Customer behavior analysis in e-commerce for the prediction of customer loyalty

This study deals with customer experience and its impact on customer loyalty to online store. Based on already existing surveys we have tried to develop a new approach for the problem of customer loyalty in Russian emerging online market. This research looks into the matter of customer purchase history provided by Russia's largest online store of tools and equipment. In the project, we carry out a review of consumer behavior. In accordance with the conducted sample, the fact that customer loyalty can be determined by the total or average order value was denied.

Key words: customer experience, customer loyalty, customer behavior analysis

1. Introduction

Customer loyalty plays a decisive role for the majority of online retailers. From the point of view of the customer the main factor determining its loyalty remains his personal experience with the company (Smith and Wheeler, 2002).

Client experience (Schmitt, 2003) is the sum of the client's experience of contact with the company that provides it with goods and services, throughout their relationship (Zomerdijk and Voss, 2010). The main incentive for the purchase in a particular online store, as shown in the works of Carbone and Berry (2007) is not a low price or delivery terms, but a positive experience (not necessarily private) in the past. The main factors for the purchase are such subjective factors as: the credibility of the store, easy to navigate, easy purchase process and delivery terms and prices (Albescu and Pugna, 2014).

Among the latest trends in the management of client experience the widespread desire of companies to personalize this experience can be distinguished (Chung and Wedel, 2014). Ball, Coelho and Vilares (2006) presented an empirical analysis of the relationship between customer loyalty and personalization. The authors formulated the idea that personalization services could partially replace the communication effect.

On the emerging Russian market of e-commerce there are a number of factors that complicate the process of finding new customers and holding old ones, for example, economic pressures, proactive competitors. In this situation, estimation of parameters, which gives the maximum response from the target audience of the company (as a main or indirect and potential) - a necessary and important task (Spiess at al., 2014). Its solution is provide a better understanding what the consumer wants, and, in turn, can be carried out the most efficient way by data analysis. The basic data for the analysis, which are available in online stores, is the purchase history of each customer.

2. Purchase history and its representation in the CRM system

The CRM-system in e-commerce is now basically stores information about products purchased by certain customers, product categories by classifier, the time of purchase, the amount and composition of the order.

According to statistics collected on purchases CRM-analytics built monthly and quarterly reports, evaluate the effectiveness of individual marketing campaigns, consider the amount of attracted and retained customers. Typically, such an analysis is performed averaging over all the company's customers and not affected each of them individually. This analysis is quite useful in terms of information for management decisions, however, we can extract more from the available data (Phillips-Wren and Hoskisson, 2015).

In addition, RFM-analysis (Recency, Frequency, and Monetary) became widespread. Proposed by Fader, Hardy and Lee (2005) RFM-analysis is used for the prediction of customer behavior based on his past actions. By construction, RFM-analysis suggests that significant factors to determine customer value (Customer Life Time Value) is the frequency of its purchases, the last date of purchase, as well as the value of the money spent, while disregarding all other factors. One of the objectives of our work is to assess how much is really frequency, recency and the total value of orders effectively determine the client's behavior and reflect the relationship between the customer and shop.

Consider the purchase history on the example of a small sample. The sample contains information about the history of purchases 1673 customers (907 B2C-client and 766 B2B-client) for the period from 4 January 2013 to 31 December 2014 (726 days). The database contains 50,326 entries.

Each database entry contains customer's identity and his belonging to the sector of the B2C or B2B, the date of registration of the order and its number, the full inventory of the order with an indication of their value and reference to the categories in a hierarchical classifier. Goods, which are not included in the final order or have not been paid, matches entries without the amount in the last column.

The presence of unfinished orders can talk about technical difficulties in placing an order, and the negative experience of the client, which may affect its loyalty to online store. Therefore, a separate task is to check how the share of successfully completed orders connected with customer loyalty and recurrence.

The main methods used in the literature to build predictive models in marketing are linear discriminant analysis, regression analysis, Bayesian classification methods (Jasrai, 2014). Besides cluster analysis is used for segmentation (Solomon, 2004). Problems of analysis of the factors affecting customer loyalty, may be considered in addition to the prediction of the value of customers (CLV, Customer Lifetime Value) using RFM-analysis and used for targeted marketing offers, special email-newsletters etc.

