BRAND LOYALTY ANALYSIS SYSTEM USING K-MEANS ALGORITHM

Ayla Sayli[a*], Isil Ozturk[b], Merve Ustunel[a]

[a]*Yildiz Technical University, Department of Mathematical Engineering, Davutpasa Campus 34010, Istanbul, Turkey,*
sayli@yildiz.edu.tr*(corresponding author), ustunelmerve@gmail.com

[b]*Migros Ticaret A.S. General Headquarter, Data WareHouse & BI, Atatürk Mah. Turgut Ozal Bulvari, No: 7, 34758 Atasehir, Istanbul, Turkey,*
isilo@migros.com.tr

Abstract

The aim of this paper is to implement a brand loyalty analysis system to find out the brand loyalty using data mining techniques. Data are increasing day by day and companies require a need for new techniques and analysis to be able to support their system automatically and intelligently by analyzing large data repositories to obtain useful information. As a specific approach, the study aims to develop a brand loyalty analysis system for the cases of general brand loyalty, item brand loyalty and categorical brand loyalty. We use the data clustering algorithm of K-means for data analysis. Our system is based on the data preparation algorithm and then it constructs the sales tables which contains sale quantity for each product. The case study is done in the stores of Migros Ticaret A.S. Our approach is based on the clustering analysis is used to provide a better knowledge about the role played by each case and emphasizes the role of attributes for the brand loyalty.

Keywords: Data mining, brand loyalty, clustering, K-means, data preparation algorithm.

1. Introduction

At a global level, the customer loyalty is generally positively related to the profitability and long-term growth of a company. The companies need to consistently make better analysis and increase business profitability (see for instance [1], [2], [3] and the references therein). In this time, the customer loyalty makes a major role and has been elaborately studied. General purpose of this study is to determine brand loyalty to achieve analysis of sales which are believed to maximize satisfaction with repetitive sales under marketing concept. Brand loyalty is a strategic objective in market management especially for customer items. As a matter of fact, conquering a new customer becomes more problematic and costly, and it is often the regular customers that assure the growth and the survival of the brand for years to come.

Henceforth, it is important to define and have a better understanding of the brand loyalty on the customers (see [6] and [8] for instance and references therein).

Data mining is an ongoing process and has attracted a great deal of attention in information industry especially in recent years that is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge (see [7] and [9]). At the same time, data mining can fit in with other strategies for understanding market and customers. Market research, customer satisfaction and the other techniques are compatible with data mining and more intensive data analysis. The key is to recognize the focus on customers and the commonality of data across the enterprise (see [4] and [10]). In this study, data mining helps for better understanding of the link between regular customers, items and brands for the brand loyalty. In this study, the loyalty to brands has been measured entirely by purchase behavior and the brand loyalty has been analyzed in three fields: (1) general brand loyalty, (2) item brand loyalty and (3) categorical brand loyalty. With this analysis, all the brand loyalty of the regular customers, which is determined by their choice of what brand or what item-based brand or what they choose to have an item of a brand in every item category (see for instance [15] and the references therein).

Clustering models in data mining focus on identifying groups of similar records and label the records according to the group to which they belong (see for instance [2] and the references therein). Clustering models are often used to create clusters or segments that are then used as inputs in subsequent analyses and its effect is determined by their ability to capture interesting groupings in the data and provide useful descriptions of those groupings (see [5], [13] and [14]). The technological advances in our study is that our system is based on K-means clustering algorithm and the analysis tables are prepared automatically using MS-SQL database. K-means clustering algorithm which is among the most popular techniques, is written on Java in this study. In addition; IBM SPSS Modeler, SAS Data Mining, RapidMiner, ORACLE Data Mining, MATLAB, WEKA etc., are the tools which can be used for clustering applications but our data preparation algorithms do not exist in them. In the Java interface, the K-means algorithm works with numerical data only and on the other hand the algorithm is scalable, so it works on large quantities of data.

The analysis of loyalty programs has been largely researched in literature related to the profit sector. The researchers mainly review the studies related to customer relationship management, customer satisfaction, customer loyalty and customer segmentation using K-means clustering algorithm based on statistics, the surveys and repetitive purchasing behaviors etc. Well-run business naturally form learn relationship with their customers. Over time, they learn more and more about their customers, and they use that knowledge to serve them better. The results are happy, loyal customers and profitable businesses (see for instance [11] and the references therein). Another requirement is ability to follow people around and tailor offers for them in order to reward desirable customers with deals that will keep them coming back (see for instance [12] and the references therein). In brief, the common approach in each research that utilizes a method is introduced that obtains the behavioral traits of customers using the information related to the customers of an organization; it then classifies the customers using the K-means algorithm and finally scores the customers in terms of their loyalty in each cluster (see for instance [1] and the references therein).

Firstly, we prepared the cumulated data for the analysis and the description of data preparation algorithms will be given in Section 3. The clustering analysis to find brand loyalty is done based on sales amounts for each brand as general brand loyalty, sales amounts based on items as item brand loyalty and based on sales amounts with hierarchical category level as categorical brand

loyalty and finally the loyalties are found, then they are compared in Section 4. The computational results on the real data are very promising and can be used to make better strategic sales plans.

2. Case study

The case study is done in the stores owned by the retailing company, Migros Ticaret A.S. which is one of the largest retail companies in Turkey, with 1,230 stores and sells over 340,090 items. Each store has over 10,000 sales transactions per day currently. The total size of the database between 1997 and 2000 was 2 TB, it is now over 12 TB. We have analyzed the data for shopping's of 814 regular customers in 2013. Our brand loyalty analysis system in Figure 1 is implemented in Java based on MS-SQL database where the data came from Teradata machine (Cabs (1), 1N6700H+1HSN, 512GB Memory, Bynet v5.0) of the company.
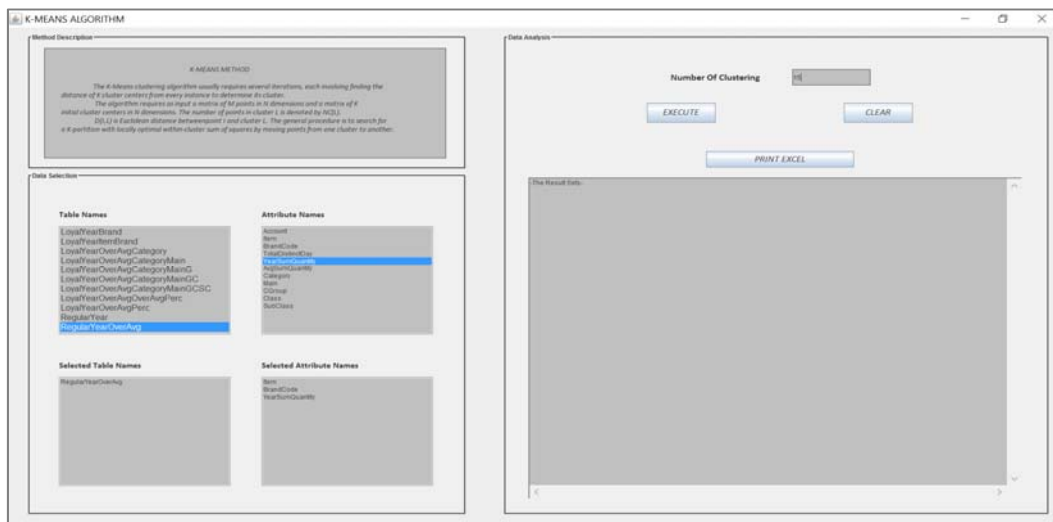

Figure 1. The interface of brand loyalty analysis system

In addition, the analysis system can be applied to small-scale companies as well as large-scale companies. Our system firstly starts to prepare the data using the algorithms given in Section 3.1 and then continues with the application of K-means algorithm described in Section 3.2.

3. Algorithms of data preparation and K-means clustering

3.1. Data preparation algorithms

In this study, the real names of brands are coded as BRAND1, BRAND2, BRAND3 and so on, likewise CUSTOMER1, CUSTOMER2, CUSTOMER3 and so on for customers and ITEM1, ITEM2, ITEM3 and so on for items, under data protection regulations of Migros Ticaret A.S. in Turkey. We have used bills of sales in 2013 for 814 regular customers. The bills contain distinct 1,628 brands and 20,407 items. Analysis will be made on the basis of that brands' sales quantity. The main table for the analysis is named as RegularYear. Its attributes are Account (used for customer number), Item (used for item number), BrandCode (used for brand code), YearDistinctDay (number of different shopping days of the item), YearSumQuantity (used for the sum quantity group by account, item, brand), Category (used for category number), Main (used for category main group number), CGroup (used for category group number), Class (used

for category class number), SubClass (used for category subclass number). The number of records in the table is 236,988. Items are organized in categorical item tree which is shown in Figure 2.
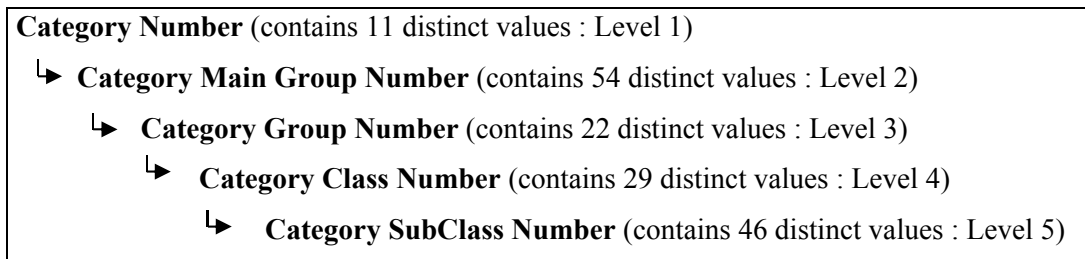
Category Number (contains 11 distinct values : Level 1)
↳ **Category Main Group Number** (contains 54 distinct values : Level 2)
　　↳ **Category Group Number** (contains 22 distinct values : Level 3)
　　　↳ **Category Class Number** (contains 29 distinct values : Level 4)
　　　　↳ **Category SubClass Number** (contains 46 distinct values : Level 5)

Figure 2. Categorical item tree

As an example, the components of categorical item tree can be respectively shown as 'Detergent-Paper-Cosmetics', 'Detergent', 'Laundry Washers', 'Liquid Detergents' and 'For Colors' or 'Delicatessen Marketing', 'Poultry', 'Chicken Products Services', 'Chicken Parts' and 'Chest and Parts' within the analysis.

For example, a prototype of the main table is given in Table 1.

