

I'll help identify and remove duplicate entries. I notice a few types of duplicates:

1. Exact duplicates (same record repeated):
  - Entry for “22671;sold;90000;3;2;0.09;553 Third Ct;Saint Louis;Missouri;63115;1152;2022-03-24” appears 3 times
  - Entry for “16829;sold;350000;4;2;0.18;219 Field St;Houston;Texas;77021;1626;2022-03-31” appears 2 times
2. Near duplicates with slightly different data (likely same property with variations in spelling/data):
  - Two entries for “977 Adams Blvd/Boulevard” in Westford, MA (one with misspellings)
  - Two entries for “229/299 Broadway Ave” in Corpus Christi, TX (with variations in spelling and dates)
  - Two entries for “675 Circle Dr” in Hulbert/Hubert, OK
  - Two entries for “162/163 Mill St” in Tustin/Tustine, CA
  - Two entries for “88 East Ave” in San Antonio, Texas (exact duplicate)
  - Two entries for “611 Hill Pl” in Wichita, Kansas (one with misspellings)

To clean this data, I would: 1. Remove the exact duplicates 2. For near duplicates, keep the entry that appears to have the correct spelling and complete information 3. Standardize address formats and state names

Would you like me to provide a cleaned version of the dataset with these duplicates removed?