LAB ASSIGNMENT-2

MACHINE LEARNING

NAME : M Jeyapathy

USN : 22BTRAD016

1. Load a dataset with outliers values (Boston Housing Dataset).

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.datasets import load\_boston

# Load Boston Housing Dataset

boston = load\_boston()

boston\_df = pd.DataFrame(boston.data, columns=boston.feature\_names)

boston\_df['MEDV'] = boston.target # MEDV is the target variable (house prices)

This shows the dataset by using boston housing dataset using the respective libraries such as matplotlib which is used to show the graphical representation of those datasets and seaborn displays the histogram graphical representation.

1. Use visualization or statistical methods to detect outliers.

Visualization - Box plot

plt.figure(figsize=(12, 6))

sns.boxplot(x=boston\_df['MEDV'])

plt.title('Box plot of MEDV (House Prices)')

plt.show()

# Statistical method - Z-score

from scipy import stats

z\_scores = np.abs(stats.zscore(boston\_df['MEDV']))

threshold = 3

outliers = np.where(z\_scores > threshold)

The visualization graphical representation shows the graphical structure of the given datasets and the statistical representation in tabular format by using scipy library which imports the packages from them.

1. Implement a strategy to handle outliers (e.g., removal and transformation).

# Strategy 1: Removal

boston\_df\_no\_outliers = boston\_df[(z\_scores < threshold)]

# Strategy 2: Transformation (log transformation)

boston\_df['MEDV\_log'] = np.log1p(boston\_df['MEDV'])

# Visualization after handling outliers

plt.figure(figsize=(12, 6))

# Before removal

plt.subplot(2, 2, 1)

sns.boxplot(x=boston\_df['MEDV'])

plt.title('Before Outlier Removal')

# After removal

plt.subplot(2, 2, 2)

sns.boxplot(x=boston\_df\_no\_outliers['MEDV'])

plt.title('After Outlier Removal')

# Before transformation

plt.subplot(2, 2, 3)

sns.boxplot(x=boston\_df['MEDV\_log'])

plt.title('Before Transformation')

# After transformation

plt.subplot(2, 2, 4)

sns.boxplot(x=boston\_df\_no\_outliers['MEDV\_log'])

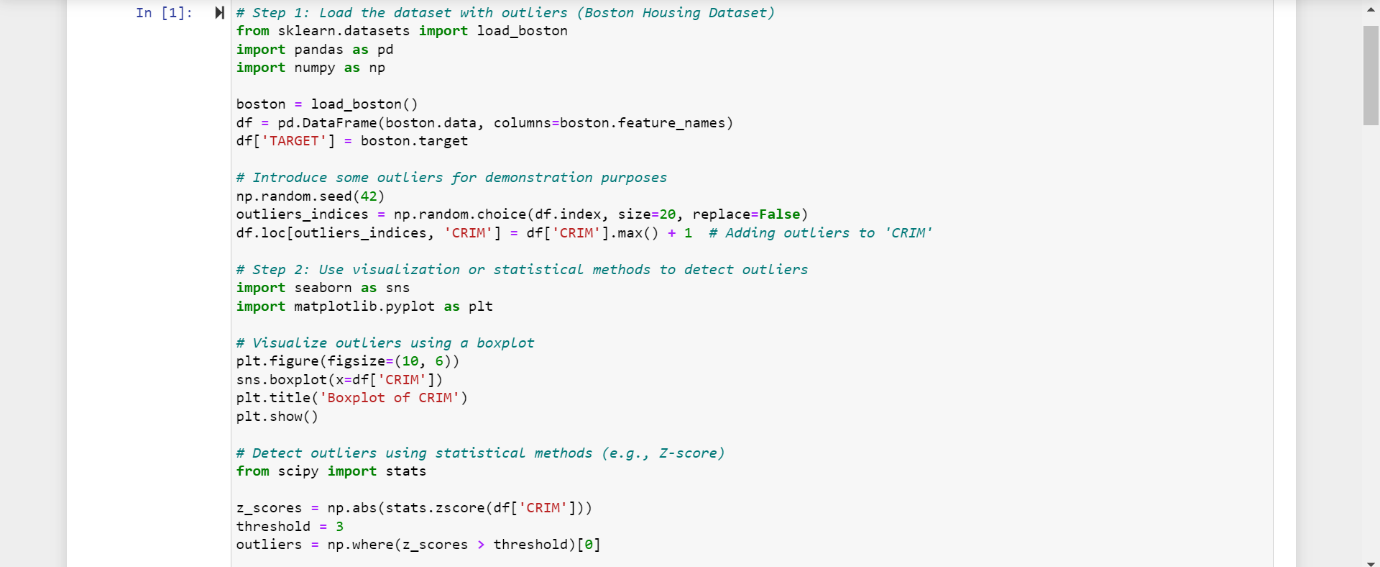
plt.title('After Transformation')

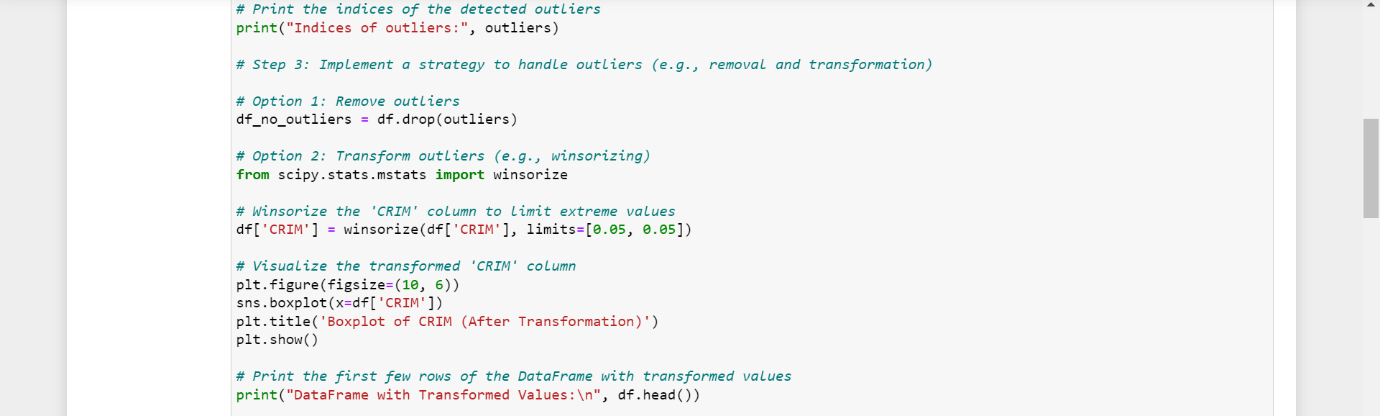
plt.tight\_layout()

plt.show()

This showcases the strategical overview of the transformation procedure to implement the handle outliers in order to show the transformation process using boston package in layout format

CODE:





OUTPUT:

