**DQR on NY Data**

Puhsin Huang

**[ Introduction ]**

**Data Name and Purpose:**

Property Valuation and Assessment Data

Data represent NYC properties assessments for purpose to calculate Property Tax, Grant eligible properties Exemptions and/or Abatements.

**Data Collection/Source:**

Data collected and entered into the system by various City employee, like Property Assessors, Property Exemption specialists, ACRIS reporting, Department of Building reporting, etc..

**Time period covered:** 2010/11

**Number of Fields:** 32

**Number of Records:** 1070994

**Data Types:** float64(13), int64(8), object(11)

**[ Summary ]**

Summary of the fields:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| RECORD | int64 | 1070994 | 100.0 | 1070994 | 0 |
| BBLE | object | 1070994 | 100.0 | 1070994 | 0 |
| B | int64 | 1070994 | 100.0 | 5 | 0 |
| BLOCK | int64 | 1070994 | 100.0 | 13984 | 0 |
| LOT | int64 | 1070994 | 100.0 | 6366 | 0 |
| EASEMENT | object | 4636 | 0.4 | 13 | 0 |
| OWNER | object | 1039249 | 97.0 | 863347 | 0 |
| BLDGCL | object | 1070994 | 100.0 | 200 | 0 |
| TAXCLASS | object | 1070994 | 100.0 | 11 | 0 |
| LTFRONT | int64 | 1070994 | 100.0 | 1297 | 169108 |
| LTDEPTH | int64 | 1070994 | 100.0 | 1370 | 170128 |
| EXT | object | 354305 | 33.1 | 4 | 0 |
| STORIES | float64 | 1014730 | 94.7 | 112 | 0 |
| FULLVAL | float64 | 1070994 | 100.0 | 109324 | 13007 |
| AVLAND | float64 | 1070994 | 100.0 | 70921 | 13009 |
| AVTOT | float64 | 1070994 | 100.0 | 112914 | 13007 |
| EXLAND | float64 | 1070994 | 100.0 | 33419 | 491699 |
| EXTOT | float64 | 1070994 | 100.0 | 64255 | 432572 |
| EXCD1 | float64 | 638488 | 59.6 | 130 | 0 |
| STADDR | object | 1070318 | 99.9 | 839281 | 0 |
| ZIP | float64 | 1041104 | 97.2 | 197 | 0 |
| EXMPTCL | object | 15579 | 1.5 | 15 | 0 |
| BLDFRONT | int64 | 1070994 | 100.0 | 612 | 228815 |
| BLDDEPTH | int64 | 1070994 | 100.0 | 621 | 228853 |
| AVLAND2 | float64 | 282726 | 26.4 | 58592 | 0 |
| AVTOT2 | float64 | 282732 | 26.4 | 111361 | 0 |
| EXLAND2 | float64 | 87449 | 8.2 | 22196 | 0 |
| EXTOT2 | float64 | 130828 | 12.2 | 48349 | 0 |
| EXCD2 | float64 | 92948 | 8.7 | 61 | 0 |
| PERIOD | object | 1070994 | 100.0 | 1 | 0 |
| YEAR | object | 1070994 | 100.0 | 1 | 0 |
| VALTYPE | object | 1070994 | 100.0 | 1 | 0 |

Statistics for numeric fields (#21):

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Field name | count | mean | std | min | 25% | 50% | 75% | max |
| LTFRONT | 1070994.0 | 36.635301 | 7.403284e+01 | 0.0 | 19.00 | 25.0 | 40.00 | 9.999000e+03 |
| LTDEPTH | 1070994.0 | 88.861594 | 7.639628e+01 | 0.0 | 80.00 | 100.0 | 100.00 | 9.999000e+03 |
| STORIES | 1014730.0 | 5.006918 | 8.365707e+00 | 1.0 | 2.00 | 2.0 | 3.00 | 1.190000e+02 |
| FULLVAL | 1070994.0 | 874264.505434 | 1.158243e+07 | 0.0 | 304000.00 | 447000.0 | 619000.00 | 6.150000e+09 |
| AVLAND | 1070994.0 | 85067.918672 | 4.057260e+06 | 0.0 | 9180.00 | 13678.0 | 19740.00 | 2.668500e+09 |
| AVTOT | 1070994.0 | 227238.168711 | 6.877529e+06 | 0.0 | 18374.00 | 25340.0 | 45438.00 | 4.668309e+09 |
| EXLAND | 1070994.0 | 36423.890692 | 3.981576e+06 | 0.0 | 0.00 | 1620.0 | 1620.00 | 2.668500e+09 |
| EXTOT | 1070994.0 | 91186.981682 | 6.508403e+06 | 0.0 | 0.00 | 1620.0 | 2090.00 | 4.668309e+09 |
| BLDFRONT | 1070994.0 | 23.042770 | 3.557970e+01 | 0.0 | 15.00 | 20.0 | 24.00 | 7.575000e+03 |
| BLDDEPTH | 1070994.0 | 39.922836 | 4.270715e+01 | 0.0 | 26.00 | 39.0 | 50.00 | 9.393000e+03 |
| AVLAND2 | 282726.0 | 246235.719265 | 6.178963e+06 | 3.0 | 5705.00 | 20145.0 | 62640.00 | 2.371005e+09 |
| AVTOT2 | 282732.0 | 713911.436173 | 1.165253e+07 | 3.0 | 33912.00 | 79962.5 | 240551.00 | 4.501180e+09 |
| EXLAND2 | 87449.0 | 351235.684273 | 1.080221e+07 | 1.0 | 2090.00 | 3048.0 | 31779.00 | 2.371005e+09 |
| EXTOT2 | 130828.0 | 656768.281904 | 1.607251e+07 | 7.0 | 2870.00 | 37062.0 | 106840.75 | 4.501180e+09 |

Most common field value of categorical fields (#11):

|  |  |  |
| --- | --- | --- |
| Field name | Field value | # times appeared |
| RECORD | Every value is unique | |
| BBLE | Every value is unique | |
| B | 4 = QUEENS | 358046 |
| BLOCK | 3944 | 3888 |
| LOT | 1 | 24367 |
| EASEMENT | E | 4148 |
| OWNER | PARKCHESTER PRESERVAT | 6020 |
| BLDGCL | R4 | 139879 |
| TAXCLASS | 1 | 660721 |
| EXT | G | 266970 |
| EXCD1 | 1017 | 425348 |
| STADDR | 501 SURF AVENUE | 902 |
| ZIP | 10314 | 24206 |
| EXMPTCL | X1 | 6912 |
| EXCD2 | 1017 | 65777 |
| PERIOD | FINAL | 1070994 |
| YEAR | 2010/11 | 1070994 |
| VALTYPE | AC-TR | 1070994 |

