Association for Information Systems

AIS Electronic Library (AISeL)

MCIS 2014 Proceedings

Mediterranean Conference on Information Systems (MCIS)

Summer 9-4-2014

A DATA MINING-BASED FRAMEWORK TO IDENTIFY SHOPPING MISSIONS

Anastasia Griva
Athens University of Economics and Business, Athens, grivaa@aueb.gr

Cleopatra Bardaki
ELTRUN, Department of Management Science and Technology, Athens University of Economics & Business, Greece, cleobar@aueb.gr

Sarantopoulos Panagiotis,
Athens University of Economics and Business, Athens, psarantopoul@aueb.gr

Dimitris, Papakiriakopoulos
Technological Educational Institute of Athens, Athens Greece, dpapak@teiath.gr

Recommended Citation

Griva, Anastasia; Bardaki, Cleopatra; Panagiotis,, Sarantopoulos; and Papakiriakopoulos, Dimitris,, "A DATA MINING-BASED FRAMEWORK TO IDENTIFY SHOPPING MISSIONS" in Mola, L., Carugati, A,. Kokkinaki, A., Pouloudi, N., (eds) (2014) *Proceedings of the 8th Mediterranean Conference on Information Systems*, Verona, Italy, September 03-05. CD-ROM. ISBN: 978-88-6787-273-2.
http://aisel.aisnet.org/mcis2014/20

A DATA MINING-BASED FRAMEWORK TO IDENTIFY SHOPPING MISSIONS

Anastasia, Griva[1], Department of Informatics, Athens University of Economics and Business, Athens Greece, grivaa@aueb.gr

Cleopatra, Bardaki[1], Department of Management Science and Technology, Athens University of Economics and Business, Athens Greece, cleobar@aueb.gr

Panagiotis, Sarantopoulos[1], Department of Management Science and Technology, Athens University of Economics and Business, Athens Greece, psarantopoul@aueb.gr

Dimitris, Papakiriakopoulos[1], Technological Educational Institute of Athens, Athens Greece, dpapak@teiath.gr

Complete Research

Abstract

The success of any business depends on the ability to understand its customers. As any other business so do retailers, understanding the reasons consumers enter their stores is playing a key role in achieving competitive advantage and retaining their market shares. Nowadays, the advent of Business Analytics has created new ways for retailers to metamorphose the vast amount of data they have into valuable knowledge, in order to gain customers' insights, and improve customer relationships. However, not enough research has been conducted to analyze point-of-sale (POS) retail data in order to investigate consumers' behavior and understand the reasons they visit retail stores. This study presents an effort to fill this gap by introducing a Data Mining-based Framework, which could be used to discover patterns in customers' visits in a supermarket, and identify their Shopping Missions. The utility of this framework is been demonstrated by applying it in real data of eight representative stores of a Greek retailer. The proposed approach is useful for both academia and retail industry. As it gives the retailers the opportunity to extract consumers' shopping missions when they visit their supermarkets, it could be used to support several decisions in the retail domain, and improve the relationships between retailers and consumers.

Keywords: Data Mining (DM), Clustering, Retail Analytics, Shopping Missions

[1] ELTRUN: The eBusiness Research Center

1 Introduction

Understanding the reasons consumers enter stores for has always been among retailers' greatest aspirations. The advent of affordable analytics and data processing capabilities (Bose, 2009) has made this knowledge attainable. Retailers can instantaneously infer a wealth of consumer behavioral insights by extracting the knowledge hidden in point-of-sale (POS) data. Associations and complementarities between product categories comprising grocery baskets convey signals that if properly analyzed can reveal the true meanings of shopping trips.

Prior marketing literature claims that consumers shopping trips have loosely shopping goals (Kahn & Schmittlein, 1989; Lee & Ariely, 2006). Many researchers have investigated the factors that make consumers enter to a store (Bell, Corsten, & Knox, 2011), and their shopping trip goals (Lee & Ariely, 2006). However, such approaches provide limited, information about the specific reasons consumers enter the store for, i.e., the actual products to be purchased. Nowadays, due to the economic crisis, consumers are trying plan their shopping visits, based on their consumption needs, moreover they visit stores more frequently and for fewer items at each time. In this context, Sarantopoulos, Theotokis, & Pramatari (2014) define shopping mission as the underlying consumption need(s) of a single store visit, which results in purchasing products that reflect this need. Under this lens, consumers may enter the store to buy products for breakfast such as milk, toast, eggs etc. and/or to prepare a meal, e.g. meat, pasta, vegetables etc. This means that they have two different shopping missions when entering the store.

