Fraud Analytics Project 1: Unsupervised Fraud Algorithm on New York Property

Team 5:

Siqi Hu

Rufeng Lin

He Sun

Shihan Wang

Xindong Zhou

2020.04.29

Table of Contents

Executive Summary

With significant real estate appreciation in New York City these years, property owners become more likely to intentionally commit tax fraud. In detail, they under-report on a tax return in an attempt to avoid paying tax obligations in full. As a result of tax fraud, the honest taxpayers are left with the burden of covering the shortages. In addition, tax fraud may lead to potential tax increases, under-funded public infrastructures, and cutbacks for certain governmental programs.

The main aim of this project is to detect potential tax fraud using the algorithmic system. To extract the property records that show a high likelihood of committing fraud, 2 types of methodology were taken and 2 fraud scores are calculated. One is about the Heuristic Function of the z scores and the other is using Autoencoder. We compared the outputs of these two approaches and surprisingly found most of the records are overlapped. Final records were manually selected based on preliminary research on those properties.

Top 10 suspicious property records that need further investigation are:

#632816	**#917942**	**#565392**	**#776306**	**#85886**
#1067360	**#230596**	**#67129**	**#750816**	**#585439**

Description of Data

Dataset Name: Property Valuation and Assessment Data

Description: This data set retrieved from the NYC OpenData website represents NYC property valuation and assessments for the purpose to calculate Property Tax, Grant eligible properties Exemptions, and/or Abatements. Data collected and entered into the system by various City employees, like Property Assessors, Property Exemption specialists, ACRIS reporting, Department of Building reporting.

Source: Department of finance, City Government

Update Frequency: annually

Time Period: 2010/11

Number of Fields: 32

Number of Records: 1,070,994

An Overview of All Variables

Index	Column	Data_Type	Data_Number	Percent Populated	Unique_values	Zero_Numbers
1	RECORD	int64	1070994	100	1070994	0
2	BBLE	object	1070994	100	1070994	0
3	B	int64	1070994	100	5	0
4	BLOCK	int64	1070994	100	13984	0
5	LOT	int64	1070994	100	6366	0
6	EASEMENT	object	4636	0.432868905	13	0
7	OWNER	object	1039251	97.03611785	863349	0
8	BLDGCL	object	1070994	100	200	0
9	TAXCLASS	object	1070994	100	11	0
10	LTFRONT	int64	1070994	100	1297	169108
11	LTDEPTH	int64	1070994	100	1370	170128
12	EXT	object	354305	33.08188468	4	0
13	STORIES	float64	1014730	94.74656254	112	0
14	FULLVAL	float64	1070994	100	109324	13007
15	AVLAND	float64	1070994	100	70921	13009
16	AVTOT	float64	1070994	100	112914	13007
17	EXLAND	float64	1070994	100	33419	491699
18	EXTOT	float64	1070994	100	64255	432572
19	EXCD1	float64	638488	59.61639374	130	0
20	STADDR	object	1070318	99.93688107	839281	0
21	ZIP	float64	1041104	97.20913469	197	0
22	EXMPTCL	object	15579	1.454629998	15	0
23	BLDFRONT	int64	1070994	100	612	228815
24	BLDDEPTH	int64	1070994	100	621	228853
25	AVLAND2	float64	282726	26.39846722	58592	0
26	AVTOT2	float64	282732	26.39902745	111361	0
27	EXLAND2	float64	87449	8.165218479	22196	0
28	EXTOT2	float64	130828	12.21556797	48349	0
29	EXCD2	float64	92948	8.678666734	61	0
30	PERIOD	object	1070994	100	1	0
31	YEAR	object	1070994	100	1	0
32	VALTYPE	object	1070994	100	1	0

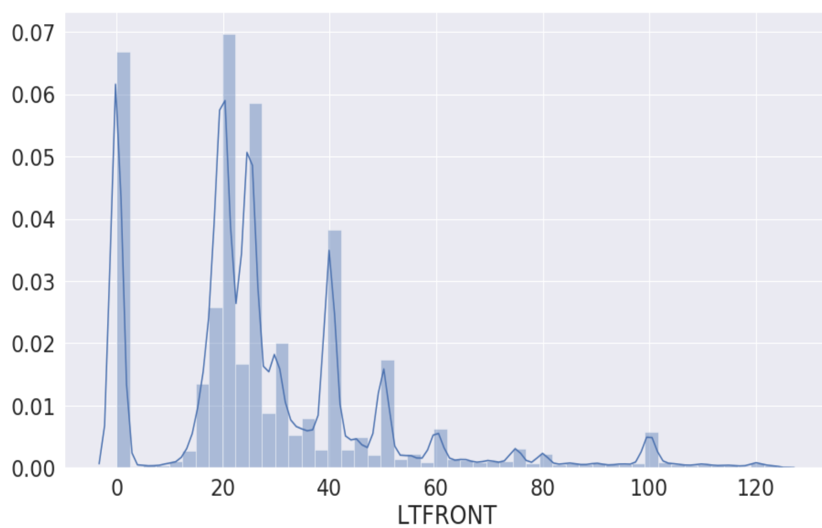
Basic Statistics for Numeric Fields

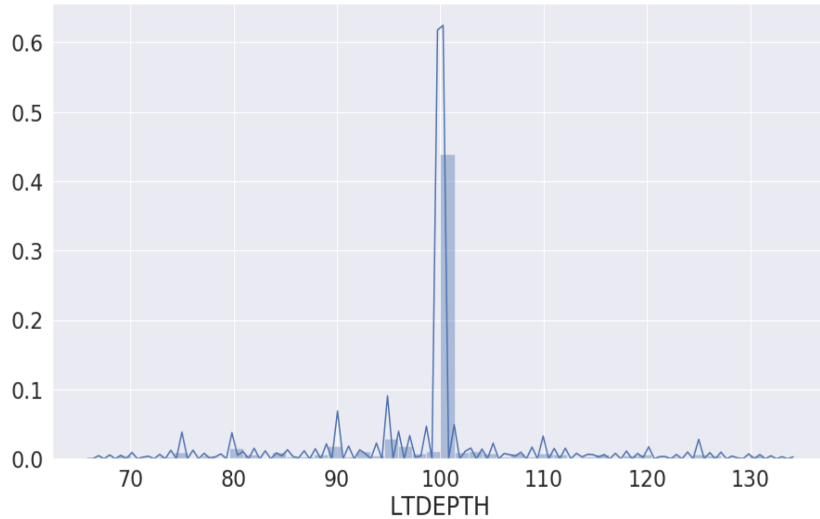
Column	count	mean	std	min	25%	50%	75%	max
RECORD	1070994	535497.5	309169.5	1	267749.3	535497.5	803245.8	1070994
LTFRONT	1070994	36.6353	74.03284	0	19	25	40	9999
LTDEPTH	1070994	88.86159	76.39628	0	80	100	100	9999
STORIES	1014730	5.006918	8.365707	1	2	2	3	119
FULLVAL	1070994	874264.5	11582431	0	304000	447000	619000	6150000000
AVLAND	1070994	85067.92	4057260	0	9180	13678	19740	2668500000
AVTOT	1070994	227238.2	6877529	0	18374	25340	45438	4668308947
EXLAND	1070994	36423.89	3981576	0	0	1620	1620	2668500000
EXTOT	1070994	91186.98	6508403	0	0	1620	2090	4668308947
EXCD1	638488	1602.014	1384.227	1010	1017	1017	1017	7170
BLDFRONT	1070994	23.04277	35.5797	0	15	20	24	7575
BLDDEPTH	1070994	39.92284	42.70715	0	26	39	50	9393
AVLAND2	282726	246235.7	6178963	3	5705	20145	62640	2371005000
AVTOT2	282732	713911.4	11652529	3	33912	79962.5	240551	4501180002
EXLAND2	87449	351235.7	10802213	1	2090	3048	31779	2371005000
EXTOT2	130828	656768.3	16072510	7	2870	37062	106840.8	4501180002
EXCD2	92948	1364.042	1094.706	1011	1017	1017	1017	7160

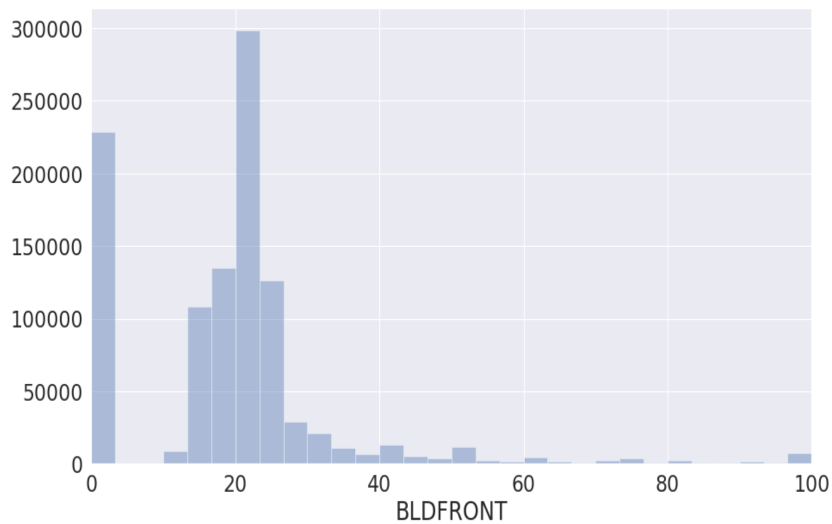
Basic Statistics for Categorical Fields

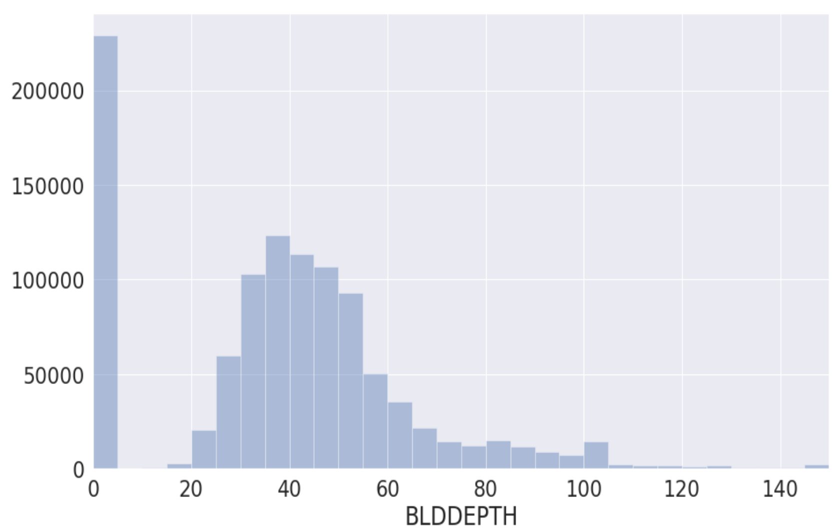
column	count	unique	top	freq
BBLE	1070994	1070994	1014191781	1
B	1070994	5	4	358046
BLOCK	1070994	13984	3944	3888
LOT	1070994	6366	1	24367
EASEMENT	4636	12	E	4148
OWNER	1039251	863348	PARKCHESTER PRESERVAT	6021
BLDGCL	1070994	200	R4	139879
TAXCLASS	1070994	11	1	660721
EXT	354305	3	G	266970
STADDR	1070318	839280	501 SURF AVENUE	902
ZIP	1.04E+06	196	10314	24606
EXMPTCL	15579	14	X1	6912
PERIOD	1070994	1	FINAL	1070994
YEAR	1070994	1	2010/11	1070994
VALTYPE	1070994	1	AC-TR	1070994