**[ Field Details ]**

Short description of the field, then a “picture”:

Numeric: a distribution, noting any outliers that are omitted

Categorical: a bar chart/histogram or a table of the top 10 to 15 field values

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| RECORD | int64 | 1070994 | 100.0 | 1070994 | 0 |

**Description:**

Number to uniquely identify each record. Values in this field are all unique, ranging from 1 to 1070994. There’s not much of indication in the graph so the graph of this field is skipped.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| BBLE | object | 1070994 | 100.0 | 1070994 | 0 |

**Description:**

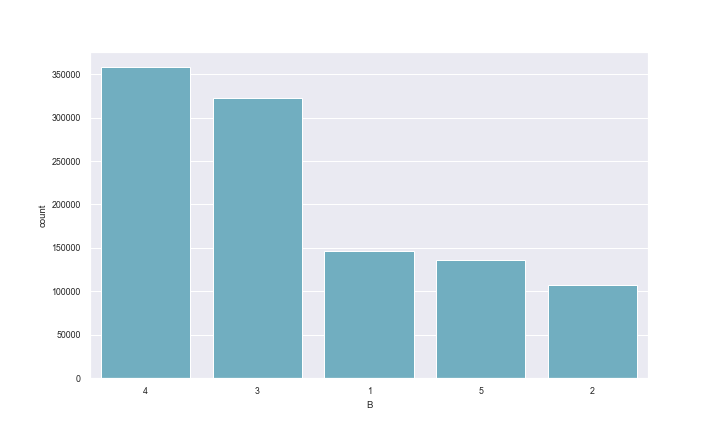
Concatenation of AV\_BORO, AV\_BLOCK, AV\_LOT, AV\_EASEMENT. Values in this field are all unique. There’s not much of indication in the graph so the graph of this field is skipped.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| B | int64 | 1070994 | 100.0 | 5 | 0 |

**Description:**

BORO CODES, there are five unique values: 1,2,3,4,5. The meaning are as follows:

|  |
| --- |
| 1 = MANHATTAN |
| 2 = BRONX |
| 3 = BROOKLYN |
| 4 = QUEENS |
| 5 = STATEN ISLAND |

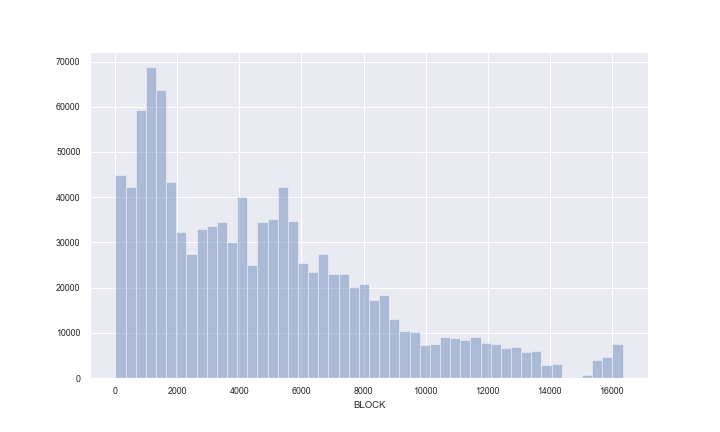


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| BLOCK | int64 | 1070994 | 100.0 | 13984 | 0 |

**Description:**

VALID BLOCK RANGES BY BORO. The meaning are as follows:

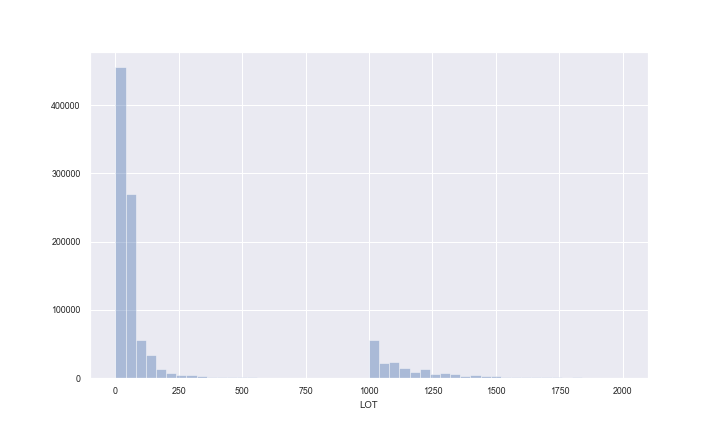
|  |
| --- |
| MANHATTAN 1 TO 2,255 |
| BRONX 2,260 TO 5,958 |
| BROOKLYN 1 TO 8,955 |
| QUEENS 1 TO 16,350 |
| STATEN ISLAND 1 TO 8,050 |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| LOT | int64 | 1070994 | 100.0 | 6366 | 0 |

**Description:**

UNIQUE # WITHIN BORO/BLOCK; noted that the values are ranging from 1 to 9978. The below graph only shows a portion of the data.