While there have been a number of solutions that utilize POS data proposed in IS literature (Ahn, 2012; Liao, Chu, & Hsiao, 2012; Shrivastava & Sahu, 2007), such a behavioral segmentation of shopping trips from POS data has been largely overlooked. To fill this gap, we follow the design science paradigm of IS research (Gregor, 2013; Hevner, March, Park, & Ram, 2004), to prescribe a framework that employs cluster analysis to extract shopping missions from POS data. Furthermore, the utility of this solution is demonstrated in practice through its application to a real dataset from eight representative stores of a major supermarket chain in Greece.

Our framework allows the metamorphosis of POS transactions from customer touch-points to customer learning-points. The common retail POS data are transformed to knowledge about the consumers' purchasing behavior and shopping goals. This research empowers the retailers and manufacturers of consumer goods to identify the shopping missions of each store, or channel, and utilize this information to make strategic plans and take effective decisions in order to acquire a larger share of consumers' shopping missions and market share. For example, retail store design may change by adjusting the products placement in shelves according to the identified shopping missions. Moreover, retailers can take clientele profiling to an even lower behavioral level by tying POS shopping missions to customer demographic information residing in loyalty schemas.

The reminder of the paper is organized as follows. Section 2 analyzes the research area, surveys the literature, and points out the research gap. The proposed framework and its evaluation are given in Section 3 and 4, respectively. Finally, Section 5 overviews the main outcomes of the paper, and presents the theoretical contribution and the practical implications of the proposed framework. Moreover, further research is highlighted.

2 Data Mining (DM) in the Retail Industry

Customer satisfaction is the key of the success of any retail store or business (Jeevananda, 2011) and has an effect on their profitability. So, Customer Relationship Management (CRM) has risen to the agenda of many organizational strategies (Bull, 2003). As a consequence, retailers need to embrace a customer-centric focus and find out innovative ways to support CRM strategies, to manage their customers and improve their relationships (Anderson, Jolly, & Fairhurst, 2007; Linoff & Berry, 2011).

The use of IT has created new ways for firms to exploit vast potentials of CRM. Business Intelligence (BI) tools are used to assist CRM systems (Phan & Vogel, 2010). Many companies have collected and stored useful data. However, they are unable to transform these data into valuable knowledge. For that reason the application of Data Mining (DM) tools in CRM is an emerging trend in global economy (Ngai, Xiu, & Chau, 2009). Data Mining applied to CRM enables in-depth analysis of datasets, and allows extracting hidden customer characteristics, and behaviors of large volumes of data (Liao et al., 2012; Ngai et al., 2009).

Data Mining-enhanced CRM could help retailers patronage customers' behavior, gain insights, and retain customer's that really add value to the business (Min, 2006). By discovering patterns in customers' behaviors, enterprise's decision making could also be empowered (Wang & Zhou, 2013). A well-known example is Tesco, which has reinvented its relation with its customers by using Loyalty Card data (Humby, Hunt, & Phillips, 2003).

Except for the traditional Business Intelligence and Data Mining applications in retailing, there are few papers in literature that are relevant to the concept of shopping mission. The most recent relevant researches are that of Cil (2012) and Borges (2003). Their main purpose is to identify the associations among categories in a supermarket, which will be used to change a store's layout. Association rules and apriori algorithms are used to analyze POS data of a Turkish and a French supermarket respectively. Also, few scholars try to understand customer's behavior in supermarkets via market-basket analysis (Ahn, 2012; Raorane, Kulkarni, & Jitkar, 2012; Shrivastava & Sahu, 2007). In these cases, apriori and nearest neighbor algorithms are used to identify the hidden associations between product categories, and to gain customer insights. But, the extracted association rules are not used to identify shopping missions or trips. Another interesting research is that of Larson, Bradlow, & Fader (2005). They used RFID (Radio Frequency Identification) in shopping carts to record customers' shopping paths in a grocery store. Then by implementing clustering they classify these paths and examine common travel behaviors.

To the best of our knowledge there is no other framework – methodology, which gives specific steps and guidelines to analyze POS retail data per basket and extract the customers' shopping missions, by identifying correlations in product categories. In contrast to the relevant works that are using association rule mining, apriori and nearest neighbor algorithms to identify the correlations between products, this is the first research that introduces clustering, as data mining model and k-means, as data mining technique to reach its goal.

3 A Framework that identifies Shopping Missions

3.1 Methodology

This research adopted the "Design Science" approach (Hevner et al., 2004). An artifact has been developed; it is a Data Mining-based framework (Figure 1) that draws on CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology (Chapman et al., 2000) and includes five phases: (A) Business and Data Understanding, sub-steps of which are: Data Acquisition, Data Exploration and Data Preparation, (B) Cluster Sampling, (C) Modeling, (D) Evaluation, and (E) Deployment. The proposed framework has been evaluated in practice, in order to realize its ability to solve the original problem. We extracted the shopping missions for eight representative stores of a big Greek retailer. The following subsections summarize each framework's phase.

