**DATA QUALITY REPORT**

**PROJECT 1:**

**NY PROPERTY** **TAX FRAUD MODELING**

**Prepared for:**

Professor Coggeshall

USC Marshall DSO 562 Fraud Analytics

Homework 1

**Prepared by:**

Nanchun Shi

January 21, 2020

**TABLE OF CONTENTS**

[**PART I - INTRODUCTION**](#_jlgkpo7oov1a) **------------------------------------------------------------------------------** 1

[**PART II - SUMMARY OF DATASET**](#_f5z4j27jbf5k) **-----------------------------------------------------------------**2

[**PART III - FIELD DESCRIPTION**](#_8lvwnqf88br7) **---------------------------------------------------------------------** 4

# 

# PART I - INTRODUCTION

The dataset of interest “NY property data”, a subset with selected fields and data of the dataset “Property Valuation and Assessment Data”, is compiled and updated annually by the Department of Finance, City of New York for purpose to calculate Property Tax, Grant eligible properties Exemptions and/or Abatements. The dataset is obtained from NYC OpenData website.

The dataset contains 32 different feature information of 1,070,994 properties in New York in November, 2010.

## **PART II - SUMMARY OF DATASET**

**Below are the summary tables for the 32 fields in the dataset:**

1. **Numerical fields:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Field Name** | **# of Records** | **% Populated** | **# Unique Values** | **# Records w/ value 0** | **Mean** | **Standard Deviation** | **Max** | **Min** |
| **RECORD** | 1070994 | 100% | 1070994 | 0 | 535497.5 | 309169.5 | 1070994 | 1 |
| **LTFRONT** | 1070994 | 100% | 1297 | 169108 | 36.6353 | 74.03284 | 9999 | 0 |
| **LTDEPTH** | 1070994 | 100% | 1370 | 170128 | 88.86159 | 76.39628 | 9999 | 0 |
| **STORIES** | 1014730 | 94.75% | 111 | 0 | 5.006918 | 8.365707 | 119 | 1 |
| **FULLVAL** | 1070994 | 100% | 109324 | 13007 | 874264.5 | 11582430 | 6150000000 | 0 |
| **AVLAND** | 1070994 | 100% | 70921 | 13009 | 85067.92 | 4057260 | 2668500000 | 0 |
| **AVTOT** | 1070994 | 100% | 112914 | 13007 | 227238.169 | 6877529 | 4668309000 | 0 |
| **EXLAND** | 1070994 | 100% | 33419 | 491699 | 36423.89 | 3981576 | 2668500000 | 0 |
| **EXTOT** | 1070994 | 100% | 64255 | 432572 | 91186.98 | 6508403 | 4668309000 | 0 |
| **BLDFRONT** | 1070994 | 100% | 612 | 228815 | 23.042 | 35.5797 | 7575 | 0 |
| **Field Name** | **# of Records** | **% Populated** | **# Unique Values** | **# Records w/ value 0** | **Mean** | **Standard Deviation** | **Max** | **Min** |
| **BLDDEPTH** | 1070994 | 100% | 621 | 228853 | 39.9228 | 42.707 | 9393 | 0 |
| **AVLAND2** | 282726 | 26.398% | 58591 | 0 | 246235.719 | 6178963 | 2371005000 | 3 |
| **AVTOT2** | 282732 | 26.399% | 111360 | 0 | 713911.436 | 11652530 | 4501180000 | 3 |
| **EXLAND2** | 87449 | 8.165% | 22195 | 0 | 351235.684 | 10902210 | 2371005000 | 1 |
| **EXTOT2** | 130828 | 12.216% | 48348 | 0 | 656768.282 | 16072510 | 4501180000 | 7 |

1. **Categorical fields:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field Name** | **# of Records** | **% Populated** | **# Unique Values** | **Most Common Value** |
| **BBLE** | 1070994 | 100% | 1070994 | - |
| **B** | 1070994 | 100% | 5 | 4 |
| **BLOCK** | 1070994 | 100% | 13984 | 3944 |
| **LOT** | 1070994 | 100% | 6366 | 1 |
| **EASEMENT** | 4636 | 0.4329% | 12 | E |
| **OWNER** | 1039249 | 97.0359% | 863346 | PARKCHESTER PRESERVAT |
| **BLDGCL** | 1070994 | 100% | 200 | R4 |
| **TAXCLASS** | 1070994 | 100% | 11 | 1 |
| **EXT** | 354305 | 33.082% | 3 | G |
| **EXCD1** | 638488 | 59.616% | 129 | 1017 |
| **STADDR** | 1070318 | 99.937% | 839280 | 501 SURF AVENUE |
| **ZIP** | 1041104 | 97.209% | 196 | 10314 |
| **EXMPTCL** | 15579 | 1.455% | 14 | X1 |
| **EXCD2** | 92948 | 8.679% | 60 | 1017 |
| **PERIOD** | 1070994 | 100% | 1 | FINAL |
| **YEAR** | 1070994 | 100% | 1 | 2010/11 |
| **VALTYPE** | 1070994 | 100% | 1 | AC-TR |

## **PART III - FIELD DESCRIPTION**

**1.**

**Field name:** RECORD

**Field type:** int64

**Description:** unique index number for each record, ascending from 1 to 1070994.

**2.**

**Field name:** BBLE

**Field type:** object

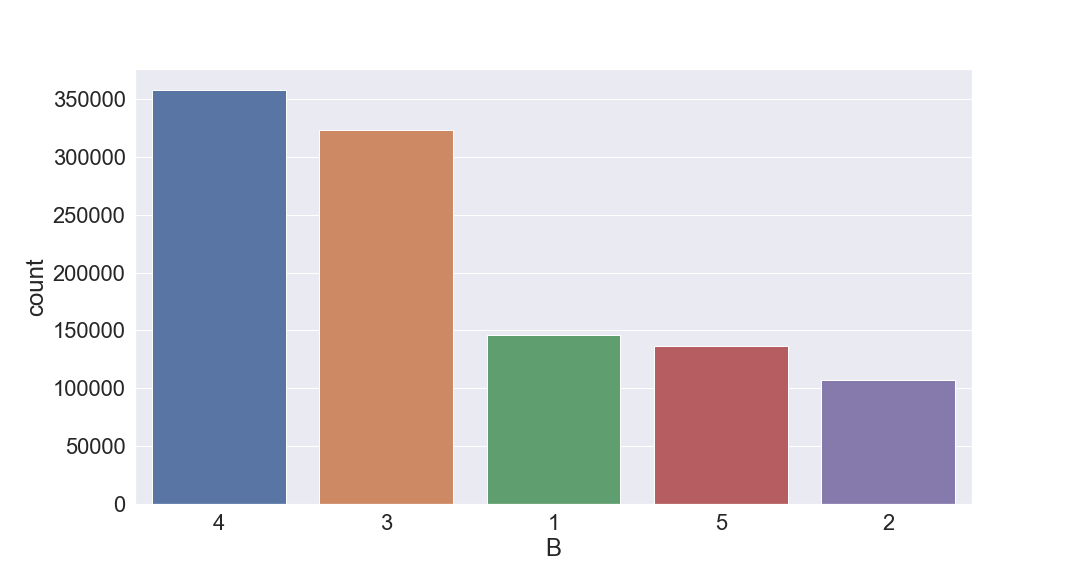
**Description:** Concatenation of BORO code, BLOCK code, LOT code, and EASEMENT code. Unique for each record.