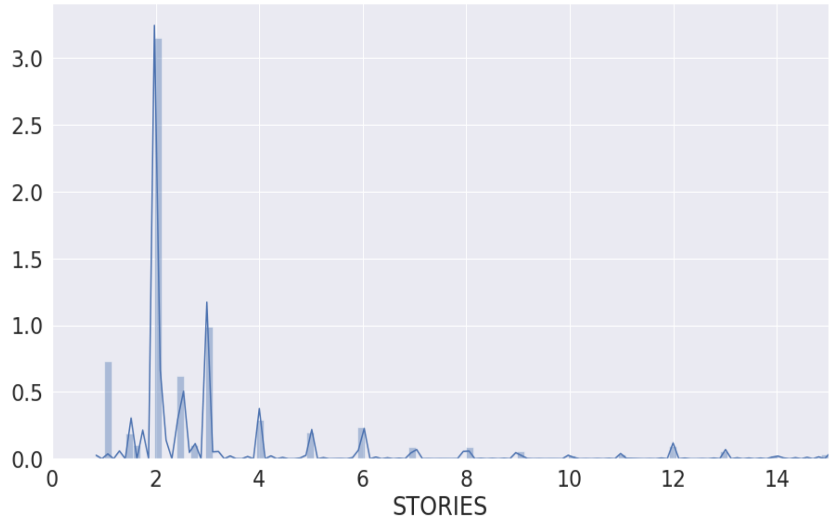
Distributions for A Couple of Important Variables

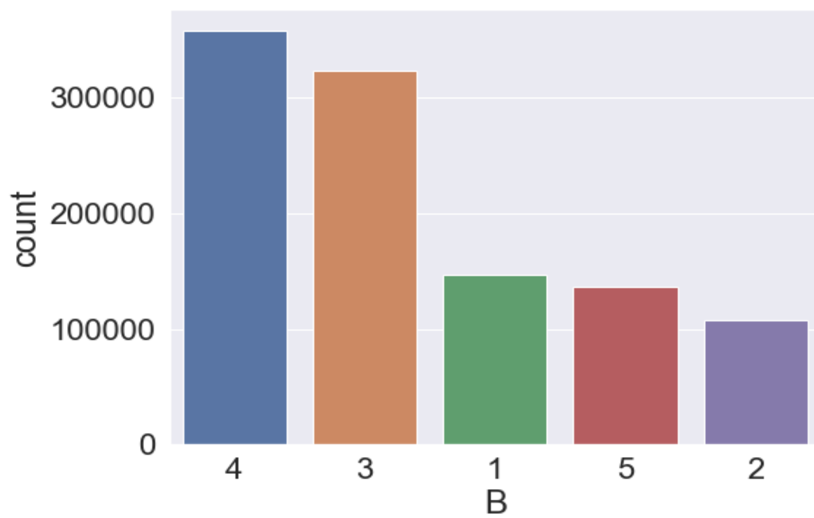


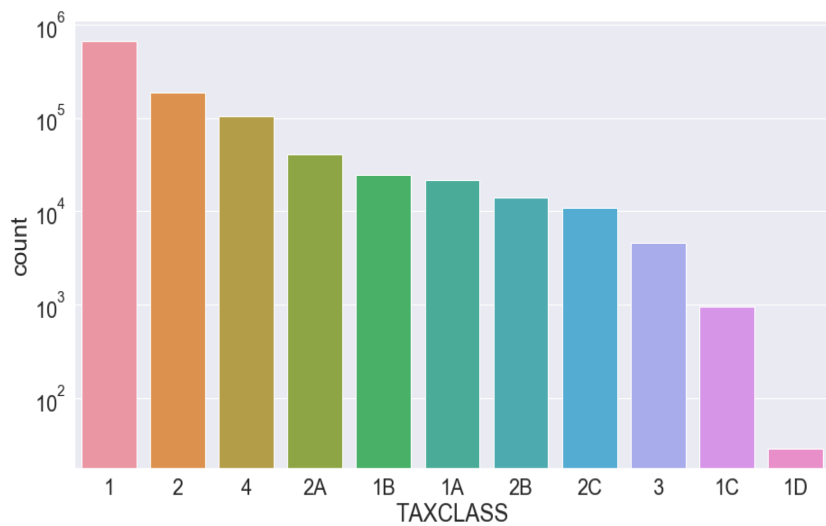

0.6
0.5
0.4
0.3
0.2
0.1
0.0
70
80
90
100
110
120
130
LTDEPTH


300000
250000
200000
150000
100000
50000
0
0
20
40
60
80
100
BLDFRONT


200000
150000
100000
50000
0
0
20
40
60
80
100
120
140
BLDDEPTH







A full Data Quality Report is attached in the appendix.

Data Cleaning

For our New York property data we will need to fill in missing values for these fields: FULLVAL, AVLAND, AVTOT, ZIP, STORIES, LTFRONT, LTDEPTH, BLDFRONT, BLDDEPTH.

LTFRONT, LTDEPTH, BLDFRONT, BLDDEPTH

For LTFRONT and LTDEPTH, we replace the zero value by random integers between 30 to 100. And for BLDFRONT and BLDDEPTH, the zero value is replaced by random integers between 20 to 40.

STORIES

There are both 0 and NA values in this field, we first transform all them to NA and fill the NA value by the average store number with the same ZIP, B, and TAXCLASS. Then for the left NA value, we first fill them by the average store number with the same ZIP, and TAXCLASS and with the same TAXCLASS until there is no NA value exists.

FULLVAL, AVLAND, AVTOT

Similar to the data process of STORIES, we fill both the 0 and NA values in these three fields by three steps until all rows have greater than 0 values

ZIP

For the NA value in ZIP field, we fill them by the same number as the record's above them or the record's below them since the dataset is sorted by the geographic location.

Variable Creation

To commence, we need to analyze the abnormal extend of the three values: FULLVAL, AVLAND, AVTOT. Since the dataset is limited, we create 3 new sizes to build more expert variables, the formulas of the sizes is:

S1 = lotarea = LTFRONT * LTDEPTH

S2 = bldarea = BLDFRONT * BLDDEPTH

S3 = bldvol = bldarea * STORIES

To better compare the values, we put each value into these three sizes so that we have more measure scales to dig out the features of them. We normalize these values by the three sizes so that we have 9 variables after this step. The formula is:

$$r_1 = \frac{V_1}{S_1} \qquad r_4 = \frac{V_2}{S_1} \qquad r_7 = \frac{V_3}{S_1}$$

$$r_2 = \frac{V_1}{S_2} \qquad r_5 = \frac{V_2}{S_2} \qquad r_8 = \frac{V_3}{S_2}$$

$$r_3 = \frac{V_1}{S_3} \qquad r_6 = \frac{V_2}{S_3} \qquad r_9 = \frac{V_3}{S_3}$$

Then we create the grouped averages of these 9 variables, grouped by the following five features: ZIP3, ZIP5, TAXCLASS, BOROUGH, ALL. Because we intend to compare the values not only in different sizes, but also to the general groups' features to have a more accurate judgement of whether they have high fraud potential. Finally, we divide each of the 9 core variables by the 5 scale factors from these groupings, which makes 9 * 5 = 45 variables. Based on the statistical characteristics of these 45 variables, we can have more information to decide which record looks more abnormal. The formula is:

$$\frac{r_1}{< r_1 >_g}, \quad \frac{r_2}{< r_2 >_g}, \quad \frac{r_3}{< r_3 >_g}, \quad \dots \quad \frac{r_9}{< r_9 >_g} \qquad g = 1, \dots, 5$$

Below is a list of all variables and their basic statistic features.

	mean	std	min	max
zip5_fullval_lotarea	1	6.025845104	2.20445E-06	2678.764802
zip5_fullval_bldarea	1	11.04330925	8.14812E-07	5162.957404
zip5_fullval_bldvol	1	10.32367319	2.06673E-07	5237.936115
zip5_avland_lotarea	1	12.75457822	1.56073E-06	7878.895693
zip5_avland_bldarea	1	24.19177735	1.81243E-07	8354.676868
zip5_avland_bldvol	1	24.17974865	5.04643E-08	9201.163738
zip5_avtot_lotarea	1	10.39805926	8.05813E-07	5673.612655
zip5_avtot_bldarea	1	20.31679325	8.91081E-07	7418.997511
zip5_avtot_bldvol	1	20.20182194	2.52977E-07	8469.408542
zip3_fullval_lotarea	1	8.565933452	2.33163E-06	5270.696279
zip3_fullval_bldarea	1	21.61432737	2.71953E-06	12842.92874
zip3_fullval_bldvol	1	15.58397718	6.12408E-07	7957.05535
zip3_avland_lotarea	1	27.40683056	7.83537E-07	20110.59348
zip3_avland_bldarea	1	107.1872068	1.39896E-06	78276.38939
zip3_avland_bldvol	1	81.58023758	2.57843E-07	43581.75314
zip3_avtot_lotarea	1	23.63065678	3.67276E-07	20870.74122
zip3_avtot_bldarea	1	89.161855	3.25218E-06	66442.10269
zip3_avtot_bldvol	1	74.45474301	8.81873E-07	59687.91197
taxclass_fullval_lotarea	1	7.72978975	1.5455E-06	3836.944456
taxclass_fullval_bldarea	1	13.02038612	4.14788E-06	6377.370936
taxclass_fullval_bldvol	1	44.22217224	2.04844E-06	44648.06263
taxclass_avland_lotarea	1	9.387585576	2.50434E-06	4978.731087
taxclass_avland_bldarea	1	27.36083797	1.85714E-06	21992.02976
taxclass_avland_bldvol	1	100.4668455	6.03938E-07	103159.5148
taxclass_avtot_lotarea	1	8.978447056	1.54054E-06	5152.852132
taxclass_avtot_bldarea	1	16.23143134	8.41405E-06	9775.939613
taxclass_avtot_bldvol	1	45.96829669	2.04843E-06	44647.84819
boro_fullval_lotarea	1	8.478506264	2.64645E-06	5270.696279
boro_fullval_bldarea	1	22.20916439	2.78928E-06	12844.58089
boro_fullval_bldvol	1	16.04112179	6.33658E-07	8301.035517
boro_avland_lotarea	1	26.32268224	7.83537E-07	20112.41021
boro_avland_bldarea	1	110.1606012	1.46282E-06	78299.66015
boro_avland_bldvol	1	81.02606678	2.72212E-07	43033.84792
boro_avtot_lotarea	1	20.34621341	3.67276E-07	16880.43145
boro_avtot_bldarea	1	99.02504923	3.34161E-06	79100.11891
boro_avtot_bldvol	1	79.63735833	9.20839E-07	67780.35006
all_fullval_lotarea	1	6.69500266	2.05574E-06	3819.108497
all_fullval_bldarea	1	24.43392323	5.10737E-06	16314.78753
all_fullval_bldvol	1	15.56965659	4.4207E-07	9635.422857
all_avland_lotarea	1	18.55451777	7.04026E-07	13016.37121
all_avland_bldarea	1	105.6089365	3.8746E-06	69547.24501
all_avland_bldvol	1	86.25281074	3.97874E-07	50537.58246
all_avtot_lotarea	1	13.23863152	2.80341E-07	8356.666908
all_avtot_bldarea	1	78.75206471	1.12759E-05	60758.30412
all_avtot_bldvol	1	89.14406473	1.4231E-06	79237.417

Scaling and Dimensionality Reduction

The scales of the input values are different and thus may exert influences on the model. In order to put all variables on the same footing, z scaling is applied to the data.

$$x' = \frac{x - \mu}{\sigma} \text{ (z scaling)}$$

Furthermore, with the goal of avoiding the curse of dimensionality and eliminating the multicollinearity among variables, dimensionality reduction is conducted through principal component analysis. To be more specific, after finding out the dominant directions in the data and rotating the coordinate system along these directions, the data was rewritten in terms of this rotated coordinate system with 45 new dimensions. Considering that all principal components (PCs) are sorted by magnitude, only the first 8 PCs are kept, accounting for 95% of the total variance. In this way, the data can be displayed in an 8-dimension space, and z scaling is applied again on it since there is no significant difference among the importance of these 8 variables.

