DATA QUALITY REPORT

NY PROPERTY DATA

By Shiyi (Letty) Hua

Section 1 – Introduction of Data

1.1 **Original Name:** Property Valuation and Assessment Data

1.2 **Agency:** Department of Finance

1.3 **Update Frequency:** Annually

1.4 **Description:** This is Real Estate Assessment Property data. To be more specific, the data represent NYC properties assessments for purpose to calculate Property Tax, Grant eligible properties Exemptions and/or Abatements. The data collected and entered into the system by various City employee, like Property Assessors, Property Exemption specialists, ACRIS reporting, Department of Building reporting, etc.

1.5 **Source:** Can be downloaded from NYC OpenData

1.6 **Time Period:** 2010/11

1.7 **Number of Fields:** 32

1.8 **Number of Records:** 1,070,994

Section 2 – Summary of Fields

Table 2.1 Information of All Fields:

Field Name	*Field Type*	*# Non-null Records*	*% Populated*	*# Zero Records*
RECORD	int64	1,070,994	100	0
BBLE	object	1,070,994	100	0
B	int64	1,070,994	100	0
BLOCK	int64	1,070,994	100	0
LOT	int64	1,070,994	100	0
EASEMENT	object	4,636	0.43	0
OWNER	object	1,039,251	97.04	0
BLDGCL	object	1,070,994	100	0
TAXCLASS	object	1,070,994	100	0
LTFRONT	int64	1,070,994	100	169,108
LTDEPTH	int64	1,070,994	100	464,541
EXT	object	354,305	33.08	0

STORIES	float64	1,014,730	94.75	0
FULLVAL	float64	1,070,994	100	13,007
AVLAND	float64	1,070,994	100	13,009
AVTOT	float64	1,070,994	100	13,007
EXLAND	float64	1,070,994	100	491,699
EXTOT	float64	1,070,994	100	432,572
EXCD1	float64	638,488	59.62	0
STADDR	object	1,070,318	99.94	0
ZIP	float64	1,041,104	97.21	0
EXMPTCL	object	15,579	1.45	0
BLDFRONT	int64	1,070,994	100	228,815
BLDDEPTH	int64	1,070,994	100	228,853
AVLAND2	float64	282,726	26.40	0
AVTOT2	float64	282,732	26.40	0
EXLAND2	float64	87,449	8.17	0
EXTOT2	float64	130,828	12.22	0
EXCD2	float64	92,948	8.68	0
PERIOD	object	1,070,994	100	0
YEAR	object	1,070,994	100	0
VALTYPE	object	1,070,994	100	0

Table 2.2 Summary Statistics of Numeric Fields:

Field Name	*Mean*	*Standard Deviation*	*Min*	*Max*
LTFRONT	36.64	74.03	0	9999
LTDEPTH	88.86	76.40	0	9999
STORIES	5.01	8.37	1	119
FULLVAL	874,264.5	11,582,430	0	6,150,000,000
AVLAND	85,067.92	4,057,260	0	2,668,500,000
AVTOT	227,238.2	6,877,529	0	4,668,309,000
EXLAND	36,423.89	3,981,576	0	2,668,500,000
EXTOT	91,186.98	6,508,403	0	4,668,309,000
EXCD1	1,602.01	1,384.23	1,010	7,170
BLDFRONT	23.04	35.58	0	7,575
BLDDEPTH	39.92	42.71	0	9,393
AVLAND2	246,235.7	6,178,963	3	2,371,005,000
AVTOT2	713,911.4	11,652,530	3	4,501,180,000
EXLAND2	351,235.7	10,802,210	1	2,371,005,000
EXTOT2	656,768.3	16,072,510	7	4,501,180,000
EXCD2	1,364.04	1,094.71	1,011	7,160

Table 2.3 Summary Information of Categorical Fields:

Field Name	*# Unique Values*	*Most Common Field Value*
BBLE	1,070,994	N/A
B	5	4
BLOCK	13,984	3944
LOT	6,366	1
EASEMENT	13	E
OWNER	863,349	PARKCHESTER PRESERVAT
BLDGCL	200	R4
TAXCLASS	11	1
EXT	4	G
STADDR	839,281	501 SURF AVENUE
ZIP	197	10314
EXMPTCL	15	X1
PERIOD	1	FINAL
YEAR	1	2010/11
VALTYPE	1	AC-TR

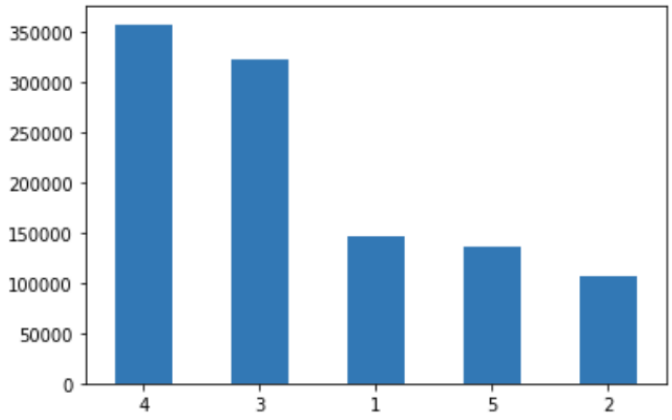
Section 3 – Description of Each Field

3.1 **RECORD:** Unique integer identifier to label each record of the dataset

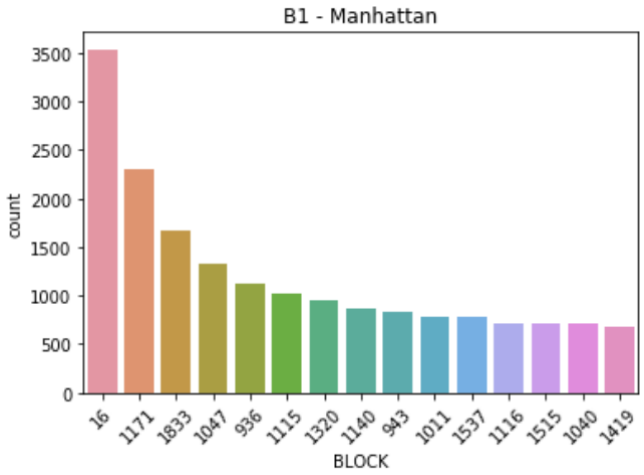
3.2 **BBLE:** File key. Concatenation of B, BLOCK, LOT, and EASEMENT. Unique BBLE for each record of the dataset. The distribution of field BBLE is that each of 1,070,994 values occurs once. Below is a table of 10 BBLE values (as examples) with their frequencies.

