Assortment planning through store clustering

Mariangela Rossi

ANR: 121777

MASTER OF SCIENCE IN COMMUNICATION AND INFORMATION SCIENCE,

MASTER TRACK DATA SCIENCE BUSINESS AND GOVERNANCE,

AT THE SCHOOL OF HUMANITIES

OF TILBURG UNIVERSITY

Thesis committee:

Max M. Louwerse

Pieter Spronck

Tilburg University

School of Humanities

Department of Communication and Information Sciences

Tilburg center for Cognition and Communication (TiCC)

Tilburg, The Netherlands

July 2016

Preface

I would like to thank the Professor Max M. Louwerse for supervised me in this experience by giving me several suggestions especially regarding the academic approach to follow.

I would like also to thank the Professor Max M. Louwerse and the Associate Professor Pieter Spronck for the time spent in evaluating my work.

Moreover, I would like to thank SAP for providing me the data set for this work and for giving me the opportunity to experience the working environment at the biggest multinational software corporation.

I would like to thank Jan Theodoor for giving me the opportunity to work in his Innovation team at SAP, Marcel de Bruin for his technical supervision about SAP technology and Dave Veerman for his supervision in the retailing domain.

Last but not least, I would like to thank my parents that gave me the opportunity to experience the international and educational environment here in Netherlands at Tilburg University.

Abstract

Despite the frequent use of clustering analysis there are some drawbacks that still affect the quality and the stability of the results of this type of analysis. The main drawbacks are addressed in this research, they are the following: the K-Means initialization problem, the high computational effort required by the hierarchical clustering techniques and the intrinsically difficult clusters evaluation. These drawbacks usually arise when using clustering algorithms individually. To address these issues a new version of two stage clustering is proposed in this work. The procedure consists in a combination of Ward's Method with K-means and SVM. While the Ward's Method has the function of finding the right estimation of the expected number clusters, of initializing the K-means with its centroids and, therefore, of making the obtained clusters more stable, the SVM algorithm is applied to evaluate the accuracy of the clusters obtained by the K-means. This procedure is applied on a data set about 240 stores of an international retailer and it has the objective of showing the strict connection between store clustering and assortment planning. From the analysis three different clusters of stores were defined in each time period: best-selling cluster, worst-selling cluster and average-selling cluster. Each cluster shows how every category of products behaves in every period. Overall, we found out that from 2014 to 2015 there was a general improvement in the selling performance of the retailer especially due to the performances of worst-selling cluster and to the average-selling cluster. In Season 1 (Winter) and Season 2 (Spring) the general trend was also positive, despite the decrease in the selling performance of the worst-selling cluster.

Contents

Introduction:

During the last years the cost of data storage faced a considerable reduction. This reduction boosted the computer performance and yielded to the popularity of computer networks. Consequently, huge volumes of information started to be produced. Such big amount of data needs mining techniques able to handle the size of this type of data and moreover able to provide useful insights from it (Chi & Yang, 2008).

Cluster analysis is one of these data mining techniques on which both academic researchers and academic applications researchers rely on developing meaningful clusters. The term cluster analysis includes different algorithms and methods for grouping objects of analogous nature into specific categories. To be able to develop those taxonomies, researchers define the groups in such a way that the degree of association between elements that belong to the same category is maximal and it is minimal if they do not belong to the same cluster. This degree of association depends on similarity or distance measures. Overall, the greater the similarity (or homogeneity) within a group and the greater the difference among clusters, the better or more distinct the clustering is (Mooi & Sarstedt, 2010). Clustering analysis does not have the objective of explaining why there are structures in the data but it has the objective to find these structures. This is why it is generally applied in the preparatory phase when there are no priori hypotheses. (Shmueli, Patel & Bruce, 2007). Cluster analysis is also called unsupervised analysis since the samples given to the learner are unlabeled. This means that there is absence of information about how many clusters are expected (Shmueli, Patel & Bruce, 2007).

Once these groups are made the next step is, often, to use them to execute inferences and therefore to predict how new instances will behave according to the original groupings. Groups can be made by any kind of objects such as people, patients, products, consumers, stores, and many others depending on the domain of data set. This flexibility makes this technique useful and powerful at the same time (Chi & Yang, 2008).

Retailing is among the fields in which clustering was, so far, applied the most. In retailing, clustering techniques are often applied to dive into the relationships between retailers and their customers.

Vassilikopoulou, Siomkos and Mylonakis (2005), for instance, analyzed the attitude of 341 consumers towards Corporate Social Responsibility. As shown also by Creyer (1997), the opinion of the customers about the corporate behavior highly affects the buying behavior of the customers, which makes this a relevant topic for every firm independently on whether they are selling products or goods. In Vassilikopoulou, Siomkos and Mylonakis (2005) consumers were clustered in three different categories (ambitious CSR consumers, fanatic CSR consumers and passive CSR consumers). Their results, along with the Creyer (1997) ones, show that consumers are willing to pay higher prices if the firm is socially responsible and only want to pay lower prices if the firm behaves irresponsibly. This was just an example of consumers' choices that can directly affect the businesses behavior.

Mendes and Cardoso (2006) noticed that there are also other forces that can affect the businesses behavior. Interestingly, they noticed that in that period the retailing sector was facing a restructuring phase that was challenging many retailers. This phase can be described by the arise of factors such as rising of consumer mobility, boosting of the e-commerce, changes of the household size, escalating consumers expectations etc. These are all factors that still affect the retailing sector nowadays. Therefore, when they clustered the outlets of a supermarket chain in Portugal, they took into account these forces by considering into their store clustering the following features: stores attributes, geographical features, demographic and socioeconomic features. Mendes and Cardoso (2006) showed that, by taking into account those features, the evaluation of stores performance became easier and more effective. Besides, they pointed out that the performance evaluation of the stores affects directly the choices related to the inventory utilization and the pricing strategies which in turn determine the design of the assortment planning. Consequently, by optimizing the stores evaluation they were able to optimize the assortment planning.

The optimization of the assortment planning is a crucial factor for retailers since it brings the expansion in productivity, the expansion of the customers' satisfaction and therefore more revenues. Eventually the retailer that optimize the assortment will be in a position of growth and strength respect to the competitors (Mendes and Cardoso, 2006).

Afterwards, Mendes and Cardoso (2006) used the clusters to predict and, therefore, to evaluate the performance and the location of new potential outlets.

Mendes and Cardoso (2006) are not the only ones that selected the clustering technique for store segmentation. Day and Heeler (1971) and Segal and Giacobbe (1994) selected this technique before them. Moreover, also Schiffman, Bednall, O'Cass, Paladino, Ward and Kanuk, L. (2008), Kolyshkina, Nankani, Simoff, and Denize, (2010) and Bilgic, Kantardzic and Cakir (2015) used clustering to achieve the grouping (these paper are examined in the following paragraphs when discussing the features selected for the store clustering).

The dataset available for this work is about 240 stores of an international retailer. Therefore, by following the literature, here we also decided to use clustering techniques to achieve the store segmentation. More specifically, by following Mendes and Cardoso (2006), we want to study how store clustering affects and boosts the assortment planning.

Besides, because of the unlabeled nature of our data, it was not possible to choose other segmentation techniques, such as classification or correlation. Indeed, the classification technique relies on the fact that, in the training set of data, the instances already have a known category membership. Therefore, since we don't have any pre-information about how many clusters are expected and about the stores labels, we decided to apply clustering analysis to group the stores.

However, in order to evaluate the stability and reliability of our clusters, we compared the clusters with the groups obtained by the SVM, which is actually a classification algorithm. This was possible because we relied on the assumption, made by Mac Queen in 1967, that the partitioning method selected for this analysis (K-means) preserves the underlying structure of the data. Therefore, we were able to consider the labels predicted by the partitioning method as proxies of the unknown natural labels and then we compared them with the labels predicted by the classification problem. Afterwards, we computed the SVM's accuracy. Moreover, to be more sure about the stability of our clusters we used also additional evaluating measures: the between-variance, the within-variance and the Silhouette coefficient. All these procedures/facts and the way of reasoning are explained in detailed in the next section: the Experimental Setup.

Thus, our main objective is to suggest, apply and empirically validate a new effective data mining procedure for store segmentation through which it is possible to boost the evaluation of the assortment planning and therefore to boost the design of the assortment planning itself.

As argued by Tan and Steinbach (2013), despite the frequent use of clustering analysis, when applying clustering techniques individually, really often researchers have to deal with some issues that affect the quality and the stability of the results of the analysis. The most common issues are the following: the choice of the number of clusters that is required while using partitioning clustering (also known as initialization of partitioning methods), the high computational effort required by the hierarchical clustering techniques and the difficulties in the interpretation and validation of the obtained clusters due to the lack of original labels. In our work we suggest a way of addressing these common issues in order to get appropriately clustered instances independently from the domain or data set on which the procedure is applied. In our domain this means that, thanks to this procedure, we are able to get appropriately clustered stores. This is very important since, as also stated by Rajagopal (2011),appropriately clustered stores and channels make the assortment planning more efficient.

A well-designed assortment plan allows the business to (Donofrio, 2009):

- Strengthen the company image while following the marketing strategies.
- Addressing as quick and efficient as possible changes in the customer demand and in the buying behavior.
- Facing the competition.
- Arranging the inventory on customers' needs.

The first step to achieve a well-designed assortment planning, as we saw from Mendes and Cardoso (2006), is to make the right choice of the variables that we want to include in the store clustering.

These variables can be divided in two different types: performance and non-performance variables. Performance clustering means grouping together the businesses according to performances features. For instance, this approach puts in the same cluster the stores with analogous sales performances. Performance clustering is also the one that takes into account competition information when it is available. Non-performance clustering consider features such as customers demographics (i.e. ethnicity, disposable income, age groups, buying preferences etc.) but also climate features, store size, number of employees, stores type. (Donofrio, 2009).

Demographic and Store features:

Day and Heeler (1971) showed how store characteristics can be used to cluster stores into homogenous strata. The analysis is set to test the effects on sales of three different price levels of a new food product. The stores are located in an average city in Middle America. Day and Heeler (1971) got homogenous clusters from a population of 58 stores available. The implementation of the clustering was realized by using attributes such as store sales volume, selling area of each store in square feet, number of employees per store, customers' demographics (e.g. average households' income of the customers, ethnic class) and other attributes that were expected to be correlated with the future sales.

Demographic, Geographic, Behavior and Psychographic features:

According to Kolyshkina, Nankani, Simoff, and Denize, (2010) and to Schiffman, Bednall, O'Cass, Paladino, Ward and Kanuk, L. (2008) the main categories of features for stores segmentation are the following: demographical features of the customers (such as age, number of households, size of the households, life cycle and job occupation, education), geographical features of the area in which the stores are placed (such as provinces, regions, countries, touristic place or not), behavioral features of the customers (such as product knowledge, usage, attitudes and response) and psychographic features (lifestyle, life values and personality). A similar selection of attributes for stores clustering has also been implemented , as we discussed earlier, by Mendes and Cardoso (2006) and also by Bilgic, Kantardzic and Cakir (2015).

One of the challenges for supermarket chain companies operating in different regions is how to create marketing strategies for each store especially when the company does not collect information about its customers. In order to solve this problem Bilgic, Kantardzic and Cakir (2015) took into account the demographic features of the population living in the trade area in which 73 stores of a Turkish supermarket chain company were located (e.g. the age, the marital status and the level of the education). This means that the population was considered to be a potential customer of the supermarket chain so that it was used as a proxy for the customers. This data was provided by the National Statistical Institute of Turkey. Moreover, by following Segal and Giacobbe (1994), they included also information about the presence of competitors close to their stores.

Competition features:

The way of competing nowadays is different from the past. Modern competition depends on productivity, it does not depend anymore on the availability of inputs, or on the size of the business. Productivity relies on the way which enterprises competes while the specific field in which they compete in is not relevant. Companies can be highly productive in any business as soon as they employ efficiently high technologies and offer original products and services. Therefore, nowadays the dynamicity of the competition pushes every company in employing advanced technology (i.e. also advanced assortment planning devices) and it also pushes companies in being knowledge intensive in order to put itself in a position of strength (Porter, 2007).

Segal and Giacobbe (1994) identified in their study a useful methodology to expose basic market segments and to analyze competitive positions. They stated and showed that having information about competition in the store clustering is important especially when combining clustering analysis with competitor analysis. Indeed, according to Segal and Giacobbe (1994) this combination is a channel through which it is possible to improve the target market selection decisions and/or positioning (repositioning) strategies particularly in highly competitive markets.

There are three extensive ways in which store clustering can affect competition: first of all, by increasing the productivity of companies through the optimization of the assortment planning; second, by boosting innovation from which the productivity growth depends; and third, by encouraging the creation of new activities that can broaden and reinforce the cluster itself. Each member of the cluster gains from belonging to it. Indeed, each member can see itself as if it had larger scale, or as if it had joined the other members without being obliged to give up on its flexibility. Businesses that belong to the same cluster can work coordinately with the other members and can be more productive in sourcing inputs, accessing information and technologies, dealing with institutions, assessing and motivating improvements (Porter, 2007).

It is not too difficult to imagine that a successful store clustering needs to consider all the performance and non-performance features discussed so far. However, this means that the company needs to collect this data, thus, it has to be able to (Donofrio, 2009):

- Preserve history and plans of the assortment
- Conserve and update algorithms and the related parameters
- Promote performance clustering
- Promote non-performance clustering
- Arrange review and revision skills

The data set available for this work includes a combination of performance and non-performance features. Indeed it includes: point of sales (POS) data of 240 stores of an international retailer, demographics of the regions where these stores are located and store information (i.e. size of the stores and number of employees per store). The chart below shows in bold the variables included in our data set besides the POS data. A more detailed description about how this data was collected can be found in the next section.

Papers	Variables selected and/or suggested by the Papers
Day and Heeler (1971)	• Selling area of the store • **Number of employees** • **Sales of the company** • **Number of items** • Ethnic groups • Average household income
Mendes and Cardoso (2006)	• **Store size** • Retail composition • Size accessibility • Competition sales area • Competition quality • **Income distribution** • **Demographic data** • Average buy • Clients preferences
Kolyshkina at al. (2010) and Schiffman et al. (2008)	• **Demographical data** • **Geographical data** (e.g. regions, countries, etc.) • Product knowledge • Costumers attitudes • Values • Personality
Bilgic, Kantardzic and Cakir (2015)	• **Store size** • Average rental • Competitor • Marital Status • **Age groups** • Amount of Low educated people

	• Amount of Middle educated people • **Amount of High educated people** • Factory area • University area • Trade area • Touristic area • Car park • Bus service

After the right selection of the variables to include in the analysis, the next step is to choose the right data mining algorithm.

In the next section it follows a more detailed description of the data available for this work, then follows the section about the proposed algorithm for store clustering through which it is possible to boost the evaluation of the assortment planning and therefore to boost the design of the assortment planning itself. Afterwards, it follows the section in which we discussed our results and the conclusion section. The last two sections are the Appendix and the References.

The dataset:

The dataset for this work is provided by SAP and it is about one of its clients. SAP is a German multinational software corporation that creates software for businesses operations and customers relations. The client that provided the data set is an international retailer that center its business on low prices, high volumes, wide variety of products and brands and therefore high flexibility. The products can belong either to the standard assortment or the changing one. The standard products are the ones that are always present in the folder, independently on the time and on the place the store is located (region/country). Only the 30% of the total products belong to the standard range while all the other products change weakly.

The dataset is about point of sales (POS) receipts and point of sales (POS) lines of 240 stores located in Netherlands, France and Germany. This POS data is about the amount sold per category of products and the revenues gained per category of products.

The data is placed in only 6 quarters of the year, in other words one year and half of data are available. It goes from the 1^{st} January of 2014 to the 30^{th} of June 2015. The POS data is stored in one instance in the SAP data center and uploaded in SAP HANA platform which is the in-memory database developed by SAP.

As we discussed already in the introduction, because of the lack of customers' data, Bilgic, Kantardzic and Cakir (2015) used demographic features of the population living in the trade area as proxies of customers' information. Since also in our study the retailer does not collect information about its customers, we collected demographic data and we used it as proxy of customers' information too. The demographic data collected is relative to the regions where the stores are located. This information was collecting by following the selection of variables made by Day and Heeler (1971), Kolyshkina at al. (2010), Schiffman et al. (2008), Mendes and Cardoso (2006) and Bilgic, Kantardzic and Cakir (2015). The demographic features collected are: total population per region, total number of households per region, disposable income per person in the region, amount of people that achieved high level of education per region, unemployment rate in the region and amount of people per age group (the age groups are: 0-15, 15-65, 65+).

Because of the international nature of the retailer the demographics were collected from Statistical Institutes across countries. The data was retrieved from the Central Bank for Statistics of Netherlands (CBS), Federal Statistical Office of Germany (Destatis) and from the French Statistics Institute (Insee). The data was collected at regional level because of the different definition of lower geographical levels (such as municipalities and provinces) applied by each of the Statistical Institutes.

