Course Name: Machine Learning

Topic: Market Basket Analysis using Association Rule Mining

Submitted by:

Akaash Chatterjee
Roll No: M24CSE002

Aman Saini
Roll No: M24CSE003

Bera Swaminath Ansuman Sabita
Roll No: M24CSE007

Department of Computer Science and Engineering
Indian Institute of Technology Jodhpur

2024

Abstract

This project leverages Market Basket Analysis (MBA) to uncover meaningful associations in transactional data from multiple regions (Germany, France, Spain, and Ireland). Using the Apriori algorithm with various support and confidence thresholds, association rules were generated to identify product co-purchasing patterns and opportunities for strategic marketing. Key metrics, such as support, confidence, and lift, guided the analysis to reveal actionable insights for region-specific marketing initiatives. For instance, common products with high lift and confidence values—like snack boxes and postage in Germany and France or themed sets in Spain and Ireland—suggest strong cross-selling and bundling potential. Additionally, items associated with high lift indicate natural product complements, providing a basis for bundled promotions or targeted event-based marketing. Across regions, postage frequently emerged as a consequential item, highlighting opportunities for its cross-selling. By applying these insights, businesses can enhance customer experience, maximize order value, and increase sales through tailored, data-driven marketing strategies. The results demonstrate how MBA, as an analytical tool, can be employed to inform retail strategies, optimize promotional campaigns, and improve inventory management.

Contents

1 **2**
1.1 Introduction . . . 2
1.2 Objective . . . 3

2 **4**
2.1 Methodology . . . 4
2.1.1 Dataset Description . . . 4
2.1.2 Data Loading and Pre-processing . . . 5
2.1.3 Exploratory Data Analysis . . . 6
2.1.4 Apriori Algorithm Application . . . 7
2.2 Results . . . 10
2.3 Analysis . . . 34
2.4 Conclusion . . . 37
2.4.1 Country-wise Conclusions . . . 37
2.4.2 Final Conclusion . . . 38

Chapter 1

1.1 Introduction

Market Basket Analysis (MBA) is a data mining technique widely used in retail and e-commerce to uncover relationships between products that customers frequently purchase together. The goal of MBA is to identify patterns or associations within transactional data that can be leveraged to improve sales strategies, product placement, inventory management, and personalized marketing. One of the most common algorithms used for this purpose is the Apriori algorithm, which efficiently identifies frequent itemsets and generates association rules based on the concepts of support, confidence, and lift.
Apriori Algorithm: The Apriori algorithm finds collections of commodities that appear frequently in a dataset of transactions. It finds smaller itemsets that satisfy a minimal support criterion and then merges them into larger ones. Iteratively creating larger and larger itemsets, the procedure continues until no more frequently occurring itemsets are discovered. By taking a "bottom-up" approach, the algorithm filters out irrelevant item sets and only considers the most common ones. Looking for patterns in transaction data, Apriori helps us determine which things are typically bought together, to put it simply.
There are four key concepts used in apriori algorithms:

- **Support:** Support is a measure of how often a group of objects, known as an itemset, appears in the dataset of transactions. If an item is too rare to be significant, it can be filtered out.

  In a mathematical context, support is defined as:

  $$\text{Support(Itemset } X) = \frac{\text{Number of transactions containing } X}{\text{Total number of transactions}}$$

  where X represents the itemset in question.

- **Confidence:** Confidence indicates the probability that a transaction including itemset A will also include itemset B. It measures the reliability of the rule.

  In a formal sense, confidence is defined as:

  $$\text{Confidence}(A \to B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

  where $A \to B$ represents the association rule from itemset A to itemset B.

- **Lift:** Lift measures how much more likely itemset B is purchased when itemset A is also purchased, compared to if they were independent. It helps identify non-trivial associations between itemsets by evaluating the increase in likelihood of purchasing B when A is purchased.

Formally, lift is defined as:

$$\text{Lift}(A \to B) = \frac{\text{Confidence}(A \to B)}{\text{Support}(B)}$$

A lift value greater than 1 indicates a positive association, meaning that items A and B are more likely to be purchased together than by chance.

- **Conviction:** Conviction measures the strength of an association rule by comparing the expected frequency of itemset B not occurring when A occurs, assuming A and B are independent. It indicates how strongly the absence of B contradicts the presence of A, making it a useful metric for identifying dependency relationships.

  Conviction is defined as:

  $$\text{Conviction}(A \to B) = \frac{1 - \text{Support}(B)}{1 - \text{Confidence}(A \to B)}$$

  A higher conviction value indicates a stronger association, especially when the confidence is high and support for B is low.

This project focuses on performing market basket analysis using the online retail dataset and applying the Apriori algorithm to uncover associations between products. The dataset contains transactional data from an online retailer with a wide range of products and customers across different regions. For this analysis, the dataset is segmented based on the "Country" column, allowing for a comparative study of customer purchasing behavior in four selected countries.
By applying the Apriori algorithm from scratch, the project aims to explore frequent product combinations and derive meaningful association rules for each country individually. These insights can be used to understand regional differences in consumer preferences, allowing businesses to tailor their marketing strategies, optimize product placements, and manage inventory more effectively. Through this study, the project seeks to provide valuable insights into cross-country purchasing behavior and offer actionable recommendations for retailers operating in diverse markets.

1.2 Objective

The objective of this project is to perform a Market Basket Analysis on the Online Retail dataset, specifically to uncover patterns in purchasing behavior across four selected countries. Using a custom-built implementation of the Apriori algorithm, this analysis aims to identify significant association rules within transaction data. By segmenting the dataset by country, the project will provide insights into country-specific purchasing trends, enabling a comparative analysis of consumer behavior across regions. The findings will aid in understanding product affinities, guiding tailored marketing strategies and inventory management for each country.

Chapter 2

2.1 Methodology

2.1.1 Dataset Description

We used the *Online Retail* dataset from the UCI Machine Learning Repository for our task This dataset contains transactional records from a UK-based non-store online retail business, covering transactions from December 1, 2010, to December 9, 2011. The retailer primarily sells unique, all-occasion gift items, with a substantial portion of customers being wholesalers. This dataset is valuable for analyzing purchasing patterns and conducting market basket analysis, as it includes detailed information on each transaction, including product details, purchase quantities, and customer locations.

Variable Name	**Role**	**Type**	**Description**	**Units**	**Missing Values**
InvoiceNo	ID	Categorical	A unique 6-digit number assigned to each transaction. Codes starting with 'C' indicate cancellations.	None	No
StockCode	ID	Categorical	A unique 5-digit code assigned to each distinct product.	None	No
Description	Feature	Categorical	Name of the product.	None	Yes
Quantity	Feature	Integer	The quantity of each product per transaction.	None	No
InvoiceDate	Feature	Date	The date and time each transaction was generated.	None	No
UnitPrice	Feature	Continuous	Price per unit of the product, in sterling.	Sterling (£)	No
CustomerID	Feature	Categorical	A unique 5-digit ID assigned to each customer.	None	Yes
Country	Feature	Categorical	Name of the country where each customer resides.	None	No

Table 2.1: Variables in the Online Retail Dataset

2.1.2 Data Loading and Pre-processing

First we import all the necessary libraries for our project

```
#We import the necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from itertools import combinations
import networkx as nx
```

Next,we load our xlsx file into our dataframe with the help of the openpyxl engine.

```
df = pd.read_excel('Online Retail.xlsx',engine='openpyxl')
```

We then check for null values in our data and subsequently drop any rows containing null values.

```
df.isnull().sum() #Check for null values in columns
```

	0
InvoiceNo	0
StockCode	0
Description	1454
Quantity	0
InvoiceDate	0
UnitPrice	0
CustomerID	135080
Country	0

dtype: int64

```
df = df.dropna(subset=['Description', 'CustomerID'])
```

```
df.shape #Check Updated no of rows and columns
```

```
(406829, 8)
```

Next,we convert the InvoicenNo column to type string and drop an rows where InvoiceNo begins with c indicating that the transaction was canceled and does not affect our analysis.We further remove blank spaces from the description column entries.

```
df['InvoiceNo'] = df['InvoiceNo'].astype(str) #Co
df = df[~df['InvoiceNo'].str.startswith('C')] #We
```

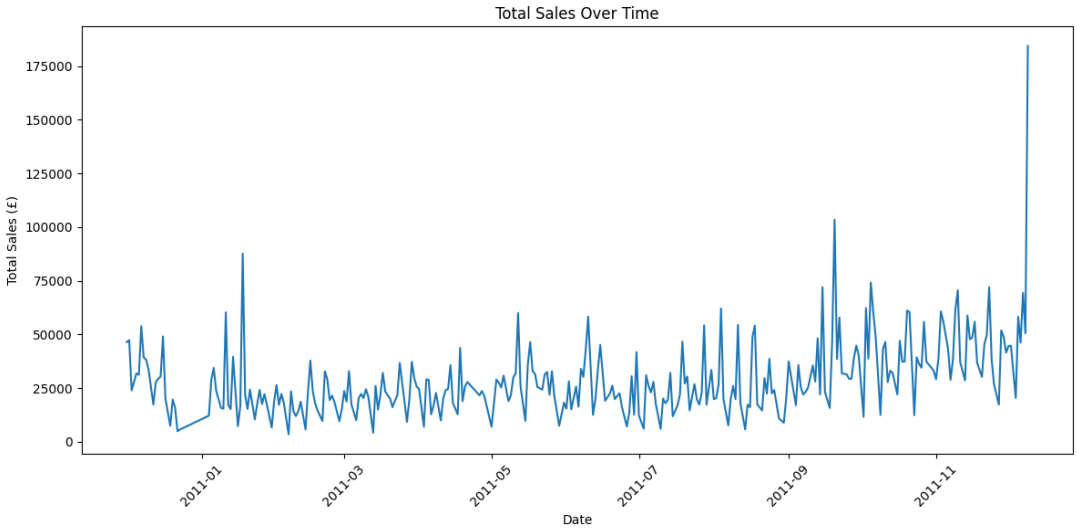
```
df.shape #Updated No of rows and columns
```

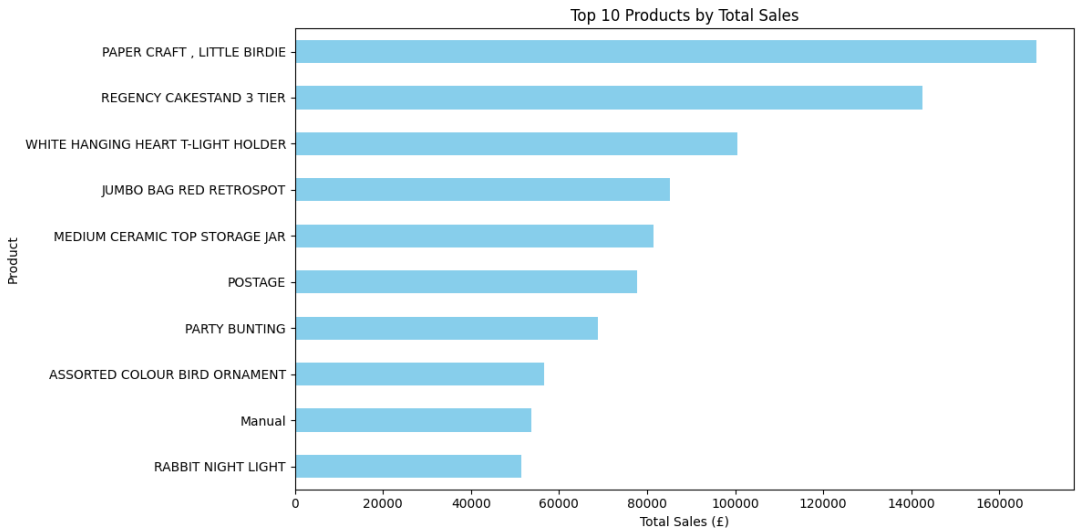
```
(397924, 8)
```

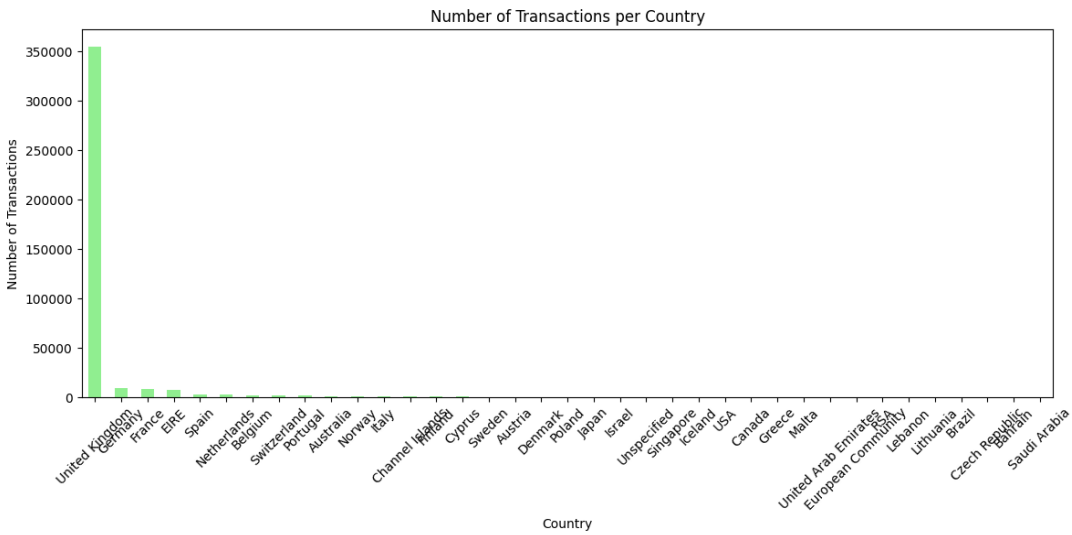
```
df['Description'] = df['Description'].str.strip()
```

2.1.3 Exploratory Data Analysis

We draw a series of necessary plots on the raw data to gain some initial insights into our data which would be further discussed in the results section.







2.1.4 Apriori Algorithm Application

Our goal is to apply apriori algorithm on the data seperated via Country into 4 such countries and generate the corresponding association rule for each.We follow a similar methodology for all 4 countries-Germany,France,Spain and Ireland(ERIN).

- Data Filtering: We filter the main dataset to include only transactions where Country is the one we are testing on.

```
[ ] df1 = df[df['Country'] == 'Germany'] #First we select all rows where the country is Germany
```

- Data Summary: We display the first few rows and data summary of the filtered dataset to verify structure and contents.
- Basket Preparation We then group the data by InvoiceNo and Description to aggregate quantities for each item in a transaction. Next we pivot the dataset to make each item a column, fill missing values with zero, and set InvoiceNo as the index.

```
[ ] #Creating the basket to apply apriori algorithm on
    basket_1 = df1.groupby(['InvoiceNo', 'Description'])['Quantity'].sum().unstack().reset_index().fillna(0).set_index('InvoiceNo')
    def encode_units(x):
        if x <= 0:
            return 0
        if x >= 1:
            return 1
```

- Binary Encoding of Quantities We then define a function to convert item quantities to binary values: 0 if quantity is zero or less, indicating the item was not purchased in the transaction and 1 if quantity is one or more, indicating the item was purchased. Next we apply this encoding to the basket data.

```
    #apply function to data using applymap.
    basket_1 = basket_1.applymap(encode_units)
    basket_1.head()
```

- Test for Various values of minimum support: We set minsupport to 0.05 and 0.10 and analyze the results and choose appropriate metrics to decide for the final rule generation procedure
- Frequent Itemset Generation (Apriori) We initialize a dictionary to store frequent itemsets and their support values. Next we generate 1-itemsets by calculating support for each item and filtering those meeting minsupport.

```python
#Checking for support = 0.05
min_support = 0.05
#APRIORI ALGORITHM
def get_frequent_itemsets(data, min_support):
    itemsets = {}
    num_transactions = len(data)

    # 1-itemsets
    for item in data.columns:
        support = data[item].sum() / num_transactions
        if support >= min_support:
            itemsets[frozenset([item])] = support

    k = 2
    current_itemsets = itemsets

    while current_itemsets:
        # Generate candidate k-itemsets from (k-1)-itemsets
        candidate_itemsets = {}
        keys = list(current_itemsets.keys())
        for i in range(len(keys)):
            for j in range(i + 1, len(keys)):
                union_set = keys[i] | keys[j]
                if len(union_set) == k:
                    support = (data[list(union_set)].min(axis=1) > 0).sum() / num_transactions
                    if support >= min_support:
                        candidate_itemsets[union_set] = support

        # Add candidate k-itemsets to frequent itemsets
        itemsets.update(candidate_itemsets)
        current_itemsets = candidate_itemsets
        k += 1

    return itemsets

frequent_itemsets = get_frequent_itemsets(basket_1, min_support)
```

- Iteratively generate k-itemsets: We combine (k-1)-itemsets to create candidate k-itemsets. next we calculate support for each candidate and retain those meeting minsupport. We repeat until no more frequent itemsets are found.
- Association Rule Generation We define a function to generate association rules from frequent itemsets with variable confidence: For each itemset with more than one item, consider subsets as possible antecedents. We calculate confidence by dividing itemset support by antecedent support. Retain the rules where confidence meets a specified threshold, and calculate the lift.

```python
# Function to generate association rules with variable confidence
def get_association_rules(frequent_itemsets, confidence):
    rules = []
    for itemset in frequent_itemsets.keys():
        if len(itemset) > 1:
            for antecedent in map(frozenset, combinations(itemset, len(itemset) - 1)):
                consequent = itemset - antecedent
                support = frequent_itemsets[itemset]
                antecedent_support = frequent_itemsets[antecedent]
                rule_confidence = support / antecedent_support
                if rule_confidence >= confidence:
                    lift = rule_confidence / frequent_itemsets[consequent]
                    rules.append((antecedent, consequent, support, rule_confidence, lift))
    return pd.DataFrame(rules, columns=['Antecedent', 'Consequent', 'Support', 'Confidence', 'Lift'])
```

- Confidence Level Testing We define a range of confidence levels (0.1 to 0.9) to test. For each confidence level we generate association rules and count the number of rules generated.

```python
# We define confidence levels to test and list to store rule counts
confidence_levels = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
rule_counts = []

# We generate rules for each confidence level and store the count
for confidence in confidence_levels:
    rules = get_association_rules(frequent_itemsets, confidence)
    rule_counts.append(len(rules))
```

- Plotting Results: We plot the number of rules generated at varying confidence levels to visualize the effect of confidence on rule count.This allows us to visualize the results and analyze the best possible metrics for the
- Support and Confidence Selection for Analysis: Based on the analysis,we set minsupport and minconfidence to appropriate values to generate a refined set of rules.
- Display and Analysis of Rules: We then perform some plots and analysis to evaluate our rules and generate necessary inferences.