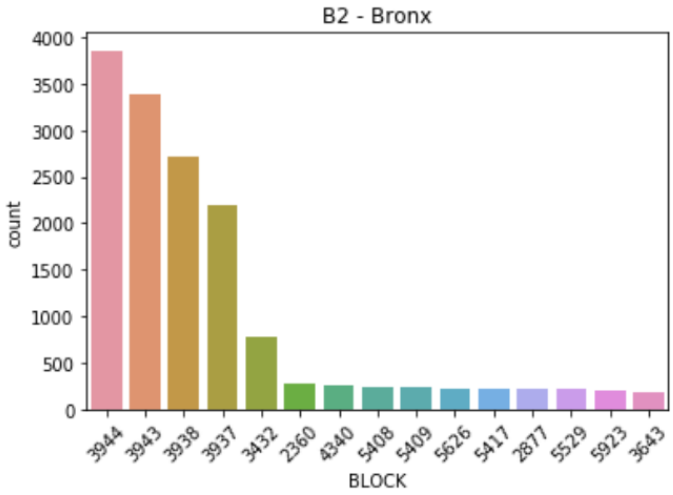
5002540061	1
4005391001	1
4051371088	1
4161131182	1
1011430020	1
2055370092	1
3018600037	1
2024280027	1
3026180027	1
1021310004	1

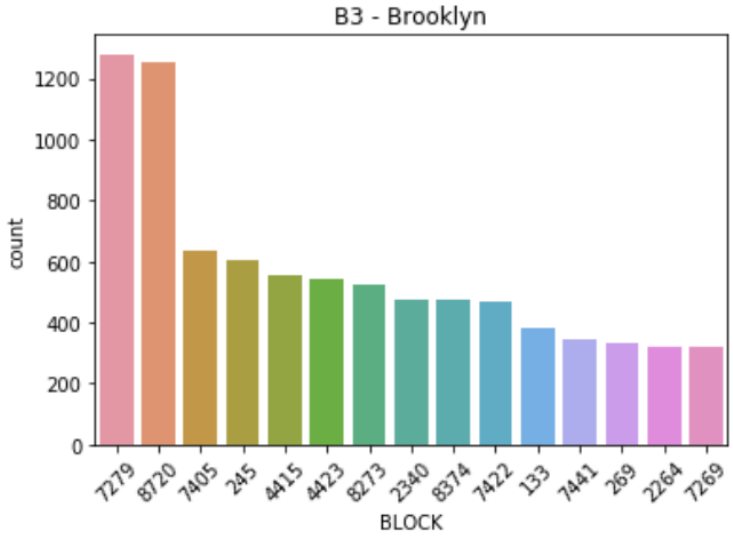
3.3 **B:** Borough codes. 1 = Manhattan, 2 = Bronx, 3 = Brooklyn, 4 = Queens, and 5 = Staten Island. Below is a bar char of each unique code with its frequency.

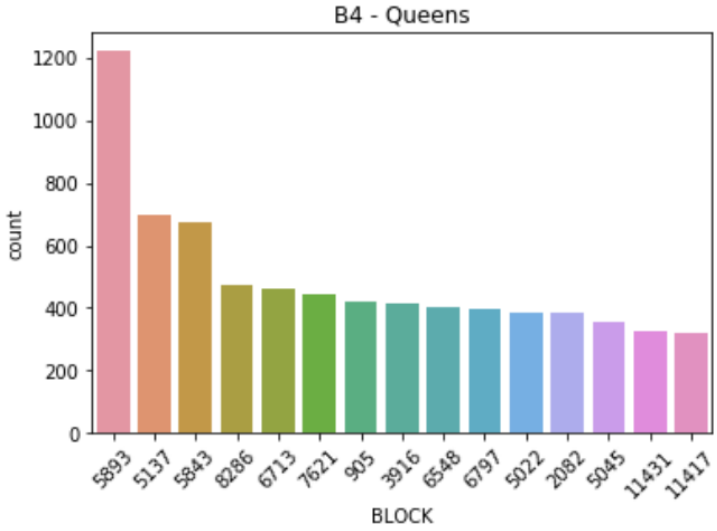


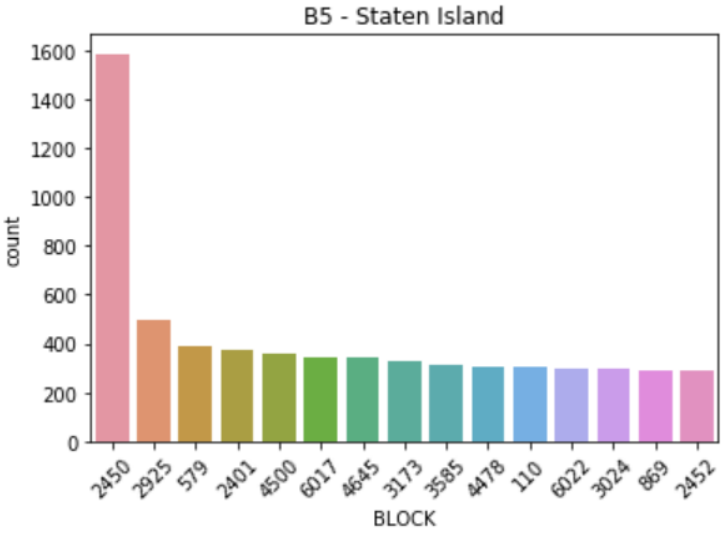
3.4 **BLOCK:** Valid Block ranges by B. Manhattan 1 to 2,255, Bronx 2,260 to 5,958, Brooklyn 1 to 8,955, Queens 1 to 16,350, and Staten Island 1 to 8,050. Below are five bar charts of top 15 blocks with their counts within each borough code.



