

9. Appendix

1. Description of the data:

The dataset is the property valuation and assessment data of New York City. Data represent NYC properties assessments for purpose to calculate Property Tax, Grant eligible properties Exemptions and/or Abatements. It is provided by the department of Finance on NYC Open Data and covers only in the time period of November 2010. There are 32 columns and 1,070,994 records.

2. Summary Tables

2.1 Numerical Value

Field Name	Field Type	# unique values	# of records with value	% populated	# of records with value zero	Mean	Standard Deviation	Min	Max
LTFRONT	numerical	1297	1070994	100.00	169108	36.64	74.03	0	9999
LTDEPTH	numerical	1370	1070994	100.00	170128	88.86	76.40	0	9999
FULLVAL	numerical	109324	1070994	100.00	13007	874264.51	11582430.99	0	6150000000
AVLAND	numerical	70921	1070994	100.00	13009	85067.92	4057260.06	0	2668500000
AVTOT	numerical	112914	1070994	100.00	13007	227238.17	6877529.31	0	4668308947
EXLAND	numerical	33419	1070994	100.00	491699	36423.89	3981575.79	0	2668500000
EXTOT	numerical	64255	1070994	100.00	432572	91186.98	6508402.82	0	4668308947
BLDFRONT	numerical	612	1070994	100.00	228815	23.04	35.58	0	7575
BLDDEPTH	numerical	621	1070994	100.00	228853	39.92	42.71	0	9393
STORIES	numerical	112	1014730	94.75	0	5.01	8.37	1	119
AVLAND2	numerical	58592	282726	26.40	0	246235.72	6178962.56	3	2371005000
AVTOT2	numerical	111361	282732	26.40	0	713911.44	11652528.95	3	4501180002
EXLAND2	numerical	22196	87449	8.17	0	351235.68	10802212.67	1	2371005000
EXTOT2	numerical	48349	130828	12.22	0	656768.28	16072510.17	7	4501180002

2.2 Categorical Value

Field Name	Field Type	# unique values	# of records with value	% populated	most common field value
RECORD	categorical	1070994	1070994	100.00	n/a
BBLE	categorical	1070994	1070994	100.00	n/a
B	categorical	5	1070994	100.00	4
BLOCK	categorical	13984	1070994	100.00	3944
LOT	categorical	6366	1070994	100.00	1
EASEMENT	categorical	13	4636	0.43	E
OWNER	categorical	863347	1039249	97.04	PARKCHESTER PRESERVAT
BLDGCL	categorical	200	1070994	100.00	R4
TAXCLASS	categorical	11	1070994	100.00	1
EXT	categorical	4	354305	33.08	G
EXCD1	categorical	130	638488	59.62	1017
STADDR	categorical	839281	1070318	99.94	501 SURF AVENUE
ZIP	categorical	197	1041104	97.21	10314
EXMPTCL	categorical	15	15579	1.45	X1
EXCD2	categorical	61	92948	8.68	1017
PERIOD	categorical	1	1070994	100.00	FINAL
YEAR	date/time	1	1070994	100.00	2010/11
VALTYPE	categorical	1	1070994	100.00	AC-TR

3. Field Description

Field 1

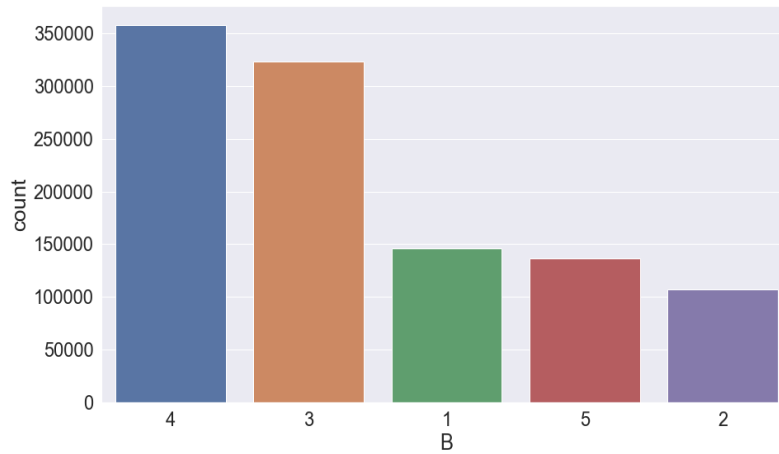
Name: RECORD

Description: It is a unique integer label for each record.

Field 2

Name: BBLE

Description: Concatenation of AV_BORO, AV_BLOCK, AV_LOT, AV_EASEMENT.
It's a unique combination for each property.

