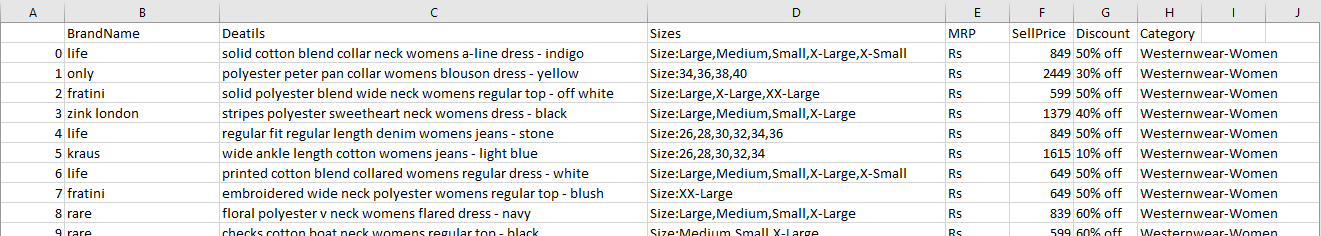
# Fashion Datasets

**Data Importation**

The dataset was imported from Kaggle in CSV format. Thereafter the dataset was then loaded into Excel.

**Observation**

A thorough examination on the dataset revealed that the data needed some cleaning before it could be used to get useful information. Some of the observations made on the dataset include the non-linearity of the index column (i.e indexing resets to zero after every 26 columns), too many info on the Details columns that could be split to multiple columns, redundancy of the word ‘size’ and ‘off’ in the Sizes and Discount columns respectively, amongst others.

****

(Picture 1)

**Data Cleaning**

To make the cleaning process easier, the dataset was first converted into a table called ‘FashionTable’. Below is a step-by-step process I followed in carrying out the data cleaning.

1. New index: A consistent index was created and the former one deleted. This new index was created using the ROWS() function and was called ‘ID’.
2. Split Column: The Details column was then split into two columns. The first column retained the name ‘Details’, while the second the column was renamed ‘Color’. The colors were simply extracted from the details of each transaction.