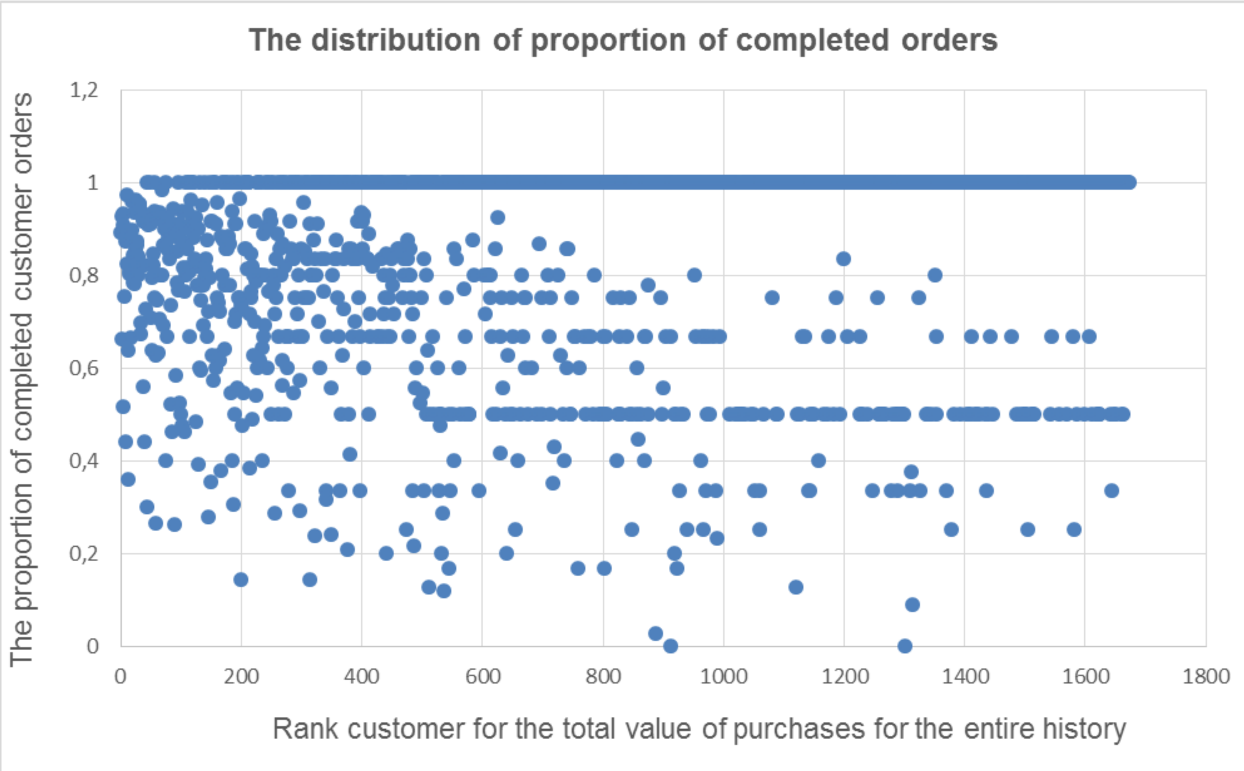

Figure 1. The distribution of share of successfully completed orders

As an indicator that characterizes loyalty, we consider the frequency of shopping, defining it through a number of orders from the first order for the current date (today's date for us is the cutoff date in database sample). We call loyal customers, which have made more than one purchase, and makes purchases more frequently than once every six months. Those who make purchases less often than once at six months, we call occasional customers; we will assume the loyalty is undefined for the other customers. Separation of clients on regular customers and occasional customers is shown in Fig. 2.

