Customer Analytics Part II

V. Kumar, W. Reinartz, *Customer Relationship Management*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-662-55381-7_6

Overview

In the previous chapter, we examined some traditional marketing metrics, various primary customer-based metrics, and explained some popular customer-based value metrics used in industry. Some of the primary customer-based metrics introduced earlier form the inputs to derive customer value – the key metric that drives decision making in the age of data-based marketing.

This chapter proceeds to conceptualize strategic metrics of customer value and introduces popular customer selection strategies and techniques to evaluate these strategies. Strategic customer-based metrics such as *recency, frequency, monetary value* (*RFM*), *past customer value, lifetime value metrics*, and the *customer equity* are forward looking measures. *RFM* value is a frequently used metric to predict, e.g., purchase behavior. The past customer *value (PCV)* assumes that the results of past transactions are an indicator of the customer's future contributions. Evaluating the long-term economic value of a customer to the firm is the goal of *lifetime value metrics*, which also form the basis for the calculation of the *customer equity*.

In addition to these metrics, some customer selection strategies are explained in ► Sect. 6.2. These techniques help firms identify the right customers in order to optimally allocate the available marketing resources. The chapter introduces three popular customer selection strategies: profiling, binary decision trees, and logistic regression. Finally, ► Sect. 6.3 discusses methods, such as misclassification rates and lift analysis, which companies can use to evaluate alternative selection strategies.

An overview of the metrics and methods discussed in this chapter is given in ◘ Table 6.1.

6

◘ Table 6.1 Metrics and methods used in customer analytics part 2

Metrics
6.1 Strategic customer-based value metrics
6.1.1 RFM value
6.1.2 Past customer value
6.1.3 Lifetime value metrics
6.1.4 Customer equity
Methods
6.2 Popular customer selection strategies
6.2.1 Profiling
6.2.2 Binary classification trees
6.2.3 Logistic regression
6.3 Techniques to evaluate alternative customer selection strategies
6.3.1 Misclassification rate
6.3.2 Lift analysis

6.1 Strategic Customer-Based Value Metrics

Strategic customer based value metrics are forward looking and aim to guide company decisions with the goal to maximize long-term profitability of the customer base. In particular we will introduce recency, frequency, and monetary value (RFM), past customer value, the lifetime value of a customer, and customer equity.

6.1.1 RFM Value

RFM stands for recency, frequency, and monetary value. This technique utilizes these three metrics to evaluate customer behavior and customer value and is often used in practice.

1. **Recency:** A measure of how long it has been since a customer last placed an order with the company.
2. **Frequency:** A measure of how often a customer orders from the company in a certain defined period.
3. **Monetary value:** The amount that a customer spends on an average transaction.

The general idea of RFM is to classify customers based on their RFM measure. The resulting groups of customers are associated with purchase behavior, e.g., likelihood to respond to a marketing campaign. RFM is similar to the transition matrix approach in that it also tracks customer

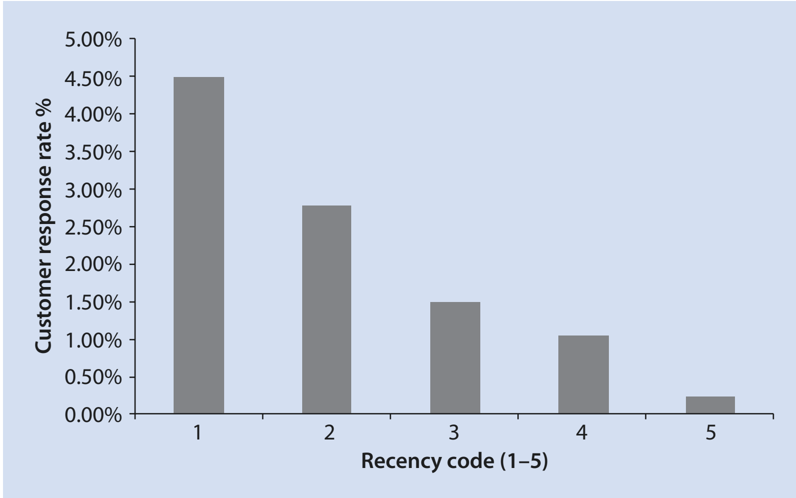


Fig. 6.1 Response and recency

behavior over time in what is called a state-space. That is, customers move over time through space with certain defined activity states.

RFM Method

For the following discussions about RFM coding, consider the example of a firm with a customer base of 400,000 customers. From this customer base, a sample of 40,000 customers is chosen. In other words, every tenth customer from the larger database of 400,000 customers was picked in order to form a test group of 40,000 customers who are representative of the whole customer base.

Also, assume this firm is planning to send a marketing mailer campaign of a $150 discount coupon to be mailed to its customers.[1]

Recency Coding

Assume this firm sends its $150 mailer campaign to the 40,000 customers in the test group, and assume that 808 customers (2.02% of 40,000) responded. In order to determine if there is any correlation between those customers who responded to the mailer campaign and their corresponding historical recency, the following analysis is done.

The test group of 40,000 customers is sorted in descending order based on the criterion of *most recent purchase date*. The earliest purchasers are listed on the top and the oldest are listed at the bottom. The sorted data are further divided into five groups of equal size (20% in each group). The top-most group is assigned a recency code of 1, the next group is assigned a code of 2, and so on, until the bottom-most group is assigned a code of 5. An analysis of the customer response data from the mailer campaign and the recency-based grouping point out that the mailer campaign received the highest response from those customers grouped in recency code 1, followed by those grouped in code 2, and so on. Figure 6.1 depicts the distribution of relative frequencies of customers who responded across the recency groups coded 1 through 5. The highest response rate (4.5%) for the campaign was from those customers in the test group who had the highest recency quintile (recency code = 1). Note the average customer response rate computed for all five groups would be none other than the actual response rate of 2.02% achieved by the campaign, i.e., (4.50% + 2.80% + 1.50% + 1.05% + 0.25%)/5 = 2.02%.

At the end of this recency coding exercise we would assign recency values of r = 1 through 5 for groups of customers, depending on the quintile that they belong to.

Frequency Coding

The frequency coding process is the same as the recency coding process just discussed. However, to sort the test group of 40,000 customers based on the frequency metric, we need to know the *average number of purchases* made by a customer per month. Of course, the choice of the appropriate time period depends on the usual frequency of

1 All numerical figures mentioned in the discussions below are hypothetical data created for instructional purposes only. However, due care has been exercised to ensure these data are fairly close to real life experiences of many firms.

purchases (e.g., weeks, months, quarters, years, etc.). In this case, customers with the highest number of purchases per month are grouped at the top, while those with lower number of purchases per month were listed below. Here again, the sorted list is grouped into five quintiles. Those in the top are assigned a code of 1 and those at the bottom a code of 5. The response rate of each of the frequency-based sorted quintiles is depicted in Fig. 6.2.

An analysis of the customer response data from the mailer campaign and the frequency-based grouping show that the mailer campaign received the highest response rate from those customers grouped in frequency code 1, followed by those grouped in code 2 (2.22%), and so on.

At the end of the frequency coding we would assign frequency values of f = 1 though 5 for groups of customers in the five frequency quintiles.

Monetary Value Coding

The monetary value coding process is exactly the same as the recency and frequency coding processes. However, to sort the test group of 40,000 customers based on the monetary value metric, we need to know the *average amount purchased per month*. As with recency and frequency, the customer data are sorted, grouped, and coded 1 to 5.

As can be seen in Fig. 6.3 the highest response rate (2.35%) for the campaign was from those customers in the test group who had the

Fig. 6.2 Response and frequency

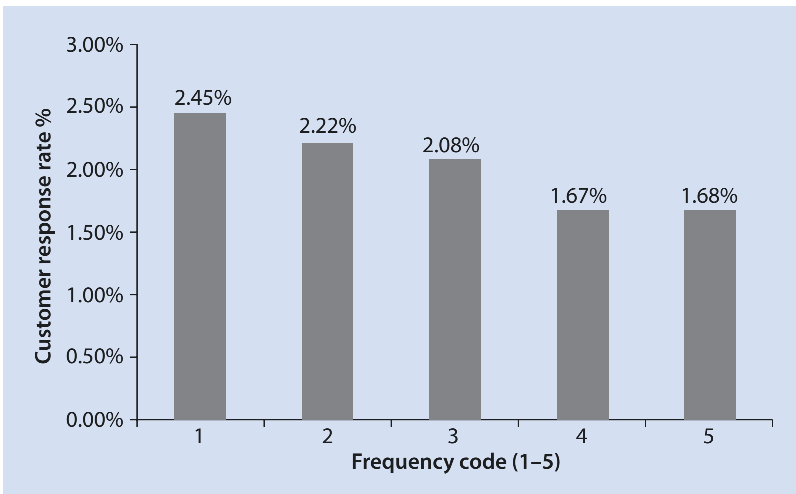
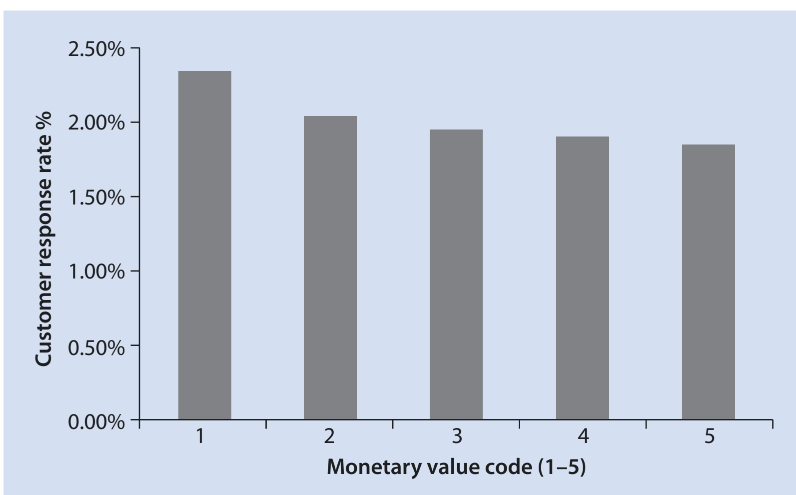


Fig. 6.3 Response and monetary value



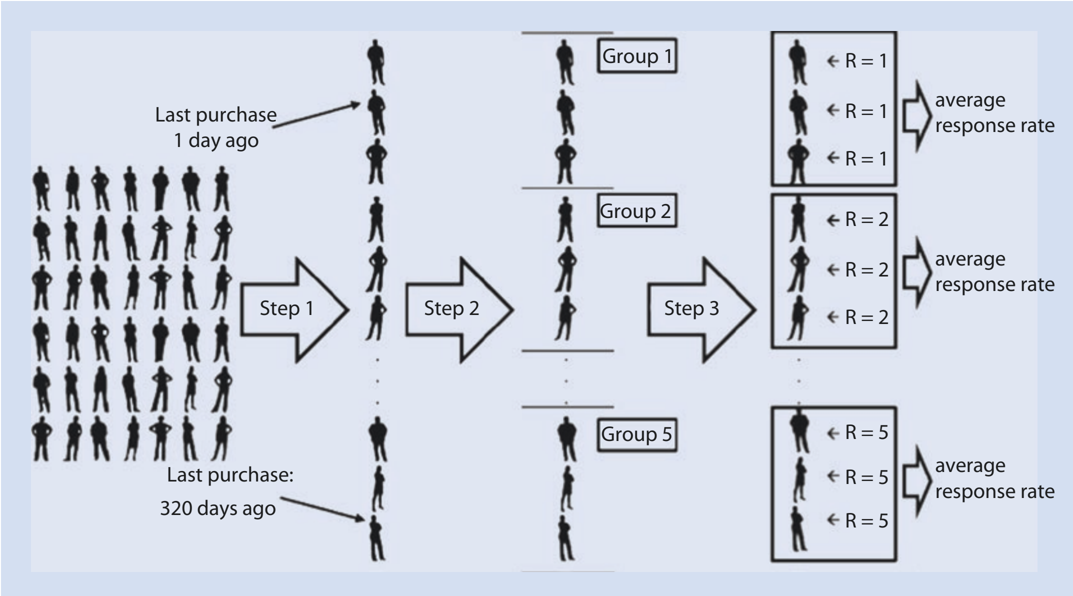


Fig. 6.4 RFM procedure

highest *monetary value* quintile (monetary value code = 1). Thus, indicating that the monetary value is also an important metric for the analysis of customer behavior.

