JUNE 17, 2024

NYC REAL ESTATE CASE STUDY

AD 571 ASSIGNMENT 4

BOSTON UNIVERSITY

PROFESSOR: GREG PAGE
STUDENT: PRABU JEYABALAN
Business Analytics Foundations

CONTENTS

EXECUTIVE SUMMARY

This report provides an in-depth analysis of residential real estate sales in Bushwick, starting from 2009, using a time series model. The data was cleaned to include only valid sale prices and gross square footage greater than zero, ensuring robustness in the historical context for predictions. The time series plot reveals distinct trends, with initial sales fluctuating around a lower average value, followed by mid-term peaks and troughs, and a sharp spike in recent quarters. This recent increase significantly impacts the overall trend, setting the stage for future forecasts and providing actionable insights into market trends.

The ETS model forecasts substantial growth for the next eight quarters, although performance metrics indicate significant challenges in capturing precise sales amounts, resulting in high uncertainty and broad prediction intervals. In contrast, the multiple regression model, which incorporates time and seasonality offers more conservative estimates with narrower prediction intervals indicating higher confidence. This model effectively captures seasonality and time trends providing more stable and confident predictions while identifying the most and least useful predictors of sale amounts.

Considering the findings from the analysis, establishing a real estate company in Bushwick at this juncture would be unwise. The market exhibits high volatility, notable prediction inaccuracies, and inflated property values all of which pose substantial risks. Furthermore, economic uncertainties and timing issues add complexity to the investment environment. Although Bushwick has growth potential present conditions suggest that the risks surpass the potential benefits. It would be wiser to consider other markets or wait for more stable circumstances before investing in the Bushwick real estate sector to ensure a safer and more profitable venture.

TIME SERIES MODEL

Analysis of the Residential Real Estate Sales in Bushwick Starting from 2009 and Forecast

The data used in this analysis was cleaned of anomalies, ensuring that only valid sale prices (greater than zero) and gross square footage (greater than zero) were included. The analysis begins from 2009 after the real estate market crash providing a robust historical context for the predictions. The time series plot of total residential real estate sales in Bushwick reveals distinct trends over time. By leveraging clean post-crash data and robust descriptive analytics this predictive analysis aims to provide actionable insights into future market trends. Initially, sales fluctuate around a lower average value with some mid-term peaks and troughs indicating periodic increases and decreases. Towards the end of the series, there is a sharp spike highlighting a significant surge in recent quarters (*Appendix 1*). This recent increase greatly impacts the overall trend and sets the stage for future forecasts.

Predictive analytics stems from descriptive analytics, where understanding historical data leads to forecasting future trends. The ETS model (ETS(M,N,M)) used here focuses on extrapolating historical data into a new model (*Appendix 2*). The forecast for the next eight quarters shows substantial growth with point forecasts for 2022 Q1 and 2023 Q4 reaching $238,482,453 and $995,329,226, respectively. The prediction intervals are broad, especially in 2022 Q4 and 2023 Q4, where the upper bounds exceed $2 billion. These wide intervals reflect the high level of uncertainty in the forecasts, largely due to the sharp recent surge in sales.

Model performance metrics indicate some significant errors, suggesting challenges in capturing the precise sales amounts. The RMSE is 67,733,639, and the MAE is 27,551,549, while other metrics include ME (9,227,454), MPE (-4.483184), MAPE (30.59027), MASE (0.7858091), and ACF1 (0.03066711). These values imply that, although the model captures the general trend,

the sharp increase towards the end introduces considerable deviations (*Appendix 4*). The forecast predicts continued high sales with substantial uncertainty, emphasizing the need for potential adjustments to the model.

MULTIPLE REGRESSION MODEL

The multiple regression model which incorporates time and seasonality, forecasts lower fit values compared to the ETS model for most quarters and offers a more conservative estimate (*Appendix 4*). It predicts $125,579,963 for Q1 2022, while the ETS model predicts $238,482,453. Additionally, the multiple regression model provides narrower prediction intervals indicating higher confidence in the predictions, such as an interval of [66,930,237, 184,229,689] for Q1 2022. In contrast, the ETS model's wide prediction intervals especially in later quarters suggest significant uncertainty with intervals like -56,674,816 to 1,888,725,608 for Q4 2022. The ETS model also shows a substantial expected increase in Q4 2022 and Q4 2023 which the multiple regression model does not reflect.

