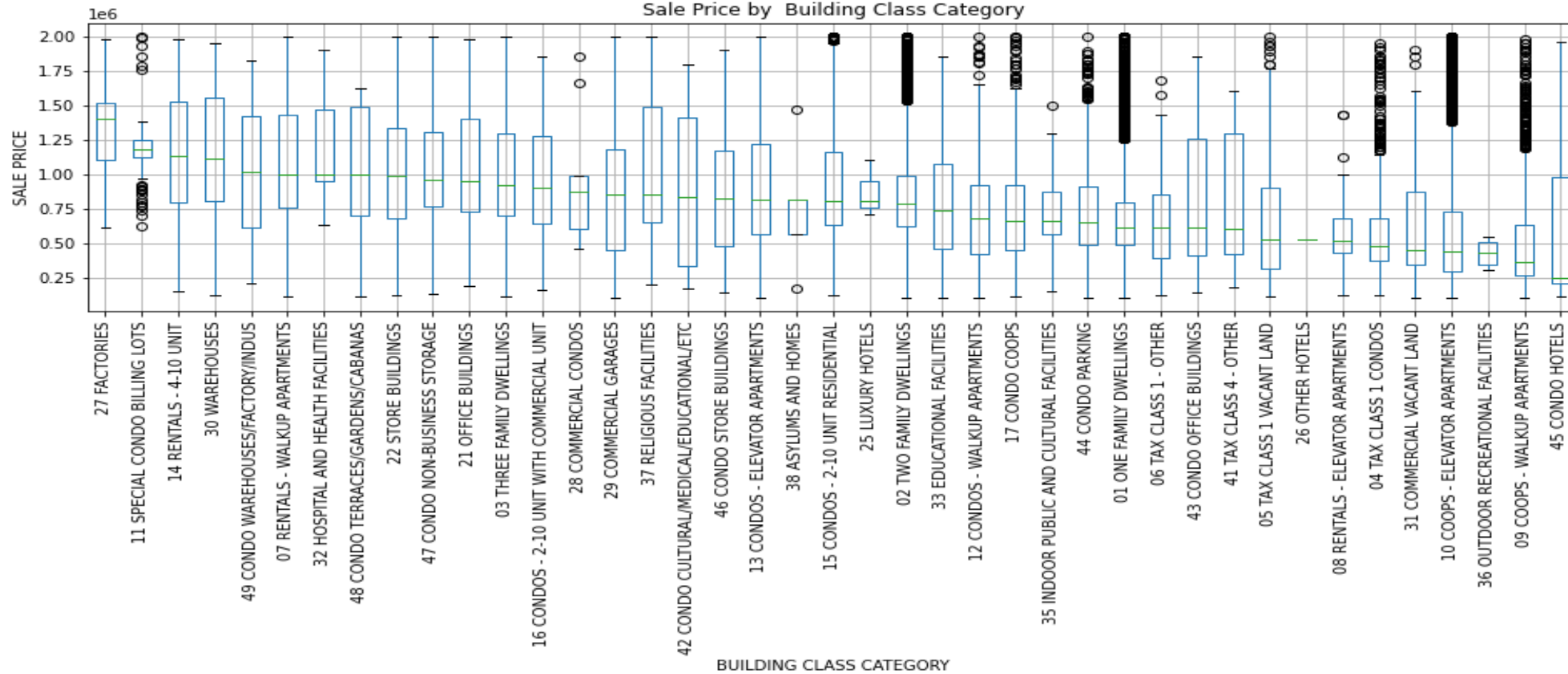
Factories had the highest sale price and condo hotels had the least median sale price.

Some building categories had less than 10 sales, so they were merged to category Others. Here is the distribution of sale price for a given building category.
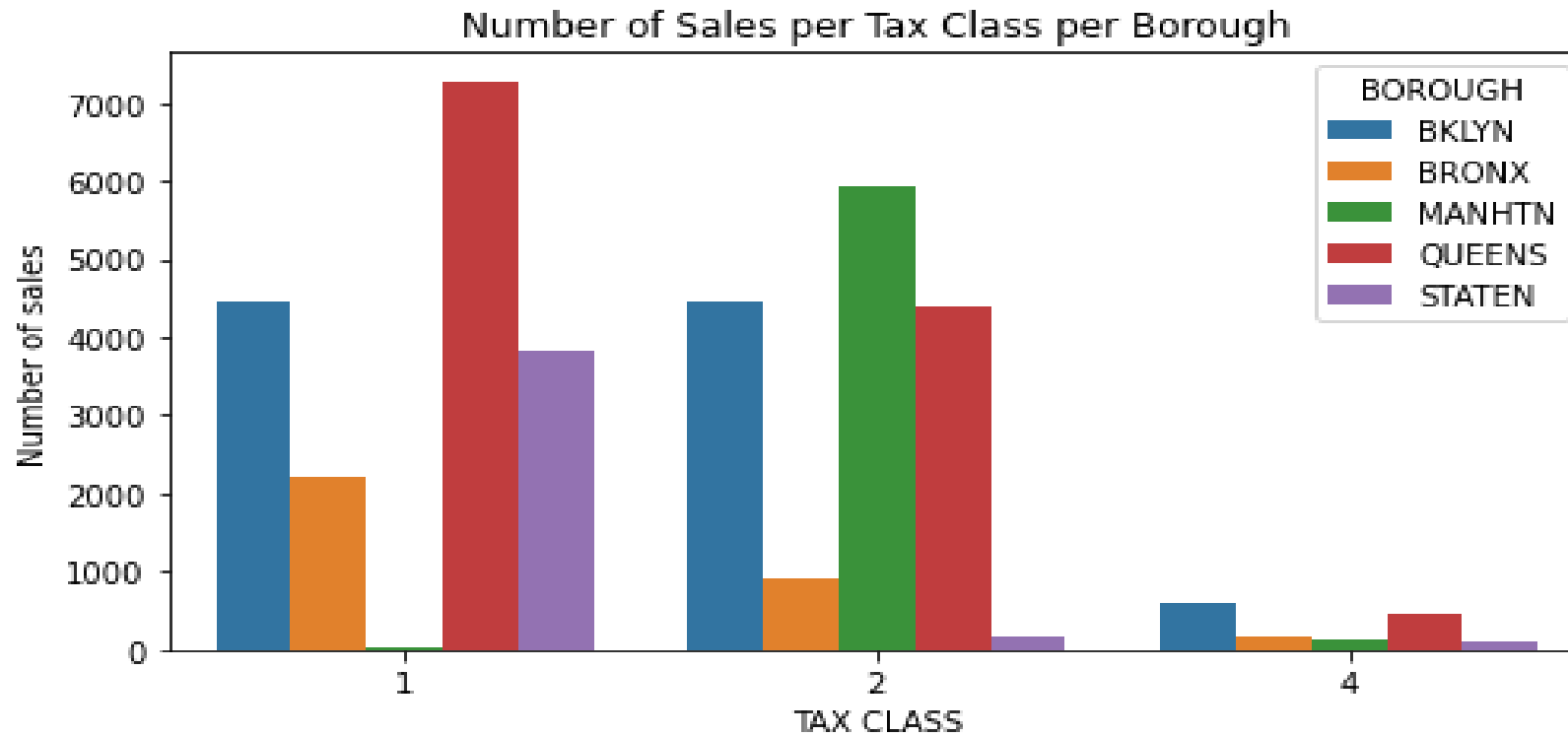
