MARKET BASKET ANALYSIS

Introduction

Curiosity struck at the most opportune time, considering the current online retail sales climate and the evolution of e-commerce companies the likes of Amazon and eBay have spearheaded. A Market Basket Analysis allows decision-makers to understand their product relationships and customer purchase behavior. In this analysis, we expect valuable insights from sales data allowing for more informed business strategies for the next year.

Data:

https://www.kaggle.com/datasets/suhanias/market-basket-analysis-data

Our analysis is based on transactional data from online retail stores from 2010-2011. This data set is sourced from Kaggle, a widely recognized platform for sharing and accessing datasets. The data has over 4,000 unique items available for sale with more than 500,000 data entries.

- BillNo: unique number for each transaction
- ItemName: identifies the unique name of the product.
- Quantity: amount purchased of the product
- Date: time of date purchased
- Price: cost of a product in USD
- Customer ID: a unique identifier for those with accounts in the store; turned into a binary value due to the number of null values in its column.
- Country: the country in which the purchase took place

Objectives:

With over 4000 unique products, 500,000 transactions total, and multiple storefronts this retailer needs to understand its product associations and customer behavior. The project seeks to identify and understand trends and patterns through the comparison of Key Performance Indicators (KPI) of the sales data. Uncovering strong item combinations that could boost sales, marketing opportunities, and Inventory Management.

Questions

- How many association rules is an effective amount? What parameters does that include?
- Exploratory data analysis? Avg spent? What are the most frequently purchased items?
- Establish standard threshold goals for item sets.
- Which products should the marketing team focus on the following year?
- Where can we improve support itemsets and encourage multiple items in transactions as much as possible?
- Notice any seasonal or other trends?

Data cleaning:

Correcting Data types

- Setting date data into a datetime data type

Formatting our null values

- Removing all null values in the 'itemname' column, as we have sufficient data points

Correcting membership values

- Changing those with customerid's to 1 value and non customerids to 0
- Allows for the retailer to identify customers who have memberships with the store or not

Preparing for Apriori model

  - Basket = rows with the same 'BillNo' get combined into a basket
  - Basket_sets= turns the basket items into binary values to either the unique item the store offers was purchased (1) or not (0)

Apriori:

The Apriori algorithm is used to mine frequent item sets and uncover important association rules of the transactional data. Setting the correct parameters allows for the identification of meaningful associations among the transactions

Key Metrics

- Support: measure how frequently an item or item set appears in that dataset
- Confidence: measure the likelihood that an item B would be purchased when Item A is purchased.
- Lift: measure how much more likely Item B would be purchased if Item A is purchased compared to if you purchased it alone.
- Antecedents (A): is the item or itemset found on the left side of the association rule and is the initial item that suggests other purchases.
- Consequents (B): is the item or item set found on the Right Side of the association rules and is the item affected by the initial item or the antecedents
- Item set(s): is a collection of one or more items that appear together

Parameters: set of conditions to refine the most impactful relationships amongst the items

Min_support: .03	Set at .03 or 3%; focused on the most frequently occurring item sets
Min_confidence: .50	Set at .50 or 50%; gave the strongest relationships critical for insights
Max_length: 2	Max length of 2; seeking item sets of 2 items or 1 by itself.

Data exploration:

Yearly

- Total entries: 522,061
- Total baskets: 20,205
- Unique items: 4,175
- Average item cost: $3.86
- Avg basket cost: $509.32
- Avg item amount in basket: 261
- Rules generated: divided by month in monthly analysis

January

- Total entries: 34,102
- Total baskets: 1,091
- Unique items: 2,529
- Average item cost: $ 3.77
- Avg basket cost: $ 611.68
- Avg item amt in basket: 340
- Rules generated: 264

April

- Total entries: 28,765
- Total baskets: 1,261
- Unique items: 2,435
- Average item cost: $3.96
- Avg basket cost: $ 418.77
- Avg item amt in basket: 239
- Rules generated: 175

August

- Total entries: 33,656
- Total baskets: 1,345
- Unique items: 2,579
- Average item cost: $3.69
- Avg basket cost: $ 537.68
- Avg item amt in basket: 303
- Rules generated: 276

December

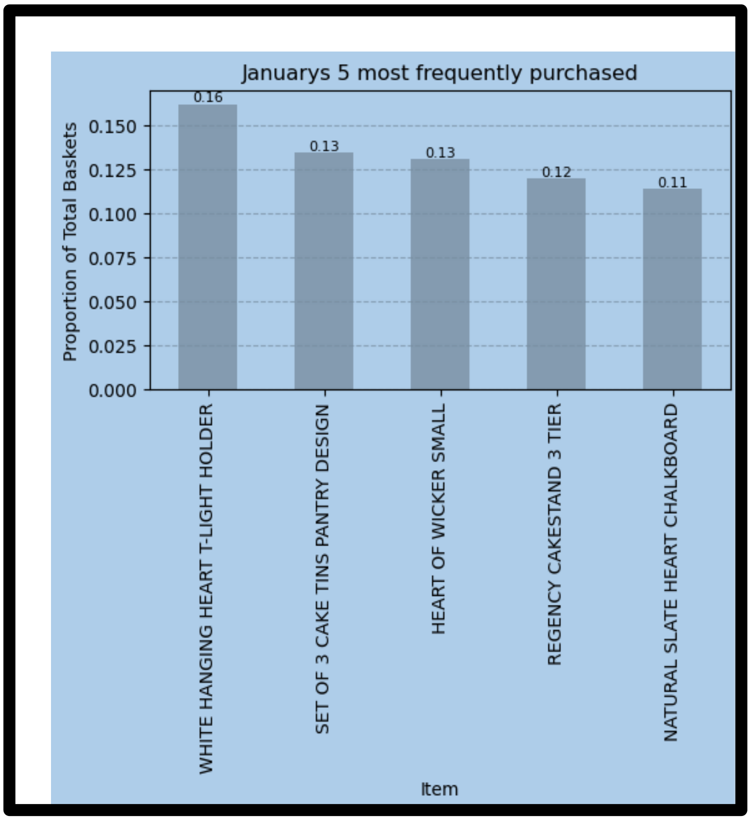
- Total entries: 65,906
- Total baskets: 2,392
- Unique items: 3,435
- Average item cost: $4.52
- Avg basket cost: $ 601.77
- Avg item amt in basket: 274
- Rules generated: 174

