Data Quality Report, Project 1

Chengjun Liu, 7116099552

# Data Description

Dataset Name: Property Valuation and Assessment Data of New York City Properties

Dataset Source: New York City Department of Finance

Release Date: November 17th, 2010

Number of Columns: 32

Number of Records: 1070994

Summary

Overview:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Data Type | Non-null Records | Description |
| RECORD | Numeric | 1070994 | Number of Records |
| BBLE | Text | 1070994 | Borough & Block codes |
| B | Number | 1070994 | Borough valid values 1-5 |
| BLOCK | Number | 1070994 | Block Name & Number |
| LOT | Number | 1070994 | Lot Number |
| EASEMENT | Text | 4636 | Whether having easement in the lot |
| OWNER | Text | 1039249 | Owner name |
| BLDGCL | Text | 1070994 | Building class |
| TAXCLASS | Text | 1070994 | Tax class |
| LTFRONT | Number | 1070994 | Lot width |
| LTDEPTH | Number | 1070994 | Lot depth |
| EXT | Text | 354305 | Extension indicator |
| STORIES | Number | 1014730 | Number of stories in building |
| FULLVAL | Number | 1070994 | Market value |
| AVLAND | Number | 1070994 | Actual land value |
| AVTOT | Number | 1070994 | Actual total value |
| EXLAND | Number | 1070994 | Actual exempt land value |
| EXTOT | Number | 1070994 | Actual exempt land total |
| EXCD1 | Number | 638488 | Exemption code 1 |
| STADDR | Text | 1070318 | Street address |
| ZIP | Number | 1041104 | Zip code |
| EXMPTCL | Text | 15579 | Exempt class |
| BLDFRONT | Number | 1070994 | Building width |
| BLDDEPTH | Number | 1070994 | Building depth |
| AVLAND2 | Number | 282726 | Transitional land value |
| AVTOT2 | Number | 282732 | Transitional total value |
| EXLAND2 | Number | 87449 | Transitional exempt land value |
| EXTOT2 | Number | 130828 | Transitional exempt land total |
| EXCD2 | Number | 92948 | Exemption code 2 |
| PERIOD | Text | 1070994 | Assessment period when file was created |
| YEAR | Text | 1070994 | Assessment year |
| VALTYPE | Text | 1070994 | Value type (No meaning, no used) |

Summary of Numeric Variable:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable Name | Non-null Record | Non-null Percentage | Mean | Standard Variation | Min | Max |
| LTFRONT | 1070994 | 1 | 36.6353 | 74.03284 | 0 | 9999 |
| LTDEPTH | 1070994 | 1 | 88.86159 | 76.39628 | 0 | 9999 |
| STORIES | 1014730 | 0.947466 | 5.006918 | 8.365707 | 1 | 119 |
| FULLVAL | 1070994 | 1 | 874264.5 | 11582430 | 0 | 6150000000 |
| AVLAND | 1070994 | 1 | 85067.92 | 4057260 | 0 | 2668500000 |
| AVTOT | 1070994 | 1 | 227238.2 | 6877529 | 0 | 4668309000 |
| EXLAND | 1070994 | 1 | 36423.89 | 3981576 | 0 | 2668500000 |
| EXTOT | 1070994 | 1 | 91186.98 | 6508403 | 0 | 4668309000 |
| BLDFRONT | 1070994 | 1 | 23.04277 | 35.5797 | 0 | 7575 |
| BLDDEPTH | 1070994 | 1 | 39.92284 | 42.70715 | 0 | 9393 |
| AVLAND2 | 282726 | 0.263985 | 246235.7 | 6178963 | 3 | 2371005000 |
| AVTOT2 | 282732 | 0.26399 | 713911.4 | 11652530 | 3 | 4501180000 |
| EXLAND2 | 87449 | 0.081652 | 351235.7 | 10802210 | 1 | 2371005000 |
| EXTOT2 | 130828 | 0.122156 | 656768.3 | 16072510 | 7 | 4501180000 |

Summary of Categorical Variables:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Name | Non-null Record | Non-null Percentage | Number of Unique Values | Most frequent category |
| RECORD | 1070994 | 1 | 1070994 | None |
| BBLE | 1070994 | 1 | 1070994 | None |
| B | 1070994 | 1 | 5 | 4 |
| BLOCK | 1070994 | 1 | 13984 | 3944 |
| LOT | 1070994 | 1 | 6366 | 1 |
| EASEMENT | 4636 | 0.004329 | 12 | E |
| OWNER | 1039249 | 0.970359 | 863346 | PARKCHESTER PRESERVAT |
| BLDGCL | 1070994 | 1 | 200 | R4 |
| TAXCLASS | 1070994 | 1 | 11 | 1 |
| EXT | 354305 | 0.330819 | 3 | G |
| EXCD1 | 638488 | 0.596164 | 129 | 1017 |
| STADDR | 1070318 | 0.999369 | 839280 | 501 SURF AVENUE |
| ZIP | 1041104 | 0.972091 | 196 | 10314 |
| EXMPTCL | 15579 | 0.014546 | 14 | X1 |
| EXCD2 | 92948 | 0.086787 | 60 | 1017 |
| PERIOD | 1070994 | 1 | 1 | FINAL |
| YEAR | 1070994 | 1 | 1 | 2010/11 |
| VALTYPE | 1070994 | 1 | 1 | AC-TR |

Data Field Exploration

Variable 1

Name: RECORD

Description: Unique number of records, used as primary key.

Variable 2

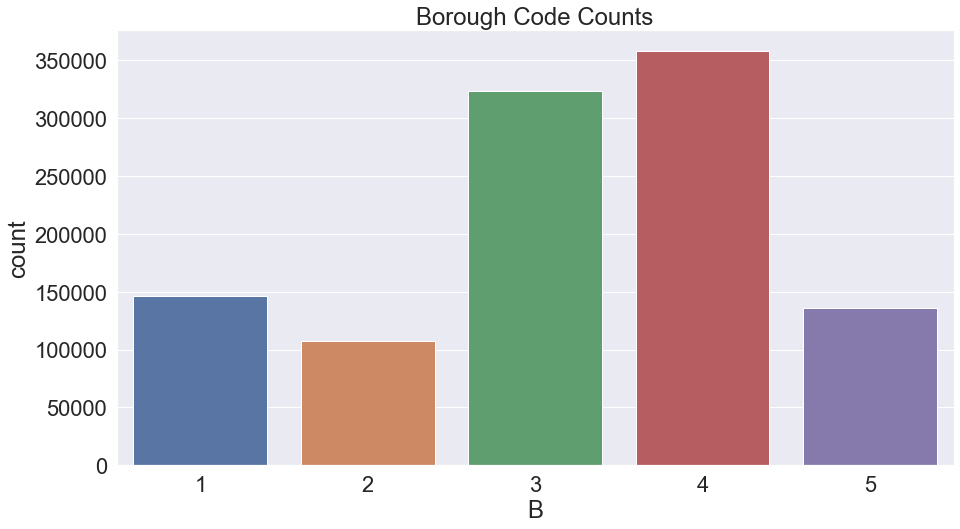
Name: BBLE

Description: Borough & block codes, consisting of borough, block code, unique number within borough, and block easement.

Variable 3:

Name: B

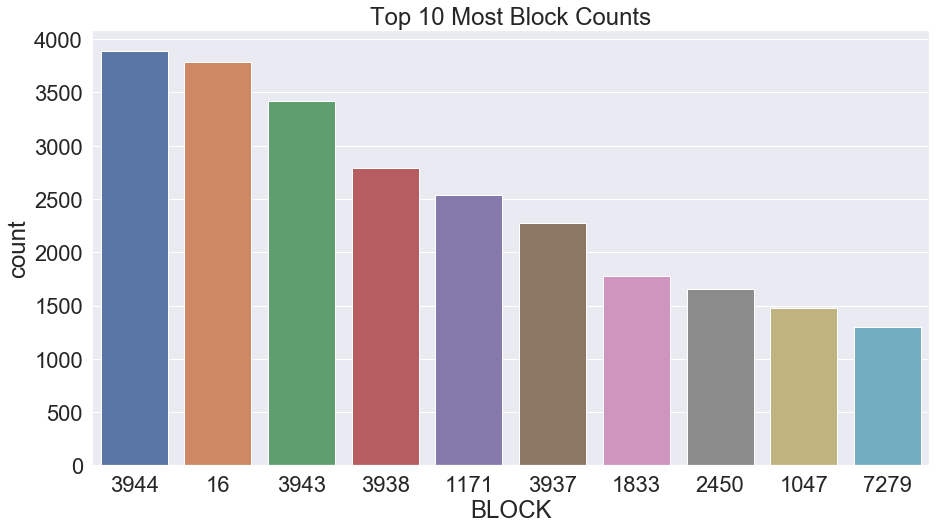
Description: Borough codes with valid values of 1-5.