At the end of this monetary value coding exercise, we would assign a monetary value of $m = 1$ through 5 for groups of customers, depending on the quintile that they fall within.

After performing the three steps (R, F, and M) you will have individual R, F, and M scores for each customer. Each customer will be assigned to one of the 125 groups such as 111, 233, 432, …, 555, based on her respective RFM code. An overview of the RFM procedure is given in Fig. 6.4.

Limitations

This method independently links customer response data with R, F, and M values and then groups customers belonging to specific RFM codes. However, this method may not produce an equal number of customers under each RFM cell. This is because the individual metrics R, F, and M are likely to be correlated. For example, someone spending more (high M) is also likely, on average, to buy more frequently (high F). However, for practical purposes, it is desirable to have exactly the same number of individuals in each RFM cell. A sorting technique ensuring equal numbers in each RFM cell is described as follows.

RFM Cell Sorting Technique

An alternative approach to applying RFM sequentially to the initial dataset is the RFM cell sorting. This is a more sophisticated sorting technique which helps to arrive at an RFM code for each customer and ensures the grouping of an equal number of customers under each RFM code. Figure 6.5 depicts a schematic diagram of the logic behind RFM cell sorting. Consider the list of 40,000 test group customers. The list is first sorted for recency and grouped into five equal groups of 8000 customers. Therefore group 1 will have 8000 customers, and so will the other groups through group 5. Now, take the 8000 customers in each group and sort them based on frequency and divide them into five equal groups of 1600 each. At the end of this stage, you will have RF codes starting from 11 through 55, with each group having 1600 customers. In the last stage, each of the RF groups is further sorted based on monetary value and divided into five equal groups of 320 customers each. Again, we will have RFM codes starting from 111 through 555, each having 320 customers. Considering each RFM code as a cell, we will have 125 cells (5 recency divisions * 5 frequency divisions * 5 monetary value divisions = 125 RFM codes).

Breakeven Value (BE)

To arrive at a decision which customers (more precisely: which customer «cells») to target, it is

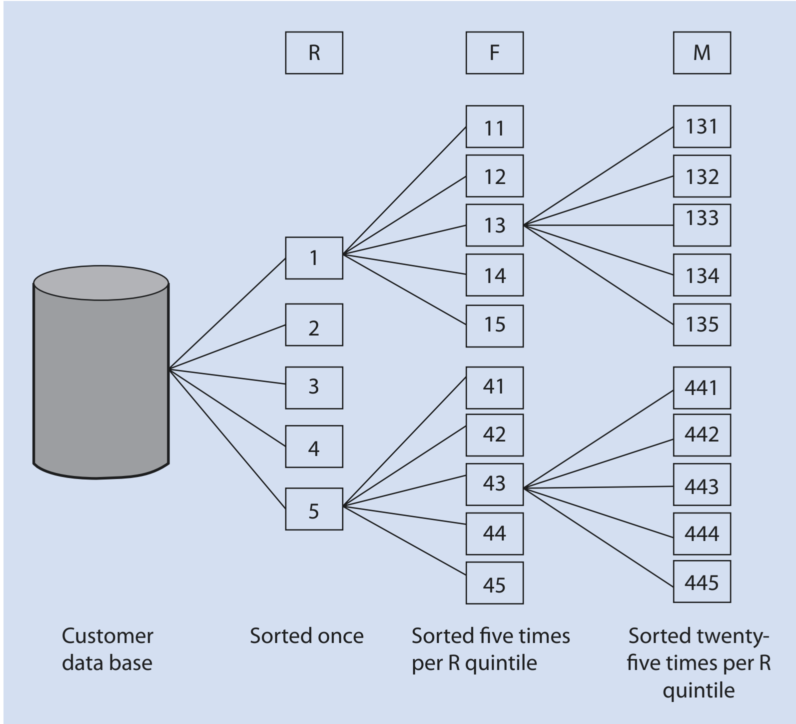


Fig. 6.5 RFM cell sorting

necessary to determine a cutoff point for the marketing campaign. This cutoff point should be based on the profitability of the customer. The breakeven value (BE) provides a simple to calculate metric for this purpose. In marketing literature *breakeven* refers to the fact that the net profit from a marketing promotion equals the cost associated with conducting the promotion. The BE is defined as

$$BE = \frac{unit\ cost\ price}{unit\ net\ profit} \tag{6.1}$$

If this ratio for a particular promotion is 1, then the promotion only broke even and did not generate any net profits. This ratio also computes the minimum response rates required in order to offset the promotional costs involved and thereby not incur any losses. Thus, we also refer to the BE as the *breakeven response rate*.

Example (Continued)

Consider the example of mailing \$150 discount coupons. Suppose the cost to mail each piece is a dollar, and the net profit (after all costs) is \$45, then the breakeven value or breakeven response rate required can be computed as BE = \$1/\$45 = 0.0222, or 2.22%. This value can be computed and then can be compared with the actual response rate of each RFM cell.

To simplify comparison, the breakeven response rate just computed could be used in computing a *breakeven index* (BEI) for every RFM cell. The BEI is calculated using the following formula.

$$BEI = \left(\frac{(Actual\ response\ rate - BE)}{BE} \right) \tag{6.2}$$

A positive BEI value indicates that some profit was made from the transaction. A BEI value of 0 indicates that the transaction just broke even, and a negative BEI value indicates that the transaction resulted in a loss.

Example (Continued)

Therefore, in the above example, if the actual response rate of a particular RFM cell was 3.5%, then BEI = ([3.5% − 2.22%]/2.22%) × 100 = 57.66.

Table 6.2 shows an excerpt from the BEI computations for 35 RFM cells. The complete table is available in Appendix III of this chapter. RFM cells with a corresponding positive BEI value are

Table 6.2 Combining RFM codes, breakeven codes, breakeven index

Cell #	RFM codes	Cost per mail ($)	Net profit per sale ($)	Breakeven (%)	Actual response (%)	Breakeven index
1	111	1	45.00	2.22	17.55	690
2	112	1	45.00	2.22	17.45	685
3	113	1	45.00	2.22	17.35	681
4	114	1	45.00	2.22	17.25	676
5	115	1	45.00	2.22	17.15	672
6	121	1	45.00	2.22	17.05	667
7	122	1	45.00	2.22	16.95	663
8	123	1	45.00	2.22	16.85	658
9	124	1	45.00	2.22	16.75	654
10	125	1	45.00	2.22	16.65	649
11	131	1	45.00	2.22	16.55	645
12	132	1	45.00	2.22	16.45	640
13	133	1	45.00	2.22	16.35	636
14	134	1	45.00	2.22	16.25	631
15	135	1	45.00	2.22	16.15	627
16	141	1	45.00	2.22	16.05	622
17	142	1	45.00	2.22	15.95	618
18	143	1	45.00	2.22	15.85	613
19	144	1	45.00	2.22	15.75	609
20	145	1	45.00	2.22	15.65	604
21	151	1	45.00	2.22	15.55	600
22	152	1	45.00	2.22	15.45	595
23	153	1	45.00	2.22	15.35	591
24	154	1	45.00	2.22	15.25	586
25	155	1	45.00	2.22	15.15	582
26	211	1	45.00	2.22	15.65	604
27	212	1	45.00	2.22	15.55	600
28	213	1	45.00	2.22	15.45	595
29	214	1	45.00	2.22	15.35	591
30	215	1	45.00	2.22	15.25	586
31	221	1	45.00	2.22	15.15	582
32	222	1	45.00	2.22	15.05	577
33	223	1	45.00	2.22	14.95	573
34	224	1	45.00	2.22	14.85	568
35	225	1	45.00	2.22	14.75	564

the groups of customers the marketing campaign should target, and all those RFM cells with corresponding negative BEI values are those customers to be avoided for this promotion.

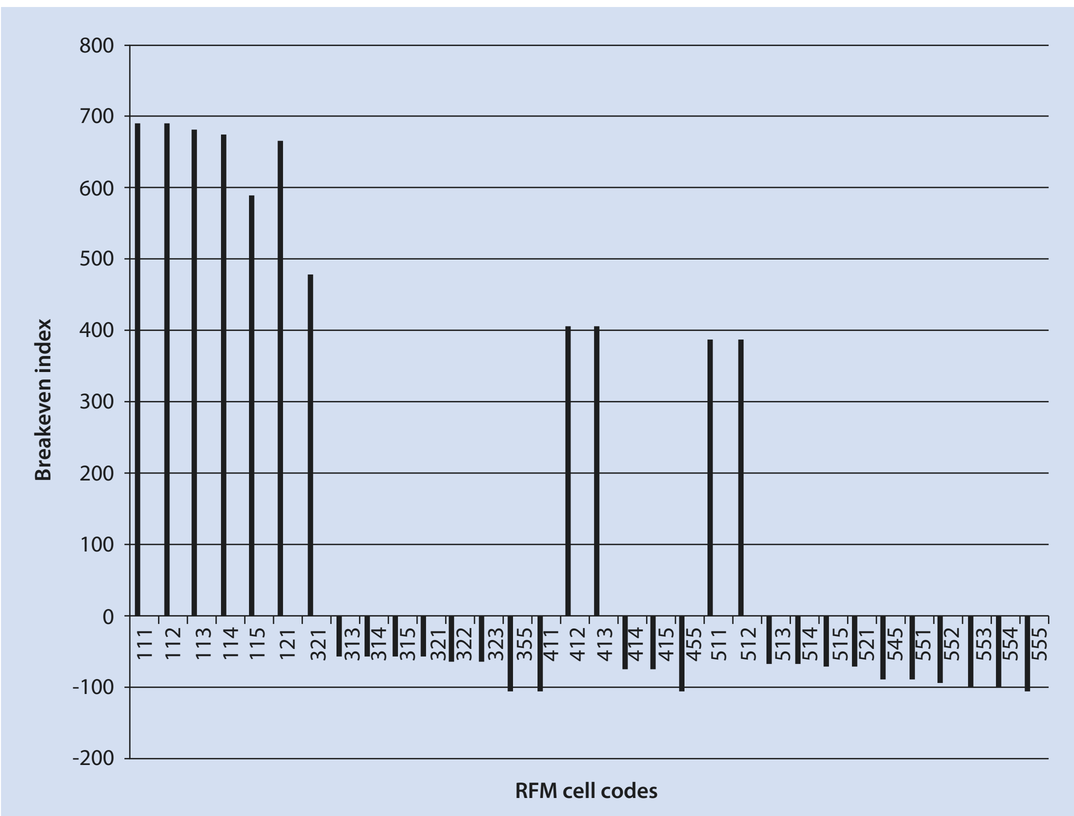
It is interesting to note that of the 125 cells, only customers within 56 cells have a higher chance of offering profitability to the firm, and the rest do not! Therefore, it becomes clear that a firm can achieve significant savings by only focusing on potentially profitable customers and not targeting the rest.

◘ Figure 6.6 plots the RFM cell codes and their corresponding BEI values. Customers with positive BEI values are to be chosen and the rest are to be left unconsidered. Note that customers with higher RFM values tend to have higher BEI values. However, at the same time, customers with a lower recency value but relatively higher F and M values also tend to have positive BEI values and hence should be considered for target mailing.

Order of Importance of R, F, and M

Most often businesses use the RFM technique in the order of recency, frequency, and monetary value. However, the order varies for different industry segments. Although the RFM order is normally acceptable, a more accurate order of coding would depend on the rapidity the customer response rate drops. The metric (R, F, or M) for which the customer response rates declines more quickly is likely to be the best predictor of future customer response and, hence, should be coded first. Once the metric of highest influence is determined, the same method of measurement can be used to determine the order of the remaining metrics.

Referring to ◘ Figs. 6.1, 6.2, and 6.3 it should be noticed that customer response rate drops more rapidly for the recency metric than the other two metrics. Similarly, the customer response rate for the frequency metric drops more rapidly than the monetary value metric. Therefore, the order of R, F, and M holds good in this case.



◘ **Fig. 6.6** RFM codes versus BEI

Table 6.3 Comparison of profits for targeting campaign test

	Test	Full customer base	RFM selection
Average response rate	2.02%	2.02%	15.25%
Number of responses	808	8808	2372.8
Average net profit/sale	$45	$45	$45
Net revenue	$36,360	$363,600	$122,976
Number of mailers sent	40,000	400,000	17,920
Cost per mailer	$1.00	$1.00	$1.00
Mailing cost	$40,000.00	$400,000.00	$17,920.00
Profits	(−$3640.00)	(−$36400.00)	$105,056.00

Example (Continued)

Table 6.3 compares the profits made by targeting all customers vs. using RFM to target selected customers. It is clear the firm will benefit significantly more by sending the mailers to select customers within selected RFM cells, than by sending it to their entire customer base. The loss of $3640 incurred in conducting the test is offset by the profits generated by sending mailers to the RFM select customers.

