

# **DATA QUALITY REPORT**

## **NEW YORK PROPERTY VALUATION AND ASSESSMENT**

**Prepared for:**

Prof. Coggeshall

**Prepared by:**

Chengyin Liu

January 24, 2019

## TABLE OF CONTENTS

<b>1.0 INTRODUCTION.....</b>	<b>1</b>
<b>2.0 FILEDS SUMMARY.....</b>	<b>1</b>
<b>3.0 DATA QUALITY ASSESSMENT .....</b>	<b>1</b>
3.01 RECORD.....	1
3.02 BBLE .....	1
3.03 B .....	1
3.04 BLOCK .....	1
3.05 LOT.....	2
3.06 EASEMENT .....	2
3.07 OWNER .....	3
3.08 BLDGCL.....	4
3.09 TAXCLASS .....	5
3.10 LTFRONT .....	6
3.11 LTDEPTH.....	7
3.12 EXT .....	7
3.13 STORIES.....	8
3.14 FULLVAL .....	9
3.15 AVLAND .....	10
3.16 AVTOT .....	11
3.17 EXLAND .....	12
3.18 EXTOT .....	13
3.19 EXCD1 .....	14
3.20 STADDR.....	14
3.21 ZIP .....	15
3.22 EXMPTCL .....	15
3.23 BLDFRONT.....	16
3.24 BLDDEPTH .....	17
3.25 AVLAND2 .....	18
3.26 AVTOT2.....	19
3.27 EXLAND2 .....	20

<b>3.28 EXTOT2 .....</b>	<b>21</b>
<b>3.29 EXCD2 .....</b>	<b>22</b>
<b>3.30 PERIOD.....</b>	<b>22</b>
<b>3.31 YEAR .....</b>	<b>22</b>
<b>3.32 VALTYPE.....</b>	<b>22</b>

## 1.0 INTRODUCTION

The name of the Dataset is Property Valuation and Assessment Data, it is updated annually and is owned by New York City (NYC). The dataset this report is going to evaluate is updated on September 10, 2018. Data represent NYC properties assessments for purpose to calculate Property Tax, Grant eligible properties Exemptions and/or Abatements. Data collected and entered into the system by various City employee,

**Source:** <https://data.cityofnewyork.us/Housing-Development/Property-Valuation-and-Assessment-Data/rgy2-tti8>

**Covered Time Period:** 2010 - 2011

**Number of Fields:** 32

## 2.0 FILEDS SUMMARY

Table 2.1 (a)

No.	Field Name	Field Type	# Records	% Populated	# Unique Values	# Records with Value NA/" "	Details
1	RECORD	ordinal	1,070,994	100.00%	1,070,994	0	From 1 to 1,070,994
2	BBLE	Categorical	1,070,994	100.00%	1,070,994	0	Concatenation of B, BLOCK, LOT, EASEMENT
3	B	Categorical	1,070,994	100.00%	5	0	1: 146,220 2: 107,285 3: 358,046 4: 358,046 5: 136,200
4	BLOCK	Categorical	1,070,994	100.00%	13,984	0	3944: 3,888 16: 3,786 3943: 3,424 ...
5	LOT	Categorical	1,070,994	100.00%	6,366	0	1: 24,367 20: 12,294 15: 12,171 ...
6	EASEMENT	Categorical	1,070,994	0.43%	13	1,066,358	E: 4,148 F: 296 G: 102 ...
7	OWNER	String	1,070,994	97.04%	863,348	31,745	PARKCHESTER PRESERVAT: 6,021 PARKS AND RECREATION: 4,255 DCAS: 2,169 ...
8	BLDGCL	Categorical	1,070,994	100.00%	200	0	R4: 139,879 A1: 123,369 A5: 96,984 ...

Table 2.1 (b)

No.	Field Name	Field Type	# Records	% Populated	# Unique Values	# Records with Value NA/" "	Details
9	TAXCLASS	Categorical	1,070,994	100.00%	11	0	1: 660,721 2: 188,612 4: 104,310 ...
10	LTFRONT	Numeric	1,070,994	100.00%	—	0	Unit: Feet Min: 0.00 Max: 9,999.00 Mean: 36.63 Std: 74.03
11	LTDEPTH	Numeric	1,070,994	100.00%	—	0	Unit: Feet Min: 0.00 Max: 9,999.00 Mean: 88.86 Std: 76.40
12	EXT	Categorical	1,070,994	33.08%	4	716,689	G: 266,970 E: 49,442 EG: 37,893
13	STORIES	Numeric	1,070,994	94.75%	111	56,264	Unit: Floor Min: 119.00 Max: 5.01 Mean: 5.01 Std: 8.37
14	FULLVAL	Numeric	1,070,994	100.00%	—	0	Unit: US Dollar Min: 0.00 Max: 6,150,000,000.00 Mean: 874,264.51 Std: 11,582,430.99
15	AVLAND	Numeric	1,070,994	100.00%	—	0	Unit: US Dollar Min: 0.00 Max: 2,668,500,000.00 Mean: 85,067.91 Std: 4,057,260.06
16	AVTOT	Numeric	1,070,994	100.00%	—	0	Unit: US Dollar Min: 0.00 Max: 4,668,308,947.00 Mean: 227,238.17 Std: 6,877,529.31
17	EXLAND	Numeric	1,070,994	100.00%	—	0	Unit: US Dollar Min: 0.00 Max: 2,668,500,000.00 Mean: 36,423.89 Std: 3,981,575.79

Table 2.1 (c)

No.	Field Name	Field Type	# Records	%Populated	# Unique Values	# Records with Value NA/" "	Details
18	EXTOT	Numeric	1,070,994	100.00%	—	0	Unit: US Dollar Min: 0.00 Max: 4,668,308,947.00 Mean: 91,186.98 Std: 6,508,402.82
19	EXCD1	Categorical	1,070,994	59.62%	130	432,506	1017: 425,348 1010: 49,756 1015: 31,323 ...
20	STADDR	String	1,070,994	99.94%	839,281	676	501 SURF AVENUE: 902 330 EAST 38 STREET: 817 322 WEST 57 STREET: 720 ...
21	ZIP	Categorical	1,070,994	97.21%	196	29,890	10314: 24,606 11234: 20,001 10312: 18,127 ...
22	EXMPTCL	Categorical	1,070,994	1.45%	15	1,055,415	X1: 6,912 X5: 5,208 X7: 820 ...
23	BLDFRONT	Numeric	1,070,994	100.00%	—	0	Unit: Feet Min: 0.00 Max: 7,575.00 Mean: 23.04 Std: 35.58
24	BLDDEPTH	Numeric	1,070,994	100.00%	—	0	Unit: Feet Min: 0.00 Max: 9,393.00 Mean: 39.92 Std: 42.71
25	AVLAND2	Numeric	1,070,994	26.40%	—	788,268	Unit: US Dollar Min: 3.00 Max: 2,371,005,000.00 Mean: 246,235.72 Std: 6,178,962.56
26	AVTOT2	Numeric	1,070,994	26.40%	—	788,262	Unit: US Dollar Min: 3.00 Max: 4,501,180,002.00 Mean: 713,911.44 Std: 11,652,528.95

Table 2.1 (d)

No.	Field Name	Field Type	# Records	% Populated	# Unique Values	# Records with Value NA/" "	Details
27	EXLAND2	Numeric	1,070,994	8.17%	—	983,545	Unit: US Dollar Min: 1.00 Max: 2,371,005,000.00 Mean: 351,235.68 Std: 10,802,212.67
28	EXTOT2	Numeric	1,070,994	12.22%	—	940,166	Unit: US Dollar Min: 7.00 Max: 4,501,180,002.0 Mean: 656,768.28 Std: 16,072,510.17
29	EXCD2	Categorical	1,070,994	8.68%	61	978,046	1017: 65,777 1015: 12,337 5112: 6867 ...
30	PERIOD	String	1,070,994	100.00%	1	0	FINAL: 1,070,994
31	YEAR	Date	1,070,994	100.00%	1	0	2010/11: 1,070,994
32	VALTYPE	String	1,070,994	100.00%	1	0	AC-TR: 1,070,994

## 3.0 DATA QUALITY ASSESSMENT

### 3.01 RECORD

FILE KEY, to uniquely identify each record of the AV Master, ranging from 1 to 1,070,994

### 3.02 BBLE

BBLE is an alphanumeric label in the length of 11, generated by concatenating the RECORD, B and BLOCK. Unique for each observation in the dataset.

### 3.03 B

B stands for borough codes

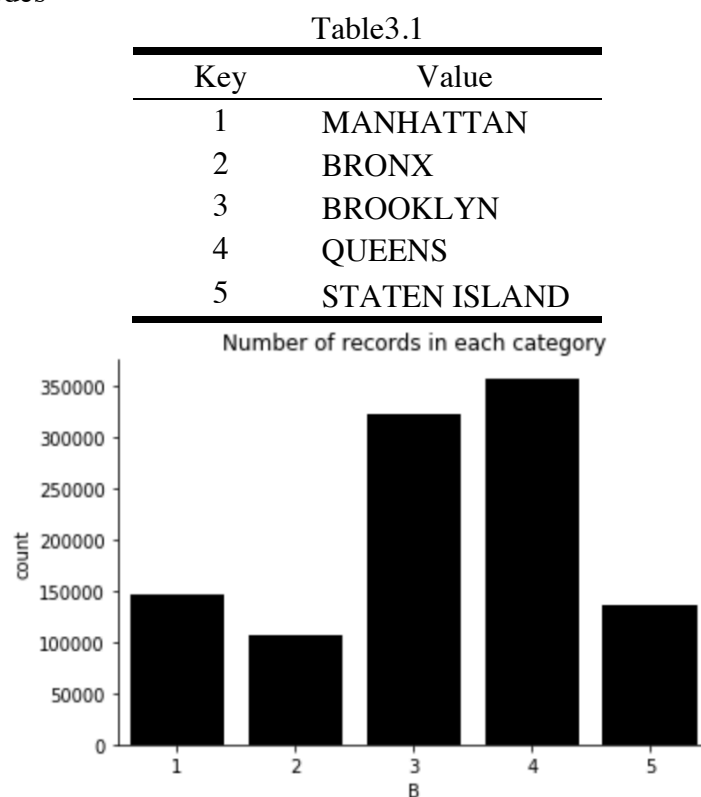


Figure 3.1

### 3.04 BLOCK

BLOCK stands for valid block ranges by borough

Table 3.2

Borough	Range
MANHATTAN	1 to 2,255
BRONX	2,260 to 5,958
BROOKLYN	1 to 8,955
QUEENS	1 to 16,350
STATEN ISLAND	1 to 8,050



### 3.05 LOT

LOT stands for unique number within borough /block

Table 3.3

LOT	Count
1	24367
20	12294
15	12171
12	12143
14	12074
16	12042
17	11982
...	...

