Data Quality Report: NYC Property Data

Shreyash Kondakindi

September 21, 2025

Section 1 – Data Overview

This dataset contains **1,070,994** property assessment records from the NYC Department of Finance, covering the Final Property Tax Assessment Roll for fiscal year 2010/11. The records span all five New York City boroughs, with each property identified by a unique BBLE code. The dataset includes **32 fields** covering location, owner, physical attributes, valuation, and exemption details. Most fields are fully populated; some have structured missingness (e.g., exemption codes). The `PERIOD` and `YEAR` fields confirm this dataset refers entirely to the **2010/11** final roll.

Section 2 – Field Summary Tables

Categorical Fields

Field Name	# Records	% Populated	# Zeros	# Unique	Most Common
RECORD	1,070,994	100.00%	0	1,070,994	1
BBLE	1,070,994	100.00%	0	1,070,994	1000010101
BORO	1,070,994	100.00%	0	5	4
BLOCK	1,070,993	100.00%	0	13,984	3944
LOT	1,070,993	100.00%	0	6,366	1
EASEMENT	4,636	0.43%	0	12	E
OWNER	1,039,249	97.04%	0	863,347	PARKCHESTER PRESERVAT
BLDGCL	1,070,993	100.00%	0	200	R4
TAXCLASS	1,070,993	100.00%	0	11	1
EXT	354,305	33.08%	0	3	G
EXCD1	638,487	59.62%	0	197	1017
STADDR	1,070,318	99.94%	0	986,263	501 SURF AVENUE
ZIP	1,041,104	97.21%	0	181	10314
EXMPTCL	15,579	1.45%	0	14	X1
EXCD2	92,947	8.68%	0	135	1017
PERIOD	1,070,994	100.00%	0	1	FINAL
YEAR	1,070,994	100.00%	0	1	2010/11
VALTYPE	1,070,994	100.00%	0	1	AC-TR

Table 1: Summary of Categorical Fields

Numeric Fields

Field Name	# Records	% Populated	# Zeros	Min	Max	Mean	Std Dev
LTFRONT	1,070,994	100.00%	169,108	0	9999	36.64	74.03
LTDEPTH	1,070,994	100.00%	170,128	0	9999	97.01	112.04
STORIES	1,014,730	94.75%	0	1	119	5.01	8.37
FULLVAL	1,070,994	100.00%	13,007	0	6,150,000,000	874,264.51	11,582,425.58
AVLAND	1,070,994	100.00%	13,009	0	2,523,000,000	85,067.92	4,057,258.16
AVTOT	1,070,994	100.00%	13,007	0	2,610,000,000	227,238.17	6,877,526.09
EXLAND	1,070,994	100.00%	491,699	0	1,980,000,000	36,423.89	3,981,573.93
EXTOT	1,070,994	100.00%	432,572	0	2,038,500,000	91,186.98	6,508,399.78
BLDFRONT	1,070,994	100.00%	228,815	0	3,000	27.65	36.54
BLDDEPTH	1,070,994	100.00%	228,853	0	3,306	39.15	49.04
AVLAND2	282,726	26.40%	0	3	2,371,005,000	246,235.72	6,178,951.64
AVTOT2	282,732	26.40%	0	5	2,450,000,000	713,911.44	11,652,508.34
EXLAND2	87,948	8.21%	0	0.01	1,963,500,000	351,235.68	10,802,150.91
EXTOT2	130,828	12.22%	0	0.01	2,104,500,000	656,768.28	16,072,448.75

Table 2: Summary of Numeric Fields

Section 3 – Field-Level Analysis

RECORD

The `RECORD` field serves as a unique, sequential identifier for each record in the dataset. Values range from 1 to 1,070,994 with no duplicates or missing values. Every entry is distinct, confirming its function as a primary key. Although stored numerically, this field does not convey any measurable or analytical information beyond row indexing. It is best treated as a categorical identifier. Because each value is unique, a distribution or frequency plot is unnecessary.

BBLE

The `BBLE` (Borough-Block-Lot-Easement) field is a concatenated unique identifier for each property parcel. It combines borough, block, lot, and easement information into a single alphanumeric string. All records contain a valid BBLE, and every value is unique. Like `RECORD`, this field is better treated as categorical. While this field is foundational to property identification and joins across datasets, its values are not analytically meaningful in isolation.

BORO

The `BORO` field indicates the borough in which the property is located. It uses integer codes:

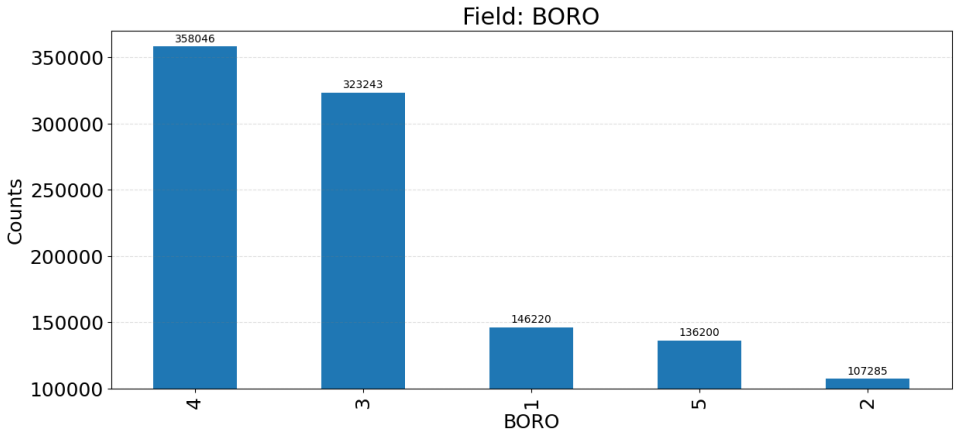

Figure 1: Distribution of Properties by Borough Code

- 1 = Manhattan
- 2 = Bronx
- 3 = Brooklyn
- 4 = Queens
- 5 = Staten Island

This field is fully populated, with all values within the valid range. The most frequent borough code is 4 (Queens), consistent with Queens having the most residential parcels. The borough distribution provides a helpful regional breakdown of the data and aligns with New York City's urban structure.

BLOCK

The `BLOCK` field identifies a tax block within each borough. Values are numeric and range from small integers up to over 16,000. A block number alone is not globally unique; it must be interpreted along with the borough code. All records have valid block numbers. The field is highly granular, with over 13,000 unique values, and most blocks contain multiple lots. We also observe that around 14000-15000, there are no block numbers.

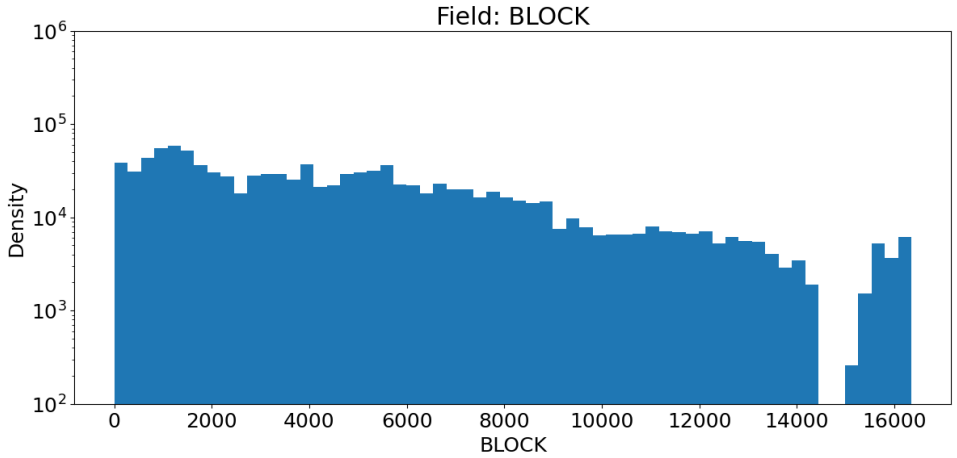

Figure 2: Distribution of BLOCK Numbers

LOT

`LOT` is a tax lot number within a given block. This field is fully populated and has a wide range of values. The most frequent lot number is 1, which appears in many different blocks. While the numbers are numeric, the combination of borough, block, and lot defines uniqueness—not the lot number alone. The distribution is consistent with NYC property structure, where each block can have several lots.

