DATA MINING BASED ASSOCIATION RULES & RFM ANALYSIS IN INDIAN RETAIL SECTOR: AN EMPIRICAL INVESTIGATION

Dr. ANSHUL SHARMA
Asst. Professor
S. D. COLLEGE OF MANAGEMENT
ISRANA – 132 107

Prof. (Dr.) M. K. KULSHRESHTHA
Director
S. D. COLLEGE OF MANAGEMENT
ISRANA – 132 107

Prof. (Dr.) ASHOK AGRWAL
Dept. of Commerce
University of Rajasthan
JAIPUR

ABSTRACT

Retail in India has emerged as one of the most dynamic and fast paced industries with several players entering the market. The data that retail business collect about their customers is one of the greatest assets of it. Buried within this vast amount of data is all sorts of valuable information that could make a significant difference to the way in which any business organization run their business, and interact with their current and prospective customers and gaining the competitive edge on their competitors.

Data Mining based Association Rules are used in Indian Retail Industry for analyzing the data that describes transactions, lists of items, etc. The association rules are derived from patterns in a large datasets to determine which products are frequently purchased together and help in understanding the buying behaviour. Association results are easy to understand and association rules are easy to use.

Association analysis, commonly referred to as Market Basket Analysis, used by retail industry to group the items into small sets (e.g. Sets of items that are purchased together) and determine the optimal locations to promote the products. It can be applied not only to items purchased concurrently but also to items purchased sequentially. Association rules are frequently used by retail industry to assist in marketing and sales promotion, shelf management/item bundling (identifying items that are bought together by sufficiently many customers by processing the point-of-sale data collected with barcode scanners to find dependencies among items), Inventory management, advertising, discount/promotion decisions, etc. Retailers organized its merchandise based on buying patterns and information about associations between products. One illustrative example has been explained with the help of a data mining tool in the retail sales data (live data).The model with an illustration in this research paper has been discussed. The research in his paper models for incorporating the Data Mining based Association Rules in Indian Retail Industry.

INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE AND MANAGEMENT

CONTENTS

Sr. No.	Article / Paper	Page No.
1.	**STRATEGIC MARKETING PRACTICES ON THE PERFORMANCE OF FIRMS IN NIGERIAN OIL AND GAS INDUSTRY** *DR. S. T. AKINYELE*	6
2.	**HUMAN RESOURCE SYSTEMS AND ORGANIZATIONAL EFFECTIVENESS: THE CASE OF INDIAN RURAL BANKING** *PROF. NEELU ROHMETRA & DR. JAYA BHASIN*	33
3.	**A COMPARATIVE STUDY ON THE PRICE MOVEMENTS BETWEEN GOLD AND CRUDE OIL BETWEEN 2006 AND 2007** *PROF. (DR.) A. OLIVER BRIGHT & KARTHIK*	55
4.	**RELATIONSHIP BETWEEN FII, SENSEX AND MARKET CAPITALISATION** *GAYATHRI DEVI. R & PROF. (DR.) MALABIKA DEO*	97
5.	**A NOVEL INDEPENDENT COMPONENT ANALYSIS APPROACH FOR BANKRUPTCY PREDICTION USING NEURO-FUZZY NETWORKS** *NIDHI ARORA & PROF. (DR.) SANJAY K. VIJ*	104
6.	**CHALLENGES FOR IFRS IMPLIMENTATIONS IN INDIA - AN ACCOUNTING REVOLUTION** *PROF. (DR.) ATUL BANSAL & DR. SHWETA BANSAL*	113
7.	**EMPLOYEE INVOLVEMENT – A TOOL FOR ORGANIZATIONAL EXCELLENCE** *DR. SMITHA SAMBRANI*	128
8.	**PREFERENTIAL TRADING AGREEMENTS: THE CASE FOR ASEAN+4 AS A POTENTIAL TRADE BLOC** *DR. VIRENDER PAL, NARESH KUMAR & BALJJIT SINGH*	136
9.	**A STUDY OF LIQUIDITY, PROFITABILITY AND RISK ANALYSIS OF CEMENT INDUSTRY IN INDIA** *MS. RAJNI SOFAT*	142
10.	**BASE RATE: THE NEW BENCHMARK RATE** *PROF. REKKHA DHIAYA, PROF. HARPREET SINGH & PROF. ANMOL SOI*	162
11.	**A STUDY OF FACTORS AFFECTING TRAINING DECISIONS OF EMPLOYEES IN SERVICE INDUSTRY: A STUDY WITH REFERENCE TO SELECTED SERVICE INDUSTRY IN NCR** *VIJIT CHATURVEDI*	171
12.	**DATA MINING BASED ASSOCIATION RULES & RFM ANALYSIS IN INDIAN RETAIL SECTOR: AN EMPIRICAL INVESTIGATION** *Dr. ANSHUL SHARMA, Prof. (Dr.) M. K. KULSHRESHTHA & Prof. (Dr.) ASHOK AGRWAL*	186
13.	**FACTORS AFFECTING INDIA'S BALANCE OF PAYMENT (BOP) AFTER LIBERALIZATION (1991)** *DEBASISH MAULIK*	204
14.	**INCOME INEQUALITY AND PROGRESSIVE INCOME TAXATION IN CHINA AND INDIA** *DR. SUNIL GUPTA, DR. VIJITA AGGARWAL & DR. ALKA MITTAL*	215
15.	**CHALLENGES FACED BY WOMEN ENTREPRENEURS IN A DEVELOPING ECONOMY** *DR. SHEFALI VERMA THAKRAL*	221
16.	**MARKET VALUE ADDED: A STUDY IN THE SELECT INDIAN SOFTWARE COMPANIES** *DR. D. KAMALAVENI & DR. S. KALAISELVI*	227
	REQUEST FOR FEEDBACK	245

CALL FOR ARTICLES/RESEARCH PAPERS

We invite original research papers in the area of finance, marketing, HRM, Banking, Insurance, and other allied subjects. The above mentioned tracks are only indicative, and not exhaustive. The journal expects unpublished and original quality research articles/papers only.

You may submit your articles/papers at the email addresses, info@ijrcm.org.in or infoijrcm@gmail.com.

GUIDELINES FOR SUBMISSION OF ARTICLE/PAPER

1. COVERING LETTER FOR SUBMISSION:

The Editor
IJRCM

Subject: **Submission of Manuscript**.

Dear Sir,

Find my submission of research paper/article for possible publications in your e-journal.

I hereby affirm that the content of this manuscript are original. Furthermore it has been neither published elsewhere fully or partially or any language nor submitted for publication (fully or partially) elsewhere simultaneously.

I affirm that the all author (s) have seen and agreed to the submitted version of the paper and their inclusion of name(s) as co-author(s).

Also, if our research paper/article accepted, I/We agree to comply with the formalities as given on the website of journal.

Name of Corresponding Author (s)

Affiliation:

Mailing address:

Mobile Number (s):

Landline Number (s):

E-mail Address (s):

2. **INTRODUCTION:** Manuscript must be in English prepared on a standard A4 size paper setting. It must be prepared on a double space and single column with 1" margin set for top, bottom, left and right. It should be typed in 12 point-Times New Roman Font with page numbers at the bottom and centre of the every page.
3. **MANUSCRIPT TITLE:** The title of the paper should be in a 14 point Times New Roman Font. It should be bold typed, centered and fully capitalised.
4. **AUTHOR NAME(S) & AFFILIATIONS:** The author(s) full name, affiliation(s), mobile/landline numbers, and email/alternate email address should be in 10-point Times New Roman. It must be centered underneath the title.
5. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 300 words. The abstract must be informative and explain background, aims, methods, results and conclusion.
6. **KEYWORDS:** Abstract must be followed by list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stop at the end.
7. **HEADINGS:** All the headings and sub-headings should be in a 12 point-Times New Roman Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.
8. **MAIN TEXT:** The main text should be in a 12 point-Times New Roman Font, single spaced, fully justified.
9. **FIGURES &TABLES:** These must be simple, centered & numbered, and table titles must be above the tables. Sources of data should be mentioned below the table.

10. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to Harvard Style of Referencing. The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow the references as per following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order…
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- Note that italics are used only for titles of books and names of journals. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- Use endnotes rather than footnotes.
- The location of endnotes within the text should be indicated by superscript numbers.
- For sources which have insufficient details to be included in the Reference, use endnotes (such as interviews, some media sources, some Internet sources).

See the following for style and punctuation in References:

Books

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio," Ohio State University.

Contributions to books

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

Journal and other articles

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.
- Kiran Ravi, Kaur Manpreet (2008), Global Competitiveness and Total Factor Productivity in Indian Manufacturing, International Journal of Indian Culture and Business Management, Vol. 1, No.4 pp. 434-449.

Conference papers

- Chandel K.S. (2009): "Ethics in Commerce Education." Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

Unpublished dissertations and theses

- Kumar S. (2006): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University.

Online resources

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

Website

- Kelkar V. (2009): Towards a New Natutal Gas Policy, Economic and Political Weekly, Viewed on 11 September 2009 http://epw.in/epw/user/viewabstract.jsp

Retail industry is adopting the use of loyalty cards. Rewarding customers who are frequent buyers encourage them to do even more of their shopping at that retail store and make them less likely to buy from the competitor. RFM analysis uses three metrics viz. Recency, Frequency and monetary to evaluate the customer behaviour and customer value. The author has done the RFM analysis on the live customer's historical large dataset of a retail store. The paper demonstrates the RFM analysis with empirical evidence.