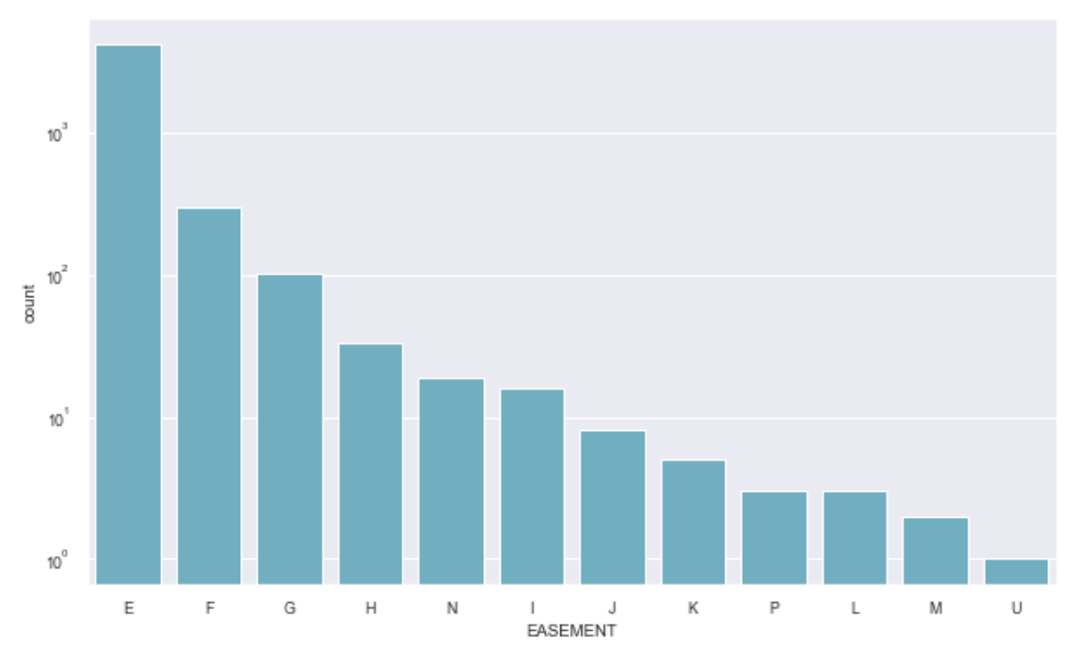
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EASEMENT | object | 4636 | 0.4 | 13 | 0 |

**Description:**

EASEMENT is a 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:

|  |
| --- |
| 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 |

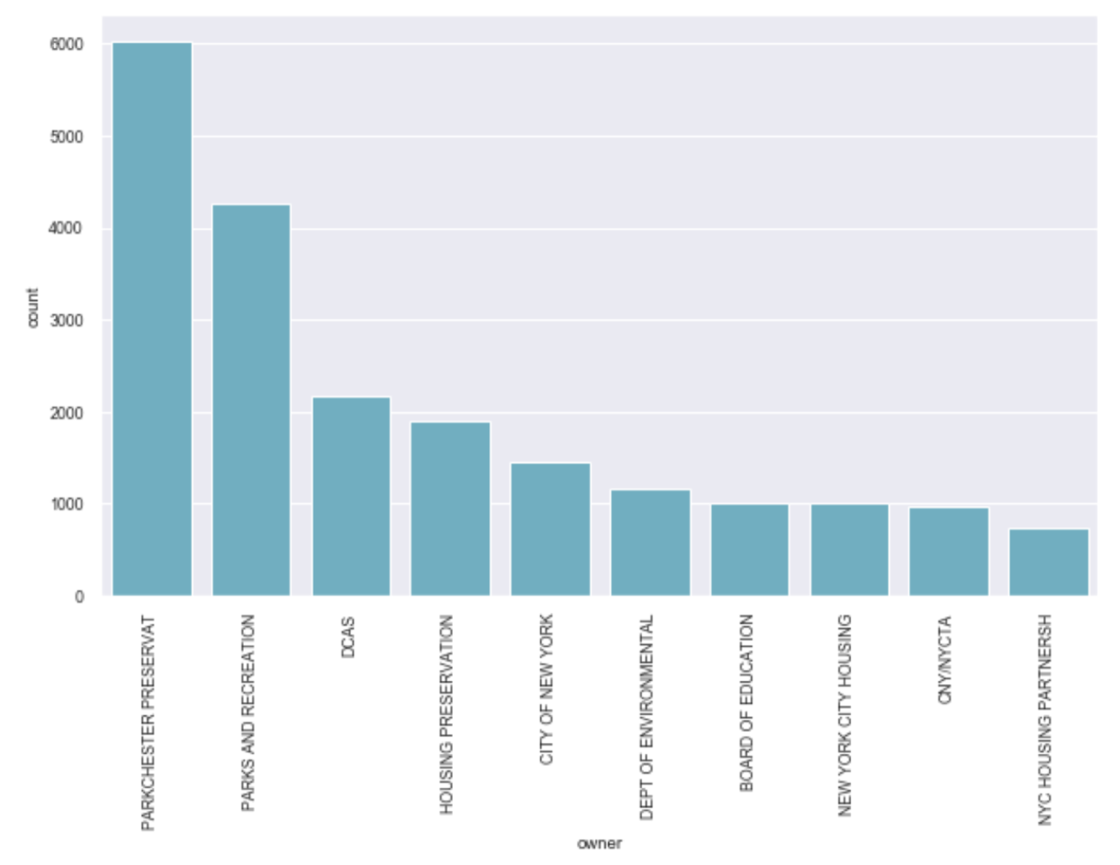
Among the 4636 lots that have easement, 4181 of them have a land easement. When we took log on y-scale, the graph became much more interpretable.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| OWNER | object | 1039249 | 97.0 | 863347 | 0 |

**Description:**

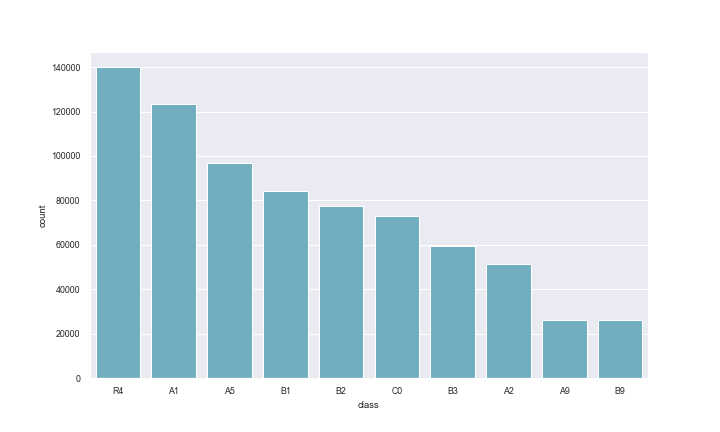
Owner’s Name. Below are the top 10 owners that appear in the dataset the most:



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| BLDGCL | object | 1070994 | 100.0 | 200 | 0 |

