Melissa Melnick
Association Rules
IEMS 308
02/15/2021

Executive Summary

Dillard’s is a major retail chain with nearly 300 stores across 29 states in the U.S; the results of this analysis are based off of all of the Dillard’s stores located in Missouri. This report aims to identify the SKUs that are the most promising candidates for rearranging the store’s planograms through association rules learning. This method is a rule-based machine learning method used to identify interesting relationships between variables within large datasets.

Problem Statement

Due to budgetary constraints, Dillard’s is only able to make 20 total moves across the entire chain of stores. A move consists of changing the location of a certain product within the store. This report aims to find the 100 SKUs with the most promising association with other products.

Methodology

Dillard’s POS data was stored in 5 separate CSV files. The first step of this analysis is to merge columns from all 5 of these files into one Dataframe with only the most necessary information to solve the problem. To begin, the amount of data is enormous, so to make for a more concise analysis, a specific subset of data was selected. Only Dillard’s stores located in Missouri were used for this study.

The next step was to filter out all of the data points that referred to a Return. If Dillard’s rearrangement of the planogram ultimately has a goal of creating more sales, it would not be helpful in this case to study the return data. Only the Purchase was considered in this analysis.

Even after carefully subsetting the data, only selecting the purchasing data, there were still nearly 250 thousand unique SKUs present in the dataset. To combat this, the 200 most popular SKUs were identified based on their frequency, and the dataset was once again pared down to transactions that only contained a purchase with one of the 200 most popular skus.
Next, we had to create a unique ID for each basket that exists within the dataset. We created this unique ID by concatenating the Store Number, the Register Number, the Transaction Number, the Sequence Number, and the Sale Date.

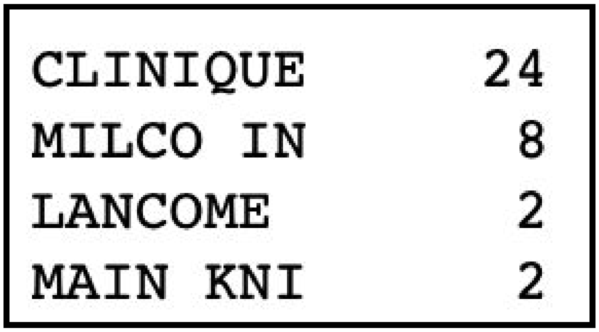
This biggest reason for only selecting the 200 most popular SKUs is because each one needs to be one-hot encoded in order to perform association rules analysis, which is the next step in this methodology. Finally, the last thing to do before performing the association rules analysis is to create a final Dataframe of just the Basket Ids and the One-Hot encoded SKU numbers.

To begin the actual Association Rules Analysis, we use the mlxtend package. First we use the Apriori function to determine item sets that are frequently seen together. We chose the minimum support level as .0002 because with approximately 225 thousand baskets, .0002 support means the set will be found in approximately 45 of those baskets, which seems like an appropriate number on which to base this analysis.

Finally, we used the frequent item sets generated in the last step to calculate the association rules for this analysis.

Key Insights

Upon investigation of the data, the most common SKU values present in the top 100 association rules referred to products by Clinique, a cosmetics company. Looking further at the association rules, we can see that customers often purchase multiple projects from the same brand. For example, two Clinique produces, or two Lancome Products. The figure below shows the most frequent brands that show up in the top association rules.



```
CLINIQUE      24
MILCO IN       8
LANCOME        2
MAIN KNI       2
```

Conclusions

In the future, when Dillard’s sets up its cosmetics department, it should be sure to highlight the products from both Clinique and Lancome.

To get a better idea of these trends, in the future more research should be performed in other regions of the United States to see if these trends are consistent across all Dillard’s stores.