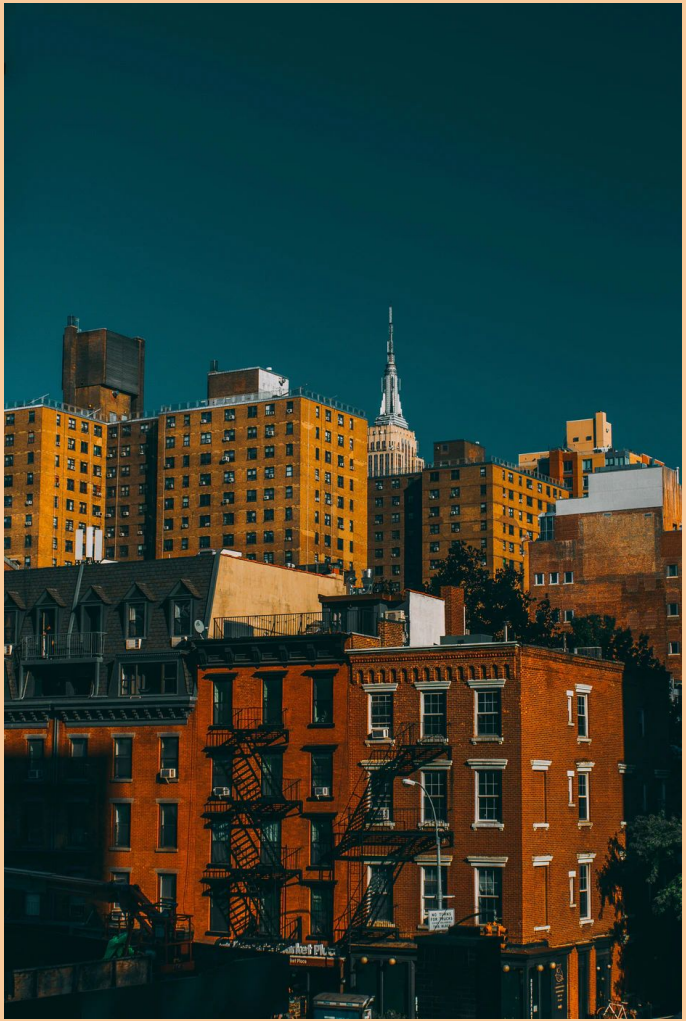


Predicting NYC Property Prices with Machine Learning

Walid, Emil, Aziz, Jorge

Surprising Facts about NY Housing

- The average New York rent is about 82% of the median American salary.
- The cheapest Manhattan neighborhood has an average rent of more than $1,600.
- Median gross rent (the base rent plus estimated utilities) rose 10 percent between 2005 and 2011, but median household income in New York City fell.

