**Modifying Planograms for Dillard’s**

Rosanne Harrison

**Executive Summary**

The goal of this analysis was to determine the best way for Dillard’s, a department store with several hundred locations, to modify their planograms in a way that will increase sales/profit. Since Dillard’s has a limited budget we needed to determine a list of 100 SKUs that are the best candidates for relocations; from this list, Dillard’s will choose 20 items to execute the moves.

The analytical tool applied in the following analysis is market-basket analysis, or association rules. This method involves using transactional data to identify items that are commonly purchased together, and eventually deducing ‘rules’ that will predict the purchase of one item based on the presence of other item(s) in a shopper’s basket. This is a data mining technique often used by retail stores, making it very appropriate for Dillard’s analysis. By identifying the strongest association rules between SKUs in Dillard’s stores, I will enable Dillard’s to modify their planograms based on concrete data. Dillard’s can then make decisions about SKUs locations’ based on their relative distance to other associated SKUs and increase the odds that customers will purchase multiple items.

After exploring, cleaning and selecting a subset of the available data for analysis, I calculated the three key statistics used in market basket analysis for each of the remaining pairs of SKUs. Based on these statistics – support, confidence and lift – I was able to identify the strongest association rules within my data. I also applied a minimum support value, to ensure that any SKUs that were selected were commonly purchased in Dillard’s stores, and took into account the profitability of each selected SKU. In the end, I selected 100 SKUs that were a part of the best association rules, and provided this list to Dillard’s for the modification of their planograms.

In analyzing the final association rule pairs, I was able to draw several conclusions about the viability of the discovered rules. I noticed that while support seemed generally low for all SKU (likely due to the massive number of SKUs sold at Dillard’s), confidence in many of our rules was extremely high. This indicated that the decisions that Dillard’s makes about these items locations could very well impact customers’ purchasing behavior. I also acknowledged that not all of the 100 SKUs were equally good candidates for planogram modification, as some rules are stronger than others, and noted that it is essential that Dillard’s relocations take into account BOTH SKUs in an associated pair, as their proximity to one another is the key aspect of the planogram modification.

**Problem Statement**

This project’s client is Dillard’s, a major department store chain in the US. They are interested in rearranging the planograms in their stores in a way that will lead to increased sales/profits. Due to budgetary constraints, Dillard’s can only afford to make up to 20 moves across the chain (where a move involves moving one SKU to a new location). Our goal is to create a list of 100 SKUs that are the best candidates for these moves.

**Assumptions**

* When a SKU is moved, it is moved in all stores across the entire Dillard’s chain.
* There are no errors or inaccuracies in data set that are significant enough to impact our analysis
* Dillard’s wants to base their planogram modifications on association rules derived between various SKUs that they stock

**Methodology**

**Data Parsing**

The first step in solving this problem was to analyze and comprehend the available data. The data was stored in five tables that contained information about Dillard’s stores, SKUs, transactions, departments and prices. Much of the data in these tables was extraneous information; determining the relevant information was a key aspect to solving the problem. According to the SKU table there are **1,048,576 SKUs** that can be sold at Dillard’s; however, based on the table that relates SKUs with stores and their prices (SKSTINFO) there are only **760,212** SKUs across Dillard’s stores within the time frame of our data. Similarly, the ‘store info’ table shows that there are **453 Dillard’s stores**, however we only have data relating **357** of these stores to SKUs. For the purpose of this analysis, we will be using the SKSTINFO table for the count of SKUs (and stores) eligible for planogram analysis.

The next step was to narrow down the SKUs before performing computationally expensive analysis. In order to do this, I calculated how many unique SKUs were sold in each Dillard’s stores. The numbers ranged from a low of 4 SKUs to a high of just over 216,000. I decided that it wouldn’t make sense to move a SKU that is only sold in a few Dillard’s stores; the ideal candidates would be sold in most Dillard’s stores so that the chain can reap maximum benefits from the new planograms. Based on this understanding, I decided to move forward only considering SKUs that are sold in at least **80%** of the 357 Dillard’s stores for which we have data. This left **30,403 SKUs** for analysis.

I now moved on to analyzing the transaction table. Several “clean-up” steps had to be applied to this table – I first pulled only the rows that corresponded to the ~30,000 SKUs I had selected for analysis, and also eliminated extraneous columns. I found that three variables – store, sale date and transaction number – are required to identify a unique transaction and thus kept all of those columns. I also noticed that the data set had not combined rows where multiples of the same item were purchased in the same transaction. To fix this, I summed the “quantity” column for all rows of identical SKUs purchased in the same transaction.

After this analysis had been performed, I noticed that the remaining data set only contained **22,234** unique SKUs – this meant that ~8,000 of my “good SKUs” had not actually been purchased over the time period of our transaction data.

