

UNLOCKING VALUE BY LINKING PROPERTY DATA AND UTILITY DATA

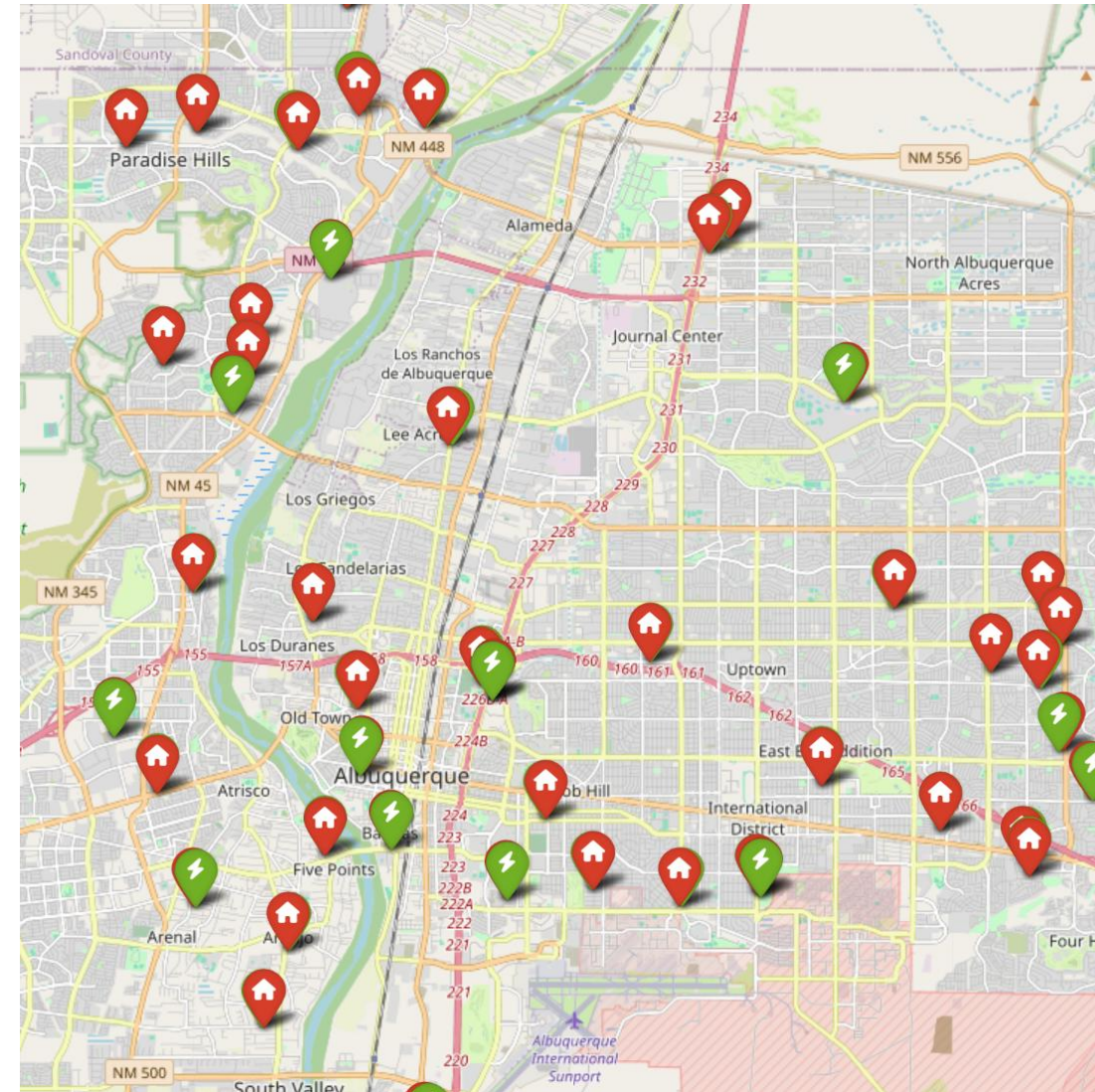
AEIC-WLRA CONFERENCE

APRIL 15, 2026



WHAT IS ADDRESS MATCHING?

- A process of linking datasets by the address record
- Link property data to premise level data (billing, meter, solar/battery usage data, etc.)
- Often paired with distance matching when location data (*lat, lon*) is available
- When location data is not available, geocoding (API) is used for address normalization and extracting a possible location (*lat, lon*)



ADDRESS MATCHING CAN PROVIDE A NEW DIMENSION TO UTILITY DATA

■ Why do property matching

- Targeting
- Energy burden and low-income customer identification
- Matching applications for evaluation
- Propensity scoring and granular forecasting

■ Projects

- *Central Hudson, PSEG LI, New Hampshire, PNM, PG&E, SCE, Pennsylvania (Commercial), Maryland (Residential)*

Property Data

- Variables can include:
 - ✓ Address
 - ✓ Latitude, Longitude
 - ✓ Square footage of building
 - ✓ Land area
 - ✓ Home value
 - ✓ Year built and year of major remodel
 - ✓ Estimated income
 - ✓ Presence of AC
 - ✓ Heating type
 - ✓ Pool info
 - ✓ Owner info
 - ✓ Mortgage info
 - ✓ HOA info
 - ✓ Other home characteristics

Created Features

- Targeting features :
 - ✓ Energy burden (electricity bill / estimated annual bill)
 - ✓ Whole home EUI
 - ✓ Cooling EUI
 - ✓ Heating EUI
 - ✓ Base load EUI
 - ✓ Energy burden percentile
 - ✓ EUI percentile
- Propensity features :
 - ✓ Annual gross consumption
 - ✓ Annual cooling kWh
 - ✓ Annual heating kWh
 - ✓ Annual base load
 - ✓ Annual cooling kWh per sqft
 - ✓ Annual heating kWh per sqft
 - ✓ Annual base load per sqft



EXAMPLE: PROPENSITY SCORING OF TECHNOLOGY ADOPTION

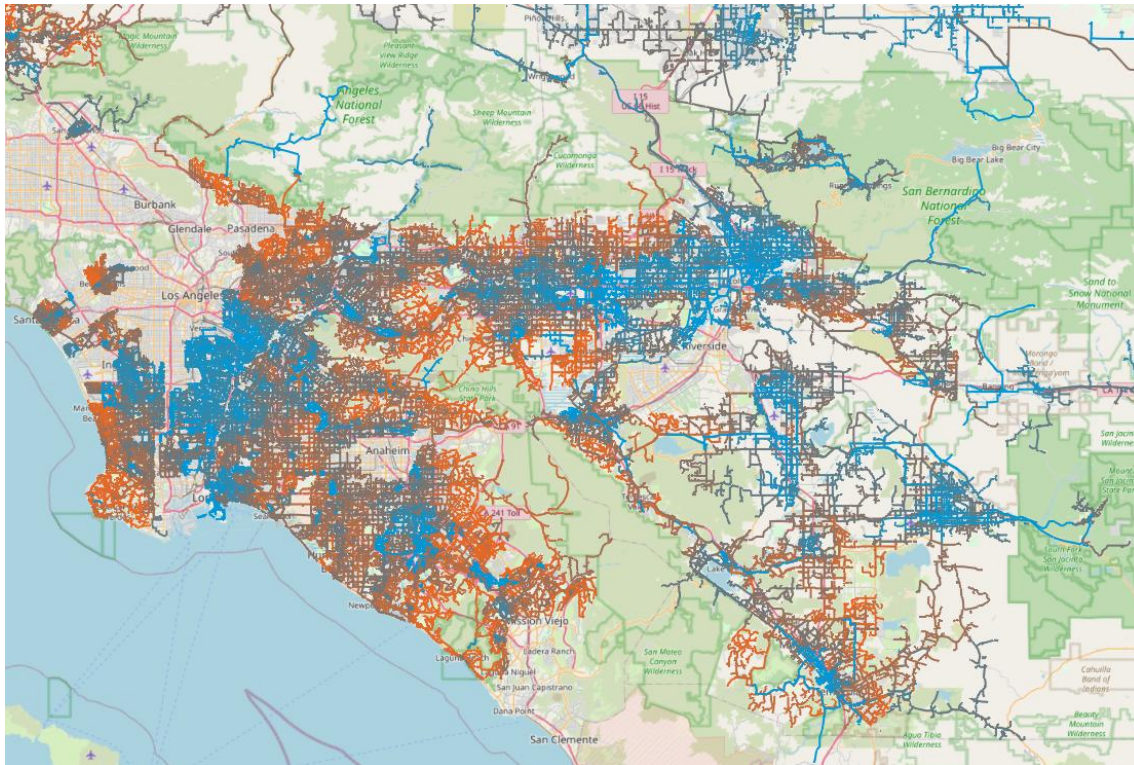


Usage variables from billing data, solar interconnection data, and property features (i.e., square footage) can be combined to create energy intensity usage (EUI) features for predicting adoption propensities

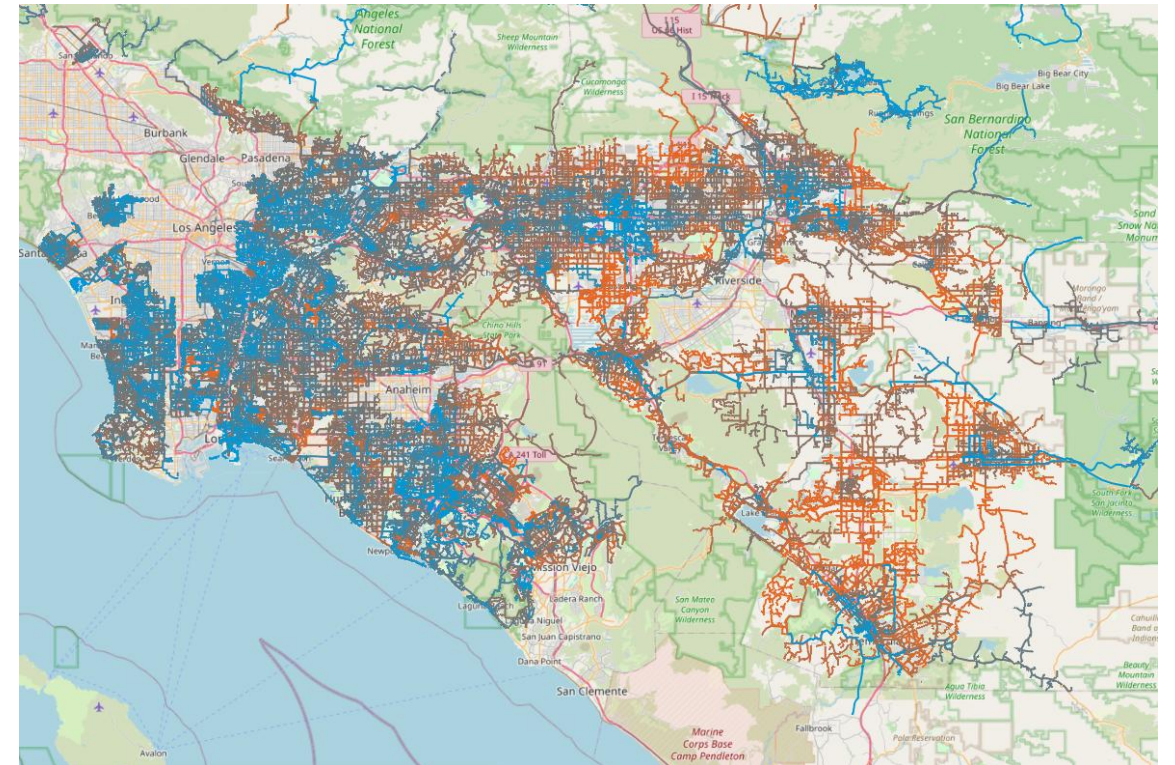


EXAMPLE: PROPENSITY SCORING OF TECHNOLOGY ADOPTION CONT.

Residential EV



Residential Solar



CURRENT DSA ADDRESS MATCHING SOLUTIONS

Classic Approach

Standardize and clean all addresses*

Merge 1:1 on full address

Find closest fuzzy match

Within city/zipcode

Distance matching (< 100m)

Flexible Approach

Standardize all addresses* and split into address components (*address #, predir, street name, posttype, postdir, unittype, unitnum, city, zip*)

Strict matching on all fields (street name, address number, unit number, zip)

For unmatched, strict matching on all fields except unit type and #

For still unmatched, matching on address # & zip, fuzzy match on street name

Distance matching (< 100m)



ADDRESS CLEANING CHALLENGES

- Address formats can vary significantly between data sources and projects
- Trying to assign each component of the address is not trivial
 - *address #, predir, street name, posttype, postdir, unittype, unitnum, city, zip*

Notable Differences	Target Address
No Post Type	119 San Leon, Irvine, CA 92606
Grouped Non-Residential Address Numbers	17053-17089 Murphy Ave, Irvine, CA 92614
Unit Type appears before Street Name, and has no address number	FITNESS, 163 Cantata, Irvine, CA 92606



ADDRESS CLEANING CHALLENGES CONT.

- Address validation
 - Ideal but difficult to do by hand
 - External APIs are expensive and impractical for millions of addresses
- Adds a preliminary step to the actual address matching process
- Can still be incorrect

Address: 6637 OASIS ROAD TWENTYNINE PALMS 92278 Region: United States [Validate]

Validated address API response

Formatted address
A post-processed, region-formatted, single-line address
6637 Oasis Road, Twentynine Palms, CA 92277-3233, USA

Verdict

Granularity
How specific is the address?

Entered Premise (building) Validated Premise (building) Geocoded Route (street/road/highway)

ArcGIS (API)

Smarty Products Solutions Resources Company Pricing

Bulk address validation tool

Standardize and validate your addresses instantly, worldwide.
For instructions, see the [documentation](#).

Step 1 of 3

Upload your file

Google (API)

Read into a DataFrame

The previous example explained how to get results as a `FeatureSet` using `as_featureset = True` parameter, and map the `FeatureSet` on the Map Widget. Next, we will convert the `FeatureSet` to a `DataFrame`, and see how the results look.

```
disney_fset = geocode("Disneyland", as_featureset = True)
disney_fset
```

<FeatureSet> 20 features

```
disney_fset.sdf.head()
```

	Loc_name	Status	Score	Match_addr	LongLabel	ShortLabel	Addr_type	Type	PlaceName
0	World	T	100	Disneyland	Disneyland, 1313 S Harbor Blvd, Anaheim, CA, 9...	Disneyland	POI	Amusement Park	Disneyland
1	World	T	100	Disneyland	Disneyland, Place des Passagers du Vent, 77700...	Disneyland	POI	Historical Monument	Disneyland
	World	T	100	Disney Land	Disney Land, Disneyland, Paris, France	Disney Land	POI	Convention	Disney Land

Smarty (Batch)



FUZZY MATCHING CHALLENGES

Key Issues

- Many cases where we should get correct matches
- Requires a m:m matching
 - For 100,000 addresses, this can mean *billions* of individual similarity calculations

