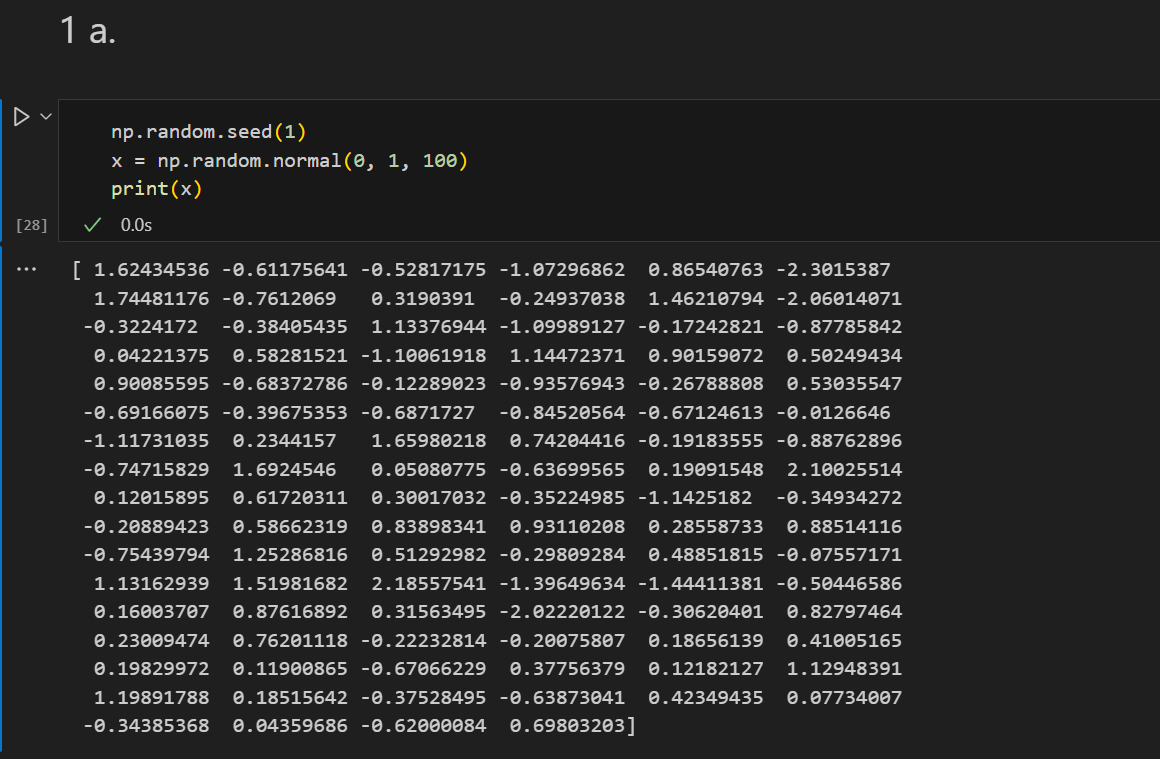
**Assignment II(Ex.No 13,15)**

1. In this exercise you will create some simulated data and will fit simple linear regression models to it. Make sure to use set.seed(1) prior to starting part (a) to ensure consistent results.rnorm

