

Data mining (Itm6285)

ASSIGNMENT 1



February 8, 2017

ANANYA AUDDY

NET ID: PY7552

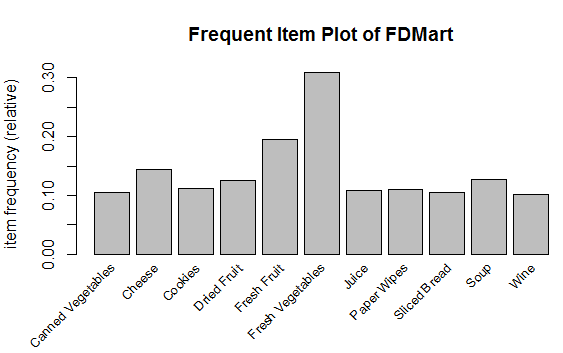
Objective: FDMart is a grocery store dealing with various products of daily needs. Customer visits the store, buys items and a transaction is generated. FDMart specializes in fresh vegetables and fruits. The store is considering expanding its product selection and wants to better understand its customers and their purchasing behaviour. We are provided with a database having 64808 transactions and need to identify interesting patterns. FDMart deals with 106 unique items and we need to do a market basket analysis to identify interesting patterns.

The Marketing Analyst has provided the following patterns as a starting point for analysing the data

* Purchase patterns related to beverages (Wine, Beer etc.)
* Canned vs Fresh
* Small vs large transactions
* Find one other interesting patterns, across two different product categories that can help increase sales of their products other than Fresh Vegetables and Fruits.

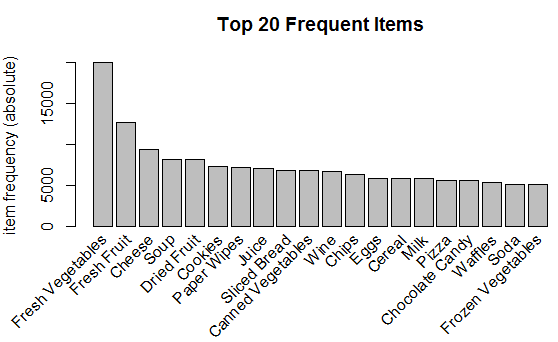
Before we start our analysis, we need to get familiarized with the transaction pattern. Let’s explore before creating any rules.

Firstly, we split the products based on transaction ID and plot the frequent items that have 10% support i.e. if there are 100 transactions, there will be 10 such transactions which will have these items (Relative Frequency).



The plot clearly reveals that the store indeed specializes in Fresh Vegetables and Fresh Fruit as stated above. Fresh Vegetables occur in 30% of the transactions and Fresh fruit in 20%.

Top 20 Frequent Item sets that occur in the transaction database(Absolute Frequency). This would help in identifying other interesting patterns, to increase sales of their products other than Fresh Vegetables and Fruits.



Next, we create a rule, having 0.5% support and 80% confidence. This resulted in a set of 733515 rules.

## BEVERAGE SELLING PATTERN

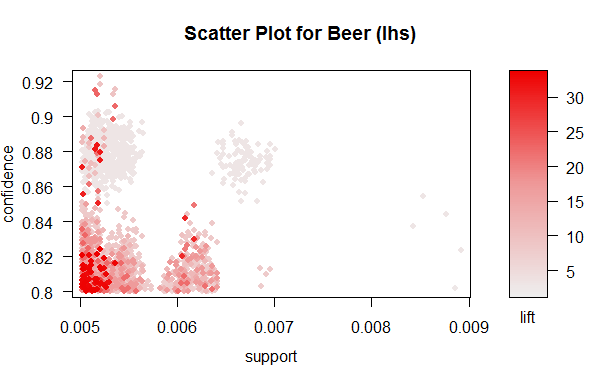
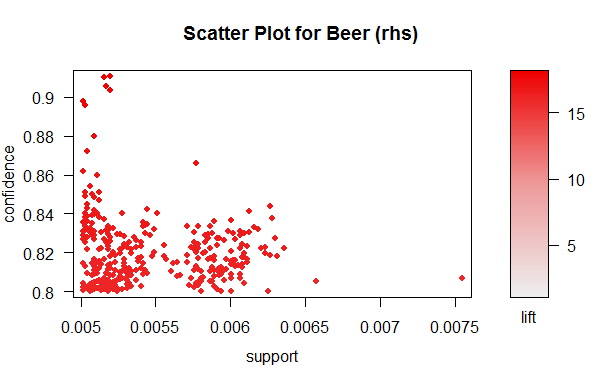
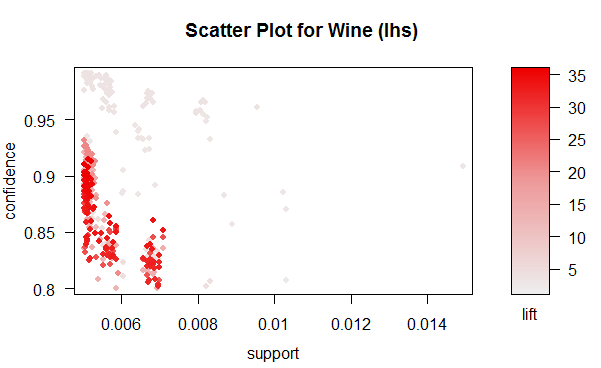
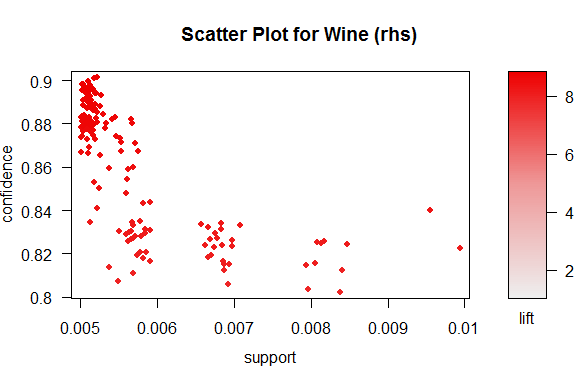
**Purchase patterns related to beverages (Wine, Beer):** Determination of Wine and Beer Association Rule.

|  |  |  |
| --- | --- | --- |
| LHS | RHS | # of Rules |
| Other Items -> | Wine | 162 |
| Wine -> | Other Items | 433 |
| Other Items -> | Beer | 354 |
| Beer -> | Other Items | 2278 |

SCATTER PLOT:

The scatter plot clearly reveals the purchasing pattern of wine and beer. Moreover, the generated rules have a potential usefulness as they have both high confidence and lift and these would play a key role in the making of important business decision.

Both Beer and Wine on the LHS has a better lift value than on RHS.



Below are the top 5 patterns of each rule. It shows how other items are related to Wine and Beer.

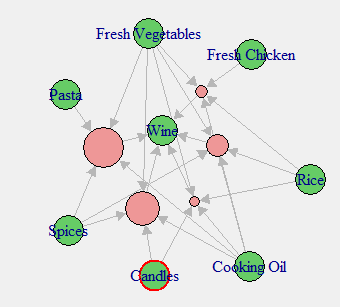
Wine(lhs): If someone is buying wine along with other items like Fresh Vegetables, Spices, Rice, Pasta, Sauces, fresh chicken in various combination, then it clearly implies that he will end up buying candles as well.

Wine(rhs): Similarly, buying Fresh Vegetables, Spices, Rice, Cooking oil, Pasta, spices, candles in various patterns implies that someone will buy wine.

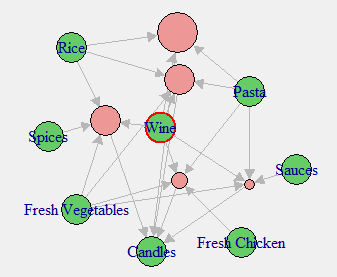
Beer(lhs): Buying Beer with other items like Cereal, Clams, Eggs etc implies buying oysters. It is to be noted that one of the rules implies Sardines as well. Both Oyster and Sardine have the same lift value.

Beer(rhs): Similarly, buying Dried meat, Nuts, Dips etc implies buying Beer.

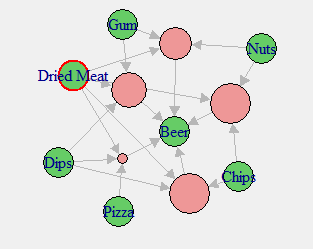
The diagram clearly explains, the top 5 transactions (marked in pink) and its associated items in Green. The implied items are those which has all converging arrows and they are Candles, Wine, Oysters, and Beer.



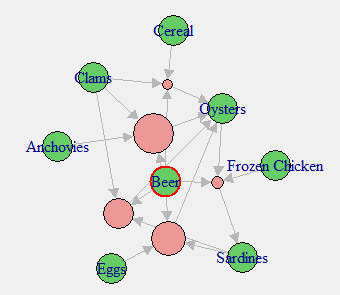
Wine(rhs)



Wine(lhs)



Beer (rhs)



Beer(lhs)

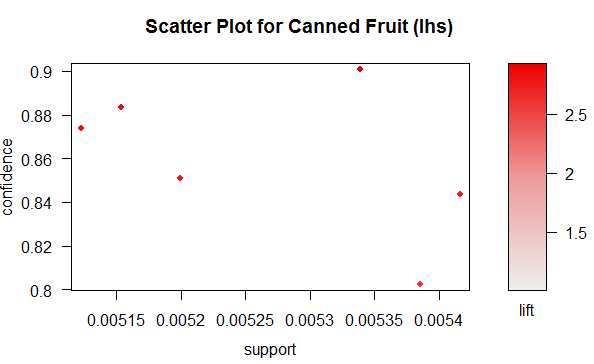
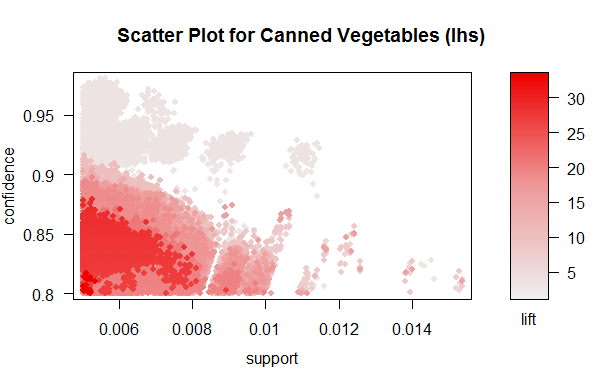
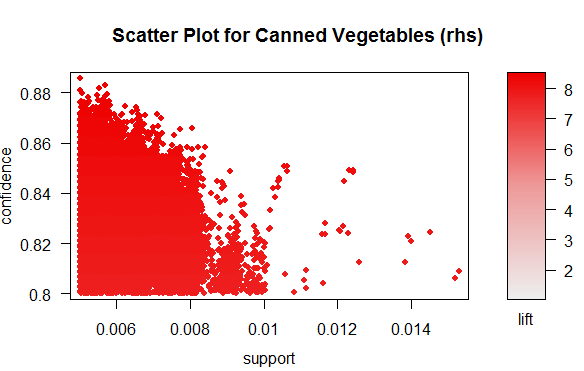
## **Canned vs Fresh**

**Purchase patterns related to canned items vs fresh items:** Determination of canned and fresh item sets association rule.

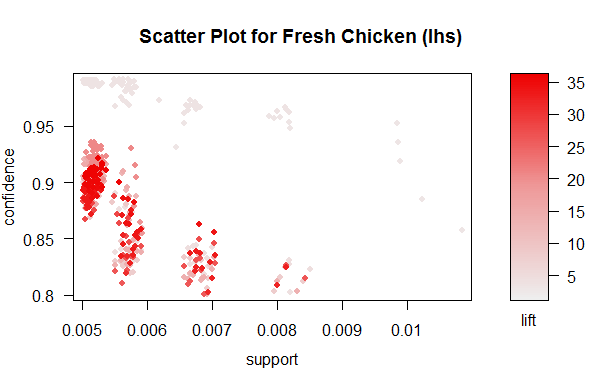
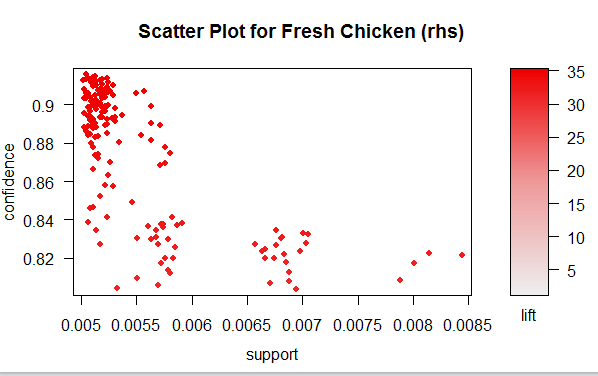
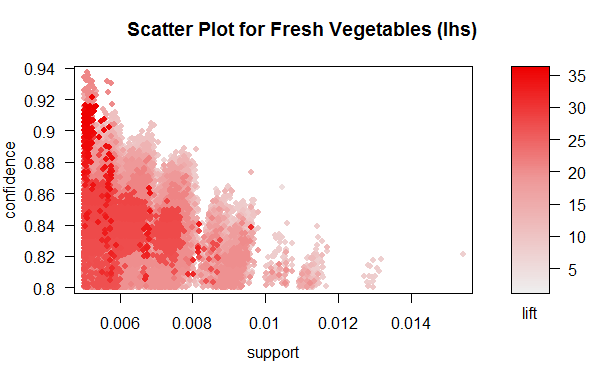
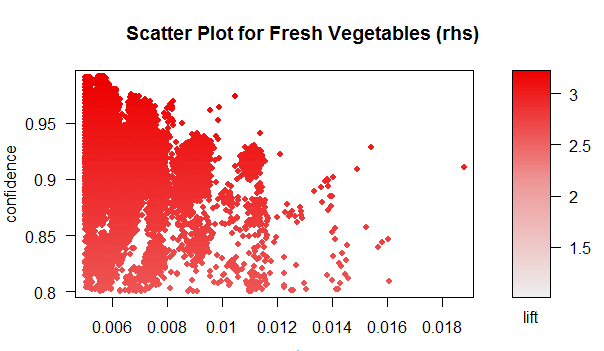
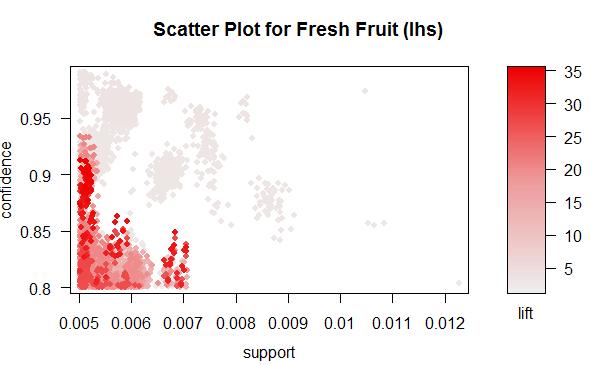
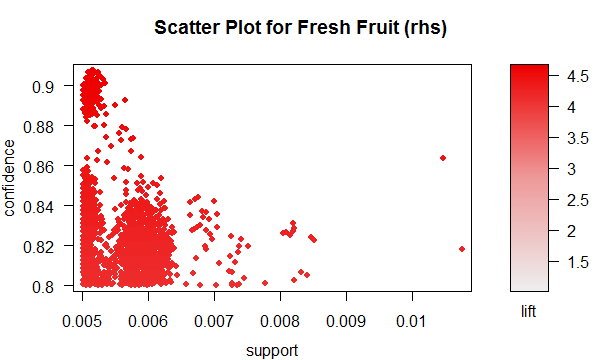
|  |  |  |
| --- | --- | --- |
| LHS | RHS | # of Rules |
| Other Items -> | Canned Vegetables | 35096 |
| Canned Vegetables -> | Other Items | 182631 |
| Other Items -> | Canned Fruit | 0 |
| Canned Fruit -> | Other Items | 6 |
| Other Items -> | Fresh Fruit | 1947 |
| Fresh Fruit -> | Other Items | 6615 |
| Other Items -> | Fresh Vegetables | 59969 |
| Fresh Vegetables -> | Other Items | 266511 |
| Other Items -> | Fresh Fish | 0 |
| Fresh Fish -> | Other Items | 0 |
| Other Items -> | Fresh Chicken | 170 |
| Fresh Chicken -> | Other Items | 604 |

SCATTER PLOT:

The scatter plot clearly reveals the purchasing pattern of canned items. There seem to be a demand of canned vegatables but there is neglible purchase of canned fruit.



The below scatter plot reveals the purchasing pattern of fresh items. It is clear from the plot that the demand of fresh fruit and vegetables is more than the canned items. No association rule was found for fresh fish.



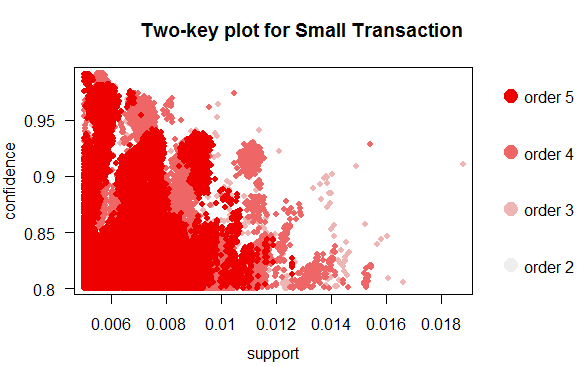
1. **Small Transaction vs Large Transaction**

## **Small vs Large Transaction**

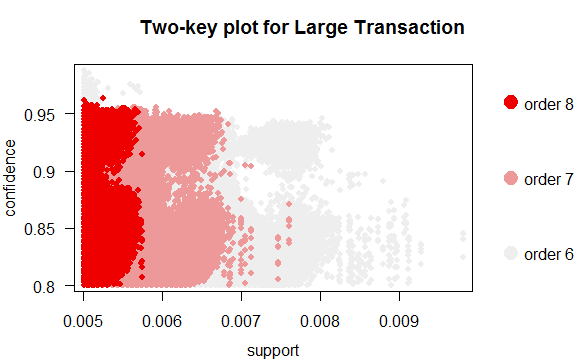
Transactions items less than equal to 5 has been marked as small transaction and greater than 5 as large.

Small Transaction Pattern: # of rules generated for Small Transaction is 114264.

The below graph depicts, the volume of small transactions are more than large transactions.



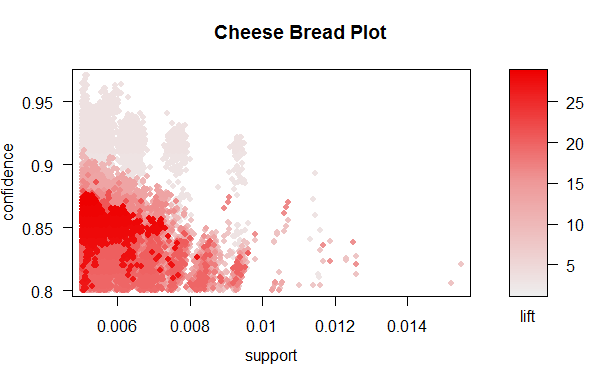
Large Transaction Pattern: # of rules generated for Large Transaction is 619251



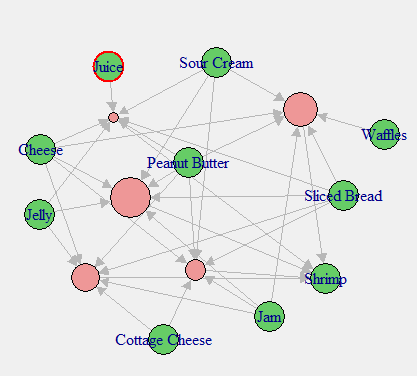
## **Interesting Pattern – Cheese & Sliced Bread**

**Interesting Pattern:** After fresh vegetables and fresh fruit, the most frequent item that occurs is Cheese. We explored and found Cheese along with Sliced Bread create 54036 rules.

Both Cheese and Sliced bread are among the top 20 frequent items. Therefore, if we club the non-frequent items along with these 2 frequent items and offer a bundle price, then the business can increase the number of items sold in a transaction and also increase its revenue. The confidence range is between 80% - 100 % and the lift value is also high.



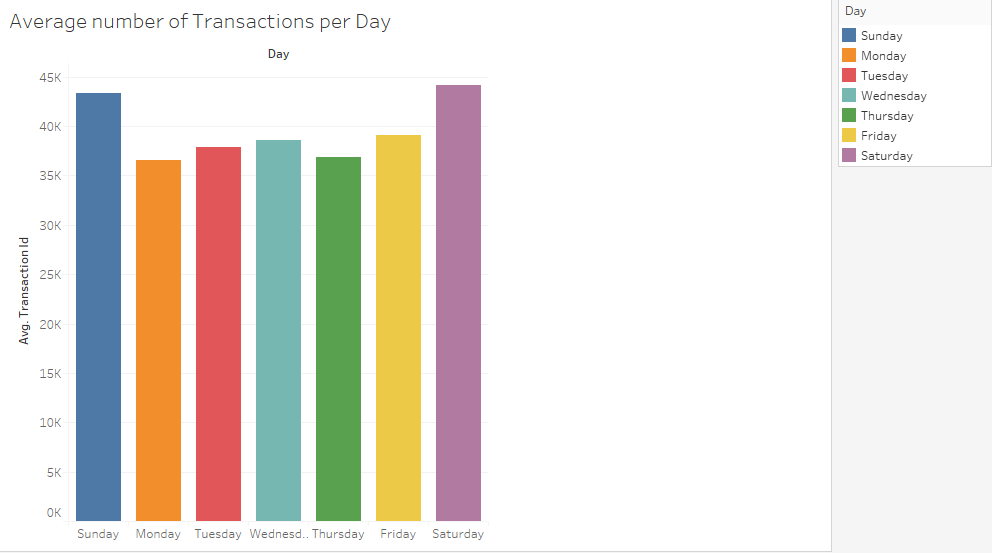
Top 5 patterns having Cheese and Sliced bread.



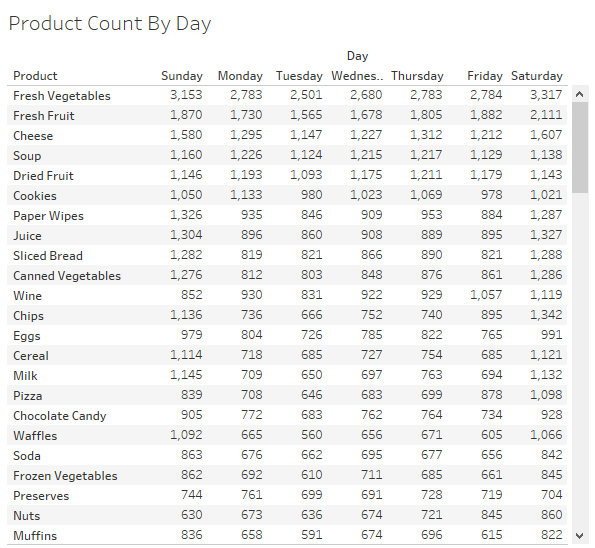
This pattern shows buying Cheese and sliced bread along with other items will imply buying shrimp.

## **Weekend vs Weekday Transaction**

Average number of transactions per day depicts there are more transactions on weekends than weekdays.



The below data shows the number of product transactions on each day and it is highest on weekends.



The boxplot shows the transaction median is more on weekends than weekdays and it also shows the maximum outlier is on weekend (Saturday & Sunday). The above two graphs along with the box plot clearly implies than there is more transaction on weekend than weekday.

