I can help you with data imputation for this dataset. There are several approaches we could take:

1. For numerical columns (price, bed, bath, acre_lot, house_size):

- Replace -1 with NaN first
- Use median imputation for continuous variables
- Use mode imputation for discrete variables (like bed and bath)

2. For categorical columns (brokered_by, status, street, city, state):

- Replace '-' and 'Unknown' with NaN
- Use mode imputation
- For street addresses, we might want to keep them as NaN since imputation might not make sense

3. For date columns (prev_sold_date):

- Replace 'Unknown' with NaN
- Use median date or forward/backward fill

Would you like me to proceed with implementing these imputation strategies? I can provide the code and resulting statistics. Also, please let me know if you have any specific requirements or preferences for the imputation methods.

Note: Given the size and complexity of the dataset, it would be helpful to know: 1. Which columns are your priority for imputation? 2. Are there any specific business rules we should consider? 3. Would you prefer simple imputation (mean/median/mode) or more sophisticated methods (like KNN or regression-based imputation)?