HW 1
Swayam Mehta - 801207470

Github: https://github.com/Swayyum/Intro-to-ML--4105/blob/main/HW1.ipynb

Part 1:

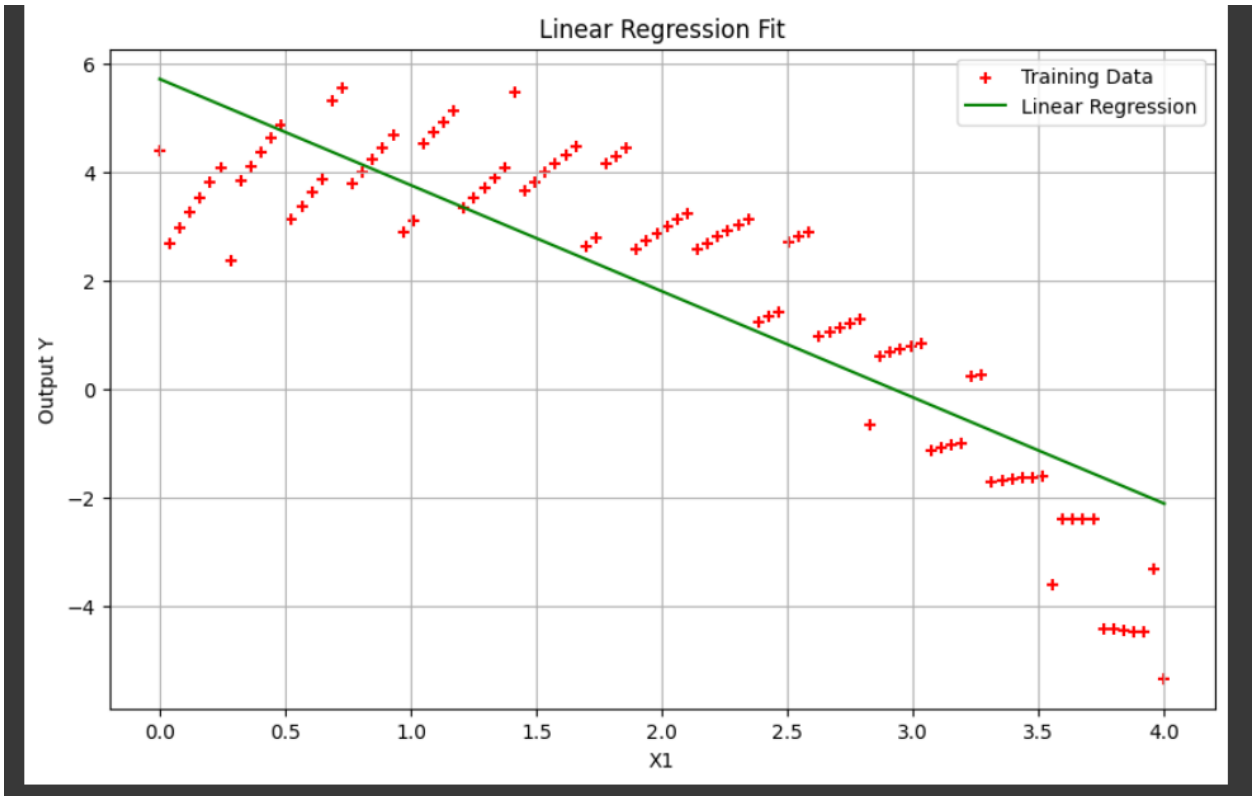
1. Report the linear model you found for each explanatory variable.

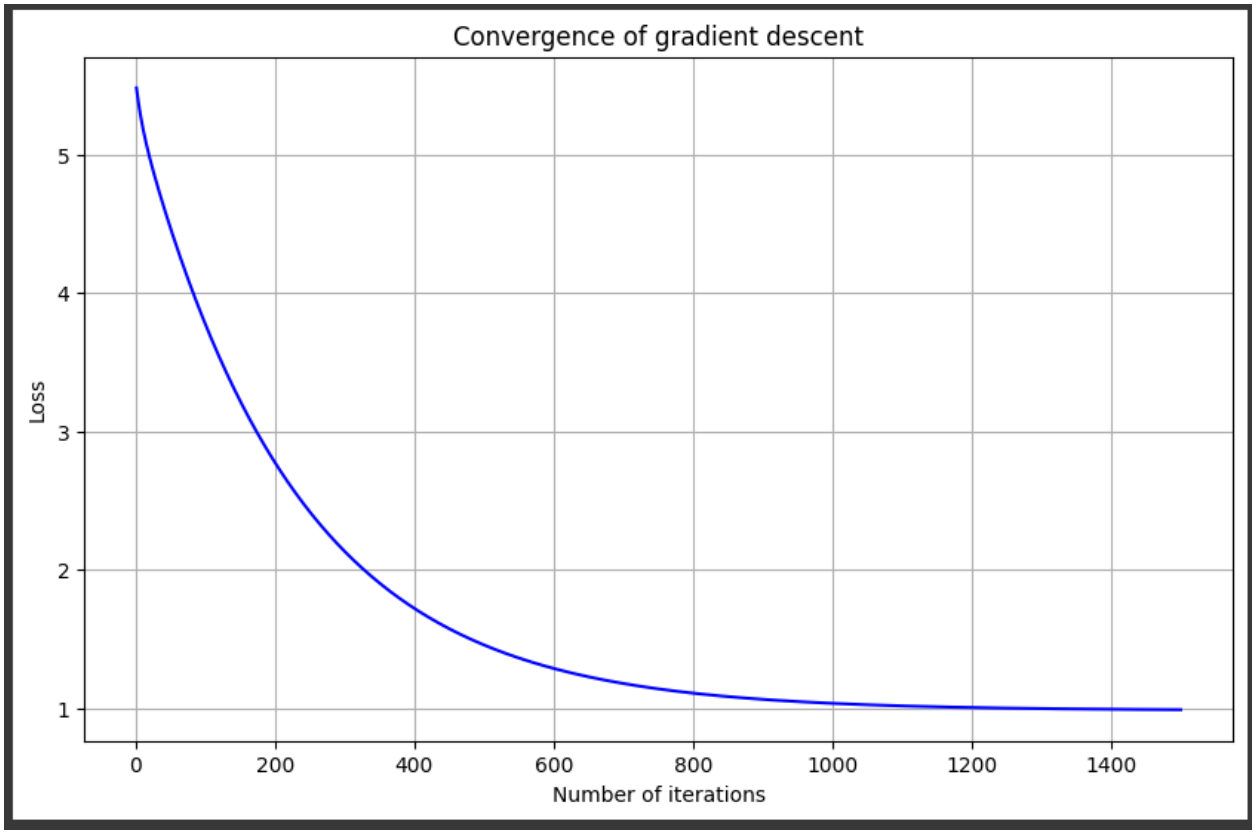
For X1: [5.48226715 5.44290965 5.40604087 ... 0.99063932 0.99061433 0.99058944]
For X2: [5.29831663 5.09909109 4.92356115 ... 3.5993997 3.59939955 3.5993994]
For X3: [5.40768785 5.30397076 5.21178297 ... 3.63053597 3.6305311 3.63052625]

