

A linked dataset of New York City residential sales, recorded deeds and mortgages, 2016–2025

Data note

Documentation version 1.0.3 · 14 October 2026

Pablo Loschi Independent researcher. ORCID 0009-0004-9455-4713. Corresponding author: Pablo Loschi (loschi.pablo@gmail.com).

Abstract

New York City publishes its property sales and its recorded property documents as separate open datasets, and the two do not join directly. The Department of Finance sales file gives a price and a date but says nothing about the deed or how the purchase was paid for. The city’s register, ACRIS, holds recorded documents and their amounts, rather than a directly corresponding transaction table. For condominiums the parcel identifier in the sales file usually names the building’s billing lot, while ACRIS records the deed and mortgage against the unit’s own lot. This dataset links the two sources once, with fixed and documented rules. Its sales table holds all 729,047 sales of one- to three-family houses, condominium units and co-operative apartments recorded in 2016–2025. Linking is attempted for the 315,530 house and condominium sales with a positive price in the four boroughs ACRIS covers. A unique deed is found for 97.2% of those house sales and 74.0% of those condominium sales; condominium coverage ranges from 78.9% in Manhattan to 44.7% in the Bronx. For houses and uniquely linked condominiums, the dataset records whether a mortgage was filed against the parcel in three windows around the sale, and its amount. A recorded mortgage near the sale date is not verified purchase finance. Each linked deed carries flags for company buyers, shared surnames, shared mailing addresses and lender take-title events. No party names are published. The links reproduce the archived financing flags and mortgage-recording-tax statistics of an earlier study, which shows that the computations agree, not that the links are accurate. The raw extracts and code are deposited, and the sales and link tables can be rebuilt offline from them. The deed flags were computed from party records that are not deposited and are provided as a fixed snapshot.

Keywords: New York City; property transactions; ACRIS; mortgages; condominiums; record linkage; transfer taxes; open government data

Specifications table

Subject	Economics, Econometrics and Finance
Specific subject area	Urban and housing economics: residential property transactions, purchase finance and deed types in New York City
Type of data	Tables (gzip-compressed CSV); codebook (Markdown); raw extracts with retrieval manifests (CSV, JSON); Python code
Data collection	Three public datasets were retrieved from NYC Open Data through its Socrata API on 26 September 2026: the Department of Finance annualized sales file and the ACRIS Real Property Master and Legals files. Each retrieval was checked against the live row count. Sales were linked to deeds by parcel, amount and date, and to mortgages by parcel and date window. Party records for the linked deeds were retrieved from ACRIS Parties, used to compute flags and not deposited.
Data source location	New York City (Manhattan, Bronx, Brooklyn, Queens and Staten Island), New York, United States. Primary sources: NYC Department of Finance; NYC Open Data.
Data accessibility	Repository name: Zenodo. Data identification number: 10.5281/zenodo.22977532 (concept DOI; version 1.0.1 is 10.5281/zenodo.23047979). Direct URL to data: https://doi.org/10.5281/zenodo.22977532 [10]. Licence: CC BY 4.0 for the code, codebook and derived tables; the source records keep the City of New York's open-data terms.
Related research article	None published in a journal. The dataset supports three preprints by the author: Loschi [5], [6], [7].

Value of the data

- It links house and condominium sales of 2016–2025 in the four ACRIS boroughs to their deeds wherever exactly one deed qualifies: 183,912 of 189,279 eligible house sales and 93,386 of 126,251 eligible condominium sales. For each linked sale, it records whether a mortgage was filed against the parcel near the sale date. Neither source provides this link.
- It matches condominium sales among the unit lots of their block, so that the deed and the mortgage are sought on the unit's own lot. A join on the sales file's parcel identifier, which usually carries the building's billing lot, misses most condominium deeds and loans. In 99.3% of unique matches the recovered unit equals the sales file's separate lot field. That is a consistency check, not a measure of match accuracy.
- Researchers can compare sales with and without a recorded mortgage near the sale date, measure recorded loan-to-price ratios, and study schedule-implied mortgage recording tax or transfer-tax notches without rebuilding the link. A recorded mortgage is not verified purchase finance, no match is not a verified cash purchase, and the tax actually paid is not observed. Results for condominiums refer to the linked subset and should be reported with its coverage (Tables 4 and 5).
- Flags for company buyers, same-surname transfers, shared addresses and lender take-title deeds let users screen out transfers that may not be arm's-length sales. Party names are not released.
- The raw extracts, their retrieval manifests and the code are deposited together. The sales and link tables can be rebuilt offline from the deposit or refreshed from the live sources. The deed flags cannot be rebuilt offline: they were computed from ACRIS party records that were fetched into a private cache and not deposited.

