Data Quality Report
Abhisheik Jadhav
MGTA463- Fraud Analytics

1. Data Description

The dataset "Property Valuation and Assessment Data," comprises **1,070,994 records** detailing real estate assessment property data for the **year 2010/11**. It encompasses **32 fields** and is updated annually by the Department of Finance. This dataset provides an extensive collection of information related to property assessments, evaluations, and condensed roll data within the city government category. Key fields include property ID, assessment value, property type, location, and other relevant property-specific information.

2. Summary Tables:

Numeric:

	Field Name	Field Type	# Records Have Values	% Populated	# Zeros	Min	Max	Mean	Standard Deviation	Most Common
0	LTFRONT	numeric	1,070,994	100.0%	169,108	0.00	9999.00	36.64	74.03	0.00
1	LTDEPTH	numeric	1,070,994	100.0%	170,128	0.00	9999.00	88.86	76.40	100.00
2	STORIES	numeric	1,014,730	94.7%	0	1.00	119.00	5.01	8.37	2.00
3	FULLVAL	numeric	1,070,994	100.0%	13,007	0.00	6150000000.00	874264.51	11582425.58	0.00
4	AVLAND	numeric	1,070,994	100.0%	13,009	0.00	2668500000.00	85067.92	4057258.16	0.00
5	AVTOT	numeric	1,070,994	100.0%	13,007	0.00	4668308947.00	227238.17	6877526.09	0.00
6	EXLAND	numeric	1,070,994	100.0%	491,699	0.00	2668500000.00	36423.89	3981573.93	0.00
7	EXTOT	numeric	1,070,994	100.0%	432,572	0.00	4668308947.00	91186.98	6508399.78	0.00
8	BLDFRONT	numeric	1,070,994	100.0%	228,815	0.00	7575.00	23.04	35.58	0.00
9	BLDDEPTH	numeric	1,070,994	100.0%	228,853	0.00	9393.00	39.92	42.71	0.00
10	AVLAND2	numeric	282,726	26.4%	0	3.00	2371005000.00	246235.72	6178951.64	2408.00
11	AVTOT2	numeric	282,732	26.4%	0	3.00	4501180002.00	713911.44	11652508.34	750.00
12	EXLAND2	numeric	87,449	8.2%	0	1.00	2371005000.00	351235.68	10802150.91	2090.00
13	EXTOT2	numeric	130,828	12.2%	0	7.00	4501180002.00	656768.28	16072448.75	2090.00

Categorical:

	Field Name	Field Type	# Records Have Values	% Populated	# Zeros	# Unique Values	Most Common
0	RECORD	categorical	1,070,994	100.0%	0	1,070,994	1
1	BBLE	categorical	1,070,994	100.0%	0	1,070,994	1000010101
2	BORO	categorical	1,070,994	100.0%	0	5	4
3	BLOCK	categorical	1,070,994	100.0%	0	13,984	3944
4	LOT	categorical	1,070,994	100.0%	0	6,366	1
5	EASEMENT	categorical	4,636	0.4%	0	12	E
6	OWNER	categorical	1,039,249	97.0%	0	863,347	PARKCHESTER PRESERVAT
7	BLDGCL	categorical	1,070,994	100.0%	0	200	R4
8	TAXCLASS	categorical	1,070,994	100.0%	0	11	1
9	EXT	categorical	354,305	33.1%	0	3	G
10	EXCD1	categorical	638,488	59.6%	0	129	1017.00
11	STADDR	categorical	1,070,318	99.9%	0	839,280	501 SURF AVENUE
12	ZIP	categorical	1,041,104	97.2%	0	196	10314.00
13	EXMPTCL	categorical	15,579	1.5%	0	14	X1
14	EXCD2	categorical	92,948	8.7%	0	60	1017.00
15	PERIOD	categorical	1,070,994	100.0%	0	1	FINAL
16	YEAR	categorical	1,070,994	100.0%	0	1	2010/11
17	VALTYPE	categorical	1,070,994	100.0%	0	1	AC-TR

3. Fields:

RECORD:

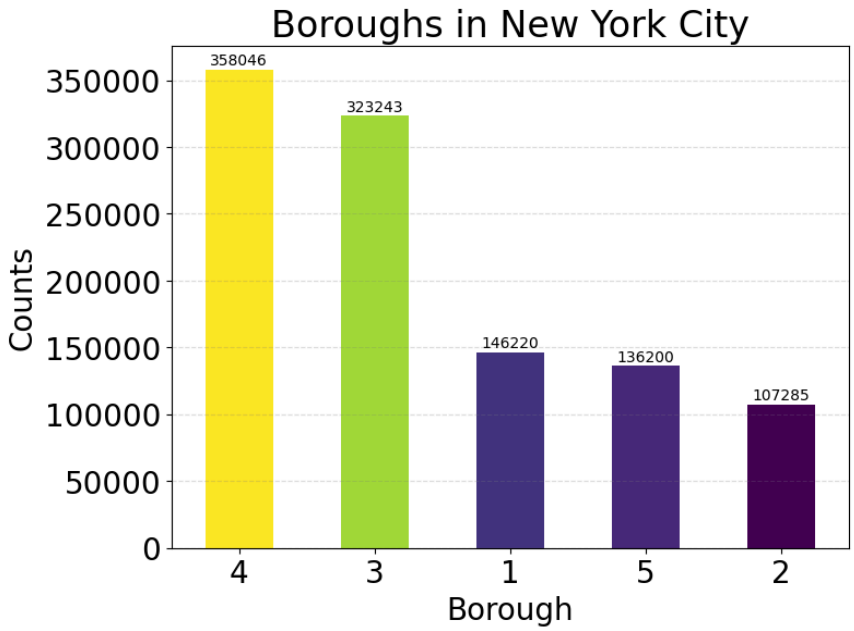
The "RECORD" field is a categorical field that uniquely identifies each of the 1,070,994 records in the dataset, with 100% population and no duplicate values.

BBLE:

The "BBLE" field is a categorical field that uniquely identifies each record as a file key, with all 1,070,994 records populated and no duplicate values.

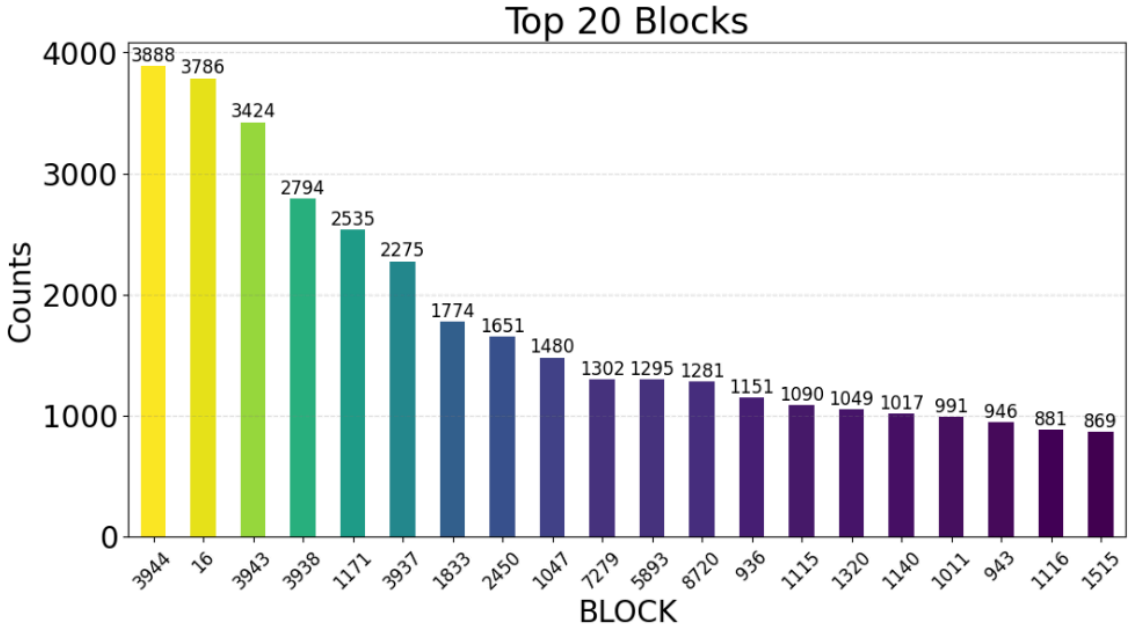
BORO:

The "BORO" field is a categorical field representing the boroughs of NYC, with all 1,070,994 records populated and no zeros. There are 5 unique values, with the most common being "4." The following graph shows the visualization of boroughs in NYC.



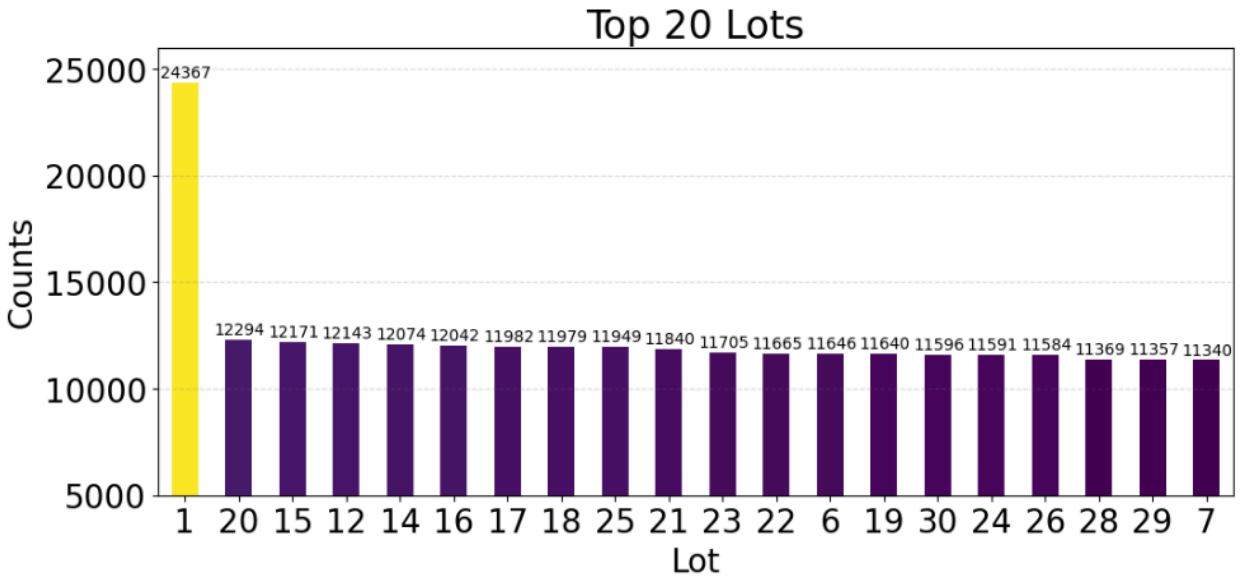
BLOCK:

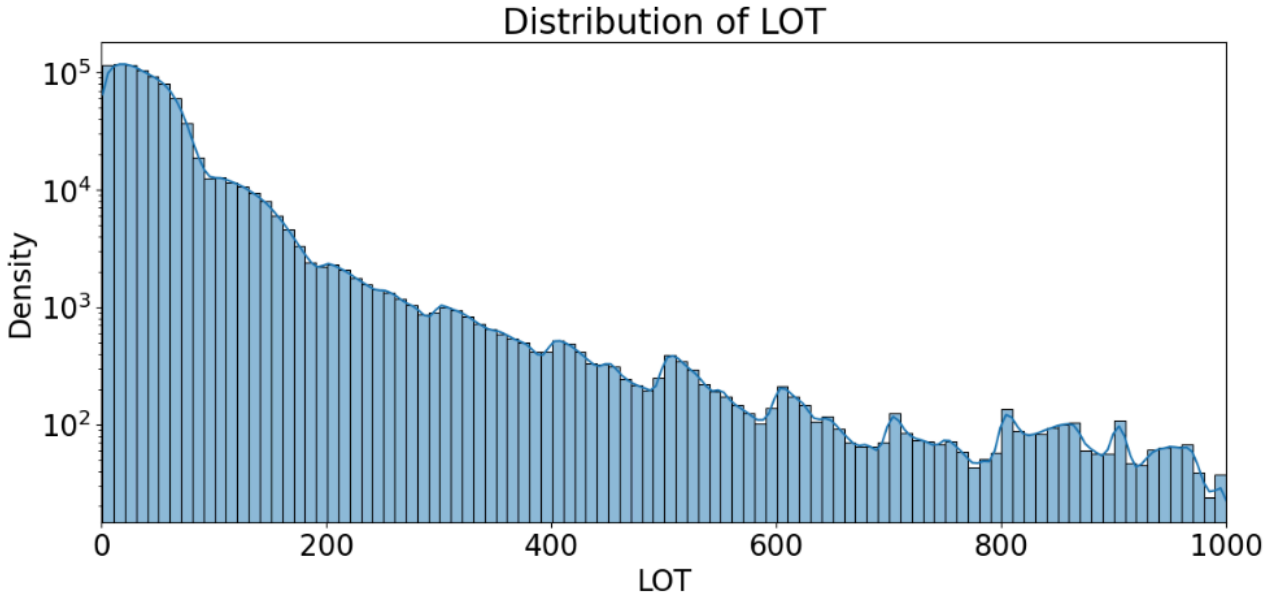
The "BLOCK" field is a categorical field representing valid block ranges by borough, with all 1,070,994 records populated and no zeros. There are 13,984 unique values, with the most common being "3944." The following graph visualizes the top 20 blocks.



LOT:

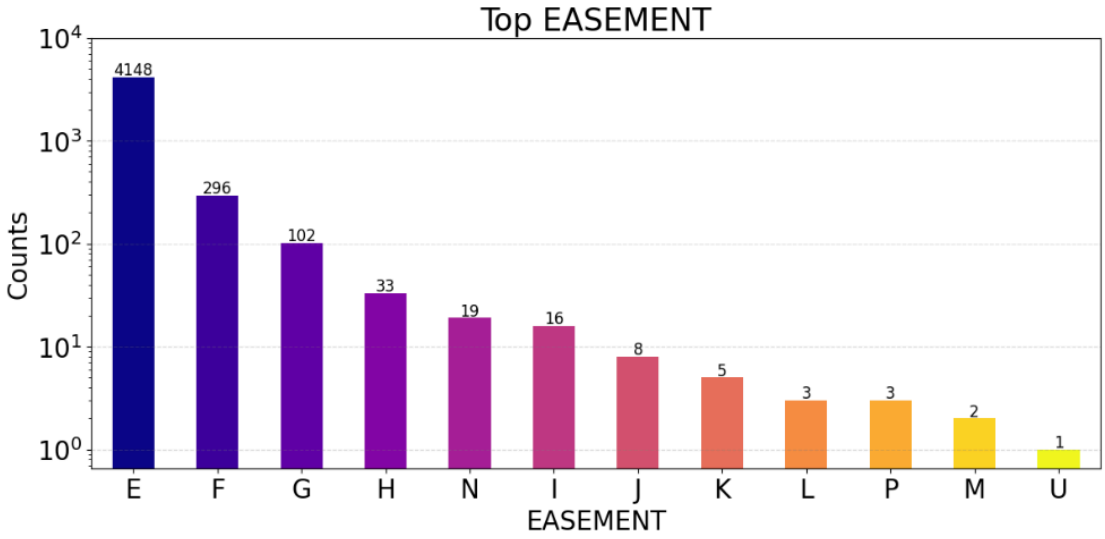
The "LOT" field, a categorical identifier with 100% coverage across 1,070,994 records and zero missing values, features 6,366 distinct categories. Notably, the most prevalent category is "1." Additionally, visualizations depicting the top 20 lots and the distribution of lots across the dataset are available for analysis.





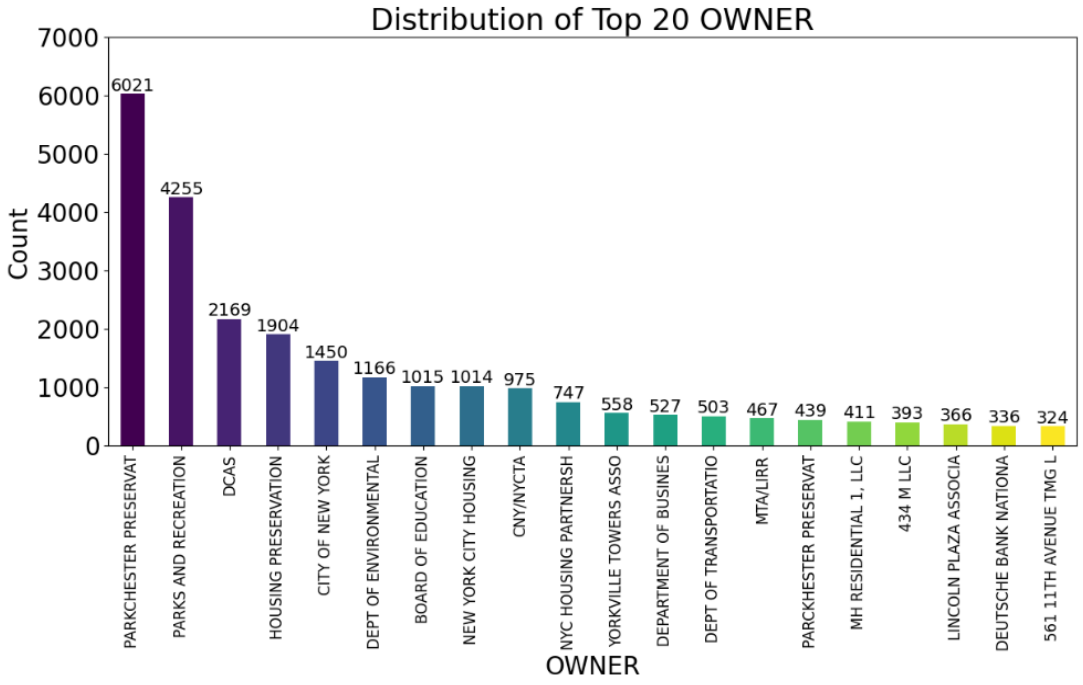
EASEMENT:

The "EASEMENT" field is categorical, present in 0.4% of the 1,070,994 records, with no instances of zero values. It comprises 12 unique categories, with "E" being the most common. An accompanying visualization showcases the distribution of the top easements within the dataset.



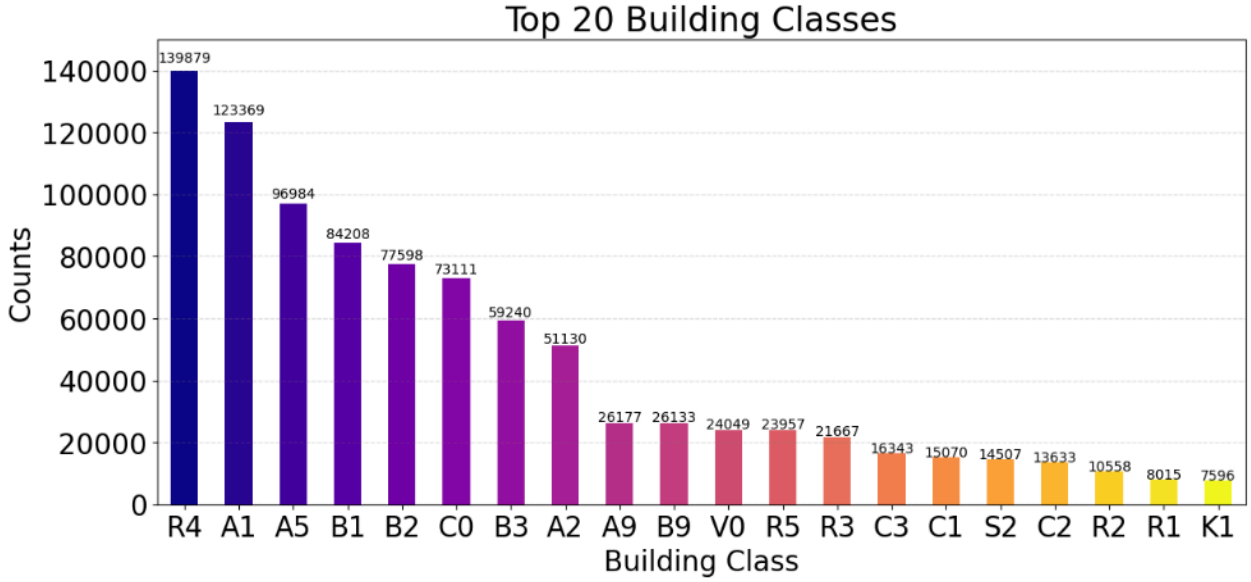
OWNER:

The "OWNER" field, capturing categorical data representing owner names, is prevalent in 97.0% of the 1,070,994 records, with no zero entries. It encompasses a wide diversity of 863,347 unique owners, with "PARKCHESTER PRESERVAT" being the most frequent. A visual representation highlights the top 20 owners for further analysis.



