Data Analysis Report

Version 1.0

Jiang Zhou, PhD

Email: Jiang.zhou@gmail.com
Telephone: 978-726-3182

Jiang Zhou

Table of Contents

EXECUTIVE SUMMARY

The purpose of this study is to solve the following business problems. Due to the cost of delivering email ACME decides to send email to only 25% of it's subscriber base for week 27.
Given the data provided:
(1) Which subscribers would you send email to?
(2) Which campaign(s) would you deliver to them?
(3) What do you expect the response rate to be?

In this analysis, I analyzed 10,000 subscribers' responses to 260,000 campaigns in 26 weeks. I assumed the data represent the original subscriber population. More than 10 derived variables were developed. The data from the first 23 weeks were used as training set based on which 10 models were created. The models used include simple cell models, logistic regressions, decision trees (CARTs), and gradient boosting tree. The remaining 3 weeks were used as testing set to compare the performances of these models. Gain charts and lift curves were calculated to compare the models. The gradient boosting tree provided the best performance and was used as the final model to provide answers to the above questions. Appendix B gives the top 30 subscribers and the campaigns that should be sent to them. The list of full 25% subscribers and their campaigns is included in a separate text file. The expected response rate for the best 25% subscribes in week 27 is 97.97% with a standard error 0.22%.

As a by-product of this study, I also analyzed the effect of training sample size on the model performance. It is shown that all the models except cell models have reasonable performance when only 30% or 10% of training sets were used. Using less training data can reduce the cost of data collection and improve the campaign's speed to the market. The following sections describe in detail the steps I took in the analysis and the results I achieved.

DATA PREPARATION

I received sq_small.txt with 2600 records and sq_large.csv.zip from www.adknowledge.com/sq_large.csv.zip with 260,000 records. I decided to go for the large data set.

The first thing I did was to unzipped the file and ran a number of Unix bash commands to get a sense of the data. Since I was using Cygwin under Microsoft Windows, I had access to most of the UNIX bash commands. I calculated the frequency for each variable using the combination of commands cat, cut, sort, and uniq. Appendix A gives the frequency tables. From the frequency tables we can see that there is no missing value in the data fields. Each week has 10,000 records and each subscriber has 26 records. Data also distribute almost evenly on user category, state, gender and response. Since the data are well balanced, I suspected they are generated through careful design. However, in the study, I assumed they are the original subscriber population.

Once I was sure that the data are clean, I loaded the file into an Oracle database schema. I generated the frequency statistics using SQL and found that they are the same as the one I generated using UNIX commands. This confirmed the file was correctly imported into Oracle.

The table 1 gives the statistics on the data. Table 1b gives the response rate by week.

Table 1. Overall Distributions of the Data

Number of Subscribers	10,000
Number of Response	110358
Number of Non Response	149642
Overall Response Rate	42.4%

Table 1b. Response Rate by Week

Week	# of Campaign	# of Response	Response Rate
1	10000	3009	30.1%
2	10000	2887	28.9%
3	10000	3271	32.7%
4	10000	3137	31.4%
5	10000	3256	32.6%
6	10000	3271	32.7%
7	10000	3281	32.8%
8	10000	3225	32.3%
9	10000	3241	32.4%
10	10000	3259	32.6%
11	10000	3870	38.7%
12	10000	3850	38.5%
13	10000	4041	40.4%
14	10000	4858	48.6%
15	10000	4849	48.5%
16	10000	4809	48.1%
17	10000	4944	49.4%
18	10000	4894	48.9%
19	10000	4860	48.6%
20	10000	4909	49.1%
21	10000	5108	51.1%
22	10000	5163	51.6%
23	10000	5238	52.4%
24	10000	5740	57.4%
25	10000	5675	56.8%
26	10000	5713	57.1%

FEATURE CALCULATION

Four data fields that come with the data may be used as inputs into models. They are, user category, state id, gender and campaign id. I also developed the following derived fields.

FRQ_W_2: number of responses in last two weeks

FRQ_W_3: number of responses in last three weeks

FRQ_W_4: number of responses in last four weeks

FRQ_W_5: number of responses in last five weeks

LIFE_TIME_RESPONSE_RATE: overall response rate from week 1 to last week

LAST_CAMPAIGN: campaign id of last week

LAST_2ND_CAMPAIGN: campaign id of last second week

IS_LAST_ A_RESPONSE: is last campaign a response? (0 No, 1 Yes)

WEEK_SINCE_LAST_RESPONSE: number of weeks since last response

The above features were calculated on per subscriber basis. I also calculated the following feature at state level for each week.

