

# Marketing Segmentation in Consumer Product Industry

Xuan Zhao<sup>1</sup>

<sup>1</sup>The Chinese University of Hong Kong, Faculty of Business Administration, Hong Kong, China

**Abstract**—Marketing segmentation is widely used for targeting a smaller group of market in consumer product industry, and is useful for decision makers to effectively focus on their consumers' purchasing behavior based on one marketing mix. In order to have a better understanding of the relationship between marketing segmentation in conventional business field and data science field, marketing segmentation is introduced from the perspectives of business and data science in this paper. In addition, on the basis of the introduction of both methods, conclusion can be drawn that data science methods are used in marketing segmentation as an assistance to present more precise output, based on the theory generated in business practice.

## 1 INTRODUCTION

Consumers have their own behavior patterns with different demands. What consumer product industry do is to meet the needs of different consumers. Market segmentation can be described as the process of partitioning a large market into the smaller groups or clusters of customers [2]. Each group gathers consumers with similar behaviors and marketing segmentation strategy can be cultivated through business and data science method. For business, consumer market segmentation divides market into small groups by using different variables, including geographical, demographic, behavior, psychographic variables. To better fit a business oriented market segmentation, market can also be divided through three variables, namely geography, consumer type and buyer behavior. What is more, data science algorithms such as decision tree and cluster are discussed in this paper. There is no doubt that market segmentation can increase the efficiency and specificity for business. Also, the actions taken after market segmentation can let consumers realize the importance that businesses attach to them.

## 2 MARKET SEGMENTATION BY USING BUSINESS METHODS

With the increase of customer diversity, marketing segmentation is widely used in companies to make a distinction between their products and services and that of the competition [9]. Moreover, products involved in marketing programs designed for particular market segments may achieve a broader acceptance than originally planned, thus revealing a basis for convergence of demand and a more generalized marketing approach [11]. Therefore, marketers use marketing segmentation to target a certain segment, maximizing profits at the lowest

costs. Two types of marketing segmentation methods will be introduced below, namely consumer market segmentation and business oriented market segmentation.

### 2.1 Consumer Market Segmentation

Consumer market segmentation aims at partitioning consumers into different segments according to different criteria, which includes geographical, demographic, behavior and psychographic.

#### 2.1.1 Geographical segmentation

Geographical segmentation divides market based on locations. Consumers who live in the same location may share similar demands, which may be different from the demand of those who live in other regions in the world [9]. And sometimes geographical segmentation depends on culture diversity. For example, vodka is loved by Russians, and can be seen very often on Russians' table. However, Chinese have their own liquor, and Japanese have their Sake as well as Korean have their Korea Shochu. Markets in those countries can be defined as different market segments, marketers can make their own strategy to target a segment, selling specific alcohol to the target country to make profit effectively and efficiently.

#### 2.1.2 Demographic segmentation

Demographic segmentation identifies consumer subgroups that are homogeneous in view of demographic factors, such as age, sex and occupation [1]. Different demographic subgroups have different wants and needs. Taking age as an example, Coke Cola may target consumers in an age group between 12 and 24, and Starbucks may target consumers beyond 30 years old. Under today's market competition, demographic

[zxuan1121181158@gmail.com](mailto:zxuan1121181158@gmail.com)

segmentation is considered as a great advantage to any company.

### *2.1.3 Behavior segmentation*

Behavior segmentation divides consumers into segments depending on their behavioral patterns, like reaction towards discounts, loyalty to a certain brand. Consumers always act distinctively towards same message. For instance, when people see promotion emails or posters, some of them may buy the product immediately if the product looks satisfactory to them. However, others may check that if other buyers are really satisfied with the product before they make the order. With behavior segmentation, some companies can target a group of customers, and build customer loyalty to their brand easily.