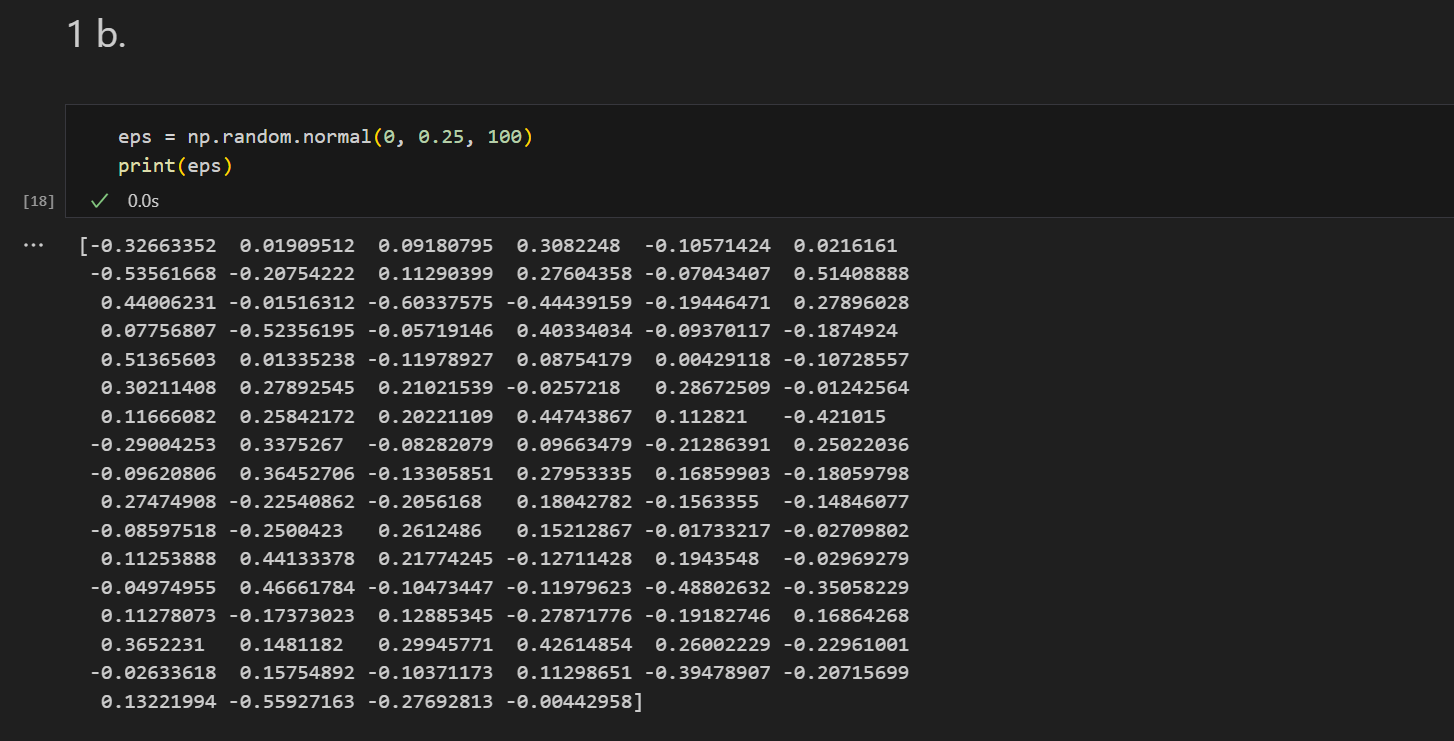
1. Using the rnorm() function, create a vector, “x”, containing 100 observations drawn from a N(0,1) distribution. This represents a feature, X.

**Answer:**

****

1. Using the rnorm() function, create a vector, “eps”, containing 100 observations drawn from a N(0,0.25) distribution.

**Answer:**

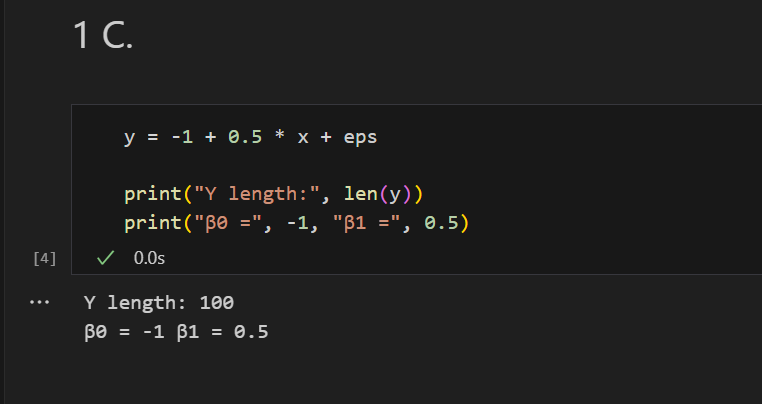
****

1. Using “x” and “eps”, generate a vector “y” according to the model

Y=−1+0.5X+ε.

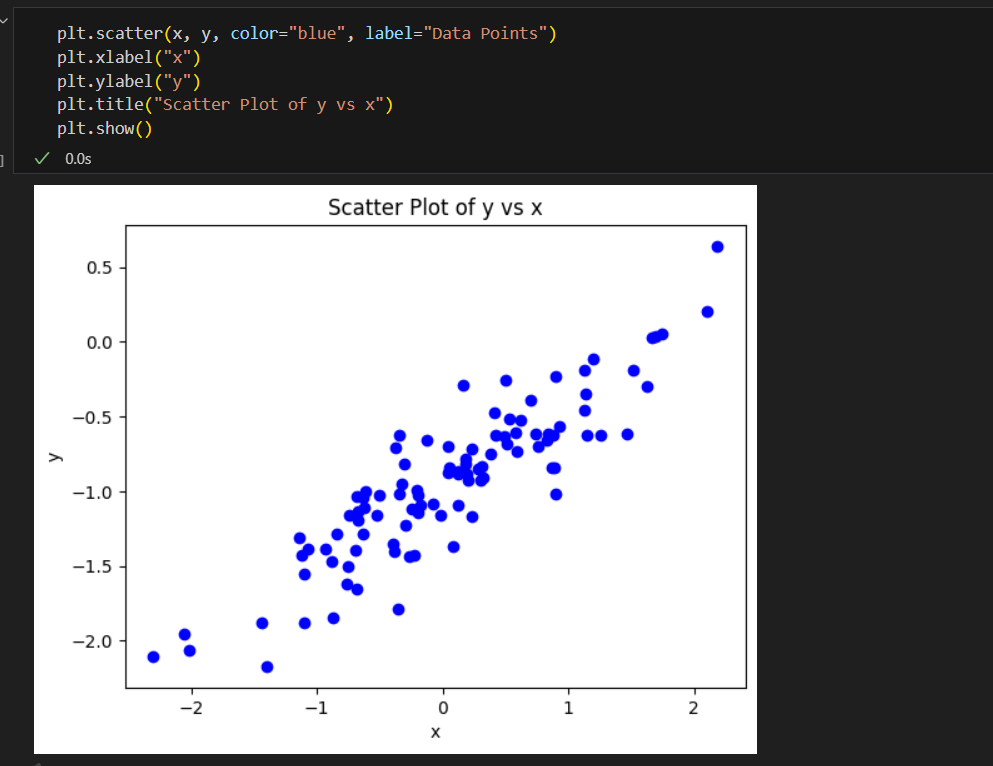
What is the length of the vector “y” ? What are the values of β0 and β1 in this linear model ?