Algorithms

Method 1: Heuristic Function of the z scores

Based on 45 variables we created, we first standard scale them to center the data and get all the scales the same. Then, in order to reduce the dimension and concentrate on the main information, we use the PCA method with 99% representative to only keep 8 PCs and z scale these 8 PCs again. For the simple distance measure of each record's 8-d space to the origin, we use Euclidean distance (n = 2) to calculate the z score, which is also the fraud score of each record.

$$s_i = \left(\sum_k |z_k^i|^n\right)^{1/n}, \quad n \text{ anything}$$

This method is a reasonable fraud score because z-scaling itself has been a good approach to detect anomaly, and we also did PCA based upon it to get a 95% representative range. Even more, we z-scaling again PCA scores, which made it a much more effective way of detecting anomaly.

Method 2: Autoencoder Error

Considering that method 1 uses a heuristic function calculating the distance to origin as the fraud score and it is hard to tell whether it is the best assessment method, another method using the autoencoder error is introduced. The autoencoder can learn the nature of the bulk of the data and reproduce the data records as well as possible, therefore, the records that are not reproduced well are unusual records. This algorithm is reasonable because it is consistent with the mechanism of detecting fraudulent records, which is finding out the outliers in results with inputs at relatively normal levels.

$$s_i = \left(\sum_k |z_k'^i - z_k^i|^n\right)^{1/n}, \quad n \text{ anything}$$

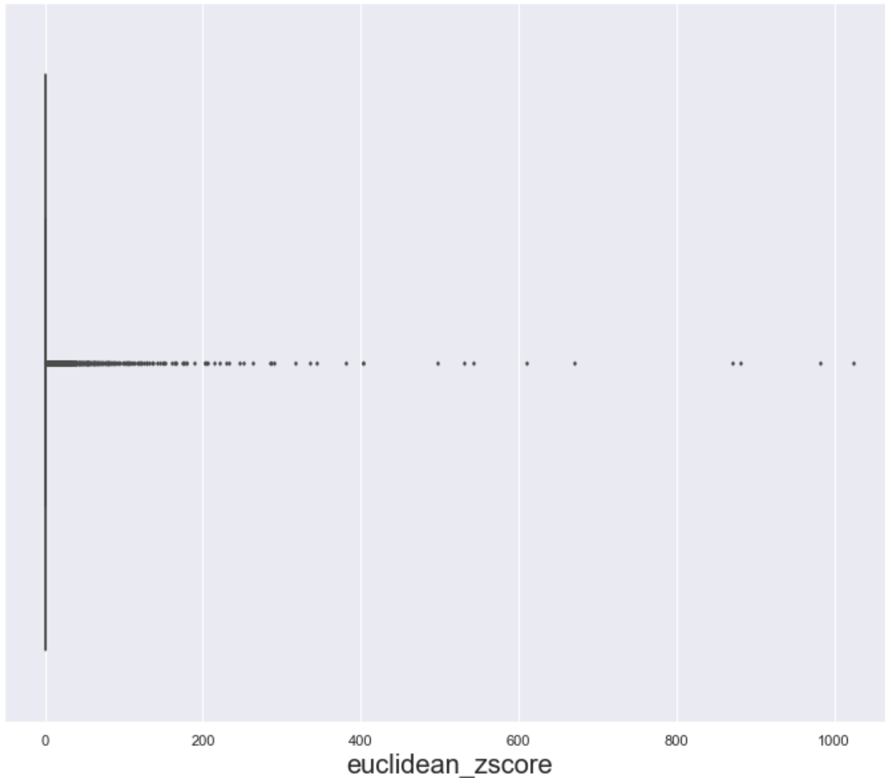
The autoencoder is implemented with an 8-layer neural network, 4 encoders and 4 decoders, built on the scaled data. For the encoder, the first, second, third and fourth layers consist of 64, 32, 16 and 8 nodes, respectively; the decoder has an inverse number of nodes. Last, the fraud scores of method 2 is the Euclidean distance between the output of the model and the fraud scores generated in Method 1.

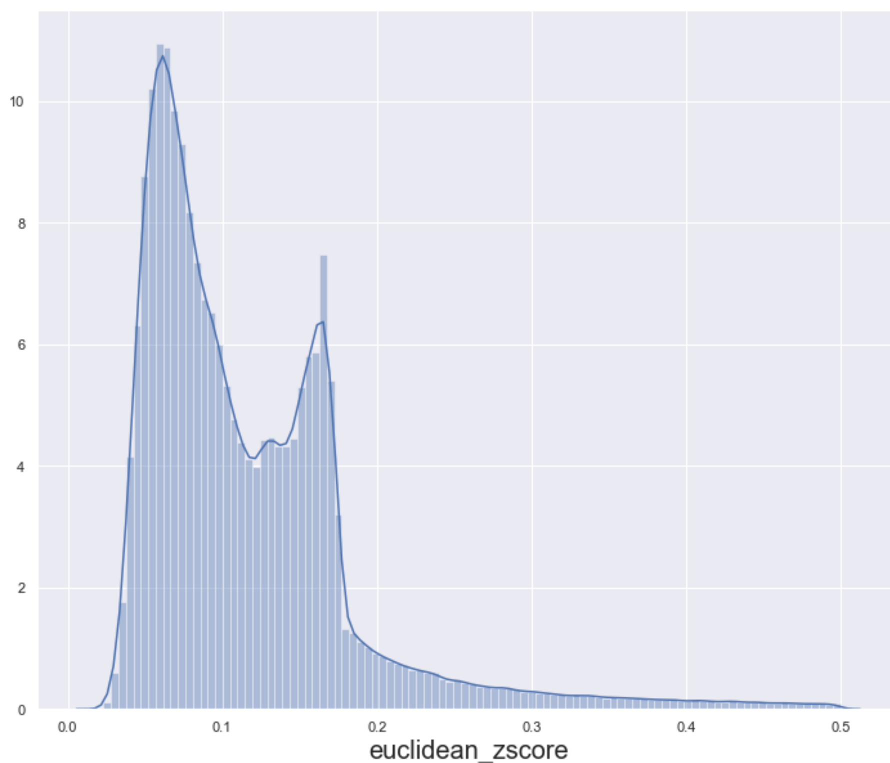
Final Score Calculation

After the comparison, we choose the Euclidean distance to calculate the fraud score in both Heuristic function and Autoencoder ways. The final fraud score is the average of the two scores.