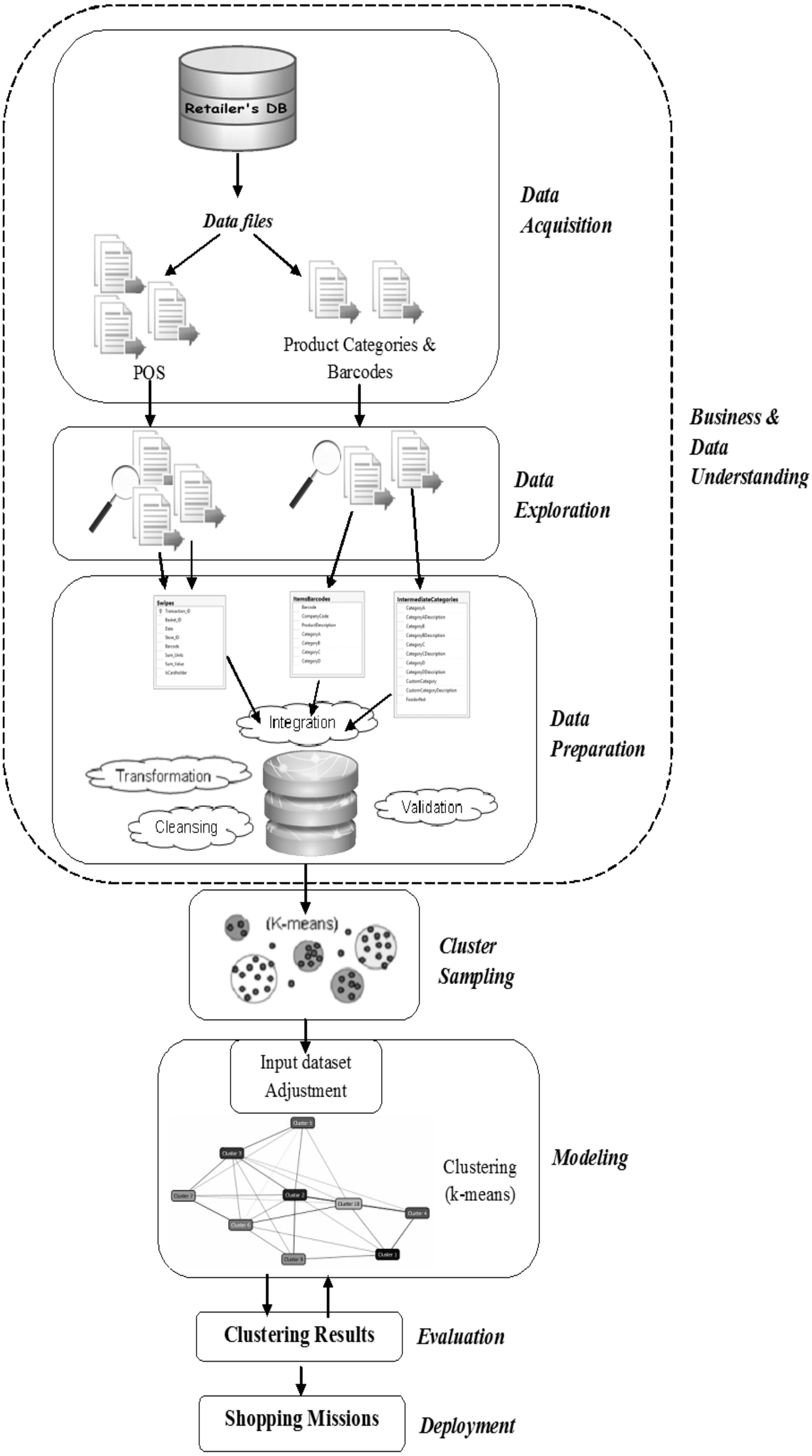

Figure 1. *Proposed Framework*

3.2 Business & Data Understanding

The initial phase focuses on obtaining, exploring and preparing the dataset, having in mind the business objective, which is to extract the shopping missions.

3.2.1 Data Acquisition

The necessary dataset is the following:
- Data that concern the consumers' daily purchases collected from the Point-Of-Sales (POS) /cashiers in the retail stores. You need to know all the transactions that occurred in the POS, with additional information about the receipt-basket they belong to.
- Data that concern the product categories hierarchy (product taxonomies) and the barcodes the retailer uses.

Moreover, you could enrich the framework by asking for additional data, such as customers' loyalty cards data. Having the latter dataset, you can extract additional information, such as demographic characteristics, about each resulting shopping mission.

3.2.2 Data Exploration

It is the beginning phase of the analysis. It is a high level of data understanding, and it is necessary to get familiar with the dataset and access whether the available data would supply data mining techniques (Shahbaba, 2012).