The Data Problem

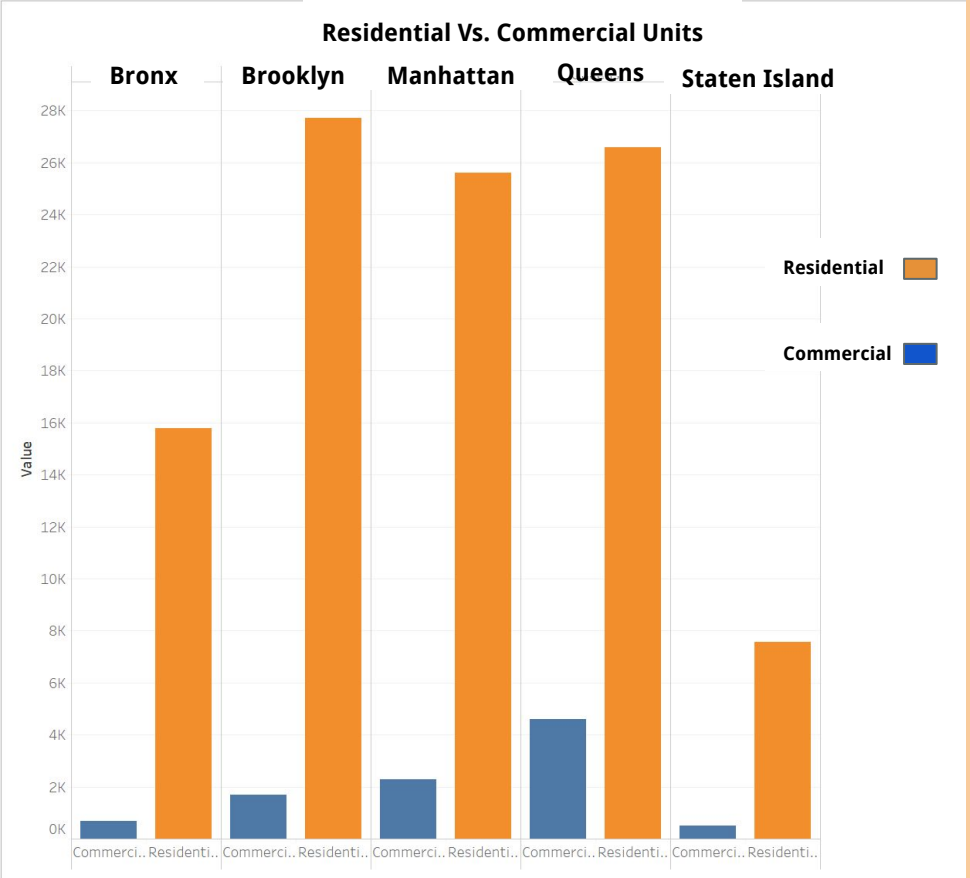
- What can you discover about New York City real estate by looking at a year's worth of raw transaction records?
- Can you spot trends in the market, or build a model that predicts sale value in the future?

Data Set

- NYC-Rolling-Sales dataset from Kaggle
- This dataset is a record of every building or building unit (apartment, etc.) sold in the New York City property market over a 12-month period (September 2016 to September 2017).
- Process:
  - Python, Jupyter Notebook (ETL)
  - Scikit Learn, Keras Regressor (Machine Learning)
  - Tableau (Data Visualization)

Residential vs Commercial Units

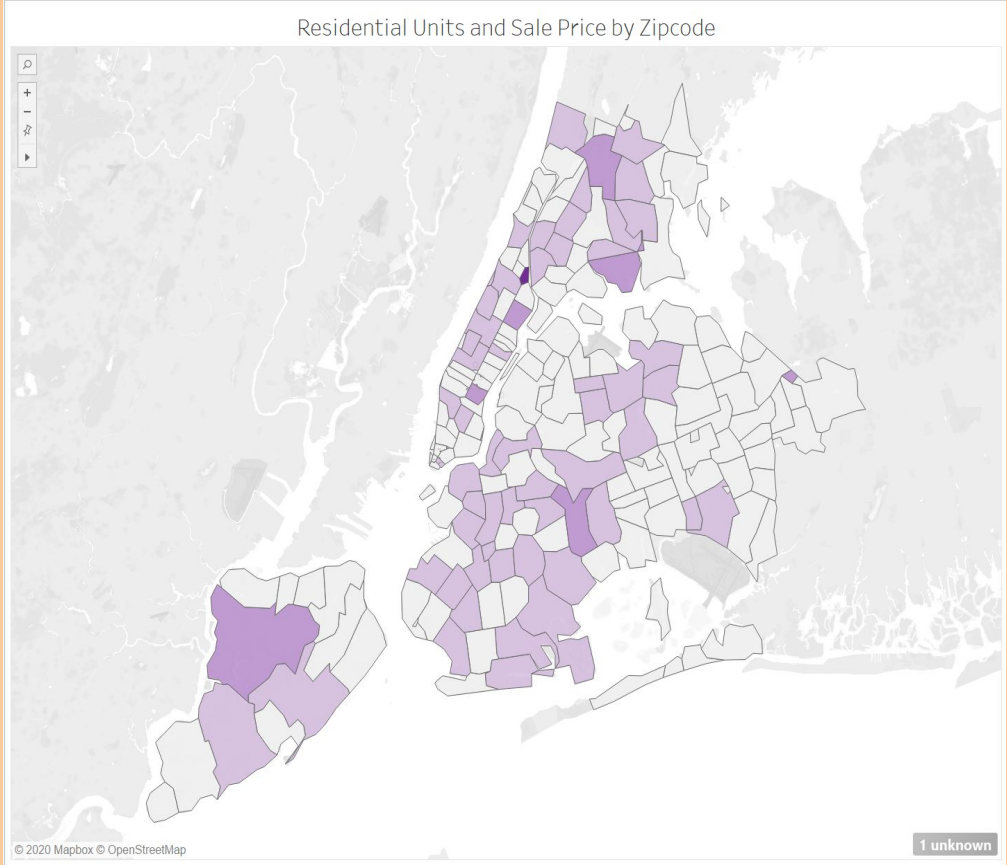
- Brooklyn has most residential units (27,724)
- Queens has the most commercial units (4,586)
- Manhattan has the highest avg sale price (combined residential and commercial) $3,337,951.
- Staten Island has the lowest avg sale price (combined residential and commercial) $543,472.



