

# NYC REAL ESTATE Sale Price Prediction



How to predict real estate sale prices  
with key metrics

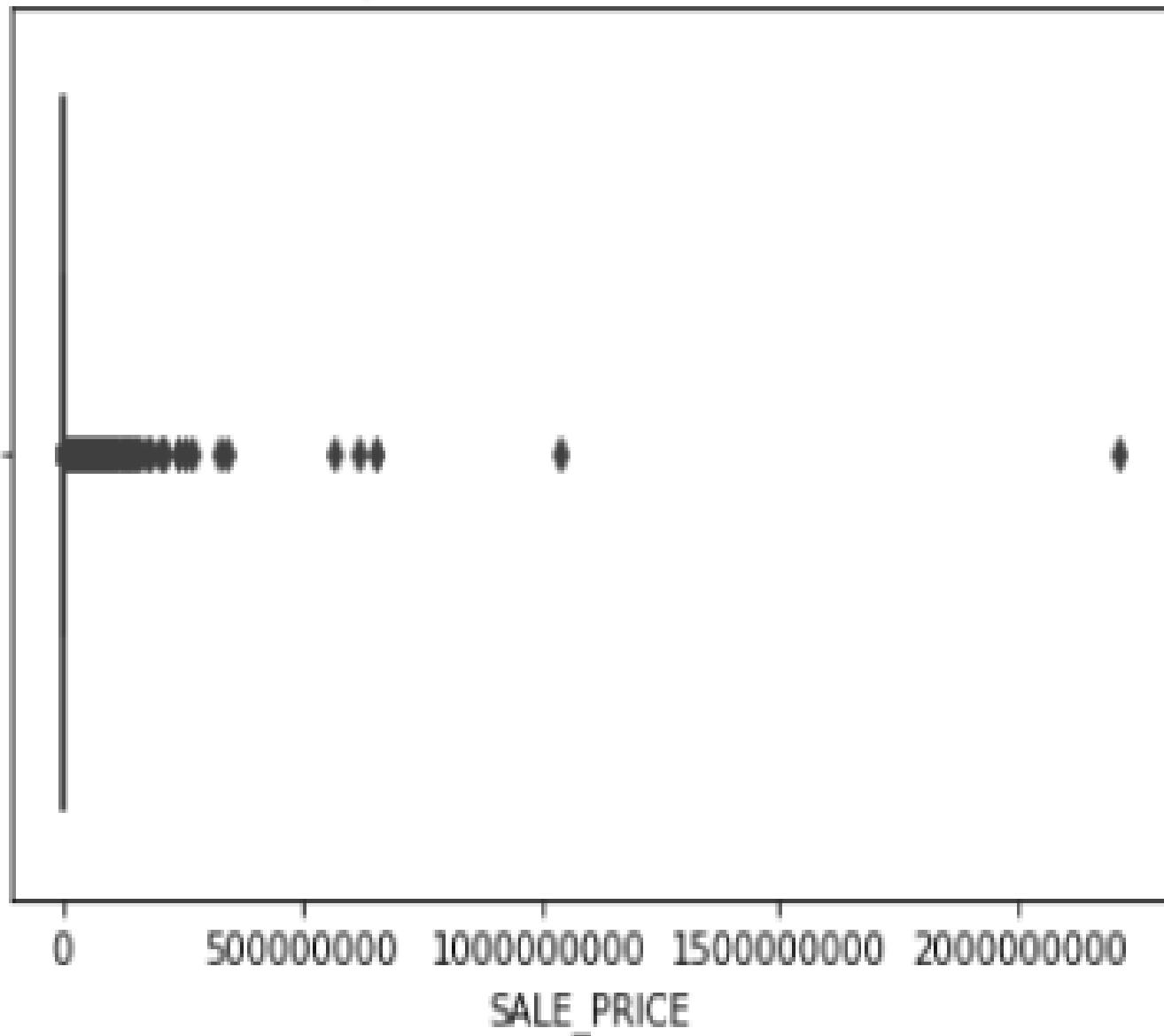


# Using Data to predict prices

- ▶ This Data Set consists of 22 columns of Data.
- ▶ 12-month period from September 2016 to September 2017
- ▶ Five Boroughs and neighborhood
- ▶ Metrics (Age, Land Size and property type)

