**Changes done in Pharmaceutical Dataset for analysing and model prediction**

* Replacing all the Unknown values with NaN so that Python can understand the missing or null values and we can apply the functions in order to treat those values
* Dropping the Ptid column as it contains only the unique value to each entry in the data
* Replacing the null values in columns Ethnicity and Ntm\_Speciality with the mode of the column
* Dropping the columns Risk\_Segment\_During\_Rx, Tscore\_Bucket\_During\_Rx, Change\_T\_Score and Change\_Risk\_Segment as the first 3 columns contain 45% missing data and the 4th one containing 65% missing values
* Changing all the categorical columns into new columns of 0s and 1s in order to build the prediction model and then dropping the original columns
* Converting Y and N in all the remaining columns to 1 and 0 respectively