**DATA QUALITY REPORT FOR PROPERTY VALUATION AND ASSESSMENT DATA**

**Description**

**Dataset Name**: Property Valuation and Assessment Data

**Data Source**: New York City Department of Finance

**Time Period**: 2010/11

**Number of Fields**: #32

**Number of Records**: 1,070,994

**Description:**

This dataset includes over one million valuation and assessment information for properties in New York City, under investigation of Department of Finance. It aims to value properties based on 32 fields as an important step in calculating proper tax bill and find potential tax fraud.

**Summary Table**

**Numeric Fields:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Field  Name | percent  Populated | Minimum  Value | Maximum  Value | Mean  Value | Standard  Deviation | percent  Zero |
| LTFRONT | 100 | 0 | 9,999 | 36.64 | 74.03 | 15.79 |
| LTDEPTH | 100 | 0 | 9,999 | 88.86 | 76.40 | 15.89 |
| STORIES | 94.75 | 1 | 119 | 5.01 | 8.37 | 0 |
| FULLVAL | 100 | 0 | 6,150,000,000 | 874,264.50 | 11,582,430 | 1.21 |
| AVLAND | 100 | 0 | 2,668,500,000 | 85,067.92 | 4,057,260 | 1.21 |
| AVTOT | 100 | 0 | 4,668,309,000 | 227,238.20 | 6,877,529 | 1.21 |
| EXLAND | 100 | 0 | 2,668,500,000 | 36,423.89 | 3,981,576 | 45.91 |
| EXTOT | 100 | 0 | 4,668,309,000 | 91,186.98 | 6,508,403 | 40.39 |
| BLDFRONT | 100 | 0 | 7,575 | 23.04 | 35.58 | 21.36 |
| BLDDEPTH | 100 | 0 | 9,393 | 39.92 | 42.71 | 21.37 |
| AVLAND2 | 26.40 | 3 | 2,371,005,000 | 246,235.70 | 6,178,963 | 0 |
| AVTOT2 | 26.40 | 3 | 4,501,180,000 | 713,911.40 | 11,652,530 | 0 |
| EXLAND2 | 8.17 | 1 | 2,371,005,000 | 351,235.70 | 10,802,210 | 0 |
| EXTOT2 | 12.22 | 7 | 4,501,180,000 | 656,768.30 | 16,072,510 | 0 |
| YEAR | 100 | 2010/11 | 2010/11 | None | None | 0 |

**Categorical Fields:**

|  |  |  |  |
| --- | --- | --- | --- |
| Field Name | percent  Populated | num of Unique Values | Most Common Value |
| RECORD | 100 | 1,070,994 | Same |
| BBLE | 100 | 1,070,994 | Same |
| BORO | 100 | 5 | 4 |
| BLOCK | 100 | 13,984 | 3,944 |
| lot | 100 | 6366 | 1 |
| EASEMENT | 0.43 | 12 | E |
| OWNER | 97.04 | 863,347 | PARKCHESTER PRESERVAT |
| BLDGCL | 100 | 200 | R4 |
| TAXCLASS | 100 | 11 | 1 |
| EXT | 33.08 | 3 | G |
| EXCD1 | 59.62 | 129 | 1,017.0 |
| STADDR | 99.94 | 839,280 | 501 SURF AVENUE |
| ZIP | 97.21 | 196 | 10,314.0 |
| EXMPTCL | 1.45 | 14 | X1 |
| EXCD2 | 8.68 | 60 | 1,017.0 |
| PERIOD | 100 | 1 | FINAL |
| VALTYPE | 100 | 1 | AC-TR |

**Data Field Exploration:**

**Field 1:**

Name: RECORD

Description: Unique identifier of each property valuation and assessment data in the dataset. This field has 1,070,994 unique values (integers) range from 1 to 1,070,994.

**Field 2:**

Name: BBLE

Description: File key, each stands for a borough, block, lot, or easement code. This field has 1,070,994 unique values.

**Field 3:**

Name: BORO

Description: Borough code, 5 unique values stand for 5 different regions:

1 = Manhattan

2 = Bronx

3 = Brooklyn

4 = Queens

5 = Staten Island

Among those numbers, 4 occurs most (358,046 times).

Chart, bar chart

Description automatically generated

**Field 4:**

Name: BLOCK

Description: Valid block ranges by borough, 13,894 unique values in 5 different ranges stand for blocks in 5 different boroughs listed in field 4:

Manhattan: 1 to 2255

Bronx: 2260 to 5958

Brooklyn: 1 to 8955

Queens: 1 to 16350

Staten Island: 1 to 8050

Among those numbers, 3944 occurs most (3,888 times).

The histogram here shows top 20 values of block count.

Chart, bar chart

Description automatically generated

**Field 5:**

Name: LOT

Description: The lot codes of the property. This field has 6366 unique values, and the most common value is 1, which occurs 24,367 times. This field does not have missing values.

The histogram here shows top 20 values of lot counts.

