Professor Stephen Coggeshall

Unsupervised Fraud Model
On NYC Data

Team 1

Kalvin Tran | Aviroop Ghosal | Mohamad Ganji
Zhaojun (Pauline) Liu | Mark Orgel | Vu Duong
Jiaqi Zhu | Saif Rehman | Zhihan (Han) Li

Feb/16/2019

Table of Contents

Executive Summary

In this paper, we will outline the process we used to develop and employ an unsupervised fraud detection model that can identify likely candidates of fraud within the NYC Department of Finance Property Valuation and Assessment database. Leveraging a weighted average of two fraud models, one developed as a heuristic function of z-scores, and the other using an autoencoder, we were able to mitigate any weaknesses introduced by either of the models used in isolation, further improving the expected accuracy of our fraud candidate predictions.

In our approach, we derived summary information from the data through exploratory data analysis and descriptive statistics. We assessed the distributions and frequency of numerical and categorical variables. We then algorithmically populated field values for incomplete records and created new variables that would give us sufficient information to distinguish between the anomalies and normal records.

From there, we normalized the variables and applied a dimensionality reduction technique to further derive a subset that contains the most relevant information about the data. Our first fraud score for each record was computed based on a summation of these values as a heuristic function of their z-scores. For our second model, we applied a neural network model that learned the representation of data, reconstructed it, and based on the reconstruction error observed, we computed our second fraud score. After rank ordering records based on fraud scores obtained from each of these techniques, we combined these two fraud score ranks to obtain our final score, based on which we identified the top fraud records.

The combined values produced a fraud score distribution with the vast majority of properties being ranked with a score of under 500, and a very small subset receiving scores over 500 - 2200, which was in line with our expectations for the dataset. We believe properties in the top end of this range are likely candidates for property fraud and recommend they be individually investigated by property experts within the NYC Department of Finance.

Description of Data

Our analysis was performed using the City of New York ("NYC") Property Valuation and Assessment Data available on the NYC Open Data Portal ("Assessor's Roll"). This data was first made publicly available on September 2, 2011 and has been updated on an as-needed basis. For the purposes of this analysis, the Assessor's Roll data reviewed was last updated on September 10, 2018.

The primary use of the Assessor's Roll is to assist in the determination of property tax bills for every new fiscal year; it acts as a record to help in determining the amount of property tax due, the entity responsible for paying it, and where to mail it. This is crucial to many local agencies as property taxes can make up a significant portion of their respective revenues. For example, in the fiscal year 2017, which ended on June 30, 2017, approximately 45 percent of all city tax revenues for NYC came from property taxes.

The data reviewed in this report can be found on Open Data. It is the condensed version of the Assessor's Value Roll which includes 32 different fields for every property within NYC. Of the 32 fields, 14 fields are numeric and 18 are categorical; as of the last update (September 10, 2018), the Assessor's Roll identified 1,070,994 properties.

An aggregate summary of the Assessor's Roll is presented in Figure 1 below.

FIGURE 1 Field Summary

Field	Field Type	Records	% Populated	Unique Values	Records with value zero	% with value zero
RECORD	Categorical	1,070,994	100.00%	1,070,994	0	0.00%
BBLE	Categorical	1,070,994	100.00%	1,070,994	0	0.00%
B	Categorical	1,070,994	100.00%	5	0	0.00%
BLOCK	Categorical	1,070,994	100.00%	13,984	0	0.00%
LOT	Categorical	1,070,994	100.00%	6,366	0	0.00%
EASEMENT	Categorical	4,636	0.43%	13	0	0.00%
OWNER	Categorical	1,039,249	97.04%	863,348	0	0.00%
BLDGCL	Categorical	1,070,994	100.00%	200	0	0.00%
TAXCLASS	Categorical	1,070,994	100.00%	11	0	0.00%
LTFRONT	Numeric	1,070,994	100.00%	1,297	169,108	15.79%
LTDEPTH	Numeric	1,070,994	100.00%	1,370	170,128	15.89%
EXT	Categorical	354,305	33.08%	4	0	0.00%
STORIES	Numeric	1,014,730	94.75%	112	0	0.00%
FULLVAL	Numeric	1,070,994	100.00%	109,324	13,007	1.21%

Field	Field Type	Records	% Populated	Unique Values	Records with value zero	% with value zero
AVLAND	Numeric	1,070,994	100.00%	70,921	13,009	1.21%
AVTOT	Numeric	1,070,994	100.00%	112,914	13,007	1.21%
EXLAND	Numeric	1,070,994	100.00%	33,419	13,007	1.21%
EXTOT	Numeric	1,070,994	100.00%	64,255	432,572	40.39%
EXCD1	Categorical	638,488	59.62%	130	0	0.00%
STADDR	Categorical	1,070,318	99.94%	839,281	0	0.00%
ZIP	Categorical	1,041,104	97.21%	197	0	0.00%
EXMPTCL	Categorical	15,579	1.45%	15	0	0.00%
BLDFRONT	Numeric	1,070,994	100.00%	612	228,815	21.36%
BLDDEPTH	Numeric	1,070,994	100.00%	621	22,853	2.13%
AVLAND2	Numeric	282,726	26.40%	58,592	0	0.00%
AVTOT2	Numeric	282,732	26.40%	111,361	0	0.00%
EXLAND2	Numeric	87,449	8.17%	44,196	0	0.00%
EXTOT2	Numeric	130,828	12.22%	48,349	0	0.00%
EXCD2	Categorical	92,948	8.68%	61	0	0.00%
PERIOD	Categorical	1,070,994	100.00%	1	0	0.00%
YEAR	Categorical	1,070,994	100.00%	1	0	0.00%
VALTYPE	Categorical	1,070,994	100.00%	1	0	0.00%

For the purposes of this report, we have mainly utilized numeric variables in our analysis, so summary statistics for the numeric variables are presented below in Figure 2.

FIGURE 2 Numeric Variables Statistics Summary

Field	Records	Unique Values	Mean	Median	Mode	Min	Max	Std. Dev.
LTFRONT	1,070,994	1,297	36.64	25	0	0	9,999	74.03
LTDEPTH	1,070,994	1,370	88.86	100	100	0	9,999	76.40
STORIES	1,014,730	112	5.01	2	2	1	119	8.37
FULLVAL	1,070,994	109,324	874,264	447,000	0	0	6.1x10^9	11,582,431
AVLAND	1,070,994	70,921	85,067	13,678	0	0	2.6x10^9	4,057,260
AVTOT	1,070,994	112,914	227,238	25,340	0	0	4.6x10^9	6,877,529
EXLAND	1,070,994	33,419	36,423	1,620	0	0	2.6x10^9	3,981,575
EXTOT	1,070,994	64,255	91,186	1,620	0	0	4.6x10^9	6,508,402
BLDFRONT	1,070,994	612	23.04	20	0	0	7,575	35.58
BLDDEPTH	1,070,994	621	39.92	39	0	0	9,393	42.71
AVLAND2	282,726	58,592	246,235	20,145	2,408	3	2.3x10^9	6,178,962
AVTOT2	282,732	111,361	713,911	79,963	750	3	4.5x10^9	11,652,529
EXLAND2	87,449	44,196	351,235	3,048	2,090	1	2.3x10^9	10,802,213
EXTOT2	130,828	48,349	656,768	37,062	2,090	7	4.5x10^9	16,072,510

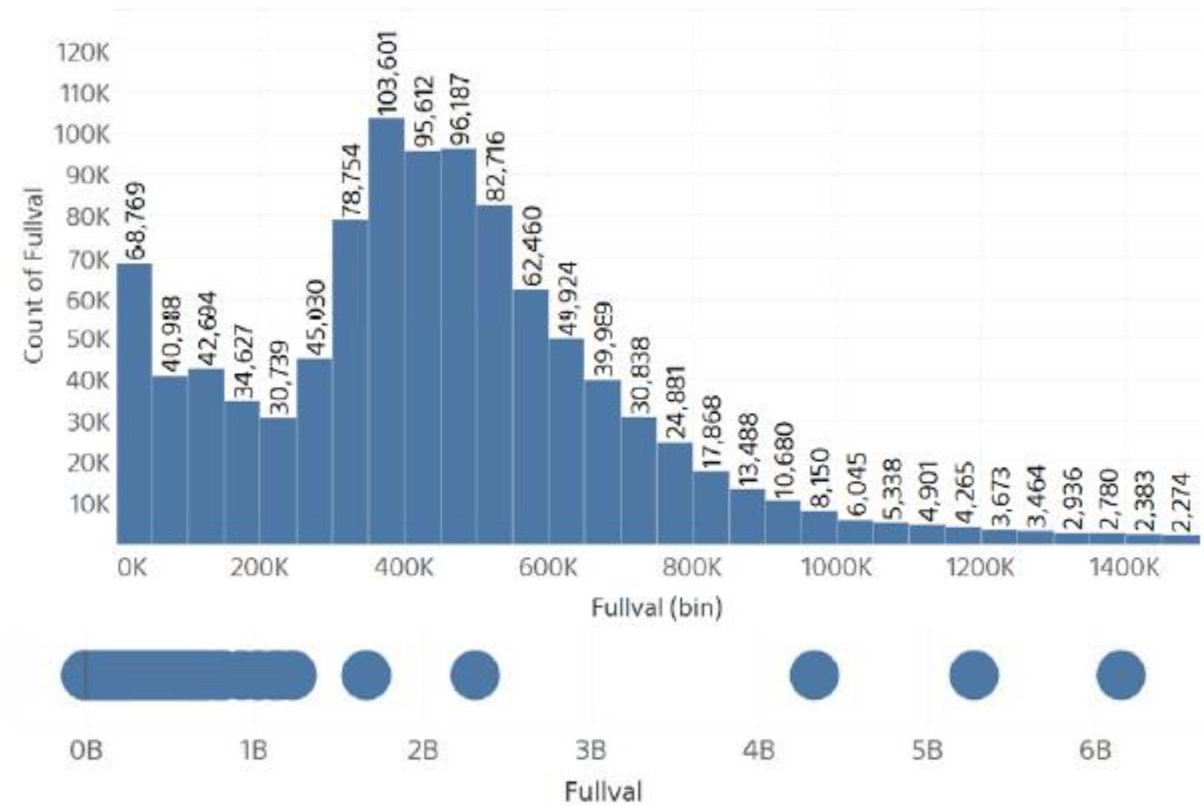
Summary Distributions of Most Important Variables

Below, we have identified several fields that were critical for our analysis: FULLVAL, AVLAND, and AVTOT. These are the primary variables that we used to identify outliers and predicted fraud. Details about these values and their distributions within the dataset are as follows:

FULLVAL

The FULLVAL field represents the total market value of a property. It is a numeric field which appears to have a bimodal distribution and is heavily right skewed, as shown in Figure 3 below. The boxplot at the bottom of Figure 3 supports the existence of a heavy right skew of the data distribution due to a few outliers in the billions. There is an upper whisker at 818,734 which results in 135,798 outliers.

FIGURE 3 FULLVAL Histogram & Boxplot

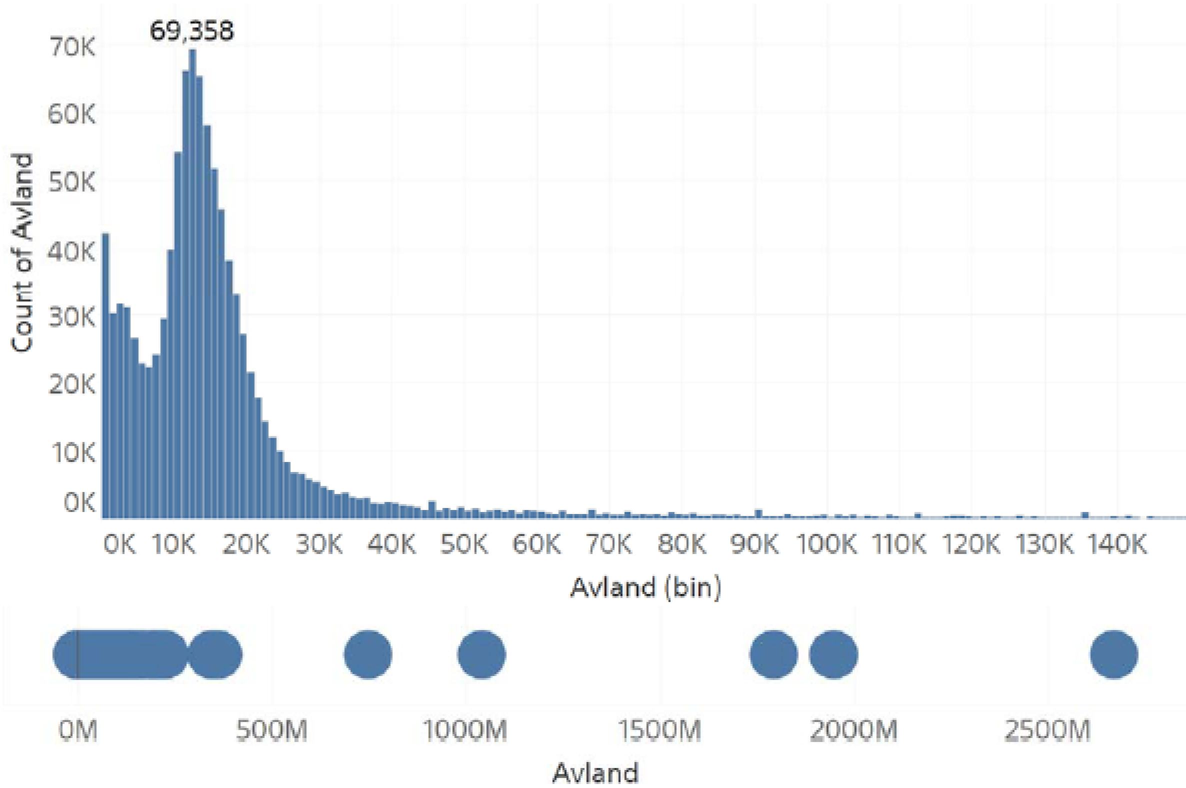


AVLAND

AVLAND is the current year's total market value of the land. It is a numeric field and appears to have a bimodal distribution and has a heavy right skew as shown in Figure 4 below.

The boxplot at the bottom of Figure 4 supports the existence of a heavy skew to the right due to a few outliers in the billions. There is an upper whisker at 192,998 which results in 39,315 outliers.

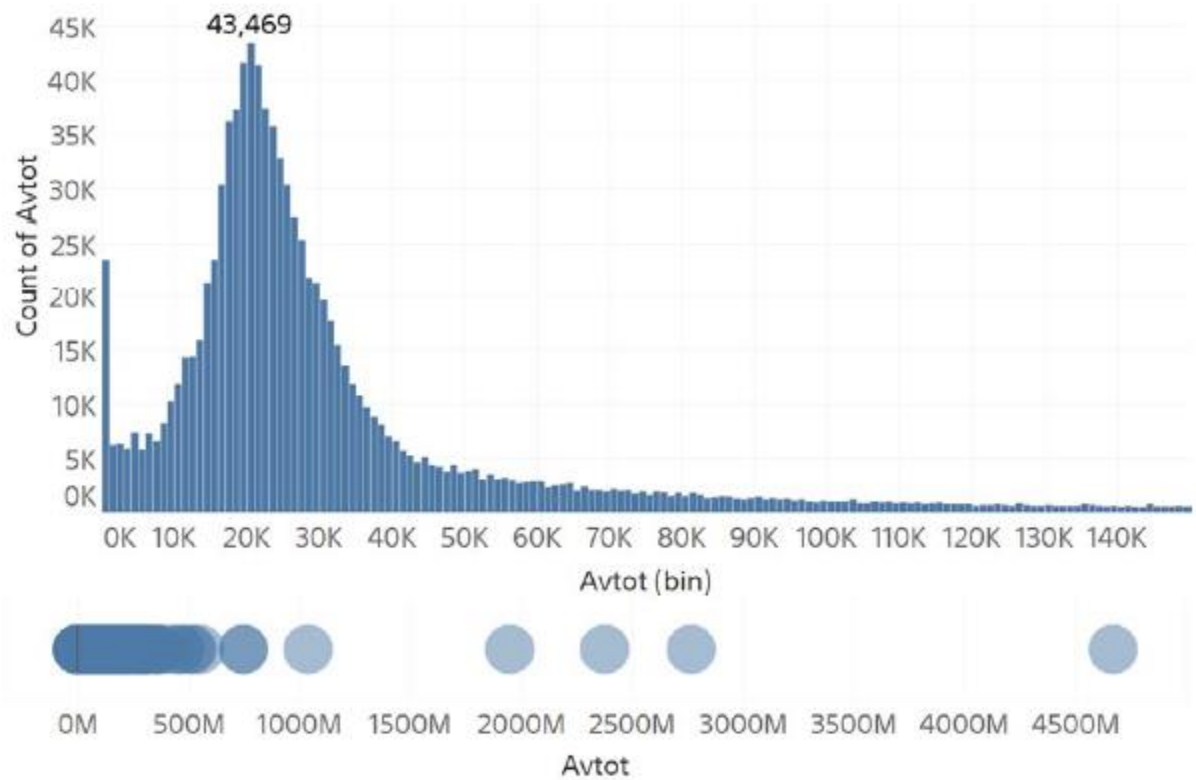
FIGURE 4 AVLAND Histogram & Boxplot



AVTOT

AVTOT is the current year's total market value of the property. It is a numeric field and appears to have a bimodal distribution and has a heavy right skew as shown in Figure 5 below. The boxplot at the bottom of Figure 5 supports the existence of a heavy skew to the right due to a few outliers in the billions. There is an upper whisker at 335,025 which results in 62,127 outliers.

FIGURE 5 AVTOT Histogram & Boxplot



Please refer to Appendix 1 for the complete Data Quality Report ("DQR") on the NYC Property Valuation and Assessment Data. The DQR presents comprehensive detailing of each of the individual variables as well as their distributions and more information about the dataset in general.

Data Cleaning

The NYC Property and Valuation data included missing values for many fields. Since the focus of the analysis is primarily related nine (9) fields (FULLVAL, AVLAND, AVTOT, STORIES, LTFRONT, LTDEPTH, BLDFRONT, BLDDEPTH, AND ZIP), extensive effort was put forth into determining reasonable estimates for missing values. Up to 251,019 properties had missing data in one of the nine (9) fields. Figure 6 shows the count of missing data for each of these fields.

FIGURE 6 Missing Data

Field	Count for Missing Data
FULLVAL	13,007
AVLAND	13,009
AVTOT	13,007
STORIES	56,264
LTFRONT	169,108
LTDEPTH	170,128
BLDFRONT	228,815
BLDDEPTH	228,853
ZIP	29,890

We utilized both Python and Visual FoxPro (VFP) in order to create new variables/fields and fill in missing values or replace zeros (0) that misrepresented the property. For example, LTDEPTH and LTFRONT did not make sense with values at zero (0) and had to be replaced with an estimate. The strategy was to group properties by their nearest neighbors of the same building class and then fill in any missing values. If there were still zeros or missing values, the groupings would escalate to the next level going from locality at street name to ZIP code then borough, and finally citywide.

Using VFP, the street number, direction, and name were parsed out into separate fields, situs_num, situs_dir, and situs_name, respectively. From there, filling in zeros and missing values followed the process below:

1. Fill in ZIP Code based on the median of groupings by Borough, Block, and Situs_name
2. Fill in ZIP Code based on the median of groupings by Borough and Block
3. Fill in FULLVAL, AVTOT, AVLAND, LTFRONT, LTDEPTH, BLDFRONT, and BLDDEPTH based on the following (every subsequent level is an escalation to fill in any remaining zeros or missing values):
    a. Groupings by Borough, Block, Situs_name, and BLDGCL
        i. For BLDGCL = A2, A3, A4, A5, and A5 us the mean
        ii. For all other BLDGCL, use the median
    b. Median by groupings by ZIP and BLDCL

c. Median by groupings by Borough and BLDGL
d. Median by groupings by BLDGCL

After applying the process to fill in missing values, we were able to reduce the number of properties with missing values in one of the nine (9) fields to zero (0).

Variable Creation

To assess the market value of property, we considered the land and property value, and total units of building as our most important variables. New features such as building area, lot area, building volume are created as these values would highlight the dimensions of the property. Then we computed 9 ratios of value per two-dimensional or three-dimensional quantity measure to assess the value of the property per unit quantity measure. Ratios are aggregated over groups such as ZIP, TAXCLASS, borough to find out the average values of these ratios across each location, district and TAXCLASS. At last, we standardize these ratios based on the obtained 5 scale factors and further form 45 variables.

FIGURE 7 Variable Creation

No.	Variable	No.	Variable	No.	Variable
1	V_1 = FULLVAL	21	$r_2/<r_2>_{g1}$	41	$r_6/<r_6>_{g1}$
2	V_2 = AVLAND	22	$r_2/<r_2>_{g2}$	42	$r_6/<r_6>_{g2}$
3	V_3 = AVTOT	23	$r_2/<r_2>_{g3}$	43	$r_6/<r_6>_{g3}$
4	S_1=LFTFRONT*LTDEPTH	24	$r_2/<r_2>_{g4}$	44	$r_6/<r_6>_{g4}$
5	S_2=BLDFRONT*BLDDEPTH	25	$r_2/<r_2>_{g5}$	45	$r_6/<r_6>_{g5}$
6	$S_3=S_2$*STORIES	26	$r_3/<r_3>_{g1}$	46	$r_7/<r_7>_{g1}$
7	$r_1=V_1/S_1$	27	$r_3/<r_3>_{g2}$	47	$r_7/<r_7>_{g2}$
8	$r_2=V_1/S_2$	28	$r_3/<r_3>_{g3}$	48	$r_7/<r_7>_{g3}$
9	$r_3=V_1/S_3$	29	$r_3/<r_3>_{g4}$	49	$r_7/<r_7>_{g4}$
10	$r_4=V_2/S_1$	30	$r_3/<r_3>_{g5}$	50	$r_7/<r_7>_{g5}$
11	$r_5=V_2/S_2$	31	$r_4/<r_4>_{g1}$	51	$r_8/<r_8>_{g1}$
12	$r_6=V_2/S_3$	32	$r_4/<r_4>_{g2}$	52	$r_8/<r_8>_{g2}$
13	$r_7=V_3/S_1$	33	$r_4/<r_4>_{g3}$	53	$r_8/<r_8>_{g3}$

No.	Variable	No.	Variable	No.	Variable
14	$r_8=V_3/S_2$	34	$r_4/\langle r_4\rangle_{g4}$	54	$r_8/\langle r_8\rangle_{g4}$
15	$r_9=V_3/V_3$	35	$r_4/\langle r_4\rangle_{g5}$	55	$r_8/\langle r_8\rangle_{g5}$
16	$r_1/\langle r_1\rangle_{g1}$	36	$r_5/\langle r_5\rangle_{g1}$	56	$r_9/\langle r_9\rangle_{g1}$
17	$r_1/\langle r_1\rangle_{g2}$	37	$r_5/\langle r_5\rangle_{g2}$	57	$r_9/\langle r_9\rangle_{g2}$
18	$r_1/\langle r_1\rangle_{g3}$	38	$r_5/\langle r_5\rangle_{g3}$	58	$r_9/\langle r_9\rangle_{g3}$
19	$r_1/\langle r_1\rangle_{g4}$	39	$r_5/\langle r_5\rangle_{g4}$	59	$r_9/\langle r_9\rangle_{g4}$
20	$r_1/\langle r_1\rangle_{g5}$	40	$r_5/\langle r_5\rangle_{g5}$	60	$r_9/\langle r_9\rangle_{g5}$

Dimensionality Reduction

The main tool used for dimensionality reduction is Principal Components Analysis (“PCA”). PCA is a common dimensionality reduction algorithm that finds the dominant directions in a given set of data and rotates the coordinate system along these directions by creating new variables, called principal components (“PC’s”). These PC’s are orthogonal and ordered by the variance explained from the original data set. Put simply, each PC is a linear combination of the original variables so it can be thought of as a rotated axis from the original data set. When doing PCA, the PC’s are ordered by the variance that they explain so PC1 explains a greater amount of the variance of the dataset than PC2 and so on. Therefore, PCA is a very useful tool because one can select a subset of the PC’s that explain the majority of the variance/information in the dataset.