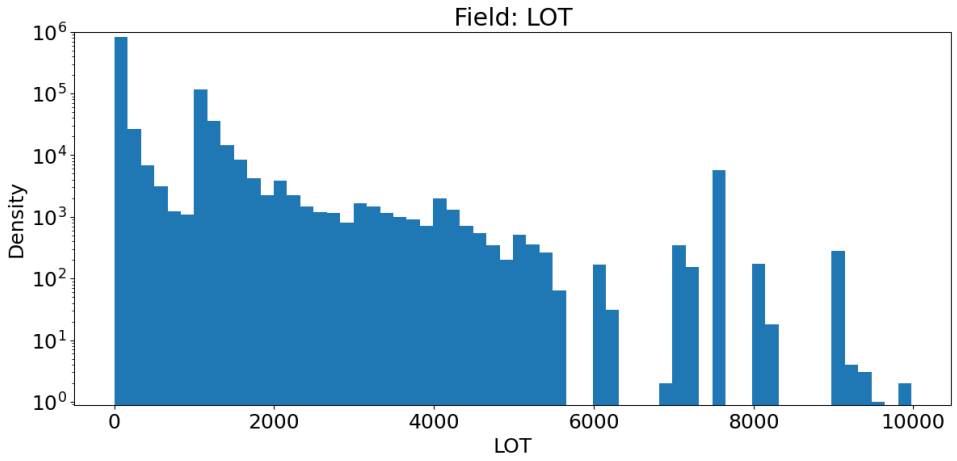

Figure 3: Distribution of LOT Numbers

EASEMENT

`EASEMENT` denotes special rights or access agreements on properties (e.g., utility easements, air rights). It is sparsely populated—only about 0.43% of the records contain a non-null value. Values are typically single letters, such as "E" (land easement) or "A" (air rights). The field is categorical and meaningful primarily for spatial planning or legal analysis. Despite the high number of missing values, the non-missing entries are consistent and valid.

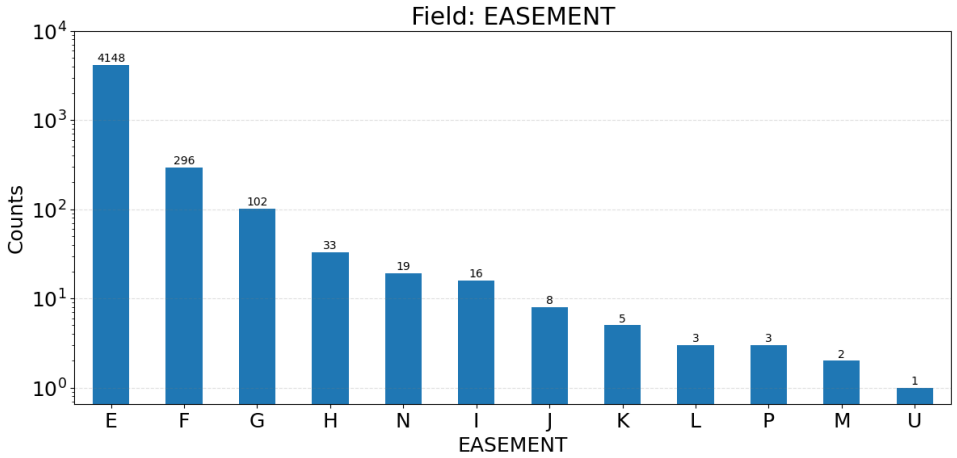

Figure 4: Frequency of EASEMENT Codes (Non-Null Only)

BLDGCL

`BLDGCL` represents the Building Class code and is a two-character alphanumeric identifier describing the property's structural use. Examples include "A1" (1-family dwelling), "C0" (vacant land), and "R4" (condominium units). The field is fully populated with over 200 unique values, reflecting the diversity of NYC's real estate. The distribution is skewed toward residential categories, especially R4 (condos) and A1 (single-family homes), consistent with the city's residential zoning. Building Class codes are crucial for tax assessment and land use classification.

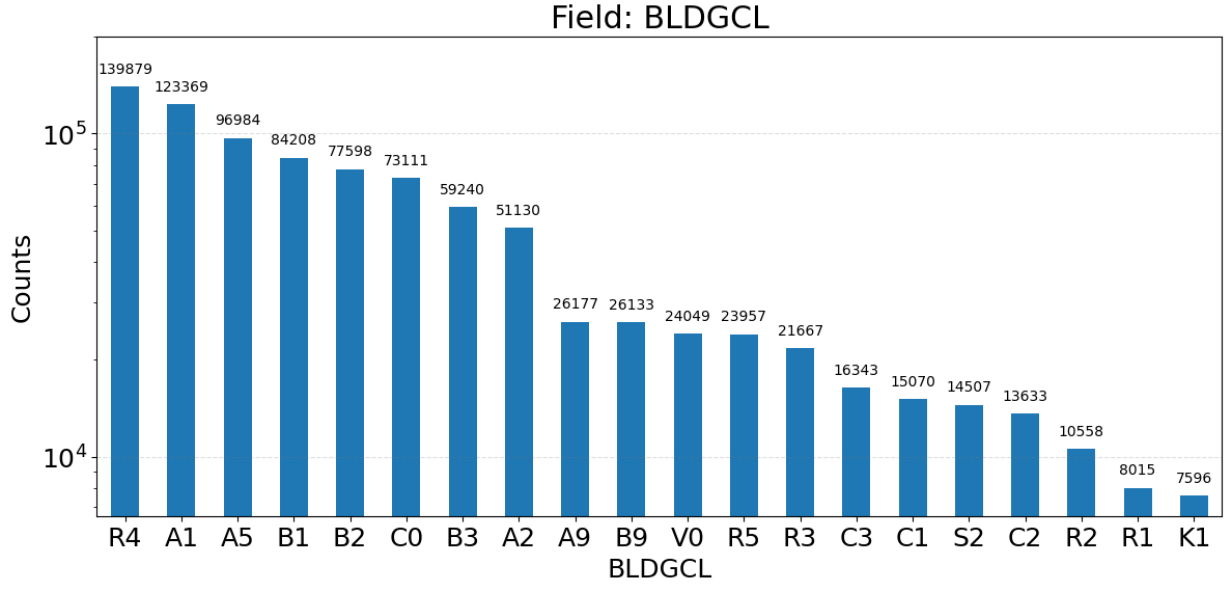

Figure 5: Top 20 Most Frequent Building Class Codes

OWNER

The `OWNER` field records the name of the property owner. It is 97% populated, with over 860,000 unique values. Many entries are institutional or corporate names (e.g., NYC Housing Authority, Parkchester Preservation). The most common name appears over 6,000 times, typically for large housing complexes. Though stored as strings, this field is best used for grouping, not statistical analysis. Values are typically uppercased and may be truncated or abbreviated due to system constraints.

