

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Create a sample dataset with outliers
data = {
    'X': np.random.normal(loc=0, scale=1, size=1000), # Inliers
    'Y': np.concatenate([np.random.normal(loc=0, scale=1, size=950),
np.random.normal(loc=5, scale=1, size=50)]) # Outliers
}

# Create DataFrame
df = pd.DataFrame(data)

# Display the original dataset
print("Original Dataset:")
print(df)

# Detecting Outliers using Z-score method
from scipy.stats import zscore

# Calculate z-scores for each data point
z_scores = np.abs(zscore(df))

# Define a threshold to identify outliers
threshold = 3

# Find the indices of outliers
outlier_indices = np.where(z_scores > threshold)

# Remove outliers
df_cleaned = df[(z_scores < threshold).all(axis=1)]

# Display the cleaned dataset
print("\nCleaned Dataset (Outliers Removed):")
print(df_cleaned)

# Visualization of Outliers
plt.figure(figsize=(10, 6))

# Scatter plot of original dataset
plt.scatter(df['X'], df['Y'], color='blue', label='Original Data')

# Highlight outliers
plt.scatter(df.iloc[outlier_indices[0]]['X'],
df.iloc[outlier_indices[0]]['Y'], color='red', label='Outliers')

```

```
plt.title('Outlier Detection using Z-score')
plt.xlabel('X')
plt.ylabel('Y')
plt.legend()
plt.grid(True)
plt.show()
```

Original Dataset:

	X	Y
0	1.905064	-1.684729
1	-0.014777	-0.499417
2	-0.300479	-1.426938
3	-0.355029	-0.932570
4	-1.892362	-1.012457
..
995	1.079931	5.034190
996	1.094167	4.218515
997	-1.094241	5.597860
998	-0.147637	4.493923
999	1.131812	4.311554

[1000 rows x 2 columns]

Cleaned Dataset (Outliers Removed):

	X	Y
0	1.905064	-1.684729
1	-0.014777	-0.499417
2	-0.300479	-1.426938
3	-0.355029	-0.932570
4	-1.892362	-1.012457
..
991	2.285601	4.313780
993	-0.913595	4.313898
994	0.809741	3.407789
996	1.094167	4.218515
999	1.131812	4.311554

[967 rows x 2 columns]

