Srikar Gunisetty
A53102026

SECTION-1: INTRODUCTION:

The New York property dataset represents the annual property valuation and assessment provided by the Department of Finance (DOF). The data is collected and entered in the system by various City employees, such as Property Assessors, Property Exemption specialists, ACRIS reporting, Department of Building reporting, and so on. There are **1070994 records** and **32 fields,** and all the records are within year 2010 and 2011.

File Name and Purpose: *Property_Valuation_and_Assessment_Data.xlsx*

Data Collection/Source: [LINK]

Time period covered: 2010/11

Data Provided by: Department of Finance (DOF)

Dataset Owner: NYC Open Data

Category: Housing & Development

Date Created: September 2nd, 2011

Last Updated: September 10th, 2018

Data Volume: 1,070,994

Fields: 32

Field Variables: RECORD, BBLE, BORO, BLOCK, LOT, EASEMENT, OWNER, BLDGCL, TAXCLASS, LTFRONT, LTDEPTH, EXT, STORIES, FULLVAL, AVLAND, AVTOT, EXLAND, EXTOT, EXCD1, STADDR, ZIP, EXMPTCL, BLDFRONT, BLDDEPTH, AVLAND2, AVTOT2, EXLAND2, EXTOT2, EXCD2, PERIOD, YEAR, and VALTYPE.

Srikar Gunisetty
A53102026

SECTION-2: SUMMARY TABLES:

Table:1 – Field Names with Description

Field Name	Data Type	Type	Description
RECORD	*int64*	*Categorical*	Record Number
BBLE	*object*	*Categorical*	Concatenation of BORO, BLOCK, LOT, EASEMENT
BORO	*int64*	*Categorical*	BORO or Borough
BLOCK	*int64*	*Categorical*	Valid block ranges by BORO codes
LOT	*int64*	*Categorical*	Unique # of the property within BORO/BLOCK
EASEMENT	*object*	*Categorical*	Used to describe easement
OWNER	*object*	*Categorical*	Owner's name
BLDGCL	*object*	*Categorical*	Building Class
TAXCLASS	*object*	*Categorical*	Tax Class
LTFRONT	*int64*	*Numerical*	Lot frontage in feet
LTDEPTH	*int64*	*Numerical*	Lot depth in feet
EXT	*object*	*Categorical*	No info
STORIES	*float64*	*Numerical*	Number of stories in the building
FULLVAL	*float64*	*Numerical*	Total market value of the property
AVLAND	*float64*	*Numerical*	Assessed land value of the property
AVTOT	*float64*	*Numerical*	Assessed total value of the property
EXLAND	*float64*	*Numerical*	Part of land value that is tax exempted
EXTOT	*float64*	*Numerical*	Part of total assessed value that is tax exempted
EXCD1	*float64*	*Categorical*	No info
STADDR	*object*	*Categorical*	Street address
ZIP	*float64*	*Categorical*	Postal zip code of the property
EXMPTCL	*object*	*Categorical*	Exempt class used for fully exempt properties
BLDFRONT	*int64*	*Numerical*	Building frontage in feet
BLDDEPTH	*int64*	*Numerical*	Building depth in feet
AVLAND2	*float64*	*Numerical*	No info
AVTOT2	*float64*	*Numerical*	No info
EXLAND2	*float64*	*Numerical*	No info
EXTOT2	*float64*	*Numerical*	No info
EXCD2	*float64*	*Categorical*	No info
PERIOD	*object*	*Categorical*	No info
YEAR	*object*	*Date/Time*	Assessment year
VALTYPE	*object*	*Categorical*	No info

Srikar Gunisetty
A53102026

Table:2 – Summary Statistics for Numeric Fields

Field Name	Count	[%] full	# Zeros	mean [μ]	std [σ]	min	25%	50%	75%	max
LTFRONT	1070994	100.0	169108	36.64	74.03	0.0	19.0	25.0	40.00	9.999000e+03
LTDEPTH	1070994	100.0	170128	88.86	76.40	0.0	80.0	100.0	100.00	9.999000e+03
STORIES	1014730	94.7	0	5.01	8.37	1.0	2.0	2.0	3.00	1.190000e+02
FULLVAL	1070994	100.0	13007	874264.51	11582430.99	0.0	304000.0	447000.0	619000.00	6.150000e+09
AVLAND	1070994	100.0	13009	85067.92	4057260.06	0.0	9180.0	13678.0	19740.00	2.668500e+09
AVTOT	1070994	100.0	13007	227238.17	6877529.31	0.0	18374.0	25340.0	45438.00	4.668309e+09
EXLAND	1070994	100.0	491699	36423.89	3981575.79	0.0	0.0	1620.0	1620.00	2.668500e+09
EXTOT	1070994	100.0	432572	91186.98	6508402.82	0.0	0.0	1620.0	2090.00	4.668309e+09
BLDFRONT	1070994	100.0	228815	23.04	35.58	0.0	15.0	20.0	24.00	7.575000e+03
BLDDEPTH	1070994	100.0	228853	39.92	42.71	0.0	26.0	39.0	50.00	9.393000e+03
AVLAND2	282726	26.4	0	246235.72	6178962.56	3.0	5705.0	20145.0	62640.00	2.371005e+09
AVTOT2	282732	26.4	0	713911.44	11652528.95	3.0	33912.0	79962.5	240551.00	4.501180e+09
EXLAND2	87449	8.2	0	351235.68	10802212.67	1.0	2090.0	3048.0	31779.00	2.371005e+09
EXTOT2	130828	12.2	0	656768.28	16072510.17	7.0	2870.0	37062.0	106840.75	4.501180e+09

YEAR – all values for this field are between 2010/2011 – constant.

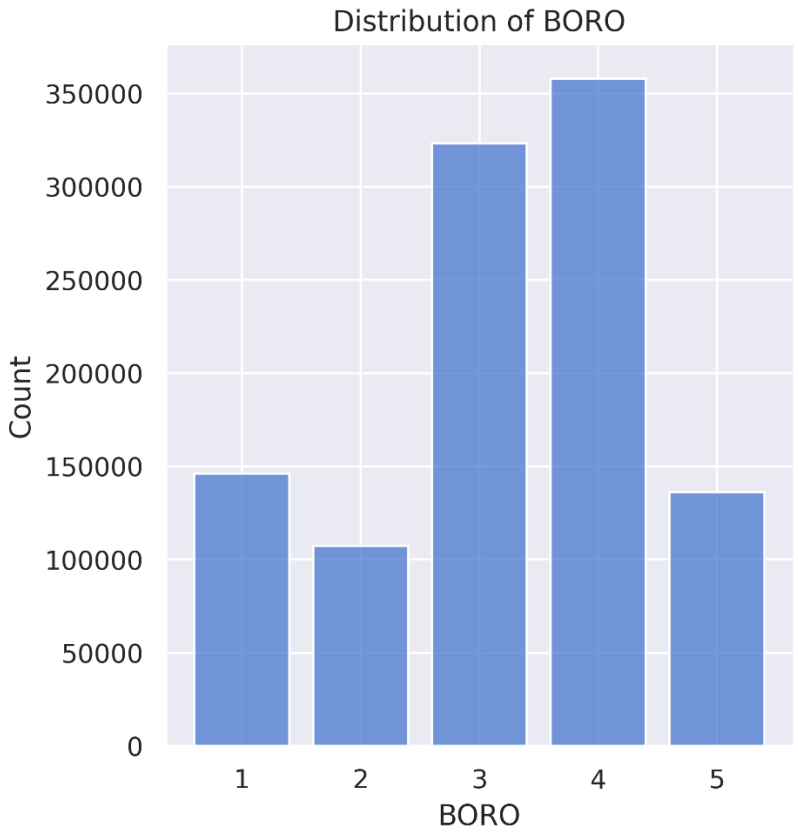
Table:3 – Summary Statistics for Categorical Fields

