

Data Quality Report

1. Data Description

The dataset appears to be a comprehensive record of real estate properties located in New York City, published by the **Department of Finance**. The data comprises various essential details about each property, including a unique identification number, location information such as borough, block, and lot for the **year 2010/11**. This data can be beneficial for businesses involved in the real estate industry to gain insights into property trends and make informed decisions related to investments, acquisitions, and sales. The dataset contains **1,070,994 records** across **32 fields**.

2. Summary Tables

Numerical Fields Table

Field Name	# Records Have Values	% Populated	% Zeros	Min	Max	Mean	Standard Deviation	Most Common
LTFRONT	1,070,994	100.00%	15.79%	0	9,999	36.63	74.03	0
LTDEPTH	1,070,994	100.00%	15.89%	0	9,999	88.86	76.4	100
STORIES	1,014,730	94.75%	0.00%	1	119	5.00	8.37	2
FULLVAL	1,070,994	100.00%	1.21%	0	6,150,000,000	874,264.50	11,582,425.58	0
AVLAND	1,070,994	100.00%	1.21%	0	2,668,500,000	85,067.91	4,057,258.16	0
AVTOT	1,070,994	100.00%	1.21%	0	4,668,308,947	227,238.16	6,877,526.09	0
EXLAND	1,070,994	100.00%	45.91%	0	2,668,500,000	36,423.89	3,981,573.93	0
EXTOT	1,070,994	100.00%	40.39%	0	4,668,308,947	91,186.98	6,508,399.78	0
BLDFRONT	1,070,994	100.00%	21.36%	0	7,575	23.04	35.58	0
BLDDEPTH	1,070,994	100.00%	21.37%	0	9,393	39.92	42.71	0
AVLAND2	282,726	26.40%	0.00%	3	2,371,005,000	246,235.71	6,178,951.64	2,408
AVTOT2	282,732	26.40%	0.00%	3	4,501,180,002	713,911.43	11,652,508.34	750
EXLAND2	87,449	8.17%	0.00%	1	2,371,005,000	351,235.68	10,802,150.91	2,090
EXTOT2	130,828	12.22%	0.00%	7	4,501,180,002	656,768.28	16,072,448.75	2,090

Categorical Fields Table

Field Name	# Records Have Values	% Populated	# Zeros	# Unique Values	Most Common
RECORD	1,070,994	100.00%	0	1,070,994	1
BBLE	1,070,994	100.00%	0	1,070,994	1000010101
BORO	1,070,994	100.00%	0	5	4
BLOCK	1,070,994	100.00%	0	13,984	3944
LOT	1,070,994	100.00%	0	6,366	1
EASEMENT	4,636	0.43%	0	12	E
OWNER	1,039,249	97.04%	0	863,347	PARKCHESTER PRESERVAT
BLDGCL	1,070,994	100.00%	0	200	R4
TAXCLASS	1,070,994	100.00%	0	11	1
EXT	354,305	33.08%	0	3	G
EXCD1	638,488	59.62%	0	129	1017
STADDR	1,070,318	99.94%	0	839,280	501 SURF AVENUE
ZIP	1,041,104	97.21%	0	196	10314
EXMPTCL	15,579	1.45%	0	14	X1
EXCD2	92,948	8.68%	0	60	1,017
PERIOD	1,070,994	100.00%	0	1	FINAL
YEAR	1,070,994	100.00%	0	1	2010/11
VALTYPE	1,070,994	100.00%	0	1	AC-TR

3. Visualization of each Field

1) Field Name: RECORD

Description: The RECORD column is a unique identifier for each property record in the dataset. It contains no zeros and has 100% populated values, with a total of 1,070,994 unique values, which signifies the total number of records in the dataset.

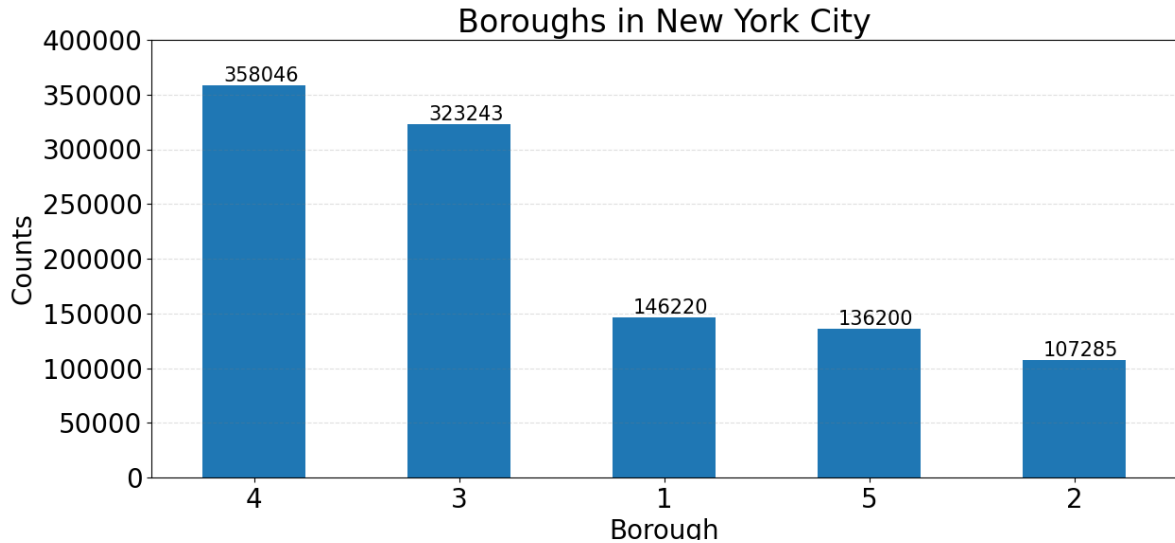
2) Field Name: BBLE

Description: The BBLE column is a categorical column that stands for "borough, block, and lot" and is a unique identifier for each property in New York City. The first digit of the BBLE represents the borough followed by the block number and the lot number. This column is used to link the property information with other datasets.

3) Field Name: BORO

Description: The BORO column contains the borough code where the property is located. There are 5 borough codes and they are assigned like 1 = Manhattan, 2 = Bronx, 3 = Brooklyn, 4 = Queens, 5 = Staten Island. This column is used for identifying the location of the property within New York City.

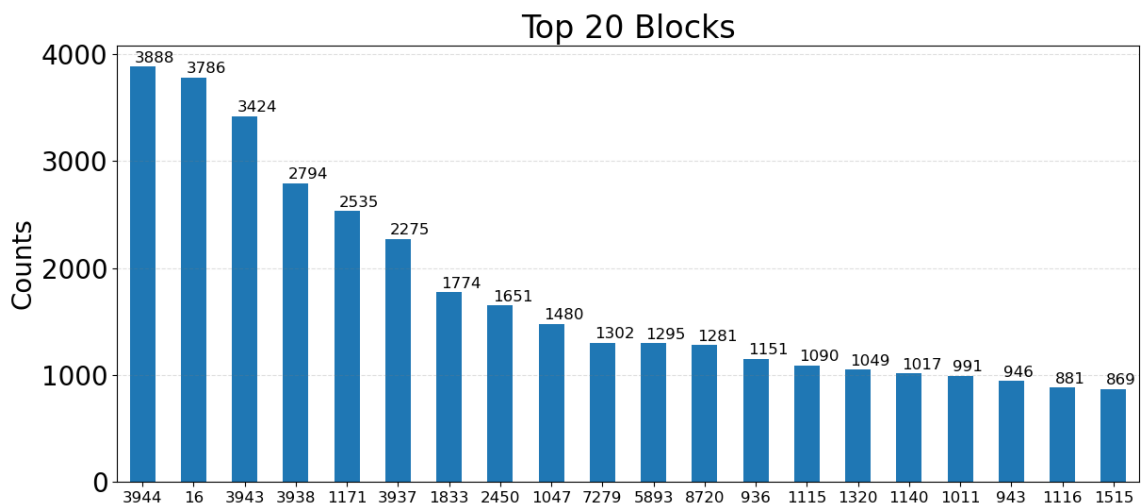
Visualization:



4) Field Name: BLOCK

Description: The BLOCK column in the dataset represents the tax block for each property, a subdivision of a borough used for identifying a group of properties. Each tax block has a unique identification number, and there are a total of 13,984 distinct tax blocks in the dataset.

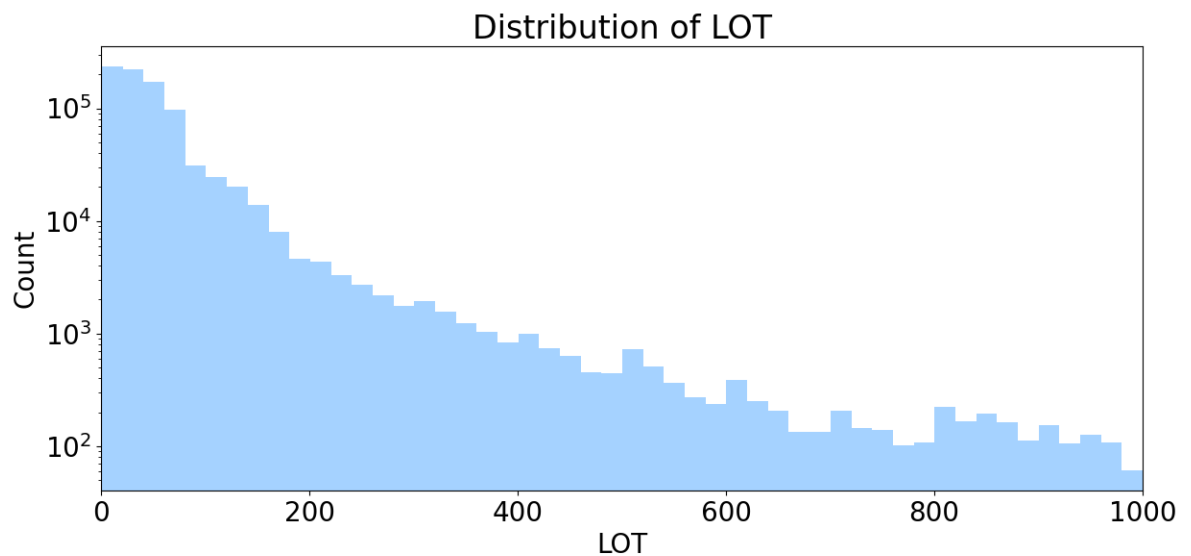
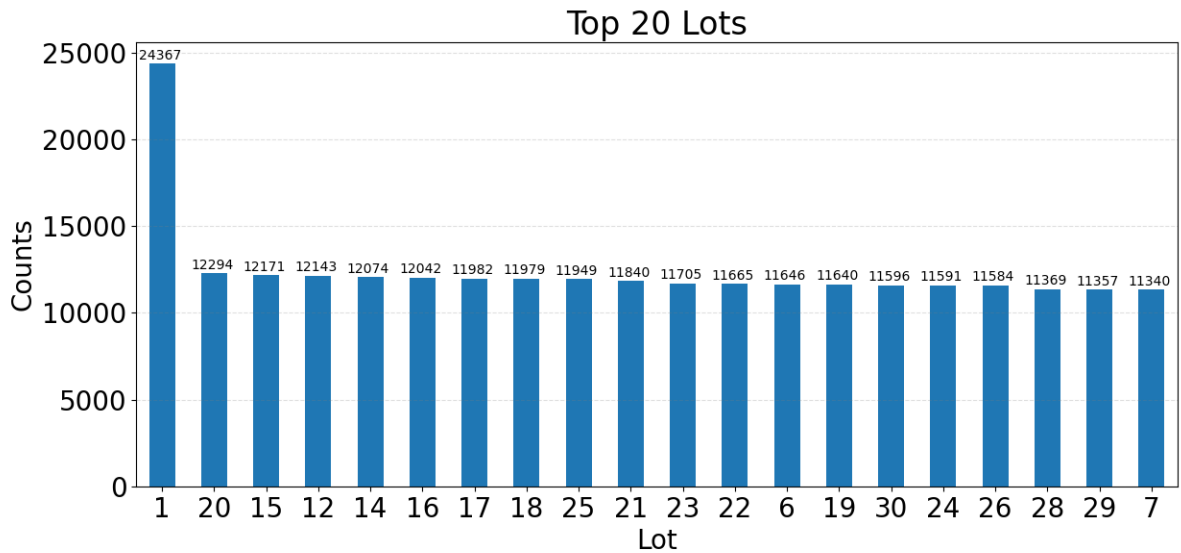
Visualization:



5) Field Name: LOT

Description: The LOT column refers to the tax lot number assigned to the property by the New York City Department of Finance. It is a unique identifier for each tax lot within a block and borough. The LOT number is used for determining ownership, assessed value, and tax liability of each property.

