Data Quality Report
for New York Property Valuation and Assessment Data

I. Dataset Summary

New York Property Valuation and Assessment Data was **provided by NYC Department of Finance (DOF) on September 2, 2011**. This Dataset has **32 fields** and **1070994 records**. Dataset represents 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, like Property Assessors, Property Exemption specialists, ACRIS reporting, Department of Building reporting, etc.

II. Field Summary

Numerical Fields

In this dataset there are 14 numerical fields. Table 1 represents some basic statistic summary for each numerical field. Fields that marked in red are fields of low percentage populated.

Field Name	Count	Unique Values	Value with Zero	% Populated	Mean	Max	Min	SD
LTFRONT	1,070,994	1,297	169,108	100.00%	3.66E+01	1.00E+04	0.00E+00	7.40E+01
LTDEPTH	1,070,994	1,370	170,128	100.00%	8.89E+01	1.00E+04	0.00E+00	7.64E+01
STORIES	1,014,730	112	n/a	94.75%	5.01E+00	1.19E+02	1.00E+00	8.37E+00
FULLVAL	1,070,994	109,324	13,007	100.00%	8.74E+05	6.15E+09	0.00E+00	1.16E+07
AVLAND	1,070,994	70,921	13,009	100.00%	8.51E+04	2.67E+09	0.00E+00	4.06E+06
AVTOT	1,070,994	112,914	13,007	100.00%	2.27E+05	4.67E+09	0.00E+00	6.88E+06
EXLAND	1,070,994	33,419	491,699	100.00%	3.64E+04	4.67E+09	0.00E+00	3.98E+06
EXTOT	1,070,994	64,255	432,572	100.00%	9.12E+04	2.67E+09	0.00E+00	6.51E+06
BLDFRONT	1,070,994	612	228,815	100.00%	2.30E+01	7.58E+03	0.00E+00	3.56E+01
BLDDEPTH	1,070,994	621	228,853	100.00%	3.99E+01	9.39E+03	0.00E+00	4.27E+01
AVLAND2	282,726	58,592	n/a	26.40%	2.46E+05	2.37E+09	3.00E+00	6.18E+06
AVTOT2	282,732	111,361	n/a	26.40%	7.14E+05	4.50E+09	3.00E+00	1.17E+07
EXLAND2	87,449	22,196	n/a	8.17%	3.51E+05	2.37E+09	1.00E+00	1.08E+07
EXTOT2	130,828	48,349	n/a	12.22%	6.57E+05	4.50E+09	7.00E+00	1.61E+07

Table 1.

Categorical Fields

There are 18 categorical fields included in this dataset. Summary information for each categorical fields is shown as Table 2. Fields that marked in red are fields of low percentage populated.

Field Name	Count	Unique Values	Most Common	% Populated
RECORD	1,070,994	1,070,994	Uniform Distributed	100.00%
BBLE	1,070,994	1,070,994	Uniform Distributed	100.00%
B	1,070,994	5	4(33.43%)	100.00%
BLOCK	1,070,994	13,984	3944(0.36%)	100.00%
LOT	1,070,994	6,366	1(2.28%)	100.00%
EASEMENT	4,636	13	E(89.47%)	0.43%
OWNER	1,039,249	863,348	PARKCHESTER PRESERVAT (0.58%)	97.04%
BLDGCL	1,070,994	200	R4(13.06%)	100.00%
TAXCLASS	1,070,994	11	1(61.69%)	100.00%
EXT	354,305	4	G(75.35%)	33.08%
EXCD1	638,488	130	1017(0.1%)	59.62%
STADDR	1,070,318	839,281	501 SURF AVENUE(0.08%)	99.94%
ZIP	1,041,104	197	10314(2.36%)	97.21%
EXMPTCL	15,579	15	X1(44.37%)	1.45%
EXCD2	92,948	61	65777(6.1%)	8.68%
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%

Table 2.

III. Fields Description

1. RECORD

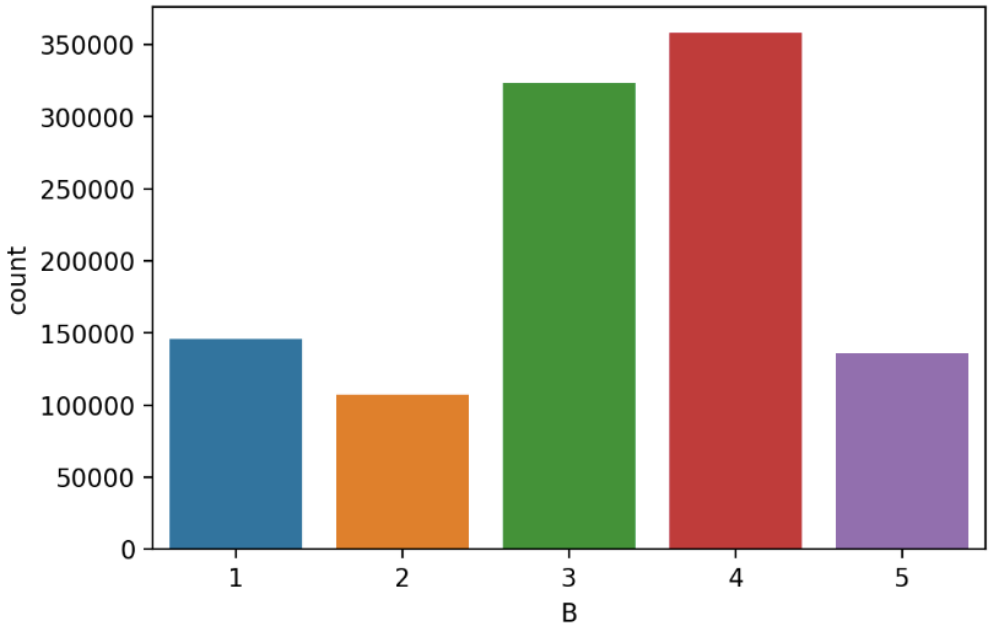
RECORD represents the numbering for each record. There are 1070994 unique values in the dataset which means each record has one corresponding **RECORD**.

2. BBLE

BBLE is the concatenation of **B**, **BLOCK**, **LOT**, and **EASEMENT**. There are 1070994 unique values of **BBLE**, which means every record is associated with one unique value.

3. B

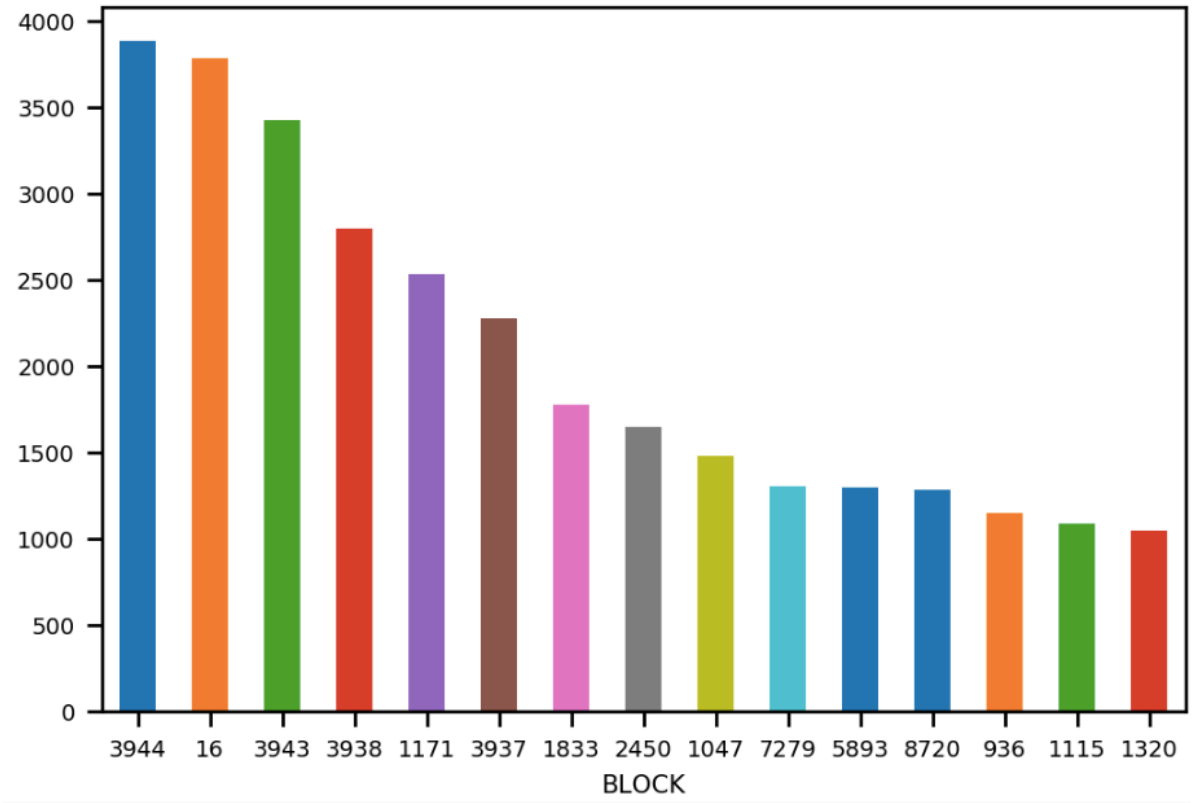
B is borough codes. (1 = Manhattan, 2 = Bronx, 3 = Brooklyn, 4 = Queens, 5 = Staten Island) The bar chart below shows the count for each borough code.