Variable 4

Name: BLOCK

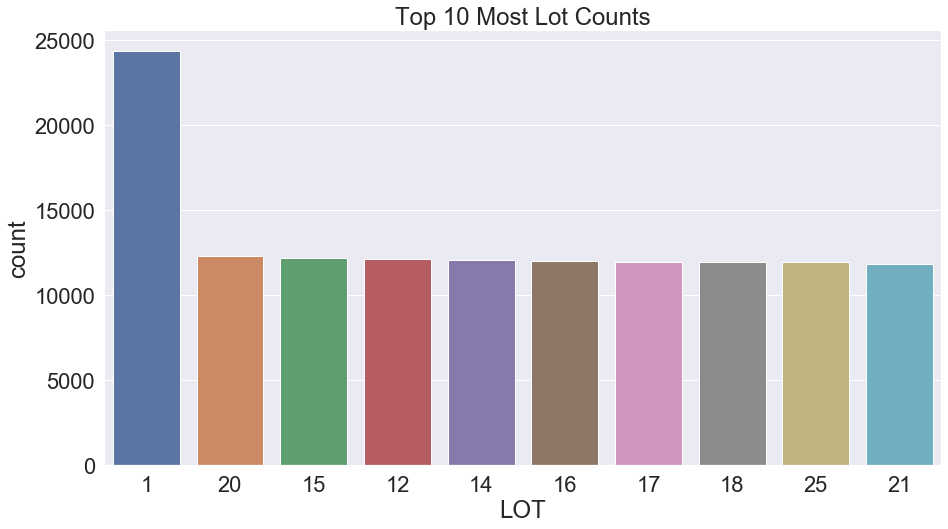
Description: Valid block number.



Variable 5

Name: LOT

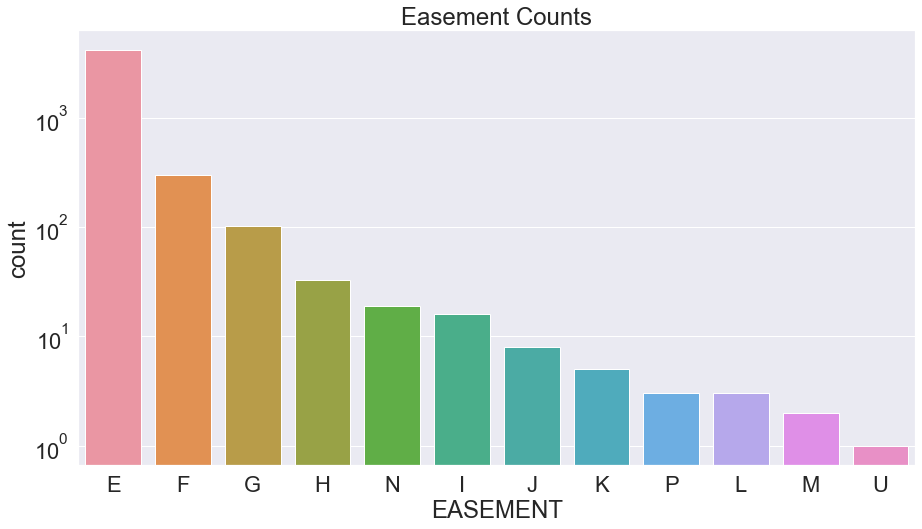
Description: Unique lot number within borough and block



Variable 6

Name: EASEMENT

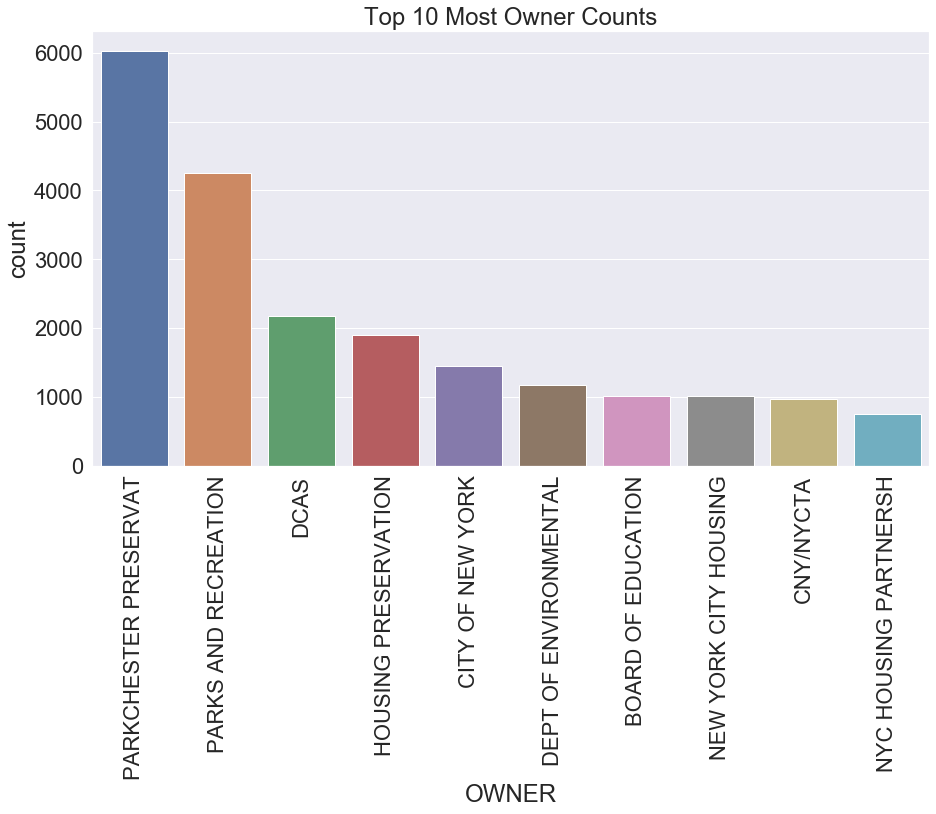
Description: Easement in the lot, marked with characters.



Variable 7

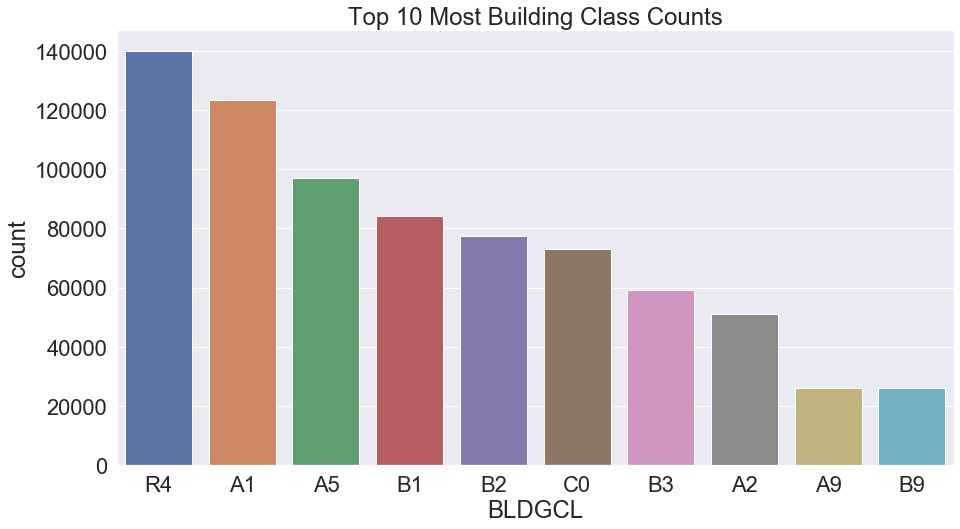
Name: OWNER

Description: Owner’s name.



Variable 8

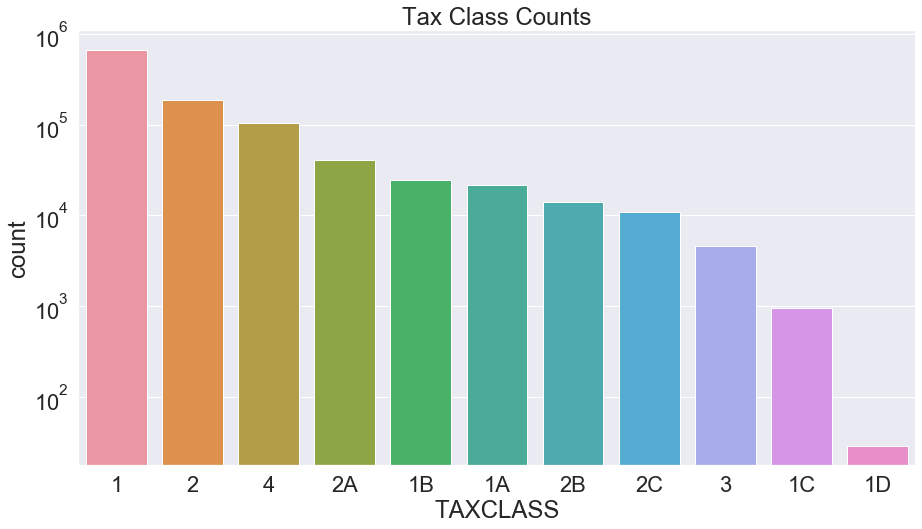
Name: BLDGCL

Description: Bu  ilding class.

