* Exploratory analysis

For the dataset we previously reviewed in the word games sales , there are so many features including 'Platform','Genre','Publisher','Year\_of\_Release','Critic\_Score','Global\_Sales'. However, we are exploring the weight of each feature in deciding the global sales. And according to the features, we try to find the possibility of a non-popular game to be a popular game. We define the popular game to be the one which sold more than one million dollars.

First we extract the features that we think are important

'Platform','Genre','Publisher','Year\_of\_Release','Critic\_Score','Global\_Sales','User\_Score'

We use to supervised learning algos for training . One is the Random Forest Classifier and another one is neural network Multi-layer Perceptron classifier. We extract the features, and split the origin data to test and training data. After train this algorithms, we evaluate the correctness of these two. The Random Forest has a correctness percentage of 0.8510691823899371, and the Multi-layer Perceptron is 0.8274213836477987.

So we use the Random Forest Classifier for training and find the rank of the features.

Then We use the whole data set to train the model then we get the the possibility of a non-popular game to be a popular game.



