# Association Rules:
# Dillards Department Store

Anna Blakley
IEMS 308
Professor Klabjan
February 14, 2020

# Executive Summary

The goal of this project was to identify the top 100 SKUs to be moved within Dillard's Department Store based on items frequently purchased together. After examining the market basket data from Dillard's Department stores, the following procedure was taken: 1. Examine and understand the data 2. Create a subset of data 3. Execute association rules algorithm on subset 4. Run association rules algorithm on test data. The data included purchases over multiple years, at multiple locations, at multiple registers. The average number of items per transaction was approximately 2. There were 111,649,093 rows/transactions and 13 columns. There were also 332 stores, and 713172 SKUs. SKUs purchased varied by store, month, and year, so the final subset of data was all purchases from store "8402" from January 2005, training, and February 2005, test. For the training data, the minimum support was set to 0.001, the minimum confidence was set to 1, and the minimum lift was set to 990. This yielded 89428 association rules which included up to seven elements. However, many of these were combinations of the same items with different combinations as the antecedents and consequents.

One of the most important findings from executing association rules on this data was that most items purchased together were from the same brand i.e. two or more items from Clinique were purchased together. The brands that were most commonly in these association rules were "Clinique", "Liz Clai", "FORCE ON", "CABERNET", and "ROUNDTRE." However, other brands and SKUs were also included in the association rules. All association rules are outlined in the file "rules_train", all 100 top SKUs are outlined in "FINALSKU", and all brands are found in "FINAL BRANDS."

The test data revealed a similar trend that most items purchased together were from the same brand, however, the primary brand that this applied to was Haskell. Several of the same SKUs were present in the test rules were the same as those in the training data.

# Summary

**Business Question:** What are the top 100 SKUs to be moved in Dillards Department Store based on Association Rules?

**Process:**

1. Explore data and understand columns
2. Identify a subset. In the case of this exploration the subset was all purchases at store 8402 in January 2005.
3. Create unique transaction values. The transaction numbers were unique to a specific register at each store, so to create unique transaction numbers I combined register number, date, and transaction number to create a unique code for the transactions.
4. Execute Association Rules on the data.
5. Identify 100 SKUs to move based on the association rules and understand brands.
6. Repeat 3-5 on test data that is a subset from the same store (8402), but in February 2005.

# Exploration

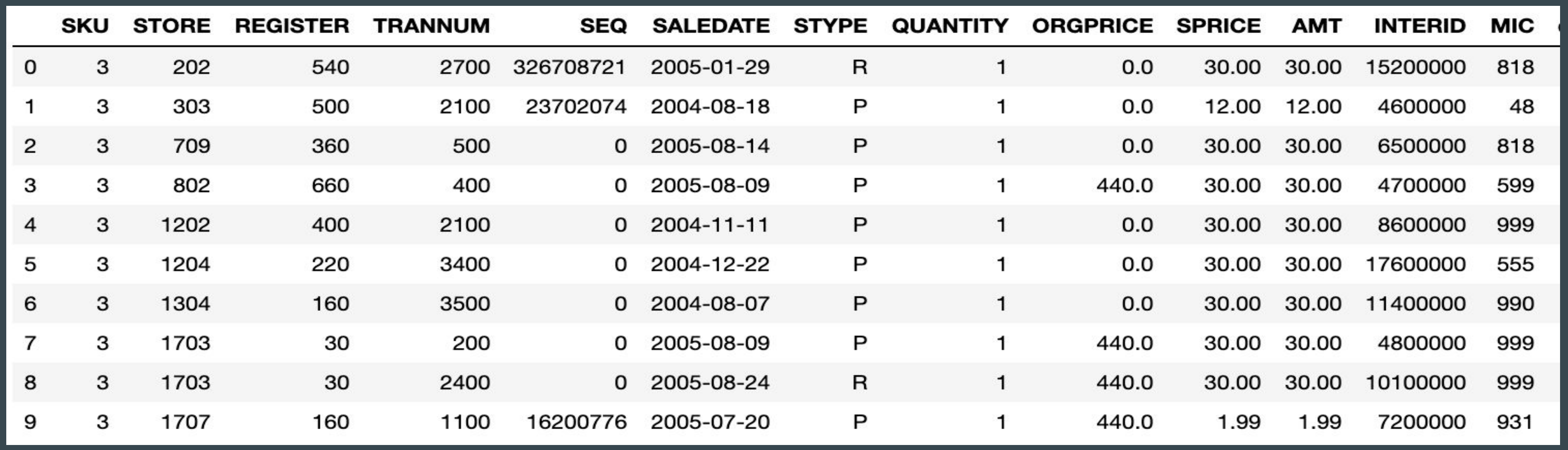
**Dataset features:**

111,649,093 Rows/Transactions

13 columns

332 Stores

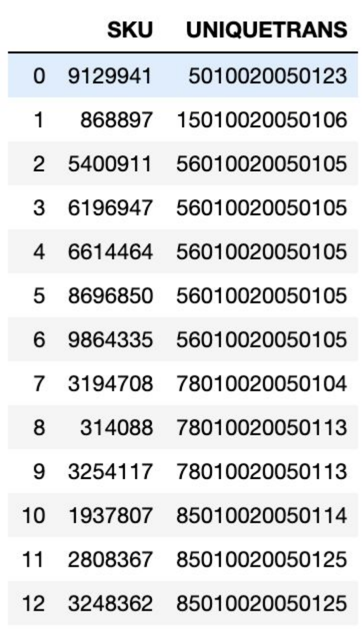
713172 Unique SKUs



| | SKU | STORE | REGISTER | TRANNUM | SEQ | SALEDATE | STYPE | QUANTITY | ORGPRICE | SPRICE | AMT | INTERID | MIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3 | 202 | 540 | 2700 | 326708721 | 2005-01-29 | R | 1 | 0.0 | 30.00 | 30.00 | 15200000 | 818 |
| 1 | 3 | 303 | 500 | 2100 | 23702074 | 2004-08-18 | P | 1 | 0.0 | 12.00 | 12.00 | 4600000 | 48 |
| 2 | 3 | 709 | 360 | 500 | 0 | 2005-08-14 | P | 1 | 0.0 | 30.00 | 30.00 | 6500000 | 818 |
| 3 | 3 | 802 | 660 | 400 | 0 | 2005-08-09 | P | 1 | 440.0 | 30.00 | 30.00 | 4700000 | 599 |
| 4 | 3 | 1202 | 400 | 2100 | 0 | 2004-11-11 | P | 1 | 0.0 | 30.00 | 30.00 | 8600000 | 999 |
| 5 | 3 | 1204 | 220 | 3400 | 0 | 2004-12-22 | P | 1 | 0.0 | 30.00 | 30.00 | 17600000 | 555 |
| 6 | 3 | 1304 | 160 | 3500 | 0 | 2004-08-07 | P | 1 | 0.0 | 30.00 | 30.00 | 11400000 | 990 |
| 7 | 3 | 1703 | 30 | 200 | 0 | 2005-08-09 | P | 1 | 440.0 | 30.00 | 30.00 | 4800000 | 999 |
| 8 | 3 | 1703 | 30 | 2400 | 0 | 2005-08-24 | R | 1 | 440.0 | 30.00 | 30.00 | 10100000 | 999 |
| 9 | 3 | 1707 | 160 | 1100 | 16200776 | 2005-07-20 | P | 1 | 440.0 | 1.99 | 1.99 | 7200000 | 931 |

Figure 1: Head of full dataset

**Within Subset:**



| | SKU | UNIQUETRANS |
|---|---|---|
| 0 | 9129941 | 5010020050123 |
| 1 | 868897 | 15010020050106 |
| 2 | 5400911 | 56010020050105 |
| 3 | 6196947 | 56010020050105 |
| 4 | 6614464 | 56010020050105 |
| 5 | 8696850 | 56010020050105 |
| 6 | 9864335 | 56010020050105 |
| 7 | 3194708 | 78010020050104 |
| 8 | 314088 | 78010020050113 |
| 9 | 3254117 | 78010020050113 |
| 10 | 1937807 | 85010020050114 |
| 11 | 2808367 | 85010020050125 |
| 12 | 3248362 | 85010020050125 |

