CLUSTERING AS A TOOL FOR OPTIMIZING THE SALE OF CONSUMER GOODS

Radek Kratochvíl[1], Mária Jánešová[2]

[1]*Czech Technical University in Prague, Faculty of Transportation Sciences, Czech Republic*
[2]*Czech Technical University in Prague, Faculty of Transportation Sciences, Czech Republic*

Abstract

Nowadays, most sales of consumer goods are realized by e-shops. For these shops, of course, it is impossible – in terms finance and/or space - to have large quantities of these goods in stock. It is therefore necessary to reduce the offered goods in a suitable and sophisticated way. In this work, based on the knowledge of historical data and using the Cluster Analysis, the sales offer of these goods is optimized. Here the supply of goods is reduced depending on the price and the sale time using the real data. With this reduction, only a certain type and quantity of goods can be stocked within a certain time period to be offered at that time period. The entire supply of goods is selected by individual clusters that determine what goods, at what prices and in what time period are at a certain moment beneficial to sell. This therefore reduces the total sales offer and hence the need to own or fund unnecessarily large warehouses. The cluster analysis is applied to the real data of laptop sales in a medium-sized e-shop in the Czech Republic for the period 2016.

Keywords: clustering, consumer goods, optimization, warehousing

---***---

1. INTRODUCTION

In this work, real data collected from the sale of consumer goods for 2016 are analysed using the cluster analysis. The test is performed on medium-moving consumer goods - laptops. Using this method, sets in a way of similar data are created, which then serve as a basis for the required elimination of the entire data file. The objective is to identify individual clusters that can be considered a certain optimum sales. Our optimum sets will be sets with maximum sales depending on individual months and sales prices. These will be three-dimensional sets with the input values, see Table 1 below.

Table -1: Real data

Goods - price (axis X)	Month of sale (axis y)	Number of pieces sold (axis z)
Laptop - 6.000 CZK	1-12	see Fig.1 (0-70)
Laptop - 8.000 CZK	1-12	see Fig.1 (0-70)
Laptop - 10.000 CZK	1-12	see Fig.1 (0-70)
Laptop - 12.000 CZK	1-12	see Fig.1 (0-70)

2. SALES OPTIMIZATION

2.1 General Requirements and Optimization Process

Before the analysis itself it is necessary to define the individual steps and, of course, to set the necessary goals. In the co-operation with the consumer goods e-shop the necessary real data needed to the analysis were obtained. A set of data variables was created from these real data. The Fig. 1 clearly indicates the plain view of the entire data set in space.

The data content was subjected to a clustered analysis that on a basis of selected algorithm that works with Gaussian normal probability distribution creates individual parameterized sets. In the used algorithm, it is possible to determine in advance the number of required clusters. When calculating, it is very important to determine the initial initialization of the calculation, in our case it will be determined by our own estimation. Subsequently, individual clusters are selected based on sales and optimization requirements. As a result, there is a certain reduction in the data content and consequently the required savings of storage costs.
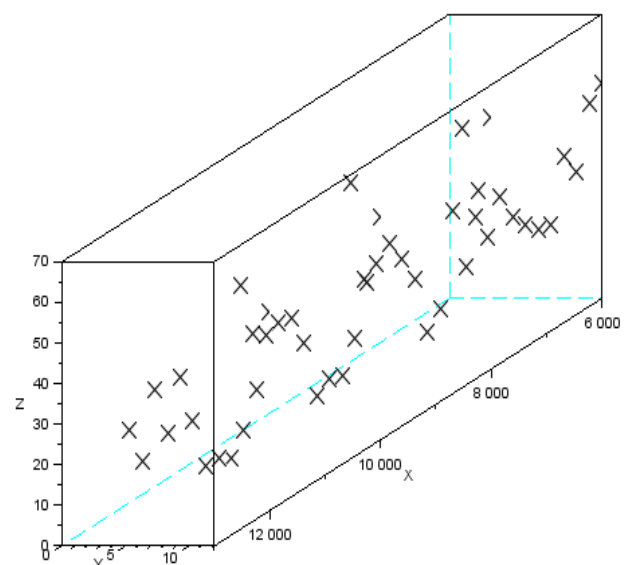

Fig -1: The plain view of the entire data set in space (Z-number of sold units, Y-month, X-price of the laptop)

Individual points of analysis:

- Real data from the sale of goods
- Cluster analysis
- Selection of individual volumes based on specified requirements
- Reduction of offered goods / sales times - saving of storage space

2.2 Cluster Analysis

After creating a data file including complete data collected from laptop sales, this dataset was subjected to Cluster Analysis. A clustering algorithm working with Gaussian probability distribution and with a predetermined number of clusters was used. For this analysis, the number of clusters 8 was selected. In Fig. 2 we can clearly see the division of the whole dataset into individual clusters. The graph also shows gradual shifts of the centres of individual cluster during individual iteration cycles of the clustering process.

