

Data imputation involves filling in missing or incomplete data in a dataset. Given the dataset you provided, we can apply several strategies to handle missing values. Here are some common imputation techniques:

1. **Mean/Median/Mode Imputation:** For numerical columns, you can fill missing values with the mean, median, or mode of the column. For categorical columns, you can use the mode.
2. **Forward/Backward Fill:** This method fills missing values using the previous or next value in the column. It is useful for time-series data.
3. **Constant Value Imputation:** Fill missing values with a constant value, such as 0 or “Unknown”.
4. **Predictive Imputation:** Use machine learning models to predict and fill missing values based on other available data.
5. **Dropping Missing Values:** In some cases, it might be appropriate to drop rows or columns with too many missing values.

Let's apply some of these techniques to your dataset:

Imputation Strategy

1. **Numerical Columns:**
 - For `price`, `bed`, `bath`, `acre_lot`, and `house_size`, use the median to fill missing values, as it is less sensitive to outliers.
 - For `zip_code`, use the mode since it is categorical.
2. **Categorical Columns:**
 - For `brokered_by`, `status`, `street`, `city`, `state`, and `prev_sold_date`, fill missing values with “Unknown”.
3. **Special Cases:**
 - If a column has a specific placeholder for missing values (e.g., `-1` or `nan`), replace these with the chosen imputation value.

Imputed Dataset

Here's how you might perform the imputation:

```
import pandas as pd
import numpy as np

# Load the dataset
data = pd.read_csv('your_dataset.csv')

# Fill numerical columns with median
numerical_cols = ['price', 'bed', 'bath', 'acre_lot', 'house_size']
for col in numerical_cols:
    data[col] = data[col].replace([-1, 'nan'], np.nan)
    data[col] = data[col].fillna(data[col].median())

# Fill categorical columns with 'Unknown'
categorical_cols = ['brokered_by', 'status', 'street', 'city', 'state', 'prev_sold_date']
for col in categorical_cols:
    data[col] = data[col].replace(['nan', -1], 'Unknown')
    data[col] = data[col].fillna('Unknown')

# Fill zip_code with mode
data['zip_code'] = data['zip_code'].replace([-1, 'nan'], np.nan)
data['zip_code'] = data['zip_code'].fillna(data['zip_code'].mode()[0])

# Display the imputed dataset
print(data.head())
```

This code snippet demonstrates how to handle missing values using simple imputation techniques. Depending on the context and the importance of the data, you might choose more sophisticated methods, such as predictive modeling, for imputation.