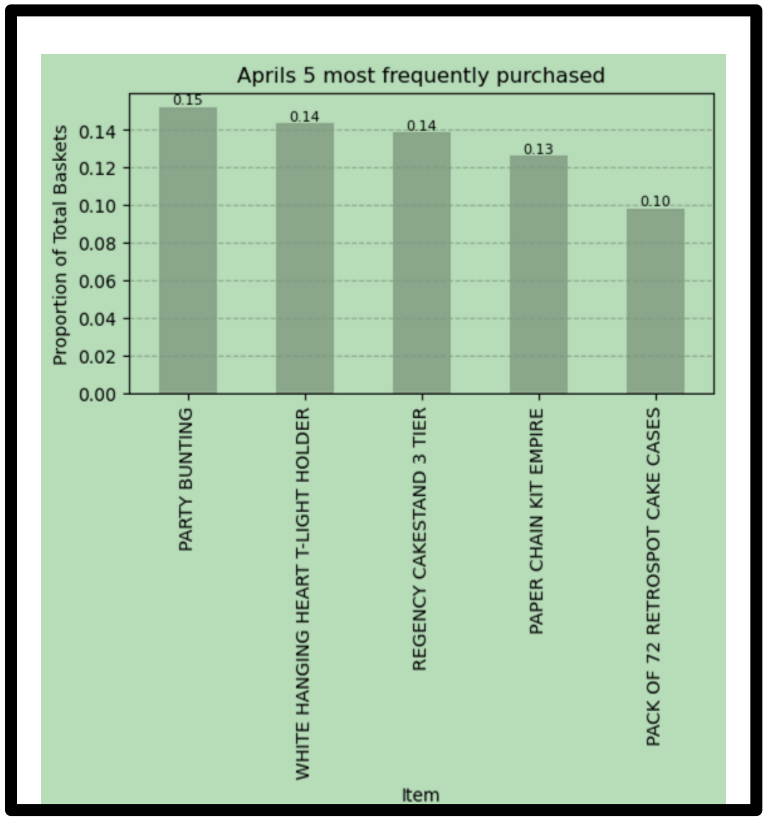
Takeaways

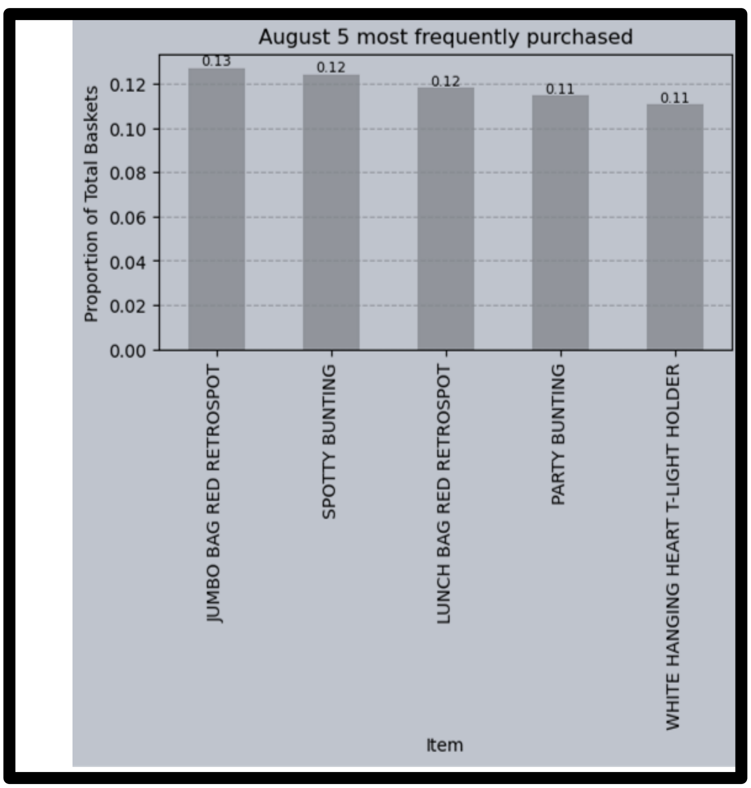
- December had the most entries with 65,0906; most baskets at 2,392; most unique items at 3,435

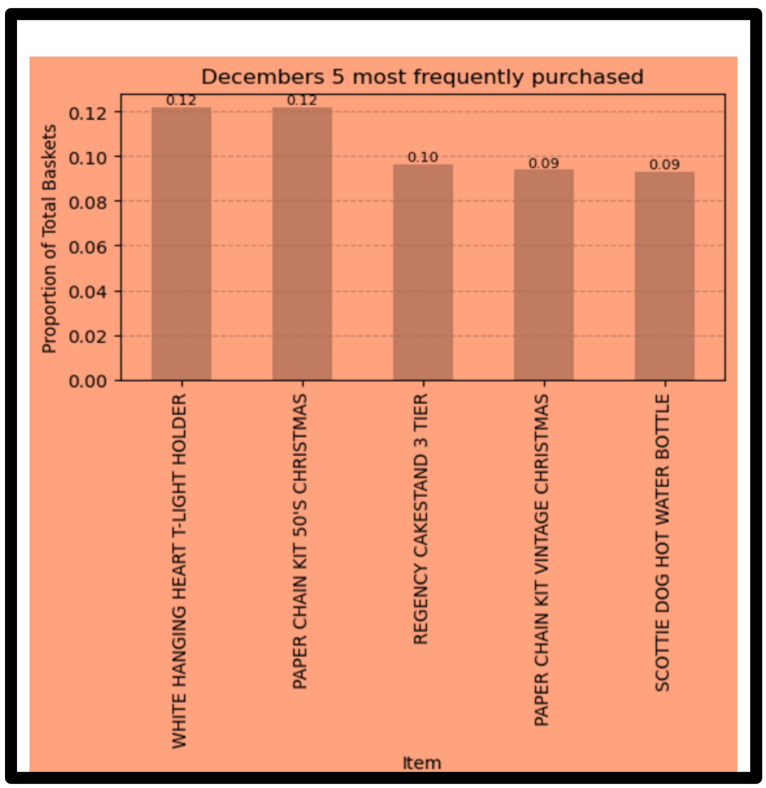
- April had the least entries with 28,765 but not least baskets suggesting a trend of purchasing a single item and not multiple

- August while having almost half the entries as December still generated double association rules


Januarys 5 most frequently purchased
Proportion of Total Baskets
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.16
0.13
0.13
0.12
0.11
WHITE HANGING HEART T-LIGHT HOLDER
SET OF 3 CAKE TINS PANTRY DESIGN
HEART OF WICKER SMALL
REGENCY CAKESTAND 3 TIER
NATURAL SLATE HEART CHALKBOARD
Item


Aprils 5 most frequently purchased
Proportion of Total Baskets
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.15
0.14
0.14
0.13
0.10
PARTY BUNTING
WHITE HANGING HEART T-LIGHT HOLDER
REGENCY CAKESTAND 3 TIER
PAPER CHAIN KIT EMPIRE
PACK OF 72 RETROSPOT CAKE CASES
Item


August 5 most frequently purchased
Proportion of Total Baskets
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.13
0.12
0.12
0.11
0.11
JUMBO BAG RED RETROSPOT
SPOTTY BUNTING
LUNCH BAG RED RETROSPOT
PARTY BUNTING
WHITE HANGING HEART T-LIGHT HOLDER
Item


Decembers 5 most frequently purchased
Proportion of Total Baskets
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.12
0.12
0.10
0.09
0.09
WHITE HANGING HEART T-LIGHT HOLDER
PAPER CHAIN KIT 50'S CHRISTMAS
REGENCY CAKESTAND 3 TIER
PAPER CHAIN KIT VINTAGE CHRISTMAS
SCOTTIE DOG HOT WATER BOTTLE
Item

Frequency Analysis:

In this section, we analyze the most frequently purchased items across four different months- January, April, August, and December. By examining these months, we can gain insights into customer purchase patterns and identify potential marketing strategies while comparing them to the yearly exploration data. For example, the White Hanging Heart T-light holder is the most frequent seller in 2 of the 4 months examined and is backed by leading in yearly sales with 2,269.

Looking to identify items with a **frequency rate above 0.10.**

January

Showed the highest frequency proportion of 0.16 White-Hanging Hearts, thus suggesting a preference for decorative items or a redecorating trend at the beginning of the year.

Supporting this pattern is the 3-Cake Tin Pantry and Regency 3-Tier Cake Stand both being at 0.13 proportion furthermore leading to a case of post-holiday redecoration and home improvement.

April

The April frequency chart also included a White-Hanging Heart Holder and a 3-Tier Cake Stand. However, spearheading the chart is the Party Bunting with a proportion of 0.15, which is the second highest frequency out of the four months examined.

Paired with the introduction of the Party Bunting, Paper Chain Kit Empire, and Cake Cases indicated a transition into party supplies and decorations for events.

