

I'll help you profile this dataset by analyzing various aspects:

1. Basic Information:

- Number of columns: 12
- Number of rows: approximately 85 records

2. Column Analysis:

brokered_by: - Contains numeric IDs and missing values - Many entries with “-1” or “nan” - Mixed data quality

status: - Main categories: “sold”, “for_sale”, “s”, “f”, “Unknown”, “-” - Contains inconsistent representations - Some missing values

price: - Numeric values ranging from about \$22,400 to 12,312,100 – *Some entries prefixed with ”* - Contains “nan” and “-1” values - Mixed format issues

bed: - Mix of numeric and text representations (“three”, “four”, “five”) - Range typically 1-12 bedrooms - Contains “-1” and missing values

bath: - Similar to bed, mixed numeric and text (“two”, “three”) - Range typically 1-12 bathrooms - Contains “nan” and missing values

acre_lot: - Numeric values ranging from very small (0.0014) to large (27.59) - Contains “-1” and missing values

street: - Mix of full addresses and partial information - Some entries with format “Street Name, Number” - Contains “Unknown” and “-” values

city: - Various US city names - Contains “Unknown” and missing values

state: - US state names and abbreviations - Inconsistent format (full names vs abbreviations) - Contains missing values and “_”

zip_code: - 5-digit US zip codes - Contains “-1” and missing values

house_size: - Numeric values in square feet - Some very small values (possibly errors) - Contains “-1” and missing values

prev_sold_date: - Mixed date formats (YYYY-MM-DD, MM/DD/YY) - Contains “Unknown” and missing values - Dates ranging from 1991 to 2022

3. Data Quality Issues:

- Inconsistent formatting across columns
- Mixed data types within same columns
- Multiple representations of missing values (“-1”, “nan”, “-”, empty)
- Inconsistent date formats
- Possible unit conversion issues in house_size
- Inconsistent state name formatting
- Various placeholder values (“Unknown”)

4. Recommendations:

- Standardize missing value representation
- Convert text numbers to numeric values
- Standardize date formats
- Normalize state names to consistent format
- Validate extreme values in numeric fields
- Clean up special characters in price values
- Standardize status categories

This dataset would benefit from significant cleaning and standardization before analysis.