By following the selection of variables made by Mendes and Cardoso (2006) and by Bilgic, Kantardzic and Cakir (2015), we asked to the retailer to provide us also stores data . This data is about store size and number of employees per store. We also collected additional performance data about the amount of standard products sold and revenues gained from standard products per store. This allowed us to make a distinction in the analysis between the general products and the standard ones (which are the ones that never change).

All these external and additional data was uploaded by a SAP specialized expert in the same instance where the POS data was already uploaded. The analysis was code in IDLE (integrated development environment for Python language) and supported by SAP Lumira (which is the tool for data visualization developed by SAP).

Segal and Giacobbe (1994) included in their store clustering also information about the competition. Indeed they showed that competition information directly affect the target market selection decisions and/or positioning (repositioning) strategies. Therefore we also tried to get information about the competitors of the retailer from Gfk and Nielsen, but it was not possible to get them for free. Since our discussion in the previous section about the importance of competition information we suggest for future researches to include this information in the scenario in which it is possible to get this data.

Experimental Setup (Method):

Because of the unlabeled nature of our data, we grouped the stores by using clustering algorithms. As argued by Tan and Steinbach (2013), despite the frequent use of clustering analysis, when using clustering techniques individually, really often researchers have to deal with some issues that affect the quality and the stability of the results of the analysis. The most common issues are the following three: the choice of the number of clusters that is required while using partitioning clustering (also known as initialization of partitioning methods), the high computational effort required by the hierarchical clustering techniques and the difficulties in the interpretation and validation of the obtained clusters due to the lack of original labels.

Let's start with the first issue. The first issue is related to the choice of the number of clusters required to initialize the partitioning methods. This issue puts often the researchers in front of a trade-off. By selecting a high number of clusters it is possible to identify more segments in the data and more differences among these segments, on the other hand, less clusters make the interpretation easier. To address this issue several methods have been proposed so far. A common approach is to use a random initialization of centroids. However, when the random initialization is applied, the partitioning algorithm is ran different times and this typically conduces to different total SSEs making the resulting clusters poor. A technique, that is commonly used to face this drawback of the random initialization, is to perform multiple runs with a different set of randomly chosen initial centroids, and then to select the set of clusters with the minimum SSE. Another possible approach is to update the centroids incrementally so update the centroids after each assignment SSE (Tan & Steinbach, 2013). Although these techniques to choose the initial number of clusters are widely used, the aim of this work is to apply a method through which it is possible to obtain the number of the expected clusters in a more sound way. Therefore, to address the initial centroids selection problem, we want to look for a method that achieves a proper estimation of these centroids instead of just choosing them randomly.

In order to avoid the randomization approach, we decided to apply a two stages clustering method through which we can get the estimation of the expected number of clusters. This method was suggested and applied firstly by Fisher (1987), then by Higgs, Bemis, Watson and Wikel (1997) and also by Meila and Heckerman (2001). Later on, it was also applied by Arai and Barakbah (2007) and by Chi and Yang (2008). All these applications showed that the accuracy and the clustering results improved. Besides, this algorithm is also designed to handle very large data sets, as is the case for the current data set.

Principally, as explained by Chi and Yang (2008), a two-stage clustering method, combines a hierarchical method with a partitioning method. Chi and Yang (2008) stated the combination of a hierarchical method with K-means (which belongs to the family of the partitioning methods) is more powerful than using the two methods individually. Indeed, this combination allows to compensate the first two drawbacks listed above: the higher computational effort required by the hierarchical method (but not required by K-means) and the selection of the initial centroids required by K-means (but not required by the hierarchical method). These are the reasons why the two-stage clustering is more powerful than using partitioning methods or hierarchical methods individually. In the first stage, a hierarchical method is applied. The function of the hierarchical method is to determine (estimate) the number of the initial groups (clusters) and to initialize the partitioning method. Moreover, hierarchical methods are very handy when there is any presence of hierarchy in the structure of the data.

As has been already said above, we want to make the choice of the initial centroids as much reliable as possible, so we chose as hierarchical method for the first stage the Ward's Minimum Variance Method. According to Punj and Stewart (1983), this method have shown superior performance respect to the other hierarchical methods. Moreover, the Ward's method has been already applied specifically for store clustering by Segal and Giacobbe (1994) and by Bilgic, Kantardzic and Cakir (2015) in their studies. This method is called like that because it minimizes the total within-cluster variance. Ward's Method belongs to the family of the agglomerative methods which means that at the first step every point represents a cluster. At each step the objective is to find the pair of clusters that once merged leads to the minimum increase in the total within-cluster variance. This increase is computed as a squared weighted distance among the centers of the clusters. The most important characteristic of the Ward's method is that it considers the loss of information that arises when the observations are merged. Indeed, when each cluster has one observation there is no loss of information, but when the observations are clustered together, then the information about the individual observation is replaced by the general information of the cluster to which it belongs. This loss of information is measured by Ward's Method with error sum of squares (also known as ESS) which measures the difference between individual observations and the group mean. Usually the ESS function takes the form of the Euclidean distance between points (Punj & Stewart, 1983).

Here, the Ward's Minimum Variance Method has the role of making the clusters obtained by the partitioning method more stable and reliable by initializing the partitioning method with the centroids of the Ward's Method itself. Hence, the Ward's Method belongs to the preliminary analysis (first stage) of our algorithm because, in combination with the partitioning method, it determines the candidate number of clusters.

Once the 'k' (number of clusters) has been estimated in the first stage, then, in the second stage, the partitioning method is employed again to implement the actual clustering process and to generate the final clusters of store.

Following Chi and Yang (2008), the partitioning method applied here is the well-known and widely appreciated method called K-means developed for the first time by Mac Queen in 1967. Besides, Pollack (nd.) showed that K-means has already been specifically applied to achieve different kind of store clustering according to the objectives of the different assortment planning (such as channel-based store clustering, sales-volumes based store clustering, store capacity-based clustering, etc.). K-means minimizes a measure of dispersion within the clusters and the variance between the clusters is maximized. In the K-means algorithm choosing the initial centers is the key to get precise results. This is the reason why we use the Ward's Method to initialize the K-means. Indeed, if the initial centers are not chosen properly they can be trapped into local minima easily and lead to incorrect clustering results. This means that determine the number of clusters that are expected is a quite sensitive step. However, since K-means belongs to the family of unsupervised algorithms, usually there is absence of information about how many clusters can be expected (Shmueli, Patel & Bruce, 2007). Therefore, we can state that the initialization by the Ward's Method makes the clusters found by the K-means more reliable. Indeed, according to Punj and Stewart (1983), K-means works very well when nonrandom starting points are used, thus, when the K-means is initialized is more reliable.

The K-means algorithm starts with an initial separation of the observations into "n" clusters. At every step each observation is reassigned to the cluster that has the closest mean to the value of the observation. This step is repeated multiple times. The algorithm only stops when by reassigning the observations it increases the between clusters variance a little but increases also the within variance which we want it to be as small as possible. Therefore this procedure minimizes the variance within each cluster. This ensures that the obtained clusters are homogeneous. The Euclidean distance is used in this algorithm as a measure of within-cluster dispersion of observations from their cluster centroids (Shmueli, Patel & Bruce, 2007).

Once the clusters are obtained the following step consists in post-processing. Post-processing can, for example, consists in merging clusters that have relatively low SSE and that are really similar to each other. However, post-processing can also mean to use the k-means results as other algorithms' initialization. Indeed, here we use K-means in combination with the classification algorithm called Support Vector Machine. This algorithm was proposed the first time by Boser, Guyon and Vapnik (1992) and then re-discussed by Vapnik V. N. and Vapnik V. (1998). Its strong mathematical foundation and high accuracy of the testing made the SVM being extensively implemented. For instance, Support Vector Machine has been implemented for supporting business decisions (Wang, Wu & Zhang, 2005).

The combination between K-means and SVM is due to the fact that, despite clustering techniques have been studied for numerous years, the crucial problem of how getting an accurate evaluation has not been solved yet. Obtaining a solid evaluation of clustering results especially on real data is genetically complicated (Färber et al. , 2010). One way to address this problem is to let the cluster analysis rely on classification algorithms (here SVM) .

When talking about classification in data mining and statistics, researchers refer to the problem of establishing to which category a new observation belongs. The classification procedure is made by two stages. The first one is building the classifier, in this stage the learning process is built. The second one is the stage in which the actual classification takes place. The second stage can also been seen as the predictive stage which uses the classifier of the previous stage for predicting the unseen data. The model is powerful with regard to its generalization abilities, which means that the model should have a good classification accuracy on both train and test sets (Yao et al. , 2013).Therefore, the classification can be done by relying on the fact that in the training set of data the instances already have a known category membership. This is the reason why classification belongs to the family of supervised learning because the learner is trained on a training set of correctly identified observations. Hence, classification is a function through which new observations are assigned to the targeted categories while precisely forecasting the label of the class for each instance of the data set (Shmueli, Patel & Bruce, 2007).

Since the K-means preserves the same distribution and structure of the original data it is possible to use it in conjunction with SVM to evaluate the final clusters.

An example of combination between SVM and K-means can be found in Gad (2016). He proposed a new algorithm called SVM-Kmeans which works as follows: after the data is preprocessed, the K-means algorithm is applied. Then, the important features are selected using Chi-square. Finally, the SVM algorithm is applied.

Following the main approach of Gad (2016) in this work we also aim to turn an unsupervised problem into a supervised one in order to address the fundamental problem of getting a valid evaluation. Indeed, we used the SVM accuracy to evaluate the K-means final clusters.

In Figure 1 and Figure 2 can be found a scheme that visually explains the structure of the algorithm proposed in this work.

As explained above this algorithm needs to be split in two consecutive stages:

Figure 1: First Stage

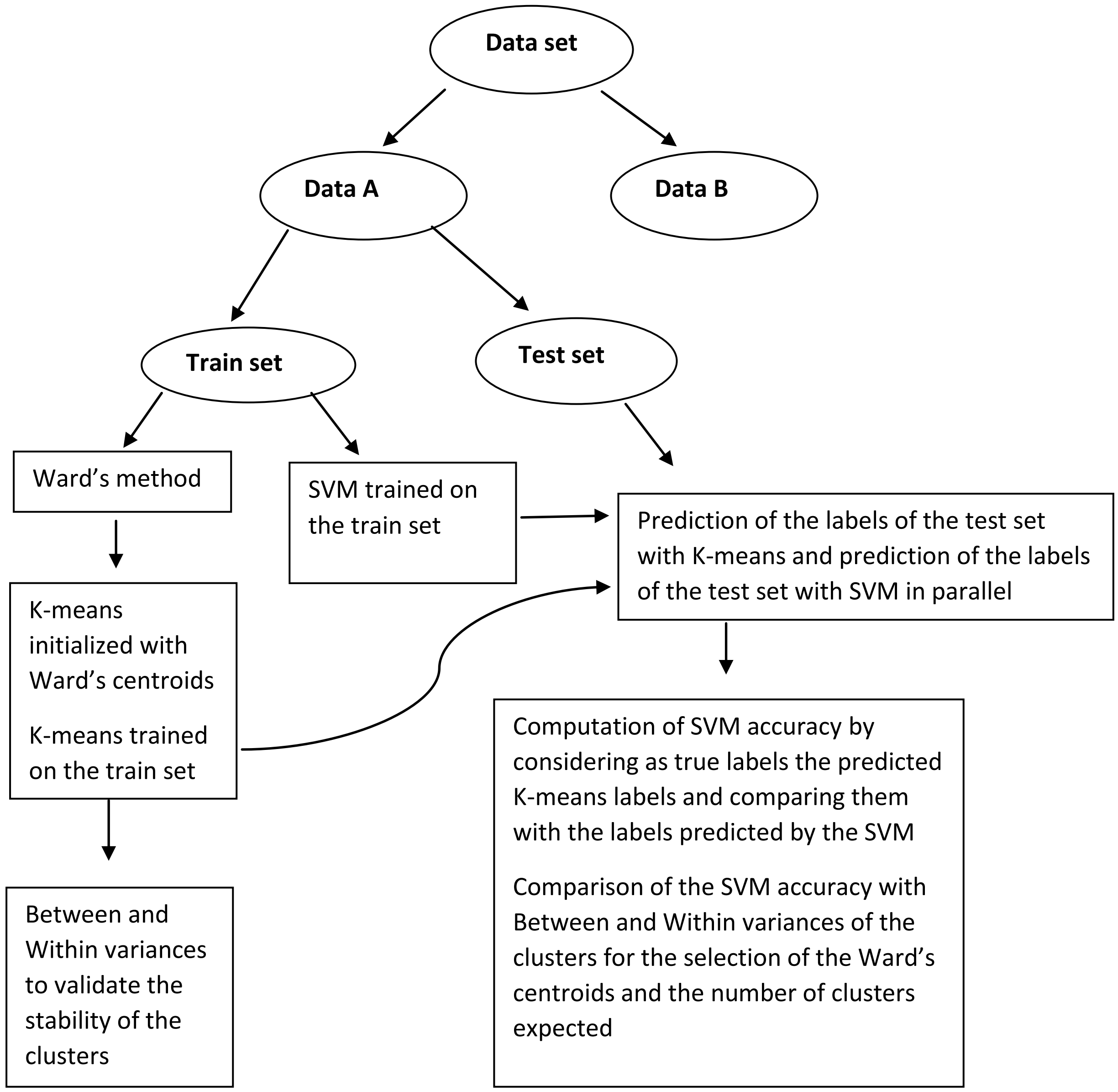


Figure 2: Second Stage

Application of the algorithms to current data set

Support Vector Machine requires all the cases being a vector of numbers. Thus, in case of categorical features, it is necessary to convert them into numeric attributes. Scaling the data is also really important for several reasons. The main benefit in scaling the data is to avoid that features with larger numeric ranges prevail those that belong to smaller numeric ranges. Besides, scaled data allow to avoid numerical complications due to the fact that the kernel values usually rely on the inner products of attributes vectors (Shmueli, Patel & Bruce, 2007).

Therefore, by following the literature, and since K-means also prefers numerical scaled data, the data has been standardized and the categorical features (such as regions and countries) transformed in numbers.

Afterwards, the initial dataset has been divided in two smaller data sets (as is shown in Figure 1 and Figure 2) by using cross validation in Python: Data A and Data B.

The first stage aims to find the right estimation for the expected number of clusters. Data set used in the first stage is Data A. In this stage the Ward's Method, the K means and the SVM are all used to achieve this estimation. To achieve this estimation, Data A has been further split into training and test sets. On the training set the Ward's method is applied. Afterwards the centroids of the Ward are used to initialize the K-means. Once the K-means has been initialized with the Ward's centroids, then K-means gets trained on the training set. The training set is also used to train the SVM. Conversely, the test set is utilized to test both the K-means and the SVM and therefore to predict the labels. The labels predicted by the K-means are being considered as proxy for the true natural labels. We can assume this by relying on the assumption made by the literature (discussed in the above paragraphs) that K-means preserve the underlying structure of the data. Consequently, the labels predicted by the K-means are compared with the labels predicted by the SVM. Thus, the accuracy of the SVM is computed between the predicted labels of the K-means and the predicted labels of the SVM. Therefore, the SVM behaves as an evaluator.

The between-variance, the within-variance and the Silhouette coefficient are computed along with the SVM accuracy with the same purpose of evaluating the stability and the effectiveness of the clusters.

The between-variance is the variance between the clusters. High level of between-variance means that the clusters are well separated among each other. On the other hand the within-variance is the variance among the observations that belong to the same cluster. Low level of within-variance indicate that the observations in the same clusters are very similar to each other and therefore that the cluster is well defined and homogeneous (Shmueli, Patel & Bruce, 2007).

The Silhouette coefficient is a way of measuring the strength of the clusters, it is used to evaluate how well the cases are being grouped. This coefficient is very handy since it puts together the information given by the between and the within variances in an index very easy to read. Indeed, its values are always between -1 and 1. Introduced by Rousseeuw (1987), it is calculated using the mean intra-cluster distance and the mean nearest-cluster distance for each observation. Values close to 1 indicate that clusters are well separated from each other and that the clusters are homogeneous (consequently all the observations are close to the center of the cluster). Values close to -1 mean the exact opposite. So when the values are close to -1 it is possible to conclude that clusters are overlapping or that the observations were assigned to the wrong cluster. Values close to 0 mean that the observations are very close to the decision boundaries. As in Rousseeuw (1987), the Silhouette coefficient in this work is computed with the standard Euclidean distance.

The first stage is run multiple times. The centroids selected in the first stage to initialize the final K-means (the final K-means is ran in the second stage), are therefore the ones that brought high accuracy, high between-variance, high Silhouette coefficient and small within-variance.

After the selection of the expected number of clusters, the second stage starts. In this stage, the final K-means has the role of making the actual store clustering. In the second stage all the data set is used. More specifically, Data A is used as a training set while Data B is used as test set. This means that both final K-means and final SVM are trained on Data A and tested on Data B. The final accuracy is computed following the same approach as in stage 1 and it expresses the overall accuracy of the stores clusters.

