Market Basket Analyis

Akbar Zeb Khan - 2012236   
*Computer Science (CS)*   
*Szabist*Karachi, Pakistan  
[bscs2012236@szabist.pk](mailto:bscs2012236@szabist.pk)

*Abstract*— Market Basket Analysis helps find links and patterns in how customers shop. This advanced data mining method works best by using the Apriori Algorithm to discover hidden item patterns. The goal of this work is to use shopping list analysis on sales data from UK, Germany and France. This will help discover relationships between products. By using steps like support, confidence and lift, the Apriori algorithm can make association rules. These rules are guides that show connections between items or actions. It also tells us about things bought together, and the rules we get can help with setting up products for selling, maker plans for advertising sales, and managing what stuff we have. The results, shown by charts, scatter plot, interactive map and heat-maps, show how the Apriori process can find useful patterns hidden inside store data. By looking at what customers buy, we learn better how they shop. This shows that retail stores can make their business better and earn more money. We want to give a better understanding of buying habits by including information from transactions across countries too.

# **Introduction**

Understanding what customers want and why they buy particular things is very important to businesses wanting staying power in today's retail world. This is where market basket analysis becomes indispensable. Market basket analysis is a detective's tool for retailers that helps them find the connections between products which consumers tend to buy together. This start is our opening into the world of Market Basket Analysis, specifically, how it can peel back the mystery around customer shopping habits and patterns.

The term Market Basket Analysis refers to the uncovering of relationships between products that are often chosen in combination. For retailers aiming to maximize the arrangement of products, design targeted marketing strategies or improve inventory control, this information is priceless. A well-known association rule mining technique is the Apriori algorithm, which we use to extract meaningful patterns from our transaction data.

The purpose of this project is to give an in-depth look at the use of Market Basket Analysis across different multi-national markets. We're not just a local project. Our Dataset includes over 38 countries. For this project, We will use item data from 3 countries. Thus, by employing the Apriori algorithm to uncover associations and then measuring their importance using support, confidence and lift metrics we hope for something like actionable insight.

Once the association rules are uncovered, we will analyze the item sets through different visualization techniques. Pie chart will be employed to visualize predicted item-sets on the basis of different metrics. First we will visualize the item set on the basis of the lift, confidence and support. Scatter plot will be used to display confidence, life and support of overall item set, per country. Heat-map will visualize item set based on per country.

In the contrary we have used the same structure in power-bi, we have showed the support basket, for the Market basket analysis. Confidence prod1 is used to show the, relationship of a certain products or item with other item in the data-set. This same is used to show case the relationship among other items as well and thus we found the confidence of the item, by making a pair. We also have showed cased this in the support basket.[[1]](#endnote-1)

The next few sections will discuss the literature review, problem statement and methodology. Then we move on to in-depth examination of code implementation. The project ends in a complete examination of results and conclusions based on the found association rules. Through this exploration, we hope to illustrate the real-world applications of Market Basket Analysis in improving retail operations and making better business decisions.

# **Literature Review**

### In the commercial world, Market Basket Analysis has become a useful tool where it lets businesses find hidden links among things they can sell. The Market Basket Analysis started in early work on researching markets and handling data. It became more famous when connection rules methods were developed for grouping items together. The reason to do this book review is to look at the main ideas, past changes and modern methods of Market Basket Analysis.

## Historical Evolution of Market Basket Analysis

This para will provide a brief overview of how the approach to basket analysis in markets has changed over time. For checking out what customers buy, market people studied their actions by seeing which items are bought together. We can use this to find out where Market Basket Analysis was first created. Ever since Agrawal et al. wrote about the Apriori algorithm in 1993, this change from studying by hand to quick thinking has become very strong. Apriori broke new ground. This tool was able to fast find out connections between big shopping data. It allowed stores to see consumer habits easily. [1]

## Foundational Concepts in Association Rule Mining

The basis of Market Basket Analysis is association rule mining, which generates rules to show the relationship between things in a transaction. The fundamental metrics guiding association rules include:

Support: Shows how often a rule is seen in the data set. Stronger connections show values more aligned with 1.0.

Confidence: A way of checking how dependable a rule is: this means, how likely is it that if one thing happens then another will too? Yes, and all these factors added together help to better understand and predict what might happen in the future. [1]

Lift: It shows how important something is when another thing happens with it, while considering its likely chance of happening.

## **Applications of Market Basket Analysis**

The use of Market Basket Analysis is larger than just stores. It also applies to places like health care, phone companies and online shopping websites. In the world of buying and selling, Market Basket Analysis affects where you put your things. It also helps with knowing how much you need for different places or people and making plans that are special just for certain customers. It's used by people in health care to find links between medical treatments and how patients are doing. This aids in creating medicine supported by evidence. [2]

## Challenges and Advancements

Market Basket Analysis has helped a lot, but the issues are still there. To handle a lot of data, speed up algorithms, and worry about privacy issues, research is always going on. With better use of smart computer skills and joins with artificial intelligence, machine learning techniques will make Market Basket Analysis more accurate and deeper. [2]

## Cross-Cultural Considerations

In a global marketplace, culture is used to find what people want. The number of people living in an area and its location play a big role. A review look at information we found is needed if we want to check retail sales in different places like Germany, the United Kingdom and France using Market Basket Analysis method. It's also important for seeing how it has changed with time. [2]

# **Problem statement**

Interpreting consumer preferences from large retail datasets is a difficulty. The use of Market Basket Analysis on transaction data from Germany, the UK, and France is necessary because traditional methods are insufficient. Establishing product relationships, maximising placement, creating marketing plans, and reducing inventory waste are the objectives. To improve overall retail operations based on client preferences, the challenge is to derive actionable rules from big datasets.

# **Methodology**

The dataset was acquired from kaggle, the dataset contains retail transactions based on each country. Our dataset contains 38 countries retail transactions.

### **Data Collection**

This dataset contains details about each transaction such as Invoice Number, Stock Code, Product Description, Quantity, Invoice Date, Unit Price, Customer ID, and the Country of the transaction.

### **Data Cleaning**

The integrity of our dataset must be preserved at all costs. As a result, the first stage involves in-depth data cleansing and preparing it for analysis. The key steps undertaken include:

* Trimming Description Spaces: To maintain consistency of the data, eliminating unnecessary spaces before and after product descriptions.
* Duplicate Invoice Removal: Eliminate duplicate invoice entries to avoid data distortion.
* Invoice Type Standardization: Consistency in handling by converting the 'InvoiceNo' type to a string.
* Cancellation Transactions Exclusion: Looking at the information, finding out and getting rid of those labeled as 'C' which shows cancelled bills. This helps keep our findings true..
* Column Pruning: Delete the redundant columns 'StockCode', 'InvoiceDate' and so on in order to have a cleaner dataset.

### **Data Transformation**

With a cleansed dataset in hand, the focus shifts to transforming it into a suitable format for Market basket analysis. This involves creating a transactional matrix where each row represents an invoice, each column represents a product, and the cell values indicate the quantity of each product in a specific invoice. The transformation process includes:

* **Grouping by Country:** Separating data for each country to allow for focused analysis.
* **Summing Product Quantities:** Aggregating product quantities for each invoice to capture purchasing patterns.
* **Handling Missing Values:** Filling in missing values with zeros to maintain the integrity of the transactional matrix.
* **Encoding Data:** Converting all positive values to 1 and non-positive values to 0, simplifying the dataset for analysis.
* **Removing 'POSTAGE':** Eliminating the 'POSTAGE' item from the analysis to focus on tangible product associations.
* **Filtering for Relevant Invoices:** Selecting only those invoices with two or more items to accentuate meaningful purchasing patterns.

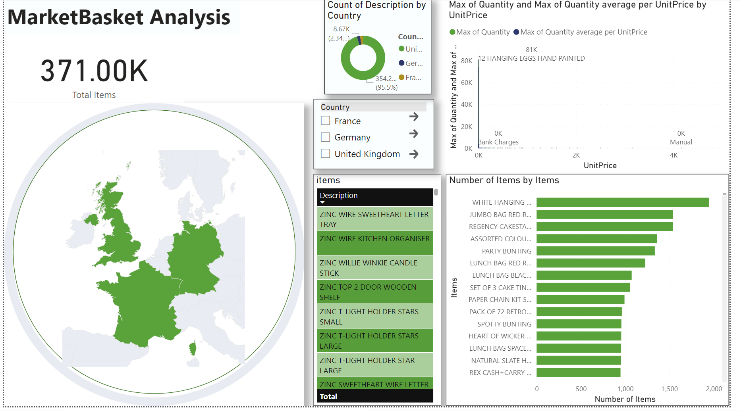
### **Apriori Algorithm and Association Rules**

We have our ready list of data. Then, we use the Apriori method to find what items show up most often together in UK, France and Germany. This helps us work out connections between these things in a detailed way. Putting a low help limit at 0.03 makes sure that we look for big groups of things. Rules are made based on the lift score, a measurement of how much more likely one product is to be bought when another is purchased. The final rules get lined up in reverse order, from highest to lowest. Then they're ready for more studying..

### **Visualization**

To enhance our understanding and interpret the pattern relationships within the dataset, we use various visualization techniques. In Python, we create compelling visualizations, including Bar Chart, Pie chart, heat-map and 3D scatter plots, to illustrate the association rules in terms of lift, support, and confidence. The visualization is done for each country separately. Predicted Analysis is done through visualization to display the frequent patterns based on various metrics.

### **Power BI Dashboard**



This user-friendly interface has maps that show items and associations in stores across three countries. It also has interactive map, the prices of individual items, and details about all the transactions. The UI on Power BI is the main place to see data in pictures and numbers. It shows us more about how people in Germany, United Kingdom and France buy things all put together better. Our board focuses on giving users a good experience. It does this with a nice look and easy ways to get around. People can easily change between parts of the information, learning something. The tool we use, called a dashboard, includes maps. These maps help us understand where retail things are bought in three different nations. They show how products are spread out over the countries. With better analysis, we give businesses future guesses and plans. This allows companies to change their ways as markets shift. Moreover, users can find out how much a product costs and it's result. This can give them a money lesson.

# **Result**

The dataset was acquired from kaggle, the dataset contains retail transactions based on each country. Our dataset contains 38 countries retail transactions.

### **UK Basket**

Fig 1.1

Top Five Generated Rules

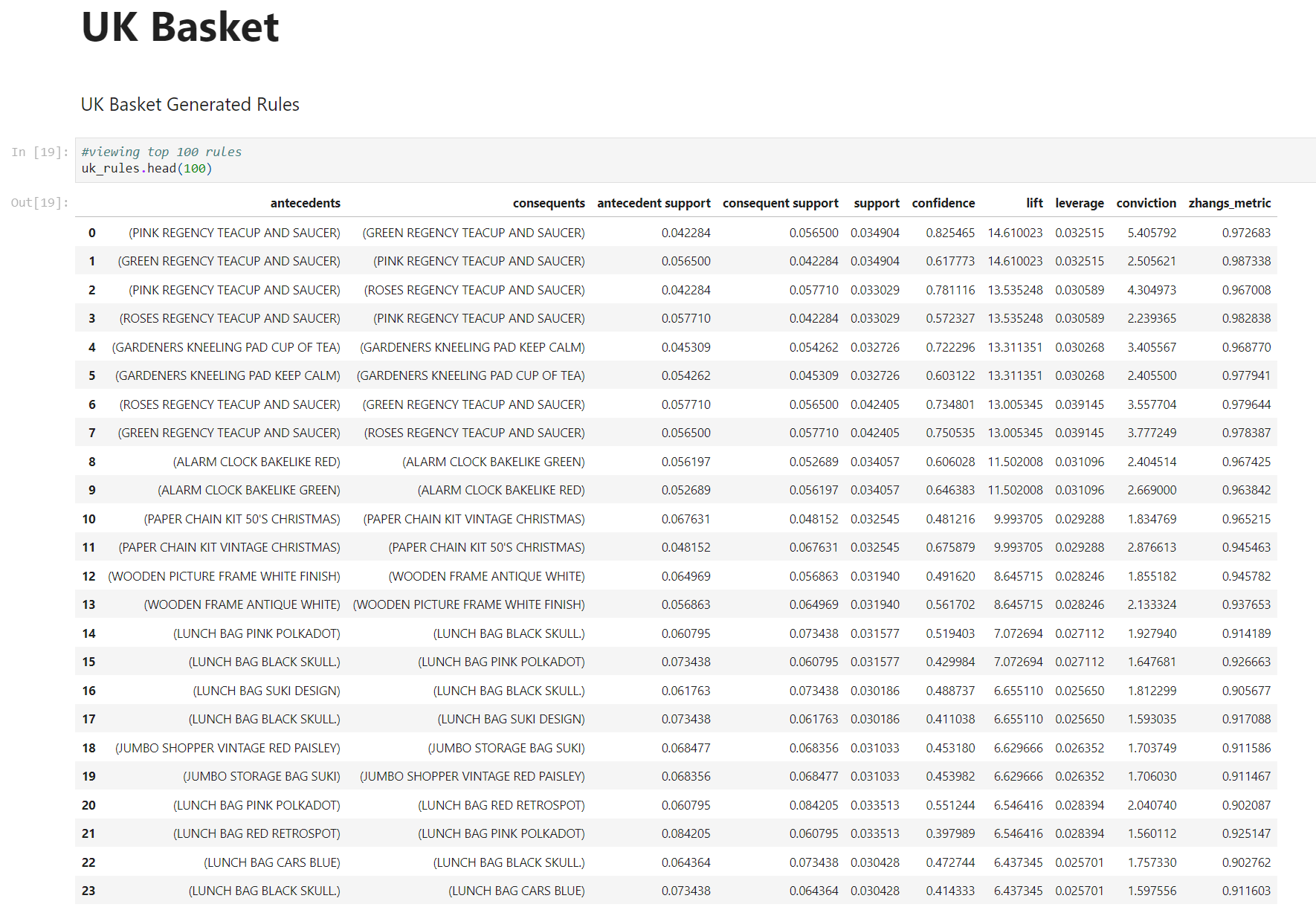
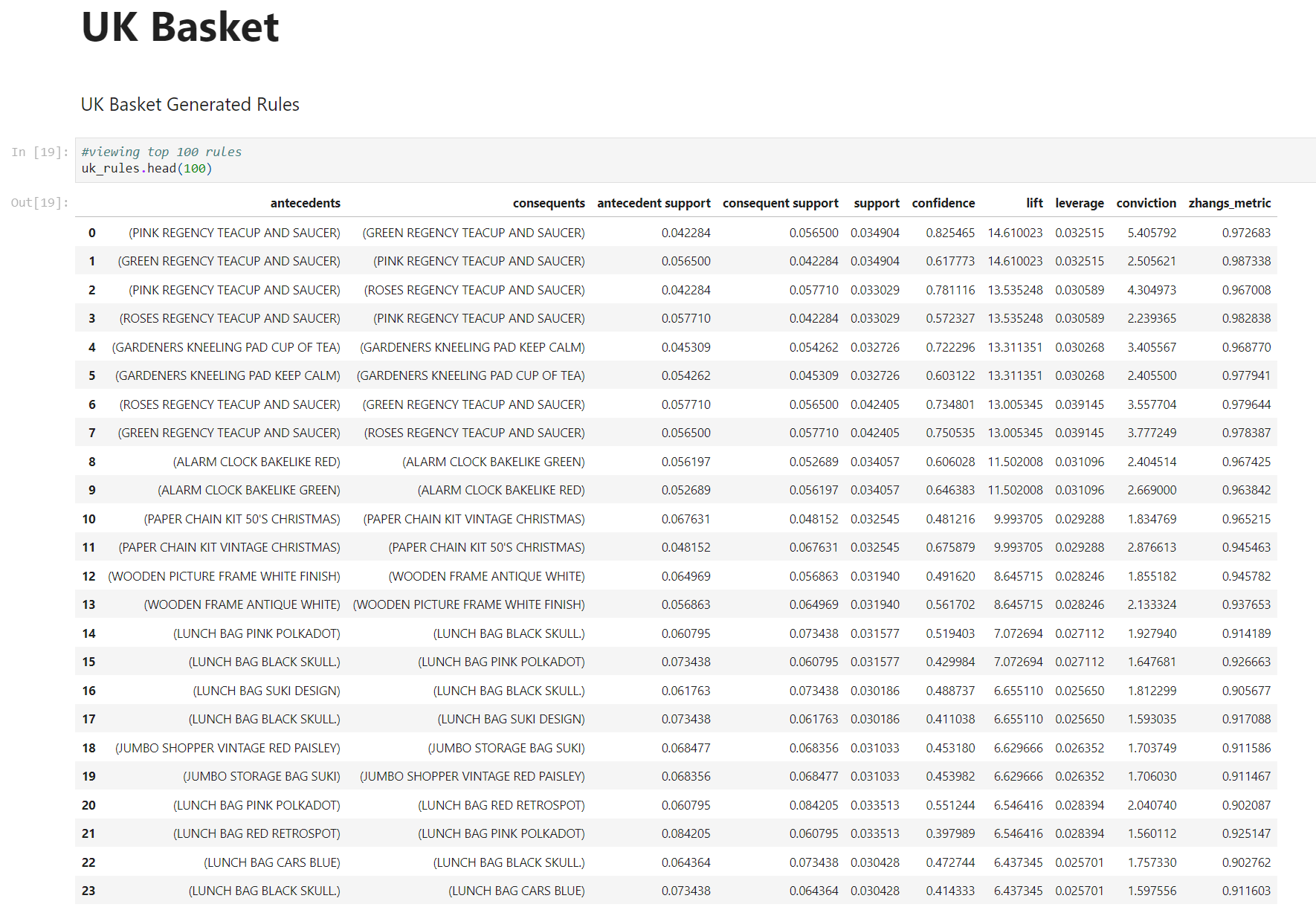


Fig 1.2

Top Five Generated Rules By confidence



### **Germany Basket**

Fig 2.1

Top Five Generated Rules

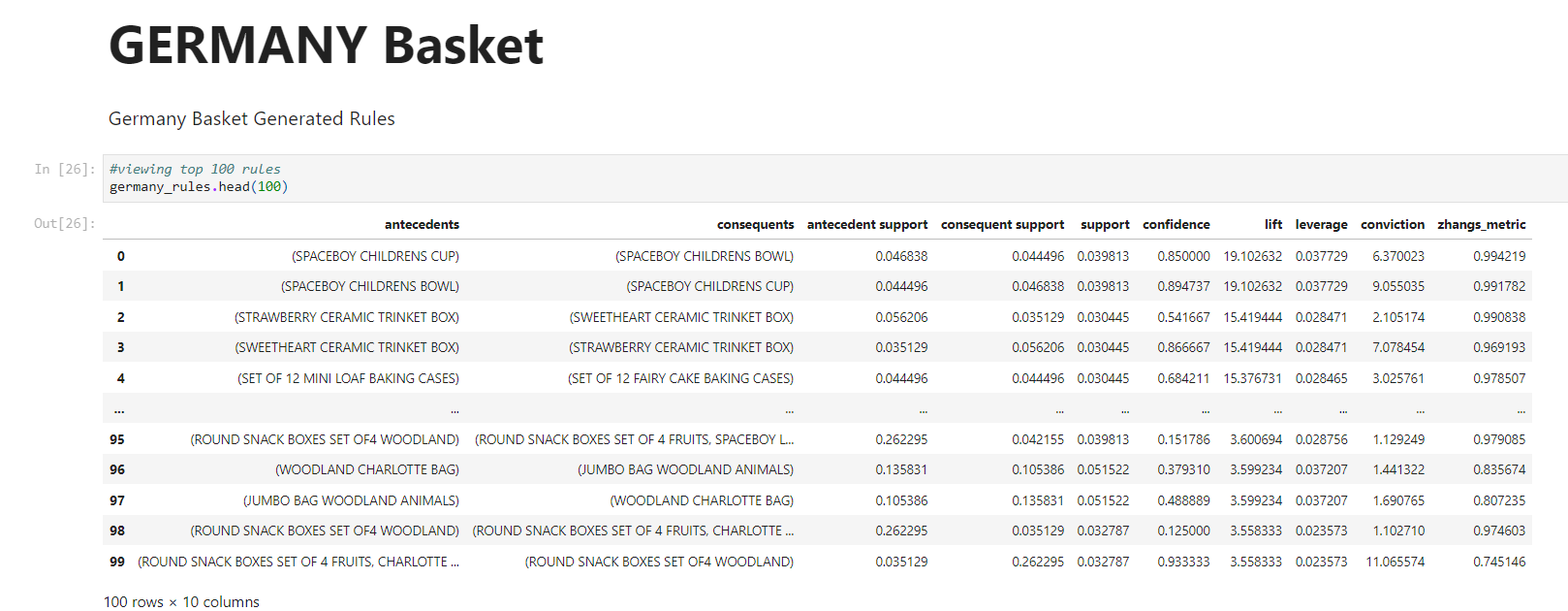
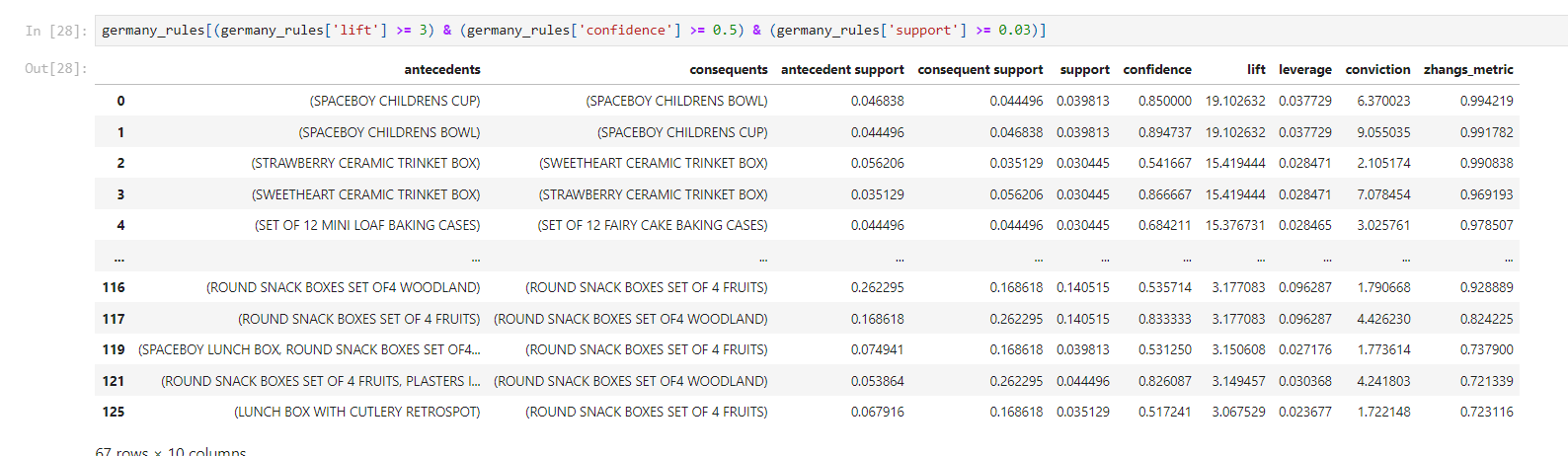


Fig 2.2

Top Five Generated Rules by confidence



### **France Basket**

Fig 3.1

Top Five Generated Rules

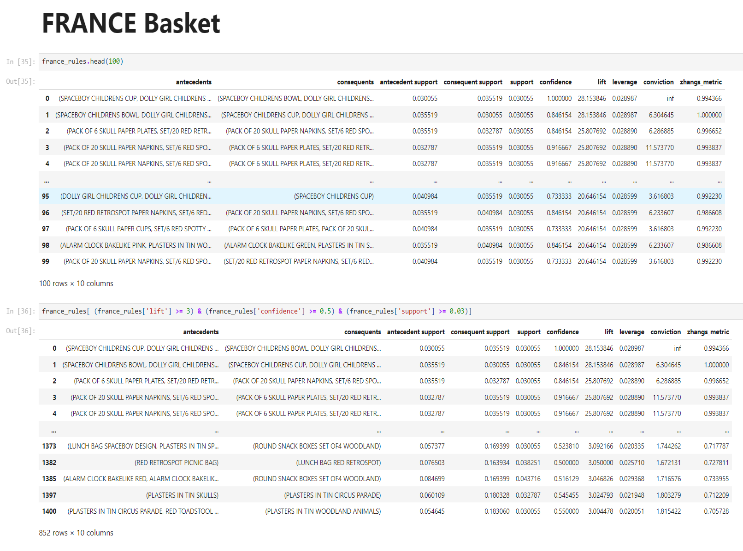
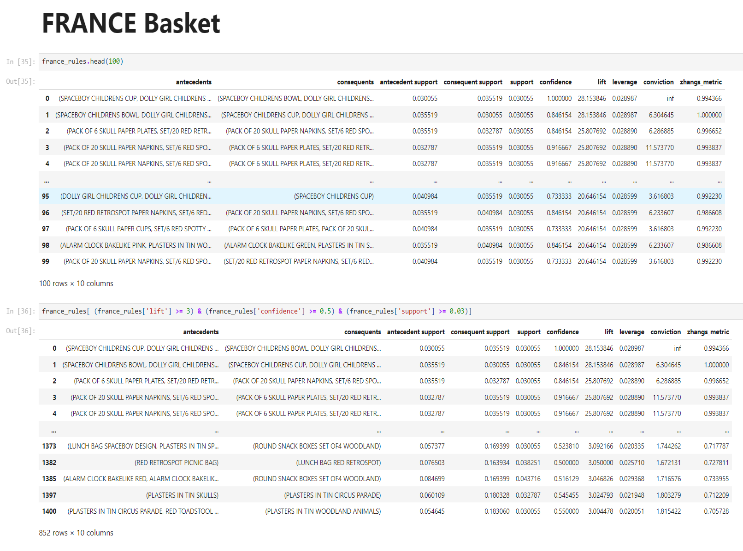


Fig 3.2

Top Five Generated Rules by confidence



# **Conclusion**

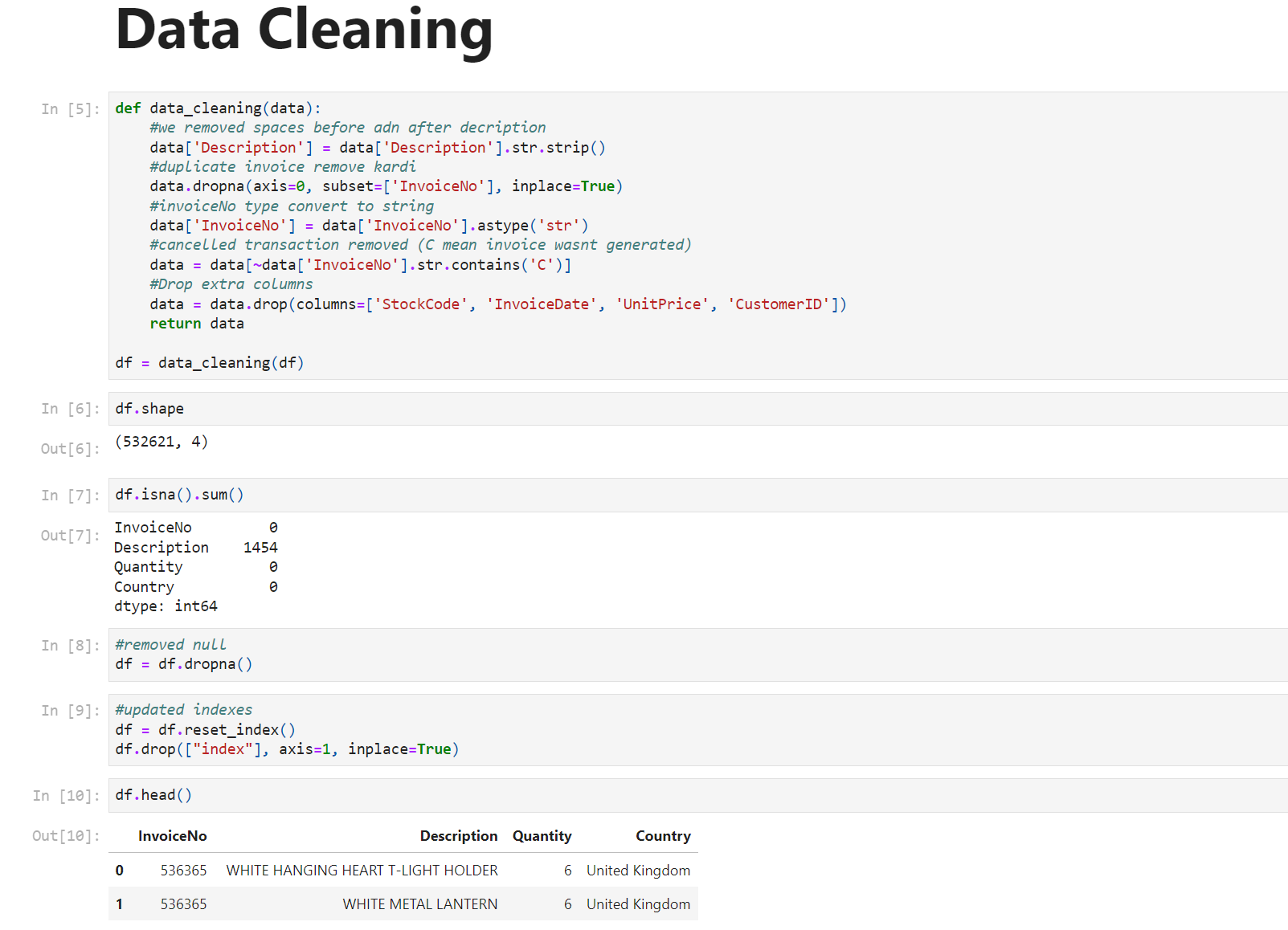
The application of Market basket analysis using Apriori algorithm on our retail data-set provided us an extensive look into item patterns and user buying habits. By going hands-on with exploratory data analysis, generating association rules, and adding a visualization through Python and Power-BI dashboard, we've gained some detailed insights into how customers purchase items and which items are purchased frequently. The findings from the retail data from the United Kingdom, Germany and France has practical benefits for retail stores. These rules can serve as a comprehensive guide for retail businesses, help them in strategic decision-making. Businesses can now stock their shelves with accuracy, ensuring popular products are readily available and what customers really want. The visualization depicts the findings in a user-friendly way and enhances the project depiction.

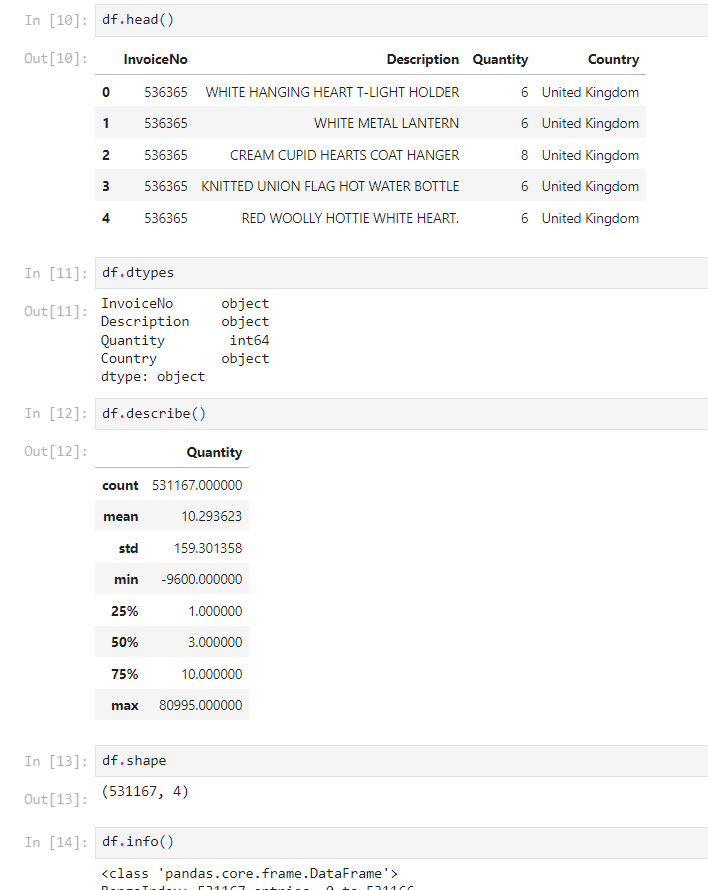
# **References**

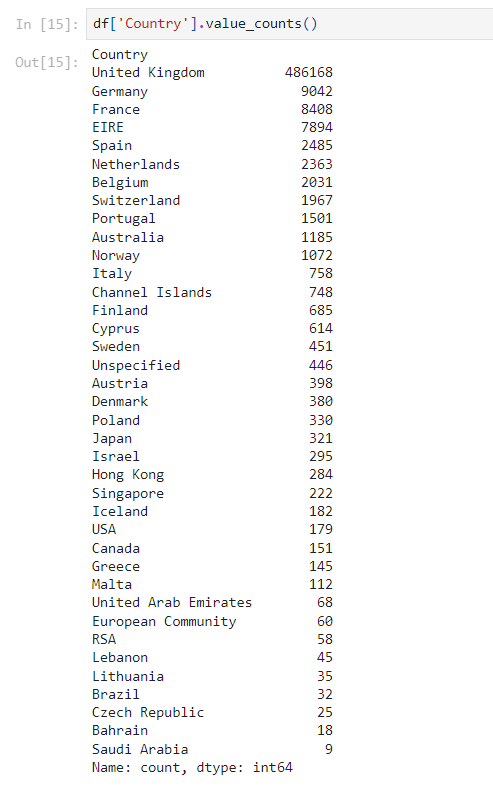
1. <https://dl.acm.org/doi/10.1145/170036.170072>
2. <https://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf>

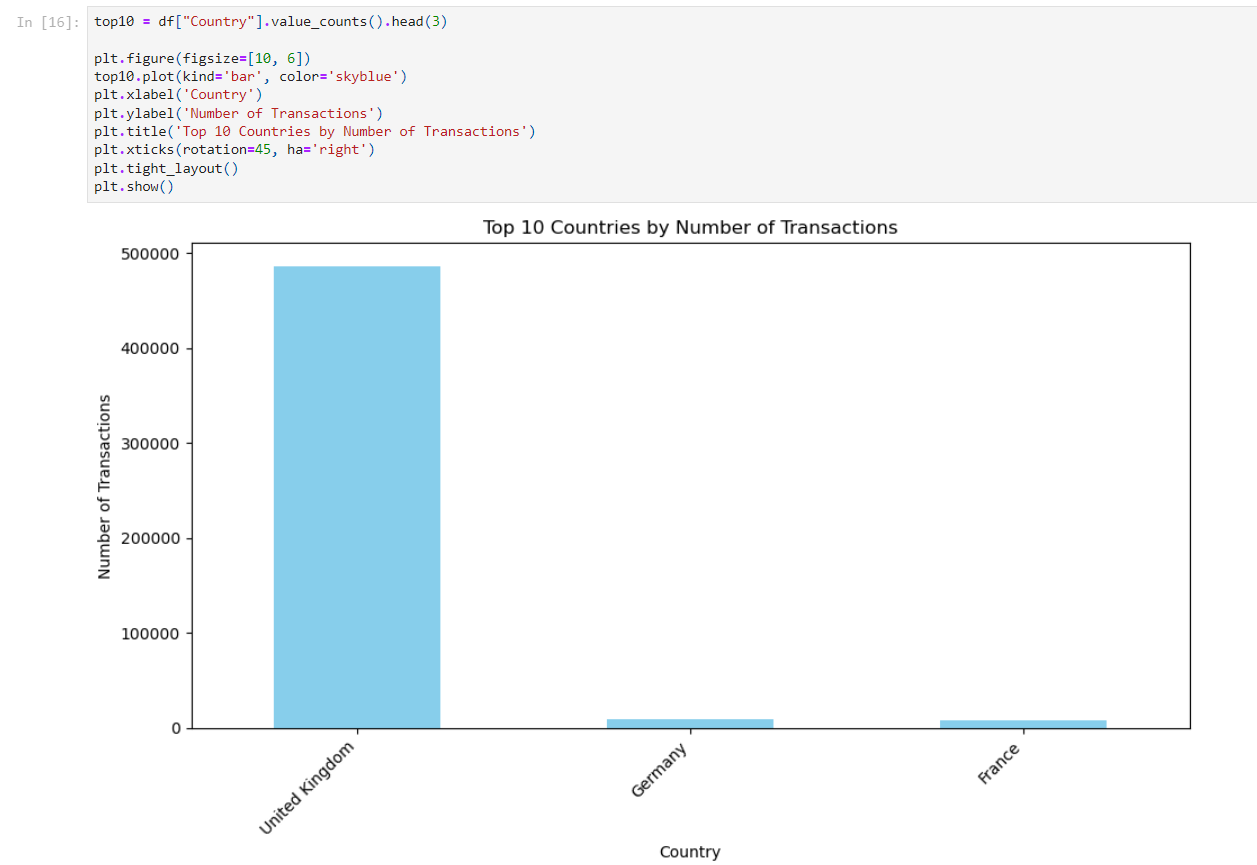
# **Appendix**

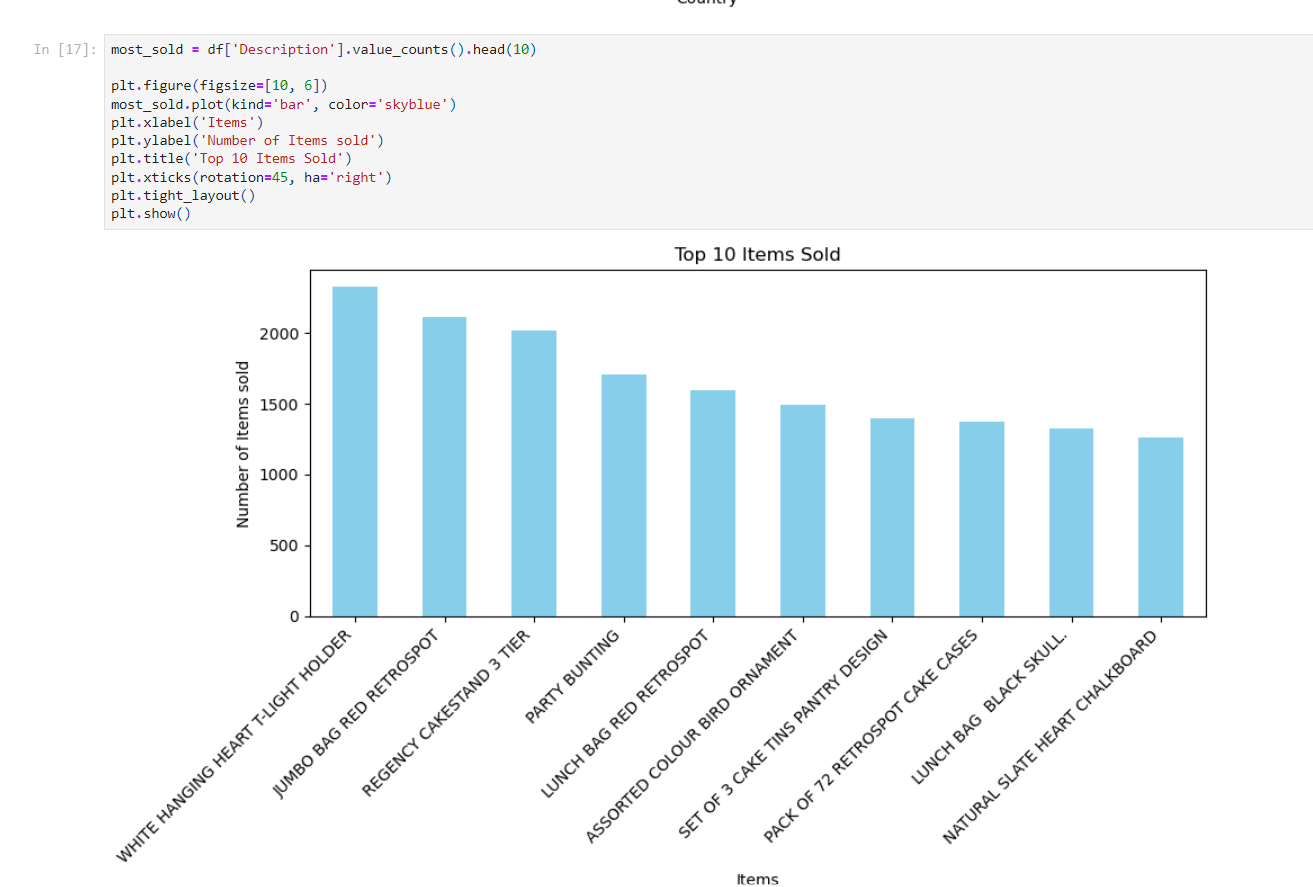


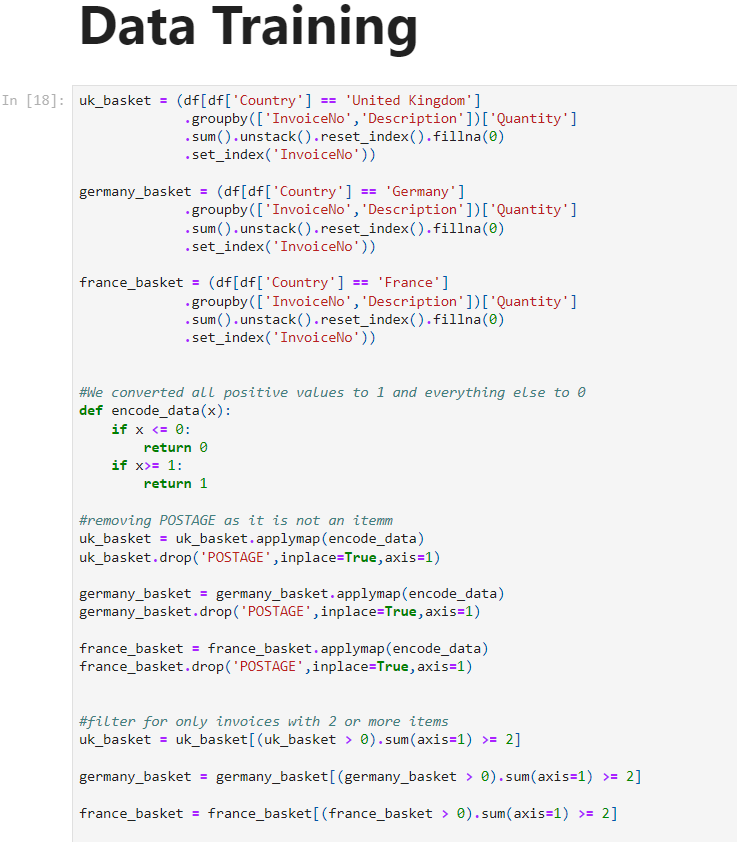


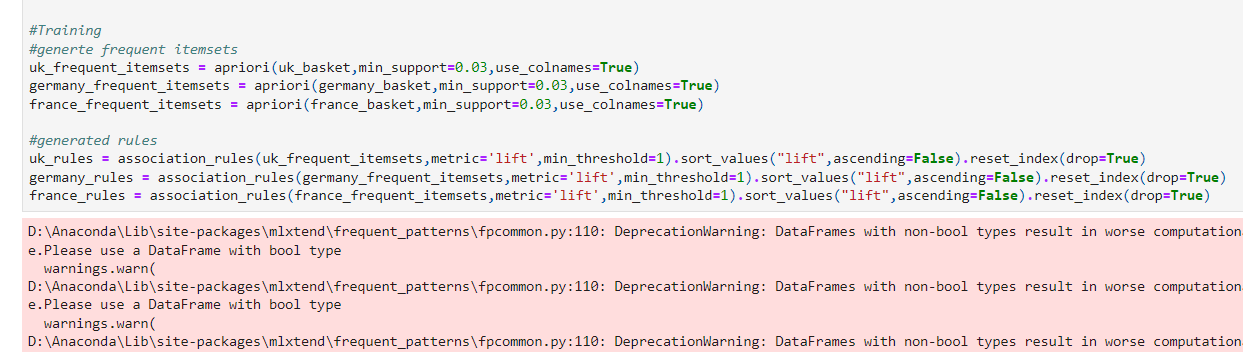


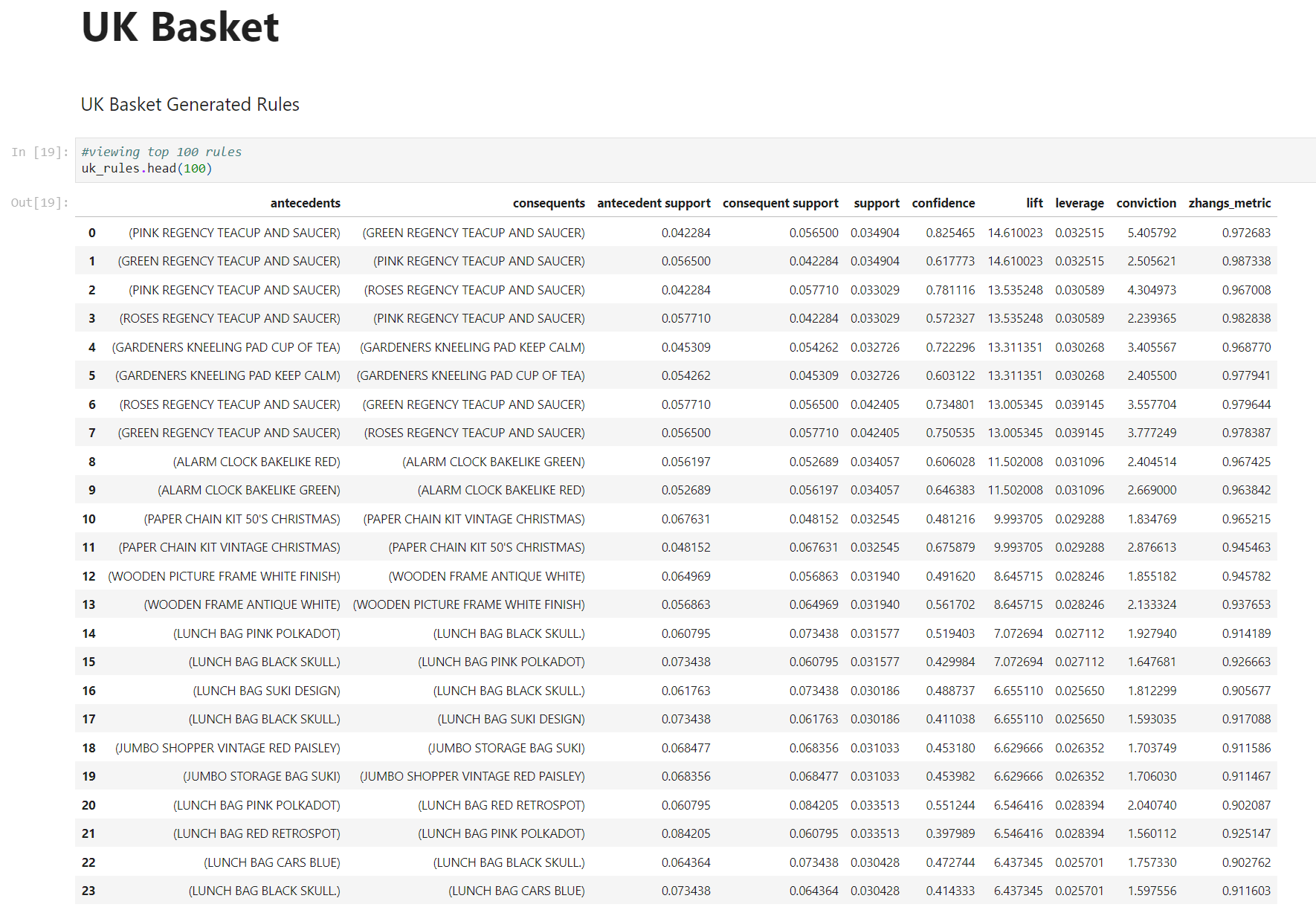






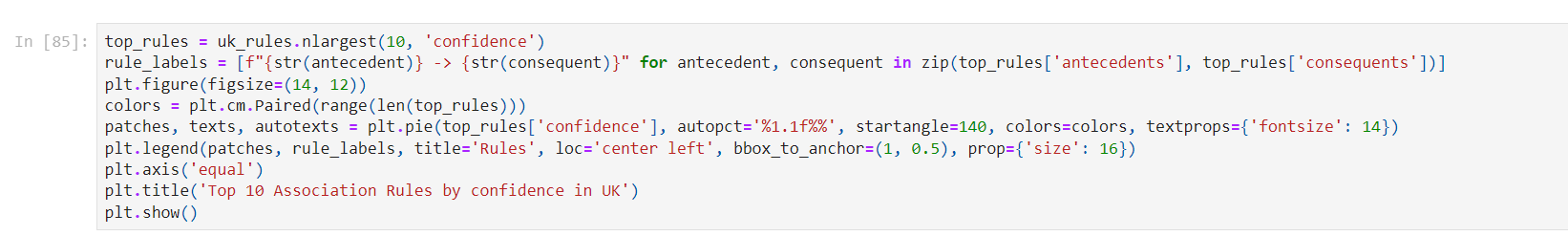


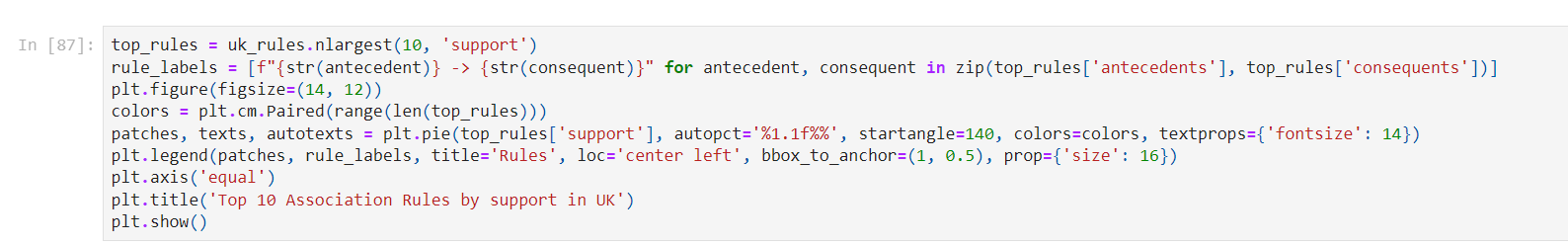


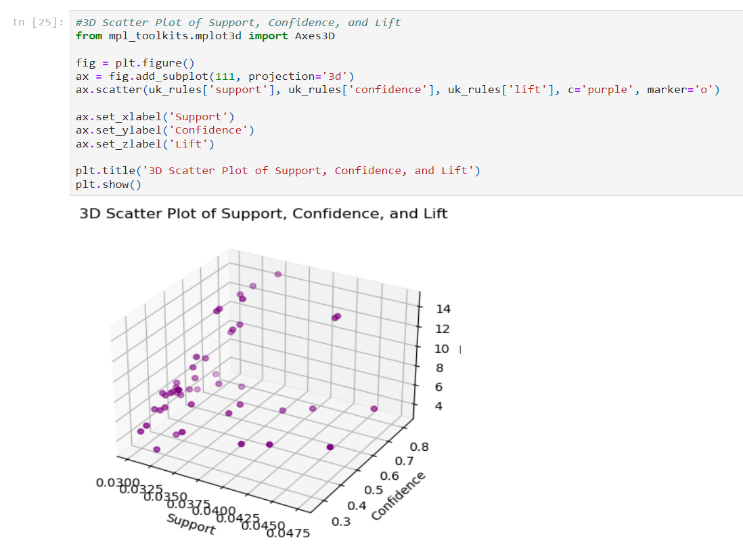


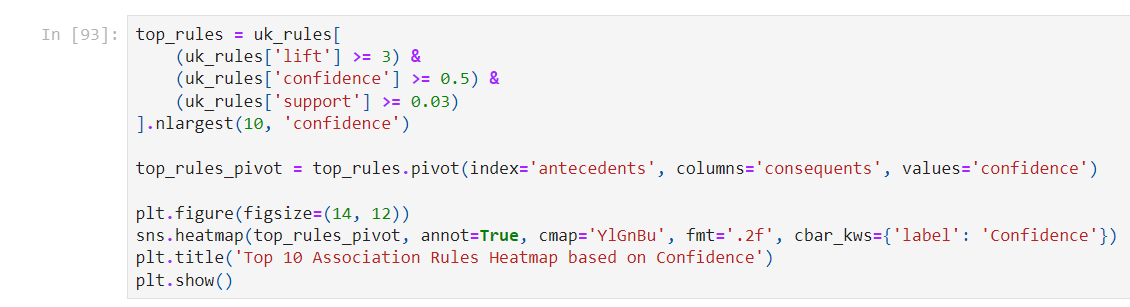


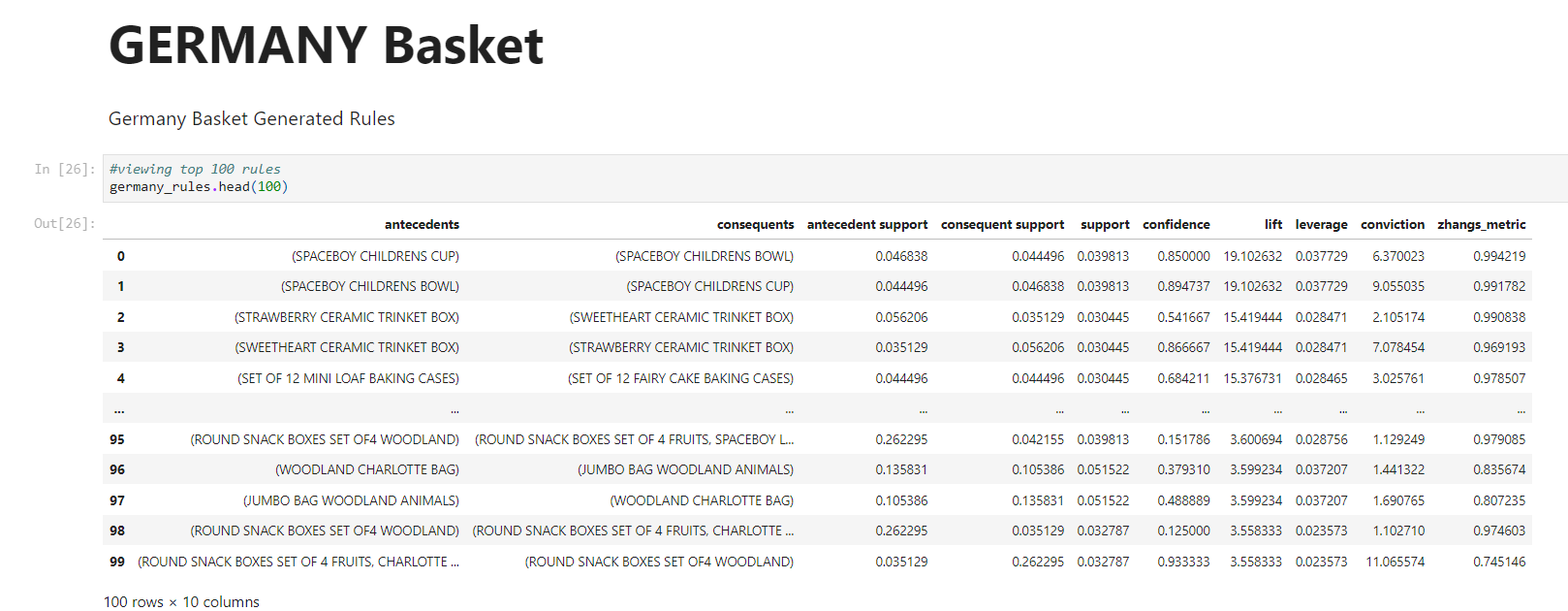


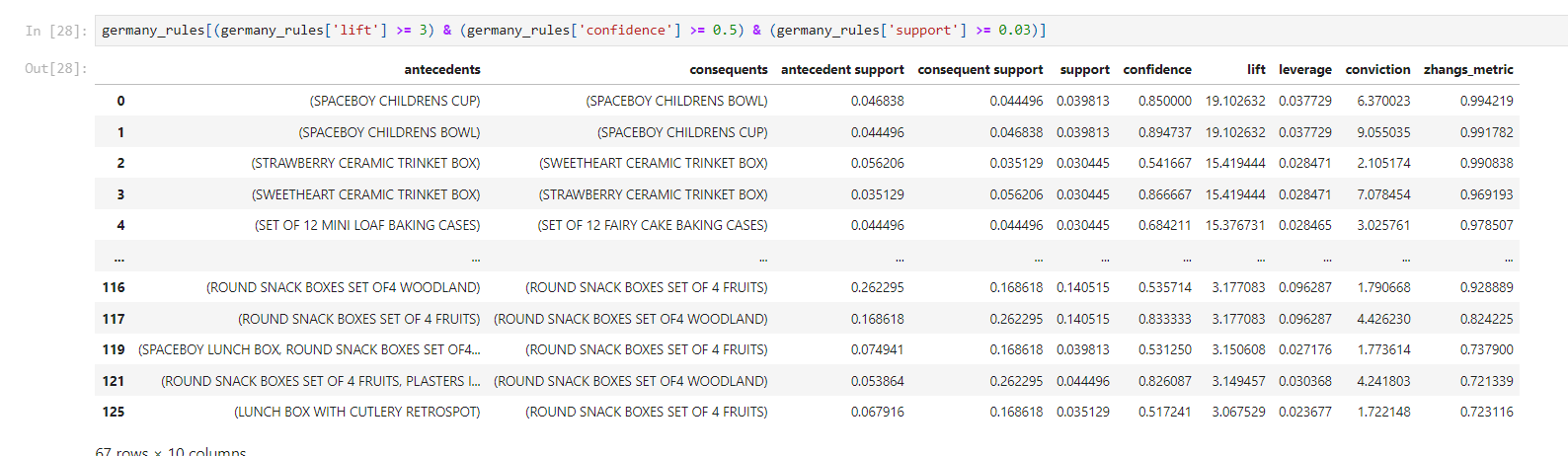


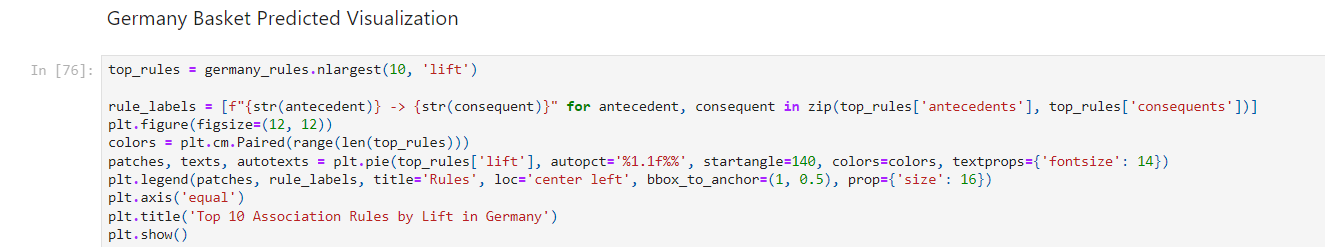


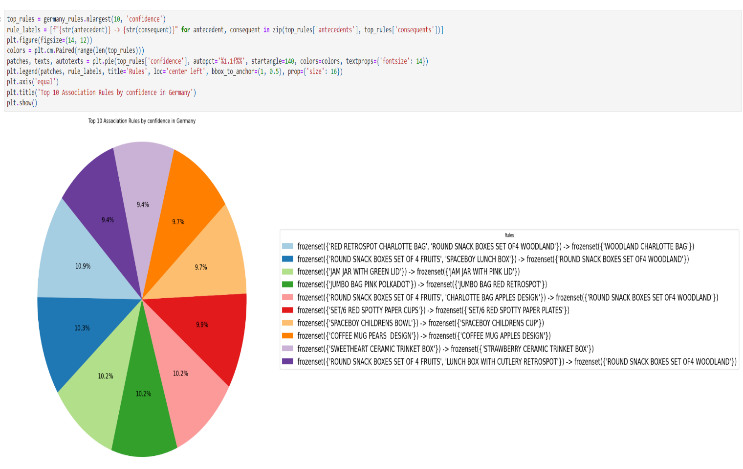


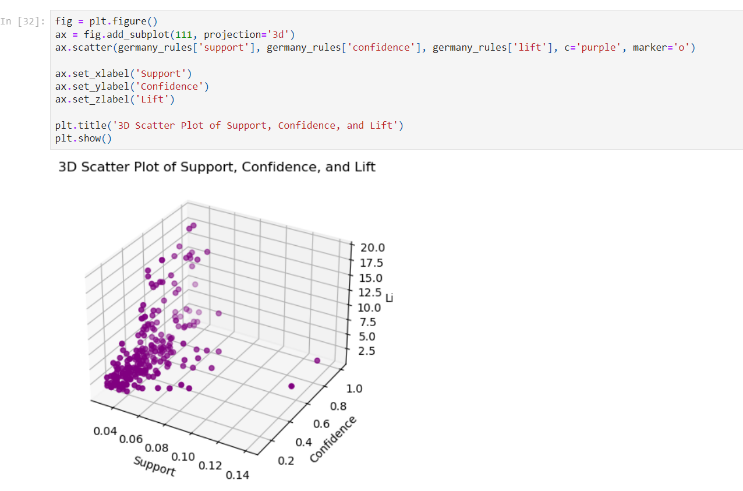




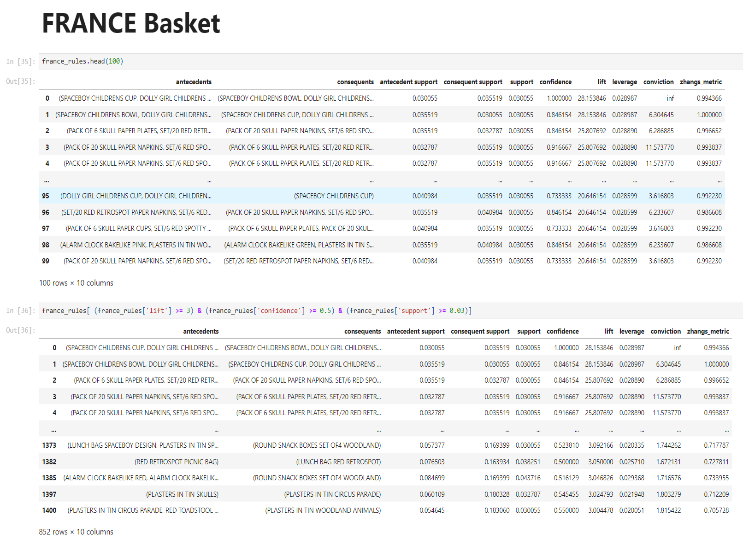


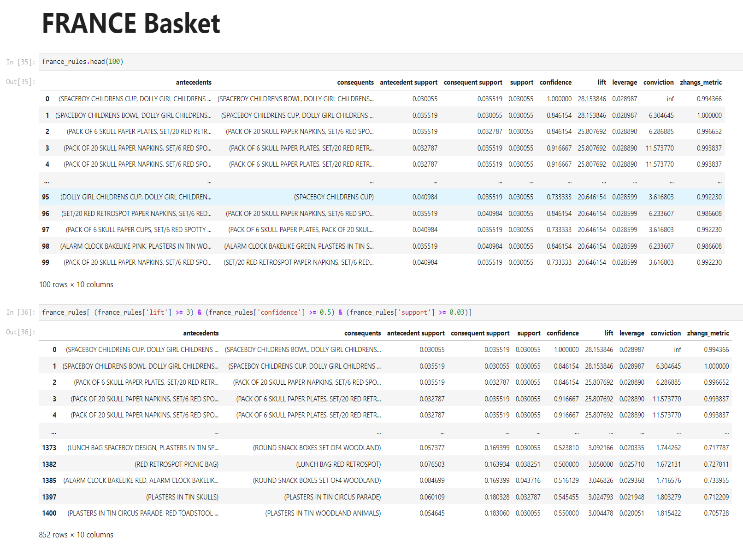


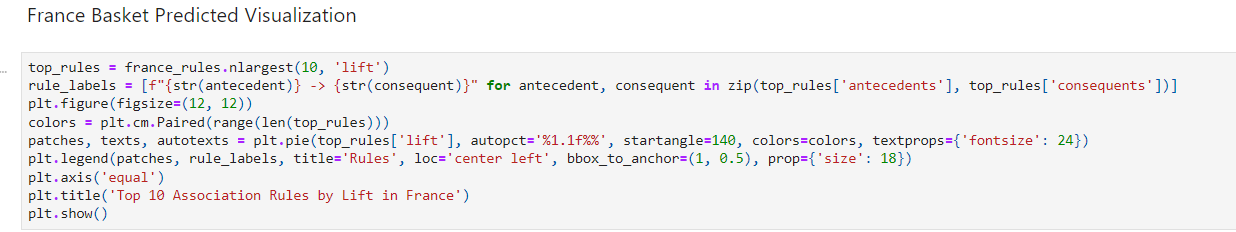


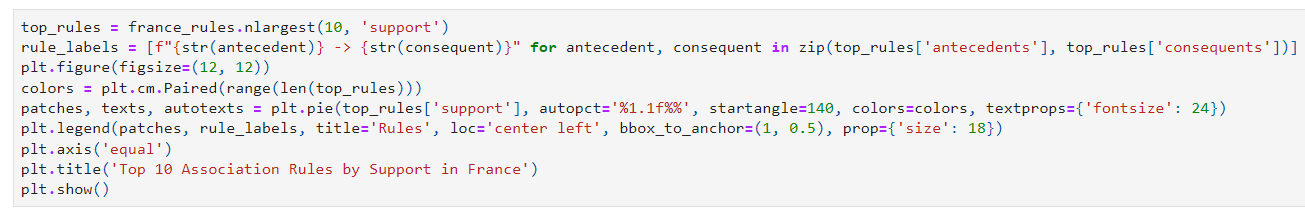


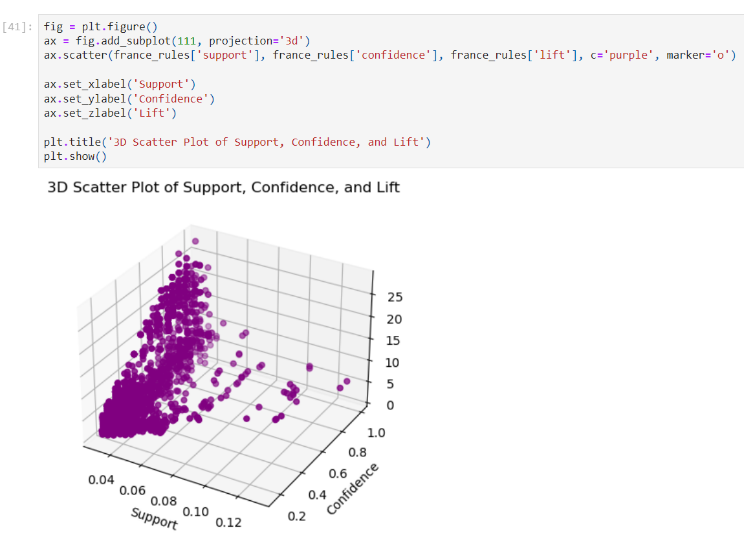




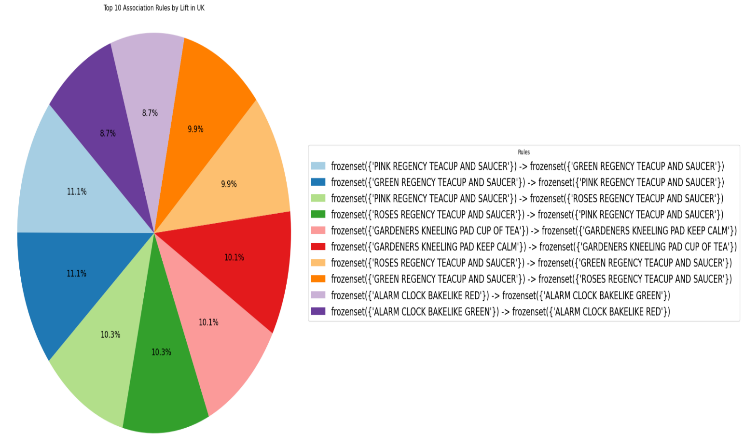


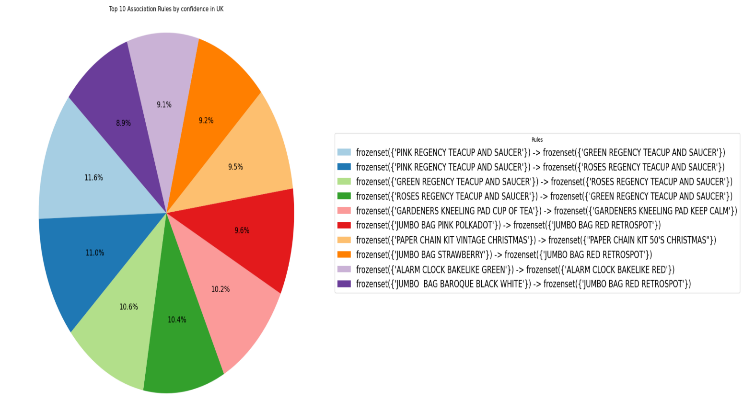


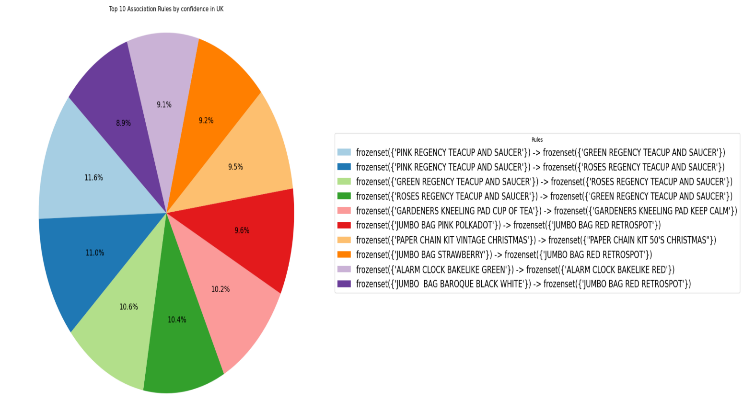


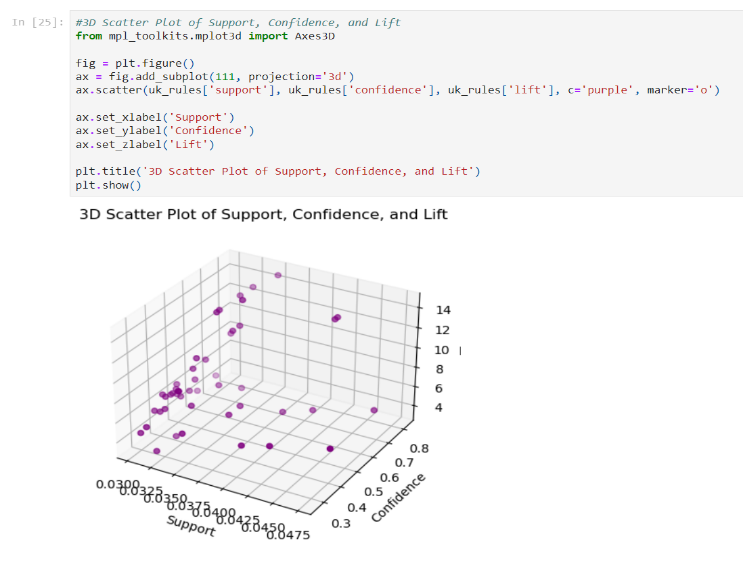


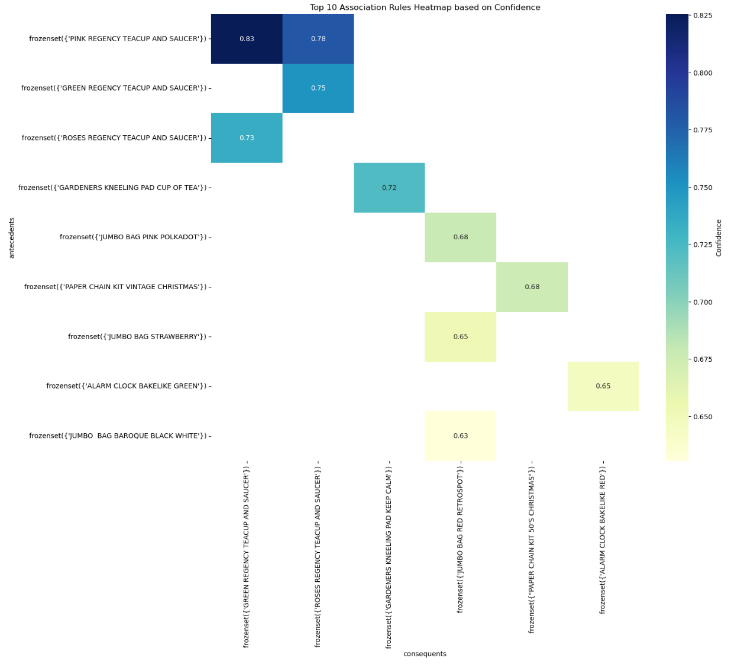


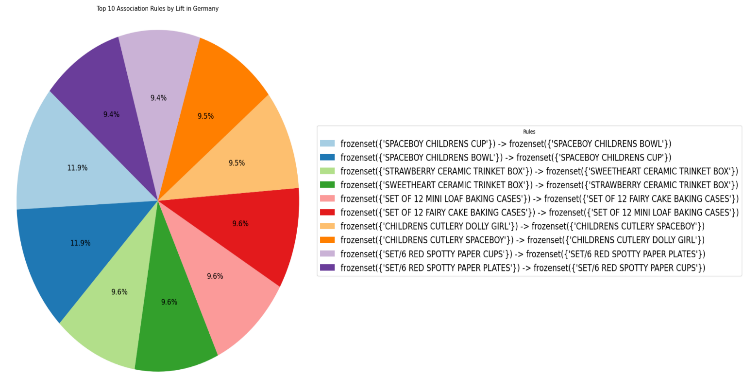


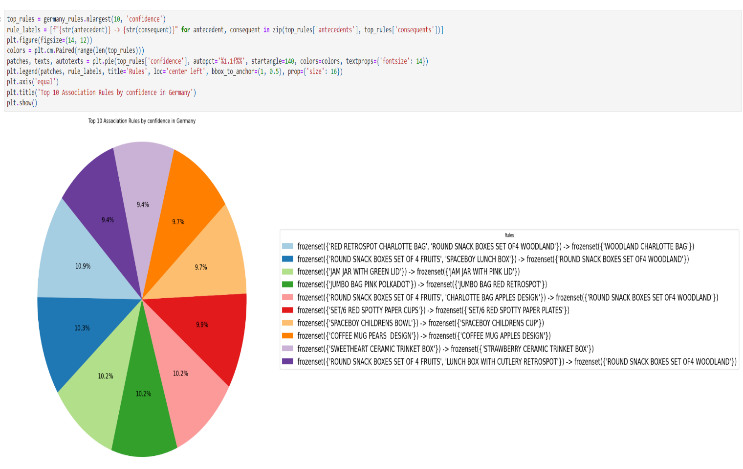


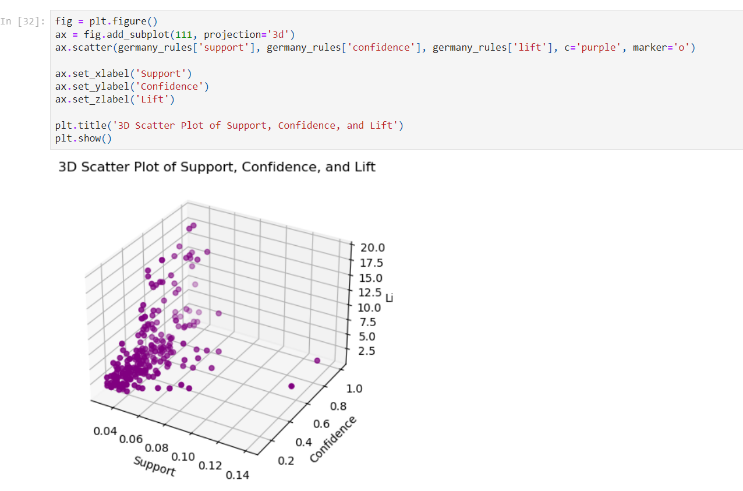


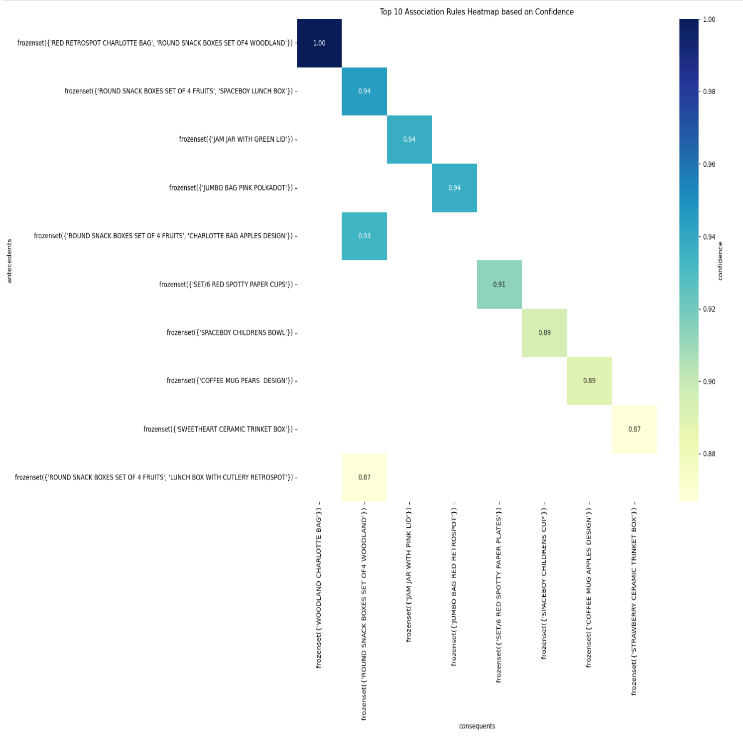


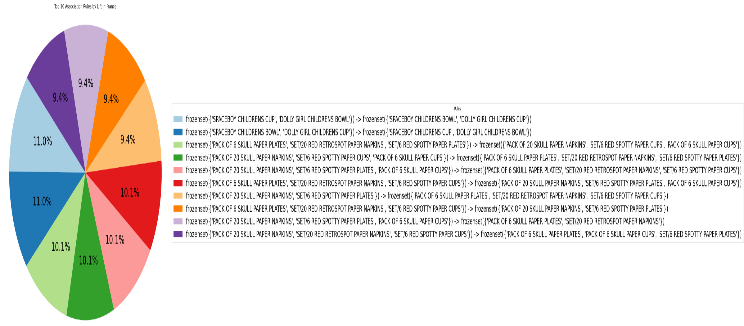


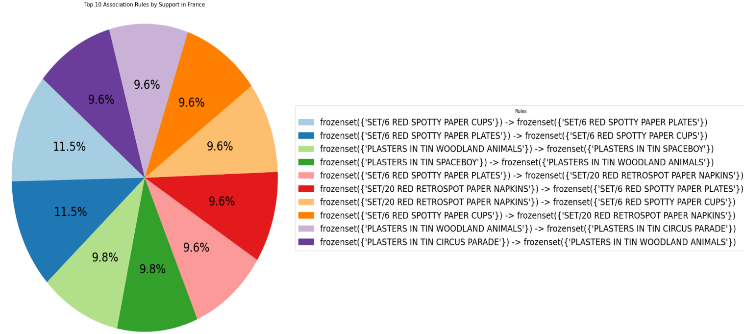


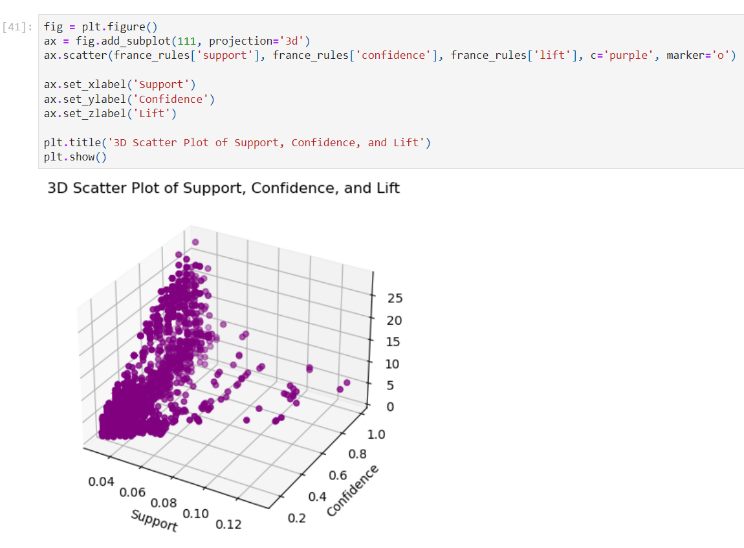














1. [↑](#endnote-ref-1)