Given the need for stable and reliable forecasts, the multiple regression model is recommended for this dataset. It effectively captures seasonality and time trends, providing more stable and confident predictions. While the ETS model highlights potential dramatic changes and recent trends, its high uncertainty and wide prediction intervals make it less suitable for precise forecasting in this context. Therefore, the multiple regression model is the best choice to provide actionable insights into future market trends in Bushwick.

MULTIPLE REGRESSION MODEL TO DETERMINE RESIDENTIAL PROPERTY SALES

This expands on the previous multiple regression model by analyzing other factors that contribute to and influence these projections. This aims to verify and gauge the effect of redundant independent variables in the multiple regression model. The factors focused on are Sale Date, Year Built, Building Type, Gross Square Feet, and Number of Units (*Appendix 6*).

MOST AND LEAST USEFUL PREDICTORS OF THE AMOUNT OF SALES

Most Useful Predictors

The most useful predictors are determined by their high significance and low p-values, including sale date (extremely significant with a p-value < 0.0000000000000002), year built (significant with a p-value of 0.003014), and several specific building types such as converted dwellings or rooming house (highly significant with a p-value of 0.000671), elevator apt; fireproof with stores (extremely significant with a p-value < 0.0000000000000002), elevator apt; fireproof without stores (p < 0.0000000000000002), five to six families (significant with a p-value of 0.002719), over six families without stores (extremely significant with a p-value of 0.00000001568482493), walk-up apt. over six families with stores (extremely significant with a p-value < 0.000000002840125), and walk-up cooperative (extremely significant with a p-value < 0.00000052754221968). Additionally, gross square feet and number of residential units are both extremely significant with p-values < 0.0000000000000002.

Least Useful Predictors

On the other hand, several predictors were found to be the least useful due to their low significance and high p-values. These include building types such as condo; resid. unit of 1-3 unit

bldg-orig class 1 (p-0.849573), condo; residential unit in 2-10 unit bldg. (p-0.542931), condo; residential unit in elevator bldg. (p-0.534336), elevator apt; artists in residence (p-0.720717), and elevator apt; miscellaneous (p-0.286500). Other less useful predictors include elevator apt; semi-fireproof with stores (marginally significant with a p-0.086130), four families (marginally significant with a p-0.066048), miscellaneous one family (p-0.831666), miscellaneous two family (p-0.352435), one family attached or semi-detached (p-0.722516), one story - permanent living quarter (p-0.876924), three families (p-0.120209), two family brick (p-0.407154), two family converted from one family (p-0.431446), two family frame (p-0.356446), and two stories - detached sm or mid (p-0.955581). These predictors, with their high p-values, indicate minimal impact on the sale amount and suggest redundancy or irrelevance in the model.

IDENTIFYING REDUNDANT INDEPENDENT VARIABLES

Predictors with high p-values indicate redundancy and multicollinearity. These redundant variables can be removed or combined to improve the model's performance and interpretability. Specifically, building types with high p-values mentioned above show minimal impact on the sale amount and suggest redundancy or irrelevance in the model. By calculating and analyzing p-values, we could identify and address redundancy in the regression model.

IDENTIFYING THE BIGGEST BARGAINS AND MOST OVERPRICED PROPERTIES

Based on the residuals from the multiple regression model, it is evident that properties were the biggest bargains (negative residuals) and which were the most overpriced (positive residuals). Residuals are the differences between the actual sale price and the predicted sale price. A large negative residual indicates a property was sold for much less than predicted (a bargain), while a large positive residual indicates a property was sold for much more than predicted (overpriced).

Biggest Bargain

The property at 1103 Gates Avenue is identified as the biggest bargain (*Appendix 7*) using residual analysis. This property, sold in 2010, was built in 1928 and falls under the description "Over Six Families Without Stores." It has a gross square footage of 56,940 and includes 83 residential units. The sale price was $4,323,727, with a residual of -35,711,040. Additional details about the property include a lot size of 22,002 square feet, building dimensions of 200 ft x 80 ft, four stories, and the last alteration was made in 1989. The substantial negative residual suggests that this property was sold for significantly less than its predicted value, making it the biggest bargain according to the model.