KEYWORDS

Association Rules, RFM, Data Mining; DM, Customers, Retail sector.

1. INTRODUCTION

In the recent years the significant changes are done in the retail industry which has important implications on DM. Retail industry is using information technology (IT) for generating, storing and analyzing mass produced data not only for operational purposes but also for enabling strategic decision making to survive in a competitive and dynamic environment. DM helps in reducing information overload along with the improved decision-making by searching for relationships and patterns from the huge dataset collected by organizations. It enables a retail industry to focus on the most important information in the database and allows retailers to make more knowledgeable decisions by predicting future trends and behaviours.

The extraction of association rules from large databases has proven beneficial for companies since such rules can be very effective in revealing actionable knowledge that leads to strategic decisions. With competition for shelf space intensifying, there is a pressing need to provide shoppers with a highly differentiated value proposition through 'right product mix in right time amount at right time' (Bala, 2008). Mining or extracting association rules from structured and unstructured data collected from customer transaction database will be tremendous importance for taking strategic decisions in retail industry. Since an organization have thousands of items, only few of them deserve management's closest attention. Knowledge of purchase pattern will be an important input for developing association rules. Association rule correlates one set of items with another set of items. Association rules generated are derived from the patterns in a particular dataset. The discovery of such association rule can help retailers to develop marketing strategies by gaining insight into, which items are frequently purchased together by customer. Data mining is used to find new, hidden or unexpected patterns from large volume of historical data, typically stored in a data warehouse (Bala, 2008). Knowledge discovered using data mining helps in more effective strategic decision making.

The arrival of retail boom caused the global technology vendors to quickly get into the marketplace with solutions that claim to make retailers' lives simpler. Retailers have to put in great efforts to really know their customers. Retail industry emphasized on quick delivery of customer focused services (offers, promos, etc) since adapting to customer needs in a very limited period of time is also very important. Retailers continuously get the advantage from information collected from customers' transactions. Hence requirements of retail, technology wise would encompass business intelligence, data mining/warehousing, and other similar technologies since using these, retailers can constantly benefit from newly observed trends based on user purchases (Sohoni, 2007). The changing consumption patterns trigger changes in shopping styles of consumers and also the factors that drive people into stores (Kaur and Singh, 2007). Hou and Tu (2008) addressed that the managers in the contemporary marketing must importantly identify potential customer relationships to positively affect corporate performance. Ranjan and Bhatnagar (2008) opinioned that the optimization of revenue can be accomplished by a

better understanding of customers, based on their purchasing patterns and demographics, and better information empowerment at all customers touch points, whether with employees or other media interfaces. With the retail boom, companies are likely to deploy IT tools that help them enhance the end-customer's experience. Jones and Ranchhod (2007) expressed that the strategic focus is required on the real complexity of the relationship that organizations are initially able to establish with customers. Sangle and Verma (2008) opinioned that the customer relationship management unites the potential of marketing strategies and IT to create profitable, long-term relationships with customers and helps in enhancing the opportunities to use data and information to both understand customers and co-create value with them.

2. LITERATURE REVIEW

Dennis (2001) explored the customer knowledge management framework for shopping centers using Data mining. Terrovitis and Vassilliadis (2003) present the architecture of a pattern base management system that can be used to efficiently store, and query patterns. Van der Aalst (2003) introduces the concept of workflow mining and presents a common format for work flow logs. van der Aalst (2003) introduce the concept of workflow mining and present a common format for work flow logs. List and Machaczek (2004) illustrates how a data warehouse can be used to facilitate a Corporate Performance Measurement System by the integration of business process performance information into a traditional data warehouse that generally represents only the functional organization. Pan (2005) find out the state of IT adoption and factors that affect IT adoption in Chinese retail companies. Wong et al. (2005) developed a method to select inventory items from association rules which gives a methodology to choose a subset of items which can give maximal profit with the consideration of cross- selling effect. Marketos and Theodoridis(2006) proposed a framework for measuring the performance in the retail industry. Tvrdikova(2007) discussed the issues of business intelligence applications to support decision making. Bala(2008) models for incorporating purchase dependencies in retail multi-item inventory management. Bhanu and Balasubramanie (2008) explored the predictive modeling of inter-transaction association rules from a business perspective. Srivastava(2008) build a picture of the changes in retail taking place in India. Cross-selling is the strategy of pushing new products to current customers based on their past purchases.

3. RESEARCH METHODOLOGY

The customer transaction data is very valuable asset for any company hence the need for research design was felt. So, the data for this paper was collected in two phase. First the primary data is collected through various sources which include personal interviews, surveys and filled questionnaire, review the available online software packages, attending conferences and seminars, etc. Secondary data is collected through studying the literature related to research that is available in various journals, books, magazine, websites, established doctoral thesis, etc. The authors got the customer transaction database of one retail firm (name masked) which is analyzed with the help of data mining tool Statisca and SPSS' Clementine. The basic objective is to study the advantages of association rules using DM in Indian retail industry with the help of an empirical study.

4. INDIAN RETAIL INDUSTRY

The increased globalization, market saturation, and increased competitiveness give rise to mergers and acquisitions. Indian retailers are seeking competitive advantages by better improving relationships with

customers which has taken on new life. Rogers (2005) addressed that the companies recognize that customer relationships are the underlying tool for building customer value, and they are finally realizing that growing customer value is the key to increasing enterprise value.

The retail sector is growing rapidly in the Indian scenario as well as globally. With the Indian retail sector booming, it brings immense opportunities for foreign as well as domestic players. The changing lifestyle of the Indian consumer makes it essential for the retailers to understand the patterns of consumption. The changing consumption patterns trigger changes in the shopping styles of consumers and also in the factors that drive people into stores (Kaur and Singh, 2007). The Indian retail has been transformed due to the attitudinal shift of the Indian consumer in terms of choice preference, value for money and the emergence of organized retail formats. Rising incomes, increased advertising, and a jump in the number of women working in the country's urban centers have made goods more attainable and enticing to a larger portion of the population. At the same time, trade liberalization and more sophisticated manufacturing techniques create goods that are less expensive and higher quality (Hanna, 2004). Pande and Collins (2007) explored to centralize the retail supply chain in India with the goal to improve overall retail business in India.

Vector (2007) explored that the Retail is India's largest industry with the market size of around US $312 billion in which organized retailing comprises only 2.8 per cent of the total retailing market and is estimated at around US$ 8.7 billion. The organized retail sector is expected to grow to US $ 70 billion by 2010. FICCI Retail Report (20007) reported that the estimates predict that the overall size of the retail sector in India is expected to touch US$427 billion by 2010 and US$637 billion by 2015 with the modern segment expected to account for 22 per cent by 2010, up from the present four per cent.

5. DATA MINING

Data Mining is a process of analyzing the data from different perspectives and presenting it in a summarized way into useful information. It extracts patterns and trends that are hidden among the data. It is often viewed as a process of extracting valid, previously unknown, non-trivial and useful information from large databases (Rao, 2003). Han and Kamber (2007) expressed that the DM is extracting or mining knowledge from large amount of data. Feelders et al. (2000) opinioned that the DM is the process of extracting information from large data sets through the use of algorithms and techniques drawn from the field of statistics, machine learning and database management systems. Noonan (2000) explained that DM is a process for sifting through lots of data to find information useful for decision making. It helps in predicting the future of the business. It can make the improvement in every industry throughout the world. The data can be mined and the results can be used to determine not only what the customers wants, but to also predict what they will do. West (2005) addressed that by relying on the power of data mining, retailers can maintain the consistency and accuracy of their underwriting decisions; they can significantly reduce the impact of fraudulent claims; and can have a better understanding of their customer's wants and needs. It can be used to control costs as well as contribute to revenue increases (Two Crows Corporation, 2005).

The DM software uses the business data as raw material using a predefined algorithm to search through the vast quantities of raw data, and group the data according to the desired criteria that can be useful for the future target marketing (Ahmed, 2004). DM involves the use of predictive modeling, forecasting and descriptive modeling techniques. By using these techniques, a retail firm can proactively manage customer retention, identify cross-sell and up-sell opportunities, profile and segment customers, set optimal pricing policies, and objectively measure and rank which suppliers are best suited for their needs (Bhasin, 2006). DM applications automate the process of searching the huge amount of data to find patterns that are good predictors of purchasing behaviors. After mining the data, marketers

must feed the results into campaign management software that manages the campaign directed at the defined market segments (Thearling, 2007).