As has already said in the description of the data set, the data of this work is placed in only six quarters of the years, from the first quarter 2014 till the second quarter 2015. In other words, we have only one and a half year of data. This means that we do not have all the seasons for the two years and we do not have the two full years to compare. Therefore, we decided to conduct the following analysis:

- *Analysis per quarters*: to get a general overview of the selling trend.
- *Analysis per years*: to analyze the difference in the selling performances between 2014 and 2015. Since we do not have the full 2015 we used as proxy for 2014 the first two quarters of 2014, and as proxy of 2015 the first two quarters of 2015.
- *Analysis per seasons*: to analyze the difference in the selling performances between Season 1(winter) and Season 2 (spring). Since we do not have all the seasons we focused on the seasons that we have in both years. Season 1 is equal to the sum of the first quarter of 2014 and the first quarter 2015, and Season 2 is equal to the sum of the second quarters of 2014 and 2015.

In the next section we are going to discuss the results of our analysis.

Results and Discussion:

As discussed earlier in the introduction, the selection of the variables included in this analysis follows the selection of variables for store clustering made by Day and Heeler (1971), Kolyshkina at al. (2010), Schiffman et al. (2008), Mendes and Cardoso (2006) and Bilgic, Kantardzic and Cakir (2015). The next chart shows these variables. In bold the demographic data collected for this work.

Papers	Variables selected and/or suggested by the Papers
Day and Heeler (1971)	• Selling area of the store • **Number of employees** • **Sales of the company** • **Number of items** • Ethnic groups • Average household income
Mendes and Cardoso (2006)	• **Store size** • Retail composition • Size accessibility • Competition sales area • Competition quality • **Income distribution** • **Demographic data** • Average buy • Clients preferences
Kolyshkina at al. (2010) and Schiffman et al. (2008)	• **Demographical data** • **Geographical data** (e.g. regions, countries, etc.) • Product knowledge • Costumers attitudes • Values • Personality
Bilgic, Kantardzic and Cakir (2015)	• **Store size** • Average rental • Competitor • Marital Status • **Age groups** • Amount of Low educated people • Amount of Middle educated people

	• **Amount of High educated people** • Factory area • University area • Trade area • Touristic area • Car park • Bus service

First of all, we collected demographic data of the population living in the trading area of the stores. This choice was due to the fact that the retailer does not collect any information about its customers. Therefore, by following the reasoning of Bilgic, Kantardzic and Cakir (2015), we used these features as proxies of customers data. The demographics included in the analysis are: population in the region where the store are located, number of households per region, disposable income in the region, amount of people with high education in the region, unemployment rate, amount of people between 0 and 15 years old, amount of people between 15 and 65 years old and people older than 65. All these variables were collected at regional level because of the international nature of the retailer. Indeed, as we already explained in the data set section, the different statistical institutes use different definitions of lower regional levels (such as municipalities and provinces).

The stores features provided by the retailer are: stores size in square meter and number of employees per store, while the performance features are related to the 15 main categories of products available in the data set. These features are: the amount sold per each main category of products and the revenues gained by each main category of products.

In each category has also been taken into account the distinction between all the products (POS) and the standard products. The standard products are the products of the assortment that do not ever change in the folder. This means that they can always be found in the stores independently on the time and on the geographic area.

Figure 3: summarizes all the variables included in the data set and their measures (to make the comparison between clusters easier some of the variables were transformed in percentage respect to the total population of the region).

Features	Measure	Clusters Comparison
Size	Meter squared	
Number of employees	Amount	
Total population per region	Amount	
Total amount of households per region	Amount	% to population

Disposable income per region	Euros	
Amount of people with high educational level	Amount	% to population
Unemployment rate in the region	Amount	
Amount of people between 0 and 15 years old	Amount	% to population
Amount of people between 15 and 65 years old	Amount	% to population
Amount of people older than 65	Amount	% to population
POS amount per each main category of products	Amount	
POS revenue per each main category of products	Euros	
Standard amount per each main category of products	Amount	
Standard revenue per each main category of products	Euros	
Standard rev. + VAT per each main category of products	Euros	

As explained above, the analysis starts with the first stage. The objective of this stage is to find the right estimation of the expected number of clusters. This stage is ran multiple times. Every time it produces a different set of centroids which are used to initialized the K-means. Then the clusters are evaluated with the between-variance, the within-variance, the Silhouette coefficient and the SVM accuracy. The right estimation of the expected number of clusters is the one that brings clusters that reported high SVM accuracy, high between-variance, high Silhouette coefficient and small within-variance.

Indeed, the between-variance is the variance between the clusters. High level of between variance means that the clusters are well separated among each other. The within-variance is the variance among the observations that belong to the same cluster. Hence, low level of within-variance indicate that the observations in the same clusters are very similar to each other and therefore that the cluster is well defined (Shmueli, Patel & Bruce, 2007).

The Silhouette coefficient introduced the first time by Rousseeuw (1987), is a way of measuring the strength of the clusters. This coefficient is very handy since it puts together the information given by the between and the within variances in an index easy to read. Indeed, its values are always between -1 and 1. Values close to 1 indicate that clusters are well separated from each other and that the clusters are homogeneous (consequently all the observations are close to the center of the cluster).

The charts presented below show the values of between-variance, within-variance, SVM accuracy and Silhouette coefficient in Season 1 (winter) and Season 2 (spring) as examples of how the selection of the estimated number of cluster was conducted.

Figure 4: Choice of the estimated number of clusters in Season1

Number_of_Clusters	Accuracy	Between_Variance	Within_Variance	Silhouette_Coefficient
2	1	30,49	3,08	0,69
3	0,93	22,49	2,65	0,36
4	0,93	20,28	2,47	0,39
5**	0,97	18,47	2,32	0,43
6	0,97	21,45	2,81	0,42
7	0,97	20,82	2,74	0,41
8	0,95	19,17	2,56	0,31
9	0,93	18,17	2,43	0,26
10	0,88	17,71	2,37	0,27
11	0,88	17,32	2,32	0,29
12	0,88	20,31	2,76	0,31
13	0,88	19,64	2,67	0,32
14	0,86	19,08	2,60	0,31
15	0,81	18,52	2,53	0,29
16	0,72	17,83	2,44	0,28
17	0,72	18,30	2,53	0,30
18	0,68	17,76	2,46	0,28
19	0,68	17,39	2,41	0,29
20	0,65	17,38	2,41	0,29

Figure 5: Choice of the estimated number of clusters in Season 2.

Number_of_Clusters	Accuracy	Between_Variance	Within_Variance	Silhouette_Coefficient
2	1,00	31,01	3,12	0,70
3	0,97	23,05	2,69	0,36
4	1,00	20,48	2,48	0,38
5**	0,88	18,58	2,33	0,41
6	0,88	16,57	2,14	0,32
7	0,88	19,64	2,63	0,32
8	0,93	19,23	2,57	0,31
9	0,97	18,64	2,50	0,32
10	0,95	18,10	2,44	0,33
11	0,93	17,24	2,32	0,26

12	0,93	20,17	2,74	0,28
13	0,90	19,52	2,65	0,29
14	0,81	18,67	2,54	0,26
15	0,81	19,26	2,65	0,28
16	0,81	18,84	2,60	0,27
17	0,81	18,50	2,54	0,25
18	0,77	18,26	2,51	0,26
19	0,75	17,98	2,46	0,27
20	0,70	17,49	2,40	0,26

In both seasons five clusters emerged.

In Season 1, five clusters were selected because they registered the second highest SVM accuracy, high between-variance, the second highest Silhouette coefficient and the lowest within-variance. Even if two cluster registered the highest SVM accuracy, the highest between-variance and the highest Silhouette coefficient, they were not selected because, in the same time, they reported also the highest within-variance . Moreover when controlling for the support of the two clusters the number of observations per cluster was unbalanced. This means that most of the stores were grouped in only one cluster and only few of them were grouped in the other one. Therefore five clusters turned to be the best choice.

In Season 2 five clusters emerged too. Indeed, they reported very low within-variance and the second highest Silhouette coefficient. Four clusters were not selected because despite the high SVM accuracy and the high between-variance, they reported also high within-variance and low Silhouette coefficient. Two clusters were not selected for the same reasons than Season 1. Moreover, selecting five clusters in Season 2 makes easier the comparison with Season 1.

In the other periods in which the analysis was conducted, the selection of the estimated number of clusters followed the same reasoning showed for Season 1 and Season 2. The tables related to the selections of the other periods can be found in the Appendix.

Once the estimation of the number of clusters was executed, the second stage was ran. In this stage the actual stores clustering was achieved in every period through the final K-means. The obtained final clusters were then evaluated with the final SVM accuracy.

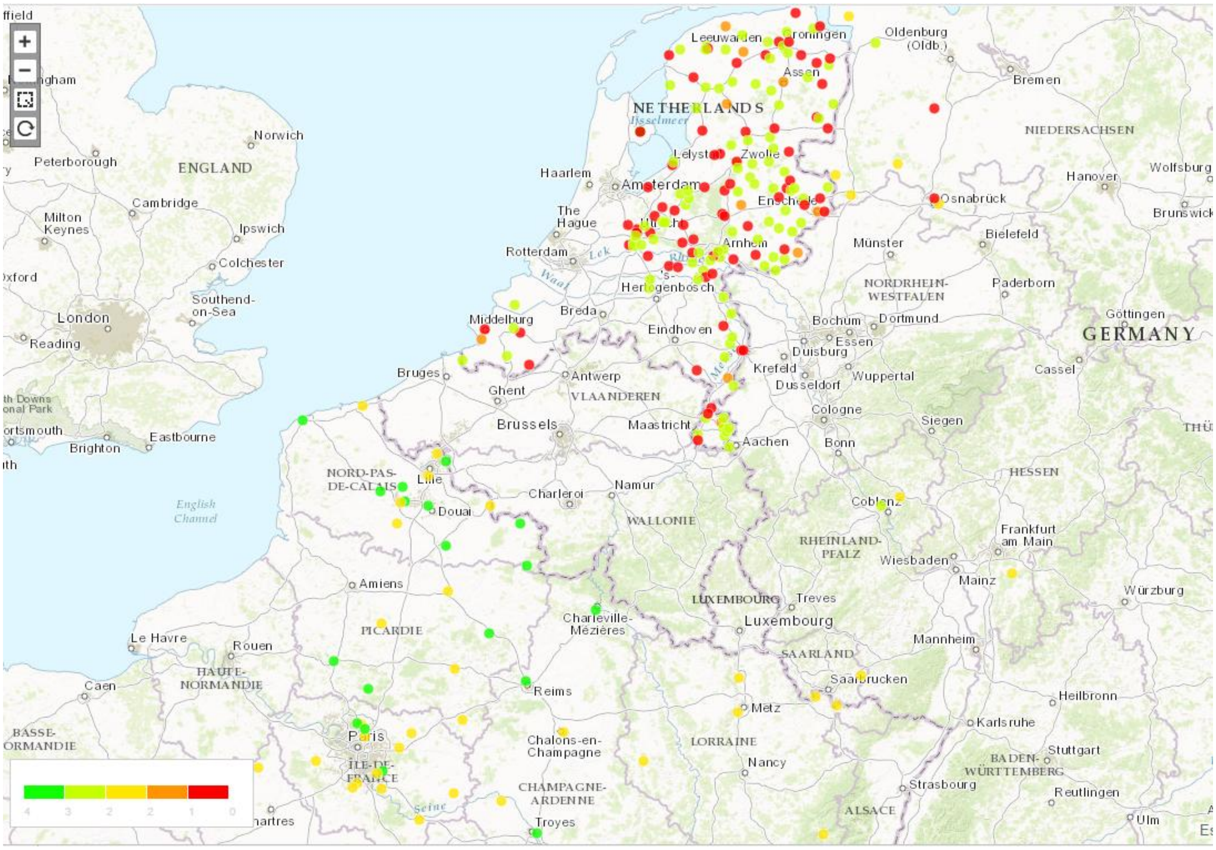
General results:

The first general results tells that independently on the time (quarters, year, season) in which the analysis was focused, the stores were always clustered according to their sales performance. Consequently, two extreme clusters always emerged: the best-selling cluster (in which the stores that sold the most in every category are grouped together) and the worst-selling cluster (in which the stores that sold the least in every category are grouped together). The other stores were grouped into different average-selling clusters between the two extreme. It is important to remember that the data are unlabeled, therefore the labels just described are the ones that came up from the interpretation of the composition of the clusters.

For efficiency reasons, in this study we focused our attention on the differences between the best-selling cluster and the worst-selling cluster, while the average-selling clusters were put together to make one big average-selling cluster and used as point of reference for the other two.

The figure below shows (as an example) the distribution of the final clusters in Season 2 (spring). The distributions of the final clusters in the other periods can be found in the Appendix.

Figure 6 shows the distribution of the clusters in Season 2.



The dark orange dots are the best-selling stores while the yellow dots represent the worst-selling ones. The other dots are the middle-clusters that in our analysis we put all together to make one big average-selling cluster. The final SVM accuracy for the Season 2 clusters is 0.975. The final SVM accuracies of the other periods can be found in the Appendix.

The second general results shows that the best-selling cluster and the worst-selling cluster have always the same (non-performance) characteristics independently from the time period:

Figure 7 shows the results of the non-performance features

Features	Best-selling cluster	Worst-selling cluster
Size	Highest	Low
Number of employees	High/highest	Lowest
Total population per region	Low/lowest	High
Total households per region	Average/high	High
Disp. income per region	High/highest	Average/Low
People with high edu. level	High/highest	Lowest
Unemp. rate in the region	Low/lowest	Average/High
People 0-15	Average	Low
People 15-65	Lowest	Average/High
People 65+	High	High
Country	NL	NL, FR & GE

From the chart we can see that size, number of employees, total population in the region, disposable income and high education level are good explanatory variables. Bigger stores with more employees sold more than smaller stores with less working employees. However, this result is not connected with the population density. Indeed, the best-selling stores are all located in Netherlands which is less densely populated than France and/or Germany. This result can be motivated by the fact that, even if Netherlands is less populated than Germany and/or France, this country is also richer. Indeed, the demographics show that in Netherlands the disposable income is higher than in the other two countries. Moreover, in this country the number of high educated people is higher and the unemployment rate is lower.

The fact that the best-selling stores are located in the richer country can seem obvious in general but in this case it is quite interesting since the target of this retailer is exactly the opposite. Thus, the retailer's target is people with low/average income. One of the possible reasons why the retailer is not totally aware of which kind of people are buying to its stores, is that it does not collect any information about its customers. Consequently, as we discussed in the introduction, by collecting this information the retailer would improve noticeably the selling performances of its stores and also the customers' satisfaction of its customers. Indeed, it will be more capable of meeting the actual demand.

It is also interesting to see that people that spend the most are either very young or old (more than 65 years old). This can explained by several factors discussed also by Williams and Page (2011). People between 15 and 65 belong to the range of people in the working age. However, these are also the people that have to face very high expense: education loans for themselves or for their children, house mortgages, maybe they want to buy also a car etc. These are all examples of large expenses that the other two age groups do not have to face because they are either too young (so they still ask money to their parents when they need to buy something) or old which means that they already faced these expenses.

The last general shows results that independently on the time period and on the type of cluster, the categories the sold the most in both changing and standard products are always the following: Do it yourself, Food&Drink, House&Inventory, Beauty and Cleaning. Conversely, the categories that always performed the worst in both changing and standard products are: Animals and Fun&Multimedia. Moreover, the categories Decorations and Fashion have different behavior in changing and standard products. Indeed, these categories performed both always badly in the standard products and better in the changing ones. These results show in which categories and types of products (changing or standard), the retailer needs to undertake changes in order to improve the categories performances.

Figure 8: summarize the level of the revenues produced by each category (independently from time and type of cluster).

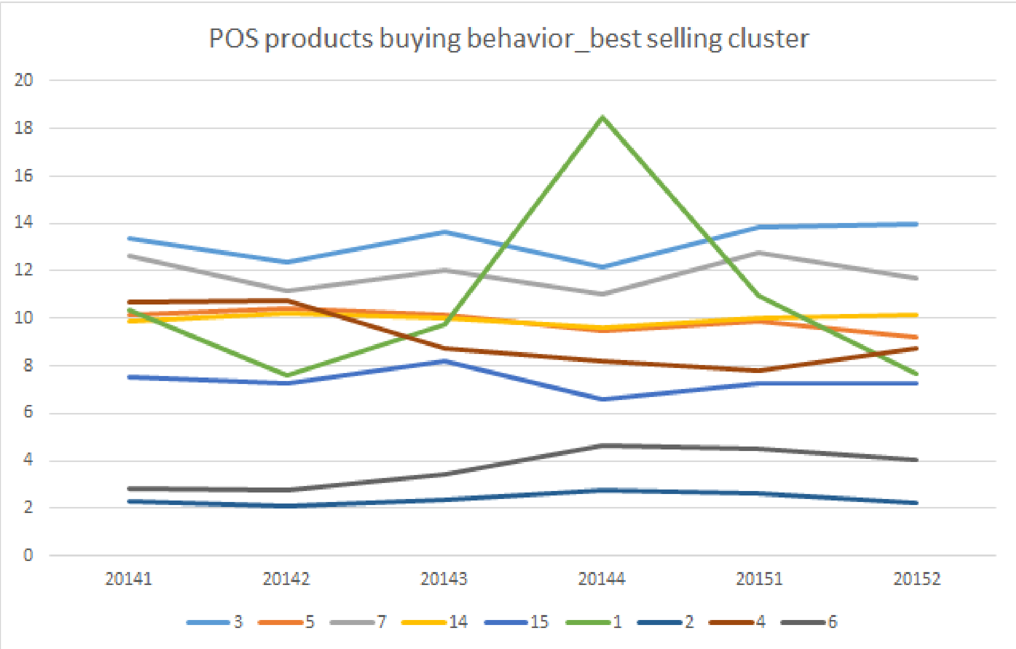
Features	Revenue produced
Do it yourself (POS)	Highest
Do it yourself (Standard)	Highest
Food&Drink (POS)	Highest
Food&Drink (Standard)	Highest
House&Inventory (POS)	Highest
House&Inventory (Standard)	Highest
Beauty (POS)	Highest
Beauty (Standard)	Highest
Cleaning (POS)	Highest
Cleaning (Standard)	Highest
Animals (POS)	Lowest
Animals (Standard)	Lowest
Fun&Multimedia (POS)	Lowest
Fun&Multimedia (Standard)	Lowest
Decoration (Standard)	Lowest
Fashion (Standard)	Lowest

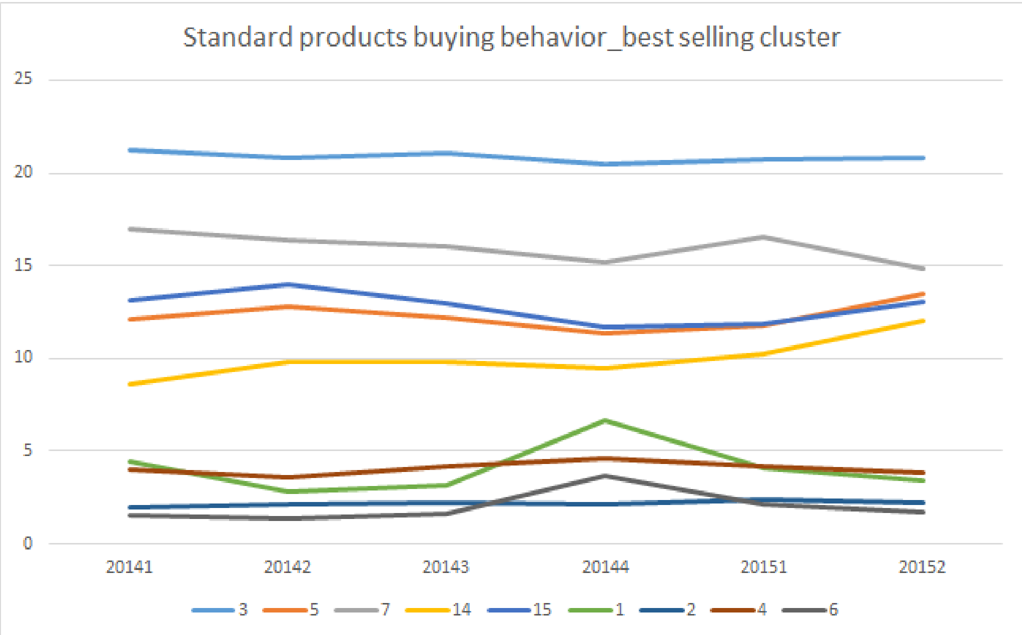
Note: Remind that POS indicates all the kind of products (changing + standard products) while Standard indicates only the products that never change in the folder. However since the standard products are only about the 30% of the assortment, the values of the POS revenues are mainly due to the selling performance of the changing products. So we can consider the trend of the POS as the trend of the changing products.

Analysis per quarters:

The following graphs shows the general buying behavior of the customers of the best-selling cluster along the all periods (the last number in the graphs next to the year indicates the quarter of the year).

The buying behavior of the customers is identified as the percentage of the total revenue achieved by each category. In other words, it indicates how much does the customers spend in each category respect to the total spent.





These graphs confirm what we just discussed in the last general result. Thus, the categories Do it yourself (3), Food&Drink(5), House&Inventory(7), Beauty(14) and Cleaning(15) are, indeed, the ones that always sell the most in both POS (changing) and standard products. Decoration(1) and Fashion(4) perform well only in POS (changing) products, while Animals(2) and Fun&Multimedia(6) perform bad in both changing and standard products.

From the graphs we can see that from summer 2014 to autumn 2014 there was an obvious decrease in the categories usually perform the best and, at the same time, an obvious increase in the minor categories. For this reason we decided to look into the significance of this changes.

The significance of the changes was computed through the Mann-Wilcoxon U Test. This is a powerful non-parametric test for differences in two independent samples of orderable data. The alpha of significance is therefore the alpha from the Mann-Wilcoxon U Test computation (Mann & Whitney, 1947).

All the numerical results the will follow are the results of an average store that belongs to the specified group.

Figure 9 shows the changes in the percentages (customers buying behavior) from 2014_3 (summer) to 2014_4 (autumn) for the best-selling cluster and at which Mann-Wilcoxon alpha level these changes are significant.

Figure 9:

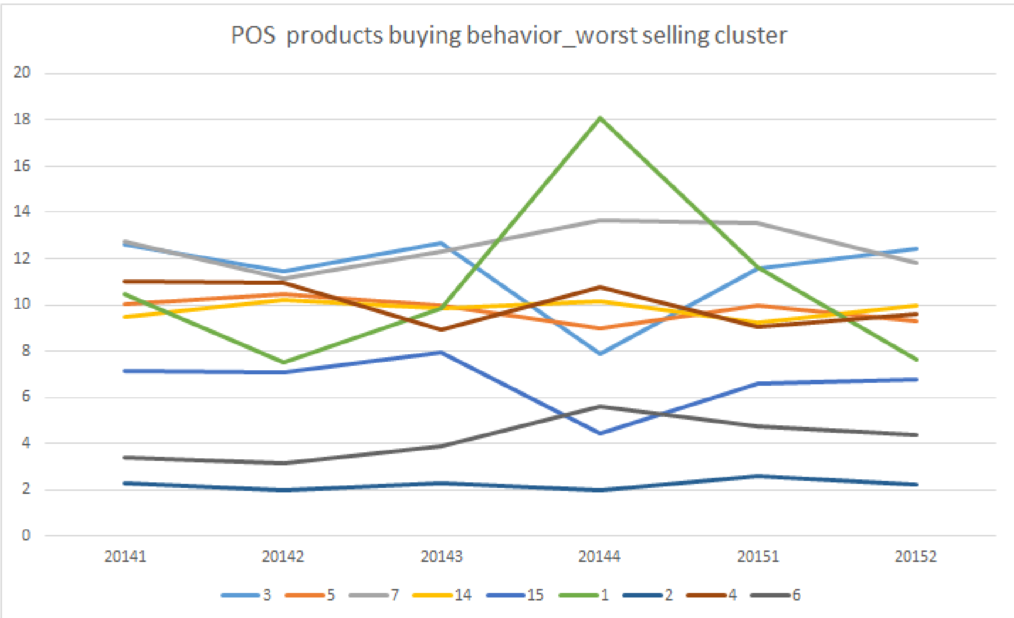
Category	Summer (%)	Autumn(%)	Difference	alpha
Categories that sell the most				
Do it yourself POS	13,75	12,12	-1,63	0,05
Do it yourself Standard	21,01	20,50	-0,51	0,2
Food&Drink POS	10,30	9,51	-0,79	0,1
Food&Drink Standard	12,16	11,32	-0,84	0,1
House&Inventory POS	12,20	11,06	-1,15	0,1
House&Inventory Standard	16,00	15,19	-0,82	>0,2
Beauty POS	10,17	9,60	-0,56	0,2
Beauty Standard	9,83	9,48	-0,35	0,2
Cleaning POS	8,25	6,57	-1,68	0,05
Cleaning Standard	13,00	11,71	-1,29	0,05
Categories that sell the least				
Decoration POS	9,78	18,46	8,68	0,05
Decoration Standard	3,19	6,65	3,46	0,05

Animals POS	2,28	2,77	0,48	0,05
Animals Standard	2,27	2,17	-0,10	0,2
Fashion POS	8,85	8,21	-0,64	>0,2
Fashion Standard	4,22	4,60	0,38	0,05
Fun&Multimedia POS	3,44	4,62	1,18	0,05
Fun&Multimedia Standard	1,62	3,71	2,09	0,05

Note: the POS percentages are computed on the total POS revenues respectively for summer 2014 and autumn 2014, while the Standard percentages are computed on the total Standard revenues respectively for summer 2014 and autumn 2014.

From the chart we can see that, in the transition from summer 2014 to autumn 2014, the categories in which people concentrated (in percentage) their shopping were: Decoration, Fashion (standard) and Fun&Multimedia. This result can be explained by the fact that in autumn there are several important celebrations (such as Christmas, New Eve etc.) that make people concentrate more in "celebration shopping". For instance, in this period people may want to decorate their house (Decoration) or buy presents (Fashion and Fun&Multimedia).

The same trends were registered by the worst-selling cluster. Although, as we can see from the following graphs, these trends were sharper in both POS and standard products than in the best-selling cluster.



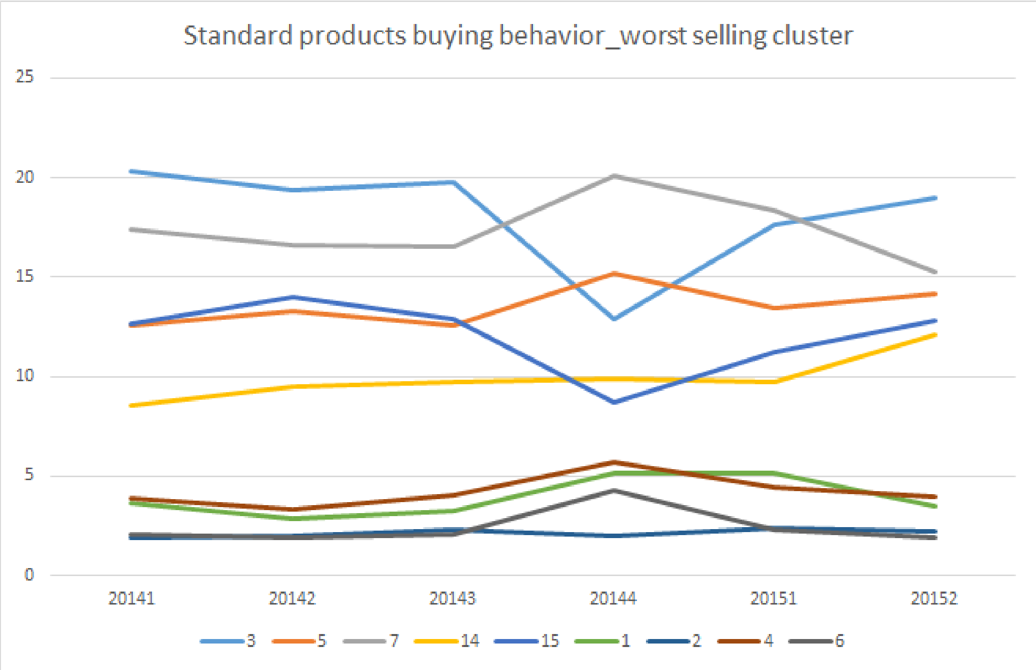


Figure 10 shows the changes in the percentages (customers buying behavior) from 2014_3 (summer) to 2014_4 (autumn) for the worst-selling cluster and at which Mann-Wilcoxon alpha level these changes are significant:

Category	Summer (%)	Autumn (%)	Difference	alpha
Categories that sell the most				
Do it yourself POS	13,35	7,89	-5,46	0,05
Do it yourself Standard	20,91	12,92	-7,99	0,05
Food&Drink POS	9,92	9,01	-0,91	>0,2
Food&Drink Standard	11,02	15,17	4,15	0,2
House&Inventory POS	12,20	13,67	1,47	>0,2
House&Inventory Standard	15,82	20,06	4,25	>0,2
Beauty POS	8,26	10,15	1,89	>0,2
Beauty Standard	8,38	9,93	1,55	>0,2
Cleaning POS	9,48	4,43	-5,06	0,05
Cleaning Standard	13,83	8,71	-5,12	0,05
Categories that sell the least				
Decoration POS	11,21	18,08	6,87	0,1
Decoration Standard	2,79	5,13	2,34	0,05
Animals POS	2,18	1,99	-0,19	>0,2
Animals Standard	2,28	1,99	-0,28	>0,2
Fashion POS	7,85	10,77	2,92	0,1

Fashion Standard	3,60	37,62	34,02	0,05
Fun&Multimedia POS	3,42	5,59	2,17	0,05
Fun&Multimedia Standard	1,89	4,24	2,35	0,05

The table shows that, in the best-selling categories, Do it yourself and Cleaning experienced a significant decrease of the buying behavior especially in the standard products, while Food&Drink POS, House&Inventory and Beauty registered a not significant change (alpha>0,2). Conversely, the less-performing categories show a general increase especially in Fashion Standard and Decoration POS. So we can conclude that also in the worst-selling group people bought the most (in percentage) in the categories that have a composition that fits to the celebration period.

The average-selling cluster respected these trends too. Overall its percentages are closer to the ones of the best-selling cluster, so they are less sharp than the trend registered by the worst selling group. In the Appendix Figure 1a shows these changes.

In order to understand better these percentage changes (i.e. the buying behavior of the customers), we conducted also an analysis on the absolute values of the revenues.

Here we start again with best-selling cluster.

Figure 11 shows the changes in the absolute values of the total revenues gained in 2014_3 (summer) and 2014_4 (autumn) for the best-selling cluster and at which Mann-Wilcoxon alpha level these changes are significant:

Category	Summer Revenues	Autumn Revenues	Difference	% Diff	alpha
Categories that sell the most					
Do it yourself POS	370843,26	382591,13	11747,87	3,17	>0,2
Do it yourself Standard	204651,26	207077,08	2425,81	1,19	>0,2
Food&Drink POS	276617,36	300366,55	23749,19	8,59	0,05
Food&Drink Standard	118481,98	114389,51	-4092,47	-3,45	>0,2
House&Inventory POS	327589,72	349053,68	21463,96	6,55	0,2
House&Inventory Standard	155903,30	153431,38	-2471,91	-1,59	>0,2
Beauty POS	272667,32	303224,68	30557,36	11,21	0,05
Beauty Standard	95705,422	95770,460	65,04	0,07	>0,2
Cleaning POS	222612,21	207422,83	-15189,38	-6,82	0,05
Cleaning Standard	126666,76	118295,57	-8371,19	-6,61	0,05
Categories that sell the least					

Decoration POS	265370,66	582682,18	317311,52	119,57	0,05
Decoration Standard	31042,59	67165,84	36123,24	116,37	0,05
Animals POS	63556,70	87456,99	23900,29	37,60	0,05
Animals Standard	22071,07	21894,93	-176,13	-0,80	>0,2
Fashion POS	238793,68	259169,31	20375,62	8,53	0,1
Fashion Standard	41134,62	46495,68	5361,06	13,03	0,05
Fun&Multimedia POS	93282,05	145799,49	52517,44	56,30	0,05
Fun&Multimedia Standard	15774,34	37446,18	21671,84	137,39	0,05

Overall the absolute values of best-selling categories did not change much from summer to autumn 2014. Indeed, the absolute values of the category Do it yourself, Food&Drink Standard, House&Inventory Standard, and Beauty standard registered a not significant change. However Food&Drink POS, House&Inventory POS and Beauty POS registered a significant increase of more than 6%. However, we saw that in these latter categories the buying behavior registered a decrement. Thus, the buying behavior and absolute values detect opposite trends. Therefore, even if people bought more in autumn than in summer in absolute values, they should have spent even more in order to maintain the same percentage of buying of the summer. Hence, in autumn people bought proportionally less in Food&Drink POS, House&Inventory POS and Beauty POS.

The category that performed the worst in this cluster was Cleaning. Indeed, it registered a significant decrement of more than 6% in both POS and Standard products and in both buying behavior and absolute values.

In autumn people focused their shopping in Decoration, Animals, Fashion and Fun&Multimedia. Their absolute values increased significantly. Decoration registered an increment of more than 115% respect to the summer, in both POS and Standard products. Fun&Multimedia registered noticeable increments too. Thus, for this cluster, the hypothesis of the “celebration shopping” is confirmed also in absolute values.

Despite the shift in the buying preferences, the general result for the best-selling cluster is positive. Indeed, the stores of this group reported an increment in the total POS revenues of 15.99% and an increment in the total Standard revenues of the 3.7%.

Unfortunately it is not possible to conclude the same for the worst-selling cluster. Indeed this group of stores reported a decrement of 48.19% in the total POS revenues and a decrement of 57.56% in the total Standard revenues.

Figure 12 shows the changes in the absolute values of the total revenue gained in 2014_3 (summer) and 2014_4 (autumn) for the group of the stores that sold the least and at which Mann-Wilcoxon alpha level these changes are significant:

Category	Summer Revenues	Autumn Revenues	Difference	% Diff	alpha
Categories that sell the most					
Do it yourself POS	89231,11	28779,83	-60451,276	-67,75	0,05
Do it yourself Standard	48723,36	13498,41	-35224,943	-72,30	0,05
Food&Drink POS	69835,19	32879,01	-36956,176	-52,92	0,05
Food&Drink Standard	27128,09	15844,99	-11283,104	-41,59	0,05
House&Inventory POS	85941,86	49886,57	-36055,292	-41,95	0,05
House&Inventory Standard	38932,00	20959,35	-17972,653	-46,16	0,05
Beauty POS	58193,31	37038,58	-21154,731	-36,35	0,05
Beauty Standard	20621,14	10371,40	-10249,652	-49,70	0,05
Cleaning POS	55998,95	16159,20	-39839,743	-71,14	0,05
Cleaning Standard	34055,42	9099,38	-24956,040	-73,28	0,05
Categories that sell the least					
Decoration POS	78963,12	65980,74	-12982,380	-16,44	0,1
Decoration Standard	6879,67	5360,30	-1519,379	-22,09	0,1
Animals POS	15348,01	7250,15	-8097,849	-52,76	0,05
Animals Standard	5604,99	2082,15	-3522,835	-62,85	0,05
Fashion POS	55308,31	39295,34	-16012,968	-28,95	0,05
Fashion Standard	8850,03	39295,34	30445,314	344,01	0,05
Fun&Multimedia POS	24051,95	20387,50	-3664,444	-15,24	0,1
Fun&Multimedia Standard	4661,79	4433,65	-228,137	-4,89	>0,2

The changes of the absolute values are significantly negative in all the categories, without distinction between categories that usually sell the most or the least. Thus, people bought in absolute values less in autumn than in summer. However, as we saw in the chart of the buying behavior, even if they spent less money they still focused their shopping preferences on the categories that usually perform less well (the decrements in these categories are on average smaller). This again confirms the “celebration shopping” hypothesis made above.

Fashion standard is the only category in this cluster that registered a noticeably significant increase. In this category also the buying behavior increased. So we can infer that people focused especially on buying clothes and fashionable articles.

The average-selling cluster behaved exactly in between the two extreme clusters. Indeed, it reported an overall decrease in the absolute values of the main categories and an overall increase in the absolute values of the minor categories (as registered also by the buying behavior). This again confirms the hypothesis that during this period the customers went for the celebration shopping. Overall, the average group registered an increment of about 7% in the total POS revenues and a decrement of 4% in the total Standard revenues, so the changing products performed better than the standard ones. Figure 1b in the Appendix shows the table of its changes in the absolute values.

Despite the bad performance of the worst-selling cluster, the retailer achieved still a positive balance. Indeed the total POS revenues from all the groups reported a +0.89% from summer to autumn. This means that the increment registered by the cluster with the best-selling stores and the cluster of the average selling stores compensated the loss reported by the worst-selling cluster with. However, since the Standard products closed the quarters with a decrement: -10.37% we can conclude that, in autumn, within the categories, the changing products performed better than the standard products.

Analysis per years:

The analysis starts with the best-selling cluster as above.

Figure 13 shows the changes in the percentages (customers buying behavior) from 2014 to 2015 of the best-selling cluster and at which Mann-Wilcoxon alpha level these changes are significant.

Category	2014 (%)	2015 (%)	Difference	alpha
Categories that sell the most				
Do it yourself POS	12,76	13,86	1,10	0,1
Do it yourself Standard	21,04	20,70	-0,34	>0,2
Food&Drink POS	10,31	9,69	-0,61	0,1
Food&Drink Standard	12,50	12,77	0,27	>0,2
House&Inventory POS	11,74	12,41	0,67	>0,2
House&Inventory Standard	16,69	15,81	-0,88	0,1
Beauty POS	10,08	9,97	-0,10	>0,2
Beauty Standard	9,24	11,08	1,84	0,05
Cleaning POS	7,39	7,43	0,04	>0,2
Cleaning Standard	13,60	12,62	-0,98	0,1
Categories that sell the least				

Decoration POS	8,62	9,59	0,97	0,05
Decoration Standard	3,59	4,21	0,62	0,05
Animals POS	2,17	2,46	0,29	0,05
Animals Standard	2,05	2,32	0,27	0,05
Fashion POS	10,72	8,23	-2,49	0,05
Fashion Standard	3,82	3,99	0,16	>0,2
Fun&Multimedia POS	2,78	4,23	1,45	0,05
Fun&Multimedia Standard	1,44	1,88	0,44	0,05

From 2014 to 2015 the stores of the best-selling cluster experienced a significant increase in the buying behavior related to the categories that usually sell the least. Conversely, despite the slight decreases in the categories Food&Drink POS, House&Inventory Standard and Cleaning Standard, the buying behavior of the best-selling categories did not change much from 2014 to 2015.

Fashion POS is the category that performed the worst in this cluster (-2.49%).

The increments and decrements reflected themselves also in the absolute values. So the categories in which people bought more/less in percentage are actually the same ones in which people bought more/less in absolute values. We can see the changes in the absolute in the next table.

Figure 14 shows the changes in the absolute values of the total revenue gained in 2014 and 20145 of the best-selling cluster and at which Mann-Wilcoxon alpha level these changes are significant:

Category	Revenues 2014	Revenues 2015	Difference	% Diff	alpha
Categories that sell the most					
Do it yourself POS	579095,48	597921,46	18825,98	3,25	0,20
Do it yourself Standard	372309,85	398362,92	26053,07	7,00	0,10
Food&Drink POS	467751,95	418158,72	-49593,23	-10,60	0,05
Food&Drink Standard	221194,38	245811,34	24616,96	11,13	0,10
House&Inventory POS	532796,29	535457,35	2661,05	0,50	>0,2
House&Inventory Standard	295341,67	304285,99	8944,32	3,03	>0,2
Beauty POS	457369,18	430305,74	-27063,43	-5,92	0,10
Beauty Standard	163506,27	213194,67	49688,40	30,39	0,05
Cleaning POS	335494,57	320647,33	-14847,23	-4,43	0,15
Cleaning Standard	240663,50	242930,58	2267,08	0,94	>0,2
Categories that sell the least					
Decoration POS	391092,68	413677,81	22585,12	5,77	0,20
Decoration Standard	63507,05	80991,24	17484,18	27,53	0,05

Animals POS	98372,73	105996,04	7623,31	7,75	0,05
Animals Standard	36210,75	44557,38	8346,62	23,05	0,05
Fashion POS	486428,8	355135,62	-131293,18	-26,99	0,05
Fashion Standard	67657,02	76735,26	9078,24	13,42	0,05
Fun&Multimedia POS	126020,32	182484,11	56463,79	44,81	0,05
Fun&Multimedia Standard	25465,87	36142,67	10676,80	41,93	0,05

Fashion reported the highest decrement in absolute values followed by Food&Drink POS (about -27% and -11% respectively). On the other hand Food&Drink Standard, Beauty standard, Animals Standard and Fashion Standard products reported a large significant increment from 2014 to 2015. Thus, POS and standard products of Fashion and Food&Drink registered completely opposite trends. On the other hand, Fun&Multimedia reported a large significant increment in both POS and Standard products.

Overall the worst-selling categories reported a positive trend while the best-selling categories had a less homogenous pattern with some increments and decrements.

Unfortunately, the overall POS increments are not enough for offsetting the total POS losses. Indeed the best-selling stores registered a decrement of about 5% on the total POS revenues from 2014 to 2015. While, the Standard products registered an increment of about 9% on the total Standard revenues. So we can conclude by saying that overall the standard products performed better than the changing ones.

Let's now check how the worst-selling cluster performed. This cluster performed actually well in all the categories in both percentages and absolute values. In absolute values the customers bought noticeably more in 2015 in all the categories. Conversely, in 2015 the percentages were only slightly larger. Thus, the distribution of the shopping (buying behavior) in 2015 was not too different to the 2014 ones. Moreover, the only changes in the buying behavior that reported a negative sign are all insignificant.

These results are shown in the next two tables.

Figure 15 shows the changes in the percentages (customers buying behavior) from 2014 to 2015 of the worst-selling cluster and at which Mann-Wilcoxon alpha level these changes are significant.

Category	2014 (%)	2015(%)	Difference	alpha
Categories that sell the most				
Do it yourself POS	9,91	12,25	2,33	0,1
Do it yourself Standard	15,64	18,80	3,15	>0,2
Food&Drink POS	10,84	9,90	-0,94	>0,2

Food&Drink Standard	17,22	13,49	-3,73	>0,2
House&Inventory POS	15,45	12,47	-2,97	>0,2
House&Inventory Standard	21,66	16,34	-5,32	>0,2
Beauty POS	9,21	10,20	0,98	0,1
Beauty Standard	7,58	11,67	4,09	0,05
Cleaning POS	4,92	7,49	2,57	0,05
Cleaning Standard	9,57	13,29	3,72	0,05
Categories that sell the least				
Decoration POS	12,53	9,52	-3,01	>0,2
Decoration Standard	5,00	4,16	-0,84	>0,2
Animals POS	1,75	2,53	0,79	0,05
Animals Standard	1,91	2,46	0,55	0,1
Fashion POS	12,54	9,54	-3,00	>0,2
Fashion Standard	3,25	4,68	1,42	0,05
Fun&Multimedia POS	3,63	4,27	0,63	0,1
Fun&Multimedia Standard	1,89	2,06	0,17	>0,2

Figure 16 shows the changes in the absolute values of the total revenue gained in 2014 and 20145 for the group of the stores that sold the least and at which Mann-Wilcoxon alpha level these changes are significant:

Category	Revenue 2014	Revenues 2015	Difference	% Diff	alpha
Categories that sell the most					
Do it yourself POS	83197,39	148258,01	65060,62	78,20	0,05
Do it yourself Standard	46497,68	97232,01	50734,32	109,11	0,05
Food&Drink POS	90966,38	119863,95	28897,56	31,77	0,2
Food&Drink Standard	51206,37	69793,60	18587,22	36,30	0,2
House&Inventory POS	129677,73	151033,62	21355,89	16,47	0,2
House&Inventory Standard	64380,81	84534,36	20153,55	31,30	0,2
Beauty POS	77331,55	123459,60	46128,04	59,65	0,05
Beauty Standard	22522,34	60369,23	37846,89	168,04	0,05
Cleaning POS	41276,32	90659,04	49382,72	119,64	0,05
Cleaning Standard	28448,45	68769,08	40320,62	141,73	0,05
Categories that sell the least					
Decoration POS	105204,44	115322,08	10117,63	9,62	>0,2
Decoration Standard	14862,49	21516,38	6653,89	44,77	>0,2

Animals POS	14681,00	30689,24	16008,24	109,04	0,05
Animals Standard	5690,02	12737,60	7047,59	123,86	0,05
Fashion POS	105255,79	115540,99	10285,19	9,77	>0,2
Fashion Standard	9674,64	24199,70	14525,06	150,14	0,05
Fun&Multimedia POS	30501,67	51675,14	21173,46	69,42	0,05
Fun&Multimedia Standard	5608,88	10642,51	5033,62	89,74	0,05

It is possible to conclude that, in absolute values, this cluster experienced an noticeable increase in the selling performances from 2014 to 2015. Indeed, the total POS revenues increased about the 33% and the Standard revenues increased even more (about 74%). So also in this cluster the revenues from the standard products increased more than the changing ones.

The average-selling cluster also performed well. People bought in all the categories more in 2015 than in 2014. From the buying behavior is possible to see that the customers focused their shopping especially on the following categories: Food&Drink Standard, House&Inventory POS, Decoration, Beauty Standard and Fun&Multimedia POS. Conversely, Do it yourself Standard and Cleaning Standard were the categories in which people bought in percentage less than 2014. As in the best-selling cluster, Fashion performed badly in both POS and standard products. Overall, also this cluster experienced an increase in revenues from 2014 to 2015. Indeed, the total POS revenues increased of about 14% and the Standard revenues increased of about 29%. In the Appendix Figure 2a and Figure 2b show the relative tables.

We can conclude by saying that from 2014 to 2015 there was an overall increment for the retailer in both POS and Standard revenues especially due to the performance of the worst-selling cluster and the average-selling cluster. Moreover the total revenues gained from the standard products reported a higher percentage of increments than the changing ones. Indeed in 2015, if the total POS revenues increased of about 11% , the total Standard revenues increased of the 28%.

Analysis per seasons:

When looking at the absolute values of the revenues gained in Season 1 (winter) and Sesaon 2 (spring), we can see that the best-selling cluster performed better in the best-selling categories (House&Inventory is the only category that showed a contraction). Conversely, it faced several significant decrements in the less performing categories (especially in Decoration and Fun&Multimedia). The same trends are visible also in the percentages of buying behavior. However, in percentage, the preference for the best-selling categories is less evident. The categories that dropped more evidently in both percentage and absolute values are: House&Inventory, Decoration and

Fun&Multimedia. Since these categories faced a decrement in both percentages and absolute values, we can infer that during spring people were less attracted by them.

We can conclude by saying that, in spring, people focused their shopping more in the best-selling categories than in the other ones.

Despite the contractions reported by some categories, the general trend of this cluster is positive. Both total POS revenues and total Standard revenues increased (respectively about 9% and 3%). Thus, the changing products performed better than the standard ones.

The following tables show these results.

Figure 17: shows the changes in the percentages (customers buying behavior) from Season 1 (winter) to Season 2 (spring) of the best-selling cluster and at which Mann-Wilcoxon alpha level these changes are significant:

Category	Winter (%)	Spring (%)	Difference	alpha
Categories that sell the most				
Do it yourself POS	13,74	13,01	-0,72	0,05
Do it yourself Standard	20,89	20,84	-0,04	>0,2
Food&Drink POS	10,20	9,95	-0,25	>0,2
Food&Drink Standard	12,08	13,19	1,10	0,05
House&Inventory POS	12,98	11,39	-1,59	0,05
House&Inventory Standard	16,88	15,60	-1,28	0,05
Beauty POS	9,93	10,18	0,26	0,05
Beauty Standard	9,39	10,98	1,58	0,05
Cleaning POS	7,65	7,27	-0,38	0,05
Cleaning Standard	12,65	13,52	0,88	0,05
Categories that sell the least				
Decoration POS	10,69	7,60	-3,09	0,05
Decoration Standard	4,74	3,11	-1,62	0,05
Animals POS	2,43	2,15	-0,28	0,05
Animals Standard	2,20	2,18	-0,02	>0,2
Fashion POS	9,06	9,96	0,90	0,05
Fashion Standard	4,10	3,72	-0,37	0,05
Fun&Multimedia POS	3,79	3,24	-0,55	0,05
Fun&Multimedia Standard	1,80	1,54	-0,26	0,05

Figure 18: shows the changes in the absolute values of the total revenue gained in Season 1 (winter) and Season 2 (spring) of the best-selling cluster and at which alpha level these changes are significant:

Category	Winter Revenues	Spring Revenues	Difference	% Diff	alpha
Categories that sell the most					
Do it yourself POS	576143,77	600873,19	24729,42	4,29	0,05
Do it yourself Standard	380057,00	390615,78	10558,79	2,78	>0,2
Food&Drink POS	426353,43	459557,25	33203,81	7,79	0,05
Food&Drink Standard	219859,52	247146,21	27286,69	12,41	0,05
House&Inventory POS	542390,95	525862,70	-16528,25	-3,05	0,05
House&Inventory Standard	307197,51	292430,16	-14767,36	-4,81	0,05
Beauty POS	417395,28	470279,65	52884,37	12,67	0,05
Beauty Standard	170944,00	205756,95	34812,95	20,37	0,05
Cleaning POS	320330,87	335811,03	15480,16	4,83	0,05
Cleaning Standard	230139,68	253454,41	23314,73	10,13	0,05
Categories that sell the least					
Decoration POS	453937,17	350833,33	-103103,84	-22,71	0,05
Decoration Standard	86173,55	58324,75	-27848,80	-32,32	0,05
Animals POS	105074,77	99294,01	-5780,76	-5,50	0,05
Animals Standard	39990,26	40777,88	787,62	1,97	0,05
Fashion POS	381640,20	459924,23	78284,02	20,51	0,05
Fashion Standard	74581,99	69810,31	-4771,68	-6,40	0,05
Fun&Multimedia POS	159091,27	149413,17	-9678,10	-6,08	0,05
Fun&Multimedia Standard	32749,58	28858,98	-3890,60	-11,88	0,05

When looking at the absolute values of the worst-selling cluster the trend is less positive than the trend of previous group of stores. Indeed, there is a significant noticeable decrement in the absolute values of most of the categories. On the other hand, the changes in the buying behavior are mostly not significant (especially for the best-selling categories). There are some categories that reported significant decrements even in both percentage and absolute values. These categories are the following: House&Inventory POS, Decoration, Animals POS, and Fashion Standard.

Therefore, in spring, House&Inventory and Decoration registered bad selling performances in both best-selling and worst-selling clusters.

The categories that perform the best in these quarters are Food&Drink Standard and Cleaning Standard. Indeed, the absolute values of their revenues increased respectively of about 14% and 18%.

Therefore, in the worst-selling cluster people bought on average less in spring than in winter. However, the distribution of their shopping preferences among the categories did not change significantly from one season to the other one. The standard products experienced a slightly better trend. Indeed, while the total POS revenues decreased of about 17%, the total Standard revenues increased of 4%. However, this increment (in percentage and in absolute values) is not enough to cover the loss.

We can conclude by saying that, in spring, in the worst-selling cluster, people spent even less money than usual and the few money that they spent was focused on the standard products.

The following two tables show these results.

Figure 19: shows the changes in the percentages (customers buying behavior) from Season 1 (winter) to Season 2 (spring) of the worst-selling cluster and at which Mann-Wilcoxon alpha level these changes are significant

Category	Winter (%)	Spring (%)	Difference	alpha
Categories that sell the most				
Do it yourself POS	8,45	8,01	-0,44	>0,2
Do it yourself Standard	12,27	12,11	-0,15	>0,2
Food&Drink POS	10,21	9,84	-0,37	>0,2
Food&Drink Standard	16,42	17,93	1,51	>0,2
House&Inventory POS	15,29	14,04	-1,25	0,20
House&Inventory Standard	21,21	19,47	-1,75	>0,2
Beauty POS	11,16	10,38	-0,79	>0,2
Beauty Standard	11,74	12,68	0,93	>0,2
Cleaning POS	5,50	5,16	-0,33	>0,2
Cleaning Standard	9,65	10,94	1,29	0,20
Categories that sell the least				
Decoration POS	10,62	8,00	-2,63	0,05
Decoration Standard	4,21	3,61	-0,60	0,10
Animals POS	2,12	1,85	-0,27	0,20
Animals Standard	2,36	2,15	-0,21	>0,2
Fashion POS	9,51	10,69	1,17	0,20
Fashion Standard	4,20	3,26	-0,95	0,05
Fun&Multimedia POS	4,95	4,72	-0,24	>0,2
Fun&Multimedia Standard	2,42	2,20	-0,22	0,10

Figure 20: shows the changes in the absolute values of the total revenue gained in Season 1 (winter) and Season 2 (spring) for the group of the stores that sold the least and at which Mann-Wilcoxon alpha level these changes are significant:

Category	Winter Revenues	Spring Revenues	Difference	% Diff	alpha
Categories that sell the most					
Do it yourself POS	69761,08	54261,66	-15499,43	-22,22	0,05
Do it yourself Standard	34318,15	35409,27	1091,12	3,18	>0,2
Food&Drink POS	84150,23	66640,42	-17509,81	-20,81	0,05
Food&Drink Standard	45940,43	52419,35	6478,92	14,10	0,15
House&Inventory POS	126413,64	95073,13	-31340,51	-24,79	0,05
House&Inventory Standard	59348,16	56893,22	-2454,94	-4,14	>0,2
Beauty POS	92176,18	70294,19	-21881,99	-23,74	0,05
Beauty Standard	32850,73	37050,07	4199,34	12,78	>0,2
Cleaning POS	45610,98	34966,70	-10644,28	-23,34	0,05
Cleaning Standard	26989,77	31961,66	4971,89	18,42	0,1
Categories that sell the least					
Decoration POS	87214,39	54153,21	-33061,18	-37,91	0,05
Decoration Standard	11777,49	10552,58	-1224,91	-10,40	0,15
Animals POS	17804,68	12551,67	-5253,01	-29,50	0,05
Animals Standard	6593,91	6274,03	-319,88	-4,85	>0,2
Fashion POS	78602,23	72384,87	-6217,36	-7,91	>0,2
Fashion Standard	11760,34	9523,33	-2237,01	-19,02	0,1
Fun&Multimedia POS	41073,63	31948,79	-9124,84	-22,22	0,05
Fun&Multimedia Standard	6782,23	6439,38	-342,85	-5,06	>0,2

The average-selling cluster performed less well than the best-selling group, but, like the best-selling group, it performed better on the categories that sell the most and less well in the last categories. On average the categories in which people spent less money in absolute values were also the ones in which the buying behavior deflated. Conversely, the categories that reported higher revenues are also the ones in which the buying behavior grew. So on average there was a clear concentration of the shopping in the best-selling categories.

The general trend registered by the average-selling cluster was positive. The total POS revenues increased from winter to spring of about 3% (which is enough, also in absolute values, to cover the 0.85% decrement of the total Standard revenues). Figure 3a and 3b in the Appendix show the results for this group of stores.

Overall, the positive performance of the best-selling cluster and of the average –selling cluster was more than enough to compensate the negative POS results of the worst-selling cluster. Indeed, the general trend for the retailer from winter to spring was positive. It registered an increase of almost 4% of the total POS revenues and an increase of 1.20% of the total Standard revenues.

Conclusion:

Despite the frequent use of clustering analysis there are some drawbacks that still affect the quality and the stability of the results of the analysis. These drawbacks are: the K-Means initialization problem, the high computational effort required by the hierarchical clustering techniques and the intrinsically difficult clusters evaluation. These drawbacks usually arise when using each of these algorithms individually. Here a new version of the two stage clustering was proposed to address these problems. The procedure consists in a combination of Ward's Method with K-means and SVM. The Ward' Method, together with the K-means, has the function of finding the right estimation of the expected number of clusters and also the function of initializing the K-means with its centroids. This means that the Ward's Method guarantees the reliability of the clusters obtained by the K-means. Besides, the SVM algorithm is applied to evaluate the accuracy of the clusters obtained by the K-means by comparing its stores groups with the stores clusters obtained by the K-means.

This procedure is applied on a data set that includes: POS data of 240 stores of an international retailer, demographics of the regions where these stores are located and store information (i.e. size of the stores and number of employees per store). This data is placed in only six quarters of the years: from the first quarter 2014 till the second quarter 2015. In other words only one and a half year of data is available.

Therefore, three different analyses were conducted:

- *Analysis per quarters*: to get a general overview of the selling trend.
- *Analysis per years*: to analyze the difference in the selling performances between 2014 and 2015. Since we do not have the full 2015 we used as proxy for 2014 the first two quarters of 2014, and as proxy of 2015 the first two quarters of 2015.
- *Analysis per seasons*: to analyze the difference in the selling performances between Season 1(winter) and Season 2 (spring). Since we do not have all the seasons we focused on the seasons that we have in both years. Season 1 is equal to the sum of the first quarter of 2014 and the first quarter 2015, and Season 2 is equal to the sum of the second quarters of 2014 and 2015.

The proposed procedure allowed us to define stable and reliable clusters. Finding stable and reliable clusters is very important because, as shown in the example of Mendes and Cardoso (2006), obtaining an accurate stores clustering means boosting the performance evaluation of the stores which directly affects the shape of the current and future assortment planning. Indeed, the performance of the stores depends on inventory utilization and on pricing strategies which in turn determine the design of the assortment planning. Consequently, optimization of the stores evaluation means optimization of the evaluation of the assortment planning. The concept of assortment planning is very important because from its optimization follows the expansion in productivity, the expansion of the customers' satisfaction and therefore higher revenues for the retailer. Eventually the retailer will be in a position of growth and strength respect to the competitors.

In each period of our analysis we obtained three big clusters. The best-selling cluster in which we can find all the stores that achieved the highest amount of revenues. The worst-selling cluster that includes all the stores that registered the worst selling performance and the average-selling cluster. The average selling cluster is the one in which are placed all the stores that reached levels of revenues in between the two extremes. For time efficiency reasons we focused our analysis on the two extreme clusters and we used the average-selling cluster as point of reference.

The variables included in our data set are: Size, Number of employees, Total population per region, Total households per region, Disposable income per region, People with high educational level, Unemployment rate in the region, Age groups (0-15, 15-65, 65+) and Country where the stores are located. We select these variables by following the selection of variables for store clustering made by Day and Heeler (1971), Kolyshkina at al. (2010), Schiffman et al. (2008), Mendes and Cardoso (2006) and Bilgic, Kantardzic and Cakir (2015).

The first general result tells that size, number of employees, total population in the region, disposable income and high education level are good explanatory variables. Bigger stores with more employees sold more than smaller stores with less working employees. However, this result is not connected with the population density. Indeed, the best-selling stores are all located in Netherlands which is less densely populated than France and/or Germany. This result can be motivated by the fact that, even if Netherlands is less populated than Germany and/or France, this country is also richer. Indeed, the demographics show that in Netherlands the disposable income is higher than in the other two countries. Moreover, in this country the number of high educated people is also higher and the unemployment rate is lower.

The fact that the best-selling stores are located in the richer country can seem obvious in general but in this case it is quite interesting since the target of this retailer is exactly the opposite. Thus, the retailer’s target is people with low/average income. One of the possible reasons why the retailer is not totally aware of which kind of people are buying to its stores, is that it does not collect any information about its customers. Consequently, as we discussed in the introduction, by collecting this information (for instance through loyalty programs) the retailer would improve noticeably the selling performances of its stores and the customers’ satisfaction will increase too. Indeed, it will be more capable of meeting the actual demand.

Interestingly, we also noticed that the people that spent the most were either very young or older than 65 years old. We saw that this can be explained by the fact that people in the working age are the ones that have to face the highest expenses (e.g. paying education loans for themselves or for their children, house mortgages etc.). Therefore, again, in contradiction with the retailer’s target, people that earned/saved more money and have a job spent more in these stores than people with less money. This further result makes even in more evidence the importance of collecting customers' data for meeting the actual demand.

If we could achieve these results by using only the demographics, we can imagine that clustering by including the actual costumers' information will boost noticeably the accuracy of the clusters and therefore the shape of the assortment planning.

The last general result shows that independently on the time period and on the type of cluster, the categories the sold the most in both changing and standard products are always the following: Do it yourself, Food&Drink, House&Inventory, Beauty and Cleaning. Conversely, the categories that always performed the worst in both changing and standard products are: Animals and Fun&Multimedia.

Moreover, the categories Decorations and Fashion have instead different behavior in changing and standard products. Indeed, these categories performed both always badly in the standard products and better in the changing ones.

Therefore, these results show, in general, in which categories and types of products (changing or standard) the retailer needs to make improvements.

From the general analysis per quarters we noticed a remarkable decrease in the buying behavior of the customers between summer 2014 and autumn 2014. These decrement is related to the best-selling categories while, in the same period, the worst-selling categories registered an noticeable increment. We noticed that, in this period, customers shift their preferences on the "celebration shopping" by focusing on categories such as Decoration, Fashion and Fun&Multimedia. This can be explained by the fact that in autumn there are important celebrations such as Christmas and New Year's Eve that affect the shopping behavior of people. This shift in preferences is noticeable especially in the best-selling cluster and the average-selling cluster while the worst-selling cluster performed badly in all the categories.

Despite the bad performance of the worst-selling cluster, the retailer achieved still a positive balance. Indeed, from summer to autumn, the total POS revenues from all the groups reported a +0.89%. This means that the increment registered by the cluster with the best-selling stores and the cluster of the average stores compensated the loss reported by the worst-selling cluster. Conversely, the Standard products closed the quarters with a decrement: -10.37%. This means that, in autumn, the changing products performed better than the standard ones.

From 2014 to 2015 the best-selling cluster experienced an in increase in both buying behavior and absolute values in some categories (such as Decoration, Do it yourself and Fun&Multimedia) but also remarkable decrements in both buying behavior and absolute values in the categories Fashion POS and

Food&Drink POS. These decrements made the cluster register a total decrement of almost 5% in the total POS revenues. Conversely, the Standard revenues registered an increment of almost 9%. The worst-selling clusters performed well in all the categories on both buying behavior and absolute values by registering increments in both total POS revenues and total Standard revenues (+33% and +74% respectively). Thus, in the worst-selling cluster people spent more in 2015 than in 2014. This is true especially for Do it yourself, Animals, Fashion Standard and Cleaning. The average-selling group also performed well. People bought more in 2015 than in 2014 in most of the categories. Fashion POS was the category that performed the worst by reporting a decrease of 7%. The result of this cluster is, in general, positive. Indeed, from 2014 to 2015, the total POS revenues increased of about 14% and the Standard revenues increased of about 29%.

We can conclude that from 2014 to 2015 Fashion was the category that performed the worst among the all clusters. However, the retailer registered a general improvement in the total selling performance especially due to the performances of worst-selling cluster and of the average-selling cluster.

From Season 1 (winter) to Season 2 (spring) the best-selling cluster performed better in the best-selling categories (the only category that showed a contraction is category House&Inventory), while it faced several significant decrements in the less performing categories (especially in Decoration and Fun&Multimedia). Thus, in spring, in this cluster people focused their shopping more on the best-selling categories than in the other ones. However, despite the contractions, the general trend of this cluster is positive. Indeed, both total POS revenues and total Standard revenues increased respectively by 9% and 3%.

When looking at the absolute values of the worst-selling cluster the trend is less positive than the trend of previous group of stores. Indeed, there is a significant noticeable decrement in the absolute values of most of the categories. Moreover, House&Inventory POS, Decoration, Animal POS and Fashion Standard registered decrements also in the buying behavior. This means that people lost interest especially in these categories. However, in the other categories, the customers' preferences did not change much, they just slightly prefer the standard products. Overall, the worst-selling cluster closed the transition from winter to spring with a negative sign. Thus, people bought even less than usual. Indeed, even if the total Standard revenues increased of 4%, the total POS revenues decreased of about 17%.

From winter to spring, the average group of stores performed less well than the best-selling stores group but, like the best-selling group, the average group performed better on the categories that sell the most and less well in the categories that usually sell the least. Moreover, the categories in which people spent less money in absolute values were, on average, also the ones in which the buying behavior deflated. Overall, the general trend for this cluster was positive. Indeed, the total POS revenues increased from winter to spring of about 3% (which is enough also in absolute values to cover the 0.85% decrement of the total Standard revenues).

The positive performance of the best-selling cluster and the average-selling cluster was more than enough to compensate the negative POS results of the less-performing stores group. Hence, the general trend from winter to spring for the retailer was positive with an increase of almost 4% of the total POS revenues and an increase of 1.20% of the total Standard revenues.

With these results we showed how strict is the connection between store clustering and assortment planning. Indeed, we showed, in different time periods, how the categories behaved in each cluster and also if, within the same categories, changing and standard products were behaving differently. Indeed, we showed which categories need to be improved in the future and which ones are already performing well. All these information can be used to optimize the assortment planning.

We can conclude that appropriately clustered stores give the retailer the possibility to get, in an easy and quick way, lot of insights about the design of the current and future assortment planning.

Despite the procedure proposed in this work was trained and tested on a specific data set in a specific domain, it can be easily applied to different situations. Indeed, the combination of the Ward's Method with the K-means and the SVM for the grouping of the data, and the employing of the SVM's accuracy, the between-variance, the within-variance and the Silhouette coefficient for the evaluation of the clusters, can be used also in other domains that differ from retailing. These are, indeed, standard algorithms that have already been applied (individually) to several other fields such as medicine, biology etc. Therefore, their combination can be applied to different fields too. Indeed, the procedure proposed in this work aims to address technical issues that arise when using individually those algorithms, independently on the domain or data set on which it is applied.

Conversely, the selection of the variables, made for achieving the store clustering, is specifically related to the retailing domain. However, this selection can also be followed by other cases that aim to achieve store clustering too. Indeed, performance and non-performance features characterize all the retailers, thus, they do not just characterize the retailer that provided the data set for this work.

We can conclude that not only we reached our main objective of suggesting, applying and empirically validating a new effective data mining procedure for store segmentation (through which it is possible to boost the evaluation of the assortment planning and the assortment planning itself), but we also suggested a procedure that can be applied to other domains too.

Appendix:

Figure 1a: This table shows the changes in percentages (customers buying behavior) from 2014_3 (summer) to 2014_4 (autumn) for the cluster of average stores.

Category	Summer (%)	Autumn (%)	Difference
Categories that sell the most			
Do it yourself POS	10,81	10,25	-0,55
Do it yourself Standard	16,47	17,15	0,69
Food&Drink POS	10,16	9,74	-0,42
Food&Drink Standard	14,46	13,36	-1,10
House&Inventory POS	13,72	11,83	-1,891
House&Inventory Standard	18,60	16,55	-2,06
Beauty POS	11,71	10,40	-1,31
Beauty Standard	10,76	10,32	-0,44
Cleaning POS	7,54	6,18	-1,36
Cleaning Standard	12,35	11,20	-1,14
Categories that sell the least			
Decoration POS	8,97	17,83	8,85
Decoration Standard	2,74	5,80	3,06
Animals POS	2,34	2,59	0,24
Animals Standard	2,47	2,28	-0,19
Fashion POS	9,04	8,64	-0,39
Fashion Standard	3,73	4,63	0,90
Fun&Multimedia POS	3,86	4,78	0,92
Fun&Multimedia Standard	1,97	3,90	1,92

Figure 1b: This table shows the changes in the absolute values of the total revenue gained in 2014_3 (summer) and 2014_4 (autumn) for the group of the average stores.

Category	Summer Revenues	Autumn Revenues	Difference	% Diff
Categories that sell the most				
Do it yourself POS	127220,549	126465,576	-754,974	-0,59
Do it yourself Standard	67407,434	65437,530	-1969,905	-2,92
Food&Drink POS	116766,819	120376,941	3610,122	3,09
Food&Drink Standard	55993,608	50662,152	-5331,456	-9,52
House&Inventory POS	156138,178	146252,590	-9885,589	-6,33

House&Inventory Standard	72303,498	62710,104	-9593,394	-13,27
Beauty POS	133289,275	128793,341	-4495,934	-3,37
Beauty Standard	42363,058	39253,869	-3109,188	-7,34
Cleaning POS	87546,045	76177,301	-11368,743	-12,99
Cleaning Standard	49322,787	42607,659	-6715,128	-13,61
Categories that sell the least				
Decoration POS	103947,757	219453,995	115506,238	111,12
Decoration Standard	11038,724	22032,857	10994,133	99,60
Animals POS	26847,376	31744,326	4896,950	18,24
Animals Standard	9640,016	8653,769	-986,247	-10,23
Fashion POS	103501,528	106538,935	3037,407	2,93
Fashion Standard	14940,085	17572,825	2632,740	17,62
Fun&Multimedia POS	44209,644	59188,621	14978,977	33,88
Fun&Multimedia Standard	7733,029	14846,186	7113,157	91,98

Figure 2a: This table shows the changes in percentages (customers buying behavior) from 2014 to 2015 for the cluster of average stores.

Category	2014 (%)	2015 (%)	Difference
Categories that sell the most			
Do it yourself POS	10,84	10,71	-0,12
Do it yourself Standard	17,53	15,81	-1,72
Food&Drink POS	10,55	10,22	-0,33
Food&Drink Standard	14,43	15,22	0,79
House&Inventory POS	12,57	13,53	0,96
House&Inventory Standard	18,33	17,64	-0,69
Beauty POS	11,05	11,13	0,08
Beauty Standard	9,92	12,38	2,45
Cleaning POS	6,91	6,72	-0,19
Cleaning Standard	12,95	11,88	-1,07
Categories that sell the least			
Decoration POS	7,99	8,77	0,78
Decoration Standard	3,19	3,78	0,58
Animals POS	2,15	2,33	0,18
Animals Standard	2,18	2,40	0,22
Fashion POS	10,81	8,78	-2,03

Fashion Standard	3,49	3,72	0,24
Fun&Multimedia POS	3,22	4,68	1,46
Fun&Multimedia Standard	1,90	2,25	0,35

Figure 2b: This table shows the changes in the absolute values of the total revenue gained in 2014 and 2015 for the group of the average stores.

Category	2014 Revenues	2015 Revenues	Difference	% Diff
Categories that sell the most				
Do it yourself POS	184149,968	212463,131	28313,164	15,375
Do it yourself Standard	112403,638	134323,929	21920,291	19,501
Food&Drink POS	179295,797	198064,792	18768,995	10,468
Food&Drink Standard	91328,882	122651,325	31322,443	34,296
House&Inventory POS	213534,258	261149,390	47615,132	22,299
House&Inventory Standard	116039,346	142945,095	26905,749	23,187
Beauty POS	188285,683	214890,464	26604,781	14,130
Beauty Standard	63251,004	100783,357	37532,354	59,339
Cleaning POS	117360,420	131780,566	14420,146	12,287
Cleaning Standard	82544,808	98478,640	15933,832	19,303
Categories that sell the least				
Decoration POS	135684,253	171406,302	35722,049	26,327
Decoration Standard	20400,860	31467,906	11067,047	54,248
Animals POS	36439,362	45470,017	9030,655	24,783
Animals Standard	13773,104	19567,976	5794,872	42,074
Fashion POS	182678,141	169848,703	-12829,438	-7,023
Fashion Standard	22221,075	30935,069	8713,993	39,215
Fun&Multimedia POS	54357,959	90660,998	36303,039	66,785
Fun&Multimedia Standard	11863,885	18307,179	6443,294	54,310

Figure 3a: This table shows the changes in percentages (customers buying behavior) from Season 1 (winter) to Season 2 (spring) for the cluster of average stores.

Category	Winter (%)	Spring (%)	Difference
Categories that sell the most			
Do it yourself POS	11,49	10,88	-0,61
Do it yourself Standard	17,42	17,19	-0,22

Food&Drink POS	10,48	10,21	-0,27
Food&Drink Standard	13,97	15,04	1,06
House&Inventory POS	13,60	12,11	-1,49
House&Inventory Standard	18,44	16,67	-1,77
Beauty POS	10,76	10,99	0,23
Beauty Standard	10,26	11,77	1,51
Cleaning POS	7,03	6,81	-0,22
Cleaning Standard	11,88	13,04	1,16
Categories that sell the least			
Decoration POS	9,99	7,10	-2,89
Decoration Standard	4,30	2,95	-1,35
Animals POS	2,51	2,14	-0,37
Animals Standard	2,38	2,32	-0,06
Fashion POS	9,23	10,09	0,86
Fashion Standard	3,97	3,37	-0,60
Fun&Multimedia POS	4,18	3,70	-0,48
Fun&Multimedia Standard	2,17	1,97	-0,20

Figure 3b: This table shows the changes in the absolute values of the total revenue gained in Season 1 (winter) and Season 2 (spring) for the group of the average stores.

Category	Winter Revenues	Spring Revenues	Difference	% Diff
Categories that sell the most				
Do it yourself POS	188075,45	184759,37	-3316,07	-1,76
Do it yourself Standard	117844,26	115840,40	-2003,86	-1,70
Food&Drink POS	173334,14	174013,88	679,74	0,39
Food&Drink Standard	95090,14	101115,40	6025,26	6,34
House&Inventory POS	225359,31	206519,91	-18839,40	-8,36
House&Inventory Standard	125268,16	112018,97	-13249,19	-10,58
Beauty POS	178854,21	187722,39	8868,19	4,96
Beauty Standard	69782,49	79262,81	9480,32	13,59
Cleaning POS	115631,53	115559,83	-71,70	-0,06
Cleaning Standard	80558,68	87573,79	7015,12	8,71
Categories that sell the least				
Decoration POS	163871,14	120589,33	-43281,81	-26,41
Decoration Standard	29049,46	19793,79	-9255,66	-31,86

Animals POS	41093,20	36228,82	-4864,38	-11,84
Animals Standard	16088,40	15523,54	-564,86	-3,51
Fashion POS	151739,48	170981,30	19241,82	12,68
Fashion Standard	26810,75	22603,91	-4206,85	-15,69
Fun&Multimedia POS	68780,52	62765,87	-6014,65	-8,74
Fun&Multimedia Standard	14654,59	13158,73	-1495,85	-10,21

Selection of the clusters for 2014:

Clusters	Accuracy	Between_Variance	Within_Variance	Silhouette_Coefficient
2	1,00	26,24	2,65	0,68
3	0,94	19,27	2,30	0,36
4	1,00	18,33	2,19	0,39
5**	1,00	17,04	2,13	0,40
6	1,00	20,26	2,62	0,40
7	0,94	18,73	2,43	0,31
8	0,94	19,57	2,49	0,33
9	0,97	18,39	2,37	0,28
10	0,97	18,10	2,34	0,30
11	0,94	20,37	2,69	0,31
12	0,97	19,76	2,60	0,26
13	0,97	19,54	2,56	0,26
14	0,92	18,72	2,46	0,24
15	0,86	18,81	2,49	0,26
16	0,86	18,93	2,53	0,26
17	0,83	18,57	2,49	0,26
18	0,75	18,00	2,41	0,26
19	0,78	18,34	2,47	0,28
20	0,75	17,92	2,42	0,30

Selection of clusters for 2015:

Clusters	Accuracy	Between_Variance	Within_Variance	Silhouette_Coefficient
2	1,00	31,15	3,14	0,71
3	1,00	23,34	2,72	0,37
4**	0,98	19,94	2,43	0,42

5	1,00	18,18	2,29	0,39
6	1,00	17,05	2,19	0,39
7	1,00	20,43	2,70	0,38
8	1,00	19,18	2,54	0,31
9	0,98	17,83	2,39	0,28
10	0,98	17,35	2,33	0,29
11	0,98	20,37	2,77	0,31
12	0,98	19,71	2,68	0,31
13	0,91	19,09	2,61	0,30
14	0,89	18,61	2,54	0,25
15	0,89	18,53	2,52	0,26
16	0,93	17,98	2,46	0,26
17	0,93	18,61	2,56	0,27
18	0,93	18,20	2,51	0,28
19	0,89	17,77	2,45	0,28
20	0,89	17,36	2,39	0,29

Selection of clusters for 2014_3 (summer):

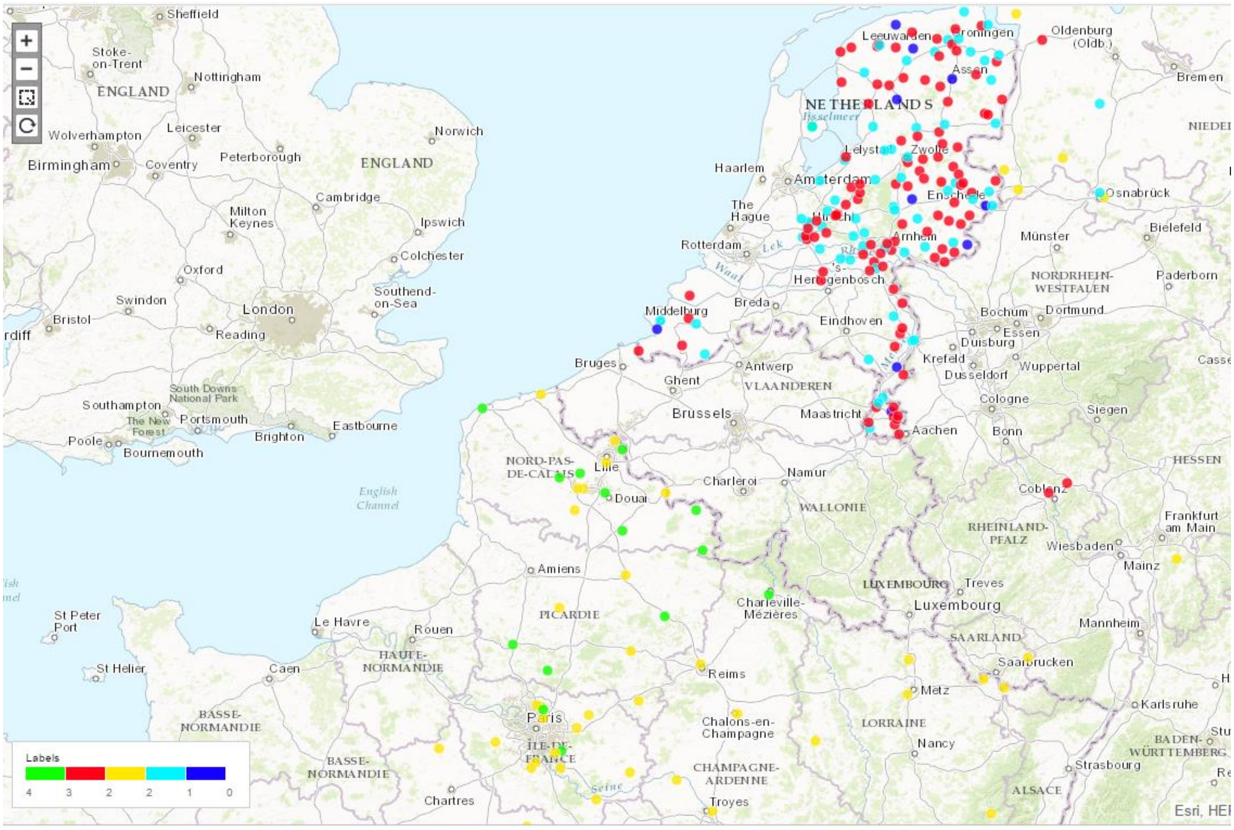
Clusters	Accuracy	Between_Variance	Within_Variance	Silhouette_Coefficient
2	0,97	32,12	3,25	0,72
3	0,95	24,06	2,81	0,33
4**	1,00	21,08	2,54	0,40
5	0,97	20,14	2,53	0,41
6	0,95	25,79	3,32	0,43
7	0,95	23,35	3,07	0,37
8	0,95	23,18	3,07	0,38
9	0,87	21,82	2,90	0,28
10	0,87	20,91	2,80	0,25
11	0,87	20,91	2,81	0,27
12	0,89	19,80	2,67	0,23
13	0,89	19,46	2,63	0,24
14	0,87	19,24	2,61	0,25
15	0,76	18,51	2,50	0,23
16	0,76	18,39	2,50	0,24
17	0,76	19,34	2,64	0,25

18	0,63	18,68	2,55	0,25
19	0,68	18,16	2,47	0,25
20	0,68	17,62	2,40	0,24

Selection of clusters for 2014_4 (autumn):

Clusters	Accuracy	Between_Variance	Within_Variance	Silhouette_Coefficient
2	1,00	29,24	2,97	0,65
3	0,98	23,45	2,72	0,39
4	0,95	19,22	2,41	0,35
5**	1,00	19,04	2,41	0,41
6	0,98	17,33	2,23	0,33
7	0,98	22,32	2,94	0,34
8	0,98	20,75	2,76	0,34
9	0,98	20,53	2,74	0,34
10	0,93	19,17	2,58	0,29
11	0,88	18,58	2,50	0,28
12	0,86	19,93	2,71	0,29
13	0,86	19,54	2,68	0,29
14	0,81	19,02	2,60	0,31
15	0,83	18,54	2,54	0,30
16	0,81	17,92	2,45	0,26
17	0,83	17,69	2,42	0,26
18	0,83	17,23	2,37	0,25
19	0,79	16,69	2,30	0,25
20	0,81	17,25	2,39	0,27

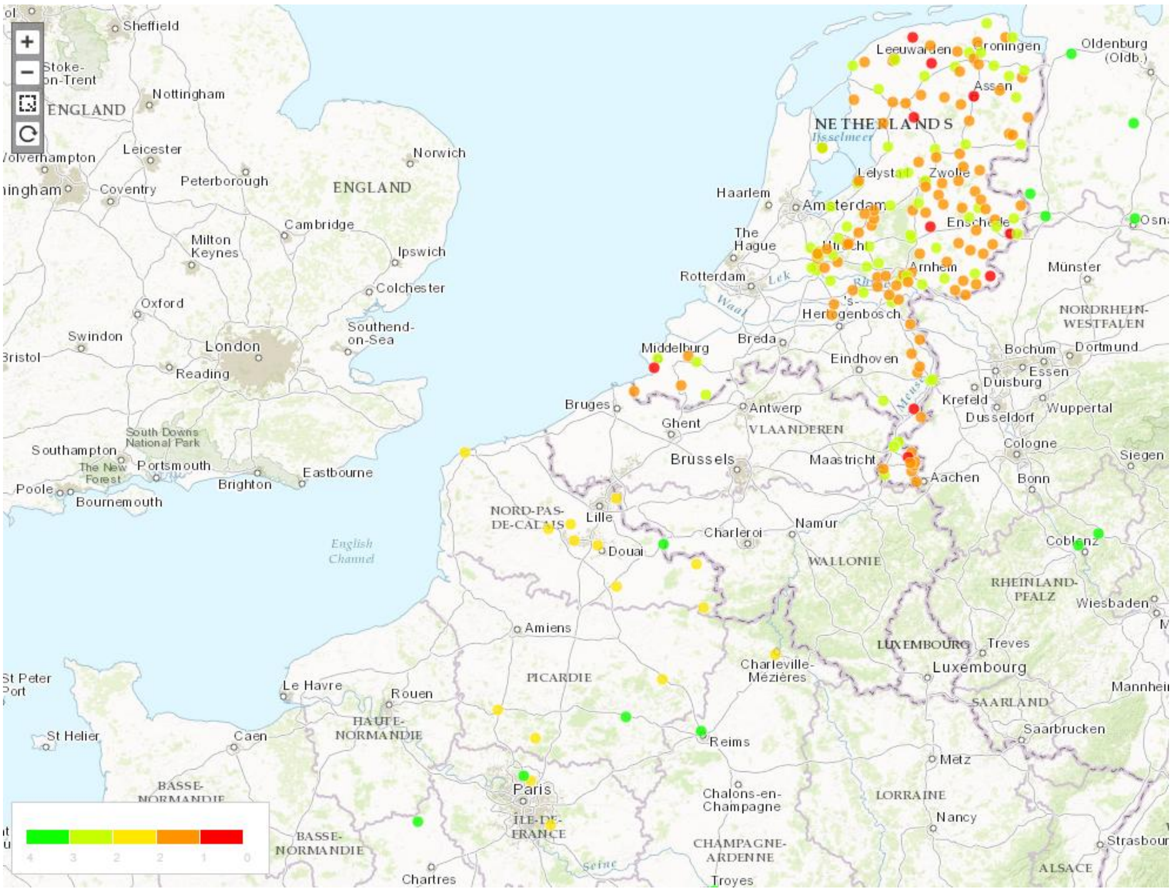
Season 1 clusters distribution:



Blue dots are stores of the best-selling cluster, yellow dots are the stores that belong to the worst-selling cluster and the other colors are the middle clusters.