Residential Units and Sale Price

- Most Units found in one zip code (10037) was in North-East Harlem, Manhattan with 4,594 units.

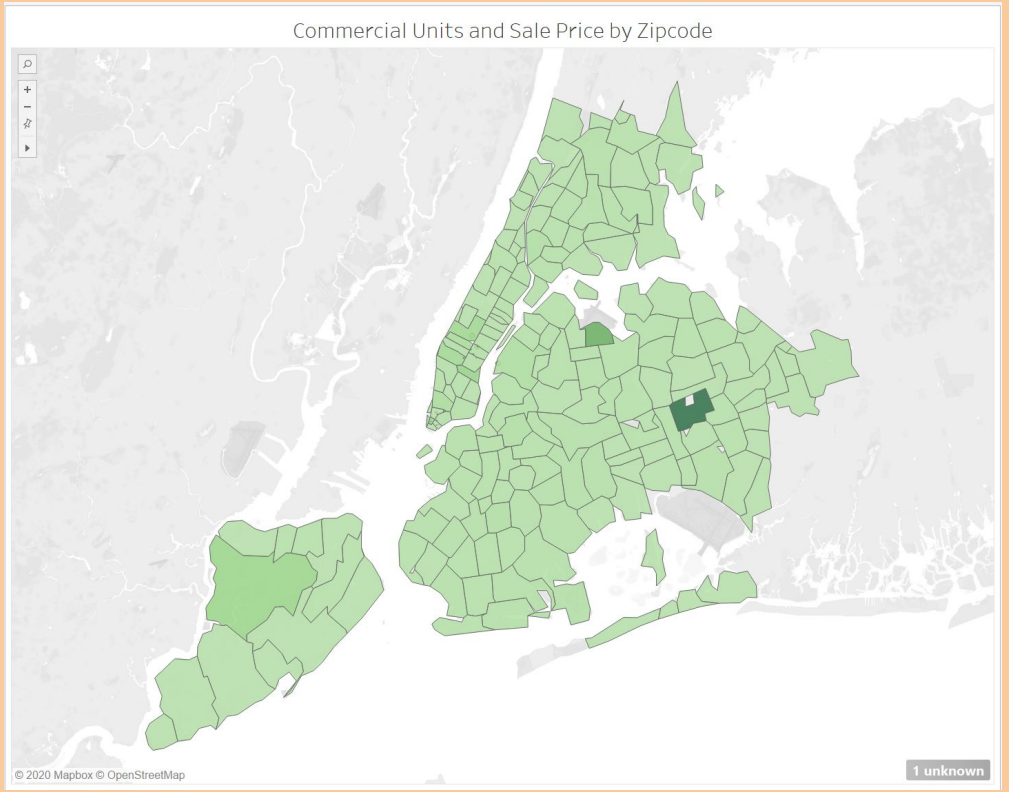
- Most Units found in Staten Island was in Richmond County (10314) with 1,392 units.