Background

The Department of Finance publishes every recorded property sale with its price, date, building class and parcel identifier (borough, block and lot, or BBL) [1]. The Automated City Register Information System (ACRIS) publishes every recorded deed, mortgage and related document for Manhattan, the Bronx, Brooklyn and Queens, with separate files for document headers, parcels and parties [2], [3], [4]. Staten Island documents are kept by the Richmond County Clerk and are not in ACRIS.

Many questions about housing need both sources. A sale price is more informative once one knows whether the buyer borrowed, whether the deed was a referee's deed after a foreclosure, or whether buyer and seller share a surname. Joining the two sources is not straightforward. The sales file has no document identifier.

Deed amounts and dates differ slightly from the sale record. For condominiums, the sales file’s BBL usually gives the building’s billing lot (7501 and above), while the deed and the mortgage are recorded against the unit lot (1001–1999). A join on BBL therefore misses most condominium documents. The file’s separate lot field usually carries the unit lot, but this dataset does not rely on it: the unit is taken from the matched deed.

The author built this link for three studies of New York housing: the gap between cash and mortgage prices [5], mortgage recording taxes and the 2019 transfer-tax thresholds [6], and foreclosure transfers in house-price indexes [7]. Each study had its own version of the join. This dataset replaces them with one set of rules, applied to all residential sales, and publishes the result so the studies and other users work from the same records.

Record-linkage research distinguishes coverage, representativeness and errors in accepted or missed links [8]. The historical error rates in that literature are not estimates for these property records. The coverage comparisons below describe sample selection; agreement with earlier files checks reproducibility. Independent source evidence is still needed to measure linkage accuracy.

Data description

The deposit contains three linked tables, a codebook, validation outputs, the raw extracts, the code and worked examples (Table 1). The tables share keys: sales joins to links on `sale_id`, and links joins to deed_flags on `deed_document_id`. A blank cell means not applicable or not measured. It never means zero. `CODEBOOK.md` defines every column.

Table 1. Files in the deposit.

File	Rows	Contents
<code>build/sales.csv.gz</code>	729,047	One row per Department of Finance sale, with location, building class, price, date and geography fields
<code>build/links.csv.gz</code>	729,047	Deed match, matched deed, financing parcel, financing flags in three windows, number and amount of mortgages, CEMA flag
<code>build/deed_flags.csv.gz</code>	276,255	One row per linked deed: party counts and screening flags
<code>build/validation.json</code>	—	Agreement with the archived data of two earlier studies
<code>build/link_stats.json</code> , <code>build/party_meta.json</code>	—	Match counts and deed-event counts
<code>raw/</code>	—	The three source extracts as retrieved, each with a manifest giving the query, the row count and the live count
<code>code/</code>	—	Five Python scripts (standard library only) and the classification rules of the earlier studies, copied unchanged
<code>examples/</code>	—	A worked example, the condominium selection diagnostics and a test (standard library only; added in version 1.0.1)
<code>CODEBOOK.md</code> , <code>README.md</code> , <code>CITATION.cff</code>	—	Documentation and citation metadata

Sales. The sales table covers all sales dated 2016–2025 in building-class categories 01–03 (one- to three-family houses), 04, 12, 13 and 15 (condominiums), and 09, 10 and 17 (co-operatives). Table 2 gives the counts by borough. All prices are kept, including the 213,457 sales recorded at \$0, whose recorded prices are not market prices. Users choose their own price screen. The flag `in_tax_price_band` marks the \$300,000–\$6,000,000 band used in [6].

Table 2. Sales by property type and borough, 2016–2025.

