Roger Mei

Planogram Restructuring Recommendations for Dillard's in Illinois

Executive Summary

The Dillard's point-of-sales data contains information about Dillard's transactions over the years 2004 and 2005 across the US. This information can be used to determine association rules between store items and their characteristics. In this study, items of different brands but part of the same overarching parent company are of primary interest. Specifically, the brands Polo Jeans Co. and Polo Ralph Lauren are analyzed.

Apriori is an algorithm that identifies frequent items in the database and extends these items to larger and larger sets as long as these item sets appear more often than a given minimum frequency level, also called minimum support. Using apriori, all combinations of the frequent item sets are tested for association rules, and only the rules with confidence above a given threshold are kept.

In this analysis, the minimum support level used is 0.05% and the minimum confidence used to cut down on the number of rules is 0.2. The 100 SKUs chosen as the best candidates to modify the planograms are the antecedents of the rules with the 100 highest lifts.
Similarly, the rules with the 20 highest lifts are used to determine the 20 moves in the stores

Taking a look at the results, all antecedent and consequent items in the 20 rules with the highest lift metric have the same size and brand. It is clear that customers that buy one article of clothing would buy another of the same size and brand. Because of this, none of the 142 determined association rules involve both Polo Jeans Co. and Polo Ralph Lauren items in the same rule. It doesn't matter that the two brands are subsidiaries of the same parent company. These two brands should be kept in separate departments with their associated items close to each other. Customers who wear Polo Jeans Co. clothes may not like Polo Ralph Lauren attire and vice versa. This may be that although they share the same parent company, they are viewed as completely different styles.

The 100 SKUs that are the best candidates to modify the planograms with are:
5773003, 3074240, 4437898, 4823869, 3263869, 3148046, 6837616, 5353003, 3904240, 3064240, 3894240, 306827, 8686826, 3287898, 5454240, 8234240, 5906249, 5666249, 7617591, 304072, 4568105, 9238597, 2324085, 2864398, 632916, 9294395, 9182555, 3986264, 456264, 8632552, 2652631, 9324395, 4662914, 6337594, 4537594, 6554238, 6684219, 8224240, 7046264, 6034219, 1814220, 1824220, 6372524, 5452548, 1324228, 6777933, 4917933, 1056223, 9068481, 5677937, 6184085, 3396339, 336339, 9146222, 4080379, 8060116, 4822018, 4862018, 9432506, 9442506, 8382552, 4256463, 9296463, 1897917, 6337897, 2264226, 2934275, 574072, 2484725, 954722, 6564238, 1824225, 984225, 7784238, 5914721, 9797919, 4618105, 9721644, 7926810, 8953989, 2870559, 6487933, 4637933, 4090379, 2307933, 6984275, 2397933, 1144228, 8783989, 2340559, 8784749, 9734273, 3907933, 3897933, 5036463, 294072, 2314085, 7727933, 8437933, 3204072.

The items with the SKUS in the antecedents column of Table 1 below should be moved next to the items in the consequents column. These should be the 20 moves made in the stores' planograms.

antecedents	consequents
(5773003)	(5353003)
(3074240)	(3904240)
(4437898)	(3287898)
(4823869)	(3263869)
(3263869)	(4823869)
(3148046)	(6837616)
(6837616)	(3148046)
(5353003)	(5773003)
(3904240)	(3074240)
(3064240)	(3894240)
(3894240)	(3064240)
(306827)	(8686826)
(8686826)	(306827)
(3287898)	(4437898)
(5454240)	(8234240)
(8234240)	(5454240)
(5906249)	(5666249)
(5666249)	(5906249)
(7617591)	(4568105, 9238597)
(304072)	(2324085)

Table 1

Problem Statement

Dillard's, a major retail chain with several stores across the US, is interested in rearranging the floors of their stores. In retail terms, this is referred to as changing their planograms. They are interested in finding the 100 SKUs of items that are the best candidates to modify the planograms. Due to a constrained budget, they can only make at most 20 moves across the entire chain.

Because of these restrictions, only stores in Illinois are to be analyzed as building a new store in Chicago may be of business interest as well. To start off, analysis of brands that are subsidiaries of the same company are looked at. Polo Jeans Company, LLC is a subsidiary of the Ralph Lauren Corporation. However, in Dillard's, the Ralph Lauren department contains mostly Polo Ralph Lauren brands and some other brands like Tommy Hilfiger but not Polo Jeans Co.

This study looks at only the brand Polo Jeans Co. and clothing in the department Ralph Lauren to see if clothing by different brands but part of the same company would have any association. If there are association rules, then clothing of the brand Polo Jeans Co. can be placed next to their associated items in the same Ralph Lauren department. In addition to modifying the locations of items, association rules for characteristics of clothing such as styles, colors, and brands can be determined.

Assumptions

- The company hasn't performed this type of analysis before and it is highly unlikely that SKUs are already appropriately close to each other.
- The data is accurate and as according to the data, most items in the Ralph Lauren department are items of the brand Polo Ralph Lauren. All items of the brand Polo Jeans Co. are currently not in the Ralph Lauren department.
- All clothing items are homogeneous. All items have similar frequencies.

Methodology

First, data exploration is performed in order to determine worthy portions of the data to analyze. This study analyzes transaction data from Illinois stores. It also looks at the brands Polo Jeans Co. and clothing in the Ralph Lauren department. Although Polo Jeans Co. is a subsidiary brand of the Ralph Lauren Corporation, no Polo Jeans Co. clothes are in the Ralph Lauren department.

Next, the infrequent items are filtered out using a minimum support of 0.05%. We can use the same minimum support level because all items are clothing and also all these clothes are similar in brand and style.

After obtaining the frequent item sets, the apriori algorithm is run to determine association rules of those item sets. A confidence level threshold of 0.2 is used, disposing of rules with confidence lower than that.

The rules are then sorted by lift. The antecedents in the top 100 rules are chosen as the best SKU candidates to modify the planograms. Items in these rules are recommended to be moved next to each other within the store and within the same department in order to increase sales.

Furthermore, the rules are searched through to see if any contained both brands to answer the question whether subsidiary brands have association rules or not. Other characteristics such as style, brand, size, and color are analyzed as well within the association rules.

Analysis

The transaction table has a composite key made up of many primary keys. Combining the sale date, store number, transaction number, and register number into a single identifier, transactions are combined and viewed more easily. Looking at the number of items in a transaction as shown in Table 2 (where the left column is transaction id and right column is number of items), every transaction seems like it has a typical amount of items. The highest number of items in a single transaction among all the kept baskets is 14. This is quite low compared to baskets of over 30 items in some transactions in the complete dataset without filtering.

Because of these low amounts of items in a basket, it may be more difficult to find association rules compared to larger datasets. Thus, a relatively low minimum support level of 0.05% is used so as to not filter out most of the transactions and important rules that could be found. The same minimum support level is used for all items in this subset of data because all these items are articles of clothing mostly from the brand Polo Jeans Co. and Polo Ralph Lauren.

317	14
1831	12
856	11
1873	8
321	8
305	8
598	8
2673	8
1911	8
230	7
3437	7
600	7
351	7
1399	7
322	7
1844	7
490	7
3341	7
621	7
2581	6

Table 2

Further looking at the frequencies of each SKU as shown in Table 3 below (left column is SKU and right is frequency), a minimum support level of 0.05% is justified. Most SKUs have a support less than 0.1%. Therefore, the minimum support level used in the apriori algorithm must be less than 0.1% in order to capture the most important association rules.

4747172	0.003653
5760331	0.003288
5540331	0.003288
5770331	0.003288
1726624	0.002922
4707172	0.002922
7604275	0.002557
6974275	0.002374
6842459	0.002374
9586806	0.002192
3082459	0.002192
3092459	0.002192
6474085	0.002192
4687172	0.002009
5332459	0.002009
3022459	0.002009
6874275	0.002009
7072459	0.001826
8784749	0.001826
1314861	0.001826
6884861	0.001826
5584721	0.001826
1736624	0.001826
5750331	0.001826
6664085	0.001644
3072459	0.001644
6694085	0.001644
1673501	0.001644
7926810	0.001644
504225	0.001644
	...
9418721	0.000183
7444861	0.000183
9466316	0.000183
1718737	0.000183
8622552	0.000183
3987933	0.000183
1184228	0.000183
7414258	0.000183
7014899	0.000183
4176261	0.000183
3664247	0.000183
2572502	0.000183
5777679	0.000183

Table 3

As it turns out, the minimum support level of 0.05% results in 142 association rules found. This is a good support level because it results in a number of rules close to the target of 100. As shown in Table 4 of the top 20 rules by lift, all supports are 0.0545%.

antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(5773003)	(5353003)	0.000545	0.000545	0.000545	1.000000	1836.0
(3074240)	(3904240)	0.000545	0.000545	0.000545	1.000000	1836.0
(4437898)	(3287898)	0.000545	0.000545	0.000545	1.000000	1836.0
(4823869)	(3263869)	0.000545	0.000545	0.000545	1.000000	1836.0
(3263869)	(4823869)	0.000545	0.000545	0.000545	1.000000	1836.0
(3148046)	(6837616)	0.000545	0.000545	0.000545	1.000000	1836.0
(6837616)	(3148046)	0.000545	0.000545	0.000545	1.000000	1836.0
(5353003)	(5773003)	0.000545	0.000545	0.000545	1.000000	1836.0
(3904240)	(3074240)	0.000545	0.000545	0.000545	1.000000	1836.0
(3064240)	(3894240)	0.000545	0.000545	0.000545	1.000000	1836.0
(3894240)	(3064240)	0.000545	0.000545	0.000545	1.000000	1836.0
(306827)	(8686826)	0.000545	0.000545	0.000545	1.000000	1836.0
(8686826)	(306827)	0.000545	0.000545	0.000545	1.000000	1836.0
(3287898)	(4437898)	0.000545	0.000545	0.000545	1.000000	1836.0
(5454240)	(8234240)	0.000545	0.000817	0.000545	1.000000	1224.0
(8234240)	(5454240)	0.000817	0.000545	0.000545	0.666667	1224.0
(5906249)	(5666249)	0.000817	0.000545	0.000545	0.666667	1224.0
(5666249)	(5906249)	0.000545	0.000817	0.000545	1.000000	1224.0
(7617591)	(4568105, 9238597)	0.000817	0.000545	0.000545	0.666667	1224.0
(304072)	(2324085)	0.000545	0.000817	0.000545	1.000000	1224.0

Table 4

24% of the 142 determined rules have a confidence of 1.0. 15% of the 142 determined rules have a confidence of 0.5. 19% of the 142 determined rules have a confidence less than 0.33. In order to further reduce the number of rules down to 100, only rules with confidence above 0.2 are kept (removing 7% of the rules) and the top 100 rules by lift are chosen.

Conclusions

As shown in Appendix A, the top 100 SKU candidates to modify the planograms are in rules that have high lifts. The top rule has a lift of 1836. This means that the combination of items with SKU 5773003 and 5353003 is found 1836 times more often than expected if bought independently. The rule at the bottom of the top 100 has a lift of 350. This is still a very high lift, meaning that all these items have high correlations in these transactions.

Most of these rules have confidence levels above 0.5 as well meaning that most of the consequent items have more than a 50% chance to be in the same basket given that the antecedent items are bought too. If moved next to each other in the store, customers will have a high chance of purchasing both items contained in the determined association rules.

Furthermore, clothing of subsidiary brands of the same company do not have association rules. Customers do not buy clothing of the brands Polo Jeans Co. and Polo Ralph Lauren together. Although the two brands have similar clothing styles, name, and even logo, none of the 100 association rules shown in Appendix A contain both brands. As shown in Table 5 below, the top 20 association rules with the highest lifts have antecedent and consequent items with the exact same sizes and brands. For example, if a customer buys a Polo Ralph Lauren item of size 24, the customer is also likely to buy another Polo Ralph Lauren item of size 24. Same sizes and brands should be kept next to each other within the same departments.

antecedents	STYLE	COLOR	SIZE	BRAND	consequents	STYLE	COLOR	SIZE	BRAND
5773003	921641N70917	RED	12	TOMMY HI	5353003	579118N70223	SNOW WHITE	12	TOMMY HI
3074240	47 RG5427	882DEVONRE	24	RALPH LA	3904240	15 RG5427	210ANWT	24	RALPH LA
4437898	62 RJ1443	PALEROSE	18	RALPH LA	3287898	19 RJ1440	PINK	18	RALPH LA
4823869	97 RG8422	168FLORAL	9	RALPH LA	3263869	96 RG8422	168FLORAL	9	RALPH LA
3263869	96 RG8422	168FLORAL	9	RALPH LA	4823869	97 RG8422	168FLORAL	9	RALPH LA
3148046	55 BG5428	110WHITE	12	RALPH LA	6837616	55 BG5428	908BEACHRY	12	RALPH LA
6837616	55 BG5428	908BEACHRY	12	RALPH LA	3148046	55 BG5428	110WHITE	12	RALPH LA
5353003	579118N70223	SNOW WHITE	12	TOMMY HI	5773003	921641N70917	RED	12	TOMMY HI
3904240	15 RG5427	210ANWT	24	RALPH LA	3074240	47 RG5427	882DEVONRE	24	RALPH LA
3064240	47 RG5427	882DEVONRE	18	RALPH LA	3894240	15 RG5427	210ANWT	18	RALPH LA
3894240	15 RG5427	210ANWT	18	RALPH LA	3064240	47 RG5427	882DEVONRE	18	RALPH LA
306827	972 33LEVF	413CLASC N	M	POLO JEA	8686826	967 33LEVF	413CLASC N	M	POLO JEA
8686826	967 33LEVF	413CLASC N	M	POLO JEA	306827	972 33LEVF	413CLASC N	M	POLO JEA
3287898	19 RJ1440	PINK	18	RALPH LA	4437898	62 RJ1443	PALEROSE	18	RALPH LA
5454240	55 RG5425	746GRMCYGR	18	RALPH LA	8234240	88 RG5427	829NEWMGNT	18	RALPH LA
8234240	88 RG5427	829NEWMGNT	18	RALPH LA	5454240	55 RG5425	746GRMCYGR	18	RALPH LA
5906249	76 BH7436	AVTRNAV	18	RALPH LA	5666249	55 BH7436	VALORRE	18	RALPH LA
5666249	55 BH7436	VALORRE	18	RALPH LA	5906249	76 BH7436	AVTRNAV	18	RALPH LA
7617591	2AT7 53UBS2	407COVE BL	4XT	POLO JEA	4568105	9AT7 53UBS2	100WHITEWS	4XT	POLO JEA
7617591	NaN	NaN	NaN	NaN	9238597	2AT7 83UBS4	720SHMR YL	4XT	POLO JEA
304072	8TB7 13UCH8	612BARN RE	4X	POLO JEA	2324085	8TB7 13UCH8	413HDSN NV	4X	POLO JEA

Table 5

Next Steps

Although items of subsidiary brands of the same parent company do not have association rules, the analysis discovered that items of the same size and brand do have association rules. For the next moves in the planograms, analysis should be performed on items of the same brand and size.

Furthermore, analysis on styles and colors should be done to see which clothing are of similar style. Knowing which styles can be grouped together allows the retail chain to make marketing campaigns and directly target customer interests. Clustering could also be done to see how many distinct styles and market segments there may be.

It'd also be interesting to see if clothing of similar brands or styles can be confused with each other. Convolutional neural networks could be used on images of similar items, like in Figure 1, to determine the similarity of articles of clothing. This information can be used to further understand the optimal spacing of items in a store.

Figure 1 - Left: Polo Jeans Co., Right: Polo Ralph Lauren

Appendix A – Best 100 SKU candidates to modify planograms

antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(5773003)	(5353003)	0.000545	0.000545	0.000545	1.000000	1836.0
(3074240)	(3904240)	0.000545	0.000545	0.000545	1.000000	1836.0
(4437898)	(3287898)	0.000545	0.000545	0.000545	1.000000	1836.0
(4823869)	(3263869)	0.000545	0.000545	0.000545	1.000000	1836.0
(3263869)	(4823869)	0.000545	0.000545	0.000545	1.000000	1836.0
(3148046)	(6837616)	0.000545	0.000545	0.000545	1.000000	1836.0
(6837616)	(3148046)	0.000545	0.000545	0.000545	1.000000	1836.0
(5353003)	(5773003)	0.000545	0.000545	0.000545	1.000000	1836.0
(3904240)	(3074240)	0.000545	0.000545	0.000545	1.000000	1836.0
(3064240)	(3894240)	0.000545	0.000545	0.000545	1.000000	1836.0
(3894240)	(3064240)	0.000545	0.000545	0.000545	1.000000	1836.0
(306827)	(8686826)	0.000545	0.000545	0.000545	1.000000	1836.0
(8686826)	(306827)	0.000545	0.000545	0.000545	1.000000	1836.0
(3287898)	(4437898)	0.000545	0.000545	0.000545	1.000000	1836.0
(5454240)	(8234240)	0.000545	0.000817	0.000545	1.000000	1224.0
(8234240)	(5454240)	0.000817	0.000545	0.000545	0.666667	1224.0
(5906249)	(5666249)	0.000817	0.000545	0.000545	0.666667	1224.0
(5666249)	(5906249)	0.000545	0.000817	0.000545	1.000000	1224.0
(7617591)	(4568105, 9238597)	0.000817	0.000545	0.000545	0.666667	1224.0
(304072)	(2324085)	0.000545	0.000817	0.000545	1.000000	1224.0
(4568105, 9238597)	(7617591)	0.000545	0.000817	0.000545	1.000000	1224.0
(2324085)	(304072)	0.000817	0.000545	0.000545	0.666667	1224.0
(9238597, 7617591)	(4568105)	0.000545	0.000817	0.000545	1.000000	1224.0
(2864398)	(632916)	0.000817	0.000545	0.000545	0.666667	1224.0
(632916)	(2864398)	0.000545	0.000817	0.000545	1.000000	1224.0
(4568105)	(9238597, 7617591)	0.000817	0.000545	0.000545	0.666667	1224.0
(9294395)	(9324395)	0.000545	0.001089	0.000545	1.000000	918.0
(9182555)	(8632552)	0.001089	0.000545	0.000545	0.500000	918.0
(3986264)	(456264)	0.000545	0.001089	0.000545	1.000000	918.0
(456264)	(3986264)	0.001089	0.000545	0.000545	0.500000	918.0
(8632552)	(9182555)	0.000545	0.001089	0.000545	1.000000	918.0

(2652631)	(4662914)	0.001089	0.000545	0.000545	0.500000	918.0
(9324395)	(9294395)	0.001089	0.000545	0.000545	0.500000	918.0
(4662914)	(2652631)	0.000545	0.001089	0.000545	1.000000	918.0
(9238597)	(4568105, 7617591)	0.001089	0.000545	0.000545	0.500000	918.0
(4568105, 7617591)	(9238597)	0.000545	0.001089	0.000545	1.000000	918.0
(4568105)	(7617591)	0.000817	0.000817	0.000545	0.666667	816.0
(7617591)	(4568105)	0.000817	0.000817	0.000545	0.666667	816.0
(6337594)	(4537594)	0.000817	0.000817	0.000545	0.666667	816.0
(4537594)	(6337594)	0.000817	0.000817	0.000545	0.666667	816.0
(6554238)	(8224240)	0.000817	0.000817	0.000545	0.666667	816.0
(6684219)	(6034219)	0.000817	0.000817	0.000545	0.666667	816.0
(8224240)	(6554238)	0.000817	0.000817	0.000545	0.666667	816.0
(7046264)	(1814220, 1824220)	0.000817	0.000817	0.000545	0.666667	816.0
(6034219)	(6684219)	0.000817	0.000817	0.000545	0.666667	816.0
(1814220, 1824220)	(7046264)	0.000817	0.000817	0.000545	0.666667	816.0
(6372524)	(5452548)	0.001362	0.000545	0.000545	0.400000	734.4
(5452548)	(6372524)	0.000545	0.001362	0.000545	1.000000	734.4
(1324228)	(6184085)	0.000545	0.001634	0.000545	1.000000	612.0
(6777933)	(4917933)	0.000545	0.001634	0.000545	1.000000	612.0
(4917933)	(6777933)	0.001634	0.000545	0.000545	0.333333	612.0
(1056223)	(9146222)	0.000817	0.001089	0.000545	0.666667	612.0
(9238597)	(7617591)	0.001089	0.000817	0.000545	0.500000	612.0
(9068481)	(5677937)	0.001634	0.000545	0.000545	0.333333	612.0
(1814220)	(7046264, 1824220)	0.001634	0.000545	0.000545	0.333333	612.0
(5677937)	(9068481)	0.000545	0.001634	0.000545	1.000000	612.0
(6184085)	(1324228)	0.001634	0.000545	0.000545	0.333333	612.0
(7046264, 1824220)	(1814220)	0.000545	0.001634	0.000545	1.000000	612.0
(3396339)	(336339)	0.000545	0.001634	0.000545	1.000000	612.0
(336339)	(3396339)	0.001634	0.000545	0.000545	0.333333	612.0
(4568105)	(9238597)	0.000817	0.001089	0.000545	0.666667	612.0
(9238597)	(4568105)	0.001089	0.000817	0.000545	0.500000	612.0
(7617591)	(9238597)	0.000817	0.001089	0.000545	0.666667	612.0

(9146222)	(1056223)	0.001089	0.000817	0.000545	0.500000	612.000000
(7046264, 1814220)	(1824220)	0.000545	0.001906	0.000545	1.000000	524.571429
(1824220)	(7046264, 1814220)	0.001906	0.000545	0.000545	0.285714	524.571429
(4080379)	(8060116)	0.001906	0.000817	0.000817	0.428571	524.571429
(8060116)	(4080379)	0.000817	0.001906	0.000817	1.000000	524.571429
(4822018)	(4862018)	0.000817	0.001362	0.000545	0.666667	489.600000
(4862018)	(4822018)	0.001362	0.000817	0.000545	0.400000	489.600000
(9432506)	(9442506)	0.001089	0.001089	0.000545	0.500000	459.000000
(9442506)	(9432506)	0.001089	0.001089	0.000545	0.500000	459.000000
(9182555)	(8382552)	0.001089	0.001089	0.000545	0.500000	459.000000
(8382552)	(9182555)	0.001089	0.001089	0.000545	0.500000	459.000000
(4256463)	(9296463)	0.001089	0.001089	0.000545	0.500000	459.000000
(9296463)	(4256463)	0.001089	0.001089	0.000545	0.500000	459.000000
(1897917)	(6337897)	0.001089	0.001089	0.000545	0.500000	459.000000
(6337897)	(1897917)	0.001089	0.001089	0.000545	0.500000	459.000000
(7046264)	(1814220)	0.000817	0.001634	0.000545	0.666667	408.000000
(2264226)	(2934275)	0.001634	0.000817	0.000545	0.333333	408.000000
(1814220)	(7046264)	0.001634	0.000817	0.000545	0.333333	408.000000
(2934275)	(2264226)	0.000817	0.001634	0.000545	0.666667	408.000000
(574072)	(2484725)	0.002451	0.000545	0.000545	0.222222	408.000000
(2484725)	(574072)	0.000545	0.002451	0.000545	1.000000	408.000000
(954722)	(5914721)	0.001089	0.001362	0.000545	0.500000	367.200000
(6564238)	(7784238)	0.001362	0.001089	0.000545	0.400000	367.200000
(1824225)	(984225)	0.001362	0.001089	0.000545	0.400000	367.200000
(984225)	(1824225)	0.001089	0.001362	0.000545	0.500000	367.200000
(7784238)	(6564238)	0.001089	0.001362	0.000545	0.500000	367.200000
(5914721)	(954722)	0.001362	0.001089	0.000545	0.400000	367.200000
(1824220)	(7046264)	0.001906	0.000817	0.000545	0.285714	349.714286
(9797919)	(4618105)	0.001906	0.000817	0.000545	0.285714	349.714286
(7046264)	(1824220)	0.000817	0.001906	0.000545	0.666667	349.714286
(4618105)	(9797919)	0.000817	0.001906	0.000545	0.666667	349.714286