* I downloaded the dataset, imported necessary library of python to interogate the dataset
* Used pandas for reading the dataset, followed by taking a glimpse of data using head method, sometimes using sample method on the dataset.
* Looking at the data types of various columns.
* Cleaning the data by imputing null values using replace method and ffill and bfill methods too wherever required. Dropping the rows having null values and playing around with various methods to make data ready to feed for machine learning algorithm.
* Performed feature engineering on columns by converting them similiar data type, performed exploratory data analysis on the dataset using groupby method in order to see how data behaves when grouped with respect to a certain variable.
* After all the rigrous preprocessing, I split the data into training, and test set with a ratio of 0.8 and 0.2 respectively, cross checked it too.
* There are 935 classes in which I need to classify our dataset and the number of points is 111k so it won't make much of sense to classify dataset into such large number of categories.
* Even after this analysis, I want to classify the target variable based on dependent variable I need to use One Hot Encoding follwed by Label Encoding, after this I train a neural network for classification. And this way I can accomplish our task, but that is very very tedious.

I start by splitting the data into training and test set for training and testing our neural network which I can build.

* Second approach for #5: First, I will be doing an SVM classification on the dataframe I have. The label is the Play column. Since it's a string column, I converted it into a categorized labels. That means I are giving a number for each Play so that the classifier can deal with it. The second thing is the features. I had 2 columns with string values as well. A simple conversion to int values is to count the number of words. This should be non representative of the dataset of course but the better option is to use neural network trained on natural language processing. At the end I dropped all NAN rows and then I trained the classifier.