Customer Relationship Management and Predictive Modeling

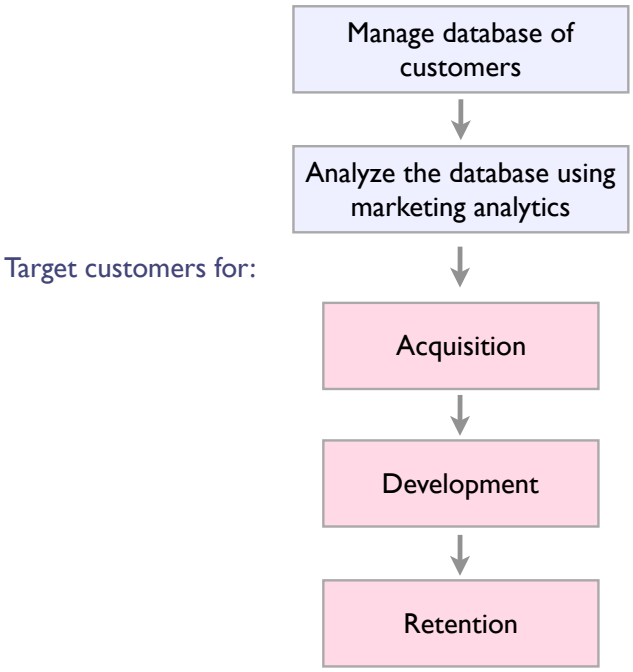
37105 Data Science for Marketing Decision Making
Günter J. Hitsch
The University of Chicago Booth School of Business

2017

Overview

1. Customer relationship management (CRM)
2. Predictive modeling
3. RFM analysis
4. Predictive modeling framework

Customer relationship management through the life cycle



Introduction

- Goal of CRM
  - Use data and marketing analytics to predict ROI's and increase customer profitability
- Customer relationship management (CRM) is closely related to
  - Database marketing
  - Direct marketing
- CRM typically addresses marketing to individual customers
  - Segments of one
- CRM is most likely the hottest field in marketing and associated with "big data" in the business press
  - Demand for marketing accountability and ROI metrics
  - Concern over low effectiveness of traditional marketing techniques
    - TV advertising

U.S. marketing spending in 2016

Traditional advertising: $ 127.7 billion

- TV: 56%
- Newspapers: 13%
- Radio: 12%
- Magazines: 12%
- Outdoor: 6%
- Cinema: 1%

Direct and digital marketing: $ 163.7 billion

- Direct mail: 29%
- Tele-services: 27%
- Search advertising: 19%
- Display advertising: 18%
- E-mail: 2%

Source: Winterberry Group

Introduction

- Technological innovations are the source of the growth in CRM activities
  - Information storage technology
  - Printing
  - Distribution
  - Analytics
- Availability of detailed customer information on a large number of actual or potential customers
- Data bases of customer transaction, contact, and interaction records

Applications of CRM

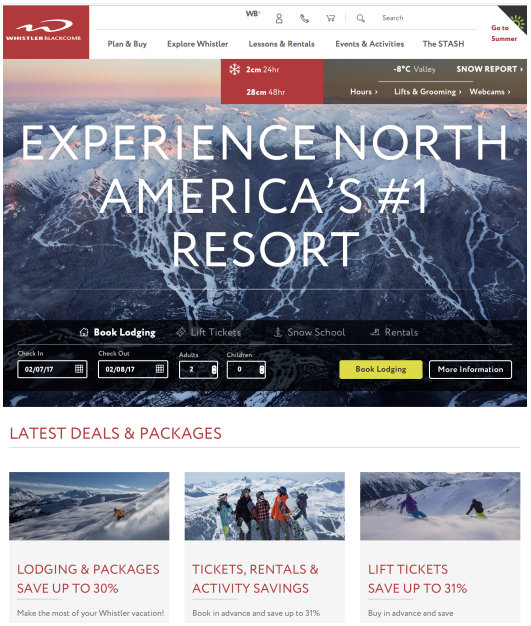
- Catalog retailing
- Travel services
- Airlines
- Credit card companies
- Insurance
- Real estate
- Luxury/specialty goods
- E-tailing
- Telecoms

Examples of CRM activities

- Catalog planning
  - Which customers in the house file should receive a catalog?
  - What mailing lists should a company buy (rent) for customer acquisition?
- Managing retail bank customers
  - Who are the most profitable customers?
  - What is the lifetime value of these customers?
  - Can we "profile" these customers in order to improve the efficiency of prospecting?
- Retaining wireless mobile communications customers
  - Which customers are likely to cancel her/his contract?
  - Can we give these customers incentives to stay?