**Answer:**

`

1. Create a scatterplot displaying the relationship between “x” and “y”. Comment on what you observe.

**Answer:**

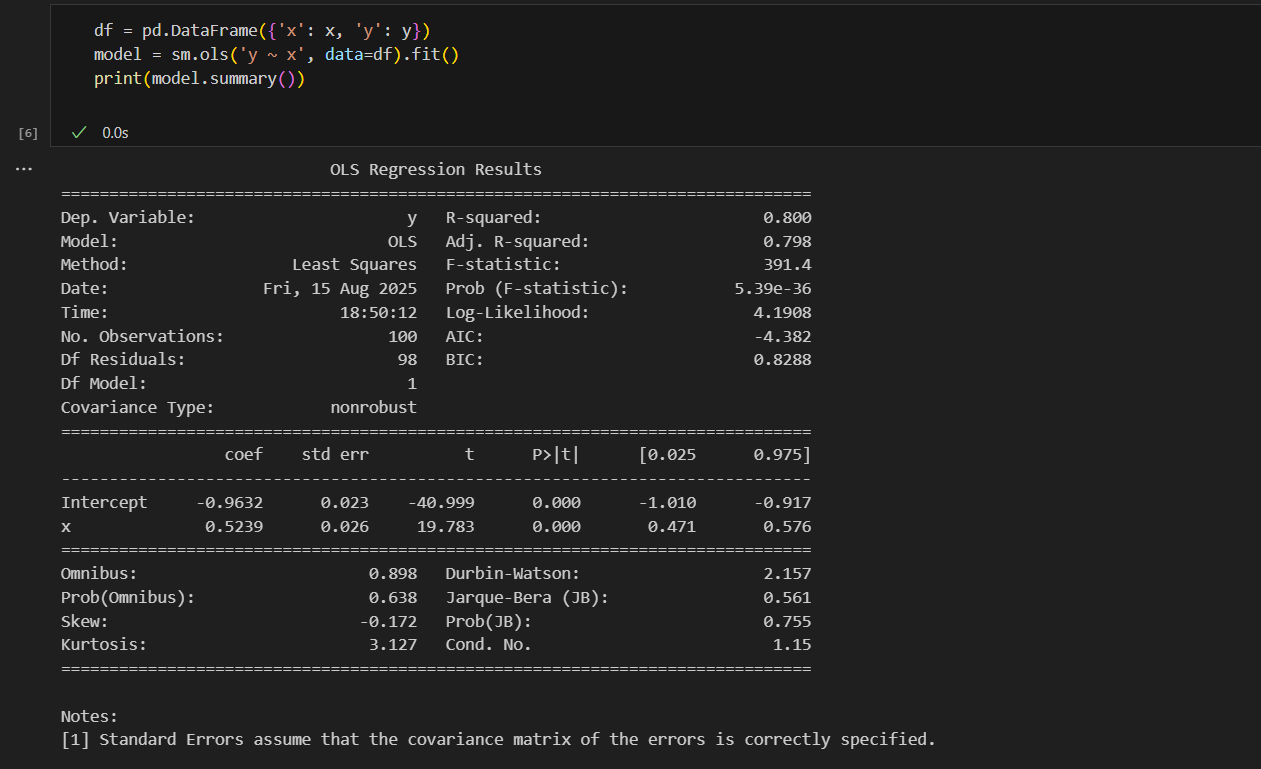


**Observation:**

* **Positive Correlation** – As the values of x increase, the values of y also tend to increase. The points are clustered along an upward-sloping trend.
* **Moderate to Strong Relationship** – The points are relatively close to an imaginary straight line, suggesting a moderately strong linear relationship.
* **Some Variability** – While most points follow the upward trend, there is some scatter around the trend line, indicating that x explains much, but not all, of the variation in y.
* **No Major Outliers** – There are no extreme points far away from the overall trend, meaning there aren’t strong anomalies affecting the relationship.
* **Symmetrical Spread** – The scatter appears fairly evenly distributed above and below the trend line, indicating no significant skew.

1. Fit a least squares linear model to predict “y” using “x”. Comment on the model obtained. How do β^0 and β^1 compare to β0 and β1 ?

**Answer:**

****

**Observation:**

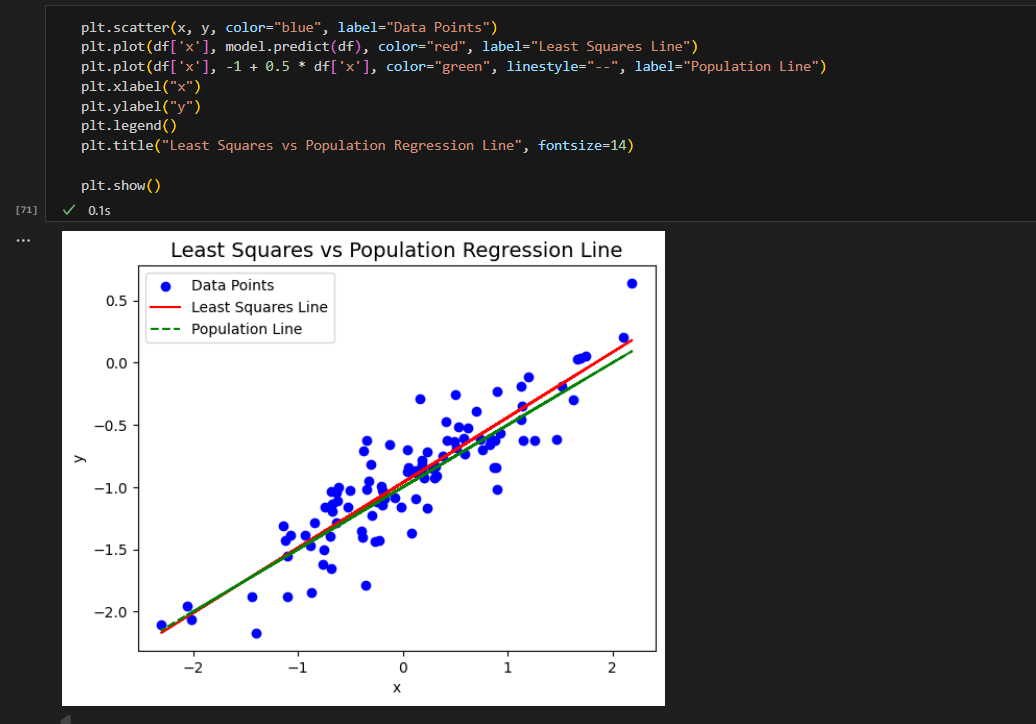
The obtained model shows a strong linear relationship between x and y, explaining 80% of the variation in the dependent variable. Both the intercept and slope coefficients are highly statistically significant (p < 0.001). The residuals pass normality tests (Jarque-Bera p = 0.755) and there’s no indication of autocorrelation (Durbin-Watson = 2.157). Overall, the model is reliable, with well-behaved residuals and a high goodness-of-fit.