This is a critical step in the analysis as the expert variables that were created were many in number and were correlated to each other. Since PCA essentially creates a new coordinate system using a linear combination of the variables and is orthogonal, mathematically, the PC’s are considering the correlations of the different expert variables and representing them in a much more succinct manner. Therefore, there is a dual-benefit of using PCA as a means of dimensionality reduction as it also takes into account the correlations between the different variables.

Specifically, for this analysis, we performed PCA on the set of expert variables that were created in the last step. To best determine the number of PC’s to utilize, we performed PCA with 20 components initially so that we can generate a scree plot and check the eigenvalues of each of the PC’s to see which number of PC’s has the best balance of explaining the most variance and reducing the dimensionality considerably.

FIGURE 8 Cumulative explained variance

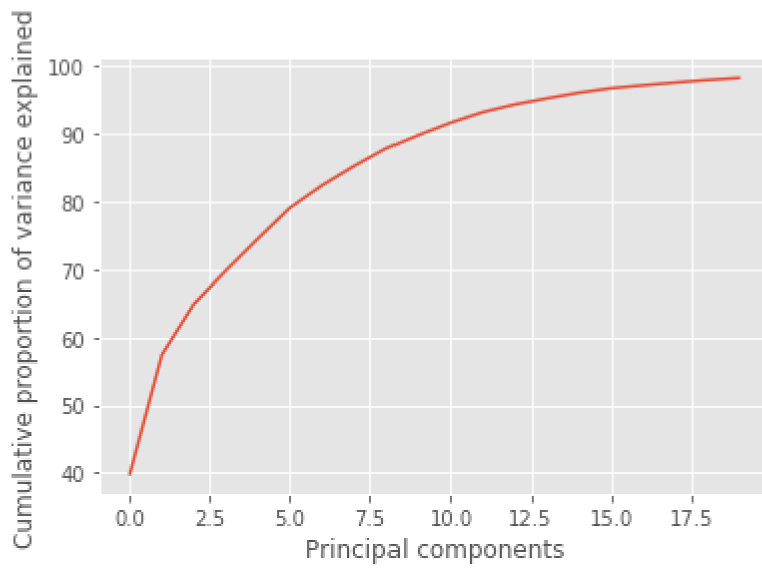
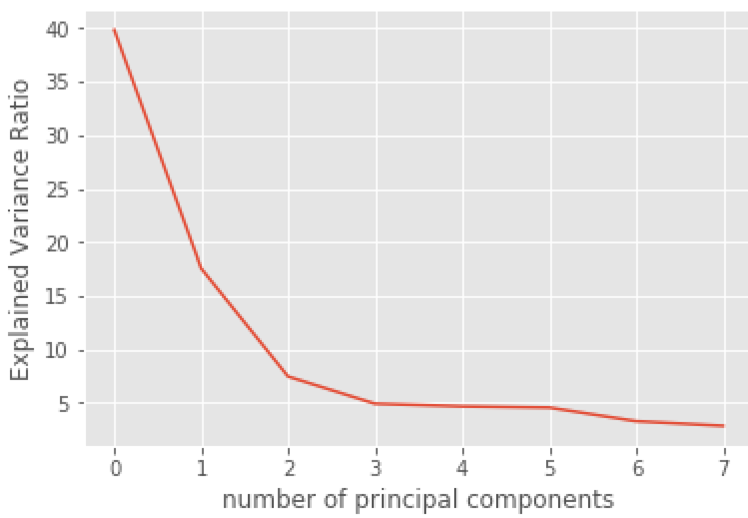


FIGURE 9 PCA Scree Plot



As shown Figures 8 and 9, the biggest drop-off in the variance explained in the PC's ranges occurs anywhere from 6 to 10 PC's. For this analysis, we determined that 8 PC's are the best trade-off between dimensionality reduction and variance explained, as 8 PC's explain around 85% of the variance in the dataset.

We then re-performed our analysis using 8 PC's to obtain a dataset with greatly reduced dimensionality versus the original dataset, while still maintaining a bulk of the information. However, due to the PCA's functionality which orders the PC's in decreasing importance, it is imperative that we reclassify the PC's to have an equal amount of weight distributed amongst themselves. To achieve this, z-scoring has been performed on the resulting PC's to rescale them, making the weight amongst the 8 various PC's equal.

Algorithms

There are two main algorithms that we utilized to calculate our fraud detection score: a heuristic function of the z-scores, and an Autoencoder. The outputs of these two algorithms were then combined using a weighted average rank order to obtain the final fraud score.

Model 1: Heuristic Function of Z-Scores

After performing the PCA and scaling the variables using z-scoring, our selected PC's are each of equal importance. One of the benefits of using z-scoring and PCA is that the current score for each of the individual PC's shows how many standard deviations the current data point is from the mean. Therefore, for the first fraud score, due to the equal weights of the PC's and the nature of the problem, the heuristic function for this analysis is a sum of the absolute values of the different components. This can be written mathematically as:

$$s_i = \sum_{8} \left| z^i{}_8 \right|$$

For the purposes of this analysis, there were other options that could be considered for the heuristic function such as taking the sum of the absolute squares of the different PC components. However, using the absolute squares approach, this would inflate the impact of any one particular PC component outlier. For this analysis, it is deemed more appropriate to take the absolute sums of the different PC's so that if there were records that were consistently marginal outliers, then the sum of those outliers would flag that record as suspicious as opposed to only flagging records where a large outlier occurred in any one PC, which alone may have been caused by a typo or recording error.

After the construction of this algorithm and applying it on the z-scored PC's, the fraud score for each of the records were computed. From there, we rank ordered the fraud scores in descending order with the highest numbers signaling records that we identified most likely to be fraudulent.

Model 2: Autoencoder

Autoencoders encode the data into a handful of significant features, essentially retains the information of the dataset in the hidden layer, and further decodes this back to a reconstructed form of the original data. Based on the reconstruction error (distance measures), we make a judgment on the presence of outliers. It's commonly used in reducing dimensionality and noise, but one of its important applications is anomaly detection. Typically, identifying data points that have different information compared to other data points could be potential fraud transactions. Usually, a given dataset would have mostly normal instances with a few anomalies, autoencoders would have a reconstruction error on these anomalies, thus making it easy for us to detect possible fraud transactions. The autoencoders are trained and tested on the same set of data points, feeding in the input data as labels.

Autoencoders comprises of input and output layers, with encoding, hidden and decoding layers in between. The encoding layers encode the large data fed into it and reduces its dimensionality or compresses it to an encoded format via multiple encoding layers, thus storing the compressed data in a hidden layer. Then, this encoded data is gradually decoded, trying to replicate the input data, but with some reconstruction error in the output.

After constructing the neural network layers using the Keras Library, we trained the model on our input data such that it learned the data encodings in an unsupervised manner and resulted in a reconstructed form on the input data. The input data values were further subtracted from the output data values to find out the records that did not get reproduced well. Thus, this measure of reconstruction error is used in our fraud score.

Final Fraud Score

For the final fraud score, since both fraud scores were quantile binned, both of the fraud scores are on equal footing. Essentially, the quantile binning consisted of a ranked list in a descending order. Therefore, for the final fraud score, a weighted average rank order is constructed. To best determine the weights of the weighted average for the final fraud score, we decided to put equal weight on our Heuristic and Autoencoder scores, since both methods are equally valid for detecting anomalies that deviate away from average. Using an equal weight of 0.5 on the autoencoder and heuristic scores, our final calculation can be written mathematically as:

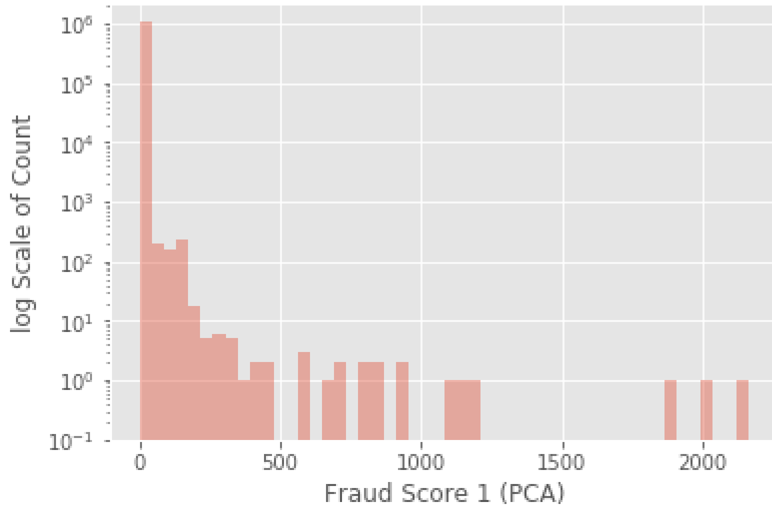
$$S_3 = w_1 s_1 + w_2 s_2$$
$$S_3 = 0.5 s_1 + 0.5 s_2$$
$$with\ w_1 + w_2 = 1$$

Results

After performing the dimensionality reduction techniques, both of the algorithms were constructed with the quantile binning/rank ordering performed.

As shown in Figure 10 below, the histogram Fraud Score #1 (Heuristic Function) is illustrated.

FIGURE 10 Distribution of Fraud Score 1 based on PCA



The histogram shows that there are a significant number of properties with fraud scores over 500. Below is a table with the top ten most fraudulent properties as identified by PCA.

FIGURE 11 Top 10 highest fraud scores 1 based on the PCA

Record	Rank	FS1	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
632816	1	2167.63	576.53	-173.31	-558.56	265.66	355.42	53.63	63.12	121.41
565392	2	2011.09	214.48	546.95	86.11	-334.52	357.34	-245.27	-18.87	-207.56
917942	3	1894.77	441.34	-37.52	697.85	423.75	127.84	51.41	20.85	94.21
111420	4	1207.13	-0.12	0.52	12.64	-95.59	82.28	284.37	621.02	-110.59
1067360	5	1131.96	34.06	378.10	-84.74	109.75	-69.53	55.64	20.03	380.11
67129	6	1093.39	186.40	51.55	191.11	-258.90	-259.38	-52.03	70.69	23.33
585118	7	947.14	169.50	-42.67	90.34	-107.49	-318.44	-10.09	80.15	128.46
565398	8	937.24	389.18	-62.13	118.64	-66.31	-109.64	-15.71	-46.24	-129.38

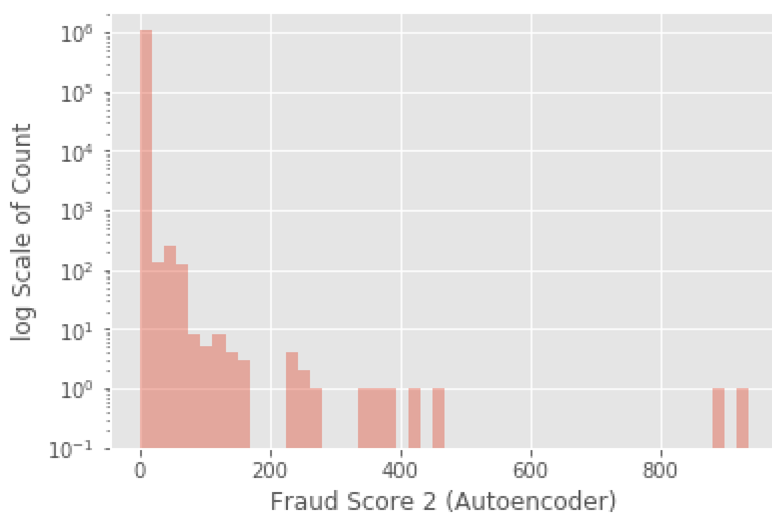
Record	Rank	FS1	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
230596	9	866.23	18.95	219.54	-53.96	59.60	-42.91	23.17	32.37	415.73
111426	10	844.34	0.01	0.97	10.35	-71.70	60.86	208.94	420.25	-71.26

FIGURE 12 Top 10 highest fraud scores 1 based on the PCA

Record	Rank	Owner	Address	FS1
632816	1070994	864163 REALTY, LLC	86-55 BROADWAY	2167.63
565392	1070993	U S GOVERNMENT OWNRD	FLATBUSH AVENUE	2011.09
917942	1070992	LOGAN PROPERTY, INC.	154-68 BROOKVILLE BOULEVARD	1894.77
111420	1070991	BOXWOOD FLTD PARNTERS	1438 3 AVENUE	1207.13
1067360	1070990		20 EMILY COURT	1131.96
67129	1070989	CULTURAL AFFAIRS	1000 5 AVENUE	1093.39
585118	1070988	NEW YORK CITY ECONOMI	28-10 QUEENS PLAZA SOUTH	947.14
565398	1070987	DEPT OF GENERAL SERVI	FLATBUSH AVENUE	937.24
230596	1070986		BELL AVENUE	866.23
111426	1070985	969 PARK CORP	969 PARK AVENUE	844.34

As shown in Figure 13 below, the histogram Fraud Score #2 is illustrated.

FIGURE 13 Distribution of Fraud Score 2 based on Autoencoder



Unlike the PCA, the fraud score based on the autoencoder shows that there aren't as many significant fraud scores over 500. Below is a table with the top ten most fraudulent properties as identified by the autoencoder.

FIGURE 14 Top 10 highest fraud scores 2 based on the Autoencoder

Record	Rank	Fraud Score 2
917941	1	937.760887
632815	2	884.909133
565391	3	467.300467
67128	4	422.655102
585117	5	384.621593
111419	6	369.916942
585438	7	339.987429
585119	8	269.9684
920627	9	257.831366
565397	10	253.580137

FIGURE 15 Top 10 highest fraud scores 2

Record	Rank	owner	Address	Fraud Score 2
917942	1070994	LOGAN PROPERTY, INC.	154-68 BROOKVILLE BOULEVARD	937.760887
632816	1070993	864163 REALTY, LLC	86-55 BROADWAY	884.909133
565392	1070992	U S GOVERNMENT OWNRD	FLATBUSH AVENUE	467.300467
67129	1070991	CULTURAL AFFAIRS	1000 5 AVENUE	422.655102
585118	1070990	NEW YORK CITY ECONOMI	28-10 QUEENS PLAZA SOUTH	384.621593
585439	1070989	11-01 43RD AVENUE REA	11-01 43 AVENUE	369.916942
565398	1070988	DEPT OF GENERAL SERVI	FLATBUSH AVENUE	339.987429
920628	1070987	PLUCHENIK, YAAKOV	7-06 ELVIRA AVENUE	269.9684
585120	1070986		28 STREET	257.831366
1067001	1070985	DRANOVSKY, VLADIMIR	238 BEDELL AVENUE	253.580137

After taking the weighted average of the rankings, we were left a new fraud score and rankings. Below is Figure 16 with the distribution of the Final Fraud Score.

FIGURE 16 Distribution of Final Fraud Score

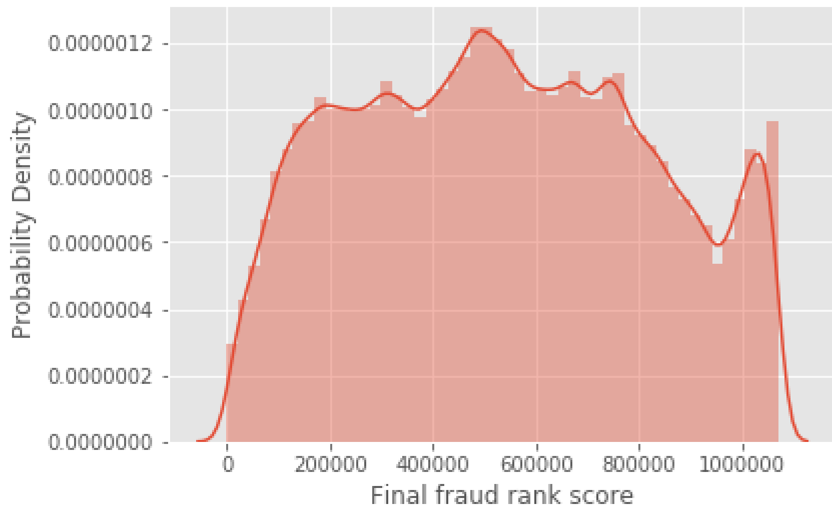


FIGURE 17 Top 10 Final Fraud Score

Record	Rank	Address	Owner	Fraud Score
632816	1	86-55 BROADWAY	864163 REALTY, LLC	1070993.5
917942	2	154-68 BROOKVILLE BOULEVARD	LOGAN PROPERTY, INC.	1070993
565392	3	FLATBUSH AVENUE	U S GOVERNMENT OWNRD	1070992.5
111420	4	1438 3 AVENUE	BOXWOOD FLTD PARNTERS	1070990
67129	5	1000 5 AVENUE	CULTURAL AFFAIRS	1070990
585118	6	28-10 QUEENS PLAZA SOUTH	NEW YORK CITY ECONOMI	1070989
585439	7	11-01 43 AVENUE	11-01 43RD AVENUE REA	1070986
565398	8	FLATBUSH AVENUE	DEPT OF GENERAL SERVI	1070986
111426	9	969 PARK AVENUE	969 PARK CORP	1070984.5
1067360	10	20 EMILY COURT		1070984

Below is a review of each of the top 10 most likely fraudulent properties based on the Final Fraud Score. Figure 18 shows the z-scores for each of the primary fields.

FIGURE 18 Z-Score Distribution for the Top 10 Fraud Scores

Record	Rank	FULLVAL	AVLAND	AVTOT	LTFRONT	LTDEPTH	BLDFRONT	BLDDEPTH	STORIES
632816	1	0.17	0.30	0.16	0.71	-0.19	-0.72	-1.12	-0.48
917942	2	31.79	441.08	673.53	35.81	-0.15	-0.24	-0.42	-0.23
565392	3	368.51	478.98	280.86	0.41	-0.09	0.39	1.27	-0.48
111420	4	-0.05	-0.02	-0.01	0.10	-0.21	158.55	186.15	3.15
67129	5	523.89	656.54	399.27	5.75	-0.14	-0.20	0.46	1.22
585118	6	0.22	0.36	0.19	1.75	2.14	-0.72	-1.12	1.82
585439	7	0.24	0.04	0.21	0.24	0.34	-0.72	-1.12	0.61
565398	8	196.80	255.84	150.01	2.99	6.75	-0.34	-0.30	-0.35
111426	9	1.66	1.11	1.29	0.84	0.55	157.77	186.05	0.85
1067360	10	-0.01	-0.01	-0.03	-0.45	-0.91	0.01	-0.24	-0.35

Below is a review of each of the top 10 most fraudulent properties based on the Final Fraud Score:

Record: 632816

Record 632816 is considered the most likely fraudulent property. Although its valuation (FULLVAL, AVLAND, and AVTOT) were not outliers, its building area (BLDFRONT*BLDDEPTH, which would be 1 due to both fields equaling 1) make the property an outlier by price per building square footage. It is highly unlikely that the building would actually be 1 square foot. A closer look at the exemption codes reveals that the property receives an exemption for school tax relief.

Record: 917942
Record 917942 is considered the second most likely fraudulent property. It's valuation far exceeds the mean and median valuations. Furthermore, LTDEPTH, BLDFRONT, and BLDDEPTH values were originally zero (0) and filled with median values through the data cleaning process. Thus, it would be reasonable to see that the valuations per area also exceed both the average and median.

However, it may be reasonable to expect that this property would be an outlier as the exemption code classifies it as an airfield. Airfields are likely to have a large lot size relative to building square footage. Although Record 917942 only has LTFRONT filled, its massive AVLAND is likely due to the actual size of the property.

Record: 565392
Record 565392 is considered the third most likely fraudulent property. Similar to Record 917942, it's valuation far exceeds the mean and median valuations. Furthermore, it also has zero (0) for building area information and replacing it with the median would only result in value per square footage far exceeding both the average and median. The exemption code classifies the property as a park, and thus, the building square footage may actually be zero square feet. Leaving the square footage at zero, however, would not change the property from being flagged as fraudulent.

Record: 111420
Record 111420 is considered the fourth most likely fraudulent property. This property has high valuations and although it has high values for a building square footage, it's low lot square footage to flag the property for extremely high valuations for lot square footage. Furthermore, this property's extremely high building square footage may be the factor that triggers it to be considered fraudulent. Unlike the other school tax relief properties, this property has building area information given. Thus, more research may be required to discern whether or not this property is truly fraudulent or if it is following a rule regarding School Tax Relief properties.

Record: 67129
Record 67129 is considered the fifth most likely fraudulent property. It has a high valuation that far exceeds the mean and median valuations. Furthermore, it has zero (0) for building area information and replacing it with the median would only result in value per square footage far exceeding both average and median. The exemption code classifies the property as a park, and thus, the building square footage may actually be zero square feet. Leaving the square footage at zero, however, would not change the property from being flagged as fraudulent.

Record: 585118
Record 585118 is considered the sixth most likely fraudulent property. Similar to record 632816, this property has high valuations and a building square footage of 1 square foot. This would result in a high valuation per square foot and flag the property as fraudulent. This property also has the same

exemption code (School Tax Relief), which may be why the building square footage was inputted at 1 for both BLDFRONT and BLDDEPTH.

Record: 585439

Record 585439 is considered the seventh most likely fraudulent property. Similar to record 632816, this property has high valuations and a building square footage of 1 square foot. This would result in a high valuation per square foot and flag the property as fraudulent. Unlike the other properties classified as School Tax Relief, this property is classified as IND/Special Ex, which is currently unknown.

Record: 565398

Record 565398 is considered the eighth most likely fraudulent property. Similar to record 917942, this property has high valuations and zero (0) in both BLDFRONT and BLDDEPTH. Thus, replacing building area with median values would result in high valuations per square foot and flag the property as fraudulent. The property is currently classified as Port Terminals and that may explain why it is valued so highly.

Record: 111426

Record 111426 is considered the ninth most likely fraudulent property. This property has a high valuation and high building square footage. It is likely flagged for having high building square footage. The relative valuations per square footage as a result, may be very low and also flag the property being fraudulent. The property is considered veteran property by the exemption code and it is currently unknown whether or not that has any implications to the estimation of the valuations.

Record: 1067360

Record 1067360 is considered the tenth most likely fraudulent property. This property has a low lot square footage of 1 square foot. This is likely to result in a high valuation per lot square foot and flag the property as fraudulent. Unlike the other school tax relief properties, this property has building area information given. Thus, more research may be required to discern whether or not this property is truly fraudulent or if it is following a rule regarding School Tax Relief properties.

Conclusions

In this project, we leveraged the unsupervised learning method to detect the top 10 possible frauds among the NY Property Valuation and Assessment database. Using dimensionality reduction technology and z-scaling approach, we calculated two fraud scores, one with the heuristic function of z-scores and other with an autoencoder, for every property to eventually generate a final fraud score.