Most Overpriced

Conversely, the property at 189 Cooper Street is identified as the most overpriced (*Appendix 8*) by assessing the residuals in regression. This property was sold in 2019, was built in 2016, and is categorized as "Elevator Apt; Fireproof Without Stores." It has a gross square footage of 10,932 and includes 13 residential units, with an additional 7 commercial units. The sale price was $2,605,714, with a residual of 11,845,738. The total square footage is 11,051. The significant positive residual indicates that this property was sold for much more than its predicted value, making it the most overpriced property according to the model.

Accounting For Disparities

The disparities between the sale prices and the model's predictions can be attributed to various factors. For the biggest bargain at 1103 Gates Avenue, market conditions at the time of sale (post-recession in 2010) likely contributed to lower property prices. Additionally, the property's physical condition, last altered in 1989 (*RealtyHop, n.d.*), might not have been fully

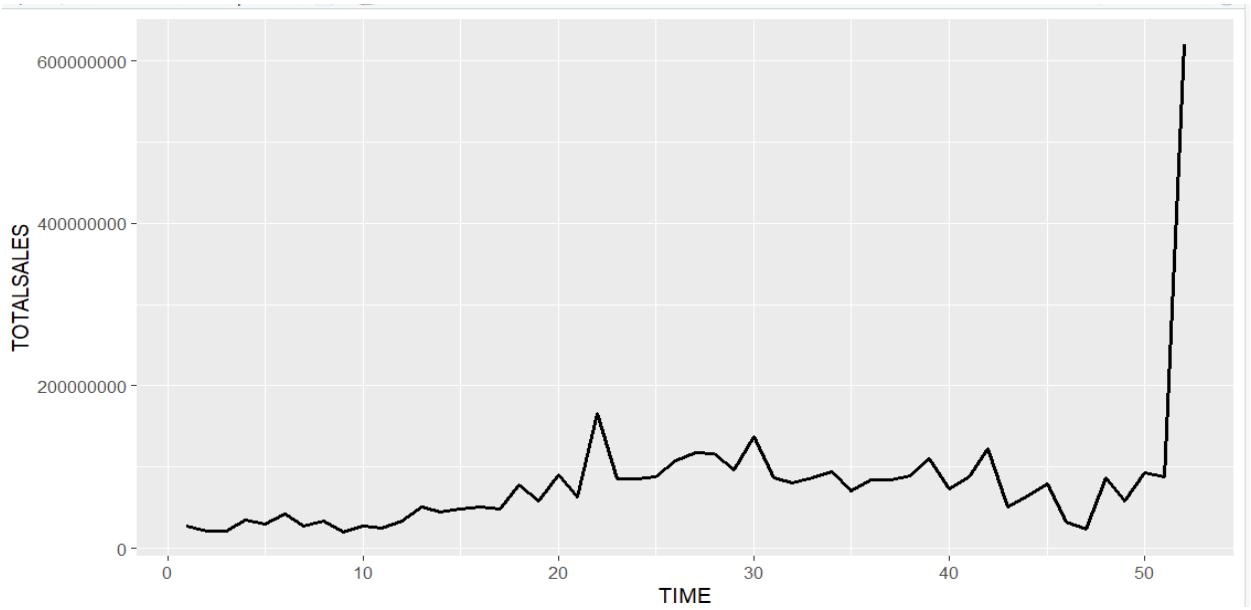
reflected in the model, potentially leading to a lower sale price. Location factors, such as neighborhood changes or proximity to amenities, could also have negatively impacted the sale price. In contrast, the most overpriced property at 189 Cooper Street benefits from recent development, as it was built in 2016 and may feature modern amenities and design elements not captured by the model. The buyer's willingness to pay a premium for such attributes, combined with market trends favoring sellers in 2019, could have driven up the price (*n.d., p. Propertyshark*). Furthermore, the presence of 7 commercial units in addition to 13 residential units enhances its value due to the potential for mixed-use revenue generation, justifying the higher sale price.

CONCLUSION

Based on the current analysis, starting a real estate company in Bushwick may not be advisable at this time. The market's high volatility, significant prediction errors, and overpriced properties present considerable risks. Additionally, economic uncertainties (*Morgan, 2023*) and timing challenges further complicate the investment landscape. While Bushwick shows potential for growth, the current conditions suggest that the risks outweigh the potential rewards. Therefore, it may be more prudent to explore other markets or wait for more stable conditions before investing in the Bushwick real estate startup.