**Comparison between Estimated (β̂⁰, β̂¹) and True (β₀, β₁) Coefficients**

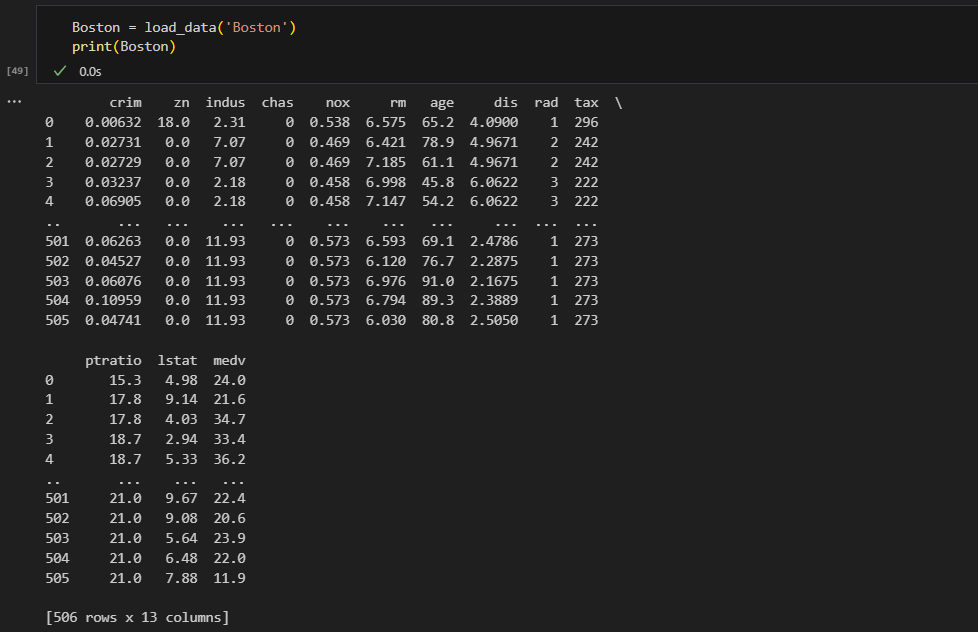
* **β̂⁰ (estimated intercept)** = -0.9632 is very close to the true β₀ (if the data was generated with β₀ ≈ -1, this matches well).
* **β̂¹ (estimated slope)** = 0.5239 is very close to the true β₁ (if the data was generated with β₁ ≈ 0.5, this is a good estimate).
* The small difference is due to **sampling variation** and **random error** in the dataset.

1. Display the least squares line on the scatterplot obtained in (d). Draw the population regression line on the plot, in a different color. Use the legend() function to create an appropriate legend.