Chart, bar chart

Description automatically generated

**Field 6:**

Name: EASEMENT

Description: Property easement, 12 unique values stand for 12 different types of easements. The most common easement type is E, which occurs 4,148 times. This filed has 99.57% null values.

The y of histogram posted here is in the log scale and the text in the graph shows the actual count amount for each type.

Chart, bar chart

Description automatically generated

**Field 7:**

Name: OWNER

Description: Owner name of the property, 863,347 unique values in this field and the most common owner is ‘PARKCHESTER PRESERVAT’, which occurs 6021 times. This filed has 2.96% null values.

The histogram here shows top 20 values of counts.

Chart

Description automatically generated

**Field 8:**

Name: BLDGCL

Description: Building Class of the property, 200 unique values in this field and the most common building class is R4, which occurs 139,879 times. This filed has no null values.

The histogram here shows top 20 values of counts.

Chart, bar chart

Description automatically generated

**Field 9:**

Name: TAXCLASS

Description: Tax Class, 11 unique values in this field can be divided into 4 different types based on the first number.

1 = 1 - 3 Unit Residence

2 = Apartments,2A = 4, 5, or 6 Units

3 = Utilities

4 = All Others

The most common tax class is 1, which occurs 660,721 times. This filed has no null values.

The y of histogram posted here is in the log scale and the text in the graph shows the actual count amount for each type.

Chart, bar chart

Description automatically generated

**Field 10:**

Name: LTFRONT

Description: Lot width, ranging from 0 to 9999 with mean equals to 36.64 and standard deviation equals to 74.03. This field has no missing value but has 169,108 zero values (accounts for 15.79%). 0 is the most common value. However, according to field’s actual meaning, 0 and unusually small values seem to be missing data.

The graph posted here only includes the lot width with smaller than or equal to 150 (1,040,554 data), accounting for 97.16% of total number. Plus, I divide the data into 100 bins in this graph.

Chart, histogram

Description automatically generated

**Field 11:**

Name: LTDEPTH

Description: Lot depth, ranging from 0 to 9999 with mean equals to 88.86 and standard deviation equals to 76.4. This field has no missing value but has 170,128 zero values (accounts for 15.89%). 100 is the most common value, occurring 464,541 times. Plus, according to field’s actual meaning, 0 and unusually small values seem to be missing data.

The graph posted here only includes the lot depth with smaller than or equal to 150 (1,027,205 data), accounting for 95.91% of total number. Plus, I divide the data into 100 bins in this graph.

Chart, histogram

Description automatically generated

**Field 12:**

Name: EXT

Description: Extension indicator, 3 unique values in this field and the most common extension indicator is G, which occurs 266,970 times. This filed has 66.92% null values.

The histogram here shows values of counts.

Chart, bar chart

Description automatically generated

**Field 13:**

Name: STORIES

Description: Number of stories in building, ranging from 1 to 119 with mean equals to 5.01 and standard deviation equals to 8.37. This field has 5.25% missing value but has no zero values. The most common value is 2, which occurs 415,092 times.

The graph posted here only includes the number of stories with smaller than or equal to 20 (960,160 data), accounting for 94.62% of total number. Plus, I divide the data into 21 bins in this graph.

Chart, histogram

Description automatically generated

**Field 14:**

Name: FULLVAL

Description: Market value of the property, ranging from 0 to 6,150,000,000 with mean equals to 874,264.5 and standard deviation equals to 11,582,430. This field has no missing value but has 1.21% zero values.

The graph posted here only includes the market value with less than or equal to 2,000,000 (1,031,498 data), accounting for 96.31% of total number. Plus, I divide the data into 100 bins in this graph.

Chart, histogram

Description automatically generated

**Field 15:**

Name: AVLAND

Description: Actual land value of property, ranging from 0 to 2,668,500,000 with mean equals to 85,067.92 and standard deviation equals to 4,057,260. This field has no missing value but has 1.21% zero values.

The graph posted here only includes the actual land value with less than or equal to 50,000 (969,541 data), accounting for 90.53% of total number. Plus, I divide the data into 100 bins in this graph.

Chart, histogram

Description automatically generated

**Field 16:**

Name: AVTOT

Description: Actual total value of property, ranging from 0 to 4,668,309,000 with mean equals to 227,238.2 and standard deviation equals to 6,877,529. This field has no missing value but has 1.21% zero values.

The graph posted here only includes the actual total value with less than or equal to 100,000 (921,616 data), accounting for 86.05% of total number. Plus, I divide the data into 100 bins in this graph.

Chart, histogram

Description automatically generated

**Field 17:**

Name: EXLAND

Description: Actual exempt land value, ranging from 0 to 2,668,500,000 with mean equals to 36,423.89 and standard deviation equals to 3,981,576. This field has no missing value but has 45.91% zero values.

The graph posted here only includes the actual exempt land value with less than or equal to 20,000 (1,036,987 data), accounting for 96.82% of total number. Plus, I divide the data into 100 bins.

Chart, histogram

Description automatically generated

**Field 18:**

Name: EXTOT

Description: Actual exempt total value, ranging from 0 to 4,668,309,000 with mean equals to 91,186.98 and standard deviation equals to 6,508,403. This field has no missing value but has 40.39% zero values.