Table 1. Prototype of RegularYear table

Account	Item	BrandCode	Year Distinct Day	Year Sum Quantit	Category	Main	CGroup	Class	SubClass
CUSTOM	ITEM	BRAND1	1	1	1	609	10	10	15
CUSTOM	ITEM	BRAND1	1	1	1	609	10	10	55
CUSTOM	ITEM	BRAND2	1	1	1	125	10	10	30

The data preparations for the analysis are given in Table 2-9 with their algorithms.

RegularYearOverAvg (Account, Item, BrandCode, TotalDistinctDay, YearSumQuantity, AvgSumQuantity, Category, Main, Group, Class, SubClass**) with 105,318 records**

Algorithm 1 :

Step1: First, taking all data from RegularYear main table　Sum (YearSumQuantity) and Avg (YearSumQuantity) values are calculated.

Step2: Data is selected if Sum (YearSumQuantity) >= Avg (YearSumQuantity) inequality to obtain if the customer buys this product over the average sales quantity of this product and then the further research should be done for this　customer-product couple to find whether the customer has a brand loyalty for this product or not.

Step3: Future analysis tables are created from this table.

Table 2. Creation of the RegularYearOverAvg table

Account	Item	Brand Code	Total Distinct Day	Year Sum Quantity	Avg Sum Quantity	Category	Main	CGroup	Class	SubClass
CUSTOMER2	ITEM4	BRAND3	1	1,42	0,88	5	311	15	10	10
CUSTOMER2	ITEM5	BRAND3	1	1,15	0,91	5	311	15	10	10
CUSTOMER2	ITEM6	BRAND4	4	12	1,6	5	311	10	20	15

LoyalYearBrand (BrandCode, BrandYearSumQuantity**) with 1,628 records**

Algorithm 2 :

Step1: Just BrandCode and YearSumQuantity attributes are selected from RegularYear OverAvg table.

Step2: Sum (YearSumQuantity) is calculated and grouped by for the BrandCode attribute.

Step3: The results are inserted to LoyalYearBrand table for the analysis.

Step4: K-means algorithm is applied for LoyalYearBrand table on the BrandYearSum Quantity attribute.

Table 3. Data Preparation for general brand loyalty

BrandCode	BrandYearSumQuantity
BRAND12	6,19
BRAND9	8,24
BRAND7	12,23

LoyalYearItemBrand (Item, BrandCode, ItemBrandYearSumQuantity**) with 20,407 records**

Algorithm 3 :

Step1: BrandCode, Item and YearSumQuantity attributes are selected from Regular YearOverAvg table.

Step2: SUM (YearSumQuantity) is calculated and grouped by for the BrandCode and Item attributes.

Step3: The results are inserted to LoyalYearItemBrand table for the analysis.

Step4: K-means is applied for LoyalYearItemBrand table on the Item BrandYearSum Quantity attribute.

Table 4. Data preparation for item based brand loyalty

Item	BrandCode	ItemBrandYearSumQuantity
ITEM7	BRAND5	38
ITEM8	BRAND6	11
ITEM9	BRAND7	2

LoyalYearOverAvgCategory (Category, BrandCode, CategoryBrandSumQuantity**) with 1,798 records**

Algorithm 4 :

Step1: Category, BrandCode and YearSumQuantity attributes are selected from Regular YearOverAvg table.

Step2: SUM (YearSumQuantity) is calculated and grouped by for the Category and BrandCode attributes.

Step3: The results are ordered by Category and then inserted to LoyalYearOverAvg Category table for the analysis.

Step4: K-means is applied for LoyalYearOverAvgCategory table on the CategoryBrand SumQuantity attribute.

Table 5. Data preparation for categorical brand loyalty: level 1

Category	BrandCode	CategoryBrandSumQuantity
0	BRAND10	3
0	BRAND11	1
0	BRAND12	269

LoyalYearOverAvgCategoryMain (Category, Main, Brand Code, CategoryMainBrand SumQuantity**) with 2,066 records**

Algorithm 5 :

Step1: Category, Main, BrandCode, YearSumQuantity attributes are selected from Regular YearOverAvg table.

Step2: SUM (YearSumQuantity) is calculated and grouped by for the Category, Main and BrandCode attributes.

Step3: The results are ordered by Category and Main, then inserted to LoyalYear OverAvgCategoryMain table for the analysis.

Step4: K-means is applied for LoyalYearOverAvgCategoryMain table on the Category MainBrandSumQuantity attribute.

Table 6. Data preparation for categorical brand loyalty: level 2

Category	Main	BrandCode	CategoryMainBrandSumQuantity
0	999	BRAND10	3
0	999	BRAND13	6
0	999	BRAND14	55

LoyalYearOverAvgCategoryMainG (Category, Main, CGroup, BrandCode, CategoryMain GroupBrandSumQuantity**) with 2,577 records**

Algorithm 6 :

Step1: Category, Main, Group, BrandCode and YearSumQuantity attributes are selected from RegularYearOverAvg table.

Step2: SUM (YearSumQuantity) is calculated and grouped by for the Category, Main, Group and BrandCode attributes.

Step3: The results are ordered by Category, Main and Group, then inserted to LoyalYearOverAvgCategoryMainG table for the analysis.

Step4: K-means is applied for LoyalYearOverAvgCategoryMainG table on the CategoryMainGroupBrandSumQuantity attribute.

Table 7. Data preparation for categorical brand loyalty: level 3

Category	Main	CGroup	BrandCode	CategoryMainGroupBrandSum Quantity
0	999	999	BRAND13	6
0	999	999	BRAND10	3
0	999	999	BRAND14	55

LoyalYearOverAvgCategoryMainGC (Category, Main, CGroup, Class, BrandCode, CategoryMainGroupClassBrand SumQuantity**) with 3,563 records**

Algorithm 7 :

Step1: Category, Main, Group, Class, BrandCode and Year SumQuantity attributes are selected from RegularYearOverAvg table.