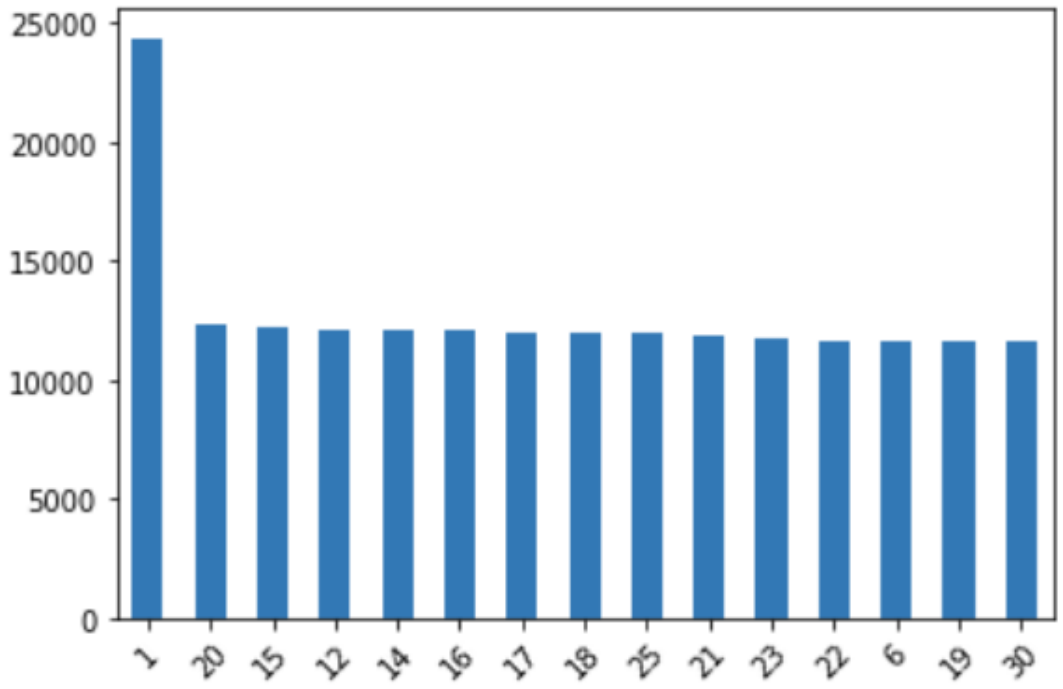



B3 - Brooklyn
count
BLOCK
7279
8720
7405
245
4415
4423
8273
2340
8374
7422
133
7441
269
2264
7269

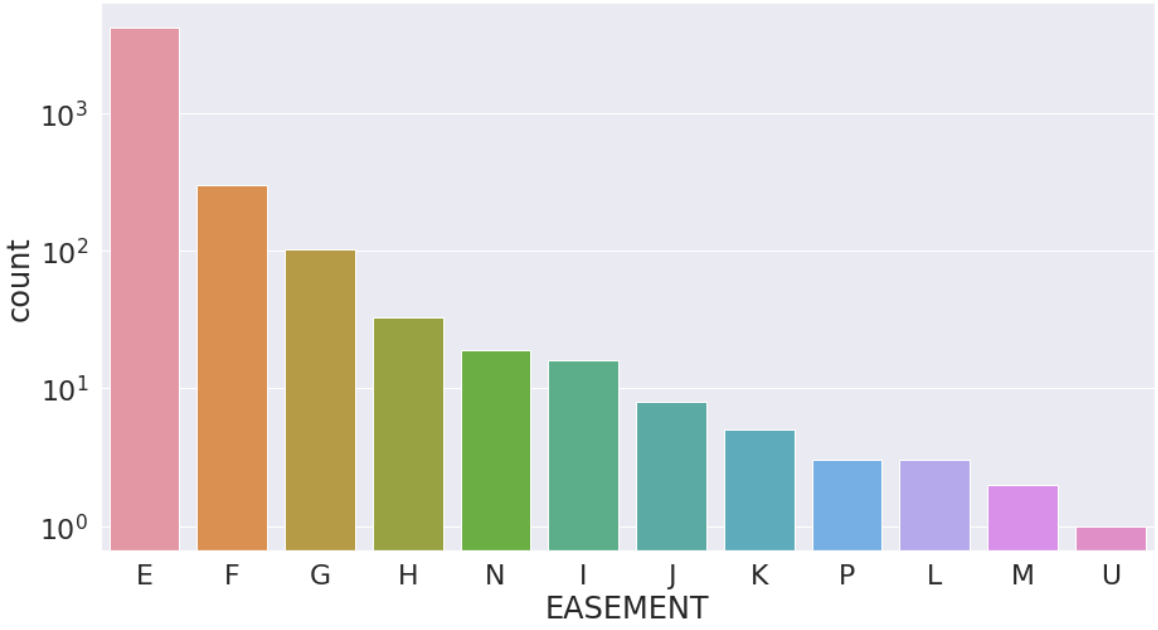

B4 - Queens
count
BLOCK
5893
5137
5843
8286
6713
7621
905
3916
6548
6797
5022
2082
5045
11431
11417


B5 - Staten Island
count
BLOCK
2450
2925
579
2401
4500
6017
4645
3173
3585
4478
110
6022
3024
869
2452

3.5 **LOT:** Unique # within B/BLOCK. Below is a bar chart of top 15 lots with their counts.