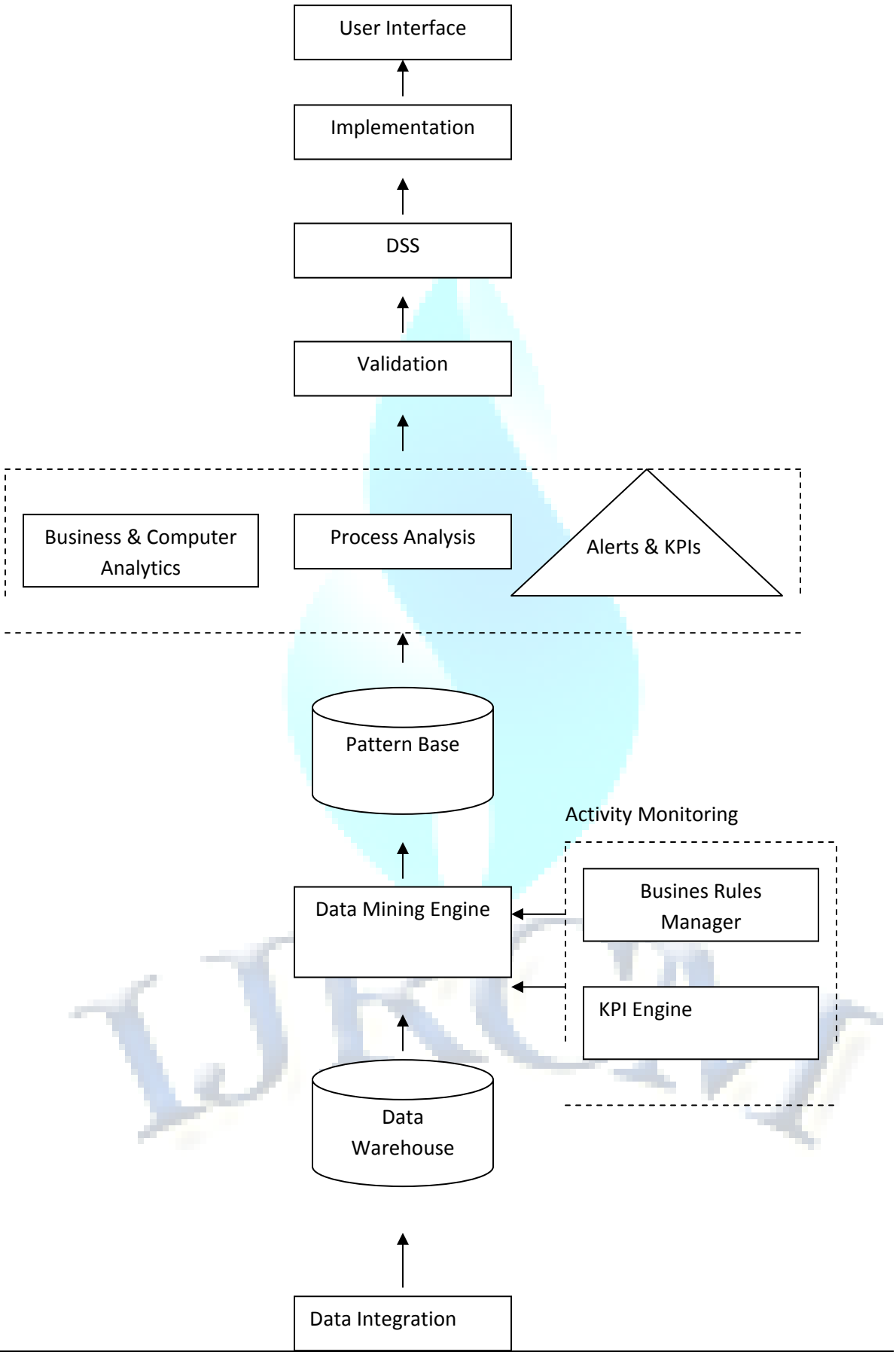
Wang and Wang (2007) pointed out that the DM techniques for the online customer segmentation helps in clustering the customers on the basis of the characteristic that they show while purchasing the product online or surfing the net. Chen, Wu and Chen (2005) effectively discovered the current spending pattern of customers and trends of behavioral change by using DM tools, which would allow management to detect in a large database potential change of customer preference, and provide products and services faster as desired by the customers to expand the client base and prevent customer attrition. Pan et al. (2007) found that the problem of classification of the customer is cost sensitive in nature. Consumer-focused companies with sizable caches of information on current and potential customers such as retailers are ideal for data mining technology (Cowley, 2005).

Chen and Liu (2005) focused on enhancing the functionality of current applications of DM. Berry and Linoff (2001) expressed that only through the application of DM techniques can a large enterprise hope to turn the myriad records in its customer databases into some sort of coherent picture of its customers. It can also be used to locate individual customers with specific interests or determine the interests of a specific group of customers (Guzman, 2002). Berman and Evans (2008) opinioned that DM is used by retail executives and other employees-and sometimes channel partners to analyze information by customer type, product category, and so forth in order to determine opportunities for tailored marketing efforts that would lead to better retailer performance.

6. MODEL DEVELOPMENT

The model proposed has been depicted in the following figure1 for measuring and managing the performance in the retail sector. The source data module containing customer database, transactional database and other databases can be used as the input to the system for mining association rules (patterns).The model can integrate data from heterogeneous sources and data warehouse can be formed. Activity monitoring module is concerned with the real time information. It controls and updates KPIs and verifies that corporate rules are satisfied by triggering Business Rules Manager. KPIs can also be verified for satisfying predefined Business Rules. Violation of business rules can trigger alerts in the enterprise portal. Data mining engine have a set of techniques and algorithms for identifying patterns on the data warehouse. Sequence of purchase, routes of purchase, identifying churners and non-churners, correlations between products, prediction of the product demand, customer segmentation, identifying the switching behaviour of the customers are few of the tasks that can be applied on the shopping transactional data. Patterns are extracted from the heterogeneous data sources by applying the data mining techniques contained in the data mining engine. Frequent item sets, association rules extracted from warehouse data are typical example of patterns. Enterprise portal containing the business and customer analytics, Process analytics and alerts and KPIs is the proposed output which is validated and then used in Decision Support System (DSS). Data mining facilitated by data warehousing, addresses the need of strategic decision making. Knowledge discovered from data mining can be used in DSS and can be used by the users in various positions after implementation.

One live example in the following section has been illustrated where association rules have been mined using a data mining tool named Statisca Data Miner. The proposed model is as follows:


User Interface
Implementation
DSS
Validation
Business & Computer Analytics
Process Analysis
Alerts & KPIs
Pattern Base
Activity Monitoring
Busines Rules Manager
Data Mining Engine
KPI Engine
Data Warehouse
Data Integration

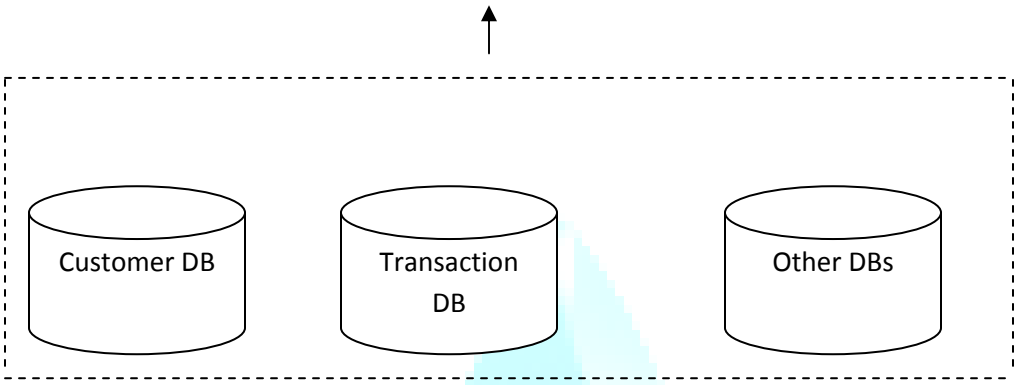


Figure1: Proposed Model

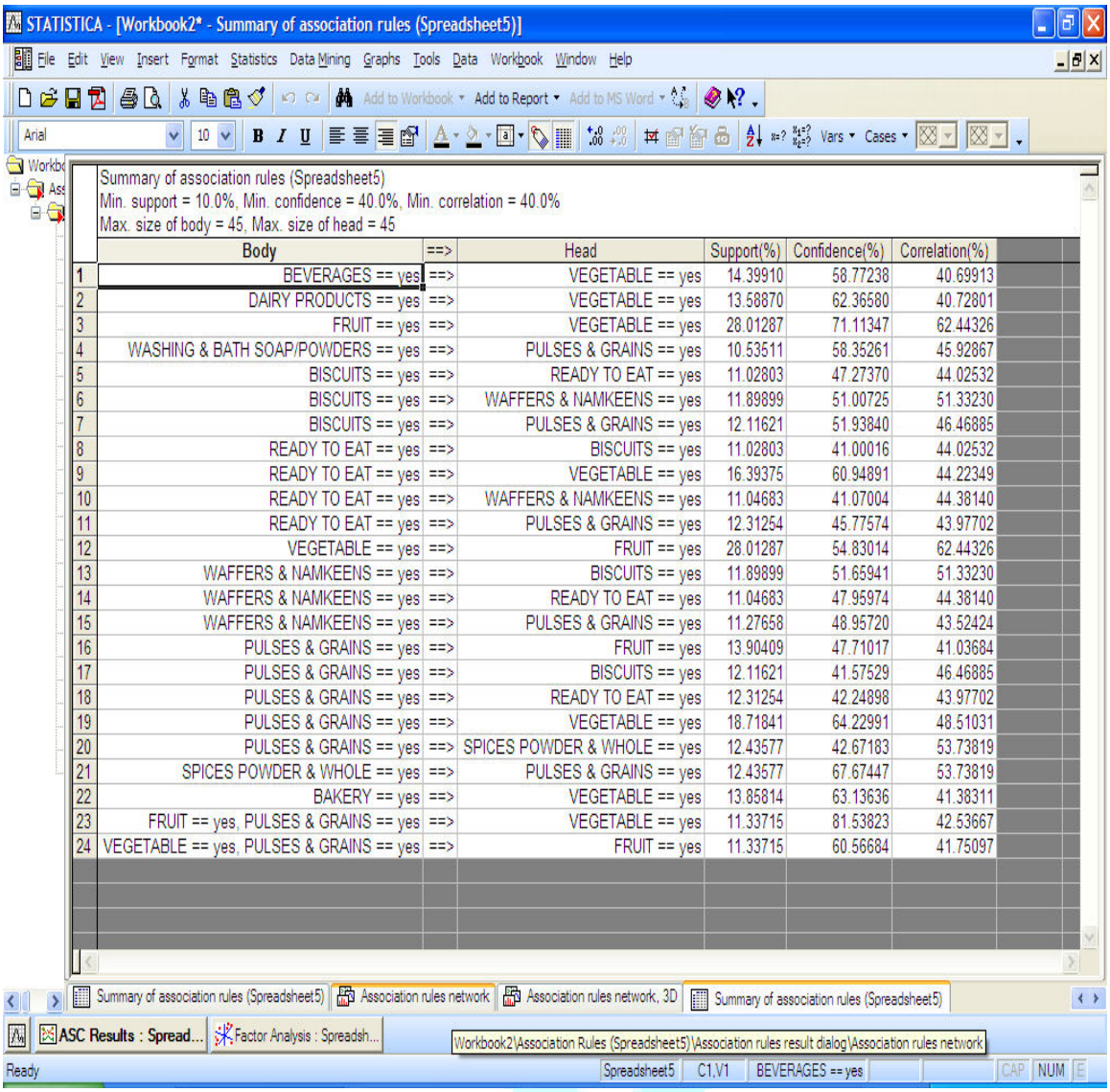
7. EMPIRICAL INVESTIGATION USING DATA MINING

The data collected from the retail firm (name masked) is analyzed with the help of a data mining tool.

7.1 Association Rules: Empirical investigation

The analysis was performed on the live large dataset of customer transactions of a retail outlet using a data mining tool Statisca Data Miner.

Association rules are generated of the general form *if Body then Head*, where *Body* and *Head* stand for single codes or text values (items) or conjunctions of codes or text values. The major statistics computed for the association rules are *Support* (relative frequency of the *Body* or *Head* of the rule), *Confidence* (conditional probability of the *Head* given the *Body* of the rule), and *Correlation* (support for *Body* and *Head*, divided by the square root of the product of the support for the *Body* and the support for the *Head*). These statistics can be summarized in a spreadsheet, as shown below:



	Body	==>	Head	Support(%)	Confidence(%)	Correlation(%)
1	BEVERAGES == yes	==>	VEGETABLE == yes	14.39910	58.77238	40.69913
2	DAIRY PRODUCTS == yes	==>	VEGETABLE == yes	13.58870	62.36580	40.72801
3	FRUIT == yes	==>	VEGETABLE == yes	28.01287	71.11347	62.44326
4	WASHING & BATH SOAP/POWDERS == yes	==>	PULSES & GRAINS == yes	10.53511	58.35261	45.92867
5	BISCUITS == yes	==>	READY TO EAT == yes	11.02803	47.27370	44.02532
6	BISCUITS == yes	==>	WAFFERS & NAMKEENS == yes	11.89899	51.00725	51.33230
7	BISCUITS == yes	==>	PULSES & GRAINS == yes	12.11621	51.93840	46.46885
8	READY TO EAT == yes	==>	BISCUITS == yes	11.02803	41.00016	44.02532
9	READY TO EAT == yes	==>	VEGETABLE == yes	16.39375	60.94891	44.22349
10	READY TO EAT == yes	==>	WAFFERS & NAMKEENS == yes	11.04683	41.07004	44.38140
11	READY TO EAT == yes	==>	PULSES & GRAINS == yes	12.31254	45.77574	43.97702
12	VEGETABLE == yes	==>	FRUIT == yes	28.01287	54.83014	62.44326
13	WAFFERS & NAMKEENS == yes	==>	BISCUITS == yes	11.89899	51.65941	51.33230
14	WAFFERS & NAMKEENS == yes	==>	READY TO EAT == yes	11.04683	47.95974	44.38140
15	WAFFERS & NAMKEENS == yes	==>	PULSES & GRAINS == yes	11.27658	48.95720	43.52424
16	PULSES & GRAINS == yes	==>	FRUIT == yes	13.90409	47.71017	41.03684
17	PULSES & GRAINS == yes	==>	BISCUITS == yes	12.11621	41.57529	46.46885
18	PULSES & GRAINS == yes	==>	READY TO EAT == yes	12.31254	42.24898	43.97702
19	PULSES & GRAINS == yes	==>	VEGETABLE == yes	18.71841	64.22991	48.51031
20	PULSES & GRAINS == yes	==>	SPICES POWDER & WHOLE == yes	12.43577	42.67183	53.73819
21	SPICES POWDER & WHOLE == yes	==>	PULSES & GRAINS == yes	12.43577	67.67447	53.73819
22	BAKERY == yes	==>	VEGETABLE == yes	13.85814	63.13636	41.38311
23	FRUIT == yes, PULSES & GRAINS == yes	==>	VEGETABLE == yes	11.33715	81.53823	42.53667
24	VEGETABLE == yes, PULSES & GRAINS == yes	==>	FRUIT == yes	11.33715	60.56684	41.75097

Figure2: Tabular Representation of Association Rules

This results spreadsheet shows an example of how association rules can be applied to mining tasks. The values for support, confidence, and correlation are expressed in percent.

From the above spreadsheet, the association rule *If (Fruit=yes AND Pulses & Grains=yes) then vegetable=yes* have the maximum confidence and the rule *If Fruit=yes then vegetable*=yes have the maximum support value and correlation value.

Though above rules is a very common combination, few unexpected rules are also extracted with a good confidence value, some of them are as follows:

If Bakery=yes then vegetables=yes,

If Dairy Products=yes then vegetables=yes,

If ready to eat=yes then vegetables=yes,

If Beverages=yes then vegetables=yes

These rules can be reviewed in the graphical format depicted as follows:

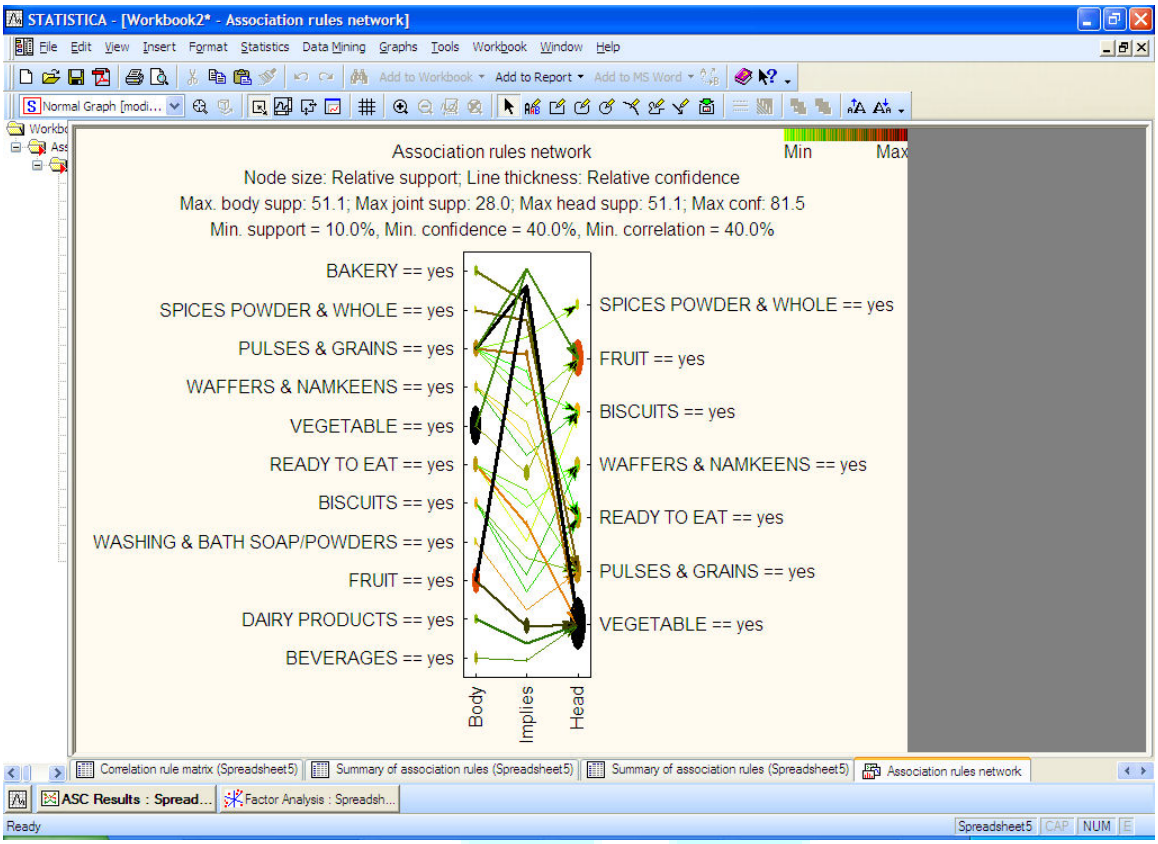


Figure3: Association Rules Network, 2D

In this graph, the support values for the *Body* and *Head* portions of each association rule are indicated by the sizes and colours of each. The thickness of each line indicates the confidence value (conditional probability of Head given Body) for the respective association rule; the sizes and colours of the circles in the centre, above the *Implies* label, indicate the joint support (for the co-occurrences) of the respective *Body* and *Head* components of the respective association rules.

Hence, in this graphical summary, the strongest support value was found for vegetable=yes which was associated with FRUITS=yes, BEVERAGES=yes, DAIRY PRODUCTS=yes, PULSES AND GRAINS=yes. The absolute frequencies with which individual codes or text values (items) occur in the data are often not reflected in the association rules; instead, only those codes or text values (items) are retained that show sufficient values for support, confidence, and correlation, i.e., that co-occur with other codes or text values (items) with sufficient relative(co-)frequency

If we increase the minimum correlation value from 40% to 50%, we get the clearer picture as shown below:

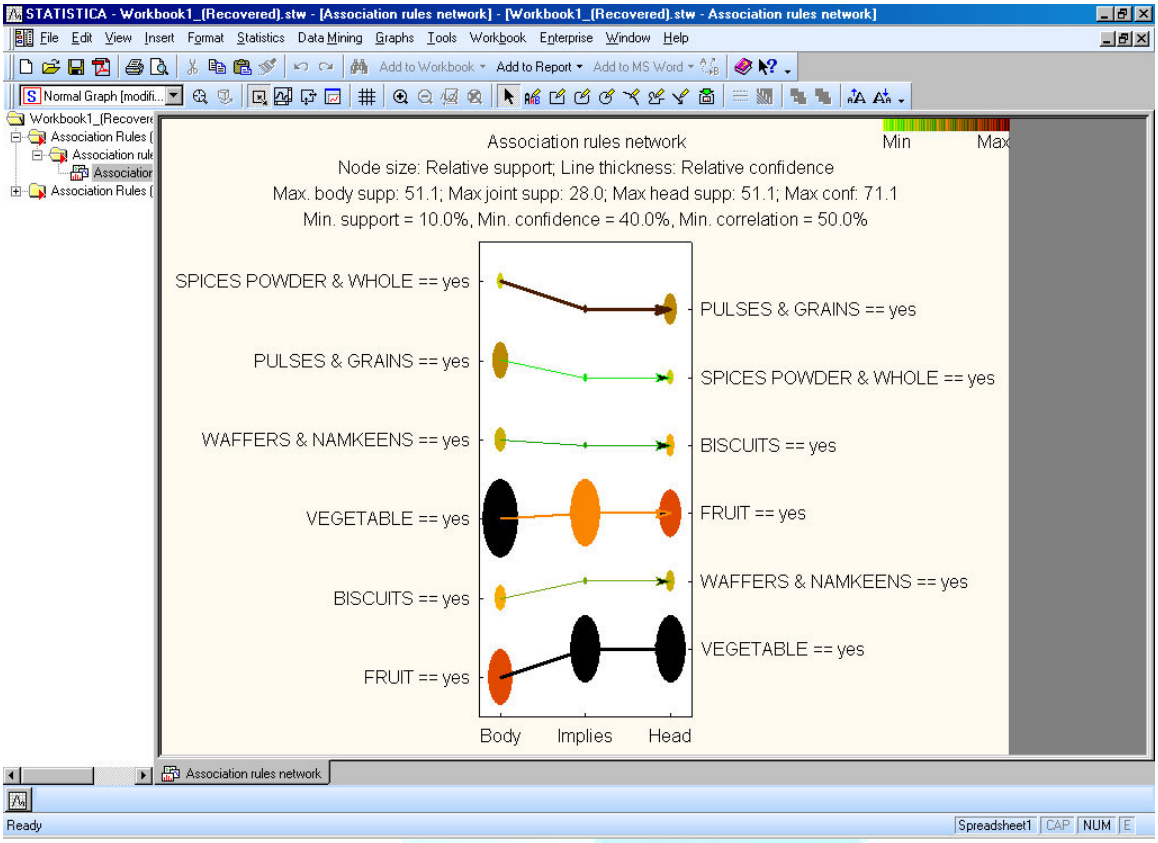


Figure4: Association rules Network

. The association rules derived can be graphically summarized in 3D association network display as well as shown below.

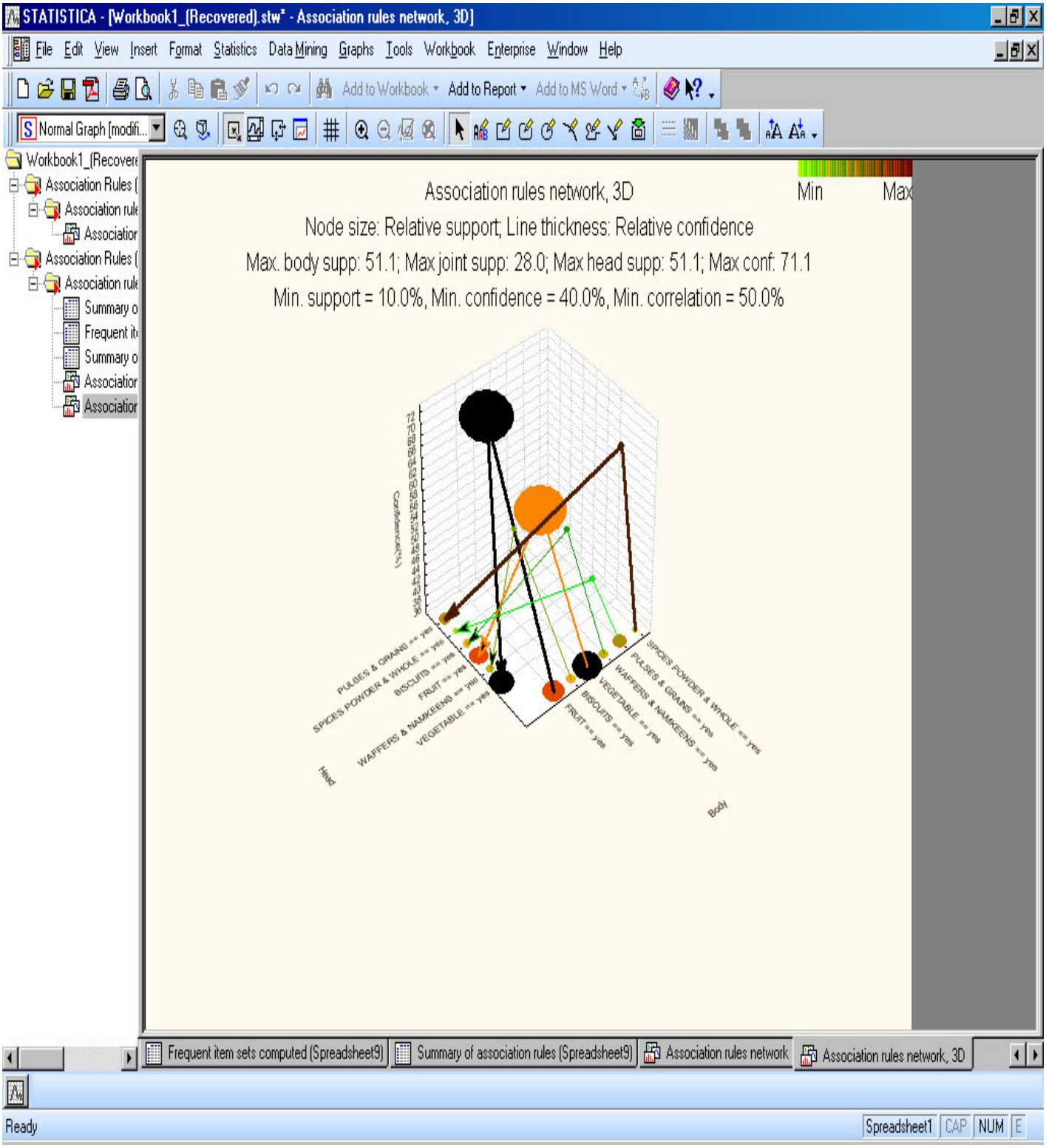
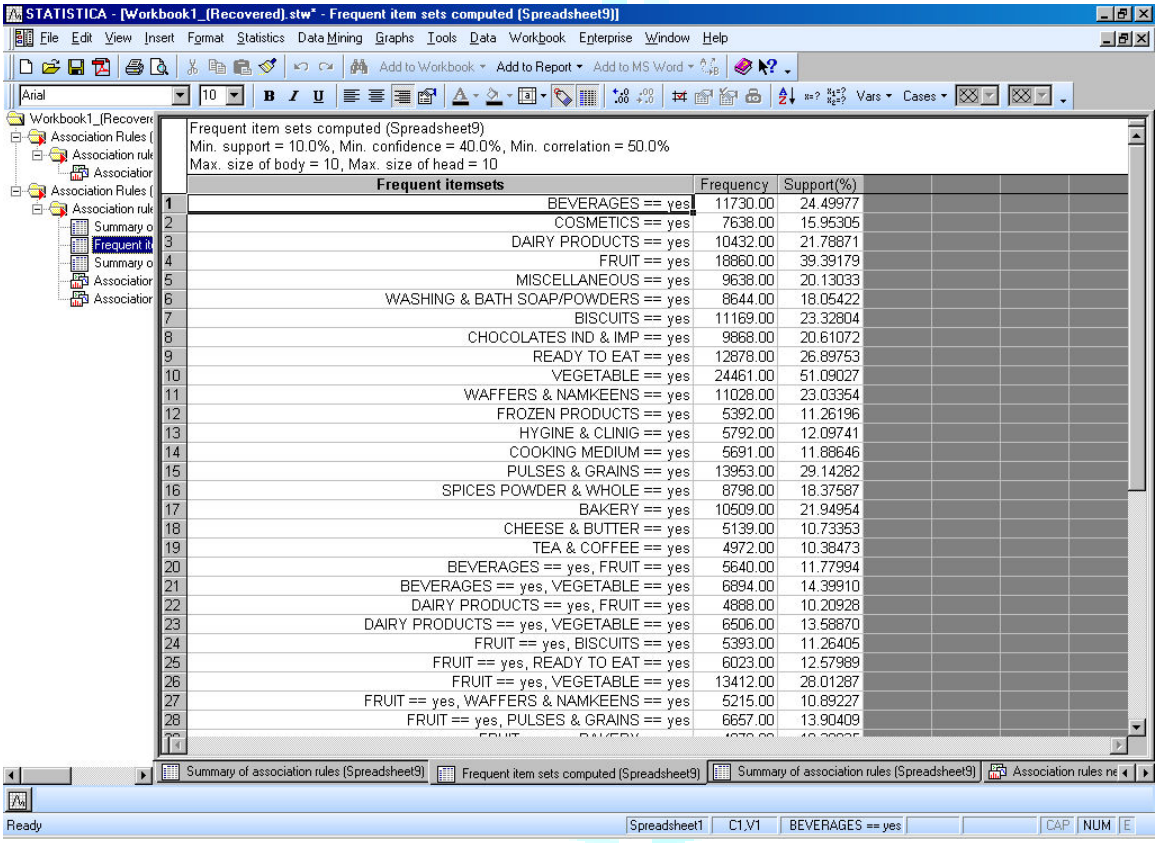


Figure5: Association Rules Network, 3D

As in the 2D Association Network, the support values for the *Body* and *Head* portions of each association rule are indicated by the sizes and colors of each circle in the 2D. The thickness of each line indicates the confidence value (joint probability) for the respective association rule; the sizes and colors of the "floating" circles plotted against the (vertical) z-axis indicate the joint support (for the co-occurences) of the respective *Body* and *Head* components of the association rules. The plot position of each circle along the vertical z - axis indicates the respective confidence value. Hence, this particular graphical summary clearly shows two simple rules: Customer who purchases vegetables also purchases fruits, and vice versa. Customer who purchases pulses and grains also purchases spices powder and whole, and vice versa.



	Frequent itemsets	Frequency	Support(%)
1	BEVERAGES == yes	11730.00	24.49977
2	COSMETICS == yes	7638.00	15.95305
3	DAIRY PRODUCTS == yes	10432.00	21.78871
4	FRUIT == yes	18860.00	39.39179
5	MISCELLANEOUS == yes	9638.00	20.13033
6	WASHING & BATH SOAP/POWDERS == yes	8644.00	18.05422
7	BISCUITS == yes	11169.00	23.32804
8	CHOCOLATES IND & IMP == yes	9868.00	20.61072
9	READY TO EAT == yes	12878.00	26.89753
10	VEGETABLE == yes	24461.00	51.09027
11	WAFFERS & NAMKEENS == yes	11028.00	23.03354
12	FROZEN PRODUCTS == yes	5392.00	11.26196
13	HYGINE & CLINIG == yes	5792.00	12.09741
14	COOKING MEDIUM == yes	5691.00	11.88646
15	PULSES & GRAINS == yes	13953.00	29.14282
16	SPICES POWDER & WHOLE == yes	8798.00	18.37587
17	BAKERY == yes	10509.00	21.94954
18	CHEESE & BUTTER == yes	5139.00	10.73353
19	TEA & COFFEE == yes	4972.00	10.38473
20	BEVERAGES == yes, FRUIT == yes	5640.00	11.77994
21	BEVERAGES == yes, VEGETABLE == yes	6894.00	14.39910
22	DAIRY PRODUCTS == yes, FRUIT == yes	4888.00	10.20928
23	DAIRY PRODUCTS == yes, VEGETABLE == yes	6506.00	13.58870
24	FRUIT == yes, BISCUITS == yes	5393.00	11.26405
25	FRUIT == yes, READY TO EAT == yes	6023.00	12.57989
26	FRUIT == yes, VEGETABLE == yes	13412.00	28.01287
27	FRUIT == yes, WAFFERS & NAMKEENS == yes	5215.00	10.89227
28	FRUIT == yes, PULSES & GRAINS == yes	6657.00	13.90409

Figure 6: Frequent Item sets

The spreadsheet above demonstrates the frequent item set with the support values. From the table it is clear that vegetables has the maximum frequencies followed by fruits, pulses & grains, biscuits, beverages, etc.

7.2. RFM ANALYSIS

RFM stands for recency, frequency, and monetary value. This technique uses these three metrics to evaluate customer behavior and customer value (Kumar, 2006). *Recency* is a measure of the time lag since the customer has purchased last from your business or how recently a customer has purchased. Recency can be measured in weeks, months, quarters, fiscal years, etc. *Frequency* is the quantity or volume of items or services purchased in a certain defined period. It can be single units or perhaps aggregated in deciles or whatever meaningful grouping. *Monetary value* is a numeric currency figure representing the value of each of the frequency units or aggregated units that were purchased.

Aggelis and Christodoulakis (2005) addressed that the RFM analysis is a method to identify high-response customers in marketing promotions, and to improve overall response rates, Smith (2006) expressed that the analysis of customer data is the key component for the success of CRM.

The RFM analysis is depicted in the following figure:

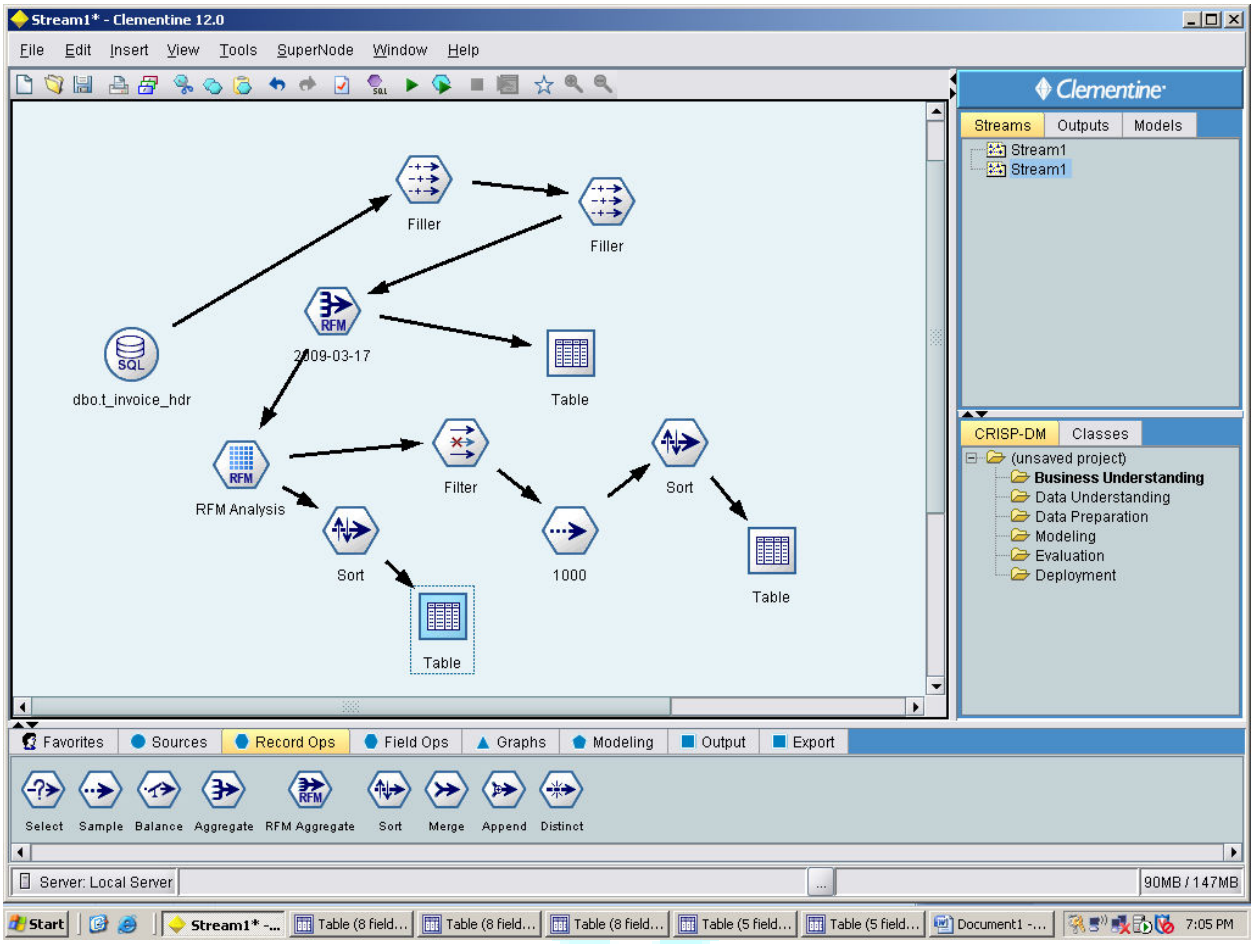


Figure 7: RFM Analysis using SPSS Clementine tool

RFM Aggregate enables to take a customer's historical transactional data and combine that data into a single row that lists when they last dealt with the retailer,how many transactions they have made and the total monetary value of those transactions.RFM Analysis then enables the retailer to further analyse this prepared data.The results of the RFM Analysis is shown in the following figure:

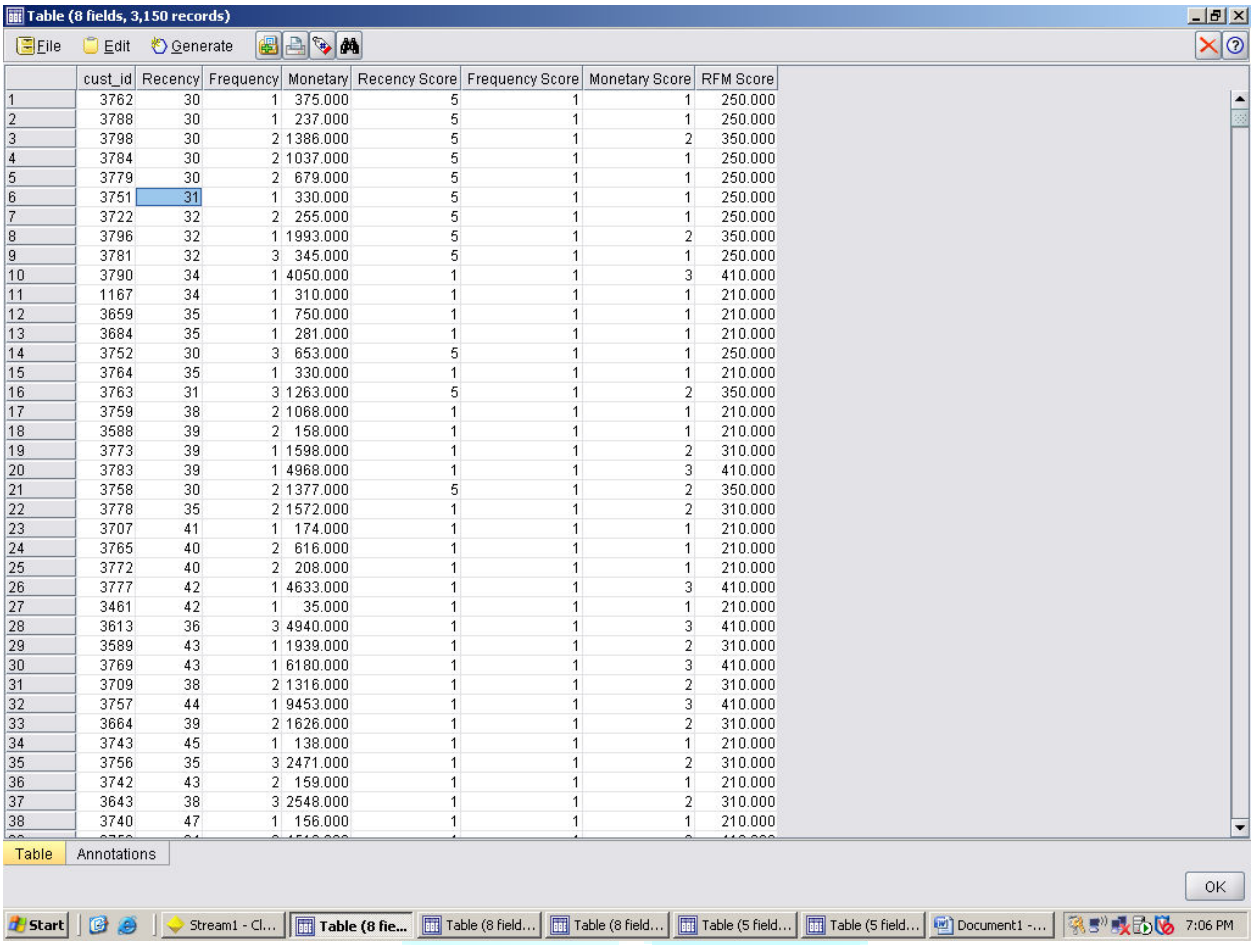


Table (8 fields, 3,150 records)

File Edit Generate

	cust_id	Recency	Frequency	Monetary	Recency Score	Frequency Score	Monetary Score	RFM Score
1	3762	30	1	375.000	5	1	1	250.000
2	3788	30	1	237.000	5	1	1	250.000
3	3798	30	2	1386.000	5	1	2	350.000
4	3784	30	2	1037.000	5	1	1	250.000
5	3779	30	2	679.000	5	1	1	250.000
6	3751	31	1	330.000	5	1	1	250.000
7	3722	32	2	255.000	5	1	1	250.000
8	3796	32	1	1993.000	5	1	2	350.000
9	3781	32	3	345.000	5	1	1	250.000
10	3790	34	1	4050.000	1	1	3	410.000
11	1167	34	1	310.000	1	1	1	210.000
12	3659	35	1	750.000	1	1	1	210.000
13	3684	35	1	281.000	1	1	1	210.000
14	3752	30	3	653.000	5	1	1	250.000
15	3764	35	1	330.000	1	1	1	210.000
16	3763	31	3	1263.000	5	1	2	350.000
17	3759	38	2	1068.000	1	1	1	210.000
18	3588	39	2	158.000	1	1	1	210.000
19	3773	39	1	1598.000	1	1	2	310.000
20	3783	39	1	4968.000	1	1	3	410.000
21	3758	30	2	1377.000	5	1	2	350.000
22	3778	35	2	1572.000	1	1	2	310.000
23	3707	41	1	174.000	1	1	1	210.000
24	3765	40	2	616.000	1	1	1	210.000
25	3772	40	2	208.000	1	1	1	210.000
26	3777	42	1	4633.000	1	1	3	410.000
27	3461	42	1	35.000	1	1	1	210.000
28	3613	36	3	4940.000	1	1	3	410.000
29	3589	43	1	1939.000	1	1	2	310.000
30	3769	43	1	6180.000	1	1	3	410.000
31	3709	38	2	1316.000	1	1	2	310.000
32	3757	44	1	9453.000	1	1	3	410.000
33	3664	39	2	1626.000	1	1	2	310.000
34	3743	45	1	138.000	1	1	1	210.000
35	3756	35	3	2471.000	1	1	2	310.000
36	3742	43	2	159.000	1	1	1	210.000
37	3643	38	3	2548.000	1	1	2	310.000
38	3740	47	1	156.000	1	1	1	210.000

Table Annotations

OK

Start Stream1 - Cl... Table (8 fie... Table (8 field... Table (8 field... Table (5 field... Table (5 field... Document1 -... 7:06 PM

Figure 8 Results

By default, the highest importance when calculating scores is given to the recency data, followed by frequency, and then monetary. If required, it can be amended by changing the weights.

Kumar (2006) explained that the relative weights of R, F, and M are computated using regression techniques and then those weights are used for calculating the combined effects of RFM. The RFM score is calculated as follows:

(Recency score x Recency weight) + (Frequency score x Frequency weight) + (Monetary score x Monetary weight). RFM score can be considered as the weighted sum of the recency, frequency, and monetary value scores for a customer. The RFM score tell the retailer which customers are more profitable to the company. The retailer can identify the promotional target customers with the highest RFM score. RFM analysis is considered significant for the retail industry. A customer who has visited a retailer Recently (R) and Frequently (F) and made a lot of Monetary Value (M) through payment and standing orders is very likely to visit and make payments again. After evaluation of the customer's behaviour using specific RFM criteria the RFM score is correlated to the retailer interest, with a high RFM score being more beneficial to the retailer currently as well as in the future.

Kahan and Kahan (1998) opinioned that the RFM is a powerful behavioral analysis technique which is easy and cost-effective, providing the customer and transactional information stored in an accessible electronic form which helps the database marketers to effectively use electronically captured information leading to three types of benefits: increased response rates; lowered cost per order; and greater profit.

Ghazanfari et al. (2008) developed a novel country segmentation methodology based on Recency (R), Frequency (F) and Monetary value (M) variables. After the variables are calculated, clustering methods are used to segment countries and compare the results of these methods by three different criteria. Customers are classified into four tiers: Top-active, Medium-active, New customer and Inactive. Then a customer pyramid is drawn and the customer value is calculated. Consequently, the data are used to analyze the relative profitability of each customer cluster and the proper strategy is determined for them. The key component to successful RFM is good record-keeping and tapping into these three measures of customer behavior - recency, frequency, and monetary - will put their increasingly rich insights to work to improve the fundraising results of a firm (BlackBaud, 2004). Kitayama (2002) explained an example of marketing method to establish customer strategies, using data mining technique based on customer profile data.

MANAGERIAL IMPLICATIONS

Managers can develop profiles of customers with certain behaviors, for example, those who purchase designer labels clothing or those who attend sales. This information can be used to focus cost–effective promotions. They can perform basket analysis through they can identify which items customers tend to purchase together. This knowledge can improve stocking, store layout strategies, and promotions. They can do Sales forecasting which helps in examining time-based patterns to make stocking decisions. There are various benefits to managers by implementing our model.

DIRECTIONS FOR FUTURE RESEARCH

The application of the proposed model in other industries could be a task of future work.

SUMMARY AND CONCLUSIONS

Data mining is a tool used to extract important information from existing data and enable better decision-making throughout retail industry. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed and the information that is captured is used throughout the organization to support decision-making. The retail industry is also realizing that data mining could give them a competitive advantage. The leading corporate in India have recognized that the business world is knowledge –intensive with innovative –or-die approaches. All corporate utilize the technology for storing and managing enterprise related data. Those retailers that have realized the utility of data mining and are in the process of building a data mining environment for their decision-making process will reap immense benefit and derive considerable competitive advantage to withstand competition in future. Data mining is a very powerful tool that should be used with utmost care for increasing customer satisfaction, providing best, safe and useful products at reasonable and economical prices. This should be used for making the business more competitive and profitable. Data mining should be used in any way that affects the privacy of common man, so that the confidentiality and individuality of

human being is preserved. It should not be used in any way that may cause undue hardship, financial loss or emotional setback.

REFERENCES

1. Aggelis, V. and Christodoulakis, D. (2005) 'Customer clustering using RFM analysis', *Proceedings of the 9th WSEAS International Conference on Computers,* World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA.
2. Ahmed, S. R. (2004), "Applications of Data Mining in Retail Business", *Proceedings of the International Conference on Information Technology: Coding and Computing,* Vol.2*,* pp. 455-459 IEEE.
3. Bala, P.K. (Oct22-24, 2008), Exploring Various forms of Purchase Dependency in Retail Sale, *Proceedings of the world congress engineering and computer science*,San Franciso, USA.
4. Bala, P.K.(2008), Retail inventory Management with Purchase Dependencies, Engineering Letter,16:4,EL_16_4_12.
5. Berman, B. and Evans, J.R. (2008) *Retail Management-A strategic Approach*, Person Publisher.
6. Berry, M.J.A. and Linoff, G.S. (2001) *Mastering Data Mining, The art of Customer relationship management*, John Wiley and Sons, Inc., New York.
7. Bhanu, D., Balasubramanie, P. (2008),Predictive modeling of inter-transaction association rules – A business perspective, *International Journal of Computer Science and Applications*,Vol.5, No.4, pp 57-69.
8. BlackBaud (2004) 'RFM: a formula for greater direct mail success', (retrieved from www.blackbaud.com on 20 April 2009).
9. Chen, S.Y. and Liu, X. (2005), 'Data mining from 1999 to 2004: an application – oriented review', *International Journal of Business Intelligence and Data Mining*, Vol. 1, No. 1, pp.4–21.
10. Cowley, S. (2005) *Data Mining*, IDG News Service, New York Bureau.
11. Dennis,C., Marsland,D., Cockett,T.(2001), Data Mining for Shopping Centers- Customer Knowledge Management Framework, *Journal of Knowledge Management*,Vol.5, No.4, pp.368-374.
12. Feelders, A. , Daniels, H. and Holsheimer, M. (2000), "Methodological and Practical Aspects of Data Mining", Information *and Management*, Volume 37, Issue 5, pp.271-281.
13. FICCI Retail Report 2007, www.ficci.com (accessed on 25 th June 2008).
14. Ghazanfari, M., Mohamadi, S.M. and Alizadeh, S. (2008) 'Data-mining application for country segmentation based on the RFM model', *International Journal of Data Analysis Techniques and Strategies*, Vol. 1, No. 2, pp.126–140.
15. Golfarelli, M., Rizzi,S.,Cella,I.(2004): Beyond Data warehousing: What's next in Business Intelligence?Proc.DOLAP'04,Washington,DC,USA
16. Guzman, I. (2002) 'A strategic decision support tool for organisations', *Strategic Management of Information Resources*, Research paper.
17. Han, J. and Kamber, M. (2007), *Data Mining*, Morgan Kaufmann Publishers.

18. *Hanna, J. (2004),* "Ground-Floor Opportunities for Retail in India", *Harvard Business School Newsletter.*
19. Hou, J.J. and Tu, H.H.J. (2008) 'Customer relationship management strategy and firm performance: an empirical study', *Int. J. Electronic Customer Relationship Management*, Vol. 2, No. 4, pp.364–375.
20. Jones, S. and Ranchhod, A. (2007) 'Marketing strategies through customer attention: beyond technology enabled customer relationship management', *International Journal of Electronic Customer Relationship Management*, Vol. 1, No. 3, pp.279–286.
21. Kahan, R. and Kahan, A. (1998) 'Using database marketing techniques to enhance your one-to-one marketing initiatives', *Journal of Consumer Marketing, Database Marketing Group, Inc*., Vol. 15, No. 5, pp.491–493, Denver, Colorado, USA.
22. Kaur, P. and Singh, R. (2007), "Uncovering retail shopping motives of Indian youth", *Young Consumers: Insight and Ideas for Responsible Marketers*, Vol. 8, No. 2, pp.128–138.
23. Kitayama, M., Matsubara, R. and Izui, Y. (2002) 'Application of data mining to customer profile analysis in the power electric industry', *Power Engineering Society Winter Meeting, IEEE*, Vol. 1, pp.632– 634.
24. Kumar, V. and Reinartz, W.J. (2006) *Customer Relationship Management: A Databased Approach*, John Wiley, New York.
25. List, B., Machaczek, K.(2004), Towards a Corporate Performance Measurement System.Proc. ACM SAC'04,Nicosia,Cyprus
26. Marketos, G. and Theodoridis, Y. (2006), Measuring the performance in the retail industry, Business Process Management Workshops, Vol. 4103, pp.129-140.
27. Marketos, G., Kotsifakos, E., Theodoridis, Y., (2007) '*Rethinking Decision Making: Measuring and Managing Performance*' Proc. Pan-Hellenic Conference on New Technologies and Marketing, Ierapetra, Crete, May.
28. Noonan, J. (2000) 'Data mining strategies', *DM Review*.
29. Pan, H.(2005), Factors affecting IT Adoption: The case of a Chinese Retail Company,0-7803-8971-9, IEEE
30. Pan, J., Yang, Q., Yang, Y., Li, L., Li, L., Li, F.T. and Li, G.W. (2007) 'Cost-sensitive-data pre-processing for mining customer relationship management databases', *IEEE Intelligent Systems*, Vol. 22 No. 1, pp.46–51.
31. Pande, S. and Collins, T. (2007), "Strategic implementation of information technology to improve retail supply chain in India", *International Journal of Logistics Systems and Management*, Vol. 3, No. 1, pp.85–100.
32. Ranjan, J. and Bhatnagar, V. (2008), 'Data Mining tools: a CRM perspective', *International Journal Electronic Customer Relationship Management*, Vol. 2, No. 4, pp.315-331.
33. Rao, I. K. R. (2003), "Data Mining and Clustering Techniques", DRTC Workshop on Semantic Web, December.
34. Rogers, M. (2005), "Customer strategy: observations from the trenches", *Journal of Marketing*, Vol. 69 No.4, pp.262.

35. Sangle, P.S. and Verma, S. (2008) 'Analysing the adoption of customer relationship management in Indian service sector: an empirical study', *Int. J. Electronic Customer Relationship Management*, Vol. 2, No. 1, pp.85–99.
36. Smith, A.D. (2006) 'CRM and customer service: strategic asset or corporate overhead?' *Handbook of Business Strategy*, 7(1), .87–93.
37. Sohoni, A. (2007), "Indian Retailers - Ready for Take Off?" available from http://www.tech2.com/biz/india/features/retail/indian-retailers-ready-for-take-off/1313/0 (accessed on 02-august-2008).
38. Srivastava, R.K. (2008), Changing Retail scene in India, *International Journal of Retail & Distribution Management*, Vol.36, No.9.
39. Terrovitis, M. and Vassilliadis, P.(2003),Architecture for Pattern-Base Management System, *In Proc. of PANDA Workshop on Pattern-base Management Systems (PBMS)*, pp. 21-30, Villa Olmo, Como, Italy.
40. Thearling, K. (2007) 'Data mining and customer relationships', available from: www.thearling.com, (accessed on 22 August 2008).
41. Tvrdikova, M.(2007), Support of Decision making by business intelligence tools, *6th International Journal on Computer Information Systems and Industrial Management Applications*,0-7695-2894-5,pp.364-368, IEEE
42. Two Crows corporation, "Introduction to Data Mining and Knowledge Discovery", available at http://www.twocrows.com/ (Accessed on 25/july/2008).
43. van der Aalst, W.M.P., van Dongen, B.F., Herbst, J.,Maruster, L.,Schimm, G., Weijters, A.J.M.M.: Work Flow mining: a survey of issues and approaches, *Data & Knowledge Engineering,* 47(2), 237-267
44. Vector, D. (2007), "Indian Retail Industry: Strategies, Trends and Opportunities 2007", available at http:// www. Marketresearch .com /product /display.asp? productid = 1497236 (accessed on 4th July 2008).
45. West, D. (2005) 'Enhancing value through data mining: Insurers can use data mining technology to improve their competitive position', *Insurance Networking News: Executive Strategies for Technology Management*, October.
46. Wang, H. and Wang, S. (2007) 'Mining purchasing sequence data for online customer segmentation', *International Journal of Service Operations and Informatics*, Vol. 2, No. 4, pp.382–390.
47. West, D. (2005) 'Enhancing value through data mining: Insurers can use data mining technology to improve their competitive position', *Insurance Networking News: Executive Strategies for Technology Management*, October.
48. Wong, W., Fu, A.W. and Wang, K. (2005), Data mining for inventory item selection with cross-selling considerations, *Data Mining and Knowledge Discovery*, 11(1), 81-112.
49. Wu, R.C., Chen, R.S. and Chen, J.Y. (2005) 'Data mining application in customer relationship management of credit card business', *Computer Software and Applications Conference, IEEE*, Vol. 2, pp.39–40.

REQUEST FOR FEEDBACK

Esteemed & Most Respected Reader,

At the very outset, International Journal of Research in Commerce and Management (IJRCM) appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to take this opportunity to request to your good self to supply your critical comments & suggestions about the material published in this issue as well as on the journal as a whole, on our E-mails i.e. **info@ijrcm.org.in** or **infoijrcm@gmail.com** for further improvements in the interest of research.

If your good-self have any queries please feel free to contact us on our E-mail **enquiryijrcm@gmail.com**.

Hoping an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Editor