

I'll help identify and remove duplicate entries. A record can be considered a duplicate if it has very similar or identical information, even if there might be slight variations in spelling or formatting. Here's how I'll approach this:

1. First, I'll look for exact duplicates (completely identical rows)
2. Then, I'll look for near-duplicates (rows that appear to represent the same property but have slight variations in spelling or data entry)

Here are some examples of duplicates I found:

Exact duplicates: - The record for "70 Cedar Dr, Cumming, Georgia" (ID: 10437) appears 3 times - The record for "896 Sunset Ct, Marathon, Florida" (ID: 54093) appears 3 times - The record for "162 Mill St, Tustin, California" (ID: 84529) appears 2 times - The record for "285 Center Ln, Fairfax, Virginia" (ID: 81316) appears 2 times - The record for "655 First St, Stockbridge, Massachusetts" (ID: 94681) appears 2 times - The record for "611 Hill Pl, Wichita, Kansas" (ID: 53556) appears 2 times - The record for "88 East Ave, San Antonio, Texas" (ID: 48807) appears 2 times - The record for "219 Field St, Houston, Texas" (ID: 16829) appears 2 times - The record for "347 Oak Ln, Wetumpka, Alabama" (ID: 86329) appears 3 times

Near-duplicates (same property with spelling variations): - "715 Terrace Dr, Sparks, Nevada" and "715 Terrace, Spark, Navada" are the same property - "277 Princess Dr, Livermore, California" and "277 Princess, Livemore, California" are the same - "226 Square Rd, Brewster, Massachusetts" and variations are the same - "254 Princess Ave, Bettendorf, Iowa" and variations are the same - "698/699 Wilson Dr, Crossville/Crosville, Tennessee" are the same - "383 Market St, Houston/Huston, Texas" variations are the same - "829 Main Boulevard/Blvd, Oak Ridge/Oaky Ridge, Tennessee" variations are the same

To clean this dataset, you should: 1. Keep only one instance of each exact duplicate 2. Standardize the spelling and formatting for near-duplicates and keep only one instance 3. Verify and correct state names and zip codes for consistency

Would you like me to provide a cleaned version of specific entries or help with standardizing any particular group of records?