Samuel Cheboi Tarus

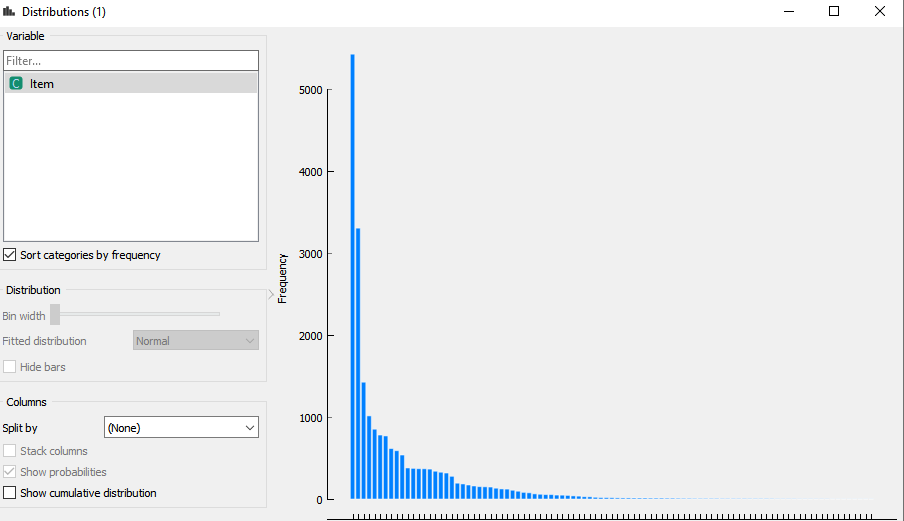
P15/100316/2017

CSC 442

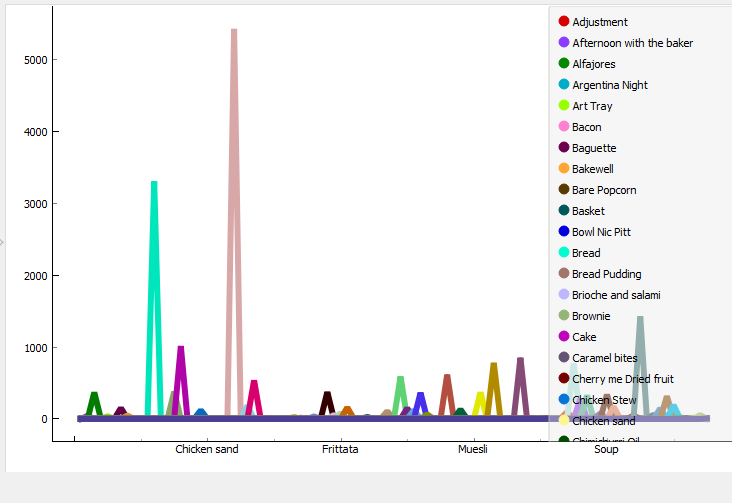
**1. Study the data and analyze it carefully. Provide a monthly summary that include frequencies, describe the timings vs the customer flow over a working day and weekend day.**

First l check the datasets : 21173 instances ( No missing data) with 3 features and 2 meta attributes.

The Graph shows the frequency of items distribution



From the data analysing from the graph using line plot it clearly shows that some of the products are highly consumer by the customers which makes one of the main business objectives to provide such as :



**2. Convert the data in a format that you can then use on WEKA (or any other ML toolkit) to enable you use the Apriori Algorithm for Association Analysis.**

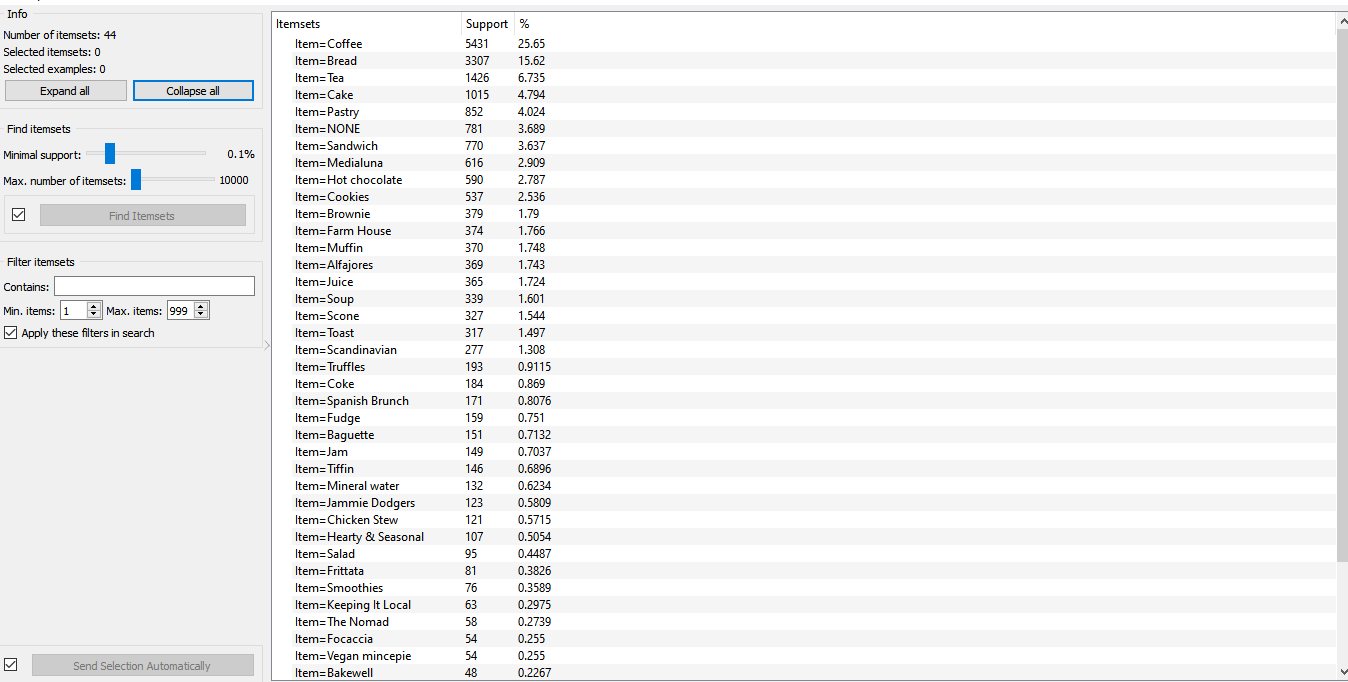
I have used the orange toolkit to analyse the data mostly and weka on determining the frequencies monthly ,weekly and weekend

**a. Identify the frequent itemsets by setting your own Support and Confidence values.**

Since there a lot of instances l took the minimal support of **0.1%**

For the frequent itemsets :

The Customers seem to be buying a whole lot of Coffee, bread and Tea. Call your marketing department or the cooking department to increase quality and serve this mostly highly consumed item which becomes the ultimate coffee house .



So if a person buys Coffee, it is most likely to buy Cake as an accompanying product group.or So if a person buys Tea, it is most likely to buy Cake as an accompanying product group.

**Association Rule**

Since the data set is larger l increase the confidence level to increase the level of accuracy.

b. Using this, generate frequent itemsets per month, working day vs weekend. What Association Rules do you find?

c. Is there a drift over the months?