**Data Quality Report**

Yi-Hsin (Amy) Chung

**Data Overview**

This Data Quality Report summarized information of “NY property data.csv”, which contains a large amount of property records in New York City. The data is provided by Department of Finance (DOF) of NYC government and can be obtained from “Property Valuation and Assessment Data” on NYC OpenData. The data file has 32 fields and 1,070,994 records, and all records are within year 2010 and 2011.

The following contents include the two parts:

* Tables for Summary Statistics of Each Field
* Field Description and Graphs

**Tables for Summary Statistics of Each Field**

Table 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **General Information** | | | | | |
| **Field Name** | **Field Type** | **#Records Populated** | **%Records Populated(%)** | **#Unique Values** | **#Records with Value Zero** |
| RECORD | Categorical | 1070994 | 100.00 | 1070994 | n/a |
| BBLE | Categorical | 1070994 | 100.00 | 1070994 | n/a |
| B | Categorical | 1070994 | 100.00 | 5 | n/a |
| BLOCK | Categorical | 1070994 | 100.00 | 13984 | n/a |
| LOT | Categorical | 1070994 | 100.00 | 6366 | n/a |
| EASEMENT | Categorical | 4636 | 0.43 | 13 | n/a |
| OWNER | Categorical | 1039249 | 97.04 | 863347 | n/a |
| BLDGCL | Categorical | 1070994 | 100.00 | 200 | n/a |
| TAXCLASS | Categorical | 1070994 | 100.00 | 11 | n/a |
| LTFRONT | Numerical | 1070994 | 100.00 | 1297 | 169108 |
| LTDEPTH | Numerical | 1070994 | 100.00 | 1370 | 170128 |
| EXT | Categorical | 354305 | 33.08 | 4 | n/a |
| STORIES | Numerical | 1014730 | 94.75 | 112 | 0 |
| FULLVAL | Numerical | 1070994 | 100.00 | 109324 | 13007 |
| AVLAND | Numerical | 1070994 | 100.00 | 70921 | 13009 |
| AVTOT | Numerical | 1070994 | 100.00 | 112914 | 13007 |
| EXLAND | Numerical | 1070994 | 100.00 | 33419 | 491699 |
| EXTOT | Numerical | 1070994 | 100.00 | 64255 | 432572 |
| EXCD1 | Categorical | 638488 | 59.62 | 130 | n/a |
| STADDR | Categorical | 1070318 | 99.94 | 839281 | n/a |
| ZIP | Categorical | 1041104 | 97.21 | 197 | n/a |
| EXMPTCL | Categorical | 15579 | 1.45 | 15 | n/a |
| BLDFRONT | Numerical | 1070994 | 100.00 | 612 | 228815 |
| BLDDEPTH | Numerical | 1070994 | 100.00 | 621 | 228853 |
| AVLAND2 | Numerical | 282726 | 26.40 | 58592 | 0 |
| AVTOT2 | Numerical | 282732 | 26.40 | 111361 | 0 |
| EXLAND2 | Numerical | 87449 | 8.17 | 22196 | 0 |
| EXTOT2 | Numerical | 130828 | 12.22 | 48349 | 0 |
| EXCD2 | Categorical | 92948 | 8.68 | 61 | n/a |
| PERIOD | Categorical | 1070994 | 100.00 | 1 | n/a |
| YEAR | Date/Time | 1070994 | 100.00 | 1 | n/a |
| VALTYPE | Categorical | 1070994 | 100.00 | 1 | n/a |

Table 2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Numerical Fields** | | | | | | | |
| **Field Name** | **Mean** | **Standard Deviation** | **Min** | **25%** | **50%** | **75%** | **Max** |
| LTFRONT | 36.64 | 74.03 | 0.00 | 19.00 | 25.00 | 40.00 | 9999.00 |
| LTDEPTH | 88.86 | 76.40 | 0.00 | 80.00 | 100.00 | 100.00 | 9999.00 |
| STORIES | 5.01 | 8.37 | 1.00 | 2.00 | 2.00 | 3.00 | 119.00 |
| FULLVAL | 874264.51 | 11582430.00 | 0.00 | 304000.00 | 447000.00 | 619000.00 | 6150000000.00 |
| AVLAND | 85067.92 | 4057260.06 | 0.00 | 9180.00 | 13678.00 | 19740.00 | 2668500000.00 |
| AVTOT | 227238.17 | 6877529.31 | 0.00 | 18374.00 | 25340.00 | 45438.00 | 4668308947.00 |
| EXLAND | 36423.89 | 3981575.79 | 0.00 | 0.00 | 1620.00 | 1620.00 | 2668500000.00 |
| EXTOT | 91186.98 | 6508402.82 | 0.00 | 0.00 | 1620.00 | 2090.00 | 4668308947.00 |
| BLDFRONT | 23.04 | 35.58 | 0.00 | 15.00 | 20.00 | 24.00 | 7575.00 |
| BLDDEPTH | 39.92 | 42.71 | 0.00 | 26.00 | 39.00 | 50.00 | 9393.00 |
| AVLAND2 | 246235.72 | 6178962.56 | 3.00 | 5705.00 | 20145.00 | 62640.00 | 2371005000.00 |
| AVTOT2 | 713911.44 | 11652528.95 | 3.00 | 33912.00 | 79962.50 | 240551.00 | 4501180002.00 |
| EXLAND2 | 351235.68 | 10802212.67 | 1.00 | 2090.00 | 3048.00 | 31779.00 | 2371005000.00 |
| EXTOT2 | 656768.28 | 16072510.17 | 7.00 | 2870.00 | 37062.00 | 106840.75 | 4501180002.00 |