2.2 Results

- For Germany:

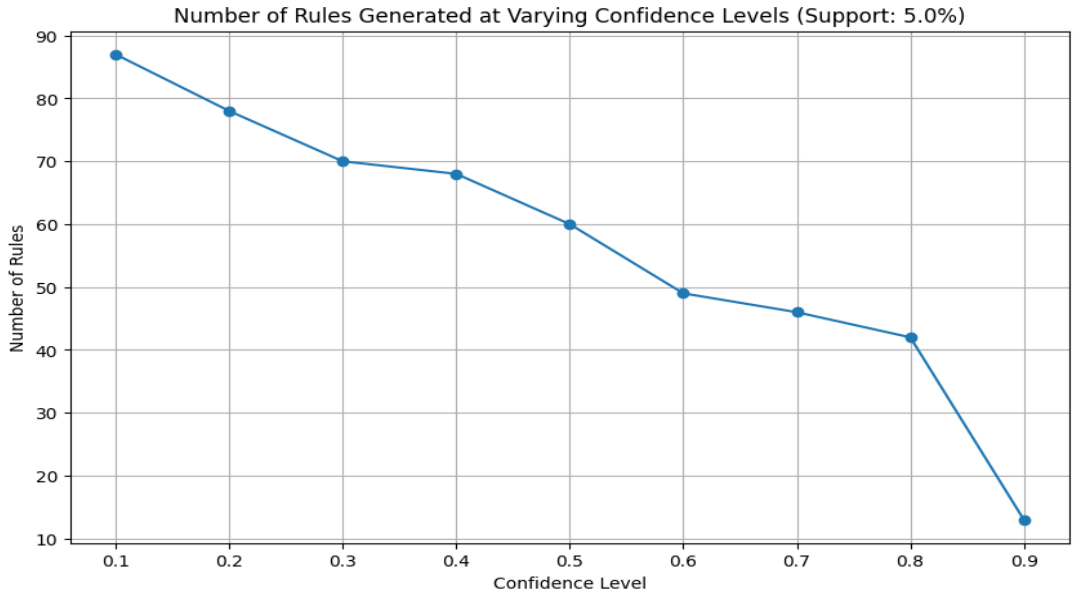


Figure 2.1: Number of rules generated in Germany with 5% support threshold.

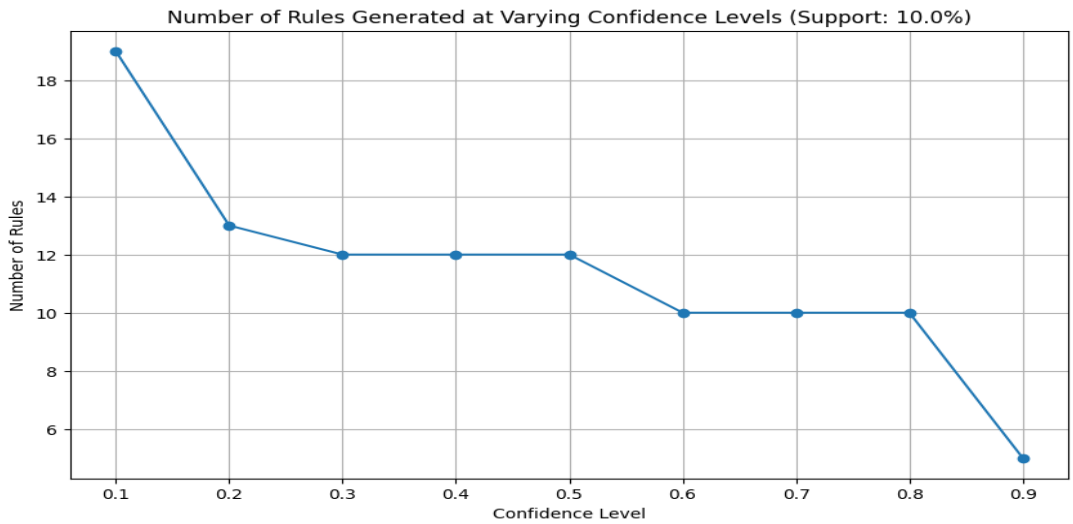


Figure 2.2: Number of rules generated in Germany with 10% support threshold.

	Antecedent	Consequent	Support	Confidence	Lift
0	(6 RIBBONS RUSTIC CHARM)	(POSTAGE)	0.091904	0.893617	1.091933
1	(ALARM CLOCK BAKELIKE PINK)	(POSTAGE)	0.050328	0.718750	0.878259
2	(CHARLOTTE BAG APPLES DESIGN)	(POSTAGE)	0.061269	0.933333	1.140463
3	(COFFEE MUG APPLES DESIGN)	(POSTAGE)	0.054705	0.892857	1.091005
4	(FAWN BLUE HOT WATER BOTTLE)	(POSTAGE)	0.059081	0.931034	1.137654
5	(GUMBALL COAT RACK)	(POSTAGE)	0.056893	0.787879	0.962729
6	(JAM JAR WITH PINK LID)	(POSTAGE)	0.054705	0.862069	1.053384
7	(JAM MAKING SET PRINTED)	(POSTAGE)	0.074398	0.809524	0.989177
8	(JUMBO BAG APPLES)	(POSTAGE)	0.052516	0.857143	1.047364
9	(JUMBO BAG RED RETROSPOT)	(POSTAGE)	0.063457	0.805556	0.984329
10	(JUMBO BAG WOODLAND ANIMALS)	(POSTAGE)	0.087527	0.869565	1.062544
11	(LUNCH BAG APPLE DESIGN)	(POSTAGE)	0.054705	0.892857	1.091005
12	(LUNCH BAG WOODLAND)	(POSTAGE)	0.067834	0.861111	1.052213
13	(LUNCH BOX WITH CUTLERY RETROSPOT)	(POSTAGE)	0.054705	0.862069	1.053384
14	(PACK OF 20 NAPKINS RED APPLES)	(POSTAGE)	0.050328	0.851852	1.040899
15	(PACK OF 72 RETROSPOT CAKE CASES)	(POSTAGE)	0.054705	0.641026	0.783285
16	(PLASTERS IN TIN CIRCUS PARADE)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.067834	0.584906	4.242887
17	(PLASTERS IN TIN CIRCUS PARADE)	(POSTAGE)	0.100656	0.867925	1.060539
18	(PLASTERS IN TIN SPACEBOY)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.061269	0.571429	4.145125
19	(PLASTERS IN TIN SPACEBOY)	(POSTAGE)	0.100656	0.938776	1.147113
20	(PLASTERS IN TIN STRONGMAN)	(POSTAGE)	0.067834	0.968750	1.183740
21	(PLASTERS IN TIN WOODLAND ANIMALS)	(POSTAGE)	0.118162	0.857143	1.047364
22	(PLASTERS IN TIN WOODLAND ANIMALS)	(ROUND SNACK BOXES SET OF4 WOODLAND)	0.074398	0.539683	2.202098

```
rules_apriori_1.shape #Checking the no of rules generated
```

```
(60, 5)
```

Figure 2.3: Distribution of rule metrics in Germany.

Top 10 Rules by Support

	Antecedent	Consequent	Support	Confidence	Lift
31	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND'})	frozenset({'POSTAGE'})	0.225382932166301 98	0.919642857142857 2	1.123734721161191 9
30	frozenset({'ROUND SNACK BOXES SET OF 4 FRUITS'})	frozenset({'POSTAGE'})	0.150984682713347 92	0.958333333333333 3	1.171011586452762 8
40	frozenset({'ROUND SNACK BOXES SET OF 4 FRUITS'})	frozenset({'ROUND SNACK BOXES SET OF 4 WOODLAND'})	0.131291028446389 48	0.833333333333333 3	3.400297619047619
39	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND'})	frozenset({'ROUND SNACK BOXES SET OF 4 FRUITS'})	0.131291028446389 48	0.535714285714285 7	3.400297619047618 6
55	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND', 'POSTAGE'})	frozenset({'ROUND SNACK BOXES SET OF 4 FRUITS'})	0.124726477024070 03	0.553398058252427 2	3.512540453074433 5
54	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND', 'ROUND SNACK BOXES SET OF 4 FRUITS'})	frozenset({'POSTAGE'})	0.124726477024070 03	0.950000000000000 1	1.160828877005347 6
53	frozenset({'POSTAGE', 'ROUND SNACK BOXES SET OF 4 FRUITS'})	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND'})	0.124726477024070 03	0.826086956521739 1	3.370729813664596 5
27	frozenset({'REGENCY CAKESTAND 3 TIER'})	frozenset({'POSTAGE'})	0.120350109409190 37	0.873015873015873	1.066760037348272 6
21	frozenset({'PLASTERS IN TIN WOODLAND ANIMALS'})	frozenset({'POSTAGE'})	0.118161925601750 55	0.857142857142857 2	1.047364400305769
36	frozenset({'WOODLAND CHARLOTTE BAG'})	frozenset({'POSTAGE'})	0.115973741794310 72	0.913793103448275 8	1.116586760095887 9

Top 10 Rules by Confidence

	Antecedent	Consequent	Support	Confidence	Lift
20	frozenset({'PLASTERS IN TIN STRONGMAN'})	frozenset({'POSTAGE'})	0.067833698030636 58	0.968750000000000 1	1.183739973620322
29	frozenset({'RETROSPOT TEA SET CERAMIC 11 PC'})	frozenset({'POSTAGE'})	0.054704595185995 62	0.961538461538461 6	1.174928013163307 3
30	frozenset({'ROUND SNACK BOXES SET OF 4 FRUITS'})	frozenset({'POSTAGE'})	0.150984682713347 92	0.958333333333333 3	1.171011586452762 8
54	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND', 'ROUND SNACK BOXES SET OF 4 FRUITS'})	frozenset({'POSTAGE'})	0.124726477024070 03	0.950000000000000 1	1.160828877005347 6
19	frozenset({'PLASTERS IN TIN SPACEBOY'})	frozenset({'POSTAGE'})	0.100656455142231 95	0.938775510204081 7	1.14711339081087
23	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	frozenset({'POSTAGE'})	0.065645514223194 74	0.9375	1.145554812834224 6
2	frozenset({'CHARLOTTE BAG APPLES DESIGN'})	frozenset({'POSTAGE'})	0.061269146608315 096	0.933333333333333	1.140463458110517
58	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND', 'WOODLAND CHARLOTTE BAG'})	frozenset({'POSTAGE'})	0.059080962800875 28	0.931034482758620 8	1.137654434814678 4
4	frozenset({'FAWN BLUE HOT WATER BOTTLE'})	frozenset({'POSTAGE'})	0.059080962800875 28	0.931034482758620 8	1.137654434814678 4
51	frozenset({'WOODLAND CHARLOTTE BAG', 'RED RETROSPOT CHARLOTTE BAG'})	frozenset({'POSTAGE'})	0.054704595185995 62	0.925925925925925 8	1.131412160823925 4

Top 10 Rules by Lift

	Antecedent	Consequent	Support	Confidence	Lift
38	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	frozenset({'WOODLAND CHARLOTTE BAG'})	0.059080962800875 28	0.843750000000001	6.648168103448277
52	frozenset({'POSTAGE', 'RED RETROSPOT CHARLOTTE BAG'})	frozenset({'WOODLAND CHARLOTTE BAG'})	0.054704595185995 62	0.833333333333334	6.560091954022989
16	frozenset({'PLASTERS IN TIN CIRCUS PARADE'})	frozenset({'PLASTERS IN TIN WOODLAND ANIMALS'})	0.067833698030636 58	0.584905660377358 5	4.242887091943696
18	frozenset({'PLASTERS IN TIN SPACEBOY'})	frozenset({'PLASTERS IN TIN WOODLAND ANIMALS'})	0.061269146608315 096	0.571428571428571 4	4.145124716553288
47	frozenset({'PLASTERS IN TIN SPACEBOY', 'POSTAGE'})	frozenset({'PLASTERS IN TIN WOODLAND ANIMALS'})	0.054704595185995 62	0.543478260869565 2	3.942374051069703
43	frozenset({'PLASTERS IN TIN CIRCUS PARADE', 'POSTAGE'})	frozenset({'PLASTERS IN TIN WOODLAND ANIMALS'})	0.054704595185995 62	0.543478260869565 2	3.942374051069703
55	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND', 'POSTAGE'})	frozenset({'ROUND SNACK BOXES SET OF 4 FRUITS'})	0.124726477024070 03	0.553398058252427 2	3.512540453074433 5
40	frozenset({'ROUND SNACK BOXES SET OF 4 FRUITS'})	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND'})	0.131291028446389 48	0.833333333333333 3	3.400297619047619
39	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND'})	frozenset({'ROUND SNACK BOXES SET OF 4 FRUITS'})	0.131291028446389 48	0.535714285714285 7	3.400297619047618 6
53	frozenset({'POSTAGE', 'ROUND SNACK BOXES SET OF 4 FRUITS'})	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND'})	0.124726477024070 03	0.826086956521739 1	3.370729813664596 5

Figure 2.4: Top 10 association rules in Germany.

Figure 2.5: Support vs. Confidence for association rules in Germany.

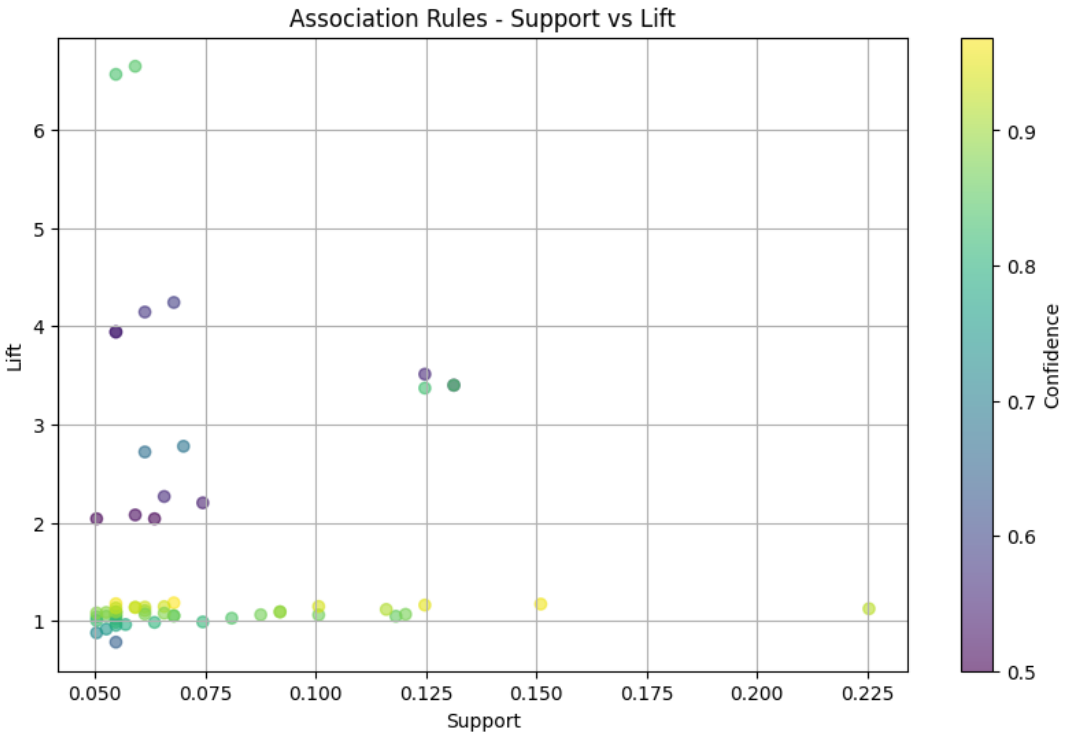


Figure 2.6: Support vs. Lift for association rules in Germany.

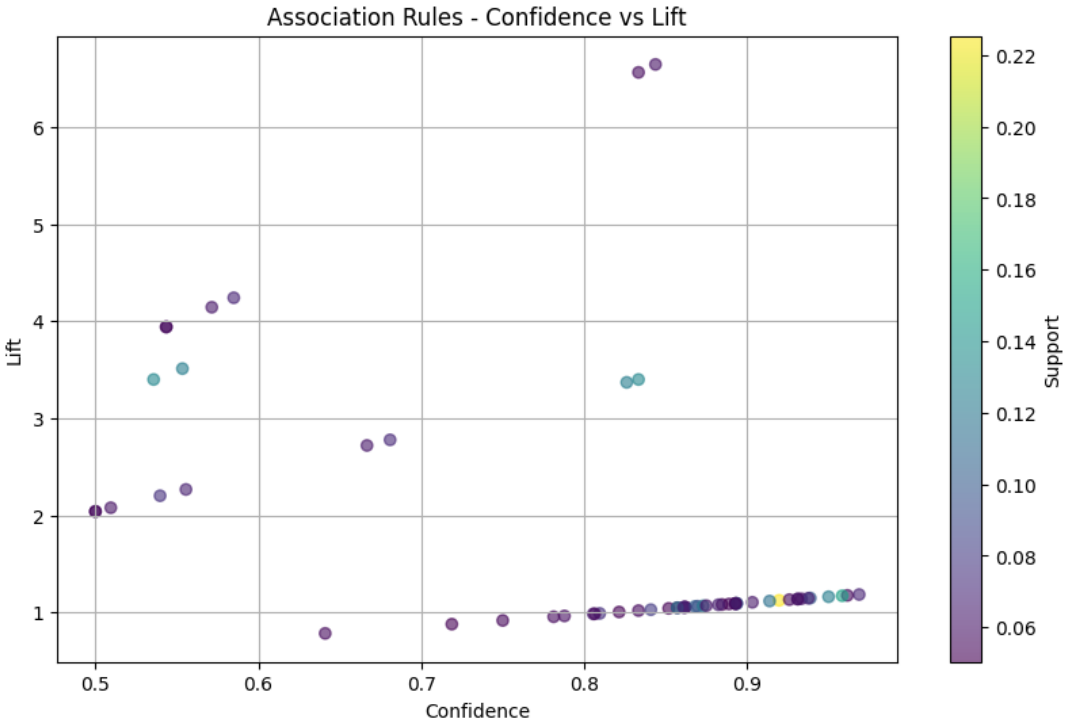


Figure 2.7: Confidence vs. Lift for association rules in Germany.

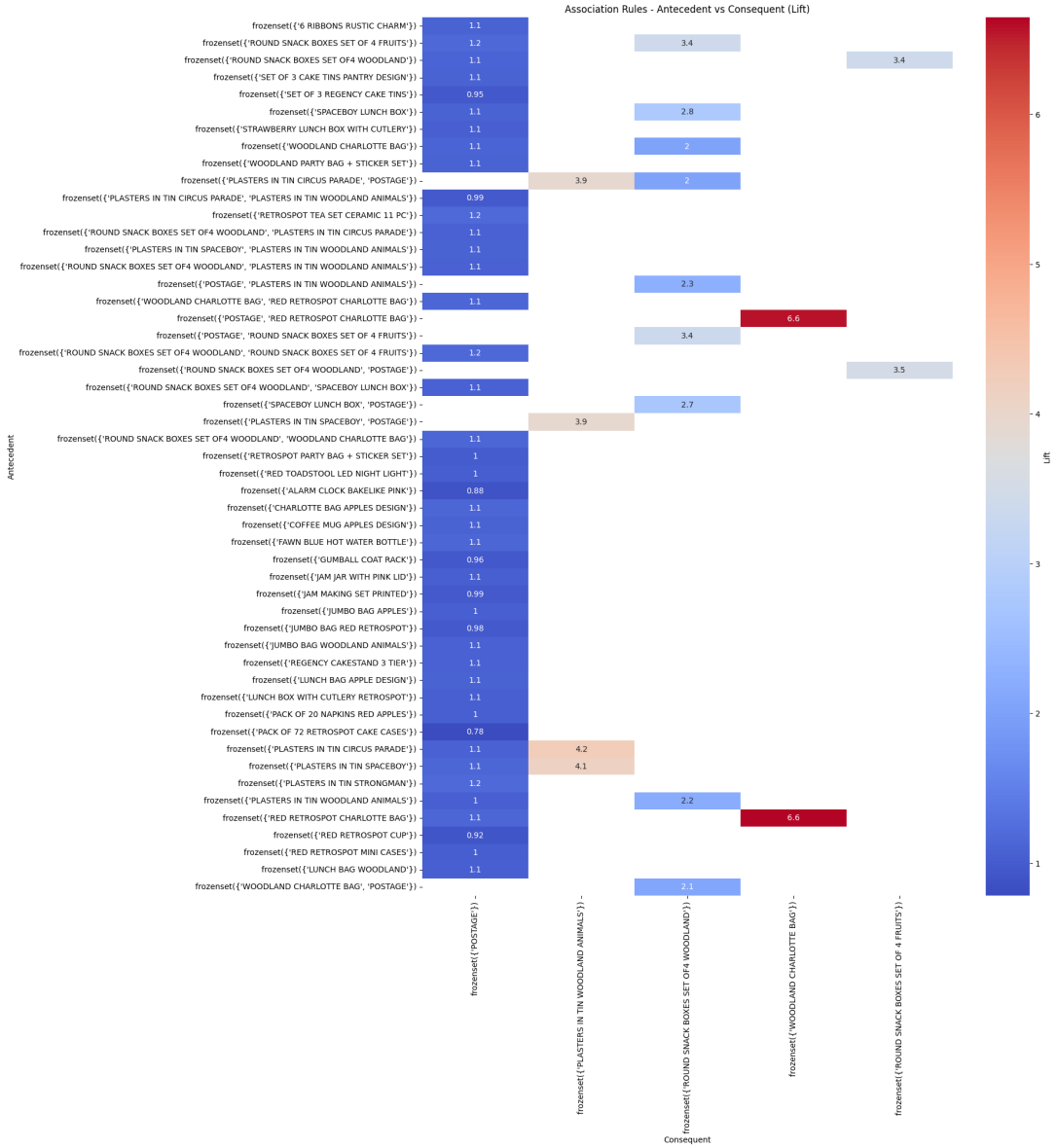


Figure 2.8: Antecedent Lift vs. Consequent Lift in Germany.

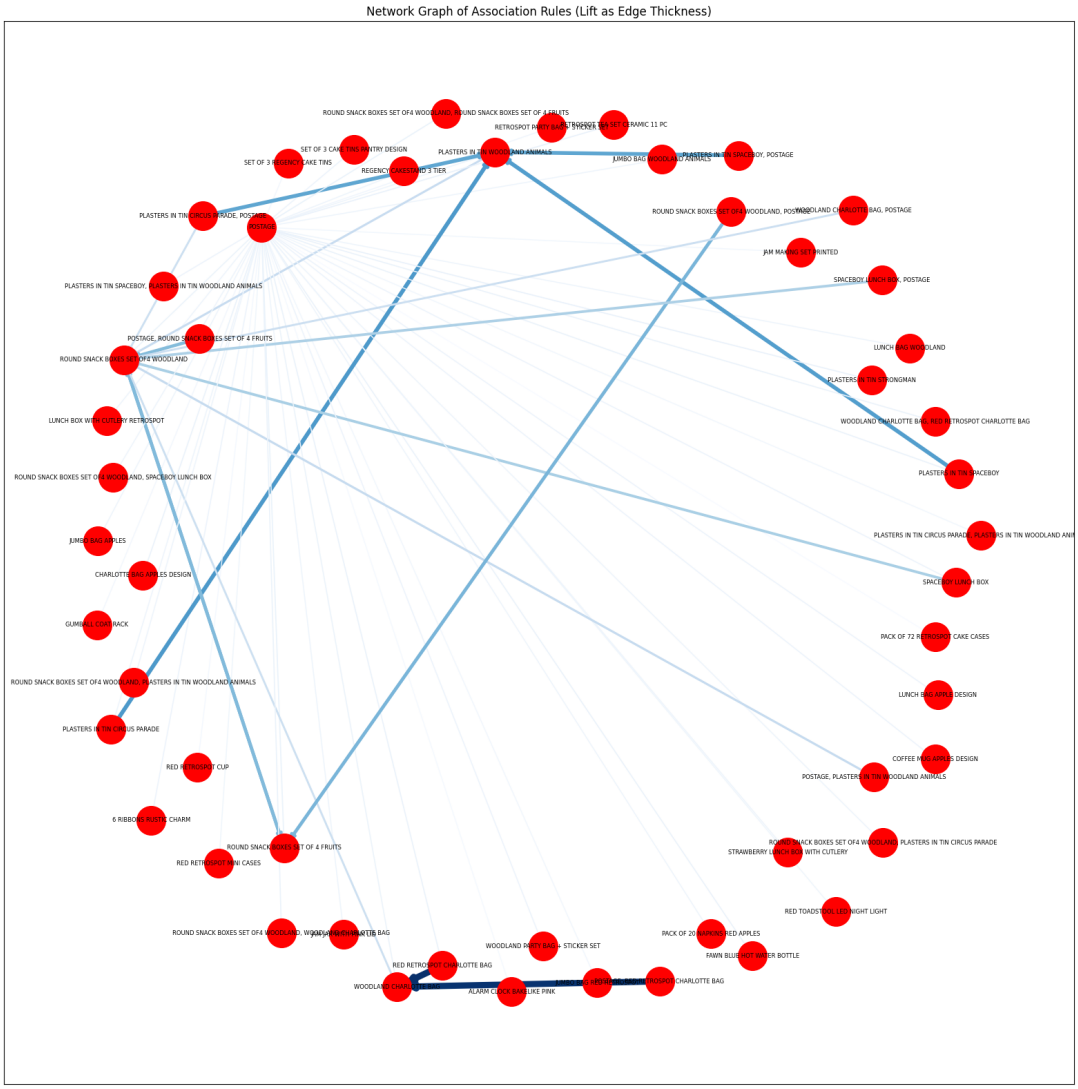


Figure 2.9: Association rule network graph for Germany.

- For France:

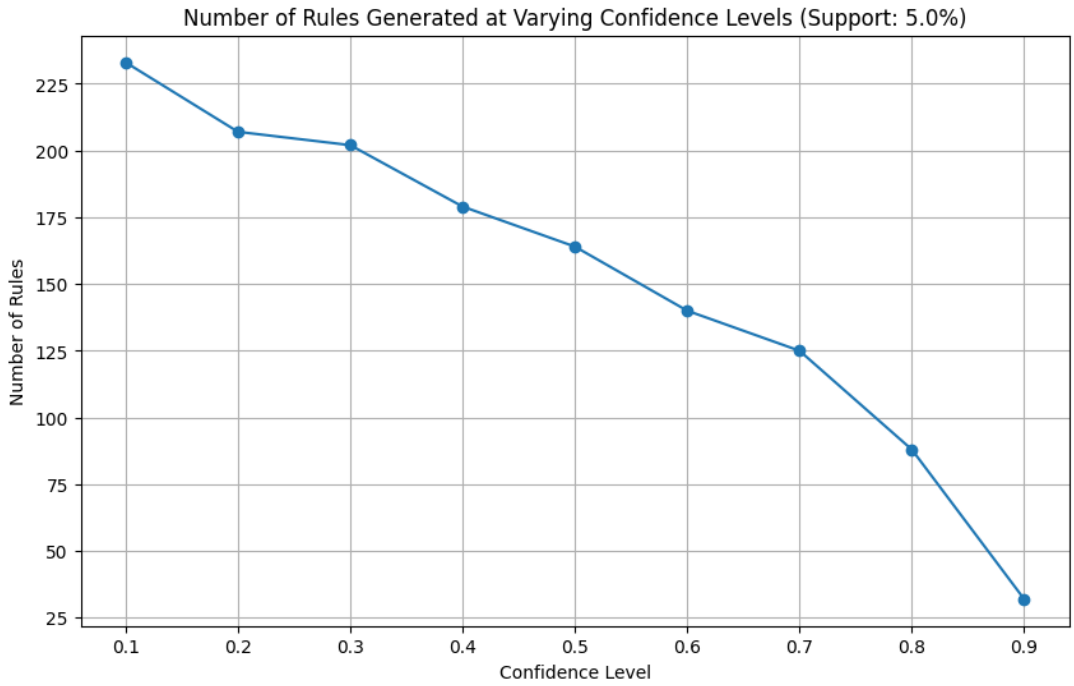


Figure 2.10: Number of rules generated in France with 5% support threshold.

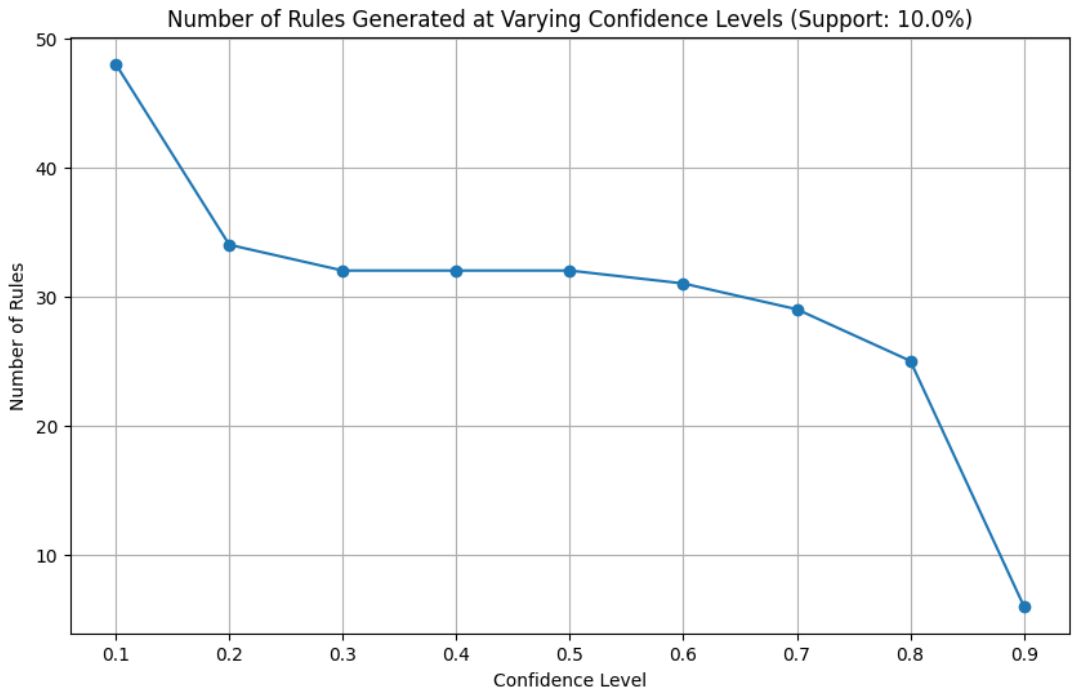


Figure 2.11: Number of rules generated in France with 10% support threshold.

	Antecedent	Consequent	Support	Confidence	Lift
0	(LUNCH BAG APPLE DESIGN)	(POSTAGE)	0.105398	0.836735	1.084966
1	(LUNCH BAG RED RETROSPOT)	(POSTAGE)	0.123393	0.800000	1.037333
2	(LUNCH BAG WOODLAND)	(POSTAGE)	0.102828	0.869565	1.127536
3	(LUNCH BOX WITH CUTLERY RETROSPOT)	(POSTAGE)	0.115681	0.803571	1.041964
4	(PLASTERS IN TIN CIRCUS PARADE)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.102828	0.606061	3.518770
5	(PLASTERS IN TIN WOODLAND ANIMALS)	(PLASTERS IN TIN CIRCUS PARADE)	0.102828	0.597015	3.518770
6	(PLASTERS IN TIN CIRCUS PARADE)	(POSTAGE)	0.149100	0.878788	1.139495
7	(PLASTERS IN TIN SPACEBOY)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.105398	0.773585	4.491411
8	(PLASTERS IN TIN WOODLAND ANIMALS)	(PLASTERS IN TIN SPACEBOY)	0.105398	0.611940	4.491411
9	(PLASTERS IN TIN SPACEBOY)	(POSTAGE)	0.115681	0.849057	1.100943
10	(PLASTERS IN TIN WOODLAND ANIMALS)	(POSTAGE)	0.138817	0.805970	1.045075
11	(RABBIT NIGHT LIGHT)	(POSTAGE)	0.167095	0.890411	1.154566
12	(RED RETROSPOT MINI CASES)	(POSTAGE)	0.115681	0.833333	1.080556
13	(RED TOADSTOOL LED NIGHT LIGHT)	(POSTAGE)	0.159383	0.885714	1.148476
14	(REGENCY CAKESTAND 3 TIER)	(POSTAGE)	0.105398	0.836735	1.084966
15	(ROUND SNACK BOXES SET OF4 WOODLAND)	(POSTAGE)	0.149100	0.935484	1.213011
16	(SET/20 RED RETROSPOT PAPER NAPKINS)	(POSTAGE)	0.110540	0.826923	1.072244
17	(SET/6 RED SPOTTY PAPER CUPS)	(POSTAGE)	0.118252	0.851852	1.104568
18	(SET/6 RED SPOTTY PAPER PLATES)	(POSTAGE)	0.107969	0.840000	1.089200
19	(STRAWBERRY LUNCH BOX WITH CUTLERY)	(POSTAGE)	0.115681	0.937500	1.215625
20	(SET/6 RED SPOTTY PAPER CUPS)	(SET/20 RED RETROSPOT PAPER NAPKINS)	0.102828	0.740741	5.541311

```
[ ]  1 rules_apriori_2.shape #Checking the no of rules generated
```

```
(32, 5)
```

Figure 2.12: Distribution of rule metrics in France.