Most of the fraudulent properties are related to public properties such as schools and parks. These types of properties have either larger than normal lot square footage or building square footage, making the actual valuation of square footage far below the average. Furthermore, some properties only have

Lot or building square footages equal to one (1). There may be a specific reason for this and may require further exploration beyond the existing data.

However, with more time, we can develop a table to assist in identifying government properties and exclude themm from the analysis. Doing so may change results into focusing on either or both commercial/industrial and residential properties. More time would also allow us to acquire external data sets to bring in and improve our analysis. Lastly, further clarification of project goals and scope may help drive us to focus on specific segments and isolate the potential frauds that we'd want to look at.

Appendix

Data Quality Report

Table of Contents

Introduction

This Data Quality Report ("DQR" or "Report") is a review of the City of New York ("NYC") Property Valuation and Assessment Data (“Assessor’s Roll”) available on the NYC Open Data portal. The Assessor's Roll was first made publicly available on September 2, 2011 and has been updated on an as-needed basis. For the purpose of this Report, the Assessor's Roll data reviewed was last updated September 10, 2018.

A. Purpose of the Assessor's Roll

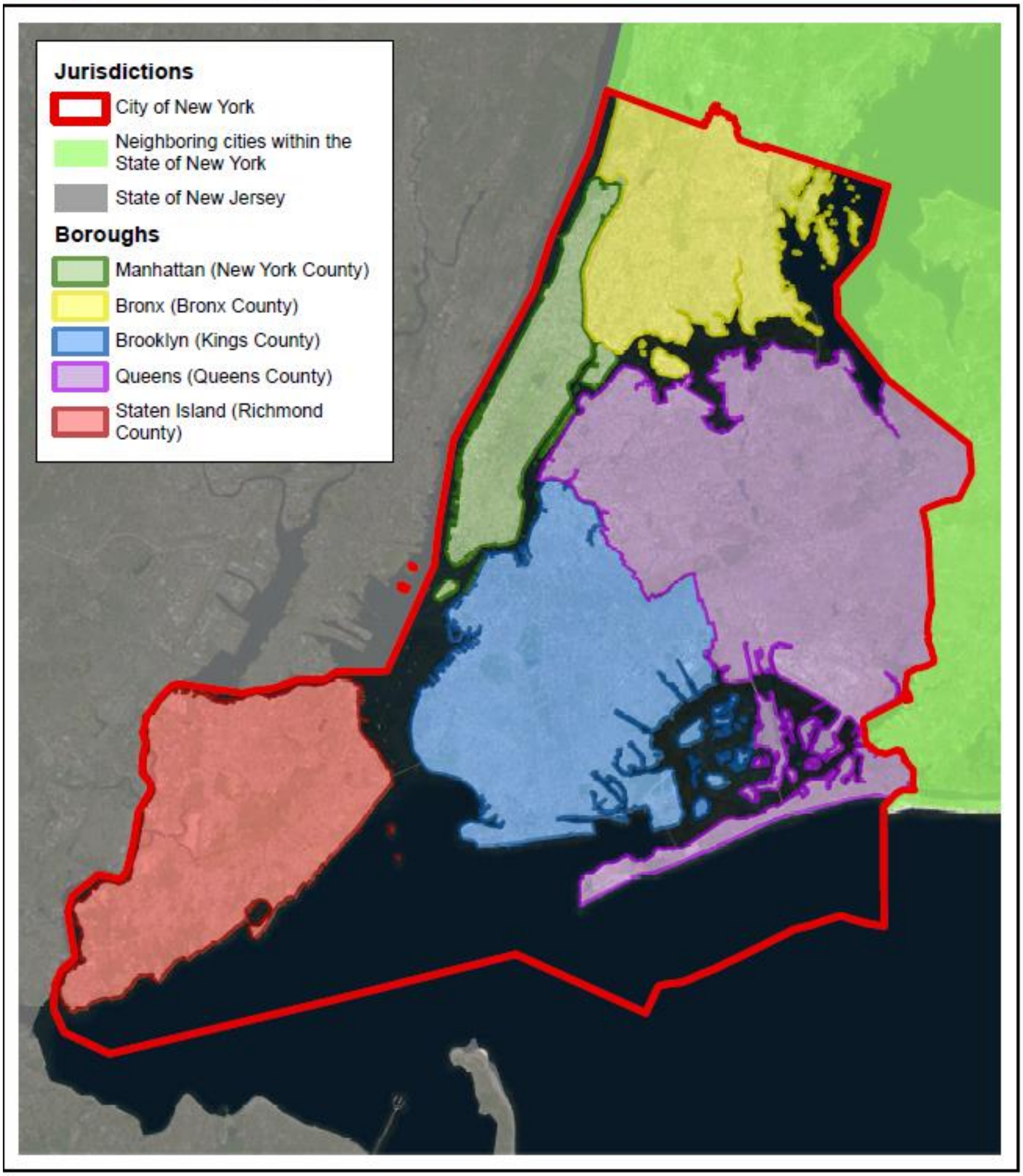
The primary use of the Assessor’s Roll is to assist in the determination of property tax bills for every new fiscal year; it acts as a record to help in determining the amount of property tax due, the entity responsible for paying it, and where to mail it. This is important to many local agencies as property taxes can make up a significant portion of their revenues. For example, in the fiscal year 2017, which ended on June 30, 2017, approximately 45 percent of all city tax revenues for NYC came from property taxes.

However, the Assessor's Roll play's an important role in other functions as well. For example, School Districts looking to raise funds for the construction of a new school can pursue issuing debt through local taxes. State, county, and/or city regulations may place limitations on taxable amounts based on assessed values which makes it important for the School District to review the Assessor's Roll for the total valuation of taxable properties within their jurisdiction. The Assessor's Roll can be useful to industries. Examples include real estate, finance, market analysis, etc.…

B. Summary of the Data

The data reviewed in this report can be found on Open Data. It is the condensed version of the Assessor's Value Roll which includes 31 different fields for every property within NYC. Of the 32 fields, 14 fields are numeric and 18 are categorical; as of the last update (September 10, 2018), the Assessor's Roll identified 1,070,994 properties. Please see Map 1 on the following page for a geographical extent of the data.

Map 1 - Geographic Profile



Field Summary

1.Aggregate Summary

Field	Field Type	Records	% Populated	Unique Values	Records with value zero	% with value zero
RECORD	Categorical	1,070,994	100.00%	1,070,994	0	0.00%
BBLE	Categorical	1,070,994	100.00%	1,070,994	0	0.00%
B	Categorical	1,070,994	100.00%	5	0	0.00%
BLOCK	Categorical	1,070,994	100.00%	13,984	0	0.00%
LOT	Categorical	1,070,994	100.00%	6,366	0	0.00%
EASEMENT	Categorical	4,636	0.43%	13	0	0.00%
OWNER	Categorical	1,039,249	97.04%	863,348	0	0.00%
BLDGCL	Categorical	1,070,994	100.00%	200	0	0.00%
TAXCLASS	Categorical	1,070,994	100.00%	11	0	0.00%
LTFRONT	Numeric	1,070,994	100.00%	1,297	169,108	15.79%
LTDEPTH	Numeric	1,070,994	100.00%	1,370	170,128	15.89%
EXT	Categorical	354,305	33.08%	4	0	0.00%
STORIES	Numeric	1,014,730	94.75%	112	0	0.00%
FULLVAL	Numeric	1,070,994	100.00%	109,324	13,007	1.21%
AVLAND	Numeric	1,070,994	100.00%	70,921	13,009	1.21%
AVTOT	Numeric	1,070,994	100.00%	112,914	13,007	1.21%
EXLAND	Numeric	1,070,994	100.00%	33,419	13,007	1.21%
EXTOT	Numeric	1,070,994	100.00%	64,255	432,572	40.39%
EXCD1	Categorical	638,488	59.62%	130	0	0.00%
STADDR	Categorical	1,070,318	99.94%	839,281	0	0.00%
ZIP	Categorical	1,041,104	97.21%	197	0	0.00%
EXMPTCL	Categorical	15,579	1.45%	15	0	0.00%
BLDFRONT	Numeric	1,070,994	100.00%	612	228,815	21.36%
BLDDEPTH	Numeric	1,070,994	100.00%	621	22,853	2.13%
AVLAND2	Numeric	282,726	26.40%	58,592	0	0.00%
AVTOT2	Numeric	282,732	26.40%	111,361	0	0.00%
EXLAND2	Numeric	87,449	8.17%	44,196	0	0.00%
EXTOT2	Numeric	130,828	12.22%	48,349	0	0.00%
EXCD2	Categorical	92,948	8.68%	61	0	0.00%
PERIOD	Categorical	1,070,994	100.00%	1	0	0.00%
YEAR	Categorical	1,070,994	100.00%	1	0	0.00%
VALTYPE	Categorical	1,070,994	100.00%	1	0	0.00%

2.Numerical Field Summary

Field	Records	Unique Values	Mean	Median	Mode	Min	Max	Std. Dev.
LTFRONT	1,070,994	1,297	36.64	25	0	0	9,999	74.03
LTDEPTH	1,070,994	1,370	88.86	100	100	0	9,999	76.40
STORIES	1,014,730	112	5.01	2	2	1	119	8.37
FULLVAL	1,070,994	109,324	874,2641	447,000	0	0	6.1x10^9	11,582,431
AVLAND	1,070,994	70,921	85,067	13,678	0	0	2.6x10^9	4,057,260
AVTOT	1,070,994	112,914	227,238	25,340	0	0	4.6x10^9	6,877,529
EXLAND	1,070,994	33,419	36,423	1,620	0	0	2.6x10^9	3,981,575
EXTOT	1,070,994	64,255	91,186	1,620	0	0	4.6x10^9	6,508,402
BLDFRONT	1,070,994	612	23.04	20	0	0	7,575	35.58
BLDDEPTH	1,070,994	621	39.92	39	0	0	9,393	42.71
AVLAND2	282,726	58,592	246,235	20,145	2,408	3	2.3x10^9	6,178,962
AVTOT2	282,732	111,361	713,911	79,963	750	3	4.5x10^9	11,652,529
EXLAND2	87,449	44,196	351,235	3,048	2,090	1	2.3x10^9	10,802,213
EXTOT2	130,828	48,349	656,768	37,062	2,090	7	4.5x10^9	16,072,510

3.Categorical Field Summary

Field	Records	Unique Values	Most Common Filed	Percentage of Total
RECORD	1,070,994	1,070,995	N/A	N/A
BBLE	1,070,994	1,070,995	N/A	N/A
B	1,070,994	5	4	33.43%
BLOCK	1,070,994	13,984	3,944	0.36%
LOT	1,070,994	6,366	1	2.28%
EASEMENT	4,636	13	E	89.47%
OWNER	1,039,249	863,348	PARKCHESTER PRESER...	0.58%
BLDGCL	1,070,994	200	R4	13.06%
TAXCLASS	1,070,994	11	1	61.69%
EXT	354,305	4	G	75.35%
EXCD1	638,488	130	1017	66.62%
STADDR	1,070,318	839,281	501 SURF AVENUE	0.08%
ZIP	1,041,104	197	10,314	2.36%
EXMPTCL	15,579	15	X1	44.37%
EXCD2	92,948	61	1017	70.77%
PERIOD	1,070,994	1	Final	100.00%
YEAR	1,070,994	1	2010/11	100.00%
VALTYPE	1,070,994	1	AC-TR	100.00%

Field Details

The Assessor's Roll as 32 different fields. Below is a short description of every field including additional information such as graphs or tables to provide greater insights into the data:

1.RECORD

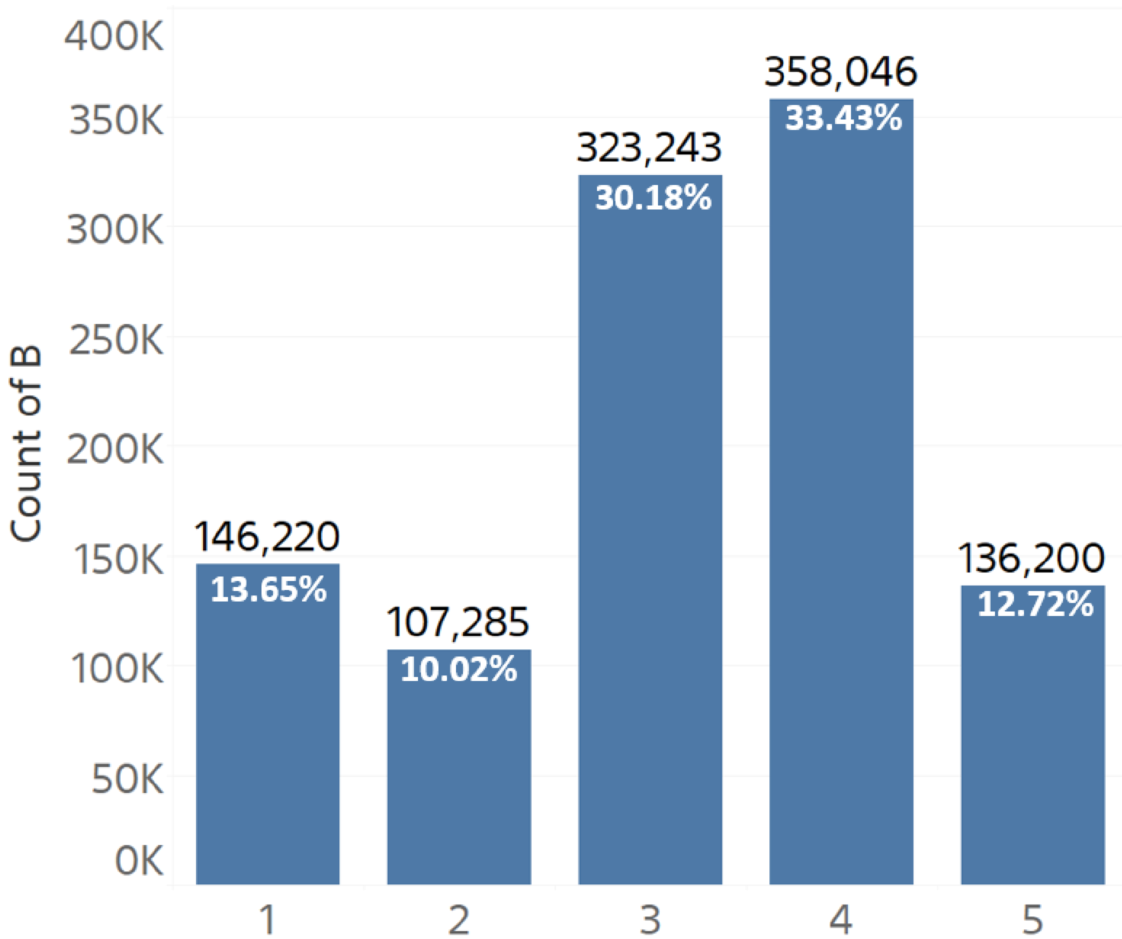
RECORD is a nominal categorical field that provides a unique ordinal identifier for each tuple. Since each tuple is assigned a single unique identifier, the field follows a uniform distribution. A graph or table would not provide any greater insight into the data and thus, was not included.

2.BBLE

The NYC Department of Finance has the Borough-Block-Lot classification system ("BBLE") as the unique identifier buildings and properties. This makes BBLE a categorical field. The structure of the BBLE number is 1 digit for Borough followed by 5 digits for Block and then 4 digits for Lot. Since each tuple is assigned a unique identifier, the field follows a uniform distribution. A graph or table would not provide any greater insight into the data and thus, was not included.

3.B

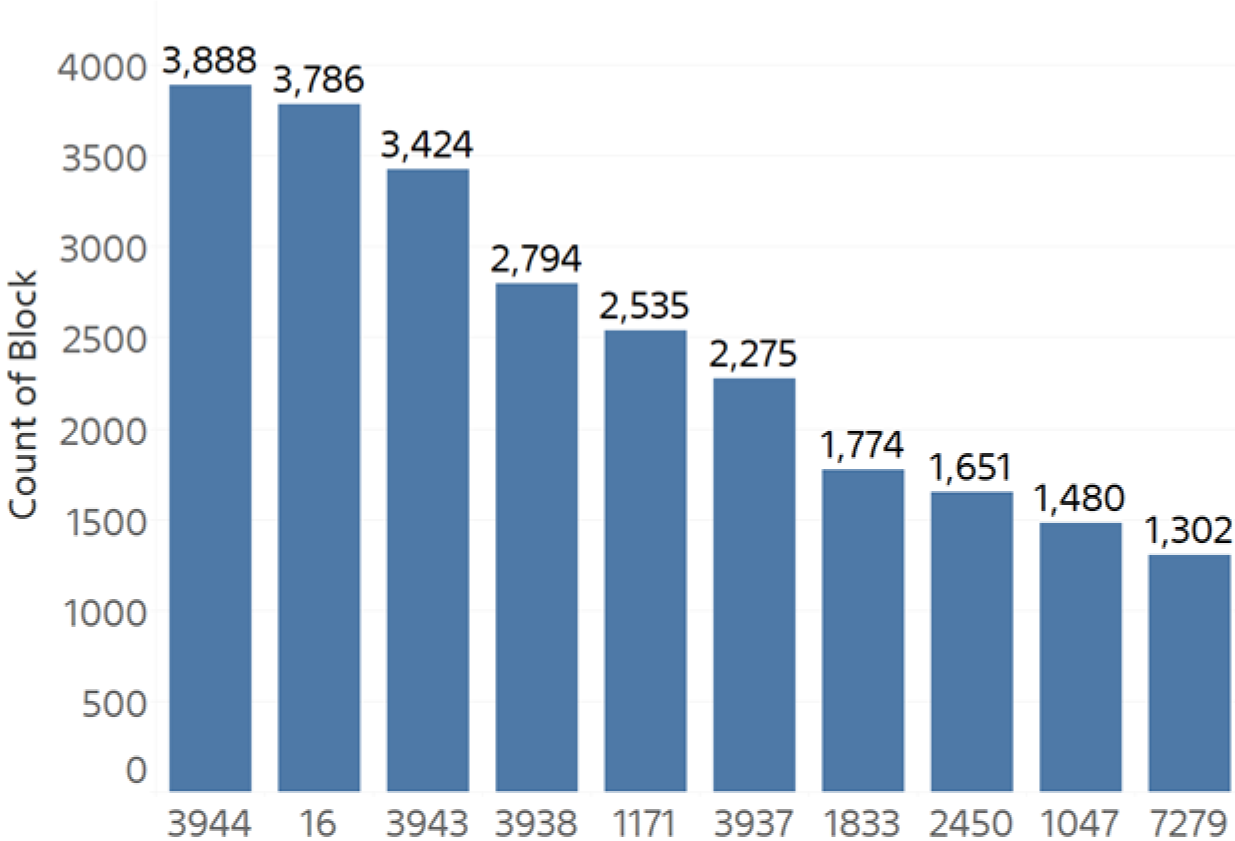
B is a categorical field that represents the borough which the parcel resides in. There are 1,070,994 records with one of five (5) unique boroughs that represent different counties: Manhattan (New York County), Bronx (Bronx County), Brooklyn (Kings County), Queens (Queens County), and Staten Island (Richmond County) which are coded 1 through 5, respectively. Graph 1 below shows a bar graph of B.



As shown in the graph above, Borough 4 (Queens) has the most tuples within the data followed by Borough 3 (Brooklyn). Together, the two make up most of the data. Skew is unimportant as the designation of Borough is more related geographically than its relation ordinally. As shown in Map 1 on page 2, Boroughs Queens and Brooklyn appear to be largest Boroughs by geographic size down on the southeastern portion of NYC.

4.BLOCK

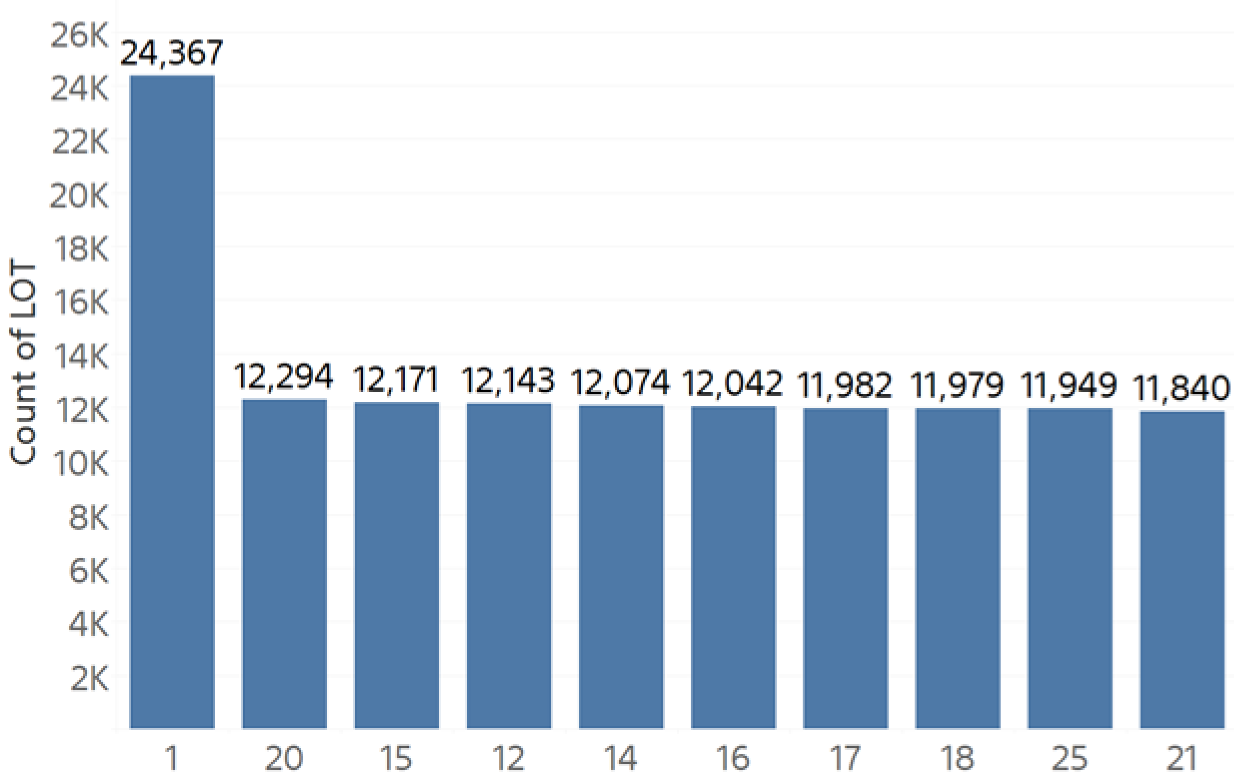
BLOCK is a categorical field that has 1,070,994 records that represent the Tax Block which a sub-division of a borough on which real properties are located. It is part of the Borough-Block-Lot classification system used by the NYC Department of Finance to identify real properties. Graph 2 on the following page shows a bar graph of BLOCK.



Based on Graph 2, it looks like the data has a slight skew to the right but not in an ordinal sense. In general, planning labels are ordinal, and it would make sense that the lower number of blocks, in general, would have on average, a higher number of samples. However, the data doesn't fully comply with that logic and further research may be needed. It is possible that some blocks were removed and replaced with blocks with a higher number.