Relative Importance of R, F, and M

In the simplest case the RFM values are assigned for each customer by sequential sorting based on the RFM metrics. However, there is an alternative method which uses regression techniques to compute the relative weights of the R, F, and M metrics, and these relative weights are used to compute the cumulative points of each customer. The pre-computed weights for R, F, and M, based on a test sample, are used to assign RFM scores to each customer (see Appendix II). The higher the computed score, the more profitable the customer is likely to be in the future. This method, unlike the earlier one, is more flexible and can be tailored to each business situation.

Example (Aaker, Kumar, & Day, 2003)

Three customers have a purchase history calculated over a 12-month period (see Tables 6.4, 6.5, 6.6, and 6.7). For every customer, numerical

Table 6.4 Recency score

Customer	Purchase number	Recency (month)	Assigned points	Weighted points
John	1	2	20	100
	2	4	10	50
	3	9	3	15
Smith	1	6	5	25
Mary	1	2	20	100
	2	4	10	50
	3	6	5	25
	4	9	3	15

Points for recency: 20 points if within past 2 months, 10 points if within past 4 months, 5 points if within past 6 months, 3 points if within past 9 months, 1 point if within past 12 months, relative weight = 5

Table 6.5 Frequency score

Customer	Purchase number	Frequency	Assigned points	Weighted points
John	1	1	3	6
	2	1	3	6
	3	1	3	6
Smith	1	2	6	12
Mary	1	1	3	6
	2	1	3	6
	3	2	6	12
	4	1	3	6

Points for frequency: 3 points for each purchase within 12 months, maximum = 15 points, relative weight = 2

Table 6.6 Monetary value score

Customer	Purchase number	Value of purchase ($)	Assigned points	Weighted points
John	1	40	4	12
	2	120	12	36
	3	60	6	18
Smith	1	400	25	75
Mary	1	90	9	27
	2	70	7	21
	3	80	8	24
	4	40	4	12

Points for monetary value: 10% of the $-value of purchase within 12 months, maximum = 25 points, relative weight = 3

points have been assigned to each transaction according to a historically derived RFM formula. The relative weight based on the importance assigned to each of the three variables, R, F, and M on the basis of an analysis carried out on past customer transactions is as follows:

$$\text{Recency} = 5,\quad \text{Frequency} = 2,\quad \text{Monetary} = 3$$

The resulting cumulative scores, 249 for John, 112 for Smith, and 308 for Mags, indicate a potential preference for Mags. In this case, John seems to be a good prospect as well, but mailing to Smith might be a misdirected marketing effort. This example illustrates a simple application of the RFM technique. In practice, however, the number of customers to be analyzed can run into millions. Regression techniques are often employed to arrive at the relative weights for RFM. See Appendix II for an overview of a regression scoring model.

Table 6.7 RFM cumulative score

Customer	Purchase number	Total weighted points	Cumulative points
John	1	118	118
	2	92	210
	3	39	249
Smith	1	112	112
Mary	1	133	133
	2	77	210
	3	61	271
	4	37	308

Evaluation

The RFM technique helps organizations significantly, not only in identifying and targeting valuable customers who have a very high chance of purchasing, but also in avoiding costly communications and campaigns to customers who have a lower probability of purchasing. Instead it helps to identify only those customers with high probabilities of purchase and to target the companies' marketing strategies and communications accordingly. A limitation is that the RFM technique can be applied only on historical customer data available and not on prospects data.

6.1.2 Past Customer Value

Past customer value (PCV) is a metric which assumes the results of past transactions are an indicator of the customer's future contributions. The value of a customer is determined based on the total contribution (toward profits) provided by the customer in the past. This modeling technique assumes that the past performance of the customer indicates the future level of profitability. Since products or services are bought at different points in time during the customer's lifetime, all transactions must be adjusted for the time value of money.

$$PCV\ of\ customer\ i = \sum_{t=0}^{T} GC_{i(t_0-t)} * (1+\delta)^t \tag{6.3}$$

Where
i = customer
t = time index
δ = applicable discount rate (for example 1.25% per month)
t_0 = current time period
T = number of time periods prior to current period that should be considered
GC_{it} = gross contribution of transactions of customer i period t

Example

If we have data on the products purchased by a customer over a period of time, the value of the purchases, and the contribution margin, we can calculate the value generated by the customer by computing all transactions in terms of their present value. Assuming a contribution margin of 0.3, a monthly discount rate of r = 1.25%, and a spending pattern as illustrated in Table 6.8, the PCV is calculated as:

$$\begin{aligned} PCV_i &= 6(1+0.0125)^0 + 9(1+0.0125)^1 \\ &\quad + 15(1+0.0125)^2 + 15(1+0.0125)^3 \\ &\quad + 240(1+0.0125)^4 = 302.01486 \end{aligned}$$

This customer is worth $302.01 expressed in net present value in May dollars.

Evaluation

By comparing the PCV of a set of customers, we arrive at a prioritization for directing future mar-

Table 6.8 Spending pattern of a customer

	January	February	March	April	May
Purchase amount ($)	800	50	50	30	20
GC	240	15	15	9	6

Gross contribution (GC) = purchase amount × contribution margin

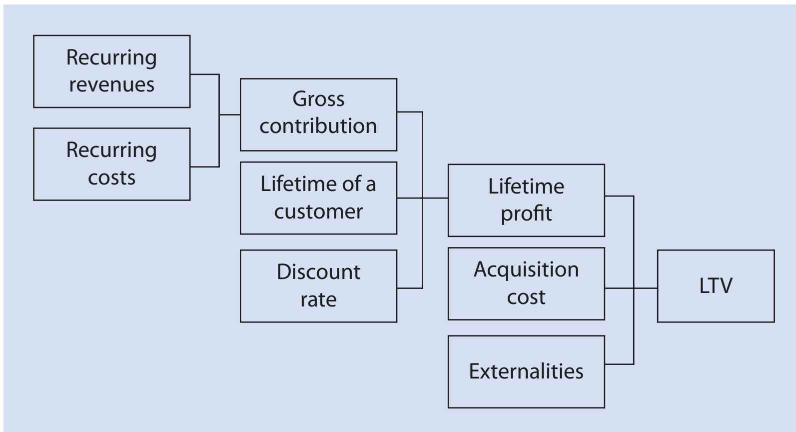


Fig. 6.7 Principles of LTV calculation

keting efforts. The underlying assumption is that the past contribution of a customer is a good predictor of her future contributions. The customers with higher values are normally those deserving greater marketing resources. This method, while extremely useful, does not incorporate other information which could help refine the process of selecting profitable customers. For instance, it does not consider whether a customer is going to be active in the future. It also does not incorporate the expected cost of maintaining the customer in the future. Concluding, it is still a backward-looking metric.

6.1.3 Lifetime Value Metrics

Evaluating the long-term economic value of a customer to the firm has seen a dramatic rise in interest. This is a direct outcome of the shift from transactional marketing to relational marketing. If a manager wants to evaluate marketing resource allocation plans targeted at improving the long-term value of customers, corresponding control measures must be put in place. Looking at profits on a per-transaction basis is not sufficient. Managers want to have an idea how the value of a client has evolved over time. The general term used to describe the long-term economic value of a customer is lifetime value (LTV), also referred to as the customer lifetime value (CLV). In very simple terms, it is a multi period evaluation of a customer's value to the firm in net present value. However, the term LTV is not without controversies. Although it is the Holy Grail for some, others call it «an elaborate fiction of presumed precision» (Jackson, 1992). In the following section we present some of the most common ways to calculate LTV. Nevertheless, the reader should be aware that there are many specific formulations.[2] It is important to present the principle in such a manner that readers can adapt the calculation to their own requirements. Conceptually, the principle of calculating LTV is represented in Fig. 6.7.

As one can easily see, there is not a single way to arrive at each component. For example, are the recurring costs comprised only of direct product

2 For details on various forms of LTV models, see: Jain and Singh (2002), Berger and Nasr (1998).

costs or also marketing, sales and service costs? Depending on many factors, such as nature of product, data availability and statistical capabilities, the inputs for the LTV calculation change. Is this problematic? It is not. First and foremost, it is important to understand the principle. Based on the general principle, one can then start to adjust the calculation to the available data. Also, one needs to adapt the formulation to the industry and company context. For example, having a defined finite lifetime duration (as for a contractual relationship such as cable subscription) makes for a different formulation as having a nonfinite relationship (as for a noncontractual relationship such as buying from a supermarket).

After discussing the basic formulations, we will highlight key issues that should be considered when employing the models. In the following discussion, we will present different formulations of the same principle.

Basic LTV Model

In the most simple definition, the lifetime value of an individual customer i is the sum of her discounted gross contribution over the respective observation horizon T.

$$LTV_i = \sum_{t=1}^{T} GC_{it} \left(\frac{1}{1+\delta} \right)^t \tag{6.4}$$

Where

- i = customer
- t = time period
- δ = interest (or discount) rate
- GC_{it} = gross contribution of customer i at time t
- T = observation time horizon
- LTV_i = lifetime value of an individual customer i in net present value at time t = 0

The resulting LTV is a measure of a single customer's worth to the firm. The gross contribution (GC) may vary, of course, across customers and across time. This formulation is primarily used for pedagogical and conceptual purposes. It is typically based on past customer behavior and may have limited diagnostic value for future decision making.

Cautionary note: If the time unit is different from a yearly basis, the interest rate δ needs to be adjusted accordingly. For example, if the yearly interest rate is 15%, the quarterly interest rate is 3.56%.

LTV with Splitted Revenues and Costs

On the next level, one can break down the gross contribution into its constituting elements.

$$LTV_i = \sum_{t=1}^{T} \left(\left(S_{it} - DC_{it} \right) - MC_{it} \right) \left(\frac{1}{1+\delta} \right)^t \tag{6.5}$$

Where

- i = customer
- t = time period
- T = observation time horizon
- δ = interest (or discount) rate
- S_{it} = sales value to customer i at time t
- DC_{it} = direct costs of products purchased by customer i at time t
- MC_{it} = marketing costs directed at customer i at time t
- LTV_i = lifetime value of an individual customer i in net present value at time t = 0

The cost element in this example is broken down into direct product-related costs and marketing costs. Depending on data availability, it can be enhanced by including service-related cost, delivery cost, or other relevant cost elements.

LTV Including Customer Retention Probabilities

So far, an assumption was that all customers under investigation remain fully active during the period of interest. However, in reality, more and more customers stop their relationship with the firm over time. The next step is therefore to consider customer retention probabilities. This relates to the fact that customers tend to remain in the relationship only with a certain probability approximated by the average retention rate Rr. Also, the AC is now subtracted from the customer's value.

$$LTV_i = \left(\left(\sum_{t=1}^{T} \left(\prod_{k=1}^{K} Rr_k \right) GC_{it} \left(\frac{1}{1+\delta} \right)^t \right) \right) - AC_i \tag{6.6}$$

Where

- i = customer
- t = time period
- T = time horizon under consideration
- δ = interest (or discount) rate
- Rr_t = average retention rate at time t (it is possible to use an individual level retention probability

Rr_{it} but usually this is difficult to obtain; see discussion in ► Sect. 5.4.2).

GC_{it} = gross contribution of customer i at time t.

AC_i = costs of acquiring customer i (acquisition costs).

Note that in this equation the term $\prod_{k=1}^{t} Rr_k$ is actually the survival rate SR_t (see ► Sect. 5.4.3).

It is also noteworthy, that if the retention rate is constant over time (i.e., it does not vary across time and thus Rr_k = Rr for all k) and thus the expression can be simplified using the identity:

$$\prod_{k=1}^{t} Rr_k = (Rr)^t \tag{6.7}$$

Although this is a common assumption, it is mostly not very realistic, as was previously discussed.

LTV with Constant Retention Rate and Gross Contribution

Under the assumption that.

$T \to \infty$, that the retention rate (Rr) is constant over time, and that the GC does not change over time, (6.6) can be simplified to the following formula:

$$LTV_i = GC_i\left(\frac{Rr}{1 - Rr + \delta}\right) - AC_i \tag{6.8}$$

We call the term

$$\frac{Rr}{1 - Rr + \delta}$$

the margin multiplier.