August

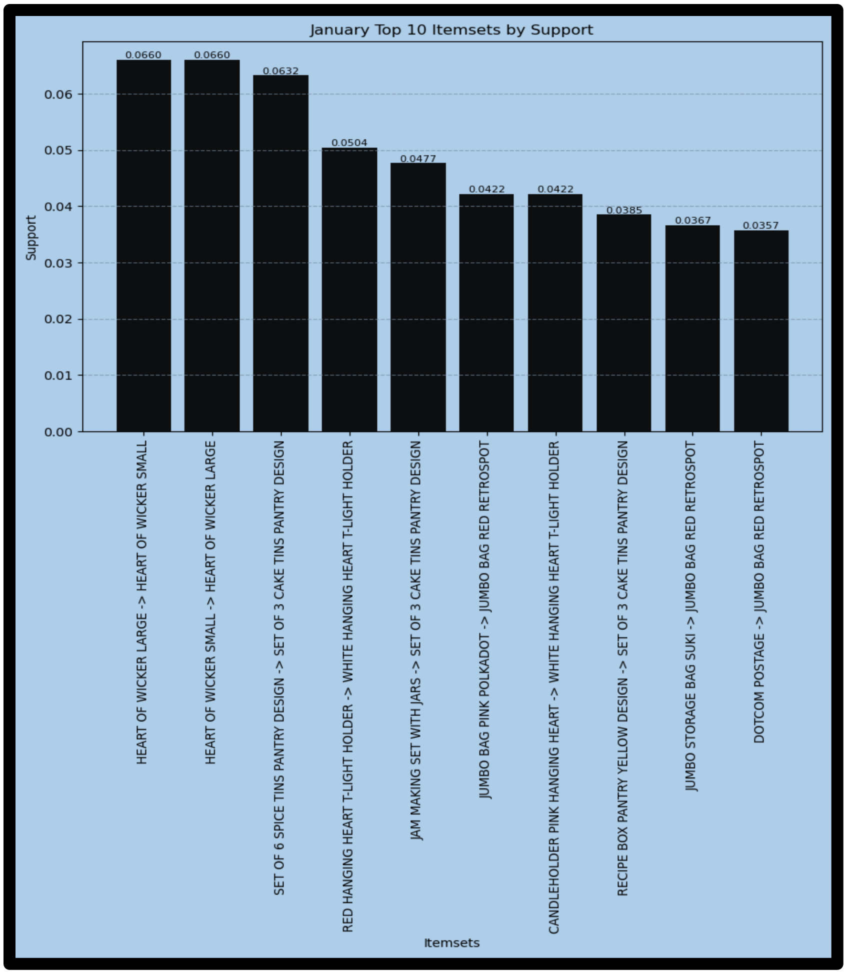
Jumbo Bag Red Retro sport led the way in August with a 0.13 frequency rate not as high as other leading items however there should be an emphasis on inventory and preparation to take full advantage of the uprise in sales.

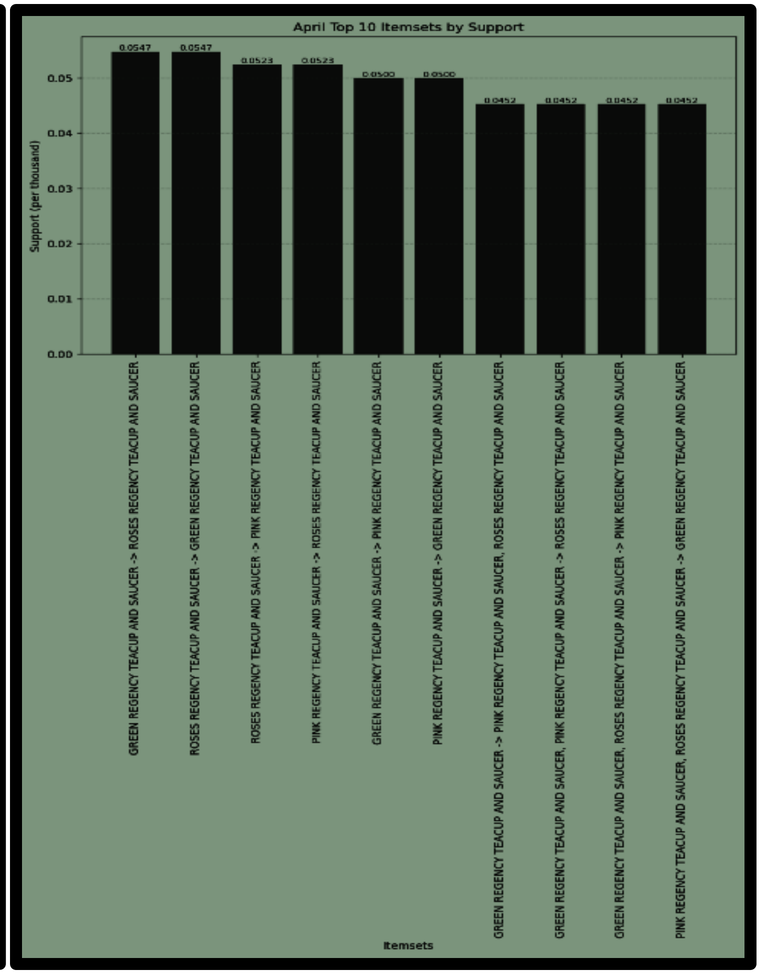
Again, the White Heart Light Holder and Party Bunting are frequently purchased confirming their consistent popularity year wide.

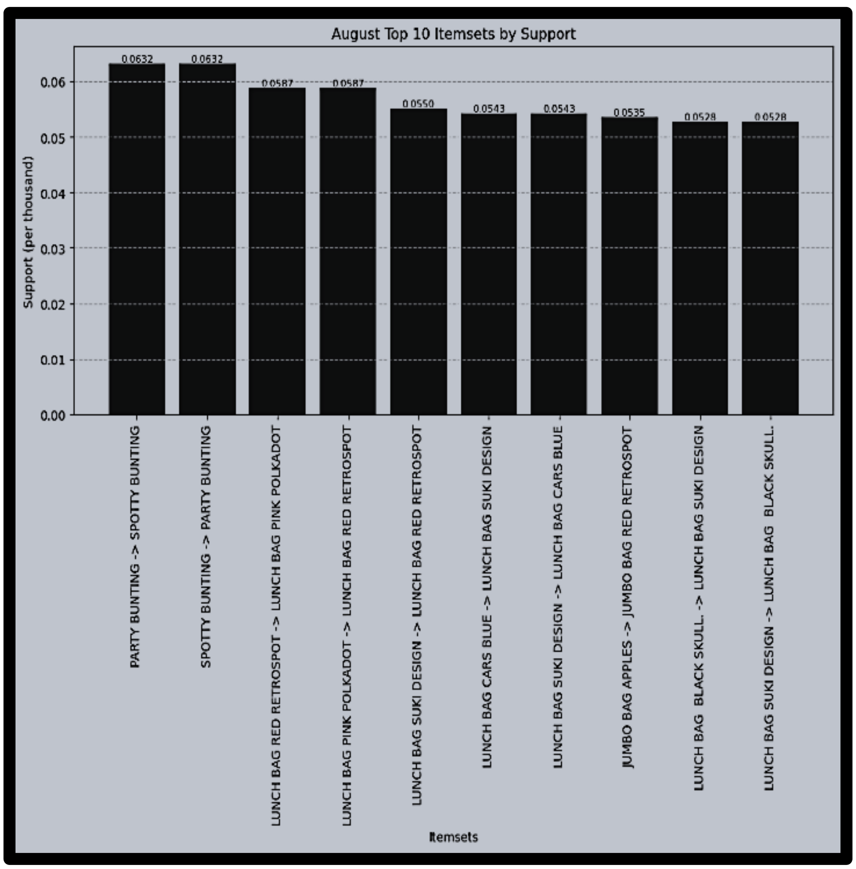
December

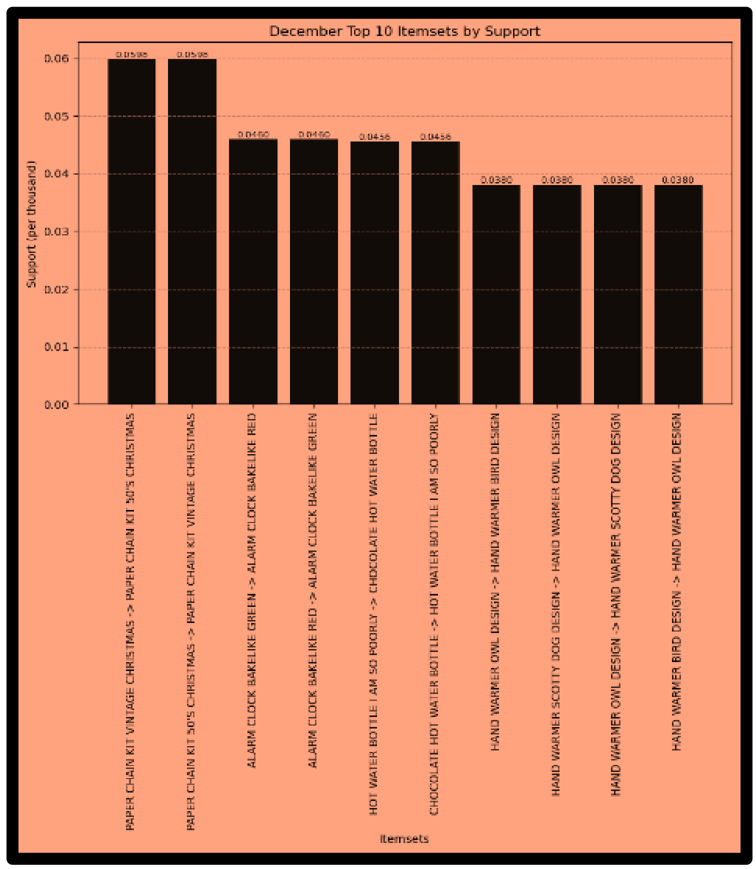
December's analysis highlighted the holidays as expected. The Hanging White Heart Light Holder and Regency Cake Stand resulted in the top 3 reflecting the focus for holiday season decorations.

Lastly, the Scottie Dog Hot Water Bottle at 0.09 proportion suggests an increase in gift purchasing. This new item in the top 5 most frequent sellers hints at the cold weather and its effect on customer purchase behavior. The winter season sees an uptick in the frequency of purchases for items to keep warm like the Scottie Dog Hot Water Bottle and other variations


January Top 10 Itemsets by Support
Support
Itemsets
0.0660
0.0660
0.0632
0.0504
0.0477
0.0422
0.0422
0.0385
0.0367
0.0357
HEART OF WICKER LARGE -> HEART OF WICKER SMALL
HEART OF WICKER SMALL -> HEART OF WICKER LARGE
SET OF 6 SPICE TINS PANTRY DESIGN -> SET OF 3 CAKE TINS PANTRY DESIGN
RED HANGING HEART T-LIGHT HOLDER -> WHITE HANGING HEART T-LIGHT HOLDER
JAM MAKING SET WITH JARS -> SET OF 3 CAKE TINS PANTRY DESIGN
JUMBO BAG PINK POLKADOT -> JUMBO BAG RED RETROSPOT
CANDLEHOLDER PINK HANGING HEART -> WHITE HANGING HEART T-LIGHT HOLDER
RECIPE BOX PANTRY YELLOW DESIGN -> SET OF 3 CAKE TINS PANTRY DESIGN
JUMBO STORAGE BAG SUKI -> JUMBO BAG RED RETROSPOT
DOTCOM POSTAGE -> JUMBO BAG RED RETROSPOT


April Top 10 Itemsets by Support
Support (per thousand)
Itemsets
0.0547
0.0547
0.0523
0.0523
0.0500
0.0500
0.0452
0.0452
0.0452
0.0452
GREEN REGENCY TEACUP AND SAUCER -> ROSES REGENCY TEACUP AND SAUCER
ROSES REGENCY TEACUP AND SAUCER -> GREEN REGENCY TEACUP AND SAUCER
ROSES REGENCY TEACUP AND SAUCER -> PINK REGENCY TEACUP AND SAUCER
PINK REGENCY TEACUP AND SAUCER -> ROSES REGENCY TEACUP AND SAUCER
GREEN REGENCY TEACUP AND SAUCER -> PINK REGENCY TEACUP AND SAUCER
PINK REGENCY TEACUP AND SAUCER -> GREEN REGENCY TEACUP AND SAUCER
GREEN REGENCY TEACUP AND SAUCER -> PINK REGENCY TEACUP AND SAUCER, ROSES REGENCY TEACUP AND SAUCER
GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER -> ROSES REGENCY TEACUP AND SAUCER
GREEN REGENCY TEACUP AND SAUCER, ROSES REGENCY TEACUP AND SAUCER -> PINK REGENCY TEACUP AND SAUCER
PINK REGENCY TEACUP AND SAUCER, ROSES REGENCY TEACUP AND SAUCER -> GREEN REGENCY TEACUP AND SAUCER


August Top 10 Itemsets by Support
Support (per thousand)
Itemsets
0.0632
0.0632
0.0587
0.0587
0.0550
0.0543
0.0543
0.0535
0.0528
0.0528
PARTY BUNTING -> SPOTTY BUNTING
SPOTTY BUNTING -> PARTY BUNTING
LUNCH BAG RED RETROSPOT -> LUNCH BAG PINK POLKADOT
LUNCH BAG PINK POLKADOT -> LUNCH BAG RED RETROSPOT
LUNCH BAG SUKI DESIGN -> LUNCH BAG RED RETROSPOT
LUNCH BAG CARS BLUE -> LUNCH BAG SUKI DESIGN
LUNCH BAG SUKI DESIGN -> LUNCH BAG CARS BLUE
JUMBO BAG APPLES -> JUMBO BAG RED RETROSPOT
LUNCH BAG BLACK SKULL. -> LUNCH BAG SUKI DESIGN
LUNCH BAG SUKI DESIGN -> LUNCH BAG BLACK SKULL.


December Top 10 Itemsets by Support
Support (per thousand)
Itemsets
0.0598
0.0598
0.0460
0.0460
0.0456
0.0456
0.0380
0.0380
0.0380
0.0380
PAPER CHAIN KIT VINTAGE CHRISTMAS -> PAPER CHAIN KIT 50'S CHRISTMAS
PAPER CHAIN KIT 50'S CHRISTMAS -> PAPER CHAIN KIT VINTAGE CHRISTMAS
ALARM CLOCK BAKELIKE GREEN -> ALARM CLOCK BAKELIKE RED
ALARM CLOCK BAKELIKE RED -> ALARM CLOCK BAKELIKE GREEN
HOT WATER BOTTLE I AM SO POORLY -> CHOCOLATE HOT WATER BOTTLE
CHOCOLATE HOT WATER BOTTLE -> HOT WATER BOTTLE I AM SO POORLY
HAND WARMER OWL DESIGN -> HAND WARMER BIRD DESIGN
HAND WARMER SCOTTY DOG DESIGN -> HAND WARMER OWL DESIGN
HAND WARMER OWL DESIGN -> HAND WARMER SCOTTY DOG DESIGN
HAND WARMER BIRD DESIGN -> HAND WARMER OWL DESIGN

High Support analysis:

From the heading, we seek to analyze the item sets with the highest support across different months-January, April, August, and December. A high support analysis not only helps identify frequent items it also helps identify popular item sets or combinations of items. Examining these High-Support item sets will lead to more informed strategic decisions for Product Placement, inventory management, and promotional campaigns. Each month's analysis highlights the top pairings while uncovering a new pattern and trend seasonally the firm can capitalize on.

Looking for **Support greater than = 0.05**

<u>January</u>

Spearheading January's top support item sets are Heart of Wicker Large → Heart of Wicker Small at 0.0660; at such a high probability of purchasing together, it is in the firm's best interest to make sure these items are marketed together.

The second most popular combination was the 6 Spice Tins Pantry Design → Set of 3 Cake Tins pantry design at 0.0632; The top 3 most popular item sets reinforce the trend of house improvement and redecoration to start the year.

<u>April</u>

April produced a different set of high-support item sets. At the top, Green Regency Teacup and Saucer→ Pink Regency Teacup and Saucer (0.0547).

The next four high-support item sets are different variations of colors of the same teacup saucer set. Thus, indicating a huge interest in the Regency Teacup and Saucers sets in April.

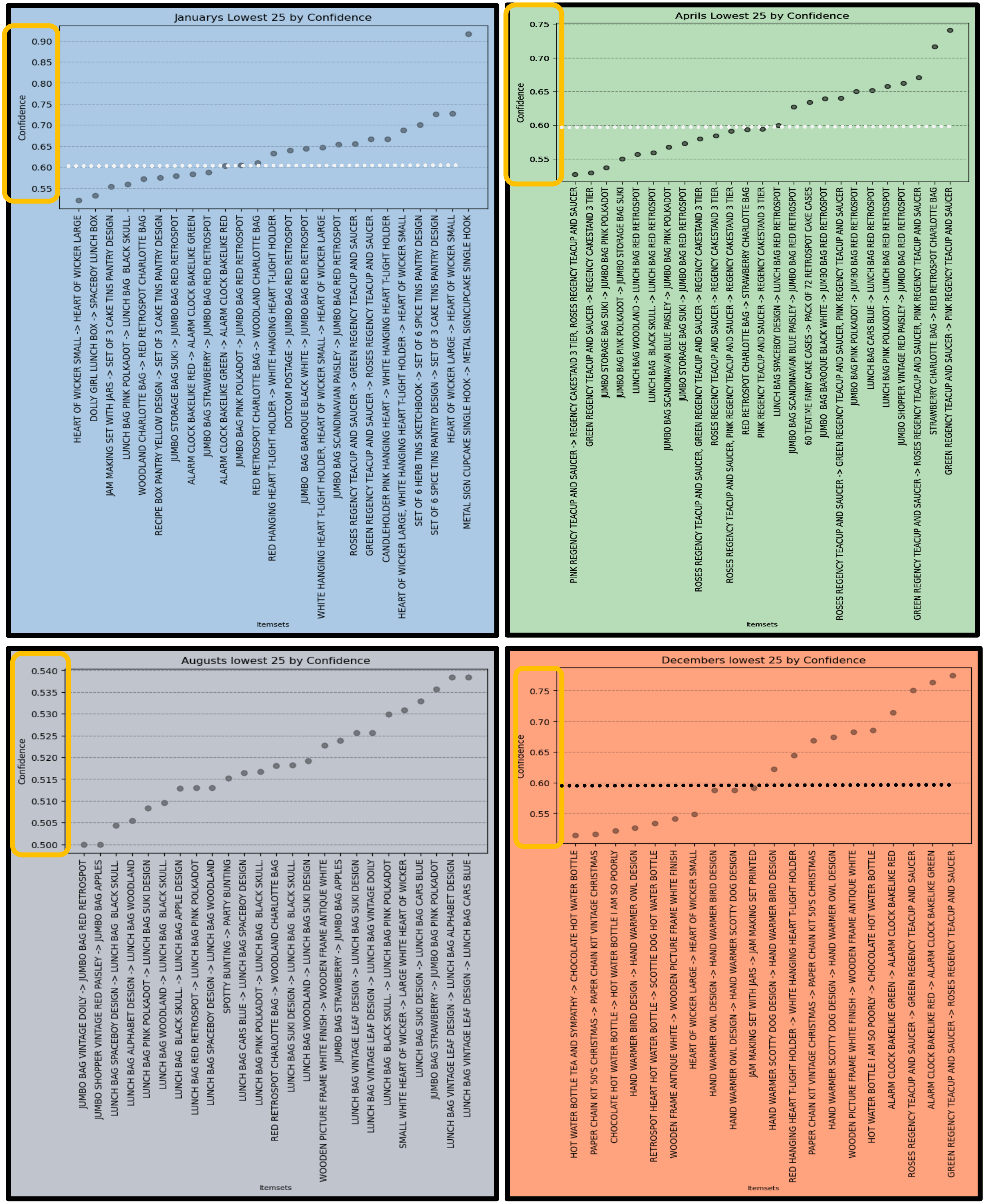
<u>August</u>

In August, Party Bunting→Spotty Bunting item set led the list at a 0.632 probability of being purchased together. Indicating a late push for summer events and transitioning into settling into the school year. For example, the next highest probability item set is Lunch Bag Red Retro Sport → Jumbo Bag Red Retro Sport.

<u>December</u>

As expected, December sales revolved around Christmas and the holiday season. At the top, the combination of Paper Chain Kit Vintage Christmas→ Paper Chain Kit 50's Christmas at 0.0598. This is a weak leading itemset for December. There should be a higher emphasis on taking advantage of the holiday season and common purchasing trends.

In other leading item sets like the handwarmer owl design → handwarmer bird design at 0.0380 is the 7th highest support itemset. However, that probability does not satisfy the goal of 0.05 in such a


Januarys Lowest 25 by Confidence
Confidence
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
HEART OF WICKER SMALL -> HEART OF WICKER LARGE
DOLLY GIRL LUNCH BOX -> SPACEBOY LUNCH BOX
JAM MAKING SET WITH JARS -> SET OF 3 CAKE TINS PANTRY DESIGN
LUNCH BAG PINK POLKADOT -> LUNCH BAG BLACK SKULL.
WOODLAND CHARLOTTE BAG -> RED RETROSPOT CHARLOTTE BAG
RECIPE BOX PANTRY YELLOW DESIGN -> SET OF 3 CAKE TINS PANTRY DESIGN
JUMBO STORAGE BAG SUKI -> JUMBO BAG RED RETROSPOT
ALARM CLOCK BAKELIKE RED -> ALARM CLOCK BAKELIKE GREEN
JUMBO BAG STRAWBERRY -> JUMBO BAG RED RETROSPOT
ALARM CLOCK BAKELIKE GREEN -> ALARM CLOCK BAKELIKE RED
JUMBO BAG PINK POLKADOT -> JUMBO BAG RED RETROSPOT
RED RETROSPOT CHARLOTTE BAG -> WOODLAND CHARLOTTE BAG
RED HANGING HEART T-LIGHT HOLDER -> WHITE HANGING HEART T-LIGHT HOLDER
DOTCOM POSTAGE -> JUMBO BAG RED RETROSPOT
JUMBO BAG BAROQUE BLACK WHITE -> JUMBO BAG RED RETROSPOT
WHITE HANGING HEART T-LIGHT HOLDER, HEART OF WICKER SMALL -> HEART OF WICKER LARGE
JUMBO BAG SCANDINAVIAN PAISLEY -> JUMBO BAG RED RETROSPOT
ROSES REGENCY TEACUP AND SAUCER -> GREEN REGENCY TEACUP AND SAUCER
GREEN REGENCY TEACUP AND SAUCER -> ROSES REGENCY TEACUP AND SAUCER
CANDLEHOLDER PINK HANGING HEART -> WHITE HANGING HEART T-LIGHT HOLDER
HEART OF WICKER LARGE, WHITE HANGING HEART T-LIGHT HOLDER -> HEART OF WICKER SMALL
SET OF 6 HERB TINS SKETCHBOOK -> SET OF 6 SPICE TINS PANTRY DESIGN
SET OF 6 SPICE TINS PANTRY DESIGN -> SET OF 3 CAKE TINS PANTRY DESIGN
HEART OF WICKER LARGE -> HEART OF WICKER SMALL
METAL SIGN CUPCAKE SINGLE HOOK -> METAL SIGNCUPCAKE SINGLE HOOK
Itemsets
Aprils Lowest 25 by Confidence
Confidence
0.75
0.70
0.65
0.60
0.55
PINK REGENCY TEACUP AND SAUCER -> REGENCY CAKESTAND 3 TIER, ROSES REGENCY TEACUP AND SAUCER
GREEN REGENCY TEACUP AND SAUCER -> REGENCY CAKESTAND 3 TIER
JUMBO STORAGE BAG SUKI -> JUMBO BAG PINK POLKADOT
JUMBO BAG PINK POLKADOT -> JUMBO STORAGE BAG SUKI
LUNCH BAG WOODLAND -> LUNCH BAG RED RETROSPOT
LUNCH BAG BLACK SKULL. -> LUNCH BAG RED RETROSPOT
JUMBO BAG SCANDINAVIAN BLUE PAISLEY -> JUMBO BAG PINK POLKADOT
JUMBO STORAGE BAG SUKI -> JUMBO BAG RED RETROSPOT
ROSES REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER -> REGENCY CAKESTAND 3 TIER
ROSES REGENCY TEACUP AND SAUCER -> REGENCY CAKESTAND 3 TIER
ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER -> REGENCY CAKESTAND 3 TIER
RED RETROSPOT CHARLOTTE BAG -> STRAWBERRY CHARLOTTE BAG
PINK REGENCY TEACUP AND SAUCER -> REGENCY CAKESTAND 3 TIER
LUNCH BAG SPACEBOY DESIGN -> LUNCH BAG RED RETROSPOT
JUMBO BAG SCANDINAVIAN BLUE PAISLEY -> JUMBO BAG RED RETROSPOT
60 TEATIME FAIRY CAKE CASES -> PACK OF 72 RETROSPOT CAKE CASES
JUMBO BAG BAROQUE BLACK WHITE -> JUMBO BAG RED RETROSPOT
ROSES REGENCY TEACUP AND SAUCER -> GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER
JUMBO BAG PINK POLKADOT -> JUMBO BAG RED RETROSPOT
LUNCH BAG CARS BLUE -> LUNCH BAG RED RETROSPOT
LUNCH BAG PINK POLKADOT -> LUNCH BAG RED RETROSPOT
JUMBO SHOPPER VINTAGE RED PAISLEY -> JUMBO BAG RED RETROSPOT
GREEN REGENCY TEACUP AND SAUCER -> ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER
STRAWBERRY CHARLOTTE BAG -> RED RETROSPOT CHARLOTTE BAG
GREEN REGENCY TEACUP AND SAUCER -> PINK REGENCY TEACUP AND SAUCER
Itemsets
Augusts lowest 25 by Confidence
Confidence
0.540
0.535
0.530
0.525
0.520
0.515
0.510
0.505
0.500
JUMBO BAG VINTAGE DOILY -> JUMBO BAG RED RETROSPOT
JUMBO SHOPPER VINTAGE RED PAISLEY -> JUMBO BAG APPLES
LUNCH BAG SPACEBOY DESIGN -> LUNCH BAG BLACK SKULL.
LUNCH BAG ALPHABET DESIGN -> LUNCH BAG WOODLAND
LUNCH BAG PINK POLKADOT -> LUNCH BAG SUKI DESIGN
LUNCH BAG WOODLAND -> LUNCH BAG BLACK SKULL.
LUNCH BAG BLACK SKULL. -> LUNCH BAG APPLE DESIGN
LUNCH BAG RED RETROSPOT -> LUNCH BAG PINK POLKADOT
LUNCH BAG SPACEBOY DESIGN -> LUNCH BAG WOODLAND
SPOTTY BUNTING -> PARTY BUNTING
LUNCH BAG CARS BLUE -> LUNCH BAG SPACEBOY DESIGN
LUNCH BAG PINK POLKADOT -> LUNCH BAG BLACK SKULL.
RED RETROSPOT CHARLOTTE BAG -> WOODLAND CHARLOTTE BAG
LUNCH BAG SUKI DESIGN -> LUNCH BAG BLACK SKULL.
LUNCH BAG WOODLAND -> LUNCH BAG SUKI DESIGN
WOODEN PICTURE FRAME WHITE FINISH -> WOODEN FRAME ANTIQUE WHITE
JUMBO BAG STRAWBERRY -> JUMBO BAG APPLES
LUNCH BAG VINTAGE LEAF DESIGN -> LUNCH BAG SUKI DESIGN
LUNCH BAG VINTAGE LEAF DESIGN -> LUNCH BAG VINTAGE DOILY
LUNCH BAG BLACK SKULL. -> LUNCH BAG PINK POLKADOT
SMALL WHITE HEART OF WICKER -> LARGE WHITE HEART OF WICKER
LUNCH BAG SUKI DESIGN -> LUNCH BAG CARS BLUE
JUMBO BAG STRAWBERRY -> JUMBO BAG PINK POLKADOT
LUNCH BAG VINTAGE LEAF DESIGN -> LUNCH BAG ALPHABET DESIGN
LUNCH BAG VINTAGE LEAF DESIGN -> LUNCH BAG CARS BLUE
Itemsets
Decembers lowest 25 by Confidence
Confidence
0.75
0.70
0.65
0.60
0.55
HOT WATER BOTTLE TEA AND SYMPATHY -> CHOCOLATE HOT WATER BOTTLE
PAPER CHAIN KIT 50'S CHRISTMAS -> PAPER CHAIN KIT VINTAGE CHRISTMAS
CHOCOLATE HOT WATER BOTTLE -> HOT WATER BOTTLE I AM SO POORLY
HAND WARMER BIRD DESIGN -> HAND WARMER OWL DESIGN
RETROSPOT HEART HOT WATER BOTTLE -> SCOTTIE DOG HOT WATER BOTTLE
WOODEN FRAME ANTIQUE WHITE -> WOODEN PICTURE FRAME WHITE FINISH
HEART OF WICKER LARGE -> HEART OF WICKER SMALL
HAND WARMER OWL DESIGN -> HAND WARMER BIRD DESIGN
HAND WARMER OWL DESIGN -> HAND WARMER SCOTTY DOG DESIGN
JAM MAKING SET WITH JARS -> JAM MAKING SET PRINTED
HAND WARMER SCOTTY DOG DESIGN -> HAND WARMER BIRD DESIGN
RED HANGING HEART T-LIGHT HOLDER -> WHITE HANGING HEART T-LIGHT HOLDER
PAPER CHAIN KIT VINTAGE CHRISTMAS -> PAPER CHAIN KIT 50'S CHRISTMAS
HAND WARMER SCOTTY DOG DESIGN -> HAND WARMER OWL DESIGN
WOODEN PICTURE FRAME WHITE FINISH -> WOODEN FRAME ANTIQUE WHITE
HOT WATER BOTTLE I AM SO POORLY -> CHOCOLATE HOT WATER BOTTLE
ALARM CLOCK BAKELIKE GREEN -> ALARM CLOCK BAKELIKE RED
ROSES REGENCY TEACUP AND SAUCER -> GREEN REGENCY TEACUP AND SAUCER
ALARM CLOCK BAKELIKE RED -> ALARM CLOCK BAKELIKE GREEN
GREEN REGENCY TEACUP AND SAUCER -> ROSES REGENCY TEACUP AND SAUCER
Itemsets

