Renee Probetts
IEMS 308
2/12/18
Association Rules

Colorado Dillard's Planogram Re-Design

Executive Summary

There are about 254 of the popular retail chain, Dillard's, in the United States, fourteen of those being in Colorado. While online shopping is becoming ever more popular in today's day and age, in store transactions are still important and Dillard's makes millions every year from them. However, they have not redesigned the floor plan for many years and are looking for an update. It is important to design these floor plans based on data. There are certain items that should be placed in the same department because if a customer buys one, it is likely they would also purchase the other. In some instances, it might actually be best to place these items in different departments. The reasoning behind this is to force customers to walk through the store, increasing the likelihood that they will find other items to purchase.
To determine these probabilities, association rules mining is used. Association rules uses probabilities to find three characteristics that a relation can have: support, confidence, and lift. With high values of all three, the relation between items purchased at Dillard's is a good one. This type of analysis is highly useful. If a company is going to spend the money and use the manpower to make large department moves, there needs to be good reasoning behind it. This information can be used to make better informed planograms to best utilize the space that Dillard's has.
The analysis of this Colorado store can be generalized to the rest of their stores nationwide. While it is likely that many of the association rules found in this store will be similar to the rules for stores across the country, it is recommended to reproduce the analysis using data from a variety of states. It is probable that any association rules that exist for the Littleton branch, do not exist in a store in New York.

Problem Statement

Dillard's is one of the most popular department stores in the United States, but with the rise of online shopping, they have seen a drop in sales. Our team, based in Colorado, was tasked with the job of redesigning the planogram in order to better utilize the physical space and increase sales. In order to move certain products around the store, we must determine which products are associated with one another to amplify the effect of this redesign.

Methodology

There are fourteen different Dillard's in the state of Colorado; however, it is inefficient and useless to analyze each different store. Our team did some preliminary analysis on the various Dillard's locations in Colorado and decided that store number 8109, located in Littleton, CO

80124. The decision to focus the analysis on this store was due to the fact that it was the largest in the state, the one with the most business. Thus, it would be a good model for the rest.
In running the analysis, much of the information about Dillard's was unnecessary and needed to be cut out until the end. A data frame was created only containing the sku, trannum, and saledate for this specific store. However, before cutting out stype, any transactions that were returns were discarded from use in the analysis. It is only imperative to look at the purchases.
Once this data was pared down, our team needed to better focus which skus to look at. We decided to discard the skus which didn't appear in at least 10% of the transactions. A sample of this table with this counter can be found in Table 2. Due to the fact that the transaction numbers are recycled based on the day, we combined the sale date and transaction number to create a new, unique variable called "baskets"; a sample can be seen in Table 1. These baskets refer to the unique transactions associated with different sku.

sku	saledate	trannum	basket
180	2005-08-03	2600	2005-08-03, 2600
180	2005-08-09	900	2005-08-09, 900
387	2004-08-15	2400	2004-08-15, 2400
387	2005-03-31	6900	2005-03-31, 6900
450	2005-07-04	4500	2005-07-04, 4500
450	2005-07-24	2200	2005-07-24, 2200
450	2005-07-28	1300	2005-07-28, 1300
450	2005-07-20	3200	2005-07-20, 3200

Table 1: Baskets

sku	baskets	n
2407335	2004-12-21, 2300	19
2407335	2005-01-17, 3400	19
2407335	2005-07-01, 300	19
2407335	2005-01-23, 500	19
2407335	2005-08-27, 4500	19

Table 2: SKU Counts

With this reduced data set, we are able to move on to the association rules analysis. Association rules mining is a useful tool in determining the probability of items being purchased simultaneously. It can be run in R or Python using the apriori algorithm, which narrows down association rules based on a minimum support threshold. Before this can be done, the data must be one-hot encoded. This is done using a function ddply, which takes a basket and returns a table that includes each basket and all of the skus associated with it. Table 3 holds a sample of such.

basket	sku list
2004-08-01,1100	994478
2004-08-01, 1200	2292671,3874099
2004-08-01, 1300	3161221,3854099
2004-08-01, 1400	1184024,1410555
2004-08-01, 1500	3898011,3968011,6971371,9259440
2004-08-01, 1600	803921,1933241,7808101,9288109

Table 3:Sku List and Transaction

Then, it is a simple task of running the apriori function and analyzing the results.

Next Steps

With the information from association rules analysis, Dillard's can move forward with redesigning the planograms. Certain items will need to be moved to different departments which can be a hassle, but will hopefully result in greater profits for the company.

Next steps that the company can take include expanding the analysis to include stores from different areas of the country. It is not necessary to run the analysis for every single store; however, having a broader variety of customers included in the transactions will provide a better idea of what might work in every single store… if they are looking to make the same moves nationwide. Otherwise, it is something that could be done regionally.

Additionally, with this information, Dillard's can plan for better promotions. The items that are associated, should not be put on promotion together. A promotion for one will likely increase the sales of the other.

There are many uses for association rules mining and rearranging the planograms is only the beginning.