This formulation is easy applicable for quick calculations and, unless the retention rate is very high, produces results very close to the more precise formulation. One only needs to multiply the GC with the margin multiplier ($Rr/[1 - Rr + \delta]$) and subtract the acquisition cost.

How Long is Lifetime Duration?

The word *lifetime* must, in many circumstances, be taken with a grain of salt. Although the term makes little sense with one-off purchases (say, for example, a house), it also seems strange to talk about LTV of a grocery shopper. Clearly, there is an actual lifetime value of a grocery shopper. However, given the long time span, this actual value is not practical. For all practical purposes, the lifetime duration is a longer-term duration used managerially (see ► Sect. 5.4.4). For example, in a direct marketing general merchandise context, managers do not look beyond a 4-year time span. Beyond that, any calculation and prediction may become difficult due to so many uncontrollable factors (e.g., the customer moves, new competitors enter the market, and so on). It is therefore important to make an educated judgment regarding a sensible duration horizon in the context of making decisions.

Incorporating Externalities in the LTV

The value a customer provides to a firm does not only consist of the revenue stream that results from her purchases of goods and services. In the era of modern telecommunication technology and the rapid growth of online social communities (e.g., Facebook, Twitter), product rating websites, and weblogs, the passing on of personal opinions about a product or brand can contribute substantially to the lifetime value of a customer. Examples are customer referrals that result in new customer acquisitions or the posting of negative product reviews that detract potential customers from buying a product. We subsume all these activities under the term word-of-mouth (WOM).

Measuring and Incorporating WOM

A first step at measuring WOM is to look at its effects on revenues and expenses. It can be expected that WOM has a direct effect on new customer acquisitions through reducing (or in case of negative WOM increasing) the AC. To incorporate the value of WOM in the LTV calculation we need to determine the savings in AC per customer due to a referral (AC savings: ACS) and the number of new customer acquisitions (n_i) that arise due to the referrals of customer i.

$$LTV_i = \left(\left(\sum_{t=1}^{T}\left(\prod_{k=1}^{T} Rr_k\right)\left(GC_{it} + n_{it} ACS_t\right)\left(\frac{1}{1+\delta}\right)^t\right)\right) - AC_i \tag{6.9}$$

Where

i = customer

t = time period

T = time horizon under consideration

δ = interest (or discount) rate

Rr_t = average retention rate at time t

GC_{it} = gross contribution of customer i at time t

n_{it} = number of new acquisitions at time t due to referrals of customer i

ACS_t = average acquisition cost savings per customer gained through referral of customer i at time t

AC_i = costs of acquiring customer i (acquisition costs)

Equation 6.9 accounts for the fact that there are opinion leaders who have a stronger influence on their peers and the differences in the size of the social network of customers through an individual level n_{it}. The downside is that this number is difficult to obtain. Surveys and questionnaires may be used but are costly, time consuming, and often not reliable. Instead network centrality metrics such as degree centrality or betweenness can help to approximate the number of referrals.[3] Still, it neglects the differences in GC of the customers that were acquired through referrals.

Alternative Ways to Account for Externalities

The situation becomes even more complex, when acknowledging that WOM not only reduces AC but also impacts the purchase frequency, purchase volume, and cross-buying and thus the LTV of the customer that was acquired because of a referral. A major problem is to whom to account the additional value of a customer j if his GC rises due to WOM by customer i. To avoid this question, the value of a customer's referrals can be separated from the LTV, for example by calculating a separate customer referral value (CRV) for each customer. A joined evaluation of both metrics helps the management to select and determine how to develop its customers (see Table 6.9). A separate calculation of a CRV from the LTV also allows acknowledging for the fact that the customers you acquire due to referrals have different GC.[4]

How to Spur Positive Referrals

WOM agent campaigns, viral marketing, opinion leader programs, or referral reward programs are activities that companies can initiate to encourage its customers to make more referrals. But companies need to bear in mind that investments in these initiatives can only pay off if customers are satisfied with the product or service offering encountered.

Table 6.9 Customer value matrix

		Average CRV after 1 year	
		Low	High
Average LTVafter 1 year	High	Affluents	Champions
	Low	Misers	Advocates

Source: This table is adapted from Kumar, Petersen, and Leone (2007)

Where Does the Information Come from?

The information on GC, sales, direct cost, and marketing cost comes from internal company records. The key issue is that this information must be known on a per-customer basis. This knowledge is not necessarily common among many firms. An increasing number of firms are installing activity-based costing (ABC) schemes. ABC methods are used to arrive at appropriate allocations of customer and process-specific costs. The observation horizon (duration of customer relationship) T are derived either from managerial judgment or come from actual purchase data (see ► Sect. 5.4.6). The retention rates can be calculated from internal records and customer tracking (see ► Sect. 5.4.2). The interest rate is a function of a firm's cost of capital and can be obtained from the financial accounting department. Information on externalities such as number of referrals or WOM can be derived from social network analyses or customer surveys.

Evaluation

LTV (or CLV) is a forward looking metric that allows for long-term decision making. It is a flexible measure that has to be adapted to the specific business context of an industry. Each individual customer is evaluated on his expected contributions to the company. Yet, one has to bear in mind that all forecasts are subject to uncertainty. Being aware of the assumptions (especially for the prediction of the retention rate and GC) is important for the correct interpretation of LTV and in

---

3 For a discussion of degree centrality and betweenness see for example Lee, Catte, and Noseworthy (2010) or Kiss and Bichler (2008).

4 Further illustrations and a practical application of the concept of CRV can be found in Kumar, Peterson, and Leone (2007).

implementing the right actions. LTV is useful in a variety of situations as for example, to evaluate the effect of a loyalty program (or an investment in customer satisfaction) on the bottom line. Other applications comprise price discrimination of customers with low LTV, better treatment of high LTV customers (e.g., Lufthansa Gold Card Members) or the decision on how much to pay for a click-through from an internet banner advertisement for new customer acquisition.

6.1.4 Customer Equity

6

Building on the definition of LTV, we can aggregate the LTV measure across customers. The resulting quantity is the customer equity (CE). This metric is an indicator of how much the firm is worth at a particular point in time as a result of the firm's customer management efforts.

$$CE = \sum_{i=1}^{I} LTV_i \tag{6.10}$$

Where

i = customer

I = all customers of a firm (or a specified customer cohort or segment)

LTV_i = lifetime value of customer i

The CE is the sum of individual lifetime values of the customer base in net present value. In this case, the CE measure gives the economic value of an entire cohort or segment of customers. As CE is based on LTV this metric requires the allocation of revenues and costs on an individual customer level. One can relax this constraint by calculating the LTV of an average customer and take the sum of these average LTVs. This is further illustrated in the comprehensive example in ► Sect. 6.1.5.

Customer Equity Share (CES)

An alternative metric to MS that takes the lifetime value of customers into account is the customer equity share (CES). The CES for a brand j can be calculated using the following formula:

$$CES_j = \frac{CE_j}{\sum_{K=1}^{K} CE_k} \tag{6.11}$$

Where

j = focal brand

K = all brands a firm offers

CE_j = customer equity of brand j

Where Does the Information Come from?

- Basically the same information as for the LTV is required.

Evaluation

The CE represents the value of the customer base to a company. This metric therefore can be seen as a link to the shareholder value of a firm. Besides the core elements of LTV (retention rate and GC), an important influencing factor for the CE is the proportion of profitable to unprofitable customers. In order to increase the CE management efforts should focus on increasing the number of highly profitable customers while reducing the number of unprofitable ones.

The CES is a relative measure of the value of a brand within a firm.

6.1.5 Comprehensive Example

The following example illustrates some aspects of the previously introduced LTV and customer equity models (see ◘ Table 6.10). The observation horizon for this example is 5 years (column 1). A company targets a list of 10,000 purchased addresses with an acquisition campaign. The company acquires 1000 customers through the target mailing; thus, the acquisition rate is 10%. At the end of the first period only 400 of the 1000 customers remain. Once acquired, a customer generates on average $120 in sales (column 2). For simplicity's sake, this level of sales is assumed to be constant over the lifetime of the customers. The margin of the firm is 30% (column 3), resulting in a constant gross margin (column 4). Marketing and service cost while alive are constant as well (column 5). The retention rate in the first period is 40% and then increases over time, as the loyal customers remain. The resulting number of remaining customers in each period is shown in column 8. The profit per customer (column 9) is computed by subtracting the marketing and service cost from the gross margin. This per-period contribution is discounted to present value with a yearly rate of 15% (column 10). Finally, the yearly discounted profit is multiplied with the number of remaining customers in each year. Then these values are summed up to the total customer equity of this group or cohort of customers (column 11).

Table 6.10 Customer equity calculation example

Year from acquisition	Sales per customer	Manufacturer margin	Manufacturer gross contribution	Marketing and servicing costs	Actual retention rate	Survival rate	Expected number of active customers	Profit per customer per period to manufacturer	Discounted profit per customer per period to manufacturer	Total discounted profits per period to manufacturer
0	120	0.3	36	20	0.4	0.4	400	16	16	6400
1	120	0.3	36	20	0.63	0.25	250	16	14	3500
2	120	0.3	36	20	0.75	0.187	187	16	12	2244
3	120	0.3	36	20	0.82	0.153	153	16	11	1683
4	120	0.3	36	20	0.85	0.131	131	16	9	1179
Total customer equity										**15,006**

6.2 Popular Customer Selection Strategies

Customer selection strategies are applied when firms want to target individual customers or groups of customers. The reason for targeting these customers can be manifold, for example, for sending out a promotion or inviting them to a special event. Finding the right targets for marketing resource allocation is at the heart of any CRM strategy. Smart targeting allows firms to spend resources judiciously and allows customers to receive messages relevant to them. Inconsiderate targeting actions destroy value by over- or under spending from the firm's perspective and by providing undesirable messages (junk mail). One step in the successful implementation of CRM is the smart deployment of targeting methodologies to maximize the benefits to firm and customer.

In particular we will discuss the techniques of *profiling*, *binary classification trees*, and *logistic regression* which can all be applied to binary (i.e., 0/1) outcome variables. To check the performance of each of these models, typically, two thirds of the available data are used to calculate the model and the remaining third is used as a hold-out sample for validation. At the end of this chapter we introduce the *misclassification rate* and *lift analysis*, which are simple techniques to compare the performance of two or more alternative models.

This is an example where a company combines analytical skills, judicious judgment, knowledge about consumer behavior, and careful targeting of customers. It also shows that targeting always happens in a business context and is not a mechanic activity.

6.2.1 Profiling

An intuitive approach to customer selection is to assume that the most profitable customers share common characteristics (i.e., profitable customers are similar to one-another). Based on this assumption the company should try to target customers with similar profiles to the currently most profitable ones. Depending on the intended goal (e.g., customer acquisition or direct-mail promotion to existing customers) «most profitable» can have different meanings, e.g., the customers who are most likely to respond to a direct-mail promotion. In the latter case RFM (see ► Sect. 6.1.1) is widely used in practice for profiling. An alternative is to use available demographics instead of recency, frequency, or monetary value to sort and group the customer base. For the identification of

CRM AT WORK 6.1

The Kroger Co

Kroger, an American grocery chain, has a very successful loyalty program in place, the Kroger Plus Card. The majority of Kroger's shoppers make purchases using the loyalty card which makes Kroger's loyalty program rated one of the highest in the United States. Its loyalty program allows the company to better understand their customers' shopping behavior. Further, they individually tailor coupons to customers based on their purchase history.

Kroger utilizes Dunnhumby, a data analytics firm, to track customer spending habits. With majority of their transactions involving the Kroger Plus Card, Kroger uses this valuable tool to collect the data Dunnhumby analyzes. Every swipe a customer makes using the loyalty card provides more insight on that shopper and any shopper that has similar habits. Kroger then logs those purchases into a database, and creates customer profiles based on past customer purchases. After the data is sent to a database, the company is then able to analyze the purchases and engage in target marketing. Kroger's target marketing involves monitoring an individual customer's shopping behavior, analyzing the data, and following up on the findings by offering coupons that cater to the customer's shopping habits. Kroger does not offer coupons that are not inline with past customer purchasing habits because they found that coupon redemption rate is higher when promotions include items customers are interested in. Kroger's tactics to cater coupons based on the individual customer has resulted in increased customer spending. In 2013, Kroger reported that their targeted marketing generated $10 billion in revenue. In 2015, Kroger earned $1.7 billion in profit on sales of $108.5 billion. Kroger's loyalty card is one of the reasons why Kroger has been able to survive in today's competitive market.