4. BLOCK

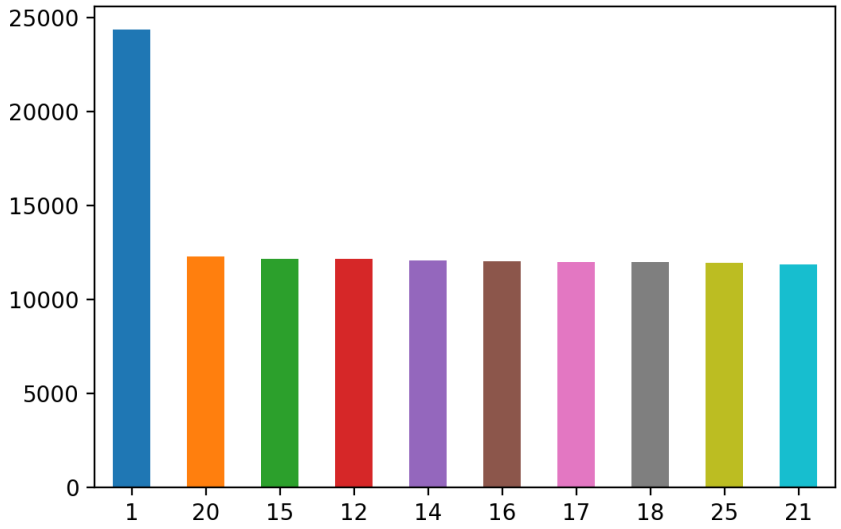
BLOCK represents the valid block ranges by borough codes. The bar chart below shows the count for top 15 most common values.

- 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



5. LOT

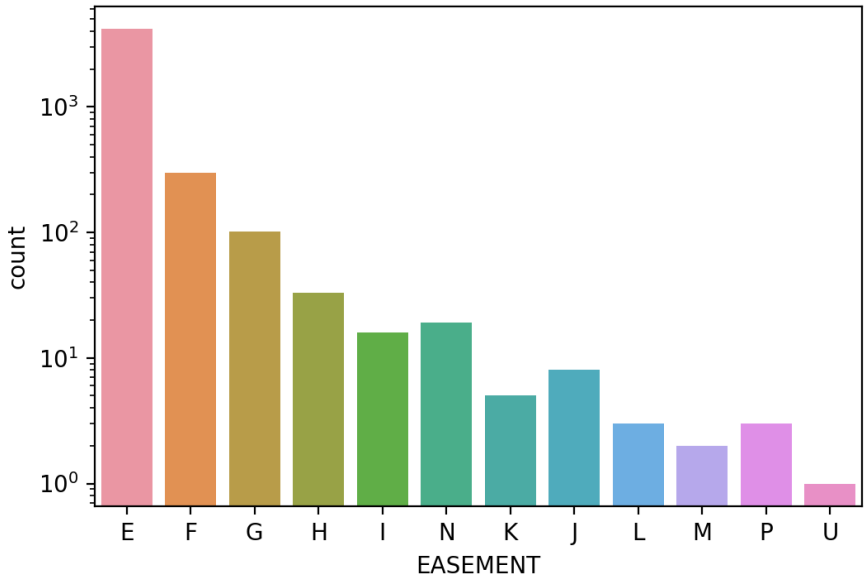
LOT is the unique code within borough codes and block ranges. The bar chart below shows the top 10 most common values in **LOT**.



6. EASEMENT

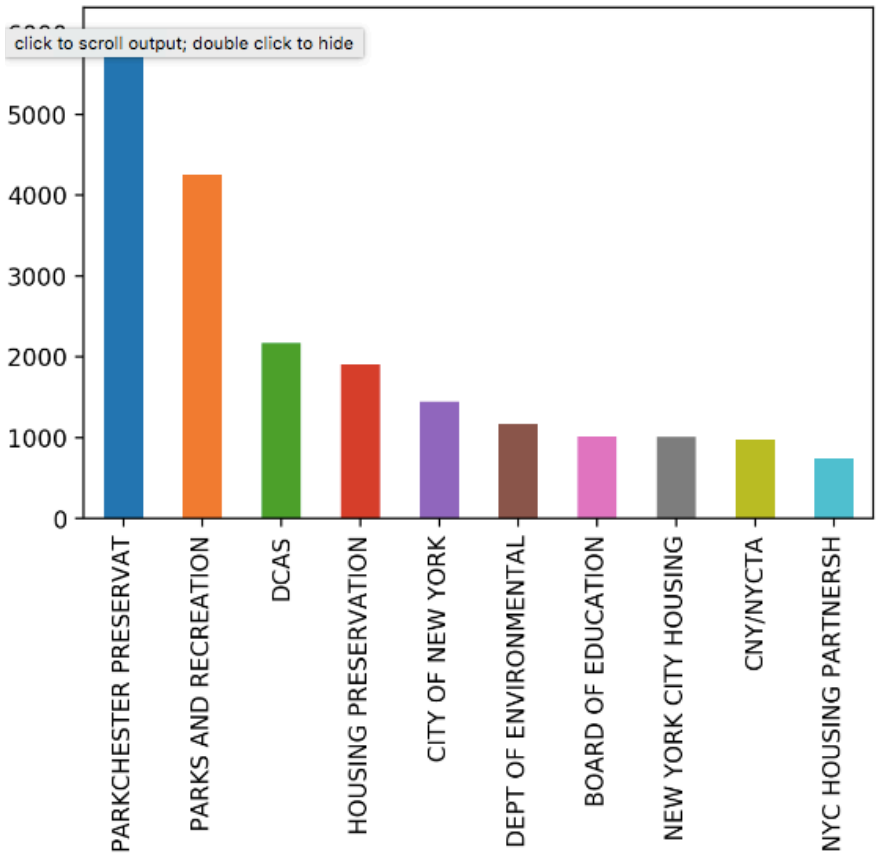
EASEMENT is a legal right to use another's land for a specific limited purpose. The bar chart below shows the log scale count for each value.

- 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



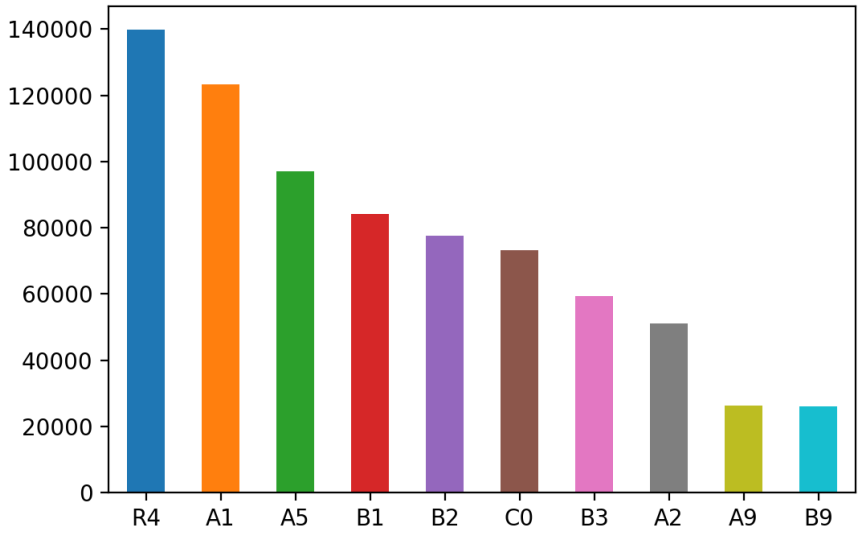
7. OWNER

OWNER represents the property owner's name. The following graph shows the count fir top 10 owners.