### *2.1.4 Psychographic segmentation*

Psychographic segmentation groups customers found on factors like personality and lifestyle [6]. Lifestyles of customers are defined by the consumption choices they made in various product categories. Hence, lifestyle can be quantitatively defined and used to recognize a market segment [8]. Likewise, personality can be reflected on their apparel, and clothing company can take personality as a segmentation criteria to sell cloths to suitable segments. Thus, psychographic variables are also effective for partitioning and targeting sub-markets.

## **2.2 Business Oriented Market Segmentation**

Business oriented market segmentation does not make much difference with consumer market segmentation, as it focuses on one specific segment to maximize profits. Taking the previous example, if geographical segmentation for business is used, company will only sell vodka, and sell it directly to Russia without considering other countries. Customer type segmentation is used when companies need to select a certain customer type. In order to allocate resources effectively and reduce

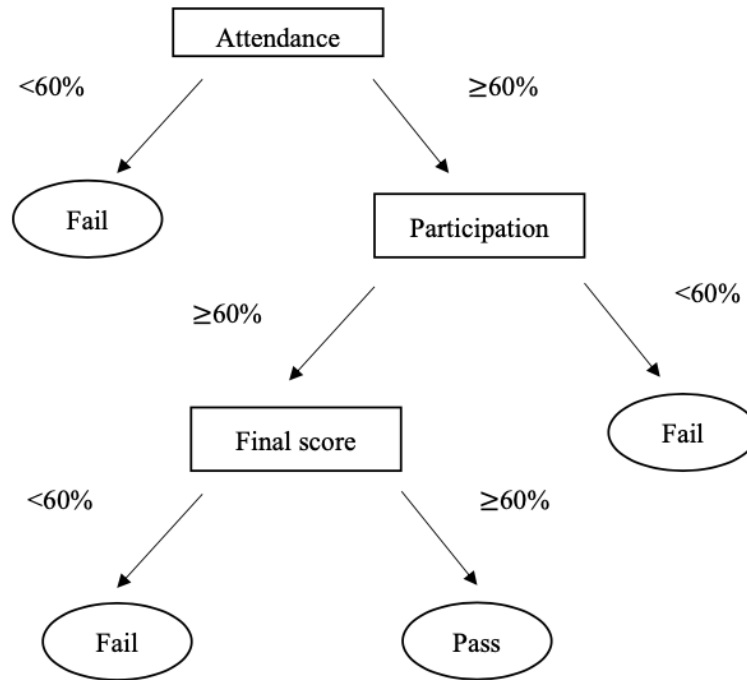
unnecessary cost to increase profits, companies choose to tailor each product and market tactic to that particular target markets [9]. Same as geographical segmentation and customer type segmentation, companies use buyer behavior segmentation to target a certain group of customers and customize sale pattern for these customers to make them more engaged in purchase, so as to enhance their loyalty to the brand.

## **3 MARKET SEGMENTATION BY USING ALGORITHMS**

With the development of computer science and statistics, marketers are eager to apply data science approaches to marketing segmentation to replace a lot of tedious work. Undoubtedly, with data science approaches, companies have been able to save a large amount of time and money in the process of collecting, processing and analyzing data. Furthermore, in order to meet the requirements of the modern companies, the outputs of marketing segmentation have also become diversified. In particular, visualization gives people who can not read the code or do not have statistics knowledge a chance to understand the output easily, then they can make decisions or market tactic adjustment according to the segmentation output directly.

### **3.1 Decision Tree**

Decision tree is a simple measure to visualize a decision, and it is the reason that decision trees have been applied in the analysis of marketing segmentation. Decision tree consists of three components, which are decision nodes, branches and terminal nodes or end nodes. Decision nodes utilize the chosen variables and implement a logical test to split up data into two or more descendant nodes. Branches connect decision nodes and descendant nodes. Terminal nodes represent the answer or output of a decision tree, which are related to the maximum number of individuals in the nodes [4].