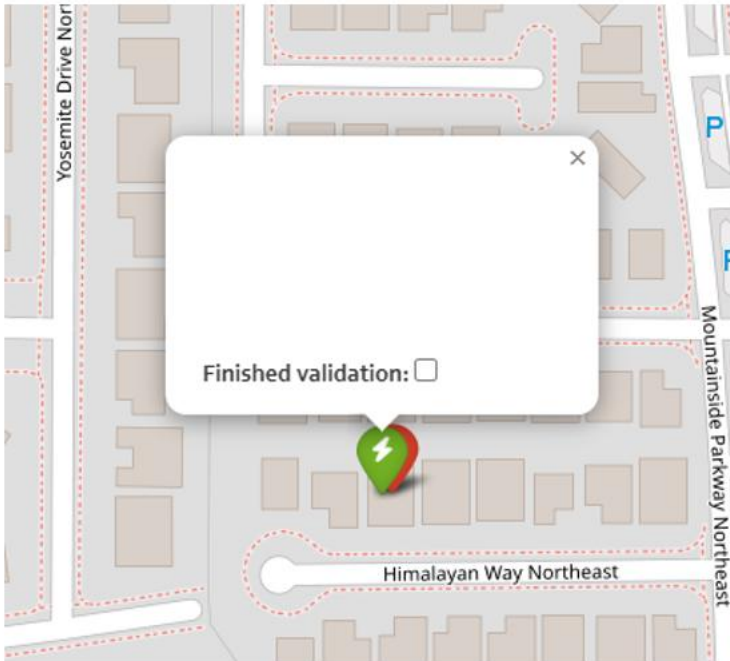
Types of High Fuzzy Matches (>97% match)

Description	Target Address	All have the same Fuzzy Match Score
Choosing the Wrong Apartment Number	742 Amsterdam Avenue, Apt 3B, New York, NY 10025	742 Amsterdam Avenue, Apt B6, New York, NY 10025
		742 Amsterdam Avenue, Apt B4, New York, NY 10025
		742 Amsterdam Avenue, Apt B3, New York, NY 10025
Issue	Target Address	Incorrect Match
Off-by-One Street Number	235 East 73rd Street, New York, NY 10021	236 East 73rd Street, New York, NY 10021
Avenue vs Street Confusion	890 3rd Avenue, New York, NY 10022	890 3rd Street, New York, NY 10022
East vs West	125 East 86th Street, New York, NY 10028	125 West 86th Street, New York, NY 10024



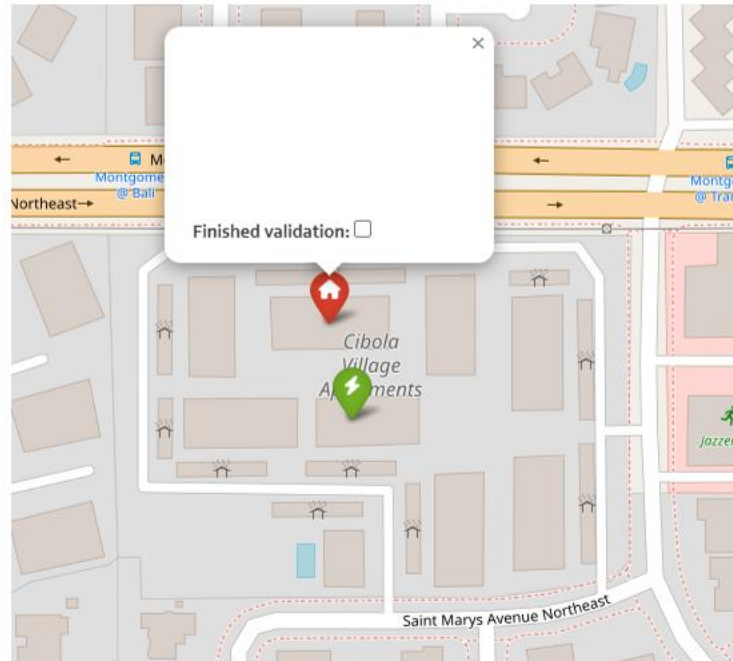
DISTANCE MATCHING CHALLENGES

Excellent Match



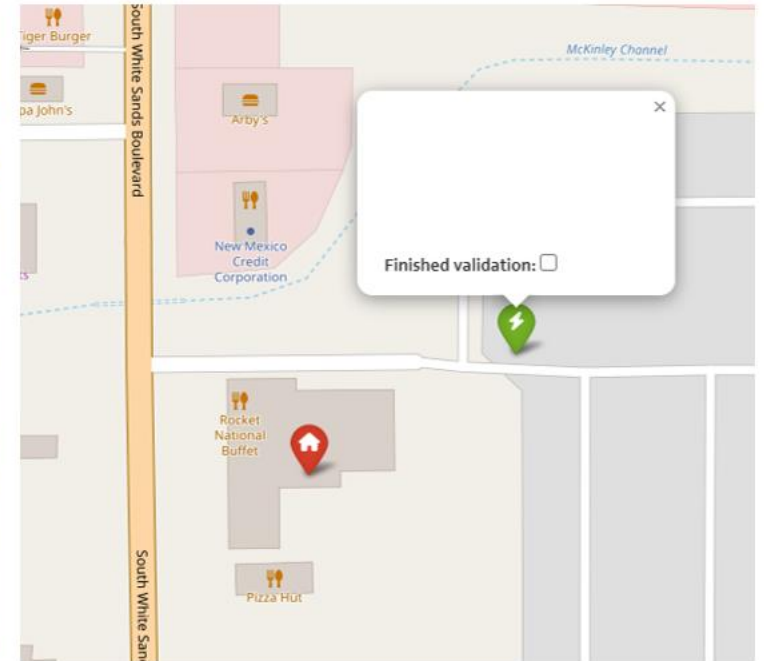
Addresses match, both premise and property markers are on the exact same building