Table 3

|  |  |  |  |
| --- | --- | --- | --- |
| **Categorical Fields** | | | |
| **Field Name** | **Most Common Value** | **Frequency** | **Percentage (%)** |
| B | 4 | 358046 | 33.43 |
| BLOCK | 3944 | 3888 | 0.36 |
| LOT | 1 | 24367 | 2.28 |
| EASEMENT | E | 4148 | 0.39 |
| OWNER | PARKCHESTER PRESERVAT | 6020 | 0.56 |
| BLDGCL | R4 | 139879 | 13.06 |
| TAXCLASS | 1 | 660721 | 61.69 |
| EXT | G | 266970 | 24.93 |
| EXCD1 | 1017 | 425348 | 39.72 |
| STADDR | 501 SURF AVENUE | 902 | 0.08 |
| ZIP | 10314 | 24606 | 2.30 |
| EXMPTCL | X1 | 6912 | 0.65 |
| EXCD2 | 1017 | 65777 | 6.14 |
| PERIOD | FINAL | 1070994 | 100.00 |
| YEAR | 2010/11 | 1070994 | 100.00 |
| VALTYPE | AC-TR | 1070994 | 100.00 |

**Field Description and Graphs**

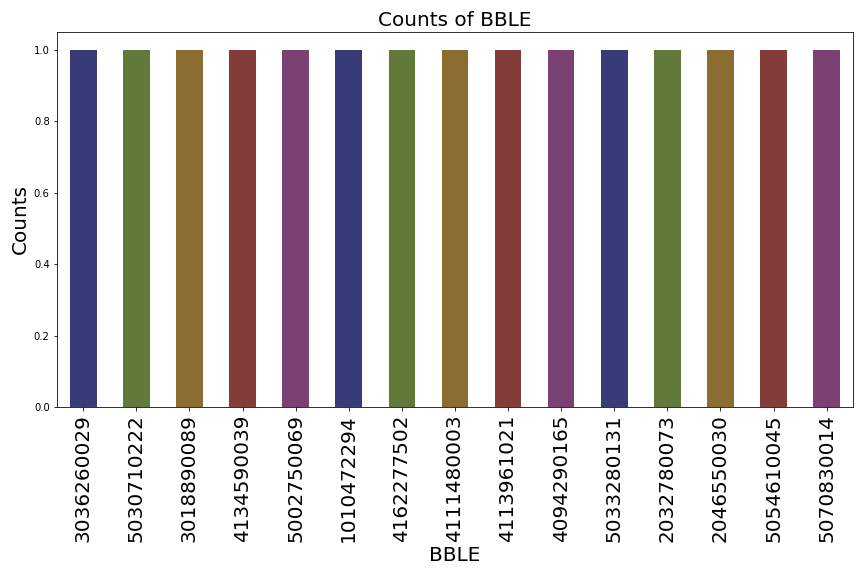
For each numerical field, I showed the distribution of the field with a histogram. For each categorical field, I showed the number of observations for top 15 categories with a bar chart.

1. RECORD

The RECORD field is a unique integer for each observation in the dataset, served for labeling each record.

1. BBLE

The BBLE field is a unique 11-digit alphanumeric text for each field. It is a concatenation of AV\_BORO, AV\_BLOCK, AV\_LOT, and AV\_EASEMENT. As you can see in the graph, each value has only one record.