Total SVM accuracy: 0.98

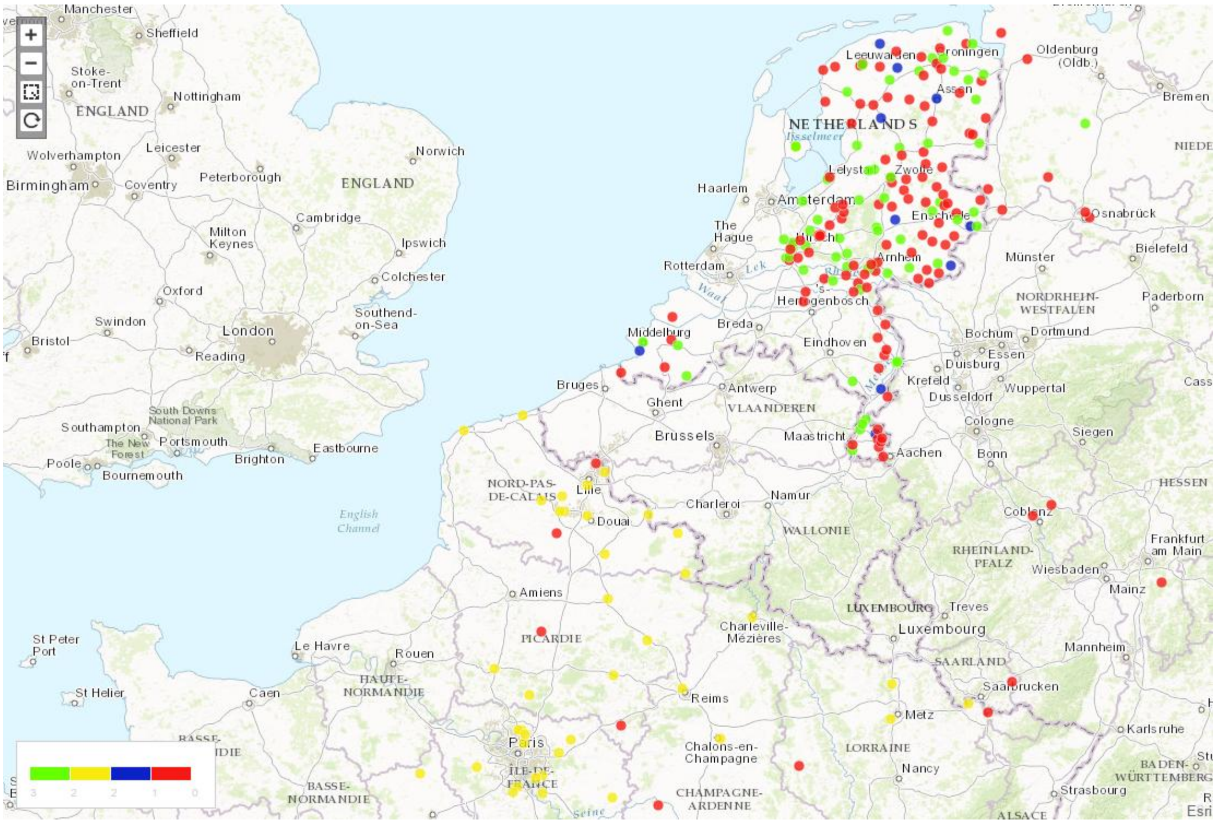
2014 clusters distribution:



The red dots are the stores that belong to best-selling cluster, the green dots are the stores that belong to the worst-selling cluster while the other colors indicate the middle clusters.

Total SVM accuracy: 0.96

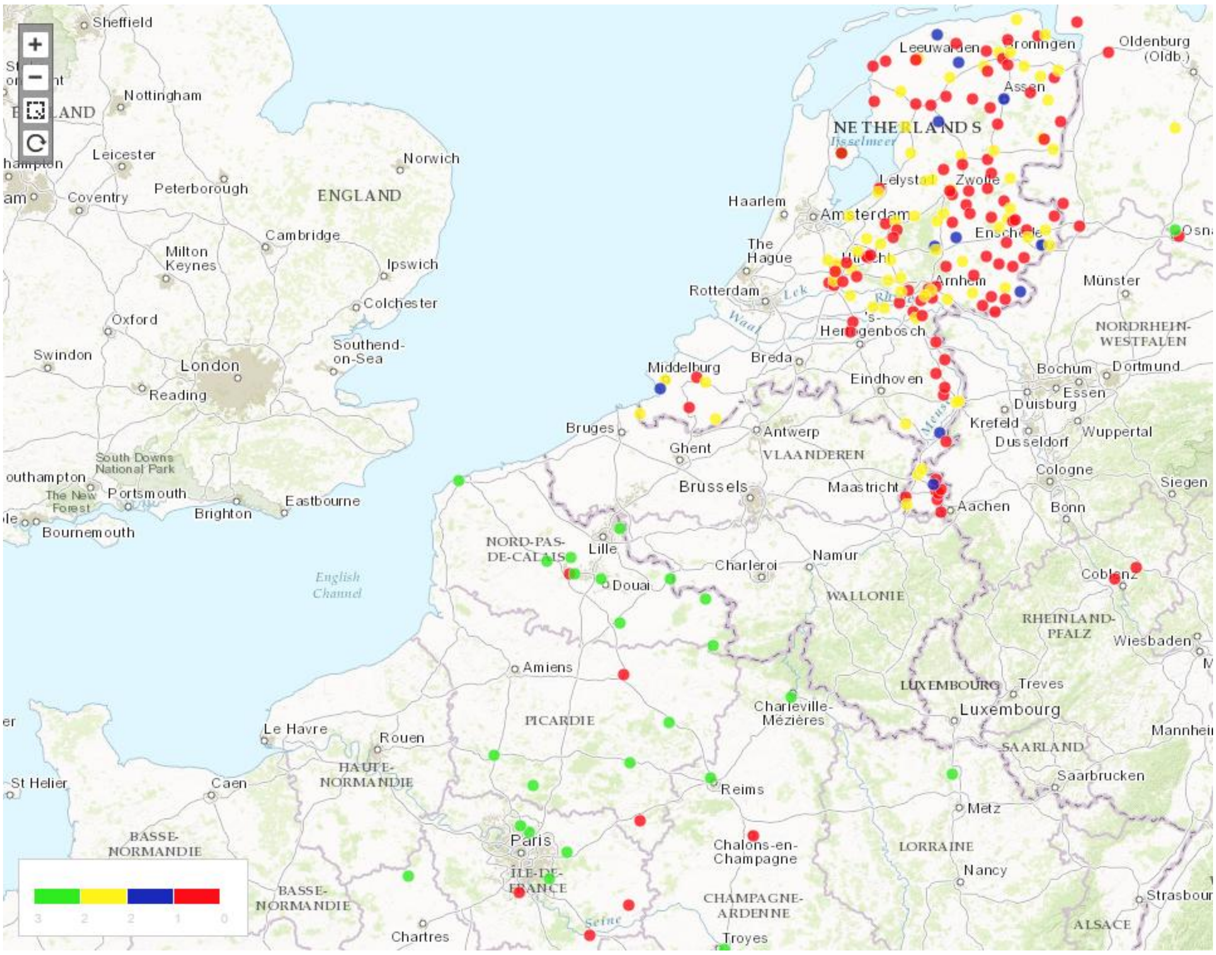
2015 cluster distribution:



The red blue dots are the stores that belong to the best-selling cluster while, the red one are the stores that belong to the worst-selling group, the other colors indicate the middle clusters.

Total SVM accuracy: 0.95

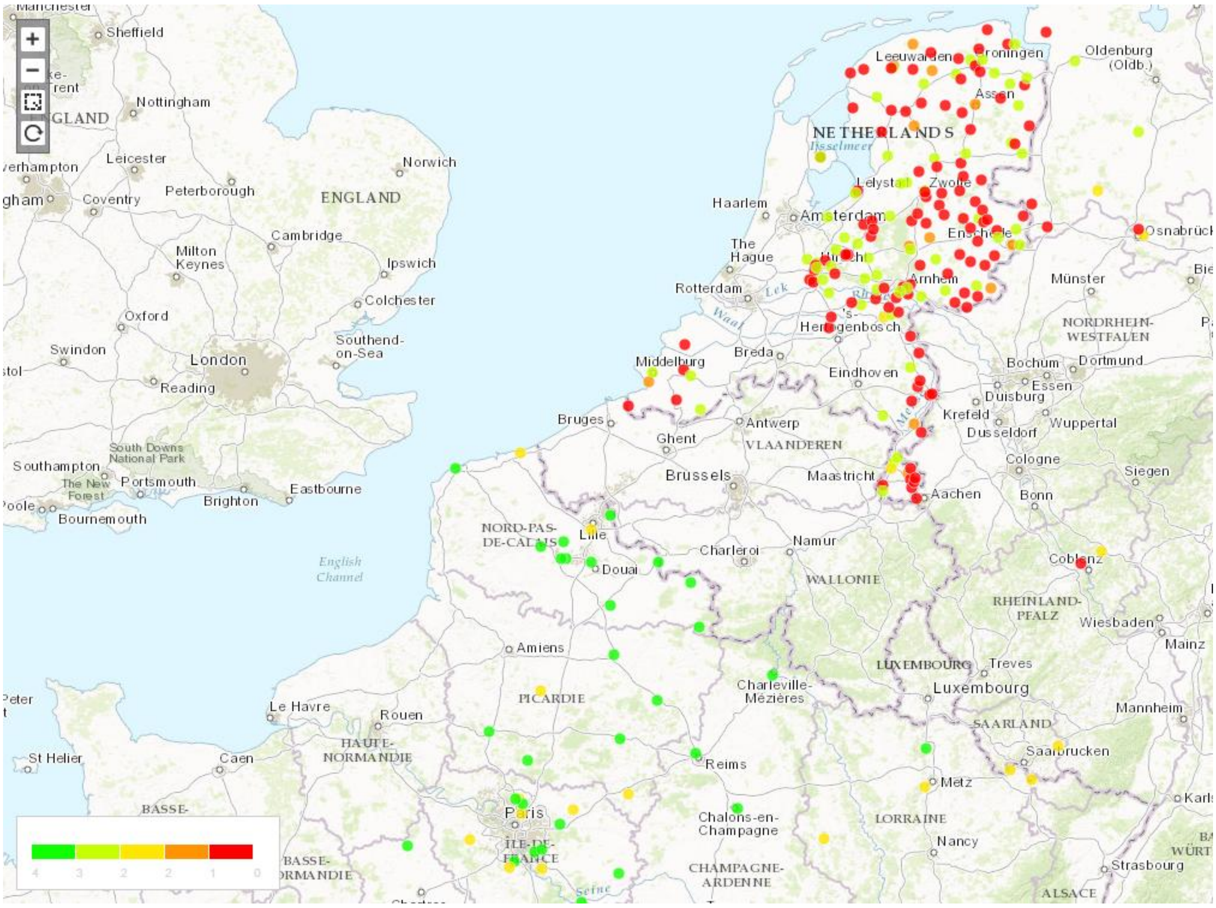
Summer 2014 clusters distribution:



The blue dots are the stores that belong to the best-selling cluster, the red dots are the stores that belong to the worst-selling group. The other colors indicate the middle clusters.

Total SVM accuracy: 0.95

Autumn 2014 clusters distributions:



The orange dots are the stores that belong to the best-selling cluster, the yellow dots are the stores that belong to the worst-selling group. The other colors indicate the middle clusters.

Total SVM accuracy: 0.97

References:

Arai, K., & Barakbah, A. R. (2007). Hierarchical K-means: an algorithm for centroids initialization for K-means. *Reports of the Faculty of Science and Engineering*, *36*(1), 25-31.

Bilgic, E., Kantardzic, M., & Cakir, O. (2015, July). Retail Store Segmentation for Target Marketing. *Lecture Notes in Computer Science Advances in Data Mining: Applications and Theoretical Aspects*, 32-44. doi:10.1007/978-3-319-20910-4_3

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* - COLT '92, 144-152. doi:10.1145/130385.130401

Chi, S. C., & Yang, C. C. (2008). A Two-stage Clustering Method Combining Ant Colony SOM and K-means. *Journal of Computing and Information Science in Engineering*, 24(5), 1445-1460.

Creyer, E. H. (1997). The influence of firm behavior on purchase intention: do consumers really care about business ethics?. *Journal of Consumer Marketing*, *14*(6), 421-432.

Day, G. S., & Heeler, R. M. (1971). Using cluster analysis to improve marketing experiments. *Journal of Marketing Research*, 340-347.

Donofrio, T. J. (2009, September 09). Advanced Planning and Optimization Part 3: Store Clustering. *Retail Info System and Services (RS&S).* Retrieved from http://risnews.edgl.com/retail-news/Advanced-Planning-and-Optimization-Part-3--Store-Clustering38904

Färber, I., Günnemann, S., Kriegel, H. P., Kröger, P., Müller, E., Schubert, E., Seidl, T. & Zimek, A. (2010, July). On using class-labels in evaluation of clusterings. In *MultiClust: 1st international workshop on Discovering, Summarizing and Using Multiple Clusterings held in conjunction with 16th ACMSIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 1-9) ,Washington, DC, USA.

Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine learning*, *2*(2), 139-172.

Gad, W. (2016). SVM-Kmeans: Support Vector Machine based on Kmeans Clustering for Breast Cancer Diagnosis. *International Journal of Computer and Information Technology*, *5*(2), 253-257.

Higgs, R. E., Bemis, K. G., Watson, I. A., & Wikel, J. H. (1997). Experimental designs for selecting molecules from large chemical databases. *Journal of Chemical Information and Computer Sciences*, *37*(5), 861-870.

Kolyshkina, I., Nankani, E., Simoff, S., & Denize, S. (2010). Retail analytics in the context of “Segmentation, Targeting, Optimisation” of the operations of convenience store franchises. In: *Anzmac*.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50-60.

Mendes, A. B., & Cardoso, M. G. (2006). Clustering Supermarkets: The role of experts. *Journal of Retailing and Consumer Services*, *13*(4), 231-247.

Mooi, E., & Sarstedt, M. (2010). *Cluster analysis* (pp. 237-284). Springer Berlin Heidelberg. doi:10.1007/978-3-642-12541-6_9

Pollack, J. (n.d.). Retail Clustering Methods. Retrieved from http://www.parkeravery.com/pov_Retail_Clustering_Methods.html

Porter, M. E. (2007). Clusters and economic policy: Aligning public policy with the new economics of competition. *White Paper (Institute for Strategy and Competitiveness, Harvard Business School, 2007).*

Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 134-148.

Rajagopal, S. (2011,November). Customer data clustering using data mining technique. *International Journal of Database Management Systems*, *3*(4), 1-11.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65.

Schiffman, L., Bednall, D., O’Cass, A., Paladino, A., Ward, S., & Kanuk, L. (2008). Consumer Behaviour, Frenchs Forest, N.S.W.: Pearson Education Australia.

Segal, M. N., & Giacobbe, R. W. (1994). Market segmentation and competitive analysis for supermarket retailing. *International Journal of Retail & Distribution Management*, *22*(1), 38-48.

Shmueli, G., Patel, N. R., & Bruce, P. C. (2007). *Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner* (pp.226-243). Hoboken, NJ: John Wiley & Sons

Vapnik, V. N., & Vapnik, V. (1998). *Statistical learning theory* (Vol. 1). New York: Wiley.

Vassilikopoulou, A. I., Siomkos, G. J., & Mylonakis, J. (2005). Clustering consumers according to their attitudes on corporate social responsibility. *International Journal of Business Governance and Ethics*, *1*(4), 317-328.

Wang, J., Wu, X., & Zhang, C. (2005). Support vector machines based on K-means clustering for real-time business intelligence systems. *International Journal of Business Intelligence and Data Mining*, *1*(1), 54-64.

Williams, K. C., & Page, R. A. (2011). Marketing to the generations. *Journal of Behavioral Studies in Business, 3,* 1.

Yao, Y., Liu, Y., Yu, Y., Xu, H., Lv, W., Li, Z., & Chen, X. (2013). K-SVM: An Effective SVM Algorithm Based on K-means Clustering. *JCP Journal of Computers,* 8(10), 2632-2639. doi:10.4304/jcp.8.10.2632-2639