Variable 9

Name: TAXCLASS

Description: Tax class.

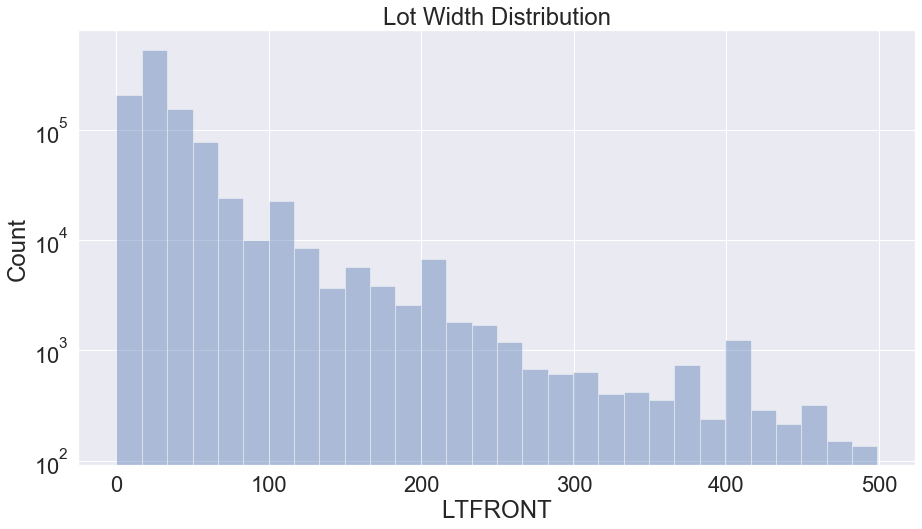


Variable 10:

Name: LTFRONT

Description: Lot width in feet.

We plot the lot width distribution under 500, accounting for 99.753% of all records.

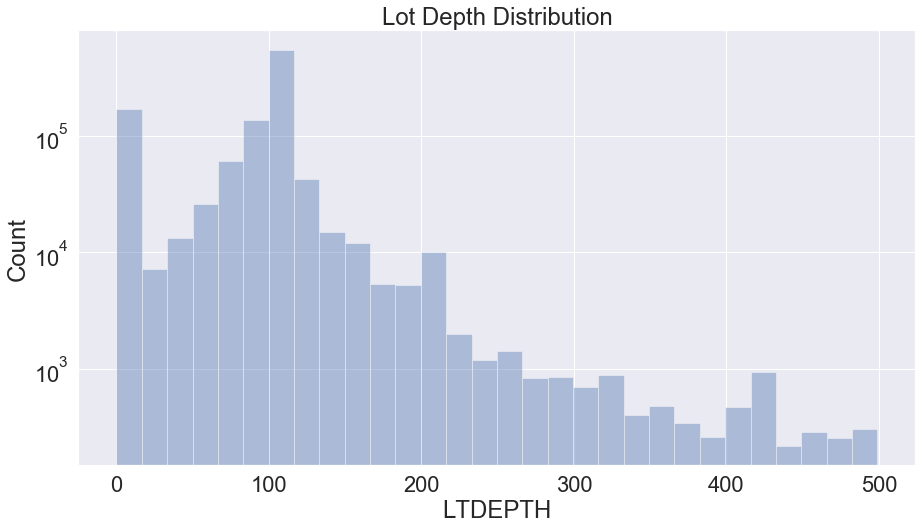


Variable 11

Name: LTDEPTH

Description: Lot depth in feet

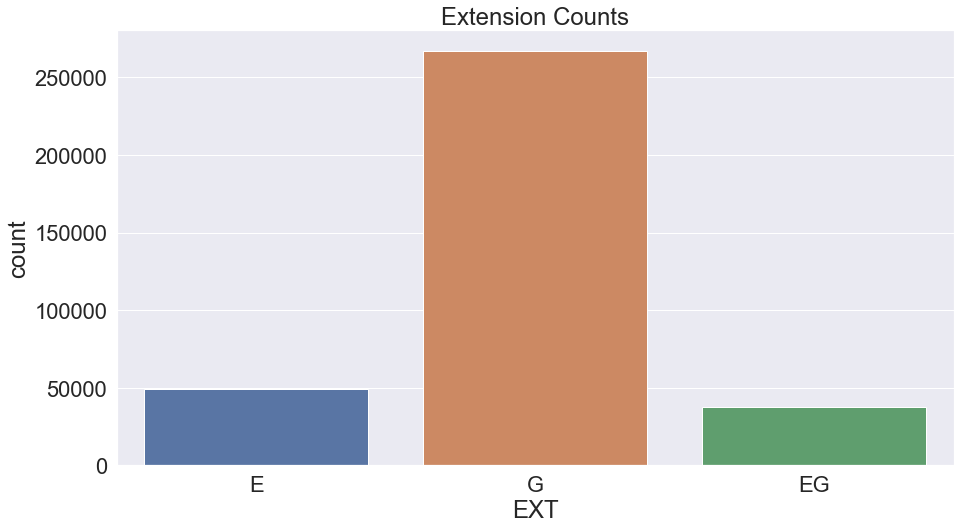
We plot the lot depth distribution under 500, accounting for 99.666% of all records



Variable 12

Name: EXT

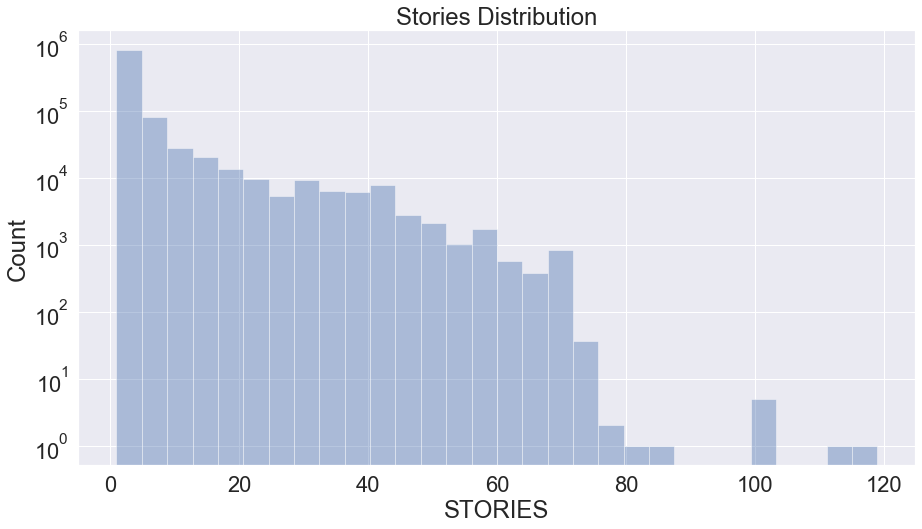
Description: Extension indicator. ‘E’: extension, ‘G’: garage, ‘EG’: extension and garage.



Variable 13

Name: STORIES

Description: Number of stories in building

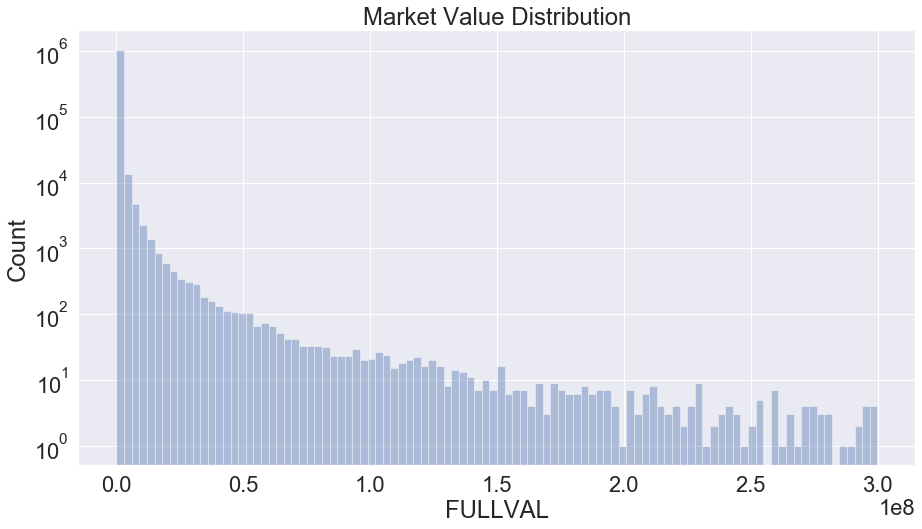
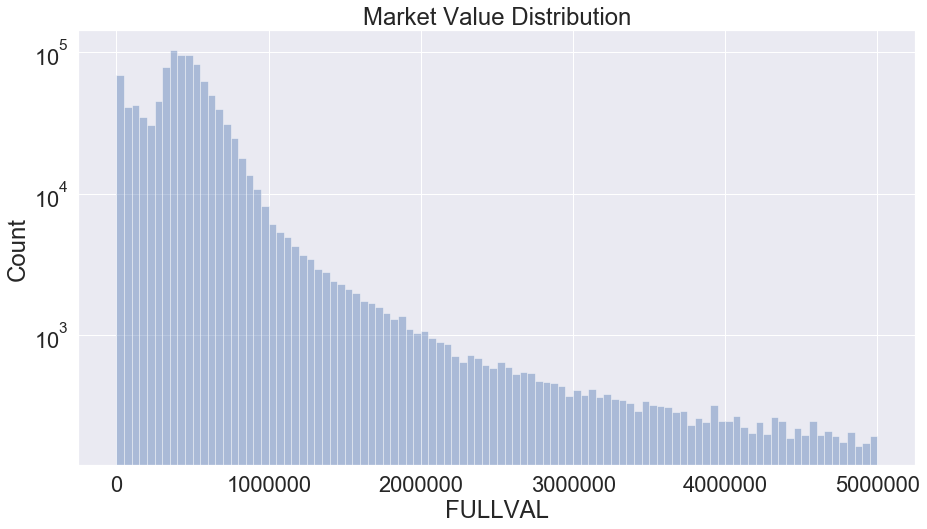


Variable 14

Name: FULLVAL

Description: Full market value

We plot the distribution of market values under 30000000/5000000, accounting for 99.778%/98.474% of all records.

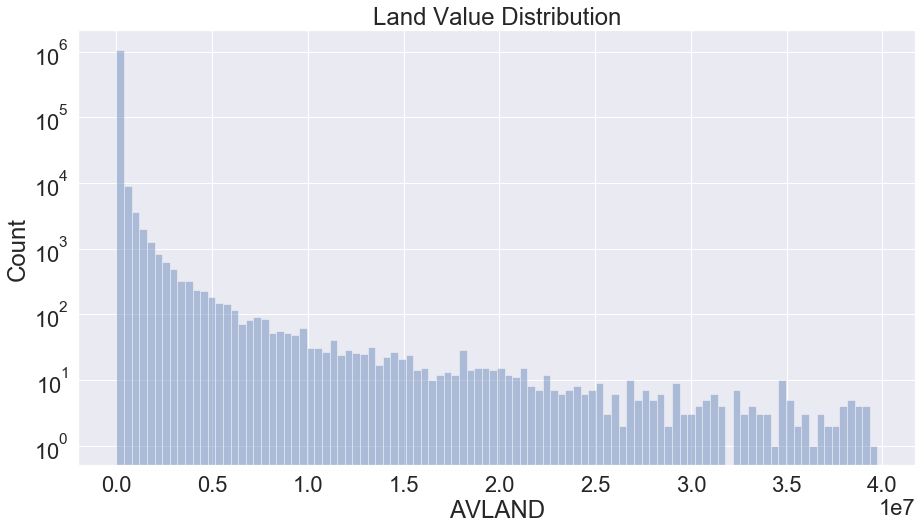
 

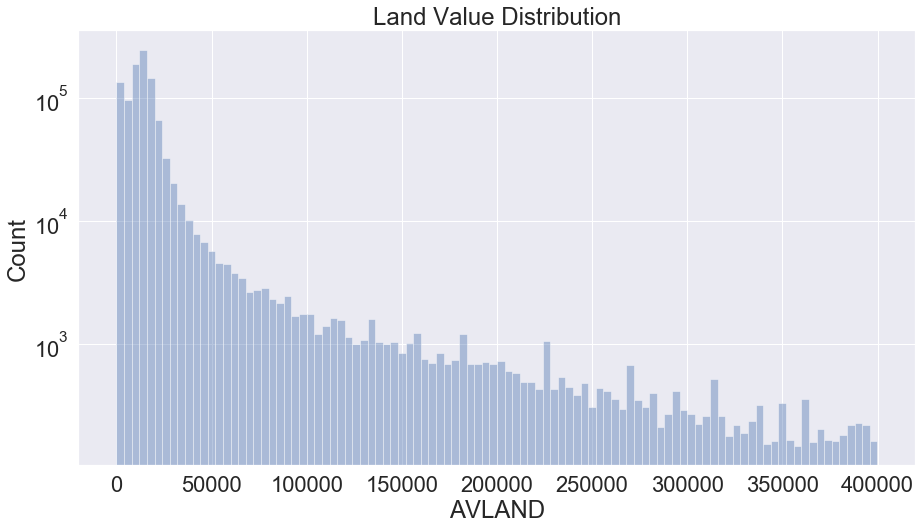
Variable 15

Name: AVLAND

Description: Actual land value

We plot the distribution of land values under 40000000/400000, accounting for 99.985%/98.030% of all records.



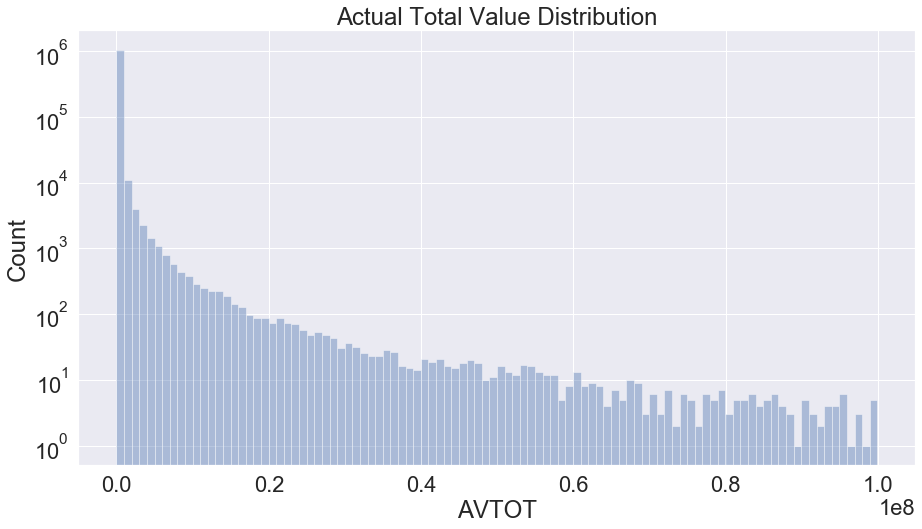
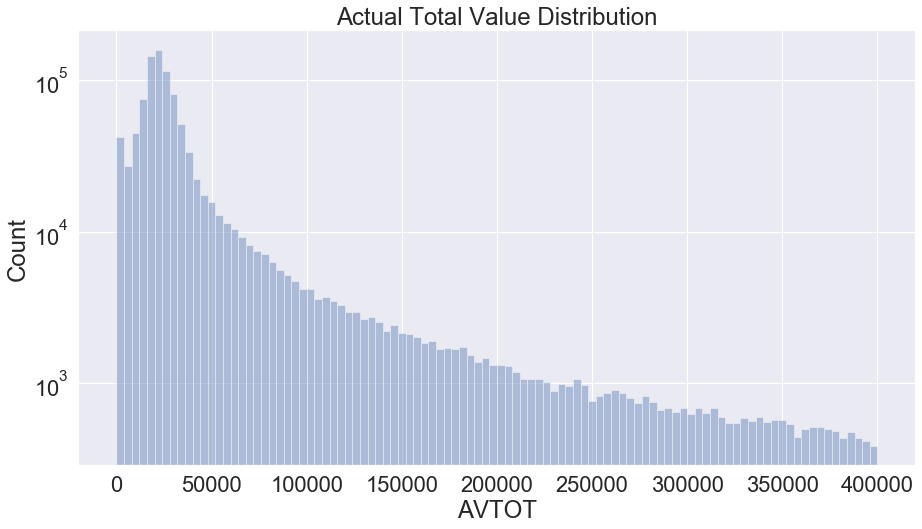


Variable 16

Name: AVTOT

Description: Actual total value

We plot the distribution of actual total values under 100000000/400000 accounting for 99.984%/94.946% of all records.