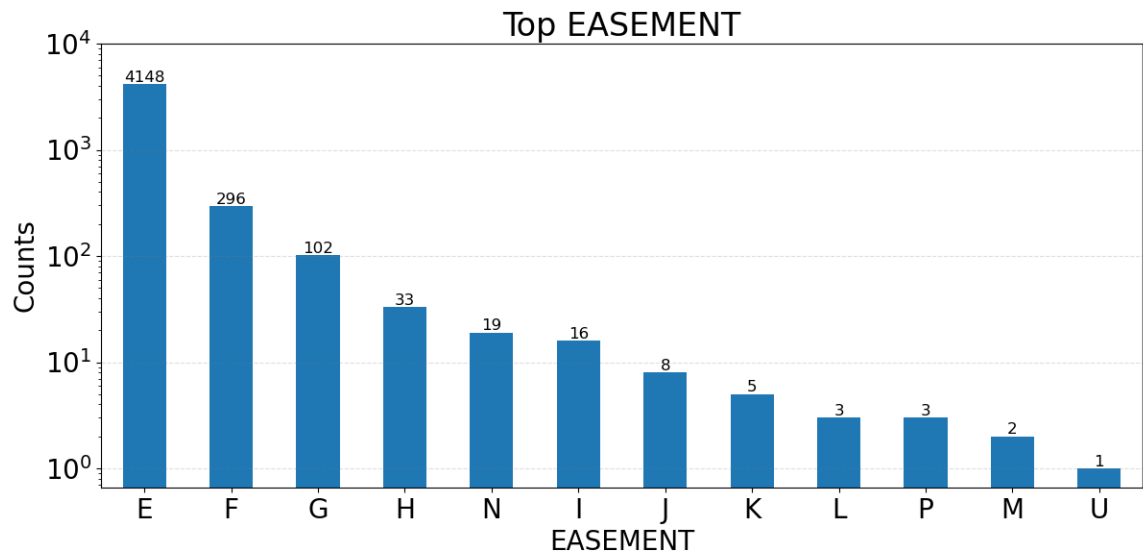
Visualization:



6) Field Name: EASEMENT

Description: The EASEMENT column contains information about any easement associated with a property. This column has 0.43% populated values, indicating that easements are present for only a small percentage of the properties in the dataset. The column has 12 unique values, with the most common value being "E".

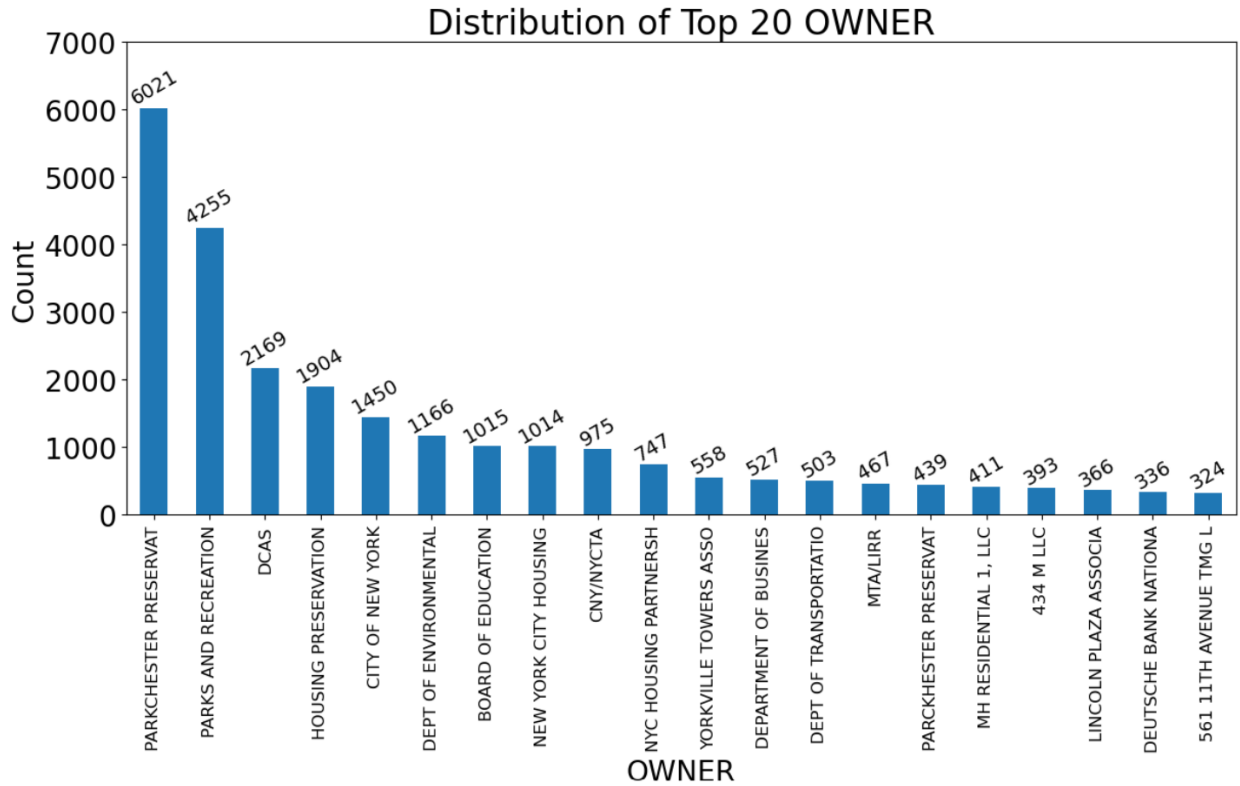
Visualization:



7) Field Name: OWNER

Description: The OWNER column in the dataset contains the name of the property owner. It has 1,039,249 records and 97.04% populated. It has 863,347 unique values, and the most common owner is "PARKCHESTER PRESERVAT".

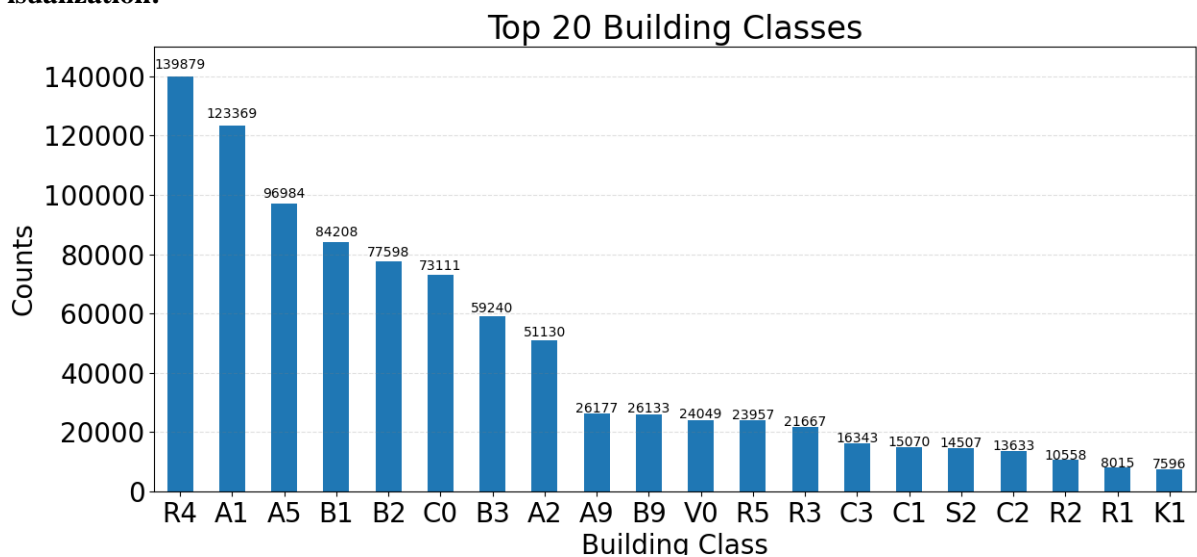
Visualization:



8) Field Name: BLDGCL

Description: The BLDGCL column represents the Building Class Code, which is a code used to classify properties based on the type of building they are. The code consists of a two to four-character alpha-numeric characters.

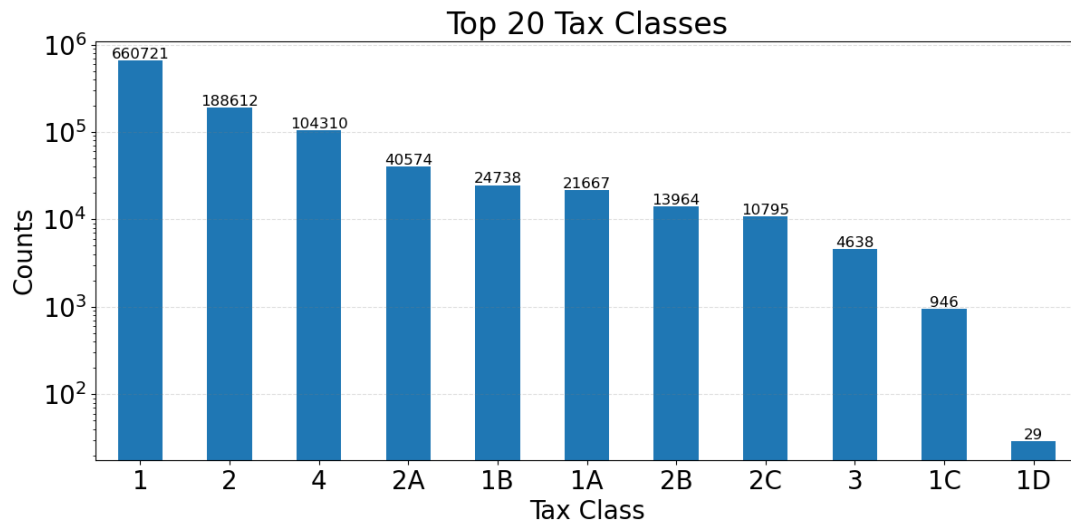
Visualization:



9) Field Name: TAXCLASS

Description: The TAXCLASS column is a categorical column that indicates the tax class of each property. There are 11 unique tax classes represented as integers from 1 to 4, and then 6, 7, 8, and 9, and letters like 'A', 'B', and 'C'.

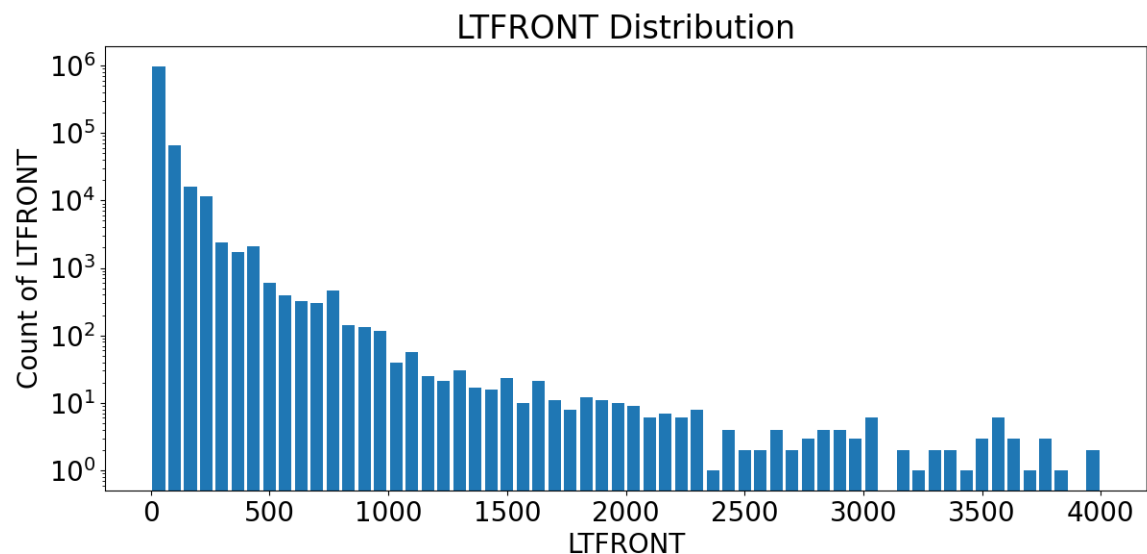
Visualization:

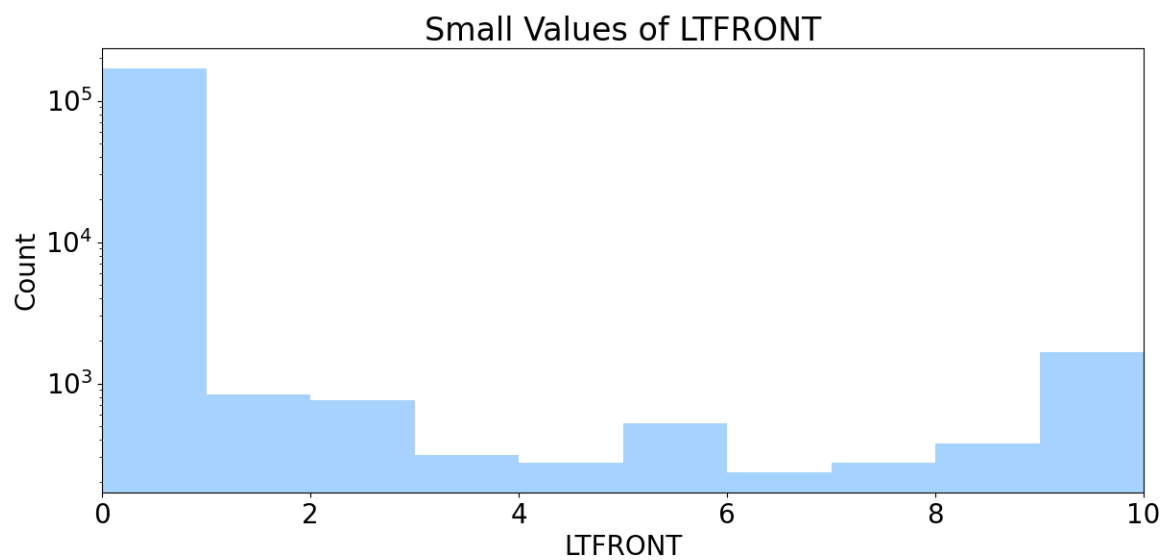
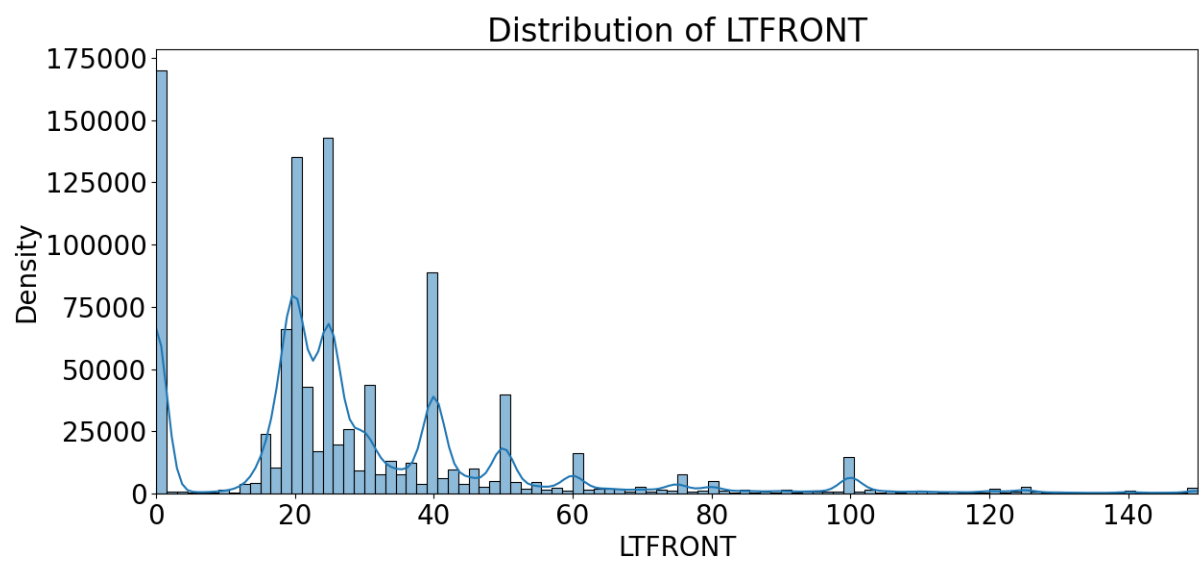


10) Field Name: LTFRONT

Description: The LTFRONT column refers to the lot frontage, the linear feet of the street front of the property. There are 1,070,994 records with values in this column, and it is 100% populated. The minimum value is zero, the maximum value is 9,999, with 15.79% of the values being zero. The mean value is 36.63 feet, and the standard deviation is 74.03 feet.

Visualization:

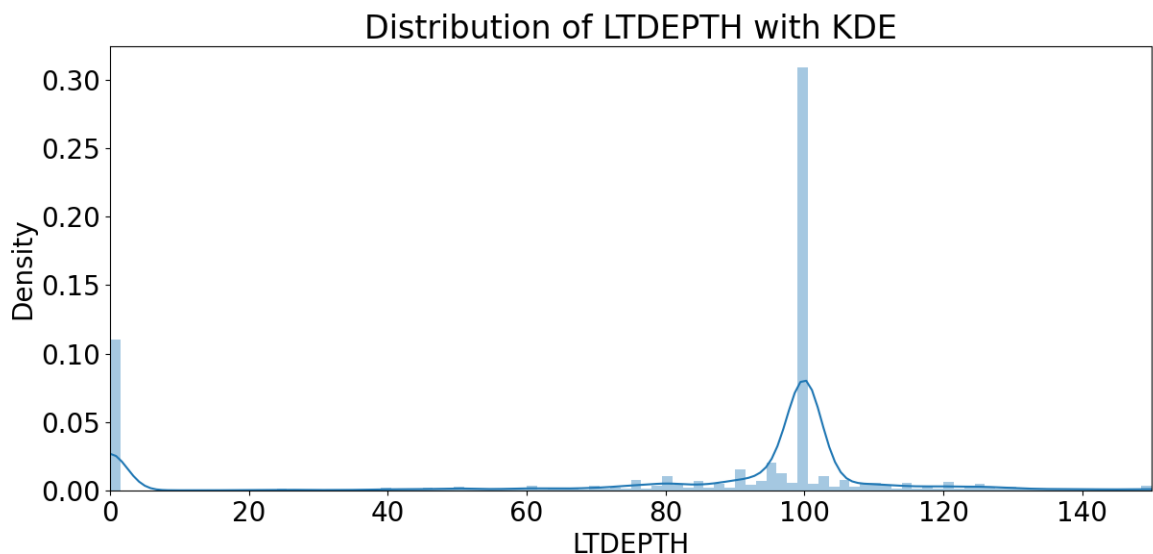
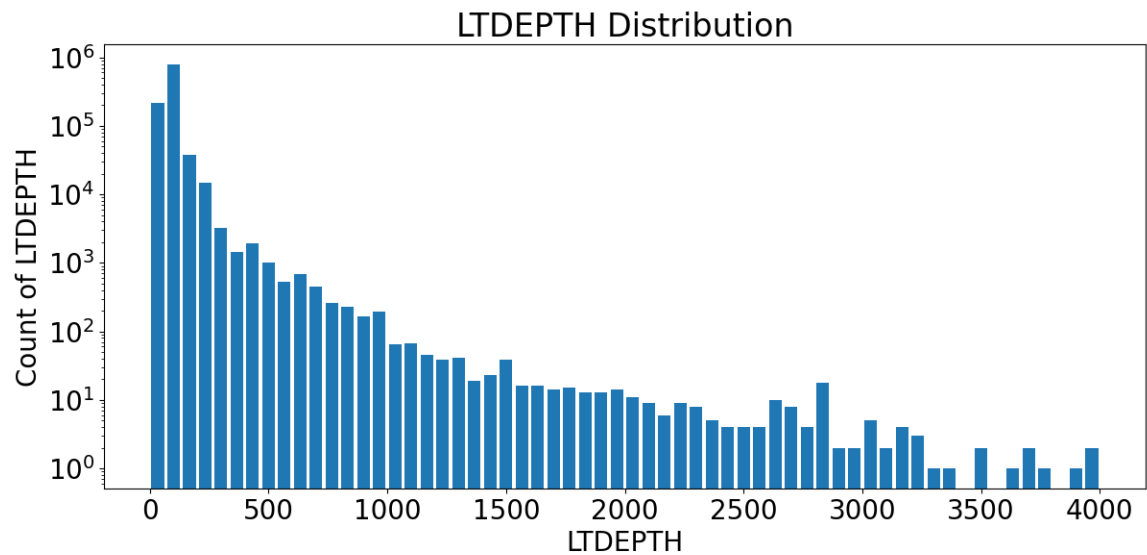


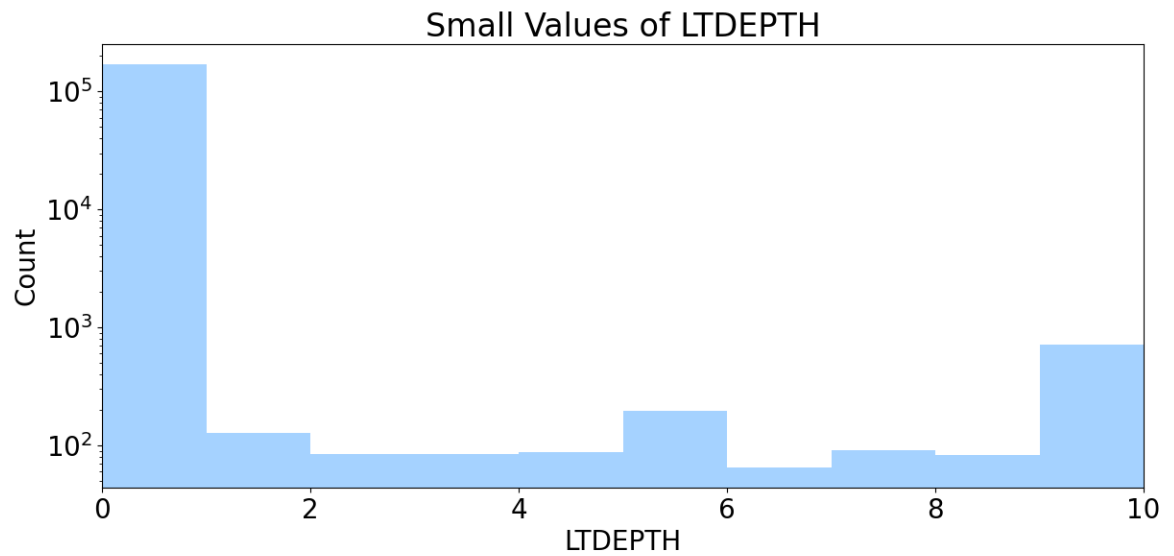


11) Field Name: LTDEPTH

Description: The LTDEPTH column represents the depth of the lot in feet. It has 1,070,994 records with values, and 100% of the records are populated. The minimum value is 0 feet, the maximum value is 9,999 feet, with 15.89% of the values being zero.. The mean depth of the lot is 88.86 feet, and the standard deviation is 76.4 feet.

