كلية الدراسات العليا

Sudan University of Science and Technology

College of Graduate Studies

Using Frequent Pattern Growth Algorithm in Analyzing Sudanese Shopping Behavior

إستخدام خوارزمية نمو النمط المتكرر في تحليل سلوك التسوق للسودانيين

A thesis submitted in partial fulfillment of the requirement of M.Sc Degree in Computer science

By: Alargum Gassim Alzain Abo Algassim

Supervised by: Dr. Wafaa Faisal Mukhtar

October 2019

الآيـــة

(وَلَسَوۡفَ يُعۡطِيكَ رَبُّكَ فَتَرۡضَىٰٓ)

صدق الله العظيم

سورة الضحى (الآية 5)

Dedication

I dedicate this work

To my dear mother

Rabaa Shaikh Eldein

Acknowledgement

First I would like to thank God – Allah for what I achieve and completing this research.

I'd like also to express my deepest thanks to the generous staff of SUST, particularly Collage of Computer Science and Information Technology staff. I would like to express my sincere gratitude to **Prof. Izzeldin Mohammed Osman, Prof. Mohammed El hafiz, Dr. Tallat M. Wahby and Dr. Abo Agalah Babiker**.

Beside those, I cannot express enough thanks to My Supervisor **Dr. Wafaa Faisal** for her guidance, contributions and overall insights in this field have made this an inspiring experience for me on this research.

Last but not least, I have to confirm that my completion of this project could not have been accomplished without the support of my family, my friends, and also my colleagues' specially Alaa Eldein, Mohammed Elfateh and Mohammed Suleiman.

Alargum Gassim

Abstract

Data mining is an important technique to discover frequent items in customer shopping basket. Such information can be used as a basis for decisions about marketing activities such as promotional support, inventory control and cross-sale campaigns.

The main objective of this research is analyzing Sudanese shopping behavior: a case study Aldooma supermarket and figuring out the commodities that are sold together, the data source is ORACLE database backup file which are collected from Aldooma supermarket's sales points system, the results performed by using Frequent Pattern Growth Algorithm in Rapidminer tool.

The researcher conducted many experiments and selected the best results which contained the longest frequent itemsets sold together. The best results represented in 12 Association rules with confidence 0.8 and support 0.004.

The techniques which applied in this research are useful for the supermarket owner or the decision maker who can use them to grow their customer base and build stronger customer relationships to turn inventory into cash.

المستخلص

يُعد تنقيب البيانات تقنية مهمة لاكتشاف السلع المتكررة في سلة تسوق العملاء. يمكن استخدام هذه المعلومات كأساس لاتخاذ القرارات المتعلقة بأنشطة التسويق مثل الدعم الترويجي ومراقبة المخزون وحملات البيع المتبادل.

الهدف الرئيسي من هذا البحث هو تحليل سلوك التسوق للسودانيين: دراسة حالة سوبر ماركت الدومة، من أجل معرفة السلع التي تم بيعها معًا ، مصدر البيانات هو ملف نسخ إحتياطي لقاعدة بيانات أوراكل تم جمعها من نظام نقاط البيع في سوبرماركت الدومة . النتائج تم الحصول عليها بإستخدام خوارزمية نمو النمط المتكرر(FP Growth) بواسطة الأداة Rapidminer.

أجرى الباحث العديد من التجارب وإختار أفضل النتائج حيث تضمنت على أطول مجموعة سلع تباع بشكل متكرر متمثلة في 12 قاعدة إرتباط بثقة 0.8 Confidence ودعم 0.004 Support.

تعتبر التقنيات التي طُبقت في هذا البحث مفيدة لمالك السوبر ماركت أو صانع القرار الذي يمكنه استخدامها لتطوير قاعدة عملائه وبناء علاقات أقوى مع العملاء لتحويل المخزون إلى نقد .

Table of Contents

List of Tables

List of Figures

CHAPTER ONE
INTRODUCTION

1.1 Research Background

Nowadays there is a vast amount of data is collected and stored daily. There is an important need to analyze this data, but without any analytical tool, that seems impossible. It's has led to the development of Knowledge Discovery in Databases (KDD) which transforms the low level data to high level knowledge. KDD consists of various processes at different steps and data mining is one of those processes.
The main aim of data mining process and techniques such as association rule mining (ARM) termed as market basket analysis is to extract information from a dataset and transform it into an understandable form in order to aid decision making especially in marketing strategic and research.

Data mining helps retailers to prevent their business from taking non calculative decisions, losing customer and falling behind competition to the peers. There are varied benefits that data mining services offers like: discovering customer shopping behavior, improve the quality of customer service, achieve better customer satisfaction and reduce the cost of business. Therefore, it could be used for predicting certain items that will be sold together (as a set) based on some criteria via a prediction model built through one of the data mining tools.

The main objective of the research is to see which frequent itemsets sold together in Aldooma supermarket and how to exploit these relations by marketing activities such as promotional support, inventory control and cross-sale campaigns. The techniques which applied in this research are useful for the supermarket owner or the decision maker who can use them to grow their customer base and build stronger customer relationships to turn inventory into cash.

1.2 Problem Statement

In the shops and supermarkets it is easy to turn cash into inventory, but the challenge is how to turn inventory into cash making use of the stored (transactional) data about sales.

Advanced technology made a possibility for retailers to gather information on their customers buying behavior and what they buy. Transactional data is used for mining useful information on co purchases and adjusting marketing activities accordingly.

The issue in question is figuring out the commodities that were usually sold together to improve products assortment and maximize profits.

1.3 Research Scope

The research is conducted on Aldooma supermarket (retail sector) to analyze the stored data in the period (JAN/2014 to MAY/2015) for finding the frequent itemsets sold together.

1.4 Research Objectives

- Determine the frequent itemsets sold together.
- Building an FP-tree model for finding frequent item sets based on analysis of past transactions performed.

1.5 Research Significance

It is very important for shopkeeper to know what their customers are buying. Some products have higher affinity to be sold together and hence the retailer can benefit from this affinity if special offers and promotions are developed for these products. Data mining techniques are highly valued for the useful information they provide so that the shopkeeper can serve customers better and generate higher profits.

1.6 Research Methodology

The present research discovers frequent item sets in Supermarket system (ORACLE backup file) using a suitable association rule mining techniques for analyzing Sudanese shopping behavior: Aldooma supermarket as a case study.

The methodology will follow the steps of knowledge discovery (KDD) as described in figure 1.1.

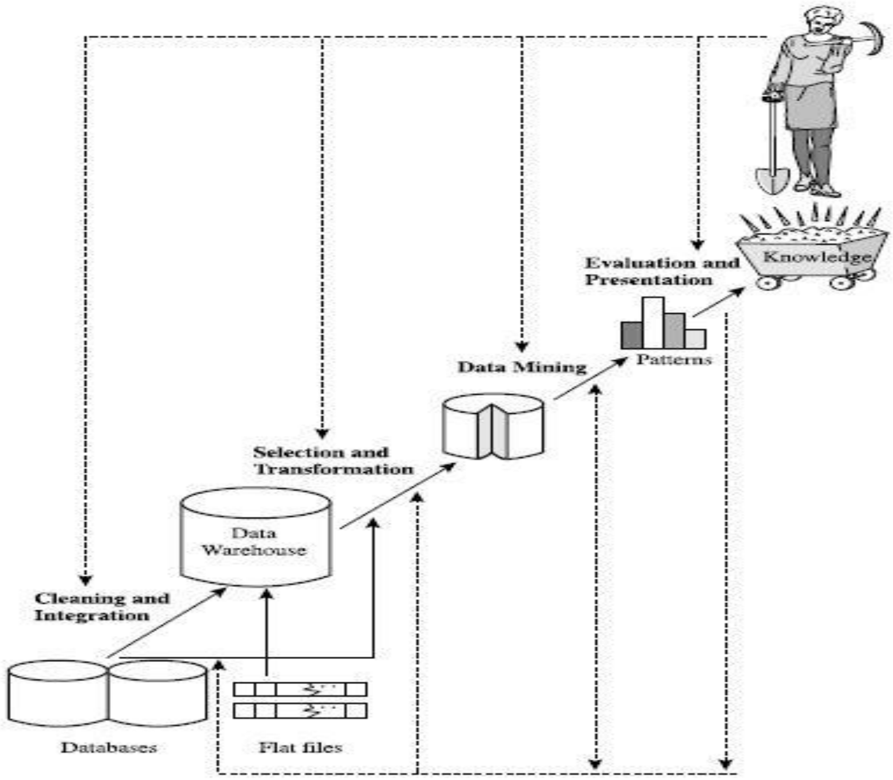


Figure 1.1: Knowledge Discovery Process

.

1.7 Research Structure

This research is structured of five chapters including the introduction in chapter one. Chapter two provides and focused on the literature review. Chapter three contains the methodology. Chapter four discusses the results and discussion. And finally, Chapter five presents the conclusions and recommendations for the future work.

CHAPTER TWO
BACKGROUND AND LITERATURE REVIEW

2.1 Introduction

The purpose of this chapter is to provide a detailed review of data mining domains and the relevant literature to the problem statement.

2.2 An Overview of Data Mining

Data mining refers to extracting or mining knowledge from large amounts of data. Thus, data mining should have been more appropriately named knowledge mining from data. Many other terms carry a similar or slightly different meaning to data mining, knowledge extraction, data/pattern analysis, data archaeology, and data dredging (Jiawei Ha, 2012).

2.3 Knowledge Discovery Steps

Many people believe in data mining as a synonym for another commonly used term the discovery of knowledge from data (Jiawei Ha, 2012). Others see data mining as a fundamental step in the knowledge discovery process which consists of an iterative sequence of the following steps:

- Data cleaning: to remove noise and inconsistent data
- Data integration: where multiple data sources may be combined
- Data selection: where data relevant to the analysis task are retrieved from the database
- Data transformation: where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations
- Data mining: an essential process where intelligent methods are applied in order to extract data patterns.
- Pattern evaluation: to identify the truly interest patterns representing knowledge based on some interestingness measures.

- Knowledge presentation: where visualization and knowledge representation techniques are used to present the mined knowledge to the user.

Data mining is the task of discovering interesting patterns from large amounts of data, where the data can be stored in databases, data warehouses, or other information repositories. It is an interdisciplinary field, drawing from areas such as database systems, data warehousing, statistics, machine learning, pattern recognition, and many application fields, such as business, economics, and bioinformatics (Jiawei Ha, 2012).

2.4 Data Mining Tasks

Data mining tasks are the kinds of data patterns that can be mined. In general it classified into two categories: the first category is prediction methods, which use some variables to predict unknown or future values of the same or other variables like (classification, regression), the second is descriptive methods which find human interpretable patterns that describe data like clustering and association rule (Rawat, 2017).

2.4.1 Regression

Predict the value of a given continuous valued variable based on the values of other variables assuming a linear or non-linear model of dependency. Regression extensively studied in the fields of statistics and neural networks (Rawat, 2017).

2.4.2 Classification

It is a data mining technique that assigns data in a collection to target classes, Classification is like bunching in a way that it likes wise fragments data records into various portions called classes, There are many common algorithms to do classification such as Decision trees, Bayesian classifiers and Support vector machine (Rawat, 2017).

2.4.3 Clustering

Determine object groupings such that objects within the same cluster are similar to each other, while objects in different groups are not, Typically objects are represented by data points in a multi dimensional space with each dimension corresponding to one or more attributes. Some clustering algorithms are k- means, cure, birch and wave cluster (Rawat, 2017).

2.4.4 Anomaly or Outlier Detection

Anomaly detection is also known as outlier detection. It is used to find date items, which do not match with items in the data set, There are various issues exist in mining information in substantial dataset such as information repetition, the estimations of traits is not particular, information is not finished and anomaly or Outlier. Anomaly detection can be used in a variety of areas like- fraud detection, system health monitoring, and intrusion detection (Rawat, 2017).

2.4.5 Association Rule

It used to extract frequent patterns, associations among sets of items in the transaction databases and data repositories, Association rules are applied in various areas such as communication networks, risk management, inventory control. Some recognize algorithm of association rules are Apriori, Eclat and FP growth (Rawat, 2017).

2.4.5.1 Apriori Algorithm

An algorithm uses a level-wise search, where k-itemsets (An item set which contains k items is known as k-itemset) are used to explore (k+1)-itemsets, to mine frequent itemsets from transactional database for Boolean association rules. In this algorithm, frequent subsets are extended one item at a time and this step is known as candidate generation process (Trupti A. Kumbhare, 2014).

2.4.5.2 Frequent Pattern Growth Algorithm (FP- Growth)

Frequent Pattern Growth requires constructing FP-tree. For that, it requires two passes. FP-growth uses divide and conquer strategy, FP growth requires two scans on the database, in first computes a list of frequent items sorted by frequency in descending order (F-List) and during its first database scan. In the second scan, the database is compressed into a FP-tree, the frequent item sets are generated with only two passes over the database and without any candidate generation process (Trupti A. Kumbhare, 2014).

There are two sub processes of frequent patterns generation process which includes: construction of the FP-tree, and generation of the frequent patterns from the FP-tree (Trupti A. Kumbhare, 2014).

FP-tree is constructed over the data-set using two passes are as follows:

Pass 1:

1) Scan the data and find support for each item.
2) Discard infrequent items.
3) Sort frequent items in descending order which is based on their support.

By using this order we can build FP-tree, so that common prefixes can be shared.

Pass 2:

1) Here nodes correspond to items and it has a counter.

2) FP-growth reads one transaction at a time and then maps it to a path.

3) Fixed order is used, so that paths can overlap when transactions share the items.

In this case, counters are incremented. Some pointers are maintained between nodes which contain the same item, by creating singly linked lists. FP-tree may fit in memory. Finally, frequent item sets are extracted from the FP-Tree. There are a unit 2 sub methods of frequent patterns generation process that includes: construction of the FP-tree, and generation of the frequent patterns from the FP-tree (Trupti A. Kumbhare, 2014).

The next is FP growth algorithm example represented in pseudocode below and figure 2.1.

The procedure fp-growth (tree T,A)

If tree T contains a single path P,

Then for each combination of the nodes in the

Path P do generate pattern B U A with

Support= minimum support of nodes in B

Else for each hi in the header of the

Tree T do

{Generate pattern B=hi U A with

Support= hi.support;

Construct B's conditional pattern base

And B's conditional FP-tree that is B;

If tree B ≠Ø

Then call FP-growth (Tree B,B)}

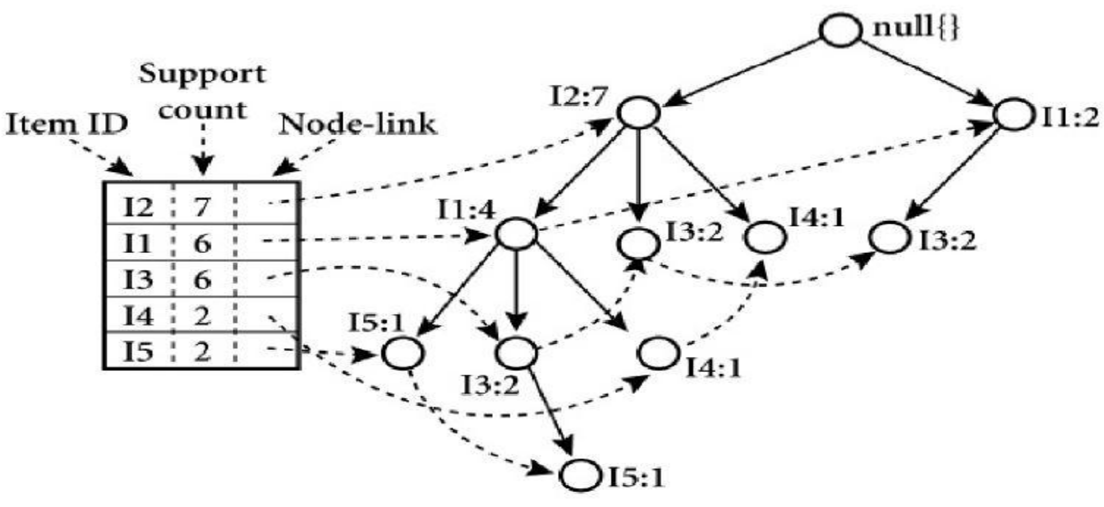


Figure 2.1: FP growth Algorithm example.

2.5 The Previous Studies

This section reviews the work done by past researchers in the field as presented below:

(Venkatachari, 2016) Explained how different products in a grocery store assortment interrelate and how to exploit these relations by marketing activities. Mining association rules from transactional data provided valuable information about co-occurrences and co-purchases of products. (Venkatachari, 2016) Collected data from Mumbai Retail Store and the sample size for the analysis is 300 transactions. Two algorithms are used: FP growth and Apriori algorithm using R tool and Rapid miner for finding association rules use. As per the study FP growth is much slower in Rapid Miner and in R Programming Apriori algorithm is faster (Venkatachari, 2016).

According to (Ayşe NurSağın, 2018) Market basket analysis was conducted to know purchasing trends from records in company databases on a five-and-a-half year data 81,384 transaction of a large hardware company operating in the retail sector. They used Apriori and FP-Growth algorithms to figure out the results 24 rule by Apriori and 721 rules by FP-Growth.

(Fachrul Kurniawan, 2018) Developed and implemented market basket analysis over association rule for definition of customer behavior in private supermarket of UIN Malang Business Center contains 891319 Transactions, they used Apriori algorithm to achieve that, with the means of confidence value of 46.69% and support value of 1.78%, and the amount of the generated rule was 30 rules. (Ulas, 2001) Described Most of the established companies had accumulated masses of data from their customers for decades. With the e-commerce applications growing rapidly, the companies will have a significant amount of data in months not in years. The aim of this thesis is to find correlations between the sales of items. The data are taken from GimaT¨urk A.S¸ a large Turkish supermarket chain, the dataset contains 756,868 transactions, the technique used to come up with the results is Apriori Algorithm and he found 25 association rules.

(Alghamdi, 2011) Presented how data mining can be applied on medical data and provided more guidance to the doctors as well as more understanding about the relationship between a doctor and a patient, the dataset contains 100 Transaction. Frequent Pattern (FP)-Growth algorithm used for building the results. He described the conceptual model for extraction rules on medical databases final result (12 rules) can guide the relationship between the different attributes presented in the data, For example, parent heart attack/angina before 50 age is depending on where the father born, if parent has asthma before at any age, and if parent has diabetes at any of the age.

(Ankur Mehay, 2013) Found the association rules among the large dataset (75000 transactions) using two algorithms: FP growth and Apriori algorithms. First found frequent itemsets using WEKA tool and Rapid-miner tool then generated the results (352 rules), This research showed FP-growth is much faster than Apriori algorithm to generate association rules when dataset is large.

Market basket analysis of sports equipment done by collecting primary data directly from retail and wholesale sports basket, applied FP Growth algorithm on the data collected directly from the sports vendors to optimize marketing and sale of sports equipment (28 rules) and most purchased sports items are tumbler, towel, running shoes, socks and back pack . Knowing different dynamics of sports equipment basket, sports vender can optimize the display of sports equipment on their premises to maximize sales (Harpreet Kaur, 2013).

(Ivan F. Videla-Cavieres, 2014) Developed a novel approach for market basket analysis based on graph mining techniques known k- means Algorithm to find sets of products sold together, demonstrate the effectiveness of approach in a wholesale supermarket chain and a retail supermarket chain, processing around 238,000,000 and 128,000,000 transactions and gain 7 clusters.

(Sally Jo Cunningham, 1999) Used 52,518 library transactions of a university by Apriori Algorithm and detected relevant books 21 rules, This information useful in directing users to additional portions of the collection that may contain documents relevant to their information need, and in determining a library’s physical layout.

Market basket analysis help to identify which items are purchased together. He collects sales data from a food joint of 400 transactions and used Apriori algorithm for determining top 4 rules with higher confidence between sold items on the menu (Priti Kulkarni, 2017).

Frequent itemset mining is an important component of analytical system in retail organizations to determine the placement of goods, designing sales promotions for different segments of customers to improve customer satisfaction and hence the profit of the supermarket. The frequent itemsets are mined from the market basket database contains 9620 transactions using the efficient K-Apriori algorithm and then 27 rules are generated (Loraine Charlet Annie, 2012).

(Sergey Brin, 1997) Presented a new algorithm for finding large itemsets which uses fewer passes over the data than classic algorithms, and showed how they produce more intuitive results than other methods, finally explained how different characteristics of real data (IBM test data 100000 Transactions), as opposed to synthetic data, can dramatically affect the performance of the system and the form of the results (23712 rules) by Apriori Algorithm.

Retailing is an industry with high level of competition, It is a customer based industry which depends on how it could be aware of what the customers' needs and requirement, The process of identifying the related products bought together from supermarket transactions (1049), the results found five category association rules was done by Apriori algorithm (Seruni, 2005).

Stored data has information that can be extracted by data mining techniques. O! Fish restaurants used Information about sales patterns to create more potential promotional strategies to boost sales by referring to items that are often purchased together, 150 transaction data used by Apriori algorithm and generating 7 rules (Yusuf Kurnia, 2019).

Product bundling is one of the most important marketing strategies used to get rid of stock by making integrated bundles of inactive products and demanded products with discount prices. She used pharmacy dataset contains 9998 transactions by two algorithms: Apriori and FP-growth are gave strong association rules are 11 by Apriori and 49 by FP-Growth, But FP-Growth algorithm it was faster in dealing data (Amira H. SHalaby, 2015).

Table 2.1: literature summary

Author and Date	**Area**	**Objectives**	**Datasets**	**Algorithm and Setting**	**Number of Results**
(Venkatachari, 2016)	Retail market	Consumer shopping baskets	300 Transactions	FP growth and Apriori Support 0.05 Confidence 0.83	2090141 by Apriori 987 by FP growth
(Fachrul Kurniawan, 2018)	Retail market	Definition of customer behavior	891319 Transactions	Apriori Support 0.01 Confidence 0.46	30

(Ulas, 2001)	Retail market	Correlations in sales of items	756,868 transactions	Apriori Support 0.001 Confidence 0.52	25
(Ayşe NurSağın, 2018)	Retail market	Extract purchasing trends	81,384 Transactions	FP growth and Apriori Support 0.5 Confidence 0.40	24 by Apriori 721 by FP growth
(Alghamdi, 2011)	Medical data	Provide more guidance	100 Transactions	FP growth Support 0.60 Confidence 0.95	12
(Ankur Mehay, 2013)	Market basket	found frequent itemsets	75000 Transactions	FP growth and Apriori Support 0.009 Confidence 0.2	352 by FP growth 19 by Apriori
(Harpreet Kaur, 2013)	Sports store	Optimize marketing and sale of sports equipment	No Details	FP growth Support 0.23 Confidence 1	28
(Ivan F. Videla-Cavieres, 2014)	Retail and Wholesale	Find sets of products	128000000 238000000	K- Means	7 Clusters
(Sergey Brin, 1997)	Retail market	finding large itemsets	100000 Transactions	Apriori Support 0.05 Confidence 0.09	23712
(Sally Jo Cunningham, 1999)	Library Circulation Data	Detect relevant books	52,518 Documents	Apriori Support 0.01 Confidence	21

				0.01	
(Priti Kulkarni, 2017)	Retail market	Improve sales of food items	400 Transactions	Apriori Support 0.92 Confidence 0.96	4
(Loraine Charlet Annie, 2012)	Retail market	Improve customer satisfaction	9620 Transactions	K- Apriori Support 0.62 Confidence 1	27 Rules 2 Cluster
(Amira H. SHalaby, 2015)	Pharmacy Data	get rid of stock	9998 Transactions	Apriori and FP Growth Support 0.3 Confidence 0.9	11 by Apriori 49 by FP growth
(Seruni, 2005)	Retail market	Products bought together	1049 Transactions	Apriori Support 0.1 Confidence 0.9	5
(Yusuf Kurnia, 2019)	Rest runt Food Market	Boost sales	150	Apriori support 0.04 confidence 0.60	7

2.4 Summary

The chapter contained a review of data mining domains and the relevant literature to the problem statement.

The research found association rules are widely used in market basket analysis, Apriori algorithm is the most commonly used Algorithm and FP growth Algorithm is suitable with a large amount of data. So he decided using FP growth Algorithm.

CHAPTER THREE
METHODOLOGY AND IMPLEMENTATION

3.1 Introduction

This chapter includes the data source, the performed data pre- processing, and the implementation of FP growth algorithm. The research methodology steps are explained in the Figure 3.1.

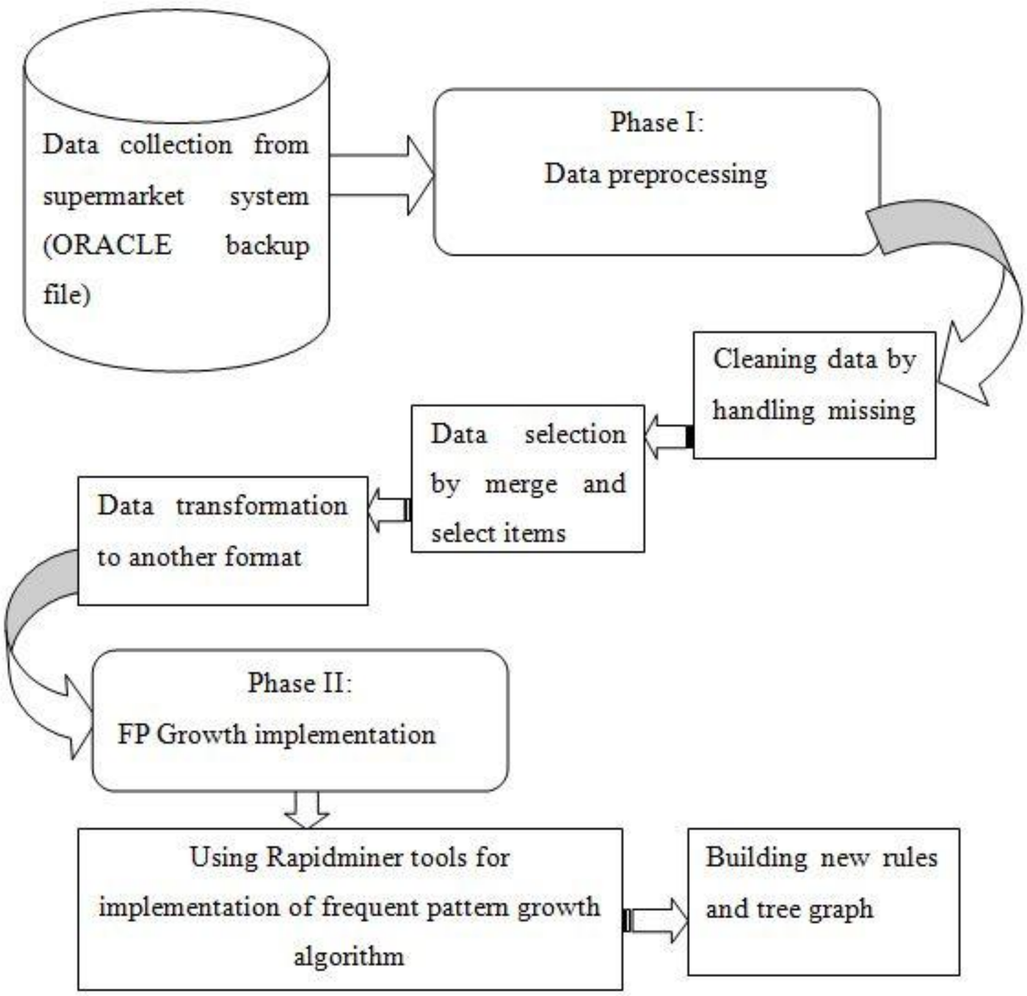


Figure 3.1: Research Methodology Steps

3.2 Data Source Description

The data source is Aldooma supermarket's oracle database backup files which are collected from sale points system. The data contains records in the period (JAN/2014 to MAY/2015). Twenty one tables were there, but only two tables out of the total were selected and used for the research purpose, namely: daily_orders, and item_mast.

'Daily_orders' table contains the data about selling transactions and the total number of records inside this table was 1,982,390, the attributes are shown in Table 3.1, Figure 3.2 shows a screenshot of daily_orders table with sample of transactions.

Table 3.1: Attributes of Daily_orders

Attribute Name	**Explanation**	**Data Type**
Cashier_code	Cashier Name	String
order_date	Sales Date	Date
Slip_no	Invoice Number	Numeric
Item_code	Code of Items	Numeric
Sell_price	Item Price	Numeric
Order_quantity	Product Quantity	Numeric
Cost_price	Cost Price	Numeric
Ext_ammount	Stock Quantity	Numeric
Order_status	Order Status	Numeric
Line_no	Point of Sales Number	Numeric
Order_shift	Order Shift	String

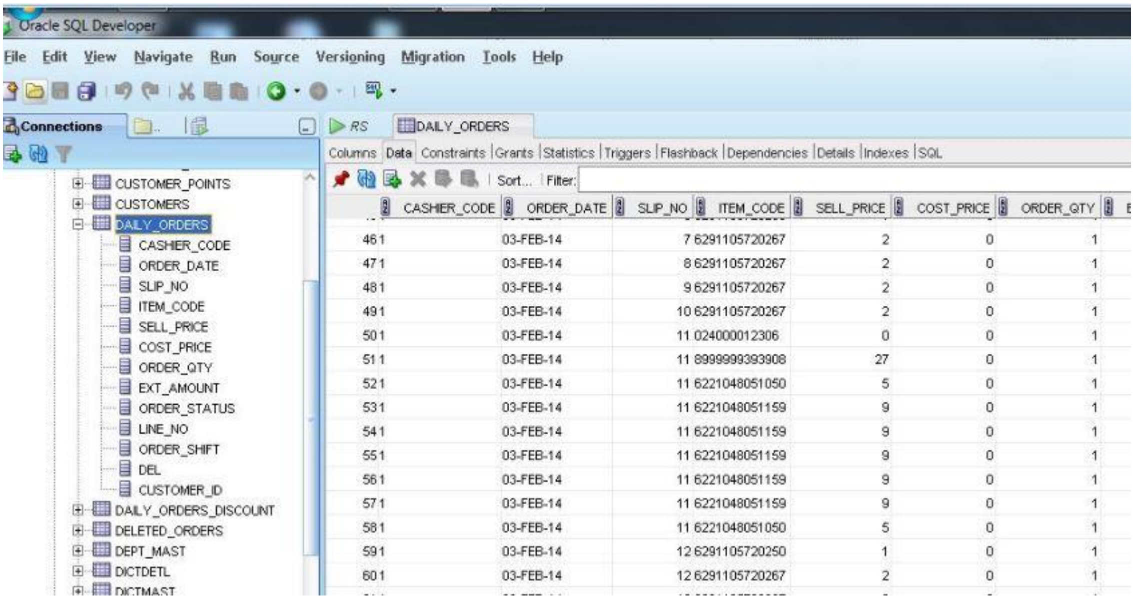


Figure 3.2: Selling transactions (Daily_orders)

Regarding the table 'Item_mast', it contains the data of Selling Items, the total number of records was 13,077and the attributes are shown in Table 3.2,

Figure 3.3 shows a screenshot of Item_mast table with sample of data.

Table 3.2: Attributes of Item_mast

Attribute Name	**Explanation**	**Data Type**
Family_code	Family Categories	String
Mark_code	Mark Name	String
Item_code	Code of Items	Numeric
Item_desc	Items Arabic Name	String
Sell_price	Item Price	Numeric
Avg_cost_price	Average Cost Price	Numeric
Supplier_id	Supplier Code	Numeric
Min_stock	Stock Quantity	Numeric
Dept_id	Department Code	Numeric
Min_expiry	Expiry Date	Date

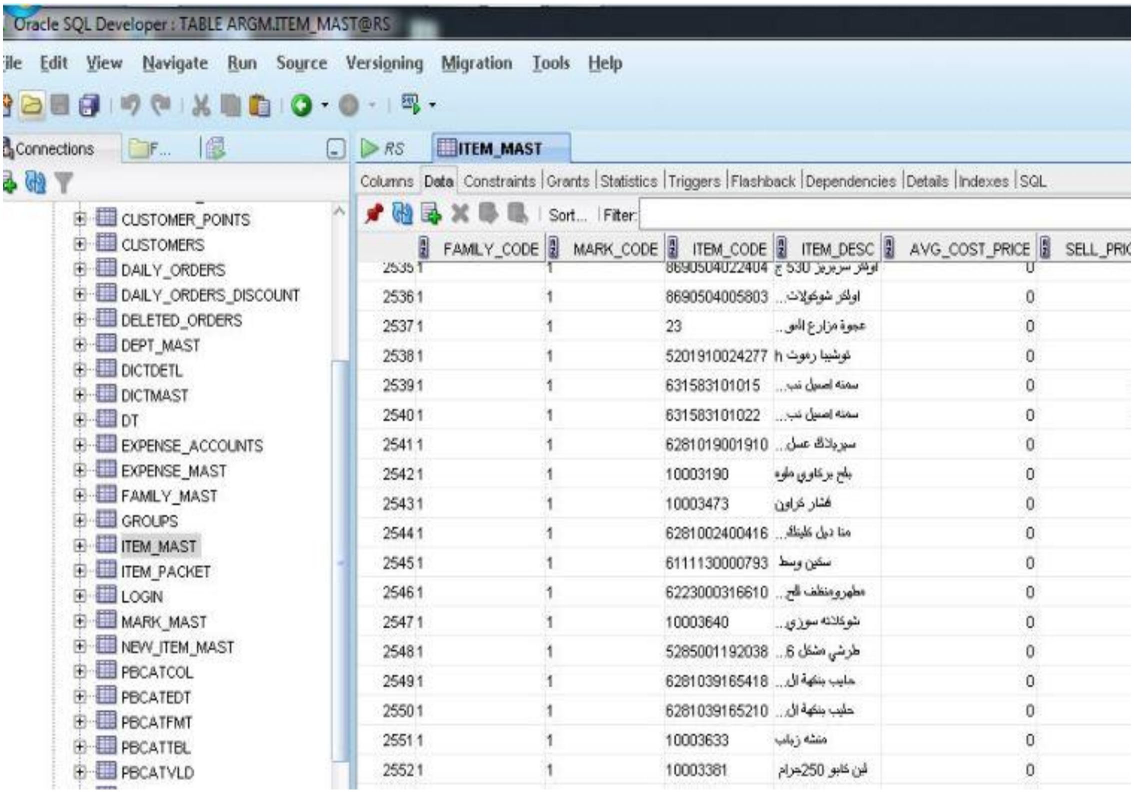


Figure 3.2: Selling transactions (Daily_orders)

Two major phases was performed in research methodology to get the required results:

Phase I: Data Pre-processing, using Microsoft Office Excel and ORACLE DBMS (Database management system) 11G by SQL developer interface.

Phase II: FP Growth Algorithm Implementation, using Rapidminer tools.

3.3 Phase I: Data Pre-processing

This phase talks about some of steps used to transform the raw data in a useful and efficient format to be ready for implementation, those processes are: Data cleaning, data selection, data transformation.

3.3.1 Data Cleaning Process

To handle the missing value (null and zero) items name from item_mast table in the database by using SQL delete query. Figure 3.4 shows a screenshot of the Cleaning query using ORACLE DBMS11g (SQL developer).

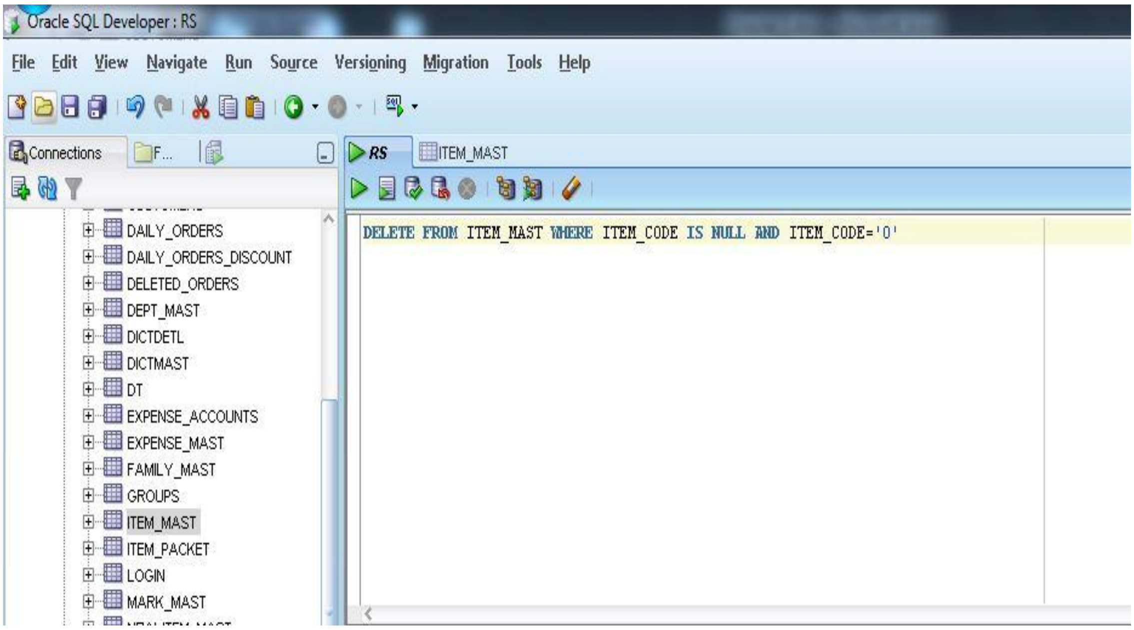


Figure 3.4: Cleaning the data

3.3.2 Data Selection Process

Because the database is too big, the researcher selected 1,463 attribute (items) from 13,077 of total items and merged the records of items name by grouping them into 24 groups. For instance: Pepsi, Miranda and cola are grouped as 'gas water' .Sugar 10kg, white sugar are grouped as 'sugar' and so on. The rest of groups are shown in the following steps below. The researcher used SQL ALTER, SELECT and CREATE queries for doing so as shown in Appendix D.

3.3.3 Data Transformation Process

In this process the data format is converted to another format (from column to rows) for creating the dataset for being ready for the implementation phase. This is done by one-hot encoding method it's commonly used for data transformation by making **1** if item bought and **0** if not bought, using SQL UPDATE QUERY in TRANS table. ORACLE SQL developer and Microsoft Excel are used in transformation for the 24 items, which are explained in Appendix E.

After performing Transformation Process the researcher transformed data from database into an excel file to be ready for implementation. In the Table 3.3 below shows the numbers of processed items in data pre-processing tasks.

Table 3.3: data pre-processing summary

Items	Numbers
Deleted	30
Merged to GasWater	514
Merged to BISCUIT	47
Merged to CAKE	12
Merged to SUGAR	9
Merged to CHEESE	6
Merged to LENTILS	8
Merged to RICE	5
Merged to JAM	10
Merged to HALVA	14
Merged to MILK	220
Merged to SOAP	56
Merged to TEA	25
Merged to ANDOMI	13
Merged to PASTA	7
Merged to SHARIEA	13
Merged to YOGURT	21
Merged to MEAT	17
Merged to FLOUR	11
Merged to JUICE	378
Merged to OIL	32
Merged to EGGS	9
Merged to **TomatoPaste**	10
Merged to CHIPS	20
Merged to BREAD	6

3.4 Phase II: FP Growth Algorithm Implementation

The tool used for the implementation of FP growth in this research is Rapid Miner (RM), Rapid Miner is an open source package provides an integrated development environment to build and apply the data mining tasks, It provides a long list of processes, operators and data sets for helping students, decision makers and researchers. RM simplifies the various scattered tasks of data mining and analysis. There is a place to load data, pre-process and prepare data, modeling data, and even visualizing outputs (Manikandan, 2018). Figure 3.5 shows a screenshot of work interface in Rapid Miner.

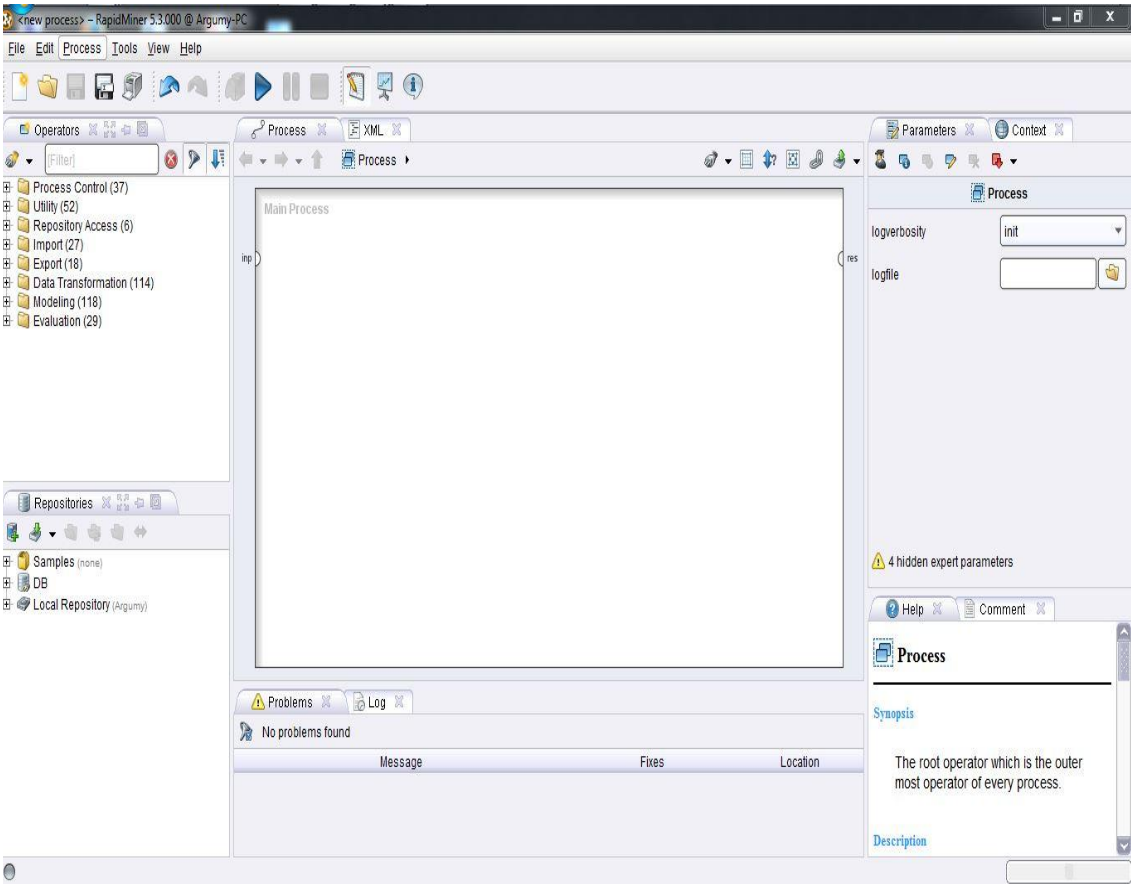


Figure 3.5: Rapid Miner Main Interface

3.4.1 FP Growth Implementation Process:

Now, the data has been selected, checked, and modified using data pre processing techniques which performed in phase I. All steps in data pre processing are depending on the applying of association mining algorithm FP growth. The explanation of the entire operators used in rapid miner and steps for the complete process of generate association rules are presented below and shown in figure 3.6:

Operator No.1: is the Reading the CSV (comma delimited) file of data from the hard disk. It use for import data from .CSV file.

Operator No.2: is a Type Converter of numerical data into binominal data. Using this operator is converting all the data values into binominal values before connecting it with the FP growth.

Operator No.3: Using FP growth algorithm, receiving the data file of sales transaction. Given support value 0.004 for extracting frequent item set. Generated frequent item set will be used by the next operator for create association rules.

Operator No.4: Finally, the rules generator operator. After extraction of frequent item set by using input data, the frequent item set will be forwarded for the generation of association rules from the given minimum confidence value 0.8.

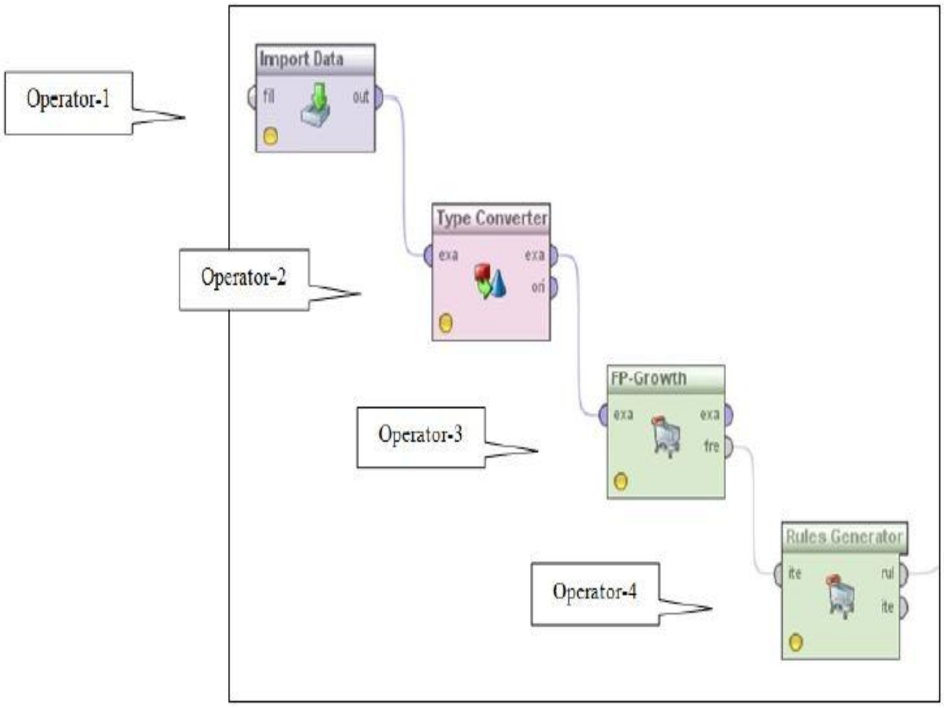


Figure 3.6: FP Growth Implementation Process.

3.5 Summary

This chapter explains the methodology steps. It started, first with description of data source, the data pre-processing tasks (cleaning, selection and transformation) and finally the implementation of FP growth algorithm for building tree graph and generating association rules. The results are presented and discussed in the next chapter.

CHAPTER FOUR
THE RESULTS AND DISCUSSION

1.4 Introduction

This chapter presents the results that the researcher reached after following the steps in the previous chapter and discussion of these results.

4.2 Results Analysis and Discussion

The results obtained by implementing FP growth algorithm Using Rapid miner tools on 42010 Transactions of Aldooma supermarket dataset to determine the frequent item sets sold and Building a model (FP-tree) for Figuring out the commodities that were sold together and how to exploit these relations by marketing activities such as promotional support, inventory control and cross-sale campaigns.

The researcher conducted many experiments by changing the values of confidence and support and reached to a set of results as shown in table 4.1.

Table 4.1: Different group of results

NO	Support	Confidence	Number of Rules
1	0.003	0.7	722
2	0.001	0.8	40909
3	0.005	0.9	8
4	0.002	0.6	45510
5	0.004	0.8	66

The 66 rules of frequent itemsets sold together and their frequent pattern tree using min support 0.004 and min confidence 0.8 are explained below in table 4.2 and figure 4.1.

Table 4.2: The results of discovered rules

Transaction item sets	frequent itemsets sold	Support	Confidence
YOGURT, MEAT	OIL	0.005	0.8
MEAT, EGGS	OIL	0.005	0.8
YOGURT, MEAT	OIL, EGGS	0.005	0.8
MEAT, EGGS	YOGURT, OIL	0.005	0.8
YOGURT, MEAT, EGGS	OIL	0.005	0.8
YOGURT, OIL	EGGS	0.008	0.8
PASTA, SHAYRIEA	ANDOMI	0.010	0.8
PASTA, FLOUR	MEAT	0.004	0.8
TomatoPaste, EGGS	MEAT	0.004	0.8
TomatoPaste, EGGS	YOGURT, MEAT	0.004	0.8
YOGURT, TomatoPaste, EGGS	MEAT	0.004	0.8
TomatoPaste, EGGS	FLOUR, MEAT	0.004	0.8
TomatoPaste, EGGS	YOGURT, FLOUR, MEAT	0.004	0.8
YOGURT, TomatoPaste, EGGS	FLOUR, MEAT	0.004	0.8
OIL, TomatoPaste	PASTA	0.004	0.8
JUICE, EGGS	OIL	0.004	0.8
OIL, BREAD	YOGURT	0.005	0.8
OIL, BREAD	EGGS	0.004	0.8
OIL, BREAD	YOGURT, EGGS	0.004	0.8
TomatoPaste, EGGS	FLOUR	0.004	0.9
TomatoPaste, EGGS	YOGURT, FLOUR	0.004	0.9
YOGURT, TomatoPaste, EGGS	FLOUR	0.004	0.9
YOGURT, FLOUR, TomatoPaste	MEAT	0.004	0.9
YOGURT, FLOUR, EGGS	MEAT	0.004	0.9
FLOUR, TomatoPaste,	MEAT	0.004	0.9

EGGS			
YOGURT, FLOUR, TomatoPaste	MEAT, EGGS	0.004	0.9
YOGURT, FLOUR, EGGS	MEAT, TomatoPaste	0.004	0.9
FLOUR, TomatoPaste, EGGS	YOGURT, MEAT	0.004	0.9
YOGURT, FLOUR, TomatoPaste, EGGS	MEAT	0.004	0.9
FLOUR, TomatoPaste	MEAT	0.004	0.9
LENTILS, FLOUR	JAM	0.011	1.0
ANDOMI, JUICE	EGGS	0.005	1.0
ANDOMI, EGGS	JUICE	0.004	1.0
PASTA, OIL	TomatoPaste	0.004	1.0
YOGURT, MEAT	EGGS	0.004	1.0
MEAT, EGGS	YOGURT	0.006	1.0
TomatoPaste, EGGS	YOGURT	0.006	1.0
EGGS, BREAD	YOGURT	0.004	1.0
MEAT, TomatoPaste	FLOUR	0.004	1.0
OIL, JUICE	EGGS	0.011	1.0
EGGS, BREAD	OIL	0.005	1.0
YOGURT, FLOUR, MEAT	TomatoPaste	0.004	1.0
YOGURT, MEAT, TomatoPaste	FLOUR	0.004	1.0
YOGURT, FLOUR, MEAT	EGGS	0.004	1.0
FLOUR, MEAT, EGGS	YOGURT	0.004	1.0
YOGURT, FLOUR, TomatoPaste	EGGS	0.004	1.0
YOGURT, FLOUR, EGGS	TomatoPaste	0.004	1.0
FLOUR, TomatoPaste,	YOGURT	0.004	1.0

EGGS			
YOGURT, MEAT, OIL	EGGS	0.004	1.0
MEAT, OIL, EGGS	YOGURT	0.005	1.0
YOGURT, MEAT, TomatoPaste	EGGS	0.005	1.0
MEAT, TomatoPaste, EGGS	YOGURT	0.004	1.0
YOGURT, OIL, BREAD	EGGS	0.004	1.0
EGGS, BREAD	YOGURT, OIL	0.004	1.0
YOGURT, EGGS, BREAD	OIL	0.004	1.0
OIL, EGGS, BREAD	YOGURT	0.004	1.0
FLOUR, MEAT, EGGS	TomatoPaste	0.004	1.0
MEAT, TomatoPaste, EGGS	FLOUR	0.004	1.0
YOGURT, FLOUR, MEAT	TomatoPaste, EGGS	0.004	1.0
YOGURT, MEAT, TomatoPaste	FLOUR, EGGS	0.004	1.0
YOGURT, FLOUR, MEAT, TomatoPaste	EGGS	0.004	1.0
FLOUR, MEAT, EGGS	YOGURT, TomatoPaste	0.004	1.0
YOGURT, FLOUR, MEAT, EGGS	TomatoPaste	0.004	1.0
MEAT, TomatoPaste, EGGS	YOGURT, FLOUR	0.004	1.0
YOGURT, MEAT, TomatoPaste, EGGS	FLOUR	0.004	1.0
FLOUR, MEAT, TomatoPaste, EGGS	YOGURT	0.004	1.0

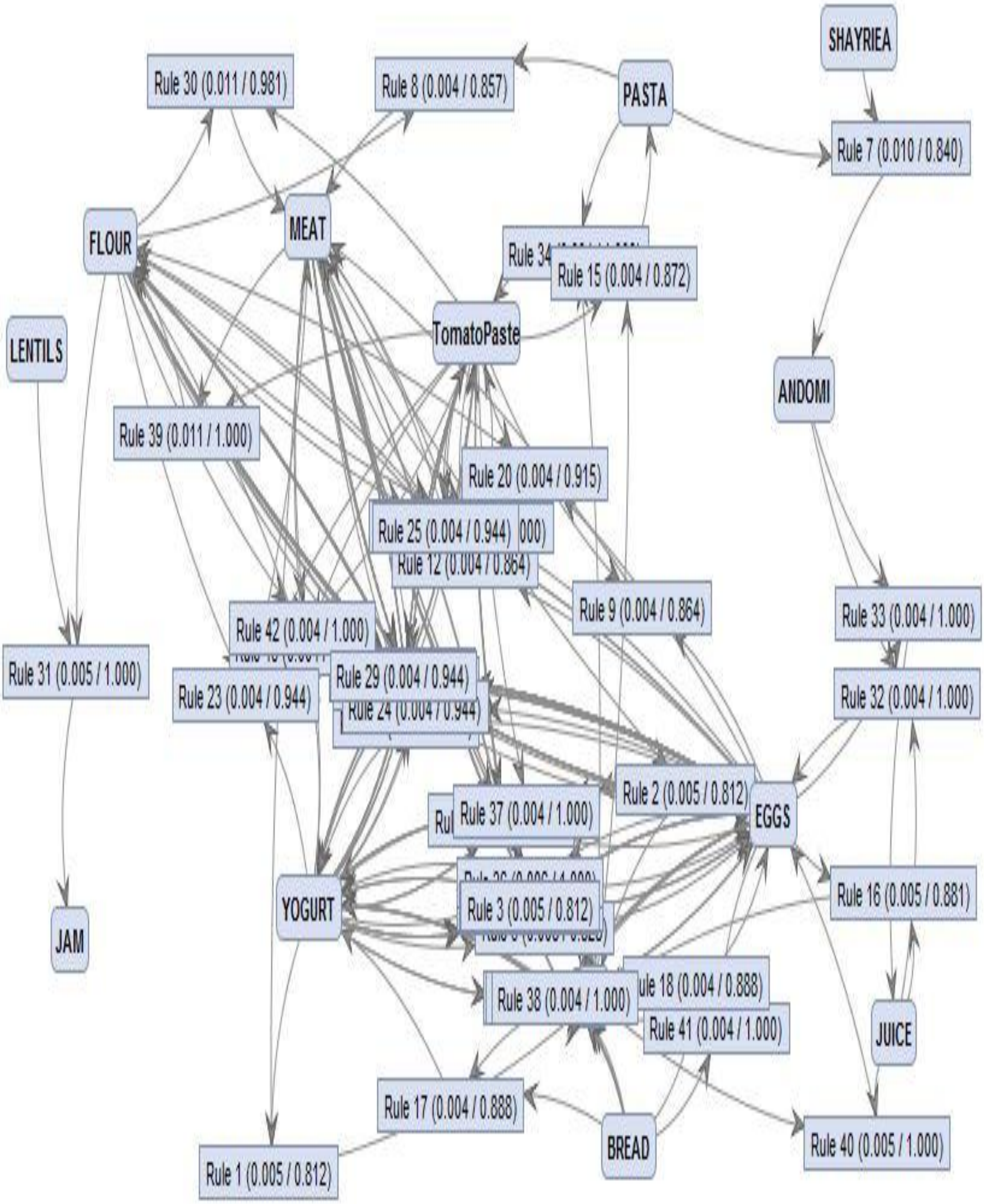


Figure 4.1: Frequent pattern tree

The researcher selected the best results for the discussion which contained the longest frequent itemsets sold together, with confidence values (0.8-0.9-1.0) and support values (0.004-0.005) which are represented in 12 Association rules as shown in Table 4.2.

The researcher conducted another experiments by Apriori Algorithm in Rapidminer extension with the same value of min confidence (0.8) and min support (0.004) that used in FP growth. There are no rules found as shown in Appendix (C), Table 4.3 presented comparison of the results between Apriori and FP growth.

Table 4.2: Best Discovered Rules.

Transaction item sets	**frequent itemsets sold**	**Support**	**Confidence**
YOGURT, MEAT	OIL, EGGS	0.005	0.8
MEAT, EGGS	YOGURT, OIL	0.005	0.8
TomatoPaste, EGGS	YOGURT, MEAT	0.004	0.8
TomatoPaste, EGGS	FLOUR, MEAT	0.004	0.8
TomatoPaste, EGGS	YOGURT, FLOUR, MEAT	0.004	0.8
YOGURT, TomatoPaste, EGGS	FLOUR, MEAT	0.004	0.8
OIL, BREAD	YOGURT, EGGS	0.004	0.8
TomatoPaste, EGGS	YOGURT, FLOUR	0.004	0.9
YOGURT, FLOUR, TomatoPaste	MEAT, EGGS	0.004	0.9
YOGURT, FLOUR, EGGS	MEAT, TomatoPaste	0.004	0.9
FLOUR, TomatoPaste, EGGS	YOGURT, MEAT	0.004	0.9
EGGS, BREAD	YOGURT, OIL	0.004	1.0
YOGURT, FLOUR, MEAT	TomatoPaste, EGGS	0.004	1.0
YOGURT, MEAT, TomatoPaste	FLOUR, EGGS	0.004	1.0
FLOUR, MEAT, EGGS	YOGURT, TomatoPaste	0.004	1.0
MEAT, TomatoPaste, EGGS	YOGURT, FLOUR	0.004	1.0

Table 4.3: Result comparison between Apriori and FP growth algorithms

Algorithm	**Performance** (time consumed in implementation)	**Accuracy** (Number of rules)
FP GROWTH	4 seconds	66 rules
APRIORI	13 seconds	0 rules

The results explained that the most basic commodity is yogurt, which appeared in 19 frequent itemsets sold out of 66 transaction itemsets. The best results the longest frequent itemsets sold together represented in 12 Association rules obtained by the researcher are:

- - If the customer bought YOGURT and MEAT will buy OIL and EGGS.
- - If the customer bought MEAT and EGGS will buy YOGURT and OIL.
- - If the customer bought TomatoPaste and EGGS will buy YOGURT and MEAT.
- - If the customer bought TomatoPaste and EGGS will buy FLOUR and MEAT.
- - If the customer bought TomatoPaste and EGGS will buy YOGURT, FLOUR and MEAT.
- - If the customer bought YOGURT, TomatoPaste and EGGS will buy FLOUR and MEAT.
- - If the customer bought OIL and BREAD will buy YOGURT and EGGS.
- - If the customer bought TomatoPaste and EGGS will buy YOGURT and FLOUR.
- - If the customer bought YOGURT, FLOUR and TomatoPaste will buy MEAT and EGGS
- - If the customer bought YOGURT, FLOUR and EGGS will buy MEAT and TomatoPaste
- - If the customer bought FLOUR, TomatoPaste and EGGS will buy YOGURT and MEAT
- - If the customer bought EGGS and BREAD will buy YOGURT and OIL.
- - If the customer bought YOGURT, FLOUR and MEAT will buy TomatoPaste and EGGS
- - If the customer bought YOGURT, MEAT and TomatoPaste will buy FLOUR and EGGS
- - If the customer bought MEAT, TomatoPaste and EGGS will buy YOGURT and FLOUR

The Top 10 items sold are (YOGURT, EGGS, MEAT, FLOUR, OIL, TomatoPaste, PASTA, JAM, JUICE and ANDOMI).

CHAPTER FIVE
CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The main purpose of the research is analyzing Sudanese shopping behavior using Aldooma supermarket as a case study and FP growth algorithm in Rapid miner tool. The research is intended to determine the frequent itemsets sold together and how to exploit these relations by marketing activities such as promotional support, inventory control and cross-sale campaigns. The techniques which applied in this research are useful for the supermarket owner or the decision maker who can use them to grow their customer base and build stronger customer relationships to turn inventory into cash.

After reviewing and analyzing the results of the research, the results explained that the most basic commodity is yogurt, which appeared in 19 frequent itemsets sold out of 66 transaction itemsets. The best results are the longest frequent itemsets sold together represented in 12 Association rules. The important of storing purchases data into a database is emerging to be used in decision-making and market research such as promotional support, inventory control and cross-sale campaigns.

5.2 Recommendation

1. Building an automated method for preparing and pre-processing data in the appropriate template for this algorithm because it took too much time like design a software or using deep learning techniques to achieve that.
2. Apply other algorithms such as Eclat and DIC to this data and compare it with the obtained results.
3. Use results and apply other data mining techniques such as classification, regression and clustering algorithms for customer segmentation or prediction of the year or quarter profits.

REFRENCES

Alghamdi, A. S. (2011). *Efficient Implementation of FP Growth Algorithm-Data Mining on Medical Data.* Jeddah: IJCSNS International Journal of Computer Science and Network Security.

Amira H. SHalaby, D. M. (2015). Suggested Marketing Strategy Using Apriori and FP-Growth Algorithms in retail sales in Egypt. Egypt: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY.

Ankur Mehay, D. K. (2013). *Analyze Market Basket Data using FP-growth and Apriori.* Patiala: International Journal on Recent and Innovation Trends in Computing and Communication.

Ayşe NurSağın, B. (2018). Determination of Association Rules with Market Basket Analysis: Application in the Retail Sector. Istanbul: Southeast Europe Journal of Soft Computing.

Fachrul Kurniawan, B. U. (2018). *Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data.* Malang: Knowledge Engineering and Data Science (KEDS).

Harpreet Kaur, K. S. (2013). *Market Basket Analysis of Sports Store using Association.* Patiala: International Journal of Recent Trends in Electrical & Electronics Engg.

Ivan F. Videla-Cavieres, S. A. (2014). Extending market basket analysis with graph mining techniques: A real case. Santiag: Elsevier Ltd.

Jiawei Ha, M. K. (2012). *Data mining concepts and techniques third edition.* San Francisco: Diane Cerra.

Loraine Charlet Annie, A. K. (2012). *Market Basket Analysis for a Supermarket based on Frequent Itemset Mining.* Trichy: IJCSI International Journal of Computer Science Issues.

Manikandan, H. (2018, October 16). *Medium.* Retrieved November 25, 2018, from A Walk-through the Rapid Miner: https://medium.com/@vshshv3/a-walk-through-the-rapid-miner-921dfaf53722

Priti Kulkarni, A. R. (2017). Use of Association Mining to Analyse Customer Behaviour for Small Scale Food Joints -A Case Study Approch. Pune: Sinhgad Institute of Management & Computer Application (SIMCA).

Rawat, K. S. (2017). Comparative Analysis of Data Mining Techniques, Tools and Machine Learning Algorithms for Efficient Data Analytics. Dharamsala: Journal of Computer Engineering.

Sally Jo Cunningham, E. F. (1999). *Market basket analysis of library circulation data.* Hamilton: International Conference on Neural Information Processing.

Sergey Brin, R. M. (1997). Dynamic Itemset Counting and Implication Rules for Market Basket Data. Stanford : SIGMOD Conference.

Seruni, I. S. (2005). DESIGN OF PRODUCT PLACEMENT LAYOUT IN RETAIL SHOP USING MARKET BASKET ANALYSIS. Indonesia : MAKARA TEKNOLOGI.

Trupti A. Kumbhare, P. S. (2014). *An Overview of Association Rule Mining Algorithms.* Pune: International Journal of Computer Science and Information Technologies.

Ulas, M. A. (2001). *MARKET BASKET ANALYSIS FOR DATA MINING.* Gima T¨urk A.S: Computer Engineering, Bo˘gazi¸ci University.

Venkatachari, P. K. (2016). *MARKET BASKET ANALYSIS USING FP GROWTH AND APRIORI ALGORITHM: A CASE STUDY OF MUMBAI RETAIL STORE.* Mumbai: Prof. Kavitha Venkatachari,"MARKET BASKET ANALYSIS USING FP GROWTH AND APRIORI ALGORITHBVIMSR's Journal of Management Research.

Yusuf Kurnia, Y. I. (2019). Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm. Jakarta: International Conference on Advance and Scientific Innovation.

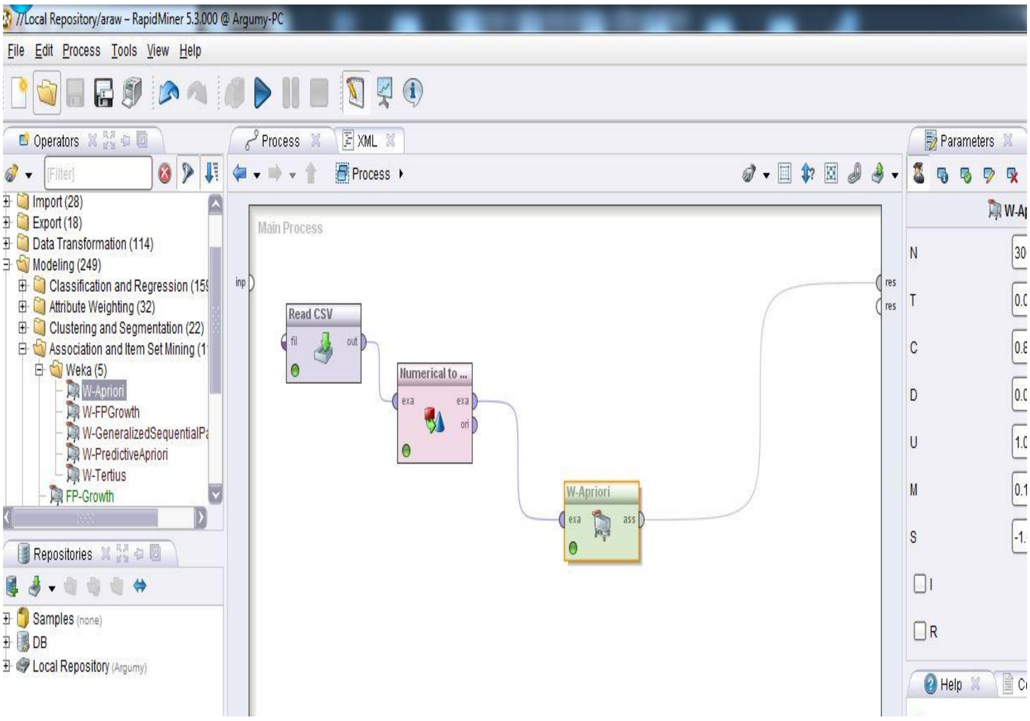
APPENDICES

Appendix A

Capture for the process of import data source into oracle 11g

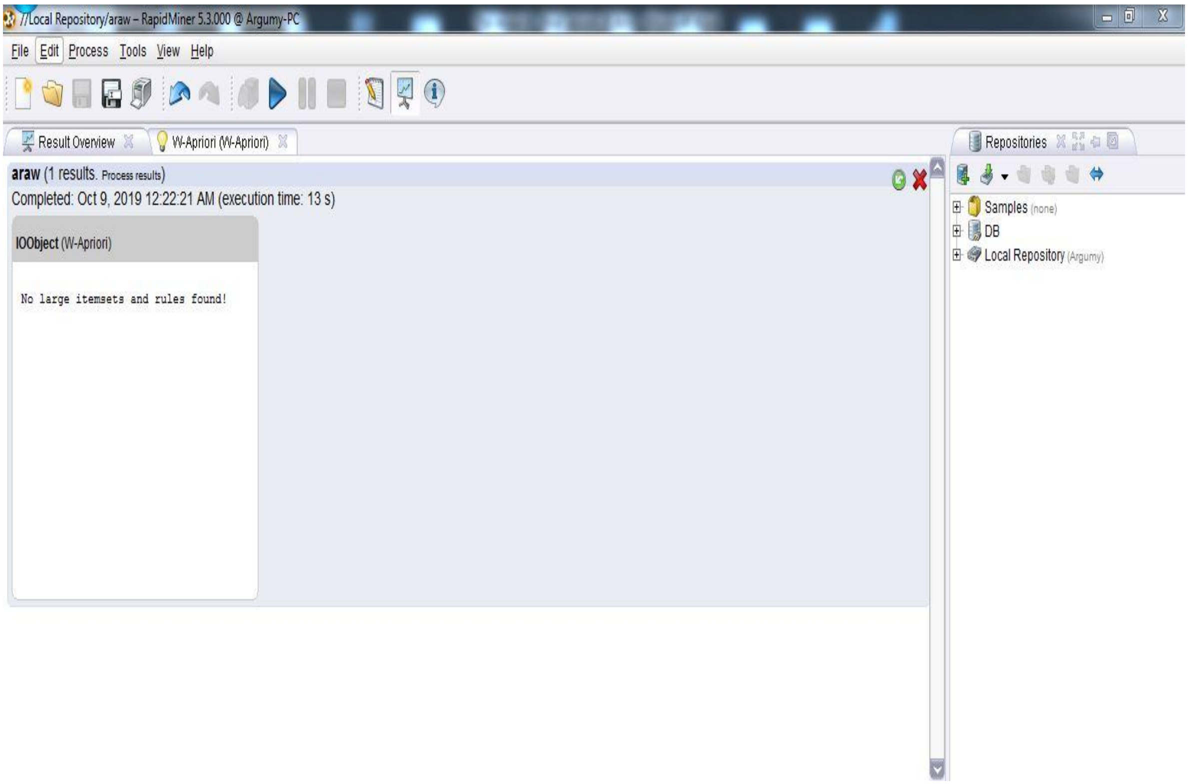
Appendix B

Aproiri Algorithm Impelemention Process Screenshot



Appendix C

Screenshot of Apriori Algorithm Results



Appendix D

STEP 1: Adding a new column (NEW_NAME) that represents 24 new items name.
ALTER TABLE **DAILY_ORDERS** ADD COLUMN **NEW_NAME** (NOT NULL, VARCHAR2 (20));

STEP 2: SELECT ITEM 1 **GasWater**
UPDATE DAILY_ORDERS SET NEW_ITEM = ‘GasWater’
WHERE ITEM_CODE IN
(‘6281034016142’,‘10001240’,‘5449000028945’,‘1120102001855’,‘012000803048’,‘01200000545’,‘012000002489',....)

STEP 3: SELECT ITEM 2 **BISCUIT**
UPDATE DAILY_ORDERS SET NEW_ITEM = ‘BISCUIT’
WHERE ITEM_CODE IN
(‘8690705055089’,‘776992031048’,‘1010’,‘4017100125515’,‘10004623’,‘84100887’,‘0162300176310044’,....)

STEP 4: SELECT ITEM 3 **CAKE**
UPDATE DAILY_ORDERS SET NEW_ITEM = ‘CAKE’
WHERE ITEM_CODE IN
(‘6281073111587’,‘4011100985346’,‘8690705002656’,‘0005’,‘03161005’, ‘041550016943’,'3017760832496 ',....)

STEP 5: SELECT ITEM 4 **SUGAR**
UPDATE DAILY_ORDERS SET NEW_ITEM = ‘SUGAR’
WHERE ITEM_CODE IN
(‘6281011411076’,‘10000069’,‘5201002001292’,‘8412224018259’,‘5010067376500’,‘6141001203008’,'4046746625974',....)

STEP 6: SELECT ITEM 5 **CHEESE**
UPDATE DAILY_ORDERS SET NEW_ITEM = ‘CHEESE’
WHERE ITEM_CODE IN
(‘5760466738095’,‘6981024012139’,‘4522001292’,‘3228021060019’,‘735420036515

9’,‘5760466738088’,‘373706624335’,....)

STEP 7: SELECT ITEM 6 **LENTILS**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘LENTILS’

WHERE ITEM_CODE IN

(‘10000076’,5000157062734,‘8697669207615’,‘2961658030086’,‘6221048821325’,‘0 18250092132667’,‘020110000101002’,....)

STEP 8: SELECT ITEM 7 **RICE**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘RICE’

WHERE ITEM_CODE IN

(‘8901144100507’,‘6281041122010’,‘071771554269’,‘9310140010533’,‘40111009840 97’,‘10006153’,‘6224000222253’,....)

STEP 9: SELECT ITEM 8 **JAM**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘JAM’

WHERE ITEM_CODE IN

(‘5027876065235’,‘617950410065’,‘10000717’,‘61795041003433’,‘7790580267186’,‘ 8412224026612’,‘4006424022693’,....)

STEP 10: SELECT ITEM 9 **HALVA**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘HALVA’

WHERE ITEM_CODE IN

(‘6281041012212’,‘10000229’,‘6223000053379’,‘6281100250456’,‘10000823’,‘62240 00719067’,‘6281041004712’,....)

STEP 11: SELECT ITEM 10 **MILK**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘MILK’

WHERE ITEM_CODE IN

(‘611106640621’,‘2305102000407’,‘6291007700015’,‘9415007014072’,‘10000304’,‘8 0001218’,‘6281022118421’,....)

STEP 12: SELECT ITEM 11 **SOAP**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘SOAP’

WHERE ITEM_CODE IN
('7610053826453','5410076732111','8901138513061','4015600939670','9501030310493','5413149805750','8853976000428,....)

STEP 13: SELECT ITEM 12 **TEA**
UPDATE DAILY_ORDERS SET NEW_ITEM = 'TEA'
WHERE ITEM_CODE IN
('070177067779','6009629720027','06281016000190','8690705025112','10000243','617950175506','4791021461005',....)

STEP 14: SELECT ITEM 13 **ANDOMI**
UPDATE DAILY_ORDERS SET NEW_ITEM = 'ANDOMI'
WHERE ITEM_CODE IN
('089686171372','5285000395164','089686120073','089686120110','089686120141','089686122169','896861209054',....)

STEP 15: SELECT ITEM 14 **PASTA**
UPDATE DAILY_ORDERS SET NEW_ITEM = 'PASTA'
WHERE ITEM_CODE IN
('8000380007219','6291007700275','6291007700015','8690705025112','5000183007457','8000350003036','6281033213023',....)

STEP 16: SELECT ITEM 15 **SHARIEA**
UPDATE DAILY_ORDERS SET NEW_ITEM = 'SHARIEA'
WHERE ITEM_CODE IN
('617950302155','5201020521512','089686120080','5285000392811','221048821295','5285000392002','8003773511560',....)

STEP 17: SELECT ITEM 16 **YOGURT**
UPDATE DAILY_ORDERS SET NEW_ITEM = 'YOGURT'
WHERE ITEM_CODE IN
('024911000027','10000656','2491830000083','2305102000117','735400365104','6223003880361','2305102000360',....)

STEP 18: SELECT ITEM 17 **MEAT**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘MEAT’

WHERE ITEM_CODE IN

(‘1976630021’,‘7290001316085’,‘019766300217’,‘6251038407116’,‘114’,‘6295130000042’,‘7896041169560’,....)

STEP 19: SELECT ITEM 18 **FLOUR**

UPDATE DAILY_ORDERS SET NEW_ITEM = ’FLOUR’

WHERE ITEM_CODE IN

(‘10000441’,‘6161106640089’,‘888970696995’,‘024913001473’,‘10000496’,‘5027876048641’,‘1114102000079’,....)

STEP 20: SELECT ITEM 19 **JUICE**

UPDATE DAILY_ORDERS SET NEW_ITEM = ’JUICE’

WHERE ITEM_CODE IN

(‘8412224044432’,‘6001240100066’,‘10000694’,‘5032619310022’,‘4006424185268’,‘5290040003283’,‘043000303993’,....)

STEP 21: SELECT ITEM 20 **OIL**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘OIL’

WHERE ITEM_CODE IN

(‘8410086751017’,‘6291003051890’,‘048327203438’,‘232300053640342’,‘8000550594631’,‘10000199’,‘6222000504171’,....)

STEP 22: SELECT ITEM 21 **EGGS**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘EGGS’

WHERE ITEM_CODE IN

(‘2279151269200’,‘10000366’,‘819052000339’,‘94150014072’,‘819052000308’,‘2626’,‘819052000315’,....)

STEP 23: SELECT ITEM 22 **TomatoPaste**

UPDATE DAILY_ORDERS SET NEW_ITEM = ‘TomatoPaste’

WHERE ITEM_CODE IN

(‘6917878007489’,‘011210600133’,‘048400152011’,‘8851978601018’,‘717273504028’,‘10000236’,‘6281022118421’,....)

STEP 24: SELECT ITEM 23 **CHIPS**

UPDATE DAILY_ORDERS SET NEW_ITEM = 'CHIPS'
WHERE ITEM_CODE IN
('5033876045085', '6281036170101' , '9556023096588', '4018077724121', '9556023674526', '5283003301113','6661036110602',....)

STEP 25: SELECT ITEM 24 **BREAD**

UPDATE DAILY_ORDERS SET NEW_ITEM = 'BREAD'
WHERE ITEM_CODE IN
('6281100321392','00004','5449000046390','5000171002808','2499153118102','617950144953','8852046100099',....)

STEP 26:

After selecting and merging 24 items from all items, now the researcher created a table contains only those 24 items.

CREATE TABLE **TRANS** AS SELECT SLIP_NO, New_Name FROM DAILY_ORDERS
WHERE New_Name IN (
GasWater, BISCUIT, CAKE, SUGAR, CHEESE, LENTILS, RICE,JAM, HALVA, MILK, SOAP, TEA, ANDOMI, PASTA, SHARIEA, YOGURT, MEAT, FLOUR, JUICE, OIL, EGGS, TomatoPaste, CHIPS, BREAD)

STEP 27:

Now adding columns of 24 items to **TRANS** table

ALTER TABLE TRANS ADD COLUMN (
GasWater VARCHAR2 (10),**BISCUIT** VARCHAR2 (10), **CAKE** VARCHAR2 (10), **SUGAR** VARCHAR2 (10), **CHEESE** VARCHAR2 (10), **LENTILS** VARCHAR2 (10), **RICE** VARCHAR2 (10),**JAM** VARCHAR2 (10), **HALVA** VARCHAR2 (10), **MILK** VARCHAR2 (10), **SOAP** VARCHAR2 (10), **TEA** VARCHAR2 (10), **ANDOMI** VARCHAR2 (10), **PASTA** VARCHAR2 (10), **SHARIEA** VARCHAR2 (10), **YOGURT** VARCHAR2 (10), **MEAT** VARCHAR2 (10), **FLOUR** VARCHAR2 (10), **JUICE** VARCHAR2 (10), **OIL** VARCHAR2 (10), **EGGS** VARCHAR2 (10),**TomatoPaste** VARCHAR2 (10),**CHIPS** VARCHAR2 (10), **BREAD** VARCHAR2 (10)).

Appendix E

ITEM 1:

UPDATE TRANS

SET **GasWater** =1

WHERE New_Name = '**GasWater**';

UPDATE TRANS

SET **GasWater** =0

WHERE New_Name != '**GasWater**';

ITEM 2:

UPDATE TRANS

SET **BISCUIT** =1

WHERE New_Name = '**BISCUIT**';

UPDATE TRANS

SET **BISCUIT** =0

WHERE New_Name != '**BISCUIT**';

ITEM 3:

UPDATE TRANS

SET **CAKE** =1

WHERE New_Name = '**CAKE**';

UPDATE TRANS

SET **CAKE** =0

WHERE New_Name != '**CAKE**';

ITEM 4:

UPDATE TRANS

SET **SUGAR** =1

WHERE New_Name = '**SUGAR**';

UPDATE TRANS

SET **SUGAR** =0

WHERE New_Name != '**SUGAR**';

ITEM 5:

UPDATE TRANS

SET **CHEESE** =1

WHERE New_Name = '**CHEESE**';

UPDATE TRANS

SET **CHEESE** =0

WHERE New_Name != ‘**CHEESE**’;

ITEM 6:

UPDATE TRANS

SET **LENTILS** =1

WHERE New_Name = ‘**LENTILS**’;

UPDATE TRANS

SET **LENTILS** =0

WHERE New_Name != ‘**LENTILS**’;

ITEM 7:

UPDATE TRANS

SET **RICE** =1

WHERE New_Name = ‘**RICE**’;

UPDATE TRANS

SET **RICE** =0

WHERE New_Name != ‘**RICE**’;

ITEM 8:

UPDATE TRANS

SET **JAM**=1

WHERE New_Name = ‘**JAM**’;

UPDATE TRANS

SET **JAM** =0

WHERE New_Name != ‘**JAM**’;

ITEM 9:

UPDATE TRANS

SET **HALVA** =1

WHERE New_Name = ‘**HALVA**’;

UPDATE TRANS

SET **HALVA** =0

WHERE New_Name != ‘**HALVA**’;

ITEM 10:

UPDATE TRANS

SET **MILK** =1

WHERE New_Name = ‘**MILK**’;

UPDATE TRANS

SET **MILK** =0

WHERE New_Name != ‘**MILK**’;

ITEM 11:

UPDATE TRANS

SET **SOAP** =1

WHERE New_Name = ‘**SOAP**’;

UPDATE TRANS

SET **SOAP** =0

WHERE New_Name != ‘**SOAP**’;

ITEM 12:

UPDATE TRANS

SET **TEA** =1

WHERE New_Name = ‘**TEA**’;

UPDATE TRANS

SET **TEA** =0

WHERE New_Name != ‘**TEA**’;

ITEM 13:

UPDATE TRANS

SET **ANDOMI** =1

WHERE New_Name = ‘**ANDOMI**’;

UPDATE TRANS

SET **ANDOMI** =0

WHERE New_Name != ‘**ANDOMI**’;

ITEM 14:

UPDATE TRANS

SET **PASTA** =1

WHERE New_Name = ‘**PASTA**’;

UPDATE TRANS

SET **PASTA** =0

WHERE New_Name != ‘**PASTA**’;

ITEM 15:

UPDATE TRANS

SET **SHARIEA** =1

WHERE New_Name = '**SHARIEA**';

UPDATE TRANS

SET **SHARIEA** =0

WHERE New_Name != '**SHARIEA**';

ITEM 16:

UPDATE TRANS

SET **YOGURT** =1

WHERE New_Name = '**YOGURT**';

UPDATE TRANS

SET **YOGURT** =0

WHERE New_Name != '**YOGURT**';

ITEM 17:

UPDATE TRANS

SET **MEAT** =1

WHERE New_Name = '**MEAT**';

UPDATE TRANS

SET **MEAT** =0

WHERE New_Name != '**MEAT**';

ITEM 18:

UPDATE TRANS

SET **FLOUR** =1

WHERE New_Name = '**FLOUR**';

UPDATE TRANS

SET **FLOUR** =0

WHERE New_Name != '**FLOUR**';

ITEM 19:

UPDATE TRANS

SET **JUICE** =1

WHERE New_Name = '**JUICE**';

UPDATE TRANS

SET **JUICE** =0

WHERE New_Name != '**JUICE**';

ITEM 20:

UPDATE TRANS

SET **OIL** =1

WHERE New_Name = ‘**OIL**’;

UPDATE TRANS

SET **OIL** =0

WHERE New_Name != ‘**OIL’**;

ITEM 21:

UPDATE TRANS

SET **EGGS** =1

WHERE New_Name = ‘**EGGS**’;

UPDATE TRANS

SET **EGGS** =0

WHERE New_Name != ‘**EGGS’**;

ITEM 22:

UPDATE TRANS

SET TomatoPaste =1

WHERE New_Name = ‘**TomatoPaste**’;

UPDATE TRANS

SET TomatoPaste =0

WHERE New_Name != ‘**TomatoPaste’**;

ITEM 23:

UPDATE TRANS

SET **CHIPS** =1

WHERE New_Name = ‘**CHIPS**’;

UPDATE TRANS

SET **CHIPS** =0

WHERE New_Name != ‘**CHIPS’**;

ITEM 24:

UPDATE TRANS

SET **BREAD** =1

WHERE New_Name = ‘**BREAD**’;

UPDATE TRANS

SET **BREAD** =0

WHERE New_Name != ‘**BREAD**’;