Low Confidence Analysis:

In this section, the analysis seeks insights into weak association item sets- highlighting opportunities for improvement. The goal remains to understand the relationship amongst the unlikely combinations and establish a plan for increased association. Establishing a goal threshold for all items will allow for a more efficient turnover rate and inventory management.

- All months show strong positive linear relationships with their itemset confidence rates
- August has the lowest of the month's confidence itemset peaks
- Identified a pattern of common items with weak associations indicating items usually sold alone.
- Highlighted low-sale trends in the surrounding months of August

<u>January</u>

January, despite having a high support value, the confidence level of purchasing the Heart of Wicker Small (A) → Heart of Wicker Large (B) is the lowest confidence rate itemset with 0.52. This helps identify a clear distinction between the two items suggested, although their association with each other is strong their relationship with other items is weak.

Another interesting item set is Jam Making Jars(A) → Set of 3 Cake Tins Pantry Design (B) with the third lowest for the month at 0.54 demonstrating a good opportunity to use these items in a baking theme bundle.

<u>April</u>

Similarly, April's lowest item set was also 0.52 with the Regency Teacups and Saucer(A) → Green Regency Teacup Saucer.

The Regency teacups, saucers, and the Jumbo Bags often appear in low confidence analysis for April. These items, though with strong support metrics seem to show a tendency to be sold by themselves. The patterns revealed are duplicate items in different colors. Recommend a stronger effort to campaign a bundling of the different colors of the same item expanding reach and versatility.

<u>August</u>

The lowest item set for the months analyzed is Lunch Bag Vintage Doily (A) → Jumbo Bag Red Retrosport (B) at a 0.5 confidence rate. August showed the most interesting results, with all items set <u>within the range of 0.50 and 0.55 confidence rates</u>.

