Basket Analysis for Online Retail

This project analyzes online retail transactions to uncover product pairings and support the marketing strategy through basket analysis. It focuses on key metrics such as support, lift, leverage, and confidence, using only 2-item sets due to the heavy computational power required for larger itemsets. By understanding which products are bought together, it enables the design of marketing strategies for online retail—with the potential to be applied to online supermarkets as well. These insights can be used for cross-selling, bundling, and targeted promotions.

Data Preparation

Cleaned product descriptions by removing commas, trimming spaces, and converting text to lowercase. Missing values were removed to ensure data quality. Items were then grouped by InvoiceNo to form transaction "baskets" totaling 9700 transactions. One-hot encoding was applied to create a binary matrix representing item presence in each basket, and support (frequency) was computed for every item to identify purchasing patterns.

Frequent Itemset Mining

Used the Apriori Algorithm (mlxtend) to analyze the data, focusing on the top 100 frequent items with a minimum support threshold of 0.005. This approach helped manage computational demands while still uncovering approximately 3000 strong item pairings of length 2.

Association Rule Mining

Metrics analyzed included support, indicating how often items appear together; confidence, the probability of purchasing B given A; lift, measuring the strength of association above chance; and leverage, the absolute increase in joint occurrence over independence. Rules were filtered for strong associations with high lift values (14+). Example insights include alarm clocks in green and red variants with a lift of approximately 14 and confidence of 71%, suggesting cross-selling them as a set, and Christmas Scandinavian decor items like hearts and stars with a lift of around 17 and confidence of 78%, indicating potential for seasonal bundling. High-lift, high-confidence rules reveal natural product bundles that can drive effective marketing campaigns.

Marketing Strategy with Association Rules
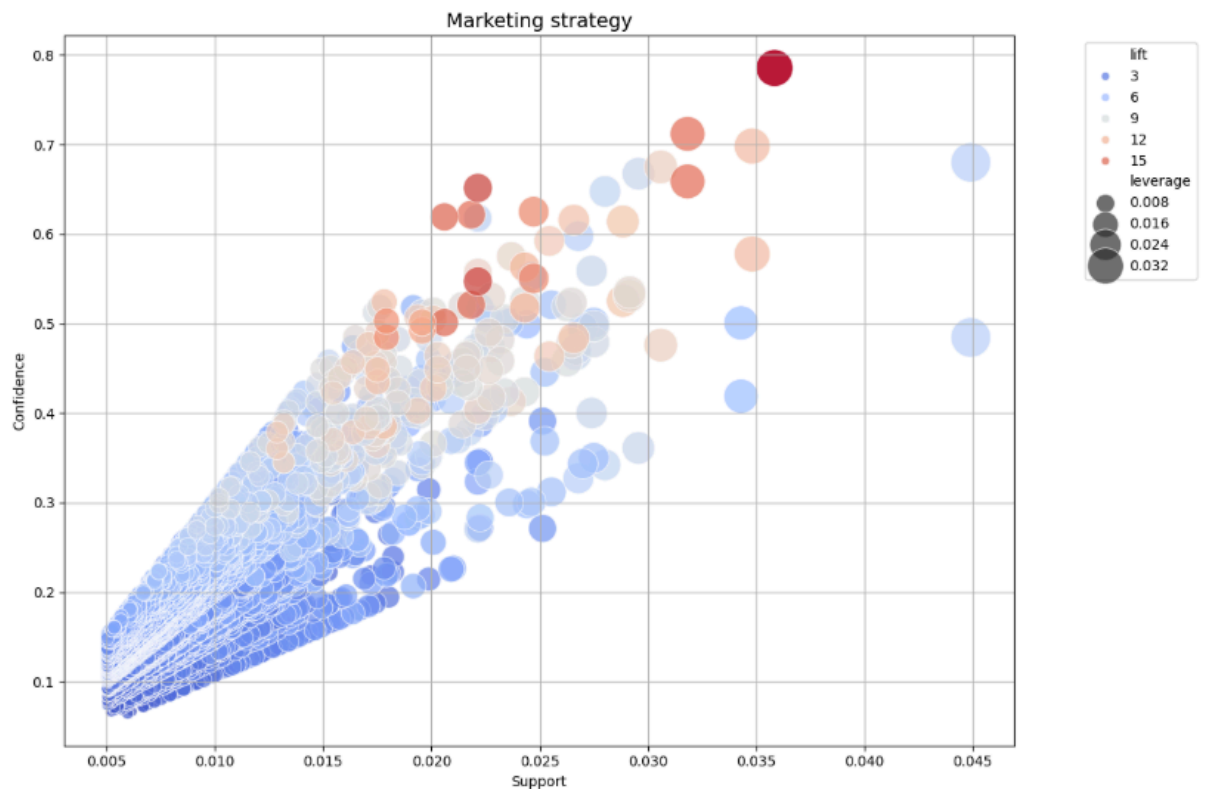
Interpretation:

Marketing Strategy: The bubble plot reveals key patterns for targeting: top-right large red bubbles indicate high support, high confidence, and high lift, making them ideal for cross-selling, bundling, and promotional focus. The middle area represents moderate rules suitable for segmented or seasonal campaigns, while the bottom-left small blue bubbles are weak rules to ignore. Overall, marketing strategy should prioritize campaigns around the strongest top-right rules, bundle and cross-sell high-lift pairs, and target niche but strong-lift, low-support pairs through email segmentation

Conclusion

Basket analysis delivers actionable marketing insights by identifying natural product bundles, supporting personalized cross-selling, and informing data-driven promotions. It helps maximize average basket size and customer lifetime value. By analyzing the final bubble plot, marketing teams can prioritize the strongest and most impactful product pairings to achieve better ROI.