Commercial Units and Sale Price

- Most Units found in one zip code is 11432 in Queens (2,304 units).

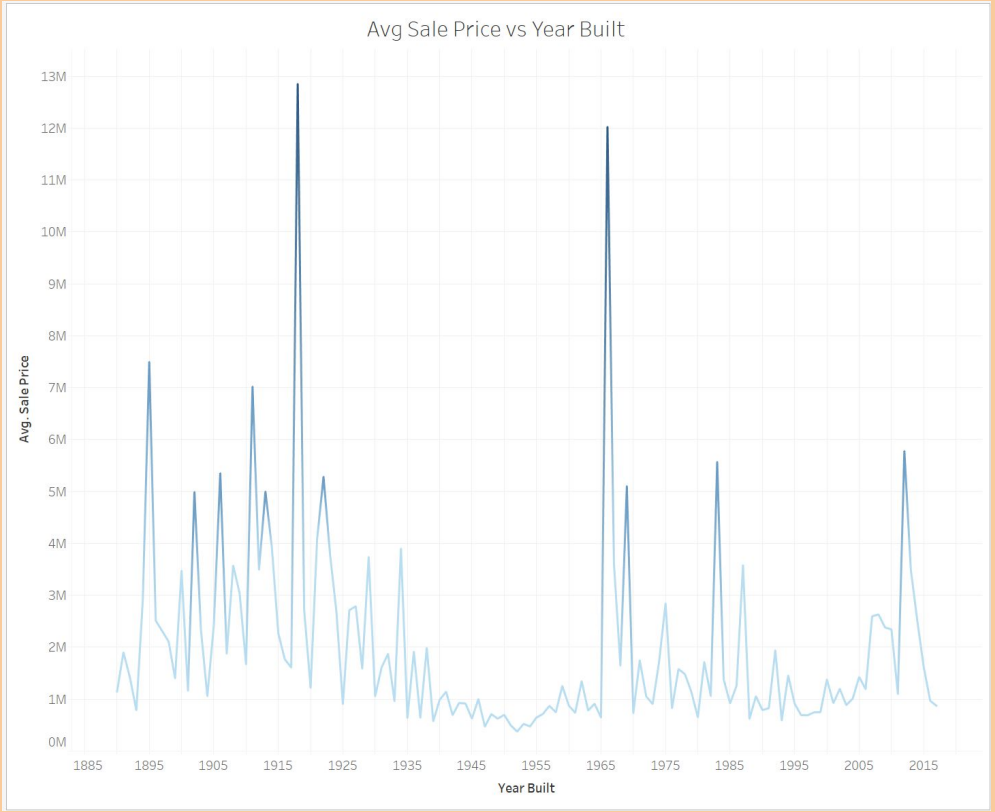
- Over 3000 Nursing Facilities and Student housing.



Avg Sale Price vs Year Built

- The boom in the 1920's that led to the great crash of 1929. Beginning of the great depression.