1. B

The B field represents 5 boroughs, and the values for each borough are shown as below,

1 = Manhattan

2 = Bronx

3 = Brooklyn

4 = Queens

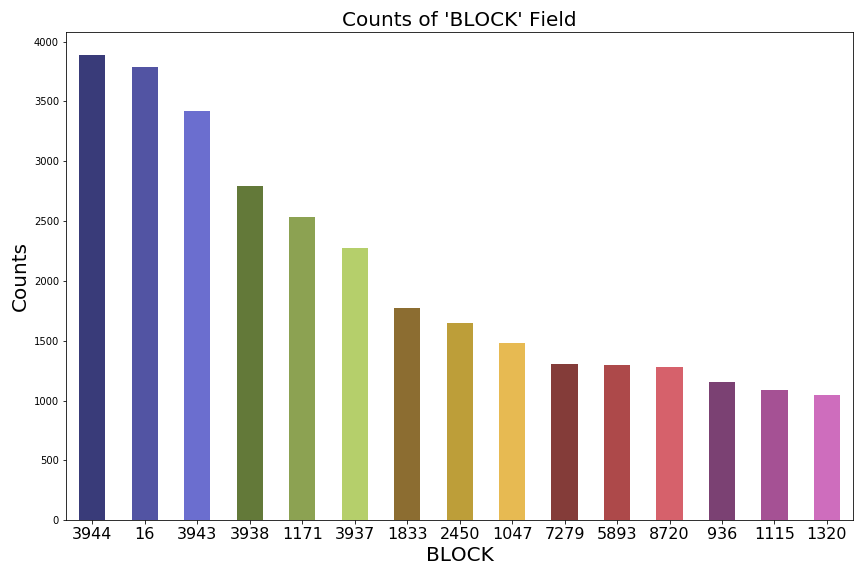
5 = Staten Island

A screenshot of a cell phone

Description automatically generated

1. BLOCK

The BLOCK field represents different block ranges within different boroughs.



1. LOT

The LOT field is a unique number within each borough or each block for every observation.

A screenshot of a cell phone

Description automatically generated

1. EASEMENT

The EASEMENT field is a variable that describes the following:

SPACE = the lot has no Easement

'A' = the portion of the Lot that has an Air Easement

'B' = Non-Air Rights

'E' = the portion of the lot that has a Land Easement

'F' THRU 'M' = Duplicates of 'E'

'N' = Non-Transit Easement

'P' = Piers

'R' = Railroads

'S' = Street

'U' = U.S. Government

A screenshot of a cell phone

Description automatically generated

1. OWNER

The OWNER field shows the name of the owner for the property.

A screenshot of a cell phone

Description automatically generated

1. BLDGCL

The BLDGCL field indicates the building class, and it has a direct correlation with TAXCLASS field.

A screenshot of a cell phone

Description automatically generated

1. TAXCLASS

The TAXCLASS field shows the current property tax class code (NYS Classification), and the first position of the tax class has a direct correlation with the building class.

1 = 1-3 UNIT RESIDENCES

1A = 1-3 STORY CONDOMINIUMS, ORIGINALLY A CONDO

1B = RESIDENTIAL VACANT LAND

1C = 1-3 UNIT CONDOMINUMS, ORIGINALLY TAX CLASS 1

1D = SELECT BUNGALOW COLONIES

2 = APARTMENTS

2A = APARTMENTS WITH 4-6 UNITS

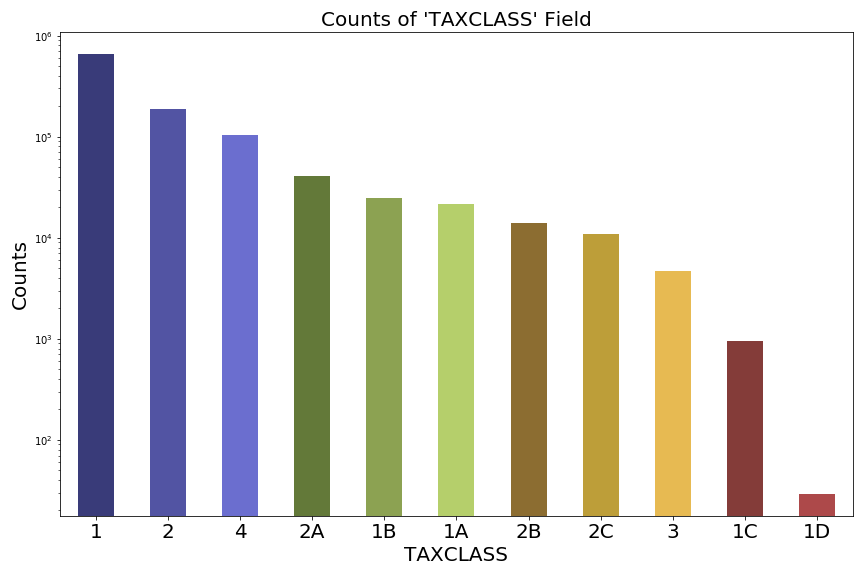
2B = APARTMENTS WITH 7-10 UNITS

2C = COOPS/CONDOS WITH 2-10 UNITS

3 = UTILITIES (EXCEPT CEILING RR)

4A = UTILITIES - CEILING RAILROADS

4 = ALL OTHERS



1. LTFRONT

The LTFRONT field is the lot width in feet. I include records with value <= 250.