	Manhattan	Bronx	Brooklyn	Queens	Staten Island	Total
Houses (1–3 family)	3,520	41,171	117,893	158,355	67,852	388,791
Condominiums	77,800	5,818	49,245	29,723	7,301	169,887
Co-operatives	80,767	12,058	27,929	48,539	1,076	170,369
Total	162,087	59,047	195,067	236,617	76,229	729,047

Links. Linking is attempted for the 315,530 house and condominium sales with a positive price in the four ACRIS boroughs. Co-operative unit sales transfer shares and a proprietary lease. Share loans are not real-property mortgages on the purchased apartment. ACRIS can record building-level mortgages and separate personal-property filings, which this linkage does not assign to individual share purchases. Table 3 reports the outcomes. Of the 315,530 sales where linking is attempted, 277,298 find a unique deed; the other 38,232 are unlinked. Almost every house sale finds a unique deed. About three in four condominium sales do. The rest either find no deed within the tolerances or find qualifying deeds on more than one unit, in which case they are left unlinked rather than guessed. Sales with a nonpositive price, co-operative sales and Staten Island sales are in the sales table but outside the linkage.

Table 3. Deed matches and recorded mortgages among sales where linking was attempted.

	Houses	Condominiums
Sales where linking was attempted	189,279	126,251
Unique deed match	183,912 (97.2%)	93,386 (74.0%)
No qualifying deed	5,367	22,030
Qualifying deeds on more than one unit (left unlinked)	—	10,835
Sales with financing flags	189,279	93,386
Mortgage recorded, strict window [–15, +30] days	134,283 (70.9%)	51,455 (55.1%)
Mortgage recorded, base window [–15, +90] days	135,741 (71.7%)	52,335 (56.0%)
Mortgage recorded, wide window [–15, +180] days	138,778 (73.3%)	53,119 (56.9%)

Notes: Houses carry financing flags whether or not a deed matched, because the mortgage is sought on the sale’s own parcel. Condominiums carry them only with a unique deed, because the deed supplies the unit lot. Percentages of financed sales are of sales with financing flags.

For the 188,076 sales with a mortgage in the base window, mortgage_amount_base gives the largest recorded mortgage and mortgage_document_id_base its ACRIS identifier. Recorded amounts include errors and loans that also cover other property: 6,822 exceed 105% of the price and 406 are 5% of the price or less. The earlier studies keep loan-to-price ratios between 0.05 and 1.05 when they use amounts. A purchase financed by assigning and consolidating the seller’s loan (a CEMA) is flagged separately. There are 108 such sales.

Condominium coverage. Linked condominium sales are a selected subset. The 126,251 condominium sales where linking is attempted are those with a positive price in the four ACRIS boroughs; the other 43,636 condominium sales in the sales table are on Staten Island (7,301) or have a nonpositive price (36,335). Table 4 compares the three match outcomes, and Table 5 gives coverage by borough and price. Median recorded prices are similar: \$1,008,329 for uniquely linked sales and \$950,000 for unlinked ones. Location is not. Half of the linked sales are in Manhattan, against 38% of the unlinked. Unique-deed coverage is 78.9% in Manhattan and 44.7% in the Bronx, where about half of eligible condominium sales find no qualifying deed. By price band it is highest at \$1 million to \$2 million (78.7%) and lowest at \$6 million or more (60.9%). By year it varies less, between 72.0% and 76.3%. The standardized difference in mean log price between linked and unlinked sales is +0.018, which hides these differences by borough and price. The table describes who is linked. It does not show why a match is missing, and it does not supply weights that would make the linked subset represent

all condominium sales. Users who estimate condominium shares should report the linked denominator and, where relevant, coverage for the borough and price range they study.

Table 4. Condominium sales where linking was attempted, by deed-match outcome, 2016–2025.

