



The Use of Data mining Techniques
for Retail Customer Behavior
Analysis

by

Benjawan Onsang

Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science
in Information Technology
Assumption University

November, 2000

The Use of Data mining Techniques for Retail Customer Behavior Analysis

by

Benjawan Onsang



**Submitted in Partial Fulfillment of
The Requirements for the Degree of
Master of Science
In Information Technology
Assumption University**

November, 2000

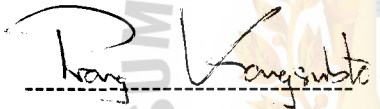
The Faculty of Science and Technology

Thesis Approval


Thesis Title	The Use of Data Mining Techniques for Retail Customer
By	Ms. Benjawan Onsang
Thesis Advisor	Dr. Prong Kongsubto
Academic Year	1/2000

The Department of Information Technology, Faculty of Science and Technology of Assumption University has approved this final report of the **twelve** credits course. **IT7000 Master Thesis**, submitted in partial fulfillment of the requirements for the degree of Master of Science in Information Technology.

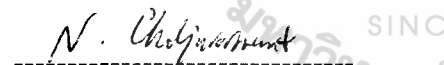
Approval Committee:



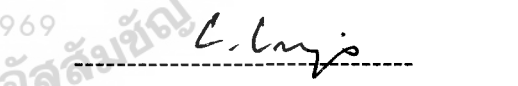
(Dr. Prong Kongsubto)
Advisor



(Dr. Thotsapon Sortrakul)
Committee

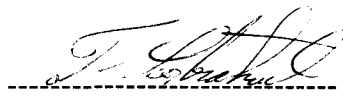


(Dr. Naruetep Choakjarernwanit)
Committee

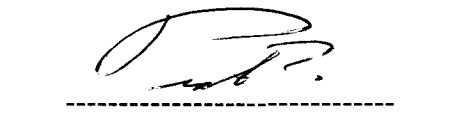


(Professor.Dr. Chiidchanok Lursinsap)
Representative of Ministry of
University Affairs

Faculty Approval:



(Dr. Thotsapon Sortrakul)
Director



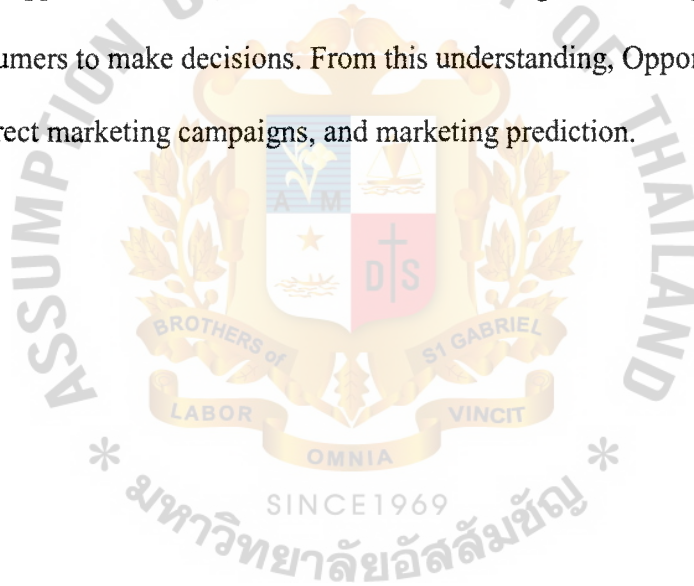
(Asst.Prof.Dr. Pratit Santiprabhob)
Dean

November /2000

ABSTRACT

The successful of current marketing strategy in which marketers try to reach the most customers' demand and expectation has been shown as an evidence. A lot of retailers try to imitate such a strategy onto their own business plan, especially the pricing method. However, it is not easy to do so because customers ' demand and expectation has changed over time and also varied per target group.

This thesis explores further how data mining technology and clear marketing strategies can be applied when there is a better understanding of the likely behaviors that drives consumers to make decisions. From this understanding, Opportunities flow for improved direct marketing campaigns, and marketing prediction.



ACKNOWLEDGEMENTS

I desire to express my most sincere appreciation and thanks to Dr. Prong Kongsombo for his encouragement and advice through the course of this study and for his help in the preparation and writing of this thesis.

I would here like to take the opportunity to thank all those without this master's thesis would neither have been completed. First of all, I would like to send thanks to The Mall Group Co., LTD. and FoodLand Supermarket LTD. supplying the source of information for this work. Many Thanks go to Ms. Kanchana Rateepon has been a coding constant of support during the long and hard months of work. Finally, I would like to dedicate this work to my beloved, whom had given me pressure.

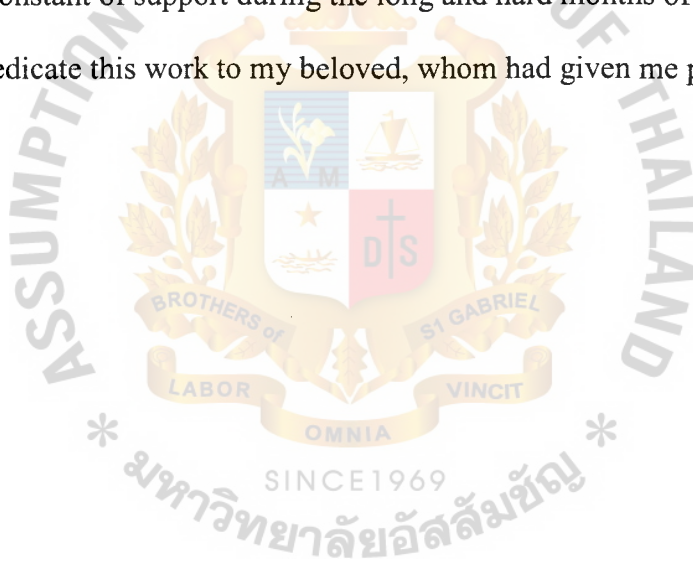


TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT	ii
LIST OF FIGURES	iii
LIST OF TABLES	v
CHAPTER 1 INTRODUCTION	
1.1 What is Data Mining?	1
1.2 How can we use the Data Mining methodology for analyzing customer behavior of Retail Supermarket?	6
1.3 Scope of work	6
1.4 Objective	6
CHAPTER 2 LITERATURE SURVEY	
2.1 Related Work	8
2.2 Customer Relation Management	13
2.3 The Problem of all data mining systems	14
CHAPTER 3 THE PROPOSE SYSTEM	
3.1 System Requirements and Specifications	16
3.2 Interface design	18
3.3 System Design	25
3.4 Mining Methodology Design	33

CHAPTER 4: SYSTEM IMPLEMENTATION

4.1 Overview of System Implementation	40
4.2 Resource Utilization	40
4.3 Data Collection and Enhancement	40
4.4 Modeling Strategies	42
4.5 Analyzing Results	46

CHAPTER 5 CONCLUSION AND FUTURE WORK

5.1 Conclusion	53
5.2 Future Work	55

REFERENCE

APPENDIX A: Installation	A
APPENDIX B: Database Entity Relation Diagram	B
APPENDIX C: Sample of Analysis Result and Comment from Retailer	C



LIST OF FIGURES

Figure 1-1	Data Collection and Associated Steps in the Data Mining Process	3
Figure 3-1	Interface Design (first Interface)	18
Figure 3-2	Interface Design (Second Interface)	19
Figure 3-3	First Interface of the Retail Customer Behavior Analysis	20
Figure 3-4	Second Interface of Retail Customer Behavior Analysis	21
Figure 3-5	SKU Listing	22
Figure 3-6	SKU Information	23
Figure 3-7	Data Period Selection Calendar	24
Figure 3-8	Retail Customer Behavior Analysis System	25
Figure 3-9	Operation Flow	26
Figure 3-10	Data Flow Diagram Level 1	30
Figure 3-11	Data Flow Diagram Level 2	31
Figure 3-12	Data Flow Diagram Level 3	32
Figure 3-13	RCBA System Overview	33
Figure 3-14	Input Process Flow 1	34
Figure 3-15	Input Process Flow 2	35
Figure 3-16	Input Process Flow 3	36
Figure 3-17	Input Process Flow 4	37
Figure 3-18	Mining Process Flow	38
Figure 3-19	Output Process Flow	39
Figure 4-1	RCBA Screen	47
Figure 4-2	2-Dimension Graph	48

Figure 4-3	3-Dimension Graph	50
Figure 4-4	RCBA Application Report by DAY (Sunday)	51
Figure 4-5	Comparison Transaction Count Graph	52
Figure A-1	Create New Data Source Screen	A-2
Figure A-2	Input ODBC Name "SQLDSN" and Description Screen	A-2
Figure A-3	ODBC Connection	A-3
Figure A-4	Select Database Name Screen	A-3
Figure A-5	ODBC Setting Screen	A-4
Figure A-6	Test create ODBC Screen	A-5
Figure A-7	Test ODBC Completed Screen	63
Figure B-1	Database ER Diagram	64



LIST OF TABLES

Table 1-1	Use of Data Mining by Modeling Objective and Learning Method	5
Table 4-1	Cross tabulate data Table	43
Table 4-2	Purchase Pattern	48



CHAPTER 1

INTRODUCTION

The real goal of business success both Retail and Wholesale is to gain the great profit that is how to apply marketing strategies appropriately. There are several approaches to reach such a purpose. According to Woolf (1996) [1], one basic approach is by switching the emphasis from item pricing to pricing based on total purchases. This is being achieved by cutting back significantly the deep-cut item pricing and instead are offering discounts varying with customers' spending behavior.

This means that we need the tool to monitor the customer activity on a daily, weekly, or monthly basis, and keep tracking the various elements of strategy without competitors knowing what changes are being made.

This thesis will apply data mining methodology to develop a tool for achieving such business goal.

1.1 What is Data Mining?

Data mining combines data analysis techniques with high-end technology for use within a process [6]. The concept of data mining is to develop usable knowledge with regard to market trends in the future. This thesis will examine the data mining process, describe the importance steps of data mining in general, and show how data mining can be used to predict customer behavior in retailing business especially supermarket.

The steps in the data mining process are:

- Identifying problem
- Collecting and Enhancing data
- Modeling strategies

- Analyzing results

a) Identifying Problem

The business process usually requires some solutions to resolve the business problem. Particularly, when we apply the data mining techniques into the business process, it is becoming mandatory to define the business problem properly as first priority. The problem's defining should neither be a discussion of the implementation nor efficacy of enabling technology. Instead, the problem definition should be emphasized on the business objective and process of rethinking.

A proper business objective must be clear, simplified language that focuses on the business problem and clearly states how the results are to be measured. In addition, the identifying problem should be included costs estimation, inaccurate prediction's result together with the advantage of making accurate analysis's result.

b) Collecting and Enhancing data

The significant aspect for analysis is algorithm of data mining can produce incomplete or biased data produce incomplete or biased models with typical blind spots. Defining the data sources should be a prominent part of details in the problem defining. As Figure 1-1 illustrates, data collection itself involves four distinct steps:

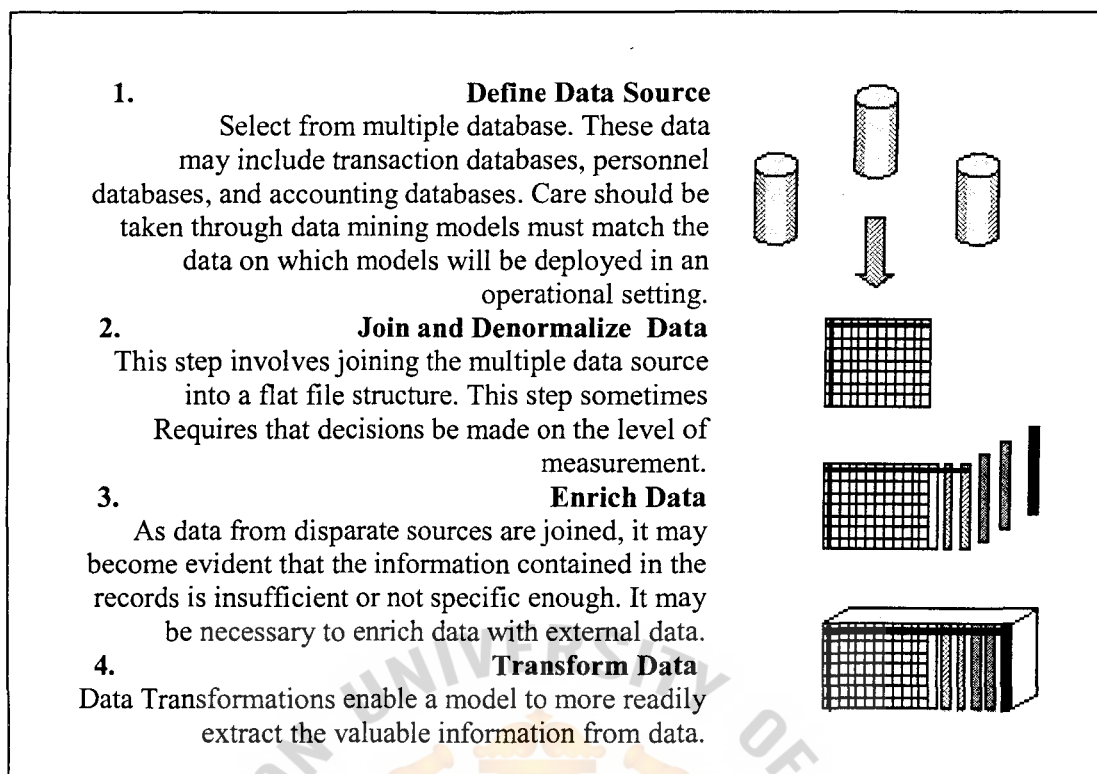


Figure 1-1 : Data Collection and Associated Steps in the Data Mining Process

c) Modeling Strategies

Data mining strategies can be classified into two categories: supervised learning and unsupervised learning [7]. Supervised learning methods are deployed when target variable exists with known values. Unsupervised learning methods are intended to develop data for which target variable does not exist with known values, but input variables still exists.

There are four modeling objectives: Prediction, Classification, Exploration, and Affinity.

- ◆ Prediction algorithms determine models or rules to predict continuous or discrete target values given input data.
- ◆ Classification algorithms determine models to predict discrete values given input data. Classification modeling tries to find models (or rules) that predict

the values of one or more variables in a data set (target) from the values of other variables in the data set (inputs).

- ◆ Exploration uncovers dimensionality in input data. For example, trying to uncover groups of similar customers based on spending habits for a large, targeted mailing is an exploration problem.
- ◆ Affinity analysis determines which events are likely to occur in conjunction with one another. Retailers use affinity analysis to analyze product purchase combinations.

A variety of the data mining technique depends on business goals and the data involved. Rarely does a data mining effort rely on a single algorithm to solve a particular business problem. In fact, multiple data mining approaches are often deployed on a single problem. Table 1-1 will map data mining techniques by modeling objective and how to use such techniques.

Modeling	Methods	Business Problems
Prediction	Regression and Logistic regression Neural Networks Decision Trees	Cash Used in ATM Cost of Hospital Stay Fraud Detection Customer Analysis
Classification	Decision Trees Neural Networks Discriminate Analysis Clustering	Segmentation Fraud Detection Customer Analysis
Exploration	Decision Trees Principle Component Clustering	Segmentation Retention Fraud Detection Customer Analysis
Affinity	Factor Analysis Sequences Associations	Market Basket Analysis

Table 1-1 : Use of Data Mining by Modeling Objective and Learning Method

d) Analyzing Results

Diagnostics use in model evaluation varies in supervised and unsupervised learning. To classify problems, analysts typically review the reports, lift and profit charts, threshold charts, confusion matrices, and statistics of fit for the training and validation sets, or for the test set.

1.2 How can we use the Data Mining methodology for analyzing customer behavior of Retail Supermarket ?

The data mining methodology helps identify products that sell together under certain condition. The Retailer correlates the actual customer shopping cart combination by item or SKU (Stock Keeping Unit) by the day of week. As a result, customer purchasing behavior can be identified accurately. Therefore, the analysis helps generate a set of business rules that guide business activities such as a targeted marketing promotion and product ordering.

1.3 Scope of work

- To develop a tool which capable of monitoring and predicting the consumer's Mining model for monitoring the retail supermarket customer behavior and their activity.
- To analyze 2 months of data from store and create purchasing pattern from the analysis result.
- To examine the result of the analysis to maximize the profits

1.4 Objective

1.4.1 Business Purpose

Customer Behavior Analysis is a powerful concept that promises to change the very dynamics of the retail business. However to fully realize its potential, Customer Behavior Analysis must be viewed in light of the Retailer's other initiatives.

Customer Analysis delivers business benefits as the following ways.

- **Efficient Forecasting**

Customer Behavior Analysis and Supply Chain Management initiates to drive the move from supply-based forecasting to demand-driven forecasting.

Retailers apply Customers' Behavior Analysis and Supply Chain Management

to better match demand with product supply and maximizing product margin contribution.

- **Efficient Brand Management**

Customer Behavior Analysis drives the synergy between promotions and even marketing and direct measurement as demonstrated by the changes in Customer behaviors. Retailers apply Customers' Behavior Analysis to better predict promotional, event effectiveness and use the predictive information to optimize the purchasing system.

- **Efficient Precision Marketing**

Precision Marketing is a business activity whose goal is to leverage the data that organization has about Customer in order to develop offerings targeting specific destination product. The Retailers learn to track the customer purchasing behaviors and then to improve marketing effectiveness.

- **For better Decision Making**

The Retailer applies the knowledge gained from Customer Behavior Analysis to make better decisions in merchandising, marketing, advertising, promotions and inventory

1.4.2 For Information Technology Students

- To understand the retailing business and business processing
- To understand Data Mining techniques and Model Strategies
- To understand the database design, especially the Relational Database concept
- To study Microsoft SQL Server version 7.0
- To enable student to improve thinking process and skills in designing and analyzing the system.

CHAPTER 2

LITERATURE SURVEY

2.1 Related Work

There are many works and systems, which related to the study and work of Data Mining Methodology. The evolution of the applications for Data mining and case study, which related are as following.

The system that will be referred first was made in 1992, Intergral Solution Ltd. initiated “Project Clementine” [26], aiming to build a comprehensive data mining system accessible to business and professional end-user such as doctors. Professionals can take part directly in data mining and thus access the implications of using data mining in safety-critical or service-critical applications. The techniques work by “learning” patterns in data and find patterns to make predictions. Every clinical act and its outcome is recorded. Patient records are being transferred to electronic form. Pharmaceutical knowledge increases daily, with new compounds, new dosage pattern. The user’s business or clinical knowledge is essential to determine what factors to consider and how the various input factors may need to be combined.

The second system concerned Police Force performance in the UK the Crime and Disorder Act (1998) [10] places a statutory responsibility on the 464 District Authorities to perform crime pattern analysis. Most police forces have geographically referenced crime data, which are developing on-line crime reporting systems with GIS (Geographical Information System), or digital map based command and control computer systems.

T.J. Watson Research Center [12], IBM Research Division, USA has developed “**UPA application**”. The UPA (Underwriting Profitability Analysis)

application embodies a new approach to mining Property & Casualty (P&C) insurance policy and claims data for the purpose of constructing predictive models for insurance risks. UPA utilizes the IBM ProbE (TM) Probabilistic Estimation based predictive modeling class library to discover risk characterization rules by analyzing large and noisy insurance data sets. Each rule defines a distinct risk group including its level of risk. To satisfy regulatory constraints, the risk groups are mutually exclusive and exhaustive. The rules generated by ProbE are statistically rigorous, interpretable, and credible from an actuarial standpoint. The ProbE library itself is scalable, extensible, and embeddable. The approach to modeling insurance risks and the implementation of that approach have been validated in an actual engagement with a P&C insurance firm. The benefit assessments of the results suggest that this methodology provides significant value to the P&C insurance risk management process.

The next system was developed by the Detection and Management of Fraud in UMTS Networks System at University of London. It is called “**ACTS project**”.

ACTS project AC095, Advanced Security for Personal Communications Technologies (ASPeCT), is engaged in the advancement of security issues for the next generation of mobile communications UMTS. One of the work packages within this project is developing fraud detection and management tools for the GSM network. Prototypes of three different fraud detection tools have been developed, and demonstrated, using Rule-Based and Neural Network technologies.

Another work has been done by Nautilus Systems, Incorporated [18], Virginia. It is a business and computer consulting firm focusing on data mining and data warehousing. It has developed the system for the Northridge, California Earthquake. The data were collected during the Northridge, California earthquake occupied

several warehouses, and ranged from magnetic media to bound copies of printed reports. Nautilus Systems personnel sorted, organized, and cataloged the materials. Documents were scanned and converted to text. Data were organized chronologically and according to situation reports, raw data, agency data, and agency reports. For example, the Department of Transportation had information on highways, street structures, airport structures, and related damage assessments.

Nautilus Systems applied its proprietary data mining techniques to extract and refine data. Geography was used to link related information, and text searches were used to group information tagged with specific names (e.g., Oakland Bay Bridge, San Mateo, Marina). The refined data were further analyzed to detect patterns, trends, associations and factors not readily apparent. At that time, there was not a seismographic timeline, but it was possible to map the disaster track to analyze the migration of damage based upon geographic location. Many types of analyses were done.

In the United State, Gas Processing Plant Project was carried out for an oil company and was based in a remote US oil field location. The process investigated was a very large gas processing plant which produces two useful products from the gas in the wells, natural gas liquids and miscible injection. NGL is mixed with crude oil and transported for refining, and MI is used to improve the viscosity of oil in the fields to improve crude oil recovery. The aim of the study was to use data mining techniques to analyze historical process data to find opportunities to increase the production rates, and hence increase the revenue generated by the process. Approximately 2000 data measurements for the process are captured every minute.

One of the country's largest banks, Bank of America with headquarters in San Francisco, executives also knows the long road to data mining. It took four years for

IT and marketing teams to go from simple queries to more sophisticated ones that help them understand their customer clearly not just of their own products.

The pictures that have emerged are prompting The Bank of America marketing executives to launch a new array of services for a wide variety of customers in the early 1998.

18648 e-3

Bank of America was interested in finding new ways to retain current checking-account customers while recruiting new ones. Rather than taking the drastic, and perhaps unprofitable, step of simply cutting prices or offering free checking for life, the bank's marketing executives wanted to find out what kind of customers tended to use which products. Further, they wanted their customer data to tell them if a different mix of products and services might better meet the needs of certain customers.

A subsidiary of the National Australia Group, the Northern Bank [14] has a new data mining application now being used in each of the 107 branches in the Province. The new system is designed to deliver financial and sales information such as volumes, margins, revenues, overheads and profits. The application consists two sub systems as a system to integrate the multiple data sources into a consolidated database and another system to deliver that information to the users in a meaningful way. The new system delivers management information in electronic form to the branch network. The information is now more accessible, paperless and timely.

Information Harvesting Inc. (IH), founded in 1994 and based in Cambridge, Mass, IH have been developed the data mining software to solve the problem deriving meaningful information from enormous amounts of complex data. It makes use of conventional statistical analysis techniques by building upon a proprietary tree-based learning algorithm that generates expert-system-like rules from data sets,

initially presented in forms such as numbers, dates, categories, codes, or any combination.

Two examples of companies using the data mining software are below :

Michael Reese Medical Associates (MRMA) employed data mining software from Information Harvesting and Vantage Point as a tool for gaining advantage in contract negotiations. The 28-doctor group had to predict trends in type, price, location, and use of service, since they must negotiate with insurance companies to provide certain services at a set monthly fee, doctors must accurately predict their per member/per month cost to break even or make a profit. Normally physicians could only make an intuitive estimate roughly based on after-the-fact evaluations of prior estimates when determining this critical figure whereas data mining offered a new approach.

The Philadelphia Police and Fire Federal Credit Union (PFFCU) used data mining to maximize their membership base by cultivating multiple relationships (e.g. consumer loans, annuities, credit cards, etc.) with members. Because the membership base is extremely homogeneous (police and fire dept. employees and their families), data have to be deeply drilled to identify segmented groups. Used in conjunction with software such as Inter Global Financial Systems' Credit Analyzer, Information Harvester identified members most and least profitable to the organization as well as those who would make attractive loan candidates. Data mining often led PFFCU to accurate but counter-intuitive results. For example, members who had filed for bankruptcy were more inclined to clear debts with the Credit Union than outside lenders. Thus, PFFCU identified members with imperfect credit histories but a strong tendency to pay, whereas these individuals would be ignored by large conventional lenders.

Customer loyalty is hard to win and even harder to keep, as consumers are increasingly pressed for time and faced with an overwhelming number of choices. Not having the right product at the right place at the right time could mean the loss of many previously loyal customers. Hallmark Card Incorporated [22] the leader in the greeting card industry in U.S is getting closer to market and making better decisions with its point of sale (POS) data warehouse and mining the data by using MicroStregy's products. Hallmark has built a warehouse and decision support system application to better understand customer purchase patterns. The data warehouse and data mining are enabling Hallmark to improve decision making for determining assortments, promotions, advertising, inventory management, pricing and product development.

MINEvision N.V. [21], Belgium has been develop Data mining system for the Retailing Industry and they demonstrates a data mining project in the retailing industry that was carried out for vendor a fully automated convenience stores in Belgium. The company faced the problem how to display its products to the customer in order to maximize cross-selling opportunities. Indeed, it was because of the limited shelf space in the convenience store that can only stock about 200 items. It is critical to offer the right mix of products at the right position in the store.

The concept of all cases study will extract and refine data to develop usable knowledge or turn the data into usable information and find pattern to make predictions and make a good decision in their business.

2.2 Customer Relation Management

Almaden Research Center [13] is the one of eight IBM Research Division facilities worldwide and a premier industrial research laboratory which is Located about 55 miles southeast of San Francisco. One of their research concerns

Customer Relation Management (CRM) and Product affinity analysis. The later is a part of CRM which identifies products that sell together under certain conditions (e.g., time of day, customer demographics, shopping purpose). With product affinity analysis, the retailer correlates the actual market basket combination by item. The analysis generates a set of business rules that guide business activities such as product ordering and replenishment, inventory control and in-store merchandising. Another aspect of analysis is a targeted marketing campaign could be launched to motivate a change in behavior.

The research has been done with a Canadian grocery chain according to product affinity analysis. The impact of small improvement in the behavior of customers on the store's profitability has been figured out. Given its fixed cost structure, it has been found that even small improvement in any one customer's behavior led to very significant profitability gains.

2.3 The Problems of all data mining systems

- The rule garbage in garbage out applies to data mining project as well. Indeed, the quality of the data is vital in all data mining projects and will eventually determine its success to a large extent. Consequently, the identification of data limitations and how can be resolved, must be carefully analyzed in advance.
- Some rules are inexplicable : Seem to have no explanation and do not suggest a course of action.
- The analysis works best when all items have approximately the same frequency in the data. Items that rarely occur are in very few transactions and will be pruned.
- The function of an algorithm is to find a pattern, perhaps a coincidence in the data, only human expertise in the form of knowledge of the meaning and context of data, can decide both how the data might reasonably be analyzed and how to

interpret and evaluate any results produced. It uniquely remains the responsibility of the professional to rigorously test any model or prediction and crucially to decide whether to act on its recommendations.

Almost of Customer Relation management are using data mining technique to explore customer's behavior and to predict and make decision for marketing strategies. In addition to Sale Transaction, CRM may need also the customer information such as their age, sex, address, etc. Then CRM analysis result can help to specify the target customer.

However, the characteristic of Supermarket-shopping goes to buy simply and quickly. This means the customers just walk in, buy goods and walk out so retailer could not identify or get the exact customer information. As a result, the available data are just the sale information that will be used for the analysis.

This thesis is going to explore Supermarket customer's behavior context of Bangkok and make purchase pattern by data mining methodology and give the result by statistics. If the result does not suitable, the reason and comments for that will be explained from the experts and the retailers.

CHAPTER 3

THE PROPOSE SYSTEM

3.1 System Requirements and Specifications

To identifying the methodology for achieving the real goal of Retail business, the study includes observing in Supermarket business, discussing with Retail Executive Management person and studying Data Mining that how it can be used to analyze Customer's behavior.

Generally, the top requirement of Retail Business is similarly to other businesses that try to gain profit as much as possible, in order to achieve such a goal, appropriate marketing strategies must be planned and used.

To clarify the problem, data mining methodology and techniques are recommended develop a model that will help retailer plan and apply with market strategy in the most effective way such as offering discounts varying with Customer's spending behavior.

3.1.1 User requirements

The Retailer will get insight into the merchandise by telling which products tend to be purchased together and which are most amenable to promotion and simply be measuring the success of previous marketing campaigns.

To perform the user requirement and full fill all the function and non-function of the customer behavior analysis.

- The system will use the transaction data for finding the specify purchase patterns.
- The result of system should be useful for decision making marketing, advertising, promotions and inventory.

- The analysis might predict promotional and event effectiveness, and use the predictive information to optimize the purchasing system.
- It must be Real-time interactive
- The system should be user friendly and uncomplicated for use.

3.1.2 System Specifications

After the requirement has been specified, the prototype was developed to complete these requirements and it was called “Retail Customer Behavior Analysis”

There are some specifications of the system in terms of software; hardware.

Software

- Application Tool should provide the environment and input screen where the retailers are able to input the interesting SKU and the data period
- After the retailer has entered specified SKU, the real-time important will be displayed on the screen. On the other hand if the SKU does not exist in the system, error message will be shown.
- For non function, system should be available for retailer to select SKU code from listing by clicking drop down at SKU field
- While the system is processing, the status must show on.
- The system will analyze the model and generate the result as two dimension graph, three dimension graph and the analysis reports.
- For summarizing the analysis result, the system will show the pattern or rule of the customer’s purchase behavior by percentage of correlation.
- All the analysis result will be printed out to printer.

Hardware

- Input device : users interact with the Retail Customer behavior analysis system by using mouse and keyboard.
- Output :
 - Visual display – regular color CRT monitor.
 - Printer – require Laser printer
- Computer : power of PC with high performance processor.

3.2 Interface design

3.2.1 Design The Virtual Environment

The interface was developed to meet user requirement designing, the first Interface will begin with the major function and non-function as shown in figure 3-1
The contain will be included 3 options :

- 1) Analysis option, to go inside of the analysis program.
- 2) About option, to show about the thesis and software.
- 3) Exit option, to terminate or exit from the program.

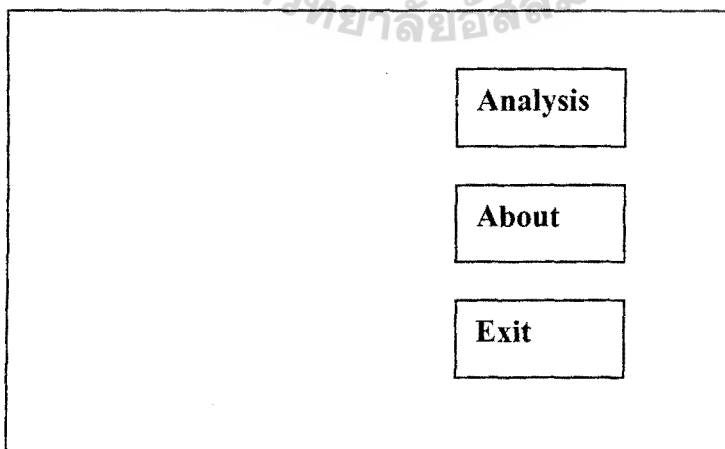


Figure 3-1 : Interface Design (first interface)

As the retailer want to see the analysis, the input screen of Analysis option will be shown as Figure 3-2. This screen will be used for input screen and process command button.

SKU :

Department :

Period : To :

Figure 3-2 : Interface Design (Second interface)

For the input field of SKU code, the system will provide the list tool for easy access to SKU and display related description before chosen. For period selection, Application will provide the calendar for more comfortable. The important command button is for starting process and status of process will be showing while processing.

3.2.2 Interface of The Retail Customer behavior analysis

This interface consists 3 main options as following :

- analyze the data
- show about thesis and software
- terminate or exit from program, as shown in Figure 3-3.

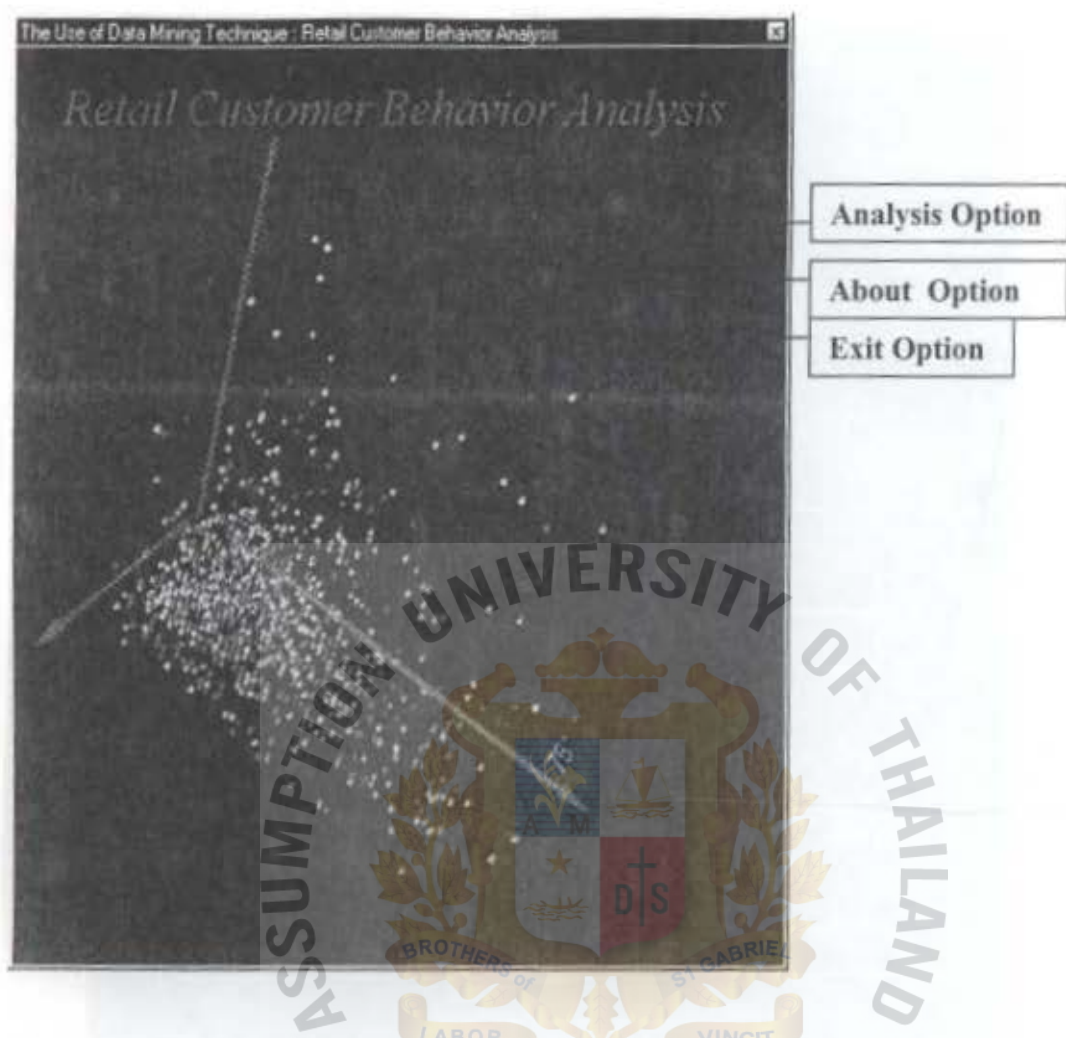


Figure 3-3 : First interface of the Retail Customer Behavior Analysis

For the second interface screen is shown as Figure 3-4 : this sub screen will show over the first screen after the retailer has select "Analysis" option. This screen is for entering or specifying SKU and after finish searching, screen will display all interested SKU's information. The next field is for choosing the data period. For the sample of this thesis, the data is available from October 1, 1999 to November 30, 1999

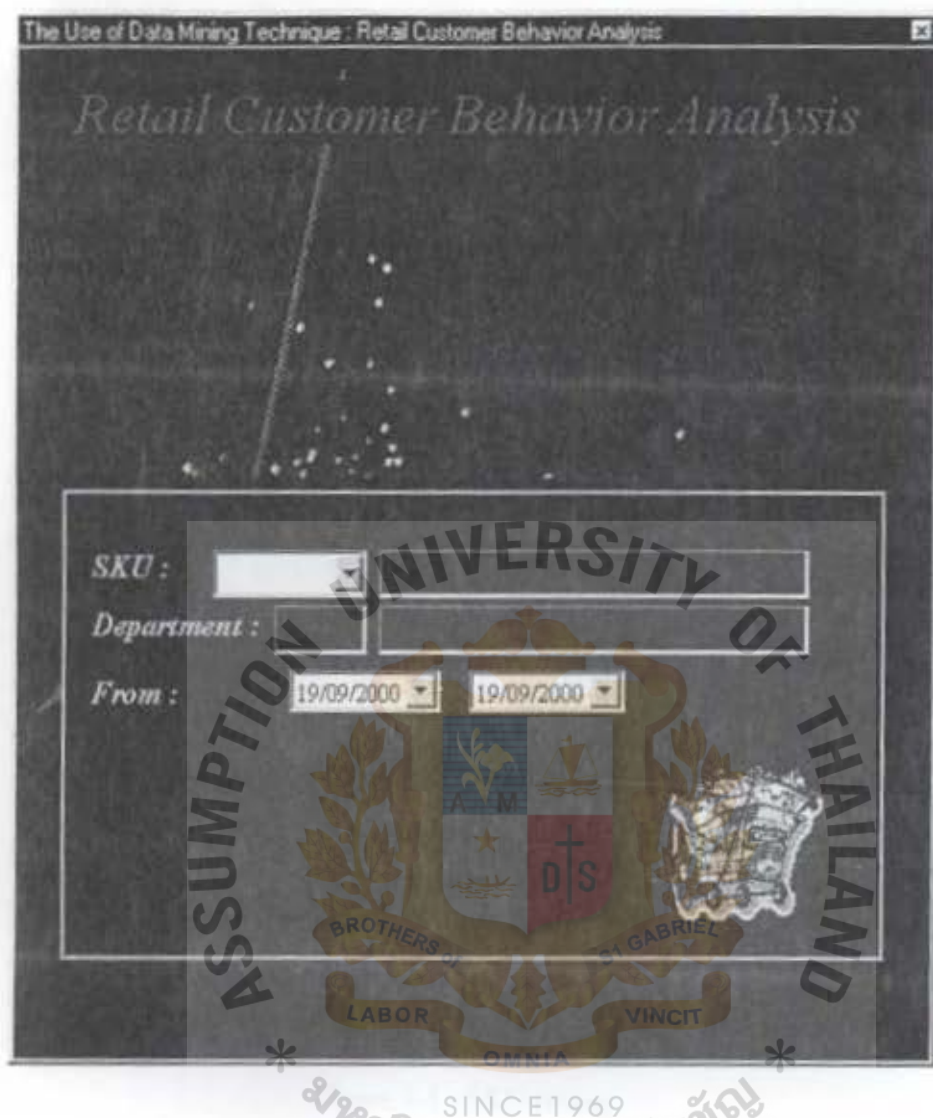


Figure 3-4 : Second Interface of Retail Customer Behavior Analysis

At SKU entry field, Application provides the retail to be able to select SKU from SKU list as Figure 3-5. In listing of SKU Table, it will be included SKU code, SKU description, Department code and Department description and if user wants to see fully list, just press Vertical scroll bar and Horizon scroll bar.

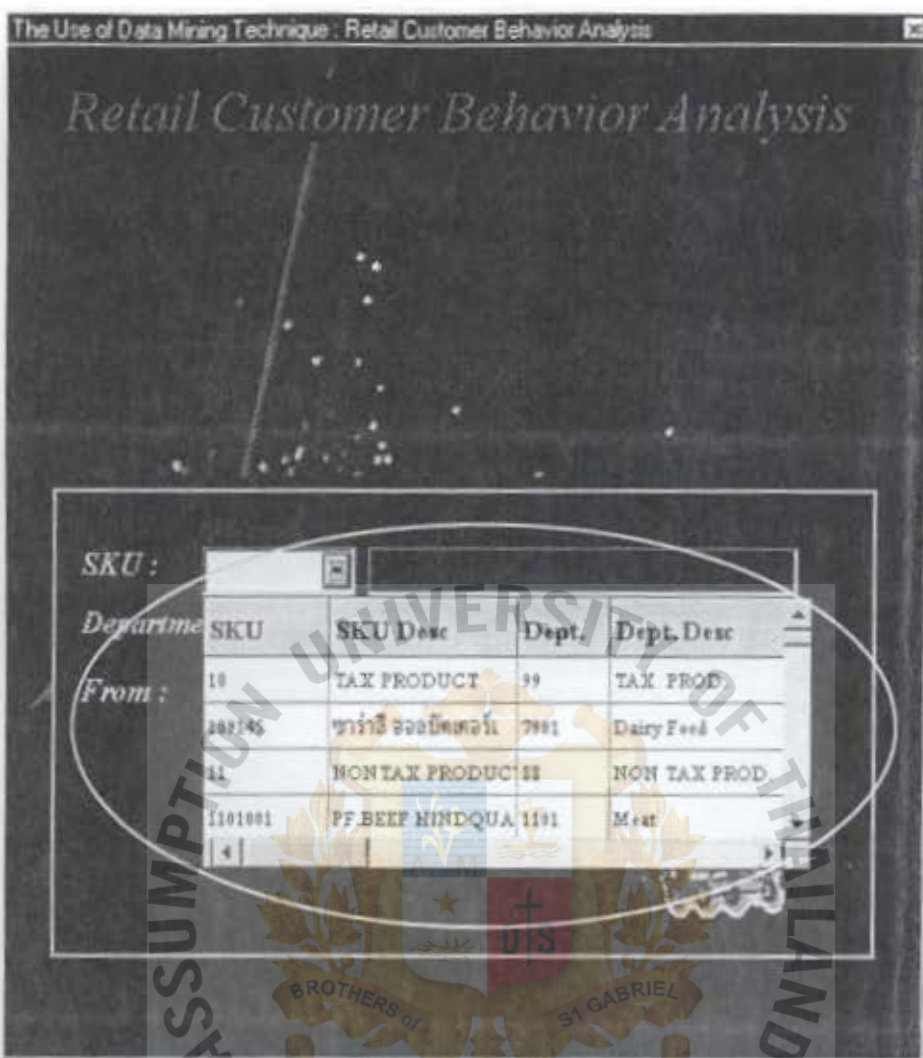


Figure 3-5 : SKU listing

In Figure 3-6 : it will show the SKU information after entering SKU. For example the retailer enters SKU code 3402237 then the information will show as SKU description "DUBBLE AMPLEX MINT", Department code 3402 and Department description "Grocery"

Retail Customer Behavior Analysis

SKU : 3402237 DUBBLE AMPLEX MINT
Department : 3402 Grocery
From : 20/09/2000 20/09/2000



Figure 3-6 : SKU Information

After entering SKU, retailer will be able to input the specify data period or select from Calendar screen which provided as Figure 3-7.

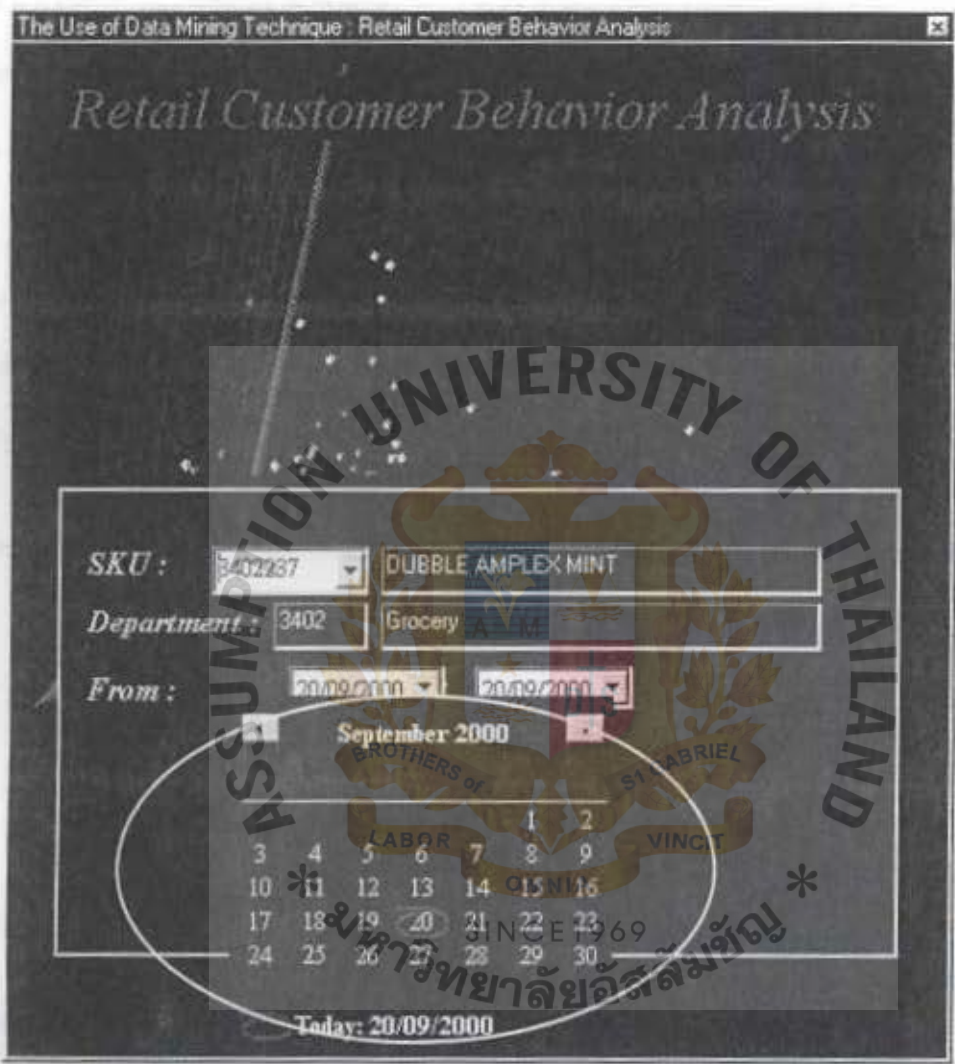


Figure 3-7 : Data period selection calendar

3.3 System Design

3.3.1 System Overview

To study and analyze the customer behavior, the system needs Prices Database which include SKU(Stock Keeping Unit) information and Department information and Sales Transaction Database from Supermarket store. After the data is completely put, it will be transformed to RCBA Database.

The Retailer can use the result that identify customer purchasing behavior from the use of data mining model. The analysis result will be represented via the Analysis Reports and Graphical Analysis. System Overview can represent by Figure 3-8.

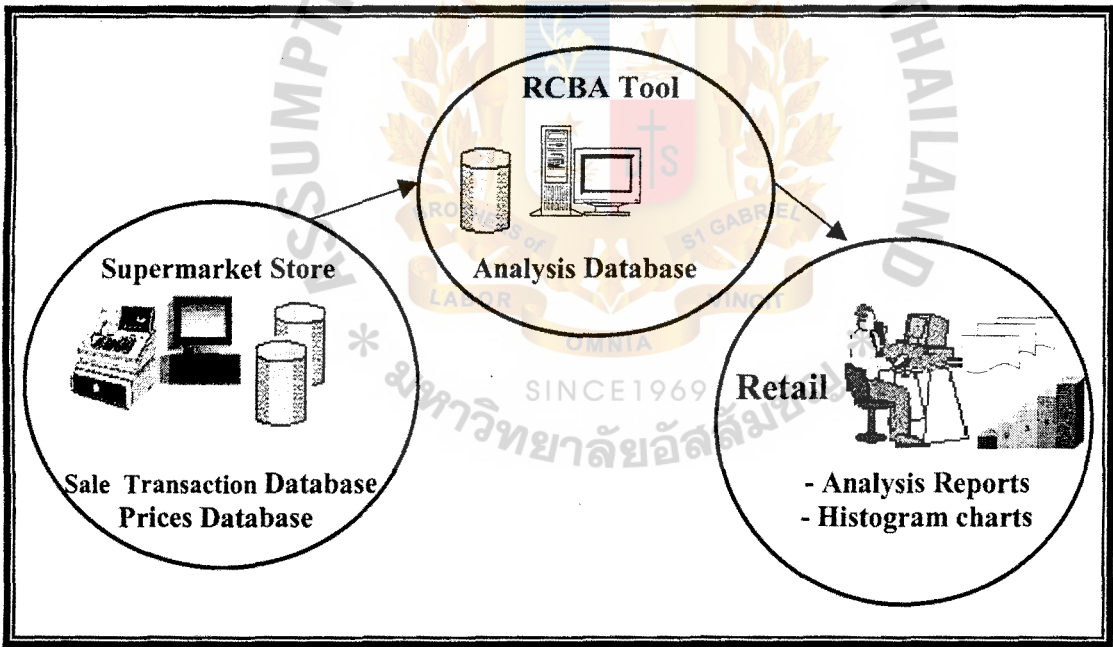


Figure 3-8 : Retail Customer Behavior Analysis System

3.3.2 Operation Flow

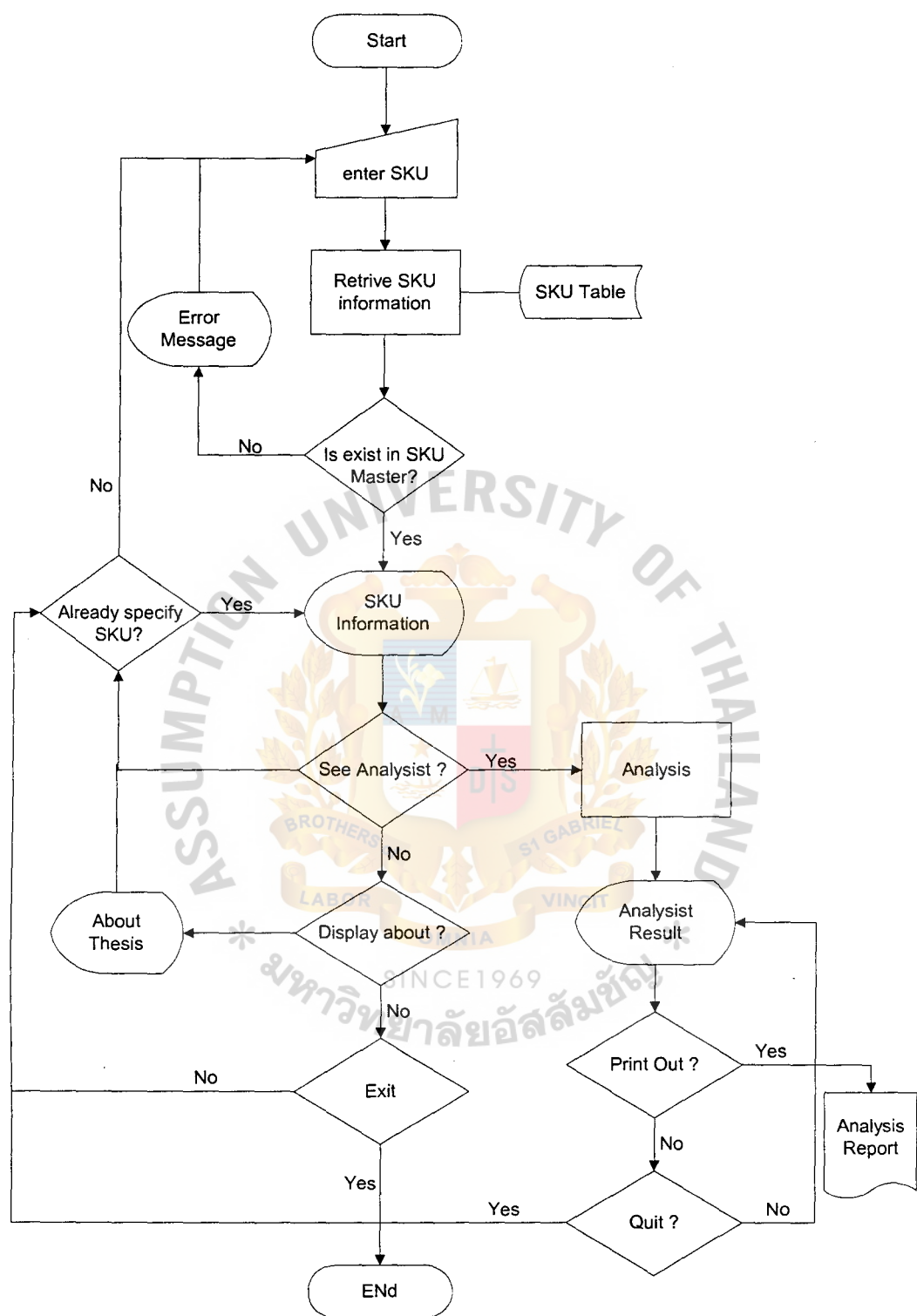


Figure 3-9 : Operation Flow

3.3.3 System Process

The System Process will be represented by Data Flow Diagram Figure 3-10 .

Data Flow Diagram consists of 4 main processing 1.0 Export Data, 2.0 Enrich Data, 3.0 Transform Data and 4.0 Mining Data, each processing will represent by a rounded rectangular symbol.

Process narrative

Process 1.0 : Export Data

This process will export data from POS (Point of Sale) System to flat text files.

Text files will separate 2 category as Data Master and Sale Transaction. The output result from this process consists of 4 flat files at the first time. Only 2 flat files are remaining next time because it is not necessary to export Data Master every time.

Data Master

- ◆ SKU files will be exported from some data field of POS SKU table by SQL command according to SKU code, description, price and their department code
- ◆ Department file will be exported from some POS Department table by SQL command according to department code and description

Sale Transaction

- ◆ Transaction Header file will be exported from some data field of POS Transaction header by SQL command according to POS number, Transaction Number, Store Number, Transaction type and Transaction Total amount.
- ◆ Transaction Detail file will be exported from some data field of POS Transaction detail by SQL command according to POS number, Transaction Number, Row Number (sequence of item of business day), line item number in the transaction, SKU code (item sold) and Quantity of item per line item.

A. Process 2.0 : Enrich data

This process is necessary for Transaction header and Transaction Detail because POS System always keep transaction data day by day. If requirement of user need to specify data date, Application should add one more field of business day to Transaction header and Detail

B. Process 3.0 : Transform data

It is vital to transform data to more readily extract valuable information data by applying SQL7.0 import wizard to RCBA Database. The result of this process at the first time, will create 4 table: skuTable for SKU master, deptClassTable for Department master, xTran for Transaction header and xItem for Transaction detail and the later time just insert for new Transaction header and Transaction detail for more data.

C. Process 4.0 : Mining data

This process is the most important for Data mining methodology that will show the analysis result from the database by 2D (2 dimension) graph, the analysis report during the period, 3D (3dimension) graph and the analysis report during the period

D-1 Process 4.1 : Input SKU

To input specify SKU (target SKU) and data period after input SKU code if the SKU code exists in RCBA database, system should show on screen in form of SKU description, department code and department description which that SKU is related to. (one department might have more than one SKU) otherwise the system will show the error message and Retail must input once again for the correct SKU code.

D-2 Process 4.2 : Select Sales Transaction

After the process 4.1 receives SKU code and data period, system will seek only the Sale Transaction which sold the specify SKU by check dTran field that can define the transaction type then the system will search in Transaction detail for the line item in the transaction.

D-3 Process 4.3 : Analysis

The Analysis process will provide the result of Analysis by Graph and Analysis report of 10 Top SKU that sold together with the Target SKU.

D-3.1 Process 4.3.1 : Count Transaction

To count transaction of item which sold together with specify SKU if any item has duplicate line item in the same transaction, the system will count one.

D-3.2 Process 4.3.2 : Select Top(10) counting

From the result of process 4.3.1, Retailers will be able to know the counting of the item that sold together and this process will select the Top ten of counting by descending.

D-3.3 Process 4.3.3 : Calculate %

This process will calculate transaction counting percentage compared with Transaction counting of target SKU to show the analysis result by 2D graph and the analysis report.

D-3.4 Process 4.3.4 : Calculate % by Day of week

This process need to use Transaction counting that separate by day of week and calculate percentage compared with transaction counting by day of week. The result of this process is a 3D graph and the analysis report by day of week.

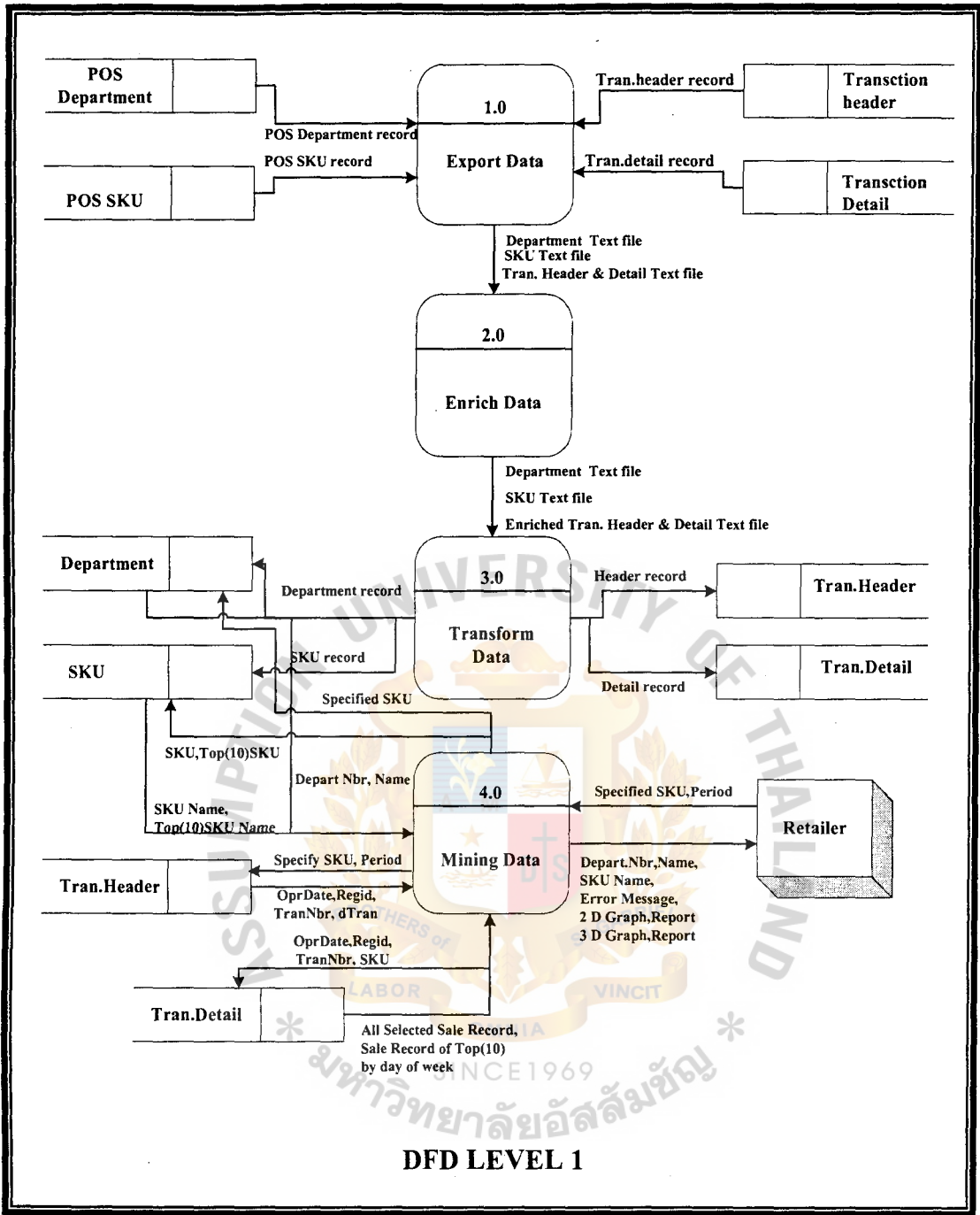


Figure 3-10 : Data Flow Diagram Level 1

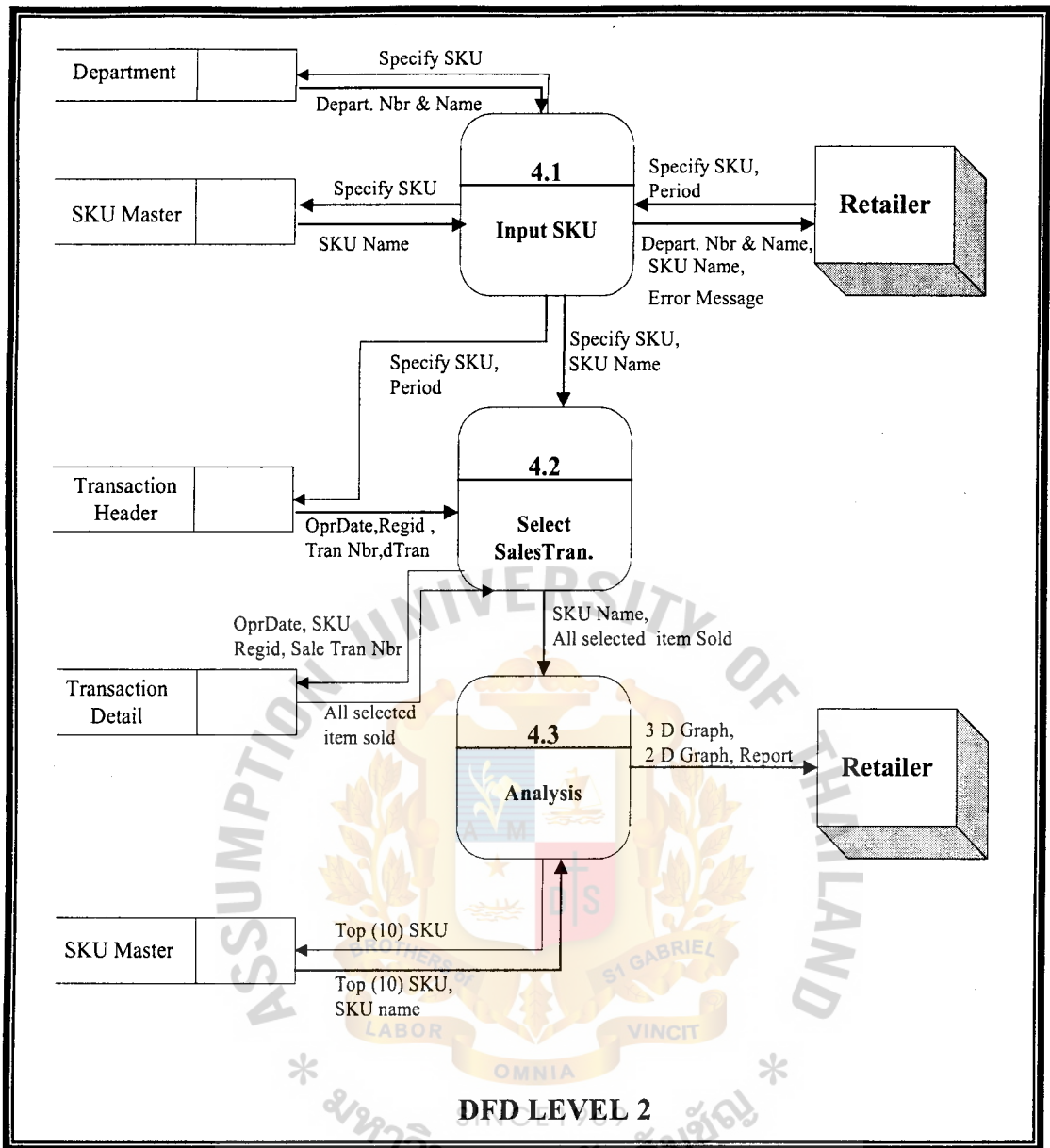


Figure 3-11 : Data Flow Diagram Level 2

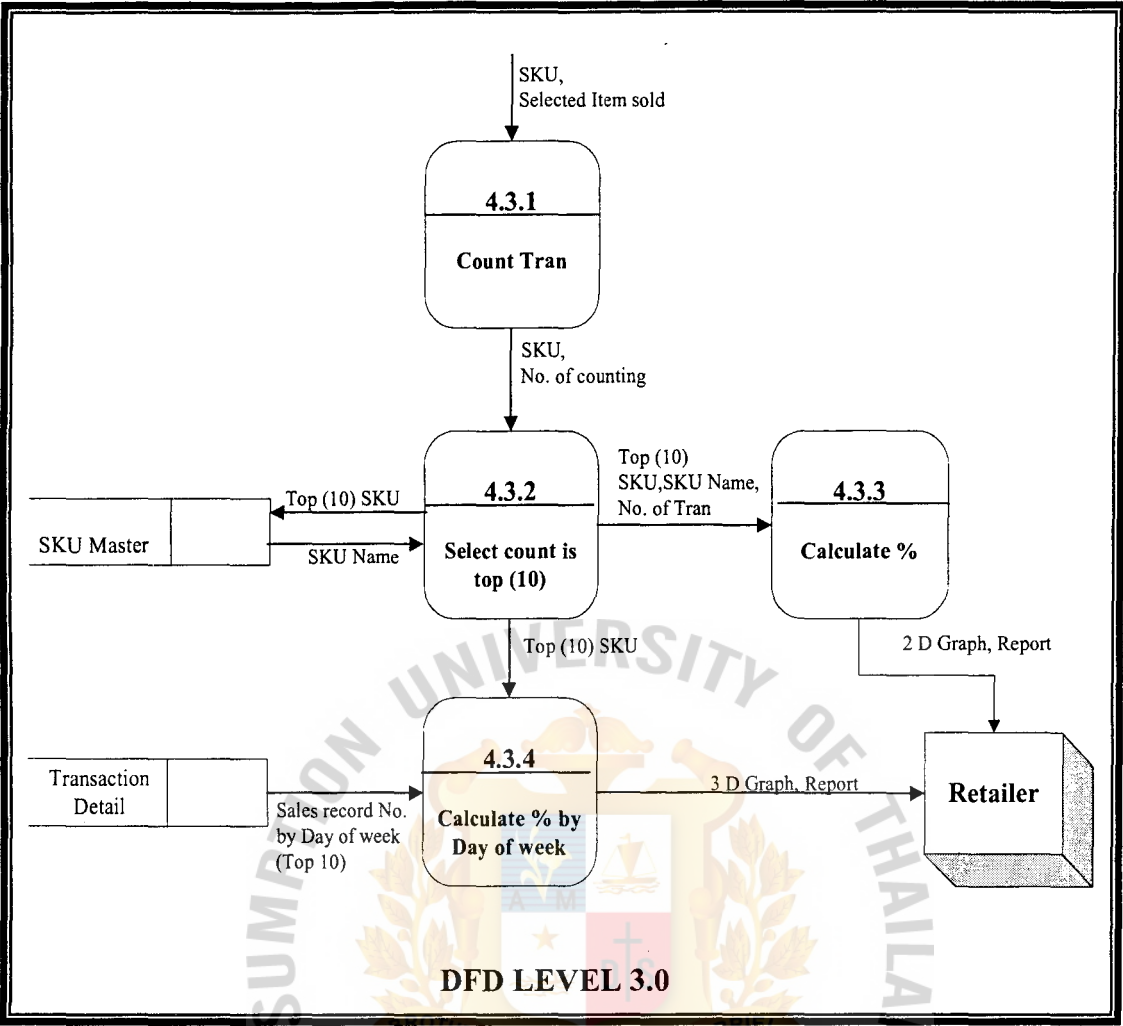


Figure 3-12 : Data Flow Diagram Level 3

3.4 Mining Methodology Design

3.4.1 Steps of Using Software

There are four important steps for using Retail Customer Behavior Analysis (RCBA) software to show how to mine data and how to generate Analysis result.

- 1) Specify interesting SKU
- 2) Select data period
- 3) Start mining process
- 4) See the Analysis Result

3.4.2 Mining Process design

The components of the system consist of Input as interesting SKU, Process and Output as Figure 3-13. To design the whole RCBA system, each component is related.

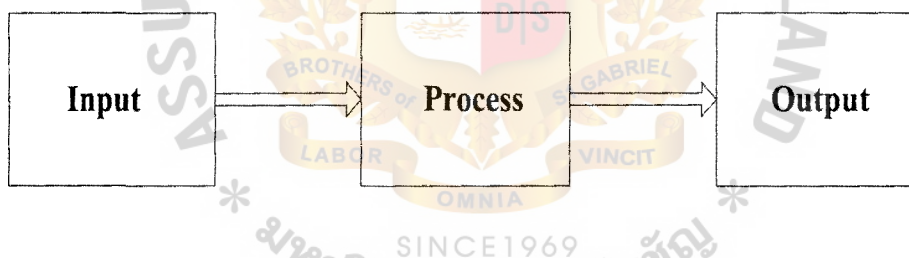


Figure 3-13 : RCBA System Overview

3.4.3 RCBA System Flow

After User enters SKU and Data period then order to start analysis processing, the system flow of RCBA program can be separated into 3 parts, first part for input process as on Figure 3-14, Figure 3-15, Figure 3-16 and Figure 3-17; Second part for mining process as on Figure 3-18. and the last for output process as in Figure 3-19.

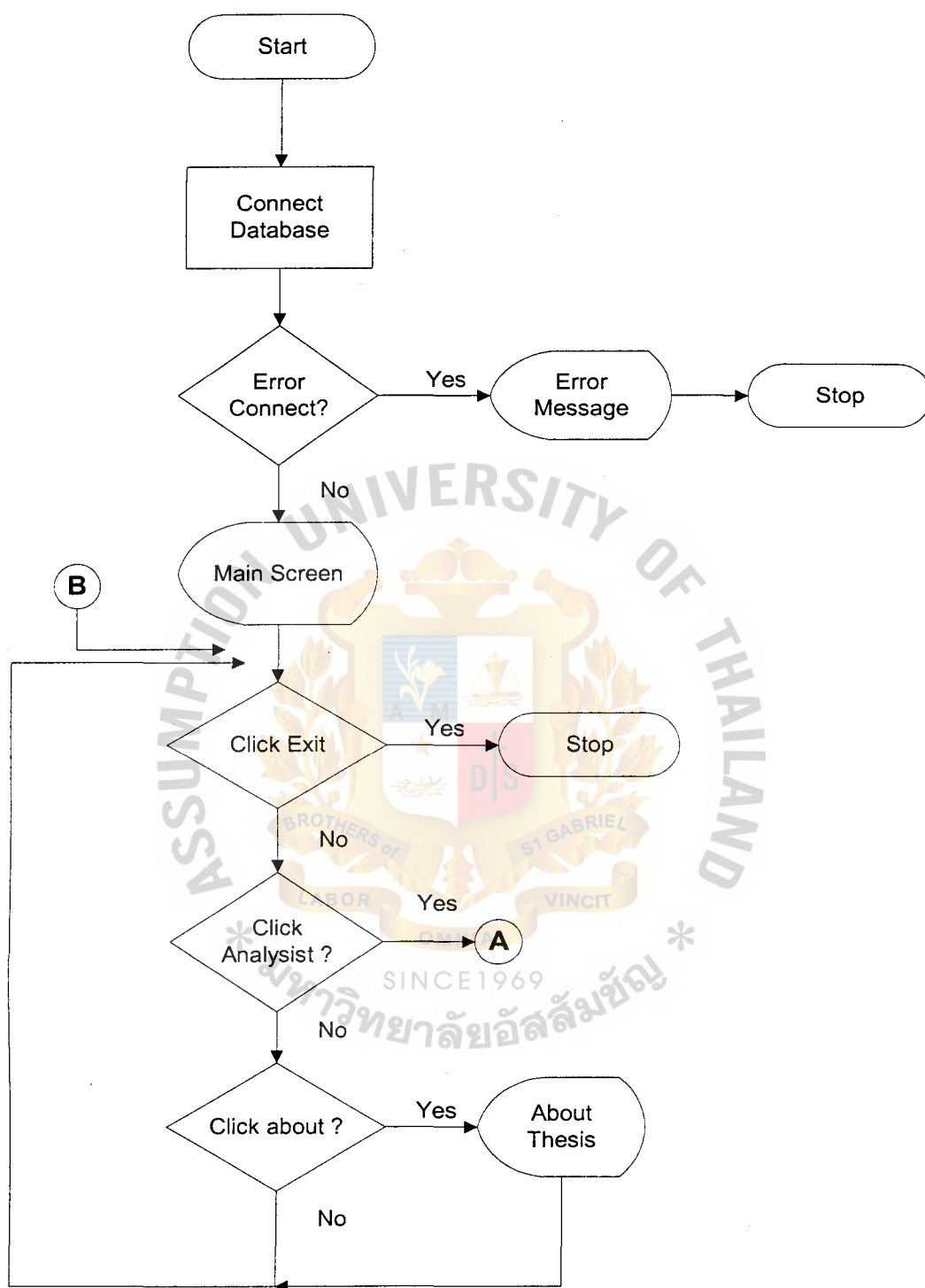


Figure 3-14 : Input Process Flow 1

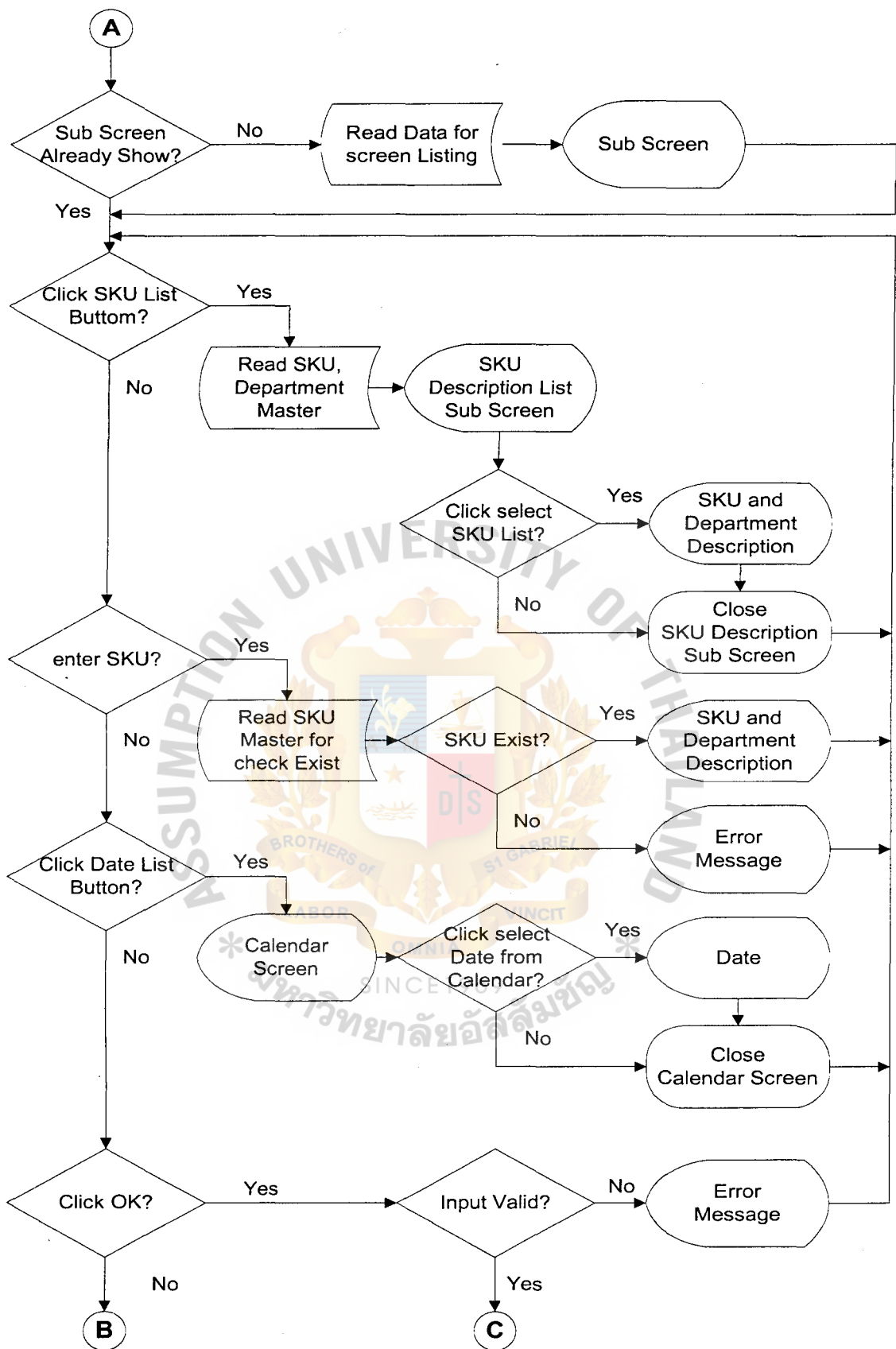


Figure 3-15 : Input Process Flow 2

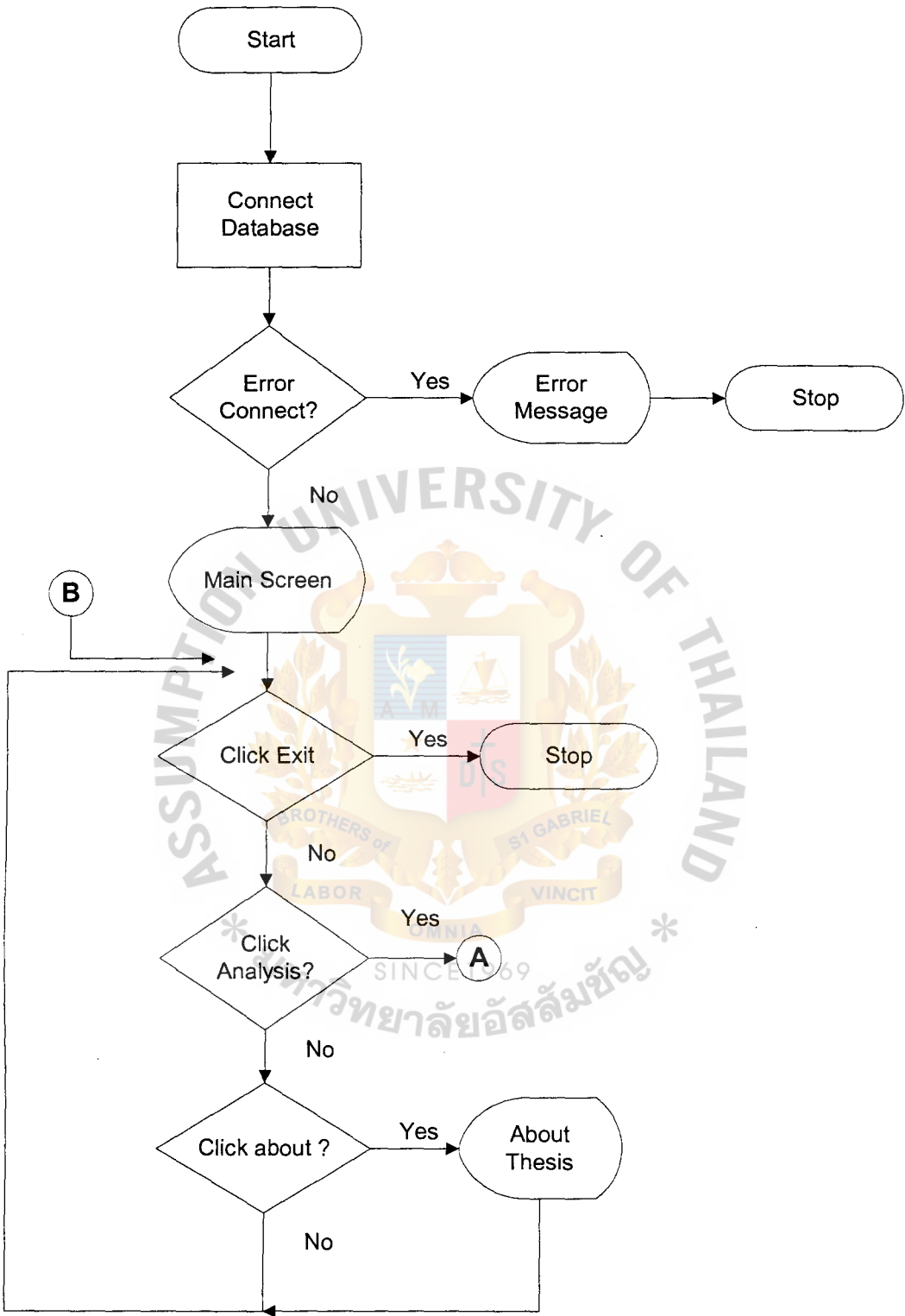


Figure 3-16 : Input Process Flow 3

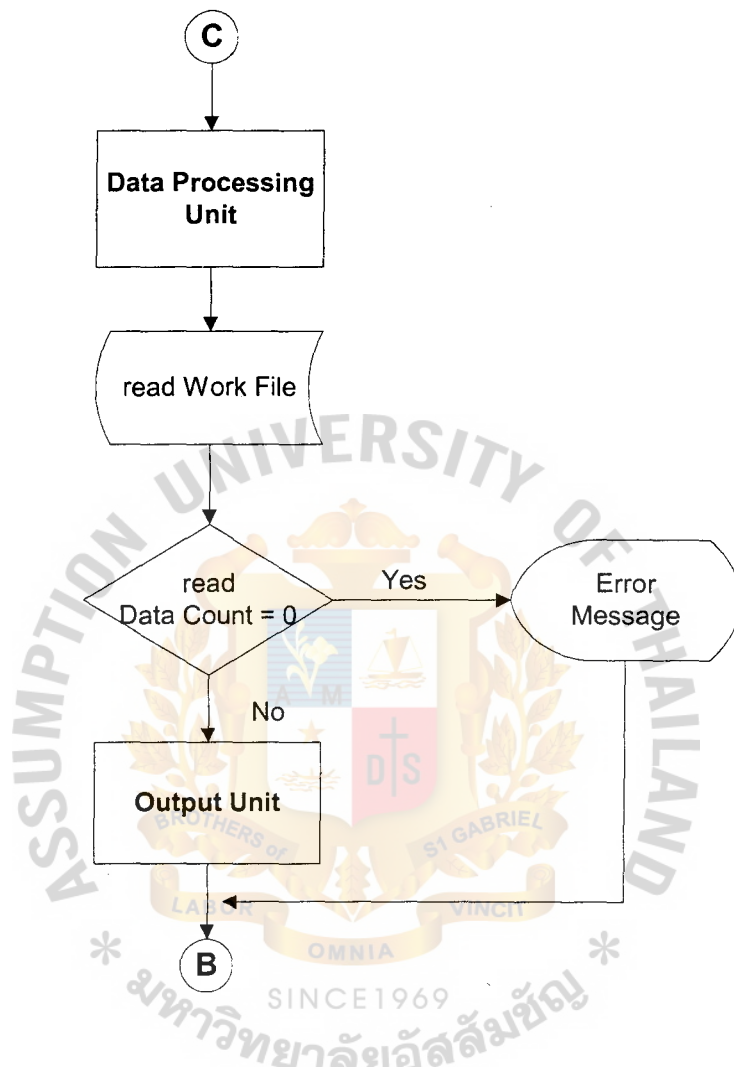


Figure 3-17 : Input Process Flow 4

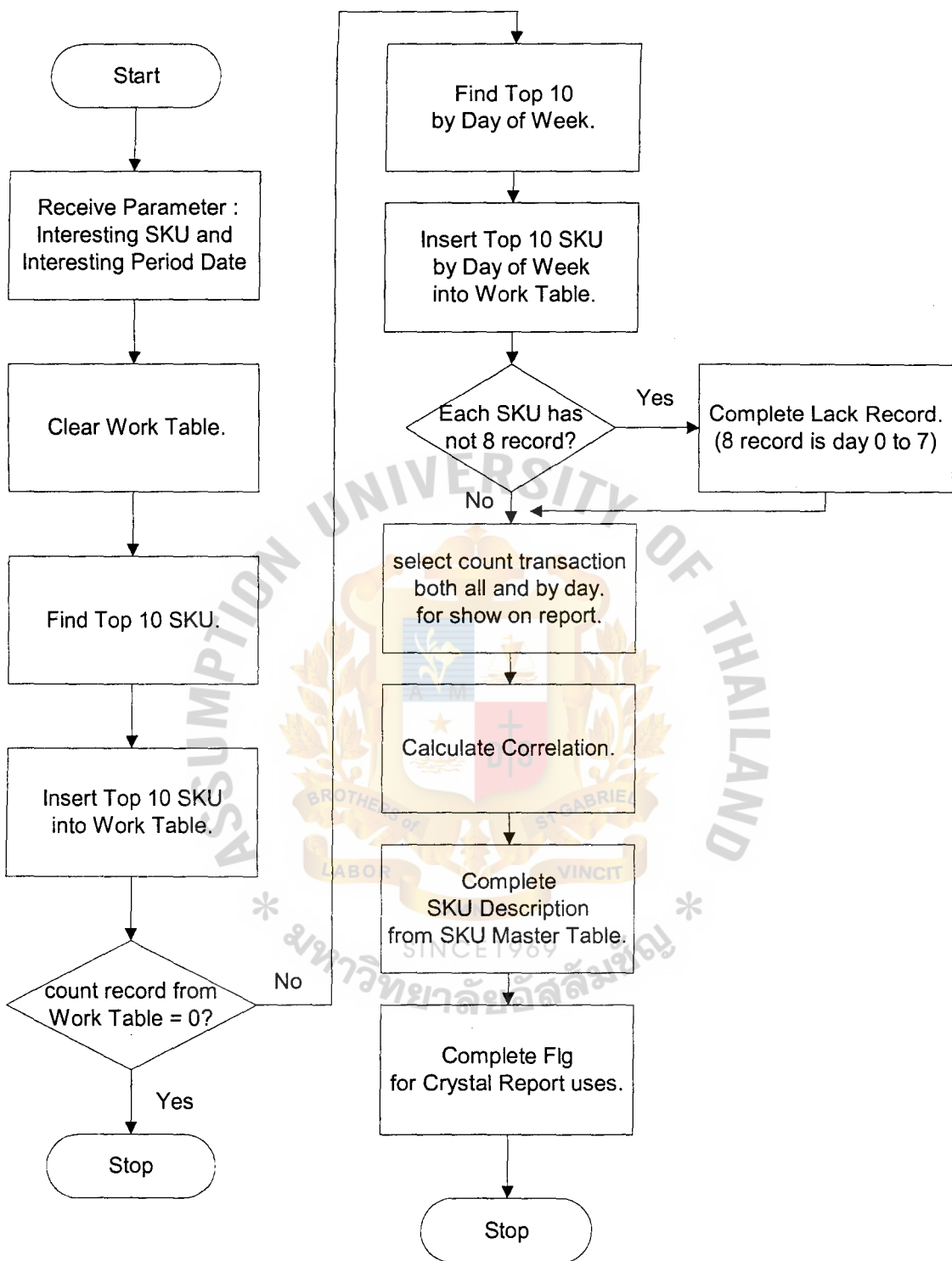


Figure 3-18 : Mining Process Flow

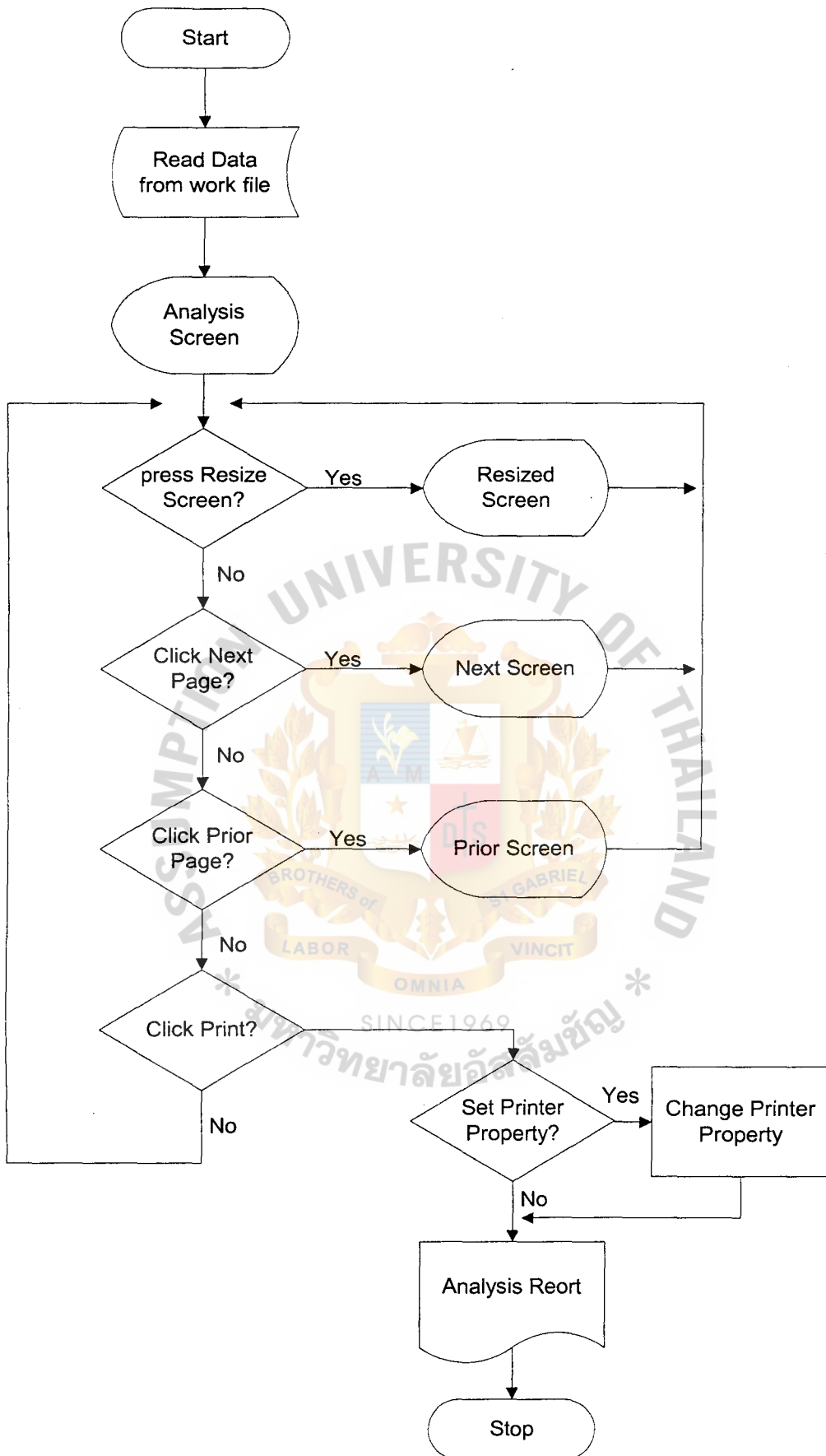


Figure 3-19 : Output Process Flow

CHAPTER 4

SYSTEM IMPLEMENTATION

4.1 Overview of System Implementation

To implement the system, SQL Server version 7.0 is needed to be installed including ODBC, RCBA programs the steps in are as following.

- **Defining Problem**, that is to determine the objective and user requirement and study the exist system then plan and design to implement.
- **Collecting and Enhancing data**, that is to prepare the enhancement data and transform data to database in new system.
- **Modeling strategies**, this is to explain the methodology or algorithm that how to get the analysis result.
- **Analyzing result**, after getting the result by statistic, then the step will generate the purchase pattern and give examples of the some results that should be reasonably acceptable including interpret and evaluate the analysis result.

4.2 Resource Utilization

1. RCBA Application
2. SQL Server version 7.0
3. Microsoft Windows 9X or latest with ODBC
4. Laser Printer

4.3 Data Collection and Enhancement

Data collection involves four distinct steps :

4.3.1 Define Data Sources

Defining the data sources should be a prominent part of details in the problem definition. The data should comes from Supermarket store and can

categorize 2 types of data as **Data Master** and **Daily Transaction data**.

4.3.2 Join and De-normalize data

After define data sources, will prepare or export data to 4 flat files.

- **SKU master** flat file will consists of SKU code, 20 character of SKU description, regular price, promotion price (future use) and department code.
- **Department master** flat file will consists of Department code and 20 character of department description
- **Transaction Header** flat file will consists of POS Number, Transaction Number, Store Number, Transaction Type and Transaction Total amount.
- **Transaction Detail** flat file will consists of POS Number, Transaction Number, Sold sequence of daily sale, Sold sequence in the Transaction, Sold quantity, SKU code, Sale amount, Tax amount and department code.

4.3.3 Enrich data

For SKU flat file and Department flat file are complete but Transaction header and Transaction detail are not specific enough, to enrich data might add one more field to specific data date and new structure will follows :

- **Transaction Header enrichment** flat file will include Operation date, Store Number, POS Number, Transaction Number, Transaction Type and Transaction amount.
- **Transaction Detail enrichment** flat file will include Operation date, POS Number, Transaction Number, Sold sequence of daily sale, Sold sequence in the Transaction, Sold quantity, SKU code, Sale amount, Tax amount and department code.

4.3.4 Transform data

To transform flat files data to SQL Server version 7.0 database by using import wizard, the data transformation enables to more readily extract the valuable information from data.

4.4 Modeling Strategies

Model can be classified to Affinity Modeling which events are likely to occur in conjunction with one another. The Model of Product Affinity Analysis will display result to identify products that sell together under certain condition.

4.4.1 Methodology

In order to perform analysis model, it is necessary first to have a list of transaction purchased in each one for this issue after data collection and enhancement step, all transaction will be ready for analysis. For simplicity, we will look at the example of convenience store customers, each of whom bought only a few items:

Transaction 1 : Beer, Milk, News Paper, Potato Chips

Transaction 2 : News Paper, Potato Chips, Coke, Milk

Transaction 3 : Marlboro, Beer, Potato Chips

Transaction 4: Coke, Beer, Marlboro

Transaction 5: Potato Chips, Coke, News Paper, Beer

Each customer purchased a different basket of items, and at first glance, there is no obvious relationship between any of the items purchased and any other items. The next step of an analysis, however, is to cross tabulate the data into a table, allowing you to see how often products occurred together. For these five supermarket store purchases, the table looks like this:

Item	Beer	Milk	News Paper	Potato Chips	Coke	Marlboro
Beer		1	2	3	2	2
Milk	1		2	2	1	0
News Paper	2	2		3	2	0
Potato Chips	3	2	3		2	0
Coke	2	1	2	2		1
Marlboro	2	0	0	0	1	

Table 4-1 : Cross tabulate data table

The central diagonal of the chart shows how often each item was purchased with itself. Though this is significant for figuring some reliability statistics, it does not show how items sell together, and can be ignored for now. Look at the first row out of the people who bought beer, one bought milk, two bought news paper, three bought potato chips, two bought Coke and two bought Marlboro. This hints the fact that beer and potato chips may sell well together, and should be placed side-by-side in the supermarket store. Looking over the rest of the table, there is nowhere else that an item sold together with another item that frequently – this is probably an actual cross-selling opportunity. Compare this to the fifth row – of people who bought milk, two bought beer, one bought milk, two bought news paper, two bought potato chips and one bought Marlboro. It seems coke sells well with everything in the store – there is probably not a good cross-selling opportunity with coke. This makes sense for a supermarket store – people often come to a convenience store for the express purpose of buying coke, and will buy it regardless of anything else they're looking for.

To perform the methodology and generate output any number of associated rules, but only the best rule is used for developing market campaigns. There are two measures of the quality of a rule, which this software will perform Correlation. Considering rules will be as following:

If **condition** then **result**.

And

If **condition** and **day of week** then **result**

Correlation is the ratio of the number of transactions with all the number of transactions with all the items in the rule to the number of transactions with just the items in the condition.

4.4.2 RCBA Software Algorithm

The explanation of RCBA Software algorithm will start since interesting SKU entry, mining process and also algorithm to generate the analysis result.

- 1) Accept interesting SKU.
- 2) Read interesting SKU from Master File.
 - 2.1) If exist **then** show SKU Code, SKU Description, Department Code, Department Description **Else** display Error Message and go to 1
- 3) Accept From Date.
- 4) Accept To Date.
- 5) Accept Click OK button.
- 6) Check Valid Data.
 - 6.1) If invalid **then** display Error Message **and** go to 1.
- 7) Find TOP 10 SKU in Sale Transaction, which occurs with interesting SKU.
 - 7.1) Select sale transaction detail from xitem table that
 - Interesting SKU exist.
 - Between Period Date (between From Date – To Date).
 - Is Sale Transaction (Xtran.dTrans = 35).
 - If any Transaction has **uplicated** same SKU Item, **Count only 1**.
 - Counting SKU is not including SKU that code is space.

- **Group by SKU.**

7.2) Summary Count record and Sort by Descending.

7.3) Cut First 10 Top and insert into Work File with define nDay = 0

(Summary Record).

8) Read Work File.

8.1) **If** Count Record = 0 **then** display Error Message and go to 1.

9) Find Selected TOP 10 SKU by Day.

9.1) Select xitem that

- Selected TOP 10 SKU.
- Between Period Date (between From Date – To Date).
- Is Sale Transaction (Xtran.dTrans = 35).
- If 1 Transaction has duplicated same SKU Item, Count only 1.
- Counting SKU is not including SKU that code is space.
- **Group by SKU and Day of Week (from OprDate).**

9.2) Insert into Work File.

10) Complete lack Record, when find Selected TOP 10 SKU by Day, someday has no interesting data and must be complete lack record with count = 0.

for Day = 1 to 7

Select Data from Work File where nDay = Day

If count SKU \leq 11 then

Compare record with record that Day = 0 (Summary Record)

to find missing record then Insert into Work File with set count = 0.

Next Day

11) Total Transaction calculation.

11.1) Count Total Transaction from xtran Table both by day and all summary.

11.2) Update Work File.

For Day = 0 to 7

If **Total transaction Count** ≤ 0 then

Set totCount = Total transaction Count

Next Day

12) Correlation calculation

12.1) Select interesting SKU each Day from Work File.

12.2) Update Work File.

For Day = 0 to 7

If **interesting SKU Transaction Count** ≤ 0 then

Correlation = record Transaction Count * 100 / Accepted SKU

Next Day

13) Select SKU Description from SKU Master Table and update to Work File.

14) Set Flag to interesting SKU record for use in Crystal Report.

14.1) Read Work file where SKU = interesting SKU

14.2) Update iFlg = '*'

15) Call Crystal Report to Show Report.

16) Go to 1.

17)

4.5 Analyzing Results

To understand clearly the result, there are rules in the form of

“if condition then result” and “if condition and day of week then result”

This is an example of the best kind of analysis result for SKU 3304024 MAMA มานะปิ่นทอง
มยุรี with data test from Supermarket Store.

1 st Step: We need to input **SKU code** as Figure 4-1

2 nd Step : After SKU code entry and SKU information is display, press **OK button**.

3 rd Step : The analysis result will display on Screen as 3 dimension Graph, 2
dimension Graph and Table of result

Last Step : Press **Print** button to print out the analysis result

Figure 4-1 : RCBA Screen

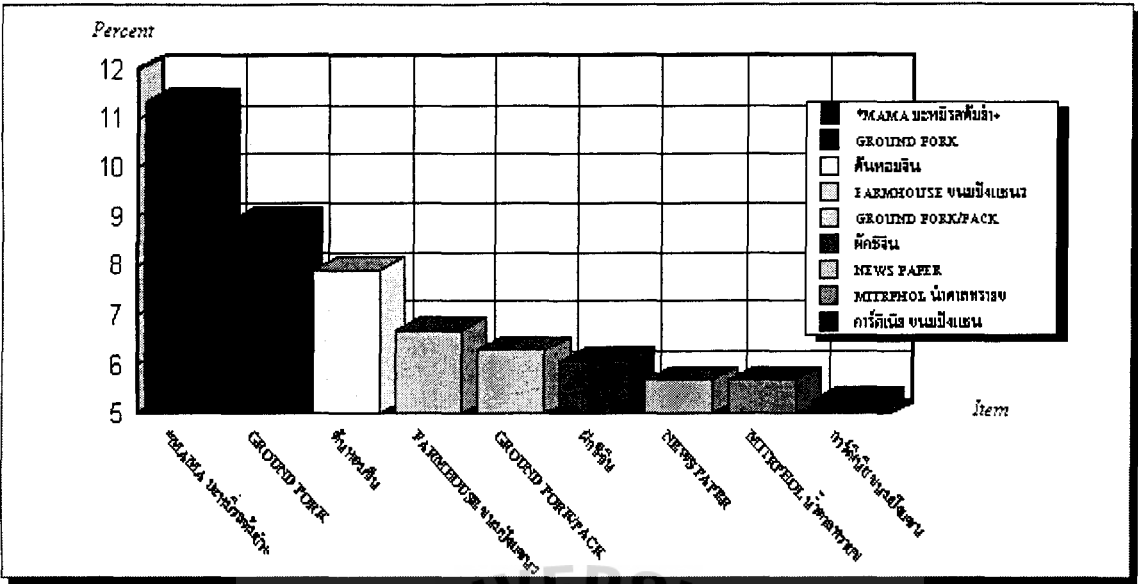


Figure 4-2 : 2 Dimension Graph

Rule	Correlation
If MAMA บะหมี่รสต้มยำ then MAMA บะหมี่รสต้มยำ	11.34%
If MAMA บะหมี่รสต้มยำ then GROUND PORK	8.91%
If MAMA บะหมี่รสต้มยำ then ต้นหอมจีน	7.89%
If MAMA บะหมี่รสต้มยำ then FARMHOUSE ขนมปังแชนว	6.68%
If MAMA บะหมี่รสต้มยำ then GROUND PORK/PACK	6.28%
If MAMA บะหมี่รสต้มยำ then ผักชีจีน	6.07%
If MAMA บะหมี่รสต้มยำ then NEWSPAPER	5.67%
If MAMA บะหมี่รสต้มยำ then MITRPHOL น้ำตาลทรายข	5.67%
If MAMA บะหมี่รสต้มยำ then การ์ดเนียขนมปังแชนว	5.29%

Table 4-2 : Purchase Pattern

For Total Transaction analysis can be summarized as Table 4-1 the table will generate the purchase pattern by select Top 10 Items, which always sell together with **MAMA บะหมี่รสต้มยำ**.

The first item is **MAMA บะหมี่รสต้มยำ** that has **Correlation** 11.34 %. Why customers always buy **MAMA บะหมี่รสต้มยำ** with **MAMA บะหมี่รสต้มยำ**, If we check the

both items, it is same brand and put on the shelf side by side, it is possible that the position is easy to get at the same time and customers may have loyalty with the brand. In this case it will be useful for Supplier's MAMA that they can know which items should be promoted together.

The second item is GROUND PORK, that has **Correlation 8.91%**, **MAMA** บะหมี่รสหมูสับ is an instant noodle mince pork flavor so the customers buy **GROUND PORK** to cook together with **MAMA** บะหมี่รสหมูสับ that is the reason why they buy together. This purchase pattern is good sound for Retailer. They should put both items in the same area or together then customer will can grasp easier, The Suppliers may have promotion or coupon if customers buy **MAMA** บะหมี่รสหมูสับ then they can get discount if they buy also **GROUND PORK**. And **GROUND PORK** can be perishable, so the retailer may use time sale promotion during closely store close time period.

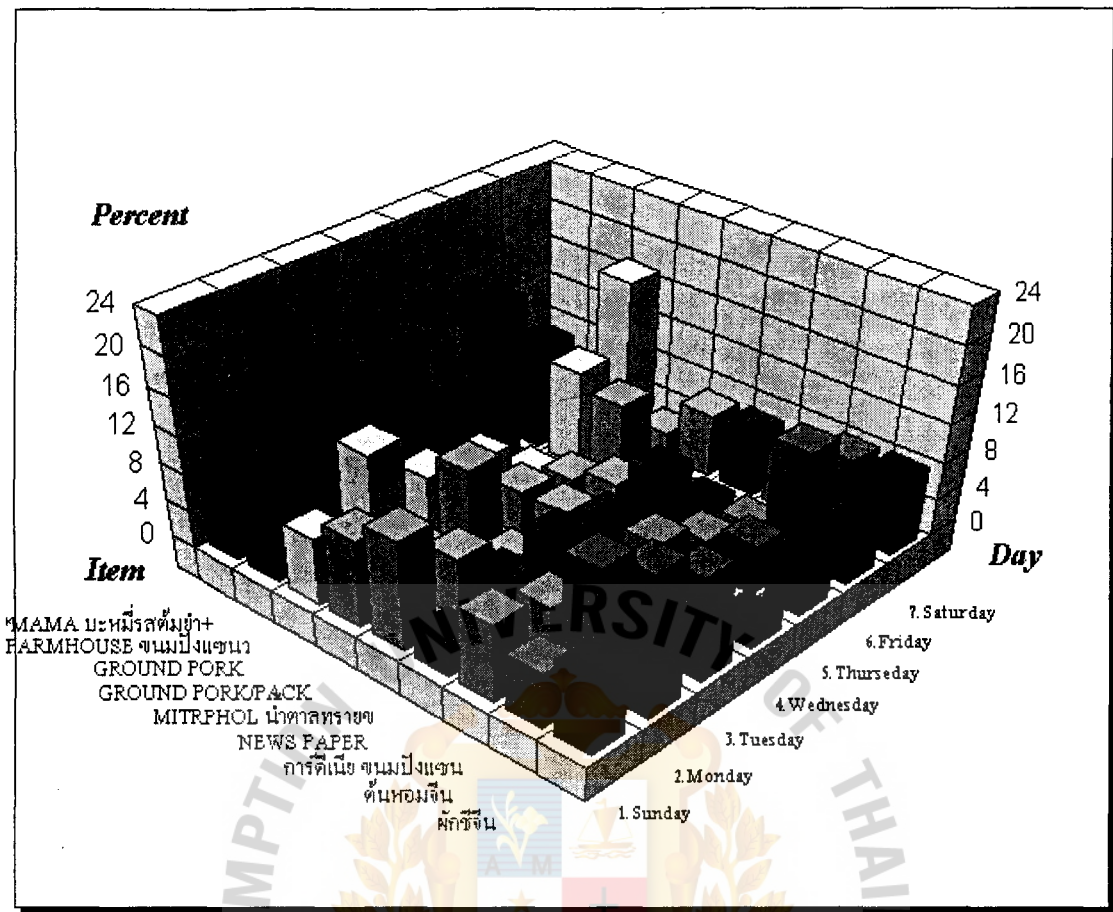


Figure 4-3 : 3 Dimension Graph

The Figure 4-3 shows the result in the form of three dimension graph by X axis, that represents the SKU name of top 10 SKU which sell together with interesting SKU, Y axis represents description of day of week and Z axis shows the percentage of Correlation.

The graph is shown that If **MAMA บะหมี่รสต้มยำ** and **Monday** then **MAMA บะหมี่รสต้มยำ** The Correlation of this pattern is 20.63%.

1.Sunday		Total Transaction Count : 42,334	
SKU	Description	Transaction Count	Correlation
3304024	*MAMA บะหมี่รสหมูดิบ	87	100.00 %
3621012	MITRPHOL น้ำตาลทรายขาว	10	11.49 %
1104036	GROUND PORK/PACK	8	9.20 %
3304025	*MAMA บะหมี่รสต้มยำ1+	8	9.20 %
1351058	ผักชีจีน	7	8.05 %
7801037	การดีเนียบ ขนหมูปิ้งแฉวน	7	8.05 %
1104035	GROUND PORK	6	6.90 %
1351043	ต้นหอมจีน	5	5.75 %
13	NEWS PAPER	3	3.45 %
7801001	FARMHOUSE ขนหมูปิ้งแฉวน	3	3.45 %

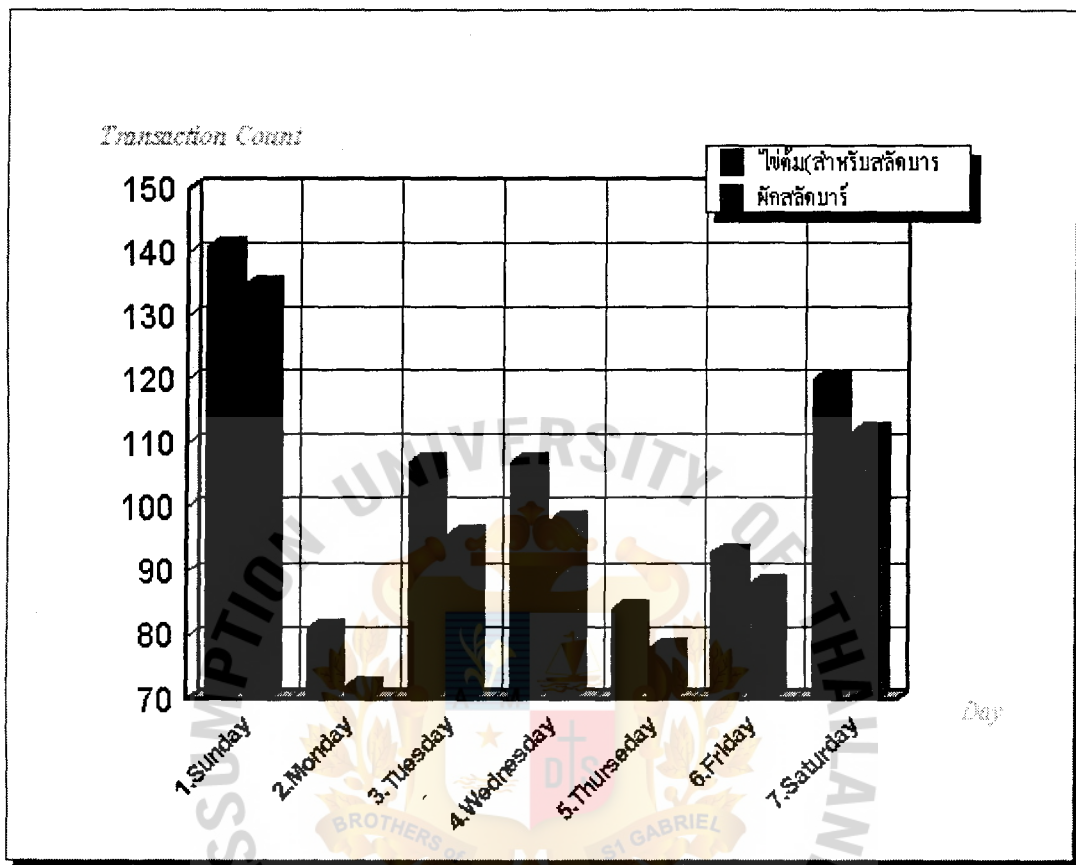
Figure 4-4 : RCBA Application Report by DAY (Sunday)

If retailers consider **MAMA บะหมี่รสหมู** only is good selling on Sunday as

Fig 4-4, they should check stock before the day and re-order on time to prevent short of stock.

The problem of this item is that this supermarket has more than one SKU code for this item some Sale count go to another code. It may effect with the analysis.

Comparison Transaction Count Graph



SKU: 1356000 ไข่ต้ม(สำหรับสลัดบาร์)
 Department: 1356 Fruit & Vegetable

Period From : 01/10/1999 To: 30/11/1999

Figure 4-5 : Comparison Transaction Count Graph

On Figure 4-5 shows Transaction Count of interesting SKU and first best selling SKU, the result shows that “ผักสลัดบาร์” always sell together with “ไข่ต้มสำหรับสลัดบาร์” And the both SKU have best selling on Sunday.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

After the thesis completion the first version of RCBA software for analysis customer behavior in supermarket retail business to generate the purchasing pattern and the analysis result helps the Retailer to make a better decision in their business.

The result of analyzing in my thesis was extracted the usable data from one supermarket in Bangkok. The data test was collected 2 month during September 1, 2000 and October 31, 2000, the result may not so accurate with any SKU, so it still need more transaction data from store for more accuracy and sometimes the result is not reasonable it still needs some expertise from Retailer to give explanation. However, this thesis summerizes the advantages and disadvantage of methodology.

5.1.1 Advantage of Methodology

Knowing which products sell together can be very useful to retail business. The most obvious effect is the increase in sales that a retail store can achieve by reorganizing its products so that products that sell together are found together. This facilitates impulse buying and helps ensure that customers who would buy a product do not forget to buy it on account of not having seen it. In addition, this has the side effect of improving customer satisfaction – once they've found one of the items they want, the customer does not have to look all over the store for something they want to buy then items that sell together should be found together.

Finally, It can be useful for operations purposes to know which products sell together in order to stock inventories. Running out of one item can affect sales of associated

items; perhaps the reorder point of a product should be based on the inventory levels of several products, rather than just one.

There are several advantages to this analysis over other types of data mining, first of all, it is undirected. It is not necessary to choose a product that you want to focus on in order to run an analysis tool. Instead, all products are considered, and the data mining software reveals which products are the most important to the analysis. In addition, the results of analysis are clear, understandable with associated rules that lend themselves to being immediately acted upon, and the individual calculations involved are simple.

5.1.2 Disadvantages of Methodology

Though being a useful and productive type of marketing data mining, analysis does have a few limitations. The first is the kind of data needed to do an effective analysis, it is necessary to have a large number of real transactions to get meaningful data, but the data's accuracy is compromised if all of the products do not occur with similar frequency. Thus, in supermarket store example, if newspaper is sold in almost every transaction, but glue only sells once or twice per month, putting both of them into the same analysis will probably generate results that look impressive without being statistically significant – acting on these results might not actually benefit profitability. With only one or two glue customers, the RCBA software will be able to very confidently state what sells well with glue but this may only be true for the one or two customers analyzed. Second, the analysis can sometimes present results that are actually due to the success of previous marketing campaigns. If the supermarket store has always been putting Coke can discount coupons on the Sprite can, the fact that Coke can and Sprite can sell well together may come as no surprise to them and it does not give any new information, just show that previously existing marketing

campaigns are already working. In fact, the previous campaign may even be overshadowing a real relationship perhaps people would normally prefer to buy Beer with Sprite can, but only buy the Coke because of the discount. In this case, the supermarket store is missing out on what could be a better promotion.

5.2 Future Work

Current thesis focuses mainly on the discovery of purchase pattern with the single item. This can be extended to the discovery of multiple items purchasing pattern and may have analysis for the item, which unlikely to occurrence, the result of analysis may use to change customer's behavior by new type promotion. If this RCBA software is developed in advance, it would improve performance to be faster in the future.



REFERENCE

- [1] Brian P. Woolf, Customer Specific Marketing, 1996. ISBN: 1-888051-02-7
- [2] Customer Relationship Management, <http://www.remedy.com>, April 22, 1999.
- [3] Database Marketing Standard for the Retail Industry,
<http://www.rtms.com/paper/dmsri.htm>
- [4] Michael Meltzer 1998: NCR Corporation, Customer Management Solutions.
- [5] Pilot Software, An Introduction to Data Mining,
<http://www.pilotsw.com/dmpaper/dmindex.htm>, March 25, 1999.
- [6] Pilot Software, Expanding Decision Support Horizon in the Retail Environment,
http://www.pilotsw.com/r_and_t/retail/contents.htm, March 25, 1999.
- [7] SAS Institute's Customer Relationship Management Solution,
<http://www.sas.com>, August 20, 1999.
- [8] Sequent UK Press Release : One-to-One Volume Marketing becomes Reality
with World's First,
http://www.sequent.com/solutions/dec_advantage/dssmb_wp1.htm, June 15,
1999.
- [9] Westphal, Christopher, Data mining Solution : methods and tools for solving
real-world problems, 1998. ISBN: 0-471-25384-7
- [10] Centrobe-Enterprise Customer Management, <http://www.centrobe.com>, October
26, 1999
- [11] Market Management Inventory, <http://www.phraimworks.com>, August 20, 1999
- [12] Data Mining Resource project, <http://www.ecs.solution>, August 16, 2000
- [13] Completing a solution for Market Basket Analysis, <http://www.anadfamilly.com>,
July 6, 1997
- [14] Data Distilleries-datamining main, <http://www.dai.ed.au.uk> , March 3, 2000

- [15] Data Mining and CRM (Kurt Thearling), <http://www3.shore.net/~kht/>
- [16] Thearling, Kurt (1995), From Data Mining to Database Marketing, white paper, Data Intelligence Group, Pilot Software
- [17] Data Mining from a Statistical Perspective John Maindonald...statistics and data mining is statistical analysis,
<http://www.maths.anu.edu.au/~johnm/dm/dmpaper.html>
- [18] Case Study : Data Mining the Northridge Earthquake,<http://www.nautitus-systems.com>
- [19] Case Study-Qantas Air-Data Mining,
http://ftp.ncr.com/product/case_studies/qantas/datamining.htm
- [20] Technology Market Basket Analysis technique White paper By Limitations Performing Market Basket Analysis Virtual Items,
<http://www.megaputer.com/html/mba.html>
- [21] Minevision N.V., Data Mining For The Retailing Industry,
<http://www.minevision.com>
- [22] MicroStrategy, For Hallmark Cards, the answer is MicroStrategy An E-Business Case Study, <http://www.microstrategy.com/ContactUs/Download/>
- [23] Monash Australia's International University : Data Mining Concepts 2000.07.24
2 Market Basket.. Data Mining Lecture material for COT5230,
<http://www.csse.monash.edu.au/courseware/cot5230/html/lectures.html>, July 24, 2000
- [24] Data Warehousing and Data Mining: Opportunities, Data Warehousing and Data Mining: Possibilities, Pitfalls,
<http://www.marketing.gsu.edu/mk8700eer/dmgoals7.htm>

[25] Michael J. A. Berry Gordon Linoff, Marketing Sales and Customer Support,
1997. ISBN: 0-471-17980-9

[26] White Paper library, Data Mining an Introduction by SPSS,

[http:// www.tradeSpeak.com](http://www.tradeSpeak.com)



APPENDIX A

Installation

Server Site

1. Import Transaction Data.
2. Run script “RCBA_Script.sql” for creates Work Table Structure.
3. Run script “RCBA_StoreProc.sql” for creates Store Procedure.

Client Site

1. Install SQL Server Client.
2. Install RCBA Application.
 - 2.1 Insert Install CD and run Setup.
 - 2.2 Select RCBA Application Path, Default is “C:\Program Files\RCBA”.
3. Set Printer.

Need at least one for show and print report. If none, Add Printer by goes to Start Menu and Select Printers. Click Add Printer to run Printer Wizard.
4. Set ODBC.
 - 4.1 Go to Start Menu and Select Settings, Control Panel and ODBC.
 - 4.2 Click Add and Input Description as FigureA-1

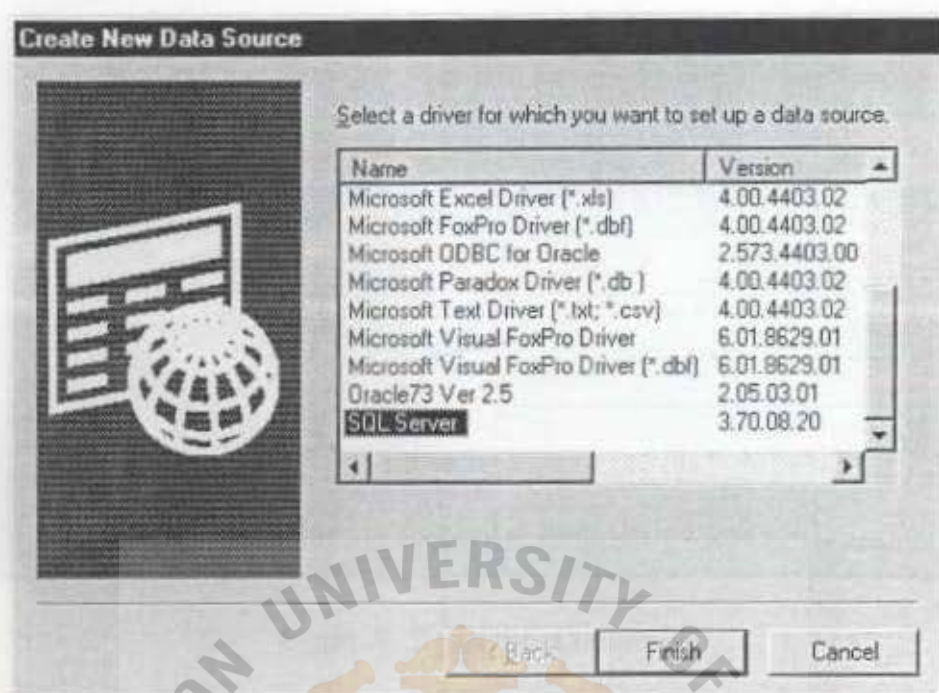


Figure A-1 : Select SQL Server Driver on Create New Data Source screen.

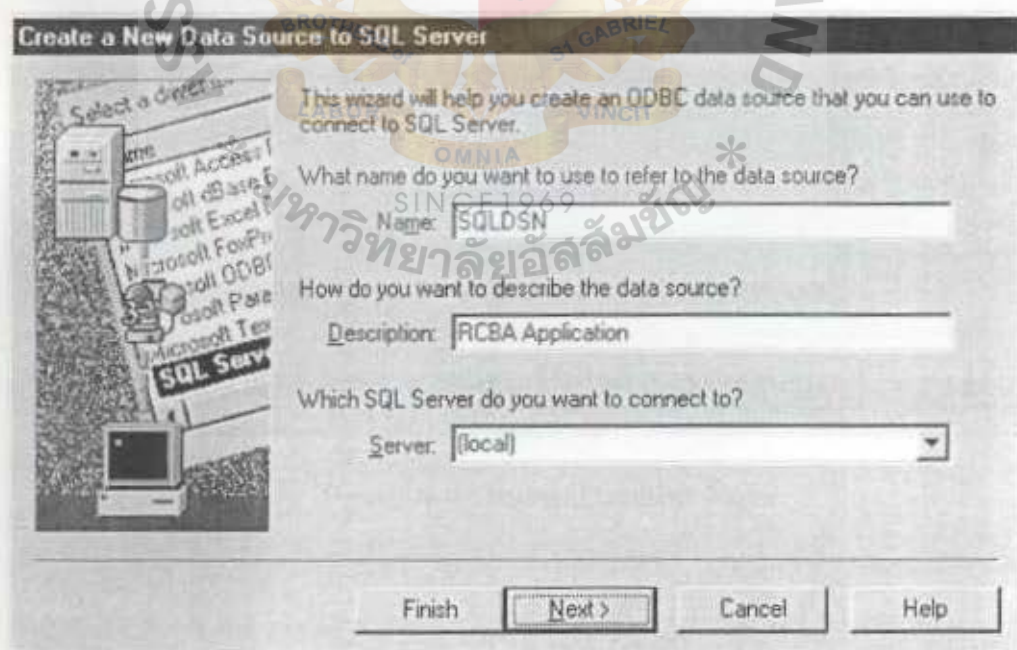


Figure A-2 : Input ODBC Name "SQLDSN" and Description as screen.

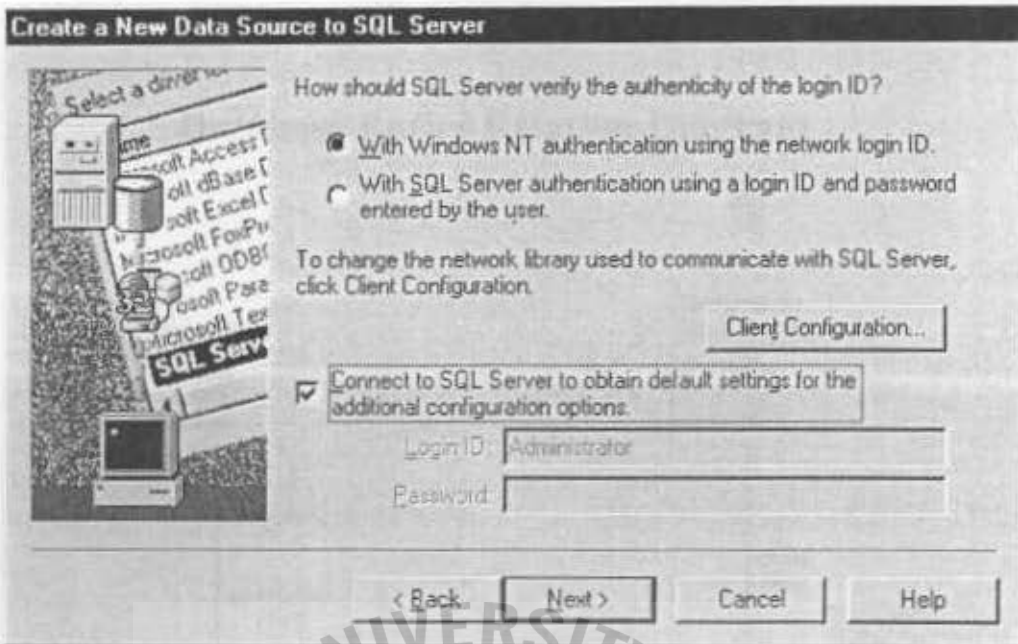


Figure A- 3 : ODBC Connection

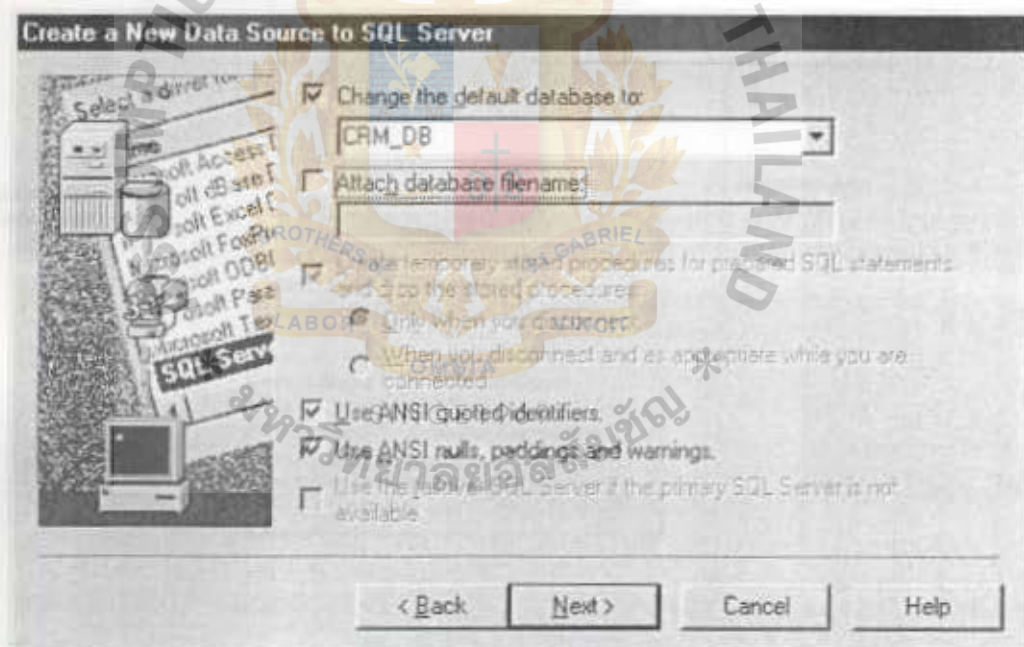
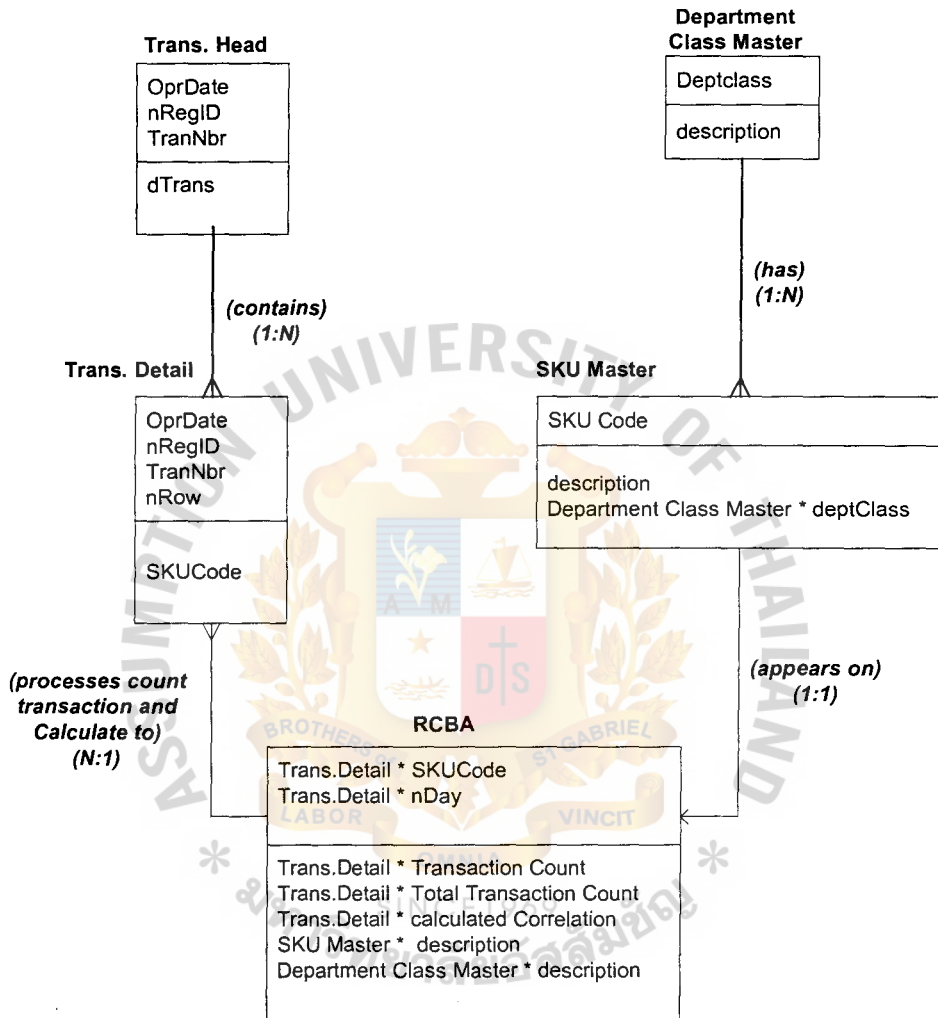


Figure A-4 : select Database Name.

APPENDIX B

Database Entity Relation Diagram



As figure, One Transaction Head contains many Transaction Detail matching by OprDate, nRegID and TranNbr. nRow of Trasaction Detail use to specify serial number of item on that Transaction Head.

Processing count many Transaction Detail and calculate correlation will makes RCBA data. RCBA's primary key uses SKU Code and nDay that from done grouping.

Finally, RCBA's SKU and Department Class Description comes from one-by-one with SKU Master and a part of Department Class Master where that SKUCode exists will give detail to RCBA.



APPENDIX C

Sample of Analysis Result and Comment from Retailer

There are several comments made by retailers who've seen the sample of Analysis result.

1. Comment from Miss. Wanna Wanapiroon, Budget Department Manager of The Mall Group CO., LTD.. she said that the analysis can let us know about unexpected information. Normally we have our promotion pattern by packing the same item and give discount such as COKE CAN 12.00 baht if the customer buy 6 cans, they have to pay only 62.00 bath. From the analysis the plan for packing different items could work but for Grocery item and Fresh food we need to take more time to consider.

2. Below is the comment from Miss. Jitima Donksumnerd, Assistant Brand Manager of Luxasia (Thailand) company LTD., responding to the result of my analysis.

"I think the results of the analysis provide useful information for retail business. However as for the perfumery counter of our store, I notice most of the customers usually look for products that have similar fragrance, for example, soap, Eau-de-cologne, shower cream or other cleansing products, etc.

Currently we have a gift set promotion package. This package includes products in the same brand which have similar fragrance. It will be useful if we have the information of what similar product to match. For example if the customer buy lady's perfume with flowery fragrance, what fragrance will a customer buy for man.

The perfumery counter in our department store has a limited space. So we need to consider displaying the products within the given space in such a way that attracts the customer most. "

Retail Customer Behavior Analysis

I am a graduate student, major on Master Information Technology at Assumption University. I am currently working on a thesis and would like to have your comment on the analysis of this thesis.

The source of this thesis is base on the Sales Transaction data from one of the Supermarket in Bangkok during Oct 1, 1999 – Nov 30, 1999

Please tick the checkbox on the sample of the result attached.

Agree Disagree

☐ ☐ **Efficient Forecasting**

- The result was shown that “SUPER LEO BEER” had a good sold on Sunday. The Retailer should order before the day that to prevent running out of Stock.
- The result was shown the Item which often to sold together with “SUPER LEO BEER” , it was “LEO BEER”.
The Retailer may need to re-order “LEO BEER” also.

☐ ☐ **Efficient Brand Management**

- The suggestion “SUPER LEO BEER” and “LEO BEER” are the same supplier. The result was shown that customer usually has the loyalty with this brand, it will be useful for supplier’s LEO that can know which items should be promote together.

☐ ☐ **Efficient Precision Marketing For Better**

- Retailer may use the analysis to improving marking by offering some discount to driven sales amount per customer.
- For second item with top selling together with “SUPER LEO” was “น้ำแข็งหลอด”, we may provide coupon for discount “น้ำแข็งหลอด” if the customer buy “SUPER LEO”. And if you notice in Supermarket “น้ำแข็งหลอด” always put beside Beverage Item, the result shows the reason in backward that we can use the analysis result to make decision for allocate store location.

COMMENT

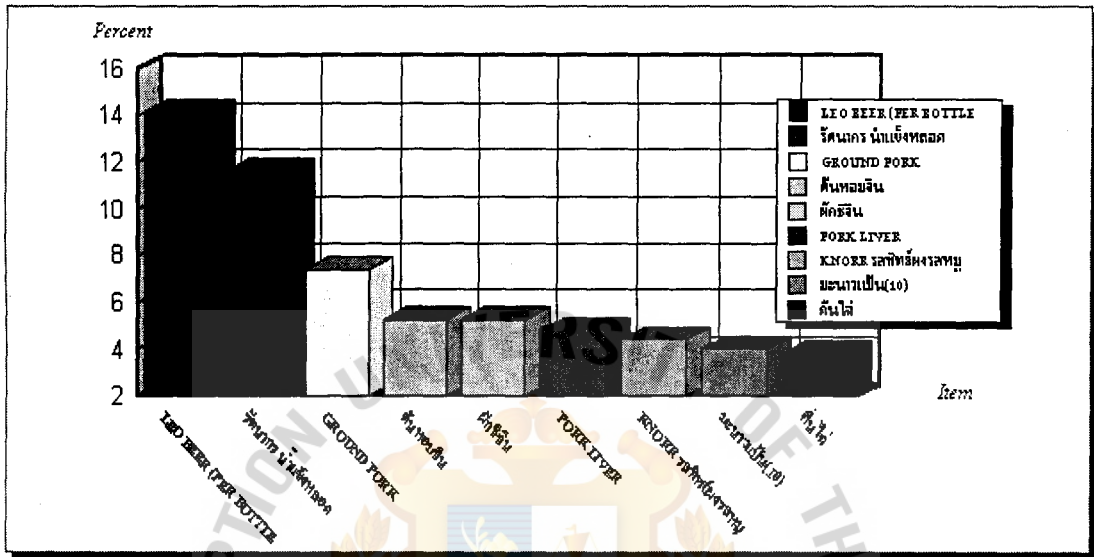
Name : _____ Position : _____

Retail Customer Behavior Analysis Report



The Use of Data Mining Technique :

Total Graph



SKU : 2443044 SUPER LEO BEER

Department : 2443 Wine & Spirit

Period From : 01/10/1999 To: 30/11/1999

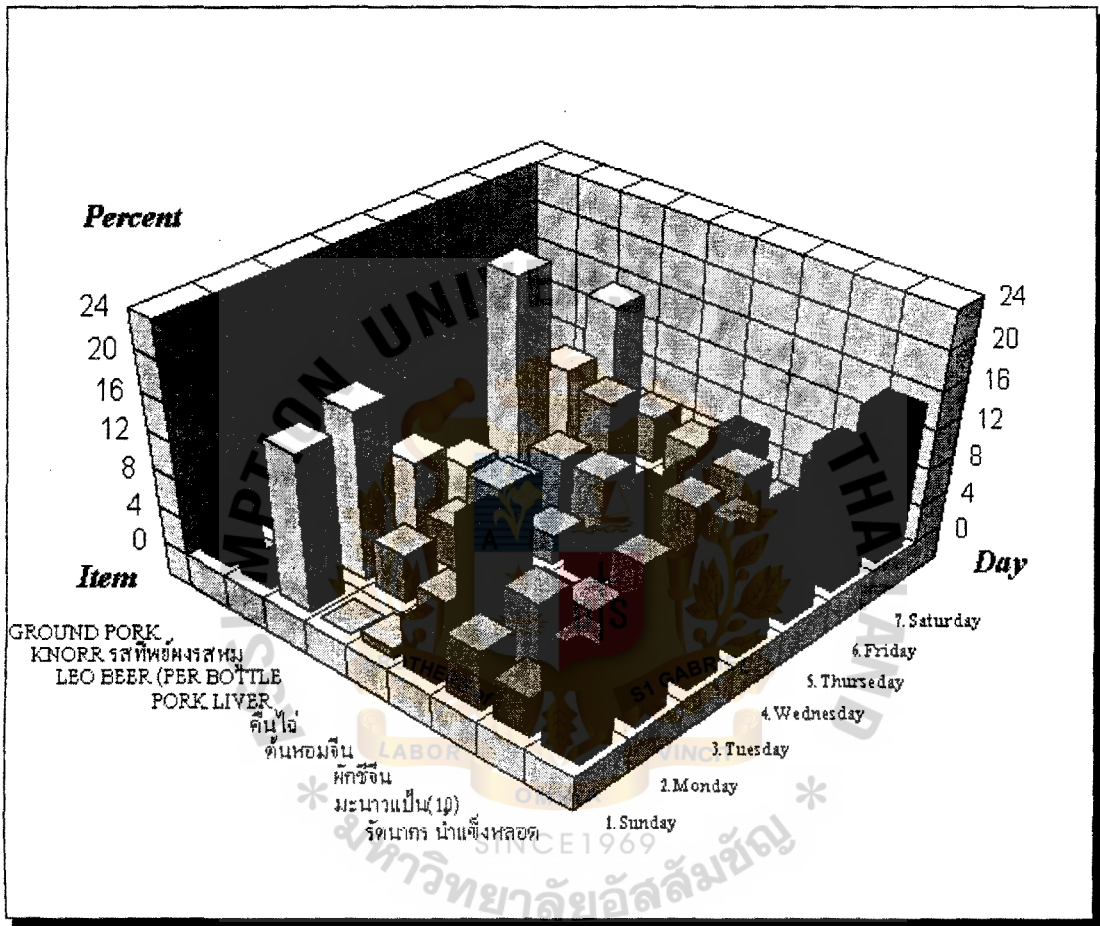
SKU	Description	Transaction Count	Correlation
2443044	SUPER LEO BEER	474	100.00 %
2443042	LEO BEER (PER BOTTLE)	67	14.14 %
7901119	รัตนกร น้ำแข็งหลอด	55	11.60 %
1104035	GROUND PORK	35	7.38 %
1351043	ต้นหอมจีน	25	5.27 %
1351058	ผักชีจีน	25	5.27 %
1104039	PORK LIVER	23	4.85 %
3617014	KNORR รสทิพย์ผงรสหมู	21	4.43 %
1351101	มะนาวแป้น(10)	19	4.01 %
1351038	ถั่วลิสง	18	3.80 %

Retail Customer Behavior Analysis Report



The Use of Data Mining Technique :

Total Graph by Day



SKU : 2443044 SUPER LEO BEER
Department : 2443 Wine & Spirit

Period From : 01/10/1999 To: 30/11/1999

Retail Customer Behavior Analysis Report



The Use of Data Mining Technique :

1.Sunday

Total Transaction Count : 42,334

SKU	Description	Transaction Count	Correlation
2443044	SUPER LEO BEER	98	100.00 %
2443042	LEO BEER (PER BOTTLE	16	16.33 %
7901119	รัตนากร น้ำแข็งหลอด	9	9.18 %
1351058	ผักชีจีน	6	6.12 %
1351043	ต้นหอมจีน	5	5.10 %
1104035	GROUND PORK	4	4.08 %
1351101	มะนาวแป้น(10)	4	4.08 %
1351038	คื่นไฉ่	1	1.02 %
3617014	KNORR รสทิพย์ผงรสหมู	1	1.02 %
1104039	PORK LIVER	0	0.00 %

2.Monday

Total Transaction Count : 36,148

SKU	Description	Transaction Count	Correlation
2443044	SUPER LEO BEER	40	100.00 %
2443042	LEO BEER (PER BOTTLE	7	17.50 %
1351043	ต้นหอมจีน	5	12.50 %
7901119	รัตนากร น้ำแข็งหลอด	4	10.00 %
1351058	ผักชีจีน	3	7.50 %
1104035	GROUND PORK	2	5.00 %
1104039	PORK LIVER	2	5.00 %
1351101	มะนาวแป้น(10)	2	5.00 %
1351038	คื่นไฉ่	1	2.50 %
3617014	KNORR รสทิพย์ผงรสหมู	1	2.50 %

Retail Customer Behavior Analysis Report



The Use of Data Mining Technique :

3. Tuesday

Total Transaction Count : 37,405

SKU	Description	Transaction Count	Correlation
2443044	SUPER LEO BEER	63	100.00 %
1351038	กิ้นไผ่	7	11.11 %
7901119	รัตนกร น้ำแข็งหลอด	7	11.11 %
1104035	GROUND PORK	6	9.52 %
2443042	LEO BEER (PER BOTTLE	6	9.52 %
1104039	PORK LIVER	3	4.76 %
1351058	ผักชีจีน	2	3.17 %
3617014	KNORR รสทิพย์ผงรสหมู	2	3.17 %
1351043	ต้นหอมจีน	1	1.59 %
1351101	มะนาวแป้น(10)	0	0.00 %

4. Wednesday

Total Transaction Count : 38,288

SKU	Description	Transaction Count	Correlation
2443044	SUPER LEO BEER	65	100.00 %
7901119	รัตนกร น้ำแข็งหลอด	7	10.77 %
1104035	GROUND PORK	6	9.23 %
3617014	KNORR รสทิพย์ผงรสหมู	6	9.23 %
1104039	PORK LIVER	4	6.15 %
2443042	LEO BEER (PER BOTTLE	4	6.15 %
1351058	ผักชีจีน	3	4.62 %
1351038	กิ้นไผ่	2	3.08 %
1351043	ต้นหอมจีน	2	3.08 %
1351101	มะนาวแป้น(10)	1	1.54 %

Retail Customer Behavior Analysis Report



The Use of Data Mining Technique :

5.Thursday

Total Transaction Count : 37,840

SKU	Description	Transaction Count	Correlation
2443044	SUPER LEO BEER	65	100.00 %
2443042	LEO BEER (PER BOTTLE	15	23.08 %
1104035	GROUND PORK	7	10.77 %
7901119	รัตนกร น้ำแข็งหลอด	7	10.77 %
1351058	ผักชีจีน	5	7.69 %
1351101	มะนาวแป้น(10)	5	7.69 %
1104039	PORK LIVER	4	6.15 %
1351038	กิ้นไฉ่	4	6.15 %
1351043	ต้นหอมจีน	4	6.15 %
3617014	KNORR วัตถุดิบผงบรณ	3	4.62 %

6.Friday

Total Transaction Count : 37,767

SKU	Description	Transaction Count	Correlation
2443044	SUPER LEO BEER	64	100.00 %
7901119	รัตนกร น้ำแข็งหลอด	9	14.06 %
2443042	LEO BEER (PER BOTTLE	7	10.94 %
1104035	GROUND PORK	6	9.38 %
1104039	PORK LIVER	6	9.38 %
1351058	ผักชีจีน	5	7.81 %
3617014	KNORR วัตถุดิบผงบรณ	5	7.81 %
1351043	ต้นหอมจีน	3	4.69 %
1351101	มะนาวแป้น(10)	1	1.56 %
1351038	กิ้นไฉ่	0	0.00 %

Retail Customer Behavior Analysis Report



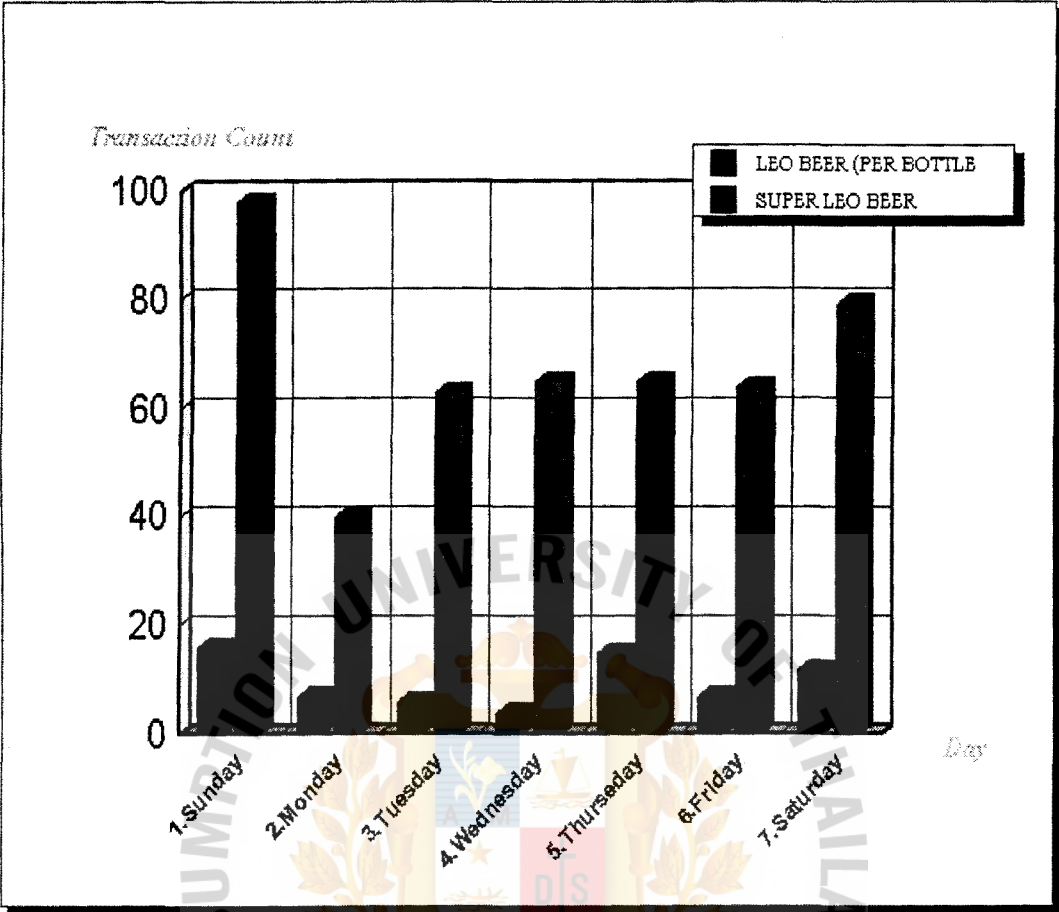
The Use of Data Mining Technique :

7.Saturday

Total Transaction Count : 41,685

SKU	Description	Transaction Count	Correlation
2443044	SUPER LEO BEER	79	100.00 %
2443042	LEO BEER (PER BOTTLE	12	15.19 %
7901119	รัตนกร น้ำแข็งหลอด	12	15.19 %
1351101	มะนาวแป้น (10)	6	7.59 %
1351043	ต้นหอมจีน	5	6.33 %
1104035	GROUND PORK	4	5.06 %
1104039	PORK LIVER	4	5.06 %
1351038	กิ้นไฉ่	3	3.80 %
3617014	KNORR รสทิพย์ผงรสหมู	3	3.80 %
1351058	ผักชีจีน	1	1.27 %





SKU : 2443044 SUPER LEO BEER
Department : 2443 Wine & Spirit

Period From : 01/10/1999 To: 30/11/1999

THE ASSUMPTION UNIVERSITY LIBRARY