Top 10 Rules by Support

	Antecedent	Consequent	Support	Confidence	Lift
11	frozenset({'RABBIT NIGHT LIGHT'})	frozenset({'POSTAGE'})	0.1670951156812339 5	0.8904109589041097	1.1545662100456624
13	frozenset({'RED TOADSTOOL LED NIGHT LIGHT'})	frozenset({'POSTAGE'})	0.1593830334190231 3	0.8857142857142857	1.1484761904761904
15	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND'})	frozenset({'POSTAGE'})	0.1491002570694087 4	0.935483870967742	1.2130107526881722
6	frozenset({'PLASTERS IN TIN CIRCUS PARADE'})	frozenset({'POSTAGE'})	0.1491002570694087 4	0.8787878787878788	1.1394949494949496
10	frozenset({'PLASTERS IN TIN WOODLAND ANIMALS'})	frozenset({'POSTAGE'})	0.1388174807197943 6	0.8059701492537313	1.0450746268656717
25	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	0.1233933161953727 5	0.9600000000000001	6.915555555555556
24	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	0.1233933161953727 5	0.8888888888888888	6.915555555555556
1	frozenset({'LUNCH BAG RED RETROSPOT'})	frozenset({'POSTAGE'})	0.1233933161953727 5	0.8	1.0373333333333334
17	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	frozenset({'POSTAGE'})	0.1182519280205655 6	0.8518518518518517	1.1045679012345677
3	frozenset({'LUNCH BOX WITH CUTLERY RETROSPOT'})	frozenset({'POSTAGE'})	0.1156812339331619 6	0.8035714285714286	1.041964285714286

Top 10 Rules by Confidence

	Antecedent	Consequent	Support	Confidence	Lift
31	frozenset({'SET/6 RED SPOTTY PAPER PLATES', 'SET/20 RED RETROSPOT PAPER NAPKINS'})	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	0.1002570694087403 7	0.975	7.023611111111111
29	frozenset({'SET/6 RED SPOTTY PAPER CUPS', 'SET/20 RED RETROSPOT PAPER NAPKINS'})	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	0.1002570694087403 7	0.975	7.585500000000001
25	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	0.1233933161953727 5	0.9600000000000001	6.915555555555556
28	frozenset({'SET/6 RED SPOTTY PAPER PLATES', 'POSTAGE'})	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	0.1028277634961439 6	0.9523809523809524	6.860670194003527
19	frozenset({'STRAWBERRY LUNCH BOX WITH CUTLERY'})	frozenset({'POSTAGE'})	0.1156812339331619 6	0.9375000000000001	1.2156250000000002
15	frozenset({'ROUND SNACK BOXES SET OF4 WOODLAND'})	frozenset({'POSTAGE'})	0.1491002570694087 4	0.935483870967742	1.2130107526881722
11	frozenset({'RABBIT NIGHT LIGHT'})	frozenset({'POSTAGE'})	0.1670951156812339 5	0.8904109589041097	1.1545662100456624
24	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	0.1233933161953727 5	0.8888888888888888	6.915555555555556
13	frozenset({'RED TOADSTOOL LED NIGHT LIGHT'})	frozenset({'POSTAGE'})	0.1593830334190231 3	0.8857142857142857	1.1484761904761904
6	frozenset({'PLASTERS IN TIN CIRCUS PARADE'})	frozenset({'POSTAGE'})	0.1491002570694087 4	0.8787878787878788	1.1394949494949496

Top 10 Rules by Lift

	Antecedent	Consequent	Support	Confidence	Lift
29	frozenset({'SET/6 RED SPOTTY PAPER CUPS', 'SET/20 RED RETROSPOT PAPER NAPKINS'})	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	0.1002570694087403 7	0.975	7.585500000000001
31	frozenset({'SET/6 RED SPOTTY PAPER PLATES', 'SET/20 RED RETROSPOT PAPER NAPKINS'})	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	0.1002570694087403 7	0.975	7.023611111111111
25	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	0.1233933161953727 5	0.9600000000000001	6.915555555555556
24	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	0.1233933161953727 5	0.8888888888888888	6.915555555555556
28	frozenset({'SET/6 RED SPOTTY PAPER PLATES', 'POSTAGE'})	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	0.1028277634961439 6	0.9523809523809524	6.860670194003527
26	frozenset({'SET/6 RED SPOTTY PAPER CUPS', 'POSTAGE'})	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	0.1028277634961439 6	0.8695652173913043	6.765217391304348
30	frozenset({'SET/6 RED SPOTTY PAPER CUPS', 'SET/6 RED SPOTTY PAPER PLATES'})	frozenset({'SET/20 RED RETROSPOT PAPER NAPKINS'})	0.1002570694087403 7	0.8125000000000001	6.078125000000002
22	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	frozenset({'SET/20 RED RETROSPOT PAPER NAPKINS'})	0.1028277634961439 6	0.8	5.984615384615386
23	frozenset({'SET/20 RED RETROSPOT PAPER NAPKINS'})	frozenset({'SET/6 RED SPOTTY PAPER PLATES'})	0.1028277634961439 6	0.7692307692307693	5.984615384615386
20	frozenset({'SET/6 RED SPOTTY PAPER CUPS'})	frozenset({'SET/20 RED RETROSPOT PAPER NAPKINS'})	0.1028277634961439 6	0.7407407407407407	5.541310541310541 5

Figure 2.13: Top 10 association rules in France.

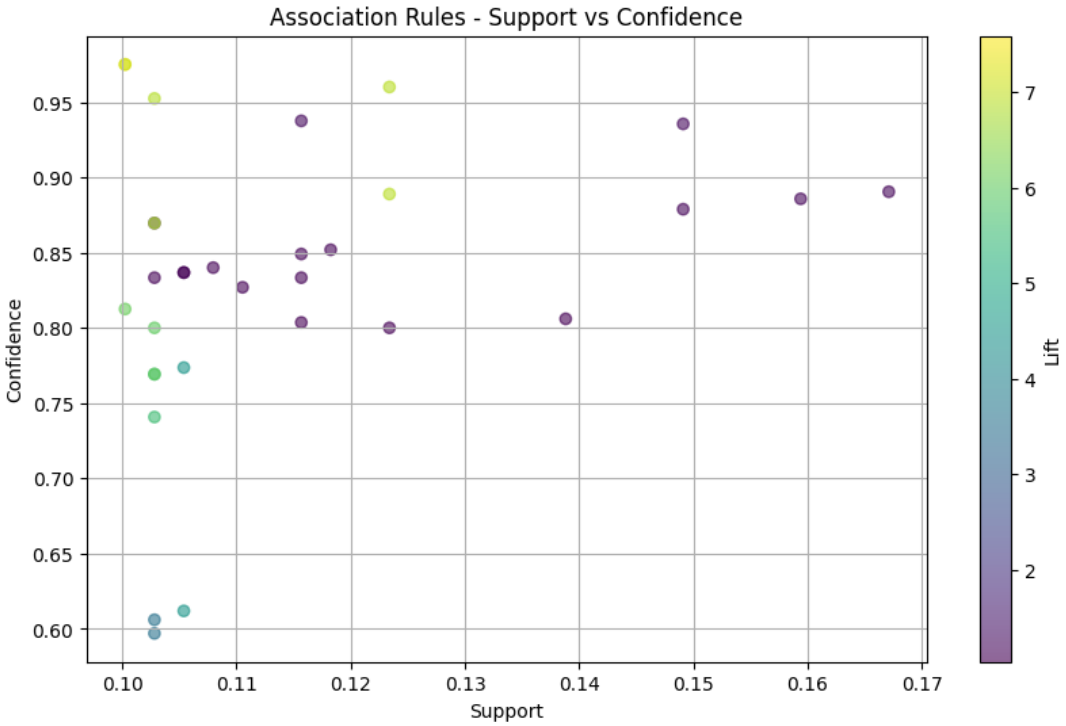


Figure 2.14: Support vs. Confidence for association rules in France.

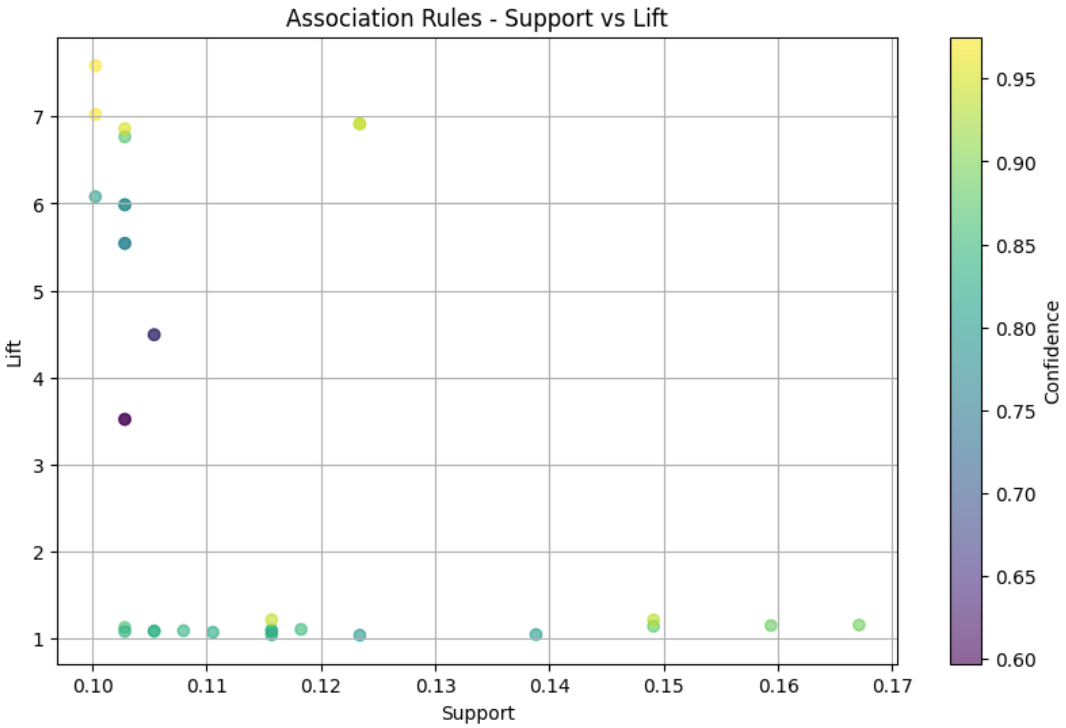


Figure 2.15: Support vs. Lift for association rules in France.

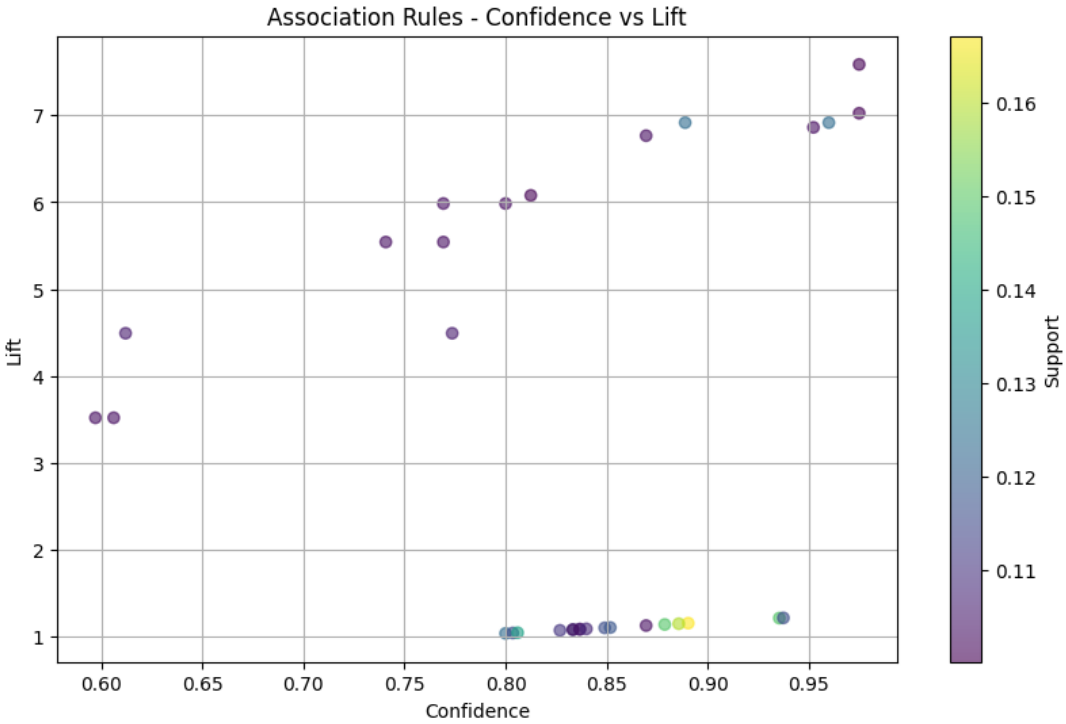


Figure 2.16: Confidence vs. Lift for association rules in France.

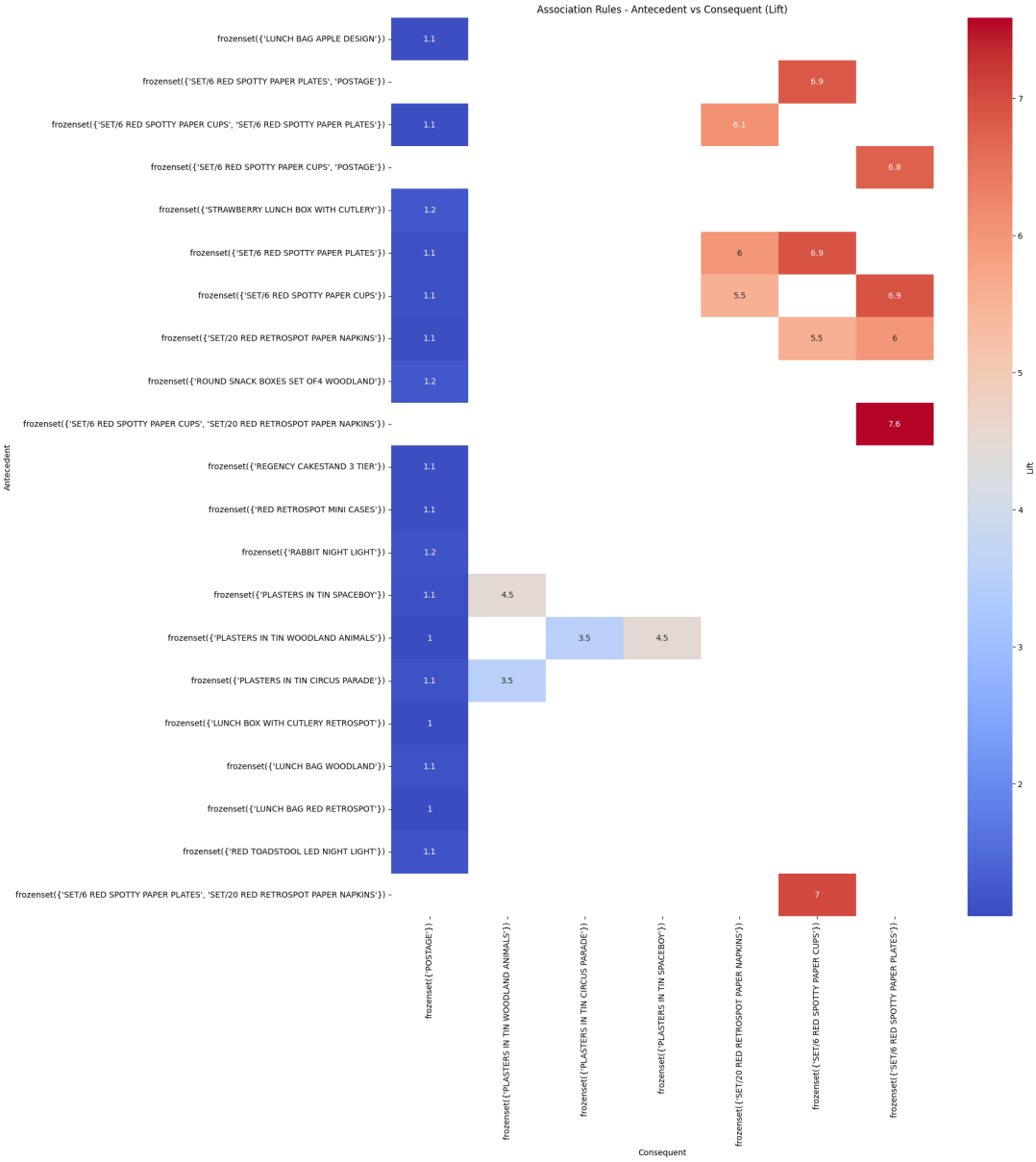


Figure 2.17: Antecedent Lift vs. Consequent Lift in France.

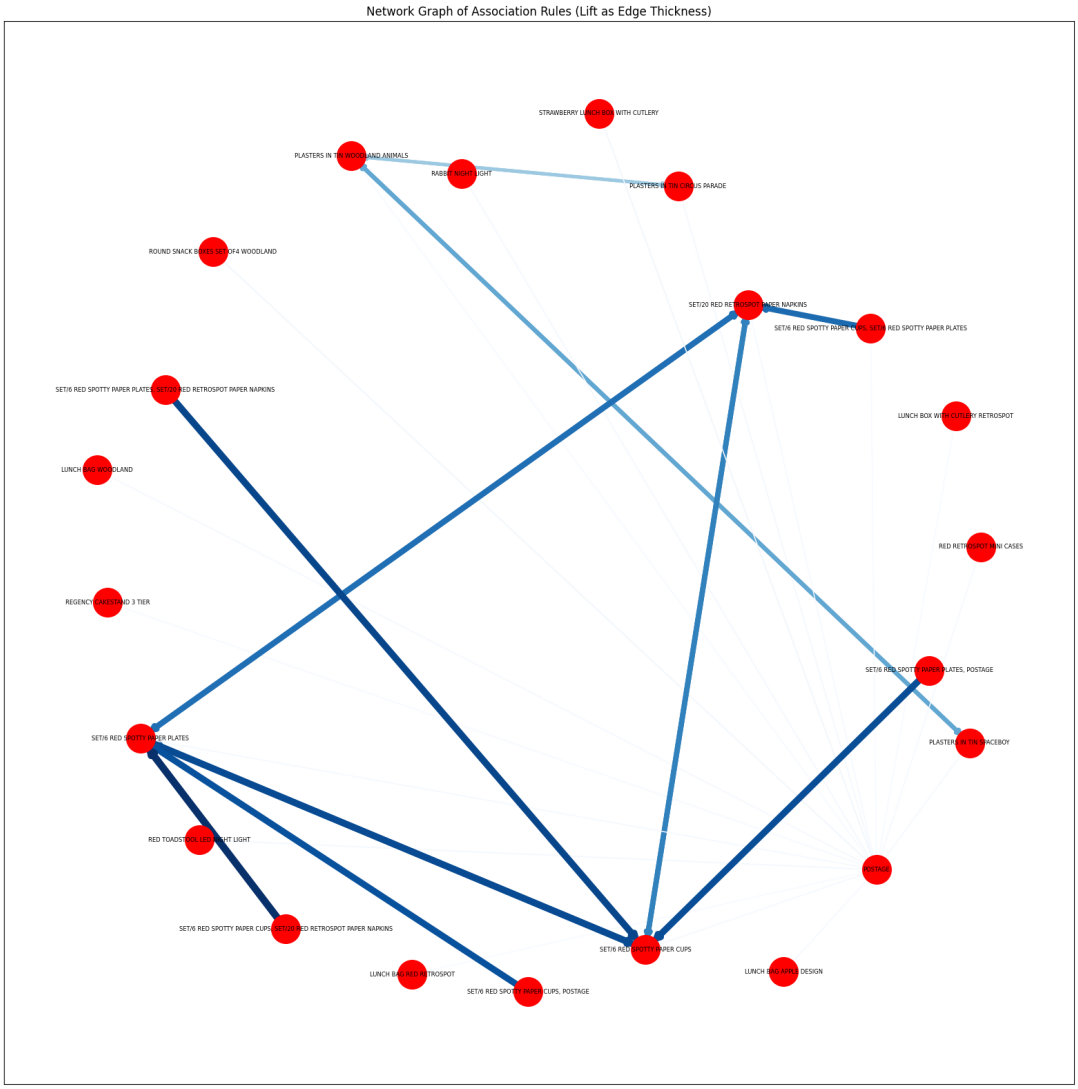


Figure 2.18: Association rule network graph for France.

- For Spain:

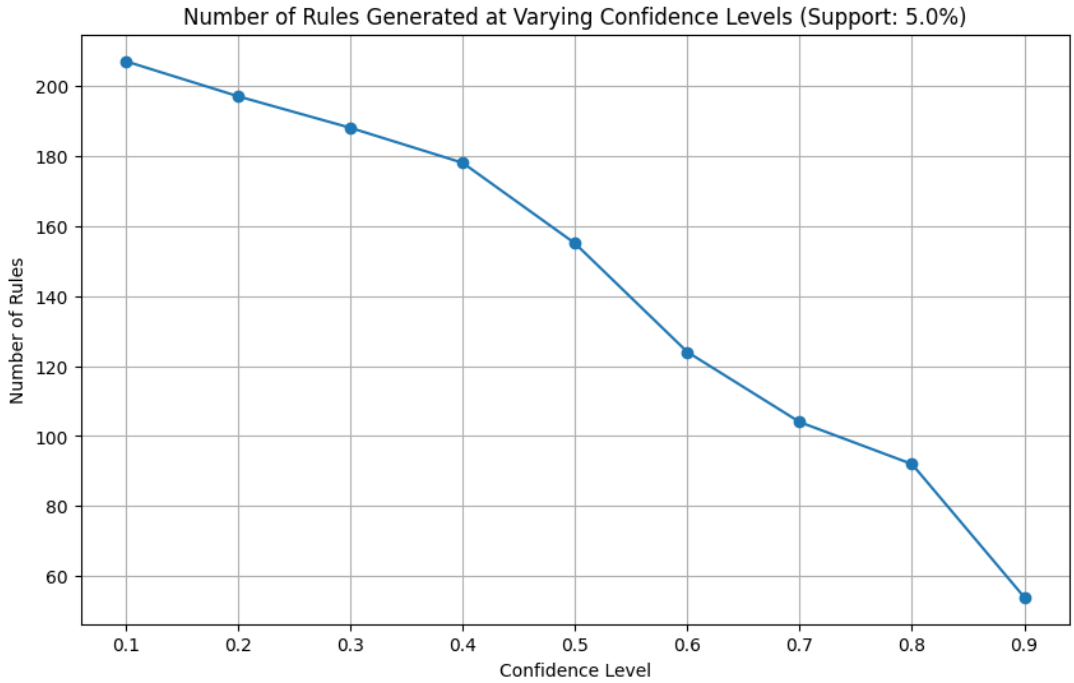


Figure 2.19: Number of rules generated in Spain with 5% support threshold.

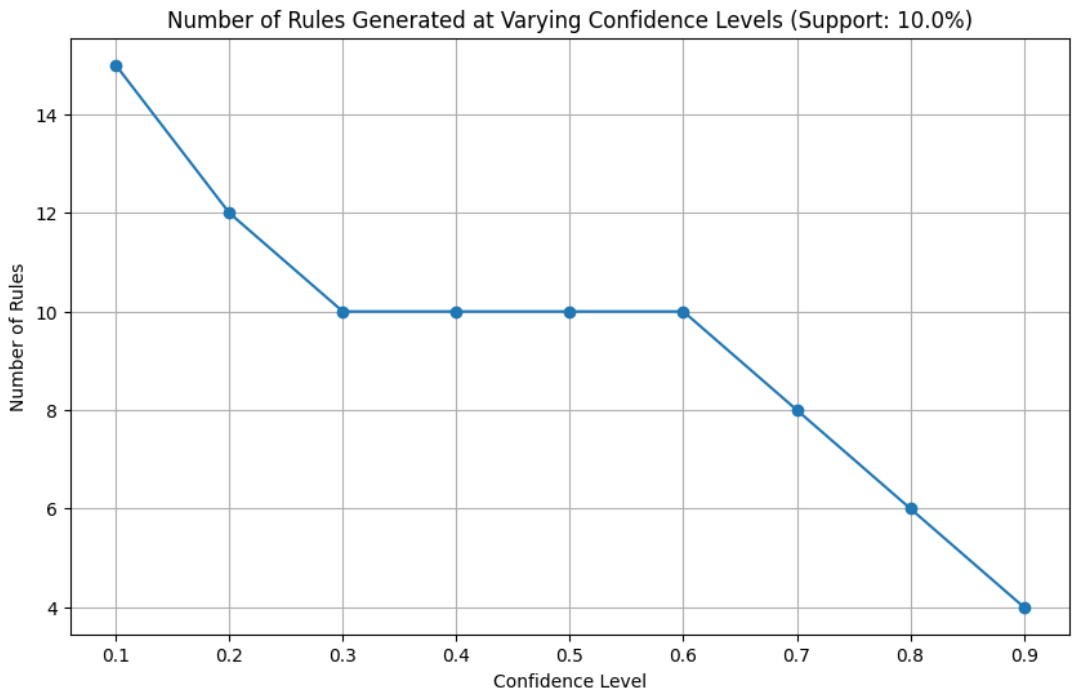


Figure 2.20: Number of rules generated in Spain with 10% support threshold.

	Antecedent	Consequent	Support	Confidence	Lift
0	(4 TRADITIONAL SPINNING TOPS)	(POSTAGE)	0.066667	1.000000	1.451613
1	(ASSORTED COLOUR BIRD ORNAMENT)	(6 RIBBONS RUSTIC CHARM)	0.100000	0.750000	4.500000
2	(6 RIBBONS RUSTIC CHARM)	(POSTAGE)	0.144444	0.866667	1.258065
3	(RED RETROSPOT TAPE)	(6 RIBBONS RUSTIC CHARM)	0.066667	0.857143	5.142857
4	(ALARM CLOCK BAKELIKE PINK)	(POSTAGE)	0.066667	1.000000	1.451613
5	(ALARM CLOCK BAKELIKE RED)	(POSTAGE)	0.055556	0.833333	1.209677
6	(PLASTERS IN TIN CIRCUS PARADE)	(ASSORTED COLOUR BIRD ORNAMENT)	0.055556	0.714286	5.357143
7	(ASSORTED COLOUR BIRD ORNAMENT)	(POSTAGE)	0.122222	0.916667	1.330645
8	(BAKING SET 9 PIECE RETROSPOT)	(POSTAGE)	0.055556	0.833333	1.209677
9	(BLUE HARMONICA IN BOX)	(POSTAGE)	0.055556	1.000000	1.451613
10	(BOX OF VINTAGE ALPHABET BLOCKS)	(POSTAGE)	0.055556	1.000000	1.451613
11	(PLASTERS IN TIN CIRCUS PARADE)	(CLASSIC METAL BIRDCAGE PLANT HOLDER)	0.055556	0.714286	6.428571
12	(CLASSIC METAL BIRDCAGE PLANT HOLDER)	(POSTAGE)	0.100000	0.900000	1.306452
13	(DANISH ROSE DECORATIVE PLATE)	(REGENCY CAKESTAND 3 TIER)	0.055556	1.000000	4.090909
14	(DOORMAT FANCY FONT HOME SWEET HOME)	(POSTAGE)	0.055556	1.000000	1.451613
15	(DOORMAT SPOTTY HOME SWEET HOME)	(POSTAGE)	0.077778	1.000000	1.451613
16	(FAIRY CAKE BIRTHDAY CANDLE SET)	(POSTAGE)	0.055556	1.000000	1.451613
17	(GIN & TONIC DIET GREETING CARD)	(POSTAGE)	0.066667	1.000000	1.451613
18	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.055556	1.000000	15.000000
19	(GREEN REGENCY TEACUP AND SAUCER)	(PINK REGENCY TEACUP AND SAUCER)	0.055556	0.833333	15.000000
20	(GREEN REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.066667	1.000000	8.181818

```
[ ]  1 rules_apriori_3.shape #Checking the no of rules generated
```

```
(104, 5)
```

Figure 2.21: Distribution of rule metrics in Spain.

Top 10 Rules by Support

	Antecedent	Consequent	Support	Confidence	Lift
47	frozenset({'REGENCY CAKESTAND 3 TIER'})	frozenset({'POSTAGE'})	0.1888888888888888	0.7727272727272727	1.121708797653958
2	frozenset({'6 RIBBONS RUSTIC CHARM'})	frozenset({'POSTAGE'})	0.1444444444444443	0.8666666666666666	1.258064516129323
33	frozenset({'PACK OF 72 RETROSPOT CAKE CASES'})	frozenset({'POSTAGE'})	0.1222222222222222	1.0	1.451612903225806 5
7	frozenset({'ASSORTED COLOUR BIRD ORNAMENT'})	frozenset({'POSTAGE'})	0.1222222222222222	0.9166666666666666	1.330645161290322 5
12	frozenset({'CLASSIC METAL BIRDCAGE PLANT HOLDER'})	frozenset({'POSTAGE'})	0.1	0.9000000000000001	1.306451612903226
66	frozenset({'6 RIBBONS RUSTIC CHARM', 'ASSORTED COLOUR BIRD ORNAMENT'})	frozenset({'POSTAGE'})	0.1	1.0	1.451612903225806 5
67	frozenset({'ASSORTED COLOUR BIRD ORNAMENT', 'POSTAGE'})	frozenset({'6 RIBBONS RUSTIC CHARM'})	0.1	0.8181818181818182	4.90909090909091
1	frozenset({'ASSORTED COLOUR BIRD ORNAMENT'})	frozenset({'6 RIBBONS RUSTIC CHARM'})	0.1	0.75	4.5
53	frozenset({'SET/20 RED RETROSPOT PAPER NAPKINS'})	frozenset({'POSTAGE'})	0.08888888888888889	0.8888888888888888	1.290322580645161 3
52	frozenset({'SET OF 72 RETROSPOT PAPER DOILIES'})	frozenset({'POSTAGE'})	0.08888888888888889	0.8	1.161290322580645 2

Top 10 Rules by Confidence

	Antecedent	Consequent	Support	Confidence	Lift
0	frozenset({'4 TRADITIONAL SPINNING TOPS'})	frozenset({'POSTAGE'})	0.06666666666666667	1.0	1.451612903225806 5
79	frozenset({'CLASSIC METAL BIRDCAGE PLANT HOLDER', 'PLASTERS IN TIN SKULLS'})	frozenset({'POSTAGE'})	0.0555555555555555	1.0	1.451612903225806 5
40	frozenset({"POPPY'S PLAYHOUSE KITCHEN"})	frozenset({"POPPY'S PLAYHOUSE BEDROOM"})	0.0777777777777778	1.0	11.25
42	frozenset({"POPPY'S PLAYHOUSE LIVINGROOM"})	frozenset({"POPPY'S PLAYHOUSE BEDROOM"})	0.0555555555555555	1.0	11.25
45	frozenset({"POPPY'S PLAYHOUSE LIVINGROOM"})	frozenset({"POPPY'S PLAYHOUSE KITCHEN"})	0.0555555555555555	1.0	12.857142857142858
58	frozenset({'WRAP DOILEY DESIGN'})	frozenset({'POSTAGE'})	0.0555555555555555	1.0	1.451612903225806 5
62	frozenset({'6 RIBBONS RUSTIC CHARM', 'PACK OF 72 RETROSPOT CAKE CASES'})	frozenset({'ASSORTED COLOUR BIRD ORNAMENT'})	0.06666666666666667	1.0	7.5
63	frozenset({'ASSORTED COLOUR BIRD ORNAMENT', 'PACK OF 72 RETROSPOT CAKE CASES'})	frozenset({'6 RIBBONS RUSTIC CHARM'})	0.06666666666666667	1.0	6.0
66	frozenset({'6 RIBBONS RUSTIC CHARM', 'ASSORTED COLOUR BIRD ORNAMENT'})	frozenset({'POSTAGE'})	0.1	1.0	1.451612903225806 5
68	frozenset({'CLASSIC METAL BIRDCAGE PLANT HOLDER', '6 RIBBONS RUSTIC CHARM'})	frozenset({'POSTAGE'})	0.0555555555555555	1.0	1.451612903225806 5

Top 10 Rules by Lift

	Antecedent	Consequent	Support	Confidence	Lift
85	frozenset({'LUNCH BAG PINK POLKADOT', 'LUNCH BAG CARS BLUE'})	frozenset({'LUNCH BAG BLACK SKULL.'})	0.0555555555555555	1.0	18.0
81	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0555555555555555	1.0	15.0
83	frozenset({'GREEN REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'PINK REGENCY TEACUP AND SAUCER'})	0.0555555555555555	0.8333333333333333	15.0
84	frozenset({'LUNCH BAG BLACK SKULL.', 'LUNCH BAG PINK POLKADOT'})	frozenset({'LUNCH BAG CARS BLUE'})	0.0555555555555555	1.0	15.0
27	frozenset({'LUNCH BAG CARS BLUE'})	frozenset({'LUNCH BAG BLACK SKULL.'})	0.0555555555555555	0.8333333333333333	15.0
26	frozenset({'LUNCH BAG BLACK SKULL.'})	frozenset({'LUNCH BAG CARS BLUE'})	0.0555555555555555	1.0	15.0
19	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	frozenset({'PINK REGENCY TEACUP AND SAUCER'})	0.0555555555555555	0.8333333333333333	15.0
18	frozenset({'PINK REGENCY TEACUP AND SAUCER'})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.0555555555555555	1.0	15.0
45	frozenset({"POPPY'S PLAYHOUSE LIVINGROOM"})	frozenset({"POPPY'S PLAYHOUSE KITCHEN"})	0.0555555555555555	1.0	12.857142857142858
93	frozenset({"POPPY'S PLAYHOUSE BEDROOM", "POPPY'S PLAYHOUSE LIVINGROOM"})	frozenset({"POPPY'S PLAYHOUSE KITCHEN"})	0.0555555555555555	1.0	12.857142857142858

Figure 2.22: Top 10 association rules in Spain.

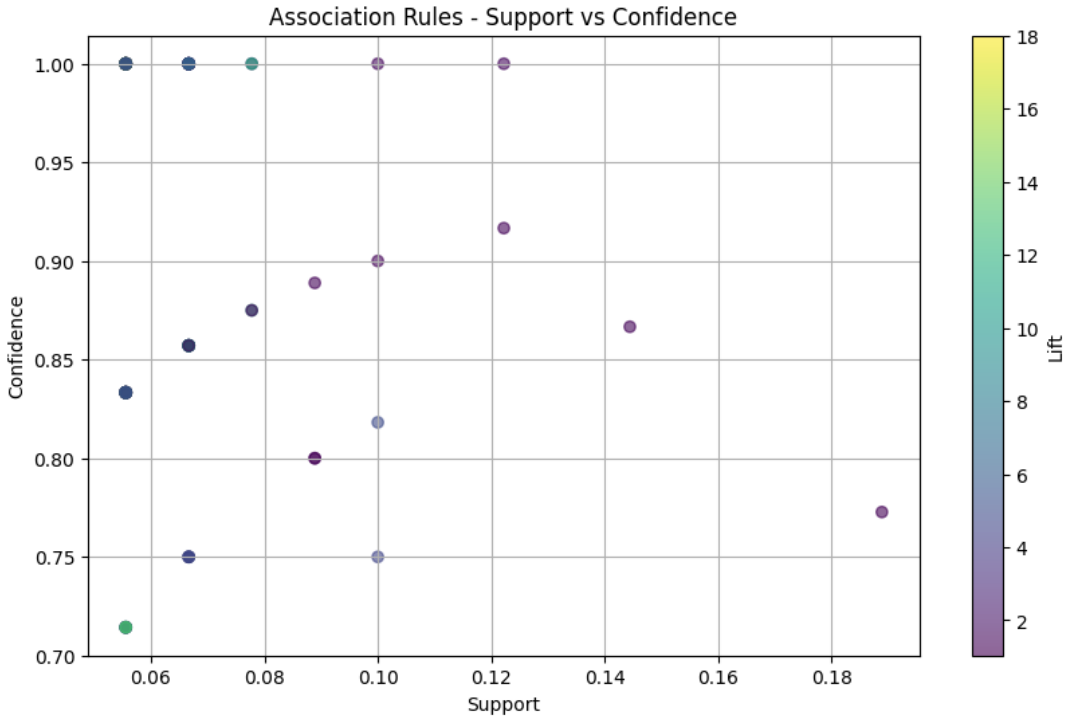


Figure 2.23: Support vs. Confidence for association rules in Spain.

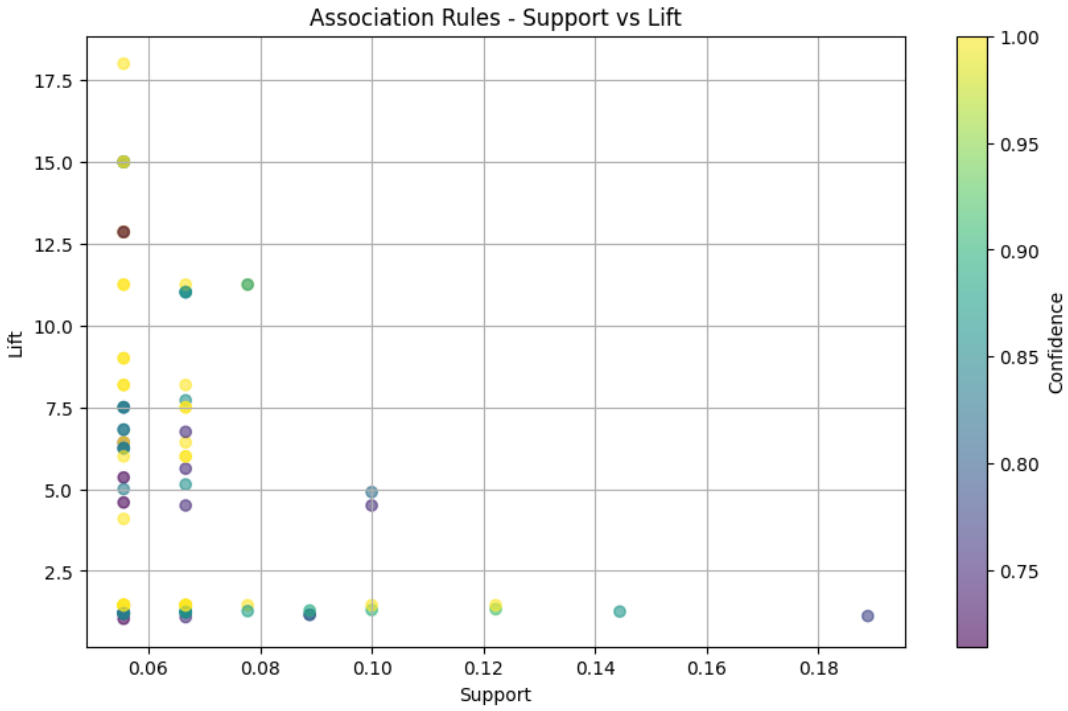


Figure 2.24: Support vs. Lift for association rules in Spain.

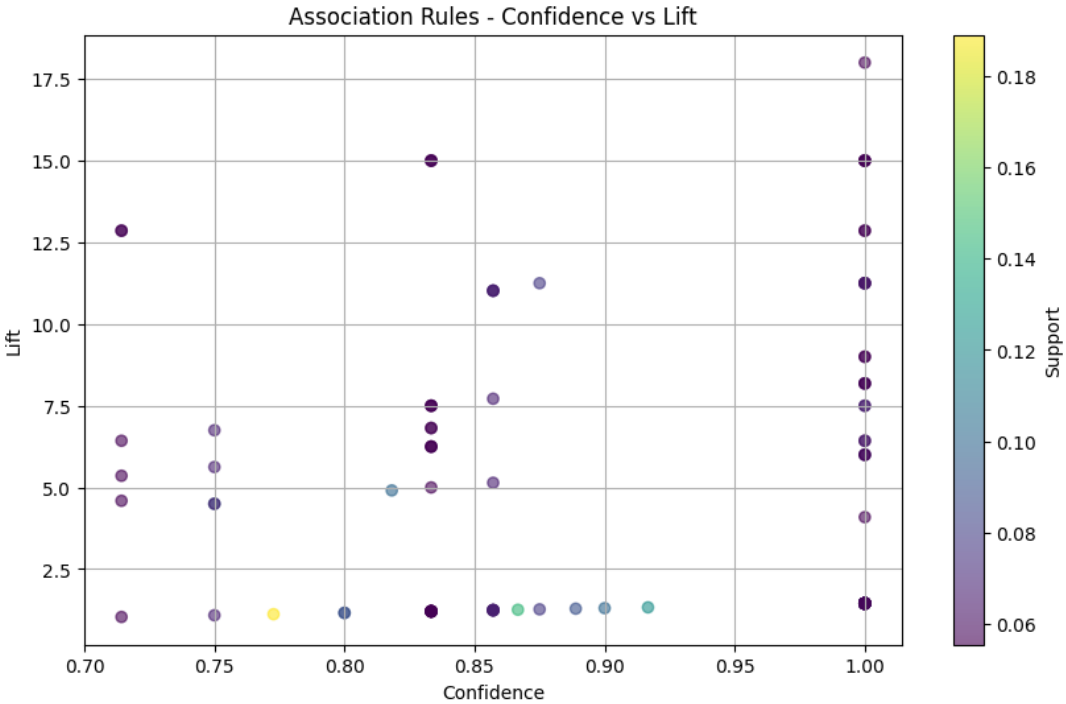


Figure 2.25: Confidence vs. Lift for association rules in Spain.

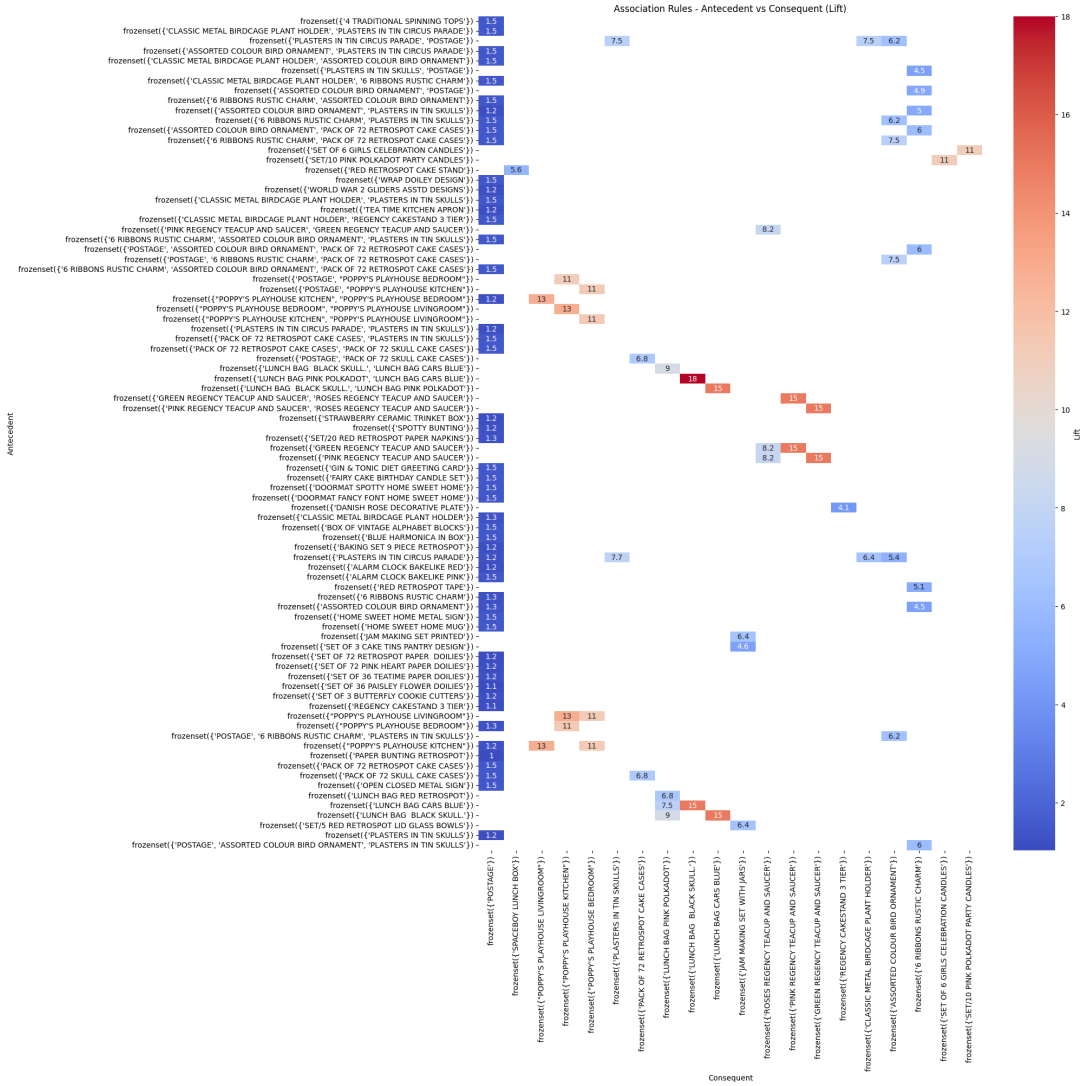


Figure 2.26: Antecedent Lift vs. Consequent Lift in Spain.

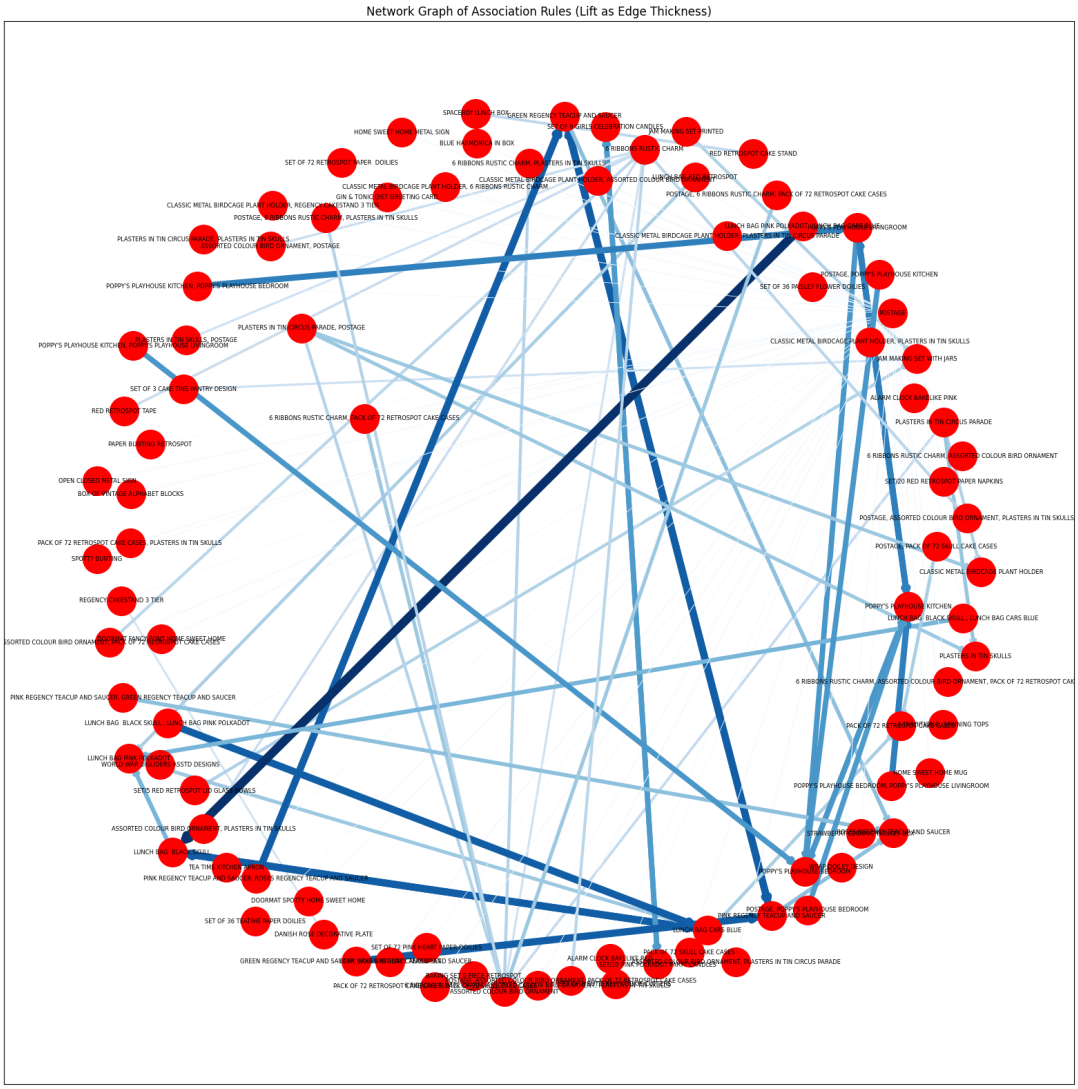


Figure 2.27: Association rule network graph for Spain.

- For Ireland:

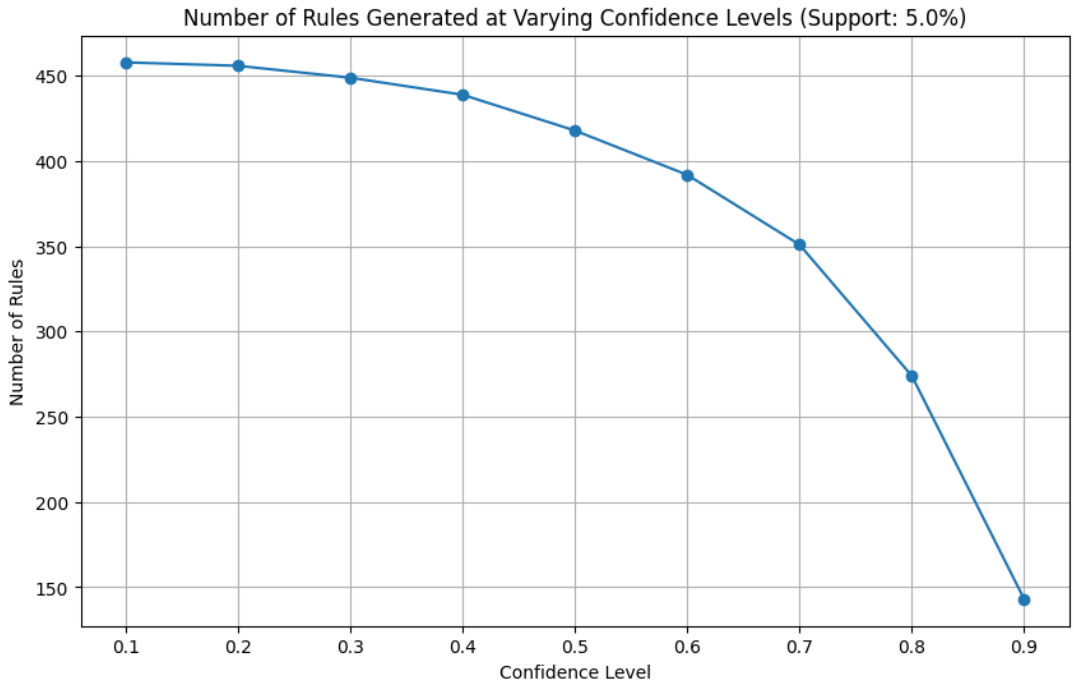


Figure 2.28: Number of rules generated in Ireland with 5% support threshold.

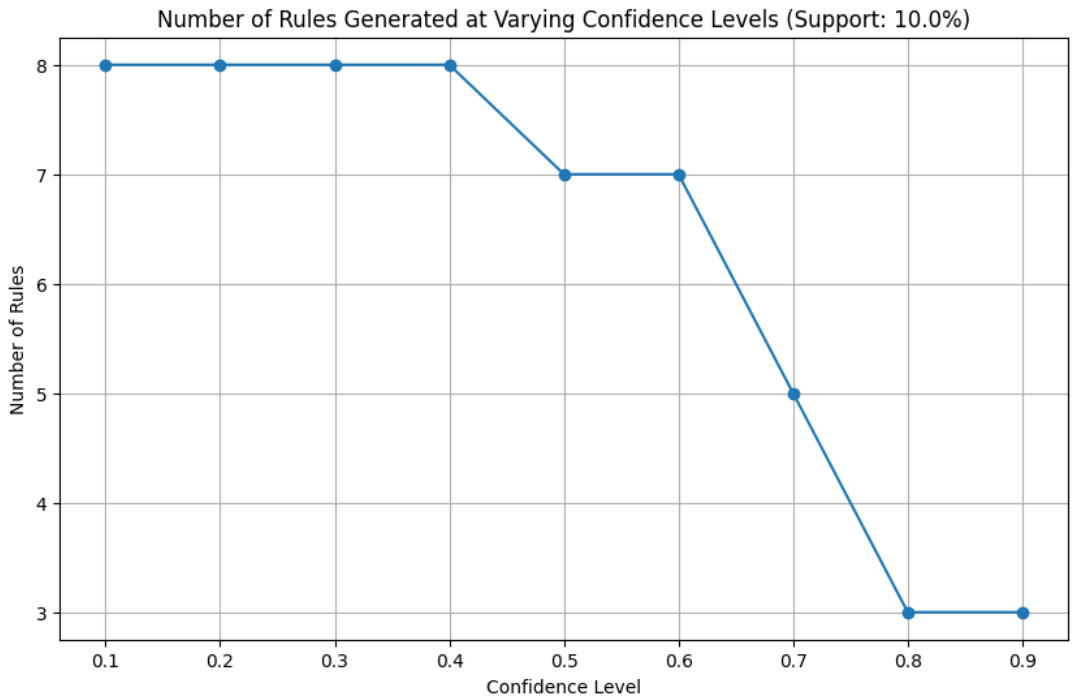


Figure 2.29: Number of rules generated in Ireland with 10% support threshold.

	Antecedent	Consequent	Support	Confidence	Lift
0	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.100000	0.928571	7.100840
1	(GREEN REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.119231	0.911765	5.387701
2	(PINK REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.103846	0.964286	5.698052
3	(REGENCY TEA PLATE GREEN)	(REGENCY CAKESTAND 3 TIER)	0.069231	0.857143	3.537415
4	(REGENCY TEAPOT ROSES)	(REGENCY SUGAR BOWL GREEN)	0.065385	0.944444	10.231481
5	(REGENCY TEA PLATE GREEN)	(REGENCY TEA PLATE PINK)	0.069231	0.857143	11.142857
6	(REGENCY TEA PLATE PINK)	(REGENCY TEA PLATE GREEN)	0.069231	0.900000	11.142857
7	(REGENCY TEA PLATE GREEN)	(REGENCY TEA PLATE ROSES)	0.073077	0.904762	9.801587
8	(REGENCY TEA PLATE PINK)	(REGENCY TEA PLATE ROSES)	0.065385	0.850000	9.208333
9	(REGENCY TEAPOT ROSES)	(ROSES REGENCY TEACUP AND SAUCER)	0.061538	0.888889	5.252525
10	(PINK REGENCY TEACUP AND SAUCER, REGENCY CAKES...	(GREEN REGENCY TEACUP AND SAUCER)	0.073077	0.950000	7.264706
11	(PINK REGENCY TEACUP AND SAUCER, REGENCY MILK ...	(GREEN REGENCY TEACUP AND SAUCER)	0.053846	0.875000	6.691176
12	(REGENCY MILK JUG PINK, GREEN REGENCY TEACUP A...	(PINK REGENCY TEACUP AND SAUCER)	0.053846	0.875000	8.125000
13	(PINK REGENCY TEACUP AND SAUCER, REGENCY SUGAR...	(GREEN REGENCY TEACUP AND SAUCER)	0.057692	0.882353	6.747405
14	(REGENCY SUGAR BOWL GREEN, GREEN REGENCY TEACU...	(PINK REGENCY TEACUP AND SAUCER)	0.057692	0.882353	8.193277
15	(PINK REGENCY TEACUP AND SAUCER, REGENCY TEA P...	(GREEN REGENCY TEACUP AND SAUCER)	0.050000	0.928571	7.100840
16	(REGENCY TEA PLATE GREEN, GREEN REGENCY TEACUP...	(PINK REGENCY TEACUP AND SAUCER)	0.050000	0.866667	8.047619
17	(PINK REGENCY TEACUP AND SAUCER, REGENCY TEA P...	(GREEN REGENCY TEACUP AND SAUCER)	0.050000	0.866667	6.627451
18	(REGENCY TEA PLATE PINK, GREEN REGENCY TEACUP ...	(PINK REGENCY TEACUP AND SAUCER)	0.050000	1.000000	9.285714
19	(PINK REGENCY TEACUP AND SAUCER, ROSES REGENCY...	(GREEN REGENCY TEACUP AND SAUCER)	0.096154	0.925926	7.080610
20	(PINK REGENCY TEACUP AND SAUCER, GREEN REGENCY...	(ROSES REGENCY TEACUP AND SAUCER)	0.096154	0.961538	5.681818

```
[ ]  1 rules_apriori_4.shape #Checking the no of rules generated
```

```
(231, 5)
```

Figure 2.30: Distribution of rule metrics in Ireland.

Top 10 Rules by Support

	Antecedent	Consequent	Support	Confidence	Lift
1	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.119230769230769 24	0.911764705882352 9	5.387700534759358
2	frozenset({'PINK REGENCY TEACUP AND SAUCER'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.103846153846153 85	0.964285714285714 3	5.698051948051948
0	frozenset({'PINK REGENCY TEACUP AND SAUCER'})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.1	0.928571428571428 6	7.100840336134453
20	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'GREEN REGENCY TEACUP AND SAUCER'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.096153846153846 16	0.961538461538461 6	5.681818181818182
19	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.096153846153846 16	0.925925925925925 9	7.080610021786492
24	frozenset({'REGENCY CAKESTAND 3 TIER', 'GREEN REGENCY TEACUP AND SAUCER'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.088461538461538 47	1.0	5.909090909090909
40	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY CAKESTAND 3 TIER'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.076923076923076 93	1.0	5.909090909090909
112	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY CAKESTAND 3 TIER', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.073076923076923 08	0.95	7.264705882352294
111	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY CAKESTAND 3 TIER', 'GREEN REGENCY TEACUP AND SAUCER'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.073076923076923 08	1.0	5.909090909090909
10	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY CAKESTAND 3 TIER'})	frozenset({'GREEN REGENCY TEACUP AND SAUCER'})	0.073076923076923 08	0.95	7.264705882352294

Top 10 Rules by Confidence

	Antecedent	Consequent	Support	Confidence	Lift
230	frozenset({'REGENCY TEA PLATE ROSES', 'REGENCY TEA PLATE GREEN', 'REGENCY TEA PLATE PINK', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY CAKESTAND 3 TIER'})	0.05	1.0	4.126984126984127
155	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY CAKESTAND 3 TIER', 'REGENCY TEA PLATE PINK'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.05	1.0	5.909090909090909
160	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY SUGAR BOWL GREEN', 'REGENCY MILK JUG PINK'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.053846153846153 85	1.0	5.909090909090909
163	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY TEA PLATE GREEN', 'REGENCY TEA PLATE PINK'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.05	1.0	5.909090909090909
167	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY TEA PLATE ROSES', 'REGENCY TEA PLATE GREEN'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.05	1.0	5.909090909090909
171	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY TEA PLATE ROSES', 'REGENCY TEA PLATE PINK'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.05	1.0	5.909090909090909
71	frozenset({'REGENCY TEA PLATE GREEN', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY CAKESTAND 3 TIER'})	0.061538461538461 54	1.0	4.126984126984127
65	frozenset({'REGENCY CAKESTAND 3 TIER', 'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.061538461538461 54	1.0	12.380952380952381
181	frozenset({'REGENCY SUGAR BOWL GREEN', 'REGENCY TEA PLATE GREEN', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY CAKESTAND 3 TIER'})	0.05	1.0	4.126984126984127
125	frozenset({'REGENCY SUGAR BOWL GREEN', 'REGENCY CAKESTAND 3 TIER', 'GREEN REGENCY TEACUP AND SAUCER'})	frozenset({'ROSES REGENCY TEACUP AND SAUCER'})	0.057692307692307696	1.0	5.909090909090909

Top 10 Rules by Lift

	Antecedent	Consequent	Support	Confidence	Lift
128	frozenset({'REGENCY TEA PLATE ROSES', 'REGENCY CAKESTAND 3 TIER', 'GREEN REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.05	1.0	12.380952380952381
227	frozenset({'REGENCY TEA PLATE ROSES', 'REGENCY CAKESTAND 3 TIER', 'REGENCY TEA PLATE PINK', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.05	1.0	12.380952380952381
186	frozenset({'REGENCY TEA PLATE ROSES', 'REGENCY CAKESTAND 3 TIER', 'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.053846153846153 85	1.0	12.380952380952381
212	frozenset({'ROSES REGENCY TEACUP AND SAUCER', 'REGENCY TEA PLATE ROSES', 'REGENCY CAKESTAND 3 TIER', 'GREEN REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.05	1.0	12.380952380952381
217	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY CAKESTAND 3 TIER', 'REGENCY TEA PLATE PINK', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.05	1.0	12.380952380952381
65	frozenset({'REGENCY CAKESTAND 3 TIER', 'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.061538461538461 54	1.0	12.380952380952381
222	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY TEA PLATE ROSES', 'REGENCY CAKESTAND 3 TIER', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.05	1.0	12.380952380952381
149	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY TEA PLATE ROSES', 'REGENCY CAKESTAND 3 TIER'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.05	1.0	12.380952380952381
191	frozenset({'REGENCY CAKESTAND 3 TIER', 'REGENCY TEA PLATE PINK', 'ROSES REGENCY TEACUP AND SAUCER'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.053846153846153 85	1.0	12.380952380952381
147	frozenset({'PINK REGENCY TEACUP AND SAUCER', 'REGENCY CAKESTAND 3 TIER', 'REGENCY TEA PLATE PINK'})	frozenset({'REGENCY TEA PLATE GREEN'})	0.05	1.0	12.380952380952381

Figure 2.31: Top 10 association rules in Ireland.

Figure 2.32: Support vs. Confidence for association rules in Ireland.

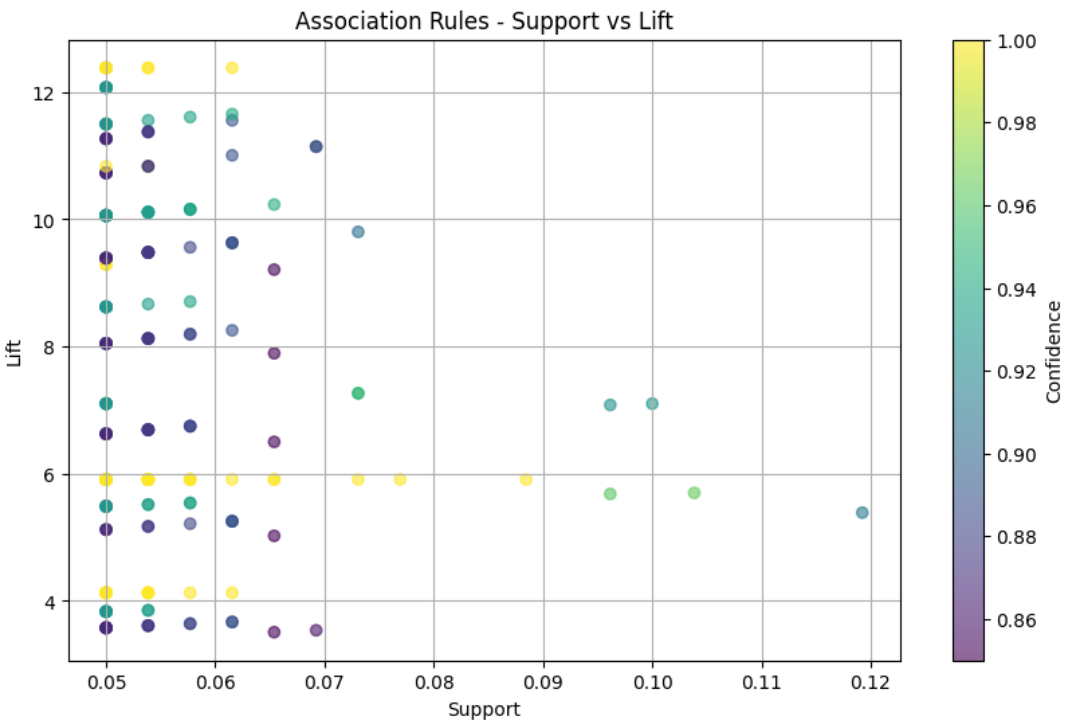


Figure 2.33: Support vs. Lift for association rules in Ireland.

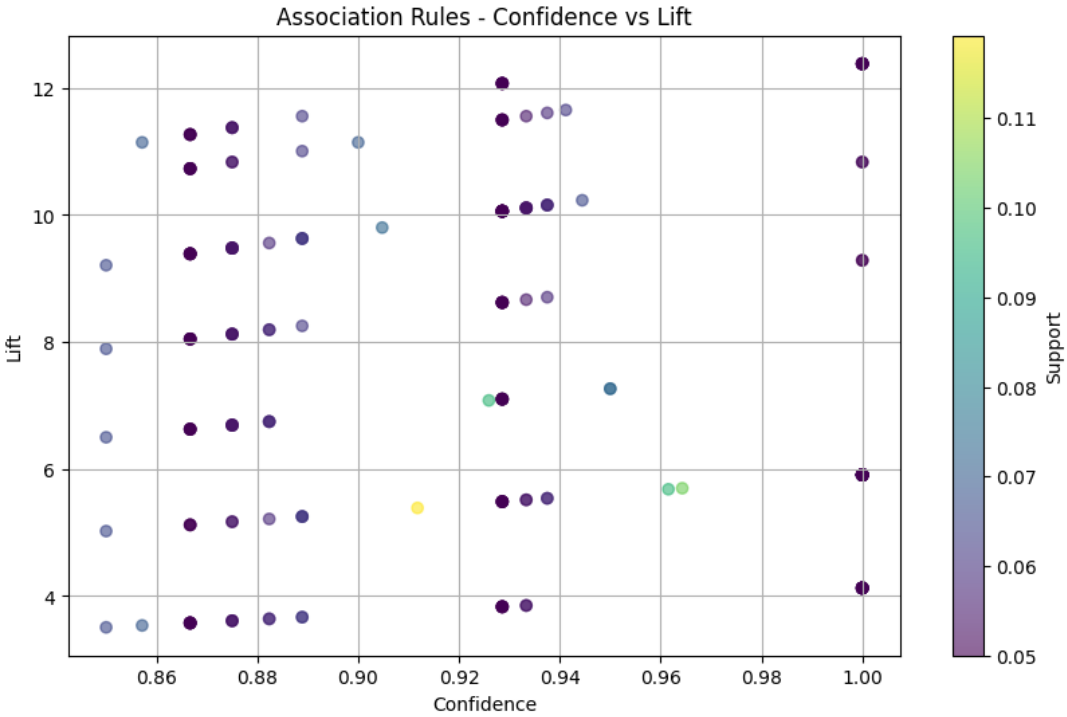


Figure 2.34: Confidence vs. Lift for association rules in Ireland.

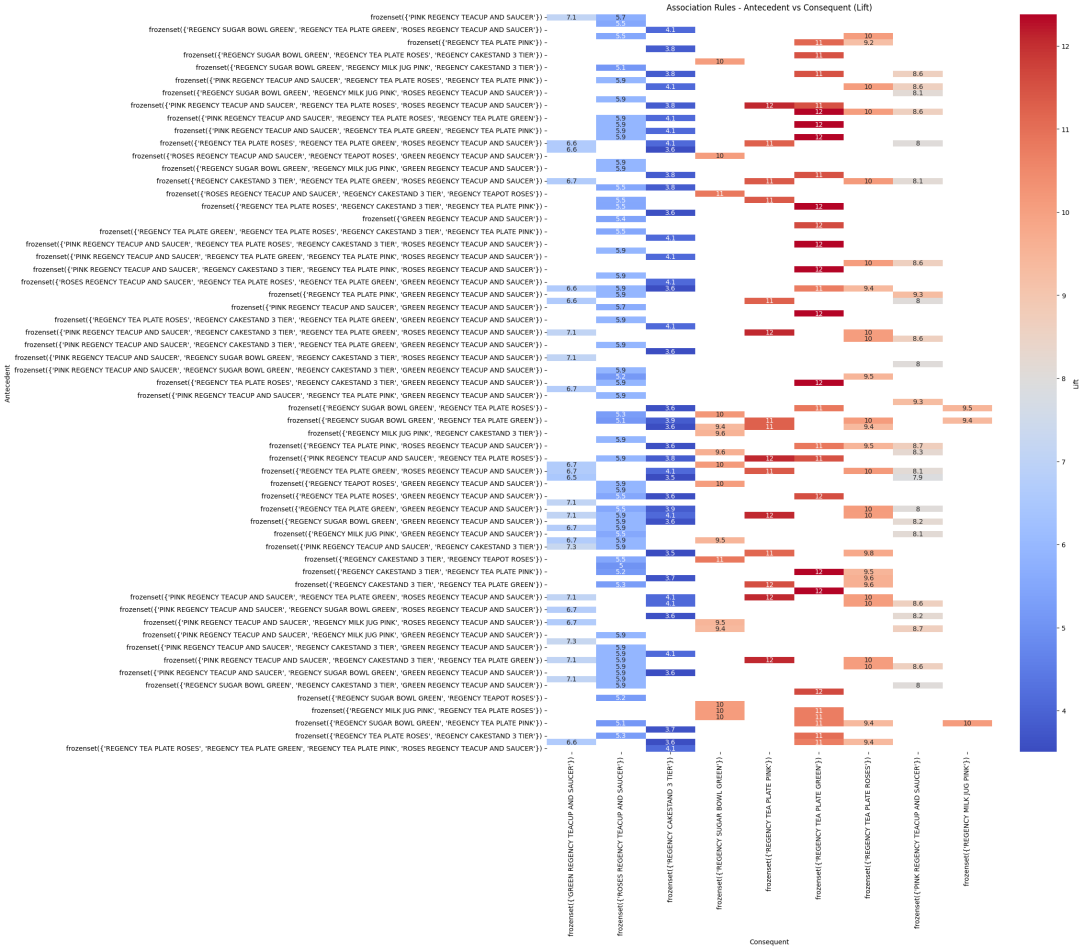


Figure 2.35: Antecedent Lift vs. Consequent Lift in Ireland.

