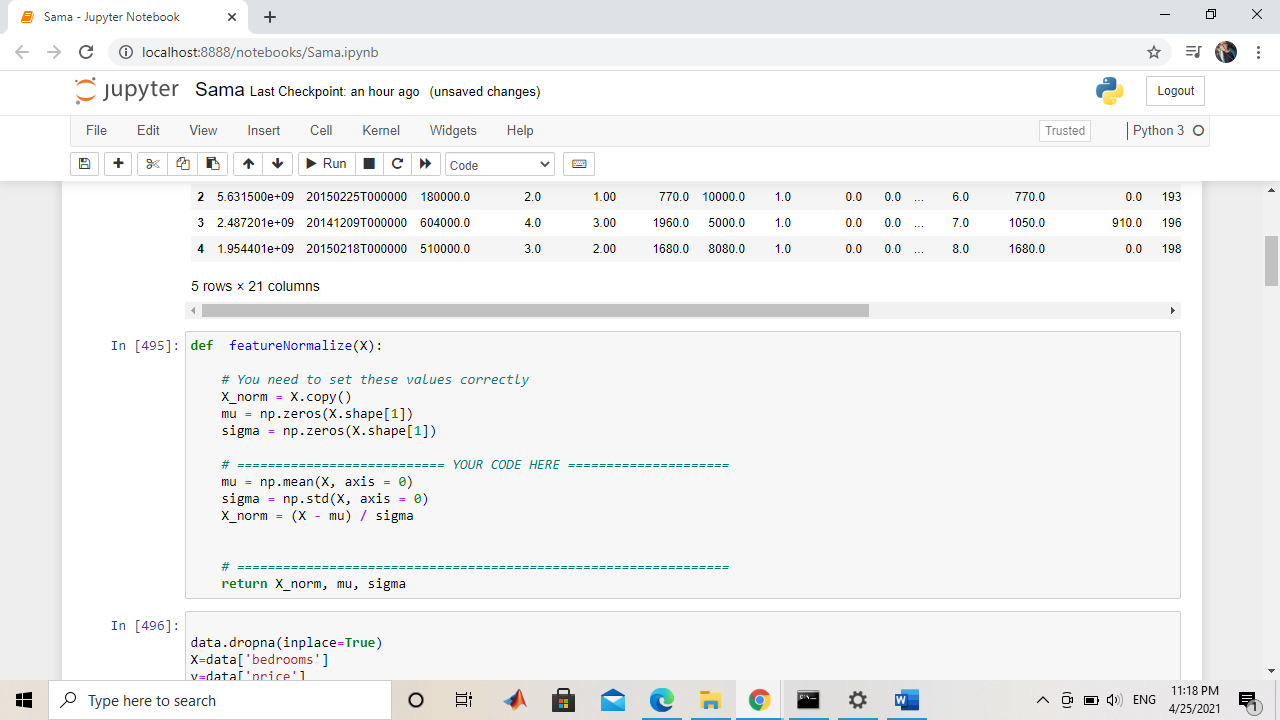
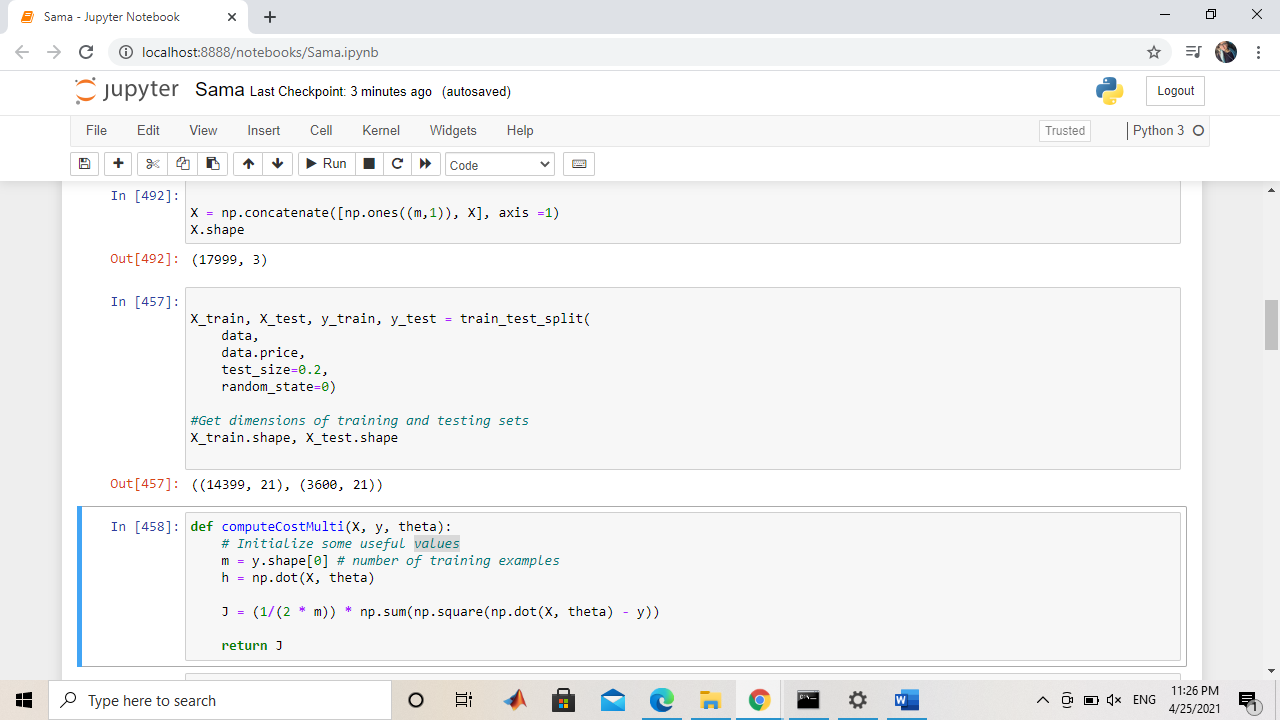
**Report**

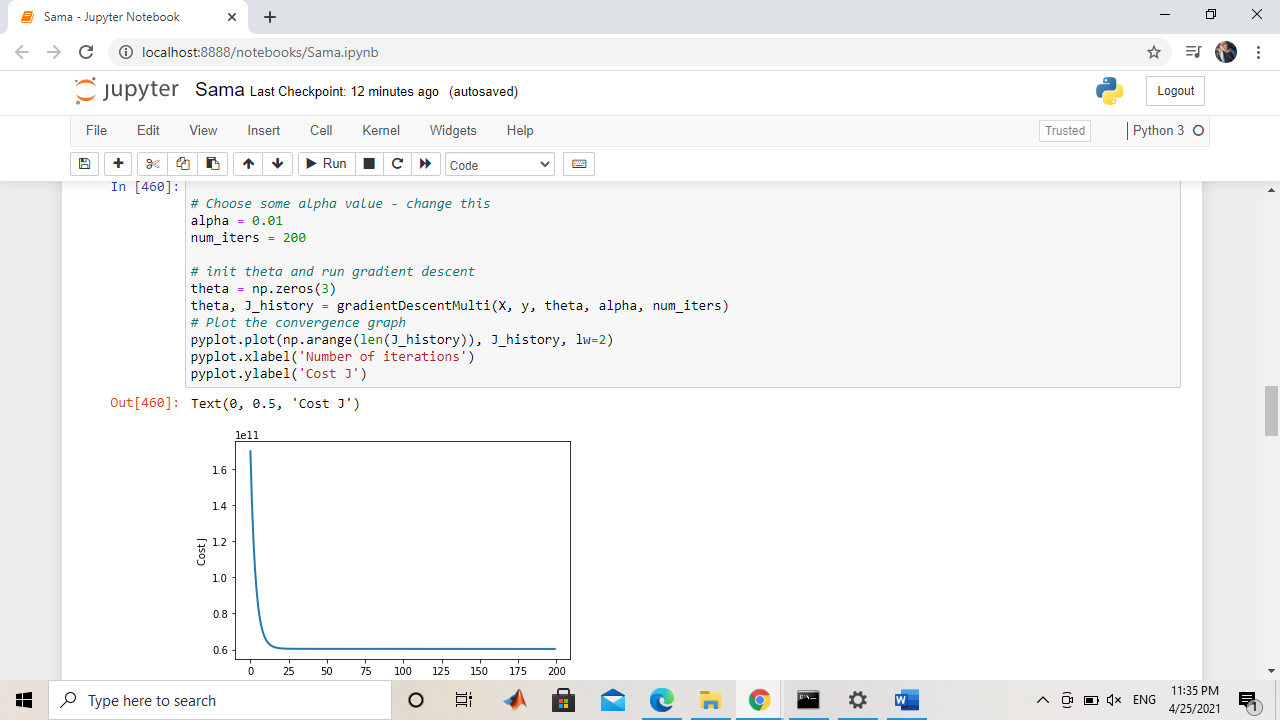
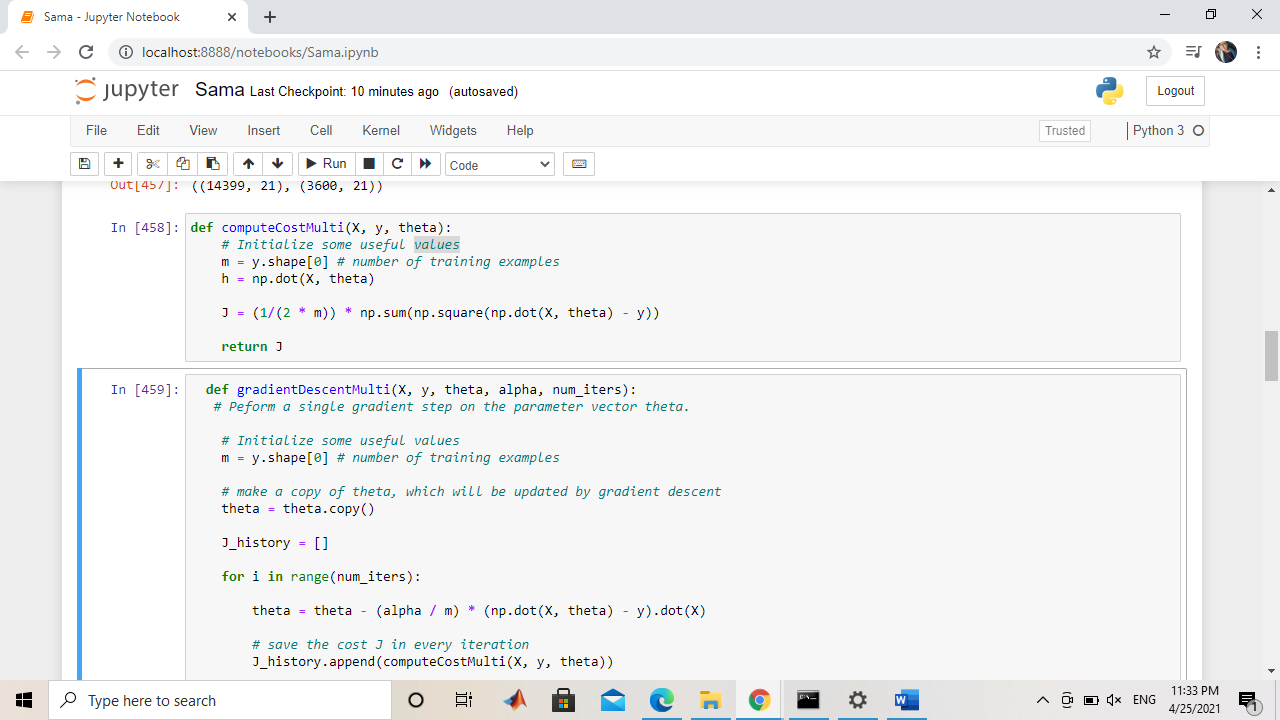
**First, we normalize to make sure that the different features take on similar ranges of values so that gradient descents can converge more quickly.**

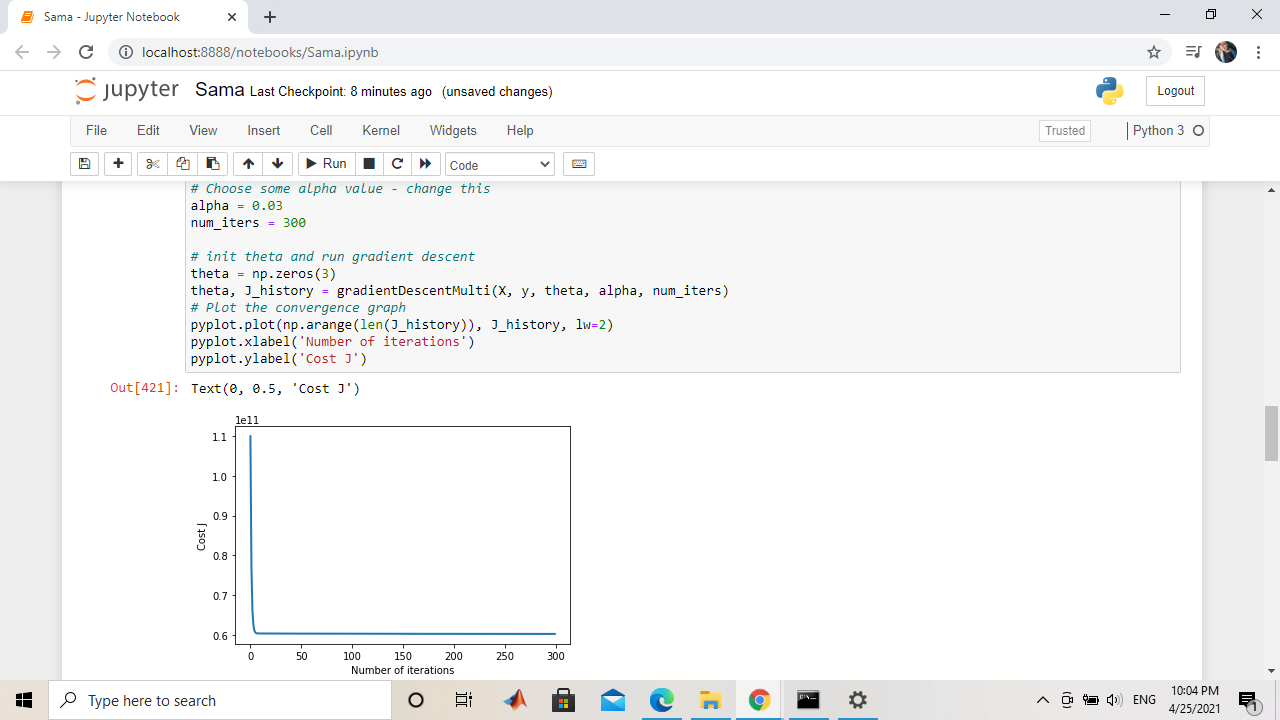
****

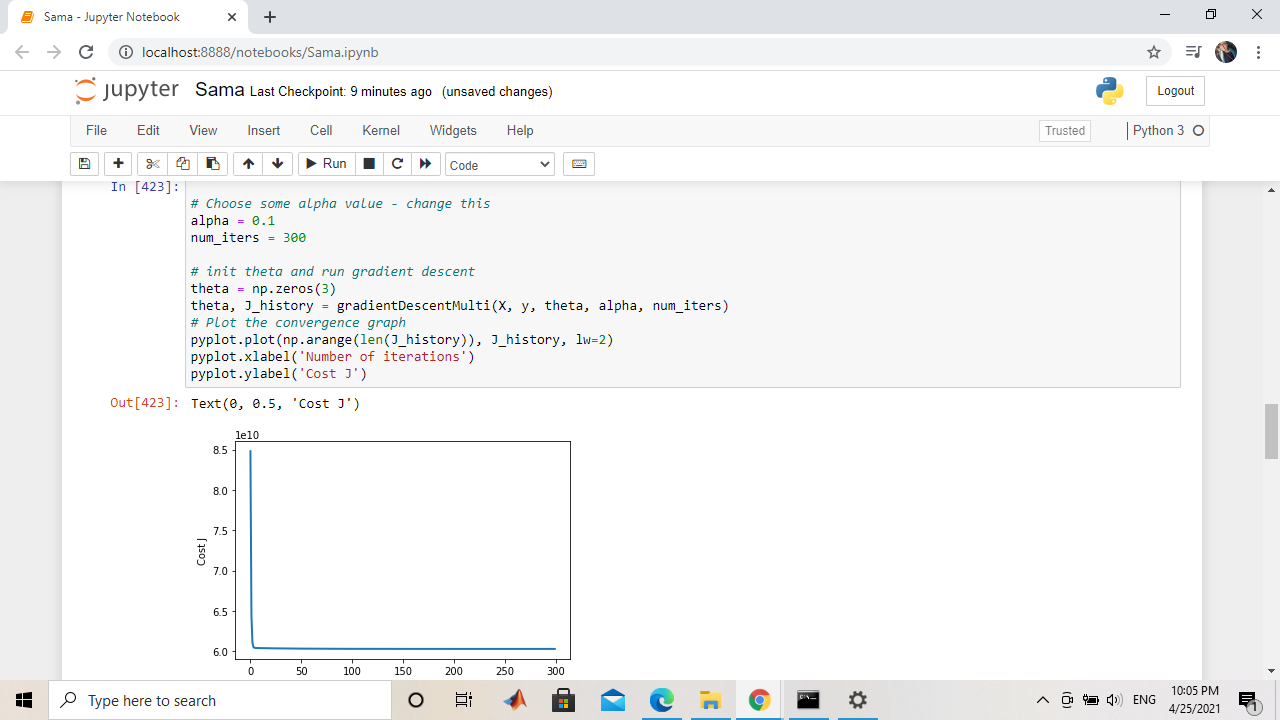
**Then we split the dataset into training set and testing set to understand the parameters in the training set better to minimize the error in the testing set**

****

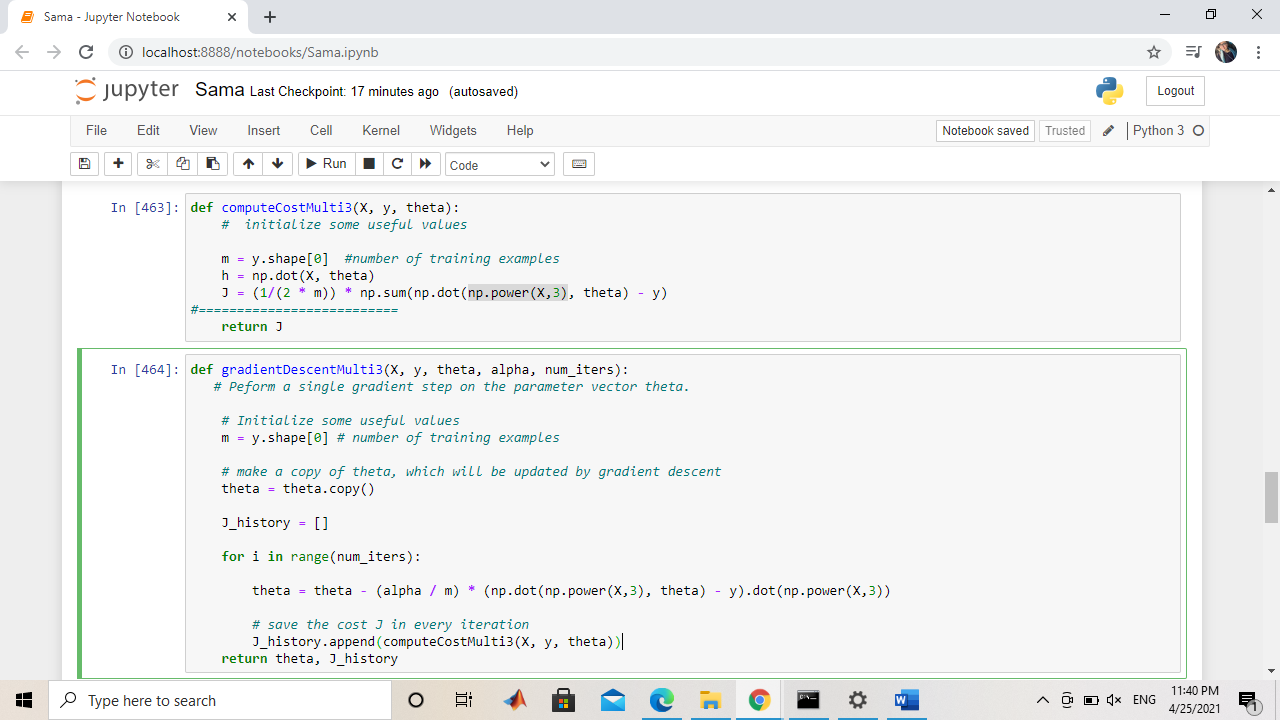
**Here we are calculating the cost and the gradient descent then plotting the results with different alpha values and different number of iterations. When we increase the alpha(the learning rate) the cost decrease.**

****

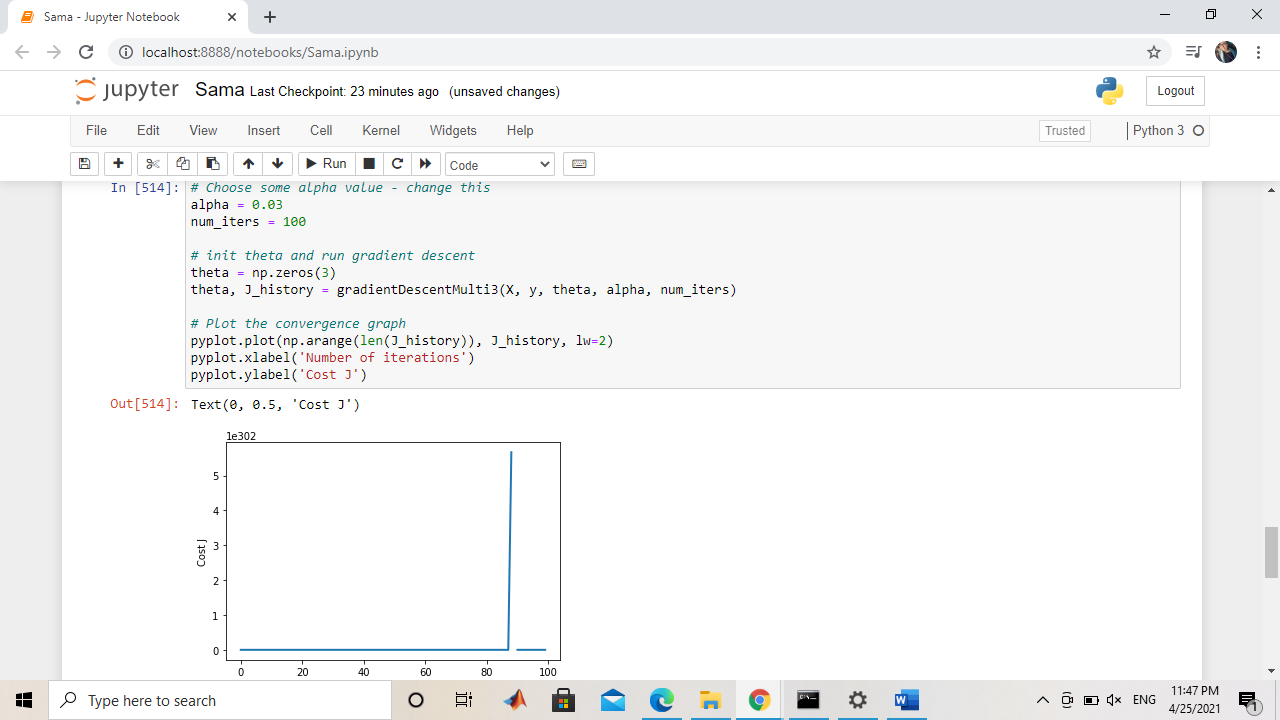
****

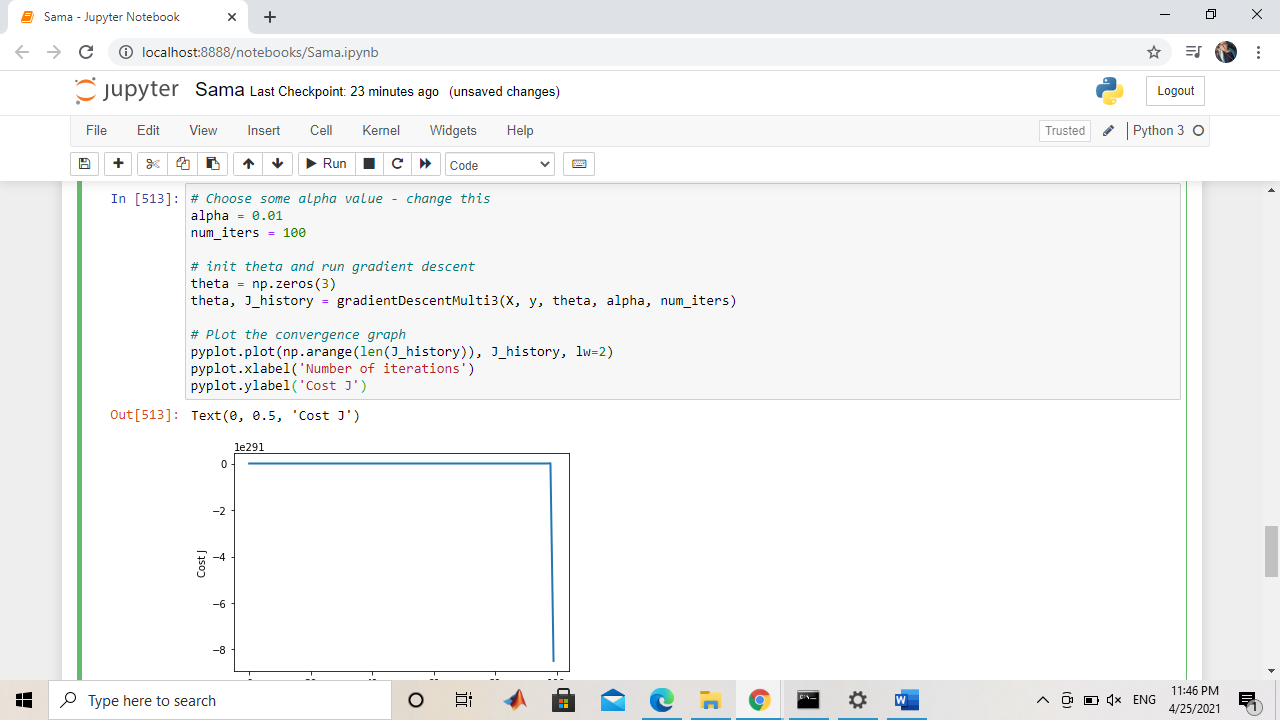
****

**Here we calculate the cost and the gradient descent with different polynomial degree to increases the number of input features. To know how much this impacts the number of features, we can perform the transform with a range of different degrees and compare the number of features in the dataset.**

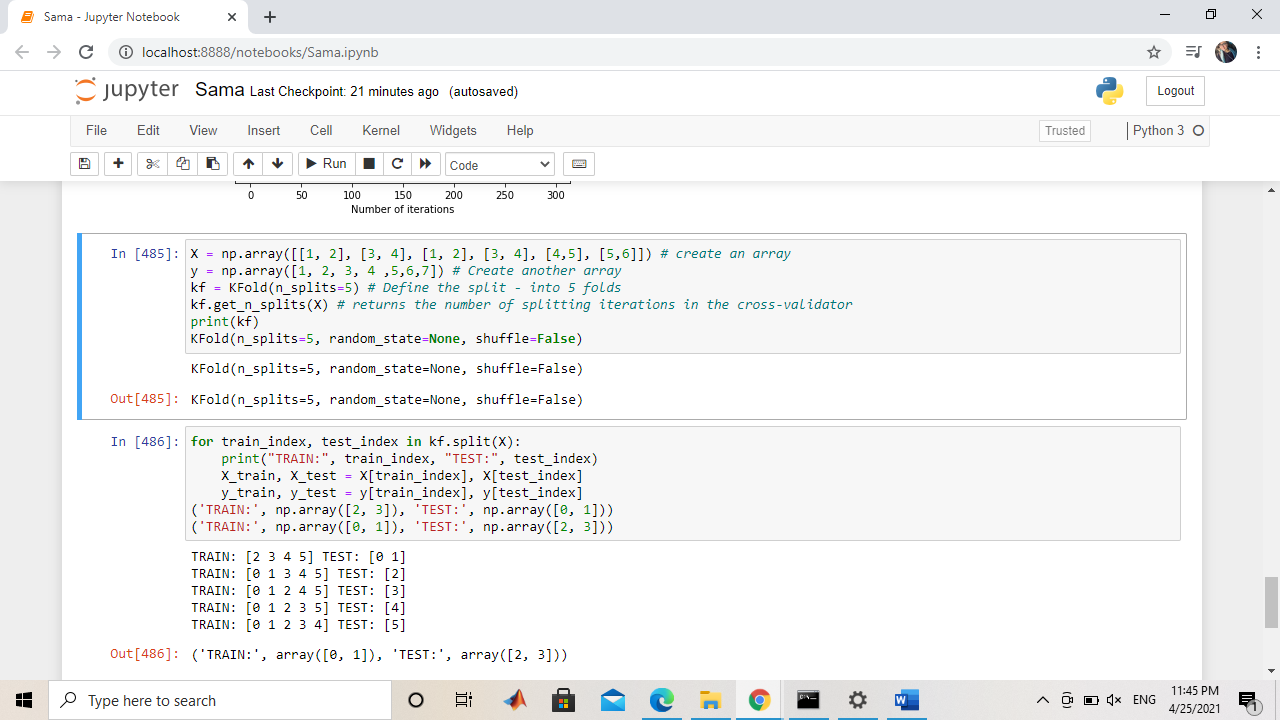
****

**The graphs with the 3rd polynomial with different alpha values**

****

****

**Here we are using the K-fold sampling to have less biased dataset because it ensures that every observation from the original dataset has the chance of appearing in training and test set.**

****