Field Name	Count	[%] full	# Unique	Most Frequent Value
RECORD	1070994	100.0	1070994	EVERY VALUE IS UNIQUE
BBLE	1070994	100.0	1070994	EVERY VALUE IS UNIQUE
BORO	1070994	100.0	5	4.0 = QUEENS
BLOCK	1070994	100.0	13984	3944
LOT	1070994	100.0	6366	1
EASEMENT	4636	0.4	12	E
OWNER	1039249	97.0	863347	PARKCHESTER PRESERVATORY
BLDGCL	1070994	100.0	200	R4
TAXCLASS	1070994	100.0	11	1
EXT	354305	33.1	3	G
EXCD1	638488	59.6	129	1017
STADDR	1070318	99.9	839280	501 SURF AVENUE
ZIP	1041104	97.2	196	10314
EXMPTCL	15579	1.5	14	X1
EXCD2	92948	8.7	60	1017
PERIOD	1070994	100.0	1	FINAL
VALTYPE	1070994	100.0	1	AC-TR

Srikar Gunisetty
A53102026

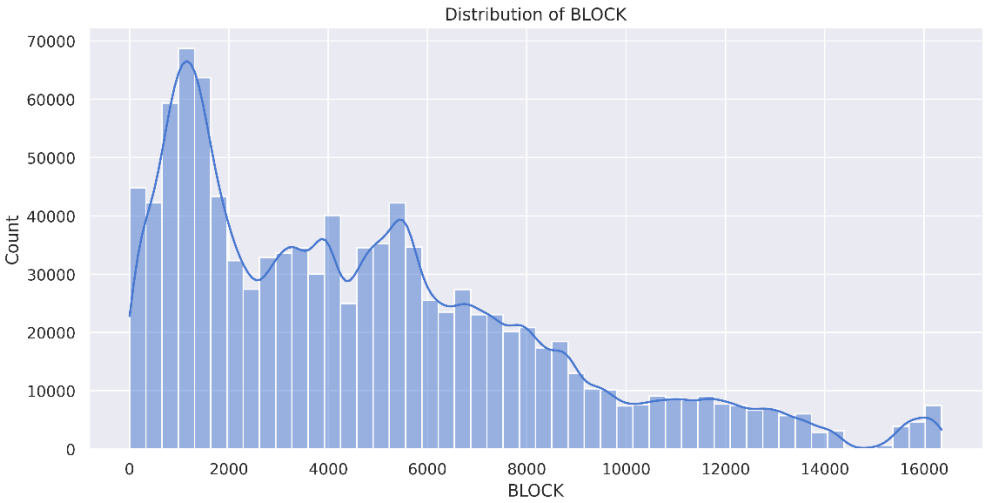
SECTION-3: FIELD DESCRIPTIONS:

1. **RECORD:**
   a. Type: Categorical
   b. Description:
      i. Number to uniquely identify each record.
      ii. All the values in this field are unique, ranging from 1 to 1070994.
      iii. Nothing major to interpret with a graph. Skipped.
2. **BBLE:**
   a. Type: Categorical
   b. Description:
      i. Concatenation of AV_BORO, AV_BLOCK, AV_LOT, AV_EASEMENT.
      ii. All the values in this field are all unique.
      iii. Nothing major to interpret with a graph. Skipped.
3. **BORO:**
   a. Type: Categorical
   b. Description:
      i. BORO CODES, there are five unique values ranging: [1,2,3,4,5].



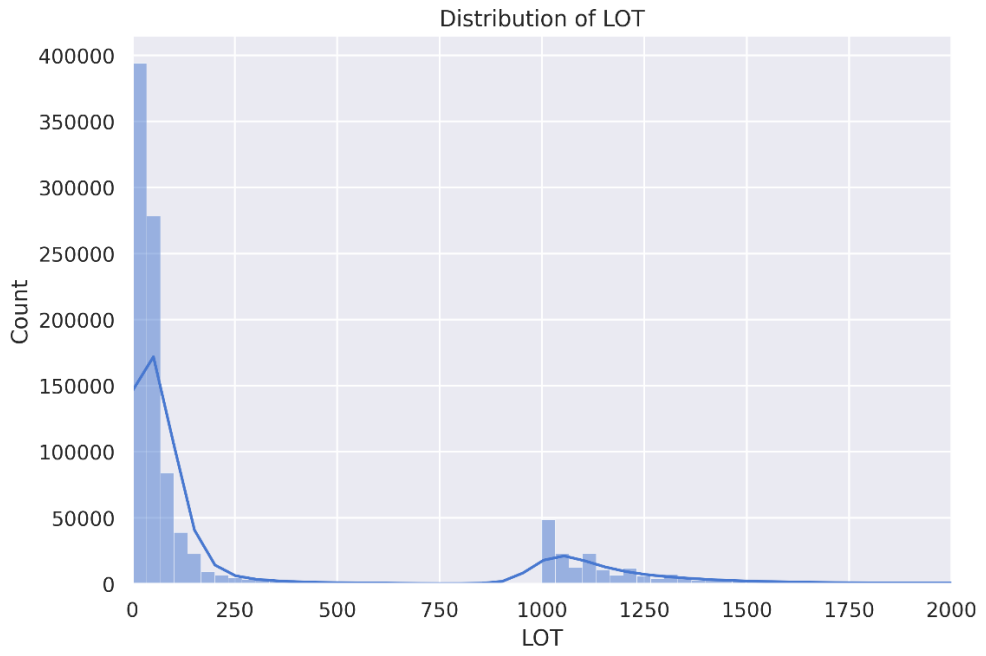
1	MANHATTAN	**2**	BRONX	**3**	BROOKLYN	**4**	QUEENS	**5**	STATEN ISLAND

4. **BLOCK:**
    a) Type: Categorical
    b) Description:
        i. Valid BLOCK ranges by BORO. The meaning are as follows:

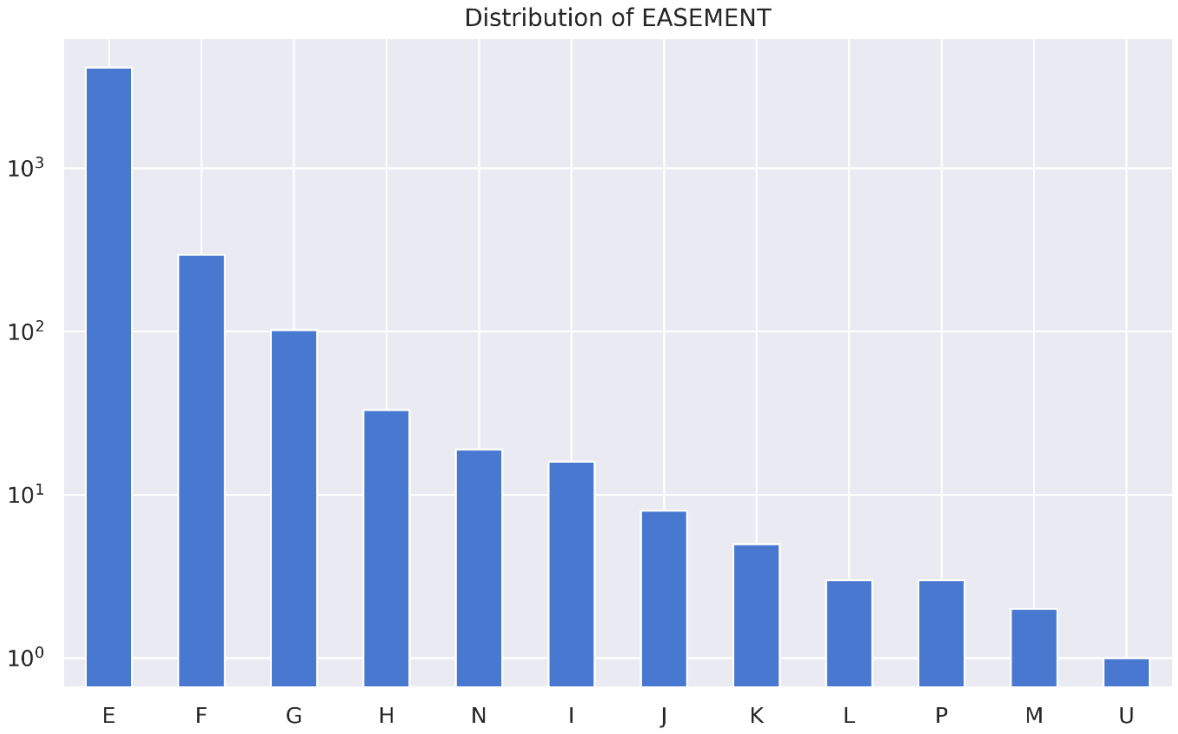


MANHATTAN	**1 to 2,255**	BROOKLYN	**1 to 8,955**	STATEN ISLAND	**1 to 8,050**
BRONX	**2,260 to 5,958**	QUEENS	**1 to 16,350**		

