# ACKNOWLEDGEMENT

First and foremost, we would like to convey our gratitude to the department of Computer Engineering for giving us such an opportunity to present our skills and ability. We have tried our level best to complete our proposed project successfully. We are greatly indebted to our project coordinator **Er. Reshma Maharjan** for the suggestions and guidance despite her busy schedule. We would also like to thank **Er. Sudeep Shakya** (Head of Department,Computer Engineering), our supervisor and year coordinator **Er. Amar** **Bahadur Gurung** and friends for giving us invaluable suggestion for decision making of this project.

# ABSTRACT

**Show Shopper** is data analysis in an inventory system. It analyzes the data to help an organization in better and informed decision making. With analyzed data, it provides recommendation to the customer for items frequently bought together and visualization to the employee which will help them in identifying the customer’s need. All the analysis is done on R: an open source data analysis tool. Apriori algorithm is used for implementation of association analysis from the dataset of transaction. This system can be implemented in any inventory market, as it also provides user with simple Management Information System (MIS) functionality which makes this system dynamic; as new transaction, by the customer, also updates the database with new patterns. Database used in this system is MySQL. This system segments customer into 4 hierarchy according to the importance towards the organization. Reference to this cluster result will give user the ability to know the valued customers and make plans according to their needs. Customer behaviors are record on the basis of RFM analysis, which will support employees in understanding the customer to the system. RFM analysis of real world data with the response of customer to a marketing scheme is used for training models like Random Forest, SVM and fast AdaBoost for finding the customer who will response to our recommendation and messages. This prediction based system will help minimize marketing budget allowing user to know the outcome before hand.

## KEYWORDS:

Data analysis, Association Analysis, RFM analysis, Apriori, Management Information System, Inventory Management, Random Forest, SVM, AdaBoost

TABLE OF CONTENT

[ACKNOWLEDGEMENT i](#_Toc489266036)

[ABSTRACT i](#_Toc489266037)

[KEYWORDS: i](#_Toc489266038)

[LIST OF FIGURE iv](#_Toc489266039)

[LIST OF ABBREVIAITON v](#_Toc489266040)

[CHAPTER 1: INTRODUCTION 1](#_Toc489266041)

[1.1 Background Theory 1](#_Toc489266042)

[1.2 Problem Statement 2](#_Toc489266043)

[1.3 Objectives 2](#_Toc489266044)

[1.4 Scope and Application 3](#_Toc489266045)

[CHAPTER 2: LITERATURE REVIEW 3](#_Toc489266046)

[CHAPTER 3: RELATED THEORY 4](#_Toc489266047)

[CHAPTER 4: METHODOLOGY 6](#_Toc489266048)

[4.1 System Planning 6](#_Toc489266049)

[4.1.1 Software Model (Incremental Model) 6](#_Toc489266050)

[4.1.2 Data Analysis Process 7](#_Toc489266051)

[4.2 System Analysis 9](#_Toc489266052)

[4.2.1 Feasibility Study 9](#_Toc489266053)

[4.3 System Design 9](#_Toc489266054)

[4.3.1 Overview of the System 9](#_Toc489266055)

[4.3.2 Data Flow Diagram 11](#_Toc489266056)

[4.3.3 Schema Diagram 13](#_Toc489266057)

[4.4 Implementation 14](#_Toc489266058)

[CHAPTER 5: EPILOGUE 19](#_Toc489266059)

[5.1 Task Completed 19](#_Toc489266060)

[5.2 Task to be Completed 19](#_Toc489266061)

[5.3 Gantt Chart 20](#_Toc489266062)

[REFERENCE 21](#_Toc489266063)

[BIBLIOGRAPHY 23](#_Toc489266064)

# LIST OF FIGURES

[Figure 1 Incremental Model 7](#_Toc485359103)

[Figure 2 Data analysis process 8](file:///C:\Users\yushg\Desktop\Show%20Shopper\Inventory%20Management.docx#_Toc485359104)

[Figure 3 Complete System Overview 10](file:///C:\Users\yushg\Desktop\Show%20Shopper\Inventory%20Management.docx#_Toc485359105)

[Figure 4 Context Diagram 11](#_Toc485359106)

[Figure 5 DFD Level 0 for show shopper 12](#_Toc485359107)

[Figure 6 Schema Diagram for show shopper 13](#_Toc485359108)

[Figure 7 Login for employee and customer 14](#_Toc485359109)

[Figure 8 List if items, with the amount left in stock 14](#_Toc485359110)

[Figure 9 Ordering Item from vendor 15](#_Toc485359111)

[Figure 10 Adding New Item into the system 15](#_Toc485359112)

[Figure 11 All the data collected 16](#_Toc485359113)

[Figure 12 Data in basket format after cleaning & preprocessing 16](#_Toc485359114)

[Figure 13 Association Rules extracted from the data set 17](#_Toc485359115)

[Figure 14 Recommendation to the customer 17](#_Toc485359116)

[Figure 15 Visualization to the employee 18](#_Toc485359117)

[Figure 16 Gantt Chart for our development 20](#_Toc485359118)

# LIST OF ABBREVIAITON

CDA: Confirmative Data Analysis

DFD: Data Flow Diagram

EDA: Exploratory Data Analysis

MIS: Management Information System

MBA: Market Basket Analysis

RDB: Relational Database

RFM: Recency Frequency Monetary

SQL: Structured Query Language

UPC: Universal Project Code

# CHAPTER 1: INTRODUCTION

Data are factual information which is used for referencing and analysis. In computing term, it is the information that is in the form efficient for storage and transfer. Data at present are stored in binary format. It has been an integral part of various fields like: marketing, telecommunication, sports and many more. All these data are termed as raw data; it is the data in their original form. This data can be valuable information, and if used properly in business can produce benefit to the user.

## 1.1 Background Theory

Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. It is a process for obtaining raw data and converting it into information useful for decision making by users. It can be divided into exploratory data analysis (EDA) and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data and CDA confirms or falsify existing hypothesis.

There are several phases in data analysis process, which are: data collection, data processing, data cleaning, exploratory data analysis, modeling algorithms and data product. Data are collection done from real world, this data that we collected are not in a formatted order therefore they must be cleaned and processed to bring it to a point where it can be analyzed.

Data analysis can be used in many field, one such field is market. Application of data analysis in market are market basket analysis, customer segmentation, customer profitability, churn analysis etc. In a market, we collect typical manner in which customer purchase good, such data are usually known as Market Basket Transaction. Each transaction is listed in a row with a unique id. Retailer are concerned with analyzing such data to produce valuable information that can be used to support business related application such as inventory management, marketing promotions and customer relationship management. Analysis in these basket formats of transaction is known as Market Basket Analysis(MBA). It is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items. For example, if you are in an English pub and you buy a pint of beer and don't buy a bar meal, you are more likely to buy chips at the same time than somebody who didn't buy beer [1].

Technique for implementing Market Basket Analysis to find relationship between items can be done with Association analysis; which is used to produce important information from large unknown set of data. Inventory management would get a boost if used along with this. All the data produced through regular transaction with the costumer, if can be analyzed would be very beneficial; resulting in the boost of retailer’s business.

Another type of analysis that can be performed in a market is customer profiling or also known to be segmentation. It uses RFM analysis where R stands for recency, F for frequency and M for monetary. It has many benefit such as increased customer retention, increased response rate, increased conversion rate and revenue [2].

To get to the customers and let them know about the new products in our Inventory is a challenging task, and needs a lot of creative ideas. Customers gets the stuff of the brand only they trust. Also, all the customers will not take the item recommendation in a positive way and if are persuaded more might cut the deal with the retailer. Also on recommending the customer about a product resources are spent. In order to find and select the customer who are likely to respond to the recommendation by our system, customer behavior is analyzed from the previous customer’s data. On the basis of RFM analysis and using predictive algorithms, such as Random Forest, SVM and AdaBoost, customers are classified according to their decision and response to the system’s recommendation. Only the customer who have high chance of responding to our system are then sent product recommendation messages and offers.

## 1.2 Problem Statement

Presently, small mart stores data but seldom use it in the business except for the verification. These data are just bits of information which rusts away in some storage device. Small marts lack the knowledge and understanding of data analysis. Therefore, have not opted for one yet. Which if used could be source of an invaluable information for business development. Retail market does not have the knowledge to base their marketing schemas and recommendation, and using this real-time data from the customer user can produce higher revenue with low expenditure.

## 1.3 Objectives

Analysis on present data of inventory to learn customer’s need and behavior, with simple MIS functionality for continuous data update.

## 1.4 Scope and Application

This system will be applicable to small and medium inventories which want to use stored data to produce result for their better business decision. It can be used to produce recommendation to the customer with frequently bought itemset suggestions. It also can produce itemset pattern visualization to the employees to help them make decision in case of schemes and advertisement. Customer’s most valued to the organization are segmented which helps in marketing new scheme to with higher response rate from the customer. This system helps the owner peak into customer’s mind and understand them.

# CHAPTER 2: LITERATURE REVIEW

Statistician John Tukey defined data analysis in 1961 as: "Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning, gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data [8]." There have been various data analysis tool for analysis, and one of them is R.

R is an integrated suite of software facilities for data manipulation, calculation and graphical display [9]. R can be regarded as an implementation of the S language which was developed at Bell Laboratories by Rick Becker, John Chambers and Allan Wilks, and also forms the basis of the S-Plus systems. The evolution of the S language is characterized by four books by John Chambers and coauthors. For R, the basic reference is The New S Language: A Programming Environment for Data Analysis and Graphics by Richard A. Becker, John M. Chambers and Allan R [10]. Wilks. The new features of the 1991 release of S are covered in Statistical Models in S edited by John M. Chambers and Trevor J. Hastie [11]. The formal methods and classes of the methods package are based on those described in Programming with Data by John M. Chambers [12].

R is a statistical tool with many classical and modern statistical techniques as packages. It has about 25 packages that are standard and recommended. One of those package is arules. It provides the generic function and the methods to abbreviate long item labels in transactions, associations (rules and itemsets) and transaction ID lists [13]. A market transaction data is cleaned, preprocessed and mined for data through arules in thesis published by Pazaras Christos, where a case study on a dataset containing 247535 records with 2037 items contained in 33701 transactions of a supermarket was performed [14]. The association rules, after the analysis can be viewed using another package in R which is arulesViz. This apriori analysis when combined with RFM analysis in R, become a great tool for market revenue generation. RFM uses sales data to segment a pool of customers based on their purchasing behavior. The resulting customer segments are neatly ordered from most valuable to least valuable. [15]. Our system brings this RFM and apriori into one system to be implemented in a small market and inventory.

# CHAPTER 3: RELATED THEORY

**Association Rule**: An association rule is an implication expression of form X🡪Y, where X and Y are disjoint itemset, i.e. X∩Y = Ø. The strength of an association rule can be measured in terms of its support and confidence. Support determines how often a rule is applicable to a given dataset, while confidence determines how frequently items in Y appear in transaction that contain X. The formal definitions of these metrics are

Support, s(X🡪Y) =

Confidence, c(X🡪Y) =

**R**: R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering…), graphical techniques, and is highly extensible. One of R’s strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.

It is available as Free Software under the terms of the Free Software Foundation’s GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS [3].

**PHP**: PHP (recursive acronym for PHP: Hypertext Preprocessor) is a widely-used open source general-purpose scripting language that is especially suited for web development and can be embedded into HTML. Instead of lots of commands to output HTML (as seen in C or Perl), PHP pages contain HTML with embedded code that does "something" (in this case, output "Hi, I'm a PHP script!"). The PHP code is enclosed in special start and end processing instructions <? php and ?> that allow you to jump into and out of "PHP mode."

What distinguishes PHP from something like client-side JavaScript is that the code is executed on the server, generating HTML which is then sent to the client. [4].

**Fusion Chart**: FusionCharts helps you create animated & interactive charts for web & enterprise applications. It is the industry’s leading enterprise-grade charting component that functions seamlessly on PCs, Macs, iPads, iPhones and a majority of other mobile devices.

It leverages Flash and JavaScript (HTML5) to create stunning charts, and works with both XML and JSON data. It can be integrated with any server-side technology (ASP, ASP.NET, PHP, JSP, ColdFusion, Ruby on Rails etc.) and database [5].

Co-founder Pallav Nadhani came up with the name "FusionCharts" because the product was a fusion of two disparate technologies, XML and Flash.

**MySQL**: MySQL, the most popular Open Source SQL database management system, is developed, distributed, and supported by Oracle Corporation [5]. Its name is a combination of "My", the name of co-founder Michael Widenius' daughter, and "SQL", the abbreviation for Structured Query Language. The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation [6].

**RFM analysis:** To perform RFM analysis, each customer is assigned a score for recency, frequency, and monetary value, and then a final RFM score is calculated.

Recency score is calculated based on the date of their most recent purchase. The scores are generally categorized based on the values. For example, a company may follow a category system of 1 to 5, score of 5 being the highest. In this case, customers who purchased within the last one month have a recency score of five, customers who purchased within the last 1-3 months have a score of four and so on. Similarly, frequency score is calculated based on the number of times the customers purchased. Customers with higher frequency receive a higher score

Finally, customers are assigned a score based on the amount they spent on their purchases. For calculating this score, you may consider the actual amount spent or the average spent per visit. By combining these three scores, a final RFM score is calculated. The customers with the highest RFM score are considered to be the ones that are most likely to respond to their offers.

RFM analysis is a powerful technique to help you identify your best customers and create better targeted campaigns. However, RFM itself is not enough and retailers should focus on creating more detailed customer profiles including their demographics, behavioral and purchase patterns and use this information in conjunction with RFM to provide better value to customers [7].

**Random Forest:** A Random Forest consists of a collection or ensemble of simple [tree](http://www.statsoft.com/textbook/classification-and-regression-trees/) predictors, each capable of producing a response when presented with a set of predictor values. For classification problems, this response takes the form of a class membership, which associates, or classifies, a set of [independent](http://www.statsoft.com/textbook/statistics-glossary/i.aspx?button=i#Independent%20vs.%20Dependent%20Variables) predictor values with one of the categories present in the [dependent variable](http://www.statsoft.com/textbook/statistics-glossary/i.aspx?button=i#Independent%20vs.%20Dependent%20Variables). Alternatively, for regression problems, the tree response is an estimate of the dependent variable given the predictors. The Random Forest algorithm was developed by Breiman.

A Random Forest consists of an arbitrary number of simple trees, which are used to determine the final outcome.  For classification problems, the ensemble of simple trees vote for the most popular class. In the regression problem, their responses are averaged to obtain an estimate of the dependent variable. Using tree ensembles can lead to significant improvement in prediction accuracy (i.e., better ability to predict new data cases).

The response of each tree depends on a set of predictor values chosen independently (with replacement) and with the same distribution for all trees in the forest, which is a subset of the predictor values of the original data set. The optimal size of the subset of predictor variables is given by log2 *M*+1, where *M* is the number of inputs.

For classification problems, given a set of simple trees and a set of random predictor variables, the Random Forest method defines a margin function that measures the extent to which the average number of votes for the correct class exceeds the average vote for any other class present in the dependent variable. This measure provides us not only with a convenient way of making predictions, but also with a way of associating a confidence measure with those predictions.

For regression problems, Random Forests are formed by growing simple trees, each capable of producing a numerical response value. Here, too, the predictor set is randomly selected from the same distribution and for all trees. Given the above, the mean-square error for a Random Forest is given by:

mean error = (observed - tree response)2

The predictions of the Random Forest are taken to be the average of the predictions of the trees:

Random Forest Predictions

where the index *k* runs over the individual trees in the forest.

Typically, Random Forests can flexibly incorporate [missing data](http://www.statsoft.com/textbook/statistics-glossary/m.aspx?button=m#Missing%20values) in the predictor variables. When missing data are encountered for a particular observation (case) during model building, the prediction made for that case is based on the last preceding (non-terminal) node in the respective tree. So, for example, if at a particular point in the sequence of trees a predictor variable is selected at the root (or other non-terminal) node for which some cases have no valid data, then the prediction for those cases is simply based on the overall mean at the root (or other non-terminal) node. Hence, there is no need to eliminate cases from the analysis if they have missing data for some of the predictors, nor is it necessary to compute surrogate split statistics.[RefRandomF].

**Support Vector Machine (SVM):** A Support Vector Machine is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes.

Support vectors are the data points nearest to the hyperplane, the points of a data set that, if removed, would alter the position of the dividing hyperplane. Because of this, they can be considered the critical elements of a data set.

The distance between the hyperplane and the nearest data point from either set is known as the margin. The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified correctly. [RefSVM]

**AdaBoost:** Classification is a machine-learning technique that uses training data to generate a model (usually a single complex rule or mathematical equation) that assigns data items to one of several distinct categories. The model can then be used to make predictions about new data items whose category is unknown. Adaptive boosting classification is a technique in which, instead of attempting to determine a single complex prediction rule, training data is used to generate a large collection of very simple crude rules of thumb. A weight for each rule of thumb is then computed. A prediction about new input is made by combining the rules of thumb, taking into account each simple rule’s weight and arriving at a consensus outcome. The term “boosting” comes from the fact that the predictive quality of the simple rules is boosted (improved) by combining them.

Adaptive boosting is a meta-heuristic, which means adaptive boosting is a set of guidelines that can be used to create a specific classification algorithm. There are many variations of adaptive boosting algorithms and there are many existing standalone tools that implement some form of adaptive boosting, so why bother to code adaptive boosting classification from scratch? Existing adaptive boosting classification tools can be difficult or impossible to customize, they might be difficult to integrate into a software system, and they may have copyright or intellectual property issues. [RefAdaBoost]

# CHAPTER 4: METHODOLOGY

## 4.1 System Planning

### 4.1.1 Software Model (Incremental Model)

In Incremental model, the whole requirement is divided into various builds. Multiple development cycles take place here, making the life cycle a multi waterfall cycle.  Cycles are divided up into smaller, more easily managed modules.  Each module passes through the requirements, design, implementation and [testing](http://istqbexamcertification.com/what-is-a-software-testing/) phases. A working version of software is produced during the first module, so you have working software early on during the [software life cycle](http://istqbexamcertification.com/what-are-the-software-development-life-cycle-phases/). Each subsequent release of the module adds function to the previous release. The process continues till the complete system is achieved.



Figure 1 Incremental Model

**Advantages of Incremental model:**

* Generates working software quickly and early during the software life cycle.
* This model is more flexible-less costly to change scope and requirements.
* It is easier to test and debug during a smaller iteration.
* In this model customer, can respond to each built.
* Lowers initial delivery cost.
* Easier to manage risk because risky pieces are identified and handled during it’d iteration.

**Disadvantages of Incremental model:**

* Needs good planning and design
* Needs a clear and complete definition of the whole system before it can be broken down and built incrementally.

### 4.1.2 Data Analysis Process

Figure 2 Data analysis process

Raw Data Collected

Data is Processed

Clean Dataset

Exploratory

Data Analysis

Models and Algorithms

Communicate

Visualize

Report

Data Product

Inventory

Make

Decisions

The above figure is the block diagram of Data Analysis process. Raw data of from the real-world transaction was collected from an open source platform with customers behavior data. These raw data were then processed and changed into the format on which we can use market basket analysis and other predictive analysis. After we got the data in basket form, an apriori algorithm was applied on this “clean dataset” to generate frequent itemset which is later used for recommendation of items in our Show Shopper system. Apriori algorithm is to be performed on R in our project. The obtained result was further processed and modeled to properly upload the generated association rules on our local server database (MySQL). The results were extracted from database in the form of related itemset and visualized properly. Then raw dataset of customers is analyzed on the basis of Recency, Frequency and Monetary (RFM), which then helped in training predictive model that predicted the response of the customer to our recommendation and schema.

## 4.2 System Analysis

### 4.2.1 Feasibility Study

**Technical Feasibility**

The above-mentioned project is technically feasible since the necessary tools and techniques are already available. All the software is easily available, we will use R and other analyzing system for statistical analysis. R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

**Schedule Feasibility**

For the development of this application the suggested process model is incremental modeling. Following the process model, we will prioritize our schedule on bare minimum functional application development. With this objective in mind with the given time frame of 6 months we will be able to deliver the product at the end of major project.

**Economic Feasibility**

Most of the project tools and technique that we will be using in this project will be open-source and these tools are easily available and are in disposal for our use. Most of the tools are free with great support, in context to our project with available resources this project is economically feasible.

## 4.3 System Design

### 4.3.1 Overview of the System

Employee

Vendor

Inventory

Customer

Data Analysis Process

Inventory

Database

Manual

Ordering

Confirm

Payment

Auto Ordering

Add Items

Sales Order

Recommend Product

Report

Make

Decision

Figure 3 Complete System Overview

The above figure shows a block diagram of our entire system: Show Shopper. Show Shopper has an inventory acting as a central unit. There are basically two main entities in the system: Customer and Employee. Each user must login into the system before accessing any features. Customers can view the item list and order the required item by adding it into the cart. During adding items to the cart, customers get a recommended list of items continually. Recommended items are shown for easy and convenient booking of items. Employee can access the customer’s cart and after payment confirmation, cart items are handled accordingly. Recommendation to the customers are given after analyzing the data set for frequent itemset, plus a RFM analysis is also done which helps understand the behavior of the customer. According to the customer behavioral data, predictive models are also trained and produced. These model will help in predicting response of customer according to the behavioral pattern.

### 4.3.2 Data Flow Diagram

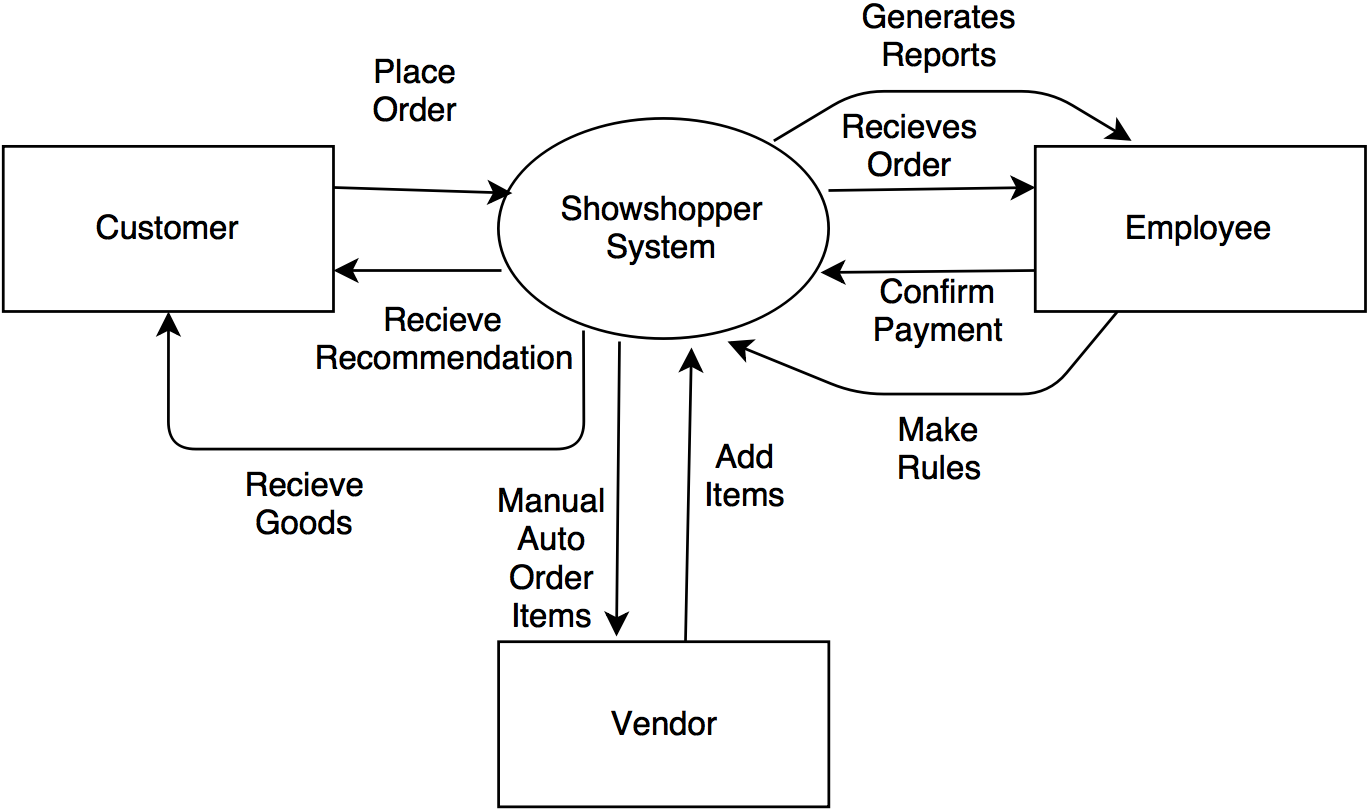


Figure 4 Context Diagram

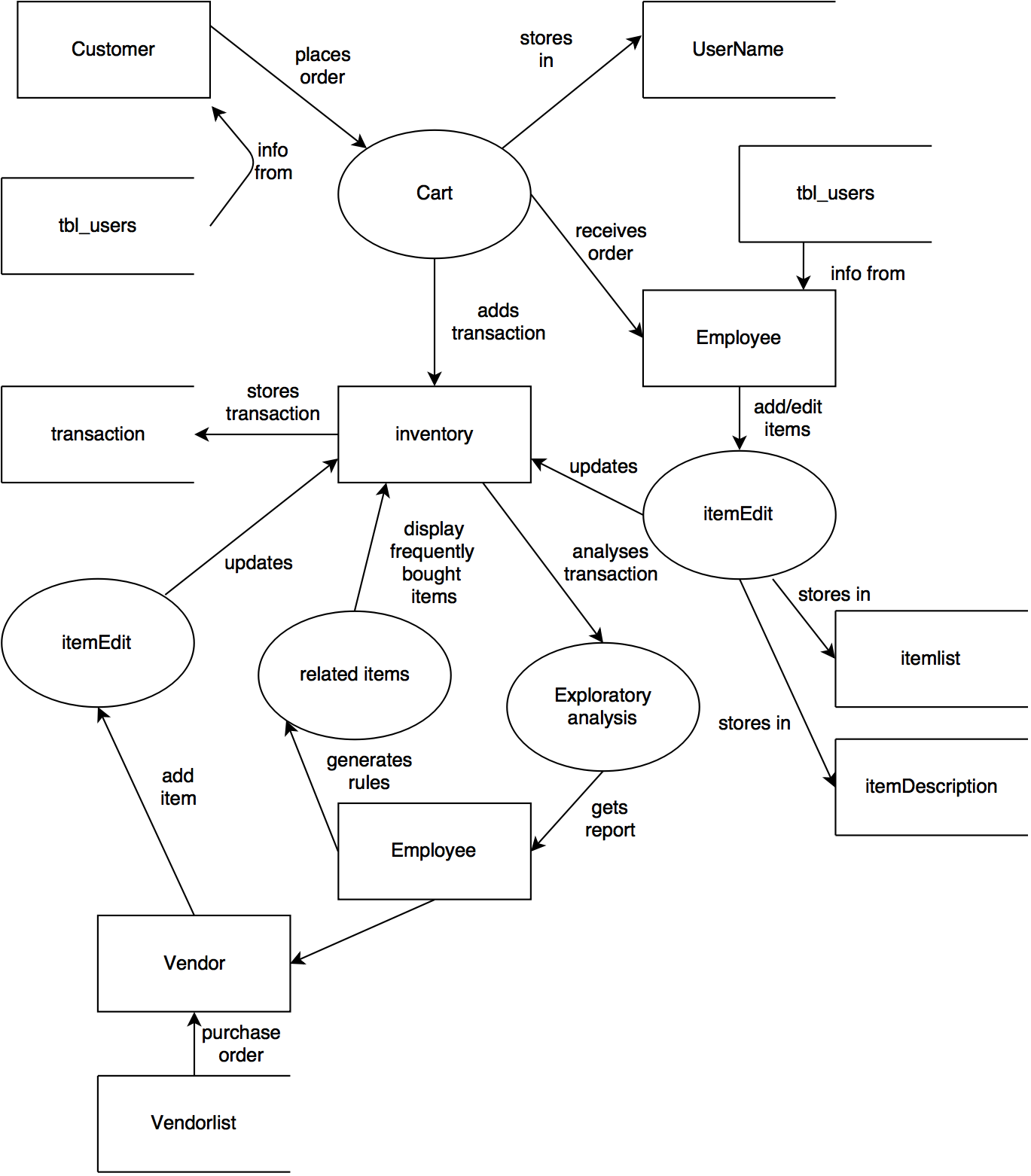


Figure 5 DFD Level 0 for show shopper

### 4.3.3 Schema Diagram

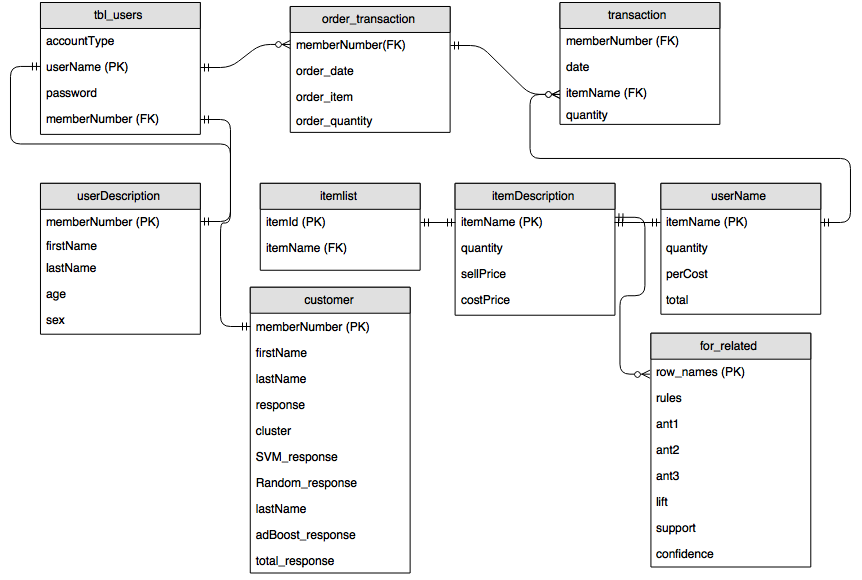


Figure 6 Schema Diagram for show shopper

## 4.4 Implementation

**Data Analysis Process**

**Data Collection:** The data collection for the analysis process was obtained through an open source platform. The data has 5000 unique customers, with 38766 transaction data. The data has 3 attributes: date, member number and item. The data sample is shown below.

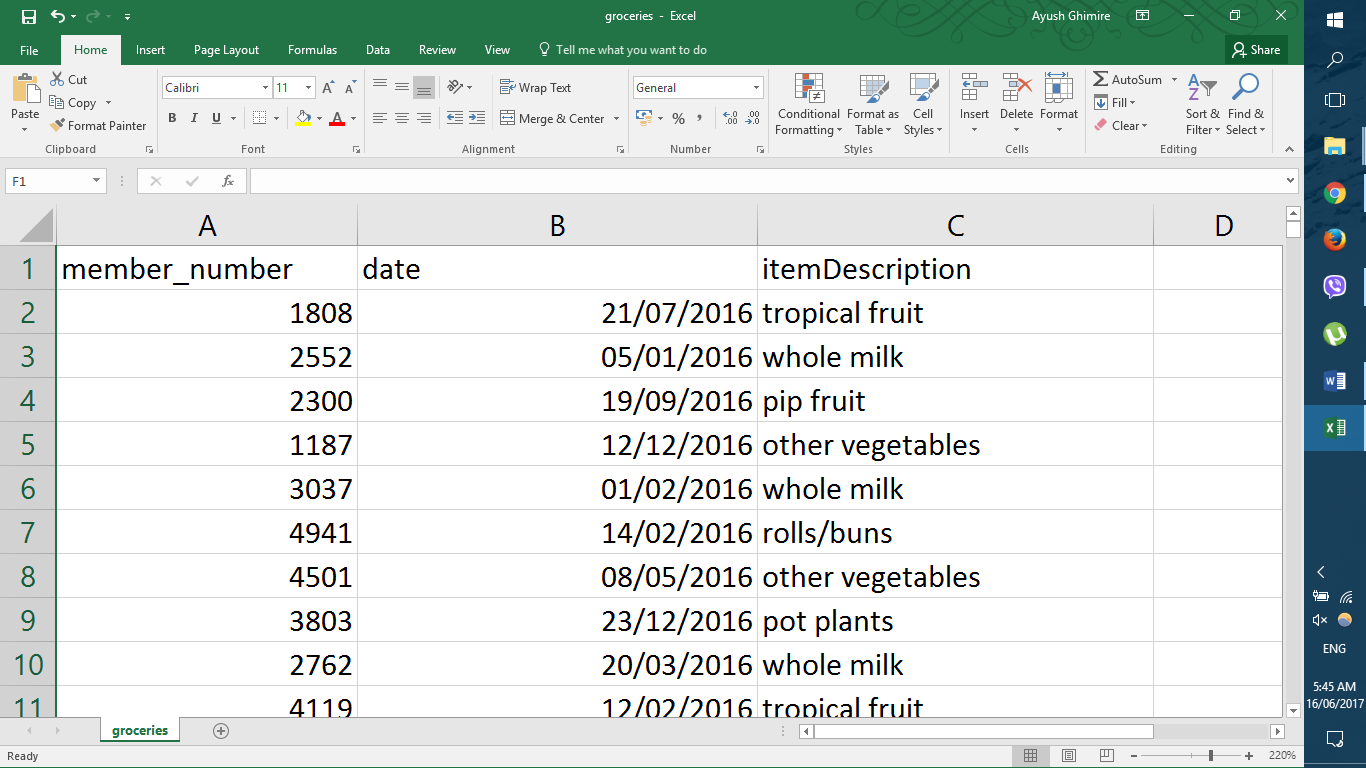


Figure 11 All the data collected

**Data cleaning & processing:** This data was imported into MySQL database where it was stored for future use. The stored data is then retrieved into R for cleaning and preprocessing. White strip before and after the data tuples where removed when reading the data from MySQL. With the help of plyr function in R, we processed the atomic data to produce transactional data. As we require the data to be in basket format for market basket analysis. Data set after processing is shown below

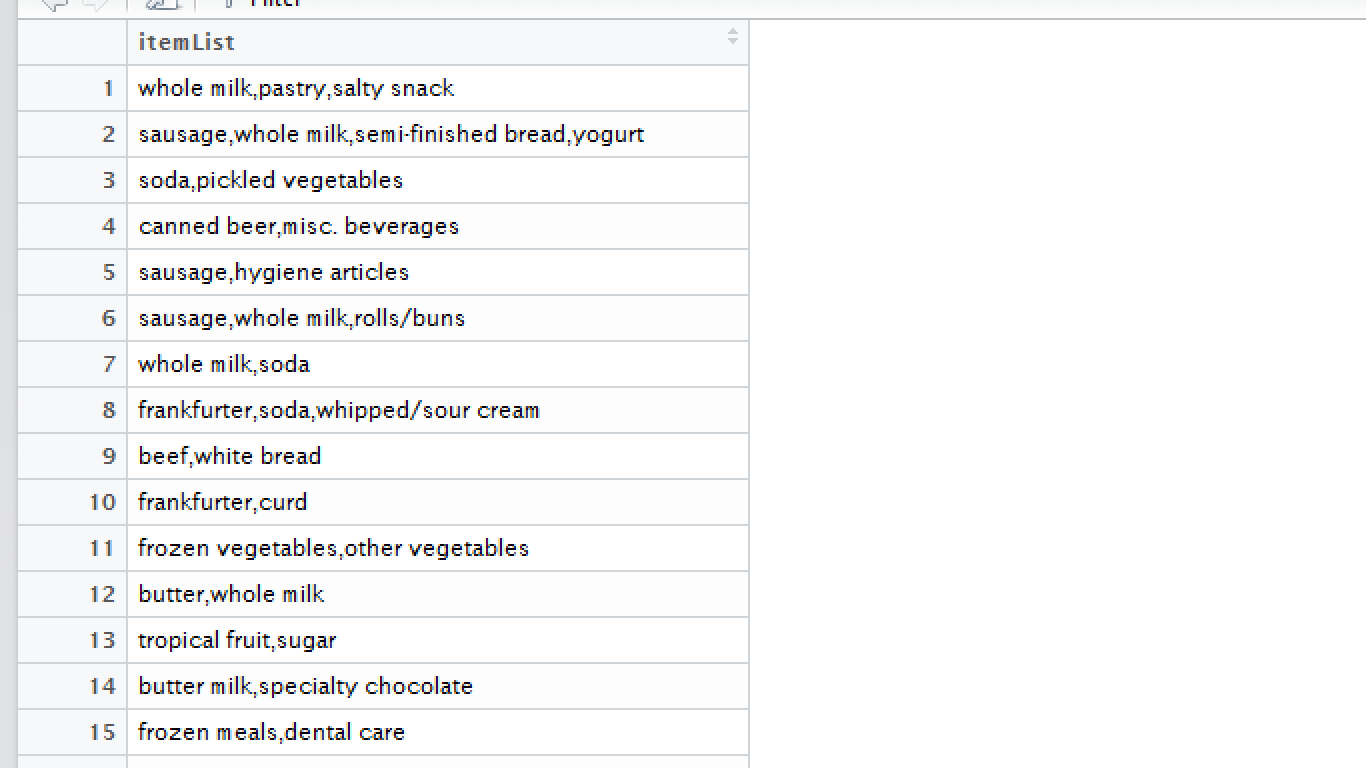


Figure 12 Data in basket format after cleaning & preprocessing

Above dataset are used for market basket analysis and then to train a model. Customers behavior on the basis of RFM is calculated and produced

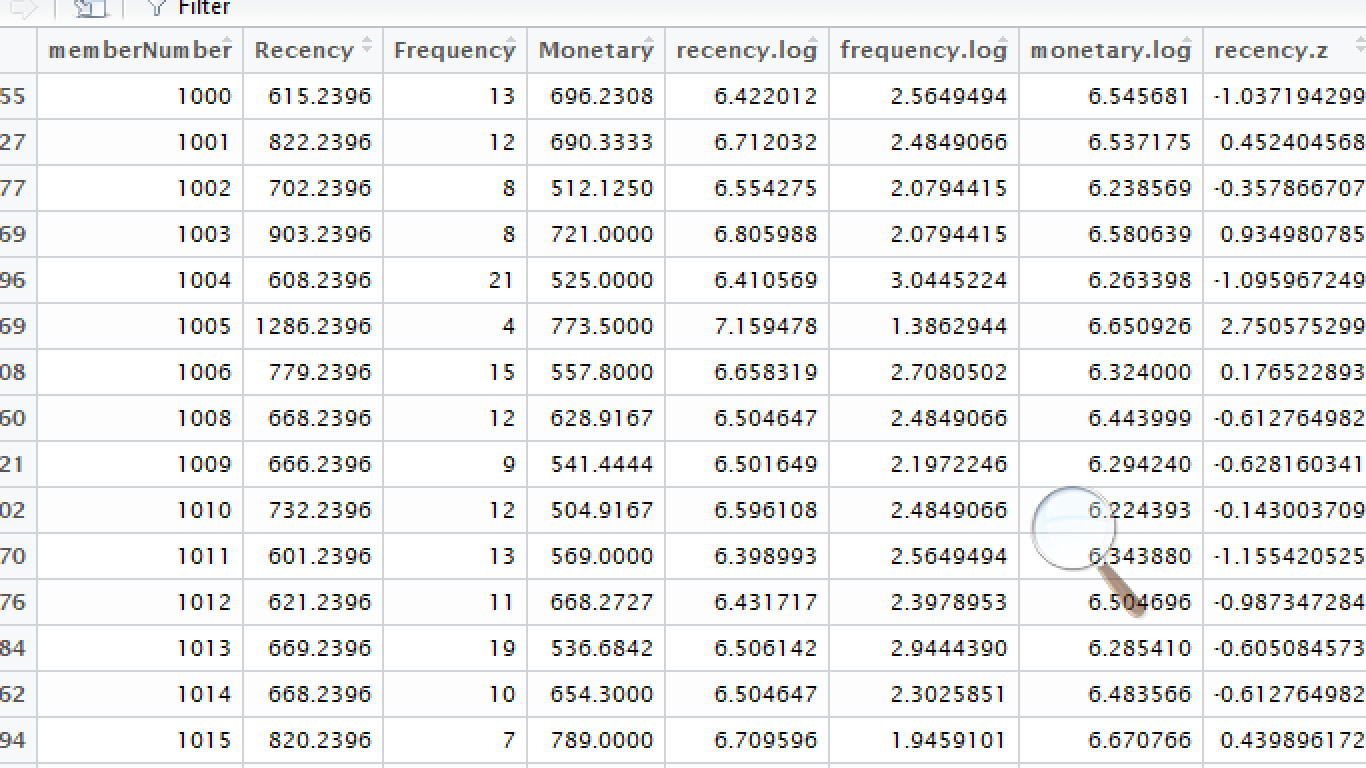


Figure 12 RFM analysis after cleaning & preprocessing

**Model & algorithm:** After the basket of item where produced from the data set, we applied the apriori analysis to produce association rules. Association rules achieved is shown below:

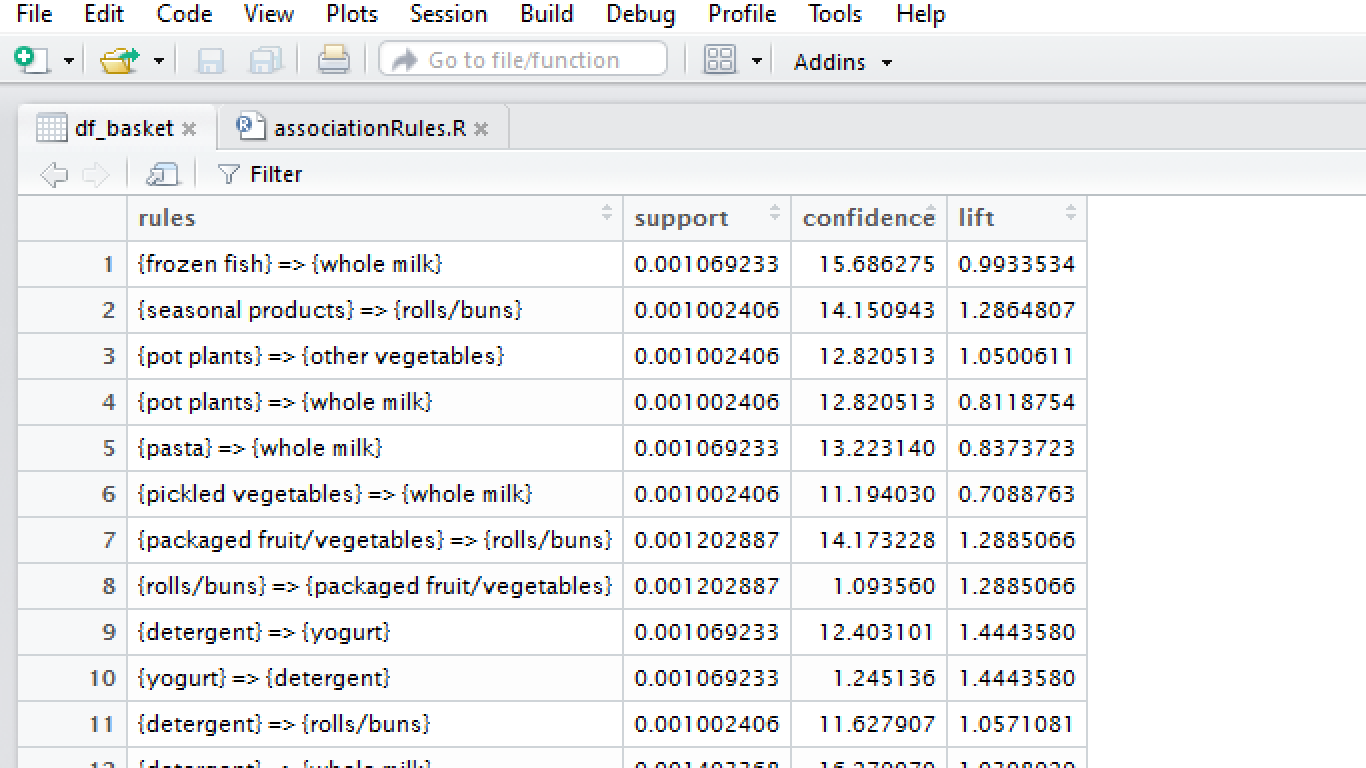


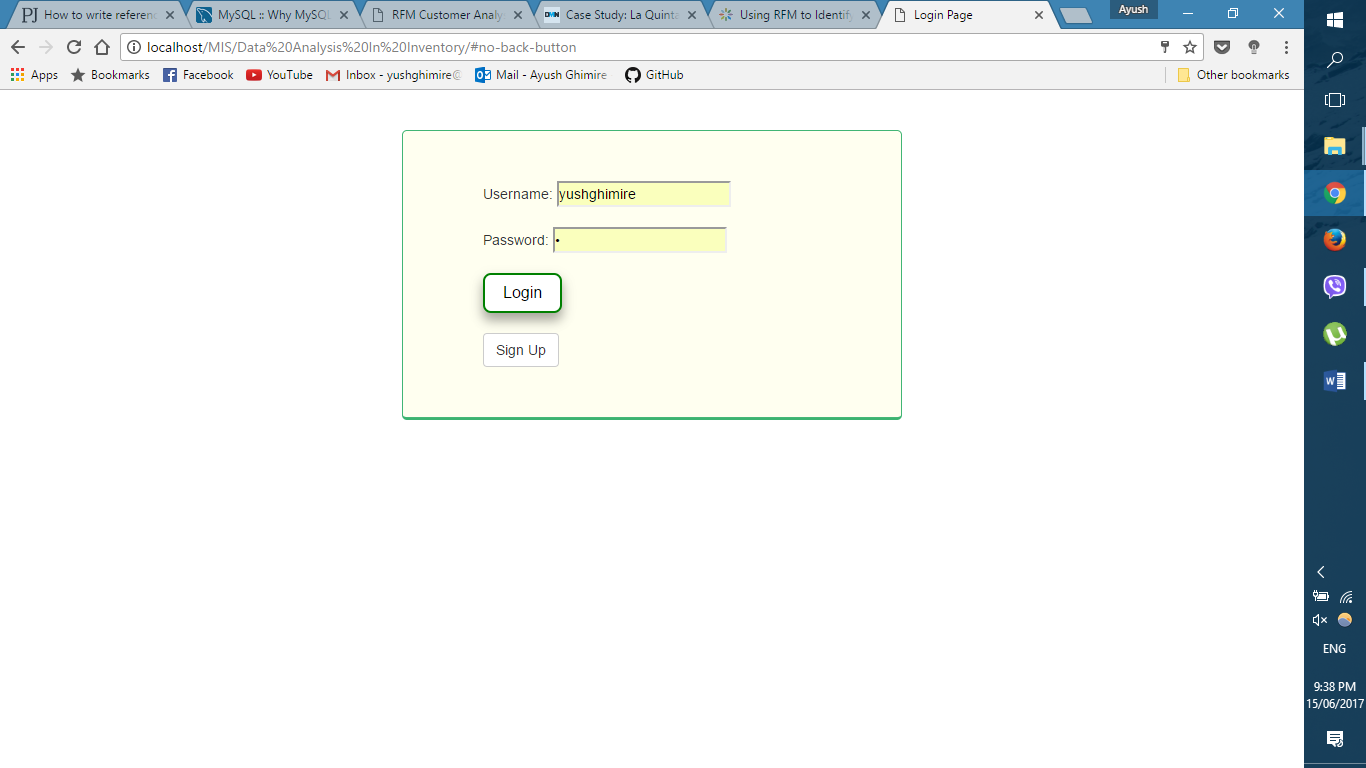
Figure 13 Association Rules extracted from the data set

Models like Random Forest, SVM, adaBoost are trained for classification. Present data set are divided into training set and testing set with 30 percent in training set and 70 percent in testing set. Model is then used to predict the response of the customer.

# Chapter 5. EPILOGUE

## 5.1 Result

**Inventory Management system:** Basic functionalities of Inventory Management system is incorporated in our system. With basic security feature like login, item searching viewing, auto ordering and booking. Our system’s MIS functionality is shown below:

Figure 7 Login for employee and customer

**Customer View:**

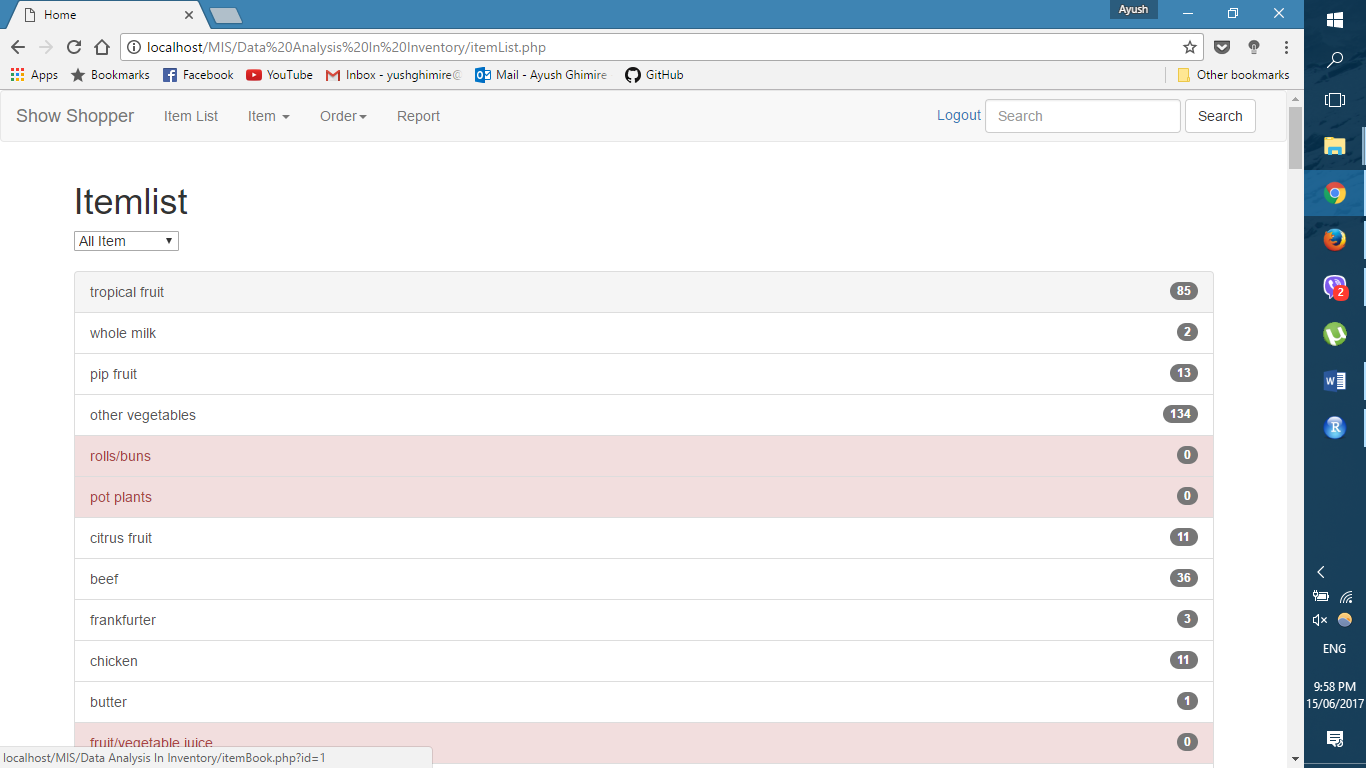


Figure 8 List of items

After the association rules are calculated customer were given recommendation for each item bought along with the item they have chosen. Below is the recommendation given to the customer who chose whole milk as the item in inventory.

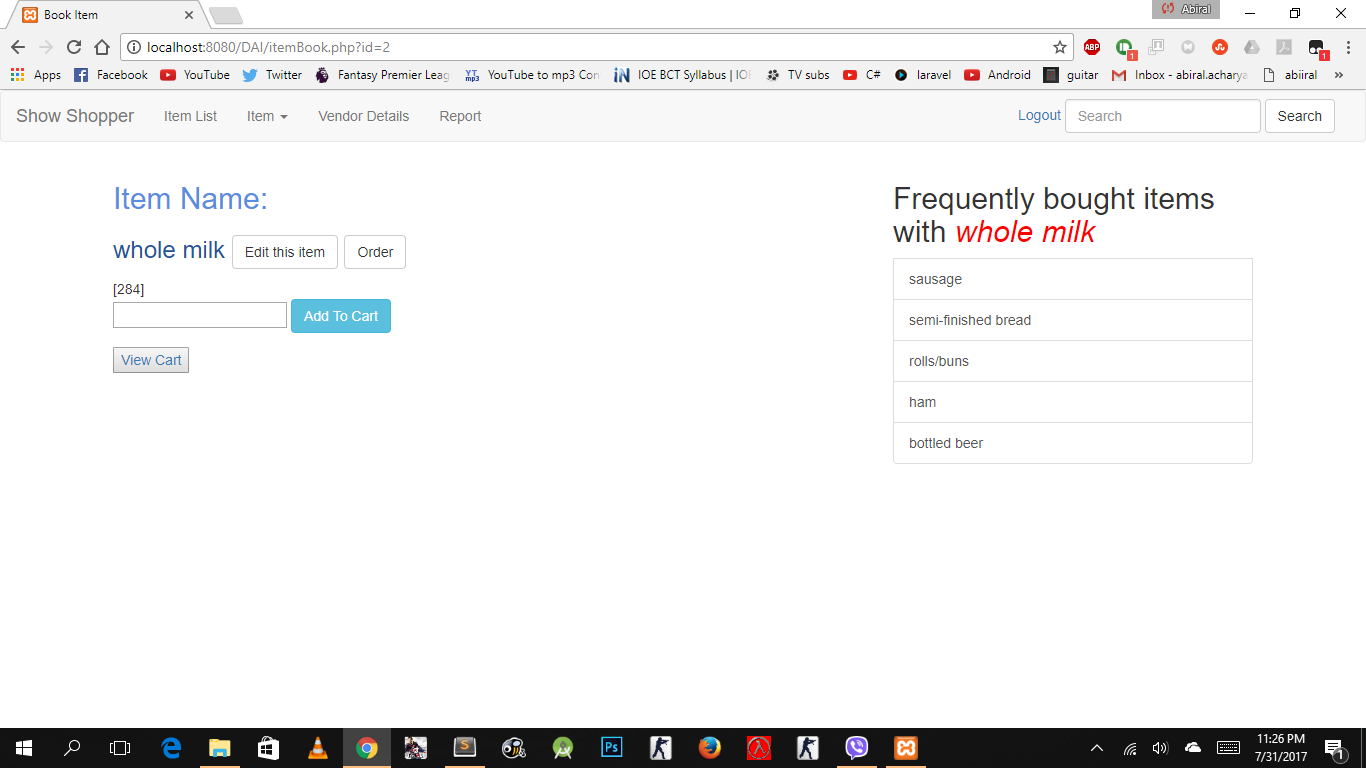


Figure Add item to cart.

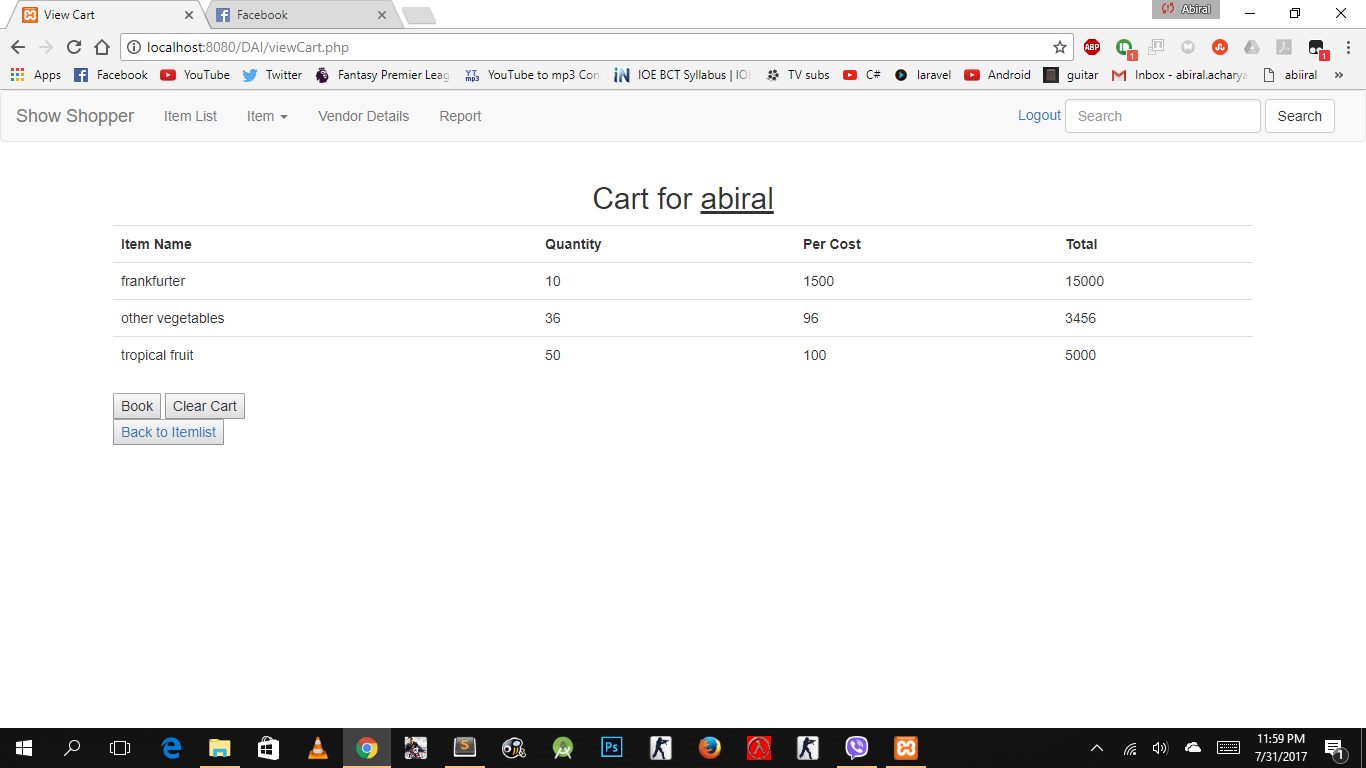


Figure Item added in cart

**Employee View:**

Employee will be able to add new item in the database, which include these fields: item name, cost price, selling price and quantity.

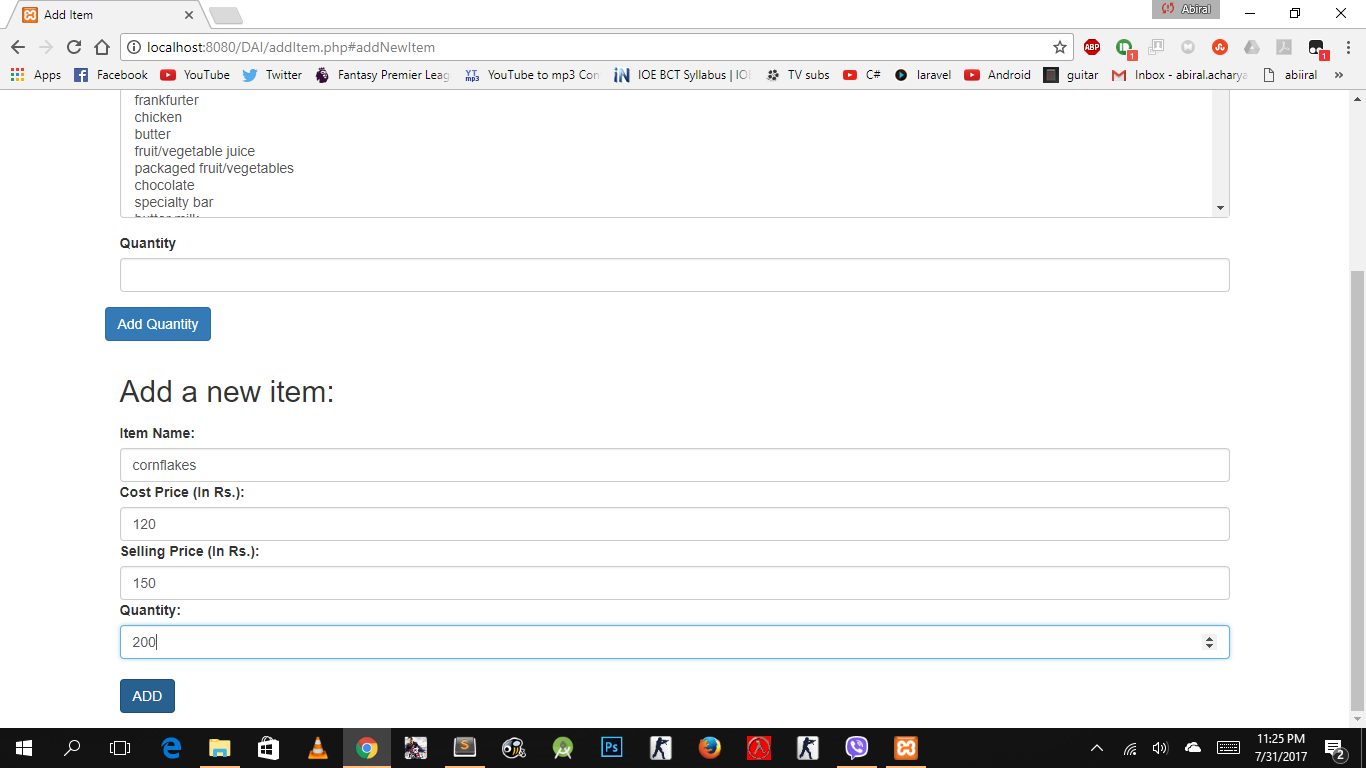


Figure Add new item

Employee will be able to change item details, updating quantity, cost price and selling price of the item.

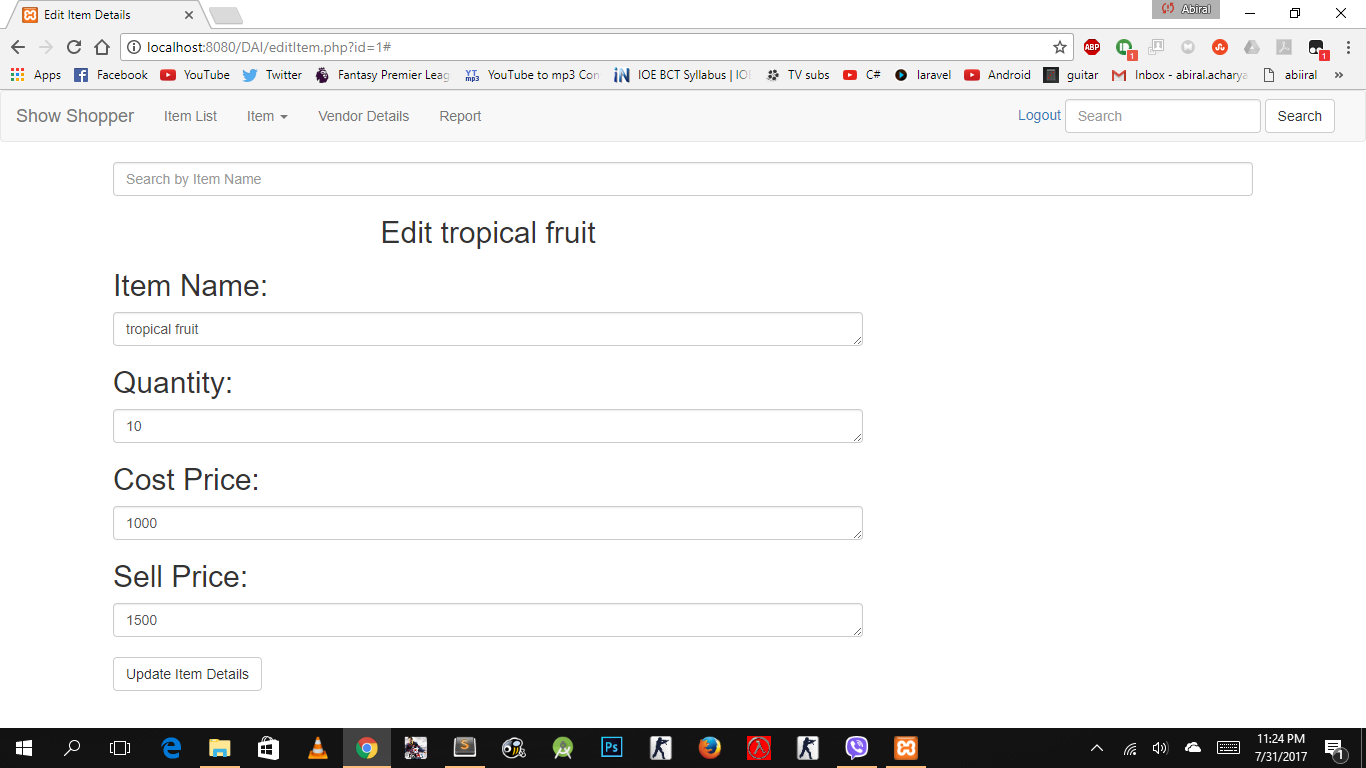


Figure Edit items

When the item runs low in stock employee are able to order the item, from the selected vendors

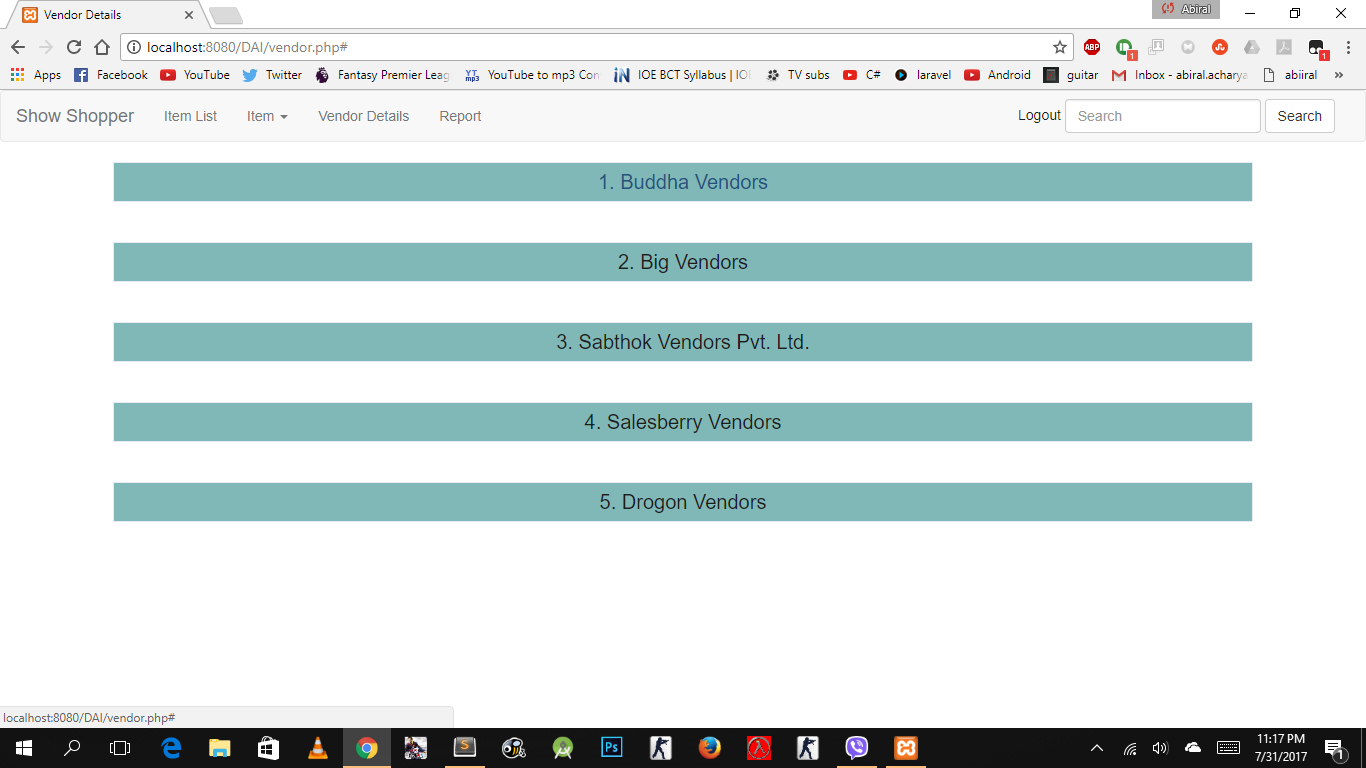


Figure Vendors List

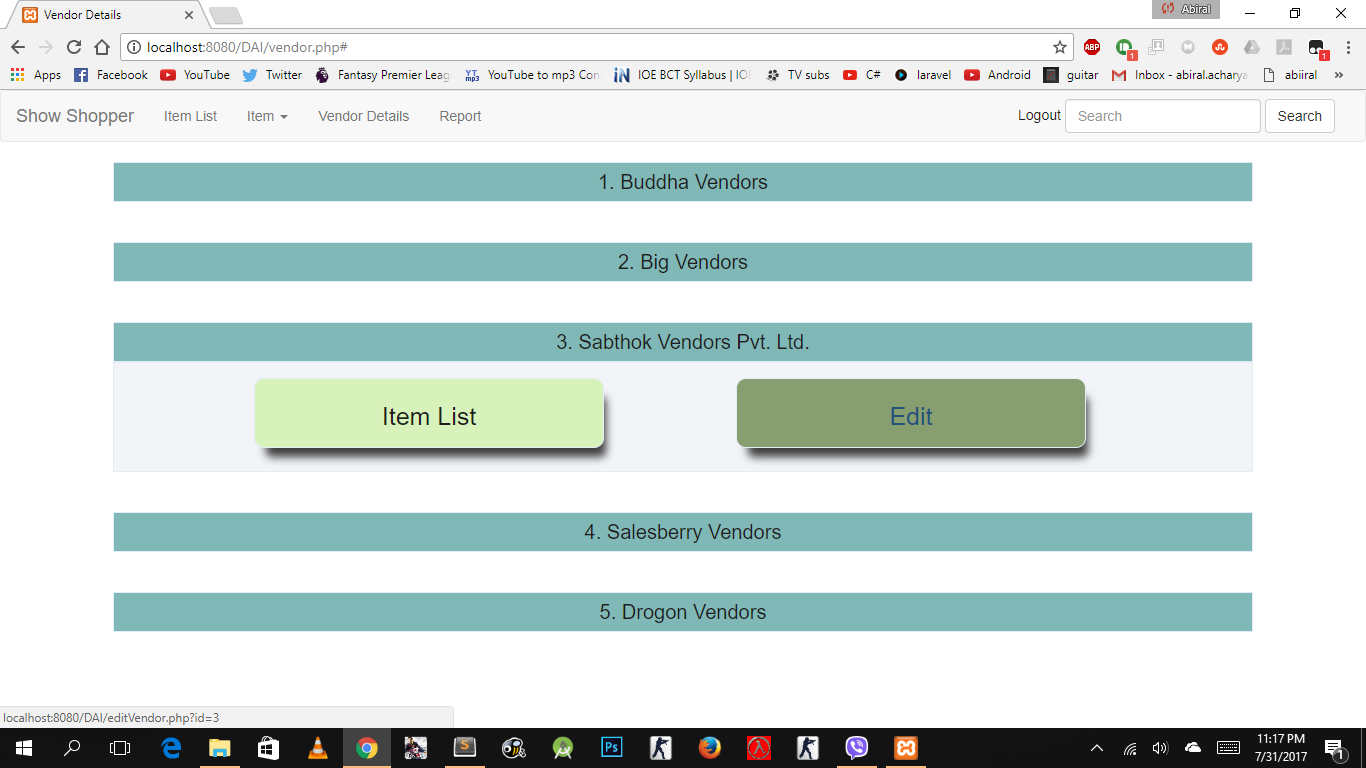


Figure Vendor option

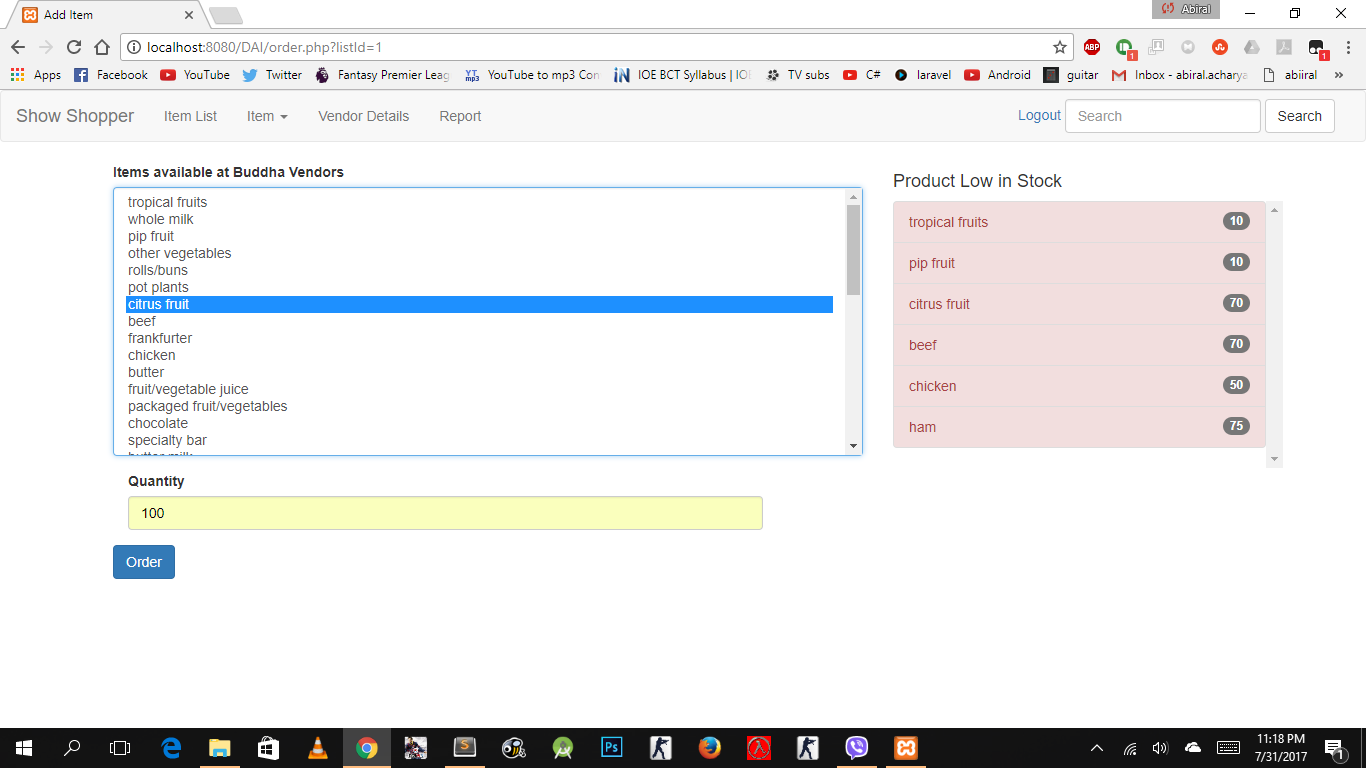


Figure Item order

The employee of the system is also given a visual representation for the combination of itemset, so that they can have the knowledge of association for scheme development. Below we can see the visualization of top 25 itemset in terms of lift.

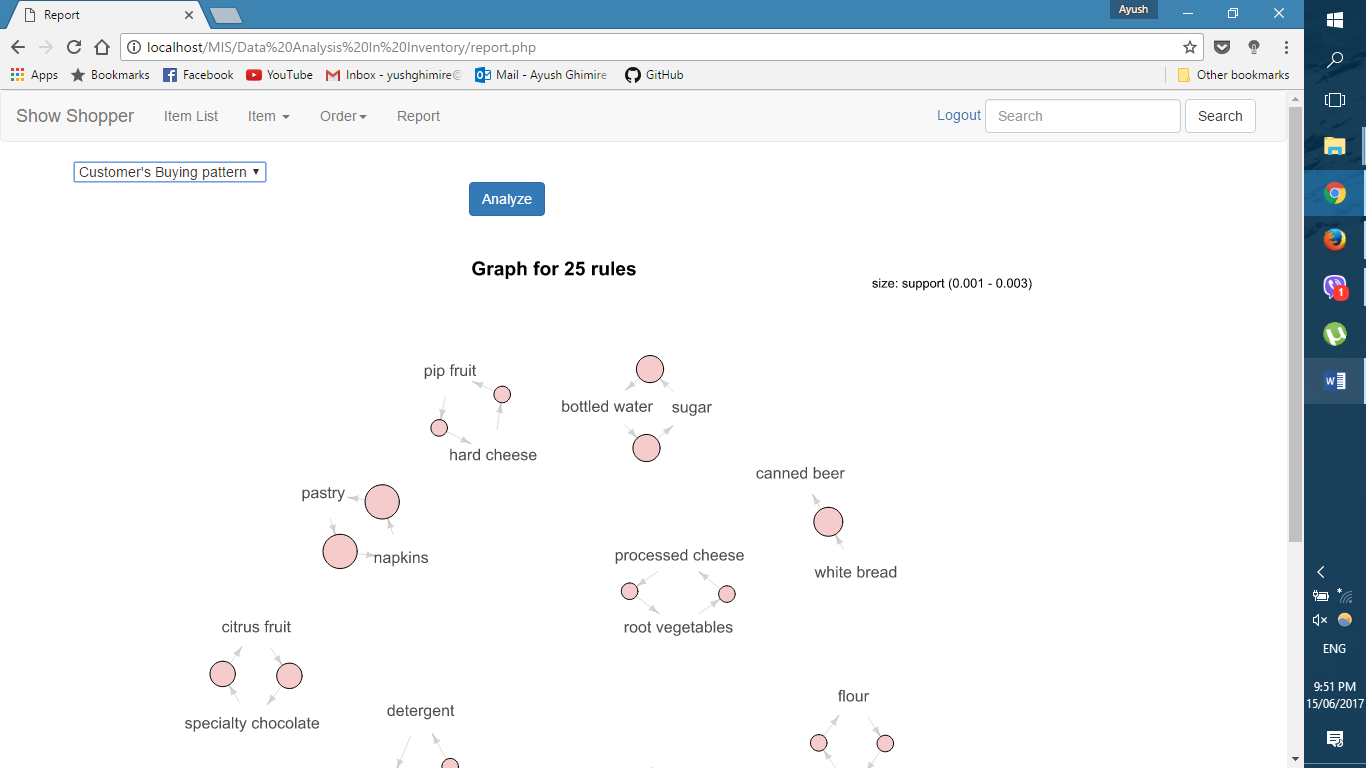


Figure 15 Visualization to the employee

Employee and Customer both can search for item using keywords Below is a search preform for ham.

## C:\Users\Abhaya\AppData\Local\Microsoft\Windows\INetCache\Content.Word\searchResult.png

Figure Search result

**Customer Data Analysis and It’s result:**

Customer behavior on the basis of Recency, Frequency and monetary data are calculated as shown below. This analysis helps understand user in case of marketing.

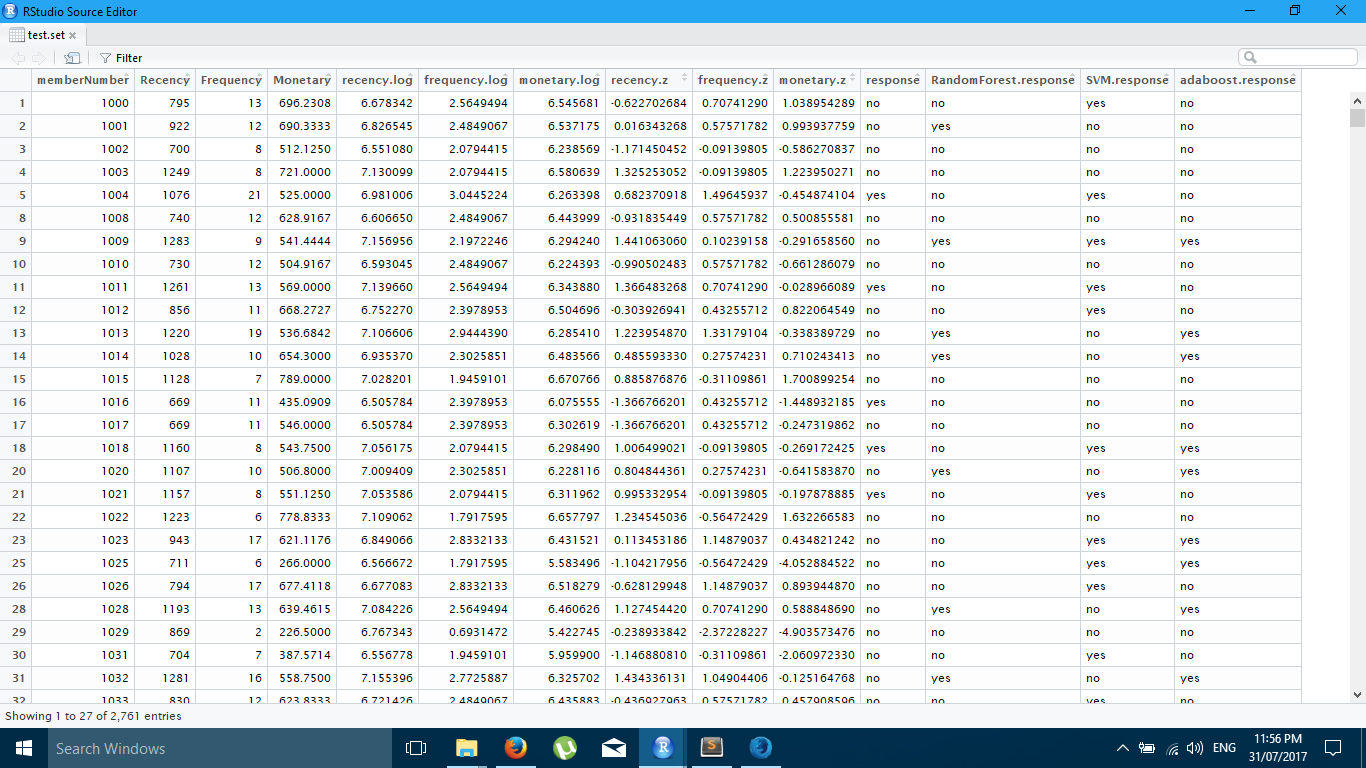


Figure RFM data analysis

After a model is trained using the train set, testing set is used to know the accuracy of our system. On performing predicitve analysis, accuracy percentage is calculated. The percentage specifies the effectiveness of using the algorithm to predicit the behaviour of customer. Random forest predicts with 70% of accuracy.

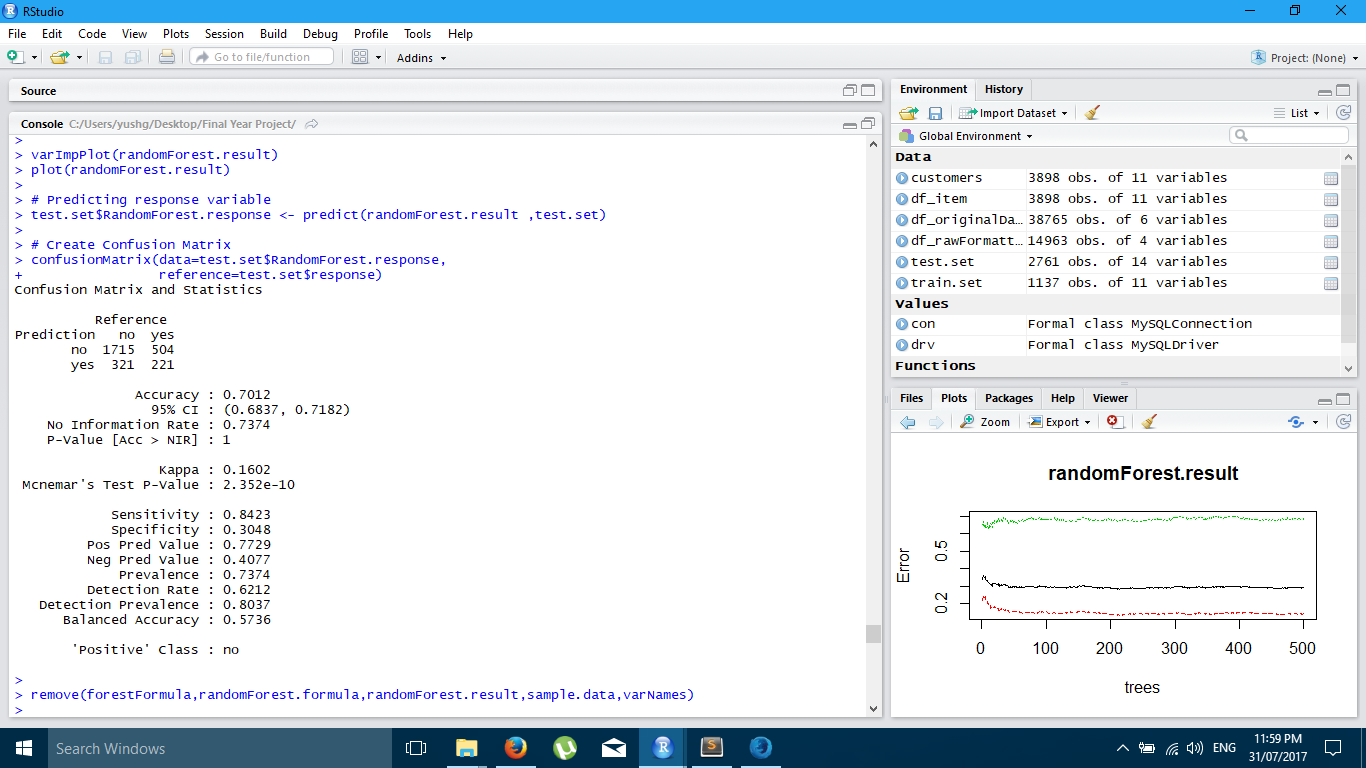


Figure Random Forest analysis

Support Vector Machine with 62.8% of accuracy.

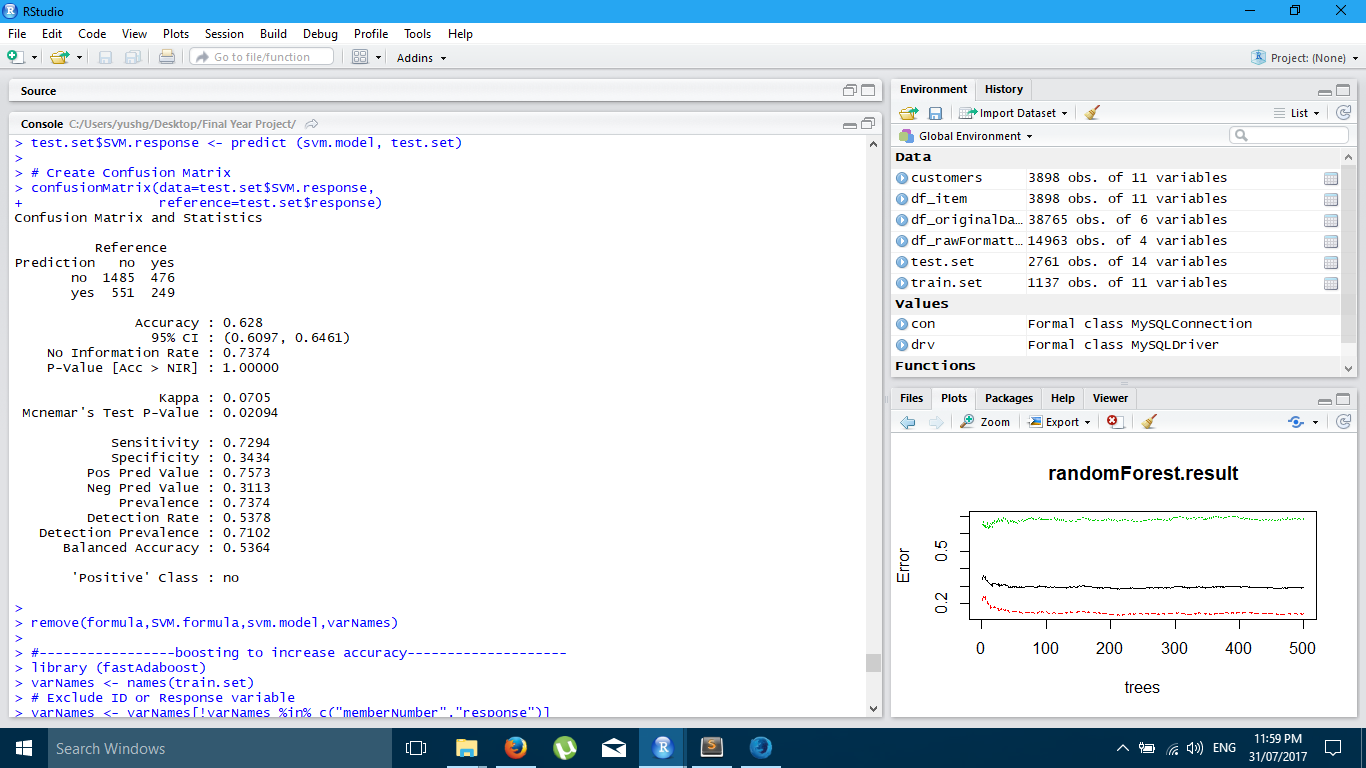


Figure Support Vector Machine

Fast AdaBoost with 67.7% of accuracy.

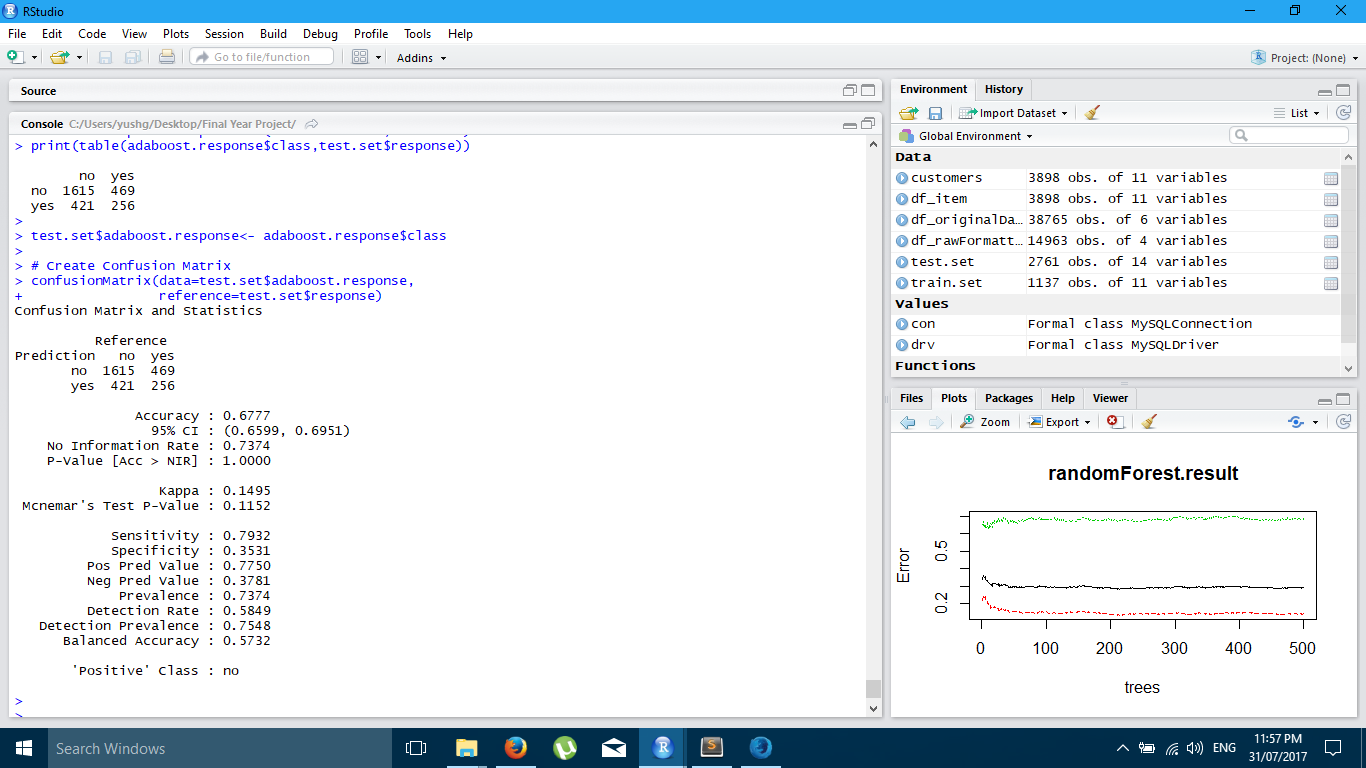


Figure Support Vector Machine

After finding the data predicted with each model where different in about 15 datasets, i.e. dataset that are wrongly predicted as No even though they are actual a Yes responder, we combined the result of all the model to increase the prediction of Yes. Result of all the system is ORed, this decrease the accuracy but minimize the wrongly prediction of customer who are likely to respond to our call.

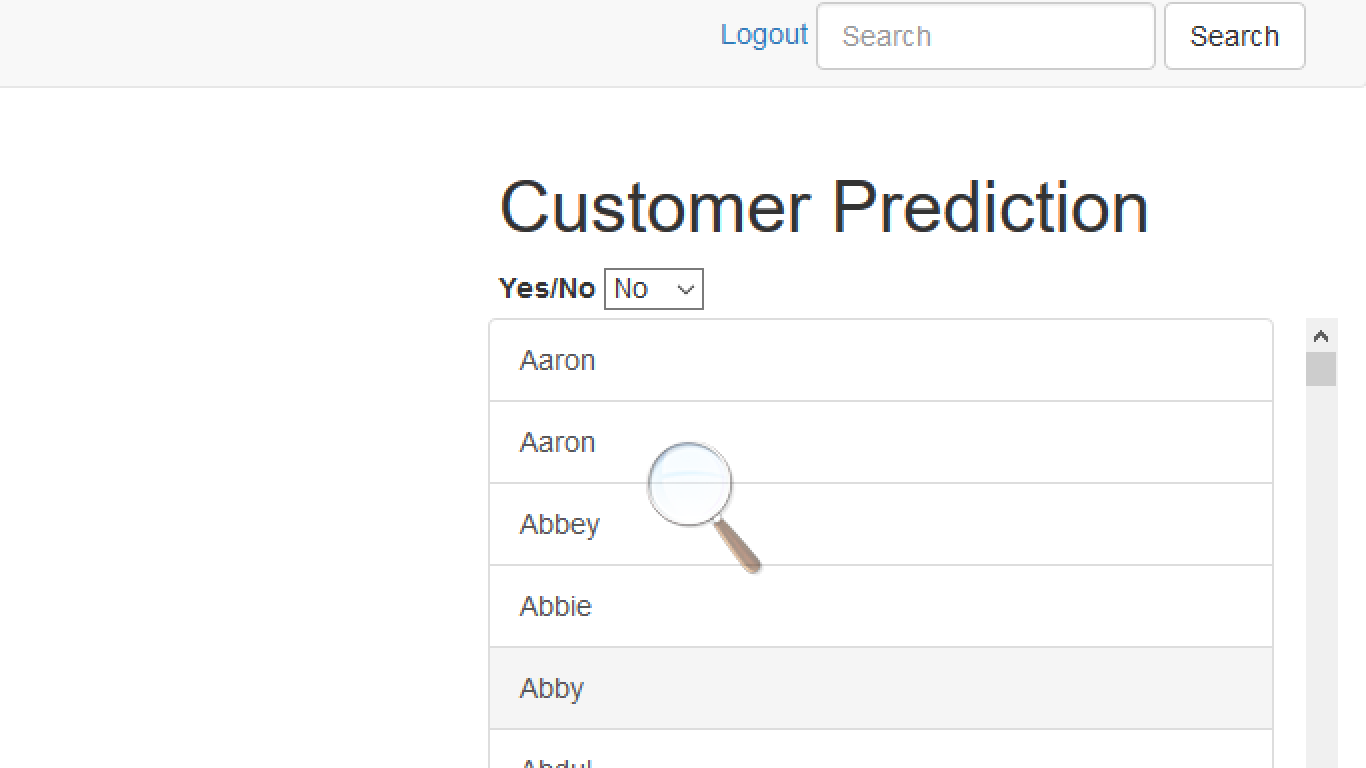
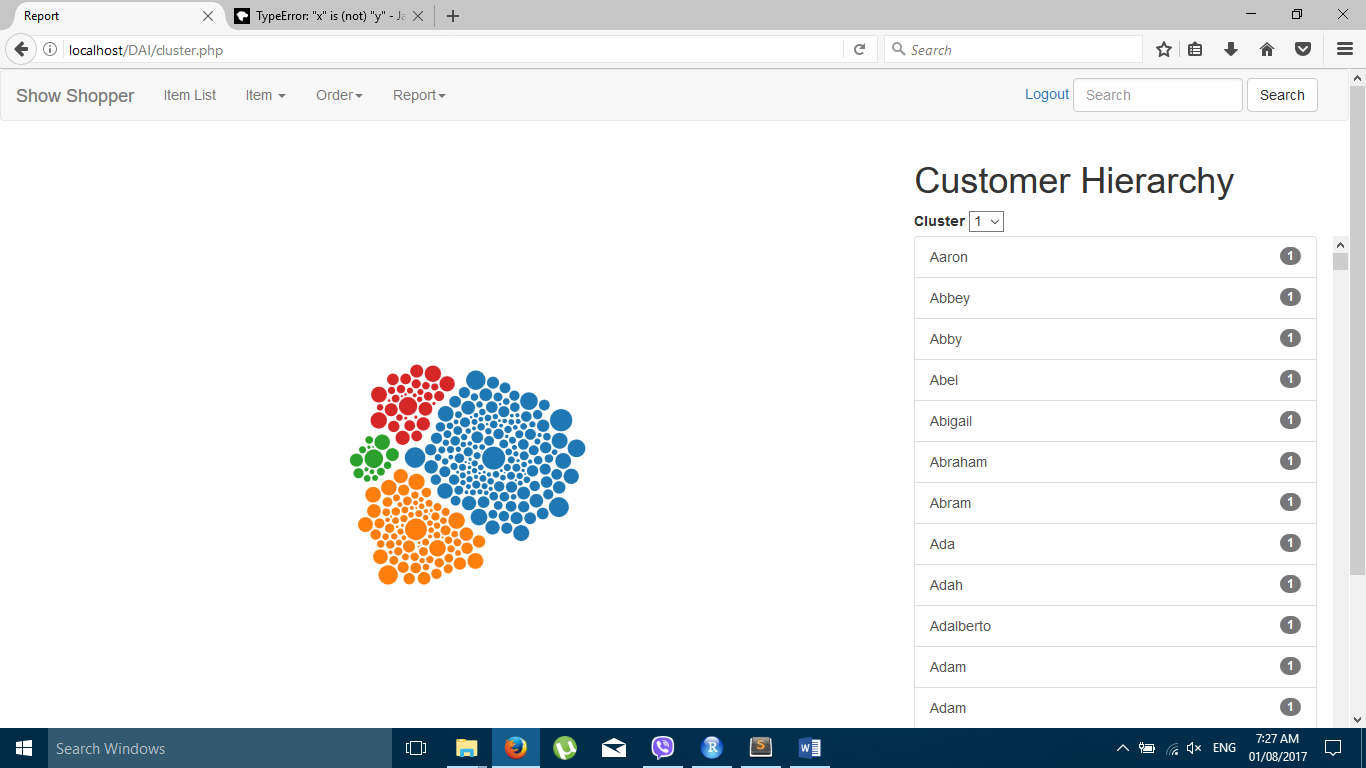


Figure List of customers predicted as non-Responder

As shown in above figure a list is presented to the user of the non-Responder customer, that helps in identifying the customers and minimizes the expenditure in market advertisement.



## 5.3 Problem Encountered

## 5.3 Gantt Chart

Figure 16 Gantt Chart for our development

# REFERENCE

[1] © Albion Research Ltd. 2017. *What is Market Basket Analysis?* [Online]. Available from: http://www.albionresearch.com/data\_mining/market\_basket.php [Accessed: 14th June, 2017].

[2] Joao Correia. *How RFM Analysis Boosts Sales*. [Online]. 2016 . Available from: http://www.blastam.com/blog/rfm-analysis-boosts-sales [Accessed: 14th June, 2017].

[3] © The R Foundation. *What is R?.*[Online]. Available from: https://www.r-project.org/about.html [Accessed: 13th March,2016].

[4] © 2001-2017 The PHP Group. *What is PHP?.* [Online]. Available from: http://php.net/manual/en/intro-whatis.php [Accessed:15th March,2016].

[5] Willis Wee. *FusionCharts: From Start-up To Serving Google, LinkedIn, and Barack Obama.* [Online]. 2011. Available from: https://www.techinasia.com/ fusioncharts [Accessed: 14th June 2017]

[6] © Oracle Corporation and/or its affiliates. *Why MySQL?.* [Online]. Avaliable from: https://www.mysql.com/why-mysql/ [ Accessed: 27th March 2016].

[7] Jack Han. *RFM Customer Analysis with R Language.* [Online]. Available from: http://www.dataapple.net/?p=84 [Accessed: 13th June 2017]

[8] © GainInsights Solutions. *Customer Segmentation Using RFM Analysis*. [Online]. 2014. Available from: http://gain-insights.com/solutions/retail-analytics/ customer-segmentation-using-rfm-analysis/ [Accessed: 14th June 2017]

[9] John Tukey. *The Future of Data Analysis*. [Online]. Princeton University. July 1961. Available from: http://projecteuclid.org/download/pdf\_1/euclid.aoms/ 1177704711 [Accessed: 25th March 2016].

[10] Richard A. Becker, John M. Chambers and Allan R. Wilks. *The New S Language*. New York . Chapman & Hall. 1988. This book is often called the “Blue Book”.

[11] John M. Chambers and Trevor J. Hastie eds. *Statistical Models in S*. New York. Chapman & Hall. 1992 This is also called the “White Book”.

[12] John M. Chambers. *Programming with Data*. New York. . Springer . 1998 This is also called the “Green Book”.

[13] Michael Hahsler, Christian Buchta, Bettina Gruen, Kurt Hornik,Ian Johnson, Christian Borgelt. *Mining Association Rules and Frequent Itemsets*.[Online]. 2017 Available from: http://mhahsler.github.io/arules/, http://lyle.smu.edu/IDA/arules [Acessed: 14th June 2017]

[14] Pazaras Christos*. DATA PREPARATION AND PREPROCESSING*

*FOR DATA MINING USING R.*  Alexander Technological Educational Institute of Thessaloniki. 2013.

[15] Sridhar Mutyala. *Using RFM to Identify Your Best Customers.*[Online]. 2017. Available from: http://www.eightleaves.com/2011/01/using-rfm-to-identify-your-best-customers [Accessed: 14th June 2017]

# BIBLIOGRAPHY

* TechTarget. *Data.* [Online]. Available from: http://searchdatamanagement. techtarget.com/definition/data. [Accessed: 1st April 2016].
* © Retalon 2016. *PREDICTIVE ANALYTICS TRANSFORMS INVENTORY MANAGEMENT IN RETAIL.* [Online]. Available from: http://retalon.com /news-updates/predictive-analytics-transforms-inventory-management-in-retail [Accessed Date: 14th June 2017].
* w3schools.com. *PHP 5 Tutorial.* [Online]. Available from: https:// www.w3schools.com/php/default.asp [Accessed: 23th March 2016].
* Thinking inside the box. *#7: C++14, R and Travis — A useful hack.* [Online]. Available from: https://www.r-bloggers.com/ [Accessed: 24th May 2017].
* Anish. *RFM Analysis For Successful Customer Segmentation.* [Online]. 2017. Available from: http://www.putler.com/rfm-analysis/ [Accessed: 14th June 2017]
* © 2013-2017 MastersInDataScience.org. *Data Science in Retail*. [Online]. Available from: http://www.mastersindatascience.org/industry/retail/ [Accessed: 1 April 2016].
* © Spotless Data Ltd. *EXPLORING DATA ANALYSIS.* [Online]. Available from: https://spotlessdata.com/blog/exploring-data-analysis [Accessed: 10th April 2016].