AN ANALYTICAL CRM SYSTEM FOR VAT REFUND FOR TOURISTS

SOMJAI SAMOEMUAN

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (COMPUTER SCIENCE) FACULTY OF GRADUATE STUDIES MAHIDOL UNIVERSITY 2009

Research Project
entitled
AN ANALYTICAL CRM SYSTEM FOR VAT REFUND FOR TOURISTS

Somjai Samoemuan
.....
Miss Somjai Samoemuan
Candidate

Songsri Tangsripairoj
.....
Lect. Songsri Tangsripairoj,
Ph.D.
Major-advisor

Charnyote Pluempitiwiriyawej
.....
Asst.Prof. Charnyote Pluempitiwiriyawej,
Ph.D.
Co-advisor

B. Mahaisavariya
.....
Prof. Banchong Mahaisavariya,
M.D.
Dean
Faculty of Graduate Studies
Mahidol University

Supachai Tangwongsan
.....
Assoc. Prof. Supachai Tangwongsan,
Ph.D.
Program Director
Master of Science Programme
in Computer Science
Faculty of Information and
Communication Technology
Mahidol University

Research Project
entitled

AN ANALYTICAL CRM SYSTEM FOR VAT REFUND FOR TOURISTS

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Computer Science)
on
October 2, 2009

Somjai Samoemuan
Miss Somjai Samoemuan
Candidate

P. Meesad
Asst.Prof. Phayung Meesad,
Ph.D.
Chair

Songsri Tangsripairoj
Lect. Songsri Tangsripairoj,
Ph.D.
Member

Charnyote Pluempitiwiriyawej
Asst.Prof. Charnyote Pluempitiwiriyawej,
Ph.D.
Member

B. Mahaisavariya
Prof. Banchong Mahaisavariya,
M.D.
Dean
Faculty of Graduate Studies
Mahidol University

Jarernsri L. Mitrpanont
Assoc. Prof. Jarernsri L. Mitrpanont,
Ph.D.
Acting Dean
Faculty of Information and
Communication Technology
Mahidol University

ACKNOWLEDGEMENTS

The success of this research project is dedicated to the extensive support and assistance from my major advisor, Dr. Songsri Tangsripairoj, for her valuable feedback and guidance, encouragement, patience and all the academic and personal advice.

I am also grateful to Asst. Prof. Charnyote Pluempitiwiriyawej and Asst. Prof. Phayung Meesad, who sit on my research project committee and provide me their insightful comments, technical review and suggestions on this research project.

I would like to thank all the lecturers of the Faculty of Information Communication and Technology. I thank my best friends in the Computer Science Program for their kind support and ceaseless encouragement.

I am deeply indebted to the Revenue Department of Thailand for providing me an opportunity to study, understanding and encouragement throughout the course of my study.

Finally, I am grateful to my family for their financial support, love, encouragement and tender care. This research project is sincerely dedicated to my father, my mother, my sister and all the teachers who had taught me well throughout my childhood.

Somjai Samoemuan

AN ANALYTICAL CRM SYSTEM FOR VAT REFUND FOR TOURISTS

SOMJAI SAMOEMUAN 4737983 ITCS/M

M.Sc. (COMPUTER SCIENCE)

RESEARCH PROJECT ADVISORY COMMITIGE: SONGSRI TANGSRIPAIROJ, Ph.D., CHARNYOTE PLUEMPITIWIRIYAWEJ, Ph.D.

ABSTRACT

VAT refund for Tourists (VRT), supervised by the Revenue Department, is a tax collection policy that supports Thai government policy to promote tourism. There are three main problems of the existing VRT system. First, making a manual report takes a long time. Second, these is only a small number of entrepreneurs and shops in the VRT system. Third, the officers in the VAT refund for tourists team (KOR-TOR) are lacking an efficient tool for analyzing goods purchase behavior in order to find new groups of VAT taxpayers that should join the VRT system.

In this research project an Analytical Customer Relationship Management System for VAT Refund for Tourists (ACRM-VRT) has been designed and developed by applying the notion of customer relationship management (CRM), data warehousing, OLAP, and data mining. The data mart is built for collecting the data for VAT refund for Tourists. The OLAP tool is developed to support the KOR-TOR work, especially for reporting and generating executive summaries. The Data mining tool, particularly an association rule mining tool, is created for studying tourist's goods purchase behavior to search for new groups of VAT taxpayers who sell the products that meet tourists' requirements.

The ACRM-VRT system is evaluated in the aspect of user satisfaction by the KOR-TOR officers. The overall results show that the user satisfactory level is good. The function that has the highest satisfactory level is the data mining tool. This is because the obtained knowledge can be used for specifying the strategies to search new groups of VRT taxpayers and facilitate the VRT system to recruit them into the system. Moreover, the obtained knowledge can be used for advertising to the VRT taxpayers in order to make tourist marketing plans. The VAT taxpayers who did not yet apply to the VRT system might be interested and volunteer to join the VRT system.

KEY WORDS: CUSTOMER RELATIONSHIP MANAGEMENT/ DATA MART/ ONLINE ANALYTICAL PROCESSING/ ASSOCIATION RULE MINING/ VAT REFUND FOR TOURISTS

194 pages

ระบบบริหารลูกค้าสัมพันธ์เชิงวิเคราะห์ สำหรับการคืนภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยว

AN ANALYTICAL CRM SYSTEM FOR VAT REFUND FOR TOURISTS

สมใจ เสมอเหมือน 4737983 ITCS/M

วท.ม.(วิทยาการคอมพิวเตอร์)

คณะกรรมการที่ปรึกษาโครงการวิจัย: ทรงศรี ตั้งศรีไพโรจน์ Ph.D., ชาญยศ ปลื้มปิติวิริยะเวช Ph.D.

บทคัดย่อ

การคืนภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยว (VAT Refund for Tourists (VRT)) เป็นนโยบายทางด้านภาษีเพื่อสนับสนุนนโยบายของรัฐบาลในการส่งเสริมการท่องเที่ยว โดยอยู่ในความรับผิดชอบของกรมสรรพากร ปัญหาในระบบงานปัจจุบันพบว่าการจัดทำรายงานที่มีลักษณะ Manual ใช้เวลามาก จำนวนผู้ประกอบการและร้านค้าที่อยู่ในระบบ VRT มีจำนวนน้อย และเจ้าหน้าที่กลุ่มงานคืนภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยว (คท.) ยังขาดเครื่องมือที่ช่วยในการวิเคราะห์ข้อมูลการซื้อสินค้าของนักท่องเที่ยว เพื่อนำไปหากลุ่มผู้ประกอบการภาษีมูลค่าเพิ่มให้สมัครเข้ามาอยู่ในเครือข่ายการขายสินค้าให้แก่นักท่องเที่ยวมากขึ้น

โครงงานวิจัยนี้ จึงได้ออกแบบและพัฒนาระบบ An Analytical CRM System for VAT Refund for tourists (ACRM-VRT) โดยนำแนวคิดของ Customer Relationship Management (CRM) ความรู้ทางด้าน Data Warehousing, OLAP และ Data Mining มาประยุกต์ใช้ ได้มีการสร้าง Data Mart สำหรับจัดเก็บข้อมูลการคืนภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยว นำ OLAP Tool มาใช้วิเคราะห์ข้อมูลการคืนภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยว เพื่อสนับสนุนการทำงานของเจ้าหน้าที่ (คท.) ในการนำเสนอรายงาน และข้อมูลที่เป็นประโยชน์ต่อผู้บริหาร (คท.) และ Data Mining Tool วิธีการของ Association Rule เพื่อศึกษาพฤติกรรมการซื้อสินค้าของนักท่องเที่ยว เพื่อหากลุ่มผู้ประกอบการ VAT ที่มีร้านค้าที่ขายสินค้าที่ตรงกับความต้องการซื้อสินค้าของนักท่องเที่ยว

ระบบ ACRM-VRT ได้ผ่านการประเมินความพึงพอใจในการใช้ฟังก์ชันงานต่าง ๆ จากเจ้าหน้าที่ (คท.) ซึ่งผลการประเมินโดยรวมออกมาอยู่ในระดับพึงพอใจมาก โดยฟังก์ชันงานที่ได้รับความพึงพอใจมากที่สุดคือ Data Mining Tool เนื่องจากความรู้ที่ได้ สามารถนำไปกำหนด กลยุทธ์ในการหากลุ่มผู้ประกอบการเป้าหมายที่น่าจะชักชวนให้สมัครเข้ามาอยู่ในระบบ VRT และสามารถนำความรู้ที่ได้ไปประชาสัมพันธ์ให้แก่ผู้ประกอบการที่อยู่ในระบบ VRT เพื่อใช้ในการวางแผนการขายสินค้าของตนเองได้ และผู้ประกอบการที่ยังไม่อยู่ในระบบ VRT อยากสมัครเข้ามาอยู่ในเครือข่ายของการคืนภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยวมากขึ้น

194 หน้า

CONTENTS

CONTENTS (cont.)

CONTENTS (cont.)

CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

CHAPTER I
INTRODUCTION

1.1 Research Motivation

A VAT Refund for tourists (VRT) is a tax policy in support of the government policy on tourism promotion aimed at making Thailand the center of tourists' commodity purchase. Apparently, this will attract foreign currencies into the country, resulting in monetary circulation in Thailand. Currently, the Revenue Department has assigned the working group for management of a VAT refund for tourists (VRT) to be responsible for providing the service of a VAT Refund to tourists at Thailand's international airports. This also includes planning for and development of the system of a VAT refund so that it is accurate, swift and efficient. In addition, there are advertisements and public relations campaigns for stores and shops to tourists in order to promote sales for entrepreneurs selling their products to tourists [1].

In fiscal year 2008, there were a total of 379,040 VAT entrepreneurs, 965 of which were in the VRT system, apart from 3,871 VRT shops. It is devious that the number of entrepreneurs in the VRT system is small, compared to that of entrepreneurs in the VAT system [2]. Besides, the VRT officers still lack necessary instruments for making real time ad hoc reports, analysis of data in support of the VRT management team's strategic planning and decision making to improve the services of tax refund to tourists and the services for VRT entrepreneurs, including drawing more VAT entrepreneurs into the VRT system.

Customer relationship management (CRM) is a marketing strategy. It seeks to understand customers' behaviors and desires and applies this knowledge to building good relations with customers, by using the technology of data analysis i.e. Online Analytical Processing (OLAP) and data mining, to help analyze customer data existing in the company for a better understanding of customers' behaviors and

desires. This, of course, assists a company to present its goods and services that correspond with customers' needs. Furthermore, the analysis result can be used for better planning and improvement of the company's performance.

In this research project, the analytical customer relationship management system for VAT refunds for tourists (ACRM-VRT) has been designed and developed through and establishment of a data mart for storing data on VAT refunds for tourists (VRT), apart from OLAP and Data Mining tool used to analyze VRT data. In this connection, OLAP tool is used to generate reports and data mining tool through the association rule technique is used to investigate tourists' goods purchase behavior in order to support the VRT officers' work performance of presenting reports and information useful for the VRT management team so that it can improve the VRT services and strategies of access to VAT entrepreneur groups owning stores selling goods meeting tourists' demand. Additionally, the knowledge gained can be disseminated to entrepreneurs coming in the VRT system in order to plan for and increase their products, making them more satisfied to be in the VRT network.

1.2 Problems Statement

Problems and obstacles of the current work system include the following.

1) Currently, reports are made manually by making Excel files and then transforming these into Html files, which typically takes a long time. Moreover, the format of reports in only one point-of-view.
2) The number of entrepreneurs in the VRT system is still small, compared to that in the current VAT system.
3) VRT officers lack equipment used to analyze data on tourists' goods purchase so as to use the data to make public relations campaigns and induce more VAT entrepreneurs into a network of selling goods to tourists.

1.3 Research Objectives

This research project has the following objectives.

1) To design and develop the analytical customer relationship management system for VAT refunds for tourists (ACRM-VRT), which includes the following.
    - Data Mart for storing data on entrepreneurs in the VRT system, VAT payment of VRT entrepreneurs, and VAT refunds to tourists, all of which will be used to make reports.
    - OLAP to enable VRT officers to access data rapidly, retrieve data in different dimensions and make ad hoc reports efficiently.
    - Data Mining for studying tourists' goods purchase behavior, and use the obtained knowledge to find entrepreneurs who should take part in the VRT system. Besides, VRT entrepreneurs can apply their knowledge, derived from the system, to planning for selling their products.
2) To evaluate VRT officers' satisfaction on the use of ACRM-VRT system.

1.4 Scope of Research

This research project has the following scope of work.

1) Data used in this study are derived from the system of VAT refunds to tourists (VRT) and VAT system.
2) Development of the ACRM-VRT system in support of VRT officers, which include:
    - Data Mart for storing data on entrepreneurs in the VRT system, VAT payment of VRT entrepreneurs, and VAT refunds to tourists.
    - Development of the web-application ACRM-VRT system, which consists of 2 data analysis tools, used to analyze VRT data, listed below.

- OLAP tool to be used to create and hoc reports as required by VRT officers.
- Data Mining tool, through the technique of association rule mining using an apriori algorithm, to be used for analysis of data on tourists' goods purchase.

1.5 Organization of the Research Project Document

The report of this research includes 7 chapters as the following:

Chapter 1: Introduction

Discuss project background, problems and obstacles, objectives and scope of research, as well as organization of the research report.

Chapter 2: VAT Refund for Tourists System

Describe the overall current system of VAT refunds for tourists (VRT), including different reports, so as to be used to analyze the requirements of system users to design and develop the ACRM-VRT system later.

Chapter 3: Literature Reviews

Discuss the concepts and theories of customer relationship management (CRM), the design and development of data warehouse, OLAP and data mining through the association rule technique, using a apriori algorithm in support of the analysis of CRM data.

Chapter 4: ACRM-VRT System Design

Describe the design of the analytical customer relationship management system for VAT refunds to tourists (ACRM-VRT), including an explanation of system components.

Chapter 5: ACRM-VRT System Implementation

Discuss the requirements for hardware and software used for the system implementation, and functions of the ACRM-VRT system that has been developed.

Chapter 6: System Evaluation and Results

Discuss an evaluation of the satisfaction of VAT refund for tourists (VRT) officers on the use of the ACRM-VRT system, which includes objectives, techniques and evaluation results.

Chapter 7: Conclusion and Future Work

The chapter summarizes the ACRM-VRT system and the future work.

CHAPTER II
VAT REFUND FOR TOURISTS SYSTEM

This chapter discusses the overview of VAT refunds to tourists, the current system of VAT Refund for Tourists, and VAT refunds to tourists in foreign countries. Details of the above issues are given below.

2.1 Overview of VAT Refund for Tourists

VAT refund to tourists are defined to support the government policy on tourism promotion and push Thailand to become, a shopping center for tourists. Apparently, this is beneficial in drawing foreign exchanges into the country, generating multiplier effects. Due to the above reason, the Revenue Department of Thailand has introduced the system of VAT refund for tourists (VRT) since 1999 onwards. The work has been assigned to the working group for management of VAT Refund for Tourists officers to implement VRT accurately, swiftly and efficiently, both for the case of VAT refunds in cash and on credit at Thailand's five international airports, namely, Suvarnabhumi, Chiangmai, Phuket, Hat Yai and U-Tapao [1].

At present, the Revenue Department has a web-application VRT system for recording data on entrepreneurs and their places of establishments, including VRT data. Based on a statistical report (Working group for management of VAT refunds to tourists, November 17, 2008) [2], in fiscal year 2008, there were 965 entrepreneurs in the VRT system, 3,871 stores, 351,467 tourists requesting for VAT refunds, 11,119.74 million Baht of tourists' goods purchases, and 778.38 million Baht of VRT amount. Detail of statistical reports between fiscal years 1999 and 2008 are illustrated in Figures 2.1 through 2.5.

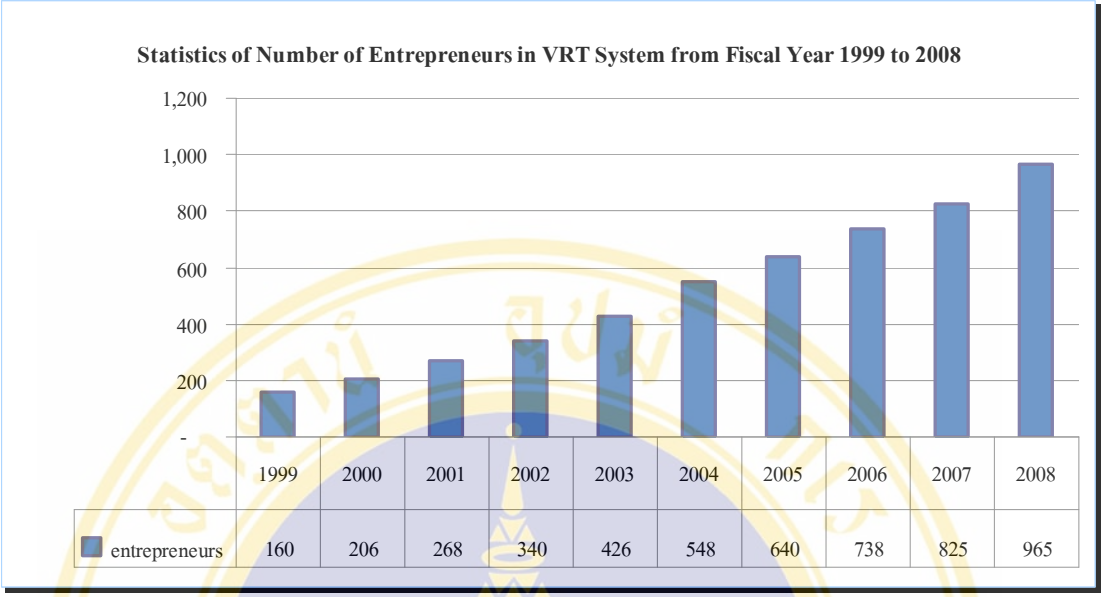


Figure 2.1 Statistics of number of entrepreneurs in VRT System from fiscal year 1999 to 2008

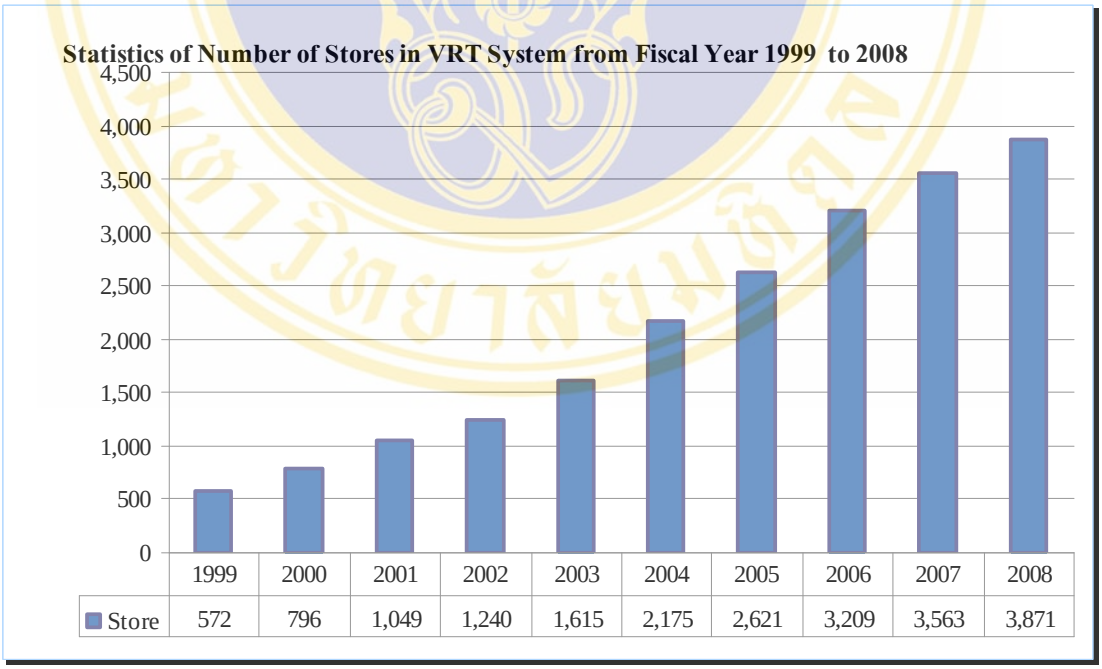


Figure 2.2 Statistics of number of stores in VRT System from fiscal year 1999 to 2008

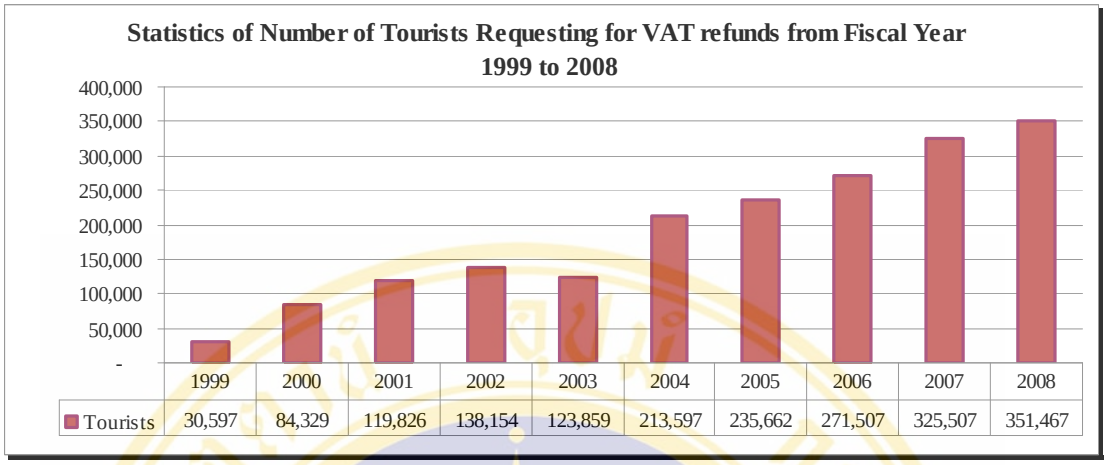


Figure 2.3 Statistics of number of tourists requesting for VAT refunds from fiscal year 1999 to 2008

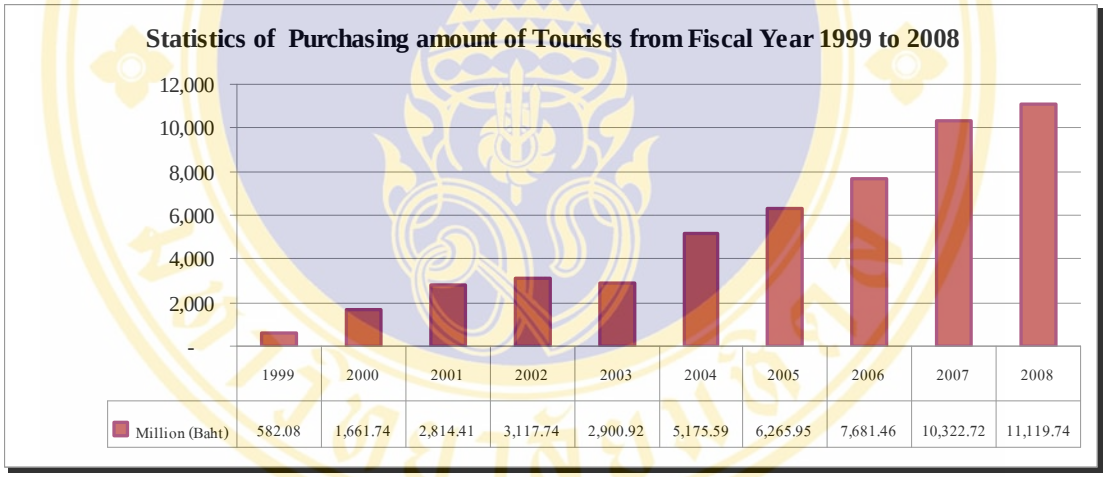


Figure 2.4 Statistics of purchasing amount of tourists from fiscal year 1999 to 2008

Figure 2.5 Statistics of VAT amount refunded to tourists from fiscal year 1999 to 2008

For VAT refunds to tourists, the Revenue Department has defined stages of tourists' request for VAT refunds, the evidences to be submitted for VAT refunds request, and qualifications of VRT entrepreneurs. Details of the above issues are given below.

2.1.1 Stages of tourists' request for VAT Refund

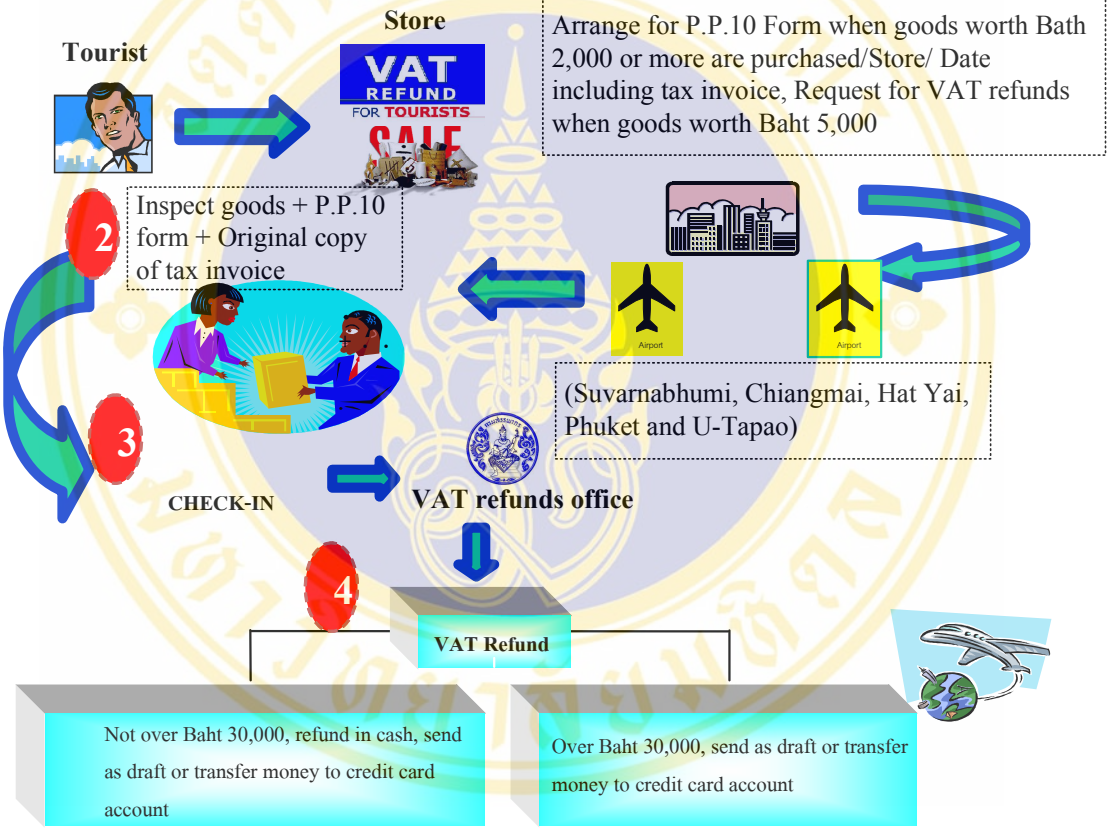


Figure 2.6 Overall picture illustrating stages of VAT refunds to tourists

Figure 2.6 illustrates stages of tourists' request for VAT refunds, of which details are described below.

1) When a tourist buys goods with a total value of no less than 2,000 baht (goods prices including VAT) from a store in the VRT system, that establishment place (store) has to arrange for a request for VAT refunds to tourists (P.P.10 Form) and give it to tourist along with a tax invoice.

The tourist may request for VAT refunds when the total amount of goods purchases from all P.P.10 Form is worth 5,000 baht or over.

2) When a tourist leaves Thailand at any one of the five international airports, he may present to the customs officers the goods bought through P.P.10 Form, and an original copy of VAT invoice for an inspection of goods, of which VATs are requested for refunds, for the person departing from the kingdom. The inspection is made at the international departure hall before the tourist checks in at a counter of an airliner.
3) When the tourist has checked in at a counter of an airliner, he has to make a request for VAT refunds at a counter of the Revenue Department. In the case that the tourist bought gemstones made into shape or gold ornaments, watches, eyeglasses or pens, he has to present such goods, worth from 10,000 baht to more, to revenue officers for VAT refunds request.
4) Tourists' request for VAT refunds.
   - In the case that the VAT amount does not exceed 30,000 baht, it may be refunded, upon request, in cash (Baht), by draft or money transfer to a tourist's credit card.
   - In the case that the VAT amount exceed 30,000 baht, it may be refunded, upon request, by draft or money transfer to a tourist's credit card.

2.1.2 Evidences for VAT Refunds Request

For a tourist's request for VAT refunds, the Revenue Department has defined the following evidences that a tourist has to submit for his request [4].

1) A request form for VAT refunds (an example illustrated in Figure 2.7).
2) An invoice of goods purchased (an example shown in Figure 2.8).
3) A tourist's passport (an example displayed in Figure 2.9).
4) Goods purchased.

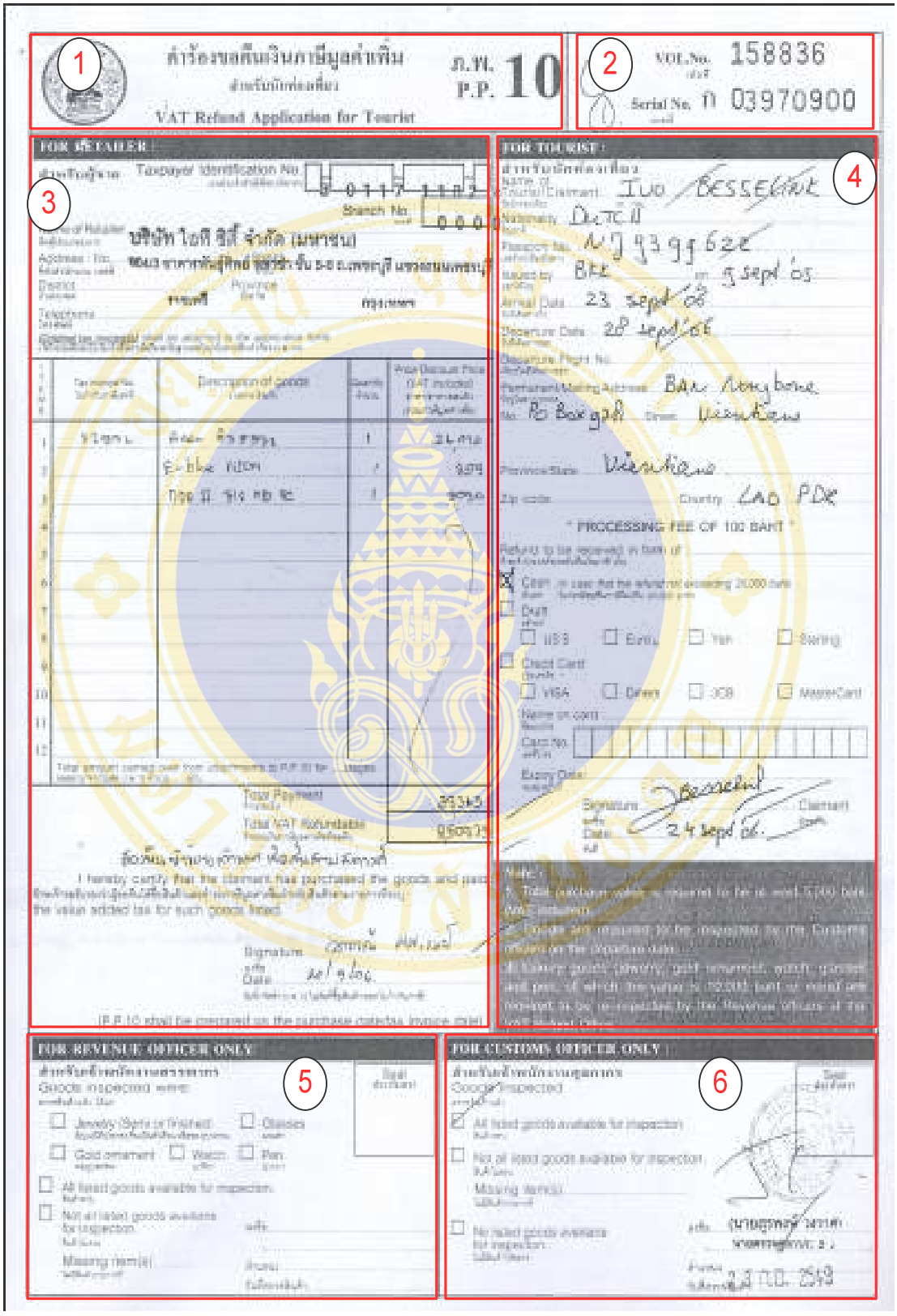


Figure 2.7 An example of a request form for tourists' VAT refund

Figure 2.7 illustrates an example of a request form for tourists' VAT refunds, consisting of,

1) Name of a request form.
2) Volume and serial numbers of a request form.
3) Details of goods entrepreneur, comprising,
   - Taxpayer identification number
   - Name of entrepreneur
   - Address of entrepreneur
   - List of goods purchased
4) Details of a tourist, including,
   - Tourist's name
   - Nationality
   - Number of passport
   - Address
   - Type of VAT refund request
5) Detail of an inspection of the Revenue Department's officers.
6) Detail of an inspection of the Customs Department's officers.

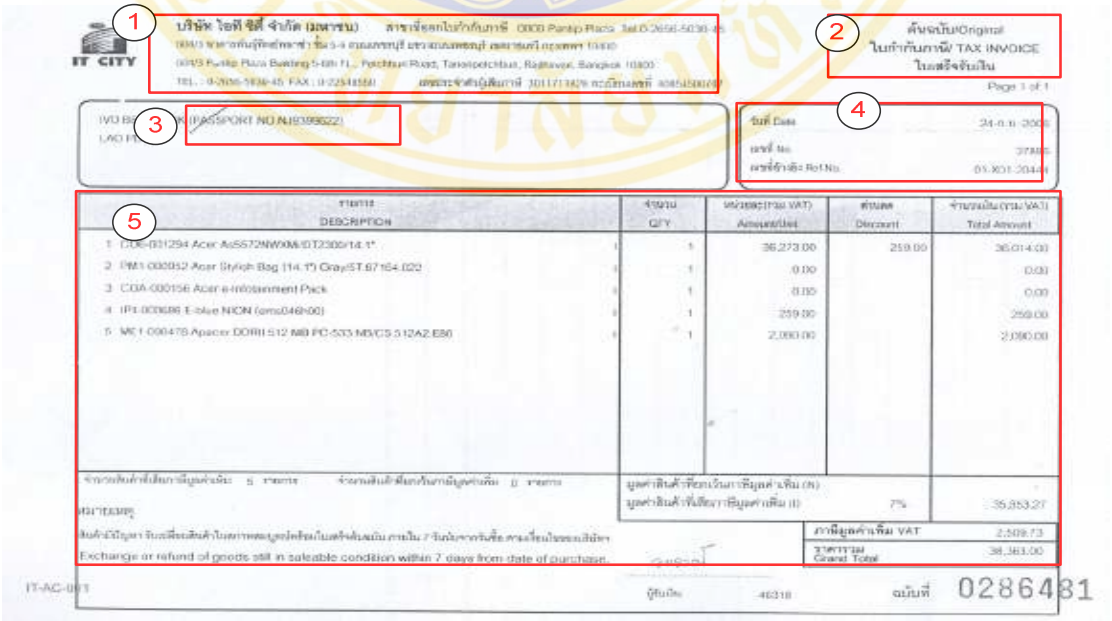


Figure 2.8 An example of a tax invoice

Figure 2.8 illustrates an example of a tax invoice, which consists of the following details.

1) Name, address and taxpayer identification number of an entrepreneur.
2) A statement indicating the word “A tax invoice”.
3) A tourist’s passport number.
4) Date of goods sale and number of tax invoice.
5) Items of goods purchased, including goods prices, VAT and the total amount of money.

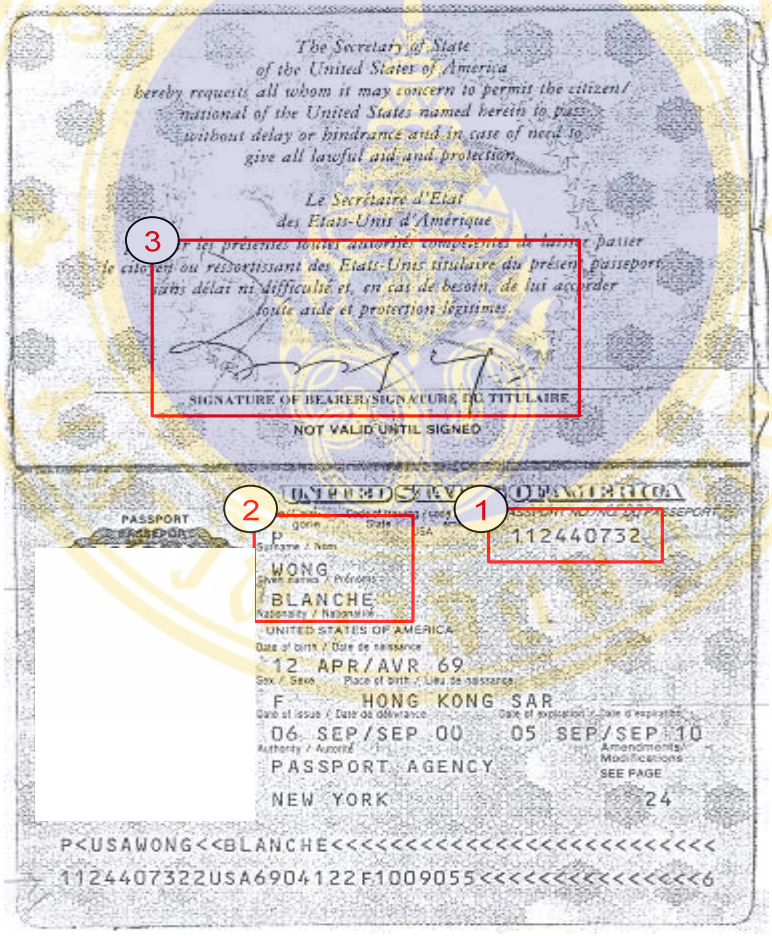


Figure 2.9 An example of a tourists’ passport

Figure 2.9 illustrates an example of a tourist’s passport including,

1) Passport number
2) A Tourist’s name
3) A Tourist’s signature

2.1.3 Qualifications of VRT Entrepreneurs

An entrepreneur who is in a VRT system has to have the following qualifications.

1) Have to be a registered VAT entrepreneur.
2) Be a company or a legal entity partnership, with the following paid-in capital.
   - No less than 2 million baht for an entrepreneur in Bangkok.
   - No less than 500,000 baht for an entrepreneur in the provincial area.
3) Be secure and continuous in the business transaction and have a good record of tax payment.

2.2 Current System of VAT Refunds for Tourists

Currently, A system of VAT refunds for tourists (VRT) has been developed through a web application. It is divided into three parts of entrepreneurs, VAT refunds and statistical reports, the detail of which are revealed in Figure 2.10 while statistical reports details are given in Table 2.1.

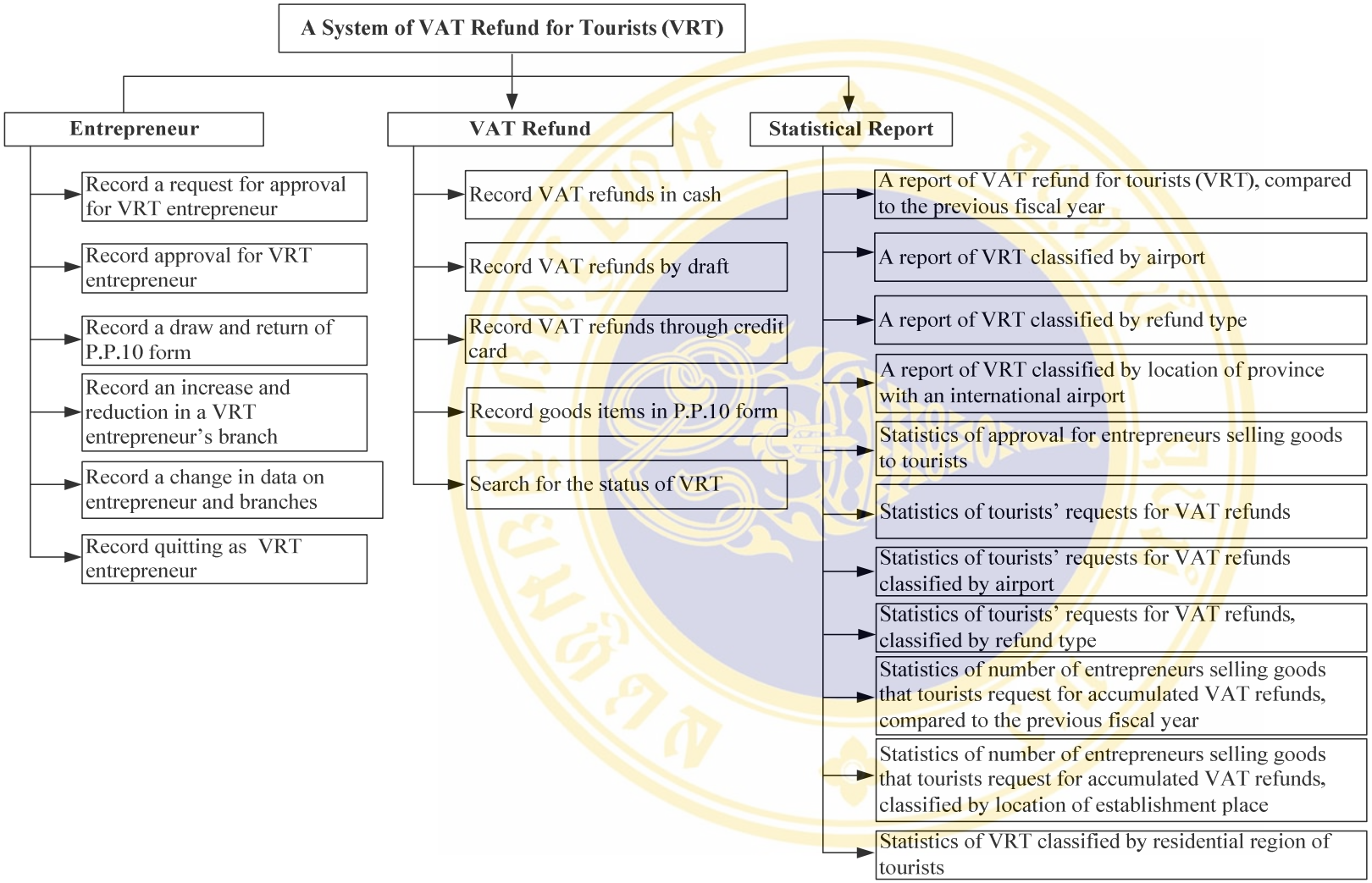


Figure 2.10 Main functions of VRT system divided into three parts of entrepreneurs, VAT refunds and statistical reports

Table 2.1 Statistical reports in the current system

Item	Processing
1. A report of VAT refund for tourists (VRT), compared to the previous fiscal year.	Monthly
2. A report of VRT classified by airport.	Monthly
3. A report of VRT classified by refund type.	Monthly
4. A report of VRT classified by location of province with an international airport.	Monthly
5. Statistics of approval for entrepreneurs selling goods to tourists.	Fiscal yearly
6. Statistics of tourists' requests for VAT refunds.	Fiscal yearly
7. Statistics of tourists' requests for VAT refunds, classified by airport.	Fiscal yearly
8. Statistics of tourists' requests for VAT refunds, classified by refund type.	Fiscal yearly
9. Statistics of number of entrepreneurs selling goods that tourists request for accumulated VAT refunds, compared to the previous fiscal year.	Fiscal yearly
10. Statistics of number of entrepreneurs selling goods that tourists request for accumulated VAT refunds, classified by location of establishment place.	Fiscal yearly
11. Statistics of VRT classified by residential region of tourists.	Fiscal yearly

2.2.1 Problems of The Current System

Problems found in the current system may be summarized as follows:

1) The current report making is carried out manually, by creating and Excel file and then transforming it into an Html file, which takes a lot of time to make a report. Besides, the report format is just one point of view (dimension).
2) The number of entrepreneurs in the VRT system is small, analytical results to that in the VAT system at present.

3) VRT officers of the Revenue Department are short of tools that help to analyze data on tourists' purchases of goods in order to launch a public relations campaign and attract more VAT entrepreneurs into a network of selling goods to tourists.

2.2.2 Requirements of The System Users

Based on a study of the requirements of users, the following can be concluded.

1) Tools for making reports and on-line data analysis, including tools to support the making of ad hoc reports.
2) Tools for analysis of data on tourists' goods purchases so that analytical results may be used to make public relations and draw a larger number of VAT entrepreneurs into a network of selling goods to tourists.
3) Tools for analysis of data on goods sales to tourists of each entrepreneur so that outcomes may be used to plan for his sale of goods in order to enhance entrepreneurs satisfaction on coming more into a network of selling goods to tourists.

2.3 VAT Refund for Tourists in Foreign Countries

There are about 40 countries worldwide that have a policy of VAT refunds to tourists [5], for instance, Singapore, Japan, South Korea, Australia, Canada and all countries in Europe. Each of these countries has different requirement and practices.

2.3.1 Models of VAT Refunds to Tourists in Foreign Countries

Based on a study of the National Tourism Alliance (NTA)[5], an agency promoting Australia's tourism, patterns of VAT refunds to tourists in countries refunding VAT on goods purchases include 3 models listed below.

1) The Government processing model (G Model)
2) The Reimbursement Agent model (RA Model)
3) The Retailer model (R Model)

Table 2.2 makes a comparison of the three above models.

Table 2.2 A comparison of patterns of VAT refunds to tourists based on a study of The National Tourism Alliance (NAT)

Approach	Pattern		
	G Model	**RA Model**	**R Model**
1. Agency responsible for VAT refunds to tourists.	Government agencies such as the Revenue Dept. and Customs Dept.	Companies that are agents and get permission for their governments to act as agents for VAT refunds to tourists, by collecting fees from tourists and asking for a reimbursement from the government for VAT refunded to tourists, for example, Global Refund Company, Premier Tax Free Company, etc.	Goods stores will refund VAT to tourists. Goods items upon which tourists request for VAT refunds are not used for VAT calculation for the government. The stores have to apply to a company acting as an agent for VAT refunds that get permission from the government, such as Global Refund Company, Premier Tax Free Company, etc.
2. Type of goods which may be requested for VAT refunds.	Goods which are calculated for VAT refunds and have to pass an inspection from customs officers of the country first.	Goods which are calculated for VAT refunds and have to pass an inspection from customs officers of the country first.	Goods which are calculated for VAT refunds and have to pass an inspection from customs officers of the country first.
3. Countries that use.	Thailand and Australia.	Iceland and Canada.	All countries in Europe, Singapore, Argentina, South Korea, Norway, Switzerland, Turkey and Lebanon.

For an approach of VAT refunds to tourists, each country has different approaches and conditions. Table 2.3 makes a comparison of an approach of VAT refunds to tourists of Thailand, Australia And Singapore [4][6][7].

Table 2.3 A Comparison of VAT refunds to tourists among Thailand, Australia and Singapore

Item	Thailand	Australia	Singapore
1. Agency responsible for VAT refunds to tourists.	Revenue department	Customs Department	- Global Refund Company - Premier Tax Free Company
2. Total prices of goods that may be requested for VAT refunds.	Baht 2,000 or more per store and a total purchase amount must not be less than Baht 5,000.	Australian dollar 300 or more.	Singaporean dollar 100 or more and not exceed Singaporean dollar 300 per store.
3. Period of time that goods may be brought to request for VAT refunds.	Not exceed 60 days.	Not exceed 30 days.	Not exceed 60 days.
4. Things that a tourist has to present to customs officers for an inspection.	- True copy of tax invoice. - Request form for VAT refunds (P.P.10 Form). - Passport. - Goods purchased.	- True copy of tax invoice. - Passport - Boarding pass - Goods purchased	- Passport - Letter of VAT refunds request. - Goods purchased
5. Services given to a store.	- Logo put at the front of the store. - Brochure of VAT	- Logo put at the front of the store. - Brochure of VAT	- Logo put at the front of the store. - Brochure of vat refunds. - Statistical report of tourists' goods

Item	Thailand	Australia	Singapore
	refunds in 4 languages, namely, Thai, English, Japanese and Chinese.	refunds in 14 languages such as English, Japanese, Thai Chinese, and French	purchases classified by goods type and country, in each store.
6. Services provided to tourists.	- Brochure suggesting an approach and stages of VAT refunds in 4 languages, namely, English, Japanese, Thai and Chinese.	- Brochure suggesting an approach and stages of VAT refunds in 14 languages, such as English, Japanese, Thai, Chinese, French, etc.	- Suggestions on an approach and stages of vat refunds in such languages as Spanish, Japanese, Chinese And French on the website of an agent company (Global Refund). - A search for stores that tourists can request for VAT refunds, on the website of an agent company (Global Refund). - Monitor the status of VAT refunds on the website of an agent company (Global Refund).

2.3.2 Example of VAT Refunds to Tourists of Global Refund Company

Global Refund Company is a company in the Netherlands. It was established in 1980 and has made a business transaction as an agent , implementing VAT refunds to tourists for countries worldwide. At presents, there are 35 countries altogether using its services, for instance, all countries in Europe, Singapore, Argentina, South Korea, Norway, Switzerland and Turkey. The company provides services to tourists by refunding VAT to 30,000 of them per day. There are about 230,000 stores across the world using its service. The company has an easy, swift and convenient approach of refunding VAT to tourists purchasing goods. It possesses a management program called Global Refund's In-store Processing Solution (GRIPS)[8], which is a system of issuing a request form for tourists' VAT refunds, including compiling statistics of goods sale of each store, which can retrieve data on its good sale to tourists. In addition, the company also provides services, on its website, of advertising its customers' stores in order that tourists may search for data on sources of goods purchases, as illustrated in Figure 2.11, Apart from services of statistical reports on tourists' goods purchases worldwide, as illustrated in Figures 2.12 and 2.13

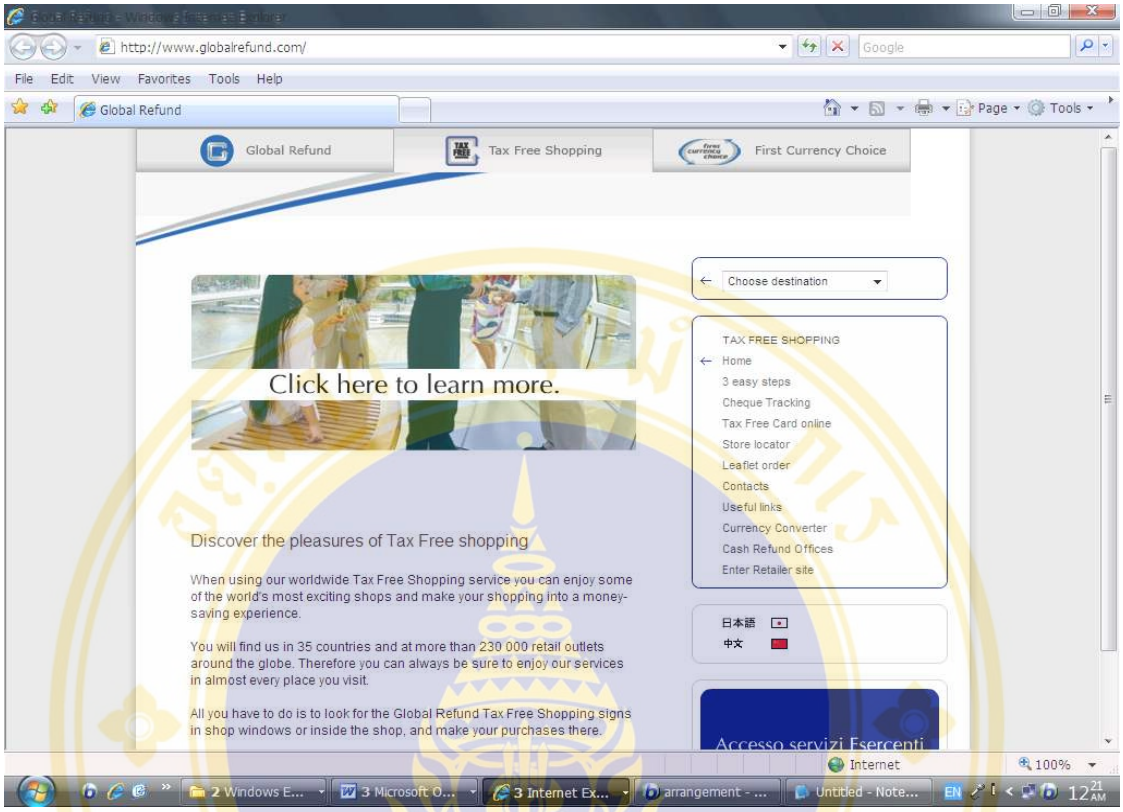


Figure 2.11 Website of Global Refund Company [8]

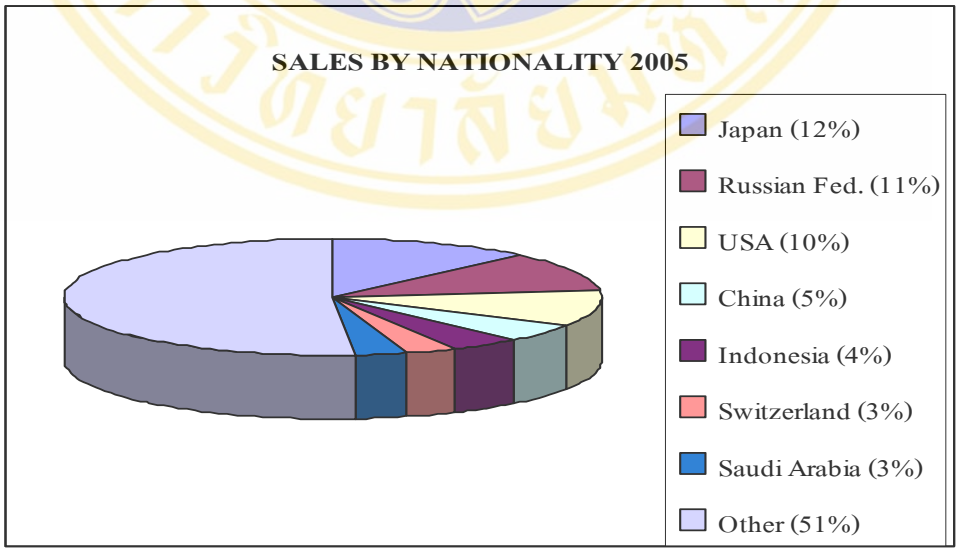


Figure 2.12 A graph showing the ratio of nationality of tourists buying goods and making a request for VAT refunds [8]

Figure 2.12 illustrates a graph displaying the proportion of tourists purchasing goods and requesting for VAT refunds in 2005, classified by nationality. The first 3 countries whose populations, as tourists, purchased the largest amounts of goods are Japan (12%), Russia (11%) and the United States (10%), respectively.

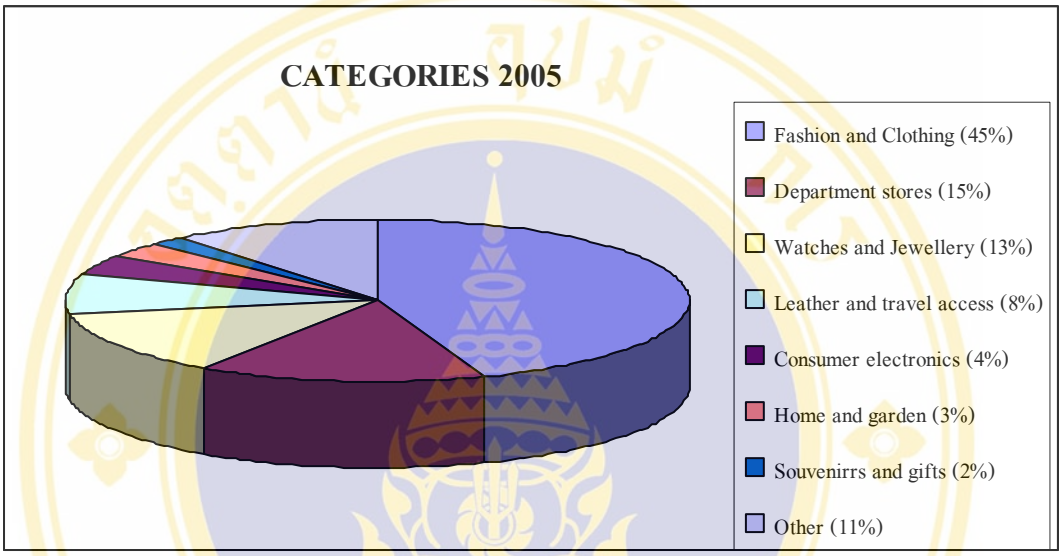


Figure 2.13 Type of goods that tourists preferred to buy in 2005 [8]

Figure 2.13 illustrates a graph revealing goods which tourists preferred to purchase, classified by goods category. The first 3 categories of goods that tourists preferred to buy the most include goods on fashion and clothing (45%), goods in department stores (15%) as well as watches and jewellery (13%), respectively.

CHAPTER III
LITERATURE REVIEWS

This chapter describes about basic knowledge of data warehousing and data mining, the idea and concept of customer relationship management (CRM). Moreover, the method of the Association Rule Mining which is called Apriori Algorithm is used to analyze our CRM data.

3.1 Data Warehousing and Data Mining

The basic knowledge of Data Warehousing and Data Mining will be discussed as the following.

3.1.1 Data Warehouse

W.H. Inmon stated that "Data Warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management' decision making process" [18].

The data warehouse definition mentioned by W.H. Inmon can be summarized 4 major characteristics of a data warehouse as shown below [10].

1) **Subject-oriented** : the data warehouse that is designed to describe the specific business subject.
2) **Integrated** : the data an integration of multiple database systems and managing them into the same standard. For example, transforming variables of the data into the same format.
3) **Time-variant** : the data in the data warehouse must record in a defined time period that relates to business process of the business unit. This is because business decision needs a specific period of time in the comparison

4) **Nonvolatile** : the data in the data warehouse should not be changed frequently in terms of updating or deletion. The user needs only to access the data.

3.1.1.1 Data Warehouse Architecture

The architecture of Data Warehouse includes components, processes and tools. The detail is shown in Figure 3.1.

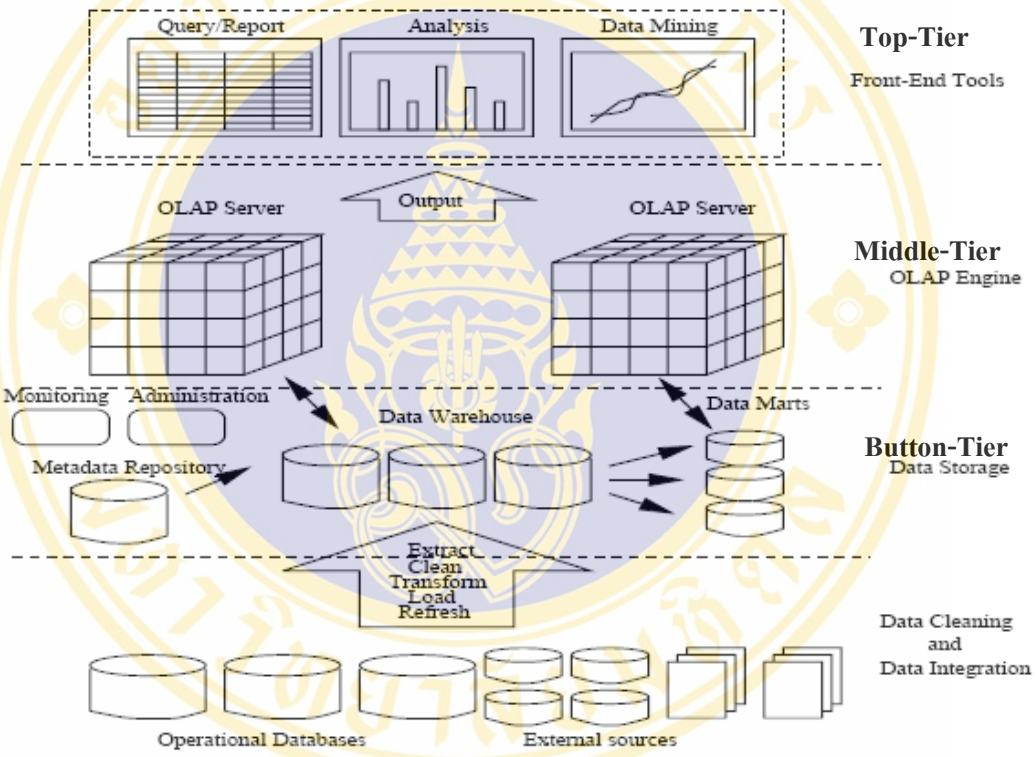


Figure 3.1 A three-tier data warehousing architecture [18]

The three-tier data warehousing architecture shown in Figure 3.1 can be described as the following [15][18][22]

1) The Bottom Tier or Data Storage is the Database Server for storing data of the Warehouse (Warehouse Database Server). Most of data is stored as Relational Database System. Data in Data Warehouse come from many sources such as Operational Database or External Sources. In the data warehouse, the data may be divided and stored in the

Data Marts for a specific purpose and a fast and convenient way in data access and management. For example, Marketing Data Mart is used for marketing information. The information includes the product information, customer information, and sales information. Moreover, the bottom tier has the repository metadata which stores the structural detail, data description detail, management detail of the Warehouse. In addition, the information describes tools for monitoring and managing the Warehouse system.

Those tools will deploy the data in the repository metadata. However, before loading information into the Data Warehouse, the data need to pass the Extract, Transform, and Load (ELT) process which has the following steps.

- Data Extraction is a process to select extract the interesting information from data sources.
- Data Transformation is a process to transform data from the source data format to the defined warehouse format. Typically, there are five processes which include.
  - Cleaning Process: the process for adjusting and correcting the data.
  - Add Attribute: the process for adding new fields for collecting and querying data.
  - Splitting/Merge: the process for splitting and merging data.
  - Creating Surrogate Key: the process for creating primary key in the case that there is no primary key field in the existing data.
  - Summarizing : the process for summarizing the data using the defined data warehouse structure.
- Load is a process of importing the correct data into the Data Warehouse.

- Refresh is a process of updating the data in the Data Warehouse as our defined time period such as daily, weekly or monthly.

2) The Middle Tier (OLAP Engine) is the online analysis server running in between the Warehouse database server and the front-end client. This tier is responsible for data analysis and multi-dimensional OLAP querying. There are three types of the OLAP server which includes MOLAP, ROLAP, and HOLAP.
3) The Top Tier or Front-End Tools are the user interface tools that interact with users and OLAP engine. This tools are capable for analyzing the data and displaying multi-dimensional information such as Analysis Tool, Query/Reporting Tool, and Data Mining Tool.

3.1.1.2 Multidimensional Data Model

Data Warehouse has been designed to be multi-dimensional model. The model is designed as data cube, therefore it is possible to view in multi-dimensional.

Multi-Dimensional data model consists of two types of database tables : Fact tables and Dimension Table.

- Fact Table is the table that stores calculated information and extracted data from the data source. The information could be summary data such as the sale amount, the number of customer, or the key for linking to the dimension table. The fact table always needs to be updated.
- Dimension Table is the table for storing entity data such as customer and product information. In Data Warehouse or Data Mart, the dimension table is the table that describes information in fact table in all perspectives.

Multi-Dimensional Model is data model for implementing data warehouse. There are 3 types including Star Schema, Snowflake Schema, and Fact Constellation.

