UNDERSTANDING CUSTOMERS - PROFILING AND SEGMENTATION

Mircea Andrei SCRIDON
Babes Bolyai University of Cluj-Napoca

Abstract:
In any industry, the first step to finding and creating profitable customers is determining what drives profitability. This leads to better prospecting and more successful customer relationship management. Any company can segment and profile their customer base to uncover those profit drivers using the knowledge of their customers, products, and markets. Or they can use data-driven techniques to find natural clusters in their customer or prospect base. Whatever the method, the process will lead to knowledge and understanding that is critical to maintaining a competitive edge.

Key words: profiling, segmentation, penetration analysis, cluster analysis.

The methodologies discussed in this paper describe several techniques and applications for understanding customers. Common sense tells us that it is a good first step to successful customer relationship management. It is also an important step for effective prospecting. In other words, once companies know what customer attributes and behaviors are currently driving their profitability, they can use these to direct their prospecting efforts as well. The first step in effective prospecting is learning how to find prospects that look like customers for the company. It is also useful to segment and profile the prospect base to assist acquisition efforts. The goal in both cases is to identify what drives customer profitability.

The second half of the paper details the process using a case study. The study is from the household industry, in which it will show some profile and penetration analyses. Next, it will illustrate the use of cluster analysis to discover segments.

What is the importance of understanding customers?

Studies show that many companies operate for years— pumping out offers for products and services—without a clue of what their best customer looks like. For every company in every industry, this is the most important first step to profitable marketing. Similar to modeling, before a company begins any profiling or segmentation project, it is important to establish their objective. This is crucial because it will affect the way in which to approach the task. The objective can be explained by reviewing the definitions of *profiling* and *segmentation*.

Profiling is exactly what it implies: the act of using data to describe or *profile* a group of customers or prospects. It can be performed on an entire database or distinct sections of the database. The distinct sections are known as segments. Typically they are mutually exclusive, which means no one can be a member of more than one segment.

Segmentation is the act of splitting a database into distinct sections or segments. There are two basic approaches to segmentation: market driven and data driven. Market-driven approaches allow managers to use characteristics that they determine to be important drivers of their business. In other words, they preselect the characteristics that define the segments. This is why defining the objective is so critical. The ultimate plans for using the segments will determine the best method for creating them. On the other hand, data-driven approaches use techniques such as cluster analysis or factor analysis to find homogenous groups. This might be useful if companies are working with data about which they have little knowledge.

Types of Profiling and Segmentation

If a company have never done any segmentation or modeling, its customer base may seem like a big blob that behaves a certain way, depending on the latest stimulus. If the company does a little digging, it will find a variety of demographic and psychographic characteristics as well as a multitude of buying behaviors, risk patterns, and levels of profitability among the members of its database. This is the beauty of segmentation and profiling. Once companies understand the distinct groups within the database, they can use this knowledge for product development, customer service customization, media and channel selection, and targeting selection.

RFM: Recency, Frequency, Monetary Value

One of the most common types of profiling originated in the catalog industry. Commonly called RFM, it is a method of segmenting customers on their buying behavior. Its use is primarily for improving the efficiency of marketing efforts to existing customers. It is a very powerful tool that involves little more than creating segments from the three groups.

Recency. This value is the number of months since the last purchase. It is typically the most powerful of the three characteristics for predicting response to a subsequent offer. This seems quite logical. It says that if a consumer has recently purchased something from a company, it is more likely to make another purchase than someone who did not recently make a purchase.

Frequency. This value is the number of purchases. It can be the total of purchases within a specific time frame or include all purchases. This characteristic is second to *recency* in predictive power for response. Again, it is quite intuitive as to why it relates to future purchases.

Monetary value. This value is the total currency amount. Similar to frequency, it can be within a specific time frame or include all purchases. Of the three, this characteristic is the least powerful when it comes to predicting response. But when used in combination, it can add another dimension of understanding.

These three characteristics can be used alone or in combination with other characteristics to assist in CRM efforts.

Demographic

The emphasis is on finding the individual who may not fit the local demographic profile. In reality, though, many people who live in the same area behave in a similar fashion. There are many sources of demographic data. Many sources are collected at the individual level with enhancements from the demographics of the surrounding geographic area. Segmenting by values such as age, gender, income, and marital status can assist in product development, creative design, and targeting. There are several methods for using demographics to segment a database and/or build customer profiles.