During this phase you have (A) to evaluate whether the data acquired satisfy the relevant requirements, (B) to select the appropriate data (Chapman et al., 2000); and (C) to process the product categories hierarchy you will use. The last step is critical for the forthcoming analysis. It includes exploring the given product taxonomies and determining new, customized product category names that will be meaningful for the identification of the shopping missions. For example, examining the product taxonomies of Figure 2, we can create two new product categories, named "toast breads" and "breads".

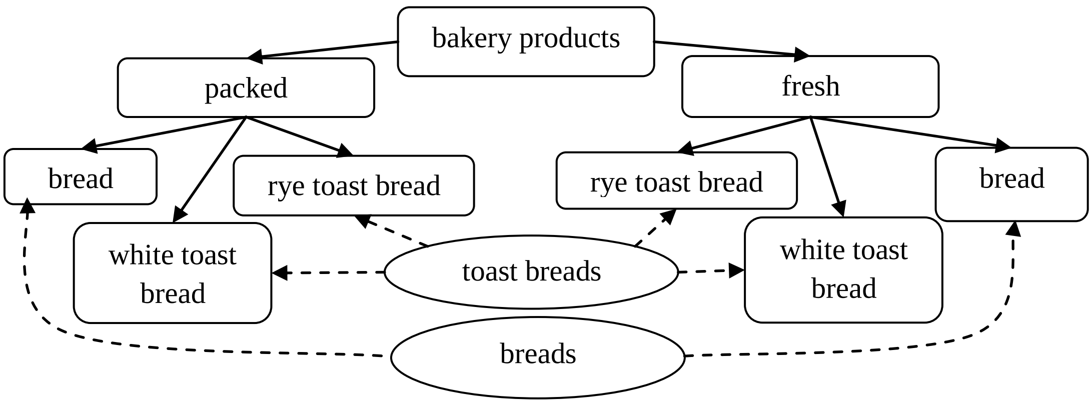

Figure 2. Creation of new customized product categories

3.2.3 Data Preparation

Data Preparation phase covers all the activities to construct the final dataset, before applying cluster sampling, and before proceeding with the modeling. The basic tasks in this framework are the following:

- *Data Integration*: Involves combining data residing in different sources (Lenzerini, 2002). It is common that the retailer sends the POS data in batches, concerning months, or quarters, etc. A basic task is to integrate all these data in a table that will concern all the POS data.

- *Data Cleansing*: Frequently, the collected data will not be in a format ready for analysis. Data Cleansing involves detecting and correcting or removing errors and inconsistencies of the data to improve data quality to improve data quality (Rahm & Do, 2000). It also contains deleting data irrelevant with your analysis (e.g. seasonal items).
- *Data Transformation*: Some of the data need to be transformed and represented in a different way, more suitable for the analysis. A common need is to transform the information from the level of transaction in POS data, to basket level. We need this information because in the next steps it will be used to implement Cluster Sampling.
- *Data Validation*: Happens after each of the above steps to consolidate the data integrity of the available datasets, based in ad hoc criteria selected by the researchers. The main tasks are: cross-check the data with the initial raw files, ensure that there are not missing data after each step's execution, search for mistakes in data types that caused corrupted records etc.

3.3 Cluster Sampling

Having considered the purpose of the research, there is no meaning in applying data mining on all the dataset. There are outliers in the dataset derived from the previous phases. Outliers are baskets with too many products in terms of volume that are related to many shopping missions; and baskets with very few products, from which you can't extract a shopping mission.

For instance, if you have a basket with two products, such as milk and detergent, this basket does not indicate a specific shopping mission. In order to proceed with the modeling, we have to identify the baskets that really matter. Basket size is the key to find out the meaningful baskets. The sampling technique we need to use is Cluster Sampling, with equal sampling weights. Cluster Sampling is a technique where the finite population is grouped into subpopulations-groups called clusters; then, a subset of these clusters is selected (Särndal, Swensson, & Wretman, 2003). The proposed algorithm to be used is k-means. K-means is one of the well-known algorithms for clustering, searching for a nearly optimal partition with a fixed number of clusters (Phan & Vogel, 2010). After receiving the clustering results, the calculation of the actual number of baskets that belong to each cluster, and the revenues derived from each cluster, are the two basic factors that will help us determine the meaningful and prevailing baskets.

3.4 Modelling

According to CRISP-DM the basic tasks in this phase are: (A) select the data mining model type and technique to build the model, (B) set the model's parameters, and (C) generate a procedure or mechanism to test model's quality (Chapman et al., 2000). The proposed framework specifies the above tasks and adds one more. This task has to do with the need to adjust the dataset in order to ensure interoperability with other tools.

