

# CONCEPTUALISATION OF MARKET SEGMENTATION AND PATTERNS FOR PRE-CHRISTMAS SALES IN AN ONLINE RETAIL STORE

Anthony O. Otiko, John A. Odey, Gabriel A. Inyang, Abhishek Puppala

October 23, 2022

## 1 The main idea

The motive of this paper is to create a market segmentation for an online retail store, which analyses the buying patterns of customers for pre-christmas sales. That is, if the customer buys a certain product, what is the chance of buying another similar product related to it with the help of hierarchical clustering.

## 2 The methodology

Here, they used the online-retail dataset which can be found at <https://archive.ics.uci.edu/ml/datasets/Online+Retail>. It is a UK based online retail company with total 541909 transactions and then reduced the dataset that contains 41753 transactions. Here the attributes used are Stock Code, Description, Quantity, Unit Price, Customer Id.

Here the processing of data is done by importing the dataset and then applied association rules. And then visualization was done for clustering with Quantity and Unit Price to create a segmentation. Here customer id is used for creating the baskets(i.e the products in customers cart). Here the apriori function in r always creates rules. By analyzing data, these rules indicates that customers who bought the item 22300 also might buy 22301 with lift 39.66. Some rules are essentially same, hence rules are pruned.

Then hierarchical clustering was performed for finding the group of items that are likely to be purchased. With complete linkage using function hclust and dendogram with 5 clusters were plotted.

## 3 The results

It is observed that, products with consecutive stock codes are associated. Because they may be very similar products which are related. And here they are obtained by creating a co-occurrence of items in customers baskets.

## 4 Recommendation

They could have also used the Country attribute. As there is some dependency, where some of the customers might buy the products that are available in their country, which are deliverable fast. Also more visualizations could have been represented which would be much more clearer to group the clusters.