Life Stage

With few exceptions, our lives follow patterns that change over time to meet our needs. These patterns are clustered into groups defined by demographics like age, gender, marital status, and presence of children to form life stage segments. Life stage segments are typically broken into young singles; couples or families; middle-aged singles, couples, or families; and older singles or couples. Additional enhancements can be achieved by overlaying financial, behavioral, and psychographic data to create well-defined homogeneous segments. Understanding these segments provides opportunities for businesses to develop relevant products and fine-tune their marketing strategies.

The market segments should pass the following rules in order to be taken into consideration:

• They should be relevant to the business objective

• They should be understandable and easy to characterize

• They should be large enough to warrant a special offering

• They should be easy to develop unique offerings

After testing which of the segments have passed the above rules, a company should evaluate the results behaviorally and financially to determine which segmentations and offerings should be expanded to the target population, and also how good did they perform against the business objectives, set by it. Managers often fail to remember the business objectives until it is too late. It is critical that they have designed the segmentations to satisfy a business objective and that they have evaluated the market tests based on those same business objectives. If not, the fall-out could be costly on other fronts, such as lower profitability, reduced product usage, or negative changes in attitude or expectations. By keeping the business objectives in perspective throughout the development, testing, and analysis stages, companies are more assured of meeting their goals, maximizing their profitability and improving their customers' long-term behavior.

Profiling and penetration analysis of a catalog company's customers

The following case study is about a company which offers household items. It has been operating for six years and now has a database of 14244 customers. It is interested in expanding its customer base. It is therefore looking for ways to identify good prospects from outside sources. The first step is to perform RFM analysis.

RFM Analysis

As it has been mentioned earlier, recency, frequency, and monetary value are typically the strongest drivers of response for a catalog company. To discover the effects of these measures on company's database, the relevant variables are identified in the database:

last purchase (lp) - Months since last purchase or *recency*.

number of purchases (np) - Number of purchases in the last 36 months or *frequency*.

total purchases (tp) - Total monetary amount of purchases in the last 36 months or *monetary value*.

The first step is to get a distribution of the customers' general patterns. The tables below provide a good overview of customer buying habits for the company:

Table 1

Recency of purchases

Lp (months)	Percent	Cumulative percent
0-2	17,1	17,1
3-5	13,6	30,7
6-8	41,2	71,9
9-12	28,1	100,0

Table 2

Number of purchases

Np	Percent	Cumulative percent
0-1	30	30
2-6	43,1	73,1
7-14	19,8	92,9
15-20	7,1	100,0

Table 3

Total purchases

Tp (€)	Percent	Cumulative percent
<100	37,9	37,9
100-200	24	61,9
200-300	10,2	72,1
300-400	5,4	77,5
400-500	4,1	81,6
500+	18,4	100,0

It can be observed that the majority of customers haven't purchased anything for at least six months. A large percentage of customers made between two and six purchases in the last year with 92,9% making fewer than seven purchases. The total monetary value of yearly total purchases is mainly below 100 €, with 77,5% below 400 €.

The next step is to look at the response rate from a recent mailing activity to see how these three drivers affect response. The database is sorted by *recency* and created quintiles from it (equal fifths of the file). By calculating the response rate for each quintile, the relationship between *recency* and *response* can be determined.

It can be seen that the measure with the strongest relationship to *response* is recency according to table 4.

Table 4

Quantile – Recency

Recency	Response rate
Quantile	
Q1	0,21
Q2	0,082
Q3	0,018
Q4	0,012
Q5	0,005

Table 5

Quantile – Frequency

Frequency	Response rate
Quantile	
Q1	0,145
Q2	0,03
Q3	0,019
Q4	0,012
Q5	0,011

Table 6

Quantile – Monetary

Monetary	Response rate
Quantile	
Q1	0,078
Q2	0,041
Q3	0,037
Q4	0,022
Q5	0,015

The following figure compares recency, frequency, and monetary value as they relate to response.