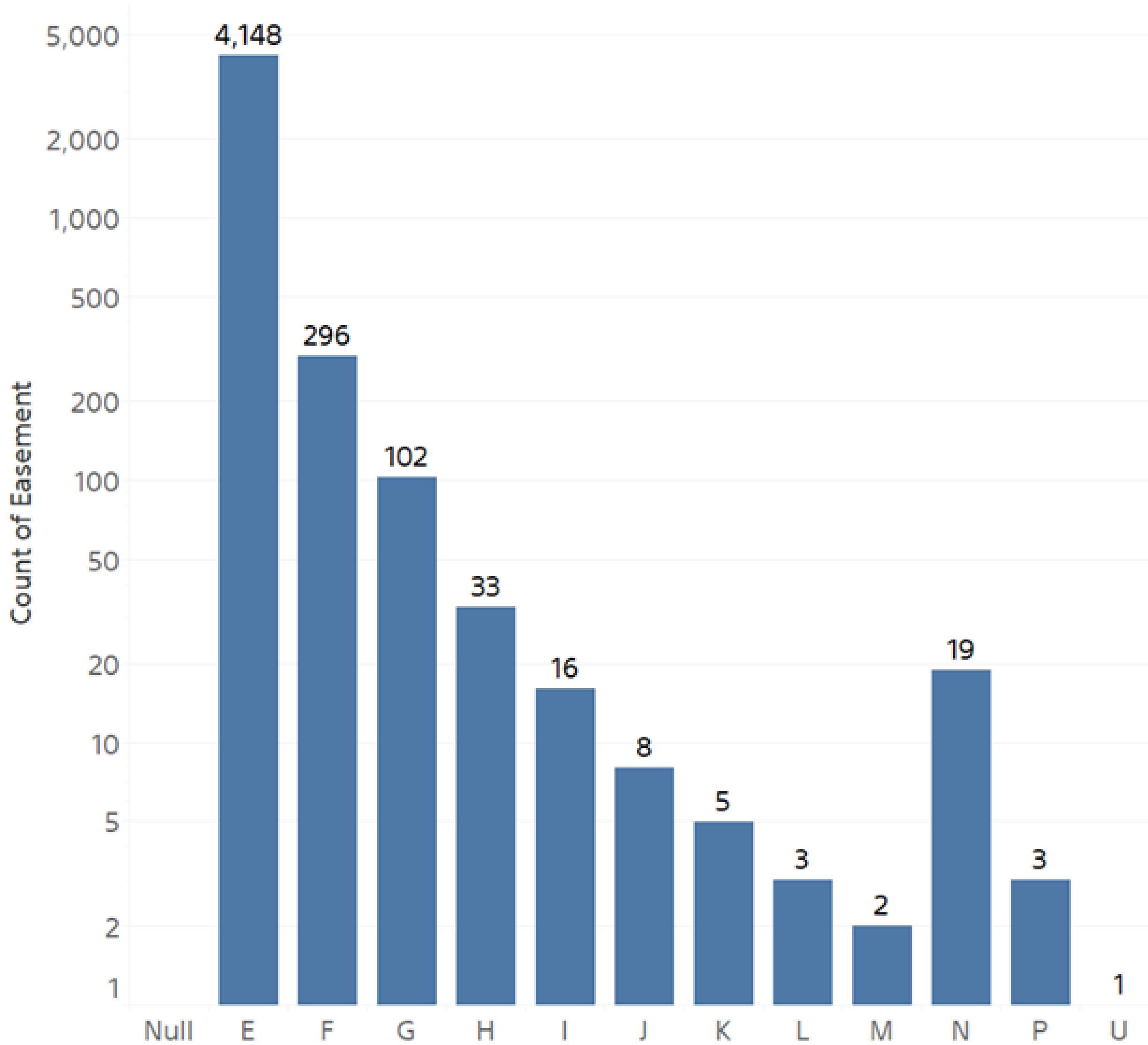
5.LOT

LOT is a categorical field that represents the Tax Lot, which is the smallest subdivision representing a unique location. There are 1,070,994 entries in the field and below is Graph 3 showing a graph of LOT:



Like BLOCK, LOT is a categorical field where the nominal assignments don't necessarily mean much based on their value. Graph 3 doesn't provide very much noteworthy information other than that lot 1 is approximately double if not greater than many other values. More information on how planners assigned the lot numbers may provide insight into the pattern observed above.

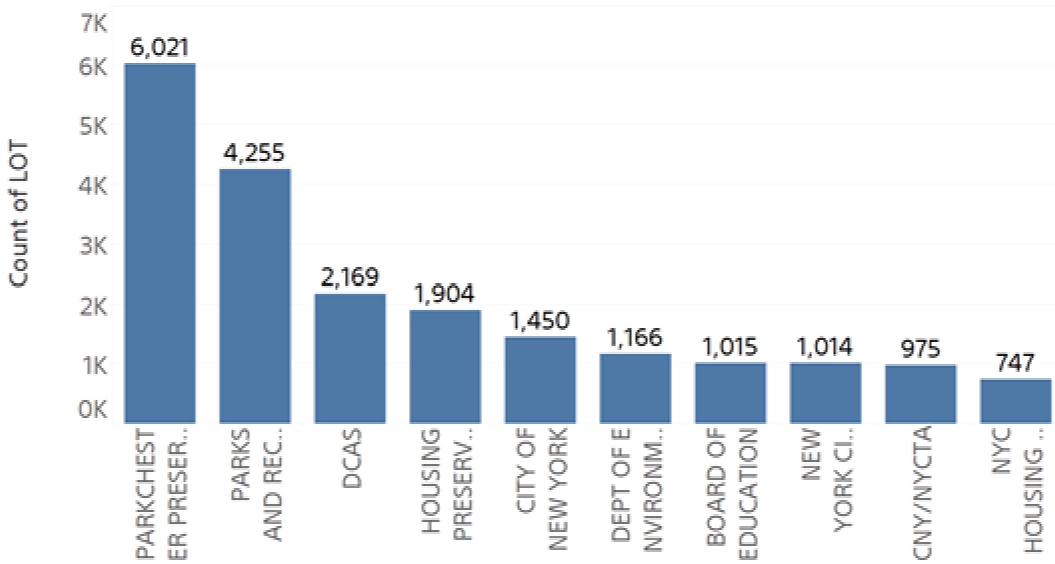
6.EASEMENT



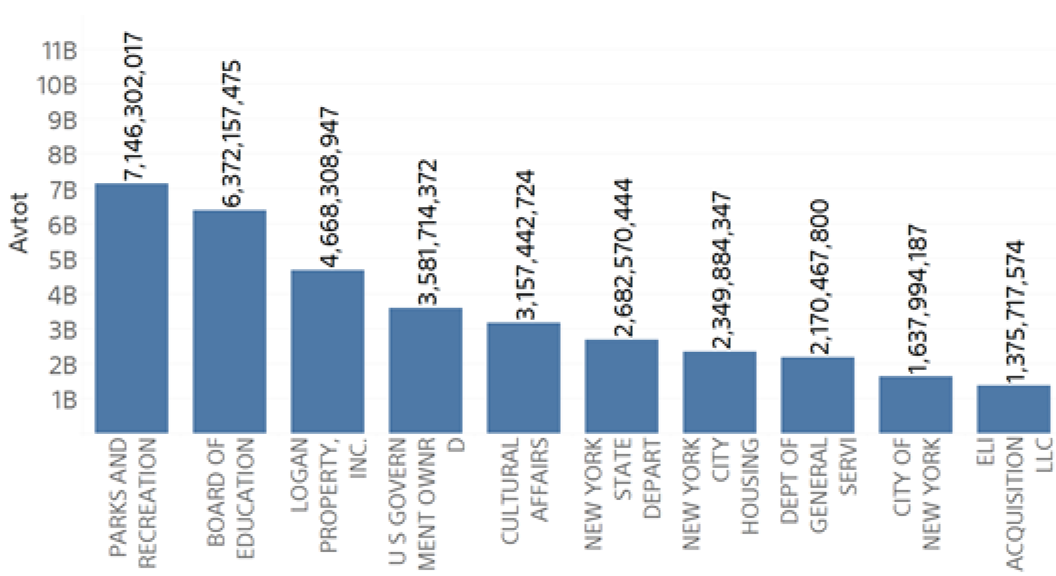
Although there is a pattern of decline as the designation descends down alphabetically, more information regarding the assignments in order to understand this trend.

7.OWNER

OWNER is a categorical field that has 1,039,249 valid records and indicates the entity responsible for the property. This field is important for when the NYC Department of Finance sends out the annual property tax bill. However, it may also be useful to aggregate properties to identify the largest landowners or property taxpayers. An example is when a local agency goes to issue a bond, it must publish a list of the top 20 largest landowners (by assessed value) in order to indicate the relative level of risk. The reason for this is that property taxes would pay for the bonds and the potential for a large landowner leaving is a risk of losing income. Graph 5 below shows the top 10 owners by lots:



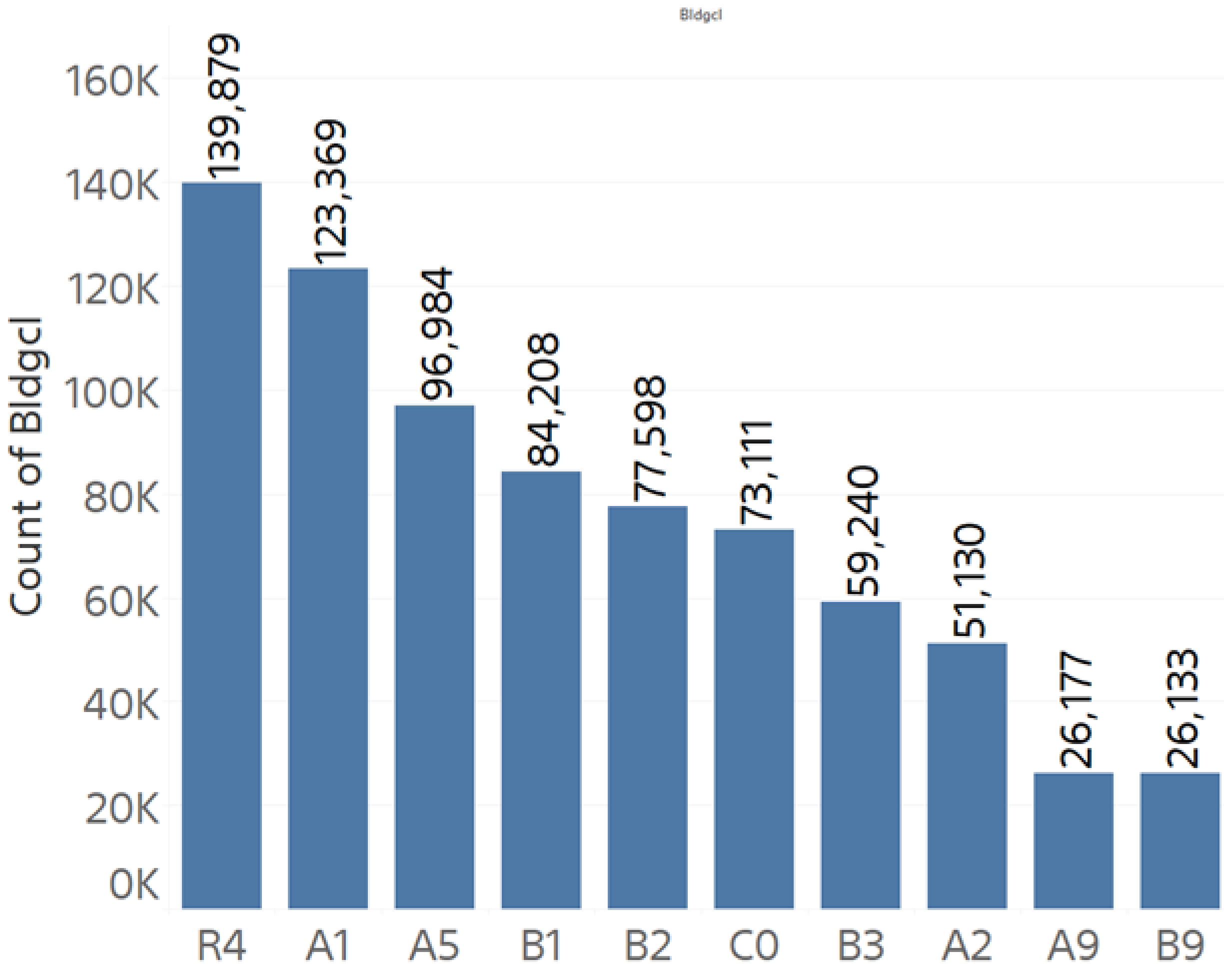
Graph 6 Below shows the top 10 owners by total assessed value:



Based on the graphs above, the number of lots and assessed value are skewed to the right.

8.BLDGCL

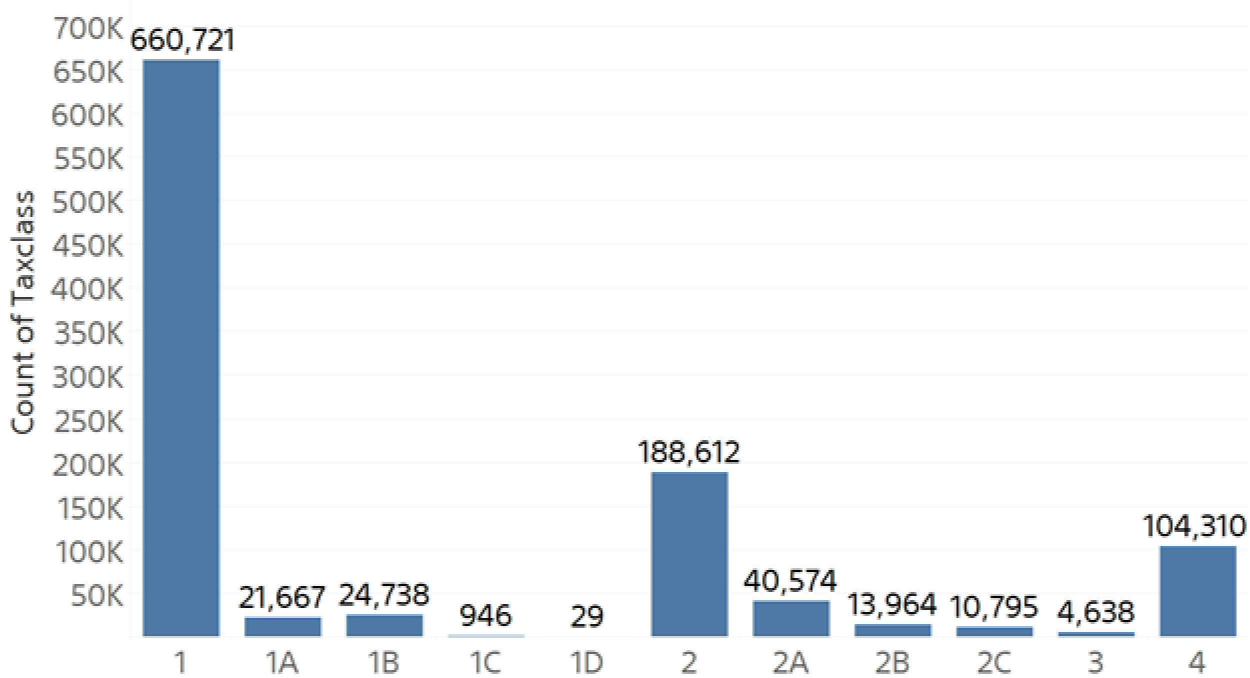
The BLDGCL field has 1,070,994 valid records and it is the building classification that provides information on a property's constructive use. The classification system is broken into two (2) parts, a letter indicating the general class of the property followed by a number providing more specific information regarding a property's use. Please see Exhibit 2 for more information regarding the NYC Building Classification. Graph 7 below shows the top 10 building classifications:



Although R4 (Condos) has the largest share of records, groups of A's and B's make up a larger share of the top 10. Based on the NYC Building Classification Codes in Exhibit 2, the A's and B's seem to be individual homes while R's are multifamily units. Thus, based on the top 10 list, at least 50.87% of the records are single family or two-family homes.

9.TAXCLASS

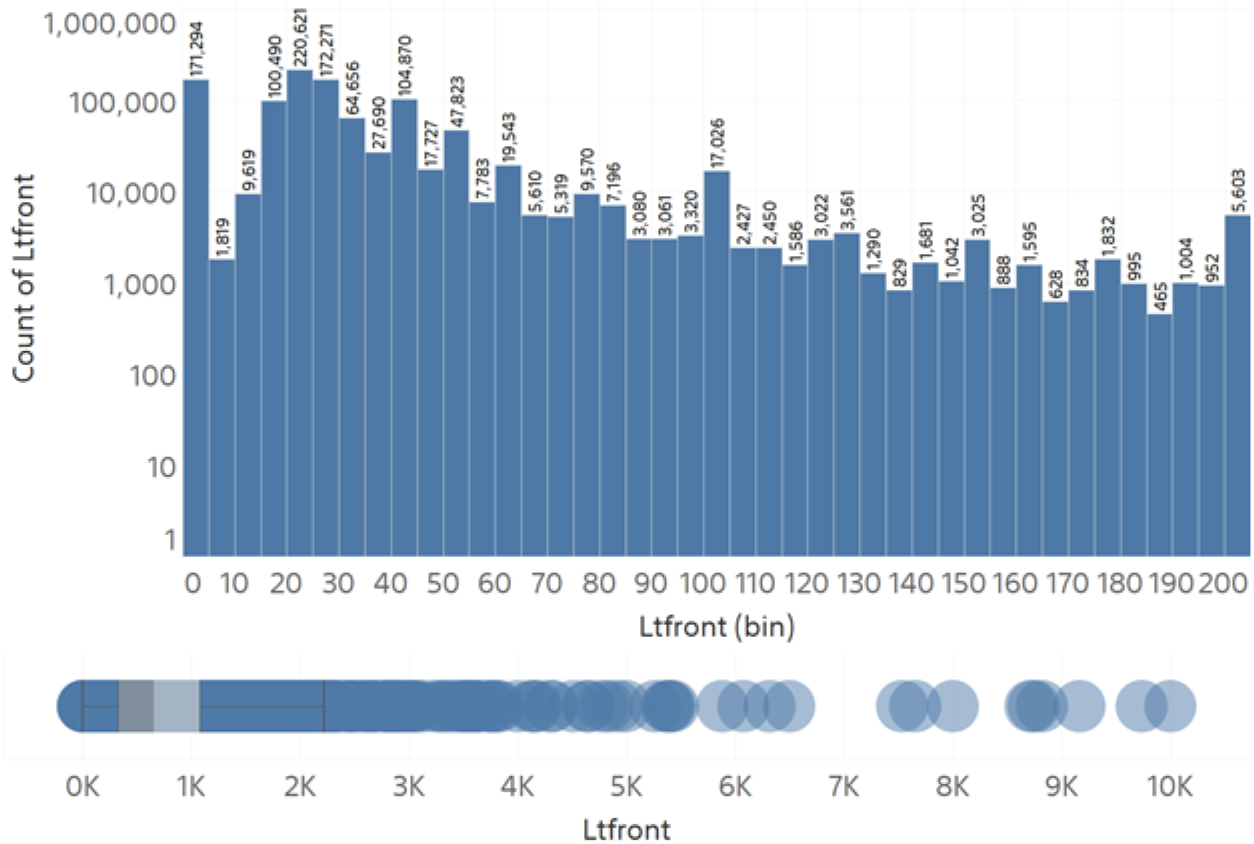
TAXCLASS is categorical field that represents the tax class designation assigned by the NYC Department of Finance and it is used to identify the correct tax rates when determining the annual property tax bill for all 1,070,994 properties/buildings. Please see Exhibit 3 for more information regarding the NYC Department of Finance tax classes. Graph 8 below shows a bar graph of the tax classes:



Based on Graph 8, most of the lots are related to residential properties since most of the tuples have values equal to or associated with 1 or 2.

10.LTFRONT

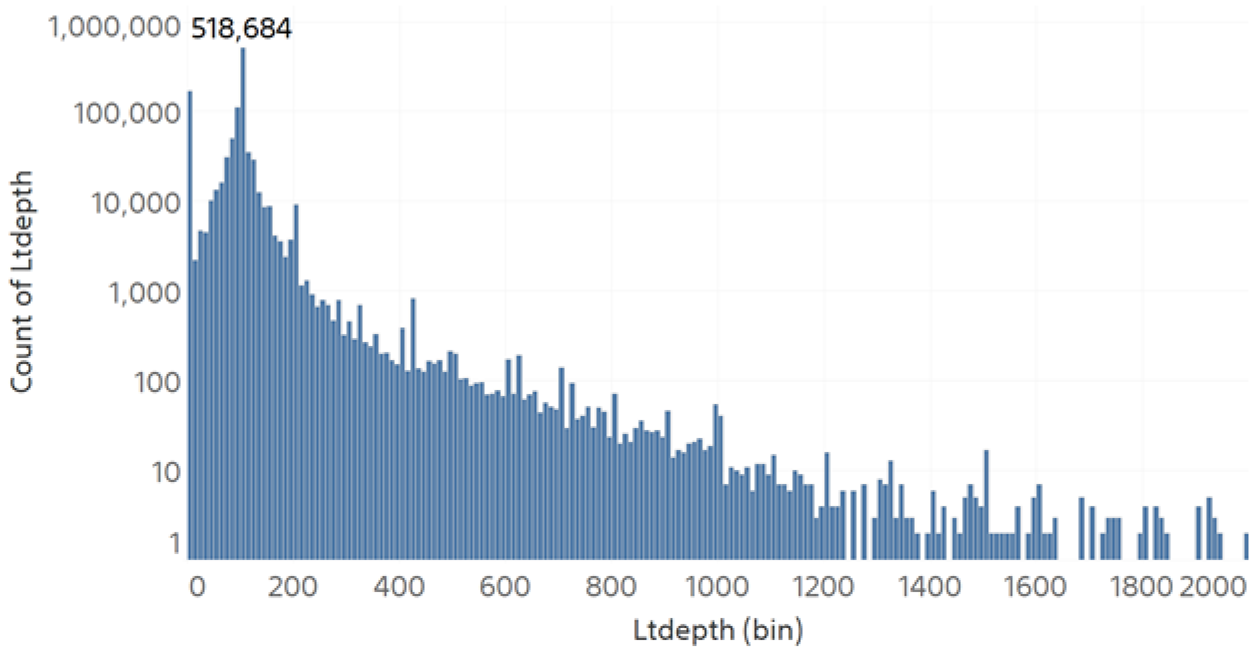
LTFRONT is the measurement of the lot width in feet. Thus, it is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 9 below shows the histogram of LTFRONT followed by its boxplot:

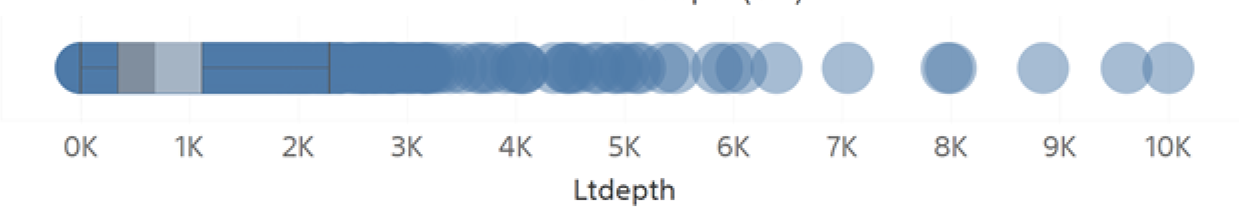


Based on Graph 9 above, most of LTFRONT seems to show a skew to the right and that is backed by the subsequent boxplot. There is an upper whisker at 2,213 which results in 109 outliers.