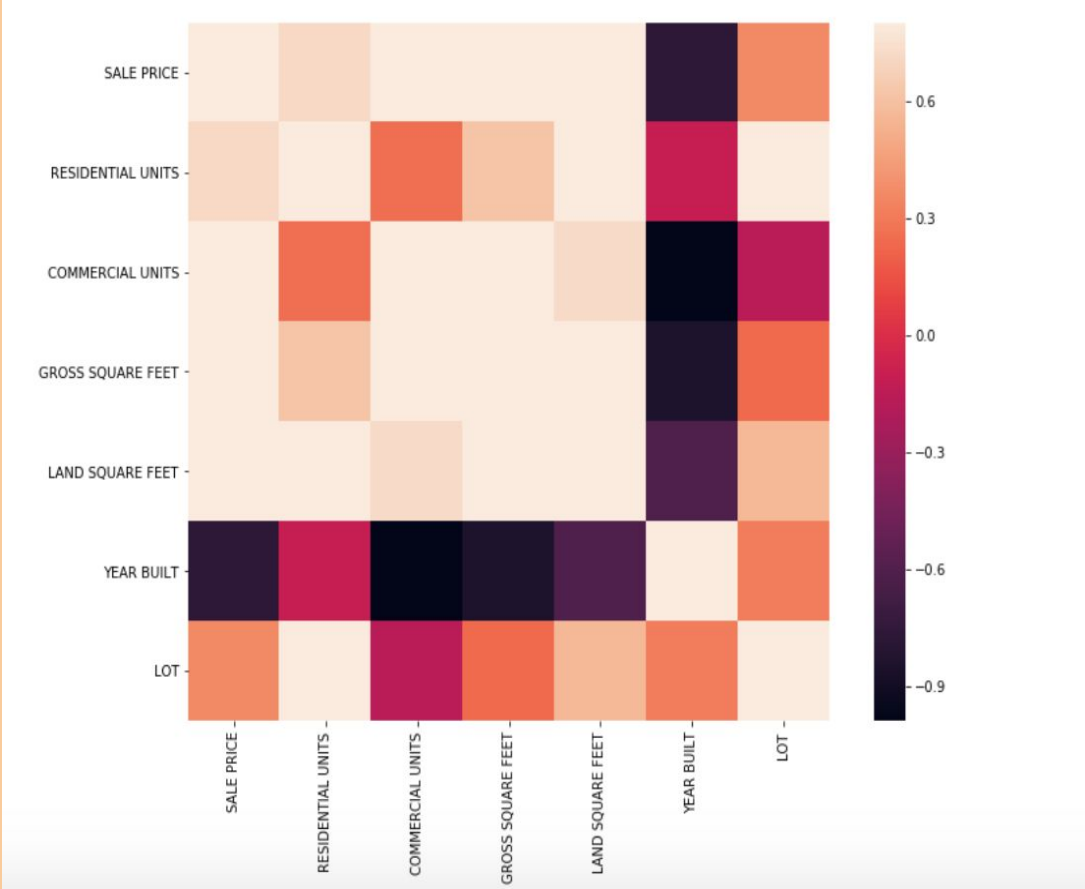
- From 1960 to **1970**, inflation rose from 1.4% to 6.5% (a 5.1% increase), which caused the median house price to almost double in the space of less than 10 years.



Sale Price Correlation

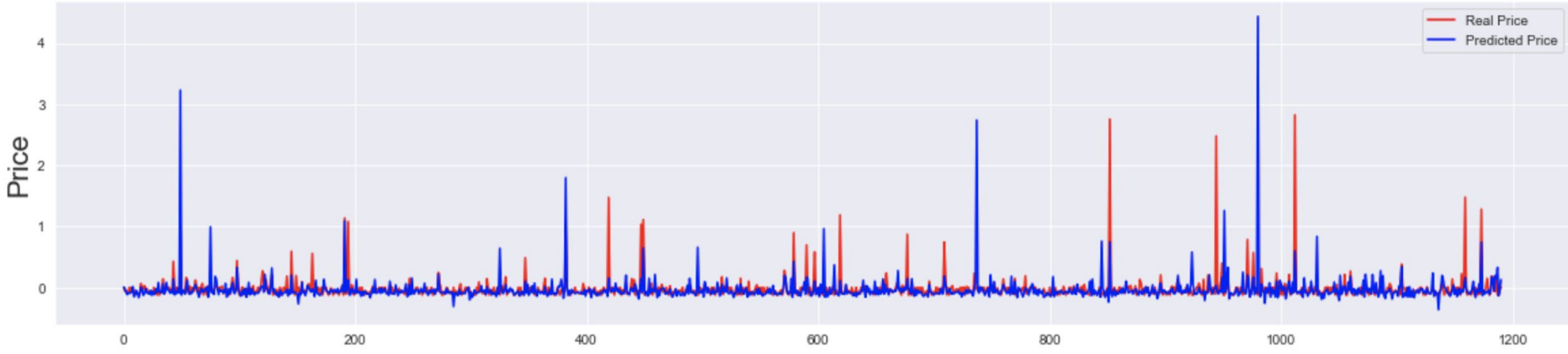
- Negative correlation with Year Built

- 80-90% correlation between Sale Price and Commercial Units & Gross Square ft & Land Square ft

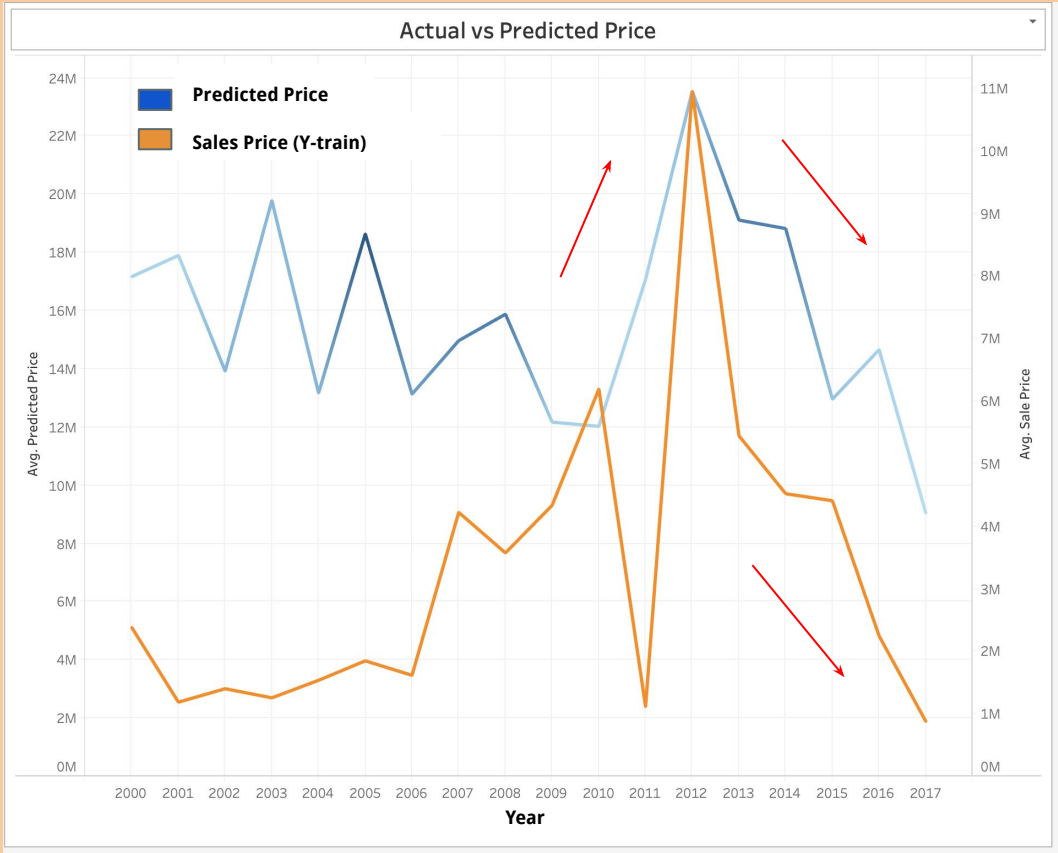


Machine Learning: SciKit-Learn and Keras

- After identifying the top coefficients from the correlation matrix, we extracted X_train and y_train data for the model prediction
- Data was scaled using Scaler (Scikit learn) to fit data into the model
- Training data was fed into Keras Linear Regression Model at Epoch = 150 for best accuracy
- Prediction price and actual price was plotted and initial model predictions looked promising



Actual Sales Price Vs Predicted Price



- We were able to visualize the total actual vs prediction price
- Interesting trends that we see is that predicted price followed the same trend of actual price overall in time
- How accurate was our prediction??

```
In [88]: pred_train= prediction
         print(np.sqrt(mean_squared_error(y_test, pred_train)))

         0.6617310633838279
```

- Mean squared error = 0.6617

Future Analysis and ML Implications

- This analysis allowed us to explore specific coefficients, their relationships and how they affect housing prices to be further explored in future predictions
- Next we can explore what other initiatives in real estate are and how other companies use coefficients in their predictions
- Help homeowners get the best price

The end