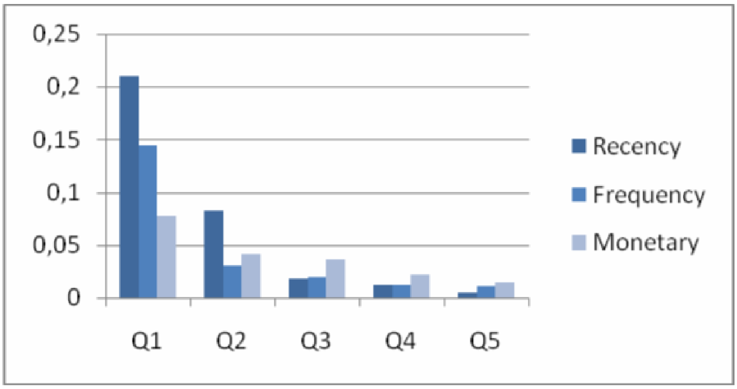

Figure 1. Quantile - Response rate

Again, it can be seen that the recency of purchase is the strongest driver. This is a valuable piece of information and can be used to target the next campaign. In fact, many companies include a new catalog in every order. This is a very inexpensive way to take advantage of recent purchase activity.

Penetration Analysis

As it has been mentioned earlier, the study company wants to explore cost-effective techniques for acquiring new customers. Penetration analysis is an effective method for comparing the distribution of the customer base to the general population. Many companies sell lists that cover a broad base of the population. The methodology is simple. First it procedure begins with a frequency distribution of some basic demographic variables.

In this case, the variables age and income are selected. The figure below shows the output from for these two variables. This gives information about the distribution of the company's customers. It can be noticed how 35%

of the customers are between the ages of 41 and 44.

Table 7

Age – Customers

Age	Percent	Cumulative percent
<25	0,3	0,3
26-30	1,7	2
31-35	10,2	12,2
36-40	28,4	40,6
41-44	35	75,6
45-54	15,3	90,9
55-64	7,6	98,5
65+	1,5	100

Table 8

Income – Customers

Income (€)	Percent	Cumulative percent
<400	1,1	1,1
400-700	26,4	27,5
700-900	27,2	54,7
900-1200	25,5	80,2
1200-1500	18,2	98,4
1500+	1,6	100

In order to make use of this information for new acquisition marketing, it is needed to compare this finding to the general population. The next table showes similar profiles for the general population:

Table 9

Age – General population

Age	Percent	Cumulative percent
<25	1,4	1,4
26-30	2,3	3,7
31-35	12	15,7
36-40	26,9	42,6
41-44	30,5	73,1
45-54	18,4	91,5
55-64	7,6	99,1
65+	0,9	100

Table 10

Income – General population

Income (€)	Percent	Cumulative percent
<400	1,6	1,6
400-700	25,2	26,8
700-900	28,4	55,2
900-1200	23,9	79,1
1200-1500	16,1	95,2
1500+	4,8	100

It can be noticed how the second figure displays the same distributions as the first table except this time they are on the general population.

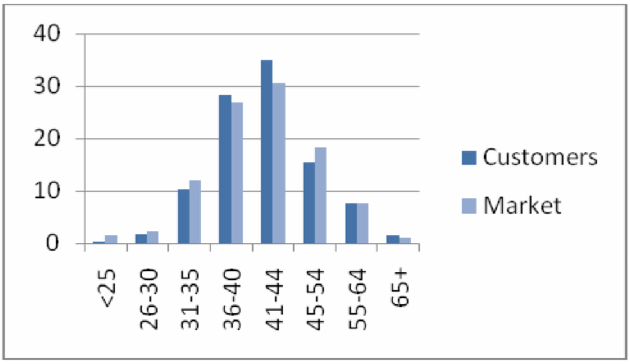

Figure 2. Age distribution among customers and general population

The figure above shows a market comparison graph of age. It provides a graphical display of the differences in distribution for the various age groupings.

The following table brings the information from the two analyses together and creates a measure called a penetration index. This is derived by dividing the customer percentage by the market percentage for each group and multiplying by 100.

Table 11

Penetration analysis – Age

Age	Percent of customers	Percent of market	Penetration index
<25	0,3	1,4	21,42 %
26-30	1,7	2,3	73,91 %
31-35	10,2	12	85 %
36-40	28,4	26,9	105,57 %
41-44	35	30,5	114,75 %
45-54	15,3	18,4	83,15 %
55-64	7,6	7,6	100 %

Table 12

Penetration analysis – Income

Income	Percent of customers	Percent of market	Penetration index
<400	1,1	1,6	68,75 %
400-700	26,4	25,2	104,76 %
700-900	27,2	28,4	95,77 %
900-1200	25,5	23,9	152 %
1200-1500	18,2	16,1	106,69 %
1500+	1,6	4,8	33,33 %

The company would be wise to see new customers in the 36–44 age group. This age range is more prominent in its customer base than in the general population. Also wise would be to take in consideration new customers whose income is situated in the 400-700 € group and 1200-1500 € group.