11.LTDEPTH

LTDEPTH is the measurement of the lot depth in feet. Thus, it is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 10 below shows the histogram of LTDEP followed by its boxplot:

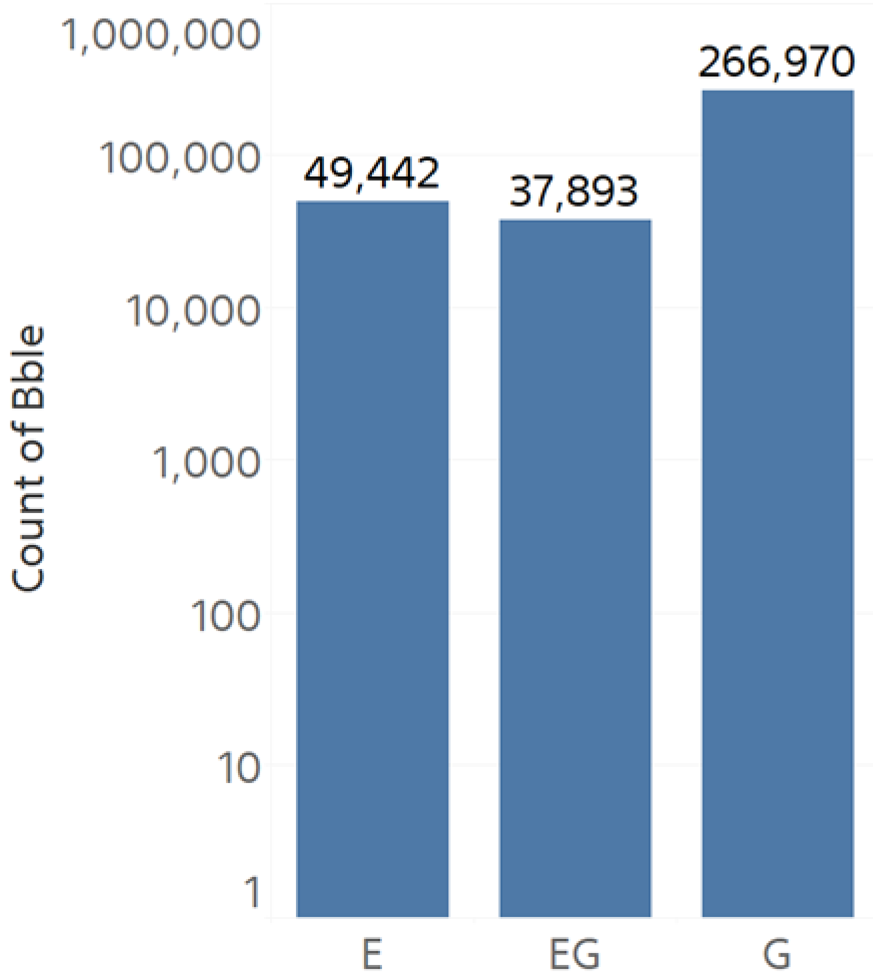




Based on Graph 10 above, most of LTDEPTH seems to show a skew to the right and that is backed by the subsequent boxplot. There is an upper whisker at 2,281 which results in 120 outliers.

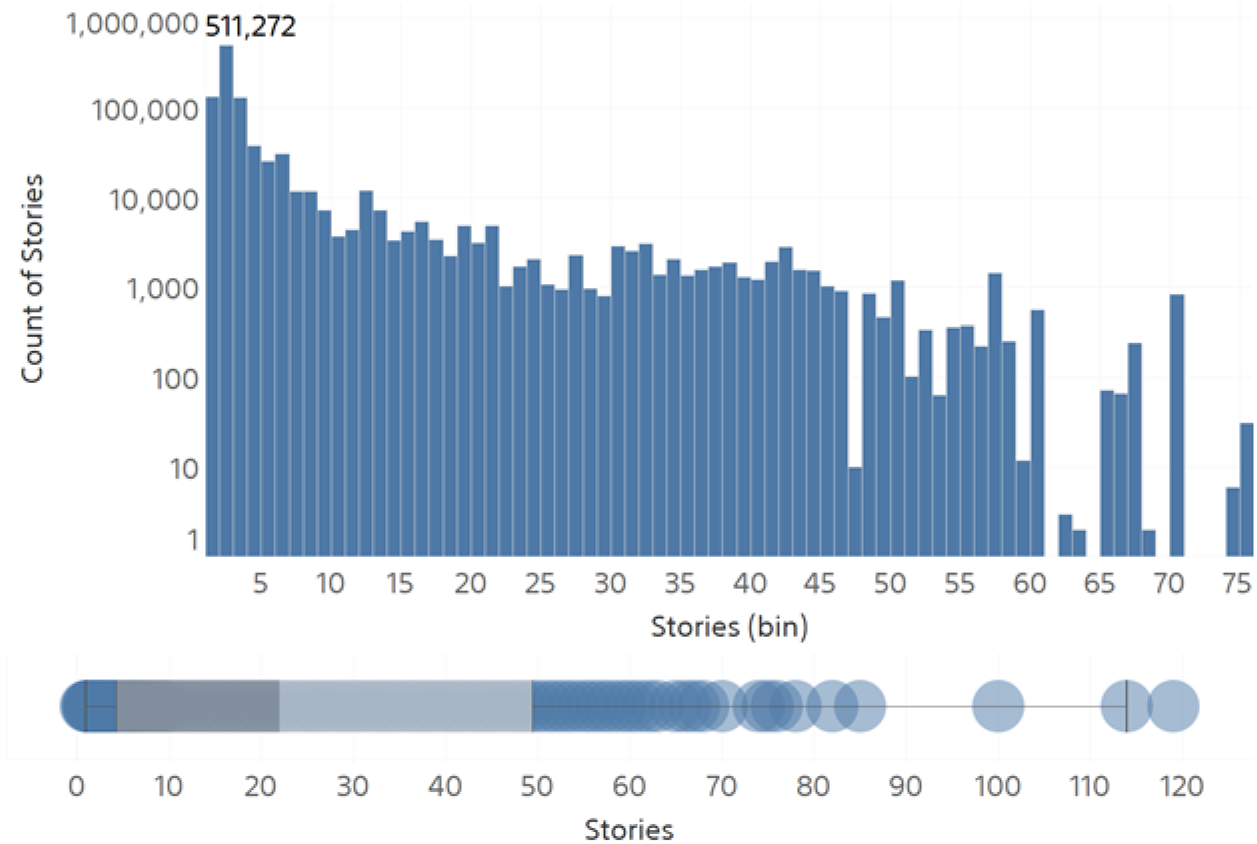
12.EXT

The EXT field only has 354,305 valid records and it is a categorical field that indicates whether or not the property has an extension and the type of extension has been made. According to the NYC Department of Finance, the field could only be filled with either E for Extension, G for Garage, or EG for Extension and Garage. Graph 11 below shows a bar graph of EXT:



13.STORIES

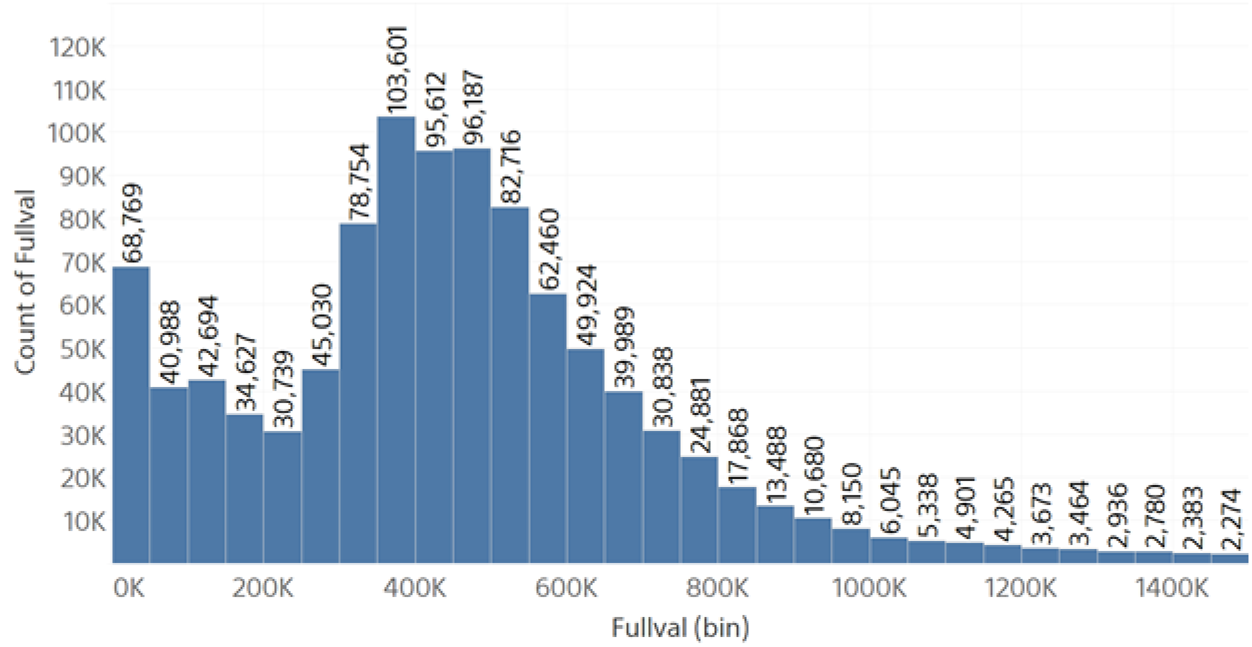
STORIES is a numeric field that indicates the number of floors for the building. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 12 below shows the histogram of STORIES followed by its boxplot:

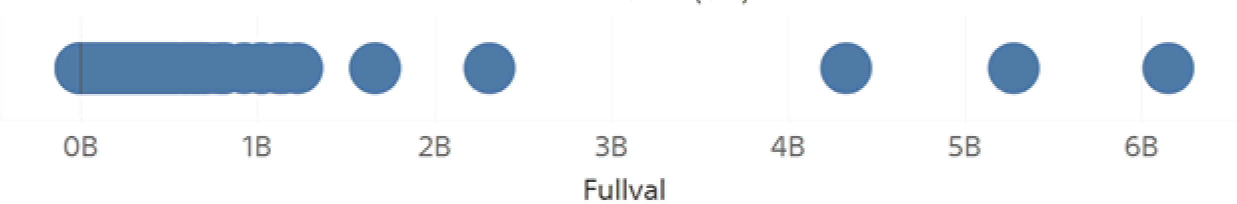


Based on Graph 12 and the subsequent boxplot, STORIES is skewed to the right There is an upper whisker at 114 which results in 1 outlier.

14.FULLVAL

The FULLVAL field states the total market value of the property. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 13 below shows the histogram of FULLVAL followed by its boxplot:

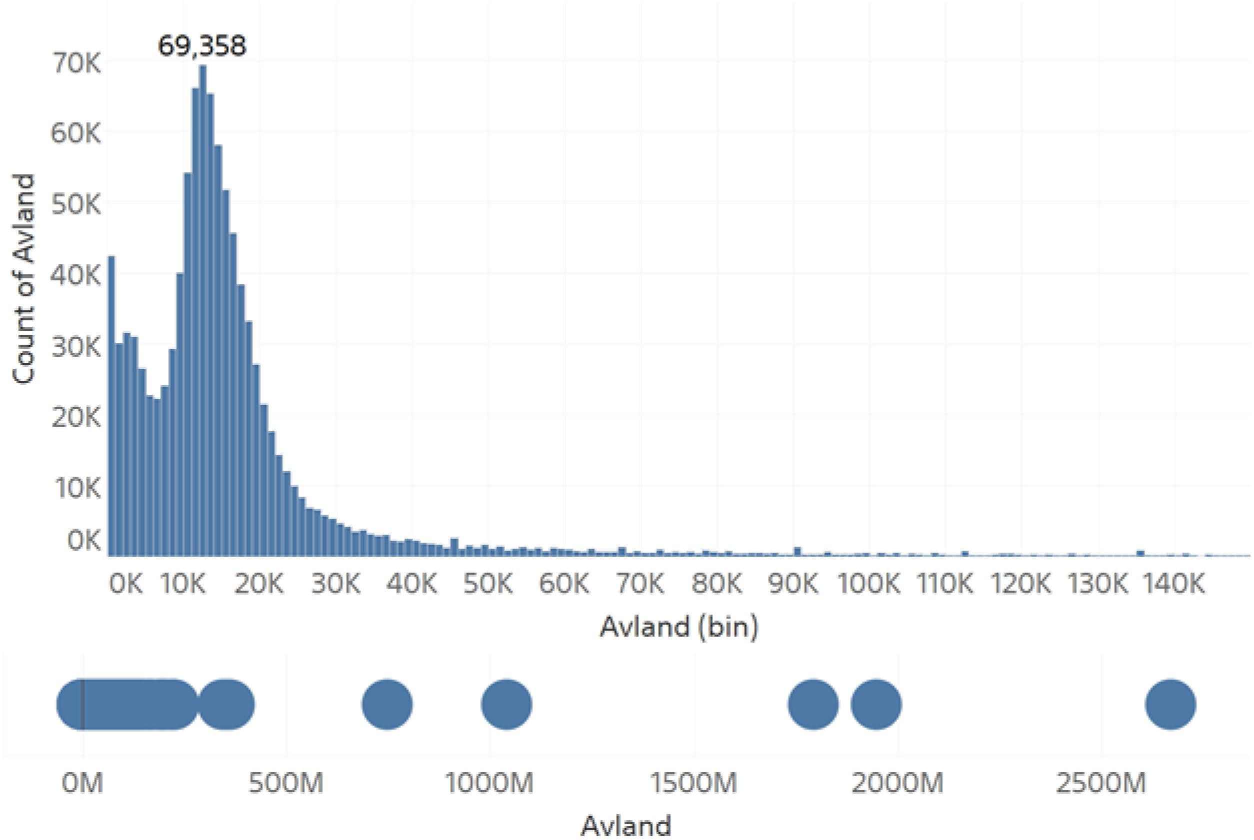




Based on Graph 13, FULLVAL appears to have a bimodal distribution and have a heavy skew to the right. The boxplot below Graph 13 supports the existence of a heavy skew due to a few outliers in the billions. There is an upper whisker at 818,734 which results in 135,798 outliers.

15.AVLAND

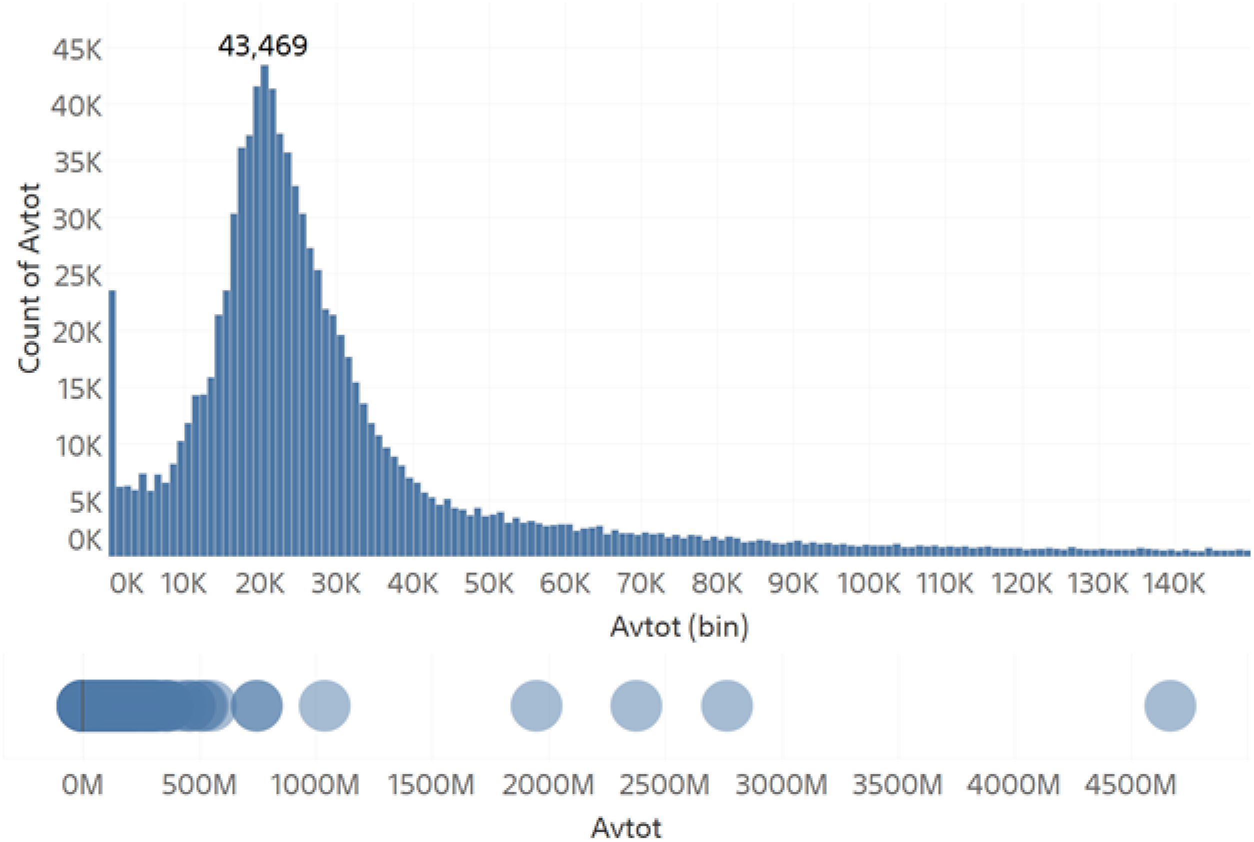
AVLAND is the current year's total market value of the land. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 14 below shows the histogram of AVLAND followed by its boxplot:



Based on Graph 14, AVLAND appears to have a bimodal distribution and have a heavy skew to the right. The boxplot below Graph 13 supports the existence of a heavy skew due to a few outliers in the billions. There is an upper whisker at 192,998 which results in 39,315 outliers.

16.AVTOT

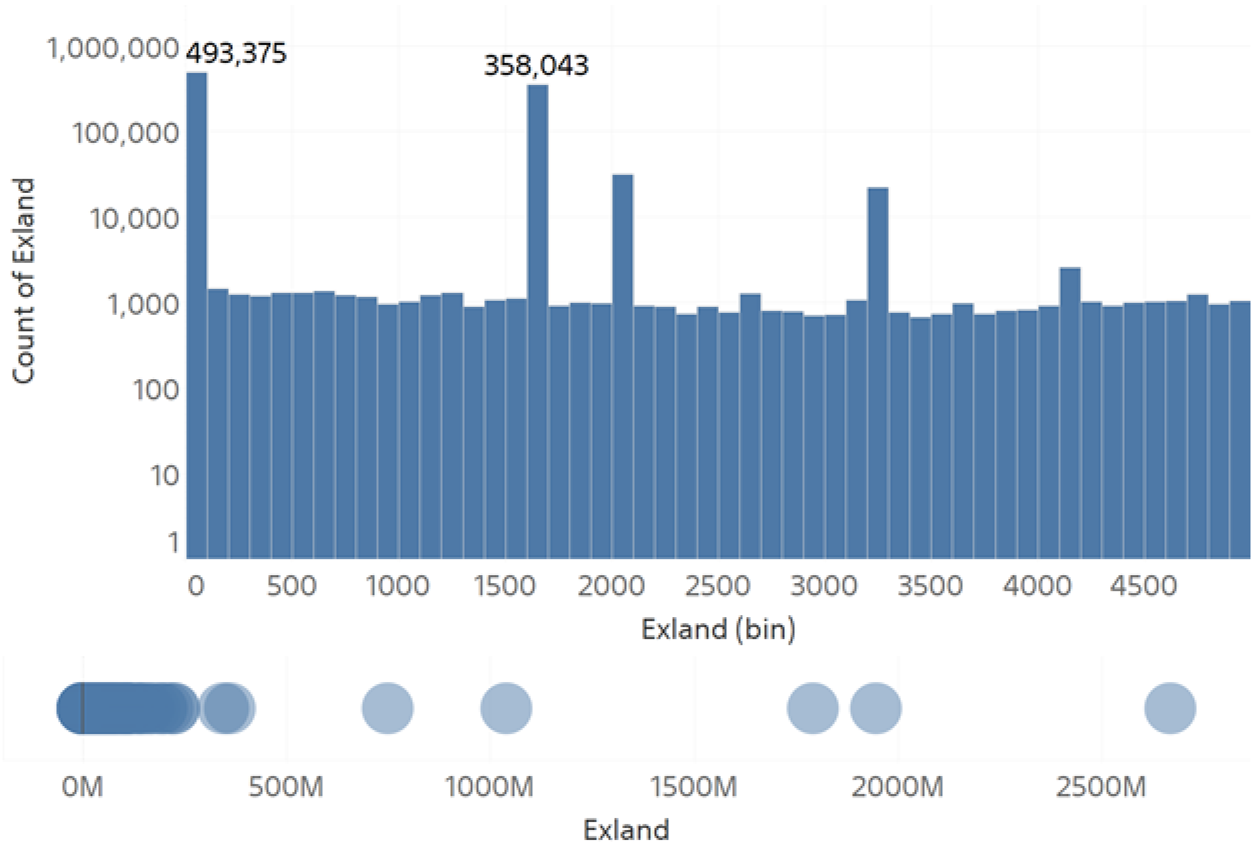
AVTOT is the current year's total market value of the property. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 15 below shows the histogram of AVTOT followed by its boxplot:



Based on Graph 15, AVTOT appears to have a heavy skew to the right. The boxplot below Graph 15 supports the existence of a heavy skew due to a few outliers in the billions. There is an upper whisker at 335,025 which results in 62,127 outliers.

17.EXLAND

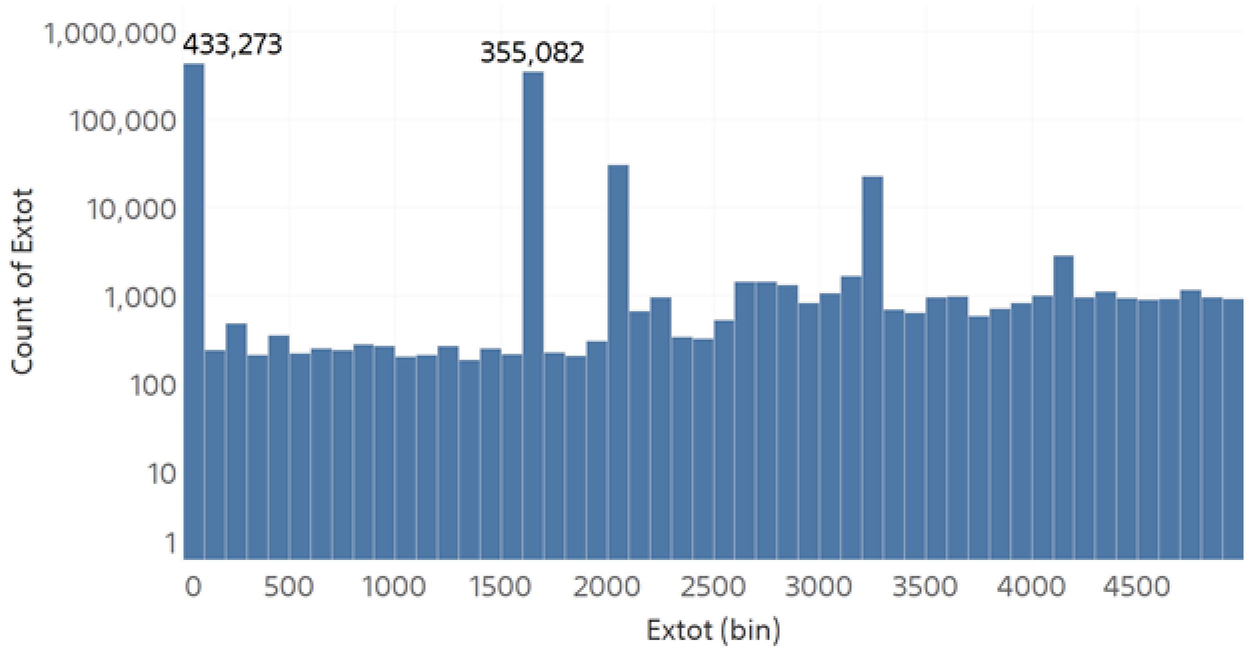
EXLAND is the current year's exempt land value. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 16 below shows the histogram of EXLAND followed by its boxplot:



Based on Graph 16, EXLAND follows a non-normal distribution. There is an upper whisker at 109,521 which results in 14,747 outliers.

18.EXTOT

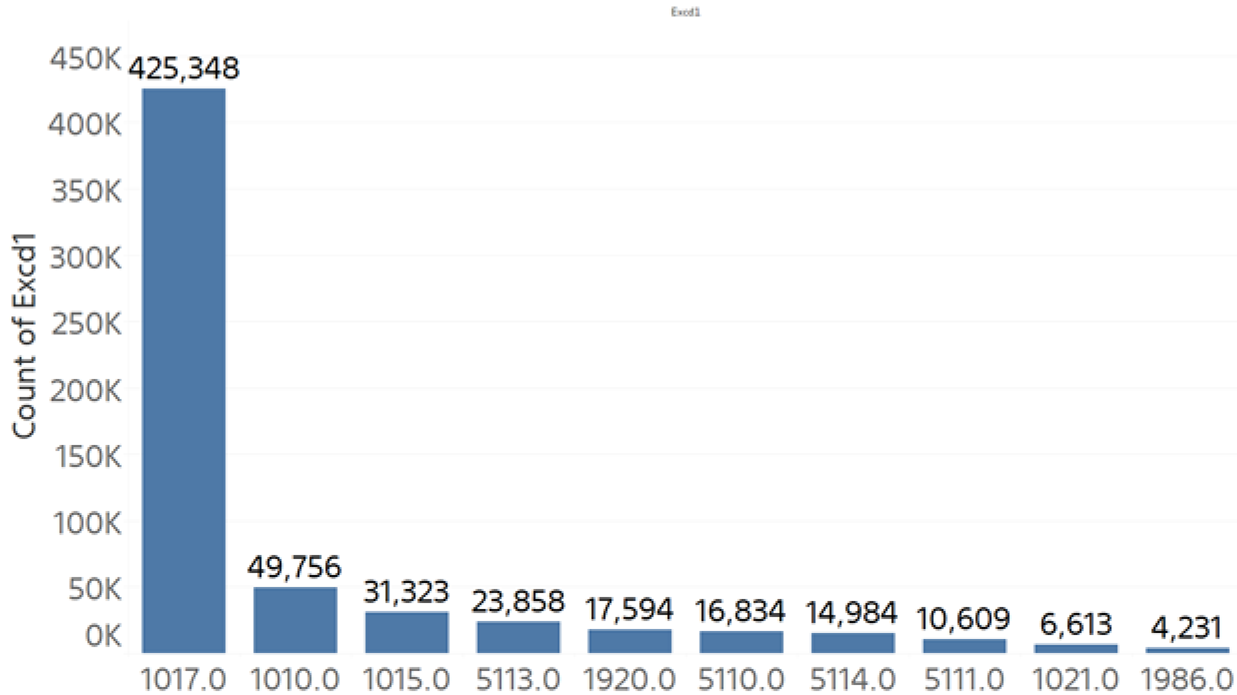
EXTOT is the current year's exempt total value. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 17 below shows the histogram of EXTOT followed by its boxplot:



Like EXLAND, Graph 17 shows that EXTOT also follows a non-normal distribution. There is an upper whisker at 297,263 which results in 18,419 outliers.

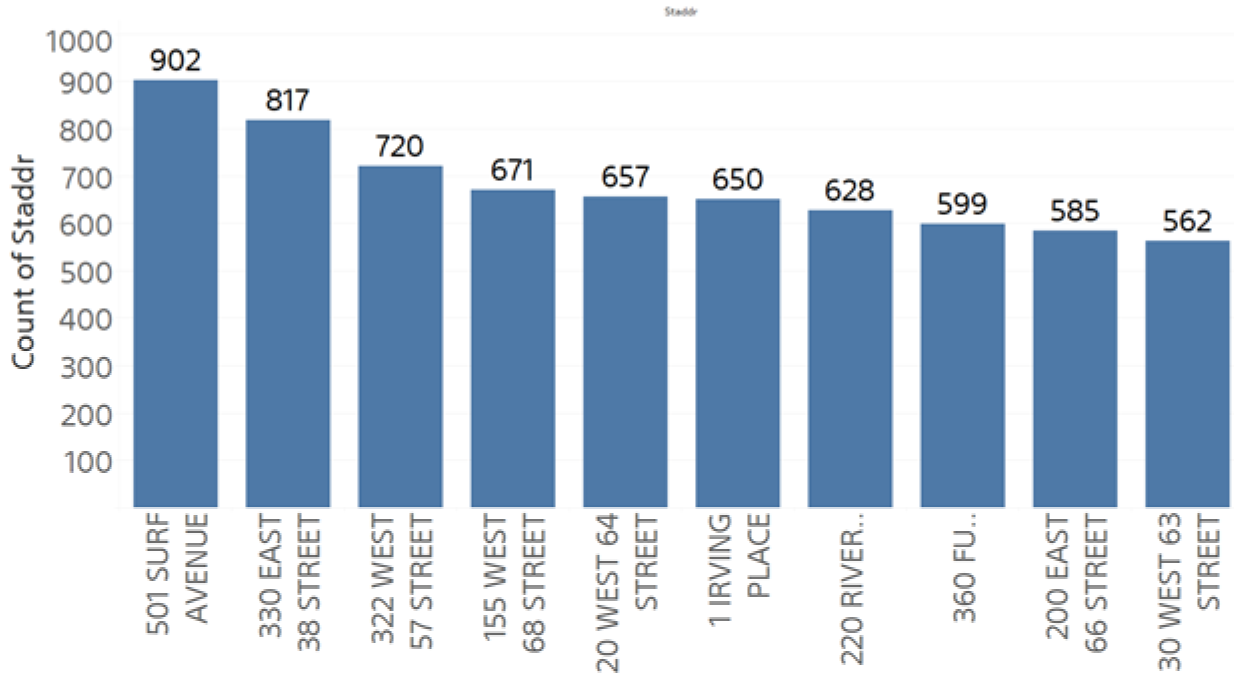
19.EXCD1

EXCD1 is a categorical field that identifies the exemption code relevant to assessed value. Please see Exhibit 4 for the NYC Exemption Classification Codes. Graph 18 below shows the bar graph of the top exemption codes:



20.STADDR

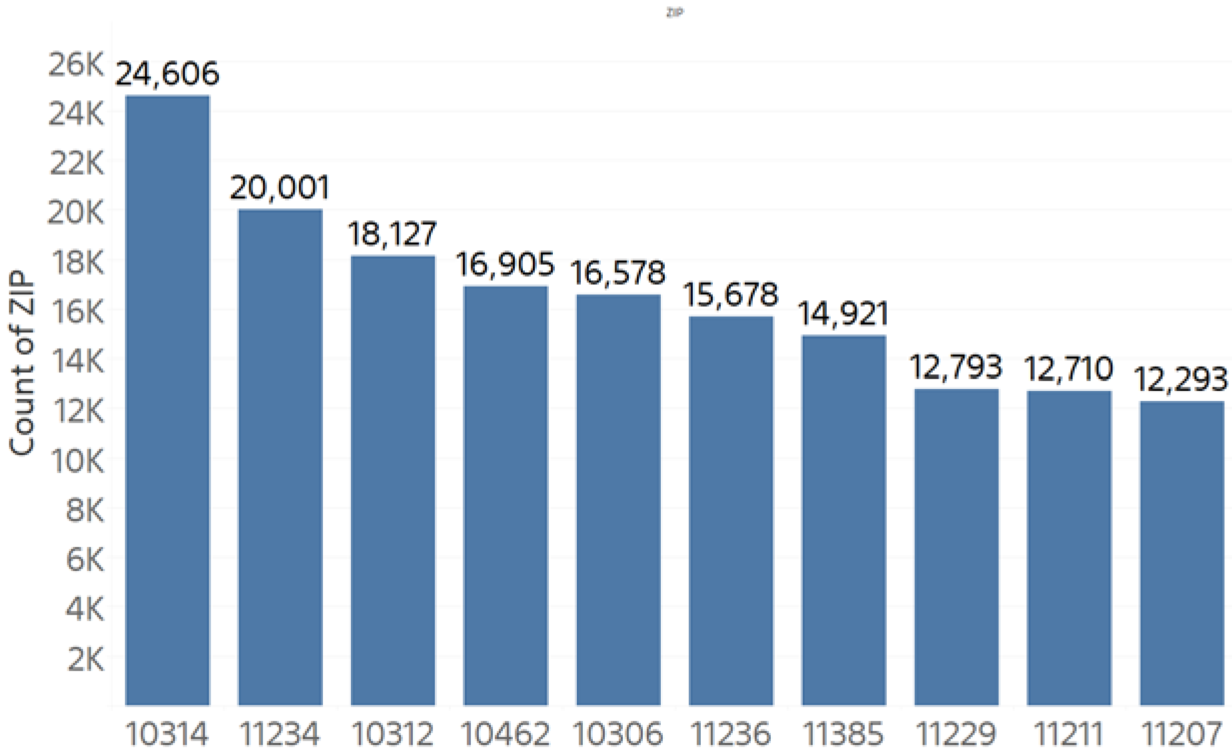
STADDR is a categorical field for street address of the property where the entire address contains the street number, direction, name, suffix, and apartment number. Graph 19 below shows the bar graph of the top street addresses:



Graph 19 shows that at least ten street addresses possess either multiple properties or homes. It is a possibility that these addresses for multi-family structures. More research is needed to further verify that.

21.ZIP

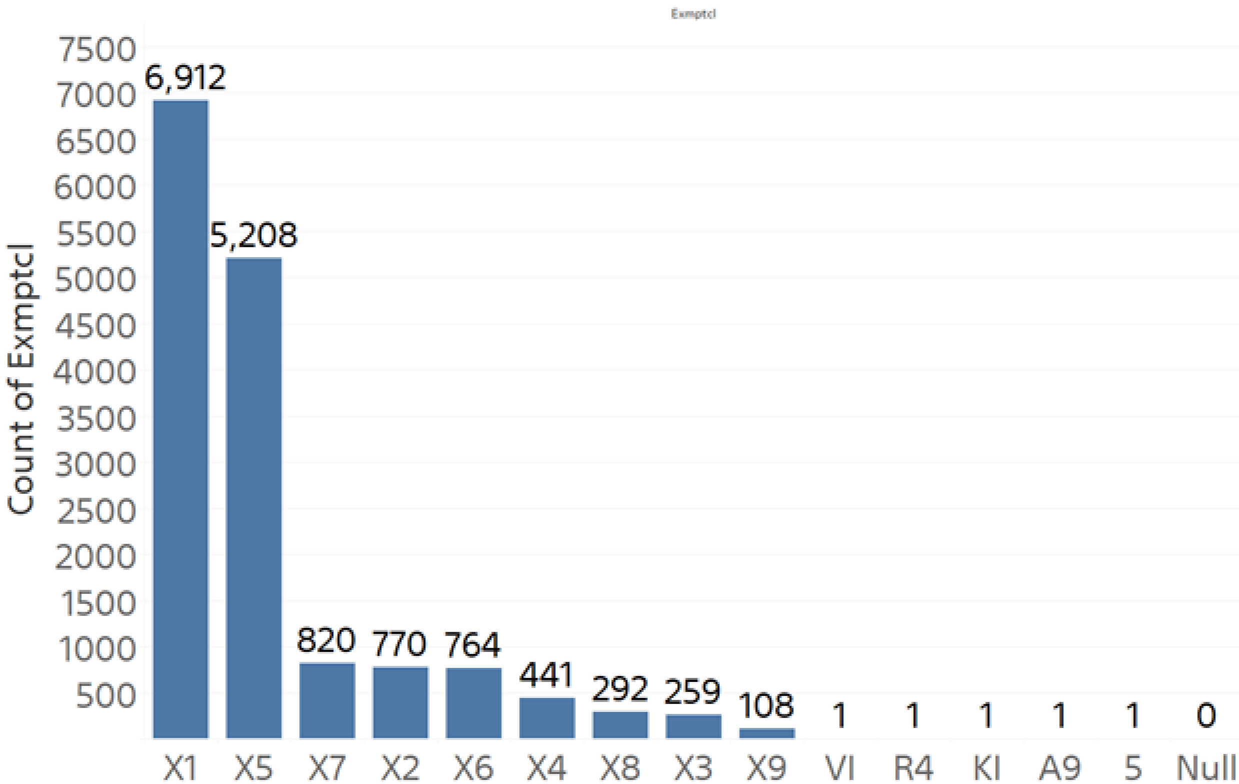
ZIP is a categorical field that is the ZIP code for the property. The ZIP code is a useful identifier in determining city and linking with other information. Graph 20 below shows the bar graph of the top ZIP codes:



Graph 20 does not reveal much other than a potential skew. Similar to the Borough map, a density map of ZIP codes could reveal insightful information and may be worthwhile to look into.

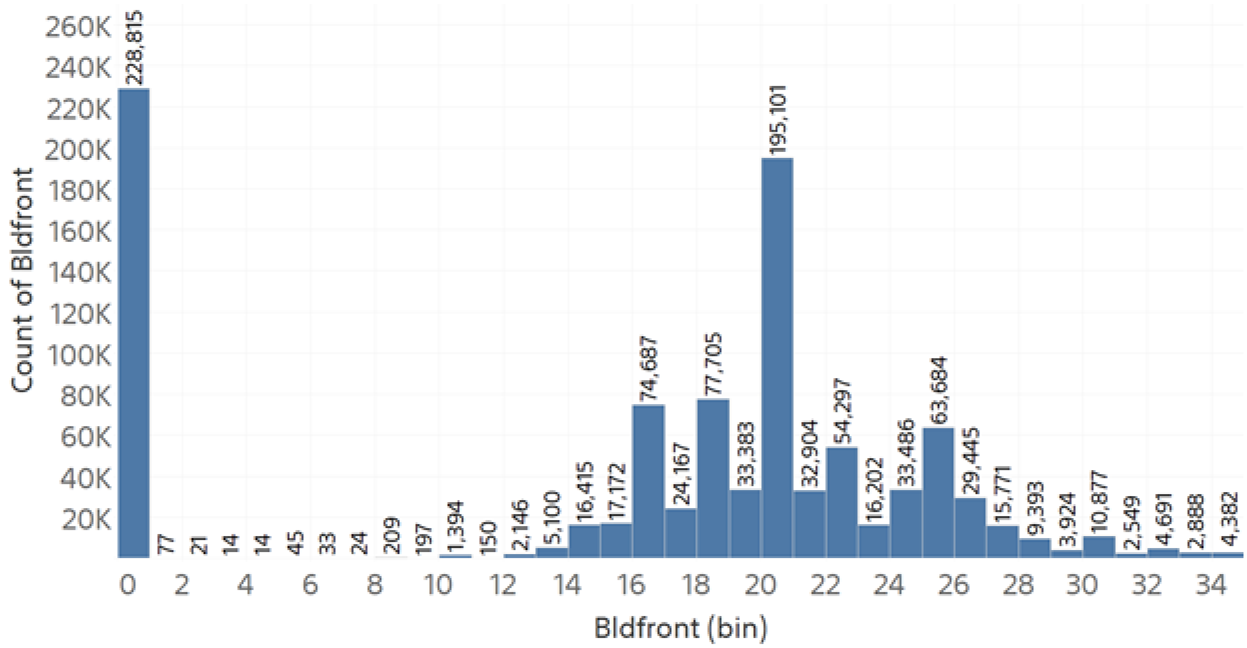
22.EXMPTCL

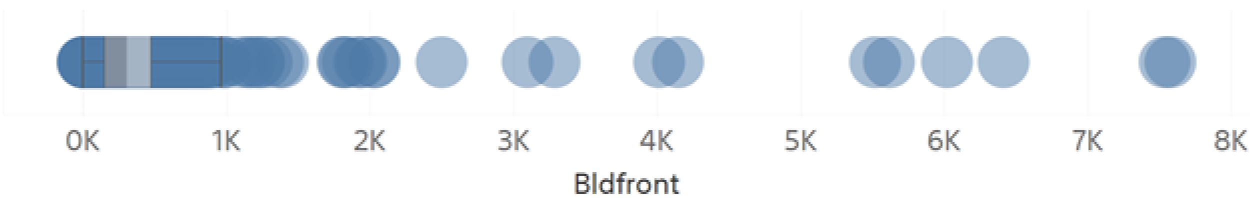
EXMPTCL is a categorical field for the exempt class which is used for fully exempt properties. If entered, the field should be between X1 and X9. Graph 21 below shows the bar graph of the top exempt class:



As shown in Graph 21, X1 and X5 are the top exemption classes.

23.BLDFRONT

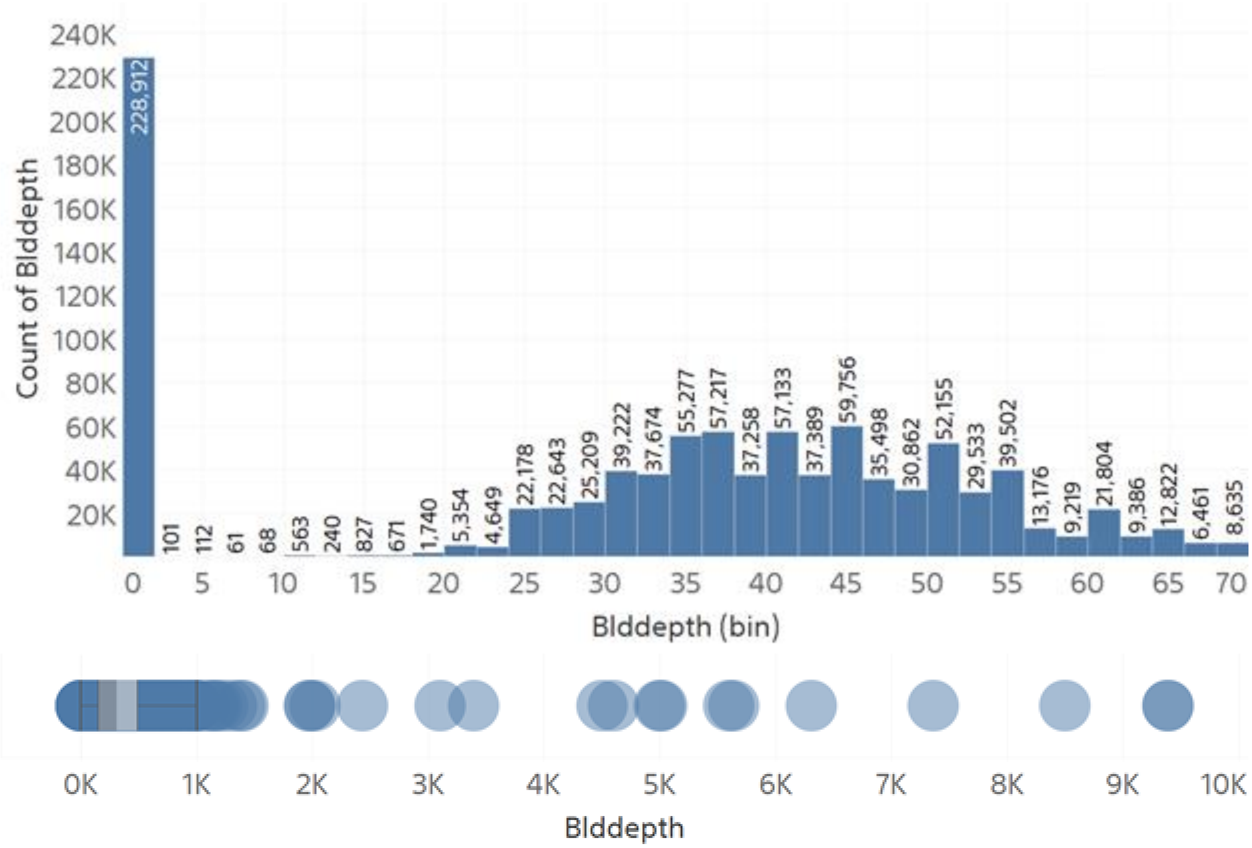




Based on Graph 22, BLDFRONT appears to potentially follow a normal distribution. However as shown in the boxplot below, it also suffers from a heavy skew to the right due to outliers. There is an upper whisker at 961 which results in 30 outliers.

24.BLDDEPTH

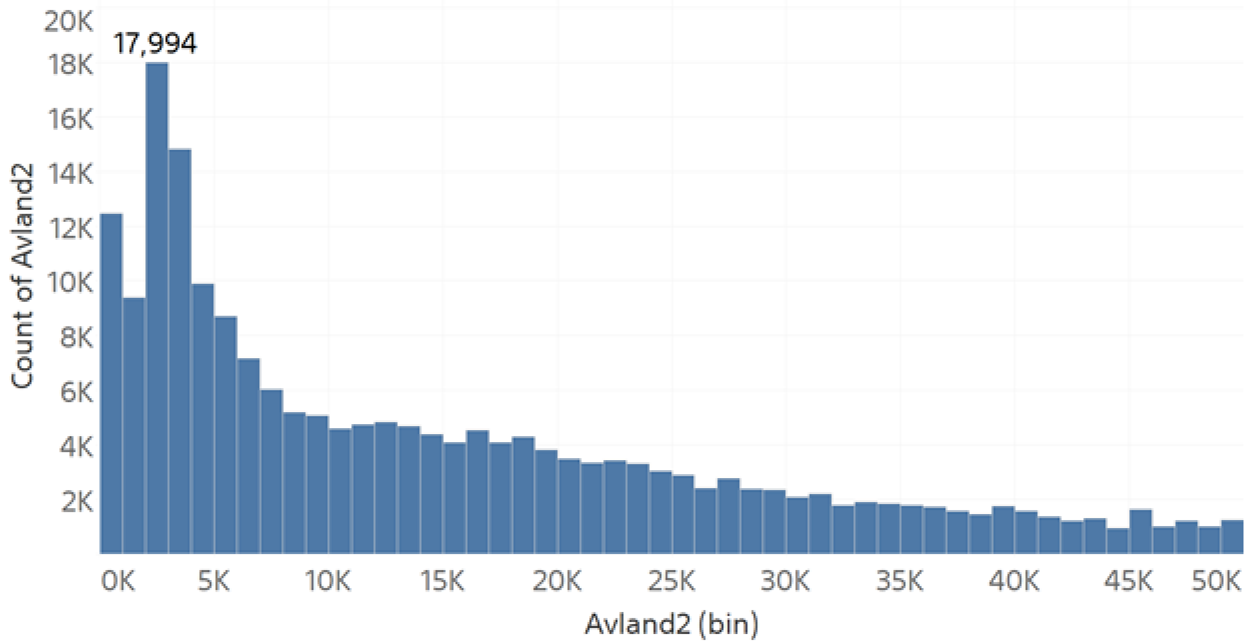
BLDDEPTH is the depth of the building in feet. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 23 below shows the histogram of BLDGFRONT followed by its boxplot:



Like BLDGFRONT, Graph 23 shows that BLDDEPTH may follow a normal distribution. It's boxplot also shows that it suffers a skew to the right due to outliers. There is an upper whisker at 1,000 which results in 28 outliers.