Visualization:

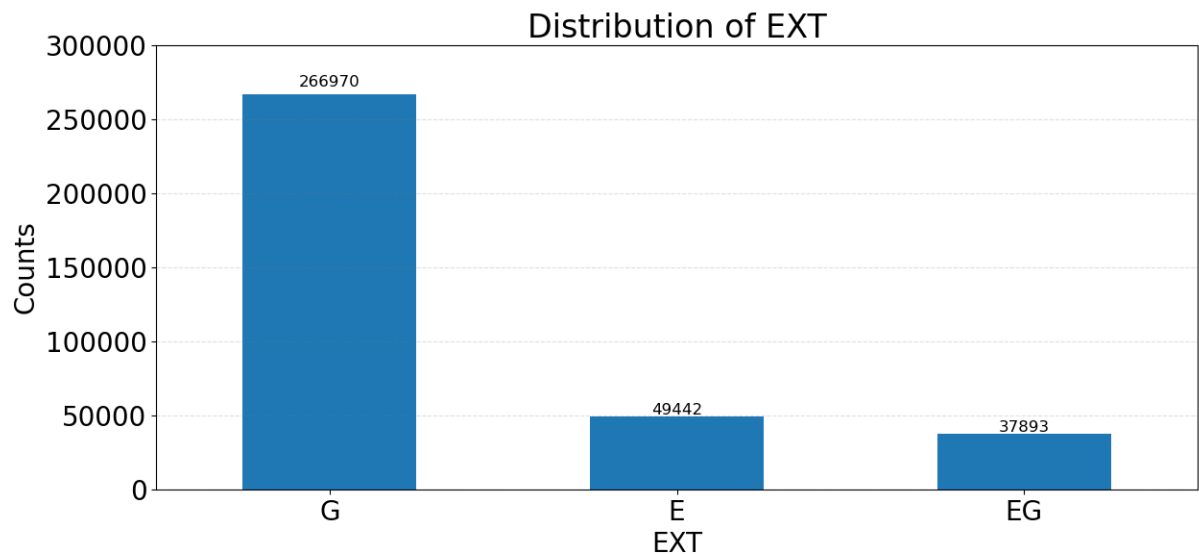




12) Field Name: EXT

Description: The EXT column in the given dataset has 354,305 records with 33.08% populated values. The column represents the Extension code of the building or the property. The unique values in this column are 'G', 'E', and ' ', where 'G' denotes a garage, 'E' denotes an extension, and ' ' denotes the absence of a garage or an extension.

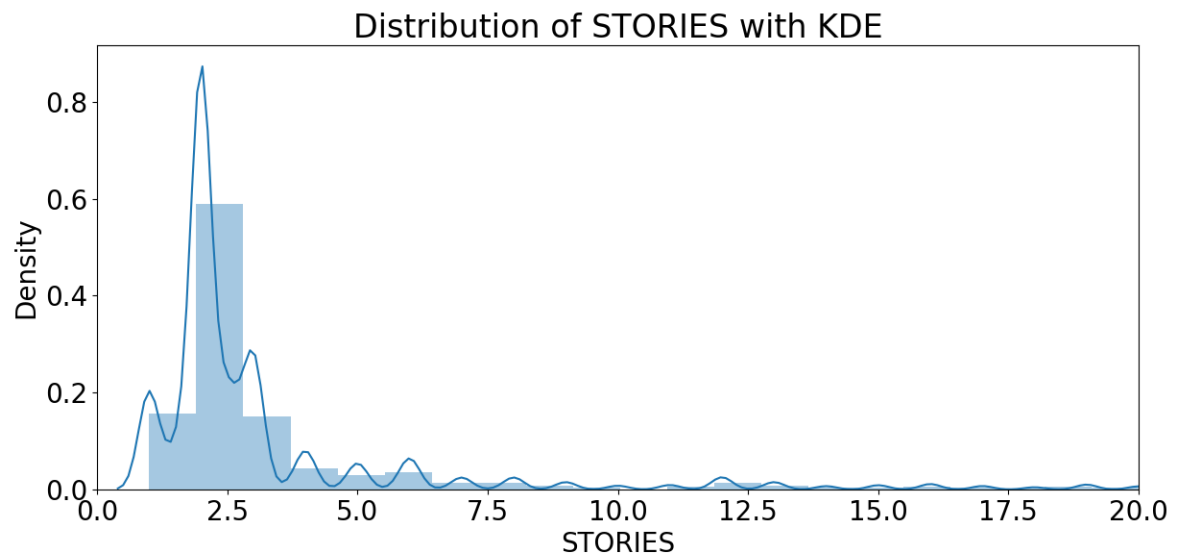
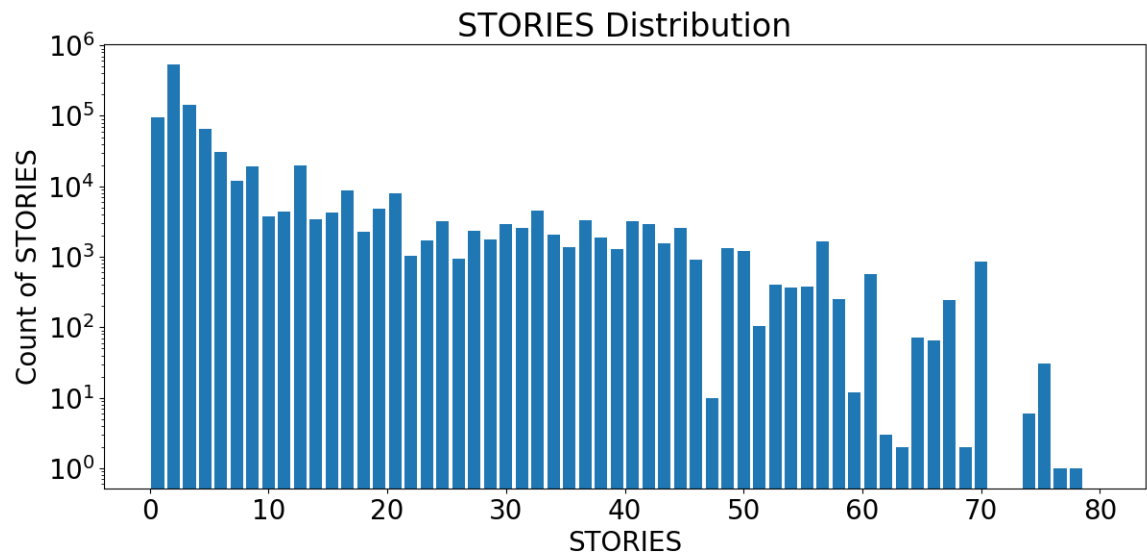
Visualization:



13) Field Name: STORIES

Description: The STORIES column in the dataset contains the number of stories in the building that the property is located in. It is a numerical column with a range of values from 1 to 119. This column has 1,014,730 records with values, and it is 94.75% populated. The most common value in this column is 2, with a mean of 5.00 and a standard deviation of 8.37.

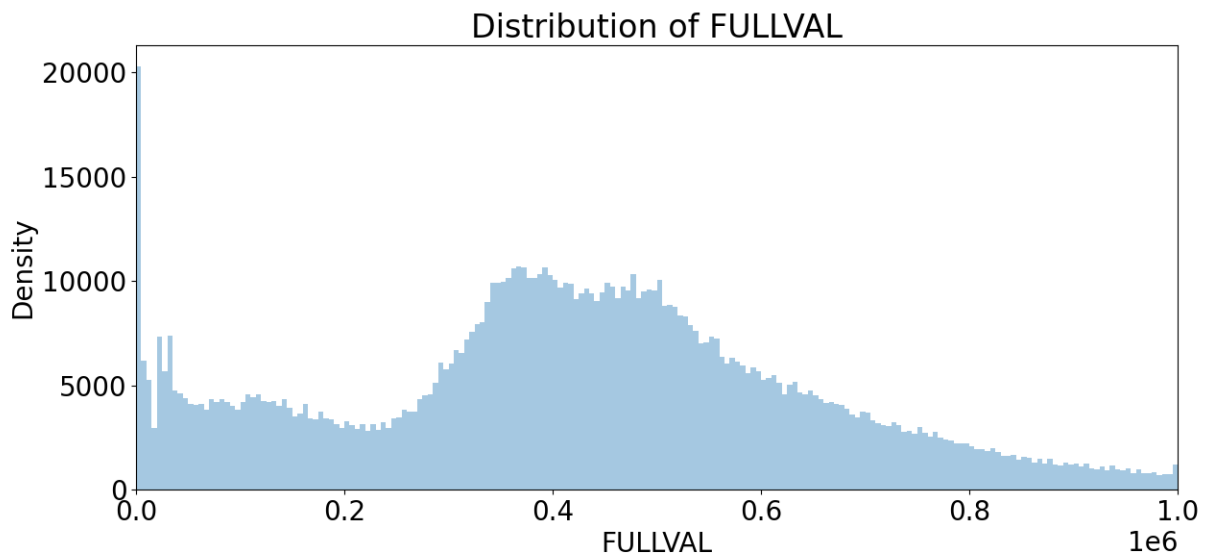
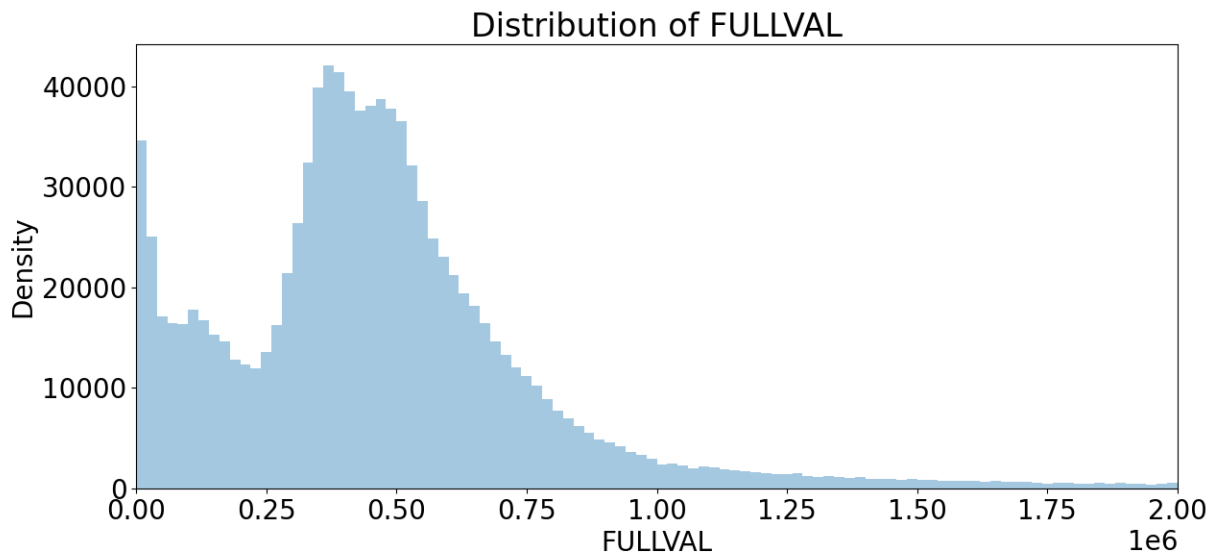
Visualization:



14) Field Name: FULLVAL

Description: The FULLVAL column contains the assessed value of the property for tax purposes. It is expressed in dollars and has a mean value of \$874,264.50 and standard deviation of \$11,582,425.58. It has a minimum value of 0 and a maximum value of \$6,150,000,000.

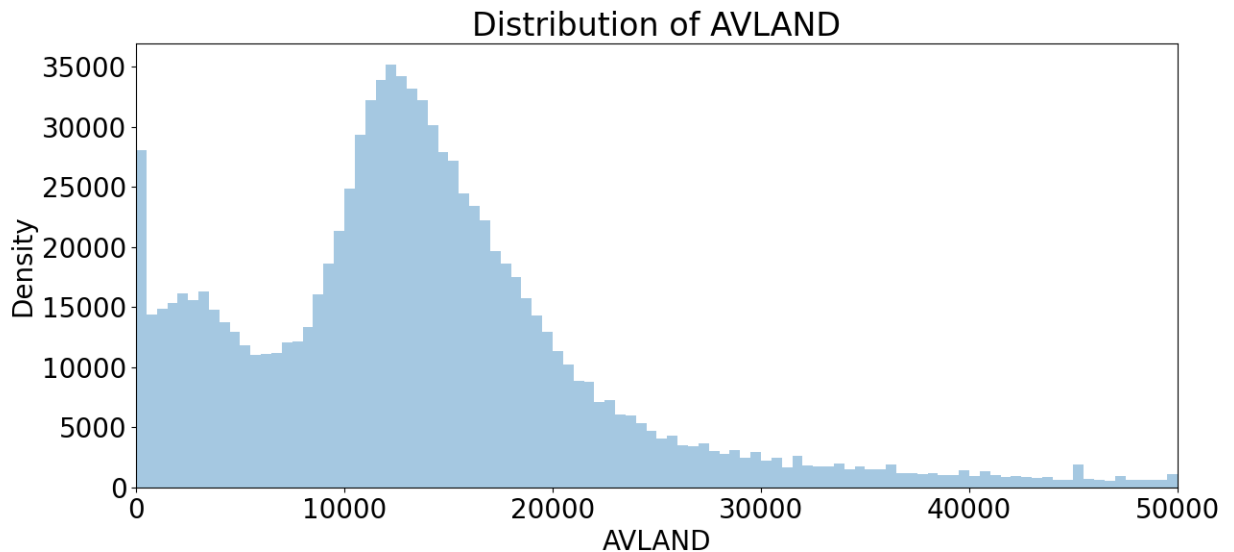
Visualization:



15) Field Name: AVLAND

Description: The AVLAND column represents the assessed value of land for the tax lot. The values range from 0 dollars to a maximum value of 2,668,500,000 dollars. The most common value in this column is also 0 dollars signifying there are many tax lots that do not have any assessed value for land.