STATE_RESPONSE_RATE: response rate for each state from week 1 to last week.

I used extreme caution to make sure that responses to current or future campaigns were NOT used in calculating the features.

Table 2 gives the original and most of derived variables for subscriber id 3.

Table 2. Original and Derived Variable for Subscriber ID 3

WEEK ID	U_C (VAR 3)	ST (VAR 4)	SEX (VAR 5)	CAMP ID (VAR6)	RESONSE (VAR7)	FRQ W 2	FRQ W 3	FRQ W 4	FRQ W 5	LIFE RES RATE	LAST CAMP ID	LAST 2ND CAMP ID	IS LAST A RES	WEEKS SINCE LAST RESPONSE
1	C	46	M	3	0									
2	C	46	M	4	1	0	0	0	0	0.0%	3		0	
3	C	46	M	5	1	1	1	1	1	50.0%	4	3	1	1
4	C	46	M	6	1	2	2	2	2	66.7%	5	4	1	1
5	C	46	M	7	1	2	3	3	3	75.0%	6	5	1	1
6	C	46	M	8	0	2	3	4	4	80.0%	7	6	1	1
7	C	46	M	9	0	1	2	3	4	66.7%	8	7	0	2
8	C	46	M	10	1	0	1	2	3	57.1%	9	8	0	3
9	C	46	M	1	0	1	1	2	3	62.5%	10	9	1	1
10	C	46	M	2	0	1	1	1	2	55.6%	1	10	0	2
11	C	46	M	3	0	0	1	1	1	50.0%	2	1	0	3
12	C	46	M	4	0	0	0	1	1	45.5%	3	2	0	4
13	C	46	M	5	1	0	0	0	1	41.7%	4	3	0	5
14	C	46	M	6	1	1	1	1	1	46.2%	5	4	1	1
15	C	46	M	7	1	2	2	2	2	50.0%	6	5	1	1
16	C	46	M	8	0	2	3	3	3	53.3%	7	6	1	1
17	C	46	M	9	0	1	2	3	3	50.0%	8	7	0	2
18	C	46	M	10	1	0	1	2	3	47.1%	9	8	0	3
19	C	46	M	1	0	1	1	2	3	50.0%	10	9	1	1
20	C	46	M	2	1	1	1	1	2	47.4%	1	10	0	2
21	C	46	M	3	0	1	2	2	2	50.0%	2	1	1	1
22	C	46	M	4	1	1	1	2	2	47.6%	3	2	0	2
23	C	46	M	5	1	1	2	2	3	50.0%	4	3	1	1
24	C	46	M	6	1	2	2	3	3	52.2%	5	4	1	1
25	C	46	M	7	1	2	3	3	4	54.2%	6	5	1	1
26	C	46	M	8	0	2	3	4	4	56.0%	7	6	1	1

Last campaign ID and last second campaign ID were created based on the thought that customer may compare current campaign with last 1 or 2 offers to decide to respond or not.

All derived variables except WEEK_SINCE_LAST_RESPONSE are bounded. I converted WEEK_SINCE_LAST_RESPONSE into a discrete variable with 6 categories using rules in table 3. By doing so, two benefits were achieved. Firstly, the variable becomes bounded. Secondly, missing value is represented by a separate category. All other variables do not contain missing value after the week 2.

Table 3. Rules to convert WEEK_SINCE_LAST_RESPONSE into Discrete Variable

Low	High	Category
Missing	Missing	1
1	1	2
2	2	3
3	3	4
4	5	5
6	Infinite	6

MODELS DEVELOPMENT

As a standard practice in a model developing process, we divided the data into two parts: records before week 24 were used as development data and the data on week 24, 25, and 26 were used as test data.

1. Cell Model:

Given the data from weeks 1-23, I calculated the response rate for each unique combination of user category, state id, gender and campaign, using the following SQL "group by" statement.

Table 4 gives the response rate, i.e., number of response divided by number of campaigns, for 20 unique combinations. For example, from week 1 to week 23, there are 38 out of 69 campaigns got responses for user category A, Stat ID 1, Female and campaign. So the responses rate is 38/69=55.1%.

Table 4. Response Rate for 20 Cells

(Unique Combinations of User Category, State ID, Gender and Campaign)

User Category	State ID	Gender	Campaign	# of Campaign	# of Responses	Response Rate
A	1	F	1	69	38	55.1%
A	1	F	2	67	28	41.8%
A	1	F	3	67	15	22.4%
A	1	F	4	66	35	53.0%
A	1	F	5	66	17	25.8%
A	1	F	6	67	23	34.3%
A	1	F	7	66	29	43.9%
A	1	F	8	65	18	27.7%
A	1	F	9	66	15	22.7%
A	1	F	10	68	31	45.6%
A	1	M	1	45	32	71.1%
A	1	M	2	45	31	68.9%
A	1	M	3	45	24	53.3%
A	1	M	4	43	33	76.7%
A	1	M	5	43	21	48.8%
A	1	M	6	45	34	75.6%
A	1	M	7	45	35	77.8%
A	1	M	8	44	31	70.5%
A	1	M	9	41	30	73.2%
A	1	M	10	41	26	63.4%

It turned out there are 4000 cells, which is precisely 50(levels of state)*4(levels of user category)*10(levels of campaign)*2(levels of gender). The average size of cells is 57.5 with a standard deviation of 15. The minimum and maximum sizes are 21 and 113. This again suggests that the data were produced based on analyst's design. It is not likely the original customer base distributes so evenly across multiply attributes.

Once I got the response rate for each cell, I could find out for each record in the weeks 24, 25, and 26 which cell it falls in and assigned the response rate to that record as its score. I did this using a SQL joining statement.

2. Logistic regression model

Since the target variable that we are trying to predicting is binary, logistic regression is the natural choice. I import the data into Splus and used glm function with binomial option to create the model. In addition to user category, state id, gender and campaign, I also included IS_LAST_A_RESPONSE, LAST_CAMPAING_ID, FRQ_W_5, WEEK_SINCE_LAST_RESPONSE and STATE_RESPONSE_RATE. I also include the interaction between user category and campaign id, the interaction between gender and campaign id. Variables such as FRQ_W_5, are not "instant" variables like gender, and they require the knowledge about responses from previous weeks. To get an accurate calculation of these "non instant" variables, data from week 1 and week 5 were excluded from model building. Of course, the starting week for calculating feature is still week 1. The trained model was applied to weeks 24, 25 and 26.

3. CART

With CART, we do not have to deal with the issues like function form specification, collinearity, etc. So in addition to variable used in logistic regression, I also included FRQ.W.2, FRQ.W.3, FRQ.W.4,LAST_2ND_CAMPAIGN. The tree was pruned using 10-fold cross validation.

4. Gradient Boosting Tree

Gradient boosting tree is described in book The Element of Statistical Learning: Data Mining, Inference, and Prediction (Hastie, T., R. Tibshirani, and J. Friedman). The gradient boosting tree I developed has 200 sub trees and with the same set of variables used in CART.

MODELS PERFORMANCE EVALUATION

Chart 1 shows the gain chart for the three models.

Chart 1. Gain Chart for Three Models on Testing Set

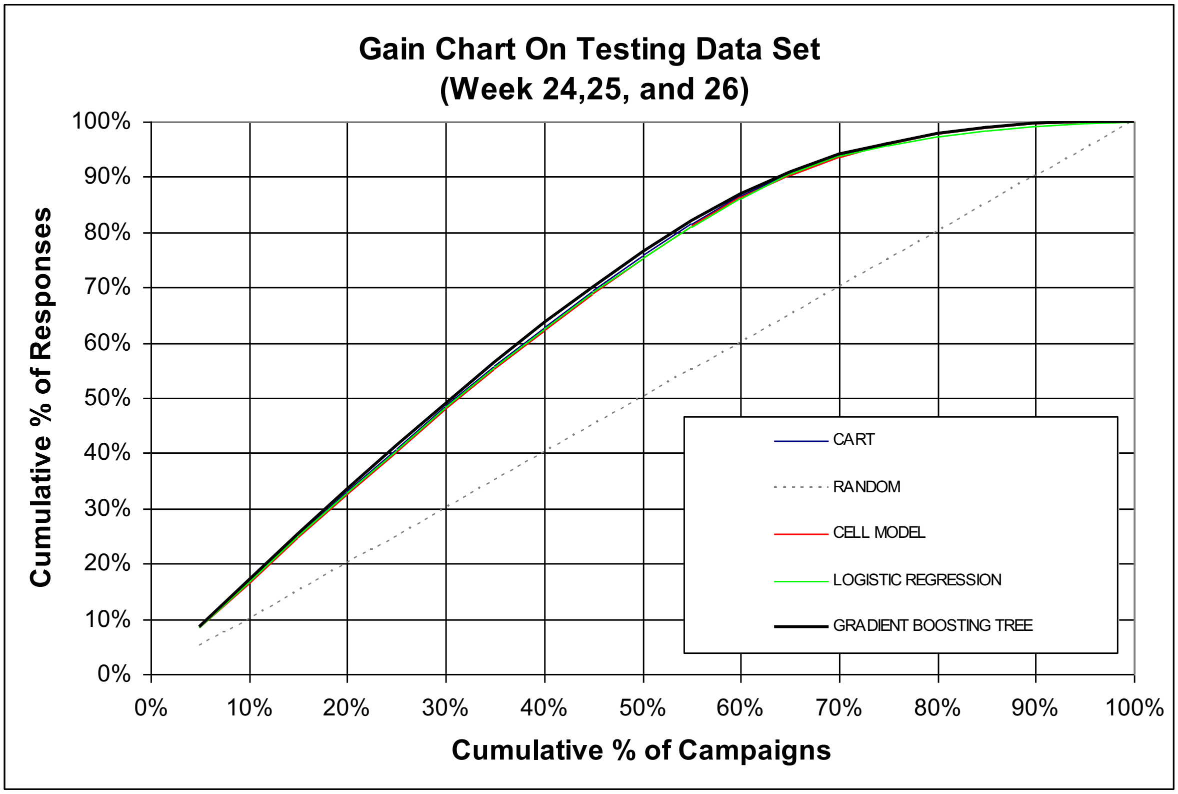
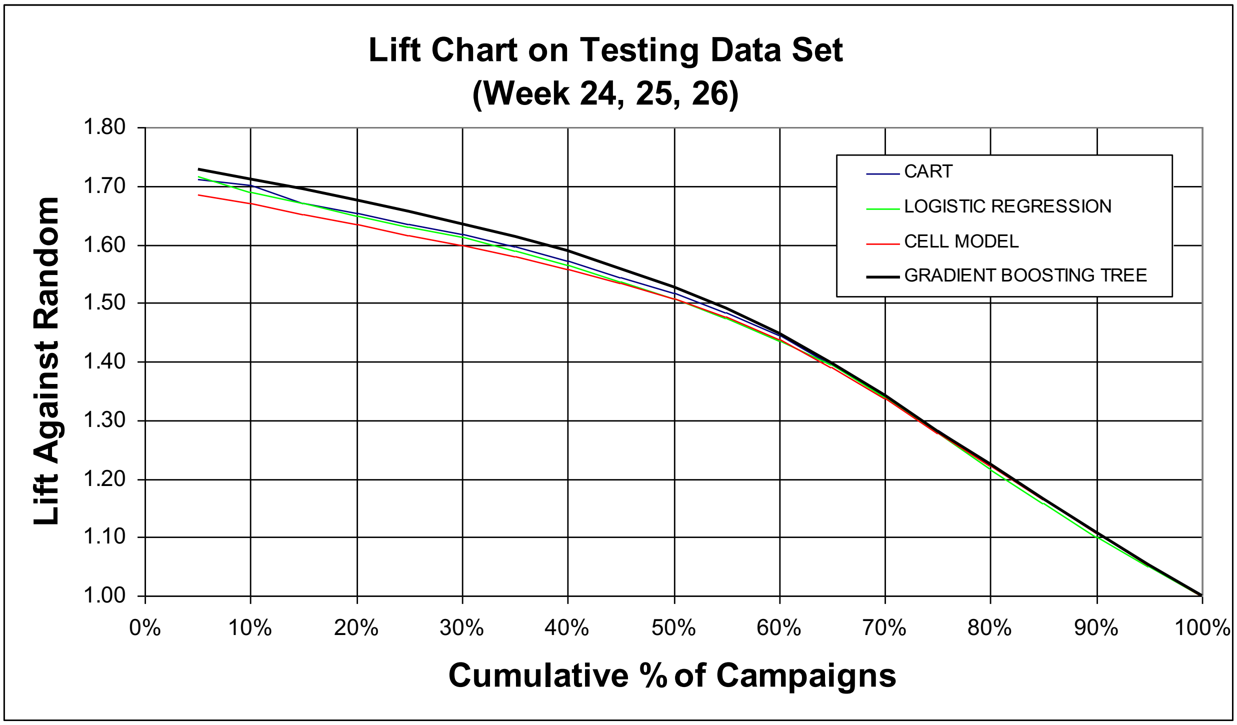


Chart 2 shows the lift curves for the three models. Lift curves are calculated by simply dividing the cumulative percentage of responses of a model by the cumulative percentage of total number of campaigns.

Chart 2. Lift Chart for Three Models on Testing Set



From the above charts we can see that gradient boosting tree provides best performance. Performances of CART and logistic regression are similar. Cell model has the worst performance. However, the performance discrepancies among these models are small. That begs the question: why we need to spend extra effort to develop more complex models than cell models where can be simply implemented in SQL? To answer this question, I did some experiments that are described in the next section.

IMPACT OF TRAINING DATA SIZES

I randomly sampled the training data set at rate of 30% and 10%, and created two cell models, two CARTs, and two logistic regression models. The testing data set are still the same. Charts 3,4, and 5 are gain charts for three types of models with different training data sets. We can see that cell model performance significantly degrade when the sample sizes are small while CART and logistic regression models perform only slightly worse. CART and logistic regression are more robust than cell model when the size of training samples is small. Using less training data can reduce the cost of data collection and improve the campaign's speed to the market.

Chart 3. Cell Models Built On Different Training Data

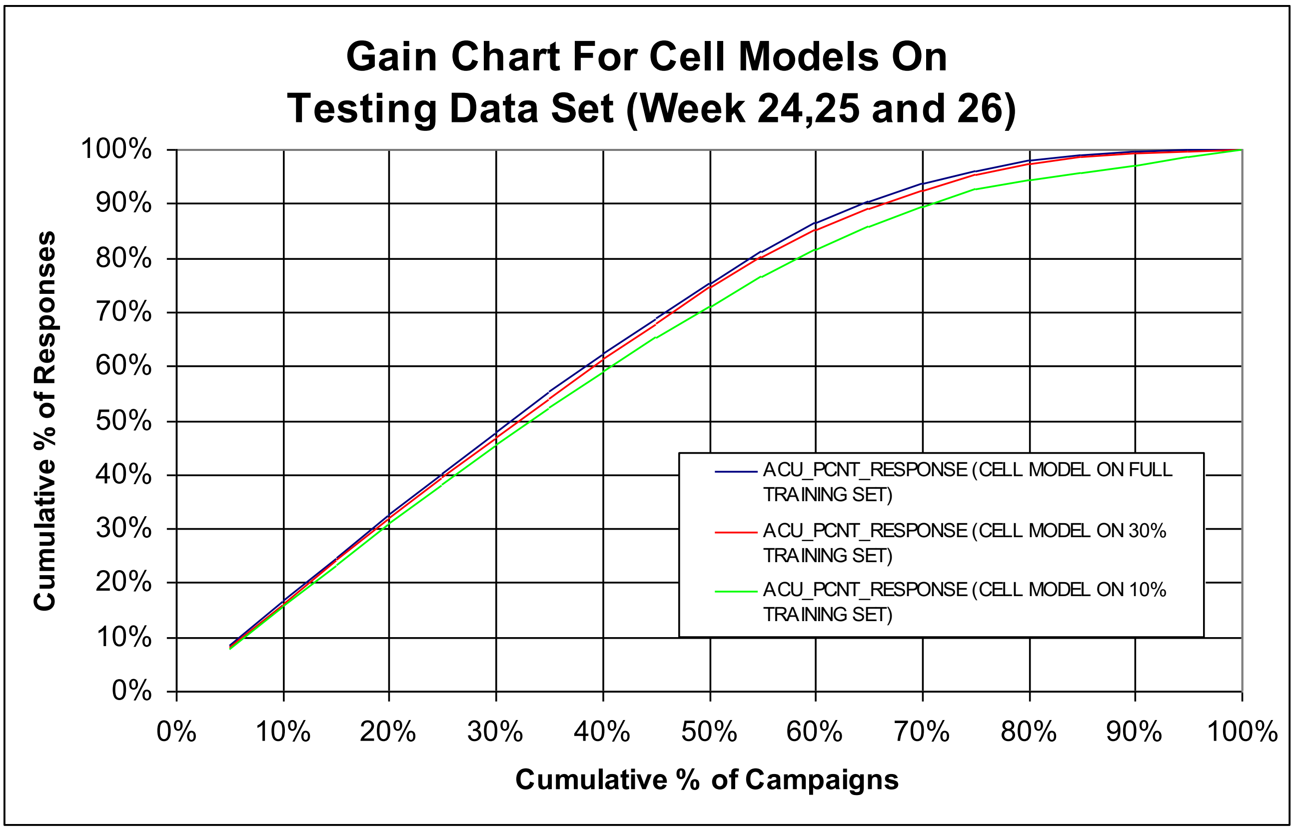


Chart 4. Logistic Regressions Built On Different Training Data

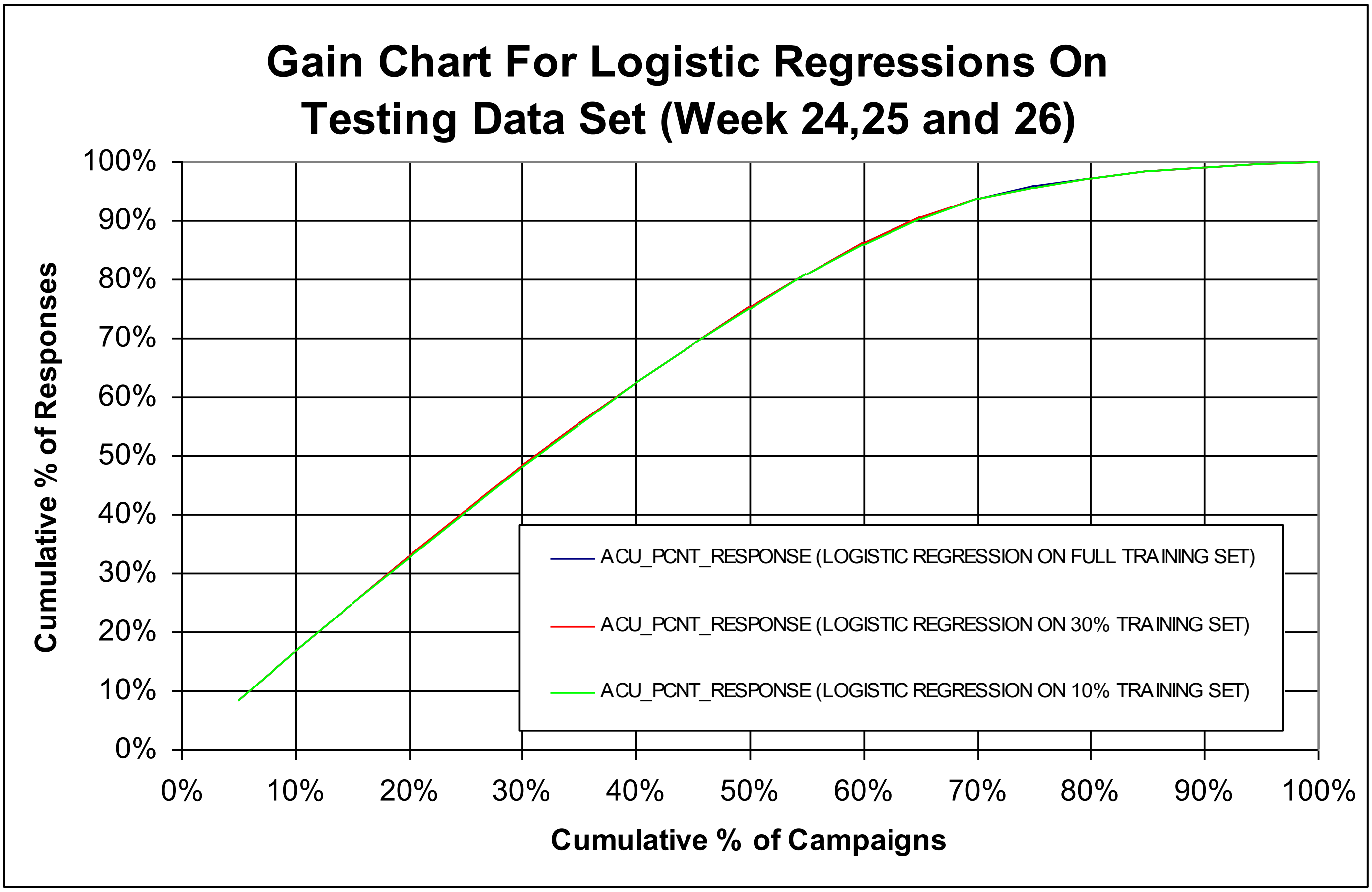
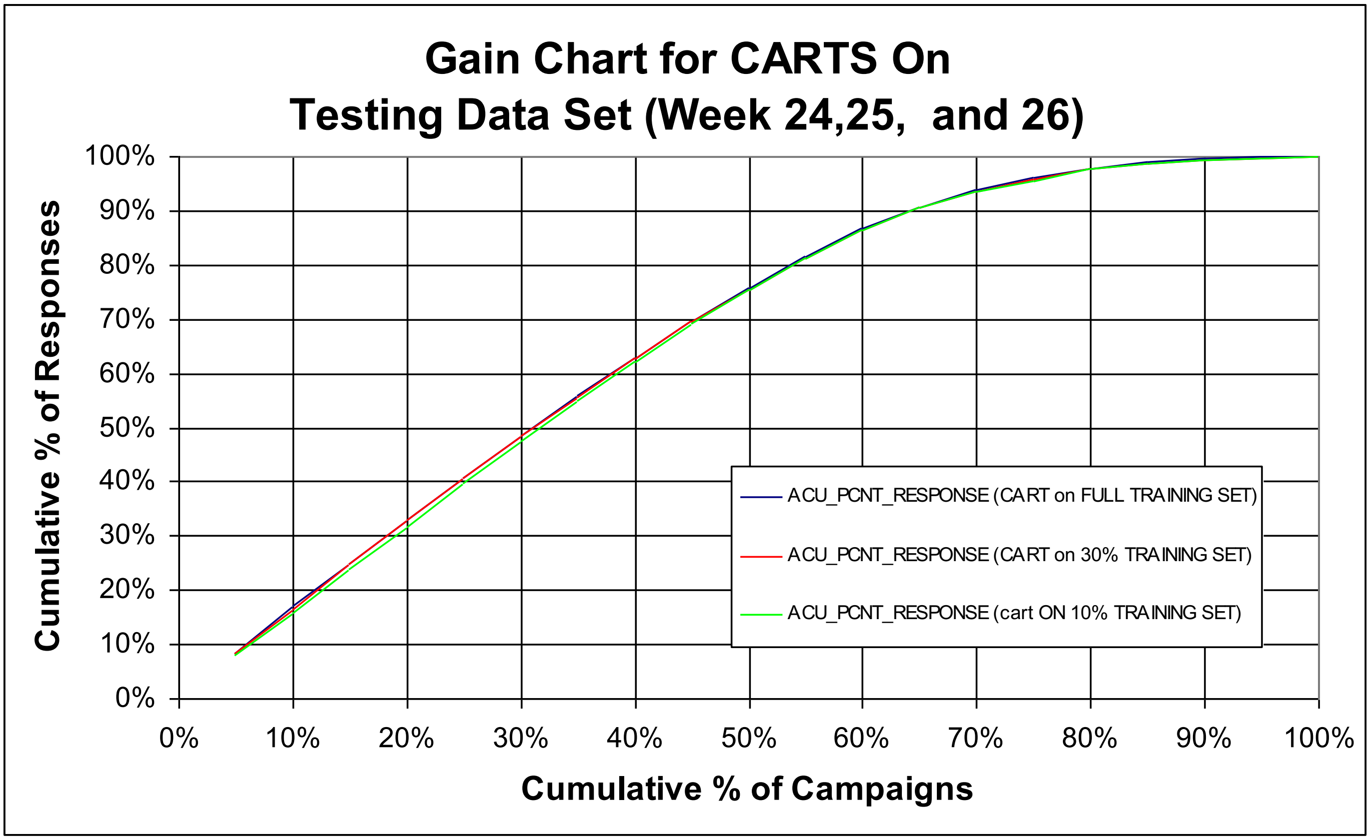