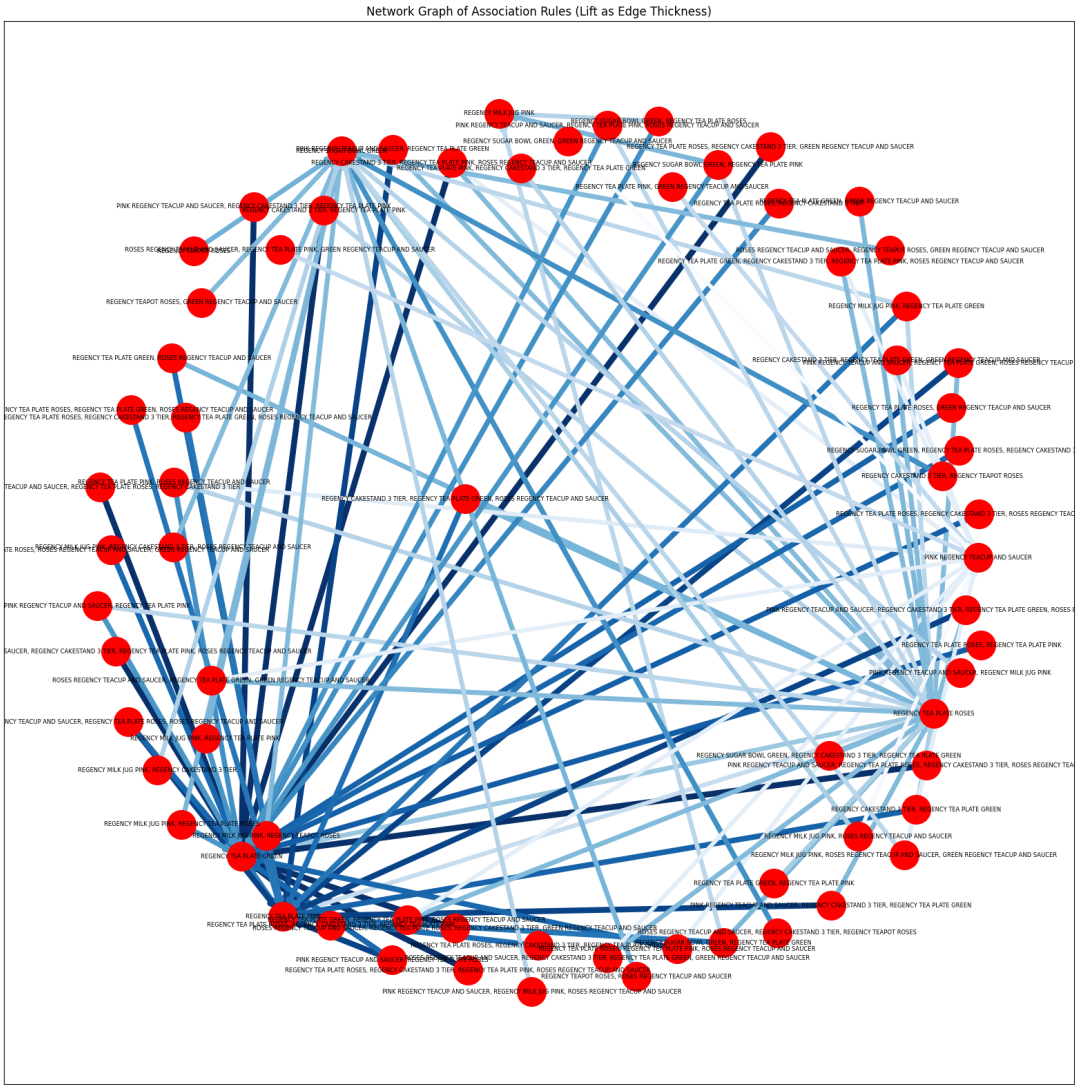


Figure 2.36: Association rule network graph for Ireland.

2.3 Analysis

Initial Data Analysis:

- From the total sales over time plot,we can infer that the highest sales occur in October and November respectively.
- Paper Craft,Little Birdie is the highest sold product followed by Regency Cakestand 3 Tier and White Hanging Heart T-Light Holder.
- In terms of Number of transactions,UK leads by a massive margin followed by Germany,France and Ireland respectively.

Analysis of the German Transactional Data:

- Based on the plots,we find support level = 0.05 and confidence = 0.5 as an appropriate metric for carrying out the rule generation
- **Analysis of Top 10 Rules by Support**
  The top rules by Support indicate products that commonly co-occur in customers' purchases:
  Products like ROUND SNACK BOXES SET OF 4 FRUITS and ROUND SNACK BOXES SET OF 4 WOODLAND frequently appear with POSTAGE in transactions, with Support values of 22.5 percent and 15.1 percent, respectively.
  The high Support for combinations like ROUND SNACK BOXES SET OF4 WOODLAND and ROUND SNACK BOXES SET OF 4 FRUITS suggests that customers tend to buy variations of snack boxes together.
  Products associated with POSTAGE likely indicate items frequently shipped together, or POSTAGE may appear as an ancillary item when certain items are ordered.
- **Analysis of Top 10 Rules by Confidence**
  Confidence indicates the likelihood that if the antecedent is in a transaction, the consequent will also be there. Higher Confidence values imply stronger predictive relationships.
  Items like PLASTERS IN TIN STRONGMAN (96.9 percent confidence) and RETROSPOT TEA SET CERAMIC 11 PC (96.2 percent confidence) indicate that these items are highly likely to be purchased with POSTAGE.
  Rules involving items such as CHARLOTTE BAG APPLES DESIGN and WOODLAND CHARLOTTE BAG show high confidence with POSTAGE, suggesting that these items are often shipped.
  High Confidence with POSTAGE may reflect commonly paired items in gift packs or multi-item purchases that require shipping, thus adding POSTAGE automatically.
- **Analysis of Top 10 Rules by Lift**
  Lift measures the strength of an association relative to random chance, with values above 1 indicating a meaningful association:
  Rules with items like RED RETROSPOT CHARLOTTE BAG and WOODLAND CHARLOTTE BAG (Lift = 6.65) or PLASTERS IN TIN CIRCUS PARADE with PLASTERS IN TIN WOODLAND ANIMALS (Lift = 4.24) suggest a strong association.
  Combinations of ROUND SNACK BOXES SET OF 4 FRUITS and ROUND SNACK BOXES SET OF4 WOODLAND show strong Lift, indicating that customers purchasing one variety are also very likely to buy another variation.
  High Lift scores often reveal items that complement each other in customer preferences, indicating potential bundling opportunities for similar items, such as themed bags or snack boxes, which may appeal to specific customer segments.

Analysis of the French Transactional Data:-

- Based on the plots,we find support level = 0.10 and confidence = 0.5 as an appropriate metric for carrying out the rule generation

- **Analysis of Top 10 Rules by Support**
  Many rules contain 'POSTAGE' as a common consequent item, suggesting that customers purchasing specific items (like 'RABBIT NIGHT LIGHT', 'PLASTERS IN TIN CIRCUS PARADE', etc.) are highly likely to also buy postage.
  The high support values (e.g., 16.7 percent for 'RABBIT NIGHT LIGHT' → 'POSTAGE') indicate that these rules apply to a substantial portion of transactions, making these pairs valuable for broad targeting strategies.
  item often associated with 'POSTAGE' may represent products frequently purchased online or as gifts, where postage is necessary. Cross-selling postage with these products could be effective.

- **Analysis of Top 10 Rules by Confidence**
  Rules in this category also largely feature 'POSTAGE' as a consequent item, reflecting high confidence scores around 80-97%.
  High-confidence rules, such as 'STRAWBERRY LUNCH BOX WITH CUTLERY' → 'POSTAGE' (93.75%), indicate a high likelihood that customers who buy one item will buy the other.
  However, rules like 'SET/6 RED SPOTTY PAPER PLATES' → 'SET/6 RED SPOTTY PAPER CUPS' (96%) highlight associations that may represent themed purchases for events or parties.
  With high confidence values, these item pairs are strong candidates for bundled promotions or suggested purchases in online shopping.

- **Analysis of Top 10 Rules by Lift**
  Lift measures the strength of association, with higher values indicating a stronger association than expected by chance.
  For example, the rule 'SET/6 RED SPOTTY PAPER CUPS' + 'SET/20 RED RETROSPOT PAPER NAPKINS' → 'SET/6 RED SPOTTY PAPER PLATES' has a very high lift (7.58), suggesting a very strong association and a likely opportunity for bundling or discounting together.
  The high lift values for themed items like 'SET/6 RED SPOTTY PAPER CUPS' and 'SET/20 RED RETROSPOT PAPER NAPKINS' with 'SET/6 RED SPOTTY PAPER PLATES' reflect customer interest in matching sets. These could be marketed as packages, especially for event-based promotions.

Analysis of the Spain Transactional Data:

- **Top 10 Rules by Support**
  The rules with the highest support often include POSTAGE as the consequent, meaning items like REGENCY CAKESTAND 3 TIER and ASSORTED COLOUR BIRD ORNAMENT frequently co-occur with POSTAGE. Example: "PACK OF 72 RETROSPOT CAKE CASES → POSTAGE" has a support of 0.122 and a confidence of 1.0, indicating high frequency in the dataset.

- **Top 10 Rules by Confidence**
  These rules show combinations where the antecedent nearly always leads to the consequent.
  Items like "POPPY'S PLAYHOUSE KITCHEN" and "POPPY'S PLAYHOUSE BEDROOM" have a confidence of 1.0, suggesting a very strong association. Example: "4

TRADITIONAL SPINNING TOPS → POSTAGE" has a confidence of 1.0, meaning whenever the antecedent is present, the consequent is also present.

- **Top 10 Rules by Lift**
  High lift values highlight strong associations that go beyond mere co-occurrence. The association between LUNCH BAG CARS BLUE and LUNCH BAG BLACK SKULL has a lift of 18.0, indicating a strong, unique pairing. Example: "LUNCH BAG PINK POLKADOT and LUNCH BAG CARS BLUE → LUNCH BAG BLACK SKULL" has a lift of 18.0, signifying a highly beneficial association in a marketing context.

Analysis of the Ireland Transactional Data:

- **Top 10 Rules by Support**:
  The highest support rule indicates that customers who bought a 'Green Regency Teacup and Saucer' also bought a 'Roses Regency Teacup and Saucer' with a support of 0.119, high confidence (0.911), and a lift of 5.39.
  Another strong rule in terms of support is 'Pink Regency Teacup and Saucer' → 'Roses Regency Teacup and Saucer' with support at 0.1038 and confidence of 0.964, highlighting frequent purchases of these complementary items.

- **Top 10 Rules by Confidence:**

  These rules have a confidence of 1.0, indicating perfect association between antecedents and consequents.
  For instance, 'Regency Tea Plate Green, Regency Cake Stand 3 Tier, Pink Regency Teacup and Saucer' → 'Roses Regency Teacup and Saucer' appears with a confidence of 1.0 and lift of 5.91, indicating customers who buy these items are very likely to purchase the 'Roses Regency Teacup and Saucer.'

- **Top 10 Rules by Lift:**

  The highest lift values (12.38) suggest exceptional association strength between certain tea plate and cake stand items, such as 'Regency Cake Stand 3 Tier, Regency Tea Plate Pink' → 'Regency Tea Plate Green'.
  This indicates that the purchase of one tea plate color and cake stand strongly suggests a high likelihood of another tea plate color being bought.

2.4 Conclusion

2.4.1 Country-wise Conclusions

Germany

- **Bundling Snack Boxes with Postage:** The high support values for items like *ROUND SNACK BOXES SET OF 4* with *POSTAGE* indicate frequent co-occurrence. Offering these items as bundles or encouraging the purchase of multiple snack box variations with postage may appeal to customers, particularly those buying gifts.
- **Target High Confidence Items:** High confidence rules show that items like *PLASTERS IN TIN STRONGMAN* and *RETROSPOT TEA SET CERAMIC* are often purchased together with postage. Marketing these items together as "frequently bought together" or in bundled offers could increase sales.
- **Strengthen Complementary Item Promotions:** High lift values, such as for *RED RETROSPOT CHARLOTTE BAG* and *WOODLAND CHARLOTTE BAG*, suggest that promoting these items together in advertisements or recommendations could attract customers interested in themed bags.

France

- **Cross-selling with Postage:** Many items, such as *RABBIT NIGHT LIGHT* and *STRAWBERRY LUNCH BOX WITH CUTLERY*, are highly associated with postage. Promoting postage alongside these items in the shopping cart or as a default add-on could encourage customers to buy both.
- **Event-based Product Bundling:** Items like *SET/6 RED SPOTTY PAPER PLATES* and *SET/6 RED SPOTTY PAPER CUPS* have high confidence and lift values when paired. Marketing these items as themed bundles for parties or events could capture a larger audience looking for complete sets.
- **Leverage Popular Themes for Targeted Marketing:** High confidence and lift scores indicate a demand for themed products, making it advantageous to offer discounts on bundled, matching items for celebrations or gatherings.

Spain

- **Popular Postage Pairings:** Items like *PACK OF 72 RETROSPOT CAKE CASES* frequently co-occur with *POSTAGE*. Suggesting postage at checkout or offering shipping-inclusive bundles could boost sales of frequently shipped items.
- **Bundling Themed Lunch Bags:** Strong lift values between items like *LUNCH BAG CARS BLUE* and *LUNCH BAG BLACK SKULL* suggest that themed lunch bags are commonly bought together. Marketing these items as bundles or offering a discount for purchasing multiple designs could appeal to customers.
- **Focus on High Support Rules for Broader Reach:** Items with high support, such as *PACK OF 72 RETROSPOT CAKE CASES*, should be targeted in marketing campaigns, as these items apply to a substantial portion of transactions, maximizing the reach of promotions.

Ireland

- **Promote Complementary Regency Teacup and Saucer Sets:** Items like *Green Regency Teacup and Saucer* and *Roses Regency Teacup and Saucer* frequently appear together in purchases, suggesting a strong customer preference for completing collections. Highlighting these items as complementary purchases in product recommendations could drive additional sales.

- **Encourage Purchase of Matching Regency Items:** High lift values between items like *Regency Cake Stand 3 Tier* and *Regency Tea Plate* (different colors) show that customers often buy variations together. Promoting complete sets or offering bundle discounts for multiple colors or items in the Regency line can appeal to those looking to expand their collection.

- **Upselling Through High Confidence Rules:** For items with perfect confidence (1.0), such as *Regency Tea Plate Green* and *Pink Regency Teacup and Saucer* with *Roses Regency Teacup and Saucer*, upselling strategies could include special offers to complete the set or package deals.

2.4.2 Final Conclusion

Across all regions, the analysis reveals clear opportunities to drive sales by leveraging item associations through bundling, cross-selling with postage, and themed promotions. Products that commonly co-occur, particularly with postage, can be marketed together to improve customer convenience and increase sales volume. Themed items, such as matching sets of tableware or snack boxes, can be bundled to attract customers interested in curated or gift-ready purchases. By implementing these strategies, businesses can boost revenue, improve customer satisfaction through convenience, and capitalize on customer purchasing patterns that suggest a strong affinity for certain item combinations.

Bibliography

[1] R. Agrawal, T. Imieliński, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 1993, pp. 207–216.

[2] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach," *Data Mining and Knowledge Discovery*, vol. 8, no. 1, pp. 53–87, 2004.

[3] M.-S. Chen, J. Han, and P. S. Yu, "Data Mining: An Overview from a Database Perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, no. 6, pp. 866–883, 1996.

[4] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic Itemset Counting and Implication Rules for Market Basket Data," in *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*, 1997, pp. 255–264.

[5] M. Hahsler, S. Chelluboina, K. Hornik, and C. Buchta, "The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets," *Journal of Machine Learning Research*, vol. 12, pp. 1977–1981, 2007.