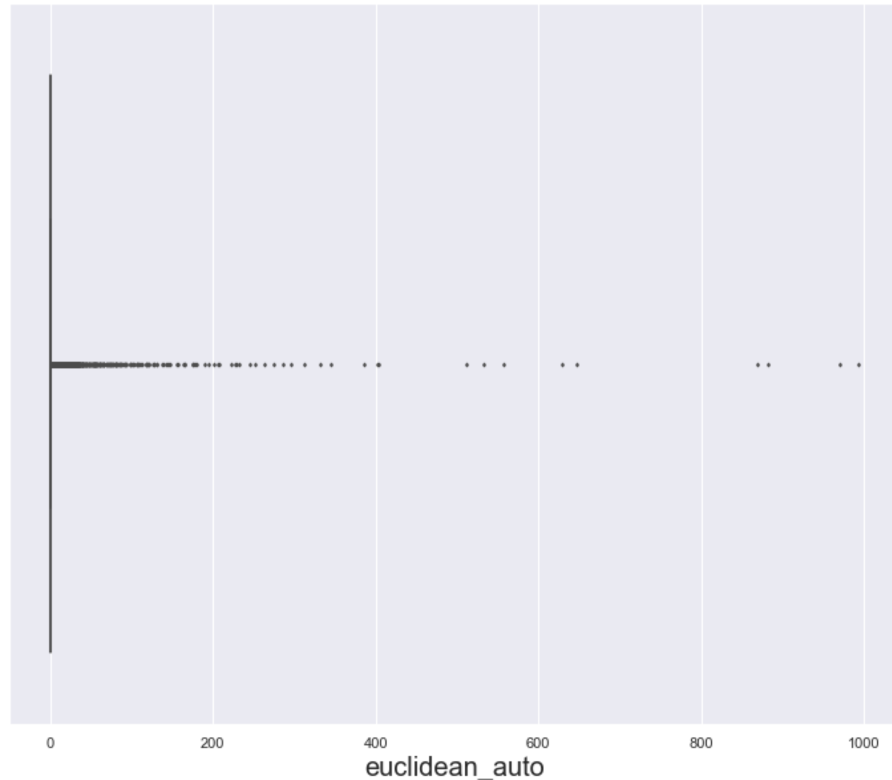
Results

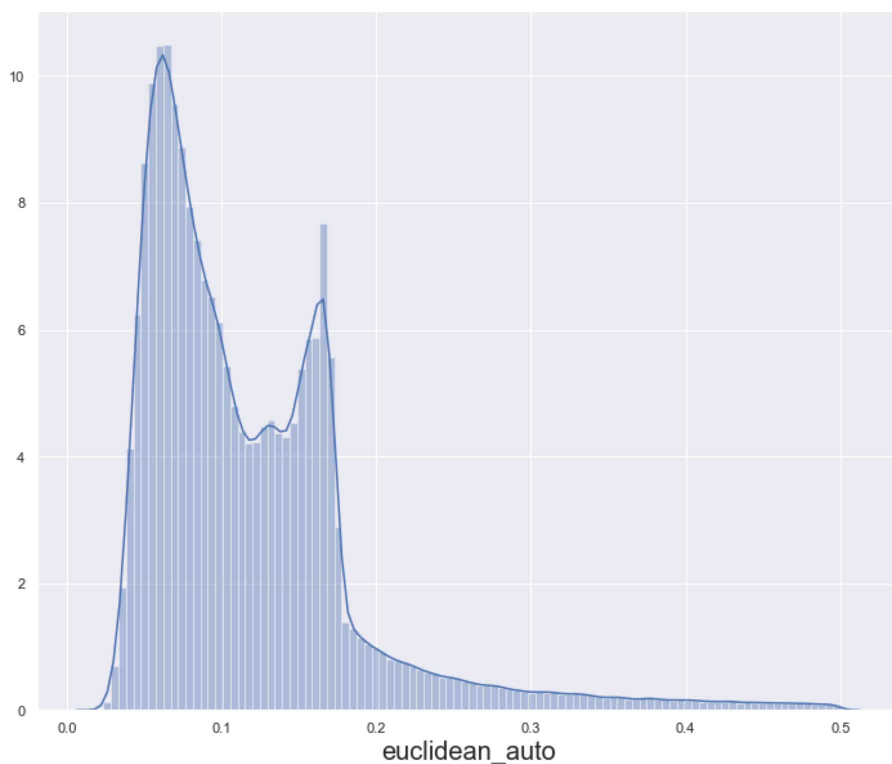
Fraud Score 1 Distribution



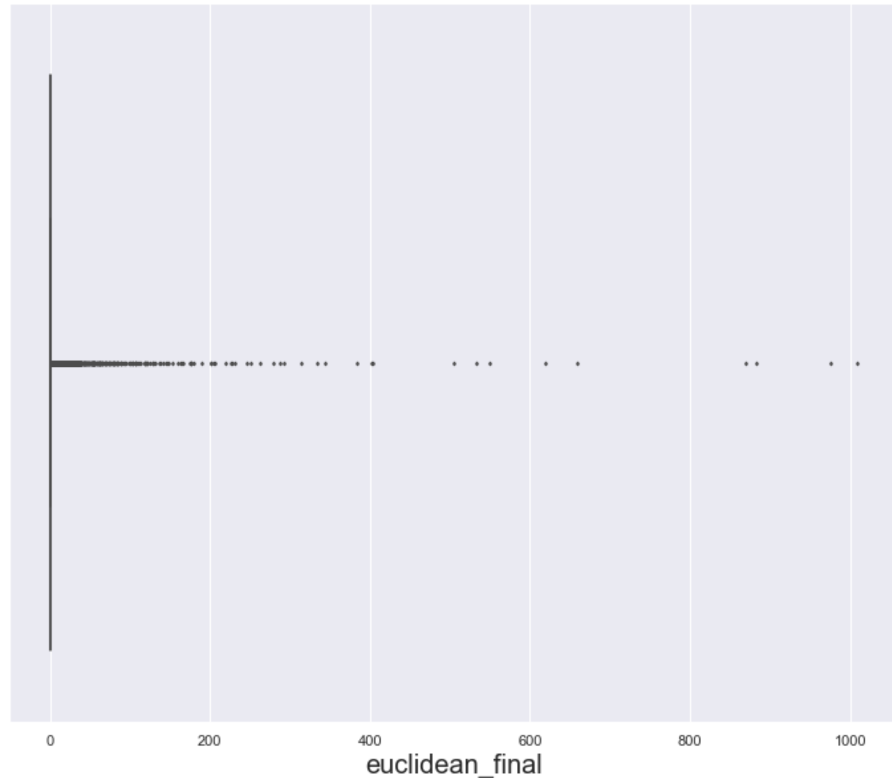


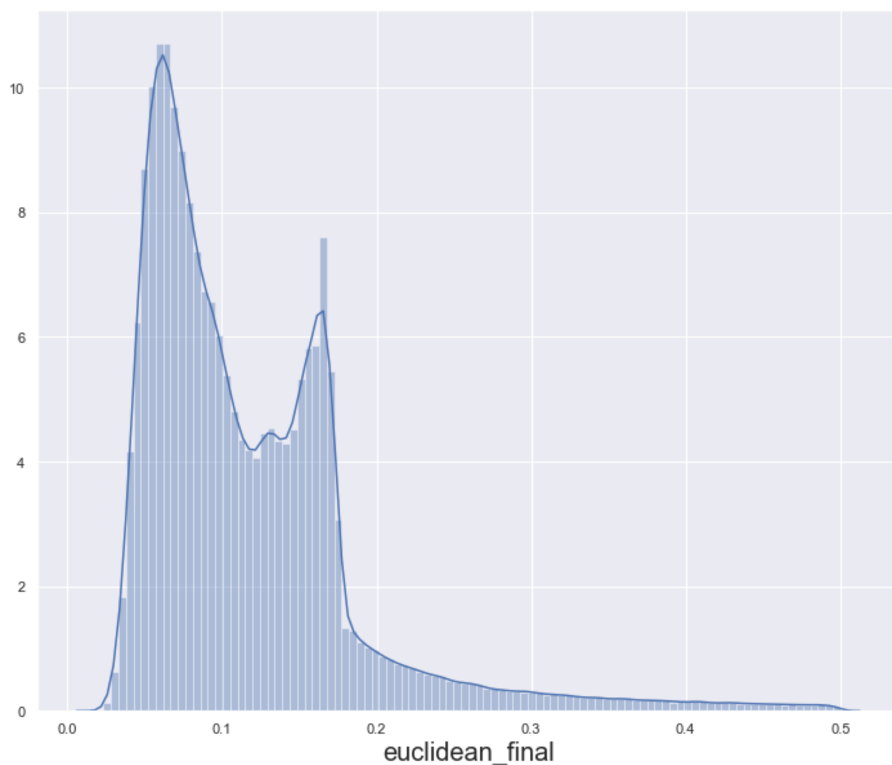
Fraud Score 2 Distribution





Final Fraud Score





Top 10 Anomalous Records

#632816. Multiple features related to the size of the property are equal to 1, which is unrealistic. The owner of this property has been changing frequently over the years, and one of the owners is the CNY/NYCTA. Further investigation is needed to make things clear.

#917942. Land value and total value are both exempted in full. We checked the address and found it belongs to a chain hotel, indicating that it is quite suspicious to have 100% tax exempted.

#565392. This property is a park, and thus its land value and total value are both exempted in full.

#776306. Multiple features related to the size of the property are equal to 1, which is unrealistic. Further investigation is needed to make things clear.

#85886. This property is a park, and thus its land value and total value are both exempted in full.

#1067360. Multiple features related to the size of the property are equal to 1, which is unrealistic. Further investigation is needed to make things clear.

#230596. For one thing, multiple features related to the size of the property are equal to 1, which is unrealistic. For another thing, all the value features of this property are exactly the same with #750816 and detailed information is missing, so the data might be manipulated. Further investigation is needed to make things clear.

#67129. This property is a national museum, and thus its land value and total value are both exempted in full.

#750816. Same as #230596.

#585439. This property is a proprietary hotel, but 85% of the property's total value was exempted. Further investigation is needed to make things clear.

Conclusions

In this tax fraud detection project, we created 45 variables and used PCA to conduct dimensionality reduction. Then, two types of algorithms were taken to unveil the anomalous records. We compare the results from two approaches and select the final records based on our preliminary investigation.

We will further polish our algorithms by trying to find more layers or neurons to the same layer to test if there is any improvement in loss value. Also, we can leverage different optimizers to further improve the model performance. Our current approach dealing with missing values and untidy data could be upgraded by using advanced and complex logics with more time. Also, we'd like to hear back from experts in property tax and then reduce the number of unusual but legitimate records being selected currently. We believe all of these can be effective techniques in optimizing our detection competence and can be accomplished in the future works.

Appendix: Data Quality Report for NY Property Data

Description of Dataset

Dataset name: NY property data

Dataset description: this is real estate assessment property data in NYC, and the Department of Finance values properties every year as one step in calculating property tax bills.

Agency name: Department of Finance

Source: Department of Finance

Date: 2010/11

Update frequency: annually

Number of fields: 32

Number of records: 1070994

Overview of Data

Below is an overview of all fields in the dataset:

	Column Name	Data Type	# Rows	Populated (%)	# Zero	# Unique Values
1	RECORD	int64	1070994	100	0	1070994
2	BBLE	object	1070994	100	0	1070994
3	B	int64	1070994	100	0	5
4	BLOCK	int64	1070994	100	0	13984
5	LOT	int64	1070994	100	0	6366
6	EASEMENT	Object	4636	0.4329	0	13
7	OWNER	Object	1039251	97.0361	0	863347
8	BLDGCL	Object	1070994	100	0	200
9	TAXCLASS	Object	1070994	100	0	11
10	LTFRONT	int64	1070994	100	169108	1297
11	LTDEPTH	int64	1070994	100	170128	1370
12	EXT	object	354305	33.0819	0	4

13	STORIES	float64	1014730	94.7466	0	112
14	FULLVAL	float64	1070994	100	13007	109324
15	AVLAND	float64	1070994	100	13009	70921
16	AVTOT	float64	1070994	100	13007	112914
17	EXLAND	float64	1070994	100	491699	33419
18	EXTOT	float64	1070994	100	432572	64255
19	EXCD1	float64	638488	59.6164	0	130
20	STADDR	Object	1070318	99.9369	0	839281
21	ZIP	float64	1041104	97.2091	0	197
22	EXMPTCL	object	15579	1.4546	0	15
23	BLDFRONT	int64	1070994	100	228815	612
24	BLDDEPTH	int64	1070994	100	228815	621
25	AVLAND2	float64	282726	26.3985	0	58592
26	AVTOT2	float64	282732	26.3990	0	111361
27	EXLAND2	float64	87449	8.1652	0	22196
28	EXTOT2	float64	130828	12.2156	0	48349
29	EXCD2	float64	92948	8.6787	0	61
30	PERIOD	object	1070994	100	0	1
31	YEAR	object	1070994	100	0	1
32	VALTYPE	object	1070994	100	0	1

Below is the basic statistics for numeric fields:

	Column Name	Count	Mean	Std	Min	25%	75%	Max
1	LTFRONT	1070994	36.6353	7.403284e+01	0	19	40	9.999000e+03
2	LTDEPTH	1070994	88.8616	7.639628e+01	0	80	100	9.999000e+03
3	STORIES	1014730	5.0069	8.365707e+00	1	2	3	1.190000e+02
4	FULLVAL	1070994	874264.5054	1.158243e+07	0	304000	619000	6.150000e+09
5	AVLAND	1070994	85067.9187	4.057260e+06	0	9180	19740	2.668500e+09

6	AVTOT	1070994	227238.1687	6.877529e+06	0	18374	45438	4.668309e+09
7	EXLAND	1070994	36423.8907	3.981576e+06	0	0	1620	2.668500e+09
8	EXTOT	1070994	91186.9817	6.508403e+06	0	0	2090	4.668309e+09
9	BLDFRONT	1070994	23.0428	3.557970e+01	0	15	24	7.575000e+03
10	BLDDEPTH	1070994	39.9228	4.270715e+01	0	26	50	9.393000e+03
11	AVLAND2	282726	246235.7193	6.178963e+06	3	5705	62640	2.371005e+09
12	AVTOT2	282732	713911.4362	1.165253e+07	3	33912	240551	4.501180e+09
13	EXLAND2	87449	351235.6843	1.080221e+07	1	2090	31779	2.371005e+09
14	EXTOT2	130828	656768.2819	1.607251e+07	7	2870	106840	4.501180e+09

Below is the basic statistics for categorical fields:

	Column Name	Count	# Unique Values	Top Value	Frequency
1	RECORD	1070994	1070994		1
2	BBLE	1070994	1070994		1
3	B	1070994	5	4	358046
4	BLOCK	1070994	13984	3944	3888
5	LOT	1070994	6366	1	24367
6	EASEMENT	4636	13	E	4148
7	OWNER	1039251	863347	PARKCHESTER PRESERVAT	6020
8	BLDGCL	1070994	200	R4	139879
9	TAXCLASS	1070994	11	1	660721
10	EXT	354305	4	G	266970
11	EXCD1	638488	130	1017	425348
12	STADDR	1070318	839281	501 SURF AVENUE	902
13	ZIP	1041104	197	10314	24606
14	EXMPTCL	15579	15	X1	6912
15	EXCD2	92948	61	1017	65777
16	PERIOD	1070994	1	FINAL	1070994

17	YEAR	1070994	1	2010/11	1070994
18	VALTYPE	1070994	1	AC-TR	1070994

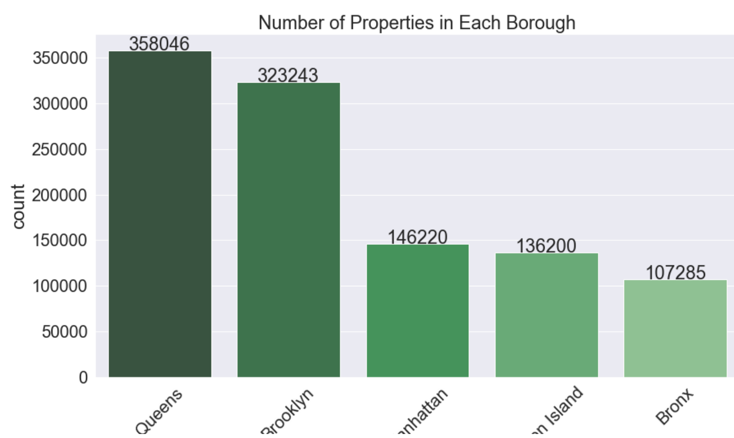
Summary of Each Column

In this section, details of each column will be given, including the description, distribution, the most popular values and their counts, and so forth.

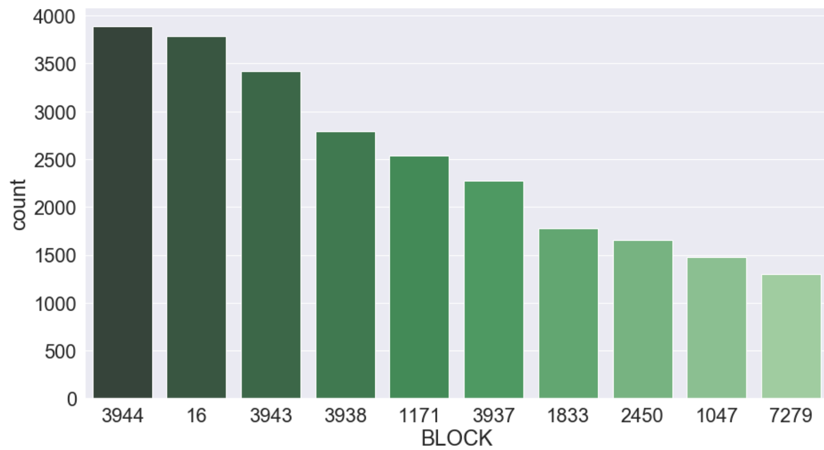
RECORD. A serial number for each record in the dataset, ranging from 1 to 1070994.

BBLE. A file key for each record, concatenating borough, block, lot, and easement. For example, the first record has an BBLE of 1000010101, and the last one has an BBLE of 5080500094.

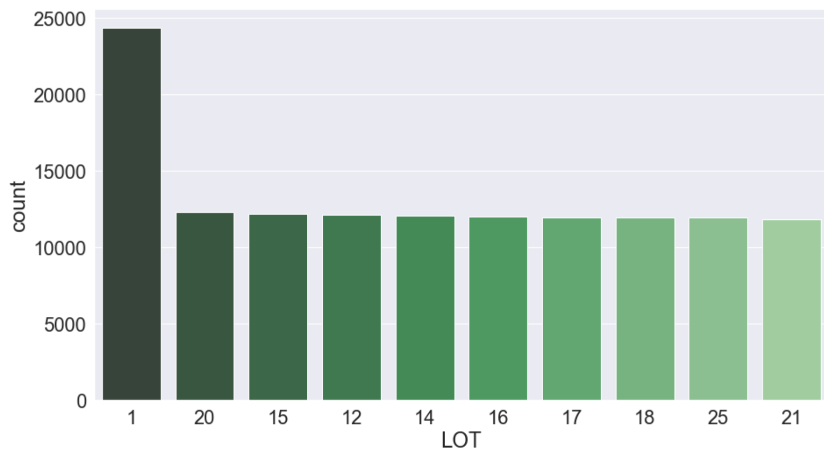
B. Borough, including Manhattan (1), Bronx (2), Brooklyn (3), Queens (4), and Staten Island (5).



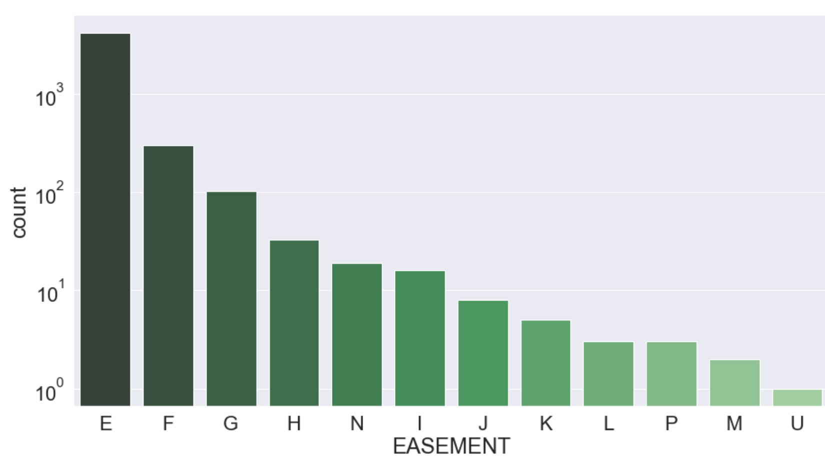
BLOCK. Valid block ranges by borough. Manhattan covers 1 to 2255, Bronx covers 2260 to 5958, Brooklyn covers 1 to 8955, Queens covers 1 to 16350, and Staten Island covers 1 to 8050.



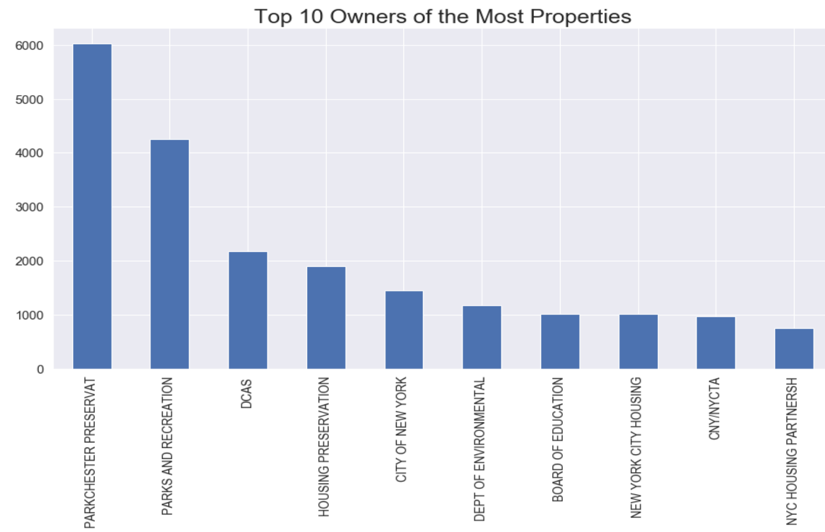
LOT. lot partitioned by block



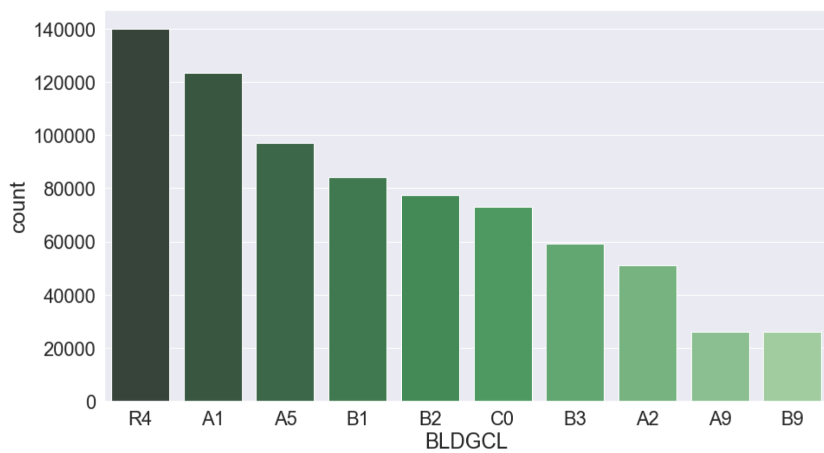
EASEMENT. type of easement. Space = No Easement, A = Air Easement, B = Non-Air Rights, E = Land Easement, F – M: duplicates of E, N = Non-Transit Easement, P = Pier, R = Railroad, S = Street, U = U. S. Government.



