ASSIGNMENT-2

MACHINE LEARNING

NAME- CHETHAN Y

USN- 22BTRAD011

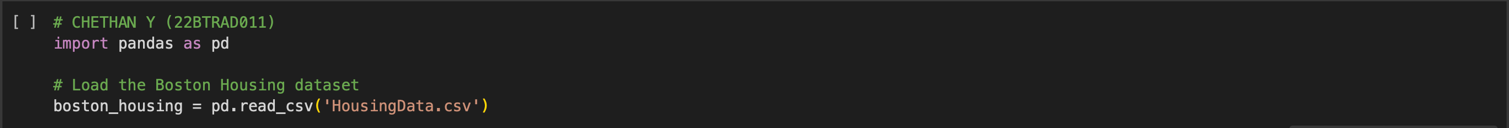
# QUESTION

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

CODE:

#CHETHAN Y (22BTRAD011)

import pandas as pd

# Load the Boston Housing dataset boston\_housing = pd.read\_csv('HousingData.csv') 

1. Use visualization or statistical methods to detect outliers.

CODE:

# CHETHAN Y (22BTRAD011)

import matplotlib.pyplot as plt

# Create box plots for each feature for col in boston\_housing.columns:

boston\_housing.boxplot(column=col) plt.title(col) plt.show() import matplotlib.pyplot as plt