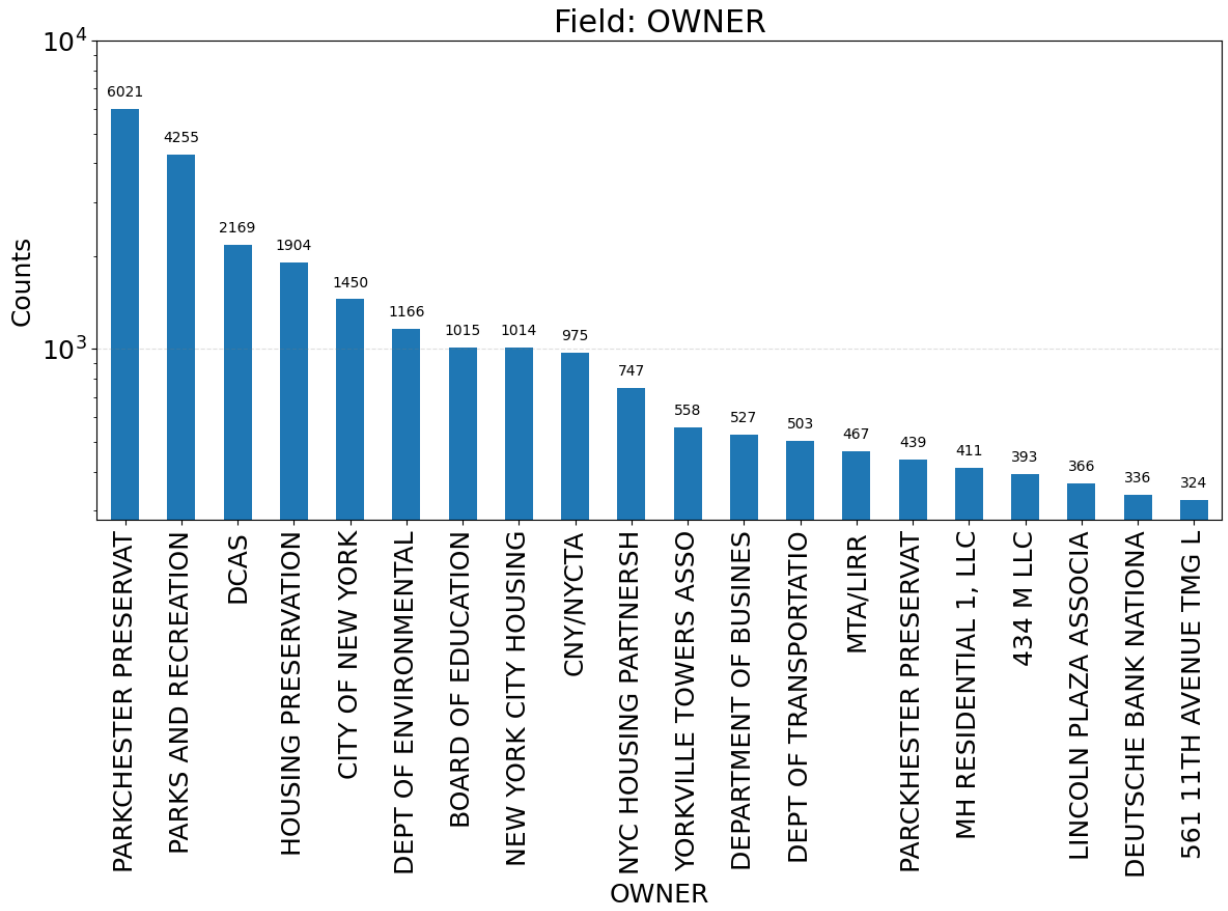

Figure 6: Top 20 Most Frequent OWNER Names

TAXCLASS

`TAXCLASS` defines how a property is taxed, with values such as:

- 1 – One-to-three family residential
- 2 – Multi-family residential (rental, condo, co-op)
- 3 – Utility
- 4 – Commercial/Industrial

The field includes some subclass distinctions (e.g., 1A, 2A), making a total of 11 unique values. The majority of properties fall into Class 1, followed by Class 2. This classification governs assessed value ratios and exemption applicability. Every record has a valid tax class, and no null or erroneous entries are present.

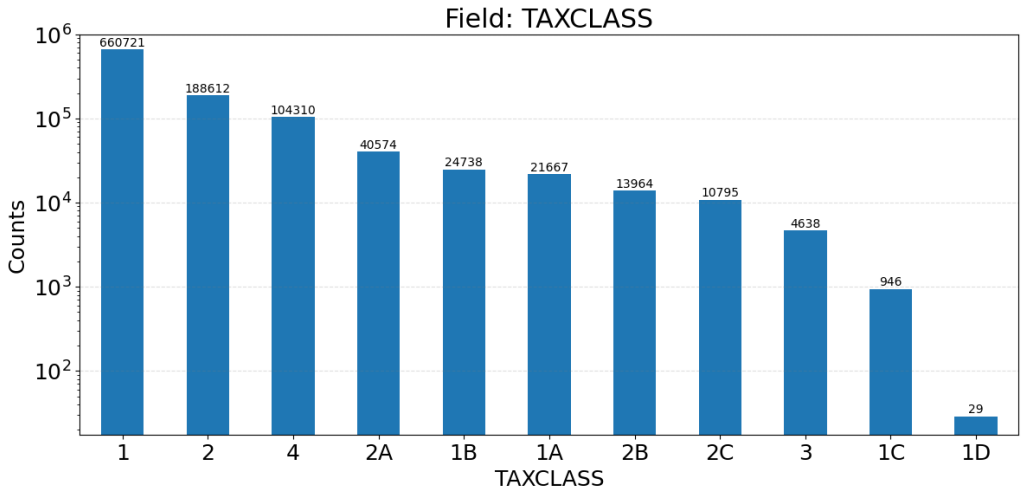

Figure 7: Distribution of TAXCLASS Values

EXT

`EXT` is a lesser-documented field likely referring to lot extension or irregularity. It contains codes like "G", "E", and "EG". Around 33% of the data has a non-empty value. These values may correspond to extended lots or special configurations (e.g., corner lots, through-lots), although detailed documentation is limited. Despite this, the consistency and limited vocabulary of the values suggest intentional and structured use.

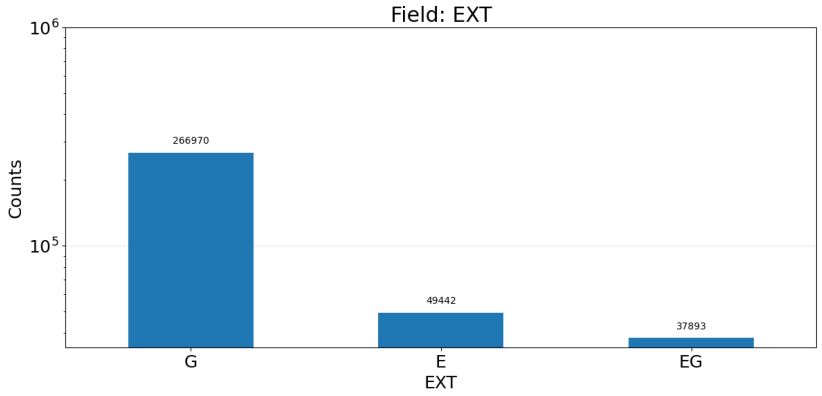

Figure 8: EXT Code Frequencies

EXCD1

`EXCD1` is the primary exemption code, indicating the type of property tax exemption applied (if any). Only about 60% of records have a value, reflecting the subset of properties that qualify for tax relief (e.g., STAR program, veterans, senior citizens). There are nearly 200 unique codes, with "1017" being overwhelmingly common—likely the STAR Basic exemption. This field is categorical, and analysis helps reveal the prevalence of exemption types.

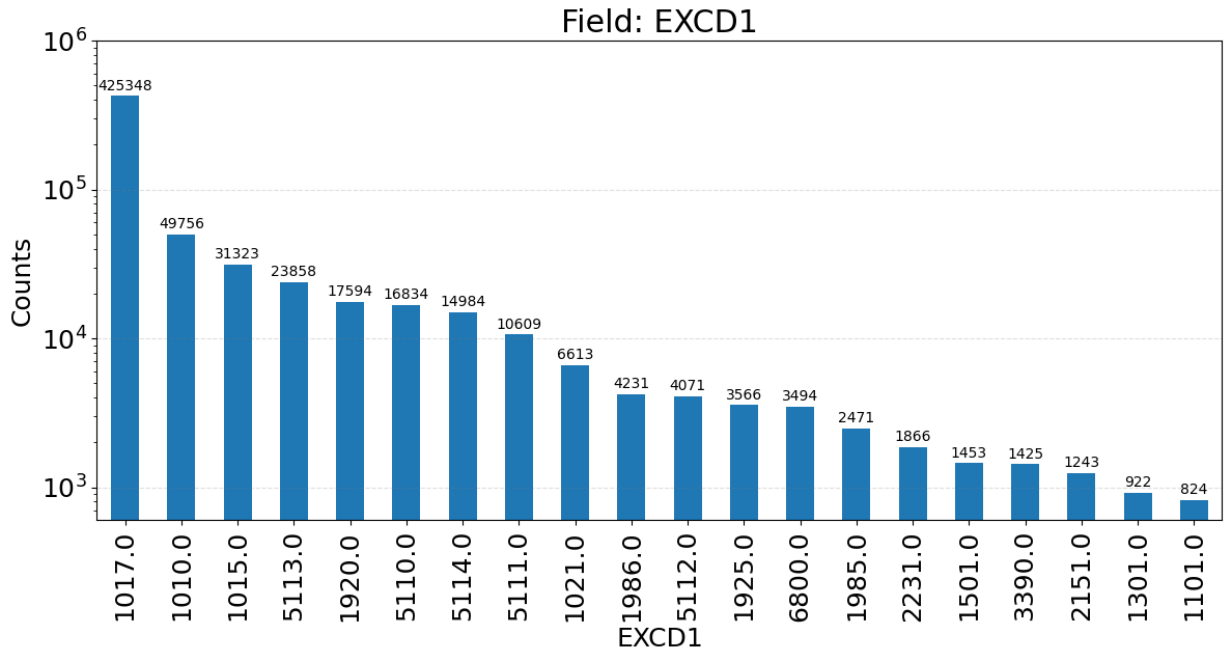

Figure 9: Top 20 EXCD1 (Primary Exemption Codes)

STADDR

The `STADDR` field holds the street address of the property, usually including street number and name (e.g., "501 SURF AVENUE"). It is over 99.9% populated and contains more than 980,000 unique values. Some addresses repeat across units in multi-unit buildings or condominiums, which is expected. Address values are uppercased and sometimes truncated due to formatting limitations. While not analytically numeric, `STADDR` is useful for mapping and aggregation.