3.4.1 Input Dataset Adjustment

A data mining pre-condition is to integrate all the tables of the database into one. This will be used as the learning dataset of the model. The above is a compulsory step to ensure the interoperability of the dataset with other data mining tools. This new table will include all the information about the database tables in basket level. Each row will represent a basket, having as columns all the attributes that basket has, and the customized product categories created in the previous steps. In Figure 3, the columns of this table are shown. All columns that concern the product categories will be filled with a binary value of (1) or (0) respectively, if the basket contained products of this product category or not.

attributes from Basket table				the categories you have created			additional data about loyal customers	
Basket id	Value	...	Number of Items	Bread	...	Detergent	Cardholder's Age	Cardholder's Sex

Figure 3. Fact Table Columns

3.4.2 Model Implementation

Clustering is proposed as the type of data mining model, and k-means as data mining technique. Clustering is the task of segmenting the objects into groups called clusters, so that the object within a cluster are "similar" to another and "dissimilar" to objects in other clusters. Similarity is defined of how close the objects are in space, based on a distance function (Phan & Vogel, 2010). Before executing the model, we need to generate a procedure or mechanism to test the model’s quality, validity and accuracy. Therefore, we split the dataset into training and testing sets. The training set will be used to implement the clustering, and the testing set to estimate its quality and accuracy (Chapman et al., 2000). The proposed percentage of training and testing set is sixty (60%) and forty (40%) percent respectively. Moreover, it is proposed not to indicate the actual number of clusters, but leave the algorithm at the early stages of the research to execute the optimum case, because on the one hand there isn’t any detailed insight about the number of shopping missions and on the other hand a better fit between the model and the data is expected. After model's implementation, we need to assess the model, by checking its accuracy.

3.5 Evaluation

Now, we evaluate the data mining results with the assistance of industry's people that are experts in this domain. We decide whether the results meet the reality, or re-execute the previous phase with changes in the input dataset. In order to evaluate the results, we make a first identification of the shopping missions. Specifically, the results of the modeling phase would be groups-clusters of product categories; each resulting cluster describes a shopping trip, and each shopping trip contains one or more shopping missions. By identifying the prevailing categories of a cluster, you can identify the shopping missions. For example, if the prevailing product categories of a cluster are: milk, cereals, coffee and sugar, this might be a shopping mission for "breakfast".

The changes in the input dataset have to do with deleting and merging some of the customized product categories we created in Data Exploration phase. Merging contiguous product categories attributes is followed as a practice to increase the internal consistency of a cluster and deliver recognizable shopping missions to the experts. In more detail, the task was to derive potential shopping missions from the available clusters and we realized that it is more effective to identify shopping missions based on generic product categories. A generic product category is formed through the merging operation of two or more product categories that usually have the same parent node. From a data mining perspective this decision decreases sample’s variability and consequently yields better performance results. For example, a cluster may contain the following categories: low-fat milk, high-fat milk, chocolate milk, low calories cereal and high calories cereal. By merging the cereal related product categories into one product category named "cereals", and the milk related product categories into another product category named "milk" (thus create more generic product categories) we acquire two benefits: (a) the clusters are more comprehensive by the experts due to dimensionality reduction and (b) the contribution of the (new) product category within cluster formation is higher. Last but not least, we decided to remove the infrequent product categories from the sample and rebuild the cluster model. To this end the evaluation process forms a dialectic process between the experts and the data mining techniques. The role of the researcher was to balance the cluster model, in order to satisfy important data mining metrics and, at the same time, deliver a readable abstraction of the marker to the experts.

3.6 Deployment

In this phase, the final and verified modeling results are explained and analyzed in depth. The final identification of the shopping missions takes place and we extract the conclusions that retailers can use to support their CRM strategies. Thus, in the last two phases, it would be useful to choose a tool to achieve a greater visualization of your results. This tool will be used as the bridge between the knowledge extracted from the data mining, and the managers. In order to proceed with the identification of the shopping missions, as in the evaluation phase, we examine the product categories, and their contribution in each cluster. It is helpful to calculate the average basket size (baskets volume) and the average number of unique product categories (baskets variety) of the baskets contained in each cluster. These two factors will support the analysis and characterization of the final shopping missions. Clusters that contain few products and product categories could be left un-characterized. Moreover, drill-down clustering could be helpful in clusters that contain abstract shopping trips. With the drill-down we will perform clustering in a single cluster; this way from an abstract shopping trip maybe more than one shopping missions will occur.

4 Framework Evaluation

The proposed framework is used in practice in order to be evaluated, namely to realize if it manages to solve the original problem. One of the biggest retail chains in Greece, in terms of both turnover and number of stores, provided us with retail data, in order to extract the shopping missions. The tools we used to upload, prepare and analyze the data are the following:

- Microsoft SQL Server 2012 of SQL Server Management Studio (SSMS),
- Microsoft SQL Server Integrated Services (SSIS) and Microsoft SQL Server Analysis Services, (SSAS) of SQL Server Data Tools (SSDT) of Visual Studio and
- Data mining add-in for Excel.