The graph posted here only includes the actual exempt total value with less than or equal to 10,000 (913,903 data), accounting for 85.33% of total number. Plus, I divide the data into 100 bins in this graph.

Chart, histogram

Description automatically generated

**Field 19:**

Name: EXCD1

Description: Exemption code 1, 129 unique values in this field and the most common exemption code is 1017.0, which occurs 425,348 times. This filed has 40.38% null values.

The histogram here shows top 20 values of counts, and the y-axis is in log scale.

Chart, bar chart

Description automatically generated

**Field 20:**

Name: STQDDR

Description: Street address, 839,280 unique values in this field and the most common street address is ‘501 SURF AVENUE’, which occurs 902 times. This filed has 0.06% null values.

The histogram here shows top 20 values of counts.

Chart, bar chart

Description automatically generated

**Field 21:**

Name: ZIP

Description: Zip code, 196 unique values in this field and the most common zip code is 10314.0, which occurs 24,606 times. This filed has 2.79% null values.

The histogram here shows top 20 values of counts.

Chart, bar chart

Description automatically generated

**Field 22:**

Name: EXMPTCL

Description: Exemption class, 14 unique values in this field and the most common exemption class is X1, which occurs 6,912 times. This filed has 98.55% null values.

The histogram here shows values of counts and y-axis is in log scale.

Chart, bar chart, histogram

Description automatically generated

**Field 23:**

Name: BLDFRONT

Description: Building width, ranging from 0 to 7,575 with mean equals to 23.04 and standard deviation equals to 35.58. This field has no missing value but has 228,815 zero values (accounts for 21.36%). According to field’s actual meaning, 0 and unusually small values seem to be missing data.

The graph posted here only includes the building width with smaller than or equal to 120 (1,049,831 data), accounting for 98.02% of total number. Plus, I divide the data into 80 bins in this graph.

Chart, histogram

Description automatically generated

**Field 24:**

Name: BLDDEPTH

Description: Building depth, ranging from 0 to 9,393 with mean equals to 39.92 and standard deviation equals to 42.71. This field has no missing value but has 228,853 zero values (accounts for 21.36%). According to field’s actual meaning, 0 and unusually small values seem to be missing data.

The graph posted here only includes the building depth with smaller than or equal to 120 (1,051,043 data), accounting for 98.14% of total number. Plus, I divide the data into 80 bins in this graph.

Chart, histogram

Description automatically generated

**Field 25:**

Name: AVLAND2

Description: Transitional land value of property, ranging from 3 to 2,371,005,000 with mean equals to 246,235.7 and standard deviation equals to 6,178,963. This field has 73.6% missing value and no zero values.

The graph posted here only includes the transitional land value with less than or equal to 250,000 (255,067 data), accounting for 90.22% of total number. Plus, I divide the data into 100 bins in this graph.

Text

Description automatically generated with low confidence

**Field 26:**

Name: AVTOT2

Description: Transitional total value of property, ranging from 3 to 4,501,180,000 with mean equals to 713,911.4 and standard deviation equals to 11,652,530. This field has 73.6% missing value and no zero values.

The graph posted here only includes the transitional total value with less than or equal to 800,000 (254,052 data), accounting for 89.86% of total number. Plus, I divide the data into 100 bins in this graph.

Text

Description automatically generated

**Field 27:**

Name: EXLAND2

Description: Transitional exempt land value, ranging from 0 to 2,371,005,000 with mean equals to 351,235.7 and standard deviation equals to 10,802,210. This field has 91.83% missing value but has no zero values.

The graph posted here only includes the transitional exempt land value with less than or equal to 50,000 (68,738 data), accounting for 78.6% of total number. Plus, I divide the data into 100 bins in this graph.

Chart, histogram

Description automatically generated

**Field 28:**

Name: EXTOT2

Description: Transitional exempt total value, ranging from 7 to 4,501,180,000 with mean equals to 656,768.3 and standard deviation equals to 16,072,510. This field has 87.78% missing value but has no zero values.

The graph posted here only includes the transitional exempt total value with less than or equal to 300,000 (113,369 data), accounting for 86.65% of total number. Plus, I divide the data into 100 bins in this graph.

Chart, histogram

Description automatically generated

**Field 29:**

Name: EXCD2

Description: Exemption code 2, 60 unique values in this field and the most common exemption code is 1017.0, which occurs 65,777 times. This filed has 91.32% null values.

The histogram here shows top 20 values of counts, and the y-axis is in log scale.

Chart, bar chart, histogram

Description automatically generated

**Field 30:**

Name: PERIOD

Description: Assessment period when date was created. This field has no null values and zero values. Only ‘FINAL’ appears in this field.

**Field 31:**

Name: YEAR

Description: Assessment year. This field has no null values and zero values. Only 2010/11 appears in this field.

**Field 32:**

Name: VALTYPE

Description: Value type of the data. This field has no null values and zero values. Only ‘AC-TR’ (Do not know exact meaning) appears in this field.