I can help you with data imputation for this dataset. There are several approaches we can take, depending on the type of data and the nature of missing values. Here's a suggested approach:

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

- Replace -1 with NaN first
- Use median imputation for these values
- For house prices, we might want to consider using more sophisticated methods like regression based on other features

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

- Replace '-' and 'Unknown' with NaN
- Use mode imputation (most frequent value)

3. For date columns (prev_sold_date):

- Replace 'Unknown' with NaN
- Use median date

4. For zip_code:

- Replace -1 with NaN
- Use mode imputation by state/city where possible

Would you like me to proceed with implementing this imputation strategy? I can provide the Python code using pandas and scikit-learn to perform these operations.

Also, before proceeding, we should consider: 1. Whether to impute all missing values or only specific columns 2. Whether to use more sophisticated imputation methods like KNN or regression-based imputation 3. Whether to handle outliers before imputation

Please let me know your preferences and I'll provide the appropriate implementation.