Good Match



Addresses match, but the units do not. This commonly occurs for multi-family complexes and businesses in strip malls.

Bad Match

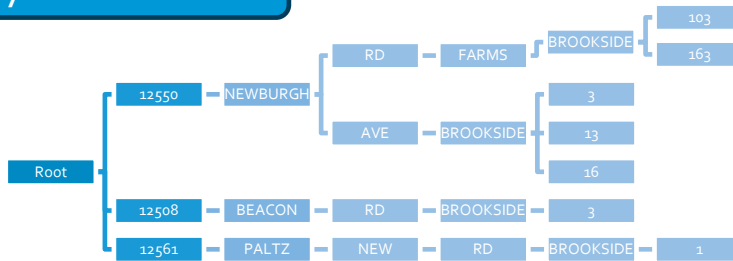


Addresses do not match. Markers refer to completely different buildings. Meter locations are guaranteed to be within the property boundaries.



UPDATED ADDRESS MATCHING PROCESS

Fuzzy Search Tree



Bag of Words Search

AVE	BEACON	BROOKSIDE	FARMS	HOOK	LN	NEW	NEWBURGH	PALTZ	RD	RED
0.000	0.000	0.717	0.000	0.000	0.000	0.000	0.343	0.000	0.174	0.000

Standardize all addresses string using off the shelf solutions

Direct match 1:1 on full address

Strict Fuzzy Match on full address

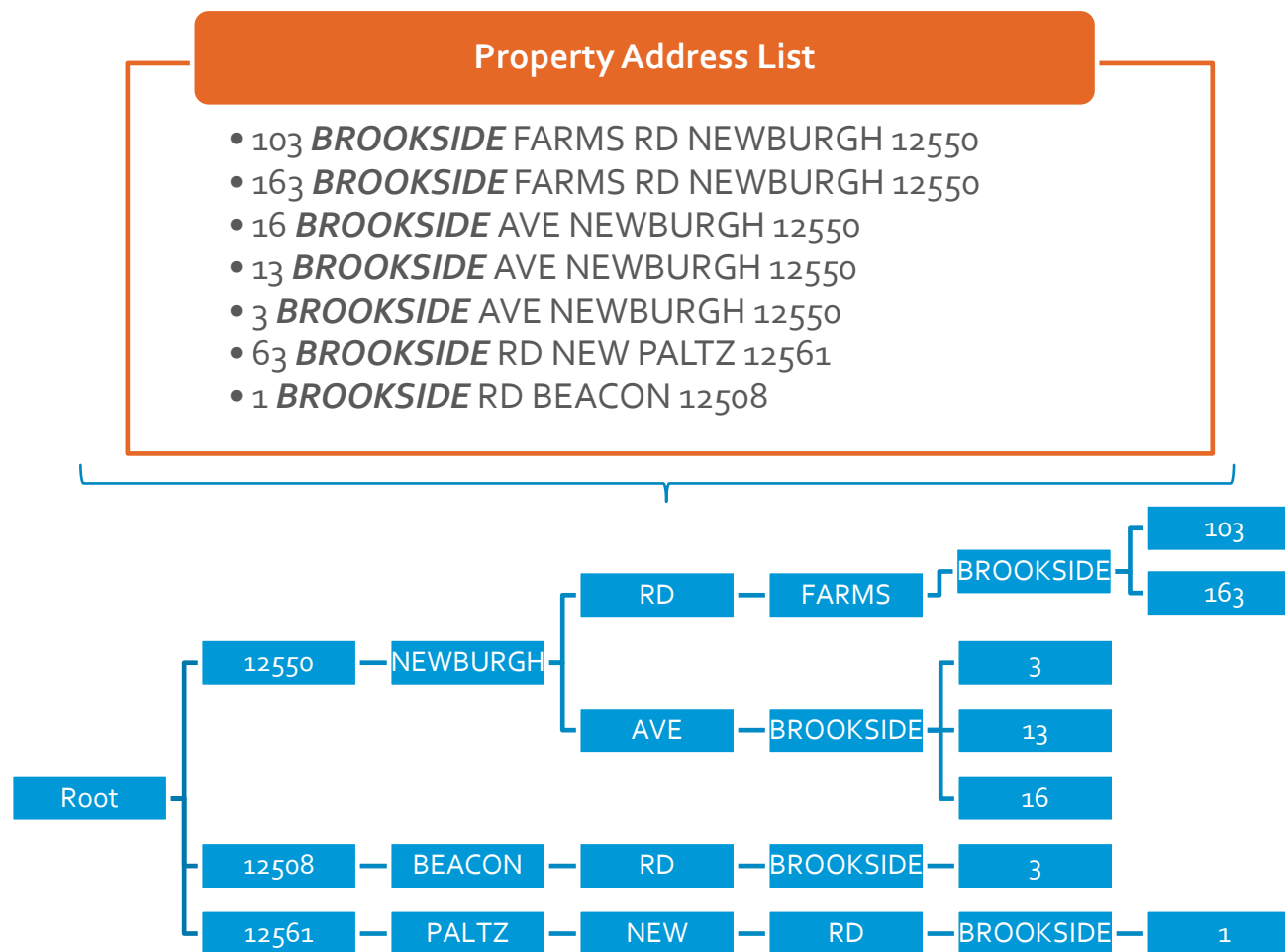
Street level match on word uniqueness

Filtered by street # and zip code

Distance matching (< 100m)

FIRST PASS: DIRECT MATCHING AND STRICT FUZZY MATCHING USING A SEARCH TREE

- The data structure known as a *trie* (search tree) is a powerful tool in address matching
- Address are stored/searched backwards
 - Reduces data size
 - Increases search speed
- Tree can be flipped for street number first search



SECOND PASS: A NEW ADDRESS MATCHING PARADIGM IS REQUIRED

Candidate Retrieval

- For each target address, identify a list of possible candidate matches
- Ensures high recall while reducing the numbers of address comparisons

2377 ROUTE 82 APT 239 LAGRANGEVILLE 12540

- 1867 RT 55 APT 1875 LAGRANGEVILLE 12540
- 2377 RT 82 APT 2391 LAGRANGEVILLE 12540
- 1131 RT 55 APT 1145 LAGRANGEVILLE 12540
- 20 AIRWAY DR APT B LAGRANGEVILLE 12540
- 9 CLUB HOUSE LN APT 15 LAGRANGEVILLE 12540
- 1807 CLOVE RD APT 1817 LAGRANGEVILLE 12540
- ...

Similarity Scoring

- Score similarity of each candidate to the target address and select the closest match
 - **Fuzzy Distance:** Minimum number of character edits required to transform the target address into the candidate address
 - **Bag-of-Words (BoW) Similarity with TD-IDF Weighting:** Measures address similarity weighted by the uniqueness of each word in the address.
 - **Inverse BoW Score:** Penalizes candidates containing words that do not exist in the target address.
- Metrics are averaged to ensure the highest possible precision



SECOND PASS: CATCHING MORE EDGE CASES WITH A BAG OF WORDS SEARCH

- Candidate list is found by simply looking up the most unique word
 - Drastically reduces the number of comparisons needed to find the best match
 - Can be done by hand when the similar score doesn't narrow down the candidates enough
- Candidates are *similarity scored*, and the best match is chosen

