**Duplicate products:**

To identify the duplicate products in Stratmed item master data in the below columns is considered

A. Generic Name

B. Form Name

C. Strength

D. Manufacture Name

E. SM Name ( Brand Name )

F. UPP

G. MRP

**Rule 1**. If in any two or more products all the columns from A\_B\_C\_D\_E\_F\_G are matching these products are **Duplicate records**.

**Changes Made:**

1. Converted all the Items into upper case
2. Replaced all NULL values by NAN

**Items not matching**: 7 Items of the duplicate list generated are not matching with the actual stratmed master file.

**Reason:** In the MRP column of the dataset, some entries have actual figure(like 44) and some are with decimals(like 44.0).The code considers them to be different items.

I have tried converting the MRP column from string to float datatype but it is giving error because MRP column has some values like NAN,#VALUE!, blank cells. So these items cannot be converted into float.

**Rule 5**. If in any two or more products all the columns A\_B\_C\_D\_E\_F are matching and column G (MRP) is not matching, products are **Duplicate records** and MRP check has to done.

**Changes Made:**

1. Converted all the Items into upper case
2. Replaced all NULL values by NAN

**Items not matching**: 1 Item of the duplicate list generated are not matching with the actual stratmed master file.

**Reason:** In the doze size column, one entry is 0.02 while some other row has entry 2%. The code considers them to be different.

**Rule 6.** If in any two or more products all the columns from A\_B\_C\_E\_F\_G are matching and column D (company Name) is not match these product are **Duplicate records and** company name needs to checked.

**Changes Made:**

1. Converted all the Items into upper case
2. Replaced all NULL values by NAN

**Items not matching**: 11 Items of the duplicate list generated are not matching with the actual stratmed master file.

**Reason:** In the MRP column of the dataset, some entries have actual figure(like 44) and some are with decimals(like 44.0).The code considers them to be different items.

I have tried converting the MRP column from string to float datatype but it is giving error because MRP column has some values like NAN,#VALUE!, blank cells. So these items cannot be converted into float.

**Rule 8**. If in any two or more products columns A\_C\_D\_E\_F\_G are matching and column B (form) is not matching products are **Duplicate records** and Form needs to checked.

**Changes Made:**

1. Converted all the Items into upper case
2. Replaced all NULL values by NAN

**Items not matching**: 8 Items of the duplicate list generated are not matching with the actual stratmed master file.

**Reason:** In the MRP column of the dataset, some entries have actual figure(like 44) and some are with decimals(like 44.0).The code considers them to be different items.

I have tried converting the MRP column from string to float datatype but it is giving error because MRP column has some values like NAN,#VALUE!, blank cells. So these items cannot be converted into float.

But there are some entries which are not matching and which are common in rule 1,6 and 8

So when I compares these items, I found total 13 items are there which are not matching and need to be changed manually for the success of the output.