Steffany Bahamon

IEMS 308, F16

Association Rules Homework

Executive Summary

Our objective was to locate the 100 best candidate store keeping units to place together in a Dillard's store. Given a database of information containing transactional information, we pared the data down to a table of selected baskets of transactions based on a list of assumptions enumerated below and performed market basket analysis on the set of transactions. The table of association rules tying candidates in the Dillard's store together is in the appendix.

Problem Statement

Dillard's is a major retail chain with several stores interested in arranging the floor of their stores; this is also referred to as changing the planograms. Dillard's can only make at most 20 planogram changes across the entire chain due to budgetary reasons. A move consists of moving one stock keeping unit, also referred to as SKU, to a different position. We aim to find the 100 SKUs that are best candidates to move by analyzing Dillard's point-of-sales (POS data mainly by performing market basket analysis.

Assumptions

In order to simplify our work, we made a series of assumptions:

- The last month of data is representative of every month of data; this is a simplification for computation purposes, but doesn't reflect seasonal changes such as buying more swim gear in warm weather or buying more things overall during the winter holiday shopping season.
- Returns give negligible benefit to proving association
- A random set of data will give us the same results as the whole set.
- Support of 500 and confidence level of 20% are sufficient for this large data set

Methodology

In order to first analyze the data, we first had to come up with a actionable data set of transactions from the POS Data. We did this by first removing all transactions from the transaction data that were returns; though this would be interesting to analyze to see which pairs are usually returned together, for the sale of planogram changes it's irrelevant. In addition, we paired down the data set to only a random set of data from the latest month in the dataset in for increased speed of computation. Afterwards, we created SKU baskets of transactions by grouping SKUs together based off of store, register, and date similarity. Lastly, we implemented the fpgrowth algorithm via the Orange-Associates package in Python to tell us which of the SKUS are best paired together.

Analysis

There were over 120 million different transactions performed over the course of a year at all Dillard's brick and mortar stores. This data was collected from more than 450 stores all across the United States, with most of the stores clustered in Southern states. We collected 143,190 different baskets across our dataset, with 40,995 different SKUs. The largest list we found was 128 SKUs, and the smallest list 1 SKU. The average size of the baskets was found to be around 8 SKUs. In order to calculate the frequent item sets for such a large file, we set the minimum number of appearances at 500. To calculate the association rules, we set the minimum confidence level at 20%; with a large dataset, we figured 20% would still give us a good list of rules while not penalizing items that were bought less frequently.

Of the 10 association rules with the highest confidence percentage, 8 were rules that showed SKU 803921 on the right hand side, indicating that 8034921 might be a good candidate to have near the front of the store since so many customers buy it so often. Most of the sets with a high confidence percentage were sets with 2 SKUS implying 1 SKU, which likely means that consumers buy these SKUS as a group. Lastly, support and in confidence did not seem to be highly correlated, which is interesting considering that confidence is partly derived from support.

Conclusions

We found some great candidate SKUs to be moved around, and obtained a much greater understanding of which SKUS are popular and which are not. Despite the drawbacks inherent with narrowing data analyzed, we were able to obtain actionable insights. In particular, recurring SKUs outlined in the table in the appendix are prime candidates for repositioning inside of the Dillard stores.

Next Steps

The most obvious next steps are to refactor the data we used to obtain more accurate and better baskets. Including the transaction number in the key, having a large initial data set, and including more dates would all increase the sample space from which we could draw association rules. In addition, analysing which items are returned together might also give insight into consumer purchasing patterns and prevent items from being sold that will be returned.

APPENDIX:

Table 1: Association rules with support > 500 and confidence over 20%

LHS	RHS	Support	Confidence
{348498, 264715, 726718}	{803921}	532	0.716981
{39633, 264715}	{803921}	565	0.709799
{29633, 264715}	{803921}	537	0.699219
{264715, 108507}	{803921}	606	0.694954
{566969, 726718}	{803921}	566	0.681107
{108507, 726718}	{803921}	747	0.678474
{39633, 726718}	{803921}	690	0.675147
{803921, 348498, 264715}	{726718}	532	0.673418
{593921, 726718}	{803921}	579	0.66782
{348498, 264715}	{803921}	790	0.66723
{29633, 726718}	{803921}	670	0.666004
{270896, 264715}	{803921}	514	0.664942
{348498, 726718}	{803921}	919	0.650389

{264715, 726718}	{803921}	1339	0.649055
{264715, 108507}	{726718}	563	0.645642
{250896, 726718}	{803921}	547	0.642773
{208362, 726718}	{803921}	550	0.641026
{270896, 726718}	{803921}	646	0.63771
{39633, 264715}	{726718}	503	0.63191
{348498, 264715}	{726718}	742	0.626689
{803921, 108507}	{726718}	747	0.62406
{803921, 566969}	{726718}	566	0.619934
{250896, 803921}	{726718}	547	0.619479
{803921, 39633}	{726718}	690	0.610619
{803921, 29633}	{726718}	670	0.6102
{726718, 106343}	{803921}	547	0.608454
{803921, 348498}	{726718}	919	0.608206
{803921, 593921}	{726718}	579	0.601871
{270896, 803921}	{726718}	646	0.59375
{803921, 208362}	{726718}	550	0.583245
{803921, 264715}	{726718}	1339	0.580659
{803921, 348498, 726718}	{264715}	532	0.57889
{803921, 106343}	{726718}	547	0.578836
{108507}	{803921}	1197	0.575758
{39171}	{803921}	548	0.57563
{539951}	{803921}	601	0.571293
{68086}	{803921}	556	0.570256
{39633}	{803921}	1130	0.566416

{19633}	{803921}	717	0.565904
{566969}	{803921}	913	0.562885
{9633}	{803921}	633	0.562167
{348498}	{803921}	1511	0.561919
{593921}	{803921}	962	0.561916
{318506}	{803921}	580	0.558767
{29633}	{803921}	1098	0.552314
{493803}	{803921}	515	0.550214
{536969}	{803921}	599	0.548033
{267565}	{803921}	558	0.547059
{144138}	{803921}	627	0.544744
{318506}	{726718}	565	0.544316
{539951}	{726718}	571	0.542776
{68086}	{726718}	529	0.542564
{59633}	{803921}	651	0.541597
{264715}	{803921}	2306	0.540047
{250896}	{803921}	883	0.539731
{7915}	{803921}	522	0.53211
{108507}	{726718}	1101	0.529582
{348498}	{726718}	1413	0.525474
{348498, 726718}	{264715}	742	0.525124
{9633}	{726718}	591	0.524867
{803921, 348498}	{264715}	790	0.522833
{536969}	{726718}	570	0.5215
{726718}	{803921}	2982	0.52051

{250896}	{726718}	851	0.520171
{270896}	{803921}	1088	0.514908
{144138}	{726718}	592	0.514335
{566969}	{726718}	831	0.51233
{39633}	{726718}	1022	0.512281
{108507, 726718}	{264715}	563	0.511353
{19633}	{726718}	644	0.508287
{593921}	{726718}	867	0.506425
{803921, 108507}	{264715}	606	0.506266
{29633}	{726718}	1006	0.506036
{267565}	{726718}	513	0.502941
{803921, 39633}	{264715}	565	0.5
{208362}	{803921}	943	0.497363
{618362}	{803921}	534	0.497207
{618362}	{726718}	532	0.495345
{59633}	{726718}	594	0.494176
{39633, 726718}	{264715}	503	0.492172
{803921, 29633}	{264715}	537	0.489071
{106343}	{803921}	945	0.486612
{803921}	{726718}	2982	0.483855
{264715}	{726718}	2063	0.483138
{270896}	{726718}	1013	0.479413
{270896, 803921}	{264715}	514	0.472426
{106343}	{726718}	899	0.462925
{208362}	{726718}	858	0.452532

{348498, 264715}	{803921, 726718}	532	0.449324
{803921, 726718}	{264715}	1339	0.449027
{348498}	{264715}	1184	0.440312
{108507}	{264715}	872	0.419432
{566969}	{264715}	678	0.418002
{250896}	{264715}	682	0.41687
{19633}	{264715}	518	0.40884
{448103}	{776350}	568	0.401413
{39633}	{264715}	796	0.398997
{593921}	{264715}	682	0.398364
{803921, 264715, 726718}	{348498}	532	0.397311
{29633}	{264715}	768	0.386318
{913783}	{776350}	500	0.38373
{348498, 726718}	{803921, 264715}	532	0.376504
{803921}	{264715}	2306	0.374168
{270896}	{264715}	773	0.365831
{106343}	{264715}	706	0.363543
{726718}	{264715}	2063	0.360098
{264715, 726718}	{348498}	742	0.35967
{108507}	{803921, 726718}	747	0.359307
{803921, 348498}	{264715, 726718}	532	0.352085
{566969}	{803921, 726718}	566	0.348952
{208362}	{264715}	660	0.348101
{39633}	{803921, 726718}	690	0.345865
{803921, 264715}	{348498}	790	0.342585

{348498}	{803921, 726718}	919	0.341763
{593921}	{803921, 726718}	579	0.338201
{29633}	{803921, 726718}	670	0.337022
{250896}	{803921, 726718}	547	0.334352
{264715}	{803921, 726718}	1339	0.313583
{803921, 726718}	{348498}	919	0.308182
{270896}	{803921, 726718}	646	0.305726
{348498}	{803921, 264715}	790	0.29379
{108507}	{803921, 264715}	606	0.291486
{208362}	{803921, 726718}	550	0.290084
{39633}	{803921, 264715}	565	0.283208
{108507}	{348498}	586	0.281866
{106343}	{803921, 726718}	547	0.281668
{29633}	{348498}	556	0.279678
{264715}	{348498}	1184	0.277283
{348498}	{264715, 726718}	742	0.275939
{264715, 726718}	{108507}	563	0.272904
{108507}	{264715, 726718}	563	0.270803
{29633}	{803921, 264715}	537	0.270121
{39633}	{348498}	533	0.267168
{803921, 264715}	{108507}	606	0.262793
{264715, 726718}	{803921, 348498}	532	0.257877
{39633}	{264715, 726718}	503	0.25213
{803921, 726718}	{108507}	747	0.250503
{270896}	{348498}	525	0.248462

{726718}	{348498}	1413	0.24664
{803921}	{348498}	1511	0.245173
{803921, 264715}	{39633}	565	0.245013
{264715, 726718}	{39633}	503	0.24382
{270896}	{803921, 264715}	514	0.243256
{726718}	{803921, 264715}	1339	0.233723
{803921, 264715}	{29633}	537	0.232871
{803921, 726718}	{39633}	690	0.231388
{803921, 264715}	{348498, 726718}	532	0.230703
{803921, 726718}	{29633}	670	0.224681
{803921, 264715}	{270896}	514	0.222897
{348498}	{108507}	586	0.217925
{803921}	{264715, 726718}	1339	0.217264
{803921, 726718}	{270896}	646	0.216633
{348498}	{29633}	556	0.206768
{264715}	{108507}	872	0.204215