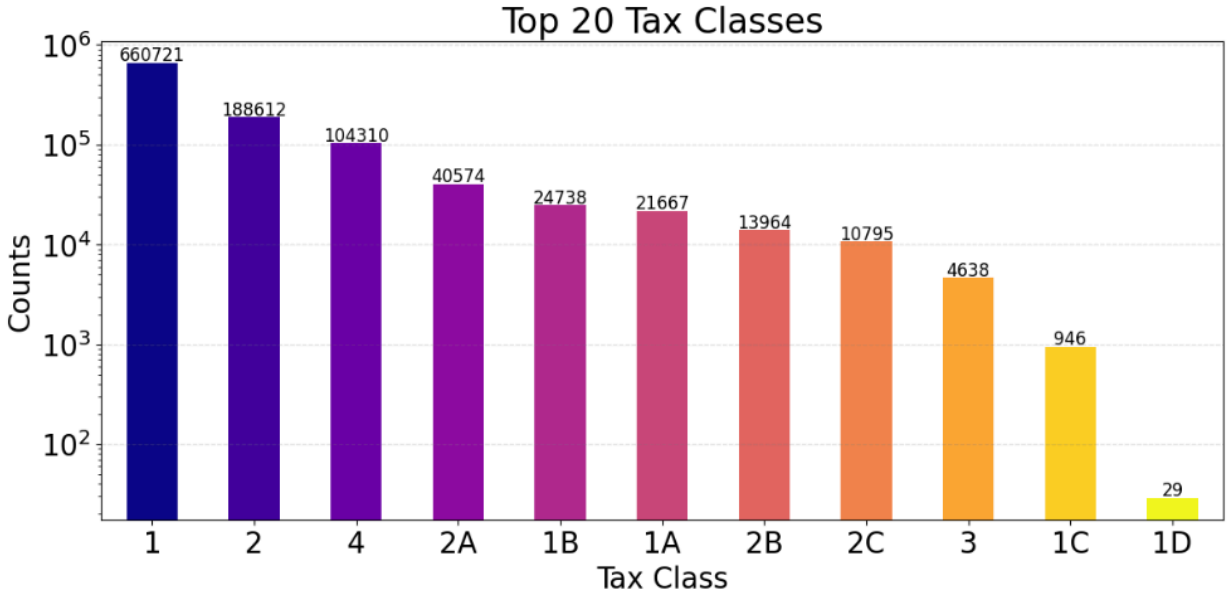
BLDGCL:

The "BLDGCL" field, categorical and present in all 1,070,994 records without any zeros, represents building classes. With 200 unique categories, "R4" emerges as the most common class. A visual overview displays the top 20 building classes.



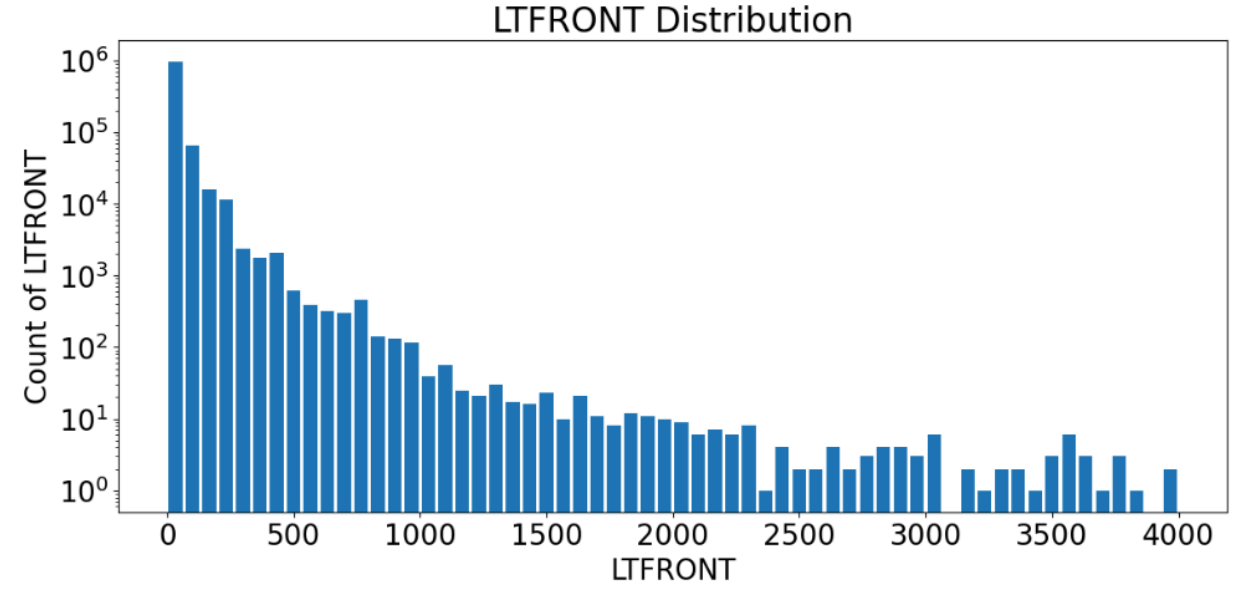
TAXCLASS:

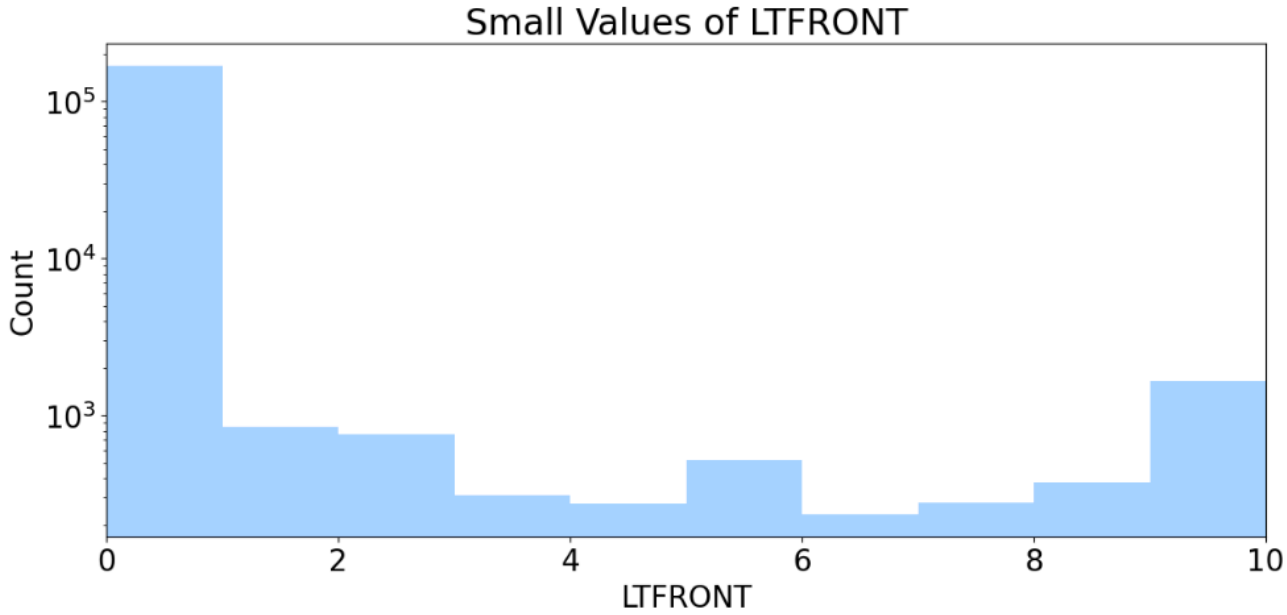
The "TAXCLASS" field, categorical and fully populated across all 1,070,994 records without any zeros, categorizes tax classes. With 11 unique classes, "1" is the most prevalent. A visual representation illustrates the distribution of the top 20 tax classes.



LTFRONT:

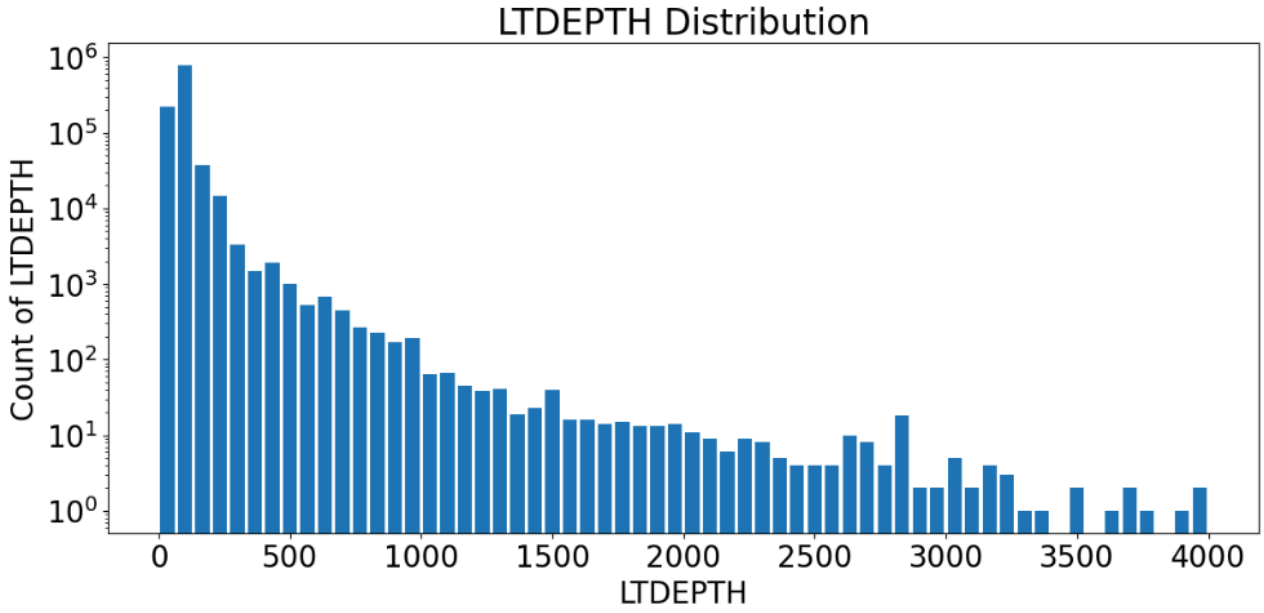
The "LTFRONT" field, a numeric attribute, is fully populated across 1,070,994 records. It denotes lot width, ranging from 0.00 to 9999.00, with an average width of 36.64 and a median of 74.03. Visualizations include distributions of widths and a focused one on smaller values.

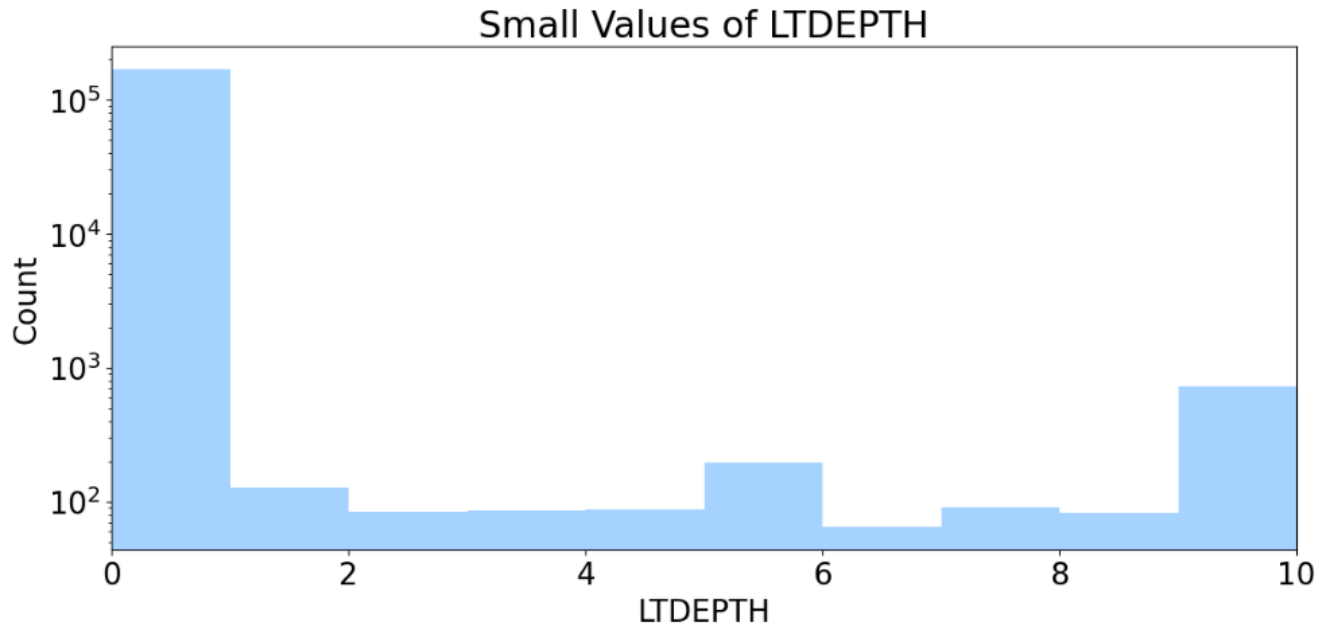




LTDEPTH:

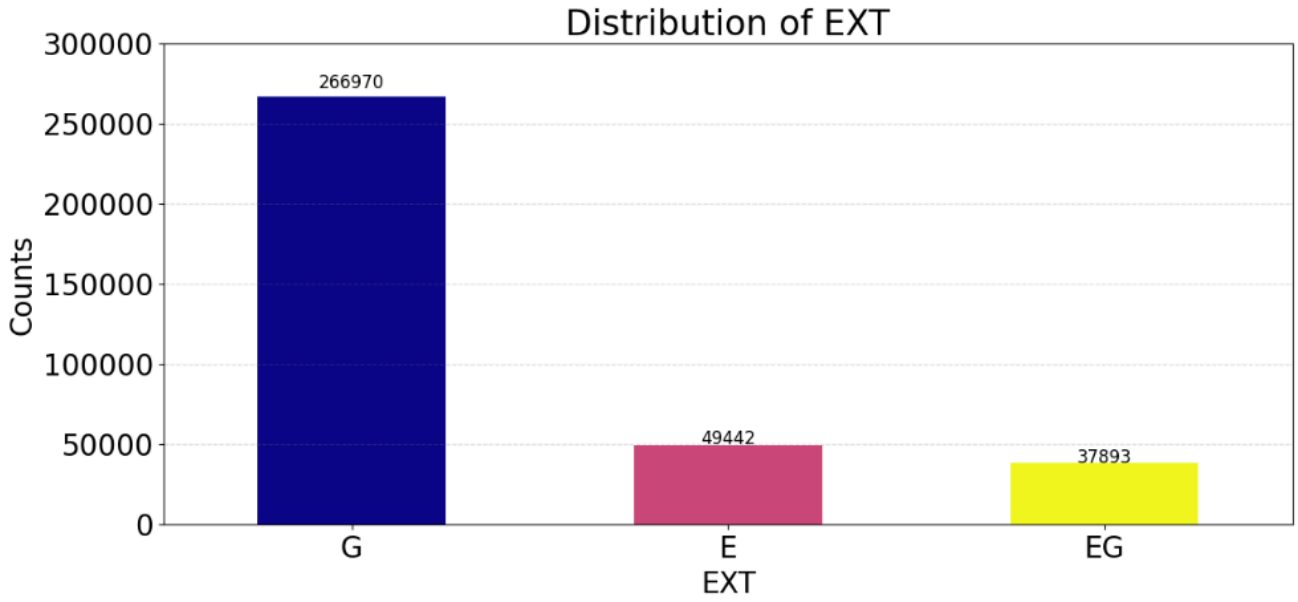
The "LTDEPTH" field, a numeric attribute, is fully populated across 1,070,994 records. It represents lot depth, ranging from 0.00 to 9999.00, with an average depth of 88.86 and a median of 76.40. Visualizations encompass distributions of depths and focus on the distribution of smaller values.





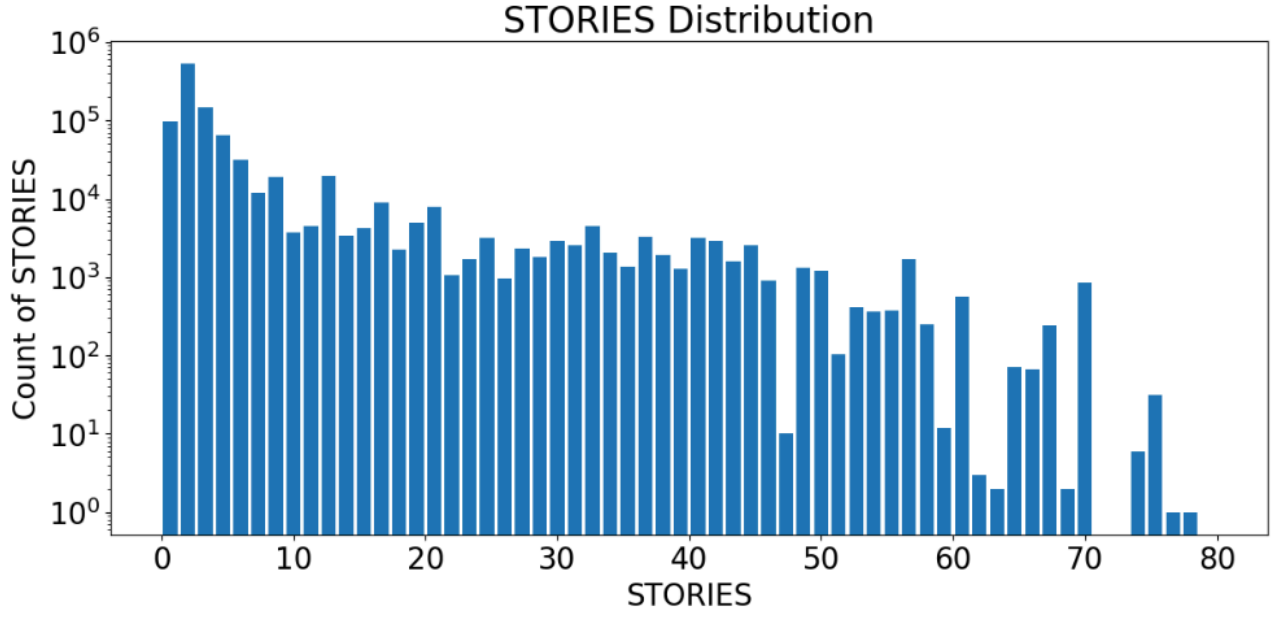
EXT:

The "EXT" field, categorical with 33.1% coverage across 354,305 records, denotes extension indicators with three unique categories. The visualization includes a distribution analysis of these extension indicators.



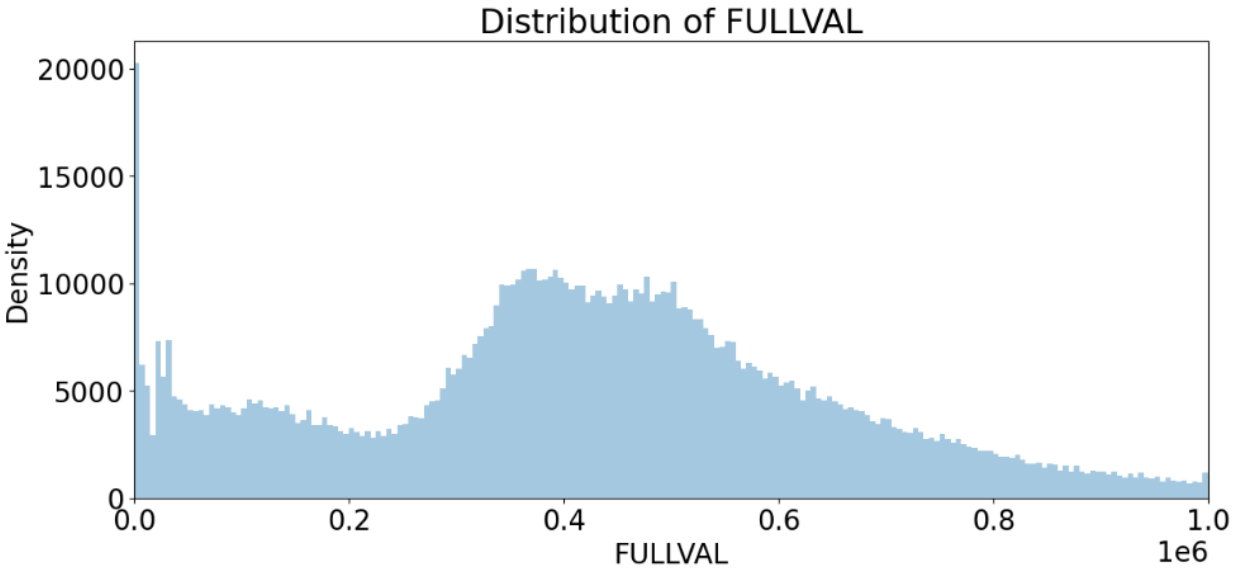
STORIES:

The "STORIES" field, a numeric attribute, is present in 94.7% of the 1,014,730 records, indicating the number of stories in a building. With values ranging from 1.00 to 119.00, the average number of stories is 5.01, with a median of 8.37. Visualizations encompass the distribution of stories within buildings.



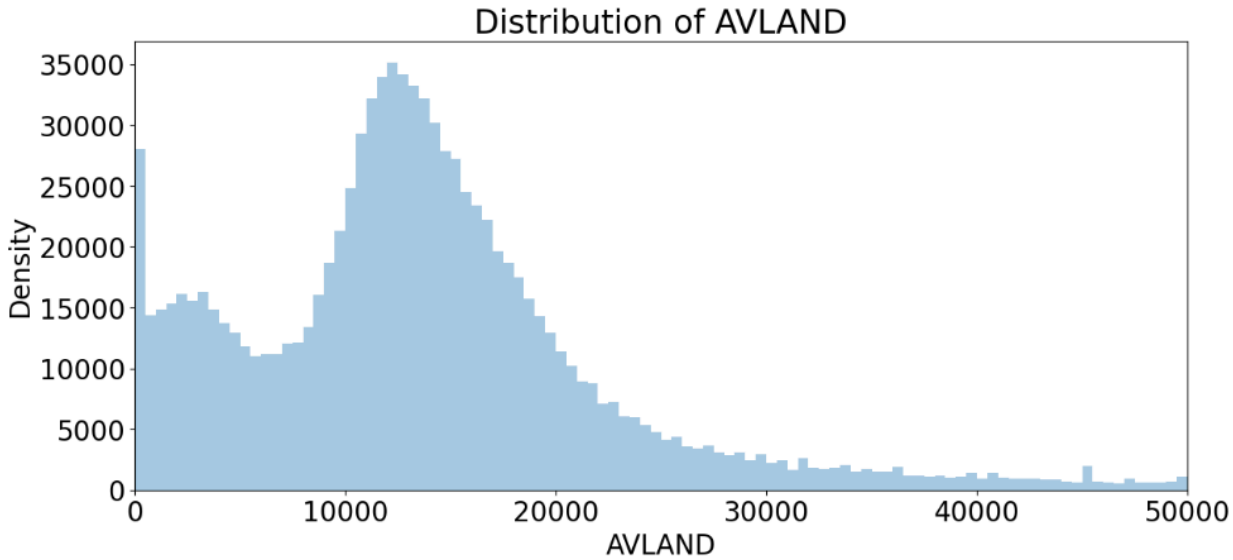
FULLVAL:

The "FULLVAL" field, a numeric attribute, is fully populated across 1,070,994 records, representing the market value. The values range from 0.00 to 6,150,000,000.00, with an average value of $874,264.51 and a median of $11,582,425.58. Visualizations showcase the distribution of market values (FULLVAL).



