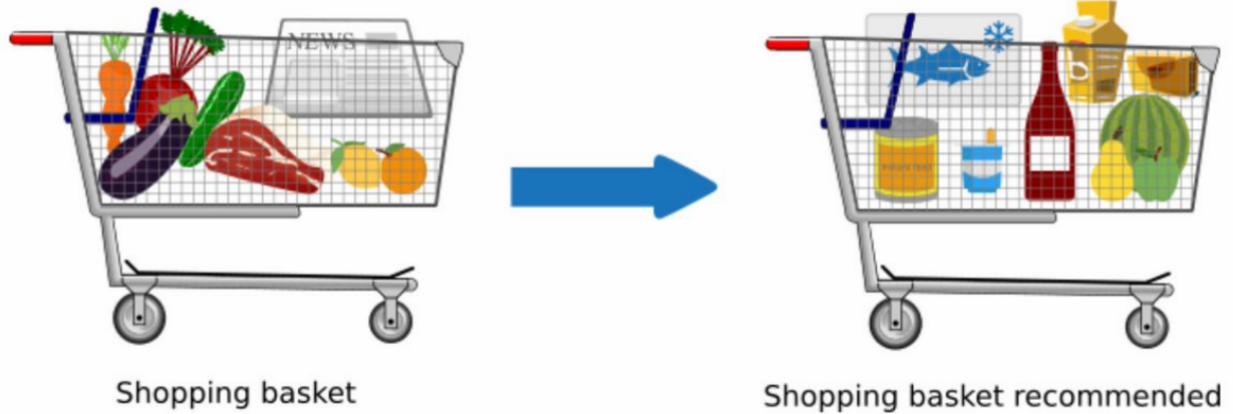


# Instacart Market Basket Analysis

## Association Rule Mining

---



### Team Members:

Omkar Nandkumar Dixit(ond170030)

Dhwani Raval (dsr170230)

Karan Yashwant Kanani (kyk170030)

### Introduction

Whether you shop from meticulously planned grocery lists or let whimsy guide your grazing, our unique food rituals define who we are. Instacart, a grocery ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review your order and do the in-store shopping and delivery for you.

Instacart's data science team plays a big part in providing this delightful shopping experience. Currently they use transactional data to develop models that predict which products a user will buy again, try for the first time, or add to their cart next during a session. Recently, Instacart open sourced this data - see their blog post on [3 Million Instacart Orders, Open Sourced](#).

---

---

In this competition, Instacart is challenging the Kaggle community to use this anonymized data on customer orders over time to predict which previously purchased products will be in a user's next order.

## File descriptions

- Aisles.csv - Description of Aisles
- Departments.csv - Description of Departments
- order\_products\_\_\*.csv - These files specify which products were purchased in each order. order\_products\_\_prior.csv contains previous order contents for all customers.
- orders.csv - This file tells to which set (prior, train, test) an order belongs.
- Products.csv - Description of Products
- Sample\_submission.csv - Sample submission file.

```
// To download use the following API  
kaggle competitions download -c instacart-market-basket-analysis
```

## Implementation Details

Libraries: arules, arulesViz, Readr, tidyverse

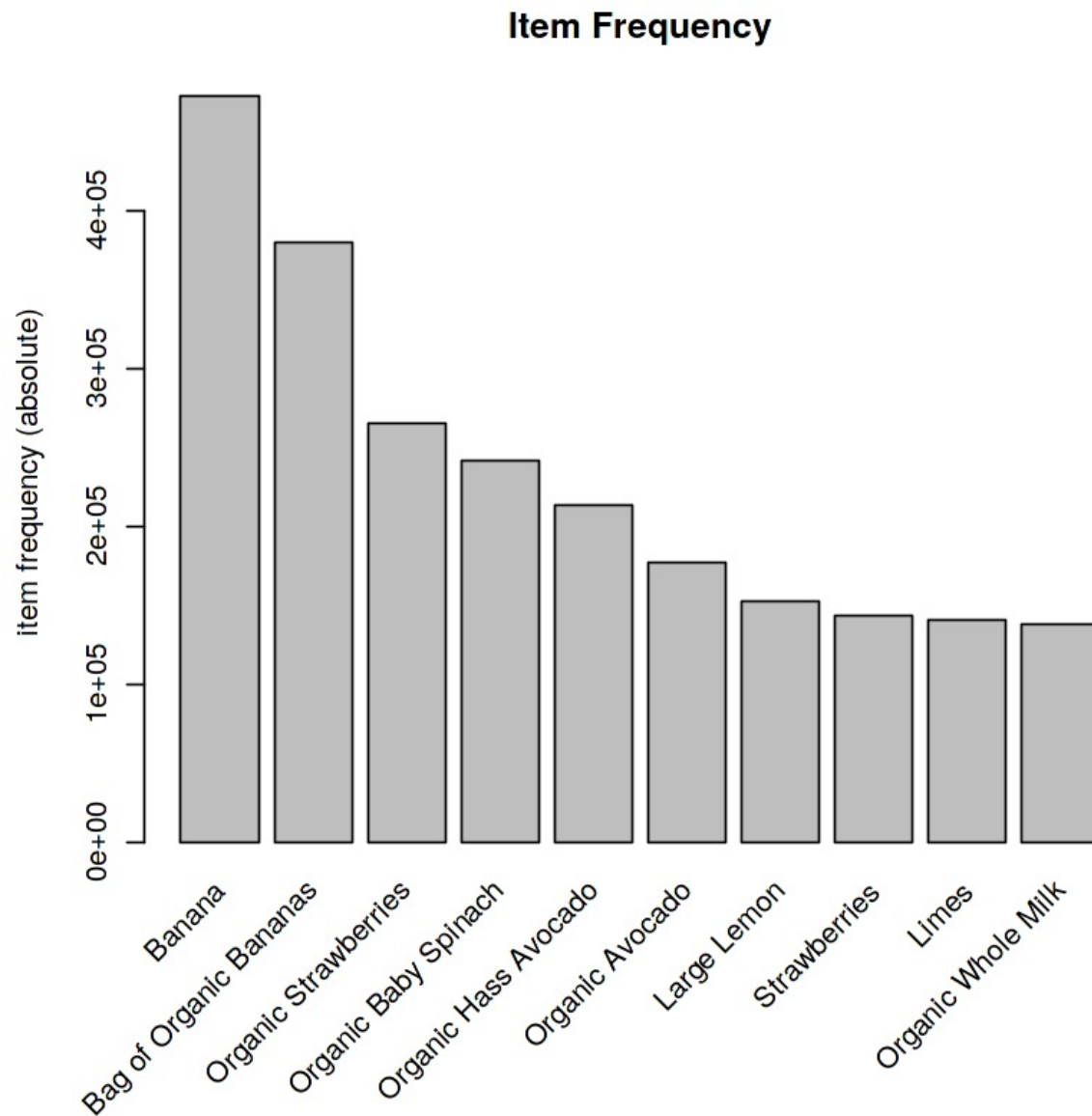
Scripting Language: R

For Calculating the frequent itemsets we have used the prior orders.

---

## Results

### Q1. Frequent itemsets for products in orders dataset



---

Eclat

parameter specification:

|          |         |        |        |          |          |
|----------|---------|--------|--------|----------|----------|
| tidLists | support | minlen | maxlen | target   | ext      |
| FALSE    | 0.01    | 1      | 80     | frequent | itemsets |

algorithmic control:

|             |         |
|-------------|---------|
| sparse sort | verbose |
| 7           | -2      |
| TRUE        |         |

Absolute minimum support count: 32148

create itemset ...

set transactions ...[49677 item(s), 3214874 transaction(s)] done [9.04s].

sorting and recoding items ... [102 item(s)] done [0.36s].

creating sparse bit matrix ... [102 row(s), 3214874 column(s)] done [0.90s].

writing ... [116 set(s)] done [2.80s].

Creating S4 object ... done [0.00s].

items

support

count

|            |   |
|------------|---|
| [1]        | {Banana,<br>Organic Fuji Apple}                   |
| 0.01055811 | 33943   |
| [2]        | {Banana,<br>Strawberries}                         |
| 0.01282539 | 41232   |
| [3]        | {Bag of Organic Bananas,<br>Organic Raspberries}  |
| 0.01259863 | 40503   |
| [4]        | {Organic Raspberries,<br>Organic Strawberries}    |
| 0.01053323 | 33863   |
| [5]        | {Banana,<br>Large Lemon}                          |
| 0.01271589 | 40880   |
| [6]        | {Banana,<br>Organic Avocado}                      |
| 0.01660874 | 53395   |
| [7]        | {Bag of Organic Bananas,<br>Organic Hass Avocado} |
| 0.01939143 | 62341   |

---

```
[8]   {Organic Hass Avocado,  
      Organic Strawberries}  
0.01268914  40794  
[9]   {Organic Baby Spinach,  
      Organic Hass Avocado}  
0.01085610  34901  
[10]  {Banana,  
      Organic Baby Spinach}  
0.01598663  51395  
[11]  {Bag of Organic Bananas,  
      Organic Baby Spinach}  
0.01566842  50372  
[12]  {Organic Baby Spinach,  
      Organic Strawberries}  
0.01186174  38134  
[13]  {Banana,  
      Organic Strawberries}  
0.01746756  56156  
[14]  {Bag of Organic Bananas,  
      Organic Strawberries}  
0.01916965  61628  
[15]  {Banana}  
0.14699332  472565  
[16]  {Bag of Organic Bananas}  
0.11802951  379450  
[17]  {Organic Strawberries}  
0.08233075  264683  
[18]  {Organic Baby Spinach}  
0.07525054  241921  
[19]  {Organic Hass Avocado}  
0.06643620  213584  
[20]  {Organic Avocado}  
0.05499905  176815  
[21]  {Large Lemon}  
0.04748460  152657  
[22]  {Limes}  
0.04374262  140627  
[23]  {Organic Raspberries}  
0.04263215  137057  
[24]  {Strawberries}  
0.04446551  142951  
[25]  {Organic Whole Milk}
```

---

---

0.04289593 137905  
[26] {Organic Yellow Onion}  
0.03528163 113426  
[27] {Organic Garlic}  
0.03414691 109778  
[28] {Organic Zucchini}  
0.03260563 104823  
[29] {Cucumber Kirby}  
0.03027024 97315  
[30] {Organic Blueberries}  
0.03112408 100060  
[31] {Organic Grape Tomatoes}  
0.02620787 84255  
[32] {Organic Lemon}  
0.02729376 87746  
[33] {Organic Cucumber}  
0.02500627 80392  
[34] {Apple Honeycrisp Organic}  
0.02644583 85020  
[35] {Organic Fuji Apple}  
0.02788041 89632  
[36] {Seedless Red Grapes}  
0.02572076 82689  
[37] {Honeycrisp Apple}  
0.02481248 79769  
[38] {Organic Baby Carrots}  
0.02391882 76896  
[39] {Fresh Cauliflower}  
0.02226650 71584  
[40] {Organic Large Extra Fancy Fuji Apple}  
0.02338039 75165  
[41] {Yellow Onions}  
0.02275890 73167  
[42] {Organic Small Bunch Celery}  
0.02117999 68091  
[43] {Carrots}  
0.02262484 72736  
[44] {Organic Cilantro}  
0.02162573 69524  
[45] {Original Hummus}  
0.02218252 71314  
[46] {Organic Baby Arugula}

---

0.02265376 72829  
[47] {Organic Gala Apples}  
0.02265905 72846  
[48] {Organic Red Onion}  
0.02083628 66986  
[49] {Sparkling Water Grapefruit}  
0.02360466 75886  
[50] {Asparagus}  
0.02092866 67283  
[51] {Michigan Organic Kale}  
0.02098029 67449  
[52] {Organic Half & Half}  
0.02375210 76360  
[53] {Organic Tomato Cluster}  
0.01999736 64289  
[54] {Organic Whole String Cheese}  
0.01856247 59676  
[55] {Organic Italian Parsley Bunch}  
0.01885642 60621  
[56] {Organic Red Bell Pepper}  
0.01788095 57485  
[57] {Half & Half}  
0.02153024 69217  
[58] {100% Whole Wheat Bread}  
0.01891707 60816  
[59] {Organic Blackberries}  
0.01920635 61746  
[60] {Organic Granny Smith Apple}  
0.01828345 58779  
[61] {Green Bell Pepper}  
0.01725853 55484  
[62] {Organic Peeled Whole Baby Carrots}  
0.01722338 55371  
[63] {Red Peppers}  
0.01719508 55280  
[64] {Organic Garnet Sweet Potato (Yam)}  
0.01615584 51939  
[65] {Organic Kiwi}  
0.01559657 50141  
[66] {Red Vine Tomato}  
0.01692415 54409  
[67] {Organic D'Anjou Pears}

---

---

0.01521522 48915  
[68] {Organic Unsweetened Almond Milk}  
0.01800848 57895  
[69] {Organic Ginger Root}  
0.01620188 52087  
[70] {Blueberries}  
0.01740224 55946  
[71] {Extra Virgin Olive Oil}  
0.01563203 50255  
[72] {Grated Parmesan}  
0.01407147 45238  
[73] {Small Hass Avocado}  
0.01524819 49021  
[74] {Bunched Cilantro}  
0.01416230 45530  
[75] {Boneless Skinless Chicken Breasts}  
0.01564012 50281  
[76] {Unsweetened Almondmilk}  
0.01541864 49569  
[77] {Uncured Genoa Salami}  
0.01290035 41473  
[78] {Organic Lacinato (Dinosaur) Kale}  
0.01339617 43067  
[79] {Jalapeno Peppers}  
0.01323878 42561  
[80] {Organic Reduced Fat 2% Milk}  
0.01488052 47839  
[81] {Organic Navel Orange}  
0.01328855 42721  
[82] {Red Onion}  
0.01337813 43009  
[83] {Lime Sparkling Water}  
0.01447833 46546  
[84] {Raspberries}  
0.01792916 57640  
[85] {Broccoli Crown}  
0.01307796 42044  
[86] {Orange Bell Pepper}  
0.01218959 39188  
[87] {Organic Black Beans}  
0.01182037 38001  
[88] {Organic Bartlett Pear}



---

0.01172861 37706  
[89] {Organic Broccoli}  
0.01204402 38720  
[90] {Roma Tomato}  
0.01250811 40212  
[91] {Clementines, Bag}  
0.01238773 39825  
[92] {Shredded Parmesan}  
0.01131149 36365  
[93] {Organic Grade A Free Range Large Brown Eggs}  
0.01245616 40045  
[94] {Spring Water}  
0.01744610 56087  
[95] {Grape White/Green Seedless}  
0.01212209 38971  
[96] {Organic Sticks Low Moisture Part Skim Mozzarella String Cheese}  
0.01064085 34209  
[97] {Organic Carrot Bunch}  
0.01114196 35820  
[98] {Large Alfresco Eggs}  
0.01255912 40376  
[99] {Garlic}  
0.01138987 36617  
[100] {Sparkling Natural Mineral Water}  
0.01335480 42934  
[101] {Organic Romaine Lettuce}  
0.01084770 34874  
[102] {Unsalted Butter}  
0.01109437 35667  
[103] {Bartlett Pears}  
0.01101536 35413  
[104] {Feta Cheese Crumbles}  
0.01006291 32351  
[105] {100% Raw Coconut Water}  
0.01181166 37973  
[106] {Hass Avocados}  
0.01562145 50221  
[107] {Organic Russet Potato}  
0.01011984 32534  
[108] {Granny Smith Apples}  
0.01125612 36187  
[109] {Organic Broccoli Florets}

```

0.01022964  32887
[110] {Organic Reduced Fat Milk}
0.01109313  35663
[111] {2% Reduced Fat Milk}
0.01153731  37091
[112] {Sparkling Lemon Water}
0.01019884  32788
[113] {Unsweetened Original Almond Breeze Almond Milk}
0.01075719  34583
[114] {Pure Irish Butter}
0.01054723  33908
[115] {Whole Milk}
0.01108379  35633
[116] {Soda}
0.01113294  35791

