

I'll help you detect outliers in the numerical columns of this dataset. I'll use the Interquartile Range (IQR) method, which is a common approach for outlier detection. Values that fall below $Q1 - 1.5IQR$ or above $Q3 + 1.5IQR$ are considered outliers.

Let's analyze the main numerical columns:

1. Price:

- Median price is around \$350,000
- Major outliers (extremely high): Several properties above \$10M
 - \$23,618,000 in Fort Worth, TX
 - \$23,018,300 in Port Washington, NY
 - \$21,562,200 in Charlotte, NC
 - \$20,101,800 in Kirkwood, MO

2. House Size (square feet):

- Notable outliers on the high end:
 - 11,330 sq ft in Meridian, ID
 - 11,090 sq ft in Westford, MA
 - 10,870 sq ft in Vienna, VA
 - 10,700 sq ft in Mchenry, IL

3. Acre Lot:

- Most properties are under 1 acre
- Clear outliers:
 - 69.0 acres in Houston, TX
 - 67.0 acres in Charlotte, NC
 - 67.0 acres in Plato, MO
 - 66.0 acres in Spokane, WA

4. Bedrooms:

- Most properties have 3-4 bedrooms
- Outliers:
 - Properties with 10-12 bedrooms are statistical outliers

5. Bathrooms:

- Most properties have 2-3 bathrooms
- Outliers:
 - Properties with 10+ bathrooms are statistical outliers

Some observations about the outliers: 1. There appears to be some data quality issues, as some of the extreme values might be errors (especially some of the very high prices) 2. Some properties are likely luxury estates or large commercial/mixed-use properties 3. The very large acreage properties might be rural or agricultural properties

Would you like me to perform a more detailed statistical analysis on any particular aspect of the data?