Field 3**Name:** B**Description:** Borough codes, from 1-5.**Field 4****Name:** BLOCK**Description:** Borough codes, from 1-5.**Top 10 common values**

BLOCK	COUNT
3944	3888
16	3786
3943	3424
3938	2794
1171	2535
3937	2275
1833	1774
2450	1651
1047	1480
7279	1302

Field 5**Name:** LOT**Description:** The lot of the property.**Top 10 common values**

LOT	COUNT
1	24367
20	12294
15	12171
12	12143
14	12074
16	12042
17	11982
18	11979
25	11949
21	11840

Field 6**Name:** EASEMENT**Description:** The right to cross/use the property land for special use.**Top 10 common values**

EASEMENT	COUNT
E	4148
F	296
G	102
H	33
N	19
I	16
J	8
K	5
L	3
P	3

Field 7**Name: OWNER****Description:** The owner of the property.**Top 10 common values**

OWNER	COUNT
PARKCHESTER PRESERVAT	6020
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

Field 8**Name: BLDGCL****Description:** The building class.**Top 10 common values**

BLDGCL	COUNT
R4	139879
A1	123369
A5	96984
B1	84208
B2	77598
C0	73111
B3	59240
A2	51130
A9	26177
B9	26133

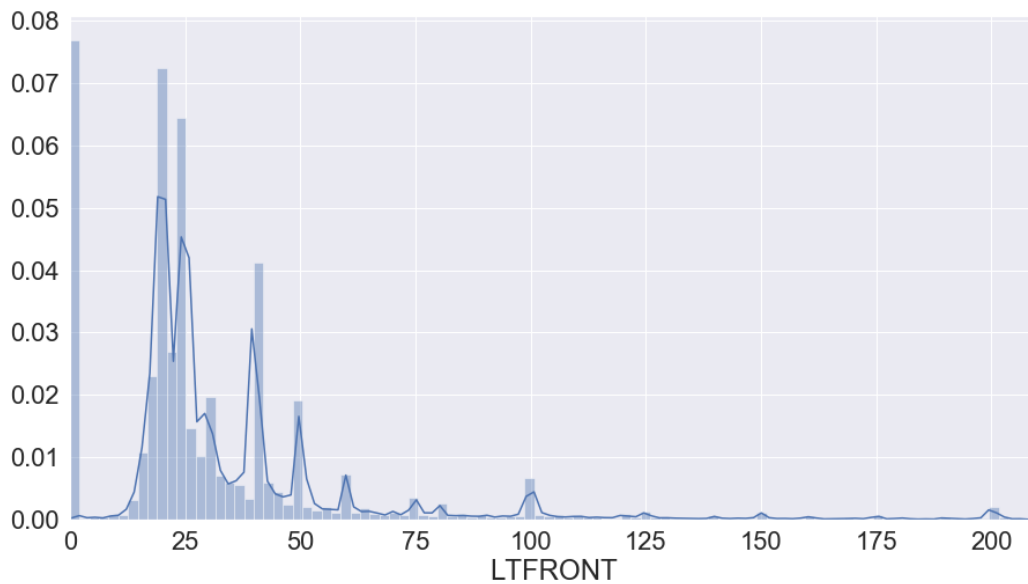
Field 9**Name:** TAXCLASS**Description:** The tax class.**Top 10 common values**

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

Field 10**Name:** LTFRONT**Description:** Lot width in feet.

Deleted outliers that have LTFRONT>210.

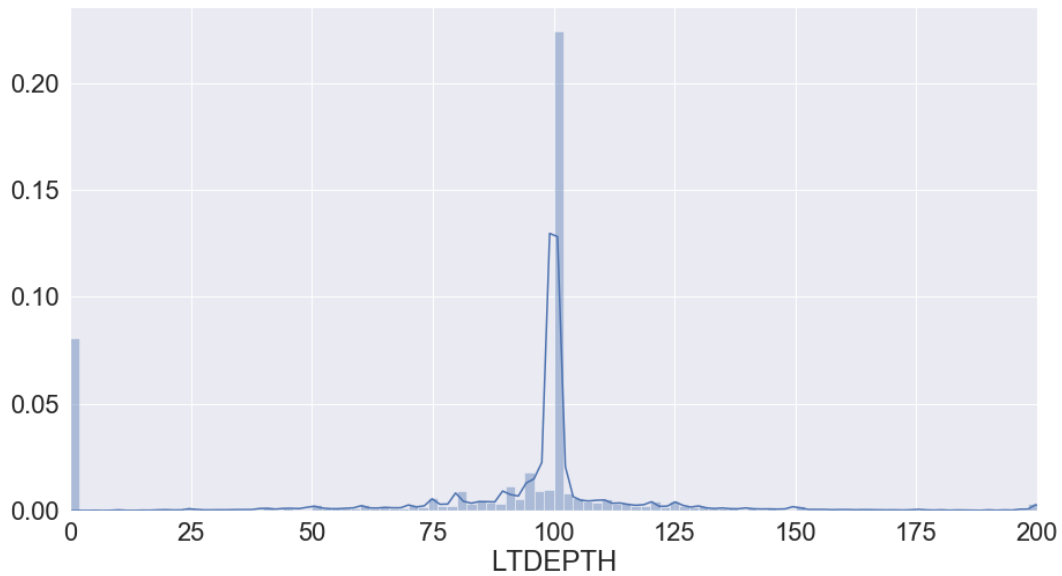
Data in the histogram is 98.67% populated among the records with values.