**Description:**

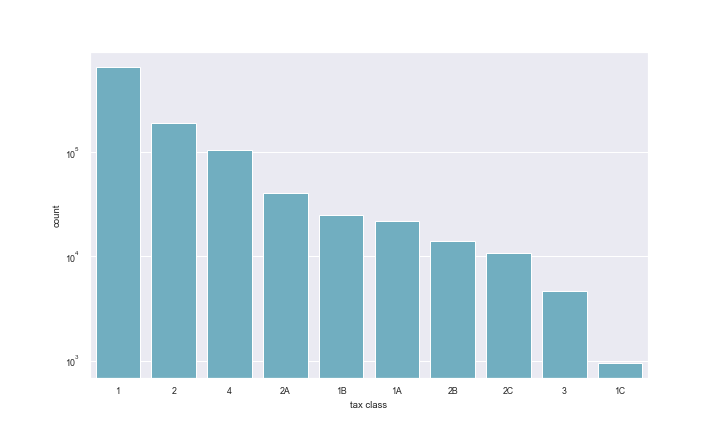
Building Class. There is a direct correlation between the Building Class and the Tax Class.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| TAXCLASS | object | 1070994 | 100.0 | 11 | 0 |

**Description:**

During the final period, contains the tax class from the CBN period. Below graph I took log on the y-axis:

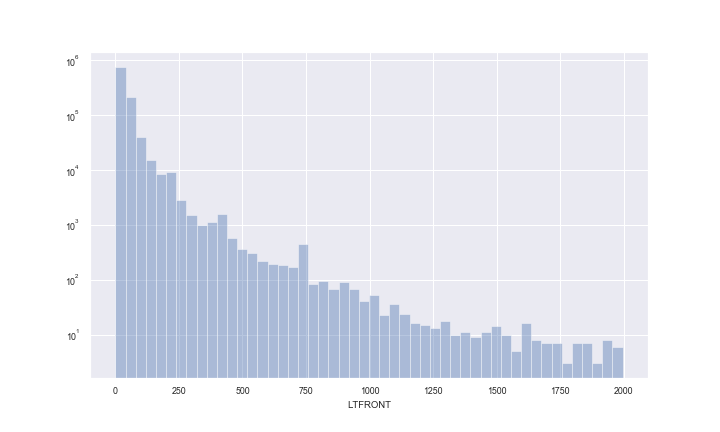


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| LTFRONT | int64 | 1070994 | 100.0 | 1297 | 169108 |

**Description:**

Lot Frontage in feet (lot width). The maximum value of this field is 9999.

To get a closer look, I limited the LFTRONT to be under 2000 and took log on y-axis.

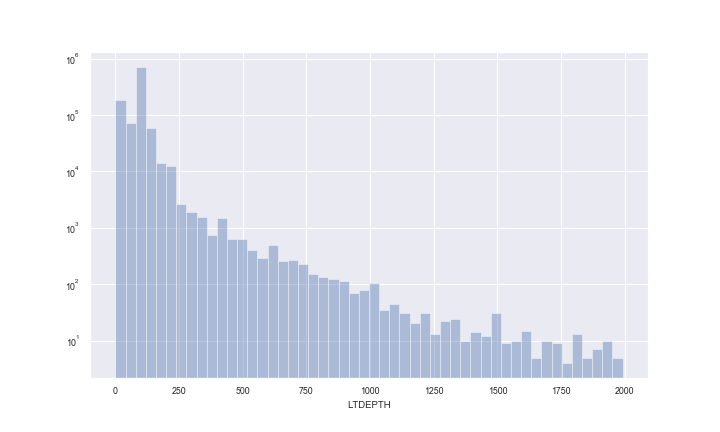


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| LTDEPTH | int64 | 1070994 | 100.0 | 1370 | 170128 |

**Description:**

Lot Depth in feet. The maximum value of this field is 9999.

To get a closer look, I limited the LTDEPTH to be under 2000 and took log on y-axis.

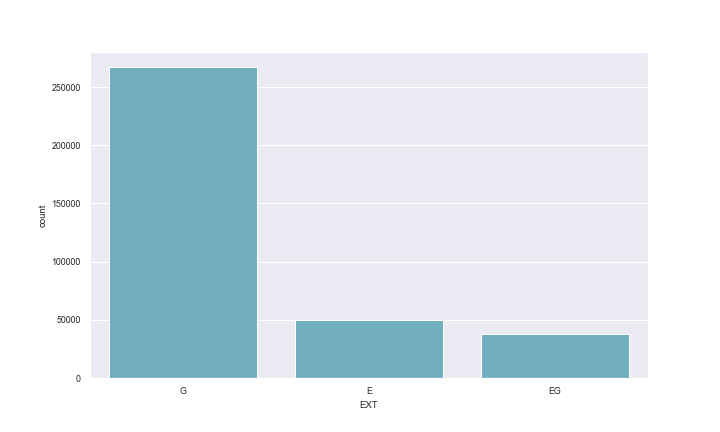


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EXT | object | 354305 | 33.1 | 4 | 0 |

**Description:**

Extension indicator; The meaning of the values are as follows:

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

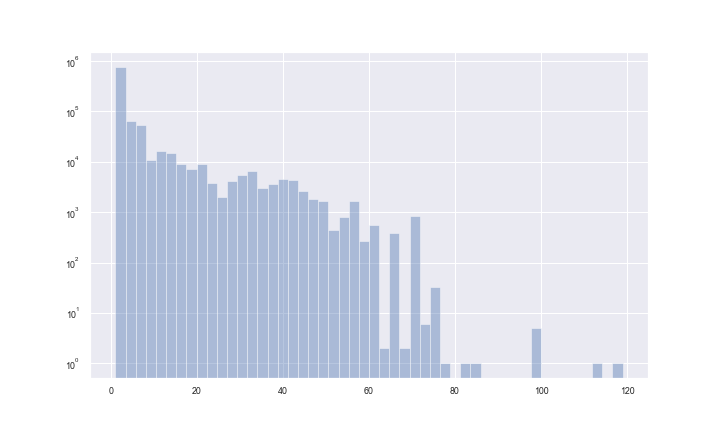


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| STORIES | float64 | 1014730 | 94.7 | 112 | 0 |

**Description:**

The number of stories in building ( # of Floors)

