

Submitted in part fulfilment of the requirements for the degree of

MSc Business Analytics

**Enhancing Online Retail Performance: Cross-Selling Techniques and Consumer Insights in Supermarkets.**

By

Aarya Urankar

6784420

Faculty of Arts and Social Sciences University of Surrey

Word Count - 13456

# **Executive Summary**

In the opening to the dissertation, we emphasized the significance of cross-selling methods in terms of increasing revenue and increasing the value that customers receive from online purchasing. In the introduction, we explain how the COVID-19 pandemic hastened the transition toward online commerce, which has made the efficient utilization of cross-selling strategies more important than it has ever been. The utilization of sophisticated analytics and large amounts of consumer data is the primary focus to comprehend and influence the purchase behaviour of customers. As a means of retaining and increasing consumer value, we highlight the developing problems that are associated with e-commerce and emphasize the necessity of providing personalized shopping experiences. The focus of this study is on online supermarkets in the United Kingdom. The research intends to investigate customer patterns, product preferences, the impact of cross-selling on sales, and the influence of seasonal fluctuations.

The literature review of the dissertation provides a thorough assessment of customer behaviour and cross-selling techniques in e-commerce, highlighting the crucial importance of advanced analytics and technology in contemporary online retail strategies. This review explores the evolution of the Internet as a marketplace and the subsequent digital transformation in marketing. It emphasizes the influence of technologies such as Electronic Data Interchange (EDI) and Electronic Funds Transfer (EFT), as well as the proliferation of different e-commerce models. This text explores the impact of the COVID-19 epidemic, the development of digital marketing, and the increasing significance of mobile commerce and social media marketing. An in-depth analysis of cross-selling methods highlights their pivotal significance in the rise of e-commerce. The review examines the difficulties associated with discovering cross-selling opportunities and evaluates the efficacy of several approaches. This highlights the significance of technology in improving client experiences and customizing cross-selling endeavours study additionally examines customer behaviour and purchase habits. This study explores the intersection of market basket analysis and data analytics, which plays a crucial role in the advancement of cross-selling techniques by comprehending customer buying habits and preferences. Finally, the review highlights a vacuum in the existing literature, specifically pointing out the little research on the impact of developing technology and customer behaviour on cross-selling in the e-commerce industry. This gap provides a chance for future research on the utilization of machine learning algorithms in the online supermarket sector in the UK, intending to improve the efficiency of cross-selling techniques.

The dissertation makes use of Saunders' Research Onion structure as its research methodology. It focuses on the e-commerce industry in the United Kingdom, namely online supermarket stores. Through the use of Market Basket Analysis (MBA) and a large dataset from an online retail company, the purpose of this project is to investigate the behaviour of consumers and the patterns of their purchases. This quantitative study makes use of data mining approaches, employing Python for data analysis, and making use of the Apriori algorithm and RFM Analysis to discover item connections and patterns of customer behaviour. The research entails meticulously preparing and cleaning the data, as well as doing exploratory data analysis. Additionally, it addresses ethical considerations and constraints relating to the extent of generalization and the privacy of the data.

A complete study of the dataset reveals fascinating patterns and trends in customer behavior in the online supermarket sector, which are discussed in the observations section of your dissertation. According to the findings of the study, there are 3,738 distinct items and 15,304 transactions completed by 3,808 different consumers. This indicates that a consistent customer base makes regular repeat purchases. According to the median transaction and price analysis, the majority of the things purchased are of lower prices, with noticeable outliers indicating that the transactions were made in bulk or due to exceptional circumstances.

The most popular products are extremely diversified, reflecting the wide range of customer interests. In terms of seasonal sales trends, the peak occurs towards the end of the year, with noteworthy rises and declines occurring in particular months. The RFM analysis classifies customers into three unique groups according to their purchase history, frequency, and monetary value. This provides a more detailed understanding of the many behaviors that customers engage in. The application of the Apriori algorithm to Market Basket Analysis results in the production of specific association rules, which demonstrates the efficacy of cross-selling methods in the realm of e-commerce.

The analysis and discussion part digs into the complex patterns that have been noticed in the behaviour of customers and the trends that have been observed in the context of the online supermarket industry. The article draws attention to the major seasonal differences that have an effect on consumer purchasing and is closely connected with the traditional seasons in the United Kingdom. In addition, RFM analysis is utilised in the research, which classifies clients into separate clusters, each of which displays distinctive purchasing behavioural tendencies.

In addition, the research highlights the strategic importance that market basket analysis has in determining product pairings and the preferences of customers. The results of this study contribute to the development of targeted marketing strategies, the improvement of the customer experience, and the optimization of inventory management, which eventually results in improved sales and increased customer satisfaction. Transactional data may be transformed into a strategic asset for business growth and customer service optimisation thanks to the insights that were collected from this analysis. These insights are crucial in enabling online supermarkets to shift from reactive to proactive tactics in inventory management and customer interaction.

This dissertation's conclusion summarises the broad investigation of cross-selling techniques in e-commerce, with a particular emphasis on the competitive landscape of the online supermarket industry in the United Kingdom. Through the examination of consumer purchasing patterns, cross-selling effects, and seasonal trends, the research successfully demonstrates the impact that data analytics may have on enhancing customer satisfaction and increasing revenue. There are certain limitations that are acknowledged, such as the high minimum support level for the Apriori technique and the tendency to concentrate on the United Kingdom geographically. A strong emphasis is placed on utilising knowledge regarding customer behaviour for strategic marketing and inventory management in the recommendations. All things considered, the dissertation makes significant contributions to the fields of digital marketing and retail management, offering insights that may be put into practice for the ever-changing landscape of e-commerce industry.

# **Declaration of Originality**

*"*I hereby declare that this thesis has been composed by myself and has not been presented or accepted in any previous application for a degree. The work, of which this is a record, has been carried out by myself unless otherwise stated and where the work is mine, it reflects personal views and values. All quotations have been distinguished by quotation marks and all sources of information have been acknowledged by means of references including those of the Internet. ***I agree that the University has the right to submit my work to the plagiarism detection sources for originality checks.***"

Date 08-01-2024

Aarya Urankar

# **Table of Contents**

Contents

[**Executive Summary** 1](#_Toc155612701)

[**Declaration of Originality** 4](#_Toc155612702)

[**Table of Contents** 5](#_Toc155612703)

[**List of Tables** **Error! Bookmark not defined.**](#_Toc155612704)

[**List of Figures** 7](#_Toc155612705)

[**List Of Abbreviations** 8](#_Toc155612706)

[**1.** **INTRODUCTION** 9](#_Toc155612707)

[**2.** **Literature review** 12](#_Toc155612708)

[**2.1.** **ECommerce and digital marketing evolution** 12](#_Toc155612709)

[**2.2.** **Cross-Selling Strategies in Ecommerce** 15](#_Toc155612710)

[**2.3.** **Customer Behaviour and Purchasing pattern** 17](#_Toc155612711)

[**2.4.** **Data Analytics and Market Basket Analysis** 19](#_Toc155612712)

[**2.5.** **Impact Of Technology on Cross-Selling** 22](#_Toc155612713)

[**3.** **Methodolgy** 26](#_Toc155612714)

[**3.1.** **Research Philosophy and Approach** 26](#_Toc155612715)

[**3.2.** **Research Strategy and choice** 26](#_Toc155612716)

[**3.3.** **Data Collection** 27](#_Toc155612717)

[**3.4.** **Data Analysis** 27](#_Toc155612718)

[**3.5.** **Data Pre-processing** 29](#_Toc155612719)

[**3.6.** **Exploratory Data Analysis** 30](#_Toc155612720)

[**3.7.** **Analysis** 30](#_Toc155612721)

[**3.8.** **Ehical consideration** 31](#_Toc155612722)

[**4.** **Observations** 32](#_Toc155612723)

[**5.** **Analysis and Discussion** 41](#_Toc155612724)

[**5.1.** **Overview of Seasonal Variation In Consumer Markets** 41](#_Toc155612725)

[**5.2.** **Introduction To Consumer Buying Pattern In Online Supermarkets in UK** 43](#_Toc155612726)

[**5.3.** **The Impact Of Bundling: Insights from Complementary Products** 46](#_Toc155612727)

[**5.4.** **Inventory Syncronization: Insights from Market Basket Analysis** 48](#_Toc155612728)

[**6.** **Conclusion** 51](#_Toc155612729)

[**6.1.** **Limitations of the Research** 52](#_Toc155612730)

[**6.2.** **Recommendations** 53](#_Toc155612731)

[**7.** **References** 55](#_Toc155612732)

**List of Tables**

Table 3.4 Dataset Fields…………………………………………………………………….29

Table 4.1 RFM metrics…………….………………………………………………………..37

Table 4.2 Association Rules………………………………………………………………...39

Table 5.3. Apriori Association Rules……………………………………………………….46

# **List of Figures**

Figure 2.1. Ecommerce Growth…………………………………………………………13

Figure 2.2 Cross-Selling in Ecommerce…………………………………………………16

Figure 2.3 TAM Model……………………………………………………………….….17

Figure 2.4.1 Introduction to Market Basket Analysis……………………………………20

Figure 2.4.2 Benefits of MBA……………………………………………………………21

Figure 2.5 Technologies to enhance cross-selling………………………………………...23

Figure 4.1 Count of Unique Values……………………………………………….………32

Figure 4.2 Unit Price Boxplot…………………………………………………………….33

Figure 4.3 Quantity Boxplot……………………………………………………………...33

Figure 4.4 Frequently Purchased Products……………………………………………….34

Figure 4.5 No. of Quantity of Products Sold……………………………………………..35

Figure 4.6 Sales Trend……………………………………………………………………36

Figure 4.7 Elbow curve for K values……………………………………………………..37

Figure 4.8 Number of Data points in each cluster………………………………………..38

Figure 5.1 Seasonal Trend………………………………………………………………..43

Figure 5.2 Clusters of Customers…………………………………………………………44

# **List Of Abbreviations**

Electronic data interchange (EDI)

Electronic funds transfer (EFT)

Business-to-business (B2B)

Business-to-consumer (B2C)

Consumer-to-consumer (C2C)

Goverment-to-business/ Consumer (G2B/C)

Recurrent neural networks (RNNs)

Long Short-Term Memory (LSTM)

Artificial neural networks (ANNs)

Deep learning (DL)

Technology Acceptance Model (TAM)

Theory of Reasoned Action (TRA)

Word-of-mouth (e-WOM)

Online consumer buying behavior (OCPB)

Market basket analysis (MBA)

Customer relationship management (CRM)

Point-of-sale (POS)

Artificial intelligence (AI)

Augmented reality (AR)

Virtual reality (VR)

# **INTRODUCTION**

When it comes to e-commerce, cross-selling is a powerful strategy that can be used to increase the value of the client and generate larger revenues. The global internet sector has been steadily expanding for the past decade as an increasing number of businesses and customers have begun to explore the world of online shopping (Gupta et al. 2023). Following the consumer's selection of a product to purchase (that is, placing an item in the shopping cart and being prepared to check out), the company makes recommendations for additional products to the consumer based on the information it has gathered about the consumer's tastes from prior purchases as well as the items that are already in the shopping cart. Following the purchase of "The Lego Batman Movie" from Amazon.com, for example, a buyer might be recommended to purchase a "Toddler backpack" from Orezi, which is a third-party retailer (Abhijeet Ghoshal, 2021). Advertising and marketing new products or services to existing customers based on their purchase history and preferences is what it comprises. A greater number of people have been forced to do their business online as a result of the COVID-19 outbreak. The use of creative cross-selling strategies has the potential to significantly increase client satisfaction, loyalty, and revenue in general (Mim and Ferdous, 2021). The pandemic caused by COVID-19 has resulted in a considerable increase in online sales, which is one of the most significant positive impacts.

In this day and age of digitalization and e-commerce, the greatest difficulty that sellers on the internet have been to able to maximize their revenue while simultaneously boosting the value that they provide to their customers (UNCTAD, 2021). The primary focus of the dissertation is on how e-commerce businesses can make use of sophisticated analytical methods and vast amounts of consumer data to foresee and influence the purchasing behavior of customers. With the development of big data technology, e-commerce platforms now have significant insights into the behavior, choices, and buying habits of their customers. These insights may be used to better serve their customers. Firms can utilize complicated algorithms to deliver individualized product recommendations, which is a crucial component of good cross-selling efforts. This is made possible by the wealth of data that firms have readily available.

In order for businesses to achieve their goals of increasing consumer satisfaction and revenue, it is essential for them to implement cutting-edge cross-selling methods in their online commercial transactions. Businesses that engage in e-commerce have the ability to create a browsing experience that is both personalized and engaging, which simply delights customers. An online retailer recently introduced a system for cross-selling based on the recommendations of McKinsey, which resulted in a 20% boost in sales (Abhijeet Ghoshal, 2021). The United Kingdom accounts for the third-largest e-commerce market in the world when it comes to online sales. In 2021, the income generated by online commerce in the United Kingdom topped 120 billion dollars. By 2024, it is anticipated that e-commerce will account for 37.5% of the retail industry in the United Kingdom. Amazon was the dominant player in the internet business in the United Kingdom in 2021, with sales of $17.1 billion (Adebisi, 2023).

In order for businesses to maintain a competitive advantage in the e-commerce sector, they need to maintain a flexible approach to their business strategy, utilizing both innovation and knowledge about their customers. Through the implementation of cutting-edge cross-marketing tactics, businesses that engage in e-commerce provide customers with an additional, more personalized and satisfying shopping experience (Falcone, Kent and Fugate, 2019). This is leading to increased levels of customer satisfaction and loyalty, as well as an increase in sales. In the quickly changing environment of e-commerce, businesses are confronted with the challenge of not only gaining new customers but also retaining existing ones (Katsikeas, Leonidou and Zeriti, 2019). However, the most important thing is to maintain and increase the value of the ones that are already there.

Businesses that successfully adopt superior cross-selling strategies gain the benefits of an aggressive advantage by providing customers with a more engaging and personalized shopping experience (Sheth, Jain and Ambika, 2023). Cross-selling has the potential to boost customer loyalty by providing additional value and more comprehensively addressing the demands of the consumer (Becker et al., 2019). A consumer who buys an electronic gadget, for instance, is grateful for a cross-sell proposition that includes the suggestion of purchasing an information drive or an accessory bag (Myasnikov and Starov, 2022).

The objective of this research is to increase our understanding of the strategies that internet retailers employ in order to engage in cross-selling. Among the primary objectives are the following: to analyse consumer purchasing trends; to determine which products customers frequently purchase together in order to ascertain their preferences; to evaluate the impact of cross-selling on sales; to investigate the application of analysis in supply chain optimization; and to investigate how seasonal trends influence consumer behaviour.

The research questions dissertation aims to answer are-

1. How do seasonal variations and market trends influence customer purchasing behavior?
2. What are the key patterns in customer purchasing behavior in an online supermarket context?
3. Which products are most frequently purchased together, and what does this suggest about customer preferences and needs, and its implication on sales?
4. Can market basket analysis inform inventory management and supply chain optimization in online retail?

The scope of this study is centered on online supermarkets, an industry sector that comprises a considerable portion of e-commerce and is distinguished by specific customer purchasing habits and operational challenges. Specifically, the analysis focuses on the demography of the United Kingdom. One of the most important aspects of this study will be the examination of market basket data and customer segmentation, which will shed light on the preferences of consumers and how those choices modify themselves in reaction to changes in the market. This emphasis ensures that the results are both extensive enough to contribute to a wealth of expertise in digital marketing and retail management and specific enough to offer in-depth insights into a crucial sector of e-commerce.

# **Literature review**

Within the context of e-commerce, this literature review takes a critical look at customer behavior and cross-selling methods. It also highlights the role that advanced analytics and technology play in influencing contemporary online retail practices.

### **ECommerce and digital marketing evolution**

During the 1990s, the internet began to develop into a marketplace, which marked a significant turning point in the history of electronic commerce (Costa and Castro, 2021). The beginnings of electronic commerce may be traced back to the late 1970s, when significant breakthroughs such as electronic data interchange (EDI) and electronic funds transfer (EFT) were introduced (Raimundo, 2021). Through the facilitation of electronic transaction execution and the exchange of information between businesses, these technologies lay the groundwork for future initiatives in the realm of electronic commerce. This shift involves the widespread use of and research into mobile services, e-commerce, and digital advertisements, which will revolutionize the way that goods and services are offered to customers (Marina Basimakopoulou, 2022). The term "the Digital Transformation" refers to the significant change that has occurred in marketing over the course of the past twenty years.

The emergence of the internet led to the development of a more sophisticated platform for electronic commerce, which merged the sharing of data with activities that were both financial and non-financial in nature. The implementation of various business-to-business (B2B), business-to-consumer (B2C), consumer-to-consumer (C2C), and Goverment-to-business/ Consumer(G2B/C) models increased the scope of e-commerce and eliminated regional limits, hence fostering worldwide trade and economic growth (Raimundo, 2021). There has been a tremendous revolution in marketing research, monitoring, planning, and management as a result of the introduction of internet analytics software, which indicates a new degree of engagement between businesses (Erdmann, 2021). Innovations in technology have resulted in the growth and development of e-commerce, which has resulted in a significant shift in marketing strategies and the behavior of consumers (Murdiana and Hajaoui, 2020).

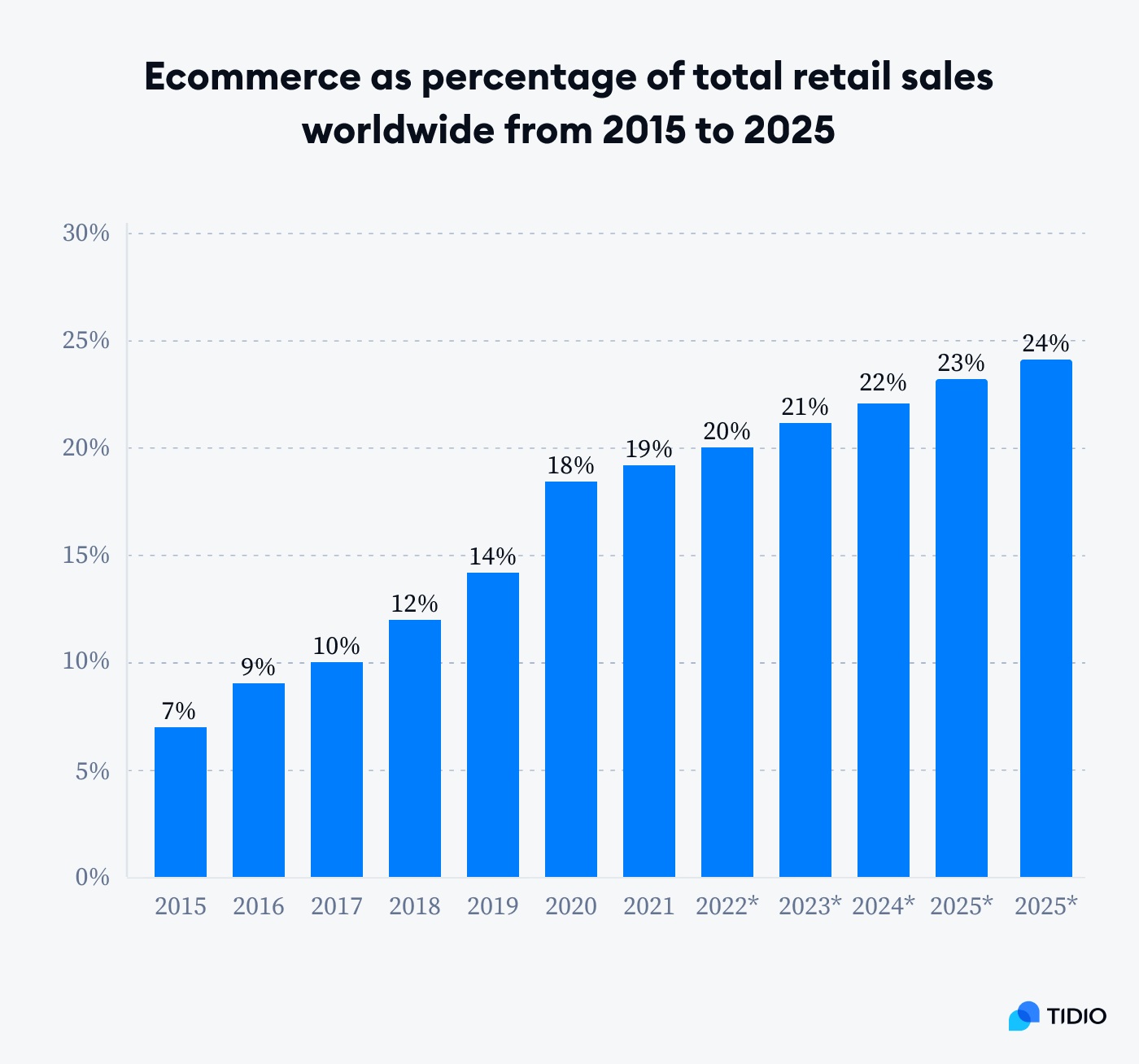


Figure 2.1. Ecommerce Growth (Fokina, 2023).

Both consumer and business behaviours were altered as a result of the COVID-19 epidemic, which significantly accelerated the development of e-commerce. Figure 2.1 Illustrates percentage growth in ecommerce industry. This development in e-commerce can be attributed to the restrictions placed on traditional retail outlets as well as the shifting preferences of consumers towards online shopping due to the convenience and security it offers. The significance of social media and mobile marketing has been brought to light as a result of this transition (Andrew N. Mason, 2021). This is because more than half of the world's population uses social media on a regular basis. Improvements in digital marketing techniques were also encouraged as a result of the pandemic, which occurred at the same time that businesses were adapting to the growing digital competition and the shifting expectations of customers (Erdmann,2021).

The number of people who are conducting transactions online is rapidly increasing, this is due to the fact that there are approximately 5 billion internet users present around the world (Khrais and Alghamdi, 2021). By gaining an understanding of their customers' requirements, supplying them with suitable products, constructing product bundles, and utilizing suggestive marketing, businesses can effectively offer their products to customers. M-commerce, also known as mobile commerce, is a crucial factor in the rise of sales made using mobile devices such as iPads and cell phones (Kasthuri Subaramaniam and Raenu Kolandaisamy, 2019). Mobile marketing has emerged as an essential digital marketing channel because it gives customers access to a large range of products and services that are both quick and economical, as well as simple to use. Additionally, users have the added benefit of being able to perform transactions nearly anywhere and at any time (Marina Basimakopoulou, 2022). Subscribers are able to obtain the things, which provide stores with a continuous source of revenue while also providing customers with the advantage of periodic replenishment.

Marketing through social media has radically changed the way in which customers and producers of goods and services connect with one another. In this day and age, social media websites have become the primary source of information on commercial products and services, particularly among younger audiences (Dubbelink Sanne Ichelle, 2021). It is becoming increasingly important for firms to incorporate a digital marketing strategy into their overall marketing plan to generate outcomes that are both profitable and sustainable over the long run (Simon Malesev, 2021).

A technique that combines digital and physical outlets to provide a seamless purchasing experience is becoming increasingly popular among businesses. This approach is known as multichannel marketing. It is anticipated that online transactions will continue to increase and eventually take the lion's share of the retail pie (Wewege and Thomsett 2019). E-commerce is projected to account for 21.8% of total retail sales by the year 2024, with global retail sales for e-commerce likely to surpass 6.388 trillion dollars by that year (Coppola, 2022).

### **Cross-Selling Strategies in Ecommerce**

The manner in which e-commerce is expanding has brought to light the significance of sales strategies that involve cross-selling. Cross-selling is a sales strategy in which a provider pushes customers to purchase additional services or products that are complementary to the transaction that they have already made (Abhijeet Ghoshal, 2021). Not only will this result in an increase in the average transaction price, but it will also improve the overall experience for customers. These technologies enable businesses to lower the amount of time that customers are required to spend searching while simultaneously increasing revenue (Abhijeet Ghoshal, 2021). This is accomplished by including a greater number of items in the purchasing process. Offering bundles or packages include not only genuine goods but also other items that are related to them. The purchaser receives value from this, and they are encouraged to buy further products that complement their purchase (Zhang and Bockstedt, 2020). A shift in perspective by emphasizing digital marketing strategies and strategic goals that are crucial to the sector of international online commerce (Sjoukje PK Goldman, 2020).

Discovering potential cross-sells might be challenging, yet, it is an essential component of expanding the sales assortment. Companies can uncover cross-selling possibilities and enhance income and client retention by studying their customers' past purchases, educating the sales personnel, evaluating customer feedback, providing package discounts, and employing technology (Dutta and Bhattacharya, 2019). As argued by Eisenreich et al., (2021), The practice of cross-selling has shown to be a game-changing strategy for successful expansion in a range of industries. It is necessary to provide consumers with cross-selling options throughout the entirety of the checkout process. This is an opportune moment to suggest additional objects associated with what the consumer is already purchasing (Siddharth et al., 2022).  Cross-selling is becoming more than just an add-on strategy, according to the findings, which point to a developing trend in which it is becoming more than just an add-on tactic. Instead, it is a significant component of digital marketing, which calls for a deeper knowledge of its workings and outcomes in both developed and emerging countries (Sjoukje PK Goldman, 2020). A customer base that is segmented according to potential, demographics, or shopping activity, and then adjusting pass-promoting strategies to each section to make the offer more relevant and enticing to the target audience.



Figure 2.2 Cross-Selling in Ecommerce (FasterCapital, n.d.).

The practice of cross-selling is a crucial strategy that internet retailers implement in order to increase their revenue and produce more money (Xu et al., 2022). Businesses must carry out effective cross-selling strategies to get the desired desired results. As criticised by Kim and Rao, (2023), personalization, complimentary items, cross-selling, and after-buying cross-selling are a few of the most effective cross-selling methods for organizations.  Electronic mail marketing makes it possible to inform past clients about linked items (Gomez et al., 2020). Through the use of cross-selling, businesses have the ability to increase their profits, strengthen their customer loyalty, increase the value of their customers over time, and increase their profitability. When businesses are aware of the benefits of cross-selling, they are better able to devise effective strategies for implementing it and gaining advantages (Bhalla, 2021). Figure 2.2 gives overview of Cross-Selling.

Both model-based and heuristic-based techniques have been the subject of a significant amount of research that has been conducted on the use of the Recommender System (Javed et al., 2021). The utilisation of recurrent neural networks (RNNs) has resulted in the transformation of these systems. The designs of these networks include Long Short-Term Memory (LSTM), artificial neural networks (ANNs), and deep learning (DL) (Şahin, 2022). These cutting-edge tactics have been shown to be effective in increasing the number of recommendations and generating opportunities for cross-selling.

### **Customer Behaviour and Purchasing pattern**

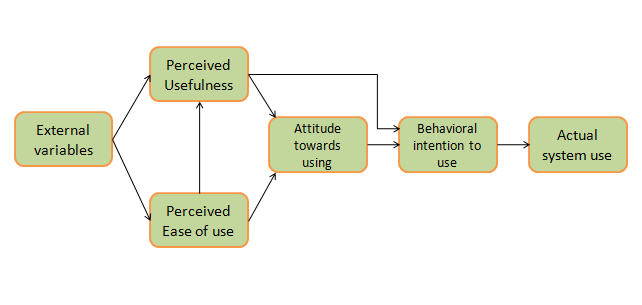
When it comes to the world of internet shopping, particularly for commodities, things are moving very quickly. The development of new retail channels, such as home delivery, click and collect, and combinations of the two has increased the variety of preferences held by consumers (Christian Brand, 2020). Conventional models of customer behaviour, such as the Theory of Planned Behaviour and the Technology Acceptance Model, illustrate how crucial it is for customers to have confidence in their purchasing decisions. Figure 2.3 expalin the TAM model. These models also highlight the significance of user-friendly interfaces, social evidence in the form of opinions, and high-quality consumer studies in influencing customer behaviour (Wewege and Thomsett 2019). In the context of the virtual market, it permits the study of aspects such as confidence in the online environment, perceived usefulness, and simplicity of use. It has been established that social influence, which has its origins in the Theory of Reasoned Action (TRA), has a significant impact on the kinds of behaviours that customers intend to engage in. There are other sources of this impact, such as the media and personal connections (Sana Sajid. Rao Muhammad Rashid, 2022).

Figure 2.3 TAM Model (Park and Park, 2020).

The pandemic caused by COVID-19 brought about substantial changes in consumer behaviour, in particular a shift towards contactless services and online shopping (Gu et al., 2021). The significance of elements including product cost and quality, customer inclinations, website layout, safety, and electronic word-of-mouth (e-WOM) in shaping online consumer buying behavior (OCPB) (Wang Q, 2023). The use of big data in online purchasing has become an effective method for identifying and satisfying the preferences of customers, particularly in the field of e-word-of-mouth data mining (Wang Q, 2023). This has contributed to the practice of purchasing groceries from supermarkets online, which is fast rising. The rise is indicative of shifting consumer habits, such as making healthier decisions by scheduling purchases in advance. On the other hand, there are problems with the quality and condition of perishable goods that are acquired online (Christian Brand, 2020). Understanding this aspect of consumer behavior is vital for gaining an understanding of purchasing patterns and developing effective cross-selling strategies in the realm of e-commerce.

The wide range of elements that influence online buying behavior, such as technology, human psychology, and cultural, demographic, and economic considerations, these variables include the way that customers view risk, their attitudes towards online buying, and how much they like browsing and making purchases online (Wang Q, 2023).  To build efficient strategies that improve customer prices and drive revenue growth in e-commerce, it is essential to have a solid understanding of customer behavior and purchasing habits within the context of pass-selling (Coppola, 2022). The impact of cross-selling on customer retention is common, even though it is generally related to rising revenue through additional income.

It was shown that even in circumstances when a purchase was not planned, browsing activity was affected by interest about the brands and products that were offered online (Sana Sajid. Rao Muhammad Rashid, 2022). When carried out effectively, cross-selling makes a significant contribution to an improved standard purchasing experience and encourages customers to investigate a significantly larger range of products (Wewege and Thomsett 2019). The act of shopping online is influenced by many primary factors, including cultural, demographic, economic, technological, and psychological aspects of an individual.

### **Data Analytics and Market Basket Analysis**

A significant contributor to the development of cross-selling tactics in online sales is the convergence of market basket analysis and data analytics (Omar et al., 2023). This convergence has resulted in enhanced sales growth and improved customer value thanks to the advancement of these methods. One of the most important aspects of data analytics is market basket analysis, which makes it possible to comprehend the purchasing patterns of various consumers (Sinha, 2021). It involves taking a look at the bundles and combinations of products that customers generally buy together.  This technique examines data on baskets to discover pertinent information regarding the consumers' intentions to make purchases (Ural Gokay ciçekli, 2021). Figure 2.4.1 shows Makrket basket analysis introduction, it analysis is a kind of data analysis that is utilised to determine the connections between the products that customers have a tendency to purchase along with one another. When businesses use market basket analysis and other data-driven techniques, they are able to more effectively target and personalise their promotional activities. This not only increases customer satisfaction and sales, but it also has a significant impact on understanding and satisfying the varied requirements of customers, which further accelerates the growth of e-commerce (Efrat, 2020). By determining the categories that are connected to one another, the study is able to gain insights about the purchasing intents of the customers. The approach is predicated on the idea that consumers' decisions to buy products from different categories may not be autonomous and instead exhibit comparable trends (Ural Gökay çiçekli, 2021). Market basket research and customer loyalty programmes are two components that should be included in any comprehensive plan for increasing customer value. Companies are able to adapt their cross-selling strategies with the assistance of this integration, which identifies the products that are likely to be ordered into a department. (Ahmad M. A. Zamila, 2020) This technique strengthens relationships with customers and increases sales, which ultimately results in the development of long-term loyalty with those customers. Through the utilization of statistical analysis, businesses have the potential to enhance the satisfaction of their customers and raise their profits. This may be accomplished by personalizing the shopping experience, recommending products that are suitable, and offering promotions.

A substantial boost in the application of market basket evaluations was brought about by the advent of digital point-of-sale (POS) systems (Sjarif et al., 2021). The process of gathering transactional details from the e-trade platform involves collecting statistics on the products that are acquired in every transaction (Addagarla and Amalanathan 2020). In light of the regulations governing affiliations, it is essential to identify chances for cross-promotion to have an understanding of the items that are frequently purchased together (Kumar Panda et al. 2020). When compared to the personal records that are kept by business owners, the digital documents that are created by point-of-sale systems make it much simpler for applications to manage and evaluate massive amounts of financial data. In order to persuade customers to acquire the package, it is possible to provide them with discounts or special offers for products that are related to the package.

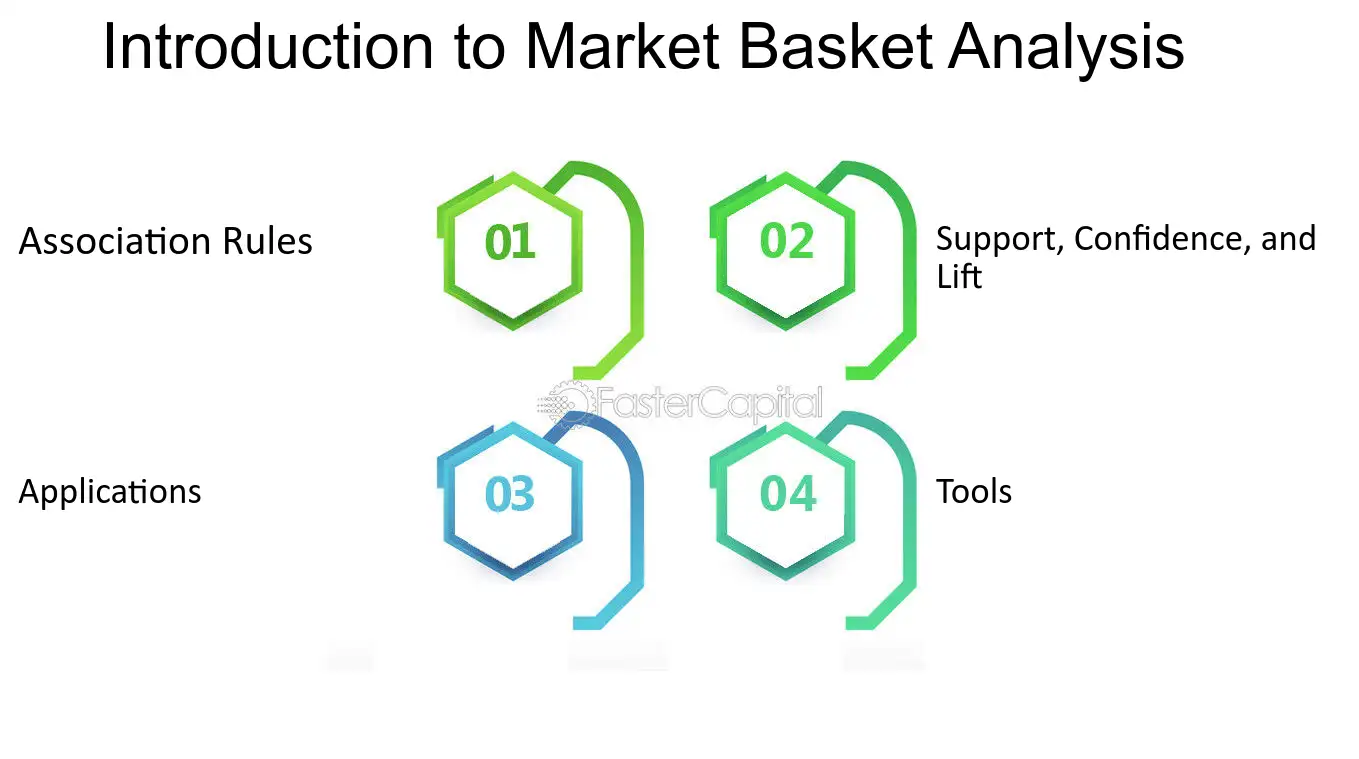


Figure 2.4.1 Introduction to Market Basket Analysis, (Faster Capital, 2023)

The Apriori set of rules provides a robust association rule that applies to an online marketplace that sells e-trade goods (Ünvan, 2020). This rule states that "If a customer purchases an excessive-give-up smartphone (X), there is a high likelihood that they will also purchase high-quality headphones (Y) and a protective case (Z)." The likelihood that customers who buy item X will also buy item Y is referred to as potential customers. When the e-commerce platform is armed with this information, it is able to strategically bundle such items, provide discounts on related merchandise, or present focused promotions to customers who have purchased the excessively give-up smartphone. It determines the extent to which the occurrence of item Y is dependent on the occurrence of item X, hence indicating the strength of the connection between the two. According to Ural Gokay Cicekli's research (2021), the antecedent (X) and the consequent (Y) in this representation reflect two different groups of things.

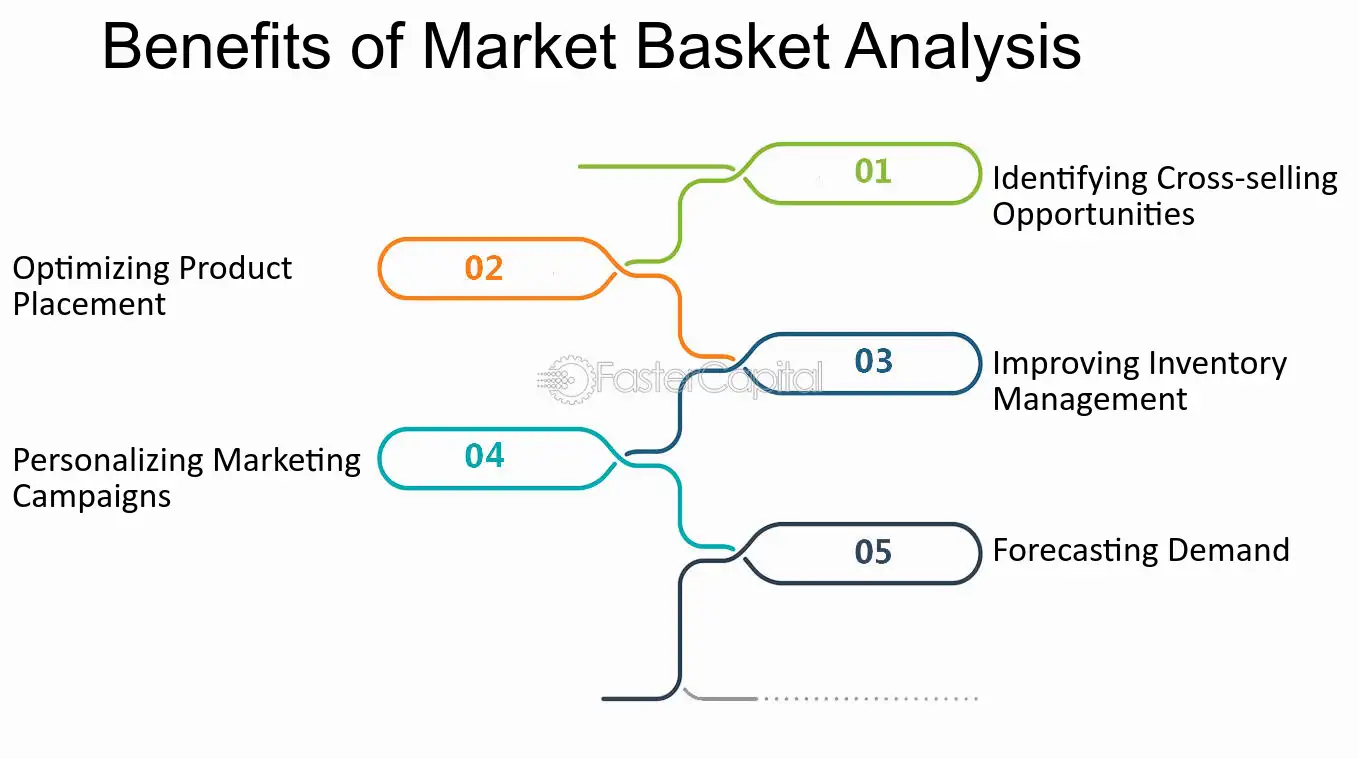


Figure 2.4.2 Benefits of MBA (Faster Capital, 2023).

Additional recent developments in MBA include the utilisation of statistical factors for the purpose of enhancing standard association rule mining approaches in order to conduct more in-depth analysis (Dogan, 2023). The use of this strategy allowed for the acquisition of more complex insights, such as the fact that when a customer adds items with a high discount rate to their shopping cart, items with a lower discount rate are suggested. The statistical enhancement of MBA enables a more nuanced understanding of the preferences and actions of consumers, which in turn makes it possible to optimise marketing strategies and product recommendations (Dogan, 2023). Text mining, deep learning, and fuzzy theory are examples of cutting-edge techniques and algorithms that are being incorporated into MBA curricula. This demonstrates how advanced and up-to-date the discipline is becoming. Because they make it possible to conduct a more in-depth analysis of market basket data, these methods are able to satisfy the ever-evolving requirements of modern retail and e-commerce environments (Dogan, 2023). Figure 2.4.2 shows some of the benefits of Market Basket Analysis.

In a nutshell the growth of e-commerce is heavily dependent on the collaboration that exists between market basket analysis and data analytics. The ability to mine customer data for analytical information that assists businesses in developing cross-selling strategies that are successful is provided by this technology. This leads to a rise in revenue growth as well as the value of customers, so establishing a foundation for sustained success in the competitive online business.

### **Impact Of Technology on Cross-Selling**

Technology plays a significant part in expanding the opportunities for cross-selling in virtual retail environments in this day and age, when the internet is becoming increasingly prevalent (Alexander and Cano, 2018). A considerable shift has occurred as a consequence of the implementation of cross-selling strategies that make use of big data and analytics. According to Charles Ntumba (2023), online retailers now have an unprecedented level of insight into the behavior, preferences, and shopping habits of their customers. This is made possible by their ability to collect, review, and analyze huge amounts of information about their customers. This data-driven strategy makes it feasible to increase the likelihood of cross-selling opportunities that are more personalised and focussed. It is now possible for companies to recommend products that are more likely to resonate with particular customers, hence increasing the likelihood that those customers would make a shopping buy. Technology-driven customer segmentation has made it possible for businesses to more precisely determine the preferences, habits, and purchase histories of their various customers (SITI ZULAIKHA, 2020).

Through the utilization of CRM (customer relationship management) platforms and POS (point-of-sale) technology, businesses have the ability to gather valuable information regarding the preferences of customers, their previous purchases, and their purchasing patterns (Mutambik et al. 2023). Afterward, this information can be put to use to personalize recommendations for cross-selling and to tailor incentives to specific customers. When a company can offer supplemental items or assistance to present customers according to their present needs or past purchases, it produces successful outcomes (Jarrah et al., 2020). In the process of doing so, the corporation is required to ensure that the proposed offers provide clients with further additional benefits. The objective of this strategy is to convince the customer to spend more money during the process of making a purchase or at the time of the transaction (Katell et al., 2020). Figure 2.5 shows Technologies that can enhance Cross-selling techniques.

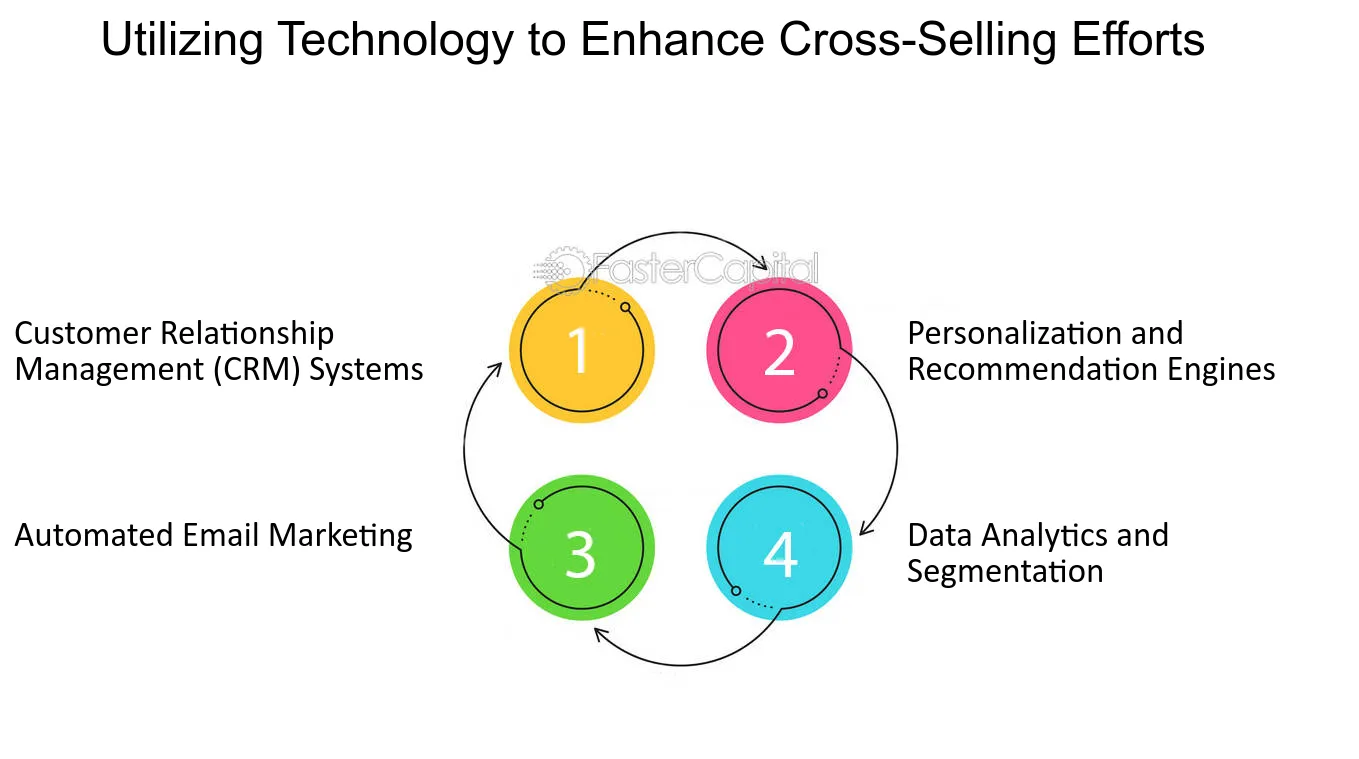


Figure 2.5 Technologies to enhance cross-selling (Faster Capital, 2023a).

The applications of artificial intelligence (AI) and machine learning (ML) have brought about a revolution in the field of personalization (Khrais, 2020). Algorithms can make product recommendations based on their capacity to anticipate the requirements of customers. Furthermore, the expansion of influencer marketing in general has affected cross-selling of products and services.Influencers regularly utilize product integration as a means of promoting brands by effectively reaching a large number of followers and encouraging consumers to make new purchases (Rajab, 2021). The use of social media platforms as a cross-selling strategy has proven increasingly effective.

New chances for cross-selling have emerged as a result of the increasing adoption of mobile applications and smartphones nowadays. Customers who shop using their smartphones provide businesses with additional information about how their customers act when they are on the move (Ami Fitri Utami and Irwan Adi Ekaputra, 2022). Smartphone shopping is convenient for clients. Cross-selling techniques can be successfully implemented through the use of push notifications and through the utilisation of app suggestions, which allow for the provision of precise and timely product recommendations (Huan Liu, 2022).

As a result of the incorporation of future technologies such as augmented reality (AR) and virtual reality (VR), it is anticipated that cross-selling strategies will be improved even more. Technological advancements have the potential to increase consumer involvement and revenue, these advancements can also enable retail encounters that are immersive and engaging (Carlos Orús, 2021). Additionally, as artificial intelligence and machine learning continue to progress, product recommendations will become more accurate and successful, as well as elevating cross-selling to a more complex and customer-focused activity. Many businesses will experience a shift in their operational patterns as a result of emerging trends such as AI-powered insights, the Internet of Things, and Intelligent Edge, or even Blockchain (Gill et al., 2022). These trends will also have an impact on the competitive dynamics of operations for many years to come.

Cross-selling in online retail is significantly impacted by technology, which has a wide-ranging and important impact. By doing so, businesses have been able to better understand and service their customers, which has led to an increase in both the worth of their customers and their income. These technological improvements, however, are not without their associated dangers, which must be carefully evaluated, particularly with regard to the protection of customer privacy and data. As innovation continues to advance, the ever-evolving realm of cross-sales in internet commerce will undoubtedly present new opportunities as well as challenges. Businesses have to find a way to strike a balance between the effectiveness of their cross-selling strategies and the duty to maintain the confidence of their customers and fulfil the requirements of privacy regulations.

**Literature Gap**

Eisenreich et al., (2021), have mentioned details regarding the development of business strategies that are not aligned with cross-selling processes specifically. This paper lacks information on the main subject of this current research. Khrais and Alghamdi, (2021), have not mentioned information on the perspectives of analytical strategies instead of proper e-commerce customer relationship strategies. In addition, research conducted by Mim and Ferdous (2021) and Adebisi (2023) emphasizes the expansion of e-commerce and the significance of cross-selling tactics, although does not extensively explore the precise influence of emerging technologies and customer behaviour.

This gap creates a potential for future research to investigate how machine learning algorithms might be customized and applied in the online supermarket industry in the United Kingdom to improve the effectiveness of cross-selling. This may encompass conducting tests with sophisticated machine learning techniques in actual online retail settings, analysing their influence on sales figures, consumer behaviour, inventory management, and personalization effectiveness.

# **Methodolgy**

Using Saunders' Research Onion structure, here's the Research Methodology section.

### **Research Philosophy and Approach**

In the rapidly increasing industry of e-commerce, particularly for online supermarkets in United Kingdom, it is essential to have a solid understanding of consumer behaviour and the patterns of their purchases. The purpose of this research is to investigate these patterns across the United Kingdom by applying Market Basket Analysis (MBA) to the dataset, which is a comprehensive database that contains transaction records from an online retail company. A rigorous identification of trends in consumer behaviour, an investigation into the efficacy of cross-selling strategies, and an evaluation of the influence these strategies have on revenue creation are the objectives of this study. Additionally, the study will investigate the intricacies of how these things function together, which will provide insight into the choices and purchasing habits of the audience as a whole. This will be done in addition to identifying items that are frequently purchased. Additionally, consideration will be given to how seasonal fluctuations and market trends influence the purchasing decisions of customers.

The proposed technique is based on quantitative research and makes use of data mining techniques to get significant findings from the extensive dataset. A substantial amount of information regarding transactions is included in the Online Retail dataset, which serves as the foundation for this investigation. Because it contains a large variety of fields, such as Bill No, Item name, Quantity, price, and information about customers, it is an excellent resource for investigating complex purchase trends.

### **Research Strategy and choice**

Our investigation will be significantly aided by Python, which is widely recognised for the robust data analysis capabilities it offers. One of the most important aspects of the research will be the preparation of the data, which will involve meticulously cleaning and preparing the relevant dataset. This method ensures that the subsequent analysis will be accurate and reliable.

It promises to guarantee both of these qualities. In the subsequent step, the investigation will make use of well-known machine learning approaches, known as the Apriori algorithm and RFM Analysis, which are recognized for their effectiveness in finding item correlations and patterns in big datasets and analyzing customer behaviour. The fact that these algorithms are capable of handling and comprehending transactional data makes them particularly well-suited for this research.

### **Data Collection**

A robust secondary dataset that was obtained through [Kaggle](https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis) serves as the foundation for the data collection strategy that was utilised in this investigation. A comprehensive collection of transaction records from an online store is included in this dataset. This particular dataset, titled "Online Retail Sales and Customer Data," is the central component of the market basket study. Its primary objective is to uncover patterns in the purchasing behaviour of customers who shop in an online retail setting.

The dataset that was gathered is quite vast and contains significant information from transactions, such as the BillNo, Itemname, Quantity, Date, Price, CustomerID, and geographic information related to the transaction. The aforementioned amount of data is essential for answering the research questions, which include gaining an understanding of the purchasing habits of consumers, identifying the products that are commonly purchased, and determining the effect that cross-selling has on revenue.

For the purpose of this investigation, direct sampling methods are not utilised because the dataset used is of a secondary origin. Despite this, the great volume and variety of transactions that are recorded make it possible for a natural stratification to occur across the various consumer segments and product categories.

### **Data Analysis**

It is a comprehensive collection of transactional data from an online retailer that is included in the dataset. Our analysis is focused on United Kingdom Demographics. Product details, transactional details, and consumer purchases are all included in the vast amount of information. Within the collection, there are a total of 4,87,622 records, which can be thought of as individual transaction entries. These records are dispersed among 7 distinct fields. The selected data is for year 2010-11 this date range does not significantly affect the validity of my research or my capacity to answer the research questions that are intended to be addressed.

Table 3.4 shows Key Fields of the Dataset:

| **Column Name** | **Description** |
| --- | --- |
| BillNo | A unique identifier for each transaction, identifying different types of transactions. Important for purchase pattern analysis and order tracking. |
| Itemname | Written description of each product, providing qualitative information about the product being sold. |
| Quantity | Number of units purchased in each transaction, vital for demand analysis, inventory planning, and understanding customer preferences. |
| Date | Date and time of each transaction, providing temporal context for evaluating customer engagement, seasonal patterns, and purchase trends. |
| Price | Cost incurred for each individual unit of the product. Essential for pricing strategies, revenue analysis, and customer perception of product value. |
| CustomerID | Unique identifier assigned to each customer, crucial for measuring customer loyalty, analyzing purchasing patterns, and implementing personalization strategies. |
| Country | Customer's country, providing geographic information for analyzing market penetration, regional preferences, and international growth prospects. |

Top of Form

Table 3.4 Dataset Fields

### **Data Pre-processing**

It was found that there were 1455 missing data in the Item name column, and there were 134041 missing data in the Customer ID column. Because these entries did not provide any information that could be of benefit, we came to the conclusion that it would be best not to impute these missing values. A total of 5,268 duplicate records are included in the dataset. We chose to eliminate these duplicates to preserve the credibility of our research. It is possible that these duplicates were errors or records that were already in existence. Cancellations are typically made for transactions that begin with the letter "C" in the invoice number. There is no consideration given to this aspect in the analysis.

As far as the quantity column was concerned, the highest possible value was 80995, and the lowest possible value was -9600. For example, mistakes or returns could be considered negative numbers. There is a range of values for the price, with the lowest being -11062.06 and the highest being 13541.33. When it comes to sales transactions, negative pricing is extremely rare and could be an indication of unethical behaviour or a specific type of transaction. The negative prices and quantities were therefore eliminated: In light of the fact that these are most likely corrections, refunds, or transactions that are not typical. In order to guarantee uniformity in the names of the items, we also standardised the cases. Fees and postage charges are included as line items in datasets; however, these entries should be eliminated because they do not represent actual products. After that, the items that were not products were taken out. In order to simplify the dataset for analysis, we encode the data using the binary transformation, which includes the values 1 for present and 0 for absent.

### **Exploratory Data Analysis**

Investigating the data and gaining an understanding of the trends is the next stage. The first section is dedicated to the task of counting and visually representing the number of distinct products, transactions, and customers that are contained within a dataset. This provides a better understanding of the structure of the dataset (IBM, 2020). This is an essential stage in gaining a grasp of the magnitude and variety of the data. We plot boxplots to understand the distribution and any potential anomalies to visualise the distribution of the number of products purchased across all transactions and to display the distribution of the item prices within the dataset. This allows us to have a better understanding of the distribution.

The next step is to create a bar graph, the objective of which is to determine which products are the most popular, taken into account the number of times they are purchased. It is beneficial for decision-makers in a firm to understand sales trends and to make informed business decisions, and this study provides them with such understanding. After that, a line graph that illustrates the progression of sales over time is plotted, so that seasonal patterns in sales can be identified and then we Plot the top 10 most sold products by quantity.

### **Analysis**

Utilising RFM analysis based on k- means clustering which allows us to analyse consumer behaviour and categorise clients into different groups. We then scale the data for clustering. The number of clusters that are present in a data set can be determined with the help of a heuristic called K-means clustering (P. Anitha, 2022). The acronym RFM is an abbreviation that stands for Recency, Frequency, and Monetary value. Each of these metrics is an important indicator of customer behaviour.

Recency (R): This metric determines how recently a consumer completed a purchase. When compared to customers who have not made a purchase in a considerable amount of time, individuals who have made a purchase in the recent past are more likely to make another buy in the future. Within the realm of practical application, this could entail classifying clients according to the date of their most recent transaction (Rendra Gustriansyah, 2020).

Frequency (F): The term "frequency" refers to the frequency with which a consumer makes a purchase regularly during a specified period. Those who make frequent purchases are typically more involved with and loyal to a brand, and retention initiatives need to have a solid grasp of their behavior. It is common practice to count the amount of purchases made by each consumer during a particular period when calculating this metric (Rendra Gustriansyah, 2020).

Monetary (M): This shows the total amount of money that a customer has spent over time and is denoted by the letter "M." Customers that spend more money are typically regarded as having a higher relative value to a company. The process entails determining the total revenue that is generated from each customer (Rendra Gustriansyah, 2020).

Then, we make use of a well-known for its application in market basket analysis, which is the process of analyzing customer transaction data to identify products that are frequently purchased together. The technique detects itemsets that recur frequently in a dataset in an iterative manner. It then expands these itemsets one item at a time, while simultaneously removing those that do not match the minimal support level (Santoso, 2021). The Apriori algorithm's ultimate objective is to provide association rules, which are implications expressed in the form of "if-then" statements (Xie, 2021). Understanding and predicting the purchase behaviour of customers is made easier with the help of these rules.

### **Ehical consideration**

Ethical concerns regarding the management and application of data are of the utmost importance in this investigation. The research will be conducted in accordance with the most stringent ethical requirements, particularly with regard to the protection of the privacy and confidentiality of any sensitive data. Given that the dataset only reflects a small portion of the online retail business, one possible limitation is the ability to generalize the results.

# **Observations**

In this section, we give a detailed analysis of the data that was seen, reflecting the results of the experiments and studies that were carried out. A synthesis of the findings is performed to attract attention to the key patterns, trends, and anomalies that were discovered through the study. It provides a quantitative perspective of the scale of the dataset by analysing the unique objects, transactions, and customers that are included in the dataset.

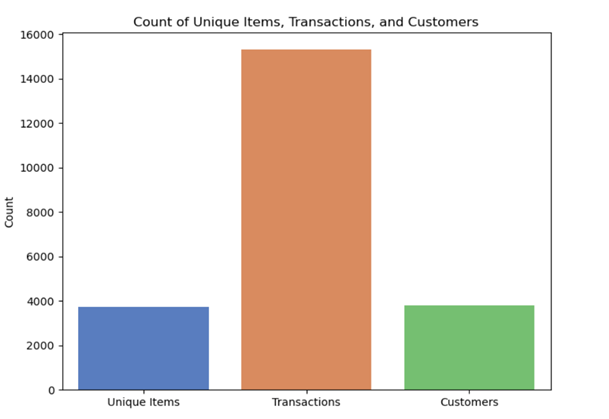


Figure 4.1 Count of Unique Values

With 3738 unique items, the retail website offers a wide selection of products for customers to choose from. According to the data, there are around 15304 unique transactions, while there are 3808 unique customers (Figure 4.1). As a result, even if there is a sizeable customer base, the number of transactions is far higher than the number of customers who are unique to the business. This may suggest that the same customers make repetitive purchases.

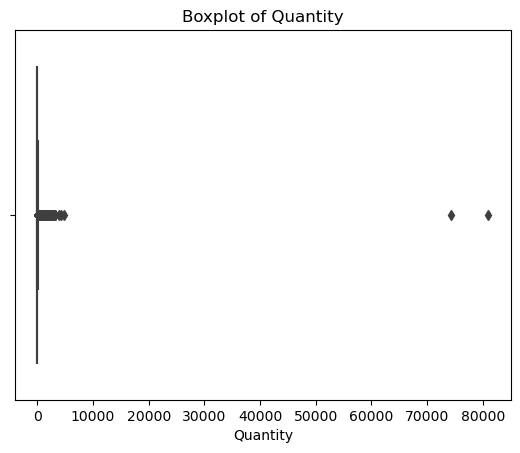
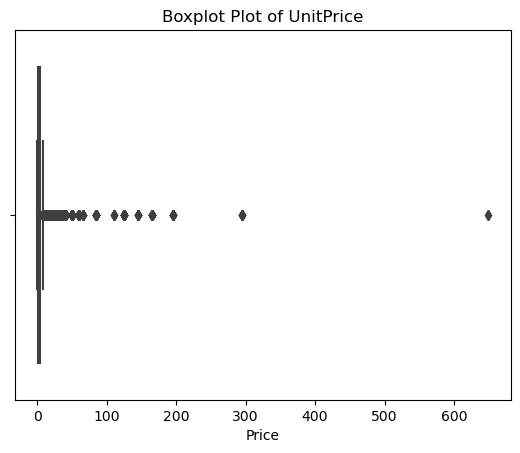


Figure 4.3 Quantity Boxplot

Figure 4.2 Unit Price Boxplot

Due to the presence of outliers in both plots, it is possible that there are some extreme numbers that could be the result of data entry errors, special circumstances, or bulk transactions that are not representative of the typical behaviour of customers. The quantity that is considered to be the median, which is the line that is located in the middle of the box, is 4 which indicates that the normal transaction contains a few items. Then again, on 1.79 units, the median price is present. At a value of 2.5, the Interquartile Range (IQR) indicates that the majority of items are priced within this range (Figure 4.2).

A skewed distribution towards lower-priced items is revealed by the boxplot and statistics for the Unit Price variable in an e-commerce dataset. Also, the median value for this variable is 1.79 units. Because of the existence of high-value outliers, the mean price is slightly higher than average, but the majority of products are clustered in a range that is cheaper (Figure 4.3).

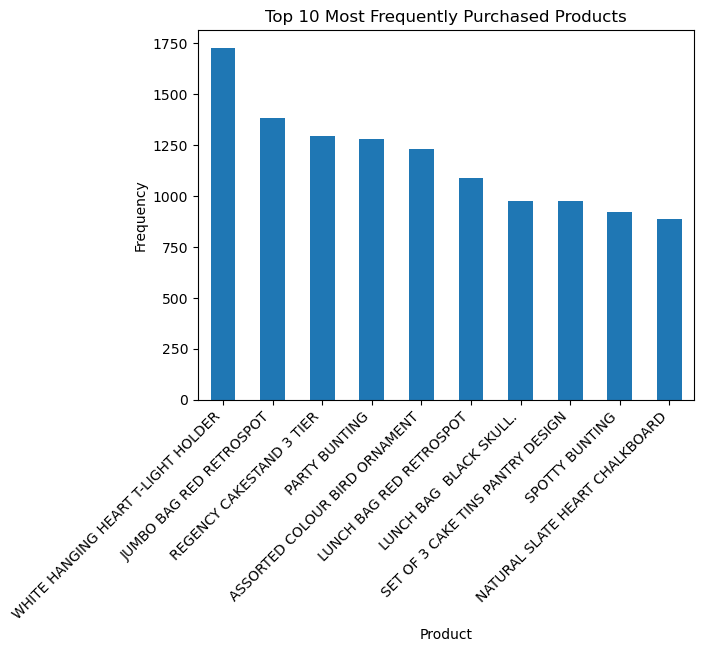


Figure 4.4 Frequently Purchased Products

A bar chart with the headline "Top 10 Most Frequently Purchased Products" is depicted in the Figure 4.4. Presented in descending order, it displays the number of times that the ten most popular products have been purchased. "WHITE HANGING HEART T-LIGHT HOLDER" looks to be the product with the highest frequency on the market, 1727. The second product on the list is referred to as "JUMBO BAG RED RETROSPOT," which has a frequency of 1384. The remaining goods have frequencies that are decreasing, with the third product having a frequency of 1294 and the tenth product, "NATURAL SLATE HEART CHALKBOARD," having a frequency of 885. The products are diverse, ranging from bags and kitchenware to decorative things and other different kinds of items. It is possible that this variability is indicative of a varied consumer base that has a wide range of priorities or interests. A firm can use the data that is reflected to make intelligent choices on stock levels, promotions, and product placement by using the information that is displayed. This makes it possible to compare the performance of different items, which can lead to strategic decisions regarding whether products should be prioritized or even discontinued.

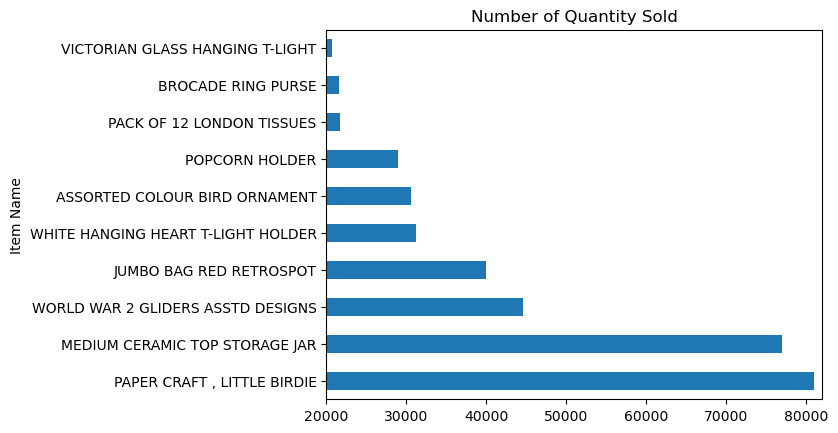
****

Figure 4.5 No. of Quantity of Products Sold

In the Figure 4.5, the quantity of items sold is shown in descending order, beginning with the most popular item and ending with the tenth most popular item. With a number of 80995, "PAPER CRAFT, LITTLE BIRDIE" is the item that has been sold the most. "MEDIUM CERAMIC TOP STORAGE JAR" looks to be the second most sold item, with a quantity of 76919, and then "WORLD WAR 2 GLIDERS ASSTD DESIGNS" comes in third place. In addition to this, the chart offers a prominent illustration of the discrepancy in sales volume that exists between the highest and lowest of the top 10 listings. The management of inventory, the development of marketing tactics, and the comprehension of client preferences could all benefit greatly from these insights.

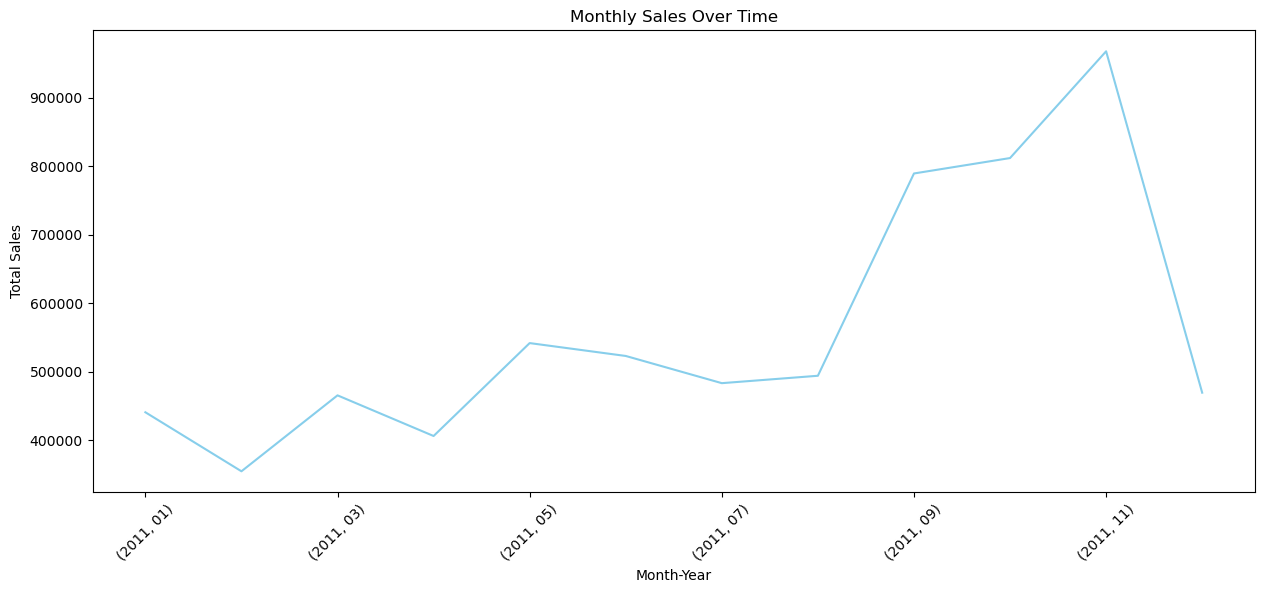


Figure 4.6 Sales Trend

Analysis of trends (Figure 4.6) provides a graphic representation of the total sales trends across a number of months. It would appear that there is a seasonal pattern involved, with sales reaching their highest point towards the end of the year. A major increase in sales appears to begin in the month of July, reach its highest point around the month of November, and then exhibit a significant fall in December.

Later, we determine the behaviour of our customers by analysing their recencies, frequencies, and monetary values. First, a range of k values is used in the k-means clustering analysis that we perform on the dataset (Figure 4.7). After testing with various values of k, we discover that k=3 produces clusters that are more useful for activities that come later in the process and that make sense for our application. We were able to develop the clustering model by utilising it. Here's an interpretation of the clusters based on the provided averages:

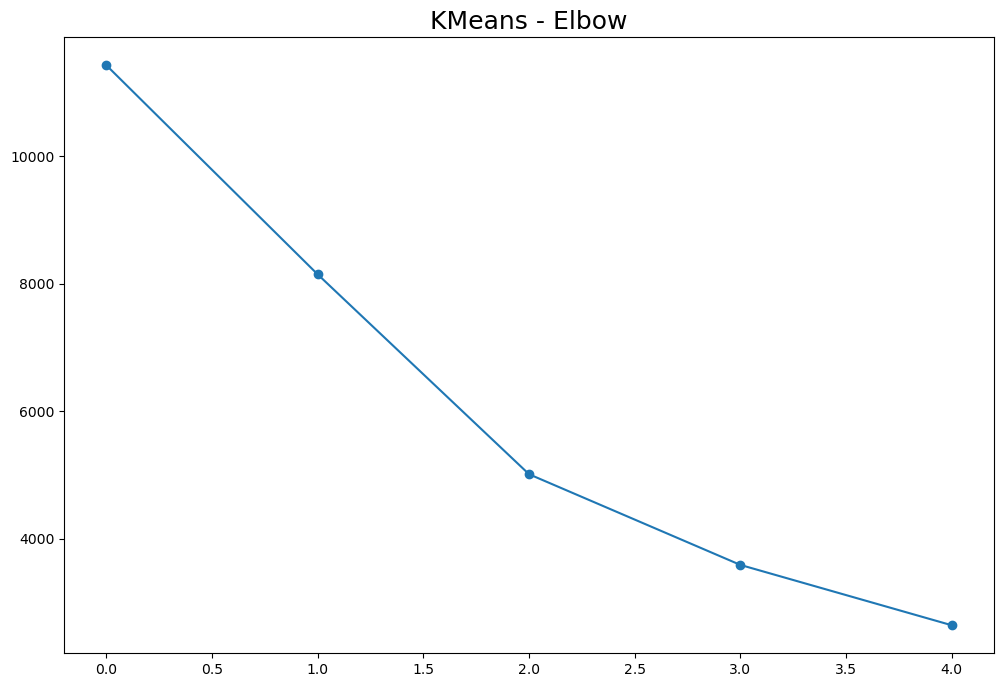


Figure 4.7 Elbow curve for K values

Cluster 1 appears to be a group of consumers that have a moderately recent purchase history (an average of 64.66 days passing since their most recent purchase), a low frequency of transactions (an average of approximately 4.57 purchases), and a relatively low monetary value (an average expenditure of 1711.53 units) as shown in Table 4.1.

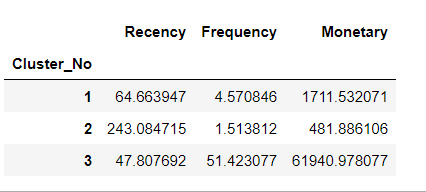


Table 4.1 RFM metrics

Cluster 2 consists of customers who have the lowest level of engagement or are the most recent customers. These customers have the highest recency value (average of 243.08 days since their last purchase), which indicates that they have not made any recent purchases; the lowest frequency (average of approximately 1.51 purchases), which suggests that they shop infrequently; and the lowest monetary value (average spend of 481.89 units), which indicates that they spend the least amount of money (Figure 4.1).

Cluster 3 consists of consumers who are the most frequent shoppers (with the highest average frequency of 51.42), who spent the most money (with the highest average monetary value of 61940.98 units), and who have made purchases more recently than those in Cluster 1 (with an average recency of 47.81 days) (Figure 4.1).

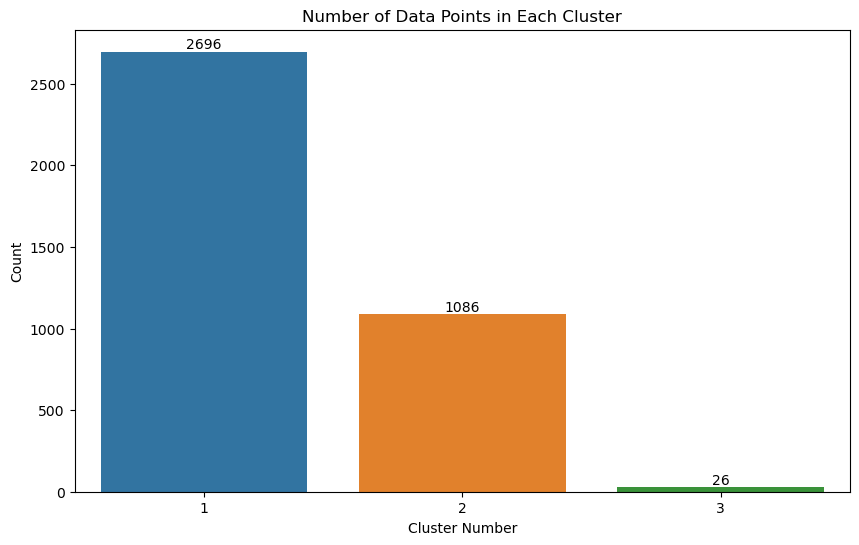


Figure 4.8 Number of Data points in each cluster

As shown in Figure 4.8, Cluster 1 contains only 26 data points, which is a very tiny number of overall data points. It's possible that this cluster is representative of a very specific type of customer. The number of data points in Cluster 2 is 1086, which is considered to be a moderate amount. Although this is a large portion, it is not the most significant. With a total of 2696 data points, Cluster 3 is the largest of the three clusters. It appears from this that the majority of the clients belong to this particular sector. The scatter plot is visualizing the same.

Later, we applied Apriori algorithm to the dataset. Two of the rules that were derived by the apriori technique are displayed in the Table 4.2. Here's a brief interpretation of the rules:

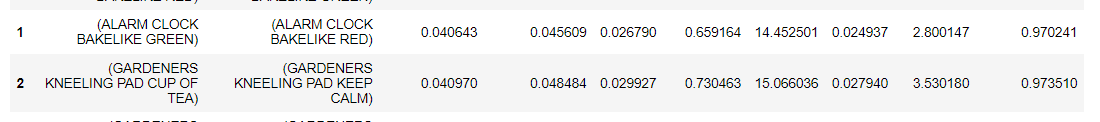


Table 4.2 Association Rules

The first rule in Table 4.2 suggests that there is a significant connection between the acquisition of "ALARM CLOCK BAKELIKE GREEN" and "ALARM CLOCK BAKELIKE RED." There is a strong level of confidence at 65.91%, even though the support for these combinations is only 0.46%, which indicates that they occur in less than half of all transactions. It can be deduced from this that when the 'ALARM CLOCK BAKELIKE GREEN' is purchased, there is a probability of 65.91% that the 'ALARM CLOCK BAKELIKE RED' is also included in the basket. The lift of 14.45 is particularly remarkable since it indicates that the likelihood of the two goods being purchased together is more than 14 times greater than the possibility of the red alarm clock being purchased on its alone.

According to the second rule, buyers who purchase a "(GARDENERS KNEELING PAD CUP OF TEA)" also have a tendency to purchase a "(GARDENERS KNEELING PAD KEEP CALM)" with a confidence level of 73.5 percent. The correlation between these items is significant, despite the fact that the frequency of this pairing in the sample is not particularly common, with a support of 0.41%. It is clear that the strength of this link is strengthened by the lift of 15.06, which indicates that the pairing occurs 15 times more frequently than what would be anticipated if the two things were functioning independently of one another. The high lift values, in particular, bring to light a potential option for cross-promotion or bundling, which might be utilised to improve sales of these complimentary products.

# **Analysis and Discussion**

In this part we discuss about our key findings.

### **Overview of Seasonal Variation In Consumer Markets**

The analysis of the monthly sales data from the online supermarket reveals significant seasonal patterns in consumer behaviour. These patterns correspond to the four traditional seasons in the United Kingdom, which are spring, summer, autumn, and winter (Anon., n.d.). The following is a breakdown of the differences in consumer behaviour and sales trends that occur with each season:

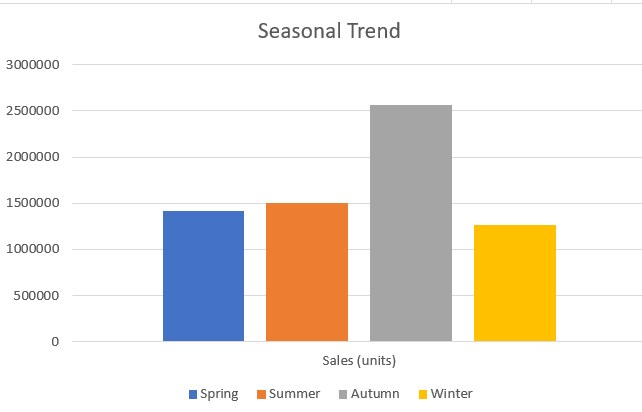


Figure 5.1 Seasonal Trend

From Figure 5.1, April marks the beginning of spring, which ushers in a period of revitalization in sales, indicating that there is a substantial seasonal impact on purchasing. There are a few possible explanations for this revival, including the introduction of new seasonal products or a general refill of stocks following the winter months. As the temperature rises and the days get longer in March, there is typically a resurgence in consumer interest in products that are associated to spring and the outdoors. This is a reflection of a change in lifestyle and preferences. On the other hand, sales are lower in the month of April. The volatility that occurred in April leads one to believe that although consumers are sensitive to the change of the seasons, there may also be a period of financial recuperation following the spending that occurred during the winter, particularly after the holiday season. In May, there is a discernible rise in the amount of consumer activity. There is a good chance that the ending of the school year and the anticipation of summer are factors that are contributing to this trend. When consumers are getting ready for summer activities and travel, this increase in sales could possibly be driven by the fact that summer holidays are getting closer and that they are looking for new experiences.

A shown in Figure 5.1, there is a continuance of the increasing trend in sales that was established in May, and it coincides with the beginning of summer. It would suggest that consumers are shifting their preferences towards goods and services that are complementary to summertime activities, such as vacations, outdoor entertainment, and leisure activities. The month of July reveals an intriguing trend of a modest decline in sales. There is a possibility that this decline can be ascribed to the commencement of vacation periods, which in turn disrupts typical purchase patterns. August, on the other hand, offers evidence of a stabilization in the purchasing behavior of consumers. This stability may be an indication that consumers have adjusted to their summer routines or that they are making purchases in preparation for the impending autumn season: both of these possibilities are possible.

The entrance of September is accompanied by a considerable increase in sales, which is most likely caused by the anticipation of the holiday season as well as the shopping for back-to-school items. At this time, consumers are actively engaged in the market, looking for new releases that are available in the autumn and taking advantage of early bird offers that are available for the Christmas season. October is often characterized by consistent sales, despite the fact that the graph does not expressly highlight this characteristic. This is because consumers are engrossed in the fall spirit, purchasing products for Halloween, and enjoying the remainder of the nice weather on the planet. Sales reach their highest point in November, which is fueled by the buildup to Black Friday, Cyber Monday, and early Christmas shopping sprees. November experiences a boom in consumer activity, with sales reaching their highest point (Figure 5.1).

The Figure 5.1 shows that there was a significant drop in December, which appears to be an unusual occurrence at first glance. This could be the result of a number of things, including a data cut-off that occurred prior to the customary Christmas shopping peak, a move towards in-store purchases for immediate holiday requirements, or a natural drop that occurred following the very intense shopping period that occurred in November. January is off to a fantastic start, most likely thanks to the New Year's resolutions and promotions that were made. Consumers may be taking advantage of sales in order to purchase products including fitness equipment, organizational tools, and other items to get the new year off to a good start. Last but not least, February reveals a small decline in sales. It's possible that this is due to customers cutting back on their spending following the holiday shopping rush, in addition to the fact that the month is getting shorter. With the approach of spring, the cycle will restart, but this dip will give you a chance to catch your breath.

### **Introduction To Consumer Buying Pattern In Online Supermarkets in UK**

When it comes to the modern environment of e-commerce, particularly within the online supermarket industry, analyzing the purchasing behavior of customers is not just enlightening; it is necessary for the continued existence and expansion of any retail firm. A strategic strategy for knowing and engaging with customers based on their purchasing behaviors can be achieved through the utilization of customer segmentation through RFM (Recency, Frequency, Monetary) research (Wilbert HJ, 2023). The RFM model divides customers into distinct groups or clusters according to the rate at which they make purchases (recency), the regularity with which they make purchases (frequency), and the amount of money they spend (monetary level). Through the lens of an RFM study, this section investigates the primary patterns of client purchase behavior that occur within the framework of an online supermarket.

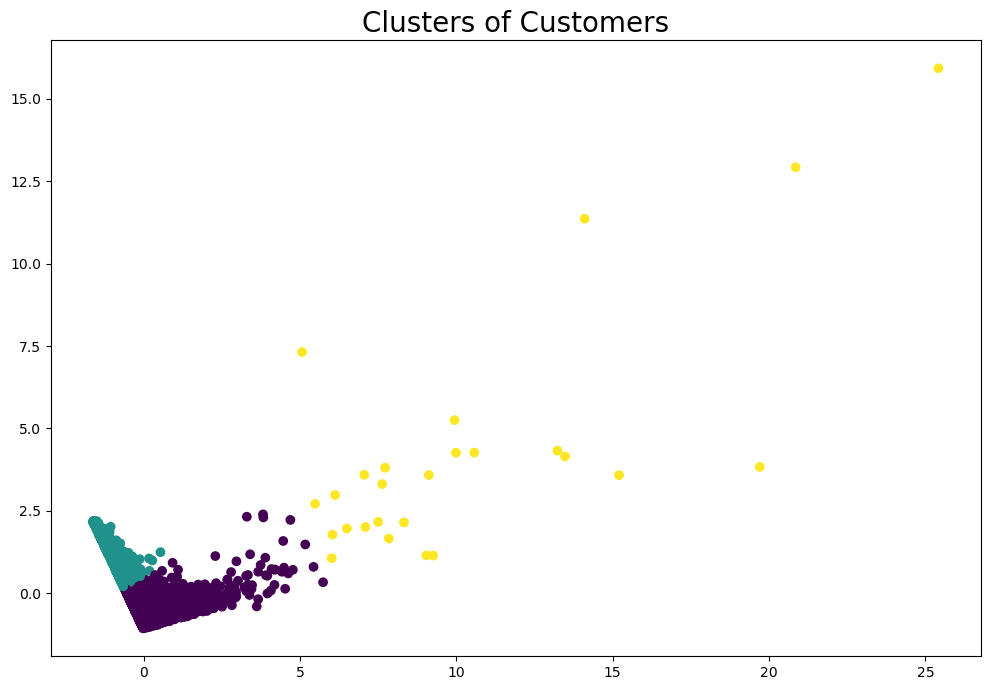


Figure 5.2 Clusters of Customers

There are three main consumer as shown in Figure 5.2, clusters that emerge from the data that was presented. These clusters provide a glimpse into the myriad of ways in which customers engage with the online supermarket and shed light on the complexities of consumer spending patterns. Cluster 1, with an average recency of 64.66 days, a frequency of 4.57 transactions, and a monetary value of 1,711.53 units, is the most populated, containing 2,696 customers. The behavior of this segment indicates that they have a consistent and reliable interaction with the supermarket, albeit they do not spend the most money or visit the shop the most frequently. These customers are regulars, and their persistent involvement with the store is probably driven by a combination of habit and happiness with the service they receive there. In addition to providing a reliable source of revenue, they are excellent candidates for loyalty programs, which have the potential to significantly increase the frequency of their purchases and the amount of money they contribute.

When we move on to Cluster 2, which is comprised of 1,086 clients, we see that there is a very different pattern. The average recency of this cluster is 243.08 days, the frequency of transactions is only 1.51 times per month, and the monetary amount is only 481.89 units. The supermarket appears to have a marginal relationship with this group, as seen by the rare and less recent interactions they have had with it. There is a possibility that they are clients who only purchase online occasionally or who have not yet completely incorporated the online supermarket into their regular shopping routine. It is vital to implement targeted re-engagement methods, such as personalised discounts and marketing that is driven by feedback, in order to capitalise on the potential of this group to increase the amount of time they spend shopping and the amount of money they spend.

Cluster 3 is the most appealing spending behaviour, despite the fact that it is the smallest group with only 26 individuals. These consumers have a high frequency of 51.42 transactions, a significant average monetary worth of 6,194.98 units, and an average recency of 47.81 days. Additionally, they have a high frequency of transactions. Their recent and frequent interactions, in addition to their substantial spending, highlight the significance of their relationship. There is a good chance that this category consists of premium customers or small enterprises that rely substantially on the store to fulfill their requirements. Given the profitable contribution they make, providing them with individualized services and private attention, possibly through a VIP program, could further solidify their loyalty and potentially increase the amount of money they spend, which is currently quite substantial.

Actionable insights can be obtained from the various numerical patterns that each cluster possesses. For example, Cluster 1, which has a moderate level of recency and frequency, could be urged to purchase more frequently using incentives that are time-sensitive. Cluster 2, which has the highest recency but the lowest frequency and spending, may require a different kind of prod, such as a personalized reintroduction to the various products and services that the store has to offer. The low recency, high frequency, and high monetary values of Cluster 3 indicate that it is of the utmost importance to sustain their high level of involvement. They could be attracted to alternatives for bulk purchasing or early access to premium products.

An in-depth image of the consumer base of the online supermarket is presented by the RFM study, which is broken down into separate purchase behaviours. The potential for focused marketing techniques, consumer interaction, and retention efforts that are available in each cluster are distinct from one another. A customer-centric strategy that ensures a high degree of customer happiness encourages returning clients, and fosters loyalty over time can be crafted with the assistance of the insights that are generated from this investigation.

### **The Impact Of Bundling: Insights from Complementary Products**

Using market basket analysis, online supermarkets have access to a wealth of information that can be used to improve their marketing, sales, and inventory strategies. This information may be gleaned from the complexities of client purchasing behaviour, which are exposed by the study. We are able to gain insights into the consumer mentality by analysing the data that captures often co-purchased items. This allows us to understand not just the specific products that interest consumers, but also the combinations that reflect larger preferences in terms of lifestyle and aesthetics.

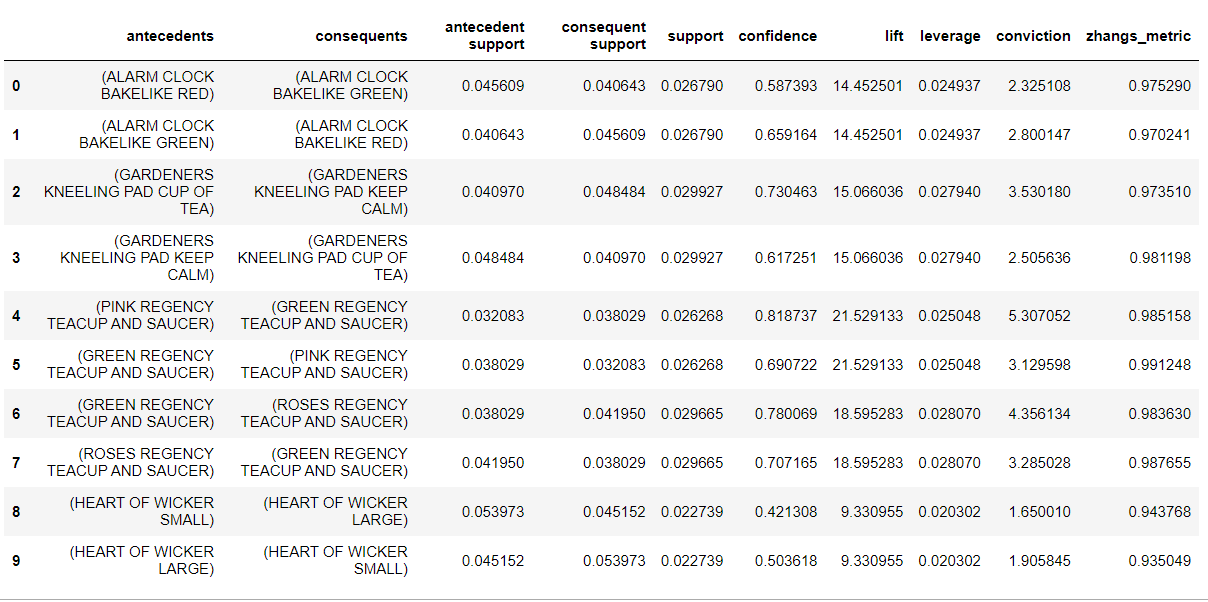


Table 5.3. Apriori Association Rules

As shown in Table 5.5, tendency for collecting or creating sets is indicated by the pairing of "Alarm Clock Bakelite Red" and "Alarm Clock Bakelite Green." This preference could be motivated by a variety of motives, such as a desire for aesthetic uniformity or the pleasure that is obtained from collecting goods that are similar to one another. When consumers opt to add a touch of antique charm to their house with one colored Bakelite alarm clock, they are extremely likely to reproduce this choice with another color variant. This is because the considerable lift value indicates that consumers are inclined to consider purchasing another color version. This indicates that buyers are not making individual judgments regarding their purchases but rather are curating collections that are reflective of a consistent theme or style among their purchases.

There is an even more obvious pattern that can be seen in the teacup and saucer sets. The "Pink Regency Teacup and Saucer" is frequently purchased in conjunction with the "Green Regency Teacup and Saucer." (Table 5.3). This could be an indication of a collector's interest in obtaining full sets for exhibition purposes, or it could be a reflection of an interest in entertaining, which requires a set that is both complete and coordinated. In this case, the high lift value is particularly illuminating because it indicates that the likelihood of purchasing these things together is almost 21 times higher than the likelihood of purchasing them separately. This is a clear indication that these items are viewed as being complementary to one another.

When these insights are translated into practical plans, it has the potential to greatly impact sales and the level of happiness experienced by customers. To give one example, if the online supermarket were to recognise the trend of collecting, it might decide to offer discounts or package deals when numerous items that are connected to one another are purchased. This would result in an increase in the value of the transaction and encourage customers to finish their sets. A personalised shopping experience could be provided by using personalised recommendations, which are powered by algorithms that identify such purchasing trends. These suggestions could propose things that other customers have commonly purchased together, so delivering a more specialised shopping experience.

In addition, gaining a grasp of the customer's proclivity for collecting or purchasing products in a sequence can provide valuable insight into more comprehensive sales methods. For instance, the online supermarket may initiate a promotion called "Collect Them All," which would encourage users to buy the whole assortment of a product collection. This not only generates sales for the products that are showcased, but it also has the potential to enhance traffic to the website and draw attention to other products that are available with the company.

Additionally, the study may reveal chances to improve the shopping experience, which, in and of itself, has the potential to exert a significant influence on sales. A deeper level of consumer engagement and a better rate of repeat purchases can be achieved, for example, through the creation of personalized shopping experiences through the use of product recommendations that are based on previous purchasing behaviour. Personalized experiences give customers the impression that they are understood and respected, which can lead to increased brand loyalty and, as a result, increased sales.

### **Inventory Syncronization: Insights from Market Basket Analysis**

By analysing the products that customers typically purchase together, market basket analysis is a strong analytical method that can be used to gain a better understanding of client purchasing patterns.

In Table 5.3, there are a number of combinations of 'Regency Teacup and Saucer' sets that demonstrate strong pairings, such as the fact that the 'Alarm Clock Bakelite Red' is frequently purchased with the 'Alarm Clock Bakelite Green'. The preference of customers for matching products and sets is reflected in these patterns, which is an important insight for inventory planning because it reveals a desire for matching things and sets. As an illustration, if one of the items in a set that is frequently purchased together gets out of stock, the sales of the item that is complementary to it may also suffer. Consequently, it is necessary to control the inventory levels of these paired products in conjunction with one another in order to guarantee that the availability of one thing can be matched by the purchase of the other item.

Furthermore, the high lift values that are linked with these product pairs, such as the teacups having a lift of over 21, indicate that customers are substantially more inclined to purchase these things together as opposed to purchasing them separately (Figure 5.3). Therefore, this has obvious repercussions for the management of supply chains. When merchants have a thorough grasp of these patterns, they are better able to foresee a rise in demand for particular products and ensure that their supply chain is able to accommodate these trends. This could entail giving more priority to the acquisition of high-lift items, making certain that the logistics are in place to deal with the corresponding demand, and modifying the reordering algorithms so that they take into consideration the purchase patterns that have been observed.

The information can also be used to inform the operations of supermarket warehouses. Due to the fact that certain commodities are commonly sold together, warehouses are able to store these things in close proximity to one another. This streamlines the process of picking and packaging, which in turn reduces the amount of time it takes from the time an order is placed until it is shipped out, thus leading to an increase in customer satisfaction through faster delivery times.

In addition, market basket research can be utilized to forecast chances for cross-selling and to develop marketing campaigns that encourage customers to acquire numerous things that are complimentary to themselves. Consequently, this may result in an increase in the average order value, which has a direct impact on the bottom line. For example, if a consumer purchases a "Heart of Wicker Small," the system may automatically recommend that they also purchase a "Heart of Wicker Large" to complete the set as shown in Figure 5.3. This would allow the system to capitalize on the trend that has been observed in the customer's purchasing behavior.

In addition, these realizations have repercussions for the procurement of products. For the purpose of ensuring a steady supply, it may be necessary to supervise suppliers for high-affinity commodities more carefully. Additionally, contracts may need to be drafted in order to take into consideration the dependency of product demand.

An online retailer's approach can be transformed from reactive to proactive with the assistance of market basket analysis, which has the potential to have a high strategic influence. Instead of reacting to stockouts after they have already occurred, retailers may utilize this information to anticipate and prevent stockouts from occurring.

In a nutshell, market basket analysis provides useful insights that have the potential to dramatically influence inventory management and supply chain optimization in the context of online shopping. Utilising the connections between items allows merchants to guarantee that they are able to fulfil the requirements of their customers, which ultimately results in more effective business operations and more sales. Through this research, transactional data is transformed into a strategic asset, which not only enables merchants to better understand their customers but also to provide them with more efficient service.

# **Conclusion**

The investigation conducted in this dissertation concerning cross-selling tactics within the domain of electronic commerce, specifically targeting the online supermarket industry in the United Kingdom, has yielded significant findings regarding the changing patterns of consumer conduct and the successful application of data analytics to improve customer satisfaction and increase revenue.

The main aims of this study were to examine consumer buying patterns in the online grocery industry, determine frequently purchased product combinations, evaluate the effect of cross-selling on sales, investigate the role of analytics in optimizing the supply chain, and explore the impact of seasonal trends on consumer behavior. The utilization of Market Basket Analysis (MBA) and RFM segmentation provided a comprehensive framework to fulfill these objectives. The results demonstrated clear trends in purchasing behavior and preferences, highlighting the success of cross-selling tactics in improving customer satisfaction and increasing income. By incorporating these analytical techniques, a more complex understanding of customer behavior was achieved, in line with the swift transformation of the online retail sector.

This study provides substantial advancements in the field of digital marketing and retail management in the United Kingdom, specifically in the realm of e-commerce. Firstly, it provides a thorough examination of how cross-selling, supported by data analytics, can be maximized in the online supermarket industry. The study emphasizes the significance of comprehending customer behavior and utilizing this insight for strategic decision-making in inventory management and marketing. Furthermore, this study enhances the current body of knowledge by presenting empirical data that demonstrates the influence of seasonal patterns on consumer buying habits. This observation is especially helpful for shops seeking to customize their strategy according to various seasons. Furthermore, the research highlights the crucial significance of technology in improving cross-selling techniques, with a focus on the potential of MBA (Market Basket Analysis) and RFM (Recency, Frequency, Monetary) in the field of electronic commerce.

From a managerial standpoint, this dissertation provides various practical insights that can be implemented. The findings highlight the significance of tailored cross-selling methods in improving customer experience and fostering loyalty. Retail managers can utilize these observations to create focused marketing campaigns and promotional deals that correspond with customer preferences and purchasing patterns. Furthermore, the study emphasizes the necessity of efficient inventory control, driven by data analytics, to guarantee the availability of products and enhance the efficiency of supply chain operations. This information can be utilized by managers to make well-informed decisions regarding stock levels, product placements, and pricing strategies. Moreover, the study's emphasis on seasonal patterns equips managers with a structure to predict shifts in customer behavior and adapt their strategy accordingly.

### **Limitations of the Research**

As this dissertation draws to a close, it is critical to recognise the following limitations that may have had an impact on the scope and results of the research:

Due to the high minimum support level of 0.02 that was used in this research, the usage of the Apriori method to build association rules for cross-selling programmes was subject to restrictions. This constraint, which was brought about by the computational constraints of the laptop that was accessible, resulted in a decreased number of association rules that were developed, which in turn limited the depth of the investigation. Given that just ten association rules were available, the breadth of the research was limited in terms of its ability to investigate a wide variety of cross-selling methods in online supermarkets. In order to overcome this constraint, increased processing power is required. This will allow for the full analysis of larger datasets, which will result in the production of a more extensive set of association rules. This will ultimately provide more in-depth insights into the purchasing behaviours of customers and allows for more sophisticated cross-selling techniques.

he fact that the research was only conducted in the United Kingdom is yet another key constraint of the investigation. There can be a significant amount of variation between regions in terms of consumer purchasing patterns, product preferences, and receptivity to cross-selling sales strategies. It is possible, for instance, that the cultural characteristics, economic conditions, and market maturity of the United Kingdom are different from other places.

In this study, rather than concentrating on a particular subset of the e-commerce business, the researchers make use of data that is broader in nature. While this technique does provide a comprehensive perspective, it may fail to take into account the intricacies and distinctive qualities that are associated with the various e-commerce industries. Every industry that falls under the umbrella of e-commerce, such as the fashion industry, the health and beauty industry, the grocery industry, and the electronics industry, has its own unique consumer behavior patterns, product categories, and purchase dynamics.

### **Recommendations**

Leveraging insights into customer behavior is essential for increasing customer value and generating revenue development in the competitive landscape of the online retail industry in the United Kingdom, notably in the supermarket sector.

An essential understanding of the levels of client interaction can be gained through RFM segmentation. Customers who have a modest level of recency and frequency but a high monetary value may be able to profit from incentive programs that encourage more frequent shopping in order to capitalize on their willingness to spend. Customers with a lower level of engagement, on the other hand, represent an unreached potential for expansion. This can be accomplished by the implementation of tailored re-engagement activities, such as personalized marketing campaigns and promotions guided by feedback, which can rekindle their interest and encourage them to become regular customers. For the high-value client sector, providing premium services and exclusive attention through VIP programs can strengthen customer loyalty and potentially increase the amount of money they spend, which is already large. This strategy recognizes the significance of their contributions and presents an opportunity further to increase the customer lifetime value of those customer relationships.

Market basket analysis is a useful tool that enhances these findings by illuminating the products that are frequently purchased together. The results of this investigation reveal that customers have a predilection for thematic or matching sets, which should be taken into consideration when developing inventory management techniques to guarantee that complimentary items are stocked in sync. The high lift values that are linked with particular product combinations indicate that shoppers are substantially more inclined to purchase these goods together. Using this information, decisions on the supply chain should be made, to ensure that logistics and inventory are responsive to the patterns of purchasing. Additionally, it enables cross-selling opportunities, which are situations in which algorithm-driven product suggestions might promote subsequent purchases, which in turn leads to an increase in the average order value achieved.

Moreover, these insights are further augmented by seasonal tendencies in purchasing behavior, which dictate a calendar for strategic marketing and stocking decisions. The ability of retail store to capitalize on seasonal peaks in consumer spending is ensured by aligning their product offerings with seasonal wants. For example, gardening supplies in the spring and holiday decorations in the winter are examples of seasonal demands.

The implementation of these strategies together can result in an improvement in customer value, which in turn can lead to a sustained growth in income for online retailer. The ultimate objective is to provide the consumer with a shopping experience that is not only personal, responsive, and profoundly rewarding but also ensures that the supermarket not only meets but also anticipates and surpasses the customer's expectations, so creating loyalty and promoting continued engagement throughout the shopping experience.

In conclusion, this dissertation presents in-depth research of cross-selling methods utilized in the online supermarket industry in UK demographics. It provides significant insights into the behavior of customers, the influence of data analytics, and the significance of technology in the realm of e-commerce. Not only do the findings fulfill the aims of the research, but they also contribute a substantial amount of value to the field. Furthermore, they provide managers with practical consequences in the marketplace that is fast expanding digitally.

# **References**

A R Efrat, R Gernowo and Farikhin3 (2020). *Consumer purchase patterns based on market basket analysis using apriori algorithms*. [online] Available at: https://iopscience.iop.org/article/10.1088/1742-6596/1524/1/012109/meta.

Addagarla, S.K. and Amalanathan, A. (2020). Probabilistic Unsupervised Machine Learning Approach for a Similar Image Recommender System for E-Commerce. *Symmetry*, 12(11), p.1783. doi:https://doi.org/10.3390/sym12111783.

Adebisi, A. (2023). *UK Retail Report, E-Commerce Trends, and Online Shopping Statistics 2023*. [online] Investingstrategy.co.uk. Available at: https://investingstrategy.co.uk/financial-news/uk-retail-report-e-commerce-trends-and-online-shopping-statistics-2023/.

Alexander, B. and Cano, M.B. (2018). Futurising the Physical Store in the Omnichannel Retail Environment. *Exploring Omnichannel Retailing*, pp.197–223. doi:https://doi.org/10.1007/978-3-319-98273-1\_9.

Anitha, P. and Patil, M.M. (2019). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, 34(5). doi:https://doi.org/10.1016/j.jksuci.2019.12.011.

Basimakopoulou, M., Theologou, K. and Tzavaras, P. (2022). A Literature Review on Digital Marketing: The Evolution of a Revolution. *Journal of Social Media Marketing*, 1(1), pp.30–40. doi:https://doi.org/10.33422/jsmm.v1i1.901.

Bhalla, T. (2021). *Online retail set to see its biggest jump in 2021*. [online] mint. Available at: https://www.livemint.com/industry/retail/online-retail-expected-to-witness-its-biggest-jump-in-2021-11625061039983.html.

Brand, C., Schwanen, T. and Anable, J. (2020). ‘Online Omnivores’ or ‘Willing but struggling’? Identifying online grocery shopping behavior segments using attitude theory. *Journal of Retailing and Consumer Services*, 57, p.102195. doi:https://doi.org/10.1016/j.jretconser.2020.102195.

ÇİÇEKLİ, U.G. and KABASAKAL, İ. (2021). Market Basket Analysis of Basket Data with Demographics: A Case Study in E-Retailing. *Alphanumeric Journal*. doi:https://doi.org/10.17093/alphanumeric.752505.

Costa, J. and Castro, R. (2021). SMEs Must Go Online—E-Commerce as an Escape Hatch for Resilience and Survivability. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(7), pp.3043–3062. doi:https://doi.org/10.3390/jtaer16070166.

Dutta, S. and Bhattacharya, S. (2019). Cross selling of investment products and services: a case study of leading financial services organisation. *International Journal of Business Forecasting and Marketing Intelligence*, 5(2), p.241. doi:https://doi.org/10.1504/ijbfmi.2019.101608.

Eisenreich, A., Füller, J. and Stuchtey, M. (2021). Circular Project Selection: How Companies Can Evaluate Circular Innovation Projects. *Sustainability*, 13(22), p.12407. doi:https://doi.org/10.3390/su132212407.

Falcone, E., Kent, J. and Fugate, B. (2019). Supply chain technologies, interorganizational network and firm performance. *International Journal of Physical Distribution & Logistics Management*, 50(3), pp.333–354. doi:https://doi.org/10.1108/ijpdlm-08-2018-0306.

Faster Capital. (2023a). *Cross selling opportunities: Leveraging Revenue Growth: Uncovering Cross Selling Opportunities*. [online] Available at: https://fastercapital.com/content/Cross-selling-opportunities--Leveraging-Revenue-Growth--Uncovering-Cross-Selling-Opportunities.html.

Faster Capital. (2023b). *Data Analytics: Unleashing the Power of Market Basket Analysis*. [online] Available at: https://fastercapital.com/content/Data-Analytics--Unleashing-the-Power-of-Market-Basket-Analysis.html.

FasterCapital. (n.d.). *Cross selling: Maximizing Sales Mix with Cross Selling Techniques*. [online] Available at: https://fastercapital.com/content/Cross-selling--Maximizing-Sales-Mix-with-Cross-Selling-Techniques.html.

Fokina, M. (2023). *Online Shopping Statistics: Ecommerce Trends for 2022*. [online] Tidio. Available at: https://www.tidio.com/blog/online-shopping-statistics/.

Ghoshal, A., Mookerjee, V.S. and Sarkar, S. (2021). Recommendations and Cross-selling: Pricing Strategies when Personalizing Firms Cross-sell. *Journal of Management Information Systems*, 38(2), pp.430–456. doi:https://doi.org/10.1080/07421222.2021.1912930.

Gill, S.S., Xu, M., Ottaviani, C., Patros, P., Bahsoon, R., Shaghaghi, A., Golec, M., Stankovski, V., Wu, H., Abraham, A., Singh, M., Mehta, H., Ghosh, S.K., Baker, T., Parlikad, A.K., Lutfiyya, H., Kanhere, S.S., Sakellariou, R., Dustdar, S. and Rana, O. (2022). AI for next generation computing: Emerging trends and future directions. *Internet of Things*, 19, p.100514. doi:https://doi.org/10.1016/j.iot.2022.100514.

Goldman, S.P., Herk, H. van , Verhagen, T. and Weltevreden, J.W. (2020). Strategic orientations and digital marketing tactics in cross-border e-commerce: Comparing developed and emerging markets. *International Small Business Journal: Researching Entrepreneurship*, [online] 39(4), p.026624262096265. doi:https://doi.org/10.1177/0266242620962658.

Gomez, H., Navarro, V.G., Badenes, R.O. and Quilis, J.A.L. (2020). Customer relationship management: digital transformation and sustainable business model innovation. *Economic Research-Ekonomska Istraživanja*, [online] 33(1), pp.2733–2750. doi:https://doi.org/10.1080/1331677x.2019.1676283.

Gu, S., Ślusarczyk, B., Hajizada, S., Kovalyova, I. and Sakhbieva, A. (2021). Impact of the COVID-19 Pandemic on Online Consumer Purchasing Behavior. *Journal of Theoretical and Applied Electronic Commerce Research*, [online] 16(6), pp.2263–2281. Available at: https://www.mdpi.com/0718-1876/16/6/125.

Gupta, S., Kushwaha, Pooja.S., Badhera, U., Chatterjee, P. and Gonzalez, E.D.R.S. (2023). Identification of Benefits, Challenges, and Pathways in E-commerce Industries: An integrated two-phase decision-making model. *Sustainable Operations and Computers*, [online] 4. doi:https://doi.org/10.1016/j.susoc.2023.08.005.

Gustriansyah, R., Suhandi, N. and Antony, F. (2019). *Clustering optimization in RFM analysis based on k-means*. [online] Available at: https://d1wqtxts1xzle7.cloudfront.net/63983488/20264-40355-1-PB20200721-118680-1taddep-libre.pdf?1595349331=&response-content-disposition=inline%3B+filename%3DClustering\_optimization\_in\_RFM\_analysis.pdf&Expires=1704727702&Signature=OCLghCBi6bD7fE19jPo93a9HYKQxryB3azyOzVovI3YqRUNDUYnMlDEWZKvibQg-H0p-3RXO3Qb6c3Grs7H97ot~NVpTP6uoBUM760Uq2CVNefPUs5Lhwog~tUW9nPsq7yhDcrNwOH63pXfiqlvjW3KvvbzNTYbo2KoT6xOWTXWnHPd4iJNclwo7mgzyLFJ6Kvh0i4dDoH9-Sd0hDZtXKdzNgkPOvYioGiJ7aRDNWfkxp54JrBdXwaj6q2J3HTt5M8SgA2kRm8QP0RpBEOp2n22cST~VdzjYD4u9WjN3A36zg9dAyFSkTndKeMG0pg3IVHouzKEojg3snv7h1767~A\_\_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA.

IBM (2020). *What Is Exploratory Data Analysis? | IBM*. [online] www.ibm.com. Available at: https://www.ibm.com/topics/exploratory-data-analysis.

Jarrah, M.A.A., Alomari, G.I., Alhawari, S.F. and Serhan, M.Al. (2020). Building a Conceptual Model for E-CRM Implementation Strategy. *International Journal of Academic Research in Business and Social Sciences*, 10(1). doi:https://doi.org/10.6007/ijarbss/v10-i1/6834.

Javed, U., Shaukat, K., Hameed, I.A., Iqbal, F., Alam, T.M. and Luo, S. (2021). A Review of Content-Based and Context-Based Recommendation Systems. *International Journal of Emerging Technologies in Learning (iJET)*, [online] 16(3), pp.274–306. Available at: https://www.learntechlib.org/p/219036/?nl=1.

Kalkan, İ.E. and Şahin, C. (2022). Evaluating cross-selling opportunities with recurrent neural networks on retail marketing. *Neural Computing and Applications*. doi:https://doi.org/10.1007/s00521-022-08019-1.

Kasthuri Subaramaniam and Raenu Kolandaisamy (2019). User Acceptance and Adaptation Towards M-Commerce. doi:https://doi.org/10.1007/978-981-13-0203-9\_28.

Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Bintz, C., Raz, D. and Krafft, P.M. (2020). Toward situated interventions for algorithmic equity. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. doi:https://doi.org/10.1145/3351095.3372874.

Katsikeas, C., Leonidou, L. and Zeriti, A. (2019). Revisiting International Marketing Strategy in a Digital Era. *International Marketing Review*, [online] 37(3), pp.405–424. doi:https://doi.org/10.1108/IMR-02-2019-0080.

Khrais, L.T. (2020). Role of Artificial Intelligence in Shaping Consumer Demand in E-Commerce. *Future Internet*, [online] 12(12), p.226. doi:https://doi.org/10.3390/fi12120226.

Khrais, L.T. and Alghamdi, A.M. (2021). The Role of Mobile Application Acceptance in Shaping E-Customer Service. *Future Internet*, [online] 13(3), p.77. doi:https://doi.org/10.3390/fi13030077.

Kim, H. and Rao, V.R. (2023). A comparison of online recommendation methods: Simultaneous versus sequential approaches. *Journal of Retailing*. doi:https://doi.org/10.1016/j.jretai.2023.02.001.

Kumar Panda, T., Kumar, A., Jakhar, S., Luthra, S., Garza-Reyes, J., Kazancoglu, I., Sonali, S. and Nayak (n.d.). *Social and environmental sustainability model on consumers’ altruism, green purchase intention, green brand loyalty and evangelism Social and environmental sustainability model on consumers’ altruism, green purchase intention, green brand loyalty and evangelism*. [online] Available at: https://repository.derby.ac.uk/download/61fe97057e4aaedf8e3f1146c1a99eb3aaeecbf8ee25b5195024b2ba4ed4af69/825973/JCLEPRO-D-19-07333R1-pages-2%2C27-60.pdf. [Accessed 8 Jan. 2024].

Liu, H. and Sese, F.J. (2021). The Impact of Mobile App Adoption on Cross-buying: The Moderating Roles of Product Category Characteristics and Adoption Timing. *Journal of Retailing*. doi:https://doi.org/10.1016/j.jretai.2021.02.002.

Malesev, S. and Cherry, M. (2021). *Digital and social media marketing - growing market share for construction SMEs*. [online] Construction Economics and Building. Available at: https://search.informit.org/doi/abs/10.3316/informit.747666622867324.

Mason, A.N., Narcum, J. and Mason, K. (2021). Social media marketing gains importance after Covid-19. *Cogent Business & Management*, [online] 8(1), p.1870797. doi:https://doi.org/10.1080/23311975.2020.1870797.

Met Office. (n.d.). *Our seasons*. [online] Available at: https://www.metoffice.gov.uk/weather/learn-about/met-office-for-schools/other-content/other-resources/our-seasons.

Mim, S. and Ferdous, M. (2021). *FACTORS INFLUENCING CUSTOMERS SATISFACTION IN HOSPITALITY INDUSTRY: FINE DINING RESTAURANTS*. [online] Available at: https://www.diva-portal.org/smash/get/diva2:1530554/FULLTEXT01.pdf.

Murdiana, R. and Hajaoui, Z. (2020). E-Commerce marketing strategies in industry 4.0. *International Journal of Business Ecosystem & Strategy (2687-2293)*, 2(1), pp.30–41. doi:https://doi.org/10.36096/ijbes.v2i1.165.

Mutambik, I., Lee, J., Almuqrin, A., Zhang, J.Z., Baihan, M. and Alkhanifer, A. (2023). Privacy Concerns in Social Commerce: The Impact of Gender. *Sustainability*, [online] 15(17), p.12771. doi:https://doi.org/10.3390/su151712771.

Ntumba, C., Aguayo, S. and Maina, K. (2023). Revolutionizing Retail: A Mini Review of E-commerce Evolution. *Journal of Digital Marketing and Communication*, [online] 3(2), pp.100–110. doi:https://doi.org/10.53623/jdmc.v3i2.365.

Omar, H., Klibi, W., Babai, M.Z. and Ducq, Y. (2023). Basket data-driven approach for omnichannel demand forecasting. *International Journal of Production Economics*, 257, p.108748. doi:https://doi.org/10.1016/j.ijpe.2022.108748.

Onur Doğan (2023). Market Basket Analysis with Statistically Improved Association Rules Considering Product Details. *Research Square (Research Square)*. doi:https://doi.org/10.21203/rs.3.rs-2581178/v1.

Orús, C., Ibáñez-Sánchez, S. and Flavián, C. (2021). Enhancing the customer experience with virtual and augmented reality: The impact of content and device type. *International Journal of Hospitality Management*, [online] 98, p.103019. doi:https://doi.org/10.1016/j.ijhm.2021.103019.

Park, E.S. and Park, M.S. (2020). Factors of the Technology Acceptance Model for Construction IT. *Applied Sciences*, [online] 10(22), p.8299. doi:https://doi.org/10.3390/app10228299.

Ponzoa, J.M. and Erdmann, A. (2021). E-Commerce Customer Attraction: Digital Marketing Techniques, Evolution and Dynamics across Firms. *Journal of Promotion Management*, [online] 27(5), pp.1–19. Available at: https://www.tandfonline.com/doi/abs/10.1080/10496491.2021.1880521.

Rajab, M. (2021). Relationship of E-Commerce, Influencer Marketing and Social Media to Digital Marketing (Marketing Management Literature Review). *Dinasti International Journal of Economics, Finance & Accounting*, 2(4), pp.443–455. doi:https://doi.org/10.38035/dijefa.v2i4.1432.

Rosário, A. and Raimundo, R. (2021). Consumer Marketing Strategy and E-Commerce in the Last Decade: a Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*, [online] 16(7), pp.3003–3024. Available at: https://www.mdpi.com/0718-1876/16/7/164.

Sajid, S., Rashid, R.M. and Haider, W. (2022). Changing Trends of Consumers’ Online Buying Behavior During COVID-19 Pandemic With Moderating Role of Payment Mode and Gender. *Frontiers in Psychology*, [online] 13. doi:https://doi.org/10.3389/fpsyg.2022.919334.

Santoso, M.H. (2021). Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom. *Brilliance: Research of Artificial Intelligence*, [online] 1(2). Available at: https://jurnal.itscience.org/index.php/brilliance/article/view/1228.

Sheth, J.N., Jain, V. and Ambika, A. (2023). The growing importance of customer-centric support services for improving customer experience. *Journal of Business Research*, [online] 164, p.113943. doi:https://doi.org/10.1016/j.jbusres.2023.113943.

Siddharth, L., Blessing, L. and Luo, J. (2022). Natural language processing in-and-for design research. *Design Science*, 8. doi:https://doi.org/10.1017/dsj.2022.16.

Sinha, A. (2021). Implying Association Rule Mining and Market Basket Analysis for Knowing Consumer Behavior and Buying Pattern in Lockdown - A Data Mining Approach. *www.preprints.org*. [online] doi:https://doi.org/10.20944/preprints202105.0102.v1.

Sjarif, N.N.A., Azmi, N.F.M., Yuhaniz, S.S. and Wong, D.H.T. (2021). A review of market basket analysis on business intelligence and data mining. *International Journal of Business Intelligence and Data Mining*, 18(3), p.383. doi:https://doi.org/10.1504/ijbidm.2021.114475.

UNCTAD (2021). *Global e-commerce jumps to $26.7 trillion, COVID-19 boosts online sales | UNCTAD*. [online] unctad.org. Available at: https://unctad.org/news/global-e-commerce-jumps-267-trillion-covid-19-boosts-online-sales.

Ünvan, Y.A. (2020). Market basket analysis with association rules. *Communications in Statistics - Theory and Methods*, 50(7), pp.1–14. doi:https://doi.org/10.1080/03610926.2020.1716255.

Utami, A.F., Ekaputra, I.A., Japutra, A. and Van Doorn, S. (2021). The role of interactivity on customer engagement in mobile e-commerce applications. *International Journal of Market Research*, p.147078532110274. doi:https://doi.org/10.1177/14707853211027483.

Wang, Q., Zhu, X., Wang, M., Zhou, F. and Cheng, S. (2023). A theoretical model of factors influencing online consumer purchasing behavior through electronic word of mouth data mining and analysis. *PLoS ONE*, 18(5), pp.e0286034–e0286034. doi:https://doi.org/10.1371/journal.pone.0286034.

Wewege, L. and Thomsett, M.C. (2019). *The Digital Banking Revolution: How Fintech Companies are Transforming the Retail Banking Industry Through Disruptive Financial Innovation*. [online] *Google Books*. Walter de Gruyter GmbH & Co KG. Available at: https://books.google.com/books?hl=en&lr=&id=YDrEDwAAQBAJ&oi=fnd&pg=PR7&dq=online+transactions+are+likely+to+rise+further+and+seize+the+lion%27s+share+of+the+retail+pie.&ots=PS4IvAg3Di&sig=MCVHKlWjD5OaeeLc4gsaozuL\_zI.

Wilbert, H.J., Hoppe, A.F., Sartori, A., Stefenon, S.F. and Silva, L.A. (2023). Recency, Frequency, Monetary Value, Clustering, and Internal and External Indices for Customer Segmentation from Retail Data. *Algorithms*, [online] 16(9), p.396. doi:https://doi.org/10.3390/a16090396.

Xie, H. (2021). Research and Case Analysis of Apriori Algorithm Based on Mining Frequent Item-Sets. *Open Journal of Social Sciences*, [online] 9(4), pp.458–468. doi:https://doi.org/10.4236/jss.2021.94034.

Xu, A.J., Loi, R., Chow, C.W.C. and Lin, V.S.Z. (2022). Driving Retail Cross-Selling. *Journal of Service Research*, p.109467052210873. doi:https://doi.org/10.1177/10946705221087399.

Zamil, A.M.A., Al Adwan and Vasista (2020). *Enhancing customer loyalty with market basket analysis using innovative methods: a python implementation approach. International Journal of Innovation, Creativity and Change,.*.

Zhang, M. and Bockstedt, J. (2020). Complements and substitutes in online product recommendations: The differential effects on consumers’ willingness to pay. *Information & Management*, 57(6), p.103341. doi:https://doi.org/10.1016/j.im.2020.103341.

ZULAIKHA, S., MOHAMED, H., KURNIAWATI, M., RUSGIANTO, S. and RUSMITA, S.A. (2020). Customer Predictive Analytics Using Artificial Intelligence. *The Singapore Economic Review*, [online] pp.1–12. doi:https://doi.org/10.1142/s0217590820480021.

# **Appendix**

!pip install mlxtend

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import mlxtend

from mlxtend.frequent\_patterns import apriori, association\_rules

import warnings

warnings.filterwarnings('ignore')

# Read data

data = pd.read\_csv('Data.csv', delimiter=';', decimal=',', encoding='utf-8-sig')

# Display 5 rows

data.head()

data.info()

data.describe()

# as we are analyzing only for the UK region

df = data[data['Country'] == 'United Kingdom']

df.shape

## Data Cleaning

# check for missing values in the dataset

missing\_values = df.isnull().sum()

missing\_values

# Remove rows where 'Itemname' is missing

cleaned\_data = df.dropna(subset=['Itemname'])

# Remove rows where 'CustomerID' is missing

cleaned\_data = df.dropna(subset=['CustomerID'])

# Recheck for missing values after removal

MissingValues\_after = cleaned\_data.isnull().sum()

cleaned\_data.shape, MissingValues\_after

# Checking for duplicate records

duplicate\_records = cleaned\_data.duplicated().sum()

# Display the number of duplicate records found

duplicate\_records

# Remove duplicate records

cleaned\_data\_no\_duplicates = cleaned\_data.drop\_duplicates()

# Confirm removal of duplicates by rechecking

duplicate\_records\_after = cleaned\_data\_no\_duplicates.duplicated().sum()

remaining\_records = cleaned\_data\_no\_duplicates.shape[0]

remaining\_records, duplicate\_records\_after

# Identifying and removing cancelled transactions

cancelled\_transactions = cleaned\_data\_no\_duplicates['BillNo'].str.contains('C', na=False)

cancelled\_transactions.head(5)

cleaned\_data\_no\_cancellations = cleaned\_data\_no\_duplicates[~cancelled\_transactions]

cleaned\_data\_no\_cancellations.shape

df=cleaned\_data\_no\_cancellations.loc[(cleaned\_data\_no\_cancellations['Itemname']!='POSTAGE')&(cleaned\_data\_no\_cancellations['Itemname']!='DOTCOM POSTAGE')&(cleaned\_data\_no\_cancellations['Itemname']!='Adjust bad debt')&

(cleaned\_data\_no\_cancellations['Itemname']!='Manual')]

df=df.loc[df['Quantity']>0]

df=df.loc[df['Price']>0]

df

df.info()

# Convert CustomerID data type to object (string)

df['CustomerID'] = df['CustomerID'].astype('object')

# Splitting data into year and month

df['Year']=df['Date'].apply(lambda x:x.split('.')[2])

df['Year']=df['Year'].apply(lambda x:x.split(' ')[0])

df['Month']=df['Date'].apply(lambda x:x.split('.')[1])

df.head()

# Creating a Total price

df['Price']=df['Price'].apply(str).str.replace(',','.').astype('float64')

df['Total price']=df.Quantity\*df.Price

df.head()

#Checking the Total price in each month.

df.groupby(['Year','Month'])['Total price'].sum()

#Delete records for 2010

df=df.loc[df['Year']!='2010']

df.shape

## EDA

# Count of Unique Items, Transactions, and Customers

unique\_items\_count = df['Itemname'].nunique()

unique\_transactions\_count = df['BillNo'].nunique()

unique\_customers\_count = df['CustomerID'].nunique()

# Print the counts

print(f"Total number of unique items: {unique\_items\_count}")

print(f"Total number of transactions: {unique\_transactions\_count}")

print(f"Total number of unique customers: {unique\_customers\_count}")

# Plotting

counts = [unique\_items\_count, unique\_transactions\_count, unique\_customers\_count]

labels = ['Unique Items', 'Transactions', 'Customers']

plt.figure(figsize=(10, 6))

sns.barplot(x=labels, y=counts, palette='muted')

plt.title('Count of Unique Items, Transactions, and Customers')

plt.ylabel('Count')

plt.show()

sns.boxplot(x=df['Quantity'])

plt.title('Boxplot of Quantity')

plt.show()

sns.boxplot(x=df['Price'])

plt.title('Boxplot Plot of UnitPrice')

plt.show()

top\_products = df['Itemname'].value\_counts().head(10)

top\_products.plot(kind='bar')

plt.title('Top 10 Most Frequently Purchased Products')

plt.ylabel('Frequency')

plt.xlabel('Product')

plt.xticks(rotation=45, ha='right') # Rotate labels for better readability

plt.show()

# Visualization of Sales Trends

plt.figure(figsize=(15, 6))

df.groupby(['Year', 'Month'])['Total price'].sum().plot(kind='line', color='skyblue')

plt.title('Monthly Sales Over Time')

plt.xlabel('Month-Year')

plt.ylabel('Total Sales')

plt.xticks(rotation=45)

plt.show()

# Top 10 most sold products by quantity

df.groupby('Itemname')['Quantity'].sum().sort\_values(ascending=False)[:10].plot(kind='barh', title='Number of Quantity Sold')

plt.ylabel('Item Name')

plt.xlim(20000, 82000)

plt.show()

## RFM Analysis

df.Date.unique()

# Let's assume that today is 01.01.2012

Today = "2012-01-01"

Today = pd.to\_datetime(Today)

df["Date"] = pd.to\_datetime(df["Date"])

# Recency

recency\_table = df.groupby(["CustomerID"]).agg({"Date": lambda x: ((Today - x.max()).days)})

recency\_table.columns = ["Recency"]

recency\_table.head(5)

# Frequency

frequency\_table = df.drop\_duplicates(subset = "BillNo").groupby(["CustomerID"])[["BillNo"]].count()

frequency\_table.columns = ["Frequency"]

frequency\_table.head(5)

# Monetary

df["Total\_Price"] = df["Quantity"] \* df["Price"]

monetary\_table = df.groupby(["CustomerID"])[["Total\_Price"]].sum()

monetary\_table.columns = ["Monetary"]

monetary\_table.head(5)

rfm = pd.concat([recency\_table, frequency\_table, monetary\_table], axis = 1)

rfm.head(5)

#Scaling the data

from sklearn.preprocessing import StandardScaler

scale = StandardScaler()

rfm\_scaled = scale.fit\_transform(rfm)

inertia = []

from sklearn.cluster import KMeans

for i in np.arange(1,6):

kmeans = KMeans(n\_clusters = i)

kmeans.fit(rfm\_scaled)

inertia.append(kmeans.inertia\_)

plt.figure(figsize = (12,8))

plt.plot(inertia, marker = "o")

plt.title("KMeans - Elbow", fontsize = 18);

kmeans = KMeans(n\_clusters = 3)

kmeans.fit(rfm\_scaled)

rfm["Cluster\_No"] = (kmeans.labels\_ + 1)

from sklearn.cluster import KMeans

# Initialize KMeans with 3 clusters

kmeans = KMeans(n\_clusters=3)

# Fit KMeans on the scaled data

kmeans.fit(rfm\_scaled)

# Assign cluster labels to your original DataFrame

rfm["Cluster\_No"] = kmeans.labels\_ + 1

rfm.head(5)

rfm.groupby(["Cluster\_No"])[["Recency", "Frequency", "Monetary"]].mean()

import matplotlib.pyplot as plt

import seaborn as sns

# Add the cluster labels to your original DataFrame

rfm['Cluster'] = rfm['Cluster\_No']

# Count the number of data points in each cluster

cluster\_counts = rfm['Cluster\_No'].value\_counts().sort\_index()

# Create the plot

plt.figure(figsize=(10, 6))

a = sns.barplot(x=cluster\_counts.index, y=cluster\_counts.values)

plt.title('Number of Data Points in Each Cluster')

plt.xlabel('Cluster Number')

plt.ylabel('Count')

for bars in a.containers:

a.bar\_label(bars)

plt.show()

from sklearn.decomposition import PCA

pca = PCA(n\_components = 2)

pca = pca.fit\_transform(rfm\_scaled)

plt.figure(figsize = (12,8))

plt.scatter(pca[:,0], pca[:,1], c = kmeans.labels\_)

plt.title("Clusters of Customers", fontsize = 20);

data = df.groupby(["BillNo", "Itemname"])[["Quantity"]].sum(

).unstack().reset\_index().fillna(0).set\_index("BillNo")

# Pivot the dataset to get a basket format

data = (df.groupby(['BillNo', 'Itemname'])['Quantity']

.sum().unstack().reset\_index().fillna(0)

.set\_index('BillNo'))

data.head(5)

# Convert quantities to 1 if an item was bought, 0 otherwise

def encode\_units(x):

return 1 if x >= 1 else 0

data\_s = data.applymap(encode\_units)

def num(x):

if x <= 0:

return 0

elif x >=1:

return 1

basket = data.applymap(num)

# Remove the postage column as it is not a product

data\_s.drop('POSTAGE', inplace=True, axis=1, errors='ignore')

data\_s.head()

basket.nunique()

from mlxtend.frequent\_patterns import apriori

# Apply the Apriori algorithm to find frequent itemsets

# Adjust the min\_support value as needed

frequent\_set = apriori(data\_sets, min\_support=0.02, use\_colnames=True)

# Display the frequent itemsets

frequent\_set.head()

from mlxtend.frequent\_patterns import association\_rules

# Generate association rules from the frequent itemsets

# Adjust the min\_threshold for confidence as needed

rules = association\_rules(frequent\_itemsets, metric="confidence", min\_threshold=0.02)

# Display the rules

rules.head(10)