FDMart Grocery Market Basket Analysis

Ignacio Faria

February 7, 2017

**FDMart Grocery** specializes in fresh vegetables and fruits, and in order to better understand our customers and their purchasing behavior, we will investigate purchasing patterns regarding:

1. Alcoholic Beverages - these only include *wine* and/or *beer* purchases
2. Canned and Fresh Vegetables and Fruits - all vegetables and fruits are grouped and none appear individually
3. Large and Small Transactions - transactions with *5* items compared to *2*
4. Quick-to-Serve Meats and Uncooked Meats - meats that are cooked, sliced (deli meats), and packaged as opposed to *frozen* or *fresh* meats (chicken and fish)

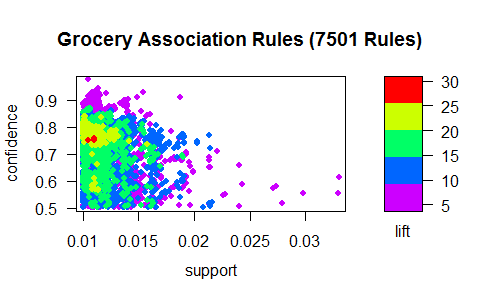
This [large transaction dataset](https://bb.csueastbay.edu/bbcswebdav/pid-2386200-dt-content-rid-5479622_1/courses/20171_ITM_6285_01_1/transactions.txt) contains over *300,000 items* purchased with *106 unique products*. The rules generated in this analysis are parametrized to *at least 1%* of each transaction, and each transaction has *at least a 50% association* with the subsequent purchase.

## Overview of the Data

Exploratory data analysis tells us that:

* The data contain *64,808* transactions, with **Fresh Vegetables** and **Fresh Fruit** as the most frequently purchased products.
* The transaction sizes are mostly spread between *3 to 6* items per purchase, with the rest of the data being heavily skewed to the right.
* The great majority of **rules** contain either *3* or *4* products.
* *Fresh Vegetables*, *Fresh Fruit*, *Canned Vegetables*, and *Wine* are of particular interest. They are among the top selling products, and are the focus of several parts of this analysis.

It is important to recognize that *Fresh Vegetables* are included in nearly 30% of transactions. Because of this, many associations in this analysis would suggest that increasing sales of nearly anything would increase sales of *Fresh Vegetables*. This is likely not the case.



# Purchacing Patterns Related to Wine and Beer

Purching patterns with *wine* and/or *beer* were not nearly as frequent as others; infact, *beer* was never found to be associated with anything. The following association rules show that the purchasing of *wine* is highly associated with the purchasing of *fresh vegetables*.

## lhs rhs support confidence lift   
## [1] {Candles,Wine} => {Fresh Vegetables} 0.01029 0.8708 2.821  
## [2] {Fresh Chicken,Wine} => {Fresh Vegetables} 0.01023 0.8852 2.868  
## [3] {Sauces,Wine} => {Fresh Vegetables} 0.01492 0.9088 2.945  
## [4] {Cooking Oil,Wine} => {Fresh Vegetables} 0.01273 0.7406 2.400  
## [5] {Rice,Wine} => {Fresh Vegetables} 0.01031 0.8077 2.617  
## [6] {Juice,Wine} => {Fresh Vegetables} 0.01025 0.7273 2.357  
## [7] {Fresh Fruit,Wine} => {Fresh Vegetables} 0.01531 0.5815 1.884

## lhs rhs support confidence lift   
## [1] {Sauces} => {Wine} 0.01642 0.5257 5.131  
## [2] {Candles,Fresh Vegetables} => {Wine} 0.01029 0.6234 6.084  
## [3] {Fresh Chicken,Fresh Vegetables} => {Wine} 0.01023 0.6357 6.204  
## [4] {Fresh Vegetables,Sauces} => {Wine} 0.01492 0.6748 6.586

Because *fresh vegetables* are so frequent in this dataset, lift would be a better measure of product purchase associations for *wine*. *Sauces*, *fresh chicken*, and *candles* appear to best predict *wine* purchases.

It appears that wine purchases are likely motivated by customers who want to have a better dinner experience.

# Canned vs Fresh Vegetables and Fruits

1. Canned Vegetables and Fruits

Unstandably, the Apriori algorithm produced many rules here. We found no rules associated with **Canned Fruit**. Also, there were far more rules regarding **Fresh Vegetables** than **Fresh Fruits**. **Canned Vegetables** were almost always paired with **Fresh Vegetables**, as was *sour cream* and *jam*.

1. Fresh Vegetables and Fruits

We have 494 rules predicting the purchase of fresh produce, and 1512 rules including fresh produce. These rules had a must lower lift than any other group, due to their frequency. **Fresh Vegetables** and **Fresh Fruits** were almost always paired together. *Pasta*, *rice*, *jelly*, *juice* and *pancake mix* were the most frequent associations here.

## lhs rhs support confidence  
## [1] {Fresh Fruit,Pasta,Rice} => {Fresh Vegetables} 0.01048 0.9742   
## [2] {Fresh Vegetables,Pasta,Rice} => {Fresh Fruit} 0.01048 0.8639   
## [3] {Jelly,Pancake Mix,Sponges} => {Fresh Vegetables} 0.01049 0.9115   
## [4] {Jelly,Juice,Sponges} => {Fresh Vegetables} 0.01054 0.8999   
## [5] {Juice,Pancake Mix,Sponges} => {Fresh Vegetables} 0.01202 0.8665   
## lift   
## [1] 3.157  
## [2] 4.429  
## [3] 2.954  
## [4] 2.916  
## [5] 2.808

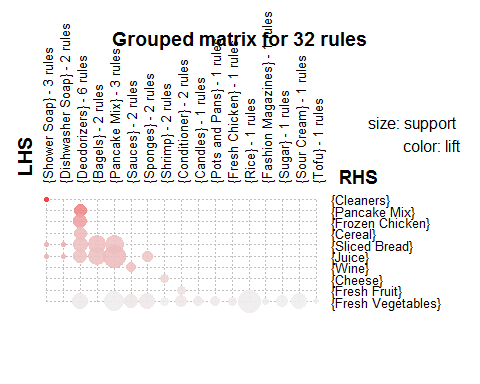
## lhs rhs support  
## [1] {Fresh Fruit,Pasta,Rice} => {Fresh Vegetables} 0.01048  
## [2] {Fresh Vegetables,Pasta,Rice} => {Fresh Fruit} 0.01048  
## [3] {Fresh Fruit,Fresh Vegetables,Pasta} => {Rice} 0.01048  
## [4] {Fresh Fruit,Fresh Vegetables,Rice} => {Pasta} 0.01048  
## [5] {Fresh Vegetables,Jelly,Sponges} => {Pancake Mix} 0.01049  
## confidence lift   
## [1] 0.9742 3.157  
## [2] 0.8639 4.429  
## [3] 0.6936 11.630  
## [4] 0.5843 9.602  
## [5] 0.7861 14.581

# Small and Large Transactions

We have 32 rules for small transactions and 400 rules for large transactions

## Small Transactions

These transactions seem to have a much lower level of confidence and lift than the previous rules we have seen. None appear to be of particular interest except for *sauces* and *wine*, which had a lift of 5, indicating a decently positive correlation.



## Large Transactions

Large transactions have much higher confidence and lift levels than the small transactions. These three large transactions in particular showed strong associations.

inspect(lrules[3:5])

## lhs rhs support confidence lift  
## [1] {Cottage Cheese,   
## Deodorizers,   
## Fresh Vegetables,   
## Sliced Bread} => {Frozen Chicken} 0.01038 0.8060 11.99  
## [2] {Deodorizers,   
## Fresh Vegetables,   
## Frozen Chicken,   
## Sliced Bread} => {Cottage Cheese} 0.01038 0.7936 14.32  
## [3] {Cottage Cheese,   
## Fresh Vegetables,   
## Frozen Chicken,   
## Sliced Bread} => {Deodorizers} 0.01038 0.8148 19.66

# Meats: Quick-To-Serve and Uncooked Meats

## Quick-to-Serve

These may not appear intuitive to most people, so the code is available to see (qmeat is "quick meats" and smeat is "slow meats").

qmeat <- c("Bologna","Deli Meats", "Dried Meat","Hot Dogs")  
smeat <- c("Fresh Chicken", "Frozen Chicken", "Hamburger", "Fresh Fish")

114 rules were generated and the only *quick-to-serve* meat that made any associations was *deli meats*. *Ice cream* and *deodorizers* seem to be most closely associated with *deli meats*.

## lhs rhs support confidence lift   
## [1] {Deli Meats,Ice Cream} => {Deodorizers} 0.01012 0.6855 16.54  
## [2] {Deli Meats,Deodorizers} => {Ice Cream} 0.01012 0.7480 12.22

## Cook-to-serve

977 rules left hand side. 347 rules right hand side. *Fresh chicken* and *frozen chicken* are the most frequent *cook-to-serve* meats with rules. *Fresh chicken* is highly associated with *wine* and *fresh vegetables*. *Frozen chicken* is very highly associated with *shrimp* and *waffles*, which are both typically sold frozen as well. *Deodorizers* also appear to be correlated with *frozen chicken*.

## lhs rhs support confidence lift  
## [1] {Deodorizers,Shrimp} => {Frozen Chicken} 0.011 0.78 12   
## [2] {Chips,Shrimp} => {Frozen Chicken} 0.011 0.73 11   
## [3] {Pancake Mix,Shrimp} => {Frozen Chicken} 0.011 0.77 11   
## [4] {Shrimp,Waffles} => {Frozen Chicken} 0.011 0.73 11   
## [5] {Canned Vegetables,Shrimp} => {Frozen Chicken} 0.011 0.72 11

# Summary

In conclusion, these associations indicate: 1. **Wine** sales were largely associated with *fresh chicken*, *candles*, and *sauces*, indicating a customer who wants to make a special dinner. No **Beer** associations were made. 2. **Canned Vegetables** were associated with *sour cream* and *jam*. No **Canned Fruit** associations were made. **Fresh Vegetables** and **Fresh Fruits** showed much stronger lift correlations and were most associated with *pasta*, *rice*, *jelly*, *juice*, and *pancake mix*. 3. **Small Transactions** gave little perspective on our customers, except for reinforcing that *sauce* was associated with *wine*. On the other hand, **Large Transactions** showed very strong correlations indicating that *frozen chicken*, *cottage cheese*, and *deodorizers* can strongly be predicted. 4. **Quick-to-Serve Meats** only associated with *deli meats*, which were strongly correlated with *ice cream* and *deodorizers* purchases. *Frozen chicken* was strongly associated with other frozen products, specifically *shrimp* and *waffles*.