# Market Basket Analysis

## Introduction

Mining frequent itemsets and association rules is a popular and well researched approach for discovering interesting relationships between variables in large databases.

Understanding buying patterns can help to increase sales in several ways. If there is a pair of items, X and Y, that are frequently bought together.

* Both X and Y can be placed on the same shelf, so that buyers of one item would be prompted to buy the other.
* Promotional discounts could be applied to just one out of the two items.
* Advertisements on X could be targeted at buyers who purchase Y.
* X and Y could be combined into a new product, such as having Y in flavors of X.

Association rules analysis is a technique to uncover how items are associated to each other. There are three common ways to measure association.

**Support**: This says how popular an itemset is, as measured by the proportion of transactions in which an itemset appears. If it is discovered that sales of items beyond a certain proportion tend to have a significant impact on profits, one might consider using that proportion as the support threshold.

**Confidence**: This says how likely item Y is purchased when item X is purchased, expressed as {X -> Y}. This is measured by the proportion of transactions with item X, in which item Y also appears

**Lift**: This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is. A lift value greater than 1 means that item Y is likely to be bought if item X is bought, while a value less than 1 means that item Y is unlikely to be bought if item X is bought.

## Methodology

We performed market basket analysis on Instacart data using R. We used the popular library arules; which is a computational environment for mining association rules and frequent item sets.

Market Basket Analysis allows us to identify items that are frequently bought together. Typically the output of an MBA is in the form of rules. The rules can be simple {A ==> B}, when a customer buys item A then it is (very) likely that the customer buys item B. More complex rules are also possible {A, B ==> D, F}, when a customer buys items A and B then it is likely that he buys items D and F.

Step 1: Find frequent items in the shopping baskets. The support is set to be 0.02