APPENDICES

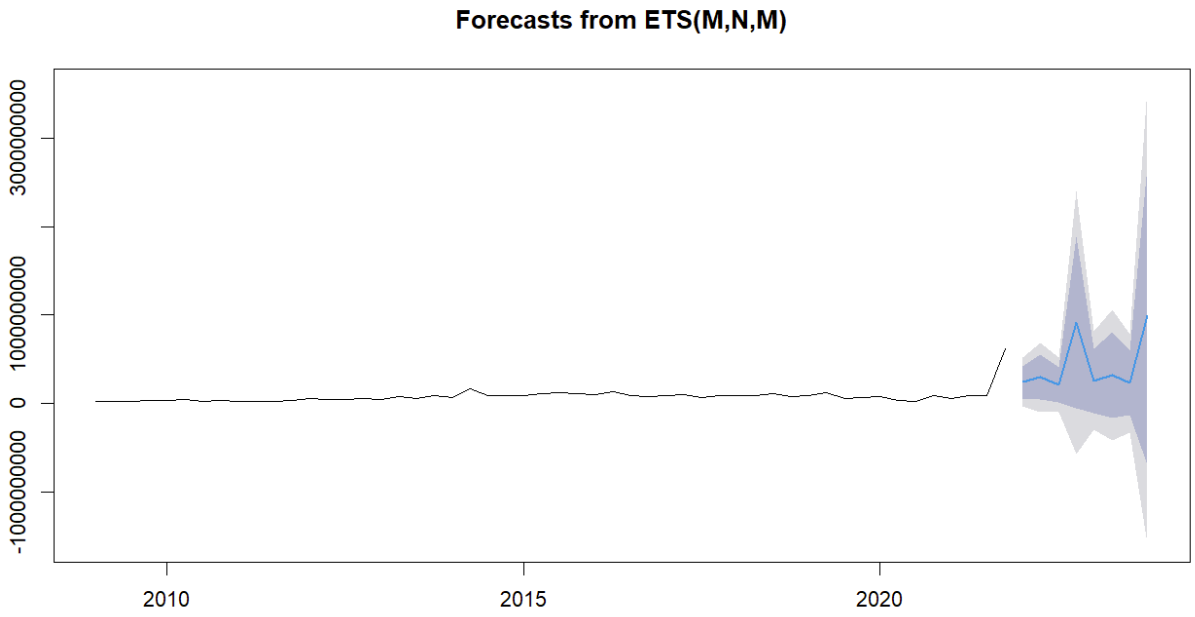
Appendix 1: ***Time series plot of total residential real estate sales in Bushwick:***



Appendix2: ***Time Series Forecast for the next 8 months in Bushwick:***

```
Forecast method: ETS(M,N,M)

Model Information:
ETS(M,N,M)

Call:
 ets(y = TIMESERIESSETUP, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.4731
    gamma = 0.5269

  Initial states:
    l = 28312006.8623
    s = 1.1258 0.7148 1.0847 1.0747

  sigma:  0.5893

     AIC     AICc      BIC
2022.269 2024.815 2035.928

Error measures:
                    ME      RMSE       MAE        MPE      MAPE      MASE        ACF1
Training set 9227454 67736369 27551549 -4.483184 30.59027 0.7858091 0.03066711

Forecasts:
        Point Forecast      Lo 80      Hi 80       Lo 95      Hi 95
2022 Q1      238482453   58369313  418595593   -36976823  513941730
2022 Q2      295343310   40866974  549819645   -93844661  684531280
2022 Q3      210262888    7734697  412791079   -99477251  520003028
2022 Q4      916025396  -56674816 1888725608  -571591205 2403641996
2023 Q1      259128794 -105272123  623529711  -298174317  816431904
2023 Q2      320912312 -161411504  803236128  -416738309 1058562933
2023 Q3      228466153 -137236905  594169210  -330828409  787760715
2023 Q4      995329226 -696449054 2687107506 -1592022377 3582680830
```

Appendix 3: ***Time series forecast Plot for the next 8 months in Bushwick:***



Appendix 4: ***Multiple Regression prediction for the next 8 quarters sales in Bushwick:***

	fit	lwr	upr
1	125579963	66930237	184229689
2	143439909	84790183	202089635
3	124782584	66132857	183432310
4	174060758	115411032	232710484
5	134451064	71751869	197150258
6	152311010	89611815	215010204
7	133653685	70954490	196352879
8	182931859	120232664	245631053

Appendix 5: ***Regression Model Output Summary:***

```
Call:
lm(formula = (TOTALSALES ~ TIME + QUARTER), data = Reg1)

Residuals:
       Min         1Q     Median         3Q        Max
-92768467 -29902224  -3863984  13937847 455167552

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8037874    27739843    0.290  0.77327
TIME          2217775      711278    3.118  0.00311 **
QUARTERQ2    15642171    30118236    0.519  0.60595
QUARTERQ3    -5232930    30143422   -0.174  0.86292
QUARTERQ4    41827469    30185352    1.386  0.17238
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 76770000 on 47 degrees of freedom
Multiple R-squared:  0.2198,    Adjusted R-squared:  0.1534
F-statistic: 3.311 on 4 and 47 DF,  p-value: 0.01805
```

Appendix 6: ***Multiple Regression Model Output Summary:***

```
Call:
lm(formula = SALE_PRICE ~ ., data = MultiRegPrep)

Residuals:
      Min        1Q    Median        3Q       Max
-35711040   -230072         0    230050  11845738

Coefficients:
                                                          Estimate        Std. Error t value             Pr(>|t|)
(Intercept)                                          -6812967.7472064  1258053.3841828  -5.415  0.00000006432252023 ***
SALE_DATE                                                  0.0033278        0.0001597  20.832 < 0.0000000000000002 ***
YEAR_BUILT                                              1114.4931522      514.3844804   2.167             0.030314 *
DESCRIPTIONCONDO; RESID.UNIT OF 1-3 UNIT BLDG-ORIG CLASS 1   186516.5461677   983349.8722757   0.190             0.849573
DESCRIPTIONCONDO; RESIDENTIAL UNIT IN 2-10 UNIT BLDG.       -475863.4116827   782113.6755478  -0.608             0.542931
DESCRIPTIONCONDO; RESIDENTIAL UNIT IN ELEVATOR BLDG.        -485642.0045368   781466.5858578  -0.621             0.534336
DESCRIPTIONCONVERTED DWELLINGS OR ROOMING HOUSE            -2947912.1352013   866102.4938707  -3.404             0.000671 ***
DESCRIPTIONELEVATOR APT; ARTISTS IN RESIDENCE                499980.4534378  1398456.1325570   0.358             0.720717
DESCRIPTIONELEVATOR APT; CONVERTED                        -27607049.3317714  1332644.0299052 -20.716 < 0.0000000000000002 ***
DESCRIPTIONELEVATOR APT; FIREPROOF WITH STORES             -11416725.2538275  1417609.4227378  -8.054  0.00000000000000102 ***
DESCRIPTIONELEVATOR APT; FIREPROOF WITHOUT STORES          -17141520.9358074   989782.4319556 -17.318 < 0.0000000000000002 ***
DESCRIPTIONELEVATOR APT; MISCELLANEOUS                     -1405144.1628919  1318197.2425445  -1.066             0.286500
DESCRIPTIONELEVATOR APT; SEMI-FIREPROOF WITH STORES          2271398.3549751  1323232.9417422   1.717             0.086130 .
DESCRIPTIONELEVATOR APT; SEMI-FIREPROOF WITHOUT STORES     -6060374.1931241  1016002.2569745  -5.965  0.00000002636020066 ***
DESCRIPTIONFIVE TO SIX FAMILIES                            -2286442.5101702   762281.7362776  -2.999             0.002719 **
DESCRIPTIONFOUR FAMILIES                                   -1414206.6968146   769198.1496030  -1.839             0.066048 .
DESCRIPTIONMISCELLANEOUS ONE FAMILY                         -163438.1479830   768837.3983526  -0.213             0.831666
DESCRIPTIONMISCELLANEOUS TWO FAMILY                         -711478.7672215   765053.1008894  -0.930             0.352435
DESCRIPTIONONE FAMILY ATTACHED OR SEMI-DETACHED             -276686.8543436   779136.1059509  -0.355             0.722516
DESCRIPTIONONE STORY - PERMANENT LIVING QUARTER              203985.3447392  1317066.8045139   0.155             0.876924
DESCRIPTIONOVER SIX FAMILIES WITHOUT STORES                -4342554.0724526   766650.6597302  -5.664  0.00000001568482493 ***
DESCRIPTIONTHREE FAMILIES                                  -1183011.0529994   761173.4974988  -1.554             0.120209
DESCRIPTIONTWO FAMILY BRICK                                 -631866.0530703   762211.6508760  -0.829             0.407154
DESCRIPTIONTWO FAMILY CONVERTED FROM ONE FAMILY             -600760.4517573   763557.7762480  -0.787             0.431446
DESCRIPTIONTWO FAMILY FRAME                                 -701856.4622807   761023.4124283  -0.922             0.356446
DESCRIPTIONTWO STORIES - DETACHED SM OR MID                   41892.2050659   806590.0492464   0.052             0.958581
DESCRIPTIONWALK-UP APT. OVER SIX FAMILIES WITH STORES       -5307408.0279007   795535.9846753  -6.671  0.00000000002840125 ***
DESCRIPTIONWALK-UP COOPERATIVE                              -4686578.4829905   932959.0564247  -5.023  0.00000052754221968 ***
GROSS_SQUARE_FEET                                             137.9756055       16.2202686   8.506 < 0.0000000000000002 ***
RESIDENTIAL_UNITS                                          444544.1268626    13104.2935753  33.924 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1075000 on 4479 degrees of freedom
Multiple R-squared:  0.9618,    Adjusted R-squared:  0.9616
F-statistic:  3892 on 29 and 4479 DF,  p-value: < 0.00000000000000022
```