5. **LOT:**
    a) Type: Categorical
    b) Description:
        i. UNIQUE # WITHIN BORO/BLOCK; noted that the values are ranging from 1 to 9978. The below graph only shows a portion of the data.

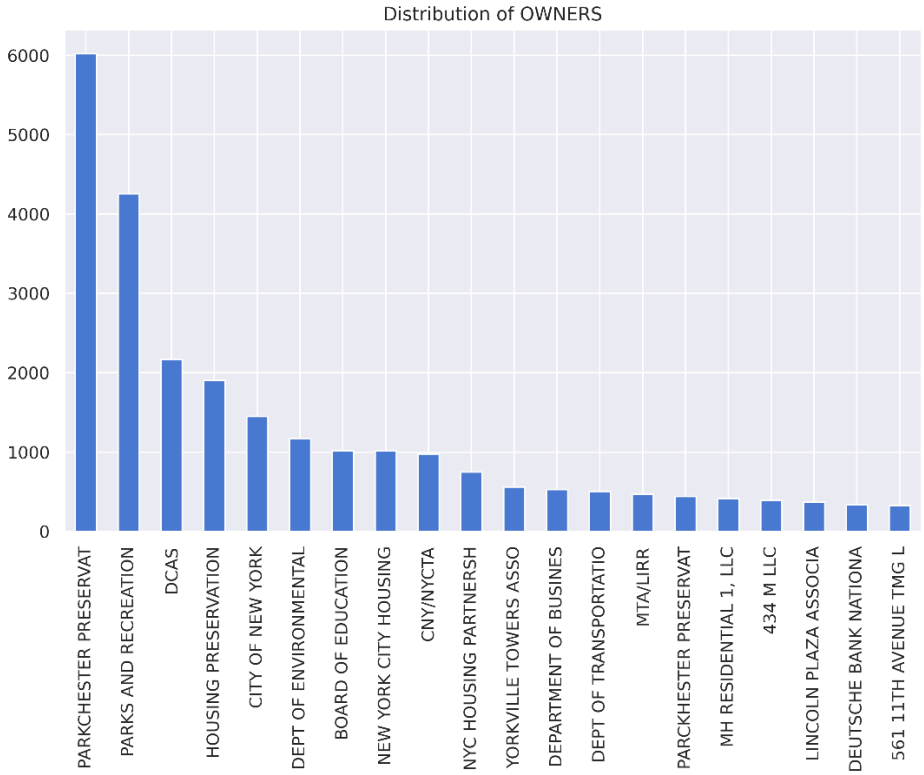


6. **EASEMENT:**
    a) Type: Categorical
    b) Description:
        i. This field that is used to describe easement. An easement is a legal right to use another's land for a specific limited purpose. The meaning are as follows:
        ii. Among the 4636 lots that have easement, 4181 of them have a land easement. When we took log on y-scale, the graph became much more interpretable.

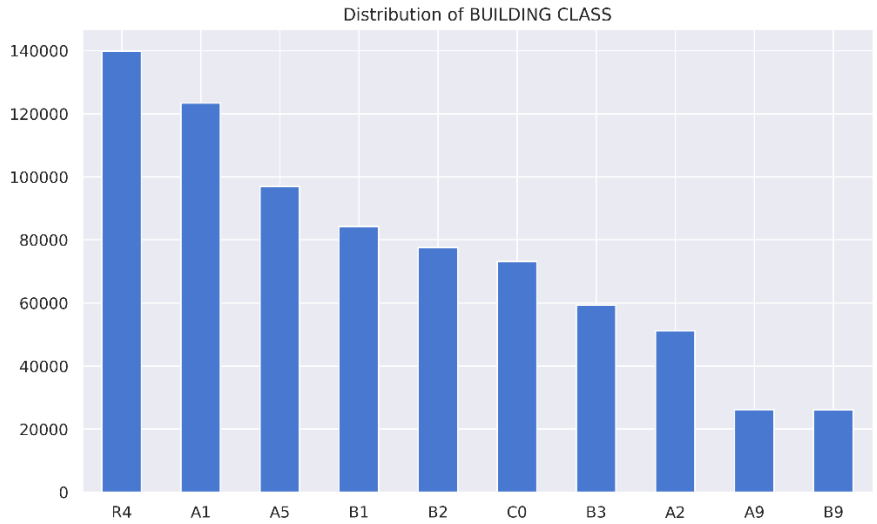


SPACE Indicates the lot has no Easement.
'A' Indicates the portion of the Lot that has an Air Easement
'B' Indicates Non-Air Rights.
'E' Indicates the portion of the lot that has a Land Easement
'F' THRU 'M' Are duplicates of 'E'.
'N' Indicates Non-Transit Easement
'P' Indicates Piers.
'R' Indicates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

7. **OWNER:**
    a) Type: Categorical
    b) Description:
        i. Below are the top 20 owners that appear in the dataset the most

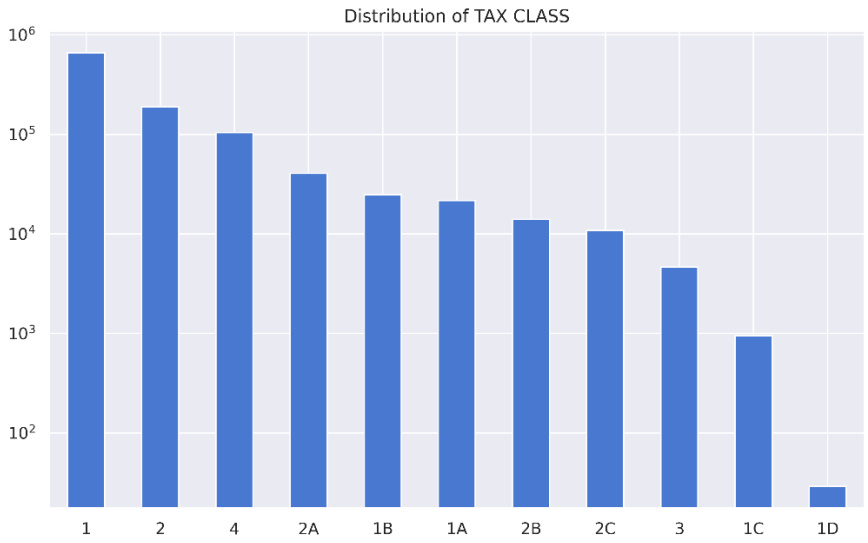


8. **BLDGC – BUILDING CLASS:**
    a) Type: Categorical
    b) Description:
        i. There is a direct correlation between the building class and Tax Class