2. Plot the final regression model and loss over the iteration per each explanatory variable.

X1

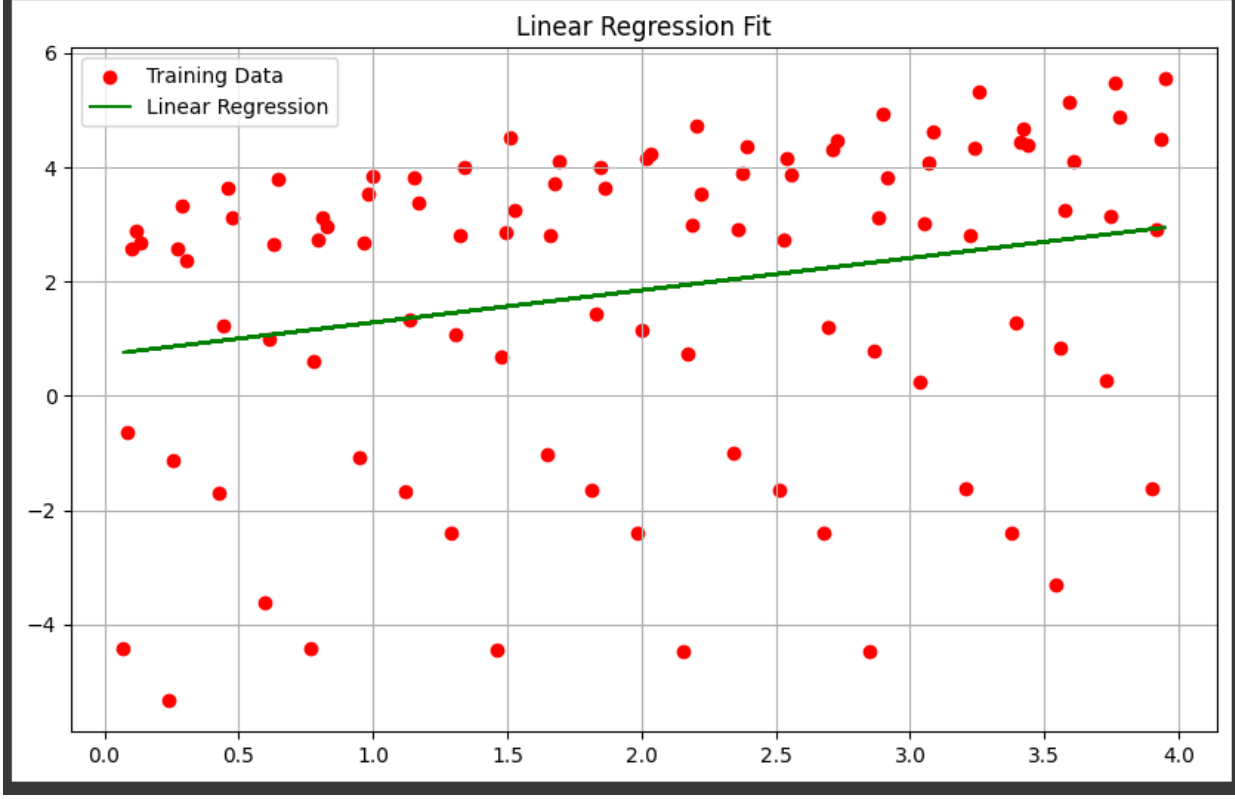
```
The cost for given values of theta_0 and theta_1 = 5.524438459196242
Final value of theta = [ 5.71850653 -1.9568206 ] cost_history =
[5.48226715 5.44290965 5.40604087 ... 0.99063932 0.99061433 0.99058944]
```

