

# Home Sale Events Data Report

## Overview

The dataset was loaded from the `raw\_home\_sale\_events` table in a PostgreSQL database. It represents real estate transaction events. Each row contains information about a specific home sale, including the event's date and other attributes related to the property.

The data was converted into a Pandas DataFrame for further exploration. Below is the summary of the columns, their descriptions, data types, and any data quality issues observed.

## Data Summary

Column Name	Description	Data Type	Contains Nulls?	Null Placeholder	Reasonable Min/Max Values
id	Unique identifier for the event	Integer	No	N/A	> 0
event_date	Date of the event	Date	No	N/A	Valid date format
price	Sale price of the property	Float	Yes	0.0 or `None`	Min: 10,000; Max: 10,000,000
bedrooms	Number of bedrooms in the house	Integer	Yes	`None` or 0	Min: 1; Max: 20
bathrooms	Number of bathrooms	Float	Yes	`None` or 0.0	Min: 0.5; Max: 10
sqft	Square footage of the property	Integer	Yes	`None` or 0	Min: 100; Max: 20,000
year_built	Year the house was built	Integer	Yes	`None`	Min: 1800; Max: Current Year
zipcode	ZIP code of the property's location	String	Yes	`None`	Valid 5-digit ZIP codes

## Data Quality Issues

1. **Missing Values:** Some columns, such as `price`, `bedrooms`, and `sqft`, contain missing values represented as `None` or zero. These should be handled to avoid skewing analysis results.

2. Outliers: Extreme values in columns like `price` and `sqft` may indicate data entry errors or truly exceptional properties. Outlier detection is recommended.
3. Date Consistency: Ensure that `event\_date` values fall within a reasonable range for home sale events.
4. Year Built: Some properties may have a `year\_built` value far in the past, which could either be correct (historical properties) or erroneous.

## SQL `CREATE TABLE` Statement

Here is a SQL statement that represents the inferred schema for the home sale events data:

```
CREATE TABLE home_sale_events (  
  id SERIAL PRIMARY KEY,  
  event_date DATE NOT NULL,  
  price FLOAT NULL,  
  bedrooms INTEGER NULL,  
  bathrooms FLOAT NULL,  
  sqft INTEGER NULL,  
  year_built INTEGER NULL,  
  zipcode VARCHAR(10) NULL  
);
```

```
CREATE TABLE home_sale_events (  
  id SERIAL PRIMARY KEY,  
  event_date DATE NOT NULL,  
  price FLOAT NULL,  
  bedrooms INTEGER NULL,  
  bathrooms FLOAT NULL,  
  sqft INTEGER NULL,  
  year_built INTEGER NULL,  
  zipcode VARCHAR(10) NULL  
);
```