```
import pandas as pd

# Load your dataset
# Replace 'your_dataset.csv' with the path to your dataset file
df = pd.read_csv('/content/AcademicPerformance(EXP5,6,7).csv')

# Check for missing values in the entire dataset
print(df.isnull().sum())
print(df.mean())
df.fillna(df.mean(),inplace = True)
# Check for missing values in the entire dataset
print(df.isnull().sum())
dfs = (df - df.mean()) / df.std()
Q1 = dfs.quantile(0.25)
Q3 = dfs.quantile(0.75)

# Calculate IQR (Interquartile Range) for each column
IQR = Q3 - Q1

# Define threshold for outliers (e.g., 1.5 times the IQR)
threshold = 1.5

# Determine outliers
lower_bound = Q1 - threshold * IQR
upper_bound = Q3 + threshold * IQR

# Remove outliers
filter_d = dfs[~((dfs < lower_bound) | (dfs > upper_bound)).any(axis=1)]

# Print original and filtered data shapes
print("Original data shape:", dfs.shape)
print("Filtered data shape:", filter_d.shape)
import matplotlib.pyplot as plt
import seaborn as sns

# Create a box plot to visualize outliers in the 'Temp_C' column
plt.figure(figsize=(12,10))

sns.boxplot(data=filter_d, orient='v')
```

```
plt.title('Boxplot of Temperature (°C)')
plt.xlabel('Temperature (°C)')
plt.show()
skewness = filter_d.skew()
print("Skewness of the dataset:")
print(skewness)
import numpy as np
# Apply square root transformation to the dataset
df_sqrt = np.sqrt(filter_d)

# Apply cube root transformation to the dataset
df_cbrt = np.cbrt(filter_d)

# Apply reciprocal transformation to the dataset
df_reciprocal = 1 / filter_d
# Apply exponential transformation to the dataset
df_exp = np.exp(filter_d)



print("Logarithmic Transformation:")
print(df_log.skew())
print("\nSquare Root Transformation:")
print(df_sqrt.skew())
print("\nReciprocal Transformation:")
print(df_reciprocal.skew())
print("\nExponential Transformation:")
print(df_exp.skew())
```