Supply Chain Optimization & Category Management : Empirical Study on an Online Grocer

Mahima Gupta
Institute of Management Technology, Ghaziabad

&

Jibin P Rajan; Milan P Panackal &Susan Thomas
Great Lakes Institute of Management, Chennai

Abstract: The focus of the research is on the lean business model that relies on a convenient mix of warehousing and cash-and-carry. How to reduce the e-grocer's dependence on capital-intensive operations and enable it to put in place a thoroughly efficient supply chain? As competition abounds in online grocery retail space, controlling cash-burning is very crucial when it comes to inventory, supply chain and logistics. The research aims at application of a model to forecast the quantity required based on the demand of the product and the lifespan of perishable goods to assure quality of products. The research also aims at collecting real data from an e-grocer to study their operations and develop optimization model to add value to their operations. The study also aims to perform analysis based on historic sales data to develop insights for category management.

Keywords : Online Grocery, Supply Chain, Retail Space

Introduction

Online grocery is a thriving business model in many developed countries like France, Japan, Switzerland, UK, US etc. Now, the market is set to become an exciting space and undergo exponential growth in India with a large portion of Indian population coming online. The online grocery business is set to grow at a rate of 25-30 per cent year-on-year in major Indian cities. Researchers predict that by 2016, the Indian grocery market would surpass Japan to become the 3rd largest market worldwide. India's e-grocery market, which is estimated to be less than \$100 million at present, is expected to cross \$25 billion by 2020. Many industry players and observers believe that the only way for grocery companies to survive in future is to become part of a highly integrated supply chain that is efficient and responsive to customer needs. Three e-grocery business models have been identified during the course of this study - the Inventory model, the Market-place model and the hybrid model.

Inventory Model

The Inventory business model (DUVAL) involves the delivery of everything your local supermarket offers. Customers order and often pay online. The order is then typically processed and shipped from a highly-automated warehouse located near the thickly populated urban area that the e-grocer serves. The order is delivered to the customer’s home by the e-grocer’s own fleet of refrigerated trucks. Online prices are competitive with traditional supermarket prices, and delivery charges are low. This business model needs strong financial support to build and stock the warehouses and purchase a fleet of delivery trucks. It has been implemented by leading e-grocers such as Big Basket. This model is distinguished by high fixed costs that require online grocers to build quickly a large customer base through expensive promotion and advertising programs.

Market Place Model

An online food retailer adopting a Market Place model is little more than an agent maintaining a website. Customers order via the e-grocer website, but the e-grocer does not generally own or store the products that they sell. All orders are automatically forwarded to wholesaler/retailer who take care of shipping the requested product to the customer. The Market Place model is a popular model because of its minimal start-up capital and human resources requirements. For example, Aaramshop.com and peppertap.com has implemented this model and is only responsible for managing the website and forwarding email and other electronic orders to wholesalers. In addition, its market place model allows it flexibility in terms of selecting and changing products to be sold on the site.

Hybrid Model

Hybrid model is characterized by the sale of groceries both online and offline, which is typical of brick-and-mortar retailers who want to secure a share of the online retail market and increase offline customer’s satisfaction by offering them a choice of online ordering and home delivery services. Reliancefreshdirect.com follows a Hybrid model.

The major challenge in an e-grocery supply chain is the perishable nature of SKUs. The normal optimization or cost effective models, which assumes infinite life of SKUs, wouldn’t be effective in case of e-grocery which sells goods like Fruits & vegetables, staples, dairy products and beverages.

Literature Review

Supply Chain Optimization

Researchers have been trying to solve the inventory problem of perishable goods. Previous work in the field of decaying inventory problems began primarily with that of Schrader and Ghare (Ghare, 1963). Back then a regular demand function was assumed and shortages or stock outs were not considered. Their work highlighted the importance of considering inventory decay and demonstrated the cost savings which could be achieved by careful inventory analysis.

Kishore R A (Kishore R A, 2011) showed that inventory problem of perishable goods can be solved by dynamic programming, if it can be broken into a number of stages and if decision can be taken stage-wise using principle of optimality. Since the life span of a perishable item is limited and its demand is not always fixed, the entire inventory period can be divided into a number of stages (small inventory periods) in which the product does not deteriorate and its demand remains almost constant. Replenishment time is generally small, therefore demand is assumed to be instantaneous at the beginning of each period and no shortage is allowed. The demand for a product may or may not be deterministic. Inventory model is likely to manage inventory of the perishable products and minimizing the total cost function in any particular inventory period of the time horizon.

A model, (Elsayed E. A, 1983) Was developed that has no real deterioration factor and allows no shortages. The total cost per week is described by

$$TC(Q) = A + CQ + hTI$$

Where A denotes the set-up cost of a production run; CQ denotes the variable cost of production; hTI the holding cost per week. Using the average weekly total cost and expressions for optimum production quantity, optimum inventory quantity and optimum cycle times are found.

Research by Morris A. Cohen (Kahn, 1975)showed the existence of an optimal order policy established for an inventory system in which fresh supply, no stock return, an m-period lifetime and FIFO issuing are identified. The components of the cost structure in their inventory are Shortage, Holding, Ordering and Outdate (wastage). Amount to be ordered is a function of age distribution of inventory on hand. This is done by taking into account the perishability and the holding cost that will be incurred as part of online grocery supply network and the cost incurred on them can be determined. By knowing the quantity and cost we can make profitable changes in the operations of the grocery chain and thereby gain the margin over the competitors.

Anna Corinna Cagliano (Anna Corinna Cagliano, 2015) has shown the application of System Dynamics (SD) to e-Grocery Supply Chain. Since the approach developed in this work is intended to assess the Supply Chain structure related to inventory, orders, and deliveries. SD proves to be useful for this purpose. SD is a modeling and simulation approach aimed at interpreting the behavior of a complex system to support policy design. This methodology enables us to graphically denote a system of interrelated stock, flow, and auxiliary variables, define the mathematical equations illustrating the relationships among them, and perform a computer-based simulation to determine the trends of the investigated variables over a preset period of time. Model validation is carried out through historical data and sensitivity analysis. To capture the interrelation between Supply Chain responsiveness and efficiency SD has been applied. It is also used to study the effects of different strategies and business model to improve them. Responsiveness to demand variability has also been tackled with the purpose of showing the effects of product availability on customer demand and proving the reinforcing feedback between demand fluctuation and the consequent adjustment of the production capacity. Furthermore, SD models have been developed to examine instabilities in SC due to actions addressing the imbalances between supply and demand like price changes, promotions, and the involvement of additional. SD proved to be beneficial in Supply Chain reengineering and to characterise the conditions under which the bullwhip effect can occur. SD models have been extensively used to evaluate the operational and economic performance of Supply Chains. M.Ferguson (Ferguson M. K., 2006) has highlighted the benefits of information sharing to a distribution centre (DC) that sells a perishable product with a fixed lifetime and is constrained to order in fixed lot sizes. It described exact policies for estimating the optimal batch size multiple at the DC for each time period and inventory state. The research proposed heuristic policies for the DC under both no information sharing and information sharing of the age of the inventory arriving upon replenishment. The heuristics are then used to measure the value of information under a wide range of parameter value settings. It found that the retailer benefits the most from information sharing when: (1) the variability of either demand or the remaining lifetime of replenished items is high, (2) product lifetimes are short, and (3) the cost of the product is high. The research also showed that information sharing is generally more beneficial when demand is satisfied with a FIFO issuing policy than with a LIFO issuing policy. A subsequent research (Ferguson M. J., 2007) applies the extension of the EOQ model for nonlinear holding cost by Weiss (Weiss, 1982) to the inventory management of perishables in small to medium–sized grocery stores, where managers frequently utilize markdowns to stabilize demand as the product's expiration date nears. The model provides significant improvement in cost vis-à-vis the classic EOQ model, with a median improvement of 40%, ranging between1% and 401%.