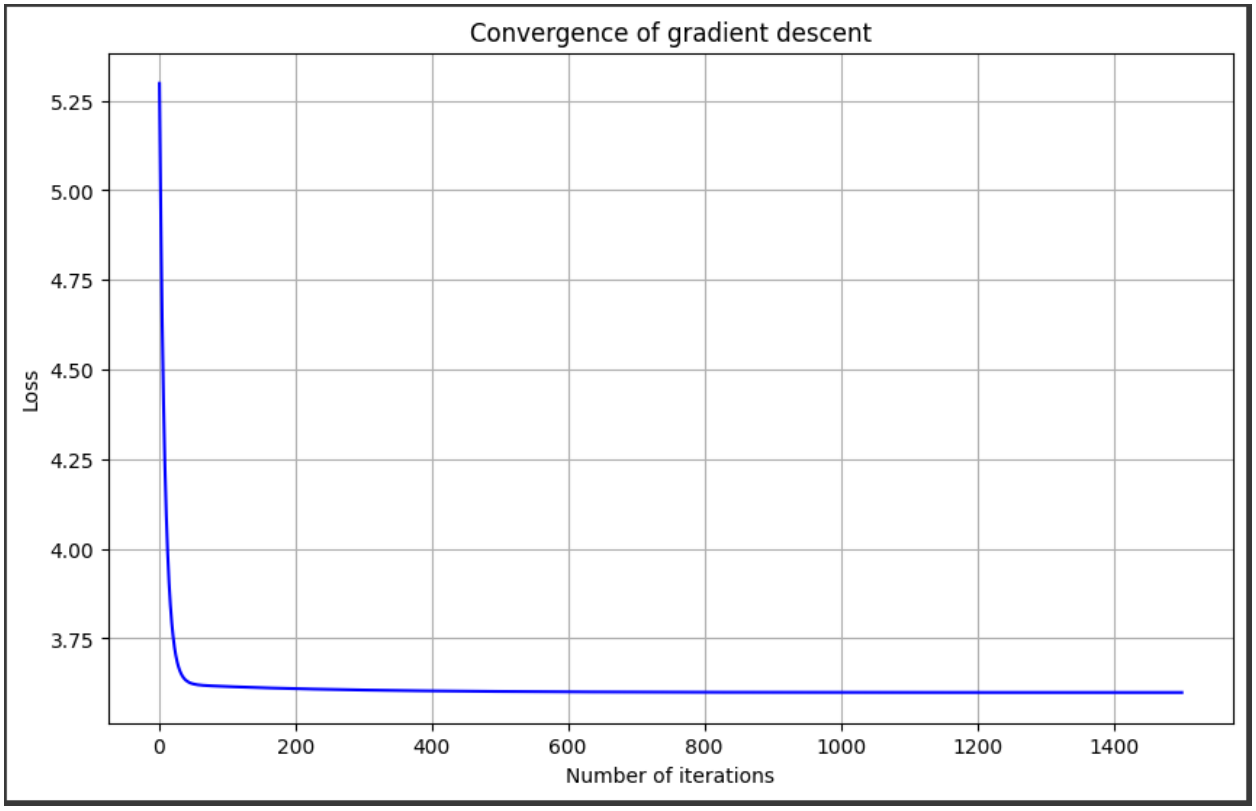



Convergence of gradient descent
Loss
5
4
3
2
1
0
200
400
600
800
1000
1200
1400
Number of iterations