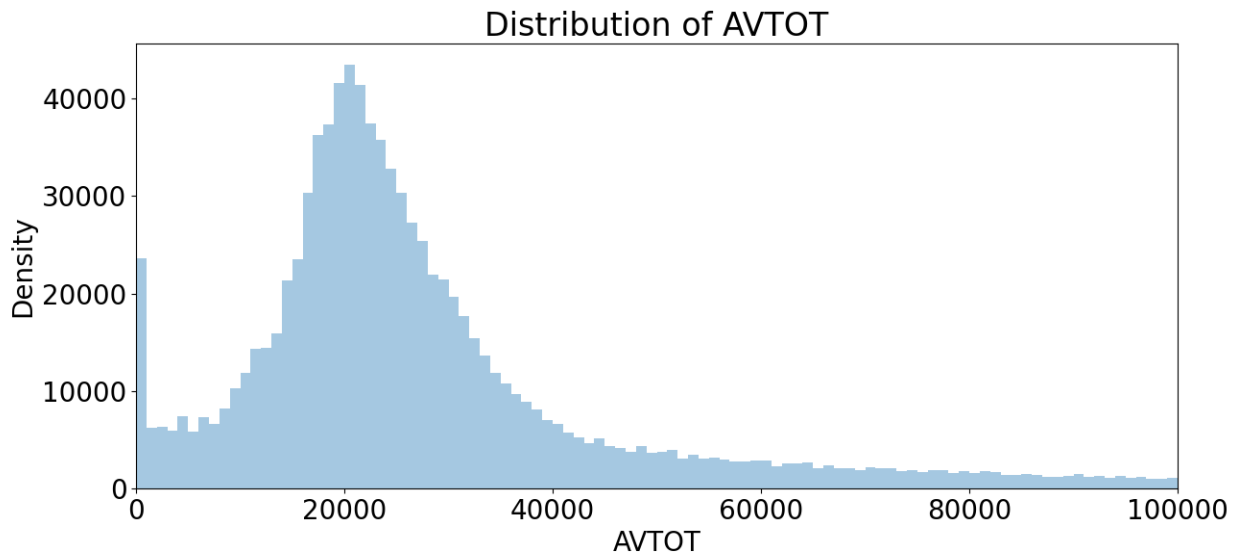
Visualization:



16) Field Name: AVTOT

Description: AVTOT column contains the assessed total value for land and building for each property. The column has 1,070,994 records, and all of them have value. The values range from 0 to 4,668,308,947, with 1.21% values being zero.

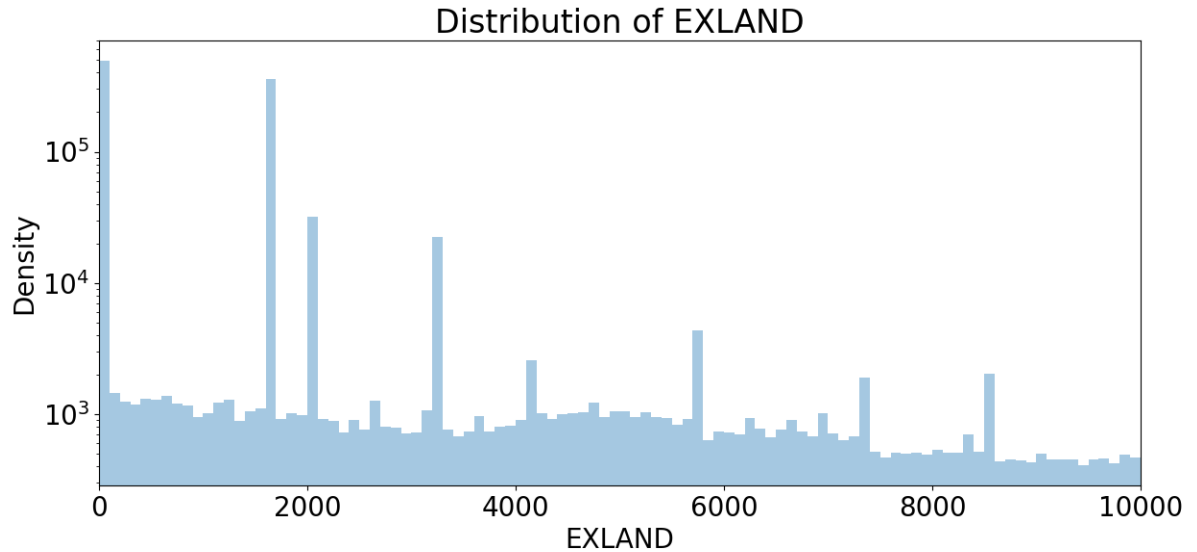
Visualization:



17) Field Name: EXLAND

Description: The EXLAND column contains the assessed value of the land excluding any buildings or structures on it. The mean value in the column is 36,423.89, and the standard deviation is 3,981,573.93.

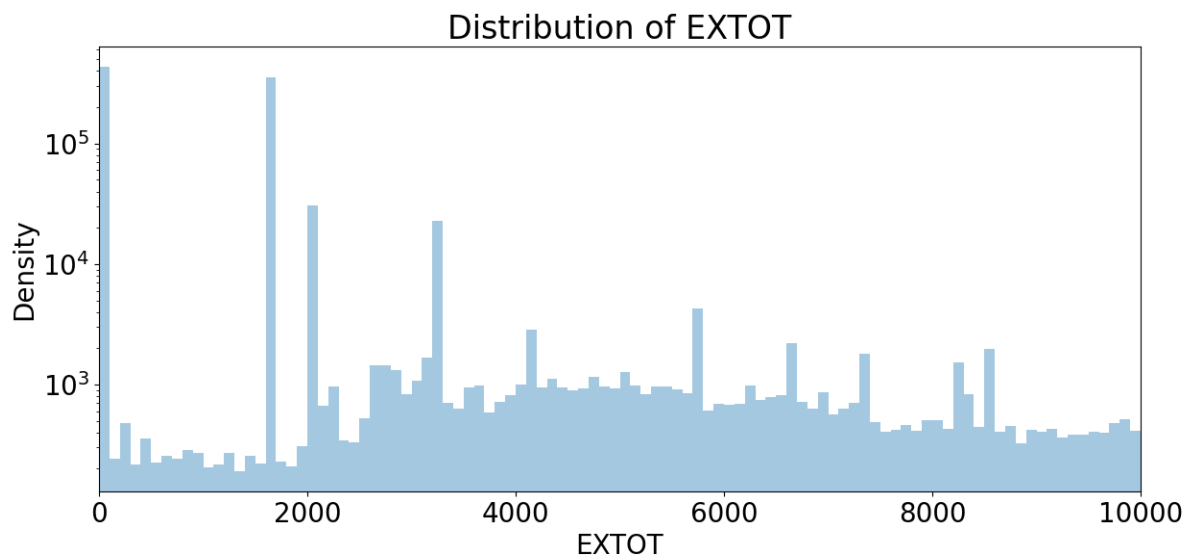
Visualization:



18) Field Name: EXTOT

Description: The EXTOT column contains the total estimated market value of the building's exterior lot area. It has 1,070,994 records, with 40.39% of them having zero values. The mean and standard deviation are 91,186.98 and 6,508,399.78, respectively.

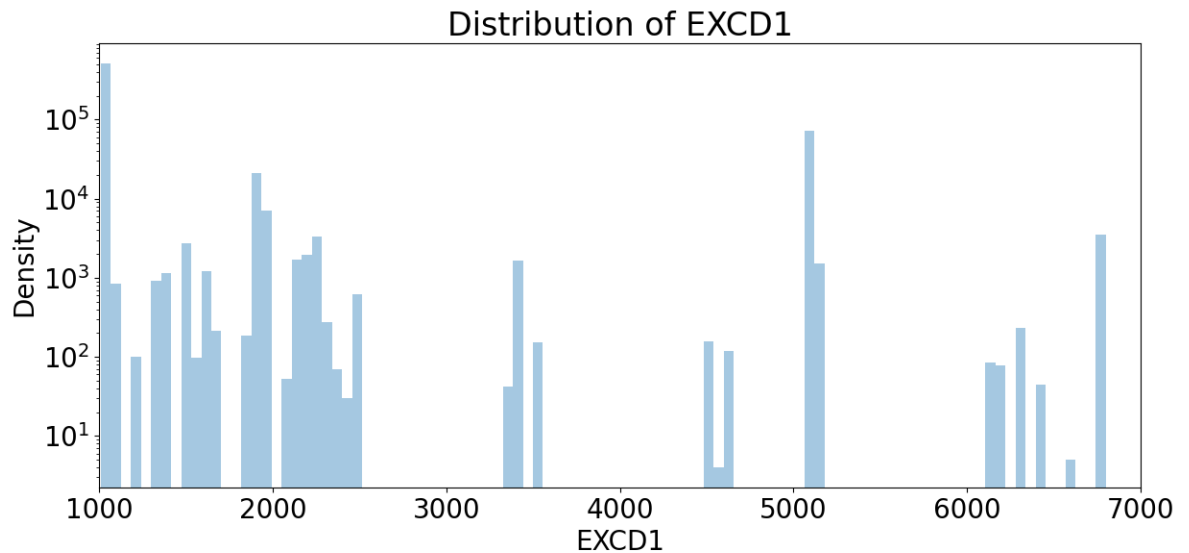
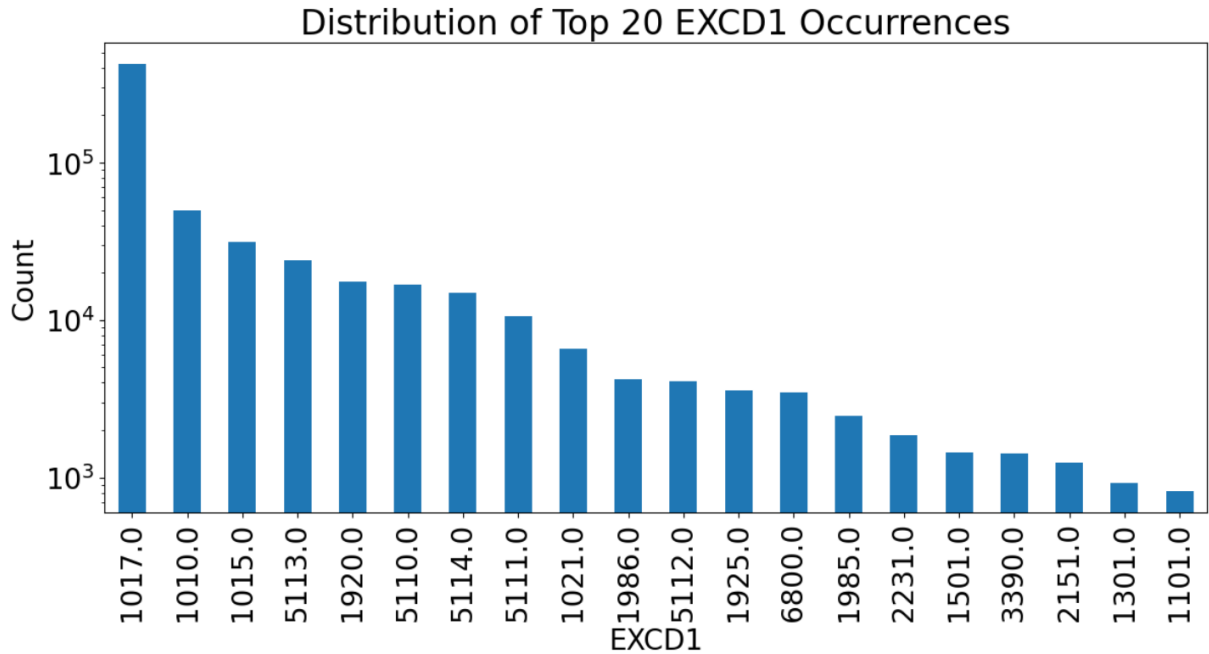
Visualization:



19) Field Name: EXCD1

Description: The EXCD1 column refers to the first exemption code, which is a numeric code that indicates the reason for a property's exemption from taxation. This column has a relatively low level of completeness, with only 45.48% of the records having non-zero values.