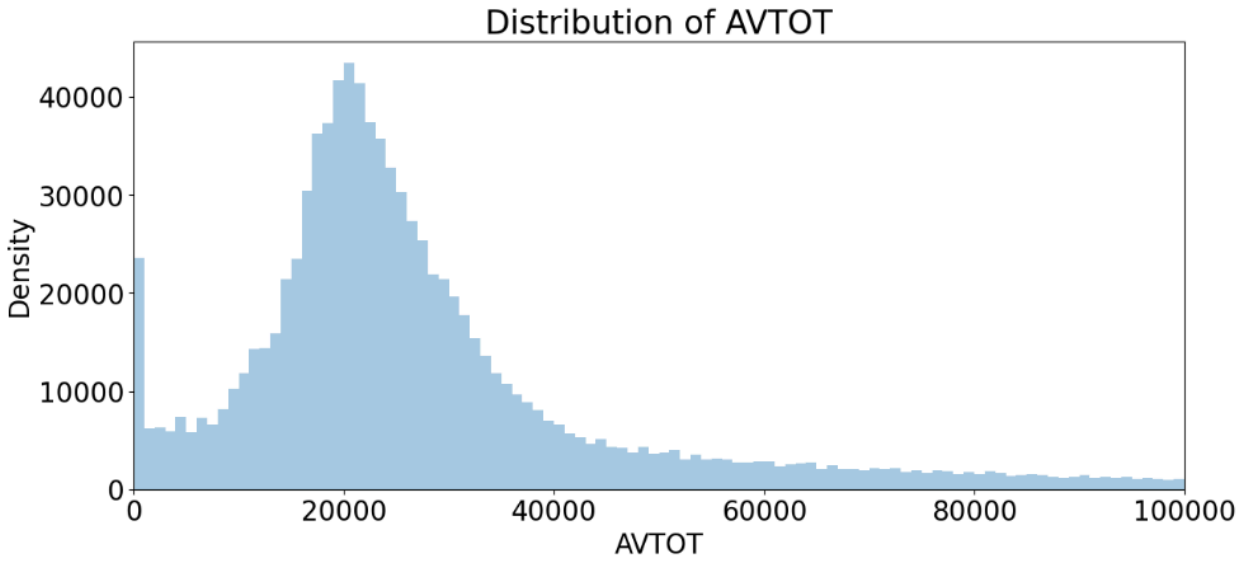
AVLAND:

The "AVLAND" field, a numeric attribute present in all 1,070,994 records, represents the actual land value. Values range from 0.00 to 2,668,500,000.00, with an average value of $85,067.92 and a median of $4,057,258.16. Visualizations illustrate the distribution of actual land values (AVLAND).



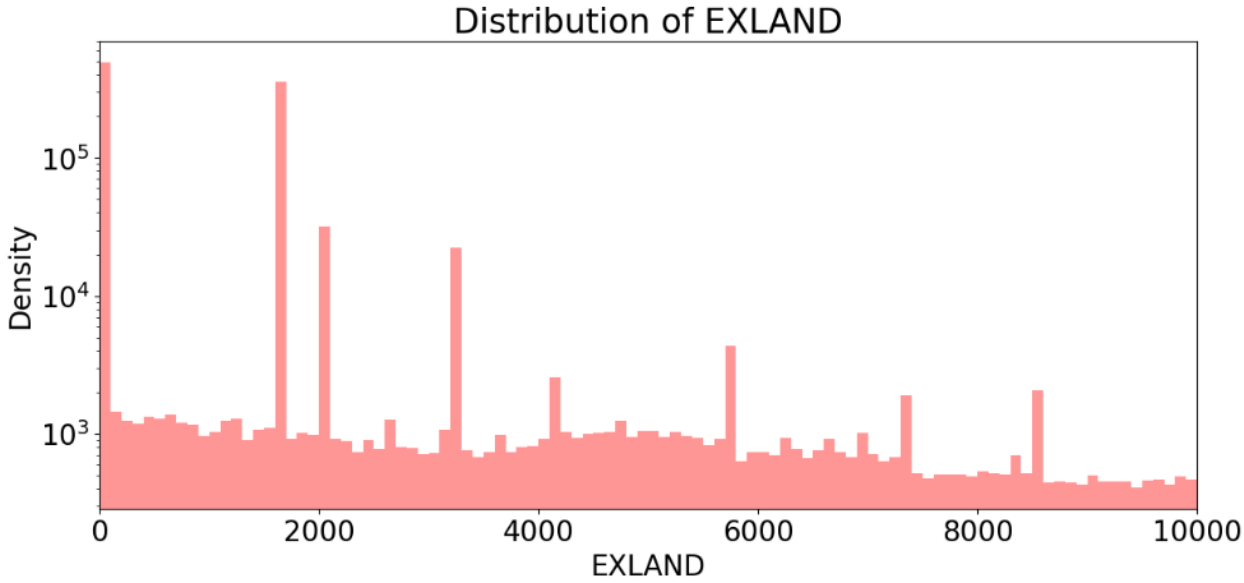
AVTOT:

The "AVTOT" field, a numeric attribute present in all 1,070,994 records, represents the actual total value. Values range from 0.00 to 4,668,308,947.00, with an average value of $227,238.17 and a median of $6,877,526.09. Visualizations illustrate the distribution of actual total values (AVTOT).



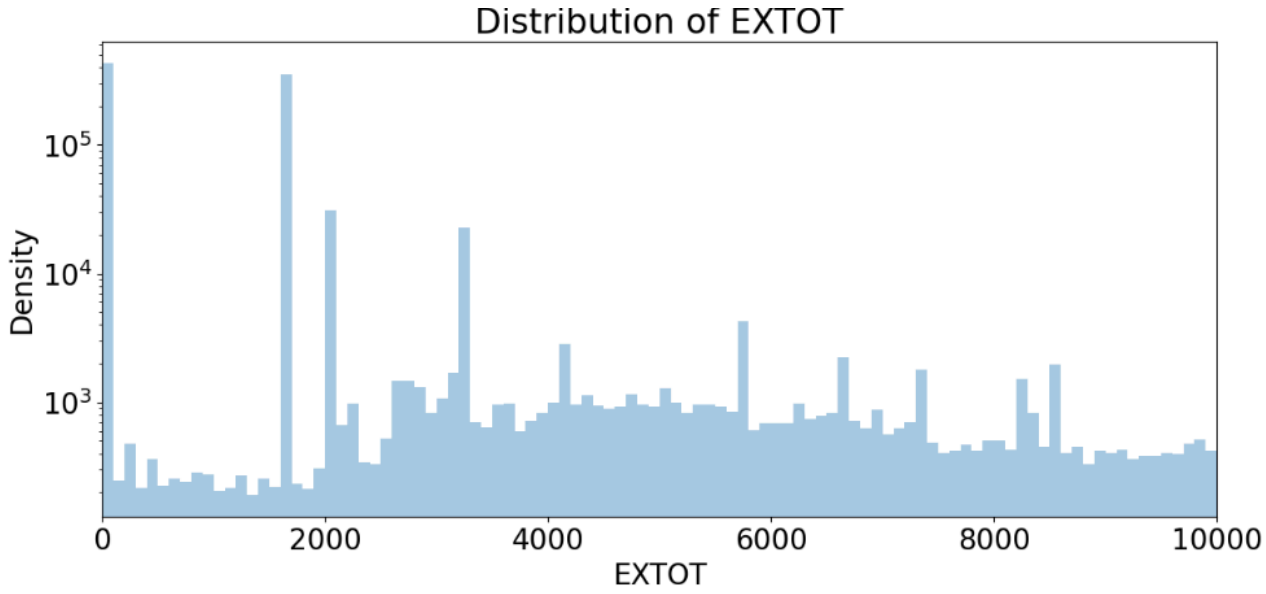
EXLAND:

The "EXLAND" field, a numeric attribute present in all 1,070,994 records, represents the actual exempt land value. Values range from 0.00 to 2,668,500,000.00, with an average value of $36,423.89 and a median of $3,981,573.93. Visualizations illustrate the distribution of actual exempt land values (EXLAND).



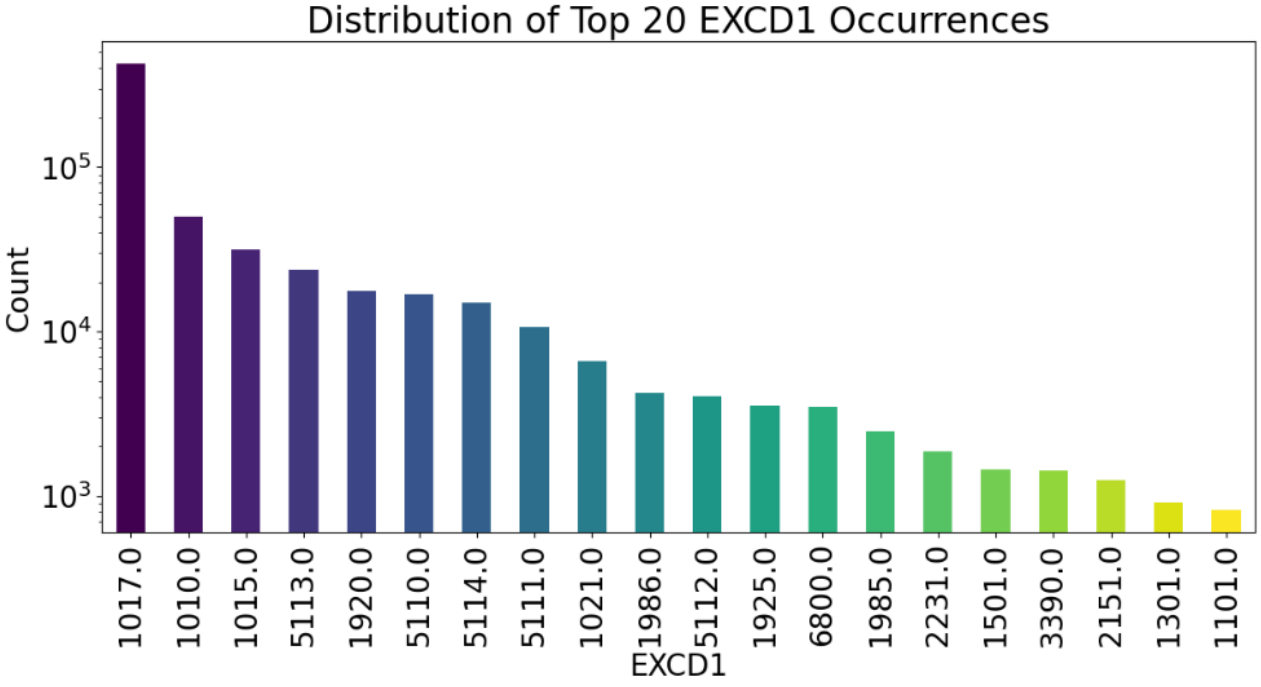
EXTOT:

The "EXTOT" field, a numeric attribute present in all 1,070,994 records, denotes the actual exempt land total. Values range from 0.00 to 4,668,308,947.00, with an average value of $91,186.98 and a median of $6,508,399.78. Visualizations depict the distribution of actual exempt land totals (EXTOT).