X2

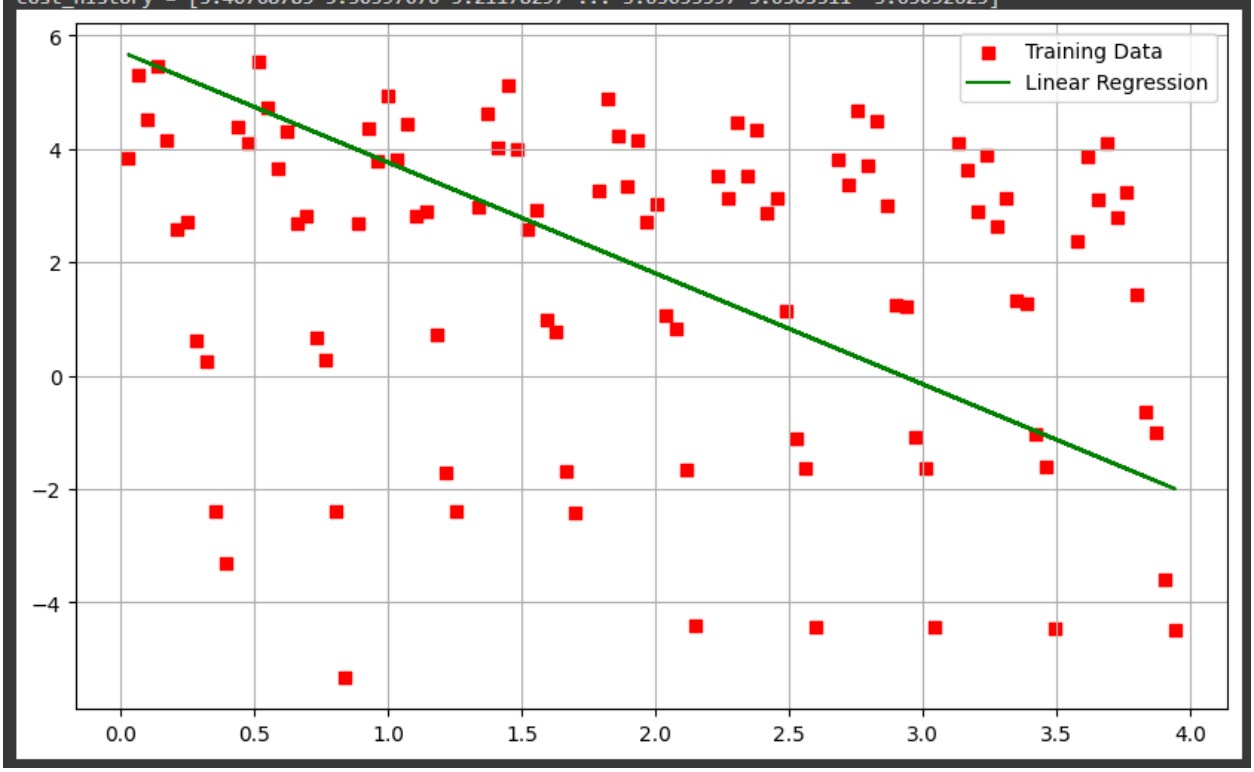
```
The cost for given values of theta_0 and theta_2 = 5.524438459196242
Final value of theta = [0.71988473 0.56390334]
cost_history = [5.29831663 5.09909109 4.92356115 ... 3.5993997  3.59939955 3.5993994 ]
```