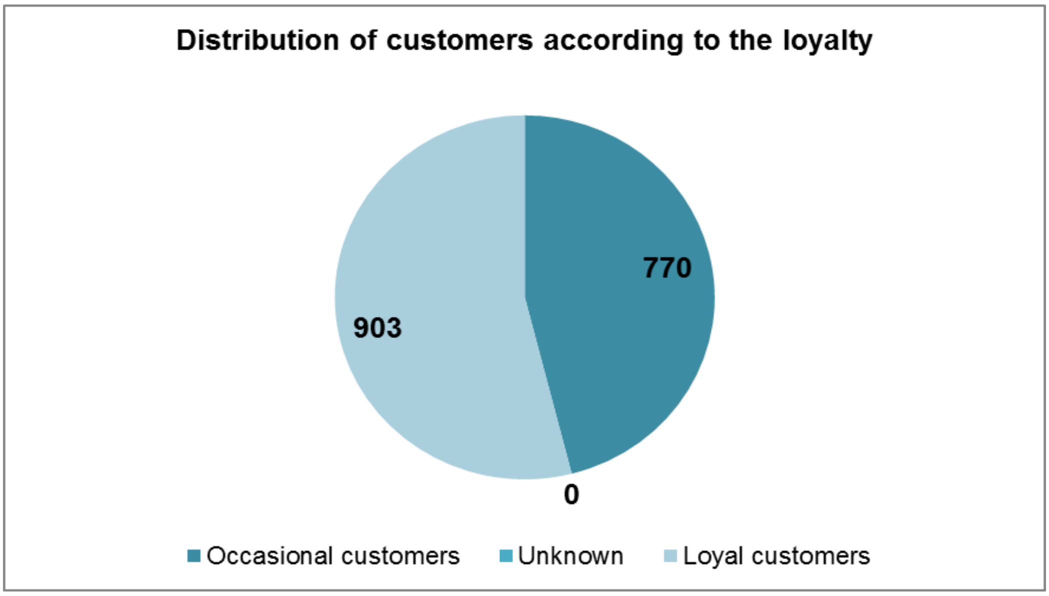

Figure 2. Regular and occasional customers

Also consider these features: the total number of orders for the period, the periods of the most frequency shopping (seasonal), categories of products that are most often purchased by classifier, the breadth of the purchased goods (number of groups presented in the categories customer orders), total and average value of orders, average order quantity (number of items in the order), the average value of the goods in the basket. Thus, instead of the original more than 50 thousand entries, we consider a table, where each customer is represented by one line, characterized by its purchasing behavior. In this table we add additional information - date of first and last order, the average time between purchases etc.

3. Analysis of consumer behavior

Now that customer data are presented in a form that allows marketers to analyze, we will build several models that analyze the behavior of customers.

The proposed method of analysis involves the following steps:

1. Based on the company's current marketing objectives and tasks determine the rate of customer loyalty, measure should take into account the frequency of purchases and buyers recurrence after the first purchase.
2. On the basis of available data on the orders of customers determine the various characteristics (such as the breadth of the categories, the average and the total value of the order, etc.).
3. Build classifying and predictive models that reflect the the connection between descriptive characteristics and indication of loyalty.
4. Based on the analysis, identify the factors affecting customer loyalty, to formulate conclusions.
5. Test the hypothesis that the unfinished orders affect customer loyalty.
6. Compare the factors with factors traditionally used in e-commerce (eg, recency, frequency, and monetary in RFM-analysis).

Here is an example of the classifying model, which gets a parameter B2B or B2C mark, the number of orders, the breadth of categories in the order, the share of successfully completed orders, average order value, the average price of goods and the number of days from the date of the first order, and the output should attributed the client to one of three classes of loyalty (label "0" corresponds to a occasional clients, the label "1" - those who can not be properly classified and "2" – to loyal customers). Here and further, for the construction of classification models and the tables with the results, we use package for data processing in the social sciences - IBM SPSS Statistics (Module Analysis - Classification).

Here we use a classification, rather than classic statistical significance analysis, in order to better identify the dependence in the analyzed data, as proceed from the assumption that the required dependencies can have a nonlinear nature and have a more complex structure. In addition, the construction of the classifying models also allows predicting the target value. The variables used in the simulation:

- order cnt - the total number of orders;
- channel - B2C or B2B-label customers;
- day_cnt - the number of days from the first order;
- cnt_cat - breadth of categories presented in orders;
- cnt_obj - the average number of positions in the order;
- order_pos_proc - the share of successfully completed orders;
- avg_object_cost - the average price of items in the order;
- avg_order_cost - the average order value.