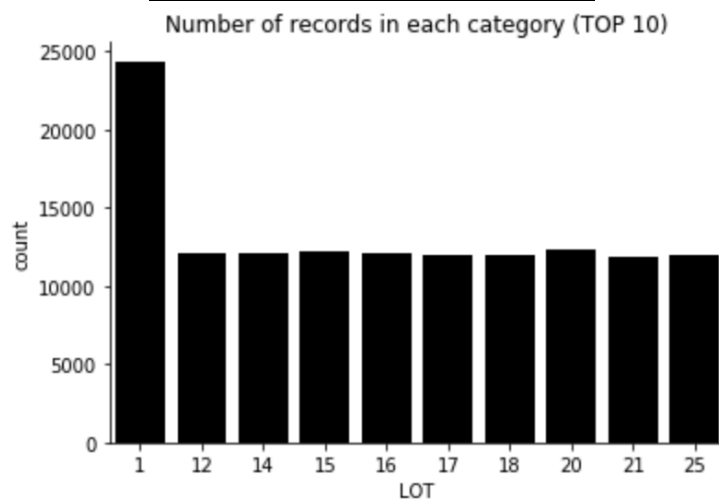


Figure 3.2

### 3.06 EASEMENT

EASEMENT is a field that is used to describe easement, contains 1,066,358 of NA values

Table 3.4

Filed	Description
Space	the lot has no Easement
A	the portion of the Lot that has an Air Easement
B	Non-Air Rights
E	the portion of the lot that has a Land Easement
F - M	duplicates of 'E'
N	Non-Transit Easement
P	Piers
R	Railroads
S	Street
U	U.S. Government

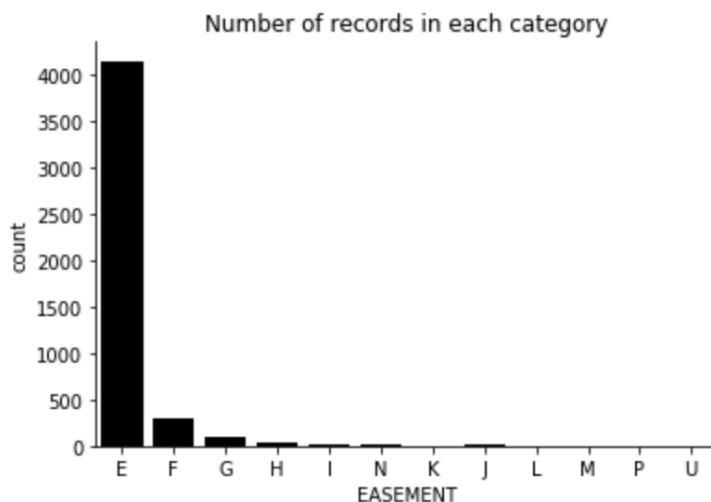


Figure 3.3

### 3.07 OWNER

WONER stands for owner's name, this field contains 31,745 of NA values.

Table 3.5

Owner	Count
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
...	...

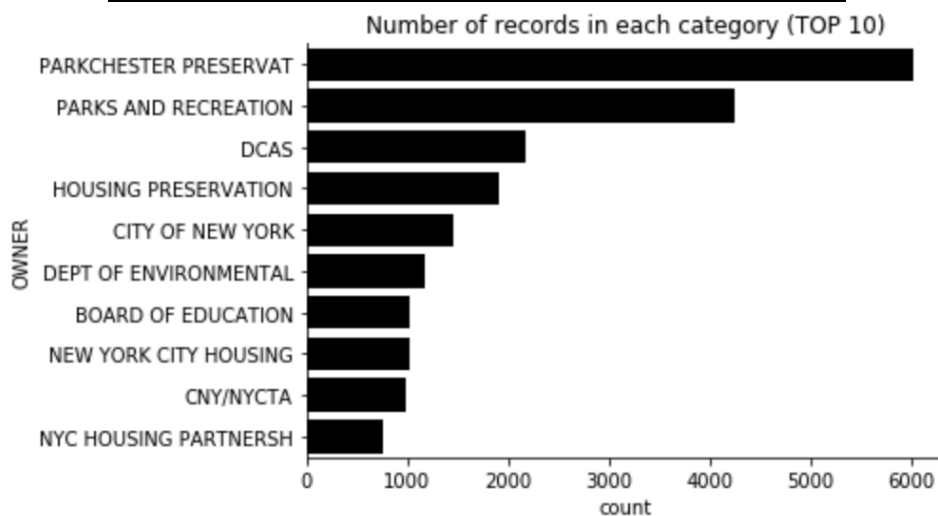


Figure 3.4

### 3.08 BLDGCL

BLDGCL stands for building class, where the first position is a letter and the second position is a number

Table 3.6

Field	Count
R4	139879
A1	123369
A5	96984
B1	84208
B2	77598
C0	73111
B3	59240
...	...

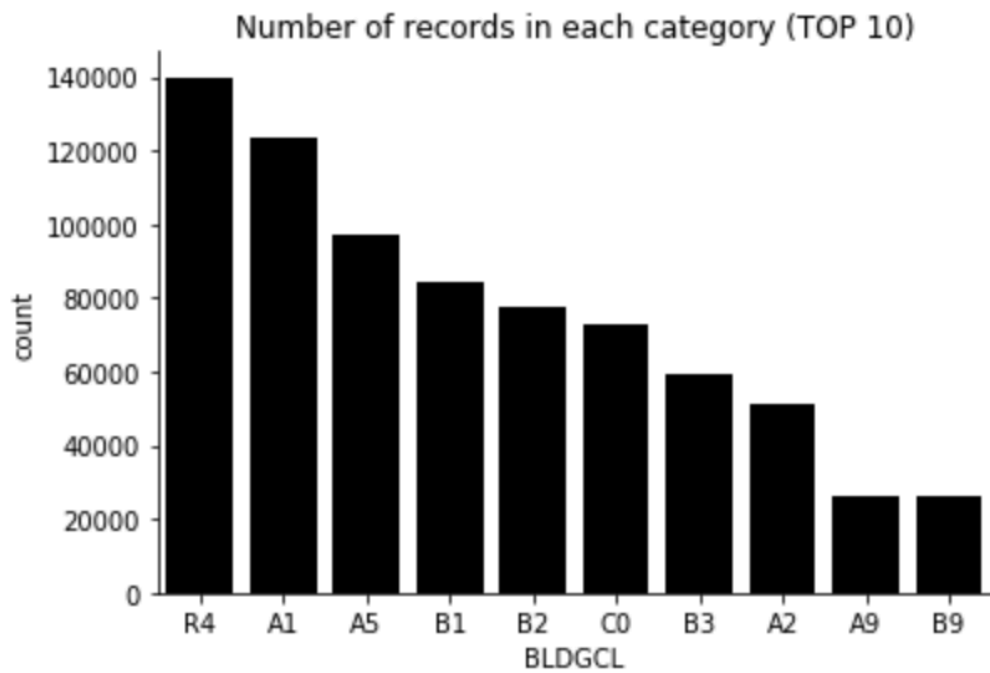


Figure 3.5

### 3.09 TAXCLASS

TAXCLASS stands for current property tax class code by NYS classification

Table 3.7

TAXCLASS	Count
1	660721
2	188612
4	104310
2A	40574
1B	24738
1A	21667
2B	13964
2C	10795
3	4638
1C	946
1D	29

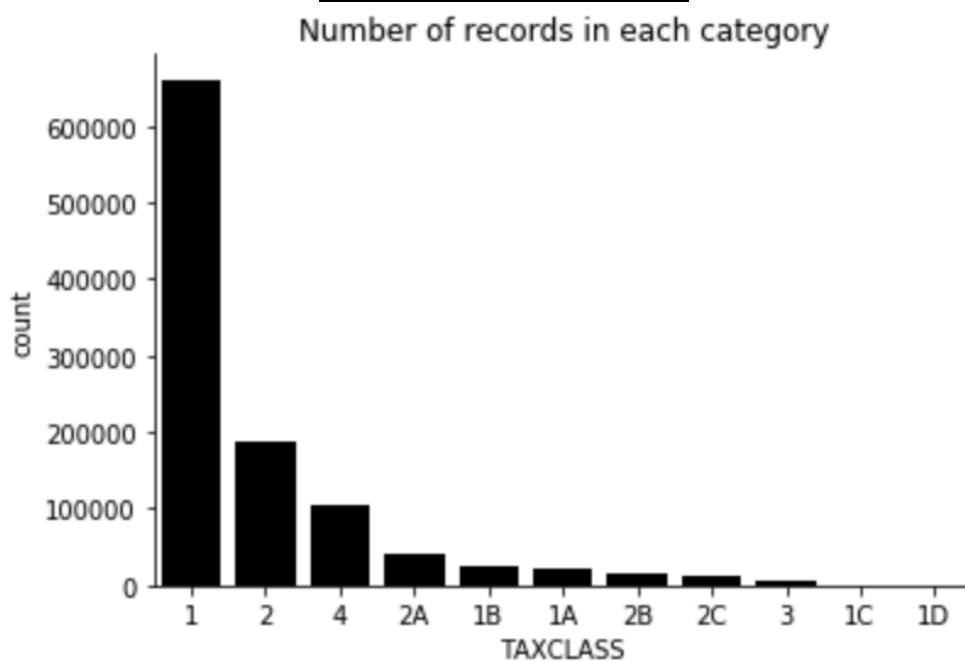


Figure 3.6

### 3.10 LTFRONT

LTFRONT stands for lot frontage (lot width) in feet.