```

## Q2. Association rules for products in orders dataset.

Apriori

Parameter specification:

```

confidence minval smax arem  aval originalSupport maxtime support minlen
      0.5      0.1      1 none FALSE                  TRUE          5      0.01      1
maxlen target  ext
      10  rules FALSE

```

Algorithmic control:

```

filter tree heap memopt load sort verbose
  0.1 TRUE TRUE  FALSE TRUE    2    TRUE

```

Absolute minimum support count: 32148

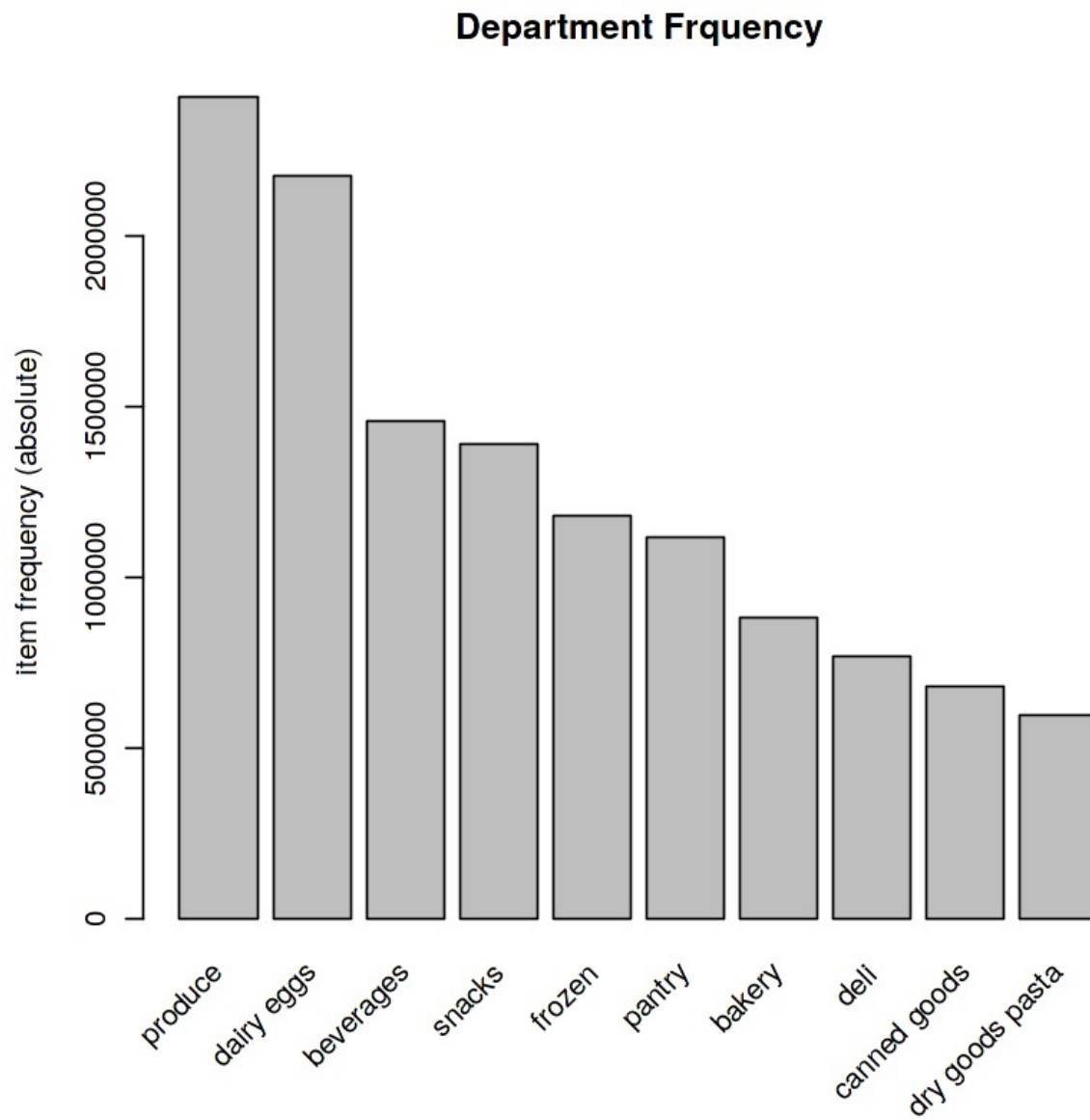
```

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[49677 item(s), 3214874 transaction(s)] done [9.05s].
sorting and recoding items ... [102 item(s)] done [0.36s].
creating transaction tree ... done [3.50s].
checking subsets of size 1 2 3 done [0.18s].
writing ... [0 rule(s)] done [0.00s].
creating S4 object ... done [1.35s].

```

---

**Q3. Frequent itemsets for departments in orders dataset**



---

Eclat

parameter specification:

```
tidLists support minlen maxlen          target  ext
  FALSE    0.1      1     80 frequent itemsets FALSE
```

algorithmic control:

```
sparse sort verbose
    7    -2    TRUE
```

Absolute minimum support count: 321487

create itemset ...

set transactions ... [21 item(s), 3214874 transaction(s)] done [0.88s].

sorting and recoding items ... [13 item(s)] done [0.30s].

creating bit matrix ... [13 row(s), 3214874 column(s)] done [0.42s].

writing ... [102 set(s)] done [0.37s].

Creating S4 object ... done [0.00s].

|      | items                                | support   | count  |
|------|--------------------------------------|-----------|--------|
| [1]  | {household,produce}                  | 0.1017561 | 327133 |
| [2]  | {dairy eggs,household}               | 0.1025857 | 329800 |
| [3]  | {breakfast,dairy eggs,produce}       | 0.1140661 | 366708 |
| [4]  | {breakfast,produce}                  | 0.1318472 | 423872 |
| [5]  | {breakfast,dairy eggs}               | 0.1350771 | 434256 |
| [6]  | {dairy eggs,meat seafood,produce}    | 0.1291578 | 415226 |
| [7]  | {meat seafood,produce}               | 0.1564752 | 503048 |
| [8]  | {dairy eggs,meat seafood}            | 0.1432821 | 460634 |
| [9]  | {dairy eggs,dry goods pasta,produce} | 0.1377208 | 442755 |
| [10] | {dry goods pasta,produce}            | 0.1617429 | 519983 |
| [11] | {dairy eggs,dry goods pasta}         | 0.1540110 | 495126 |
| [12] | {dry goods pasta,snacks}             | 0.1009327 | 324486 |
| [13] | {canned goods,dairy eggs,produce}    | 0.1536741 | 494043 |
| [14] | {canned goods,produce}               | 0.1863062 | 598951 |
| [15] | {canned goods,dairy eggs}            | 0.1699849 | 546480 |
| [16] | {canned goods,snacks}                | 0.1095060 | 352048 |
| [17] | {beverages,canned goods}             | 0.1037898 | 333671 |
| [18] | {canned goods,frozen}                | 0.1058325 | 340238 |
| [19] | {canned goods,pantry}                | 0.1089001 | 350100 |
| [20] | {deli,frozen,produce}                | 0.1034722 | 332650 |
| [21] | {beverages,deli,produce}             | 0.1053755 | 338769 |
| [22] | {beverages,dairy eggs,deli}          | 0.1019866 | 327874 |
| [23] | {dairy eggs,deli,produce,snacks}     | 0.1042573 | 335174 |

---

|      |                                       |           |        |
|------|---------------------------------------|-----------|--------|
| [24] | {deli,produce,snacks}                 | 0.1203721 | 386981 |
| [25] | {dairy eggs,deli,snacks}              | 0.1158842 | 372553 |
| [26] | {dairy eggs,deli,produce}             | 0.1717775 | 552243 |
| [27] | {deli,produce}                        | 0.2060591 | 662454 |
| [28] | {dairy eggs,deli}                     | 0.1930604 | 620665 |
| [29] | {deli,snacks}                         | 0.1376356 | 442481 |
| [30] | {beverages,deli}                      | 0.1217080 | 391276 |
| [31] | {deli,frozen}                         | 0.1166861 | 375131 |
| [32] | {deli,pantry}                         | 0.1047323 | 336701 |
| [33] | {bakery,pantry,produce}               | 0.1082727 | 348083 |
| [34] | {bakery,dairy eggs,pantry}            | 0.1073034 | 344967 |
| [35] | {bakery,dairy eggs,frozen,produce}    | 0.1014345 | 326099 |
| [36] | {bakery,frozen,produce}               | 0.1143784 | 367712 |
| [37] | {bakery,dairy eggs,frozen}            | 0.1147759 | 368990 |
| [38] | {bakery,beverages,dairy eggs,produce} | 0.1034009 | 332421 |
| [39] | {bakery,beverages,produce}            | 0.1178587 | 378901 |
| [40] | {bakery,beverages,dairy eggs}         | 0.1194672 | 384072 |
| [41] | {bakery,dairy eggs,produce,snacks}    | 0.1126784 | 362247 |
| [42] | {bakery,produce,snacks}               | 0.1274607 | 409770 |
| [43] | {bakery,dairy eggs,snacks}            | 0.1280374 | 411624 |
| [44] | {bakery,dairy eggs,produce}           | 0.1945964 | 625603 |
| [45] | {bakery,produce}                      | 0.2289679 | 736103 |
| [46] | {bakery,dairy eggs}                   | 0.2249973 | 723338 |
| [47] | {bakery,snacks}                       | 0.1486783 | 477982 |
| [48] | {bakery,beverages}                    | 0.1403287 | 451139 |
| [49] | {bakery,frozen}                       | 0.1325561 | 426151 |
| [50] | {bakery,pantry}                       | 0.1237871 | 397960 |
| [51] | {dairy eggs,frozen,pantry,produce}    | 0.1175595 | 377939 |
| [52] | {frozen,pantry,produce}               | 0.1366579 | 439338 |
| [53] | {dairy eggs,frozen,pantry}            | 0.1321103 | 424718 |
| [54] | {beverages,dairy eggs,pantry,produce} | 0.1192137 | 383257 |
| [55] | {beverages,pantry,produce}            | 0.1413468 | 454412 |
| [56] | {beverages,dairy eggs,pantry}         | 0.1376511 | 442531 |
| [57] | {beverages,pantry,snacks}             | 0.1000736 | 321724 |
| [58] | {dairy eggs,pantry,produce,snacks}    | 0.1293192 | 415745 |
| [59] | {pantry,produce,snacks}               | 0.1510482 | 485601 |
| [60] | {dairy eggs,pantry,snacks}            | 0.1471218 | 472978 |
| [61] | {dairy eggs,pantry,produce}           | 0.2324586 | 747325 |
| [62] | {pantry,produce}                      | 0.2869403 | 922477 |
| [63] | {dairy eggs,pantry}                   | 0.2692616 | 865642 |
| [64] | {pantry,snacks}                       | 0.1779613 | 572123 |
| [65] | {beverages,pantry}                    | 0.1697867 | 545843 |

---

---

|       |                                       |           |         |
|-------|---------------------------------------|-----------|---------|
| [66]  | {frozen,pantry}                       | 0.1578426 | 507444  |
| [67]  | {beverages,dairy eggs,frozen,produce} | 0.1282284 | 412238  |
| [68]  | {beverages,frozen,produce}            | 0.1520959 | 488969  |
| [69]  | {beverages,dairy eggs,frozen}         | 0.1498706 | 481815  |
| [70]  | {beverages,frozen,snacks}             | 0.1123067 | 361052  |
| [71]  | {dairy eggs,frozen,produce,snacks}    | 0.1400503 | 450244  |
| [72]  | {frozen,produce,snacks}               | 0.1634985 | 525627  |
| [73]  | {dairy eggs,frozen,snacks}            | 0.1607562 | 516811  |
| [74]  | {dairy eggs,frozen,produce}           | 0.2425800 | 779864  |
| [75]  | {frozen,produce}                      | 0.2981184 | 958413  |
| [76]  | {dairy eggs,frozen}                   | 0.2833797 | 911030  |
| [77]  | {frozen,snacks}                       | 0.1952356 | 627658  |
| [78]  | {beverages,frozen}                    | 0.1865012 | 599578  |
| [79]  | {beverages,dairy eggs,produce,snacks} | 0.1492936 | 479960  |
| [80]  | {beverages,produce,snacks}            | 0.1796139 | 577436  |
| [81]  | {beverages,dairy eggs,snacks}         | 0.1781777 | 572819  |
| [82]  | {beverages,dairy eggs,produce}        | 0.2611630 | 839606  |
| [83]  | {beverages,produce}                   | 0.3335636 | 1072365 |
| [84]  | {beverages,dairy eggs}                | 0.3207379 | 1031132 |
| [85]  | {beverages,snacks}                    | 0.2291035 | 736539  |
| [86]  | {dairy eggs,produce,snacks}           | 0.2704660 | 869514  |
| [87]  | {produce,snacks}                      | 0.3372617 | 1084254 |
| [88]  | {dairy eggs,snacks}                   | 0.3225479 | 1036951 |
| [89]  | {dairy eggs,produce}                  | 0.5504835 | 1769735 |
| [90]  | {produce}                             | 0.7494291 | 2409320 |
| [91]  | {dairy eggs}                          | 0.6772701 | 2177338 |
| [92]  | {snacks}                              | 0.4328154 | 1391447 |
| [93]  | {beverages}                           | 0.4533151 | 1457351 |
| [94]  | {frozen}                              | 0.3673606 | 1181018 |
| [95]  | {pantry}                              | 0.3477250 | 1117892 |
| [96]  | {bakery}                              | 0.2742117 | 881556  |
| [97]  | {deli}                                | 0.2396050 | 770300  |
| [98]  | {canned goods}                        | 0.2119228 | 681305  |
| [99]  | {dry goods pasta}                     | 0.1859675 | 597862  |
| [100] | {meat seafood}                        | 0.1787725 | 574731  |
| [101] | {breakfast}                           | 0.1633619 | 525188  |
| [102] | {household}                           | 0.1464381 | 470780  |

---

#### Q4. Association rules for departments in orders dataset

| hs  | => | rhs          | support    | confidence | lift     | count |
|---|----|--------------|------------|------------|----------|-------|
| [1] {bakery,<br>breakfast,<br>dry goods pasta,<br>pantry,<br>produce,<br>snacks}      | => | {dairy eggs} | 0.01005452 | 0.9618807  | 1.420232 | 32324 |
| [2] {bakery,<br>breakfast,<br>dry goods pasta,<br>frozen,<br>produce,<br>snacks}      | => | {dairy eggs} | 0.01114538 | 0.9618544  | 1.420193 | 35831 |
| [3] {bakery,<br>breakfast,<br>deli,<br>dry goods pasta,<br>produce}                   | => | {dairy eggs} | 0.01008127 | 0.9617211  | 1.419996 | 32410 |
| [4] {bakery,<br>breakfast,<br>deli,<br>pantry,<br>produce,<br>snacks}                 | => | {dairy eggs} | 0.01124834 | 0.9611929  | 1.419217 | 36162 |
| [5] {bakery,<br>breakfast,<br>deli,<br>frozen,<br>produce,<br>snacks}                 | => | {dairy eggs} | 0.01230188 | 0.9609067  | 1.418794 | 39549 |
| [6] {bakery,<br>beverages,<br>breakfast,<br>frozen,<br>pantry,<br>produce,<br>snacks} | => | {dairy eggs} | 0.01024550 | 0.9608518  | 1.418713 | 32938 |