Matthew Paras
IEMS 308
2/10/19
Association Rules Report

Executive Summary

Dillard's is a department store chain in the United States with almost 300 locations in 29 states. In this report, we explore the purchasing patterns via association rules at the 5 Dillard's locations located in Iowa, in an attempt to highlight the most promising SKUs used for rearrangement of the store's planograms. Association rule learning is a rule-based machine learning method for discovering interesting relations between variables. In this case, we are applying it to market baskets, where we look at the cooccurrences of products in a transaction in order to see what products are more likely to be purchased together, and if any products imply the purchase of another product. Our results show that customers at Dillard's exhibit a great deal of brand loyalty, very often purchasing multiple different items from the same brand. In particular, luxury cosmetics from brands such as Clinique and Lancome are extremely popular, as well as clothing from brands such as Hue and Kays. In addition, we were able to compile a list of 100 rules linking different SKUs, as well as a list of 44 unique SKUs for potential movement within the Dillard's chain.

Problem Statement

Dillard's is interested in finding 100 SKUs that are the best candidates for making planogram (store layout) changes across the chain. Dillard's can only make at most 20 planogram changes, where a change/move consists of moving one SKU from one location in the store to another. While we are to provide a list of promising SKUs, we also aim to explore the brands associated with those SKUs to provide a more reasonable alternative to just a pure list of SKU numbers.

Assumptions

1. The POS data provided is accurate and that there are no significant flaws with the reporting of the data.
2. Baskets were only selected if there was an item in the top 250 purchased items from the selected locations. This way we were able to select a reasonable number of SKUs to analyze.
3. The selected stores, the stores located in Illinois, have similar sales patterns and can be compared to one another.
4. Returned products are not important in the analysis and are discarded.
5. The original locations of products within the stores are random (given assumption).

Methodology

We first set out to explore the data and understand what the POS data represented. In order to do this, the data was loaded into a SQL database, where queries could be performed, and the tables could be further understood. The 5 tables provided included information about transactions, stores, skus, and departments. The transaction table proved to be of key interest, with over 120 million lines of sales information. Using the provided schema guide, the columns of the tables were labeled, and data sub-setting could then be performed. In order to properly perform the association rules analysis, a smaller subset of the data needed to be selected. We decided to analyze the sales data from the 5 stores located in Iowa. We also removed the transactions that were not purchases (i.e. removed return data) since we are mostly interested in what items are purchased together.

The transaction data was presented in such a way that each row corresponded to one SKU. In order to perform the analysis, we needed to map each individual item sold to the transaction and basket that it occurred in. In order to do this, we created a unique identifier by concatenating the store id, the register id, the transaction id, and the sale data into one column. From here, we then extracted the data from SQL and continued molding the data for analysis.

In order to select a reasonable number of baskets and SKU, only baskets containing the most popular 250 items were selected. The reasoning behind this is that we can then analyze the associations with popular items, which is what Dillard's is most interested in. It also reduces the number of baskets and SKUs to a reasonable amount such that the analysis can even be performed. Once the list of baskets containing at least one of the most popular SKUs was selected, the baskets were one hot encoded with the SKU numbers that they contained.

Using the mlxtend package, association rules were generated with a minimum lift value of 15. Many of the rules have extremely high lift values, suggesting the significance of these rules. In order to understand more about the rules, the SKU numbers were mapped to the associated brands. For more information regarding the data exploration, data analysis, and list of rules, see the accompanying jupyter notebook (Association Rules.ipynb). Full details regarding the data set size and information can be found there as well.

Analysis and Conclusions

In these association rules, we find that the same brands are purchased together almost exclusively. These brands include:

- Clinique – skincare, cosmetics
- Lancome – luxury cosmetics
- Hue – clothing
- Hue/Kays – clothing

An interesting trend is that these brands are purchased extremely often – and out of the 44 unique SKUs present in the 100 different rules, 26 of those SKUs are from Clinique. Here we can see the rules in Figures 1 and 2:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	antecedents_names	consequents_names
60	(7218011)	(7248011)	0.000859	0.000979	0.000412	0.480000	490.324211	0.000411	1.921194	[CLINIQUE]	[CLINIQUE]
61	(7248011)	(7218011)	0.000979	0.000859	0.000412	0.421053	490.324211	0.000411	1.725789	[CLINIQUE]	[CLINIQUE]
8	(5872825)	(1446155)	0.002559	0.000550	0.000532	0.208054	378.566695	0.000531	1.262018	[HUE]	[HUE]
9	(1446155)	(5872825)	0.000550	0.002559	0.000532	0.968750	378.566695	0.000531	31.918112	[HUE]	[HUE]
11	(2266446)	(7351914)	0.000756	0.003710	0.000756	1.000000	269.564815	0.000753	inf	[LANCOME]	[LANCOME]
10	(7351914)	(2266446)	0.003710	0.000756	0.000756	0.203704	269.564815	0.000753	1.254865	[LANCOME]	[LANCOME]
34	(4472217)	(7351914)	0.000635	0.003710	0.000618	0.972973	262.279279	0.000616	36.862742	[LANCOME]	[LANCOME]
35	(7351914)	(4472217)	0.003710	0.000635	0.000618	0.166667	262.279279	0.000616	1.199237	[LANCOME]	[LANCOME]
13	(2276446)	(7351914)	0.000584	0.003710	0.000567	0.970588	261.636438	0.000565	33.873871	[LANCOME]	[LANCOME]
12	(7351914)	(2276446)	0.003710	0.000584	0.000567	0.152778	261.636438	0.000565	1.179639	[LANCOME]	[LANCOME]
36	(4552217)	(7351914)	0.000550	0.003710	0.000532	0.968750	261.140914	0.000530	31.881290	[LANCOME]	[LANCOME]
37	(7351914)	(4552217)	0.003710	0.000550	0.000532	0.143519	261.140914	0.000530	1.166926	[LANCOME]	[LANCOME]
16	(7351914)	(2406446)	0.003710	0.000429	0.000412	0.111111	258.782222	0.000411	1.124517	[LANCOME]	[LANCOME]
17	(2406446)	(7351914)	0.000429	0.003710	0.000412	0.960000	258.782222	0.000411	24.907258	[LANCOME]	[LANCOME]
5	(270789)	(7351914)	0.000773	0.003710	0.000704	0.911111	245.603498	0.000701	11.208266	[LANCOME]	[LANCOME]
4	(7351914)	(270789)	0.003710	0.000773	0.000704	0.189815	245.603498	0.000701	1.233332	[LANCOME]	[LANCOME]
42	(5453386)	(7218011)	0.003589	0.000859	0.000515	0.143541	167.155981	0.000512	1.166595	[CLINIQUE]	[CLINIQUE]
43	(7218011)	(5453386)	0.000859	0.003589	0.000515	0.600000	167.155981	0.000512	2.491026	[CLINIQUE]	[CLINIQUE]
62	(7636135)	(7596135)	0.000739	0.004483	0.000515	0.697674	155.642876	0.000512	3.292865	[HUE/KAYS]	[HUE/KAYS]
63	(7596135)	(7636135)	0.004483	0.000739	0.000515	0.114943	155.642876	0.000512	1.129036	[HUE/KAYS]	[HUE/KAYS]
14	(3663723)	(2313723)	0.000584	0.005685	0.000464	0.794118	139.692731	0.000460	4.829531	[ROCHESTE]	[ROCHESTE]
15	(2313723)	(3663723)	0.005685	0.000584	0.000464	0.081571	139.692731	0.000460	1.088180	[ROCHESTE]	[ROCHESTE]
49	(6656135)	(6706135)	0.006887	0.000979	0.000859	0.124688	127.370171	0.000852	1.141332	[HUE/KAYS]	[HUE/KAYS]
48	(6706135)	(6656135)	0.000979	0.006887	0.000859	0.877193	127.370171	0.000852	8.086778	[HUE/KAYS]	[HUE/KAYS]
45	(7248011)	(5453386)	0.000979	0.003589	0.000429	0.438596	122.190044	0.000426	1.774856	[CLINIQUE]	[CLINIQUE]
44	(5453386)	(7248011)	0.003589	0.000979	0.000429	0.119617	122.190044	0.000426	1.134758	[CLINIQUE]	[CLINIQUE]
46	(6696135)	(6656135)	0.001030	0.006887	0.000824	0.800000	116.161596	0.000817	4.965565	[HUE/KAYS]	[HUE/KAYS]
47	(6656135)	(6696135)	0.006887	0.001030	0.000824	0.119701	116.161596	0.000817	1.134807	[HUE/KAYS]	[HUE/KAYS]
55	(7596135)	(6696135)	0.004483	0.001030	0.000515	0.114943	111.544061	0.000511	1.128706	[HUE/KAYS]	[HUE/KAYS]
54	(6696135)	(7596135)	0.001030	0.004483	0.000515	0.500000	111.544061	0.000511	1.991035	[HUE/KAYS]	[HUE/KAYS]

Figure 1: Association Rules with SKUs and associated brands (part 1)

2	(108507)	(1400555)	0.004568	0.003744	0.000515	0.112782	30.123129	0.000498	1.122899	[CLINIQUE]	[CLINIQUE]
3	(1400555)	(108507)	0.003744	0.004568	0.000515	0.137615	30.123129	0.000498	1.154277	[CLINIQUE]	[CLINIQUE]
90	(3968011, 3898011)	(3690654)	0.003504	0.014032	0.001340	0.382353	27.249550	0.001290	1.596330	[CLINIQUE , CLINIQUE]	[CLINIQUE]
95	(3690654)	(3968011, 3898011)	0.014032	0.003504	0.001340	0.095471	27.249550	0.001290	1.101675	[CLINIQUE]	[CLINIQUE , CLINIQUE]
19	(3908011)	(2716578)	0.005616	0.009034	0.001305	0.232416	25.727468	0.001255	1.291020	[CLINIQUE]	[CLINIQUE]
18	(2716578)	(3908011)	0.009034	0.005616	0.001305	0.144487	25.727468	0.001255	1.162324	[CLINIQUE]	[CLINIQUE]
88	(3978011)	(3524026, 3898011)	0.030244	0.002439	0.001855	0.061329	25.147395	0.001781	1.062738	[CLINIQUE]	[CLINIQUE , CLINIQUE]
87	(3524026, 3898011)	(3978011)	0.002439	0.030244	0.001855	0.760563	25.147395	0.001781	4.050156	[CLINIQUE , CLINIQUE]	[CLINIQUE]
21	(3988011)	(2716578)	0.007900	0.009034	0.001666	0.210870	23.342379	0.001595	1.255770	[CLINIQUE]	[CLINIQUE]
20	(2716578)	(3988011)	0.009034	0.007900	0.001666	0.184411	23.342379	0.001595	1.216421	[CLINIQUE]	[CLINIQUE]
23	(3908011)	(2726578)	0.005616	0.015697	0.001992	0.354740	22.598572	0.001904	1.525436	[CLINIQUE]	[CLINIQUE]
22	(2726578)	(3908011)	0.015697	0.005616	0.001992	0.126915	22.598572	0.001904	1.138931	[CLINIQUE]	[CLINIQUE]
32	(3998011)	(3908011)	0.014547	0.005616	0.001838	0.126328	22.494149	0.001756	1.138166	[CLINIQUE]	[CLINIQUE]
33	(3908011)	(3998011)	0.005616	0.014547	0.001838	0.327217	22.494149	0.001756	1.464742	[CLINIQUE]	[CLINIQUE]
7	(348498)	(3582465)	0.005908	0.004654	0.000618	0.104651	22.484940	0.000591	1.111685	[CLINIQUE]	[CLINIQUE]
6	(3582465)	(348498)	0.004654	0.005908	0.000618	0.132841	22.484940	0.000591	1.146378	[CLINIQUE]	[CLINIQUE]
24	(3631365)	(3581365)	0.007574	0.003263	0.000532	0.070295	21.542022	0.000508	1.072100	[CLINIQUE]	[CLINIQUE]
25	(3581365)	(3631365)	0.003263	0.007574	0.000532	0.163158	21.542022	0.000508	1.185918	[CLINIQUE]	[CLINIQUE]
97	(4108011, 3898011)	(3978011)	0.000859	0.030244	0.000550	0.640000	21.161068	0.000524	2.693766	[CLINIQUE , CLINIQUE]	[CLINIQUE]
98	(3978011)	(4108011, 3898011)	0.030244	0.000859	0.000550	0.018171	21.161068	0.000524	1.017633	[CLINIQUE]	[CLINIQUE , CLINIQUE]
27	(3581365)	(3968356)	0.003263	0.007608	0.000498	0.152632	20.061233	0.000473	1.171146	[CLINIQUE]	[CLINIQUE]
26	(3968356)	(3581365)	0.007608	0.003263	0.000498	0.065463	20.061233	0.000473	1.066557	[CLINIQUE]	[CLINIQUE]
77	(2726578, 3898011)	(3978011)	0.001357	0.030244	0.000790	0.582278	19.252554	0.000749	2.321537	[CLINIQUE , CLINIQUE]	[CLINIQUE]
78	(3978011)	(2726578, 3898011)	0.030244	0.001357	0.000790	0.026122	19.252554	0.000749	1.025429	[CLINIQUE]	[CLINIQUE , CLINIQUE]
96	(3978011, 4108011)	(3898011)	0.002559	0.014135	0.000550	0.214765	15.194305	0.000513	1.255504	[CLINIQUE , CLINIQUE]	[CLINIQUE]
99	(3898011)	(3978011, 4108011)	0.014135	0.002559	0.000550	0.038882	15.194305	0.000513	1.037793	[CLINIQUE]	[CLINIQUE , CLINIQUE]
67	(2716578)	(3968011, 3898011)	0.009034	0.003504	0.000481	0.053232	15.193544	0.000449	1.052524	[CLINIQUE]	[CLINIQUE , CLINIQUE]
66	(3968011, 3898011)	(2716578)	0.003504	0.009034	0.000481	0.137255	15.193544	0.000449	1.148620	[CLINIQUE , CLINIQUE]	[CLINIQUE]
28	(3968011)	(3898011)	0.016367	0.014135	0.003504	0.214061	15.144481	0.003272	1.254379	[CLINIQUE]	[CLINIQUE]
29	(3898011)	(3968011)	0.014135	0.016367	0.003504	0.247874	15.144481	0.003272	1.307802	[CLINIQUE]	[CLINIQUE]

Figure 1: Association Rules with SKUs and associated brands (part 2)

From these rules we can establish both that health and beauty products should be placed together, with brands like Clinique and Lancome taking up much of the space, since these products are very popular. They both have many unique products within each brand, and care should be taken it to place as many products from each brand together as possible. In addition, Hue and Kays are brands of clothing that complement each other well, and customers are loyal to purchasing multiple items of these brands together. These association rules have high lift values as well, which suggest that they are powerful indicators of purchasing patterns. Many also have high confidence values (many near the top have close to values of 1). The full list of SKU numbers can be found in rules_df.csv, in the antecedents and consequents columns.

Recommendations and Next Steps

- Dillard's should focus on creating a cosmetics department that highlights all of the Clinique and Lancome products. While we did not explicitly find rules that suggest customers will purchase both brands, customers are inclined to purchase multiple products from one of the brands.
- Clinique products should be placed near each other, since customers are inclined to purchase multiple Clinique items at a time.
- Lancome products should also be placed near each other.
- Dillard's should also highlight brands such as Hue and Key in their clothing department, in an effort to promote sales of different types of clothing.

Further analysis should be taken on this data in different regions of the country. It is possible that there are certain regional sales patterns that can be exploited. In addition, it would be pertinent to select a different subset of items (not just baskets that include items in the top 250 purchased) in order to see items that are more expensive and less purchased (e.g. expensive designer clothing, high fashion items).