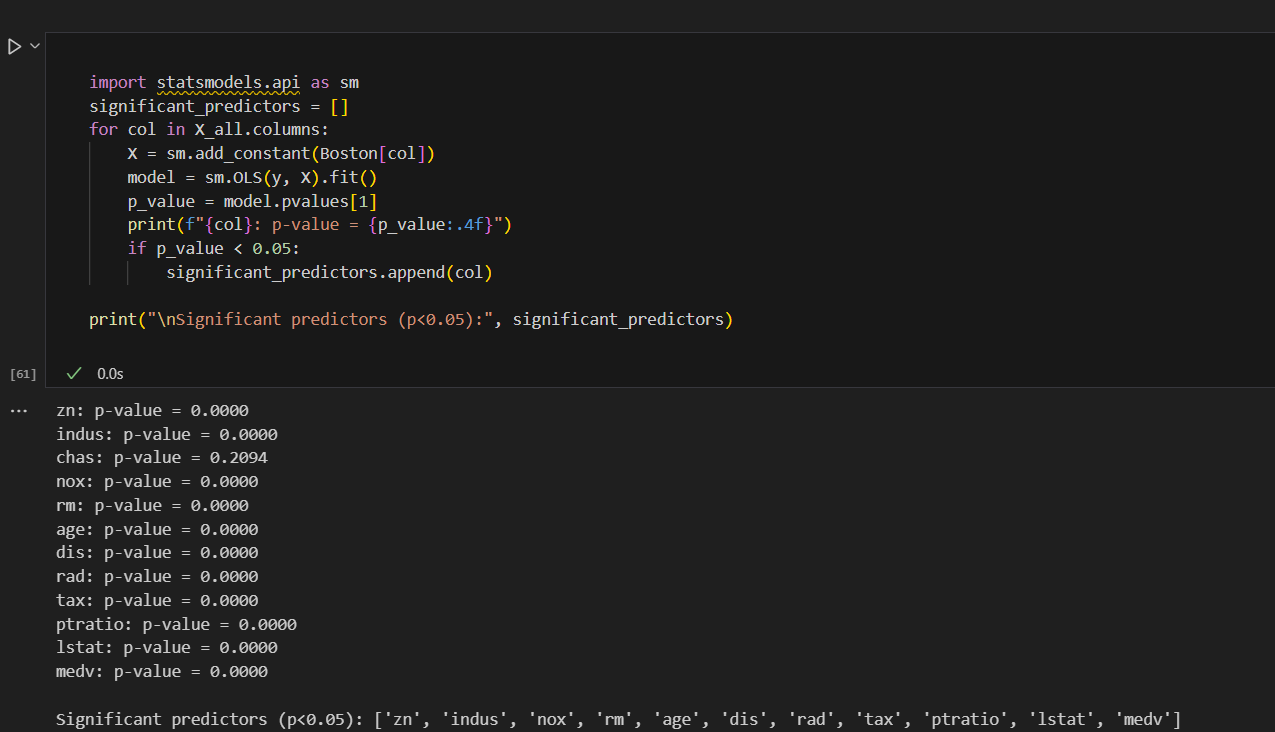
**Answer:**



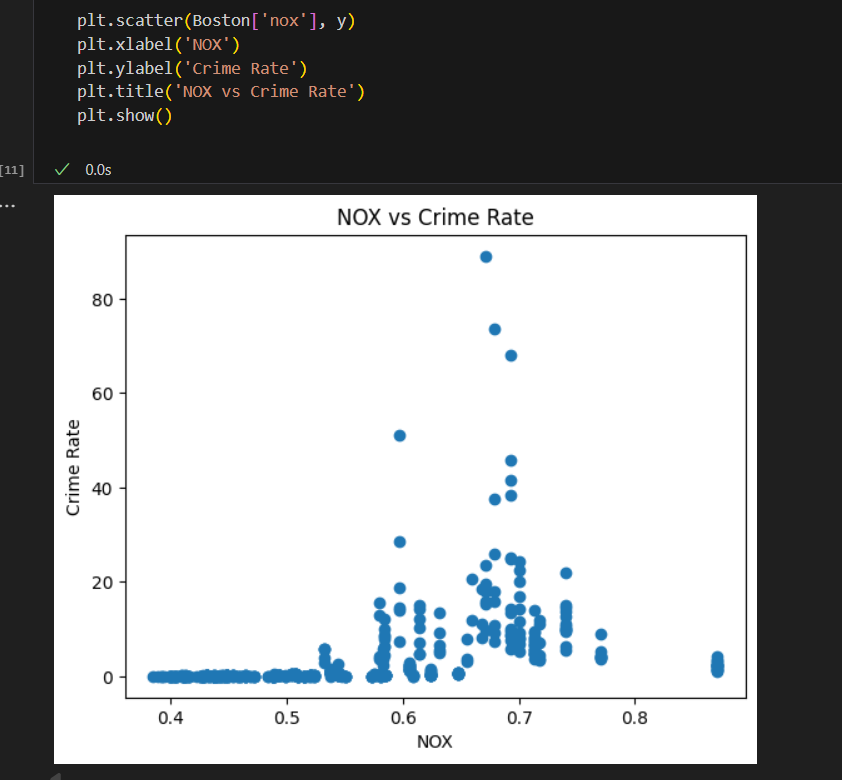
**2.** This problem involves the “Boston” data set, which we saw in the lab for this chapter. We will now try to predict per capita crime rate using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors.