1) Star Schema

Star schema looks like a star shape. It consists of one fact table and many dimension tables. The primary key of dimension table is in the fact table and used for referencing the data in the Fact table as shown in Figure 3.2

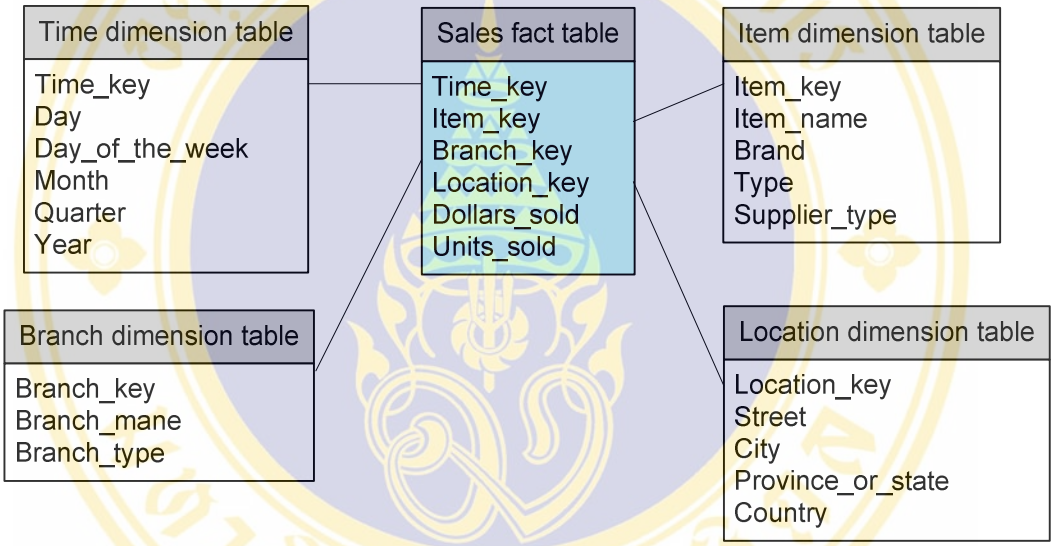


Figure 3.2 The structure of star schema [18]

Figure 3.2 consists of sale order table which is the fact table. The other dimension tables are Time, Item, Branch and Location tables.

2) Snowflake schema

Snowflake schema looks like a snowflake-molecule shape. It is similar to the Star schema, but its dimension table is divided in order to reduce data duplication as shown in Figure 3.3

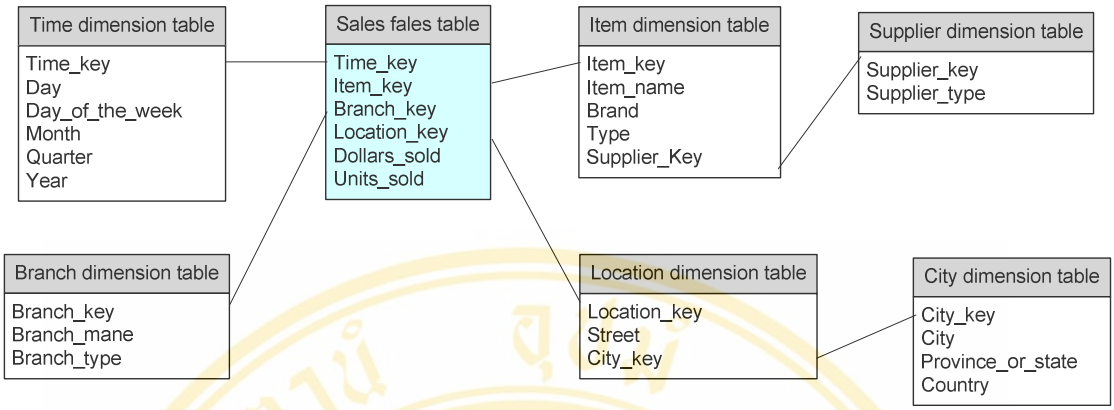


Figure 3.3 The structure of snowflake schema [18]

The figure 3.3 shows the dimension tables that are related to other dimension tables. The Item table relates to the Supplier table and the Location table relates to the City table.

3) Fact constellation schema

Fact constellation schema or Galaxy schema. There are more than one of the Fact table and the Dimension table is called at the same time shown in Figure 3.4

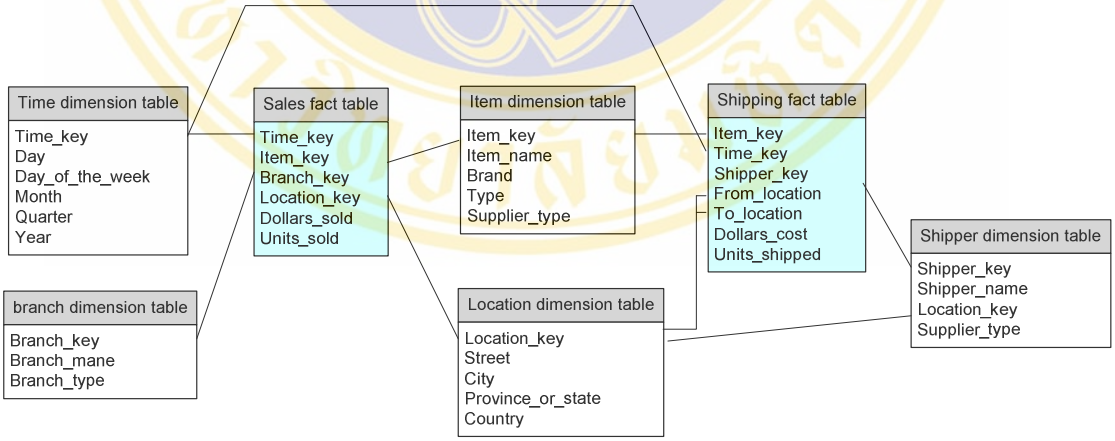


Figure 3.4 The structure of fact constellation schema [18]

The figure 3.4 shows the structure of Fact Constellation Schema. There are two fact tables including Sales and Shipping which are simultaneously using the same dimension tables that are Time and Location tables.

3.1.1.3 Data Mart

Data Mart is the subset of data warehouse implemented to serve each department or business unit in order to reduce the query access time. There are two types of data marts.

1) **Dependent Data Mart**

   Dependent data Mart is the small data warehouse that is designed for storing data in a specific subject. Most of the data comes from the existing Data Warehouse System.

2) **Independent Data Mart**

   Independent Data Mart is the small data warehouse using in the department level. Most of the data comes directly from multiple data-sources in order to reduce the access time.

3.1.2 Online Analytical Processing (OLAP)

OLAP is the online analytical processing. It is a technology that includes tools for complex querying from data warehouse system. There is a capability to analysis information in multi-dimensional tables. OLAP is designed to query the answer business questions from executive managers that need answer for some business decision. The feature to compare the information in different views is desired to summarize the results of the query to the executive users.

3.1.2.1 OLAP Functionalities

By default, the OLAP functions include Roll up, Drill down, Slice and Dice, and Pivot as shown in Figure 3.5

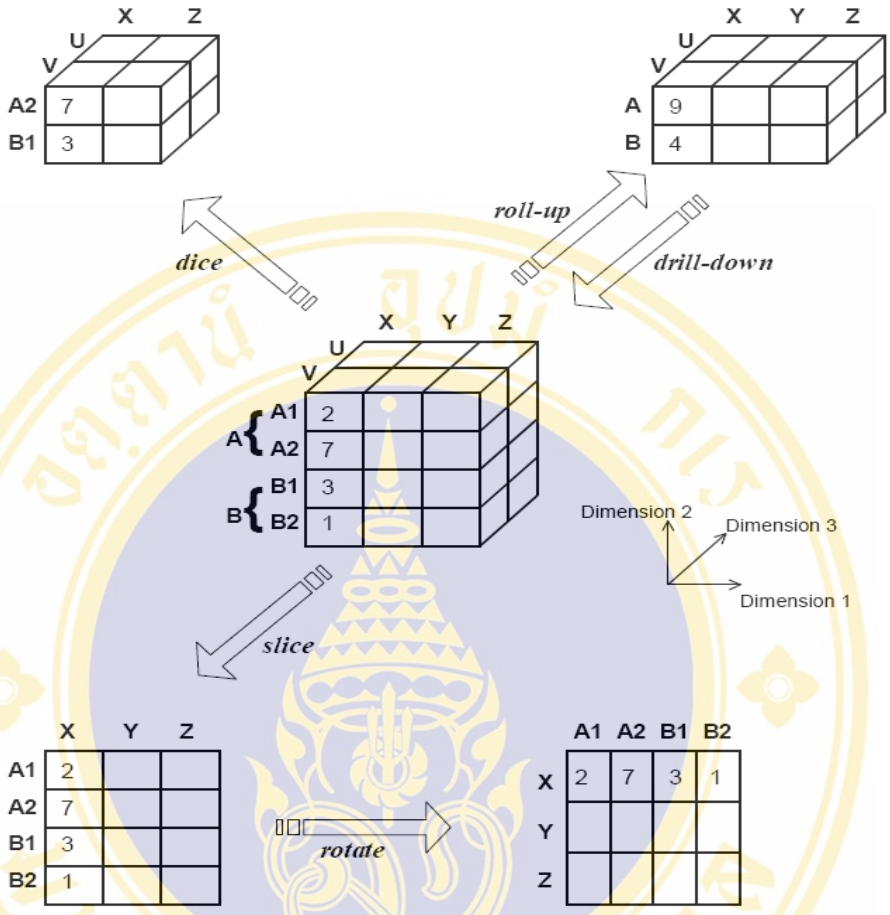


Figure 3.5 The model of OLAP procedures [18]

From the figure 3.5, the OLAP procedures can be described as shown below.

1) Roll up is the process of summarizing the data from the details to the summary
2) Drill Down is the process of expanding the data to show the details from the summary. This process is the opposite of the roll up.
3) Slice and Dice is the process of cube separation into sub-cubes. Slice is cube separation using only one dimension. Dice is cube separation using more than one dimension.
4) Pivot (rotate) is changes in the viewpoint of the data display

3.1.2.2 Types of OLAP Servers

There are three structures of the OLAP tools including MOLAP, ROLAP, and HOLAP [21].

1) MOLAP

MOLAP stands for Multidimensional Online Analytical Processing. The data will be loaded from Data Warehouse for summarizing data in advance. Then, it will be stored in the pre-defined structure as Multidimensional Cubes at MOLAP Server. The summarized values in the MOLAP server will be updated when the data in Data Warehouse has been changed. If user would like to query data from the MOLAP server, it can directly return the result without processing at RDBMS. The pro is fast processing and the con is that there is no searching in the detail data (lowest level) as shown in Figure 3.6

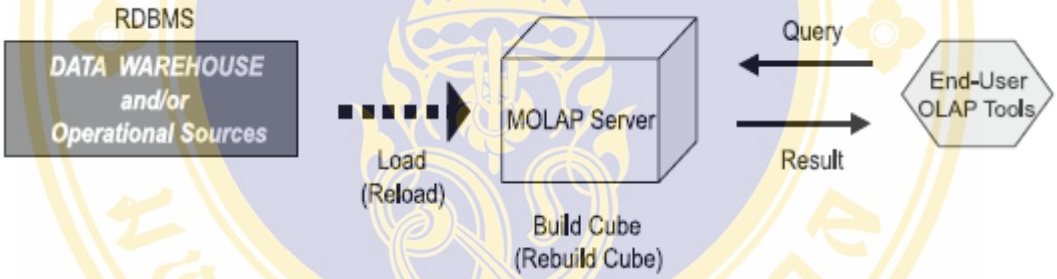


Figure 3.6 Typical MOLAP architecture [21]

2) ROLAP

ROLAP stands for Relational Online Analytical Processing, the data will be stored in relational tables at data warehouse in RDBMS. When users would like to query information, the ROLAP Server will send queries to process at RDBMS before return to users. The pro is that user can look into the lowest level and the con is it takes longer time than that of MOLAP. The diagram is shown in Figure 3.7

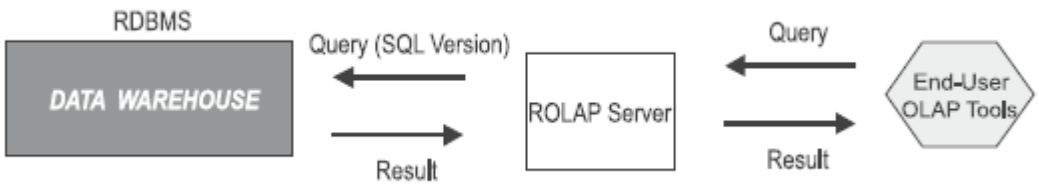


Figure 3.7 Typical ROLAP architecture [21]

3) HOLAP

HOLAP stands for Hybrid Online Analytical Processing. It is the mixed mode together between ROLAP and MOLAP. If users would like to query the lowest level, the query will be sent to RDBMS. On the other hand, query will be sent to cube. Therefore, HOLAP will have fast respond and be able to query in the lowest level.

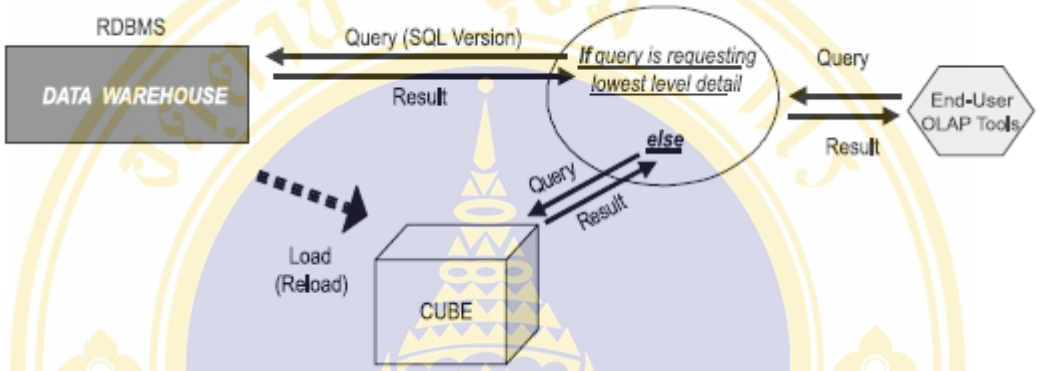


Figure 3.8 Typical HOLAP architecture [21]

3.1.3 Data Mining

Data Mining is the process of searching and analyzing data in large database system in order to find the rules and relationship of the data that is generated automatically or semi-automatically.

3.1.3.1 Types of Data Mining Tasks

In order to use data mining, we have to understand the problems and objectives of our study. There are six types of data mining tasks [9] as shown.

- Classification
- Estimation
- Prediction
- Association Rule
- Clustering
- Profiling

The details of each type of task are discussed below:

1) Classification

Classification divides data into categories as we defined their characteristics. We can use the data property to classify each group. For example, in business, the credit card companies use risk level (low, medium or high) for approving customer's credit line.

2) Estimation

Estimation uses to estimate data or some factors that we don't know its exact value by using the historical data for analysis. For example, the estimated profit from customers can be done by using customer's buying history for the future, as well as the estimated lifetime to keep the customer.

3) Prediction

Prediction uses for predicting or estimating future data. For example, we can predict which customer will no longer be our customers in the next six months.

4) Association Rule

Association rule shows the relationship of incident or object that always happens at the same time. For example, in marketing, the collected data from the point of sale (POS) can be used to analyze the selling data and consider the products that the buyers will buy together. In order to use this knowledge in shelving those products. By doing this, the company can increase more chances to sell the goods (cross-selling). They can also used the data for planning, grouping their goods and services for their promotion.

5) Clustering

Clustering is similar to classification but there is not specific details for each category. In clustering, the data will be grouped and the details of each group will be drawn. For example, in marketing, clustering can be used for grouping customers to make a promotion for selling goods or providing services that are suitable for each user.

6) Profiling

Profiling explains profile or what is the current situation and help the researchers understand the data better. For example, in marketing, profiling is used to explain customer characteristics and behavior, also called Customer Profiling.

This type of work helps the company to understand the customer need and provide good and service that is suitable for the customer. Therefore, the company can keep the customers longer.

3.1.3.2 CRISP-DM Process

CRISP-DM stands for Cross-Industry Standard Process for Data Mining [19] which is the standard for Data Mining. The purpose is to develop the Data Mining System in the same direction which is the cooperation of three companies: DaimlerChrysler, SPSS, and NCR (The National Cash Register Company). There are six phases as shown in figure 3.9.

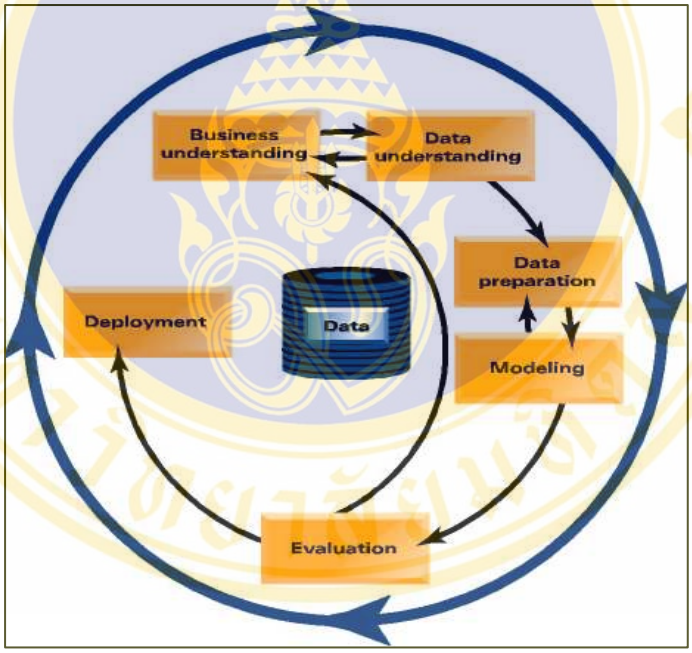


Figure 3.9 Six phases : CRISP-DM [19]

1) Business Understanding

This step aims to understand the objective of the project and the need from business point of view and to transform knowledge to the need of data mining. The details are as the following.

- **Determine business objectives**

Understanding the problems, the objectives and the business's need, then define the real need of business.

- **Assess situation**

Studying and analyzing the existing data, limitation, hypothesis that is possible, factors that effect to the project, and project plan.

- **Determine data mining goals**

Defining the goals for data mining by transforming the goal and business need into the goals for data mining techniques. For example, business goal is to increase the sale value by catalogue from the existing customers. We can transform the Data Mining goal to predict the amount of potential customers that want to buy the products from catalogue by using customer's buying history, customers and price of each item from the past 3 years.

- **Produce project plan**

Creating a project planning and development to accomplish as the purpose and time specified.

2) Data Understanding

Data Understanding is the step to understand where is the source of data and what are the component.

- Collect Initial Data by explaining where is the source of data such as internal or external organization.
- Data Description by explaining the data, such as the amount of data table, the amount of record in each table and field in the table.
- Explore Data by specifying the data that brought to analyze in order to accomplish the target.
- Verify Data Quality by checking the correctness and completeness of data that brought to analyze, such as Null, Duplicated Data and how to solve the problem.

3) Data Preparation

Data selection is the step for choosing fields and discussing the reasons. Randomly selecting records that used to build the model and records that used to test the model.

- Clean Data to manage the lost data, incorrect data and excessive data

- Construct Data, create new field from the old field for the efficient use and knowledge.
- Integrate Data, gathering data from many data table or from many fields to create a new data that more efficient.
- Format Data, transform data into the good format model.

4) Modeling

Analysis Techniques must be appropriate to our purpose and data property. If the data is not appropriated with our techniques, we need to adjust data, or if the selected techniques isn't suitable, we have to adjust those techniques too.

5) Evaluation

Determine the result from the model whether it relevant to the requirements of the system and knowledge from the model can be useful for the organization or not. If not, the whole processes would be performed again.

6) Deployment

If knowledge from the model meets the requirements, it must be illustrated in multi-dimensional views graphs that are easy to understand. Furthermore, there must have the monitoring and evaluation processes to determine the overall efficiency as well as the summary should be generated.

3.2 Customer Relationship Management (CRM)

In this section, the meaning of CRM architecture, the information analysis used for CRM and Data Mining in CRM will be described

3.2.1 The Meaning of CRM

CRM is the business approach studying the existing customer information so as to understand the customer behaviors. This approach contributes to the four benefits namely exploring new customers, maintaining existing customers, making customer satisfaction and increase revenue from the existing customers. [13]

CRM consists of two components presented in an equation as below:

CRM = customer understanding + relationship management [14]

The equation indicates that CRM is the combination of the Customer understanding and the Relationship Management. To manage the customer relationship, the analysis of the customer information is required. From this, the business will perceive the requirements of each customer. Consequently, the business can manage customer effectively.

3.2.2 CRM Architecture

CRM architecture is divided into three types the detail is shown in Figure 3.10.

1) Operational CRM System: is the front-end system used to manage the business activities regarding to the customer including sales and marketing or services such as customer contact information, price proposal, sale management, the gathering of customer behavior information and customer service system.

2) Analytical CRM System: is the part that responsible for collecting knowledge and analyzing customer information so as to understand the requirements of each customer. From this, the tool – Data Warehousing, Online Analytical Processing: OLAP and Data Mining- will be used to analyze the information.

3) Collaborative CRM System: the system assisting the interaction between the business and the customers in various channels such as private contact, letter, fax, telephone, website and email.

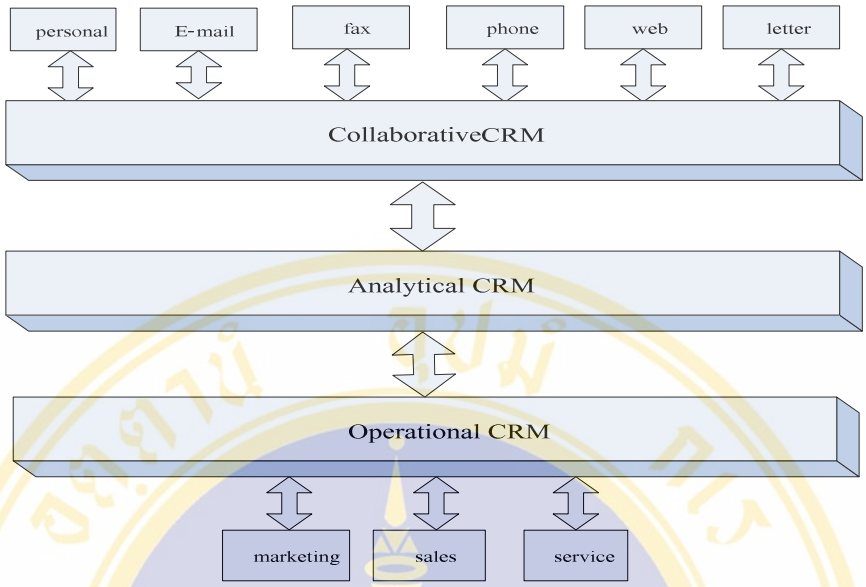


Figure 3.10 CRM architecture

3.2.3 Analytical Customer Relationship Management

Analytical CRM (ACRM) is the customer information management using the analytical technologies. This brings about the approach for developing and building the customer relationship described as follows [14]:

1) Customer Segmentation

Customer segmentation is the process to divide the customers into the small segments according to the similar characteristic or behavior. For example, the customers are clustered by the order amount. Then the products and services will be categorized to meet the customer requirements in each group [11][14][16][17].

Generally customer information is stored in the form of customer profile composing of three parts as follows [14]:

(1) Identity: is the unique information of the customer such as name and address.

(2) Characteristics: is information indicating the customer characteristics such as gender, age, status (single or married), education, number of children and revenue.

(3) Behavior: is the information demonstrating the customer behavior for instance mobile usage behavior – tend to use mobile in the evening and take approximately four hours using mobile in a week.

There are two forms of the customer segmentation divided by the customer profile showed as follows:

(1) Divide the customers using the identical characteristics and find out the common behavior in the group.

(2) Divide the customers using the similar behavior and determine the common characteristics.

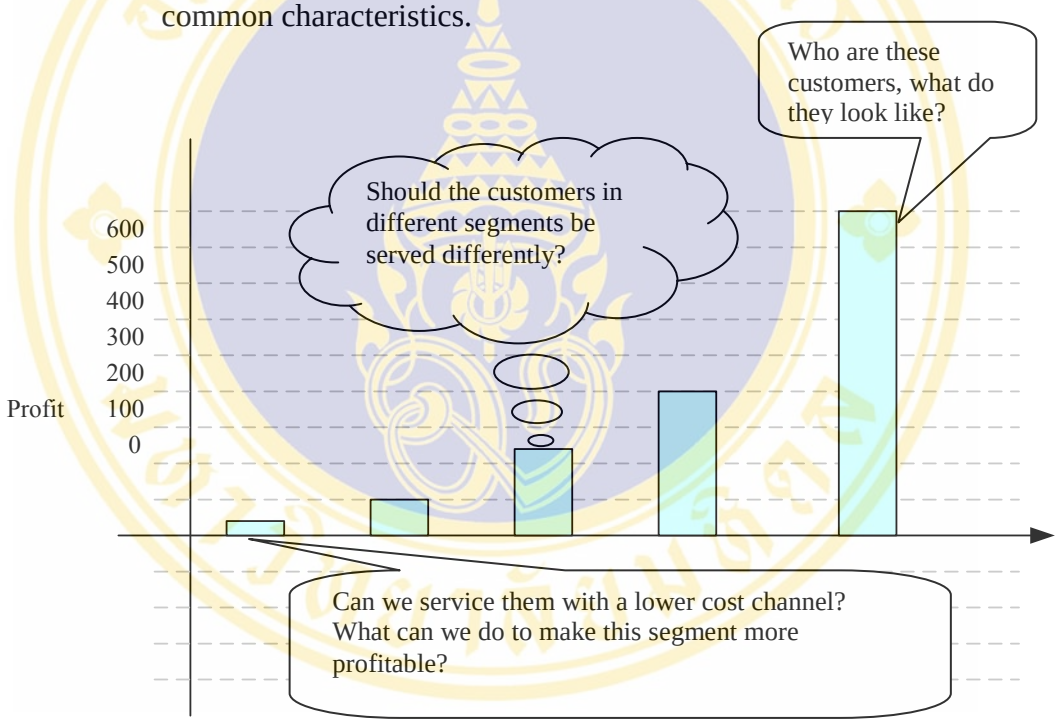


Figure 3.11 Segmentation of customers by profitability [14]

Figure 3.11 shows the customers group divided by the profit. There are many different ways to maintain each existing customer group. The figure above illustrates the example of questions used to ask each customer group to bring about the strategy to maintain the customer and make the good relations with the customers.

2) Customer Retentions

Customer retention is the customer preservation making the customer to have the faithfulness with the business as the cost for exploring the new customers is higher than the cost for maintaining the existing customers.

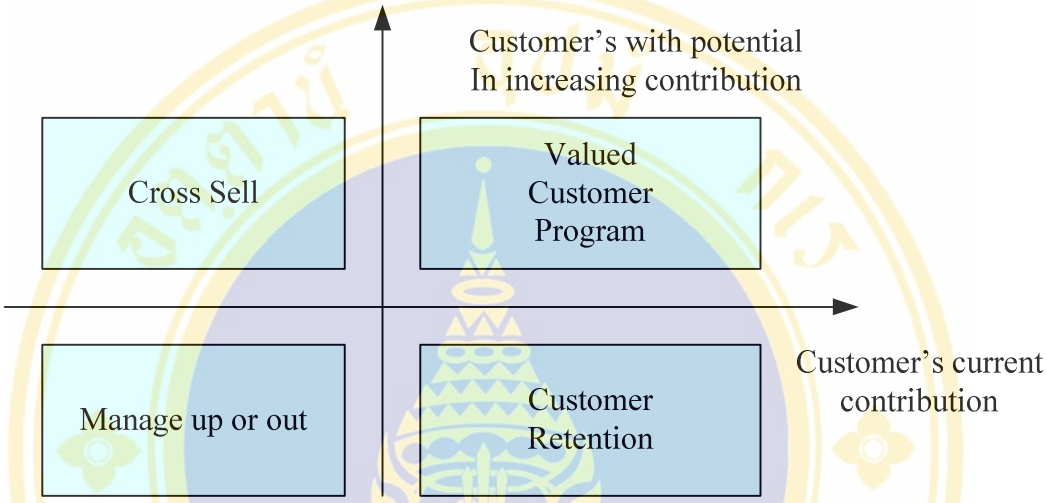


Figure 3.12 Treatment of various customer segments [14]

Figure 3.12 indicates that the business has the various perspectives regarding to the customer. The business considers the customer groups making the profits at the current time and in the future. The customer groups in the right side should be specified as the groups that should be maintained especially the groups in the top right should be strengthen the relationship as these groups are most likely to make the profit to the business. To reach the success of the customer retention, the business has to find out the opportunities to response the customers need with the products and services as soon as possible. For example, the customer goes to the ATM to get the money but the machine refutes to pay as there is not enough money in customer's account. The bank should offer the unsecured personal loans to credit-worthy customers in the next day. In practical, this strategy is the best practice and it can also enhance the customer's allegiance.

However, the customer retention is the means to analyze the customer segment, the analysis performed to build the customer royalty should be done individually in each customer not in group [14].

3) Customer Royalty

Customer royalty is the process performed after the target customer groups have been defined. The process could satisfy requirement and build satisfaction of each customer continually. Moreover the process can also identify the proper time to strengthen the relationship with the customers.

Figure 3.13 Lifetime impact of customer loyalty [14]

Figure 3.13 shows the concept of customers' tracking in order to specify the customers' activities during their lifetime. As the varieties of activities, companies have many chances to make good relationship with their customers. For example, sending a birthday card is one of the methods that has low cost, but high in efficiency.

3.2.4 Data Mining for CRM

In general, making a good relationship with customers, we will need to use the existing customer data that show each customer behavior such as type of purchased products, the amount of products, the number of times that customers purchase, and the last shopping period. This data will be analyzed to evaluate customer value and customers' requirement for the future and used for developing and improving the products and services. Data mining is one of the important analyzing tool. The work in data mining and CRM will be presented in the following part [12]:

1) Matching Campaigns to Customers

Data mining is used for choosing or defining campaigns that meet each customer requirement. In stead of offering all products to the customers, the company will only advertise products or services that meet each customer's requirement. By doing this, the companies save cost for their marketing.

2) Segmenting The Customer Base

Customer segmentation is important the data mining area. This method help the companies to maintain their relationship with customers by improving the quality of good and service to meet customers' requirement. The way of customers based on their behavior is called Clustering techniques. After clustering, the next step is classifying each customer into different groups, up to customer's basement data. In general, This method is called Classification model in data mining. This model classify data based on the specific characteristics of the customers.

3) Reducing Exposure to Credit Risk

In the real business, the company will reduce the customers that have high risk to reduce the opportunity of losing income. It is important to build model for forecasting the customer's behavior or make a prediction for customers who are likely to be in debt with the company (Prediction Model). The companies has high risk if they allow their customers to trial their good and service and let them pay such as telephone, electricity, and credit card.

4) Determining Customer Value

The evaluation of customer value and the profit is the outcome of net income minus expenses of advertising. Data Mining can be used for evaluating the profit from the customers in each time during the customers' lifetime that can be used to set the marketing plan.

5) Cross-selling, Up-selling, and Making Recommendations

Making relationship with the existing customer can help in increasing the profit by selling different products. The increase in selling is called "Cross-selling". This higher number in the net sale is called "up-selling". Data mining help the companies to analyze which and when the products and services should be offered in order to meet the customer's requirement, including adjusting their products and

services to suit their customers. Data mining technique that used for grouping the products for each customer is called “Association Rule”

3.3 Association Rule Mining

3.3.1 Principle of Association Rule Mining

Association Rule Mining, one of the data mining technique, is a kind of linking data. This principle is searching the relationship of data. This is to analyze and predict events when customers making their orders. The way to analyze is called “Market Basket Analysis”. The results will give knowledge to design catalog of products and launching products.

The recommended book system on Amazon web site is an example of this method. The large whole orders of the database on the Amazon web site will be processed to find out the relationship among data. The key is when customers buy any book they will buy another book also. For this reason, this relationship can help the company predict which book will be recommended for the second book and so on when making order. For instances, the customer who buying a database book will be likely to buy a data mining book as shown in the following picture.

Database -> data mining [support= 80% , confidence=60%]

Support 80% means the number of transactions that purchased database and data mining textbooks.

Confidence 60% means that 60% of the customers who purchased database also data mining textbooks.

In association rule mining process, there are 2 steps as follows:

1) Frequent itemsets generation. This step is to find itemsets that have frequency or supporting value greater than or equal to minimum supporting value.

2) Association rule generation. The itemsets from the first step will be used to create the association rule. Also, the association rule will be acceptable when

the rule must contain confidence value greater than or equal to the minimum confidence value

At first, frequent itemsets and supporting value will be calculated. After that, the frequent itemsets will be used to create the rule including finding out the confidence value for each rule. Consequently, the rule which has confidence value greater than or equal to the minimum confidence value will be only acceptable.

The following show the support value calculation:

$$\text{Support (A)} = \frac{\text{Transactions which consist of itemset A}}{\text{all transactions in the database}}$$

$$\text{Support (A} \cup \text{B)} = \frac{\text{Transactions which consist of itemset A and itemset B}}{\text{all transaction in the database}}$$

Then support (A) and support (A∪B) will be used to find confidence (A→B) value as shown in the following:

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\text{Support (A} \cup \text{B)}}{\text{Support (A)}}$$

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\text{transactions which consist of itemset A and itemset B}}{\text{transactions which consist of itemset A}}$$

3.3.2 Apriori Algorithm for Frequent Itemsets Generation

“Apriori is a basic algorithm for mining frequent itemsets for Boolean association rules. Apriori employs an iterative approach known as a level-wise search, where k-itemsets are used to explore (k + 1)-itemsets” [18].

Pseudo code : Apriori Algorithm [18]

Input: Database, D, of transactions; minimum support threshold, min_sup

Output: L, frequent itemsets in D.

Method:

L_1 = find_frequent 1-itemsets(D);

for (k:=2;$L_{k-1} \neq \varnothing$; k++) {

C_k = **apriori_gen**(L_{k-1}, *min_sup*);

for each transaction t $\in D$ {//scan D for counts

C_t = subset(C_k, t); //get the subsets of t that are candidates

for each candidate c $\in C_t$

c.count++;

}

$L_k = \{c \in C_k | c.count \geq min\_sup\}$

}

return L = $U_k L_k$;

procedure apriori_gen(L_{k-1} : frequent (k-1)-itemsets; min_sup: minimum support threshold)

for each itemset $l_1 \in L_{k-1}$

if ($l_1[1] = l_2[1] \wedge l_1[2] = l_2[2] \wedge \ldots \wedge (l_1[k-2] = l_2[k-2]) \wedge (l_1[k-1] = l_2[k-1])$ then {

c = l_1 join l_2; // join step: generate candidates

if has_infrequent_subset(c, L_{k-1}) then

delete c;// prune step: remove unfruitful candidate

else and c to C_k;

}

return C_k;

procedure has_infrequent_subset(c: candidate k-itemset; L_{k-1}: frequent(k-1)-itemsets);

// use prior knowledge

for each (k - 1)-subset s of c

if s $\notin$ L_{k-1} **then**

return TRUE;

return FALSE;

Figure 3.14 shows an example of apriori algorithm for ordering transactions (where minsup. Count=3)

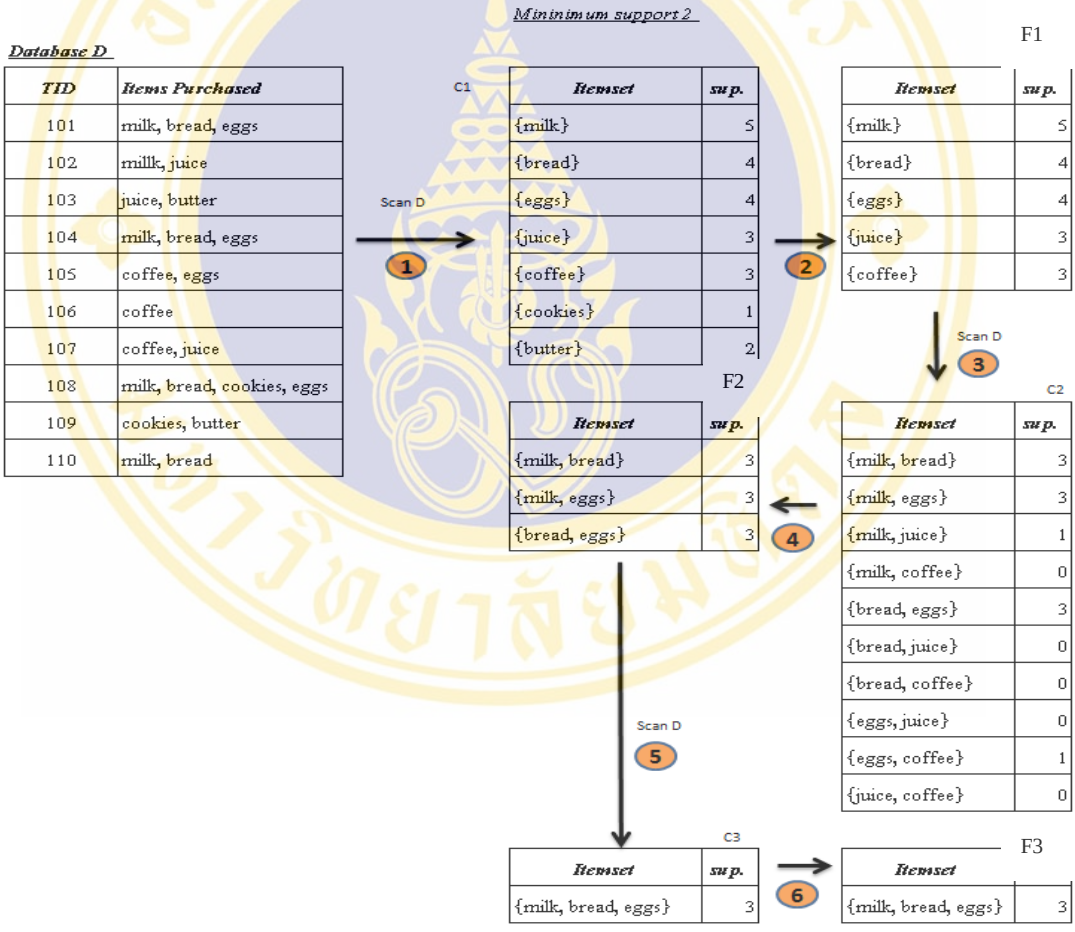


Figure 3.14 An example of Apriori Algorithm

3.3.3 Association Rule Generation

Association rule generation is the step to generate all association rules that has confidence value above minconf. [23].

Figure 3.15 shows the association rule generation algorithm.

```
Algorithm genRules(F)                    // F is the set of all frequent itemsets
1   for each frequent k-itemset f_k in F, k ≥ 2 do
2       output every 1-item consequent rule of f_k with confidence ≥ minconf and
            support ← f_k.count / n    // n is the total number of transactions in T
3       H_1 ← {consequents of all 1-item consequent rules derived from f_k above};
4       ap-genRules(f_k, H_1);
5   endfor

Procedure ap-genRules(f_k, H_m)          // H_m is the set of m-item consequents
1   if (k > m + 1) AND (H_m ≠ ∅) then
2      H_{m+1} ← candidate-gen(H_m);
3      for each h_{m+1} in H_{m+1} do
4          conf ← f_k.count / (f_k − h_{m+1}).count;
5          if (conf ≥ minconf) then
6              output the rule (f_k − h_{m+1}) → h_{m+1} with confidence = conf and
               support = f_k.count / n;    // n is the total number of transactions in T
7          else
8              delete h_{m+1} from H_{m+1};
9      endfor
10     ap-genRules(f_k, H_{m+1});
11  endif
```

Figure 3.15 The association rule generation algorithm [23]

For example of association rule generation: Suppose we use a frequent 3-itemset ({milk, bread, eggs}) to generate rules and define minconf. = 70%. The following lists all possible association rules with their support and confident values.

Rule 1: **milk , bread → eggs [confidence = 3/4 = 75%]**

Rule 2: **milk , eggs → bread [confidence=3/4 = 75%]**

Rule 3: bread, eggs → milk [confidence=3/5 = 60%]

Rule 4: **milk → bread, eggs [confidence=3/3 = 100%]**

Rule 5: **bread → milk, eggs [confidence=3/3 = 100%]**

Rule 6: **eggs → milk, bread [confidence=3/3 = 100%]**

There are five association rules, namely Rule 1, Rule 2, Rule 4, Rule 5 and Rule 6 that have confidence values greater than minconf. Therefore, these five rules are interesting for users.

CHAPTER IV
ACRM-VRT SYSTEM DESIGN

4.1 ACRM-VRT System Architecture

The ACRM-VRT system architecture is illustrated in Figure 4.1.

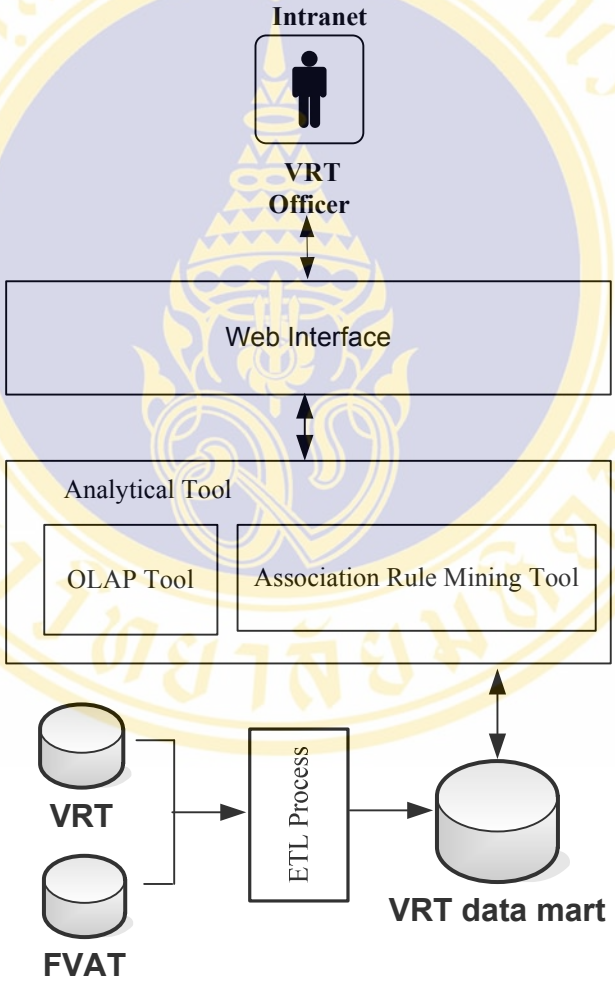


Figure 4.1 ACRM-VRT System architecture

Figure 4.1 ACRM-VRT architecture consists of 5 components as follows:

1) Data sources for building Data Mart include VRT and FVAT (Full Value Added Tax) Database

2) ETL process.

3) VRT data mart

4) Analytical Tool composed of:

- OLAP Tool is a tool for analyze multidimensional data online
- Association Rule Mining Tool is a tool for discovering the association rules representing tourist's goods purchase behaviors.

5) Web Interface is used to interact between ACRM-VRT system and users (VRT officers). The officers working at VRT office can access the system via Intranet system.

4.2 Data Sources of Data Mart

Source of Data Mart is the combination of data from two systems described as follows:

1) FVAT: Data is stored in the form of relational database in database management system named DB2 on mainframe. VAT registrant data and VAT filing data are stored in this system.

2) VRT: Data is stored in the form of relational database in DB2. VAT refunds for tourists, VRT entrepreneurs and VRT shops data are stored in this system.

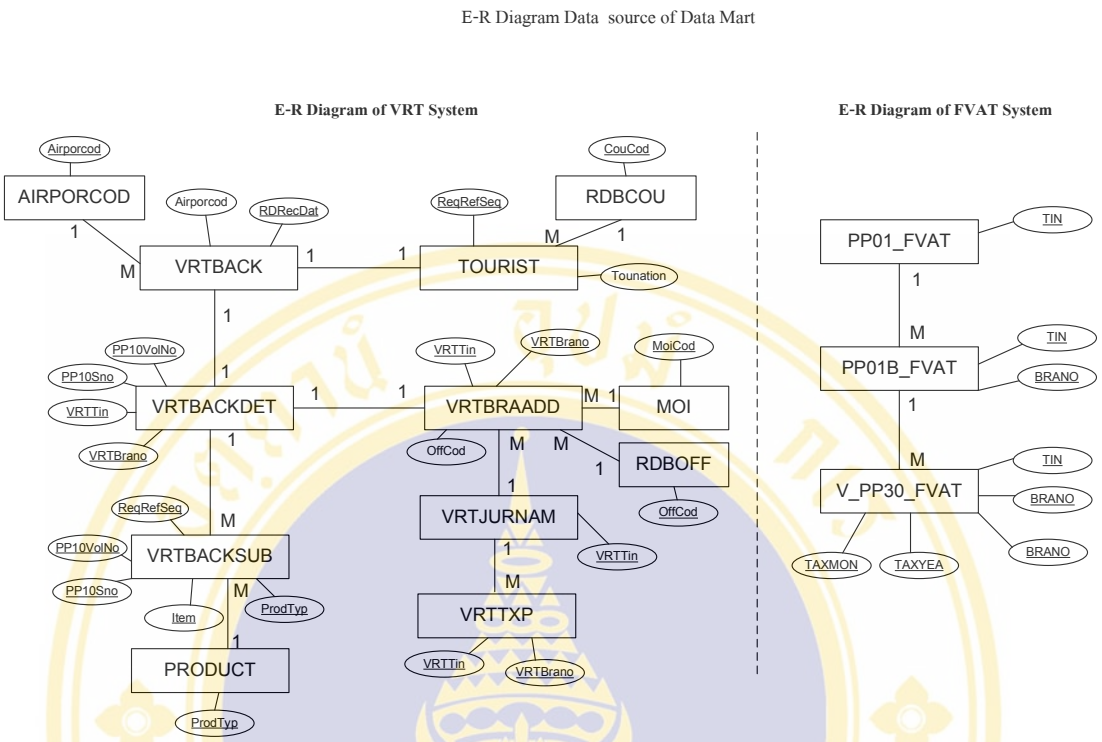


Figure 4.2 E-R diagram data source of Data Mart

Figure 4.2 shows E-R diagram of VAT registrant and VRT filing data. Only the core files used to build the Data Mart of the ACRM-VRT system are displayed in the figure. They consist of twelve tables from VRT system and three tables from FVAT system. The last updated of these data is June 30, 2009. (from January 1, 1999 to June 30, 2009. These tables are described as follows:

Table 4.1 The summary table of data sources for VRT data mart

No	Table Name	Description	Table Type / Database	Approx. Volume
1	TOURIST	Records of tourists information	Transaction / VRT	12,567,989 (≈20,000 / Month)
2	VRTBACK	Records of VAT Refund Information	Transaction / VRT	14,890,890 (≈20,000 /Month)

No	Table Name	Description	Table Type / Database	Approx. Volume
3	VRTBACKDET	Records of VAT Refund Detail	Transaction / VRT	14,897,997 (≈20,000 / Month)
4	VRTBACKSUB	Records of Product Detail	Transaction / VRT	1,151,162 (≈37,000 / Month)
5	VRTJURNAM	Records of VRT registrant Information	Master / VRT	1,036
6	VRTTXP	Records of application detail of VAT registrant shop	Master / VRT	4,212
7	VRTBRAADD	Records of address of business location in VRT system	Master / VRT	4,212
8	AIRPORCOD	Records of airport's reference code	Table file / VRT	9
9	RDBCOU	Records of country's reference code and nationality of the tourist	Table file / VRT	244
10	MOI	Records of tambol, amphur and province reference code	Table file / VRT	8,429
11	PRODUCT	Records of product type	Table file / VRT	35
12	RDBOFF	Records of regional revenue office code and area revenue office code	Table file / VRT	133
13	PP01_FVAT	Records of registration information of VAT registrant	Master / FVAT	369,945
14	PP01B_FVAT	Records of registration information of	Master / FVAT	447,808

No	Table Name	Description	Table Type / Database	Approx. Volume
		VAT registrant branch		
15	V_PP30_FVAT	Records of PP30 filing information	Transaction / FVAT	454,499,560 (≈350,000 / Month)

1. Table name : Tourist

   Description : Tourist Information.

   Data source : VRT System

No	Field Name	Data Type	Description
1	REQREFSEQ	CHAR(15)	Request reference sequence number
2	AIRPORCOD	NUMERIC(2)	International airport Code
3	REDRECDAT	DATE	Date for Revenue receive refund application form
4	PASSPORTNO	CHAR(25)	Tourist passport number
5	TOUTITNAM	VARCHAR(40)	Tourist title name
6	TOUFIRNAM	VARCHAR(80)	Tourist first name
7	TOULASNAM	VARCHAR(80)	Tourist last name
8	NATION	CHAR(5)	Tourists nationality
9	DATIN	DATE	Arrival date
10	DATOUT	DATE	Departure date
11	ADDRESSNO	VARCHAR(30)	Address number
12	STREET	VARCHAR(40)	Street name
13	DISTRICT	VARCHAR(40)	District name
14	PROVINCE	VARCHAR(40)	Province name
15	STATE	VARCHAR(40)	State name
16	COUNTRY	VARCHAR(40)	Country name
17	ZIPCODE	VARCHAR(5)	Zip code
18	UPDDATE	TIMESTAMP	Update time stamp
19	USERID	CHAR(8)	User ID

2. Table name : VRTBACK

Description : VAT Refund Information.

Data source : VRT System

No	Field Name	Data Type	Description
1	REQREFSEQ	CHAR(15)	Request reference sequence number
2	AIRPORCOD	NUMERIC(2)	International airport code
3	RDRECDAT	DATE	Date for Revenue receive refund application form
4	OFFICERID	CHAR(8)	Revenue officer ID
5	SHIFTNO	CHAR(1)	Shift number
6	TERMNO	CHAR(2)	Terminal number
7	REFPAIDAT	DATE	VAT refund payment date
8	REQREFCOD	CHAR(1)	Request refund type code
9	REFPAICOD	CHAR(1)	VAT Refund payment code
10	REQPURAMO	NUMERIC(9,2)	Tourist purchase amount
11	REQVATAMO	NUMERIC(9,2)	VAT refund amount
12	TOTPURAMO	NUMERIC(9,2)	Total purchase computed amount
13	CALVATAMO	NUMERIC(9,2)	Calculate VAT amount
14	AUDPURAMO	NUMERIC(9,2)	Audit purchase amount
15	AUDVATAMO	NUMERIC(9,2)	Audit VAT amount
16	REFFEEAMO	NUMERIC(9,2)	VAT refund fee amount
17	DRAFTNO	VARCHAR(50)	Draft number
18	DRAFTEXP	NUMERIC(9,2)	Expenses of Draft
19	CRCARACC	VARCHAR(50)	Credit card number
20	CREXPDATE	DATE	Expire date of credit card
21	CRCARCOP	VARCHAR(50)	Create card corporation
22	TRANFEXP	NUMERIC(9,2)	Transfer expenses
23	ACCSTATUS	CHAR(1)	Account status
24	REFPAIAMO	NUMERIC(9,2)	Amount of VAT Refund payment
25	OPERUSER	CHAR(8)	Operator user

No	Field Name	Data Type	Description
26	OPERSTATUS	CHAR(1)	Refund status

3. Table name : VRTBACKDET

Description : VAT Refund Detail.

Data source : VRT System

No	Field Name	Data Type	Description
1	REQREFSEQ	CHAR(15)	Request refund type code
2	PP10VOLNO	NUMERIC(7)	P.P.10 Volume Number
3	PP10SNO	NUMERIC(7)	P.P.10 Serial Number
4	VRTTIN	CHAR(10)	VRT Taxpayer identification number
5	VRTBRANO	NUMERIC(4)	VRT branch number
6	REQVATAMO	NUMERIC(9,2)	VAT refund amount
7	TOTPURAMO	NUMERIC(9,2)	Total purchase compute amount
8	CALVATAMO	NUMERIC(9,2)	Calculate VAT amount
9	AUDPURAMO	NUMERIC(9,2)	Audit purchase amount
10	AUDVATAMO	NUMERIC(9,2)	Audit VAT amount
11	MGRCOMD	CHAR(1)	Manager command code
12	REFPAIAMO	NUMERIC(9,2)	Amount of VAT Refund payment
13	PP10MIS	CHAR(1)	P.P. 10 not found in database
14	DENYCOD	CHAR(3)	Deny Code
15	ITEMNUM	NUMERIC(3)	Amount of item goods
16	UPDDATE	TIMESTAMP	Update time stamp
17	USERID	CHAR(8)	User id
18	PP10USTD	CHAR(1)	Incorrect P.P.10 form code

4. Table name : VRTBACKSUB

Description : Product Detail.

Data source : VRT System

No	Field Name	Data Type	Description
1	REQREFSEQ	CHAR(15)	Request VAT Number
2	PP10VOLNO	NUMERIC(7)	P.P. 10 Volume Number
3	PP10SNO	NUMERIC(7)	P.P. 10 Serial Number
4	ITEM	NUMERIC(1)	Goods Item
5	PRODTYP	CHAR(6)	Goods Code
6	PURAMO	NUMERIC(9,2)	Purchase Amount
7	VATAMO	NUMERIC(9,2)	VAT Amount
8	CUSCHK	CHAR(1)	Custom checking goods code
9	RDCHK	CHAR(1)	Revenue officer checking goods code

5. Table name : VRTJURNAM

Description : VRT Registrant Information.

Data source : VRT System

No	Field Name	Data Type	Description
1	VRTTIN	CHAR(10)	VRT Taxpayer identification number
2	LASTDATE	DATE	Last Date of VRT System
3	FIRSTDATE	DATE	First Date of VRT System
4	TITCOD	CHAR(8)	VRT taxpayer title name
5	VRTNAMTHA	VARCHAR(160)	VRT Taxpayer Thai name
6	VRTNAMENG	VARCHAR(160)	VRT Taxpayer English name
7	VRTPIN	CHAR(13)	VRT Personal identification number
8	VATDATE	DATE	VAT registration date

No	Field Name	Data Type	Description
9	JURDATE	DATE	Corporation registration date
10	PAIDCAPITAL	NUMERIC(6,2)	Paid capital amount

6. Table name : VRTTXP

Description : Application detail of VAT registrant shop.

Data source : VRT System

No	Field Name	Data Type	Description
1	VRTTIN	CHAR(10)	VRT Taxpayer identification number
2	VRTBRANO	NUMERIC(4)	VRT Branch Number
3	LASTDATE	DATE	Last Date of VRT System
4	FIRSTDATE	DATE	First Date of VRT System
5	CANCOD	CHAR(2)	VAT cancel code
6	DISDATE	DATE	VAT disinheritance code
7	DISDOCNO	VARCHAR(15)	VAT disinherit document number
8	DISDOCNODAE	DATE	VAT disinherit document date
9	ISIC	NUMERIC(1)	International Standard Industrial Classification code
10	CANCODVRT	CHAR(2)	VRT cancel code
11	DISDATEVRT	DATE	VRT disinherit date
12	DISDOCNOVRT	VARCHAR(15)	VAT disinherit document number
13	DISDOCNODATE VRT	DATE	VAT disinherit document date
14	ISICVRT	NUMERIC(1)	International Standard Industrial Classification Code in VRT System
15	GOODSTYP	CHAR(2)	Goods type code

7. Table name : VRTBRAADD

Description : Address of Business Location in VRT System

Data source : VRT System

No	Field Name	Data Type	Description
1	VRTTIN	CHAR(10)	VRT Taxpayer identification number
2	VRTBRANO	NUMERIC(4)	Branch number
3	LASTDATE	DATE	Last date of VRT system
4	FIRSTDATE	DATE	First date of VRT system
5	TITCOD	CHAR(8)	VRT title code
6	BRANAM	VARCHAR(160)	VRT shop Thai name
7	BRANAM_ENG	VARCHAR(160)	VRT shop English name
8	BLDGNAME	VARCHAR(40)	Building name
9	VILLAGE	VARCHAR(40)	Village name
10	ROOMNO	VARCHAR(20)	Room number
11	FLOORNO	VARCHAR(20)	Floor number
12	ADDNO	VARCHAR(30)	Address number
13	MOONO	VARCHAR(2)	Moo number
14	SOINAM	VARCHAR(40)	Soi name
15	THNNAM	VARCHAR(40)	Street name
16	TAMCOD	CHAR(6)	Sub district code
17	AMPCOD	CHAR(6)	District code
18	PROVCOD	CHAR(6)	Province code
19	POSCOD	VARCHAR(5)	Zip code
20	TELNO	VARCHAR(15)	Telephone number
21	FAXNO	VARCHAR(15)	FAX number
22	OFFCOD	CHAR(8)	Revenue office code
23	EMAIL	VARCHAR(30)	E-Mail address

8. Table name : AIRPORCOD

Description : Airport's Reference Code

Data source : VRT System

No	Field Name	Data Type	Description
1	AIRPORCOD	DECIMAL(2)	International airport code
2	AIRPORCODTXT	VARCHAR(100)	International airport Thai name
3	AIRPORCODENG	VARCHAR(100)	International airport English name
4	AIRPORCODSHT	VARCHAR(3)	Shot name of International airport
5	OFFCODE	CHARACTER(8)	Revenue office code

9. Table name : RDBCOU

Description : Country's Reference Code and Nationality of The Tourist.

Data source : VRT System

No	Field Name	Data Type	Description
1	COUCOD	CHARACTER(5)	Country code
2	COUNAM	VARCHAR(50)	Country name
3	NATNAM	VARCHAR(50)	Nationality name
4	CONTINENT	CHARACTER(1)	Continent code

10. Table name : MOI

Description : Tambol, Amphur and Province Reference Code.

Data source : VRT System

No	Field Name	Data Type	Description
1	MOICOD	CHARACTER(6)	Code for sub-district, district, province
2	MOINAME	VARCHAR(50)	Thai name for sub-district, district, province
3	MOINAMEENG	VARCHAR(50)	English name for sub-

No	Field Name	Data Type	Description
			district, district, province
4	OFFCODE	CHARACTER(8)	Revenue office code

11. Table name : PRODUCT
Description : Product Type
Data source : VRT System

No	Field Name	Data Type	Description
1	PRODTYP	CHARACTER(4)	Goods type code
2	PRODTXT	VARCHAR(100)	Goods Thai name
3	PRODENG	VARCHAR(100)	Goods English name
4	REFUNDTYP	CHARACTER(1)	R=Refund , D=Deny

12. Table name : RDBOFF
Description : Regional Revenue Office Code and Area Revenue Office Code
Data source : VRT System

No	Field Name	Data Type	Description
1	OFFCOD	CHARACTER(8)	Revenue office code
2	OFFNAME	VARCHAR(50)	Revenue office name
3	SUPOFFCOD	CHARACTER(8)	Revenue supper office code

13. Table name : PP01_FVAT
Description : Registration Information of VAT Registrant.
Data source : FVAT System

No	Field Name	Data Type	Description
1	TIN	CHAR(13)	Taxpayer identification number
2	TITNAM	VARCHAR(40)	Taxpayer title name
3	NAME	VARCHAR(160)	Taxpayer name

No	Field Name	Data Type	Description
4	SURNAME	VARCHAR(80)	Taxpayer surname
5	BLDGNAME	VARCHAR(40)	Building name
6	VILLAGE	VARCHAR(40)	Village name
7	ROOMNO	VARCHAR(20)	Room number
8	FLOORNO	VARCHAR(20)	Floor number
9	ADDNO	VARCHAR(30)	Address number
10	MOONO	VARCHAR(2)	Moo number
11	SOINAM	VARCHAR(40)	Soi name
12	THNNAM	VARCHAR(40)	Street name
13	TAMCOD	CHAR(6)	Sub-district code
14	AMPCOD	CHAR(6)	District code
15	PROVCOD	CHAR(6)	Province code
16	POSCOD	VARCHAR(5)	Post code
17	TELNO	VARCHAR(15)	Telephone number
18	FORRECDAT	DATE	Form received date
19	APRDAT	DATE	Approved date
20	FIRDATTYP	CHAR(1)	First date type
21	BUSFIRDAT	DATE	Business first date
22	PIN	CHAR(15)	Personal identification number
23	CORREGNO	CHAR(15)	Corporate registration number
24	CORREGDAT	DATE	Corporate registration date
25	BUSCPTAMO	DECIMAL(15,2)	Paid capital amount
26	ESTMONINCAMO	DECIMAL(15,2)	Estimate income amount per month
27	BUSTYPCOD1	CHAR(1)	Business type code 1
28	GOOTYPCOD1	CHAR(6)	Goods type code 1
29	BUSTYPCOD2	CHAR(1)	Business type code 2

No	Field Name	Data Type	Description
30	GOOTYPCOD2	CHAR(6)	Goods type code 2
31	BUSTYPCOD3	CHAR(1)	Business type code 3
32	GOOTYPCOD3	CHAR(6)	Goods type code 3
33	VATCOD	CHAR(1)	VAT rate code
34	OFFCOD	CHAR(8)	Area Revenue branch office code
35	OFFNM	CHAR(45)	Area Revenue branch office name
36	SUPNM	CHAR(45)	Area Revenue Office name
37	TAXGOODTYPE	CHAR(1)	Good Taxpayer type
38	JFBSTA	CHAR(1)	Joint filling Status
39	JFBNO	DECIMAL(5)	Joint filling branch number (P.P.30)
40	LTOSTA	CHAR(1)	Large Business Taxpayer status
41	TAMCOD	CHAR(1)	Sub-district code
42	AMPCOD	CHAR(6)	District code
43	PROVCOD	CHAR(6)	Province code

14. Table name : PP01B_FVAT

Description : Registration information of VAT Registrant Branch

Data source : FVAT System

No	Field Name	Data Type	Description
1	TIN	CHAR(13)	Taxpayer identification number
2	BRANO	DECIMAL(5,0)	Branch number
3	TITNAM	VARCHAR(40)	VAT Taxpayer title name
4	BRANAM	VARCHAR(160)	VAT Taxpayer branch name
5	BLDGNAME	VARCHAR(40)	Building name
6	ROOMNO	VARCHAR(20)	Room number

No	Field Name	Data Type	Description
7	FLOORNO	VARCHAR(20)	Floor number
8	ADDNO	VARCHAR(30)	Address number
9	MOONO	VARCHAR(2)	Moo number
10	SOINAM	VARCHAR(40)	Soi name
11	THNNAM	VARCHAR(40)	Street name
12	TAMCOD	CHAR(6)	Sub-district code
13	AMPCOD	CHAR(6)	District code
14	PROVCOD	CHAR(6)	Province code
15	POSCOD	VARCHAR(5)	Zip code
16	TELNO	VARCHAR(15)	Telephone number
17	FORRECDAT	DATE	Received form date
18	APRDAT	DATE	Approved date
19	BUSFIRDAT	DATE	Business first date
20	OFFCOD	CHAR(8)	Area Revenue branch office code
21	OFFNM	CHAR(45)	Area Revenue branch office name
22	SUPNM	CHAR(45)	Area Revenue office code
23	TAMCOD	CHAR(1)	Sub-district code
24	AMPCOD	CHAR(6)	District code
25	PROVCOD	CHAR(6)	Province code

15. Table name : V_PP30_FVAT

Description : PP30 Filing Information.

Data source : FVAT System

No	Field Name	Data Type	Description
1	TIN	CHAR(13)	Taxpayer identification number
2	BRANO	DECIMAL(5,0)	Branch number

No	Field Name	Data Type	Description
3	TAXYEA	DECIMAL(4,0)	Tax year
4	TAXMON	DECIMAL(2,0)	Tax month
5	SLEAMO	CHAR(8)	Sales amount this month
6	SLETAXAMO	DECIMAL(15,2)	This month's net sales tax
7	PURAMO	CHAR(1)	Purchase amount that is entitled to deduction of input tax from output tax in this month's computation
8	PURTAXAMO	DECIMAL(15,2)	This month's net purchase tax
9	PABTAXAMO	DECIMAL(15,2)	This month's tax payable
10	REBTAXAMO	DECIMAL(15,2)	This month's excess tax payable
11	REGTST	TIMESTAMP	Record register time stamp

4.3 ETL Process

Processes of importing data to Data mart are illustrated in Figure 4.3.

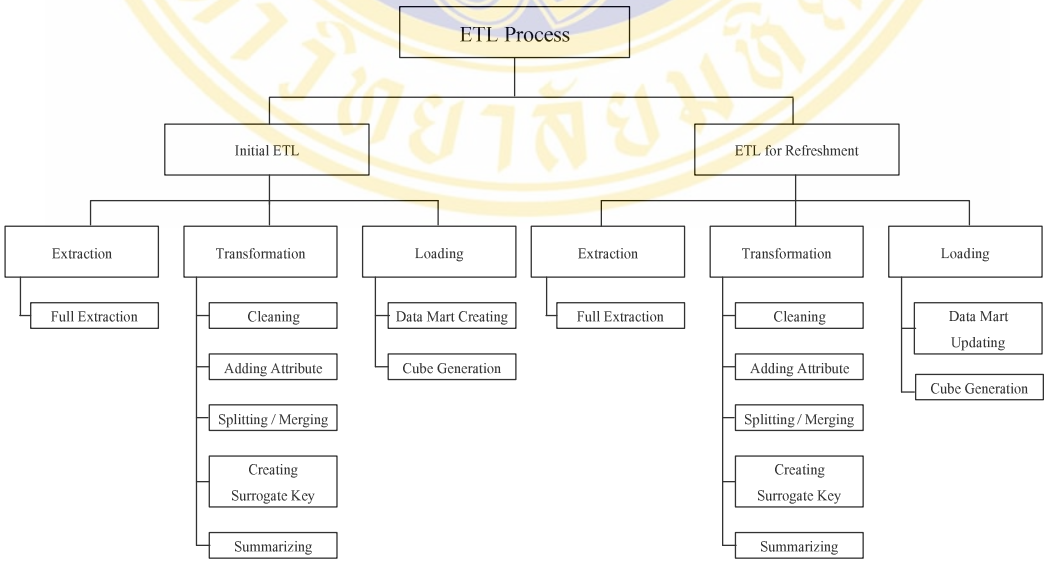


Figure 4.3 Structure Chart of ETL Process

Figure 4.3 shows the ETL's process for preparing and importing data to Data Mart. These processes are described as follows:

1) Initial ETL: is the initial process in building Data Mart and Data Cube.

2) ETL for Refreshment: is a process to update the Data Mart for ACRM-VRT system. Data will be updated monthly and be used to create monthly reports.

Initial ETL and ETL for Refreshment is the combination of these processes described as follows:

- **Extraction**

Full Extraction Steps to select data from data sources.

- **Transformation**

Cleaning Steps to update data to be consistent and in the defined forms.

Add Attribute Steps to add data field to be easy to integration and retrieve data.

Splitting / Merge Step to split data such as split purchase date in date format into separated date, month and year and merge data such as sum the total orders in each day.

Creating Surrogate Key Step to define the Primary Key in Dimension Table which is the integer type generated automatically. This key is used to link to Fact Table. Consequently, the size of the Fact Table would not be large.

Summarizing Step to summarize data for Fact Table Fact Table

- **Loading**

Data Mart Creation Step to import transformed data to Dimension Tables and Fact Tables according to the defined structure.

Data Mart Updating Step to update data in Data Mart to be up-to-dated by adding data generated in the new month.

Cube Generation Step to import and update Data Cube.

4.4 Data Mart Design

The design of a data mart is divided into three parts including data granularity declaration, concept hierarchies definition, and structure of data mart.

4.4.1 Data Granularity Declaration

There are two levels of Data Granularity namely day and month described as follows:

1) Day Level Granularity: Store daily information, including VAT refunds for tourists data. This data are used to display the amount of tourists that come to request for VAT refund in each day. From this, the VRT officer can be arranged for providing service to the tourists.

2) Month Level Granularity : Store monthly information, including VRT filing data.

4.4.2 Concept Hierarchies Definition

Showing the data perspectives used to retrieve data are defined as follows:

1) RD Office : Show area of responsibility of each Revenue Office that the business's stores are located.

2) Country : Show the country of tourists requesting for VAT refund.

3) Province : Show the business location according to the map of Thailand.

4) Pay VAT Time : Show month and year that payment occurs.

5) Refund VAT Time : Show year month day of requesting for VAT refund.

6) Refund Type : Show refund type of the tourist.

Table 4.2 Show concept hierarchy design

Dimension	Level
1. RD Office	Regional --> Area --> Branch
2. Country	Continent -- > Country
3. Province	Province -- > District -- > Sub district
4. Pay VAT Time	Year --> Semester -- > Quarter --> Month Fiscal Year -- > Fiscal Semester --> Fiscal Quarter --> Fiscal Month

Dimension	Level
5. Refund VAT Time	Year --> Semester -- > Quarter --> Month -- > Day Fiscal Year -- > Fiscal Semester --> Fiscal Quarter --> Fiscal Month
6. Refund Type	Refund Group -- > Refund Name

4.4.3 Data Mart Structure

Data Mart structure is depicted in Figure 4.4. The structure composes of 17 dimension tables and 4 fact tables described in Table 4.3

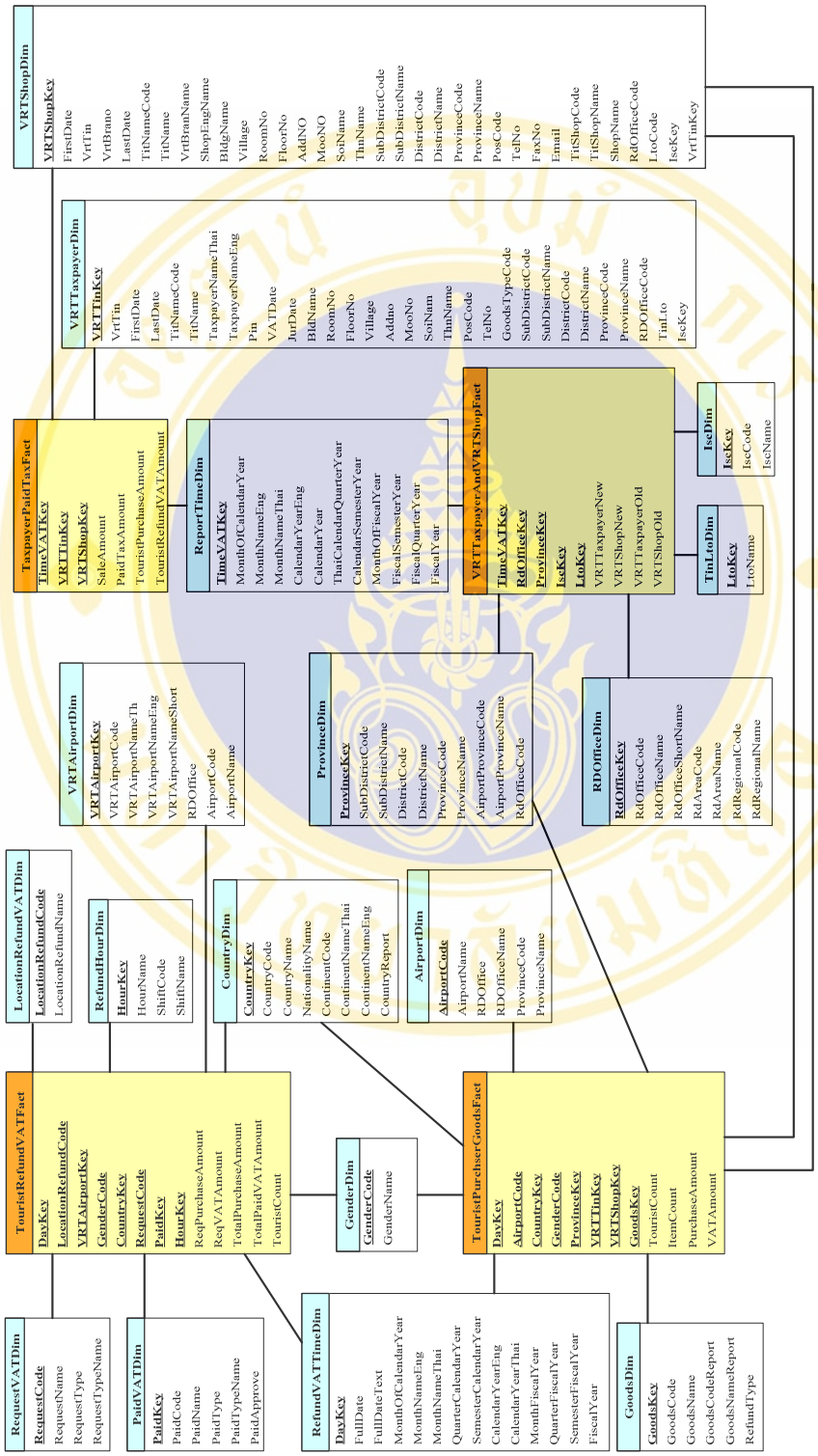


Figure 4.4 Data mart structure

Table 4.3 Tables in data mart

No	Table Name	Description	Type
1	TaxpayerPaidTaxFact	Records of summarized submitted VAT30 form data (VAT Taxpayer in VRT system)	Fact
2	TouristPurchserGoodsFact	Records of summarized tourists purchase goods data	Fact
3	TouristRefundVATFact	Records of summarized tourists refund VAT data	Fact
4	VRTTaxpayerAndVRTShopFact	Records of summarized VRT Taxpayer and VRT Shop data	Fact
5	AirportDim	Records of International airport	Dimension
6	CountryDim	Records of country	Dimension
7	GenderDim	Records of gender	Dimension
8	GoodsDim	Records of goods type	Dimension
9	IscDim	Records of International Standard Industrial Classification Code	Dimension
10	LocationRefundVATDim	Records of location refund VAT	Dimension
11	PaidVATDim	Records of paid VAT	Dimension
12	ProvinceDim	Records of province	Dimension
13	RDOfficeDim	Records of Revenue office	Dimension
14	RefundHourDim	Records of tourists refund VAT per hour	Dimension
15	RefundVATTimeDim	Records of tourists refund VAT date	Dimension
16	ReportTimeDim	Records of process report time	Dimension

No	Table Name	Description	Type
17	RequestVATDim	Records of tourist request VAT type	Dimension
18	TinLtoDim	Records of large Business Taxpayer	Dimension
19	VRTAirportDim	Records of International airport	Dimension
20	VRTShopDim	Records of VRT shop	Dimension
21	VRTTaxpayerDim	Records of VRT Taxpayer	Dimension

1. Table Name : TaxpayerPaidTaxFact
   Description : Filing Information of VAT Registrant in VRT System
   Table Type : Fact Table

No	Field Name	Description	Type	Size	Key
1	TimeVATKey	Time key	smallint	2	PK
2	VRTTinKey	VRT Taxpayer identification number key	smallint	2	PK
3	VRTShopKey	VRT shop key	smallint	2	PK
4	SaleAmount	Amount of sales goods from P.P.30 form	float	4	
5	PaidTaxAmount	Amount of tax payment	float	4	
6	TouristPurchaseAmount	Amount of tourist purchase goods	float	4	
7	TouristRefundVATAmount	Amount of tourist refund VAT	float	4	

2. Table Name : TouristPurchserGoodsFact

Description : Purchasing Transaction of the Tourist

Table Type : Fact

No	Field Name	Description	Type	Size	Key
1	DayKey	Day key	int	4	PK
2	AirportCode	International airport code	char	2	PK
3	CountryKey	Country code	smallint	2	PK
4	GenderCode	Gender code	char	1	PK
5	ProvinceKey	Province key	smallint	2	PK
6	VRTTinKey	VRT Taxpayer identification number	smallint	2	PK
7	VRTShopKey	VRT shop key	smallint	2	PK
8	GoodsKey	Goods key	tinyint	1	PK
9	TouristCount	Total of Tourist	float	4	
10	ItemCount	Total of purchase items	float	4	
11	PurchaseAmount	Amount of purchase goods	float	4	
12	VATAmount	Amount of VAT	float	4	

3. Table Name : TouristRefundVATFact

Description : Request for VAT Refund Transaction of the Tourist.

Table Type : Fact

No	Field Name	Description	Type	Size	Key
1	DayKey	Day key	int	4	PK
2	LocationRefundCode	VAT refund location code	char	1	PK
3	VRTAirportKey	International airport of VRT system	tinyint	1	PK
4	GenderCode	Gender code	char	1	PK
5	CountryKey	Country key	smallint	2	PK

No	Field Name	Description	Type	Size	Key
6	RequestCode	VAT Refund code	char	1	PK
7	PaidKey	Refund payment key	tinyint	1	PK
8	HourKey	Hour key	tinyint	1	PK
9	ReqPurchaseAmount	Amount of tourists purchase goods	float	4	
10	ReqVATAmount	Amount of tourists refund VAT	float	4	
11	TotalPurchaseAmount	Amount of VAT refundable	float	4	
12	TotalPaidVATAmount	Amount of VAT payment for tourists	float	4	
13	TouristCount	Total of tourists	float	4	

4. Table Name : VRTTaxpayerAndVRTShopFact
   Description : Number of VAT Registrant and Business's Shop in VRT System
   Table Type : Fact Table

No	Field Name	Description	Type	Size	Key
1	TimeVATKey	Key of time	smallint	2	PK
2	RdOfficeKey	Revenue office code	smallint	2	PK
3	ProvinceKey	Province code	smallint	2	PK
4	IscKey	International Standard Industrial Classification key	smallint	2	PK
5	LtoKey	Large Business taxpayer code	char	1	PK
6	VRTTaxpayerNew	Total of new VRT Taxpayer	int	4	
7	VRTShopNew	Total of new VRT	int	4	

No	Field Name	Description	Type	Size	Key
		shops			
8	VRTTaxpayerOld	Total of old VRT taxpayer	int	4	
9	VRTShopOld	Total of old VRT shops	int	4	

5. Table Name: AirportDim
   Description : Airport Code
   Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	AirportCode	Airport code	char	2	PK
2	AirportName	Airport name	nvarchar	50	
3	RDOffice	Revenue office code return VAT for tourists	char	8	
4	RDOfficeName	Revenue office name return VAT for tourists	nvarchar	50	
5	ProvinceCode	Province code of international airport	char	6	
6	ProvinceName	Province name of international airport	nvarchar	50	

6. Table Name: CountryDim
   Description : Country Code
   Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	CountryKey	Country key	smallint	2	PK
2	CountryCode	Country code	char	5	
3	CountryName	Country name	nvarchar	150	
4	NationalityName	Nationality name	nvarchar	150	

No	Field Name	Description	Type	Size	Key
5	ContinentCode	Continent code	char	1	
6	ContinentNameThai	Continent Thai name	nvarchar	50	
7	ContinentNameEng	Continent English name	nvarchar	50	
8	CountryReport	Country name for report	nvarchar	150	

7. Table Name: GenderDim
   Description : Gender Code
   Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	GenderCode	Gender code	char	1	PK
2	GenderName	Gender name	char	10	

8. Table Name: GoodsDim
   Description : Product Type Code
   Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	GoodsKey	Goods key	tinyint	1	PK
2	GoodsCode	Goods code	char	4	
3	GoodsName	Goods name	nvarchar	150	
4	GoodsCodeReport	Goods code for report	char	2	
5	GoodsNameReport	Goods name for report	nvarchar	150	
6	RefundType	VAT refund type name (D=Deny , R=Refund)	char	1	

9. Table Name: IscDim
   Description : Business Type
   Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	IscKey	International Standard Industrial Classification key	smallint	2	PK
2	IscCode	International Standard Industrial Classification Code	char	6	
3	IscName	International Standard Industrial Classification name	nvarchar	150	

10. Table Name: LocationRefundVATDim
    Description : VAT Refund Location
    Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	LocationRefundCode	Location submitted to claim for the VAT refund code	char	1	PK
2	LocationRefundName	Location submitted to claim for the VAT refund （B=Box at the international airport，C=Counter at the international airport，P=Postal service）	nvarchar	50	

11. Table Name: PaidVATDim

Description : Refund Type

Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	PaidKey	VAT refund payment key	tinyint	1	PK
2	PaidCode	Paid refund tax code (C=Credit , M=Money , N=Deny , D= Draft , U = on process)	char	1	
3	PaidName	Paid refund tax name	nvarchar	50	
4	PaidType	Paid VAT type (C=Credit , M=Money, N=Deny)	char	1	
5	PaidTypeName	Paid VAT type name	nvarchar	50	
6	PaidApprove	Approved / not Approved paid VAT	nvarchar	50	

12. Table Name: ProvinceDim

Description : Province Information.

Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	ProvinceKey	Key of Province	smallint	2	PK
2	SubDistrictCode	Sub-district code	char	6	
3	SubDistrictName	Sub-district code	nvarchar	50	
4	DistrictCode	District code	char	6	
5	DistrictName	District name	nvarchar	50	
6	ProvinceCode	Province code	char	6	
7	ProvinceName	Province name	nvarchar	50	

No	Field Name	Description	Type	Size	Key
8	AirportProvinceCode	Province code have international airport	char	2	
9	AirportProvinceName	Province name have international airport	nvarchar	50	
10	RdOfficeCode	Revenue office code	char	8	

13. Table Name: RDOfficeDim
Description : RD's office.
Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	RdOfficeKey	Revenue office code	smallint	2	PK
2	RdOfficeCode	Area Revenue branch office code	char	8	
3	RdOfficeName	Area Revenue branch office name	nvarchar	100	
4	RdOfficeShortName	Area Revenue branch office short name	nvarchar	50	
5	RdAreaCode	Area Revenue office code	char	8	
6	RdAreaName	Area Revenue office name	nvarchar	20	
7	RdRegionalCode	Region Revenue office code	char	8	
8	RdRegionalName	Region Revenue office name	nvarchar	15	

14. Table Name: RefundHourDim
    Description : Number of Refund Hour at the Airport
    Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	HourKey	Key of Hour	tinyint	1	PK
2	HourName	Hour per day	nvarchar	50	
3	ShiftCode	Shift code at international airport	char	1	
4	ShiftName	Shift name at international airport	char	10	

15. Table Name: RefundVATTimeDim
    Description : Request Date Information
    Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	DayKey	Key of day	int	4	PK
2	FullDate	Calendar date in date format	datetime	4	
3	FullDateText	Calendar date in character format	char	10	
4	MonthOfCalendarYear	Month of calendar year	nvarchar	15	
5	MonthNameEng	English month name	char	10	
6	MonthNameThai	Thai month name	char	10	
7	QuarterCalendarYear	Calendar quarter	char	7	
8	SemesterCalendarYear	Calendar semester	char	7	
9	CalendarYearEng	Calendar year (Eng.)	smallint	2	
10	CalendarYearThai	Calendar year (Thai)	smallint	2	
11	MonthFiscalYear	Fiscal month	char	7	
12	QuarterFiscalYear	Fiscal quarter	char	7	

No	Field Name	Description	Type	Size	Key
13	SemesterFiscalYear	Fiscal Semester	char	7	
14	FiscalYear	Fiscal year	smallint	2	

16. Table Name: ReportTimeDim

Description : Date for Generating Report

Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	TimeVATKey	Time Key	smallint	2	PK
2	MonthOfCalendarYear	Month of calendar year	tinyint	1	
3	MonthNameEng	English month name	nvarchar	50	
4	MonthNameThai	Thai month name	nvarchar	50	
5	CalendarYearEng	Calendar year	smallint	2	
6	CalendarYearThai	Calendar year	smallint	2	
7	CalendarQuarterYear	Calendar quarter	char	7	
8	CalendarSemesterYear	Calendar semester	char	7	
9	MonthOfFiscalYear	Fiscal month	char	7	
10	FiscalSemesterYear	Fiscal semester	char	7	
11	FiscalQuarterYear	Fiscal quarter	char	7	
12	FiscalYear	Fiscal year	smallint	2	

17. Table Name: RequestVATDim

Description : Request Type

Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	RequestCode	VAT refund description code (C= Credit , M=Money, D=Draft)	char	1	PK

No	Field Name	Description	Type	Size	Key
2	RequestName	VAT refund description name	nvarchar	50	
3	RequestType	VAT refund type (C=Credit , M=Money)	char	1	
4	RequestTypeName	VAT Refund type name	nvarchar	50	

18. Table Name: TinLtoDim
    Description : Large Business Information
    Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	LtoKey	Large business taxpayer code (0= non LTO , 1= LTO)	char	1	PK
2	LtoName	Large business taxpayer name	char	10	

19. Table Name: VRTAirportDim
    Description : Airport Information divided by The Structure of VRT System
    Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	VRTAirportKey	International airport key of VRT system	tinyint	1	PK
2	VRTAirportCode	International airport code of VRT system	char	2	
3	VRTAirportNameTh	International airport Thai name of VRT system	nvarchar	50	
4	VRTAirportNameEng	International airport English name of VRT system	nvarchar	50	

No	Field Name	Description	Type	Size	Key
5	VRTAirportNameShort	International airport short name for VRT system	nvarchar	50	
6	RDOffice	Revenue office code	nvarchar	50	
7	AirportCode	International airport code	char	2	
8	AirportName	International airport name	nvarchar	50	

20. Table Name: VRTShopDim
Description : Shop in VRT System
Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	VrtShopKey	VRT shop key	smallint	2	PK
2	FirstDate	VRT first date	datetime	4	
3	VrtTin	VRT Taxpayer identification number	char	10	
4	VrtBrano	VRT Branch number	smallint	2	
5	LastDate	VRT last date	datetime	4	
6	TitNameCode	VRT title code	char	8	
7	TitName	VRT title name	nvarchar	25	
8	VrtBranName	VRT business branch Thai name	nvarchar	160	
9	ShopEngName	VRT business branch English name	nvarchar	160	
10	BldgName	Building name	nvarchar	40	
11	Village	Village name	nvarchar	40	
12	RoomNo	Room number	nvarchar	20	
13	FloorNo	Floor number	nvarchar	20	

No	Field Name	Description	Type	Size	Key
14	AddNo	Address number	nvarchar	30	
15	MooNo	Moo number	char	2	
16	SoiName	Soi name	nvarchar	40	
17	ThnName	Street name	nvarchar	40	
18	SubDistrictCode	Sub-district code	char	6	
19	SubDistrictName	Sub-district name	nvarchar	50	
20	DistrictCode	District code	char	6	
21	DistrictName	District name	nvarchar	50	
22	ProvinceCode	Province code	char	6	
23	ProvinceName	Province name	nvarchar	50	
24	PosCode	Zip code	char	6	
25	TelNo	Telephone number	nvarchar	25	
26	FaxNo	FAX number	nvarchar	25	
27	Email	E_mail address	nvarchar	30	
28	TitShopCode	Shop title code	nvarchar	8	
29	TitShopName	Shop title name	nvarchar	25	
30	ShopName	Shop name	nvarchar	160	
31	RdOfficeCode	Revenue office code	nvarchar	80	
32	LtoCode	Large Business Taxpayer code	char	1	
33	IscKey	International Standard Industrial Classification Key	smallint	2	
34	VrtTinKey	VRT Taxpayer ID. number key	smallint	2	

21. Table Name: VRTTaxpayerDim

Description : VAT Registrant in VRT System

Table Type : Dimension Table

No	Field Name	Description	Type	Size	Key
1	VrtTinKey	VRT Taxpayer identification number Key	smallint	2	PK
2	VrtTin	VRT Taxpayer identification number	char	10	
3	FirstDate	First date of VRT system	datetime	4	
4	LastDate	Last date of VRT system	datetime	4	
5	TitNameCode	VRT Taxpayer title code	char	10	
6	TitName	VRT Taxpayer title name	nvarchar	25	
7	TaxpayerNameThai	VRT Taxpayer Thai name	nvarchar	160	
8	TaxpayerNameEng	VRT Taxpayer English name	nvarchar	160	
9	Pin	Personal identification number	nvarchar	15	
10	VATDate	VAT registration date	datetime	4	
11	JurDate	Corporation registration date	datetime	4	
12	BldName	Building name	nvarchar	40	
13	RoomNo	Room number	nvarchar	20	
14	FloorNo	Floor number	nvarchar	20	
15	Village	Village name	nvarchar	40	

No	Field Name	Description	Type	Size	Key
16	Addno	Address number	nvarchar	30	
17	MooNo	Moo number	nvarchar	2	
18	SoiNam	Soi name	nvarchar	40	
19	ThnName	Street name	nvarchar	40	
20	PosCode	Zip code	nvarchar	5	
21	TelNo	Telephone number	nvarchar	25	
22	GoodsTypeCode	Goods type code	char	6	
23	SubDistrictCode	Sub-district code	char	6	
24	SubDistrictName	Sub-district name	nvarchar	50	
25	DistrictCode	District code	char	6	
26	DistrictName	District name	nvarchar	50	
27	ProvinceCode	Province code	char	6	
28	ProvinceName	Province name	nvarchar	50	
29	RDOfficeCode	Revenue office code	char	8	
30	TinLto	Taxpayer identification number of Large Business Taxpayer code	char	1	
31	IscKey	Business type code	smallint	2	

4.5 Data Cube Design

Data Cube of ACRM-VRT system is the combination of four data cubes described in Table 4.4

Table 4.4 Summary of data cube design

No	Cube Name	Fact Table	Dimension Table
1	TouristRefundVAT Cube	TouristRefundVATFact	RefundVATTimeDim , VRTAirportDim , LocationRefundVATDim , RequestVATDim , PaidVATDim , CountryDim , RefundHourDim , GenderDim
2	TouristPurchaseGoodsCube	TouristPurchaseGoodsFact	ReportTimeDim , VRTShopDim , VRTTaxpayerDim , IscDim , GenderDim , GoodsDim , AirportDim, ProvinceDim, CountryDim
3	VRTTaxpayerAndShop Cube	VRTTaxpayerAndShopFact	ProvinceDim , TinLtoDim , RDOfficeDim , ReportTimeDim , IscDim
4	TaxpayerPaidTax Cube	TaxpayerPaidTaxFact	ReportTimeDim , VRTShopDim , VRTTaxpayerDim

1. TouristRefundVAT Cube

Data Cube of TouristRefundVAT composes of eight dimension tables and 1 fact table (showed in Figure 4.5) can be used to generate eight reports as below.

1) VAT Refund for Tourists report of the current fiscal year. Compared with the previous fiscal year (monthly).
2) VAT Refund for Tourists report of the current fiscal year. Compared with the previous fiscal year by airports.
3) VRT classified by location of province with an international airport report.
4) VAT Refund for Tourists report, by money for annual fiscal year.
5) Statistics of tourists to refund VAT, by residence of tourists.

6) Statistics of tourists to refund VAT, by airports.
7) Statistics refund by type of refund.
8) Statistics of VAT refund to tourists. By pan-residence of tourists.

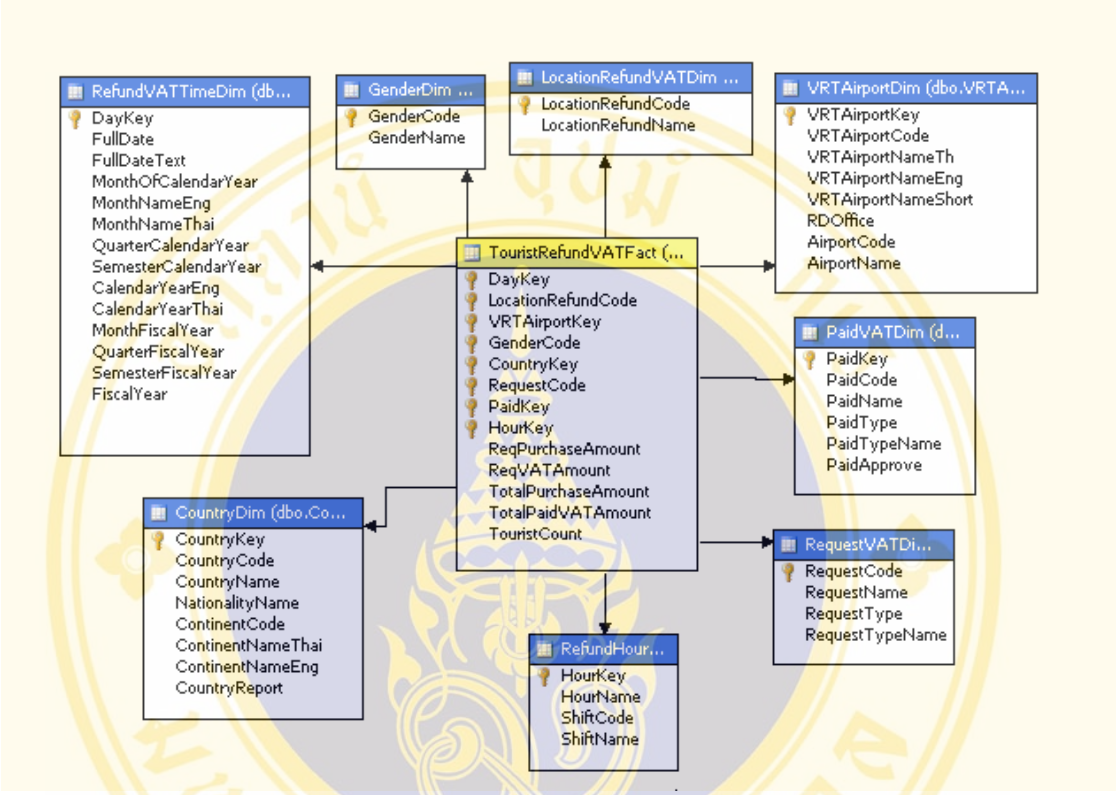


Figure 4.5 Data Cube of VAT Refund for Tourist

2. TouristPurchaseGoods Cube

Data Cube of tourists' purchasing consists of eight dimension tables and 1 fact table (shown in Figure 4.6). These data can be used to generate three reports as follows:

1) VAT refund for tourists, by type of product getting VAT return.
2) Product type which the tourists buy frequently and request for VAT refund.
3) Top ten product type which the tourists buy frequently and request for VAT.

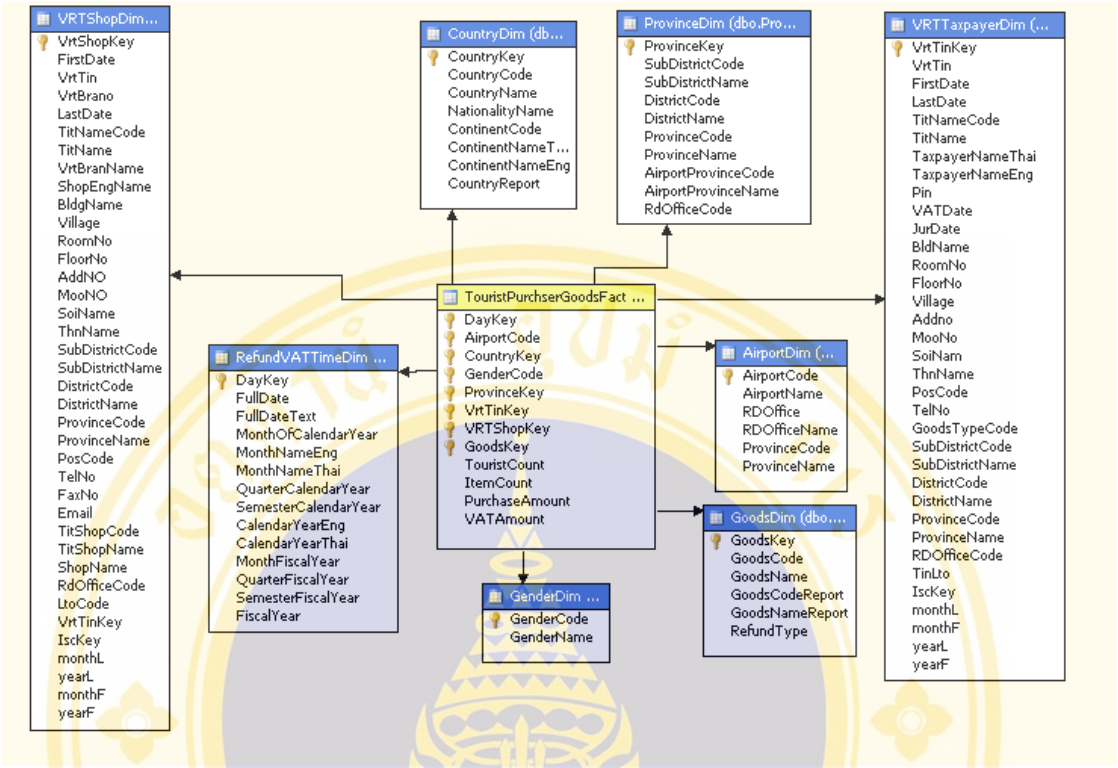


Figure 4.6 Tourists' purchasing

3. VRTTaxpayerAndShop Cube

Number of business and shop in VRT system data cube consists of five dimension tables and 1 fact table (shown in Figure 4.7) There are three types of report can be generated from this data cube described as follows.

1) Statistics of approval for entrepreneurs selling goods to tourists.
2) Number of business requesting for VAT refund to tourists. Set by the business location of the provincial international airports.
3) Accumulated statistic of VAT refund for tourist, compared with the previous fiscal year. (monthly)

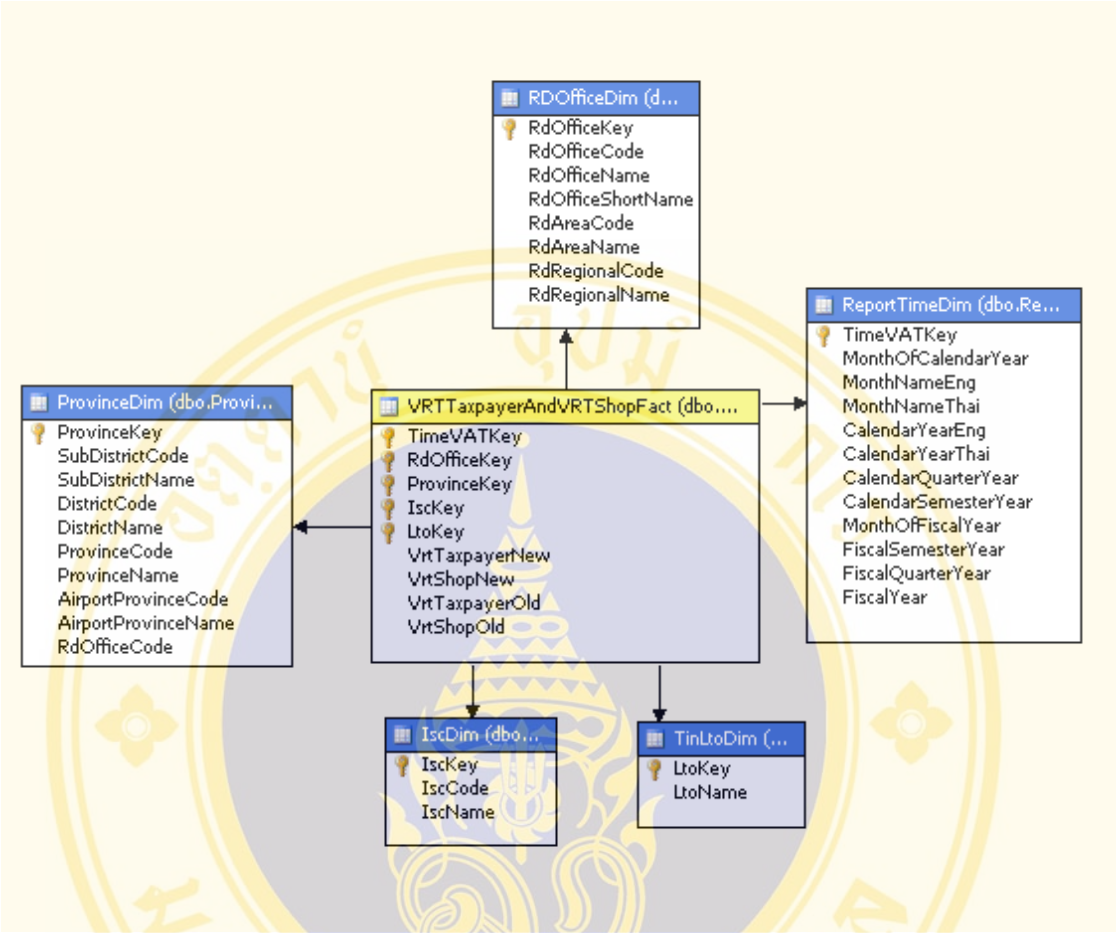


Figure 4.7 Number of business in VRT system

4. TaxpayerPaidTax Cube

Data Cube of the balance of taxation of entrepreneurs consists of three Dimension Tables and 1 Fact Table (shown in Figure 4.8) which can be used to generate monthly and annually statistic report of business's tax payment.

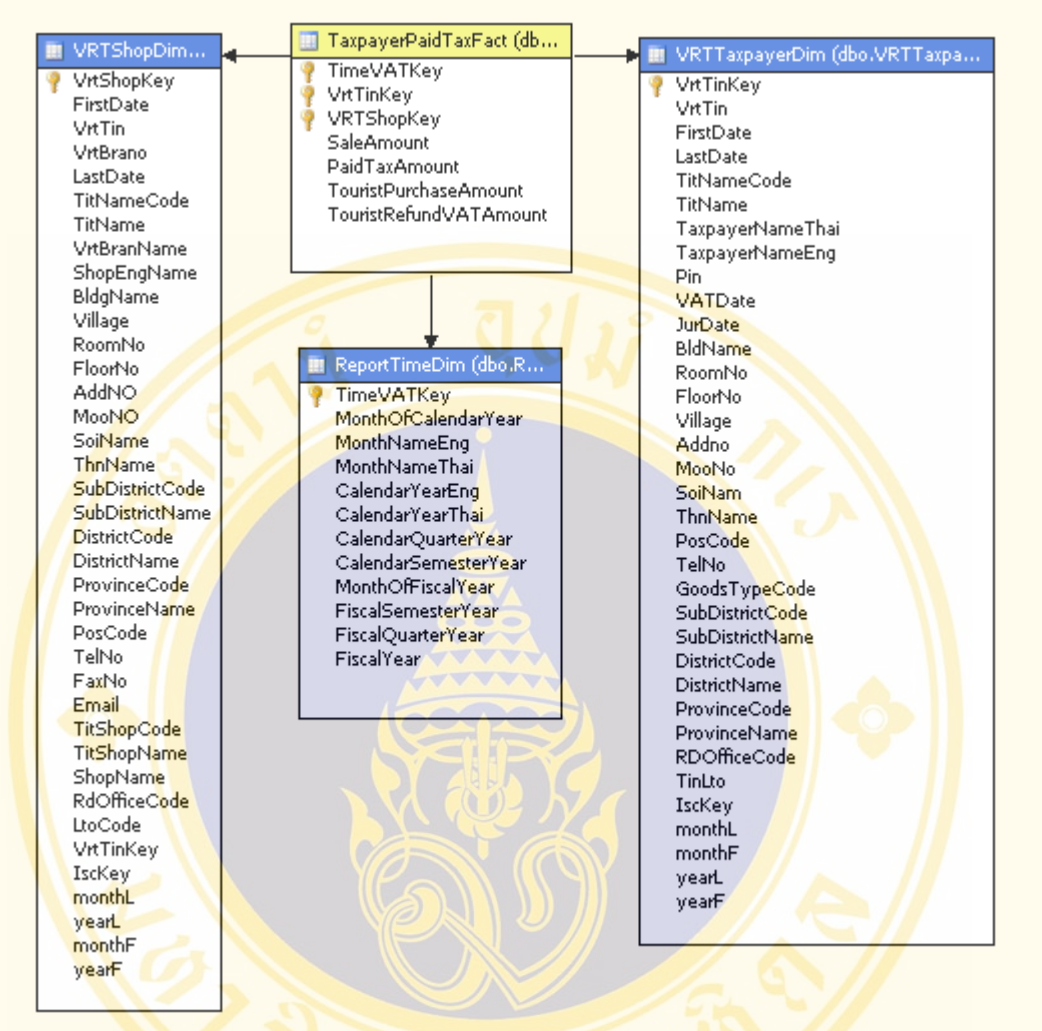


Figure 4.8 Monthly and annually statistic report of business's tax payment

4.6 OLAP Tool Design

OLAP Tool provides the service. VRT officers can view data in various dimensions via Web Interface. Moreover VRT officers can also Pivot, Rollup, Drill Down, Slice and Dice data. The working structure of this engine is illustrated in Figure 4.9.

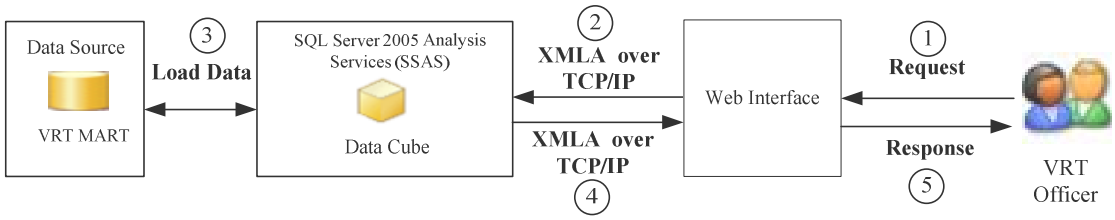


Figure 4.9 Working structure of OLAP tool

Figure 4.9 shows the working Structure of OLAP Tool. Starting with VRT officer send request to Web Interface and then it will transform request into MDX (MDX stands for Multidimensional Expressions Language. MDX is a query language that is used to interact and perform tasks with multidimensional databases (also called: OLAP Cubes). The MDX language was originally developed by Microsoft in the late 1990s, and has been adopted by many other vendors of Multidimensional Databases.[20]) and send XMLA over TCP/IP (XML for Analysis (XMLA) is the native protocol for Microsoft SQL Server 2005 Analysis Services (SSAS), used for all interaction between a client application and an instance of Analysis Services [20]) to request data from SQL Server 2005 Analysis Services (SSAS) where Data Cube is stored. After that SSAS will load data from VRT MART and send the result via XMLA over TCP/IP and return Web Interface Response to VRT officer.

4.7 Association Rule Mining Tool

Association Rule Mining Tool is the tools using to analyze purchasing data and sale data. VRT officer can request through Web Interface. (Figure 4.10)

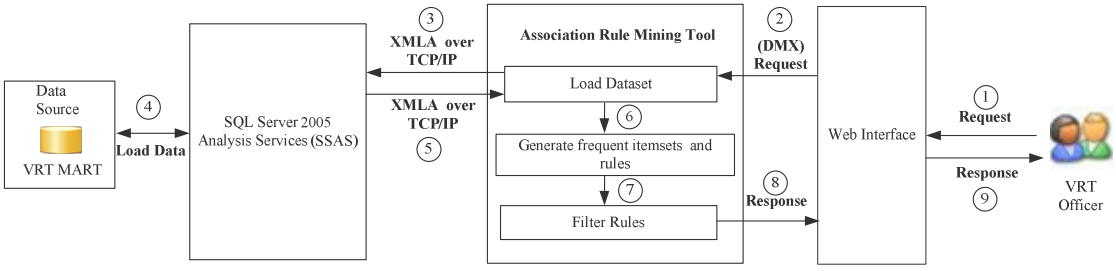


Figure 4.10 Working structure of association rule mining tool

Figure 4.10 shows the working processes of Association Rule Mining Tool. Starting from VRT officer defines the conditions of Data Set and Parameter for

building Rule via Web Interface. Then, Web Interface will interpret the statement to DMX (Data Mining Expressions Language) and send to Association Rule Mining Tool. After that Association Rule Mining Tool will send data in the form of XMLA over TCP/IP to connect and run Microsoft Association Rule Algorithm and retrieve data from SQL Server 2005 Analysis Services (SSAS). Then SSAS will load data from VRT MART and forward the result via XMLA over TCP/IP. Next, SSAS will return the result to Association Rule Mining Tool to build the rule and Display Rule according to the defined conditions and it will be sent the result to display on Web Interface.

Working Processes Association Rule Mining Tool is illustrated in Figure 4.11

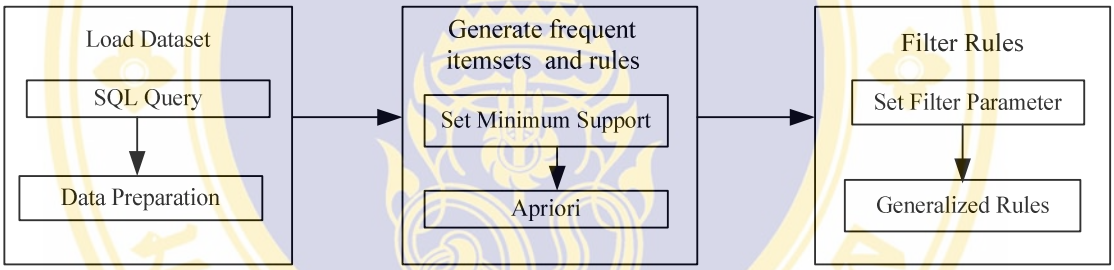


Figure 4.11 Working process of association rule mining tool

Figure 4.11 shows the working processes of the Association Rule Mining Tool consists of three components as follows:

1) Load Dataset is responsible for preparing data used to build Rule. Its working process is divided into two sub processes.

- SQL Query fetches statement from user and transforms it into the SQL form used to retrieve data from Data Mart and send it to Data Preparation.
- Data Preparation manages data to be in the form of structure and conditions defined by the user.

2) Generate frequent itemsets and rules are responsible for exploring the relational rules from received data using Apriori Algorithm. There are two main steps in this algorithm described as follows:

- Set Minimum Support is responsible for fetching Minimum Support value. This used to build Rule and defined by user and sends the result to Apriori.
- Apriori is responsible for building Rule according to Minimum Support value and the result from Apriori Algorithm.

3) Filter Rules is responsible for display the Rule satisfying conditions defined by the user. There are two main steps in this algorithm described as follows:

- Set Filter Parameter is responsible for getting Minimum Confident and/or keyword defined by user to search Rule satisfying defined conditions and transform into DMX language for searching Rule and forward the result to Generalized Rules.
- Generalized Rules are responsible for searching and displaying Rules according to DMX language and forward the result to user.

CHAPTER V
ACRM-VRT SYSTEM DEVELOPMENT

This chapter explanation about the Hardware and Software use for develop ACRM-VRT system. And explanation about operation of all functions of the ACRM-VRT system which consist of Statistic Report, Searching VRT Taxpayer Data, OLAP Tool and Data Mining Tool and example how to create report by using OLAP Tool and Data Mining Tool.

5.1 Hardware and Software Used for Implementation

Hardware and Software use for develop ACRM-VRT system.

1) Hardware

- Laptop computer simulating as Application Server and Data Mart Server
- Processor: Intel Pentium III 1.3 GHz.
- RAM: 1.5 GB

2) Software

- Windows Server 2003 Enterprise Edition
- Internet Explorer 6.0
- Microsoft SQL Server 2005
- Microsoft SQL Server 2005 Analysis Service
- Microsoft SQL Server 2005 Reporting Service
- Microsoft Virtual Studio 2005 Version 8.0
- Microsoft .Net Framework Version 2.0

5.2 User Interface of ACRM-VRT System

This part explained about system components and system functions, as follow.

5.2.1 System Login

User input username and password, then click OK button to login.

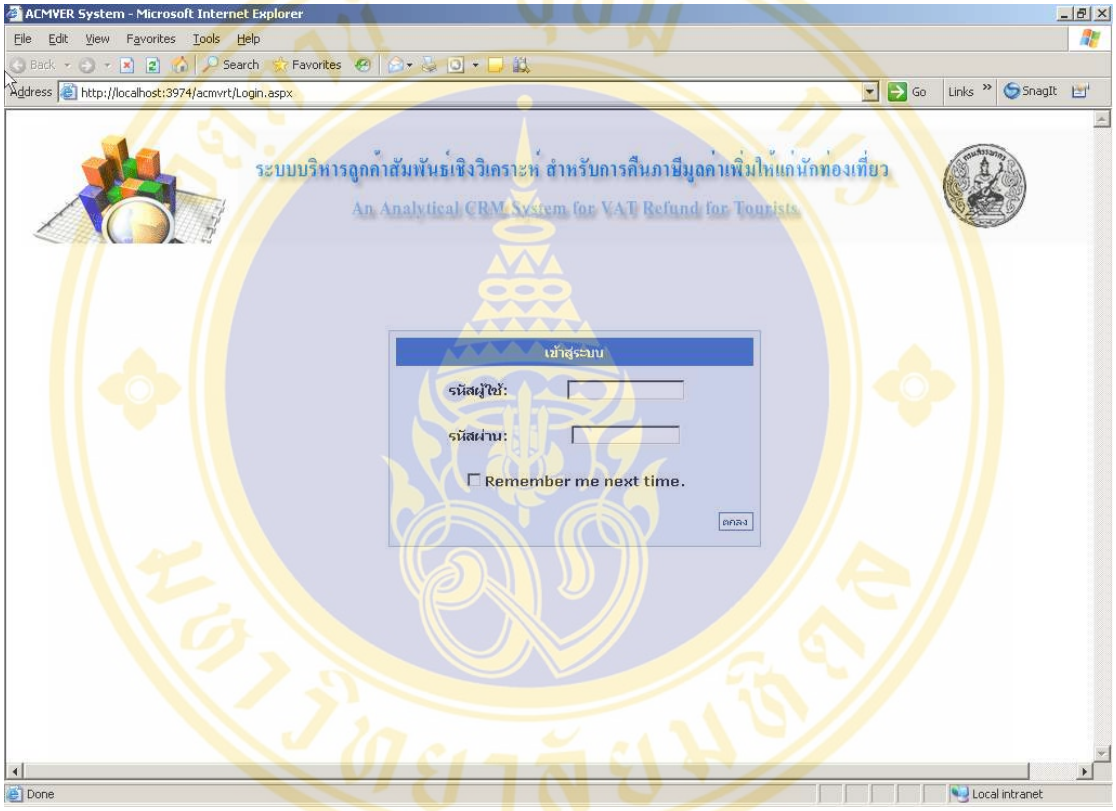


Figure 5.1 Login Screen

5.2.2 Main Menu

After login success, the Main Menu screen will appear (Figure 5.2). Main Menu is contained five functions:

1) Statistic Report
2) Searching VRT Taxpayer Data
3) OLAP Tool
4) Data Mining Tool
5) Log out

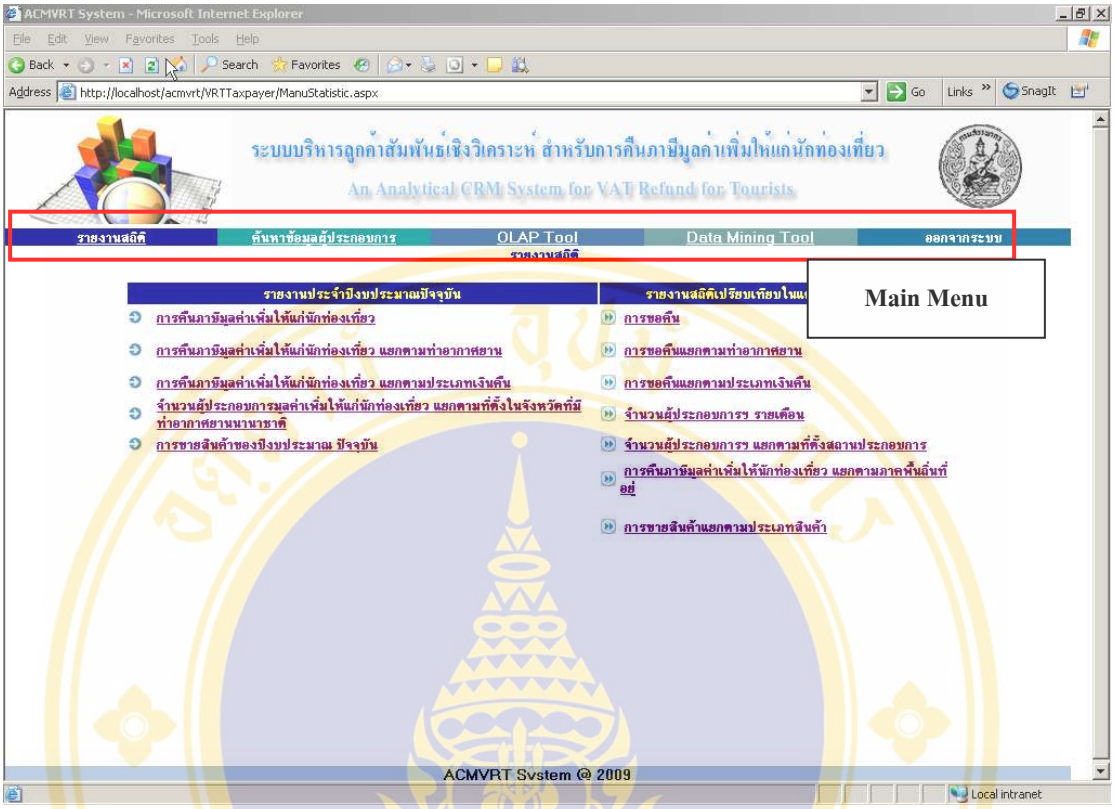


Figure 5.2 Main Menu Screen

5.2.3 Statistic Report

Users can create and view reports from this function. Reports consist of current fiscal year's reports and comparison reports between fiscal years. The detail is as follow:

Year reports contain 5 reports:

- VAT refund for tourist Report
- VAT refund for tourist Report divided by airport
- VAT refund for tourist Report divided by refund type
- Number of VRT Entrepreneurs Report divided by province which the International Airport located
- Sales Report divided by fiscal year

Comparison Reports contain 7 reports:

- Refund Statistic
- Refund Statistic divided by airport
- Refund Statistic divided by refund type

- Number of VRT Entrepreneurs Monthly Statistic
- Number of VRT Entrepreneurs Statistic divided by province which the International Airport located
- Refund Statistic divided by tourist residence's continent
- Sales Report divided by goods type

How to view report: User Click on each report name (Figure 5.3) to view the report screen (Figure 5.4)

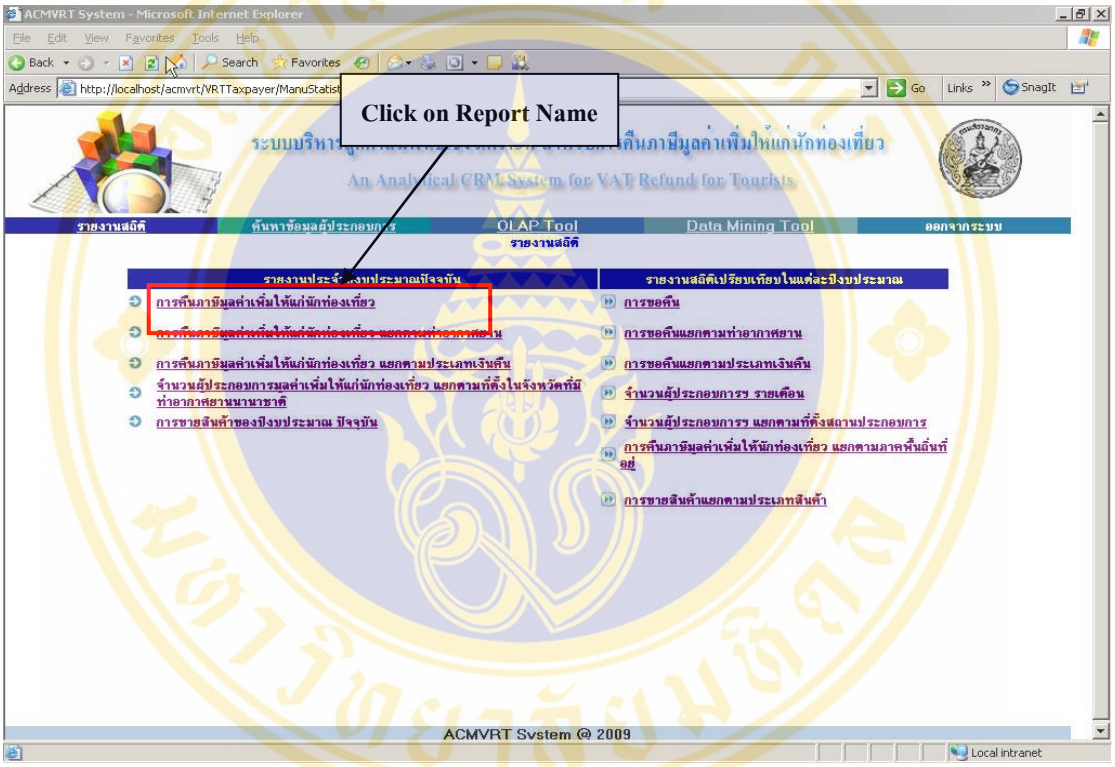


Figure 5.3 Show how to select report

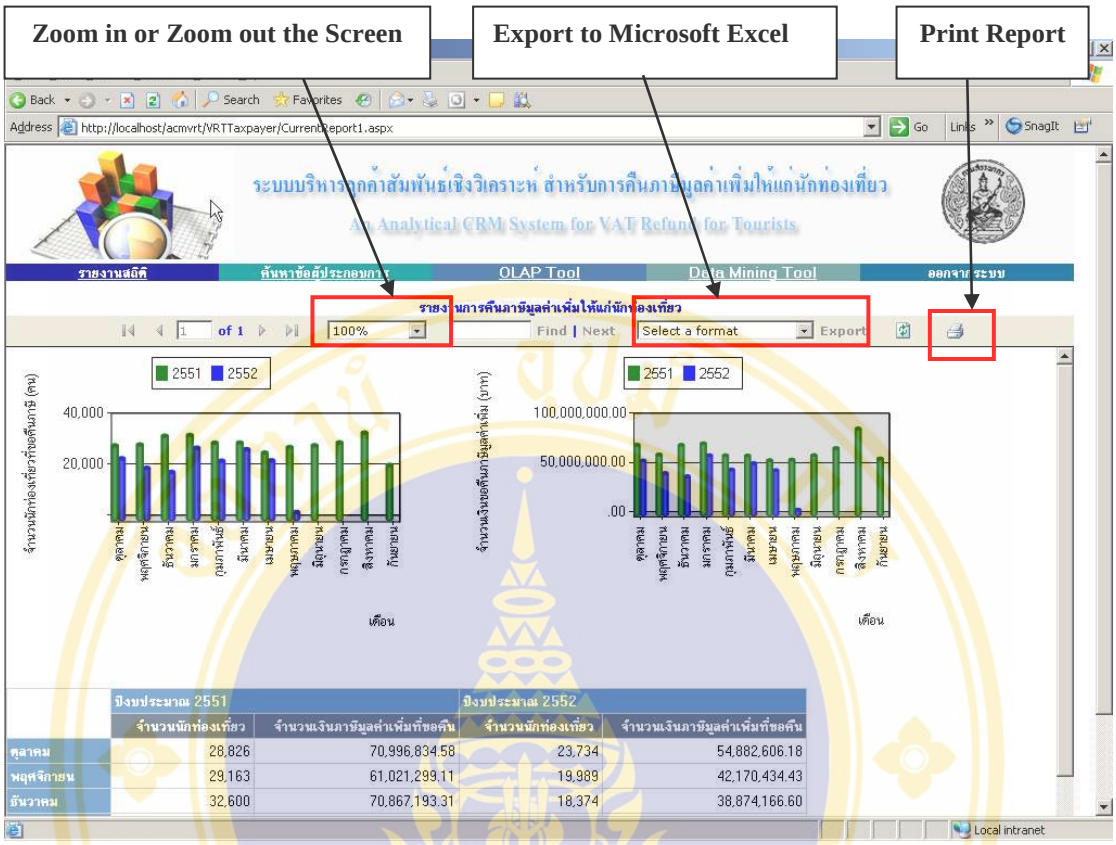


Figure 5.4 Report screen

Figure 5.4 after user selects and clicks on report name, the report screen will show up. User can zoom in or zoom out the screen, export to Microsoft Excel files and print report.

5.2.4 Searching VRT Taxpayer Data

User can search VRT Entrepreneurs by input TIN and click “View report” button. (Figure 5.5) The system will show entrepreneurs name, address, branch, VRT registered date, VRT expired date on screen (see Figure 5.6).

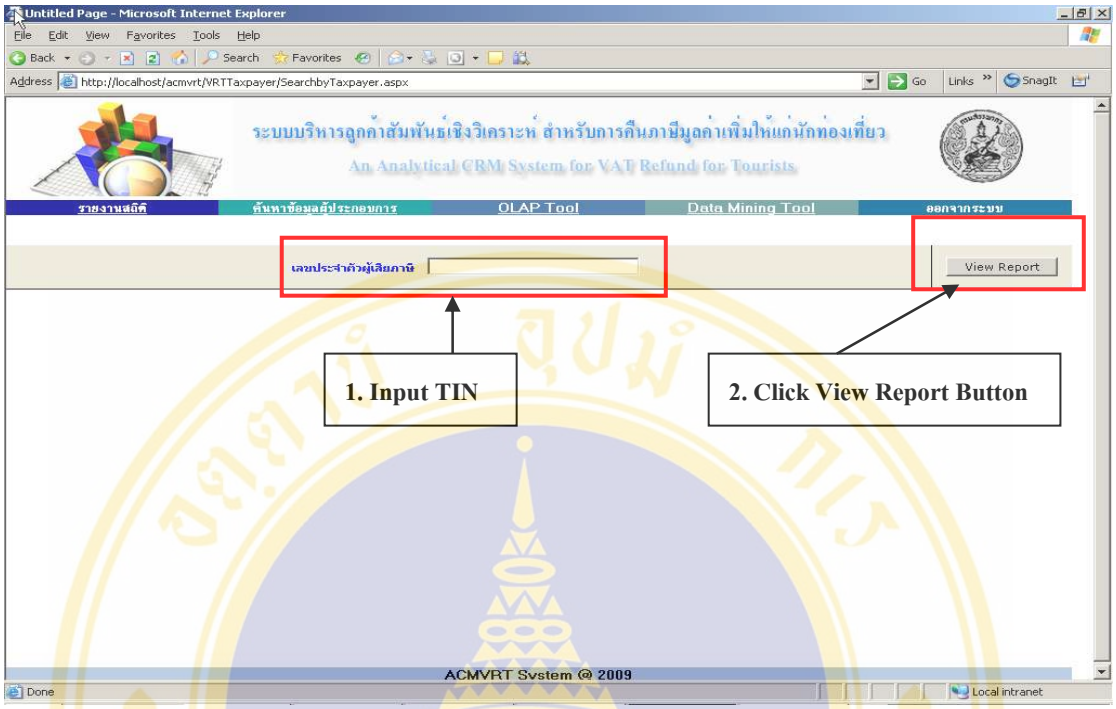


Figure 5.5 Search VRT Taxpayer data screen

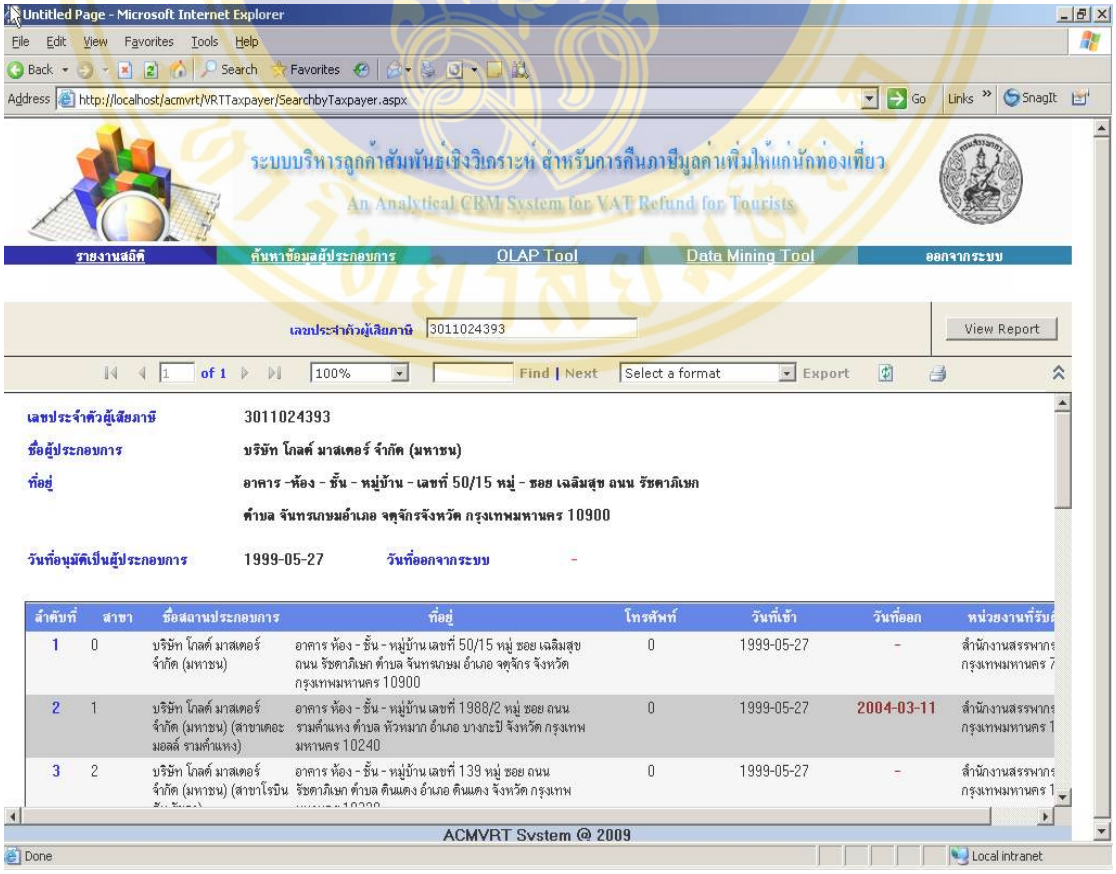


Figure 5.6 VRT Taxpayer information screen

5.2.5 OLAP Tool

Users can browse report by using OLAP Tool. Reports will display in graph form, and data table. And Users can define conditions to view only data which they want. OLAP Tool contains several components, as shown in Figure 5.7.