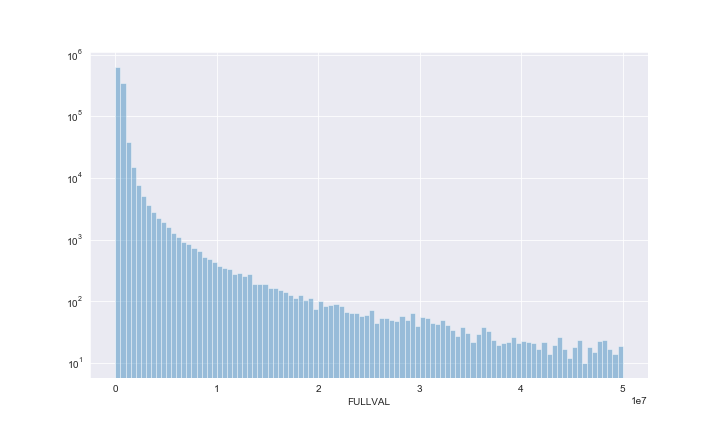
After taking log on y-axis, the graph looks like below:



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| FULLVAL | float64 | 1070994 | 100.0 | 109324 | 13007 |

**Description:**

Market value. I took log on y-axis to make the graph more interpretable. Also, to get a closer look, I limited the FULLVAL to be under 50000000.

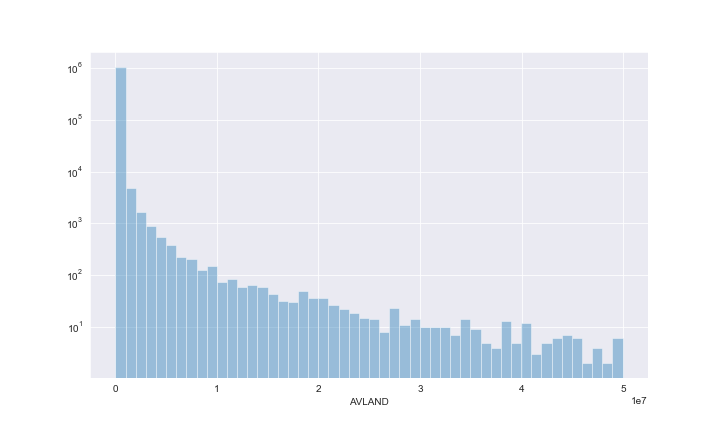


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| AVLAND | float64 | 1070994 | 100.0 | 70921 | 13009 |

**Description:**

Actual land value. I took log on y-axis to make the graph more interpretable.

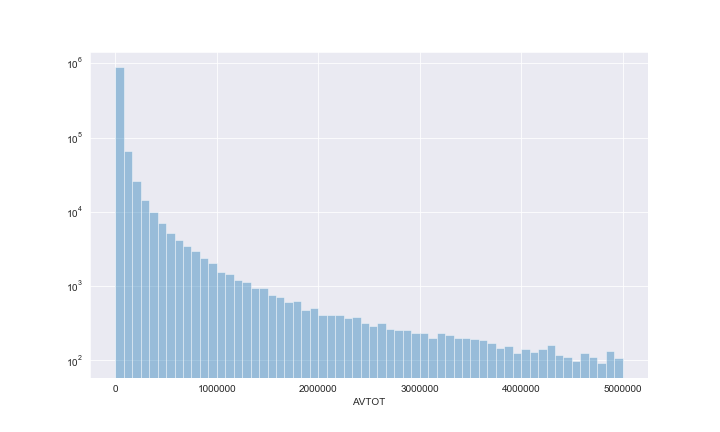
In addition, to get a closer look, I limited the AVLAND to be under 50000000.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| AVTOT | float64 | 1070994 | 100.0 | 112914 | 13007 |

**Description:**

Actual total value. I took log on y-axis to make the graph more interpretable. Also, to get a closer look, I limited the AVTOT to be under 5000000.

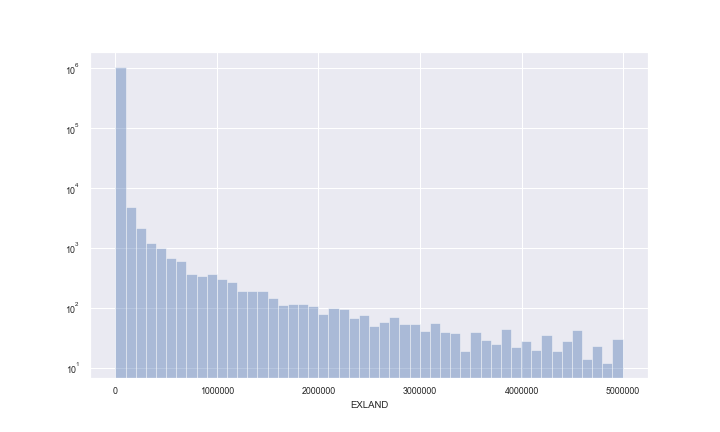


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EXLAND | float64 | 1070994 | 100.0 | 33419 | 491699 |

**Description:**

Actual exempt land value. I took log on y-axis to make the graph more interpretable.

Also, to get a closer look, I limited the EXLAND to be under 5000000.

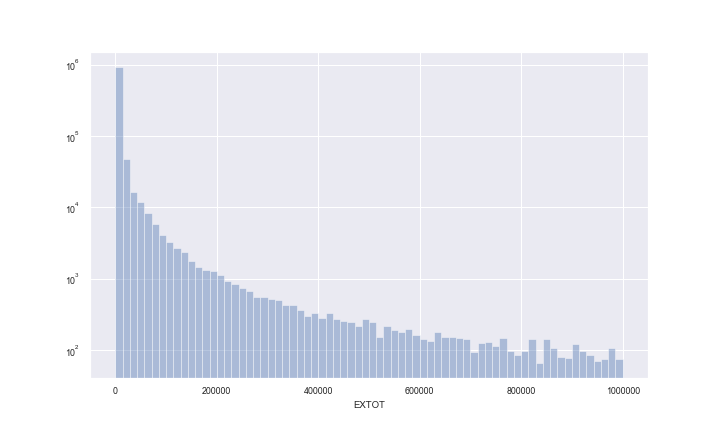


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EXTOT | float64 | 1070994 | 100.0 | 64255 | 432572 |

**Description:**

Actual exempt land total. I took log on y-axis to make the graph more interpretable.