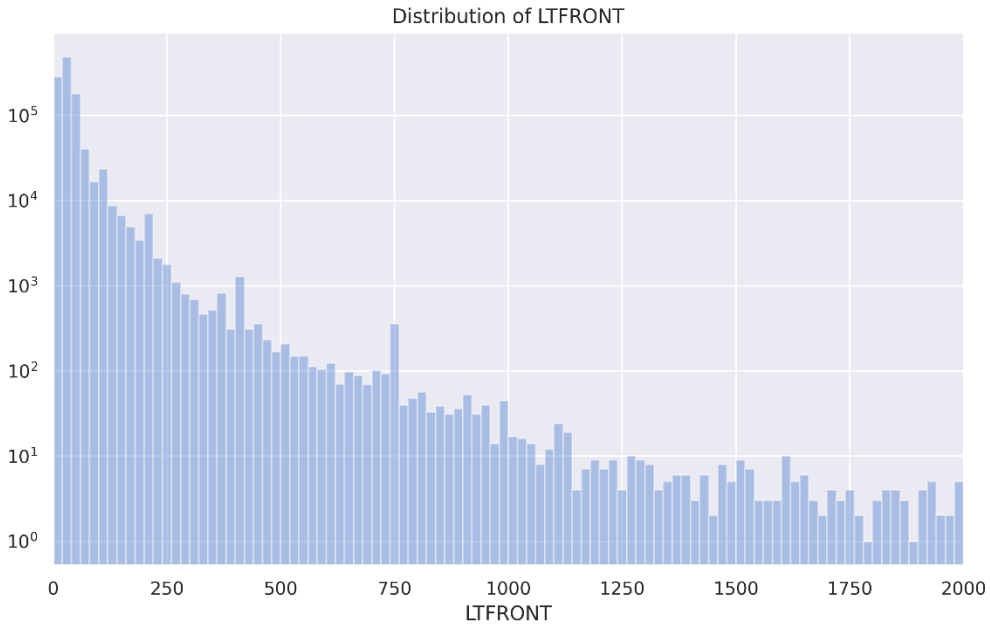
9. TAXCLASS:

a) Type: Categorical
b) Description:
   i. During the final period, contains the tax class from the CBN period. Below graph I took log on the y-axis:



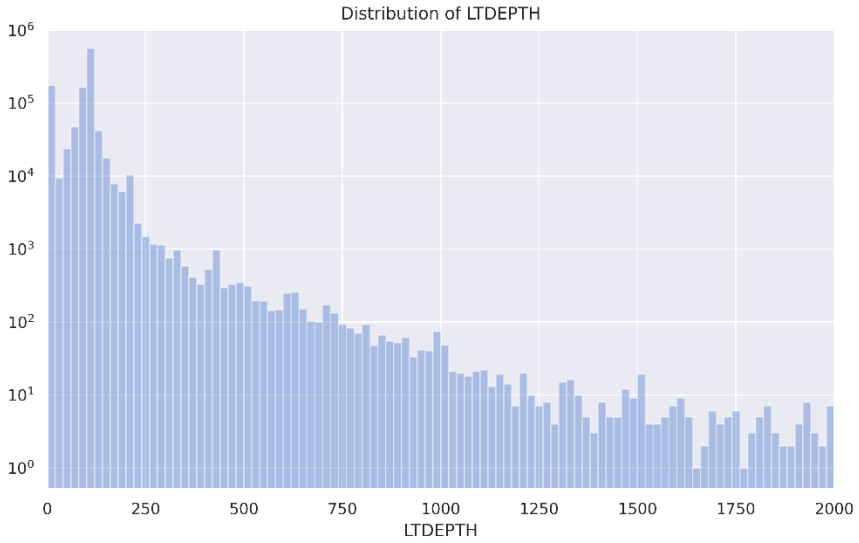
10. LTFRONT:

a) Type: Numerical
b) Description:
   i. Lot Frontage in feet (lot width). The maximum value of this field is 9999.
   ii. To get a closer look, I limited the LFTRONT on log on y-axis.



11. LTDEPTH:

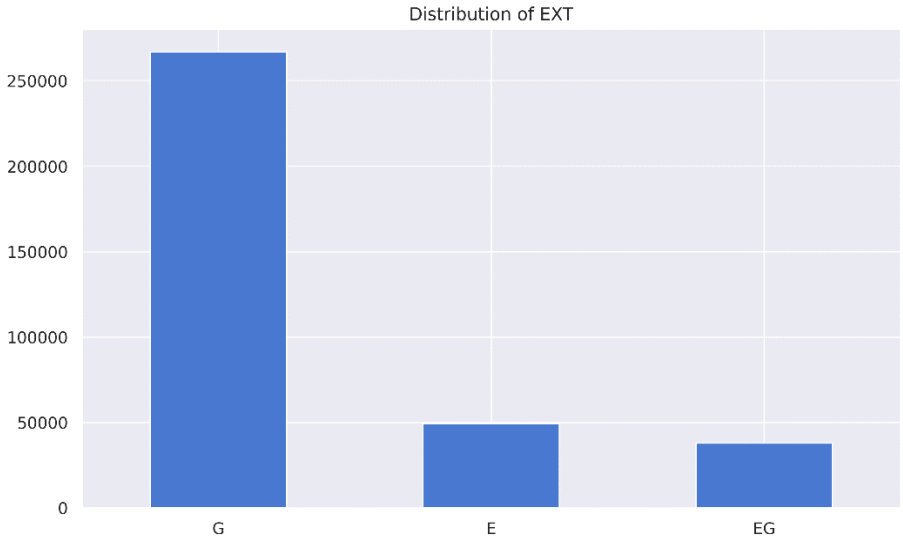
a) Type: Numerical
b) Description:
   i. Lot Depth in feet (lot depth).
   ii. To get a closer look, I limited the LTDEPTH on log on y-axis.



12. EXT:

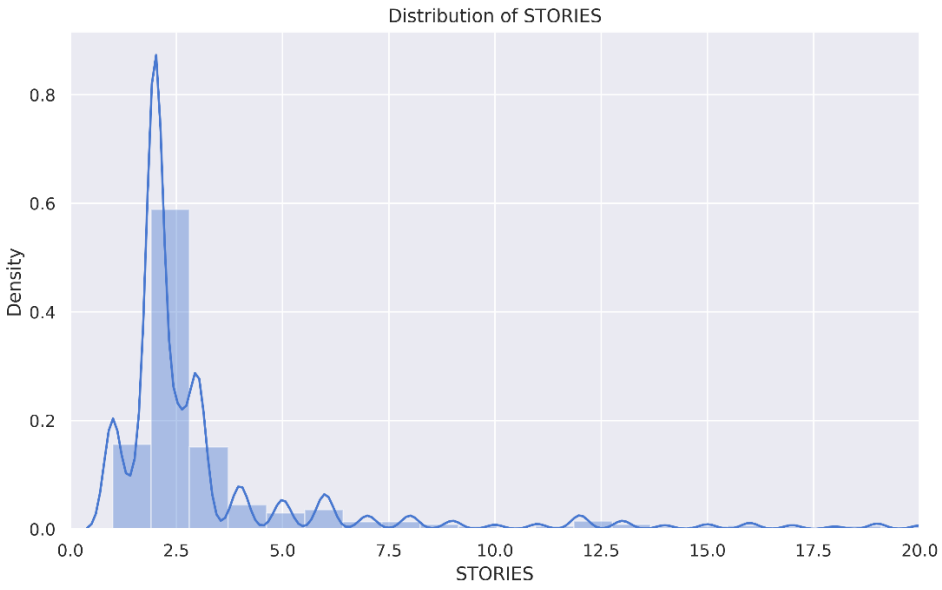
a) Type: Categorical
b) Description:
   i. Extension indicator: The meaning of the values are as follows:

'E' = EXTENSION
'G' = GARAGE
'EG' = EXTENSION AND GARAGE



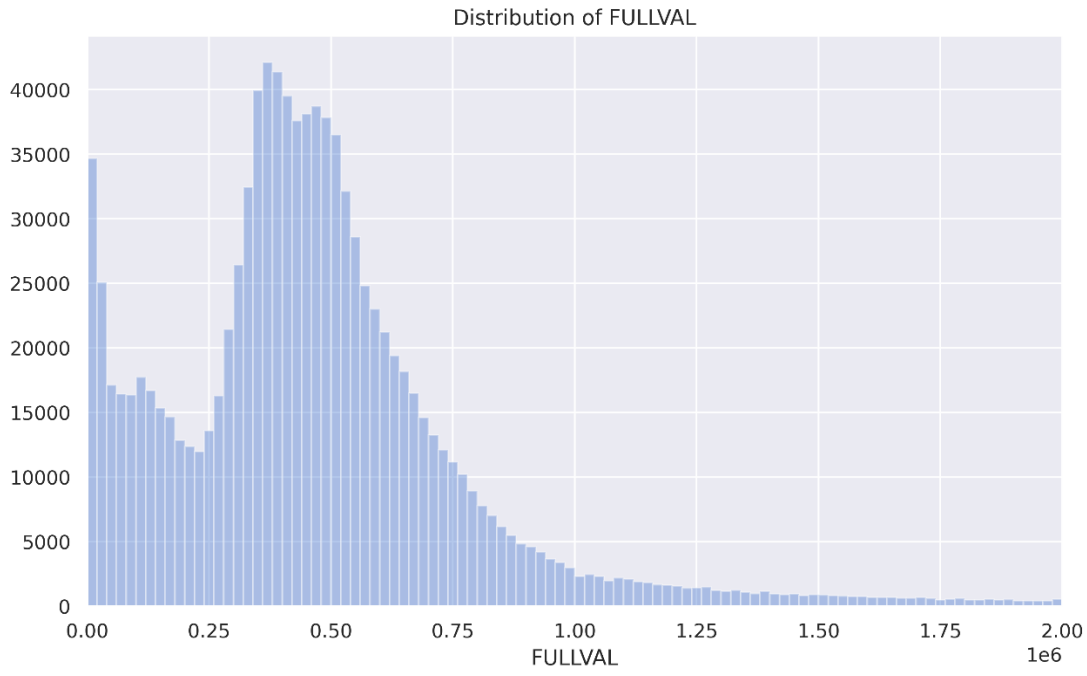
13. STORIES:

a) Type: Numerical
b) Description:
   i. The number of stories in building (# of Floors)



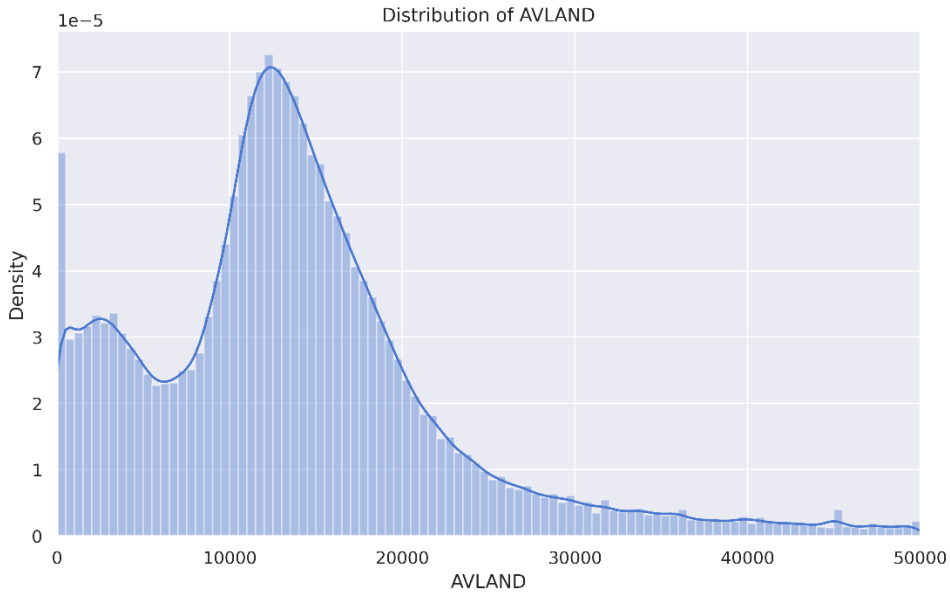
14. FULLVAL:

a) Type: Numerical
b) Description:
   i. Market value of the property.
   ii. In addition, to get a closer look, I limited the AVLAND to be under 200000.



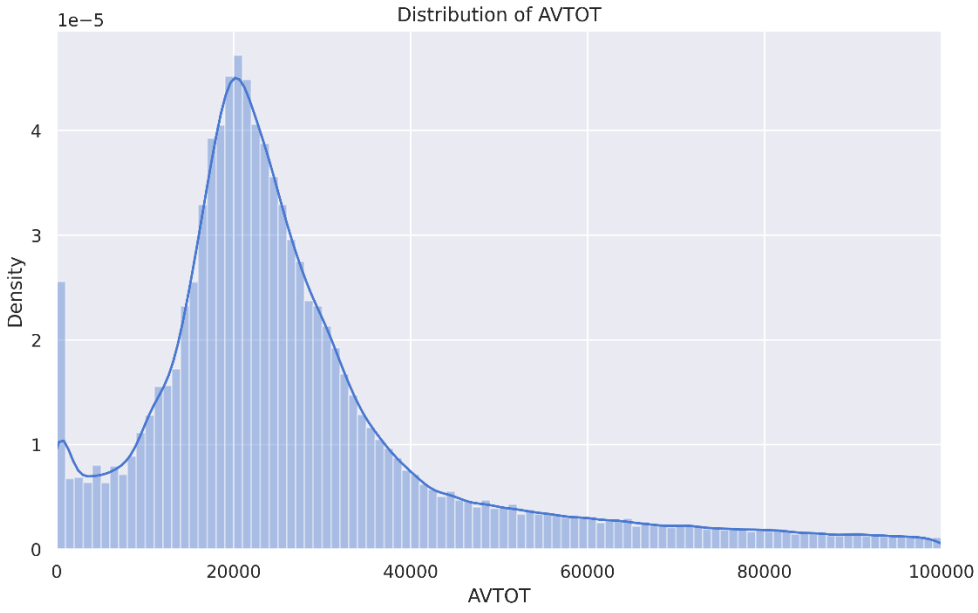
15. AVLAND:

a) Type: Numerical
b) Description:
   i. Actual land value.
   ii. In addition, to get a closer look, I limited the AVLAND to be under 50000.



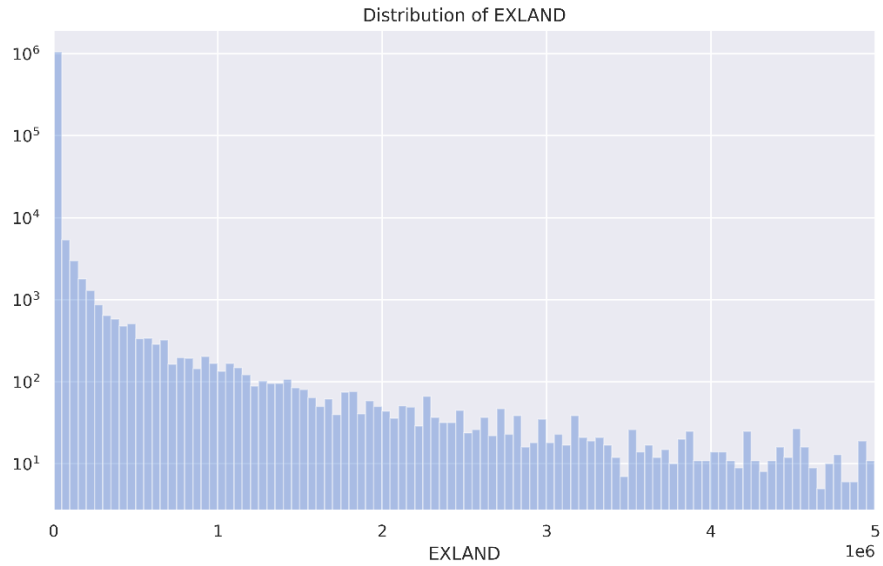
16. AVTOT:

a) Type: Numerical
b) Description:
   i. Actual Total value.
   ii. In addition, to get a closer look, I limited the AVTOT to be under 100000.