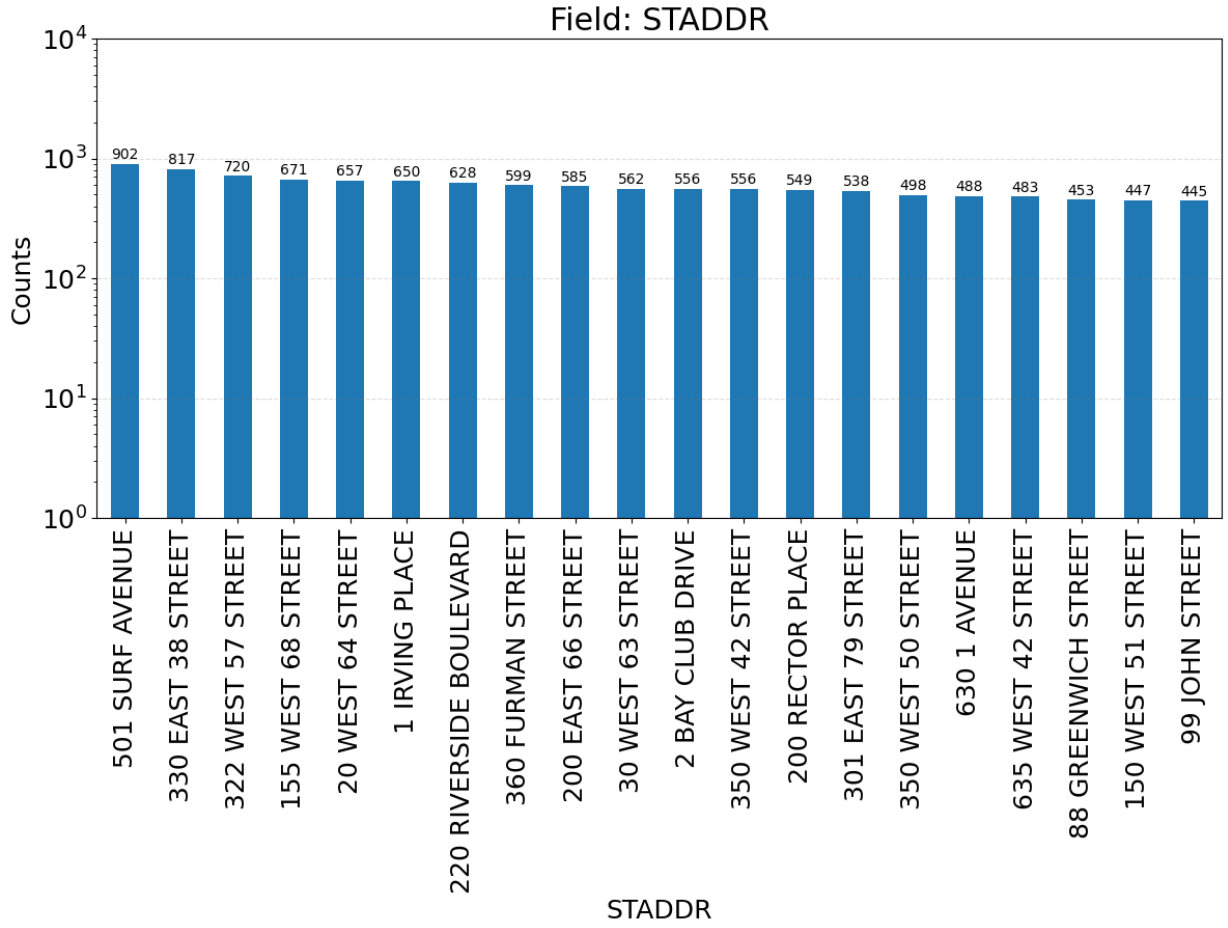

Figure 10: Top 20 Most Frequent STADDR Entries

ZIP

`ZIP` indicates the postal ZIP code for the property address. It is populated in over 97% of the dataset and includes 181 unique ZIP codes, consistent with the granularity of NYC neighborhoods. The most frequent ZIP code is 10314 (Staten Island), followed by various Brooklyn and Queens locations. Missing ZIPs are primarily associated with special-use parcels or those without typical mailing addresses. The field is well-structured and matches real NYC postal areas.

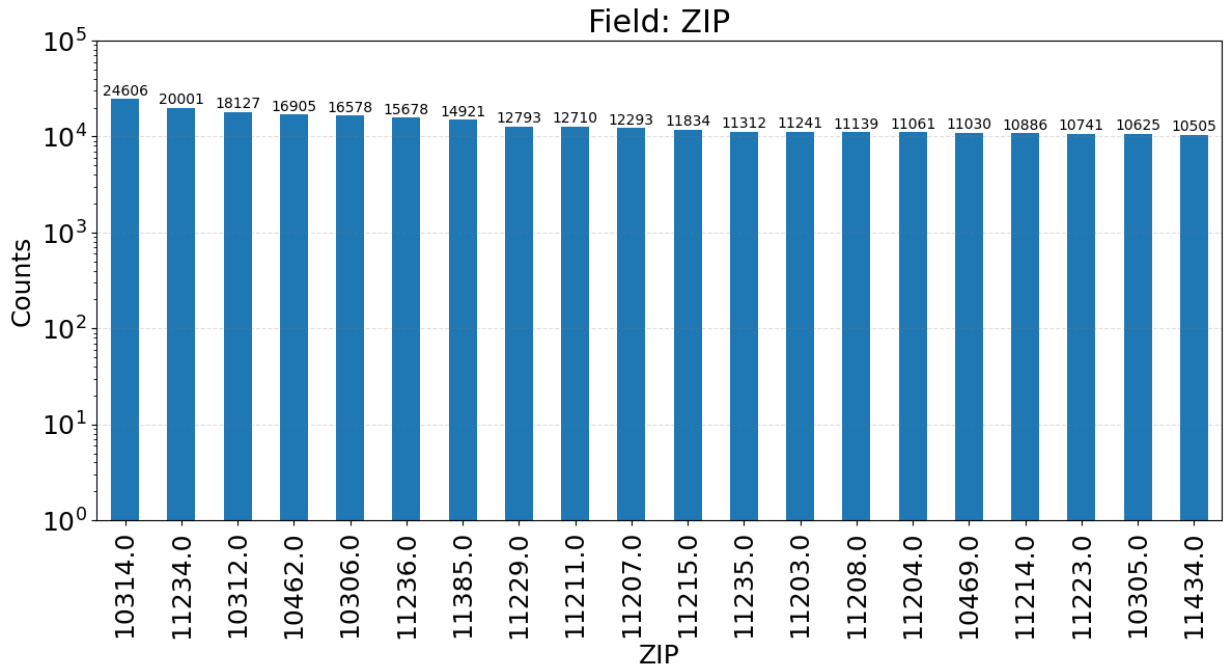

Figure 11: Top 20 Most Frequent ZIP Codes

EXMPTCL

`EXMPTCL` stands for Exemption Class and indicates whether a property is entirely tax-exempt. Only about 1.45% of records have a value, such as "X1" or "Y2", which correspond to government-owned or institutionally exempt parcels. Blank entries imply taxable properties. This field is useful for separating public/institutional properties from the taxable base and often aligns with ownership fields like "CITY OF NEW YORK".

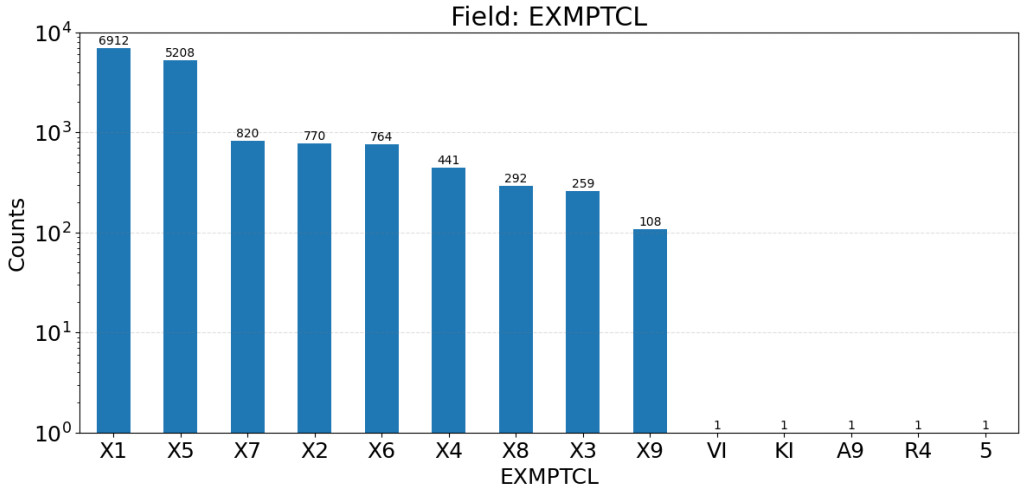

Figure 12: Distribution of Exemption Class (EXMPTCL)

EXCD2

`EXCD2` is a secondary exemption code, indicating an additional exemption applied to the same property. It is populated in about 8.7% of records. The most frequent code here is again 1017, suggesting some properties have dual STAR-type exemptions or overlapping programs (e.g., veterans + senior exemptions). While not as widely used as `EXCD1`, it plays an important role in understanding the full exemption profile of a parcel.

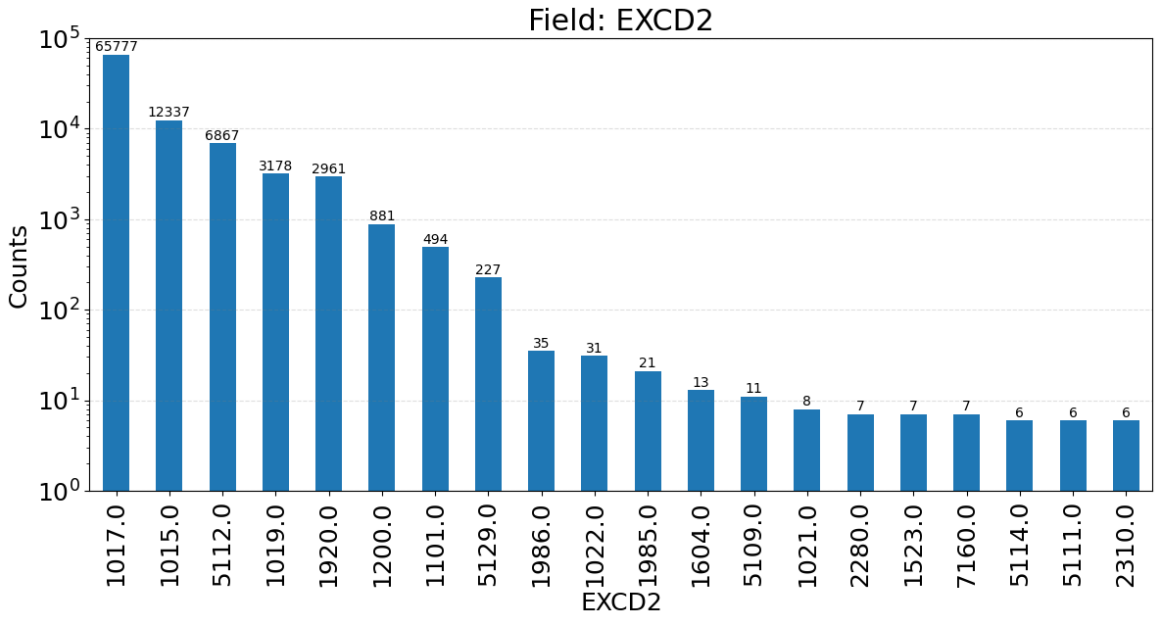

Figure 13: Top 20 EXCD2 (Secondary Exemption Codes)

PERIOD

`PERIOD` designates the type of tax roll (e.g., "FINAL" or "TENTATIVE"). In this dataset, every record has the value "FINAL", since this dataset exclusively captures the 2010/11 final tax assessment roll. As such, it is a constant field with no variance. While not analytically valuable on its own in this case, it would be critical if data from multiple assessment periods were combined.

YEAR

`YEAR` captures the fiscal year of the assessment roll. Like `PERIOD`, it is constant across all records ("2010/11"). It reaffirms the temporal scope of the data and is vital when aggregating data across multiple years. In this standalone dataset, however, it adds no variance or analytical value.

VALTYPE

`VALTYPE` stands for Valuation Type and indicates which kinds of assessed values are available. Every record contains the value "AC-TR", meaning both Actual and Transitional values are included. This field would be useful for cross-year or comparative analyses, but is redundant here as there is no variation.

LTFRONT

`LTFRONT` represents the width of the property lot in feet. It is a numeric field, fully populated, but approximately 15.8% of values are zero. A zero value may indicate either a missing or undefined dimension—common in cases such as condominium units or irregular lots. The most frequent non-zero values cluster around 20–25 feet, typical of NYC rowhouses. Some values are unusually high (up to 9,999), suggesting outliers or placeholder codes.

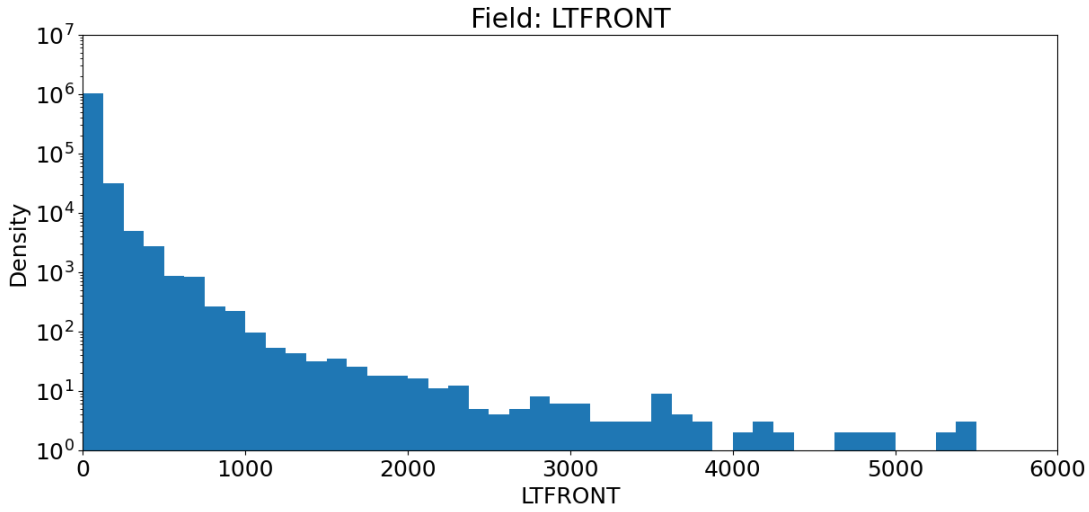

Figure 14: Histogram of LTFRONT (Lot Width) - excluding outliers beyond 6,000 feet

LTDEPTH

`LTDEPTH` measures the depth of the property lot in feet. Similar to `LTFRONT`, it is fully populated, with around 15.9% of entries equal to zero. Typical values for residential lots fall between 80–100 feet. High outlier values (e.g., 9,999) likely indicate large or irregular parcels or use as a data placeholder. Despite these quirks, the distribution for most properties is consistent with NYC planning norms.

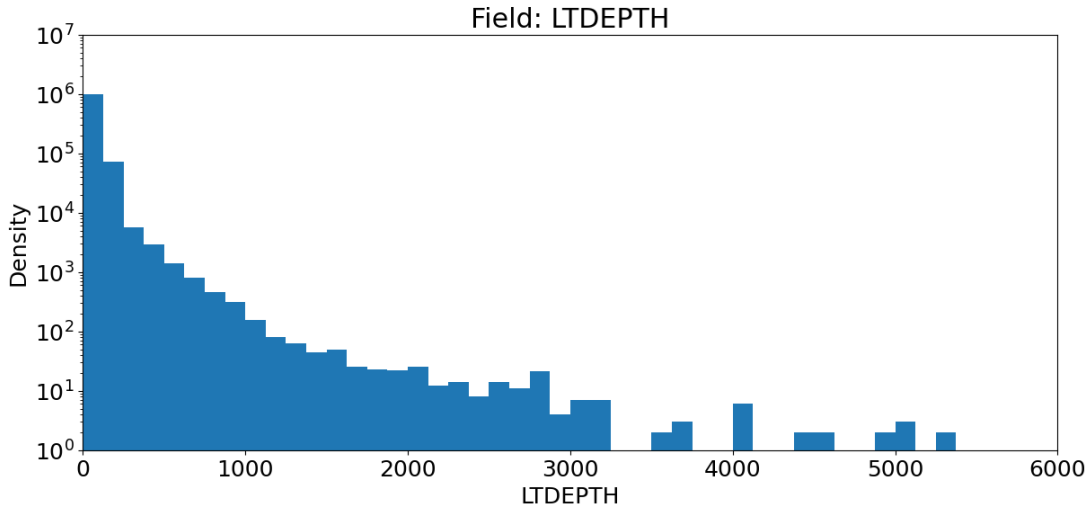

Figure 15: Distribution of LTDEPTH (Lot Depth) - excluding outliers beyond 6,000 feet

STORIES

`STORIES` captures the number of above-ground stories in a building. About 5.25% of records are missing. The most frequent values are whole numbers such as 1, 2, and 3, which represent one- to three-story buildings common in residential neighborhoods. Values like 1.5 and 2.5 appear frequently for homes with attics or partial floors. A small number of extreme values (e.g., 119) represent skyscrapers or data entry anomalies. A few fields include strange decimal values (e.g., 1.7, 3.3) which are likely erroneous.