Sales of laptops, determining 6 clusters:

Cluster 1	center [6705,5,24]	frequency 0.1861994
Cluster 2	center [6410,8,32]	frequency 0.1205446
Cluster 3	center [8730,4,29]	frequency 0.1428864
Cluster 4	center [8010,9,36]	frequency 0.0679961
Cluster 5	center [7850,11,50]	frequency 0.0508226
Cluster 6	center [11050,6,23]	frequency 0.3239716
Cluster 7	center [10130,11,42]	frequency 0.0559873
Cluster 8	center [11850,11,36]	frequency 0.0515921

From the analysis, we can consequently determine what are the sales in the individual clusters, respectively sales dependency on the price of the goods and the time of sale. After reduction and subsequent calculation, it is possible to find out that after removing 2-3 clusters out of a total number of 8, we only reduce total sales by only units to tens of percent. Thus, we eliminate a substantial supply of goods to the detriment of a relatively small percentage of sales.
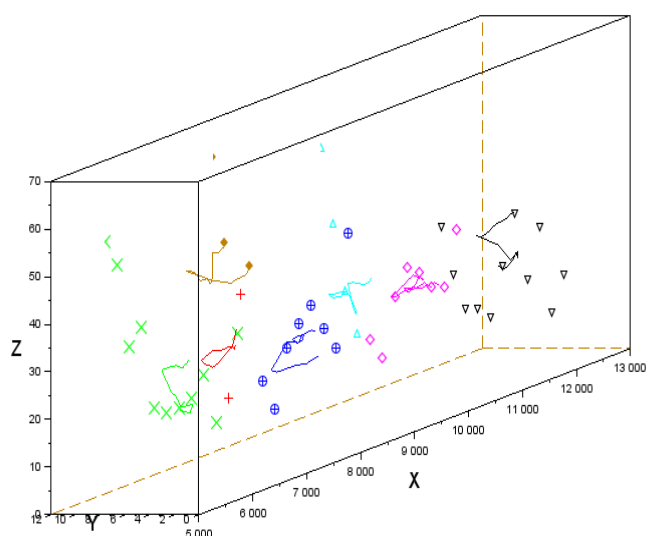

Fig -2: Data file after cluster analysis (Z-number of sold units, Y-month, X-price of the laptop)

3. CONCLUSION

Generally speaking, all kinds of notebooks are sold in large quantities in the end of the year, especially those in more expensive categories. This is mainly caused by the end of the fiscal year for most companies in the Czech Republic, and also for Christmas. Cheaper notebooks are better sold in the beginning of the year, probably due to discounts in shops and further in August and September. This is probably due to the fact that these cheaper laptops are primarily intended for students, who start their school at those time periods.

Through cluster analysis, we have verified that the goods can be appropriately split into several groups according to predefined parameters, and then these sets can be appropriately selected. The performed cluster analysis and the subsequent expert reduction indicates that that it is possible in this way to identify groups of goods that are suitable to be sold only in certain months of the year, and this optimization will not have a significant impact on sales revenue. There will be considerable financial savings due to the minimization of the need to store the goods in the warehouses. Goods will not be stored in stores unnecessarily at a time when their sales are not sufficient. At the same time, it will be ensured that a given range of goods will be symmetrically divided according to current demand.

This article only provides very short conclusions from the extensive research on the problem of reducing storage premises and the related financial costs.

REFERENCES

[1]. NAGY I. *Stochastické systémy - Ivan Nagy* [online]. Ivan Nagy [viewed. 2016-06-30]. Available from: http://nagy.rudolfpohl.cz.
[2]. PECHERKOVÁ P., NAGY I. *Vybrané funkce v programu Scilab z oblasti pravděpodobnost a statistiky*. Praha: 2015, learning material.
[3]. KRATOCHVIL R. *Disertační práce - Ekonomický model komunikačního webového portálu*. Praha: 2016. CVUT FD.

BIOGRAPHIES

Radek Kratochvíl, Ph.D. is a researcher at the Czech Institute of Academic Education and a member of other scientific organizations.

Ph Dr. Mária Jánešová is associate professor at the Czech Technical University in Prague.