name: **siddhesh shende**

div:A **rollno: 59** batch: A3

sub: ML

**assignment no: 3**

**code: linear regression**

import numpy as np

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error

# Load data skipping the first row if it contains headers

column\_names = ['longitude', 'latitude', 'housing\_median\_age', 'total\_rooms',

'total\_bedrooms', 'population', 'households', 'median\_income',

'median\_house\_value']

bos1 = pd.read\_csv('california\_housing\_test.csv', delimiter=',', skiprows=1, names=column\_names)

path = pd.read\_csv('california\_housing\_test.csv')

print(path)

# Display first few rows to verify data loading

print(bos1.head(10))

# Preprocessing: Handling NaN values if present

bos1.isna().sum()

# Splitting data

X = bos1.drop(columns=['median\_house\_value']) # Features

Y = bos1['median\_house\_value'] # Target variable

x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.30, random\_state=5)

# Using Linear Regression Model

lr = LinearRegression()

lr.fit(x\_train, y\_train)

pred\_lr = lr.predict(x\_test)

# Model Evaluation

mse\_lr = mean\_squared\_error(y\_test, pred\_lr, squared=False)

print("Root Mean Squared Error for Linear Regression = {:.2f}".format(mse\_lr))

**output:**