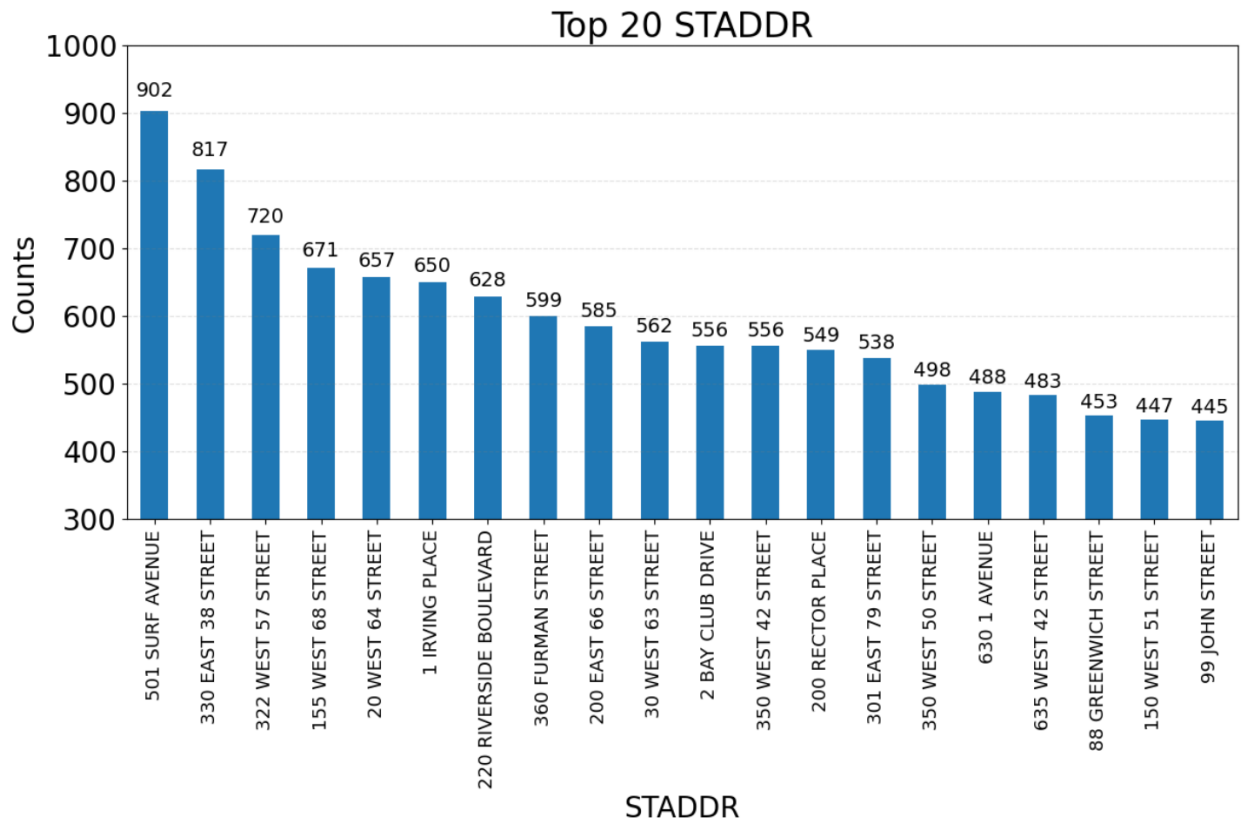
Visualization:



20) Field Name: STADDR

Description: The STADDR column likely represents the street address of each property. Majority of the values are populated with the most common value being “501 SURF AVENUE” appearing 902 times.

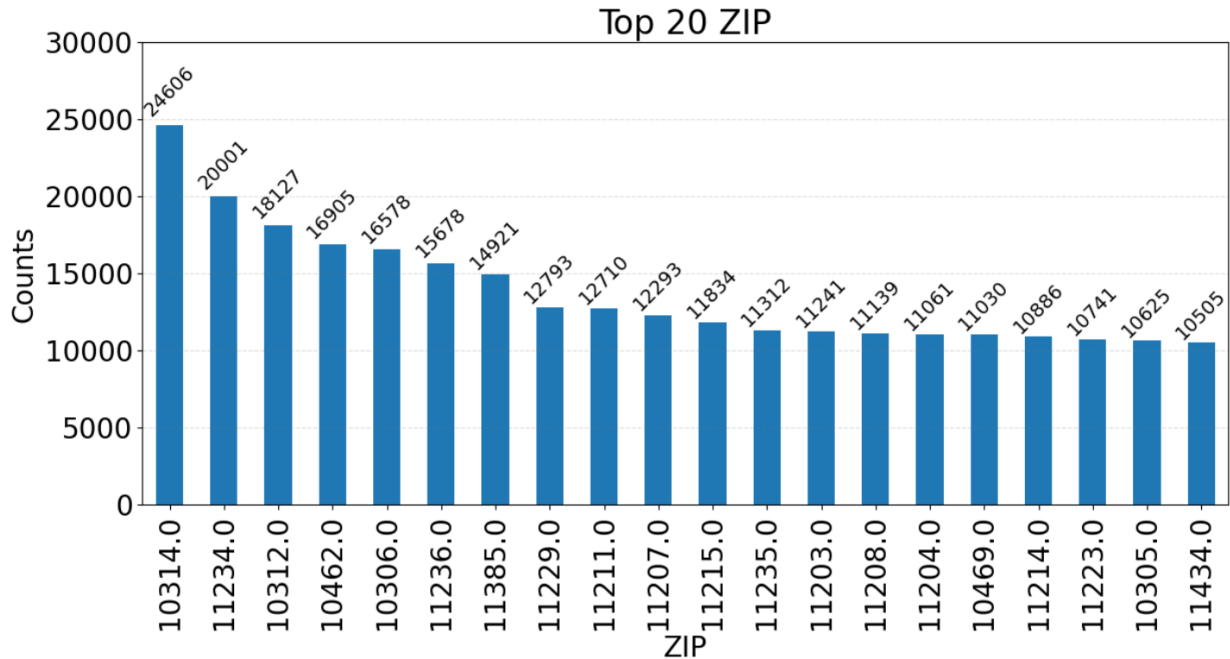
Visualization:



21) Field Name: ZIP

Description: The ZIP column in the dataset contains the postal code associated with each property in New York City. The most common zip code is 10314 with a count of 24606.

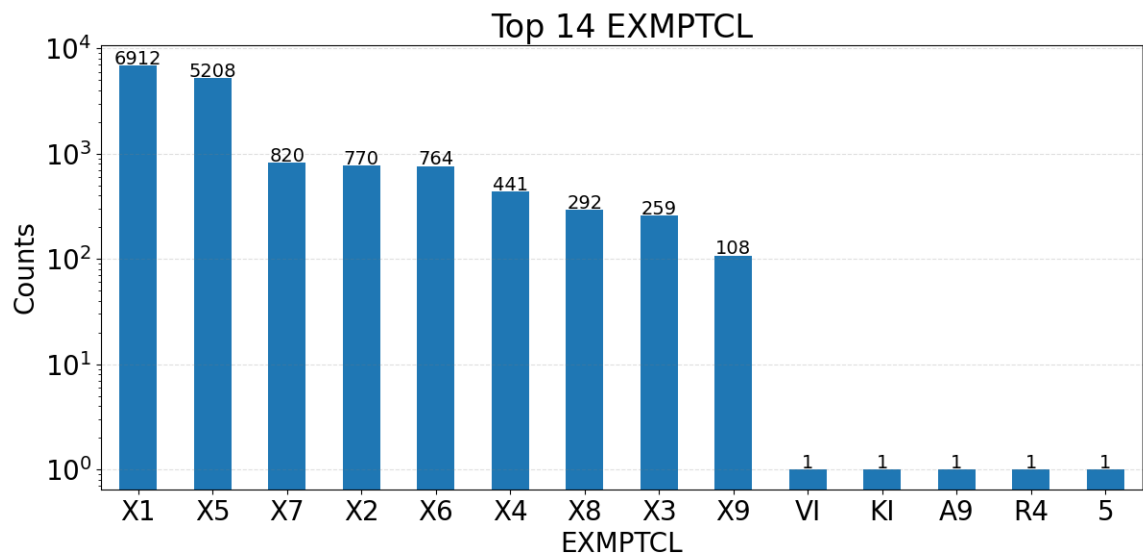
Visualization:



22) Field Name: EXMPTCL

Description: The EXMPTCL column in the dataset refers to the exempt class of the property. The values represent a two-digit code that specifies the type of exemption, such as "1A" for a property owned by a government entity, "2A" for a property used for religious purposes and so on. The most common value is X1 with a total count of 6912.

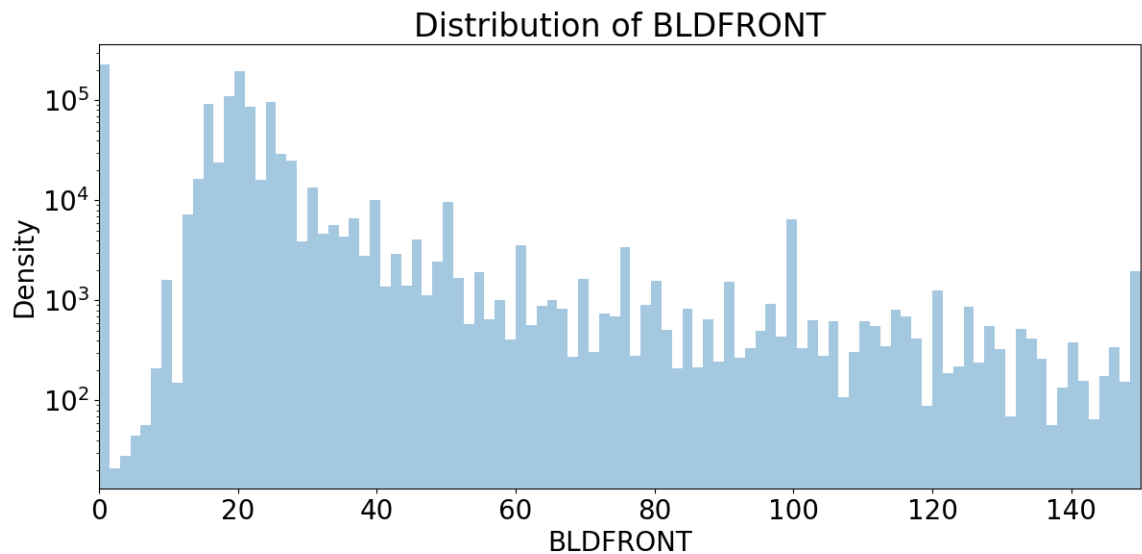
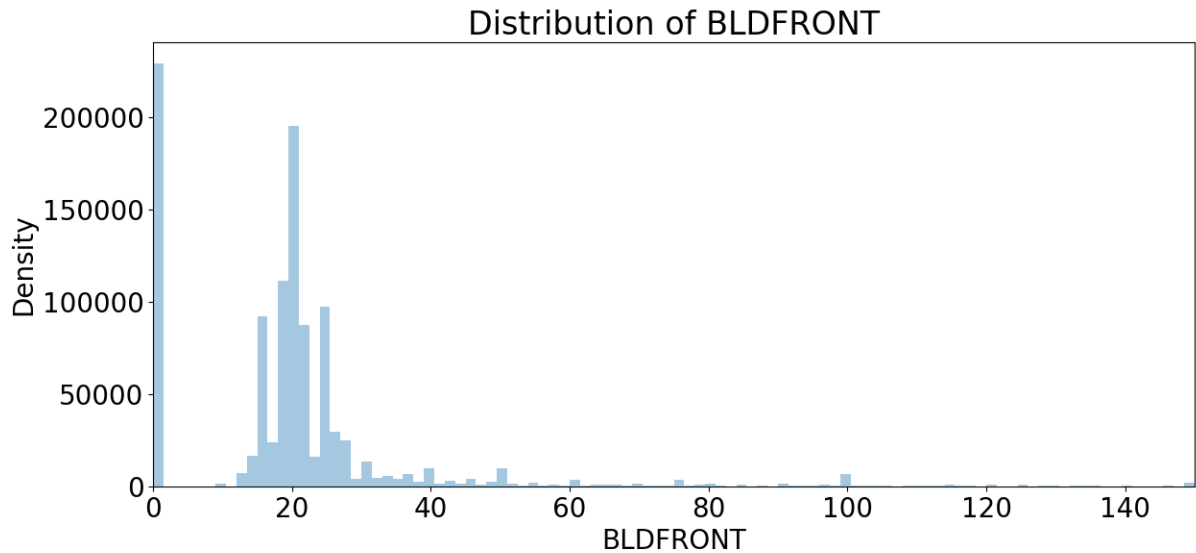
Visualization:

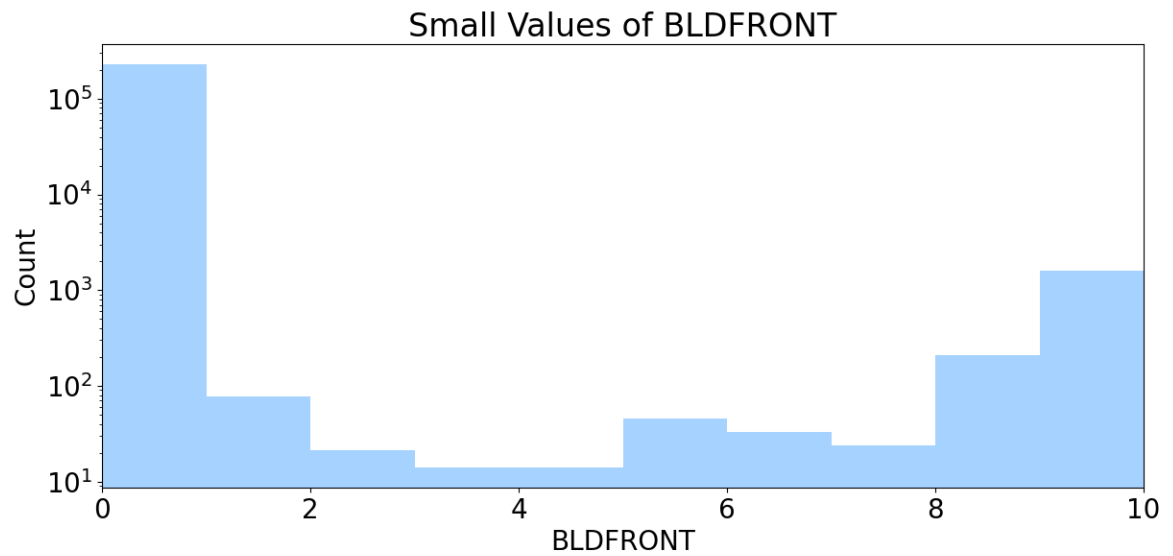


23) Field Name: BLDFRONT

Description: The BLDFRONT column in the dataset contains the number of feet of the building facing the street. It is a continuous numerical column ranging from 0 to 7,575 feet with 1,070,994 records, and 21.36% of zeros. The mean value is 23.04 feet, and the standard deviation is 35.58 feet.

Visualization:

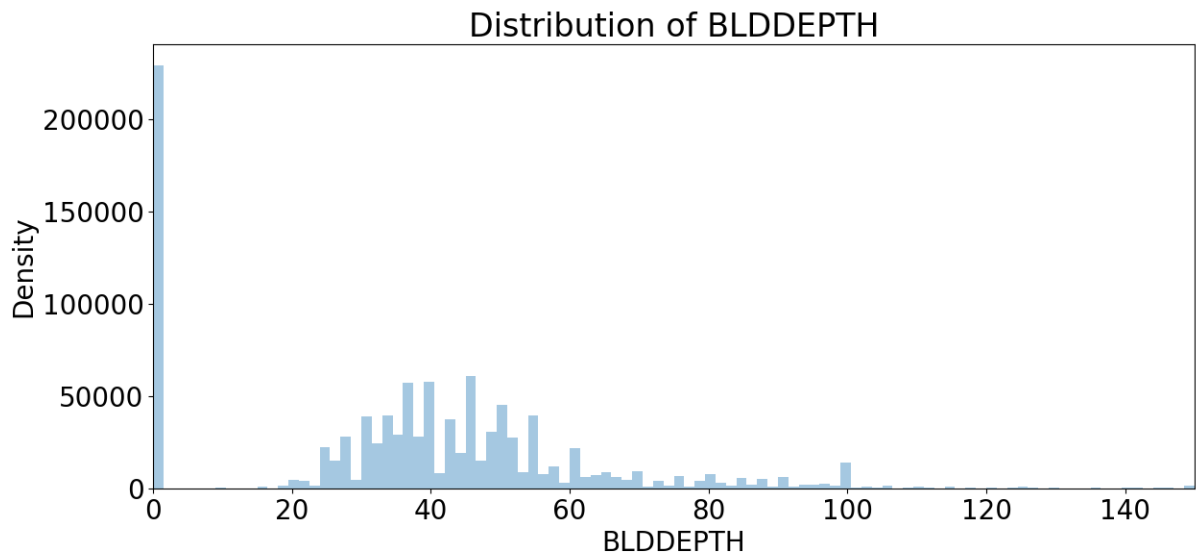


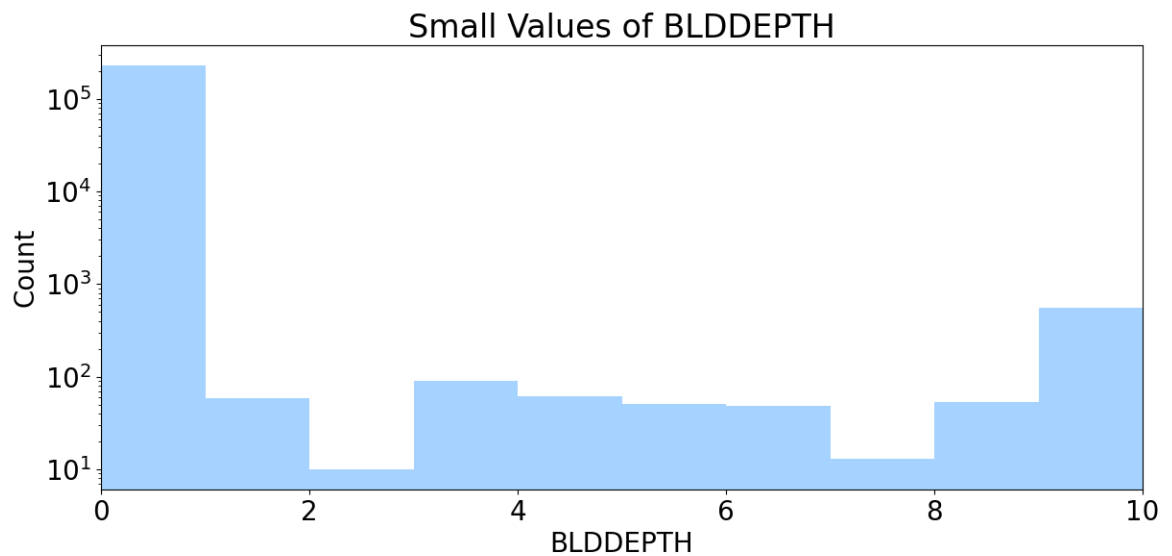
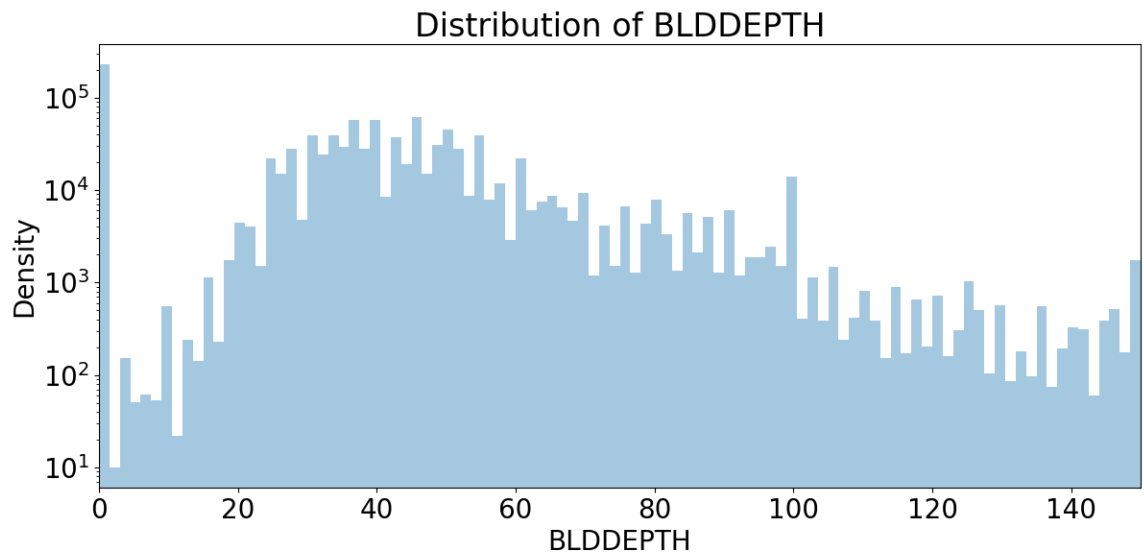


24) Field Name: BLDDEPTH

Description: The BLDDEPTH column contains the building depth in feet. The data in this column is 100% populated, and it has a mean value of 39.92 feet and a standard deviation of 42.71 feet. The minimum value in this column is 0 feet, which suggests that some records have missing values or an unusual entry.

Visualization:

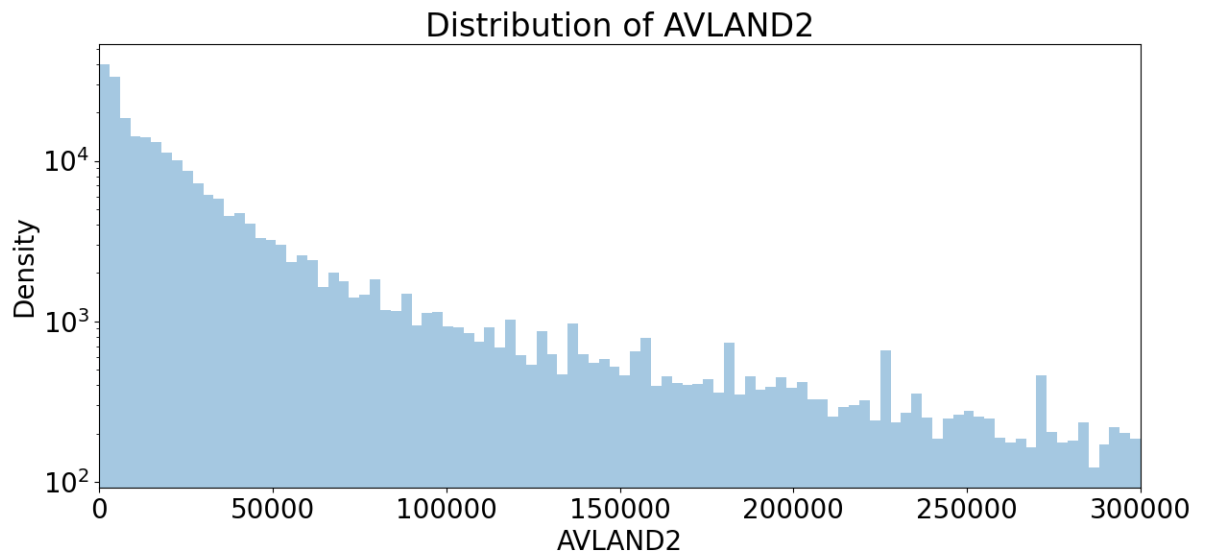




25) Field Name: AVLAND2

Description: The AVLAND2 column contains the assessed land value of the tax lot from the second assessment roll. It has 282,726 records with 26.40% populated and a minimum value of 3 and a maximum value of 2,371,005,000. The mean and standard deviation of the column are 246,235.71 and 6,178,951.64, respectively.

