ASSIGNMENT-1

MACHINE LEARNING

NAME- MUTTA DATTA SAI VISHNU MOHAN

USN- 22BTRAD026

QUESTION 1:

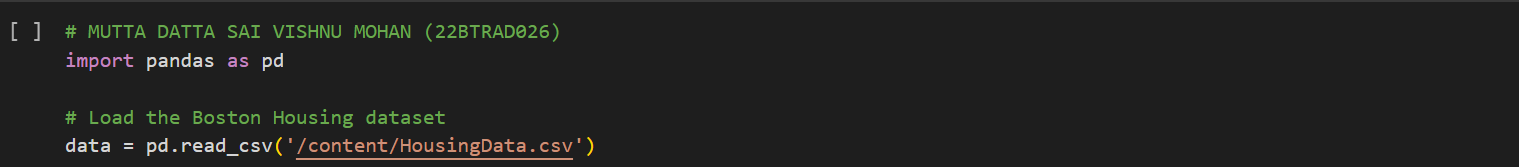
1. Load a dataset with missing values (Boston Housing Dataset).

# MUTTA DATTA SAI VISHNU MOHAN (22BTRAD026)

import pandas as pd

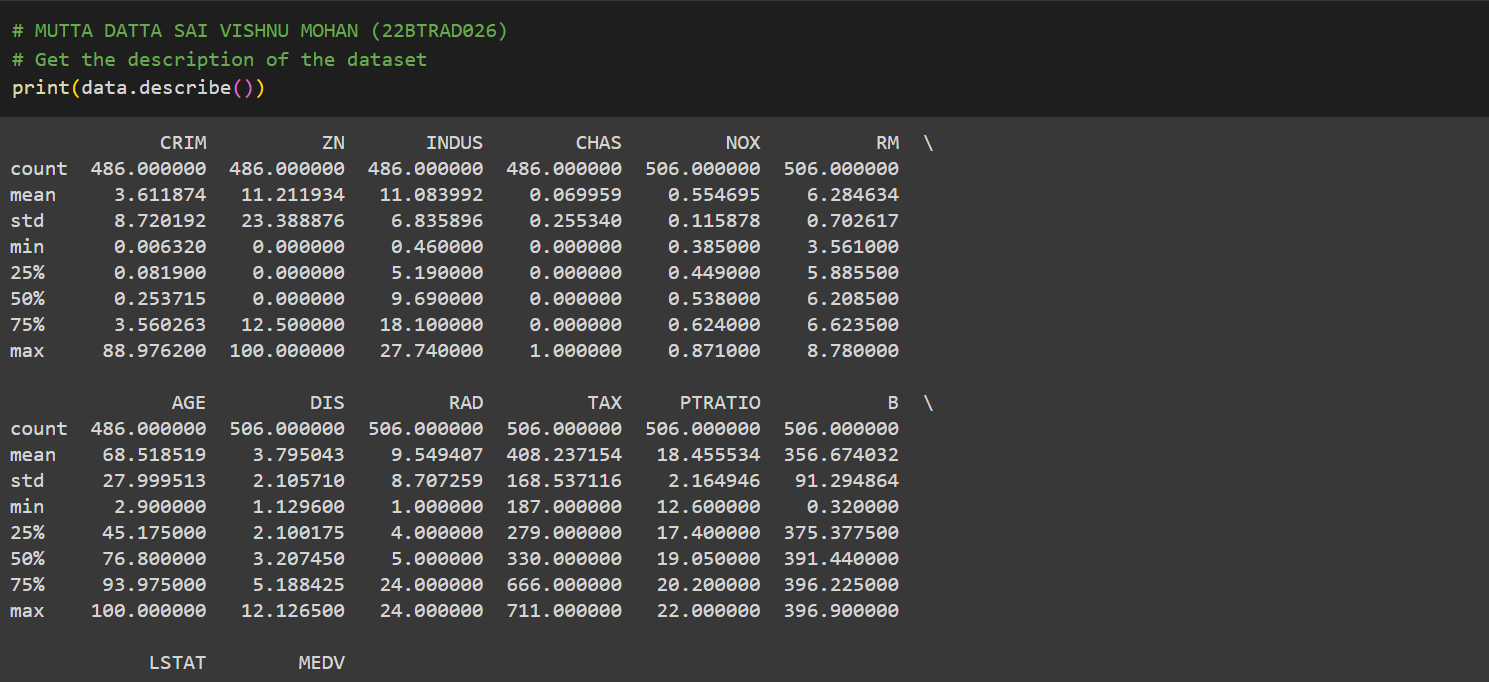
# Load the Boston Housing dataset

data = pd.read\_csv('/content/HousingData.csv')