Step2: SUM (YearSumQuantity) is calculated and grouped by for the Category, Main, Group, Class and BrandCode attributes.

Step3: The results are ordered by Category, Main, Group and Class, then inserted to LoyalYearOverAvgCategoryMainGC table for the analysis.

Step4: K-means is applied for LoyalYearOverAvgCategoryMainGC table on the Category MainGroupClassBrandSumQuantity attribute.

Table 8. Data preparation for categorical brand loyalty: level 4

Categor y	Mai n	CGrou p	Clas s	BrandCod e	CategoryMainGroupClassBrandSumQuant ity
0	999	999	999	BRAND13	6
0	999	999	999	BRAND14	55
0	999	999	999	BRAND15	36

LoyalYearOverAvgCategoryMainGCSC (Category, Main, CGroup, Class, SubClass, BrandCode, CategoryMainGroup ClassSubClassBrandSumQuantity**) with 5,928 records**

Algorithm 8 :

Step1: Category, Main, Group, Class, SubClass, BrandCode and YearSumQuantity attributes are selected from RegularYearOverAvg table.

Step2: SUM (YearSumQuantity) is calculated and grouped by for the Category, Main, Group, Class, SubClass and BrandCode attributes.

Step3: The results are ordered by Category, Main, Group, Class and SubClass, then inserted to LoyalYearOverAvgCategoryMainGCSC table for the analysis.

Step4: K-means is applied for LoyalYearOverAvgCategoryMainGCSC table on the CategoryMainGroupClassSubClassBrandSumQuantity attribute.

Table 9. Data preparation for categorical brand loyalty: level 5

Catego ry	Mai n	CGro up	Cla ss	SubCl ass	BrandCode	CategoryMainGroupClassSubClass BrandSumQuantity
0	999	999	999	999	BRAND13	6
0	999	999	999	999	BRAND14	55
0	999	999	999	999	BRAND 16	1

3.2 Data preparation algorithms

K-means clustering algorithm by (see [4] and [10]) usually requires several iterations and initial points are generally randomly chosen. The data set is divided into k clusters. The middle point of each group is the center point of the cluster which it belongs to. Steps of the algorithm is given below:

- **Calculate the Euclidean distance from each object to each cluster.**

All instances are assigned to their closest cluster centers according to the ordinary Euclidean distance metric calculated by equation (1) and *N* is number of attributes of data set.

$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \cdots + |x_{iN} - x_{jN}|^2} \qquad (1)$$

- **Calculate the sum of squared error of clustering.**

The assignment mechanism of K-means algorithm allows all data to be only in one cluster. The Sum of Squared Error(SSE) is widely used to determine the number of clusters and evaluation. The clustering with the lowest value of SSE gives the best result. The sum of the square from objects to central points of clusters they belong to is calculated by equation (2) where d is the Euclidean distance; M is the mean vector, X is any instance of each cluster. The algorithm will try to determine the part k to reduce SSE.

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} d^2(x - M_i) \qquad (2)$$

- **Determine a new center point for each cluster.**

The mean vector is recalculated each time members of a cluster is changed.

- **Repeat step 2, 3 and 4 until the members of clusters are not changed.**

Iterations continue until the same points are assigned to each cluster in consecutive rounds, at which stage the cluster centers have stabilized and will remain the same.

4. Computational results

As the approach of the brand loyalty analysis system; the clustering algorithm is applied for 814 regular customer for 2013 which is mentioned before. For each entry in the user interface analysis, BRAND database and K-means analysis are selected. During clustering, K-means algorithm has been applied in a variety of purposes and the results are calculated by the distance on multiple attributes. The results are obtained by analyzing general brand loyalty, item brand loyalty and categorical brand loyalty. In order to give the clusters in detail, the first two clusters are chosen for examples and the related figures are obtained. Then they are analyzed and interpreted. For the analyses in this section, the number of clusters is selected according to our preferences considering the presentation of the results in the paper and the data in each cluster. However this preference can easily be changed in the interface by any user and the algorithm can be executed.

4.1 General brand loyalty

Considering the total sales of brands, K-means clustering algorithm (k=25) is applied to the LoyalYearBrand table shown in Table 3 and clustering results are obtained calculating the distance on BrandYearSumQuantity attribute. Referring to Figure 3 and Figure 4, the highest brand loyalties are shown in Cluster 0 and Cluster 1.
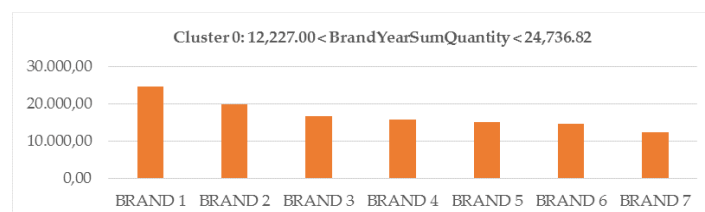

Figure 3. General brand loyalty / Cluster 0

Figure 4. General brand loyalty / Cluster 1

Taking the first place in the general brand loyalty among the brands, it can be said that these brands have the high brand values. The number of brands in each cluster according to their distances, or in other words, the distributions of the number of brands, are shown in Table 10.

Table 10. Clusters of general brand loyalty

ClusterNo	Sum Quantity Intervals	Number Of Brands
Cluster 0	12,227.0 – 24,736.82	7
Cluster 1	5,311.00 - 8,714.81	7
Cluster 2	3,199.00 – 4,689.00	13
Cluster 3	2,355.02 – 2,829.00	16
Cluster 4	1,532.00 – 2,094.00	23
Cluster 5	1,132.00 – 1,481.00	22
Cluster 6	852,00 – 1,114.75	26
Cluster 7	658,00 - 840,00	36
Cluster 8	458,00 - 635,00	48
Cluster 9	324,00 - 445,00	59
Cluster 10	243,00 - 317,00	45
Cluster 11	184,00 - 235,00	34
Cluster 12	136,00 - 181,00	49
Cluster 13	99,00 - 134,00	68
Cluster 14	71,00 - 97,33	67
Cluster 15	50,82 - 70,11	86
Cluster 16	34,00 - 50,00	105
Cluster 17	22,00 - 33,19	108
Cluster 18	14,00 - 21,00	140
Cluster 19	8,00 - 13,18	146
Cluster 20	4,00 - 7,00	200
Cluster 21	2,00 - 3,63	179
Cluster 22	1,39 - 1,85	6
Cluster 23	0,68 - 1,01	124
Cluster 24	0 - 0,60	14

Looking at the distribution of the brands for all clusters, 16% is observed in the first 10 clusters. From Table 10, the number of brands in these clusters decreases oppositely with the sum quantity intervals. The clusters can be very useful for companies to understand the worthiness of their brand and the customer's choices ; however, this analysis may not be enough to see which item of a brand has higher loyalty. Therefore, the next analysis is done for this purpose.

4.2 Item brand loyalty

The status of the brand selection is examined with item-based in this analysis. On the LoyalYearItemBrand table in Table 4, K-means clustering algorithm (k=15) is applied and the distance is calculated for ItemBrandYearSumQuantity attribute. In Figure 5 and 6, similarly the first two clusters are chosen for detail.
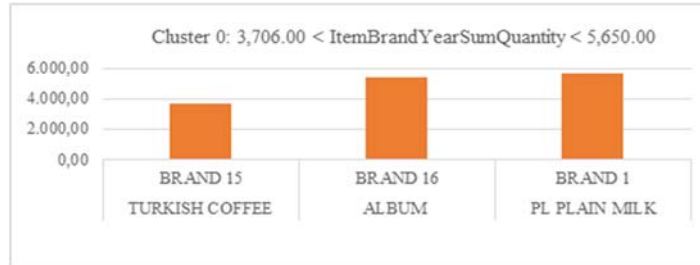

Figure 5. Item brand loyalty / Cluster 0

The continuing brand ranking is shown in Figure 6. The BRAND3 branded milk is shown as two distinct bars because the table contains two different codes for item in this brand.
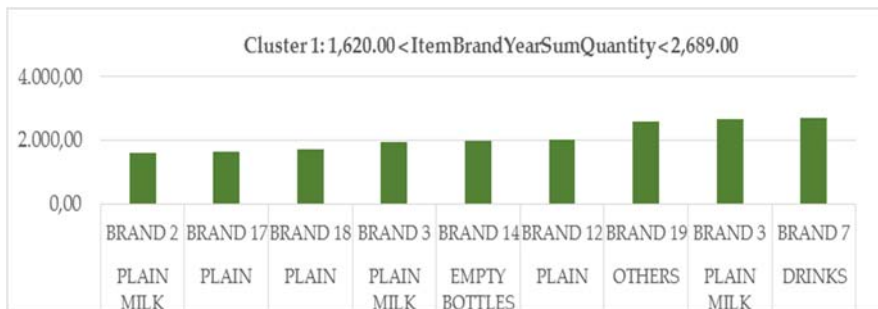

Figure 6. Item brand loyalty / Cluster 1

The item brand loyalty for all cluster distributions are shown in Table 11 and the highest item based brand loyalties were found for the brands of egg, milk, fruit juice, cigarette and bread.

Table 11. Clusters of item brand loyalty

Cluster No	Sum Quantity Intervals	Number of Distinct Brands	Number of Items
Cluster 0	3,706.00 –	3	3
Cluster 1	1,620.00 –	8	9
Cluster 2	932,00 – 1,551.00	16	24
Cluster 3	578,00 - 892,00	27	42
Cluster 4	376,00 - 557,00	40	74
Cluster 5	261,00 - 373,00	66	113
Cluster 6	181,00 - 259,00	95	194
Cluster 7	126,00 - 180,00	128	290
Cluster 8	84,00 - 125,00	183	482
Cluster 9	52,00 - 83,00	247	732
Cluster 10	30,00 - 51,81	363	1159
Cluster 11	15,00 - 29,49	526	1953
Cluster 12	5,97 - 14,60	797	3713
Cluster 13	1,94 - 5,78	1162	7760
Cluster 14	0,00 - 1,90	888	3859

Looking at clustering results of the item brand loyalty, 50% of the brands and 10% of the items took place in the first 10 clusters. On the other hand, some of these items in several brands of jam, black olive, live plants, hair dyes and skin care creams etc. in the last four clusters, were not preferred.

4.3 Categorical brand loyalty

In this section, the brand loyalty is analyzed categorically according to each level of category item tree which was given in Table 5-9. The results are shown below.

4.3.1. Categorical brand loyalty (Level 1: Category number)

In order to implement the K-means algorithm (k=10), the distance is calculated to CategoryBrandSumQuantity attribute on the LoyalYearOverAvgCategory given in Table 5. The results of the first two clusters in Figure 7 and Figure 8 are similarly chosen to give the detail about the highest loyalties.

Figure 7. Categorical brand loyalty (level 1: category number) / Cluster 0
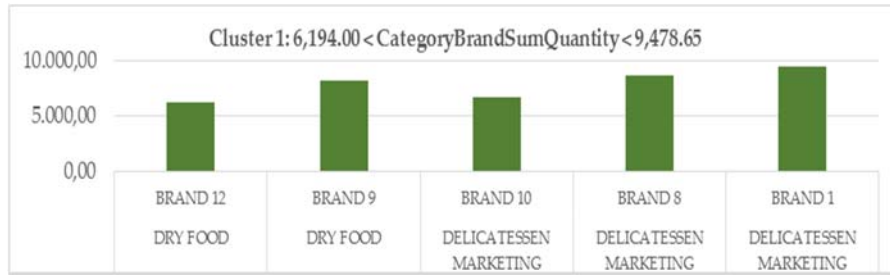
Figure 8. Categorical brand loyalty (level 1: category number) / Cluster 1

Clustering results are given in Table 12. In the categorical status has continued in the Tobacco and Alcoholic Beverages and Detergent-Paper-Cosmetics categories.

Table 12. Clusters of categorical brand loyalty (level 1: category number)

ClusterNo	Sum Quantity Intervals	Number Of Brands
Cluster 0	12,059.00 – 17,412.33	7
Cluster 1	6,194.00 – 9,478.65	5
Cluster 2	3,547.00 – 5,670.00	12
Cluster 3	2,201.40 – 3,447.00	19
Cluster 4	1,378.00 – 2,089.00	30
Cluster 5	852,00 – 1,347.00	41
Cluster 6	495,00 - 840,00	79
Cluster 7	242,00 - 490,00	129
Cluster 8	74,91 - 235,00	216
Cluster 9	0,00 - 74,00	1189

4.3.2. Categorical brand loyalty (level 2: category main group number)

From the categorical item tree, the LoyalYearOverAvgCategoryMain table is created in Table 6 and K-means algorithm (k=10) is applied. The distance is calculated on CategoryMainBrandSumQuantity attribute and two clusters' details for examples are shown in Figure 9 and Figure 10.

Figure 9. Categorical brand loyalty (level 2: category main group number) / Cluster 0
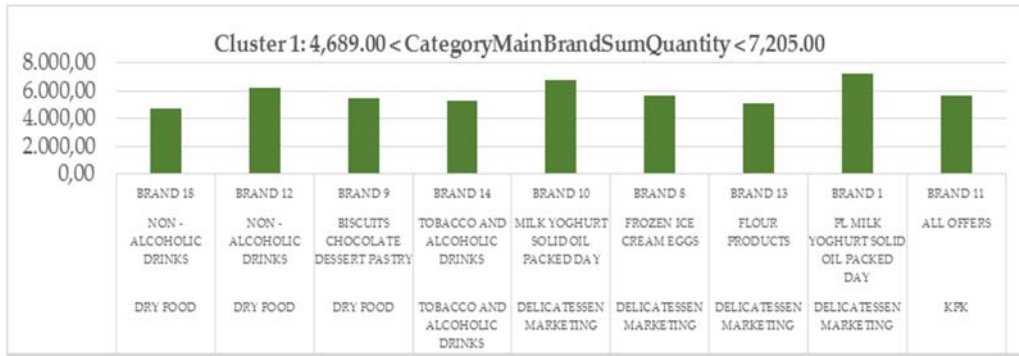
Figure 10. Categorical brand loyalty (level 2: category main group number) / Cluster 1

The distribution of the clusters is given in Table 13.

Table 13. Clusters of categorical brand loyalty (level 2: category main group number)

ClusterNo	Sum Quantity Intervals	Number Of Brands
Cluster 0	11,879.00 – 16,411.00	6
Cluster 1	4,689.00 – 7,205.00	9
Cluster 2	2,954.814 – 4,299.00	13
Cluster 3	1,930.644 – 2,829.00	24
Cluster 4	1,278.994 – 1,862.00	29
Cluster 5	820,00 – 1,231.00	39
Cluster 6	488,00 - 807,00	73
Cluster 7	242,00 - 484,00	143
Cluster 8	76,00 - 240,24	239
Cluster 9	0,00 - 75,00	1259

With the distribution, it was observed that multiple category main groups of brands taken place. The last main group in terms of sales categorical clustering has many brands that range between 0.00 and 75.00. This case showed that there are some brands without the high brand value for the categorical main group.

4.3.3. Categorical brand loyalty (level 3: category group number)

On LoyalYearOverAvgCategoryMainG table in Table 7, K-means algorithm (k=15) is applied for the creation of more accurate clustering results and the distance is calculated on CategoryMainGroupBrandSumQuantity attribute. The details of two clusters for the categorical brand loyalties are shown in Figure 11 and Figure 12.

Figure 11. Categorical brand loyalty (level 3: category group number) / Cluster 0
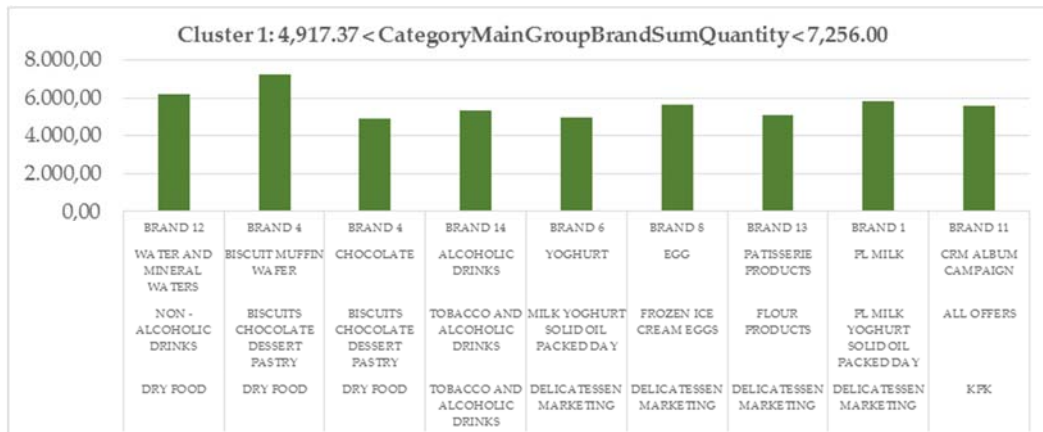
Figure 12. Categorical brand loyalty (level 3: category group number) / Cluster 1

The brand loyalty distribution for all clusters is shown in Table 14.

Table 14. Clusters of categorical brand loyalty (level 3: category group number)

ClusterNo	Sum Quantity Intervals	Number Of Brands
Cluster 0	10,266.00 – 13,809.00	4
Cluster 1	4,917.37 – 7,256.00	9
Cluster 2	2,965.00 – 4,244.00	14
Cluster 3	1,967.00 – 2,806.00	25
Cluster 4	1,369.00 – 1,919.00	23
Cluster 5	952,25 – 1,331.00	34
Cluster 6	662,00 - 943,00	41
Cluster 7	455,00 - 636,00	60
Cluster 8	315,00 - 446, 52	80
Cluster 9	216,27 - 309,44	83
Cluster 10	141,00 - 212,00	93
Cluster 11	87,00 - 139,00	124
Cluster 12	48,00 - 86,00	191
Cluster 13	18,00 - 47,00	359
Cluster 14	0,00 - 17,78	978

4.3.4. Categorical brand loyalty (level 4: category class number)

The distance is calculated for CategoryMainGroupClassBrandSumQuantity attribute on LoyalYearOverAvgCategoryMainGC in Table 8 and K-means algorithm (k=15) is applied. The first two cluster results are shown for detail, in Figure 13 and Figure 14. From the results, it can be said that the items could change in clusters but the brands remained the same.

121

Figure 13. Categorical brand loyalty (level 4: category class number) / Cluster 0

Figure 14. Categorical brand loyalty (level 4: category class number) / Cluster 1

The distribution of the clusters is shown in Table 15.

Table 15. Clusters of categorical brand loyalty (level 4: category class number)

ClusterNo	Sum Quantity	Number Of
Cluster 0	8,137.00 – 12,227.00	4
Cluster 1	5,035.00 – 6,194.00	6
Cluster 2	2,807.00 – 4,012.00	13
Cluster 3	1,890.00 – 2,725.00	20
Cluster 4	1,322.00 – 1,794.00	22
Cluster 5	923,00 – 1,299.00	40
Cluster 6	656,00 - 899,00	42
Cluster 7	452,00 - 639,00	62
Cluster 8	317,00 - 445,00	74
Cluster 9	221,00 - 315,00	97
Cluster 10	148,00 - 219,00	104
Cluster 11	93,00 - 146,00	142
Cluster 12	50,00 - 92,00	223
Cluster 13	18,00 - 49,41	415
Cluster 14	0,00 - 17,78	1099

In the analysis performed by the category main groups in the categorical level, BRAND3 is ranked as the first brand and BRAND2 is ranked the second brand in the milk group. By drilling down into classes in the milk group, BRAND2's Durable Milk lower class category and

BRAND3's Pasteurized Milk is more preferred. In general, the competition is continued in the milk class between BRAND2 and BRAND3 brands.

4.3.5. Categorical brand loyalty (level 5: category subclass number)

Finally in the categorical item tree, K-means clustering (k=25) analysis is performed on LoyalYearOverAvgCategoryMainGCSC table in Table 9 for the last level of the category item tree. The distance is calculated for ClassSubClassBrandSumQuantity attribute and the first two clusters are shown in Figure 15 and Figure 16.

Figure 15. Categorical brand loyalty (level 5: category subclass number) / Cluster 0

Figure 16. Categorical brand loyalty (level 5: category subclass number) / Cluster 1

Referring to Table 16 for the distribution of the clusters, the cigarette class of Tobacco Products group has been observed that the American Harman subclass has the brand loyalty.

Table 16. Clusters of categorical brand loyalty (level 5: category subclass number)

ClusterNo	Sum Quantity Interval	Number Of Brands
Cluster 0	7,853.00 – 12,227.00	2
Cluster 1	5,035.00 – 6,194.00	6
Cluster 2	3,289.00 – 4,381.00	7
Cluster 3	1,948.00 – 2,883.00	15
Cluster 4	1,331.00 – 1,883.00	19
Cluster 5	952,00 – 1,299.00	31
Cluster 6	717,00 - 929,00	24
Cluster 7	541,00 - 701,00	39
Cluster 8	420,00 - 533,00	45
Cluster 9	320,00 - 416,00	70
Cluster 10	240,00 - 317,00	70
Cluster 11	181,73 - 239,00	79
Cluster 12	136,00 - 181,00	100
Cluster 13	100,00 - 135,00	120
Cluster 14	70,82 - 99,00	160
Cluster 15	49,00 - 70,00	199
Cluster 16	33,00 - 48,00	243
Cluster 17	21,00 - 32,74	286
Cluster 18	11,97 - 20,81	387
Cluster 19	5,97 - 11,74	477
Cluster 20	2,96 - 5,78	498
Cluster 21	1,51 - 2,91	425
Cluster 22	0,80 - 1,49	410
Cluster 23	0,40 - 0,78	28
Cluster 24	0,00 - 0,38	26

In categorical analyzes, Dry Food and Delicatessen Marketing categories of products are found to be in the first place for each level of the tree. BRAND5 brands of the Dry Food category, BRAND2 brands of the Delicatessen Marketing category passed the BRAND1 brand.

5. Conclusion

The purpose of this study is to investigate the brand loyalty by looking at the outcomes of purchasing behavior of regular customers. In particular, three distinct analysis were performed as general brand loyalty, item brand loyalty and categorical brand loyalty. While analyzing the data, it was found that general brand loyalty symbolize worth of brands and customer choices regardless of items or the other effects in the market. Based on the above computational results, it can be said that BRAND1 is well ahead from the other brands. When we examined the result based on items, we also saw that the first cluster contains the distribution of various brands and it shows us that the item selection makes an important role in the satisfaction of brands. It was found that, the most loyal item groups contained plain milk, Turkish coffee, drinks, eggs, bread and chocolate and so on; however, last clusters for loyal item groups contained jam, black olive, skin care creams and so on. On the other hand, by drilling down into the sub-levels in categorical analysis, it was found that first brand name always changes for each level. Especially, while BRAND1 the first brand in the general and item brand analysis, it was not seen this brand take a part as the first brand in the categorical analysis. Also, we saw that tobacco and alcoholic drinks category is located in first two clusters for the second and third level analysis of categorical item tree. Therefore the researching for the brand loyalty should be done with all aspects including item's categories because its behavior could change when customer-item-brand were analyzed deeply.

For the use of results, they can be shared with the companies with brand-item clusters in knowledge about their brands-items where they are strong or weak about the future planning of brand-item-customer purposes. For example, it can be made to customers in the loyalty on a specific brand by informing public messages to promote items of this brand, new opportunities or campaigns can be planned. In this regard, we believe that the analysis conducted in this study contributes to the brand loyalty in corporate sector. Concretely, in order to determine the brand loyalty in the future, big data tools can be studied by utilizing a larger scale data.

References

[1] R. Chinomona, and M. Sandada, Customer Satisfaction, "Trust and Loyalty as Predictors of Customer Intention to Re-Purchase South African Retailing Industry, Mediterranean." Journal of Social Sciences, MCSER Publishing, Rome-Italy 14 (2013).

[2] Shital H. Bhojani, Dr. Nirav Bhatt, "Data Mining Techniques and Trends – A Review. " Global Journal for Research Analysis (2016) 252-254.

[3] R. Chinomona and D. Dubihlela, "Does Customer Satisfaction Lead to Customer Trust, Loyalty and Repurchase Intention of Local Store Brands, Mediterranean Journal of Social Sciences, MCSER Publishing, Rome-Italy, 9 (2014).

[4] J. Han and M. Kamber, "Data Mining Concepts and Techniques, (2006).

[5] A.K. Jain, "Data clustering: 50 years beyond K-means, Pattern Recognition Letters 31 (2010) 651–666.

[6] S.M.S.H. Hosseini, A. Maleki and M.R. Gholamian, "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, Expert Systems with Applications 37 (2010) 5259-5264.

[7] K. Mercier, B. Richards and R. Shockley, Analytics: "The Real-World Use of Big Data in Retail" IBM Global Business Services, Business Analytics and Optimization, Executive Report, (2012).

[8] Sofie De Cnudde and David Martens. "A Data Mining Analysis of a Public Service Loyalty Program" Decision Support Systems 73 (2015): 74–84

[9] C.L. Philip Chen and Chun-Yang Zhang, "Data-intensive applications, challenges, techniques and technologies. " A survey on Big Data, Information Sciences 275 (2014) 314–347.

[10] Ian H. Witten and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques" Second Edition, (2005).

[11] Gordon S. Linoff and Michael J. A. Berry, "Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management", Third Edition, (2011).

[12] J. Turow, L. McGuigan and E. R. Maris, "Making Data Mining a Natural Part of Life: Physical Retailing, Customer Surveillance and the 21st Century Social Imaginary" European Journal of Cultural Studie 18 (2015) 464-478.

[13] Mohammed J. Zaki, Wagner Meira, "Data Mining and Analysis: Fundamental Concepts and Algorithms", (2014).

[14] Manish Verma, Mauly Srivastava, Neha Chack, Atul Kumar Diswar, Nidhi Gupta, A Comparative, "Study of Various Clustering Algorithms in Data Mining" International Journal of Engineering Research and Applications (IJERA) 2 (2013): 1379-1384

[15] Shuihua Han, Yongjie Ye, Xin Fu and Zhilong Chen, "Category role aided market segmentation approach to convenience store chain category management" Decision Support Systems 57 (2014): 296–308.