Predictive modeling: Example

- Whistler Blackcomb (Vail Resorts) mails an offer for vacation packages to 1.5 million potential customers
  - Names obtained from a list of sports and ski magazine subscribers
- Data
  - No. of prospects: 1.5 mill.
  - Cost per mailing: \$1.5
  - Profit contribution: \$200
  - Response rate of list: 0.8%
  - Cost per mailing includes printing, distribution, and list rental
- What is the return on this direct mail campaign?



ROI calculation

Number of prospects	1,500,000	1,500,000
Cost per mailing (\$)	1.5	1.5
Profit contribution (\$)	200	200
Response rate of list	0.8%	0.6%
Expected no. responses	12,000	9,000
Total cost of mailing	2,250,000	2,250,000
Profit contribution (\$)	2,400,000	1,800,000
Profit (\$)	150,000	-450,000
ROI	6.7%	-20.0%

- This direct mail campaign has an element of targeting
  - Whistler Blackcomb bought a list of ski and sports enthusiasts
  - Response rate on mass mailing would be much less than 0.1%
- But could Whistler Blackcomb further improve the ROI?

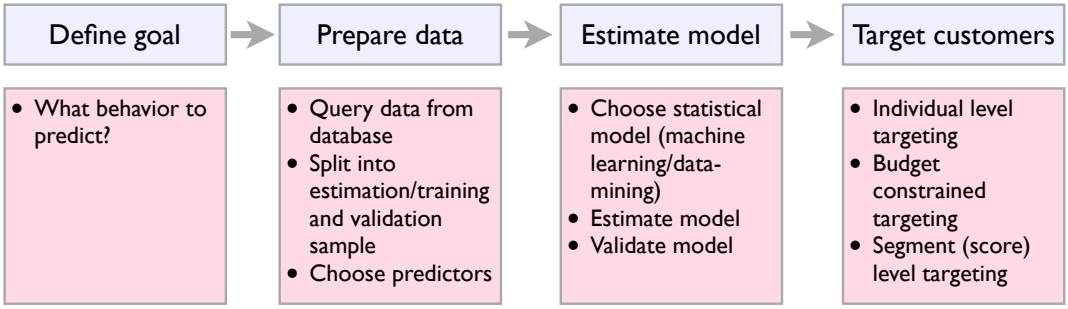
- Use database of past mailing
  - Find variables that predict responsiveness to mailing
  - Target by decile according to response rate

Decile (score)	Response rate	No. responses	Cost	Profit	ROI	Cum. profit	Cum. ROI
10	3%	4,500	225,000	675,000	300.0%	675,000	300.0%
9	1.7%	2,550	225,000	285,000	126.7%	960,000	213.3%
8	1%	1,500	225,000	75,000	33.3%	1,035,000	153.3%
7	0.8%	1,200	225,000	15,000	6.7%	1,050,000	116.7%
6	0.6%	900	225,000	-45,000	-20.0%	1,005,000	89.3%
5	0.4%	600	225,000	-105,000	-46.7%	900,000	66.7%
4	0.2%	300	225,000	-165,000	-73.3%	735,000	46.7%
3	0.15%	225	225,000	-180,000	-80.0%	555,000	30.8%
2	0.1%	150	225,000	-195,000	-86.7%	360,000	17.8%
1	0.05%	75	225,000	-210,000	-93.3%	150,000	6.7%

Lessons

- Predictive modeling can dramatically improve profitability compared to targeting that is not precisely focused (e.g. mass mailings)
- Develop systematic framework for predictive modeling
- Use a modern statistics/machine learning approach for predictive modeling

The predictive modeling approach



In-house data

- Contrast two types of data that can be used for predictive modeling:
  1. Who the customers are
     - Demographics, zip code, life-style indicators, ...
  2. What the customers do
     - Use a lot of credit, respond to e-mails we send them, . . .
- Latter information: house file of contact and transaction data
- Predictive modeling usually works best for house file data (why?)
  - However: Limited to existing customers — no prospects

RFM analysis

- Oldest CRM technique to segment customers and predict buying behavior for each segment
  - Has been around since the early 1960's
  - Practical value: Easy to understand and communicate to someone with little or no training in statistics
- Makes use of powerful metrics on customer behavior that are contained in most house files
  - **Recency**
    - How long ago did the customer make a purchase
  - **Frequency**
    - How many purchases has the customer made (in total or in specific time period)
  - **Monetary value**
    - How much has the customer spent in total (in total or in specific time period)

