Market Basket Analysis on supermarket

BY

K Nikitha Reddy

Problem Statement

Rengas is a supermarket located in Coimbatore, Tamil Nadu. Nowadays people buy daily goods from super market nearby. There are many supermarkets that provide goods to their customer. The problem many retailers face is the placement of the items. They are unaware of the purchasing habits of the customer so they don't know which items should be placed together in their store. With the help of this application shop managers can determine the strong relationships between the items which ultimately helps them to put products that co-occur together close to one another. Also decisions like which item to stock more, cross selling, up selling, store shelf arrangement are determined.

Variables

Variable	Description
Bill Num	Unique Bill id
Date	The date at which item was purchased
Product Id	ID of the product
Items	Item name with SKU
Quantity	Quantity purchased
Price	Price of the item
Total price	Total price (Quantity*Price)
GST	GST percentage charged
Final Price	Final price after taxes

Objectives

The objectives of this market basket analysis are:

- The main objective of Market Basket Analysis is to get better efficiency of market and sales strategy using consumer transactional data collected during the sales transaction.
- To spot the frequent items on or after the transaction on the basis of support and confidence.
- To generate the association rules from the frequent item sets.

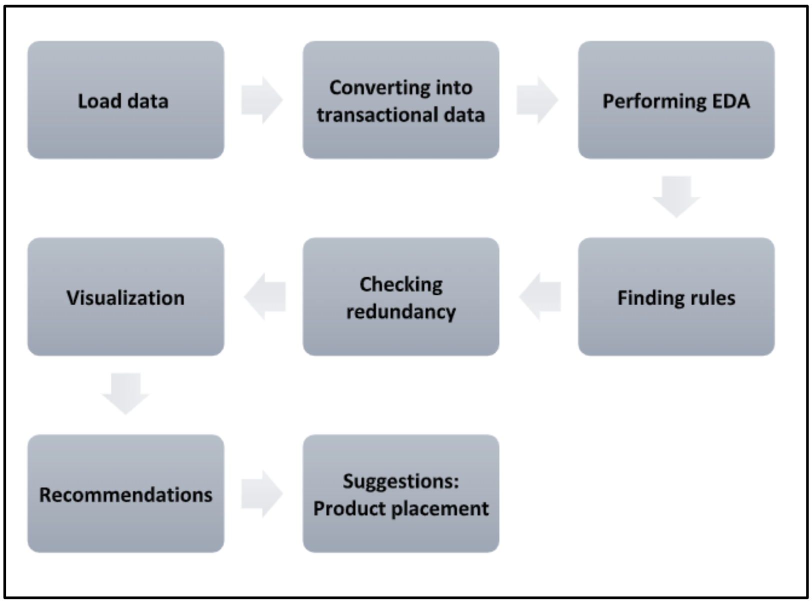
Assumptions

Here are several assumptions that are commonly made when performing market basket analysis:

- **Independence:** Market basket analysis assumes that the items in a transaction are independent of each other. However, in reality, items are often related to each other in some way, and this should be taken into account when interpreting the results.

- **Transactional data:** Market basket analysis is typically performed on transactional data, which contains information about the products that were purchased, the date of the transaction, and the customer who made the purchase.
- **No missing values:** The data is cleaned and preprocessed to remove outliers, errors and missing values. If there are missing values, it can lead to inaccurate analysis.
- **Time frame:** The data is collected within a certain time frame, as the customer behavior can change over time.

<u>Procedure followed</u>



1. DATA UNDERSTANDING

Data Collection

The data collection is primary in nature. It was collected by contacting the superstore's owner and the purpose was clearly stated. The data received was from June 2022 to August 2022. The certificate is attached below:

rengas
Life is beautiful

Coimbatore
24.01.2023

To whomsoever, it may concern,

I am writing this letter to confirm that, A.Rangaswamy Chettiar Sons & Co has approved the use of our point-of-sale (POS) data for the study purpose to Mr. Ajaysuryaa K S and his team, who are pursuing a Master of Business Administration (MBA) at Christ University, Bangalore. We are pleased to be able to support your efforts and believe that our POS system and data will be valuable resources for his work.

If you have any questions or concerns, please do not hesitate to contact us. We look forward to collaborating with you on this project and are confident that our POS system and data will be valuable assets.

Sincerely,

For A. RANGASWAMY CHETTIAR SONS & CO.,
B. Saravanan
PARTNER

A. Rangaswamy Chettiar Sons & Co.
108, D.B. Road, R.S. Puram, Coimbatore - 641 002, Tamil Nadu, INDIA
Tel : +91 422 4033908 | E-mail : rengas37@gmail.com
Web : www.rengas.in | GST : 33AAEFR3142A1ZX
super market

Data Exploration

Structure

```
'data.frame':   855058 obs. of  [illegible] variables:
 $ Billnum: chr  "2206F1S0007404" "2206C5S0007776" "2206C5S0007777" "2206C5S0007778" ...
 $ Date   : Date, format: "2022-06-01" "2022-06-01" "2022-06-01" "2022-06-01" ...
 $ Item   : chr  "TOMATO APPLE" "NEST MUNCH MAHA 25GM" "PLASTOBAG GARBAGE LARGE 15NO" "SAFFRON KESARI 500MG" ...
```

The structure here is changed according to the datatype which signifies that particular column. Since, bill number is alpha-numeric so it is considered as character here. The item is entered with the SKU. The original data had many columns but we extracted the data which is required to us for the analysis. There are total of 8,55,058 observations and 3 variables.

Summary

```
  Billnum                  Date                       Item
 Length:855058      Min.    :2022-06-01   Length:855058
 Class :character   1st Qu.:2022-06-24   Class :character
 Mode  :character   Median :2022-07-16   Mode  :character
                    Mean   :2022-07-16
                    3rd Qu.:2022-08-08
                    Max.   :2022-08-31
```

The summary statics tells us about the mean, median and interquartile range of the numeric variables. We can further explore them using density plots and boxplots.

Assessing Data Quality

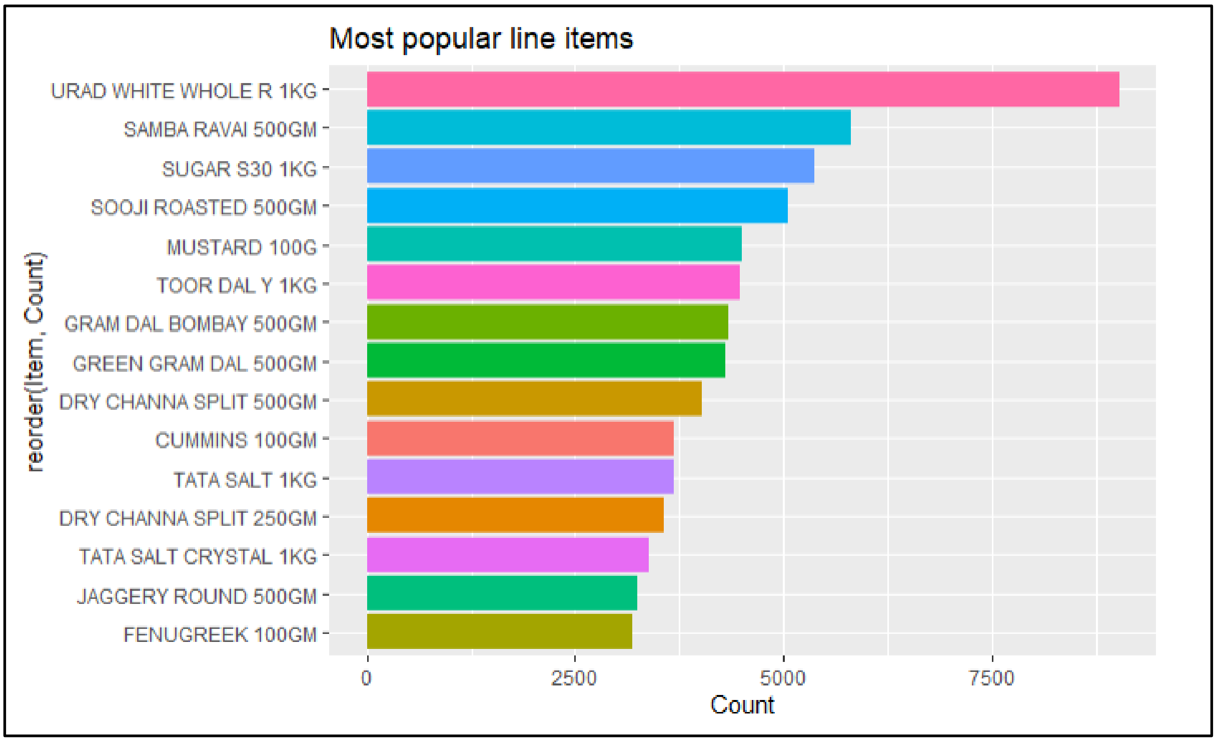
Data completeness, accuracy, consistency, reliability and correct format are all used to determine the quality of the data. After assessing the quality of the data, following comments can be made:

- The data is complete. There are no missing values.
- There are no duplicate values in the dataset.
- The date is in wrong format (character data type). This needs correction which is done.
- The data needs to be converted to transactional data.

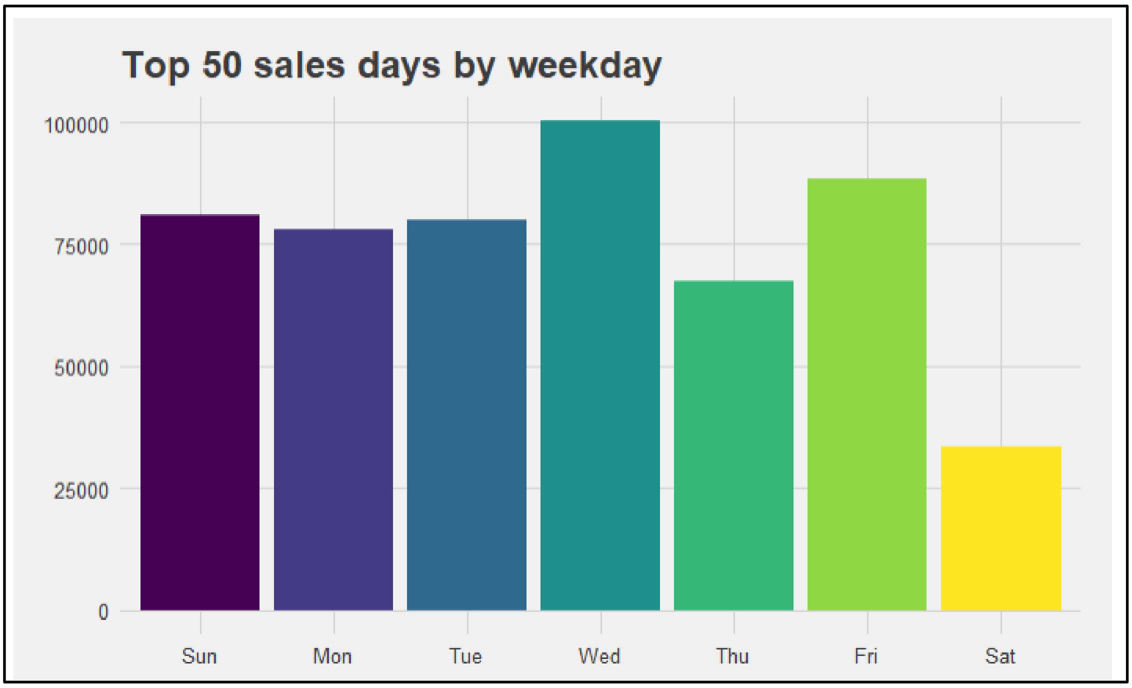
```
#Market Basket Analysis

y <- read.transactions("Rengas_pos.csv",
                       format="single",
                       cols=c(1,3),
                       sep=","
                       )
```

```
## transactions in sparse format with
##  48168 transactions (rows) and
##  17639 items (columns)
```

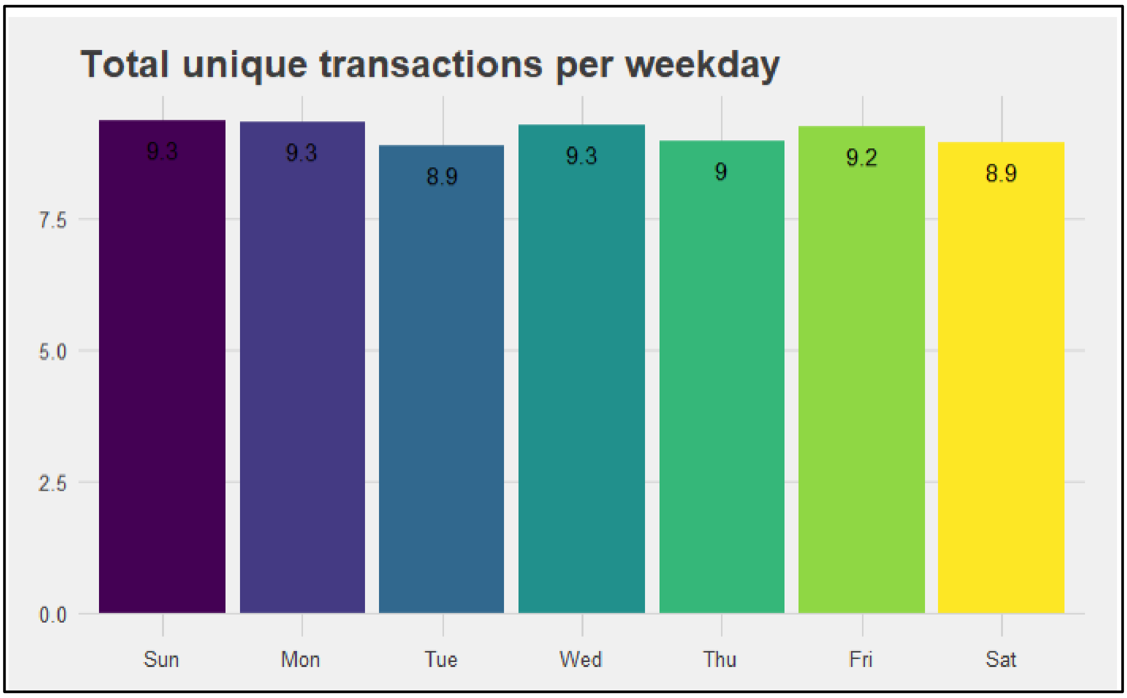
2. EDA



Here we observe that, the most 10 popular line item is Urad white whole R 1Kg. Although the SKU's are changing, we can get to know here that most customers are buying which type of SKU.



Most of the top transaction days 50 transaction days occur only during Wednesday's. Next top 50 transaction days occur during Friday. We observe that, there are few sales days on Saturday.



This gives us the number of unique transactions that takes place every week. The count of unique transactions are almost the same everyday. But this count seems interesting since people are not buying the same item which rest of them bought in the same day.

3. MARKET BASKET ANALYSIS

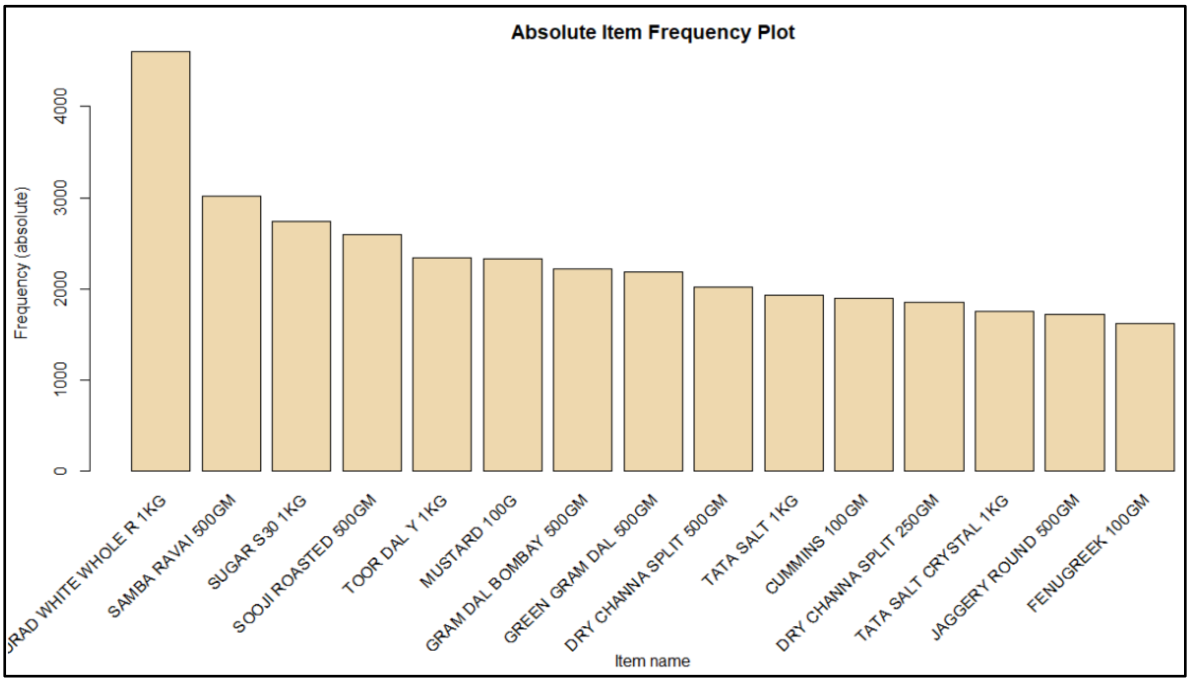
Summary

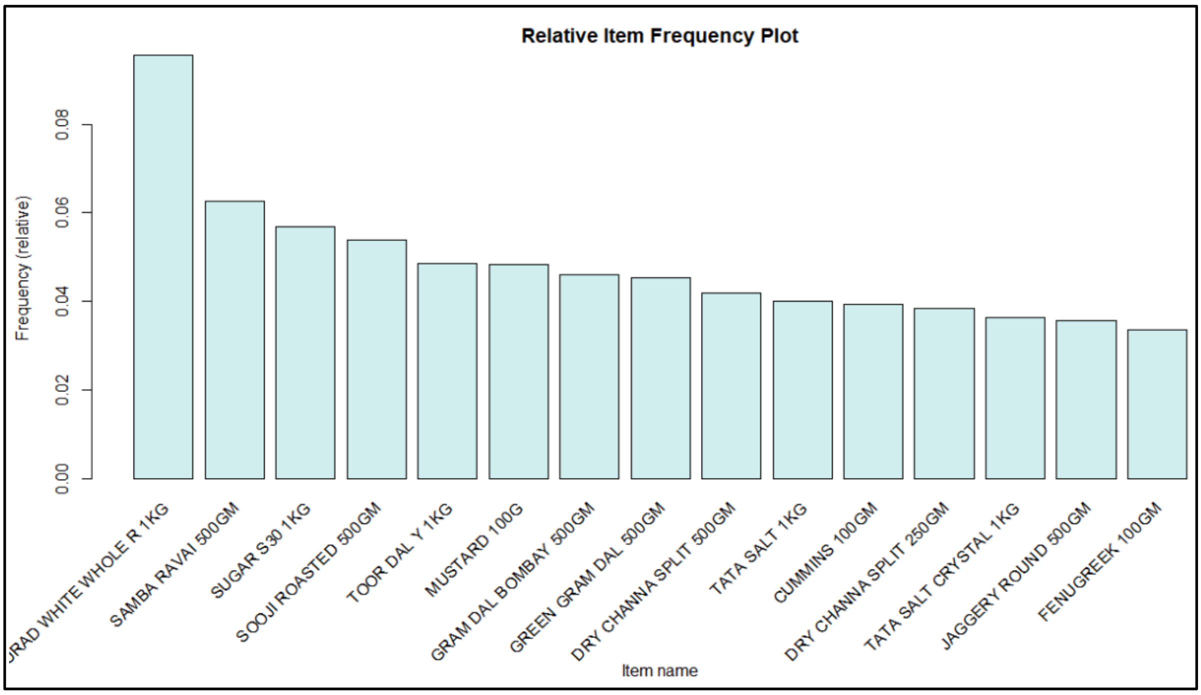
```
## transactions as itemMatrix in sparse format with
##  48168 rows (elements/itemsets/transactions) and
##  17639 columns (items) and a density of 0.0005119914
##
## most frequent items:
## URAD WHITE WHOLE R 1KG      SAMBA RAVAI 500GM          SUGAR S30 1KG
##                   4606                   3020                   2742
##    SOOJI ROASTED 500GM         TOOR DAL Y 1KG                (Other)
##                   2600                   2336                 419702
##
## element (itemset/transaction) length distribution:
## sizes
##     1     2     3     4     5     6     7     8     9    10    11    12    13
## 12807  6662  4322  3172  2489  1926  1641  1363  1142  1010   885   804   682
##    14    15    16    17    18    19    20    21    22    23    24    25    26
##   635   555   508   454   390   360   327   303   291   307   244   235   239
##    27    28    29    30    31    32    33    34    35    36    37    38    39
##   218   231   207   188   184   182   140   148   149   142   142   135   120
##    40    41    42    43    44    45    46    47    48    49    50    51    52
##   127   111   103   112    88    96   103    88    73    67    80    79    53
##    53    54    55    56    57    58    59    60    61    62    63    64    65
##    53    57    46    45    48    54    49    39    34    35    35    27    32
##    66    67    68    69    70    71    72    73    74    75    76    77    78
##    26    20    30    27    24    16    23    14    19    13    16    17    16
##    79    80    81    82    83    84    85    86    87    88    89    90    91
##    15    13    15    11     9    10    10     7     9     5    10     5     9
##    92    93    94    95    96    97    98   100   101   102   103   104   105
##     5     6     8     7     4     6     3     3     3     4     5     3     2
##   106   107   108   109   110   111   112   113   114   115   118   120   121
##     2     4     2     1     3     2     4     4     3     2     2     3     2
##   122   123   125   127   128   137   138   139   148   154   156
##     1     2     1     1     1     1     2     1     1     1     1
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   4.000   9.031  10.000 156.000
##
## includes extended item information - examples:
##                           labels
## 1      24 MANTRA 7GRAIN ATTA 1KG
## 2 24 MANTRA BENGAL GRAM DAL 500G
## 3    24 MANTRA BROWN CHANNA 500G
##
## includes extended transaction information - examples:
##    transactionID
## 1 2201C9R0008703
## 2 2201C9R0008704
## 3 2201C9R0008709
```

We see our most frequent items agree with what we found above. And that most transactions are either one or two items. If we want to visualize the frequent items using arules we can. We need to import our transaction data, and we can go ahead and reaffirm our previous findings by using the item Frequency Plot function to find the top 10 most popular.

```
# Glimpse
glimpse(trans)
```

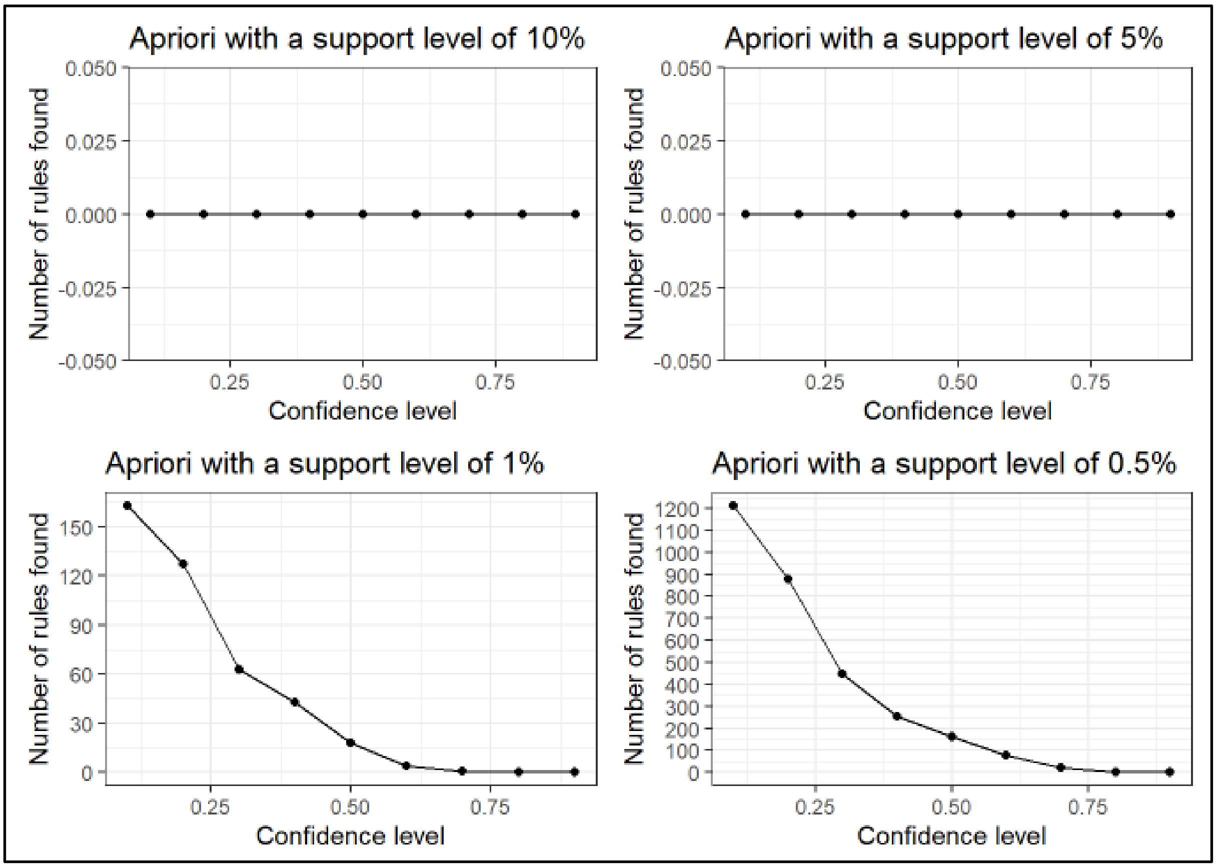
```
## Formal class 'transactions' [package "arules"] with 3 slots
##   ..@ data       :Formal class 'ngCMatrix' [package "Matrix"] with 5 slots
##   ..@ itemInfo   :'data.frame':  17639 obs. of  1 variable:
##   .. ..$ labels: chr [1:17639] "24 MANTRA 7GRAIN ATTA 1KG" "24 MANTRA BENGAL GRAM DAL 500G" "24 MANTRA BROWN C
HANNA 500G" "24 MANTRA BURA SUGAR 500G" ...
##   ..@ itemsetInfo:'data.frame':  48168 obs. of  1 variable:
##   .. ..$ transactionID: chr [1:48168] "2201C9R0008703" "2201C9R0008704" "2201C9R0008709" "2201C9R0008710" ...
```



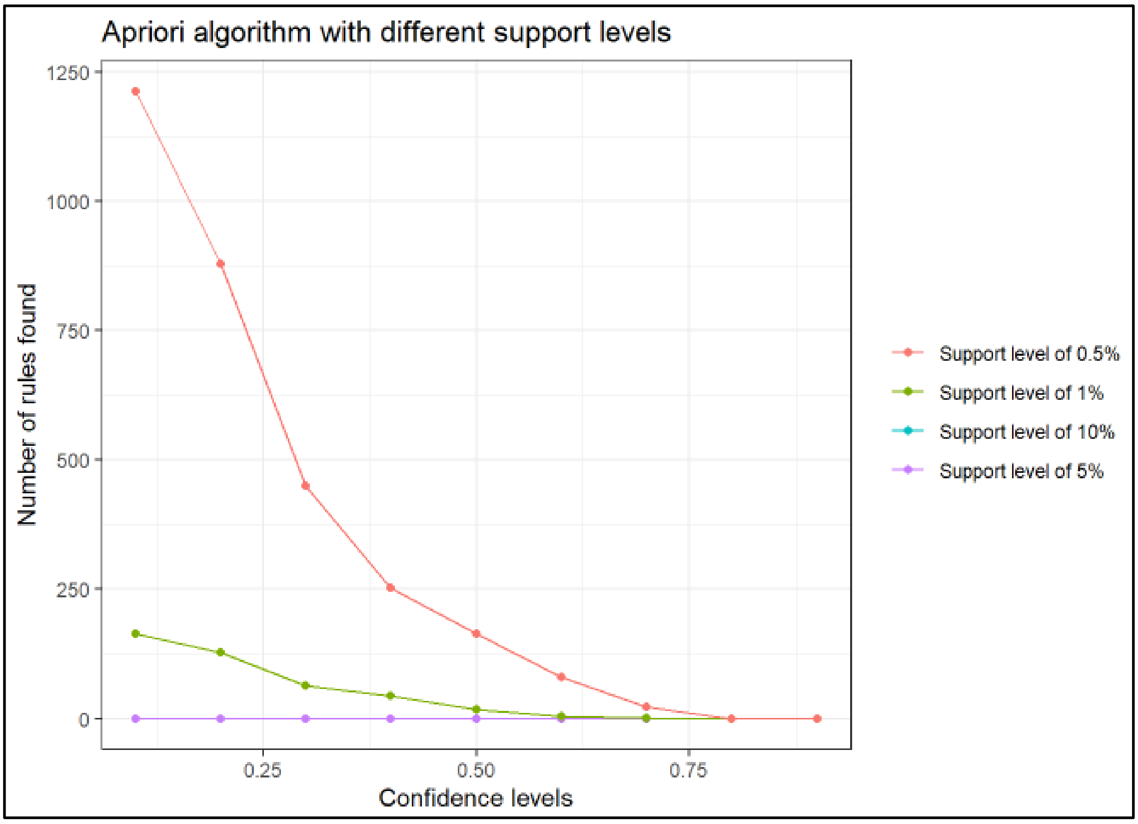


Here, both types of plots are used to visualize the item frequency data, but the AIF plot is used to understand the popularity of the items and RIF plot is used to understand the association between items by showing the relative popularity.

For example, if an AIF plot shows that urad white whole has a frequency of 5625 and samba ravai has a frequency of 3000, it's clear that urad white whole is more popular than samba ravai. On the other hand, if an RIF plot shows first has 11% than second which is 6% it's clear that the popularity of first item is 1.83 times more than second item.



The first step in order to create a set of association rules is to determine the optimal thresholds for support and confidence.



To analyze the results,

- **Support level of 10%**. We only identify no rules with no confidence levels. This means that there are no relatively frequent associations in our data set. We can't choose this value, the resulting rules are unrepresentative and has no rules.

- **Support level of 5%**. We only identify no rules with no confidence levels. It seems that we have to look for support levels below 5% to obtain a number of rules with a reasonable confidence than no rules at all.

- **Support level of 1%**. We started to get dozens of rules, of which 18 have a confidence of at least 50%.

- **Support level of 0.5%**. Here we observe that there are too many rules to analyze.

To sum up, we are going to use a support level of 1% and a confidence level of 50%.

<u>Analyzing Metrics</u>

Now before we perform our Market basket analysis, we need to determine reasonable values for the support and confidence values. This can be obtained by doing a trail and error on the metrics above to see how many rules it generates. To understand the metrics we have:

Support: An itemset with high support means it appears in our transactions frequently. High confidence is directly related to the frequency an item appears. For example, urad white whole 1kg probably has very high support.

Confidence: It is the strength of our association rule. For example, a confidence of 1 implies that when the LHS item is purchased, the RHS item is purchased 100% of the time.

Lift: It is the ratio of Confidence to Expected Confidence. A lift ratio larger than 1.0 implies the relationship between LHS and RHS is significant and not simply chance. The larger the lift ratio, the more significant the relationship between LHS and RHS.

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
2.076e-05 4.152e-05 8.304e-05 5.120e-04 2.699e-04 9.562e-02
[1] 1
```

Here we observe that the maximum support allowed is given by maximum frequency. Hence from the above graphs and this summary stats we take the support value of 1% and a confidence level of 50% from the graphs show above.

Assigning rules

```
##      lhs                           rhs                          support confidence   coverage      lift count
## [1]  {TOOR DAL Y 1KG}          => {URAD WHITE WHOLE R 1KG} 0.02943863  0.6070205 0.04849693  6.348017  1418
## [2]  {GRAM DAL BOMBAY 500GM}   => {URAD WHITE WHOLE R 1KG} 0.02360488  0.5137822 0.04594336  5.372962  1137
## [3]  {DRY CHANNA SPLIT 500GM} => {URAD WHITE WHOLE R 1KG} 0.02188175  0.5230769 0.04183275  5.470163  1054
## [4]  {TATA SALT CRYSTAL 1KG}   => {URAD WHITE WHOLE R 1KG} 0.01860156  0.5128792 0.03626889  5.363519   896
## [5]  {FENUGREEK 100GM}         => {MUSTARD 100G}           0.01723136  0.5126621 0.03361153 10.607350   830
## [6]  {GREEN GRAM DAL 1KG}      => {URAD WHITE WHOLE R 1KG} 0.01550822  0.6088020 0.02547334  6.366646   747
## [7]  {MUSTARD 250G}            => {URAD WHITE WHOLE R 1KG} 0.01507225  0.5377778 0.02802691  5.623899   726
## [8]  {TOOR DAL AG 1KG}         => {URAD WHITE WHOLE R 1KG} 0.01432486  0.5740433 0.02495433  6.003151   690
## [9]  {SOOJI ROASTED 1KG}       => {URAD WHITE WHOLE R 1KG} 0.01403421  0.5192012 0.02703039  5.429632   676
## [10] {PEPPER 100GM}            => {URAD WHITE WHOLE R 1KG} 0.01372280  0.5208826 0.02634529  5.447215   661
## [11] {BUSH BEANS}              => {CARROT}                 0.01316227  0.7684848 0.01712755 27.562456   634
## [12] {CUMMINS 250GM}           => {URAD WHITE WHOLE R 1KG} 0.01307922  0.5440415 0.02404086  5.689403   630
## [13] {FENUGREEK 250gm}         => {URAD WHITE WHOLE R 1KG} 0.01222804  0.5349682 0.02285750  5.594518   589
## [14] {SOOJI ROASTED 500GM,
##       URAD WHITE WHOLE R 1KG} => {SAMBA RAVAI 500GM}      0.01185434  0.5306691 0.02233848  8.463997   571
## [15] {SAMBA RAVAI 500GM,
##       URAD WHITE WHOLE R 1KG} => {SOOJI ROASTED 500GM}    0.01185434  0.5080071 0.02333499  9.411418   571
## [16] {SUGAR S30 1KG,
##       TOOR DAL Y 1KG}         => {URAD WHITE WHOLE R 1KG} 0.01017273  0.6960227 0.01461551  7.278772   490
## [17] {CUMMINS 100GM,
##       MUSTARD 100G}           => {URAD WHITE WHOLE R 1KG} 0.01013121  0.5589920 0.01812407  5.845750   488
## [18] {CUMMINS 100GM,
##       URAD WHITE WHOLE R 1KG} => {MUSTARD 100G}           0.01013121  0.5197018 0.01949427 10.753006   488
```

By sorting through **support**, we see that Toor dal, Gram dal bombay, Dry channa split and tata salt crystal are most commonly bought in tandem with Urad white whole R 1kg. Oddly, pepper cummins and sugar is also frequently bought with Urad white whole. We can only assume that this is due to one person buying for another.

```
##      lhs                           rhs                          support confidence   coverage      lift count
## [1]  {BUSH BEANS}              => {CARROT}                 0.01316227  0.7684848 0.01712755 27.562456   634
## [2]  {SUGAR S30 1KG,
##       TOOR DAL Y 1KG}         => {URAD WHITE WHOLE R 1KG} 0.01017273  0.6960227 0.01461551  7.278772   490
## [3]  {GREEN GRAM DAL 1KG}      => {URAD WHITE WHOLE R 1KG} 0.01550822  0.6088020 0.02547334  6.366646   747
## [4]  {TOOR DAL Y 1KG}          => {URAD WHITE WHOLE R 1KG} 0.02943863  0.6070205 0.04849693  6.348017  1418
## [5]  {TOOR DAL AG 1KG}         => {URAD WHITE WHOLE R 1KG} 0.01432486  0.5740433 0.02495433  6.003151   690
## [6]  {CUMMINS 100GM,
##       MUSTARD 100G}           => {URAD WHITE WHOLE R 1KG} 0.01013121  0.5589920 0.01812407  5.845750   488
## [7]  {CUMMINS 250GM}           => {URAD WHITE WHOLE R 1KG} 0.01307922  0.5440415 0.02404086  5.689403   630
## [8]  {MUSTARD 250G}            => {URAD WHITE WHOLE R 1KG} 0.01507225  0.5377778 0.02802691  5.623899   726
## [9]  {FENUGREEK 250gm}         => {URAD WHITE WHOLE R 1KG} 0.01222804  0.5349682 0.02285750  5.594518   589
## [10] {SOOJI ROASTED 500GM,
##       URAD WHITE WHOLE R 1KG} => {SAMBA RAVAI 500GM}      0.01185434  0.5306691 0.02233848  8.463997   571
## [11] {DRY CHANNA SPLIT 500GM} => {URAD WHITE WHOLE R 1KG} 0.02188175  0.5230769 0.04183275  5.470163  1054
## [12] {PEPPER 100GM}            => {URAD WHITE WHOLE R 1KG} 0.01372280  0.5208826 0.02634529  5.447215   661
## [13] {CUMMINS 100GM,
##       URAD WHITE WHOLE R 1KG} => {MUSTARD 100G}           0.01013121  0.5197018 0.01949427 10.753006   488
## [14] {SOOJI ROASTED 1KG}       => {URAD WHITE WHOLE R 1KG} 0.01403421  0.5192012 0.02703039  5.429632   676
## [15] {GRAM DAL BOMBAY 500GM}   => {URAD WHITE WHOLE R 1KG} 0.02360488  0.5137822 0.04594336  5.372962  1137
## [16] {TATA SALT CRYSTAL 1KG}   => {URAD WHITE WHOLE R 1KG} 0.01860156  0.5128792 0.03626889  5.363519   896
## [17] {FENUGREEK 100GM}         => {MUSTARD 100G}           0.01723136  0.5126621 0.03361153 10.607350   830
## [18] {SAMBA RAVAI 500GM,
##       URAD WHITE WHOLE R 1KG} => {SOOJI ROASTED 500GM}    0.01185434  0.5080071 0.02333499  9.411418   571
```

More interesting are the association rules. Sorting by **confidence** gives us items on the LHS that give a high chance of being bought together with the item on the RHS. When looking at purchases that contain Bush beans, 76% of them also contain Carrot. Considering the support or count values, the association of {sugar s30 1kg, toor dal y 1kg}=> {urad white whole r 1kg} may be more useful.

```
##      lhs                          rhs                          support confidence   coverage      lift count
## [1]  {BUSH BEANS}              => {CARROT}                  0.01316227  0.7684848 0.01712755 27.562456   634
## [2]  {CUMMINS 100GM,
##       URAD WHITE WHOLE R 1KG} => {MUSTARD 100G}             0.01013121  0.5197018 0.01949427 10.753006   488
## [3]  {FENUGREEK 100GM}         => {MUSTARD 100G}             0.01723136  0.5126621 0.03361153 10.607350   830
## [4]  {SAMBA RAVAI 500GM,
##       URAD WHITE WHOLE R 1KG} => {SOOJI ROASTED 500GM}      0.01185434  0.5080071 0.02333499  9.411418   571
## [5]  {SOOJI ROASTED 500GM,
##       URAD WHITE WHOLE R 1KG} => {SAMBA RAVAI 500GM}        0.01185434  0.5306691 0.02233848  8.463997   571
## [6]  {SUGAR S30 1KG,
##       TOOR DAL Y 1KG}         => {URAD WHITE WHOLE R 1KG} 0.01017273  0.6960227 0.01461551  7.278772   490
## [7]  {GREEN GRAM DAL 1KG}      => {URAD WHITE WHOLE R 1KG} 0.01550822  0.6088020 0.02547334  6.366646   747
## [8]  {TOOR DAL Y 1KG}          => {URAD WHITE WHOLE R 1KG} 0.02943863  0.6070205 0.04849693  6.348017  1418
## [9]  {TOOR DAL AG 1KG}         => {URAD WHITE WHOLE R 1KG} 0.01432486  0.5740433 0.02495433  6.003151   690
## [10] {CUMMINS 100GM,
##       MUSTARD 100G}           => {URAD WHITE WHOLE R 1KG} 0.01013121  0.5589920 0.01812407  5.845750   488
## [11] {CUMMINS 250GM}           => {URAD WHITE WHOLE R 1KG} 0.01307922  0.5440415 0.02404086  5.689403   630
## [12] {MUSTARD 250G}            => {URAD WHITE WHOLE R 1KG} 0.01507225  0.5377778 0.02802691  5.623899   726
## [13] {FENUGREEK 250gm}         => {URAD WHITE WHOLE R 1KG} 0.01222804  0.5349682 0.02285750  5.594518   589
## [14] {DRY CHANNA SPLIT 500GM} => {URAD WHITE WHOLE R 1KG} 0.02188175  0.5230769 0.04183275  5.470163  1054
## [15] {PEPPER 100GM}            => {URAD WHITE WHOLE R 1KG} 0.01372280  0.5208826 0.02634529  5.447215   661
## [16] {SOOJI ROASTED 1KG}       => {URAD WHITE WHOLE R 1KG} 0.01403421  0.5192012 0.02703039  5.429632   676
## [17] {GRAM DAL BOMBAY 500GM}  => {URAD WHITE WHOLE R 1KG} 0.02360488  0.5137822 0.04594336  5.372962  1137
## [18] {TATA SALT CRYSTAL 1KG}  => {URAD WHITE WHOLE R 1KG} 0.01860156  0.5128792 0.03626889  5.363519   896
```

When sorting rules by **lift ratio**, the goal is to identify the rules with the highest lift ratios, as these are the rules that have the strongest associations. For example, bush beans with a lift ratio of 27.56 is found, it means that the occurrence of this item in the rule is 27 times more likely if the carrot is also present in the same transaction.

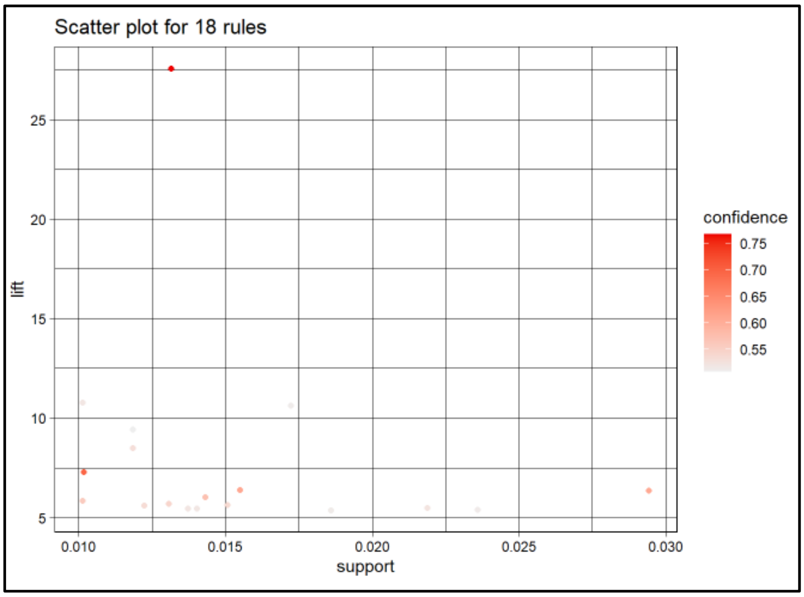
Checking Of Finding Redundancy

```
rules.pruned<-rules[!redundant]
rules.pruned<-sort(rules.pruned,by="lift")
inspect(rules.pruned)
```

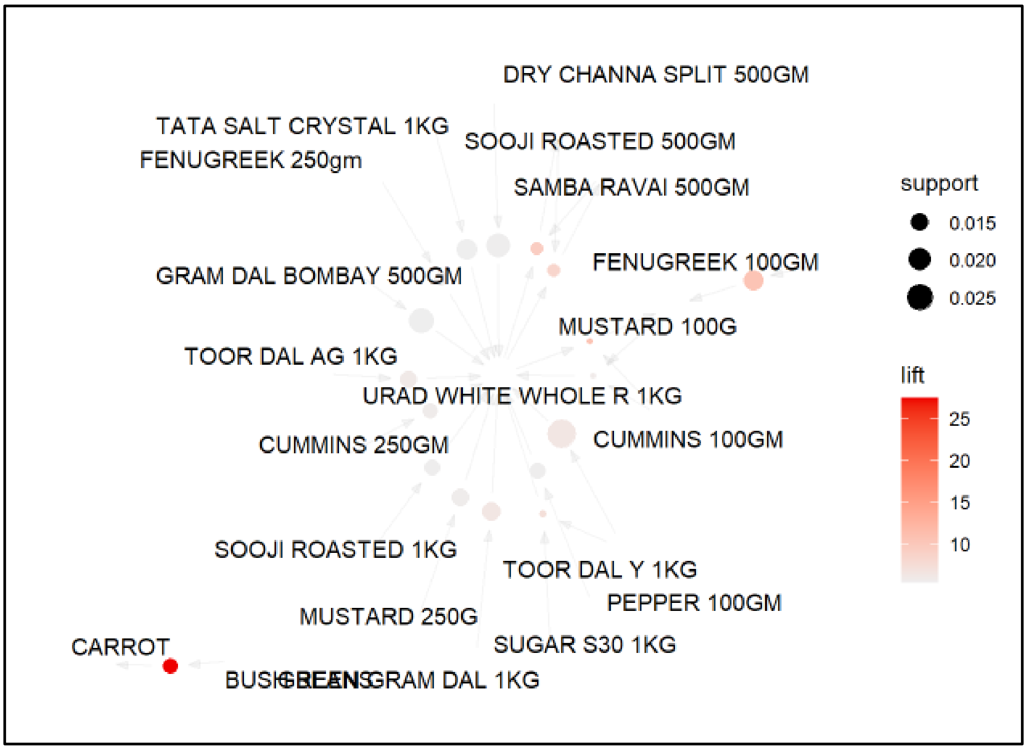
```
##      lhs                          rhs                          support confidence   coverage      lift count
## [1]  {BUSH BEANS}              => {CARROT}                  0.01316227  0.7684848 0.01712755 27.562456   634
## [2]  {CUMMINS 100GM,
##       URAD WHITE WHOLE R 1KG} => {MUSTARD 100G}             0.01013121  0.5197018 0.01949427 10.753006   488
## [3]  {FENUGREEK 100GM}         => {MUSTARD 100G}             0.01723136  0.5126621 0.03361153 10.607350   830
## [4]  {SAMBA RAVAI 500GM,
##       URAD WHITE WHOLE R 1KG} => {SOOJI ROASTED 500GM}      0.01185434  0.5080071 0.02333499  9.411418   571
## [5]  {SOOJI ROASTED 500GM,
##       URAD WHITE WHOLE R 1KG} => {SAMBA RAVAI 500GM}        0.01185434  0.5306691 0.02233848  8.463997   571
## [6]  {SUGAR S30 1KG,
##       TOOR DAL Y 1KG}         => {URAD WHITE WHOLE R 1KG} 0.01017273  0.6960227 0.01461551  7.278772   490
## [7]  {GREEN GRAM DAL 1KG}      => {URAD WHITE WHOLE R 1KG} 0.01550822  0.6088020 0.02547334  6.366646   747
## [8]  {TOOR DAL Y 1KG}          => {URAD WHITE WHOLE R 1KG} 0.02943863  0.6070205 0.04849693  6.348017  1418
## [9]  {TOOR DAL AG 1KG}         => {URAD WHITE WHOLE R 1KG} 0.01432486  0.5740433 0.02495433  6.003151   690
## [10] {CUMMINS 100GM,
##       MUSTARD 100G}           => {URAD WHITE WHOLE R 1KG} 0.01013121  0.5589920 0.01812407  5.845750   488
## [11] {CUMMINS 250GM}           => {URAD WHITE WHOLE R 1KG} 0.01307922  0.5440415 0.02404086  5.689403   630
## [12] {MUSTARD 250G}            => {URAD WHITE WHOLE R 1KG} 0.01507225  0.5377778 0.02802691  5.623899   726
## [13] {FENUGREEK 250gm}         => {URAD WHITE WHOLE R 1KG} 0.01222804  0.5349682 0.02285750  5.594518   589
## [14] {DRY CHANNA SPLIT 500GM} => {URAD WHITE WHOLE R 1KG} 0.02188175  0.5230769 0.04183275  5.470163  1054
## [15] {PEPPER 100GM}            => {URAD WHITE WHOLE R 1KG} 0.01372280  0.5208826 0.02634529  5.447215   661
## [16] {SOOJI ROASTED 1KG}       => {URAD WHITE WHOLE R 1KG} 0.01403421  0.5192012 0.02703039  5.429632   676
## [17] {GRAM DAL BOMBAY 500GM}  => {URAD WHITE WHOLE R 1KG} 0.02360488  0.5137822 0.04594336  5.372962  1137
## [18] {TATA SALT CRYSTAL 1KG}  => {URAD WHITE WHOLE R 1KG} 0.01860156  0.5128792 0.03626889  5.363519   896
```

Here we can observe that after checking for redundancy, the number of rules remain the same.

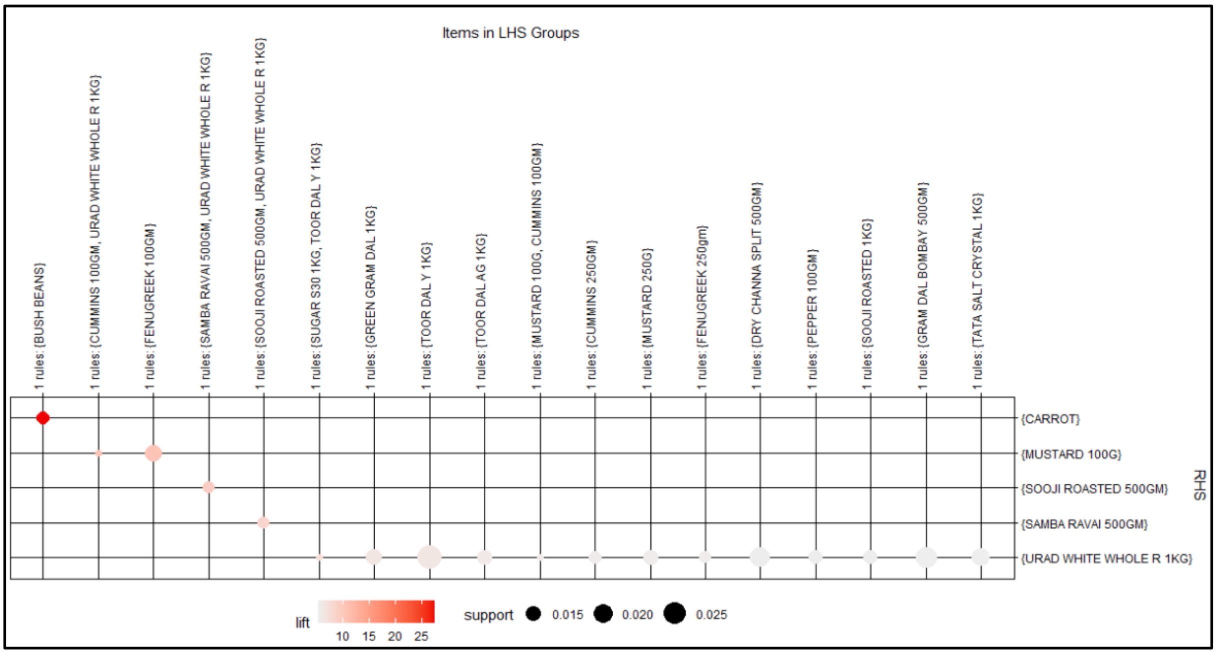
Visualization



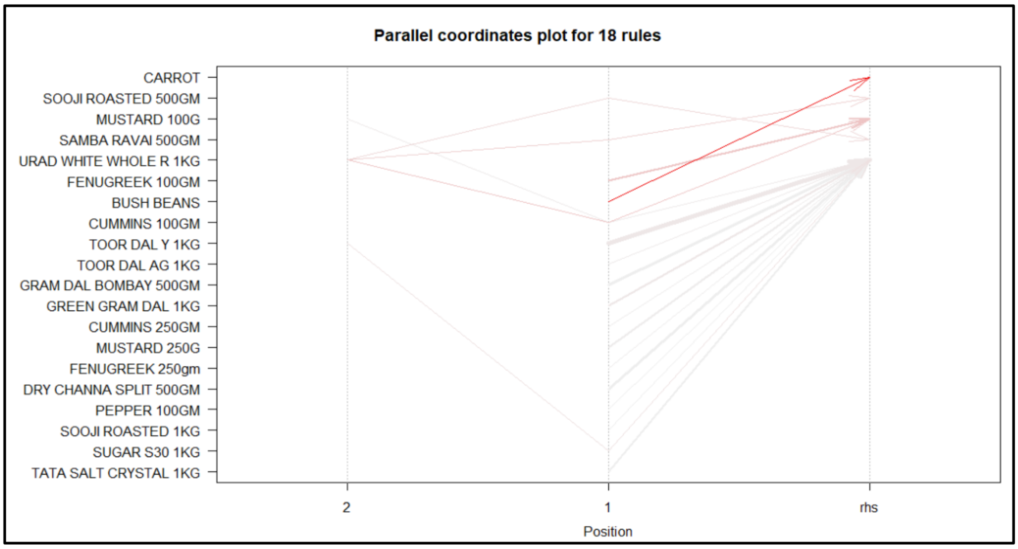
The scatter plot for all 18 rules generated in the above figure. We see that there is one rule which has got both high support and high lift ratio. This implies that, the rule generated means that the association between the two items is strong and frequent. This can provide valuable insights for retailers to make strategic decisions about product placement, promotions, and product development.



The above visualization represents the rules as a graph with items as labelled vertices, and rules represented as vertices connected to items using arrows. We see here also that urad white has a clear connection with rest of the items purchased in a network. Since this isn't that clear, we plot few more visualizations.



In the above figure we observe that, the items are represented in a matrix form. The items on the lhs and rhs are used to determine how much items is associated with the other items. The round circle on this plot gives us the lift ratio color and size of circle is by support so that it helps us understand which items have high lift and support both. Here carrot and bush beans have higher lift and support. One such item with highest support is urad dal which associates well with toor dal. In this case it's a useful visualization since there are 5 items on rhs.



The parallel coordinate plot allows for a visual comparison of the characteristics of multiple rules at once. By looking at the position and color of the lines, we can tell how many items are interconnected. For example, here we can tell that carrot and bush beans are frequently purchased together. The position refers to the location of an item or a rule on the vertical axis and it is used to represent the characteristics of the rules such as support, confidence, or lift.

4. RECOMMENDATIONS

- One of the key inference that can be made from market basket analysis is identifying which products are often purchased together. For example, the analysis reveals that bush beans and carrot are often purchased together. With this information, the supermarket can make sure that these products are always stocked and displayed together in the store, potentially increasing the likelihood that customers will purchase both items.

- From the analysis we see that urad white whole 1kg is the most popular item hence the supermarket can optimize inventory management by ensuring that they always have enough of the product in stock. They can also schedule staff and promotions to match peak demand, which can help to increase revenue.

- **Bundling popular items:** By identifying which products are often purchased together, the supermarket can bundle these items together as a package deal to increase revenue. For example, if toor dal and urad white whole are often purchased together, the supermarket can create a bundle deal that includes both items at a discounted price or buy one get one kind of deal.

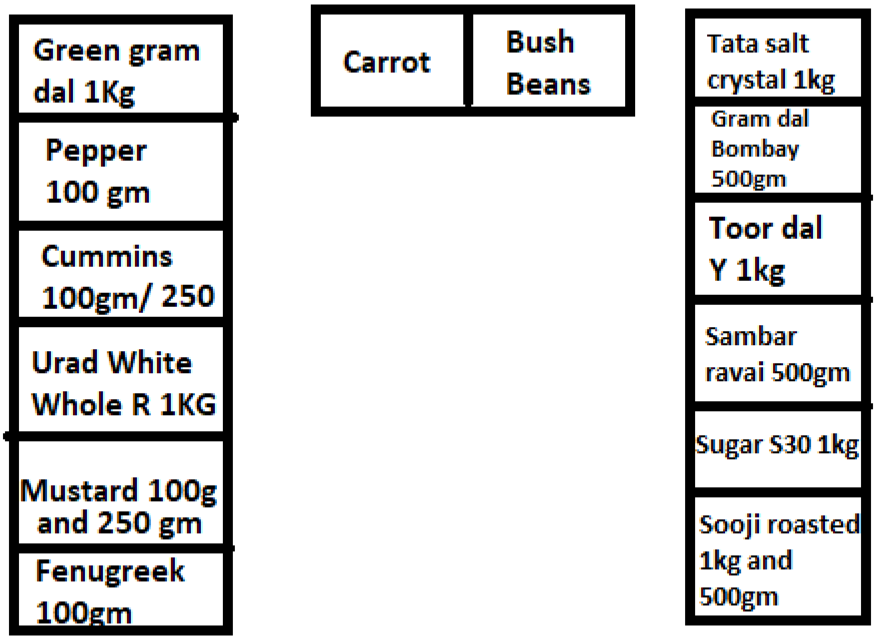
- **Optimizing product placement:** By identifying the most popular products and times of day/week for sales, the supermarket can optimize product placement to ensure that these products are easy to find and purchase. For example, here urad white whole is a popular item, the supermarket can place them near the front of the store, close to the mustard, sooji and samba ravai to make it more convenient for customers to purchase both items.

- **Managing inventory:** By understanding the most popular products and time of the day/week for sales, the supermarket can manage their inventory efficiently. They can ensure that they always have enough of the most popular products in stock, and can schedule their staff and inventory accordingly.

- **Upselling:** By understanding the associations between products, the supermarket can recommend products that complement each other to customers. For example, if a customer is buying urad white whole, the supermarket can recommend fenugreek or mustard to them, encouraging them to purchase both items.

Overall, market basket analysis can provide valuable insights for a supermarket, allowing the business to make data-driven decisions that can increase revenue and improve performance. By implementing these recommendations, the supermarket can improve customer satisfaction and increase their revenue.

5. PRODUCT PLACEMENT



The product placements from the association rule mining can be done by the help of the above diagram. The SKU's of different products can be placed in the same placement because it will be easier to customer to pick on the quantity which they desire to buy.

CONCLUSION

In conclusion, market basket analysis can provide valuable insights for supermarket that can help the business increase revenue and improve performance. By identifying popular products, times of day/week for sales, and customer demographics, the supermarket can make data-driven decisions such as bundling popular items, optimizing product placement, targeting promotions, managing inventory, and upselling. By implementing these recommendations, the supermarket can improve customer satisfaction and increase their revenue. Overall, the market basket analysis is an effective tool for understanding customer behavior and making data-driven decisions that can help a business to improve their performance and revenue.

REFERENCES:

Association Rule Learning - Javatpoint. (n.d.). www.javatpoint.com. https://www.javatpoint.com/association-rule-learning

Data mining – Lift in an association rule. (n.d.). © Copyright IBM Corporation 2014. https://www.ibm.com/docs/en/db2/10.5?topic=SSEPGG_10.5.0/com.ibm.im.model.doc/c_lift_in_an_association_rule.htm

GeeksforGeeks. (2023, January 11). *Association Rule*. https://www.geeksforgeeks.org/association-rule/

Jena, M. (2023, January 12). *Association Rule Mining Simplified 101*. Learn | Hevo. https://hevodata.com/learn/association-rule-mining/

Lutkevich, B. (2020, September 9). *association rules*. Business Analytics. https://www.techtarget.com/searchbusinessanalytics/definition/association-rules-in-data-mining