Unnamed: 0	BOROUGH	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AT PRESENT	BLOCK	LOT	BUILDING CLASS AT PRESENT	ADDRESS	ZIP CODE	...	LAND SQUARE FEET	GROSS SQUARE FEET	YEAR BUILT	CLASS AT TIME OF SALE	BUILDING CLASS AT TIME OF SALE	SALE PRICE	SALE DATE	SALE_PRICE	LAND_SQUARE_FEET	GROSS_SQUARE_FEET	
0	4	1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	392	6	C2	153 AVENUE B	10009	...	1633	6440	1900	2	C2	6625000	2017-07-19 00:00:00	6625000	1633	6440
3	7	1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	402	21	C4	154 EAST 7TH STREET	10009	...	2272	6794	1913	2	C4	3936272	2016-09-23 00:00:00	3936272	2272	6794
4	8	1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	404	55	C2	301 EAST 10TH STREET	10009	...	2369	4615	1900	2	C2	8000000	2016-11-17 00:00:00	8000000	2369	4615

Boxplot of SALE PRICE in USD



## Original Sale Price

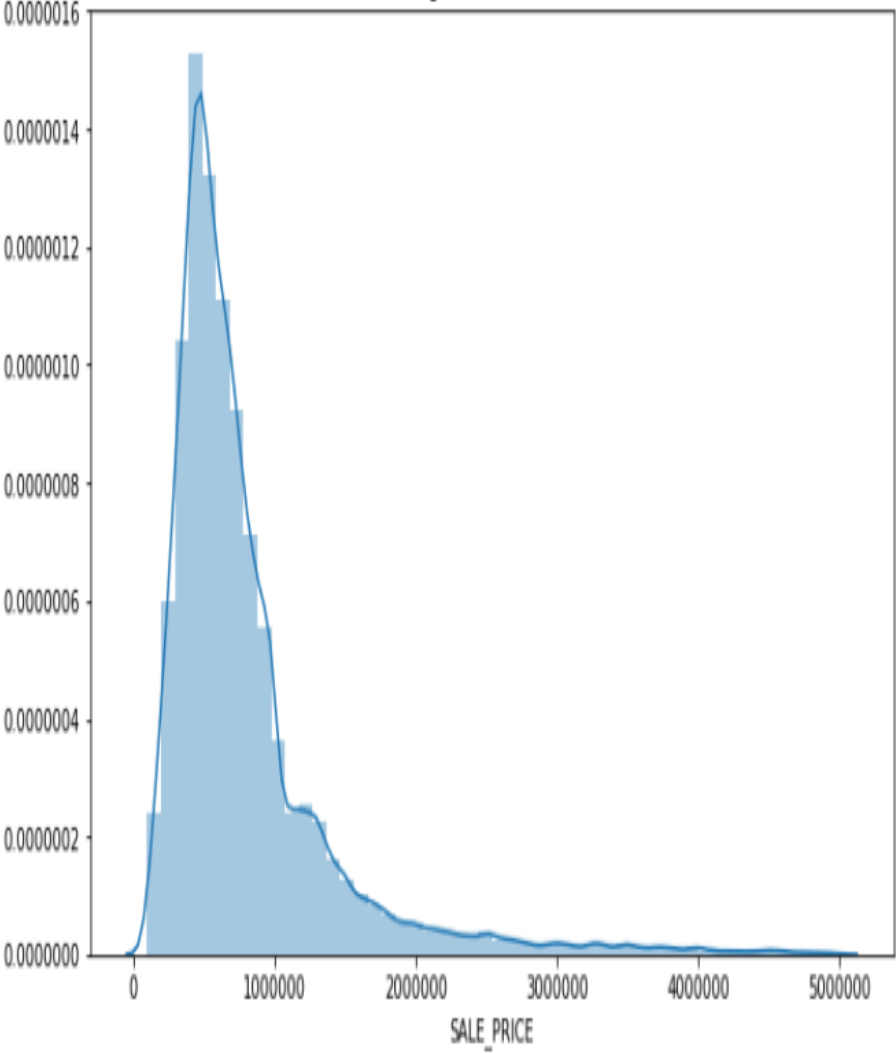
- ▶ Ranged from 0 to 20million
- ▶ Null values
- ▶ Missing values
- ▶ Outliers



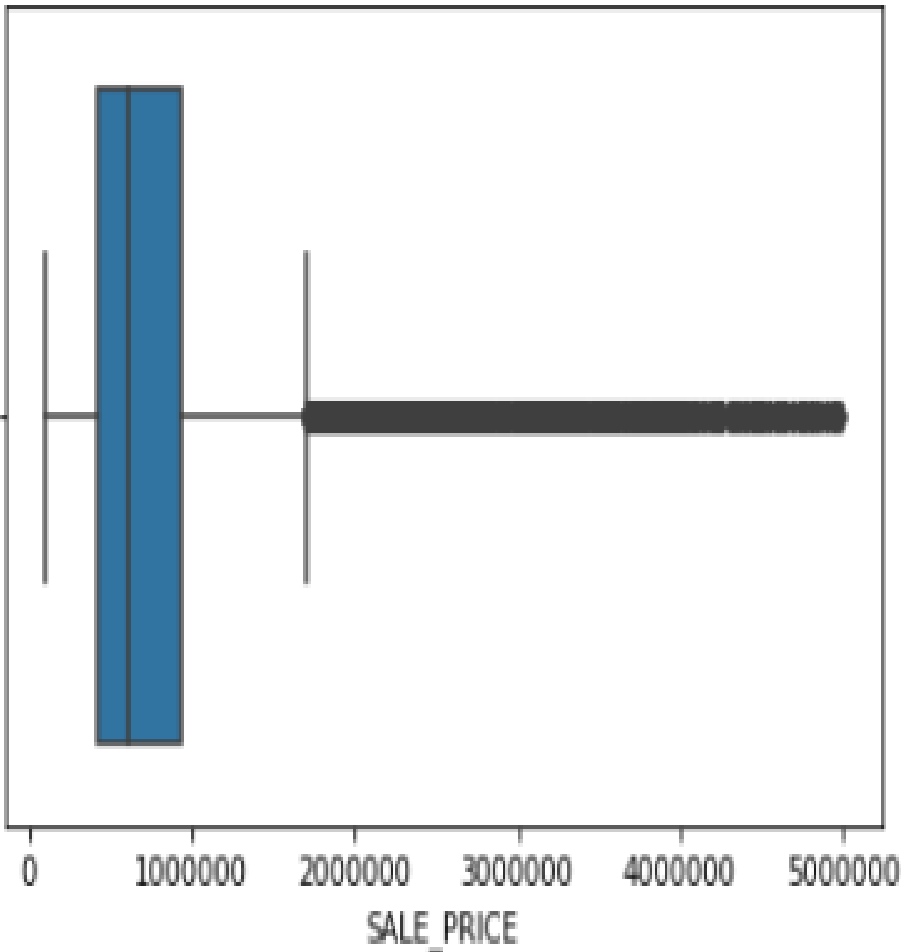
# Adjusted Sale Price

- \$100,000 to 5 million dollars
- Mean \$790,000

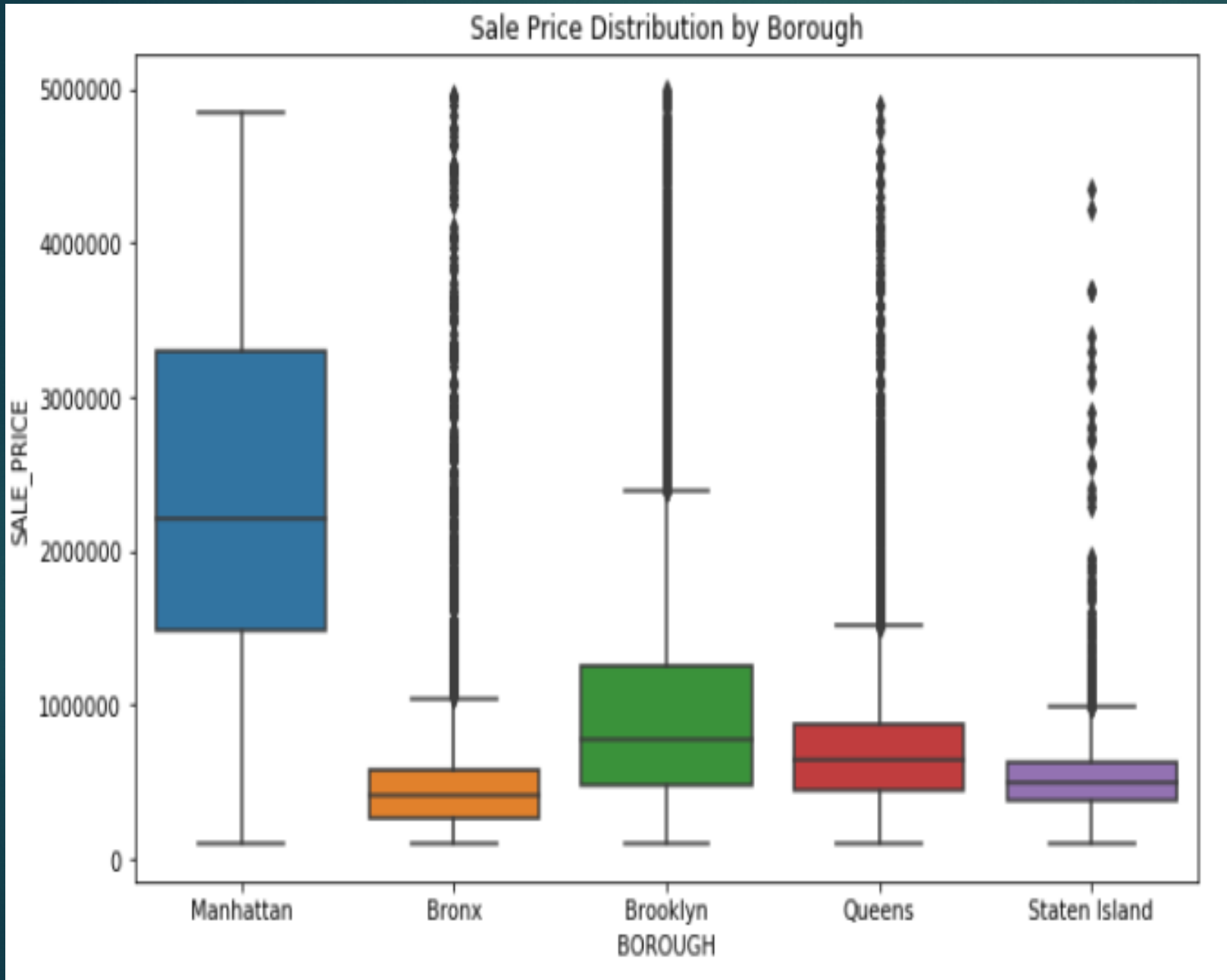
Histogram of SALE PRICE



Boxplot of SALE PRICE in USD

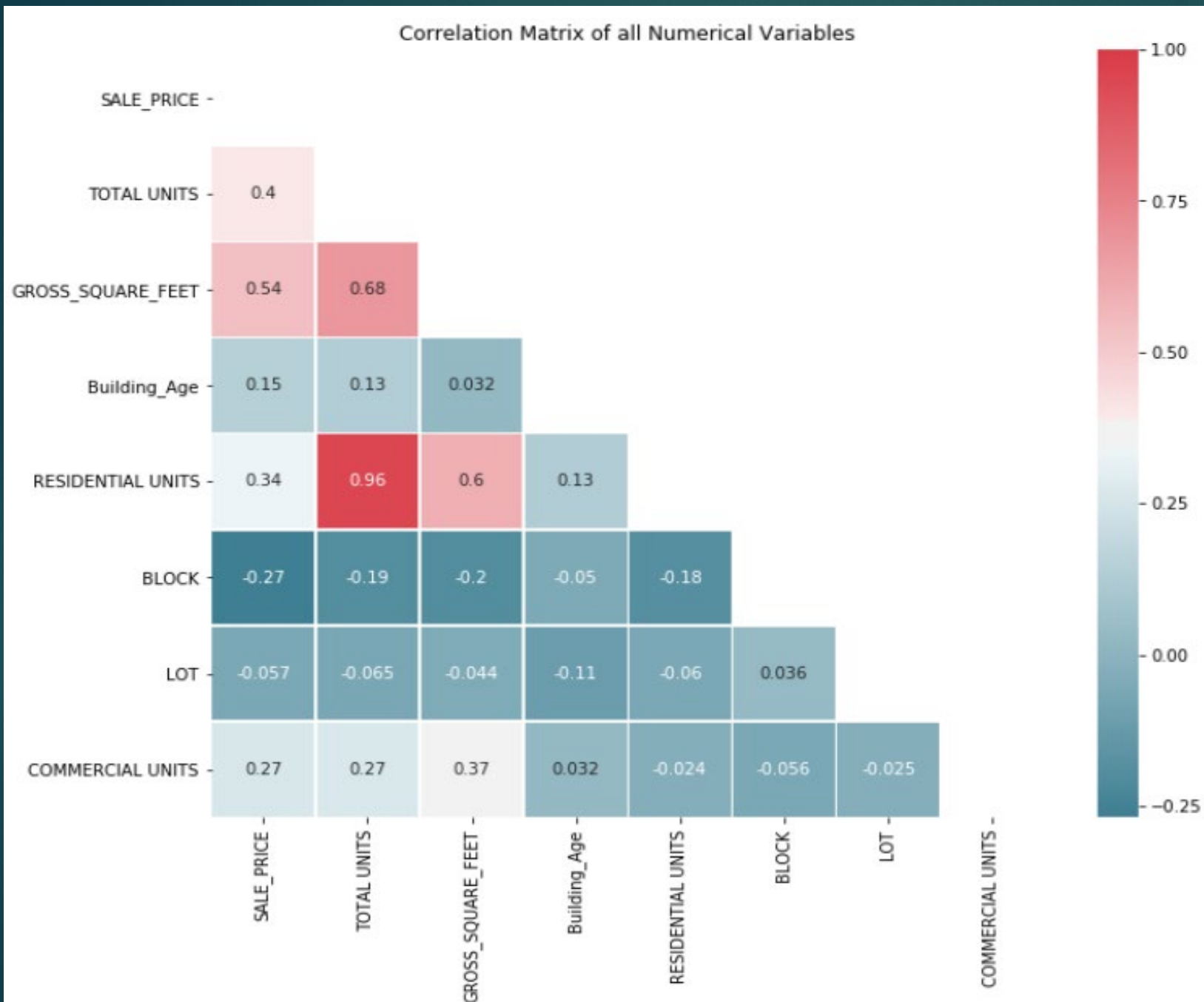






# Locations And Sale Price

- Manhattan Mean: 2.1 Million
- Bronx Mean: \$325,000
- Brooklyn Mean:\$789,000
- Queens Mean: \$645,000
- Staten Island Mean: \$410,000



## Strongest Correlations

- Total Units
- Gross Square feet
- Building Age

# Finalizing The data

- HOT ENCODING (NEIGHBORHOOD, BOROUGH CODE)
- REMOVING VALUES
- FEATURE ENGINEERING (GROSS SQUARE FEET, TOTAL UNITS)

Drop	Keep
Block	Gross Square Feet
Lot	Building Age
Zip Code	Borough Code
Residential Units	Total Units
Commercial units	Neighborhood
Land Square Feet	
Building Class Category	



### Predicting Sale Price initial results

	Random Forest	Ridge	Lasso	Linear
Root Mean Square Error	\$509,087	\$524,223	\$524,223	\$524,223
Cross Validation	0.37	0.30	0.30	0.30

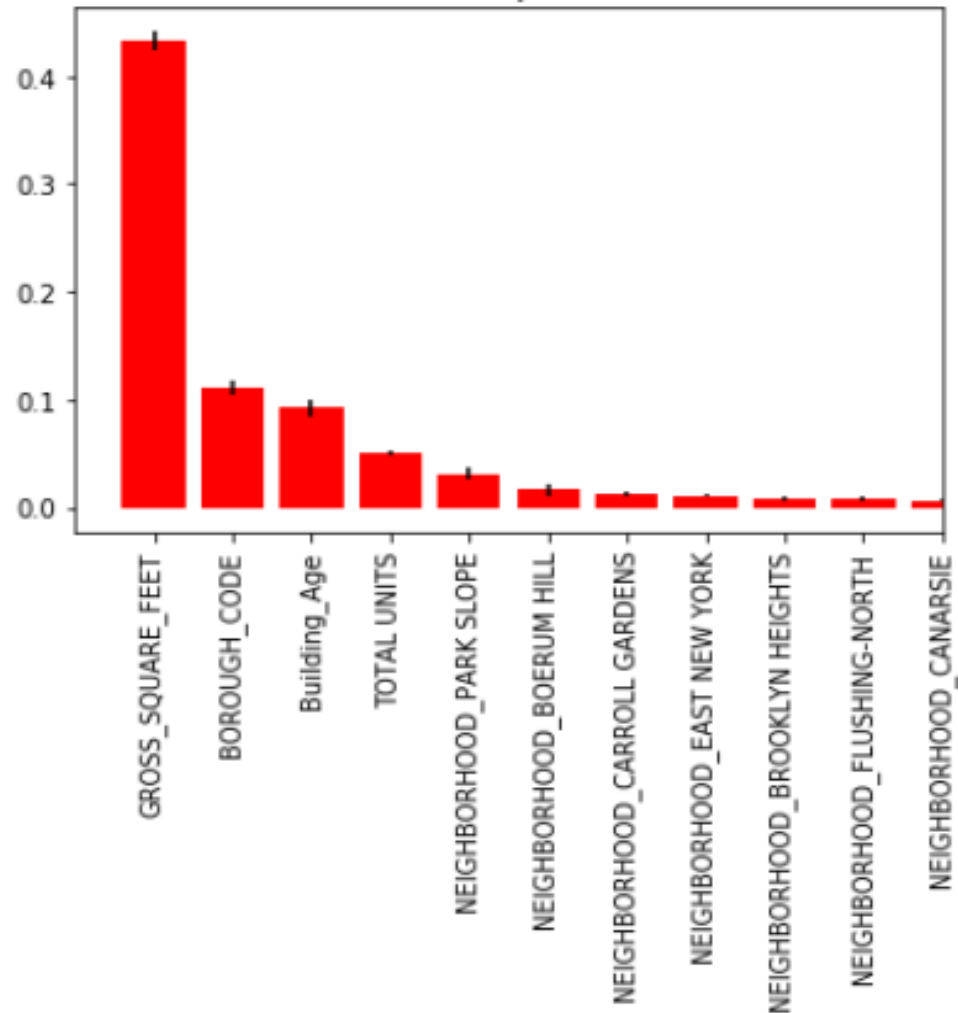
### Predicting Sale Price final results

	Random Forest	Ridge	Lasso	Linear
Root Mean Square Error	\$422,375	\$387,686	\$388,002	\$389,299
Cross Validation	0.57	0.56	0.56	0.62

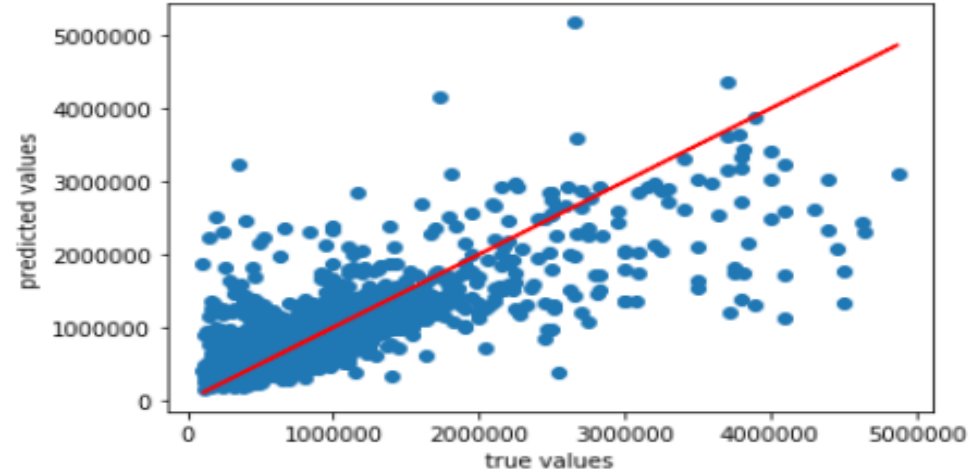


# Putting things together

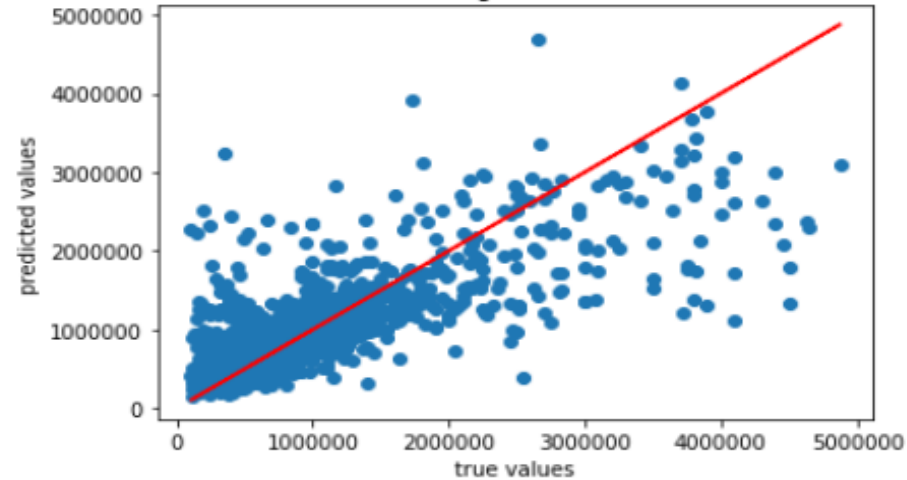
Feature importances



Lasso



Ridge Model



# Who will benefit from this project

This project is targeted towards following parties in the real estate industry.

- ▶ Real Estate investors
- ▶ Real Estate Developers
- ▶ Real Estate Agents