The RFM approach

- Use data extract from (typically) in-house data
- Choose recency, frequency, and monetary value variables
- Decide into how many groups (quantiles) to classify costumers
  - Typically pick 5 groups — split consumers into quintiles
  - If you pick 10 groups — get deciles
  - But of course conceptually works for an arbitrary number of N groups
- Assign customers to segments
  - Rank observations according to RFM values, split into groups that have an (approximately) equal number of observations
  - If there are large "lumps" of observations with equal values group sizes will be unequal (example: 37 percent of customers have frequency $= 1$)
  - Common approach: Most desirable group for each variable is in group 1, least desirable group in group N, but approach works equally well if this convention is ignored

General purpose function that is robust to "lumps" in the data:

```
# createBins ---------------------------------------------------------
# Inputs: x, a vector of numbers (or a column of a data frame/table)
#         N, the number of bins (groups) to create

createBins <- function(x, N) {
   cut_points = quantile(x, probs = seq(1/N, 1-1/N, by = 1/N), type = 2)
   cut_points = unique(cut_points)
   bins = cut(x, c(-Inf, cut_points, +Inf),
              label = 1:(length(cut_points)+1))

   return(as.numeric(bins))
}
```

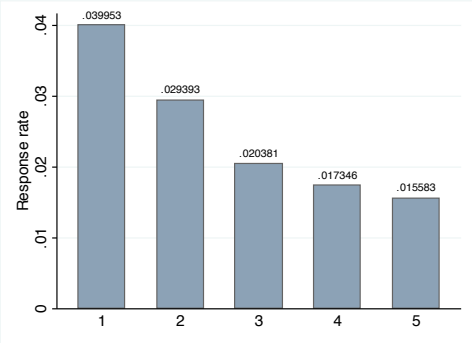

Estimating segment-level response

- First, segment customers
- Estimate average response (outcome) for each segment
  - In the example below the outcome is a purchase indicator, hence the average response is a response rate (probability)

```
customer_DT[, recency := createBins(days_since_last_purchase, 5)]

response_DT = customer_DT[, .(response_rate = mean(purchase),
                              N_obs         = .N),
                          by = recency]

response_DT[order(recency)]
```

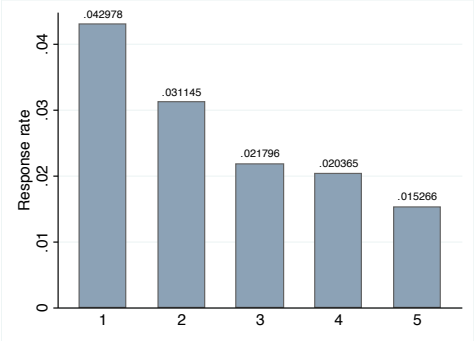
```
   recency response_rate N_obs
1:       1    0.03995269 19448
2:       2    0.02939340 19222
3:       3    0.02038079 19381
4:       4    0.01734583 19313
5:       5    0.01558347 19187
```

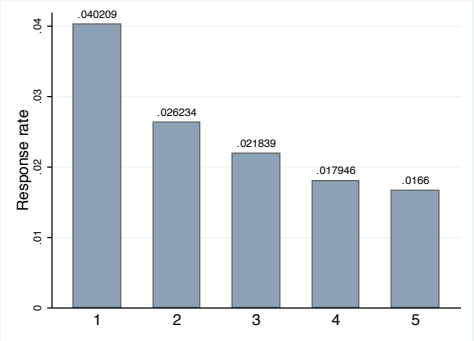
Response by recency



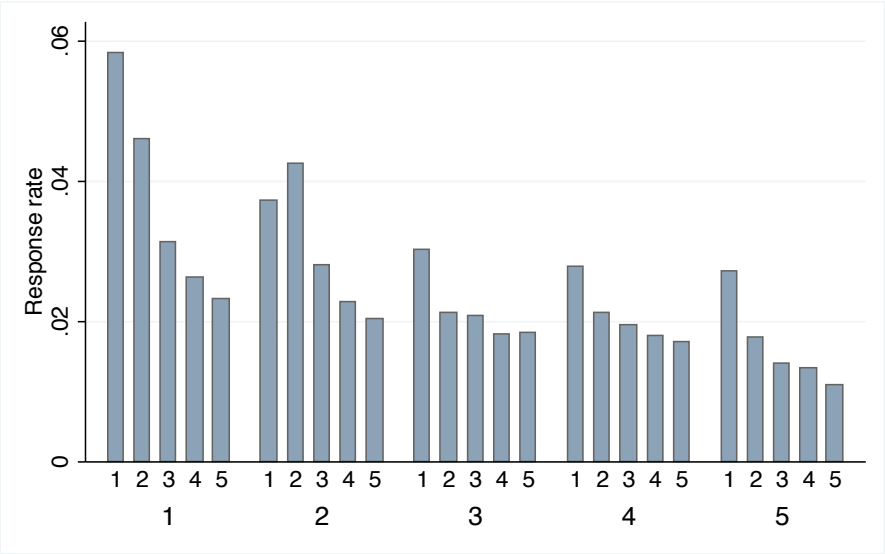
- Here: Monotonically declining relationship between recency quintile and response rate
- Sometimes we find an "inverse U-shape"
  - Why?