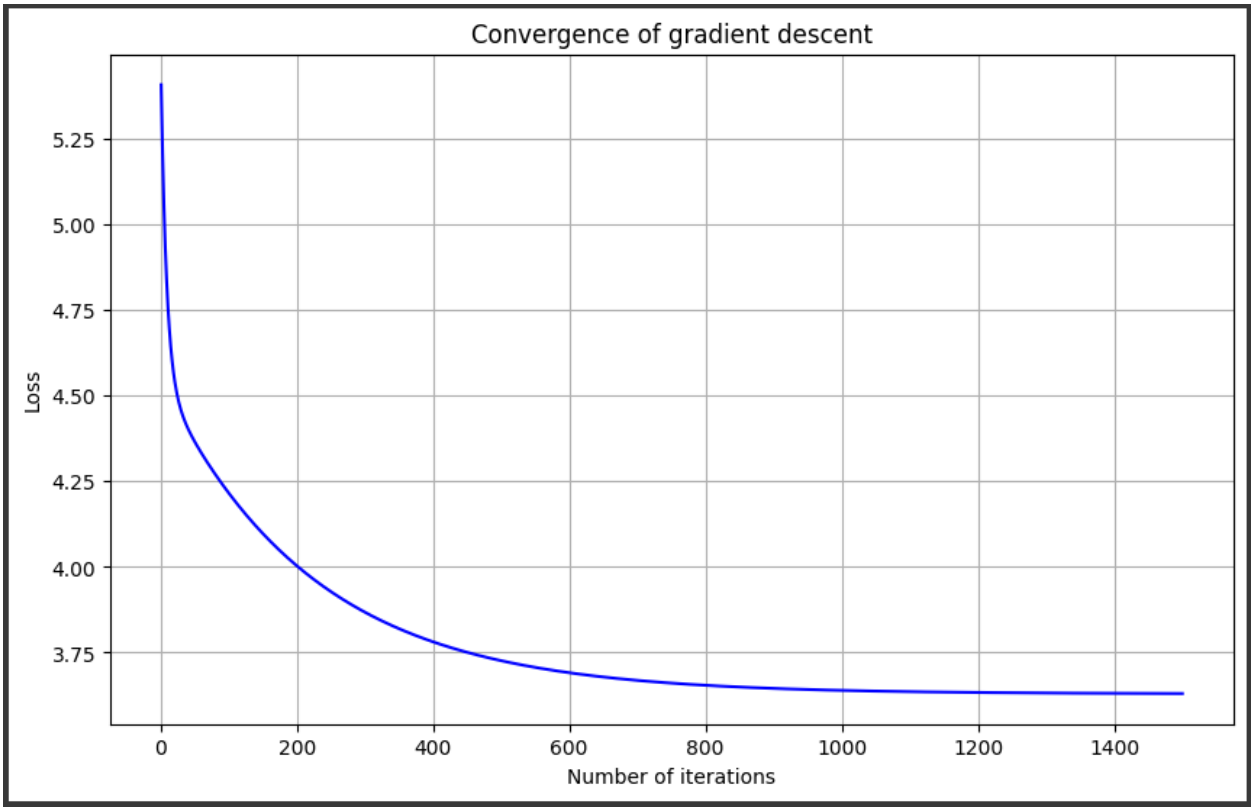



Convergence of gradient descent
Loss
5.25
5.00
4.75
4.50
4.25
4.00
3.75
0
200
400
600
800
1000
1200
1400
Number of iterations

X3

```
The cost for given values of theta_0 and theta_3 = 5.524438459196242
Final value of theta = [ 2.78048129 -0.48451631]
cost_history = [5.40768785 5.30397076 5.21178297 ... 3.63053597 3.6305311  3.63052625]
```





3. Which explanatory variable has the lower loss (cost) for explaining the output (Y)?

The lowest cost among these variables is associated with X1, which has a cost of approximately 0.99058944. Therefore, for explaining the output (Y), X1 has the lowest loss (cost) and provides the best fit according to the linear regression model.