Target Address

163 BROOKSIDE FARM RD NEWBURGH 12550

Indexed by most unique word

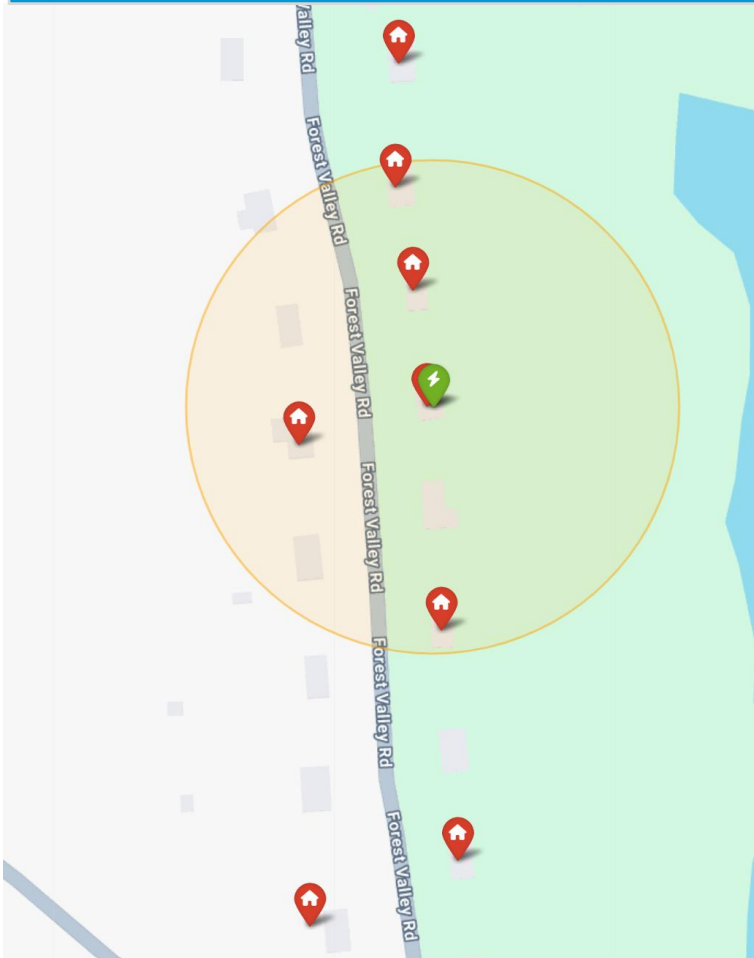
Closest Match

	AVE	BEACON	BROOKSIDE	FARMS	HOOK	LN	NEW	NEWBURGH	PALTZ	RD	RED
163 BROOKSIDE FARM RD NEWBURGH 12550	0.000	0.000	0.717	0.000	0.000	0.000	0.000	0.343	0.000	0.174	0.000
163 BROOKSIDE FARMS RD NEWBURGH 12550	0.000	0.000	0.543	0.527	0.000	0.000	0.065	0.260	0.000	0.132	0.000
103 BROOKSIDE FARMS RD NEWBURGH 12550	0.000	0.000	0.557	0.541	0.000	0.000	0.067	0.266	0.000	0.135	0.000
16 BROOKSIDE AVE NEWBURGH 12550	0.282	0.000	0.693	0.000	0.000	0.000	0.083	0.331	0.000	0.000	0.000
13 BROOKSIDE AVE NEWBURGH 12550	0.279	0.000	0.686	0.000	0.000	0.000	0.082	0.328	0.000	0.000	0.000
3 BROOKSIDE AVE NEWBURGH 12550	0.283	0.000	0.697	0.000	0.000	0.000	0.083	0.333	0.000	0.000	0.000
63 BROOKSIDE RD NEW PALTZ 12561	0.000	0.000	0.630	0.000	0.000	0.000	0.151	0.000	0.379	0.153	0.000
1 BROOKSIDE RD BEACON 12508	0.000	0.407	0.665	0.000	0.000	0.000	0.080	0.000	0.000	0.161	0.000
16 BROOKSIDE LN RED HOOK 12571	0.000	0.000	0.594	0.000	0.377	0.249	0.071	0.000	0.000	0.000	0.369

