Report for Assignment 1

[Question 1]:

Step 1: Calculated root mean squared error

Step 2: Calculated the mean value of a list of numbers

Step 3: Calculated covariance between x and y

Step 4: Calculated the variance of a list of numbers

Step 5: Calculated coefficients

Step 6: Applied Linear regression algorithm

Step 7: Evaluate the algorithm using a train/test split

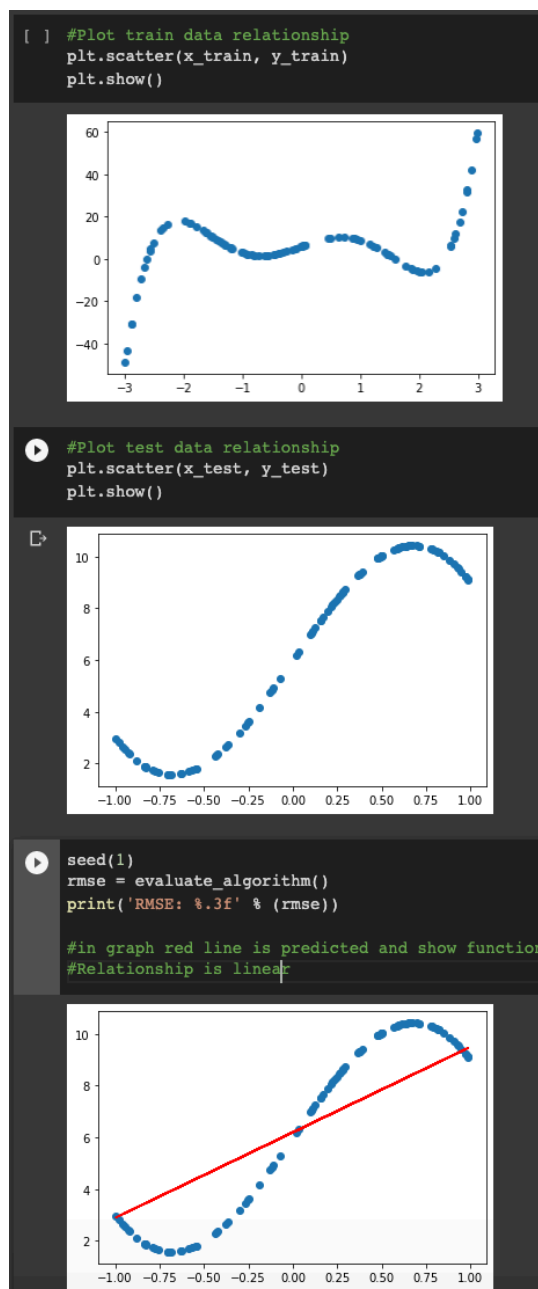Step 8: Plot train data relationship

a.   Is the relationship linear? - YES

b. Do I need feature engineering to add any non - linearity? - YES

i.   If so, how can I engineer these features? - Polynomial Basis

ii. What are some functions that I can try? - RBF is used for approximating particular solutions related to specific loads

1. Plot each of them individually to verify!

```
[ ]  #Plot train data relationship
     plt.scatter(x_train, y_train)
     plt.show()
```



```
#Plot test data relationship
plt.scatter(x_test, y_test)
plt.show()
```



```
seed(1)
rmse = evaluate_algorithm()
print('RMSE: %.3f' % (rmse))

#in graph red line is predicted and show function
#Relationship is linear
```

[Question 2]:

1.What is the average least squares error for the given data using your simple linear regression model?
=> Error = 214.31463619845528

2. Which factor has the most effect on the final value? How do you know this? Can you use only this feature to predict the price?
=> Most Affecting Factor = # Rooms has the most effect on the final value. We know this by a) create a list of the dict's keys and values; b) return the key with the max value. yes, we can use only this feature to predict the price. When we see the correlation value it is more than 60%.

3. Which factor has the least effect on the final value? How do you know this? What effect does removing this feature have on the performance?
=> Least Affecting Factor = Architecture type has the least effect on the final value. We know this by a) create a list of the dict's keys and values; b) return the key with the min value