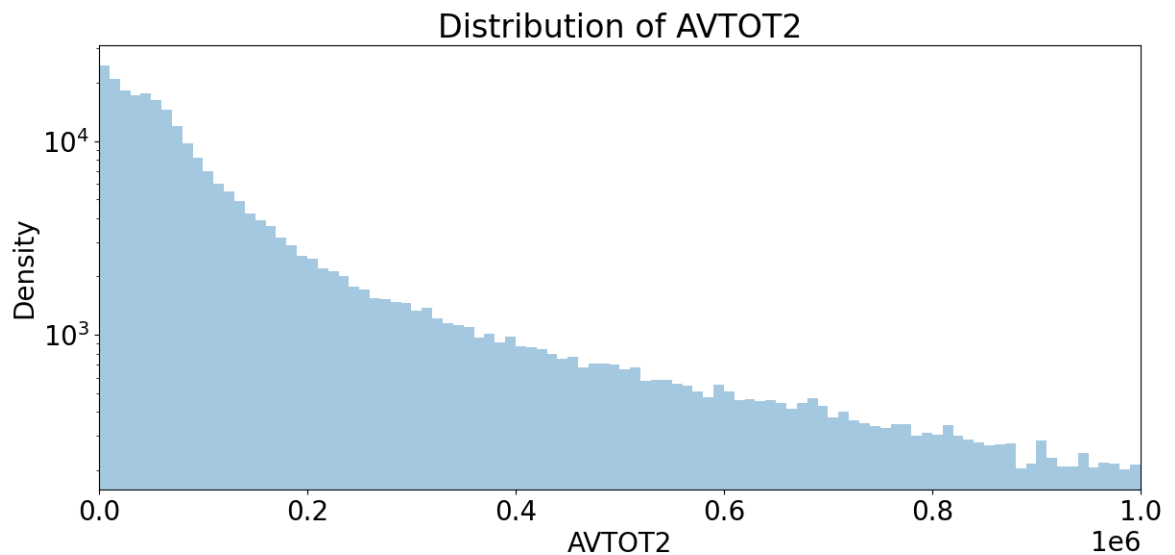
Visualization:



26) Field Name: AVTOT2

Description: The AVTOT2 column refer to the tentative assessed total value for properties with final assessment rolls. It has 282,732 records, with a population percentage of 26.40%. The minimum value is 3, and the maximum value is 4,501,180,002, with a mean value of 713,911.43 and a standard deviation of 11,652,508.34.

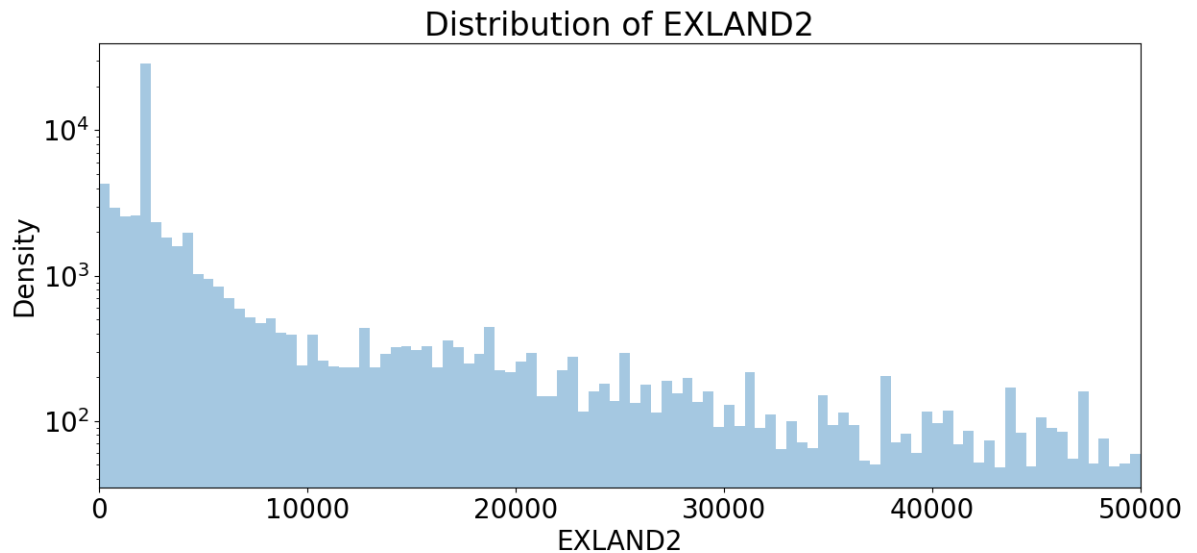
Visualization:



27) Field Name: EXLAND2

Description: The EXLAND2 column refers to the assessed value of the land excluding exempt amounts for properties with multiple units, condominiums, and cooperatives. It has 87,449 records, with a data completeness percentage of 8.17% and a maximum value of 2,371,005,000. The average assessed value is \$351,235.68, and the standard deviation is \$10,802,150.91.

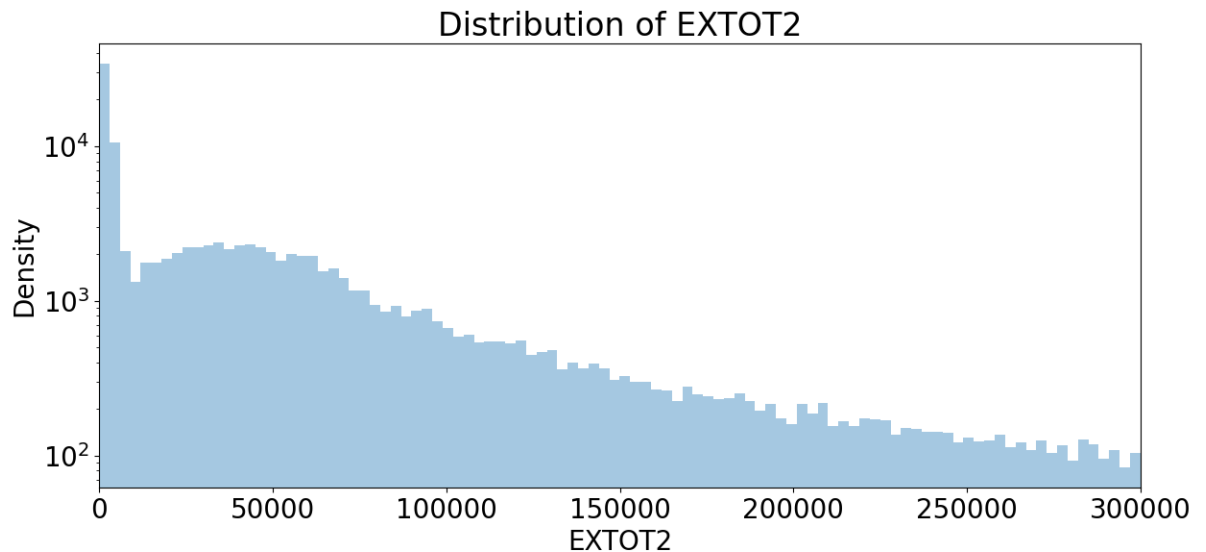
Visualization:



28) Field Name: EXTOT2

Description: The EXTOT2 column contains the assessed value of the property's exterior total area in the second assessment roll. It contains 130,828 non-null values and has a data population of 12.22%. The values are expressed in US dollars and range from 7 to 4,501,180,002.

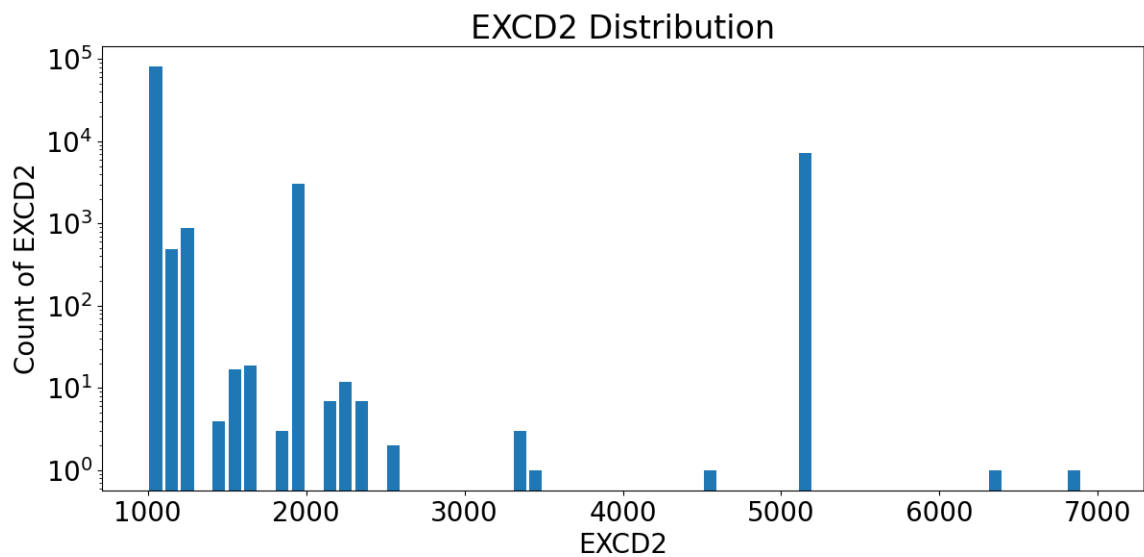
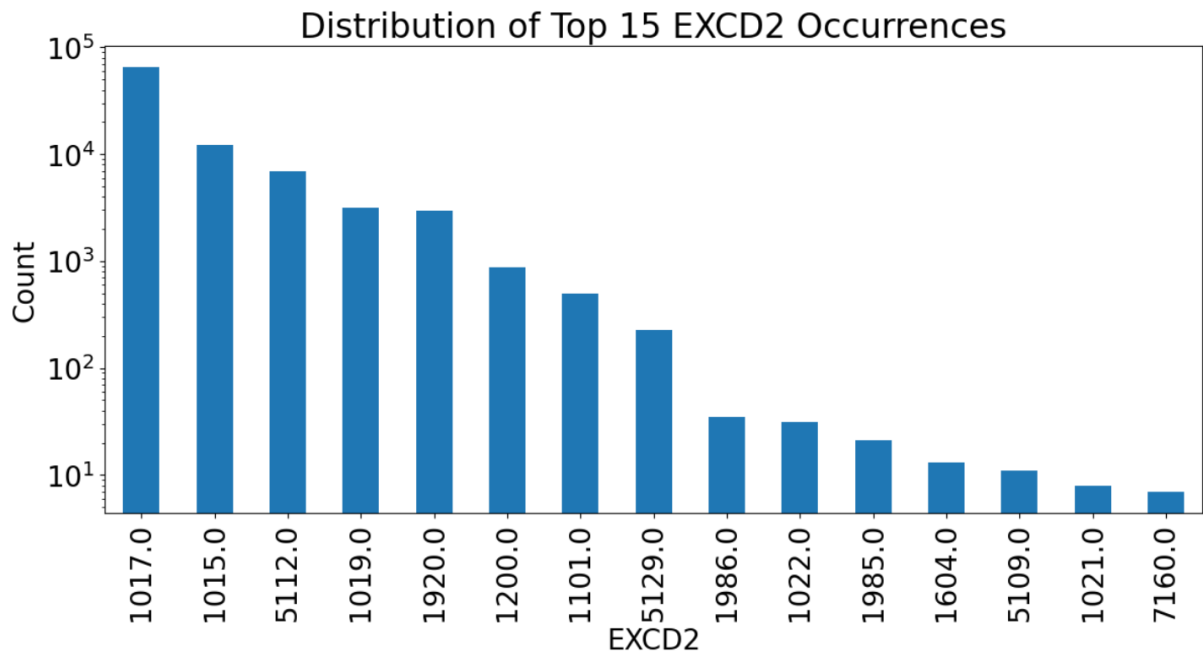
Visualization:



29) Field Name: EXCD2

Description: The EXCD2 column contains the second extended code description, which provides additional information on the tax class of the property. It is a numerical column that ranges from 101 to 7170 and has 166 unique values. It has 226,837 non-null values and 844,157 null values.

Visualization:



30) Field Name: PERIOD

Description: The PERIOD column indicates the fiscal period for which the property's assessed value is being reported. It has only one unique value, which is "FINAL".

31) Field Name: YEAR

Description: The YEAR column contains information about the fiscal year that the property was assessed. This column has 100% populated values and 0 zero values. It has only one unique value, which is "2010/11".

32) Field Name: VALTYPE

Description: The VALTYPE column contains information about the type of value associated with each property record. This column is fully populated and there are no zero values in this column. It has only one unique value, which is "AC-TR" which suggests that all the property records are based on the actual sale transactions and not on assessed values.