simple and multiple-linear regression

TP 1: Simple Linear Regression using the Boston Housingdataset:

1- We have the correlation between the two variables equals 0.69 which means that there is a moderate dependence between the two variables.

2- The r2 score changes dramatically when using different independent variables and we can see that the two independent variables, RM and LSTAT, with 48% and 56%, respectively, are the variables that have the highest r2 score.

3- Simple linear regression depends on several assumptions, the most important of which is that it assumes that the dependent variable has a linear relationship with the independent variables, and the r2 score helps us determine how much variance in the data is explained by the linear relationship between the dependent and independent variables. On the other hand, linear regression is a robust and consistent method, which means it gives a decent result even when one of its assumptions is not met and its result does not change dramatically with different data.

TP 2: Comparing Simple and Multiple Linear RegressionModels:

7- The two models have a difference of 12% in the r2 score, simple and multiple linear regression with 45% and 57%, respectively, which means that adding more independent variables increases the r2 score by 26%. This result emphasizes the importance of using multiple independent variables, which can sometimes decrease the r2 score by adding more variance, but in our case, the other independent variables provide a more linear explanation to the variance of the data.

8- simple linear regression is a simple and easy-to-implement model that can givea robust and consistent result, but in the majority of real-world data, one independent variable is rarely sufficient to predict the target or the dependent variable. In this case, using multiple independent variables can

make a big difference in terms of decreasing the mean squared error and increasing the r2score, which was proven to be true for our data in the lab.

9- As I have mentioned in question 8, simple linear regression is a great model that is robust, consistent, easy to implement, and easy to interpret, but the fact that it relies on the assumption of linear combination between the dependent variable and the independent variables and the independence of the explicative variables make it very weak, as the real world data is way more complex and dependent on each other.