DRIVING SMART BASKET SIZE

Customer Analytics at Bigbasket

Nicole Hong

India's largest online supermarket

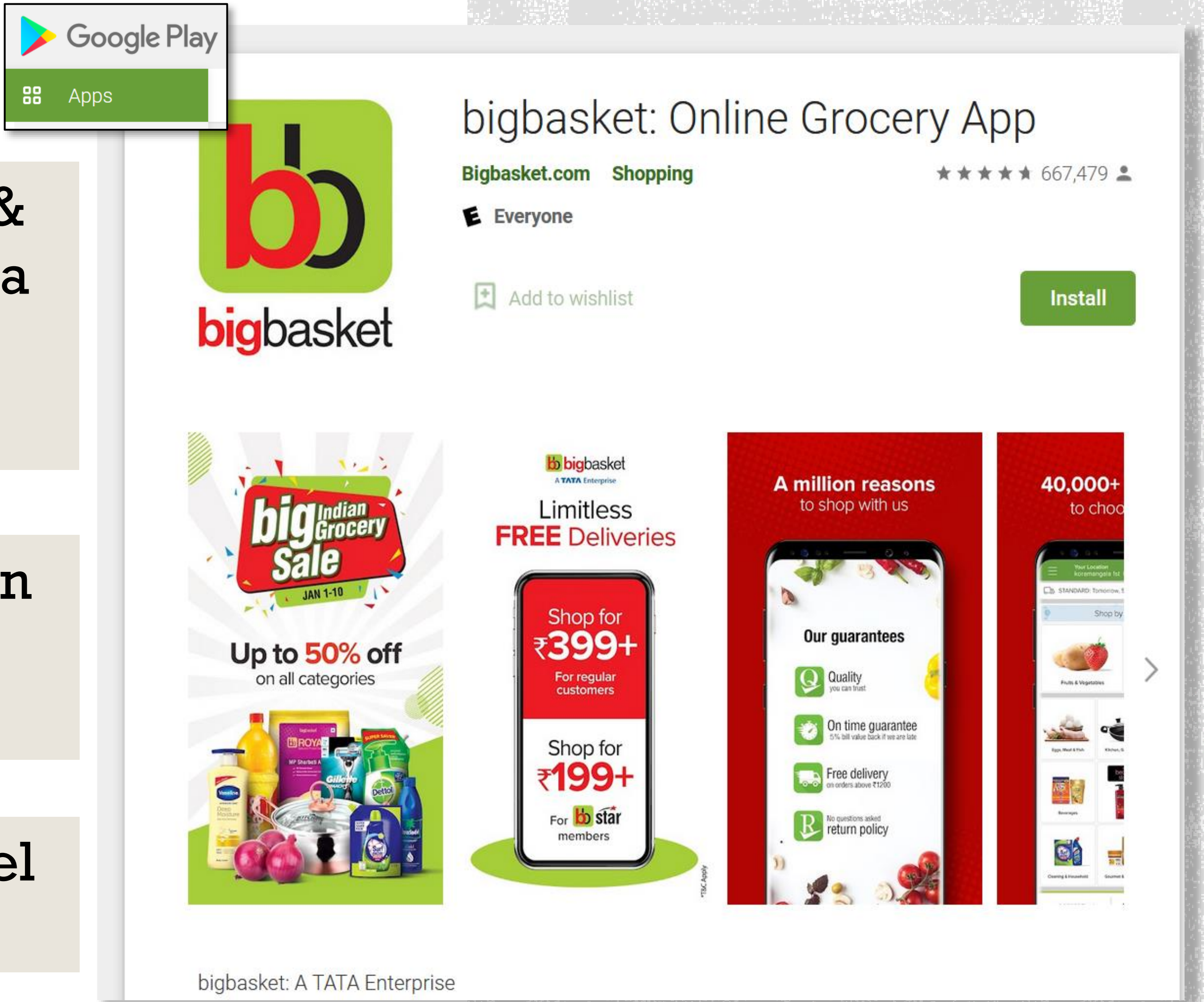
bigbasket
A TATA Enterprise

- INDIA'S LARGEST ONLINE GROCERY/FOOD STORE

- In 2021, selling 20,000+ products & 1,000+ brands in 25+ cities in India
- 100+ million customers, targeting tech-savvy, young generation

- Online grocery delivery service on home essentials, convenience stores, & food supplies to its users

- Revenue based on Inventory model & Hyper-local delivery



RECOMMENDATION

Business Issues

- Poor customer experience
- High operational costs
- Reduced profit

Analytics Approach

Features:

- Your Top 20
- Smart Basket
- "Did You Forget"

Improve the customer experience:

- Increase average basket size from 6 to 10 items, with expected sales lift of 9%[1]
- Increase average order growth rate from 30% to 50% every month
- Shorten shop to checkout time from 30 minutes to 10 minutes
- Lower delivery costs by reducing from 4 to 2 deliveries per day to the same location, increasing the efficiency by 50%.

1. https://news.ufl.edu/articles/2018/09/how-helpful-are-product-recommendations-really.html

2. For detailed estimations, see Slide 21 in Appendix.

MARKET BASKET ANALYSIS IN R

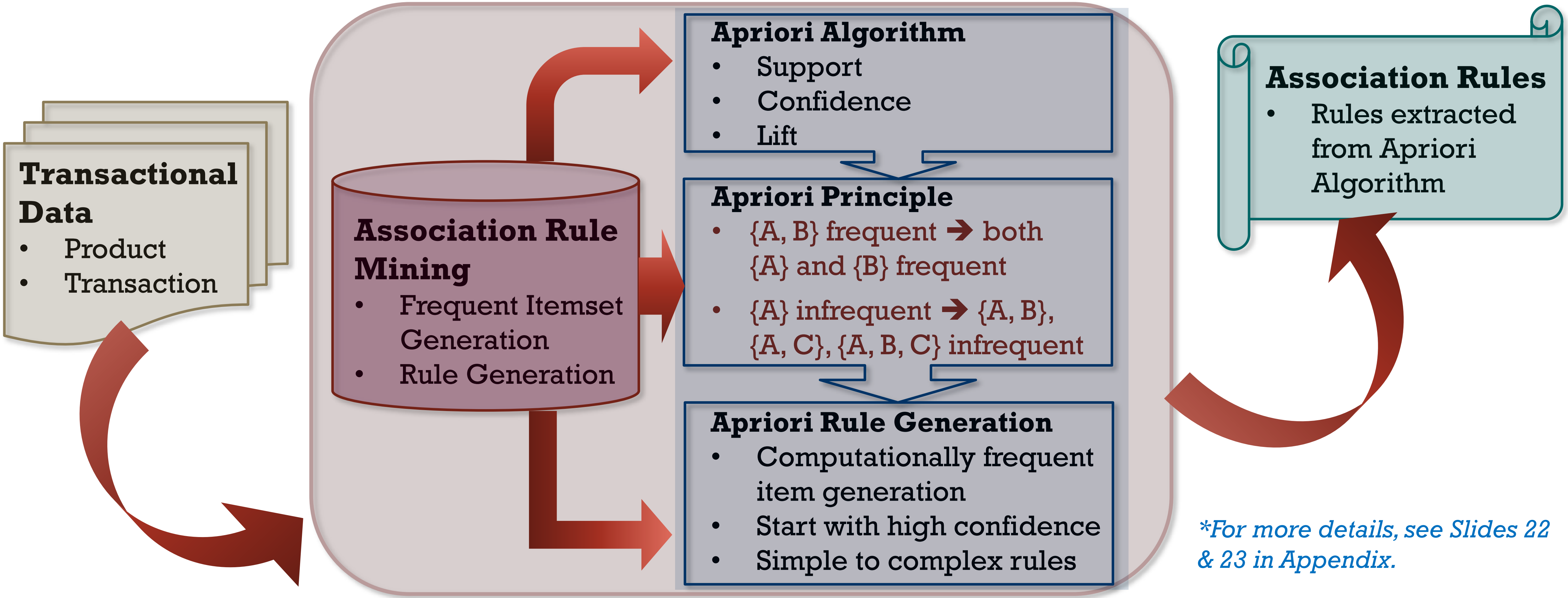
Transform Data → Item Frequency → Extract Rules → Association Rules → Recommenders

Problems → **Solutions**

Transform Data	Item Frequency	Extract Rules	Association Rules	Recommenders
Raw Dataset Data Preprocessing Conversion to Transactional Data	Product Variable (SKU & Description): ------------------------------ • Absolute Item Frequency Plot • Relative Item Frequency Plot	Density Plot Apriori Algorithm: ------------------------------ • Set the following levels: - Support - Confidence - Lift	• 194 non-redundant rules • Top 20 Rules with the highest lift • Observe the rules with the visualizations	• Top 20 SKU • Smart Basket • “Did you forget?” Engine

ASSOCIATION RULES

The rules extracted based in the inference on frequently co-occurring associations among different items, which are determined by market basket analysis (i.e. association rule mining)



**For more details, see Slides 22 & 23 in Appendix.*

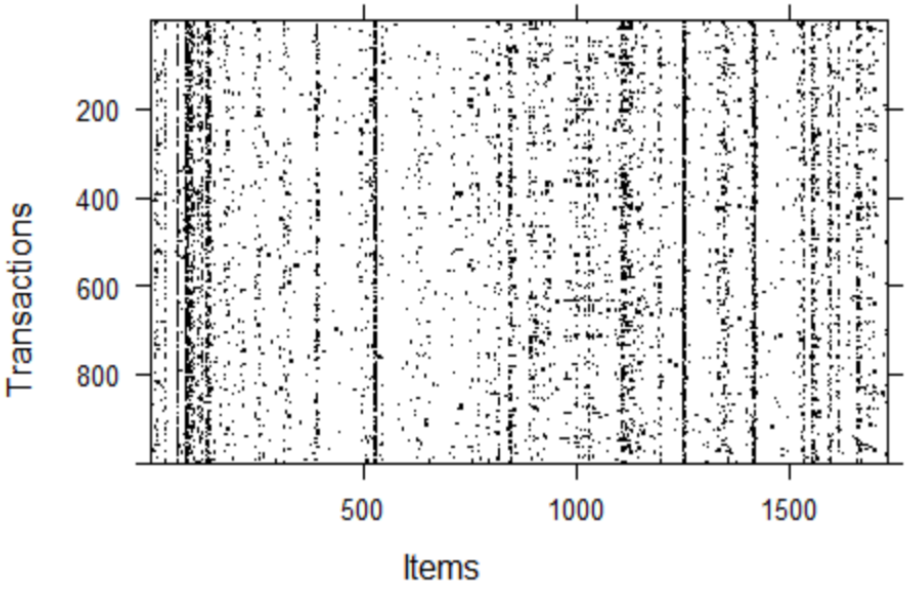
RECOMMENDER #1

"Smart Basket"

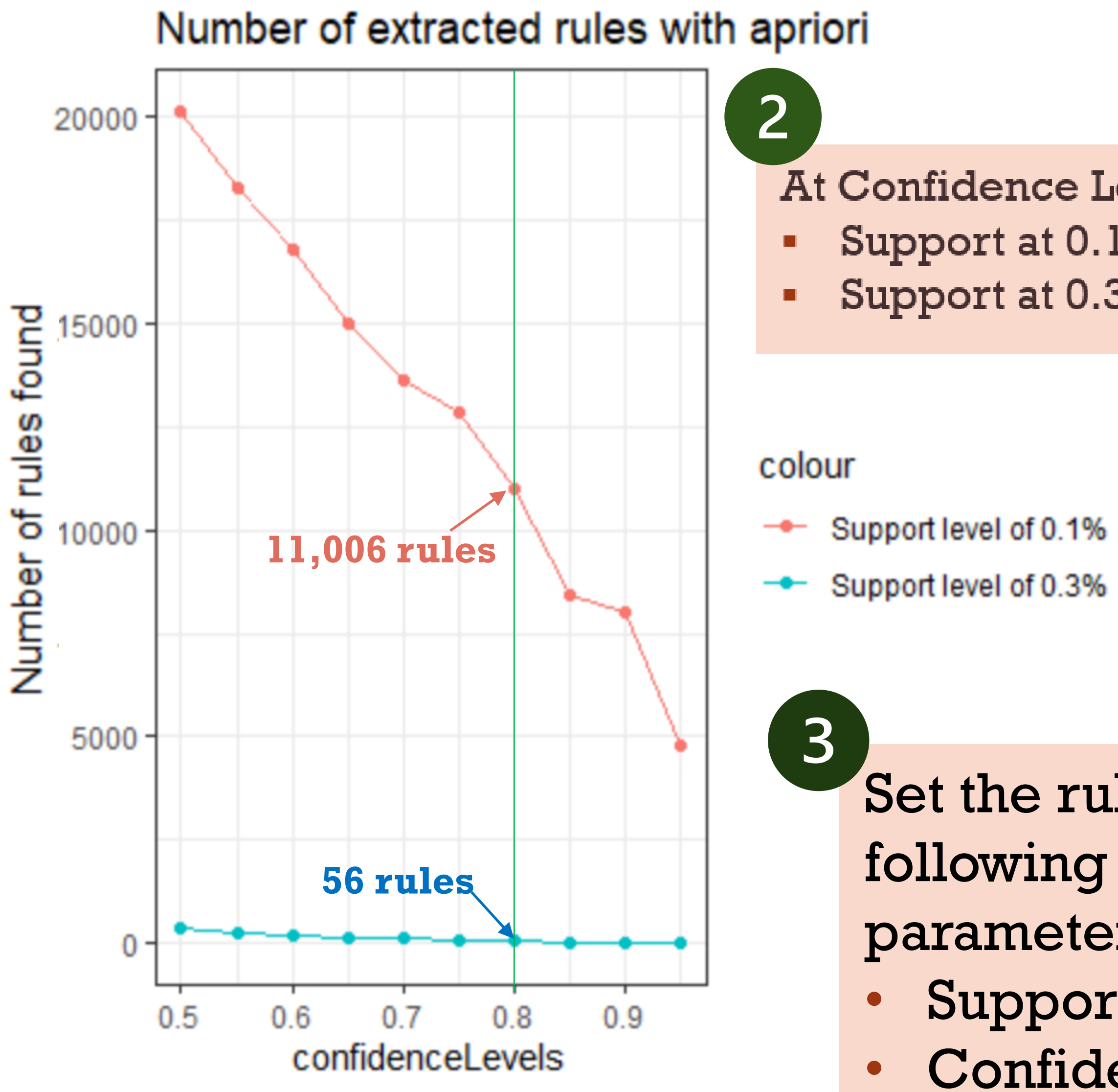
- **Association Rules** for the recommender feature
- Out of 248 rules, removed the redundancy to extract the finalized 194 rules
- Order the items under the RHS and LHS based on the highest Lift
- Demo 5 recommender rules to test "Smart Basket" feature

EXTRACTING THE RULES

1 Density of 0.43%



**For more details, see Slide 26 in Appendix.*



2 At Confidence Level of 80%:

- Support at 0.1% = 11,006 rules
- Support at 0.3% = 56 rules

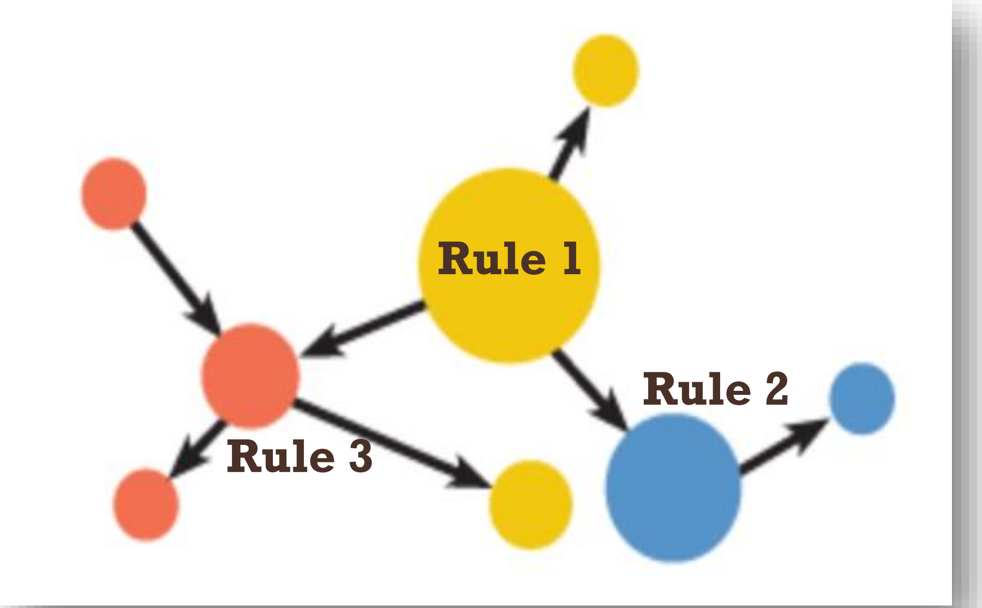
3 Set the rule with the following minimum parameters:

- Support = 0.25%
- Confidence = 80%

"SMART BASKET" – ASSOCIATION RULES

Shopping Basket (LHS)

Basket Recommended (RHS)



Rules	LHS	RHS
Rule 1	Beans_Haricot	Brinjals_Eggplant
	Exotic Vegetables_Broccoli	
	Other Vegetables_Onion Flower	
	Other Vegetables_Bottle of Pickle #2	
	Root Vegetables_Beet Root	

Rule 1	
Support	0.25%
Confidence	84%
Lift	83.86
Count	21

For more details, see Slides 34, 35, 36, 37, 38 & 39 in Appendix.

RECOMMENDER #2

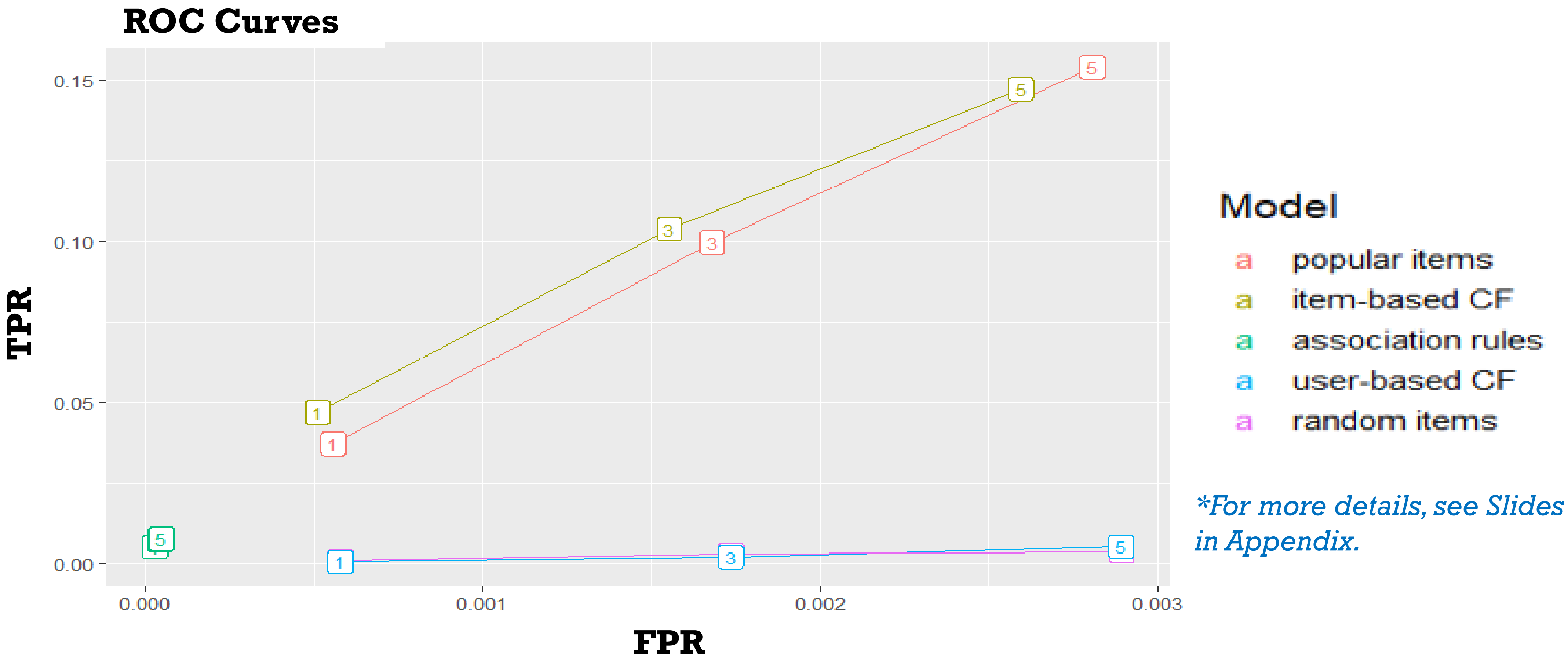
"Did you forget"

- Recommenderlab to evaluate 5 recommenders using cross validation: AR, **IBCF**, UBCF, Popular & Random

- Select recommender with best performance per ROC and TPR/FPR plots

- Demo 3 shopping scenarios to test recommender suggestions

ROC CURVES TO EVALUATE "DID YOU FORGET" RECOMMENDERS



**For more details, see Slides 24 & 25 in Appendix.*

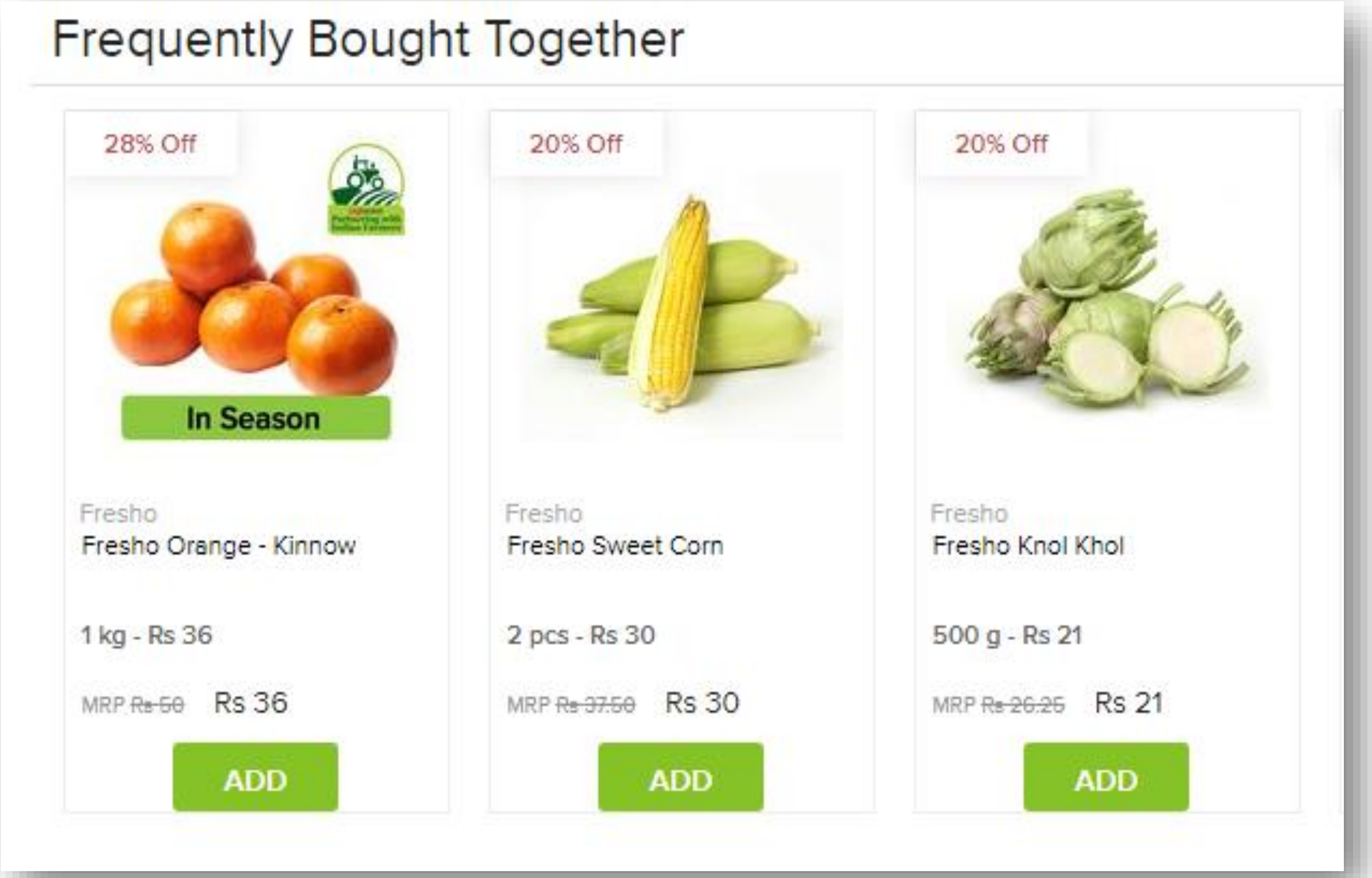
SCENARIO ANALYSIS

"Smart Basket"

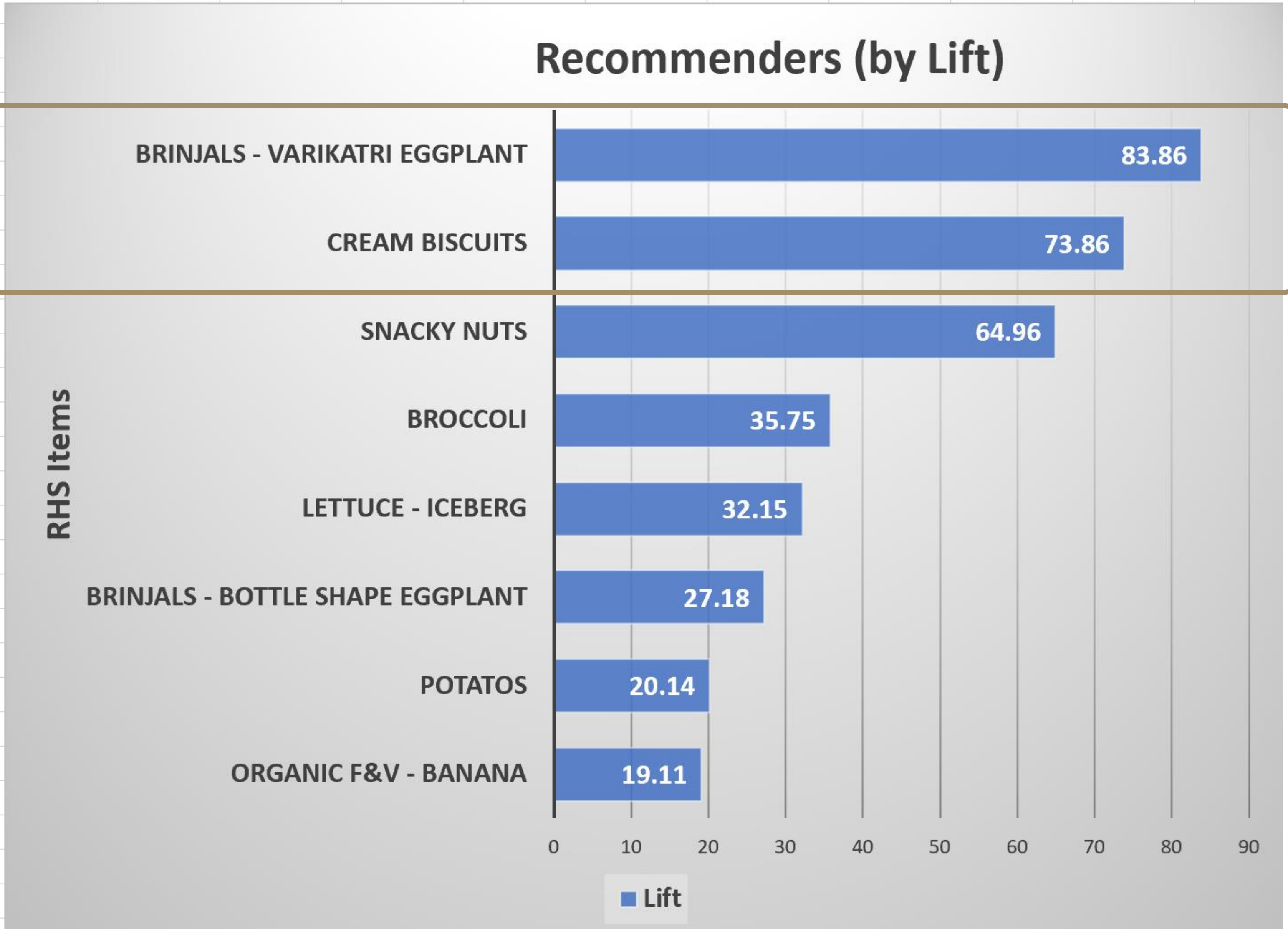
"Did You Forget?“ Engine

"SMART BASKET"

- Based on the highest Lift level, identified the obtain Top 5 association rules
- Recommendation notification popup

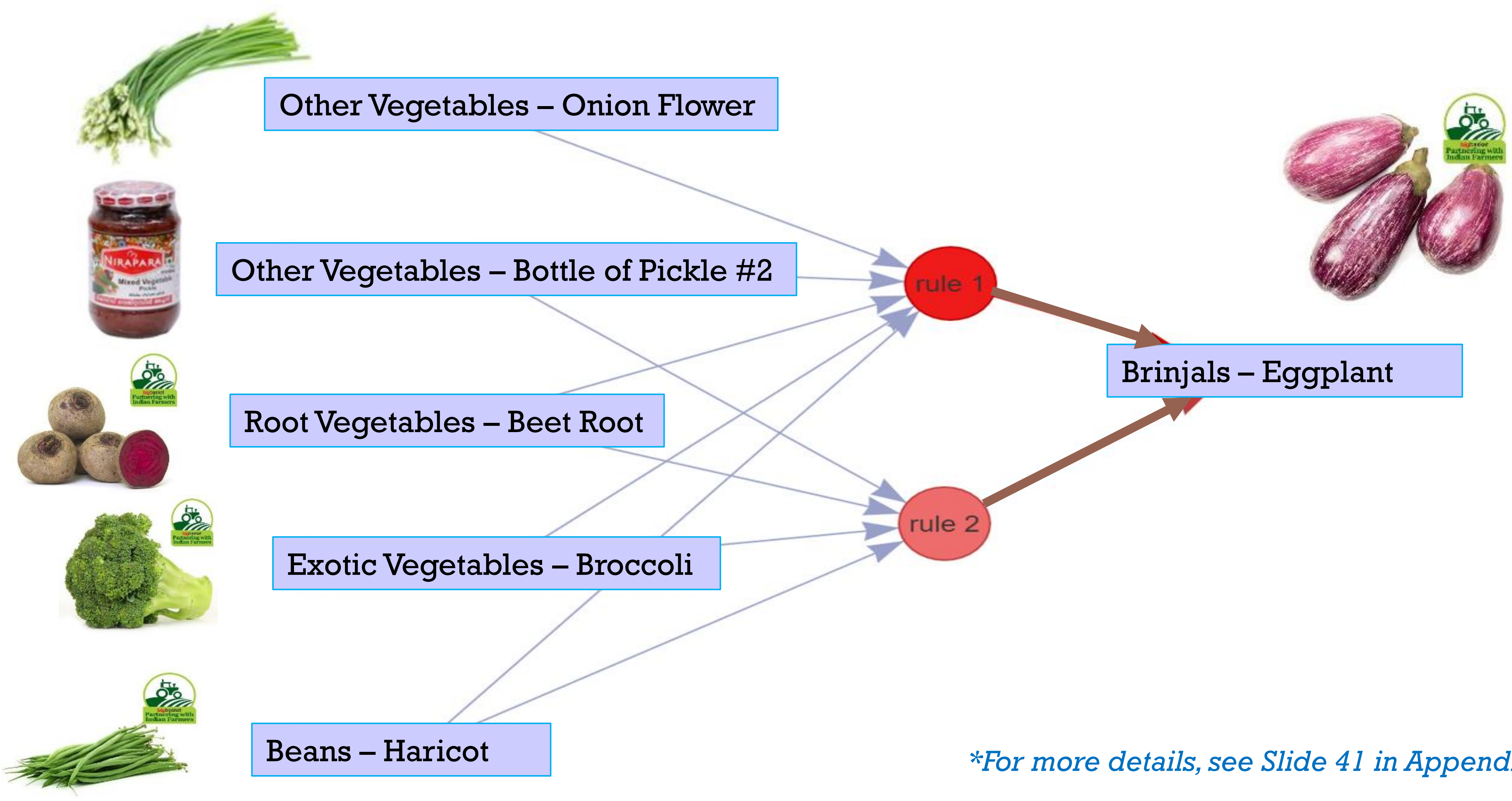


"SMART BASKET" - TOP RECOMMENDERS



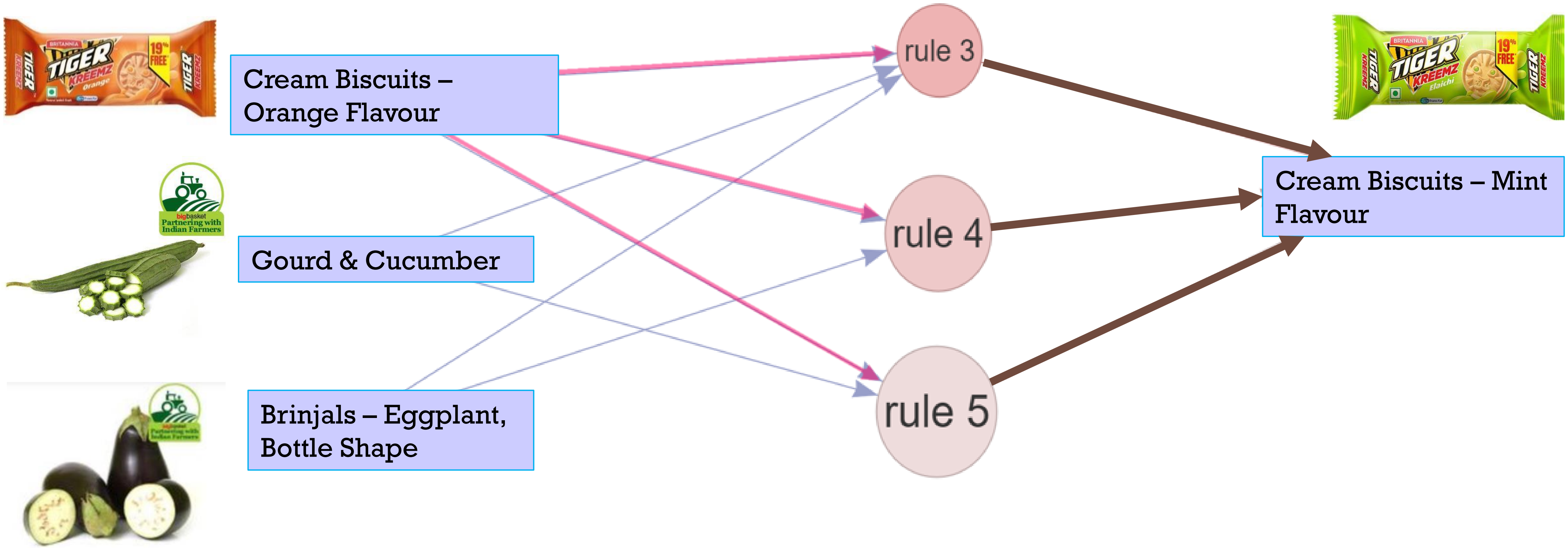
**For more details, see Slide 40 in Appendix.*

PEOPLE ARE ALSO BUYING…



For more details, see Slide 41 in Appendix.

CROSS-SELLING OPPORTUNITY



For more details, see Slide 43 in Appendix.

"DID YOU FORGET?" RECOMMENDER

- Suggests top 5 items that may have been forgotten, based on basket content at check out

"DID YOU FORGET?" DEMO AT CHECK-OUT

Scenario: Customer is shopping for ingredients for a Masala recipe

Organic Masalas & Spices

Other Sauces

**For more details, see Slide 45 in Appendix.*

Top 5 "Did you forget?" suggested items

Garam Sauce

Cashews

Rice Products

Turmeric Masalas & Spices

Organic Salt &Sugar & Jaggery

NEW CUSTOMER ORDER FLOW

1. Start with Your Top 20?
2. Yes! User adds 6+ items to their basket
3. User manually adds other items to basket
4. Smart Basket recommends items as they shop
5. “Did You Forget” raises 5 potential items that the user may have forgotten
6. Add some of these from “Did You Forget” list

Thank You!

Nicole Hong
hnicole01@yahoo.ca

APPENDIX

CONCLUSION – ESTIMATIONS BASED ON ASSUMPTIONS

1 Estimated Costs:
- Analysis & Model dev: 100 hours @ $100/hr = $10,000
- Deployment: 50 hours @ $100/hr = $5,000

- Total estimated cost = $15,000

2 Average order growth rate:
- Increase average basket size from 6 to 10 items
- % Increase: (10 items – 6 items) / 6 items = 67%
- 30% growth x (1 + 0.67) = 50% growth every month

3 Delivery efficiency rate:
- Lower delivery costs by reducing from 4 to 2 deliveries per day
- % Decrease: (4 deliveries – 2 deliveries) / 4 deliveries = 50%

Note:

Due to the lack of data factual to the case, our estimations required a number of assumptions to validate the efficiency rate (%) such as the number of items increased in average basket and the number of deliveries made per day.

ASSOCIATION RULE MINING

To discover relationships between items in a large transactional database

Divided into two subtasks:

- **Frequent itemset generation**: *determine all frequent itemsets of a potentially large database of transactions, satisfying a minimum support threshold*
- **Rule generation**: *from the frequent itemset generation above, generate association rules with confidence above a minimum confidence threshold*

Apriori Algorithm:

- Belong to the class of the association rule mining algorithms
- Using the bottom-up approach, generate candidate itemsets through the apriori principle

Apriori Principle:

- For a frequent itemset, all of its subsets must also be frequent.
  e.g. If {A, B} is frequent, then both {A} and {B} are frequent.
- For an infrequent itemset, all its super-sets are infrequent.
  e.g. If {A} is infrequent, then {A, B}, {A, C} and {A, B, C} are infrequent.

Apriori Rule Generation → Association Rules

- After computationally expensive frequent item generation, apriori generates rules
- Start with high-confidence rules with single precedent: {A, C} → {B}
- Build more complex rules, with more items on the right-hand side: {A, C} → {B, D}

METRICS IN ASSOCIATION RULE MINING

Elements of the Apriori Function

Support

- Popularity of an itemset

Support (A) = Fraction of transactions that contain itemset A

Support (A∪B) = Fraction of transactions with both A and B

Confidence

- How often the rule is true
- Shows % in which B is bought with A

Confidence(A->B) = P(A∩B) / P(A)
= frequency(A,B) / frequency(A)

Lift

- How strong is the association
- Lift > 1: B is likely to be bought with A
- Lift <1: B is unlikely to be bought if A is bought

Lift(A->B) =
*Support (A∪B) / (Supp(A) * Supp(B))*

TID	Transaction
1	{**Bread, Butter**, Cheese, Wine}
2	{**Bread, Butter**, Wine}
3	{**Bread, Butter**}
4	{Butter, Cheese, Wine}
5	{Butter, Cheese}
6	{Cheese, Wine}
7	{Butter, Wine}

Support Measure:
- supp({Bread}) = 3/7 = 42%
- supp({Bread} U {Butter}) = 3/7 = 42%

Confidence Measure:

A = {Bread} B = {Butter}

conf (A -> B) = (3/7) / (3/7) = 100%

Lift Measure:

A = {Bread} B = {Butter}

lift (A -> B) = $\frac{3/7}{(3/7)*(6/7)} = \frac{7}{6}$ ~ 1.16

ROC PERFORMANCE METRICS WITH TPR/FPR PLOTS

- AUC - ROC curve: performance measurement for the classification problems at various threshold settings

  ➔ ROC: probability curve

  ➔ AUC: the degree or measure of separability that tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model

- TPR = True Positive Rate

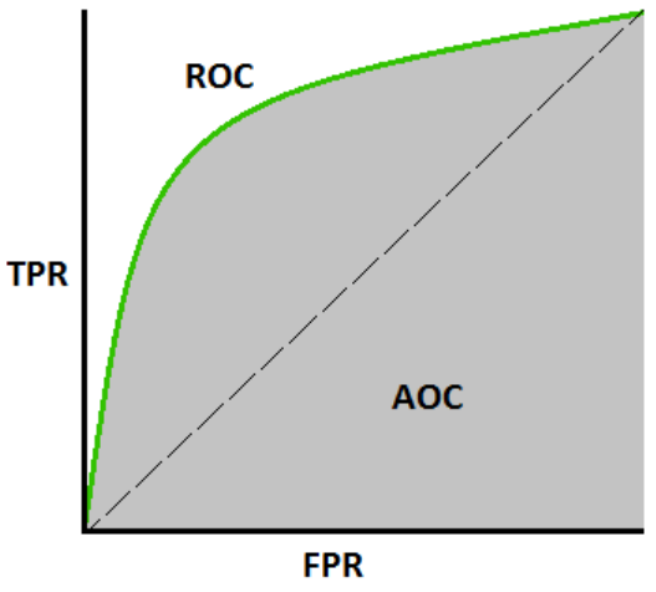
$$\text{TPR /Recall / Sensitivity} = \frac{TP}{TP + FN}$$

- FPR = False Positive Rate

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{FPR} = 1 - \text{Specificity} = \frac{FP}{TN + FP}$$

- Sensitivity and Specificity are inversely proportional to each other – when Sensitivity increases, Specificity decreases, and vice versa.
- Similarly, when the threshold increases, more negative values, thus higher specificity and lower sensitivity.
- As FPR = 1 – specificity, when TPR increases, FPR also increases and vice versa.



The ROC curve is plotted with TPR against FPR, where TPR is on the y-axis and FPR is on the x-axis.

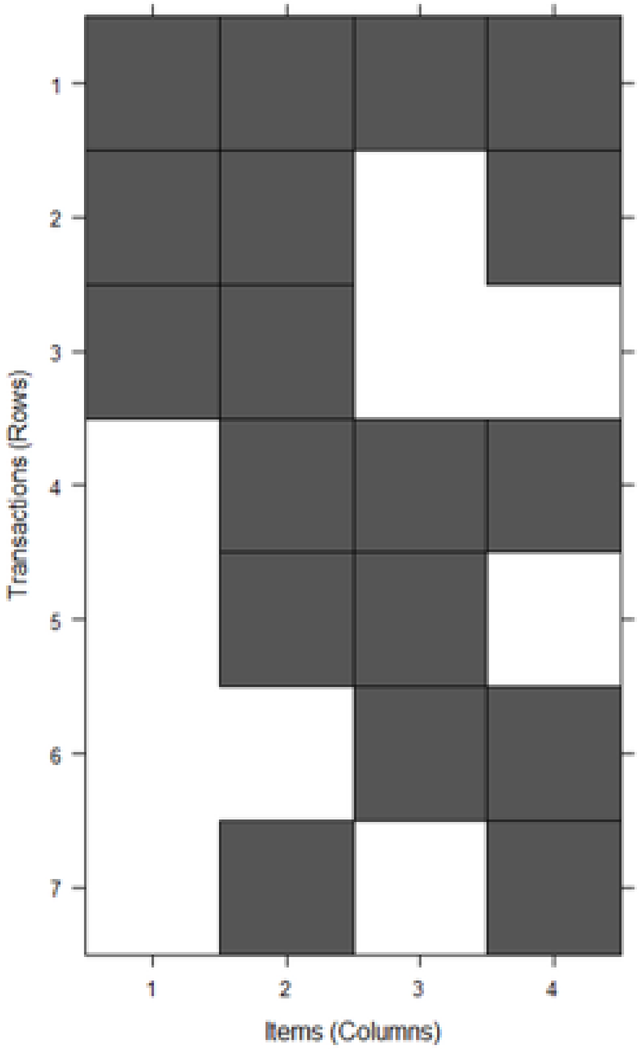
"DID YOU FORGET" – 5 RECOMMENDER ALGORITHMS

- **Item-based collaborative filtering:** a type of recommendation system that is based on the similarity between items calculated using the rating users have given to items. It helps solve issues that user-based collaborative filters suffer from such as when the system has many items with fewer items rated.
- **Popularity based filtering (Popular items):** as the name suggests, this system works with the trend. It basically uses the items which are in trend right now. For example, if any product which is usually bought by every new user then there are chances that it may suggest that item to the user who just signed up.
- **User-based collaborative filtering:** a technique used to predict the items that a user might like on the basis of ratings given to that item by the other users who have similar taste with that of the target user. Many websites use collaborative filtering for building their recommendation system.
- **Randomizable filter classifier:** This method employed an arbitrary classifier on data that has been passed through an arbitrary filter. Similar to the classifier, the structure of the filter worked exclusively on the training data and test instances will be processed by the filter without altering their structure. In using randomizable filter (RF) as an ensemble base classifier, each base classifier is built using a different random number seed (but based on the same data). The final prediction is the average of the predictions generated by the individual base classifiers.

Model

a item-based CF
a popular items
a association rules
a user-based CF
a random items

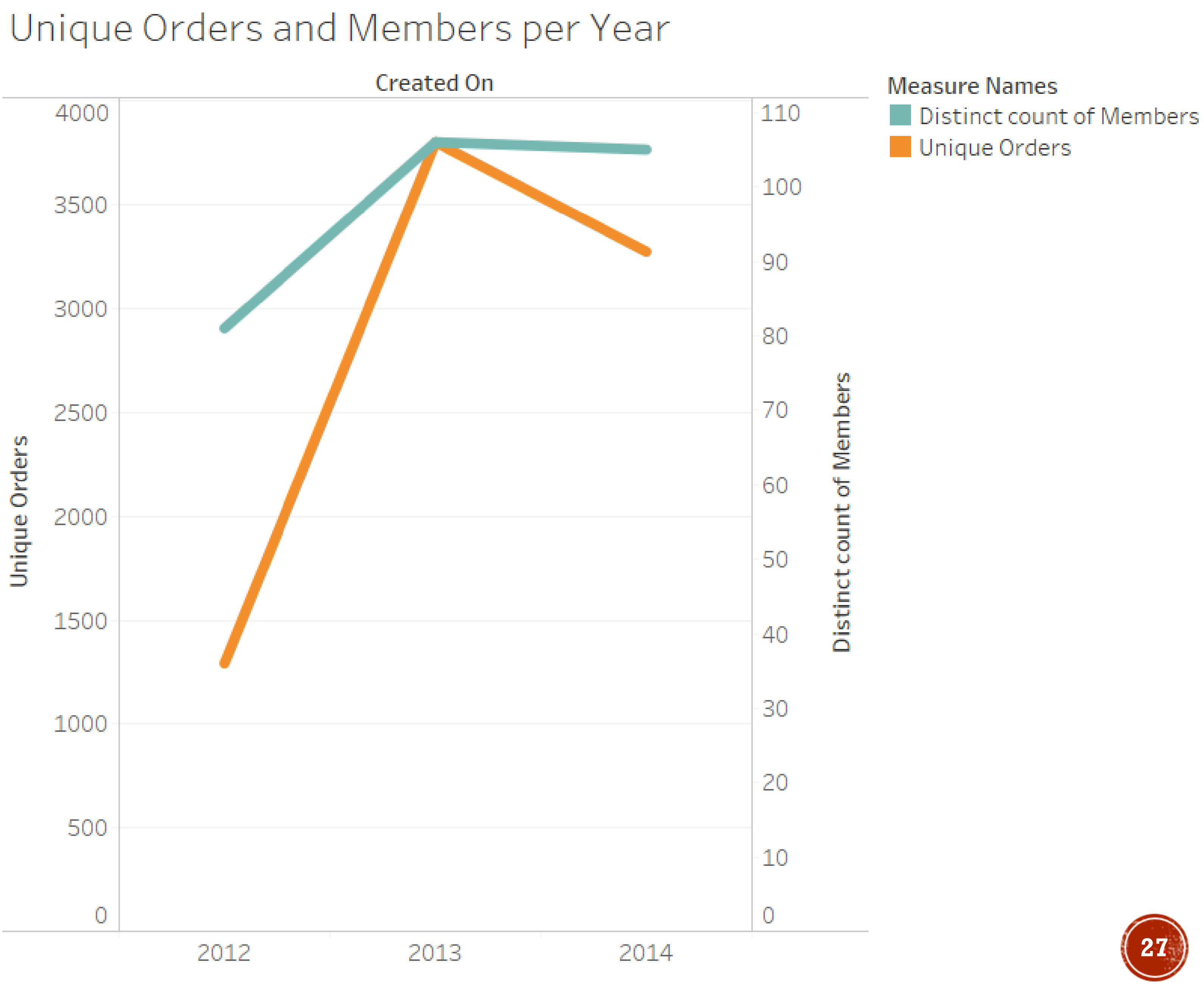
ITEM-MATRIX & DENSITY PLOT



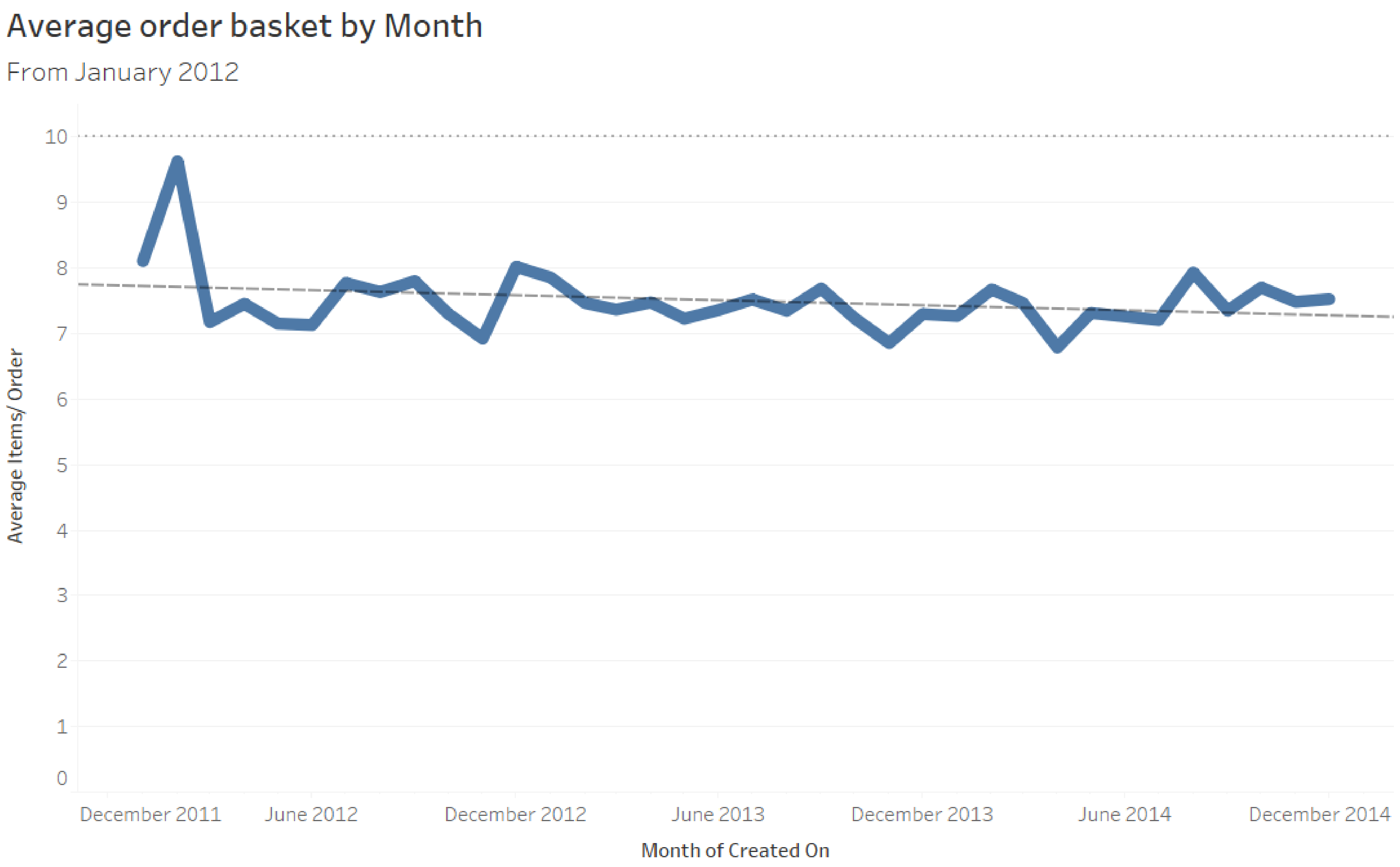
- **Density:** measures % of non-zero cells in a sparse matrix. It is the total number of items that are purchased divided by the possible number of items in that matrix
- **Density Map:** plot of ItemMatrix
- Useful to identify:
  ➔ Patterns in the transactions
  ➔ Sparsity in the data

Columns = 4; Rows = 7
Total 28 items in ItemMatrix
Total 18 non-zero cells
Thus, Density = 18 / 28 = 0.64

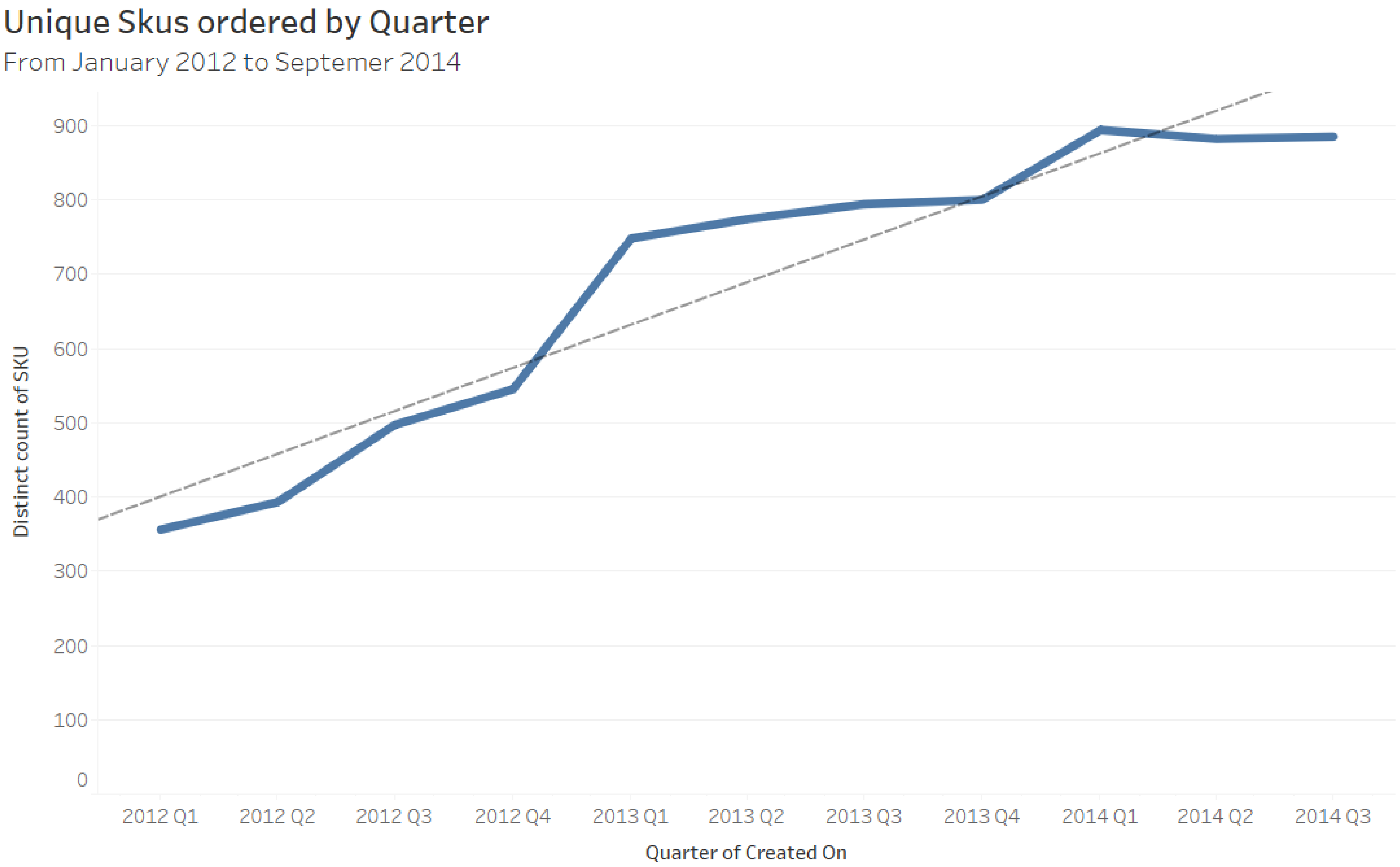
MEMBER AND ORDER GROWTH HAS STALLED



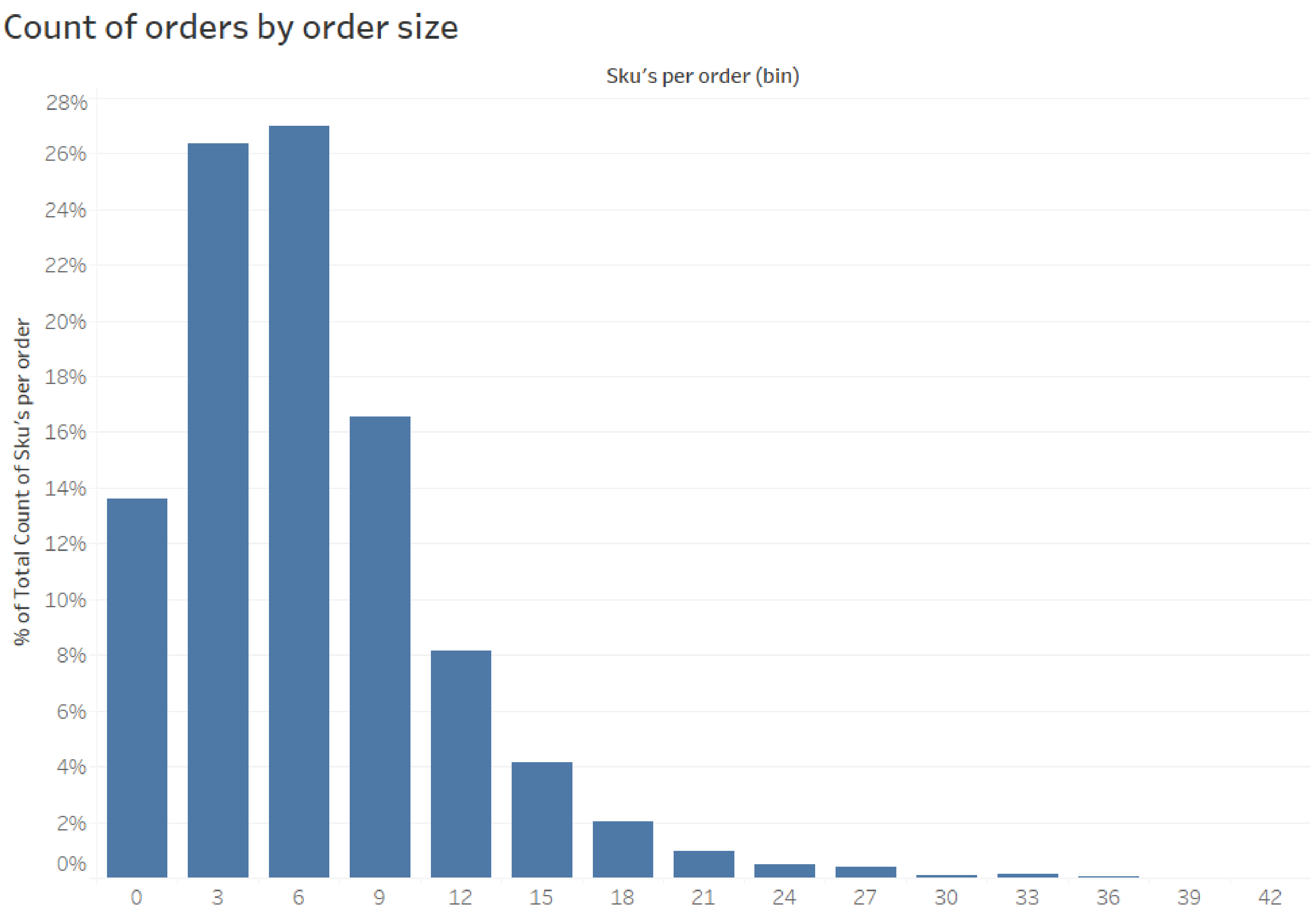
AVERAGE BASKET SIZE IS FALLING



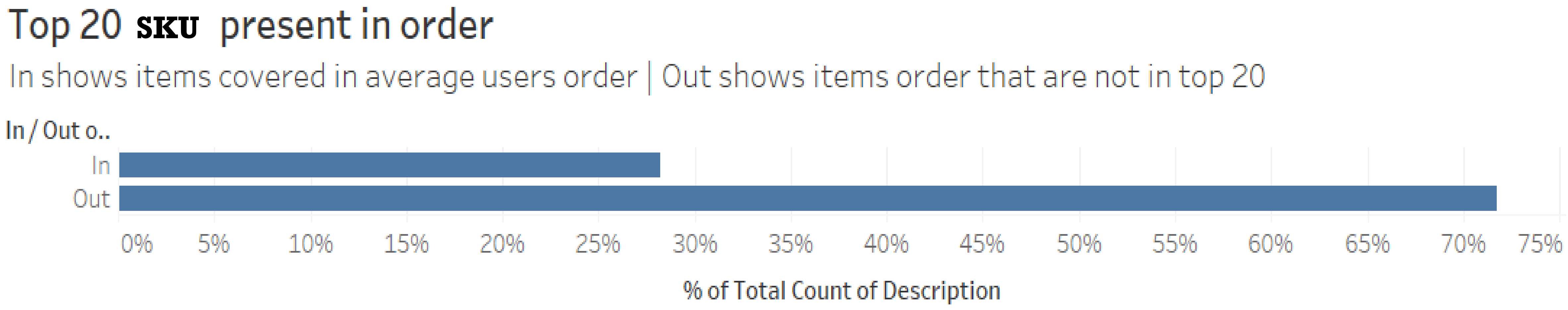
PRODUCT SKUS ARE GROWING



SMALL BASKETS ARE THE NORM



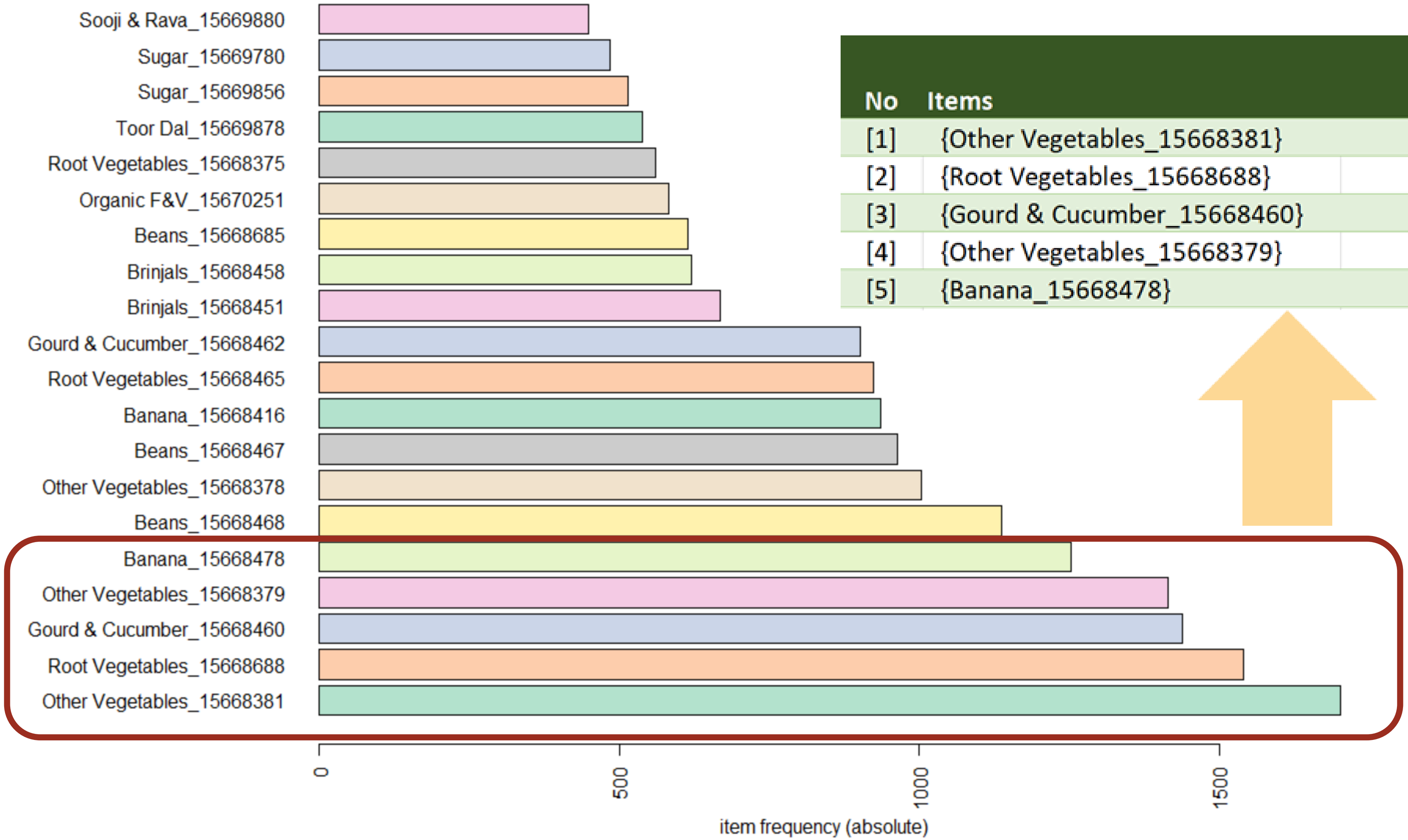
START WITH YOUR TOP 20?



START WITH YOUR TOP 20? IMPACT

- Shows users products that they often buy
- Drive ease in purchasing their common items quickly
- Increase minimum basket size to 6
- Provides a basket to build recommendations from

POPULAR PRODUCTS



Absolute Item Frequency Plot

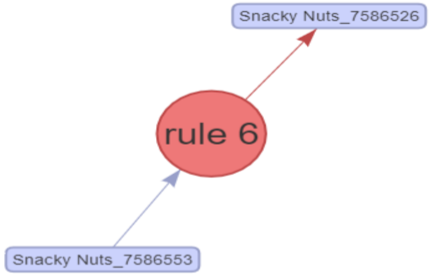
No	Items	Support	Trans Identical To Itemsets	Count
[1]	{Other Vegetables_15668381}	0.20	0.02	1,702
[2]	{Root Vegetables_15668688}	0.18	0.03	1,540
[3]	{Gourd & Cucumber_15668460}	0.17	0.02	1,439
[4]	{Other Vegetables_15668379}	0.17	0.03	1,415
[5]	{Banana_15668478}	0.15	0.04	1,252

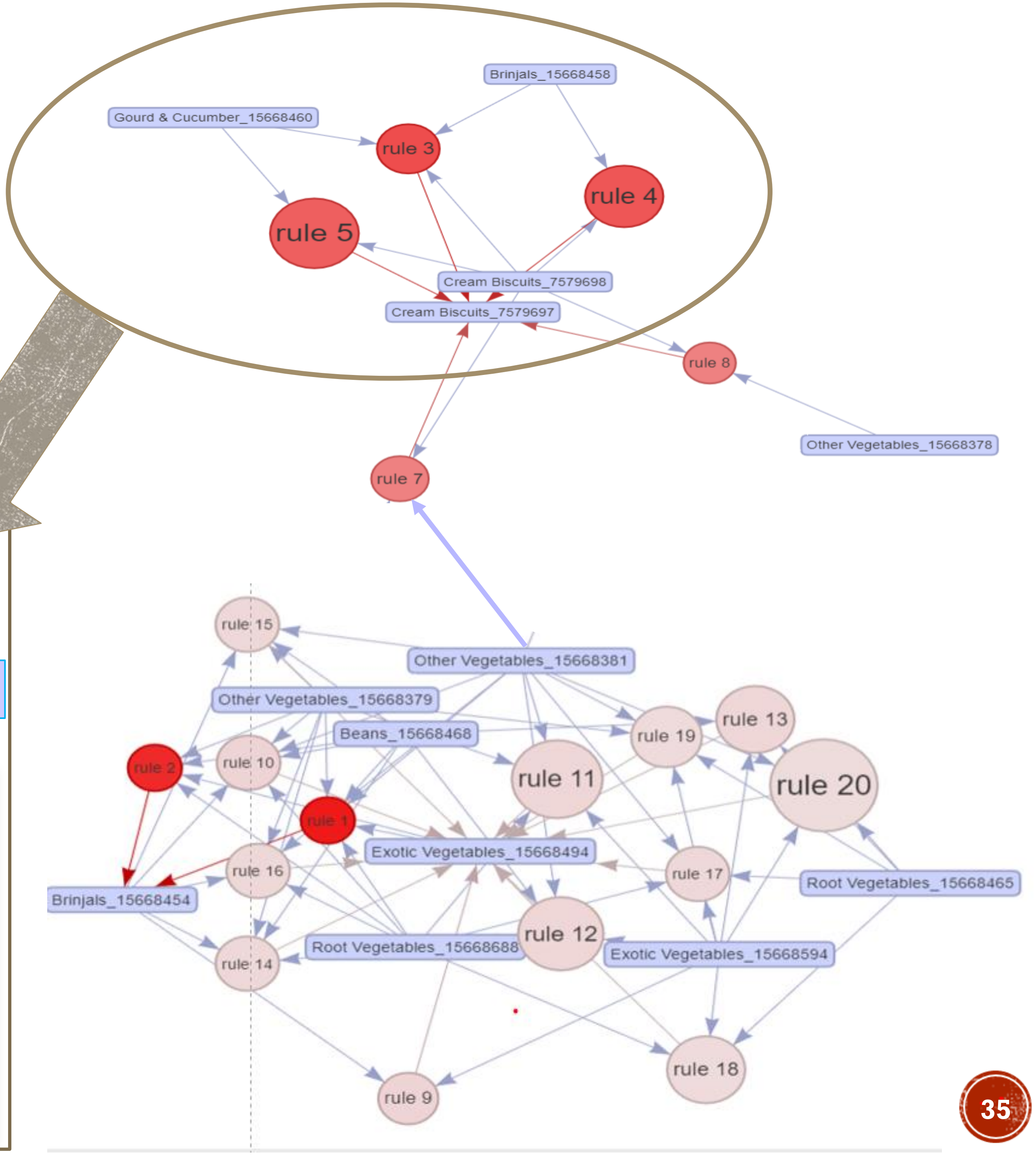
"SMART BASKET" – ASSOCIATION RULES IN TABLE

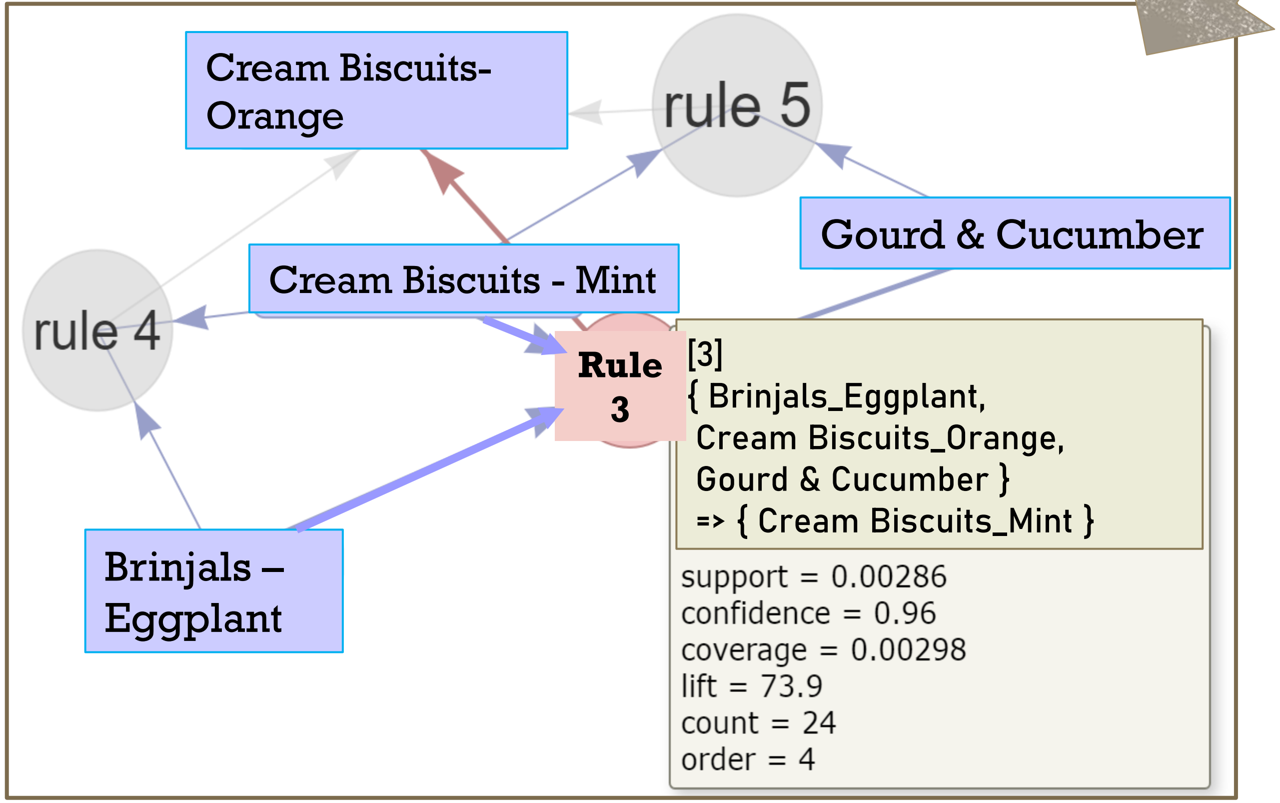
Rules	LHS	RHS	Support	Confidence	Lift	Count	Coverage
[1]	Beans_15668468						
[1]	Exotic Vegetables_15668494						
[1]	Other Vegetables_15668379						
[1]	Other Vegetables_15668381						
[1]	Root Vegetables_15668688	Brinjals_15668454	0.25%	0.84	83.86	21	0.30%
[2]	Beans_15668468						
[2]	Exotic Vegetables_15668494						
[2]	Other Vegetables_15668379						
[2]	Root Vegetables_15668688	Brinjals_15668454	0.25%	0.81	80.63	21	0.31%
[3]	Brinjals_15668458						
[3]	Cream Biscuits_7579698						
[3]	Gourd & Cucumber_15668460	Cream Biscuits_7579697	0.29%	0.96	73.86	24	0.30%
[4]	Brinjals_15668458						
[4]	Cream Biscuits_7579698	Cream Biscuits_7579697	0.32%	0.93	71.63	27	0.35%
[5]	Cream Biscuits_7579698						
[5]	Gourd & Cucumber_15668460	Cream Biscuits_7579697	0.35%	0.91	69.72	29	0.38%
[6]	Snacky Nuts_7586553	Snacky Nuts_7586526	0.35%	0.91	64.96	29	0.38%
[7]	Cream Biscuits_7579698						
[7]	Other Vegetables_15668381	Cream Biscuits_7579697	0.27%	0.82	63.20	23	0.33%
[8]	Cream Biscuits_7579698						
[8]	Other Vegetables_15668378	Cream Biscuits_7579697	0.26%	0.81	62.69	22	0.32%
[9]	Brinjals_15668454						
[9]	Exotic Vegetables_15668594	Exotic Vegetables_15668494	0.26%	0.92	35.75	22	0.29%
[10]	Beans_15668468						
[10]	Brinjals_15668454						
[10]	Other Vegetables_15668379						
[10]	Other Vegetables_15668381						
[10]	Root Vegetables_15668688	Exotic Vegetables_15668494	0.25%	0.91	35.61	21	0.27%

Inspecting 10 rules out of 194 rules with the highest Lift

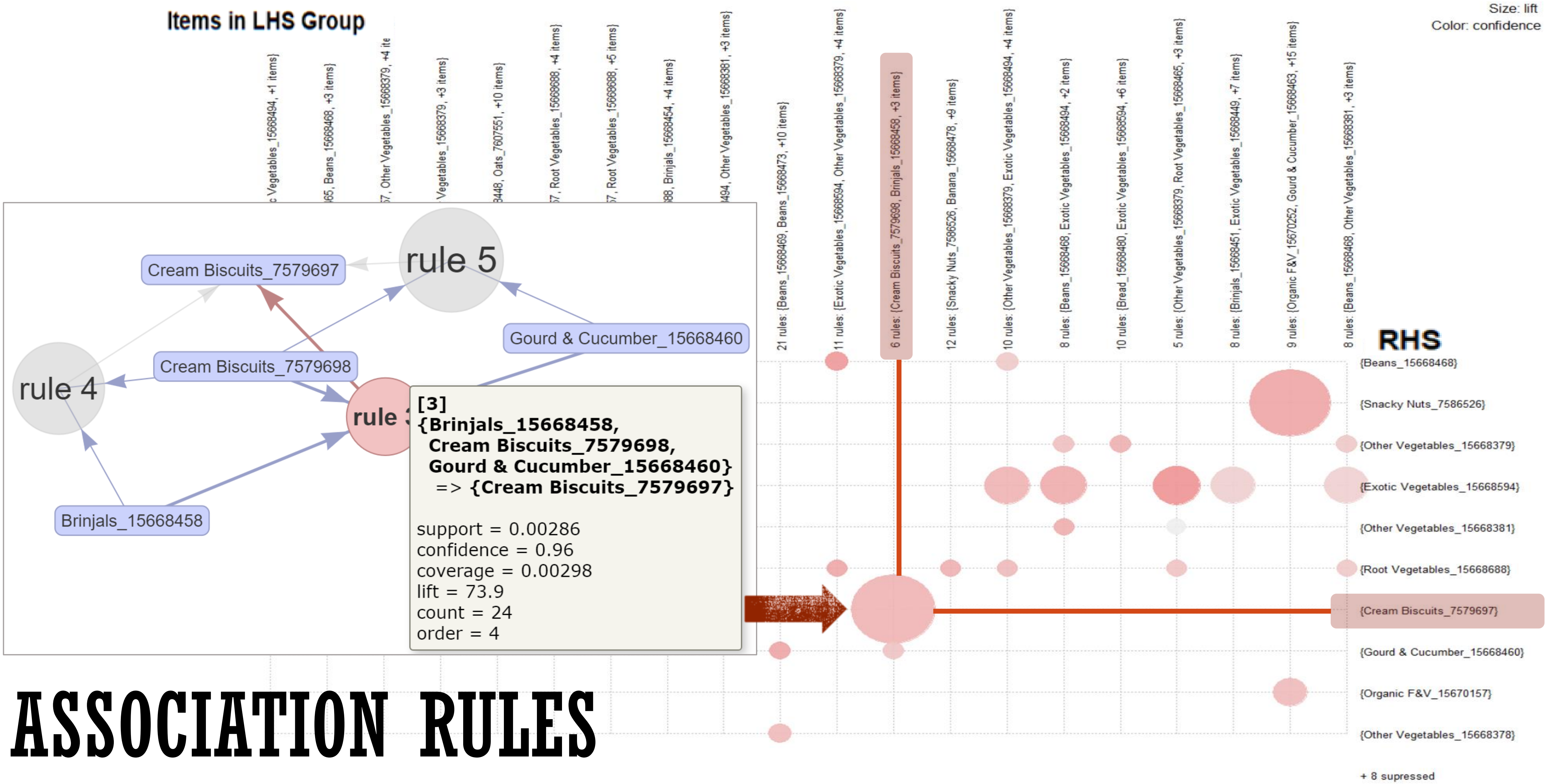
"SMART BASKET" – ASSOCIATION RULES





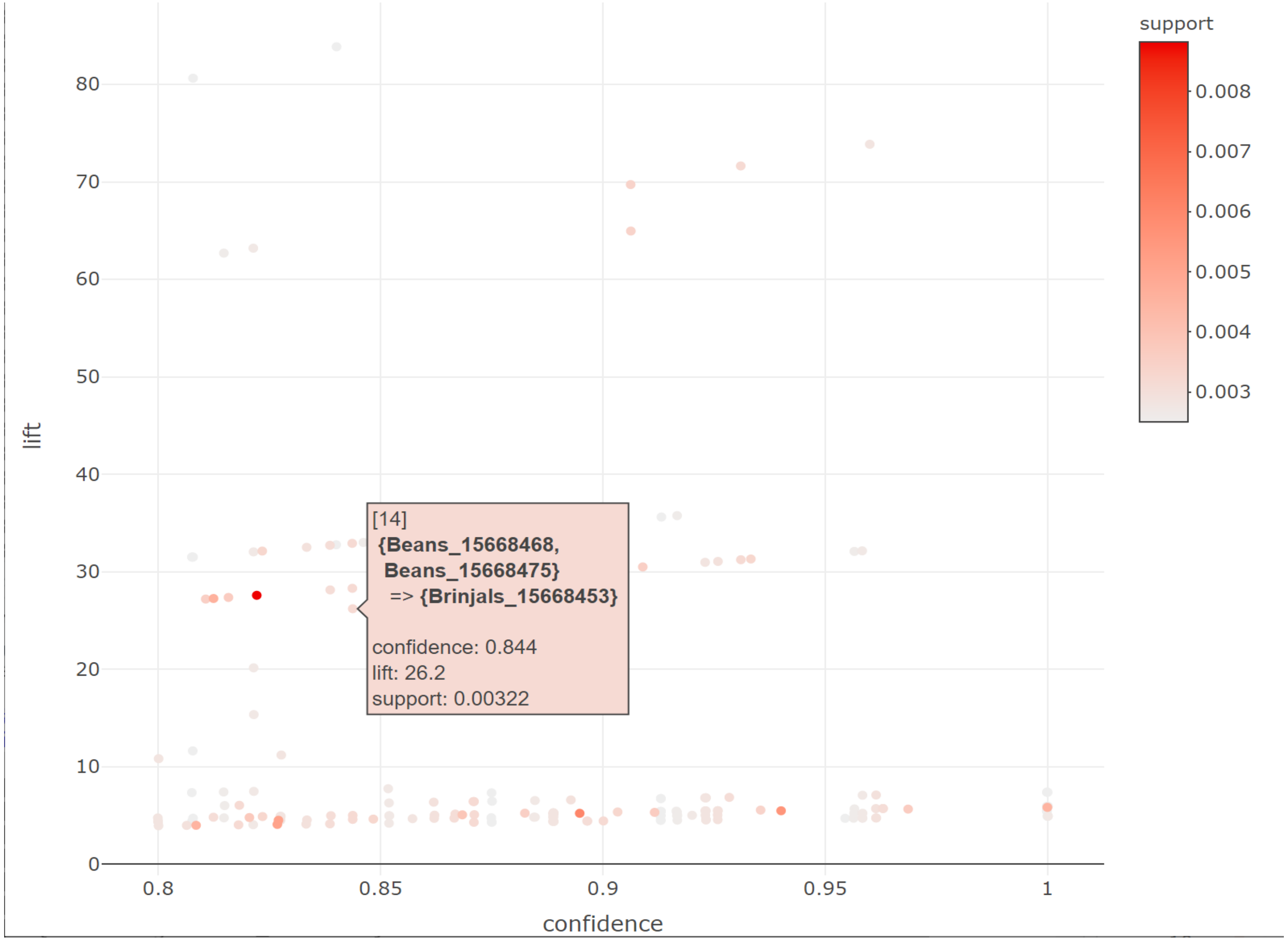


Grouped Matrix for 194 Rules



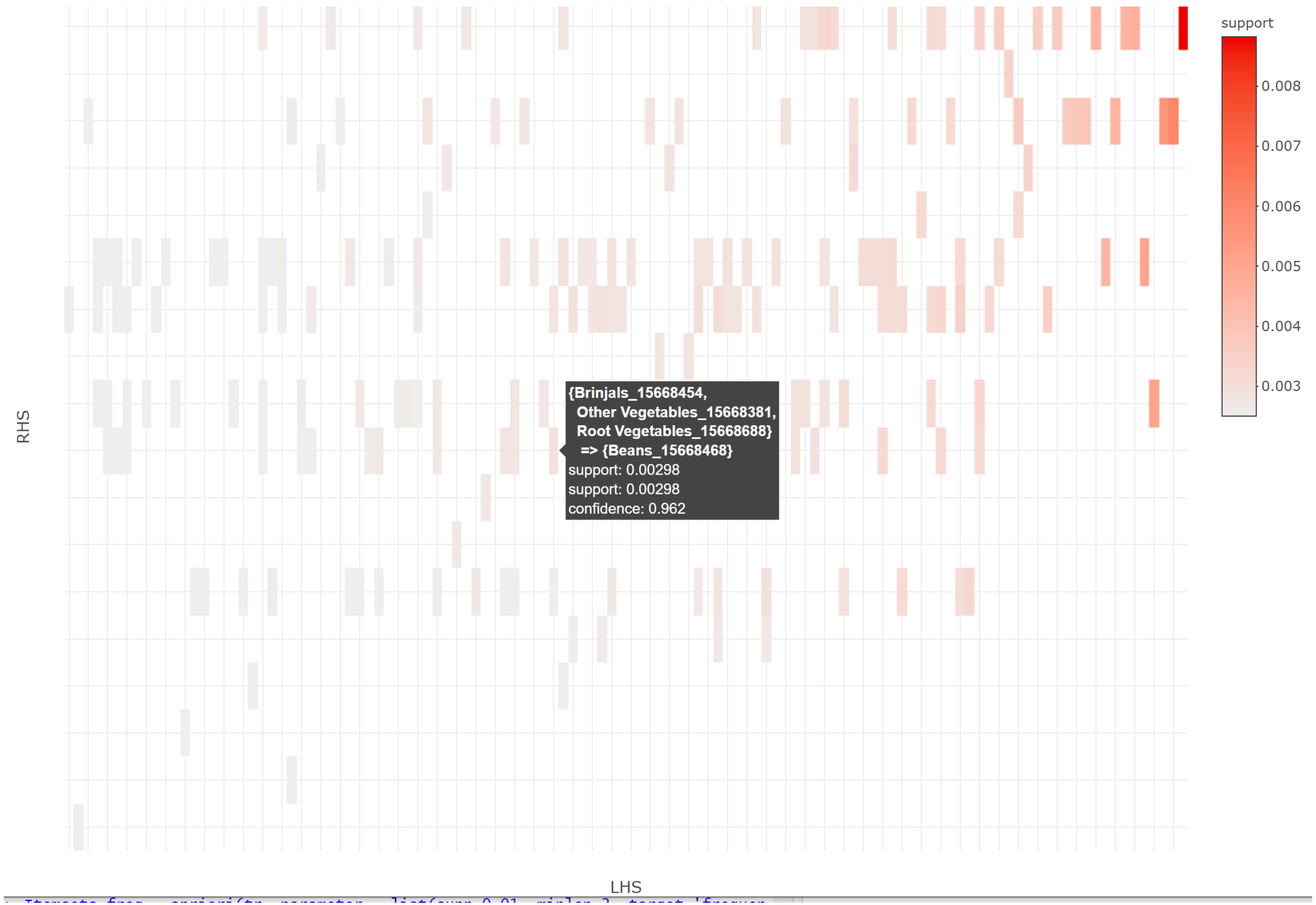
ASSOCIATION RULES

"SMART BASKET" – SCATTERED PLOT



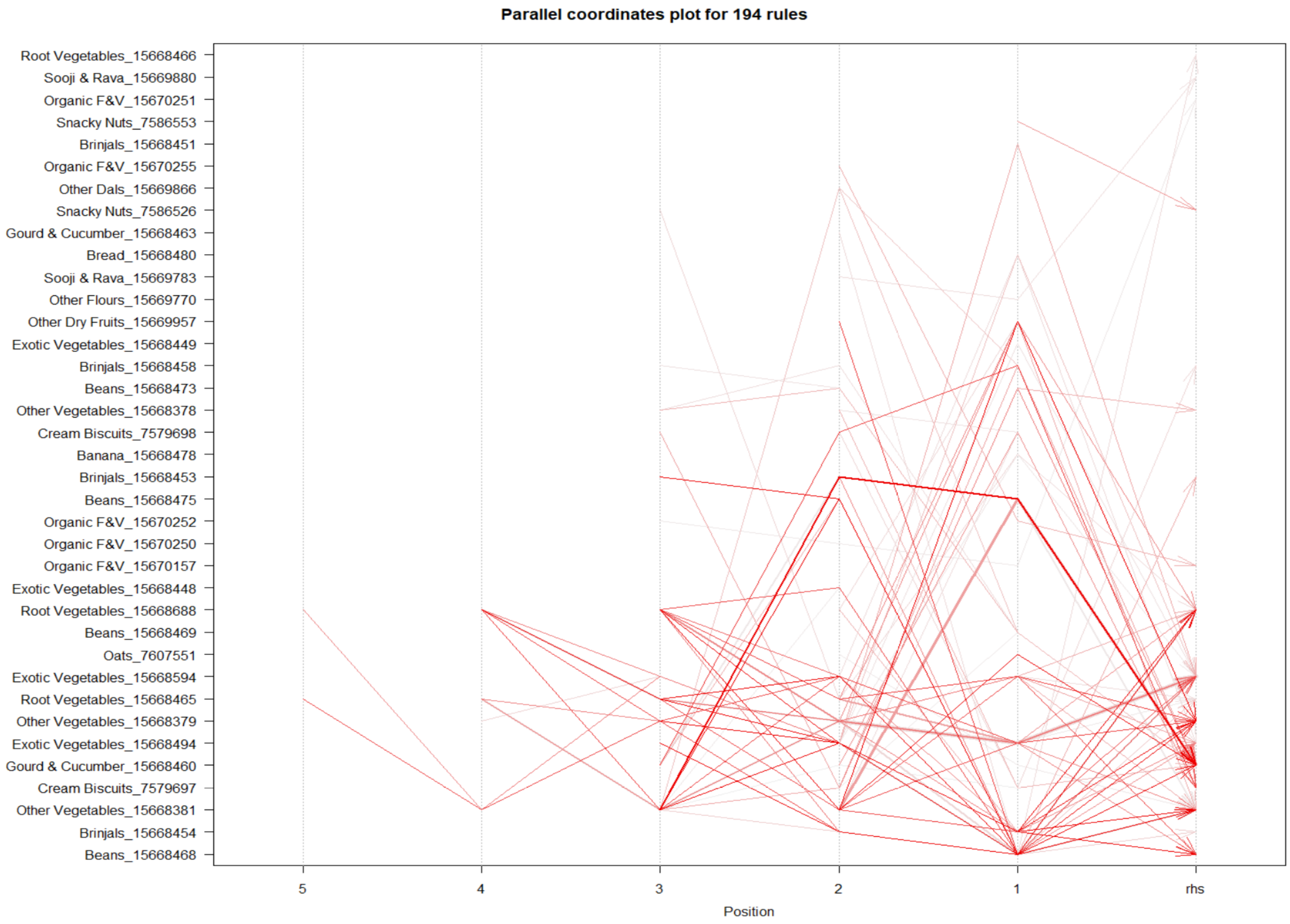
Scattered Plot for the extracted 194 rules with varying degree of Confidence, Lift and Support Levels.

"SMART BASKET" – MATRIX PLOT



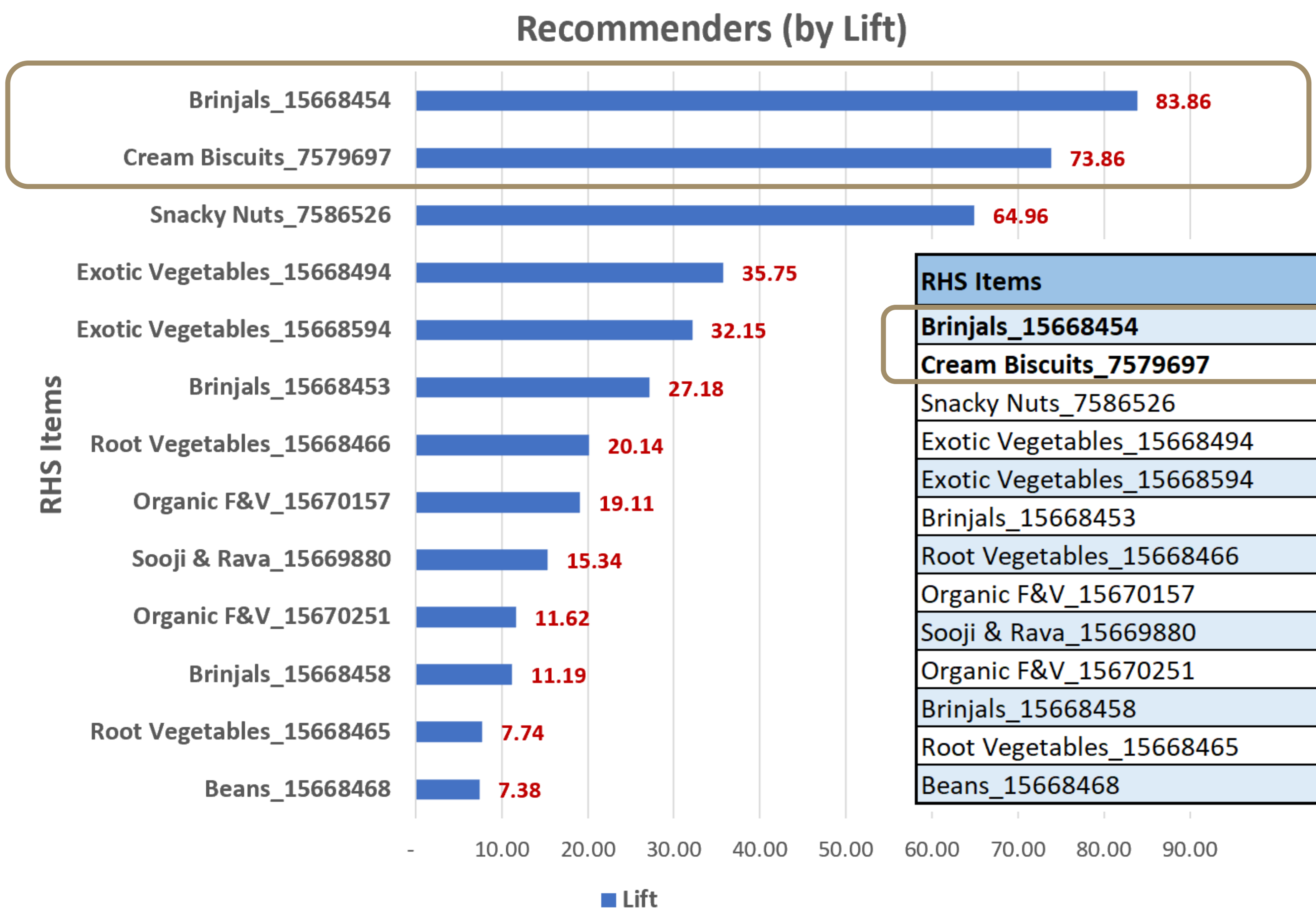
This plot shows the extracted 194 rules with RHS & LHS based on the varying degree of the support level

"SMART BASKET"



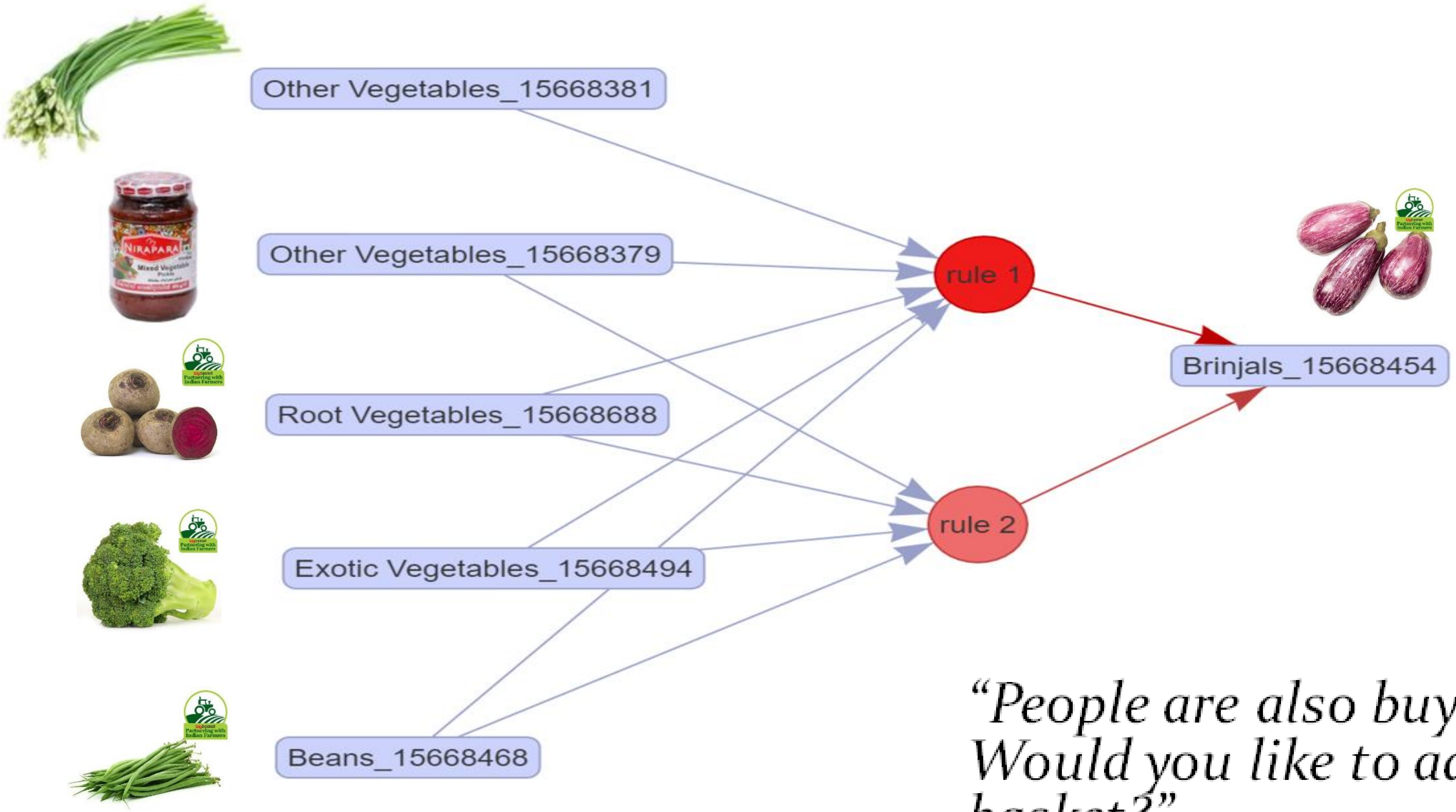
Parallel coordinates plot for the 194 rules

"SMART BASKET" - TOP RECOMMENDERS



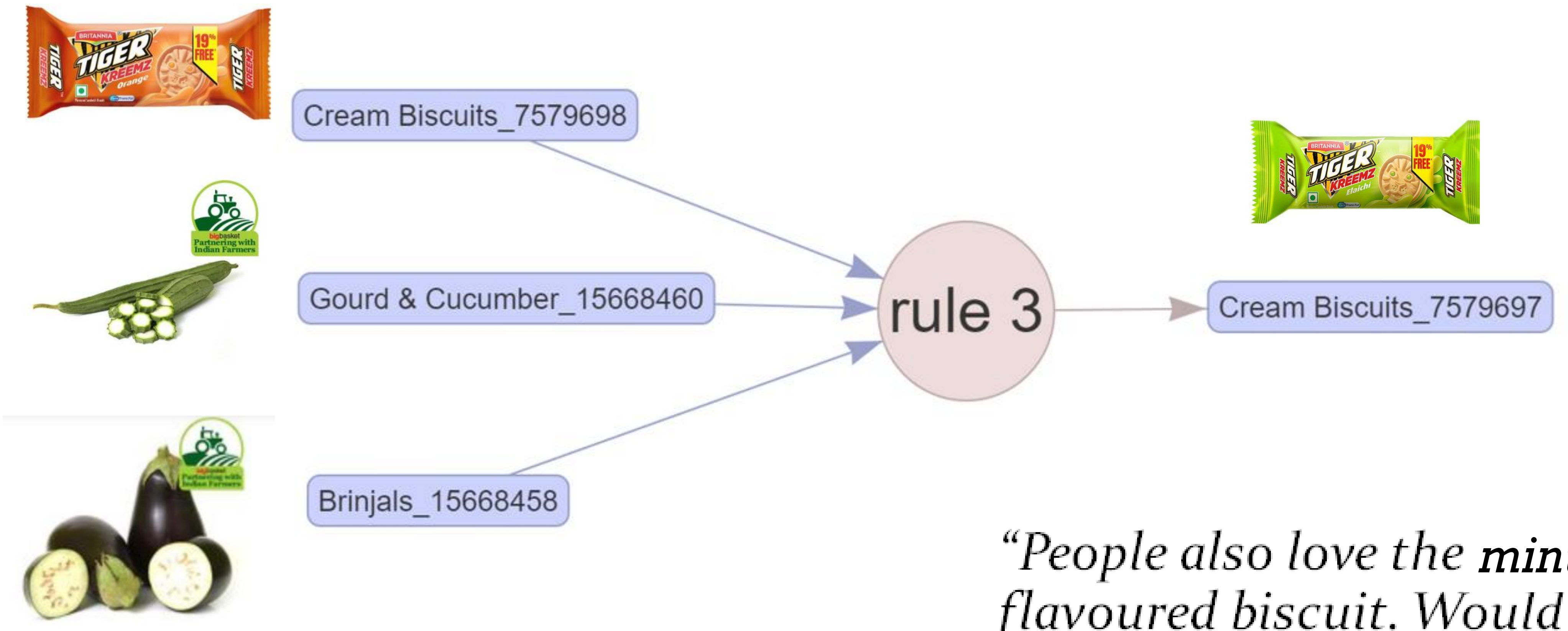
RHS Items	Description of Items	Lift	Count
Brinjals_15668454	**Brinjals - Varikatri Eggplant**	**83.86**	**2**
Cream Biscuits_7579697	**Cream Biscuits**	**73.86**	**5**
Snacky Nuts_7586526	Snacky Nuts	64.96	1
Exotic Vegetables_15668494	Broccoli	35.75	20
Exotic Vegetables_15668594	Lettuce - Iceberg	32.15	21
Brinjals_15668453	Brinjals - Bottle Shape Eggplant	27.18	3
Root Vegetables_15668466	Potatos	20.14	1
Organic F&V_15670157	Organic F&V - Banana	19.11	1
Sooji & Rava_15669880	Sooji & Rava	15.34	1
Organic F&V_15670251	Organic F&V - Tomatos	11.62	1
Brinjals_15668458	Brinjals - Long Green Eggplant	11.19	2
Root Vegetables_15668465	Root Vegetables - Beat Root	7.74	4
Beans_15668468	Beans - Haricot	7.38	20

PEOPLE ARE ALSO BUYING…



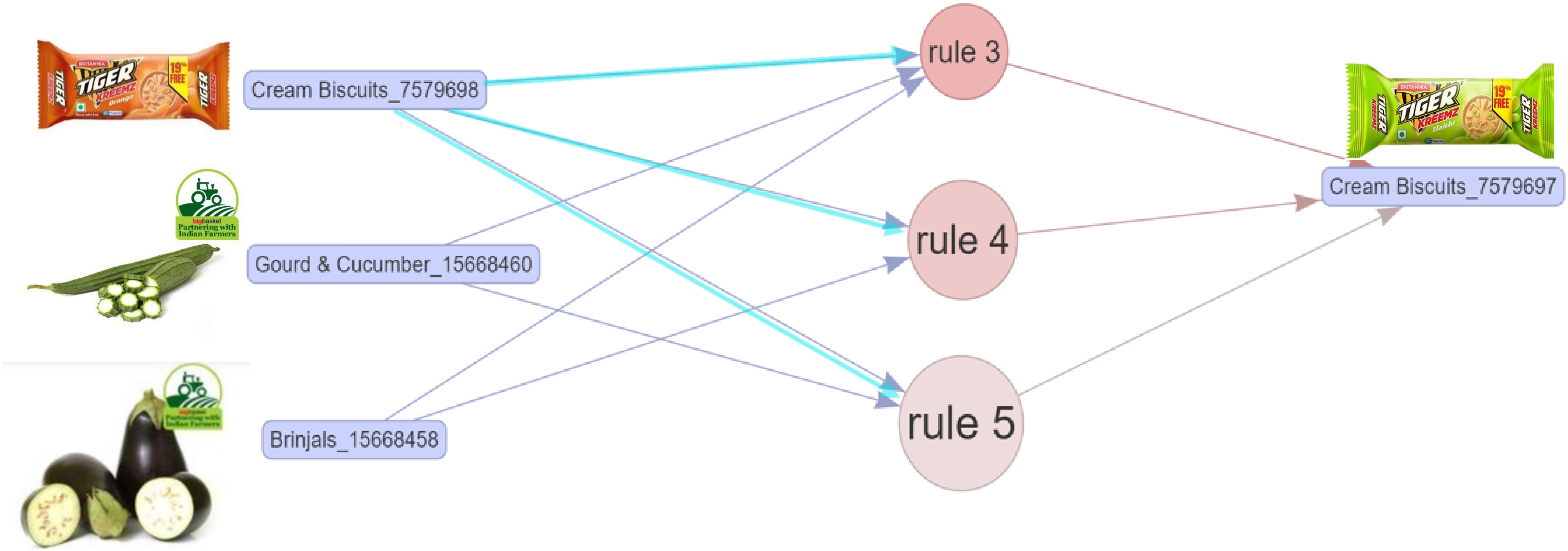
"People are also buying Brinjals. Would you like to add it to your basket?"

CROSS-SELLING OPPORTUNITY



"People also love the mint *flavoured biscuit. Would you like to add this to your basket?"*

CROSS-SELLING OPPORTUNITY



To further prove our cross-selling opportunity, Rule 4 & 5 are baskets with different combination of vegetables from Rule#3 , but they all have the same SKU of biscuit on the left leading to the same SKU of biscuit on the right. Our Rule #3 takes care of all combination of the previously mentioned items.

"SMART BASKET" IMPACT

- Create cross-selling & upselling opportunities by showing related items
- Increase the average basket size of orders
- Improve the user experience
- Make easier to find items
- Shorten the shopping time

"DID YOU FORGET?" DEMO AT CHECK-OUT

Scenario: Customer is shopping for ingredients for a Masala recipe

Organic Masalas & Spices_7587687

Other Sauces_34993740

Top 5 "Did you forget?" suggested items

Other Sauces_34993759

Cashews_15669800

Other Rice Products_3498944

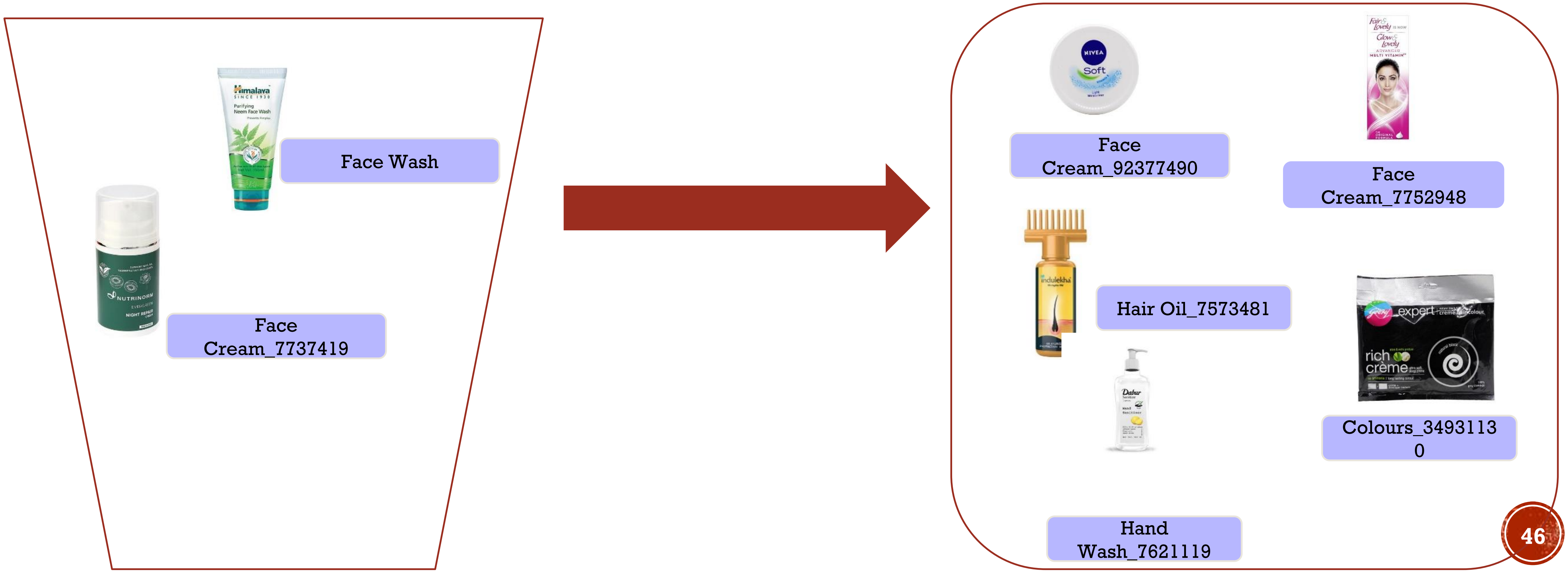
Organic Masalas & Spices_7587688

Organic Salt &Sugar & Jaggery_7587492

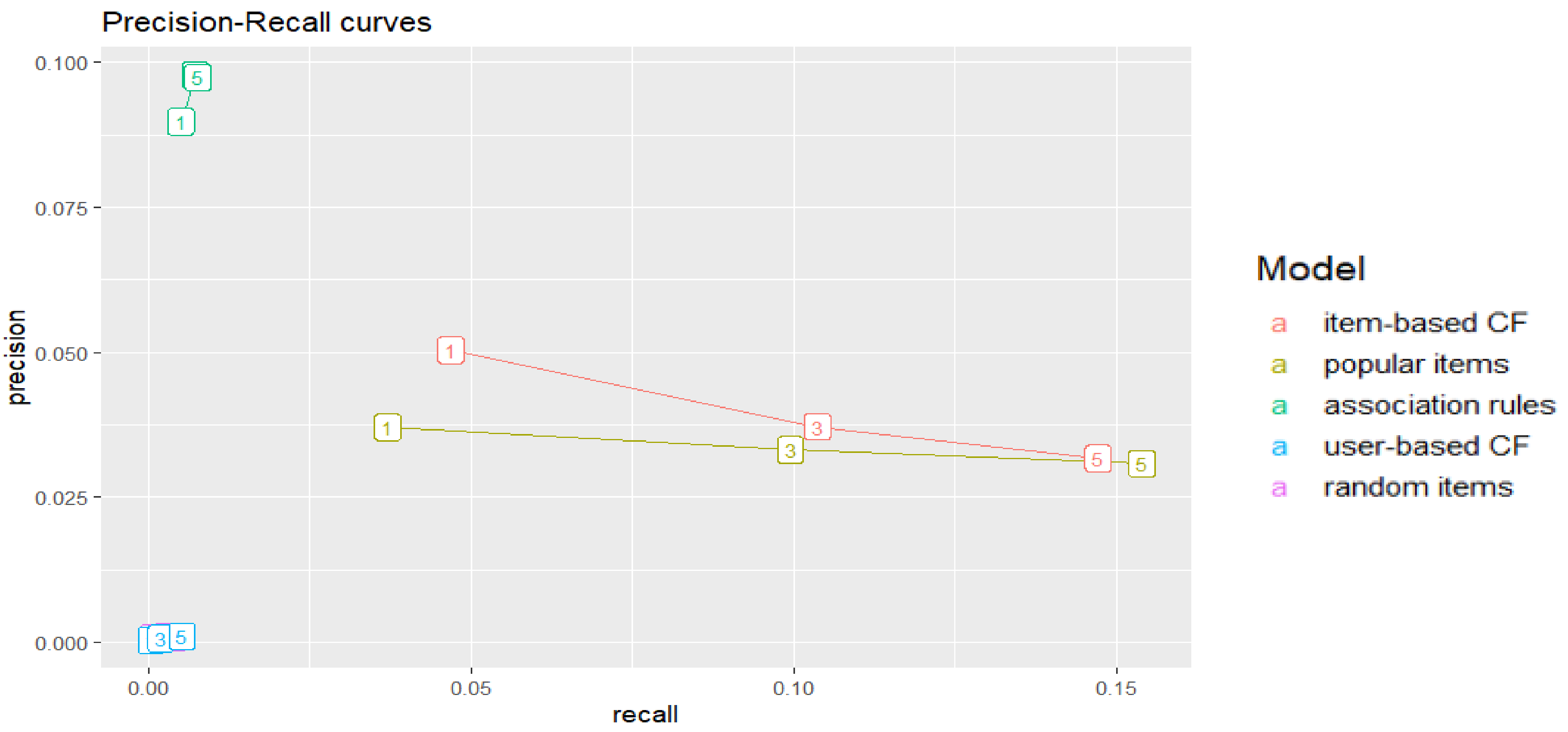
"DID YOU FORGET?" DEMO AT CHECK-OUT

Scenario: Customer's is shopping for personal care items

Top 5 "Did you forget?" suggested items



PRECISION-RECALL CURVES FOR "DID YOU FORGET" RECOMMENDERS



"DID YOU FORGET" IMPACT

Potential increase in sales of recommended items of 9% [1]

Reduce delivery costs by reducing number of weekly deliveries per customer due to forgotten items

Track sales of recommended items to determine actual sales lift and impact on number of weekly deliveries per customer

1. https://news.ufl.edu/articles/2018/09/how-helpful-are-product-recommendations-really.html