**Figure 1.** Decision tree for grade system

\*Squares: Decision nodes; Arrows: Branches; Circles: End nodes

In figure 1, multistage decision making processes are performed by decision tree. Decision tree is a supervised machine learning method, and independent variables in each decision node need to be cut into next nodes according to their dependent variables. Take binary decision tree as an example, the cutpoints, also known as the thresholds, are valued 60% in figure 1. To some extent, the performance of decision tree depends on the goodness of the thresholds, which is measured by the impurity of the resulting two children nodes:

$$I = I_{n1} + I_{n2} \quad (1)$$

where  $I$  is the calculated impurity,  $I_1$ ,  $I_2$  representing the impurity of each node, and  $n_1$ ,  $n_2$  are the number of individuals in each node.

Decision tree is common in the consumer segmentation analysis, and it allows decision makers or marketers to have an insight into each segments, and their suitable products or services. However, decision tree is interpretable up to a point, and it becomes more difficult to interpret as the tree grows deeper. Additionally, overfitting is a problem that cannot be ignored.

### 3.2 K-means Clustering

Clustering analysis is a machine learning method, and it automatically finds natural grouping in data. Many clustering methods use the distance between two independent variables to measure the similarity between a pair of objects, and there may exist centroid for each cluster. Any instance that is closer to the cluster than other clusters will be grouped together, which means the  $d(x_i, x_{c_j})$  is the smallest, where  $d(x_i, x_{c_j})$  denotes the distance between any object and the centroids.

K-means is a clustering method to minimize error, which is also used to find a clustering structure with the

minimal certain error criterion, and the most well-known error criterion is Sum of Squared Error (SSE) [10].  $K$  means partitions data into  $K$  clusters, where  $K$  can be determined by using silhouette score, and the centers of each cluster are calculated as the mean of all the objects in that cluster:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

where  $\mu$  is the average of all instances in the cluster, and  $N$  stands for the number of instances in the cluster.

K-means starts with  $K$  pre-set centroids. In each iteration, instances are allocated to the closest cluster according to their Euclidean distance with the averages, after which the averages of each clusters are re-calculated. Iterations will stop when the SSE is not reduced by the new set of clusters, or meet some pre-requisites, for instance, exceeding a certain number of iterations:

$$E(\mu_1, \mu_2, \dots, \mu_K) = \sum_{i=1}^N \sum_{m=1}^K I(x_i \in C_m) |x_i - \mu_m| \quad (3)$$

where  $\mu$  is the mean of each cluster,  $N$  stands for the number of all instances,  $K$  is the number of clusters, and  $C_m$  is the cluster dataset that contains  $x_i$ ,  $I(P)$  is 1 if  $P$  is true and otherwise.

Moreover, global k-means method can also be used in clustering analysis without randomly selecting values for all centroids. This method starts with one cluster ( $m=1$ ) in its optimal position related to the centroid of the whole data set, and for two clusters ( $m=2$ ),  $N$  executions need to be performed. In execution  $n$ ,  $x_n$  is the center of second cluster, and the centroid of the first cluster is always placed at the optimal position for  $m=1$ , and the best solution obtained after  $N$  executions is the new optimal

position for two clusters [7]. Same operation happens when a new cluster added until the number of clusters reaches the number required.

K-means algorithm is widely used in segmentation analysis, especially with some programming packages like Wordcloud in Python, therefore decision makers can see the characteristics of customers more intuitively from the clusters.

From a case study of customer segmentation analysis by using machine learning methods, it is obvious to see the relationship between business and data science method in marketing segmentation field. The case covered an e-commerce database listing purchase made by approximately 4,000 customers, and it divided customers into 11 segments in customer categories section. Instead of clustering customers directly, the study created several variables, for instance, minimum, maximum, average amount and total amount spent, and product categories according to precede product analysis, through the clustering, customer types are distinguished by those feature. Apart from customer clustering, which is just a section in customer analysis, the study also provided that how data was processed, and which variable were to be selected, beyond the study, what should companies do after they have the customer clustering results [3]. Obviously, conventional business methods that are used in marketing segmentation offer a guideline for marketing segmentation analysis today, and data science methods are introduced into this field as an assist to make the results more reliable.