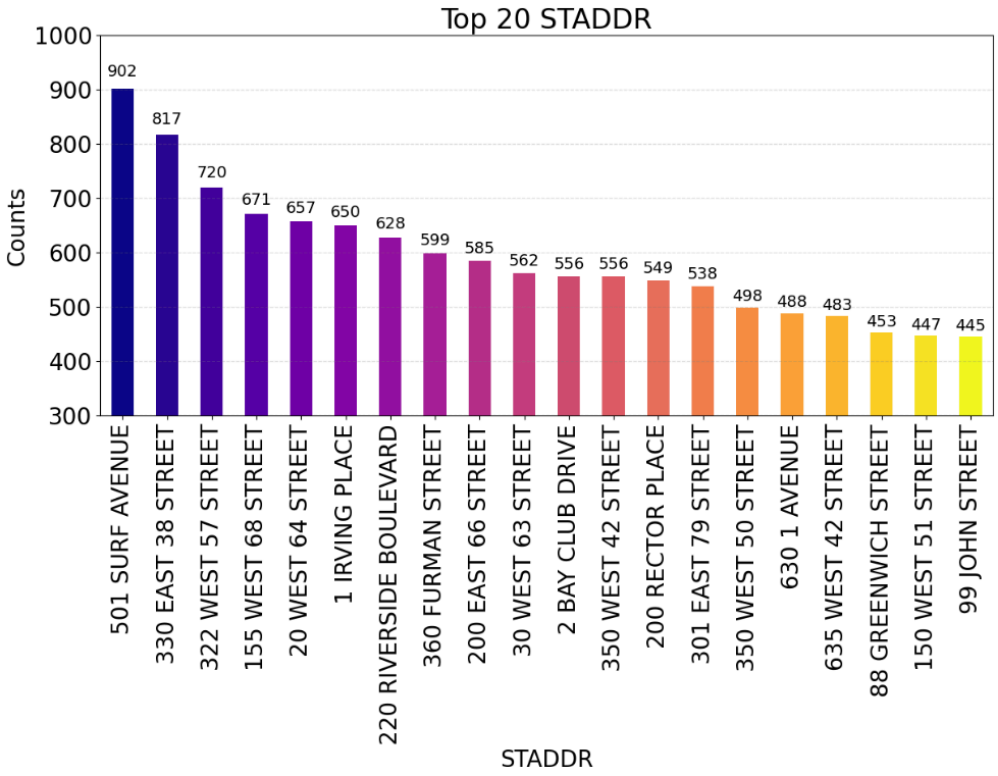
EXCD1:

The "EXCD1" field, a categorical attribute present in 59.6% of the 638,488 records, denotes exemption codes. There are 129 unique codes, with the top 20 occurrences visualized for analysis.



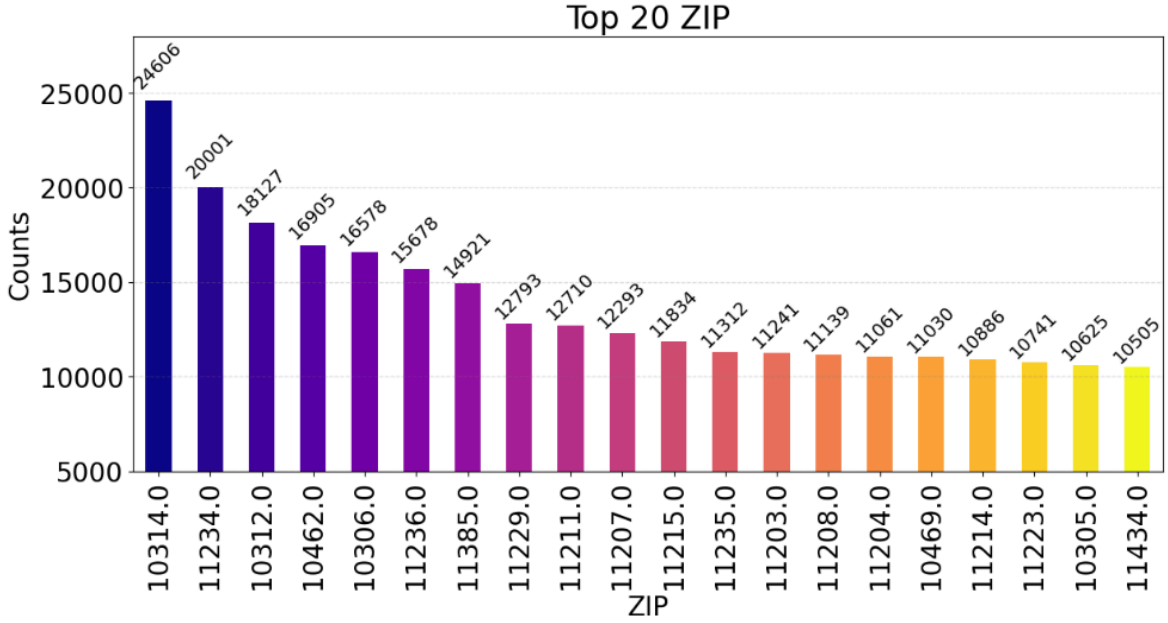
STADDR:

The "STADDR" field, a categorical attribute, is present in 99.9% of the 1,070,318 records, representing street addresses. With 839,280 unique addresses, the top 20 occurrences are visualized for analysis.



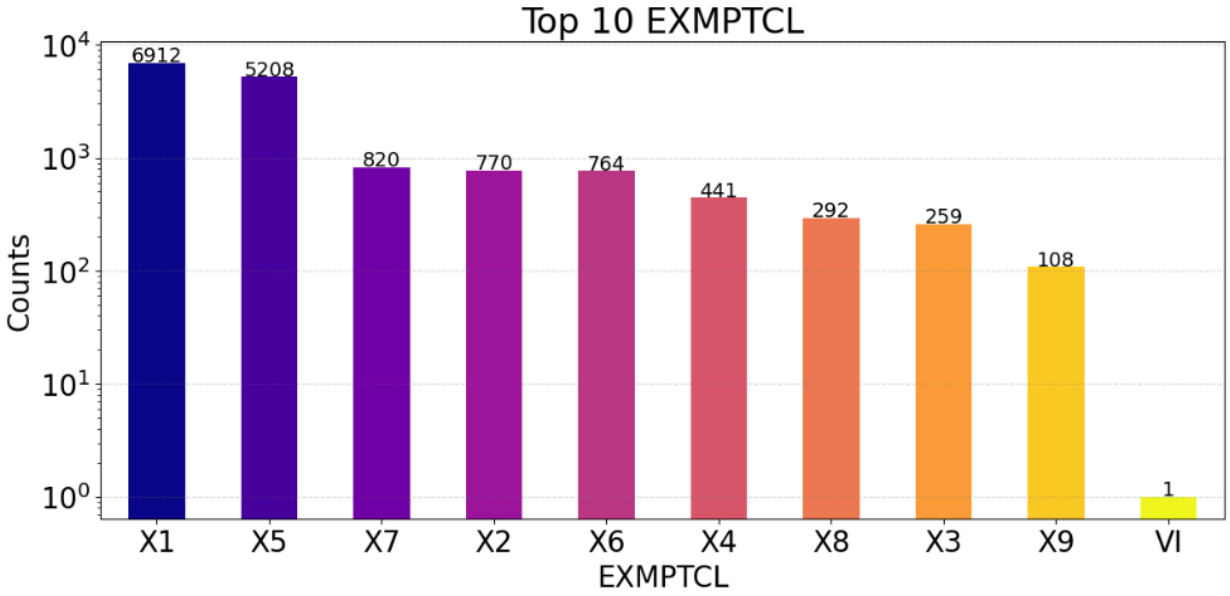
ZIP:

The "ZIP" field, a categorical attribute, is present in 97.2% of the 1,041,104 records, indicating zip codes. With 196 unique zip codes, the top 20 occurrences are visualized for analysis.



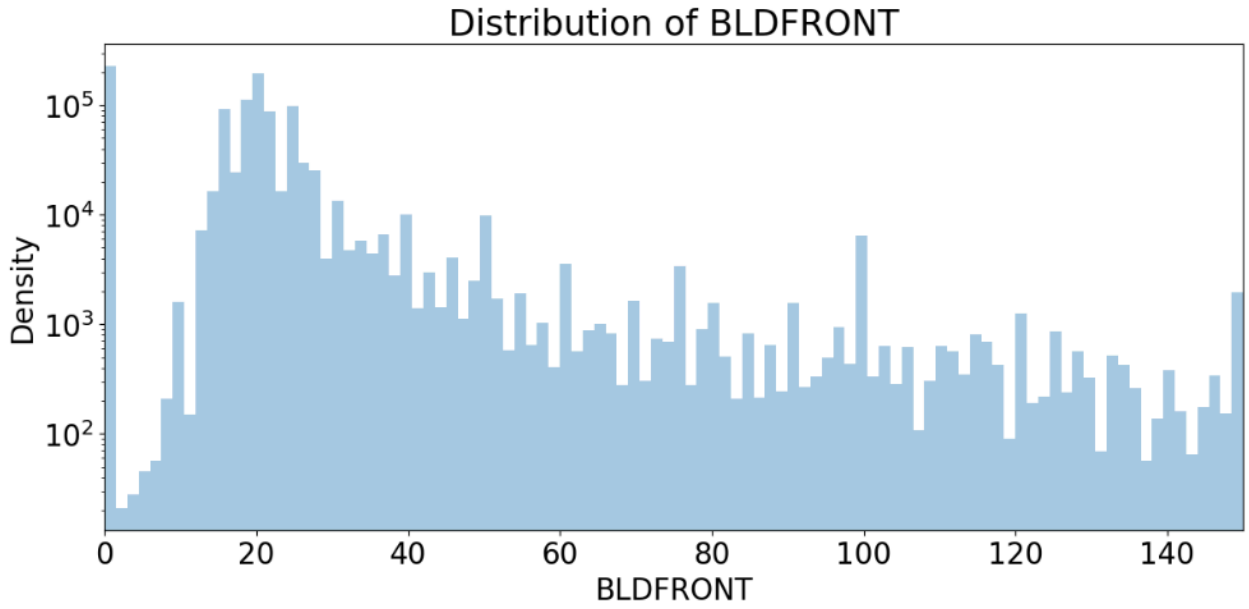
EXMPTCL:

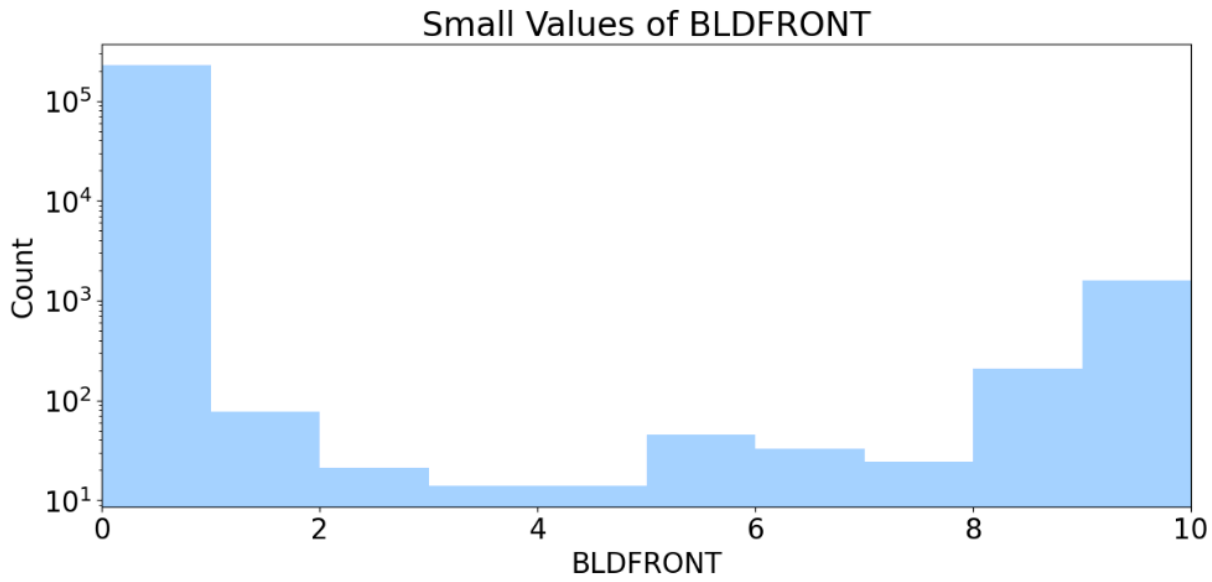
The "EXMPTCL" field, a categorical attribute, is present in 1.5% of the 15,579 records, representing exemption classes. With 14 unique classes, The visual representation showcases the top 10 occurrences for the "EXMPTCL" field.



BLDFRONT:

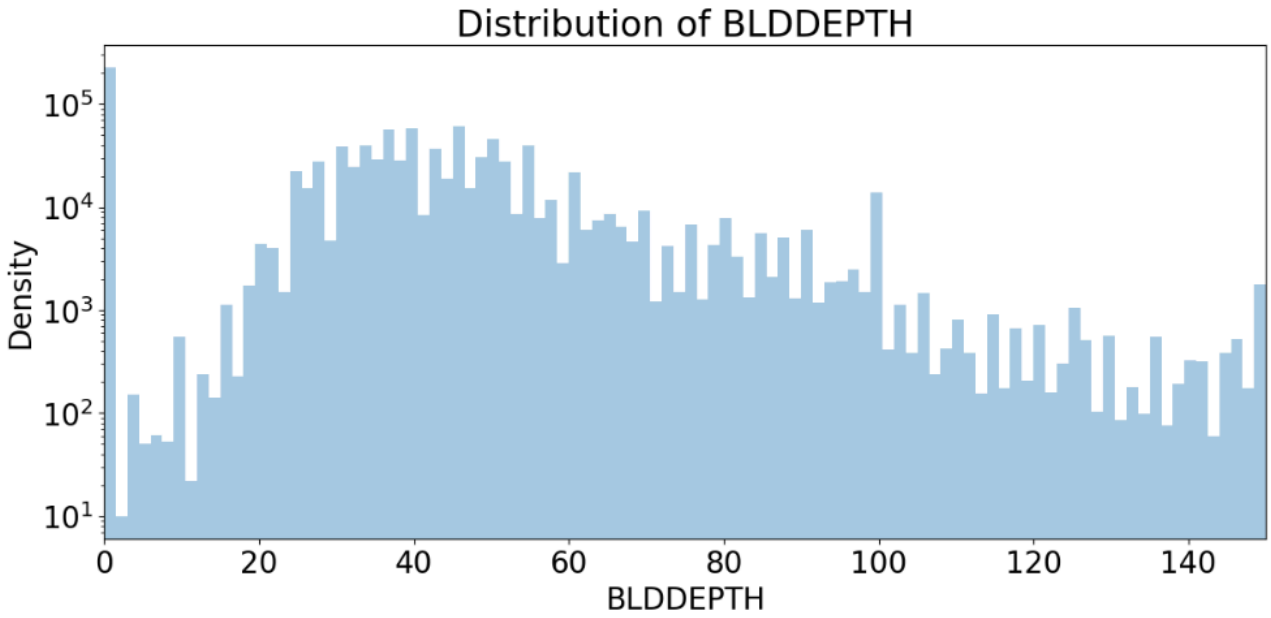
The "BLDFRONT" field, a numeric attribute, is fully populated across 1,070,994 records, indicating the building width. Ranging from 0.00 to 7575.00, with an average width of 23.04 and a median of 35.58, visualizations encompass distributions of building widths (BLDFRONT) and focus on the distribution of smaller values.

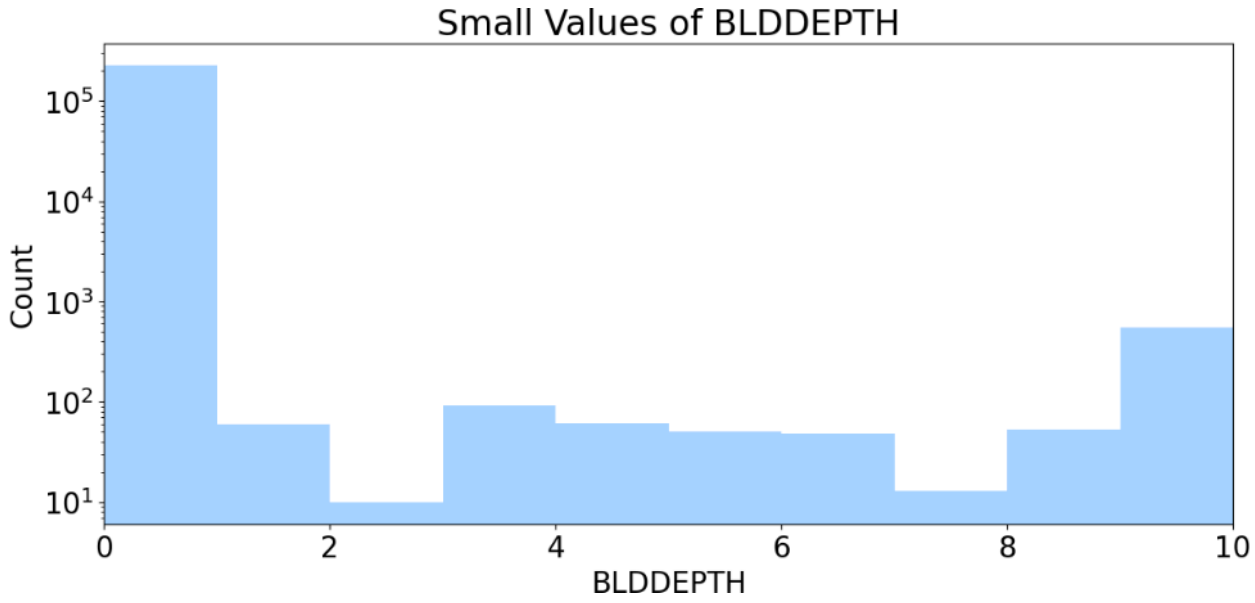




BLDDEPTH:

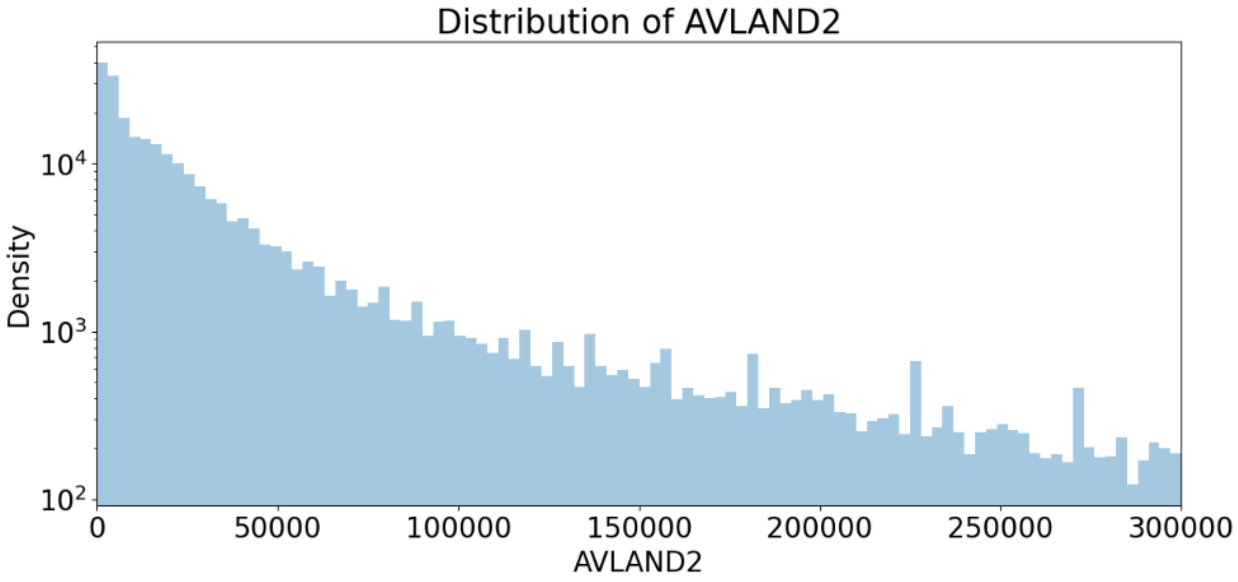
The "BLDDEPTH" field, a numeric attribute, is fully populated across all 1,070,994 records, representing building depth. Ranging from 0.00 to 9393.00, with an average depth of 39.92 and a median of 42.71, visualizations showcase the distribution of building depths (BLDDEPTH) and focus on smaller value distributions.