1. Explore the description of the dataset.

print(data.describe())

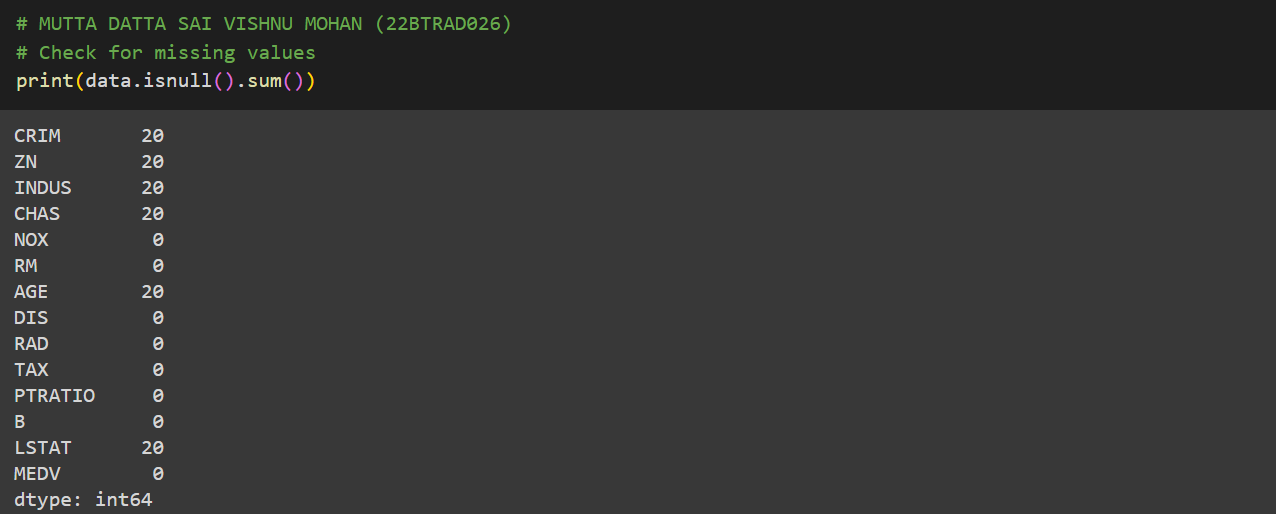


1. Identify the number of missing values corresponding to each feature.

# MUTTA DATTA SAI VISHNU MOHAN (22BTRAD026)

# Check for missing values

print(data.isnull().sum())



1. Explore and visualize the missing data patterns.

# MUTTA DATTA SAI VISHNU MOHAN (22BTRAD026)

import matplotlib.pyplot as plt

# Check for missing values and visualize them

for col in data.columns:

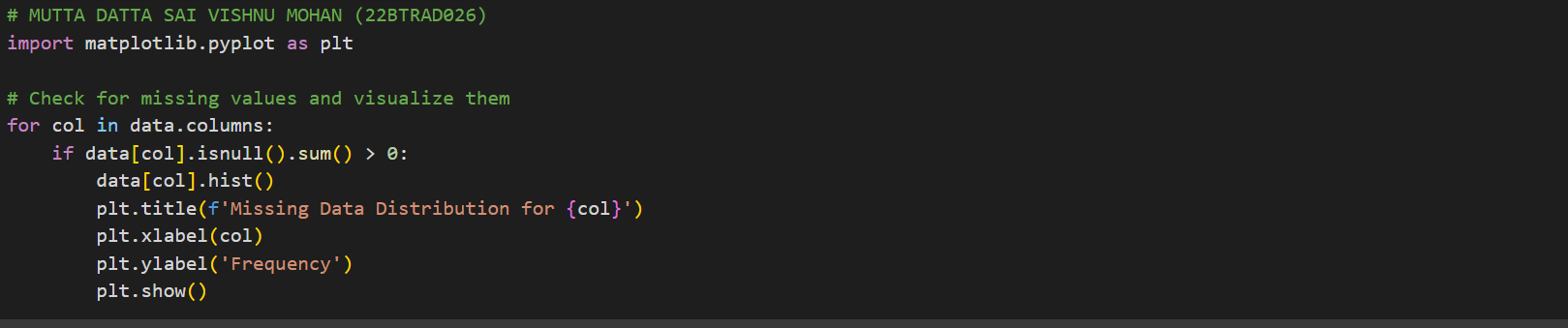
if data[col].isnull().sum() > 0:

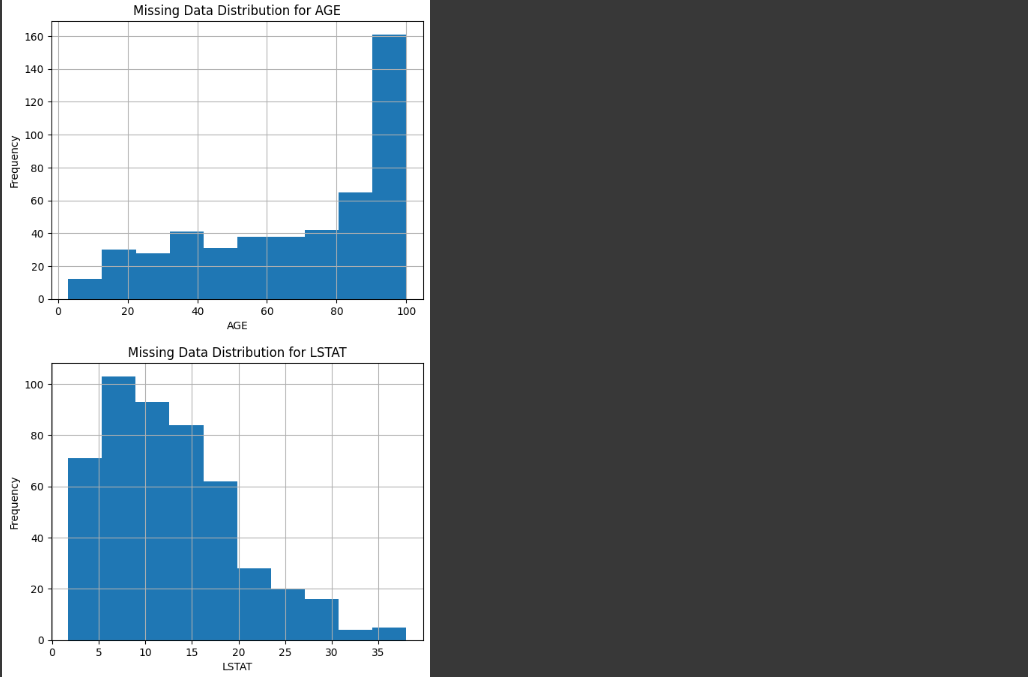
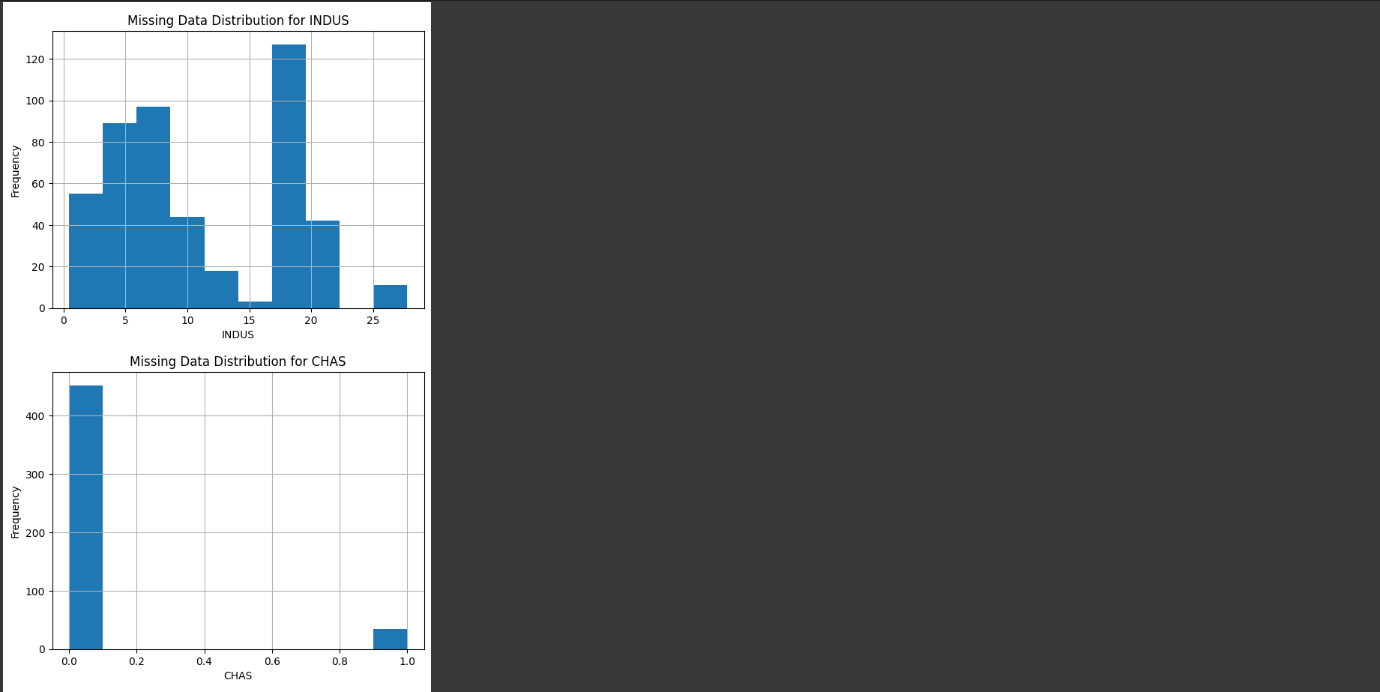