8. BLDGCL

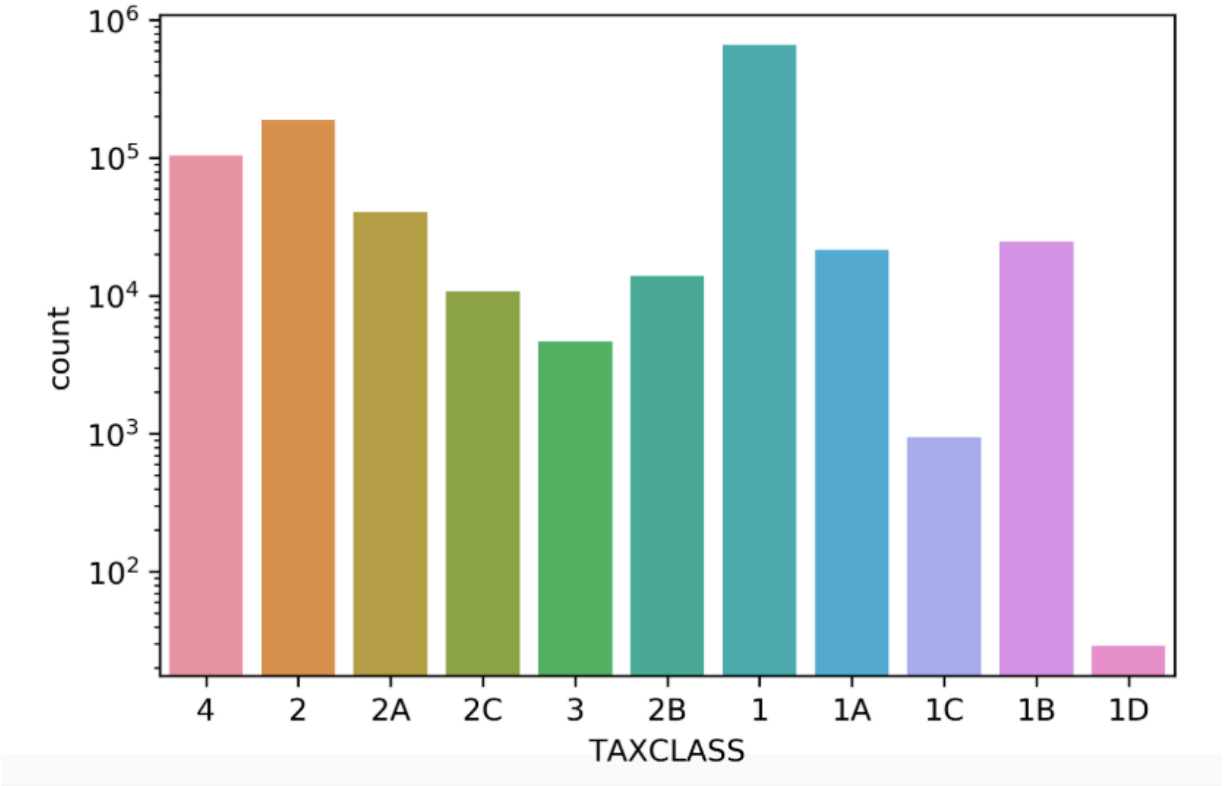
BLDGCL indicates the Building Class. Besides, there is a direct correlation between the Building Class and Tax Class. Bar chart below shows the count of top 10 Building Class.



9. TAXCLASS

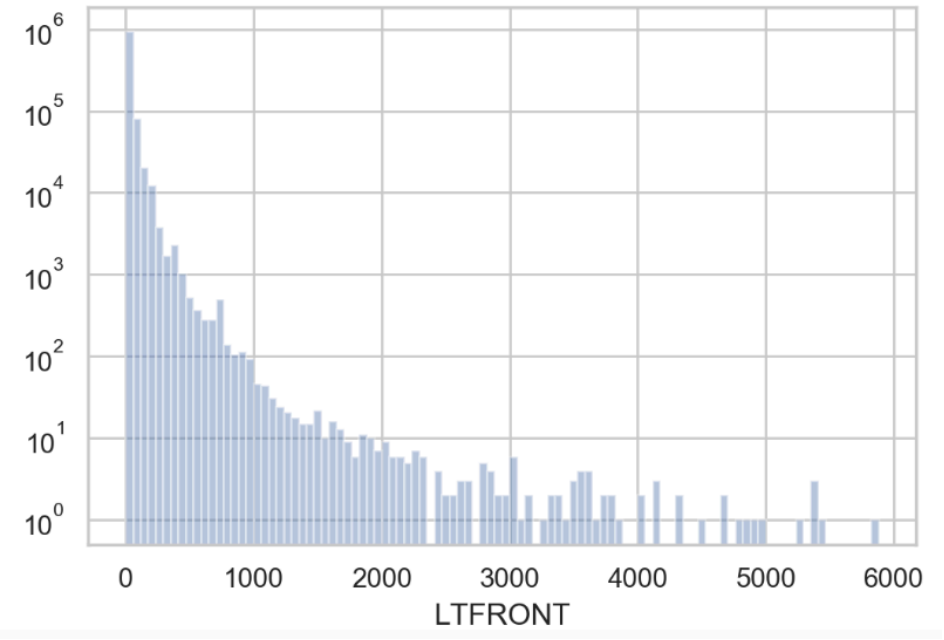
TAXCLASS is the current Property Tax Class Code (NYS classification). The following graph shows count of all Tax Class.

- Tax Class 1 = 1-3 unit residences
- Tax Class 1A = 1-3 story condominiums originally a condo
- Tax Class 1B = residential vacant land
- Tax Class 1C = 1-3 unit 1-3 condominiums originally Tax Class 1
- Tax Class 1D = select bungalow colonies
- Tax Class 2 = apartments
- Tax Class 2A = apartments with 4-6 units
- Tax Class 2B = apartments with 7-10 units
- Tax Class 2C = coops/ condos with 2-10 units
- Tax Class 3 = utilities (except Ceiling Rail Roads)
- Tax Class 4A = utilities - Ceiling Rail Roads
- Tax Class 4 = all others



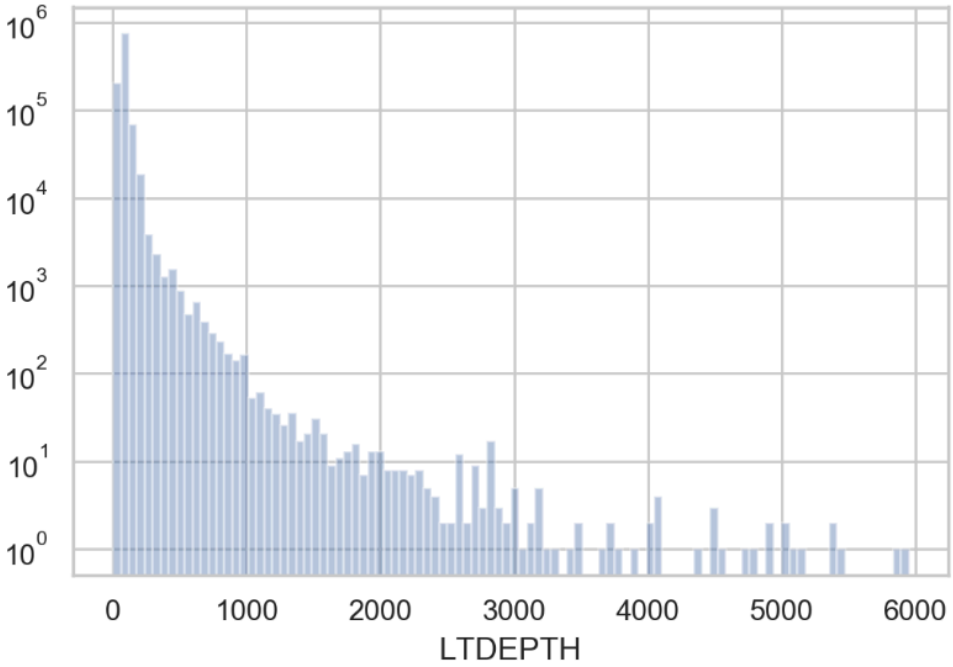
10. LTFRONT

LTFRONT indicates lot frontage in feet. The graph below is the log scale density plot of **LTFRONT.**