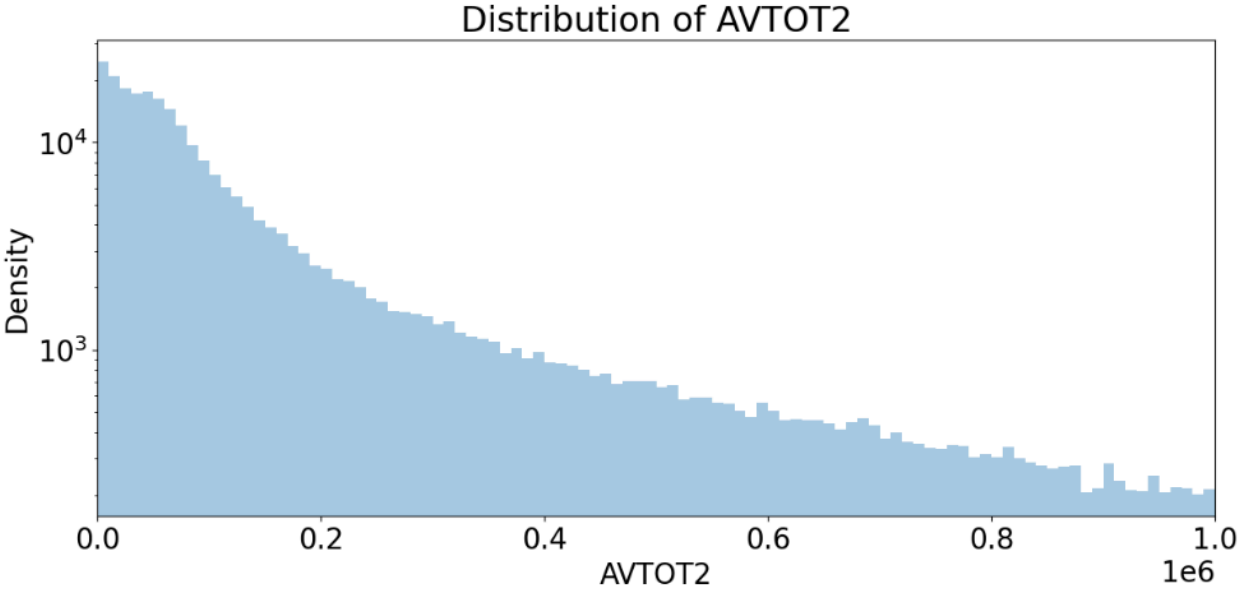
AVLAND2:

The "AVLAND2" field, a numeric attribute present in 26.4% of the 282,726 records, signifies transitional land values. Ranging from 3.00 to 2,371,005,000.00, with an average value of $246,235.72 and a median of $6,178,951.64, the visualizations depict the distribution of transitional land values (AVLAND2).



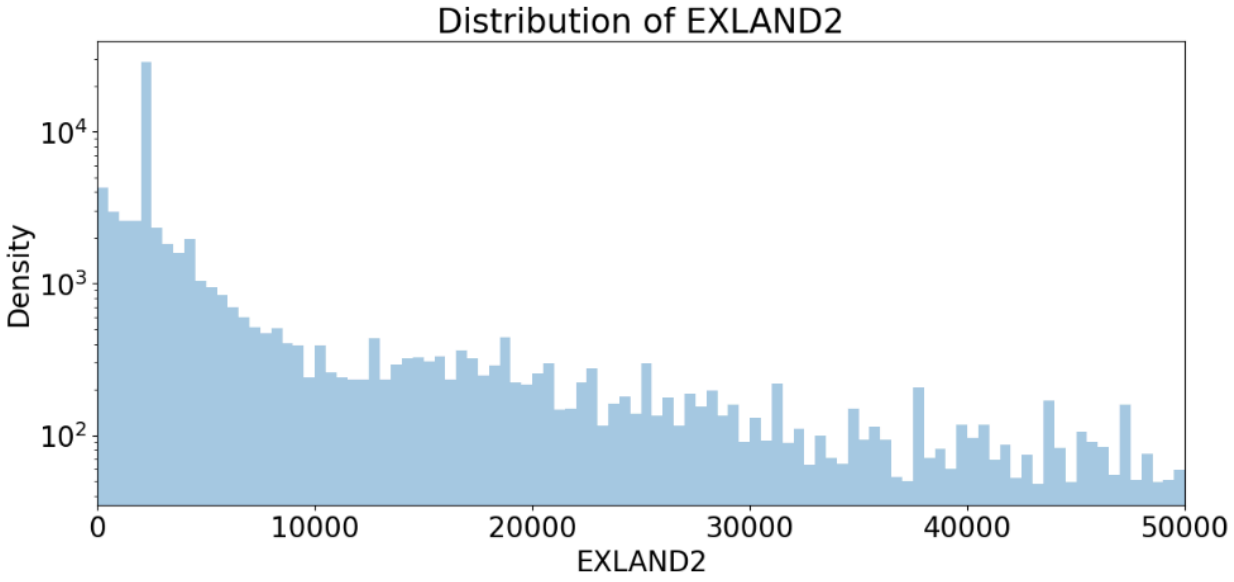
AVTOT2:

The "AVTOT2" field, a numeric attribute present in 26.4% of the 282,732 records, represents transitional total values. Ranging from 3.00 to 4,501,180,002.00, with an average value of $713,911.44 and a median of $11,652,508.34, visualizations illustrate the distribution of transitional total values (AVTOT2).



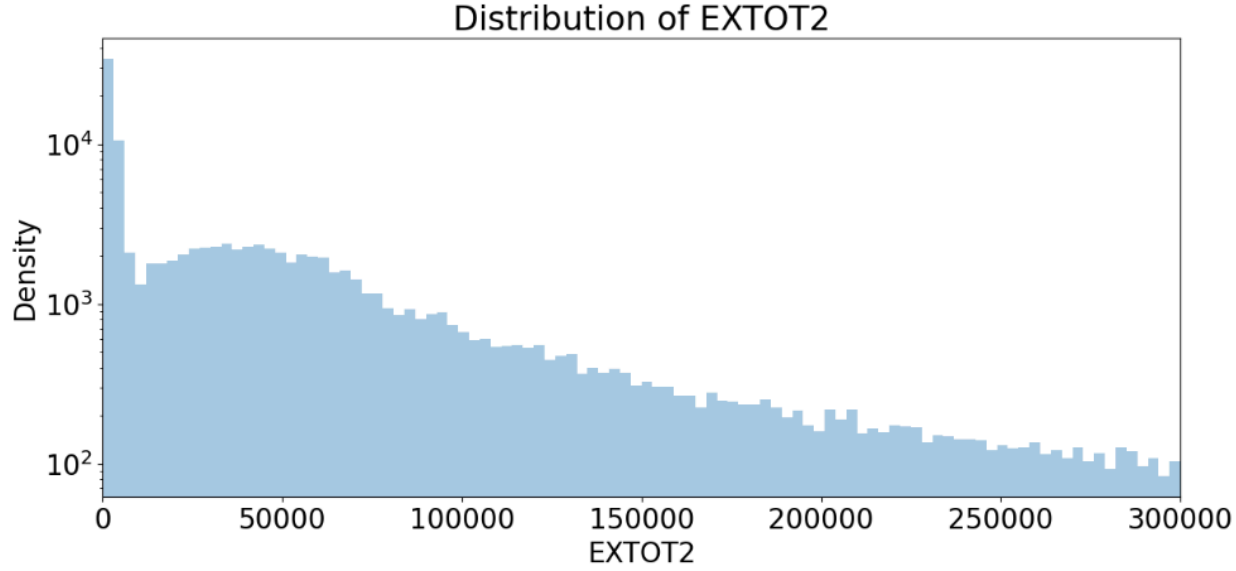
EXLAND2:

The "EXLAND2" field, a numeric attribute present in 8.2% of the 87,449 records, signifies transitional exemption land values. Ranging from 1.00 to 2,371,005,000.00, with an average value of $351,235.68 and a median of $10,802,150.91, visualizations illustrate the distribution of transitional exemption land values (EXLAND2).



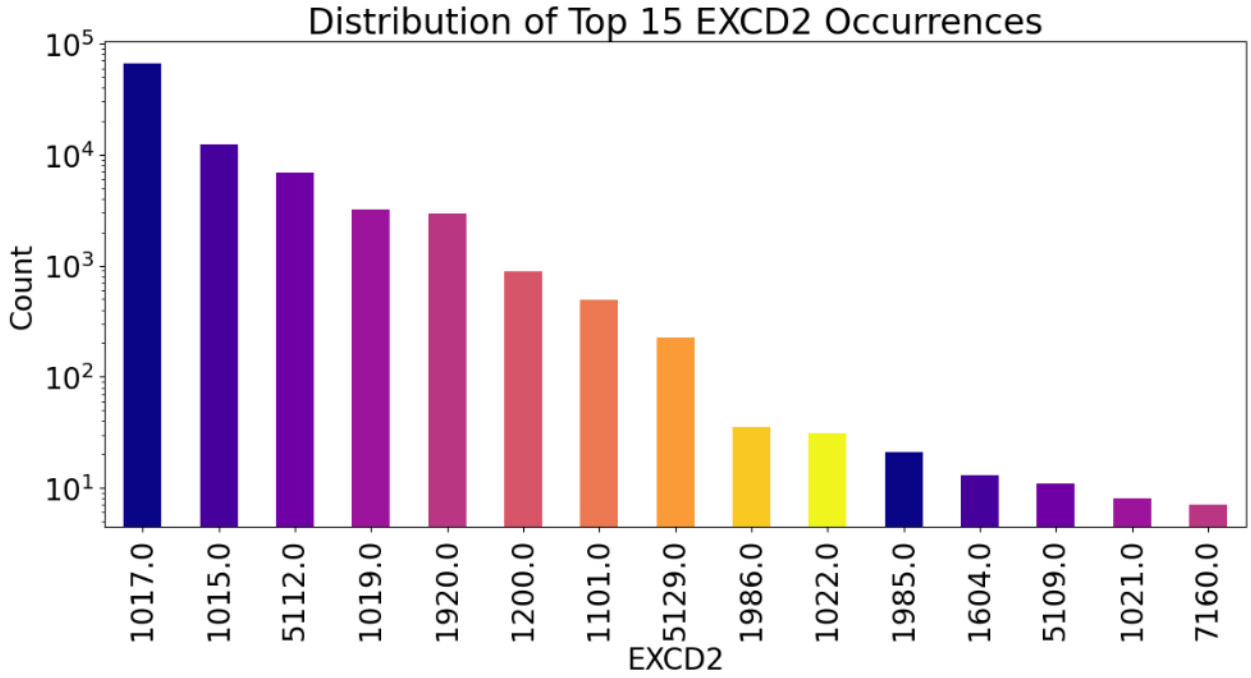
EXTOT2:

The "EXTOT2" field, a numeric attribute present in 12.2% of the 130,828 records, represents transitional exemption land totals. Ranging from 7.00 to 4,501,180,002.00, with an average value of $656,768.28 and a median of $16,072,448.75, visualizations illustrate the distribution of transitional exemption land totals (EXTOT2).



EXCD2:

The "EXCD2" field, a categorical attribute present in 8.7% of the 92,948 records, represents exemption codes (second). With 60 unique codes, the top 15 occurrences are visualized for analysis.



PERIOD:

The "PERIOD" field, a categorical attribute present in all 1,070,994 records, denotes the assessment period. With only one unique value, "FINAL," visualizations may not be necessary due to the lack of variation.

YEAR:

The "YEAR" field, a categorical attribute present in all 1,070,994 records, indicates the assessment year. With only one unique value, "2010/11," visualizations may not be necessary due to the lack of variation.

VALTYPE:

The "VALTYPE" field, a categorical attribute present in all 1,070,994 records, denotes the valuation type. With only one unique value, "AC-TR," visualizations may not be necessary due to the lack of variation.