Source: Peterson and Lutz (2015), Groenfeldt (2013), Coolidge (2016).

variables that best characterize profitable customers classification trees or regression models can be used. A disadvantage of profiling is that only customers are considered for targeting that are similar to the existing ones. Profitable customer segments that do not match the current customer base might be missed.

Example
Consider the case of a bank which wants to acquire new, profitable customers. Profiling consists of identifying profitable customers in the bank's current mass-market segment and then to target similar profiles in the prospect pool.

Since the objective is to acquire prospects likely to be high-value customers, the bank must rely on customer characteristics common to both current customers (the basis for establishing the critical profile) and the prospects (scored on the basis of their profile). The process is shown in the ◘ Fig. 6.8.

Let's say the response variable for current customers is the GC (field A). The company sorted customers by GC and chose to profile the top 20% of customers. Transaction information (field B) is not available for prospects. This is why the bank has to rely on information available for both existing customers and prospects. One type of information is geodemographic data, such as socioeconomic status of a region, average age, type of housing, and so on. These data are provided from direct marketing agencies that specialize in the collection of geodemographics. They can be purchased and then appended to individual records of existing customers. That is, depending on name or ZIP code, geodemographic data are added to existing customer records. The model attempts to predict GC as the dependent variable with geodemographic information as the independent variables. The rationale behind this process is to find the profile that best characterizes high-value clients, which is subsequently applied to prospects' information. Finally, prospects with a high expected GC are targeted for the acquisition campaign.

6.2.2 Binary Classification Trees

Using classification (or decision) trees is a methodology that can be used for finding the best predictors of a 0/1 dependent variable. For example, a company wants to know the differential demographic characteristics of loyalty program members versus nonmembers. Classification trees are especially useful when there is a large set of potential predictors for a model. In such a case, it may be difficult to determine which predictors are the most important or what the relationships between the predictors and the target (dependent) variable are. Classification tree algorithms can be used to iteratively search through the data to find out which predictor best separates the two categories of a binary (or more generally categorical) target variable.

◘ **Fig. 6.8** Using profiling for new customer acquisition

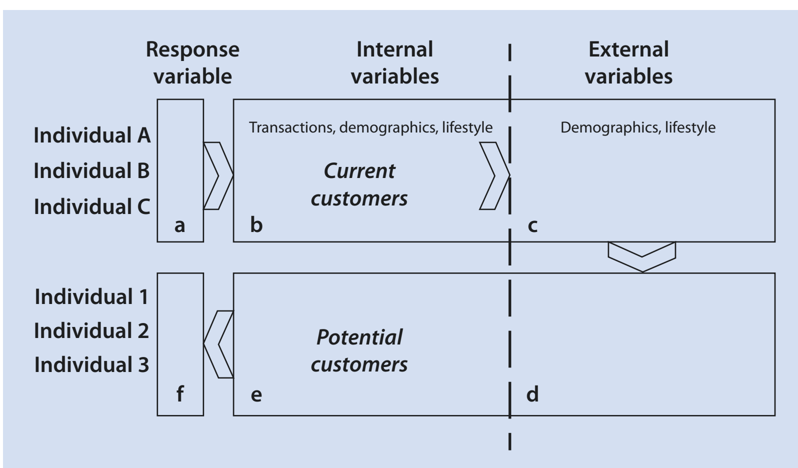


Table 6.11 Classification of potential hockey equipment buyers

	Male		Female		Total	
	Bought hockey	Did not buy hockey	Bought hockey	Did not buy hockey	Bought hockey	Did not buy hockey
Bought scuba	60	1140	50	1550	110	2690
Did not buy scuba	1540	2860	80	1320	1620	4180
Total	1600	4000	130	2870	1730	6870

Classification Tree Algorithm

Assume that Y is a binary outcome variable, i.e., $Y \in \{0,1\}$, explained by a set of explanatory variables $X_1,\ldots,X_p$ which are also binary.[5] The algorithm proceeds as follows: (1) To find out which of the explanatory variables X_i best explains the outcome Y, calculate the number of misclassifications (i.e., the number of not correctly predicted outcomes, for example if you want to predict credit card ownership and use gender as the predictor variable the misclassification rate is the number of male persons that do not own a credit card plus the number of female persons that do own a credit card) for each predictor variable X_i. (2) Use the variable X_i with the lowest misclassification rate to[6] separate the customer base. (3) This process can be repeated for each sub segment, until the misclassification rate drops below a tolerable threshold or all of the predictors have been applied to the model.

Example

Consider customer data for purchases of hockey equipment from a sporting goods catalog. For simplicity, there are only two predictor variables given, gender and whether a customer has previously bought scuba equipment (see Table 6.11). There are 8600 customers in total. 1730 bought the hockey equipment, 3000 are female and 5600 are male. It is also known that of the 8600 customers 2800 have bought scuba equipment in the past and 5800 have not bought scuba equipment.

Step 1

To determine the optimum approach how to separate the customers, we calculate the number of misclassifications for both predictor variables (gender and scuba equipment; see Fig. 6.9). Using gender as predictor of hockey sales we assume that all male persons buy hockey equipment, whereas no female buys. For the predictor variable gender we get a misclassification rate of $(4000 + 130)/8600 = 0.48$. When using the scuba equipment predictor variable (everyone who bought scuba is going to buy hockey equipment) we obtain $(2690 + 1620)/8600 = 0.50$. Thus, separating the customers based on gender is the optimal first step.

Step 2

We split the customer base by gender to obtain the classification tree as depicted in Fig. 6.10.

Step 3

If further predictor variables were available besides the scuba indicator, we could continue to identify the optimal predictor for separation in each subsegment and restart with step 1. For example, we could find that it is best to separate female customers by marital status and male customers by whether they have bought scuba equipment. The resulting tree could look like Fig. 6.11.

When this process is complete, a tree has developed in which segments are nested within segments. The profitable segments can then be identified for use as target markets.

5 If X_i is not binary one can find an optimal (in the sense that it best separates Y on the basis of classification of X_i) cutoff point to divide the domain of X_i in two parts and thus reduce X_i to a binary variable. For a further discussion see Hastie, Tibshirani, and Friedman (2009).

6 A discussion of further optimal splitting rules can be found in Blattberg, Kim, and Neslin (2008).

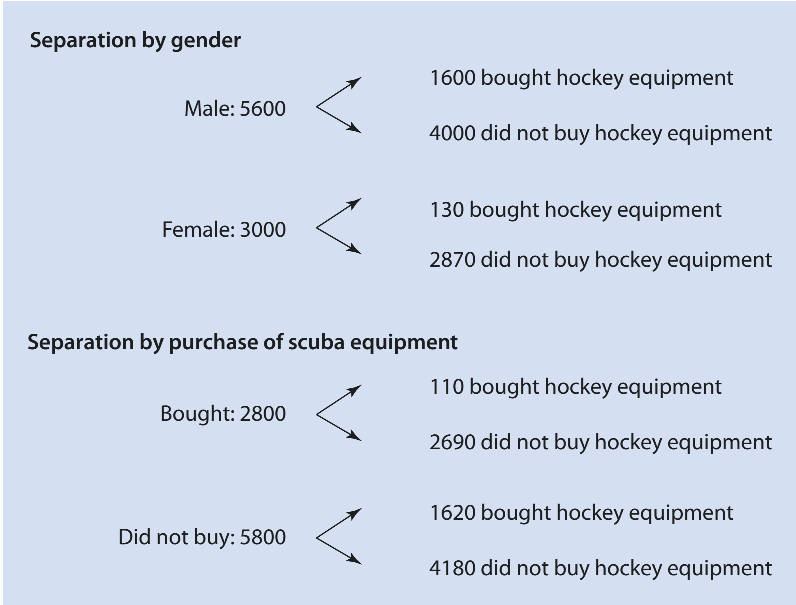


Fig. 6.9 Possible separations of potential hockey equipment buyers

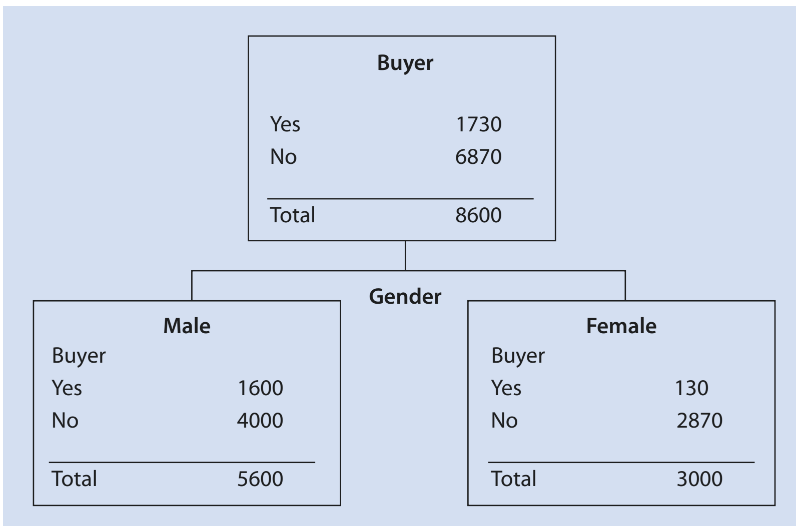


Fig. 6.10 Classification of hockey buyers by gender

Evaluation

One problem with the decision tree approach is that it is prone to overfitting, whereby segments are tailored to very small segments, (based upon the dataset that was used to create the tree) and as a result, the model developed will not perform nearly as well on a separate dataset. A hold out sample (typically 1/3 of the data set, not used for model calibration) can be used for model validation (see Sect. 6.3). If the results show a large discrepancy with what was expected, then the model will need to be reevaluated. If the results are within the range of what is predicted from the model, it is likely that the model is a good fit.

6.2.3 Logistic Regression

Linear regression starts with the specification of the dependent variable and the independent variable. For example the number of people entering a store on a Saturday is the dependent variable of interest and the amount of money the store spent on advertising on Friday is the independent variable. We

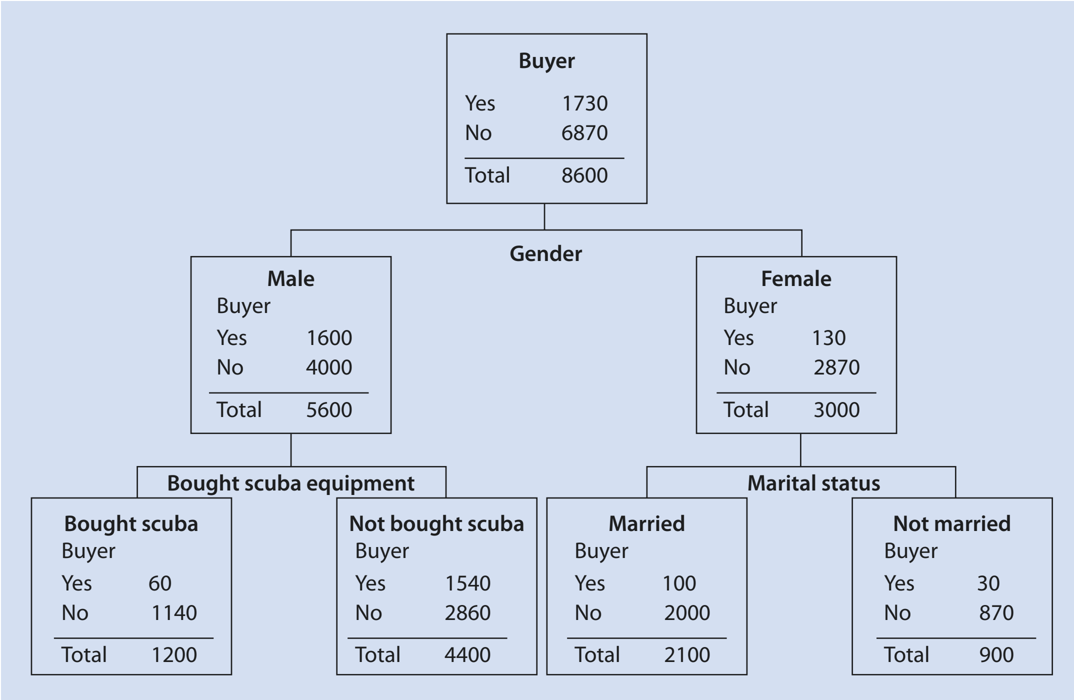


Fig. 6.11 Classification tree for hockey equipment buyers

expect to see a linear relationship between the two variables. A regression analysis creates an estimate of the coefficient that represents the effect of advertising on store traffic. In this situation we note that store traffic (the dependent variable) can take on a large range of values. However, in marketing we often encounter situations where the dependent variable is binary. For example, in a situation where we are interested in whether a customer bought a product or not, we assign the value 0 to this variable when the customer does not buy, and the value 1 when the customer buys. Regression models that allow for such a data structure are linear probability models, probit models, and logit models. Logit models also referred to as *logistic regression* are the most frequently applied method when the dependent variable is binary and assumes only two discrete values. For example:

- Whether a customer responded to a marketing campaign or not.
- Whether a person bought a car or not.