**3.**

**Field name:** B

**Field type:** int64

**Description:** BOROUGH code; 1 = MANHATTAN, 2 = BRONX, 3 = BROOKLYN, 4 = QUEENS, 5 = STATEN ISLAND.



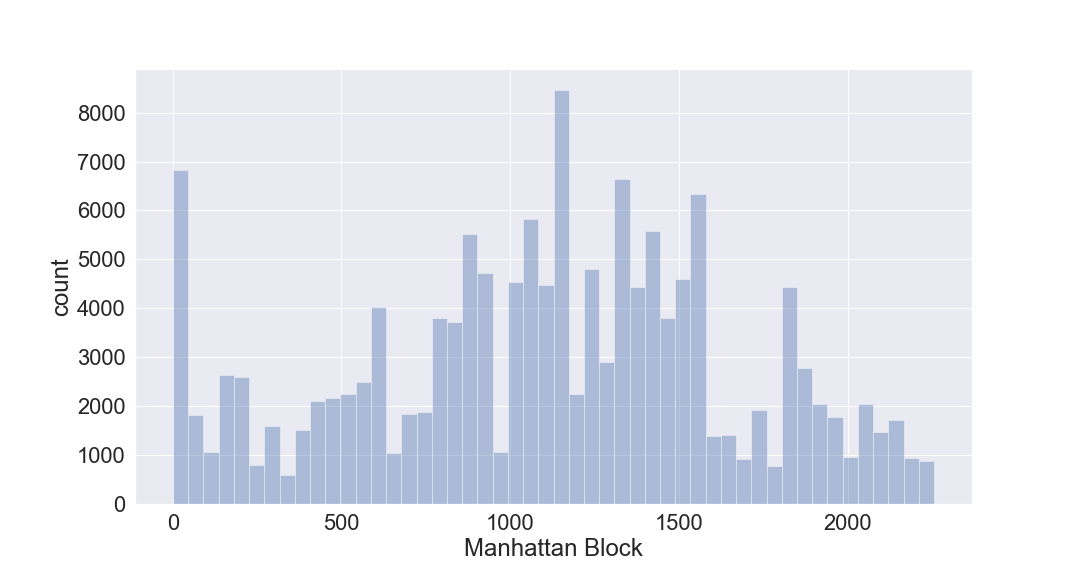
**4.**

**Field name:** BLOCK

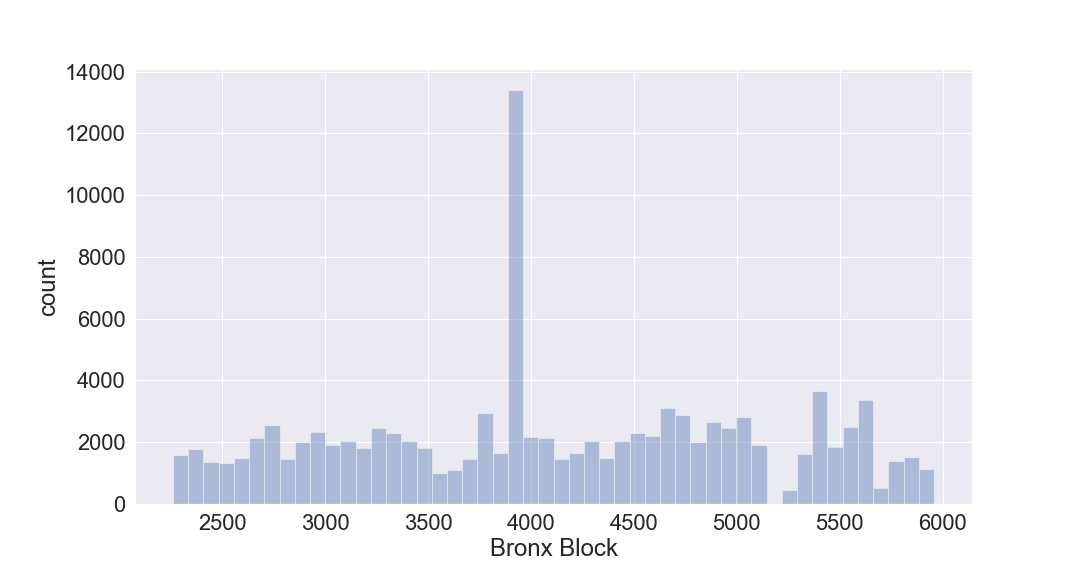
**Field type:** int64

**Description:** valid block ranges by BORO; 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.

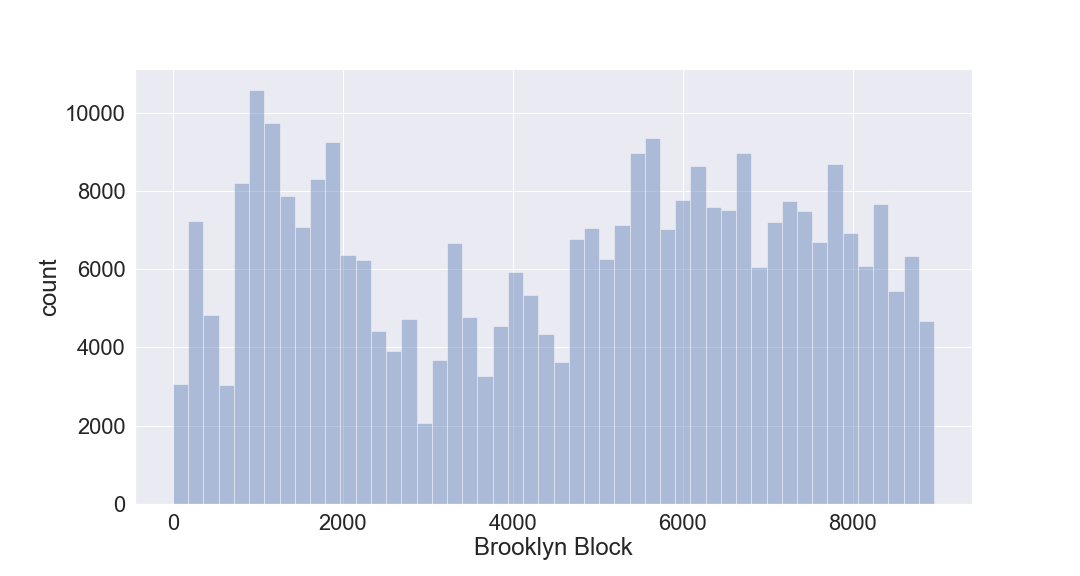
Manhattan:

****

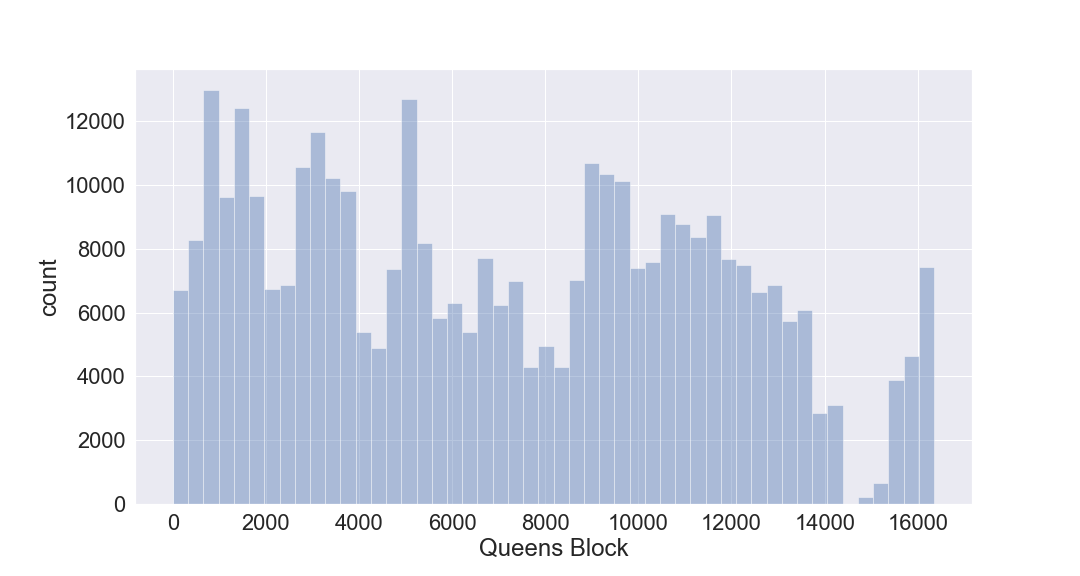
Bronx:

****

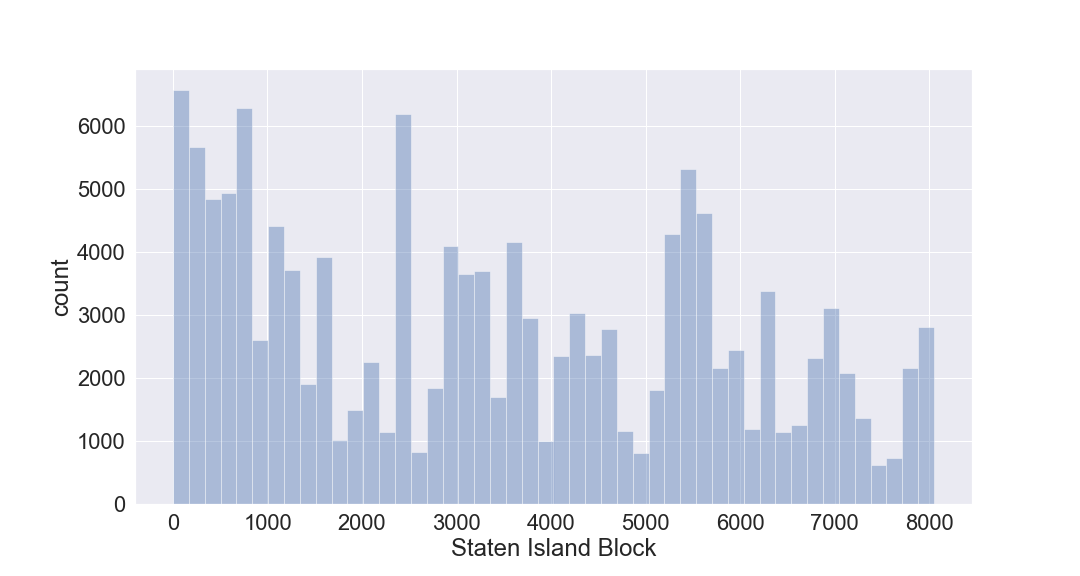
Brooklyn:

****

Queens:

****

Staten Island:

****

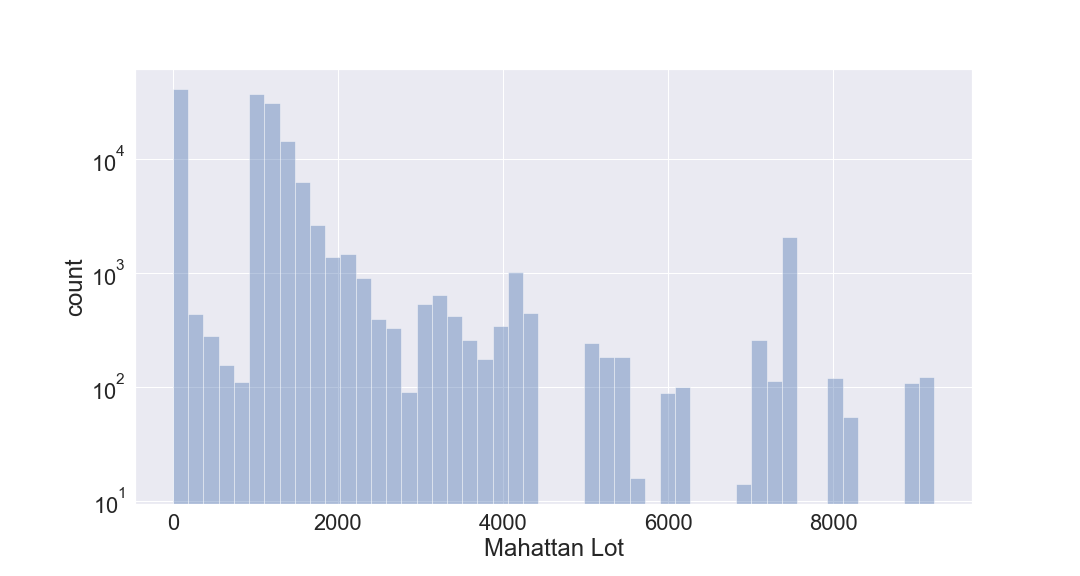
**5.**

**Field name:** LOT

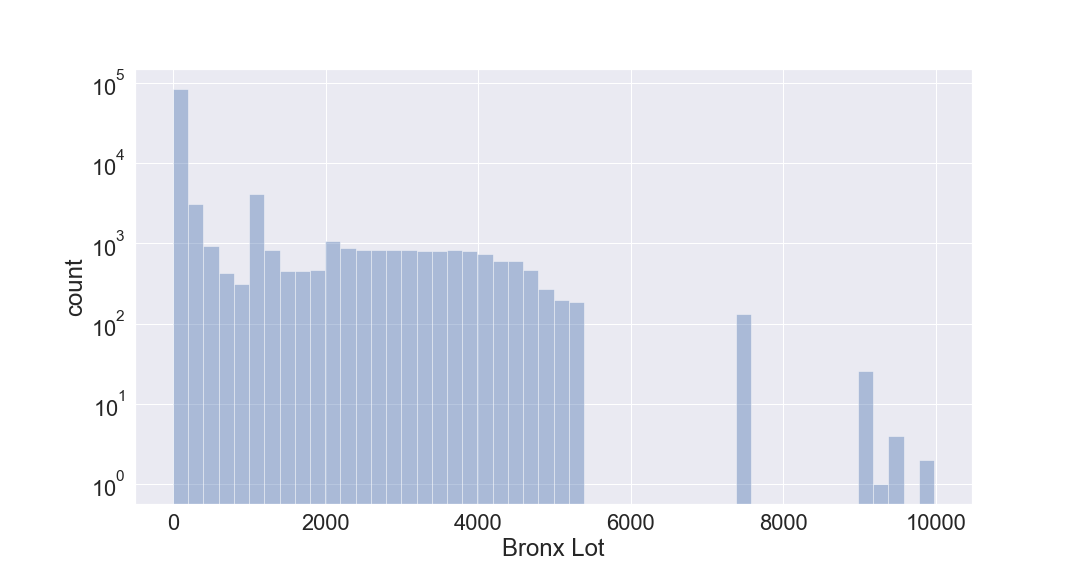
**Field type:** int64

**Description:** unique number within BORO/BLOCK

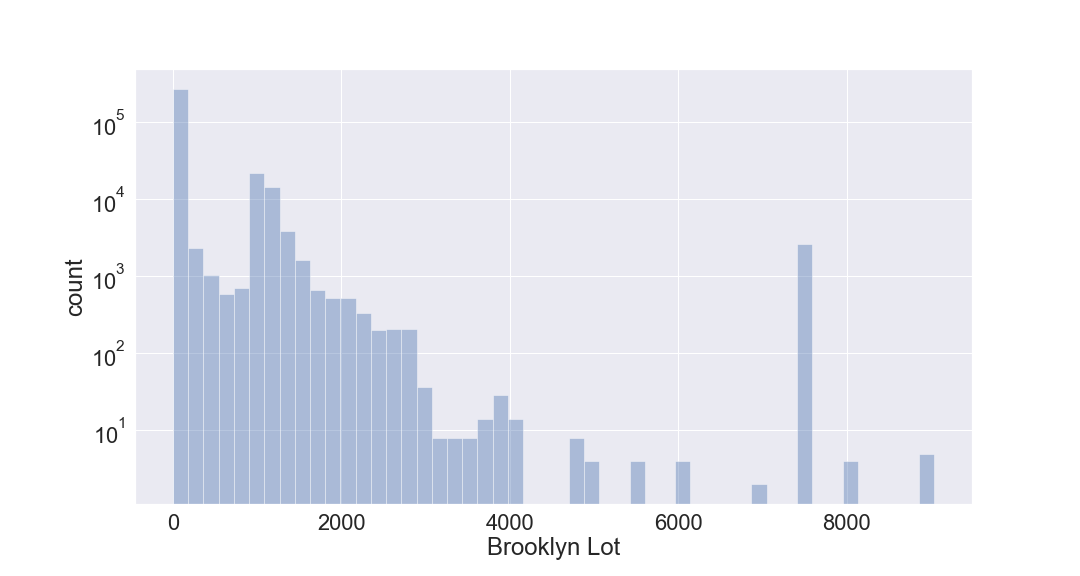
Manhattan:



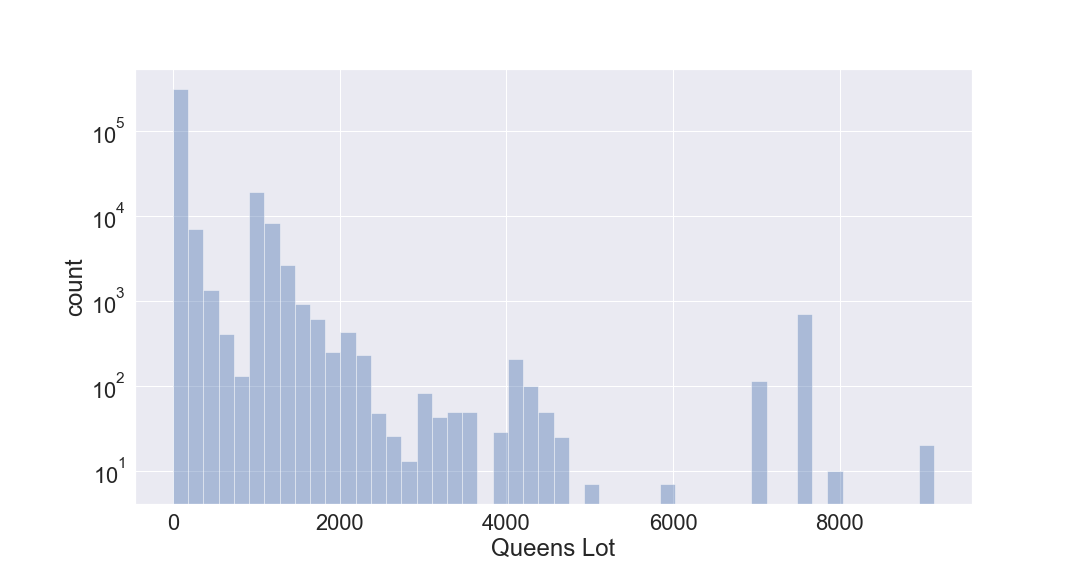
Bronx:



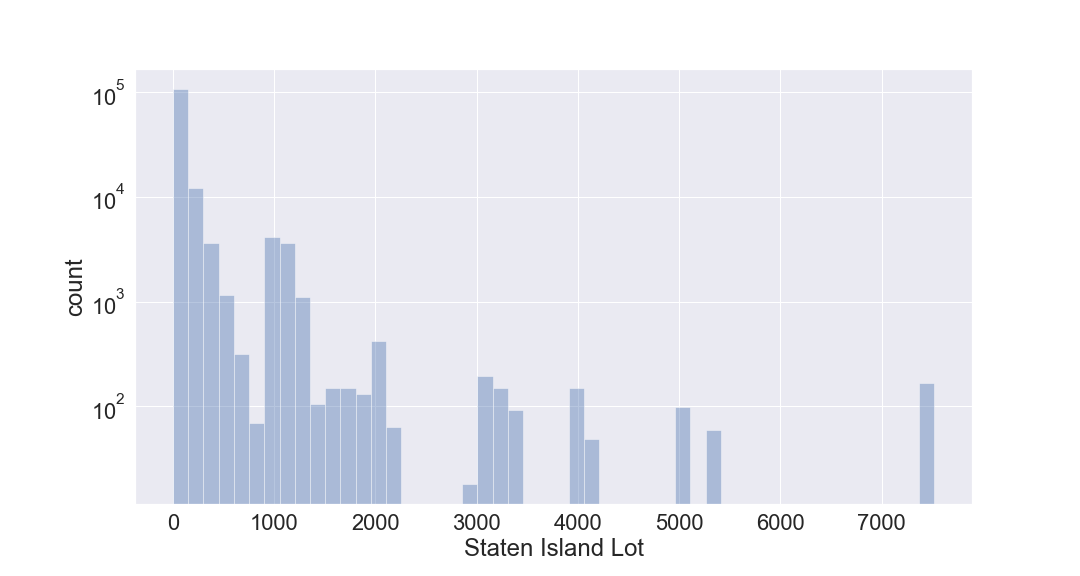
Brooklyn:



Queens:



Staten Island:



**6.**

**Field name:** EASEMENT

**Field type:** object

**Description:** 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'.

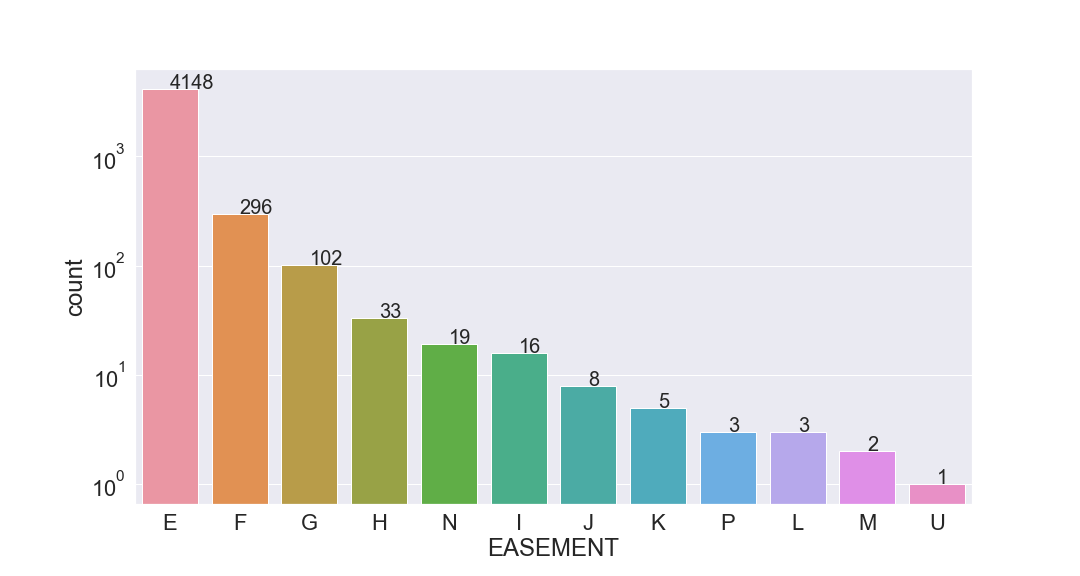
'N' Indicates Non-Transit Easement

'P' Indicates Piers.

'R' Indicates Railroads.