11. LTDEPTH

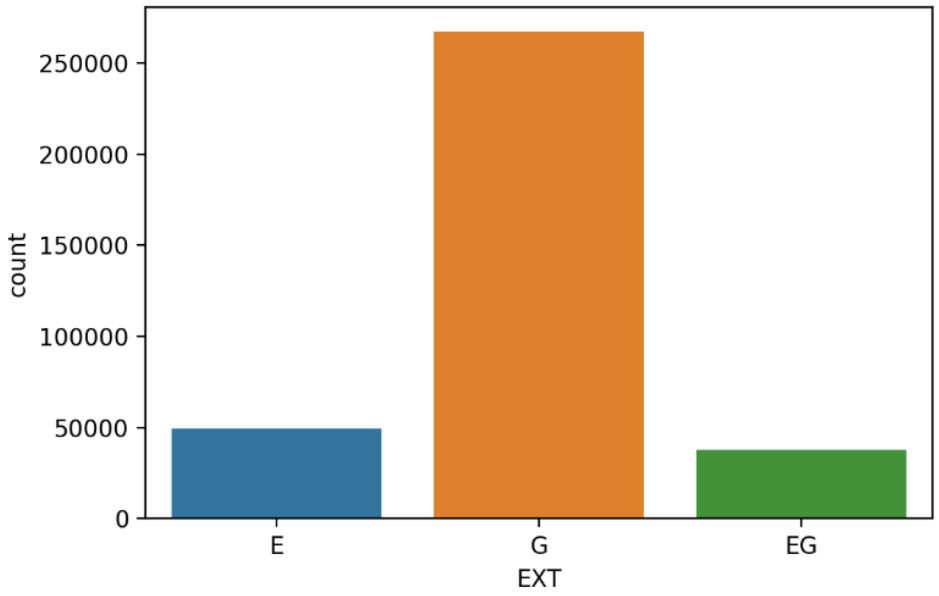
LTDEPTH indicates lot depth in feet. The graph below is the density plot of **LTDEPTH** with depth limited to 600 feet.



12. EXT

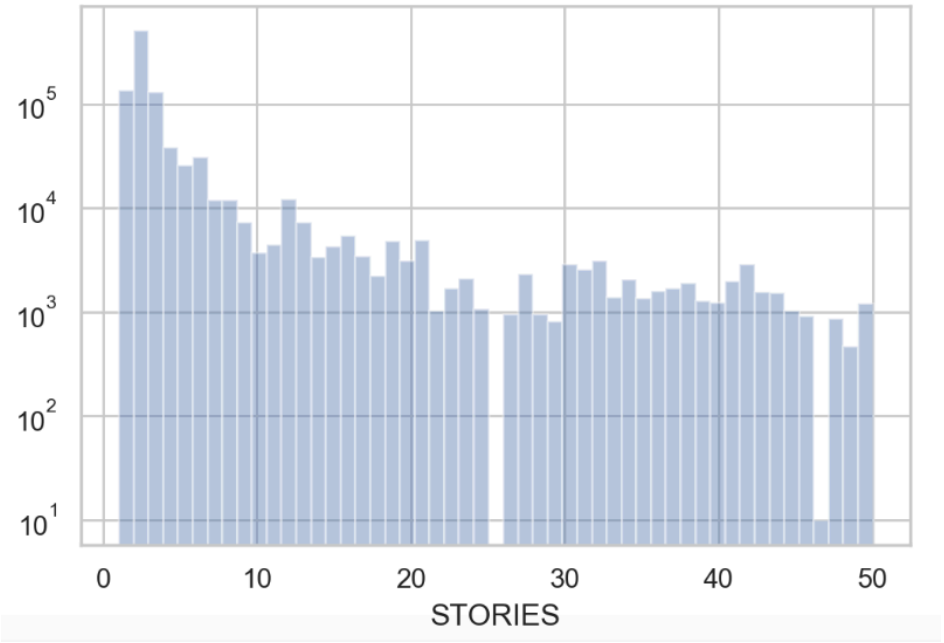
EXT means property extension. The bar chart below shows count for all extension types.

- E = extension
- G = garage
- EG = extension and garage



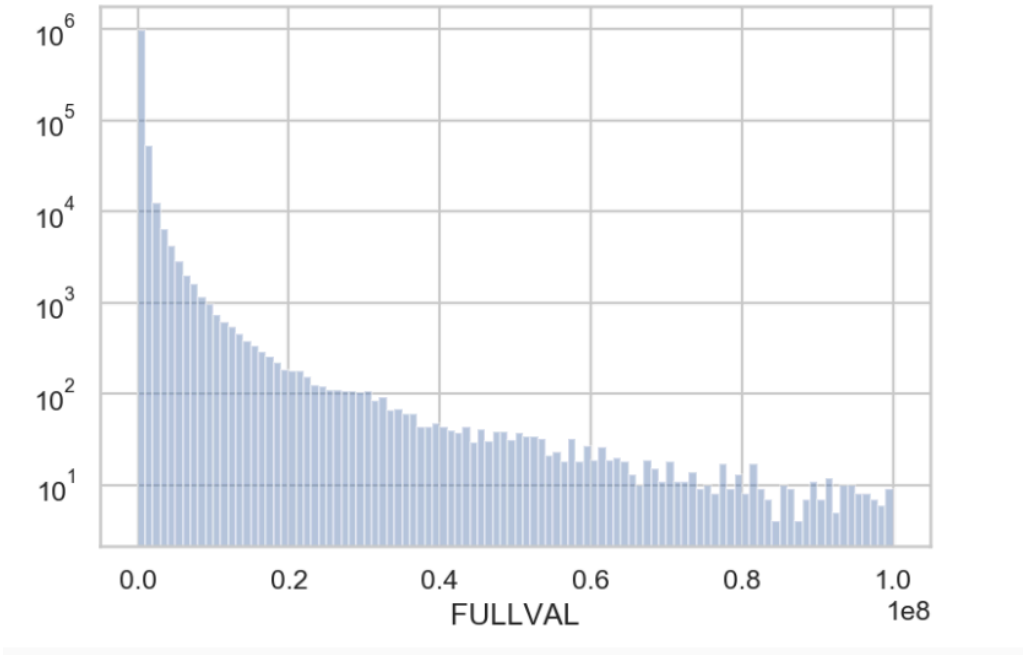
13. STORIES

STORIES represents the number of floors for the building. The following graph is the distribution of **STORIES.**



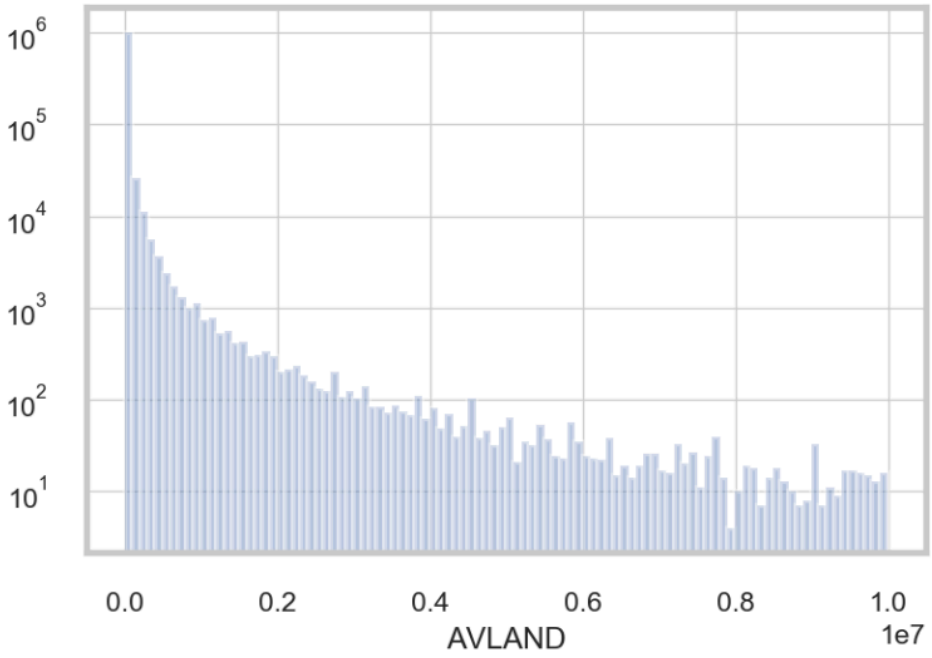
14. FULLVAL

FULLVAL represents total market value of the property. The graph shows the log scale distribution of FULLVAL with value limited to one billion.