Response by frequency and monetary value



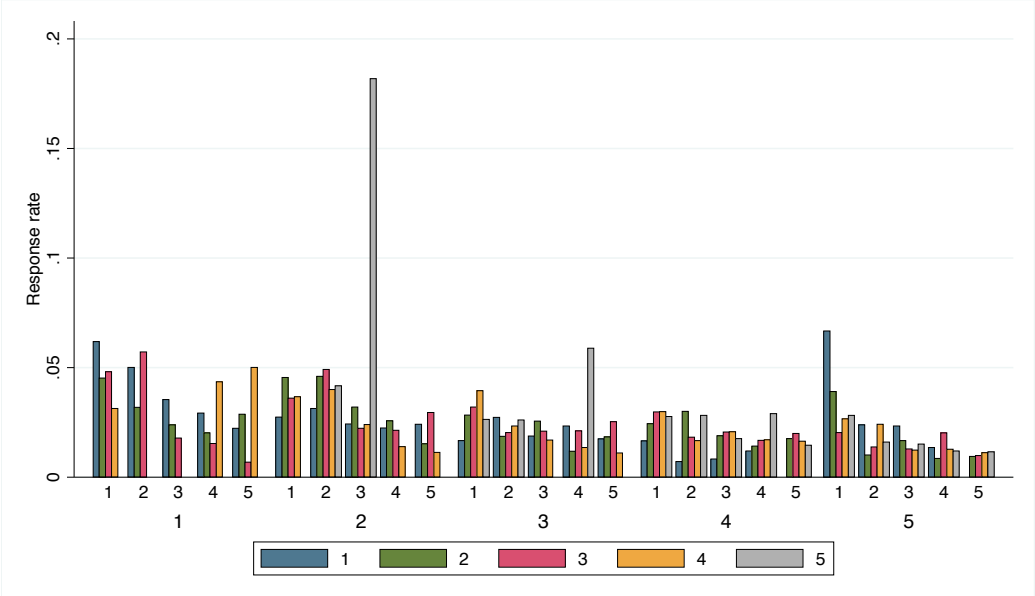


Segmentation by recency and frequency



- Top: recency segments, bottom: frequency segments

Segmentation by recency, frequency, and monetary value



- Top: recency segments, middle: frequency segments, colors: monetary value segments

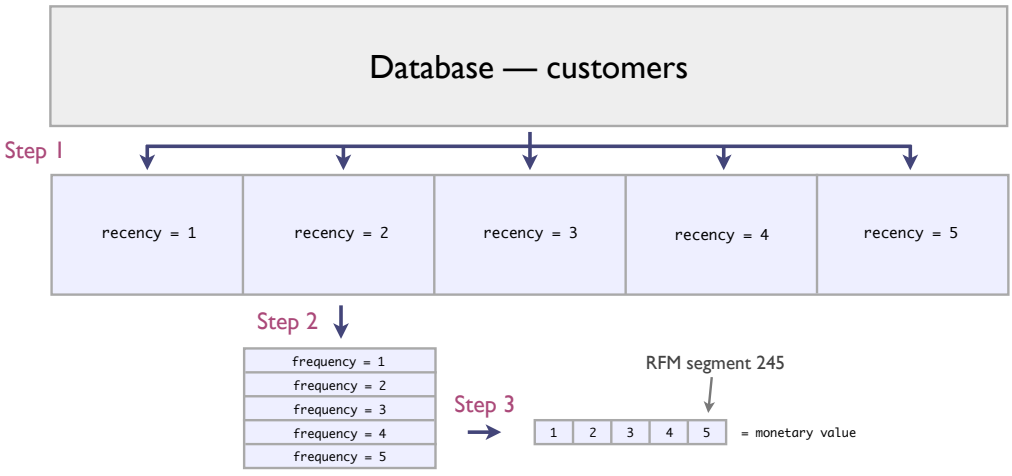
Create RFM segments

- Assign each customer a 3-digit code
  - Example: Segment 125 corresponds to recency $= 1$, frequency $= 2$, monetary value $= 5$
  - We obtain $5 \cdot 5 \cdot 5 = 125$ segments if the segmentation is based on quintiles
  - If we use deciles, we get $10^3 = 1{,}000$ segments
- Calculate the RFM index using the following formula:

$$\texttt{rfm\_index} = 100 * \texttt{r\_index} + 10 * \texttt{f\_index} + \texttt{m\_index}$$

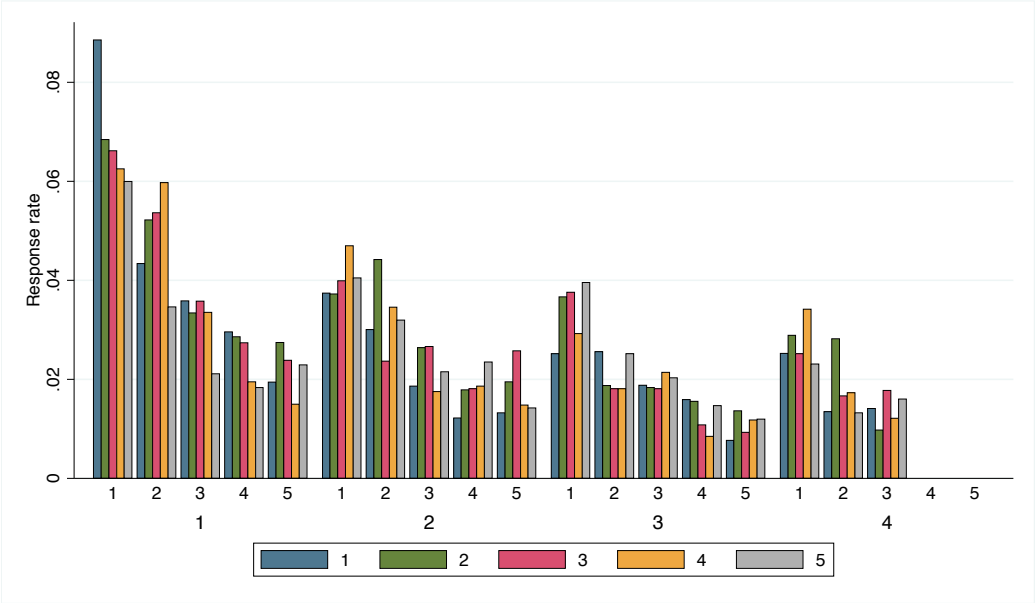
  - This is somewhat of "cheap trick"; instead you could create groups based on `.GRP` in data.table (, although these groups would not be as easily interpretable (which doesn't matter for prediction)

Sequential segmentation approach



- Sequential segmentation approach can provide a better segmentation than the independent segmentation approach we used before
  - Greater degree of within-segment homogeneity

- Sequential segmentation approach delivers a cleaner segmentation than the independent approach



- Top: recency segments, middle: frequency segments, colors: monetary value segments

Comparison: RFM analysis and regression

RFM analysis provides a segmentation into a total of N groups.

Let $X_{in} = 0, 1$ be an indicator that equals 1 if observation i was assigned to group/segment n.

The response prediction in RFM analysis is equivalent to running the following regression:

$$Y_i = \sum_{n=1}^{N} \beta_n X_{in} + \epsilon_i$$

No intercept!

$\hat{\beta}_n$ will be the estimated response rate in group n.

Lesson: RFM analysis is a simple way to explain data analytics to your client without having to go into the details of regression analysis (although this is becoming less and less important)

Limitations of RFM analysis

Ideally we would segment customers along many variables.
Example: RFM variables for two major departments and 20 region indicators. If we segment into 5 RFM groups, we end up with $5^6 \cdot 20 = 312,500$ segments. With 10 RFM groups: $10^6 \cdot 20 = 20,000,000$ segments.

In either case RFM is infeasible unless we have a huge sample. This is called the *curse of dimensionality* : With N groups and k segments in each group the total number of segments rises exponentially in N:

$$\text{total no. segments} = k^N$$

Regression analysis and modern machine learning tools (LASSO, random forest, ...) allow us to stay true to the spirit of RFM analysis (essentially a non-parametric estimator) but make the analysis feasible.

Predictive modeling steps

1. Extract a sample of customers
2. Split the sample into a *training sample* and a *validation sample*
   - Training sample: Perform statistical analysis
   - Validation sample: Investigate predictive power of the model
   - Training sample also called estimation sample
3. Estimate the response model using the training sample
4. Predict response probabilities for the validation sample
5. Evaluate the model
   - Compare predicted response probabilities in validation sample to actual response behavior in validation sample
   - Widely used metrics in CRM: Lift and gains tables and charts
6. Develop targeting strategy

Example

- Company with data base of more than 35 million customers
- Goal: Mail a spring tabloid featuring women's clothes and shoes
- Key question:
  - Can we improve the ROI from the mailing using predictive modeling?