25.AVLAND2

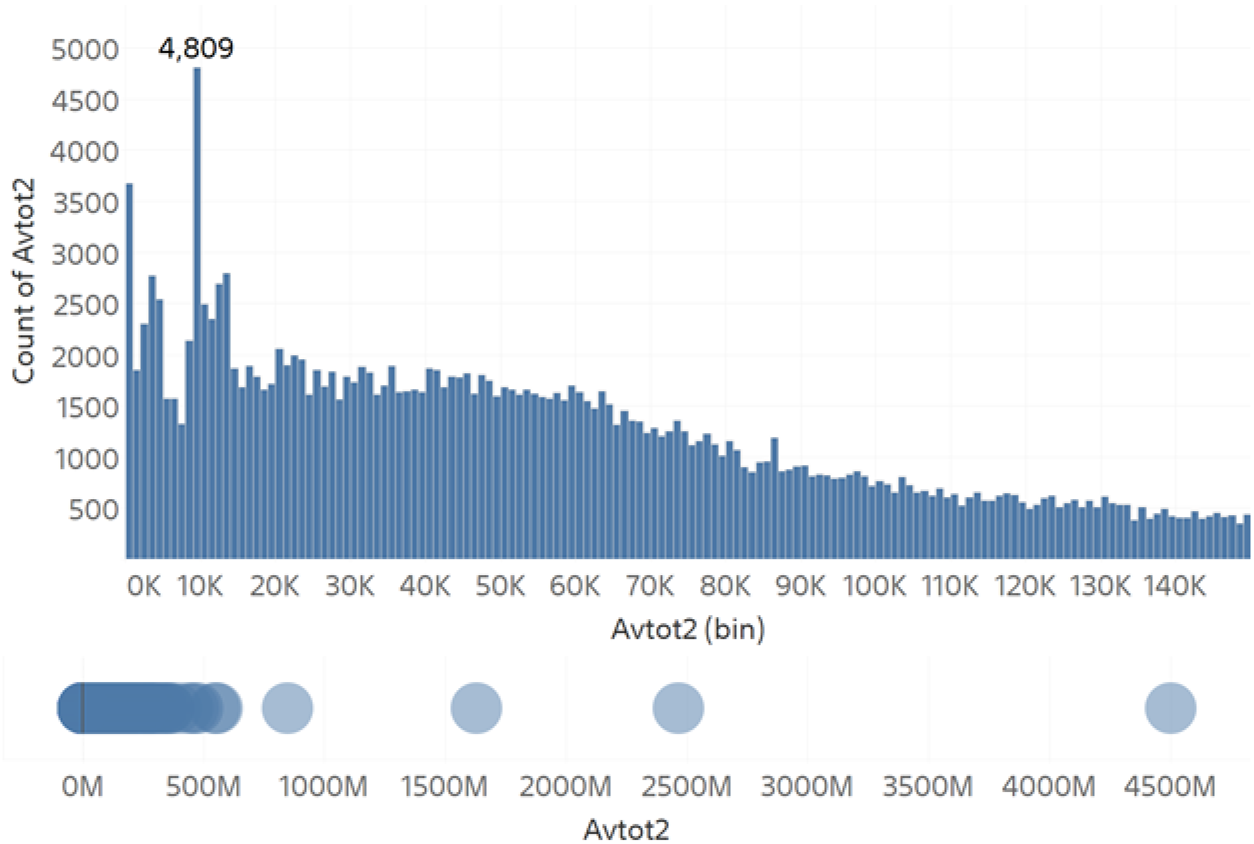
Current year's transitional assessed land value. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 24 below shows the histogram of AVLAND2 followed by its boxplot:



Based on Graph 24, AVLAND2 may have a bimodal distribution but the key takeaway is that it is heavily skewed to the right. This is supported by the boxplot showing outliers in hundreds of millions to billions. There is an upper whisker at 233,910 which results in 29,028 outliers.

26.AVTOT2

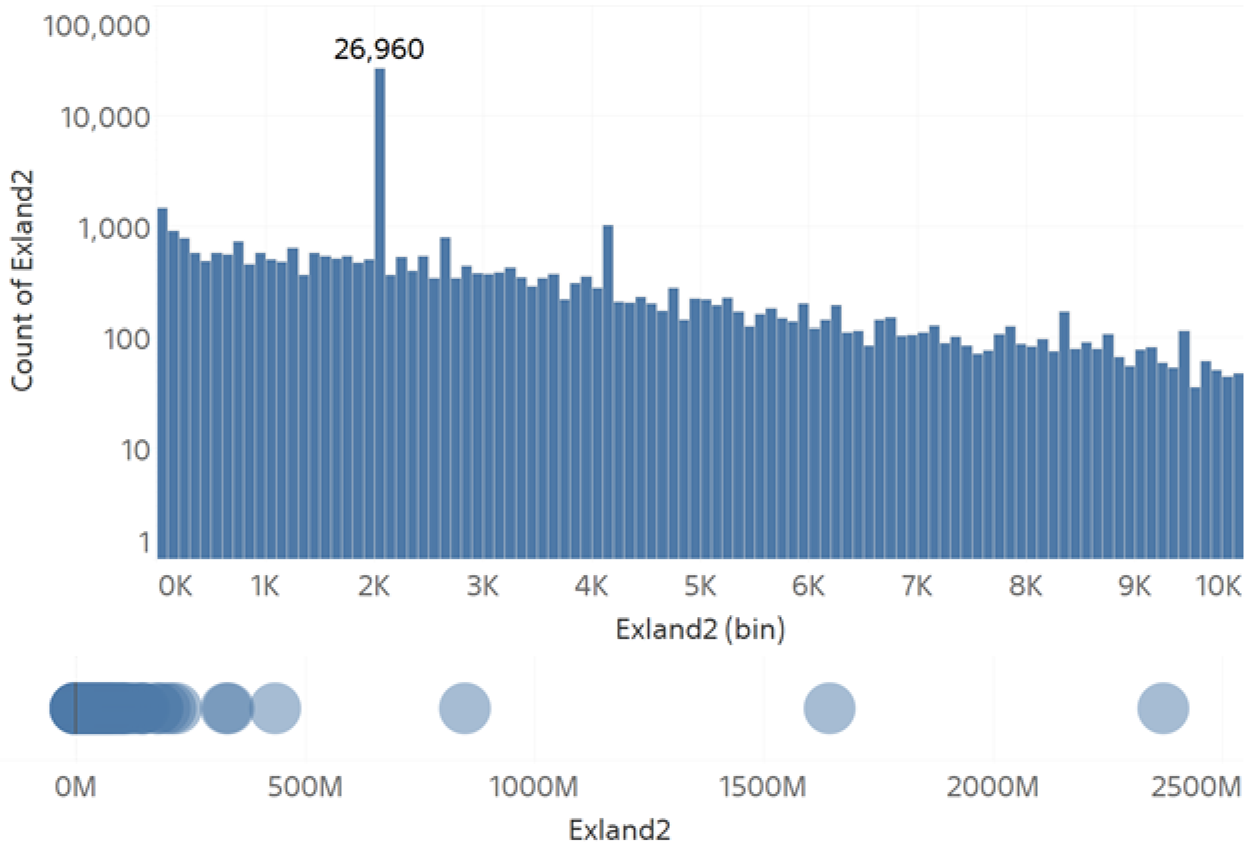
Current year's transitional assessed total value. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 25 below shows the histogram of AVTOT2 followed by its boxplot:



Based on Graph 25, AVTOT2 may have a binomial distribution but the key takeaway is that it has a skew to the right. This is supported by the boxplot showing outliers in hundreds of millions to billions. There is an upper whisker at 860,760 which results in 26,880 outliers.

27.EXLAND2

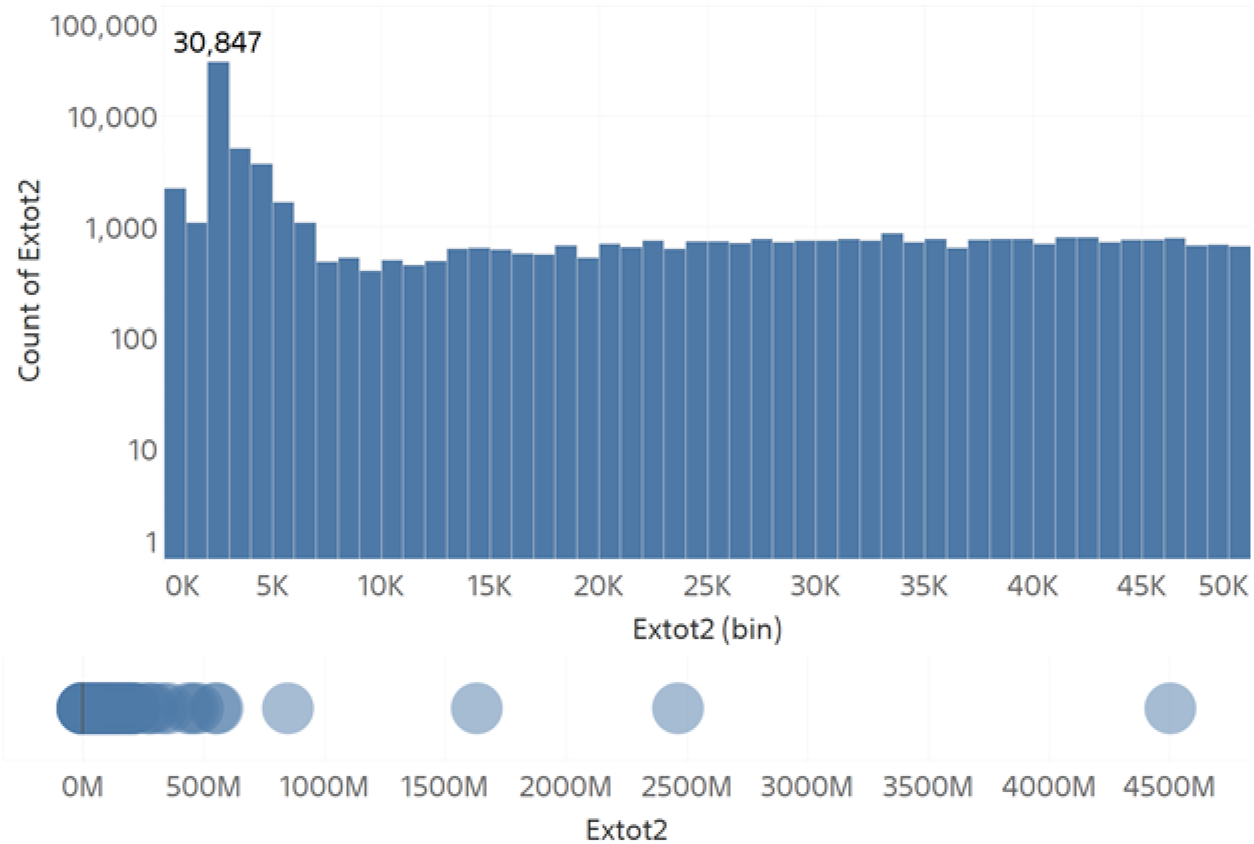
Current year's transitional exempt land value. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 26 below shows the histogram of EXLAND2 followed by its boxplot:



Based on Graph 26, EXLAND2 appears to have a long tail to the right. This is supported by its boxplot with values in the millions to billions. There is an upper whisker at 443,250 which results in 6,348 outliers.

28.EXTOT2

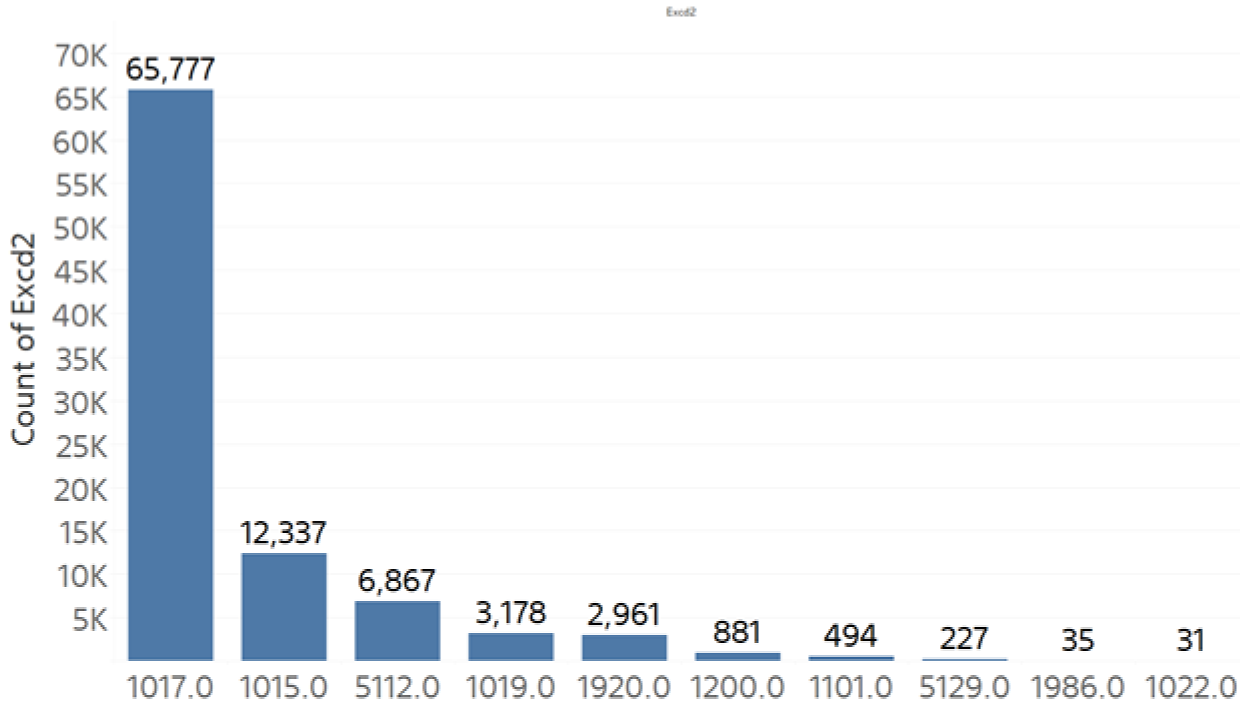
Current year's transitional exempt total value. It is a numeric field and extra insight may be provided from a histogram and boxplot. Graph 27 below shows the histogram of EXTOT2 followed by its boxplot:



Based on Graph 27, it is hard to determine the much of a pattern after the initial peak and subsequent decline. However, the boxplot helps to reveal a skew to the right due to outliers. There is an upper whisker at 650,853 which results in 11,241 outliers.

29.EXCD2

EXCD2 is a categorical field that identifies the exemption code relevant to transitional assessed value. Please see Exhibit 4 for the NYC Exemption Classification Codes. Graph 28 below shows the bar graph of the top exempt class:



Based on Graph 28, 1017 is the most popular exemption code for transitional assessed value; it is associated with school tax relief.

30.PERIOD

Period is a categorical field that might identify with the status of the Assessor's Roll for tax release purposes. Although there is uncertainty in the meaning of the field, every tuple has the same value (Final) and thus the field would only follow a uniform distribution. A graph or table would not provide any greater insight into the data and thus, was not included.

31.YEAR

YEAR is a categorical field that is likely associated with the initial fiscal year it came into the roll. In this case, all tuples are filled with the value '2010/11,' which likely to represent the 2010/2011 fiscal year (which began July 1, 2010 and ended June 30, 2011). This results in the field following a uniform distribution. A graph or table would not provide any greater insight into the data and thus, was not included.

32.VALTYPE

This field identifies what values are provided in the data. Every tuple has the same value, AC-TR, which stands for Actual-Transitional for Actual and Transitional Assessed Values. Since all the tuples have the same value, the field has a uniform distribution. A graph or table would not provide any greater insight into the data and thus, was not included.

Exhibit 1: Sources

NYC Department of Finance (2019). Building Classification Codes.
Retrieved January 20, 2019 from https://www1.nyc.gov/assets/finance/jump/hlpbldgcode.html

NYC Department of Finance (2015). Data Dictionary for the RPAD Master File Fixed Portion Only CD-ROM Format. Retrieved January 19, 2019 from
https://www1.nyc.gov/assets/finance/downloads/tar/tarfieldcodes.pdf

NYC OpenData (2019). Exemption Classification Codes. Retrieved January 20, 2019 from
https://data.cityofnewyork.us/City-Government/Exemption-Classification-Codes/myn9-hwsy

NYC Department of Finance (2007). Glossary Terms for Property Sales Files.
Retrieved January 19, 2019 from
https://www1.nyc.gov/assets/finance/downloads/pdf/07pdf/glossary_rsf071607.pdf

NYC Department of Finance (2017). NYC Residential property taxes: class one.
Retrieved January 20, 2019, from
https://www1.nyc.gov/assets/finance/downloads/pdf/brochures/class_1_guide.pdf

NYC OpenData (2019). Property Valuation and Assessment Data. Retrieved January 17, 2019 from
https://data.cityofnewyork.us/Housing-Development/Property-Valuation-and-Assessment-Data/rgy2-tti8

Exhibit 2: NYC Building Classifications

Building Code	Description
A0	CAPE COD
A1	TWO STORIES - DETACHED SM OR MID
A2	ONE STORY - PERMANENT LIVING QUARTER
A3	LARGE SUBURBAN RESIDENCE
A4	CITY RESIDENCE ONE FAMILY
A5	ONE FAMILY ATTACHED OR SEMI-DETACHED
A6	SUMMER COTTAGE
A7	MANSION TYPE OR TOWNHOUSE
A8	BUNGALOW COLONY - COOPERATIVELY OWNED LAND
A9	MISCELLANEOUS ONE FAMILY
B1	TWO FAMILY BRICK
B2	TWO FAMILY FRAME
B3	TWO FAMILY CONVERTED FROM ONE FAMILY
B9	MISCELLANEOUS TWO FAMILY
C0	THREE FAMILIES
C1	OVER SIX FAMILIES WITHOUT STORES
C2	FIVE TO SIX FAMILIES
C3	FOUR FAMILIES
C4	OLD LAW TENEMENT
C5	CONVERTED DWELLINGS OR ROOMING HOUSE
C6	WALK-UP COOPERATIVE
C7	WALK-UP APT. OVER SIX FAMILIES WITH STORES
C8	WALK-UP CO-OP; CONVERSION FROM LOFT/WAREHOUSE
C9	GARDEN APARTMENTS
CM	MOBILE HOMES/TRAILER PARKS
D0	ELEVATOR CO-OP; CONVERSION FROM LOFT/WAREHOUSE
D1	ELEVATOR APT; SEMI-FIREPROOF WITHOUT STORES
D2	ELEVATOR APT; ARTISTS IN RESIDENCE
D3	ELEVATOR APT; FIREPROOF WITHOUT STORES
D4	ELEVATOR COOPERATIVE
D5	ELEVATOR APT; CONVERTED
D6	ELEVATOR APT; FIREPROOF WITH STORES
D7	ELEVATOR APT; SEMI-FIREPROOF WITH STORES

Building Code	Description
D8	ELEVATOR APT; LUXURY TYPE
D9	ELEVATOR APT; MISCELLANEOUS
E1	FIREPROOF WAREHOUSE
E2	CONTRACTORS WAREHOUSE
E3	SEMI-FIREPROOF WAREHOUSE
E4	METAL FRAME WAREHOUSE
E7	SELF-STORAGE WAREHOUSES
E9	MISCELLANEOUS WAREHOUSE
F1	FACTORY; HEAVY MANUFACTURING - FIREPROOF
F2	FACTORY; SPECIAL CONSTRUCTION - FIREPROOF
F4	FACTORY; INDUSTRIAL SEMI-FIREPROOF
F5	FACTORY; LIGHT MANUFACTURING
F8	FACTORY; TANK FARM
F9	FACTORY; INDUSTRIAL-MISCELLANEOUS
G0	GARAGE; RESIDENTIAL TAX CLASS 1
G1	ALL PARKING GARAGES
G2	AUTO BODY/COLLISION OR AUTO REPAIR
G3	GAS STATION WITH RETAIL STORE
G4	GAS STATION WITH SERVICE/AUTO REPAIR
G5	GAS STATION ONLY WITH/WITHOUT SMALL KIOSK
G6	LICENSED PARKING LOT
G7	UNLICENSED PARKING LOT
G8	CAR SALES/RENTAL WITH SHOWROOM
G9	MISCELLANEOUS GARAGE OR GAS STATION
GU	CAR SALES/RENTAL WITHOUT SHOWROOM
G9	CAR WASH OR LUBRITORIUM FACILITY
HB	BOUTIQUE: 10-100 ROOMS, W/LUXURY FACILITIES
HH	HOSTELS- BED RENTALS IN DORMITORY
HR	SRO- 1 OR 2 PEOPLE HOUSED IN INDIVIDUAL ROOMS (AFFORDABLE HOUSING)
HS	EXTENDED STAY/SUITE: AMENITIES SIMILAR TO APT
H1	LUXURY HOTEL
H2	FULL SERVICE HOTEL
H3	LIMITED SERVICE; MANY AFFILIATED WITH NATIONAL CHAIN
H4	MOTEL
H5	HOTEL; PRIVATE CLUB, LUXURY TYPE

Building Code	Description
H6	APARTMENT HOTEL
H7	APARTMENT HOTEL - COOPERATIVELY OWNED
H8	DORMITORY
H9	MISCELLANEOUS HOTEL
I1	HOSPITAL, SANITARIUM, MENTAL INSTITUTION
I2	INFIRMARY
I3	DISPENSARY
I4	HOSPITAL; STAFF FACILITY
I5	HEALTH CENTER, CHILD CENTER, CLINIC
I6	NURSING HOME
I7	ADULT CARE FACILITY
I9	MISCELLANEOUS HOSPITAL, HEALTH CARE FACILITY
J1	THEATRE; ART TYPE LESS THAN 400 SEATS
J2	THEATRE; ART TYPE MORE THAN 400 SEATS
J3	MOTION PICTURE THEATRE WITH BALCONY
J4	LEGITIMATE THEATRE, SOLE USE
J5	THEATRE IN MIXED-USE BUILDING
J6	TELEVISION STUDIO
J7	OFF BROADWAY TYPE THEATRE
J8	MULTIPLEX PICTURE THEATRE
J9	MISCELLANEOUS THEATRE
K1	ONE STORY RETAIL BUILDING
K2	MULTI-STORY RETAIL BUILDING (2 OR MORE)
K3	MULTI-STORY DEPARTMENT STORE
K4	PREDOMINANT RETAIL WITH OTHER USES
K5	STAND-ALONE FOOD ESTABLISHMENT
K6	SHOPPING CENTER WITH OR WITHOUT PARKING
K7	BANKING FACILITIES WITH OR WITHOUT PARKING
K8	BIG BOX RETAIL: NOT AFFIXED & STANDING ON OWN LOT W/PARKING
K9	MISCELLANEOUS STORE BUILDING
L1	LOFT; OVER 8 STORIES (MID MANH. TYPE)
L2	LOFT; FIREPROOF AND STORAGE TYPE WITHOUT STORES
L3	LOFT; SEMI-FIREPROOF
L8	LOFT; WITH RETAIL STORES OTHER THAN TYPE ONE
L9	MISCELLANEOUS LOFT