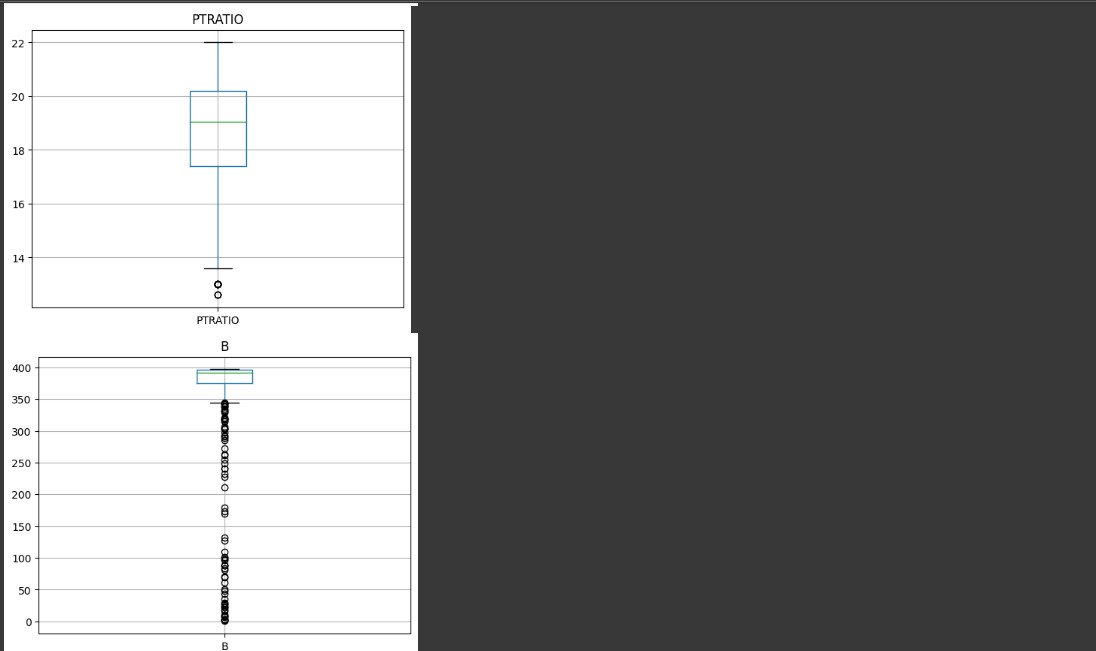
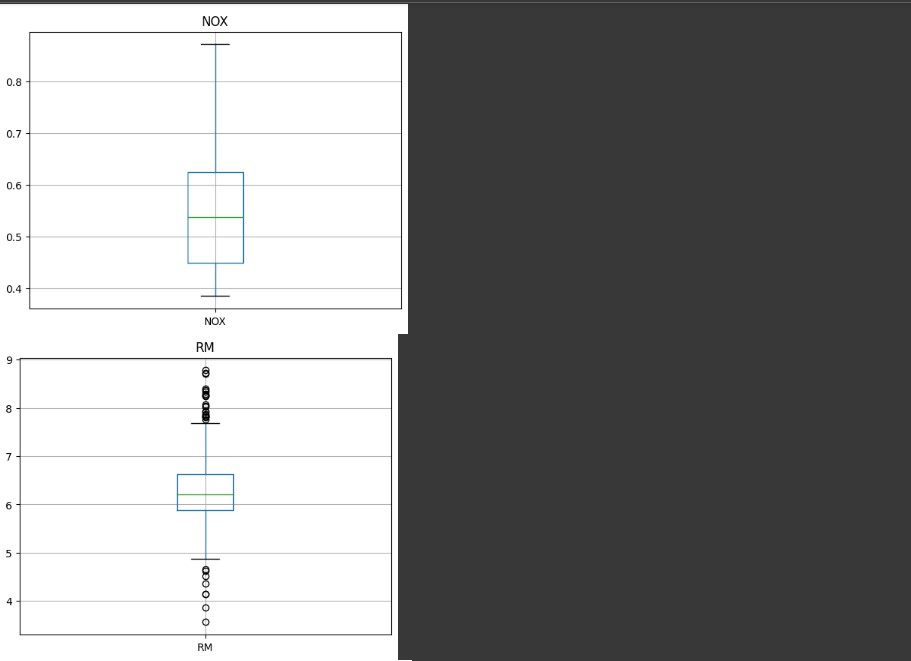
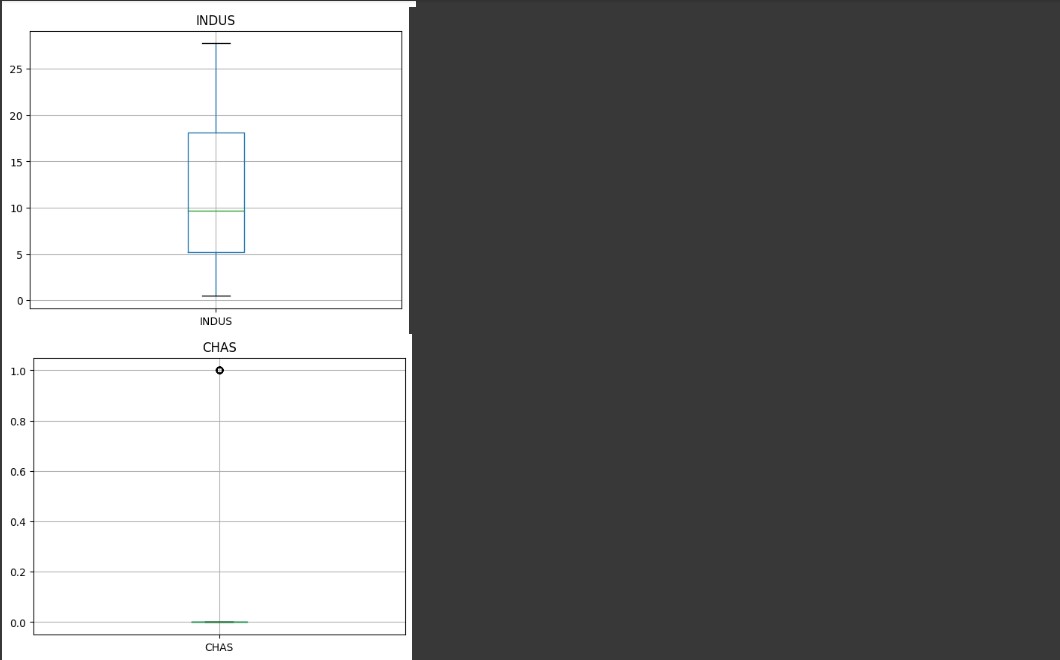