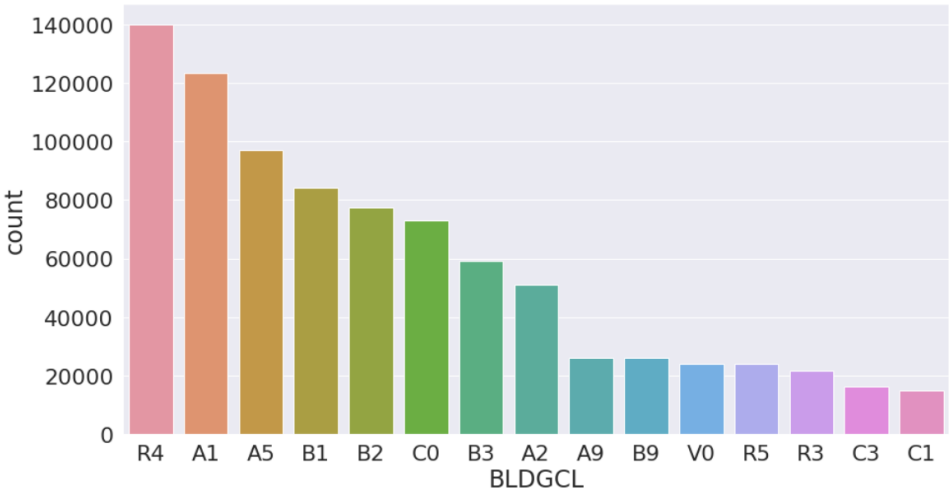
3.6 **EASEMENT:** A field that is used to describe easement. SPACE indicates the lot has no easement. 'A' indicates the portion of the lot that has an air easement. 'B' indicates non-air rights. 'E' indicates the portion of the lot that has a land easement. 'F' through 'M' are duplicates of 'E'. 'N' indicates non-transit easement. 'P' indicates piers. 'R' indicates railroads. 'S' indicates street. 'U' indicates U.S. government. Below is a bar chart of. Below is a bar chart of each unique easement with its frequency.



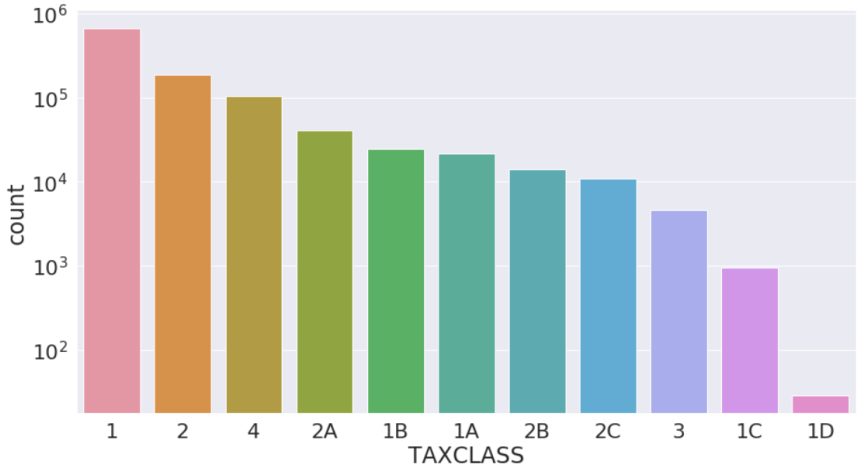
3.7 **OWNER:** Owner's Name. Below is a table of top 10 owner names (as examples) with their frequencies.

PARKCHESTER PRESERVAT	6021
PARKS AND RECREATION	4255
DCAS	2169
HOUSING PRESERVATION	1904
CITY OF NEW YORK	1450
DEPT OF ENVIRONMENTAL	1166
BOARD OF EDUCATION	1015
NEW YORK CITY HOUSING	1014
CNY/NYCTA	975
NYC HOUSING PARTNERSH	747

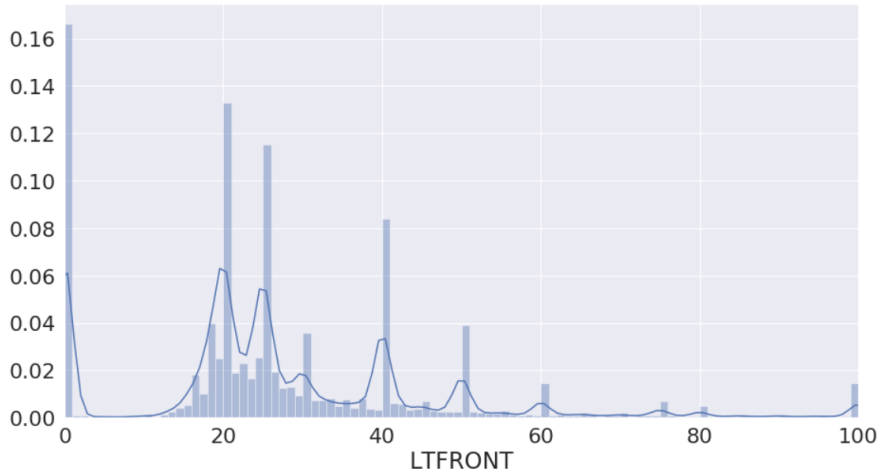
3.8 **BLDGCL:** Building class. Position 1 = ALPHA. Position 2 = NUMERIC. Below is a bar chart of top building classes with their frequencies.



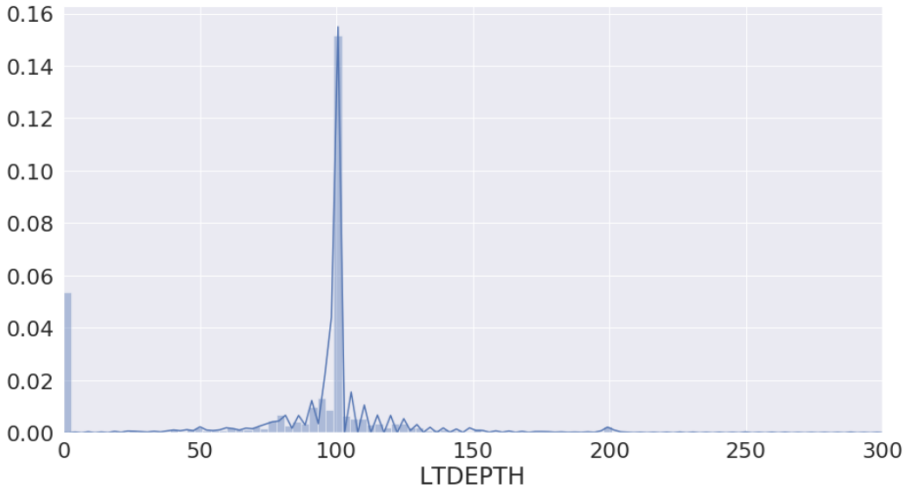
3.9 **TAXCLASS:** Current property tax class code (NYS Classification). Below is a bar chart of each unique property tax class with its frequency.