Table 3.8

Unit	Max	Min	Mean	Std
Feet	0.00	9,999.00	36.63	74.03

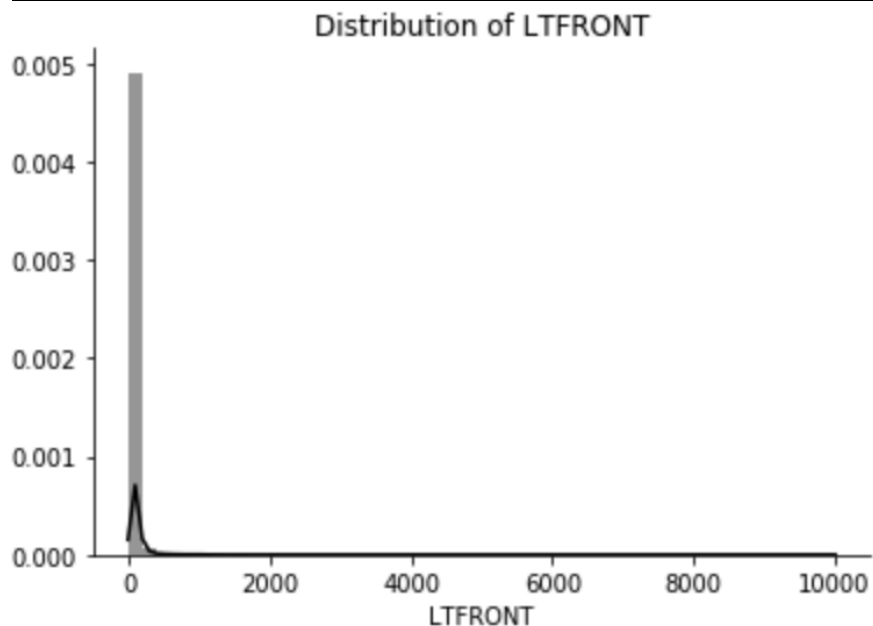


Figure 3.7 (a)

As we can see that there are outliers in the graph above, which make us cannot see the details on the left, so we try to plot the data without the outliers. (LTFRONT  $\leq$  250)

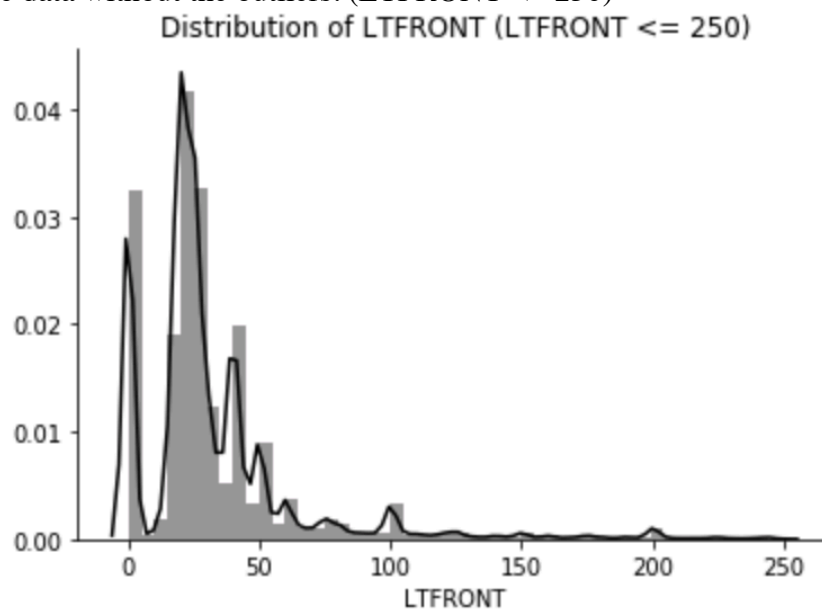


Figure 3.7 (b)

### 3.11 LTDEPTH

LTDEPTH stands for lot depth in feet.

Table 3.9

Unit	Max	Min	Mean	Std
Feet	0.00	9,999.00	88.86	76.40

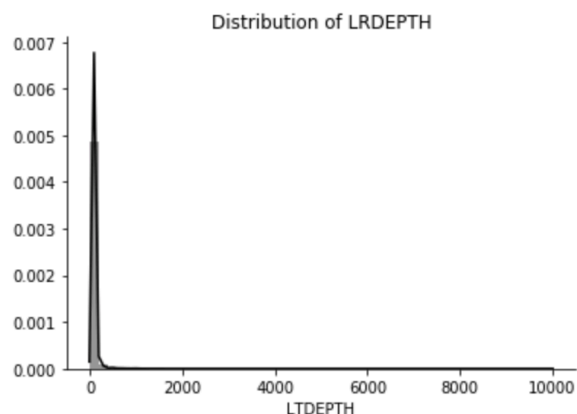


Figure 3.8 (a)

Also, because of the outliers, we cannot see details, so the plot below excludes them. (LTDEPTH <= 250)

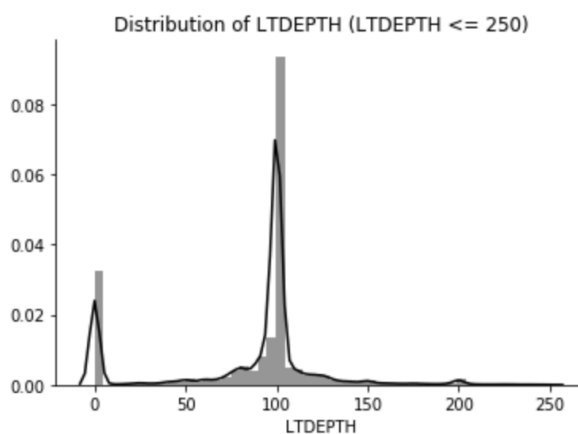


Figure 3.8 (b)

### 3.12 EXT

EXT stands for extension indicator. It contains NA values of 716,689

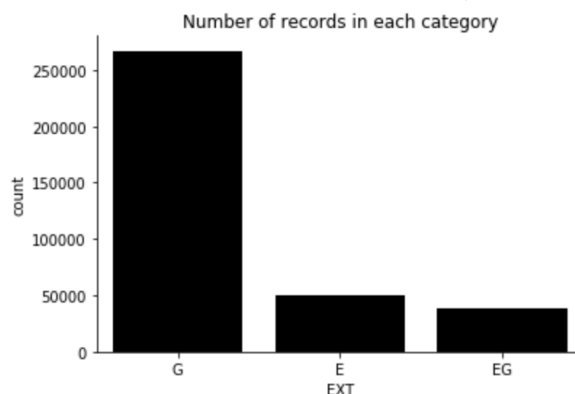


Figure 3.9

### 3.13 STORIES

STORIES stands for the number of stories (floors) for the building. It contains NA values of 56,264

Table 3.10

Unit	Max	Min	Mean	Std
Floor	119.00	1.00	5.01	8.37

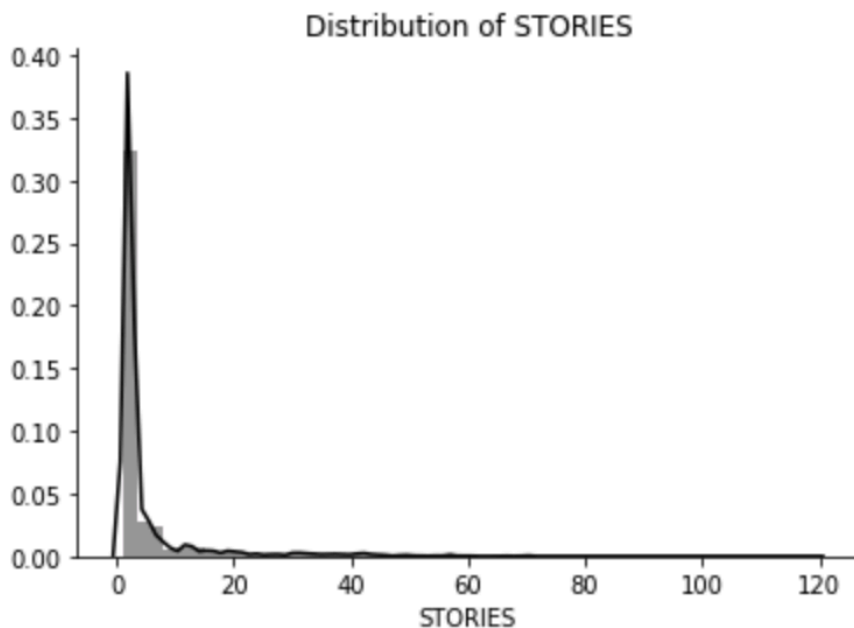


Figure 3.10 (a)

Without outliers (STORIES  $\leq 20$ ):

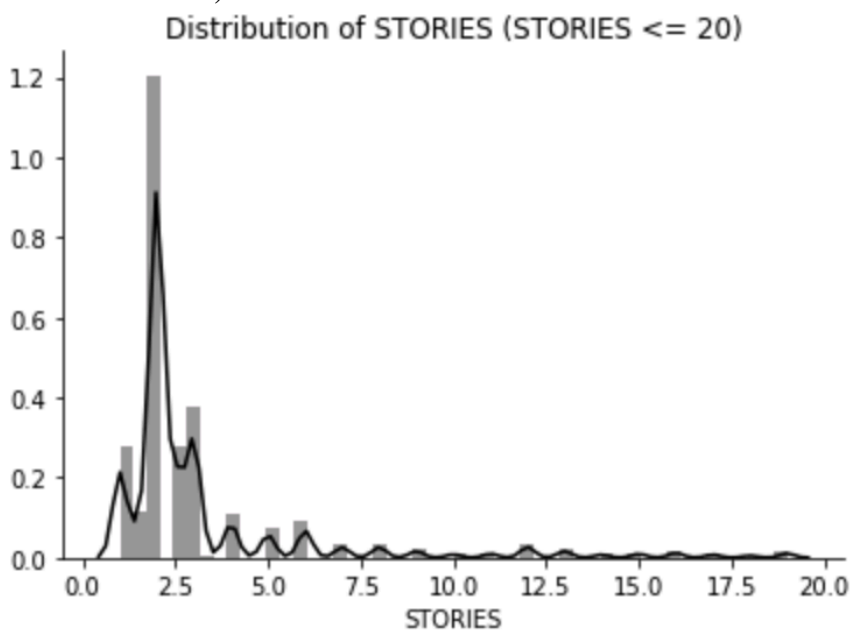


Figure 3.10 (b)

### 3.14 FULLVAL

FULLVAL stands for market value.