Predictive modeling and customer economics

Notation: Y_i is customer-level spending, and $\pi(Y_i)$ is the profit contribution for a given level of customer spending. For example, if the margin m is constant, then $\pi(Y_i) = mY_i$. $\boldsymbol{x}_i$ includes all observed customer attributes.

Expected profit when targeting a customer, not including the targeting cost:

$$\mathbb{E}(\pi(Y_i)|\boldsymbol{x}_i) = \Pr\{Y_i > 0|\boldsymbol{x}_i\} \cdot \mathbb{E}(\pi(Y_i)|Y_i > 0, \boldsymbol{x}_i)$$

c_i is the cost of targeting customer i. Then the expected ROI from a targeting effort is

$$\mathbb{E}(\text{ROI}_i|\boldsymbol{x}_i) = \frac{\mathbb{E}(\pi(Y_i)|\boldsymbol{x}_i) - c_i}{c_i}$$

Here we made an important implicit assumption: Spending is 0 when the customer is not targeted. We will revisit this crucial assumption later.

Customer economics: Implementation

One possible targeting rule: Target customers with positive expected profits, $\mathbb{E}(\text{ROI}_i|\boldsymbol{x}_i) > 0$

Implementation:

- Data: Cost of targeting, c_i
- Prediction:
  - Conditional expected profits
  - Expected response rate
  - Expected profit conditional on a response

Step 1: Extract a sample of customers

Obtain data from in-house customer data base

- Customers who received spring tabloids featuring women's clothes and shoes in the past
- Purchase information and customer attributes

Variable name	Definition
customer_no	Customer id, can be linked to address
buytabw	Dependent variable (1 = purchase, 0 = no purchase)
tabordrs	Total orders from tabloids
divsords	Orders with shoe division
divwords	Orders with women's division
spgtabord	Total spring tabloid orders
moslsdvs	Months since last shoe order
moslsdvw	Months since last women's order
moslstab	Months since last tabloid order
orders	Total orders

Step 2: Split into training and validation sample

- Create a training or validation sample indicator variable
- Always set the seed to initialize the random number generator to a specific state, otherwise you cannot exactly replicate the results later on

```
set.seed(5807)
target_DT[, training_sample := rbinom(nrow(target_DT),
                                      size = 1, prob = 0.6)]

table(target_DT$training_sample)/nrow(target_DT)


        0         1
0.4018035 0.5981965
```


Step 3: Estimate response model in training sample

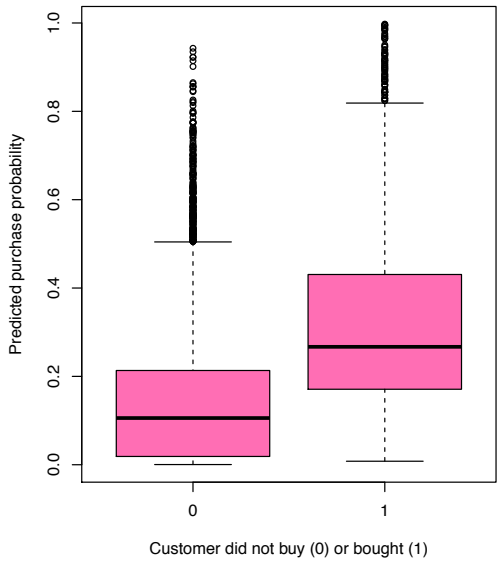
```
Call:
glm(formula = buytabw ~ . - customer_no - training_sample, family = binomial(),
    data = target_DT[training_sample == 1])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2276  -0.6381  -0.3645  -0.1193   3.0816

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.970187   0.084903 -11.427  < 2e-16 ***
tabordrs     0.044068   0.012584   3.502 0.000462 ***
divsords    -0.014717   0.014574  -1.010 0.312581
divwords     0.099861   0.007434  13.434  < 2e-16 ***
spgtabord    0.095833   0.017425   5.500 3.80e-08 ***
moslsdvs    -0.009755   0.002001  -4.876 1.08e-06 ***
moslsdvw    -0.071548   0.004915 -14.558  < 2e-16 ***
moslstab    -0.048652   0.004265 -11.408  < 2e-16 ***
orders      -0.044747   0.005295  -8.451  < 2e-16 ***
---
```

Step 4: Predict response probabilities

Step 5: Evaluate the model

Compare predicted response probabilities to the actual buying behavior in the validation sample