'S' Indicates Street

'U' Indicates U.S. Government

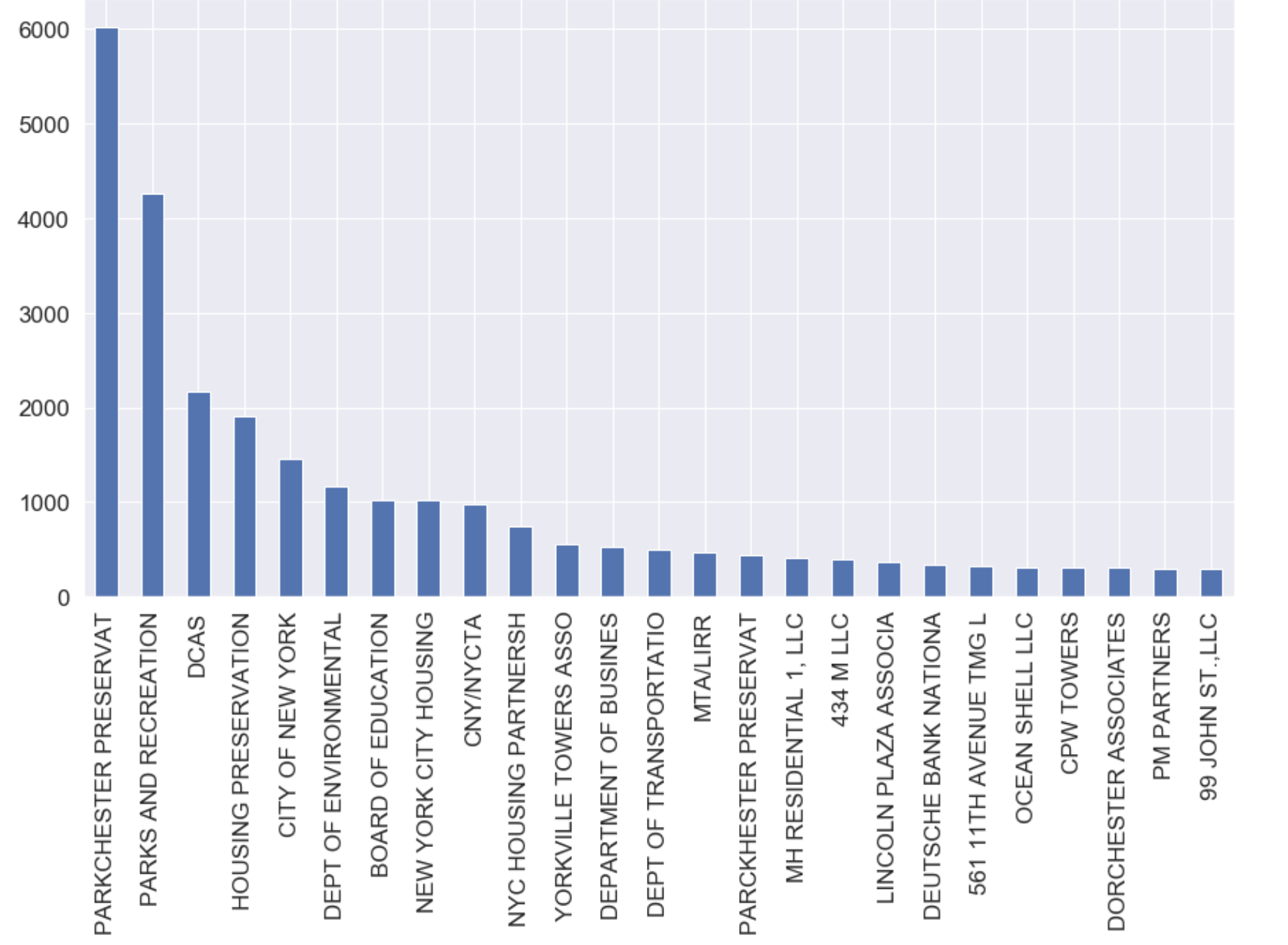
****

**7.**

**Field name:** OWNER

**Field type:** object

**Description:** owner names of the properties.

Top 25 owners with most properties:

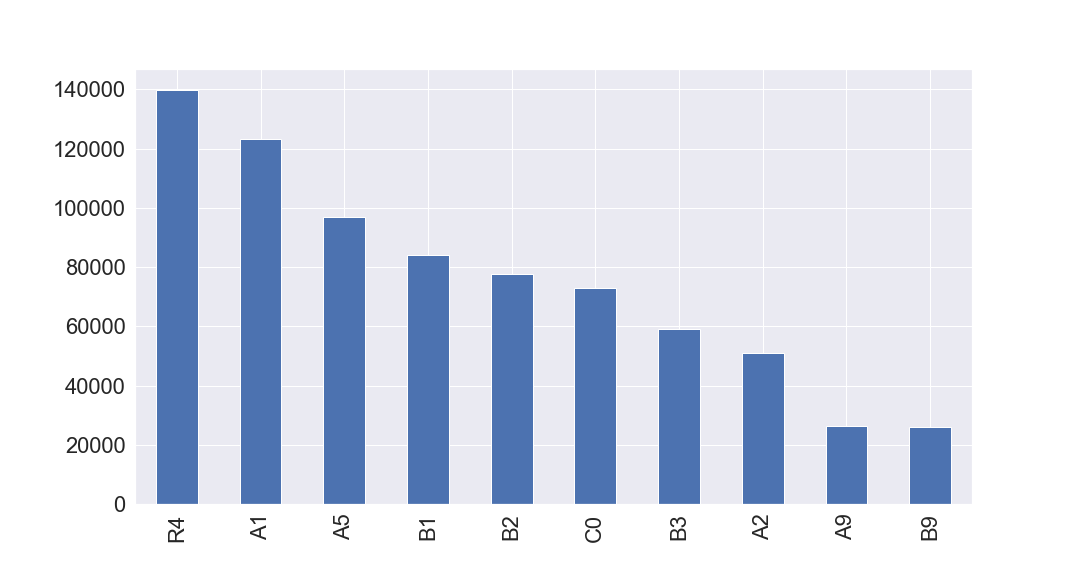
**8.**

**Field name:** BLDGCL

**Field type:** object

**Description:** building class. There is a direct correlation between the Building Class and the Tax Class.

Top 10 frequent building classes:



**9.**

**Field name:** TAXCLASS

**Field type:** object

**Description:** current Property Tax Class Code (NYS Classification);

tax class 1 = 1-3 unit residences

tax class 1a = 1-3 story condominiums originally a condo

tax class 1b = residential vacant land

tax class 1c = 1-3 unit condominiums originally tax class 1

tax class 1d = select bungalow colonies

tax class 2 = apartments

tax class 2a = apartments with 4-6 units

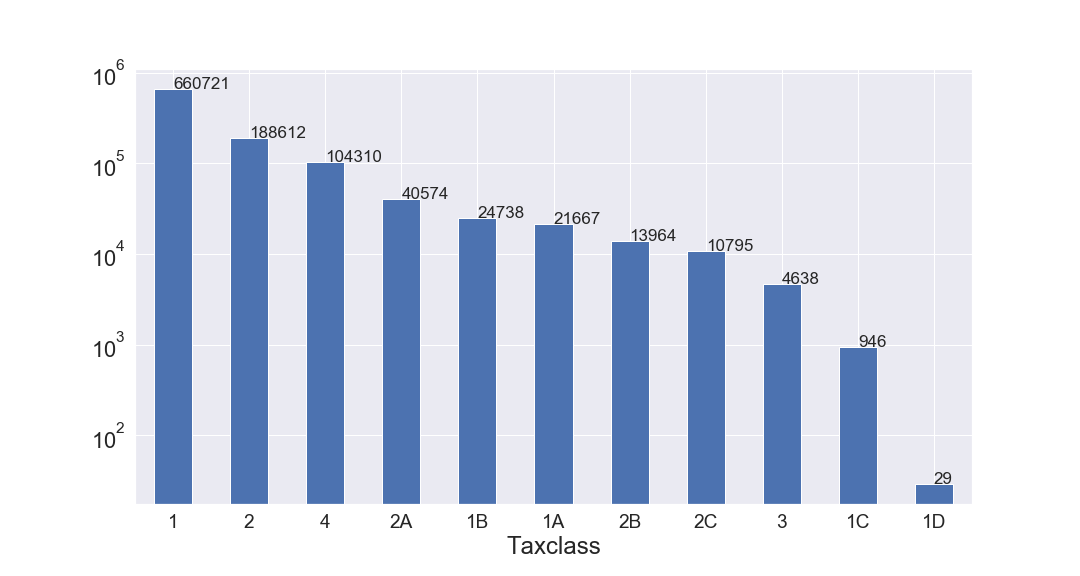
tax class 2b = apartments with 7-10 units

tax class 2c = coops/condos with 2-10 units

tax class 3 = utilities (except ceiling rr)

tax class 4a = utilities - ceiling railroads

tax class 4 = all others



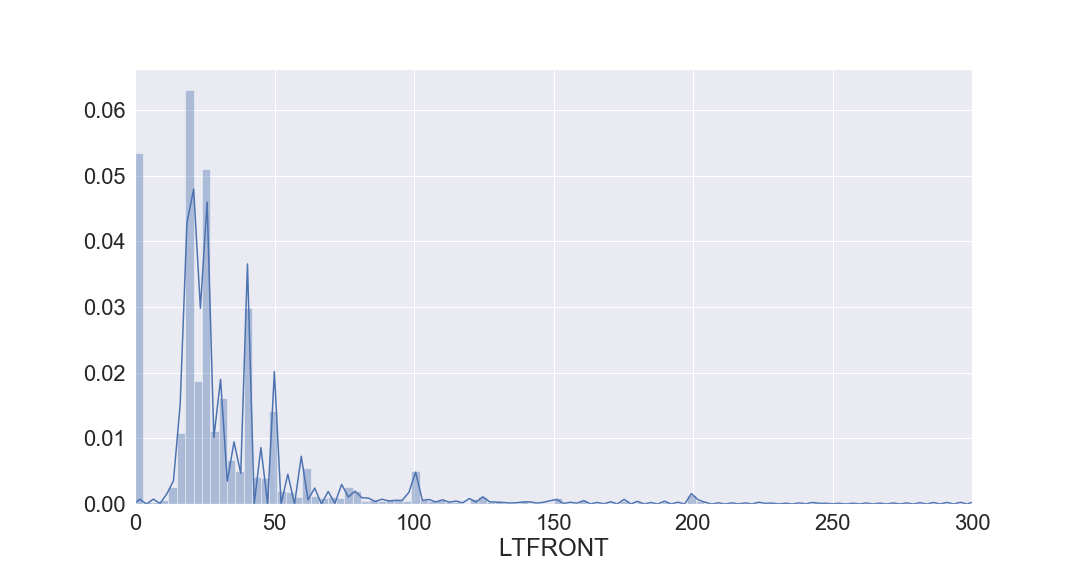
**10.**

**Field name:** LTFRONT

**Field type:** int64

**Description:** lot frontage in feet.

Properties that have lot frontage < 300:



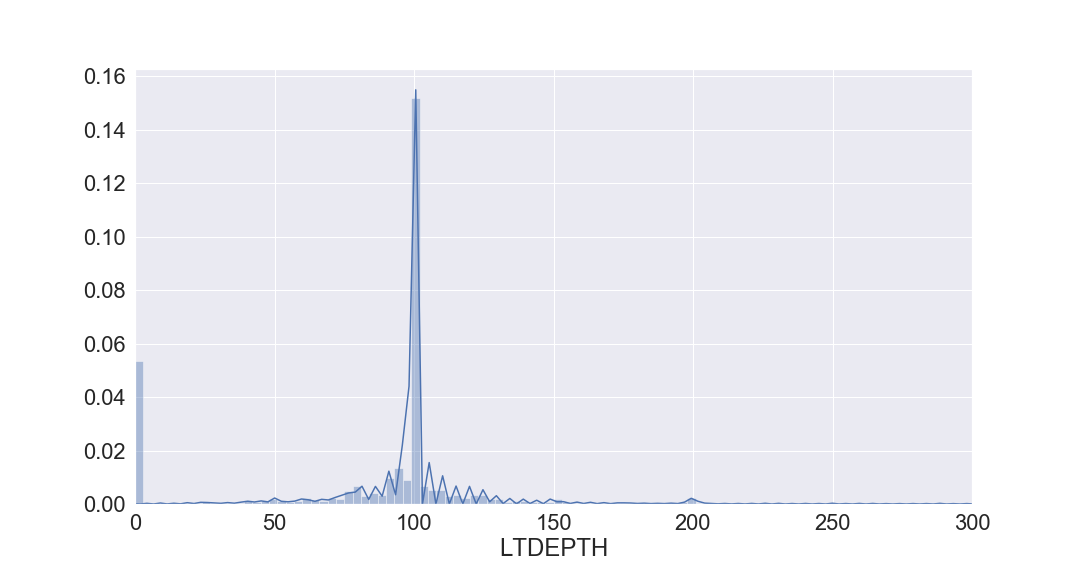
**11.**

**Field name:** LTDEPTH

**Field type:** int64

**Description:** lot depth in feet.

Properties that have lot depth < 300:

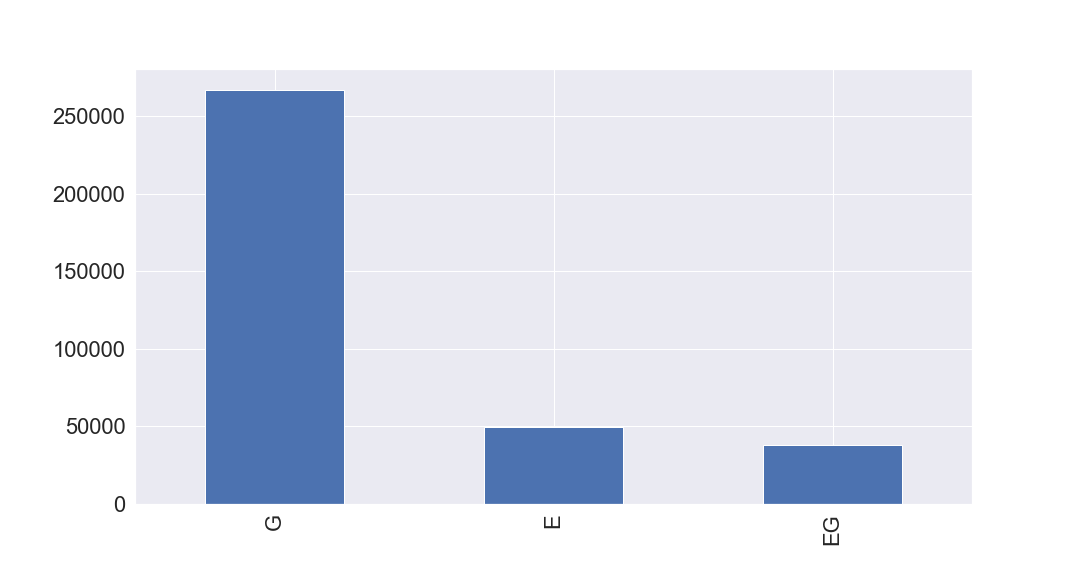


**12.**

**Field name:** EXT

**Field type:** object

**Description:** extension indicator; 'E' = extension, 'G' = garage, 'EG' = extension and garage



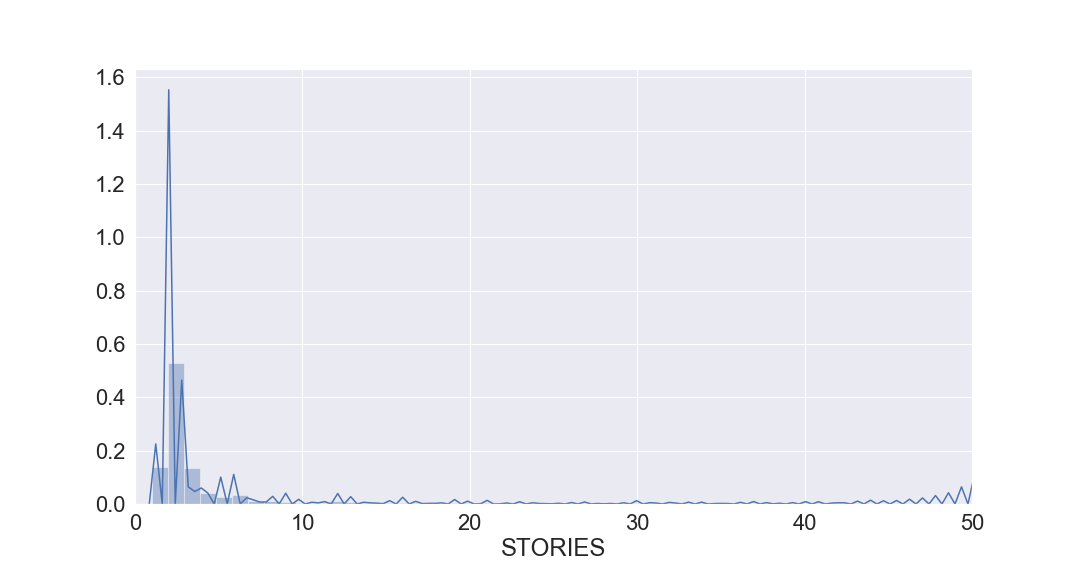
**13.**

**Field name:** STORIES

**Field type:** float64

**Description:** The number of stories for the building ( # of Floors).

Properties that have stories < 50:



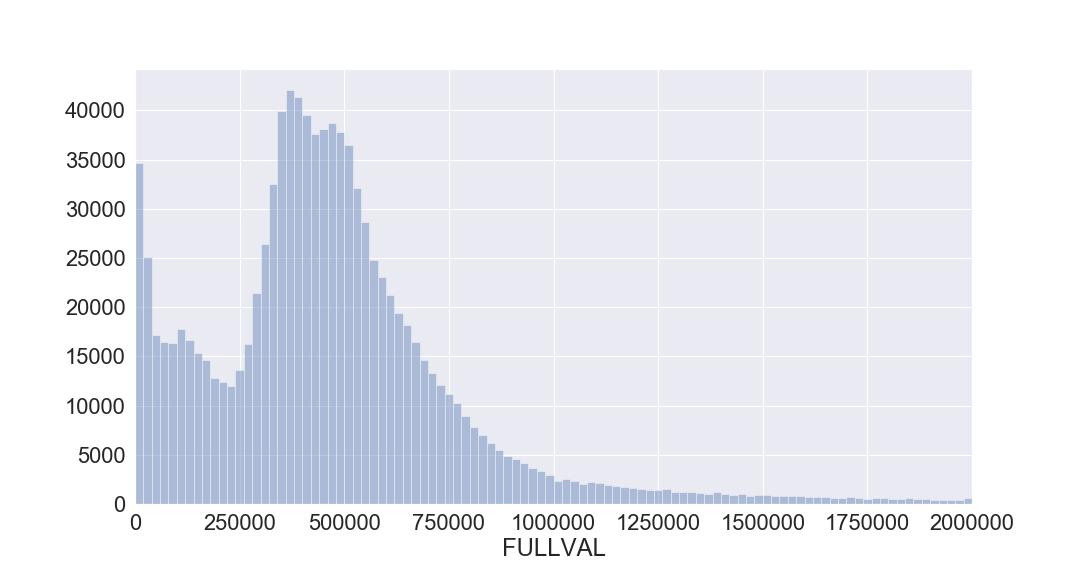
**14.**

**Field name:** FULLVAL

**Field type:** float64

**Description:** total market value.

Properties that have total market value < 2000000:



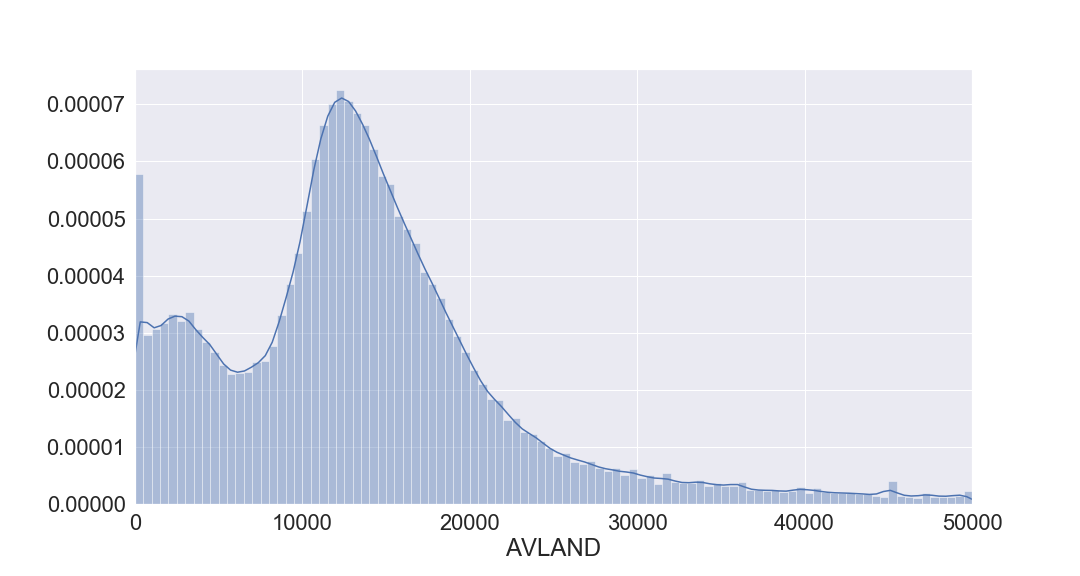
**15.**

**Field name:** AVLAND

**Field type:** float64

**Description:** actual land value.

Properties that have transitional assessed land value < 50000:



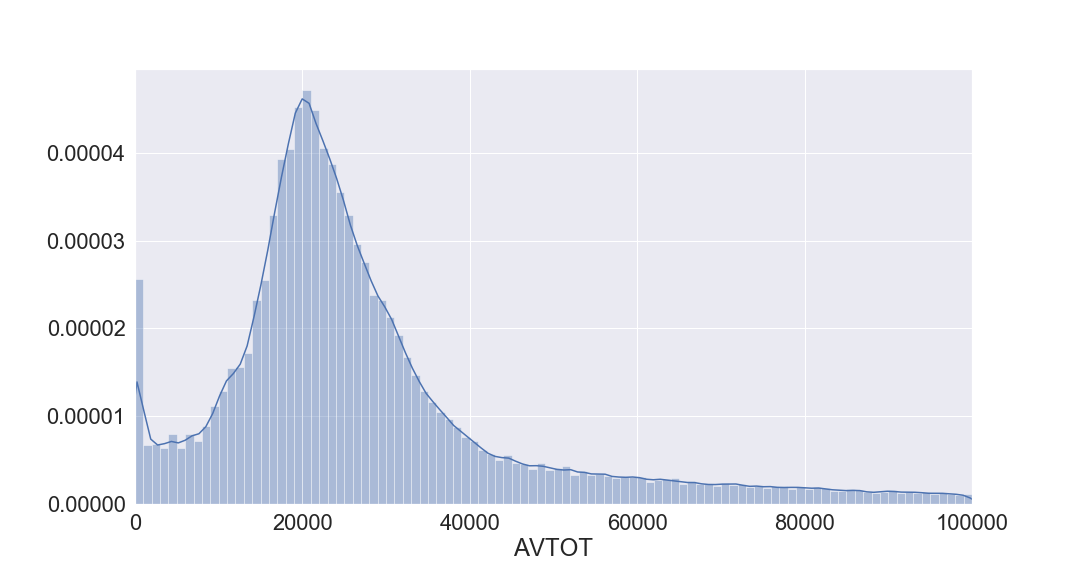
**16.**

**Field name:** AVTOT

**Field type:** float64

**Description:** actual total value.

Properties that have transitional assessed total value < 100000



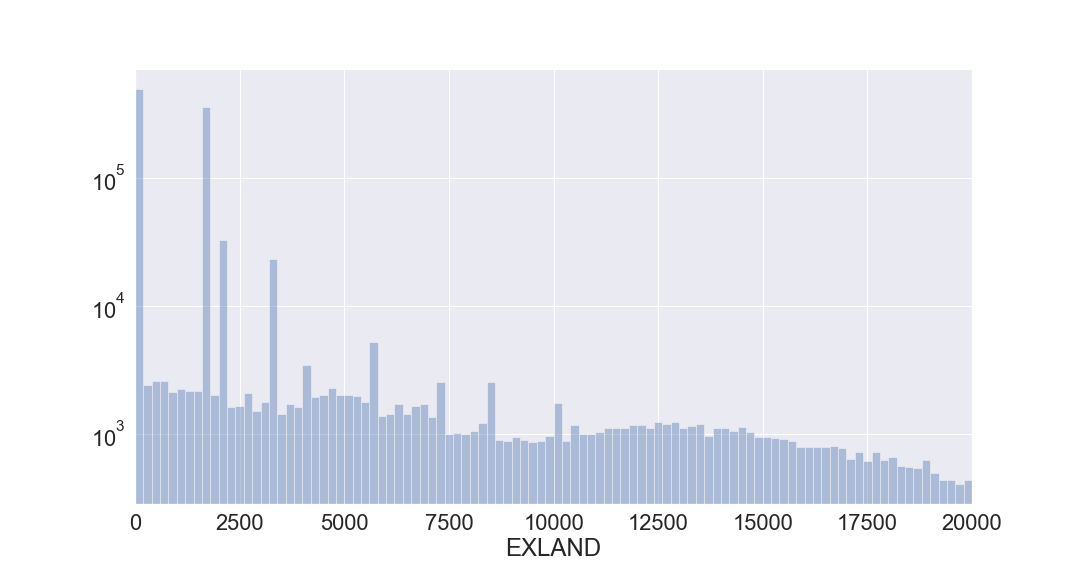
**17.**

**Field name:** EXLAND

**Field type:** float64

**Description:** actual exempt land value.

Properties that have transitional exempt land value < 20000:



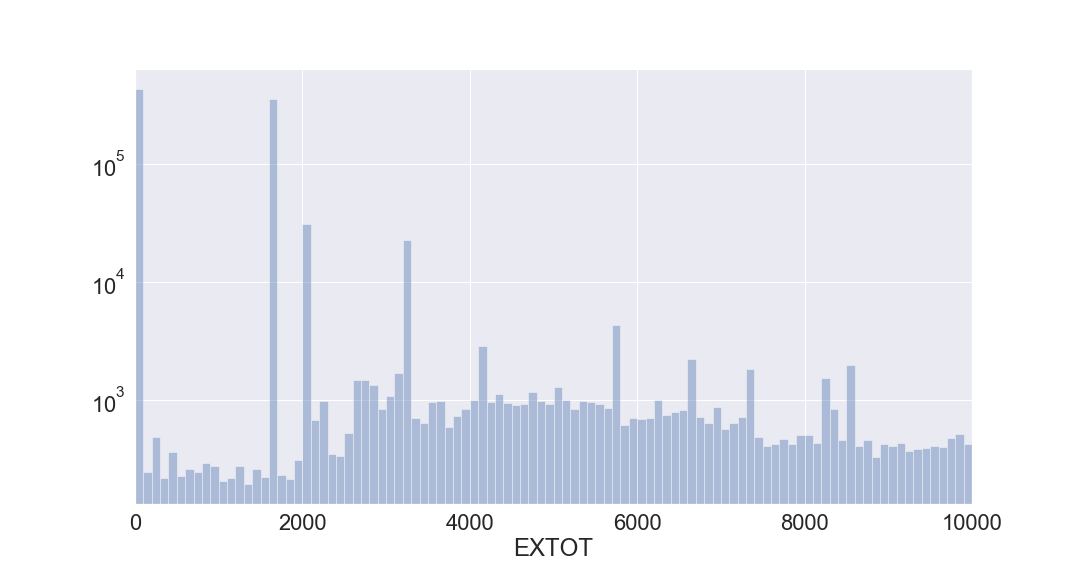
**18.**

**Field name:** EXTOT

**Field type:** float64

**Description:** actual exempt land total.

Properties that have transitional exempt total value < 20000:

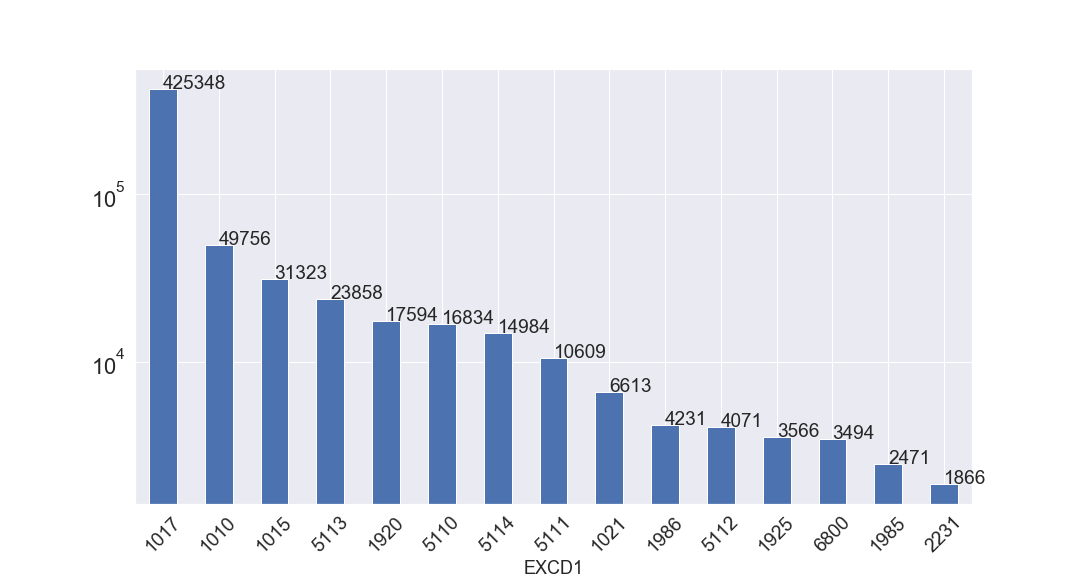


**19.**

**Field name:** EXCD1

**Field type:** float64

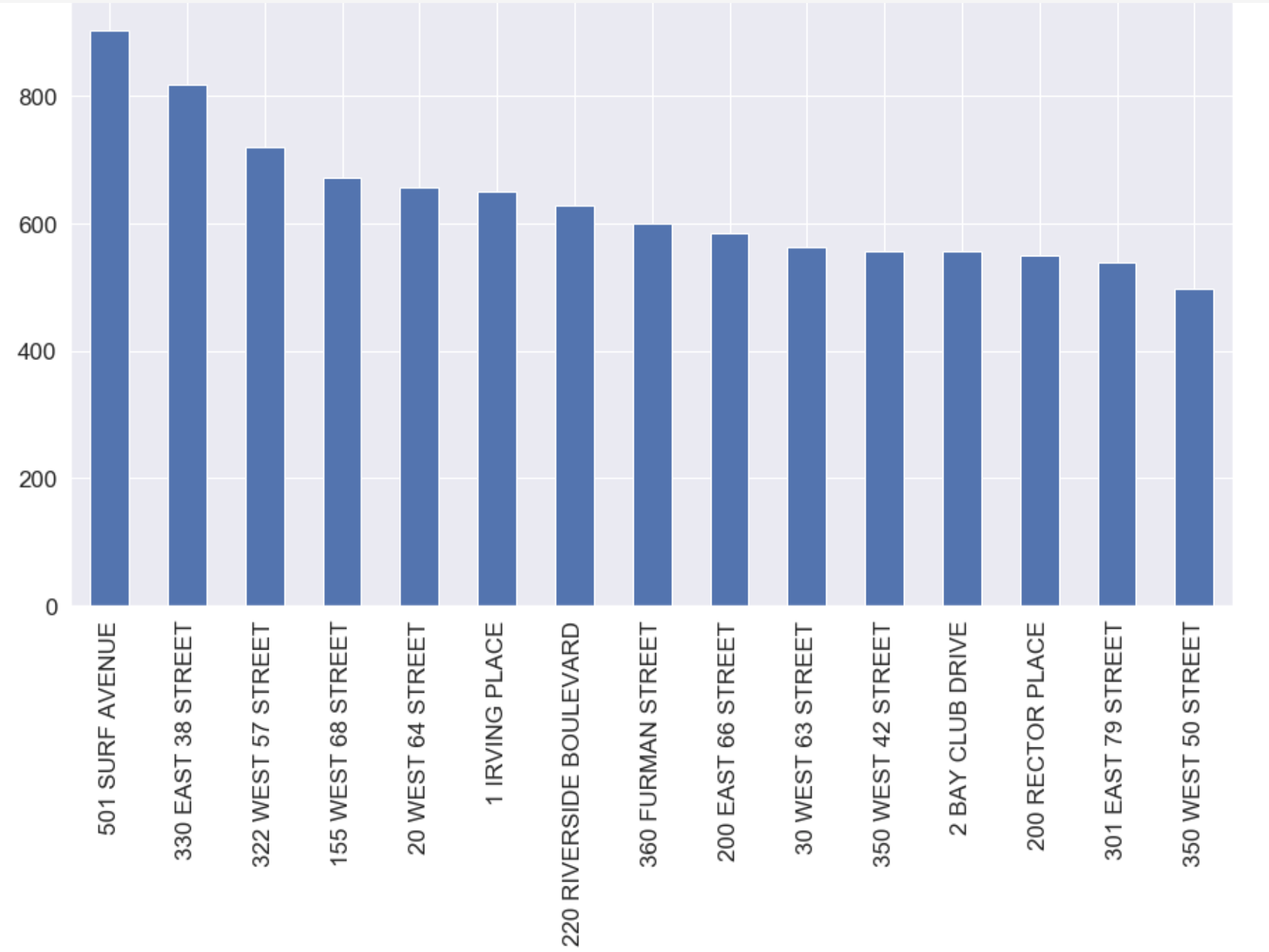
**Description:** exemption code 1.