Table 3.11

Unit	Max	Min	Mean	Std
US Dollar	6,150,000,000.00	0.00	874,264.51	11,582,430.99

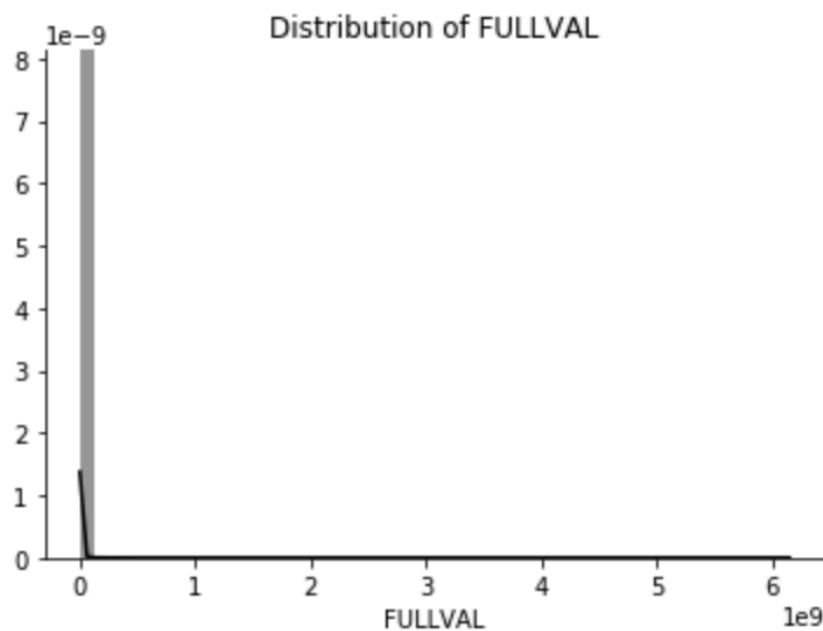


Figure 3.11 (a)

Without outliers (FULLVAL <= 5E+6):

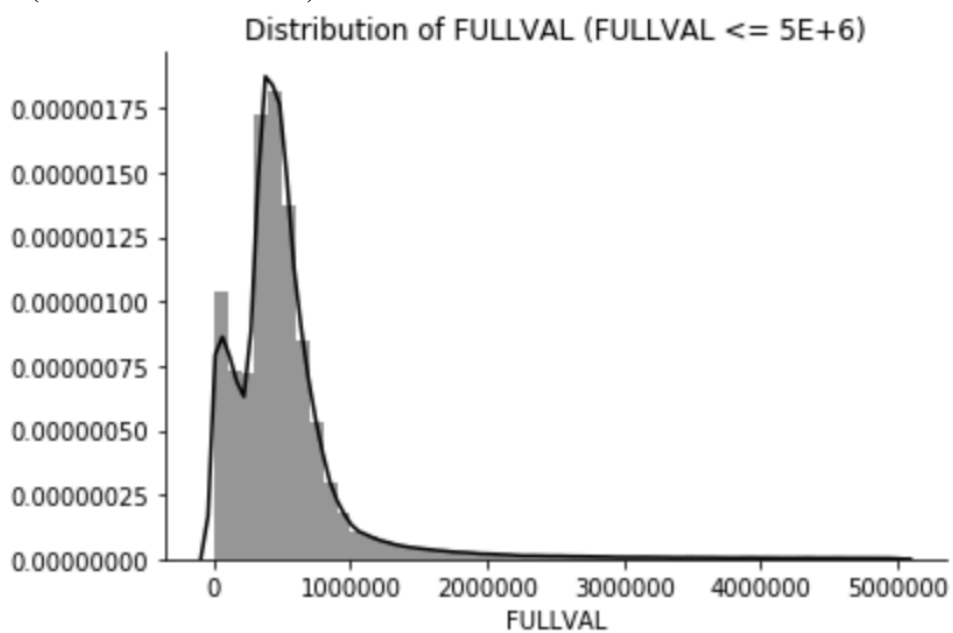


Figure 3.11 (b)



### 3.15 AVLAND

AVLAND stands for actual land value.

Table 3.12

Unit	Max	Min	Mean	Std
US Dollar	2,668,500,000.00	0.00	85,067.91	4,057,260.06

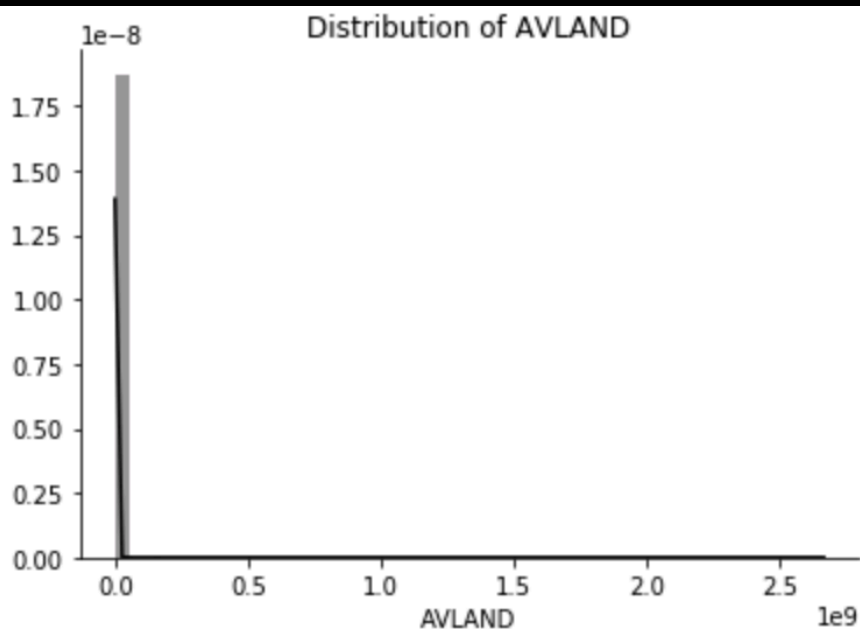


Figure 3.12 (a)

Without outliers (AVLAND  $\leq 1.5E+6$ ):

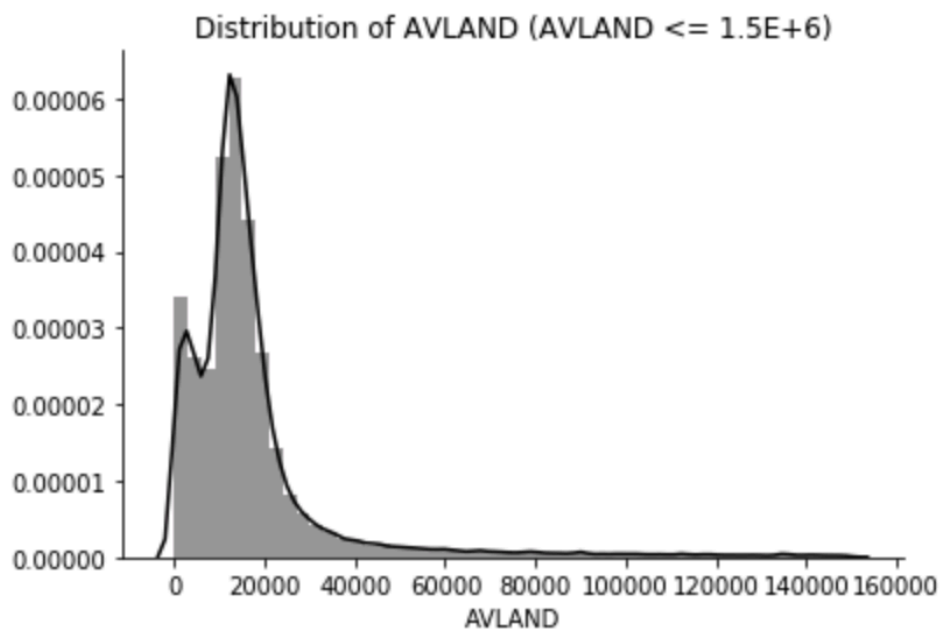


Figure 3.12 (b)

### 3.16 AVTOT

AVTOT stands for actual total value.

Table 3.13

Unit	Max	Min	Mean	Std
US Dollar	4,668,308,947.00	0.00	227,238.17	6,877,529.31

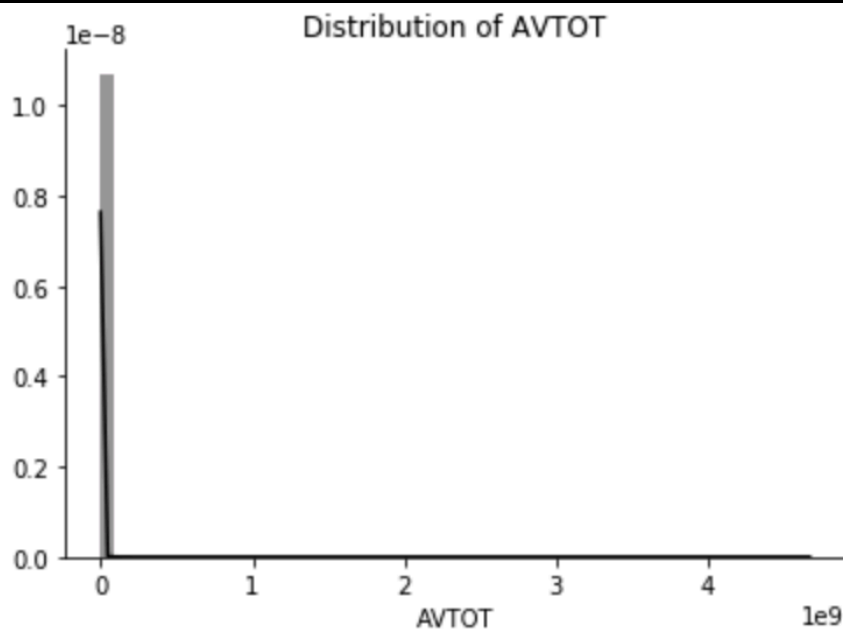


Figure 3.13 (a)

Without outliers (AVTOT  $\leq 1.5E+6$ ):

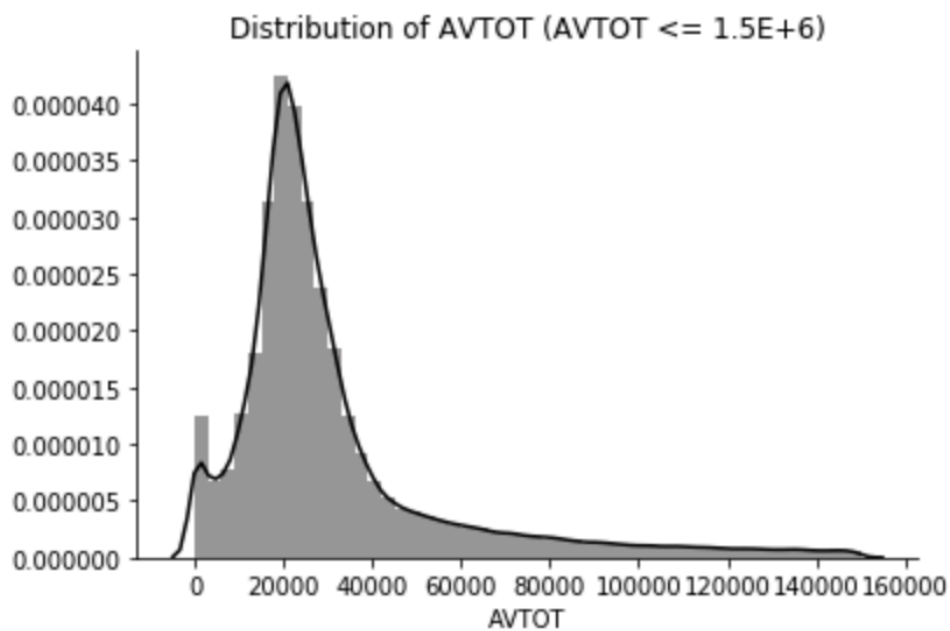


Figure 3.13 (b)

### 3.17 EXLAND

EXLAND stands for actual exempt land value.

Table 3.14

Unit	Max	Min	Mean	Std
US Dollar	2,668,500,000.00	0.00	36,423.89	3,981,575.79

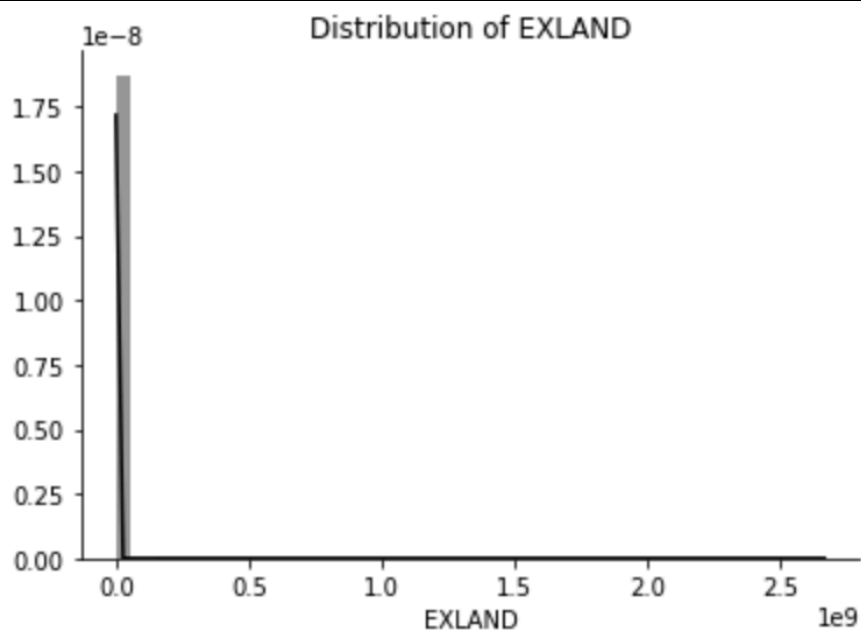


Figure 3.14 (a)

Without outliers (EXLAND  $\leq 5000$ ):

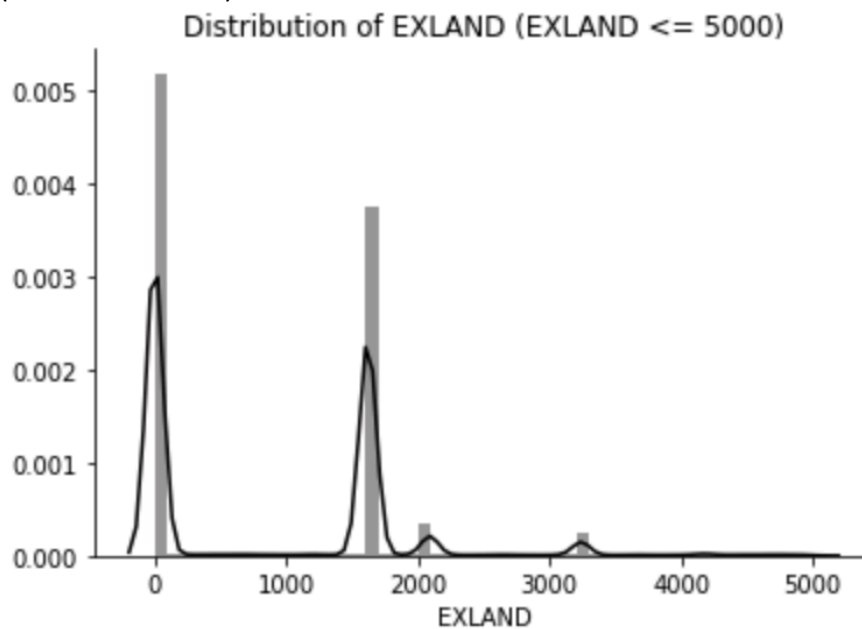


Figure 3.14 (b)

### 3.18 EXTOT

EXTOT stands for actual exempt land total.

Table 3.15

Unit	Max	Min	Mean	Std
US Dollar	4,668,308,947.00	0.00	91,186.98	6,508,402.82

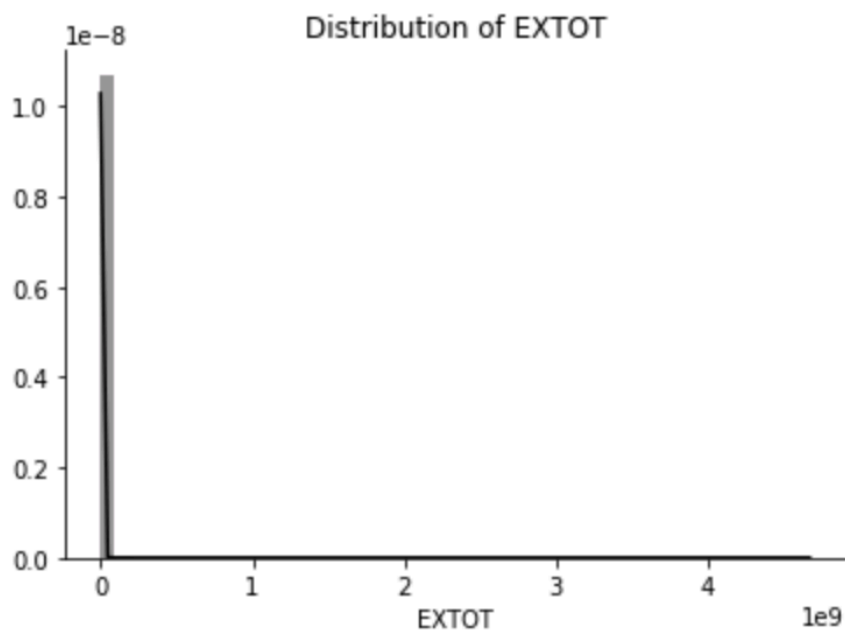


Figure 3.15 (a)

Without outliers (EXTOT  $\leq 5000$ ):

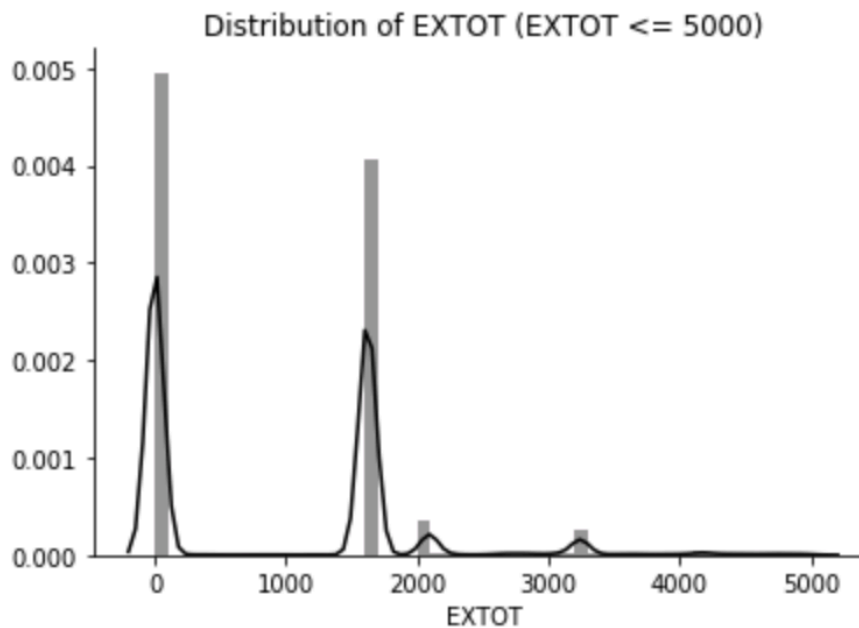


Figure 3.15 (b)

### 3.19 EXCD1

EXCD1 stands for exemption code 1, it contains NA values of 432,506

Table 3.16

EXCD1	count
1017	425348
1010	49756
1015	31323
5113	23858
1920	17594
5110	16834
5114	14984
...	...

### 3.20 STADDR

STADDR stands for street address, it contains NA values of 676

Table 3.17

STADDR	count
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
...	...

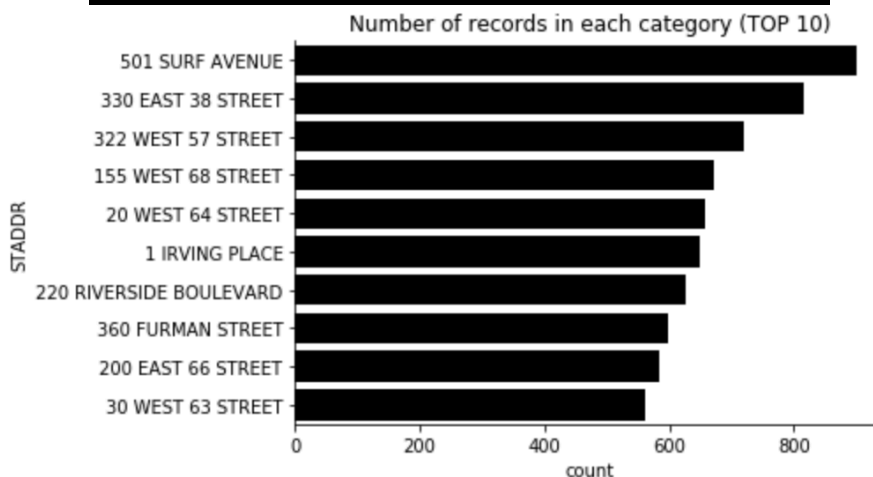


Figure 3.16

### 3.21 ZIP

ZIP stands for the postal zip code of the property. It contains NA values of 29,890

Table 3.18

ZIP	count
10314	24606
11234	20001
10312	18127
10462	16905
10306	16578
11236	15678
11385	14921
...	...

### 3.22 EXMPTCL

EXMPTCL stands for the exempt class. It contains NA values of 1,055,415

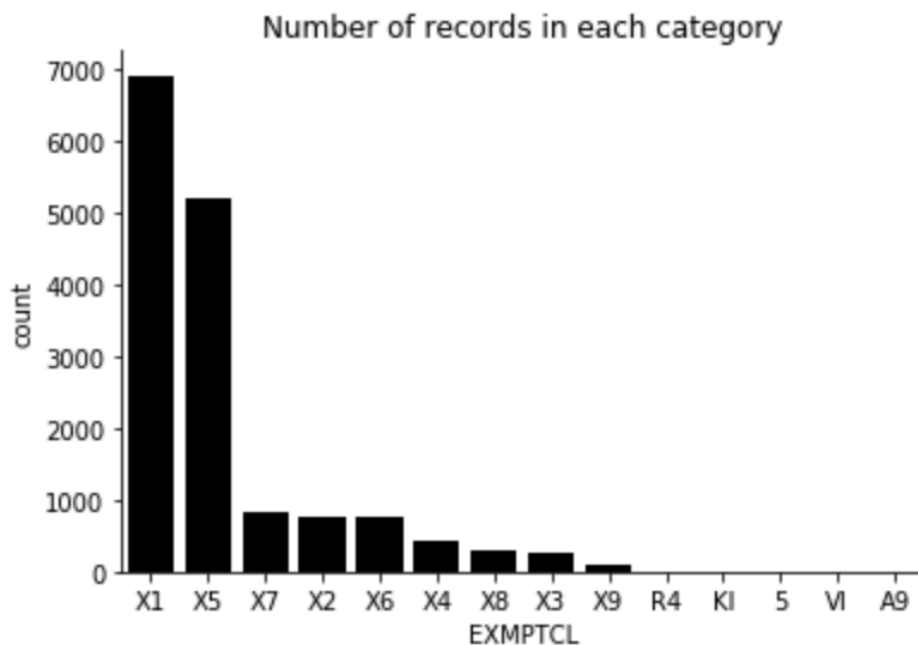


Figure 3.17

### 3.23 BLDFRONT

BLDFRONT stands for building frontage (width) in feet.

Table 3.19

Unit	Max	Min	Mean	Std
Feet	7,575.00	0.00	23.04	35.58

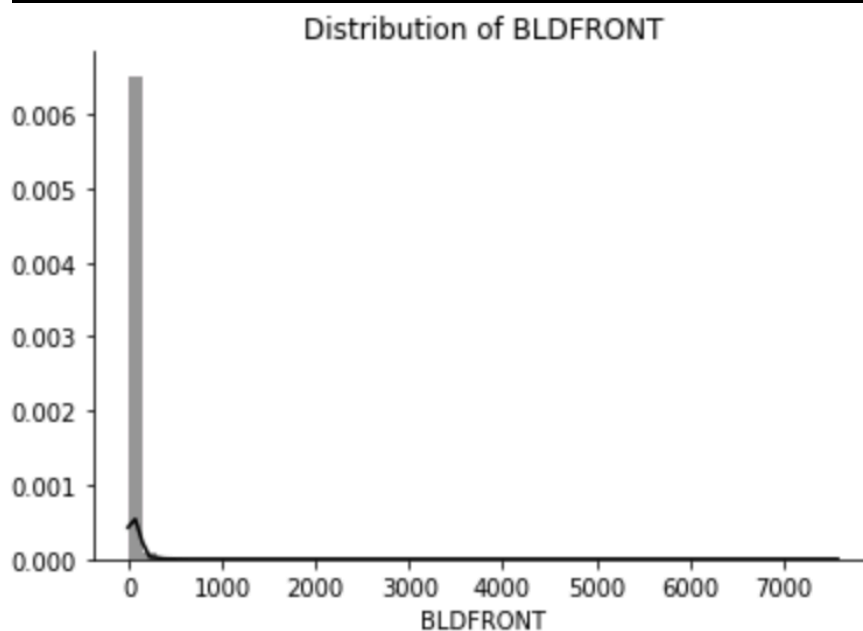


Figure 3.18 (a)

Without outliers (BLDFRONT  $\leq 200$ ):

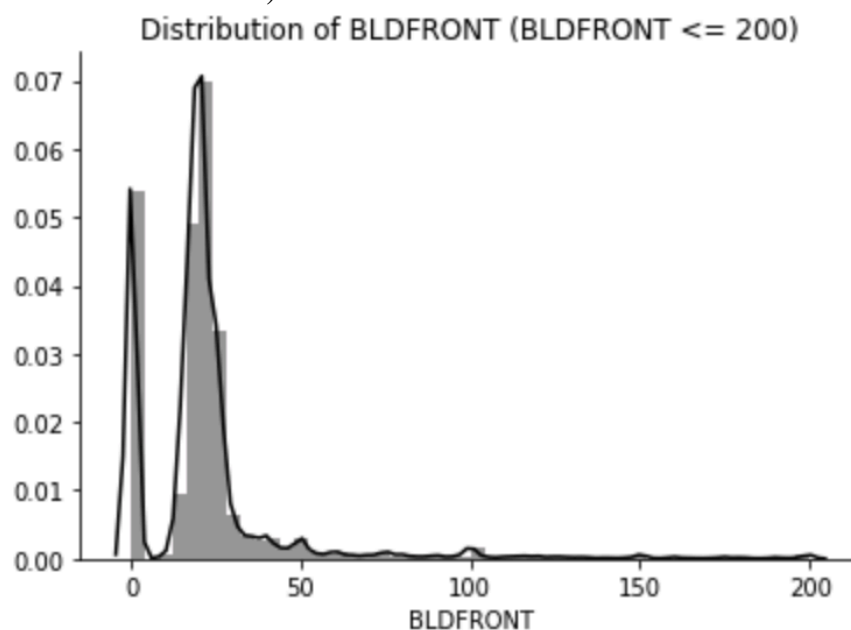


Figure 3.18 (b)

### 3.24 BLDDEPTH

BLDDEPTH stands for building depth in feet.

Table 3.20

Unit	Max	Min	Mean	Std
Feet	9,393.00	0.00	39.92	42.71

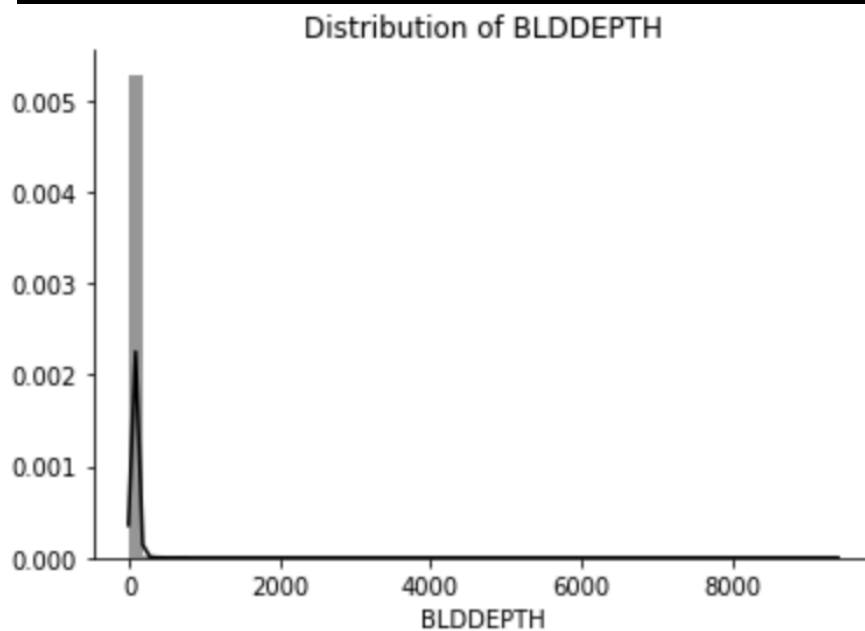


Figure 3.19 (a)

Without outliers (BLDDEPTH  $\leq 200$ ):

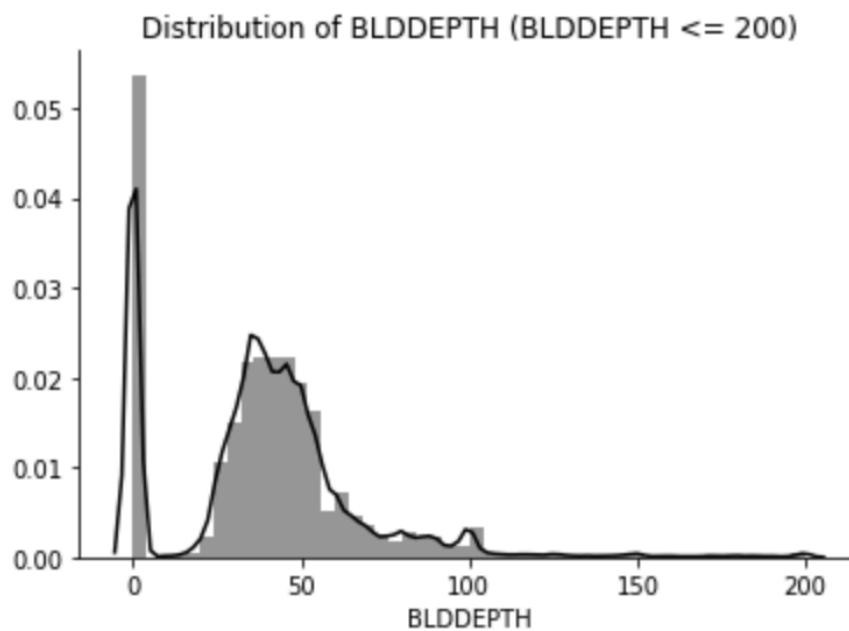


Figure 3.19 (b)



### 3.25 AVLAND2

AVLAND2 stands for transitional land value, it contains NA values of 788,268

Table 3.21

Unit	Max	Min	Mean	Std
US Dollar	2,371,005,000.00	0.00	246,235.72	6,178,962.56

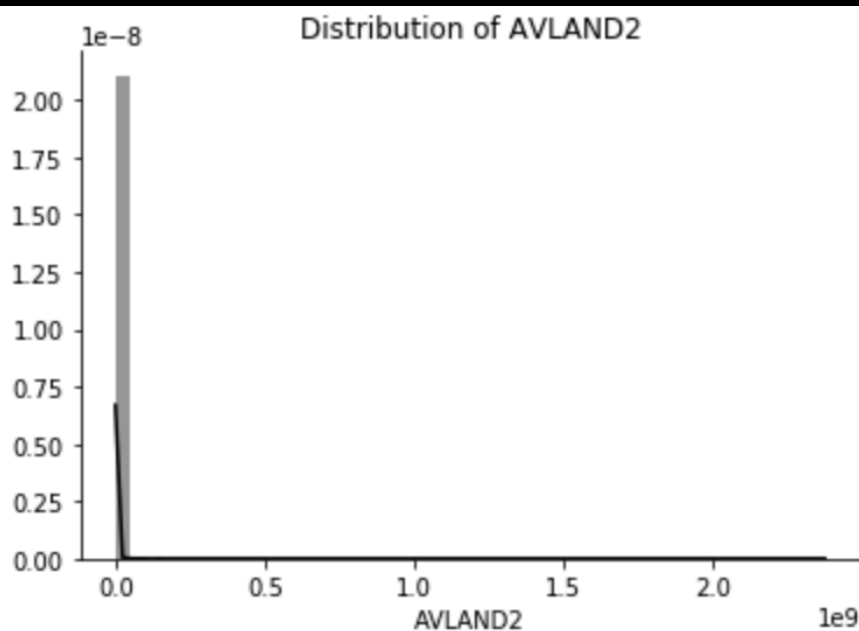


Figure 3.20 (a)

Without outliers (AVLAND2  $\leq 3E+5$ ):

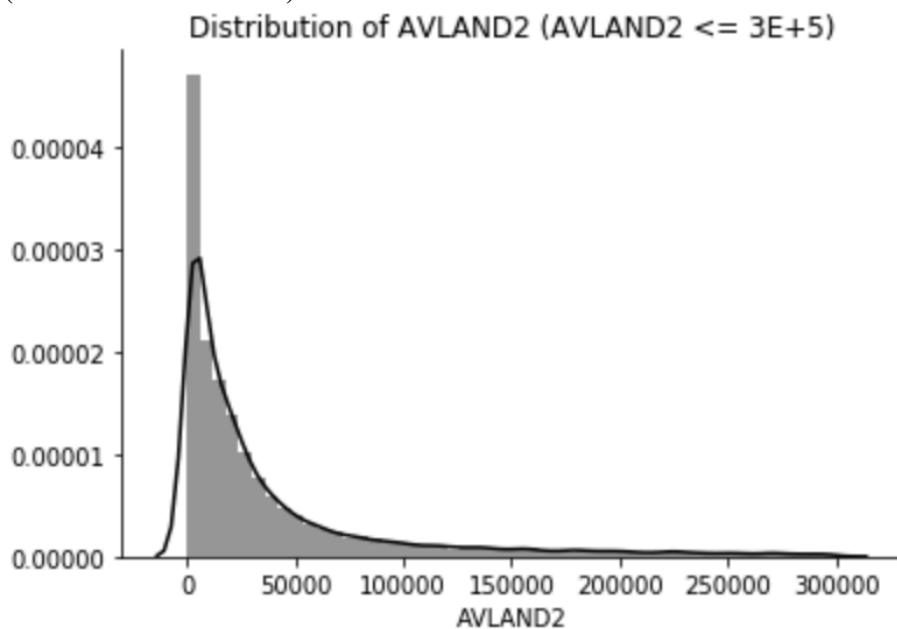


Figure 3.20 (b)

### 3.26 AVTOT2

AVTOT2 stands for transitional total value, it contains NA values of 788,262.

Table 3.22

Unit	Max	Min	Mean	Std
US Dollar	4,501,180,002.00	0.00	713,911.44	11,652,528.95

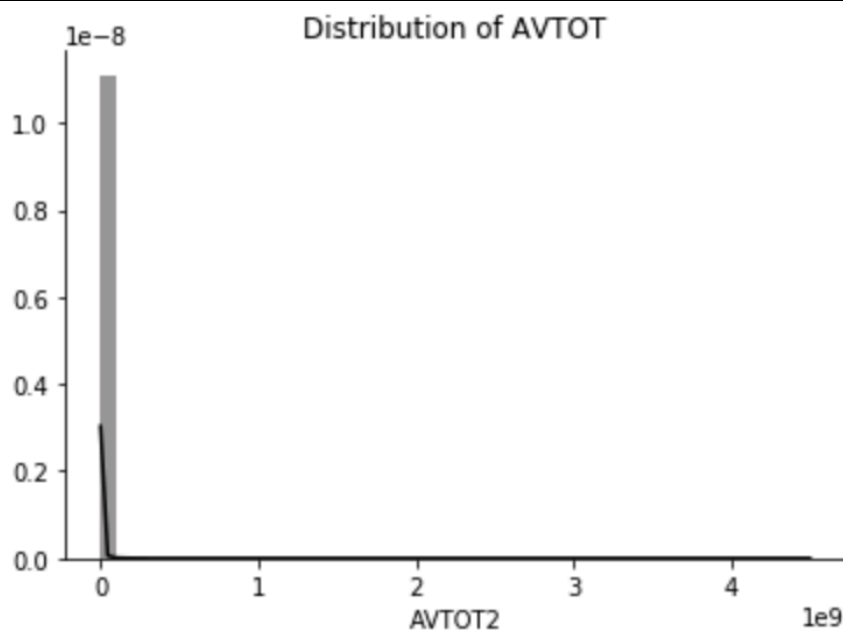


Figure 3.21 (a)

Without outliers (AVTOT2  $\leq 8E+5$ ):

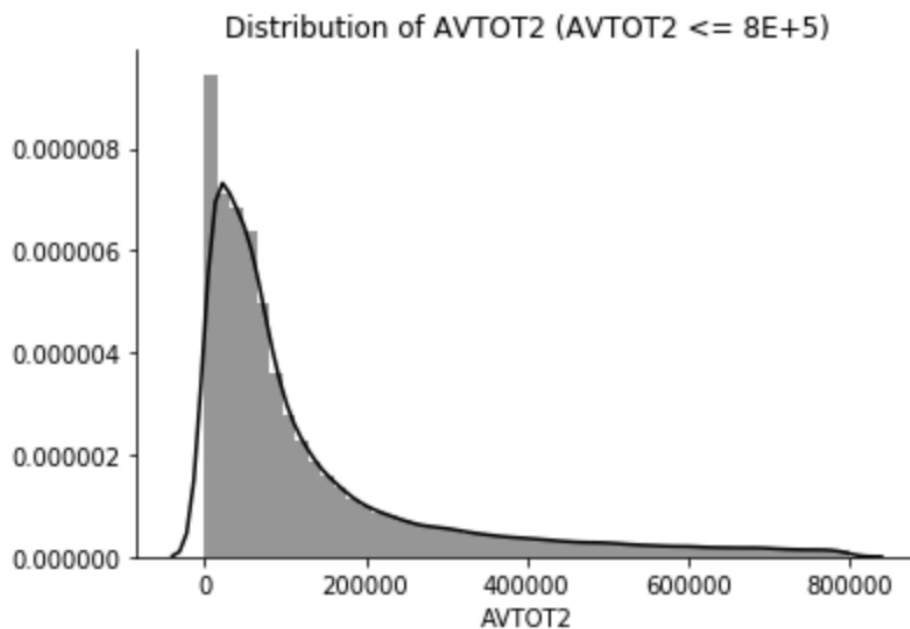


Figure 3.21 (b)

### 3.27 EXLAND2

EXLAND2 stands for transitional exempt land value, it contains NA values of 983,545.

Table 3.23

Unit	Max	Min	Mean	Std
US Dollar	2,371,005,000.00	0.00	351,235.68	10,802,212.67

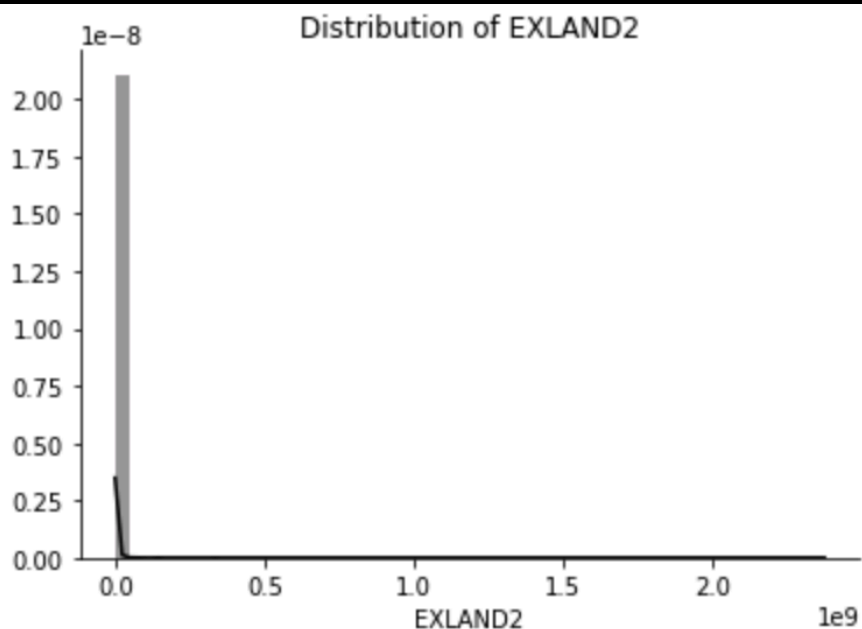


Figure 3.22 (a)

Without outliers (EXLAND2  $\leq$  30,000):

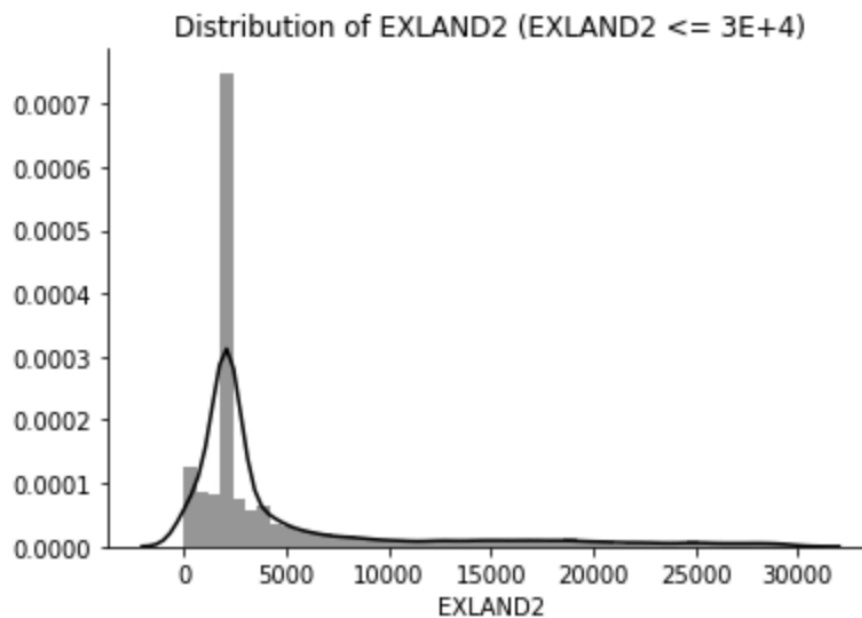


Figure 3.22 (b)

### 3.28 EXTOT2

EXTOT2 stands for transitional exempt land total, it contains NA values of 940,166.

Table 3.24

Unit	Max	Min	Mean	Std
US Dollar	4,501,180,002.00	0.00	656,768.28	16,072,510.17

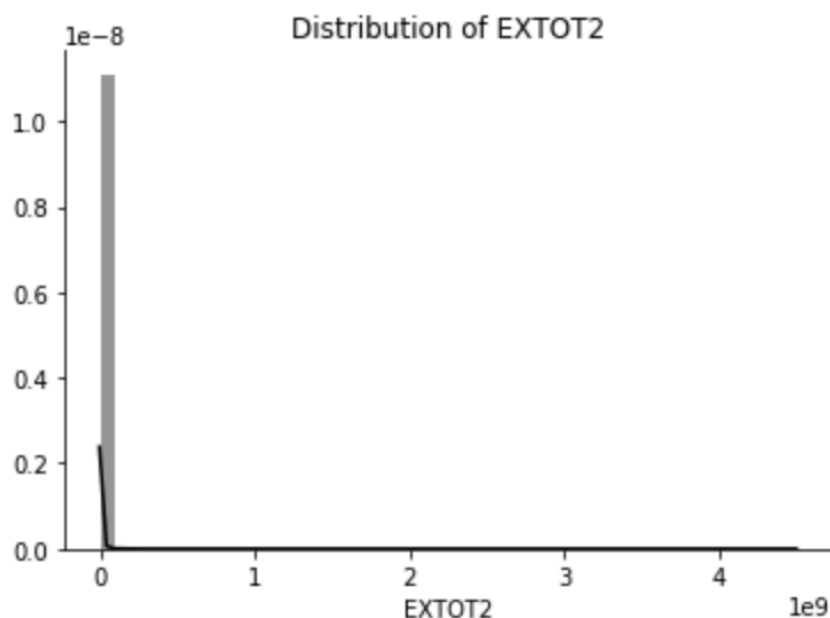


Figure 3.23 (a)

Without outliers (EXTOT2  $\leq 1E+5$ ):

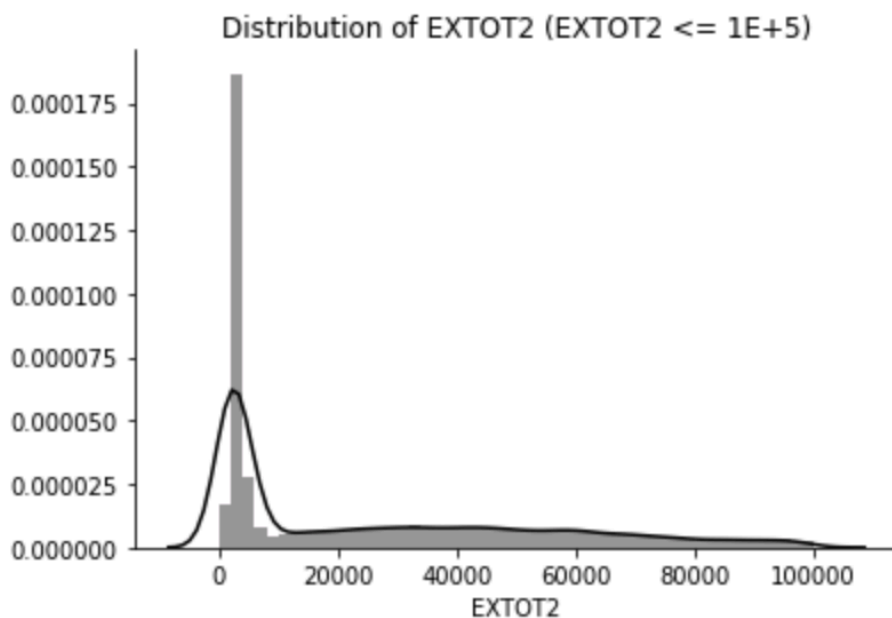


Figure 3.23 (b)

### 3.29 EXCD2

EXCD2 stands for exemption code 2, it contains NA values of 978,046

Table 3.25

EXCD2	count
1017	65777
1015	12337
5112	6867
1019	3178
1920	2961
1200	881
1101	494
...	...

### 3.30 PERIOD

Period stands for assessment period when file was created

### 3.31 YEAR

YEAR stands assessment year. In this dataset, we only have 2010/11 for this field

### 3.32 VALTYPE

In this dataset, we only have VALTYPE of “AC-TR”