Aneesh Kudaravalli
IEMS 308 HW 2
Association Rules

Item Purchase Association Rules at Dillard's

Business Question:

Our goal is to find the 100 target SKU's that's purchase depends strongly on purchasing of other items in the store. Dillard's can strategically place and price the item dependending on the associated SKUs place and price to incentivize increased consumer spending.

Methodology:

Performing association rules on the most profitable items in customer baskets to understand what items are bought together. Strong associations are rated based on their support (how frequently the items appear in the data), confidence (how often the rule or association has been found to be true), and lift (the dependency of the items being bought together). We used the dataset given to us by Dillard's.

Exploration and Pre Processing:

Association rules perform best on datasets of smaller sizes, so one of our priorities when exploring the data was to understand a reasonable way to subset it (including too many SKUs or transactions creates a combinatorial explosion). We decided on first subsetting the data to just one store.

Considering the associations are dependent on store layout, this is a good way to understand associations of items in just one store, and the rules found from this one store can be tested on other stores to see if they remain true across other Dillard's. To pick a target store, we identified the top stores in terms of their total costs of SKUs from the SKSTINFO table. We described these stores further with their location from the STRINFO table and (Appendix A). *We selected the transactions (from TRNSACT table) of the highest cost store located in Overland Park, Kansas to perform association rules.* Although this store has the highest costs, it does not have the highest retail values, so we think it's a good place to start in terms of driving more profits. Note: We also checked the data from the SKSTINFO table for outliers (as it was unclear if data was recorded by a machine or not), but the costs and retail values for items were proportional and reasonable further supporting our subsetting method (Appendix B).

To group each individual transaction on a SKU from the Kansas store into one customer basket we grouped the data by the combination of the register number and transaction number. Each SKU was turned into a binary categorical variable indicating if that SKU appeared in the customer. At this point there did not appear to be any outliers or missing values in the transaction data due to the fact they were recorded by a computer system.

Model Fitting:

Initial attempts at using all the SKUs sold in the Overland Park store to create association rules created combinatorial explosions. We then tried subsetting the transactions to only include SKUs that appeared 500 or more times in the transaction data. Although this allowed us to perform association rules, we were missing infrequently purchased items that could still be valuable.

To understand which specific SKUs would be most valuable to the Overland Park store, we multiplied the amount of times a SKU appeared in the transaction data in Overland Park by the expected profit of that SKU (retail - cost from Kansas SKSTINFO). We used the top 500 of these valuable SKUs to be the SKUs that association rules were performed on. This allowed us to understand relationships between purchases of valuable products, and we still didn't miss rare items as the frequency of these valuable SKUs appearing in the transaction data ranged from 1 to 1853 (Appendix C).

We started with performing associations of max length 2 (or 1 to 1 dependencies of products purchased) to avoid too many combinations. This returned around 31,000 rules that had a minimum support of .01 and a minimum lift of 1. Applying the minimum item support rule to ensure that the support of a rule is greater than the support of either its antecedent purchase or consequent purchase individually resulted in only 2 valid rules (Appendix D). We only had information on 2 of the 4 items that appear in these rules, and they both are make-up products.

To increase the amount of our associations, we applied a new association rules algorithm with max length of 3 (1 to 2 purchase dependencies or 2 to 1 purchase dependencies). This returned around 2.5 million rules with a minimum support of .01 and a minimum lift of 1. We modified the minimum item support algorithm (to account for inevitably low rule support when working with large datasets) to only include rules that have support that is either greater than the support for their antecedent purchase(s) or greater than the support for their consequent purchases(s). This returned around 12.5 thousand rules that had reasonable support values (Appendix E).

To subset these rules to the items that have the strongest dependencies on another item we only viewed rules that had a lift of at least 22 (meaning the combination of these high value items are found in baskets about 22 times more often than expected if the items were bought independently) and a confidence of at least .25 (meaning that percentage of the time the antecedent purchase(s) appear in the basket, the consequent purchase(s) also appear in that basket). Most of the confidence levels on rules that had lifts of at least 22 were 1. This resulted in 212 rules that showed strong dependencies between high value items purchased at the Overland Park Dillard's (Appendix F).