Figure 2: Head of subset

32163 Unique Transactions

60421 Total Items Purchased

For the Subset, ~1.879 Items Per Transactions

Structure of Unique Transactions: REGISTER|TRANNUM|SALEDATE

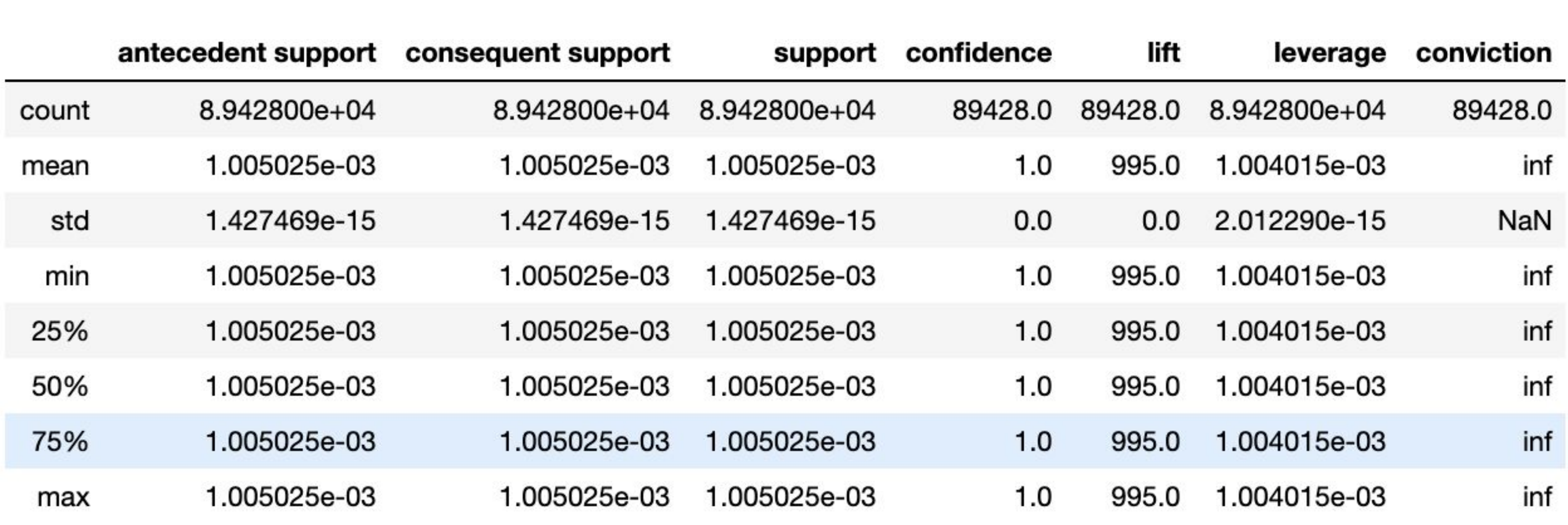
# Association Rules

Minimum Support: 0.001

Minimum Confidence : 1

Minimum Lift: 990

Total Association Rules: 89428 with up to seven elements many of which are combinations of the same items with different combinations as the antecedents and consequents



| | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|---|---|---|---|---|---|---|---|
| count | 8.942800e+04 | 8.942800e+04 | 8.942800e+04 | 89428.0 | 89428.0 | 8.942800e+04 | 89428.0 |
| mean | 1.005025e-03 | 1.005025e-03 | 1.005025e-03 | 1.0 | 995.0 | 1.004015e-03 | inf |
| std | 1.427469e-15 | 1.427469e-15 | 1.427469e-15 | 0.0 | 0.0 | 2.012290e-15 | NaN |
| min | 1.005025e-03 | 1.005025e-03 | 1.005025e-03 | 1.0 | 995.0 | 1.004015e-03 | inf |
| 25% | 1.005025e-03 | 1.005025e-03 | 1.005025e-03 | 1.0 | 995.0 | 1.004015e-03 | inf |
| 50% | 1.005025e-03 | 1.005025e-03 | 1.005025e-03 | 1.0 | 995.0 | 1.004015e-03 | inf |
| 75% | 1.005025e-03 | 1.005025e-03 | 1.005025e-03 | 1.0 | 995.0 | 1.004015e-03 | inf |
| max | 1.005025e-03 | 1.005025e-03 | 1.005025e-03 | 1.0 | 995.0 | 1.004015e-03 | inf |

Figure 3: Statistical Description of Association Rules

These rules were used to outline the top 100 SKUs that are candidates to move within Dillards. These can be seen to the right or in the file 'FINALSKU". The other files provided do not include descriptions of the items aside from size, brand, color, and codes that cannot be understood without a code. I used the brands to identify the items. These brands, with duplicates removed can also be seen on the right or in the file "FINALBRANDS"

| SKU | |
|---|---|
| 5888065 | 4751496 |
| 126170 | 474515 |
| 6163107 | 3896862 |
| 7044853 | 486689 |
| 146997 | 756689 |
| 2598084 | 7497591 |
| 6416562 | 4432751 |
| 148061 | 423714 |
| 8226562 | 7256846 |
| 8976562 | 458020 |
| 160720 | 9744811 |
| 5367556 | 366117 |
| 6036119 | 9896116 |
| 176017 | 3924024 |
| 186170 | 5317384 |
| 708367 | 293277 |
| 5706016 | 6536659 |
| 656698 | 2779816 |
| 264286 | 309665 |
| 1754130 | 204684 |
| 7376697 | 2161039 |
| 8836697 | 4594108 |
| 264715 | 231057 |
| 3254117 | 9411343 |
| 314088 | 250896 |
| 8077274 | 1937918 |
| 316667 | 5378213 |
| 3237302 | 7546562 |
| 324286 | 7886562 |
| 1576017 | 4380542 |
| 346017 | 44522 |
| 346689 | 957390 |
| 1276689 | 78355 |
| 3528835 | 3437218 |
| 3638835 | 122118 |
| 9469364 | 7.08E+12 |
| 354814 | 9633 |
| 9600684 | 9932313 |
| 386625 | 26691 |
| 7176700 | 2966979 |
| 7446105 | 1724045 |
| 8686732 | 64045 |
| 5766710 | 112212 |
| 407541 | 4313732 |
| 1394179 | 86170 |
| 416740 | 4008011 |
| 2376291 | 1842285 |
| 466823 | 4108011 |
| 2626299 | 9323130 |
| 3406291 | 9186625 |

| BRAND |
|---|
| CLINIQUE |
| AUGUST H |
| KASPER A |
| KORET OF |
| LANCOME |
| CABERNET |
| LIZ CLAI |
| ROUNDTRE |
| FORCE ON |
| NAPIER/V |
| I.C. ISA |
| SIGRID O |
| ALPHA GA |
| NAUTICA |
| HART SCH |
| POLO FAS |
| ROYCE HO |
| THE SAK |
| DIANE GI |
| E.W.L. S |
| CM SHAPE |
| HUE/KAYS |
| FISHMAN |
| HEARTBRE |
| L.SCOTT |
| G. H. BA |
| M.M. & R |
| DAX CORP |
| GREAT AM |
| TOO SHY |
| POLO JEA |
| BIJOUX G |
| MOBILE E |
| NYGARD |
| CAROLE H |
| NEXT ERA |
| IT JEANS |
| CALVIN K |

# Business Insights

- One of the most important findings from executing association rules on this data was that most items purchased together were from the same brand i.e. two or more items from Clinique were purchased together.
- The brands that very often were purchased with items from the same brand were:
  - Clinique
  - Liz Clai
  - Force On
  - Cabernet
  - Roundtre
- The brands mentioned above as well as Koret Of, Next Era, IT Jeans, and Calvin Klein appeared repeatedly throughout the association rules
- The test data had different brands appear in the association rules, more commonly Haskell, but revealed the same takeaways that items of the same brand were most frequently purchased together. The different SKUs is likely due to frequent inventory changes.
- The lift is very high and confidence is one, meaning that these association rules are meaningful.
- **Takeaway:** The best approach to moving items, if only 20 items can be moved, is to group by brand. Most likely, people have significant brand loyalty or brands create products that work together. Placing items of the same brand together will push shoppers to buy multiple items. However, the 100 SKUs listed are all good candidates.

**Outputs:**
Rules_train.csv: Rules for training data (analysis focused on this)
FINALSKU.csv: Top 100 SKUs
FINALBRANDS.csv: Brands associated with top 100 SKUs

**All code to execute association rules(took too long to restart every kernel everytime):**
SKU_Brands.ipynb: Dataframe of all SKUs and Brands
Association Rules.ipynb: Exploration of full dataset
Association Rules pt 2.ipynb: Creating subset
Association Rules 3.ipynb: Creating 1,0 coding
Association Rules 4.ipynb: Executes association rules
SKU Identification.ipynb: Analyzes association rules and identify top 100 SKUs and brands

**Test data:**
TEST.zip: Code and outputs for Test data