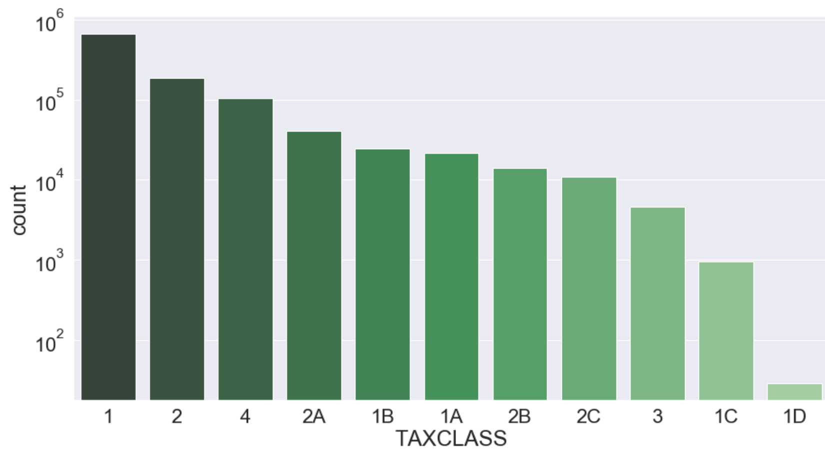
OWNER. name of the organization who owns the property



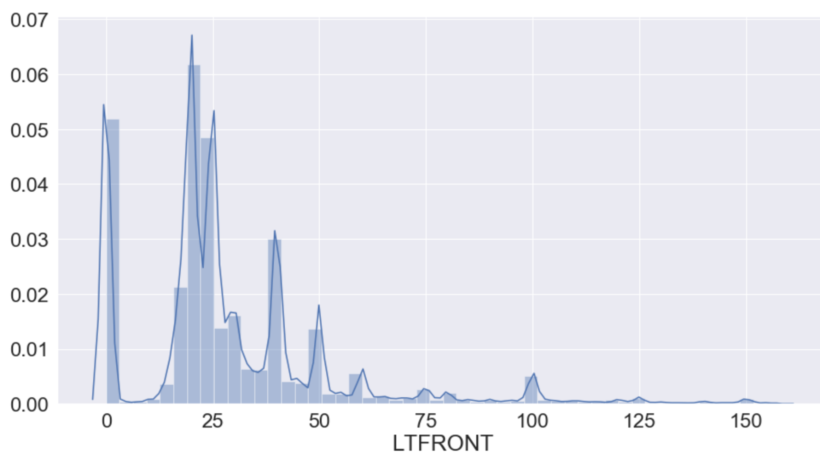
BLDGCL. building class



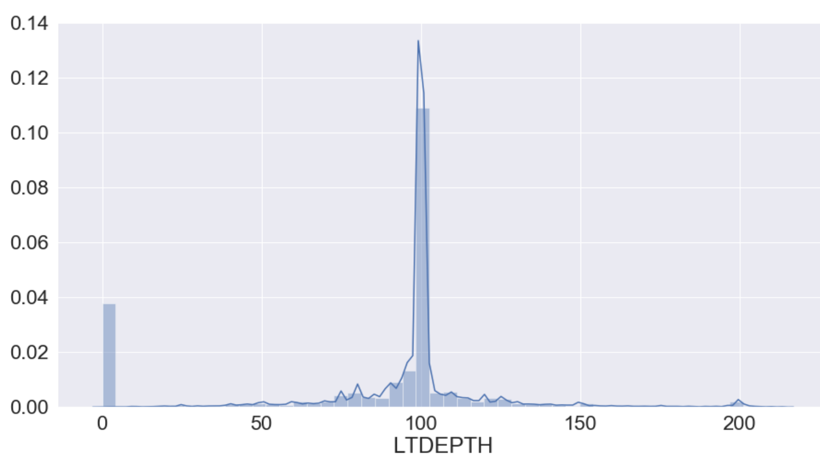
TAXCLASS. tax class. 1 = 1 - 3 unit residence, 2 = Apartments, 2A = 4 - 6 unit residence, 3 = utilities, 4 = all others.



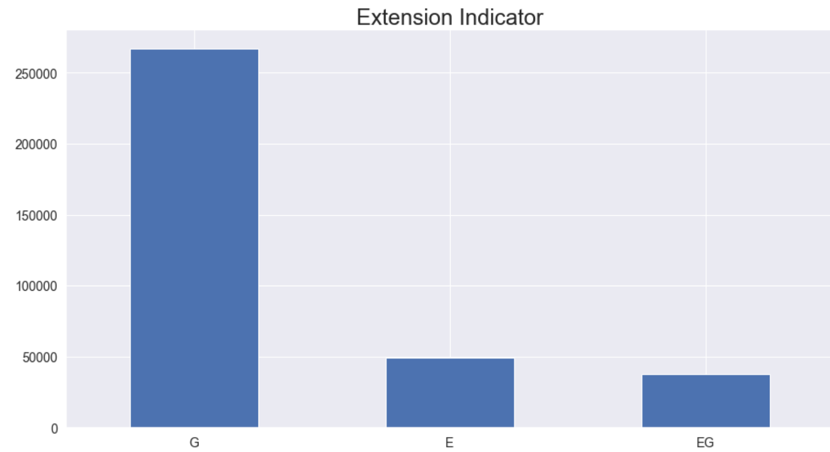
LTFRONT: lot width



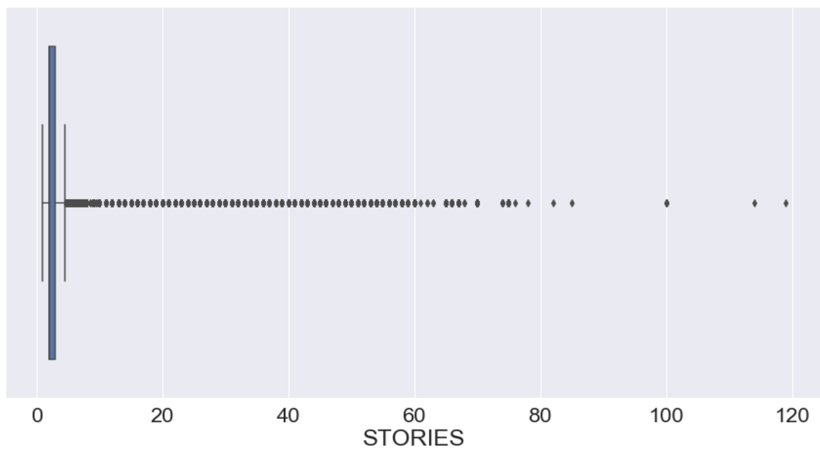
LTDEPTH: lot depth

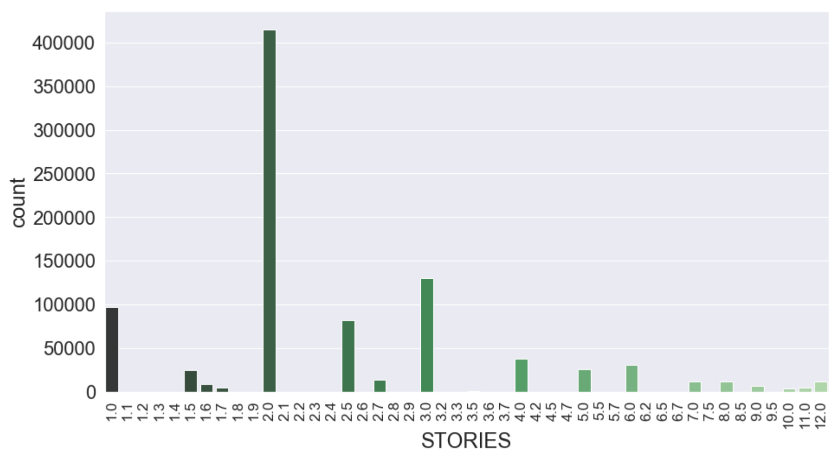


EXT: extension indicator

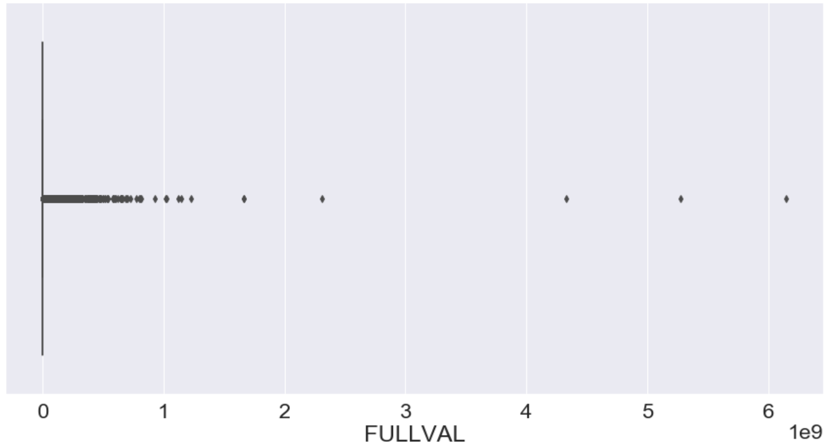


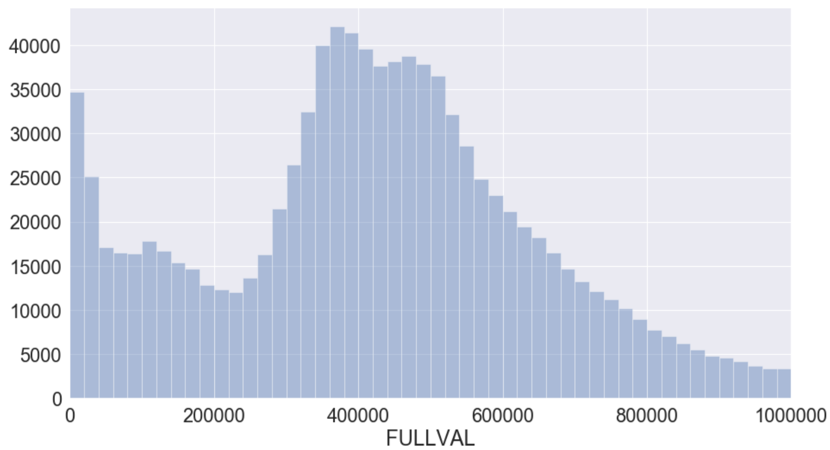
STORIES: number of stories of the building