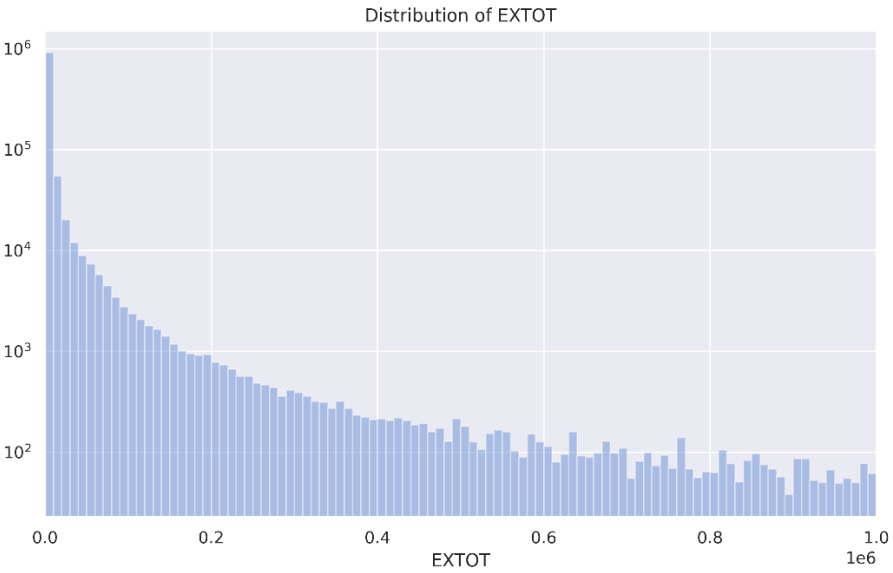
17. EXLAND:

a) Type: Numerical
b) Description:
   i. Actual exempt land value.
   ii. I took log on y-axis to make the graph more interpretable.
   iii. To get a refined look, I limited the EXLAND to be under 5000000.



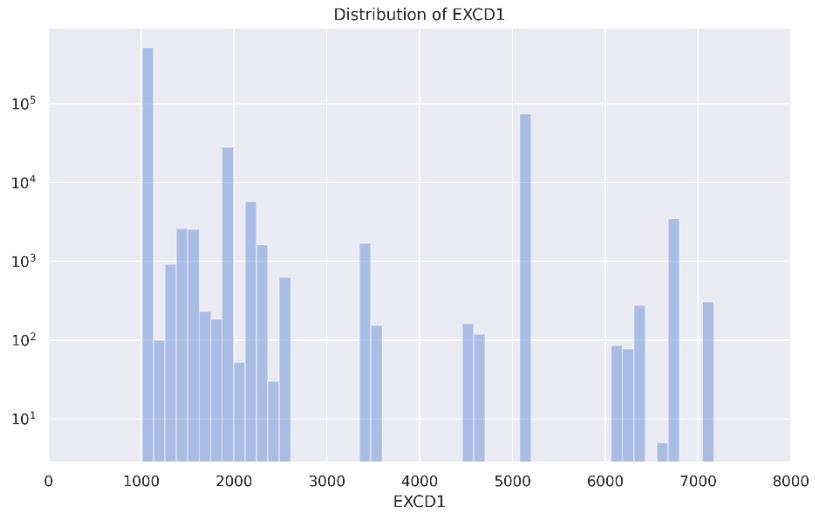
18. EXTOT:

a) Type: Numerical
b) Description:
   i. Actual exempt land total.
   ii. I took log on y-axis to make the graph more interpretable.
   iii. To get a refined look, I limited the EXTOT to be under 1000000.



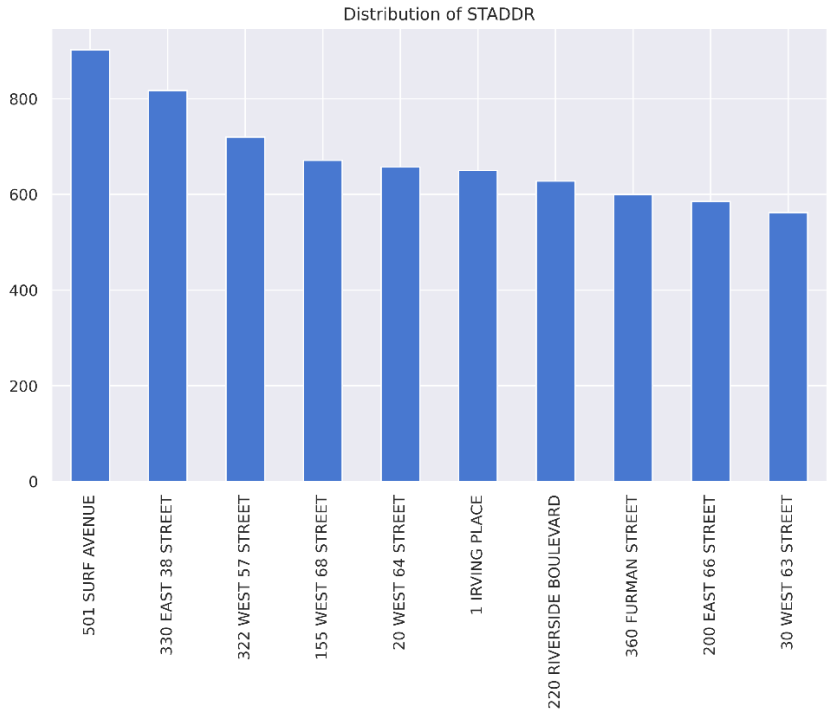
19. EXCD1:

a) Type: Numerical
b) Description:
   i. Exemption code 1
   ii. I took log on y-axis to make the graph more interpretable.



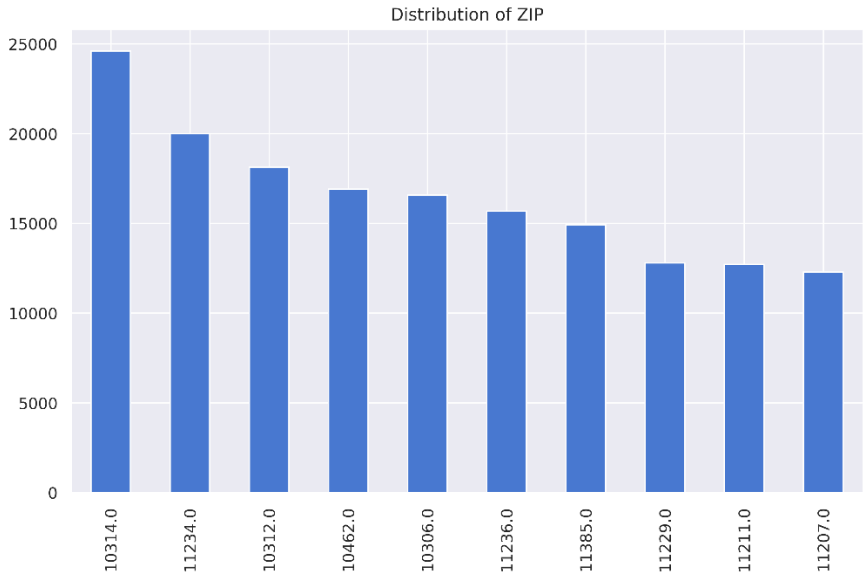
20. STADDR:

a) Type: Categorical
b) Description:
   i. Street addresses.
   ii. Below graph shows the top 10 addresses that appear the most in the dataset.