Box plot:



Scoring and segmentation

- Score along the predicted response
- Assign a score from 1 to 10 to each customer in the validation sample
  - 1: lowest 10% (decile) of predicted response
  - 10: highest 10% (decile) of predicted response
- For the resulting 10 segments calculate:
  - Number of all customers in segment
  - Number of customers who bought in segment
  - (Mean) predicted response rate in segment

Lift and gains table: Auxiliary information

Score	No. of observations	No. of buyers	Predicted response rate	Actual response rate	Cum. percent mailed	Cum. no. of observations	Cum. no. of buyers	Actual cum. response rate	Percent of all buyers in sample reached	Cum. percent of all buyers in sample reached
10	821	411	0.557	0.501	10	821	411	0.501	28.6	28.6
9	821	279	0.329	0.34	20	1642	690	0.42	19.4	48
8	822	194	0.251	0.236	30	2464	884	0.359	13.5	61.5
7	821	187	0.2	0.228	40	3285	1071	0.326	13	74.5
6	822	174	0.16	0.212	50	4107	1245	0.303	12.1	86.6
5	821	117	0.119	0.143	60	4928	1362	0.276	8.1	94.8
4	821	65	0.077	0.079	70	5749	1427	0.248	4.5	99.3
3	822	8	0.036	0.01	80	6571	1435	0.218	0.6	99.9
2	821	1	0.013	0.001	90	7392	1436	0.194	0.1	99.9
1	822	1	0.004	0.001	100	8214	1437	0.175	0.1	100

Actual response rate	No. of buyers / no. of observations in decile
Cumulative percent mailed	Percent of all customers mailed if we target score 10 first, score 9 second, ...
Cumulative number of observations	
Cumulative number of buyers	
Actual cumulative response rate in segment (score) n	Cum. no. of buyers / cum. no. of observations
Percent of all buyers in sample reached	Percent of all buyers (across all segments) within a given decile (need to calculate the sum of the number of buyers across all 10 deciles)
Cumulative percent of all buyers in sample reached	

Lift and cumulative lift

Widely used metrics in CRM to evaluate the fit or predictive power of a model

Create segments (scores) of customers according to predicted response

Lift factor for segment k:

$$\text{lift}_k = 100 \cdot \frac{\text{response rate in segment } k}{\text{avg. response rate in sample}}$$

- The lift is a measure of the predictive power of the model compared to a random targeting approach
- Claim: If there is no predictive power $\rightarrow$ lift $= 100$ for all segments

Cumulative lift:

$$\text{cum. lift}_n = 100 \cdot \frac{\text{respones rate among scores } k \geq n}{\text{avg. response rate in sample}}$$

- Response rate that we get by targeting the top scores (according to their predicted response rate) relative to random targeting

(Cumulative) Gains

Cumulative gains for score or segment n: percent of all responders (buyers) reached if we target all customers with a score of n or above

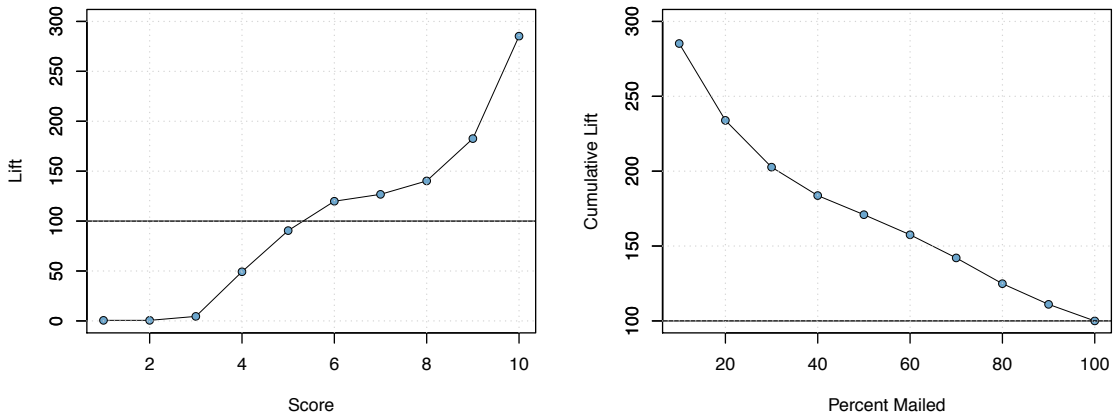
- Shows us how many of all potential customers we can skim off by targeting the top scores according to the predictive model

Note: All lift and gains calculations should be based on the response rates in the validation sample—evaluation should be out-of-sample, not in-sample