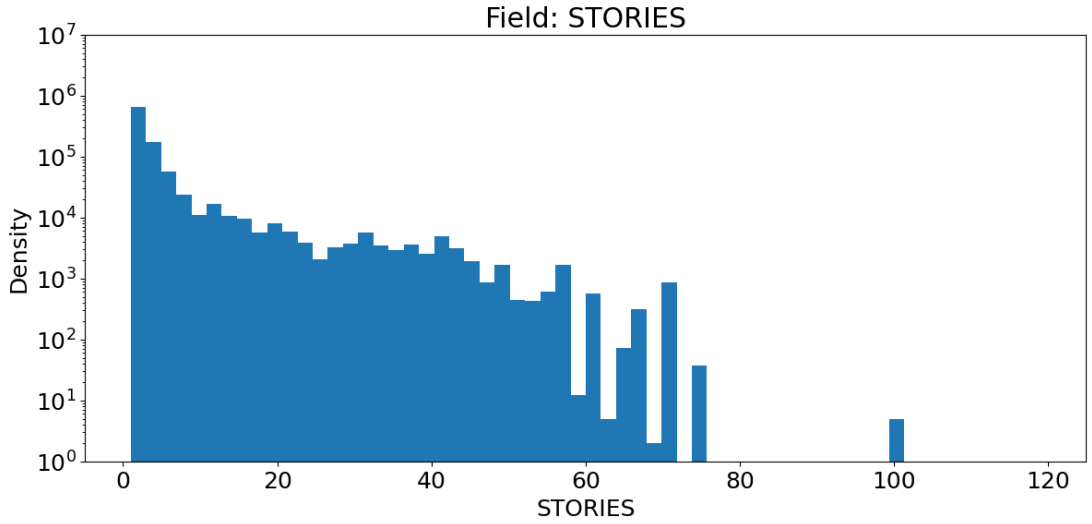

Figure 16: Distribution of STORIES (Building Height)

FULLVAL

`FULLVAL` is the full market value of the property, expressed in dollars. It is fully populated, with approximately 1.2% of records having a value of zero. The distribution is highly right-skewed; most properties are valued under $1 million, but some (e.g., large commercial properties) exceed $6 billion. The average value is about $874k. A log transformation reveals a near-normal distribution, suggesting a log-normal nature typical of real estate data.

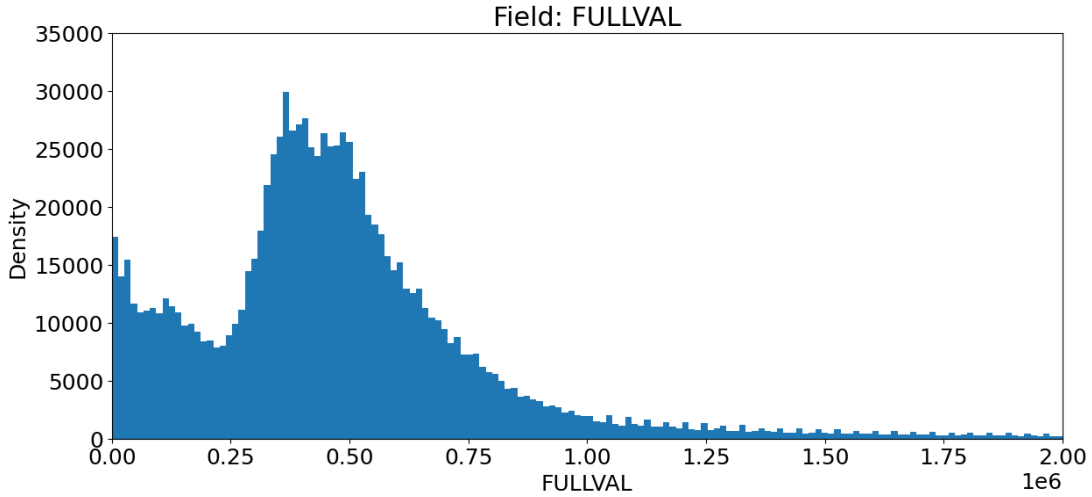

Figure 17: Distribution of FULLVAL (Full Market Value) - excluding values beyong $2 million as majority of the properties lie in this range

AVLAND

`AVLAND` is the assessed value of the land portion of the property, excluding improvements. This field is populated for all records. Values mirror the distribution of `FULLVAL` but are consistently lower due to assessment ratios and exemptions. Around 13k entries are zero. The average land assessment is $85k, and the maximum exceeds $2.5 billion. For many Class 1 residential properties, land is the dominant component of value.

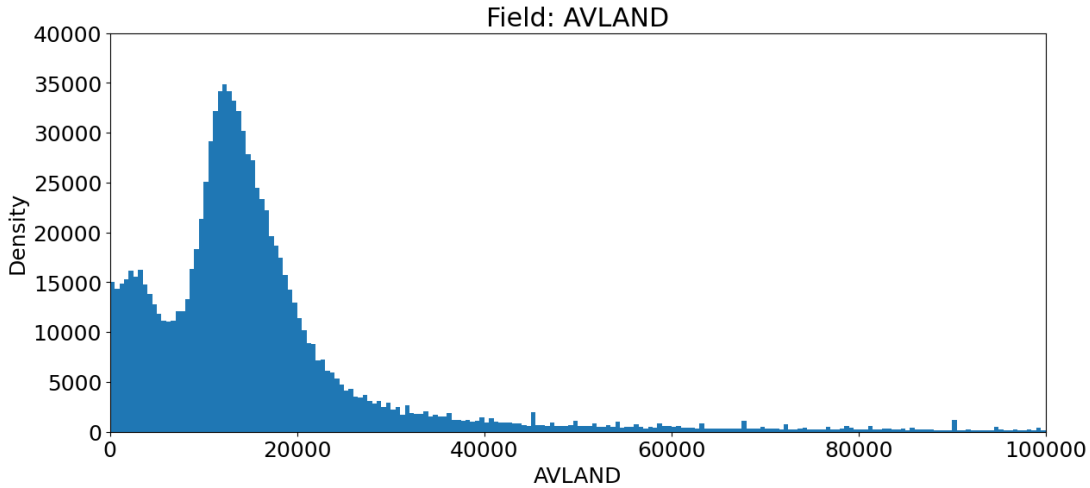

Figure 18: Distribution of AVLAND (Assessed Land Value) - excluding values beyong $100k as majority of the properties lie in this range

AVTOT

`AVTOT` is the total assessed value of the property (land + improvements). It is fully populated and, in most cases, greater than or equal to `AVLAND`. The average assessed total is about $227k. Like `FULLVAL`, this field is heavily skewed due to the wide range of property types in NYC. A small number of properties show an assessed value of zero, typically those that are fully tax-exempt or vacant.

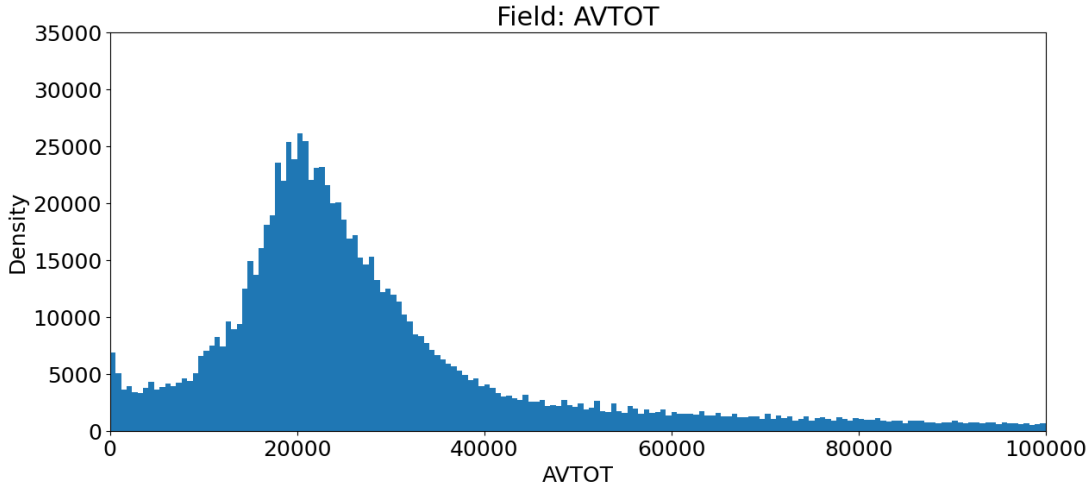

Figure 19: Distribution of AVTOT (Total Assessed Value) - excluding values beyong $100k as majority of the properties lie in this range

EXLAND

`EXLAND` is the exempt portion of the land's assessed value. This field is numeric and fully populated, though 45.9% of entries are zero. A non-zero value indicates partial exemption from property taxes on land value. Values range up to nearly $2 billion. For properties eligible for tax exemptions (e.g., veterans, senior citizens, religious organizations), this field reflects the portion of land value excluded from taxation.