Building Code	Description
M1	CHURCH, SYNAGOGUE, CHAPEL
M2	MISSION HOUSE (NON-RESIDENTIAL)
M3	PARSONAGE, RECTORY
M4	CONVENT
M9	MISCELLANEOUS RELIGIOUS FACILITY
N1	ASYLUM
N2	HOME FOR INDIGENT CHILDREN, AGED, HOMELESS
N3	ORPHANAGE
N4	DETENTION HOUSE FOR WAYWARD GIRLS
N9	MISCELLANEOUS ASYLUM, HOME
O1	OFFICE ONLY - 1 STORY
O2	OFFICE ONLY 2 - 6 STORIES
O3	OFFICE ONLY 7 - 19 STORIES
O4	OFFICE ONLY WITH OR WITHOUT COMM - 20 STORIES OR MORE
O5	OFFICE WITH COMM - 1 TO 6 STORIES
O6	OFFICE WITH COMM 7 - 19 STORIES
O7	PROFESSIONAL BUILDINGS/STAND ALONE FUNERAL HOMES
O8	OFFICE WITH APARTMENTS ONLY (NO COMM)
O9	MISCELLANEOUS AND OLD STYLE BANK BLDGS.
P1	CONCERT HALL
P2	LODGE ROOM
P3	YWCA, YMCA, YWHA, YMHA, PAL
P4	BEACH CLUB
P5	COMMUNITY CENTER
P6	AMUSEMENT PLACE, BATH HOUSE, BOAT HOUSE
P7	MUSEUM
P8	LIBRARY
P9	MISCELLANEOUS INDOOR PUBLIC ASSEMBLY
Q1	PARKS/RECREATION FACILTY
Q2	PLAYGROUND
Q3	OUTDOOR POOL
Q4	BEACH
Q5	GOLF COURSE
Q6	STADIUM, RACE TRACK, BASEBALL FIELD

Building Code	Description
Q7	TENNIS COURT
Q8	MARINA, YACHT CLUB
Q9	MISCELLANEOUS OUTDOOR RECREATIONAL FACILITY
RA	CULTURAL, MEDICAL, EDUCATIONAL, ETC.
RB	OFFICE SPACE
RG	INDOOR PARKING
RH	HOTEL/BOATEL
RK	RETAIL SPACE
RP	OUTDOOR PARKING
RR	CONDOMINIUM RENTALS
RS	NON-BUSINESS STORAGE SPACE
RT	TERRACES/GARDENS/CABANAS
RW	WAREHOUSE/FACTORY/INDUSTRIAL
R0	SPECIAL CONDOMINIUM BILLING LOT
R1	CONDO; RESIDENTIAL UNIT IN 2-10 UNIT BLDG.
R2	CONDO; RESIDENTIAL UNIT IN WALK-UP BLDG.
R3	CONDO; RESIDENTIAL UNIT IN 1-3 STORY BLDG.
R4	CONDO; RESIDENTIAL UNIT IN ELEVATOR BLDG.
R5	MISCELLANEOUS COMMERCIAL
R6	CONDO; RESID.UNIT OF 1-3 UNIT BLDG-ORIG CLASS 1
R7	CONDO; COMML.UNIT OF 1-3 UNIT BLDG-ORIG CLASS 1
R8	CONDO; COMML.UNIT OF 2-10 UNIT BLDG.
R9	CO-OP WITHIN A CONDOMINIUM
RR	CONDO RENTALS
S0	PRIMARILY 1 FAMILY WITH 2 STORES OR OFFICES
S1	PRIMARILY 1 FAMILY WITH 1 STORE OR OFFICE
S2	PRIMARILY 2 FAMILY WITH 1 STORE OR OFFICE
S3	PRIMARILY 3 FAMILY WITH 1 STORE OR OFFICE
S4	PRIMARILY 4 FAMILY WITH 1 STORE OROFFICE
S5	PRIMARILY 5-6 FAMILY WITH 1 STORE OR OFFICE
S9	SINGLE OR MULTIPLE DWELLING WITH STORES OR OFFICES
T1	AIRPORT, AIRFIELD, TERMINAL
T2	PIER, DOCK, BULKHEAD
T9	MISCELLANEOUS TRANSPORTATION FACILITY

Building Code	Description
U0	UTILITY COMPANY LAND AND BUILDING
U1	BRIDGE, TUNNEL, HIGHWAY
U2	GAS OR ELECTRIC UTILITY
U3	CEILING RAILROAD
U4	TELEPHONE UTILITY
U5	COMMUNICATION FACILITY OTHER THAN TELEPHONE
U6	RAILROAD - PRIVATE OWNERSHIP
U7	TRANSPORTATION - PUBLIC OWNERSHIP
U8	REVOCABLE CONSENT
U9	MISCELLANEOUS UTILITY PROPERTY
V0	ZONED RESIDENTIAL; NOT MANHATTAN
V1	ZONED COMMERCIAL OR MANHATTAN RESIDENTIAL
V2	ZONED COMMERCIAL ADJACENT TO CLASS 1 DWELLING: NOT MANHATTAN
V3	ZONED PRIMARILY RESIDENTIAL; NOT MANHATTAN
V4	POLICE OR FIRE DEPARTMENT
V5	SCHOOL SITE OR YARD
V6	LIBRARY, HOSPITAL OR MUSEUM
V7	PORT AUTHORITY OF NEW YORK AND NEW JERSEY
V8	NEW YORK STATE OR US GOVERNMENT
V9	MISCELLANEOUS VACANT LAND
W1	PUBLIC ELEMENTARY, JUNIOR OR SENIOR HIGH
W2	PAROCHIAL SCHOOL, YESHIVA
W3	SCHOOL OR ACADEMY
W4	TRAINING SCHOOL
W5	CITY UNIVERSITY
W6	OTHER COLLEGE AND UNIVERSITY
W7	THEOLOGICAL SEMINARY
W8	OTHER PRIVATE SCHOOL
W9	MISCELLANEOUS EDUCATIONAL FACILITY
Y1	FIRE DEPARTMENT
Y2	POLICE DEPARTMENT
Y3	PRISON, JAIL, HOUSE OF DETENTION
Y4	MILITARY AND NAVAL INSTALLATION
Y5	DEPARTMENT OF REAL ESTATE
Y6	DEPARTMENT OF SANITATION
Y7	DEPARTMENT OF PORTS AND TERMINALS

Building Code	Description
Y8	DEPARTMENT OF PUBLIC WORKS
Y9	DEPARTMENT OF ENVIRONMENTAL PROTECTION
Z0	TENNIS COURT, POOL, SHED, ETC.
Z1	COURT HOUSE
Z2	PUBLIC PARKING AREA
Z3	POST OFFICE
Z4	FOREIGN GOVERNMENT
Z5	UNITED NATIONS
Z7	EASEMENT
Z8	CEMETERY
Z9	OTHER MISCELLANEOUS

Exhibit 3: NYC Tax Classes

Every property within NYC is assigned one of four tax classes (1, 2, 3, or 4) based on the use of the property. Below is the designation and their associated uses:

Class 1: Most residential property of up to three units, vacant land that is zoned for residential use, and most condominiums that are not more than three stories.

Class 2: All other property that is primary residential, such as cooperatives and condominiums.

Class 3: Property with equipment owned by gas, telephone, and/or electric company.

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

Each Tax Class is associated with different groups of building classes. Below are the groups of building classes that fall under each Tax Class:

Tax Class	Building Classes
1	A0 - A9, B1 - B9, C0, G0, R3, R6, R7, S0 - S2, V0, V2, V3, Z0
2	C1 - C9, D0 - D9, R0, R1, R2, R4, R8, R9, S3, S4, S5, S9
3	U1 - U2, U4 - U9
4	All Other Building Classes

Exhibit 4: NYC Exemption Classification Codes

Exempt Code	SDEA Code	Description	Status	Legal Ref
1010	41101	VETERAN	ACTIVE	RPTL § 458
1010	41121	New Law Veteran	ACTIVE	RPTL § 458
1010	41131	New Law Veteran	ACTIVE	RPTL § 458
1010	41141	New Law Veteran	ACTIVE	RPTL § 458
1011	41300	SER DISABLED VET	ACTIVE	RPTL §458(3)
1015	41800	SENIOR CITIZEN	ACTIVE	RPTL §467
1016	41910	CRIME VICTIMS	ACTIVE	RPTL §459-b
1017	41836	SCHOOL TAX RELIEF	ACTIVE	RPTL § 425

1017	41856	SCHOOL TAX RELIEF	ACTIVE	RPTL § 425
1019	41930	DISABLE HOMEOWNER	ACTIVE	RPTL § 459C
1021	25110	HOUSE OF WORSHIP	ACTIVE	RPTL § 420-a
1022	25120	RELIGIOUS-SCHOOL	ACTIVE	RPTL § 420-a
1023	25120	RELIGIOUS-DORMITORY	ACTIVE	RPTL § 420-a
1101	21600	PARSONAGE	ACTIVE	RPTL § 462
1102	21600	RELIGIOUS MISSIONS	ACTIVE	RPTL § 462
1200	41400	CLERGY	ACTIVE	RPTL § 460
1301	25800	420C HOUSING	ACTIVE	RPTL § 420-C
1401	25210	HOSPITAL	ACTIVE	RPTL § 420-a
1402	25210	HEALTH CENTER	ACTIVE	RPTL § 420-a
1403	25210	NURSING HOME	ACTIVE	RPTL § 420-a
1404	25210	HOSPITAL STAFF HSG	ACTIVE	RPTL § 420-a
1501	25130	CHARITABLE	ACTIVE	RPTL § 420-a
1502	25130	CHARITABLE	ACTIVE	RPTL § 420-a
1503	25130	CHARITABLE	ACTIVE	RPTL § 420-a
1504	25130	CHARITABLE HOUSING	ACTIVE	RPTL § 420-a
1505	25130	NFP-CONTEMP USE	ACTIVE	RPTL § 420-a
1511	25400	FRATERNAL ORGANIZ	ACTIVE	RPTL § 428
1521	25230	MENTAL-MORAL IMPROVE	ACTIVE	RPTL § 420-a
1522	25110	SALVATION ARMY	ACTIVE	RPTL § 420-a
1523	25130	CHARTBLE PHILAN	ACTIVE	RPTL § 420-a
1561	46450	VOL. FIRE CO.	ACTIVE	RPTL § 466
1562	26350	PATROL SALVAGE	ACTIVE	RPTL § 468
1571	26100	A.L.,VFW,CWV,JWV,ETC	ACTIVE	RPTL § 452
1572	29650	MEMORIAL ASSN	ACTIVE	RPTL § 442
1601	25120	COLLEGE-UNIVERSITY	ACTIVE	RPTL § 420-a
1602	25120	SCHOOL-ELEM,HS,ACAD	ACTIVE	RPTL § 420-a
1603	25120	STUDENT DORMITORY	ACTIVE	RPTL § 420-a
1604	25120	FACULTY STUDENT HSG	ACTIVE	RPTL § 420-a
1605	25120	MUSEUM	ACTIVE	RPTL § 420-a
1606	26000	COOPER UNION	ACTIVE	C279-59
1620	26500	INST OF ARTS SCI	ACTIVE	RPTL § 424
1630	29150	OPERA HOUSE	ACTIVE	RPTL § 426

Exempt Code	SDEA Code	Description	Status	Legal Ref
1640	29500	PERF ARTS BLDG	ACTIVE	RPTL § 427
1650	49200	THEATRICAL CORP	ACTIVE	RPTL § 432
1660	29450	ACADEMY OF MUSIC	ACTIVE	RPTL § 434
1700	27350	CEMETERY (PRIVATE	ACTIVE	RPTL §446
1840	25300	BIBLE	ACTIVE	RPTL § 420-B
1841	25300	TRACT	ACTIVE	RPTL § 420-B
1850	25300	BENEVOLENT	ACTIVE	RPTL § 420-B
1860	25300	INFIRMARY	ACTIVE	RPTL § 420-B
1870	25300	PUBLIC PLAYGROUND	ACTIVE	RPTL § 420-B
1871	25300	SUPVD. SPORTSMANSHIP	ACTIVE	RPTL § 420-B
1872	25300	ENF/LAW/CHILD/ANIMAL	ACTIVE	RPTL § 420-B
1880	25300	SCIENTIFIC	ACTIVE	RPTL § 420-B
1881	25300	LITERARY	ACTIVE	RPTL § 420-B
1882	25300	LIBRARY	ACTIVE	RPTL § 420-B
1890	25300	PATRIOTIC	ACTIVE	RPTL § 420-B
1891	25300	HISTORICAL	ACTIVE	RPTL § 420-B
1901	29600	PASSENGER TERMINAL	ACTIVE	RPTL § 476
1902	47200	CONSTRUCTION	ACTIVE	RPTL § 476
1903	47200	RAILROAD CEILING	ACTIVE	RPTL § 489
1904	47200	RAILROAD PASSENGER	ACTIVE	RPTL § 489
1905	47200	COMMUTER CEILING	ACTIVE	RPTL § 489
1920	48070	J-51 ALTERATION	ACTIVE	RPTL § 489
1925	48820	421g RES. CONVERS. LOW MANH	ACTIVE	RPTL § 421G
1930	47700	FALLOUT SHELTER	ACTIVE	RPTL § 479
1940	27500	JAMAICA WATER	ACTIVE	RPTL §485-d
1950	25500	NON-PROFIT MED DENT	ACTIVE	RPTL § 486
1951	25500	HIP CENTER	ACTIVE	RPTL § 486
1961	49530	INDUST WASTE FACIL	ACTIVE	RPTL § 477
1963	47900	ENVIRON PROT EX	ACTIVE	RPTL § 481
1964	47900	ENVIRON PROT EX	ACTIVE	RPTL § 481
1965	49500	SOLAR/WIND ENERGY	ACTIVE	RPTL § 487
1971	47750	LIMIT.ON TEL TEL EQ	ACTIVE	RPTL § 470

Exempt Code	SDEA Code	Description	Status	Legal Ref
1972	47760	TELECOM EQUIP	ACTIVE	S-471
1975	0	PVT PROP ON US LAND	ACTIVE	BUCK ACT
1981	47650	NEW IND OR COMM BLDG	ACTIVE	RPTL § 489
1982	47650	ALT IND OR COMM BLDG	ACTIVE	RPTL § 489
1984	47660	ICIP DEFERED PAYMENT	ACTIVE	RPTL § 489
1985	47660	ICIP REG. COMML EX	ACTIVE	RPTL § 489
1986	47660	ICIP IND/SPECIAL EX	ACTIVE	RPTL § 489
1988	47680	MIXED-USE - LOW MANH	ACTIVE	RPTL § 489A
1990	13940	TRUST FOR CULT RSRCE	ACTIVE	Art-Cult L Section 21
1992	49000	PROF. MAJ LEA SPORTS	ACTIVE	RPTL § 429
2100	13350	BOROUGH PRESIDENT	ACTIVE	RPTL § 406(1)
2120	13350	ARMORY	ACTIVE	RPTL § 406(1)
2131	13350	DEPT OF CORRECTION	ACTIVE	RPTL § 406(1)
2132	13350	POLICE DEPT	ACTIVE	RPTL § 406(1)
2133	13350	FIRE DEPT	ACTIVE	RPTL § 406(1)
2134	13350	POLICE FIRE	ACTIVE	RPTL § 406(1)
2151	13800	DEPT. OF EDUCATION	ACTIVE	RPTL § 408
2152	13350	BOARD HIGHER EDUC (CUNY)	ACTIVE	RPTL § 406(1)
2171	13350	DEPT OF SANITATION	ACTIVE	RPTL § 406(1)
2172	13350	DEPT WATER RESOURCES	ACTIVE	RPTL § 406(1)
2191	13350	PORTS TERMINALS	ACTIVE	RPTL § 406(1)
2198	14620	AIRFIELD -DO NOT USE	ACTIVE	RPTL § 406(1)
2201	13350	DEPT OF PUBLIC WORKS (DEP	ACTIVE	RPTL § 406(1)
2202	13350	DEPT OF REAL ESTATE (DCAS	ACTIVE	RPTL § 406(1)
2220	13350	DEPT OF SOC SERVICES (HRA	ACTIVE	RPTL § 406(1)
2231	13350	PARK	ACTIVE	RPTL § 406(1)
2232	13350	PUBLIC LIBRARY	ACTIVE	RPTL § 406(1)
2233	13350	PUBLIC MUSEUM	ACTIVE	RPTL § 406(1)
2234	13350	PUBLIC BEACH	ACTIVE	RPTL § 406(1)
2251	13350	DEPT OF HEALTH	ACTIVE	RPTL § 406(1)
2252	13950	HEALTH & HOSPITALS C	ACTIVE	McK U Con L Section 7400
2261	13350	DEPT OF TRAFFIC	ACTIVE	RPTL § 406(1)

Exempt Code	SDEA Code	Description	Status	Legal Ref
2262	13350	DEPT OF HIGHWAYS	ACTIVE	RPTL § 406(1)
2280	13350	HOUSE PRES DEVL ADM	ACTIVE	RPTL § 406(1)
2310	18040	URBAN RENEWAL	ACTIVE	Gen Muni Law § 506, 555, 560
2350	13350	NYC ECON DEV CORP	ACTIVE	RPTL § 406(1)
2351	13350	EDC - PILOT	ACTIVE	RPTL § 406(1)
2400	13920	NYC EDUC CONST FUND	ACTIVE	Education Law §468
2500	18020	NYC INDUSTRIAL DEV	ACTIVE	RPTL §412a
2501	18020	IDA - PILOT	ACTIVE	RPTL §412a
2600	13350	NYC EMPL RETIRE SYS	ACTIVE	RPTL §404(2)
3360	12100	MILITARY	ACTIVE	RPTL §400(1)/State Law §54
3380	12100	STATE HOSPITAL	ACTIVE	RPTL §404(1)
3390	12100	STATE LANDS BLDGS	ACTIVE	RPTL §404(1)
3400	12100	STATE PUBLIC WORKS	ACTIVE	RPTL §404(1)
3410	18180	NYS URBAN DEV (ESDC)	ACTIVE	RPTL §404(1)
3420	12100	ROOSEVELT ISLAND	ACTIVE	McK U Con Law § 6395
3500	12350	DORMITORY AUTHORITY	ACTIVE	Pub Auth Law §1685
3600	12150	STATE RETIRE SYSTEM	ACTIVE	RPTL §404(2)
3700	12450	NYS MED CARE FAC FIN	ACTIVE	McK U Con L Section 7421
3800	14000	SCHOOL CONST AUTH	ACTIVE	Pub Auth Law §1742
4500	14120	ARMED FORCES	ACTIVE	State L §59-g
4520	14110	POST OFFICE	ACTIVE	State L §54
4530	14100	LIGHTHOUSE	ACTIVE	RPTL §400(1)
4540	14100	FED CEMETERY	ACTIVE	RPTL §446
4550	14110	FED HOSPITAL	ACTIVE	State L §54
4600	14100	FED GOVT LAND BLDGS	ACTIVE	RPTL §400(1)
4650	27250	AMTRAK (FED SUBSI RR	ACTIVE	45 USC §546b
5090	18120	NYS Housing Finance Agency	ACTIVE	PHFL §45a
5100	48460	LTD PROFIT HSNG CO:LEASED	ACTIVE	PHFL §33(2), (3)
5101	18080	NYC HOUSING AUTH	ACTIVE	Pub Hsng L §§52(3), 52(5), 52(6)
5102	38260	STATE AIDED PUB HSG	ACTIVE	Pub Hsng L §§52(4), 52(5)
5103	18080	NYC HOUS AUTH-DUPLIC	ACTIVE	Pub Hsng L §§52(3), 52(5), 52(6)
5104	18080	DWELLING FROM FED	ACTIVE	Pub Hsng L §§52(3), 52(5), 52(6)

Exempt Code	SDEA Code	Description	Status	Legal Ref
5105	28120	ST ASSISTED PRIV HSG	ACTIVE	RPTL §422
5106	48660	HSG DEV FUND COMPANY (SRT)	ACTIVE	PHFL §577(3)
5107	48540	LIMITED DIVIDEND (SRT)	ACTIVE	PHFL §93(6)
5108	48670	REDEVELOPMENT (SRT)	ACTIVE	PHFL §§125, 127
5109	48650	MITCHELL-LAMA (SRT)	ACTIVE	PHFL §33(1)(c)(d)
5110	48800	421a (10 yr not cap)	ACTIVE	RPTL § 421A
5111	41950	421b	ACTIVE	RPTL § 421B
5112	48100	URBAN DEV. ACT PROJ.	ACTIVE	Gen Muny L §696
5113	48800	421a (15 yr not cap)	ACTIVE	RPTL § 421A
5114	48800	421a (25 yr not cap)	ACTIVE	RPTL § 421A
5115	48690	REDEV. PHASE OUT	ACTIVE	RPTL § 423
5116	48800	421a (20 yr not cap)	ACTIVE	RPTL § 421A
5117	48800	421a (10 yr cap)	ACTIVE	RPTL § 421A
5118	48800	421a (15 yr cap)	ACTIVE	RPTL § 421A
5129	48743	DIV OF ALT MGMT PROG	ACTIVE	PHFL §1106-h
5130	48743	Special Initiatives Program (SIP)	ACTIVE	PHFL §1106-h
6110	14610	PORT AUTH-WORLD TRAD	ACTIVE	McK U Con L §6611
6120	14640	PORT AUTH - INVALID	ACTIVE	
6130	14610	PATH-DO NOT USE	ACTIVE	McK U Con L §6611
6140	14620	PORT AUTH-AIR TERMIN	ACTIVE	McK U Con L §6635
6150	14630	PORT AUTH-BUS FACILT	ACTIVE	McK U Con L §7210
6160	14600	PORT AUTH-NARROWS BR	ACTIVE	McK U Con L §6562
6170	17000	PORT AUTH-INDUST DEV	ACTIVE	McK U Con L §7181
6180	14640	PORT AUTH-BRIDGE/TUN	ACTIVE	McK U Con L §6515
6200	12360	MTA - BRIDGE/TUNNEL	ACTIVE	Pub Auth Law §566
6320	12360	MTA - NYC TRANSIT	ACTIVE	Pub Auth Law §1216
6400	12360	MTA - LIRR/MN	ACTIVE	Pub Auth Law §1275
6500	19950	MUNICIPAL RAILROAD	ACTIVE	RPTL §456
6600	12350	NY STATE POWER AUTH	ACTIVE	Pub Auth L §1012
6700	12350	NY JOB DEVEL AUTH	ACTIVE	Pub Auth L §1806
6800	12350	BATTERY PARK AUTH	ACTIVE	Pub Auth L §1981
7120	14210	FOREIGN CONSULATE	ACTIVE	Vienna Convention on Consular Relations, Article 32
7150	14400	UNITED NATIONS	ACTIVE	RPTL §416
7160	14220	FOREIGN MISSION	ACTIVE	Vienna Convention on Consular Relations, Article

Exempt Code	SDEA Code	Description	Status	Legal Ref
7165	14200	FOREIGN EMBASSY	ACTIVE	RPTL §418
7166	14200	FOREIGN STAFF HSING	ACTIVE	22 U.S.C. 4305c
7170	14410	UN DEVELOP CORP	ACTIVE	McK U Con L §9613