Field 11**Name: LTDEPTH****Description:** Lot depth in feet.

Deleted outliers that have LTFRONT>200.

Data in the histogram is 97.62% populated among the records with values.

**Field 12****Name: EXT****Description:** Extension Indicator**Top common values**

EXT	COUNT
G	266970
E	49442
EG	37893

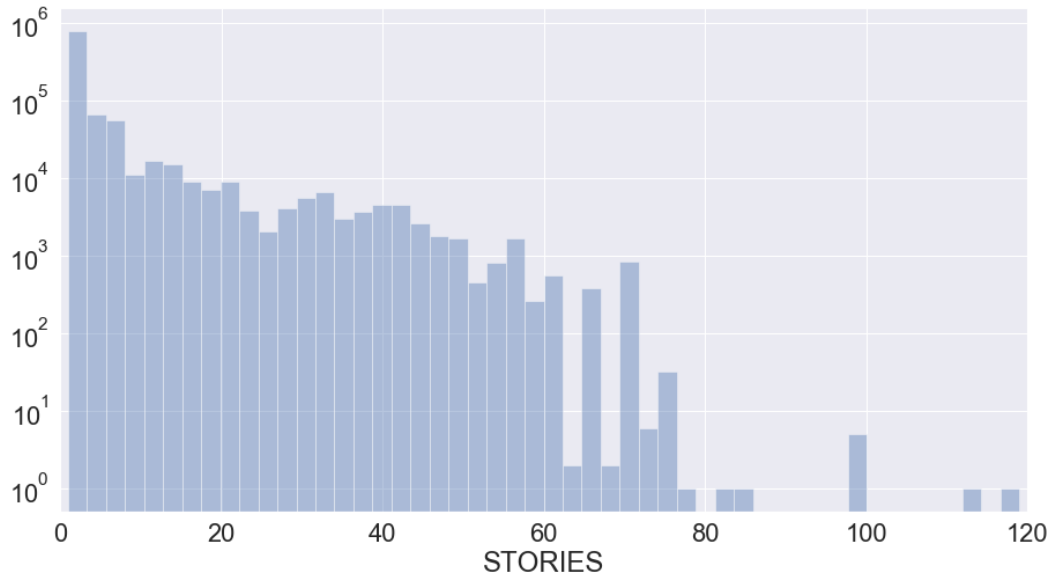
Field 13

Name: STORIES

Description: Number of Stories in Building

Didn't delete any outlier

Data in the histogram is 100% populated among the records with values.