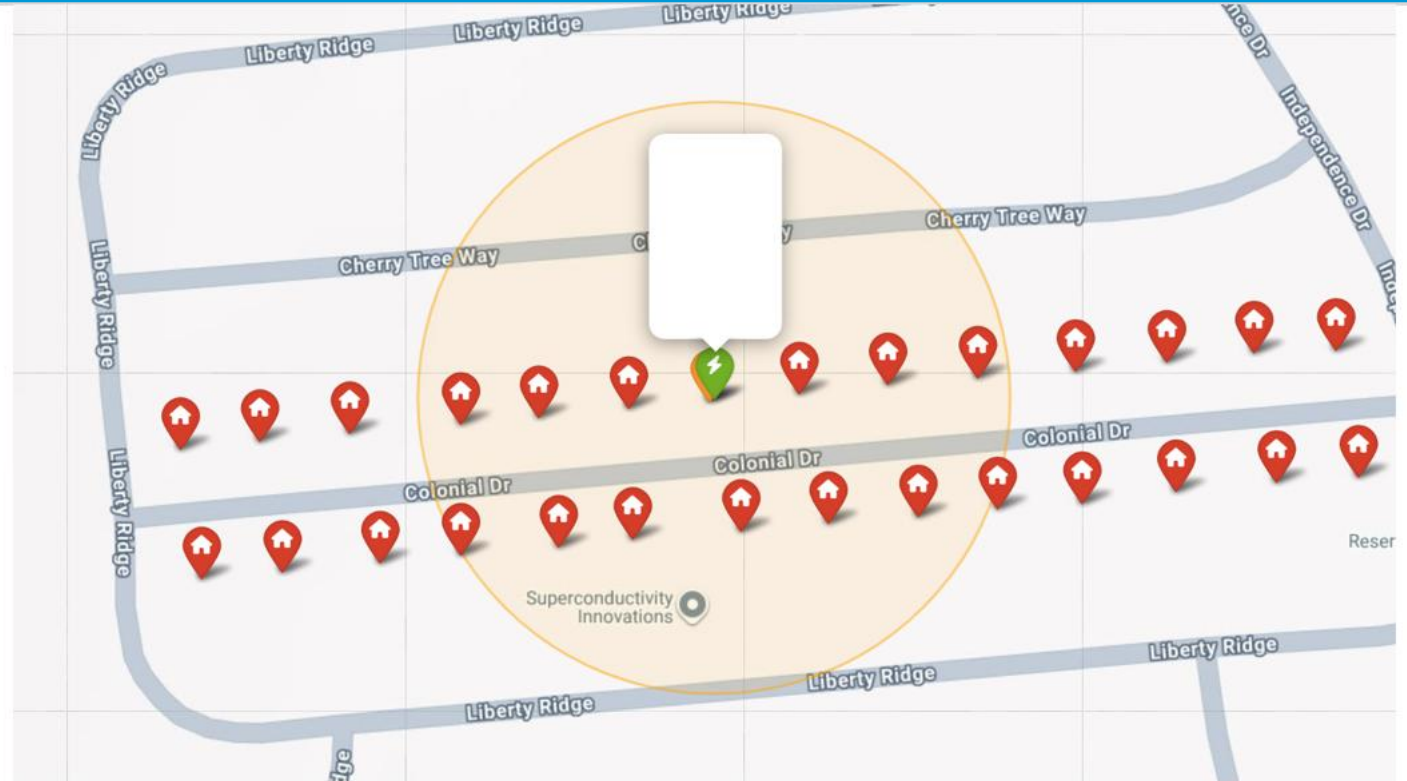


EXAMPLES OF HOW STREET LEVEL MATCHES ARE SELECTED FROM CANDIDATE LIST

Sample 1



Sample 2



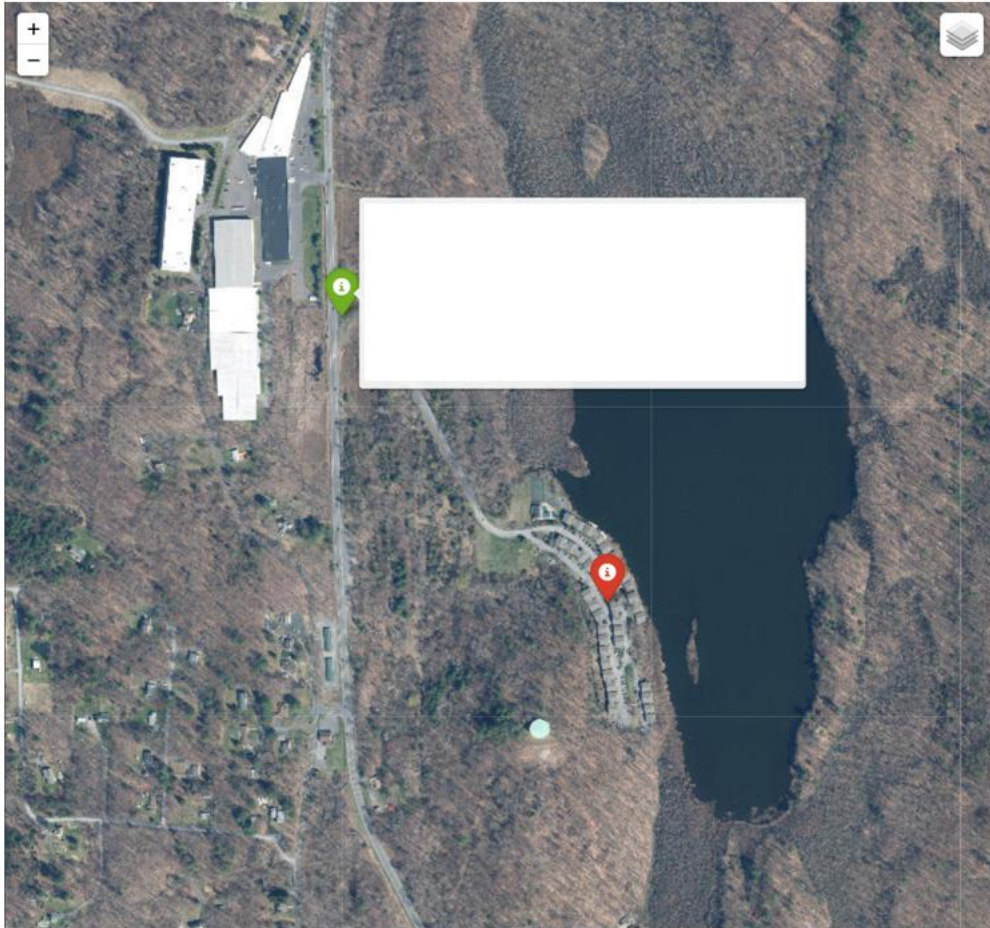
MATCHING RESULTS FROM SCE 2025 PROPENSITIES PROJECT

Match Type	Solar %	New Heat Pump %	Prior Heat Pump %
Direct Match	81.12	67.77	43.13
Strict Fuzzy Match	10.84	23.62	30.59
Street-Level Match	4.14	4.27	5.79
Distance	2.23	3.24	6.16
No Matched	1.67	1.09	14.33

- Successfully matched most of the addresses in each dataset with high confidence
- Remaining unmatched addresses are likely
 - located outside the service territory
 - incomplete/improperly formatted addresses



ADDRESS VALIDATION IS OFTEN DONE BY HAND (EXAMPLE)



This match was manually graded as unsuccessful because:

- Property data does not have apartment info and square footage is for the whole apartment complex
- Property does not contain apt number
- Green and red markers far apart - Address lat/lon is at entrance, which is distant from the property
- The CHGE customer and property towns do not match
- Billing address is residential, property match is commercial



KEY TAKEAWAYS

1. Address matching provides a new dimension to utility data
 1. Targeting and matching applications
 2. Energy burden and low-income customer identification
 3. Propensities and granular forecasting
2. Linking address records by components (*address #, street name, etc.*) can improve speed and accuracy
3. Using only address matching, we have been able to property match over 90% of the account level utility data across a wide range of projects
4. Address standardization and validation are key considerations in the matching process

QUESTIONS?



David Pojunas
Senior Quantitative Analyst
dpojunas@demandsideanalytics.com

Josh Bode
Partner
jbode@demandsideanalytics.com