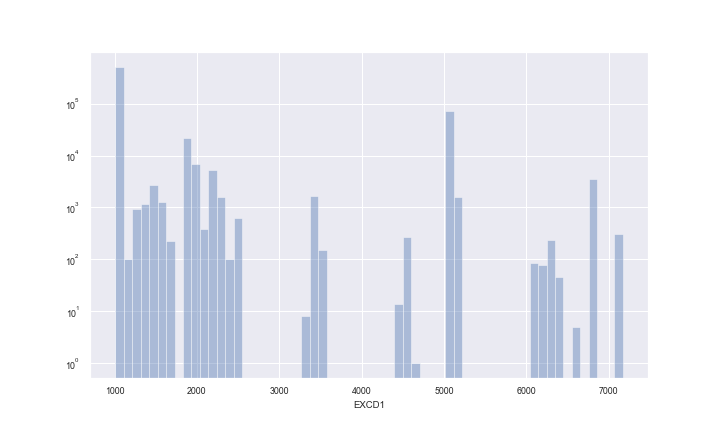
Also, to get a closer look, I limited the EXTOT to be under 1000000.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EXCD1 | float64 | 638488 | 59.6 | 130 | 0 |

**Description:**

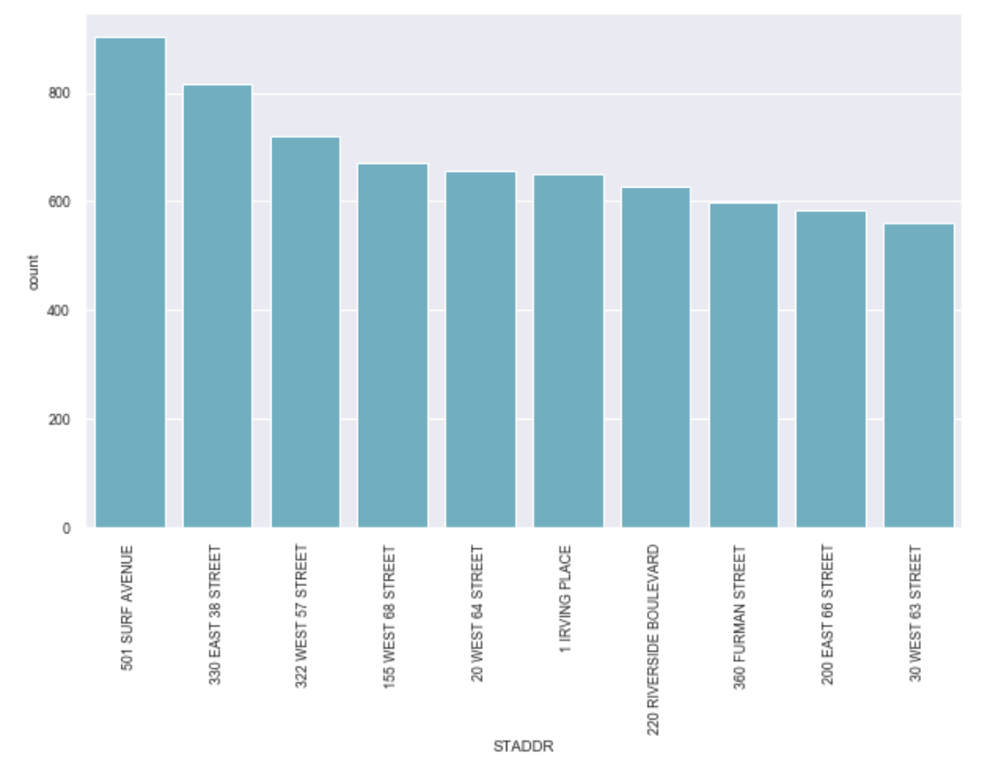
Exemption code 1. I took log on y-axis to make the graph more interpretable.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| STADDR | object | 1070318 | 99.9 | 839281 | 0 |

**Description:**

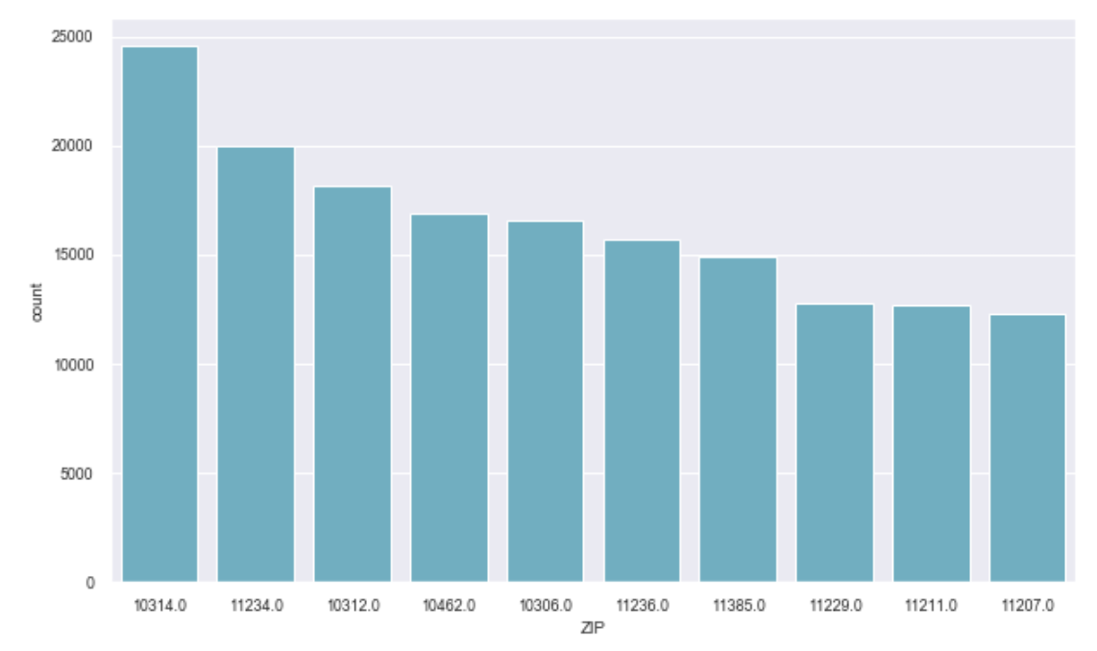
Street addresses. Below graph shows the top 10 addresses that appear the most in the dataset.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| ZIP | float64 | 1041104 | 97.2 | 197 | 0 |

