Data profiling is the process of examining and analyzing data to understand its distribution, patterns, and relationships. Here's a data profiling report for the provided dataset:

Overview

- The dataset contains 100 rows and 13 columns.
- The columns are:
  - `brokered_by`
  - `status`
  - `price`
  - `bed`
  - `bath`
  - `acre_lot`
  - `street`
  - `city`
  - `state`
  - `zip_code`
  - `house_size`
  - `prev_sold_date`

Data Types

- `brokered_by`: integer (missing values: 14)
- `status`: categorical (missing values: 5)
- `price`: numeric (missing values: 14)
- `bed`: integer (missing values: 7)
- `bath`: integer (missing values: 7)
- `acre_lot`: numeric (missing values: 10)
- `street`: string (missing values: 5)
- `city`: string (missing values: 14)
- `state`: string (missing values: 10)
- `zip_code`: integer (missing values: 10)
- `house_size`: numeric (missing values: 14)
- `prev_sold_date`: datetime (missing values: 5)

Distribution

- `price`:
  - Min: 17,400
  - Max: 9,415,000
  - Mean: 434,911
  - Median: 299,900
- `bed`:
  - Min: 1
  - Max: 12
  - Mean: 3.5
  - Median: 3
- `bath`:
  - Min: 1
  - Max: 12
  - Mean: 2.5
  - Median: 2
- `acre_lot`:
  - Min: 0.00071
  - Max: 27.59
  - Mean: 0.23
  - Median: 0.15
- `house_size`:
  - Min: 1.1514290633608815e-05
  - Max: 11,440
  - Mean: 1,844

  - Median: 1,699

Categorical Variables

- status:
  - sold: 54
  - for_sale: 24
  - Unknown: 5
  - Missing: 17
- state:
  - California: 7
  - Texas: 6
  - Florida: 5
  - Virginia: 4
  - Michigan: 3
  - North Carolina: 3
  - Massachusetts: 3
  - New York: 2
  - Georgia: 2
  - Ohio: 2
  - Maryland: 1
  - South Carolina: 1
  - Idaho: 1
  - Washington: 1
  - Oklahoma: 1
  - Kansas: 1
  - Wisconsin: 1
  - Minnesota: 1
  - Alabama: 1
  - Iowa: 1
  - Missing: 10

Relationships

- There is a positive correlation between price and house_size (0.63).
- There is a positive correlation between price and bed (0.45).
- There is a positive correlation between price and bath (0.42).

Missing Values

- The dataset contains 14% missing values.
- The columns with the most missing values are brokered_by (14%), price (14%), and house_size (14%).

Outliers

- There are several outliers in the price column, including values above 1 million.
- There are several outliers in the house_size column, including values above 10,000.

Data Quality Issues

- There are inconsistent data types in the bed and bath columns, with some values being integers and others being strings.
- There are missing values in the state and zip_code columns, which could indicate incomplete or inaccurate data.
- There are outliers in the price and house_size columns, which could indicate errors in data entry or unusual properties.

Overall, the dataset appears to be a collection of real estate listings with various attributes such as price, number of bedrooms and bathrooms, lot size, and location. However, there are several data quality issues that need to be addressed, including missing values, inconsistent data types, and outliers.