**Market Basket Analysis: Analysis of Association Rules in Different Time Periods**By

Teng Wen Foong

A PROPOSAL

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology

(Kampar Campus)

JUNE 2023

**DECLARATION OF ORIGINALITY**

I declare that this report entitled “**Market Basket Analysis: Analysis of Association Rules in Different Time Periods**” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : 

Name : \_\_\_\_Teng Wen Foong\_\_\_

Date : \_\_\_5/9/2023\_\_\_\_

**ACKNOWLEDGEMENTS**

I would like to express my sincere thanks and appreciation to my supervisor, Dr Kh’ng Xin Yi for invaluable guidance and unwavering support throughout my final year project. Insightful feedback, constructive critiques, and thoughtful suggestions from Dr Kh’ng have consistently propelled me forward, helping me navigate the complexities of my project with confidence. Again, much appreciation for the exceptional mentorship, encouragement, and support. I look forward to continuing to learn and grow under her guidance as I move forward in my academic and professional endeavours.

**ABSTRACT**

Transaction data is a set of records associated with the sales-purchase activities of a

business. As a result of digitalization, the transaction data stored in databases have been

used by the analyst to extract valuable information. The most widely used method to

discover information from transaction data is market basket analysis. Market basket

analysis is a technique that determines the relationship between items and the

relationship is known as the association rule. The association rules could be further

studied to identify commonly purchased items and gain a deeper understanding of

customer purchasing behaviors. In this study, performing market basket analysis is not

the only objective but the variation of association rules over time will also be focused

on. Therefore, market basket analysis will be conducted in different time periods to

study the regular cyclic variation of association rules over time and determine the

interaction between association rules and time. After analysing the trend of the

customers purchasing behaviours from the association rules, the interesting association

relationships between large quantities of business transaction data could assist the

businesses to design marketing strategies, promote cross-selling, and other business

decision-making to boost their profit. An optimal period to conduct market basket

analysis will be recommended to ensure the effectiveness of the association rules is

maximized all the time and allow the businesses to adjust their marketing strategies from time to time.

**Table of Contents**

[CHAPTER 1 Introduction 1](#_Toc144845707)

[1.1 Problem Statement and Motivation 1](#_Toc144845708)

[1.2 Research Objectives 2](#_Toc144845709)

[1.2.1 To conduct market basket analysis in different time periods. 2](#_Toc144845710)

[1.2.2 To analyze the fluctuation of the customer purchasing behaviors with the changes in time. 2](#_Toc144845711)

[1.2.3 To provide actionable recommendations for optimal time period for market basket analysis. 3](#_Toc144845712)

[1.3 Project Scope and Direction 3](#_Toc144845713)

[1.4 Contributions 4](#_Toc144845714)

[1.5 Report Organization 5](#_Toc144845715)

[CHAPTER 2 Literature Reviews 6](#_Toc144845716)

[2.1 Market Basket Analysis 6](#_Toc144845717)

[2.1.1 Association Rule 7](#_Toc144845718)

[2.1.2 Support 8](#_Toc144845719)

[2.1.3 Confidence 8](#_Toc144845720)

[2.1.4 Lift 9](#_Toc144845721)

[2.2 Apriori Algorithm 10](#_Toc144845722)

[2.3 Comparison between Apriori and FP Growth Algorithms 11](#_Toc144845723)

[2.4 Time Variations of the Association Rule 13](#_Toc144845724)

[CHAPTER 3 Proposed Method/Approach 16](#_Toc144845725)

[3.1 Design Specification 16](#_Toc144845726)

[3.1.1 Methodologies and general work procedures 16](#_Toc144845727)

[3.1.2 Resource requirements 19](#_Toc144845728)

[3.1.3 System performance definition 21](#_Toc144845729)

[3.2 System Design / Overview 21](#_Toc144845730)

[3.2.1 Functional modules 22](#_Toc144845731)

[3.2.2 Error handling 28](#_Toc144845732)

[3.3 Implementation Issue and Challenges 28](#_Toc144845733)

[3.4 Timeline 30](#_Toc144845734)

[3.4.1 FYP 1 Timeline 30](#_Toc144845735)

[3.4.2 FYP 2 Timeline 31](#_Toc144845736)

[CHAPTER 4 Preliminary Work 32](#_Toc144845737)

[4.1 Monthly result 32](#_Toc144845738)

[4.2 Biweekly result 45](#_Toc144845739)

[CHAPTER 5 Conclusion 70](#_Toc144845740)

[REFERENCES 71](#_Toc144845741)

[APPENDIX 1](#_Toc144845742)

[POSTER 1](#_Toc144845743)

[PLAGIARISM CHECK RESULT 1](#_Toc144845744)

[FYP 1 CHECKLIST 1](#_Toc144845745)

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure Number** | **Title** | **Page** |
|  |  |  |
| Figure 2.2.1 | Pseudo code for Apriori algorithm | 10 |
| Figure 2.2.2 | Product affinity results | 11 |
| Figure 2.3.1 | Comparison of execution time of Apriori and FP Growth algorithm | 12 |
| Figure 2.4.1 | Monthly development of association rules measurement indicators | 14 |
| Figure 3.1.1 | System flow diagram | 17 |
| Figure 3.2.1.1 | System module diagram | 22 |
| Figure 3.2.1.2 | Structure of the transactional data | 23 |
| Figure 3.2.1.3 | Pre-processing of monthly data | 24 |
| Figure 3.2.1.4 | Pre-processing of biweekly data | 24 |
| Figure 3.2.1.5 | One-hot encoding of the data | 25 |
| Figure 3.2.1.6 | Equation for support | 26 |
| Figure 3.2.1.7 | Equation for confidence | 26 |
| Figure 3.2.1.8 | Equation for lift | 26 |
| Figure 3.2.1.9 | Generation of association rules | 27 |
| Figure 3.2.1.10 | Exporting the results to excel file | 28 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table Number** | **Title** | **Page** |
|  |  |  |
| Table 2.4.1 | Rules for 1st period | 15 |
| Table 2.4.2 | Rules for 2nd period | 15 |
| Table 2.4.3 | Rules for 3rd period | 15 |
| Table 2.4.4 | Rules for 4th period | 15 |
| Table 3.1.2.1 | Hardware specification | 19 |
| Table 4.1.1 | Top 15 highest frequency itemsets in the month of October 2011 | 33 |
| Table 4.1.2 | Product affinity analysis for the month of October 2011 | 35 |
| Table 4.1.3 | Top 15 highest frequency itemsets in the month of November 2011 | 37 |
| Table 4.1.4 | Product affinity analysis for the month of November 2011 | 39 |
| Table 4.1.5 | Top 15 highest frequency itemsets in the month of December 2011 | 41 |
| Table 4.1.6 | Product affinity analysis for the month of December 2011 | 43 |
| Table 4.2.1 | Top 15 highest frequency itemsets in Biweek 1 (9/24/2011 - 10/7/2011) | 46 |
| Table 4.2.2 | Product affinity analysis for Biweek 1 (9/24/2011 - 10/7/2011) | 48 |
| Table 4.2.3 | Top 15 highest frequency itemsets in Biweek 2 (10/8/2011 - 10/21/2011) | 50 |
| Table 4.2.4 | Product affinity analysis for Biweek 2 (10/8/2011 - 10/21/2011) | 52 |
| Table 4.2.5 | Top 15 highest frequency itemsets in Biweek 3 (10/22/2011 - 11/4/2011) | 54 |
| Table 4.2.6 | Product affinity analysis for Biweek 3 (10/22/2011 - 11/4/2011) | 56 |
| Table 4.2.7 | Top 15 highest frequency itemsets in Biweek 4 (11/5/2011 - 11/18/2011) | 58 |
| Table 4.2.8 | Product affinity analysis for Biweek 4 (11/5/2011 - 11/18/2011) | 60 |
| Table 4.2.9 | Top 15 highest frequency itemsets in Biweek 5 (11/19/2011 - 12/2/2011) | 62 |
| Table 4.2.10 | Product affinity analysis for Biweek 5 (11/19/2011 - 12/2/2011) | 64 |
| Table 4.2.11 | Top 15 highest frequency itemsets in Biweek 6 (12/3/2011 - 12/15/2011) | 66 |
| Table 4.2.12 | Product affinity analysis for Biweek 6 (12/3/2011 - 12/15/2011) | 68 |

# CHAPTER 1 Introduction

## 1.1 Problem Statement and Motivation

Most of the business owners have ignored the need of tracking the changes of association rules over time and this circumstance might cause them to miss out on the

adaptions needed to be made in order to retain the effectiveness of the association rules.

The association rules could be affected by the case of certain existing items are removed

or some new items are added to the inventory. This circumstance may lead to customers

start purchasing the new items which will result in a new association rule or the existing

association rule might lose its effectiveness due to the item no longer existing. Nevertheless, customer preferences might change due to introduction of new trend. This could happen when some of the product companies put a lot of effort to advertise their products for better exposure to the public. For instance, release of a scientific report which stated that banana milkshake will benefit human body in many ways will attract more customers to buy banana and milk together. As a result, customers may follow the new trend and change their buying behaviour. Additionally, external factors such as weather will also affect the effectiveness of market basket analysis. Demand of seasonal products like clothing changes over time. For instance, thick and warm clothing such as scarves, gloves, and earmuffs will have higher demand during the winter season. Moreover, competition between businesses in market will also affect the market basket analysis over time. This is due to new competitors joining the market may offer several promotions and discounts for certain product combinations to attract customers and boost their sales. For instance, customers who buy milk and egg together will receive a discount of 20%. Eventually, it will lead to a new association rule and it might not be found if market basket analysis is not conducted periodically. Likewise, external events such as global pandemic could be recognised as one of the key factors that affect market basket analysis over time as well. For instance, the global pandemic COVID-19 which happened in these few years causes the high demand of items that protect human body from the virus. For instance, the frequency of item combination like medical mask and hand sanitiser has increased dramatically compared to other groceries during pandemic. In conclusion, there are several factors require businesses to conduct market basket analysis periodically to guarantee the elimination of insignificant association rules and the discovery of new potential association rules.

By conducting market basket analysis and analysing association rules in different time periods, businesses and retailers are allowed to track changes of customer behavior over time. Subsequently, businesses that are facing a regression of their sales will benefit from the market basket analysis. Additionally, it will help businesses to adapt their marketing strategies by focusing on certain products instead of retaining the same old strategies. Moreover, conducting market basket analysis in different time periods will help the businesses to discover a new trend or opportunities to increase their revenues. Assisting businesses to survive in the market is not only the motivation for conducting the analysis, customers will also benefit from it. For instance, businesses may place items that have association rules together to help customers with reduced mobility tofind them without needing to move around the store to look for them. In conclusion, market basket analysis is a result of the rise of information technology and it could benefit society from both customer and businesses perspectives.

## 1.2 Research Objectives

### 1.2.1 To conduct market basket analysis in different time periods.

Unlike common market basket analysis, the project will conduct a more specific market

basket analysis with an additional factor which is the time period. Afterward, an analysis of transaction data from different time periods such as months, seasons, or years, will be carried out using market basket analysis. The project will be initiated by

conducting a market basket analysis using a dataset consisting of transactions in different time periods to determine the association rules. When conducting MBA, the execution time will be observed to assess the efficiency and performance of the analysis

process. The execution time can vary depending on various factors, such as the size of

the dataset in different time periods, the complexity of the analysis algorithm and the computational resources available for MBA.

### 1.2.2 To analyze the fluctuation of the customer purchasing behaviors with the changes in time.

By comparing the statistics of association rules over different time periods, a deeper understanding and better insight into why and how customer behaviour changes over time will be obtained. The parameters of the association rules such as confidence, lift, and support will be visualized in the forms of graphs, charts, or plots for the observation of which of the values change over time. The hypothesis could be proven by comparing the real transaction data to check if the fluctuation of the transaction frequency. Next, the variation of customer purchasing behaviours with the changes in time will be analysed to understand how they change and identify if any external factors that may be influencing these changes. Customer purchasing behaviours could change in several ways such as shifting of product preferences, fluctuation of purchasing amount, and so on.

### 1.2.3 To provide actionable recommendations for optimal time period for market basket analysis.

Product recommendations will be provided to the businesses, along with the justifications, for tracking and analyzing the changes in customer behaviour more effectively. The insights gained from the project will benefit businesses to adjust their marketing strategies from time to time (e.g. product placement, and pricing of the products) after understanding how the association rules variate over time. Nevertheless, recommendations for cross-selling and up-selling could be made by considering the external factors affecting the fluctuation of customer purchasing behaviours such as the frequency of customer purchases, length of time between purchases, and the seasonality of product demand. An optimal time period for conducting market basket analysis will guide the businesses to maximize the effectiveness of their marketing strategies from time to time. Eventually, the revenue of the businesses and customer satisfaction will be improved.

## 1.3 Project Scope and Direction

The project will focus on the analysis of association rules in different time periods using market basket analysis. Market basket analysis will be conducted using a dataset of customer transactions which are divided into different time periods such as daily, weekly, monthly, and even annually. First of all, the data collected from the retail store will undergo a data preprocessing process which involves data cleaning and preparation of the data for analysis later on. In this phase, the data will be checked for completeness and consistency to guarantee quality of data. Any missing values will be input, and the outliers will be removed. Subsequently, market basket analysis will be performed using the prepared data to discover association rules in different time periods. The discovered association rules will then undergo a filtering process which will disqualify those that do not meet the minimum requirement in terms of the value of the important parameters in market basket analysis. The association rules which fulfill the requirements will be visualized in the form of graphs, charts, or tables using their parameters such as lift, support, and confidence. By presenting the data visualization, changes of customer behaviour could be observed based on the variation of the values in different time periods. Next, data interpretation will focus on the time period with the most significant variation of values to figure out the factors affecting the association rules. Further analysis will be carried out to investigate the best time period to conduct market basket analysis in order to track changes of association rules. Eventually, a recommendation with detailed justification will be provided for the businesses to refer for improving their product sales and boosting their revenues. However, there might be some constraints when carrying out the project. For instance, the large size of the dataset which included transaction data in different time periods might be constrained by an available source such as time. The deliverables of the project are association rules for different time periods, analysis of trends and patterns in association rules over time, analysis of factors affecting the variation of association rules, and recommendations to conduct market basket analysis in terms of the time period to adjust marketing strategies and improve sales.

## 1.4 Contributions

The main beneficiary of the project will be the businesses. The market basket analysis to be conducted in different time periods will assist them to make more strategic business decisions. First of all, businesses will gain better insight into the placement of their products in stores. The association rules from market basket analysis help the businesses to identify the popular combination of products and place them together. This will allow customers to find their desired products more conveniently and encourage them to make additional purchase to boost sales. Moreover, a new pricing strategy may be discovered as a result of the project. For instance, the businesses could offer a bundled promotion for the products involved in the association rules to attract customers to buy the products together. By conducting the analysis, businesses could improve their inventory management by stocking up on products that are frequently purchased together, reducing the possibility of stockouts to maximize profits. Last but not least, businesses will be able to provide a better customer experience by recommending products that are frequently purchased together, based on the association rules. The customers might accept the recommended products that they did not initially think of, eventually boosting the product sales.

## 1.5 Report Organization

The report is organized into five distinct chapters, which are introduction, literature review, proposed method / approach, preliminary work, and conclusion. The first chapter outlines the project's aims, importance, and context, which could be the foundation for the upcoming chapters by determining research inquiries and objectives. The second chapter, literature review will discuss the related work and past researches, then study further based on the works done by others. This section will address limitations of each work and the limitations will be taken into account in current project to resolve them. Subsequently, the following chapter, proposed method / approach will explain the methodology in detailed, elucidating how the project harnesses market basket analysis across various time spans to pinpoint the most efficient period for association rule mining. This chapter would also cover the algorithms to be used, tools, and data processing techniques deployed. To offer practical insights, the chapter of preliminary work investigates the operations that will be performed in the project, including data gathering, preprocessing, and initial findings, illustrating progress and laying the groundwork for the conclusive chapter. Last but not least, the conclusion chapter summarizes project outcomes, underscoring pivotal discoveries, trends, and implications arising from the association rule mining analysis. It also reflects on the importance of the optimal time period findings and explores potential avenues for future research and practical implementation.

# CHAPTER 2 Literature Reviews

## 2.1 Market Basket Analysis

Market basket analysis is known as a data mining technique that analyses the patterns of co-occurrence of certain products and investigates the relationship and strength of the link between the products that have been purchased together by the customers. The main objective of market basket analysis is to identify the purchasing behaviour of the customers in order to help retailers to understand better which are the popular products and try to get more stocks of those certain items in the inventory. In addition, retailers could find out the best strategic location to display their products by placing the products that are being bought together frequently in the same area. By having a more effective product placement, retailers could boost their sales by a significant amount. The concept of market basket analysis could be interpreted using an example from daily life. When people buy a loaf of bread, they tend to buy fruit jam as well. The relationship between bread and jam is recognised as an if-then conditional statement which implies that if bread is bought, then the jam will be bought with bread. Market basket analysis can be categorised into three types which are descriptive market basket analysis, predictive market basket analysis, and differential market basket analysis. This type of market basket analysis is done by analysing historical data and identifying the relationship between the products using statistical techniques. In fact, there will be no prediction on the future trend of the buying behaviour of customers will be made and this kind of analysis is referred as an unsupervised technique. Next, predictive market basket analysis works reversely with descriptive market basket analysis. It can be interpreted by its literal meaning “predictive”, analysis of historical and current existing data will be carried out before predicting the future trend. This kind of analysis makes use of supervised machine learning models such as classification and regression to obtain accurate predictions based on the provided data. However, this approach is less commonly to be used when compared to descriptive market basket analysis. The last type of market basket analysis is differential market basket analysis. This type of analysis could help retailers to understand why customers will purchase a product from a certain store even though there are other stores with similar product prices and quality. In this analysis, the customer decisions will be studied based on several factors such as seasons, time variations, and even delivery time and user experience for that online shopping on e-commerce platforms [1].

### 2.1.1 Association Rule

Association rule is the core concept of market basket analysis which focused on determining insignificant and hidden relationships between two items in a large data set. The data set to be analysed consists of customer transaction activity and one of the sources for the data is supermarkets. The transaction activity data will list the items that have been bought by customers in a single transaction. Analogously, association rule is similar to market basket analysis such that association rule is also categorised into four types which are multi-relational association rules, generalised association rule, interval information association rule, and qualitative association rule. Multi-relational association rule (MRAR) is derived from multiple different relational databases. Each rule under this type will contain an entity that represents the indirect relationships between entities by having different relationships. The generalised association rule is the most commonly used association rule in data mining to get a fundamental idea about the hidden patterns in data [2]. Kristiana et al. in 2020 have highlighted that the idea of the association rule is to identify all the possible connections between items and select the most frequent combination of items that will most likely be indicators of dependence [3]. The statement of association rules is developed in two stages as proposed by Rusawati, Gufroni, and Rianto in 2018 [4]. The initial stage to develop association rules is analysing patterns of high frequency. In this phase, the combination of items that meet the minimum support value that has been determined will be figured out. The association rule is supported by a mathematical model. In the application of the model, different products are recognised as objects in a set, I = {i1, i2, i3, i4, … …, in}. Each transaction is expressed as T and it corresponds to objects in the set I, which could be defined as T ⊂ I. Assuming that A and B are the object sets that fulfill the conditions A ⊂ I, B ⊂ I, and A∩B = Ø, then the circumstance has achieved the association rule which can be expressed as A ⇒ B [5]. Based on the definition of the association rule, A is the antecedent (if) and B is recognised as the consequent (then). After determining all the high-frequency item sets, the development of the association rule proceeds to the next stage which is association rules establishment. The association rule, A ⇒ B is only considered as confident after it meets the minimum requirement of confidence value. The terms, “support” and “confidence” will be further discussed in the following subchapters.

### 2.1.2 Support

Support is one of the parameters that defined the market basket analysis. Support is the percentage of the combination of certain items in a large amount of data. The support value shows how frequently the combination of items appears in transaction data. In fact, association rules with a higher support value will be preferred due to they are more likely to show the pattern in large transaction data. The support value is crucial for the first stage of developing association rules which have been mentioned earlier. The high-frequency analysis will only select the combination that reaches the minimum support value. Nevertheless, support could also be referred to as a probability. For instance, if bread is having a supporting value of 20 out of 100 transactions, then it indicates that the probability of the bread being sold is 0.2. In terms of mathematics, the supporting value could be obtained through the formula of dividing the amount of the transaction that contains the specified item by total transactions and multiplying the value by a hundred percent. The mathematical formula can be expressed as the following shows:

where:  
S = Support value  
∑ (𝑇𝑎 + 𝑇b) = Transactions that contains item A (antecedent) and item B (consequent)∑(𝑇) = Total transactions

### 2.1.3 Confidence

Confidence is another important measure for market basket analysis. It works similarly to support but it is more focused on conditional probability. The confidence value emphasises the strength of relationships between two certain items in association rules. For instance, the analyst could obtain the confidence value by studying how frequent the occurrence of the combination of bread and jam is among the total frequency of bread. Subsequently, the confidence value will be applied in the second stage of developing association rules which are establishing association rules by disqualifying the combinations that do not meet minimum confidence requirements. In 2018, Sagin and Ayvaz have developed association rules mining with market basket analysis for a large hardware company operating in the retailed sector [6]. The project aimed to identify product groups that tended to be sold together and eventually they figured out the rule of {Chisels and Cutters, Bits Tips Adapters} → {Drill Bits} with a 64% confidence value which is considered high. The company was totally unaware of the hidden rule before the project is being carried out . In terms of mathematics, confidence value has applied conditional probability and it is expressed as the following formula:

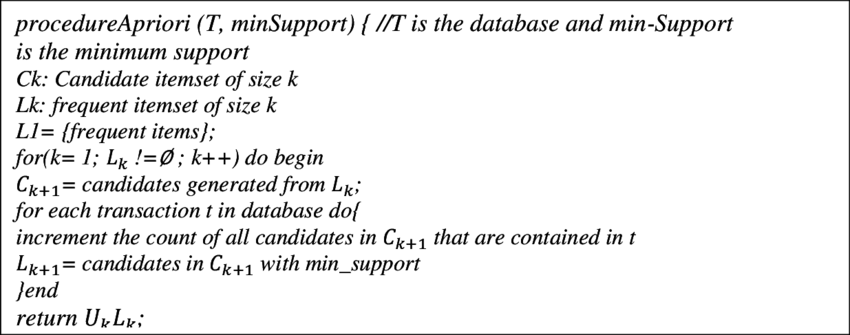
where:  
C = Confidence value  
∑ (𝑇𝑎 + 𝑇b) = Transactions that contains item A (antecedent) and item B (consequent)∑(𝑇) = Total transaction that contains A

### 2.1.4 Lift

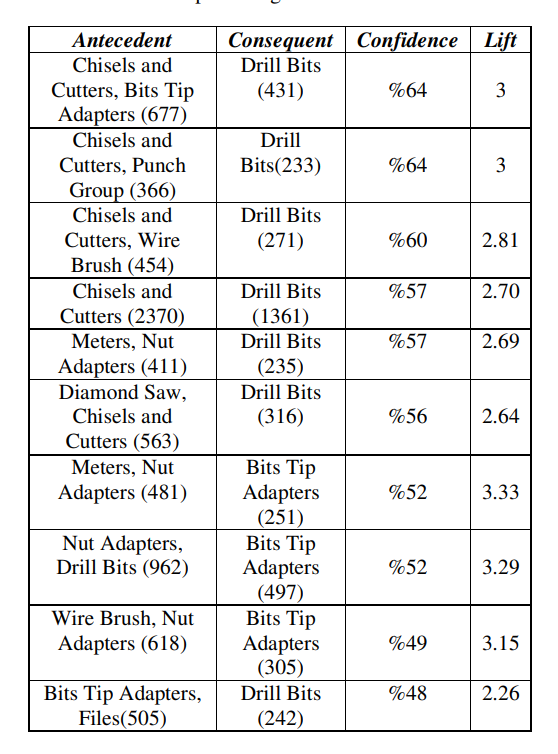
In market basket analysis, a lift is a metric that is used to measure the strength of the relationship between two items as a combination. Lift is calculated by dividing the confidence percent by the support percent. A lift value of 1 indicates that the association rule does not have any effect on the outcome which means that there is completely no association between the two items. If the value of lift is greater than 1, it implies a positive association which shows the association rule enhances the chances of the outcome, while a lift value lower than 1 corresponds to the situation of a negative association. A higher lift value indicates a higher possibility of the presence of one item in a transaction coming with another item in the same transaction. The mathematical formula for calculating lift is shown below:

## 2.2 Apriori Algorithm

The Apriori algorithm is the first and most widely used data mining technique that was proposed by R. Agrawal and R. Srikant in 1994 for frequent itemset mining [7]. It is a bottom-up approach that uses a level-wise search for identifying the frequent item sets. There are two main steps in the Apriori algorithm which are the join step and the prune step. The join step generates (k+1)-itemsets from k-itemsets by joining each item with itself while the prune step determines the frequent itemsets that meet minimum support and disqualify those itemsets that are recognised as infrequent. In the first iteration of the algorithm, a 1-itemsets candidate is found by scanning the database in which the occurrence of the itemset is satisfying the minimum support threshold. Next, 2-itemsets that meet minimum support are determined using previous frequent 1-itemsets and this process will continue until frequent k-itemsets are found [8]. In 2021, Nurmayanti et.al. applied the Apriori algorithm to determine association rules between several sales attributes based on the transaction data of outdoor sales of goods [9]. As a result, they have found that the transaction patterns in the sale of outdoor goods that are generated with the algorithm are as many as 10 rules and the strength of the rules is supported by 0.296 minimum support, 0.774 confidence, and 1.49 lift value. They could assume that customers who buy Portuguese stove items will also be likely to purchase portable gas items. The pseudocode for the Apriori algorithm is shown in the figure below [10]:

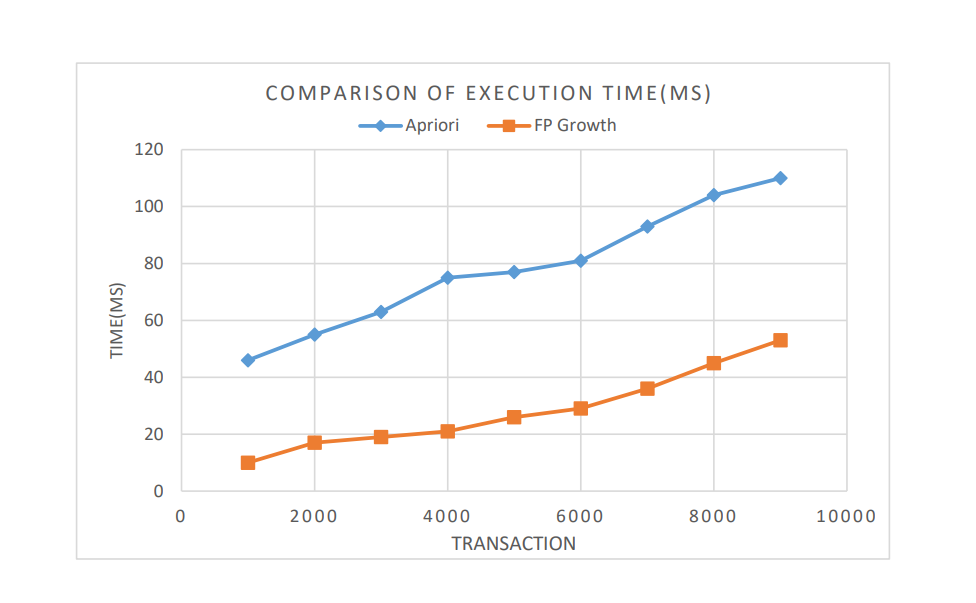
  
Figure 2.2.1. Pseudo code for Apriori algorithm [10]

In 2018, Ayvaz and Sagin conducted market basket analysis to determine association rules for a large hardware company by using Apriori algorithm [6]. As a result, they successfully develop the first rule and it is shown in the figure below. They found that there are 431 out of 667 transactions containing the combination of {Chisels and Cutters, Bits Tips Adapters} also contain product {Drill Bits}. This rule indicates that customers who purchase the group products {Chisels and Cutters} and {Bits Tips Adapters} are having the 64% probability of buying {Drill Bits} as well.

  
Figure 2.2.2. Product affinity results [6]

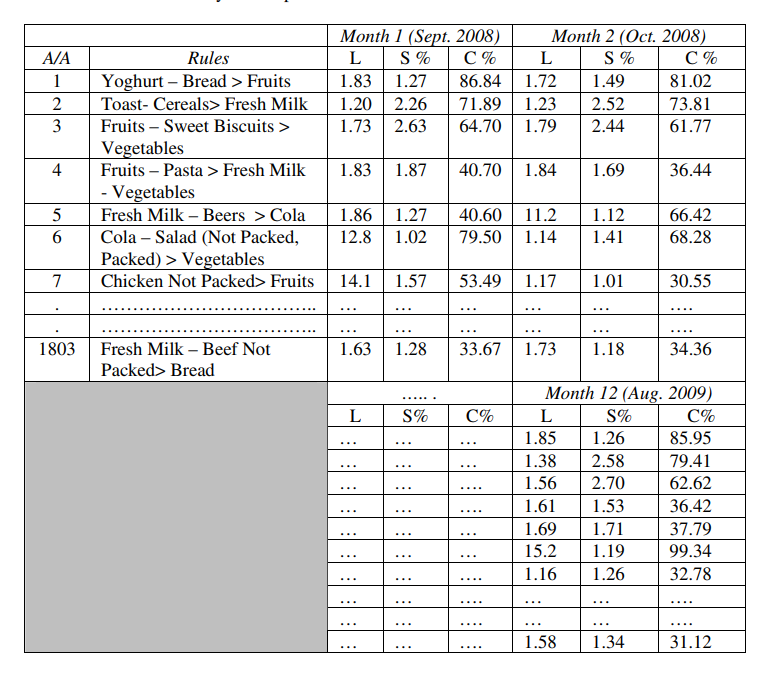
## 2.3 Comparison between Apriori and FP Growth Algorithms

The Frequent Pattern Growth (FP Growth) algorithm is an alternative data mining algorithm for market basket analysis. The algorithm could be recognised as an improved version of the Apriori algorithm since it overcomes the flaws of the Apriori algorithm by eliminating the join and prune steps. The first step of the algorithm is identical to the first step of Apriori, which is scanning the database to obtain the frequent itemsets. The non-frequent itemsets are filtered out based on the minimum support and the remaining itemsets are listed in order of size. Next, a tree with only a root is created and the frequent itemsets are added as nodes into the tree one by one. The construction of an FP-Tree is considered complete once all the frequent itemsets have been added to the tree. Now, it is much faster to get information on the most frequent itemsets by traversing the tree [11]. There are several differences between the Apriori algorithm and the FP Growth algorithm. Regarding the generation of frequent patterns, Apriori obtains itemsets by pairing such as single itemset, double itemset, and triple itemset, while FP Growth constructs an FP-Tree. In addition, Apriori uses frequent itemsets to extend frequent subsets one item at a time whereas FP Growth constructs conditional FP-Tree for every item in the database. In the aspect of time consumption, Apriori takes longer time compared to FP Growth since it scans the database at each step. In contrast, FP Growth scans the database only once at the beginning so it saves more time. In the context of space, Apriori saves a converted version of the database in memory, while FP Growth saves the set of conditional FP-tree for each item in the memory. Last but not least, Apriori applies bread-first search to search for items while FP Growth uses depth-first search due to its data structure [12]. In 2021, Aldino et al. carried out a market basket analysis to determine customer purchasing patterns using FP Growth and Apriori algorithm [13]. They found that FP Growth had a faster running time of 6 seconds which is 5 times shorter than Apriori which required 30 seconds. The last conclusion drawn from their studies is that FP Growth generates more than 19 association rules while Apriori only produces 6 rules. Hossain et.al conducted market basket analysis based on a bakery shop dataset using Apriori and FP Growth algorithms respectively [14]. They found that the required time for the FP Growth algorithm is much less than that for Apriori algorithm. The graph of comparison of execution time is shown as below:

  
Figure 2.3.1. Comparison of execution time of Apriori and FP Growth algorithms [14]

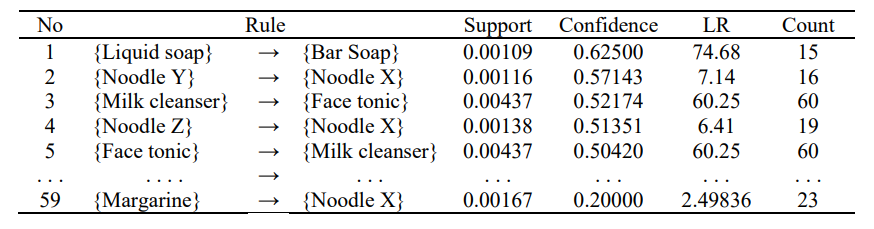
## 2.4 Time Variations of the Association Rule

Time variations of the association rule are referred to as the possible changes in the association rule over time. Customer behaviour may be different due to several factors such as market trends, product availability, and even changes in environmental issues. For instance, people are buying warm and thick clothes to wear during winter and looking for light and thin clothes in summer. Consequently, it is crucial to detect any significant changes in association rules over time to monitor customer behaviour and adjust the marketing strategy accordingly. Computing association rules based on monthly transaction data will help the analyst to recognise seasonal variation where some of the rules indicate a pattern in approximately the same month each year. Analogously, different time variations such as hourly, daily, weekly, and even annually might dig out some hidden association rules. In 2019, Ansari did a study that analyses the trend of association rules in different time periods. He worked on transactional data of a Belgian retail company and analysed the results that will assist the company to adjust time period-specific marketing strategies to boost the revenue of the company [15]. He also stated that most of the retail stores are prioritizing what the customers will purchase but they have ignored another crucial factor that is affecting their behaviours. In fact, they should not only focus on the popular combinations of items but also pay attention to the fact about when they buy them. Ansari also presented the statement of the need of focusing on the problem that shows regular variations in association rules over time. In his studies, he applied 3 months periods to discover top association rules and compare them at different time periods which will allow the company to perform up-selling and cross-selling in different time lags and maximise the profit over time. In 2011, Papavasileiou and Tsadiras presented how variations of association rules occur over time by using market basket analysis [16]. They studied the connection between variability of association rules with changes in purchasing habit over time. Annual data with 12 subsets of data corresponding to each month of a year was collected for analysis purpose and a significant change was observed during first two months. The variation of association rules was determined by three key factors: lift, support, and confidence. Based on the figure below, the table shows that rule 5, 6, and 7 were having a massive variation in terms of confidence. Rule 5 was having a growing confidence while confidence of rules 6 and 7 were decreasing.

  
Figure 2.4.1 Monthly Development of Association Rules Measurement Indicators [16].

Alfiqra and Khasanah in 2020 developed a model to implement market basket analysis based on Overall Variability of Association Rule (OCVR) on product marketing strategy [17]. Their study used monthly transaction data of a retail store in Yogyakarta to produce rules for four periods, which corresponded to weeks. As a result, they found 59 rules for the first period, 48 rules for the second period, 54 rules for the third period, and 58 rules for the last period. Based on the following figures, it can be observed that rules obtained from each period had a slight difference while the attributes in terms of support, confidence, and lift were changing over the four periods. This indicated that customer behaviour changed over time, hence, marketing strategy should be adjusted from time to time to maximize the benefits from market basket analysis.

Table 2.4.1 Rules for 1st period [17]

  
Table 2.4.2. Rules for 2nd period [17]

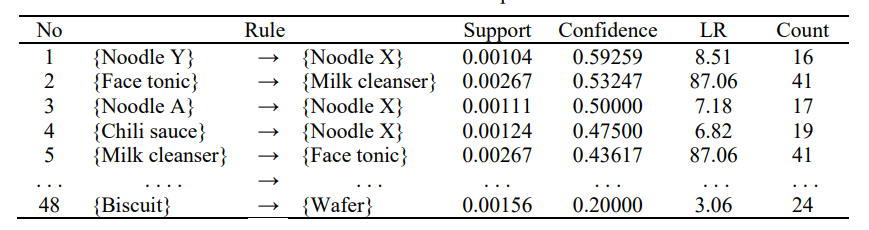
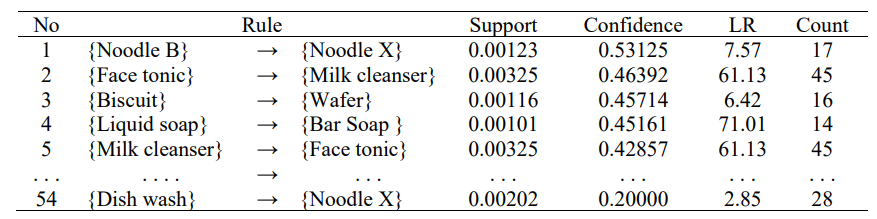
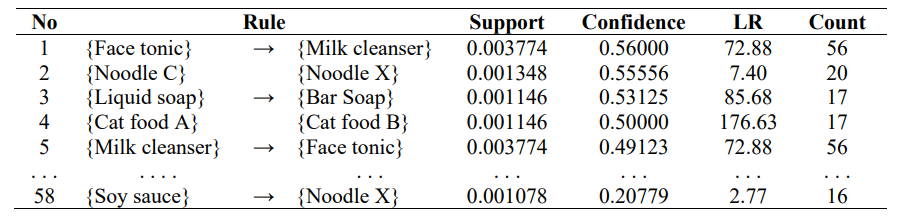
  
Table 2.4.3. Rules for 3rd period [17]  


Table 2.4.4. Rules for 4th period [17]  


# CHAPTER 3 Proposed Method/Approach

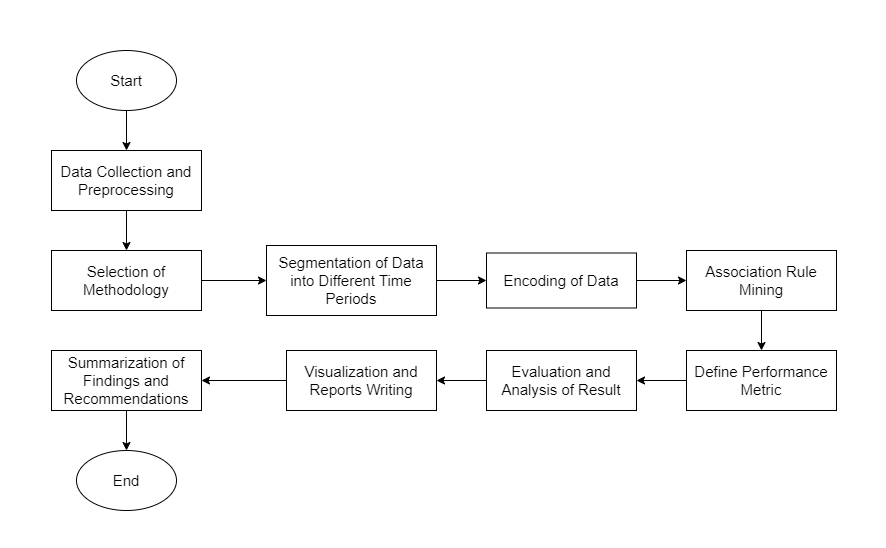
## 3.1 Design Specification

Within this section, an intricate layout and the essential technical prerequisites of the project's resolution will be expounded. Comprehensive direction for crafting the solution such as encompassing the solution's architecture, constituents, data progression, and other intricate technical facets will be provided. This meticulous guidance aims to preempt any potential misinterpretations that could arise from ambiguities in requirements. Moreover, the design specification serves to guarantee that project advancement remains in alignment with its objectives, design principles, coding conventions, and other established norms. The subchapter spans three primary aspects: methodologies and general work procedures, resource requirements, and system performance definition.

### 3.1.1 Methodologies and general work procedures

The project methodology will involve the utilization of both the Apriori and FP-Growth algorithms for conducting association rule mining. Although these algorithms yield identical association rule outputs, however, their performance varies in terms of processing time. Therefore, a comparative analysis of their processing times will be conducted to observe the impact of this distinction. The Apriori algorithm is a traditional association rule mining technique which its main objective is to mine frequent itemsets within transactional datasets. It operates iteratively by generating candidate itemsets and pruning those that do not meet a defined minimum support threshold. However, the algorithm's repeated data scans and candidate generation can render it computationally intensive for sizable datasets.

In contrast, the FP-Growth algorithm is a sophisticated association rule mining technique designed to overcome Apriori's limitations such as its multiple dataset scans and candidate generation stages. FP-Growth constructs a compact data structure named the FP-tree which encapsulates item frequency and relationships within the dataset. This construction obviates the necessity for candidate generation. Instead, the algorithm recursively mines frequent itemsets directly from the FP-tree which is considered curtailing computational overhead. FP-Growth excels with datasets featuring numerous transactions and items due to it necessitates fewer data passes and eliminates the need to generate an extensive array of candidates. This attribute renders FP-Growth an influential technique for scalable association rule mining.

  
Figure 3.1.1 System flow diagram

The initial step of the project involves initiating data collection and preprocessing procedures. The collection phase entails procuring transactional data essential for conducting market basket analysis. This dataset will encompass records of customer purchases and unveils valuable insights into customer buying behaviors. Possible sources of this data include point-of-sale systems, e-commerce platforms, customer databases, and loyalty programs. To be considered valid, transactional data must encompass key attributes such as invoice number, item stock code, item description, quantity, invoice date, and unit price. The preprocessing phase ensues once the raw data is amassed. To facilitate subsequent association rule mining, data transformation into an appropriate format is requisite and integration becomes necessary if the data originates from diverse sources. Essential data-cleaning procedures are undertaken to expunge duplicates, and irrelevant entries, and rectify missing values through imputation or elimination.

The project then proceeds to segment the data into distinct time periods to align with the project's objectives. The specified intervals consist of monthly and biweekly spans aimed at extracting meaningful insights. The dataset is organized chronologically based on the invoice date timestamps. Due to memory constraints, only the three consecutive months displaying the highest transaction frequency will be selected. Transactions within these months are grouped, and transactions sharing the same invoice number are considered a single transaction. Following the preparation of monthly transactions, biweekly transactions are derived from the preprocessed monthly data. The prepared data then undergoes encoding. Data encoding is a critical preprocessing step that transforms categorical data into a numerical format, amenable to algorithmic analysis. For market basket analysis, this involves converting item names into binary representations indicating item occurrences in transactions. In this project, one-hot encoding will be employed to ensure each unique item is allocated a binary column in the dataset. Transactions form rows in such a way that "1" denotes an item's presence and "0" its absence. This conversion enhances data processing efficiency for algorithms like Apriori and facilitates the discovery of frequent item sets and association rules.

With the encoded data prepared, the algorithms can be employed for association rule mining. This method unveils connections between items frequently co-occurring in transactions and reveals insights into customer buying patterns. Frequent item sets are identified, and association rules are deduced from these sets to depict conditional relationships between items. Each rule comprises an antecedent (premise) and a consequent (outcome) item set and they are accompanied by metrics such as support, confidence, and lift that denote the rule's strength and significance. Various metrics are utilized to gauge rule quality and relevance. Support, lift, and confidence metrics will be used to assess the reliability of item relationships, hence, guiding informed decision-making. A judicious minimum support threshold is defined to sieve out insignificant rules.

Subsequently, an evaluation and analysis of the association rules is conducted to extract meaningful insights. This entails scrutinizing the discovered rules and singling out high-confidence and high-lift rules that illuminate customer behavior. A cross-comparison of rules across different months and biweekly periods is executed to pinpoint trends and shifts in customer preferences. Visual aids such as bar charts serve to effectively communicate patterns and trends embedded in the identified rules. The ensuing report employs succinct yet informative narratives to expound upon rule significance, and spotlight significant findings, temporal changes, and potential implications for business strategies. The report's coherence ensures comprehension among both technical and non-technical stakeholders, empowering them to make business-influencing decisions rooted in mined association rules.

The findings are subsequently distilled into concise, actionable takeaways, summarizing complex insights. They synthesize pivotal patterns, trends, and variations identified through rule mining, accentuating their potential impact on business objectives. In this analysis, insightful recommendations are formulated and suggestions in terms of strategies for marketing campaigns, inventory management, or customer segmentation will be recommended. This summary streamlines comprehension of the analysis results and facilitates strategic decision-making guided by data-driven insights. This approach optimizes resource allocation, augments business performance, and enhances overall operational efficiency.

### 3.1.2 Resource requirements

Resource requirements pertain to the distinct tools, materials, technologies, equipment, and personnel indispensable for the proficient execution of a project or undertaking. These resources play a pivotal role in facilitating diverse tasks and realizing the objectives outlined in the project blueprint. Encompassing both concrete elements such as hardware and intangible assets like software, resource requirements hold paramount importance. Effectively recognizing and overseeing these requisites assume a crucial role in guaranteeing the seamless progression of a project, the attainment of its set goals, and its successful culmination within the pre-established boundaries of scope and timeline.

Hardware:

Table 3.1.2.1 Hardware specification

|  |  |
| --- | --- |
| Operating system | Windows 10 64-bit |
| Processor | Intel(R) Core (TM) i5-8265U CPU @ 1.60GHz 1.80 GHz |
| RAM | 12 GB |
| Graphic card | NVIDIA GeForce MX250 |

The specified hardware configuration of a laptop, featuring 12 GB of RAM, an Intel i5-8265U CPU, and an NVIDIA GeForce MX250 graphics card enables the establishment of a solid groundwork for a diverse array of data analysis operations and computational undertakings. The 12 GB RAM allocation furnishes ample memory capacity which could effectively manage moderately sized datasets and intricate calculations with notable efficiency. This aspect is particularly advantageous when engaged in association rule mining for market basket analysis where memory-intensive algorithms proficiently process expansive transactional datasets without encountering performance hindrances. The Intel i5-8265U CPU, hailing from Intel's 8th generation and possessing four cores achieves an equilibrium between performance prowess and energy efficiency. Boasting a base clock speed of approximately 1.6 GHz makes it capable of managing an assortment of computational responsibilities, encompassing data preprocessing, algorithmic execution, and visualization rendering. Meanwhile, although the NVIDIA GeForce MX250 graphics card is not classified as top-tier, however, it still furnishes substantial graphical processing potency for a laptop. This dedicated GPU significantly elevates the efficiency of data visualization tasks, facilitating smoother rendering of graphs, charts, and intricate visual depictions of association rules.

Software

1. Jupyter Notebook

Jupyter Notebook, an integrated development environment (IDE) is chosen due to its ability to present an interactive and collaborative platform for seamless code development, documentation, and execution. The heart of this software stack is Python which is a versatile and extensively employed programming language. Python provides an extensive array of libraries catering to data manipulation, analysis, and visualization needs, constituting the foundational toolkit.

2. Seaborn, Matplotlib, and Pandas

The integration of libraries such as Seaborn, Matplotlib, and Pandas elevates the analytical process. Seaborn and Matplotlib facilitate data visualization by enabling the crafting of insightful plots and graphs that effectively illustrate association rules and trends in transactional data. Pandas is a robust data manipulation library that could be used to simplify the preprocessing of data using the functions provided. This streamlines the cleansing, transformation, and structuring of transactional data before it is channeled into the association rule mining algorithms.

3. Apriori and FP-Growth

In terms of algorithms, both Apriori and FP-Growth algorithms are used in the project. Apriori excels in unearthing association rules through the identification of frequent itemsets whereas FP-Growth's forte lies in efficiently managing extensive datasets through a tree-based structure. These algorithms are harnessed within the Python environment, and analyze transactional data, eventually unveiling meaningful patterns, and interrelations between items. This analysis stands as the focal point of the project's core endeavors.

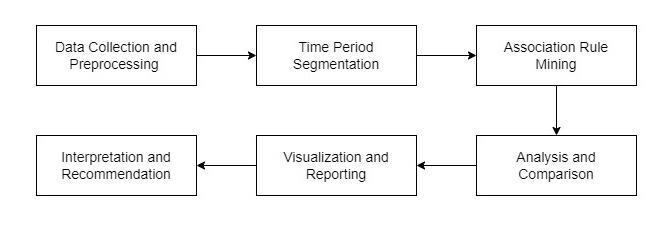
### 3.1.3 System performance definition

The definition of system performance within the project revolves around the precision and caliber of the association rules unearthed throughout the analysis. These rules must accurately portray meaningful patterns within customer purchasing behavior across distinct periods. Their significance, pertinence, and practical applicability stand as pivotal determinants of project success. Additionally, the efficacy of the Apriori and FP-Growth algorithms holds utmost importance. These algorithms drive the association rule mining process that might influence the project's capability to manage voluminous transactional datasets especially when dealing with different segmented time intervals. Equally vital is the efficiency aspect of system performance. This encompasses the computational efficiency of the algorithms to ensure the analysis's completion within reasonable timeframes. Given the potential complexity of item combinations to explore, the algorithms' speed profoundly impacts the project's feasibility. Furthermore, the judicious allocation of system resources, such as memory and processing prowess, fosters seamless execution and averts resource depletion during the analysis phase.

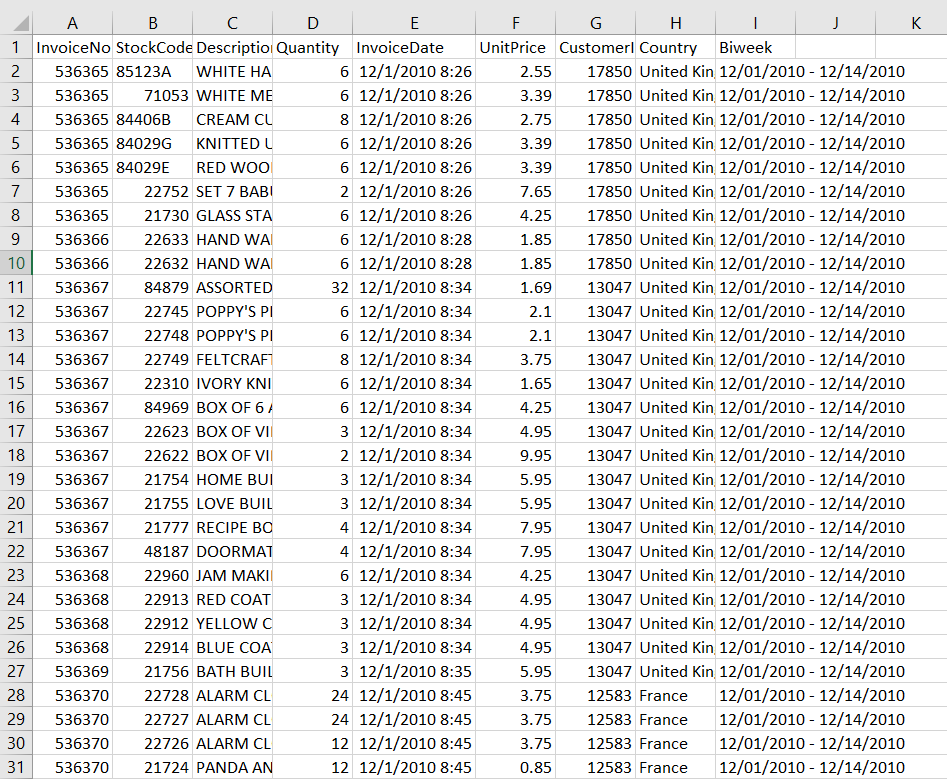
## 3.2 System Design / Overview

This subchapter will outline a detailed and methodical portrayal of the project's architecture, constituents, and interconnections will be presented. This presentation offers a high-level perspective of the amalgamation of diverse elements to attain the project's aims. The overview of the system design functions as a blueprint, elucidating the configuration of functional modules and the protocols for managing errors.

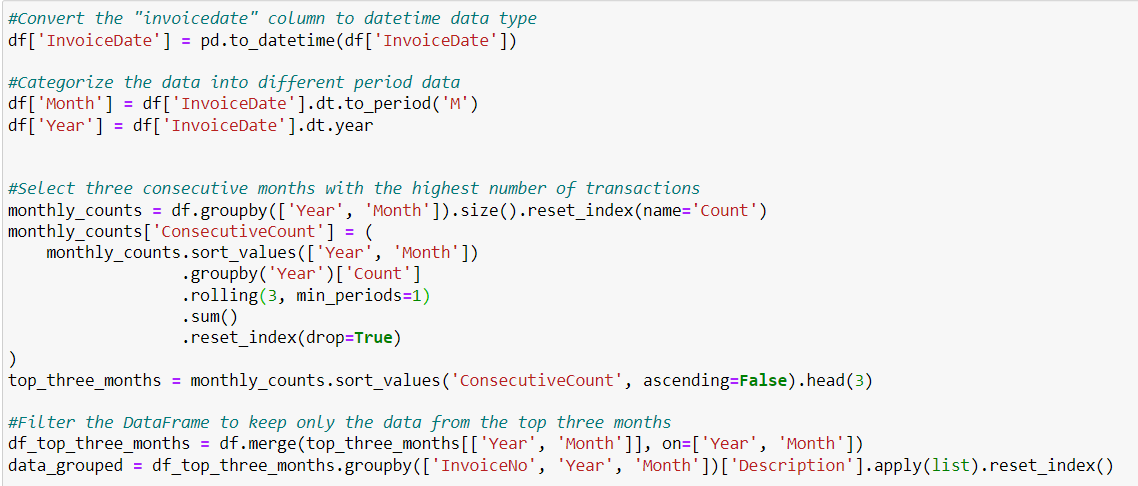
### 3.2.1 Functional modules

****Figure 3.2.1.1 System module diagram

Primarily, the data collection and pre-processing module will concentrate on acquiring raw transactional data from pertinent sources such as retail databases, e-commerce platforms, or loyalty programs. This phase encompasses data extraction, refinement, transformation, and potentially integrating data from diverse origins. This module guarantees the precision, uniformity, and structured formatting of transactional data, rendering it suitable for subsequent analytical phases. The dataset to be used is a transactional dataset which contains all the transactions occurring between December of 2010 and 2011 and the source of the dataset originated from a UK-based and registered non-store online retail. The dataset contains a total of 541909 rows and 9 columns which correspond to invoice number, stock code, description of product, quantity, invoice date, customer code, country, and biweek.

  
Figure 3.2.1.2 Structure of the transactional data

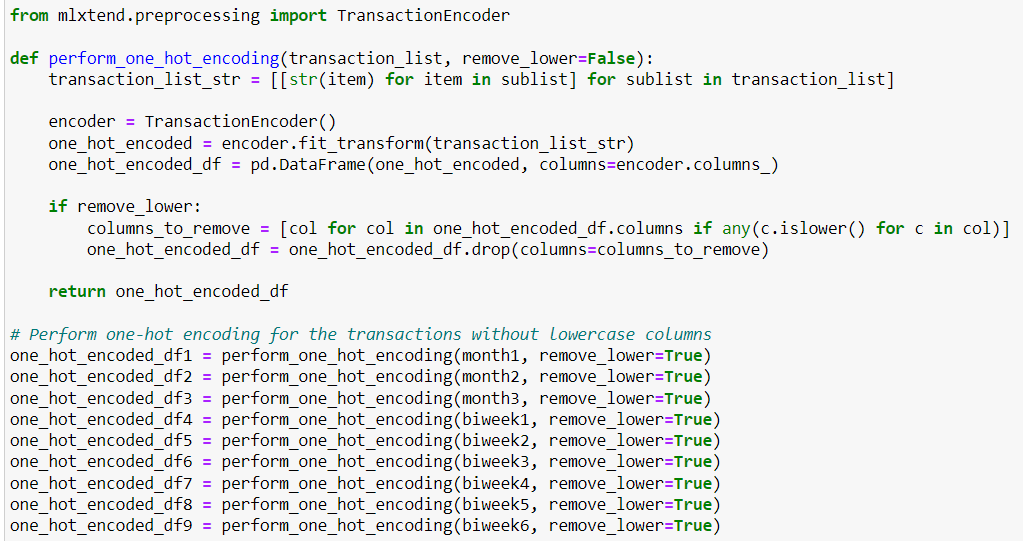
For the pre-processing steps, the ‘InvoiceDate’ column is converted to a datetime data type to perform further date-based operations. To discover the most significant association rules, three consecutive months with the highest number of transactions are selected for the monthly MBA. The DataFrame ‘df’ is firstly grouped by both 'Year' and 'Month' and counts the number of transactions in each month. The result is stored in the ‘monthly\_counts’ DataFrame and the cumulative sum of transaction counts for every three consecutive months for each year is then calculated by using the ‘rolling’ method with a window size of 3. Eventually, the top three consecutive months with the highest cumulative transaction counts are identified and selected using the ‘ConsecutiveCount' column. The DataFrame ‘df’ is further filtered into another DataFrame, ‘top\_three\_months’ that stores information about the top three consecutive months with the highest transaction counts which include the ‘Year’ and ‘Month’ to retain only the rows that match the 'Year' and 'Month' combinations present in the recognized top three months. After filtering, the ‘df\_top\_three\_months’ DataFrame is grouped by ‘InvoiceNo’, ‘Year’, and ‘Month’. The ‘list’ function is applied to the 'Description' column within each group for effectively aggregating all descriptions (items) associated with each invoice within the same month-year combination. The resulting DataFrame, ‘data\_grouped’ now contains information about each invoice in the top three months and the 'Description' column stores a list of items in that invoice.

  
Figure 3.2.1.3 Pre-processing of monthly data

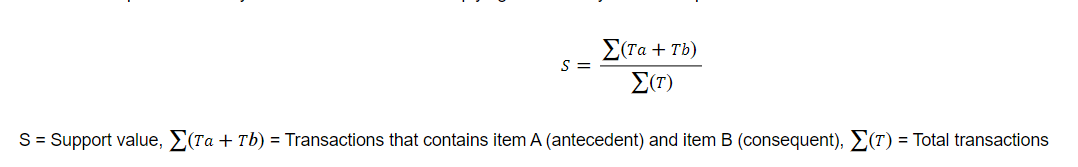
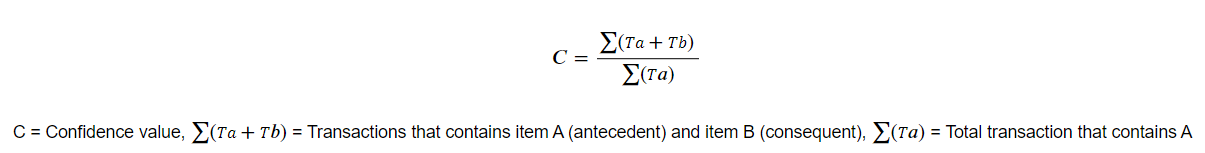
Regarding the pre-processing of biweekly data, the transactional data for biweekly MBA is selected from the top three consecutive months with the highest cumulative transaction counts which were identified previously. To realize the preparation of the biweekly data, the ‘df\_top\_three\_months’ DataFrame is grouped by 'InvoiceNo' and 'Biweek', and the ‘list’ function is applied for each unique combination of 'InvoiceNo' and 'Biweek’ to the 'Description' column. The purpose of that operation is to aggregate all the descriptions (items) associated with each invoice within the same biweek. After applying the aggregation function, it resets the index of the resulting DataFrame, creating a new DataFrame named ‘biweekly\_data\_grouped’.

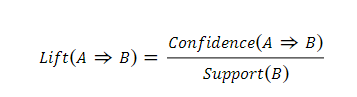
  
Figure 3.2.1.4 Pre-processing of biweekly data

To be mentioned that, there is a pre-requisite step for performing association rule mining, which is one-hot encoding on transaction data. One-hot encoding is a popular technique used in data pre-processing to convert categorical variables into a binary representation that can be used by machine learning algorithms. To perform one-hot encoding, a function ‘perform\_one\_hot\_encoding’ is defined to perform one-hot encoding on a given list of transactions. On top of that, a flag, ‘remove\_lower’ is set to True, it checks each column name in the DataFrame to determine if it contains irrelevant information. Columns with lowercase characters in their names will be removed. After defining the function for one-hot encoding, transactions of months and biweeks are ready to undergo the encoding process.

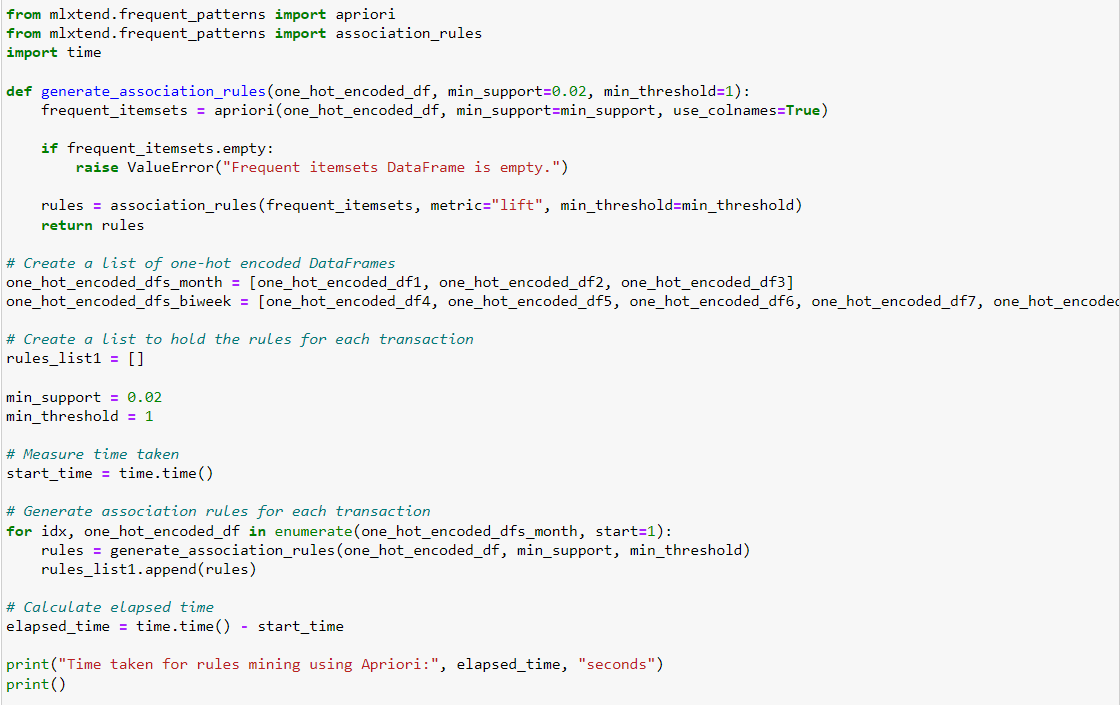
  
Figure 3.2.1.5 One-hot encoding of the data

At the core of the project, the association rule mining module employs algorithms such as Apriori and FP-Growth to extract association rules within each segmented time frame. This phase involves unearthing frequent itemsets and generating association rules based on predefined support and confidence thresholds. These processes unveil insights into item co-occurrence trends within transactions and uncover rules emblematic of purchasing behaviours specific to distinct periods. In the subsequent analysis and comparison module, the derived association rules from diverse time segments undergo assessment and juxtaposition. This module evaluates the weight, potency, and patterns of the unearthed rules. By contrasting rules across disparate temporal divisions, trends, and deviations within customer purchasing conduct are revealed. This evaluation assumes paramount importance in pinpointing the optimal period for efficient association rule mining. The Apriori algorithm is a classic data mining technique used for association rule mining to discover interesting relationships between items in large datasets, particularly in Market Basket Analysis. The algorithm works based on the principle that if an itemset is frequent, then all of its subsets must also be frequent. The key metrics used in Apriori are support, confidence, and lift. Support measures the proportion of transactions in the dataset that contain a specific itemset by calculating the ratio of transactions containing the itemset to the total number of transactions. Confidence assesses the likelihood that an item Y is purchased when item X is bought. It is computed by discovering the ratio of the support of the itemset {X, Y} to the support of itemset X. Lastly, lift measures the strength of association between items X and Y by considering how much more likely Y is purchased when X is bought, compared to the case where they are bought independently. The equations for computing support, confidence, and lift is provided below.

  
Figure 3.2.1.6 Equation for support  
Figure 3.2.1.7 Equation for confidence

  
Figure 3.2.1.8 Equation for lift

First of all, a function that generates association rules from a given one-hot encoded DataFrame is defined as ‘generate\_association\_rules’. The minimum support threshold is set to 0.02 and the minimum lift threshold is set to 1. As mentioned previously, both Apriori and FpGrowth algorithms will be used to discover frequent item sets from the one-hot encoded DataFrame. An empty list called ‘rules\_list1’ is created to store the association rules for each set of transactions. Moreover, the time taken to generate the association rules will be recorded to compare the performance of the algorithms.

  
Figure 3.2.1.9 Generation of association rules

Subsequently, the visualization and reporting module transforms the findings into comprehensible visual representations and reports. The generated reports provide a concise overview of the analysis results to facilitate decision-makers in gleaning insights and actionable recommendations. Subsequent to the analysis, the last module will interpret the ramifications of association rules uncovered within distinct time frames. This translation of patterns into meaningful insights establishes a foundation for formulating recommendations. The module identifies potential strategies for refining marketing endeavours, amplifying customer engagement, or honing inventory management strategies. The code below is responsible for exporting the results of association rule mining to an Excel file. It iterates through the association rules generated for the top three months and prints out the top 15 association rules along with additional information such as support, confidence, lift, total receipts, total sales quantity, and total sales value. The analysis will be carried out by focusing on the changes in the metrics such as lift and support. Any significant fluctuation of the values will be observed and meaningful insights are extracted.

  
Figure 3.2.1.10 Exporting the results to excel file

### 3.2.2 Error handling

In the domain of association rule mining, the minimum support threshold holds essential significance as it governs the frequency criterion that a set of items must fulfil to qualify as a frequent item set. Opting for a lower minimum support threshold can give rise to an abundance of candidate itemsets which will yield a substantial volume of association rules. However, this strategy can introduce several challenges such as notably memory space errors. To adeptly address these errors, meticulous adjustment of the minimum support threshold is essential. By slightly elevating the threshold, the proliferation of generated rules can be managed, hence, reducing the likelihood of memory space errors. This course of action ensures an equilibrium between capturing substantial insights and upholding the stability of the system.

## 3.3 Implementation Issue and Challenges

Issues arising during project implementation can delay the progress, compromise result integrity, and necessitate meticulous attention and resolution to ensure project fruition. One of the most substantial issues in this project is the minimum support threshold selection during the generation of association rules. Setting it too low can result in an excessive number of rules which makes it challenging to find meaningful insights. Conversely, setting it too high may lead to the exclusion of valuable rules. Hence, it took much time to find the right balance between comprehensiveness and relevance. Moreover, the preparation of the monthly data delays the overall progress of the project. This could be explained by the top three consecutive months with the highest cumulative transaction counts that might span across different years which might cause inaccurate extraction of data. The condition of allowing the months to span across years has to be specified in the coding work to ensure the accuracy of the data.

## 3.4 Timeline

### 3.4.1 FYP 1 Timeline

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Task | Week 1 | Week 2 | Week 3 | Week 4 | Week 5 | Week 6 | Week 7 | Week 8 | Week 9 | Week 10 | Week 11 | Week 12 | Week 13 |
| 1 | Meet up with supervisor regarding FYP 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Preliminary work |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Modification of coding |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Report writing |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Report submission |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Presentation |  |  |  |  |  |  |  |  |  |  |  |  |  |

### 3.4.2 FYP 2 Timeline

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Task | Week 1 | Week 2 | Week 3 | Week 4 | Week 5 | Week 6 | Week 7 | Week 8 | Week 9 | Week 10 | Week 11 | Week 12 | Week 13 |
| 1 | Meet up with supervisor regarding FYP 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | System design reevaluation |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | System testing |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | System debugging |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Report writing |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Report submission |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Presentation |  |  |  |  |  |  |  |  |  |  |  |  |  |

# CHAPTER 4 Preliminary Work

In this chapter, results such as the top 15 highest frequency and product affinity analysis for the top three months with the highest transaction counts will be displayed. Comparison and analysis of the association rules generated for different time periods will be conducted to study the patterns, changes, or trends in customer purchasing behaviours over time. Metrics of association rules such as support, lift, and confidence will be observed to evaluate the strength and significance of the rules. The results are shown in tables form below.

## 4.1 Monthly result

The top three consecutive months with the highest cumulative transaction counts chosen from the dataset are October, November, and December of 2011. The three months offer a fascinating opportunity to delve into the dynamics of consumer behavior and product affinities during a critical period. There is a total of 2637 transactions that has been processed in October. This robust activity likely corresponds to various factors, such as seasonality, holiday preparations, or special promotions. It's a promising starting point for understanding what was driving this high transaction volume. In November, the number of transactions ascended to 3462. The sharp increase suggests a surge in consumer activity might be due to the onset of holiday shopping, Black Friday deals, or other seasonal events. On the other hand, an interesting contrast presented in December even though the holiday season being in full swing, the transaction count dropped notably to 1015. This could be explained by the dataset likely only captures transactions up to the middle of December. The analysis will be performed to obtain the top 15 highest frequency itemsets in the months and biweeks. Product affinity analysis will also be carried out for the result of the months and biweeks to extract insights. It will help identify the most popular products during these peak transaction periods and reveals consumer preferences and potential cross-selling opportunities. Moreover, the product affinity analysis to be conducted will provide a deeper understanding of how products are associated with one another in customers' shopping baskets. Researching into these consecutive high-transaction months presents a rich opportunity to gain insights into customer behaviour and product relationships during a critical period of the year.

Table 4.1.1 below shows the top 15 highest frequency itemsets in the month of October 2011.

Table 4.1.1 Top 15 highest frequency itemsets in the month of October 2011

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | PAPER CHAIN KIT 50'S CHRISTMAS | 0.0899 | 237 | 3189 | 9419.39 |
| 2 | JUMBO BAG RED RETROSPOT | 0.0770 | 203 | 4803 | 9409.95 |
| 3 | HOT WATER BOTTLE KEEP CALM | 0.0717 | 189 | 1315 | 6066.69 |
| 4 | REGENCY CAKESTAND 3 TIER | 0.0686 | 181 | 1027 | 12660.84 |
| 5 | WHITE HANGING HEART T-LIGHT HOLDER | 0.0656 | 173 | 1700 | 4842.25 |
| 6 | DOORMAT KEEP CALM AND COME IN | 0.0637 | 168 | 1330 | 10111.95 |
| 7 | SET OF 3 CAKE TINS PANTRY DESIGN | 0.0611 | 161 | 559 | 2968.47 |
| 8 | JUMBO BAG 50'S CHRISTMAS | 0.0611 | 161 | 2273 | 4620.59 |
| 9 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0607 | 160 | 1645 | 4631.67 |
| 10 | POPCORN HOLDER | 0.0584 | 154 | 5865 | 4778.77 |
| 11 | ASSORTED COLOUR BIRD ORNAMENT | 0.0580 | 153 | 2736 | 4645.92 |
| 12 | BAKING SET 9 PIECE RETROSPOT | 0.0557 | 147 | 617 | 3111.75 |
| 13 | POSTAGE | 0.0550 | 145 | 330 | 7536.78 |
| 14 | WOODEN HEART CHRISTMAS SCANDINAVIAN | 0.0546 | 144 | 3028 | 1039.32 |
| 15 | JUMBO BAG VINTAGE DOILY | 0.0546 | 144 | 2047 | 4006.71 |

The analysis conducted in October unveils some compelling insights into the sales performance of the top 15 items for that period. Notably, "PAPER CHAIN KIT 50'S CHRISTMAS" stands out as a remarkable performer among these items and it demonstrated robust sales trends. This item amassed a remarkable 237 total receipts which reflects its consistent appeal to customers. In the context of sales quantity, it reached a substantial 3189 units and the high sales volume translated into a significant sales value of 9419.39 which makes it a top revenue generator. The support value which is the measure of the item's prevalence in transactions reached notably high at 0.0899. Similarly, the item "JUMBO BAG RED RETROSPOT" showcased impressive performance with 203 total receipts and a sales quantity of 4803. This high sales volume contributed to total sales amounting to 9409.95 as its significance in generating revenue. Another notable item which is “HOT WATER BOTTLE KEEP CALM" secured 189 total receipts with a sales quantity of 1315 and a sales value of 6066.69 which resulted in a support value of 0.0717. In the context of total sales values, the item “REGENCY CAKESTAND 3 TIER” took the credit by accumulating up to 12660.84 which was the highest sales value among the items in October 2011 even though it did not outperform the other items in terms of sales quantity and total receipt count. These findings indicated dynamic shifts in customer preferences as certain items gained substantial popularity while others exhibited varying levels of sales traction.

Table 4.1.2 below shows the product affinity analysis for the month of October 2011.

Table 4.1.2 Product affinity analysis for the month of October 2011

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | PINK REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0201 | 0.8030 | 27.8630 | 399 | 2792 | 8880.52 |
| 2 | GREEN REGENCY TEACUP AND SAUCER | PINK REGENCY TEACUP AND SAUCER | 0.0201 | 0.6974 | 27.8630 | 399 | 2792 | 8880.52 |
| 3 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS | 0.0231 | 0.8026 | 23.7813 | 494 | 2659 | 3652.55 |
| 4 | SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATIONS | 0.0231 | 0.6854 | 23.7813 | 494 | 2659 | 3652.55 |
| 5 | PINK REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0209 | 0.83333 | 23.3777 | 422 | 2957 | 9333.63 |
| 6 | ROSES REGENCY TEACUP AND SAUCER | PINK REGENCY TEACUP AND SAUCER | 0.0209 | 0.5851 | 23.3777 | 422 | 2957 | 9333.63 |
| 7 | GREEN REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0239 | 0.8289 | 23.2546 | 445 | 3107 | 9900.57 |
| 8 | ROSES REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0239 | 0.6702 | 23.2546 | 445 | 3107 | 9900.57 |
| 9 | ROLL WRAP 50'S CHRISTMAS | ROLL WRAP 50'S RED CHRISTMAS | 0.0209 | 0.5500 | 23.0214 | 468 | 3968 | 5624.77 |
| 10 | ROLL WRAP 50'S RED CHRISTMAS | ROLL WRAP 50'S CHRISTMAS | 0.0209 | 0.8730 | 23.0214 | 468 | 3968 | 5624.77 |
| 11 | SET 12 COLOUR PENCILS DOLLY GIRL | SET 12 COLOUR PENCILS SPACEBOY | 0.0235 | 0.6889 | 20.1844 | 478 | 5072 | 3781.50 |
| 12 | SET 12 COLOUR PENCILS SPACEBOY | SET 12 COLOUR PENCILS DOLLY GIRL | 0.0235 | 0.6889 | 20.1844 | 478 | 5072 | 3781.50 |
| 13 | SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN SLEIGH DECORATION | 0.0224 | 0.6629 | 19.6418 | 500 | 2867 | 3962.12 |
| 14 | SET OF 3 WOODEN SLEIGH DECORATION | SET OF 3 WOODEN STOCKING DECORATION | 0.0224 | 0.6629 | 19.6418 | 500 | 2867 | 3962.12 |
| 15 | SET OF 3 WOODEN HEART DECORATION | SET OF 3 WOODEN TREE DECORATION | 0.0220 | 0.5179 | 17.9683 | 549 | 3490 | 4795.83 |

During the initial month, the analysis of the top 15 association rules uncovers intriguing insights, shedding light on distinct patterns of product associations. Notable relationships emerged among certain products. For example, the association rule featuring the antecedent "PINK REGENCY TEACUP AND SAUCER" and consequent "GREEN REGENCY TEACUP AND SAUCER" demonstrated a substantial confidence level of 0.8030 and an extraordinary lift value of 27.8630. Despite its relatively modest support at 0.0201, this rule indicated a robust and compelling connection between these two items. Similarly, the rule involving "SET OF 3 WOODEN TREE DECORATIONS" and "SET OF 3 WOODEN STOCKING DECORATION" displayed a confidence of 0.8026 and a lift of 23.7813 which suggests a propensity for customers to purchase both sets together. Remarkably, numerous rules showcased akin products in varying colours such as tea cups and saucers underscoring the considerable impact of colour preferences on buying decisions. For instance, the “PINK REGENCY TEACUP AND SAUCER” and “GREEN REGENCY TEACUP AND SAUCER” have higher lift compared to “PINK REGENCY TEACUP AND SAUCER”, and “GREEN REGENCY TEACUP AND SAUCER”, so people have a higher tendency to buy “PINK REGENCY TEACUP AND SAUCER” and “GREEN REGENCY TEACUP AND SAUCER” together. As suggestions, businesses should be more concerned about colour assortments of their products as they can effectively influence buying decisions and customer behaviour. Retailers could adjust their offerings by understanding and utilizing colour preferences to match the dynamic tastes and preferences of customers which could result in higher sales and customer satisfaction.

Table 4.1.3 below shows the top 15 highest frequency itemsets in the month of November 2011.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | RABBIT NIGHT LIGHT | 0.1511 | 523 | 14883 | 34422.09 |
| 2 | PAPER CHAIN KIT 50'S CHRISTMAS | 0.1176 | 407 | 7868 | 28883.04 |
| 3 | HOT WATER BOTTLE KEEP CALM | 0.0887 | 307 | 2285 | 11500.87 |
| 4 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0823 | 285 | 3786 | 12831.54 |
| 5 | JUMBO BAG 50'S CHRISTMAS | 0.0797 | 276 | 3971 | 10913.10 |
| 6 | WHITE HANGING HEART T-LIGHT HOLDER | 0.0789 | 273 | 4938 | 14345.22 |
| 7 | POPCORN HOLDER | 0.0774 | 268 | 12159 | 13698.74 |
| 8 | WOODEN STAR CHRISTMAS SCANDINAVIAN | 0.0739 | 256 | 5365 | 1787.21 |
| 9 | JUMBO BAG RED RETROSPOT | 0.0731 | 253 | 5807 | 11458.21 |
| 10 | JUMBO BAG PAISLEY PARK | 0.0719 | 249 | 3072 | 7830.34 |
| 11 | GARDENERS KNEELING PAD KEEP CALM | 0.0690 | 239 | 2763 | 5863.75 |
| 12 | WOODEN HEART CHRISTMAS SCANDINAVIAN | 0.0685 | 237 | 5003 | 1658.29 |
| 13 | HAND WARMER OWL DESIGN | 0.0679 | 235 | 2530 | 5756.39 |
| 14 | LUNCH BAG PAISLEY PARK | 0.0676 | 234 | 2325 | 3758.73 |
| 15 | JUMBO BAG VINTAGE DOILY | 0.0673 | 233 | 3602 | 8996.75 |

Table 4.1.3 Top 15 highest frequency itemsets in the month of November 2011

As evident from the provided table, several noteworthy alterations can be discerned in the itemsets when considering their metrics. Notably, the item “PAPER CHAIN KIT 50’S CHRISTMAS” not only maintained its presence on the top 15 ranking but also demonstrated significant growth in terms of support value. The support value escalated from 0.0899 to 0.1176 in the second month. The noteworthy 31% increase in support implies a substantial rise in the prevalence of this item within transactions. The robust support value of 0.1176 further accentuated its significance. Moreover, "HOT WATER BOTTLE KEEP CALM" displayed increased figures in terms of total receipts, support, sales quantity, and value. Its support value ascended from 0.0717 to 0.0887 which indicates growing interest from consumers. In contrast, " REGENCY CAKESTAND 3 TIER " which was initially positioned in rank 6 in the first month experienced a decline in total receipts which naturally led to a reduction in its support value, eventually, it was eliminated from the top 15 ranking in the second month. These divergent trajectories strongly indicated shifting consumer preferences between the two months, with the item "PAPER CHAIN KIT 50'S CHRISTMAS" notably encountering a pronounced surge in demand.

Table 4.1.4 below shows the product affinity analysis for the month of November 2011.

Table 4.1.4 Product affinity analysis for the month of November 2011

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | PINK REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0202 | 0.8046 | 28.1365 | 399 | 2792 | 8880.52 |
| 2 | GREEN REGENCY TEACUP AND SAUCER | PINK REGENCY TEACUP AND SAUCER | 0.0202 | 0.7071 | 28.1365 | 399 | 2792 | 8880.52 |
| 3 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATION, SET OF 3 WOODEN HEART DECORATIONS | 0.0217 | 0.6410 | 26.4194 | 805 | 4890 | 6712.63 |
| 4 | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATIONS | 0.0217 | 0.8929 | 26.4194 | 805 | 4890 | 6712.63 |
| 5 | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS | 0.0217 | 0.9036 | 26.0693 | 805 | 4890 | 6712.63 |
| 6 | SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN TREE DECORATIONS | 0.0217 | 0.6250 | 26.0693 | 805 | 4890 | 6712.63 |
| 7 | GREEN REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0202 | 0.7071 | 23.9988 | 445 | 3107 | 9900.57 |
| 8 | ROSES REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0202 | 0.6863 | 23.9988 | 445 | 3107 | 9900.57 |
| 9 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS | 0.02560 | 0.7692 | 22.1923 | 494 | 2659 | 3652.55 |
| 10 | SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATIONS | 0.02560 | 0.7500 | 22.1923 | 494 | 2659 | 3652.55 |
| 11 | CHARLOTTE BAG PINK POLKADOT | RED RETROSPOT CHARLOTTE BAG | 0.0217 | 0.7732 | 21.4144 | 469 | 6742 | 9502.03 |
| 12 | RED RETROSPOT CHARLOTTE BAG | CHARLOTTE BAG PINK POLKADOT | 0.0217 | 0.6000 | 21.4144 | 469 | 6742 | 9502.03 |
| 13 | SET OF 3 WOODEN SLEIGH DECORATIONS | SET OF 3 WOODEN TREE DECORATION | 0.0240 | 0.7217 | 21.3561 | 482 | 2726 | 3781.07 |
| 14 | SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN SLEIGH DECORATIONS | 0.0240 | 0.7094 | 21.3561 | 482 | 2726 | 3781.07 |
| 15 | SET OF 3 WOODEN SLEIGH DECORATION | SET OF 3 WOODEN STOCKING DECORATION | 0.0234 | 0.7043 | 20.3204 | 500 | 2867 | 3962.12 |

Intriguing shifts in rankings come to the forefront in the comparative analysis of association rules between October and November in 2011. the comparative analysis of association rules between the initial and second months of the dataset. The association rule featuring "PINK REGENCY TEACUP AND SAUCER" and "GREEN REGENCY TEACUP AND SAUCER" maintains its commanding position, holding the top spot with an impressive confidence level of 0.8046 and a lift value of 28.1365. This indicates that customers consistently tend to purchase these two items together and it reflects a strong and reliable association between them. There are obvious changes in rules related to different sets of wooden Christmas decorations. For instance, there was a rule involving 'SET OF 3 WOODEN TREE DECORATIONS' and 'SET OF 3 WOODEN STOCKING DECORATION' with has higher support but 'SET OF 3 WOODEN TREE DECORATIONS', 'SET OF 3 WOODEN STOCKING DECORATION' and 'SET OF 3 WOODEN HEART DECORATIONS has a higher lift, forming more complex rules which possibly suggest that evolving purchasing behaviour and perhaps a broader interest in related Christmas decoration sets in November. The higher lift value in this rule suggests a more substantial association or correlation between these items even if they are not as frequently purchased together. Based on the result, a hypothesis of customers become more interested in comprehensive Christmas decoration collections rather than individual items could be made. These fluctuations in rule rankings underscore the dynamic nature of consumer conduct and the evolving dynamics of product affiliations over time.

Table 4.1.5 below shows the top 15 highest frequency itemsets in the month of December 2011.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | RABBIT NIGHT LIGHT | 0.1369 | 139 | 4298 | 9618.01 |
| 2 | PAPER CHAIN KIT 50'S CHRISTMAS | 0.1192 | 121 | 2089 | 6870.71 |
| 3 | HOT WATER BOTTLE KEEP CALM | 0.0916 | 93 | 680 | 3974.68 |
| 4 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0887 | 90 | 862 | 2763.56 |
| 5 | HAND WARMER OWL DESIGN | 0.0808 | 82 | 812 | 1756.11 |
| 6 | JUMBO BAG 50'S CHRISTMAS | 0.0739 | 75 | 1075 | 2224.15 |
| 7 | BICYCLE PUNCTURE REPAIR KIT | 0.0729 | 74 | 1240 | 2595.75 |
| 8 | GARDENERS KNEELING PAD KEEP CALM | 0.0719 | 73 | 699 | 1649.27 |
| 9 | POPCORN HOLDER | 0.0709 | 72 | 6198 | 5409.86 |
| 10 | CHOCOLATE HOT WATER BOTTLE | 0.0680 | 69 | 575 | 3470.24 |
| 11 | REGENCY CAKESTAND 3 TIERS | 0.0650 | 66 | 441 | 5902.92 |
| 12 | HAND WARMER RED LOVE HEART | 0.0631 | 64 | 474 | 1011.64 |
| 13 | HOT WATER BOTTLE TEA AND SYMPATHY | 0.0631 | 64 | 608 | 3734.80 |
| 14 | WHITE HANGING HEART T-LIGHT HOLDER | 0.0621 | 63 | 777 | 2204.07 |
| 15 | SCOTTIE DOG HOT WATER BOTTLE | 0.0611 | 62 | 403 | 2120.11 |

Table 4.1.5 Top 15 highest frequency itemsets in the month of December 2011

In the analysis of sales trends over the three consecutive month, in the analysis of sales trends over the three consecutive months could be observed such that certain items maintained their popularity while others surged in demand as the holiday season progressed. Firstly, the item "PAPER CHAIN KIT 50'S CHRISTMAS" maintained its position among the top items, its popularity remained with support of 0.1192 and generated a decent sales value of 6870.71 which indicates that customers were still interested in Christmas-related items during the holiday month. This observation underscores the enduring appeal of holiday-themed products and this insight allows businesses aiming to capitalize on seasonal trends. Simultaneously, "HOT WATER BOTTLE KEEP CALM" continued to perform well in December with a support of 0.0916 and this fact suggests that customers sought comfort and warmth during the winter season. Notably, "HAND WARMER OWL DESIGN" made its way into the top 15 items in December which implies a shift in customer preferences towards products that provided warmth and comfort during the cold weather. Customers may resonate with the owl design which highlights the significance of product design and aesthetics in influencing purchasing decisions. This dynamic interplay of metrics indicates the sway of evolving seasonal preferences and market fluctuations, steering shifts in the performance of these top items throughout the three consecutive months.

Table 4.1.6 below shows the product affinity analysis for the month of December 2011.

Table 4.1.6 Product affinity analysis for the month of December 2011

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS | 0.0217 | 0.8800 | 30.8000 | 494 | 2659 | 3652.55 |
| 2 | SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATIONS | 0.0217 | 0.7586 | 30.8000 | 494 | 2659 | 3652.55 |
| 3 | ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0227 | 0.8846 | 28.0589 | 633 | 4428 | 14057.36 |
| 4 | GREEN REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER | 0.0227 | 0.7188 | 28.0589 | 633 | 4428 | 14057.36 |
| 5 | GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0227 | 0.9200 | 27.4647 | 633 | 4428 | 14057.36 |
| 6 | ROSES REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER | 0.0227 | 0.6765 | 27.4617 | 633 | 4428 | 14057.36 |
| 7 | CHARLIE + LOLA PINK HOT WATER BOTTLE | CHARLIE + LOLA RED HOT WATER BOTTLE | 0.0227 | 0.8519 | 26.2009 | 259 | 958 | 2266.94 |
| 8 | CHARLIE + LOLA RED HOT WATER BOTTLE | CHARLIE+LOLA PINK HOT WATER BOTTLE | 0.0227 | 0.6970 | 26.2009 | 259 | 958 | 2266.94 |
| 9 | GREEN REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0276 | 0.8750 | 26.1213 | 445 | 3107 | 9900.57 |
| 10 | ROSES REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0276 | 0.8235 | 26.1213 | 445 | 3107 | 9900.57 |
| 11 | ROSES REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER | PINK REGENCY TEACUP AND SAUCER | 0.0227 | 0.82141 | 26.0547 | 633 | 4428 | 14057.36 |
| 12 | PINK REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER | 0.0227 | 0.7188 | 26.0547 | 633 | 4428 | 14057.36 |
| 13 | GREEN REGENCY TEACUP AND SAUCER | PINK REGENCY TEACUP AND SAUCER | 0.0246 | 0.7813 | 24.7803 | 399 | 2792 | 8880.52 |
| 14 | PINK REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0246 | 0.7813 | 24.7803 | 399 | 2792 | 8880.52 |
| 15 | PINK REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0256 | 0.8125 | 24.2555 | 422 | 2957 | 9333.63 |

In a comparison of the rules across the previous two months, discernible shifts in rankings and alterations in metrics become evident. A notable observation is the heightened confidence and lift values seen in several rules, indicative of stronger associations between antecedents and consequents. An example is the rule involving "SET OF 3 WOODEN TREE DECORATIONS" and "SET OF 3 WOODEN STOCKING DECORATION" which has experienced a substantial elevation in confidence and lift. The lift value of that rule has experienced a decent increment from 23.78134 to 30.8 across three months which indicates a continued tendency for customers to buy the items together. This suggests a more pronounced inclination among customers to purchase these items concurrently. It's noteworthy that the rules encompassing "CHARLIE+LOLA PINK HOT WATER BOTTLE" and "CHARLIE + LOLA RED HOT WATER BOTTLE" have preserved their positions with remarkable confidence and lift values which indicates a consistent pattern of joint purchase for both variations. Moreover, there are significant rank changes with some association rules moving up or down the list. For instance, the rule ("PINK REGENCY TEACUP AND SAUCER" -> "GREEN REGENCY TEACUP AND SAUCER") and ("GREEN REGENCY TEACUP AND SAUCER" -> "PINK REGENCY TEACUP AND SAUCER") held the top positions in previous months, however, these rules dropped in rank as other rules took precedence. Overall, these fluctuations in metrics mirror the evolving purchasing behaviours and preferences of customers over time, eventually spotlighting the dynamic nature of product associations within the realm of retail.

In summary, several insights could be obtained based on the results from the three months. Seasonal products especially those related to Christmas have experienced significant changes in support and rankings as the holiday season progressed. On the other hand, the "RABBIT NIGHT LIGHT" consistently performed well in terms of popularity and this possibly indicates its year-round appeal. Additionally, Customer preferences and purchasing behaviour were strongly influenced by the time of year which emphasizes the importance of seasonality in retail. In the context of association rules, one of the most prominent patterns observed throughout these three months was the strong association between the 'PINK REGENCY TEACUP AND SAUCER' and the 'GREEN REGENCY TEACUP AND SAUCER. These two items appeared together frequently in customer purchases and the combination of the two items remains a confidence level exceeding 80% in each month. This suggests that customers had a preference for buying these teacup and saucer sets as a pair. Another consistent finding was related to sets of wooden decorations. Items such as 'SET OF 3 WOODEN TREE DECORATIONS', 'SET OF 3 WOODEN STOCKING DECORATION', and ‘SET OF 3 WOODEN HEART DECORATIONS’ show strong associations with one another across all three months. Customers who purchased one set of wooden decorations had having very high possibility to buy another set as well. In conclusion, these insights can guide businesses on inventory stocking decisions and marketing efforts to cater to customer preferences more effectively and enhance their overall sales strategies.

## 4.2 Biweekly result

Table 4.2.1 below shows the top 15 highest frequency itemsets in Biweek 1 (9/24/2011 – 10/7/2011). At the top of the list was the "JUMBO BAG RED RETROSPOT” which accumulated a total sales quantity of 2489 and generated a substantial total sales value of 4609.33. The support value of 0.0832 indicates its popularity among customers. This data explains that customers were particularly drawn to this item during this period which indicates that it held a strong appeal while it is possibly due to its design, size, or practicality. Also, its high sales volume and support value suggest that it was a top choice for shoppers during this specific period. It was followed closely by the "PAPER CHAIN KIT 50'S CHRISTMAS" with 53 receipts, 931 units sold, and a total sales value of 2588.93, and contributed significantly to sales during this period with a support of 0.0802. The fact that it ranked second on the list even though being a Christmas-themed product and this implies that customers were already beginning to prepare for the upcoming holiday season. This insight can be valuable for businesses in terms of inventory planning and marketing strategies to ensure they cater to seasonal demands effectively. Another notable item was the "DOORMAT KEEP CALM AND COME IN". Although it did not exhibit the support value as high as the previous items, however, it generated the highest total sales value among the items with a value of 5046. 24. This suggests that while it may not have been the most frequently purchased item but it commanded a higher price point and it is possibly due to its unique design or functionality

Table 4.2.1 Top 15 highest frequency itemsets in Biweek 1 (9/24/2011 - 10/7/2011)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | JUMBO BAG RED RETROSPOT | 0.0832 | 55 | 2489 | 4609.33 |
| 2 | PAPER CHAIN KIT 50'S CHRISTMAS | 0.0802 | 53 | 931 | 2588.93 |
| 3 | DOORMAT KEEP CALM AND COME IN | 0.0772 | 51 | 695 | 5046.24 |
| 4 | PLAYING CARDS KEEP CALM & CARRY ON | 0.0711 | 47 | 445 | 558.67 |
| 5 | LUNCH BAG SPACEBOY DESIGN | 0.0666 | 44 | 288 | 521.53 |
| 6 | HOT WATER BOTTLE KEEP CALM | 0.0651 | 43 | 564 | 2509.64 |
| 7 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0651 | 43 | 519 | 1403.77 |
| 8 | LUNCH BAG VINTAGE DOILY | 0.0635 | 42 | 322 | 536.46 |
| 9 | REGENCY CAKESTAND 3 TIER | 0.0635 | 42 | 393 | 4715.58 |
| 10 | JUMBO BAG VINTAGE DOILY | 0.0635 | 42 | 752 | 1423.26 |
| 11 | ASSORTED COLOUR BIRD ORNAMENT | 0.0620 | 41 | 588 | 993.72 |
| 12 | JUMBO BAG 50'S CHRISTMAS | 0.0620 | 41 | 466 | 954.63 |
| 13 | SPOTTY BUNTING | 0.0605 | 40 | 178 | 969.58 |
| 14 | PARTY BUNTING | 0.0605 | 40 | 152 | 1020.88 |
| 15 | LARGE WHITE HEART OF WICKER | 0.0590 | 39 | 292 | 718.98 |

Table 4.2.2 below shows the product affinity analysis for Biweek 1 (9/24/2011 - 10/7/2011). The result highlighted the strong association between 'GLASS APOTHECARY BOTTLE TONIC' and 'GLASS APOTHECARY BOTTLE PERFUME' which exhibited a perfect confidence of 1.0000 which implies that whenever a customer purchased 'GLASS APOTHECARY BOTTLE TONIC', they also purchased 'GLASS APOTHECARY BOTTLE PERFUME'. The lift value for that rule is notably high, 34.7895 which suggests the likelihood of customers buying these two items together is significantly higher than if they were chosen at random. Furthermore, the strong association between ‘FELTCRAFT PRINCESS CHARLOTTE DOLL' and 'FELTCRAFT PRINCESS LOLA DOLL’ is supported by a perfect confidence of 1.0000 and a lift of 34.7895. The value of the metrics for the two rules mentioned are the same but they were varied in terms of total sales value such that the total sales value of the dolls was double compared to the combination of tonic and perfume. In this case, the dolls generate significantly more revenue while both pairs of products are highly associated. Thus, businesses might prioritize promoting and bundling the dolls as a strategic move to maximize profits.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | GLASS APOTHECARY BOTTLE TONIC | GLASS APOTHECARY BOTTLE PERFUME | 0.0212 | 1.0000 | 34.7895 | 256 | 1007 | 4022.43 |
| 2 | FELTCRAFT PRINCESS CHARLOTTE DOLL, FELTCRAFT PRINCESS LOLA DOLL | FELTCRAFT PRINCESS OLIVIA DOLL | 0.0212 | 1.0000 | 34.7895 | 488 | 1998 | 8091.36 |
| 3 | GLASS APOTHECARY BOTTLE PERFUME | GLASS APOTHECARY BOTTLE TONIC | 0.0212 | 0.7368 | 34.7895 | 256 | 1007 | 4022.43 |
| 4 | FELTCRAFT PRINCESS OLIVIA DOLL | FELTCRAFT PRINCESS CHARLOTTE DOLL, FELTCRAFT PRINCESS LOLA DOLL | 0.0212 | 0.7368 | 34.7895 | 488 | 1998 | 8091.36 |
| 5 | FELTCRAFT PRINCESS CHARLOTTE DOLL, FELTCRAFT PRINCESS OLIVIA DOLL | FELTCRAFT PRINCESS LOLA DOLL | 0.0212 | 0.8750 | 32.1319 | 488 | 1998 | 8091.36 |
| 6 | FELTCRAFT PRINCESS LOLA DOLL | FELTCRAFT PRINCESS CHARLOTTE DOLL, FELTCRAFT PRINCESS OLIVIA DOLL | 0.0212 | 0.7778 | 32.1319 | 488 | 1998 | 8091.36 |
| 7 | FELTCRAFT PRINCESS OLIVIA DOLL | FELTCRAFT PRINCESS LOLA DOLL | 0.0242 | 0.8421 | 30.9240 | 310 | 1094 | 4284.39 |
| 8 | FELTCRAFT PRINCESS LOLA DOLL | FELTCRAFT PRINCESS OLIVIA DOLL | 0.0242 | 0.8889 | 30.9240 | 310 | 1094 | 4284.39 |
| 9 | FELTCRAFT PRINCESS LOLA DOLL, FELTCRAFT PRINCESS OLIVIA DOLL | FELTCRAFT PRINCESS CHARLOTTE DOLL | 0.0212 | 0.8750 | 27.5417 | 488 | 1998 | 8091.36 |
| 10 | FELTCRAFT PRINCESS CHARLOTTE DOLL | FELTCRAFT PRINCESS LOLA DOLL, FELTCRAFT PRINCESS OLIVIA DOLL | 0.0212 | 0.6667 | 27.5417 | 488 | 1998 | 8091.36 |
| 11 | FELTCRAFT PRINCESS OLIVIA DOLL | FELTCRAFT PRINCESS CHARLOTTE DOLL | 0.0242 | 0.8421 | 26.5063 | 312 | 1332 | 5497.96 |
| 12 | FELTCRAFT PRINCESS CHARLOTTE DOLL | FELTCRAFT PRINCESS OLIVIA DOLL | 0.0242 | 0.7619 | 26.5063 | 312 | 1332 | 5497.96 |
| 13 | ROSES REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0257 | 0.7391 | 25.7140 | 445 | 3107 | 9900.57 |
| 14 | GREEN REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0257 | 0.8947 | 25.7140 | 445 | 3107 | 9900.57 |
| 15 | FELTCRAFT PRINCESS LOLA DOLL | FELTCRAFT PRINCESS CHARLOTTE DOLL | 0.0212 | 0.7778 | 24.4815 | 354 | 1570 | 6400.37 |

Table 4.2.2 Product affinity analysis for Biweek 1 (9/24/2011 - 10/7/2011)

Table 4.2.3 below shows the top 15 highest frequency itemsets in Biweek 2 (10/8/2011 - 10/21/2011). Once again, the item ‘PAPER CHAIN KIT 50’S CHRISTMAS’ still emerged as the top-performing item with a support value of 0.0941. This high support value indicates that this item was prevalent in a significant portion of the transactions during this period. Customers showed a consistent interest in this Christmas-themed product and solidified its popularity during this time. Following closely, the ‘JUMBO BAG RED RETROSPOT’ exhibited robust performance as well with a slightly lower support value of 0.0825 compared to the top item. This suggests that the 'JUMBO BAG RED RETROSPOT' was a highly sought-after item during this period and it might be the outcome of promotional efforts. The ‘REGENCY CAKE STAND 3 TIER’ also made an impressive appearance with a support of 0.0816. It garnered significant attention from customers and contributed a total sales quantity of 443 and a total sales value of 2112.29 which highlighted its significance in this specific time period. This suggests that the 'REGENCY CAKE STAND 3 TIER' was not just a popular item but also a financially significant one as it was the key contributor to the overall sales performance. In fact, the significance of the items is not only concerning about high support value but also investigating the total sales quantity and total sales value.

Table 4.2.3 Top 15 highest frequency itemsets in Biweek 2 (10/8/2011 - 10/21/2011)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | PAPER CHAIN KIT 50'S CHRISTMAS | 0. 0941 | 106 | 1508 | 4399.32 |
| 2 | JUMBO BAG RED RETROSPOT | 0.0825 | 93 | 1703 | 3409.39 |
| 3 | REGENCY CAKESTAND 3 TIER | 0.0816 | 92 | 454 | 5615.73 |
| 4 | HOT WATER BOTTLE KEEP CALM | 0.0799 | 90 | 443 | 2112.29 |
| 5 | SET OF 3 CAKE TINS PANTRY DESIGN | 0.0665 | 75 | 337 | 1739.67 |
| 6 | JUMBO BAG 50'S CHRISTMAS | 0.0657 | 74 | 844 | 1848.37 |
| 7 | DOORMAT KEEP CALM AND COME IN | 0.0630 | 71 | 450 | 3450.38 |
| 8 | WOODEN STAR CHRISTMAS SCANDINAVIAN | 0.0630 | 71 | 1240 | 412.68 |
| 9 | WHITE HANGING HEART T-LIGHT HOLDER | 0.0630 | 71 | 879 | 2483.65 |
| 10 | WOODEN HEART CHRISTMAS SCANDINAVIAN | 0.0621 | 70 | 1516 | 493.76 |
| 11 | POPCORN HOLDER | 0.0612 | 69 | 2159 | 1830.80 |
| 12 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0603 | 68 | 754 | 2092.90 |
| 13 | JUMBO BAG VINTAGE DOILY | 0.0603 | 68 | 669 | 1345.82 |
| 14 | BAKING SET 9 PIECE RETROSPOT | 0.0594 | 67 | 241 | 1241.11 |
| 15 | RETROSPOT TEA SET CERAMIC 11 PC | 0..0586 | 66 | 344 | 1763.36 |

Table 4.2.4 below shows the product affinity analysis for Biweek 2 (10/8/2011 - 10/21/2011). One of the most noteworthy findings is the strong positive relationship between the ‘SET OF 3 WOODEN HEART DECORATIONS’ and the ‘SET OF 3 WOODEN TREE DECORATIONS’ as antecedents, leading to the consequent purchase of the ‘SET OF 3 WOODEN STOCKING DECORATION’. This association is explained by the high confidence level of 95.45%. This suggests that these three items complement each other in terms of holiday decorating and customers tend to purchase them together to create a cohesive theme or look for their Christmas decorations. Conversely, the rules with the ‘SET OF 3 WOODEN STOCKING DECORATION’ as antecedent exhibited a lower confidence level of 68.85% compared to the previous rule. Some customers who buy stockings may not necessarily buy the heart and tree decorations. Some customers who buy stockings may not necessarily buy the heart and tree decorations. A notable observation is that many of the association rules identified in this analysis are Christmas-related which indicates people start to purchase Christmas decorations in the middle of October. This insight can help businesses plan their seasonal marketing campaigns and stock inventory accordingly to ensure they could meet the early demand for holiday-related products.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | SET OF 3 WOODEN HEART DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATION | 0.0251 | 0.9545 | 26.1952 | 805 | 4890 | 6712.63 |
| 2 | SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN HEART DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | 0.0251 | 0.6885 | 26.1952 | 805 | 4890 | 6712.63 |
| 3 | SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN STOCKING DECORATIONS | 0.0251 | 0.6563 | 24.4125 | 805 | 4890 | 6712.63 |
| 4 | SET OF 3 WOODEN HEART DECORATIONS, SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATION | 0.0251 | 0.9333 | 24.4125 | 805 | 4890 | 6712.63 |
| 5 | GREEN REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0209 | 0.7292 | 23.0307 | 445 | 3107 | 9900.57 |
| 6 | ROSES REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0209 | 0.6604 | 23.0307 | 445 | 3107 | 9900.57 |
| 7 | SET OF 3 WOODEN SLEIGH DECORATIONS, SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATION | 0.0209 | 0.8537 | 22.3285 | 738 | 4126 | 5697.87 |
| 8 | SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN SLEIGH DECORATIONS, SET OF 3 WOODEN STOCKING DECORATIONS | 0.0209 | 0.5469 | 22.3285 | 738 | 4126 | 5697.87 |
| 9 | SET OF 3 WOODEN SLEIGH DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATION | 0.0209 | 0.7955 | 21.8294 | 738 | 4126 | 5697.87 |
| 10 | SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN SLEIGH DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | 0.0209 | 0.5738 | 21.8294 | 738 | 4126 | 5697.87 |
| 11 | SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN TREE DECORATION | 0.0293 | 0.8033 | 21.0108 | 494 | 2659 | 3652.55 |
| 12 | SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN STOCKING DECORATION | 0.0293 | 0.7656 | 21.0108 | 494 | 2659 | 3652.55 |
| 13 | CHARLOTTE BAG PINK POLKADOT | RED RETROSPOT CHARLOTTE BAG | 0.0221 | 0.8222 | 20.8545 | 469 | 6742 | 9502.03 |
| 14 | RED RETROSPOT CHARLOTTE BAG | CHARLOTTE BAG PINK POLKADOT | 0.0221 | 0.5606 | 20.8545 | 469 | 6742 | 9502.03 |
| 15 | SET OF 3 WOODEN STOCKING DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN SLEIGH DECORATION | 0.0209 | 0.7143 | 20.6158 | 738 | 4126 | 5697.87 |

Table 4.2.4 Product affinity analysis for Biweek 2 (10/8/2011 - 10/21/2011)

Table 4.2.5 below displays the top 15 highest frequency itemsets in Biweek 3 (10/22/2011 - 11/4/2011). The item ‘RABBIT NIGHT LIGHT’ emerged as the top-selling item and it witnessed 131 transactions which resulted in a remarkable total sales quantity of 7,863 and a total sales value of 15,235.77. The high level of popularity may be attributed to several factors. It's possible that this item was seasonally relevant, perhaps being purchased as a Halloween or autumn-themed decoration. The 'PAPER CHAIN KIT 50'S CHRISTMAS' continued to perform well in this period and its notable presence might be attributed to the upcoming holiday season as customers seem to be preparing for Christmas. As Christmas was approaching, shoppers might have been purchasing items like paper chain kits to decorate their homes or prepare for holiday gatherings. Another item that gained significant attention during this period was the 'HOT WATER BOTTLE KEEP CALM.' This could be attributed to the changing season as cooler weather prompting customers to seek comfort and warmth. Hot water bottles are commonly used to keep warm during colder months and their popularity during this time is understandable. Surprisingly, ‘POPCORN HOLDER’ and ‘JUMBO BAG PAISLEY PARK’ also received notable attention from customers possibly indicating their relevance to gatherings or events during this period. Popcorn holders are often associated with movie nights or social gatherings while jumbo bags can be used for various purposes such as carrying items for events.

Table 4.2.5 Top 15 highest frequency itemsets in Biweek 3 (10/22/2011 - 11/4/2011)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | RABBIT NIGHT LIGHT | 0. 0980 | 131 | 7863 | 15235.77 |
| 2 | PAPER CHAIN KIT 50'S CHRISTMAS | 0.0927 | 124 | 2309 | 6757.31 |
| 3 | WHITE HANGING HEART T-LIGHT HOLDER | 0.0733 | 98 | 2131 | 6470.21 |
| 4 | HOT WATER BOTTLE KEEP CALM | 0.0666 | 89 | 844 | 3758.20 |
| 5 | ASSORTED COLOUR BIRD ORNAMENT | 0.0636 | 85 | 1668 | 2879.40 |
| 6 | POPCORN HOLDER | 0.0636 | 85 | 3609 | 2836.67 |
| 7 | JUMBO BAG PAISLEY PARK | 0.0636 | 85 | 1418 | 2738.89 |
| 8 | WOODEN STAR CHRISTMAS SCANDINAVIAN | 0.0628 | 84 | 1302 | 436.88 |
| 9 | GARDENERS KNEELING PAD KEEP CALM | 0.0621 | 83 | 590 | 1067.70 |
| 10 | JUMBO BAG RED RETROSPOT | 0.0621 | 83 | 1447 | 2979.36 |
| 11 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0606 | 81 | 840 | 2430.08 |
| 12 | REGENCY CAKESTAND 3 TIER | 0.0598 | 80 | 331 | 4199.46 |
| 13 | LUNCH BAG PAISLEY PARK | 0.0591 | 79 | 939 | 1491.83 |
| 14 | CHOCOLATE HOT WATER BOTTLE | 0.0591 | 79 | 393 | 2095.67 |
| 15 | WOODEN HEART CHRISTMAS SCANDINAVIAN | 0.0583 | 78 | 1635 | 586.05 |

Table 4.2.6 below illustrates the product affinity analysis for Biweek 3 (10/22/2011 - 11/4/2011). One of the most significant rules discovered from the analysis is the association between ‘PINK REGENCY TEACUP AND SAUCER’ and ‘GREEN REGENCY TEACUP AND SAUCER’. This rule signifies that customer who purchase one of these teacup and saucer sets are very likely to buy the other. This strong association is reflected in the high confidence of 86.05%. This strong association is further validated by a lift value of 27.1806 which indicates that these two products were frequently purchased together. Based on the insights gained from the result, businesses can capitalize on this association by promoting these sets as a bundle or suggesting one when the other is in a customer's cart. Similarly, the ‘SET OF 3 WOODEN STOCKING DECORATION’ and ‘SET OF 3 WOODEN TREE DECORATION’ remained their top performance by exhibiting a confidence level of 67.92% and a lift of 24.8696. This rule highlights that customers who selected one of these Christmas decoration sets often opted for the other as well. They are considered complementary and suitable for decorating during the holiday season. Businesses can leverage this insight to bundle or cross-promote these items effectively and enhance the customer shopping experience and increasing sales.

Table 4.2.6 Product affinity analysis for Biweek 3 (10/22/2011 - 11/4/2011)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | PINK REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0230 | 0.8605 | 27.1806 | 399 | 2792 | 8880.52 |
| 2 | GREEN REGENCY TEACUP AND SAUCER | PINK REGENCY TEACUP AND SAUCER | 0.0230 | 0.7255 | 27.1806 | 399 | 2792 | 8880.52 |
| 3 | SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN TREE DECORATION | 0.0223 | 0.6792 | 24.8696 | 494 | 2659 | 3652.55 |
| 4 | SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN STOCKING DECORATION | 0.0223 | 0.8182 | 24.8696 | 494 | 2659 | 3652.55 |
| 5 | ROSES REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0217 | 0.7609 | 24.0345 | 445 | 3107 | 9900.57 |
| 6 | GREEN REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER | 0.0217 | 0.6863 | 24.0345 | 445 | 3107 | 9900.57 |
| 7 | CHARLOTTE BAG PINK POLKADOT | RED RETROSPOT CHARLOTTE BAG | 0.0211 | 0.7907 | 23.5891 | 469 | 6742 | 9502.03 |
| 8 | RED RETROSPOT CHARLOTTE BAG | CHARLOTTE BAG PINK POLKADOT | 0.0211 | 0.6296 | 23.5891 | 469 | 6742 | 9502.03 |
| 9 | SET OF 3 WOODEN SLEIGH DECORATION | SET OF 3 WOODEN STOCKING DECORATION | 0.0205 | 0.7174 | 21.8060 | 500 | 2867 | 3962.12 |
| 10 | SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN SLEIGH DECORATION | 0.0205 | 0.6226 | 21.8060 | 500 | 2867 | 3962.12 |
| 11 | SPACEBOY LUNCH BOX | DOLLY GIRL LUNCH BOX | 0.0230 | 0.6727 | 20.4484 | 441 | 4835 | 8939.10 |
| 12 | DOLLY GIRL LUNCH BOX | SPACEBOY LUNCH BOX | 0.0230 | 0.6981 | 20.4484 | 441 | 4835 | 8939.10 |
| 13 | SET 12 COLOUR PENCILS DOLLY GIRL | SET 12 COLOUR PENCILS SPACEBOY | 0.0230 | 0.7551 | 20.2745 | 478 | 5072 | 3781.50 |
| 14 | SET 12 COLOUR PENCILS SPACEBOY | SET 12 COLOUR PENCILS DOLLY GIRL | 0.0230 | 0.6167 | 20.2745 | 478 | 5072 | 3781.50 |
| 15 | CLASSIC BICYCLE CLIPS | BICYCLE PUNCTURE REPAIR KIT | 0.0279 | 0.7895 | 20.1880 | 230 | 3508 | 6262.47 |

Table 4.2.7 reveals the top 15 highest frequency itemsets in Biweek 4 (11/5/2011 - 11/18/2011). The item ‘RABBIT NIGHT LIGHT’ emerged as the undisputed champion of this period with a remarkable 264 transactions. A staggering 7,184 of these charming night lights found new homes, illuminating a total sales value of 18,685.06. This product boasted an impressive support rate of 15.77% which implies its widespread popularity among customers. The high number of transactions and support rate suggest that it resonated with a broad customer base and it is possibly owing to its universal appeal as a night light or a decorative item. Additionally, the allure of nostalgia was evident in the popularity of ‘PAPER CHAIN KIT VINTAGE CHRISTMAS’ with 136 transactions resulting in 1,816 sales and a total sales value of 7,034.32. This item's popularity likely stems from its vintage charm that evokes memories of traditional holiday decorations. The support rate of this product at 8.12 % shows that even though it was not as prevalent as the 'RABBIT NIGHT LIGHT' but it still held a significant place in customers' hearts and homes. The items ‘JUMBO BAG PAISLEY PARK’ and ‘JUMBO BAG 50’S CHRISTMAS’ proved that practicality and style can go hand in hand with 135 transactions each, they both achieved sales in the thousands and total sales values of 5,630.48 and 7,838.61 respectively. Jumbo bags are known for their versatility and their popularity during this period suggests that customers were likely preparing for various activities or events.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | RABBIT NIGHT LIGHT | 0.1577 | 264 | 7184 | 18685.06 |
| 2 | PAPER CHAIN KIT 50'S CHRISTMAS | 0.1123 | 188 | 3397 | 15262.31 |
| 3 | HOT WATER BOTTLE KEEP CALM | 0.0872 | 146 | 1072 | 5639.12 |
| 4 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0812 | 136 | 1816 | 7034.32 |
| 5 | JUMBO BAG PAISLEY PARK | 0.0806 | 135 | 2020 | 5630.48 |
| 6 | JUMBO BAG 50'S CHRISTMAS | 0.0806 | 135 | 2543 | 7838.61 |
| 7 | JUMBO BAG RED RETROSPOT | 0.0795 | 133 | 3735 | 7167.05 |
| 8 | WHITE HANGING HEART T-LIGHT HOLDER | 0.0783 | 131 | 1900 | 5647.04 |
| 9 | WOODEN STAR CHRISTMAS SCANDINAVIAN | 0.0765 | 128 | 2617 | 863.37 |
| 10 | LUNCH BAG PAISLEY PARK | 0.0723 | 121 | 1277 | 2042.25 |
| 11 | WOODEN HEART CHRISTMAS SCANDINAVIAN | 0.0711 | 119 | 2287 | 750.63 |
| 12 | JUMBO BAG VINTAGE DOILY | 0.0711 | 119 | 2463 | 6609.78 |
| 13 | BAKING SET 9 PIECE RETROSPOT | 0.0711 | 119 | 562 | 2808.46 |
| 14 | HAND WARMER OWL DESIGN | 0.0681 | 114 | 1423 | 3443.78 |
| 15 | POPCORN HOLDER | 0..0657 | 110 | 7030 | 9469.71 |

Table 4.2.7 Top 15 highest frequency itemsets in Biweek 4 (11/5/2011 - 11/18/2011)

Table 4.2.8 provided below shows the product affinity analysis for Biweek 4 (11/5/2011 - 11/18/2011). The top rule connects the purchase of 'SET OF 3 WOODEN SLEIGH DECORATIONS' and 'SET OF 3 WOODEN STOCKING DECORATION' to the consequent acquisition of 'SET OF 3 WOODEN TREE DECORATIONS’. The rule is supported by a support rate of 2.09%, a confidence level of 77.78%, and a lift factor of 24.76, and unveils a fascinating pattern in customer choices. Although the support value is not extremely prevalent, this association still occurs in a notable number of transactions that indicates a specific subset of customers interested in these items. The result also showcases the charm of 'SET OF 3 WOODEN HEART DECORATIONS' and 'SET OF 3 WOODEN SLEIGH DECORATIONS’ and proves that they are a compelling combination for acquiring 'SET OF 3 WOODEN STOCKING DECORATION’. The 87.88% confidence level here suggests a strong affinity among shoppers for these festive items. This high confidence level indicates that when customers opt for 'SET OF 3 WOODEN HEART DECORATIONS' and 'SET OF 3 WOODEN SLEIGH DECORATIONS', there is a very strong likelihood they will also include 'SET OF 3 WOODEN STOCKING DECORATION' in their purchase.

Table 4.2.8 Product affinity analysis for Biweek 4 (11/5/2011 - 11/18/2011)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | SET OF 3 WOODEN SLEIGH DECORATION, SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN TREE DECORATION | 0.0209 | 0.7778 | 24.7593 | 738 | 4126 | 5697.87 |
| 2 | SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN SLEIGH DECORATION, SET OF 3 WOODEN STOCKING DECORATION | 0.0209 | 0.6667 | 24.7593 | 739 | 4126 | 5697.87 |
| 3 | SET OF 3 WOODEN STOCKING DECORATION, SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN SLEIGH DECORATION | 0.0209 | 0.8485 | 23.6338 | 738 | 4126 | 5697.87 |
| 4 | SET OF 3 WOODEN SLEIGH DECORATION | SET OF 3 WOODEN STOCKING DECORATION, SET OF 3 WOODEN TREE DECORATION | 0.0209 | 0.5833 | 23.6338 | 738 | 4126 | 5697.87 |
| 5 | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN SLEIGH DECORATION | SET OF 3 WOODEN STOCKING DECORATION | 0.0217 | 0.8788 | 23.4988 | 811 | 5098 | 7022.20 |
| 6 | SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN SLEIGH DECORATION | 0.0217 | 0.5800 | 23.4988 | 811 | 5098 | 7022.20 |
| 7 | SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN SLEIGH DECORATION, SET OF 3 WOODEN TREE DECORATION | 0.0209 | 0.5600 | 22.6885 | 738 | 4126 | 5697.87 |
| 8 | SET OF 3 WOODEN SLEIGH DECORATION, SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN STOCKING DECORATION | 0.0209 | 0.8485 | 22.6885 | 738 | 4126 | 5697.87 |
| 9 | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN STOCKING DECORATION | SET OF 3 WOODEN TREE DECORATION | 0.0202 | 0.7105 | 22.6184 | 805 | 4890 | 6712.63 |
| 10 | SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN STOCKING DECORATION | 0.0202 | 0.6429 | 22.6184 | 805 | 4890 | 6712.63 |
| 11 | CHARLOTTE BAG PINK POLKADOT | RED RETROSPOT CHARLOTTE BAG | 0.0202 | 0.7500 | 22.2833 | 469 | 6742 | 9502.03 |
| 12 | RED RETROSPOT CHARLOTTE BAG | CHARLOTTE BAG PINK POLKADOT | 0.0202 | 0.6000 | 22.2833 | 496 | 6742 | 9502.03 |
| 13 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN SLEIGH DECORATIONS | 0.0247 | 0.7857 | 21.8854 | 482 | 2726 | 3781.07 |
| 14 | SET OF 3 WOODEN SLEIGH DECORATIONS | SET OF 3 WOODEN TREE DECORATIONS | 0.0247 | 0.6875 | 21.8854 | 482 | 2726 | 3781.07 |
| 15 | SET OF 3 WOODEN HEART DECORATION, SET OF 3 WOODEN TREE DECORATION | SET OF 3 WOODEN STOCKING DECORATIONS | 0.0202 | 0.8182 | 21.8782 | 805 | 4890 | 6712.63 |

Table 4.2.9 below shows the top 15 highest frequency itemsets in Biweek 5 (11/19/2011 - 12/2/2011). This period covers the latter half of November that offers intriguing insights into consumer preferences during the transition from late autumn to early winter. The popularity of the item ‘RABBIT NIGHT LIGHT’ continued until the end of November. This enchanting night light earned a remarkable support score of 14.65% which signifies its immense popularity among shoppers seeking cosy and adorable lighting solutions. The ‘PAPER CHAIN KIT 50’S CHRISTMAS’ took a festive second place with 208 receipts, 3456 units sold, and a total sales value of 10,836.28. The high sales quantity suggests that customers were actively engaged in crafting and decorating for the holidays and its contribution to total sales value highlights its significance during this period. The third spot is taken by the ‘HOW WATER BOTTLE KEEP CALM’ with 155 receipts, 877 units sold, and a total sales value of 4,620.15 as this comforting item provided not just warmth in the chilling days of late November but also contributed significantly with a support value of 0.0962. Despite its simple nature, this item contributed significantly to sales revenue of 4620.15. These findings illuminate the shifting consumer preferences as the season transitions from autumn to early winter while 'RABBIT NIGHT LIGHT' remained a favourite.

Table 4.2.9 Top 15 highest frequency itemsets in Biweek 5 (11/19/2011 - 12/2/2011)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | RABBIT NIGHT LIGHT | 0.1465 | 236 | 6687 | 14025.41 |
| 2 | PAPER CHAIN KIT 50'S CHRISTMAS | 0.1291 | 208 | 3456 | 10836.28 |
| 3 | HOT WATER BOTTLE KEEP CALM | 0.0962 | 155 | 877 | 4620.15 |
| 4 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0906 | 146 | 1698 | 5191.32 |
| 5 | JUMBO BAG 50'S CHRISTMAS | 0.0875 | 141 | 1459 | 3081.97 |
| 6 | POPCORN HOLDER | 0.0863 | 139 | 6120 | 5275.14 |
| 7 | WHITE HANGING HEART T-LIGHT HOLDER | 0.0788 | 127 | 1727 | 4713.85 |
| 8 | HAND WARMER OWL DESIGN | 0.0701 | 113 | 932 | 1971.34 |
| 9 | HOT WATER BOTTLE TEA AND SYMPATHY | 0.0695 | 112 | 751 | 4572.69 |
| 10 | GARDENERS KNEELING PAD KEEP CALM | 0.0695 | 112 | 1115 | 2376.31 |
| 11 | VINTAGE DOILY TRAVEL SEWING KIT | 0.0664 | 107 | 818 | 1682.50 |
| 12 | JUMBO BAG RED RETROSPOT | 0.0664 | 107 | 1877 | 3901.56 |
| 13 | REGENCY CAKESTAND 3 TIER | 0.0658 | 106 | 459 | 6192.84 |
| 14 | 60 CAKE CASES VINTAGE CHRISTMAS | 0.0639 | 103 | 1656 | 1123.66 |
| 15 | WOODEN STAR CHRISTMAS SCANDINAVIAN | 0.0621 | 100 | 2386 | 864.28 |

Table 4.2.10 below displays the product affinity analysis for Biweek 5 (11/19/2011 - 12/2/2011). The ‘REGENCY TEA PLATE GREEN’ AND ‘REGENCY TEA PLATE ROSES’ lead the symphony of association rules The support of 0.0213 underlines their combined popularity while the confidence of 88.89% reveals that when one graces a tea table, the other is likely to follow suit. The lift score of 31.31 indicates a strong correlation between 'REGENCY TEA PLATE GREEN' and 'REGENCY TEA PLATE ROSES'. Next, the trio of ‘ROSES REGENCY TEACUP AND SAUCER’, ‘GREEN REGENCY TEACUP AND SAUCER’, and ‘PINK REGENCY TEACUP AND SAUCER’ forms a delightful tea party. This association suggests that customers who appreciate the 'GREEN' and 'ROSES' varieties of teacups are also inclined to complete the set with the 'PINK' version. This association suggests that customers who appreciate the 'GREEN' and 'ROSES' varieties of teacups are also inclined to complete the set with the 'PINK' version. The strong associations between specific tea plates and teacup varieties reflect the importance of product synergy and aesthetic appeal in customer choices. By understanding these affinities, businesses could optimize their product offerings, bundling strategies, and marketing efforts to cater to customers' desires for complementary or coordinated items.

Table 4.2.10 Product affinity analysis for Biweek 5 (11/19/2011 - 12/2/2011)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | REGENCY TEA PLATE GREEN | REGENCY TEA PLATE ROSES | 0.0213 | 0.8889 | 31.3056 | 239 | 2222 | 3908.34 |
| 2 | REGENCY TEA PLATE ROSES | REGENCY TEA PLATE GREEN | 0.0213 | 0.7500 | 31.3056 | 239 | 2222 | 3908.34 |
| 3 | ROSES REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER | PINK REGENCY TEACUP AND SAUCER | 0.0213 | 0.8276 | 28.2633 | 633 | 4428 | 14057.36 |
| 4 | PINK REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER | 0.0213 | 0.7273 | 28.2633 | 633 | 4428 | 14057.36 |
| 5 | ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0213 | 0.8571 | 26.1081 | 633 | 4428 | 14057.36 |
| 6 | GREEN REGENCY TEACUP AND SAUCER | ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER | 0.0213 | 0.6486 | 26.1081 | 633 | 4428 | 14057.36 |
| 7 | PINK REGENCY TEACUP AND SAUCER | GREEN REGENCY TEACUP AND SAUCER | 0.0248 | 0.8485 | 25.8444 | 399 | 2792 | 8880.52 |
| 8 | GREEN REGENCY TEACUP AND SAUCER | PINK REGENCY TEACUP AND SAUCER | 0.0248 | 0.7568 | 25.8444 | 399 | 2792 | 8880.52 |
| 9 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS | 0.0248 | 0.8235 | 24.4241 | 494 | 2659 | 3652.55 |
| 10 | SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATIONS | 0.0248 | 0.7368 | 24.4241 | 494 | 2659 | 3652.55 |
| 11 | ROLL WRAP 50'S RED CHRISTMAS | ROLL WRAP 50'S CHRISTMAS | 0.0240 | 0.8710 | 23.9410 | 468 | 3968 | 5624.27 |
| 12 | ROLL WRAP 50'S CHRISTMAS | ROLL WRAP 50'S RED CHRISTMAS | 0.0240 | 0.6585 | 23.9410 | 468 | 3968 | 5624.27 |
| 13 | SET 12 COLOUR PENCILS SPACEBOY | SET 12 COLOUR PENCILS DOLLY GIRL | 0.0257 | 0.7632 | 23.8911 | 478 | 5072 | 3781.50 |
| 14 | SET 12 COLOUR PENCILS DOLLY GIRL | SET 12 COLOUR PENCILS SPACEBOY | 0.0257 | 0.8056 | 23.8911 | 478 | 5072 | 3781.50 |
| 15 | CHRISTMAS HANGING STAR WITH BELL | HANGING HEART WITH BELL | 0.0204 | 0.6970 | 23.8026 | 357 | 8560 | 3877.30 |

Table 4.2.11 shows the top 15 highest frequency itemsets in Biweek 6 (12/3/2011 - 12/15/2011). In the hustle and bustle of the holiday shopping season, certain items emerged as the shining stars of retail. The ‘RABBIT NIGHT LIGHT’ appeared in 93 receipts and this delightful night light hopped into the hearts of customers which resulted in 3380 sales and an impressive total sales value of 7584.48. Speaking of nostalgia, the PAPER CHAIN KIT 50'S CHRISTMAS took shoppers on a trip down memory lane. Found in 86 receipts, this kit allowed customers to create their retro holiday decorations. The impressive sales quantity of 1545 suggests that customers were eager to create their own vintage holiday decorations. The 'PAPER CHAIN KIT 50'S CHRISTMAS' contributed significantly to holiday sales with a substantial total sales value. The ‘HAND WARMER OWL DESIGN’ was a warm and fuzzy addition to this list. It kept 689 shoppers’ cosy during the cold season and its total sales value of 1487.66 and a support value of 0.0952 proved that staying warm can be a stylish affair. These insights into the top-performing items during this period demonstrate the convergence of practicality, nostalgia, and holiday spirit in customer preferences. The 'RABBIT NIGHT LIGHT' added a touch of whimsy to holiday decor, the 'PAPER CHAIN KIT 50'S CHRISTMAS' brought back cherished memories, and the 'HAND WARMER OWL DESIGN' combined style with warmth while all of these contributed to a vibrant holiday shopping season.

Table 4.2.11 Top 15 highest frequency itemsets in Biweek 6 (12/3/2011 - 12/15/2011)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Itemsets | Support | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | RABBIT NIGHT LIGHT | 0.1321 | 93 | 3380 | 7584.48 |
| 2 | PAPER CHAIN KIT 50'S CHRISTMAS | 0.1222 | 86 | 1545 | 5328.99 |
| 3 | HAND WARMER OWL DESIGN | 0.0952 | 67 | 689 | 1487.66 |
| 4 | HOT WATER BOTTLE KEEP CALM | 0.0938 | 66 | 480 | 2902.84 |
| 5 | PAPER CHAIN KIT VINTAGE CHRISTMAS | 0.0866 | 61 | 666 | 2074.38 |
| 6 | GARDENERS KNEELING PAD KEEP CALM | 0.0810 | 57 | 573 | 1372.97 |
| 7 | CHOCOLATE HOT WATER BOTTLE | 0.0795 | 56 | 360 | 2367.55 |
| 8 | BICYCLE PUNCTURE REPAIR KIT | 0.0795 | 56 | 1109 | 2265.91 |
| 9 | POPCORN HOLDER | 0.0739 | 52 | 4563 | 4084.27 |
| 10 | HAND WARMER RED LOVE HEART | 0.0724 | 51 | 305 | 654.71 |
| 11 | HAND WARMER BIRD DESIGN | 0.0710 | 50 | 668 | 1375.92 |
| 12 | JUMBO BAG 50'S CHRISTMAS | 0.0710 | 50 | 783 | 1636.89 |
| 13 | SCOTTIE DOG HOT WATER BOTTLE | 0.0682 | 48 | 304 | 1587.65 |
| 14 | REGENCY CAKESTAND 3 TIER | 0.0682 | 48 | 388 | 5170.50 |
| 15 | GARDENERS KNEELING PAD CUP OF TEA | 0.0682 | 48 | 380 | 706.44 |

The last table in this chapter, table 4.2.12 illustrates the product affinity analysis for Biweek 6 (12/3/2011 - 12/15/2011). The ‘PINK DOG BOWL’ and ‘PINK CAT BOWL’ shared a delightful association rule with a support of 2.13% which explains that pet lovers who bought one were quite likely to purchase the other. The confidence level of 88.24% indicates a strong connection. In the realm of holiday decorations, the association rules revealed some interesting patterns. The trio of ‘SET OF 3 WOODEN HEART DECORATIONS’, ‘SET OF 3 WOODEN STOCKING DECORATION’, and ‘SET OF 3 WOODEN TREE DECORATIONS’ showcased a symphony of festive spirit. This suggests that customers decorating their homes for the holidays preferred to purchase all three types of wooden decorations. For those with a penchant for baking, an intriguing association was observed between ‘SET OF 6 SNACK LOAF BAKING CASES’ and ‘SET OF 12 MINI LOAF BAKING CASES’. The confidence level of 94.12% suggested that bakers who opted for one type of baking case often expanded their collection with the other. This association reflects the preferences of bakers who appreciated having a variety of baking case sizes at their disposal. These product affinity patterns offer valuable insights into customer behaviour during the holiday season. Pet owners showed a penchant for matching pet bowls, holiday decorators preferred a coordinated set of wooden decorations, and baking enthusiasts embraced a range of baking case options.

Table 4.2.12 Product affinity analysis for Biweek 6 (12/3/2011 - 12/15/2011)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Antecedent | Consequent | Support | Confidence | Lift | Total Receipt Count | Total Sales Quantity | Total Sales Value |
| 1 | PINK DOG BOWL | PINK CAT BOWL | 0.0213 | 0.8824 | 34.5098 | 110 | 329 | 483.06 |
| 2 | PINK CAT BOWL | PINK DOG BOWL | 0.0213 | 0.8333 | 34.5098 | 110 | 329 | 483.06 |
| 3 | SET OF 3 WOODEN HEART DECORATIONS, SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATIONS | 0.0213 | 0.9375 | 30.0000 | 805 | 4890 | 6712.63 |
| 4 | SET OF 3 WOODEN SLEIGH DECORATIONS, SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN TREE DECORATIONS | 0.0213 | 0.9375 | 30.0000 | 738 | 4126 | 5697.87 |
| 5 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN HEART DECORATIONS, SET OF 3 WOODEN STOCKING DECORATIONS | 0.0213 | 0.6818 | 30.0000 | 805 | 4890 | 6712.63 |
| 6 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN SLEIGH DECORATIONS, SET OF 3 WOODEN STOCKING DECORATIONS | 0.0213 | 0.6818 | 30.0000 | 738 | 4126 | 5697.87 |
| 7 | SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN HEART DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | 0.0213 | 0.6522 | 28.6957 | 805 | 4890 | 6712.43 |
| 8 | SET OF 3 WOODEN STOCKING DECORATIONS | SET OF 3 WOODEN SLEIGH DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | 0.0213 | 0.6522 | 28.6957 | 738 | 4126 | 5697.87 |
| 9 | SET OF 3 WOODEN HEART DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS | 0.0213 | 0.9375 | 28.6957 | 805 | 4890 | 6712.63 |
| 10 | SET OF 3 WOODEN SLEIGH DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS | 0.0213 | 0.9375 | 28.6957 | 738 | 4126 | 5697.87 |
| 11 | SET OF 6 SNACK LOAF BAKING CASES | SET OF 12 MINI LOAF BAKING CASES | 0.0227 | 0.9412 | 26.5035 | 388 | 3503 | 3042.40 |
| 12 | SET OF 12 MINI LOAF BAKING CASES | SET OF 6 SNACK LOAF BAKING CASES | 0.0277 | 0.6400 | 26.5035 | 388 | 3503 | 3042.40 |
| 13 | SET OF 3 WOODEN STOCKING DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN HEART DECORATIONS | 0.0213 | 0.7895 | 26.4662 | 805 | 4890 | 6712.63 |
| 14 | SET OF 3 WOODEN HEART DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS, SET OF 3 WOODEN TREE DECORATIONS | 0.0213 | 0.7143 | 26.4462 | 805 | 4890 | 6712.63 |
| 15 | SET OF 3 WOODEN TREE DECORATIONS | SET OF 3 WOODEN STOCKING DECORATIONS | 0.0270 | 0.8636 | 26.4348 | 495 | 2659 | 3652.55 |

According to the results of the six biweeks, certain items such as 'RABBIT NIGHT LIGHT' and 'PAPER CHAIN KIT 50'S CHRISTMAS' consistently emerged as top performers. Their popularity were maintained across the entire period by exhibiting high support values. Also, 'JUMBO BAG RED RETROSPOT' and 'REGENCY CAKESTAND 3 TIER' consistently ranked among the top items which reflects sustained customer interest. In terms of support, some items remained popular across biweeks. For instance, the support value of ‘RABBIT NIGHT LIGHT' reached 0.1577 in the fourth biweek. In fact, seasonality played a significant role in shaping customer preferences. Christmas-related items like 'PAPER CHAIN KIT 50'S CHRISTMAS' experienced rising in popularity during the holiday season. In the context of price point influence, items like 'WOODEN STAR CHRISTMAS SCANDINAVIAN' and 'LARGE WHITE HEART OF WICKER' stood out with high total sales values which implies that they might have had higher price points or were purchased in larger quantities.

For the discussion of the association rules across the six biweeks, association rules involve items related to dolls, glass apothecary bottles, and teacup and saucer sets dominated the ranking of the top association rules initially. Starting from the second biweek, the association rules revolve around wooden decorative items. There are high-confidence rules indicating that when customers purchase specific combinations of wooden decorations. This trend lasted for few biweeks and brought significant boosting of sales. Overall, the findings suggest that customers exhibit certain consistent shopping patterns especially the purchasing of related items such as dolls, glass bottles, wooden decorations, and teacup and saucer sets.

# CHAPTER 5 Conclusion

In the domain of studying customer purchasing behavior, it is essential to discover and understand the changing pattern of the behavior over time. This project has initiated the path of unveiling the optimal time period to conduct a Market Basket Analysis to perform effective mining of association rules. First and foremost, the association rules derived from the transactional data reflects the customer purchasing behavior and the consideration of different time period is an addition to the study. A comparison of association rules from different time periods will be carried out to discover any hidden trends in each period. This could reveal the importance of regularly revisiting and updating strategies for product placement, cross-selling, and marketing campaigns. By leveraging the time period of conducting an MBA, businesses are allowed to implement more agile and effective strategies such as tailoring marketing campaigns to align with changing consumer preferences to ensure promotional efforts resonate with customers in that particular period. In conclusion, this project illustrates the dynamic nature of consumer behavior and the potential benefits of adapting to these changes. This iterative process of analysis, interpretation, and action is vital for businesses striving to stay competitive and relevant. Businesses could position themselves to thrive in the ever-evolving retail landscape by embracing data analytics and harnessing the power of association rule mining.

For current progress for the implementation of the project, several milestones have been reached. For instance, the preparation of the monthly and biweekly data is carried out successfully and data pre-processing operations are performed on the extracted data. Subsequently, the mining of association rules for both monthly and biweekly data is completed as well and the results are visualized in the form of tables. In the context of analysis, the results of monthly data such as the top 15 highest frequency itemsets and product affinity analysis are interpreted in detail and significant insights are highlighted. However, the analysis of the results from biweekly data is not yet completed due to time constraints and the work will be continued in the next report, which is FYP2.

# REFERENCES

1. S. Chaudhary, “Understanding market basket analysis in data mining,” *Understanding Market Basket Analysis in Data Mining*, 11-Feb-2022. [Online]. Available: https://www.turing.com/kb/market-basket-analysis. [Accessed: 28-Feb-2023].

2. “Association rule mining: An overview and its applications,” *upGrad blog*, 23-Nov-2022. [Online]. Available: https://www.upgrad.com/blog/association-rule-mining-an-overview-and-its-applications/. [Accessed: 28-Feb-2023].

3. T. Kristiana, S. A. Putri, Nurmalasari, R. I. Handayani, N. Merlina, and N. Yunita, “Association rule implementation using algorithm apriori to analize fishing pattern in Indonesia,” *Journal of Physics: Conference Series*, vol. 1641, no. 1, p. 012072, 2020.

4. R. Ruswati, A. I. Gufroni, and R. Rianto, “Associative analysis data mining pattern against traffic accidents using apriori algorithm,” *Scientific Journal of Informatics*, vol. 5, no. 2, pp. 91–104, 2018.

5. K. J. Cios, R. W. Swiniarki, W. Pedrycz, and L. A. Kurgan, “Data Mining: A Knowledge Discovery Approach,” 2007.

6. A. N. Sagin and B. Ayvaz, “Determination of association rules with Market Basket Analysis: Application in the retail sector,” *Southeast Europe Journal of Soft Computing*, vol. 7, no. 1, 2018.

7. R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” *IBM Almaden Research Center*, vol. 1215, pp. 487–499, 1994.

8. J. Han, M. Kamber, and J. Pei, “Data mining concepts and techniques third edition,” *The Morgan Kaufmann Series in Data Management Systems*, 2011

9. W. P. Nurmayanti, H. M. Sastriana, A. Rahim, M. Gazali, R. H. Hirzi, Z. Ramdani, and M. Malthuf, “Market basket analysis with apriori algorithm and frequent pattern growth (FP-growth) on outdoor product sales data,” *International Journal of Educational Research & Social Sciences*, vol. 2, no. 1, pp. 132–139, 2021.

10. Hiding Association Rules over Privacy Preserving Distributed Data Mining - Scientific Figure on ResearchGate. Available: https://www.researchgate.net/figure/Figure-2-Pseudo-code-for-Apriori-algorithm-11\_fig1\_287993846 [Accessed: 6-Mar-2023]

11. J. Korstanje, “The FP growth algorithm,” Medium, 28-Sep-2021. [Online]. Available: https://towardsdatascience.com/the-fp-growth-algorithm-1ffa20e839b8. [Accessed: 06-Mar-2023].

12. Y. Verma, “Apriori vs FP-growth in Market Basket Analysis - A Comparative Guide,” Analytics India Magazine, 11-Nov-2021. [Online]. Available: https://analyticsindiamag.com/apriori-vs-fp-growth-in-market-basket-analysis-a-comparative-guide/#:~:text=Apriori%20uses%20candidate%20generation%20where,number%20of%20items%20is%20larger. [Accessed: 06-Mar-2023].

13 A. A. Aldino, E. D. Pratiwi, Setiawansyah, S. Sintaro, and A. Dwi Putra, “Comparison of market basket analysis to determine customer purchasing patterns using FP-growth and APRIORI ALGORITHM,” *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, 2021.

14. M. Hossain, A. H. Sattar, and M. K. Paul, “Market basket analysis using Apriori and FP growth algorithm,” *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019.

15. S. Z. Ansari, “Market basket analysis : Trend analysis of association rules in different time periods,” *Handle Proxy*, 30-Jul-2019. [Online]. Available: http://hdl.handle.net/10362/80955. [Accessed: 06-Mar-2023].

16. V. Papavasileiou and A. Tsadiras, “Time Variations of Association rules in Market Basket Analysis,” *IFIP Advances in Information and Communication Technology*, pp. 36–44, 2011.

17. Alfiqra and A. U. Khasanah, “Implementation of market basket analysis based on overall variability of association rule (OCVR) on product marketing strategy,” *IOP Conference Series: Materials Science and Engineering*, vol. 722, no. 1, p. 012068, 2020.

# APPENDIX

**FINAL YEAR PROJECT WEEKLY REPORT**

*(Project I)*

|  |  |
| --- | --- |
| **Trimester, Year: Y3S1** | **Study week no.: 5** |
| **Student Name & ID: Teng Wen Foong, 20ACB01778** | |
| **Supervisor: Dr Kh’ng Xin Yi** | |
| **Project Title: Market Basket Analysis: Analysis of Association Rules in Different Time Periods** | |

A close-up of a signature

Description automatically generated

|  |
| --- |
| **1. WORK DONE**  The coding for conducting market basket analysis is completed using Jupyter Notebook in Python programming language. It is included importing the required libraries, data preprocessing, data visualization, and the Apriori algorithm. A test run of the coding has also been carried out using dummy datasets found online. |
| **2. WORK TO BE DONE**  For further work, an improvement on the coding will be carried out to make the overall coding looks better and tidy. One of the possible improvements will be the reengineering of the code for data preprocessing part since there is no fixed format for the dataset. |
| **3. PROBLEMS ENCOUNTERED**  The issue that has been encountered is the inconsistent format of the dataset to be the input of the market basket analysis. In this case, the available dummy datasets that fit the coding is not that many. |
| **4. SELF EVALUATION OF THE PROGRESS**  The project's advancement is proceeding at a steady pace, aligning perfectly with the projected timeline and milestones |



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor’s signature Student’s signature

**FINAL YEAR PROJECT WEEKLY REPORT**

*(Project I)*

|  |  |
| --- | --- |
| **Trimester, Year: Y3S1** | **Study week no.: 7** |
| **Student Name & ID: Teng Wen Foong, 20ACB01778** | |
| **Supervisor: Dr Kh’ng Xin Yi** | |
| **Project Title: Market Basket Analysis: Analysis of Association Rules in Different Time Periods** | |

A close-up of a signature

Description automatically generated

|  |
| --- |
| **1. WORK DONE**  Restructuring of the coding is completed as well as the format of the dataset is fixed. |
| **2. WORK TO BE DONE**  FP-Growth algorithm will be applied to compare its performance with Apriori algorithm and visualization have to be advanced for better insights, |
| **3. PROBLEMS ENCOUNTERED**  The performance of Apriori algorithms is not satisfying in terms of processing time. |
| **4. SELF EVALUATION OF THE PROGRESS**  The project's progress remains on track and is in accordance with the established goals |



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor’s signature Student’s signature

**FINAL YEAR PROJECT WEEKLY REPORT**

*(Project I)*

|  |  |
| --- | --- |
| **Trimester, Year: Y3S1** | **Study week no.: 9** |
| **Student Name & ID: Teng Wen Foong, 20ACB01778** | |
| **Supervisor: Dr Kh’ng Xin Yi** | |
| **Project Title: Market Basket Analysis: Analysis of Association Rules in Different Time Periods** | |

A close-up of a signature

Description automatically generated

|  |
| --- |
| **1. WORK DONE**  The coding for applying FP-Growth algorithm is done and better visualization is performed as well. |
| **2. WORK TO BE DONE**  For further work, a real dataset will be given to run on the completed coding and observe if any modification of the coding is needed. |
| **3. PROBLEMS ENCOUNTERED**  No. |
| **4. SELF EVALUATION OF THE PROGRESS**  The project has managed to maintain a consistent pace of progress |



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor’s signature Student’s signature

**FINAL YEAR PROJECT WEEKLY REPORT**

*(Project I)*

|  |  |
| --- | --- |
| **Trimester, Year: Y3S1** | **Study week no.: 10** |
| **Student Name & ID: Teng Wen Foong, 20ACB01778** | |
| **Supervisor: Dr Kh’ng Xin Yi** | |
| **Project Title: Market Basket Analysis: Analysis of Association Rules in Different Time Periods** | |

A close-up of a signature

Description automatically generated

|  |
| --- |
| **1. WORK DONE**  Some modifications of the coding are done to suit the real dataset better. |
| **2. WORK TO BE DONE**  Visualization of the association rules mined has to be performed as further works and the rules are going to be exported to an excel file. |
| **3. PROBLEMS ENCOUNTERED**  Failed to extract biweekly transaction directly from the dataset. |
| **4. SELF EVALUATION OF THE PROGRESS**  The project has encountered some delays that are impacting the original timeline |



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor’s signature Student’s signature

**FINAL YEAR PROJECT WEEKLY REPORT**

*(Project I)*

|  |  |
| --- | --- |
| **Trimester, Year: Y3S1** | **Study week no.: 11** |
| **Student Name & ID: Teng Wen Foong, 20ACB01778** | |
| **Supervisor: Dr Kh’ng Xin Yi** | |
| **Project Title: Market Basket Analysis: Analysis of Association Rules in Different Time Periods** | |

A close-up of a signature

Description automatically generated

|  |
| --- |
| **1. WORK DONE**  The visualization of the rules is accomplished and an excel file is produced. |
| **2. WORK TO BE DONE**  Analysis of the rules are going to be performed to figure out if any significant insights and trends, finish the report writing. |
| **3. PROBLEMS ENCOUNTERED**  No. |
| **4. SELF EVALUATION OF THE PROGRESS**  The project is making steady progress, adhering to the anticipated timeline. |



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor’s signature Student’s signature

**FINAL YEAR PROJECT WEEKLY REPORT**

*(Project I)*

|  |  |
| --- | --- |
| **Trimester, Year: Y3S1** | **Study week no.: 12** |
| **Student Name & ID: Teng Wen Foong, 20ACB01778** | |
| **Supervisor: Dr Kh’ng Xin Yi** | |
| **Project Title: Market Basket Analysis: Analysis of Association Rules in Different Time Periods** | |

A close-up of a signature

Description automatically generated

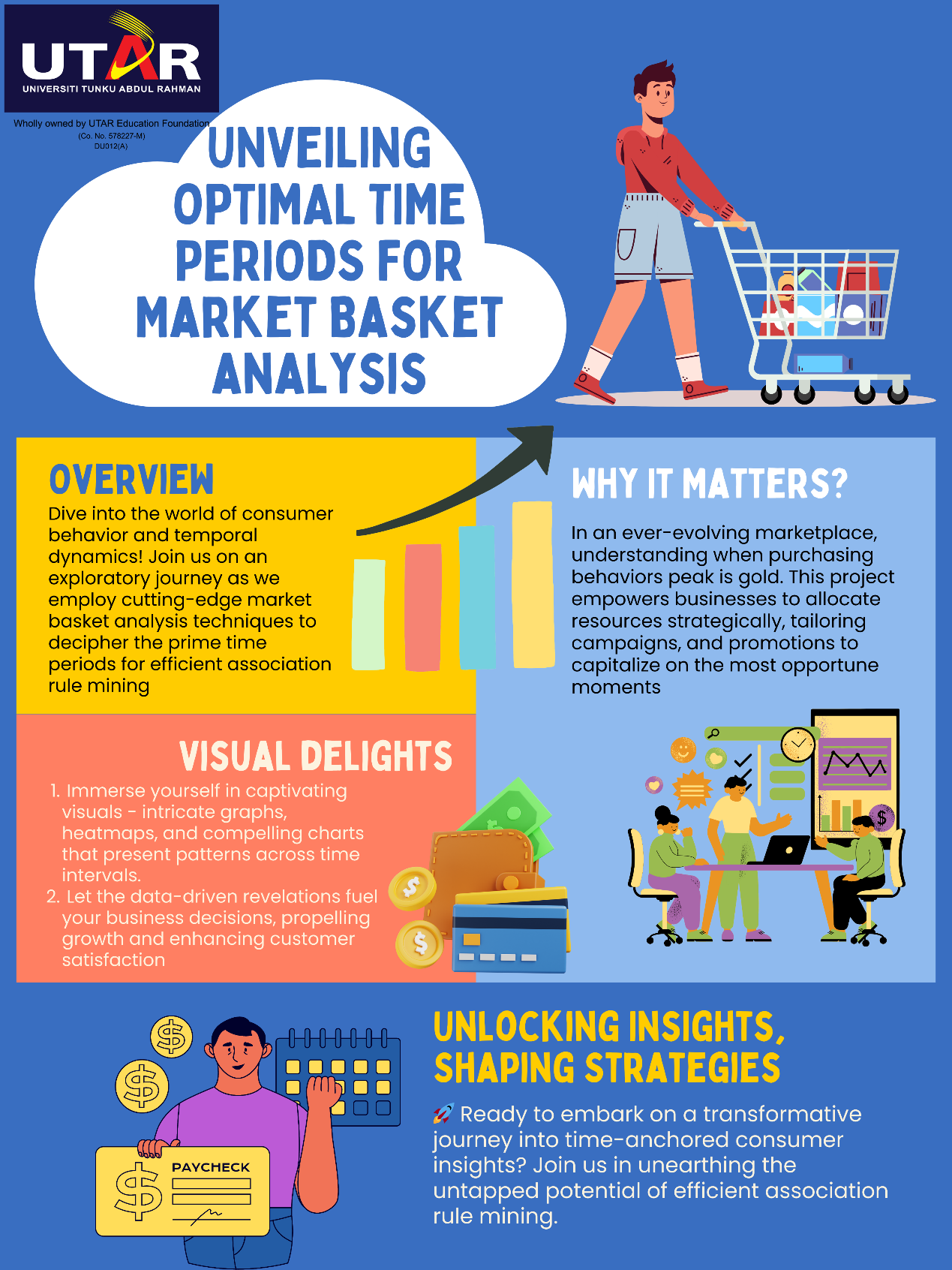
|  |
| --- |
| **1. WORK DONE**  Analysis of the result is done and valuable insights are extracted. Also, the writing of the report is completed. |
| **2. WORK TO BE DONE**  No. |
| **3. PROBLEMS ENCOUNTERED**  No. |
| **4. SELF EVALUATION OF THE PROGRESS**  The overall progress is favorably, with substantial accomplishments in each stage. |



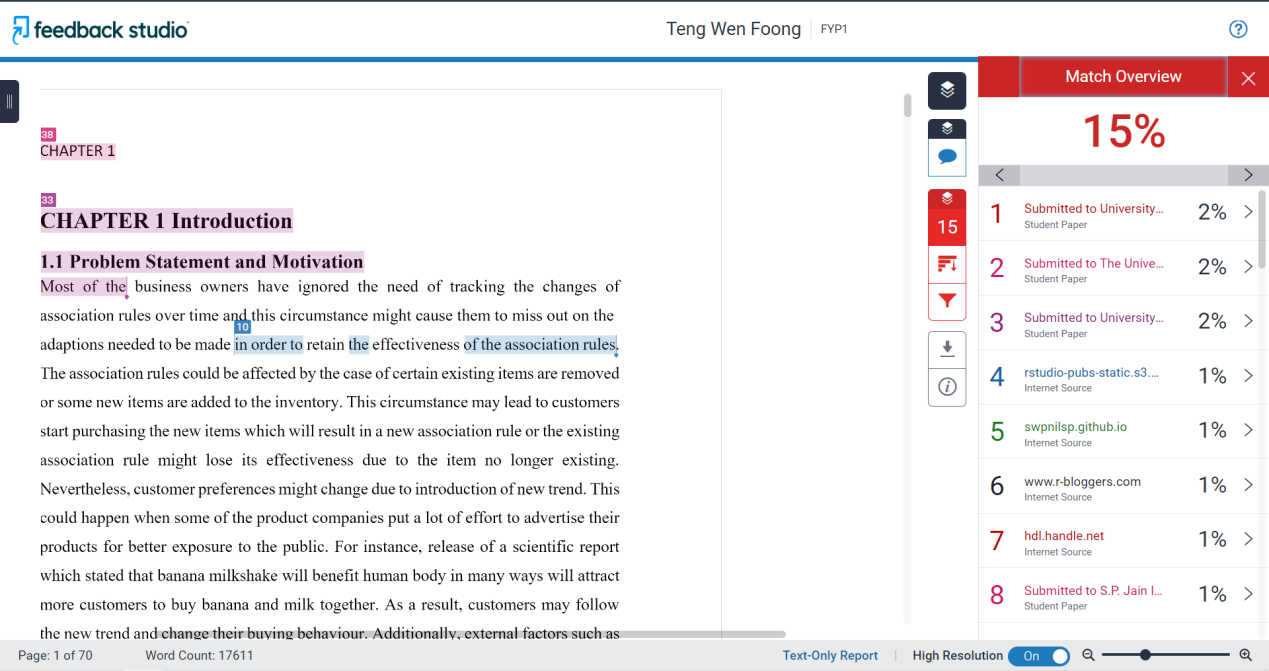
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor’s signature Student’s signature

# POSTER

****

# PLAGIARISM CHECK RESULT



|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| **Form Title: Supervisor’s Comments on Originality Report Generated by Turnitin**  **for Submission of Final Year Project Report (for Undergraduate Programmes)** | | | |
| Form Number: FM-IAD-005 | Rev No.: 0 | Effective Date: 01/10/2013 | Page No.: 1of 1 |



**FACULTY OF INFORMATION AND COMMUNICATION**

**TECHNOLOGY**

|  |  |
| --- | --- |
| **Full Name(s) of**  **Candidate(s)** | Teng Wen Foong |
| **ID Number(s)** | 20ACB01778 |
| **Programme / Course** | Bachelor of Computer Science (Honours) |
| **Title of Final Year Project** | Market Basket Analysis: Analysis of Association Rules in Different Time Period |

|  |  |
| --- | --- |
| **Similarity** | **Supervisor’s Comments**  **(Compulsory if parameters of originality exceed the limits approved by UTAR)** |
| **Overall similarity index: 15 % Similarity by source**  Internet Sources: 10 % Publications: 2 %  Student Papers: 12 % |  |
| **Number of individual sources listed** of more than 3% similarity: 0 |  |
| **Parameters of originality required, and limits approved by UTAR are as Follows:**  **(i) Overall similarity index is 20% and below, and**  **(ii) Matching of individual sources listed must be less than 3% each, and**  **(iii) Matching texts in continuous block must not exceed 8 words**  ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.*** | |

Note: Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

***Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.***

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |
| --- | --- | --- |
| Signature of Supervisor |  | Signature of Co-Supervisor |
| Name: \_\_Kh’ng Xin Yi\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  | Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Date: \_\_6/9/2023\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  | Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

# FYP 1 CHECKLIST

****

**UNIVERSITI TUNKU ABDUL RAHMAN**

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

**CHECKLIST FOR FYP1 THESIS SUBMISSION**

|  |  |
| --- | --- |
| Student ID | 20ACB01778 |
| Student Name | Teng Wen Foong |
| Supervisor Name | Dr. Kh’ng Xin Yi |

|  |  |
| --- | --- |
| **TICK (√)** | **DOCUMENT ITEMS**  Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item. |
| **√** | Title Page |
| **√** | Signed form of the Declaration of Originality |
| **√** | Acknowledgment |
| **√** | Abstract |
| **√** | Table of Contents |
| **√** | List of Figures (if applicable) |
| **√** | List of Tables (if applicable) |
| **√** | List of Symbols (if applicable) |
| **√** | List of Abbreviations (if applicable) |
| **√** | Chapters / Content |
| **√** | Bibliography (or References) |
| **√** | All references in bibliography are cited in the thesis, especially in the chapter of literature review |
| **√** | Appendices (if applicable) |
| **√** | Poster |
| **√** | Signed Turnitin Report (Plagiarism Check Result – Form Number: FM-IAD-005) |
| **√** | I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report. |

\*Include this form (checklist) in the thesis (Bind together as the last page)

|  |
| --- |
| I, the author, have checked and confirmed all the items listed in the table are included in my report.  \_\_\_\_\_\_\_\_\_\_\_\_  (Signature of Student)  Date: 5/9/2023 |