	Unique deed	No qualifying deed	More than one unit	All unlinked
Sales	93,386	22,030	10,835	32,865
Share of eligible sales	74.0%	17.4%	8.6%	26.0%
Median recorded price	\$1,008,329	\$926,662	\$973,500	\$950,000
Recorded price $\geq$ \$2 million	23.4%	28.2%	18.1%	24.9%
In Manhattan	50.4%	37.7%	39.5%	38.3%
Financing flags	measured	not measured	not measured	not measured

Notes: Eligible sales are condominium sales with a positive recorded price in Manhattan, the Bronx, Brooklyn and Queens (126,251). Recorded prices are as filed and include nominal amounts; 2,242 eligible sales are recorded below \$100,000. Condominiums carry financing flags only with a unique deed, because the deed supplies the unit lot, so financing is unknown, not zero, for the 32,865 unlinked sales. Source: `condo_selection.py` on the deposited `build/` files.

Table 5. Unique-deed coverage of eligible condominium sales, by borough and recorded price, 2016–2025.

	Eligible sales	Unique deed	Coverage
<i>By borough</i>			
Manhattan	59,647	47,061	78.9%
Bronx	4,311	1,929	44.7%
Brooklyn	39,514	29,496	74.6%
Queens	22,779	14,900	65.4%
<i>By recorded price</i>			
Under \$100,000	2,242	1,508	67.3%
\$100,000–\$499,999	17,077	11,516	67.4%
\$500,000–\$999,999	45,260	33,594	74.2%
\$1–2 million	31,594	24,874	78.7%
\$2–6 million	23,232	17,722	76.3%
\$6 million or more	6,846	4,172	60.9%

Figure 1 shows, by year, the share of sales with a mortgage recorded in the base window. For houses the denominator is every sale where linking was attempted; for condominiums it is sales with a unique deed. The share peaked in 2021 for both houses (78%) and condominiums (66%) and fell in 2023, after mortgage rates rose in 2022. The fall is steeper for condominiums, which ended the period below 47%. The condominium shares describe linked sales only.

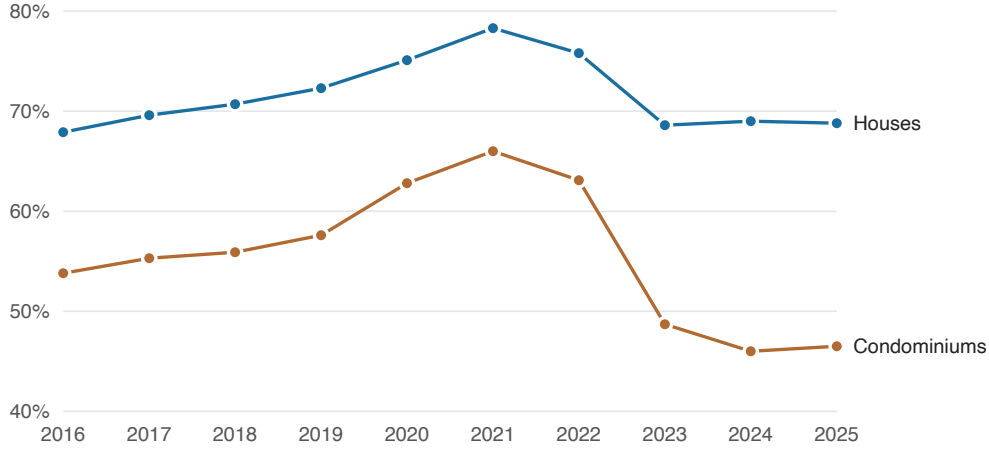


Figure 1. Share of linked sales with a mortgage recorded within the base window ($[-15, +90]$ days), by year of sale. Houses: all sales where linking was attempted. Condominiums: sales with a unique deed match.

Deed flags. Each of the 276,255 linked deeds carries flags computed from its grantor and grantee records in ACRIS. The flags mark a company buyer (62,563 deeds), a trust or estate buyer (7,976), a natural-person grantor and grantee who share a surname (10,790), a grantor and grantee with the same normalized mailing address (42,989), a party name that matches a list of identifiable lenders (5,226), and a deed event from the classifier of [7]. That classifier labels 4,003 deeds as events whose recorded amount can follow a legal rule rather than a negotiated price: a referee’s deed to the foreclosing lender (1,038), another conveyance to a lender such as a deed in lieu of foreclosure (2,787), or a transfer between lenders (178). The flags are screens built from names. None has a measured error rate.

Experimental design, materials and methods

Retrieval. `code/01_fetch.py` retrieved three NYC Open Data resources through the Socrata API on 26 September 2026: the Department of Finance annualized sales file (w2pb-icbu; 729,047 rows) [1], the ACRIS Real Property Master file restricted to deeds (DEED), mortgages (MTGE) and consolidation agreements (M&CON) recorded in 2015–2026 (bnx9-e6tj; 1,394,960 rows) [2], and the ACRIS Real Property Legals file for documents numbered 2015–2026 (8h5j-fqxa; 5,500,768 rows) [3]. Each extract was paged in a fixed key order and its row count was compared with the live count returned by the API. The counts matched for all three. The query, the counts and the retrieval time are stored in a manifest beside each extract.

Documents. Each ACRIS document is attributed to the parcels in its Legals records. A document that names more than one parcel cannot be attributed to a single sale and is dropped. This removes 255,975 documents. A document’s date is its document date, or its recording date where the document date is missing.

Deed rule. `code/02_link.py` matches a sale to a deed when the deed amount is within 0.5% of the sale price and the deed date is within 45 days of the sale date. House sales are matched on their own BBL. Condominium sales are matched on borough and block among unit lots 1001–1999, which recovers the unit lot even when the sales file gives only the billing lot. The recovered unit equals the sales file’s own `lot` field for 92,697 of the 93,386 unique condominium matches (99.3%). The remaining 689 may be parking or storage units, renumbered lots or wrong matches. The match must be unique. If deeds on two different parcels qualify, the sale is marked `ambiguous` and left unlinked. If one parcel has several qualifying deeds, such as an original and a correction deed, the lowest document identifier is kept. The sales file sometimes lists one transfer more than once, so 878 deeds are linked to two or more sales (1,921 sales in all).

Financing rule. A sale is financed in a window if a mortgage is recorded against its financing parcel with a date from 15 days before to 30 (strict), 90 (base) or 180 (wide) days after the sale. The financing parcel is the sale’s BBL for houses and the matched unit lot for condominiums. These are the rules used in [6]. The deed and financing rules are deterministic. The dataset does not use probabilistic record linkage of the kind described by Fellegi and Sunter [8] or Christen [9], because shared parcel identifiers make a rule-based match feasible and easy to audit.

Deed flags. `code/03_parties.py` retrieves the party records for each linked deed from ACRIS Parties (636b–3b5g) [4] into a private cache, computes the flags and writes `deed_flags.csv.gz`. The name rules are those of [5] and [7], copied unchanged into `code/vendor/`. A company is a name with a company form (LLC, INC, CORP, REALTY, HOLDINGS and similar). A same-surname pair is a natural-person grantor and grantee with the same last name; a strict version ignores the 40 most frequent surnames among the linked deeds. A shared address is a match on normalized street, city and five-digit ZIP code after post-office boxes and care-of addresses are dropped. The cache holds names and addresses and is not deposited. The deposited `deed_flags.csv.gz` is therefore a fixed snapshot. The same script can fetch party records again from the live source, but ACRIS is updated continually, so a new retrieval is a new computation and may not reproduce every flag. The list of frequent surnames used by the strict same-surname flag is derived from names and is also withheld.

Agreement with earlier studies. `code/04_validate.py` and `code/05_mrt_check.py` compare the dataset with archived data from two of the earlier studies (Table 6). The archived purchase panel of [6] contains 249,395 purchases in the four ACRIS boroughs. All are found here, and 249,084 can be compared one to one. The other 311 share a parcel, date and price with another sale. The financing flags agree for every compared house and for all but 4 of 82,348 compared condominiums, in all three windows. Mortgage amounts agree for all 128,412 financed houses. Recomputing the mortgage-recording-tax table of [6] from this dataset reproduces every reported statistic to the published precision, including the condominium figures, whose loan amounts were not in that study’s original deposit. Since its version 2.5, that study takes its condominium loan amounts from this dataset, so the condominium row is no longer an independent comparison. Against the repeat-sale pairs of [5], the matched deed is the same at 15,902 endpoints and different at one. The other study links the nearest deed by date without requiring the amount to match, so it links 224 deeds that this dataset leaves unmatched.

Table 6. Agreement with the archived data of earlier studies.