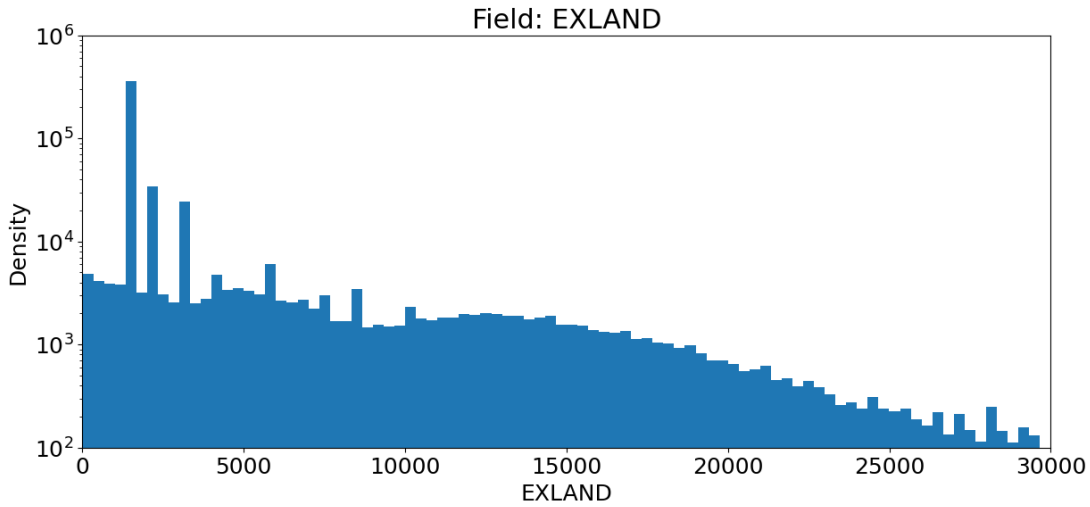

Figure 20: Distribution of EXLAND (Exempt Land Value, Log Scale) - excluding values beyond $40k

EXTOT

`EXTOT` is the exempt portion of the total assessed value (land + improvements). About 40.4% of records show a zero value, while others reflect varying degrees of tax relief. Values span a broad range and can exceed $2 billion. Like `EXLAND`, this field is essential in calculating net taxable value for a property.

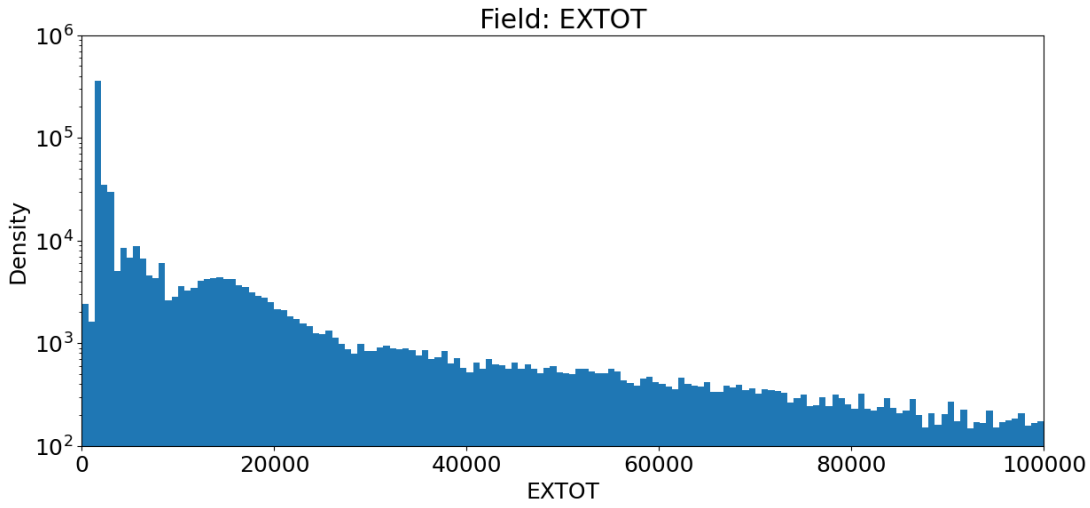

Figure 21: Distribution of EXTOT (Total Exempt Value, Log Scale) - excluding values beyond $100k

BLDFRONT

`BLDFRONT` indicates the width of the building in feet. It is fully populated, but over 21% of values are zero, likely due to properties like condos or vacant land that do not independently list a building footprint. For non-zero entries, typical widths range from 20 to 50 feet, reflecting standard townhouse and apartment widths. Some outliers (e.g., 3,000 feet) may be errors or placeholders.

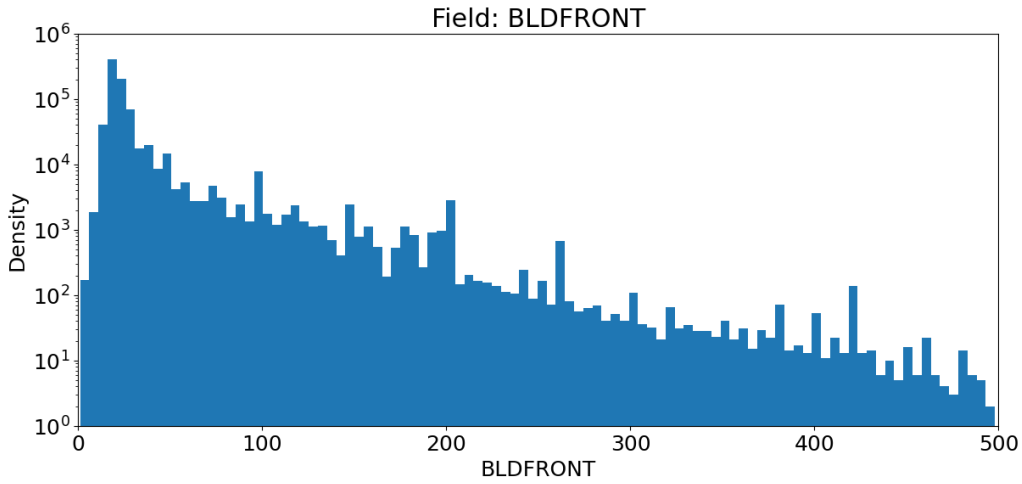

Figure 22: Distribution of BLDFRONT (Building Frontage) - excluding outliers beyond 500 feet

BLDDEPTH

`BLDDEPTH` measures the depth of a building in feet. As with `BLDFRONT`, many records have a value of zero (over 21%), especially for properties without individual buildings (e.g., condo units). Common values cluster between 30 and 100 feet. Outliers (up to 3,306 feet) may indicate industrial properties or irregular lots.

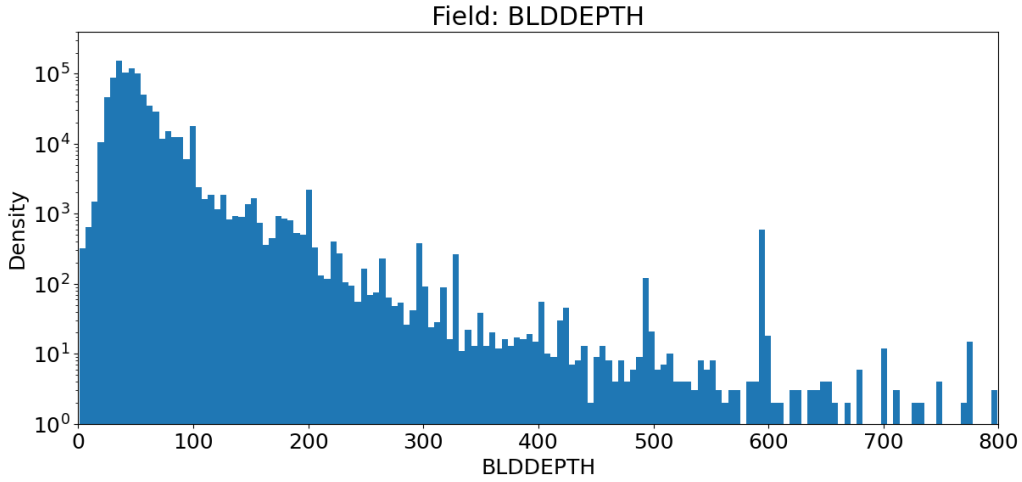

Figure 23: Distribution of BLDDEPTH (Building Depth) - excluding values beyond 800 feet

AVLAND2

`AVLAND2` is the transitional assessed land value, used for phasing in tax increases. It is populated for about 26.4% of records, primarily in Classes 2 and 4. These values often differ from `AVLAND` due to assessment caps and smoothing. Typical values mirror actual assessed land but are adjusted downward for transitional relief. The most common transitional land assessment is $2,408.