Another research (Ketzenberg, 2000) explores the value of information (VOI) for inventory replenishment of a perishable product. VOI is measured as the improvement in profit contribution that a facility is able to achieve using shared information in replenishment as compared to the classic case in which information is not shared. Beyond product perishability, essential characteristics of the problem include stochastic demand, lost sales, and ordering restrictions in the form of case-packs. They modeled a three echelon arbores cent supply chain consisting of multiple, interdependent facilities. There were several implications. First, VOI can be considerable for perishable products. The level of VOI that facilities realize depends on several factors. The three most significant determinants are the product lifetime, the level of demand uncertainty a facility experiences, and a facility's cost structure. VOI increases with the perishability of the product. Retailers generally experience the highest VOI even though they do not use information directly in their own replenishment. VOI for the manufacturer is lowest since it resides at the top of the supply chain. Since the manufacturer is ultimately responsible for the timing of replenishment into the system, it has the greatest impact on product freshness for downstream facilities and consumers. Distributors are able to achieve a moderate level of VOI and their use of information can impact both upstream and downstream facilities. They can use information to improve product freshness for retailers and thereby reduce retailer outdating. Another contribution of the research is the formulation of the information case and classic case as Markov Decision Processes. With optimal policies we not only provide exact treatment for VOI using our modeling assumptions, but also use them to benchmark simpler, easy-to-compute policies. The adoption of RFID technology (Edmund Prater, 2005) and its attendant supply chain management techniques will add to the VOI across all the echelons.

The availability and synchronization of both cost and production data provides valuable information to managers because the behavior of production systems directly affects costs, and costs directly affect operational decisions. The integrated cost model supported by an APS-based real-time order management system provides a methodology for the order planning decision that captures the physical constraints in the operational functions as well as the expenses associated with the consumption of those resources throughout the production chain. A feasible model based on current ERP and APS technology can be implemented in practice to help firms improve their profitability in the order acceptance process. (Srivastava, 2009)

Category Management

A case analysis (Taariq Lewis, 2009) attempted to categorize groups of products into distinct product performance groups from transaction data. Those groups are analyzed for distinct characteristics and then tested whether new product performance can be forecasted based on the identified group characteristics.

Another study by Puneet (Puneet Manchanda, 1999) developed a model for Multicategory purchase incidence decisions. The choice of one category may affect the selection of another category due to the complementary nature. It can also be sue to similar purchase cycles (e.g. beer and diapers) or because of a host of other unobserved factors. The paper says about a development of a conceptual framework that incorporates complementarity, co-incidence and heterogeneity.

Wagner A. Kamakura (Kamakura, 2012) introduced a new approach called Sequential market basket analysis. The practice of market basket analysis has its origin in the data-mining literature, with the introduction of association-rule discovery. By far, the most common practice in market basket analysis is the identification of association-rules (Brijs, 2004). Each pair of products A and B is evaluated on three measures: Support—the joint probability of finding the pair AB across all baskets. A low support means that the pair is not relevant because it is not purchased frequently enough.

Confidence—the conditional probability $\rho(B|A)=\rho(A\eta B)/\rho(A)$ which is often interpreted as the probability that purchase of A will lead to purchase of B.

Interest—the ratio between the joint probability and the probability of joint occurrence under independence $\frac{p(A \cap B)}{p(A) \times p(B)}$. This measure discounts the joint probability by the "popularity" ofthe individual items in all baskets.

Product B is a good recommendation for shoppers who just added A to their basket if interest and support is high (i.e., they tend to occur jointly in most baskets), and confidence, the conditional probability of purchasing B given A, is also high.

Hypothesis

1. Given access to ecommerce transaction data over a discrete time period and location details of each transaction, one can identify the current cost structure and provide suggestions for improved benefit.

2. Given access to ecommerce transaction data over a discrete time period and a discrete catalogue of products, the categories can be divided into logical clusters based on the sales pattern

3. Given access to ecommerce transaction data over a discrete time period and a discrete catalogue of products, one can apply an association rule algorithm to identify purchasing behaviour.

Method

As part of the empirical research, we have collaborated with a small scale e-grocery to identify the issues in the supply chain. The company will be called 'ABC e-grocer' (Actual name not disclosed as per the confidentiality agreement) hereafter. The aim of the empirical investigation is to get information about ABC e-grocery's supply chain which includes information about its suppliers, procurement, inventory, and logistics and hence find a means to optimize their supply chain model. Also, as the business is in initial phase we would also attempt to manage the product portfolio or category. The following question was given priority in our analysis:

- What are the possible areas of improvement in the supply chain of ABC e-grocer?
- How can we manage several categories of the products listed in the ecommerce website of the ABC e-grocer?

Introduction to ABC e-grocer

ABC e-grocer is an online grocery startup based in Chennai. They were operational from January 2014. The marketing is outsourced to a consultancy. They use mostly digital marketing through Facebook, WhatsApp etc. They have around 450 customers with approximately 5 to 10 deliveries per normal day and approx. 10 to 12 deliveries in peak days. They follow a hybrid model of operations and have a manpower capacity of 2 people for procurement, sorting, packing and loading and 1 person for delivery. 50% of their customer base is in the proximity of the warehouse. They have 1 vehicle which delivers to the entire customer. They have two types of suppliers – Wholesalers and Super markets. All the items in the inventory are purchased from wholesalers.

They are mostly non-perishable items. Other items which are perishable & not available in inventory are purchased Just-in-time from nearby super markets which they have already tied up with. This reduces the wastage and helps in customer retention.

Data Collection

Interviews

The type of research relevant for this empirical study is descriptive research, as descriptive research involves the collection of data in order to test hypotheses or to answer questions concerning the current status of the subjects of the study(here the issues concerning the supply chain of ABC e-grocer). For the descriptive analysis we have decided to focus on the stakeholders of ABC e-grocer. ABC owners/employees/suppliers will be asked the question of "Where do you see the top five weaknesses or improvement areas in the supply chain of the company? "Interviewees will thus asked to list top five improvement areas, and further prioritize by criticality, and evaluate each improvement area for estimated cost for required changes and implementation difficulty.