These observed values for the dependent variable take on only two values and are usually represented using a 0–1 dummy variable. The mean of a 0–1 dummy variable is equal to the proportion of observations with a value of 1, and can be interpreted as a probability. Furthermore, the predicted values in a logistic regression fall between 0 and 1 and are also interpreted as probabilities. For example, home ownership as a function of income can be modeled whereby ownership is delineated by a 1 and non ownership by 0. The predicted value based on the model is interpreted as the probability that the individual is a home-owner. With a positive correlation between increasing income and increasing probability of ownership, we can expect to see results where the predicted probability of ownership is, for example, 0.22 for a person with an income of $35,000, and 0.95 for a person with a $250,000 income.

Example

Consider the upgrade of a credit card. Logistic regression can be used to identify potential targets for marketing credit card offers to existing customers of a bank.

- Dependent variable – whether or not the customer signed up for a gold card offer.
- Predictor variables – other bank services the customer used plus financial and demographic customer information.

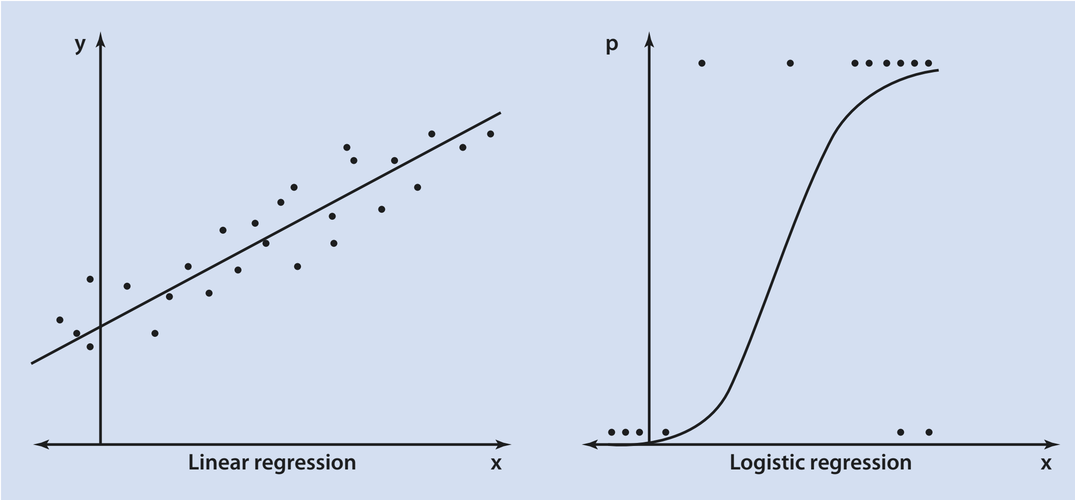


Fig. 6.12 Comparison of linear and logistic regression

The goal is to estimate the logistic regression on a sample of customers that were offered a gold card. Inputting values of the predictor variables for each potential target customer, the logistic model will yield a predicted probability for the target customer to sign up for the gold card offer. Customers with high predicted probabilities may be chosen to receive the offer because they seem more likely to respond positively.

Dots represent observations of the dependent variable y. In case of linear regression y can take any value whereas in case of logistic regression y is either 0 or 1. The logistic regression curve reflects the predicted probability p of the event y at given levels of x.

Mathematically, linear regression takes the form

$$y = \alpha + \beta x + \varepsilon$$

Where

y = dependent variable

x = predictor variable

α = constant (which is estimated by linear regression and often called intercept)

β = the effect of x on y (also estimated by linear regression)

ε = error term.

Transformation Process

In this sort of regression, y can take on any value between negative infinity and positive infinity. However, as we noted earlier, in many instances we observe only binary activity represented by 0 or 1. Sample plots of observations with fitted linear and logistic regression curves are illustrated in Fig. 6.12. Hence, if the actually observed dependent variable has to be constrained between 0 and 1 to indicate the probability of an event occurring, a transformation is necessary. This transformation is the basis of logistic regression. The steps of the transformation are given as follows:

Step 1

If p represents the probability of an event occurring, consider the ratio $\frac{p}{1-p}$. Since p is a positive quantity less than 1, the range of this expression is 0 to infinity.

Step 2

Take the logarithm of this ratio

$$\log\left(\frac{p}{1-p}\right)$$

This transformation allows the range of values for this expression to lie between negative infinity and positive infinity.

Step 3

The value

$$z = \log\left(\frac{p}{1-p}\right)$$

can now be considered as the dependent variable. A linear relationship of this value with predictor variables in the form z = α + βx + ε can be written out. The α and β coefficients can be estimated by maximum likelihood implemented in standard software packages.

Step 4

In order to obtain the predicted probability p, the following back transformation is necessary:

Since

$$\log\left(\frac{p}{1-p}\right) = z = \alpha + \beta x + \varepsilon$$

we can write

$$\left(\frac{p}{1-p}\right) = e^{z}$$

This allows us to calculate the probability p of an event occurring, the variable of interest, as

$$p = \left(\frac{e^{z}}{1+e^{z}}\right) = \left(\frac{1}{1+e^{-z}}\right) \tag{6.12}$$

Interpretation of Coefficients

Care has to be taken when interpreting the coefficients. The interpretation of coefficients differs from simple linear regression. If the logit regression coefficient β = 2.303 then the log odds ratio is

$$e^{\beta} = e^{2.303} = 10,$$

which says that when the independent variable x increases one unit, the odds that the dependent variable equals 1 increases by a factor of 10, keeping all other variables constant. Usually, besides the coefficients, «odds» are reported, i.e.,

$$\left(\frac{p}{1-p}\right) = e^{z}$$

is called the «odds» and represents the chance of an event occurring to the chance of the event not occurring, where p is the probability of the event.

Table 6.12 Sample data for contract extension and sales

Contract extension	Number of sales calls
1	2
1	4
0	6
0	2
1	4
1	8
0	3
0	0
0	2
0	5
0	0
1	2
1	8
1	4

Example

Consider the extension of a service contract in a B2B setting. The service company wants to know the probability that a client extends his existing contract. Therefore, it tracks the number of sales calls and whether a client extends his contract (see Table 6.12).

It expects that the sales calls impact the probability of contract extension. Thus, it estimates the model:

$$Prob(contract\ extension) = \frac{1}{1+e^{-\alpha-\beta(sales\ calls)}} \tag{6.13}$$

The resulting estimates are α = −1.37 and β = 0.39. The odds and the probability of extension are illustrated in Table 6.13. In this example the odds represent the relative likelihood that a customer will extend his contract to not extending his contract. If no sales calls are made the odds are 0.253 and in case of a sales call the odds are 0.375. The log-odds are $e^{0.39}$ = 1.477 meaning that for

Table 6.13 Odds of logistic regression example

	Sales calls = 0	Sales calls = 1
Odds (exp(α + β × sales calls))	0.253	0.375
Probability of contract extension	0.202	0.273
Difference in probabilities	0.071	

each additional sales call the odds for contract extension change by a factor of 1.477. The probabilities of contract extension are 0.202 and 0.273 for no and for one sales call, respectively. Thus, the difference in probabilities for a contract extension of making one sales call over not making a call is 0.07. Note that the difference of an additional sales call is not constant. For example making four instead of three sales calls changes the probability for a contract extension by 0.098.

Evaluation

Unlike in linear regression where the effect of one unit change in the independent variable on the dependent variable is assumed to be a constant represented by the slope of a straight line, for logistic regression the effect of a one-unit increase in the predictor variable varies along an s-shaped curve (see Fig. 6.12). This means that at the extremes, a one-unit change has very little effect, but in the center a one-unit change has a fairly large effect. In the case of the income versus home ownership example, the difference in the likelihood that an individual owns a home may not change much as income increases from \$10,000 to \$30,000 or from \$1000,000 to \$1,020,000, but may increase considerably if income increases from \$50,000 to \$70,000.

6.3 Techniques to Evaluate Alternative Customer Selection Strategies

A critical step to decide which model to use to select and target potential customers is to evaluate alternative selection strategies. When several alternative models are available for customer selection, we need to compare their relative quality of prediction. To compare the predictive capabilities of models, we typically divide the data into a training (calibration) dataset (2/3 of the data) and a holdout (test) sample (1/3 of the data). The models are estimated based on the training dataset. Using these estimates predictions are made for the holdout data. We can then compare the predictive performance of the models. Some models will have more predictive power than others, and we will select the model that generalizes best from the training to the test data.

The most common way to assess a model's performance is by comparing their respective misclassification rates obtained on the test set. An alternative method for getting an idea about model performance is lift analysis.

6.3.1 Misclassification Rate

To obtain the misclassification rate, we estimate the different models on two thirds of the available data and calculate the misclassification rate on the remaining third to check the model performance. The misclassification rate is the number of false predictions divided by the total number of predictions made.

A misclassification error rate or confusion matrix could look like Table 6.14. The number of false predictions is the sum of the off-diagonal entries (56 + 173 = 229). Thus, the misclassification rate in this example is 229/1459 = 15.7%.

6.3.2 LIFT Analysis

A lift chart shows how much better the current model performs against the results expected if

Table 6.14 Confusion matrix

		Predicted		
		1	0	Totals
Observed	1	726	56	782
	0	173	504	677
Totals		899	560	1459

no model was used (base model). This gives a baseline measure of how good the model is. As an example, consider 1000 prospects, of which 100 have purchased. A good predictive model helps increase the relative amount of purchasers in the selected group: a random-based selection of 100 prospects would contain about 10 purchasers, whereas a model-based selection of 100 prospects could result in 30 purchasers. This is what lift charts help to visualize. The models with the highest lift are candidates for final selection.

Lifts can be used to compare two or more alternative models, track a model's performance over time, or to compare a model's performance on different samples. To calculate lifts for different deciles of data, we need the following information about a model with regard to the set of customer data.

- **Cumulative number of customers:** The number of total customers up to and including that decile.
- **Cumulative percentage of customers:** The percent of total customers up to and including that decile.
- **Cumulative number of buyers:** The number of buyers up to and including that decile.
- **Actual response rate for each decile:** Computed by dividing the number of buyers by the number of customers for each decile.
- **OR predicted response rate based on the model for each decile:** Computed by dividing the predicted number of buyers by the number of customers for each decile.

With this information we can calculate:

$$\boldsymbol{Lift\%} = (\text{Response rate for each decile}) \div (\text{Overall response rate})^{*}\,100$$

and

$$\boldsymbol{Cumulative\ lift\%} = (\text{Cumulative response rate}) \div (\text{Overall response rate})^{*}\,100$$

Where

$$\boldsymbol{Cumulative\ response\ rate} = \text{Cumulative number of buyers} \div \text{Number of customers per decile.}$$

A comprehensive example is given in Table 6.15.