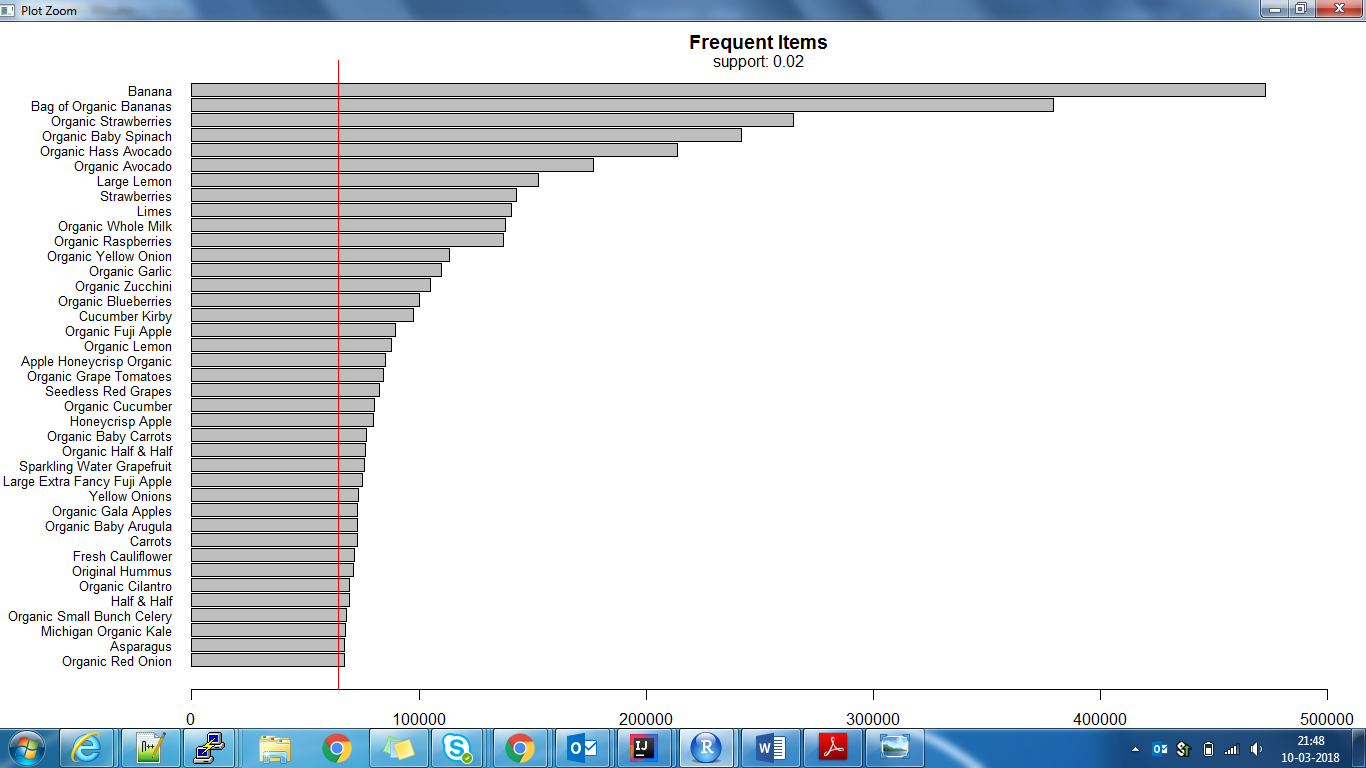


Figure 1: Frequent Items: Support=0.02

Banana is the most favorite item followed by Strawberries! Clearly vegetables and fruits are the most ordered products.

Step 2: Now we apply apriori algorithm to compute the frequent itemsets. We decrease the support threshold to take into account the small probability of observing a frequent itemset of at least size 2.

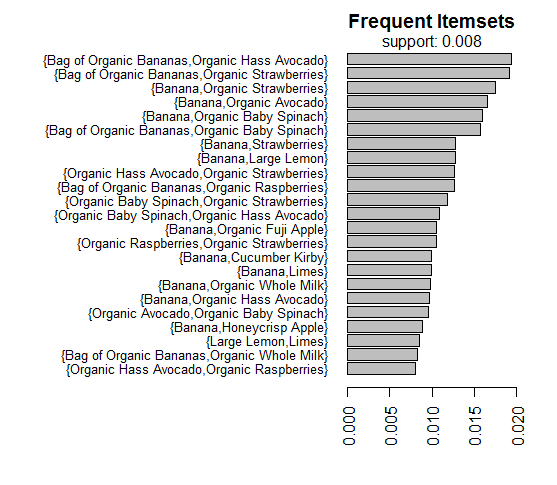
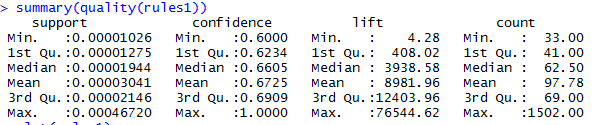


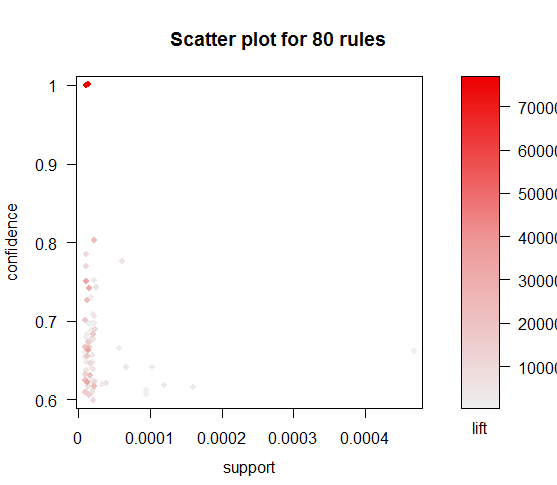
Figure 2: Frequent Itemsets; Support=0.008

Bananas being the most favorite item, rules the show here too! It is part of almost every frequent item set.

Step 3: Lets mine some association rules. First, we use a low support threshold and a high confidence to generate strong rules even for items that are less frequent.

**support = 0.00001, confidence = 0.6**





There are some rules with a large value of lift indicating a strong association between the items. Let's further investigate those critical rules.



Figure 3; Top 10 Rules by Lift



Figure 4: Top 10 rules by Confidence

Step 4: Next, we increase the support and decrease confidence to get rules of some more frequent items but with less confidence.

**support = 0.001, confidence = 0.4**

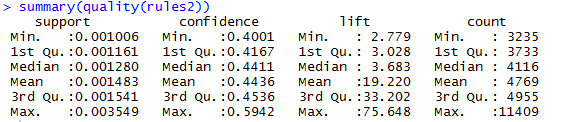




Figure 5: Top 10 rules by Lift

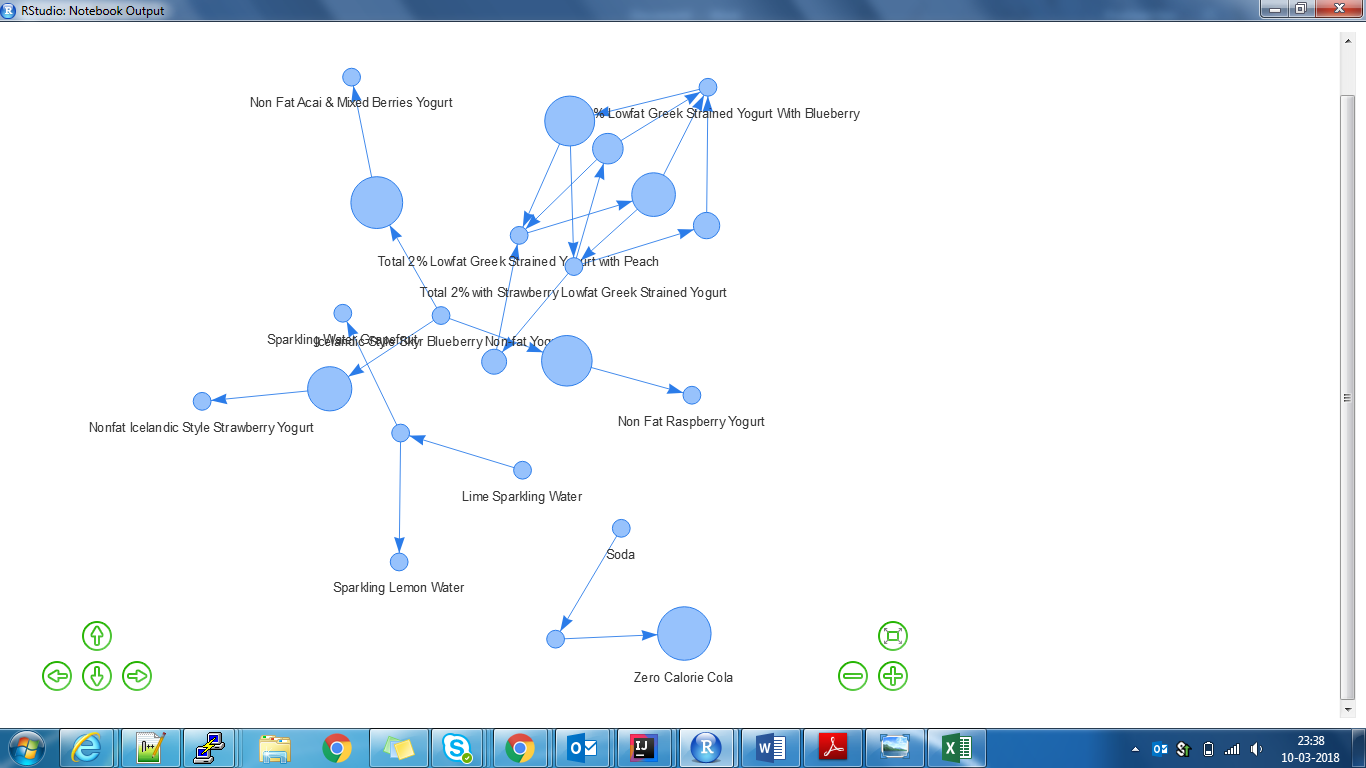
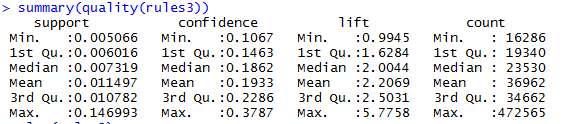


Figure 6: Network visualization of rules

Step 5: Finally, lets further increase support and decrease confidence.

**support = 0.005, confidence = 0.1**



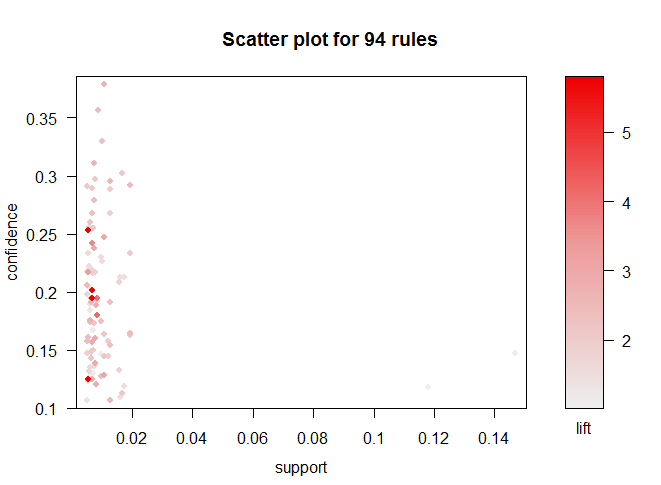




Figure 7: Top 10 rules by Lift

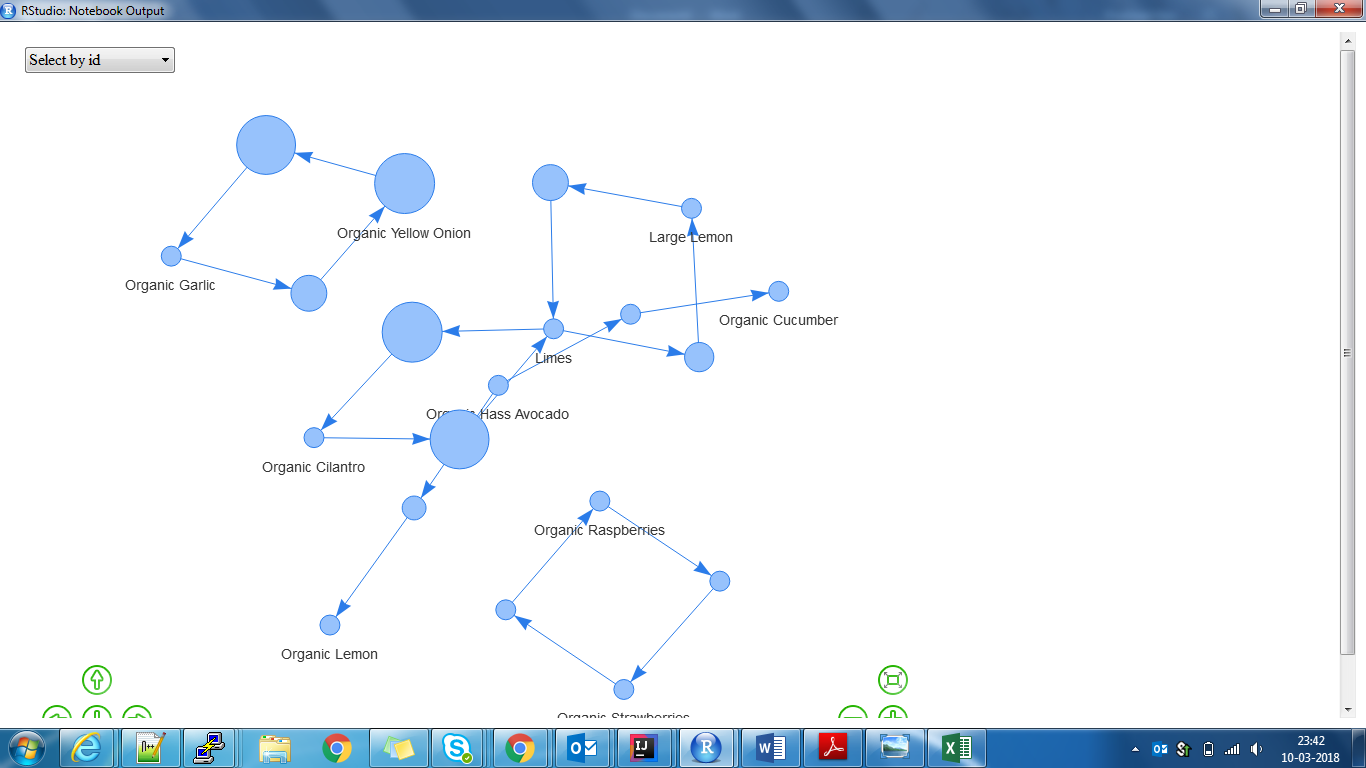




Figure 8: Top 10 rules by Confidence

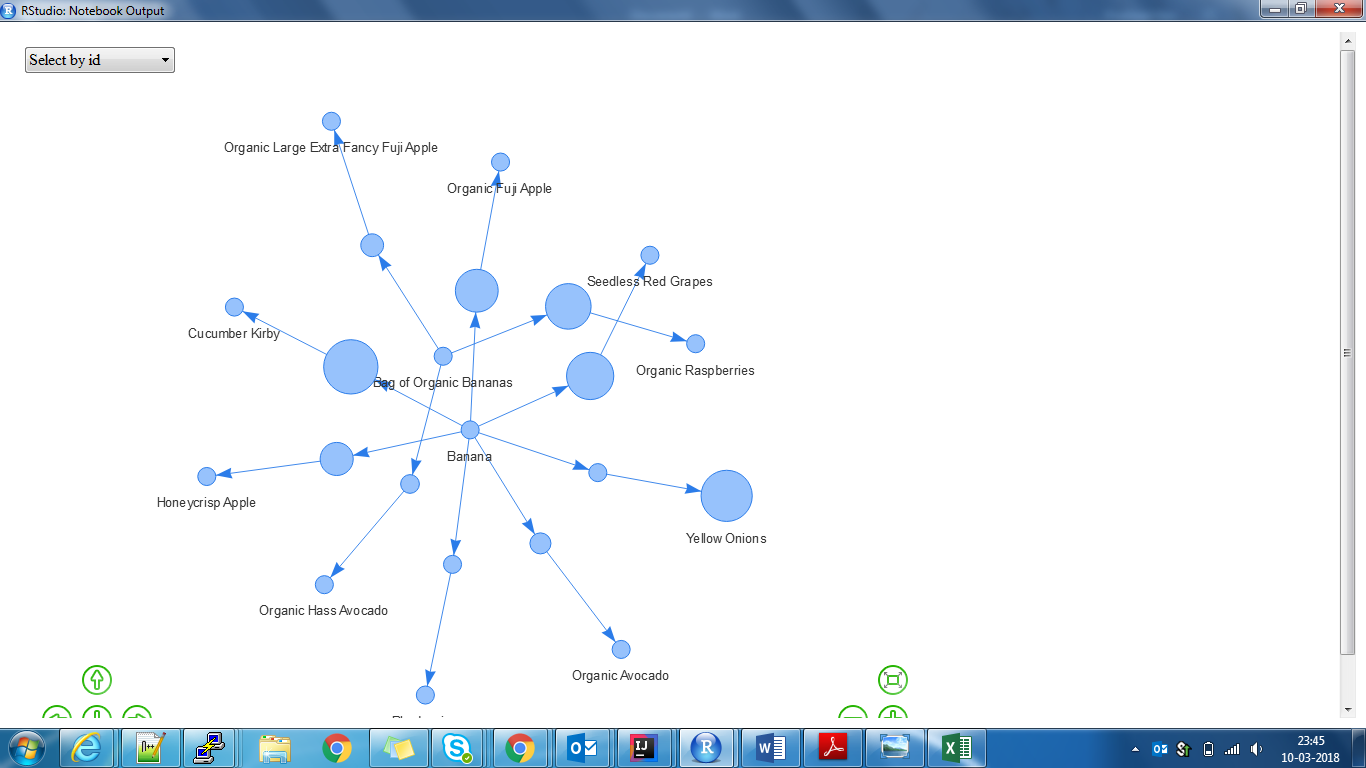


Figure 9: Network Visualization of top 10 rules by confidence