Ecommerce System Data & Other documents

The ecommerce system used by the ABC e-grocer has all the data required for trend analysis, statistical analysis. The ecommerce system has many inbuilt functions which aids the analysis on further improvement in operations, marketing and financial data. A typical ecommerce system will have following functions:

- Registration
- Basket
- Payment
- Product management
- Orders management
- VAT and shipping costs

Data is captured in each stage and is available for analysis. The data, in skilled hands, will prove to be very useful for the future prospects of the e-commerce/e-grocery company. The system has inbuilt modules to generate reports on the data. But for supply chain optimization, merely looking at the reports wouldn't add much value. We will have to analyze the raw data to come up with optimization techniques and areas of improvement. The following data were requested from the ABC e-grocer:

- Customer data (Name, Address details)
- Sales Data
- Products Portfolio
- Supplier data
- Inventory data
- Financial Statements

After so much persuasion and signing of a Non-disclosure agreement, we have come to a consensus. The e-grocer has agreed to share the transaction data from their ecommerce system.

Preparing data

The raw data was received in an unstructured format. There were six sheets, each corresponding to one database table. The sheets contained the following data:

1. Customer details (Customer_ID)
2. Order details (Order_ ID)
3. Order-Product details (Order Product_ID)
4. Product Details (Product_ ID)
5. Product-Category details (Product_ ID, Category_ ID)
6. Category details (Category_ ID)

The sheets had to be merged to form a single dataset containing all details. The sheets were merged using Microsoft Query into a new sheet. The merged sheet had the following format :

order_id	order_product_id	product_id	product_desc	Category_desc	quantity	price	item_total	customer_id	shipping_postcode	order_total	Date	Time
468	4667	303040106	Whole Wheat Atta	Atta	1	358	358	186	600097	871	6/7/2015	12:21

The resulting data set consisted of 6823 unique rows representing 555 unique transactions. The data ranged from Oct 2014 to Dec 2015 (15 months). But, as the company is in the process of picking up customers, there is very little data in Oct, Nov 2014. We had to remove the data to avoid skewness.

Cost-Benefit Analysis

In the data provided the order destination and the order quantity where provided. In order to optimize the logistics we decided to compare the cost by allocating the top 10 order location as the source and thereby calculating the cost benefit of shifting the base. We decided to use R to find the distance between the pin codes. There we 81 different order location and all were taken into consideration to find the apt location of warehouse.

Cluster Analysis

The objective here was to understand how pervasive are distribution similarity across different groups of products. We used a non-hierarchical approach to clustering knows as K-Means. The k-means approach to clustering "divides a sample into a predetermined number k of non-overlapping clusters so that clusters are as homogenous as possible with respect to the measurements used". Pre-determining the number of k clusters into which we partition the sample was done manually.

The data for twelve months (Jan 2015 – Dec 2015) was taken for this analysis. There were a total of 1509 different products across 245 categories sold in the 12 month period. Based on initial analysis, we found that some of the categories were only purchased only once or twice in a year. So, we had to remove those data to create logical clusters. The resulting data set contained 156 categories which has been ordered at least four times in the year. A new dataset with monthly sales of these categories was created. Monthly sales were normalized across the total of 12 months so that the cumulative sales patterns of various categories could be examined as groups. In this way the volume of sales wouldn't be a critical factor in clustering.

Market Basket Analysis

Market Basket Analysis (MBA) attempts to analyze customer ordering information to create a set of inferences on ordering behavior. For retail merchandizing this information is very important to decide on the physical grouping of product categories in stacks. In case of online retailing, MBA is used to generate associations. Based on association rule mining on historic transaction data, categories which are added to the cart together are found and used for recommendation. Part of the significance of market basket analysis is focus. MBA is used to locate good "rules". It is desirable to locate rules that are not trivial.

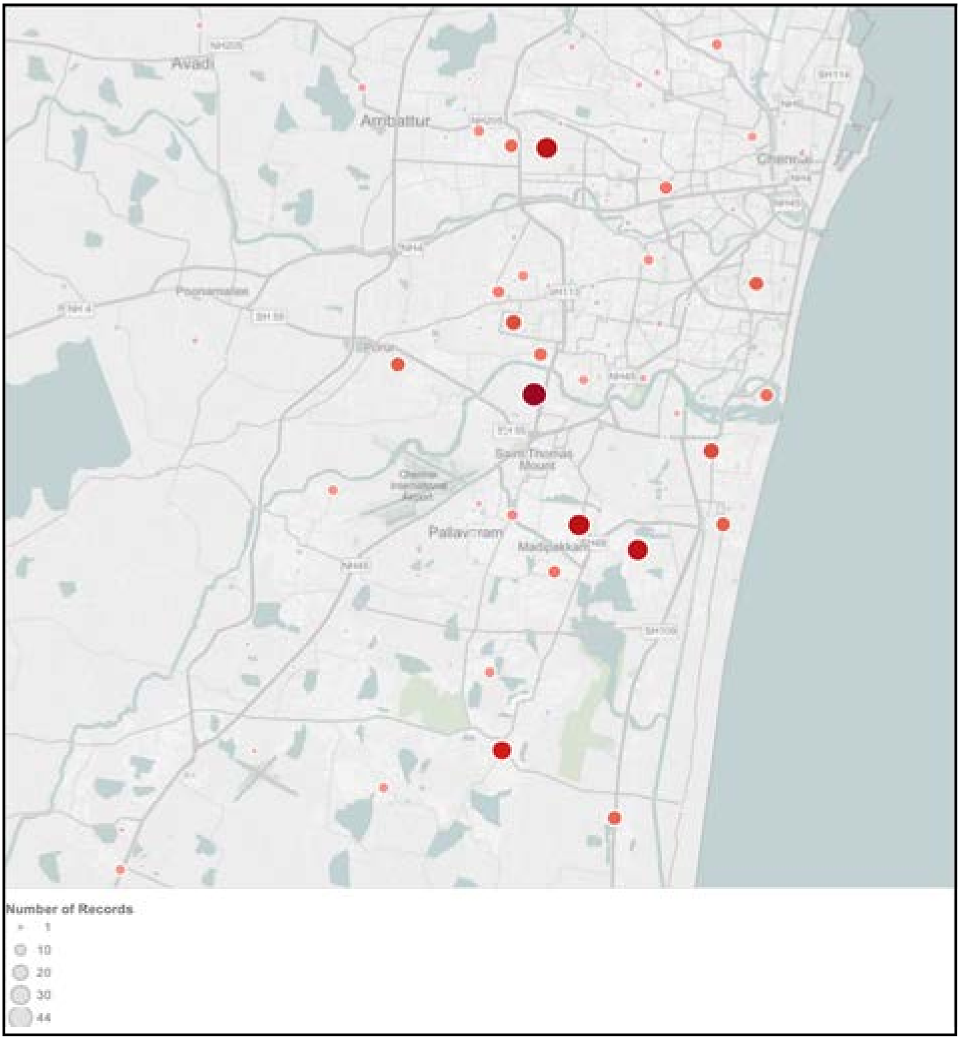
We decided to use the whole 13 months data available. The data had to be formatted to contain only two columns – Transaction_ID & Category_Name. The analytics tool, R, insisted on the format to generate association rules. The data set contained 6823 rows and 250 categories. We would also attempt a more drilled down analysis by creating association rule mapping specific products in place of categories.

RESULTS

Cost-Benefit Analysis

Figure 1 is the distribution of the order location of the online store. As we can see the order is distributed throughout Chennai and hence we have done a cost benefit analysis to find the apt location for warehouse.

Figure 1



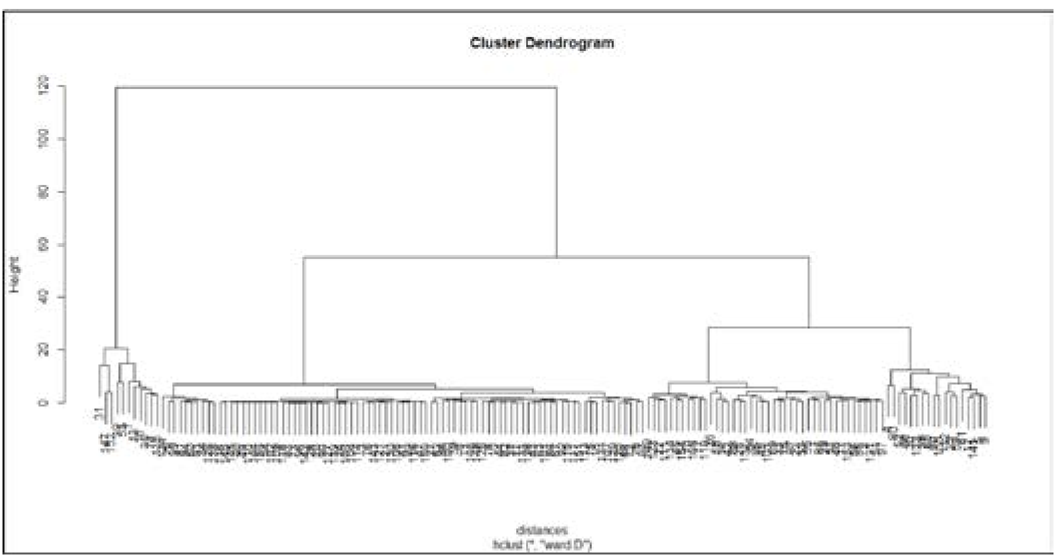
The analytical tool R has function 'mapdist' inbuilt into the library ggmap which was used in finding the distance between the source and destination. By taking the top 10 frequent order location as source the distance was calculated using R between the source and different order points. By assuming price of petrol to be Rs 5.91 per km, the total cost spend on logistics were calculated below is the total cost calculated by multiplying the distance to cost of petrol (assumed to be Rs 5.91 per km) and the total order.

Table 1	
Source Pin Code	Total Cost
600040	**49873297.86**
600042	49888653.29
600092	49900201.32
600116	49967930.1
600096	**49969092.05**
600097	50052493.85
600100	50102823.17
600119	50254635.95
600004	189762197.8
600101	189826695.3

Cluster Analysis

The analytical tool, R, had inbuilt functions to estimate the number of clusters based on optimum average silhouette width. The number came out to be 2. Figure 2 is the Clustering Dendro gram which the most important result of clustering. It lists all samples and indicates at what level of similarity any two clusters were joined. The diagram indicates only the possibility of 2 clusters.

Figure 2

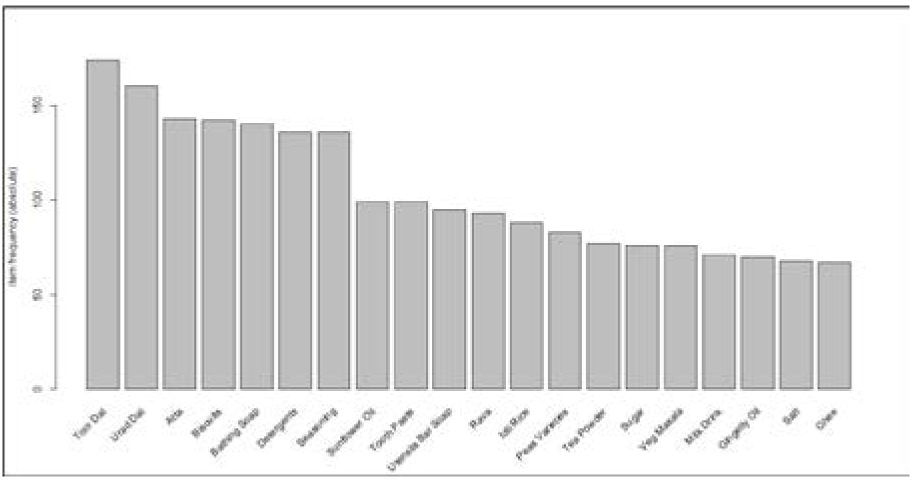


The k-means clustering created two clusters containing 12 and 144 categories (rows) each**.**

Market Basket Analysis

Appropriate value for minimum support and confidence had to be chosen before running the Association Rule algorithm. We had chosen the minimum confidence to be 60% based on qualitative analysis. Figure 3 is the plot for the top 20 frequently purchased category from the e-grocer website. Toor Dal and Urad Dal tops the list followed by Atta, Biscuits, Bathing soap, Detergents and Seasonings. Results are expected to be focused on these frequent items.

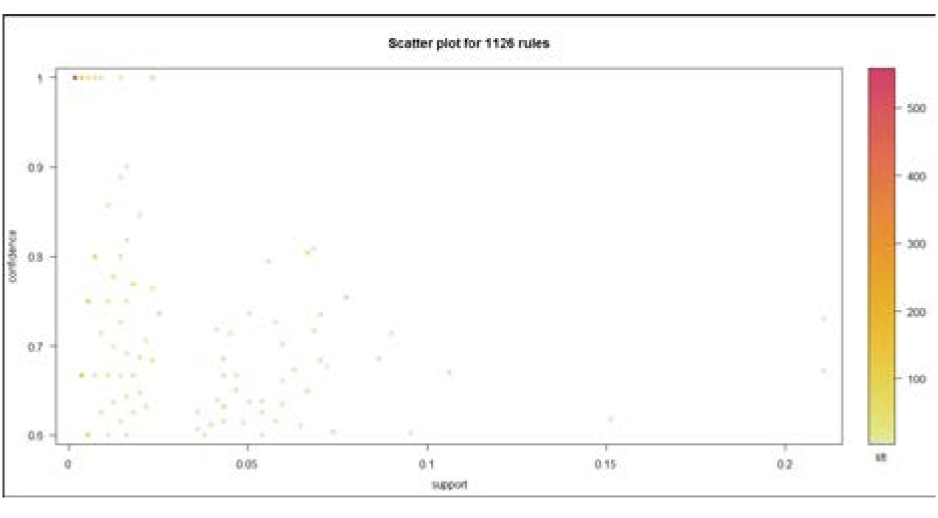
Figure 3



Category Association (One-one mapping)

Association rules are generated for one-one mapping of various categories. Figure 4 plots 1126 rules generated from the data for a minimum support of 0.1 %. Minimum support for further analysis was chosen to be 5% as it generated significant number of rules. Also, 5% in 555 transaction means a substantial 27 transactions (at the least) forms the basis for a single rule.

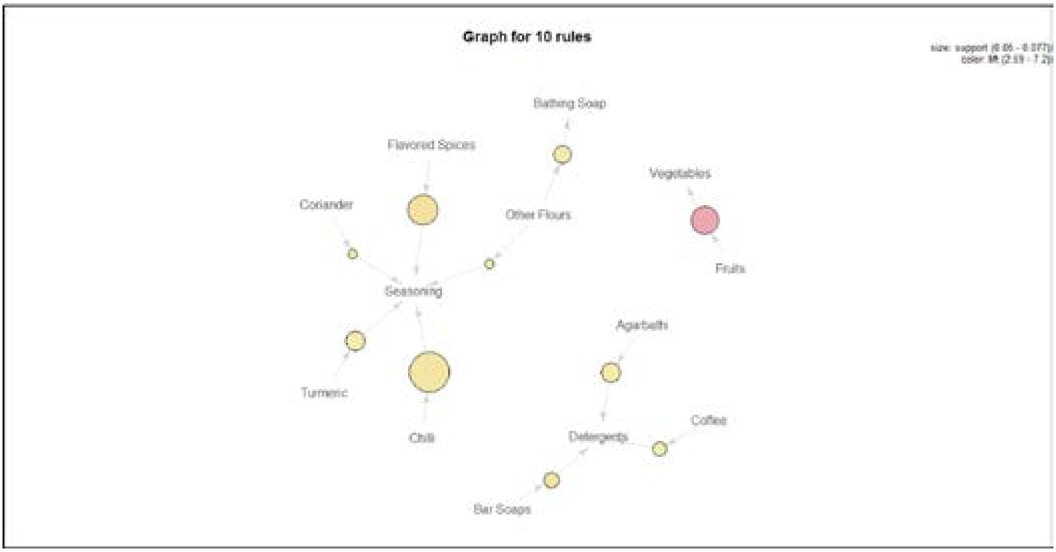
Figure 4



The Error! Reference source not found. displays the 32 rules generated by the association algorithm in the decreasing order of confidence. Figure 5 visualizes the association rules using vertices and edges where vertices typically represent items or item sets and edges indicate relationship in rules.

Category A	Category B	Support	Confidence	Lift
Flavored Spices	Seasoning	0.06846847	0.8085106	3.299437
Fruits	Vegetables	0.06666667	0.8043478	7.20021
Bar Soaps	Detergents	0.05585586	0.7948718	3.243778
Chilli	Seasoning	0.07747748	0.754386	3.07856
Coriander	Seasoning	0.05045045	0.7368421	3.006966
Fried Gram	Urad Dal	0.07027027	0.7358491	2.552476
Urad Dal	Toor Dal	0.21081081	0.73125	2.332435
Other Flours	Bathing Soap	0.05765766	0.7272727	2.883117
Fried Gram	Toor Dal	0.06846847	0.7169811	2.286923
Gingelly Oil	Urad Dal	0.09009009	0.7142857	2.477679
Coffee	Urad Dal	0.05945946	0.7021277	2.435505
Gingelly Oil	Toor Dal	0.08648649	0.6857143	2.187192
Chilli	Urad Dal	0.07027027	0.6842105	2.373355
Gram Dal	Urad Dal	0.07207207	0.6779661	2.351695
Turmeric	Toor Dal	0.06306306	0.6730769	2.146883
Toor Dal	Urad Dal	0.21081081	0.6724138	2.332435
Idli Rice	Urad Dal	0.10630631	0.6704545	2.325639
Agarbathi	Detergents	0.05945946	0.66	2.693382
Chilli	Toor Dal	0.06666667	0.6491228	2.070478
Coffee	Detergents	0.05405405	0.6382979	2.604819
Shampoo	Bathing Soap	0.05045045	0.6363636	2.522727
Other Flours	Seasoning	0.05045045	0.6363636	2.596925
Turmeric	Seasoning	0.05945946	0.6346154	2.589791
Asafoetida	Toor Dal	0.05405405	0.625	1.993534
Seasoning	Urad Dal	0.15135135	0.6176471	2.142463
Turmeric	Detergents	0.05765766	0.6153846	2.511312
Gram Dal	Toor Dal	0.06486486	0.6101695	1.94623
Salt	Urad Dal	0.07387387	0.6029412	2.091452
Salt	Toor Dal	0.07387387	0.6029412	1.923174
Idli Rice	Toor Dal	0.0954955	0.6022727	1.921042
Agarbathi	Urad Dal	0.05405405	0.6	2.08125
Agarbathi	Toor Dal	0.05405405	0.6	1.913793

Figure 5



Category Bundling (Many-One mapping)

Association rule mining is used to generate incidence patterns for more than one categories. Figure 6 plots 31207 rules with minimum support of 1%. Based on analysis, the minimum support value is selected as 6% and confidence as 70% to get more accurate rules. Table 3 list the 52 rules generated by the algorithm sorted by confidence. Figure 7 enhances matrix-based visualization using grouping of rules via clustering to handle a larger number of rules.

Figure 6

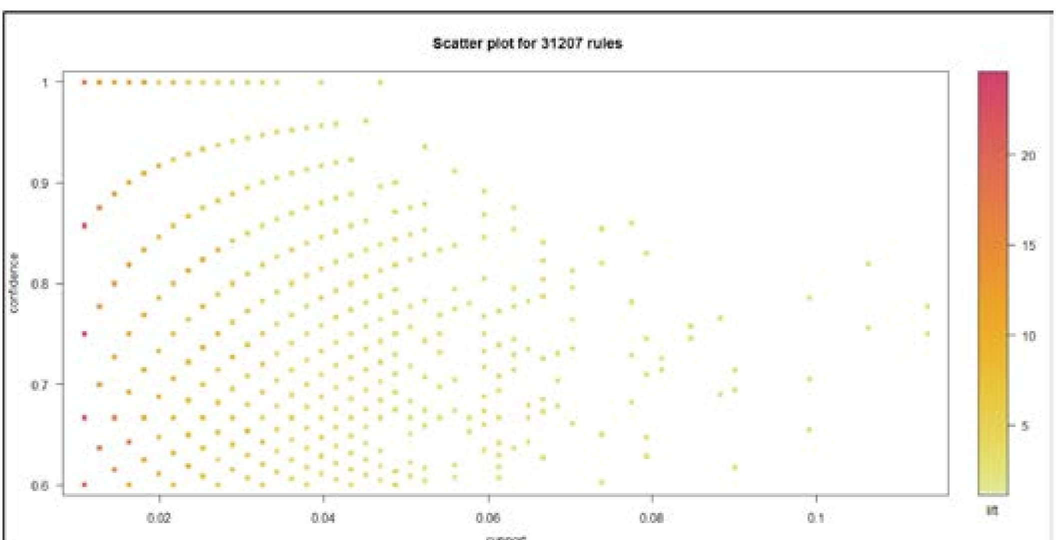


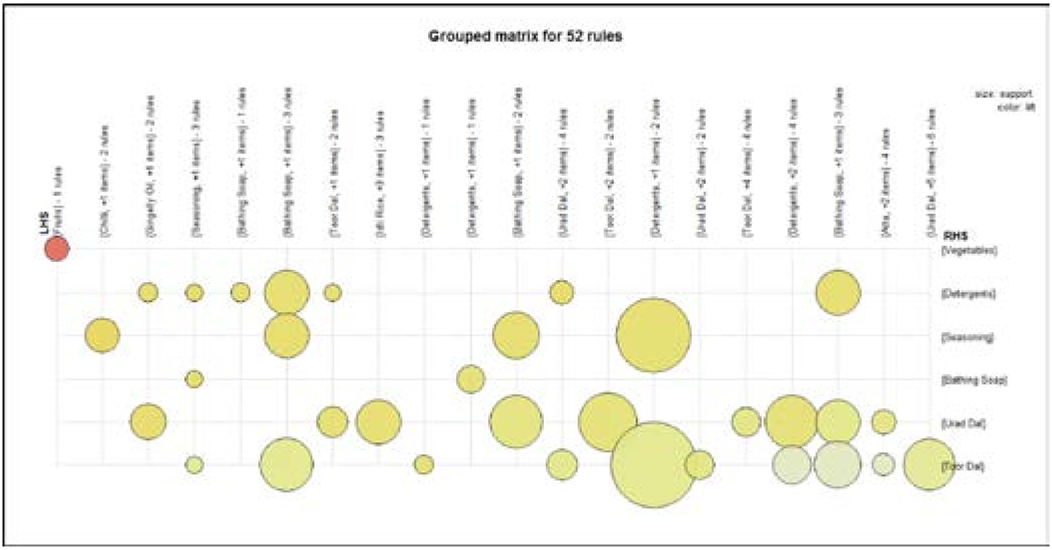
Table 3

Categories	Category B	Support	Confidence	Lift
Detergents, Gingelly Oil	Toor Dal	0.06306306	0.875	2.790948
Gingelly Oil, Toor Dal	Urad Dal	0.07387387	0.8541667	2.962891
Idli Rice,Toor Dal	Urad Dal	0.07927928	0.8301887	2.879717
Gingelly Oil, Urad Dal	Toor Dal	0.07387387	0.82	2.615517
Detergents, Toor Dal	Urad Dal	0.10630631	0.8194444	2.842448
Detergents, Idli Rice	Urad Dal	0.07027027	0.8125	2.818359
Toor Dal, Utensils Bar Soap	Urad Dal	0.07027027	0.8125	2.818359
Flavored Spices	Seasoning	0.06846847	0.8085106	3.299437
Fruits	Vegetables	0.06666667	0.8043478	7.20021
Idli Rice, Seasoning	Urad Dal	0.06666667	0.8043478	2.790082
Tooth Paste, Urad Dal	Toor Dal	0.07027027	0.7959184	2.538705
Detergents, Seasoning	Urad Dal	0.0990991	0.7857143	2.725446
Rava,Urad Dal	Toor Dal	0.06486486	0.7826087	2.496252
Peas Varieties,Toor Dal	Urad Dal	0.06306306	0.7777778	2.697917
Seasoning, Toor Dal	Urad Dal	0.11351351	0.7777778	2.697917
Atta,Urad Dal	Toor Dal	0.08828829	0.765625	2.44208
Urad Dal,Utensils Bar Soap	Toor Dal	0.07027027	0.7647059	2.439148
Bathing Soap, Urad Dal	Toor Dal	0.08468468	0.7580645	2.417964
Detergents, Urad Dal	Toor Dal	0.10630631	0.7564103	2.412688
Chilli	Seasoning	0.07747748	0.754386	3.07856
Seasoning, Urad Dal	Toor Dal	0.11351351	0.75	2.392241

Bathing Soap, Toor Dal	Urad Dal	0.08468468	0.7460317	2.587798
Idli Rice,Urad Dal	Toor Dal	0.07927928	0.7457627	2.378726
Peas Variet-ies,Urad Dal	Toor Dal	0.06306306	0.7446809	2.375275
Seasoning,Tooth Paste	Bathing Soap	0.06126126	0.7391304	2.930124
Seasoning,Tooth Paste	Detergents	0.06126126	0.7391304	3.016304
Seasoning,Tooth Paste	Toor Dal	0.06126126	0.7391304	2.357571
Fried Gram	Urad Dal	0.07027027	0.7358491	2.552476
Toor Dal,Tooth Paste	Urad Dal	0.07027027	0.7358491	2.552476
Rava,Toor Dal	Urad Dal	0.06486486	0.7346939	2.548469
Tooth Paste,Urad Dal	Detergents	0.06486486	0.7346939	2.998199
Urad Dal	Toor Dal	0.21081081	0.73125	2.332435
Sunflower Oil,Toor Dal	Urad Dal	0.06846847	0.7307692	2.534856
Gingelly Oil,Toor Dal	Detergents	0.06306306	0.7291667	2.975643
Atta,Detergents	Urad Dal	0.06306306	0.7291667	2.529297
Bathing Soap,Sea-soning	Toor Dal	0.08108108	0.7258065	2.315072
Urad Dal,Utensils Bar Soap	Detergents	0.06666667	0.7254902	2.96064
Fried Gram	Toor Dal	0.06846847	0.7169811	2.286923
Gingelly Oil	Urad Dal	0.09009009	0.7142857	2.477679
Bathing Soap,Utensils Bar Soap	Detergents	0.06306306	0.7142857	2.914916
Bathing Soap,Toor Dal	Seasoning	0.08108108	0.7142857	2.914916
Detergents,Sea-soning	Toor Dal	0.09009009	0.7142857	2.278325
Bathing Soap,Sea-soning	Detergents	0.07927928	0.7096774	2.89611

Bathing Soap,Urad Dal	Detergents	0.07927928	0.7096774	2.89611
Bathing Soap,Seasoning	Urad Dal	0.07927928	0.7096774	2.461694
Bathing Soap,Urad Dal	Seasoning	0.07927928	0.7096774	2.89611
Detergents,Idli Rice	Toor Dal	0.06126126	0.7083333	2.259339
Toor Dal,Utensils Bar Soap	Detergents	0.06126126	0.7083333	2.890625
Atta,Detergents	Toor Dal	0.06126126	0.7083333	2.259339
Detergents,Urad Dal	Seasoning	0.0990991	0.7051282	2.877545
Sunflower Oil,Urad Dal	Toor Dal	0.06846847	0.7037037	2.244572
Detergents,Tooth Paste	Bathing Soap	0.06846847	0.7037037	2.789683

Figure 7



Product Association

Association rule was used to generate rules for products which are frequently bought together. Figure 8 is the frequency distribution of 20 products from the 13 months data. The minimum support and confidence was decided to be 3% and 60% based on qualitative analysis. Table 4 is the output of the algorithm which generated 36 rules for product incidence. Figure 9 is the grouped matrix created for easy visualization of the association rule.

Figure 8

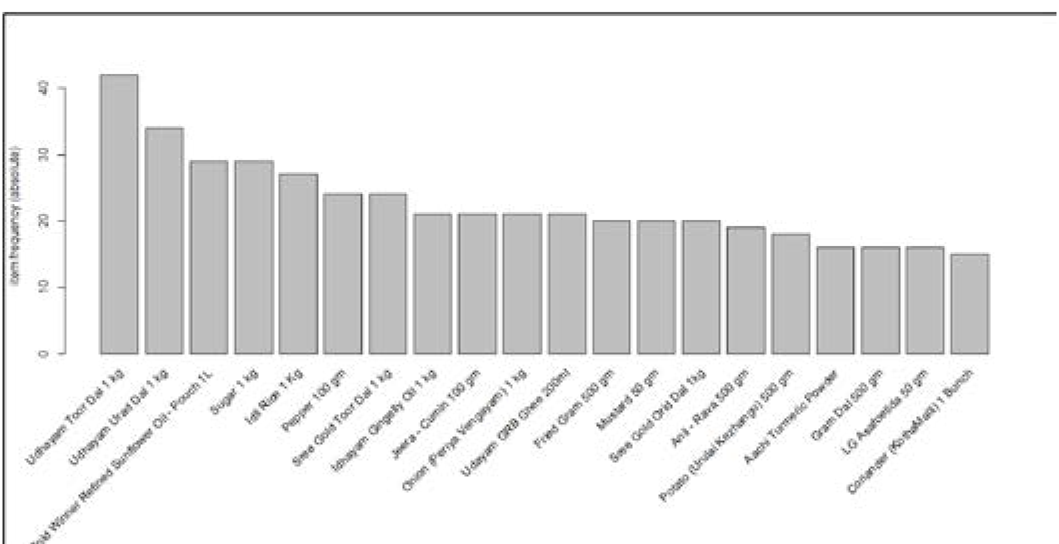


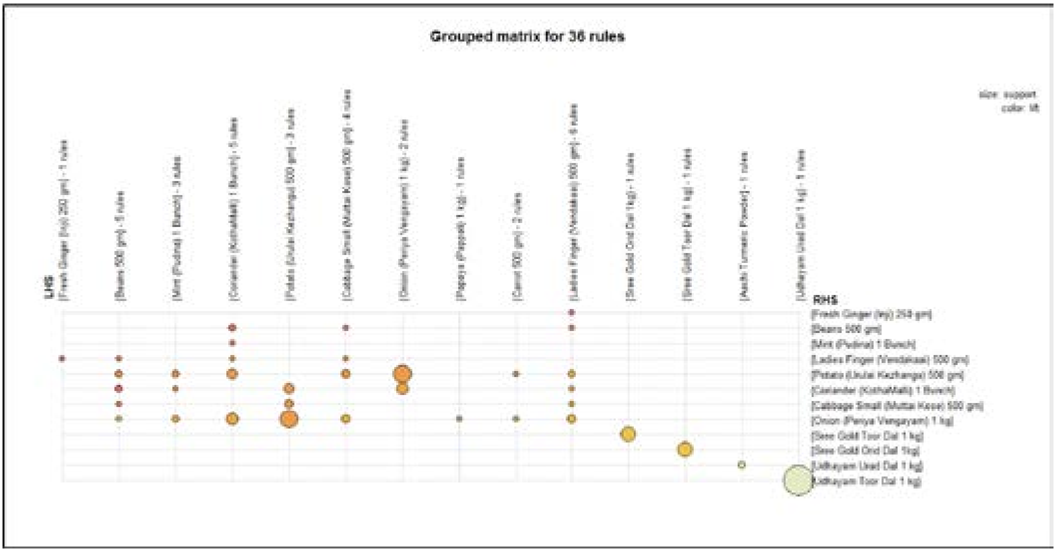
Table 4

Table 4 Product A	Product B	Support	Confidence	Lift
Potato (Urulai Kezhangu) 500gm	Onion (Periya Vengayam) 1 kg	0.06319703	0.9444444	12.097884
Papaya (Pappali) 1 kg	Onion (Periya Vengayam) 1 kg	0.03345725	0.9	11.528571
Coriander (KothaMalli) 1 Bunch	Onion (Periya Vengayam) 1 kg	0.04832714	0.8666667	11.101587
Mint (Pudina) 1 Bunch	Potato (Urulai Kezhangu) 500 gm	0.03717472	0.8333333	12.453704
Mint (Pudina) 1 Bunch	Onion (Periya Vengayam) 1 kg	0.03717472	0.8333333	10.674603
Beans 500 gm	Coriander (KothaMalli) 1 Bunch	0.03717472	0.8333333	14.944444
Beans 500 gm	Potato (Urulai Kezhangu) 500 gm	0.03717472	0.8333333	12.453704
Carrot 500 gm	Potato (Urulai Kezhangu) 500 gm	0.03345725	0.8181818	12.227273
Carrot 500 gm	Onion (Periya Vengayam) 1 kg	0.03345725	0.8181818	10.480519
Fresh Ginger (Inji) 250 gm	Ladies Finger (Vendakaai) 500 gm	0.03345725	0.8181818	14.672727
Onion (Periya Vengayam) 1 kg	Potato (Urulai Kezhangu) 500 gm	0.06319703	0.8095238	12.097884
Coriander (KothaMalli) 1 Bunch	Potato (Urulai Kezhangu) 500 gm	0.04460967	0.8	11.955556
Cabbage Small (Muttai Kose) 500 gm	Potato (Urulai Kezhangu) 500 gm	0.04089219	0.7857143	11.742063
Cabbage Small (Muttai Kose) 500 gm	Onion (Periya Vengayam) 1 kg	0.04089219	0.7857143	10.064626

Mint (Pudina) 1 Bunch	Coriander (KothaMalli) 1 Bunch	0.03345725	0.75	13.45
Beans 500 gm	Cabbage Small (Muttai Kose) 500 gm	0.03345725	0.75	14.410714
Beans 500 gm	Ladies Finger (Vendakaai) 500 gm	0.03345725	0.75	13.45
Beans 500 gm	Onion (Periya Vengayam) 1 kg	0.03345725	0.75	9.607143
Sree Gold Orid Dal 1kg	Sree Gold Toor Dal 1 kg	0.05576208	0.75	8.40625
Udhayam Urad Dal 1 kg	Udhayam Toor Dal 1 kg	0.0929368	0.7352941	4.709384
Ladies Finger (Vendakaai) 500 gm	Onion (Periya Vengayam) 1 kg	0.04089219	0.7333333	9.393651
Coriander (KothaMalli) 1 Bunch	Beans 500 gm	0.03717472	0.6666667	14.944444
Potato (Urulai Kezhangu) 500 gm	Coriander (KothaMalli) 1 Bunch	0.04460967	0.6666667	11.955556
Ladies Finger (Vendakaai) 500 gm	Potato (Urulai Kezhangu) 500 gm	0.03717472	0.6666667	9.962963
Cabbage Small (Muttai Kose) 500 gm	Beans 500 gm	0.03345725	0.6428571	14.410714
Cabbage Small (Muttai Kose) 500 gm	Ladies Finger (Vendakaai) 500 gm	0.03345725	0.6428571	11.528571
Sree Gold Toor Dal 1 kg	Sree Gold Orid Dal 1kg	0.05576208	0.625	8.40625
Aachi Turmeric Powder	Udhayam Urad Dal 1 kg	0.03717472	0.625	4.944853

Onion (Periya Vengayam) 1 kg	Coriander (KothaMalli) 1 Bunch	0.04832714	0.6190476	11.101587
Potato (Urulai Kezhangu) 500 gm	Cabbage Small (Muttai Kose) 500 gm	0.04089219	0.6111111	11.742063
Coriander (KothaMalli) 1 Bunch	Mint (Pudina) 1 Bunch	0.03345725	0.6	13.45
Ladies Finger (Vendakaai) 500 gm	Fresh Ginger (Inji) 250 gm	0.03345725	0.6	14.672727
Ladies Finger (Vendakaai) 500 gm	Beans 500 gm	0.03345725	0.6	13.45
Ladies Finger (Vendakaai) 500 gm	Cabbage Small (Muttai Kose) 500 gm	0.03345725	0.6	11.528571
Coriander (KothaMalli) 1 Bunch	Ladies Finger (Vendakaai) 500 gm	0.03345725	0.6	10.76
Ladies Finger (Vendakaai) 500 gm	Coriander (KothaMalli) 1 Bunch	0.03345725	0.6	10.76

Figure 9



Discussion

Cost Benefit Analysis

The warehouse is presently located in Thoraipakkam in Chennai as per the calculations we have seen that the logistics cost from Thoraipakkam would cost a minimum of Rs.4,99,69,092/- per year on logistics. As per calculation we have found that money on logistics could be saved if they shift their operation to the location with pin codes 600040, 600042, 600092, 600116. By shifting the warehouse to Anna Nagar (600040) a total of Rs 95794.19 can be saved in a year.

Hypothesis 1 stated about the Cost-Benefit Analysis. Hypothesis hasn't failed as we were able to analyze the supply chain of the e-grocer and provide an overview of their costs. We have found out the comparative benefit of optimizing their supply chain by shifting their warehouse to customer locations.

Cluster Analysis

Hypothesis 2, which mentioned about clustering of categories, has failed. The cluster analysis has failed to generate any logical results. The original research was done on consumer durable products, whereas our study was on non-durable products or FMCG products. The product life cycle is totally different in case of durables and non-durables. This may be one of the reason which led to the insignificant result in cluster analysis. Also, as the company is in growing stage, the sales data was not consistent across the months and majority of products didn't have any sales in some months. Once the ecommerce company has a consistent sales data, cluster analysis can be attempted upon again, probably in a different study.

Market Basket Analysis

Market Basket Analysis (Hypothesis 3) turned out to be fruitful for the given data. We were able to generate several association rules category wise and product wise. The results can be used for cross-selling, bundling and product positioning. One-one category mapping provided rules which otherwise would go unnoticed. The table can be directly referred for designing the category layouts in the ecommerce website. Many-one mapping can be more useful in bundling. ABC e-grocer has the practice of providing combo packs to customers. This data would be useful for tuning their combo packs so that the buying probability increases. Though the product mapping generated obvious results mostly on vegetables it can be used for product bundling.

REFERENCES

Anna Corinna Cagliano, A. D. (2015). Analysing the diffusion of a mobile service supporting the e-grocery supply chain. *Business Process Management Journal,* Vol. 21 Iss: 4, 928 – 963.

Brijs, T. S. (2004). Building an association rules framework to improve product assortment decisions. Data Mining and Knowledge Discovery, 8(1), 7–23.

Edmund Prater, G. V. (2005). Future impacts of RFID on e-supply chains in grocery retailing. Supply Chain Management: *An International Journal*, Vol. 10 Iss: 2, 134 - 142.

Elsayed E. A, T. C. (1983). Analysis of inventory systems with deteriorating items. *International Journal of Production* Research, Vol. 21 Issue 4, 449.

Ferguson, M. J. (2007). An application of the EOQ model with nonlinear holding cost to inventory management of perishables. *European Journal of Operational Research,* 180(1), 485.

Ferguson, M. K. (2006). Information sharing to improve retail product freshness of perishables. Production and Operations Management, 15(1), 57-73.

Ghare, P. M. (1963). A model for exponentially decaying inventory. *Journal of Industrial Engineering* 14.5, 238-243.

Kahn, M. A. (1975). Analysis Of Critical Number Ordering Policy For Perishable Inventories. University of Pennsylvania.

Kamakura, W. A. (2012). Sequential market basket analysis. Springer Science+Business Media, LLC 2012.

Ketzenberg, M. E. (2000). Information in supply chain management of perishables.

Kishore R A, T. R. (2011). N-Period Dynamic Deterministic Inventory Model for Perishable Goods. *The IUP Journal of Operations Management,* Vol. X No. 1.

Puneet Manchanda, A. A. (1999). The 'Shopping Basket': A Model for Multi-Category Purchase Incidence Decision. Marketing Science, 18(2), 95-114.

Srivastava, R. &. (2009). A Real-time Order Management and Production Planning Model in An Integrated Supply Chain. *International Journal of Business Insights & Transformation,* 2(2).

Taariq Lewis, B. L. (2009). Case Analysis Studies Of Diffusion Models on E-Commerce Transaction Data. MASSACHUSETTS: MIT Sloan School of Management.

Weiss, H. (1982). Economic Order Quantity Models with Nonlinear Holding Costs. *European Journal of Operational Research* 9, 56–60.