# MKT 552 Market Basket Analysis Assignment

Aridania Gerardo

2022-07-30

## Installing Packages

```
#packages for analysis
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
library(arules)
```

```
## Loading required package: Matrix

##
## Attaching package: 'arules'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following objects are masked from 'package:base':
##
##     abbreviate, write
```

```
library(arulesViz)
library(scales)
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:readr':
##
##     col_factor
```

## Uploading data

```
Order<-read.csv("https://raw.githubusercontent.com/aridaniaG/InstacartMKTBasket/main/Basket_Raw.csv", n
Products<-read.csv("https://raw.githubusercontent.com/aridaniaG/InstacartMKTBasket/main/products.csv")
DepartmentName<-read.csv("https://raw.githubusercontent.com/aridaniaG/InstacartMKTBasket/main/department
```

## Preprocessing data

```
#extracting columns for df
OrderId<-Order[,1:2]
ProductDepartmentName<-Products[,c(1:2,4)]


#transforming df to transactions for analysis
ProductDepartmentName<-merge(ProductDepartmentName, DepartmentName, by="department_id")
TransData<-merge(OrderId, ProductDepartmentName, by="product_id")
TransData<-arrange(TransData, OrderId)


#data for analysis
ProductAnaly<-TransData[,c(2,4)]
DepartmentAnaly<-TransData[,c(2,5)]


#data for both department and specific products
TransPData <- split(ProductAnaly$product_name, ProductAnaly$order_id)
TransDData <- split(DepartmentAnaly$department, DepartmentAnaly$order_id)


ProTran = as(TransPData,"transactions")
DepTran = as(TransDData,"transactions")
```

```
## Warning in asMethod(object): removing duplicated items in transactions
```

## EDA

### Specific Products

The five most frequent products bought on instacart are Bananas, Bag of Organic Bananas, Organic Straw-berries, Organic Baby Spinach, and Large Lemon's.

```
#frequent items for specific prodcuts
frequentItems <- eclat(ProTran, parameter = list(supp = 0.07, maxlen = 15))
```

```
## Eclat
##
## parameter specification:
##  tidLists support minlen maxlen          target  ext
##     FALSE    0.07      1     15 frequent itemsets TRUE
##
## algorithmic control:
```
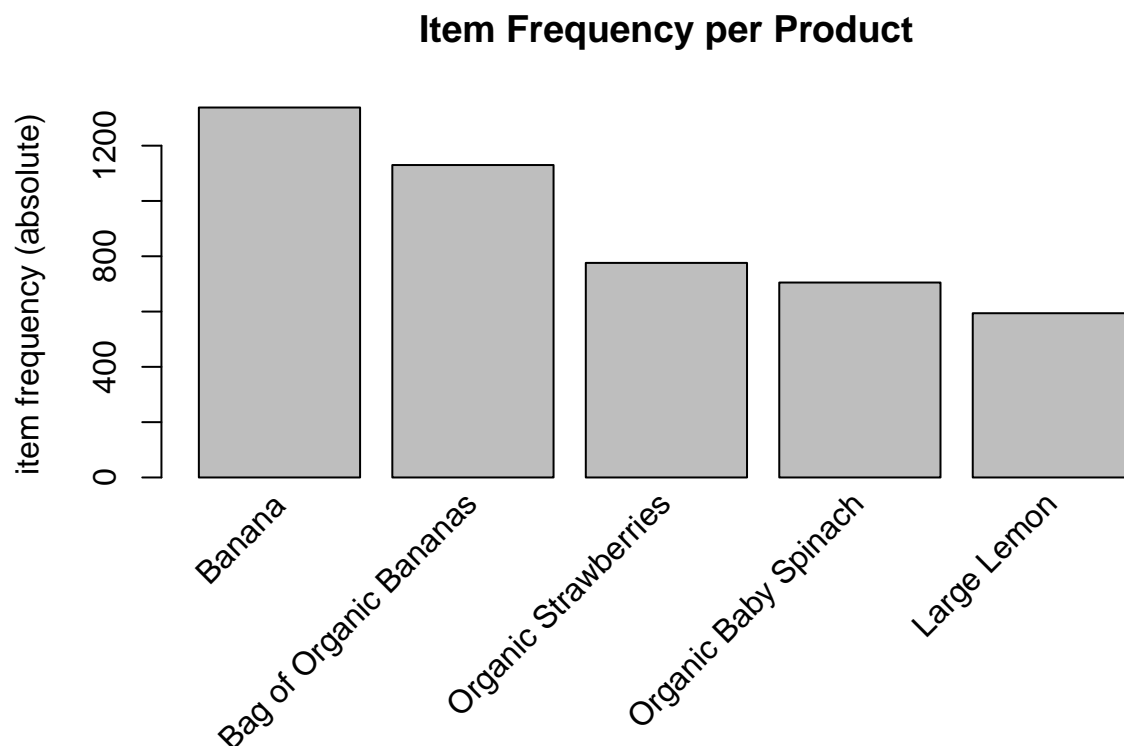
```
##   sparse sort verbose
##      7   -2    TRUE
##
## Absolute minimum support count: 663
##
## create itemset ...
## set transactions ...[16738 item(s), 9477 transaction(s)] done [0.11s].
## sorting and recoding items ... [4 item(s)] done [0.00s].
## creating sparse bit matrix ... [4 row(s), 9477 column(s)] done [0.00s].
## writing  ... [4 set(s)] done [0.00s].
## Creating S4 object  ... done [0.00s].
```

```
inspect(frequentItems)
```

```
##       items                   support    count
## [1] {Banana}                0.14118392 1338
## [2] {Bag of Organic Bananas} 0.11923605 1130
## [3] {Organic Strawberries}   0.08188245  776
## [4] {Organic Baby Spinach}   0.07439063  705
```

```r
#plot frequency
itemFrequencyPlot(ProTran, topN=5, type="absolute", main="Item Frequency per Product")
```

## Item Frequency per Product



Organic Hass Avocado, Organic Raspberries, and Organic Unsweetened Almond Milk are most likely to be purchased with a Bag of Organic Bananas with a confidence level of 0.83.

```r
#rules for Products
rules <- apriori(ProTran, parameter = list(supp = 0.001, conf = 0.5, maxlen=15))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##        0.5    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##      15  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[16738 item(s), 9477 transaction(s)] done [0.11s].
## sorting and recoding items ... [1953 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.05s].
## writing ... [141 rule(s)] done [0.01s].
## creating S4 object  ... done [0.00s].
```

```r
#confidence
rules_conf <- sort(rules, by="confidence", decreasing=TRUE)
inspect(head(rules_conf))
```

```
##      lhs                              rhs                       support confidence   coverage
## [1] {Organic Hass Avocado,
##      Organic Raspberries,
##      Organic Unsweetened Almond Milk} => {Bag of Organic Bananas} 0.001055186  0.8333333 0.001266223
## [2] {Organic Cucumber,
##      Organic Hass Avocado,
##      Organic Strawberries}          => {Bag of Organic Bananas} 0.001582779  0.8333333 0.001899335
## [3] {Organic Strawberries,
##      Organic Unsalted Butter}        => {Bag of Organic Bananas} 0.001160705  0.7857143 0.001477261
## [4] {Bag of Organic Bananas,
##      Organic Raspberries,
##      Organic Unsweetened Almond Milk} => {Organic Hass Avocado}  0.001055186  0.7692308 0.001371742
## [5] {Organic Hass Avocado,
##      Organic Raspberries,
##      Organic Strawberries}          => {Bag of Organic Bananas} 0.002004854  0.7600000 0.002637966
## [6] {Organic Ginger Root,
##      Organic Raspberries}            => {Bag of Organic Bananas} 0.001266223  0.7500000 0.001688298
```

### Departments

The top five most popular categories that are bought from are produce, dairy eggs, beverages, snacks, and frozen items.

```r
#frequent items in departments
frequentItems <- eclat(DepTran, parameter = list(supp = 0.07, maxlen = 15))
```

```
## Eclat
```

```
## 
## parameter specification:
##  tidLists support minlen maxlen          target  ext
##     FALSE    0.07      1     15 frequent itemsets TRUE
## 
## algorithmic control:
##  sparse sort verbose
##       7   -2    TRUE
## 
## Absolute minimum support count: 663
## 
## create itemset ...
## set transactions ...[21 item(s), 9477 transaction(s)] done [0.00s].
## sorting and recoding items ... [14 item(s)] done [0.00s].
## creating bit matrix ... [14 row(s), 9477 column(s)] done [0.00s].
## writing  ... [217 set(s)] done [0.00s].
## Creating S4 object  ... done [0.00s].
```

inspect(frequentItems)

```
##      items                                     support     count
## [1]  {personal care, produce}                  0.07903345   749
## [2]  {dairy eggs, personal care}               0.07586789   719
## [3]  {dairy eggs, household, produce}          0.08631423   818
## [4]  {household, produce}                      0.10794555  1023
## [5]  {dairy eggs, household}                   0.10942281  1037
## [6]  {beverages, household}                    0.09232880   875
## [7]  {household, snacks}                       0.08114382   769
## [8]  {frozen, household}                       0.07206922   683
## [9]  {breakfast, frozen, produce}              0.07428511   704
## [10] {breakfast, dairy eggs, frozen}           0.07628997   723
## [11] {breakfast, dairy eggs, produce, snacks}  0.07671204   727
## [12] {breakfast, produce, snacks}              0.08684183   823
## [13] {breakfast, dairy eggs, snacks}           0.08853013   839
## [14] {beverages, breakfast, produce}           0.07681756   728
## [15] {beverages, breakfast, dairy eggs}        0.08019415   760
## [16] {breakfast, dairy eggs, produce}          0.11659808  1105
## [17] {breakfast, produce}                      0.13411417  1271
## [18] {breakfast, dairy eggs}                   0.13706869  1299
## [19] {beverages, breakfast}                    0.09338398   885
## [20] {breakfast, snacks}                       0.10330273   979
## [21] {breakfast, frozen}                       0.08673631   822
## [22] {breakfast, pantry}                       0.07534030   714
## [23] {dairy eggs, meat seafood, pantry, produce} 0.07101403 673
## [24] {meat seafood, pantry, produce}           0.08124934   770
## [25] {dairy eggs, meat seafood, pantry}        0.07692308   729
## [26] {dairy eggs, frozen, meat seafood, produce} 0.07586789 719
## [27] {frozen, meat seafood, produce}           0.08610320   816
## [28] {dairy eggs, frozen, meat seafood}        0.08167142   774
## [29] {dairy eggs, meat seafood, produce, snacks} 0.07280785 690
## [30] {meat seafood, produce, snacks}           0.08325419   789
## [31] {dairy eggs, meat seafood, snacks}        0.07829482   742
## [32] {beverages, dairy eggs, meat seafood, produce} 0.07238578 686
## [33] {beverages, meat seafood, produce}        0.08430938   799
```
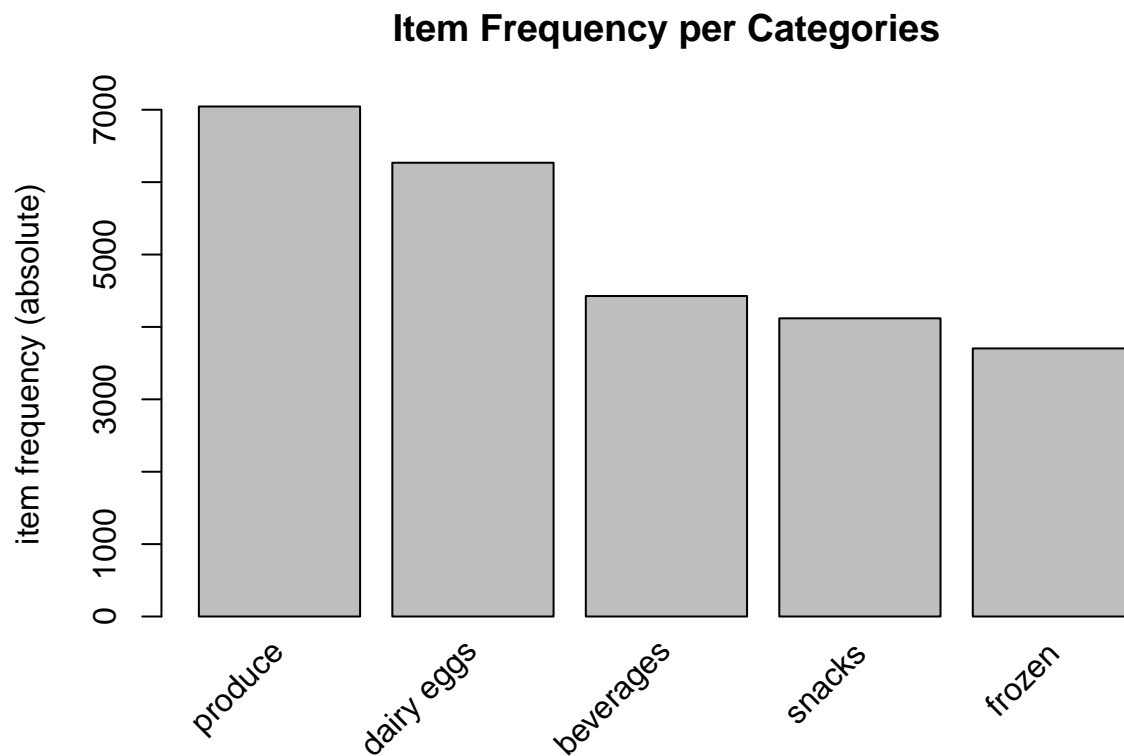
```
## [34]   {beverages, dairy eggs, meat seafood}                 0.07977208  756
## [35]   {dairy eggs, meat seafood, produce}                   0.13538039 1283
## [36]   {meat seafood, produce}                               0.16313179 1546
## [37]   {dairy eggs, meat seafood}                            0.14814815 1404
## [38]   {beverages, meat seafood}                             0.09602195  910
## [39]   {meat seafood, snacks}                                0.09222328  874
## [40]   {frozen, meat seafood}                                0.09538884  904
## [41]   {meat seafood, pantry}                                0.08937427  847
## [42]   {bakery, meat seafood}                                0.07111955  674
## [43]   {canned goods, dry goods pasta, produce}              0.07512926  712
## [44]   {canned goods, dairy eggs, dry goods pasta}           0.07291337  691
## [45]   {bakery, dry goods pasta, produce}                    0.07122507  675
## [46]   {dairy eggs, dry goods pasta, pantry, produce}        0.08335971  790
## [47]   {dry goods pasta, pantry, produce}                    0.09359502  887
## [48]   {dairy eggs, dry goods pasta, pantry}                 0.09116809  864
## [49]   {dairy eggs, dry goods pasta, frozen, produce}        0.08905772  844
## [50]   {dry goods pasta, frozen, produce}                    0.10066477  954
## [51]   {dairy eggs, dry goods pasta, frozen}                 0.09855439  934
## [52]   {dry goods pasta, frozen, snacks}                     0.07016988  665
## [53]   {dairy eggs, dry goods pasta, produce, snacks}        0.08547009  810
## [54]   {dry goods pasta, produce, snacks}                    0.09612747  911
## [55]   {dairy eggs, dry goods pasta, snacks}                 0.09412261  892
## [56]   {beverages, dairy eggs, dry goods pasta, produce} 0.08051071  763
## [57]   {beverages, dry goods pasta, produce}                 0.09095705  862
## [58]   {beverages, dairy eggs, dry goods pasta}              0.08947979  848
## [59]   {dairy eggs, dry goods pasta, produce}                0.14814815 1404
## [60]   {dry goods pasta, produce}                            0.17378917 1647
## [61]   {dairy eggs, dry goods pasta}                         0.16503113 1564
## [62]   {beverages, dry goods pasta}                          0.10361929  982
## [63]   {dry goods pasta, snacks}                             0.10857866 1029
## [64]   {dry goods pasta, frozen}                             0.11385460 1079
## [65]   {dry goods pasta, pantry}                             0.10435792  989
## [66]   {bakery, dry goods pasta}                             0.07850586  744
## [67]   {deli, dry goods pasta}                               0.07481270  709
## [68]   {canned goods, dry goods pasta}                       0.08388731  795
## [69]   {bakery, canned goods, produce}                       0.07723963  732
## [70]   {bakery, canned goods, dairy eggs}                    0.07639548  724
## [71]   {canned goods, dairy eggs, pantry, produce}           0.09623299  912
## [72]   {canned goods, pantry, produce}                       0.11100559 1052
## [73]   {canned goods, dairy eggs, pantry}                    0.10351377  981
## [74]   {canned goods, dairy eggs, frozen, produce}           0.09106257  863
## [75]   {canned goods, frozen, produce}                       0.10393584  985
## [76]   {canned goods, dairy eggs, frozen}                    0.10150892  962
## [77]   {canned goods, dairy eggs, produce, snacks}           0.09064050  859
## [78]   {canned goods, produce, snacks}                       0.10309170  977
## [79]   {canned goods, dairy eggs, snacks}                    0.09844888  933
## [80]   {beverages, canned goods, dairy eggs, produce}        0.08620872  817
## [81]   {beverages, canned goods, produce}                    0.10098132  957
## [82]   {beverages, canned goods, dairy eggs}                 0.09697162  919
## [83]   {canned goods, dairy eggs, produce}                   0.16292076 1544
## [84]   {canned goods, produce}                               0.19837501 1880
## [85]   {canned goods, dairy eggs}                            0.18212515 1726
## [86]   {beverages, canned goods}                             0.11628152 1102
## [87]   {canned goods, snacks}                                0.11469874 1087
```

```
## [88]  {canned goods, frozen}                        0.11796982 1118
## [89]  {canned goods, pantry}                         0.12134642 1150
## [90]  {bakery, canned goods}                         0.08504801  806
## [91]  {canned goods, deli}                           0.07555133  716
## [92]  {bakery, dairy eggs, deli, produce}            0.08029967  761
## [93]  {bakery, deli, produce}                        0.08863564  840
## [94]  {bakery, dairy eggs, deli}                     0.08789701  833
## [95]  {dairy eggs, deli, pantry, produce}            0.08779150  832
## [96]  {deli, pantry, produce}                        0.10098132  957
## [97]  {dairy eggs, deli, pantry}                     0.09454469  896
## [98]  {deli, pantry, snacks}                         0.07006437  664
## [99]  {deli, frozen, produce, snacks}                0.07291337  691
## [100] {dairy eggs, deli, frozen, snacks}             0.07164715  679
## [101] {dairy eggs, deli, frozen, produce}            0.10087580  956
## [102] {deli, frozen, produce}                        0.11469874 1087
## [103] {dairy eggs, deli, frozen}                     0.11016144 1044
## [104] {beverages, deli, frozen}                      0.07249129  687
## [105] {deli, frozen, snacks}                         0.08156590  773
## [106] {beverages, deli, produce, snacks}             0.07449615  706
## [107] {beverages, dairy eggs, deli, snacks}          0.07354648  697
## [108] {dairy eggs, deli, produce, snacks}            0.10847315 1028
## [109] {deli, produce, snacks}                        0.12514509 1186
## [110] {dairy eggs, deli, snacks}                     0.12092434 1146
## [111] {beverages, deli, snacks}                      0.08547009  810
## [112] {beverages, dairy eggs, deli, produce}         0.09559987  906
## [113] {beverages, deli, produce}                     0.11311596 1072
## [114] {beverages, dairy eggs, deli}                  0.10826211 1026
## [115] {dairy eggs, deli, produce}                    0.17779888 1685
## [116] {deli, produce}                                0.21388625 2027
## [117] {dairy eggs, deli}                             0.19932468 1889
## [118] {beverages, deli}                              0.13020998 1234
## [119] {deli, snacks}                                 0.14382188 1363
## [120] {deli, frozen}                                 0.12831065 1216
## [121] {deli, pantry}                                 0.11216630 1063
## [122] {bakery, deli}                                 0.09918751  940
## [123] {bakery, dairy eggs, pantry, produce}          0.09475572  898
## [124] {bakery, pantry, produce}                      0.10720692 1016
## [125] {bakery, dairy eggs, pantry}                   0.10488551  994
## [126] {bakery, pantry, snacks}                       0.07597341  720
## [127] {bakery, frozen, produce, snacks}              0.07660652  726
## [128] {bakery, dairy eggs, frozen, snacks}           0.07650100  725
## [129] {bakery, beverages, frozen, produce}           0.07038092  667
## [130] {bakery, dairy eggs, frozen, produce}          0.10762900 1020
## [131] {bakery, frozen, produce}                      0.12081882 1145
## [132] {bakery, dairy eggs, frozen}                   0.11976364 1135
## [133] {bakery, beverages, frozen}                    0.07924449  751
## [134] {bakery, frozen, snacks}                       0.08726390  827
## [135] {bakery, beverages, produce, snacks}           0.07787274  738
## [136] {bakery, beverages, dairy eggs, snacks}        0.07755619  735
## [137] {bakery, dairy eggs, produce, snacks}          0.11533186 1093
## [138] {bakery, produce, snacks}                      0.13042102 1236
## [139] {bakery, dairy eggs, snacks}                   0.12968239 1229
## [140] {bakery, beverages, snacks}                    0.08958531  849
## [141] {bakery, beverages, dairy eggs, produce}       0.10372481  983
```

```
## [142] {bakery, beverages, produce}                        0.12081882 1145
## [143] {bakery, beverages, dairy eggs}                      0.11860293 1124
## [144] {bakery, dairy eggs, produce}                        0.19320460 1831
## [145] {bakery, produce}                                    0.22633745 2145
## [146] {bakery, dairy eggs}                                 0.22000633 2085
## [147] {bakery, beverages}                                  0.14192255 1345
## [148] {bakery, snacks}                                     0.15120819 1433
## [149] {bakery, frozen}                                     0.13896803 1317
## [150] {bakery, pantry}                                     0.12102986 1147
## [151] {dairy eggs, frozen, pantry, produce, snacks}        0.07966656  755
## [152] {frozen, pantry, produce, snacks}                    0.08905772  844
## [153] {dairy eggs, frozen, pantry, snacks}                 0.08736942  828
## [154] {beverages, dairy eggs, frozen, pantry, produce}     0.07280785  690
## [155] {beverages, frozen, pantry, produce}                 0.08124934  770
## [156] {beverages, dairy eggs, frozen, pantry}              0.08051071  763
## [157] {dairy eggs, frozen, pantry, produce}                0.12915480 1224
## [158] {frozen, pantry, produce}                            0.14878126 1410
## [159] {dairy eggs, frozen, pantry}                         0.14413844 1366
## [160] {beverages, frozen, pantry}                          0.09180120  870
## [161] {frozen, pantry, snacks}                             0.10098132  957
## [162] {beverages, dairy eggs, pantry, produce, snacks}     0.07935001  752
## [163] {beverages, pantry, produce, snacks}                 0.09032394  856
## [164] {beverages, dairy eggs, pantry, snacks}              0.08779150  832
## [165] {dairy eggs, pantry, produce, snacks}                0.13421969 1272
## [166] {pantry, produce, snacks}                            0.15711723 1489
## [167] {dairy eggs, pantry, snacks}                         0.14930885 1415
## [168] {beverages, pantry, snacks}                          0.10330273  979
## [169] {beverages, dairy eggs, pantry, produce}             0.12430094 1178
## [170] {beverages, pantry, produce}                         0.14656537 1389
## [171] {beverages, dairy eggs, pantry}                      0.14202807 1346
## [172] {dairy eggs, pantry, produce}                        0.24068798 2281
## [173] {pantry, produce}                                    0.29619078 2807
## [174] {dairy eggs, pantry}                                 0.27529809 2609
## [175] {beverages, pantry}                                  0.17452780 1654
## [176] {pantry, snacks}                                     0.18318033 1736
## [177] {frozen, pantry}                                     0.17241743 1634
## [178] {beverages, dairy eggs, frozen, produce, snacks}     0.09148465  867
## [179] {beverages, frozen, produce, snacks}                 0.10541311  999
## [180] {beverages, dairy eggs, frozen, snacks}              0.10319721  978
## [181] {dairy eggs, frozen, produce, snacks}                0.14867574 1409
## [182] {frozen, produce, snacks}                            0.17357814 1645
## [183] {dairy eggs, frozen, snacks}                         0.16851324 1597
## [184] {beverages, frozen, snacks}                          0.12356231 1171
## [185] {beverages, dairy eggs, frozen, produce}             0.14308325 1356
## [186] {beverages, frozen, produce}                         0.16914635 1603
## [187] {beverages, dairy eggs, frozen}                      0.16313179 1546
## [188] {dairy eggs, frozen, produce}                        0.26210826 2484
## [189] {frozen, produce}                                    0.31993247 3032
## [190] {dairy eggs, frozen}                                 0.30062256 2849
## [191] {beverages, frozen}                                  0.20196265 1914
## [192] {frozen, snacks}                                     0.20639443 1956
## [193] {beverages, dairy eggs, produce, snacks}             0.15352960 1455
## [194] {beverages, produce, snacks}                         0.18740108 1776
## [195] {beverages, dairy eggs, snacks}                      0.18244170 1729
```

```
## [196] {dairy eggs, produce, snacks}                    0.26981112 2557
## [197] {produce, snacks}                                 0.33987549 3221
## [198] {dairy eggs, snacks}                              0.31877176 3021
## [199] {beverages, snacks}                               0.23794450 2255
## [200] {beverages, dairy eggs, produce}                  0.26632901 2524
## [201] {beverages, produce}                              0.34462383 3266
## [202] {beverages, dairy eggs}                           0.32362562 3067
## [203] {dairy eggs, produce}                             0.54257676 5142
## [204] {produce}                                         0.74337871 7045
## [205] {dairy eggs}                                      0.66139074 6268
## [206] {beverages}                                       0.46713095 4427
## [207] {snacks}                                          0.43463121 4119
## [208] {frozen}                                          0.39073546 3703
## [209] {pantry}                                          0.35907988 3403
## [210] {bakery}                                          0.26907249 2550
## [211] {deli}                                            0.24765221 2347
## [212] {canned goods}                                    0.22760367 2157
## [213] {dry goods pasta}                                 0.19995779 1895
## [214] {meat seafood}                                    0.18497415 1753
## [215] {breakfast}                                       0.16344835 1549
## [216] {household}                                       0.16133798 1529
## [217] {personal care}                                   0.11206078 1062
```

```
#plot frequency
itemFrequencyPlot(DepTran, topN=5, type="absolute", main="Item Frequency per Categories")
```



Purchase of dairy eggs is associated with bulk, personal care, and snacks. Purchase of produce is associated with bulk, dairy eggs, personal care.

```
#rules for departments
rules <- apriori(DepTran, parameter = list(supp = 0.001, conf = 0.5, maxlen=15))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##      15  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[21 item(s), 9477 transaction(s)] done [0.01s].
## sorting and recoding items ... [21 item(s)] done [0.00s].
## creating transaction tree ... done [0.01s].
## checking subsets of size 1 2 3 4 5 6 7 8 9 10 11 done [0.07s].
## writing ... [105593 rule(s)] done [0.06s].
## creating S4 object  ... done [0.06s].
```

```
#confidence
rules_conf <- sort(rules, by="confidence", decreasing=TRUE)
inspect(head(rules_conf))
```

```
##     lhs                              rhs          support     confidence
## [1] {bulk, personal care, snacks}     => {dairy eggs} 0.001055186 1
## [2] {bulk, personal care, snacks}     => {produce}    0.001055186 1
## [3] {bulk, dairy eggs, personal care} => {produce}    0.001160705 1
## [4] {bulk, personal care, produce}    => {dairy eggs} 0.001160705 1
## [5] {breakfast, bulk, deli}           => {produce}    0.001055186 1
## [6] {breakfast, bulk, frozen}         => {produce}    0.001055186 1
##     coverage    lift     count
## [1] 0.001055186 1.511966 10
## [2] 0.001055186 1.345209 10
## [3] 0.001160705 1.345209 11
## [4] 0.001160705 1.511966 11
## [5] 0.001055186 1.345209 10
## [6] 0.001055186 1.345209 10
```

## Analysis Summary

Items that are likely to be bought together are from the produce department such as dairy eggs and produce with snacks and personal care.

If consumers buy two products together, the leader of the basket is most likely to be an item from the produce department sand the add-on is most likely to be itmes that would accompany such as beverages and snacks.