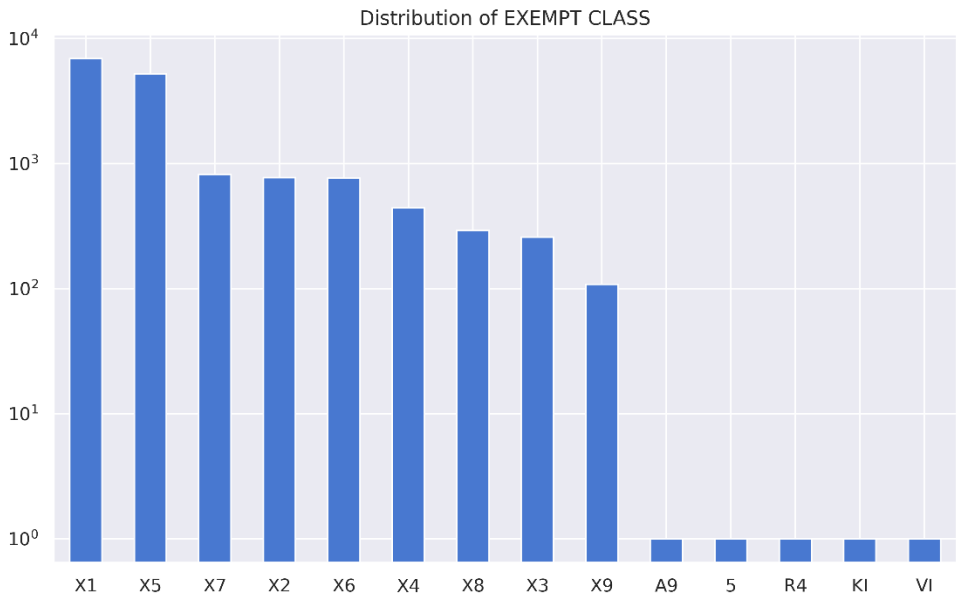
21. ZIP:

a) Type: Categorical
b) Description:
   i. Postal Zip code of the property.
   ii. Although zip code has a data type of float65, it is a categorical rather than a numerical field. Hence, I use count plot rather than distplot here. Below graph shows the top 10 Zip code that appear the most in the dataset.



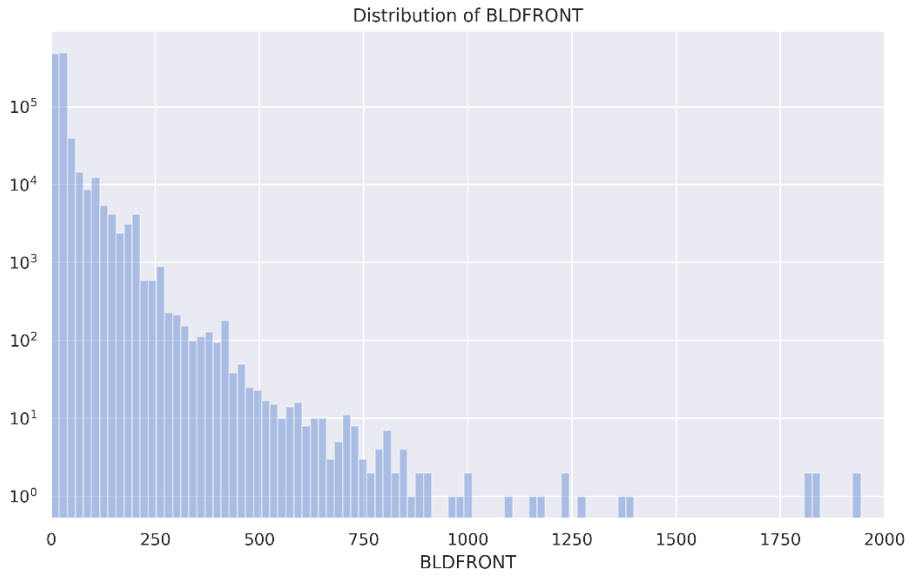
22. EXMPTCL:

a) Type: Categorical
b) Description:
   i. Exempt Class ('X1 - X9') used for fully exempt properties only.



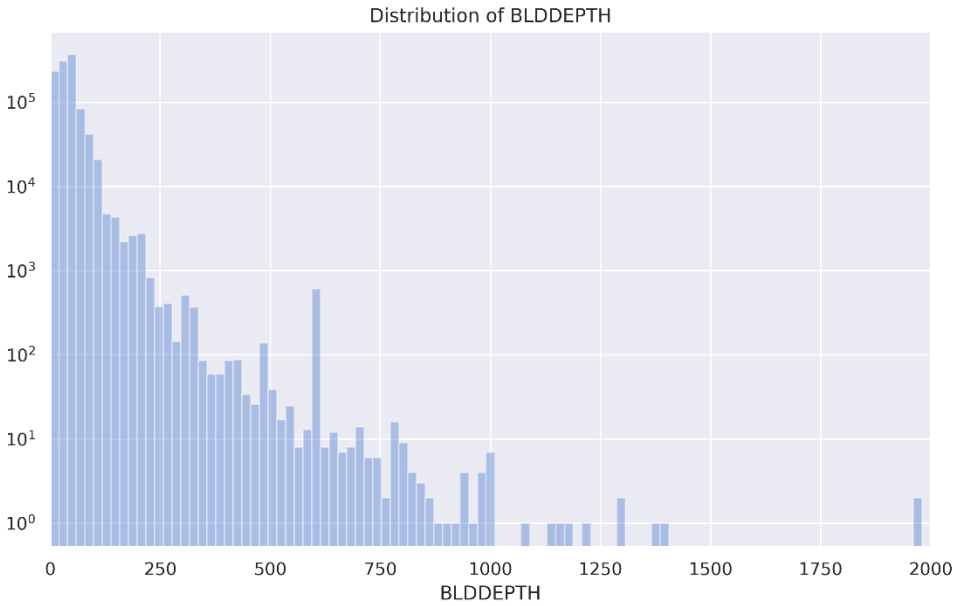
23. BLDFRONT:

a) Type: Numerical
b) Description:
   i. Building Frontage (width) in feet. I took log on y-axis to make the graph more interpretable.
   ii. To get a closer look, I limited the BLDFRONT to be under 2000.



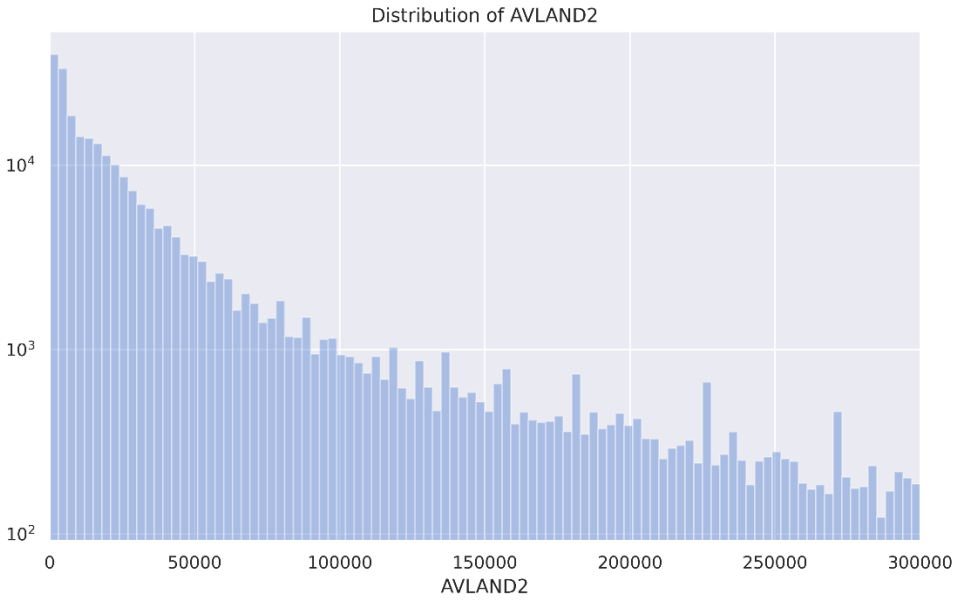
24. BLDDEPTH:

a) Type: Numerical
b) Description:
   i. Lot Depth in feet. I took log on y-axis to make the graph more interpretable.
   ii. To get a closer look, I limited the BLDDEPTH to be under 2000.