Comparison	Compared	Agree
Financing flags, houses, all three windows [6]	166,736	166,736
Financing flags, condominiums, each window [6]	82,348	82,344
Mortgage amount, financed houses [6]	128,412	128,412
Deed at repeat-sale endpoints where both link a deed [5]	15,903	15,902
Base financing flag at repeat-sale endpoints [5]	16,398	16,398
Company-buyer and same-surname flags on commonly linked deeds [5]	15,910	15,910

Notes: Mortgage-recording-tax statistics of [6], rebuilt from this dataset: houses, 124,952 loans (paper: 125,023), mean loan-to-value 0.745, mean borrower-side tax 1.41% of price; condominiums, 47,704 loans (paper: 47,610), mean loan-to-value 0.678, mean tax 1.28% of price. All means and percentiles match that earlier comparison at its reported precision. The counts in this row refer to the historical validation output, not to the later v2.6 tax calculation, which uses 125,023 house loans and 47,702 condominium loans in its own filtered panel. Reusing this dataset for that later calculation is not independent validation.

These comparisons measure computational agreement. The released rules reproduce outputs of earlier code that was applied to overlapping sources with related rules, so agreement shows that the computation is consistent, not that its links are correct. Accuracy would require a verified record, for the same sale, of the deed, the parties, the collateral and the purpose of any loan. The release contains no such record and reports no overall accuracy rate. The same applies to the 99.3% agreement between the recovered condominium unit and the sales file’s `lot` field. Coverage, the share of eligible sales that find a unique deed (Tables 3–5), is a third and separate quantity.

Worked example. `examples/recorded_mortgage_example.py` computes an annual table of condominium linkage and recorded mortgages from `build/sales.csv.gz` and `build/links.csv.gz`. It uses only the Python standard library, needs no network access or party records, and runs in a few seconds:

```
python examples/recorded_mortgage_example.py build --output annual_condo_coverage.csv
```

For each year it reports eligible condominium sales, their deed-match outcomes, recorded-mortgage matches among uniquely linked sales and the number of sales whose financing is not measured. It stops with an error if an eligible sale is missing from the links table, if a join key is duplicated, or if an unlinked condominium carries a financing flag, and it never converts a blank flag to zero. In 2021, 12,760 of 17,734 eligible condominium sales link to a unique deed (72.0%). Of those 12,760, 8,425 have a mortgage recorded in the base window (66.0%). Financing is not measured for the other 4,974 eligible sales, so 66.0% is a share of linked sales, not of all condominium sales. In 2025 the figures are 9,530 of 12,497 linked (76.3%) and 4,428 mortgage matches (46.5% of linked). A match means an MTGE on the unit lot dated from 15 days before to 90 days after the sale; it does not establish that the loan financed the purchase. `examples/test_example.py` checks these rules on a small synthetic file.

Reproduction. The scripts need Python 3.10 or later and no other packages. Run in order, `02_link.py` rebuilds `sales.csv.gz`, `links.csv.gz` and `link_stats.json` offline from `raw/`, and the worked example and `condo_selection.py` run offline on those files. Three outputs cannot be rebuilt from the deposit alone. `deed_flags.csv.gz` and `party_meta.json` require ACRIS party records, which `03_parties.py` fetches from the live source. `validation.json` requires archived files from the replication packages of [5] and [6], which are deposited separately. A fresh retrieval of any source will differ slightly from the deposited extracts, because the Department of Finance revises sales and ACRIS adds late recordings.

Limitations

- A financed flag of 1 means that a mortgage was recorded against the financing parcel in the window. It does not show that the loan financed the purchase. A flag of 0 means that no mortgage was recorded in the window. It is not proof of a cash purchase. A loan recorded late, recorded against another parcel or recorded on a multi-parcel document is missed.
- Co-operative and Staten Island sales are included in the sales table but are outside the linkage. Co-operative sales convey shares, and share loans are not recorded as real-property mortgages; Staten Island deeds are recorded by the Richmond County Clerk, not in ACRIS.
- 22,030 eligible condominium sales find no qualifying deed and 10,835 find qualifying deeds on more than one unit, so condominium coverage (74.0%) is lower than house coverage (97.2%). Coverage varies from 78.9% in Manhattan to 44.7% in the Bronx and differs by price range (Tables 4 and 5). Financing is not measured for unlinked condominiums. Estimates from linked condominium sales describe that subset.
- The company, same-surname, shared-address, lender and deed-event flags are screens based on names. None has a measured error rate. A shared surname can join unrelated people and miss relatives with different surnames. A shared address includes common mailing agents.
- Coverage is the share of eligible records linked. Agreement is consistency with an earlier implementation using related source records. Accuracy would require checking whether the linked deed or mortgage is the correct instrument for the transaction. The release reports coverage and agreement with earlier code. It does not report the accuracy of deed links, financing flags or deed flags against verified records.
- Recorded amounts are as filed and include errors.
- The `sale_id` key is stable within this release only. Use `dof_row_id` or `deed_document_id` to join across releases.

Ethics statement

The dataset is built entirely from public records that the City of New York publishes as open data. It involves no human subjects research, no interaction with individuals and no data from social media. ACRIS party names and mailing addresses were used only to compute aggregate flags on each deed. They are not published in the dataset, and the list of frequent surnames used by one flag is withheld because it is derived from names. The published document identifiers lead back to public ACRIS records, which do name the parties; users who follow them should respect the City's terms of use and applicable privacy law. The source data remain subject to the City of New York's open-data terms of use.

CRediT author statement

Pablo Loschi: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review and editing.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Declaration of competing interests

The author founded HolaCasaClara, a free comparator of Argentine mortgage loans, which has no connection to this dataset or New York housing records. The author reports no other competing interests.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used Claude (Anthropic) and ChatGPT (OpenAI) to write and check code and to draft and edit text. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

- [1] NYC Department of Finance, Citywide Annualized Calendar Sales Update, NYC Open Data, dataset w2pb-icbu. <https://data.cityofnewyork.us/d/w2pb-icbu> (accessed 26 September 2026).
- [2] NYC Department of Finance, ACRIS – Real Property Master, NYC Open Data, dataset bnx9-e6tj. <https://data.cityofnewyork.us/d/bnx9-e6tj> (accessed 26 September 2026).
- [3] NYC Department of Finance, ACRIS – Real Property Legals, NYC Open Data, dataset 8h5j-fqxa. <https://data.cityofnewyork.us/d/8h5j-fqxa> (accessed 26 September 2026).
- [4] NYC Department of Finance, ACRIS – Real Property Parties, NYC Open Data, dataset 636b-3b5g. <https://data.cityofnewyork.us/d/636b-3b5g> (accessed 26 September 2026).
- [5] P. Loschi, Who Gets the House? Related-Party Transfers and the Cash–Mortgage Price Gap in New York City, Zenodo preprint, 2026. <https://doi.org/10.5281/zenodo.22421850>
- [6] P. Loschi, The Shape of the Tax: Transaction Shifting at New York City’s 2019 Transfer-Tax Notches and the Mortgage Recording Tax, Zenodo preprint, 2026. <https://doi.org/10.5281/zenodo.22925301>
- [7] P. Loschi, When a Deed Is Not a Market Sale: Foreclosure Transfers, Statutory Consideration, and Repeat-Sales House-Price Measurement, Zenodo preprint, 2026. <https://doi.org/10.5281/zenodo.22925383>
- [8] I.P. Fellegi, A.B. Sunter, A theory for record linkage, *J. Am. Stat. Assoc.* 64 (1969) 1183–1210. <https://doi.org/10.1080/01621459.1969.10501049>
- [9] P. Christen, *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*, Springer, Berlin, 2012. <https://doi.org/10.1007/978-3-642-31164-2>
- [10] P. Loschi, NYC residential sales linked to ACRIS deeds and mortgages, 2016–2025, version 1.0.1, Zenodo, 2026. <https://doi.org/10.5281/zenodo.22977532>
- [8] M. J. Bailey, C. Cole, M. Henderson and C. Massey, How Well Do Automated Linking Methods Perform? Lessons from US Historical Data, *Journal of Economic Literature* 58(4), 997–1044 (2020). <https://doi.org/10.1257/jel.20191526>