Performing cluster analysis to discover customer segments

Cluster analysis is a family of mathematical and statistical techniques that divides data into groups with similar characteristics. Clustering uses Euclidean distance to group observations together that are similar across several characteristics, while attempting to separate observations that are dissimilar across those same characteristics. It is a process with many opportunities for guidance and interpretation.

Several algorithms are used in clustering. In this case study, a clustering algorithm is used that is designed for large data sets. It begins by randomly assigning cluster seeds or centers. The number of seeds is equal to the number of clusters requested. Each observation is assigned to the nearest seed. The seed is then reassigned to the mean in each cluster. The process is repeated until the change in the seed becomes sufficiently small.

To illustrate the methodology, two variables (age and income) are used from the company data in this case study. Before the cluster analysis begins, the variables must be standardized. Because the clustering algorithm used depends on distance between variable values, the scales of the variables must be similar. Otherwise, the variable with the largest scale will dominate the clustering procedure.

In the next figure, the output displays the distance from the seeds to the farthest point as well as the distances between clusters. The cluster means do show a notable difference in values for age and income.

Table 13

Cluster summary

Cluster	Frequency	RMS Std. deviation	Maximum distance from seed to observation	Nearest cluster	Distance between cluster centroids
1	10185	0,71	3,21	2	1,62
2	3512	0,82	3,56	1	1,62
3	547	0,98	4,02	1	3,73

Statistics for variables

Variable	Total STD	Within STD	R-squared
Age	1	0,58	0,52
Income	1	0,69	0,47
Over-all	1	0,75	0,51

Pseudo F statisitc = 24321,25
Approximate expexted over-all R-squared = 0,586

Cluster means

Cluster	Age	Income
1	0,54247	0,10581
2	1,31591	-0,42131
3	-0,21656	2,98939

The company can now tailor their marketing campaigns to each group separately. Similar to the profile analysis, understanding the segments can improve targeting and provide insights for marketers to create relevant offers.

Conclusions

A sucesful profiling and segmentation process demands that a company should define its business objectives. At the start of any segmentation process, management should agree on and clearly state their goals using language that reflects targeting and measurement. Business objectives can be (1) new account, sales, or usage driven; (2) new product driven; (3) profitability driven; or (4) product or service positioning driven.

Furthermore types of data could include survey, geo-demographic overlays, and transactional behavior. Data must be relevant to the business objectives. The process involves reviewing all data to determine only the necessary elements because collecting and analyzing data on all customers or prospects is very time-consuming and expensive.

The segmentation process means selecting a method that is appropriate for the situation. There are three segmentation methods that could be employed: predefined segmentation, statistical segmentation, or hybrid segmentation.

The predefined segmentation method allows the analyst to create the segment definitions based on prior experience and analysis. In this case, the data is known, the work involves a limited number of variables, and a limited number of segments are determined. The appropriate segments will be defined and selected based on the business objective and the knowledge of the customer base.

The statistical method should be employed when there are many segments involved and there is little or no experience with the population being investigated.

Hybrid segmentation allows the analyst to combine predefined segmentation with statistical segmentation, in any order, based on the success in deriving segments.

REFERENCES

[1] Buiga, Anuța, Florea I., Lazăr, Dorina, Parpucea, Ilie, *Statistică inferențială*, Presa Universitară Clujeană, Cluj-Napoca, 2000.

[2] Hand, David, Mannila, Heikki, Smyth, Padhraic, *Principles of Data Mining*, The MIT Press, Cambridge, MA, 2001.

[3] Hughes, Arthur M, *Strategic Database Mining*, Probus Publishing, Chicago, 1994.

[4] Pop, Marcel C., *Implicațiile pieței asupra planificării strategice de marketing*, Risoprint, Cluj-Napoca, 2005.

[5] Rud, Olivia Parr, *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management*, John Wiley & Sons, New York, 2001.