

Data imputation involves filling in missing values in a dataset. Here are some common strategies for handling missing data:

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. **Interpolation:** This method estimates missing values based on other available data points. It is often used for time series data.
4. **Predictive Modeling:** Use machine learning models to predict missing values based on other features in the dataset.
5. **Domain-Specific Imputation:** Use domain knowledge to fill in missing values with reasonable estimates.

Let's apply some of these strategies to the dataset:

Numerical Columns

- **Price, Bed, Bath, Acre Lot, House Size:** Use median imputation for these columns, as they are numerical and may have outliers that could skew the mean.

Categorical Columns

- **Brokered By, Status, Street, City, State, Zip Code, Prev Sold Date:** Use mode imputation for these columns, as they are categorical.

Imputation Steps

1. **Price, Bed, Bath, Acre Lot, House Size:** Fill missing values with the median of each column.
2. **Brokered By, Status, Street, City, State, Zip Code, Prev Sold Date:** Fill missing values with the mode of each column.

Here's how you can perform these imputations in Python using pandas:

```
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].fillna(data[col].median(), inplace=True)

# Fill categorical columns with mode
categorical_cols = ['brokered_by', 'status', 'street', 'city', 'state', 'zip_code', 'prev_sold_date']
for col in categorical_cols:
    data[col].fillna(data[col].mode()[0], inplace=True)

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

This code will fill in the missing values in the dataset using the specified imputation strategies. Note that this is a basic approach, and more sophisticated methods may be needed depending on the context and the importance of the missing data.