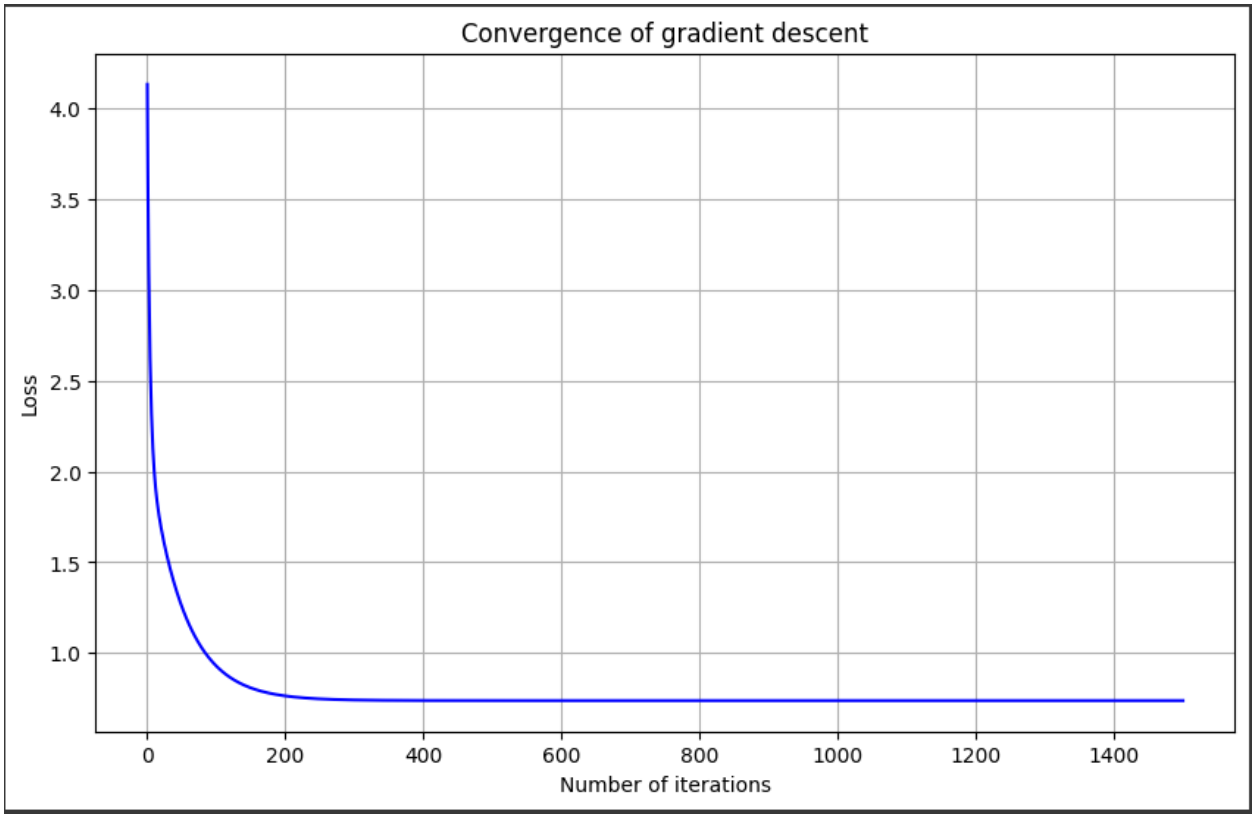
4. Based on your training observations, describe the impact of the different learning rates on the final loss and number of training iterations.

Learning Rate = 0.1
Final Loss: With a learning rate of 0.1, the final loss converged to relatively low values for all explanatory variables, as seen in the cost values provided earlier. It reached convergence after a relatively small number of iterations.
Number of Training Iterations: The algorithm reached convergence quickly, typically within a few hundred iterations.

Learning Rate = 0.01
Final Loss: With a learning rate of 0.01, the final loss also converged to relatively low values for all explanatory variables, though it may take more iterations compared to a learning rate of 0.1.
Number of Training Iterations: The algorithm required a larger number of iterations to converge compared to a learning rate of 0.1. The convergence process was slower but still effective.

Part 2:

1. Report the final linear model you found the best.
   The final value of theta for X4 = [4.15118728 -1.8394291 0.72473856 -0.09513266]
2. Plot loss over the iteration.



3. Based on your training observations, describe the impact of the different learning rates on the final loss and number of training iterations.
   The impact should be the same as when we had the three variables plotted individually.
   Learning Rate = 0.1
   Final Loss: With a learning rate of 0.1, the final loss converged to relatively low values for all explanatory variables, as seen in the cost values provided earlier. It reached convergence after a relatively small number of iterations.
   Number of Training Iterations: The algorithm reached convergence quickly, typically within a few hundred iterations.

   Learning Rate = 0.01
   Final Loss: With a learning rate of 0.01, the final loss also converged to relatively low values for all explanatory variables, though it may take more iterations compared to a learning rate of 0.1.
   Number of Training Iterations: The algorithm required a larger number of iterations to converge compared to a learning rate of 0.1. The convergence process was slower but still effective.

4. Predict the value of y for new (X1, X2, X3) values (1, 1, 1), for (2, 0, 4), and for (3, 2, 1)
   Predicted y for (1, 1, 1) : 3.577408529345462

Predicted y for (1, 1, 1) : 0.2443209702176523
Predicted y for (1, 1, 1) : 0.1025340197359193