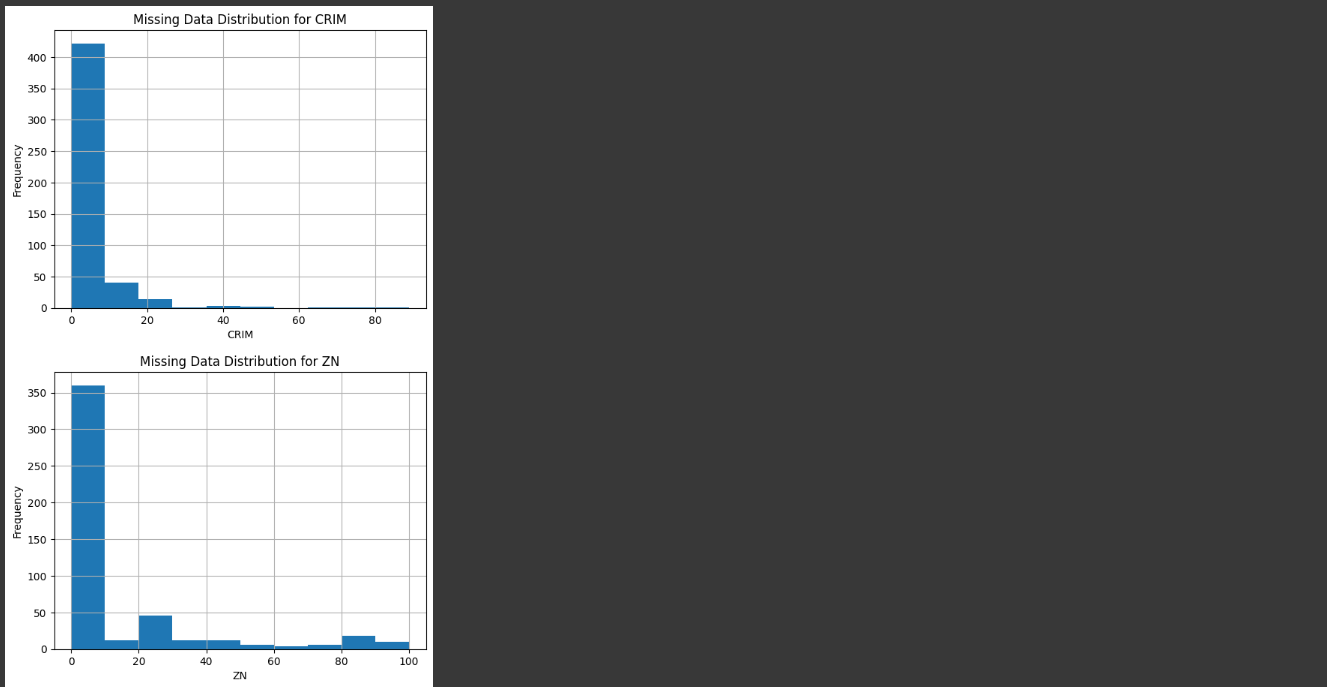
data[col].hist()