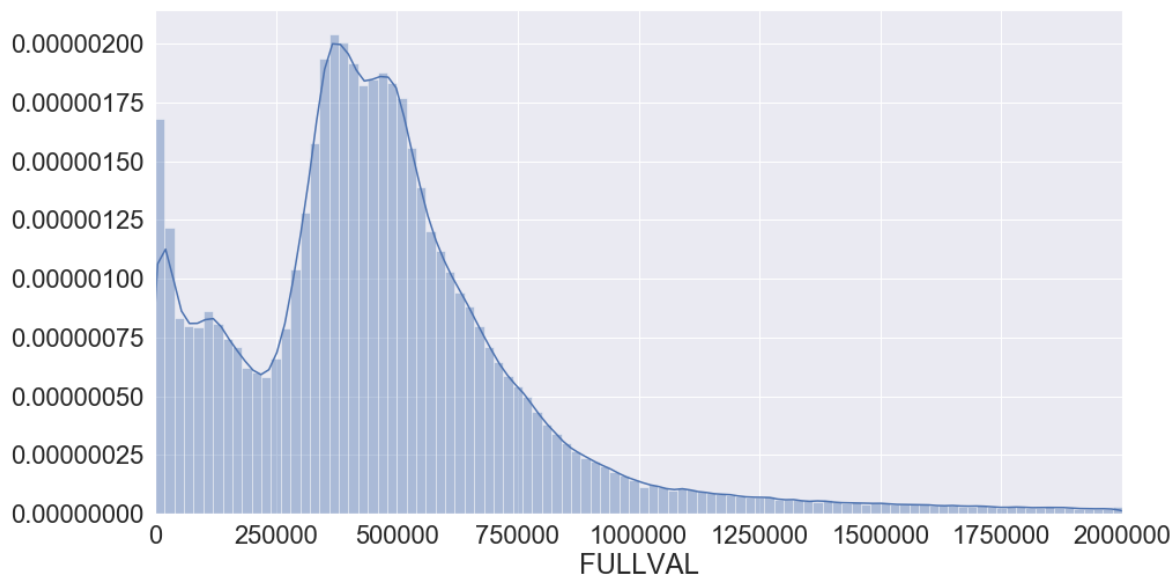
Field 14

Name: FULLVAL

Description: Market Value

Deleted outliers that have fullval > 2000000

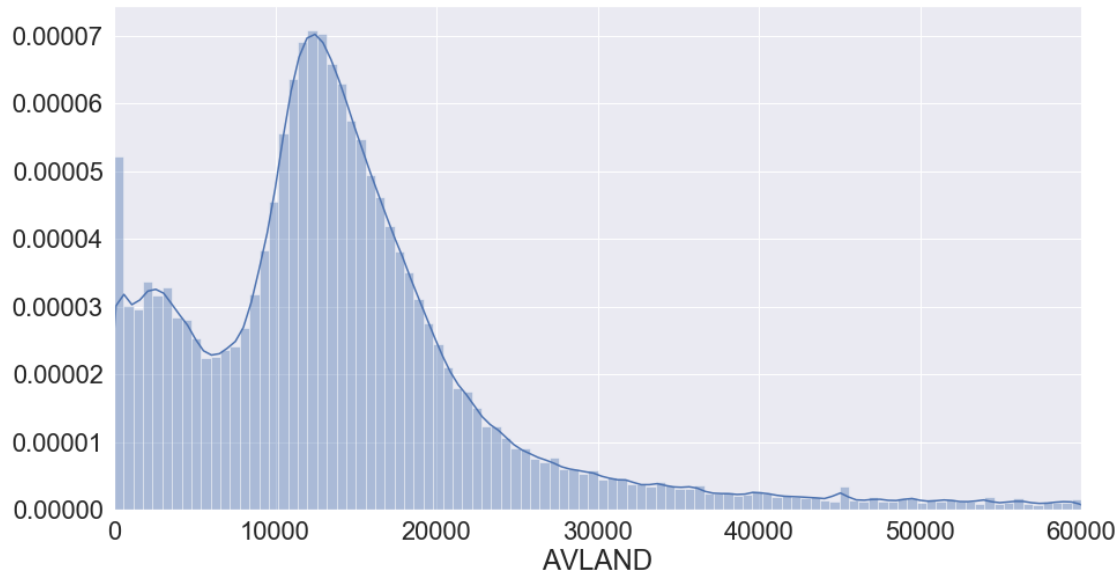
Data in the histogram is 96.29% populated among the records with values.



Field 15**Name: AVLAND****Description: Actual Land Value**

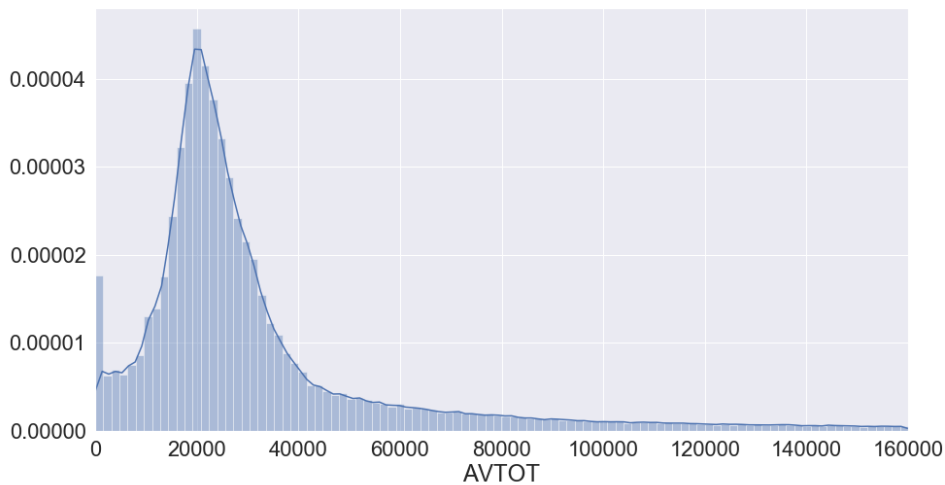
Deleted outliers that have AVLAND>60000

Data in the histogram is 91.61% populated among the records with values.

**Field 16****Name: AVTOT****Description: Actual Total Value**

Deleted outliers that have AVTOT>160000.

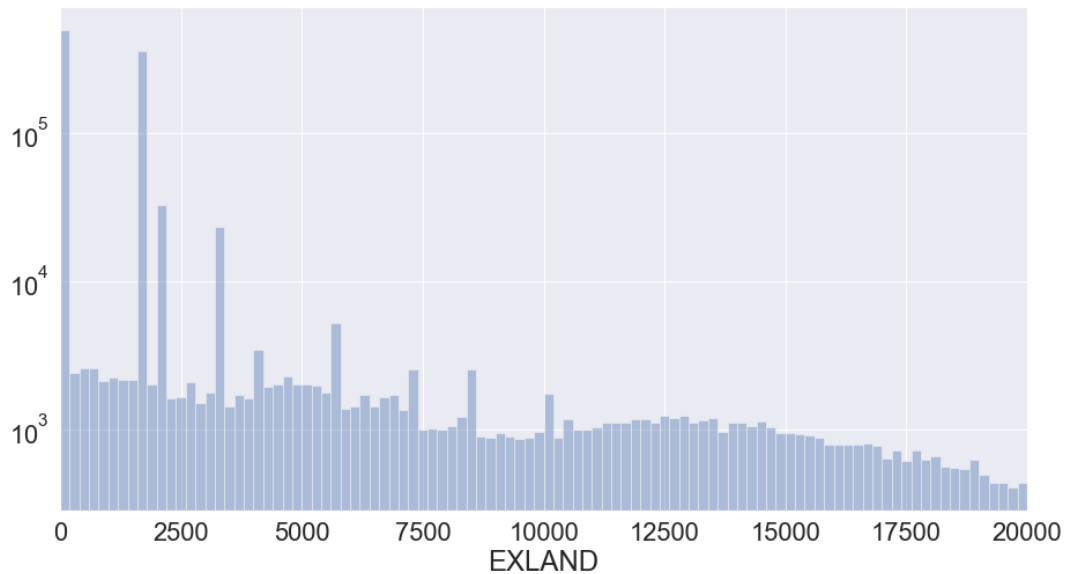
Data in the histogram is 90.05% populated among the records with values.



Field 17**Name: EXLAND****Description:** Actual Exempt Land Value

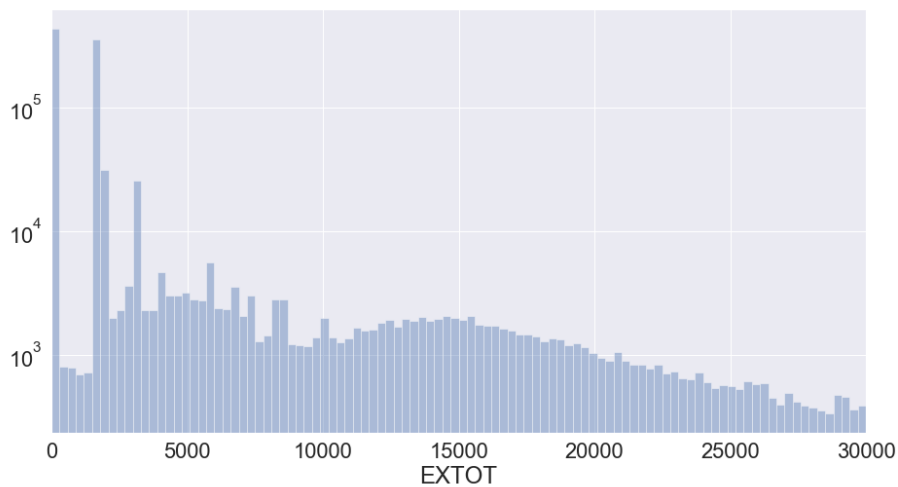
Deleted outliers that have EXLAND>20000

Data in the histogram is 96.82% populated among the records with values.

**Field 18****Name: EXTOT****Description:** Actual Exempt Land Total

Deleted outliers that have EXTOT>30000

Data in the histogram is 97.63% populated among the records with values.



Field 19**Name: EXCD1****Description:** Exemption Code 1**Top 10 common values**

EXCD1	COUNT
1017	425348
1010	49756
1015	31323
5113	23858
1920	17594
5110	16834
5114	14984
5111	10609
1021	6613
1986	4231

Field 20**Name: STADDR****Description:** Street Address of the property**Top 10 common values**

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
360 FURMAN STREET	599
200 EAST 66 STREET	585
30 WEST 63 STREET	562

Field 21**Name:** ZIP**Description:** zip code of the property**Top 10 common values**

ZIP	COUNT
10314	24606
11234	20001
10312	18127
10462	16905
10306	16578
11236	15678
11385	14921
11229	12793
11211	12710
11207	12293

Field 22**Name:** EXMPTCL**Description:** Exemption Class of the property**Top 10 common values**

EXMPTCL	COUNT
X1	6912
X5	5208
X7	820
X2	770
X6	764
X4	441
X8	292
X3	259
X9	108
R4	1

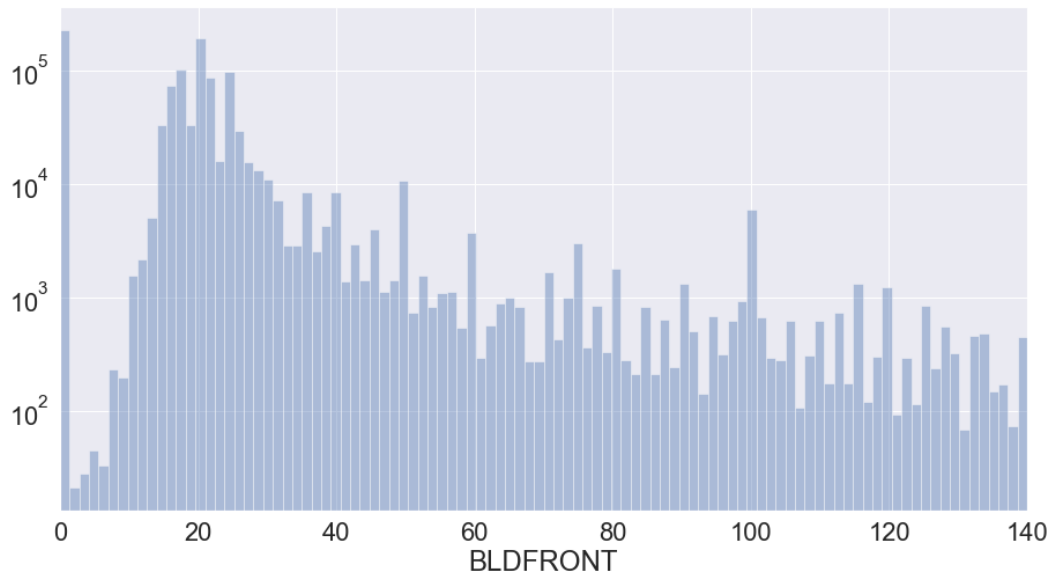
Field 23

Name: BLDFRONT

Description: building width

Deleted outliers that have $\text{BLDFRONT} > 140$

Data in the histogram is 98.39% populated among the records with values.



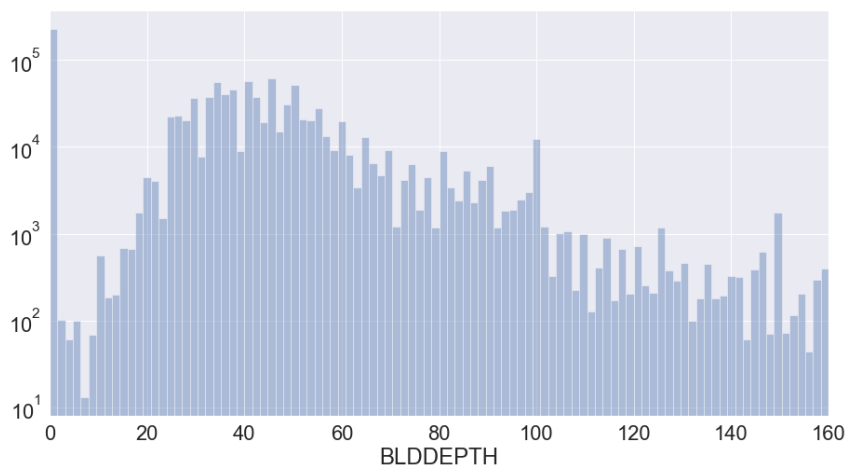
Field 24

Name: BLDDEPTH

Description: building depth

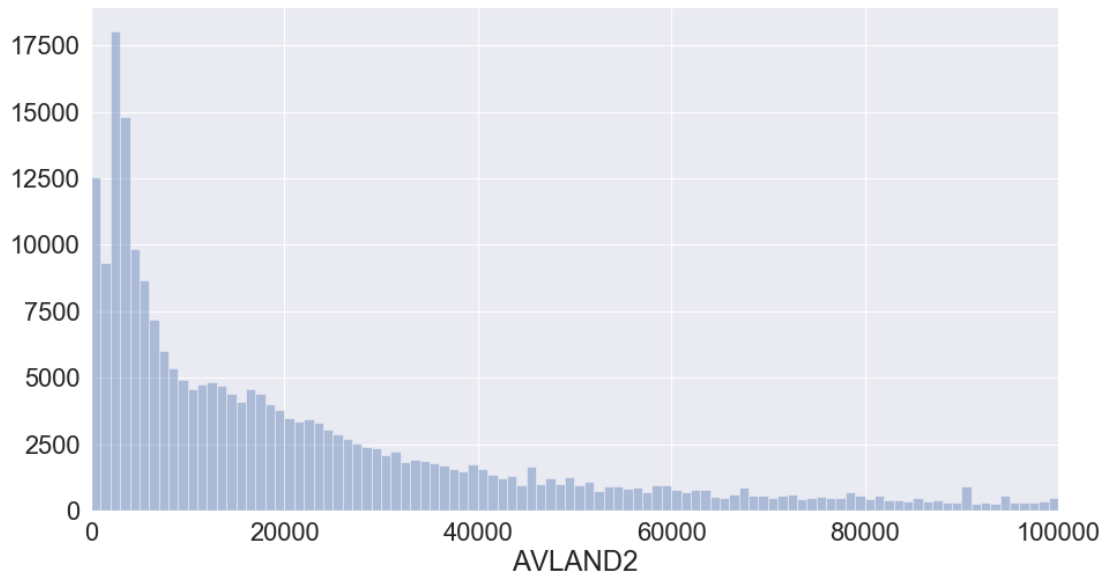
Deleted outliers that have $\text{BLDDEPTH} > 160$

Data in the histogram is 98.91% populated among the records with values.

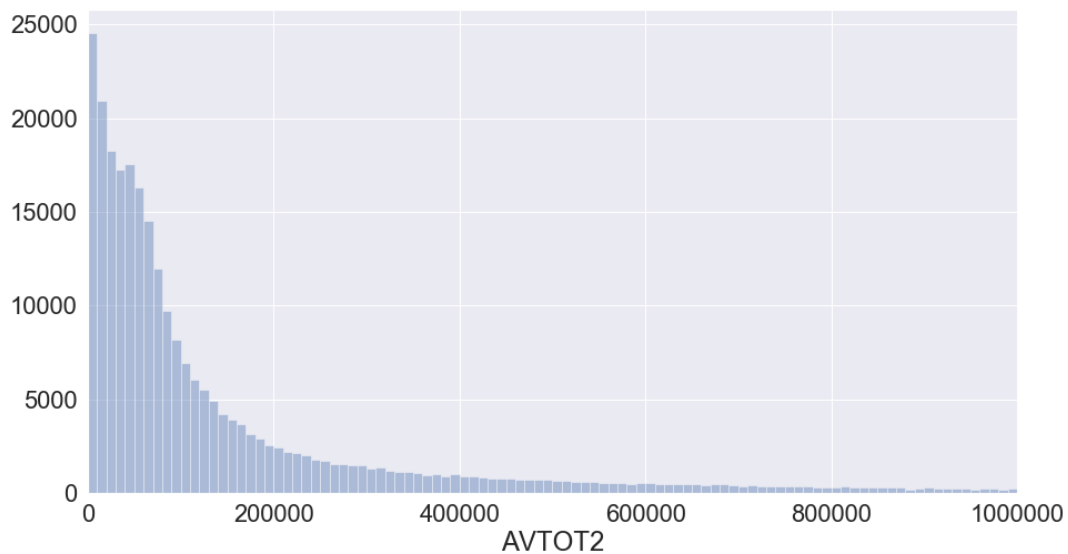


Field 25**Name: AVLAND2****Description:** Transitional Land ValueDeleted outliers that have $AVLAND2 > 100000$

Data in the histogram is 81.33% populated among the records with values.

**Field 26****Name: AVTOT2****Description:** Transitional Total ValueDeleted outliers that have $AVTOT2 > 1000000$

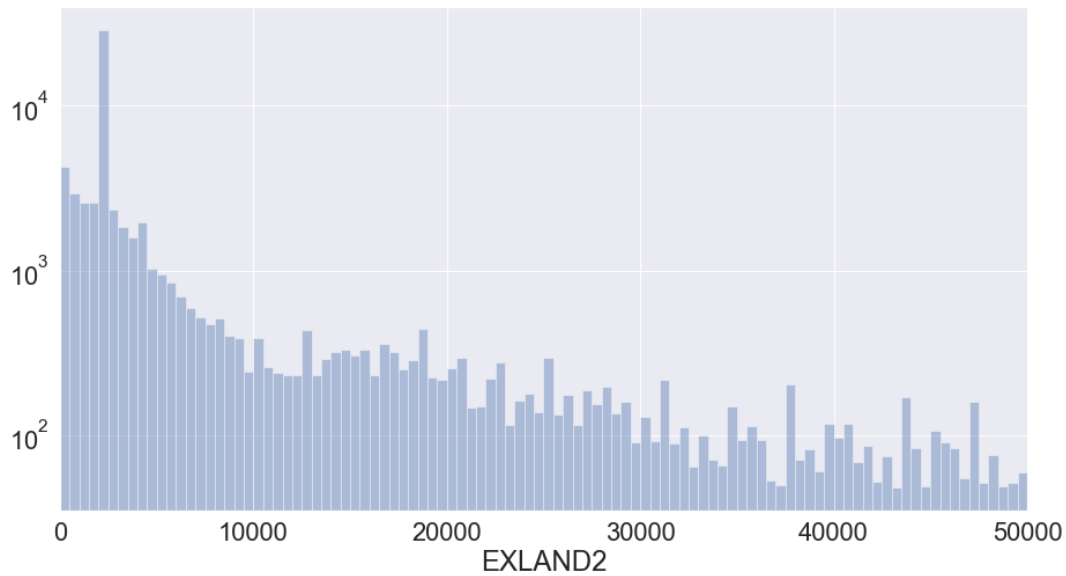
Data in the histogram is 91.61% populated among the records with values.



Field 27**Name: EXLAND2****Description:** Transitional Exemption Land Value

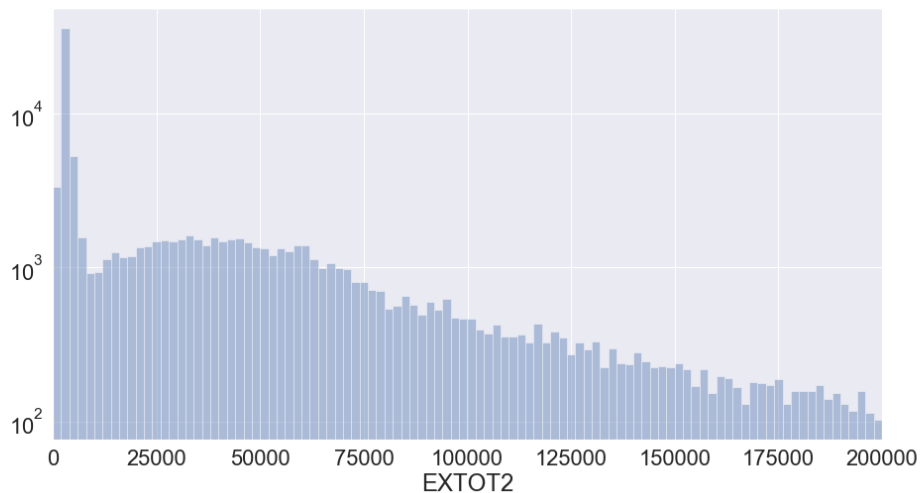
Deleted outliers that have EXLAND2>50000

Data in the histogram is 78.55% populated among the records with values.

**Field 28****Name: EXTOT2****Description:** Transitional Exemption Land Total

Deleted outliers that have EXTOT2>200000

Data in the histogram is 83.14% populated among the records with values.



Field 29**Name:** EXCD2**Description:** Exemption Code 2**Top 10 common values**

EXCD2	COUNT
1017	65777
1015	12337
5112	6867
1019	3178
1920	2961
1200	881
1101	494
5129	227
1986	35
1022	31

Field 30**Name:** PERIOD**Description:** Assessment Period:

Only one unique record: Final

Field 31**Name:** YEAR**Description:** Assessment year:

Only one unique record: 2010/11 (November 2010)

Field 32**Name:** VALTYPE**Description:** Value type:

Only one unique record: AC-TR