A screenshot of a cell phone

Description automatically generated

1. LTDEPTH

The LTDEPTH field is the lot depth in feet. I include records with value <= 220.

A screenshot of a social media post

Description automatically generated

1. EXT

The EXT field indicates the information for extension.

'E' = EXTENSION

'G' = GARAGE

'EG' = EXTENSION AND GARAGE

A screenshot of a cell phone

Description automatically generated

1. STORIES

The STORIES field represents the number of stories for the building (# of Floors). I include records with value <= 50.

A screenshot of a social media post

Description automatically generated

1. FULLVAL

The FULLVAL field is the market value of the property. I include records with value <= 1600000.

A picture containing screenshot

Description automatically generated

1. AVLAND

The AVLAND field is the actual land value of the property. I include records with value <= 50000.

A screenshot of a cell phone

Description automatically generated

1. AVTOT

The AVTOT field is the actual total value of the property. I include records with value <= 100000.

A screenshot of a social media post

Description automatically generated

1. EXLAND

The EXLAND field is the actual exempt land value of the property. I include records with value <= 10000.

A screenshot of a cell phone

Description automatically generated

1. EXTOT

The EXTOT field is the actual exempt total value of the property. I include records with value <= 20000.

A screenshot of a cell phone

Description automatically generated

1. EXCD1

The EXCD1 field represents the exemption code 1 for each property.

A screenshot of a cell phone

Description automatically generated

1. STADDR

The STADDR field is the street address for the property.

A screenshot of a cell phone

Description automatically generated

1. ZIP

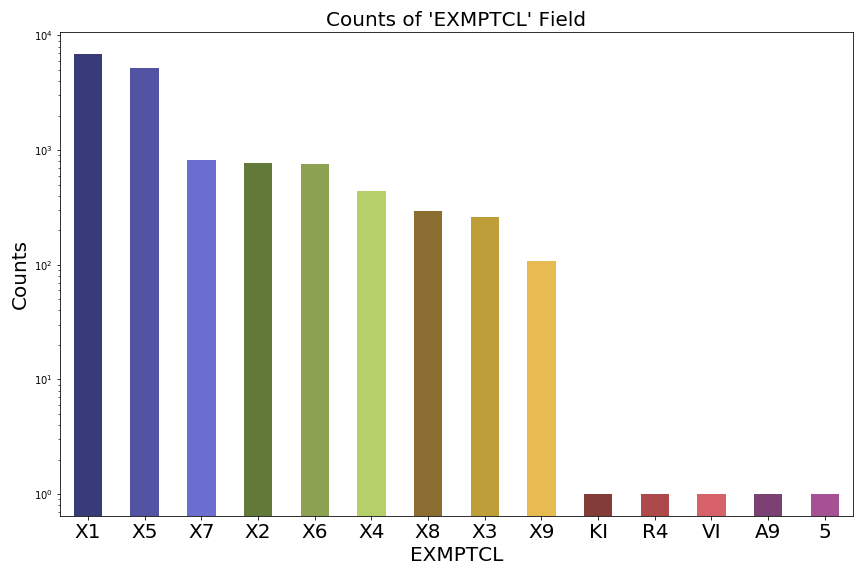
The ZIP field shows the zip code for the property.

A screenshot of a cell phone

Description automatically generated

1. EXMPTCL

The EXMPTCL field indicates the exempt class for fully exempt properties.



1. BLDFRONT

The BLDFRONT field is the building width in feet. I include records with value <= 210.

A screenshot of a social media post

Description automatically generated

1. BLDDEPTH

The BLDDEPTH field is the building depth in feet. I include records with value <= 210.

A screenshot of a cell phone

Description automatically generated

1. AVLAND2

The AVLAND2 field is the transitional land value of the property. I include records with value <= 150000.

A picture containing screenshot

Description automatically generated

1. AVTOT2

The AVTOT2 field is the transitional total value of the property. I include records with value <= 700000.

A picture containing screenshot

Description automatically generated

1. EXLAND2

The EXLAND2 field is the transitional exempt land value of the property. I include records with value <= 300000.

A screenshot of a cell phone

Description automatically generated

1. EXTOT2

The EXTOT2 field is the transitional exempt total value of the property. I include records with value <= 300000.

A screenshot of a cell phone

Description automatically generated

1. EXCD2

The EXCD2 field represents the exemption code 2 for each property.

A screenshot of a cell phone

Description automatically generated

1. PERIOD

The PERIOD field shows the assessment period when the file was created. The field is filled with ‘FINAL’ for all records.

A screenshot of a cell phone

Description automatically generated

1. YEAR

The YEAR field indicates the year of assessment, which is filled with ‘2010/11’ for all records.

A screenshot of a cell phone

Description automatically generated

1. VALTYPE

The VALTYPE field shows the value type, which is filled with ‘AC-TR’ for all records.

A screenshot of a cell phone

Description automatically generated