

I'll help you clean and analyze this dataset. Here are some common data wrangling steps we can perform:

1. First, let's handle the missing values (NaN):

```
# Remove rows with all NaN values
df = df.dropna(how='all')
```

```
# Check remaining missing values
missing_values = df.isnull().sum()
```

2. Convert data types appropriately:

```
# Convert price to numeric, removing any currency symbols
df['price'] = pd.to_numeric(df['price'], errors='coerce')
```

```
# Convert dates to datetime
df['prev_sold_date'] = pd.to_datetime(df['prev_sold_date'])
```

```
# Convert numeric columns to appropriate types
numeric_columns = ['brokered_by', 'acre_lot', 'zip_code', 'house_size']
df[numeric_columns] = df[numeric_columns].apply(pd.to_numeric, errors='coerce')
```

3. Split bedrooms\_bathrooms into separate columns:

```
# Split the bedrooms_bathrooms column
df[['bedrooms', 'bathrooms']] = df['bedrooms_bathrooms'].str.split(',', expand=True)
df['bedrooms'] = pd.to_numeric(df['bedrooms'], errors='coerce')
df['bathrooms'] = pd.to_numeric(df['bathrooms'], errors='coerce')
```

4. Clean up the status column:

```
df['status'] = df['status'].str.lower()
```

5. Create some useful derived features:

```
# Price per square foot
df['price_per_sqft'] = df['price'] / df['house_size']

# Days since last sold (as of today)
df['days_since_sold'] = (pd.Timestamp.now() - df['prev_sold_date']).dt.days
```

6. Basic statistics:

```
summary_stats = df.describe()

# Group by status and calculate mean prices
price_by_status = df.groupby('status')['price'].agg(['mean', 'count'])
```

7. Clean up state information from address:

```
df['state'] = df['address'].str.split(',').str[-1].str.strip()
```

Would you like me to perform any specific analysis or cleaning on particular columns? Or would you like to see the results of any of these operations in detail?