Appendix 7: ***1103 Gates Avenue the Biggest Bargain Property in Bushwick:***

	SALE_DATE	YEAR_BUILT	DESCRIPTION	GROSS_SQUARE_FEET	RESIDENTIAL_UNITS	SALE_PRICE	ADDRESS	RESIDUAL
1	2010-11-01	1928	OVER SIX FAMILIES WITHOUT STORES	56940	83	4323727	1103 GATES AVENUE	-35711040.3
2	2015-03-11	2015	ELEVATOR APT; FIREPROOF WITH STORES	58617	42	100000	810 FLUSHING AVENUE	-15420075.1
3	2010-06-30	1924	ELEVATOR APT; FIREPROOF WITHOUT STORES	49913	54	98000	1028 BUSHWICK AVENUE	-13236355.9
4	2010-01-08	1931	WALK-UP APT. OVER SIX FAMILIES WITH STORES	21180	36	10	889 BROADWAY	-13160272.2
5	2019-06-20	1932	OVER SIX FAMILIES WITHOUT STORES	19000	32	1987332	116-120 GROVE STREET	-11051894.6
6	2016-10-06	1931	WALK-UP COOPERATIVE	16000	28	52208	890 FLUSHING AVENUE	-10165995.5
7	2012-06-27	2012	OVER SIX FAMILIES WITHOUT STORES	29564	24	150000	422 MELROSE STREET	-10146722.4

Appendix 8: ***189 Cooper Street the Most Overpriced Property in Bushwick:***

	SALE_DATE	YEAR_BUILT	DESCRIPTION	GROSS_SQUARE_FEET	RESIDENTIAL_UNITS	SALE_PRICE	ADDRESS	RESIDUAL
1	2019-05-01	2016	ELEVATOR APT; FIREPROOF WITHOUT STORES	10932	13	2605714	189 COOPER STREET	11845738.4
2	2009-10-07	1910	FIVE TO SIX FAMILIES	4725	6	10431074	1275 DECATUR STREET	9906682.3
3	2010-04-14	1910	FIVE TO SIX FAMILIES	4125	6	10298710	26 SCHAEFER STREET	9802762.6
4	2010-01-14	1931	FOUR FAMILIES	3192	4	9890573	11 TROUTMAN STREET	9542681.6
5	2021-12-02	2018	ELEVATOR APT; FIREPROOF WITH STORES	388023	443	249205000	54 NOLL STREET	9262673.5
6	2010-06-15	1907	FIVE TO SIX FAMILIES	5025	6	9197418	248 CORNELIA STREET	8562809.9
7	2016-07-19	1931	OVER SIX FAMILIES WITHOUT STORES	6435	12	9250000	263 LINDEN STREET	7142928.8

REFERENCES

1. RealtyHop. (n.d.). *1103 Gates Avenue*. Realtyhop.com. https://www.realtyhop.com/building/1103-gates-avenue-brooklyn-ny-11221

2. https://www.propertyshark.com/mason/Property/151796616/189-Cooper-St-Brooklyn-NY-11207/

3. Morgan, J. P. (2023, December 22). 2024 economic outlook: Insights & trends. Jpmorgan.com; J.P. Morgan. https://www.jpmorgan.com/insights/outlook/economic-outlook/economic-trends