**Description:**

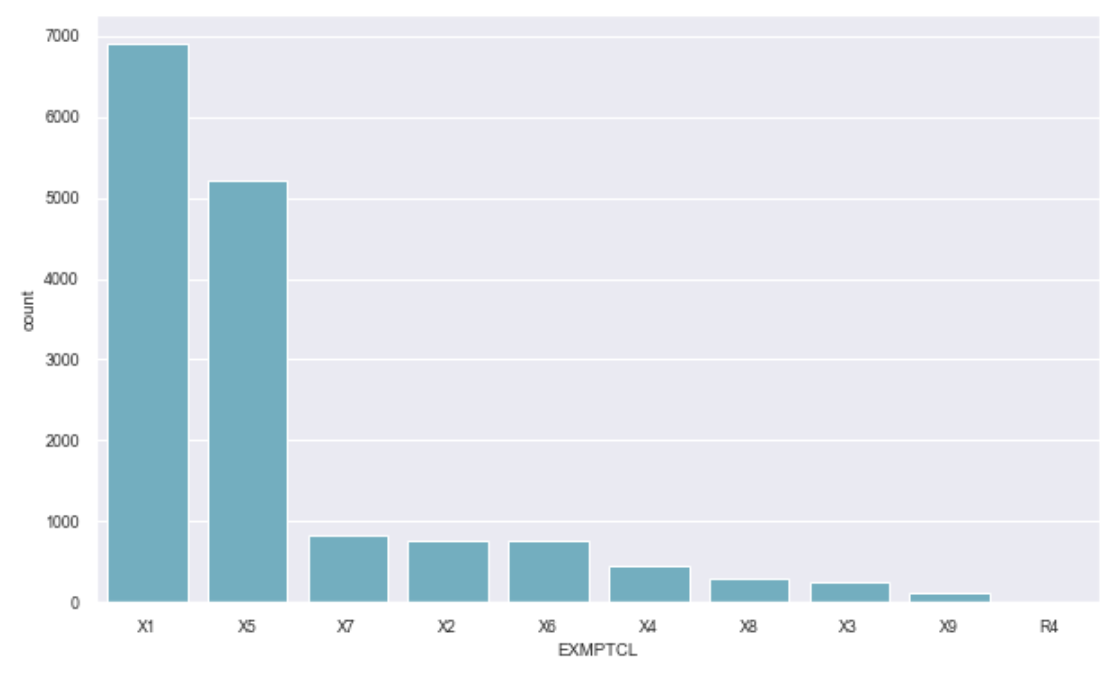
Postal Zip code of the property. Although zip code has a data type of float65, it is a categorical rather than a numerical field. Hence, I use countplot rather than distplot here. Below graph shows the top 10 Zip code that appear the most in the dataset.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EXMPTCL | object | 15579 | 1.5 | 15 | 0 |

**Description:**

Exempt Class ('X1 - X9') used for fully exempt properties only.

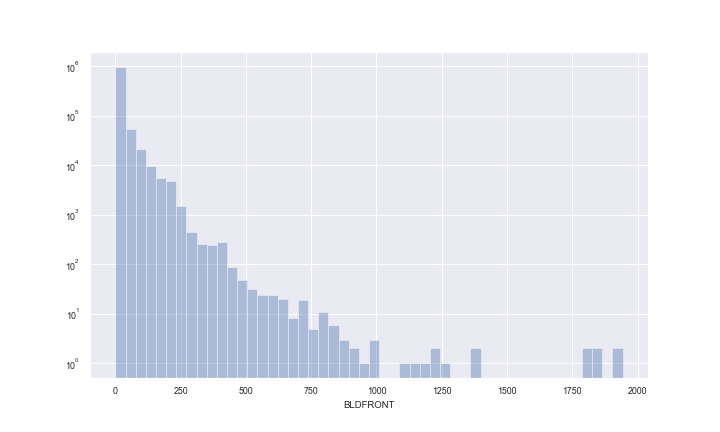


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| BLDFRONT | int64 | 1070994 | 100.0 | 612 | 228815 |

**Description:**

Building Frontage (width) in feet. I took log on y-axis to make the graph more interpretable.

To get a closer look, I limited the BLDFRONT to be under 2000.

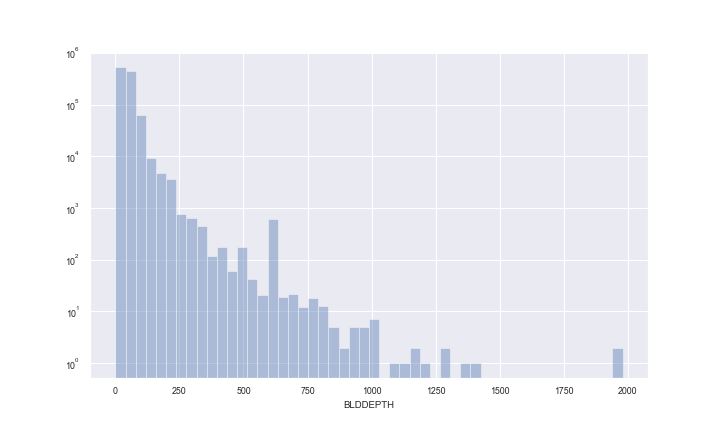


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| BLDDEPTH | int64 | 1070994 | 100.0 | 621 | 228853 |

**Description:**

Lot Depth in feet. I took log on y-axis to make the graph more interpretable.

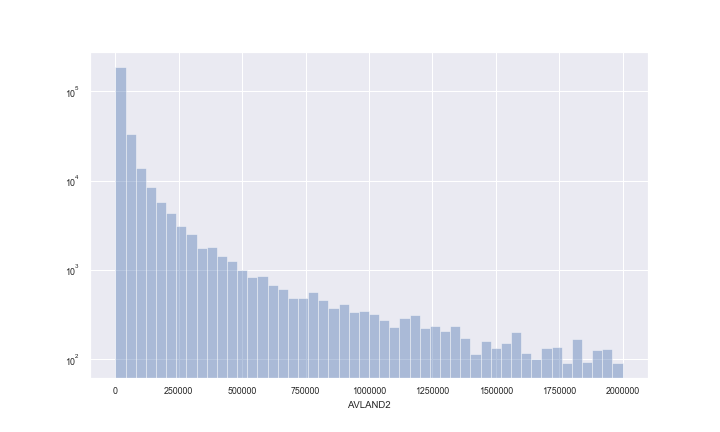
To get a closer look, I limited the BLDDEPTH to be under 2000.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| AVLAND2 | float64 | 282726 | 26.4 | 58592 | 0 |

**Description:**

Transitional land value. I took log on y-axis to make the graph more interpretable.

