Association Rules:

*Zi Mei*

Executive summary:

Association rules are widely used in marketing analysis especially in retail chains to figure out customers’ purchases preferences. It is important for store owners to uncover what are consumers shopping habits and thus improve their shopping experience by offering more convenient services such as rearrange shelf items.

This analysis focus on association rules, trying to find out the items that are bought together and select the 20 pairs to help Dillard’s change the planograms across whole retail chains. For a big department store chain like Dillard’s, which opens 453 stores across the whole country, it is necessary to understand buying patterns and apply the association to sales promotion strategies and planograms and so on. For instance, Dillard’s could put the items usually bought together close to each other, suggesting customers to buy together.

RStudio is used, and the main methodology is association rules. After the analysis, I did find the 20 pairs candidates and the high confidence and lift indicate these pairs are more likely to be bought together rather than purely impulsive shopping.

Problem Statement:

Dillard’s wants to change the planograms across its whole retail chains to better serve customers, and due to budget limit, at most 20 moves could be made. The 20 pairs from 100 candidate pairs should be ready for Dillard’s to review and check.

Assumptions:

* All the data are correctly loaded into the database, not heavily skewed and I assume the shopping behaviors are similar even in different cities. Therefore, the association rules could be applied universally.
* Only purchased data are examined, since return items can not be ensured that customers want to buy them together at the beginning.

Methodology:

Association rules in RStudio is used to analyze this problem. since the original dataset is huge, I decided to go with the top 5 stores which generates most revenue, focusing on purchase data only and sampled 100,000 records. After some data exploring, I think sku, store, register, trannum and saledate could be used to identify the items which bought together. And the analysis steps are following:

Put store, register, trannum and saledate together as ‘busket’, and followed with sku then converted them to type factor to be prepared for association rules.

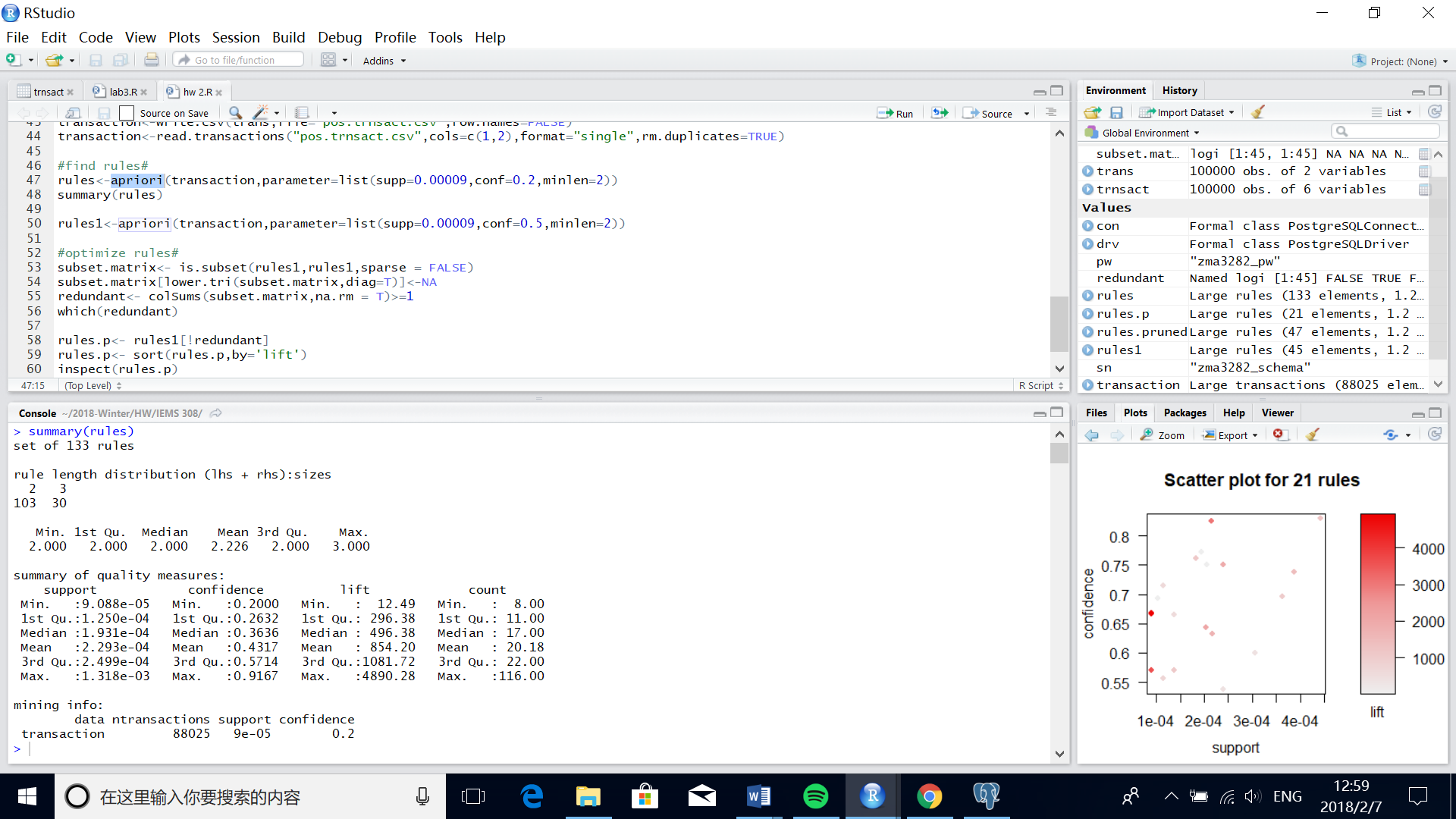
Find rules and inspect them, adjust parameters and find new rules

Check redundant and optimize the rules

Decide the final rules and plot them

Analysis:

At first, I applied apriori function, and set the parameters in confident at 0.2, it gave me over 100 pairs, and after checked the summary table found out that the mean confidence is 0.43, and I decided to lift the bar and set confidence level to 0.5.



After I changed the confidence standard to 0.5, RStudio generated 45 rules (output see appendix table 1). While after carefully examined them, I found some redundant records and decided to clean and optimize the output. And after clean redundant, there are 21 rules and I sorted them by lift. Due to Dillard’s budget limit, I think we can pick the top 20 since these items have larger lift number, indicating these items are more likely to be purchased together.

lhs rhs support confidence lift count

[1] {,"3468968} => {,"3498968} 9.088327e-05 0.6666667 4890.27778 8

[2] {,"3340314} => {,"3360314} 9.088327e-05 0.6666667 4191.66667 8

[3] {,"3370314} => {,"3360314} 9.088327e-05 0.5714286 3592.85714 8

[4] {,"3383572} => {,"3393572} 2.158478e-04 0.8260870 3029.84601 19

[5] {,"3546076} => {,"3526076} 2.044874e-04 0.6428571 2020.98214 18

[6] {,"3526076} => {,"3536076} 2.385686e-04 0.7500000 1941.72794 21

[7] {,"3503306} => {,"3523306} 2.158478e-04 0.6333333 1689.36869 19

[8] {,"3546076} => {,"3536076} 2.044874e-04 0.6428571 1664.33824 18

[9] {,"3346055} => {,"3406055} 3.862539e-04 0.7391304 1384.29695 34

[10] {,"3406055} => {,"3386055} 4.430560e-04 0.8297872 1178.09712 39

[11] {,"3336055} => {,"3386055} 1.817665e-04 0.7619048 1081.72043 16

[12] {,"3336055} => {,"3406055} 1.363249e-04 0.5714286 1070.21277 12

[13] {,"3346055} => {,"3386055} 3.635331e-04 0.6956522 987.65778 32

[14] {,"3438365,,"3468365} => {,"3458365} 1.136041e-04 0.5555556 708.73591 10

[15] {,"3338480,,"3528480} => {,"3388480} 1.136041e-04 0.7142857 610.43689 10

[16] {,"3328480,,"3518480} => {,"3378480} 1.363249e-04 0.6666667 543.36420 12

[17] {,"3513090,,"3563090} => {,"3463090} 3.067310e-04 0.6000000 369.33566 27

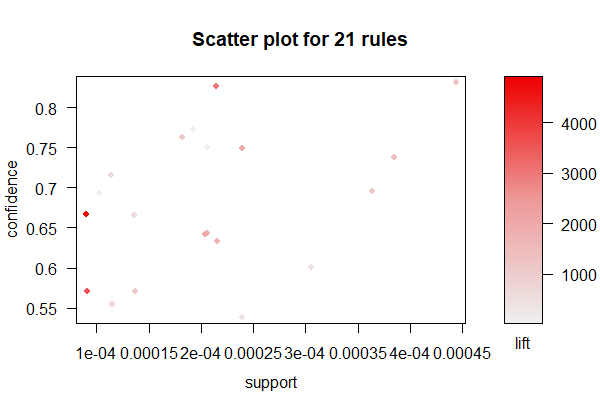
[18] {,"3483090,,"3533090} => {,"3573090} 2.385686e-04 0.5384615 367.42695 21

[19] {,"3528979,,"3547981} => {,"3537981} 1.931270e-04 0.7727273 44.16839 17

[20] {,"3517981,,"3547981} => {,"3537981} 2.044874e-04 0.7500000 42.86932 18

[21] {,"3547981,,"3558979} => {,"3537981} 1.022437e-04 0.6923077 39.57168 9

We can also see from the following plot, most of the rules have confidence over 0.6, which is good and the lift also suggest that these items are usually purchased together rather than impulsive shopping.



Conclusions:

After done analysis, I found the 20 pairs that are usually bought together and I suggest Dillard’s could change the planograms of those items, putting them together or very close, implying customers could buy them together.

In addition to change planograms, the association rules could be applied to promotion. Dillard’s could have one item in rules on sale instead of both of them have discounts. Also, items in rules could be used for bundle sale to generate more profits.

Next step:

The next step could focus on specific region, such as states and cities, to generate more detailed understanding of the purchase pattern of local residents. For instance, residents in Chicago may have a different purchase preference compared to residents in Miami, it could be interesting to study these two cities and make comparison to see any difference and make more customization on marketing strategies.

Appendix: table 1

lhs rhs support confidence lift count

[1] {,"3468968} => {,"3498968} 9.088327e-05 0.6666667 4890.27778 8

[2] {,"3498968} => {,"3468968} 9.088327e-05 0.6666667 4890.27778 8

[3] {,"3340314} => {,"3360314} 9.088327e-05 0.6666667 4191.66667 8

[4] {,"3360314} => {,"3340314} 9.088327e-05 0.5714286 4191.66667 8

[5] {,"3370314} => {,"3360314} 9.088327e-05 0.5714286 3592.85714 8

[6] {,"3360314} => {,"3370314} 9.088327e-05 0.5714286 3592.85714 8

[7] {,"3383572} => {,"3393572} 2.158478e-04 0.8260870 3029.84601 19

[8] {,"3393572} => {,"3383572} 2.158478e-04 0.7916667 3029.84601 19

[9] {,"3503306} => {,"3523306} 2.158478e-04 0.6333333 1689.36869 19

[10] {,"3523306} => {,"3503306} 2.158478e-04 0.5757576 1689.36869 19

[11] {,"3336055} => {,"3406055} 1.363249e-04 0.5714286 1070.21277 12

[12] {,"3336055} => {,"3386055} 1.817665e-04 0.7619048 1081.72043 16

[13] {,"3546076} => {,"3526076} 2.044874e-04 0.6428571 2020.98214 18

[14] {,"3526076} => {,"3546076} 2.044874e-04 0.6428571 2020.98214 18

[15] {,"3546076} => {,"3536076} 2.044874e-04 0.6428571 1664.33824 18

[16] {,"3536076} => {,"3546076} 2.044874e-04 0.5294118 1664.33824 18

[17] {,"3526076} => {,"3536076} 2.385686e-04 0.7500000 1941.72794 21

[18] {,"3536076} => {,"3526076} 2.385686e-04 0.6176471 1941.72794 21

[19] {,"3346055} => {,"3406055} 3.862539e-04 0.7391304 1384.29695 34

[20] {,"3406055} => {,"3346055} 3.862539e-04 0.7234043 1384.29695 34

[21] {,"3346055} => {,"3386055} 3.635331e-04 0.6956522 987.65778 32

[22] {,"3386055} => {,"3346055} 3.635331e-04 0.5161290 987.65778 32

[23] {,"3406055} => {,"3386055} 4.430560e-04 0.8297872 1178.09712 39

[24] {,"3386055} => {,"3406055} 4.430560e-04 0.6290323 1178.09712 39

[25] {,"3336055,,"3406055} => {,"3386055} 1.249645e-04 0.9166667 1301.44489 11

[26] {,"3336055,,"3386055} => {,"3406055} 1.249645e-04 0.6875000 1287.59973 11

[27] {,"3526076,,"3546076} => {,"3536076} 1.704061e-04 0.8333333 2157.47549 15

[28] {,"3536076,,"3546076} => {,"3526076} 1.704061e-04 0.8333333 2619.79167 15

[29] {,"3526076,,"3536076} => {,"3546076} 1.704061e-04 0.7142857 2245.53571 15

[30] {,"3338480,,"3528480} => {,"3388480} 1.136041e-04 0.7142857 610.43689 10

[31] {,"3346055,,"3406055} => {,"3386055} 3.294519e-04 0.8529412 1210.97011 29

[32] {,"3346055,,"3386055} => {,"3406055} 3.294519e-04 0.9062500 1697.29056 29

[33] {,"3386055,,"3406055} => {,"3346055} 3.294519e-04 0.7435897 1422.92363 29

[34] {,"3438365,,"3468365} => {,"3458365} 1.136041e-04 0.5555556 708.73591 10

[35] {,"3328480,,"3518480} => {,"3378480} 1.363249e-04 0.6666667 543.36420 12

[36] {,"3378480,,"3518480} => {,"3328480} 1.363249e-04 0.5454545 600.17045 12

[37] {,"3547981,,"3558979} => {,"3537981} 1.022437e-04 0.6923077 39.57168 9

[38] {,"3537981,,"3558979} => {,"3547981} 1.022437e-04 0.5625000 144.35587 9

[39] {,"3483090,,"3533090} => {,"3573090} 2.385686e-04 0.5384615 367.42695 21

[40] {,"3533090,,"3573090} => {,"3483090} 2.385686e-04 0.5833333 446.50362 21

[41] {,"3528979,,"3547981} => {,"3537981} 1.931270e-04 0.7727273 44.16839 17

[42] {,"3528979,,"3537981} => {,"3547981} 1.931270e-04 0.5483871 140.73404 17

[43] {,"3513090,,"3563090} => {,"3463090} 3.067310e-04 0.6000000 369.33566 27

[44] {,"3463090,,"3513090} => {,"3563090} 3.067310e-04 0.5192308 341.08424 27

[45] {,"3517981,,"3547981} => {,"3537981} 2.044874e-04 0.7500000 42.86932 18