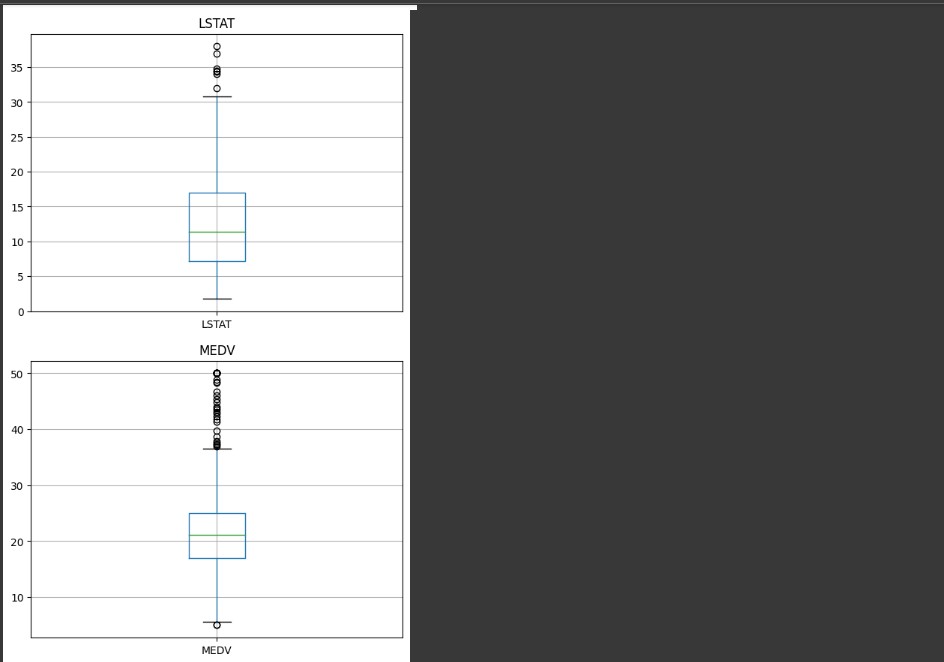
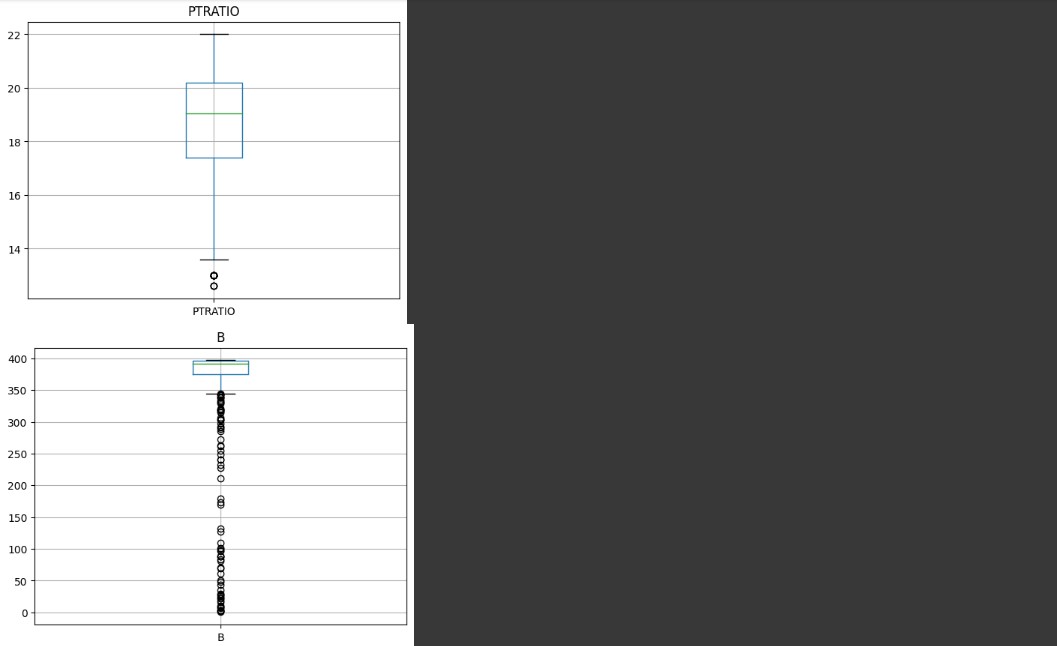
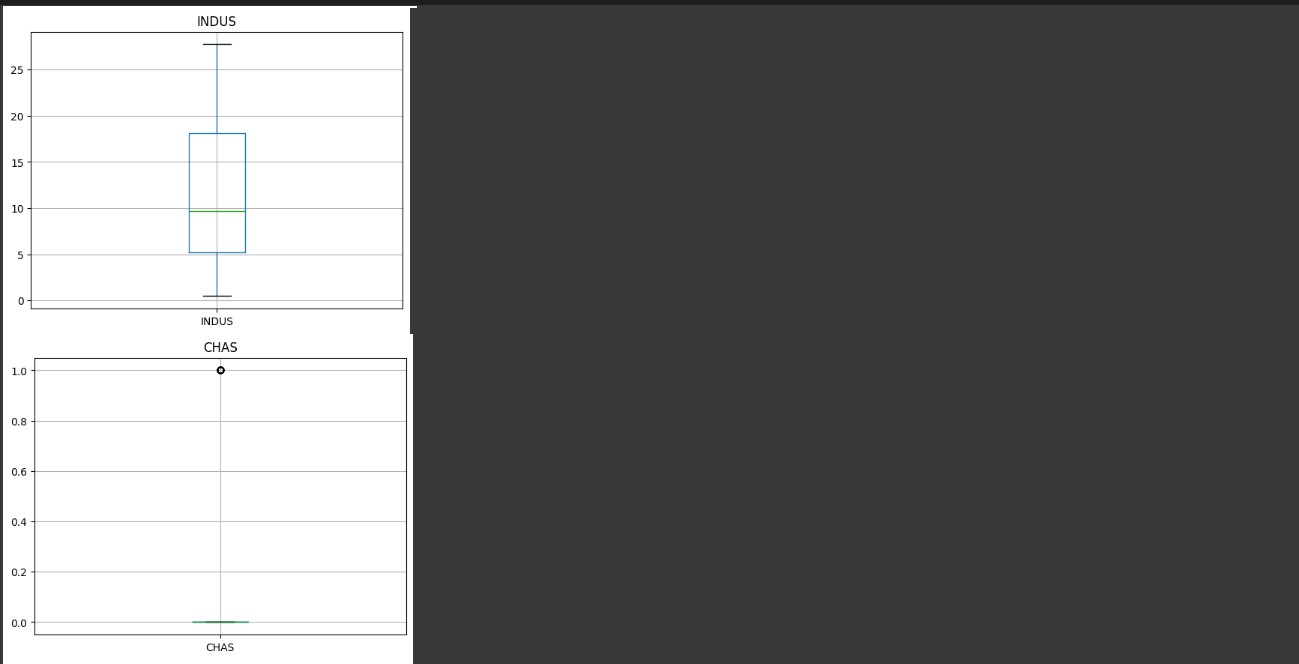