Table 6.15 Lift and cumulative lift

Number of decile	Number of customers	Number of buyers	Response rate (%)	Lift	Cumulative lift
1	5000	1759	35.18	3.09	3.09
2	5000	1126	22.52	1.98	5.07
3	5000	998	19.96	1.75	6.82
4	5000	554	11.08	0.97	7.80
5	5000	449	8.98	0.79	8.59
6	5000	337	6.74	0.59	9.18
7	5000	221	4.42	0.39	9.57
8	5000	113	2.26	0.20	9.76
9	5000	89	1.78	0.16	9.92
10	5000	45	0.90	0.08	10.00
Total	50,000	5691	11.38		

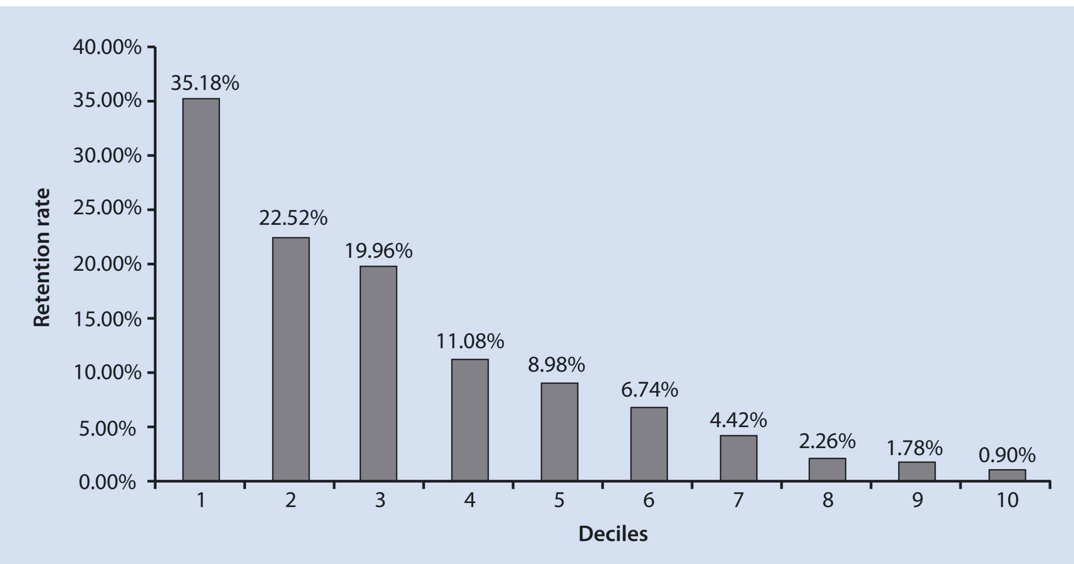
Lift Performance Illustration

Initially, the model (such as RFM, or logistic regression) which one wants to evaluate needs to be run and the customer base should be sorted accordingly. On the basis of the sorted customer list, customers are distributed into 10 equal-sized groups (see ◘ Table 6.15). In a model that performs well, customers in the first decile exhibit the highest response rate and the response rate continuously drops as we proceed along the deciles. (see ◘ Fig. 6.13).

For the top decile in this case, the lift is 3.09 (see ◘ Fig. 6.14). This indicates that by targeting only these customers we would expect to yield 3.09 times the number of buyers found by randomly mailing the same number of customers. In contrast, the last decile (decile 10) attracts only 0.08 times the number of buyers as one would expect in a random sample of the same size. Lifts that exceed 1 indicate better than average performance of that particular decile, and those that are less than 1 indicate a poorer than average performance. Also keep in mind that lift is a relative index to a baseline measure, in this case the average response rate for the entire sample.

The cumulative lifts for the model in ◘ Fig. 6.15 reveal the proportion of responders we can expect to gain from targeting a specific percent of customers using the model. If we choose the customers from the top three deciles (i.e., the top 30% of the customers), we will obtain 68% of the total responders. The larger the distance between the model and no model lines, the stronger or more powerful is the model. The slope of the cumulative lift curve reflects the lift. In a nutshell, lifts can be used to compare two or more alternative models, track a model's performance over time, or to compare a model's performance on different samples.

Past experience shows that logistic models tend to provide the best lift performance when used as a tool for customer selection. This is seen from the topmost curve in the lift chart in ◘ Fig. 6.16. The model is better able to identify the best customers and group them in the first few deciles. The past customer value approach provides the next best performance, whereas the traditional RFM approach, though by no means redundant, exhibits the poorest performance.



◘ **Fig. 6.13** Decile analysis

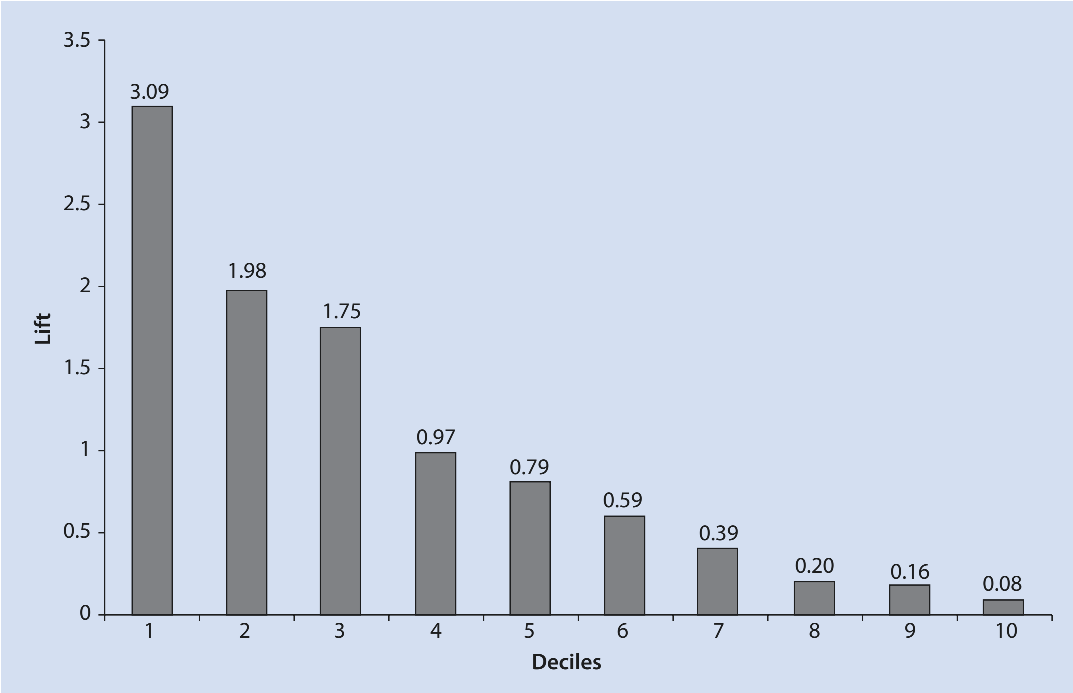


Fig. 6.14 Lift analysis

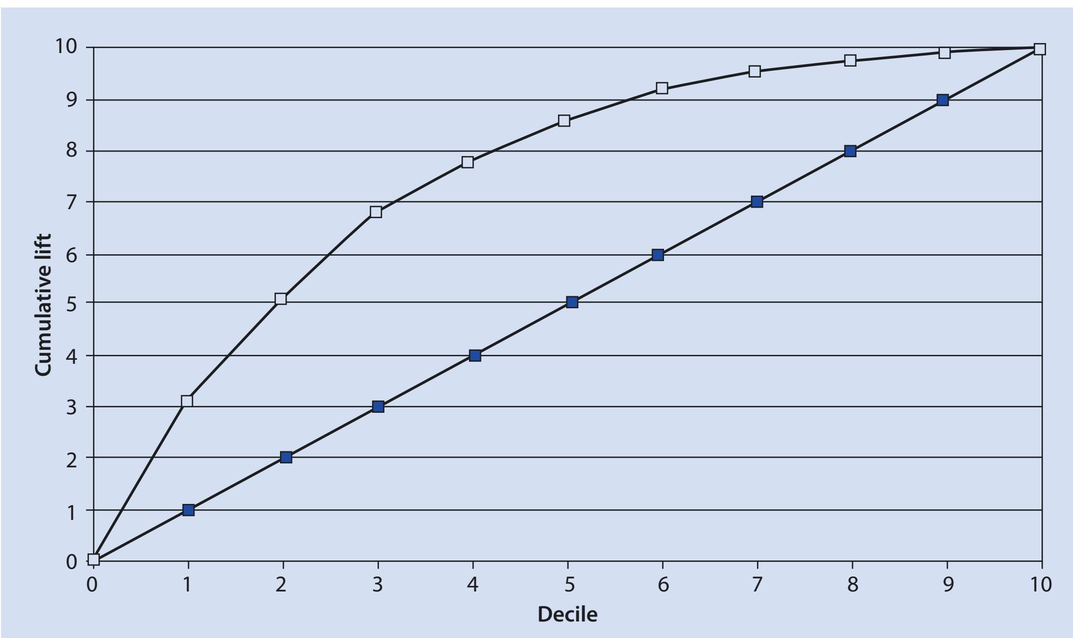


Fig. 6.15 Cumulative lift analysis

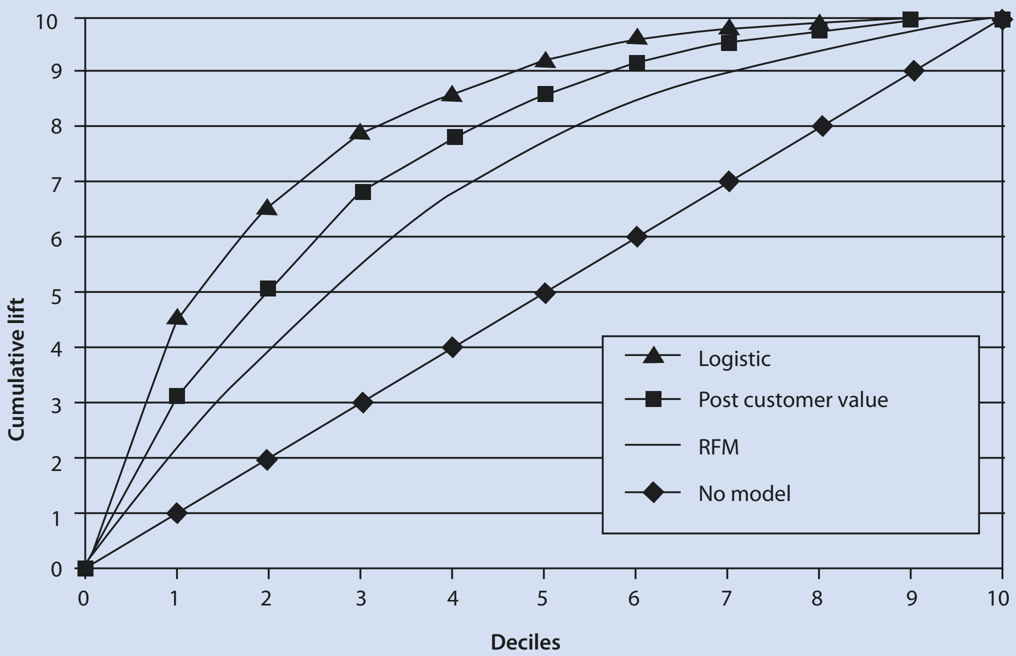


Fig. 6.16 Model comparison using lift analysis

Minicase 6.1

Differentiating Customer Service According to Customer Value at United Parcel Service (UPS)
UPS, headquartered in Atlanta, Georgia, is one of the world's largest package delivery companies. UPS—sometimes referred to as the «Big Brown»—has grown astronomically compared to its modest roots as a bicycle courier service founded in 1907. Today, UPS is a global leader in logistics. Over the years, UPS has acquired more than 40 companies which has enabled them to penetrate into sectors such as retail shipping, air freight, and business services. Now serving over 220 countries worldwide, the acquisitions have also given UPS the opportunity to develop strong relationships with many customers across the world.

In an effort to achieve their expansion goals, UPS decided to create service levels for managing temperature-sensitive healthcare products. The service levels were an addition to an already existing Temperature True portfolio. The change allowed healthcare companies to choose from tiered service levels based on the degree of temperature control needed. The service levels were as follows:

UPS Temperature True Service Portfolio:

- *UPS Temperature True Plus:* The highest level of service that is ideal for air freight shipments that require very strict temperature ranges in transit. UPS's existing service includes processes for managing shipments with the highest degree of risk mitigation. In addition, this service level includes monitoring throughout the entire shipping process.
- *UPS Temperature True Standard:* An offering ideal for shipments that need lower levels of in-transit monitoring, while still helping healthcare companies maintain product compliance and peace of mind.
- *UPS Temperature True Saver:* A standardized ocean freight solution for temperature-sensitive shipments. This option is ideal for large volume shipments where cost management and regulatory compliance are both paramount.

The expansions to the UPS Temperature True global portfolio offers a range of services that provide the healthcare industry with a more improved shipping solution for temperature-sensitive healthcare products. Services such as quality assurance support and 24/7 control tower monitoring can be customized based on the company's needs. In addition, the service options include packaging consulting services for temperature-sensitive products as well as best-in-class technology, proactive monitoring and intervention services, risk management and contingency planning as well as deep regulatory compliance expertise.

UPS decided to offer these service levels after receiving feedback from customers in the healthcare industry. The customers vocalized their healthcare product shipment needs, and UPS met those needs by improving services offered by an existing solution and allowing customers to customize based on their needs.

