

To perform outlier detection on the given dataset, we'll use the Interquartile Range (IQR) method and the Z-score method. We'll focus on the numerical columns: price, bed, bath, acre\_lot, house\_size.

```
import pandas as pd
import numpy as np

# Load the dataset
data = {
    # ... your data here ...
}

df = pd.DataFrame(data)

# Define the numerical columns
numerical_cols = ['price', 'bed', 'bath', 'acre_lot', 'house_size']

# Calculate Q1 and Q3 for each numerical column
Q1 = df[numerical_cols].quantile(0.25)
Q3 = df[numerical_cols].quantile(0.75)

# Calculate IQR for each numerical column
IQR = Q3 - Q1

# Define the lower and upper bounds for outliers
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

# Identify outliers using IQR
outliers_iqr = df[(df[numerical_cols] < lower_bound).any(axis=1) | (df[numerical_cols] > upper_bound).any(axis=1)]

# Calculate Z-scores for each numerical column
z_scores = np.abs((df[numerical_cols] - df[numerical_cols].mean()) / df[numerical_cols].std())

# Identify outliers using Z-scores (threshold: 3 standard deviations)
outliers_z_score = df[(z_scores > 3).any(axis=1)]

# Print the results
print("Outliers (IQR method):")
print(outliers_iqr)
print("\nOutliers (Z-score method):")
print(outliers_z_score)
```

This code will output the rows that are identified as outliers using both the IQR method and the Z-score method.

#### Interpretation:

- The IQR method identifies outliers as data points that fall below  $Q1 - 1.5 * IQR$  or above  $Q3 + 1.5 * IQR$ .
- The Z-score method identifies outliers as data points that are more than 3 standard deviations away from the mean.

#### Note:

- The choice of method depends on the distribution of the data. If the data is normally distributed, the Z-score method may be more suitable. If the data is skewed or has outliers, the IQR method may be more robust.
- The threshold values ( $1.5 * IQR$  and 3 standard deviations) can be adjusted based on the specific requirements of the analysis.

#### Example Use Case:

Suppose we want to identify outliers in the price column. We can use the IQR method to detect outliers:

```
price_outliers_iqr = df[(df['price'] < lower_bound['price']) | (df['price'] > upper_bound['price'])]
print(price_outliers_iqr)
```

This will output the rows where the price value is an outlier according to the IQR method.