# Python code to illustrate

# regression using data set

import matplotlib

matplotlib.use('GTKAgg')

import matplotlib.pyplot as plt

import numpy as np

from sklearn import datasets, linear\_model

import pandas as pd

# Load CSV and columns

df = pd.read\_csv("Housing.csv")

Y = df['price']

X = df['lotsize']

X=X.values.reshape(len(X),1)

Y=Y.values.reshape(len(Y),1)

# Split the data into training/testing sets

X\_train = X[:-250]

X\_test = X[-250:]

# Split the targets into training/testing sets

Y\_train = Y[:-250]

Y\_test = Y[-250:]

# Plot outputs

plt.scatter(X\_test, Y\_test, color='black')

plt.title('Test Data')

plt.xlabel('Size')

plt.ylabel('Price')

plt.xticks(())

plt.yticks(())

# Create linear regression object

regr = linear\_model.LinearRegression()

# Train the model using the training sets

regr.fit(X\_train, Y\_train)

# Plot outputs

plt.plot(X\_test, regr.predict(X\_test), color='red',linewidth=3)

plt.show()