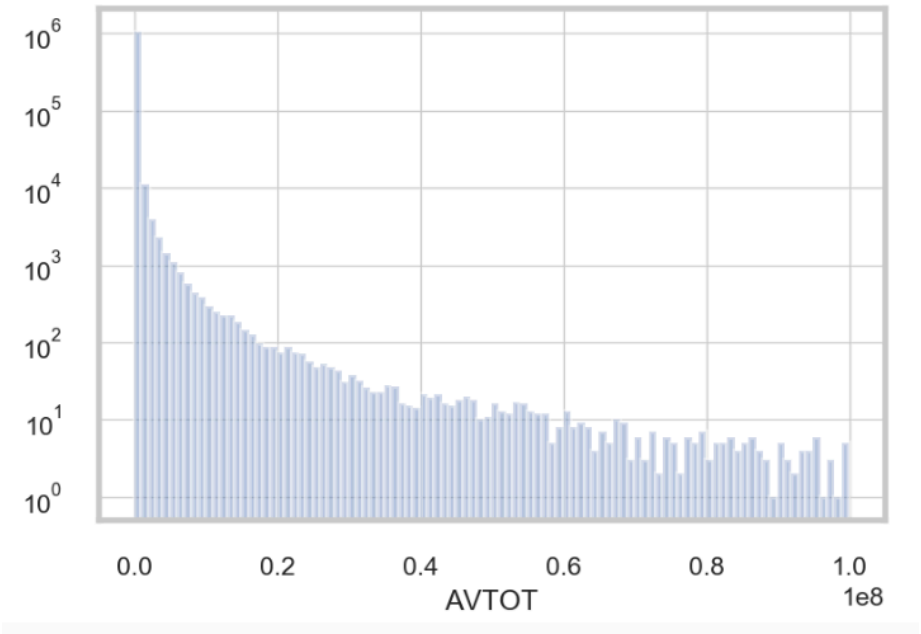
15. AVLAND

AVLAND is the Assessed land value. The following is the log scale distribution of **AVLAND** with value limited to 10 million.



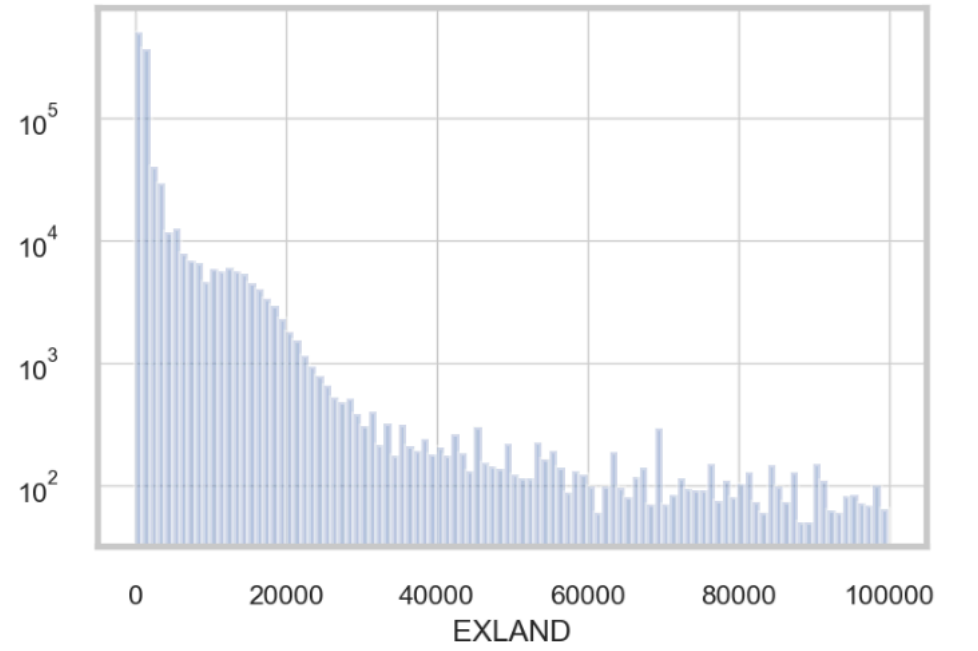
16. AVTOT

AVTOT is the Assessed Values, which is typically less than the market value. The following is the log scale distribution of **AVTOT** with value limited to one billion.



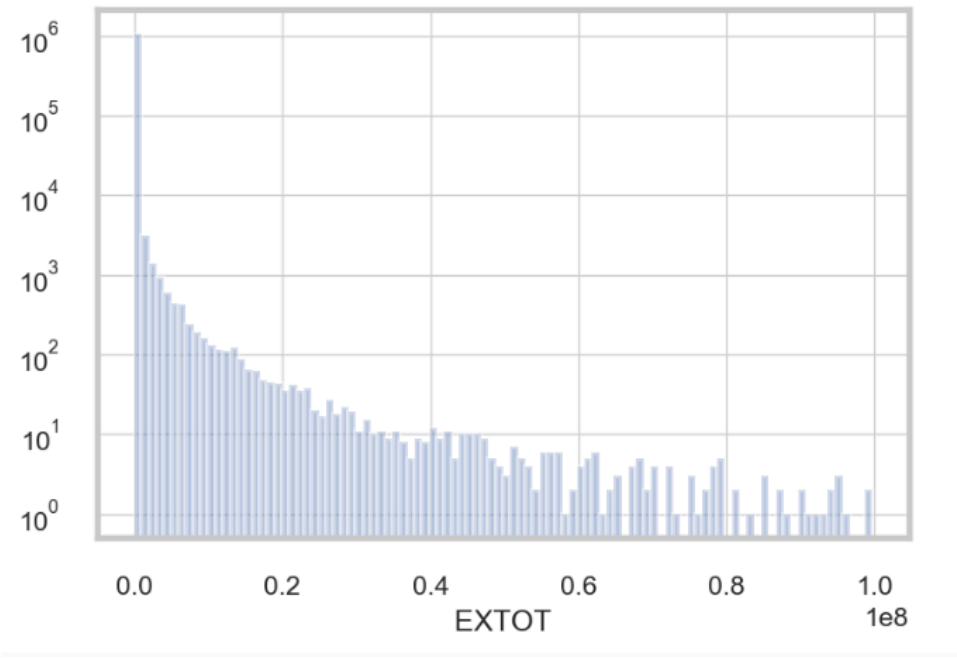
17. EXLAND

EXLAND represents exempt land value. The following graph shows log scale **EXLAND** with land value limited to 100,000.



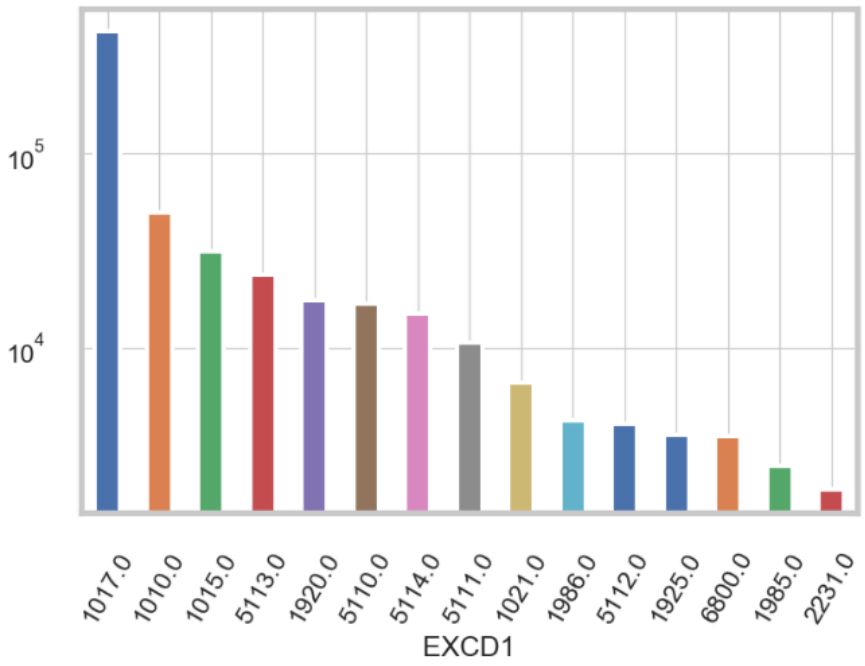
18. EXTOT

EXTOT shows exempt total value. The following graph shows log scale EXTOT with land value limited to one billion.



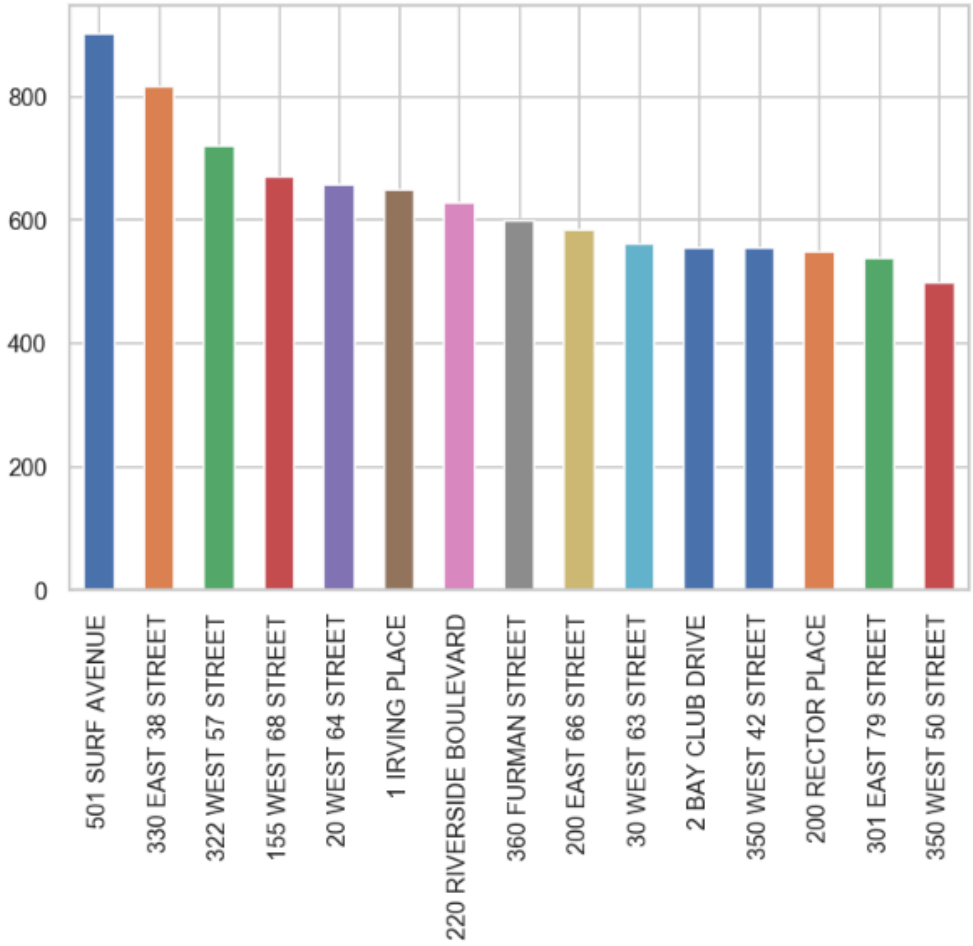
19. EXCD1

The Bar chart below shows the log scale count for top 15 values of **EXCD1.**



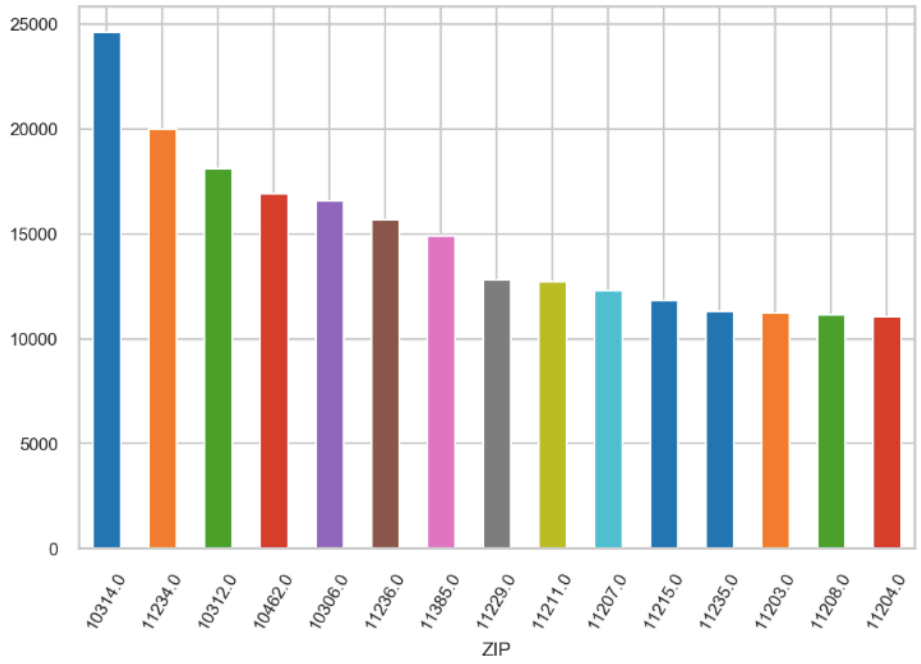
20. STADDR

STADDR is the street name for the property. The bar chart below shows the count of top 15 most common street names for STADDR.



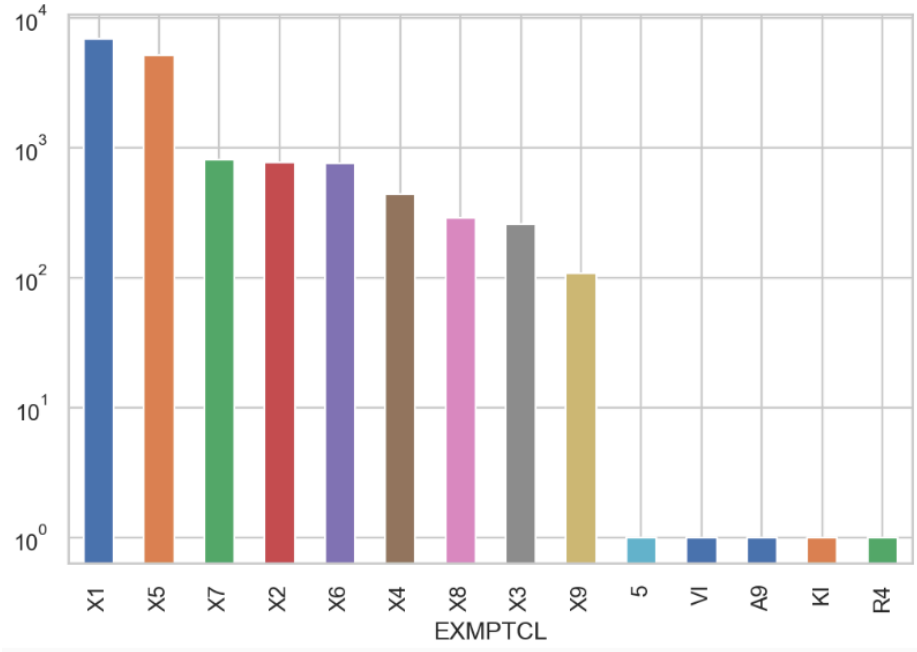
21. ZIP

ZIP is postal zip code for the property. The below graph shows the count for top 15 most common zip codes.



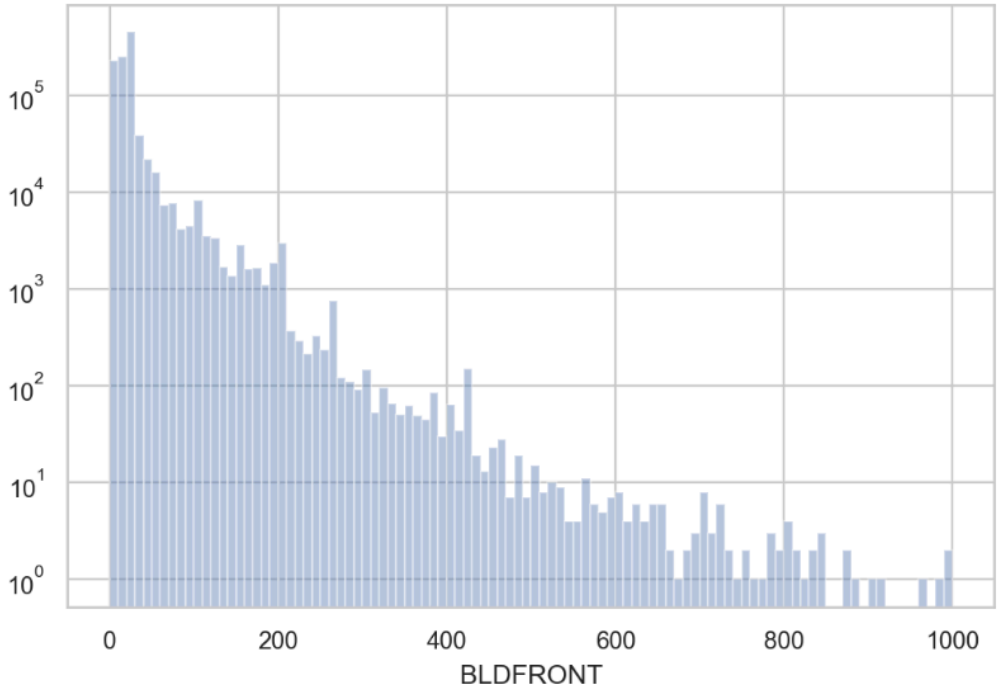
22. EXMPTCL

EXMPTCL represents exempt class, which is used for fully exempt properties only. Bar chart below shows count for all exempt classes.



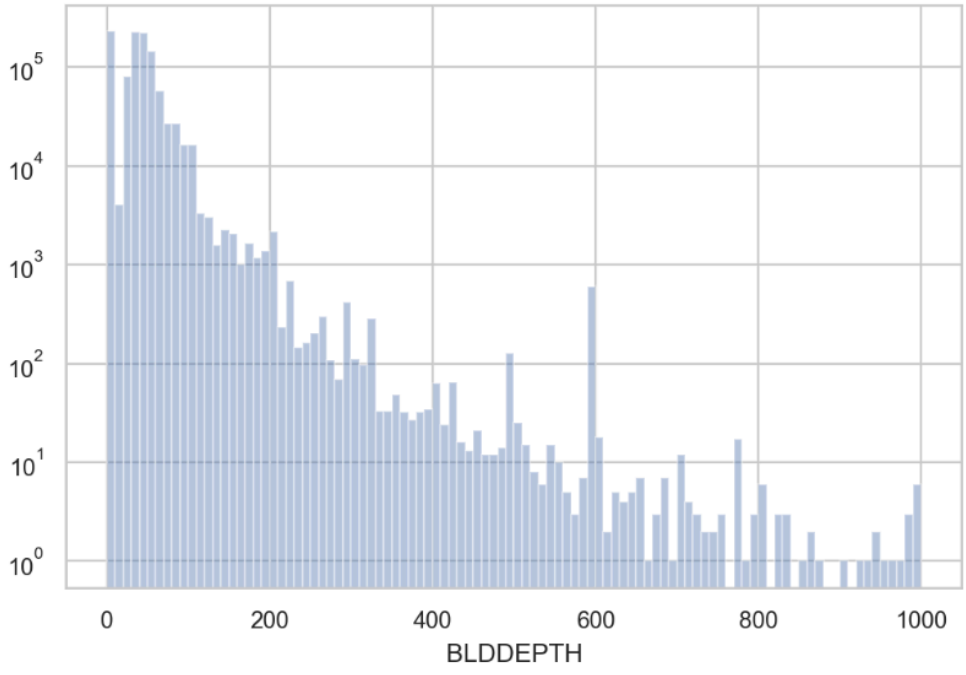
23. BLDFRONT

BLDFRONT indicates the building frontage in feet. Graph below shows the log scale distribution of **BLDFRONT** with frontage feet limited to 1,000.



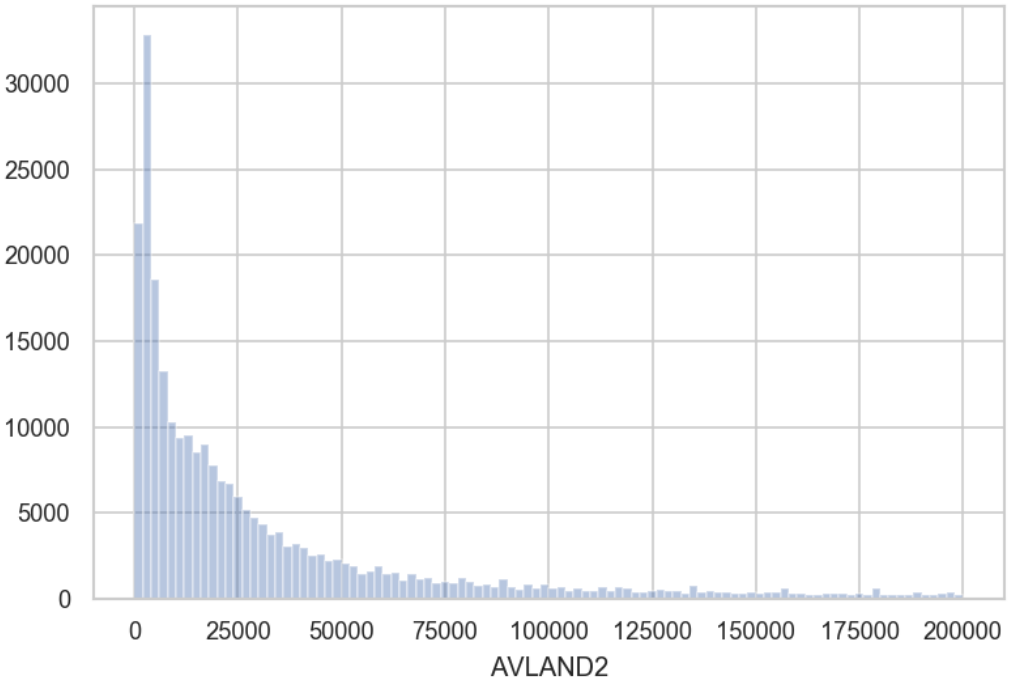
24. BLDDEPTH

BLDDEPTH indicates building depth in feet. The following graph shows the log scale distribution of **BLDDEPTH** with depth limited to 300 feet.



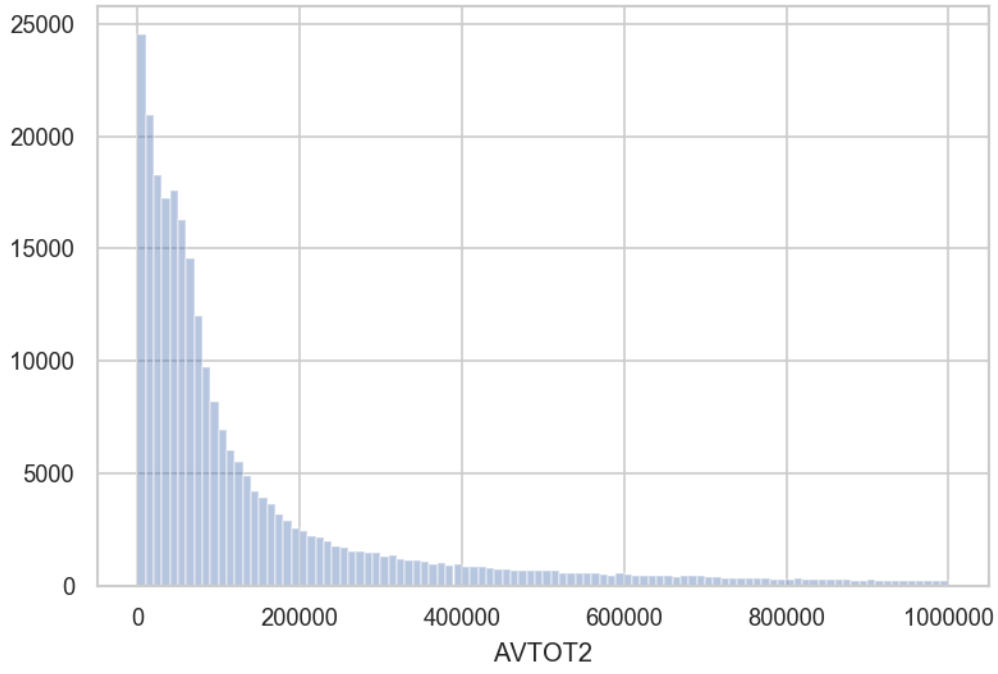
25. AVLAND2

AVLAND is the land value. The following graph is the distribution of **AVLAND2** with value limited to 200,000.