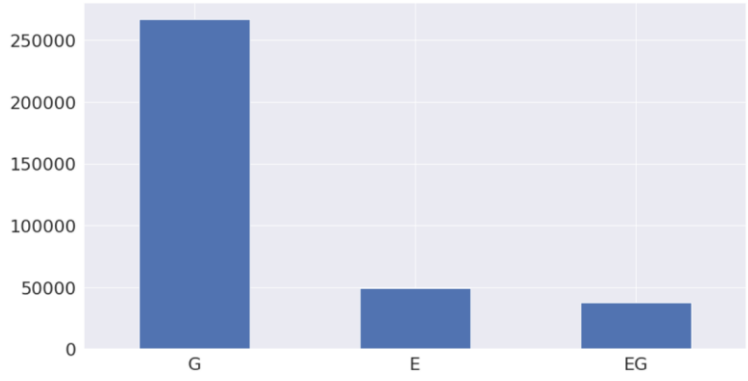
3.10 **LTFRONT:** Lot frontage in feet. Below is a distribution plot of LTFRONT with a gaussian kernel density estimate. Selected values are under 100.



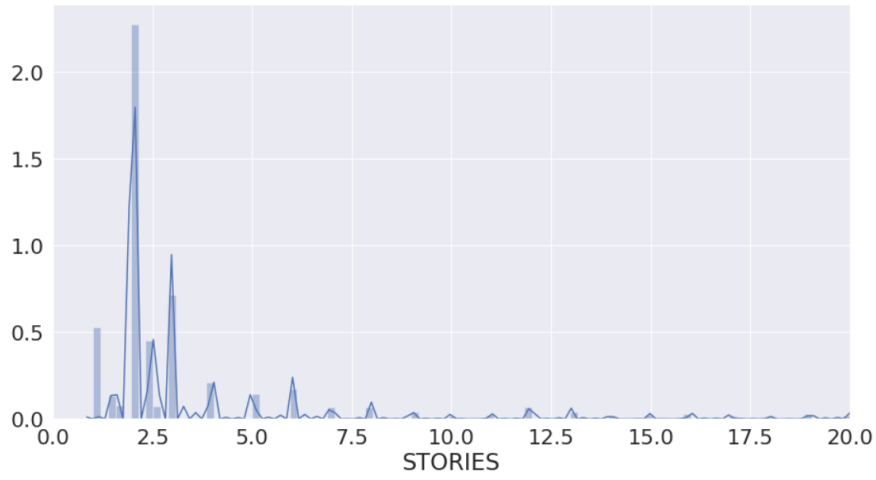
3.11 **LTDEPTH:** Lot depth in feet. Below is a distribution plot of LTDEPTH with a gaussian kernel density estimate. Selected values are under 300.



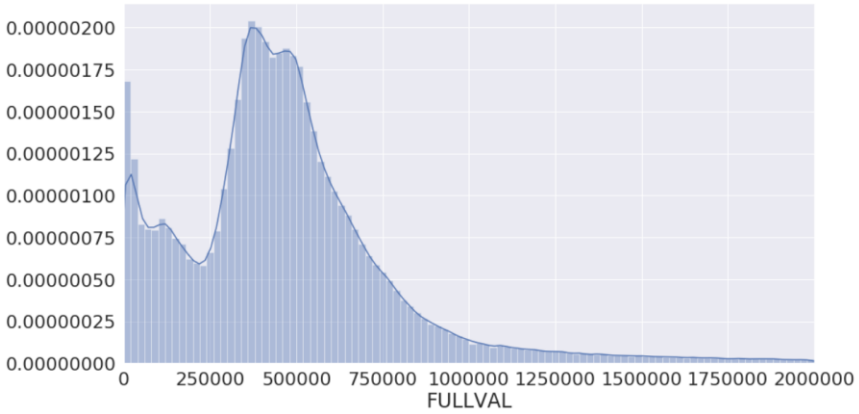
3.12 **EXT:** Extension indicator. 'E' = Extension. 'G' = Garage. 'EG' = Extension and Garage. Below is a bar chart of each unique EXT with its frequency.



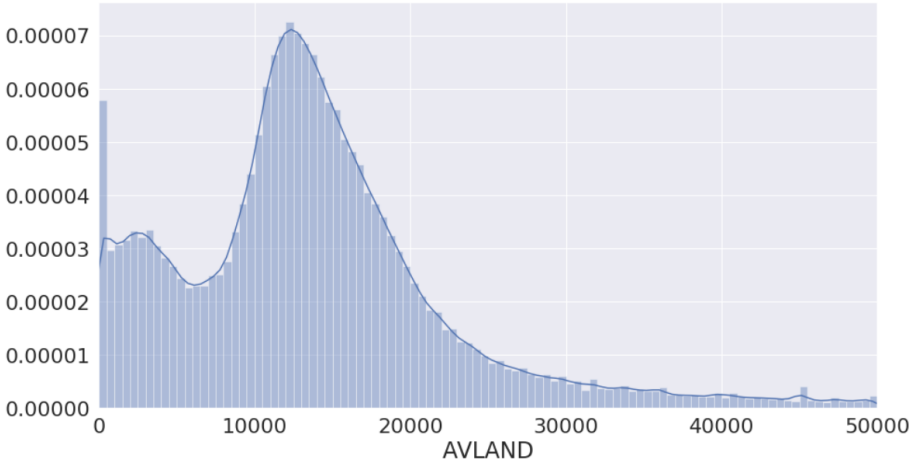
3.13 **STORIES:** The number of stories for the building (# of floors). Below is a distribution plot of STORIES with a gaussian kernel density estimate. Selected values are under 20.



