**1. Objectives**

The domain is a bakery store that sells foods and drinks. From association rule mining, they can decide which needs to be promoted, placement of products and recommendation for customers depending on the customers’ previous choices. After that, the datasets are converted into factor so that Apriori algorithm can be applied on it.

**2. Data Set Description**

There are datasets of 1000, 5000, 20000 and 75000 transactions. The transactions are further classified into sparse vector and full binary vector. For pre-processing, because the transaction is in item code format, the item code is matched with its respective names and replaced in the data frame. This is done so that later on the rules are easier to understand. Transaction IDs are ignored because it is not important in here as number of rows is same as number of transactions. The cleaned data frame are then exported into a new csv file to be processed later.

1. **Rule mining process**

|  |  |  |  |
| --- | --- | --- | --- |
| File | MinLen | MinSup | MinConf |
| 1000-out1.csv | 2 | 0.03 | 0.5 |
| 5000-out1.csv | 2 | 0.03 | 0.7 |
| 20000-out1.csv | 2 | 0.03 | 0.7 |
| 75000-out1.csv | 2 | 0.033 | 0.7 |

*Table 1 Parameter Setting*

The parameter setting value shown as above, Table 1. Support of an association rule determines its coverage: What percentage of all market baskets the rule affects. We want to find association rules with high support, because such rules will about transaction/market baskets that commonly occur.

Confidence of an association rule determines its predictability. For example, how often it occurs among the affected market baskets. We want to find association rules with high confidence, because such rules represent strong relation-ships between items.

The choice of algorithm we use is **Apriori** algorithm. Before running the algorithm, we match the item code with the item name. Then we convert the entire column into factors and save it as a new file. This is because we want to read the new file as transaction to perform Apriori algorithm.

Result for 1000 transactions:

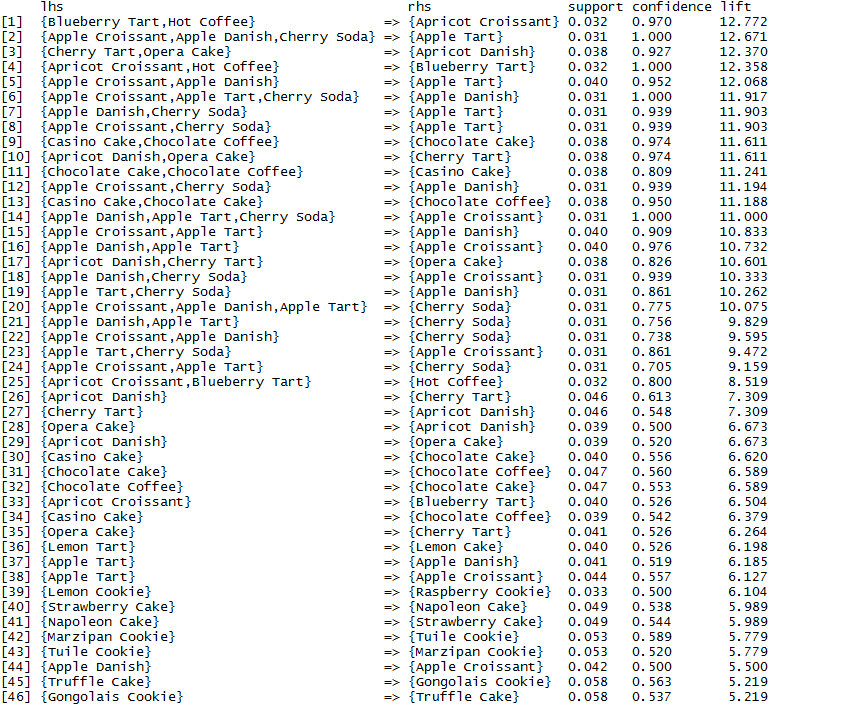


Figure 1 Results after Apriori algorithm for 1000 transactions

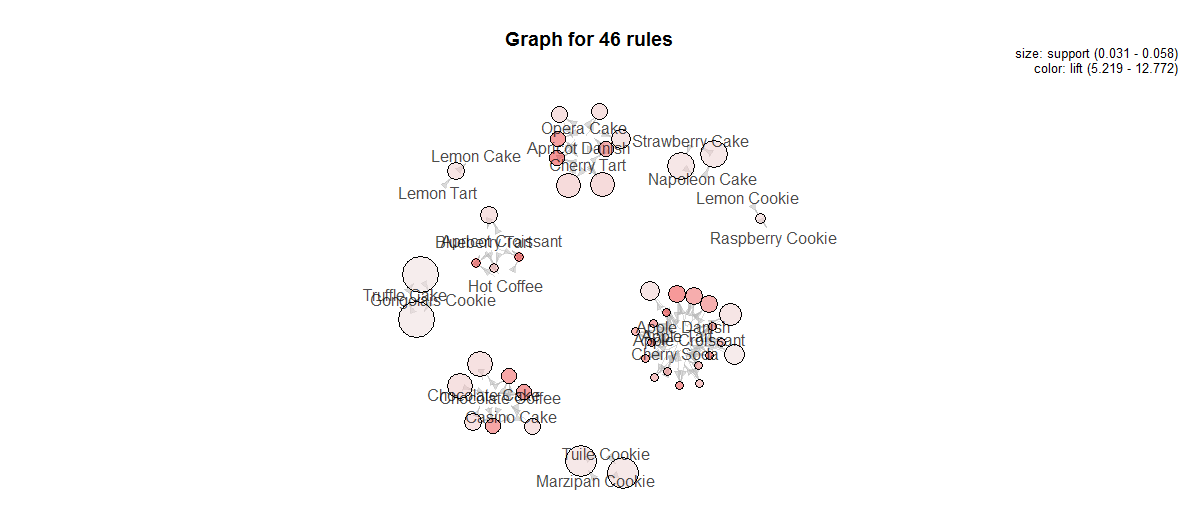


Figure 2 Graph of rules for 1000 transactions

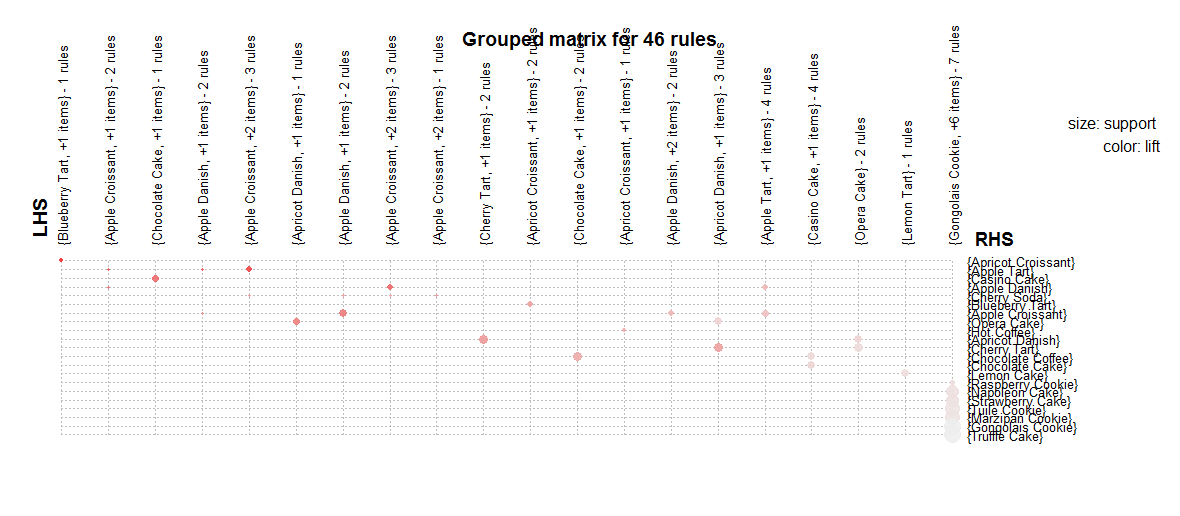


Figure 3 Grouped matrix of rules for 1000 transactions

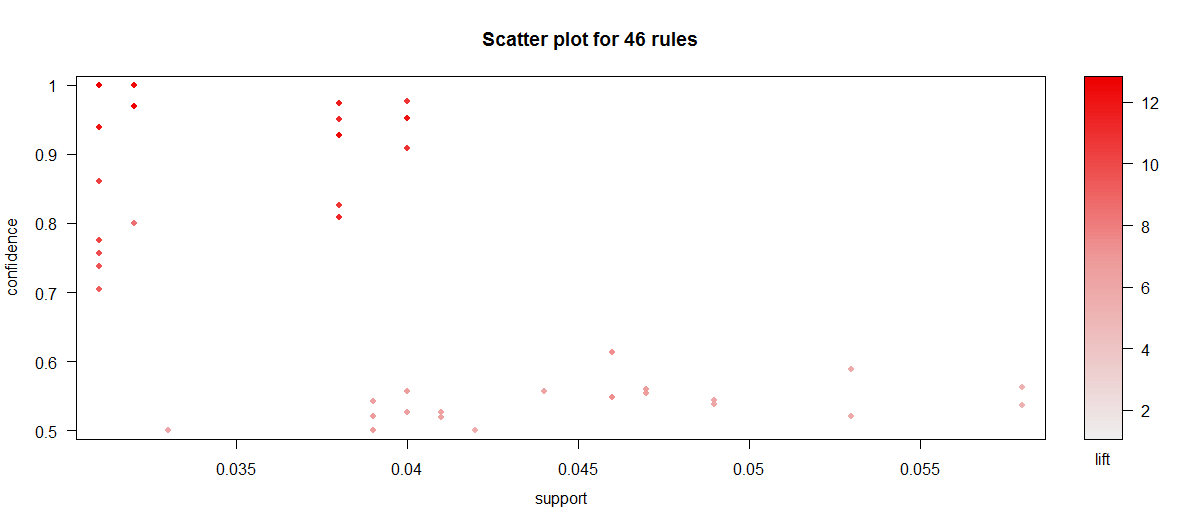


Figure 4 Scatter plot of rules for 1000 transactions

Result for 5000 transactions:

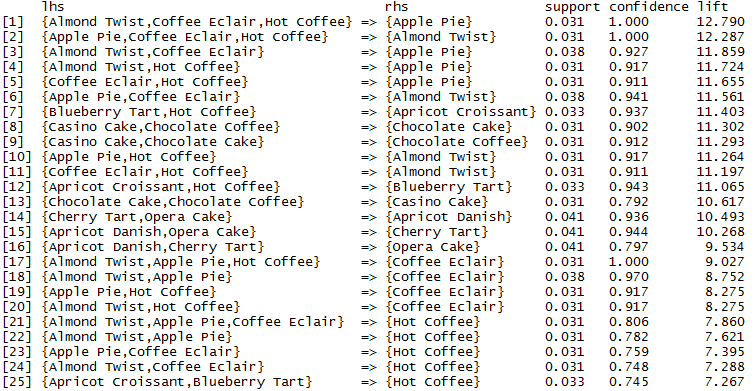


Figure 5 Results after Apriori algorithm for 5000 transactions

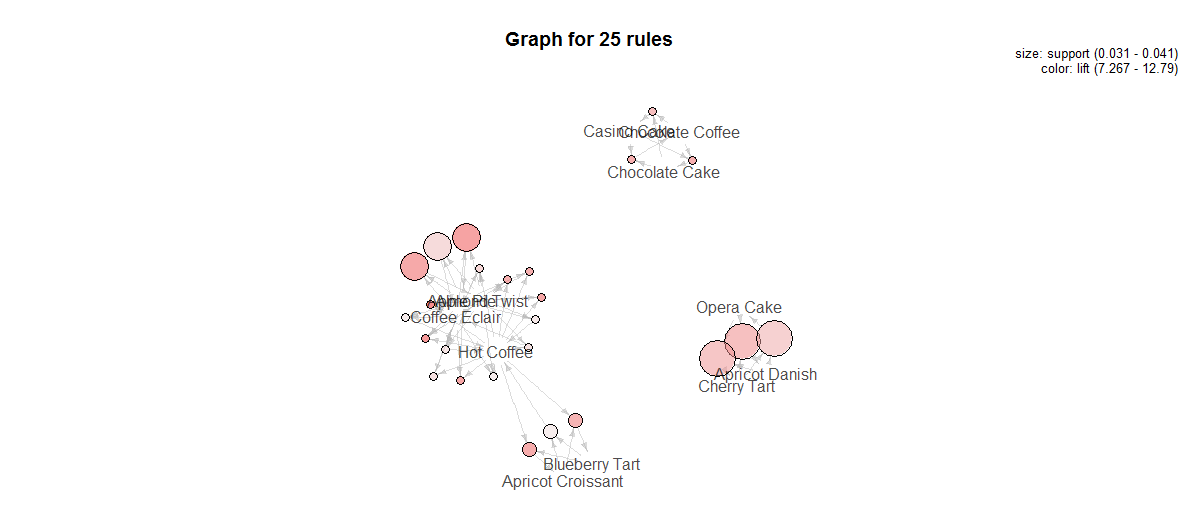


Figure 6 Graph of rules for 5000 transactions

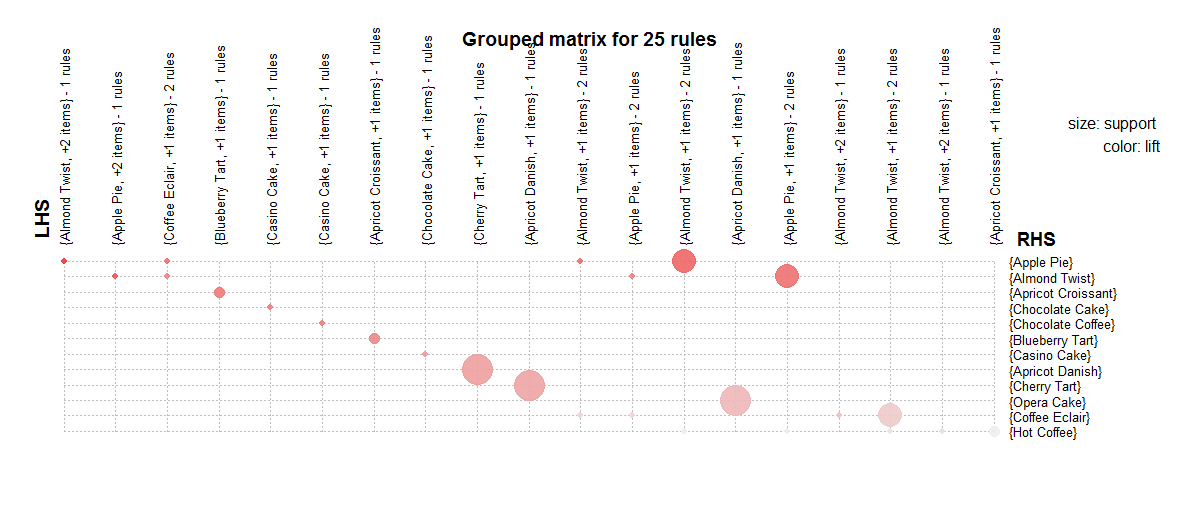


Figure 7 Grouped matrix of rules for 5000 transactions

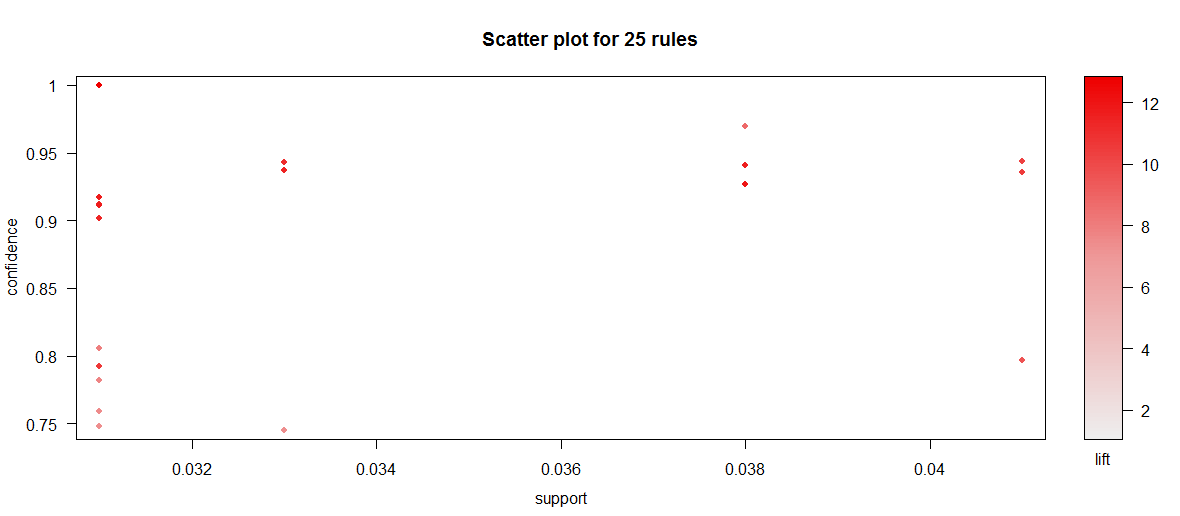


Figure 8 Scatter plot of rules for 5000 transactions

Result for 20000 transactions:

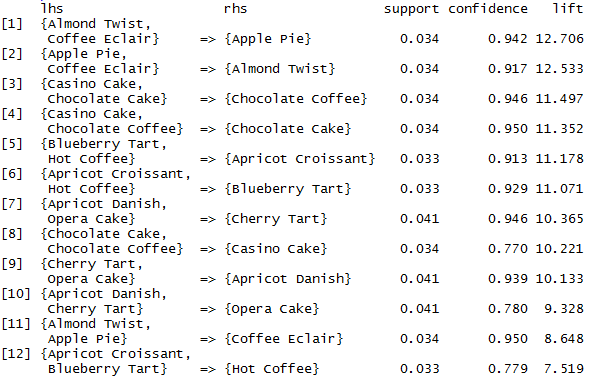


Figure 9 Results after Apriori algorithm for 20000 transactions

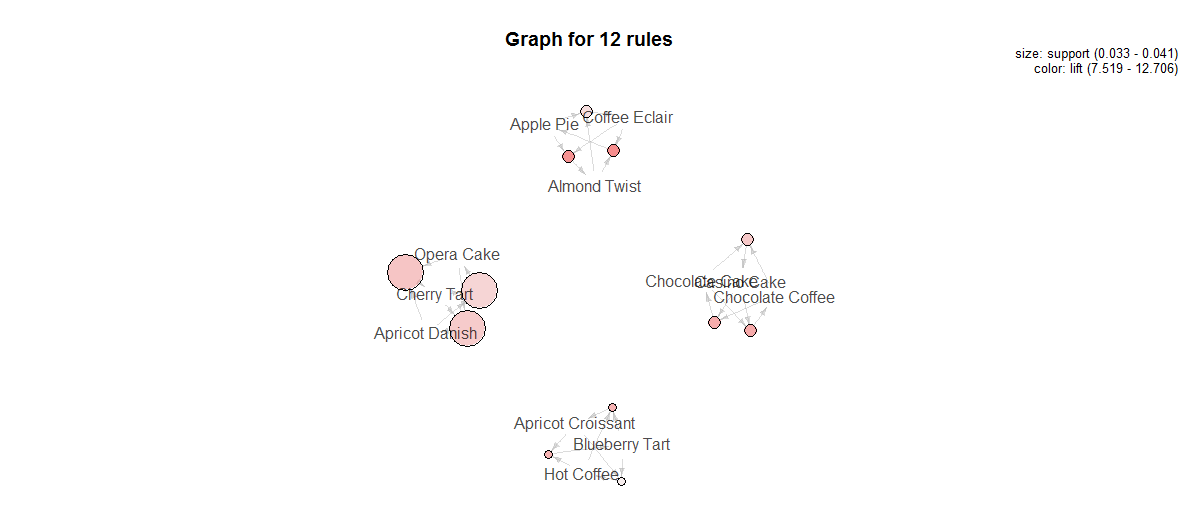


Figure 10 Graph of rules for 20000 transactions

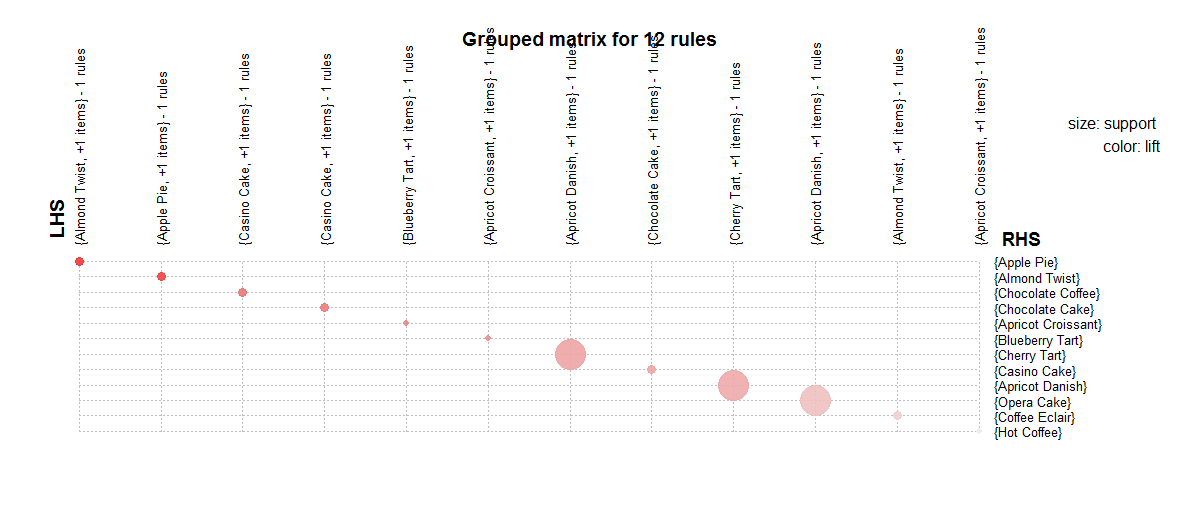


Figure 11 Grouped matrix of rules for 20000 transactions

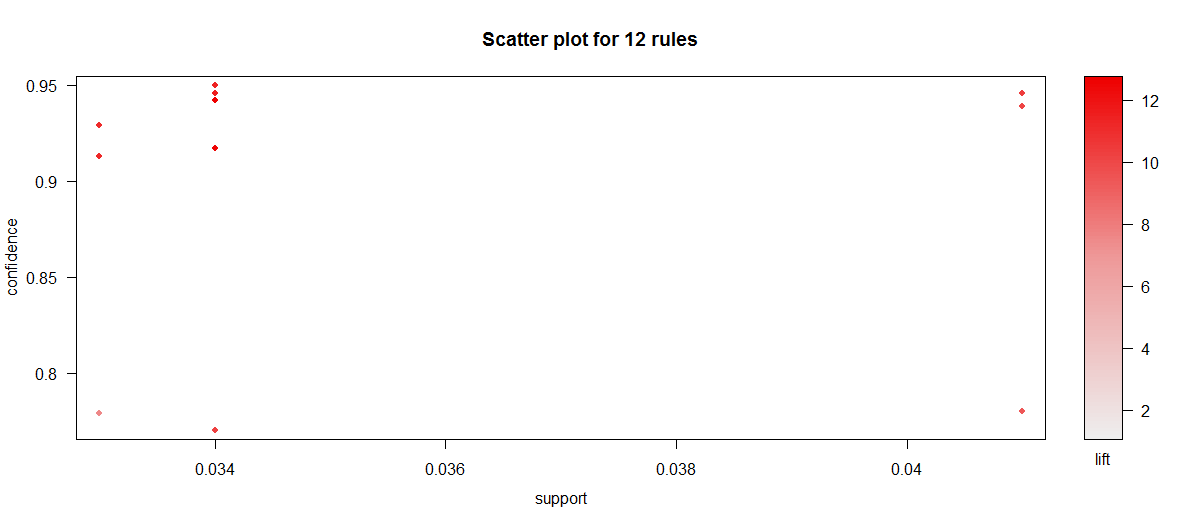


Figure 12 Scatter plot of rules for 20000 transactions

Result for 75000 transactions:

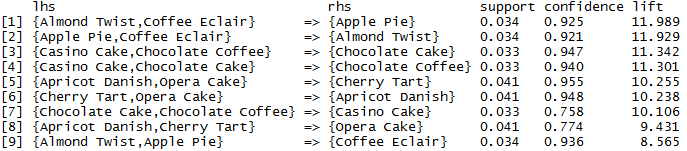


Figure 13 Results after Apriori algorithm for 75000 transactions

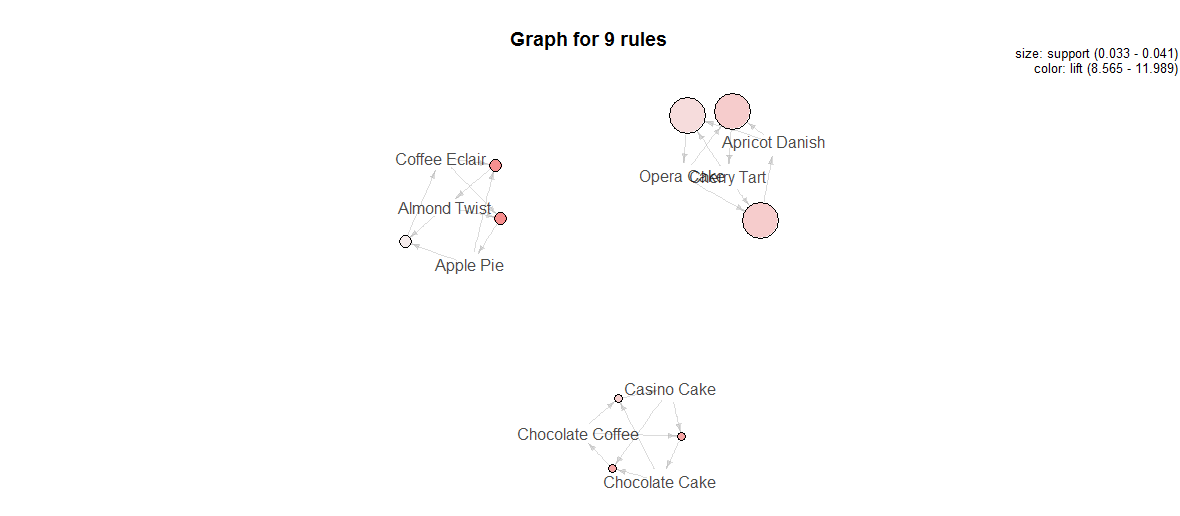


Figure 14 Graph of rules for 75000 transactions

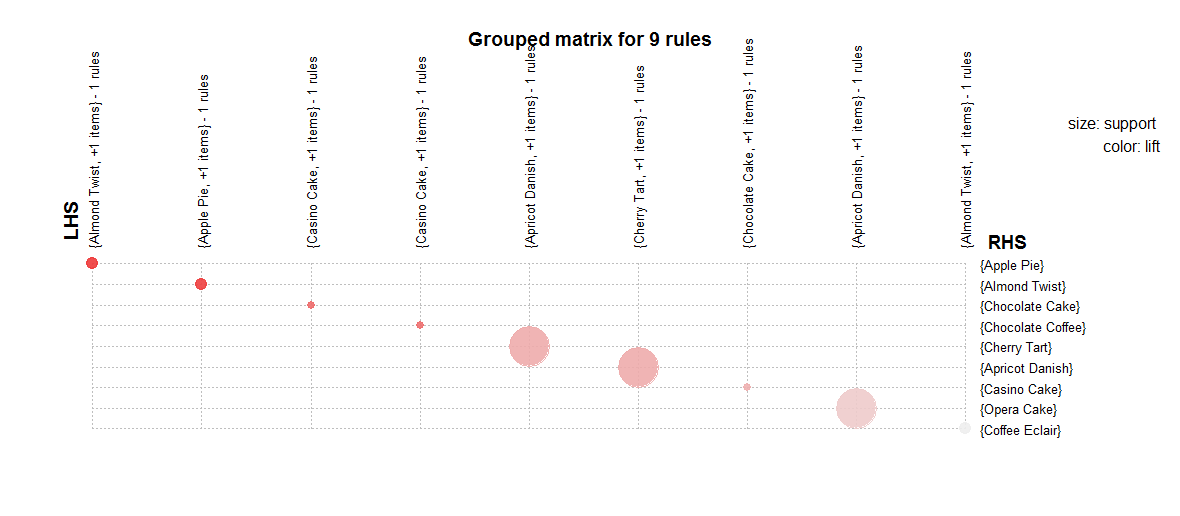


Figure 15 Grouped matrix of rules for 75000 transactions

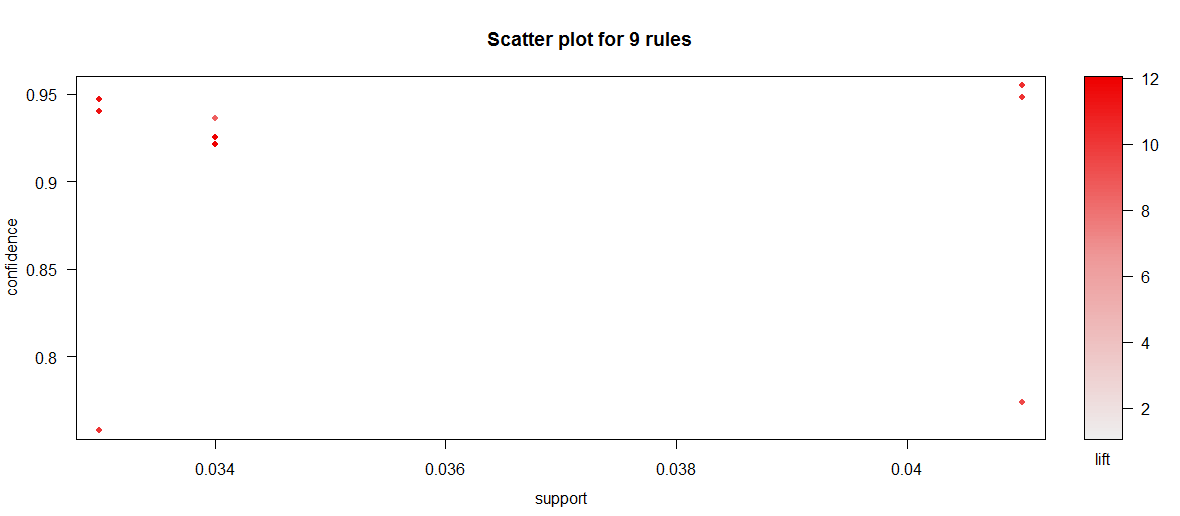


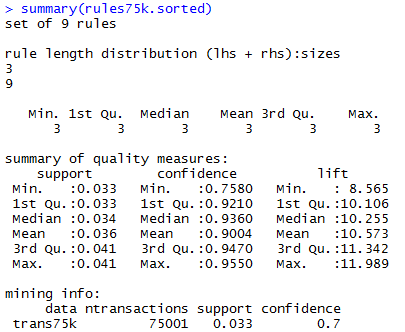
Figure 16 Scatter plot of rules for 75000 transactions

**Time required** running the process recorded as below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | 1000 | 5000 | 20000 | 75000 |
| User time | 2.61 | 3.19 | 5.56 | 15.93 |
| System time | 0.69 | 0.46 | 0.44 | 0.46 |
| Elapsed time | 36.03 | 42.39 | 36.06 | 50.03 |

**4. Resulting rules**

75000 transaction would be chosen as more sample as it will be more accurate. Summary of the association rules for 75000 transaction is shown below:



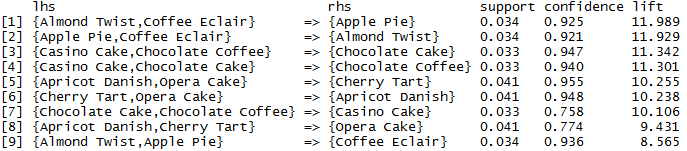


Figure 17 Summary of rules for 75000 transactions

These 9 combination of items will be shown to client as they has the highest support, confidence and lift out of all rules.

**5. Recommendations: What should the client do because of the rules discovered?**

The rules can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements.