**Data cleaning for final project**

**Product Backorder**

The data set for the Product Backorder was divided in 2, one training and one test dataset. I started working on the training part.

My first action was to check the dataset and make sure it was on a tidy format. I looked at the columns names and checked if I could make them simpler, shorter and more descriptive. Fortunately, the columns names were descriptive enough.

The 2nd step was to seek Nan values within the dataset. After finding out Nan values, I had to decide what to do with them.

Therefore, the following step was to replace the Nan values by 0.

Finally, after further investigations, I realized I had 7 columns filled with either Yes or No. I had to replace the character values with numerical values. I decided to create dummy variables by changing No/Yes to 0/1.