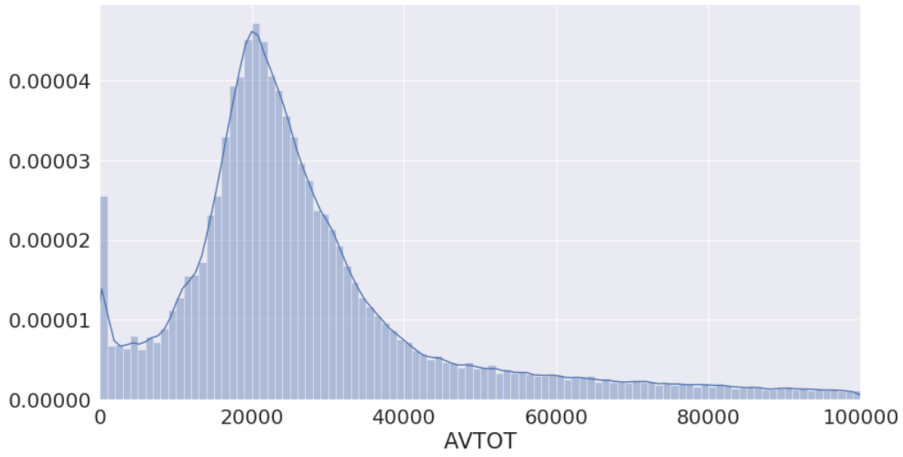
3.14 **FULLVAL:** Total market value of property. Below is a distribution plot of FULLVAL with a gaussian kernel density estimate. Selected values are under 2000000.



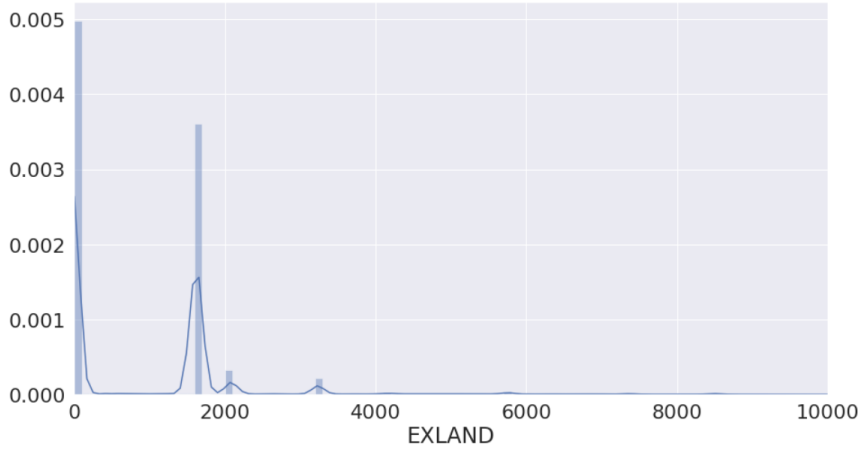
3.15 **AVLAND:** If not zero, the total land area. Below is a distribution plot of AVLAND with a gaussian kernel density estimate. Selected values are under 50000.



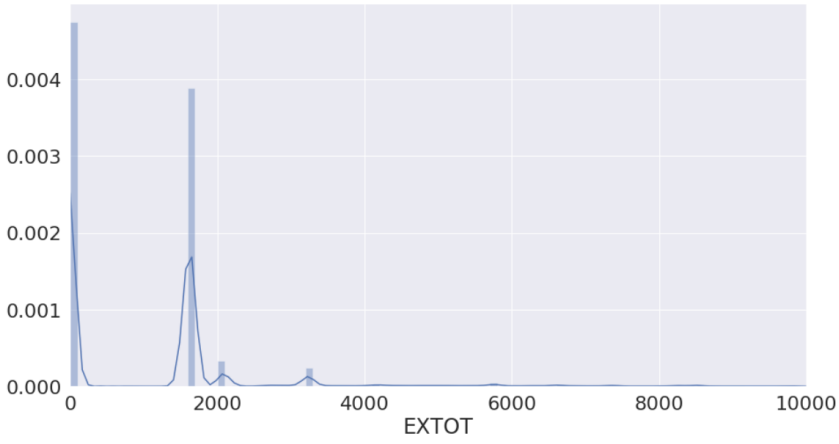
3.16 **AVTOT:** Total number of units in the building. Below is a distribution plot of AVTOT with a gaussian kernel density estimate. Selected values are under 100000.



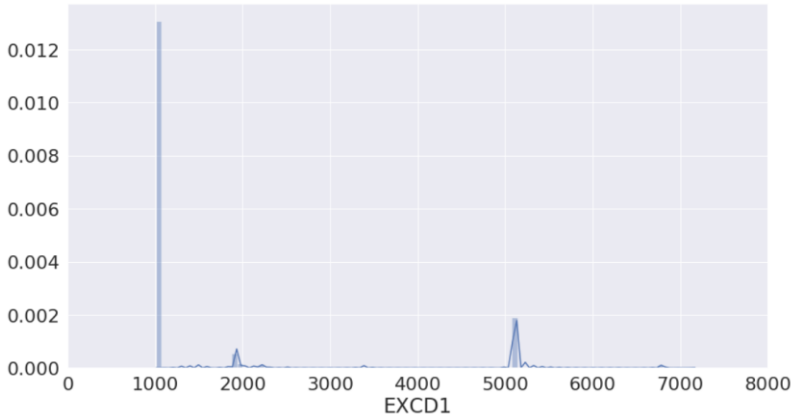
3.17 **EXLAND:** The exempt land value. Below is a distribution plot of EXLAND with a gaussian kernel density estimate. Selected values are under 10000.



3.18 **EXTOT:** The exempt number of units in the building. Below is a distribution plot of EXTOT with a gaussian kernel density estimate. Selected values are under 10000.



3.19 **EXCD1:** Exemption code 1. Below is a distribution plot of EXCD1 with a gaussian kernel density estimate. Selected values are under 8000.



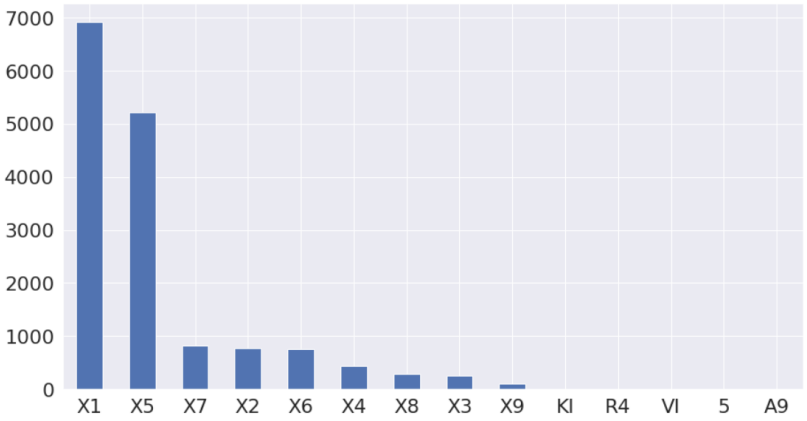
3.20 **STADDR:** Street name of the property. Below is a table of top 10 street names (as examples) with their frequencies.

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

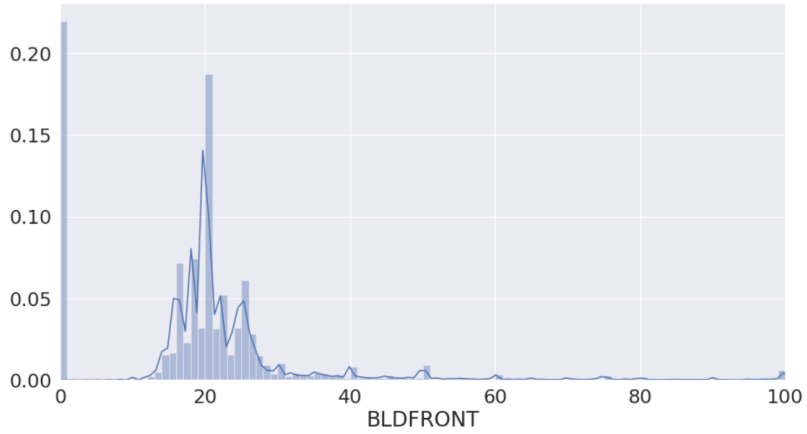
3.21 **ZIP:** Postal zip code of the property. Below is a table of top 10 zip codes (as examples) with their frequencies.

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

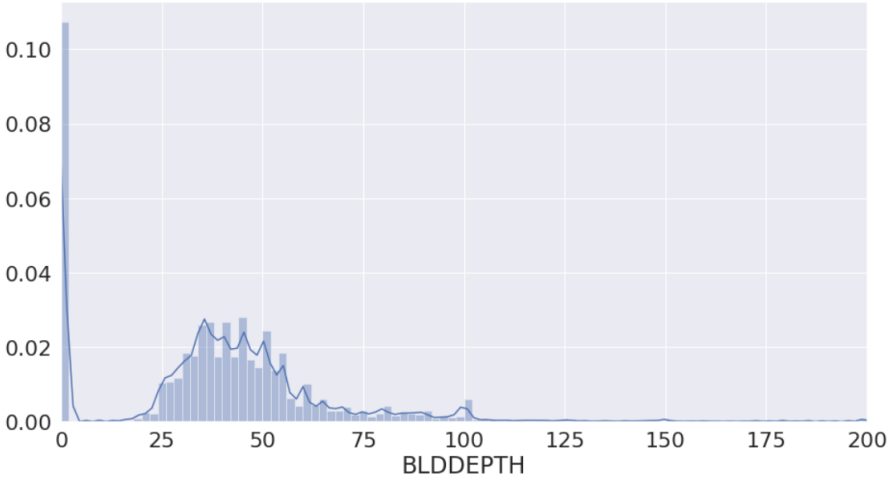
3.22 **EXMPTCL:** Exempt class used for fully exempt properties only. Below is a bar chart of each unique exempt class with its frequency.



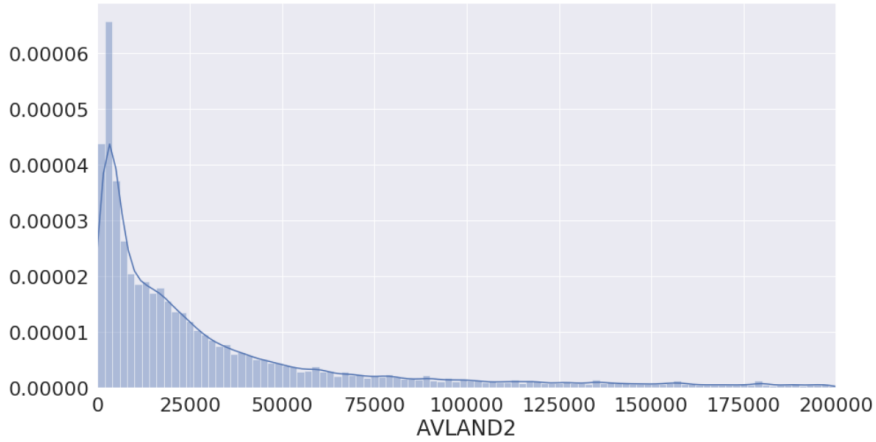
3.23 **BLDFRONT:** Building frontage in feet. Below is a distribution plot of BLDFRONT with a gaussian kernel density estimate. Selected values are under 100.



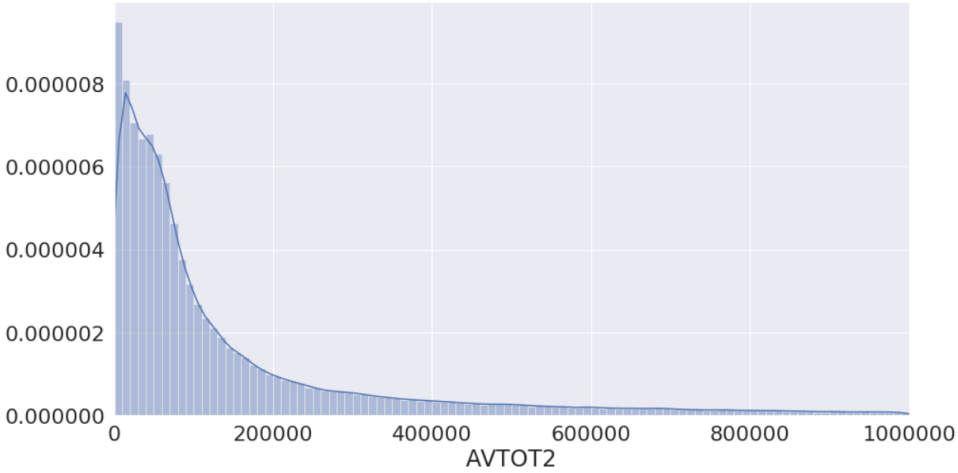
3.24 **BLDDEPTH:** Lot depth in feet. Below is a distribution plot of BLDDEPTH with a gaussian kernel density estimate. Selected values are under 200.



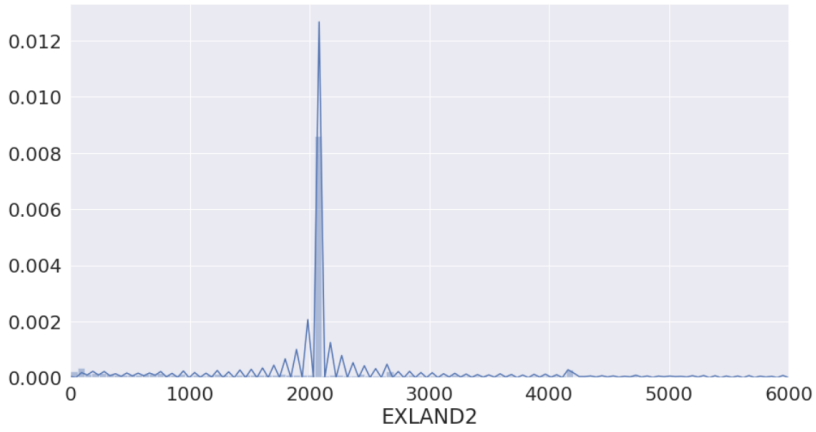
3.25 **AVLAND2:** The total land area. Below is a distribution plot of AVLAND2 with a gaussian kernel density estimate. Selected values are under 200000.



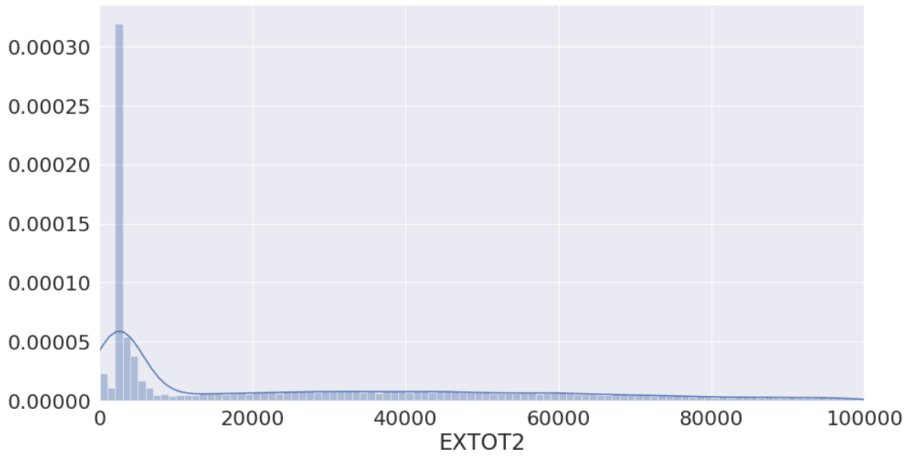
3.26 **AVTOT2:** Total number of units in the building. Below is a distribution plot of AVTOT2 with a gaussian kernel density estimate. Selected values are under 1000000.



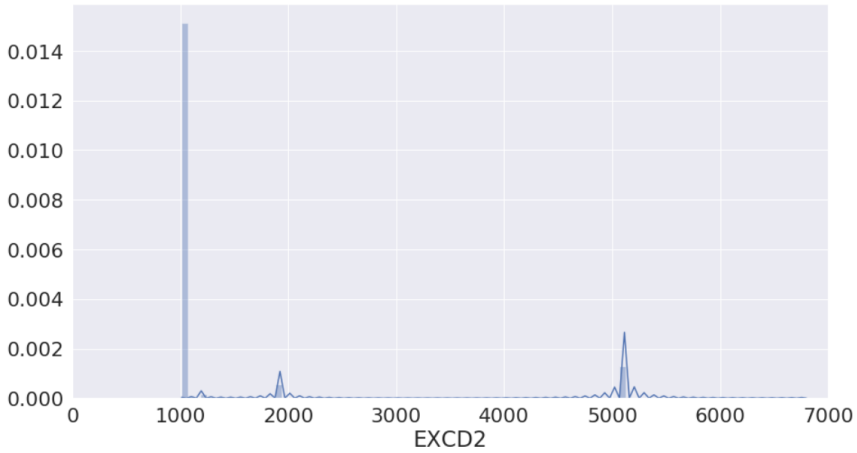
3.27 **EXLAND2:** The exempt land value. Below is a distribution plot of EXLAND2 with a gaussian kernel density estimate. Selected values are under 6000.



3.28 **EXTOT2:** The exempt number of units in the building. Below is a distribution plot of EXTOT2 with a gaussian kernel density estimate. Selected values are under 100000.



3.29 **EXCD2:** Exemption code 2. Below is a distribution plot of EXCD2 with a gaussian kernel density estimate. Selected values are under 8000.



3.30 **PERIOD:** Identical for all records of the dataset. Below is the PERIOD value with its frequency.

FINAL	1070994

3.31 **YEAR:** Identical for all records of the dataset. Below is the YEAR value with its frequency.

2010/11	1070994

3.32 **VALTYPE:** Identical for all records of the dataset. Below is the VALTYPE value with its frequency.

AC-TR	1070994