**Data Processing Assignment**

**Data**

The following data contains a short sample of hair care products from an e-retailer site.

It has Date, Categories/Subcategories, product title, translated product title, actual and discounted price, productid, ratings info.

**Task**

Write a python script (preferably maximize the use of pandas) to:

* Extract the volume from the **Title** column. (Column **J**).
* Extract the quantity from the **Title** column. If the product is a single bottle/packet, then quantity is 1. If the product has multiple Bottles (for e.g. pack of 2) , then quantity is 2.
* Bucket the pricing information in the most appropriate form. Maximum use 5 buckets.
  + example: there can be 4 pricing buckets: (0-500), (500-2000), (2000-5000), (5000+). If a product price is 4000rs the possible pricing bucket will be (2000-5000). Use statistics of the dataset to identify the best split of the buckets.
* Bucket the volume information in the most appropriate form. Maximum use 5 buckets. (i.e. if a product price is 200ml the possible pricing bucket could 100-250ml). Use statistics to identify the best split of the buckets.
* Cluster the datasets using the attribute information (there will be 4 attributes, one from each of the above points – volume, etc.). Form a competition mapping for each product.
  + For e.g Dove 200ml could have Rejoice 230 ml as a competition.
  + A 200ml product and a 500ml product cannot be competitors because the volumes are totally different.