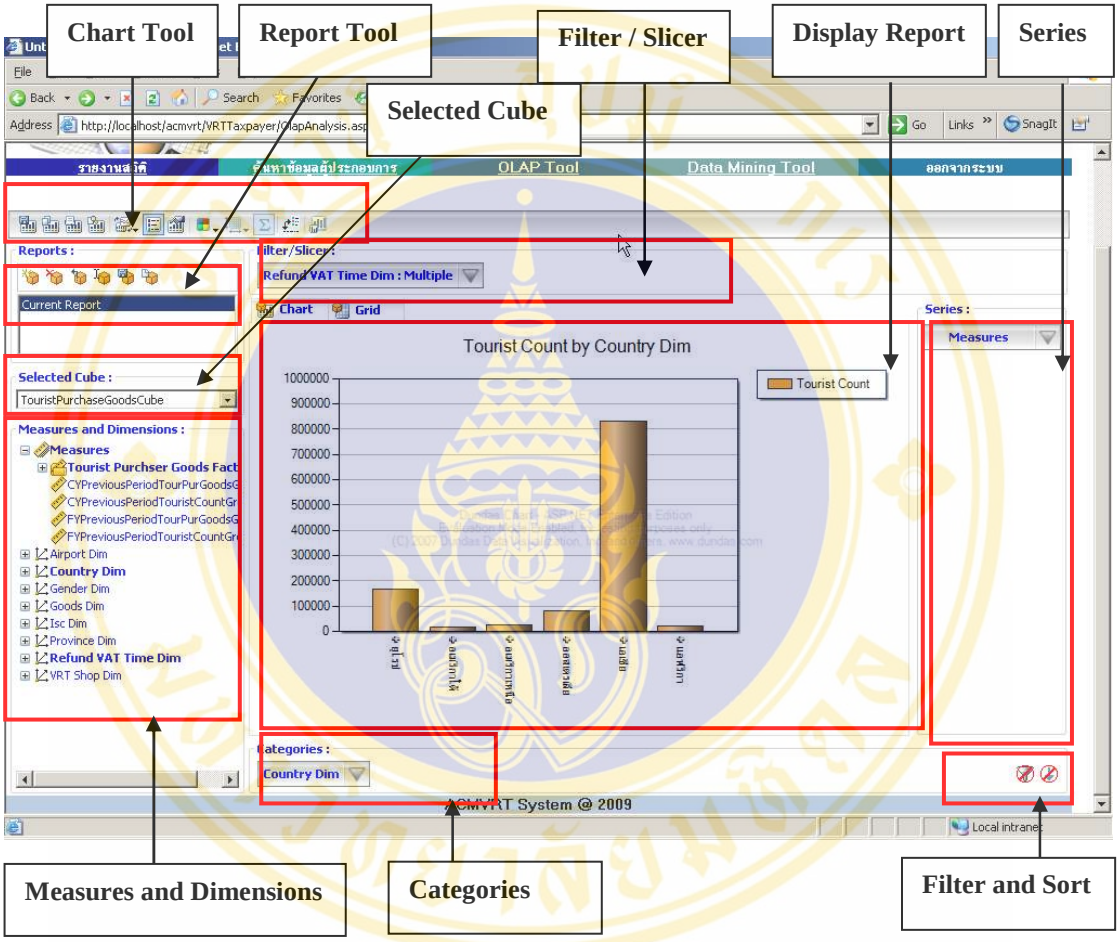


Figure 5.7 Components of online data analysis browsing screen

Detail of OLAP Tool components:

- Graph Tool is the tool for set graph forms and graph details.
- Report Tool is the tool for manage report such as create, delete and save.
- Selected Tool is the component used for select Data Cube.
- Filter/Slicer is the panel to place data which used for the specified data.
- Series is the panel to paste the specified data group.

- Measures and Dimensions is the panel for display fields in Data Cube which is Measures and Dimensions type.
- Categories is the panel for pasted and displayed required data type.
- Filter and Sort is the panel for filter and sort data.
- Display Report is the panel for display graph and data table.

Example of Report Generation

For example: Assumption that user wants to know the top 10 goods which are popular for tourists in fiscal year 2551 (or 2008). We analyze data step by step as follow (see Figure 5.8)

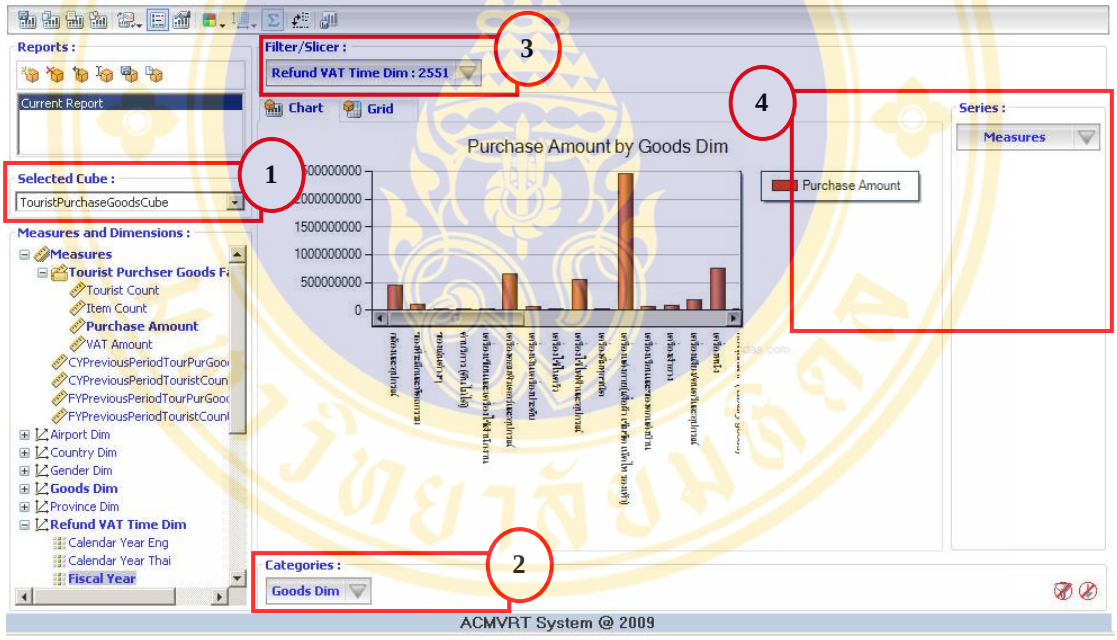


Figure 5.8 Show selection data top 10 popular product types of year 2008

1) At Selected Cube panel select TourisPurchaseGoodsCube.
2) At Measures and Dimensions panel select Goods Dim then drag GoodsName and drop at Categories panels.
3) At Measures and Dimensions panel select Refund VAT time Dim then drag Fiscal Year and drop in Filter/Slicer and select year 2551 (or 2008).
4) At measures panel select Purchase Amount.

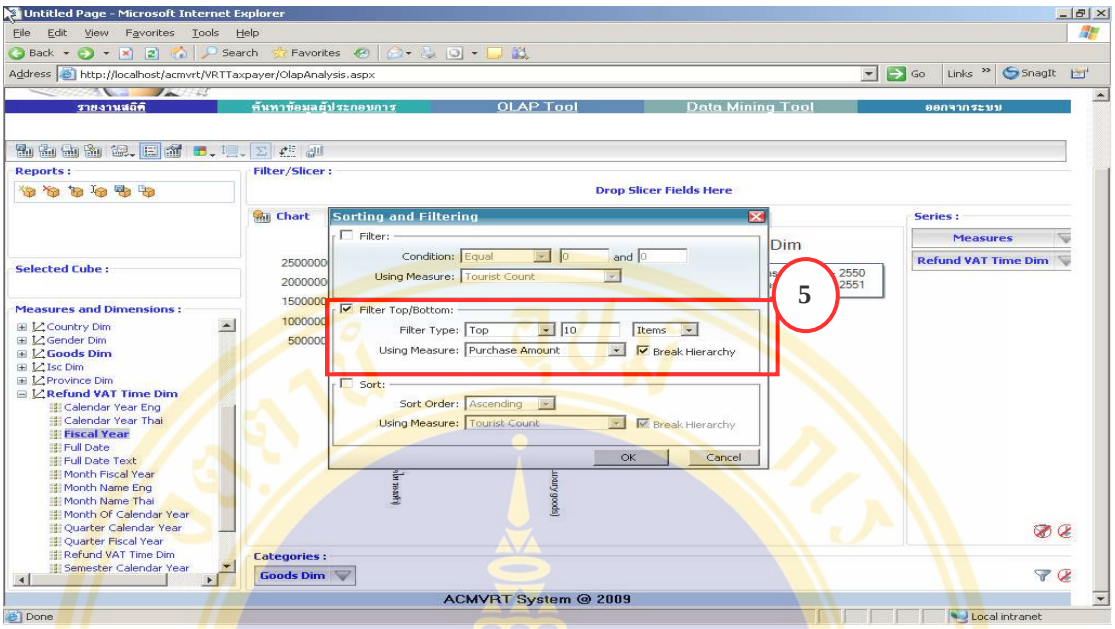


Figure 5.9 Filtering data

5) Click select Filter Top / Button :

Filter Type: select Top and 10

Using measure: select Purchase Amount and click OK button (see Figure 5.9)

6) Click Icon on Graph Tool bar to modify graph title or report title by type name in Title Text and then click OK (see Figure 5.10)

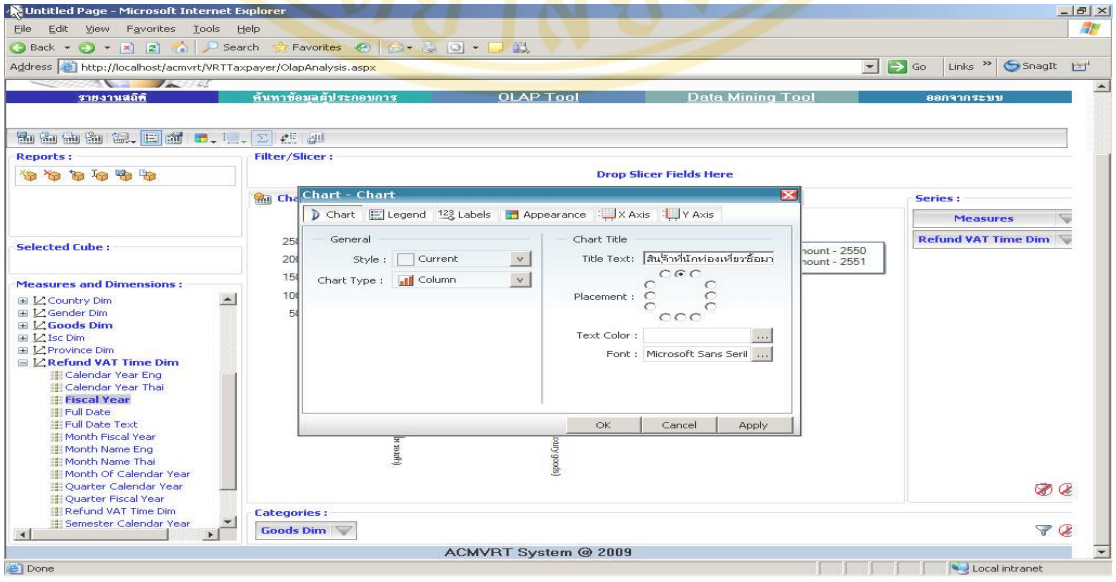


Figure 5.10 Change graph title

7) Click 

tab to changed report view from graph to data table. (see Figure 5.11)

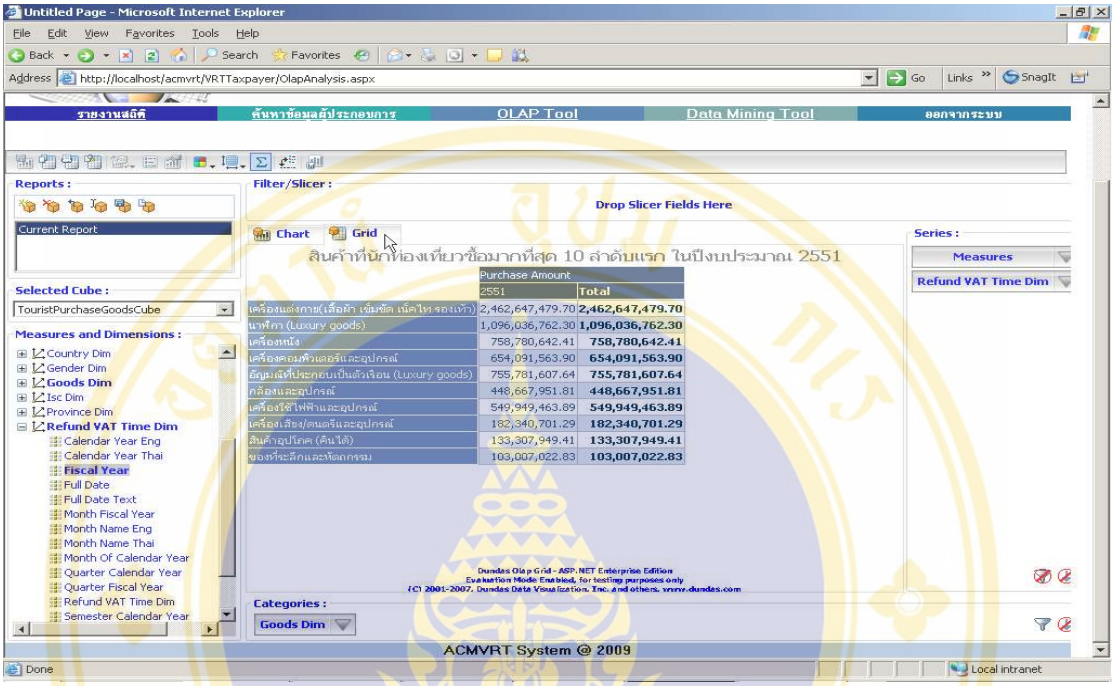


Figure 5.11 Change display from graph to table

8) Click  icon on Report Tool bar to save report. Type file name in Save As box the click save button, the saved file is XML file type. (see Figure 5.12)

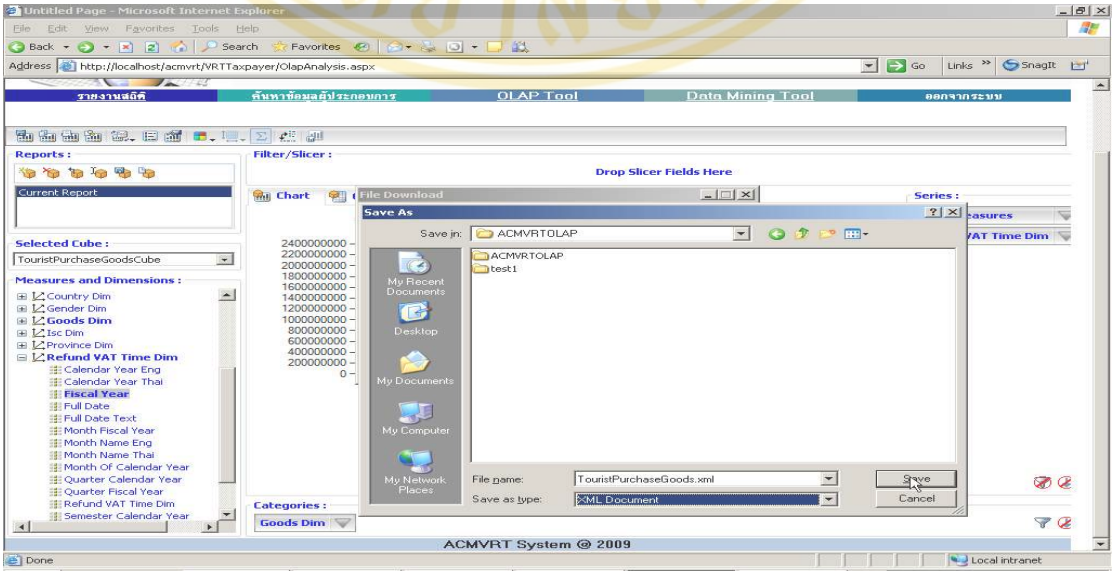


Figure 5.12 Save report

9) To open report file, click  icon on Report Tool bar. Browse saved file from Load Reports box and click OK. (see Figure 5.13)

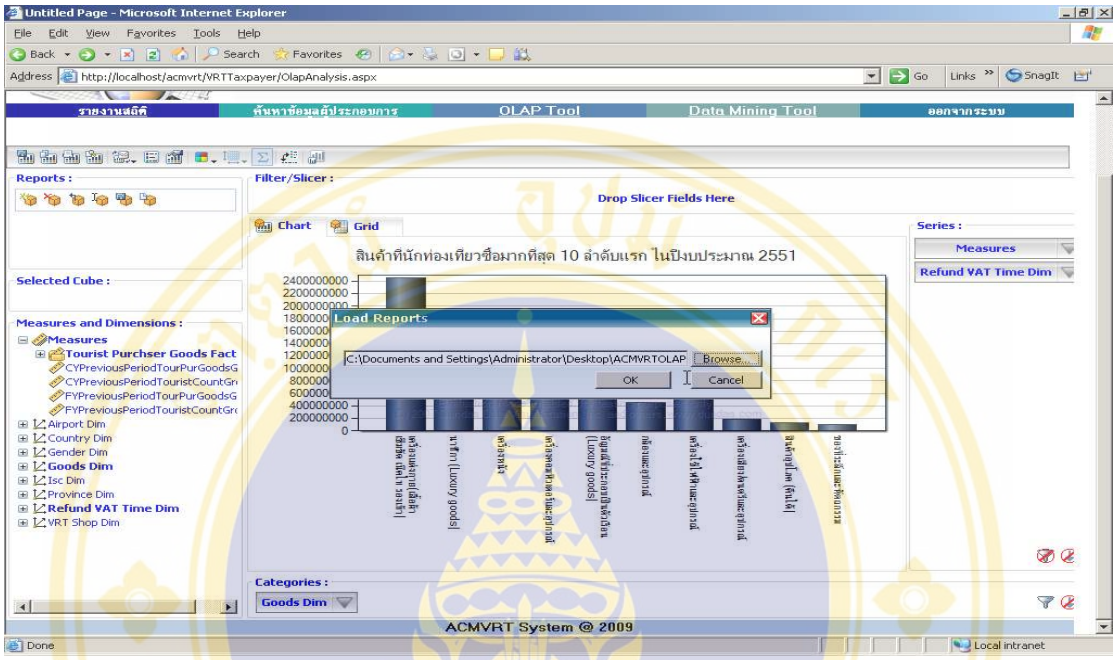


Figure 5.13 Loading report

5.2.6 Data Mining Tool

Data Mining Tool is use for analyze sales data. User can customize variables for analyze, moreover, the system provided sample cases to help user learn how to customize variables for use to analyze data.

Click Data Mining menu to start this function, system will display the Data Mining screen (see figure 5.14), there are two main parts: Conditioned part and Display part.

Conditioned part use for create rule consist of:

1) **Conditions:** used for set condition to create Rule, user can use provided sample cases or customize own conditions.
2) **Year selected:** select year of data for process, in current database we have data from year 2549 to 2552 (2006 to 2008).
3) **Set Minimum Support value:** used to set number of Records of transaction for creating Rule, can set from 0.00 to 1.00 .
4) **Where:** used to set the terms of 'Data Set' for creating rule, such as Japanese tourists.

5) **Field:** to set Attribute of Data Set for creating rule.

6) **Input:** to set Attribute value to be Antecedence of rule.

7) **Predict:** to set Attribute to be Consequence of rule.

8) **Data View:** for view list of each Attribute (see Figure 5.15) display part.

9) **Filter Rule:** used to set display condition of Rule consist of:

- Minimum Probability: used to set Confidence values of Rule
- Filter: used to Filter Rule the value set in Rule
- แสดงกราฟ display Rule's graph in bar graph form (Figure 5.16) or pie graph form (Figure 5.17)

10) **Display Rule**: consist of:

- to explained the meaning of Rule. (see Figure 5.18)
- show number, name, address of shops which are members of VRT system and not member of VRT system. Officer can find new members of VRT system from this information. (see Figure 5.19)

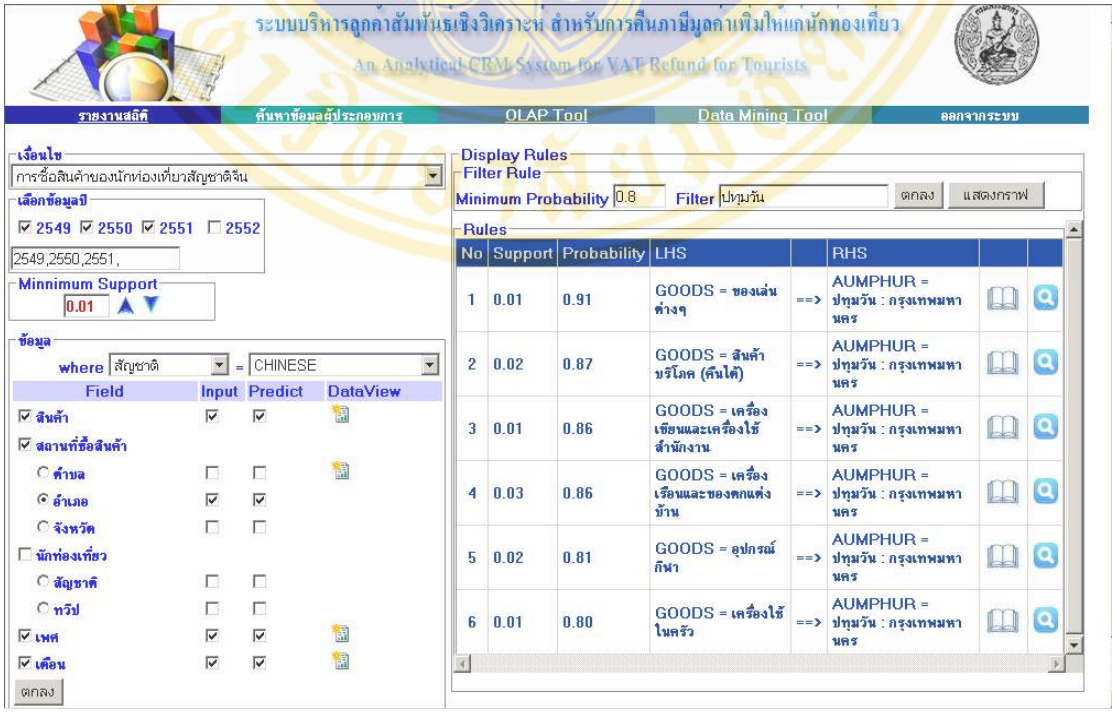
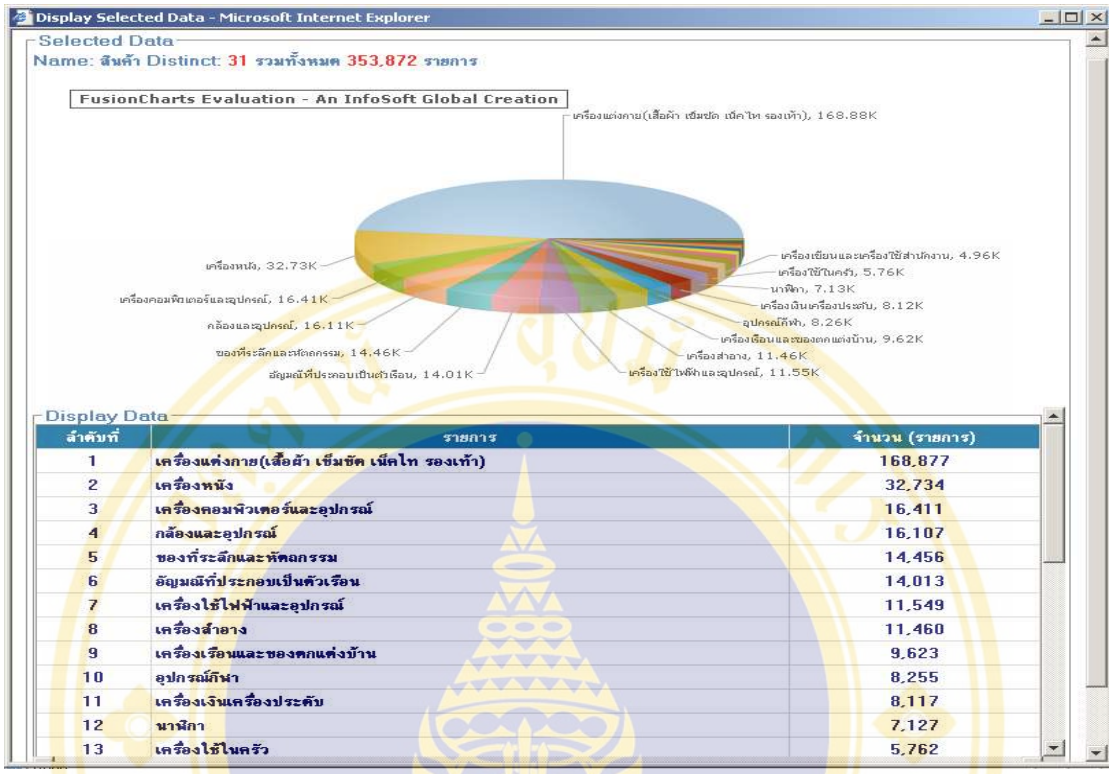


Figure 5.14 Components of Data Mining Tool



ลำดับที่	รายการ	จำนวน (รายการ)
1	เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)	168,877
2	เครื่องหนัง	32,734
3	เครื่องคอมพิวเตอร์และอุปกรณ์	16,411
4	กล้องและอุปกรณ์	16,107
5	ของที่ระลึกและหัตถกรรม	14,456
6	อัญมณีที่ประกอบเป็นตัวเรือน	14,013
7	เครื่องใช้ไฟฟ้าและอุปกรณ์	11,549
8	เครื่องสำอาง	11,460
9	เครื่องเรือนและของตกแต่งบ้าน	9,623
10	อุปกรณ์กีฬา	8,255
11	เครื่องเงินเครื่องประดับ	8,117
12	นาฬิกา	7,127
13	เครื่องใช้ในครัว	5,762

Figure 5.15 Data View

Figure 5.15 Data View used for display graph and table of Attribute which use to create Rule, consist of unique type of data item, yearly summary of data and setting conditions.

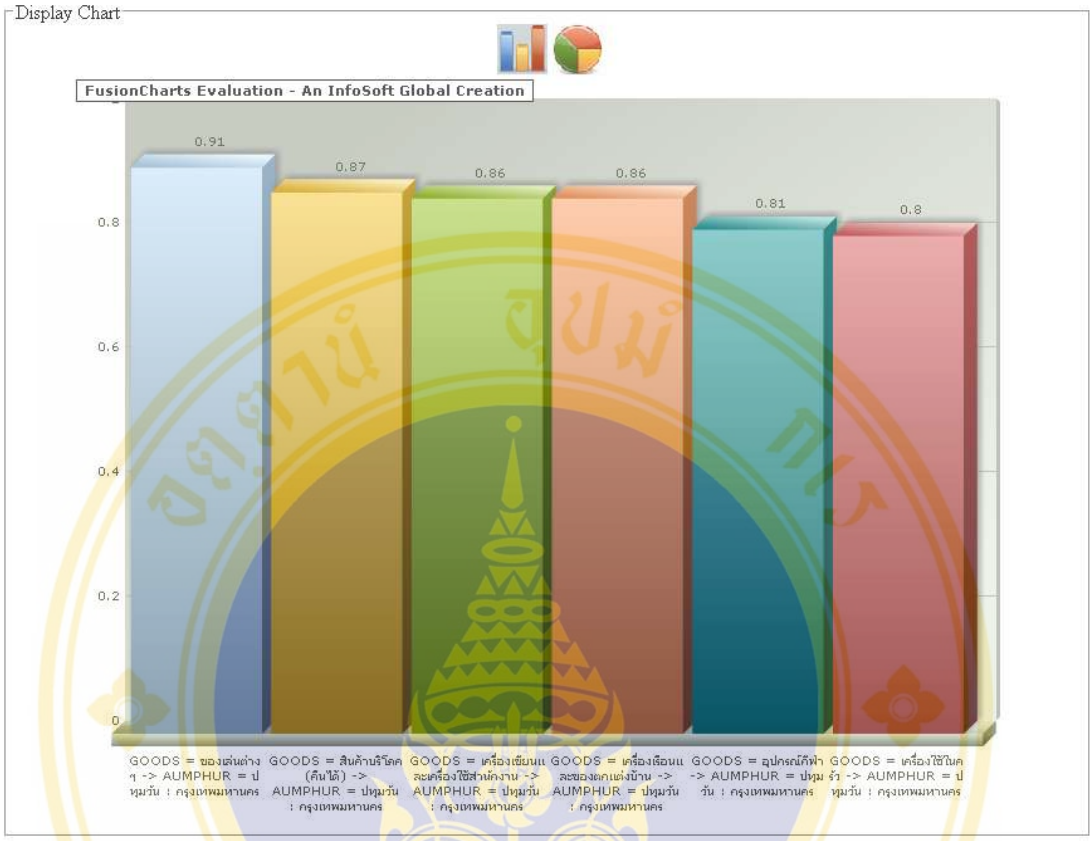


Figure 5.16 Result of rules in bar graph

Figure 5.16 Display result of Rule in bar graph and sort by Probability value descending.

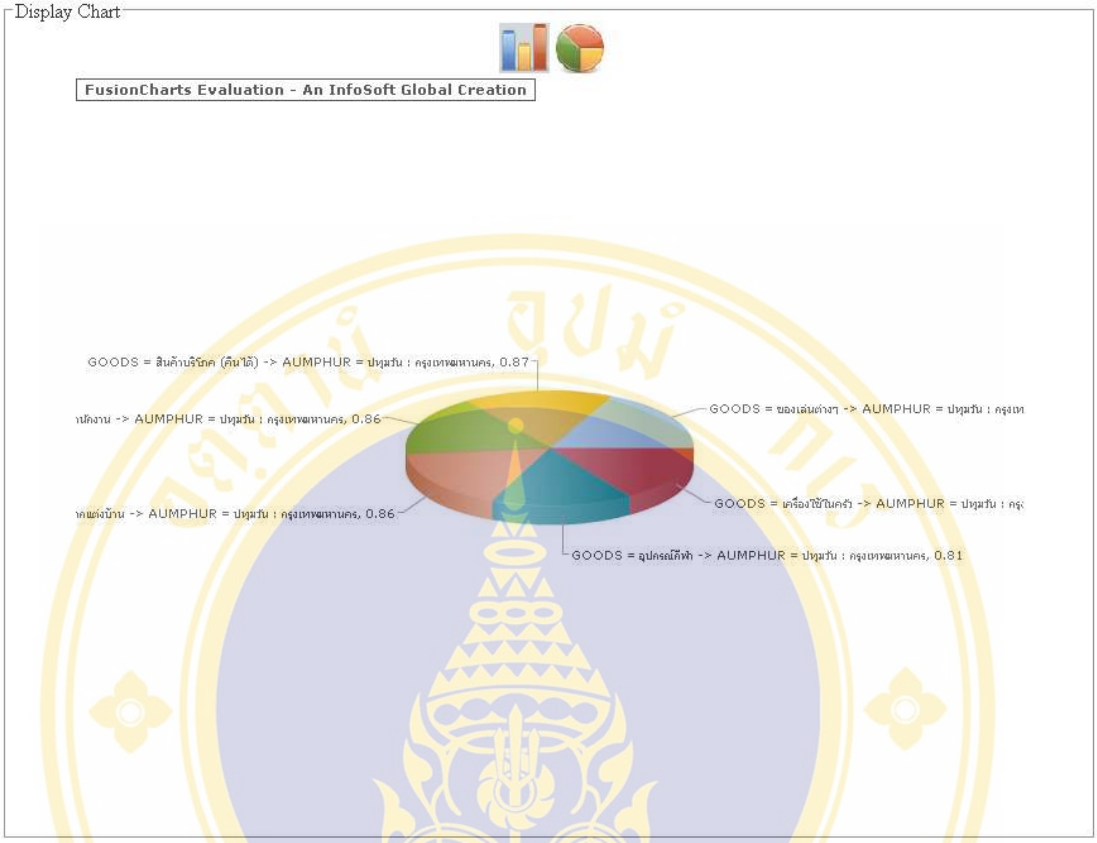


Figure 5.17 Rules' result in pie graph form

Figure 5.17 Display result of Rule in pie graph.

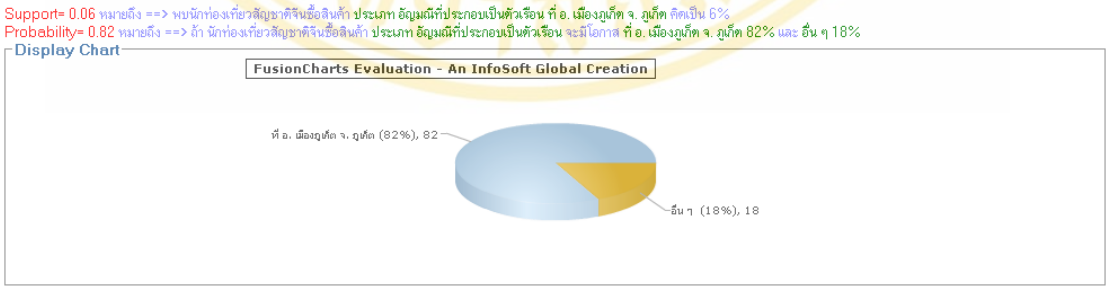


Figure 5.18 Rules' meaning

Figure 5.18 Display the meaning of Rule according to user selection, helping use to understand Rule.

For example: Analysis purchasing behavior of Chinese tourists from purchasing data from year 2549-2551.

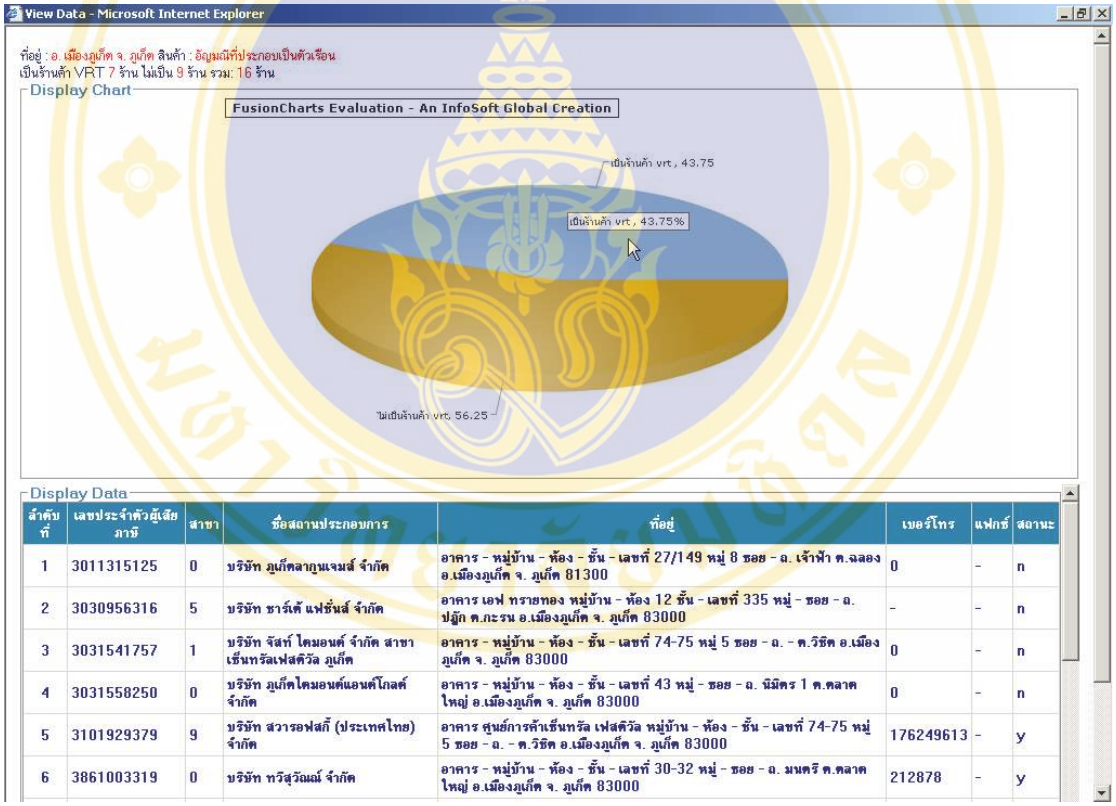
Rule

Support = 0.06 Probability = 0.82 LHS=”GOODS=Jewelly” RHS=”TUMBOL=Rassada:Muang:Phuket”

Meaning

Support = 0.06 means Chinese tourists purchased Jewelly at Amphur Muang Phuket 6%

Probability = 0.82 means if Chinese tourists probably purchased Jewelly at Muang Phuket 82% and purchase other goods 18%



ลำดับที่	เลขประจำตัวผู้เสียภาษี	สาขา	ชื่อสถานประกอบการ	ที่อยู่	เบอร์โทร	แฟกซ์	สถานะ
1	3011315125	0	บริษัท ภูเก็ตลากูนเจมส์ จำกัด	อาคาร - หมู่บ้าน - ห้อง - ชั้น - เลขที่ 27/149 หมู่ 8 ซอย - ถ. เจ้าฟ้า ต.ฉลอง อ.เมืองภูเก็ต จ. ภูเก็ต 81300	0	-	n
2	3030956316	5	บริษัท ชาร์เต้ แฟชั่นส์ จำกัด	อาคาร เอฟ ทรายทอง หมู่บ้าน - ห้อง 12 ชั้น - เลขที่ 335 หมู่ - ซอย - ถ. ปฏัก ต.กะรน อ.เมืองภูเก็ต จ. ภูเก็ต 83000	-	-	n
3	3031541757	1	บริษัท จัสท์ โดมอนด์ จำกัด สาขา เซ็นทรัลเฟสติวัล ภูเก็ต	อาคาร - หมู่บ้าน - ห้อง - ชั้น - เลขที่ 74-75 หมู่ 5 ซอย - ถ. - ต.วิชิต อ.เมืองภูเก็ต จ. ภูเก็ต 83000	0	-	n
4	3031558250	0	บริษัท ภูเก็ตโดมอนด์แอนด์โกลด์ จำกัด	อาคาร - หมู่บ้าน - ห้อง - ชั้น - เลขที่ 43 หมู่ - ซอย - ถ. นิมิตร 1 ต.ตลาดใหญ่ อ.เมืองภูเก็ต จ. ภูเก็ต 83000	0	-	n
5	3101929379	9	บริษัท สวารอฟสกี้ (ประเทศไทย) จำกัด	อาคาร ศูนย์การค้าเซ็นทรัล เฟสติวัล หมู่บ้าน - ห้อง - ชั้น - เลขที่ 74-75 หมู่ 5 ซอย - ถ. - ต.วิชิต อ.เมืองภูเก็ต จ. ภูเก็ต 83000	176249613	-	y
6	3861003319	0	บริษัท ทวิสุวัฒณ์ จำกัด	อาคาร - หมู่บ้าน - ห้อง - ชั้น - เลขที่ 30-32 หมู่ - ซอย - ถ. มนตรี ต.ตลาดใหญ่ อ.เมืองภูเก็ต จ. ภูเก็ต 83000	212878	-	y

Figure 5.19 Information of shops in/not in VRT system of each rule

Figure 5.19 Display information of VAT Entrepreneurs’ shops whish sales jewellery both are member and not member of VRT system according to Rule in figure 5.18. Officers use this information for induce Entrepreneurs to be member of VRT system.

5.2.7 System Log out

User click “ออกจากระบบ” to Log out from ACRM-VRT.

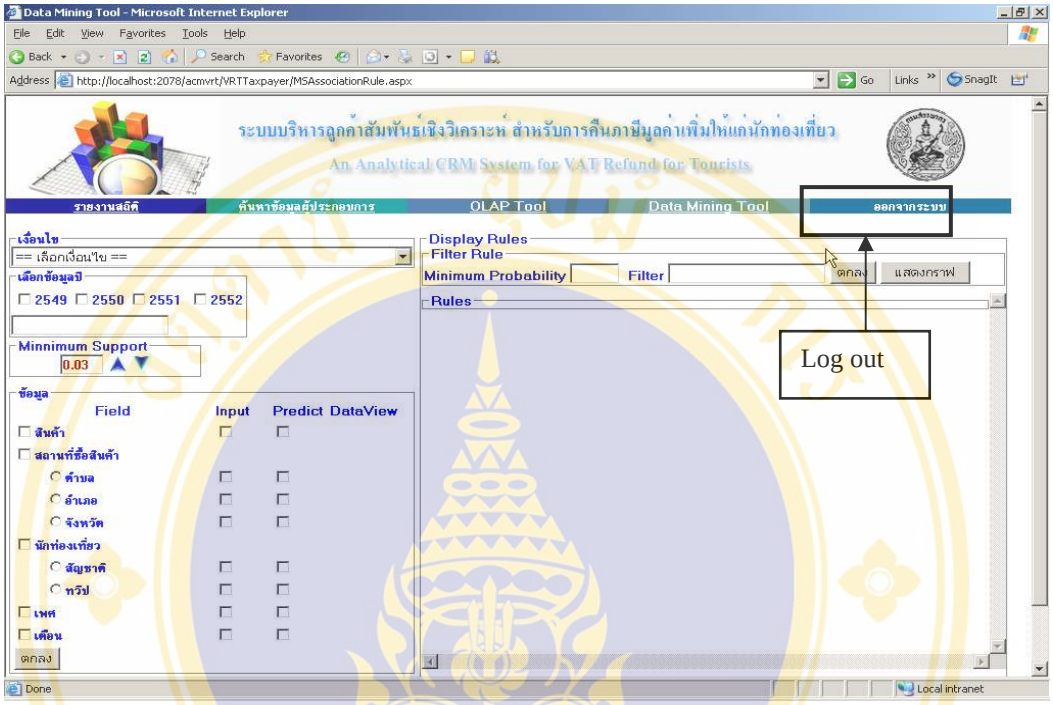


Figure 5.20 Log out

5.3 Example of Using OLAP Tool and Data Mining Tool

This part shows examples of using OLAP Tool and Data Mining Tool. OLAP Tool has 13 examples and Data Mining Tool has 4 examples.

5.3.1 Examples of Using OLAP Tool

OLAP Tool can present the result both in graph form and data table. Users can Drill Down and Roll Up data according to they need. Next are 13 examples of how to use OLAP Tool create reports according to the executive needs.

Example 1: The executive want to compared between the increase and the decrease ratio of number of Entrepreneurs and shops in VRT system backward 5 years from year 2547 to year 2551 (or 2004 to 2008). These use the results make a plan to look for new Entrepreneurs and new shop to VRT system.

Step:

1) In Selected Cube select “VRTEntrepreneursAndVRTShopCountCube” (see Figure 5.21)
2) In Measures and Dimensions select Report Time Dim then drag Fiscal Year and drop in Categories: then select fiscal year 2547 to fiscal year 2551 (or 2004 to 2008) (see Figure 5.22)
3) Select “PreviousPeriodEntrepreneursGrowth” and “PreviousPeriodShopGrowth” in Measures (see Figure 5.23)
4) The result shows n graph form (see Figure 5.24) and shows in data table (see Figure 5.25)

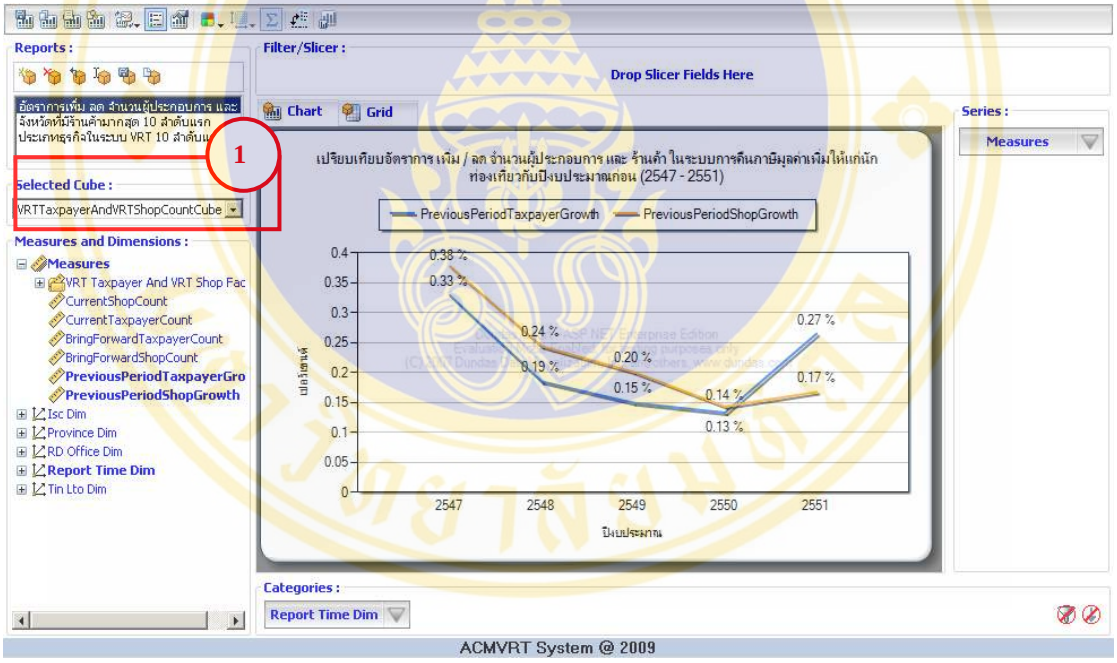


Figure 5.21 Select VRTEntrepreneursAndVRTShopCountCube

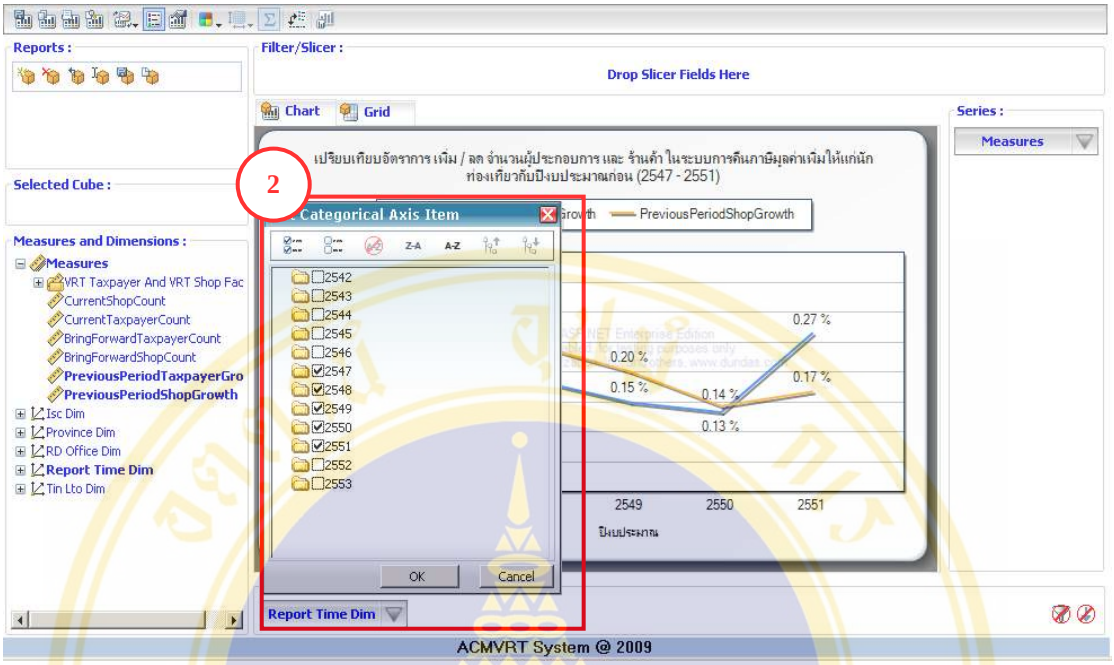


Figure 5.22 Select Year in Report Time Dim Dimension

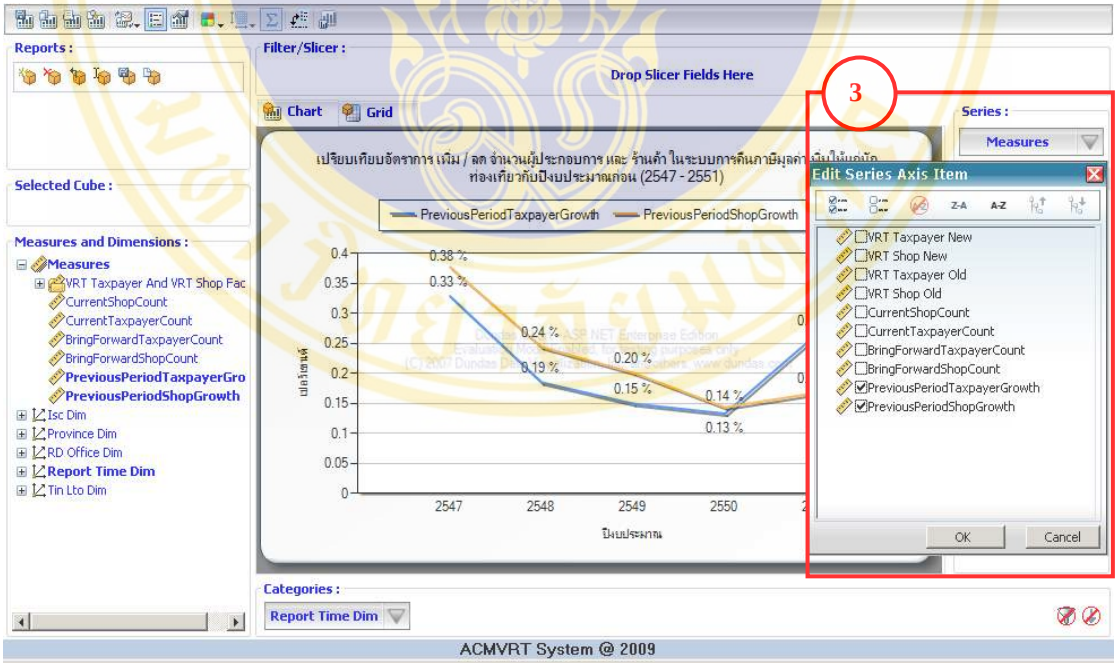


Figure 5.23 Select Measures PreviousPeriodEntrepreneursGrowth and PreviousPeriodShopGrowth

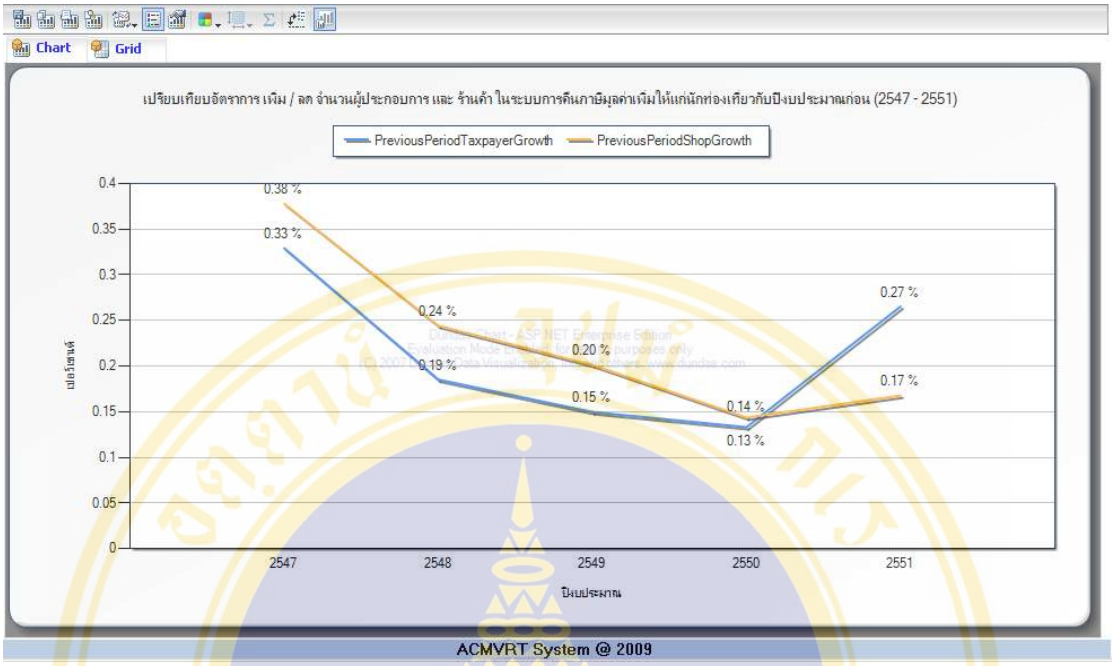


Figure 5.24 Graph of increase/decrease ratio of entrepreneurs and shops in VRT System

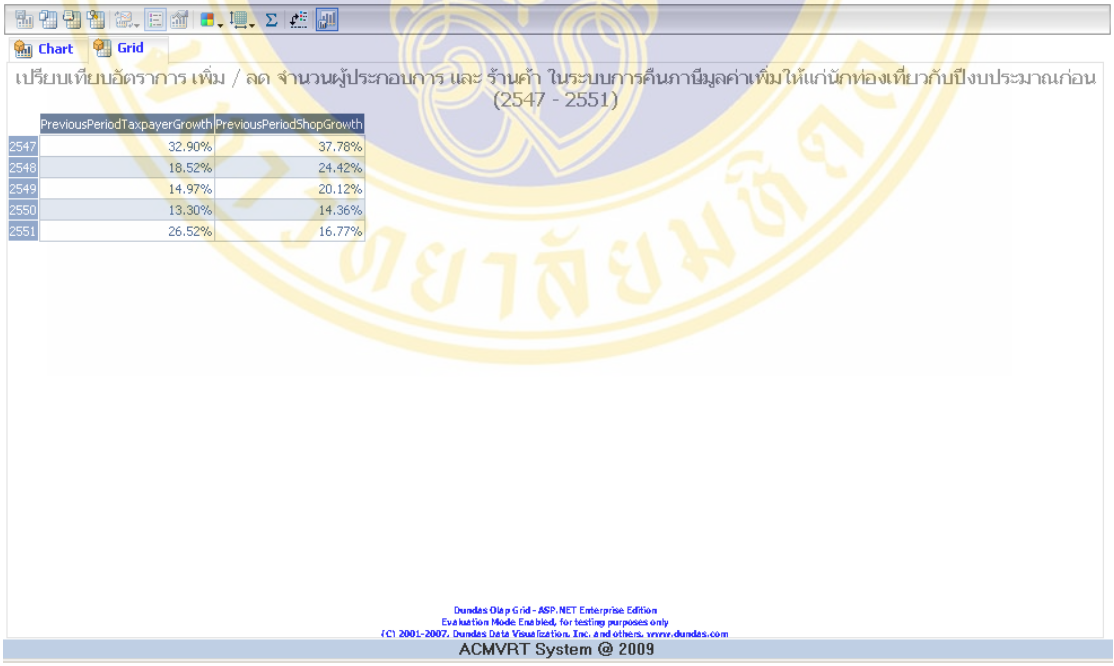


Figure 5.25 Table of increase/decrease ratio of entrepreneurs and shops in VRT system

Example 2: The executive want to know top most 10 provinces which have shops in VRT system in current year. And use the results planning to increase number of entrepreneurs and shops in VRT system.

Step:

1) At Selected Cube select “VRTEntrepreneursAndVRTShopCountCube” (see Figure 5.26).
2) At Measures and Dimensions select “Report Time Dim” then drag Fiscal Year and drop in Filter/Slicer : the select fiscal year 2552 (or 2009) (see Figure 5.27).
3) At Measures and Dimensions select “Province Dim” then drag ProvinceHiearchy and drop in Categories: (see Figure 5.28).
4) At Measures select Current Shop Count (see Figure 5.28).
5) Right click on Sort and Filter select Filter Top/Bottom and at Filter Type : select Top and 10 (see Figure 5.29).
6) The result shows n graph form (see Figure 5.30) and shows in data table (see Figure 5.31).

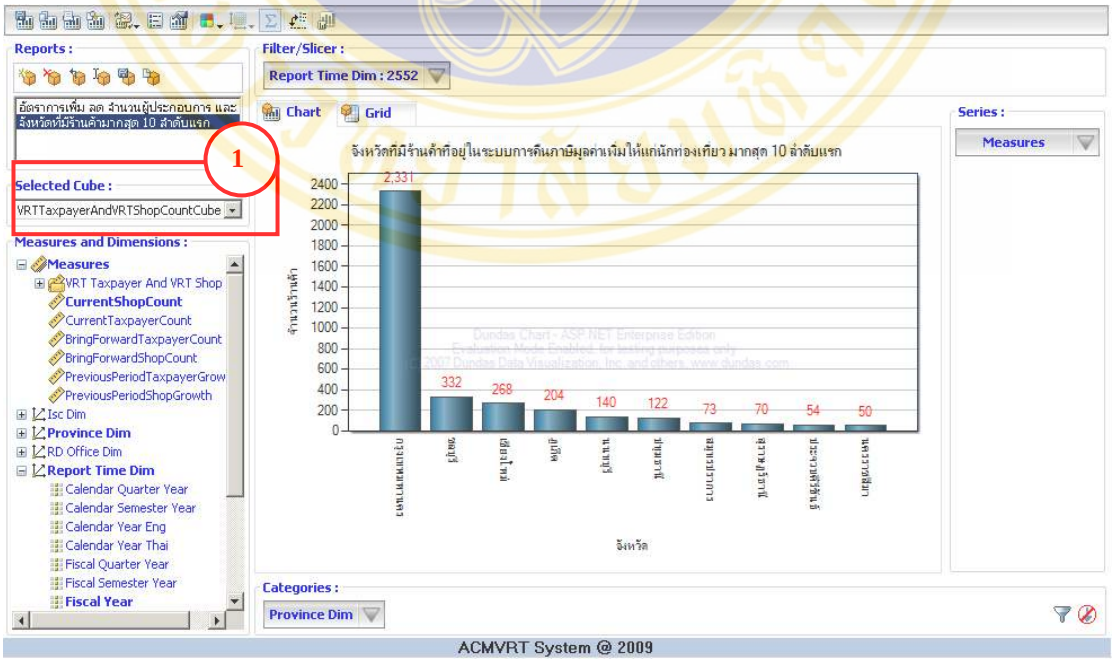


Figure 5.26 Select VRTEntrepreneursAndVRTShopCountCube

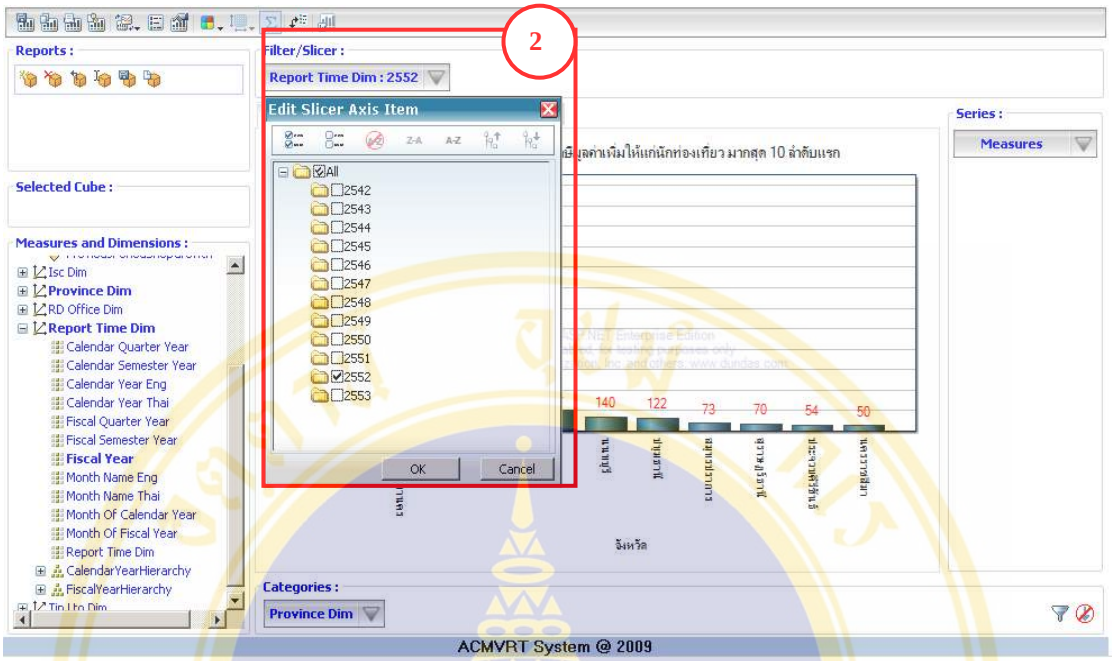


Figure 5.27 Select Report Time Dim

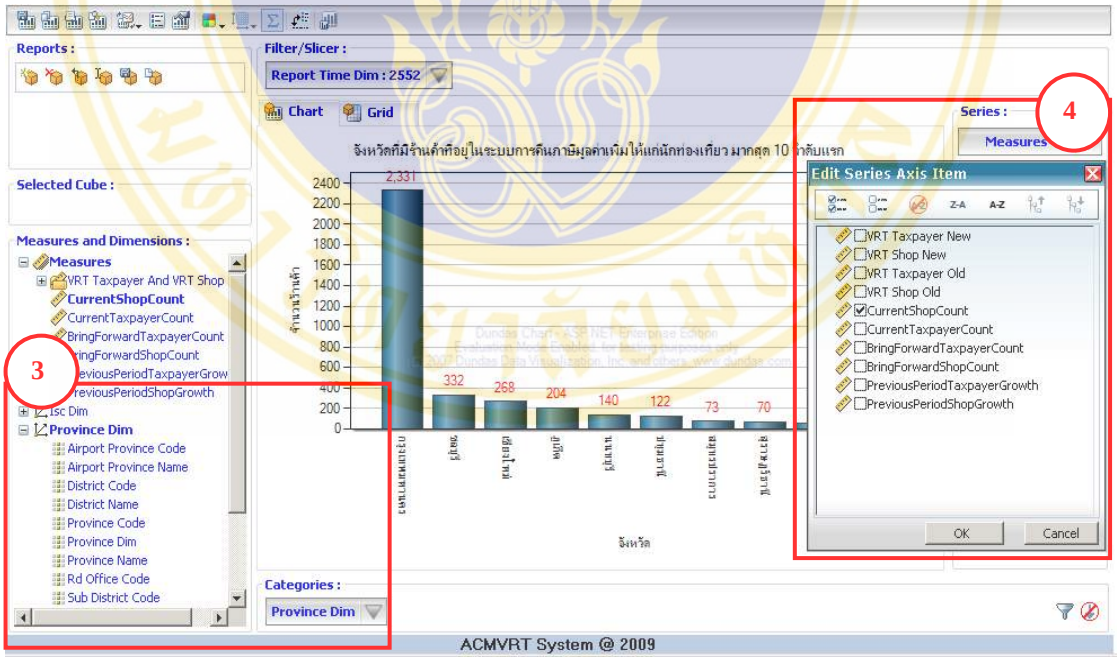


Figure 5.28 Select Province Dim and Current Shop Count at measures

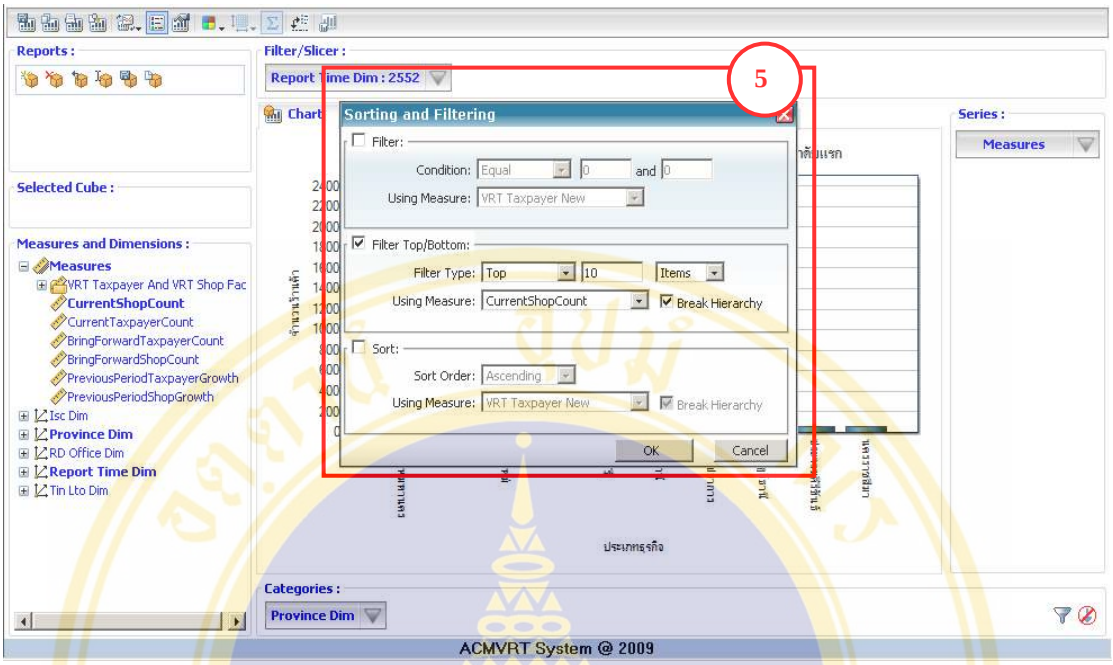


Figure 5.29 Method of filtering top 10 provinces which have the most VRT shops

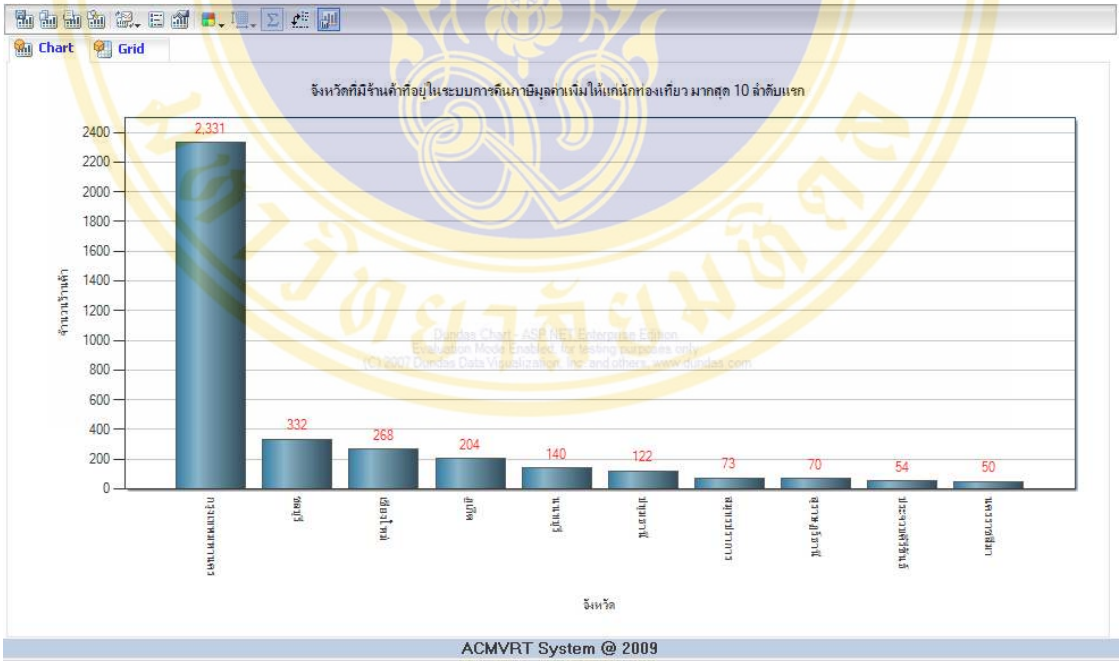


Figure 5.30 Screen of top 10 provinces which have the most VRT shops graph

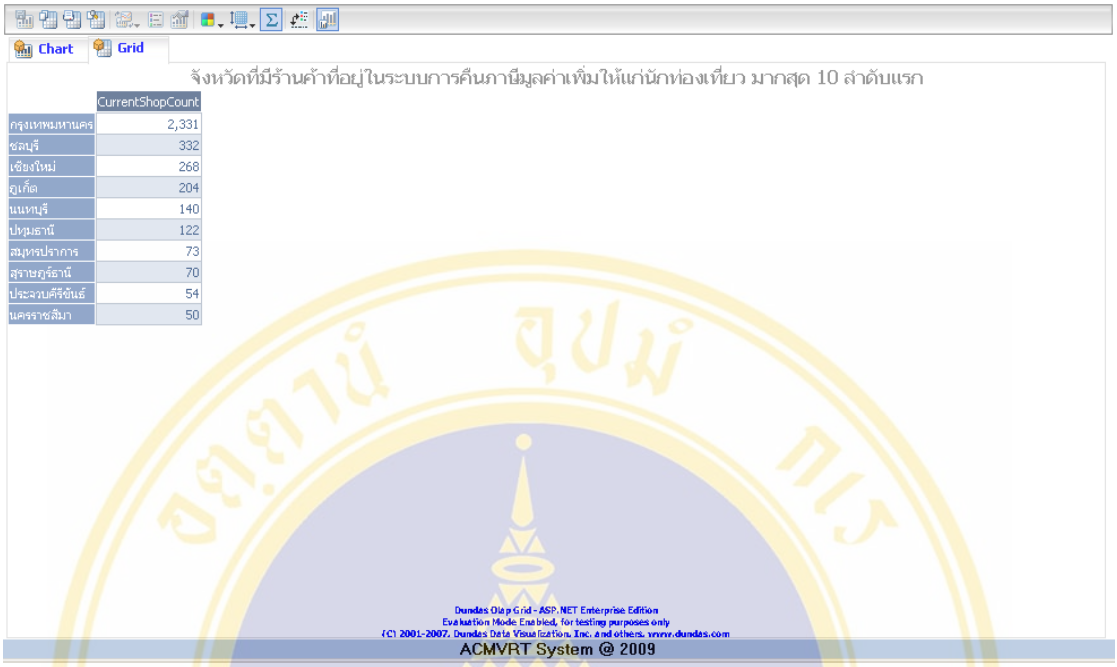


Figure 5.31 Screen of top 10 provinces which have the most VRT shops table

Example 3: executive want to know top ten business type in VRT system. And use the results planning to increase number of entrepreneurs and shops in VRT system.

Step:

1) At Selected Cube select “VRTEntrepreneursAndVRTShopCountCube”.
2) At Measures and Dimensions select “Report Time Dim” then drag Fiscal Year and drop in Filter/Slicer : then select fiscal year 2552 (see Figure 5.33)
3) At Measures and Dimensions select “Isc Dim” then drag Isc Name and drop in Categories: (see Figure 5.34)
4) At Measures select Current Shop Count (see Figure 5.34)
5) Right click Sort and Filter select Filter Top/Bottom
   - At Filter Type : select Top and 10
   - At Using Measure : select “Current Shop Count” then click OK (see Figure 5.35)
6) The result shows in graph form (see Figure 5.36) and shows in data table (see Figure 5.37)

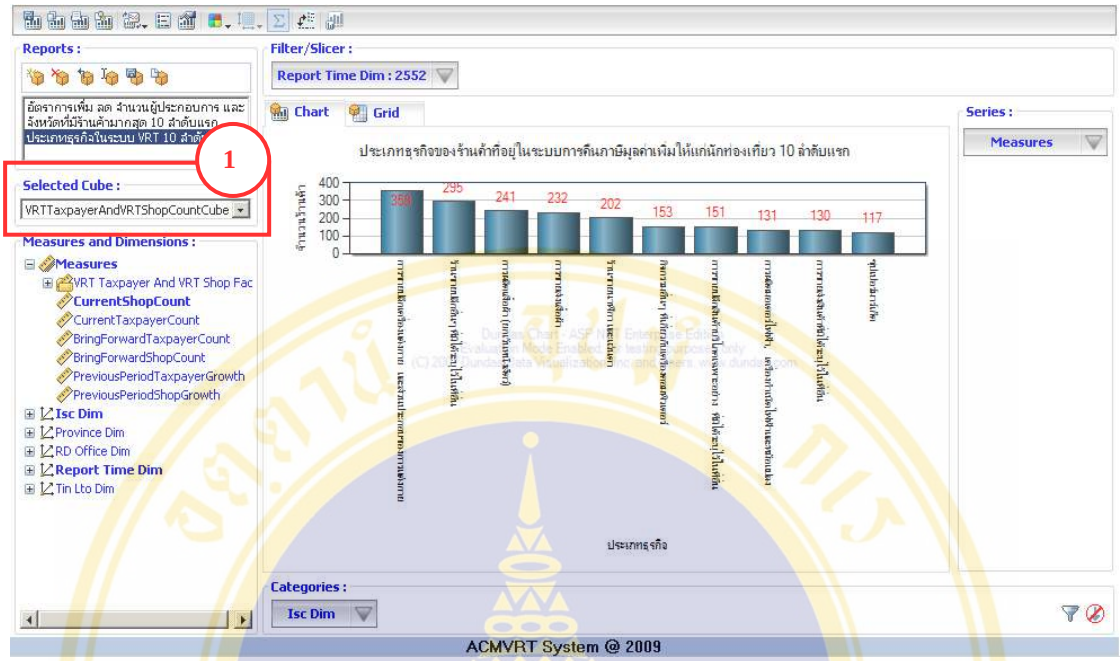


Figure 5.32 Select VRTEntrepreneursAndVRTShopCountCube

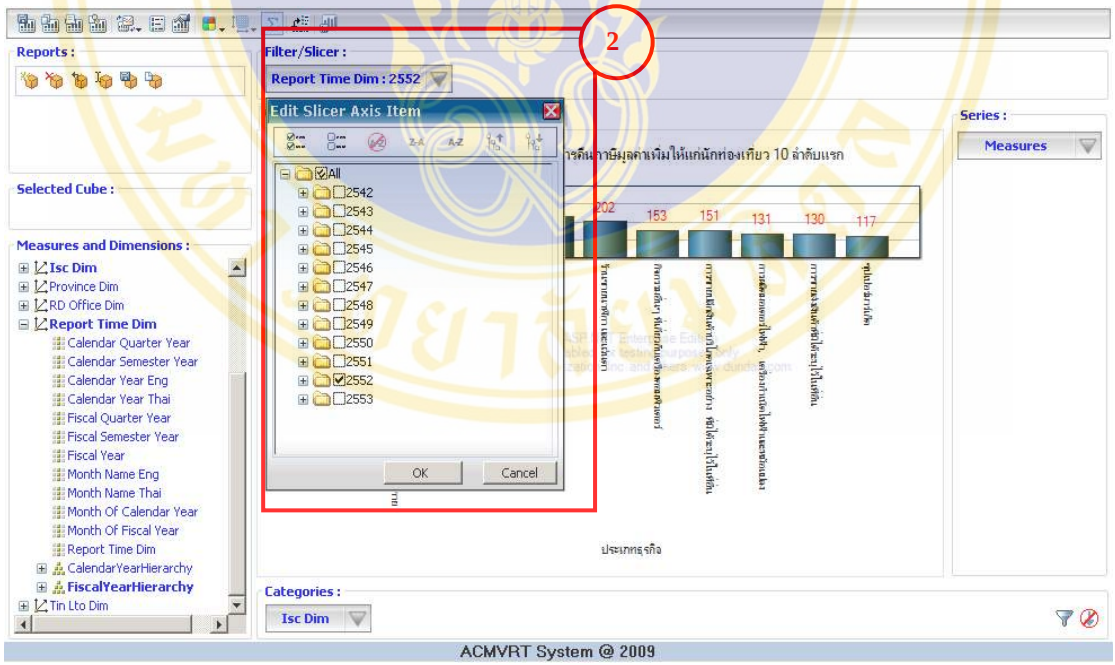


Figure 5.33 Select Report Time Dim

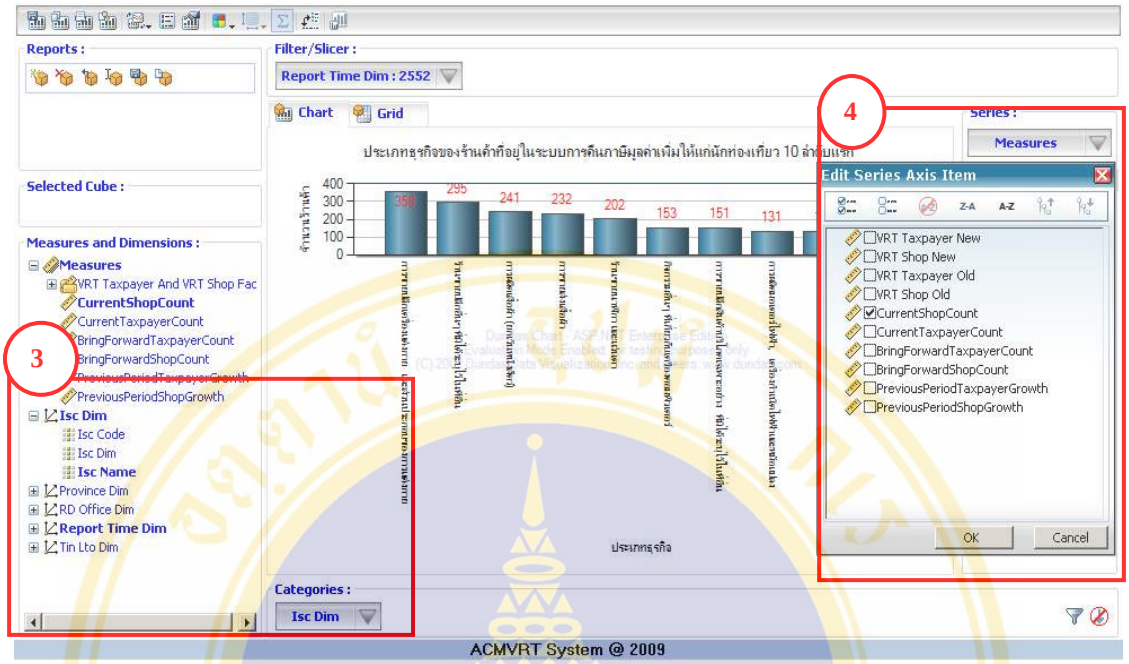


Figure 5.34 Select Isc Dim and Current Shop Count

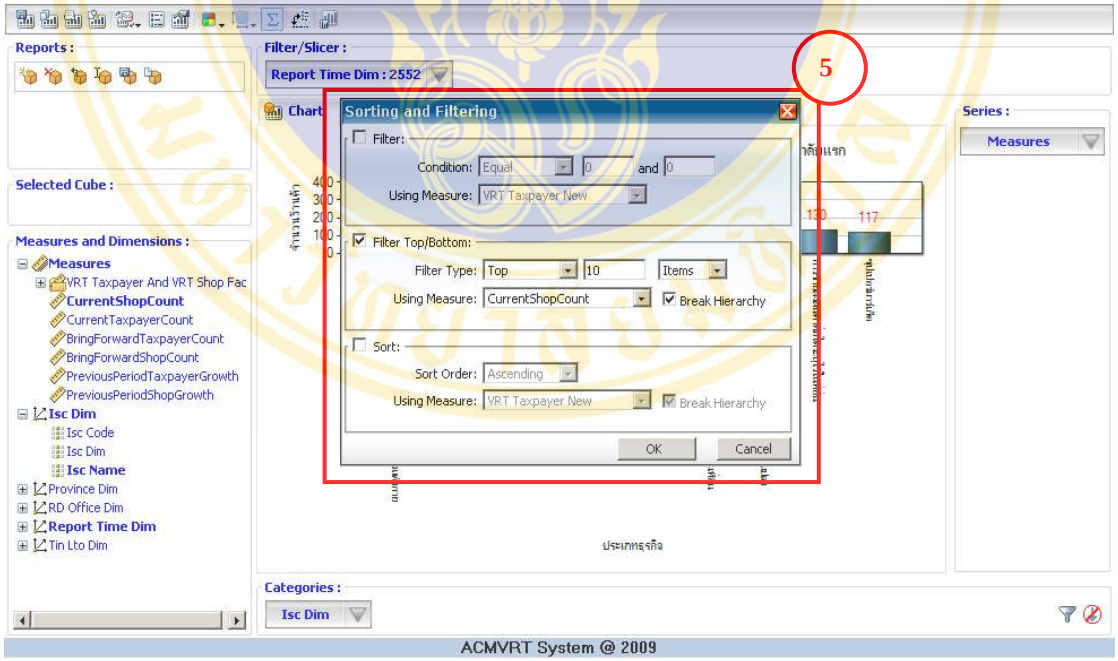


Figure 5.35 Filtering top 10 business types which have the most shops

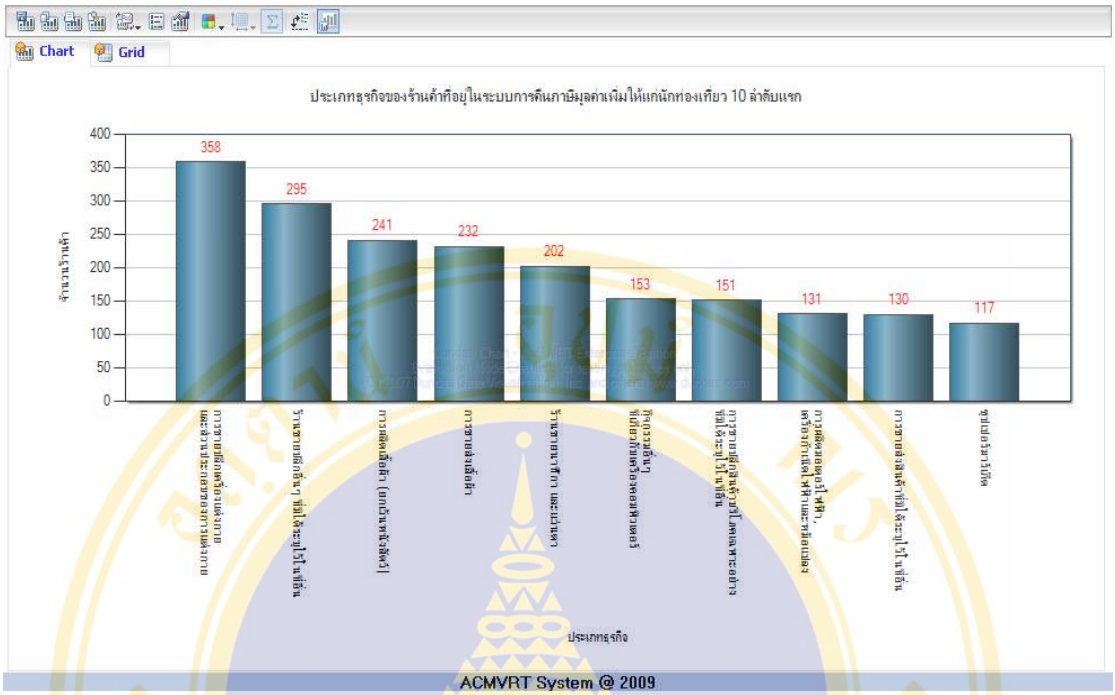


Figure 5.36 Graph show top 10 business types which have the most shops

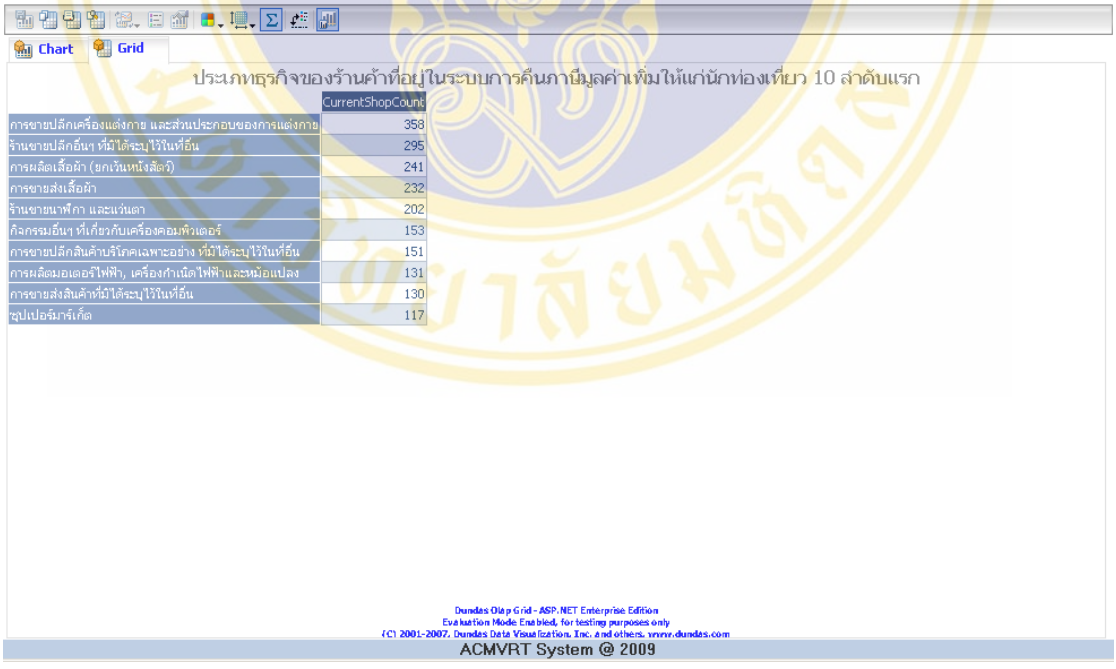


Figure 5.37 Table show top 10 business types which have the most shops

Example 4: The executive want to know total amount of VAT refund to tourists 2 years backward by compared between fiscal year 2550 (or 2007) with fiscal year 2551 (or 2008). And use the results planning to bring the budget to pay VAT refund for tourists.

Step:

1) At Selected Cube select "TouristRefundVATCube" (see Figure 5.38).
2) At Measures and Dimensions select "Refund VAT Time Dim" then drag FiscalYearHierarchy and drop in Categories: (see Figure 5.39).
3) At Measures and Dimensions select "Paid VAT Dim" then drag Paid Name and drop in Series: the select Credit Cash and Draft (see Figure 5.40).
4) At Measures select Total Paid VAT Amount (see Figure 5.41).
5) The result shows in graph form (see Figure 5.42) and shows in data table (see Figure 5.43).

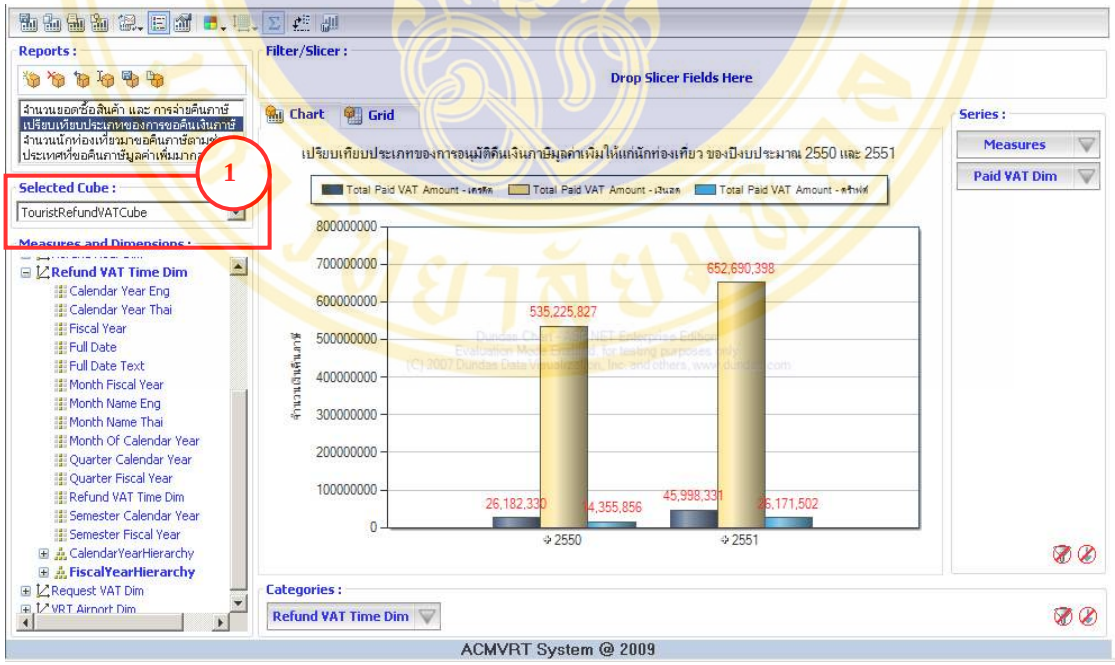


Figure 5.38 Select TouristRefundVATCube

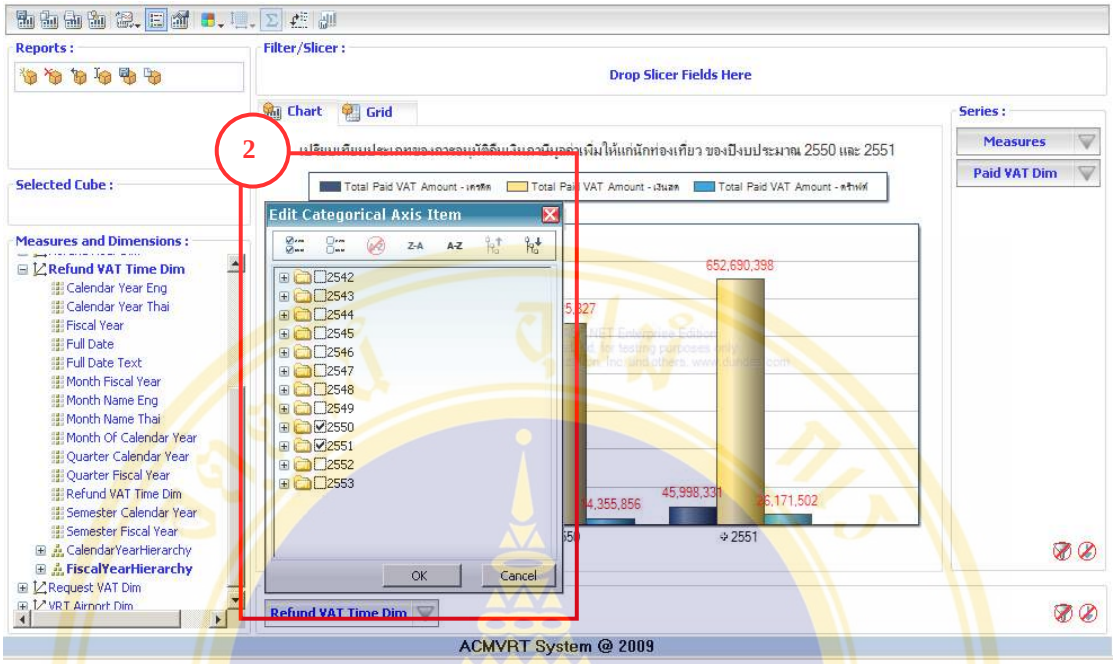


Figure 5.39 Select Refund VAT Time Dim and FiscalYearHierarchy

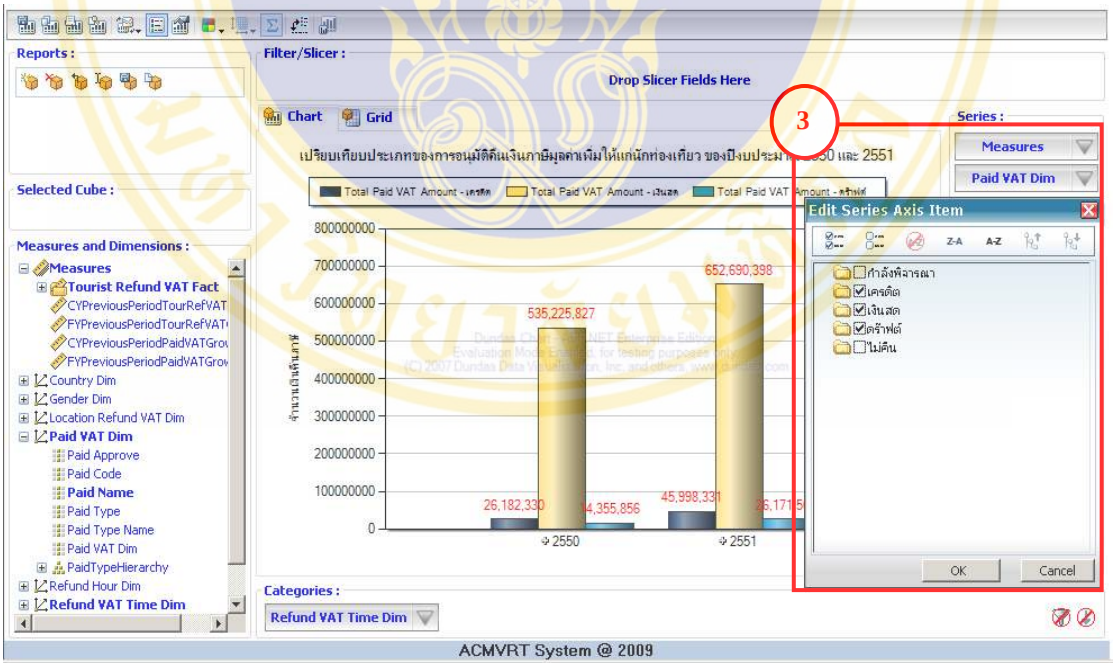


Figure 5.40 Select Paid VAT Dim and Paid Name

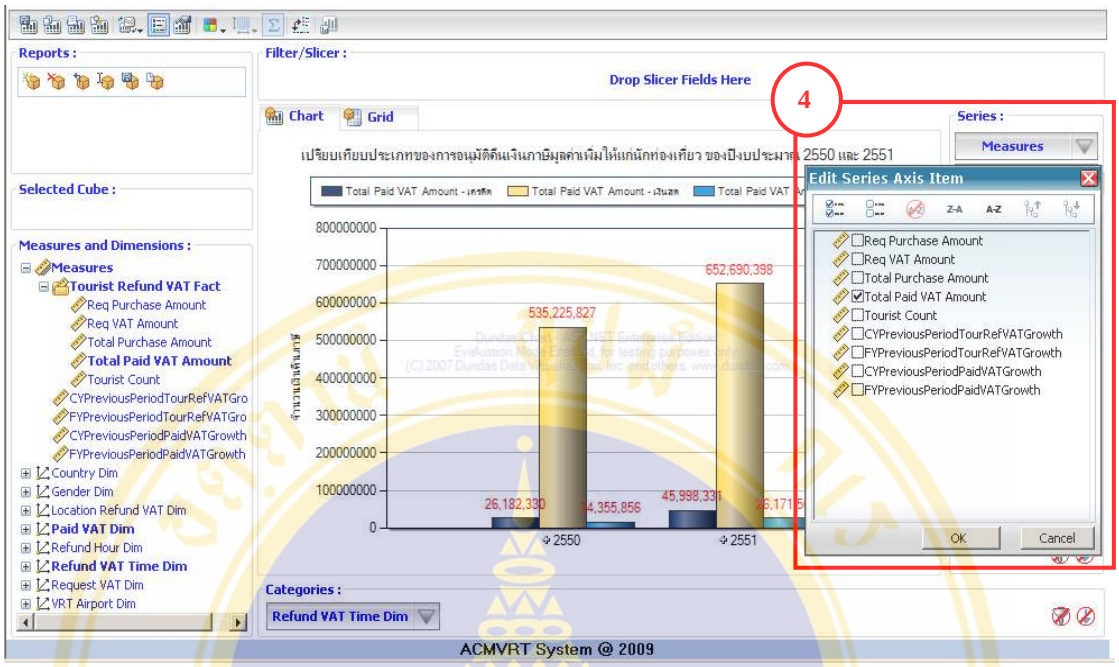


Figure 5.41 Select Total Paid VAT Amount

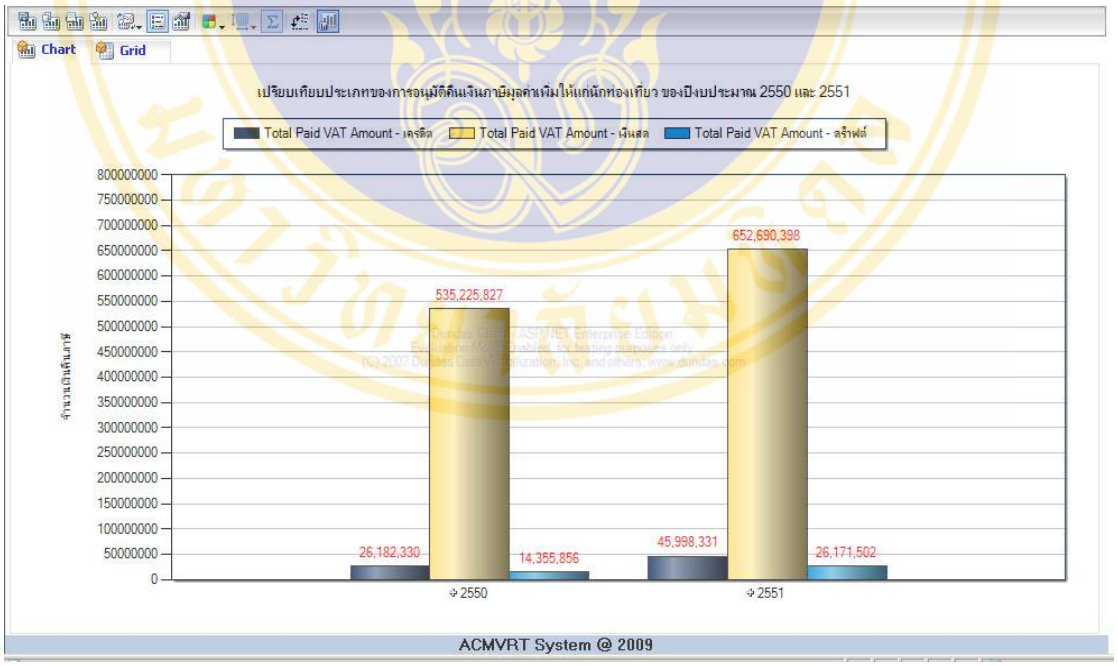


Figure 5.42 Graph show refund type of year 2008 to 2009

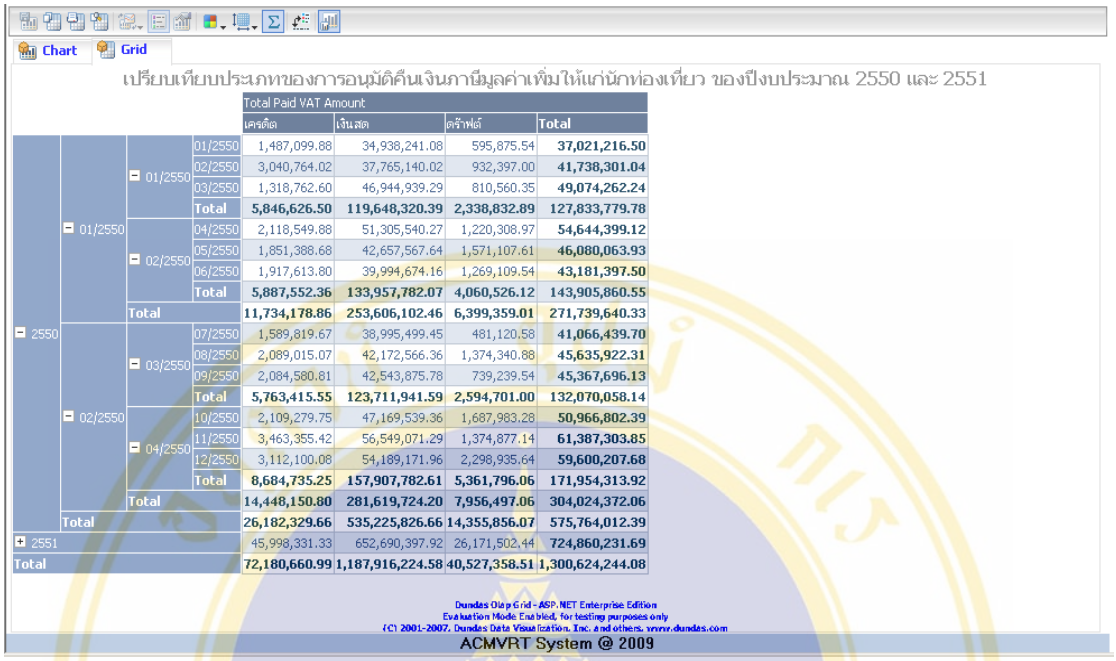


Chart Grid

เปรียบเทียบประเภทของการอนุมัติคืนเงินภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยว ของปีงบประมาณ 2550 และ 2551

				Total Paid VAT Amount			
				เครดิต	เงินสด	ดร้าฟต์	Total
2550	01/2550	01/2550	01/2550	1,487,099.88	34,938,241.08	595,875.54	37,021,216.50
			02/2550	3,040,764.02	37,765,140.02	932,397.00	41,738,301.04
			03/2550	1,318,762.60	46,944,939.29	810,560.35	49,074,262.24
			Total	5,846,626.50	119,648,320.39	2,338,832.89	127,833,779.78
		02/2550	04/2550	2,118,549.88	51,305,540.27	1,220,308.97	54,644,399.12
			05/2550	1,851,388.68	42,657,567.64	1,571,107.61	46,080,063.93
			06/2550	1,917,613.80	39,994,674.16	1,269,109.54	43,181,397.50
			Total	5,887,552.36	133,957,782.07	4,060,526.12	143,905,860.55
		Total		11,734,178.86	253,606,102.46	6,399,359.01	271,739,640.33
	02/2550	03/2550	07/2550	1,589,819.67	38,995,499.45	481,120.58	41,066,439.70
			08/2550	2,089,015.07	42,172,566.36	1,374,340.88	45,635,922.31
			09/2550	2,084,580.81	42,543,875.78	739,239.54	45,367,696.13
			Total	5,763,415.55	123,711,941.59	2,594,701.00	132,070,058.14
		04/2550	10/2550	2,109,279.75	47,169,539.36	1,687,983.28	50,966,802.39
			11/2550	3,463,355.42	56,549,071.29	1,374,877.14	61,387,303.85
			12/2550	3,112,100.08	54,189,171.96	2,298,935.64	59,600,207.68
			Total	8,684,735.25	157,907,782.61	5,361,796.06	171,954,313.92
		Total		14,448,150.80	281,619,724.20	7,956,497.06	304,024,372.06
	Total			26,182,329.66	535,225,826.66	14,355,856.07	575,764,012.39
2551				45,998,331.33	652,690,397.92	26,171,502.44	724,860,231.69
Total				72,180,660.99	1,187,916,224.58	40,527,358.51	1,300,624,244.08

Dundas Data Grid - ASP.NET Enterprise Edition
Evaluation Mode Enabled, for testing purposes only
(C) 2001-2007. Dundas Data Visualization, Inc. and others. www.dundas.com

ACMVRT System @ 2009

Figure 5.43 Table show refund type of year 2008 to 2009

Example 5: The executive want to know total amount of VAT refund to tourists divided by continent and fiscal year of 2 years backward by compared between fiscal year 2551 (or 2008) with fiscal year 2552 (or 2009). And use the results planning to bring the budget to pay VAT refund for tourists.

Step:

1) At Selected Cube select TouristRefundVATCube (see Figure 5.44).
2) At Measures and Dimensions select “Country Dim” then drag CountryHierarchy and drop in Categories (see Figure 5.45).
3) At Measures and Dimensions select “RefundVATTime Dim” then drag FiscalYearHierarchy and drop in Series: then select year 2550 and 2551 (or 2007 and 2008) (see Figure 5.46).
4) At Measures select Total Paid VAT Amount (see Figure 5.47) .
5) The result shows in graph form (see Figure 5.48) and shows in data table (see Figure 5.49) and can drill-down by continent and country.

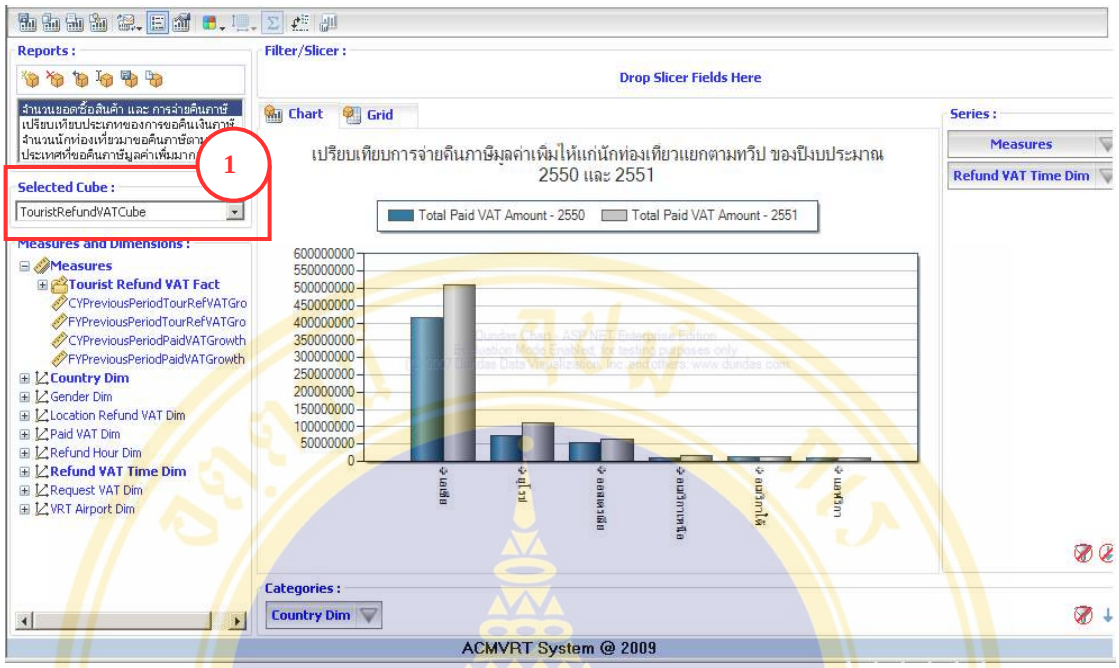


Figure 5.44 Select TouristRefundVATCube

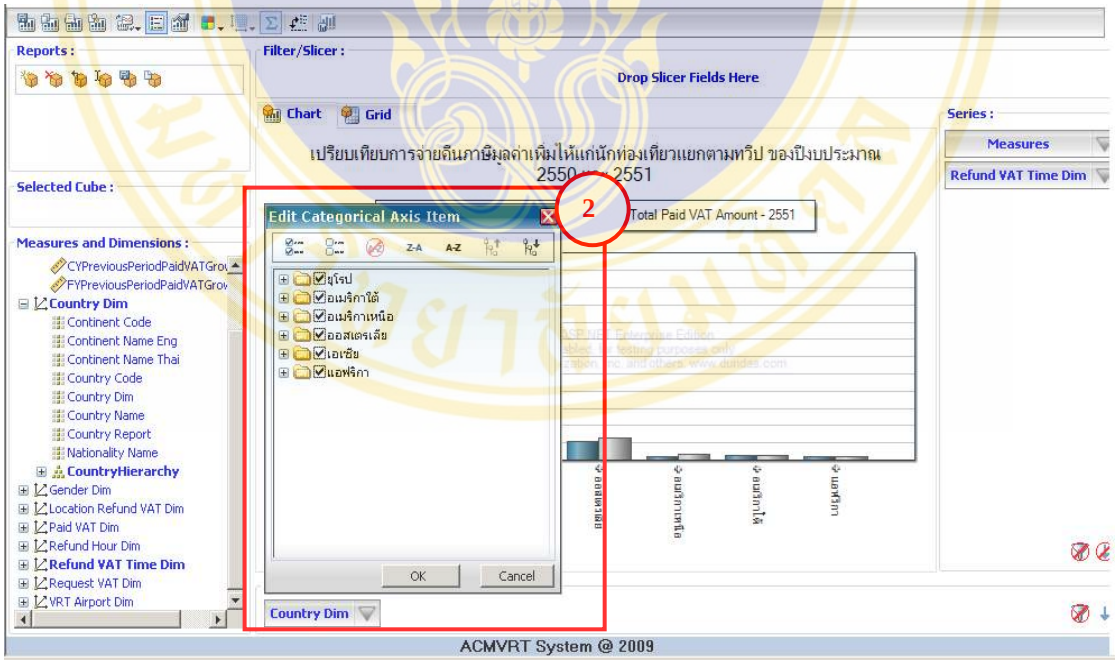


Figure 5.45 Select Dim Country and CountryHierarchy

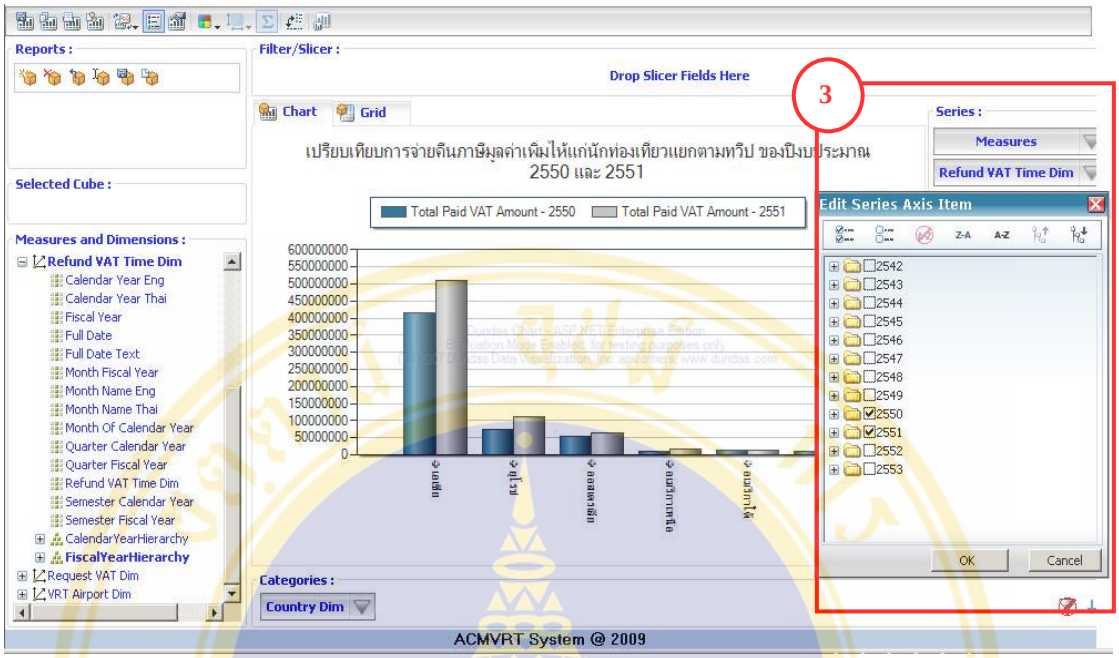


Figure 5.46 Select Dim RefundVATTime and FiscalYearHierarchy

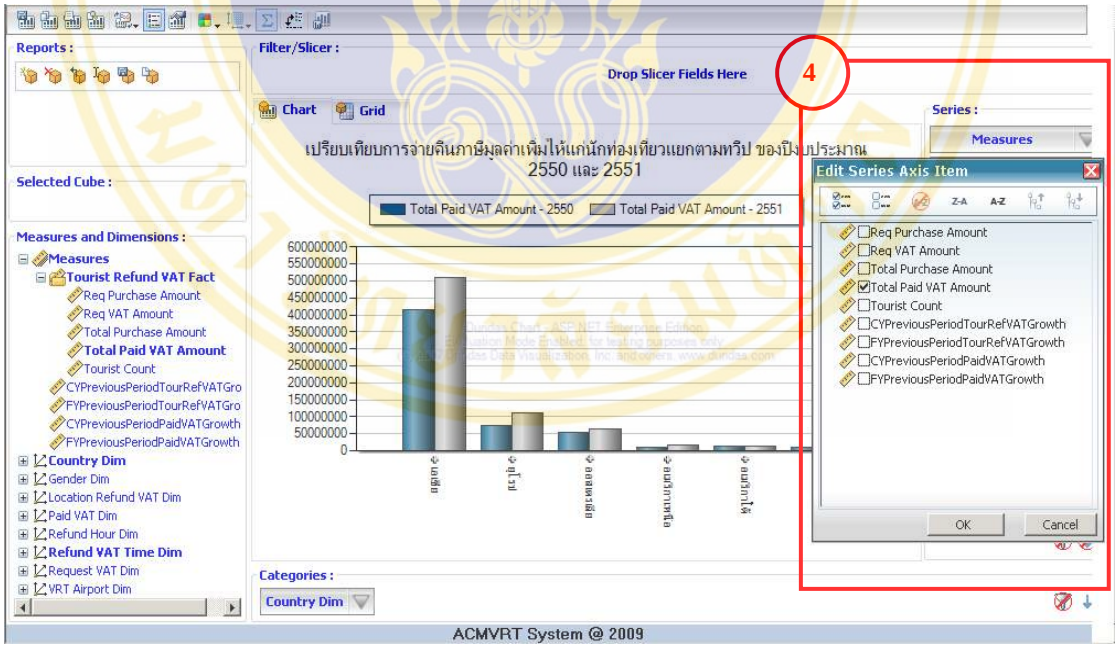


Figure 5.47 Select Total Paid VAT Amount

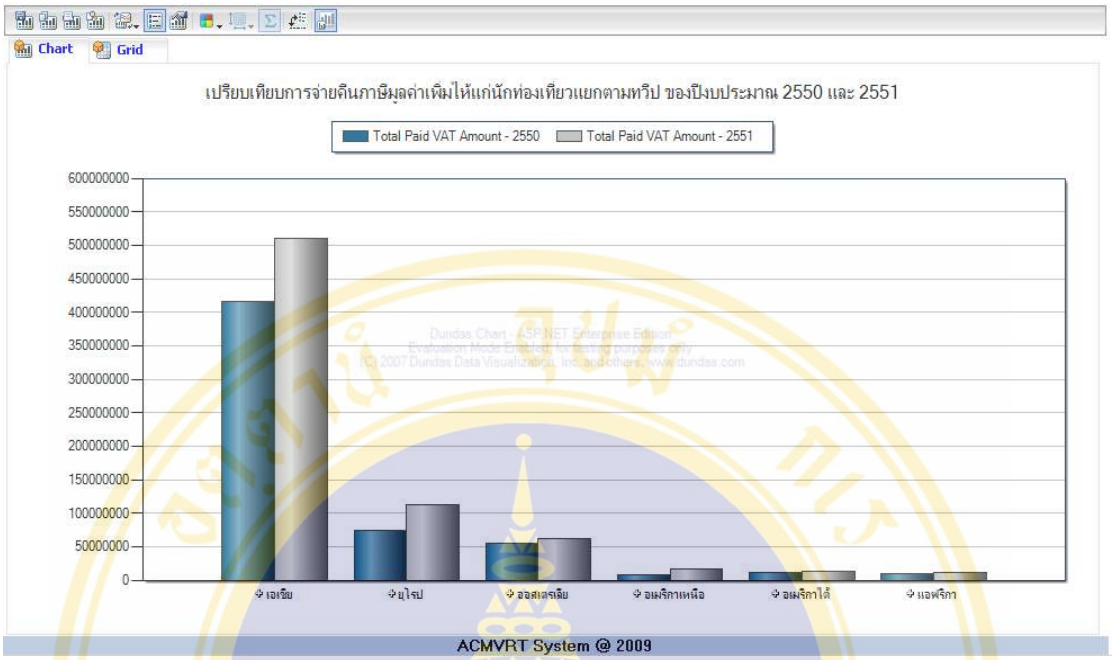


Figure 5.48 Graph show the comparative of VAT refund to tourists from each continent

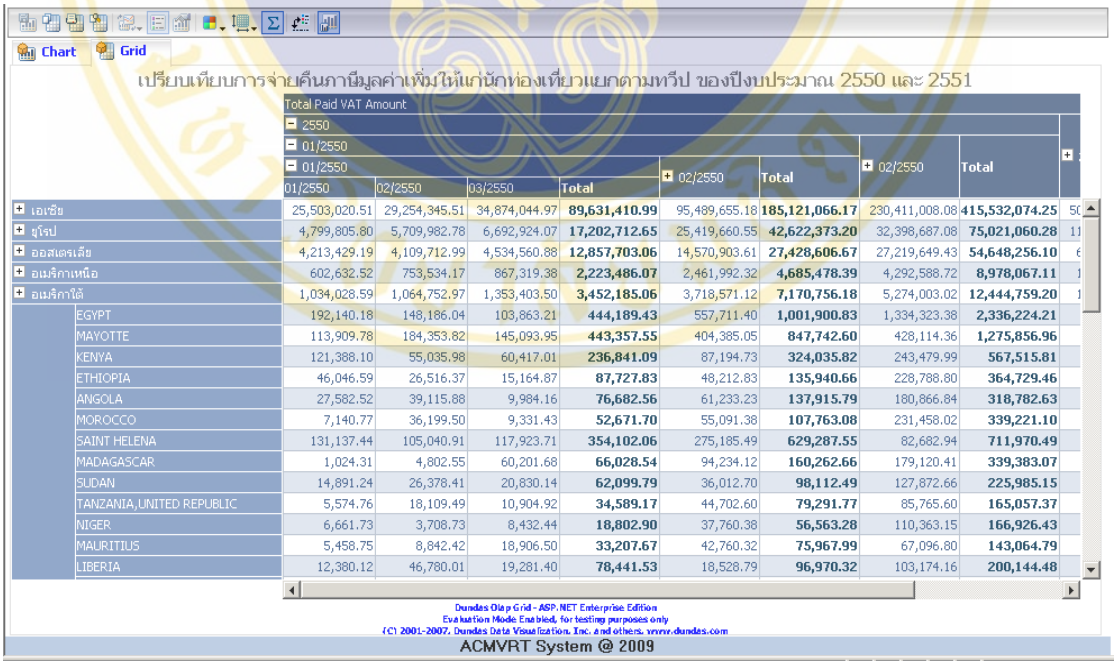


Figure 5.49 Table show data of comparative of VAT refund to tourists from each continent

Example 6: The executive want to know to ten countries which have most total amount VAT refund in year 2551 (or 2008). And use the analysis results plan to pay VAT refund for tourists.

1) At Selected Cube select “TouristRefundVATCube” (see Figure 5.50).
2) At Filter/Slicer: select “Refund VAT Time Dim” then drag Fiscal Year and drop down, then select year 2551 (or 2008) (see Figure 5.51).
3) At Measures and Dimensions select “Country Dim” then drag Country Name and drop in Categories: (see Figure 5.52).
4) At Measures select Tourist Count (see Figure 5.53).
5) At Filter/Slicer click to open Sorting and Filter, then select Top/Bottom :
   - At Filter Type : select Top and 10.
   - At Using Measure : select Req VAT Amount then click OK (see Figure 5.54).
6) The result shows in graph form (see Figure 5.55) and shows in data table (see Figure 5.56).

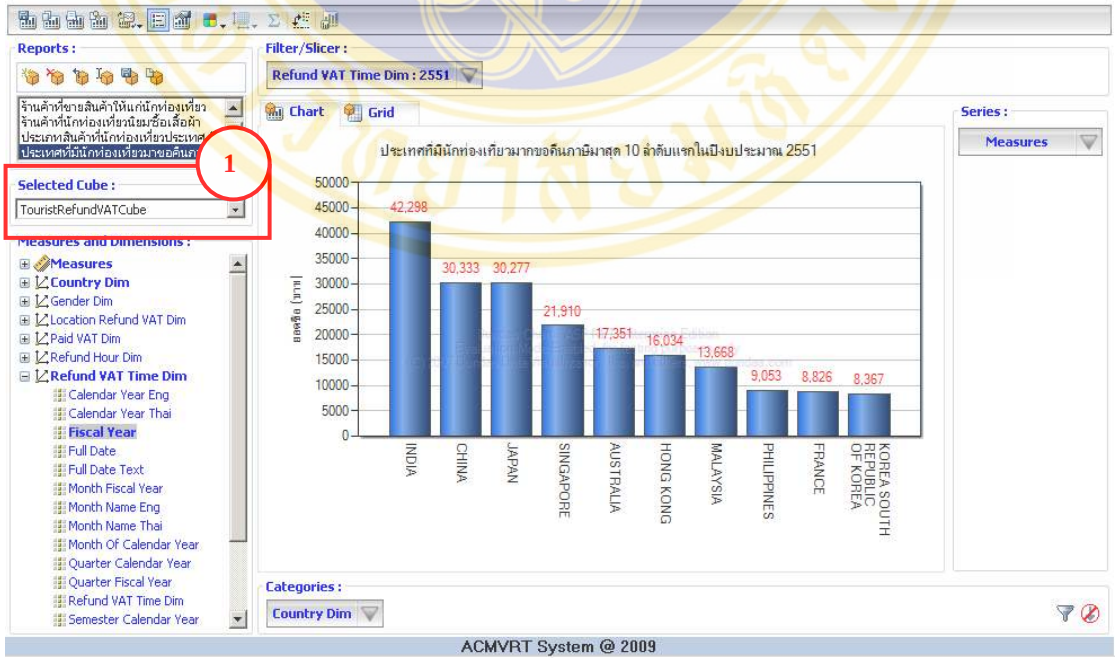


Figure 5.50 Select TouristRefundVATCube

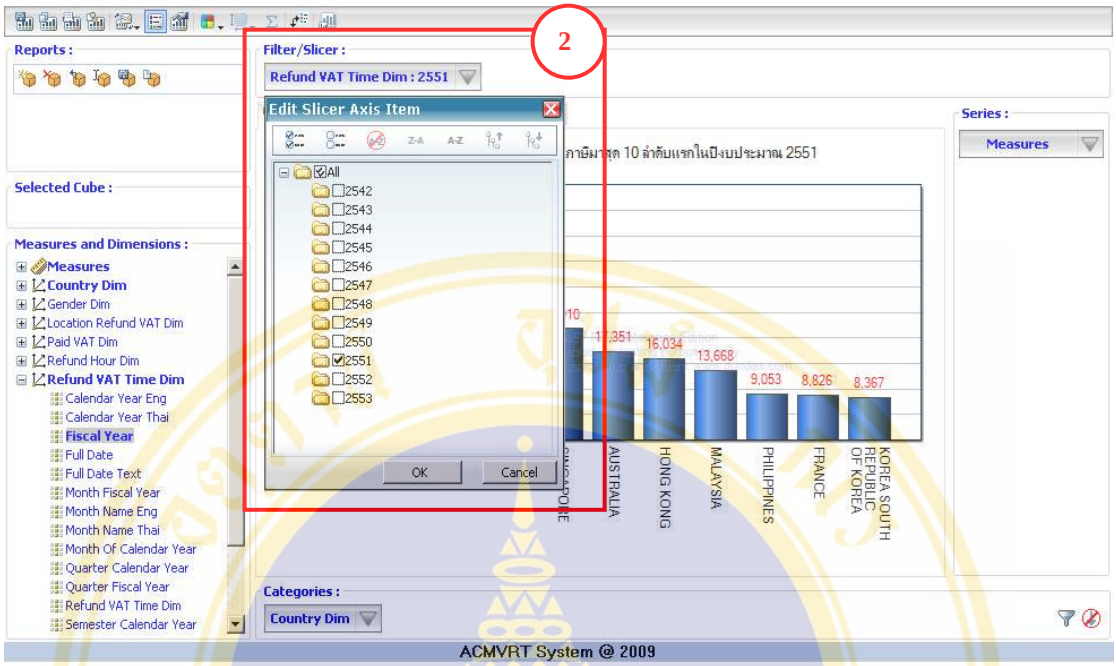


Figure 5.51 Select Refund VAT Time Dim

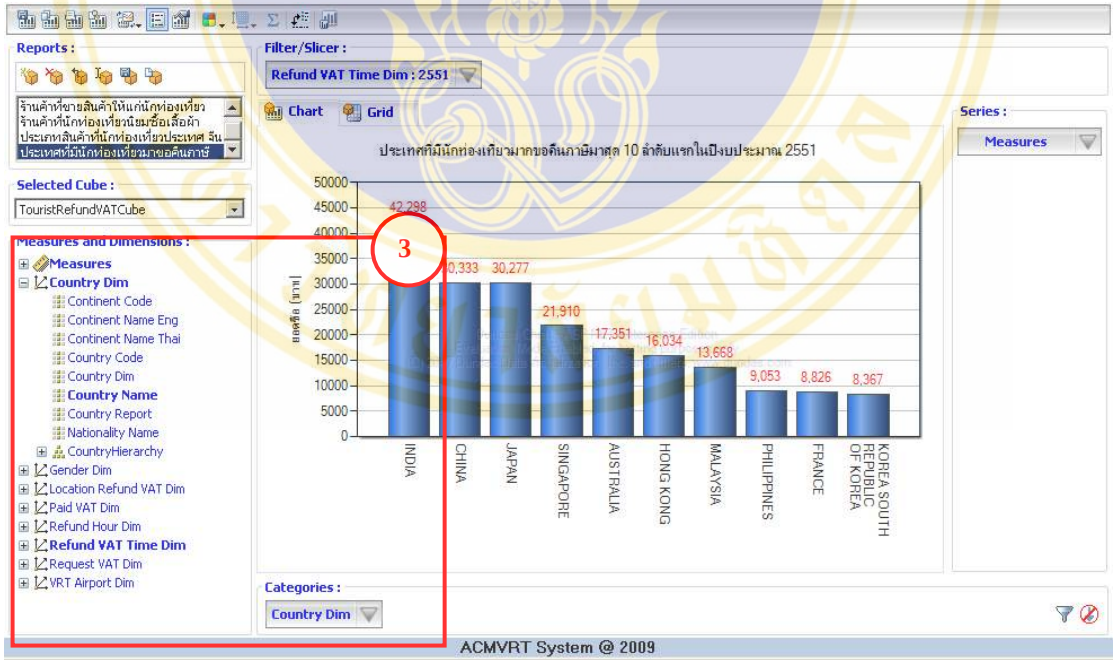


Figure 5.52 Select Country Name

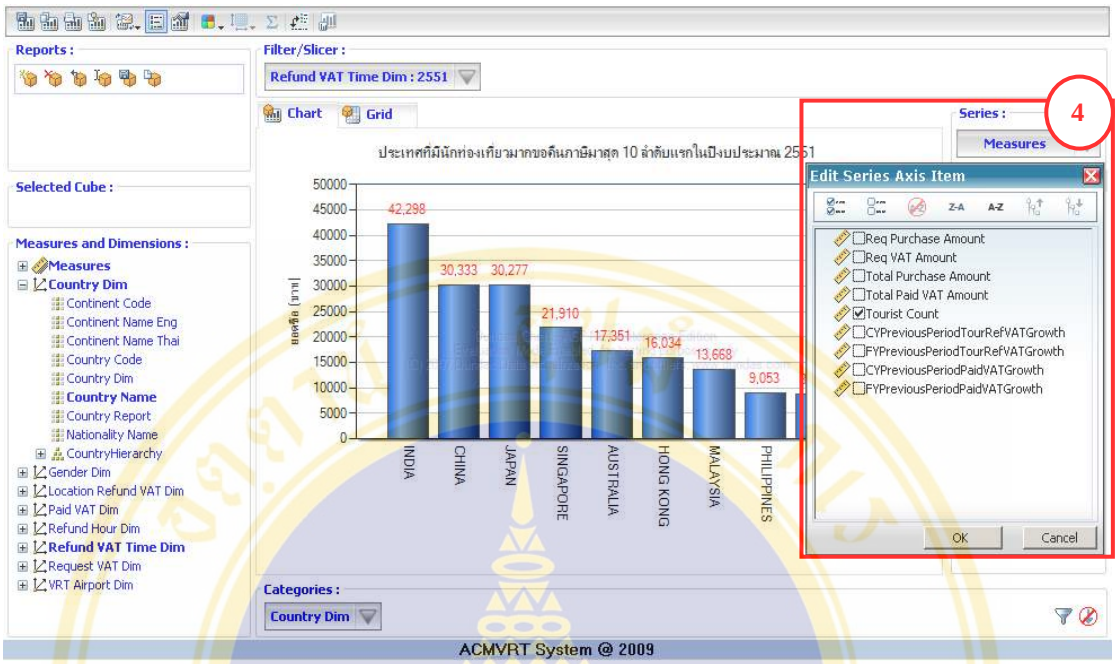


Figure 5.53 Select Measures : Tourist Count

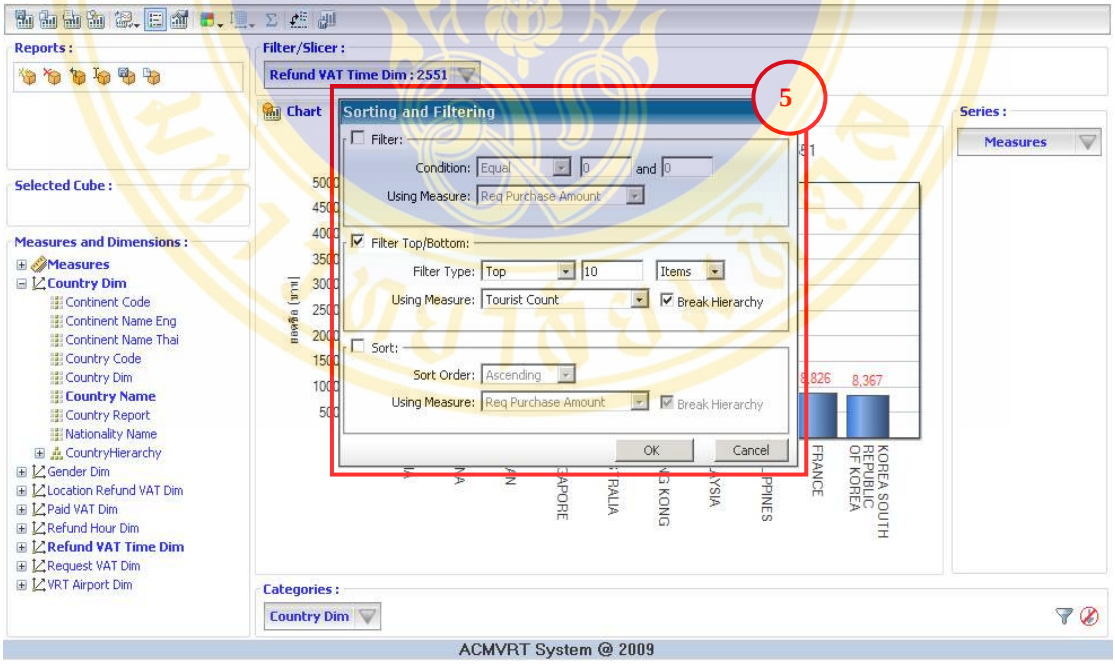


Figure 5.54 Method of filtering data of countries which have most VAT refund

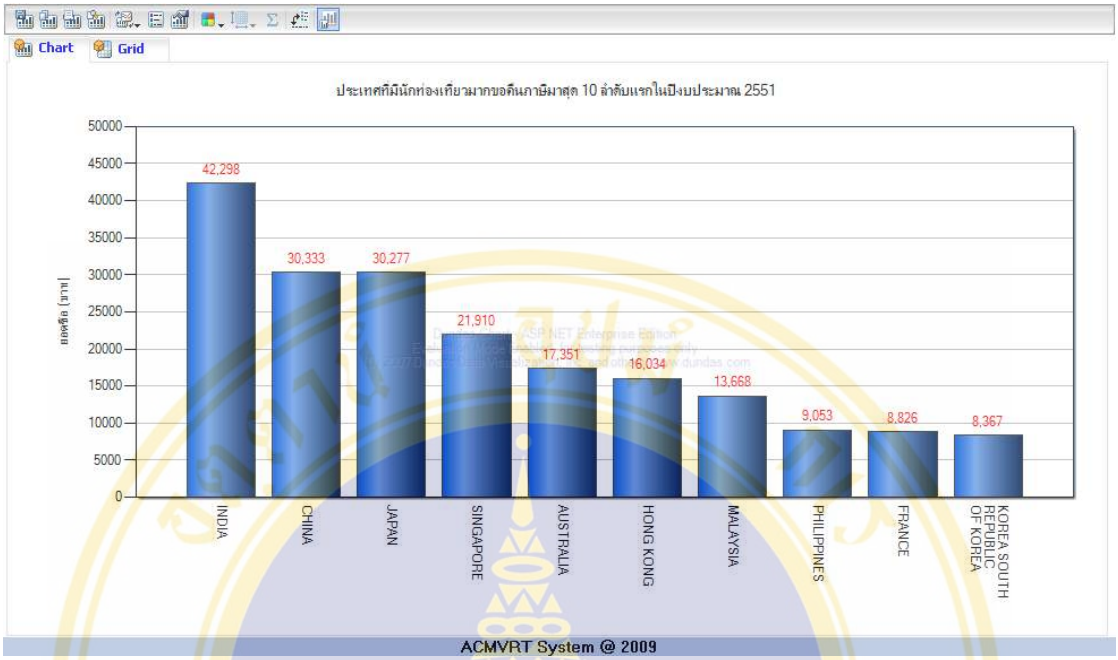


Figure 5.55 Graph show top 10 of countries have most VAT refund

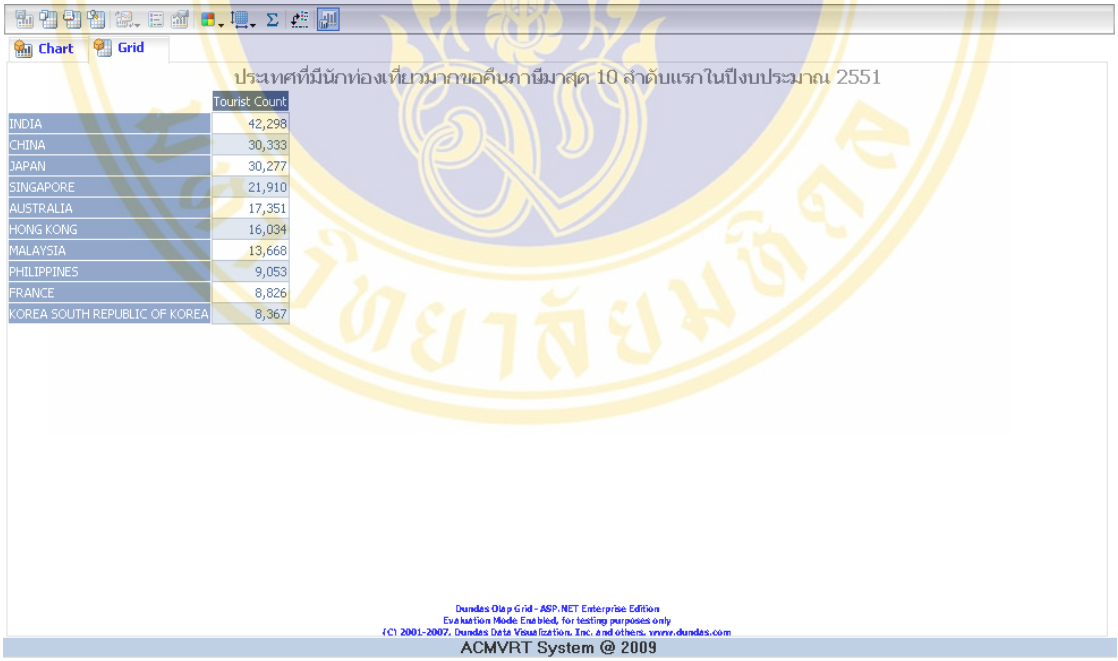


Figure 5.56 Table show top 10 of countries have most VAT refund

Example 7: The executive want to know number of tourists comes to claim VAT refund with VRT-counter at Suwannabhumi Airport in each hour of the day in year 2551 (or 2008), order from more to less. And use the results planning to organise officers for service tourists.

Step:

1) At Selected Cube select “TouristRefundVATCube” (see Figure 5.57).
2) At Filter/Slicer : select “Refund VAT Time Dim” then drag Calendar Year Thai and drop then select year 2551 (or 2008) (see Figure 5.58).
3) At Measures and Dimensions select “Refund Hour Dim” then drag Hour Name and drop in Categories (see Figure 5.59).
4) At Measures and Dimensions select “VRT Airport Dim” then drag AirportHierarchy and drop in Series: (see Figure 5.60).
5) At Measures select Tourist Count (see Figure 5.61).
6) Click on to open Sorting and Filter, then select Sort:
   - At Sort Order : select Descending .
   - At Using Measure : select Req VAT Amount then click OK (see Figure 5.62).
7) The result shows in graph form (see Figure 5.63) and shows in data table (see Figure 5.64).

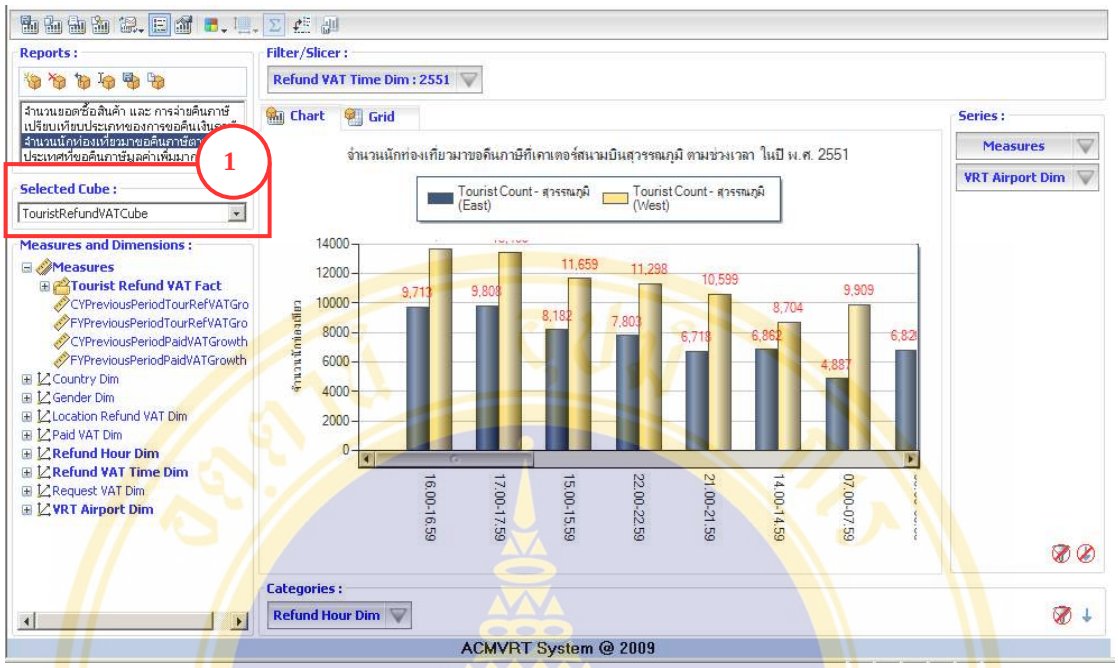


Figure 5.57 Select TouristRefundVATCube

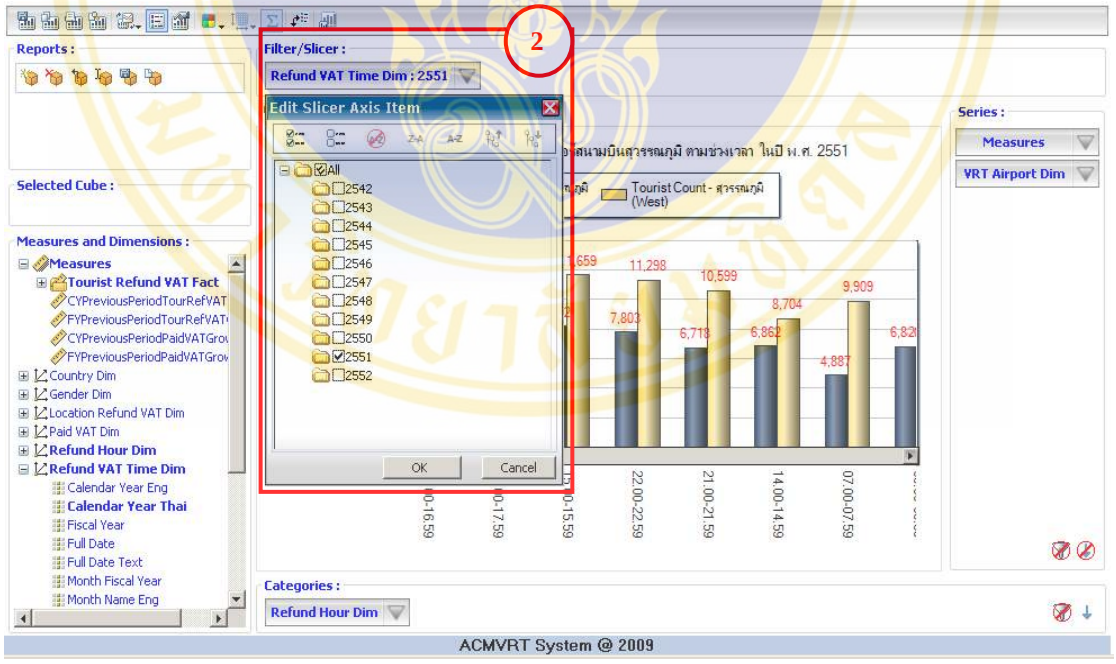


Figure 5.58 Select Refund VAT Time Dim

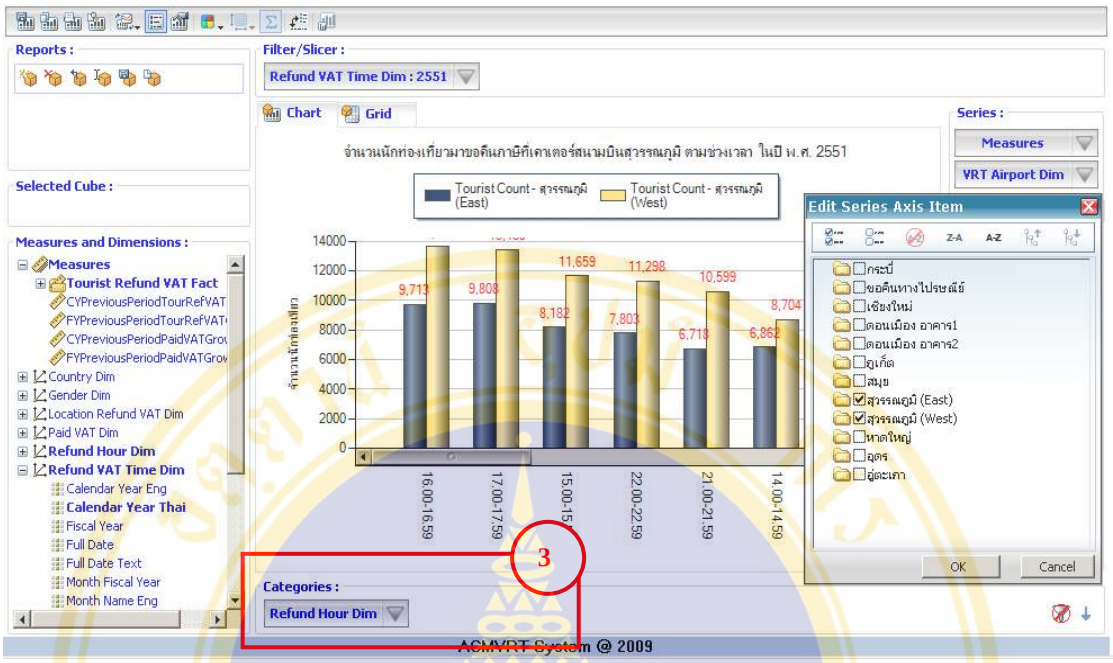


Figure 5.59 Select Refund Hour Dim

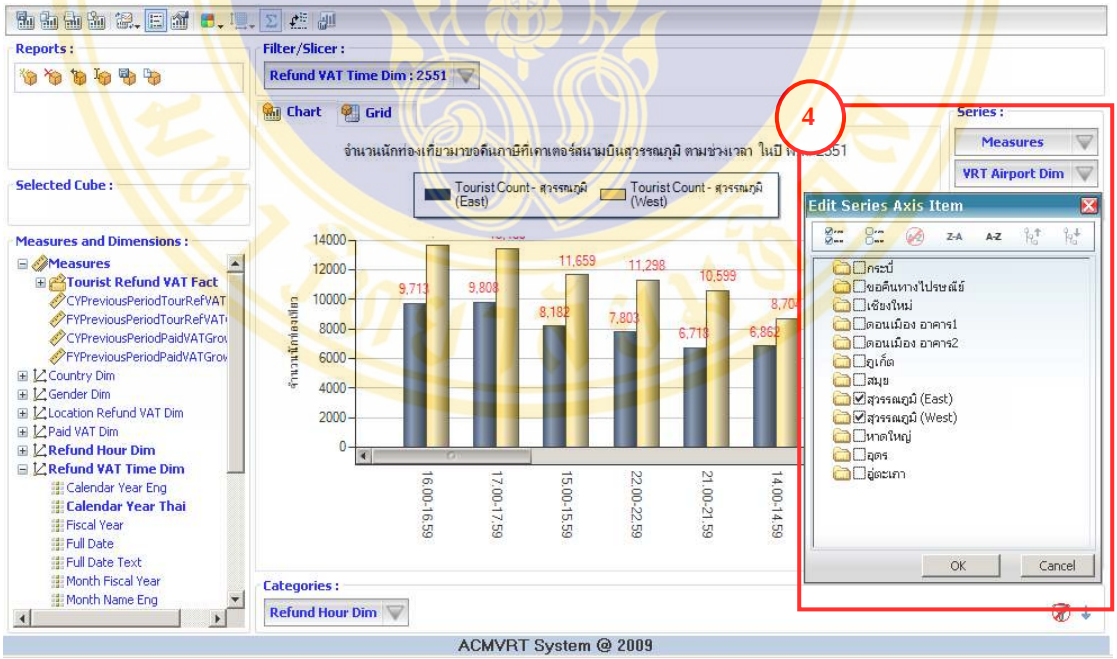


Figure 5.60 Select VRT Airport Dim

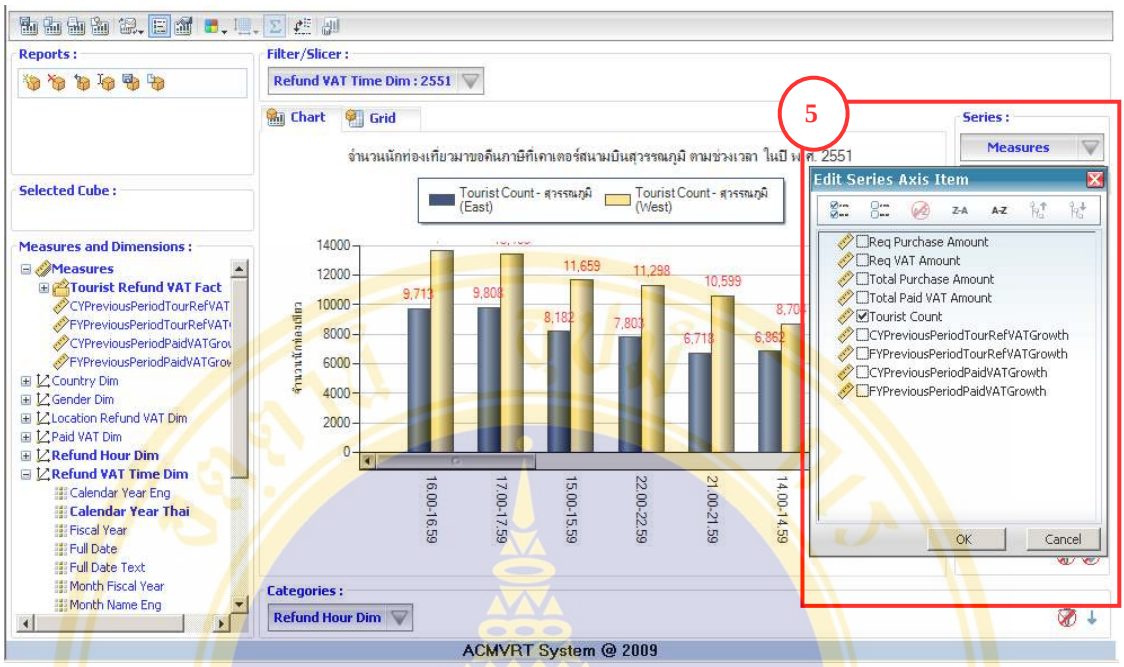


Figure 5.61 Select Measures: Tourist Count

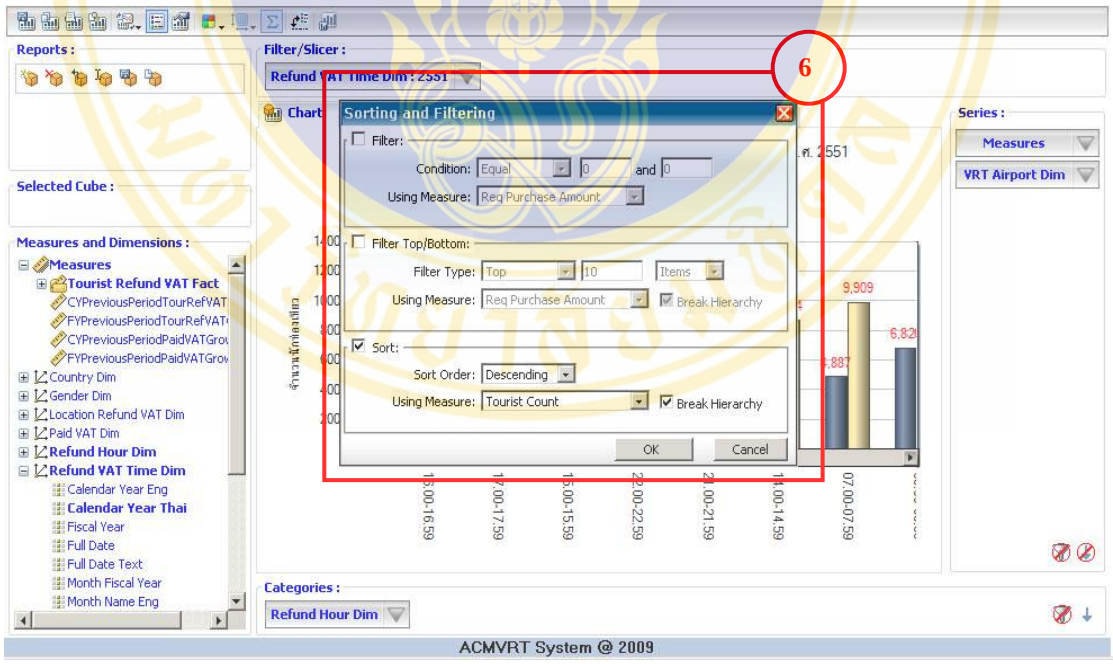


Figure 5.62 Show method to order number of tourists refund VAT in each period from more to less

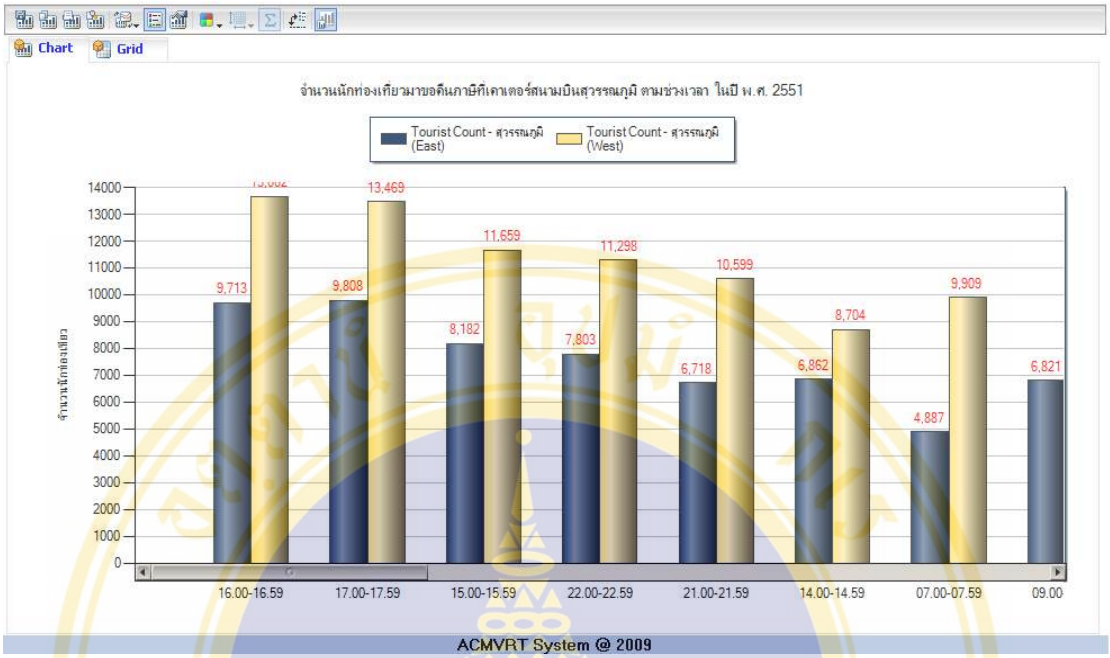


Figure 5.63 Graph show number of tourists refund VAT in each period (1 hour) in one day, order from more to less

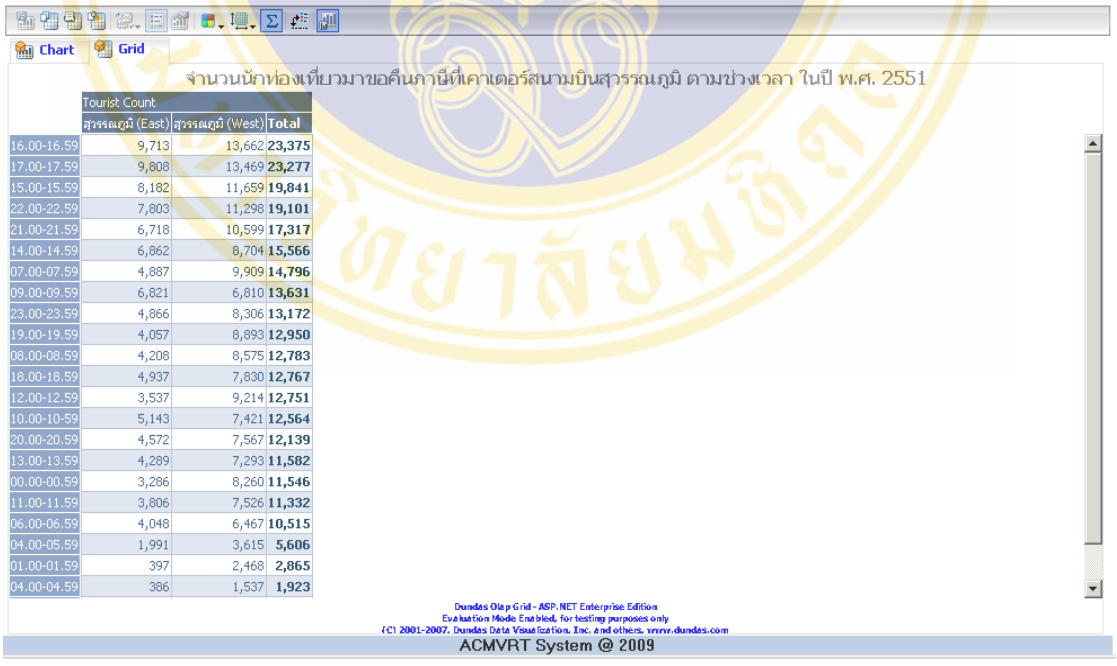


Figure 5.64 Table show number of tourists refund VAT in each period (1 hour) in one day, order from more to less

Example 8: The executive want to know the 1-10 top most sales amount goods in year 2550 and 2551 (or 2007 and 2008). Use this result to arrange the seminar and provided information to entrepreneurs and shops in VRT system and to invite new entrepreneurs to be member of VRT system.

Step:

1) At Selected Cube select "TourisPurchaseGoodsCube" (see Figure 5.65).
2) At Measures and Dimensions select "Goods Dim" the drag Goods Name Report and drop in Categories (see Figure 5.66).
3) At Measures and Dimensions select "Refund VAT Time Dim" then drag Calendar Year Thai and drop in Series: then select year 2550 and 2551 (or 2007 and 2008) (see Figure 5.67).
4) At Measures select Purchase Amount (see Figure 5.68).
5) Click select Filter Top / Button :
   - At Filter Type : select Top and 10.
   - At Using Measure : select Purchase Amount then click OK (see Figure 5.69).
6) The result shows in graph form (see Figure 5.70) and shows in data table (see Figure 5.71).

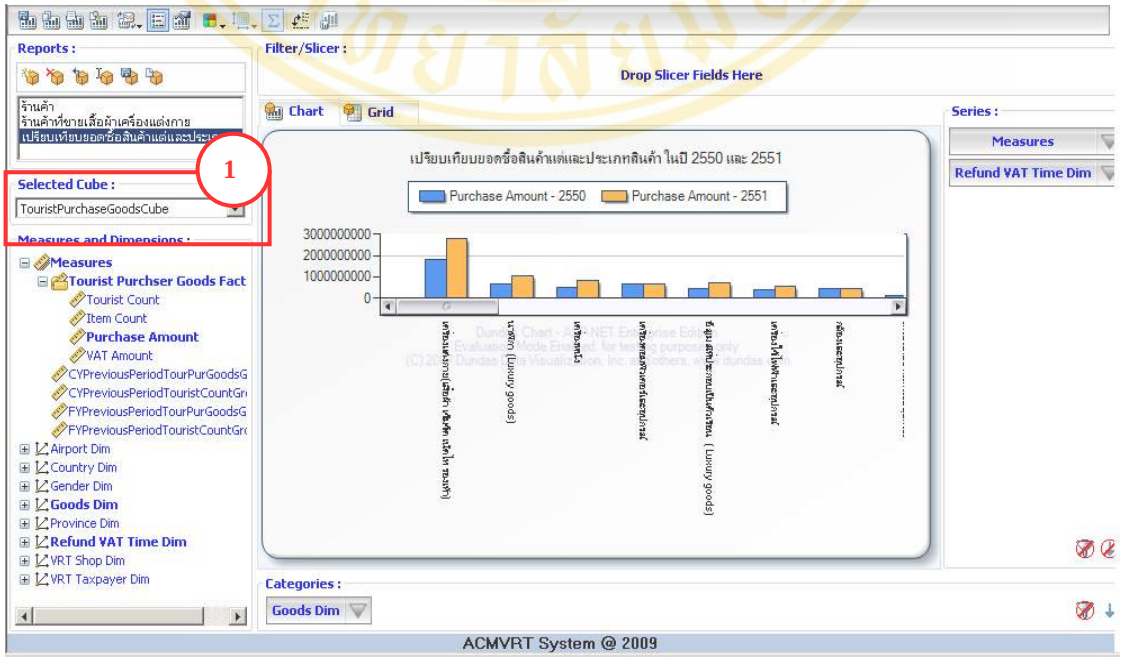


Figure 5.65 Select TourisPurchaseGoodsCube

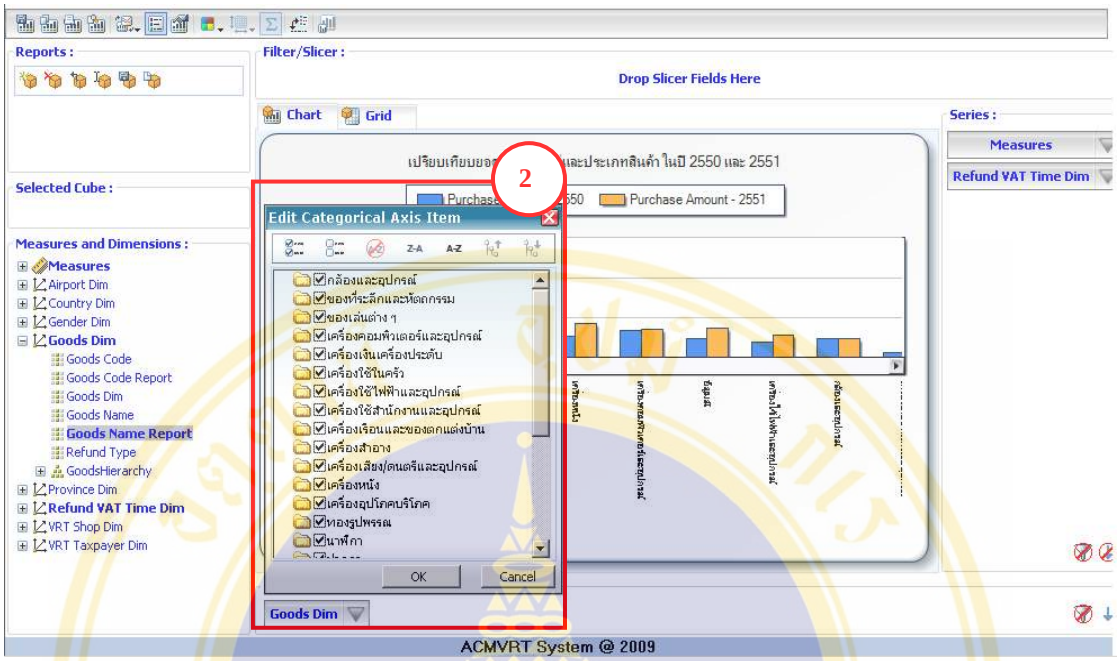


Figure 5.66 Select Goods Dim and Goods Name Report

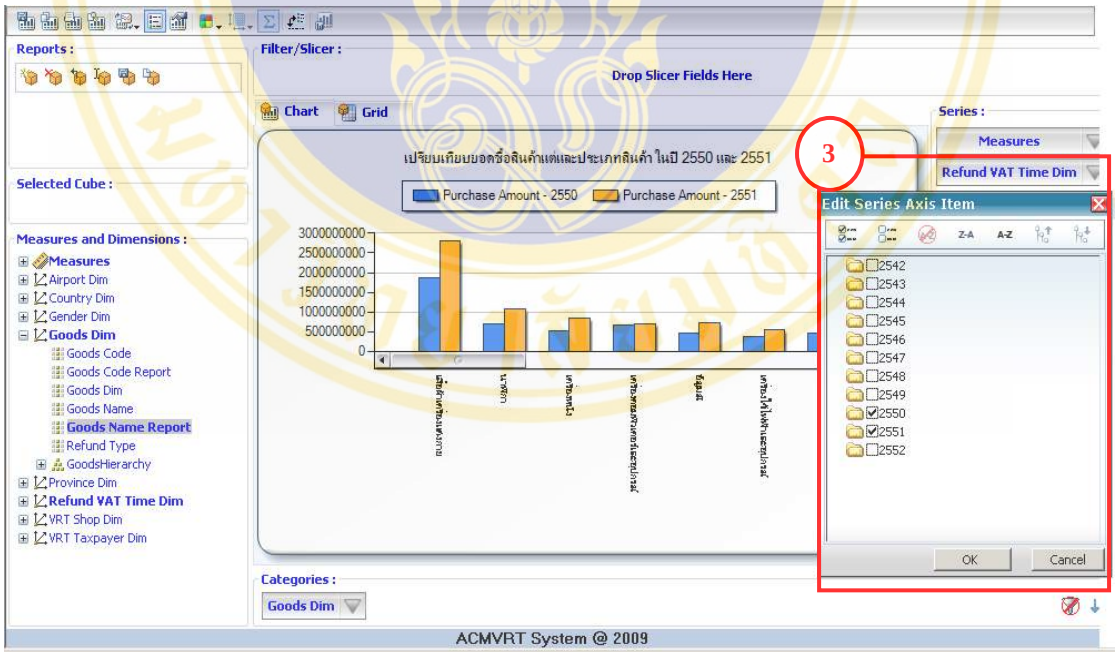


Figure 5.67 Select Refund VAT Time Dim and Calendar Year Thai

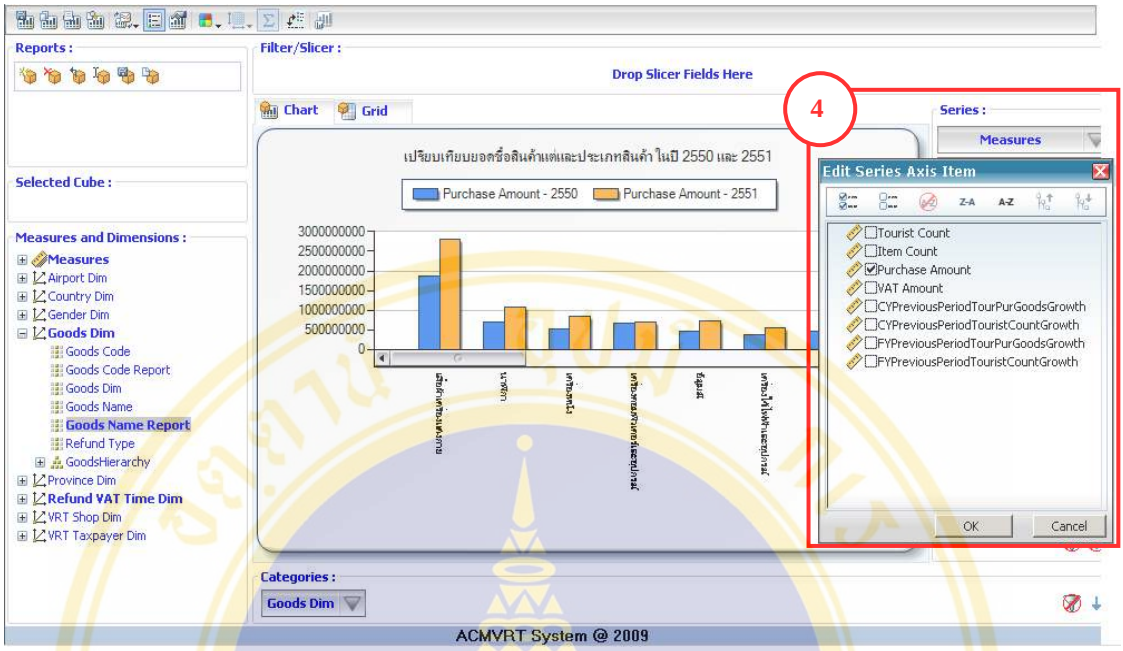


Figure 5.68 Select Measures and Purchase Amount

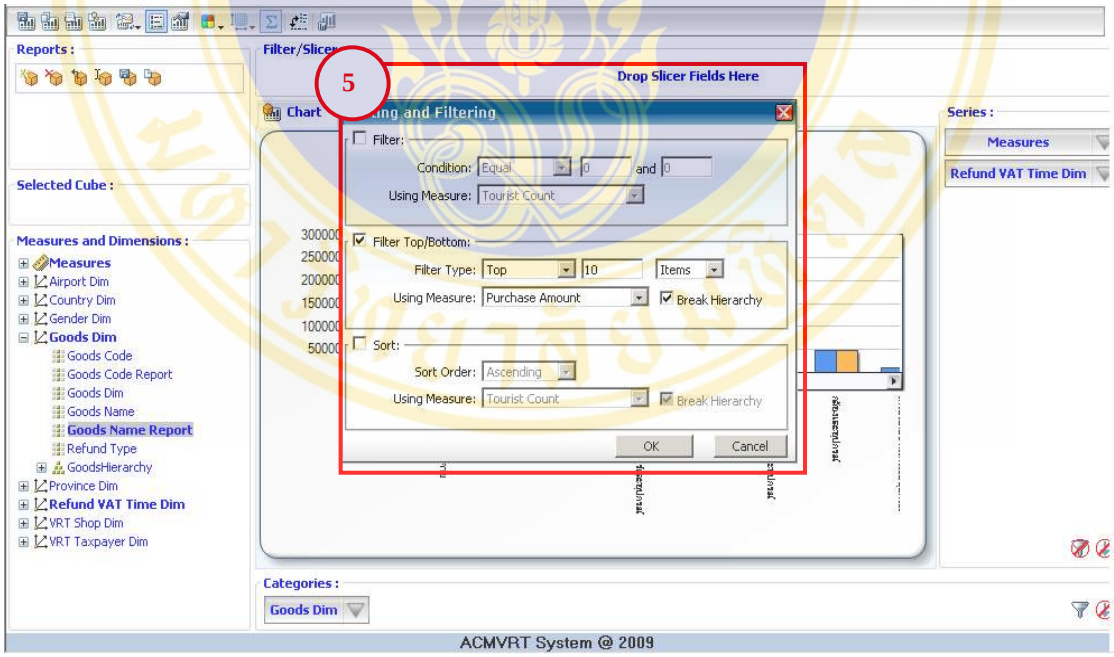


Figure 5.69 Method to sort top 10 popular products

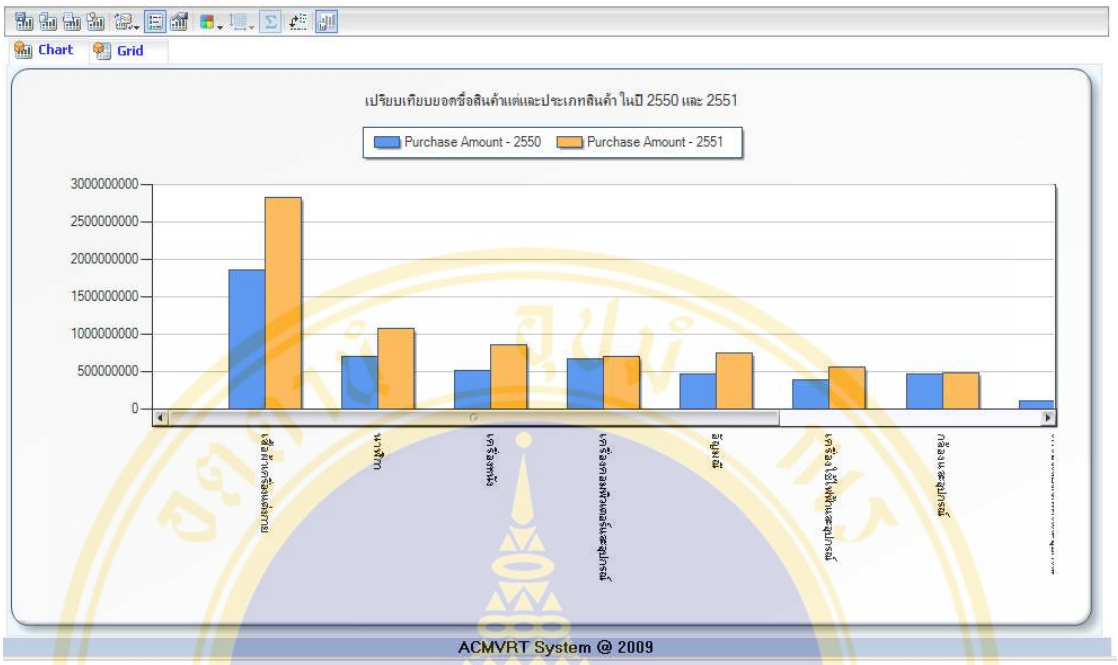
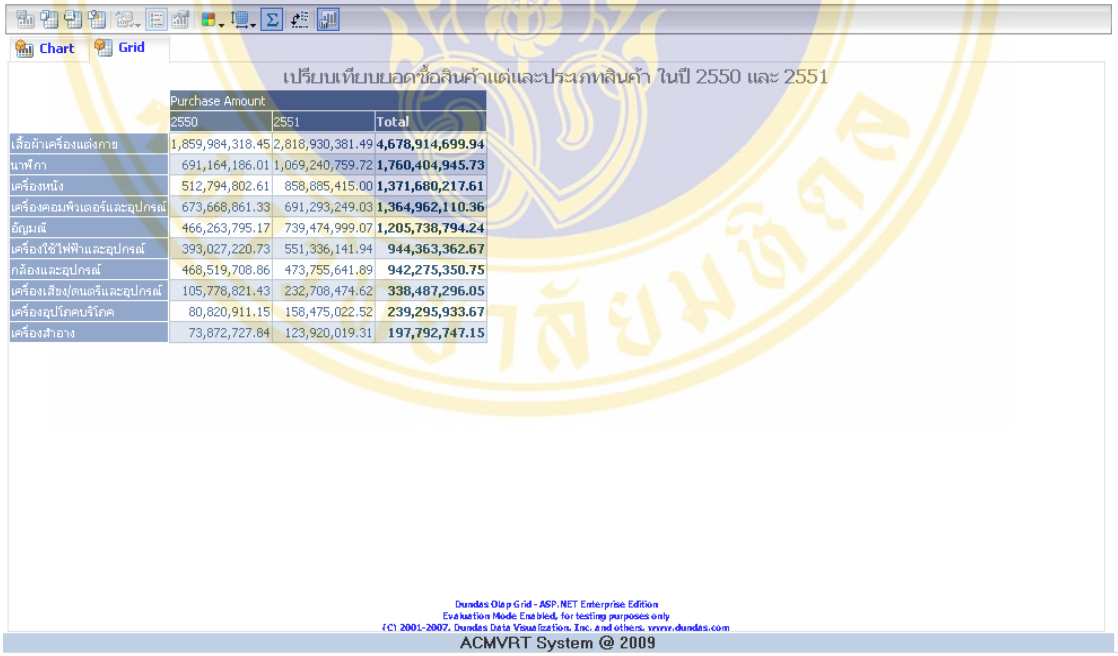


Figure 5.70 Graph of top 10 popular products



	Purchase Amount		
	2550	2551	Total
เสื้อผ้าเครื่องแต่งกาย	1,859,984,318.45	2,818,930,381.49	**4,678,914,699.94**
นาฬิกา	691,164,186.01	1,069,240,759.72	**1,760,404,945.73**
เครื่องหนัง	512,794,802.61	858,885,415.00	**1,371,680,217.61**
เครื่องคอมพิวเตอร์และอุปกรณ์	673,668,861.33	691,293,249.03	**1,364,962,110.36**
อัญมณี	466,263,795.17	739,474,999.07	**1,205,738,794.24**
เครื่องใช้ไฟฟ้าและอุปกรณ์	393,027,220.73	551,336,141.94	**944,363,362.67**
กล้องและอุปกรณ์	468,519,708.86	473,755,641.89	**942,275,350.75**
เครื่องเสียง/ดนตรีและอุปกรณ์	105,778,821.43	232,708,474.62	**338,487,296.05**
เครื่องอุปโภคบริโภค	80,820,911.15	158,475,022.52	**239,295,933.67**
เครื่องสำอาง	73,872,727.84	123,920,019.31	**197,792,747.15**

Figure 5.71 Table of top 10 popular products

Example 9: The executive want to know which shops are popular to tourists for apparel first ten orders in year 2551 (or 2008). And use the result planning public relation to VRT entrepreneurs.

Step:

1) At Selected Cube select "TourisPurchaseGoodsCube" (see Figure 5.72).
2) At Measures and Dimensions select "Refund VAT Time Dim" then drag Calendar Year Thai and drop in Filter/Slicer: Then select year 2551 (or 2008) (see Figure 5.73).
3) At Measures and Dimensions select "Goods Dim" then drag Goods Name and drop in Filter/Slicer: select "เครื่องแต่งกาย" (see Figure 5.74).
4) At Measures and Dimensions select "VRT Shop Dim" then drag Vrt Bran Name and drop in Categories: (see Figure 5.75).
5) At Measures select Purchase Amount (see Figure 5.76).
6) Click select Filter Top / Button : (see Figure 5.77).
   - At Filter Type : Select Top and 10.
   - At Using Measure : select Purchase Amount then click OK .
7) The result shows in graph form (see Figure 5.78) and shows in data table (see Figure 5.79).

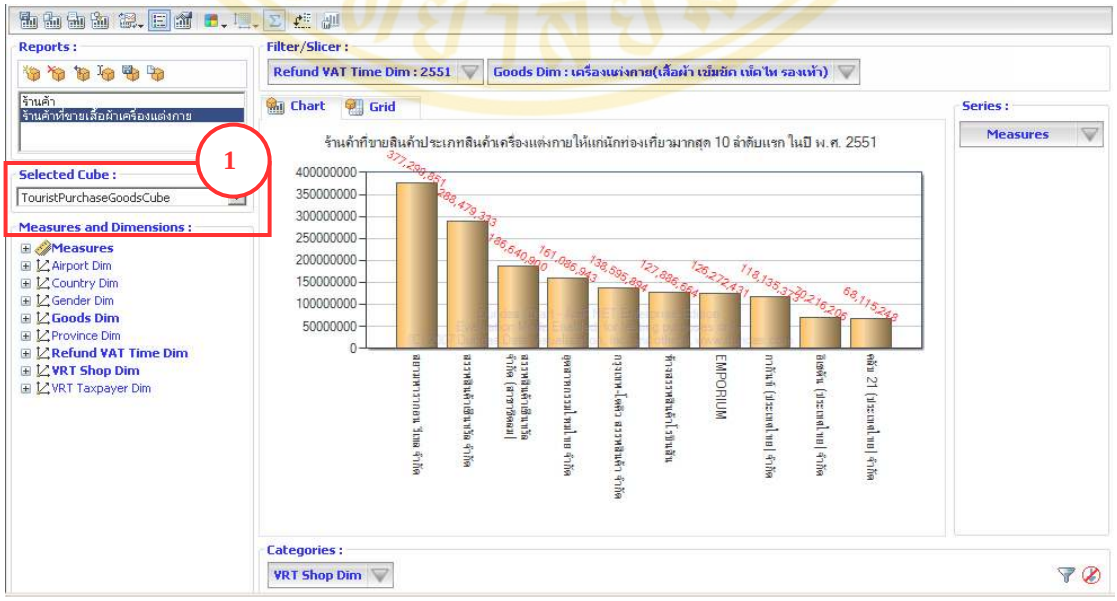


Figure 5.72 Select TourisPurchaseGoodsCube

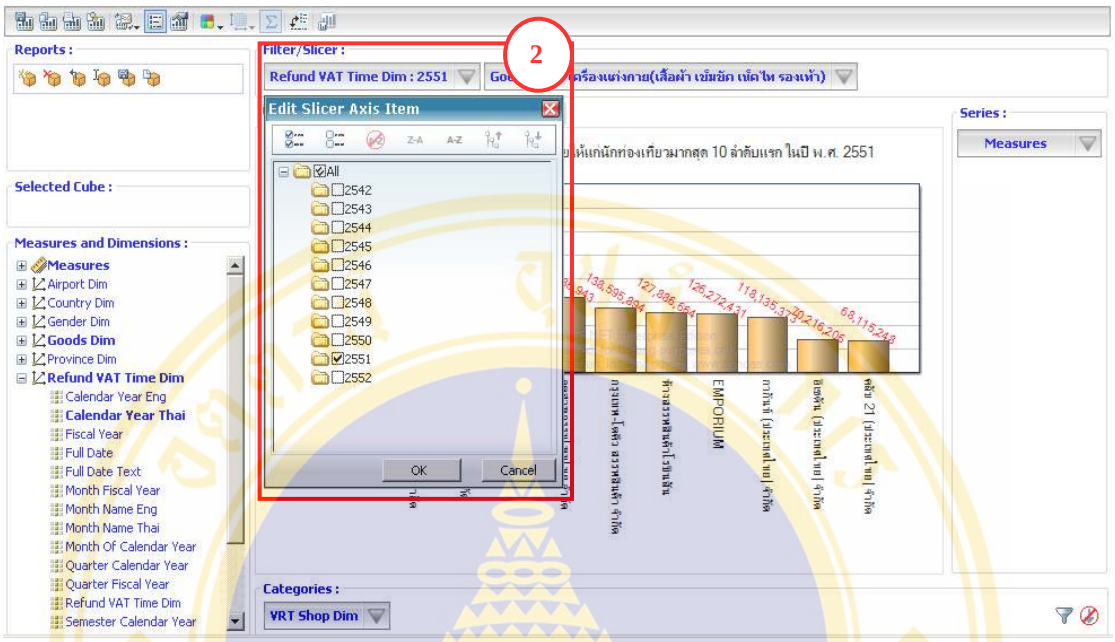


Figure 5.73 Select Refund VAT Time Dim and drag Calendar Year Thai

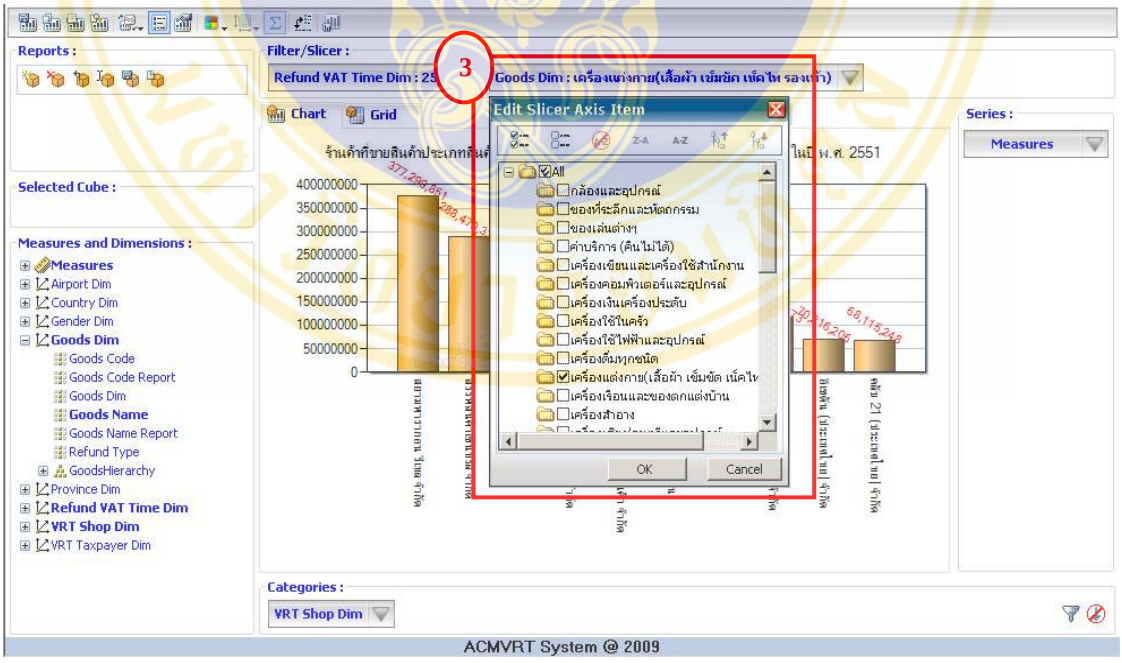


Figure 5.74 Select Goods Dim and drag Goods Name

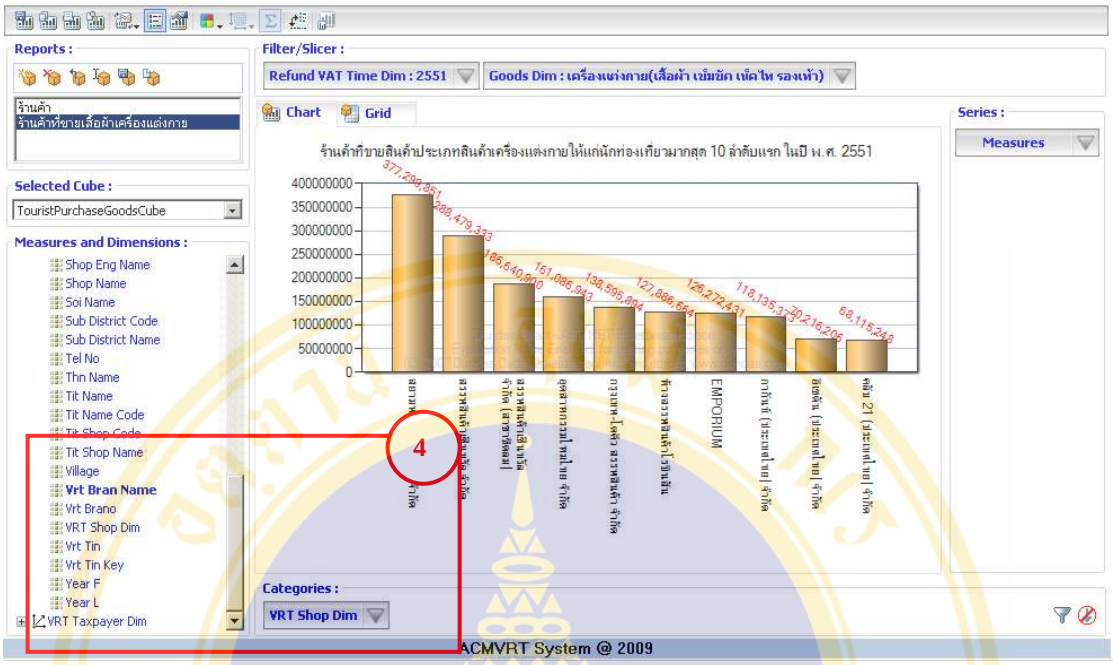


Figure 5.75 VRT Shop Dim and Vrt Bran Name

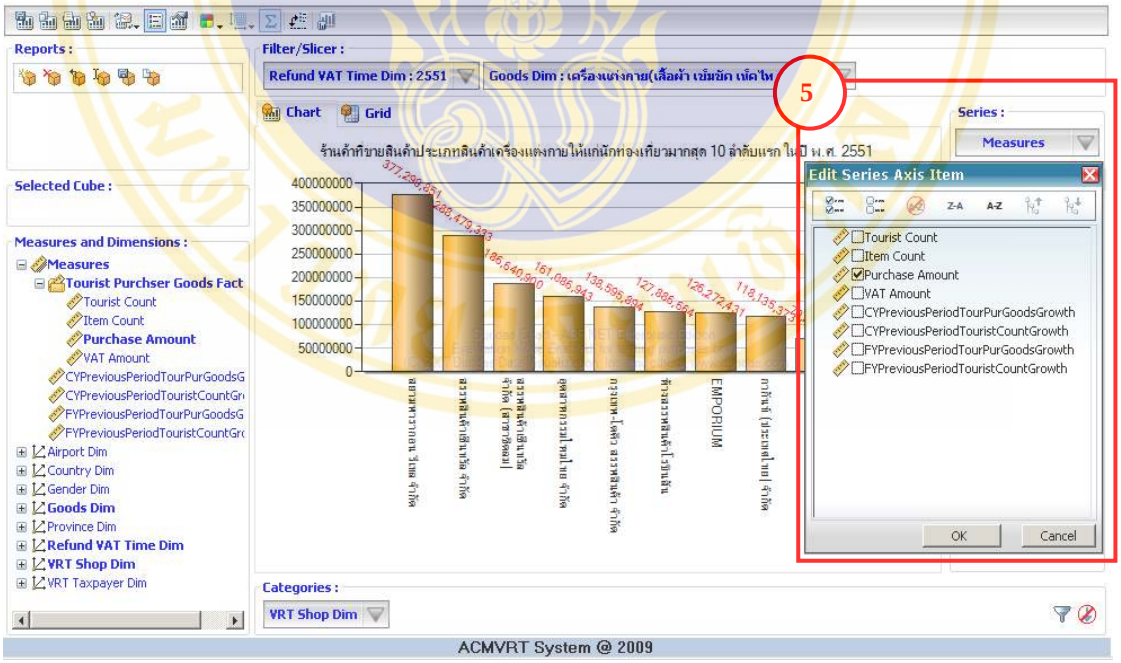


Figure 5.76 At Measures select Purchase Amount

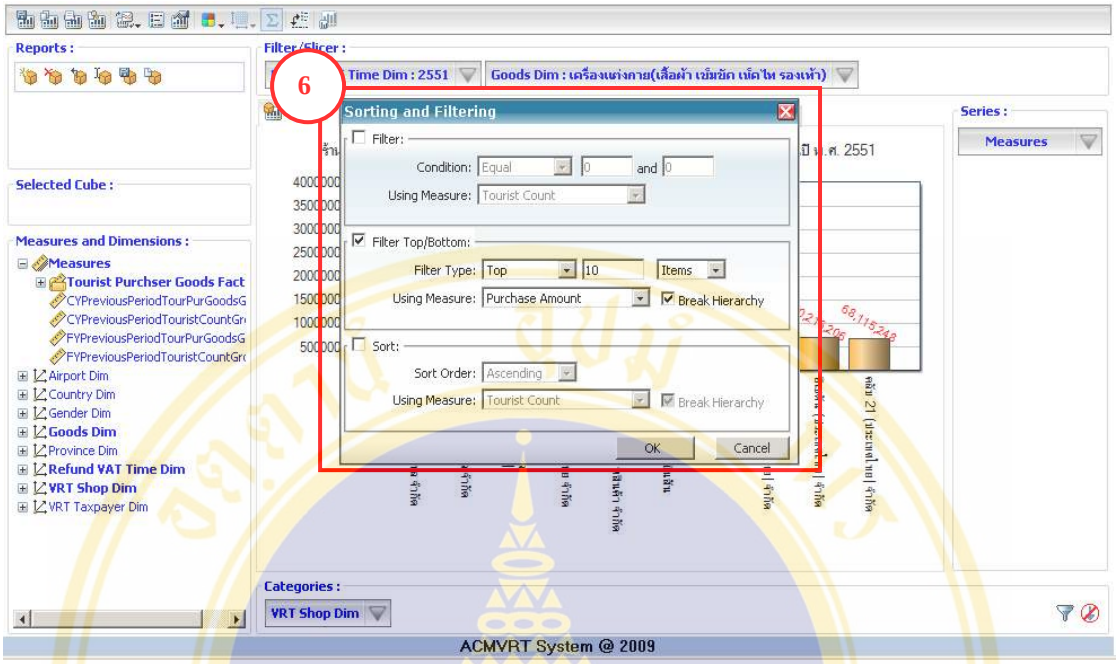


Figure 5.77 Select Purchase Amounts of apparel shops

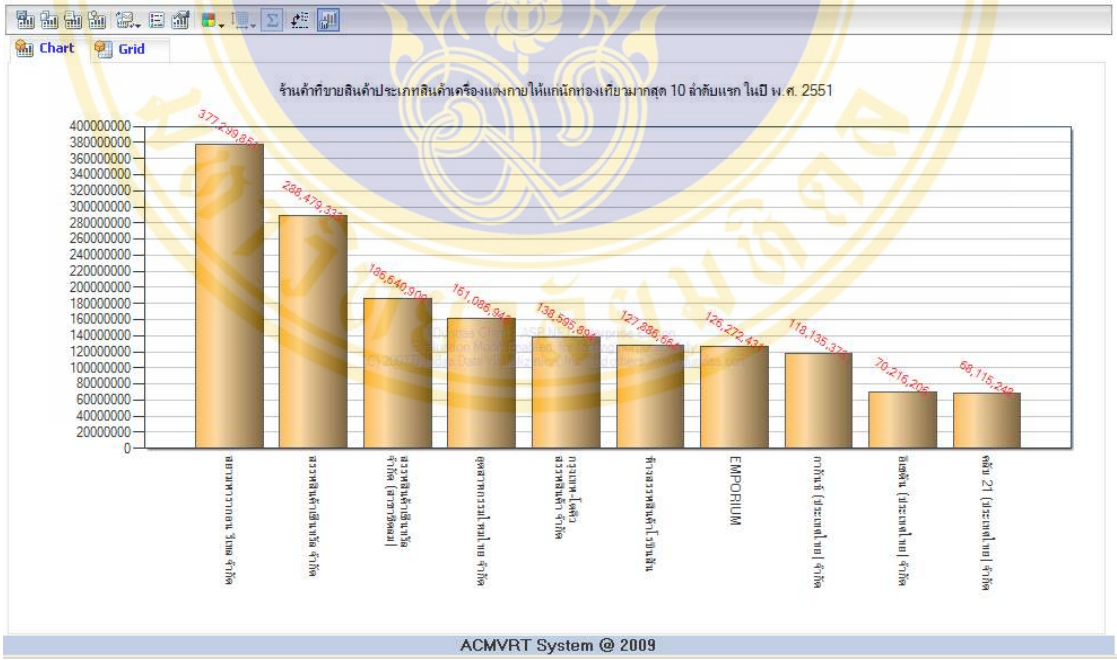


Figure 5.78 Graph of top 10 sales amounts of apparel shops

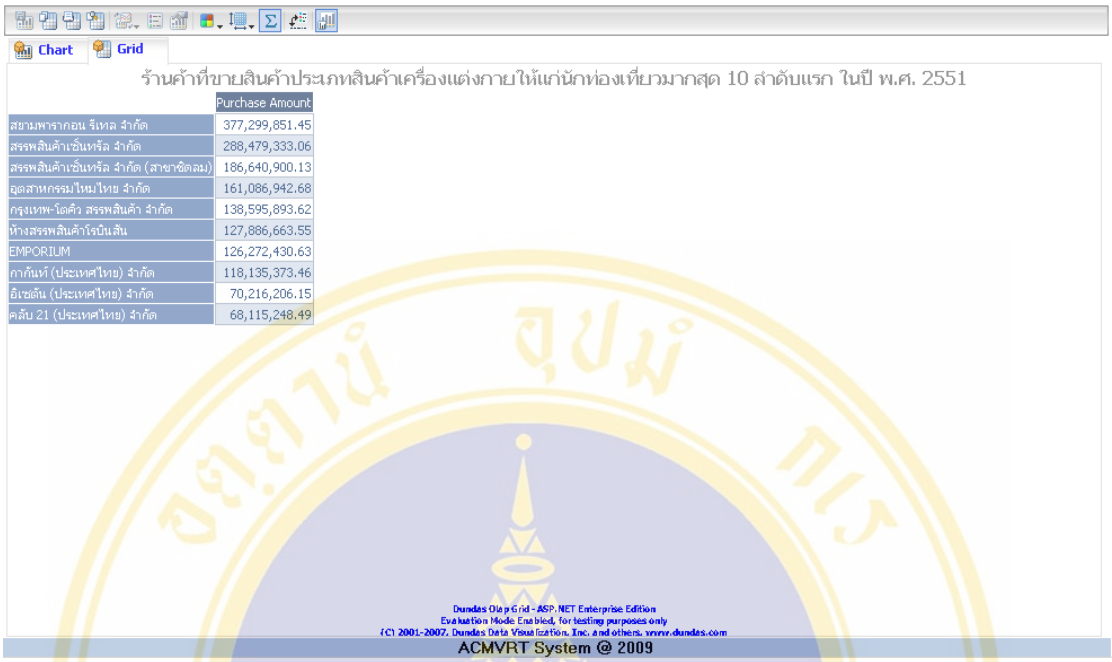


Figure 5.79 Table of top 10 sales amounts of apparel shops

Example 10: The executive want to know the first five types of goods which tourists from China India and Japan like to purchase in year 2551 (or 2008). Use this result to arrange the seminar and provided information to VRT-entrepreneurs and VRT-shops.

Step:

1) At Selected Cube select "TourisPurchaseGoodsCube" (see Figure 5.80).
2) At Measures and Dimensions select "Refund VAT Time Dim" then drag Calendar Year Thai and drop in Filter/Slicer: then select year 2551 (or 2008) (see Figure 5.81).
3) At Measures and Dimensions select "Goods Dim" then drag Goods Name and drop in Categories (see Figure 5.82).
4) At Measures and Dimensions select "Country Dim" the drag Country Name and drop in Series then select India China and Japan (see Figure 5.83).
5) At Measures select Purchase Amount (see Figure 5.84).
6) Click select Filter Top / Button (see Figure 5.85).
   - At Filter Type : select Top and 5.

- At Using Measure : select Purchase Amount then click OK .

7) The result shows in graph form (see Figure 5.86) and shows in data table (see Figure 5.87).

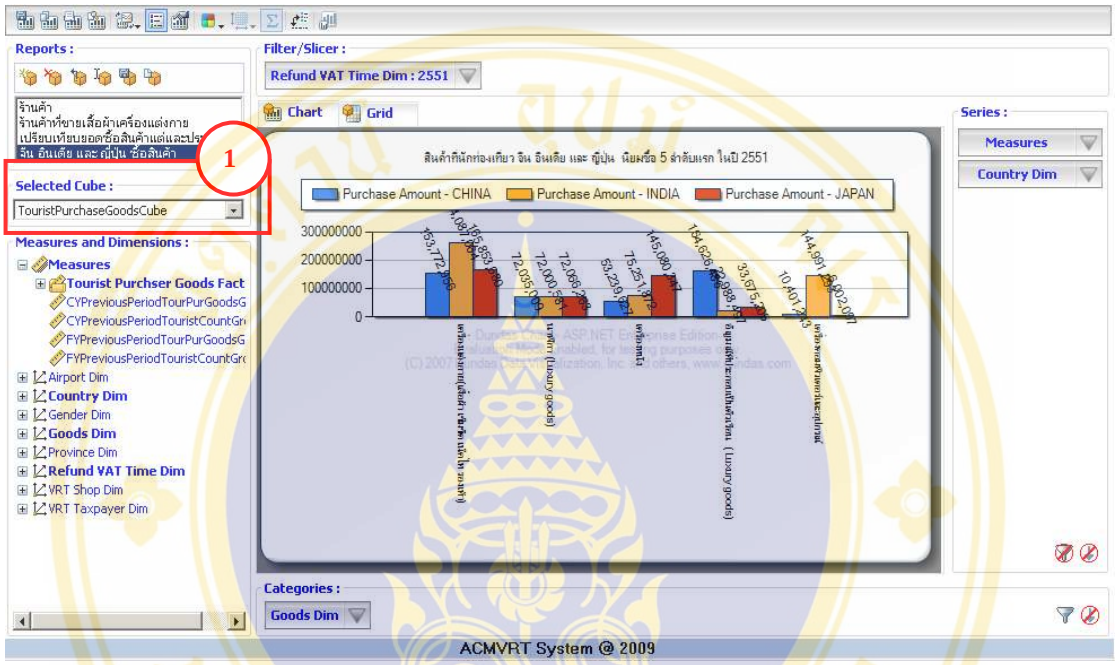


Figure 5.80 Select TourisPurchaseGoodsCube

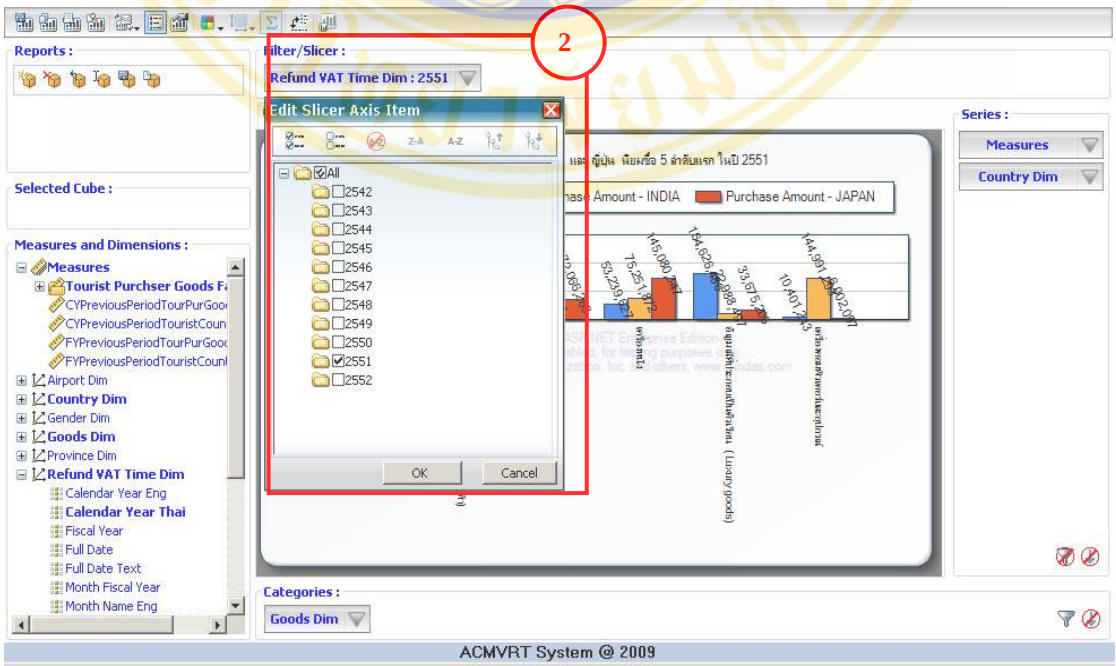


Figure 5.81 Select Refund VAT Time Dim and Calendar Year Thai

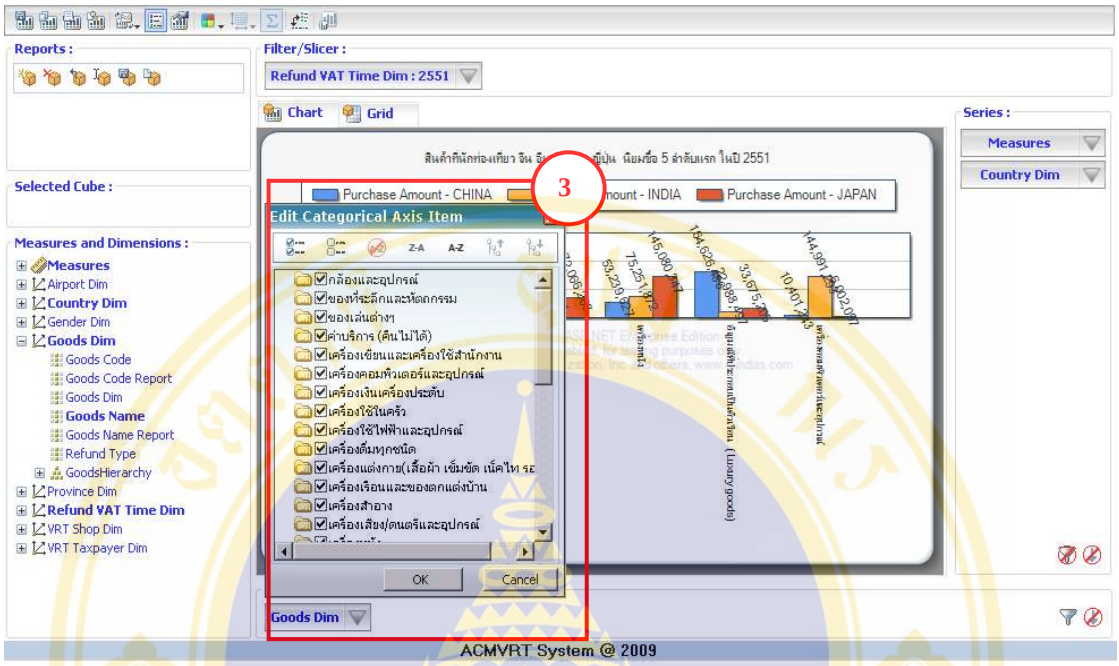


Figure 5.82 Select Goods Dim and Goods Name

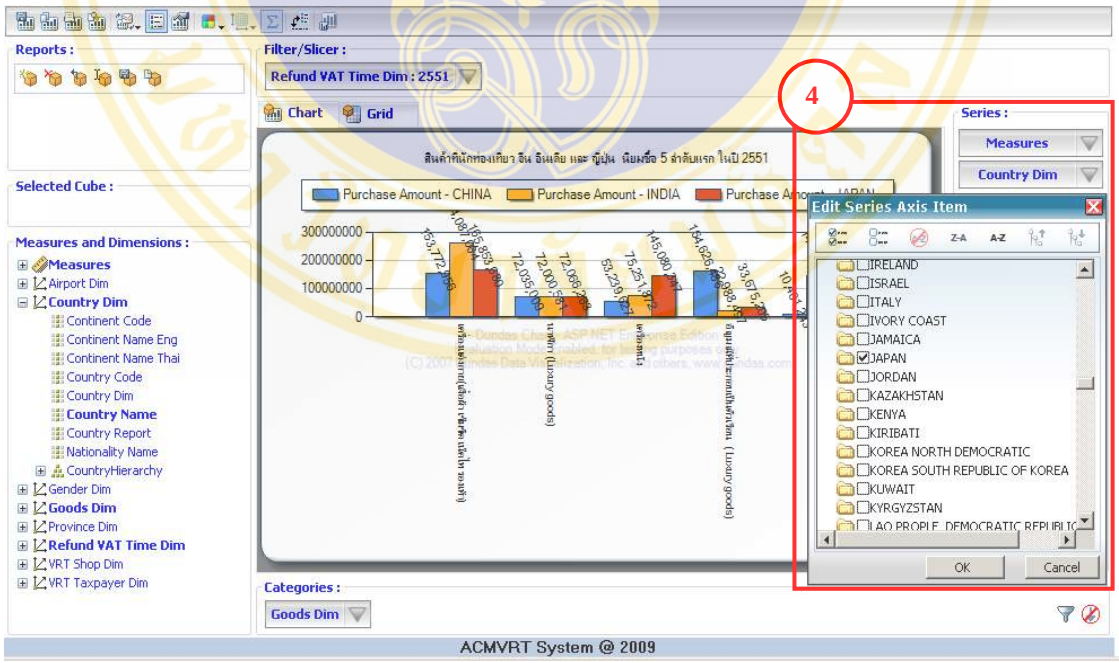


Figure 5.83 Select Country Dim and Country Name

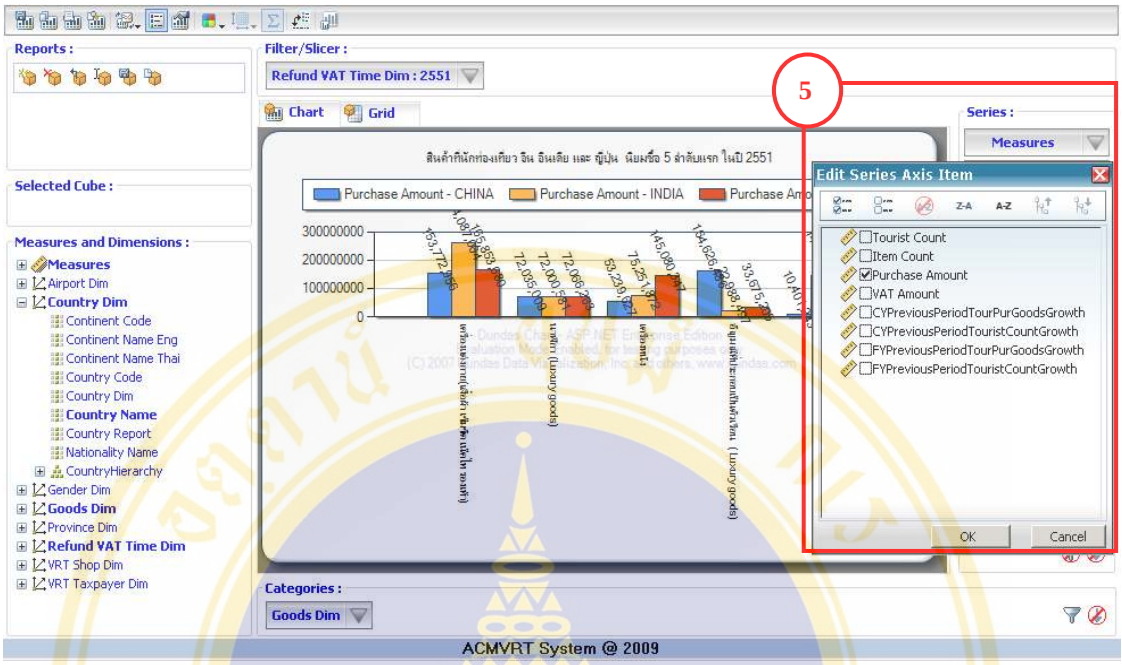


Figure 5.84 Select Measures and Purchase Amount

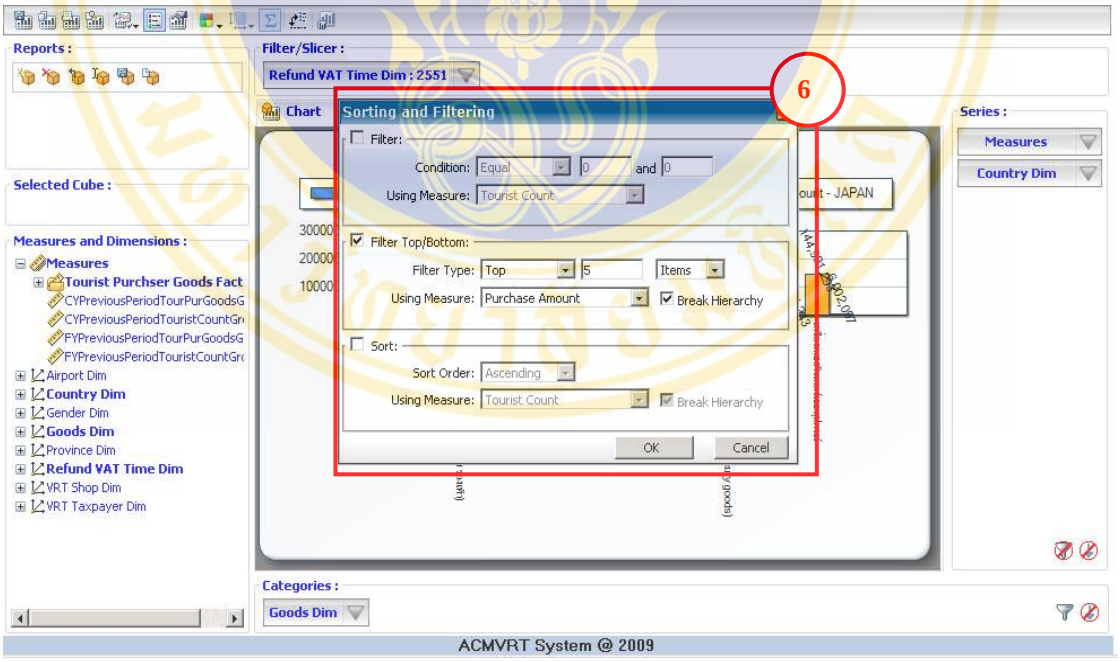


Figure 5.85 Filter top 5 highest sales amount products divided by country

Figure 5.86 Graph of popular products of tourists from each country

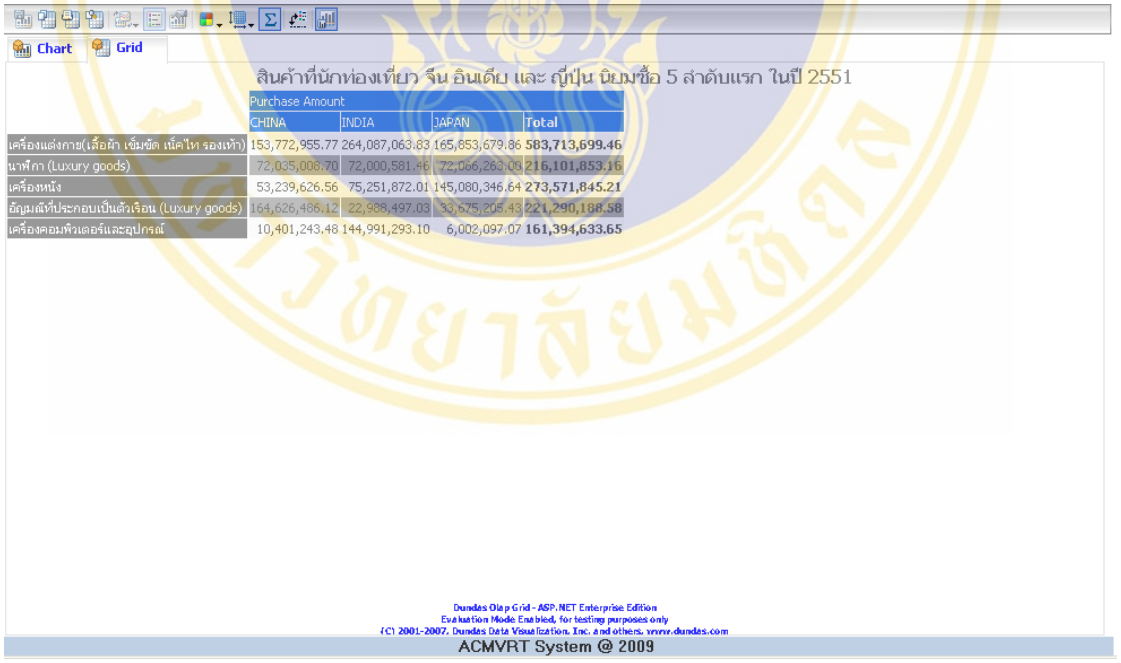


Figure 5.87 Table of popular products of tourists from each country

Example 11: The executive want to know first ten shops where Indian tourists purchase most in year 2550 and 2551 (or 2007 and 2008), because from statistic Indian tourists are maximum number of tourists who claim for VAT refund. User can use information planning to suggestion entrepreneurs and continue audit on business.

Step:

1) At Selected Cube select "TouristPurchaseGoodsCube" (see Figure 5.88).
2) At Measures and Dimensions select "Country Dim" then drag Country Name and drop in Filter/Slicer : then select India (see Figure 5.89).
3) At Measures and Dimensions select "VRT Shop Dim" then drag Shop Name and drop in Categories (see Figure 5.90).
4) At Measures and Dimensions select "Refund VAT Time Dim" then drag Calendar Year Thai and drop Series: then select 2550 and 2551 (or 2007 and 2008) (see Figure 5.91).
5) At Measures select Purchase Amount (see Figure 5.92).
6) Click select Filter Top / Button (see Figure 5.93).
   - At Filter Type : select Top and 10 .
   - At Using Measure : select Purchase Amount then click OK .
7) The result shows in graph form (see Figure 5.94) and shows in data table (see Figure 5.95).

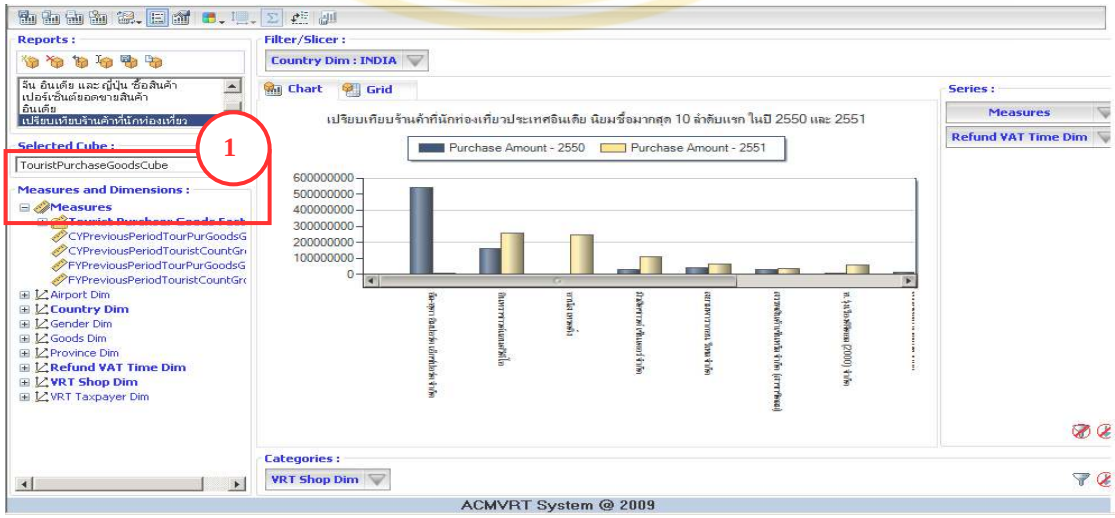


Figure 5.88 Select TouristPurchaseGoodsCube

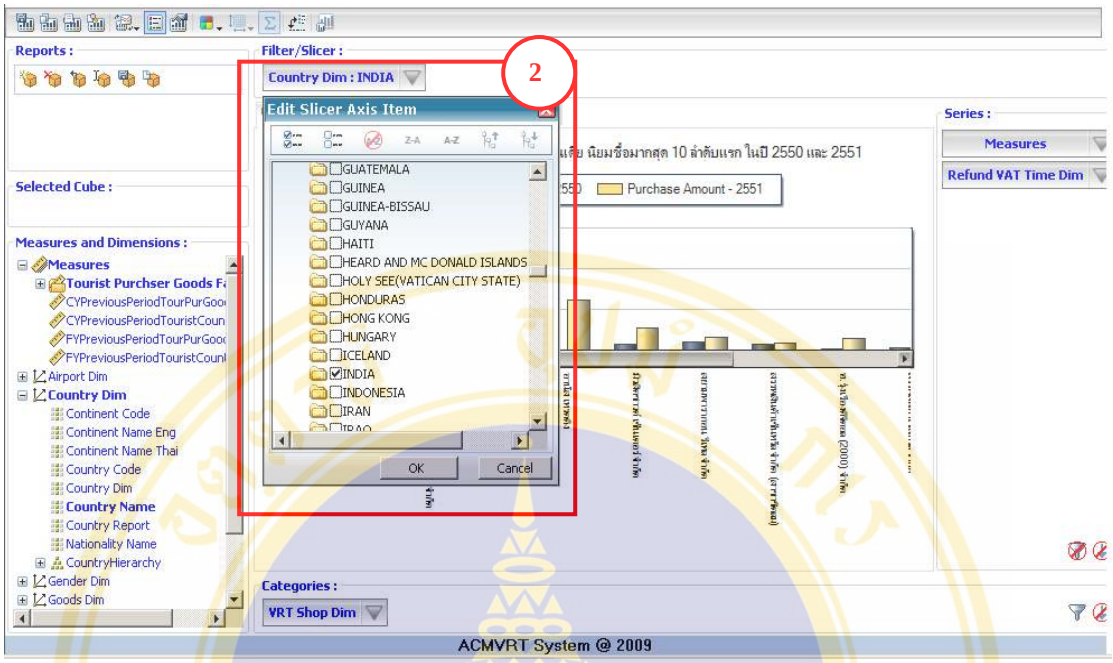


Figure 5.89 Select Country Dim and Country Name

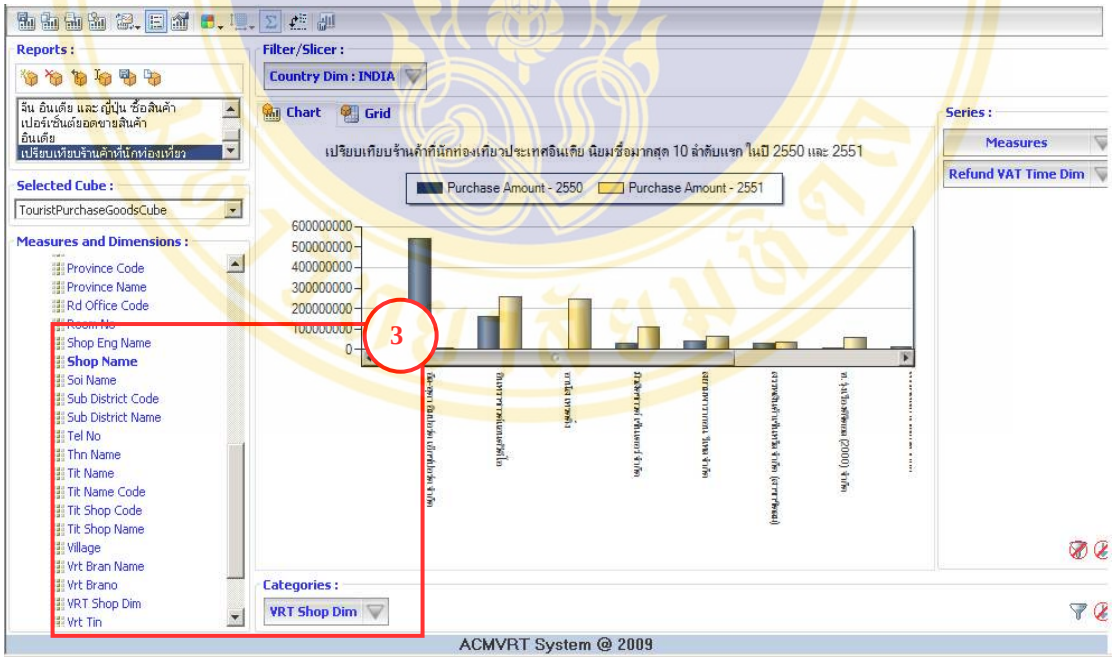


Figure 5.90 VRT Shop Dim and Shop Name

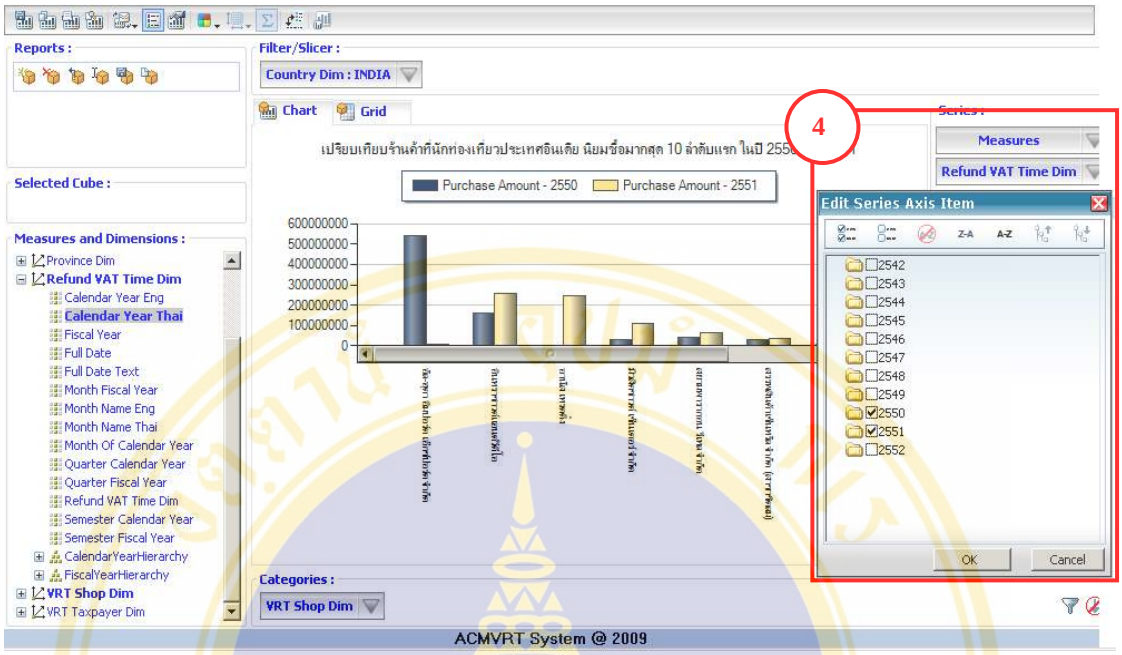


Figure 5.91 Select Refund VAT Time Dim and Calendar Year Thai

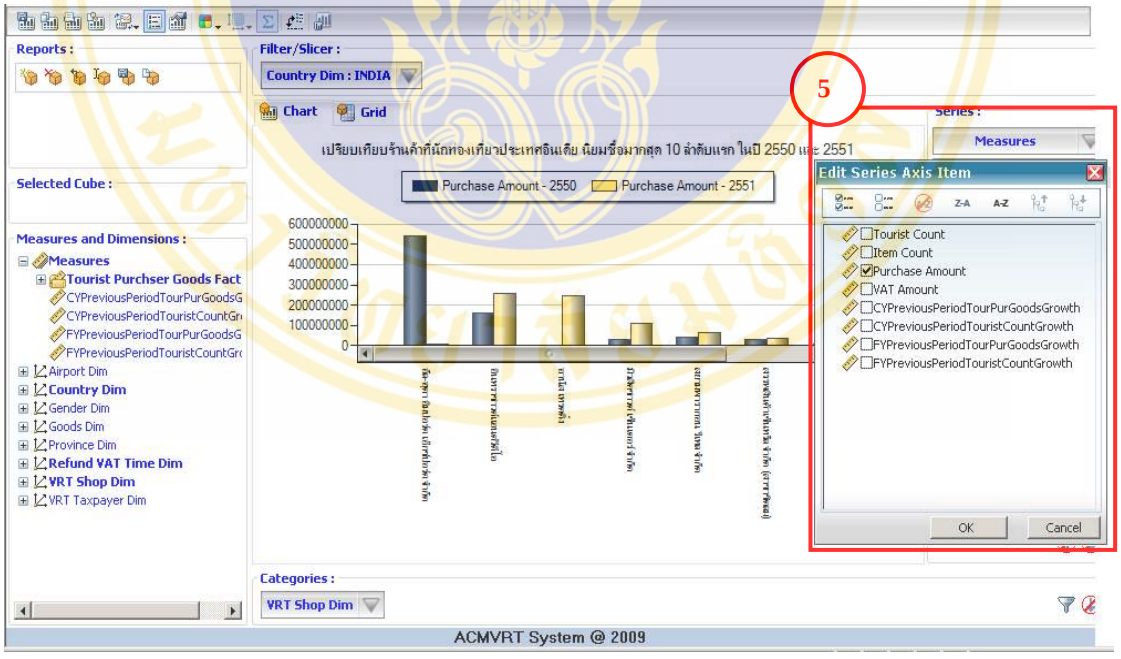


Figure 5.92 Select Measures and Purchase Amount

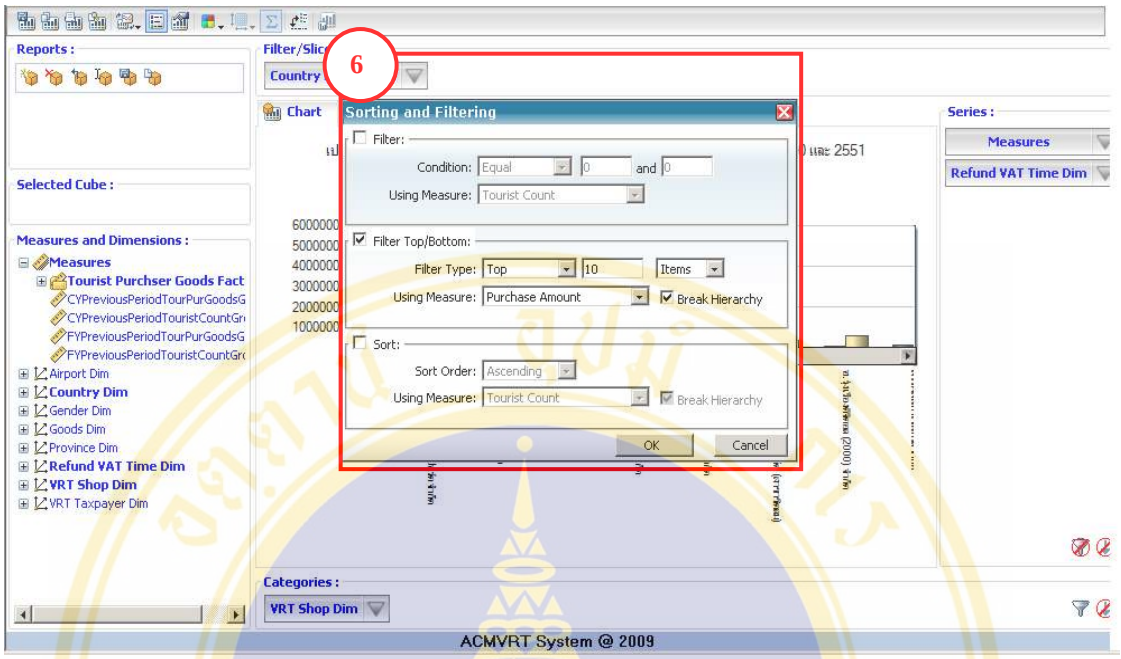


Figure 5.93 Filter top 10 VRT shops' sales amount of Indian tourists

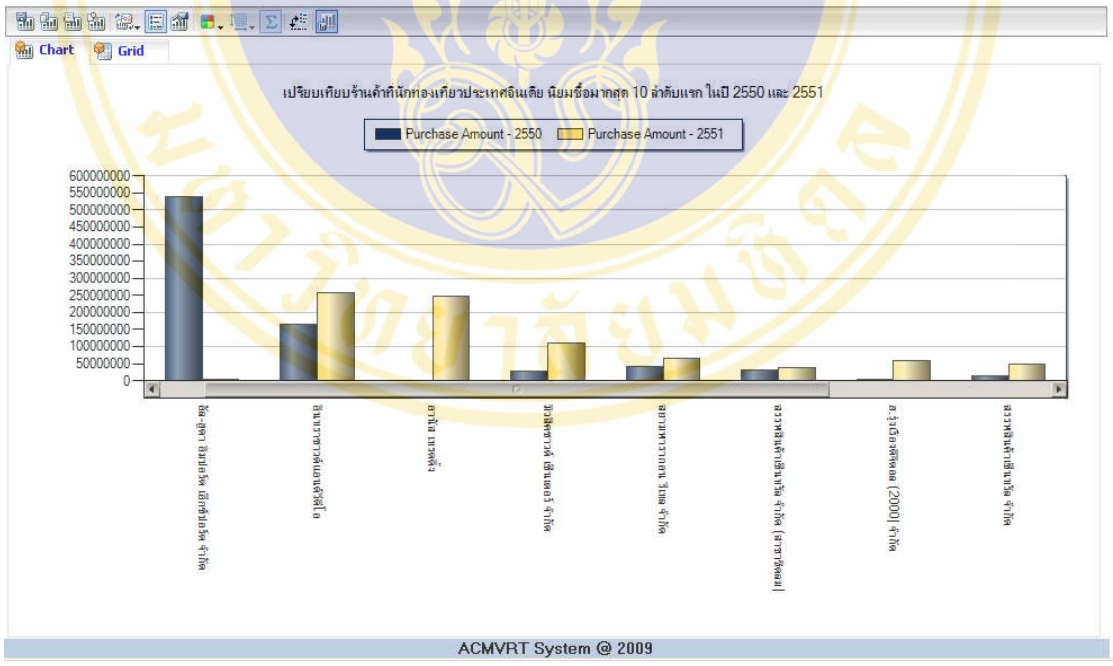


Figure 5.94 Graph of top 10 popular shops for Indian tourists

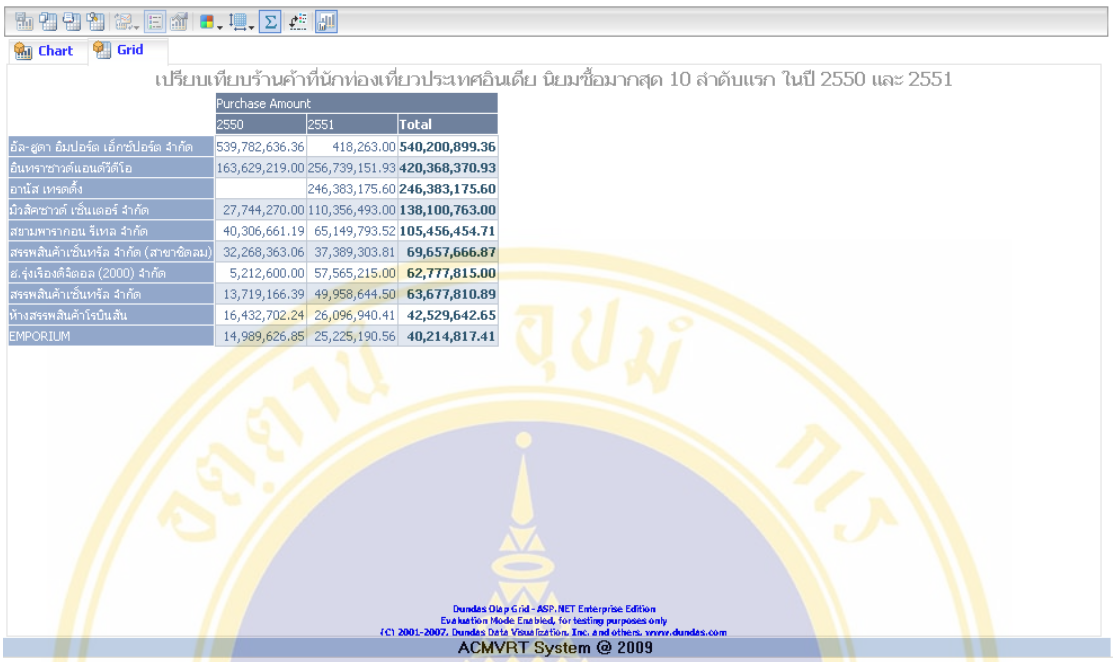


Chart Grid

เปรียบเทียบร้านค้าที่นักท่องเที่ยวประเทศอินเดีย นิยมซื้อมากสุด 10 ลำดับแรก ในปี 2550 และ 2551

	Purchase Amount		
	2550	2551	Total
อัล-ฮูดา อิมปอร์ต เอ็กซ์ปอร์ต จำกัด	539,782,636.36	418,263.00	**540,200,899.36**
อินทราชาวด์แอนด์วีดีโอ	163,629,219.00	256,739,151.93	**420,368,370.93**
อานัส เทรดดิ้ง		246,383,175.60	**246,383,175.60**
นิวสิคชาวด์ เซ็นเตอร์ จำกัด	27,744,270.00	110,356,493.00	**138,100,763.00**
สยามพารากอน รีเทล จำกัด	40,306,661.19	65,149,793.52	**105,456,454.71**
สรรพสินค้าเซ็นทรัล จำกัด (สาขาชิดลม)	32,268,363.06	37,389,303.81	**69,657,666.87**
อ.รุ่งเรืองดิจิตอล (2000) จำกัด	5,212,600.00	57,565,215.00	**62,777,815.00**
สรรพสินค้าเซ็นทรัล จำกัด	13,719,166.39	49,958,644.50	**63,677,810.89**
ห้างสรรพสินค้าโรบินสัน	16,432,702.24	26,096,940.41	**42,529,642.65**
EMPORIUM	14,989,626.85	25,225,190.56	**40,214,817.41**

Dundas Olap Grid - ASP.NET Enterprise Edition
Evaluation Mode Enabled, for testing purposes only
(C) 2001-2007, Dundas Data Visualization, Inc. and others. www.dundas.com
ACMVRT System @ 2009

Figure 5.95 Table of top 10 popular shops for Indian tourists

Example 12: The executive want to know first 10 shops which have the most sales amount in year 2551 (or 2008). Use this report to audit regulatory taxation of VAT entrepreneurs.

Step:

1) At Selected Cube select “TourisPurchaseGoodsCube” (see Figure 5.96).
2) At Measures and Dimensions select “Refund VAT Time Dim” then drag. Calendar Year Thai and droop Filter/Slicer: then select year 2551 (or 2008) (see Figure 5.97).
3) At Measures and Dimensions select “VRT Shop Dim” then drag Vrt Bran Name and drop in Categories (see Figure 5.98).
4) At Measures select Purchase Amount (see Figure 5.99).
5) Click then select Filter Top / Button (see Figure 5.100).
   - At Filter Type : select Top and 10.
   - At Using Measure : select Purchase Amount then click OK .
6) The result shows in graph form (see Figure 5.101) and shows in data table (see Figure 5.102).

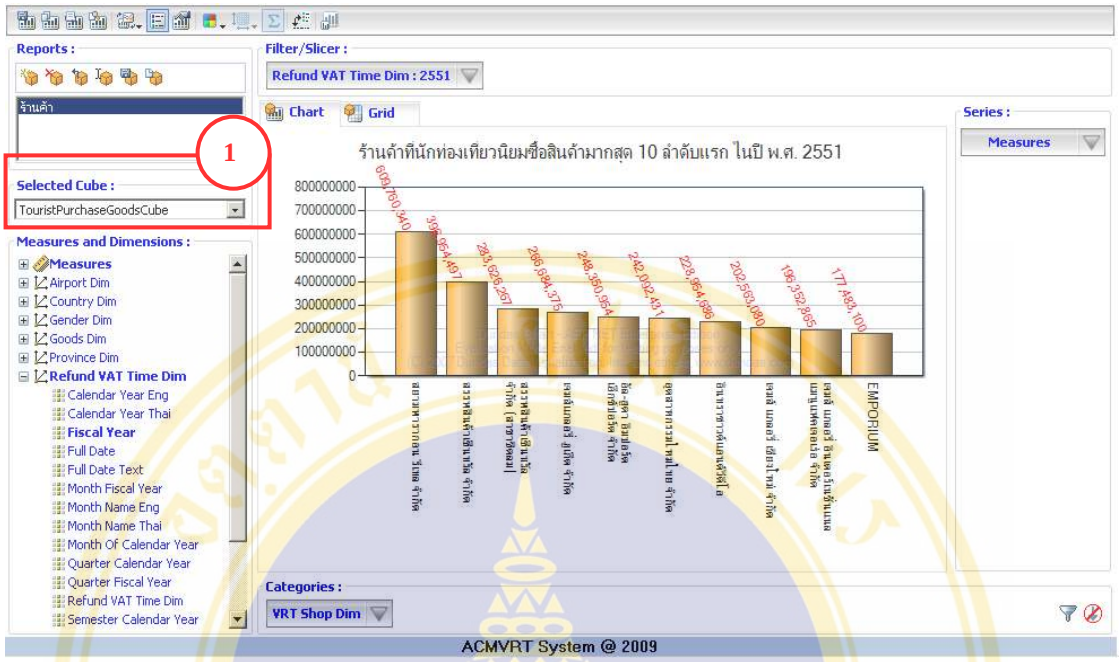


Figure 5.96 Select TourisPurchaseGoodsCube

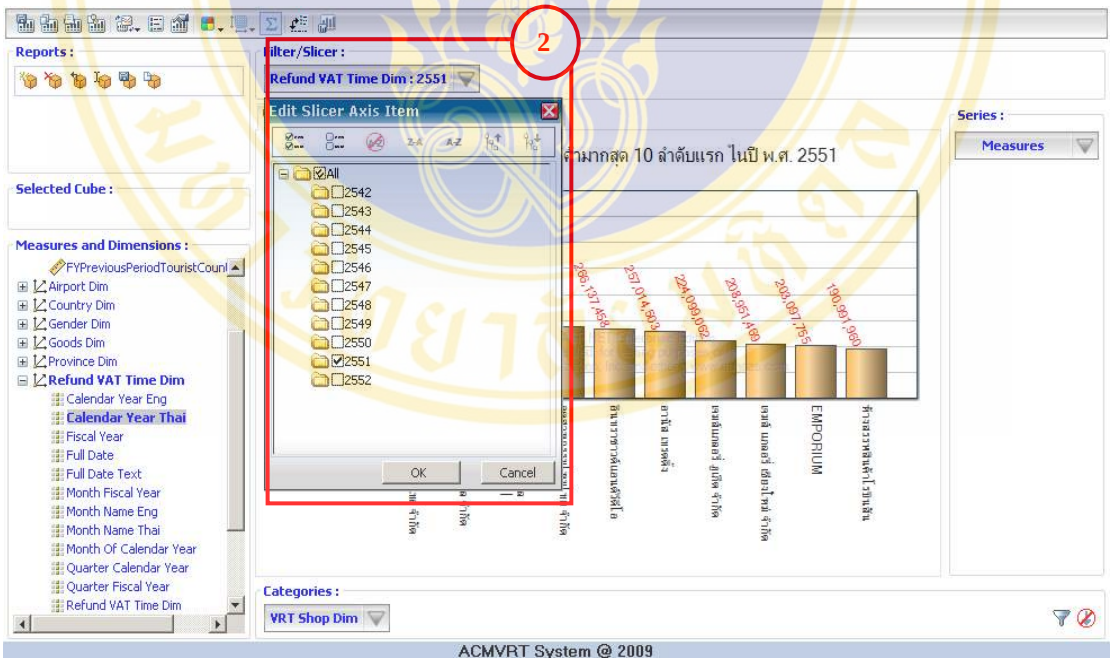


Figure 5.97 Select Refund VAT Time Dim and Calendar Year Thai

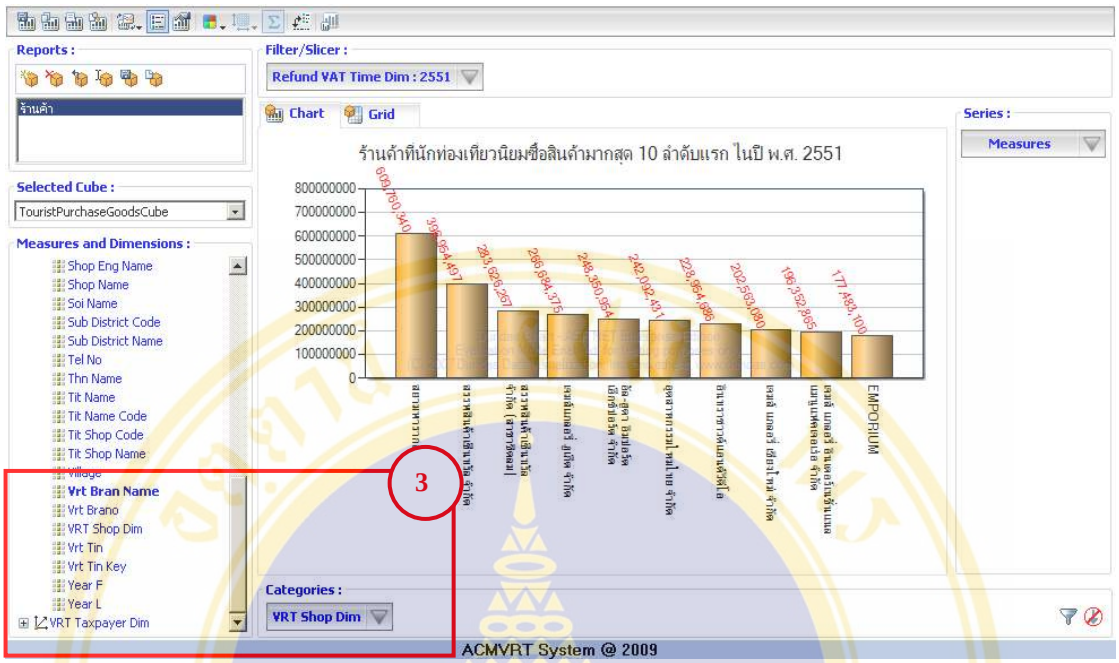


Figure 5.98 Select VRT Shop Dim and Vrt Bran Name

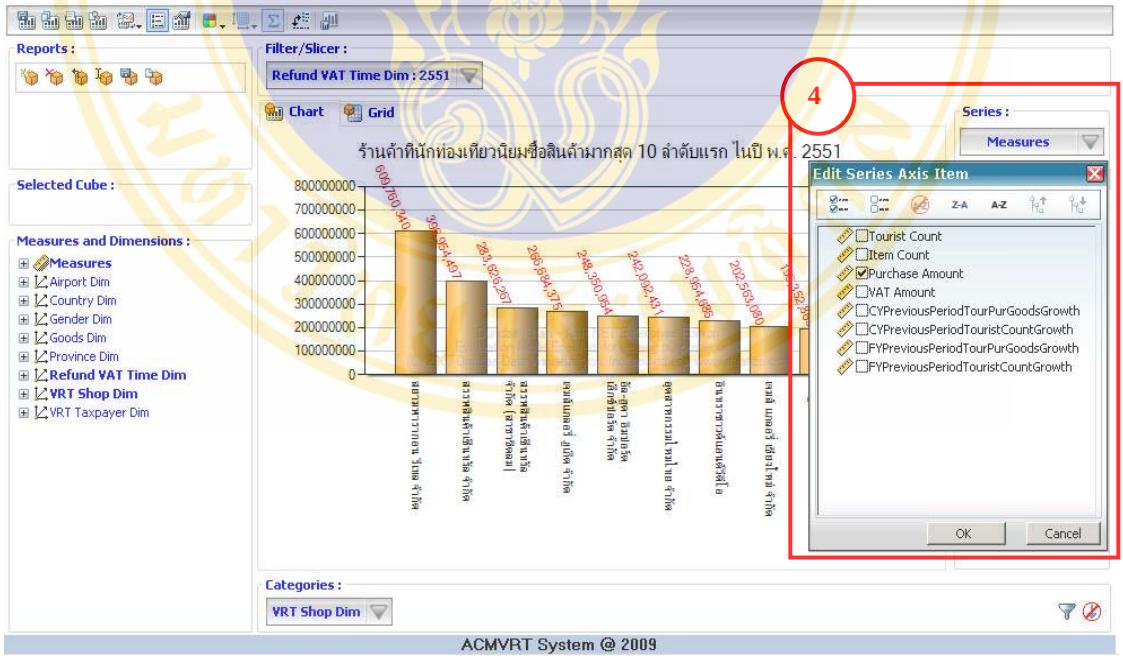


Figure 5.99 Select Measures and Purchase Amount

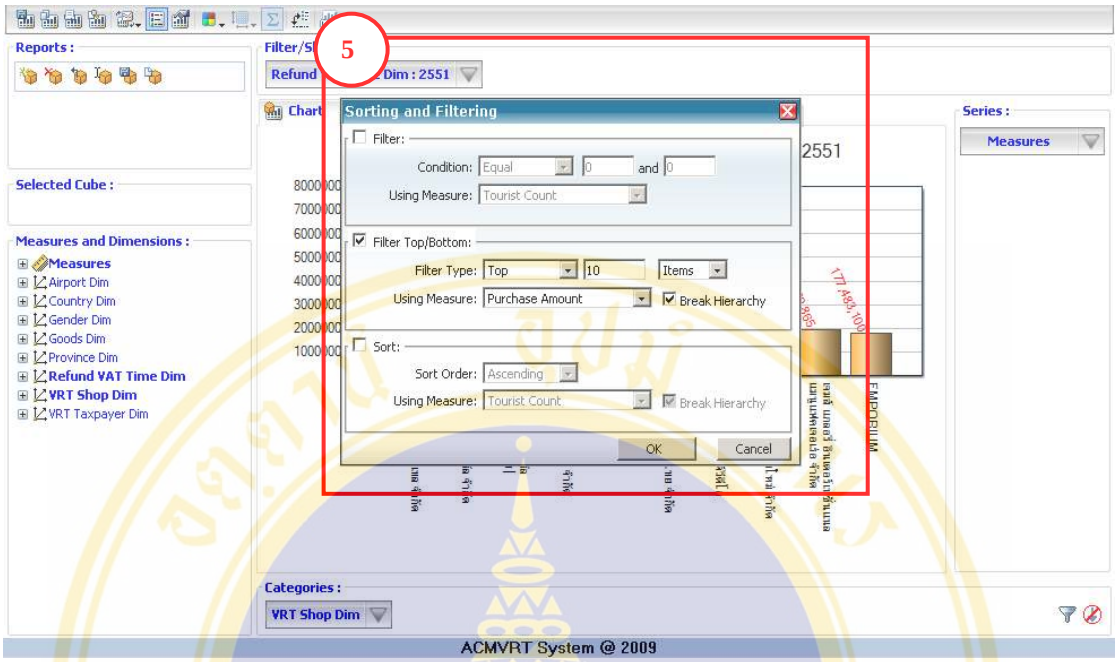


Figure 5.100 Filter top 10 popular shops for tourists

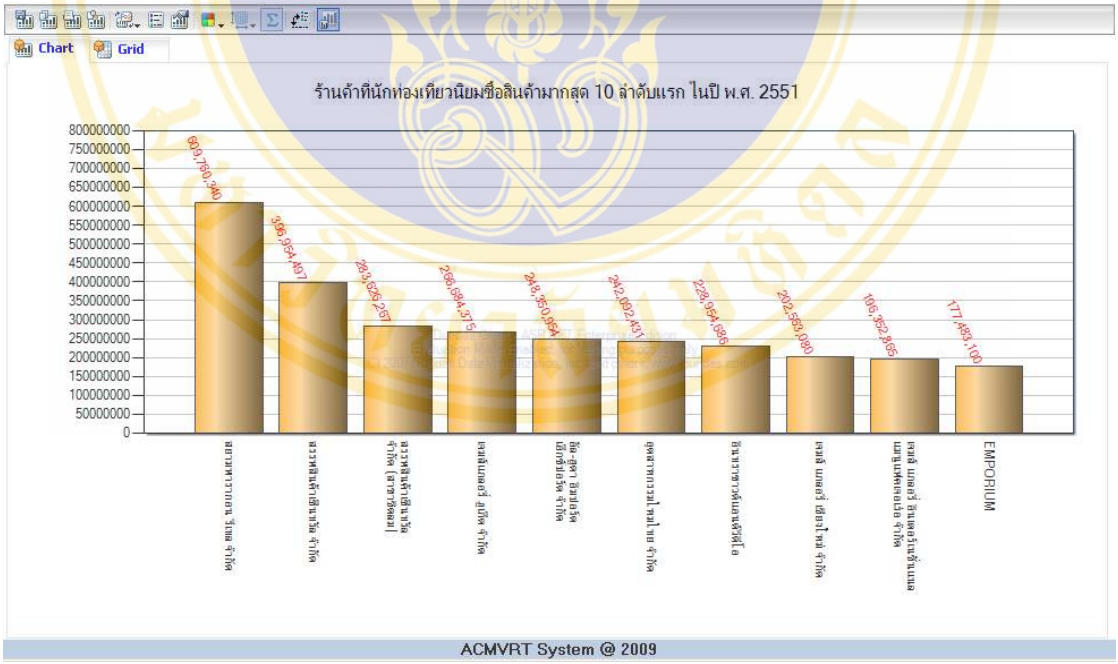


Figure 5.101 Graph of top 10 highest tourists' sales amount shops

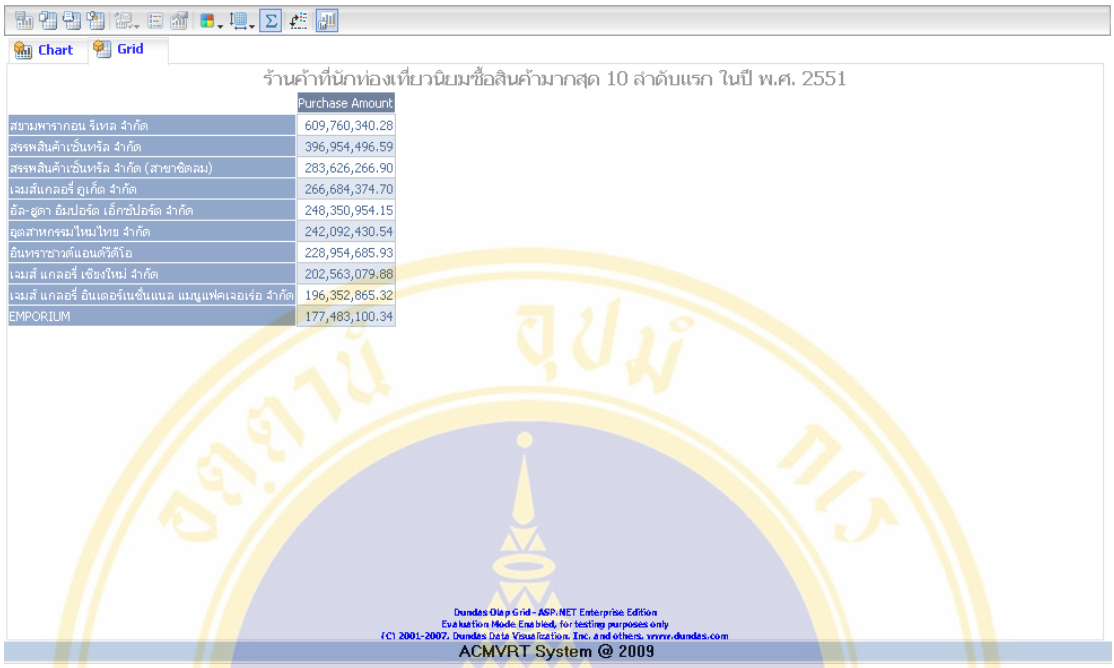


Figure 5.102 Table of top 10 highest tourists' sales amount shops

Example 13: The executive want to know top 10 of entrepreneurs who have percentage of sales amount to tourists compared with total sales amount which filed tax in year 2551 (or 2008).. To bring information to verify regulatory taxation of entrepreneurs VAT.

Step:

1) At Selected Cube select "EntrepreneursPaidTaxCube" (see Figure 5.103).
2) At Measures and Dimensions select "Refund VAT Time Dim" then drag Calendar Year Thai and drop in Filter / Slicer : then select year 2551 (or 2008) (see Figure 5.104).
3) At Measures and Dimensions select "VRT Shop Dim" then drag Shop Name and drop in Categories: (see Figure 5.105).
4) At Measures select SaleTourisPercentageOfSaleAmount (see Figure 5.106).
5) Click select Filter Top / Button (see Figure 5.107).
   - At Filter Type : select Top and 10 .

- At Using Measure : select SaleTourisPercentageOfSaleAmount then click OK.

6) The result shows in graph form (see Figure 5.108) and shows in data table (see Figure 5.109)

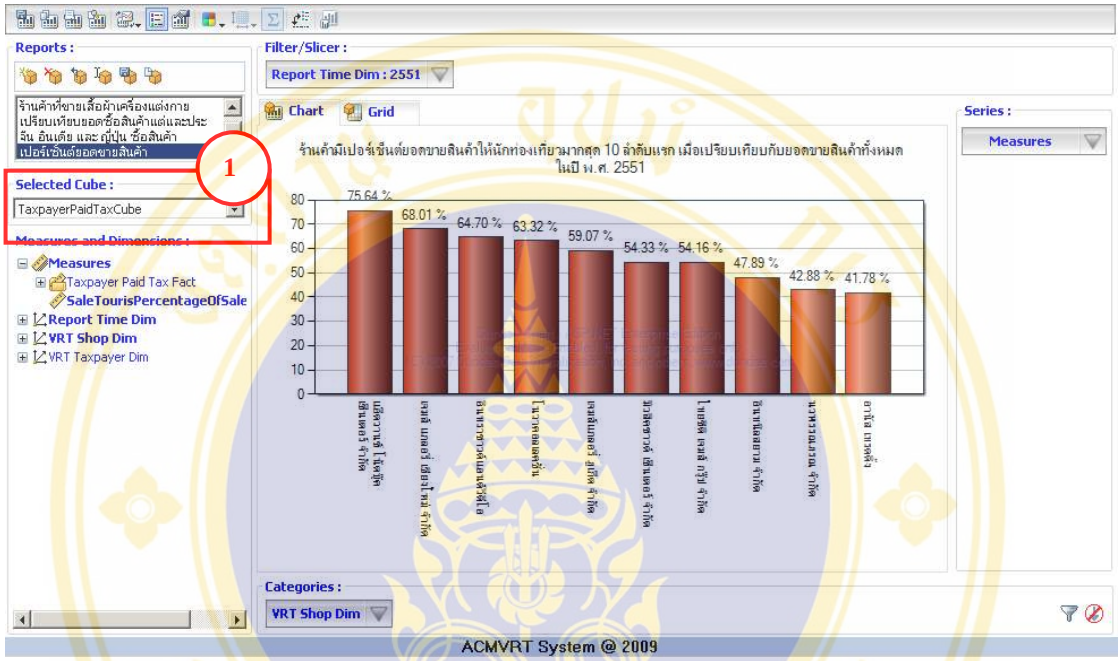


Figure 5.103 Select EntrepreneursPaidTaxCube

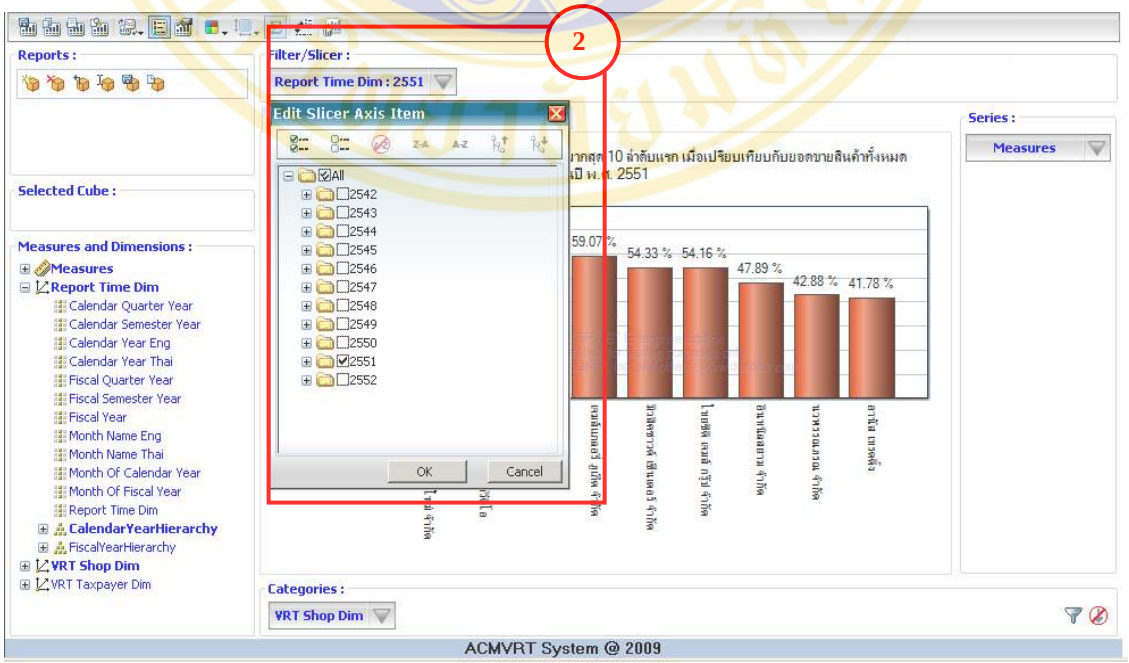


Figure 5.104 Select Refund VAT Time Dim and Calendar Year Thai

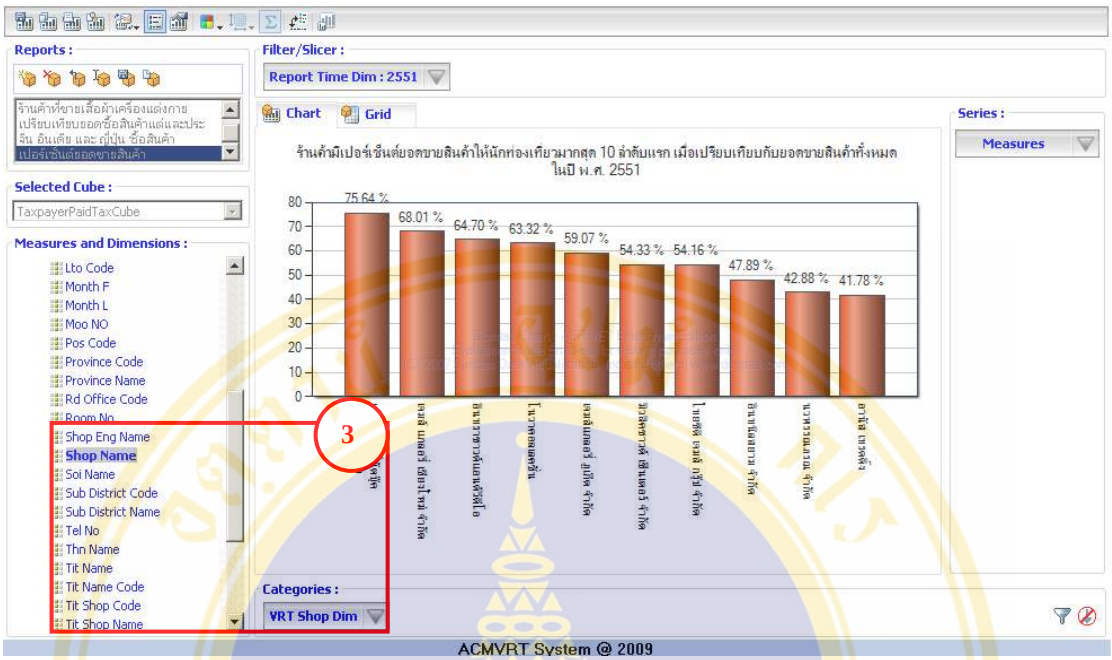


Figure 5.105 Select VRT Shop Dim and drag Shop Name

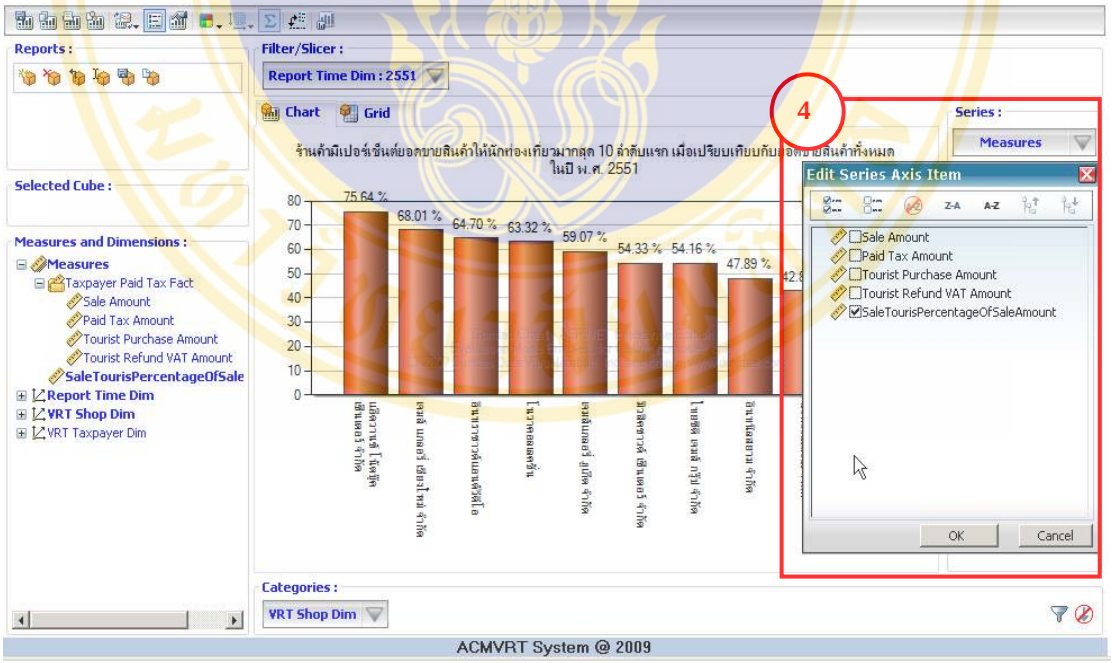


Figure 5.106 Select Measures and SaleTourisPercentageOfSaleAmount

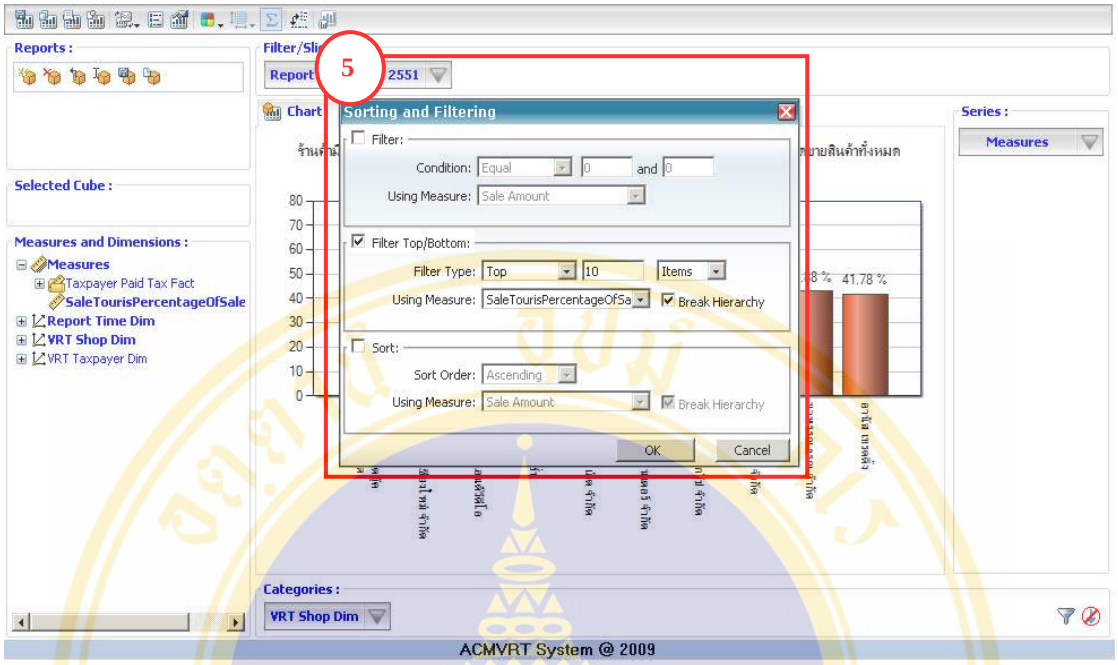


Figure 5.107 Filter top 10 shops which have highest percentage of tourists sales amount compare with total sales amount

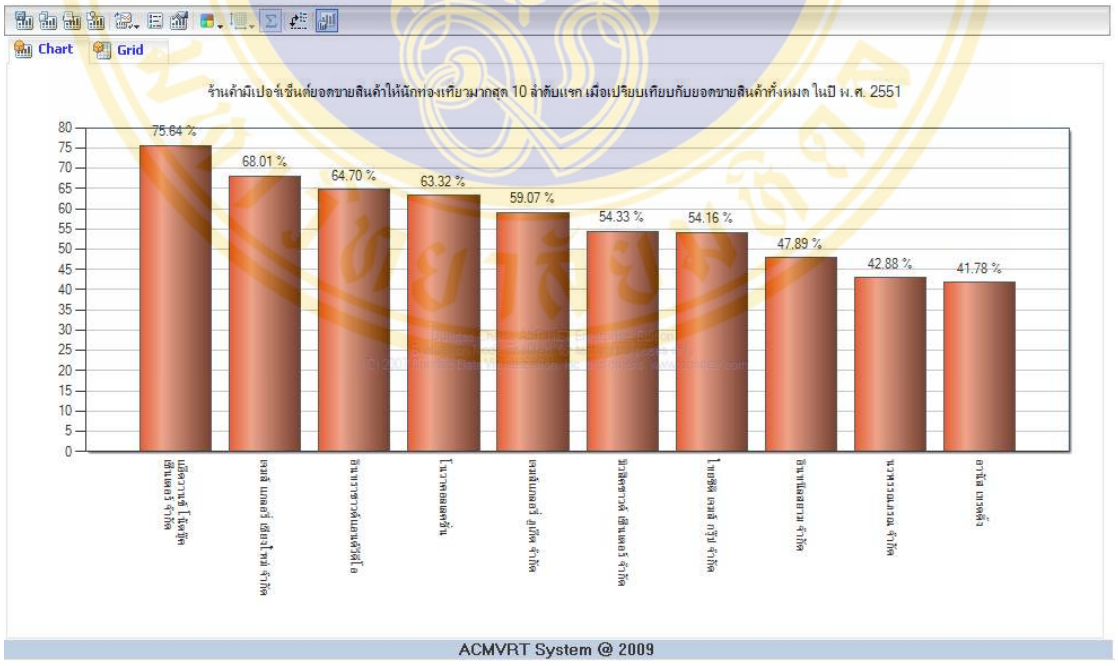


Figure 5.108 Graph of top 10 shops which have highest percentage of tourists sales amount compare with total sales amount

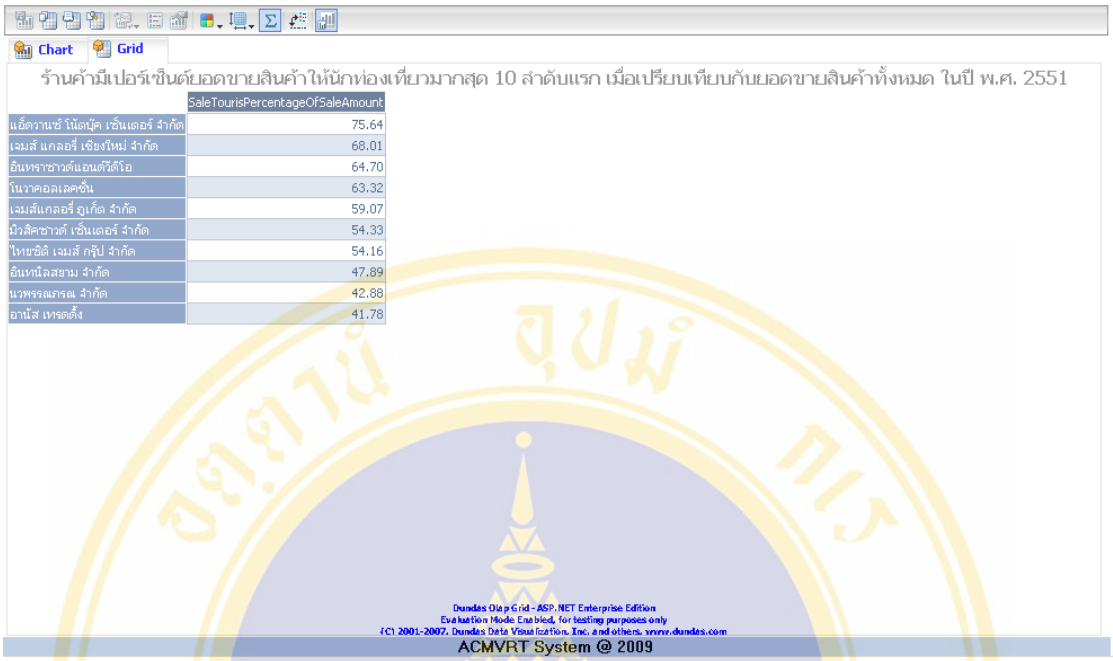


Figure 5.109 Table of top 10 shops which have highest percentage of tourists sales amount compare with total sales amount

5.3.2 Examples of Using Data Mining Tool

Data Mining Tool use for analysis tourists purchasing behaviors and use information to increase new entrepreneurs whose shops located in tourism area, and to support tourists' demand and promote VAT entrepreneurs products. In this section we have shown four example of using Data Mining Tool.

Minimum Support value configuration is to configure the minimum number of transactions which have a given itemset compared to total number of transactions before creating the Rule, can be configured from 0.00-1.00, in this case the default value is 0.01. The configured of Minimum Support value will affect the processing time, if configured less the Minimum Support value will take long processing time.

Minimum Probability value configuration is to set reliability (Confident) of the Rule, which can be configured values ranging from 0.00-1.00. Minimum Probability value will affect the number of Rule; the high numbers will create less Rule.

Example 1: Officers want to study the relationship of tourists and products. And release the information to entrepreneurs in VRT system. See figure 5.110 for more details.

1) Select condition “ความสัมพันธ์ระหว่างนักท่องเที่ยวแต่ละประเทศกับสินค้า”
2) System will selected and show data automatically
3) Set Minimum Probability value
4) Click “ตกลง”
5) Display occurred rules

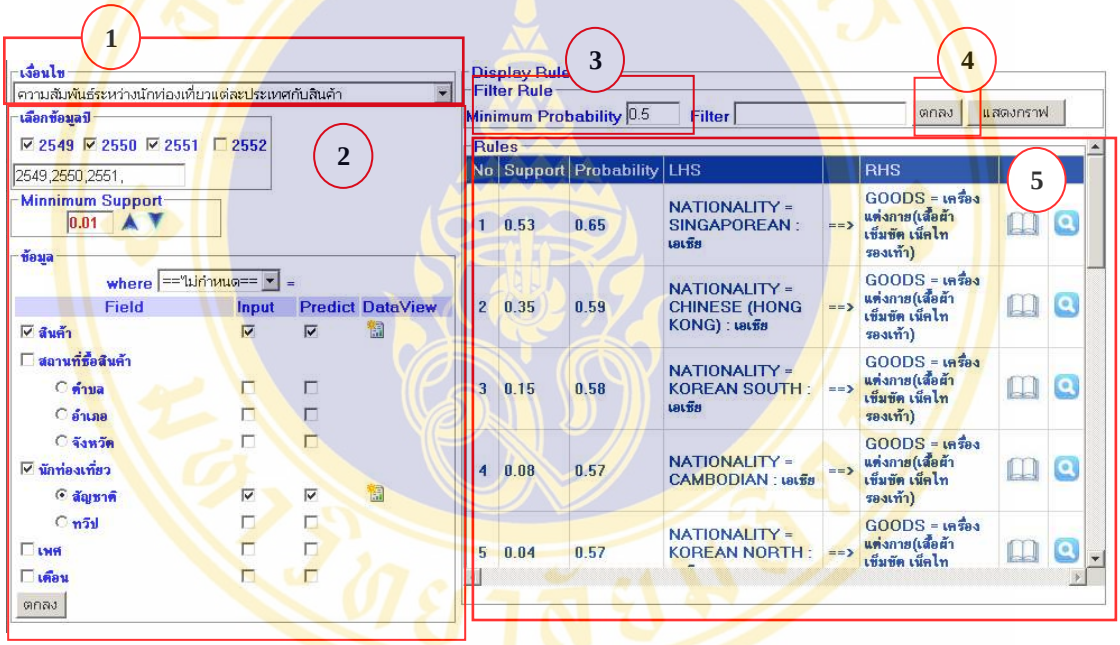


Figure 5.110 Show steps of creating rules of relationship between tourists from different country and products

Table 5.1 Top 10 rules of relationship between tourists from different country and products from data of year 2006-2008

No	Support	Probability	LHS		RHS
1	0.53	0.65	NATIONALITY = SINGAPOREAN : เอเชีย	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)

No	Support	Probability	LHS		RHS
2	0.35	0.59	NATIONALITY = CHINESE (HONG KONG) : เอเชีย	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
3	0.15	0.58	NATIONALITY = KOREAN SOUTH : เอเชีย	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
4	0.08	0.57	NATIONALITY = CAMBODIAN : เอเชีย	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
5	0.04	0.57	NATIONALITY = KOREAN NORTH : เอเชีย	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
6	0.27	0.56	NATIONALITY = MALAYSIAN : เอเชีย	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
7	0.14	0.54	NATIONALITY = RUSSIA : ยุโรป	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
8	0.01	0.53	NATIONALITY = UKRAINE : ยุโรป	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
9	0.12	0.53	NATIONALITY =	==>	GOODS = เครื่องแต่ง

No	Support	Probability	LHS		RHS
			TAIWAN : เอเชีย		กาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
10	0.10	0.52	NATIONALITY = INDONESIAN : เอเชีย	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)

From table 5.1: Display Rules of relationship between tourists from different country and products use data from year 2549-2551 (or 2006-2008) and set Minimum Support value = 0.01 and Minimum Probability value = 0.5

The result shows that the popular products among international tourists are apparel (cloth, belt, necktie and shoe). The SINGAPOREAN has Support value = 0.53, and highest Probability (Confident) value = 0.65, the second is CHINESE (HONG KONG) has Support value = 0.35 and Probability (Confident) value = 0.59, and the third is SOUTH KOREAN has Support value = 0.15, and Probability (Confident) value = 0.58.

We can conclude that the popular product types for tourists are the apparel. Therefore, we should invite entrepreneurs who sell apparel join with the VRT system.

Example 2: User want to study tourists behavior who purchase apparel products, and provided information to invite new apparel entrepreneurs to be member of VRT system (See Figure 5.111).

1) Select condition “ความสัมพันธ์ของนักท่องเที่ยวที่ซื้อเสื้อผ้าเครื่องแต่งกาย”.
2) System will selected and show data automatically.
3) Set Minimum Probability value.
4) Click “ตกลง”.
5) Display occurred rules.

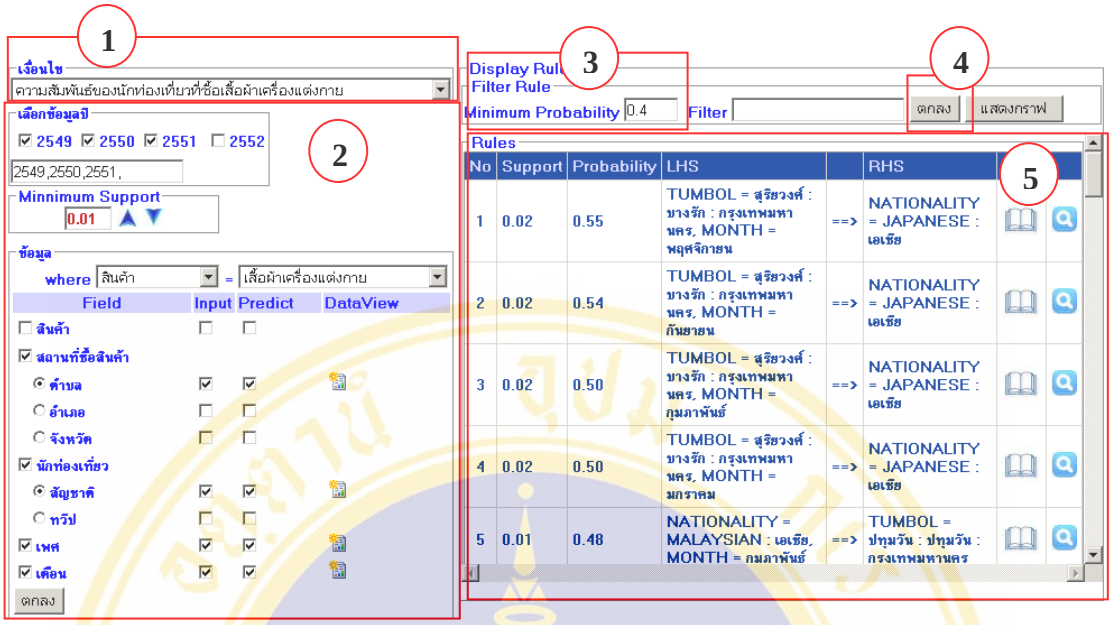


Figure 5.111 Show steps of creating steps rules of relationship of tourists and apparel products

Table 5.2 Top 10 rules of relationship of tourists and apparel products from data of year 2006-2008

No	Support	Probability	LHS		RHS
1	0.02	0.55	TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร, MONTH = พฤศจิกายน	==>	NATIONALITY = JAPANESE : เอเชีย
2	0.02	0.54	TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร, MONTH = กันยายน	==>	NATIONALITY = JAPANESE : เอเชีย
3	0.02	0.5	TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร, MONTH = กุมภาพันธ์	==>	NATIONALITY = JAPANESE : เอเชีย
4	0.02	0.5	TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร,	==>	NATIONALITY = JAPANESE : เอเชีย

No	Support	Probability	LHS		RHS
			MONTH = มกราคม		
5	0.01	0.48	NATIONALITY = MALAYSIAN : เอเชีย , MONTH = กุมภาพันธ์	==>	TUMBOL = ปทุมวัน : ปทุมวัน : กรุงเทพมหานคร
6	0.01	0.48	NATIONALITY = MALAYSIAN : เอเชีย , MONTH = มีนาคม	==>	TUMBOL = ปทุมวัน : ปทุมวัน : กรุงเทพมหานคร
7	0.19	0.47	TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร	==>	NATIONALITY = JAPANESE : เอเชีย
8	0.02	0.47	NATIONALITY = CHINESE (HONG KONG) : เอเชีย, MONTH = กุมภาพันธ์	==>	TUMBOL = ปทุมวัน : ปทุมวัน : กรุงเทพมหานคร
9	0.02	0.47	NATIONALITY = CHINESE (HONG KONG) : เอเชีย, MONTH = เมษายน	==>	TUMBOL = ปทุมวัน : ปทุมวัน : กรุงเทพมหานคร
10	0.01	0.46	NATIONALITY = CHINESE (HONG KONG) : เอเชีย, MONTH = พฤษภาคม	==>	TUMBOL = ปทุมวัน : ปทุมวัน : กรุงเทพมหานคร

From table 5.2 the Rule which created from the relationship between tourists and apparel products from data of year 2549-2551 (or 2006-2008). By set Minimum Support value = 0.01 and Minimum Probability value = 0.4.

Table show that JAPANESE tourists purchase products from shops in Suriyawong sub-district, Bangkok district, Bangkok most in November with Support value = 0.02 and Probability (Confident) value = 0.55, the second order month is September with Support value = 0.02 and Probability (Confident) value = 0.54 and the third order month is February with Support value = 0.02 and Probability (Confident)value = 0.5.

There for, we can conclude that JAPANESE tourists purchased apparel products at Suriyawong:Bangrak:Bangkok. Officers can provide this information to VRT shops in this location plan to launch apparel products in November, September and February.

Example 3: To study purchase behavior of Asian tourists and bring information to entrepreneurs who have Asian customer prepare their marketing plan, and to increase number of entrepreneurs in VRT system (See Figure 5.112).

1) Select condition “การซื้อสินค้าของนักท่องเที่ยวทวีปเอเชีย”.

2) System will selected and show data automatically.

3) Set Minimum Probability value.

4) Click “ตกลง”.

5) Display occurred rules.

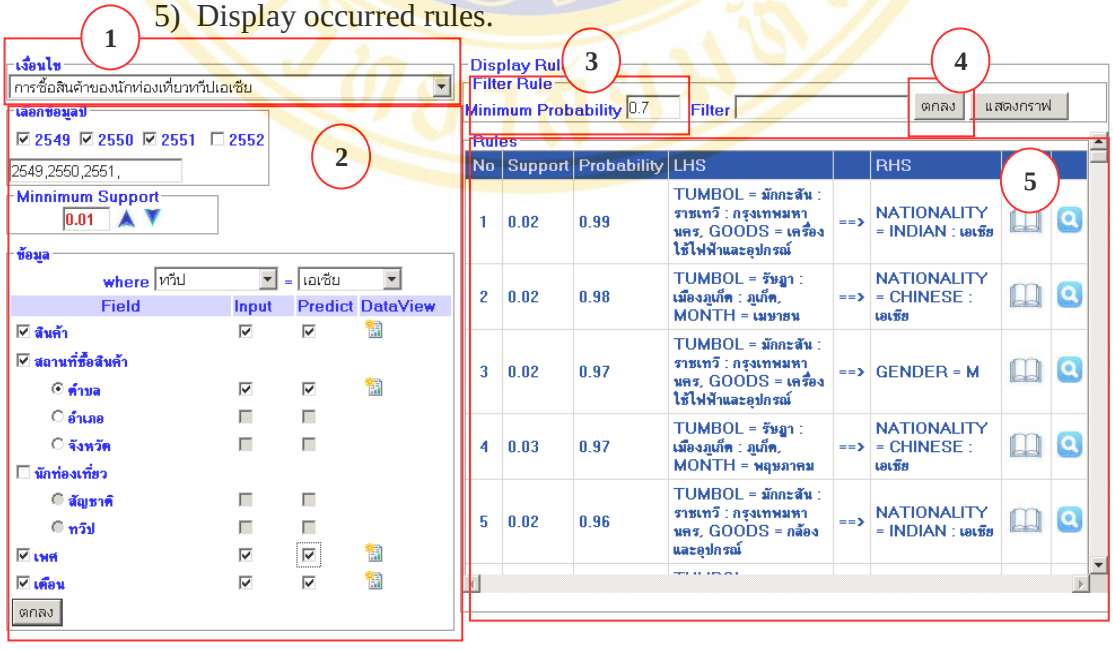


Figure 5.112 Show steps of creating rules for analyze Asian tourists purchase behavior

Table 5.3 Top 10 rules from relationship among Asian tourists from data of year 2006-2008

No	Support	Probability	LHS		RHS
1	0.02	0.99	TUMBOL = มักกะสัน : ราชเทวี : กรุงเทพมหานคร, GOODS = เครื่องใช้ไฟฟ้าและอุปกรณ์	==>	NATIONALITY = INDIAN : เอเชีย
2	0.02	0.94	TUMBOL = จักรวรรดิ : สัมพันธวงศ์ : กรุงเทพมหานคร, GOODS = เครื่องใช้ไฟฟ้าและอุปกรณ์	==>	NATIONALITY = INDIAN : เอเชีย
3	0.02	0.94	NATIONALITY = TAIWAN : เอเชีย, GOODS = อัญมณีที่ประกอบเป็นตัวเรือน	==>	TUMBOL = สันกลาง : สันกำแพง : เชียงใหม่
4	0.03	0.93	TUMBOL = ถนนพญาไท : ราชเทวี : กรุงเทพมหานคร, GOODS = เครื่องใช้ไฟฟ้าและอุปกรณ์	==>	NATIONALITY = INDIAN : เอเชีย
5	0.02	0.92	GOODS = เครื่องเงินเครื่องประดับ, TUMBOL = รัษฎา : เมืองภูเก็ต : ภูเก็ต	==>	NATIONALITY = CHINESE : เอเชีย
6	0.09	0.91	NATIONALITY = CHINA (MACAO) :	==>	TUMBOL = รัษฎา : เมืองภูเก็ต : ภูเก็ต

No	Support	Probability	LHS		RHS
			เอเชีย, GOODS = อัญมณีที่ประกอบเป็นตัวเรือน		
7	0.04	0.87	TUMBOL = วังใหม่ : ปทุมวัน : กรุงเทพมหานคร, NATIONALITY = SINGAPOREAN : เอเชีย	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
8	0.05	0.82	NATIONALITY = MALAYSIAN : เอเชีย, TUMBOL = วังใหม่ : ปทุมวัน : กรุงเทพมหานคร	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
9	0.03	0.81	TUMBOL = สันกลาง : สันกำแพง : เชียงใหม่, NATIONALITY = MALAYSIAN : เอเชีย	==>	GOODS = อัญมณีที่ประกอบเป็นตัวเรือน
10	0.02	0.81	TUMBOL = รัษฎา : เมืองภูเก็ต : ภูเก็ต, GOODS = เครื่องหนัง	==>	NATIONALITY = CHINESE : เอเชีย

From table 5.3 show to 10 Rules, result from analysis Asian tourists purchasing data of year 2549-2551 (or 2006-2008) and set Minimum Support value = 0.01 and Minimum Probability value = 0.7. We found that INDIAN tourists purchase electric appliances at Makkasan:Ratchadewi:Bangkok most with Support value = 0.02 and Probability (Confident) value = 0.99 . The second place is Chakkawat:Sampantawong:Bangkok with Support value = 0.02 and Probability (Confident) value = 0.94, and the third place is Phayathai Road:Ratchadewi:Bangkok with Support value = 0.03 and Probability (Confident) value = 0.93.

TAIWAN tourists purchase gem jewellery from shops located in Sanlang:Sankampaeng:Chiengmai with Support value = 0.02 and Probability (Confident) value = 0.94.

CHINESE tourists purchase silver and jewellery from shops located in Rassada: Muangphuket:Phuket with Support value = 0.02 and Probability (Confident) value = 0.92.

From this example we realize that which product is popular in Asian tourists and where these shops are located. VRT officers can plan to invite these shops to be member of VRT system.

Example 4: To study Japanese tourists behavior and products they like to buy, from data of years 2549-2551 (or 2006 to 2008), and bring the information to entrepreneurs use for marketing plan.

1) Select condition “การซื้อสินค้าของนักท่องเที่ยวสัญชาติญี่ปุ่น” .
2) System will selected and show data automatically.
3) Set Minimum Probability value.
4) Click “ตกลง”.
5) Display occurred rules.

Figure 5.113 Show steps of creating rules to analyses Japanese purchase behavior

Table 5.4 Top 10 Rules: result of analysis Japanese purchase behavior from data of year 2006-2008

No	Support	Probability	LHS		RHS
1	0.02	0.69	TUMBOL = คลองตันเหนือ : วัฒนา : กรุงเทพมหานคร	==>	GOODS = เครื่องหนัง
2	0.02	0.67	MONTH = มกราคม, TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
3	0.03	0.67	GOODS = ผ้า(ยังไม่แปรรูป)	==>	TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร
4	0.02	0.66	MONTH = กุมภาพันธ์, TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
5	0.02	0.65	MONTH = เมษายน, TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
6	0.02	0.64	MONTH = มีนาคม, TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
7	0.01	0.63	MONTH = พฤษภาคม, TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
8	0.19	0.62	TUMBOL = สุริยวงศ์ : บาง	==>	GOODS = เครื่องแต่ง

No	Support	Probability	LHS		RHS
			รัก : กรุงเทพมหานคร		กาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
9	0.02	0.61	MONTH = กันยายน, TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)
10	0.02	0.59	MONTH = ตุลาคม, TUMBOL = สุริยวงศ์ : บางรัก : กรุงเทพมหานคร	==>	GOODS = เครื่องแต่งกาย(เสื้อผ้า เข็มขัด เน็คไท รองเท้า)

From Table 5.4 shows result of analyze Japanese tourists purchase behavior using year 2549-2551 (or 2006-2008) data with Minimum Support value = 0.01 and Minimum Probability value = 0.5, the three most Probability value Rules are:

- Japanese tourists purchase leather product at Klongtan Nua:Wattana:Bangkok with Support value = 0.02 and Probability value = 0.69 .
- Japanese tourists purchase apparel products in January at Suriyawong:Bangrak:Bangkok with Support value = 0.02 and Probability value = 0.67 .
- Japanese tourists purchase cloth in February at Suriyawong:Bangrak:Bangkok with Support value = 0.03 and Probability value = 0.67 .

Officers can provide information of Japanese purchase behavior to entrepreneurs to make marketing plan and increase sales' opportunity.

CHAPTER VI
SYSTEM EVALUATION AND RESULTS

This chapter describes the evaluation of ACRM-VRT system by measuring the satisfaction of the ACRM-VRT users. The chapter consists of evaluation objectives, evaluation methods and results.

6.1 Evaluation Objective

The evaluation of the ACRM-VRT system aims at measuring the satisfaction of the ACRM-VRT users which are VRT officers. This evaluation is used for examining the accuracy of the ACRM-VRT development and agreement of the users' requirement. We, therefore, can deploy the results to improve the VRT system in the future.

6.2 Evaluation Method

The evaluation method for the ACRM-VRT system is to measure the satisfaction of the ACRM-VRT users. The VRT users will work with each function in the ACRM-VRT systems and complete those questionnaires to measure the satisfaction for those functions. The details are below.

1) A group of evaluators consists of 22 VRT officers who are:
   - 12 officers in the suggestion and auditing team.
   - 10 officers in the information and technology team.
2) Tools for the ACRM-VRT system evaluation are:
   - Laptop Computer installed the ACRM-VRT system, the specification is:
     - Processor : Intel Pentium III 1.3 GHz
     - RAM: 1.5 GB
     - OS: Windows Server 2003 Enterprise Edition
   - The questionnaire is an online one that can be divided into 2 steps :

- The ACRM-VRT evaluator Login : the evaluators login by using the officer ID as shown in Figure 6.1.
- The ACRM-VRT system questionnaire : as shown in Figure 6.2 This questionnaire is separated into 2 parts :
  (a) General Information for the evaluators.
  (b) The satisfaction for utilizing the function of the ACRM-VRT system.

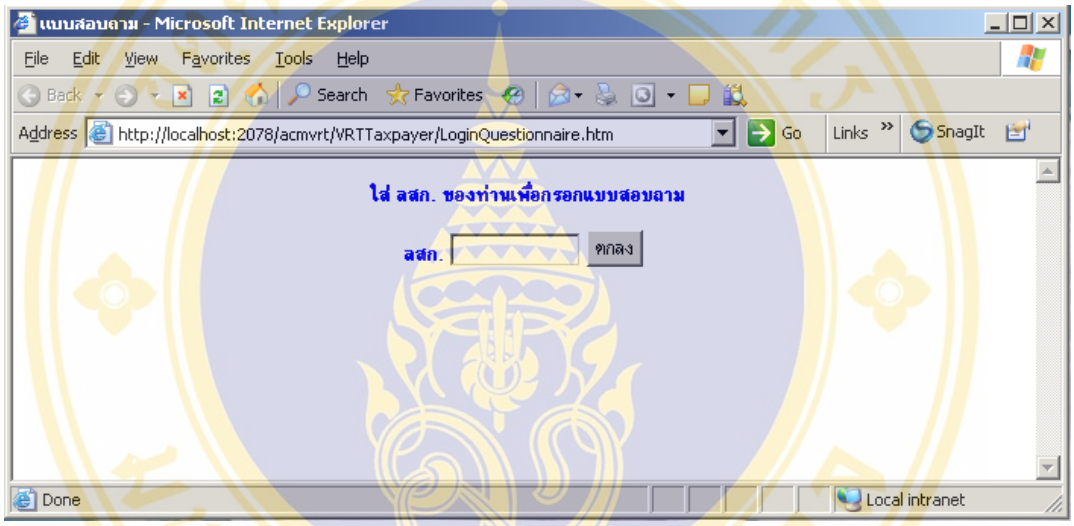


Figure 6.1 Display the login screen for the evaluators for the ACRM-VRT questionnaires

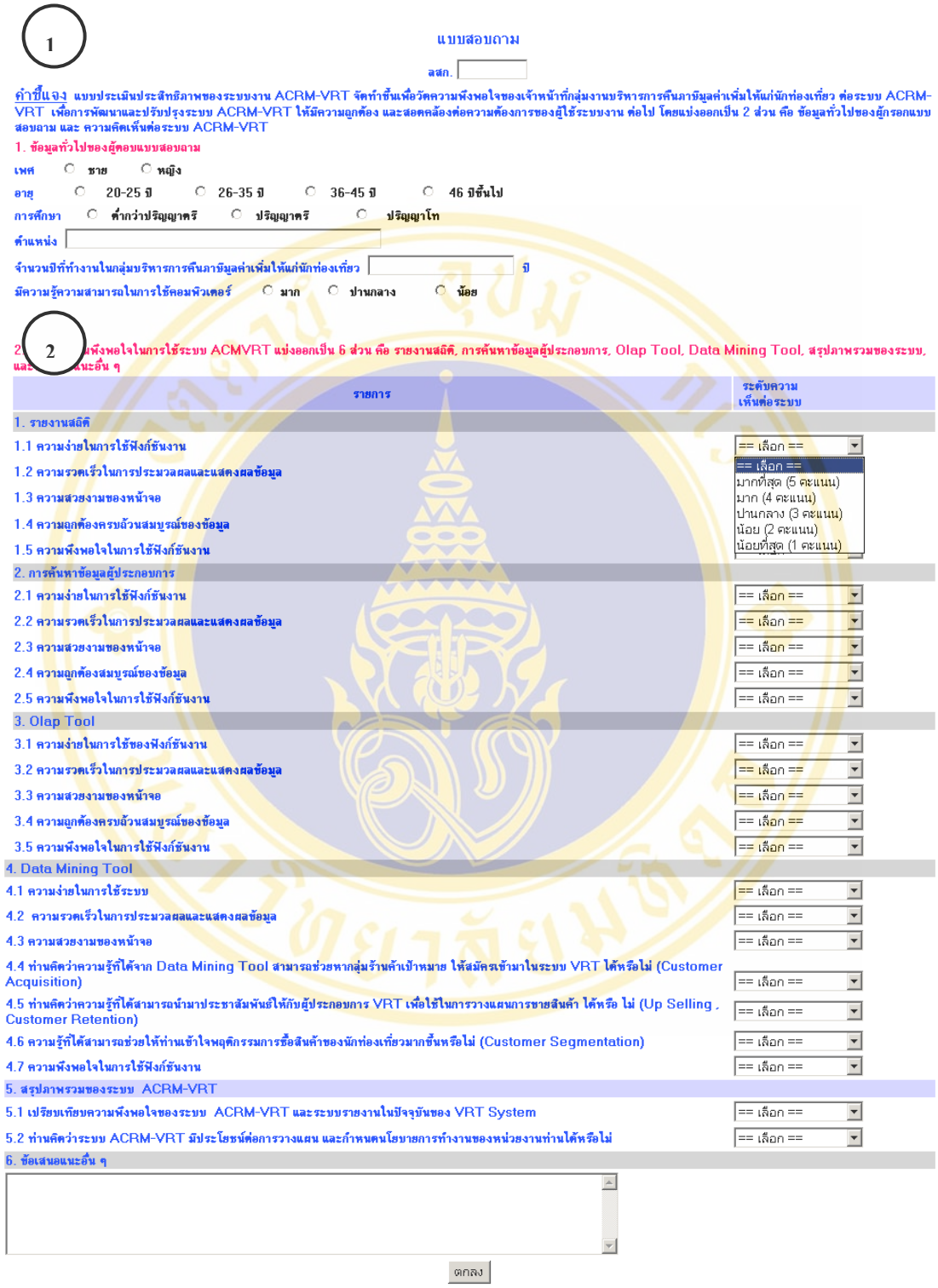


Figure 6.2 Display screen for the ACRM-VRT questionnaire

3) Evaluation step are:
   - Divide all evaluators into 3 persons per group.
   - Explain the objective and the working step of the ACRM-VRT system, then the evaluators work and test each function. Each group spends one hour to test the system.
   - After that the evaluators complete the questionnaire.
4) Data Collection, The data is collected from the online questionnaire from 22 evaluators.
5) Data Analysis are:
   - Analyzed general information based on frequency and ratio.
   - Analyzed the satisfaction of the ACRM-VRT system by using average ($\overline{X}$) and standard deviation.

Formula for Average

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Formula for Standard Deviation

$$S.D. = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$

Criteria for average interpretation

The criteria for evaluation, or rating scale, has 5 levels. The criteria for average interpretation follows Best's Model (Best 1981: 179-187) as shown.

Table 6.1 Criteria for average interpretation

No	Average	Meaning
1	4.50 - 5.00	Excellent
2	3.50 - 4.49	Good
3	2.50 - 3.49	Fairly good
4	1.50 - 2.49	Fairly poor
5	1.00 - 1.49	Poor

6.3 Evaluation Results

From the evaluation of the ACRM-VRT results, we found that there are 22 of the VRT officers, out of 25, since there are three officers that did not participate in the assessment. The evaluation results are shown as follows.

6.3.1 Summary of Evaluators

1) Sex : the majority of evaluators is female. There are 18 female officers, about 82 percent.

2) Age : mainly in the age of 26-35 years. There are 8 officers in the age of 26-35 years, about 36 percent.

3) Education : mainly at the undergraduate level. There are 16 officers in the undergraduate level, about 73 percent.

4) Rank : mainly in the rank of tax auditor senior level. There are 16 officers in the tax auditor at senior level, about 73 percent.

5) Working experience in the VRT system : mainly work with the system for more than 9 years There are 8 experienced officers, about 36 percent.

6) Computer knowledge and skill: moderate level. There are 15 officers, about 68 percent.

Table 6.2 The number and percent of the system evaluators

No	Total	Percent
Sex :		
- Male	4	18
- Female	18	82
Age :		
20-25 year olds	0	0
26-35 year olds	8	36
36-45 year olds	7	32
more than 46 year olds	7	32
Education :		
Less Than Bachelor's Degree	2	9

No	Total	Percent
Bachelor's Degree	16	73
Master's Degree	4	18
Doctorate's Degree	0	0
Rank :		
Revenue Officer Practitioner Level	3	14
Revenue Officer Professional Level	3	14
Tax auditor senior level	16	73
Working experience in the VRT system :		
Lest than one year	4	18
1-3 year	3	14
4-6 year	4	18
7-9 year	3	14
More than nine year	8	36
Computer knowledge and skill :		
High	2	9
Moderate	15	68
Low	5	23

6.3.2 Summary of The Evaluator Satisfaction

The result is summarized by using the function utilization as shown in the following.

- **Statistic Report**

The satisfaction of function utilization is in the good level. The average is 4.14. The beauty of the display is high and the lowest average is 3.82.

- **Searching VRT Taxpayer Data**

The satisfaction of function utilization is in the good level. The average is 4.00. The beauty of the display is high and the lowest average is 3.59.

- **OLAP Tool**

The satisfaction of function utilization is in the good level. The average is 4.00. The simplicity of the function utilization is good and the lowest average is 3.73.

- **Data Mining Tool**

The satisfaction of function utilization is in the good level. The average is 4.23. The simplicity of the function utilization is good and the lowest average is 3.59.

- **Overall Results of The ACRM-VRT System**

This part detail in the following:

- Comparing with the existing VRT system, the user satisfaction of the developed VRT system is in a good level and the average is 4.27.

- The satisfaction for planning and defining the VRT policy is in a good level and the average is 4.18.

Table 6.3 Summary for the User Satisfaction for Each Function

Question	5 Excellent	4 Good	3 Fairly good	2 Fairly poor	1 Poor	Avg.	Avg. (SD)	Result
1. Statistic Report								
1.1 Simplicity of function utilization	2	16	4	0	0	3.91	0.53	Good (78.18%)
1.2 Speed of processing and result display	3	16	3	0	0	4.00	0.53	Good (80.00%)
1.3 Beauty of the display	2	14	6	0	0	3.82	0.59	Good (76.36%)
1.4 Accuracy and completeness	3	16	3	0	0	4.00	0.53	Good (80.00%)
1.5 Satisfaction of the function utilization	6	13	3	0	0	4.14	0.64	Good (82.73%)
Total	**16**	**75**	**19**	**0**	**0**	**3.97**	**0.57**	**Good (79.45%)**
2. Searching VRT Taxpayer Data								
2.1 Simplicity of function utilization	6	14	2	0	0	4.18	0.59	Good (83.64%)
2.2 Speed of processing and result display	3	14	5	0	0	3.91	0.61	Good (78.18%)
2.3 Beauty of the display	1	11	10	0	0	3.59	0.59	Good (71.82%)
2.4 Accuracy and completeness	5	17	3	0	0	4.64	0.49	Excellent (92.73%)
2.5 Satisfaction of the function utilization	4	14	4	0	0	4.00	0.62	Good (80.00%)
Total	**19**	**70**	**24**	**0**	**0**	**3.96**	**0.60**	**Good (79.12%)**

Question	5 Excellent	4 Good	3 Fairly good	2 Fairly poor	1 Poor	Avg.	Avg. (SD)	Result
3. OLAP Tool								
3.1 Simplicity of function utilization	1	12	9	0	0	3.64	0.58	Good (72.73%)
3.2 Speed of processing and result display	3	18	1	0	0	4.09	0.43	Good (81.82%)
3.3 Beauty of the display	4	15	3	0	0	4.05	0.58	Good (80.91%)
3.4 Accuracy and completeness	7	13	2	0	0	4.23	0.61	Good (84.55%)
3.5 Satisfaction of the function utilization	6	10	6	0	0	4.00	0.76	Good (80.00%)
Total	**21**	**68**	**21**	**0**	**0**	**4.00**	**0.62**	**Good (80.00%)**
4. Data Mining Tool								
4.1 Simplicity of function utilization	2	12	8	0	0	3.73	0.63	Good (74.55%)
4.2 Speed of processing and result display	3	17	2	0	0	4.05	0.49	Good (80.91%)
4.3 Beauty of the display	2	17	3	0	0	3.95	0.49	Good (79.09%)
4.4 Do you think that the obtained information from data mining tool can help the RD to search for the VRT taxpayer targets in order to recruit them into the VRT system?	9	10	3	0	0	4.27	0.70	Good (85.45%)
4.5 Do you think the obtained information can be used for advertising to the VRT taxpayers to	11	8	3	0	0	4.36	0.73	Good (87.27%)

Question	5 Excellent	4 Good	3 Fairly good	2 Fairly poor	1 Poor	Avg.	Avg. (SD)	Result
help them make their marketing plan ?								
4.6 Do you think the obtained information can help the VRT taxpayers to better understand the tourist's consumption behavior?	8	11	2	1	0	4.18	0.80	Good (83.64%)
4.7 The satisfaction of function utilization	8	11	3	0	0	4.23	0.69	Good (84.55%)
Total	**43**	**86**	**24**	**1**	**0**	**4.11**	**0.67**	**Good (82.21%)**
5. Overall results for the ACRM-VRT system								
5.1 Comparing the ACRM-VRT system and the current report system in the VRT system	8	12	2	0	0	4.27	0.63	Good (85.45%)
5.2 The satisfaction of the ACRM-VRT system for making a plan and specifying the action plan for the VRT team	8	11	2	1	0	4.18	0.80	Good (83.64%)
Total	**86**	**139**	**36**	**3**	**0**	**4.17**	**0.71**	**Good (83.33%)**

6.3.3 Overall Satisfaction of The Evaluators

From the overall satisfaction for each function, the results show that data mining tool has the highest satisfactory level, about 82.21 percent. The second one is the OLAP tool that has about 80 percent. The statistic report is about 79.12 and the taxpayer searching is about 79.12 percent in respectively, as shown in Figure 6.3

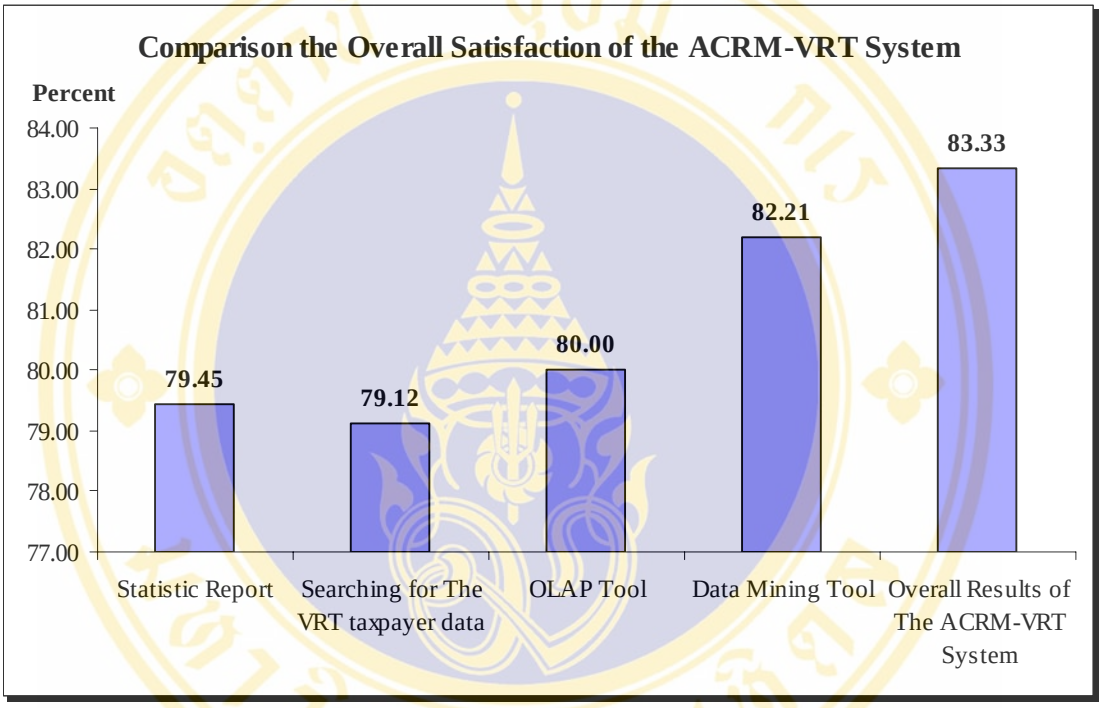


Figure 6.3 Comparison the overall satisfaction of the ACRM-VRT System

From figure 6.3, from the overall satisfaction for each function, the evaluators are highly satisfied with the data mining tool. This is because the officers can use the obtained information for searching the target taxpayers who sell the products that meet tourists' needs and who locate in the tourist shopping areas (Customer Acquisition). The obtained information can help the VRT officers to have better understanding about the tourists' good purchase behavior (Customer Segmentation). This useful information can also use for workshop and advertisement to the VRT taxpayers for making their marketing and shelving plan (Up selling, Customer retention).The second tool is OLAP tool. This tool facilitates for generating different kind of reports which are corresponding to users' needs (Ad hoc report). The

VRT officers don't have to wait for the VRT software developers to generate a new report. Finally overall results for the ACRM-VRT system is good, about 83.33 percent.

6.3.4 Discussion

From the evaluation result of the user satisfaction we observed that some parts of the system have low value of user satisfaction as discussed below.

1) Statistic Report Function: Results from the survey show that the satisfaction level is Good, but less of the average, which contains the beauty of the display has the lowest average is 3.82, and the simplicity of function utilization with an average 3.91.
2) Searching VRT Taxpayer Data Function: Results from the survey show that the satisfaction level is good, but less of the average, which contains the beauty of the display has the lowest average is 3.59, and the speed of processing and result display with an average 3.91.
3) OLAP Tool Function: Results from the survey show that the Satisfaction level is good, but the simplicity of function utilization less of the average with 3.64. Because the majority of evaluators have computer knowledge and skill moderate level. And no basic knowledge in online data analysis.
4) Data Mining Tool Function: Results from the survey show that the satisfaction level is good, but less of the average, which contains the simplicity of function utilization with an average 3.73 and the beauty of the display has an average 3.95. Furthermore numbers of rules make officer can not use the knowledge immediately.

From the above observations, we should bring the information to improve and develop ACRMVRT systems to look beauty and easy to use and suitable for basic knowledge of the user.

CHAPTER VII
CONCLUSION AND FUTURE WORK

The chapter provides a conclusion of this research project and the future work.

7.1 Conclusion

This research project applies the concept of CRM, data warehousing, OLAP, and data mining for designing and developing the ACRM-VRT system which supports the VRT officers at the RD for making static and online report for VAT refund for tourists. The ACRM-VRT system can also analyze the tourists' consumption behavior and use the obtained information to find new groups of VAT taxpayers that should be invited or persuaded to join the VRT system. The details are as the following.

1) Building the Data Mart for collecting the VRT taxpayer data and the data for VAT refund for tourists. The data is imported from the VRT system and FVAT system.

2) Building the ACRM-VRT system that has four main functions:

- **Statistic Report** : This function presents report for the number of taxpayers, stores, VAT refund for tourists. There are about 12 reports, including 5 of the annual report of the current fiscal year and 7 of the statistical report that comparing different fiscal years.
- **Searching for The VRT Taxpayer** : This function is used for searching and showing the data of VRT taxpayer and entrepreneur.
- **OLAP Tool** : This function is used for the VRT officers for analyzing the online data. The report can be displayed in multidimensional and generated as the VRT officers need.

- **Data Mining Tool** : This function is in an analysis tool. The association rule mining method, or apiori algorithm, is applied for analyzing the tourists' good purchase behavior.

The ACRM-VRT system is a web application software that is developed by using ASP.NET. The data mart is developed by the Microsoft SQL server 2005 and the Microsoft SQL 2005 Analysis Service is brought in for organizing and summarizing data into the data cube format. Data Cube supports for querying data from the Statistic Report function, the Searching for the VRT taxpayer function, and the OLAP Tool function. The Microsoft SQL 2005 Reporting Service is used for designing and generating the report. The Microsoft Association Rule Algorithm of the Microsoft SQL 2005 Analysis Service is used in the Data Mining function for analyzing the data for generating association rules.

In the evaluation for user satisfaction in terms of function utilization, 22 of the VRT officers are highly satisfied with all four functions. The function that has the highest satisfactory level is data mining tool with an average of 4.11. The OLAP tool has an average of 4.00. The statistic report has an average of 3.97. The searching for the VRT taxpayer has an average of 3.96 respectively.

The contribution of this research project are as the following.

1) The revenue department has got a central data mart and tools that can analyze and work online for VAT refund for tourists. The VRT officers can reduce the time period for making report and analyzing the data.

2) This study helps the VRT officers to manage their VAT refund. The VRT officers can take the obtained information to improve their advertising and increase their accessibility to the VRT taxpayers in order to make a network connection to the other VRT taxpayers.

3) This study helps the VRT officers to use the obtained information for shelving and marketing plan to increase their sales. The study also satisfies the member of the VRT network connection.

7.2 Future work

To improve the ACRM-VRT system, the following features will be developed in the future.

1) Develop the ETL process that can work automatically in order to reduce the working steps and facilitate for the data mart administrator. In this study, the part of data mart for the ETL process is still manually processed.

2) Develop a new function that can categorize the products into different groups. In the current system, the tourist shopping data is classified into 35 groups, but it is still unclear. For example, from the ACRM-VRT system, we found that the costume products are the most favorite for tourists. If we can define a kind of clothing such as pant, skirt, shirt, and shoe. The rule will be more beneficial to the VRT officers.

3) Develop web application that recommends the products and shops for the tourist needs by using the obtained information from the data mining tool. This will increase the marketing channel for the VRT taxpayers and persuade the taxpayers to apply for the VRT system.

4) Improve the ACRM-VRT system screen more beautiful and easy to use and suitable for basic knowledge of users. By the principles of HCI (Human Computer Interaction) to analysis and design system to serve more satisfaction to VRT officers.

5) Improve the reporting allow VRT officers to select models such as linear-graph, pie-graph and bar-graph to meet their needs. Furthermore, the system can automatically choose appropriate graph type matching the amount of information that occurs, for example if the data are many the system display linear-graph automatically.

6) Improve the OLAP Tool and Data Mining Tool functions more easy and convenient to use. The user can enter question to the system. Then, the system processed and displayed results automatically. For example:

- OLAP Tool function: The executive want to know number of tourists which have refunded tax in year 2007-2009 and trends of number of tourists will refund tax in fiscal year 2010. The system will display the graph shows trends of tourists' number in fiscal year 2010 automatically.

- Data Mining Tool functions: VRT officers want to know the Asian tourists' popular products, when and where which they were purchased those products. VRT officers enter a question that they want to know the system. Then, the system will configure the minimum support value, minimum confident value then processed and displayed the Rules' results automatically.

7) Add Data Mining Tool functions to help analyzing data for show the trends value of information that occur by using Microsoft Time Series algorithm. The executive can use the result to setup management strategies to better management, such as forecasting that the number of tourists will be refunded tax, forecast the tourists' sales amount etc.

REFERENCES

1. VAT Refund for Tourists in The Revenue Department. เกี่ยวกับการคืนภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยว [Online]. 2005. Available from URL:http://rdsrv.rd.go.th/22120.0.html [Accessed 2007 Feb 14].
2. VAT Refund for Tourists in The Revenue Department. รายงานการคืนภาษีมูลค่าเพิ่มให้แก่นักท่องเที่ยว [Online]. 2009. Available from URL:http://rdsrv.rd.go.th/18520.0.html [Accessed 2009 Oct 20].
3. VAT Refund for Tourists in The Revenue Department. แบบพิมพ์เกี่ยวกับระบบงานกรรมวิธี[Online]. 2007. Available from URL:http://rdsrv.rd.go.th/18520.0.html [Accessed 2007 Nov 14.
4. VAT Refund for Tourists in The Revenue Department. VAT Refund for Tourists [Online]. 2009. Available from URL:http://www.rd.go.th/VRT/ [Accessed 2009 Feb 14].
5. The National Tourism Alliance (NTA). Review of Tourist Shopping Arrangements in Australia [Online]. Available from URL:http://www.tourismalliance.org/tourismalliance/PDFs/NTA_TourismShoppingReview_Appendices.pdf [Accessed 2007 Feb 14].
6. Singapore Retailers Association. GST Tourist Refund Schemes [Online]. Available from URL:http://www.retail.org.sg/ [Accessed 2007 Feb 15].
7. Australian Customs Service. Tourists Refund Scheme [Online]. Available from URL: http://www.customs.gov.au/site/page.cfm?u=4366#e1161[Accessed 2007 Feb 14].
8. Global Refund Group. About Global Refund [Online]. Available from URL: http://www.globalrefund.com [Accessed 2007 Feb 20].

9. Michael J.A. Berry and Gordon S. Linoff. Mastering data mining the art and science of customer relationship management. New York: John Wiley & Sons; 2000.

10. Efraim Turban, Jay E.Aronson, and Ting-Peng Liang. Decision support system and intelligent system. New Jersey: Pearson Prentice Hall; 2005.

11. Catherien Bounsaythip and Esa Rinta-Runsala. Overview of data mining for customer behavior modeling. VTT Information Technology 2001 June 29; Version 1:1-49.

12. Yun Chen, Guozheng Zhang, Dengfeng Hu, and Shanshan Wang. Customer segmentation in customer relationship management based on data mining. International Federation for Information Processing (IFIP) 2006; Volume 207:288-293.

13. Malte Geib, Annette Reichold, Lutx Kolbe, and Walter Brenner. Architecture for customer relationship management approaches in financial services. IEEE Proceedings of the 38th Annual Hawaii International Conference; 2005 Jan 3-6; Hawaii, USA; 2005. p. 1-5.

14. Jaideep Srivastava, Jau-Hwang Wang, Ee-Peng Lim, and San-Yih Hwang. A case for analytical customer relationship management. ACM Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining; 2002 May 12; Taipei, Taiwan; 2002. p. 14-17.

15. Colleen Cunningham, ll-Yeol Song, and Peter P. Chen. Data warehouse design to support customer relationship management analyses. Proceedings of the 7th ACM international workshop on Data warehousing and OLAP; 2004 Nov 12; Washington DC, USA; 2004. p. 14-22.

16. Gediminas Adomavicius, and Alexander Tuzhilin. Using data mining methods to build customer profiles. The flagship magazine of the IEEE Computer Society 2001; Volume 34, Issue 2:74-82.

17. Fatma E. Giha, Y.P. Singh, and H.T. Ewe. Mining generalized customer profiles. AIML 06 International Conference; 2006 Jun 13-15; Berlin, Germany; 2006. p. 141-147.

18. Jaiwei Han and Micheline Kamber. Data mining concepts and techniques. San Francisco: Morgan Kaufmann; 2001.

19. CRoss Industry Standard Process for Data Mining. CRISP-DM 1.0 Step-by-step data mining guide [Online]. Available from URL:http://www.crisp-dm.org/CRISPWP-0800.pdf [Accessed 2007 Feb 20].

20. Using XML for Analysis in Analysis Services. Microsoft Corporation [Online]. Available from URL:http://msdn.microsoft.com/en-us/library/ms186654 (SQL.90).aspx [Accessed 2009 Aug 21].

21. Frada Burstein and Clyde W. Holsapple. Handbook on decision support systems 1 basic themes. Berlin:Springer Berlin Heidelberg; 2008.

22. Surajit Chaudhuri and Umeshwar Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record 1997; Volume 26, Issue 1: 65-74.

23. Bing Liu. Association rules and sequential patterns. Web data mining exploring hyperlinks, contents, and usage data. Berlin: Springer Berlin Heidelberg; 2007. p. 13-54.

BIOGRAPHY

NAME	Miss Somjai Samoemuan
DATE	February 9, 1977
PLACE OF BIRTH	Singburi, Thailand
INSTITUTIONS ATTENDED	Sripatum University, 1997-1999: Bachelor of Business Administration (Business Computer) Mahidol University, 2004-2009: Master of Science (Computer Science)
POSITION & OFFICE	2001 – Present, Revenue Department, Bangkok, Thailand. Position: Computer Technical officer
ADDRESS	103, Moo 9, Huaphai, Muang singburi, Singburi 16000, Thailand
E-MAIL	somjai.sa@rd.go.th