Source: UPS Pressroom (2013, 2017); Berman (2013).

Questions on Minicase 6.1

1. How can an organization compute client-level profitability?
2. What types of systems and processes are needed to document client profitability in a systematic and ongoing fashion?
3. For the customers that were accustomed to the previous shipping system, how can UPS urge them to adopt the new service tiers which may mean an increase in shipping costs depending on the customers' needs?
4. Describe a real world scenario that is similar to this case. If you were a manager, how would you address aligning your business offerings with customer needs without harming company profits?

6

Summary

Since customer value management involves allocating resources differently for individual customers based on their economic value, understanding value contribution from each of the customers to the firm is very important. The availability of customer-level data helps firms utilize a new set of metrics beyond traditional and popular marketing metrics introduced in ► Chap. 5. These strategic customer-based metrics enable the assignment of value to each individual customer.

Popular strategic customer-based value metrics are RFM, past customer value, lifetime value, and customer equity. RFM is a composite score of recency, frequency, and monetary value. Two methods used for computing and applying RFM are sorting techniques and calculation of relative weights for R, F, and M. In the sorting technique the RFM values are assigned for each customer by sequential sorting based on the RFM metrics and then used to group customers, belonging to specific RFM codes. However, this method might not produce an equal number of customers under each RFM cell. A hierarchical sorting technique ensures equal numbers in each RFM cell. This is done by first sorting the list of customers by recency and dividing them into equal groups. Each of these groups is then sorted based on frequency and divided into groups of equal size. Each of these subgroups is sorted on monetary value and divided into groups of equal size. This way, each RFM cell will have an equal number of customers. An alternate method for RFM metric uses linear regression to arrive at the relative weights of the R, F, and M metrics, and these relative weights are used to compute the RFM score of each customer. The higher the computed RFM score, the more profitable the customer is likely to be in the future. This method has the advantage of being flexible, so that it can be tailored to each business situation. In order to decide which RFM cells to target the break-even value (BE) and break even index (BEI) can be calculated for each RFM cell. With respect to a promotion, breakeven refers to the fact that the net profit from a marketing promotion equals the cost associated with conducting the promotion. The BE can be used in computing a BEI, calculated as Breakeven index (BEI) = ([Actual response rate − BE]/BE) × 100. A positive BEI indicates some profit was made from the transaction. A BEI value of 0 indicates the transaction only broke even, and a negative BEI value indicates the transaction resulted in a loss.

Another important customer-based metric is past customer value (PCV), in which the value of a customer is determined based on the total contribution (toward profits) pro-

vided by the customer in the past after adjusting for the time value of money. The lifetime value (LTV), in contrast, reflects the long-term economic value of a customer and is calculated as the sum of the discounted expected contribution margins over the respective observation horizon. The contribution margin can be computed knowing the values of sales, direct costs, and marketing costs. The sum of the lifetime value of all the customers of a firm represents the customer equity (CE) of a firm. It is an indicator of how much the firm is worth at a particular point in time as a result of the firm's customer management efforts. This metric therefore can be seen as a link to the shareholder value of a firm.

Firms employ different customer selection strategies to target the right customers. Some of the popular customer selection strategies are profiling, binary classification trees, and logistic regression. Profiling rests on the assumption that profitable customers share common characteristics. Profiling consists of identifying profitable customers in the existing customer base and then to target similar customers from the prospect pool. The construction of binary classification trees relies on a recursive partitioning algorithm used for finding the predictors which best separate the two categories of a binary (0/1) dependent variable. To avoid over fitting, this search is performed on two thirds of the available data with one third of the data reserved for testing the eventual model that will be developed. Logistic regression is a statistical tool to predict the probability of a binary outcome using predictor variables.

To evaluate alternative selection strategies, firms can use techniques such as misclassification rates and lift analysis. The misclassification rate is the number of false predictions divided by the total number of predictions made. Lifts indicate how much better a model performs than no model or average performance, and are calculated as (Response rate for each decile) ÷ (Overall response rate) × 100. Based on lifts we can perform a decile analysis. Therefore, customers are sorted based on the model to be evaluated and grouped into deciles. For a good model, customers in the first deciles have a significantly higher response rate. Cumulative lift is calculated as (Cumulative response rate) ÷ (Overall response rate) * 100. It reveals the proportion of responders we can expect to gain from targeting a specific percent of customers using the model.

International Perspectives: Did You Know?

1. Companies are becoming more customer focused in today's world because they have discovered how important customer oriented decisions can be when trying to survive in today's competitive market. In China, SAS, a business analytics provider, teamed up with Radica, a marketing firm headquartered in Hong Kong, to develop a CRM analytics software. This software enables companies to pull data from social media platforms, and make decisions based on the company's findings. Essentially, companies will now be able to mine data from social media sources, and use the findings to develop better customer strategies (Onag, 2015).
2. Can data analytics and science be combined for an even greater good? Laing O'Rourke, an engineering company headquartered in the United Kingdom, believed it was possible to merge the two and come up with very useful information. Recently, the company used Microsoft's Internet of Things technology to collect data on their field workers. The field workers were asked to wear sensors while on a job. The sensors were used to capture the worker's vitals and the surrounding environment conditions while the employee was working. The company's ultimate goal is to use the collected data to predict conditions that would lead to health issues before the issue happens. Since work conditions are being monitored, the company would be able to reduce employee absenteeism due to health issues related to job conditions (Merritt, 2015).

6

Exercise Questions

1. A hotel chain wants to analyze its customer base with RFM. Describe the data fields (variables) in the database necessary to do this.
2. Whatever RFM analysis can do, regression analysis can do as well. Evaluate this statement.
3. How will you use lift charts to determine future marketing action?
4. Describe three business situations where you would consider using logistic regression as the preferred technique for analysis and decision making.
5. What is the link between customer lifetime value and the profitability of an organization?

Appendix I Notation Key

Notation	Explanation
a	Coefficient of acquisition
AC	Acquisition costs
ACS	Acquisition costs savings
Ar	Acquisition rate
c	Category
CE	Customer equity
Dr	Defection rate
GC	Gross contribution
i	Individual customer
I	Total number of buyers with a focal firm
j	Firm
J	Total number of firms in a market
LTV	Lifetime value
MC	Marketing costs
n	Customer in cohort
N	Cohort size
r	Coefficient of retention
Rr	Retention rate
Rrc	Retention rate ceiling
S	Sales (value)
Sr	Survival rate
t	Time period
T	Length of time horizon
V	Sales (volume)
δ	Applicable discount rate

Appendix II Regression Scoring Models

Scoring models is the process of evaluating potential customer behavior on the basis of test results. Typically, a test is conducted in a limited market or in an experimental set up on a small subset of customers. This subset of customers is exposed to a marketing campaign and a product offering. The purpose of this test is to assign to each of the remaining customers a value which is extrapolated from the results of this test. These values typically reflect the prospective customer's likelihood of purchasing the test marketed product. The process of regression scoring can be represented in the following steps:

1. Draw a random sample from the overall population of prospective customers.
2. Obtain data from the sample that profiles individual consumer characteristics. The R, F, and M scores are variables which profile behavioral characteristics of a customer and are typically used in this procedure, along with other relevant variables.
3. Initiate a marketing campaign directed at the random sample, and record the individuals who become customers.
4. With that information, develop a regression scoring model to obtain a series of weighted variables that either predict which prospects are more likely to become customers or the value of profits that each customer is likely to provide, based on their individual characteristics.
5. By applying these weights to individual characteristics of prospective customers, we can arrive at a value for each customer which indicates how likely it is that the customer will purchase a product, or how much profit the customer will generate, if exposed to the tested marketing campaign.

Appendix III BEI Computations for RFM Cells

Cell #	RFM codes	Cost per mail ($)	Net profit per sale ($)	Breakeven (%)	Actual response (%)	Breakeven index
1	111	1	45	2.22	17.55	690
2	112	1	45	2.22	17.45	685
3	113	1	45	2.22	17.35	681
4	114	1	45	2.22	17.25	676
5	115	1	45	2.22	17.15	572
6	121	1	45	2.22	17.05	667
…						
52	312	1	45	2.22	12.91	481
53	313	1	45	2.22	0.98	−56
54	314	1	45	2.22	0.94	−58
55	375	1	45	2.22	0.90	−60
56	321	1	45	2.22	0.136	−61
57	322	1	45	2.22	0.82	−63
58	323	1	45	2.22	0.78	−65
…						
75	355	1	45	2.22	−0.15	−107
76	411	1	45	2.22	11.25	−107
77	412	1	45	2.22	11.22	406
78	413	1	45	2.22	0.55	405
79	414	1	45	2.22	0.52	−75
80	415	1	45	2.22	0.49	−77
…						
100	455	1	45	2.22	−0.11	−105
101	511	1	45	2.22	10.88	390
102	512	1	45	2.22	10.85	388
103	513	1	45	2.22	0.78	−65
104	514	1	45	2.22	0.73	−67
105	515	1	45	2.22	0.70	−69
106	521	1	45	2.22	0.67	−70
…						
120	545	1	45	2.22	0.25	−89
121	551	1	45	2.22	0.22	−90
122	552	1	45	2.22	0.19	−91
123	553	1	45	2.22	0.10	−96
124	554	1	45	2.22	0.01	−100
125	555	1	45	2.22	−0.08	−104

References

Aaker, D. A., Kumar, V., & Day, G. S. (2003). *Marketing research* (8th ed.). New York: Wiley.

Berger, P. D., & Nasr, N. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing, 12*(1), 17–30.

Berman, J. (2013, October 9). UPS Rolls out New Service Level Offerings for Temperature True Products. Logistics Management. http://www.logisticsmgmt.com/article/ups_rolls_out_new_service_level_offerings_for_temperature_true_products. Accessed March 23, 2017.

Blattberg, R., Kim, B. D., & Neslin, S. A. (2008). *Database marketing – Analyzing and managing customers*. New York: Springer. 427ff.

Coolidge, A. (2016, March 2). *Kroger expected to book largest-ever profit*. http://www.cincinnati.com/story/money/2016/03/02/kroger-expected-book-largest-ever-profit/81177650/. Accessed March 23, 2017.

Groenfeldt, T. (2013, October 30). Kroger Knows Your Shopping Patterns Better Than You Do. *Forbes Magazine*. https://www.forbes.com/sites/tomgroenfeldt/2013/10/28/kroger-knows-your-shopping-patterns-better-than-you-do/#3a0bd81a746a. Accessed March 23, 2017.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Elements of statistical learning – Data mining, inference, and prediction* (2nd ed.). Stanford, CA: Springer.

Jackson, D. R. (1992). In quest of the grail: Breaking the barriers to customer valuation. *Direct Marketing, 54*(11), 44–48.

Jain, D., & Singh, S. (2002). Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing, 16*(2), 34–46.

Kiss, C., & Bichler, M. (2008). Identification of influencer – Measuring influence in customer networks. *Decision Support Systems, 46*(1), 233–253.

Kumar, V., Petersen, J. A., & Leone, R. P. (2007). How valuable is word of mouth. *Harvard Business Review, 85*(10), 139–146.

Lee, S. H., Catte, J., & Noseworthy, T. J. (2010). The role of network centrality in the flow of consumer influence. *Journal of Consumer Psychology, 20*(1), 66–77.

Merritt, R. (2015, November 15). Laing O'Rourke pilots IoT tech that monitors field workers' health. CIO. http://www.cio.com.au/article/589173/laing-o-rourke-pilots-iot-tech-monitors-field-workers-health/. Accessed March 24, 2017.

Onag, G. (2015, March 11). SAS eyes CRM analytics, partners with Radica in Hong Kong. Computerworld Hong Kong. http://cw.com.hk/news/sas-eyes-crm-analytics-partners-radica-hong-kong. Accessed March 24, 2017.

Peterson, H., & Lutz, A. (2015, March 6). Why Kroger Is America's Most Underrated Grocery Store. *Business Insider*. http://www.businessinsider.com/why-people-love-kroger-2015-3. Accessed March 23, 2017.

UPS. (2013, October 2). UPS Introduces New Shipment Service Levels For Its Temperature True Portfolio. *UPS Pressroom*. https://pressroom.ups.com/pressroom/ContentDetailsViewer.page?ConceptType=PressReleases&id=1426329802231-238. Accessed March 23, 2017.

UPS. (2017). History Timeline. *UPS Pressroom*. https://pressroom.ups.com/pressroom/about/HistoryStackList.page?countrylang=US-English. Accessed March 23, 2017.