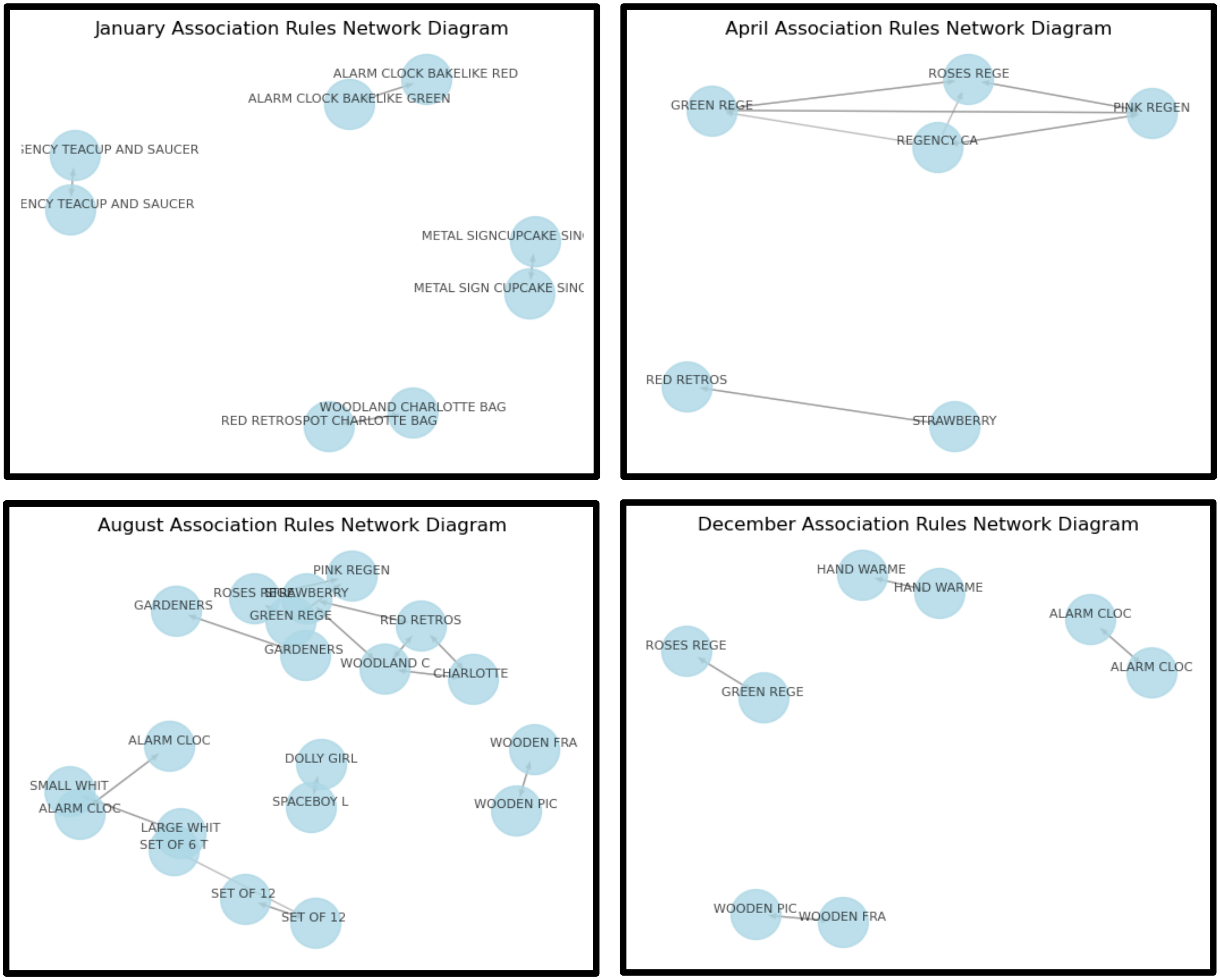
Outside of the variation of lunch bags the Party Bunting (A) → Spotty Bunting (B) at 0.53.

<u>December</u>

With the holiday season in mind, December highlighted unique new items indicative of the season. The weakest relationships of the month included different variations of hot water bottles and cold weather accessories. Along with larger party decorative items such as the paper chain kit 50's Christmas (A)→ paper chain kit vintage (B) at 0.51.

Itemset Paper Chain Kit 50's Christmas (A) → Pack of 72 Retro spot Cake Cases (B) at 0.56 seeks improvement through marketing these items together. Maybe with a Christmas party theme promotion; promoting baking throughout the holidays.

Network Diagram Comparison



Takeaways:

All clusters, Nodes, and Edges are from the strongest relationships with a **9.0 lift threshold**. This signifies almost complete compliance together. Such a high threshold being met indicates these items are best sold together.

August has the most nodes and edges while having the least amount of data entries indicating when customers come in, they are more likely to buy these item sets together

Insights and Recommendations

January

The beginning of the new year followed with a theme of re-invention as highlighted by the decorative home items spike in sales. January is the start of the new year and the transition into home improvement. Its frequency analysis saw items such as a White hanging heart T-light Holder, Heart of Wicker, and Natural Slate Heart Chalkboard. The high-support analysis enforced this theory with strong item sets of Heart of Wicker Small and Large. Low confidence analysis gave clarity on areas of improvement such as the item sets with Heart of Wicker; although a clear distinction between the two items suggested, although their association with each other is strong their relationship with other items is weak.

<u>Recommend</u>: Cross-selling opportunity

April

The transition to April is also the transition into party decorations. Due to the season of April, the party decorations trend could be attributed to graduation season and easter. The initial evidence proves the most frequently purchased item is Party Bunting versus the White Hanging Hearts of January. Secondly, the high support saw various item sets including Regency Teacups and Saucers suggesting a consistent demand. Finally, low-confidence item sets for April highlight the discrepancy of items that are brought frequently but not as pairs.

<u>Recommend</u>: Spring Sale promotion

August

August had the most unique results compared to the other months, suggesting a deeper understanding of summer months. The frequency analysis introduces the Jumbo Bag Red Retro Sport changing the theme of home décor and improvement to more summer energy. The high-support item sets confirm back-to-school season with the uptick in Lunch Bags in all kinds of variations and colors. However, despite the high support for lunch bags the low confidence analysis concluded that these item sets with lunch bags do not always lead to other purchases.

<u>Recommend:</u> Back to School Bundle

December

December saw double the entries and baskets examined, proving to be the best-selling month. The month of December is emblematic of the holiday season including winter weather. Items such as Paper Chain Kits 50's Christmas and Scotties Hot Water Bottle are in the top 5 frequent items, giving clarity to change in customer behavior. Nevertheless, the high-support analysis showed areas to leverage Christmas/holiday-themed items.

<u>Recommend</u>: Holiday-themed promotion

Appendix for charting

Yeary frequency chart

```
plt.figure(figsize=(14, 8))

top_items = transaction_data['Itemname'].value_counts().head(15)

top_items.plot(kind='bar', color='grey')

plt.title('This Years Top 15 Most Purchased Items')

plt.xlabel('Item Name')

plt.ylabel('Frequency')

plt.xticks(rotation=45, ha='right')

for i, freq in enumerate(top_items):

   plt.text(i, freq, str(freq), ha='center', va='bottom', fontsize=8)

plt.show()
```

Frequency chart # 1

```
plt.figure(figsize=(8, 8))

ax = items_frequency[:5].plot(kind='bar', color='skyblue')

plt.title('Item Frequency as Proportion of Total Baskets')

plt.xlabel('Item')

plt.ylabel('Proportion of Total Baskets')

plt.xticks(rotation=90)

for p in ax.patches:

   ax.annotate(str(round(p.get_height(), 2)), (p.get_x() + p.get_width() / 2., p.get_height()),
ha='center', va='center', xytext=(0, 10), textcoords='offset points')

plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.tight_layout()

plt.show()
```

Frequency chart # 2

```
plt.figure(figsize=(6, 3))

ax = items_frequency[:5].plot(kind='bar', color='skyblue', width=0.5)

plt.title('Aprils 5 most frequently purchased')

plt.xlabel('Item')

plt.ylabel('Proportion of Total Baskets')
```

```
plt.xticks(rotation=90)
for i, v in enumerate(items_frequency[:5]):
    ax.text(i, v, f'{v:.2f}', ha='center', va='bottom', fontsize=8, color='black')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

Frequency chart # 3

```
plt.figure(figsize=(6, 3))
ax = items_frequency[:5].plot(kind='bar', color='skyblue', width=0.5)
plt.title('August 5 most frequently purchased')
plt.xlabel('Item')
plt.ylabel('Proportion of Total Baskets')
plt.xticks(rotation=90)
for i, v in enumerate(items_frequency[:5]):
    ax.text(i, v, f'{v:.2f}', ha='center', va='bottom', fontsize=8, color='black')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

Frequency chart # 4

```
plt.figure(figsize=(6, 3))
ax = items_frequency[:5].plot(kind='bar', color='skyblue', width=0.5)
plt.title('Decembers 5 most frequently purchased')
plt.xlabel('Item')
plt.ylabel('Proportion of Total Baskets')
plt.xticks(rotation=90)
for i, v in enumerate(items_frequency[:5]):
    ax.text(i, v, f'{v:.2f}', ha='center', va='bottom', fontsize=8, color='black')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

High Support Chart # 1

```
top_n_itemsets = rules.sort_values(by='support', ascending=False).head(10)
plt.figure(figsize=(10, 6))
bars = plt.bar(top_n_itemsets['itemsets'], top_n_itemsets['support'], color='blue')
for bar, support in zip(bars, top_n_itemsets['support']):
    plt.text(bar.get_x() + bar.get_width() / 2, bar.get_height(), f'{support * 1:.4f}',
            ha='center', va='bottom', color='black', fontsize=8)
plt.title('January Top 10 Itemsets by Support')
plt.xlabel('Itemsets')
plt.ylabel('Support')
plt.xticks(rotation=90)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

High Support Chart # 2

```
top_n_itemsets = rules.sort_values(by='support', ascending=False).head(10)
plt.figure(figsize=(10, 6))
bars = plt.bar(top_n_itemsets['itemsets'], top_n_itemsets['support'], color='blue')
for bar, support in zip(bars, top_n_itemsets['support']):
    plt.text(bar.get_x() + bar.get_width() / 2, bar.get_height(), f'{support * 1:.4f}',
            ha='center', va='bottom', color='black', fontsize=8)
plt.title('April Top 10 Itemsets by Support')
plt.xlabel('Itemsets')
plt.ylabel('Support (per thousand)')
plt.xticks(rotation=90)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

High Support Chart # 3

```
top_n_itemsets = rules.sort_values(by='support', ascending=False).head(10)
```

```
plt.figure(figsize=(10, 9))
bars = plt.bar(top_n_itemsets['itemsets'], top_n_itemsets['support'], color='blue')
for bar, support in zip(bars, top_n_itemsets['support']):
    plt.text(bar.get_x() + bar.get_width() / 2, bar.get_height(), f'{support * 1:.4f}',
             ha='center', va='bottom', color='black', fontsize=8)
plt.title('August Top 10 Itemsets by Support')
plt.xlabel('Itemsets')
plt.ylabel('Support (per thousand)')
plt.xticks(rotation=90)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

High Support Chart # 4

```
top_n_itemsets = rules.sort_values(by='support', ascending=False).head(10)
plt.figure(figsize=(10, 6))
bars = plt.bar(top_n_itemsets['itemsets'], top_n_itemsets['support'], color='blue')
for bar, support in zip(bars, top_n_itemsets['support']):
    plt.text(bar.get_x() + bar.get_width() / 2, bar.get_height(), f'{support * 1:.4f}',
             ha='center', va='bottom', color='black', fontsize=8)
plt.title('December Top 10 Itemsets by Support')
plt.xlabel('Itemsets')
plt.ylabel('Support (per thousand)')
plt.xticks(rotation=90)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

Low-Confidence chart # 1

```
lowest_25_rules = rules.sort_values(by='confidence', ascending=True).head(25)
plt.figure(figsize=(8, 4))
```

```
plt.scatter(lowest_25_rules['itemsets'], lowest_25_rules['confidence'], color='red', alpha=0.5)

plt.xlabel('Itemsets', fontsize=8)

plt.ylabel('Confidence')

plt.title('Januarys Lowest 25 by Confidence')

plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.xticks(rotation=90)

plt.tight_layout()

plt.show()
```

<u>Low-Confidence chart # 2</u>

```
lowest_25_rules = rules.sort_values(by='confidence', ascending=True).head(25)

plt.figure(figsize=(8, 4))

plt.scatter(lowest_25_rules['itemsets'], lowest_25_rules['confidence'], color='red', alpha=0.5)

plt.xlabel('Itemsets', fontsize=8)

plt.ylabel('Confidence')

plt.title('Aprils Lowest 25 by Confidence')

plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.xticks(rotation=90)

plt.tight_layout()

plt.show()
```

<u>Low-Confidence chart # 3</u>

```
lowest_25_rules = rules.sort_values(by='confidence', ascending=True).head(25)

plt.figure(figsize=(8, 4))

plt.scatter(lowest_25_rules['itemsets'], lowest_25_rules['confidence'], color='red', alpha=0.5)  # Adjust color and transparency as needed

plt.xlabel('Itemsets', fontsize=8)

plt.ylabel('Confidence')

plt.title('Augusts lowest 25 by Confidence')

plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.xticks(rotation=90)

plt.tight_layout()
```

plt.show()

Low-Confidence chart # 4

```
lowest_25_rules = rules.sort_values(by='confidence', ascending=True).head(25)

plt.figure(figsize=(8, 4))

plt.scatter(lowest_25_rules['itemsets'], lowest_25_rules['confidence'], color='red', alpha=0.5)  #
Adjust color and transparency as needed

plt.xlabel('Itemsets', fontsize=8)

plt.ylabel('Confidence')

plt.title('Decembers lowest 25 by Confidence')

plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.xticks(rotation=90)

plt.tight_layout()

plt.show()
```

Network Diagram # 1

```
threshold_lift = 9

filtered_rules = rules[rules['lift'] >= threshold_lift]

G = nx.DiGraph()

for _, row in filtered_rules.iterrows():

    for item in row['antecedents']:

        G.add_node(item)

    for item in row['consequents']:

        G.add_node(item)

for _, row in filtered_rules.iterrows():

    for antecedent in row['antecedents']:

        for consequent in row['consequents']:

            G.add_edge(antecedent, consequent)

pos = nx.spring_layout(G, scale=1000, k=1/len(G)**.1)

nx.draw_networkx_nodes(G, pos, node_color='lightblue', node_size=1000, alpha=0.8)

nx.draw_networkx_edges(G, pos, edge_color='gray', width=1, arrowsize=8, alpha=0.5)

nx.draw_networkx_labels(G, pos, font_size=8, font_family='sans-serif', alpha=0.7,
verticalalignment='bottom')
```

```
plt.title('January Association Rules Network Diagram')
plt.axis('off')
plt.show()
```

Network Diagram # 2

```
threshold_lift = 9
filtered_rules = rules[rules['lift'] >= threshold_lift]
G = nx.DiGraph()
for _, row in filtered_rules.iterrows():
  for item in row['antecedents']:
    G.add_node(item)
  for item in row['consequents']:
    G.add_node(item)
for _, row in filtered_rules.iterrows():
  for antecedent in row['antecedents']:
    for consequent in row['consequents']:
      G.add_edge(antecedent, consequent)
pos = nx.spring_layout(G, scale=1000, k=1/len(G)**.1)
nx.draw_networkx_nodes(G, pos, node_color='lightblue', node_size=1000, alpha=0.8)
nx.draw_networkx_edges(G, pos, edge_color='gray', width=1, arrowsize=8, alpha=0.5)
nx.draw_networkx_labels(G, pos, labels={node: node[:10] for node in G.nodes()}, font_size=8,
font_family='sans-serif', alpha=0.7, verticalalignment='bottom')
plt.title('April Association Rules Network Diagram')
plt.axis('off')
plt.show()
```

Network Diagram # 3

```
threshold_lift = 9
filtered_rules = rules[rules['lift'] >= threshold_lift]
G = nx.DiGraph()
for _, row in filtered_rules.iterrows():
```

```
  for item in row['antecedents']:
    G.add_node(item)
  for item in row['consequents']:
    G.add_node(item)
for _, row in filtered_rules.iterrows():
  for antecedent in row['antecedents']:
    for consequent in row['consequents']:
      G.add_edge(antecedent, consequent)
pos = nx.spring_layout(G, scale=1000, k=1/len(G)**.1)
nx.draw_networkx_nodes(G, pos, node_color='lightblue', node_size=1000, alpha=0.8)
nx.draw_networkx_edges(G, pos, edge_color='gray', width=1, arrowsize=8, alpha=0.5)
nx.draw_networkx_labels(G, pos, labels={node: node[:10] for node in G.nodes()}, font_size=8,
font_family='sans-serif', alpha=0.7, verticalalignment='bottom')
plt.title('August Association Rules Network Diagram')
plt.axis('off')
plt.show()
```

Network Diagram # 4

```
threshold_lift = 9
filtered_rules = rules[rules['lift'] >= threshold_lift]
G = nx.DiGraph()
for _, row in filtered_rules.iterrows():
  for item in row['antecedents']:
    G.add_node(item)
  for item in row['consequents']:
    G.add_node(item)
for _, row in filtered_rules.iterrows():
  for antecedent in row['antecedents']:
    for consequent in row['consequents']:
      G.add_edge(antecedent, consequent)
pos = nx.spring_layout(G, scale=1000, k=1/len(G)**.1)
```

```
nx.draw_networkx_nodes(G, pos, node_color='lightblue', node_size=1000, alpha=0.8)

nx.draw_networkx_edges(G, pos, edge_color='gray', width=1, arrowsize=8, alpha=0.5)

nx.draw_networkx_labels(G, pos, labels={node: node[:10] for node in G.nodes()}, font_size=8,
font_family='sans-serif', alpha=0.7, verticalalignment='bottom')

plt.title('December Association Rules Network Diagram')

plt.axis('off')

plt.show()
```