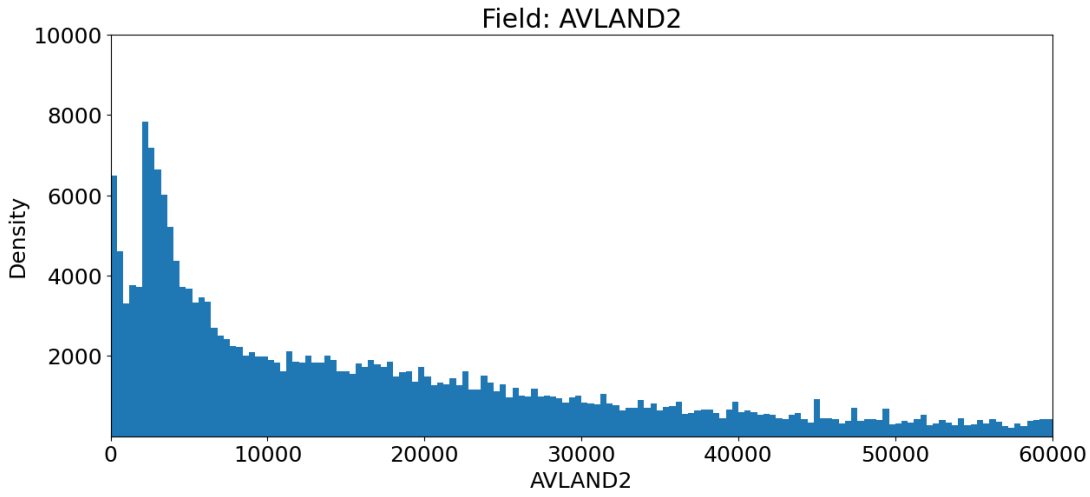

Figure 24: Distribution of AVLAND2 (Transitional Land Value) - excluding values beyond $60k

AVTOT2

`AVTOT2` is the transitional total assessed value. Like `AVLAND2`, it is used to phase in tax changes gradually. It is present in the same 26% of records. The values are generally less than or equal to `AVTOT`, although some records show equal values, indicating no transitional adjustment. The most common value is $750, and the mean is significantly higher due to high-value commercial properties.

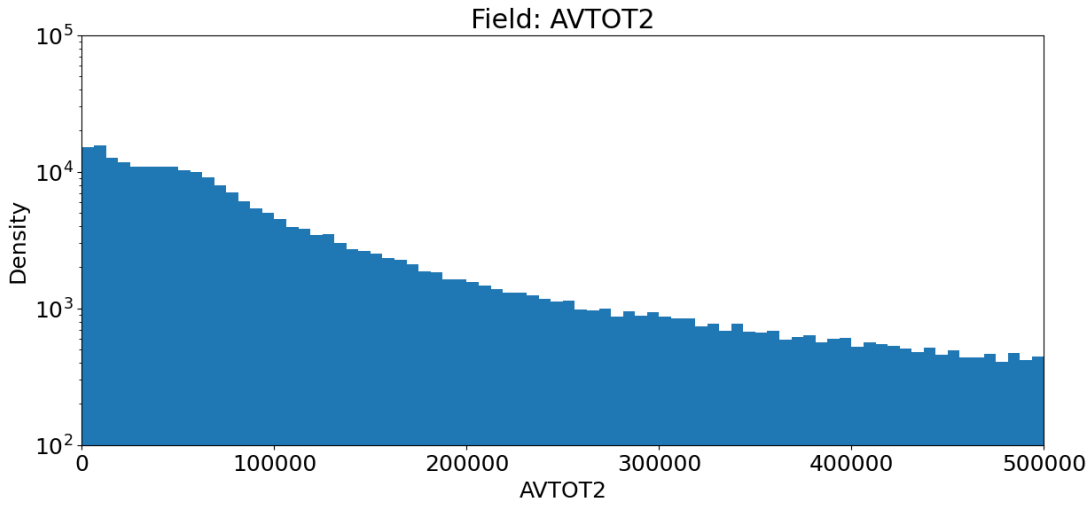

Figure 25: Distribution of AVTOT2 (Transitional Total Value, Log Scale) - excluding values beyond $500k

EXLAND2

`EXLAND2` is the exempt portion of transitional land value. Only 8.2% of records are populated, which corresponds to properties both exempt and under transitional assessment rules. The most frequent value is $2,090, aligning with standard deductions (e.g., STAR). Distribution is sparse but consistent with exemption policy.

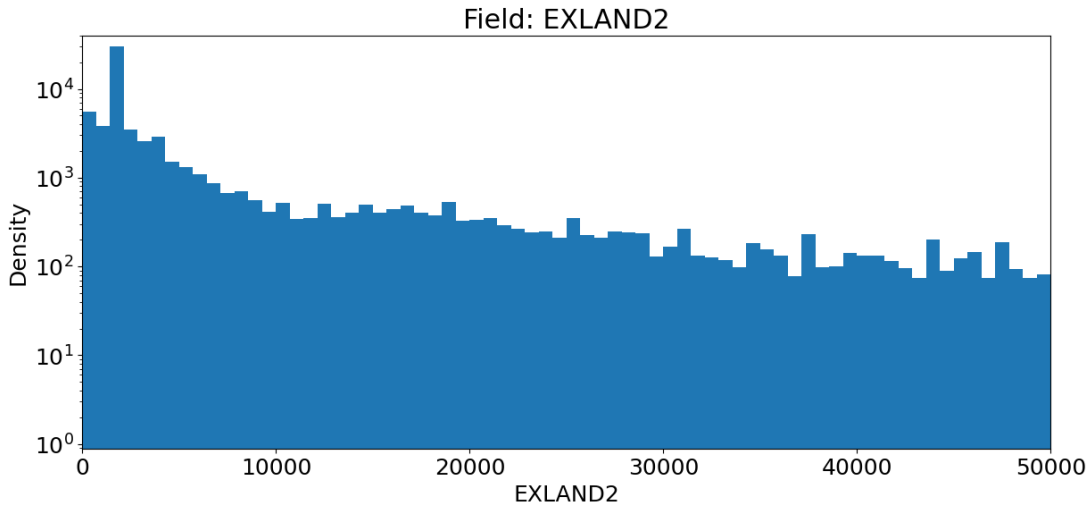

Figure 26: Distribution of EXLAND2 (Transitional Exempt Land Value, Log Scale) - excluding values beyond $50k

EXTOT2

`EXTOT2` reflects the total transitional exempt value. Populated in about 12.2% of records, it corresponds with exemption codes and class-specific transitions. The values range widely, with a common value again at $2,090. This field provides insight into phase-in relief granted to owners with exemption status.

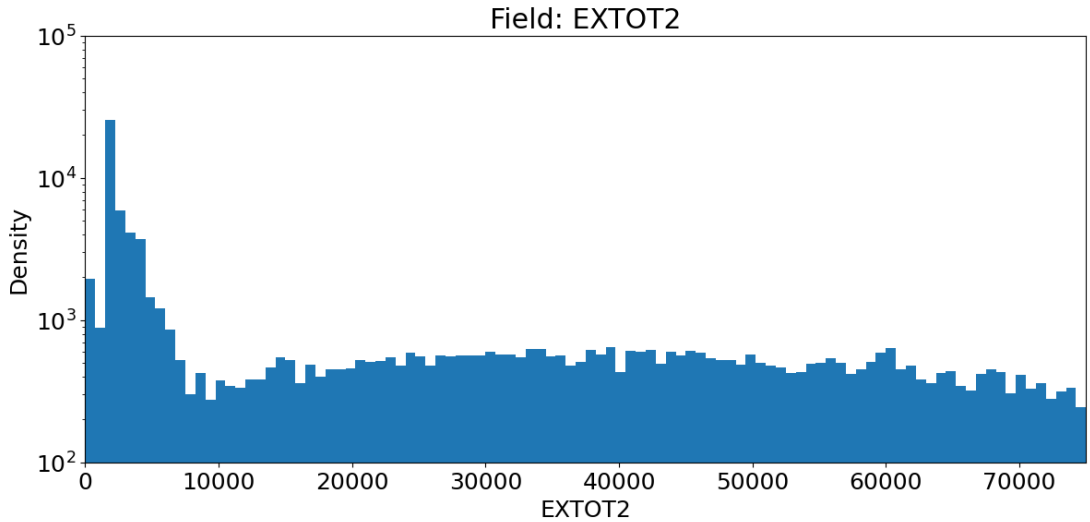

Figure 27: Distribution of EXTOT2 (Transitional Total Exempt Value, Log Scale) - excluding outliers beyond $80k