Since many of the rules were simply opposites of each other, these rules were removed to further isolate SKU's. The resulting 130 rules and 56 unique SKUs are attached in separate csv files (StrongRules.csv and DependentSKUS.csv). These rules can be used by the store to either price or place these products to incentivize further purchasing. It also allows us to comfortably answer the question of 20 ideal moves that can be made in the planogram to incentivize spending.

Analysis:

To rank the 20 best SKUs from all the dependencies, we used the same most valuable SKUs ranking system that we initially used to cut down to 500 SKUs. The resulting 20 best SKUs are saved and sorted by their value in the attached top20SKUs.csv file. These 20 best SKUs appear as the consequents of 124 rules which are saved in the attached targ20associatedrules.csv file. Dillard's can use this information to potentially put "sales" on antecedents of these 124 rules and then raise the price of the consequents which are the most valuable SKUs.

It should be noted at this point that further analysis was limited due to the SKUINFO table missing information for many of the SKUs that were included in our association rules. However, we found information on 35 of the 56 unique SKUs, and it turns out they are all Lancome makeup products (Appendix G). We recommend that Dillard's do further analysis onto what type of products are the remaining SKUs that we don't have info on.

With the products that we identified as Lancome makeup products, the SKU info given to us was not enough to determine the specific type of makeup products that are contained in these associations (for instance maybe Lancome sells suites of products designed to be bought together), so we recommend Dillard's also do further analysis on how these SKUs relate to each other in terms of type of makeup.

These rules were also tested on the store with the second biggest costs located in Metaire, Louisiana (Appendix A). We followed the same process of subsetting and grouping the transaction data to baskets from that store. We then categorically defined whether or not the top 500 most profitable SKUs were in a customer basket. We performed association rules with a similar minimum support of 0.1 and a minimum lift of 1. An examination of the top rules ranked according to the same lift, confidence, and support values from the Kansas store resulted in a few rules that showed similar strong dependencies between items (Appendix H). The SKUs contained in these rules were individually examined, and we similarly discovered they were mostly made of makeup products, this time including the Clinique brand (Appendix H)

However, knowing that the majority of valuable SKUs that are highly dependent on other valuable SKUs are makeup products is a valuable insight. It tells us that the beauty section contains some of Dillard's most valuable products, and the dependencies within the products in the beauty section can be leveraged as a sales growth driver. One strategy mentioned earlier was raising prices on consequents and putting sales or false sales on the antecedents of the associations. This may subconsciously get the consumer to be willing to spend more on the consequent. Another strategy is to have sales on makeup products that are in the same suite, so one buys all the items of a bundle of beauty products to incentivize larger purchases.

In terms of placement, it would be inconvenient for a consumer to have to walk through different sections of the store to find all the beauty products they need. Rather, it makes more sense for Dillard's to have a dedicated beauty section perhaps far away from the entrance (so the customer has to walk through much of the store to find it). We recommend organizing the beauty section by brand putting products in the same suite together and creating an immersive, fun beauty experience around the Lancome brand of products. This would create a growth driver around their most valuable SKUs and it would also allow Dillard's to compete with other beauty stores who are employing a strategy of experiential retail. Furthermore, either pricing strategy has the potential to work with this design.

Conclusion/Recommendations:

- Association rules conducted on the most valuable products sold in the Overland Park store showed that the purchase of Lancome and other beauty products tend to be highly dependent on the purchase of some specific other items.

- Top 20 SKUs and the purchases of which they are likely dependent on are shared in the attached files (top20SKUs.csv and targ20associatedrules.csv)

- We recommend Dillard’s do further analysis on the SKUs that we’re missing information on, but are still in top association rules

- We tested these rules and similar conclusions were shown in the Metaire store, but we recommend further validating the rules on other large Dillard’s stores.

- We recommend very strategically designing the beauty sections of Dillard’s so that they are in the center of the store, so that they encourage immersive beauty experiences, and so that the beauty products that are purchase dependent on each other are strategically priced to match Dillard’s goals.

- We believe this is the most valuable use of the manpower to make 20 moves across the store, as these beauty products are the biggest profit driver for Dillard’s.

Appendix A Top Stores According to Costs of all SKU's in SKSTINFO table and the location of these stores from the STRINFO table

STORE	SKU	COST	RETAIL
2203	529636267213	3310780.46	5.597186e+06
8402	510038991471	3204057.65	5.794243e+06
209	471654699461	2946128.77	5.465774e+06
7507	405954119197	2913697.15	5.575492e+06

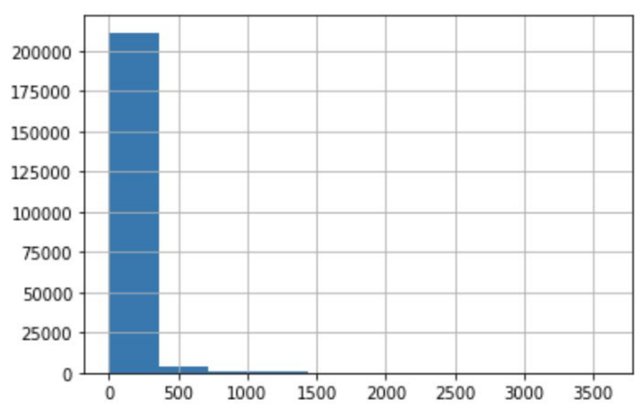
	STORE	**CITY**	**STATE**	**ZIP**
21	209	SCOTTSDALE	AZ	85251
38	504	LITTLE ROCK	AR	72205
109	2102	ORLANDO	FL	32809
113	2109	ALBUQUERQUE	NM	87110

116	2203	OVERLAND PARK	KS	66214
348	7507	HOUSTON	TX	77056
376	8109	LITTLETON	CO	80124
385	8402	METAIRIE	LA	70002
416	9103	LOUISVILLE	KY	40207
426	9304	OKLAHOMA CITY	OK	7311

Appendix B Outlier Check on Kansas SKSTINFO

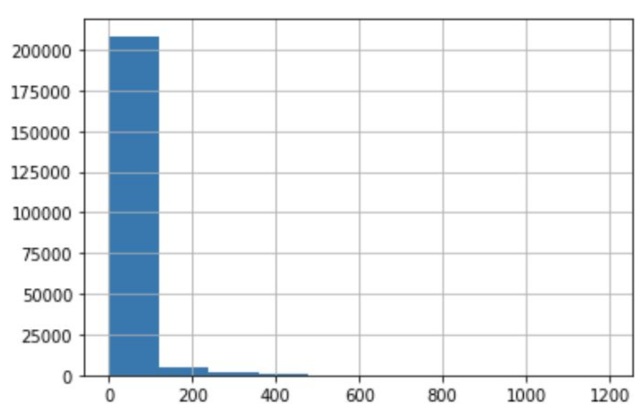
```
skstKS.RETAIL.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1154066d0>
```




```
skstKS.COST.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x174526990>
```



SKU	STORE	COST	RETAIL	EXTRA	
4727553	1200149	2203	914.10	3300.0	2385.90
18043622	4530148	2203	1048.68	3360.0	2311.32
18919867	4750148	2203	1048.68	3360.0	2311.32
24290345	6100148	2203	1071.96	3360.0	2288.04
30007239	7590148	2203	1193.88	3600.0	2406.12
37509438	9560148	2203	899.82	3060	

Appendix C Top 500 Valuable SKUs and their counts in Kansas transaction data

3524026	1853
5528349	1846
3978011	1839
2698353	1292
3161221	1070
...	
258604	2
2877851	2
5397432	2
9397666	2
1776023	1

Name: SKU, Length: 500

Appendix D Minimum Item Support applied to 1 to 1 Association Rules and SKU information from these rules

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
8210	(SKU_1699998)	(SKU_3524026)	0.014390	0.141220	0.01439	1.000000	7.081174	0.012358	inf
8211	(SKU_3524026)	(SKU_1699998)	0.141220	0.014390	0.01439	0.101900	7.081174	0.012358	1.097439
28116	(SKU_5528349)	(SKU_5901685)	0.072927	0.010000	0.01000	0.137124	13.712375	0.009271	1.147326
28117	(SKU_5901685)	(SKU_5528349)	0.010000	0.072927	0.01000	1.000000	13.712375	0.009271	inf

	SKU	DEPT	CLASSID	UPC	STYLE	COLOR	SIZE	PACKSIZE	VENDOR	BRAND
285254	3524026	800	001	4.000040e+11	68LE	DDML PUMP	4.2 OZ	6	5511283	CLINIQUE
437598	5528349	2200	006	4.000083e+11	2410	01-BLACK	01-BLACK	3	0113645	LANCOM

Appendix E Modified Minimum Support Rule to subset new Association rules length (3)

```
rules[(rules['support'] >= rules['antecedent support']) | (rules['support'] >= rules['consequent support'])].de
```

	antecedent support	consequent support	support	confidence	lift	leverage	conviction
count	12554.000000	12554.000000	12554.000000	12554.000000	12554.000000	12554.000000	1.255400e+04
mean	0.064628	0.064628	0.012548	0.557305	9.262350	0.011068	inf
std	0.055427	0.055427	0.002425	0.443599	3.297688	0.002068	NaN
min	0.010000	0.010000	0.010000	0.070812	7.081174	0.008588	1.065446e+00
25%	0.011951	0.011951	0.010732	0.106538	7.081174	0.009530	1.104510e+00
50%	0.031951	0.031951	0.011951	0.702797	7.592593	0.010528	inf
75%	0.131707	0.131707	0.013659	1.000000	10.732984	0.012071	inf
max	0.141220	0.141220	0.030976	1.000000	30.370370	0.026601	inf

Appendix F Summary Statistics on Target Rules (good lift, support, and confidence)

	antecedent support	consequent support	support	confidence	lift	leverage	conviction
count	212.000000	212.000000	212.000000	212.000000	212.000000	212.000000	212.000000
mean	0.023012	0.029419	0.011269	0.710003	24.506008	0.010804	inf
std	0.014633	0.015147	0.001189	0.352425	2.360518	0.001131	NaN
min	0.010000	0.010000	0.010000	0.251397	22.043011	0.009546	1.321159
25%	0.010488	0.011707	0.010244	0.284879	22.905028	0.009867	1.380886
50%	0.012195	0.036829	0.010976	1.000000	23.727963	0.010529	inf
75%	0.040732	0.043659	0.011951	1.000000	27.152318	0.011429	inf
max	0.045366	0.045366	0.017561	1.000000	30.370370	0.016764	inf

Appendix G Target Association Rules SKU Info

	SKU	DEPT	CLASSID	UPC	STYLE	COLOR	SIZE	PACKSIZE	VENDOR	BRAND	EXTRA	EXTRA	EXTRA
61868	397046	2200	1	4.000070e+11	2442	REN LIFT	*DISC	3	113645	LANCOME	0	NaN	NaN
69885	448103	2200	6	4.000080e+11	2416	NUDE	EYE BASE	3	113645	LANCOME	0	NaN	NaN
102542	656219	2200	1	4.000060e+11	3251	ABSOLUE LTN	ABS LTN	3	113645	LANCOME	0	NaN	NaN
105493	674109	2200	6	4.000040e+11	2411	11-ICECARO	ARTLINER	3	113645	LANCOME	0	NaN	NaN
121048	776350	2200	6	4.000060e+11	8864	01-BLACK	FLEXTEN	3	113645	LANCOME	0	NaN	NaN
166861	1069332	2200	1	4.000090e+11	4798	NOR/C/RESOLU	1.7 OUNCES	3	113645	LANCOME	0	NaN	NaN
185102	1184024	2200	5	4.000040e+11	2310	18-MIEL GLAC	BLUSH SUB	3	113645	LANCOME	0	NaN	NaN
208885	1337310	2200	8	4.000070e+11	6367	00-HOLIDAY	NO SIZE	6	113645	LANCOME	0	NaN	NaN
215704	1380638	2200	7	4.000010e+11	2704	08-BROWNIE	ONE	3	113645	LANCOME	0	NaN	NaN
247334	1583514	2200	6	4.000040e+11	2137	08-NATURAL	2137-08	3	113645	LANCOME	0	NaN	NaN
306264	1968367	2200	4	4.000080e+11	2127	03 CLAIRE II	CONCEALOR	3	113645	LANCOME	0	NaN	NaN
337940	2170445	2200	9	4.000000e+11	9552	00-S'05	GROOVE PWP	12	113645	LANCOME	0	NaN	NaN
426906	2734854	2200	4	4.000050e+11	2105	07IVOIRE	MAQUICOMPL	3	113645	LANCOME	0	NaN	NaN
428467	2744854	2200	4	4.000050e+11	2105	08LIGHT BUFF	MAQUICOMPL	3	113645	LANCOME	0	NaN	NaN
430022	2754854	2200	4	4.000050e+11	2105	09CAMEE	MAQUICOMPL	3	113645	LANCOME	0	NaN	NaN
466577	2988370	2200	1	4.000080e+11	1447	RENERGIE	EYE CREAM	3	113645	LANCOME	0	NaN	NaN
520877	3332423	2200	7	4.000020e+11	2534	07-BRONZELLE	LE LIP	3	113645	LANCOME	0	NaN	NaN
552642	3537981	2200	4	4.000080e+11	589	FOUND #2	BRUSH	3	113645	LANCOME	0	NaN	NaN
554197	3547981	2200	4	4.000080e+11	591	CONCEAL #8	BRUSH	1	113645	LANCOME	0	NaN	NaN
574049	3672270	2200	1	4.000020e+11	3687	MOUSSE	CLARTE	3	113645	LANCOME	0	NaN	NaN
588777	3763770	2200	5	4.000040e+11	4512	37-PINK POOL	SHIMMER	3	113645	LANCOME	0	NaN	NaN

Appendix H Louisiana Test Data Top Ranked Rules and some sample SKU info

In [434]:
```
testrules[((testrules['support'] >= testrules['antecedent support']) | (testrules['support'] >= testrules['consequent s
```

Out[434]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
588532	(SKU_4697737, SKU_593921)	(SKU_2938210)	0.010727	0.045292	0.010727	1.000000	22.078947	0.010241	inf
1141289	(SKU_2170445, SKU_2744854)	(SKU_7567468)	0.012157	0.036234	0.012157	1.000000	27.598684	0.011717	inf
1141292	(SKU_7567468)	(SKU_2170445, SKU_2744854)	0.036234	0.012157	0.012157	0.335526	27.598684	0.011717	1.486654
1141715	(SKU_4674761, SKU_2170445)	(SKU_7567468)	0.010489	0.036234	0.010489	1.000000	27.598684	0.010109	inf
1141718	(SKU_7567468)	(SKU_4674761, SKU_2170445)	0.036234	0.010489	0.010489	0.289474	27.598684	0.010109	1.392646
1614815	(SKU_6678353, SKU_2734854)	(SKU_7567468)	0.011681	0.036234	0.011681	1.000000	27.598684	0.011257	inf
1614818	(SKU_7567468)	(SKU_6678353, SKU_2734854)	0.036234	0.011681	0.011681	0.322368	27.598684	0.011257	1.458491

In [435]:
```
skuuu[skuuu['SKU'] == 2938210]
```
Out[435]:

	SKU	DEPT	CLASSID	UPC	STYLE	COLOR	SIZE	PACKSIZE	VENDOR	BRAND	EXTRA	EXTRA	EXTRA
458600	2938210	800	7	4.000080e+11	6FCR80	HOLIDAY PWP	HOL04 PWP		36	5511283	CLINIQUE	0	NaN	NaN

In [437]:
```
skuuu[skuuu['SKU'] == 2744854]
```
Out[437]:

	SKU	DEPT	CLASSID	UPC	STYLE	COLOR	SIZE	PACKSIZE	VENDOR	BRAND	EXTRA	EXTRA	EXTRA
428467	2744854	2200	4	4.000050e+11	2105	08LIGHT BUFF	MAQUICOMPL	3	113645	LANCOME	0	NaN	NaN

In [439]:
```
skuuu[skuuu['SKU'] == 7567468]
```