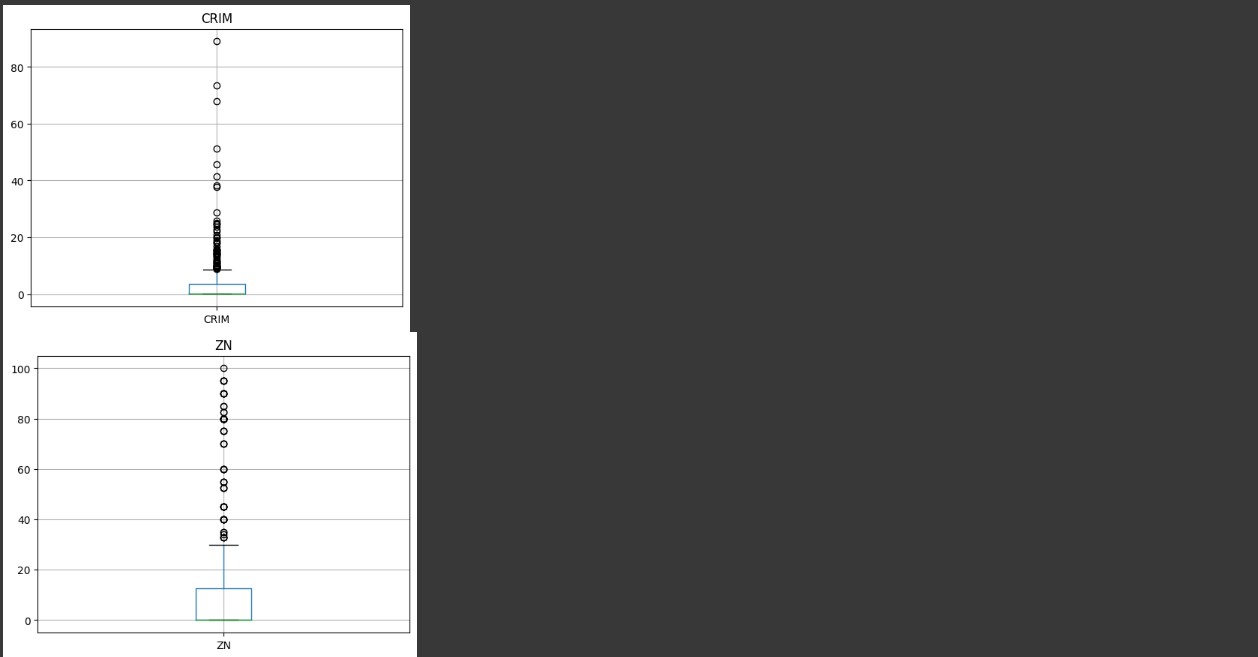
# Create box plots for each feature for col in boston\_housing.columns:

boston\_housing.boxplot(column=col)

plt.title(col) plt.show()



OUTPUT:



CODE:

# CHETHAN Y (22BTRAD011)

# Calculate the z-score for each feature for col in boston\_housing.columns:

z = (boston\_housing[col] - boston\_housing[col].mean()) / boston\_housing[col].std() outliers = boston\_housing[abs(z) > 3] print(f"Number of outliers for {col}: {len(outliers)}")

OUTPUT: 

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

CODE:

# CHETHAN Y (22BTRAD011)

import pandas as pd

# Load the Boston Housing dataset

boston\_housing = pd.read\_csv('HousingData.csv')

# Calculate the IQR for each feature for col in boston\_housing.columns:

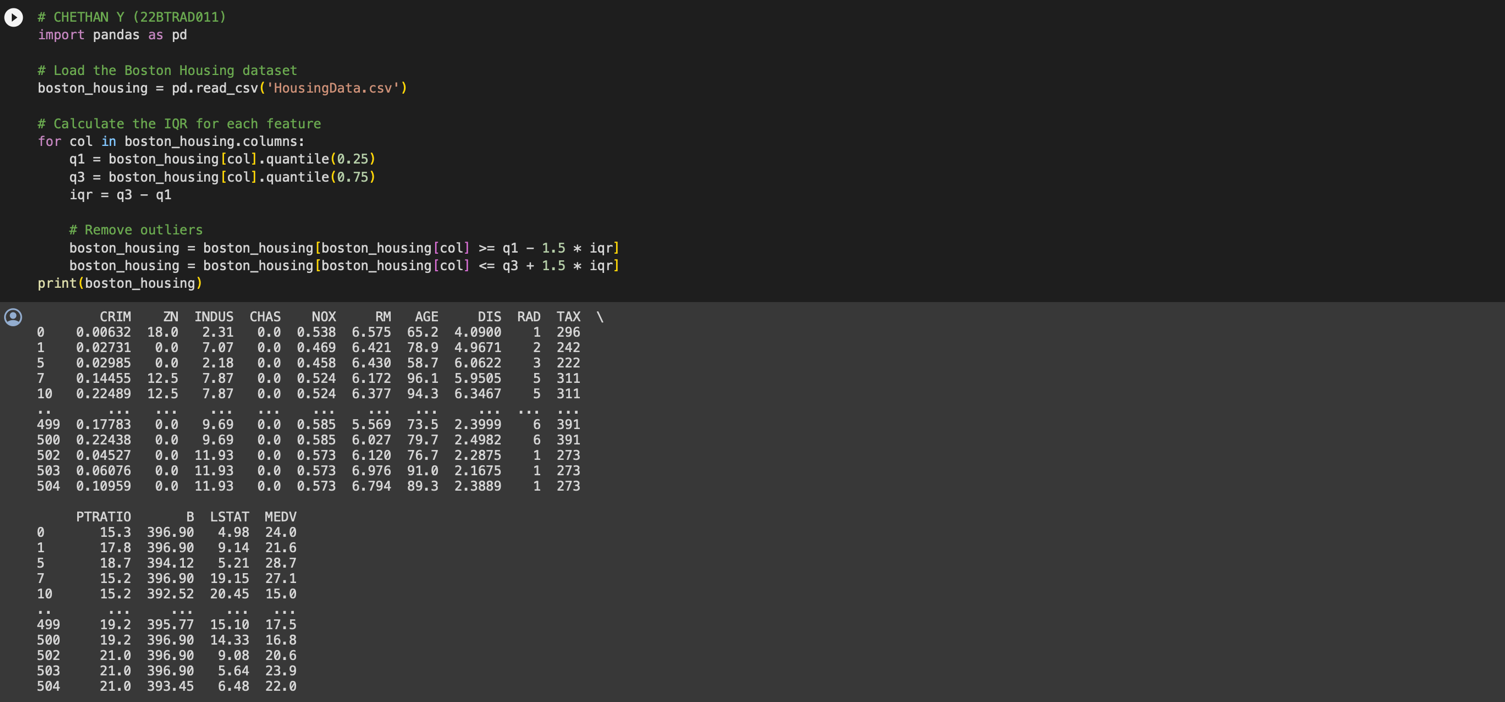
q1 = boston\_housing[col].quantile(0.25) q3 = boston\_housing[col].quantile(0.75) iqr = q3 - q1

# Remove outliers

boston\_housing = boston\_housing[boston\_housing[col] >= q1 - 1.5 \* iqr]

boston\_housing = boston\_housing[boston\_housing[col] <= q3 + 1.5 \* iqr] print(boston\_housing)

OUTPUT:



CODE:

# CHETHAN Y (22BTRAD011)

import pandas as pd import numpy as np

# Load the Boston Housing dataset

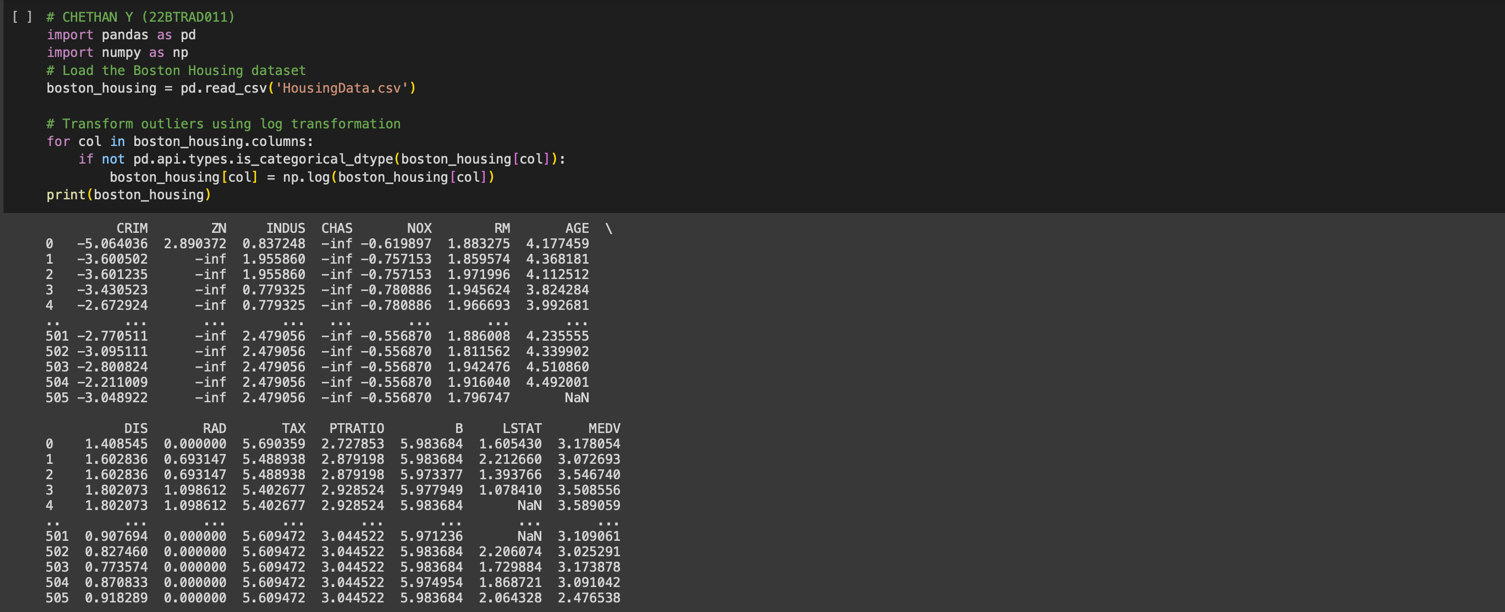
boston\_housing = pd.read\_csv('HousingData.csv') # Transform outliers using log transformation for col in boston\_housing.columns: if not pd.api.types.is\_categorical\_dtype(boston\_housing[col]):

boston\_housing[col] = np.log(boston\_housing[col])

**|** P a g e

print(boston\_housing)

OUTPUT:



GITHUB:

https://github.com/CHETHAN /MACHINE-LEARNING

**|** P a g e