Lift and gains table

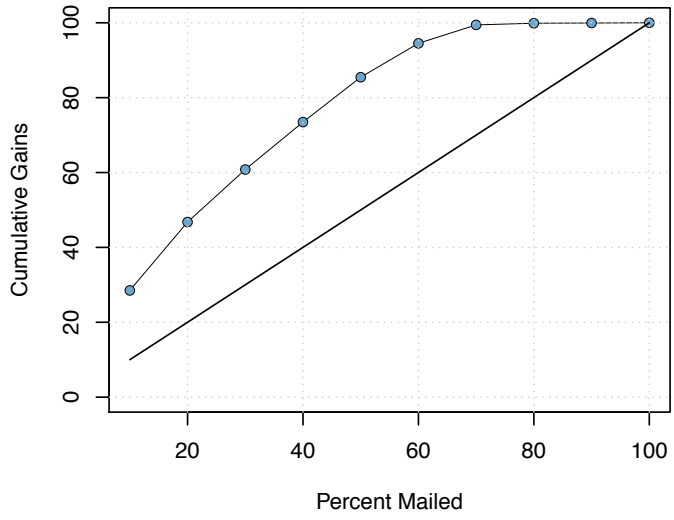
Score	Cum. percent mailed	Predicted response rate	Actual response rate	Percent of all buyers in sample	Lift	Cumulative lift	Cumulative gains
10	10	0.557	0.501	28.6	286.2	286.2	28.6
9	20	0.329	0.34	19.4	194.2	240.2	48
8	30	0.251	0.236	13.5	134.9	205.1	61.5
7	40	0.2	0.228	13	130.2	186.4	74.5
6	50	0.16	0.212	12.1	121	173.3	86.6
5	60	0.119	0.143	8.1	81.5	158	94.8
4	70	0.077	0.079	4.5	45.3	141.9	99.3
3	80	0.036	0.01	0.6	5.6	124.8	99.9
2	90	0.013	0.001	0.1	0.7	111	99.9
1	100	0.004	0.001	0.1	0.7	100	100
		Avg. actual response rate:	0.175				

Lift and cumulative lift charts



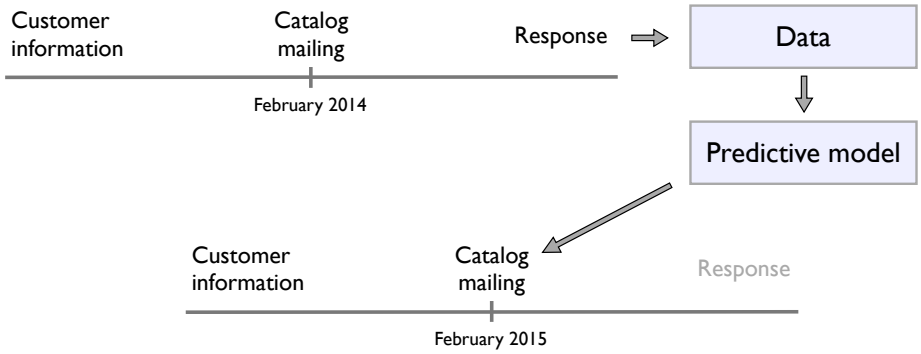
Evidence that model can predict buying behavior in validation sample

Cumulative gains chart



Step 6: Develop targeting strategy

- Based on step 5, does the model appear valid?
  - No
    - Check each step of your analysis
    - Collect experimental data: Send catalog to a random sample of customers and record their response
  - Yes: Develop a precise targeting strategy



Possible targeting strategies

$$\mathbb{E}(\text{profit}_i|\boldsymbol{x}_i) = \Pr\{Y_i > 0|\boldsymbol{x}_i\} \cdot \mathbb{E}(\pi(Y_i)|Y_i > 0, \boldsymbol{x}_i) - c_i$$

- Individual level targeting
  1. Predict conditional expected profit or ROI for each customer
  2. Target customers with expected profit or ROI that is positive or above threshold

- Budget constrained targeting
  1. Rank customers according to their expected profitability
  2. Target customers according to their expected profitability until the budget constraint is exhausted or expected profits are negative

- Segment (score) level targeting
  1. Predict profit or ROI for each segment (score)
  2. Target segments with profit/ROI that is positive or above a chosen threshold

- Note: Targeting at the individual level is typically best

Summary

- Use predictive modeling to focus marketing efforts on customers (segments) with high expected ROI's
- Predicting response
  - Segmentation based on RFM analysis
  - Individual level predictions (segments of one) using predictive modeling
- The predictive modeling approach: