Project 3: Preference and Choice / Market Basket Analysis

DA 420

Marjorie Blanco

Part 1

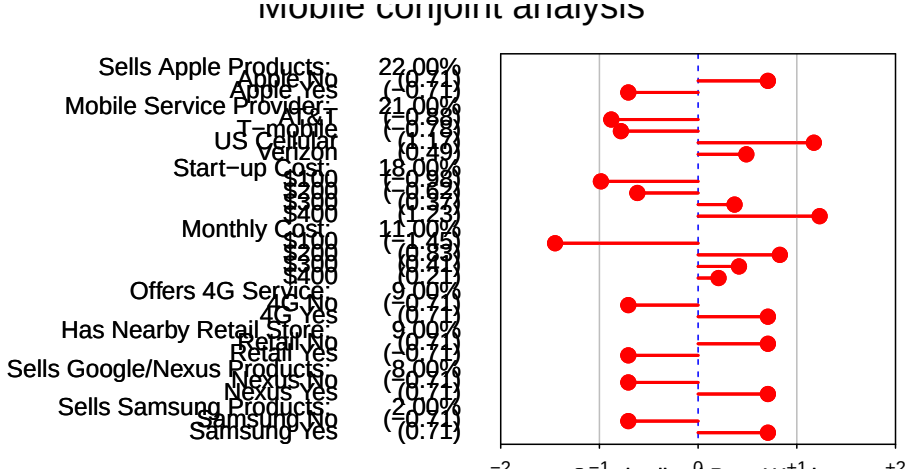
Read mobile data

brand	startup	monthly	service	retail	apple	samsung	google	ranking
AT&T	$400	$300	4G No	Retail Yes	Apple Yes	Samsung No	Nexus Yes	4
Verizon	$100	$200	4G Yes	Retail No	Apple Yes	Samsung No	Nexus Yes	8
AT&T	$200	$400	4G Yes	Retail No	Apple Yes	Samsung Yes	Nexus No	3
T-mobile	$200	$200	4G No	Retail Yes	Apple Yes	Samsung No	Nexus No	1
Verizon	$300	$100	4G No	Retail Yes	Apple Yes	Samsung Yes	Nexus No	2
T-mobile	$400	$100	4G Yes	Retail No	Apple Yes	Samsung Yes	Nexus Yes	9

```
##
## Call:
## lm.default(formula = main.effects.model, data = conjoint.data.frame)
##
## Residuals:
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
## -1 -1 -1  1 -1  1  1  1  1 -1  1 -1 -1  1  1 -1
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.500      1.000   8.500   0.0746 .
## brand1        -2.250      1.732  -1.299   0.4177
## brand2        -2.000      1.732  -1.155   0.4544
## brand3         3.000      1.732   1.732   0.3333
## startup1      -2.000      1.732  -1.155   0.4544
## startup2      -1.250      1.732  -0.722   0.6020
## startup3       0.750      1.732   0.433   0.7399
## monthly1      -1.750      1.732  -1.010   0.4967
## monthly2       1.000      1.732   0.577   0.6667
## monthly3       0.500      1.732   0.289   0.8211
## service1      -1.125      1.000  -1.125   0.4626
## retail1        1.125      1.000   1.125   0.4626
## apple1         2.750      1.000   2.750   0.2220
## samsung1      -0.250      1.000  -0.250   0.8440
## google1       -1.000      1.000  -1.000   0.5000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4 on 1 degrees of freedom
## Multiple R-squared:  0.9529, Adjusted R-squared:  0.2941
## F-statistic: 1.446 on 14 and 1 DF,  p-value: 0.5803
```

```
##
##
## apple Levels:  Apple No Apple Yes
##   Part-Worths:  2.75 -2.75
##   Standardized Part-Worths:  0.71 -0.71
##   Attribute Importance:  22.00
##
## brand Levels:  AT&T T-mobile US Cellular Verizon
##   Part-Worths:  -2.25 -2.00 3.00 1.25
##   Standardized Part-Worths:  -0.88 -0.78 1.17 0.49
##   Attribute Importance:  21.00
##
## startup Levels:  $100 $200 $300 $400
##   Part-Worths:  -2.00 -1.25 0.75 2.50
##   Standardized Part-Worths:  -0.98 -0.62 0.37 1.23
##   Attribute Importance:  18.00
##
## monthly Levels:  $100 $200 $300 $400
##   Part-Worths:  -1.75 1.00 0.50 0.25
##   Standardized Part-Worths:  -1.45 0.83 0.41 0.21
##   Attribute Importance:  11.00
##
## service Levels:  4G No 4G Yes
##   Part-Worths:  -1.13 1.13
##   Standardized Part-Worths:  -0.71 0.71
##   Attribute Importance:  9.00
##
## retail Levels:  Retail No Retail Yes
##   Part-Worths:  1.12 -1.12
##   Standardized Part-Worths:  0.71 -0.71
##   Attribute Importance:  9.00
##
## google Levels:  Nexus No Nexus Yes
##   Part-Worths:  -1.00 1.00
##   Standardized Part-Worths:  -0.71 0.71
##   Attribute Importance:  8.00
##
## samsung Levels:  Samsung No Samsung Yes
##   Part-Worths:  -0.25 0.25
##   Standardized Part-Worths:  -0.71 0.71
##   Attribute Importance:  2.00
```

```r
# plotting of spine chart begins here
spine.chart(conjoint.results)
```




```
## pdf
##   2
```

Part 2

The data set consists of 9835 market baskets across 169 generically-labeled grocery items.

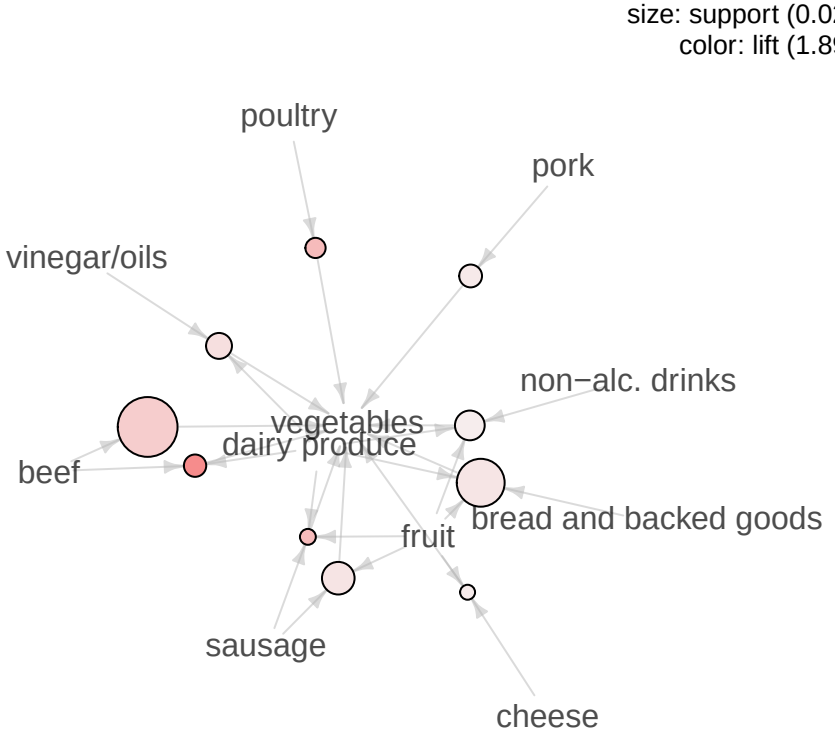
```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##        0.05    0.1    1 none FALSE            TRUE       5   0.025      1
##  maxlen target   ext
##      10  rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 245
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[55 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [32 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [344 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
## set of 344 rules
##
## rule length distribution (lhs + rhs):sizes
##   1   2   3   4
##  21 162 129  32
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.0     2.0     2.0     2.5     3.0     4.0
##
## summary of quality measures:
##     support          confidence          lift            count
##  Min.   :0.02542   Min.   :0.05043   Min.   :0.6669   Min.   : 250.0
##  1st Qu.:0.03030   1st Qu.:0.18202   1st Qu.:1.2498   1st Qu.: 298.0
##  Median :0.03854   Median :0.39522   Median :1.4770   Median : 379.0
##  Mean   :0.05276   Mean   :0.37658   Mean   :1.4831   Mean   : 518.9
##  3rd Qu.:0.05236   3rd Qu.:0.51271   3rd Qu.:1.7094   3rd Qu.: 515.0
##  Max.   :0.44301   Max.   :0.79841   Max.   :2.4073   Max.   :4357.0
##
## mining info:
##       data ntransactions support confidence
##  groceries          9835   0.025       0.05
```

Vegetables

```
plotRule(rules, "vegetables", "market_basket_vegetables_rules.pdf")
```

```
##      lhs                          rhs             support confidence     lift count
## [1]  {beef,
##       dairy produce}          => {vegetables} 0.02989324  0.6074380 2.225010   294
## [2]  {poultry}                => {vegetables} 0.02897814  0.5745968 2.104715   285
## [3]  {dairy produce,
##       fruit,
##       sausage}                => {vegetables} 0.02714794  0.5741935 2.103238   267
## [4]  {beef}                   => {vegetables} 0.04585663  0.5595533 2.049612   451
## [5]  {dairy produce,
##       vinegar/oils}           => {vegetables} 0.03141840  0.5355286 1.961610   309
## [6]  {fruit,
##       sausage}                => {vegetables} 0.03426538  0.5290424 1.937852   337
## [7]  {bread and backed goods,
##       dairy produce,
##       fruit}                  => {vegetables} 0.04077275  0.5276316 1.932684   401
## [8]  {pork}                   => {vegetables} 0.03009659  0.5220459 1.912224   296
## [9]  {cheese,
##       fruit}                  => {vegetables} 0.02674123  0.5197628 1.903861   263
## [10] {dairy produce,
##       fruit,
##       non-alc. drinks}        => {vegetables} 0.03304525  0.5183413 1.898654   325
## Rules for vegetables
```




```
## set of 10 rules
```

Top 3 Rules

- Customers who buy beef and dairy produce are 2.2 times as likely to buy vegetables as customers from the entire data set.
- Customers who buy poultry are 2.1 times as likely to buy vegetables as customers from the entire data set.
- Customers who buy dairy produce, fruit, and sausage are 2.1 times as likely to buy vegetables as customers from the entire data set.

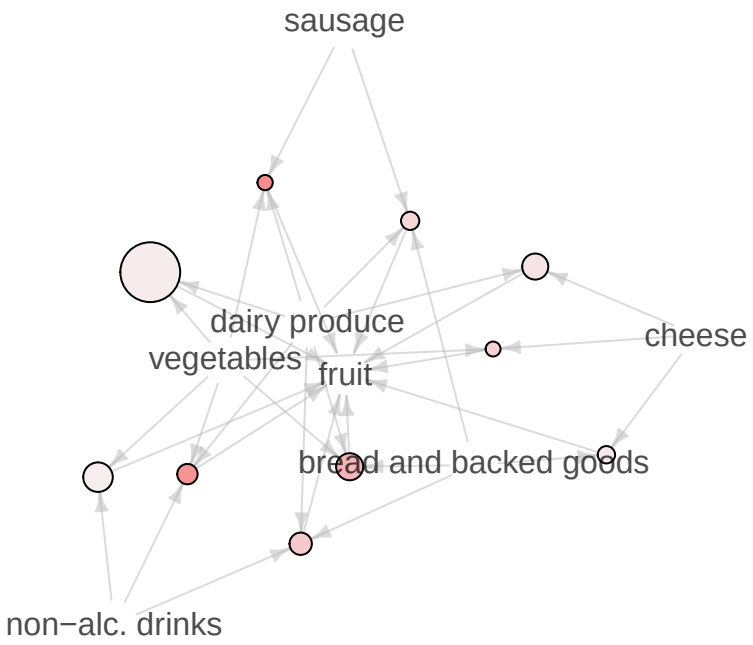
To increase profits, I recomend to promote a dinner kit containing meat (beef, poultry or sausage) with vegetable mix.

Fruit

```
##      lhs                          rhs        support confidence     lift count
## [1]  {dairy produce,
##       sausage,
##       vegetables}             => {fruit} 0.02714794  0.5154440 2.069140   267
## [2]  {dairy produce,
##       non-alc. drinks,
```

```
##       vegetables}             => {fruit} 0.03304525  0.5126183 2.057796   325
## [3]  {bread and backed goods,
##       dairy produce,
##       vegetables}             => {fruit} 0.04077275  0.4975186 1.997182   401
## [4]  {bread and backed goods,
##       dairy produce,
##       non-alc. drinks}        => {fruit} 0.03528216  0.4880450 1.959152   347
## [5]  {cheese,
##       vegetables}             => {fruit} 0.02674123  0.4834559 1.940730   263
## [6]  {bread and backed goods,
##       dairy produce,
##       sausage}                => {fruit} 0.03060498  0.4785374 1.920986   301
## [7]  {cheese,
##       dairy produce}          => {fruit} 0.03965430  0.4687500 1.881696   390
## [8]  {bread and backed goods,
##       cheese}                 => {fruit} 0.02958821  0.4648562 1.866066   291
## [9]  {dairy produce,
##       vegetables}             => {fruit} 0.07869853  0.4618138 1.853853   774
## [10] {non-alc. drinks,
##       vegetables}             => {fruit} 0.04361973  0.4612903 1.851751   429
## Rules for fruit
```

size: support (0.027 − 0.079)
color: lift (1.852 − 2.069)




```
## set of 10 rules
```

Top 3 Rules

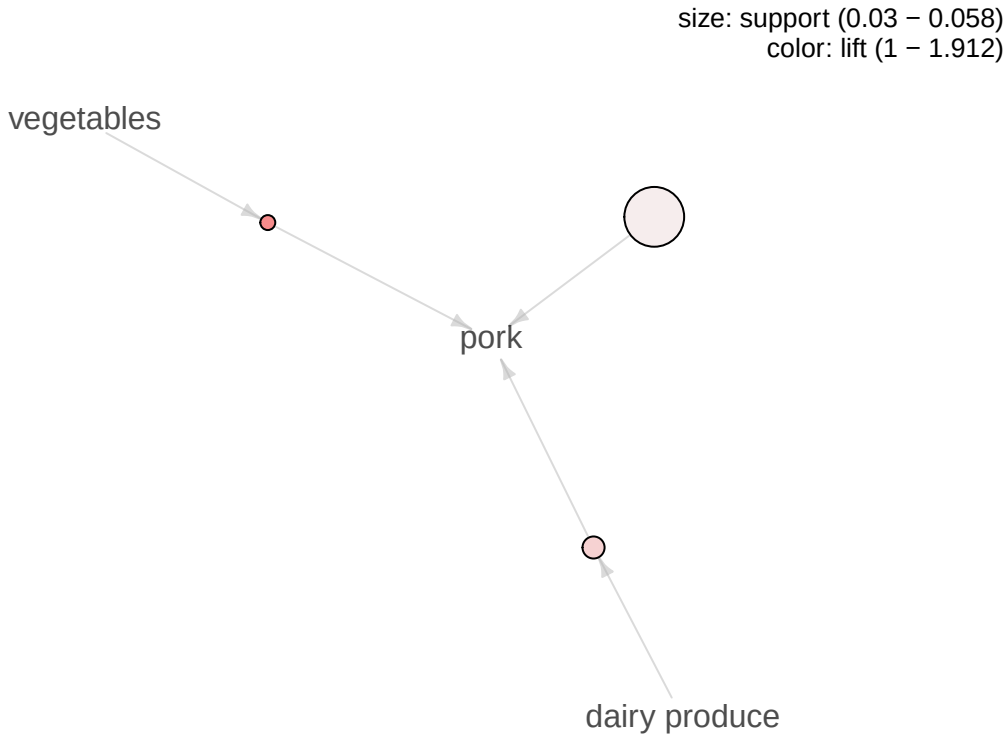
- Customers who buy dairy produce, sausage, and vegetables are 2.0 times as likely to buy fruit as customers from the entire data set.
- Customers who buy dairy produce, non-alc. drinks, and vegetables are 2.0 times as likely to buy fruit as customers from the entire data set.
- Customers who buy bread and backed goods, dairy produce, and vegetables are 1.9 times as likely to buy fruit as customers from the entire data set.

To increase profits, I recomend to promote a fruit, cheese and vegetables platter.

Meat: Pork

```
plotRule(rules, "pork", "market_basket_pork_rules.pdf")
```

```
##     lhs                rhs    support    confidence lift     count
## [1] {vegetables}    => {pork} 0.03009659 0.11024209 1.912224 296
## [2] {dairy produce} => {pork} 0.03446873 0.07780583 1.349595 339
## [3] {}              => {pork} 0.05765125 0.05765125 1.000000 567
## Rules for pork
```




```
## set of 3 rules
```

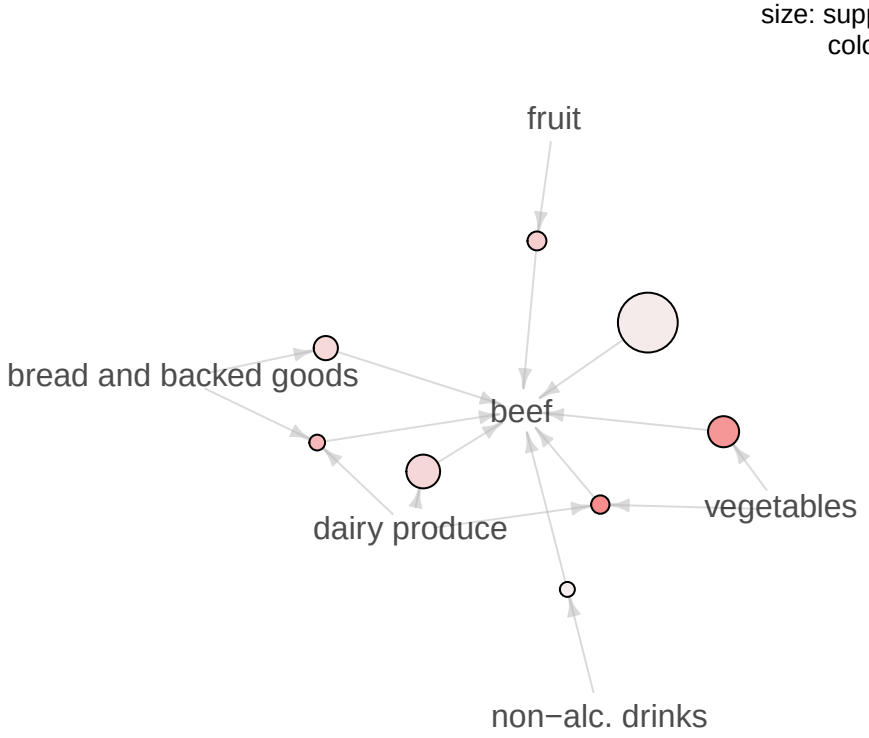
- Customers who buy vegetables are 1.9 times as likely to buy pork as customers from the entire data set.

To increase profits, I recomend to promote a dinner kit containing meat (pork) with vegetable mix.

Meat: Beef

```
plotRule(rules, "beef", "market_basket_beef_rules.pdf")
```

```
##     lhs                                        rhs    support    confidence
## [1] {dairy produce,vegetables}              => {beef} 0.02989324 0.17541766
## [2] {vegetables}                            => {beef} 0.04585663 0.16797020
## [3] {bread and backed goods,dairy produce} => {beef} 0.02663955 0.14192849
## [4] {fruit}                                 => {beef} 0.03060498 0.12285714
## [5] {dairy produce}                         => {beef} 0.04921200 0.11108561
## [6] {bread and backed goods}                => {beef} 0.03721403 0.10771042
## [7] {}                                      => {beef} 0.08195221 0.08195221
## [8] {non-alc. drinks}                       => {beef} 0.02562278 0.08058842
##     lift      count
## [1] 2.1404872 294
## [2] 2.0496116 451
## [3] 1.7318446 262
## [4] 1.4991315 301
## [5] 1.3554925 484
## [6] 1.3143076 366
## [7] 1.0000000 806
## [8] 0.9833587 252
## Rules for beef
```




```
## set of 8 rules
```

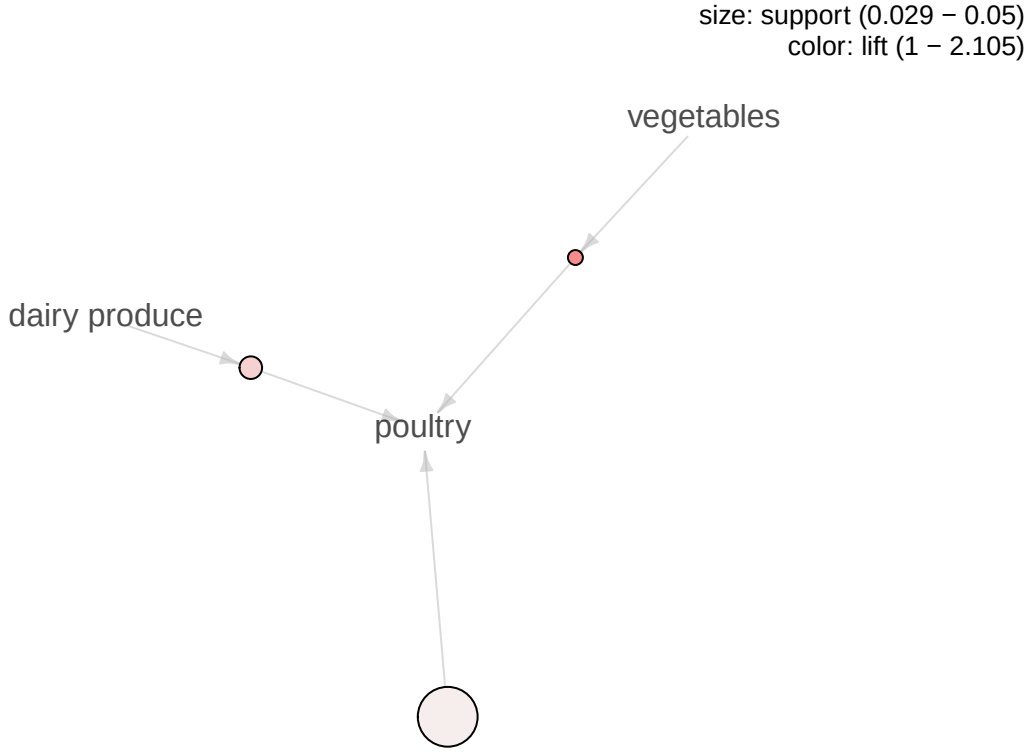
Top 3 Rules

- Customers who buy dairy produce and vegetables are 2.1 times as likely to buy beef as customers from the entire data set.
- Customers who buy vegetables are 2.0 times as likely to buy beef as customers from the entire data set.
- Customers who buy bread and backed goods and dairy produce are 1.7 times as likely to buy beef as customers from the entire data set.

To increase profits, I recomend to promote a dinner kit containing meat (beef) with vegetable mix.

```
plotRule(rules, "poultry", "market_basket_poultry_rules.pdf")
```

```
##     lhs                rhs       support    confidence lift     count
## [1] {vegetables}    => {poultry} 0.02897814 0.10614525 2.104715 285
## [2] {dairy produce} => {poultry} 0.03263854 0.07367455 1.460865 321
## [3] {}              => {poultry} 0.05043213 0.05043213 1.000000 496
## Rules for poultry
```




```
## set of 3 rules
```

- Customers who buy vegetables are 2.1 times as likely to buy poultry as customers from the entire data set.

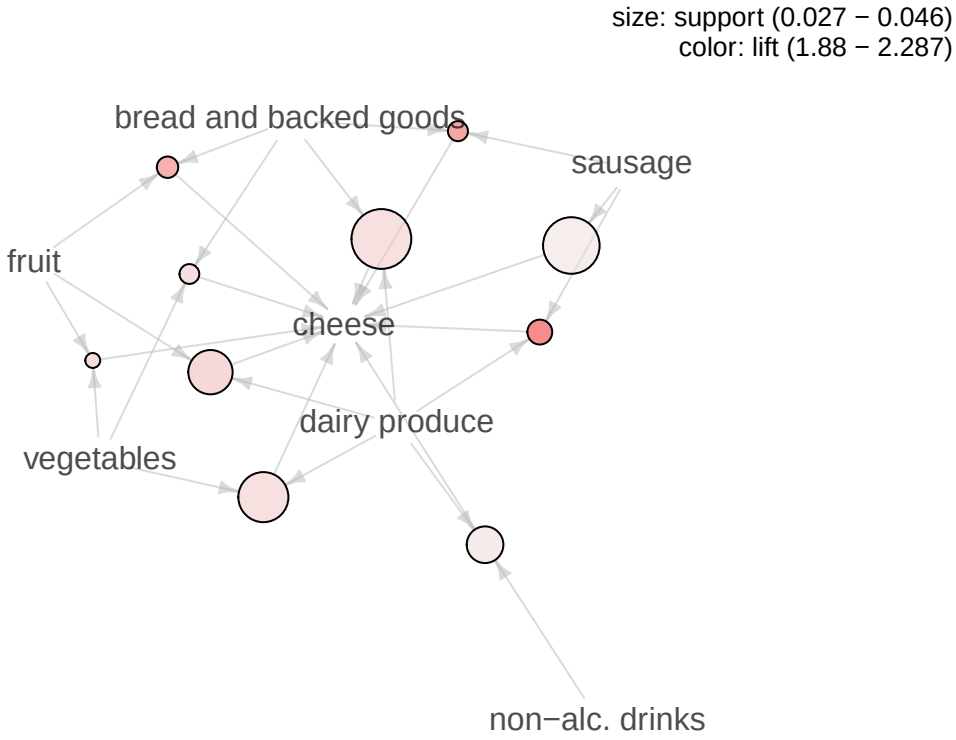
To increase profits, I recomend to promote a dinner kit containing meat (poultry) with vegetable mix.

Dairy producer

```
plotRule(rules, "cheese", "market_basket_cheese_rules.pdf")
```

```
##      lhs                                      rhs      support
## [1]  {dairy produce,sausage}                => {cheese} 0.03111337
## [2]  {bread and backed goods,sausage}       => {cheese} 0.02897814
## [3]  {bread and backed goods,fruit}         => {cheese} 0.02958821
## [4]  {dairy produce,fruit}                  => {cheese} 0.03965430
## [5]  {fruit,vegetables}                     => {cheese} 0.02674123
## [6]  {bread and backed goods,vegetables}    => {cheese} 0.02887646
## [7]  {dairy produce,vegetables}             => {cheese} 0.04219624
## [8]  {bread and backed goods,dairy produce} => {cheese} 0.04646670
## [9]  {dairy produce,non-alc. drinks}        => {cheese} 0.03629893
## [10] {sausage}                              => {cheese} 0.04504321
##      confidence lift     count
```

```
## [1]  0.2897727  2.287251 306
## [2]  0.2796860  2.207634 285
## [3]  0.2750473  2.171019 291
## [4]  0.2535761  2.001541 390
## [5]  0.2497626  1.971441 263
## [6]  0.2484689  1.961230 284
## [7]  0.2476134  1.954476 415
## [8]  0.2475623  1.954073 457
## [9]  0.2389558  1.886140 357
## [10] 0.2381720  1.879953 443
## Rules for cheese
```



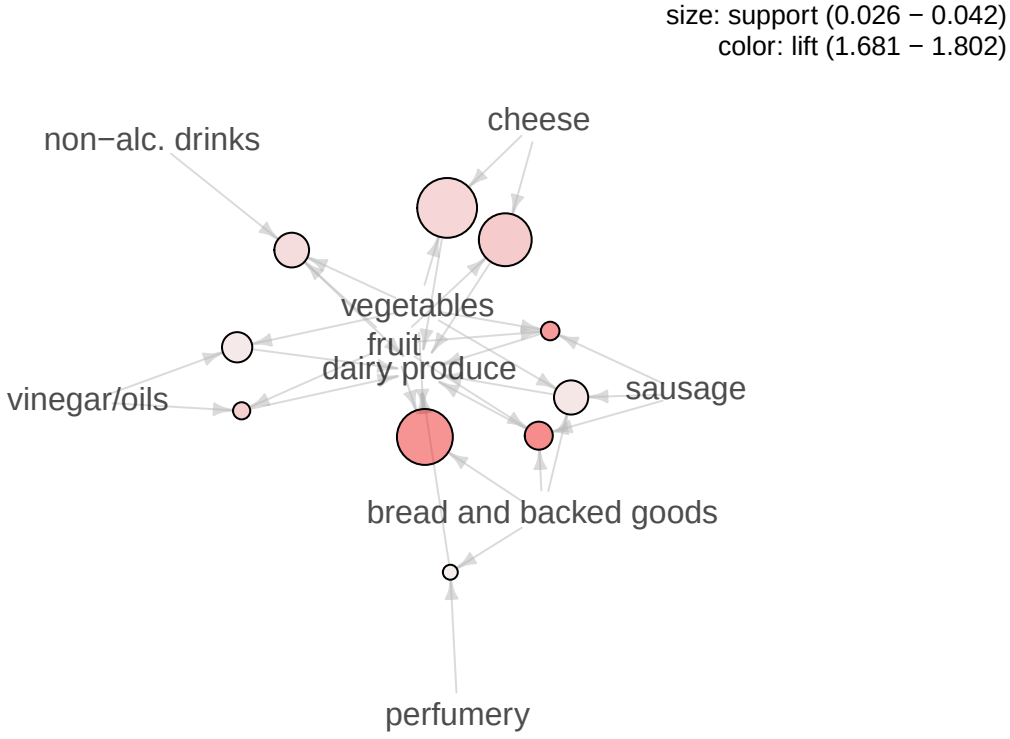

```
## set of 10 rules
```

Top 3 Rules

- Customers who buy dairy produce and sausage are 2.28 times as likely to buy cheese as customers from the entire data set.
- Customers who buy bread and backed goods, and sausage are 2.2 times as likely to buy cheese as customers from the entire data set.
- Customers who buy bread and backed goods, and fruit are 2.17 times as likely to buy cheese as customers from the entire data set.

```
plotRule(rules, "dairy produce", "market_basket_dairy_produce_rules.pdf")
```

```
##      lhs                          rhs                support confidence     lift count
## [1]  {bread and backed goods,
##       fruit,
##       sausage}                => {dairy produce} 0.03060498  0.7984085 1.802237   301
## [2]  {bread and backed goods,
##       fruit,
##       vegetables}             => {dairy produce} 0.04077275  0.7956349 1.795976   401
## [3]  {fruit,
##       sausage,
##       vegetables}             => {dairy produce} 0.02714794  0.7922849 1.788414   267
## [4]  {cheese,
##       fruit}                  => {dairy produce} 0.03965430  0.7707510 1.739806   390
## [5]  {fruit,
##       vinegar/oils}           => {dairy produce} 0.02663955  0.7683284 1.734338   262
## [6]  {cheese,
##       vegetables}             => {dairy produce} 0.04219624  0.7628676 1.722011   415
## [7]  {fruit,
##       non-alc. drinks,
##       vegetables}             => {dairy produce} 0.03304525  0.7575758 1.710066   325
## [8]  {bread and backed goods,
##       sausage,
##       vegetables}             => {dairy produce} 0.03284189  0.7494200 1.691656   323
## [9]  {vegetables,
##       vinegar/oils}           => {dairy produce} 0.03141840  0.7481840 1.688866   309
## [10] {bread and backed goods,
##       perfumery}              => {dairy produce} 0.02582613  0.7448680 1.681381   254
## Rules for dairy produce
```




```
## set of 10 rules
```

Top 3 Rules

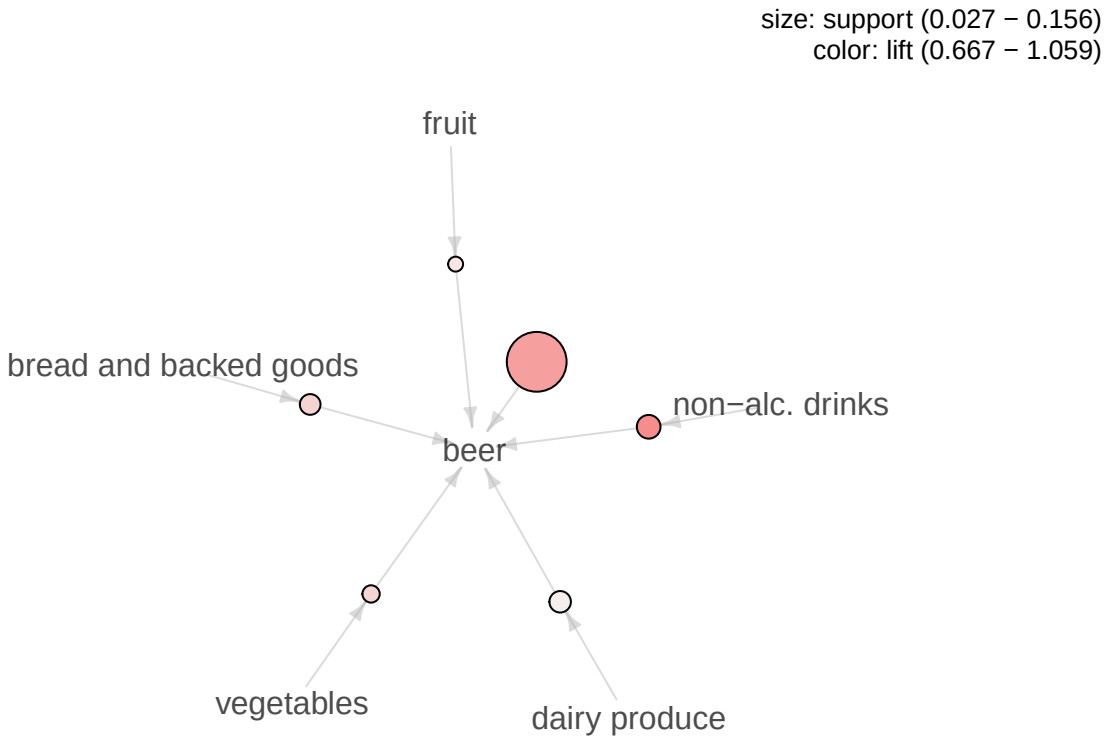
- Customers who buy bread and backed goods, fruit, and sausage are 1.8 times as likely to buy dairy produce as customers from the entire data set.
- Customers who buy bread and backed goods, fruit, and vegetable are 1.79 times as likely to buy dairy produce as customers from the entire data set.
- Customers who buy fruit, sausage, and vegetables are 1.78 times as likely to buy dairy produce as customers from the entire data set.

Beer

```
plotRule(rules, "beer", "market_basket_beer_rules.pdf")
```

```
##     lhs                          rhs    support    confidence lift
## [1] {non-alc. drinks}        => {beer} 0.05236401 0.1646946  1.0586741
## [2] {}                       => {beer} 0.15556685 0.1555669  1.0000000
## [3] {bread and backed goods} => {beer} 0.04372140 0.1265450  0.8134447
## [4] {vegetables}             => {beer} 0.03406202 0.1247672  0.8020168
## [5] {fruit}                  => {beer} 0.02724962 0.1093878  0.7031559
```

```
## [6] {dairy produce}          => {beer} 0.04595831 0.1037411  0.6668587
##     count
## [1]  515
## [2] 1530
## [3]  430
## [4]  335
## [5]  268
## [6]  452
## Rules for beer
```




```
## set of 6 rules
```

In general, association rules with lift values close to 1 will are not interesting nor useful than rules with lift values higher than 1.

Recomendation