## 4 DISCUSSION

Marketing segmentation partitions market into several segments, it is useful for marketers to target consumers and make their own marketing tactic. In the era of underdeveloped information, marketers used to get information from consumers through questionnaires, phone calls and emails. It is found in the information that customers act differently in their buying behaviors, and different customers tend to show their favor to different products. In further investigation, there are specific factors like age, reactions toward discount and regions, which can divide those customers into different groups, later marketers can partition market based on those factors, and this is the prototype of marketing segmentation. However, limited by the data science skills at that time, only rough results can be obtained, although they cost lots of money and human resources.

With the development of data analysis, statistics, math and programming skills are used in the marketing segmentation filed to make sure that precise output can be obtained. Those categories used in business, such as geographic and demographic factors, are considered as features in data analysis, and marketing segmentation can be achieved as stated by those features. Nevertheless, there is no perfect algorithm for marketing segmentation. Take k-means as an example, since means of clusters need to be defined, k-means can only deal with data with quantitative attributes. Furthermore, k-means method is not suitable for clusters with nonconvex shape or clusters

of very different size. Finally, k-means is very sensitive to outliers, which can affect the means significantly and produce wrong output [5].

## 5 CONCLUSION

Nowadays, data science methods such as machine learning methods are widely used in business field, not only the results can be more precise, but also the money and human resources can be saved and reallocated. Nonetheless, people who only master data science skills are not well qualified for this job. Marketing segmentation by using business methods has developed a complete system, involving typical insights of consumers meanwhile having a better understanding of business. Marketing segmentation can also be helpful for those who partition consumers by applying algorithms. Simultaneously, the difficulty of selecting staffs also increases notably. Marketing segmentation by using algorithms have been widely adopted by companies, and it is recommended that business methods for marketing segmentation should be learned as it is the traditional way used in marketing segmentation. The trend of marketing segmentation is to combine business and data science methods. Business methods are used to determine the features that can best cluster consumers, then data science methods are performed to get accurate output based on collected database.

## ACKNOWLEDGEMENT

I am very grateful to Professor Stephen Coggeshall, teaching assistant Parker Jiang and Jane Zhang for their guidance and help during the preparation and writing. Without them, I could not finish this paper.

## REFERENCES

1. Boslaugh, S. E., Kreuter, M. W., Nicholson, R. A. & Naleid, K. (2004, November 30). Comparing demographic, health status and psychosocial strategies of audience segmentation to promote physical activity. *Health Education Research*, 20(4), pp. 430-438.
2. Croft, M. J. (1994). *Market segmentation: A step-by-step guide to profitable new business*. London, New York: Routledge.
3. Daniel, F. (2017). Customer Segmentation.
4. Florez, R. & Ramon, J. M. (2009, February). Marketing Segmentation Through Machine Learning Models: An Approach Based on Customer Relationship Management and Customer Profitability Accounting. *Social Science Computer Review*, 27(1), pp. 96-117.
5. Huang, J. J., Tzeng, G. H., & Ong, C. S. (2007, February). Marketing segmentation using support vector clustering. *Expert Systems with Applications*, 32(2), pp. 313-317.

6. Kotler, P. (1994). *Marketing management, analysis, planning, implementation, and control*. New Jersey: Prentice Hall.
7. Likas, A., Vlassis, N., & Verbeek, J. (2003, February). The global k-means clustering algorithm. *Pattern Recognition*, *36*(2), pp. 451-461.
8. Lin, C. F. (2002). Segmenting customer brand preference: demographic or psychographic. *Journal of Product & Brand Management*, *11*(4), pp. 249-268.
9. Martin, G. (2011, June). The Importance of Marketing Segmentation. *American Journal of Business Education*, *4*(6), pp. 15-18.
10. Rokach, L., & Maimon, O. (2005). Clustering Methods. In L. & Rokach, *Data Mining and Knowledge Discovery Handbook* (pp. 321-352). Boston: Springer.
11. Smith, W. R. (1956). Product differentiation and market segmentation as an alternative marketing strategy. *Journal of Marketing*, *21*(1), pp. 3-8.