****

1. For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response ? Create some plots to back up your assertions.

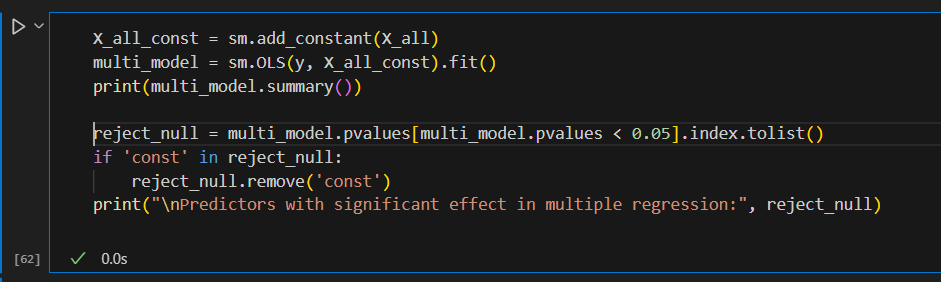
**Answer: **

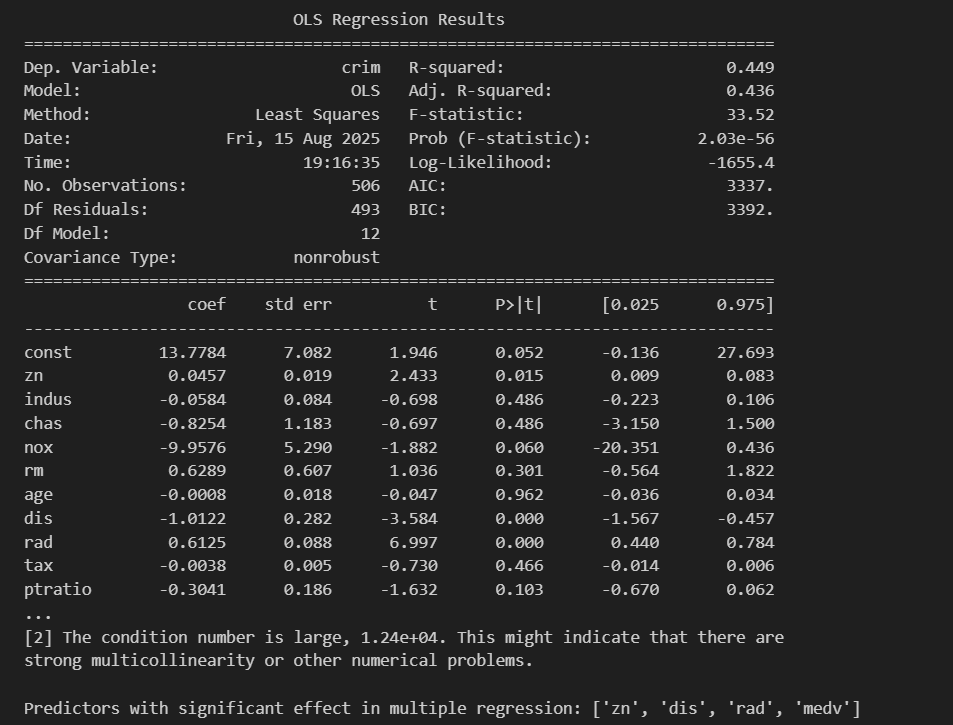
**Scatter Plot:**

****

1. Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis H0:βj=0 ?

**Answer:**

****

****

**By**

**KAVINPRASATH.S.M**

**1P24CS011**