

I'll help you clean this dataset. Here are the main issues I've identified and how we can clean them:

1. Standardize status values:
 - Convert 'f', 's', 'Unknown', '-' to proper values ('for_sale' or 'sold')
 - Remove any invalid status values
2. Price formatting:
 - Remove '\$' symbols
 - Convert string prices to numeric
 - Handle 'nan' and '-1' values
3. Bed/Bath counts:
 - Convert text numbers ('two', 'three', 'four') to numeric values
 - Handle 'nan' and '-1' values
4. Address standardization:
 - Handle missing street addresses (marked as 'Unknown' or empty)
 - Fix inconsistent formatting (like "Garden Ct, 614")
5. Location data:
 - Handle missing/unknown cities and states
 - Standardize state abbreviations (TX → Texas)
 - Handle invalid zip codes (-1)
6. Date standardization:
 - Convert different date formats (12/30/21, 03/24/22) to YYYY-MM-DD
 - Handle missing dates
7. Numeric fields (acre_lot, house_size):
 - Convert to appropriate numeric types
 - Handle negative values
 - Handle extremely small or large values that might be errors
8. Remove duplicate entries:
 - Check for exact duplicates
 - Check for near-duplicates with slightly different information

Would you like me to implement these cleaning steps with specific code, or would you prefer to focus on particular aspects of the cleaning process?