Bakery transaction data
장바구니 분석

2조 임수연 장진원 최형석

목표

- 연관 규칙 이론 학습
- Apriori algorithm 구현
- 해당 Bakery의 데이터 분석가임을 가정하고, 연관 분석을 통해 마케팅 전략 도출

데이터 소개

데이터 구조

```
> str(data)
'data.frame':   21293 obs. of  4 variables:
 $ Date       : Factor w/ 159 levels "2016-10-30"
 $ Time       : Factor w/ 8240 levels "01:21:05",
 $ Transaction: int  1 2 2 3 3 3 4 5 5 5 ...
 $ Item       : Factor w/ 95 levels "Adjustment",
```

```
> head(data)
        Date     Time Transaction          Item
1 2016-10-30 09:58:11           1         Bread
2 2016-10-30 10:05:34           2  Scandinavian
3 2016-10-30 10:05:34           2  Scandinavian
4 2016-10-30 10:07:57           3 Hot chocolate
5 2016-10-30 10:07:57           3           Jam
6 2016-10-30 10:07:57           3       Cookies
```

데이터 구조

- `Date`는 Date 자료형으로, `Time`은 Period 형으로 변환

```
> str(data)
'data.frame':   21293 obs. of  4 variables:
 $ Date       : Date, format: "2016-10-30" "2016-10-30" ...
 $ Time       :Formal class 'Period' [package "lubridate"] with 6 slots
  .. ..@ .Data : num  11 34 34 57 57 57 41 3 3 3 ...
  .. ..@ year  : num  0 0 0 0 0 0 0 0 0 0 ...
  .. ..@ month : num  0 0 0 0 0 0 0 0 0 0 ...
  .. ..@ day   : num  0 0 0 0 0 0 0 0 0 0 ...
  .. ..@ hour  : num  9 10 10 10 10 10 10 10 10 10 ...
  .. ..@ minute: num  58 5 5 7 7 7 8 13 13 13 ...
 $ Transaction: int  1 2 2 3 3 3 4 5 5 5 ...
 $ Item       : Factor w/ 95 levels "Adjustment","Afternoon with the baker",..
```

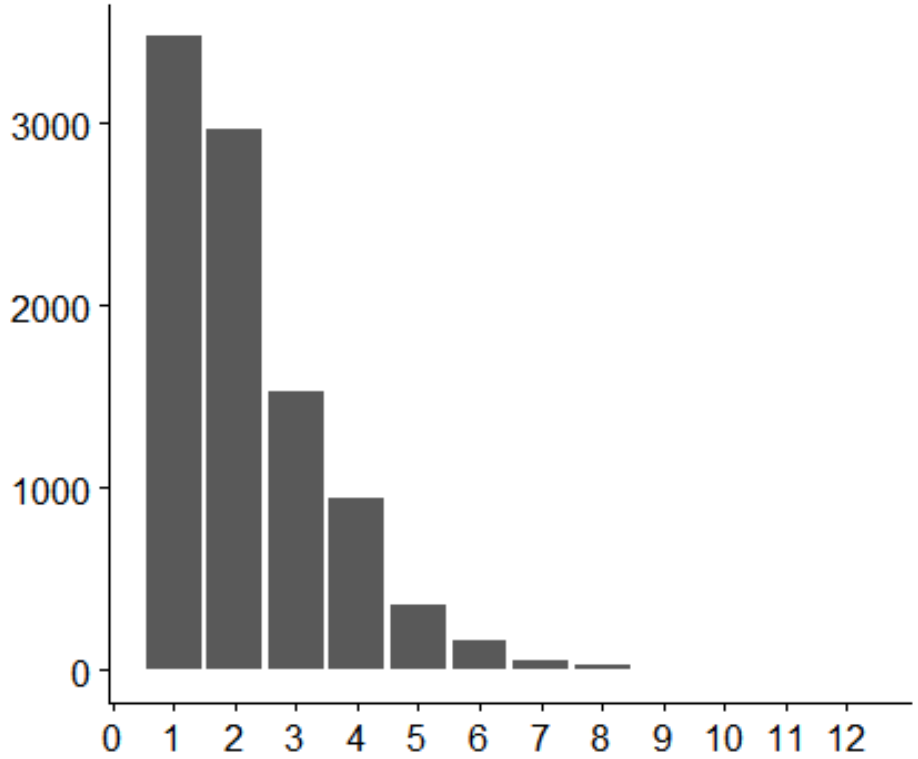
데이터 구조

```
> summary(data)
      Date                  Time
 Min.   :2016-10-30   Min.   :1H 21M 5S
 1st Qu.:2016-12-03   1st Qu.:10H 56M 40S
 Median :2017-01-21   Median :12H 37M 30S
 Mean   :2017-01-17   Mean   :12H 45M 1.401258629593945
 3rd Qu.:2017-02-28   3rd Qu.:14H 29M 37S
 Max.   :2017-04-09   Max.   :23H 38M 41S


  Transaction         Item
 Min.   :   1   Coffee :5471
 1st Qu.:2548   Bread  :3325
 Median :5067   Tea    :1435
 Mean   :4952   Cake   :1025
 3rd Qu.:7329   Pastry : 856
 Max.   :9684   NONE   : 786
                (Other):8395
```

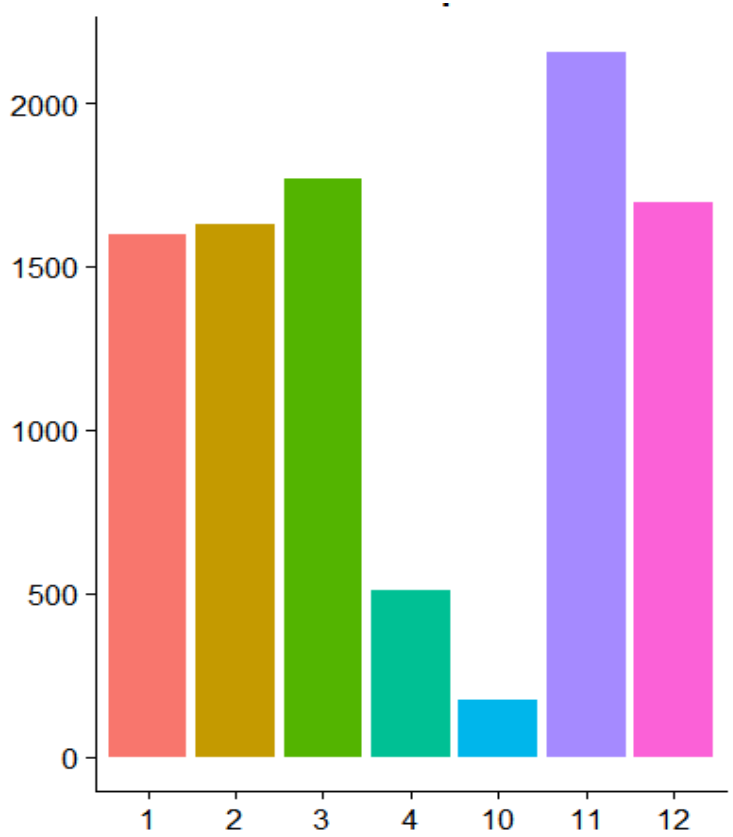
EDA

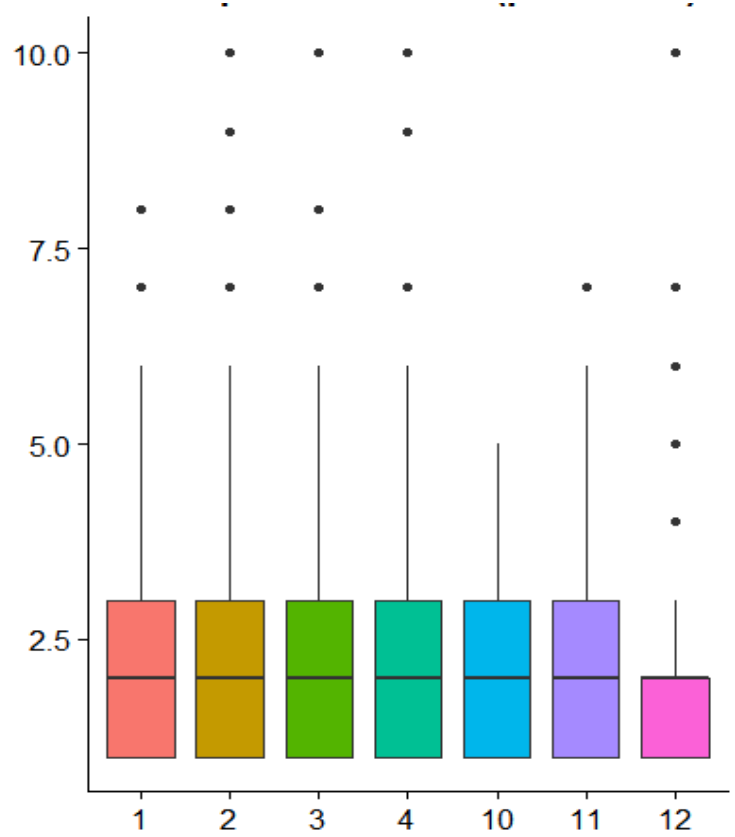
구매단위 분포

```
> summary(trans$Freq)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.000   2.000   2.234   3.000  12.000
```

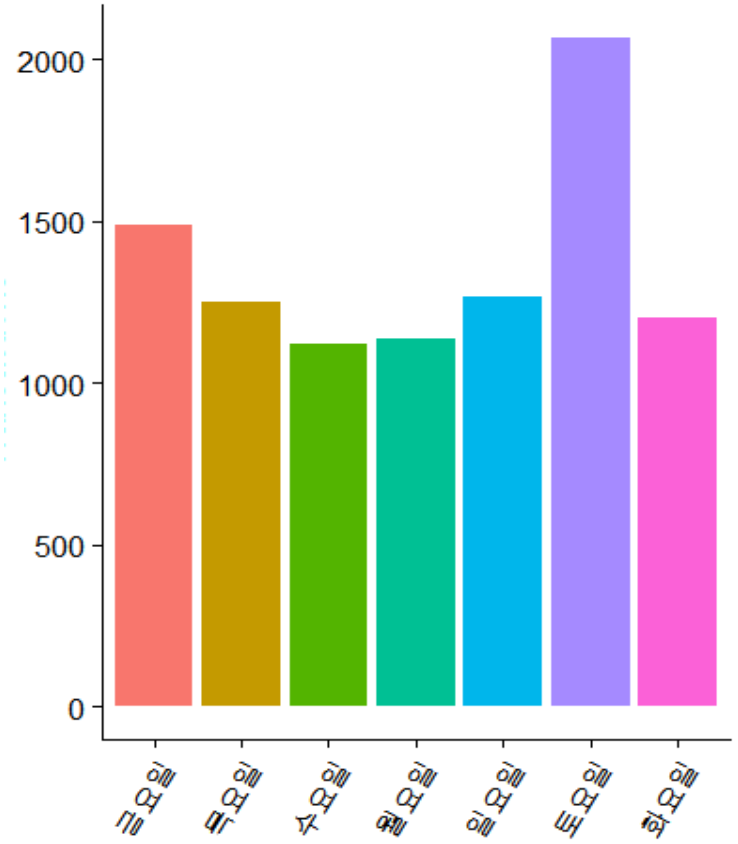


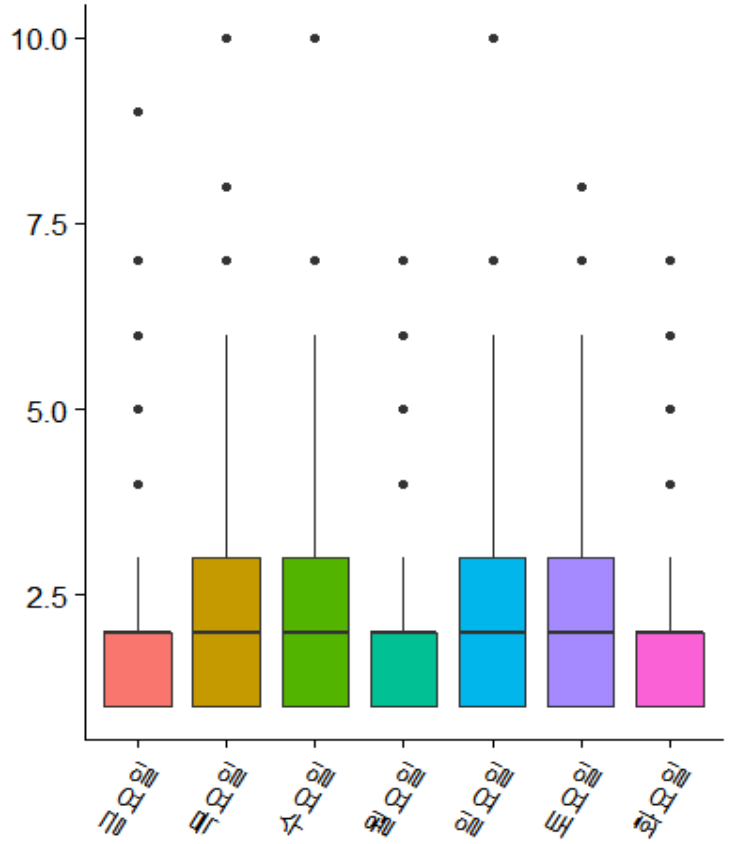
월별 판매량 및 구매단위 수 분포



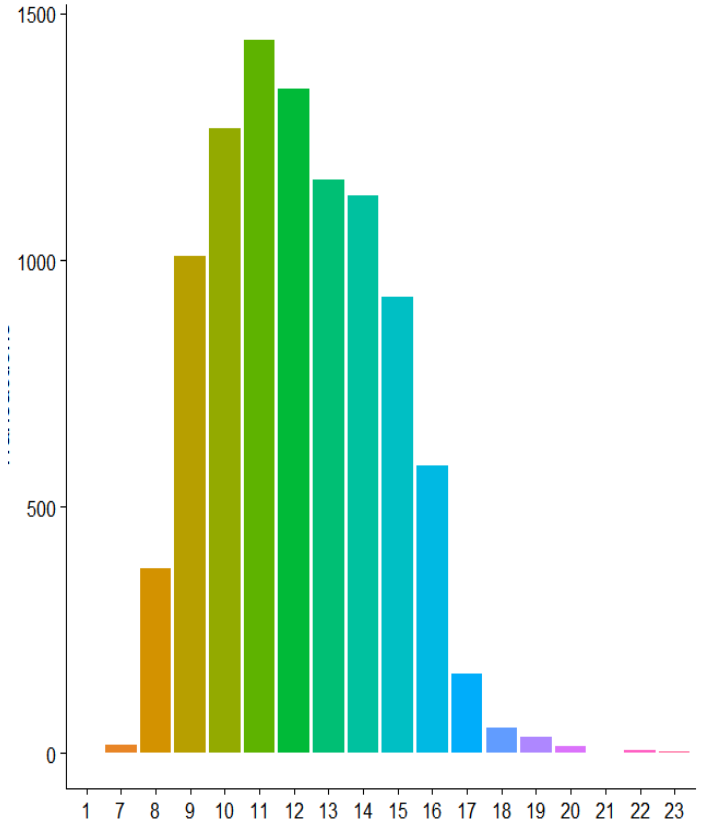


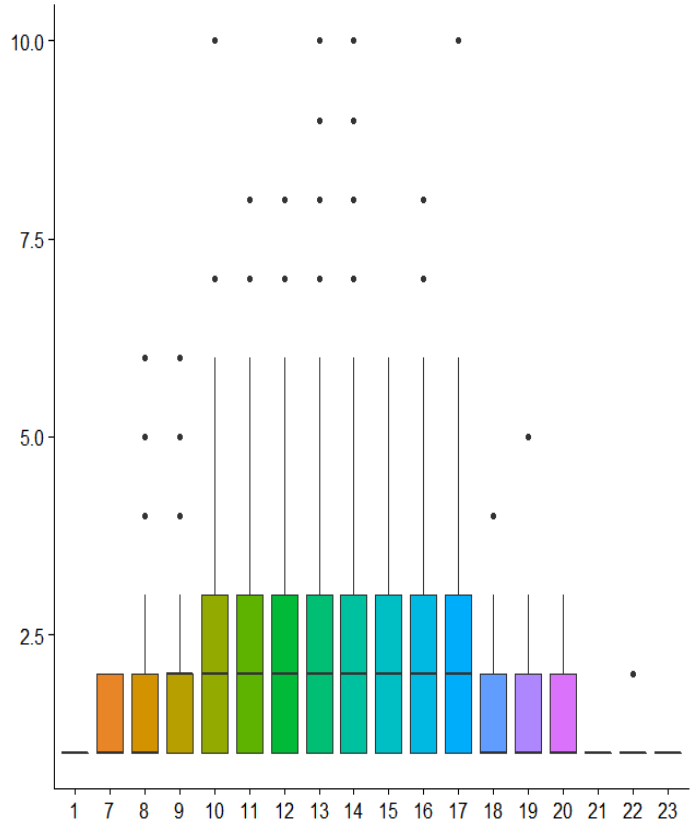
요일별 판매량 및 구매단위 수 분포



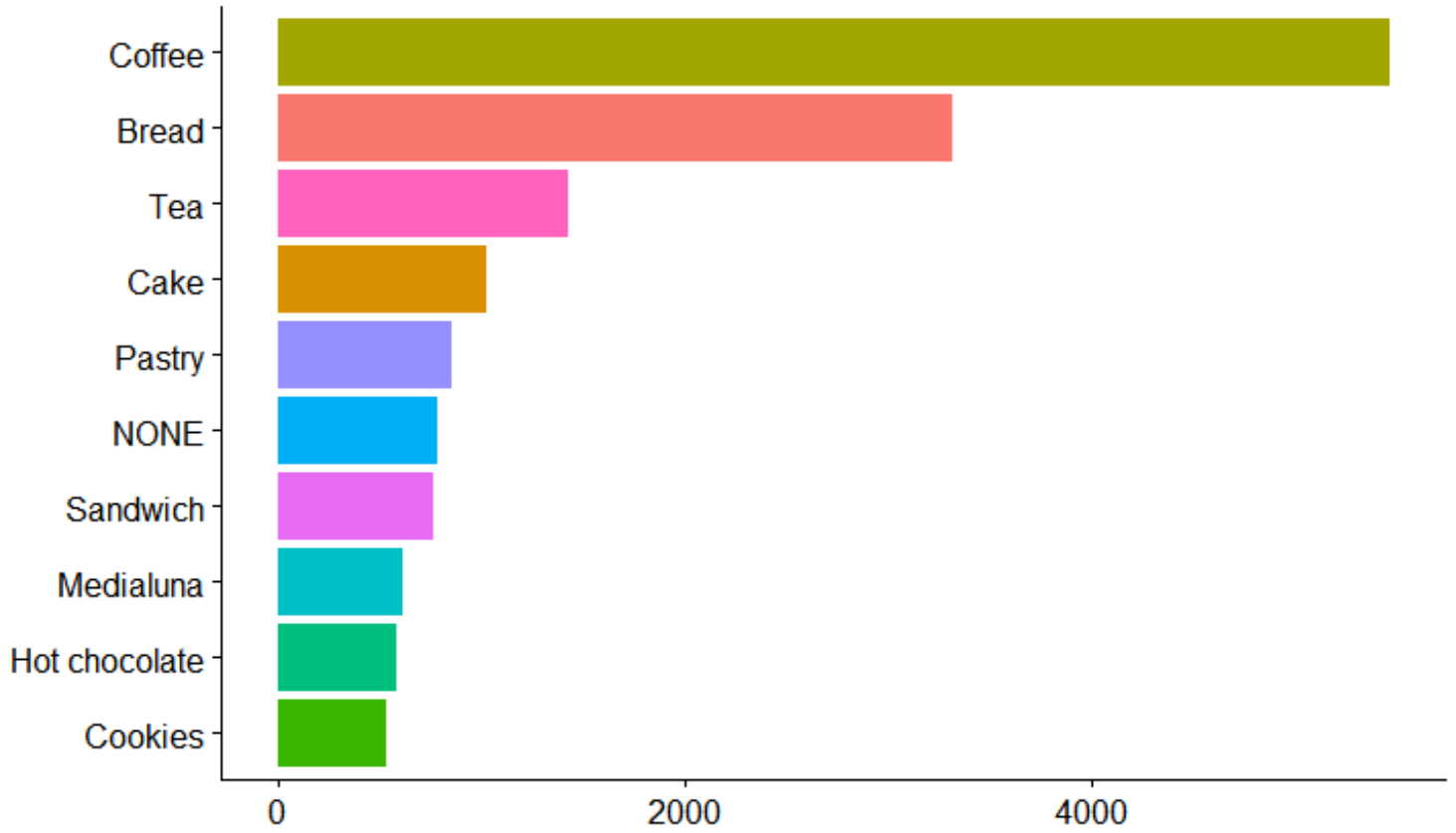


시간대별 판매량 및 구매단위 수 분포

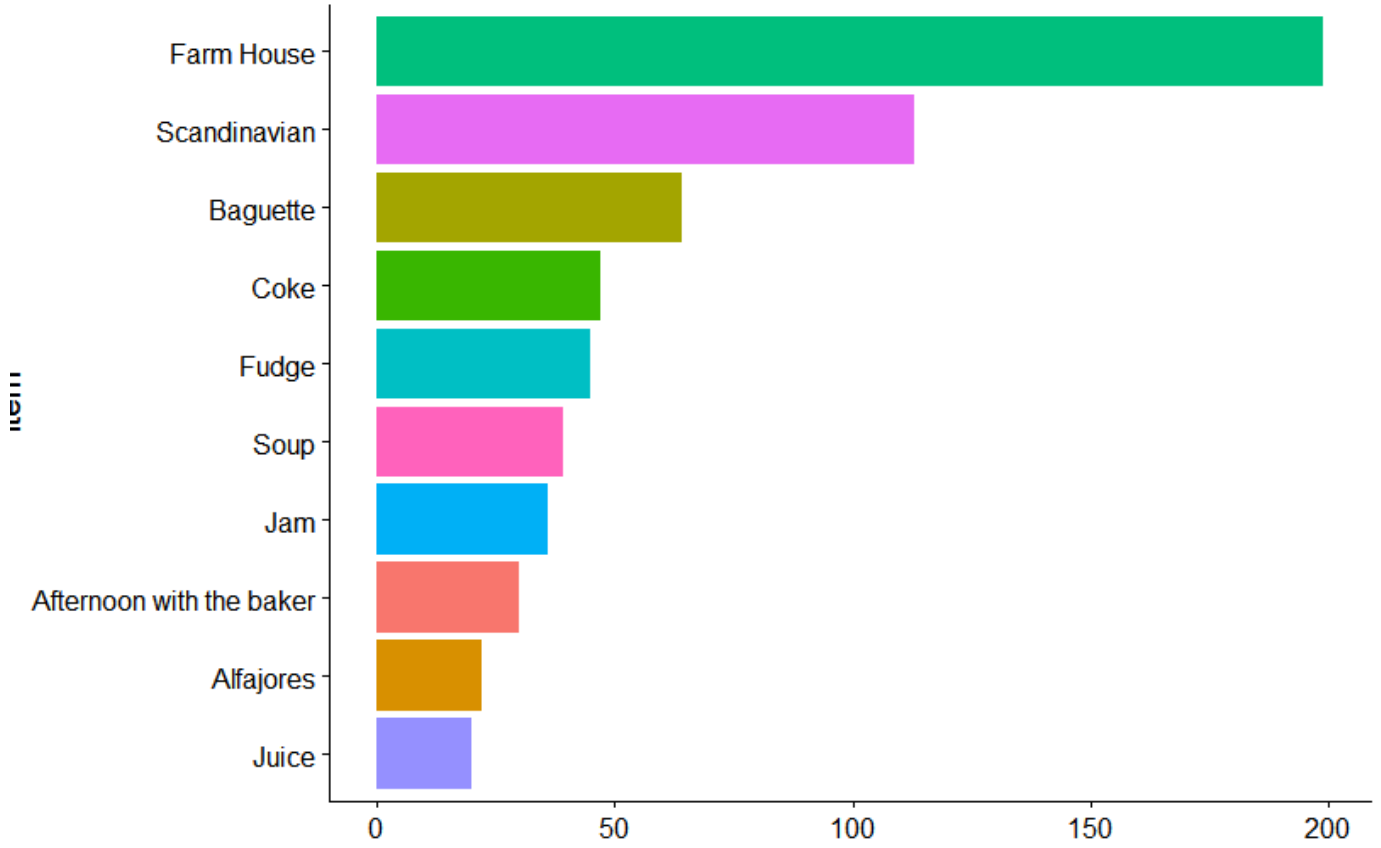




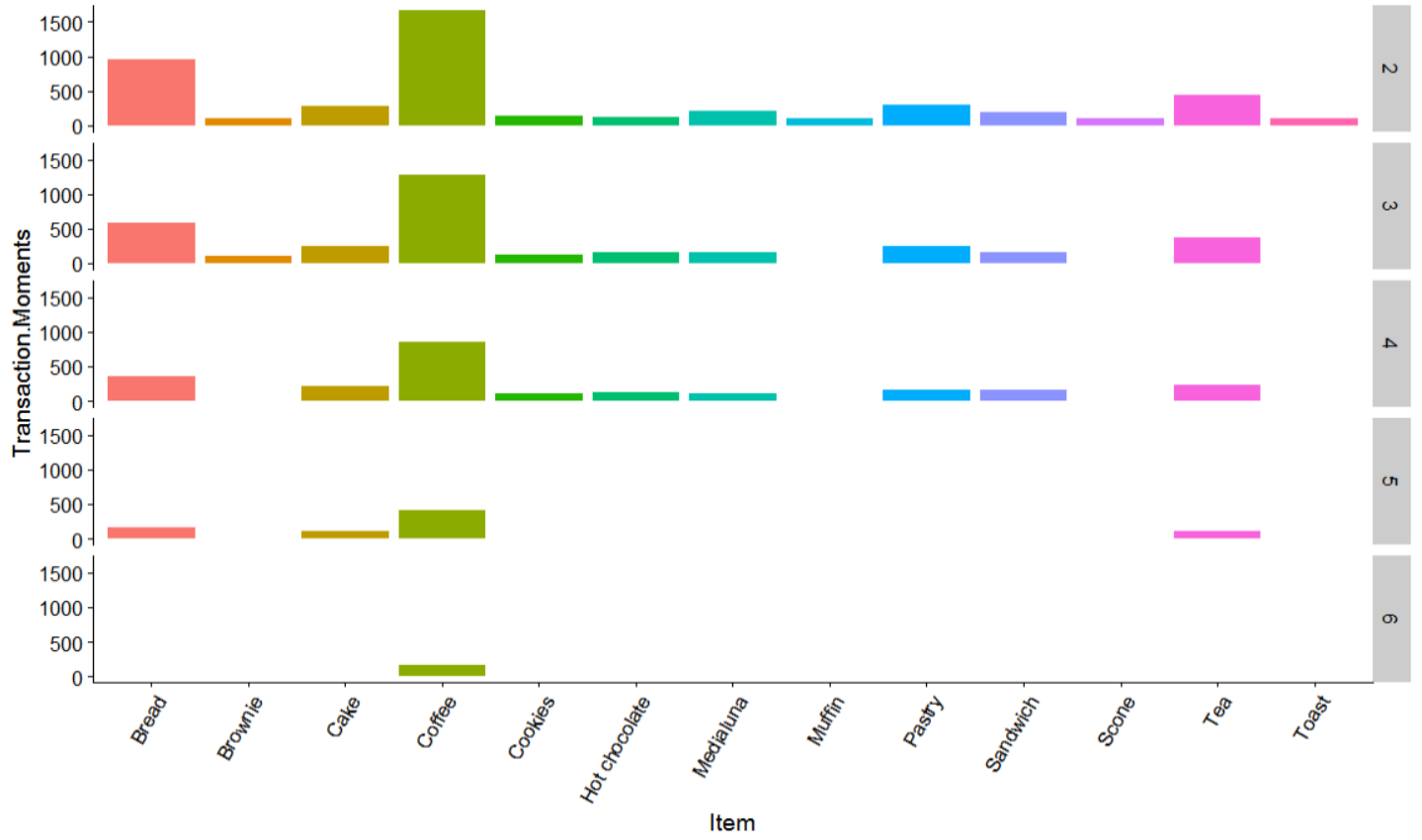
구매품목별 판매량 분포


Coffee
Bread
Tea
Cake
Pastry
NONE
Sandwich
Medialuna
Hot chocolate
Cookies
0
2000
4000

가장 많이 판매된 상품 (단독 구매)

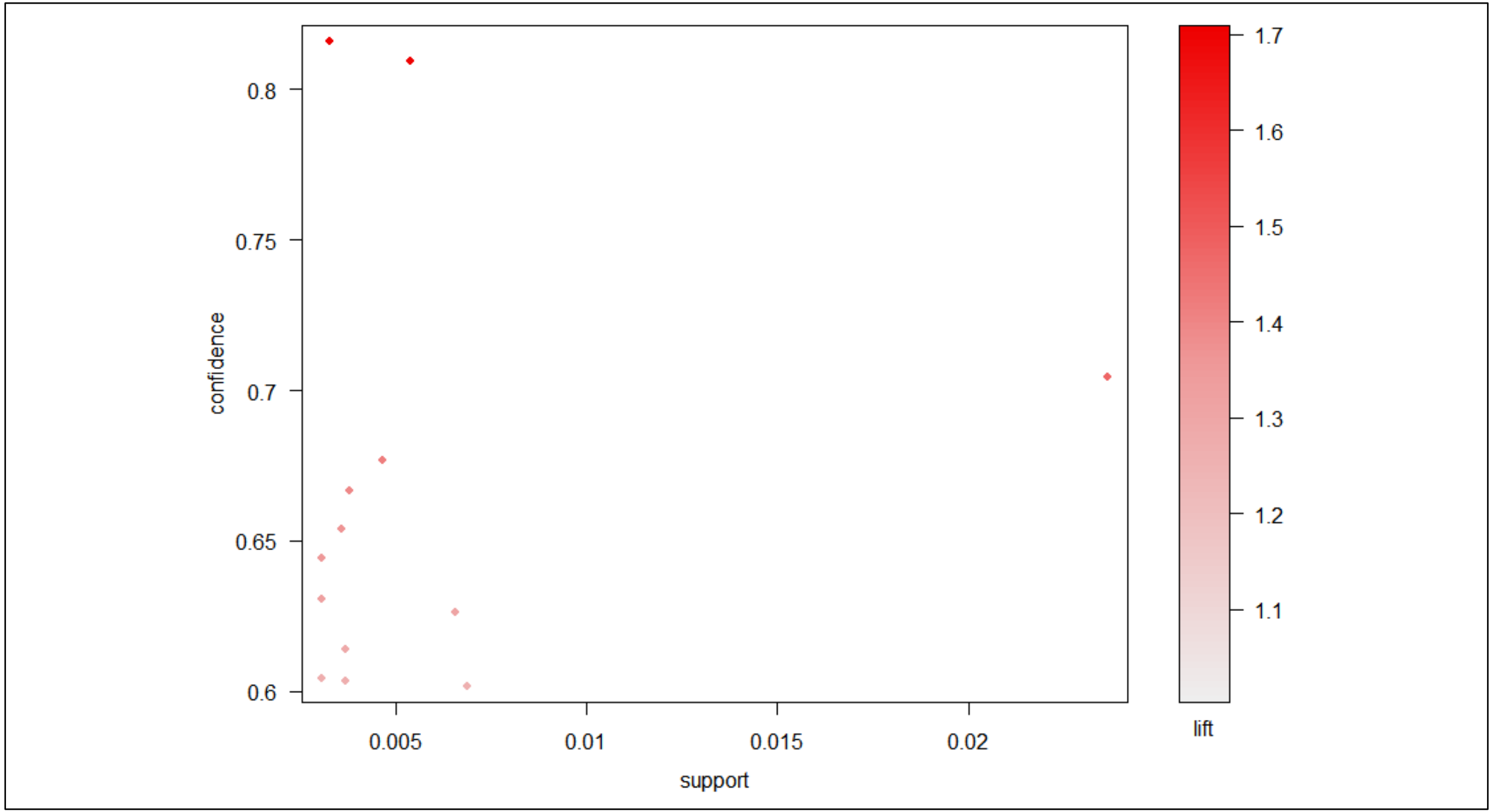


가장 많이 판매된 상품 (2개 이상 구매)



ASSOCIATION RULE

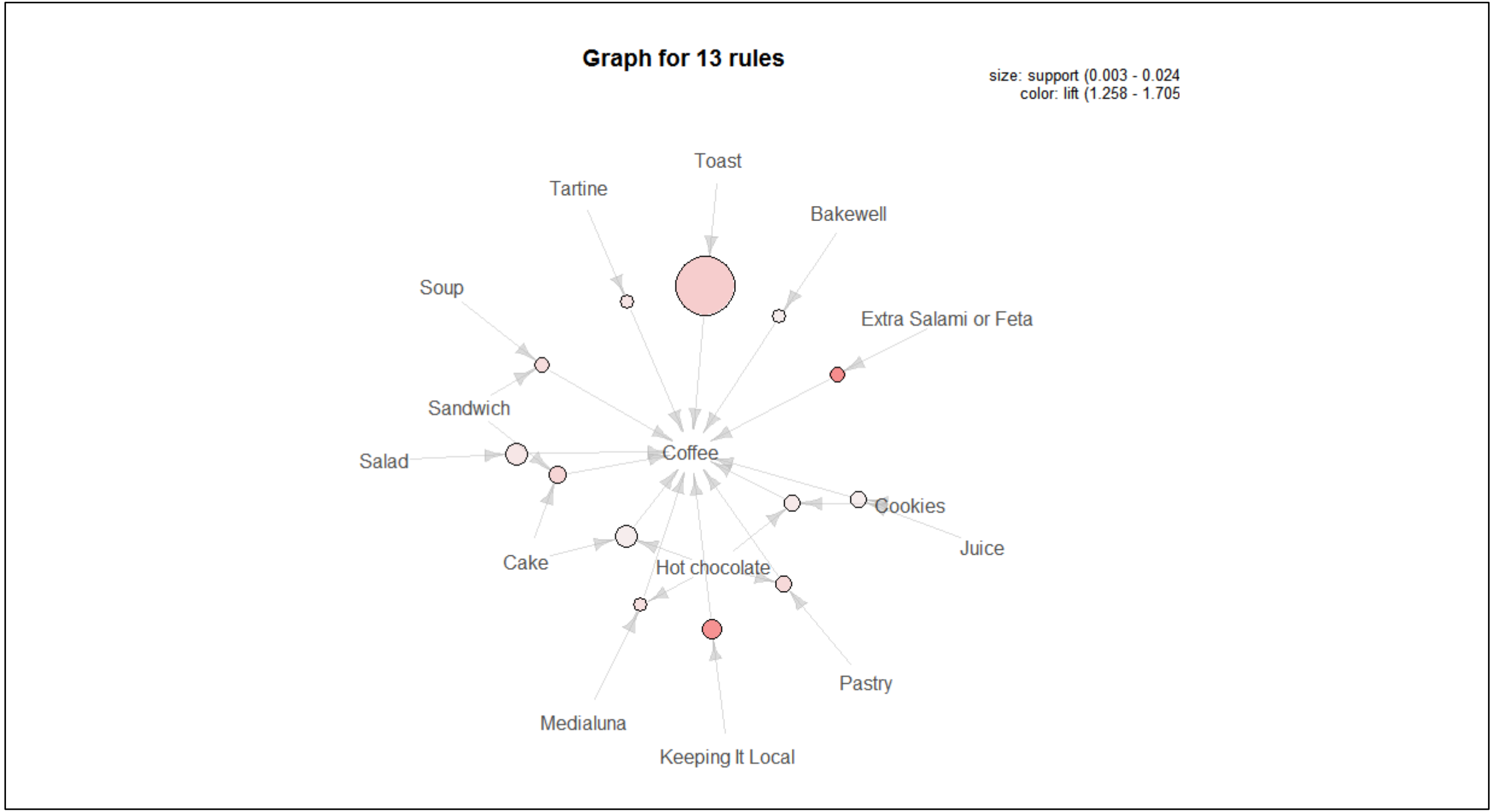
```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.6    0.1    1 none FALSE            TRUE       5   0.003      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 28

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[94 item(s), 9465 transaction(s)] done [0.01s].
sorting and recoding items ... [41 item(s)] done [0.00s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [13 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```


confidence
0.8
0.75
0.7
0.65
0.6
0.005
0.01
0.015
0.02
support
lift
1.7
1.6
1.5
1.4
1.3
1.2
1.1

	lhs		rhs	support	confidence	lift	count
[1]	{Toast}	=>	{Coffee}	0.023666138	0.7044025	1.472431	224
[2]	{Cake,Hot chocolate}	=>	{Coffee}	0.006867406	0.6018519	1.258067	65
[3]	{Salad}	=>	{Coffee}	0.006550449	0.6262626	1.309094	62
[4]	{Keeping It Local}	=>	{Coffee}	0.005388273	0.8095238	1.692169	51
[5]	{Cake,Sandwich}	=>	{Coffee}	0.004648706	0.6769231	1.414990	44
[6]	{Hot chocolate,Pastry}	=>	{Coffee}	0.003803487	0.6666667	1.393551	36
[7]	{Cookies,Juice}	=>	{Coffee}	0.003697834	0.6034483	1.261404	35
[8]	{Cookies,Hot chocolate}	=>	{Coffee}	0.003697834	0.6140351	1.283534	35
[9]	{Sandwich,Soup}	=>	{Coffee}	0.003592182	0.6538462	1.366752	34
[10]	{Extra Salami or Feta}	=>	{Coffee}	0.003275225	0.8157895	1.705267	31
[11]	{Tartine}	=>	{Coffee}	0.003063920	0.6304348	1.317815	29
[12]	{Bakewell}	=>	{Coffee}	0.003063920	0.6041667	1.262906	29
[13]	{Hot chocolate,Medialuna}	=>	{Coffee}	0.003063920	0.6444444	1.347100	29

	lhs		rhs	support	confidence	lift	count
[1]	{Extra Salami or Feta}	=>	{Coffee}	0.00327522	0.8157895	1.705267	31
[2]	{Keeping It Local}	=>	{Coffee}	0.00538827	0.8095238	1.692169	51
[3]	{Toast}	=>	{Coffee}	0.02366613	0.7044025	1.472431	224
[4]	{Cake,Sandwich}	=>	{Coffee}	0.00464870	0.6769231	1.414990	44
[5]	{Hot chocolate,Pastry}	=>	{Coffee}	0.00380348	0.6666667	1.393551	36
[6]	{Sandwich,Soup}	=>	{Coffee}	0.00359218	0.6538462	1.366752	34
[7]	{Hot chocolate,Medialuna}	=>	{Coffee}	0.00306392	0.6444444	1.347100	29
[8]	{Tartine}	=>	{Coffee}	0.00306392	0.6304348	1.317815	29
[9]	{Salad}	=>	{Coffee}	0.00655044	0.6262626	1.309094	62
[10]	{Cookies,Hot chocolate}	=>	{Coffee}	0.00369783	0.6140351	1.283534	35
[11]	{Bakewell}	=>	{Coffee}	0.00306392	0.6041667	1.262906	29
[12]	{Cookies,Juice}	=>	{Coffee}	0.00369783	0.6034483	1.261404	35
[13]	{Cake,Hot chocolate}	=>	{Coffee}	0.00686740	0.6018519	1.258067	65

	lhs		rhs	support	confidence	lift	count
[1]	{Extra Salami or Feta}	=>	{Coffee}	0.003275225	0.8157895	1.705267	31
[2]	{Keeping It Local}	=>	{Coffee}	0.005388273	0.8095238	1.692169	51
[3]	{Toast}	=>	{Coffee}	0.023666138	0.7044025	1.472431	224
[4]	{Cake,Sandwich}	=>	{Coffee}	0.004648706	0.6769231	1.414990	44
[5]	{Hot chocolate,Pastry}	=>	{Coffee}	0.003803487	0.6666667	1.393551	36
[6]	{Sandwich,Soup}	=>	{Coffee}	0.003592182	0.6538462	1.366752	34
[7]	{Hot chocolate,Medialuna}	=>	{Coffee}	0.003063920	0.6444444	1.347100	29
[8]	{Tartine}	=>	{Coffee}	0.003063920	0.6304348	1.317815	29
[9]	{Salad}	=>	{Coffee}	0.006550449	0.6262626	1.309094	62
[10]	{Cookies,Hot chocolate}	=>	{Coffee}	0.003697834	0.6140351	1.283534	35
[11]	{Bakewell}	=>	{Coffee}	0.003063920	0.6041667	1.262906	29
[12]	{Cookies,Juice}	=>	{Coffee}	0.003697834	0.6034483	1.261404	35
[13]	{Cake,Hot chocolate}	=>	{Coffee}	0.006867406	0.6018519	1.258067	65


Graph for 13 rules
size: support (0.003 - 0.024
color: lift (1.258 - 1.705
Toast
Tartine
Bakewell
Soup
Extra Salami or Feta
Sandwich
Salad
Coffee
Cookies
Juice
Cake
Hot chocolate
Pastry
Medialuna
Keeping It Local

IS측도 $= \sqrt{supp \times lift}$

서포트와 리프트간의 상이한 값을 조정해주는 값

	Items	ISmeasure
1	{Toast}	0.18667289
2	{Cake,Hot chocolate}	0.09294976
3	{Salad}	0.09260211
4	{Keeping It Local}	0.09548754
5	{Cake,Sandwich}	0.08110410
6	{Hot chocolate,Pastry}	0.07280353
7	{Cookies,Juice}	0.06829688
8	{Cookies,Hot chocolate}	0.06889337
9	{Sandwich,Soup}	0.07006870
10	{Extra Salami or Feta}	0.07473373
11	{Tartine}	0.06354273
12	{Bakewell}	0.06220484
13	{Hot chocolate,Medialuna}	0.06424488

시사점

한계

- 데이터가 편향되어 있고, 샘플 수가 부족
  구체적인 분석 결과 도출이 어려움
- 고객데이터가 없어서 다양한 마케팅 전략 도출 실패

의의

- 추천시스템 개발에 필요한 supp, conf, lift 학습
- 연관규칙 분석을 위해 트랜잭션 데이터를 다뤄보는 경험