Distribution of vacant land purchases in 2020
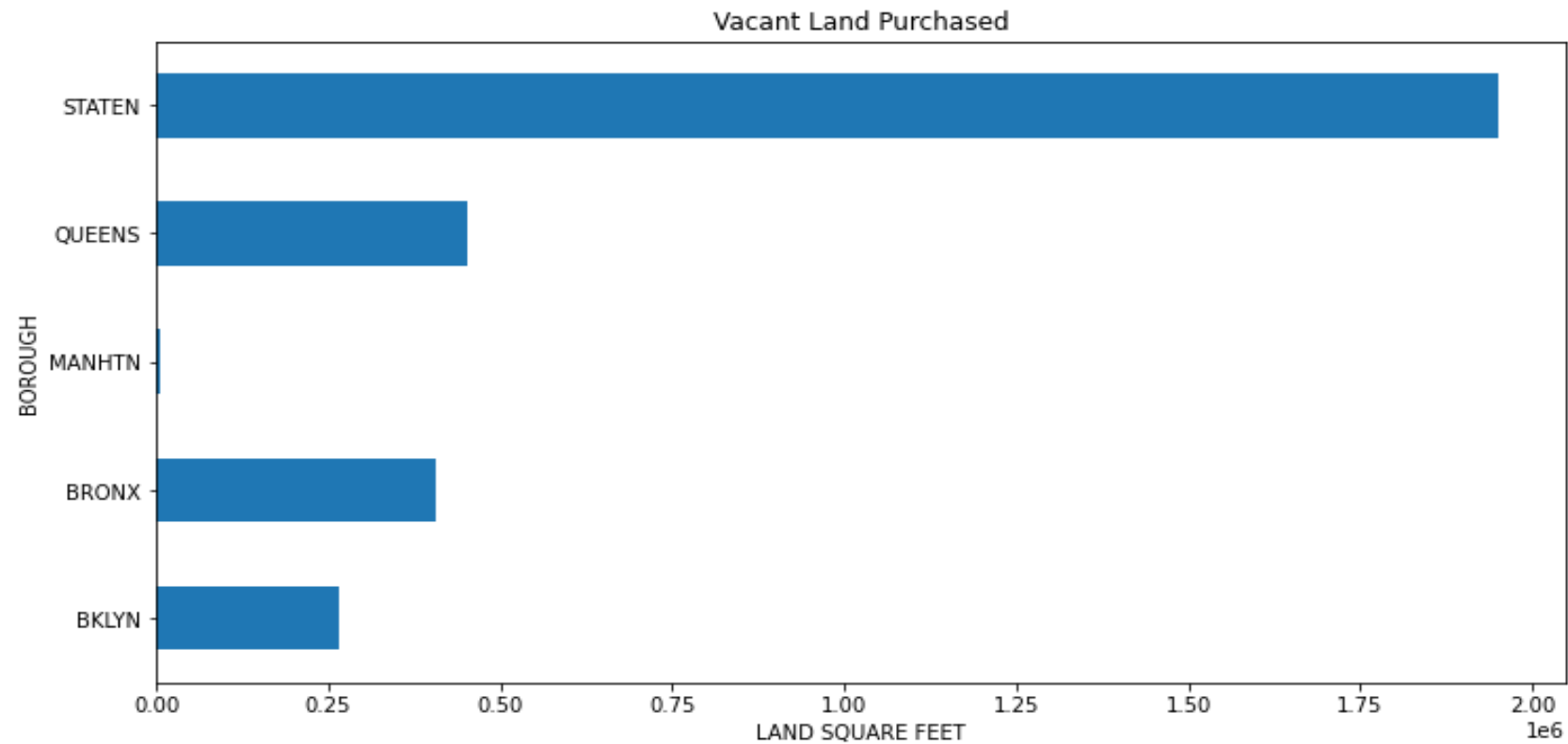
There are more chances of new housing and commercial construction in Staten Island than in other boroughs and the least in Manhattan. Same in terms of neighborhood.
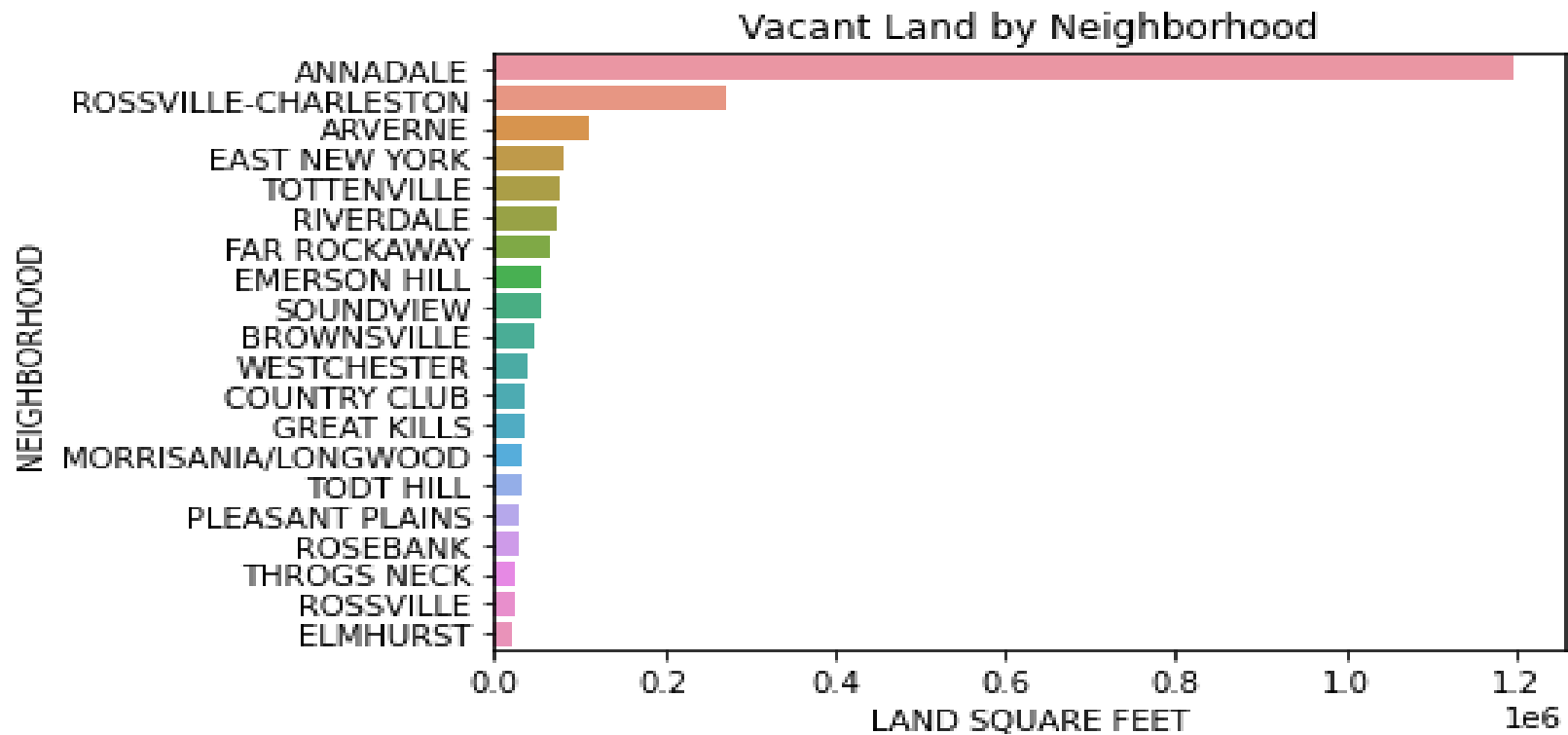
The top 20 neighborhoods where new construction is possible. Most of them are in Statenisland

Number of Sales per Tax Class per Borough

More homes of higher tax brackets were sold in Queens and none in Manhattan in 2020. Also, very few commercial and industrial properties were sold last year probably due to COVID. Most of the sales for Manhattan was in Tax class 2 which is not surprising as it has many properties which has more than 3 units.

Few houses in Brooklyn are incredibly old built in 1800's and their price is comparable with newer houses elsewhere. Bronx has comparatively lesser priced houses than other boroughs. Manhattan has more expensive houses.

Sale Price by Year Built

Sale Price by Residential Units
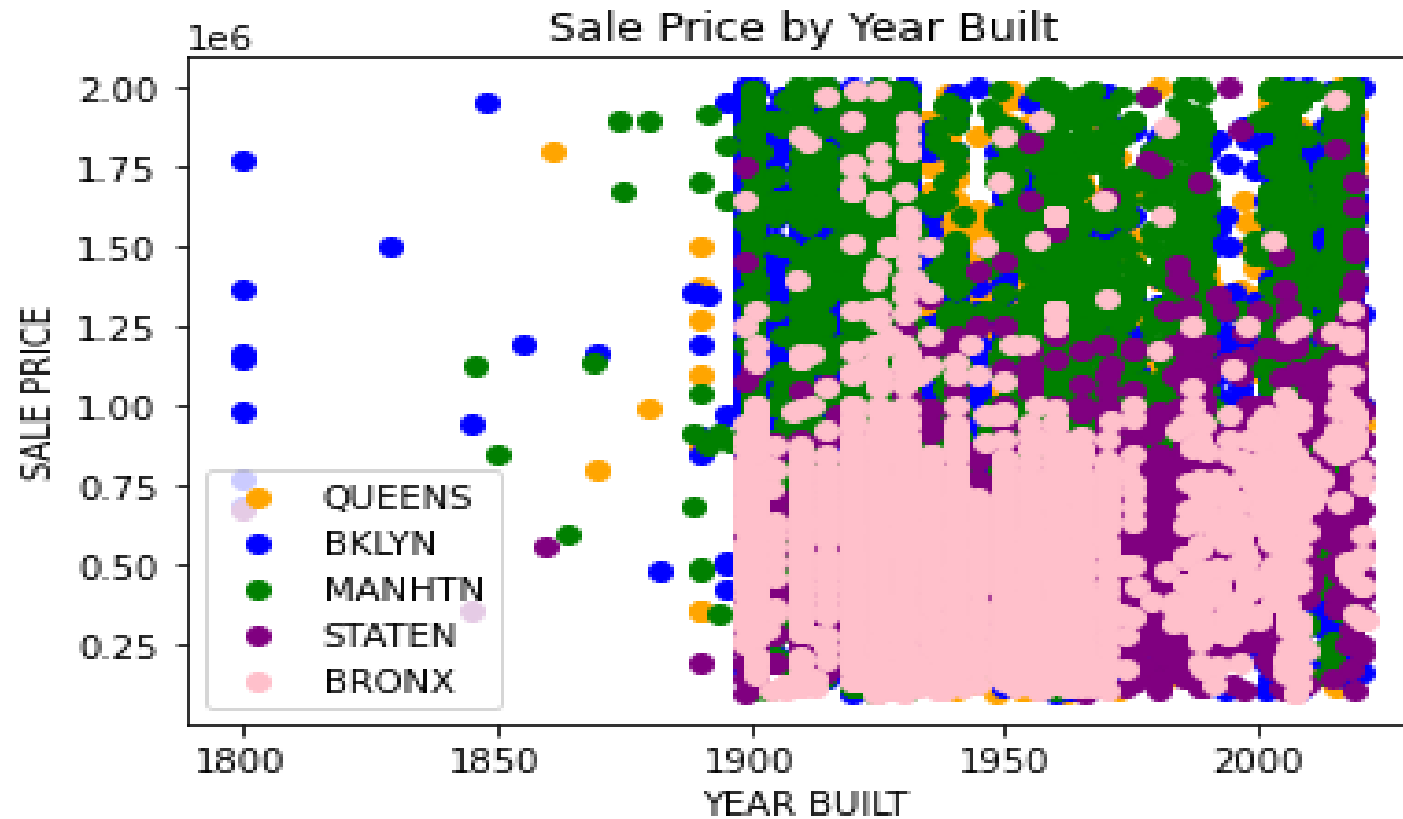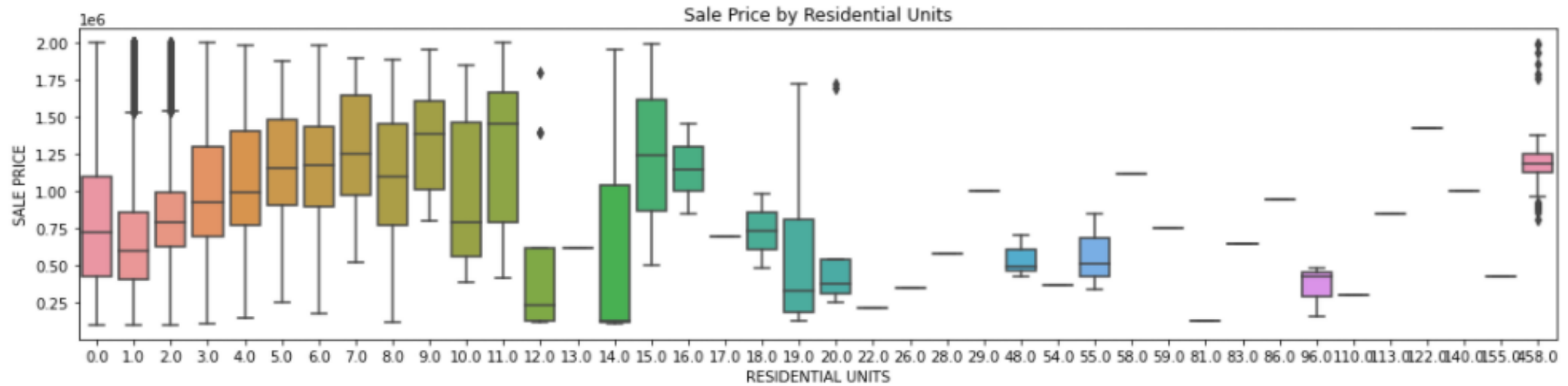
Residences that had 9 units had the highest median sale price compared to even higher units. The multi-unit building could be in Manhattan and the single unit could be in Brooklyn. Similarly, for commercial buildings below has 7 units as highly priced sales

Sale Price by Commercial Units
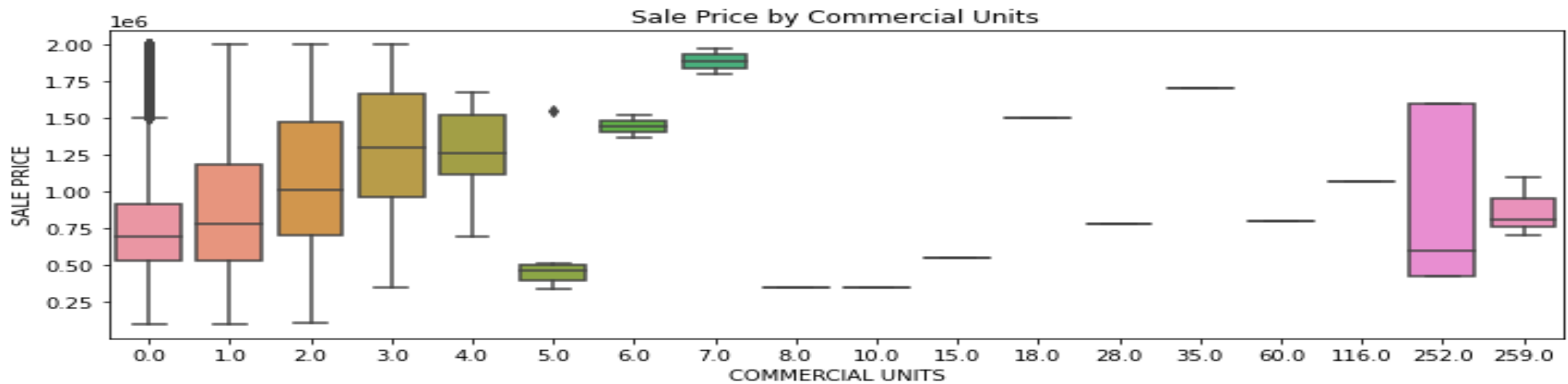
The table has the models that was run on the data. The numeric variables were standardized for some models.
It is evident that Gradient Boosting is the best model compared to all other models. The hyperparameters that produced the best results were
learning_rate: 0.05, max_depth: 7 and n_estimators: 1000

Model	R2 Score	Mean Square Error
Linear Regression	-66989904064163466185592128.00	10189285197753888768.00
SVM	0.20	352024.54
Random Forest	0.46	288921.96
KNN	0.47	286146.00
Gradient Boosting	0.57	256759.93

Future Research

As the data is limited, we could only obtain so much accuracy from public dataset. But this can be further improved by further approaches as listed here to yield better results.

- Segment dataset for each borough and a separate modelling for each borough and further for each Tax class would yield better results.

- Segment Building categories into groups and separate model for each of them

- Add more metrics like number of floors, bedrooms, bathrooms, flooring, kitchen and many more.