25. AVLAND2:

a) Type: Numerical
b) Description:
   i. Transitional land value. I took log on y-axis to make the graph more interpretable. Limited the AVLAND2 to be under 3000000

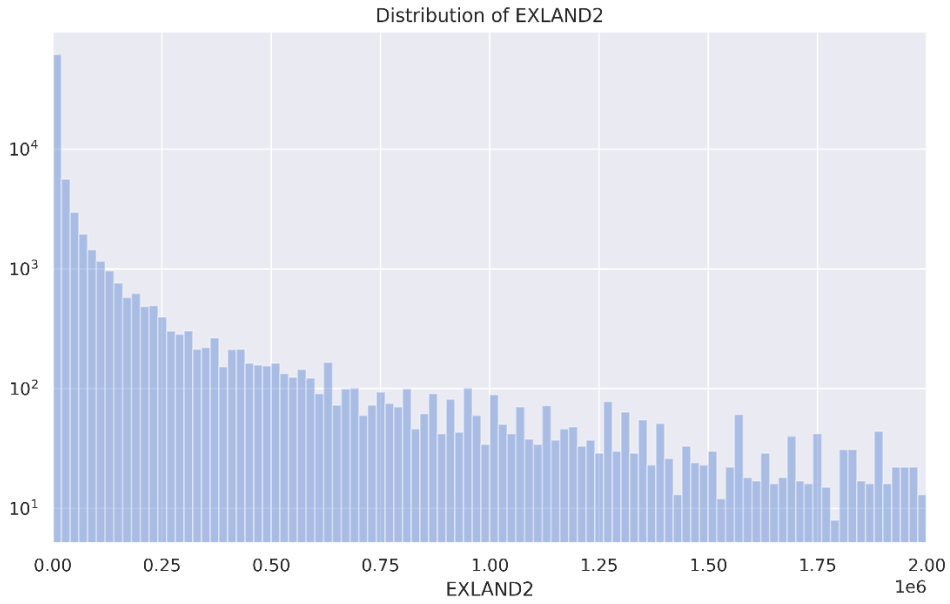


26. AVLAND2:

a) Type: Numerical
b) Description:
   i. Transitional total value. I took log on y-axis to make the graph more interpretable. Limited the AVLAND2 to be under 1000000

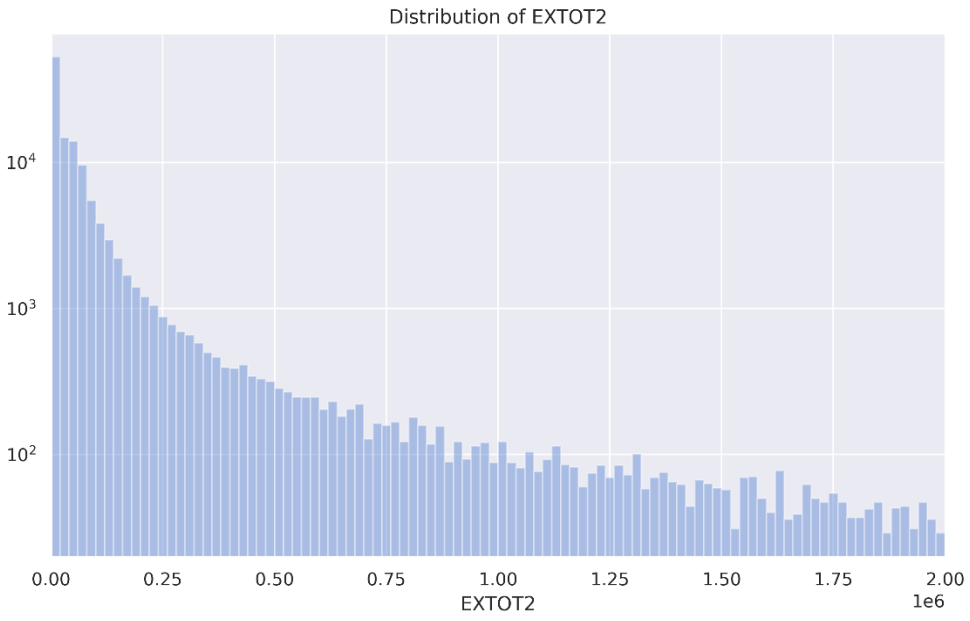
27. EXLAND2:

a) Type: Numerical
b) Description:
   i. Transitional exempt land value. I took log on y-axis to make the graph more interpretable. Limited the EXLAND2 to be under 2000000



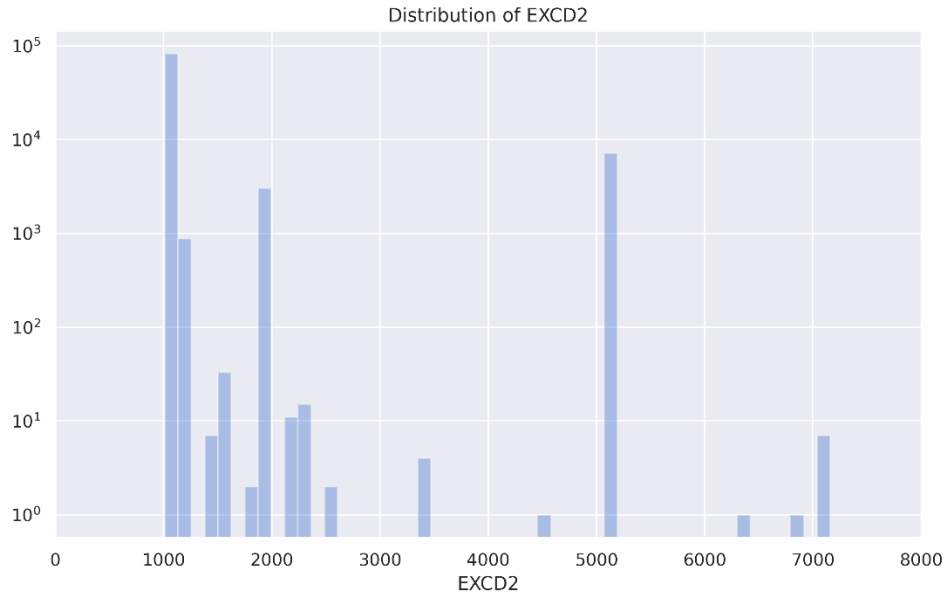
28. EXTOT2:

a) Type: Numerical
b) Description:
   i. Transitional exempt land total. I took log on y-axis to make the graph more interpretable. Limited the EXTOT2 to be under 2000000



29. EXCD2:

a) Type: Numerical
b) Description:
   i. Exemption Code-2. I took log on y-axis to make the graph more interpretable. Limited the EXCD2 to be under 8000



30. PERIOD:

a) Description:
   i. Assessment period when file was created.
   ii. Only one unique value - Final
   iii. Nothing major to interpret with a graph. Skipped.

31. YEAR:

b) Description:
   i. Assessment year.
   ii. Only one unique value – 2010/2011
   iii. Nothing major to interpret with a graph. Skipped.

32. VALTYPE:

c) Description:
   i. Need to clarify the definition of this field with the client.
   ii. Only one unique value – AC-TR
   iii. Nothing major to interpret with a graph. Skipped.