The final step of the data preparation phase was to determine the profitability of each SKU. This information will be used later on to ensure that the SKUs that are moved are relatively profitable items. This calculation involved subtracting the cost of each item from the retail price of the item (retail - cost in the SKSTINFO table). For SKUs that had multiple prices or costs across different stores, a weighted average was taken to calculate the average profit for each SKU.

**Association Rule Analysis**

At this stage, it was time to begin the association rule analysis. To determine the quality of an association rule, three metrics must be checked – support, confidence and lift. Support is generally the first of these measures to be analyzed.

The first step of analyzing support was to use some “minimum support” value to eliminate rarely-purchased SKUs from the data set before beginning to analyze the relationships between various SKUs. Ideally I would have liked to use **MIS**, which assigns each SKU a ‘minimum item support’ value that takes each item’s profitability into account when determining the minimum required support. However, due to the enormous size of the data set and absence of a knowledgeable store manager who could provide information on the MIS ratio, I decided to simply use an overall ‘minsup’ value.

That being said, I still wanted to take profit into account in some way; obviously, relocating a very profitable item to its optimal position would be more beneficial than moving an low-profit item. This was especially important because I found that some SKUs actually brought in negative profits – most likely loss-leaders for Dillard’s. Therefore, I decided to only analyze SKUs that brought in the **top 50% of profits**. This still left **11,117 SKUs** for analysis, and also ensured that any SKUs we chose to relocate would bring in mid-to-high profits (ranging from $9 to $393, with an average profit of about $25).

After taking profits into account, I determined what ‘minsup’ value to use. I tried several values, starting with 1000 transactions; however, it ended up that there weren’t enough SKUs that met this standard. I eventually settled on a ‘minsup’ value of 500 transactions. I pulled out all SKUs that appeared in at least 500 transactions, and then cross-joined those SKUs with each other and found each *pair* of SKUs that appeared in the data at least 500 times. This resulted in **206 unique pairs**. This number was still relatively small compared to what I had expected, showing that support was generally low within the data set. Based on this observation, I decided not to look into groupings of three or more SKUS, as the number of remaining SKUs was already small enough to analyzed manually.

The next step of the process was to calculate confidence and lift for all pairs of SKUs. Since each pair showed up twice, with the order of the SKUs reversed, I had 412 pairs to analyze. All three statistics were found by counting the number of transactions that contained “SKU 1”, “SKU 2” and both SKUs together, as well as finding the total number of transactions (just over 7.1 million). These values were then used to find the probabilities associated with each SKU and SKU pair, and then were combined to find support, lift and confidence. The specific calculations can be seen in the accompanying Excel file.

Our final step was to use these statistics to determine the 100 *best* SKUs for modifying Dillard’s planograms. Obviously, the ideal items to be moved are the SKUs with the strongest association rules; by adjusting the proximity of associated SKUs to one another, Dillard’s can hopefully improve their sales of these items. There is no exact science to determining which association rules are the strongest, as support, lift and confidence often don’t agree.

I decided to rely most on confidence and lift for the selection of my top 100 SKUs – this was due to support being generally low for all pairings, and the fact that I had already applied I minimum support value. All 100 of the SKUs I chose were in the top 100 pairings ranked by either support or confidence, and many were in the top 100 for both. I chose to include both SKUs in each ‘top’ pair, as we aren’t assuming that any SKUs are already appropriately placed, and thus Dillard’s may want to move both items to a more prominent location, etc.

