**IHLP – Final Project**

**Predicting items in a shopping cart using R Programming**

**By: Joel A Ulahanna**

**UCID: jau3**

**Instructor: Rahman Moshiur**

**Introduction**

R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical and graphical techniques in order to display the data in a meaningful and understandable way. It is designed to seamlessly incorporate compiled code which gives you all the benefits of an interactive language while allowing you to capitalize on the speed of compiled code.

We used R programming to analyze supermarket transactions based on the patterns and trends. The primary goal is to find the most frequent item-sets and to use that information to provide a suggestion to customers about an item they might want to buy or perhaps is missing from their cart.

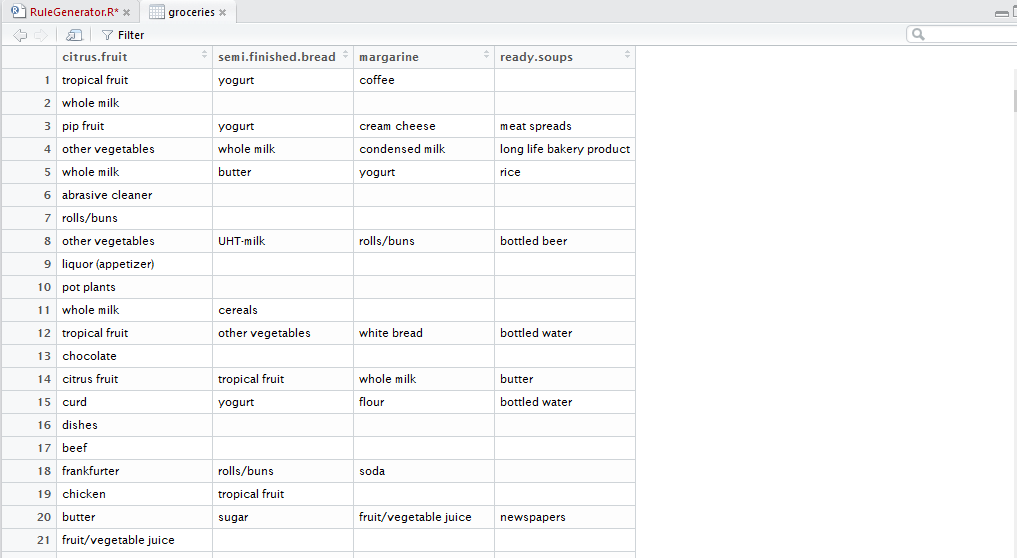
**Data Collection**

The raw data set was downloaded from the groceries dataset link on Github.com.

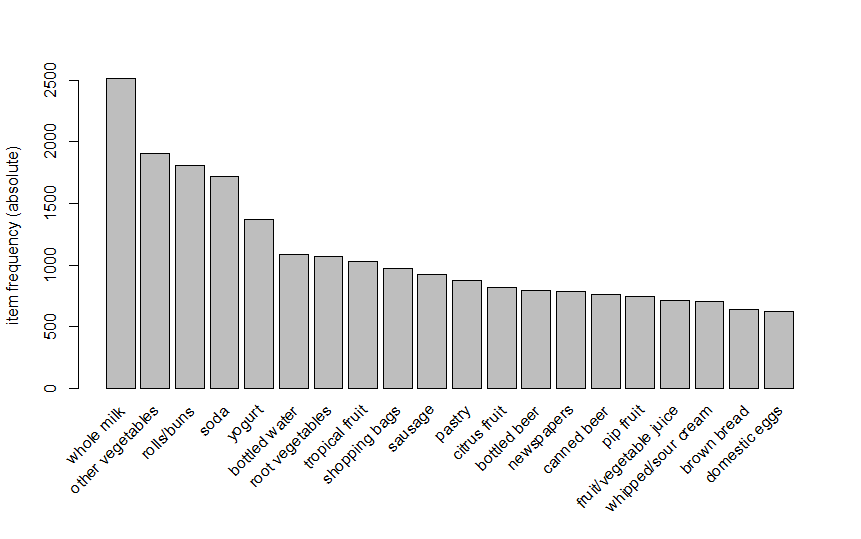
**Dataset Preparation**

The data set was in a random format to be entered into the R programming and each transaction and varying number of items. The data set was pulled from the csv file in a basket format with comma as a separator. Once loaded the data was now ready to be plotted or inspected.

**Dataset Example:**



**Data set plot:**



**Frequent item-sets and association rules generation**

This dataset was fed to an Apriori algorithm to generate frequent item-sets and association rules.

The confidence was set initially to 0.6 and support to 0.001.

The number of rules generated was large.

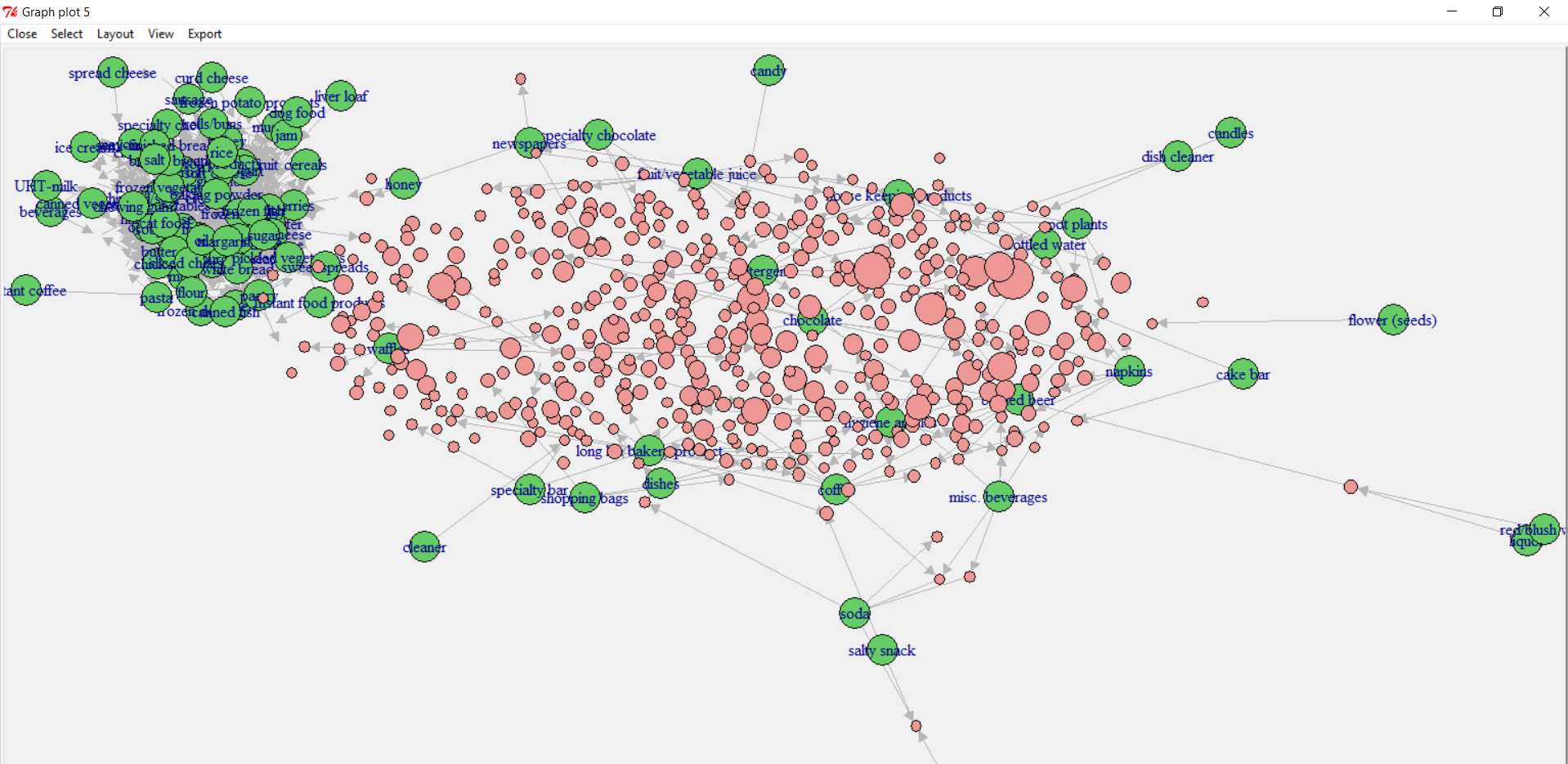
The confidence was ten set to 1 and only those rules with 100% confidence were returned.

After tuning we found that a confidence of 0.9 was ideal as it returned substantial set of rules that have 90% confidence.

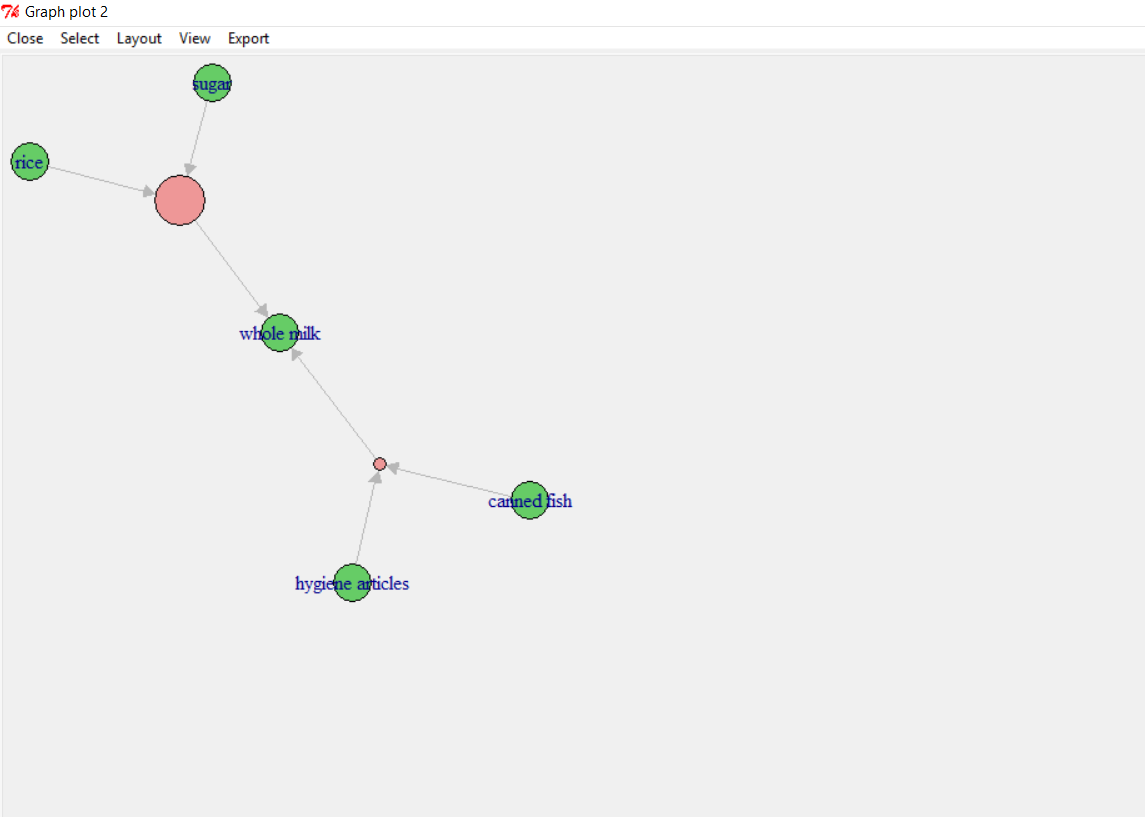
The rules were also pruned to remove any redundant rules.

Later the rules were restricted to a length of 3, i.e two antecedents and one consequent for the scope of the project.

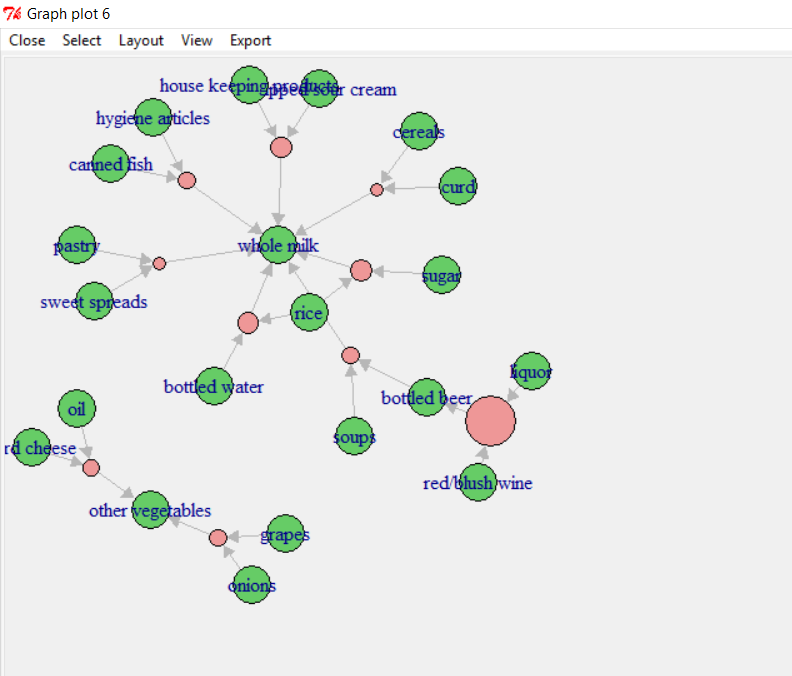
**Rules generated with Confidence 0.6:**



**Rules generated with Confidence 1:**



**Rules generated with confidence of 0.9:**



**Feeding rules to Database and Web application**

The rules were exported to a CSV and fed to MySQL database.

A jsp web application was used to demonstrate the effectiveness of a rule in action.

The jsp website basically extracted the items a customer has bought and compared them with the rules in the database for a match. If there was a match the consequent item was suggested to the user.

**Conclusion**

Thus, with frequent item sets and association rules of high confidence we can conclude that if a rule holds 90% confidence, it is likely to hold for the next customer. It helps customers in buying items they may often forget. Or items they may need for a purpose but are unaware of.

The rules generated were of a large scale and needed to be pruned and tuned to obtain a workable set of rules. Also, a much larger dataset would guarantee rules with much confidence and help for a better prediction.

**References**

* <https://github.com/stedy/Machine-Learning-with-R-datasets>
* <https://www.r-project.org>
* [https://www.**mysql**.com/](https://www.mysql.com/)
* <http://www.oracle.com/technetwork/java/javaee/jsp>