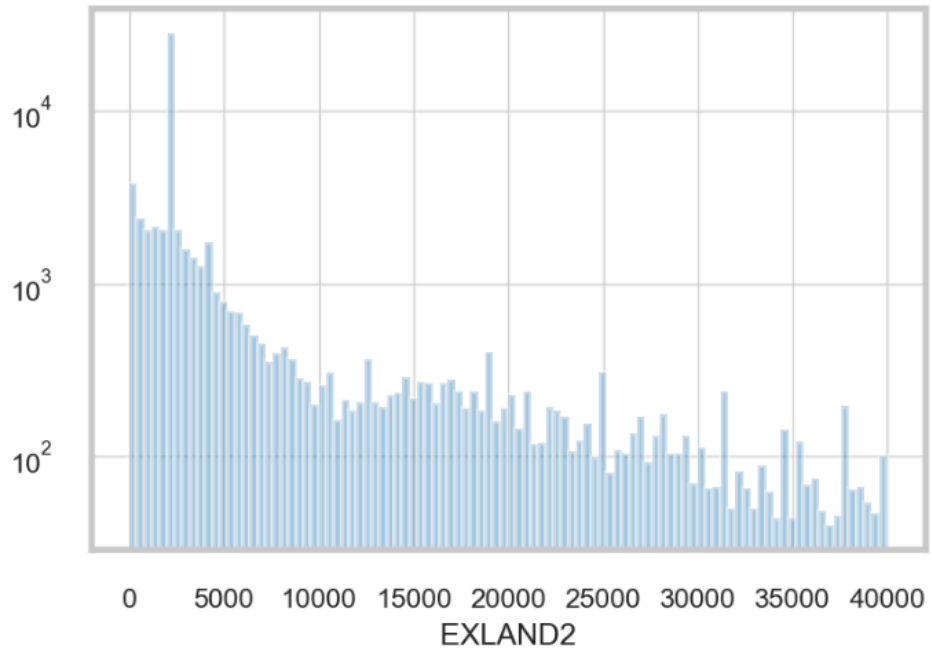
26. AVTOT2

The following graph is the distribution of **AVTOT2** with value limited to 1,000,000.



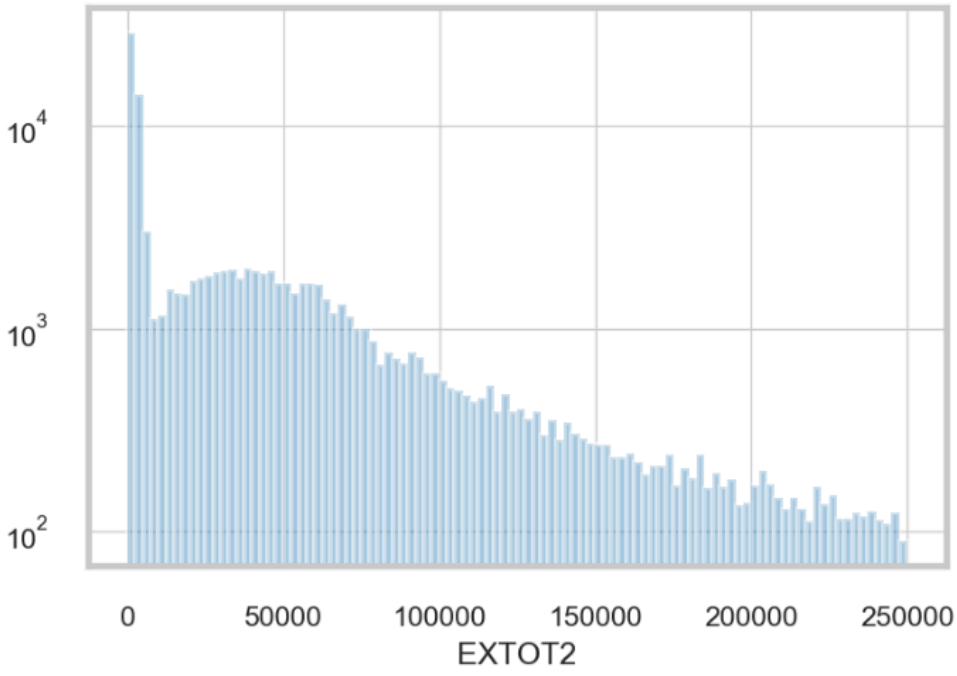
27. EXLAND2

EXLAND2 represents exempt land value. The following graph shows log scale **EXLAND**2 with land value limited to 40,000.



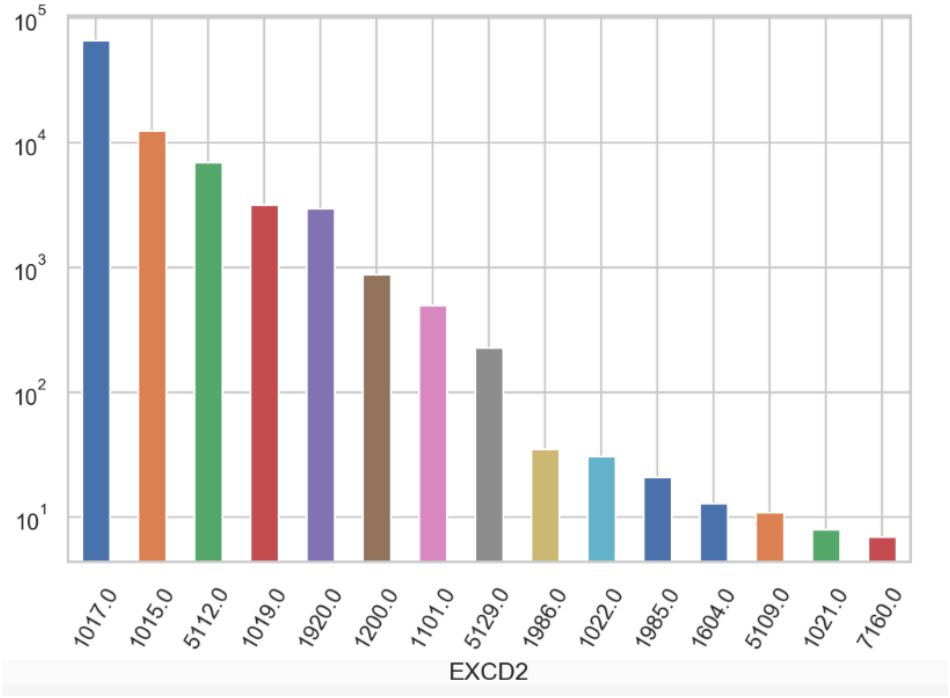
28. EXTOT2

The following graph shows log scale distribution of **EXTOT**2 with value limited to 250,000.



29. EXCD2

The Bar chart below shows the log scale count for top 15 most common values of **EXCD2.**



30. PERIOD

PERIOD contains only one value which is “FINAL”. All 1070994 records are belonging to “FINAL”.

31. YEAR

YEAR is the four-digits year of the file. The only value is “2010/11”. All 1070994 records are under “2010/11”.

32. VALTYPE

VALTYPE includes only one value, which is “AC-TR”. All 1070994 records are under “AC-TR”.