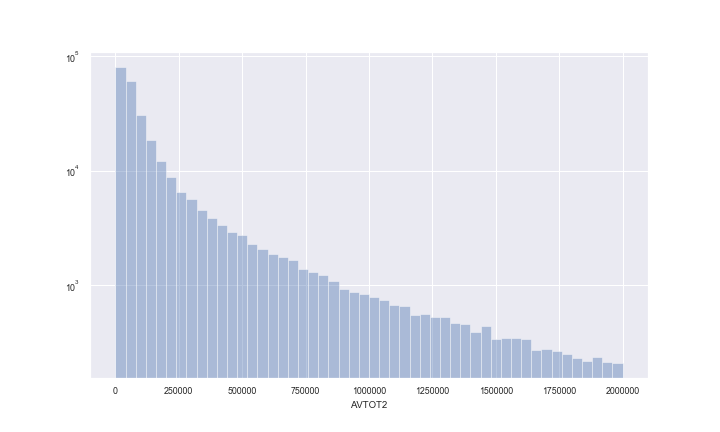
To get a closer look, I limited the AVLAND2 to be under 2000000

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| AVTOT2 | float64 | 282732 | 26.4 | 111361 | 0 |

**Description:**

Transitional total value. I took log on y-axis to make the graph more interpretable.

To get a closer look, I limited the AVTOT2 to be under 2000000.

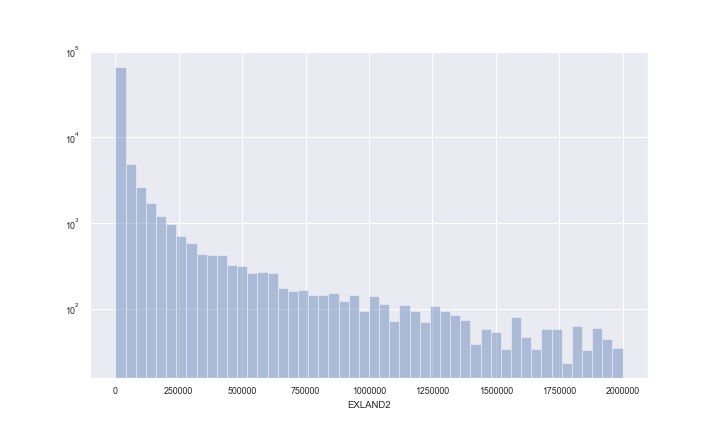


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EXLAND2 | float64 | 87449 | 8.2 | 22196 | 0 |

**Description:**

Transitional exempt land value. I took log on y-axis to make the graph more interpretable.

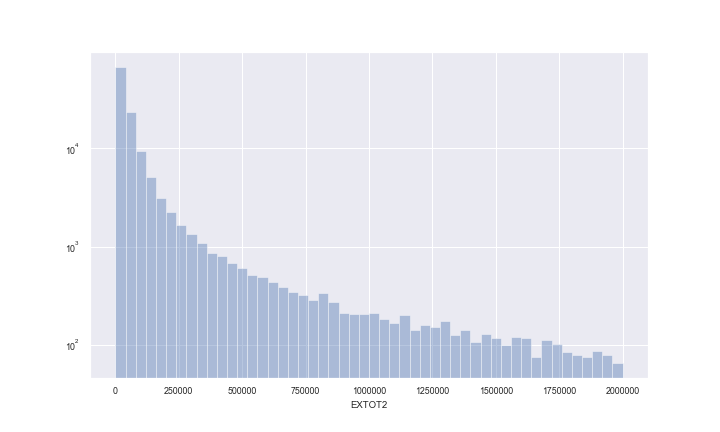
To get a closer look, I limited the EXLAND2 to be under 2000000.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EXTOT2 | float64 | 130828 | 12.2 | 48349 | 0 |

**Description:**

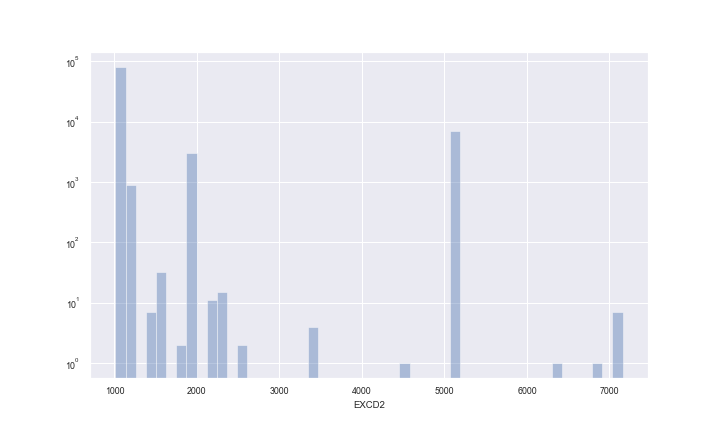
Transitional exempt land total. I took log on y-axis to make the graph more interpretable. To get a closer look, I limited the EXTOT2 to be under 2000000.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| EXCD2 | float64 | 92948 | 8.7 | 61 | 0 |

**Description:**

Exemption code 2. I took log on y-axis to make the graph more interpretable.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| PERIOD | object | 1070994 | 100.0 | 1 | 0 |

**Description:**

Assessment period when file was created. Only one unique value - Final. The graph is skipped here.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| YEAR | object | 1070994 | 100.0 | 1 | 0 |

**Description:**

Assessment year. Only one unique value – 2010/11. The graph is skipped here.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field name | Field type | # records with a value | % populated | # unique values | # records with value zero |
| VALTYPE\* | object | 1070994 | 100.0 | 1 | 0 |

Need to clarify the definition of this field with the client. Only one unique value - AC-TR. The graph is skipped here.