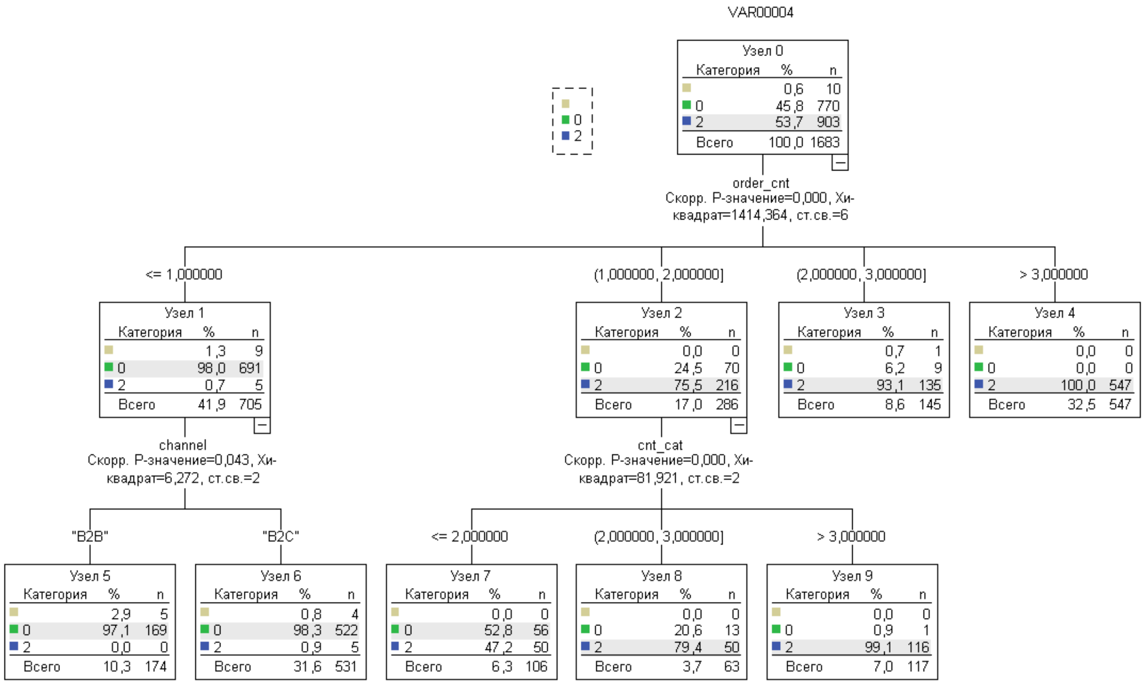

Figure 3. A decision tree with three classes for solving the task of customer classification by loyalty.

As a classification algorithm, we used the standard implementation of a well-known method for constructing decision tree (Murthy, 1998). We have built the decision tree uses as the basic dividing variable - the number of orders, it is quite logical result and a clear factor. Among those who did only one order model predicts 98% nonreturn customers, among those who have made at least 2 orders only a quarter is nonreturn. Starting with 4 orders 100% of customers are loyal. As additional parameters model also uses the tag «B2B / B2C» and the breadth of the categories presented in the order of goods. These parameters, instead of timing and monetary are helped to clarify the classification. The quality classification and the results are presented in Table 1.

Table 1. Quality of classification using the decision tree

Risk

evaluation	standard error
,052	,005

design method: CHAID
dependent variable: VAR00004

Classification

observational	predicted			
		0	2	The percentage of properly classified customers
	0	9	1	0,0%
0	0	747	23	97,0%
2	0	55	848	93,9%
The total percentage	0,0%	48,2%	51,8%	94,8%

design method: CHAID
dependent variable: VAR00004

The table confirms the applicability of the method for this problem, the data confirm the acceptability of the table as a model.

More formally, we can conclude that if the model for the prediction of loyalty uses the number of orders or frequency, as in the classic RFM-analysis, than it is pointless to recency or monetary characteristics. Instead, the relevant factors are the breadth of provided in orders and customer type categories (B2C or B2B).

The question of how to change the simulation results, if we exclude from the model number of orders. Next - discriminant model was built without the use of this parameter.

Table 2 shows the structural matrix, which reflects the correlation between the used variables and the resulting discriminant function.

Table 2. The structural matrix of correlations for the resulting discriminant function

	Correlations of variables and values of the discriminant function
	1
day_cnt	,976
cnt_cat	,566
cnt_obj	,298
order_pos_proc	-,154
avg_object_cost	-,040
avg_order_cost	-,004

The table shows that the largest contribution to the classification makes the number of days from the first order, followed by a breadth of categories and the average number of positions in the order. Note that the number of successfully completed orders begins to significantly affect the loyalty classification only if we exclude from the model the number of orders and information on the time passed from the first order. And even in this case, splitting comes at first by a breadth of categories, and only then by the share of completed orders. Thus, the share of completed orders has virtually nothing to do with customer loyalty.

4. Conclusion

The article provides a method of working with data on customers purchase history on a data sample provided by Russia's largest online store of tools and equipment. Significant factors in the analysis of consumer behavior are shown; particularities of the definition and construction of features of customers are described. The results of the statistical analysis considered data (customers order history at the online store).

Result of the analysis was the identification of important factors that determine the recurrence customers - is the total number of orders (in this connection is pretty obvious), the duration of the purchase, as well as the breadth of categories represented in the orders. The hypothesis that the share of successfully completed orders significantly affects to a positive customer experience and customer loyalty was denied. As well as the fact that loyalty can be determined by the total or average order value. Advantage of the methods proposed in this paper is that the work is carried out with standard CRM-system that does not contain any additional user attributive descriptions.

References

1. Albescu F., Pugna I. (2014) Marketing Intelligence - The Last Frontier of Business Information Technologies. *Romanian Journal of Marketing*, Issue 3, p55-68.
2. Ball, D., Coelho, P. S., & Vilares, M. J. (2006). Service personalization and loyalty. *Journal of Services Marketing*, 20(6), 391-403.
3. Carbone L.P., Berry L.L. (2007) Build Loyalty Through Experience Management. *Quality Progress*, Vol. 40, No. 9, pp. 26-32
4. Chung, T. S., & Wedel, M. (2014). 18 adaptive personalization of mobile information services. *Handbook of Service Marketing Research*, 395.
5. Fader, P. S., Hardie, B. G., & Lee, K. L. (2005) RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), 415-430.
6. L. Jasrai (2014) Predicting Customer Satisfaction Towards Mobile Value-Added Services: An Application of Multiple Regression. *The IUP Journal of Marketing Management*, Vol. XIII, No. 1.
7. Murthy S. (1998). Automatic construction of decision trees from data: A multidisciplinary survey. *Data Mining and Knowledge Discovery*. Volume 2, Issue 4 , pp. 345-389.
8. Phillips-Wren G., Hoskisson A. (2015) An Analytical Journey Towards Big Data. *Journal of Decision Systems*. Vol. 24, No. 1, 87–102
9. Schmitt B.H. (2003) Customer Experience Management: A Revolutionary Approach to Connecting with Your Customers. John Wiley & Sons.
10. Shaun Smith and Joe Wheeler (2002) Managing the Customer Experience: Turning customers into advocates. Financial Times Press. Paperback, 272 pages.
11. Solomon, M. R. Consumer (2004) Behavior. Buying, Having, and Being, Pearson Prenctice Hall. Saddle River. 621 p.
12. Spiess J., T'Joens Y., Dragnea R., Spencer P., Philippart L. (2014) Using Big Data to Improve Customer Experience and Business Performance. *Bell Labs Technical Journal*, 18(4), 3–17
13. Zomerdijk, L.G. and Voss C.A. (2010) Service Design for Experience-Centric Services. *Journal of Service Research.* 13 (1) Pp. 67-82.