4.1 Business & Data Understanding

The collaborating retailer extracted from the corporate database product categories, barcodes and POS data from January 2012 to May 2013, from eight representative stores in the Attica region. The stores had common characteristics in pairs; we had data about two convenience stores, two supermarkets, two mini-hyper markets and two hyper markets. Except for these necessary data, the retailer also gave us customers' loyalty cards data. These data contained demographic characteristics about the retailer's cardholders. After a first data exploration, 104 new-customized categories had been created.

4.2 Cluster Sampling

It has been decided to study the stores and extract the outliers in pairs. For that reason, four different views of the data had been created, one for each store type. Non-scalable k-means algorithm has been used as the clustering method. We also calculated actual number of baskets that belong to each cluster, and the revenues derived from each cluster in order to help us with the outliers' extraction. In Table 1 there is a summary of the selections that have been made. Second row represents the range (from, to) of the basket size of each store type, which will be used in the forthcoming analysis. The two last columns include the percentage of baskets which will be used for the next step of the process, compared with the initial number of baskets, and the percentage of revenues these baskets cause compared with the total sales. In flag-hyper stores there exist baskets with too many products, as people commonly make bulk purchases in these types of stores. For that purpose, and in order to ensure the uniformity of the selected dataset, we decided to use the 34.91% of the total number of basket that causes the 53.48% of the total revenue.

Store Type	Basket Size Range Sample	Percentage of the total baskets used	Percentage of revenue used
Convenience	2-24	78.64%	86.05%
Supermarket	3-40	75.49%	83.42%
Mini-Hyper	3-51	79.75%	84.54%
Flag-Hyper	13-56	34.91%	53.48%

Table 1. Summarized results of Cluster Sampling

4.3 Modeling

Each store has been analyzed separately because identical shopping missions will not necessary result from the same store types. For that reason, Java code was used in order to create eight (8) tables, one per store. Clustering and non-scalable k-means had been used. Eight models have been created and executed, one per table that contains the information about each store. All the models were proved valid, since they scored 91% to 99% accuracy.

4.4 Evaluation

According to the first clustering results, industry people realized an important omission in the resulting clusters. None of the clusters included product categories related to "meat". For that reason we merged categories such as pork, beef, lamb etc. into one. Moreover, about 15 product categories had been deleted, as they appeared in an average percentage less than 2% in all clusters. Finally, the clustering process was continued with only 75 of 104 product categories. Both in this and the next phase, excel data mining add-in had been used to demonstrate the extracted shopping missions to the managers, as it is an adequate visualization means of the data mining results.

4.5 Deployment

The final cluster diagram for a supermarket is shown in Figure 4. The more densely populated clusters have darker color. The intensity of the line’s shading that connects one cluster to another represents the strength of the similarity of the clusters. The average basket size (baskets volume), and the average unique product categories (basket variety) each cluster contained, had been calculated to help us with the analysis. For instance, as far as concerns cluster number 2, we calculated that contains baskets with on average six products that belong to on average three product categories.

According to the clustering results, cluster number 2 contains, in high percentages, product categories related to fresh vegetables, red meat, chicken, white cheese, pasta, eggs, bread, oil, vinegar etc. So, according to the contribution of the percentages that these categories have, this cluster refers to a "Main Course" shopping mission. Alike, cluster 3 contained dominant categories such as milk, baked goods, juice, coffee, tea, and cereals, so it is a "Breakfast" shopping mission. Similarly, in cluster 4 biscuits, chocolates, soft drinks, and chips were the persistent categories, so we called it as a shopping mission for "Snack". Cluster 5 is the "Detergents and Hygiene" shopping mission, as it contains dominant categories such as powders, dish washing, bathroom cleaners, paper rolls, shampoos, body creams, oral hygiene etc. Cluster 6 contains the shopping mission for "Toast with packed products", as the dominant categories are: packed cheese, packed cold cuts, packed bakery products. Cluster 8 contains almost the same categories as cluster 6, with the difference that this time they are not packed, but fresh cutting products. So, this is a shopping mission for "Toast with serviced products". Cluster 7 represents a shopping mission for "Light meal", as it contains as pasta, rice, pulses and canned food.

Clusters 9 and 10 indicated more abstract shopping trips. Drill-down was performed in both these clusters, but the results did not reveal any easy to interpret shopping missions. The occurring sub-clusters either contained the same shopping missions as those that we mentioned before, or they didn't indicate a certain shopping purpose. Last but not least, cluster number 1 could not be characterized, since it contains baskets with very few product categories.

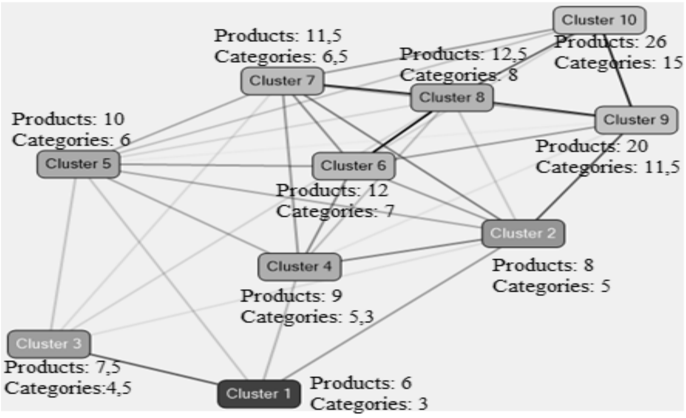

Figure 4. Cluster Diagram for a Supermarket

The results of each store type resembled a lot. According to the results each shopping trip contains from one to three different shopping missions. The shopping trips are becoming more abstract, as the size of the store grows. Thus, the shopping trips, of the two hyper stores were more abstract than those of the other store types. In hyper stores there were many abstract shopping trips which contained a lot of food-based shopping missions, such as "main course", "breakfast", "snack" etc,. There were also shopping trips with both mixed food and non-food shopping missions, such as "household and cleaning", and "main course". Last but not least, this kind of stores had too abstract shopping trips that even by drill down couldn't be characterized. The above happens because customers commonly visit this type of stores to make bulk purchases of several products. Hence, their basket contains a lot of product categories without a certain shopping purpose

After all, the given loyalty data were not used. This kind of data could enrich our framework, as we could correlate each resulting shopping mission with specific groups of customers, based on their demographic characteristics. However, while exploring the given loyalty card data, it has been noticed that many customers didn't fill their personal data such as age, sex, household size etc., so data quality was poor. Moreover, the percentage of purchases that loyal customers had made in the period covered by the given POS data, was significantly low.

5 Conclusions & Discussion

Many enterprises have realized the importance of applying new technological trends to support decision making and satisfy their customers. This research puts an effort to assist retailers in satisfying their demanding customers, via exploiting data mining techniques. We propose a data mining-based framework that discovers patterns in customers' behavior. This framework can change/ improve the nature of retailer-consumer relationships, by giving the retailers the opportunity to realize the shopping missions of consumers during their shopping trip in supermarkets.

To the best of our knowledge, there is no other framework which indicates how to extract shopping missions from retail data, by identifying correlations in product categories, using clustering. This proposed approach differs from the well-known researches in several aspects. First of all, others researchers (Ahn, 2012; Borges, 2003; Chen, Chen, & Tung, 2006; Cil, 2012; Raorane et al., 2012; Shrivastava & Sahu, 2007) make category correlations in retail stores data by using association rule mining, apriori and nearest neighbor algorithms. This research introduces clustering and k-means as data mining model and technique. Last but not least, by enriching this framework with loyal customers' data, it introduces a new way to study and analyze customers in groups.

From a practical point of view, the shopping missions extracted from this framework can support several decisions, which relate to the consumers in the retail context. Specifically, managers could modify and re-design the supermarket's layout, based on the customers' usual shopping missions. Namely, they can change the products placement; products that belong to the same shopping mission can be put in nearby supermarket's aisles and shelves. This allows retailers to cluster products around customers buying habits and appeal to busy customers. Since it is more convenient for consumers to find what they need by spending the least amount of time in the supermarket, customer satisfaction increases, but retailers profit as well. Alike, the display layout of the products in the supermarket's on-line catalog can be altered, in order to increase the basket size of the online shoppers.

Moreover, the extracted knowledge may support commercial decision making in retailing e.g. design of new marketing campaigns and promotions for products that belong to the same shopping mission. One more implication is that the framework's results can help marketers to increase customer loyalty, as they could design cross-coupon programs. Furthermore, a recommendation system for real time purchases in a supermarket can be developed. It will propose to customers the products that they may have forgotten to buy, according to their identified shopping mission(s).

Further research could be conducted to apply other data mining techniques, such as association rules, and compare the resulting shopping missions with those that had been derived from clustering. Moreover, it would be of great interest to compare the shopping missions of loyal to those of non-loyal customers. Additionally, alternative technological means could be used to identify the shopping missions. For instance, via using RFID in shopping carts to record customers' shopping paths in a store, or via conducting consumers' survey for their purchases while they are in a store. Then, it would be interesting to compare the resulting shopping missions of these different approaches.

Last but not least, it could be examined the implementation of cluster sampling based on the number of categories (basket variety) each basket has, rather than on the number of products (basket volume). This may be useful, since the whole analysis and the identification of shopping missions are based on products categories, and not on each unique product. In the proposed approach, there have been extracted baskets with many products, without taking into consideration if these products belong to fewer product categories. For instance, there have been extracted baskets containing more than twenty-five products, without examining if these baskets include products from a small number of product categories i.e. five that could be useful to identify a shopping mission.

REFERENCES

Ahn, K.-I. 2012. Effective product assignment based on association rule mining in retail. Expert Systems with Applications, 39(16): 12551-12556.

Anderson, J. L., Jolly, L. D., & Fairhurst, A. E. 2007. Customer relationship management in retailing: A content analysis of retail trade journals. Journal of Retailing and Consumer Services, 14(6): 394-399.

Bell, D. R., Corsten, D., & Knox, G. 2011. From point of purchase to path to purchase: how preshopping factors drive unplanned buying. Journal of Marketing, 75(1): 31-45.

Borges, A. 2003. Toward a new supermarket layout: from industrial categories to one stop shopping organization through a data mining approach. Paper presented at the Proceedings of the 2003 Society for Marketing Advances Annual Symposium on Retail Patronage and Strategy, Montreal.

Bose, R. 2009. Advanced analytics: opportunities and challenges. Industrial Management & Data Systems, 109(2): 155-172.

Bull, C. 2003. Strategic issues in customer relationship management (CRM) implementation. Business Process Management Journal, 9(Emerald): 592-602.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., & Shearer, C. 2000. CRISP-DM 1.0 Step-by-step data mining guide, CRISP-DM Consortium.

Chen, Y.-L., Chen, J.-M., & Tung, C.-W. 2006. A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. Decision Support Systems, 42(3): 1503-1520.

Cil, I. 2012. Consumption universes based supermarket layout through association rule mining and multidimensional scaling. Expert Systems with Applications, 39(10): 8611-8625.

Gregor, S. A. R. 2013. Positioning and Presenting Design Science Research for Maximum Impact, Vol. 37: 337-A336: MIS Quarterly & The Society for Information Management.

Hevner, A. R., March, S. T., Park, J., & Ram, S. 2004. Design science in information systems research. MIS Quartely, 28(1): 75-105.

Humby, C., Hunt, T., & Phillips, T. 2003. Scoring Points: How Tesco is winning customer loyalty: Kogan Page, London Hardback.

Jeevananda, S. 2011. Study on Customer Satisfaction Level at Hypermarkets in Indian Retail Industry. The International Journal's – Research Journal of Social Science and Management, 1(3): 2.

Kahn, B. E., & Schmittlein, D. C. 1989. Shopping trip behavior: an empirical investigation. Marketing Letters, 1(1): 55-69.

Larson, J. S., Bradlow, E. T., & Fader, P. S. 2005. An exploratory look at supermarket shopping paths. International Journal of Research in Marketing, 22(4): 395-414.

Lee, L., & Ariely, D. 2006. Shopping goals, goal concreteness, and conditional promotions. Journal of Consumer Research, 33(1): 60-70.

Lenzerini, M. 2002. Data integration: a theoretical perspective, Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems: 233-246. Madison, Wisconsin: ACM.

Liao, S.-H., Chu, P.-H., & Hsiao, P.-Y. 2012. Data mining techniques and applications - A decade review from 2000 to 2011. Expert Systems with Applications, 39(12): 11303-11311.

Linoff, G. S., & Berry, M. J. 2011. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management (3rd ed.): Wiley.

Min, H. 2006. Developing the profiles of supermarket customers through data mining. The Service Industries Journal, 26(7): 747-763.

Ngai, E. W. T., Xiu, L., & Chau, D. C. K. 2009. Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36(2, Part 2): 2592-2602.

Phan, D. D., & Vogel, D. R. 2010. A model of customer relationship management and business intelligence systems for catalogue and online retailers. Information & Management, 47(2): 69-77.

Rahm, E., & Do, H.-H. 2000. Data Cleaning: Problems and Current Approaches. IEEE Data Engineering Bulletin, 23(4): 3-13.

Raorane, A., Kulkarni, R., & Jitkar, B. 2012. Association Rule–Extracting Knowledge Using Market Basket Analysis. Research Journal of Recent Sciences, 1: 19-27.

Sarantopoulos, P., Theotokis, A., & Pramatari, K. 2014. Identifying Shopping Missions by Mining Point of Sale Data. In D. Grewal, J. Nordfält, & A. L. Roggeveen (Eds.), Shopper Marketing: In-store, On-line, Social, and Mobile Conference. Stockholm.

Särndal, C.-E., Swensson, B., & Wretman, J. 2003. Model Assisted Survey Sampling (Springer Series in Statistics): Springer.

Shahbaba, B. 2012. Data Exploration, Biostatistics with R: 17-59: Springer New York.

Shrivastava, A., & Sahu, R. 2007. Efficient Association Rule Mining for Market Basket Analysis. Global Journal of e-Business & Knowledge Management, 3(1): 21-25.

Wang, Y., & Zhou, T. 2013. Research of Data Mining in Customer Relationship Management. In Y. Yang, & M. Ma (Eds.), Proceedings of the 2nd International Conference on Green Communications and Networks 2012 (GCN 2012): Volume 2, Vol. 224: 163-170: Springer Berlin Heidelberg.