**20.**

**Field name:** STADDR

**Field type:** object

**Description:** street address. Top 15 addresses:

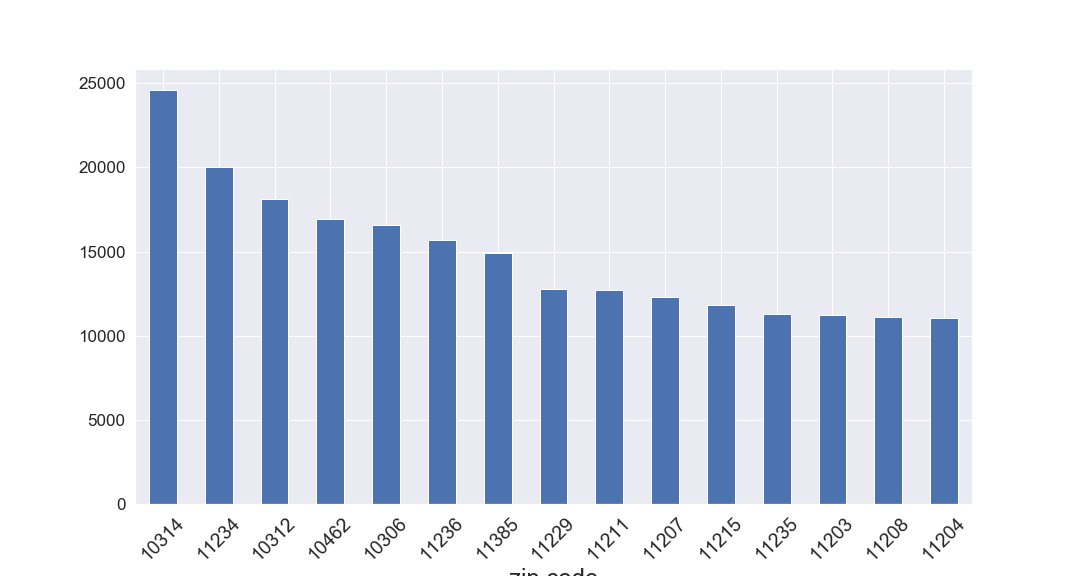
**21.**

**Field name:** ZIP

**Field type:** float64

**Description:** zip code.

Top 15 zip codes:

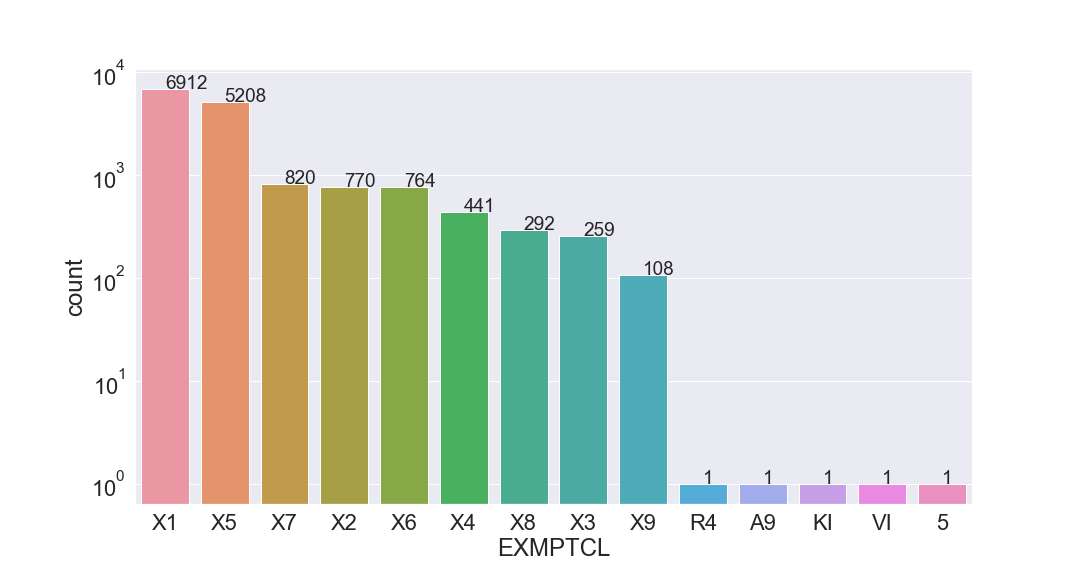


**22.**

**Field name:** EXMPTCL

**Field type:** object

**Description:** exempt class used for fully exempt properties only.

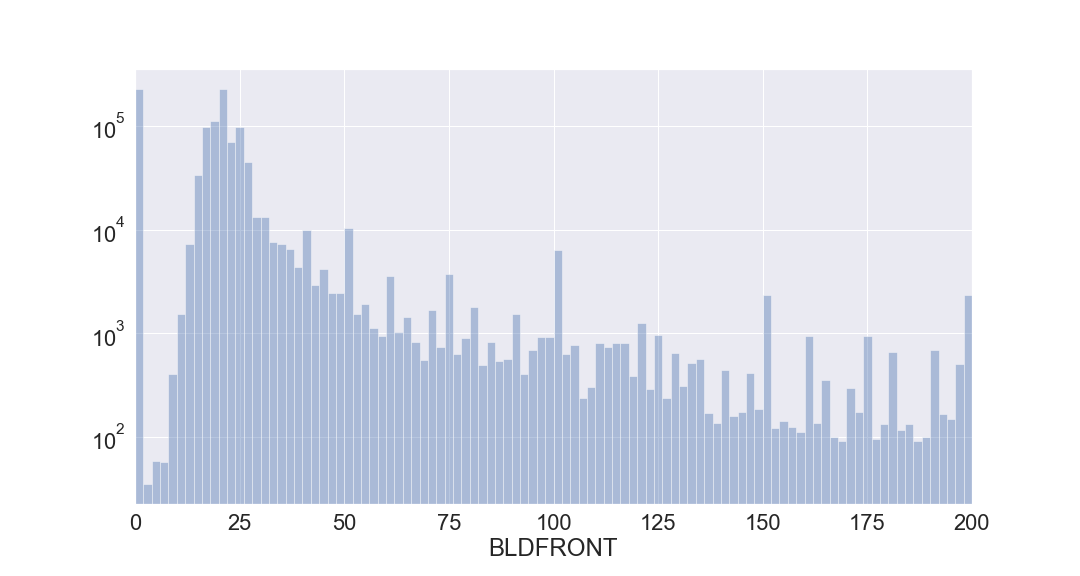


**23.**

**Field name:** BLDFRONT

**Field type:** int64

**Description:** building frontage in feet.

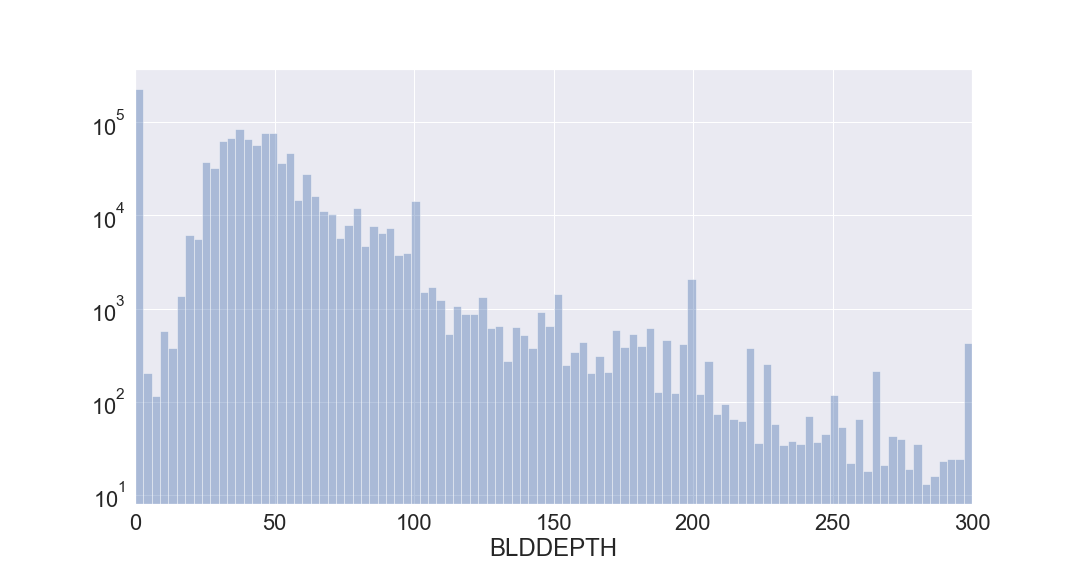


**24.**

**Field name:** BLDDEPTH

**Field type:** int64

**Description:** building depth in feet.

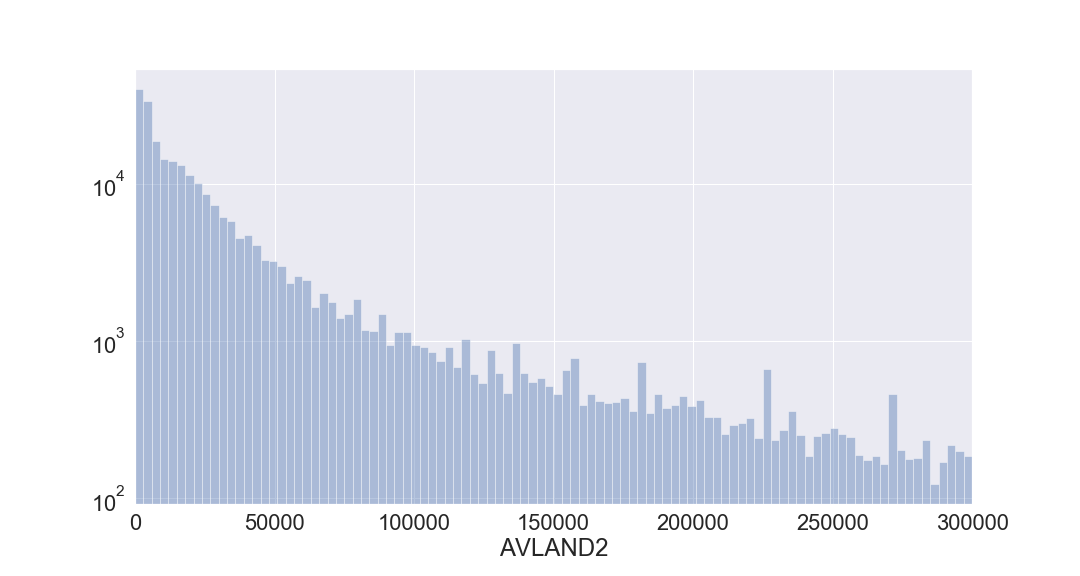


**25.**

**Field name:** AVLAND2

**Field type:** float64

**Description:** transitional land value.

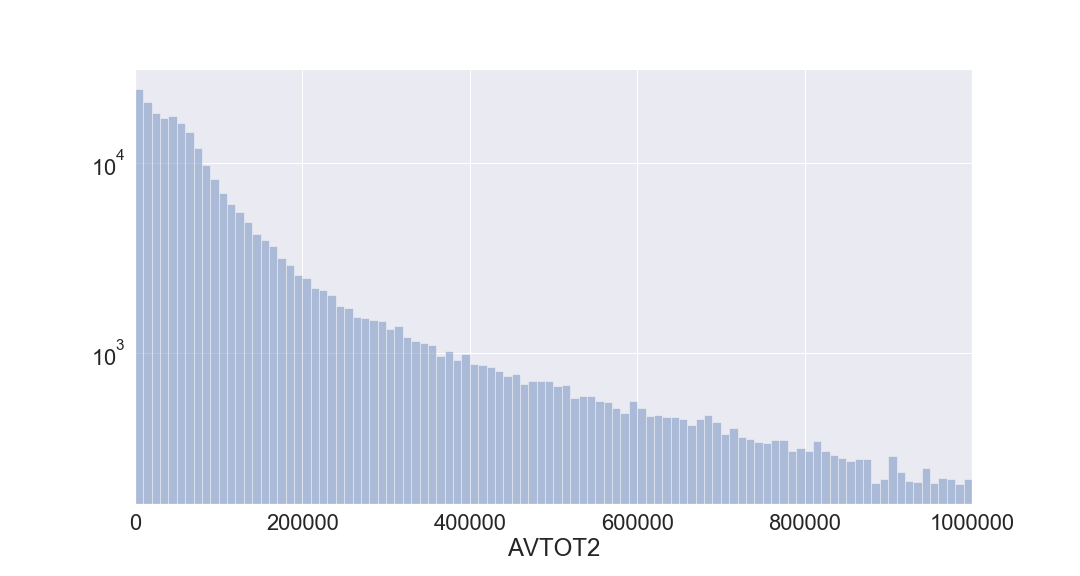
****

**26.**

**Field name:** AVTOT2

**Field type:** float64

**Description:** transitional total value.

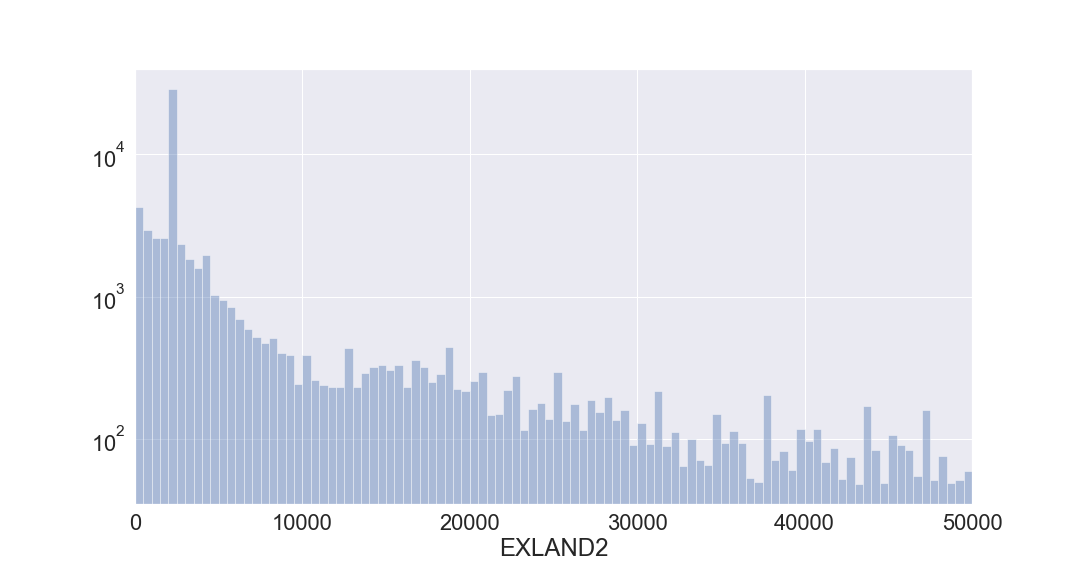
****

**27.**

**Field name:** EXLAND2

**Field type:** float64

**Description:** transitional exempt land value.

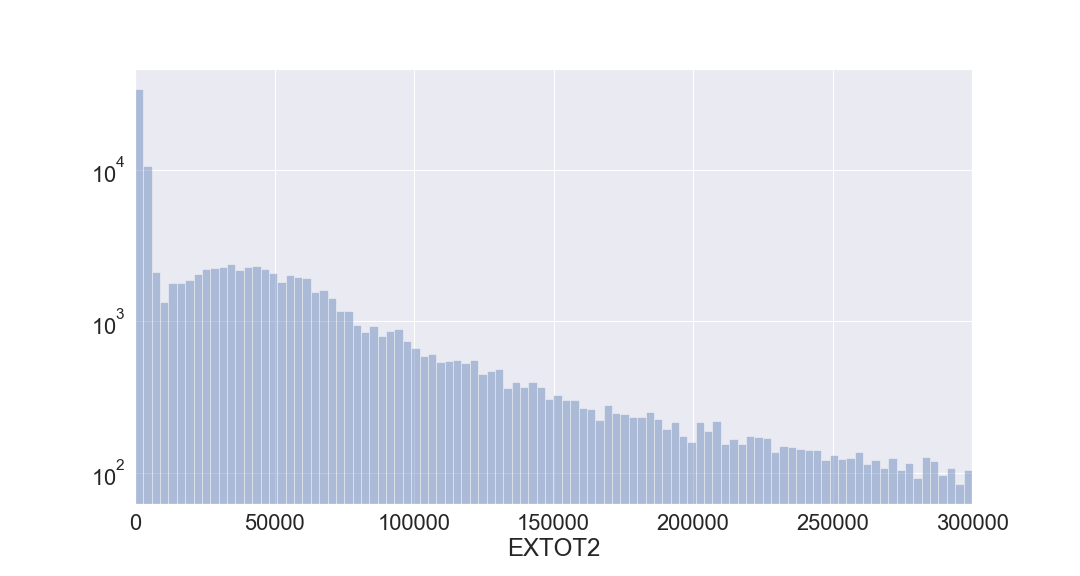
****

**28.**

**Field name:** EXTOT2

**Field type:** float64

**Description:** transitional exempt land total.

****

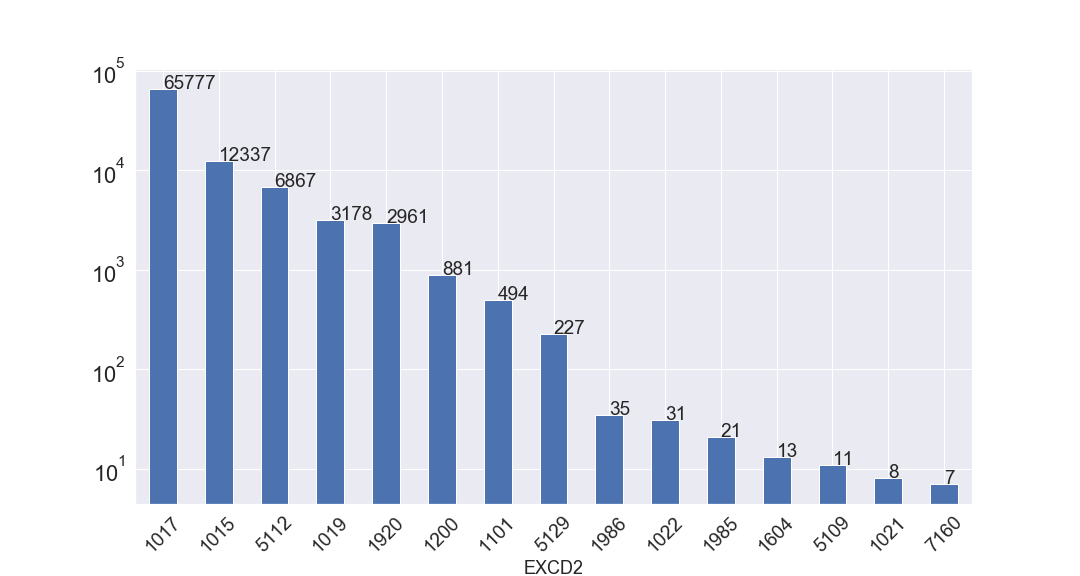
**29.**

**Field name:** EXCD2

**Field type:** float64

**Description:** exemption code 2.

Top 15 Exemption codes:

****

**30.**

**Field name:** PERIOD

**Field type:** object

**Description:** assessment period when file was created; FINAL.

**31.**

**Field name:** YEAR

**Field type:** object

**Description:** assessment year; 2010.

**32.**

**Field name:** VALTYPE

**Field type:** object

**Description:** AC-TR.