plt.title(f'Missing Data Distribution for {col}')

plt.xlabel(col)

plt.ylabel('Frequency')

plt.show()



1. Handle missing values using imputation method for a specific feature.

# MUTTA DATTA SAI VISHNU MOHAN (22BTRAD026)

from sklearn.impute import SimpleImputer

# Load the Boston Housing dataset

data = pd.read\_csv('HousingData.csv')

# Check for missing values

print(data.isnull().sum())

# Impute missing values using mean imputation for 'CHAS' feature

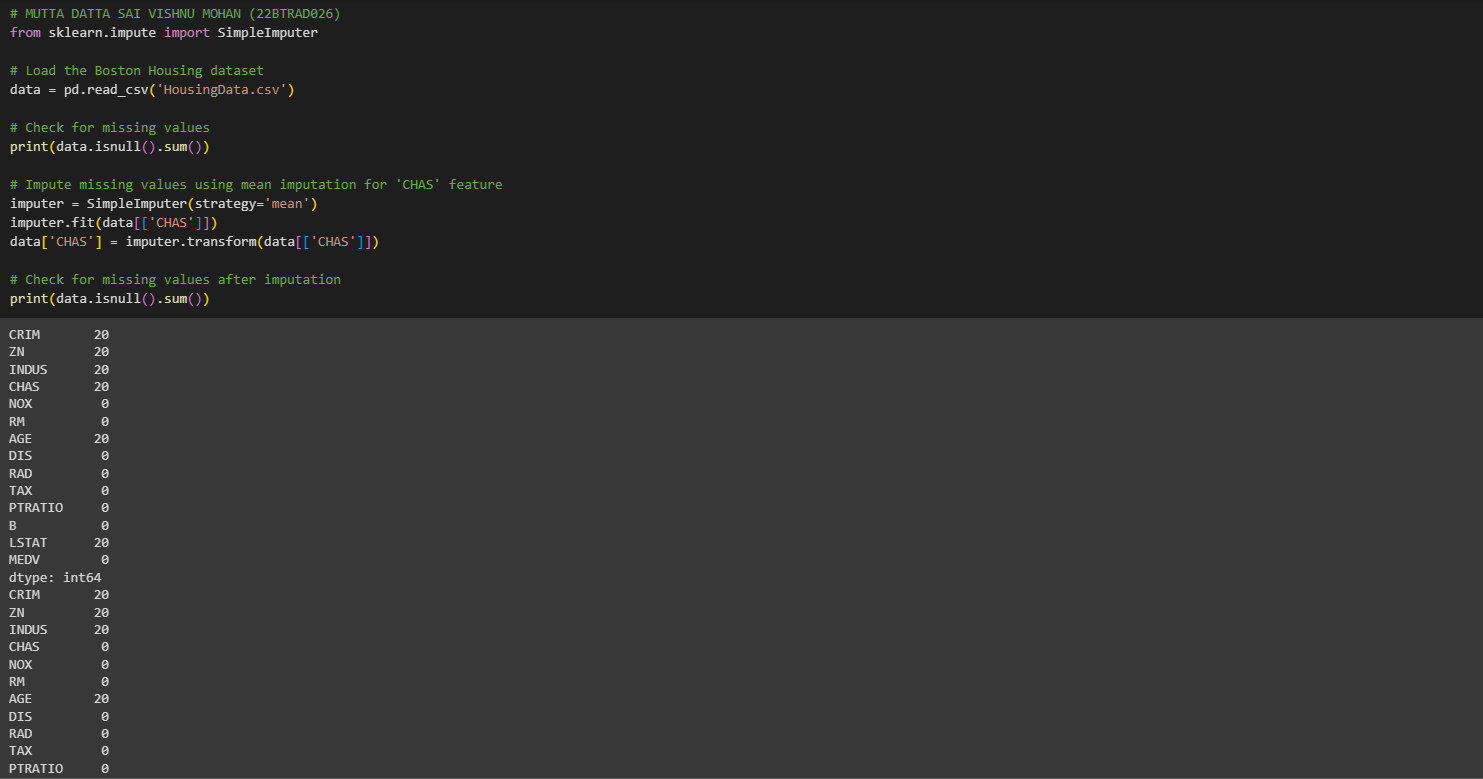
imputer = SimpleImputer(strategy='mean')

imputer.fit(data[['CHAS']])

data['CHAS'] = imputer.transform(data[['CHAS']])

# Check for missing values after imputation

print(data.isnull().sum())



1. Handle missing values using tuple removal method.

# MUTTA DATTA SAI VISHNU MOHAN (22BTRAD026)

import pandas as pd

# Load the Boston Housing dataset

data = pd.read\_csv('HousingData.csv')

# Drop rows with missing values

data\_no\_missing = data.dropna()

# Show the output

print(data\_no\_missing)

