Market Basket Analysis in correlation with Product Assortment Strategies in Supermarkets

Casey Portelli Frankie Inguanez
Institute of Information & Communication Technology
Malta College or Arts, Science & Technology
Corradino Hill
Paola PLA 9032
casey.portelli.h64005@mcast.edu.mt

***Abstract*—This study utilises Python for data analysis to examine the relationship between market basket analysis and product assortment methods. Using online data sets and a survey to investigate customer behaviour, I investigate how market basket research might help with strategic product placement in supermarkets. Insights for improving retail performance and customer experience are provided by the findings, which point to a strong association between market basket analysis and product selection methods. This study emphasises how crucial it is for supermarket managers to use data-driven methods to guide their decision-making.**

***Index Terms*—Market Basket Analysis, Product Assortment Strategies, Apriori Algorithm, Zhang's Metric, Association and Disassociation**

I. Introduction

Comprehending consumer behaviour and refining product assortment strategies are essential for sustaining competitiveness and enhancing profitability in the dynamic retail industry. As key participants in the retail industry, supermarkets are always looking for new and creative ways to improve the shopping experience and increase sales. In this effort, market basket analysis proves to be a useful tool since it provides insights into the preferences and buying habits of consumers.

In order to clarify their relationship and implications for supermarket management, this research investigates the relationship between market basket analysis and product assortment techniques. I explore market basket analysis using Python-based research approaches, making use online data sets and performing an extensive survey to gain a deeper understanding of customer behaviour.

Through this research I seek to address key questions such as:

- In what ways does market basket analysis help supermarket assortment plans be improved and optimised to better serve customers and increase sales?
- How can cross-selling opportunities in supermarkets be improved and complimentary products found using market basket analysis?
- How do external factors like pricing strategies of competitors, consumer preferences, and economic situations effect the understanding of market basket analysis in supermarkets?

This study intends to give supermarket managers useful insights by clarifying the connection between market basket analysis and product assortment methods. These insights will help supermarket managers make well-informed decisions about product placement, inventory control, and marketing tactics. In the end, this study adds to the conversation on retail management practices by emphasising the role that data-driven strategies play in promoting operational excellence and customer satisfaction in the grocery store sector.

II. Literature Review

Researchers have investigated many approaches to optimise assortment methods in the retail industry, with the goal of achieving equilibrium between income generating and consumer pleasure. The results of multiple important investigations are summarised in this review, which provides insights into how assortment optimisation is developing.

Firstly, is Santos's inquiry into online grocery retail assortment optimisation [1], specifically in the rice subcategory, wherein the importance of SKU selection in optimising sales without sacrificing revenue is emphasised. Santos emphasises the value of data-driven decision-making in assortment planning by using transactional logs and involving decision-makers.

Building on Santos's research, Shaikh, Sultan, and Asim explore the topic of big data analytics' influence on optimising assortment strategy in Pakistan's organised retail sector [2] . Their research emphasises how important it is to have highly skilled data scientists and advanced algorithms in order to fully utilise the potential of data. Nonetheless, obstacles including restricted sample sizes and unavailability of data highlight the necessity of additional improvement in subsequent research.

Karki's hybrid strategy for managing retail assortments [3] expands on the ideas presented by Santos, Shaikh, Sultan, and Asim. Karki provides a sophisticated grasp of assortment management strategies by assigning products to customer segments based on their behaviour. Nonetheless, the report notes that consumer preferences are dynamic and that there may be biases in the data, indicating prospects for continued refinement.

Kanagaraj and Venkatesh delve deeper into the topic of data-driven decision-making by examining how analytics tools are used in specialty retail industries [4]. Their research emphasises how crucial data are to creating focused marketing campaigns and improving sales performance. However, issues with privacy and data quality highlight the necessity of ongoing research and development into advanced analytics methods.

Lastly, Tamhankar et al.'s investigation into clustering strategies to raise the accuracy of sales predictions for assortment planning [5] builds a link between conceptual models and real-world implementations. The paper provides practical insights into improving prediction accuracy by combining clustering algorithms with regression and classification-type prediction models. This emphasises how crucial it is to choose the right models depending on item characteristics and demand trends, adding to the current assortment optimization.

Collectively, these studies provide a comprehensive grasp of assortment optimisation in the retail industry, highlighting the importance of data-driven decision-making, ongoing methodology improvement, and a sophisticated comprehension of customer behaviour.

Comparison of studies:

Santos concentrated on simplifying the selection in the rice subcategory for an online grocery store. To do this, he used transactional logs to integrate SKU stock-out levels and figure out how many different SKUs would maximise sales. On the other hand, Shaikh et al. investigated the importance of big data analytics in Pakistan's retail sector, evaluating the function of data scientists and sophisticated algorithms in assortment strategy optimisation through quota sampling and interviews.

Karki's research attempted to classify products according to customer behaviour in order to create a hybrid strategy for assortment management. The study used association rule mining and unsupervised clustering to improve assortment tactics and discover the best consumer segments by utilising the UCI Machine Learning Repository dataset. In a similar vein, Tamhankar et al. examined the influence of sales prediction accuracy through clustering, examining the efficacy of clustering algorithms, regressions, and classification models in assortment planning using a dataset from a national store.

Kanagaraj et al., on the other hand, studied sales and distribution analytics in specialty retail industries. They did this by using DAX Query Language and business analytics tools to create personalised marketing strategies and enable data-driven sales success. The precise dataset used for this investigation was not made public.

III. Research Methodology

The research hypothesis is that particular links between product purchases can be identified by employing Python for market basket analysis and conducting a survey catered to supermarket customers, thereby determining the best assortment tactics to boost sales and consumer satisfaction. With the help of a pipeline this research methodology was conducted effectively.

Phase 1:Initial Research and Setup

- Examine the literature that is currently available on market basket analysis, supermarket assortment methods, and how these affect sales and consumer satisfaction.
- Find studies, utilising Python for market basket analysis.
- Configuring the system

Phase 2:Data Collection

- Gather the information needed to respond to the suggested research questions from the datasets.
- To learn more about existing assortment strategies and issues, conduct structured interviews with supermarket owners or managers. Alternatively, survey customers to learn more about market basket analysis.

Phase 3:Experimentation:

- Apply market basket analysis algorithms using Python.
- Discover association rules to uncover patterns in customer purchasing behaviour.

Phase 4:Analysis and Findings

Market Basket Analysis:

- Analyse the results obtained from the market basket analysis algorithms used.

Interview and survey analysis:

- Summarize key findings from the interviews with supermarket stakeholders, or from the survey conducted
- Understand current assortment methods, customer preferences, and areas for improvement.

Phase 5:Discussion

- Analyse the results in light of the hypothesis and goal of the study.
- Address the possible methods that supermarket selection strategies can be improved and informed by market basket analysis.
- Discuss how the findings might affect raising sales and enhancing customer satisfaction.
- Determine the shortcomings and enhancements
- Summarise the contributions and findings of the research.

Research Methods proposed:

[6]The First Approach:

Supermarkets utilise market basket analysis as a key tool to find hidden trends in the purchase behaviour of their customers. Supermarkets may optimise their product assortments, customise marketing campaigns, and improve the entire shopping experience for customers by recognising items that are frequently purchased together. This research study explores the use of the Apriori algorithm, a technique frequently used for market basket analysis, and discusses how metrics like lift, support, and confidence are important for understanding customer behaviour and helping businesses make decisions.

Confidence

- In market basket analysis, confidence is essential as it offers information about the probability of customers buying a certain item based on their previous purchases.
- The Apriori technique is used in this analysis to calculate confidence, which allows us to identify client preferences

and purchase patterns by determining the strength of correlations between product pairs.
- Supermarkets may increase sales and consumer satisfaction by creating customised recommendations, optimising product placements, and creating focused marketing efforts based on confidence levels.

Lift

- Lift is a crucial statistic in market basket analysis that assesses the importance of items in boosting sales and gauges the degree of link between them.
- By using lift analysis, we can determine which product combinations are more likely to be bought together than would be predicted by chance. This helps to discover areas where cross-selling and promotional activities might be used.
- Supermarkets may improve the overall shopping experience for customers, optimise their product assortments, and increase the effectiveness of their promotions by leveraging the insights gleaned by lift analysis.

Support

- Support measures how frequently a specific item or group of items appears in transactions, offering important information about how well-liked they are among clients.
- We can determine frequently purchased item combinations by analysing support values, which helps us make informed judgements about product selection and inventory management.
- Supermarkets may boost customer satisfaction, generate loyalty, and boost income by making sure that popular item combinations are easily accessible to customers.

Supermarkets may get useful insights into the behaviour and preferences of their customers by using the Apriori algorithm to analyse measures like confidence, lift, and support. These findings provide the basis for creating focused marketing campaigns, optimizing product assortment strategies, and eventually driving the business to success in the competitive marketplace of today.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(bottled water)	(UHT-milk)	0.060683	0.021386	0.001069	0.017621	0.823954	-0.000228	0.996168
1	(UHT-milk)	(bottled water)	0.021386	0.060683	0.001069	0.050000	0.823954	-0.000228	0.988755
2	(UHT-milk)	(other vegetables)	0.021386	0.122101	0.002139	0.100000	0.818993	-0.000473	0.975443
3	(other vegetables)	(UHT-milk)	0.122101	0.021386	0.002139	0.017515	0.818993	-0.000473	0.996060
4	(sausage)	(UHT-milk)	0.060349	0.021386	0.001136	0.018826	0.880298	-0.000154	0.997391

Fig. 1. Using the imported Apriori algorithm

[7]The Second approach:

Makes use of the Apriori Algorithm, incorporates a wide range of analytical tools in addition to just using algorithms. Standard and customised metrics, association rules, aggregation and trimming strategies, and advanced visualisation approaches are a few of these tools. Especially:

- Zhang's metric is noteworthy for its thorough construction based on support criteria, making it a comprehensive and easily logical measure.
- When trying to understand the complex interactions between a big number of rules and a relatively small number of antecedents and consequents, heat maps are an important tool.
- Throughout the rulebook, scatterplots prove to be invaluable instruments for evaluating broad trends and behavioural patterns. They give crucial insights into the underlying structure and dynamics of the dataset by pointing out natural thresholds in the data and offering a comprehensive picture of the complete dataset.

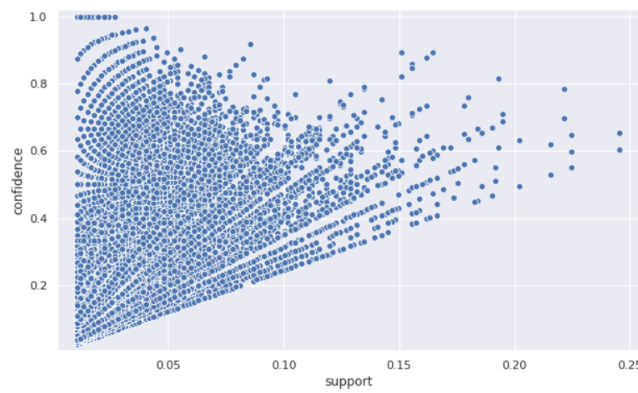

Fig. 2. Scatterplots: An example of one visualisation tools used

- Different association rules including: Multi-antecedent, multi-consequent and multi- antecedent and consequent.
- Association and dissociation following with Zhang's metric.
- Advanced aggregation and filtering with advanced Apriori pruning.
- Confidence, Lift and Support metrics

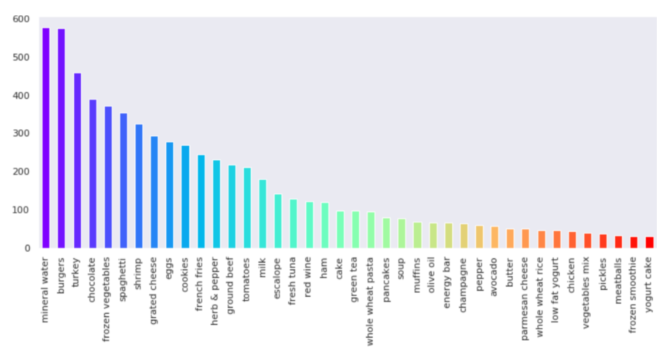

Fig. 3. Frequency of most popular items

Alternative research conducted:

Apart from the python Apriori algorithms used, I have conducted a survey with 67 participants, in order to know more about customer preferences and behaviour, which would shed some light on the market basket analysis being conducted.

Resources used: In this research Python was the primary tool, to carry out market basket analysis and the datasets used were provided from online tutorials. In addition to Python, Google forms was used to conduct the survey. The data was analysed using an HP laptop that had Visual Studio Code and the required extensions installed. The utilisation of libraries like pandas and NumPy yielded significant findings regarding purchasing patterns and correlations, which in turn facilitated the enhancement of product assortment methods in retail settings.

IV. Findings & Discussion of Results

In regards to the ***first research method*** we sorted the dataset according to metrics for lift, confidence, and support in order to perform a thorough analysis and determine the most frequent item combinations in the dataset.

The following table reveals the top four product combinations that customers most usually purchase together:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
622	(rolls/buns)	(whole milk)	0.110005	0.157923	0.013968	0.126974	0.804028	-0.003404	0.964550
623	(whole milk)	(rolls/buns)	0.157923	0.110005	0.013968	0.088447	0.804028	-0.003404	0.976350
694	(yogurt)	(whole milk)	0.085879	0.157923	0.011161	0.129961	0.822940	-0.002401	0.967861
695	(whole milk)	(yogurt)	0.157923	0.085879	0.011161	0.070673	0.822940	-0.002401	0.983638
550	(soda)	(other vegetables)	0.097106	0.122101	0.009691	0.099794	0.817302	-0.002166	0.975219
551	(other vegetables)	(soda)	0.122101	0.097106	0.009691	0.079365	0.817302	-0.002166	0.980729
648	(sausage)	(whole milk)	0.060349	0.157923	0.008955	0.148394	0.939663	-0.000575	0.988811
649	(whole milk)	(sausage)	0.157923	0.060349	0.008955	0.056708	0.939663	-0.000575	0.996140

Fig. 4. A sorted table of the most frequent item combinations in the entire dataset

- Rolls and milk
- Yoghurt and milk
- Sausages and milk
- Soda and Vegetables

This might be the result of a promotion that the grocery shop ran on certain products together or because they were arranged in a line of sight to increase sales.

Second research method: Since this research involved two different data sets we are left with two different results. For the first dataset, each line of code was refined on order to uncover more detailed data, by first computing the support, then refining the support with confidence and a further refinement with lift. Conviction was calculated suggesting that the rule if burgers then french fries is supported. After computing association and dissociation, once again the association rule if burgers then french fries proved consistent.It had a positive value for Zhang's metric, indicating that the two food items are not dissociated.Moreover, after applying Zhang's metric the following results emerged.

	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhang
count	14520.000000	14520.000000	14520.000000	14520.000000	14520.000000	14520.000000	1.440000e+04	14400.000000
mean	0.040611	0.040611	0.001906	0.052663	1.467719	0.000335	inf	-0.011728
std	0.097141	0.097141	0.007505	0.108745	1.864950	0.001148	NaN	0.621009
min	0.000133	0.000133	0.000000	0.000000	0.000000	-0.011697	7.616318e-01	-1.000000
25%	0.007732	0.007732	0.000133	0.004975	0.500009	-0.000046	9.953340e-01	-0.517778
50%	0.015731	0.015731	0.000400	0.021849	1.214494	0.000079	1.003948e+00	0.192710
75%	0.042528	0.042528	0.001333	0.058140	1.858384	0.000361	1.020828e+00	0.483074
max	1.000000	1.000000	0.238368	1.000000	45.460606	0.022088	inf	1.000000

Fig. 5. Zhang's metric results

From the table above we notice that most items were dissociated, which suggests that they would have been a poor choice to pair together for promotional purposes.

In regards to the *second dataset used*, after performing aggregation for the candles, bags and boxes the function was written to simplify the task. This outputs a DataFrame that indicates whether each transaction includes items from that category.

TABLE I
Table of Results

Share of Bags	0.41
Share of Boxes	0.39
Share of Candles	0.11

The aggregation exercise that was performed for the online retailer proved helpful. It offered a starting point for understanding which categories of items appear frequently in transactions. The retailer wanted to explore the individual items themselves to find out which are frequent.Furthermore the Apriori algorithm was applied followed by pruning with lift and confidence. Lastly Zhang's rule was used yet again to select those rules with a high and positive association.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric	zhang
25	(box, sign)	(candle)	0.126159	0.112010	0.040309	0.319510	2.852525	0.026178	1.304928	0.743194	0.743194
26	(candle, sign)	(box)	0.046254	0.385731	0.040309	0.871465	2.259255	0.022467	4.779011	0.584408	0.584408
27	(box)	(candle, sign)	0.385731	0.046254	0.040309	0.104501	2.259255	0.022467	1.065043	0.907382	0.907382
28	(candle)	(box, sign)	0.112010	0.126159	0.040309	0.359873	2.852525	0.026178	1.365104	0.731352	0.731352
29	(sign)	(box, candle)	0.202021	0.082878	0.040309	0.199529	2.407518	0.023566	1.145729	0.732644	0.732644

Fig. 6. Final Results

Survey Results:

Based on the results of the survey, we notice that most people go to the supermarket once a week and the main reason for their visit is to buy groceries. Even though some customers are clearly brand loyal, there is a still an inclination to explore, as evidenced by the 55.2% of consumers who say they are very interested in finding new or unique products when they go shopping. This shows the critical role that supermarket assortment strategies and organisation have. More than half of the participants state that they are "extremely important".

When product correlations are examined, cleaning supplies and toiletries come out on top, closely followed by fruits and vegetables (61.2%) and snacks, and confectionery items (55.2%). Customers' inclination to spend more than €100 per visit suggests that they like to buy in bulk, which is consistent with the frequency of weekly shopping that has been noted. Notably, special offers and promotions have a significant impact, increasing shopping activity for 50.7% of respondents.

Customers still place a high value on quality, as shown by the 50.7% of them who occasionally choose particular brands. Additionally, a resounding majority (88.1%) attests to the critical influence that product placement and display had on their decision to buy, emphasising the necessity for supermarkets to develop efficient product assortment strategies.

Although strategically placed products at the checkout may encourage impulsive purchases, only 9.1% of consumers regularly give in to these temptations, suggesting that this strategy has limited effectiveness. On the other hand, a considerable percentage (55.2%) of participants find it appealing when speciality items are assigned to certain sections or displays, however this is not always the case.

Additionally, 53% of respondents indicated that they preferred a consistent product variety for the entire year, highlighting the need of keeping a steady inventory. Finally, approximately 30% of participants believe that having a large selection of produce is essential, highlighting the critical role that assortment techniques have in satisfying the various demands and preferences of consumers.

V. CONCLUSION

The study's conclusions emphasise how critical it is to comprehend customer behaviour and optimise supermarket selection tactics. Significant patterns in consumer purchase behaviour were found by sorting the dataset according to measures for lift, confidence, and support. These patterns offered possible approaches for marketing related products and optimising product placements. Furthermore, product connections and dissociations were revealed by applying the Apriori algorithm and sophisticated analytical tools, which presented chances to improve advertising campaigns and product assortments.

The survey findings emphasised how important it is for promotional efforts, clever product placement, and high-quality products to have an impact on consumers' purchase decisions. Furthermore, desires for a wide range of produce options and a consistent supply of new products highlighted how crucial efficient assortment strategies are to satisfying customer needs.

In conclusion, these results demonstrate the importance of market basket research and data-driven strategies in raising customer happiness and boosting sales in supermarkets, which in turn improves company performance and competitiveness in the retail sector.

APPENDIX A
SUPPORTING MATERIAL

Survey Questions

Market Basket Analysis Research

I am conducting research on Market Basket Analysis in order to correlate how consumer behaviour in shopping affects product assortment strategies in supermarkets.

The data collected can assist supermarkets in optimizing their offerings, improving customer experiences, and driving customer loyalty.

Your feedback is essential in helping me better understand how customers like you make purchasing decisions and navigate their shopping experiences.

Age Group

- Under 25
- 25-34
- 35-44
- 45-54
- 54+

Gender

- Male
- Female

Marital or Parental Status

- Single
- Married
- Married with children
- Single with Children

How often do you visit supermarkets in a typical week?

- Once a week
- 2-3 times a week
- Almost daily
- Less than once a week

What are your primary reasons for visiting supermarkets? (select all that apply)

- Grocery Essentials
- Snacks and Beverages
- Household Supplies (e.g., cleaning products , toiletries)
- Meats and Seafood
- Bakery and Confectionery Items

When shopping for groceries, do you prefer specific brands?

- Yes, always
- Yes. sometimes
- No, I choose based on price or promotions

How important is organisation of the supermarket to you?

- 1:Extremely important
- 2
- 3
- 4
- 5:Not important at all

Do you enjoy exploring new or unique products when shopping?

- Yes, always
- Sometimes
- No, I stick to familiar products
-

Which of these product combinations do you often purchase together? (select all that apply)

- Bread and drinks
- Meat and pasta
- Snacks and confectionary items
- Cleaning products and toiletries
- Fruits and vegetables
- Frozen foods and canned foods

How much do you typically spend per supermarket visit?

- Less than €20
- €20-€50
- €50-€100
- Over €100

Do you tend to buy more during promotions or special offers?

- Yes, always
- Yes, sometimes
- No, I buy what I need

What most influences your purchasing decisions?

- Price
- Product quality

- Brand reputation
- Promotions and discounts
- Variety of products offered

Does the way products are organised and displayed in a supermarket influence your purchasing decisions?

- Yes, it helps me find what I need easily
- No, I don not pay attention to product displays

How often do you buy items that are located at the counter?

- Always
- Sometimes
- Rarely
- Never
-

How likely are you to explore speciality sections or displays(e.g., organic, international foods) within a supermarket?

- 1:Very Likely
- 2
- 3:Not Likely

Are you more likely to shop at a supermarket that offers seasonal and speciality products?

- Yes, it enhances my shopping experience
- No, I prefer a consistent selection of products year-round

Which product categories do you think are most important for a supermarket to offer a wide assortment of?

- Produce
- Dairy and refrigerated items
- Snacks and beverages
- Meat and seafood
- Personal care and household essentials
- Other:

REFERENCES

[1] H. B. O. Santos, "Assortment optimization in online retail," 2019.

[2] S. Shaikh, F. Sultan, and M. Asim, "Assortment on the bases of big-data analytics: A quantitative analysis on retail industry," *JISR management and social sciences & economics*, vol. 19, no. 2, pp. 154–167, 2021.

[3] D. Karki, "A hybrid approach for managing retail assortment by a categorizing products based on consumer behavior. dublin, national college of ireland," Ph.D. dissertation, Ph. D. thesis, 2018.

[4] K. Kanagaraj and R. Venkatesh, "Enhancing business performance: A comprehensive study of sales and distribution analytics in speciality retail sectors," in *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*. IEEE, 2023, pp. 1–5.

[5] R. Tamhankar, S. Khattar, X. Che, S. Zhu, and M. A. Lanham, "Optimal clustering of products for regression-type and classification-type predictive modeling for assortment planning."

[6] N. Selvaraj, "How to perform market basket analysis," https://365datascience.com/tutorials/python-tutorials/market-basket-analysis/, Nov. 2022, accessed: 2024-5-13.

[7] "Market basket analysis in python," https://goldinlocks.github.io/Market-Basket-Analysis-in-Python/, accessed: 2024-5-13.