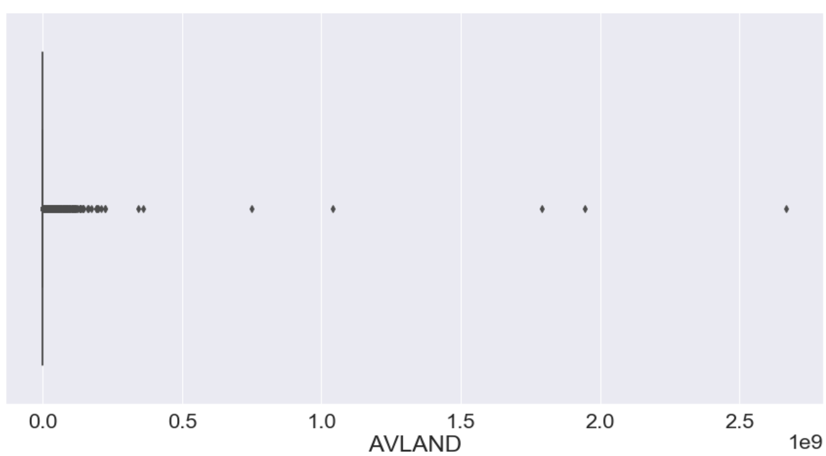


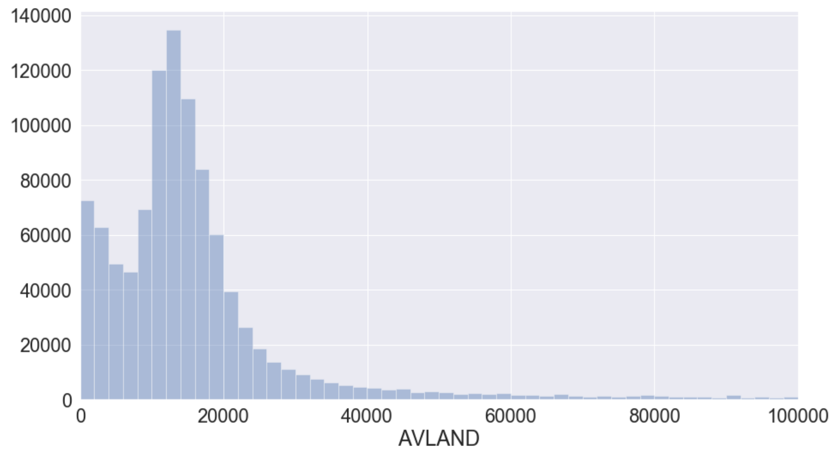
FULLVAL: market value



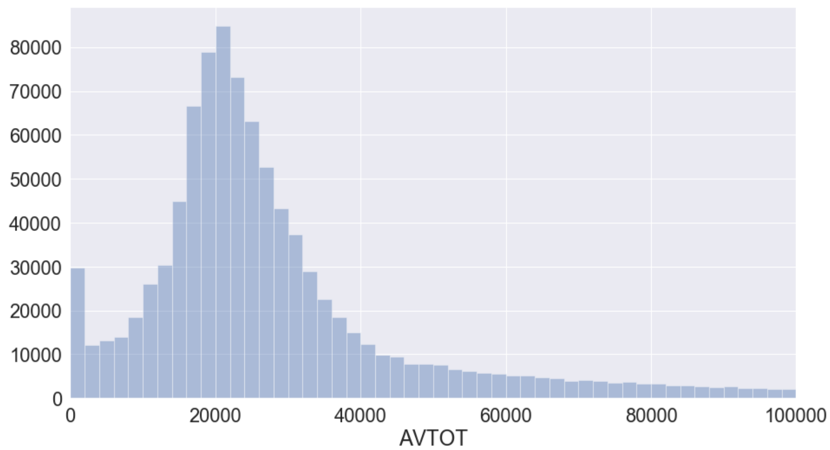


AVLAND: actual land value

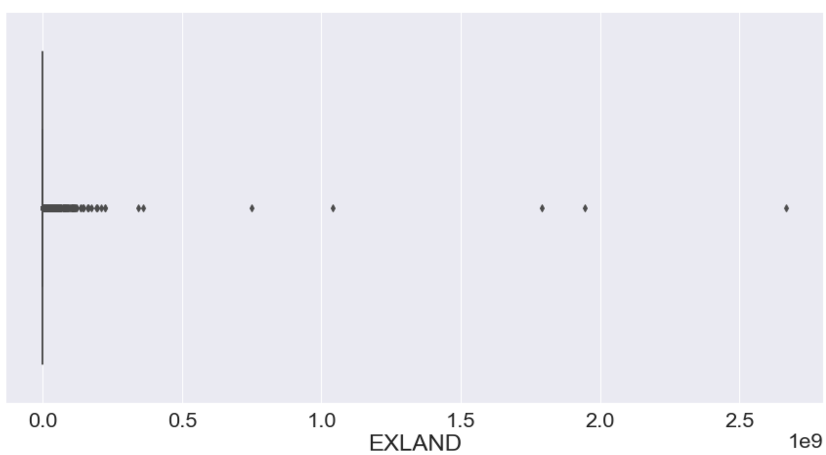




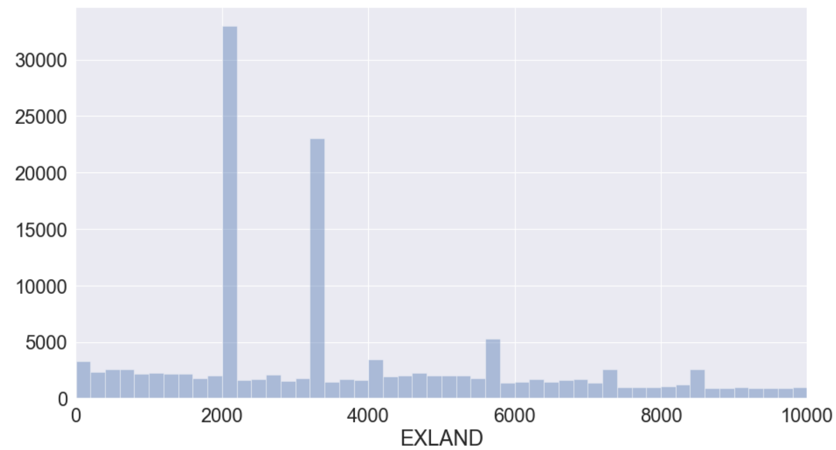
AVTOT: actual total value



EXLAND: actual exempt land value

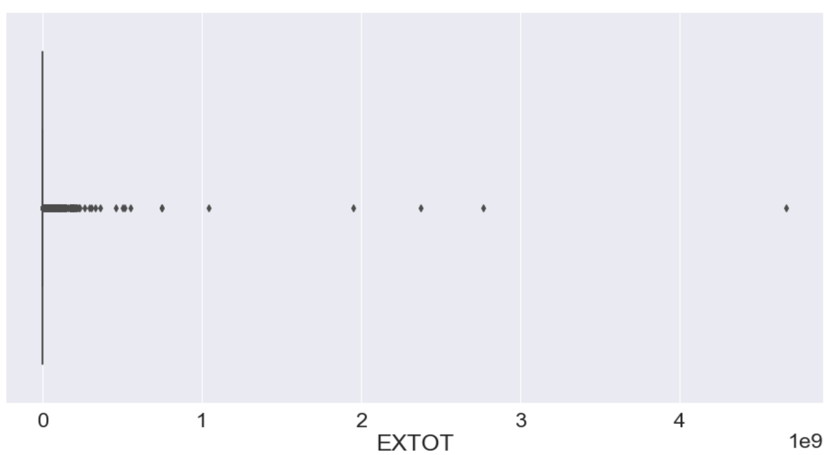


```
0.0        491699
1620.0     357182
2090.0      31112
3240.0      21519
5760.0       3560
8520.0       1613
4180.0       1323
7380.0       1279
10140.0       797
72.0          437
Name: EXLAND, dtype: int64
```

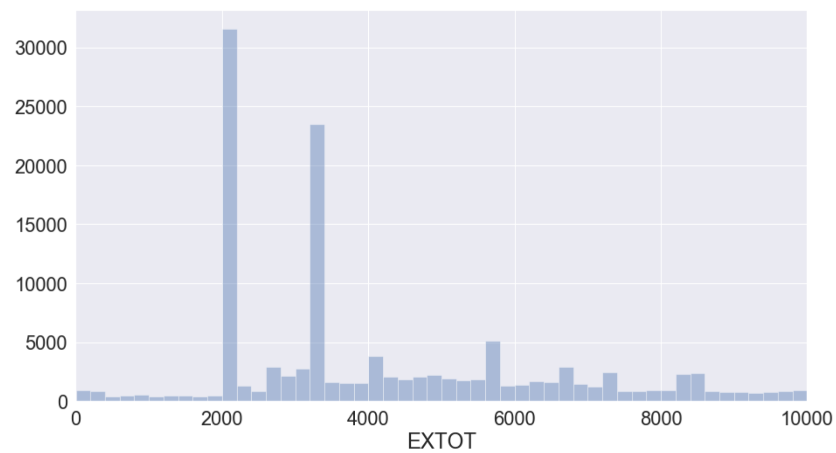


Since there are too many 0s and 1629s, they have been taken out when doing this visualization, otherwise the distribution of other values would be too inconspicuous.

EXTOT: actual exempt total value

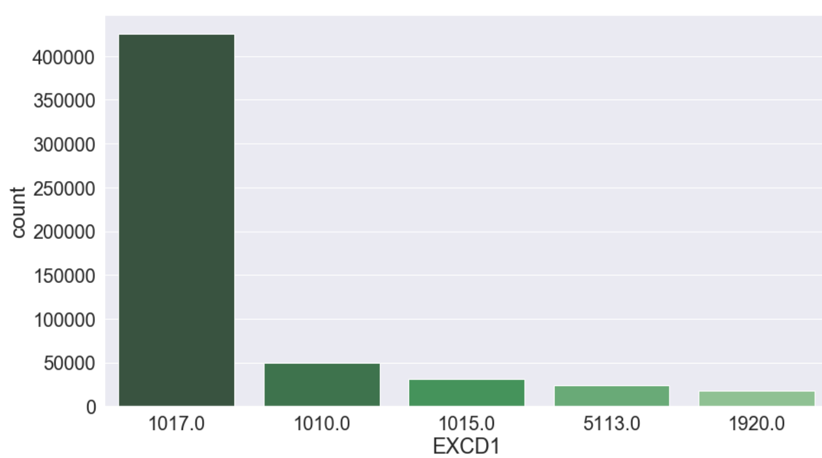


```
0.0       432572
1620.0    354880
2090.0     30069
3240.0     21803
5760.0      3549
4180.0      1624
8520.0      1614
6620.0      1475
7380.0      1277
8240.0      1060
Name: EXTOT, dtype: int64
```



Since there are too many 0s and 1629s, they have been taken out when doing this visualization, otherwise the distribution of other values would be too inconspicuous.

EXCD1: exemption code 1



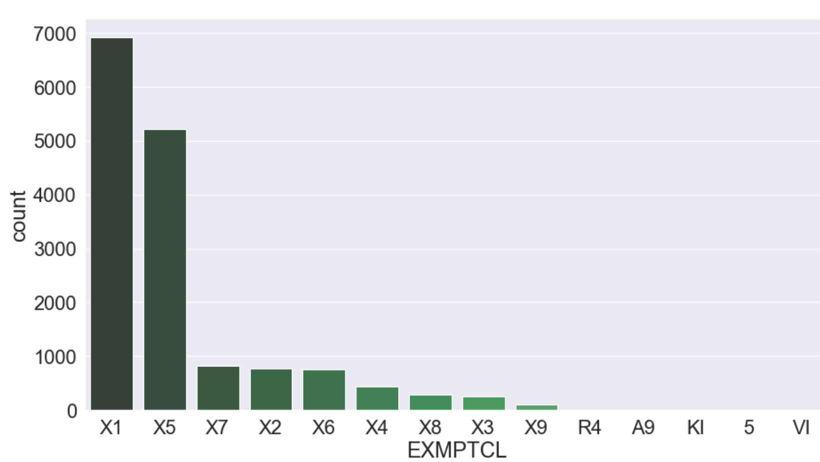
STADDR: street address

```
501 SURF AVENUE             902
330 EAST 38 STREET          817
322 WEST 57 STREET          720
155 WEST 68 STREET          671
20 WEST 64 STREET           657
1 IRVING PLACE              650
220 RIVERSIDE BOULEVARD     628
360 FURMAN STREET           599
200 EAST 66 STREET          585
30 WEST 63 STREET           562
2 BAY CLUB DRIVE            556
350 WEST 42 STREET          556
200 RECTOR PLACE            549
301 EAST 79 STREET          538
350 WEST 50 STREET          498
Name: STADDR, dtype: int64
```