The table below contains my top 100 SKUs for planogram modification:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TOP 100 SKUs (1-50)** |  | **TOP 100 SKUs (51-100)** |
| 1 | **2488302** | 51 | **994478** |
| 2 | **3181454** | 52 | **3824833** |
| 3 | **2448302** | 53 | **33993** |
| 4 | **924337** | 54 | **43993** |
| 5 | **2478302** | 55 | **5307532** |
| 6 | **1958302** | 56 | **6802574** |
| 7 | **108803** | 57 | **1679519** |
| 8 | **1998302** | 58 | **6812574** |
| 9 | **698460** | 59 | **1619519** |
| 10 | **94337** | 60 | **5317532** |
| 11 | **8158459** | 61 | **3083676** |
| 12 | **2504337** | 62 | **6732574** |
| 13 | **1988302** | 63 | **173088** |
| 14 | **248803** | 64 | **2593676** |
| 15 | **9904336** | 65 | **844455** |
| 16 | **3693199** | 66 | **874455** |
| 17 | **238803** | 67 | **4397410** |
| 18 | **894337** | 68 | **233088** |
| 19 | **49927** | 69 | **8178673** |
| 20 | **9578459** | 70 | **8238673** |
| 21 | **59927** | 71 | **4673560** |
| 22 | **68065** | 72 | **2954874** |
| 23 | **7673560** | 73 | **1004478** |
| 24 | **3753199** | 74 | **8809152** |
| 25 | **208065** | 75 | **3343560** |
| 26 | **128803** | 76 | **223088** |
| 27 | **8584832** | 77 | **4103560** |
| 28 | **2444337** | 78 | **6438963** |
| 29 | **1741454** | 79 | **5358963** |
| 30 | **3743199** | 80 | **5017532** |
| 31 | **6713560** | 81 | **3528979** |
| 32 | **9183560** | 82 | **6499231** |
| 33 | **5453561** | 83 | **4997532** |
| 34 | **1639519** | 84 | **9009152** |
| 35 | **1709519** | 85 | **3537981** |
| 36 | **8467157** | 86 | **6359231** |
| 37 | **1014478** | 87 | **1262504** |
| 38 | **7561108** | 88 | **4342850** |
| 39 | **9629925** | 89 | **8888965** |
| 40 | **1689519** | 90 | **9460196** |
| 41 | **4887932** | 91 | **3542850** |
| 42 | **1629519** | 92 | **4118498** |
| 43 | **4897932** | 93 | **3949538** |
| 44 | **9609925** | 94 | **9358964** |
| 45 | **148065** | 95 | **9549158** |
| 46 | **2863676** | 96 | **348498** |
| 47 | **2463676** | 97 | **2272504** |
| 48 | **4387410** | 98 | **9607293** |
| 49 | **6722574** | 99 | **1697200** |
| 50 | **7977157** | 100 | **9617293** |

**Conclusions**

There are several key business insights that we can take away from this association rule analysis:

**Not all 100 of the SKUs are equally as good candidates for modifying planograms.** Some associations are stronger and those SKUs should be prioritized. Additionally, the most financially optimal action is likely to move only *one* SKU in each associated pair; the key to moving these SKUs is not their absolute placement in the store, but rather their location relative to the SKU(s) they are associated with.

**The support for ALL associated pairs is generally low.** Even the ‘highest support’ pairing has a support of 0.11%, which is extremely weak. This is due to the massive amount of SKUs that are present in the data set, even after removing a large number of SKUs. Additionally, the final transaction table contains ~7.14 million transactions, but only ~1.03 million of those transactions contained multiple SKUs.

**Confidence is extremely high for many of the associated pairs, indicating that their association is not coincidental.** The highest confidence pair has about ~71% confidence; this number is extremely high for market basket analysis in general, and is especially impressive when we consider the massive number of transactions (over 7 million) in this analysis. For the confidence value to be that high indicates that there is most definitely a co-occurrence pattern between the two SKUs in question, and makes us confident that the relocation of these SKUs could significantly improve sales.

**For many pairs, it is not clear which direction the association goes (e.g. whether A 🡪 B or**

**B 🡪 A).** When we reverse the direction of some pairs, the confidence value stays very high, sometimes even maintaining almost the exact same value. This makes us unsure about which SKU indicates the other; it is even possible that some of the associations go both ways. While the direction of the association is imperative for some marketing schemes (such as discounting one item to boost sales of the other), it is less important for planogram modification. Since we mostly care about SKUs’ proximities to their associated match, understanding the directional relationship is non-essential.

**Next** **Steps**

* For Dillard’s, the next step involves determining which 20 of the 100 SKUs listed in this report are **best** for planogram modification. This will likely involve looking back at the raw data for the top association rules. They will also need to determine whether they are going to move *both*SKUs in an associated pair, or simply move one of the SKUs. This will determine whether they should look at the top 10 or top 20 rules to select their optimal SKUs.
* Dillard’s also needs to determine which methodology they will be using when relocating the SKUs. One common approach is to place associated SKUs next to each other; when a customer goes to buy one item in the pair, they will be reminded of the associated item and will find it convenient to grab and purchase. This would lead to improved profits via more sales of those related SKUs. On the other hand, other methods indicate that placing the SKUs far apart from each other can increase sales, as customers will be forced to walk around the store to fill their basket and may be inspired to pick up additional items. Obviously, this decision will be very important for Dillard’s.
* With regards to next steps that could lead to an improved analysis, one key step would be to utilize MIS (or Minimum Item Support) when selecting SKUs that pass the minimum support threshold. While this analysis would require more computing power and the expertise of someone at Dillard’s, it would allow us to perform the analysis in a way that more equally balances profitability with number of occurrences of a SKU. By eliminating some of the “cutting down” of the data set, we can eliminate any bias in our solution and get the optimal results.