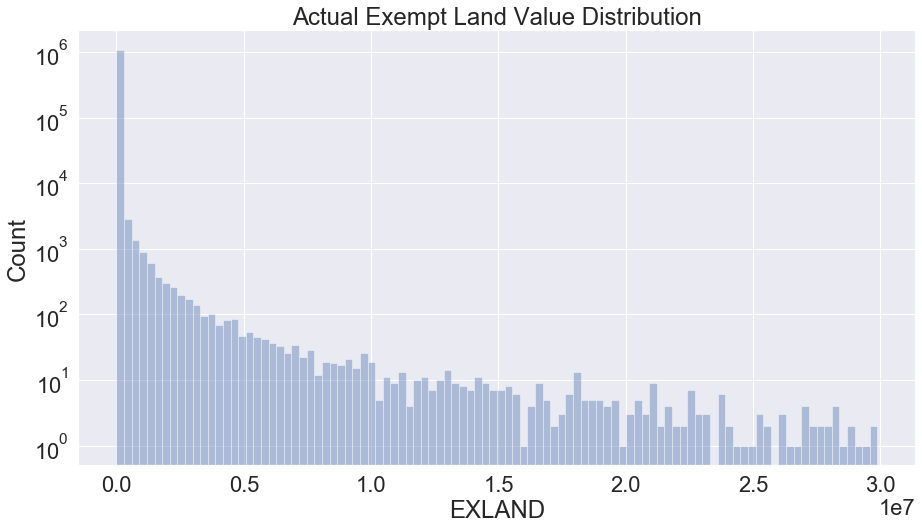
 

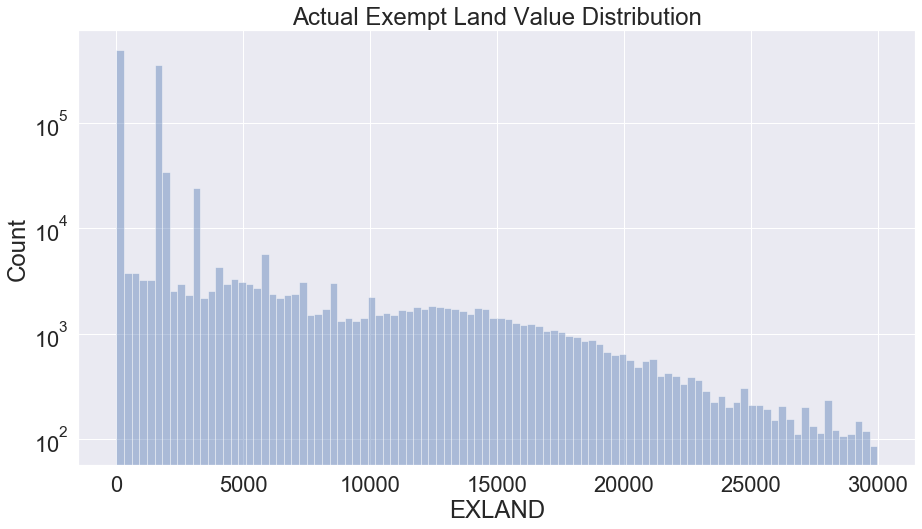
Variable 17

Name: EXLAND

Description: Tentative actual exempt land value

We plot the distribution of actual exempt land values under 10000000/30000 accounting for 99.979%/97.639% of all records.



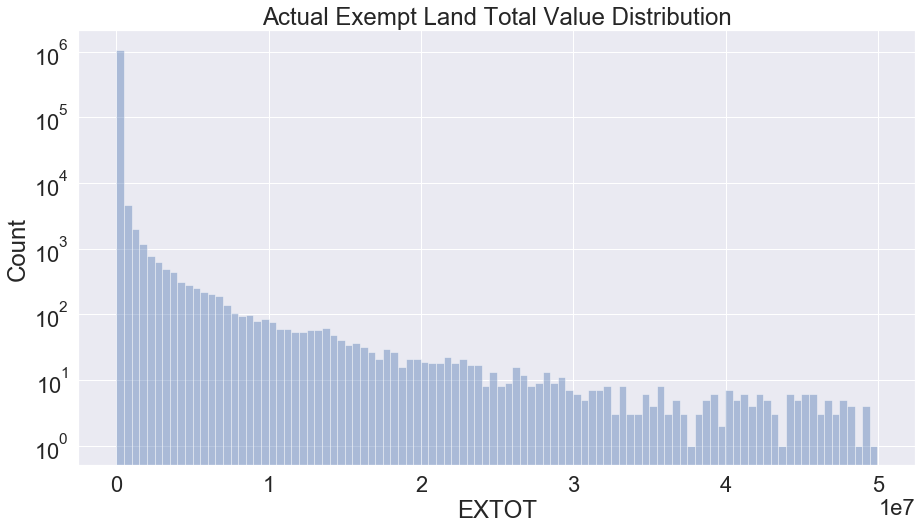


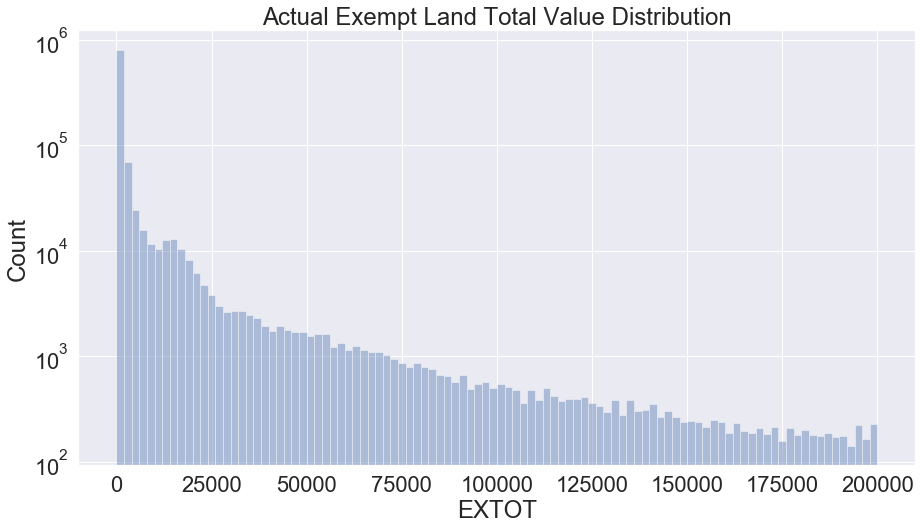
Variable 18

Name: EXTOT

Description: Tentative exempt land total value.

We plot the distribution of actual exempt land total values under 50000000/200000 accounting for 99.981%/97.779% of all records.

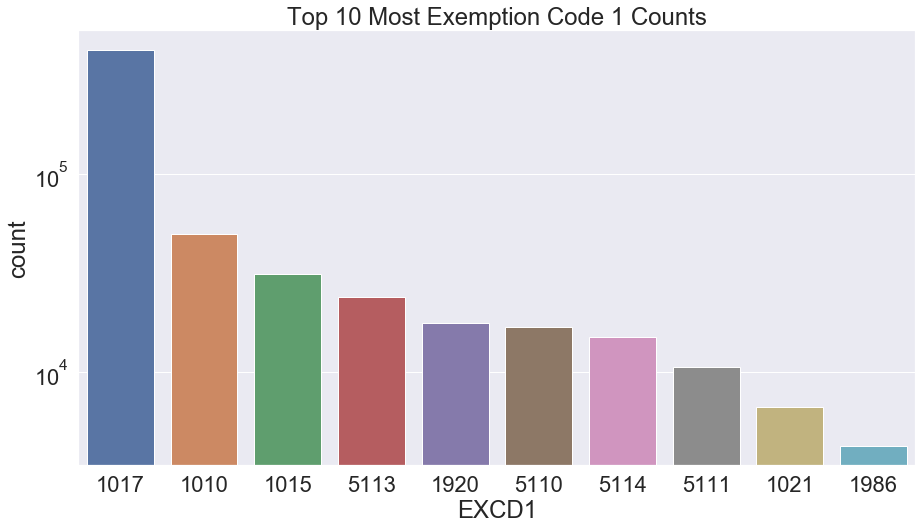




Variable 19

Name: EXCD1

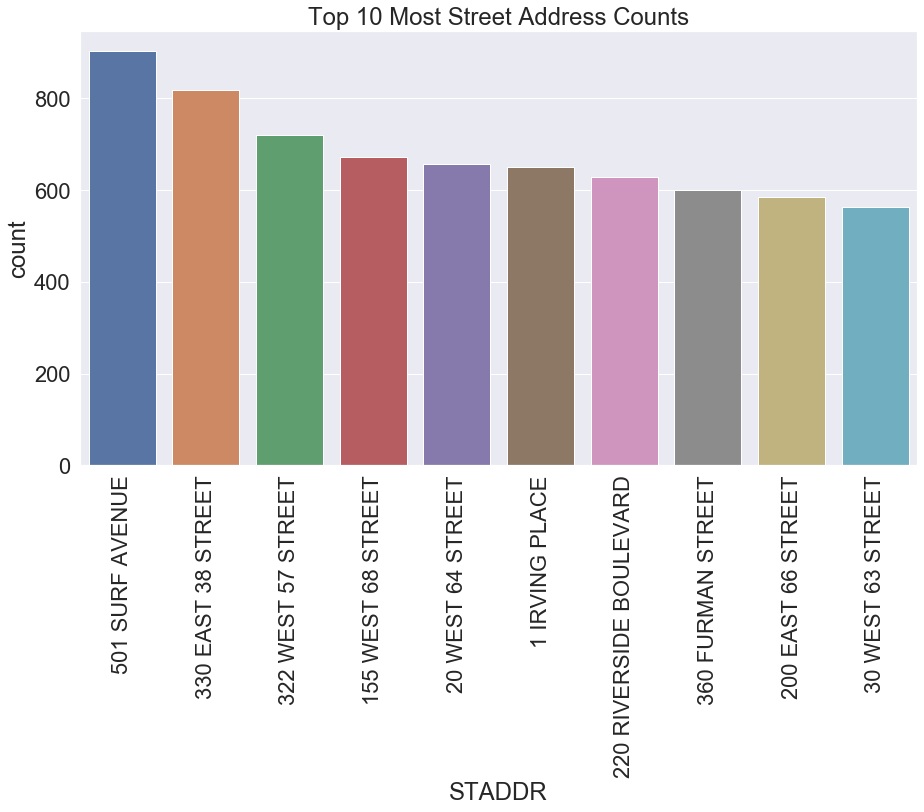
Description: Exemption code 1



Variable 20

Name: STADDR

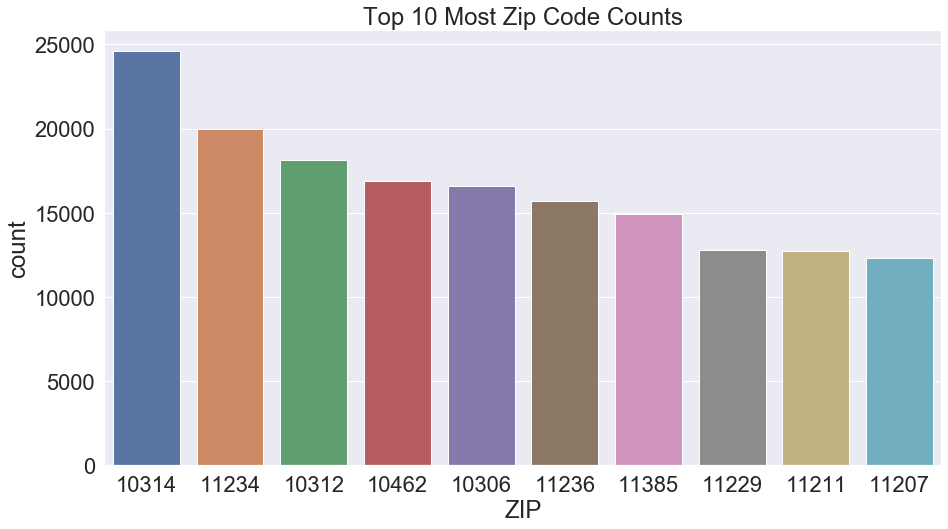
Description: Street address



Variable 21

Name: ZIP

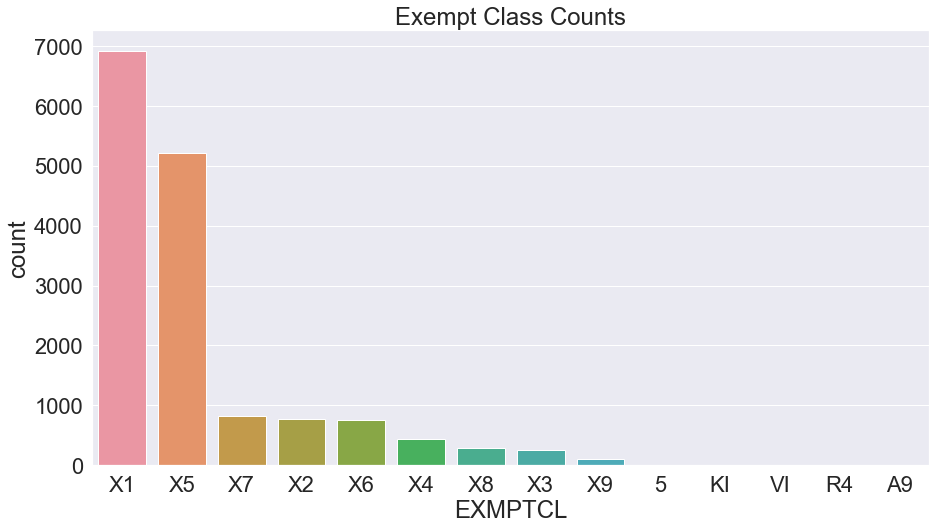
Description: Postal zip code



Variable 22

Name: EXMPTCL

Description: Exempt class for full exempt properties

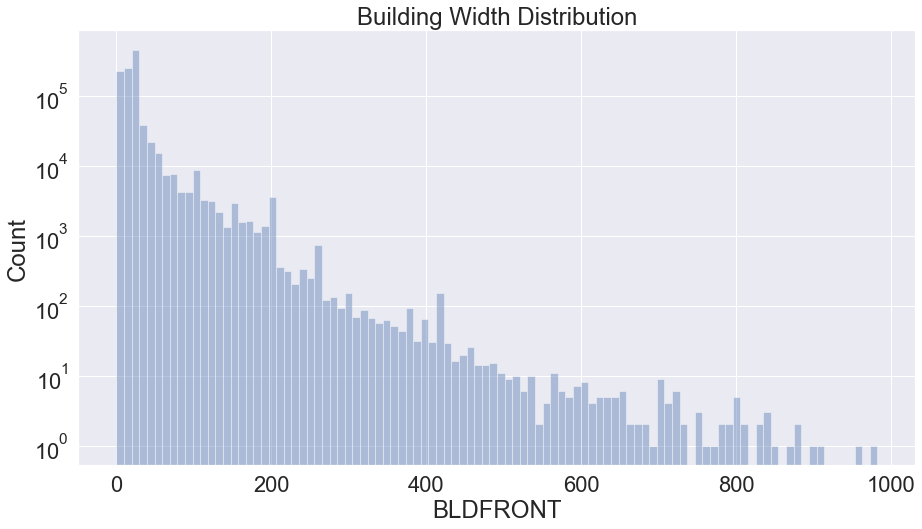


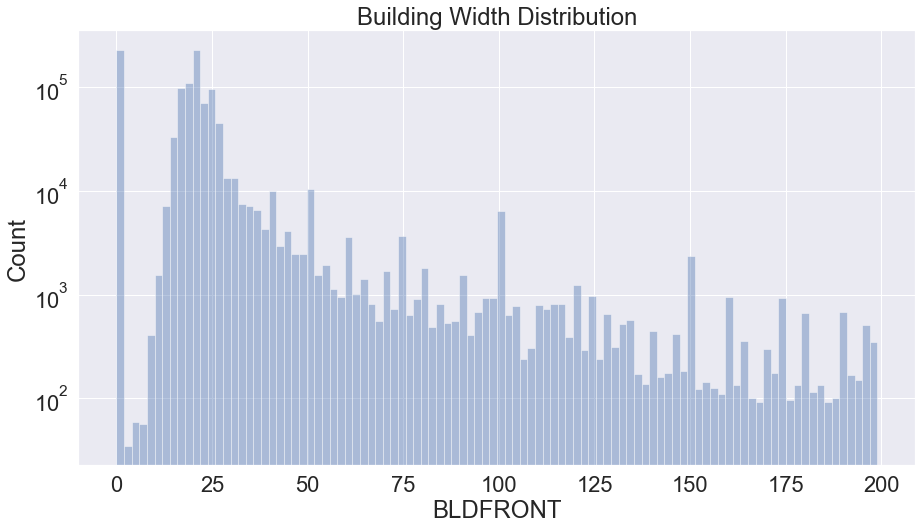
Variable 23

Name: BLDFRONT

Description: Building width in feet.

We plot the distribution of building width under 1000/200 accounting for 99.997%/99.368% of all records.

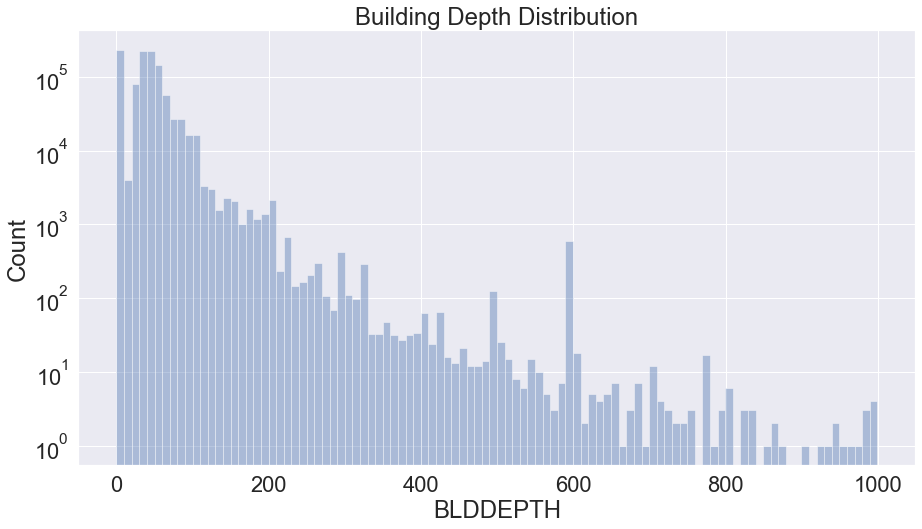


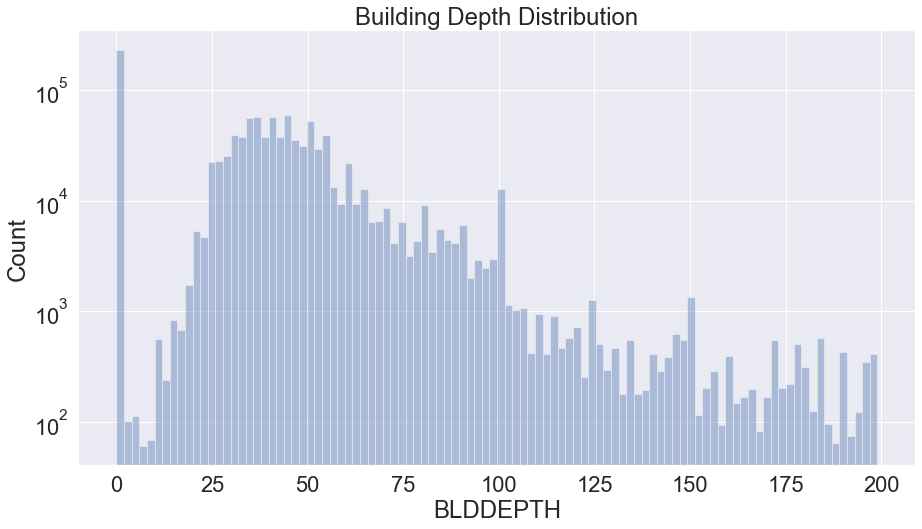


Variable 24

Name: BLDDEPTH

Description: Building depth in feet.

We plot the distribution of building depth under 1000/200 accounting for 99.997%/99.401% of all records. 

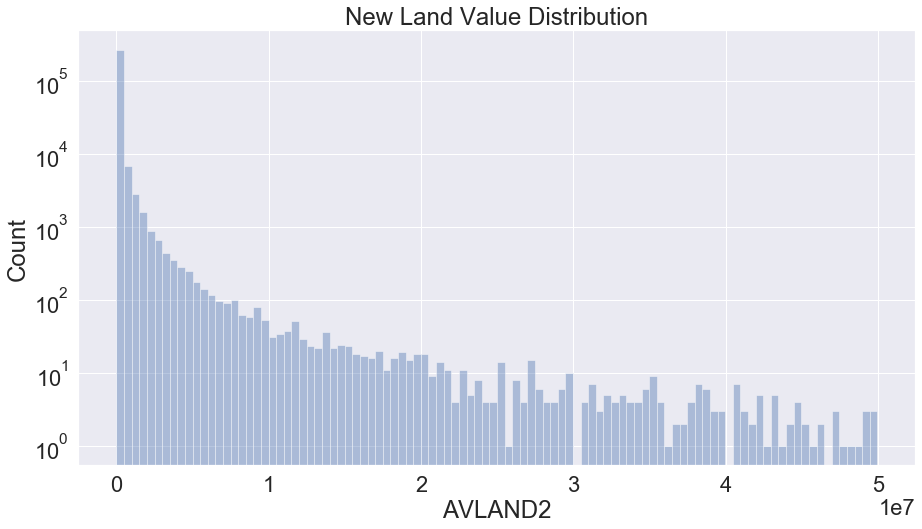
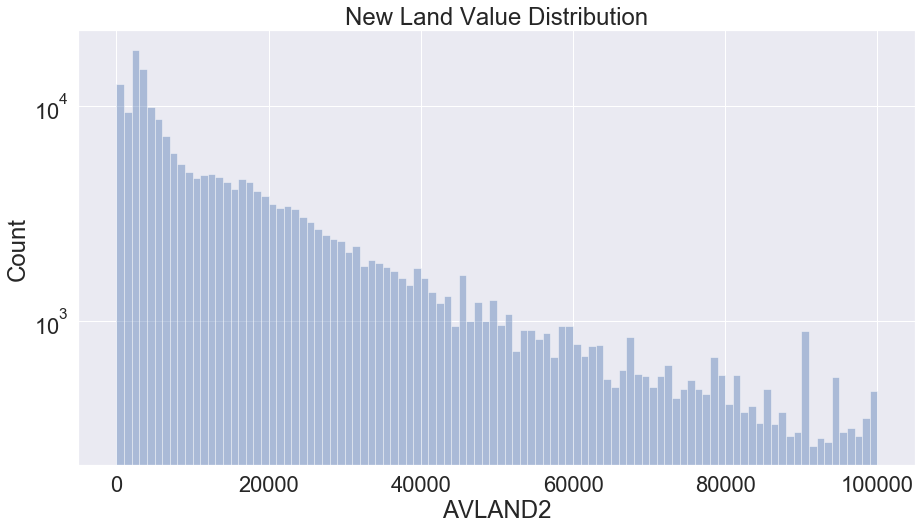


Variable 25

Name: AVLAND2

Description: New land market values.

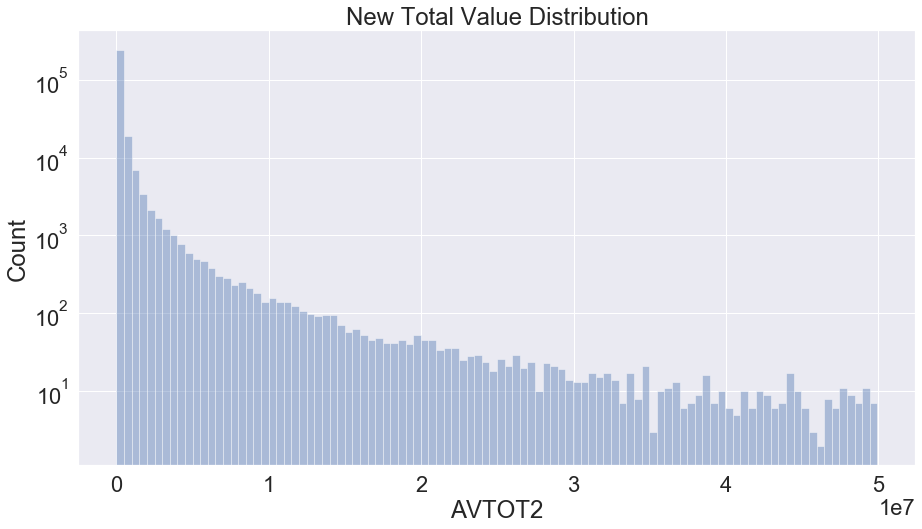
We plot the distribution of new land values under 50000000/100000 accounting for 99.966%/81.342% of all records.

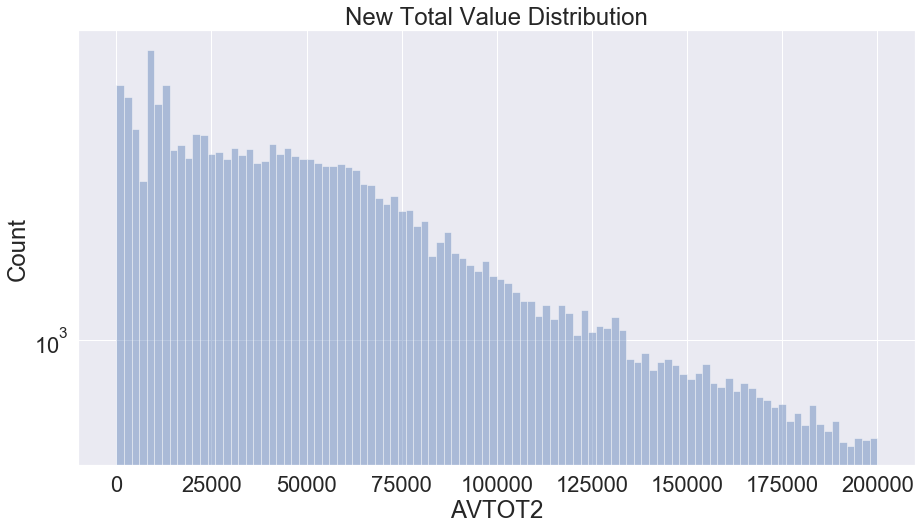
 

Variable 26

Name: AVTOT2

Description: New total market values.

We plot the distribution of new total values under 50000000/100000 accounting for 99.841%/71.851% of all records.

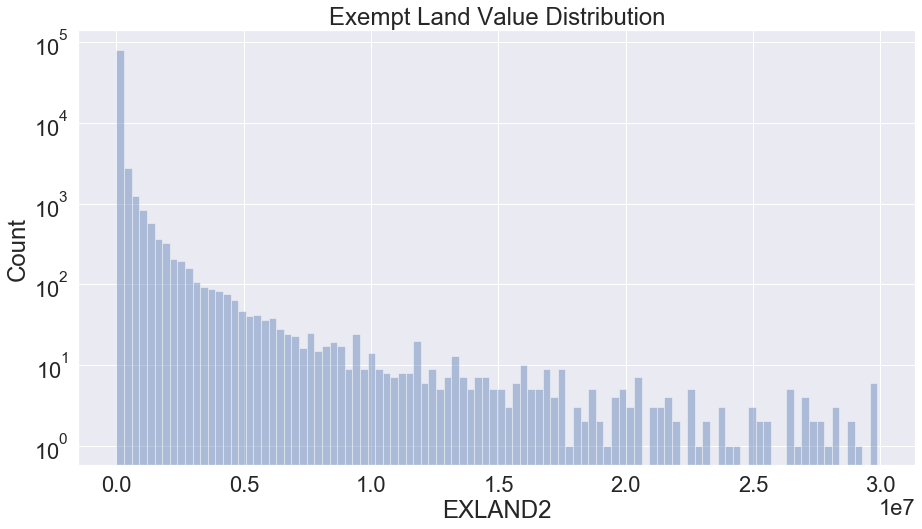
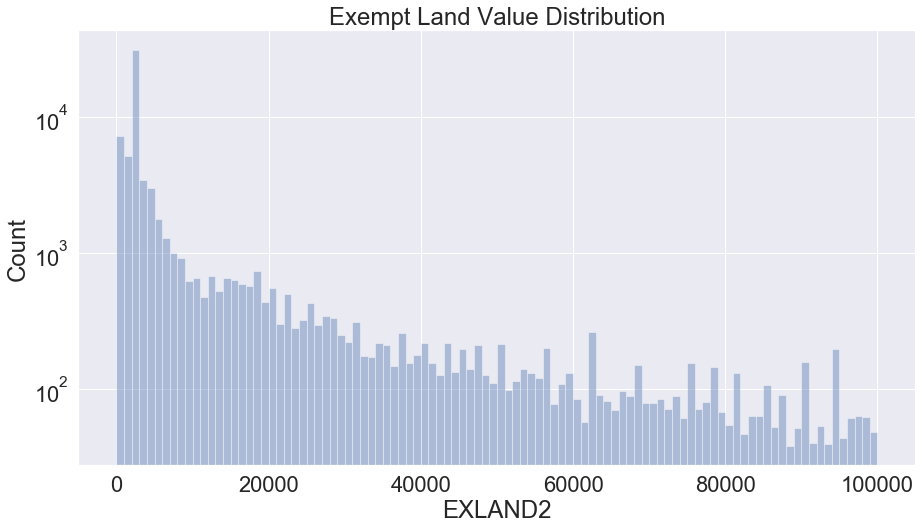


Variable 27

Name: EXLAND2

Description: New transitional exempt land values.

We plot the distribution of exempt land values under 30000000/100000 accounting for 99.873%/84.008% of all records.

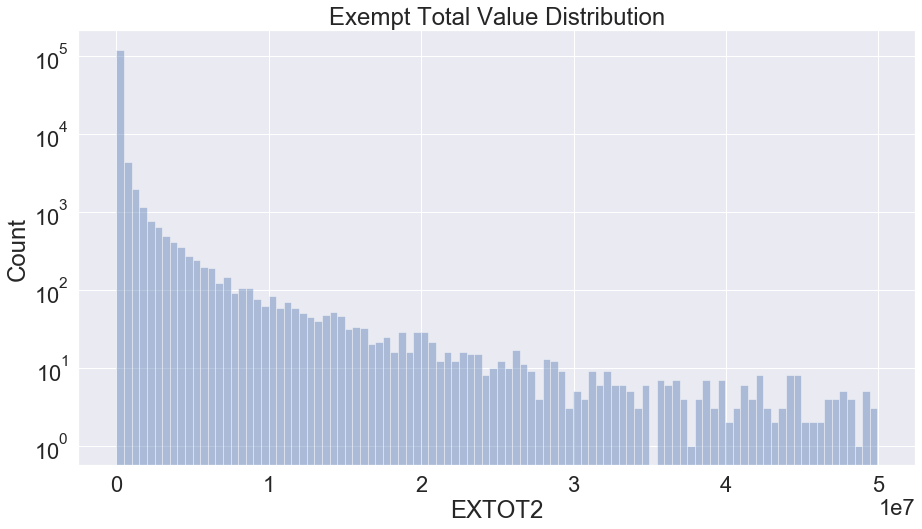
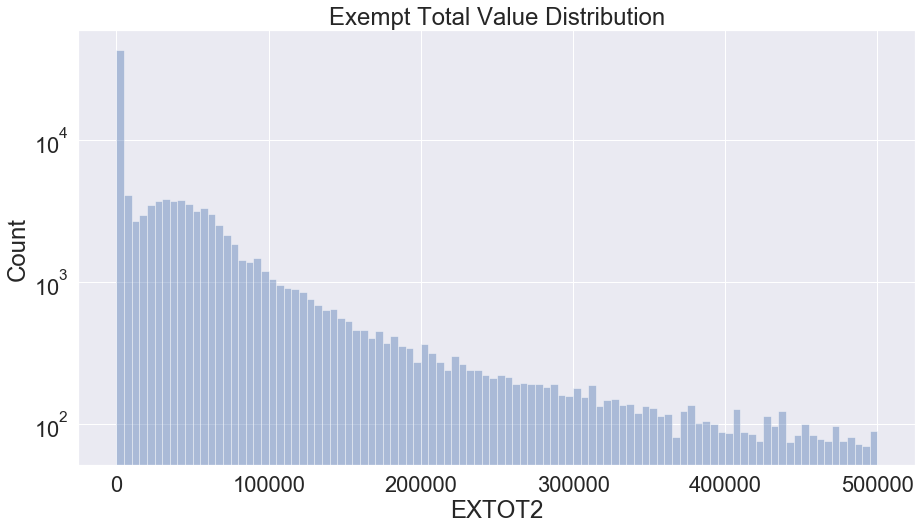
 

Variable 28

Name: EXTOT2

Description: New transitional exempt total values.

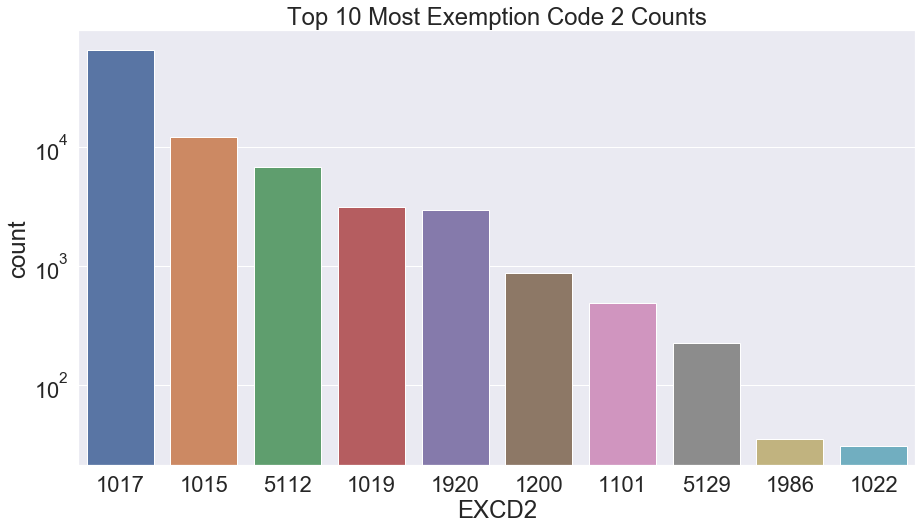
We plot the distribution of exempt total values under 50000000/500000 accounting for 99.866%/90.005% of all records.

Variable 29

Name: EXCD2

Description: Exemption code 2.



Variable 30

Name: PERIOD

Description: Assessment period when file was created.

Variable 31

Name: YEAR

Description: Assessment year.

Variable 32

Name: VALTYPE

Description: Value type (No meaning, not used).