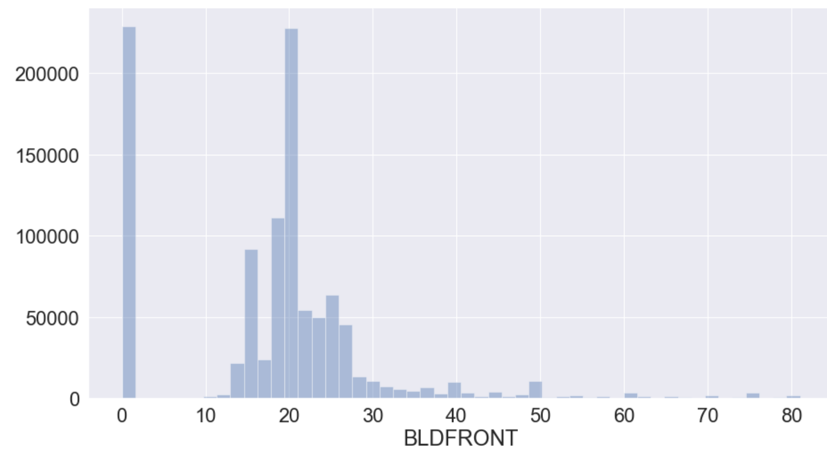
ZIP: zip code

```
10314.0    24606
11234.0    20001
10312.0    18127
10462.0    16905
10306.0    16578
11236.0    15678
11385.0    14921
11229.0    12793
11211.0    12710
11207.0    12293
11215.0    11834
11235.0    11312
11203.0    11241
11208.0    11139
11204.0    11061
10469.0    11030
11214.0    10886
11223.0    10741
10305.0    10625
11434.0    10505
Name: ZIP, dtype: int64
```

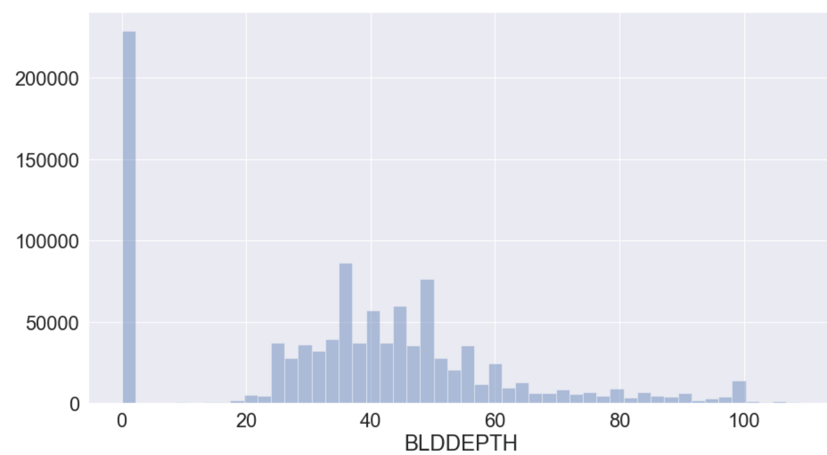
EXMPTCL: exemption class used for fully exempt properties



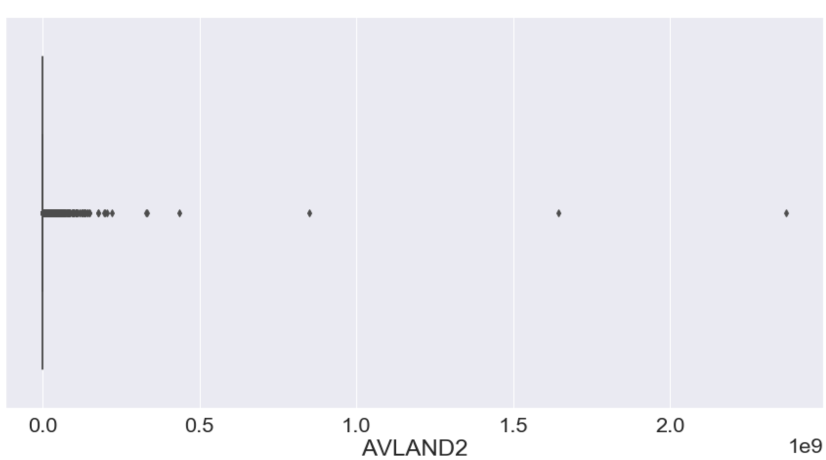
BLDFRONT: building width



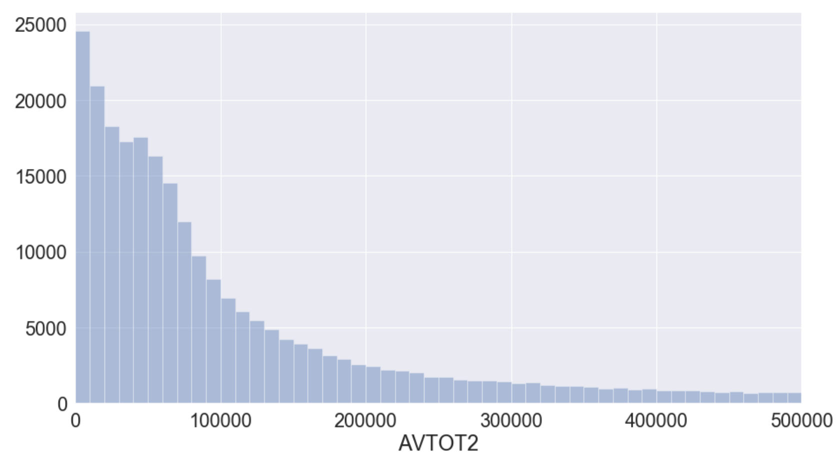
BLDDEPTH: building depth



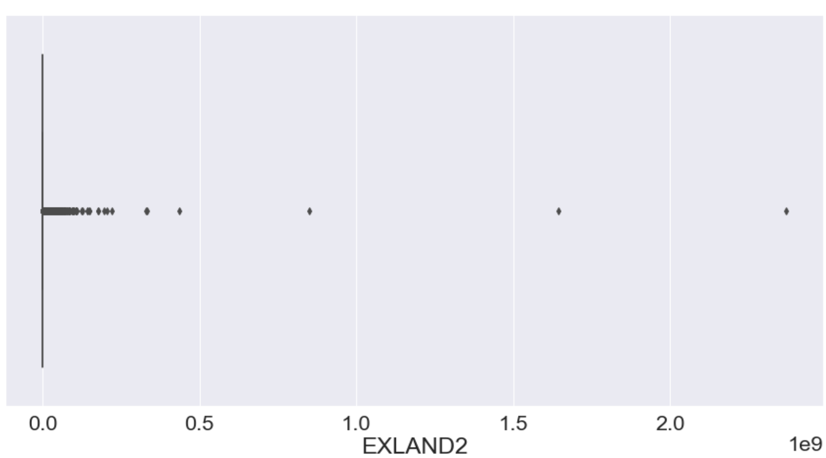
AVLAND2: transitional land value

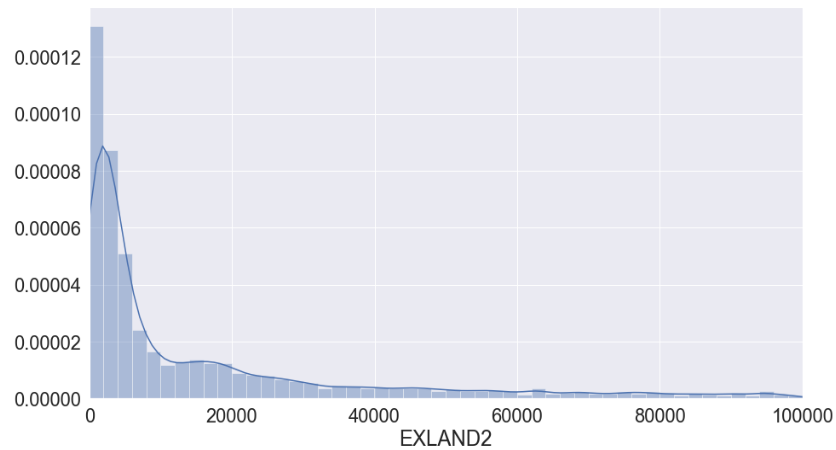


AVTOT2: transitional total value

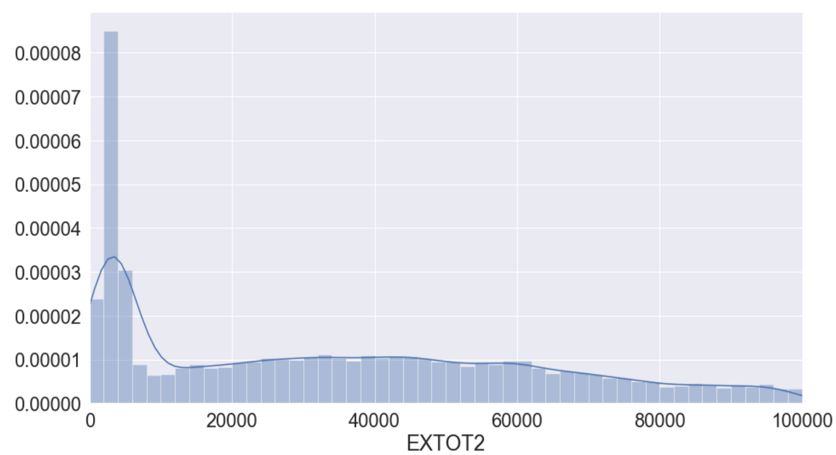


EXLAND2: transitional exempt land value

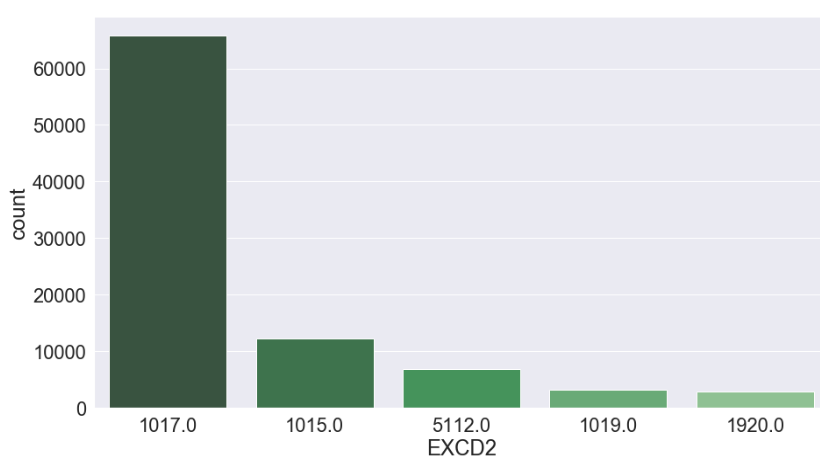




EXTOT2: transitional exempt total value



EXCD2: exemption code 2



PERIOD. Assessment period when data was created. All data in this dataset is at the fi nal period.

YEAR. Properties in this dataset are assessed in 2010/11.

VALTYPE. All data has a valtype of AC-TR.