Chart 5. CARTS Built On Different Training Data



FINAL MODEL

The gradient boosting tree where used as the final model. The table 5 gives the distribution of responses and non responses in testing data set. Base on this table, I could find out the corresponding response rate for each score bucket.

Table 5. Score Bucket Distribution in Testing Data Set

Score Bucket	# of Campaign	# of Response	Low Score	High Score	Interval Response Rate	Cumulative # of Campaign	Cumulative # of Responses	Cumulative % of Campaign	Cumulative % of Response	Cumulative Response Rate	Lift
49	373	371	1.000	1.167	99.5%	373	371	1.2%	2.2%	99.5%	1.74
48	327	325	0.979	1.000	99.4%	700	696	2.3%	4.1%	99.4%	1.74
47	435	426	0.957	0.979	97.9%	1135	1122	3.8%	6.6%	98.9%	1.73
46	652	641	0.936	0.957	98.3%	1787	1763	6.0%	10.3%	98.7%	1.73
45	848	818	0.914	0.936	96.5%	2635	2581	8.8%	15.1%	98.0%	1.72
44	940	900	0.893	0.914	95.7%	3575	3481	11.9%	20.3%	97.4%	1.71
43	1023	965	0.871	0.893	94.3%	4598	4446	15.3%	26.0%	96.7%	1.69
42	1051	974	0.850	0.871	92.7%	5649	5420	18.8%	31.6%	95.9%	1.68
41	1003	909	0.828	0.850	90.6%	6652	6329	22.2%	37.0%	95.1%	1.67
40	1079	967	0.807	0.828	89.6%	7731	7296	25.8%	42.6%	94.4%	1.65
39	999	873	0.785	0.807	87.4%	8730	8169	29.1%	47.7%	93.6%	1.64
38	938	804	0.764	0.785	85.7%	9668	8973	32.2%	52.4%	92.8%	1.63
37	880	744	0.743	0.764	84.5%	10548	9717	35.2%	56.7%	92.1%	1.61
36	779	646	0.721	0.742	82.9%	11327	10363	37.8%	60.5%	91.5%	1.60
35	774	598	0.700	0.721	77.3%	12101	10961	40.3%	64.0%	90.6%	1.59
34	668	507	0.678	0.700	75.9%	12769	11468	42.6%	67.0%	89.8%	1.57
33	659	496	0.657	0.678	75.3%	13428	11964	44.8%	69.9%	89.1%	1.56
32	596	425	0.635	0.657	71.3%	14024	12389	46.7%	72.3%	88.3%	1.55
31	508	372	0.614	0.635	73.2%	14532	12761	48.4%	74.5%	87.8%	1.54
30	510	338	0.592	0.614	66.3%	15042	13099	50.1%	76.5%	87.1%	1.53
29	453	320	0.571	0.592	70.6%	15495	13419	51.7%	78.3%	86.6%	1.52
28	388	249	0.549	0.571	64.2%	15883	13668	52.9%	79.8%	86.1%	1.51
27	365	224	0.528	0.549	61.4%	16248	13892	54.2%	81.1%	85.5%	1.50
26	392	232	0.507	0.528	59.2%	16640	14124	55.5%	82.5%	84.9%	1.49
25	360	195	0.485	0.506	54.2%	17000	14319	56.7%	83.6%	84.2%	1.48
24	410	215	0.464	0.485	52.4%	17410	14534	58.0%	84.9%	83.5%	1.46
23	390	224	0.442	0.463	57.4%	17800	14758	59.3%	86.2%	82.9%	1.45
22	442	223	0.421	0.442	50.5%	18242	14981	60.8%	87.5%	82.1%	1.44
21	452	226	0.399	0.421	50.0%	18694	15207	62.3%	88.8%	81.3%	1.42
20	437	202	0.378	0.399	46.2%	19131	15409	63.8%	90.0%	80.5%	1.41
19	408	166	0.356	0.378	40.7%	19539	15575	65.1%	90.9%	79.7%	1.40
18	366	158	0.335	0.356	43.2%	19905	15733	66.4%	91.9%	79.0%	1.38
17	326	132	0.313	0.335	40.5%	20231	15865	67.4%	92.6%	78.4%	1.37
16	321	113	0.292	0.313	35.2%	20552	15978	68.5%	93.3%	77.7%	1.36
15	381	107	0.271	0.292	28.1%	20933	16085	69.8%	93.9%	76.8%	1.35
14	488	115	0.249	0.270	23.6%	21421	16200	71.4%	94.6%	75.6%	1.32
13	535	121	0.228	0.249	22.6%	21956	16321	73.2%	95.3%	74.3%	1.30

Score Bucket	# of Campaign	# of Response	Low Score	High Score	Interval Response Rate	Cumulative # of Campaign	Cumulative # of Responses	Cumulative % of Campaign	Cumulative % of Response	Cumulative Response Rate	*Lift*
12	751	188	0.206	0.227	25.0%	22707	16509	75.7%	96.4%	72.7%	1.27
11	738	167	0.185	0.206	22.6%	23445	16676	78.2%	97.4%	71.1%	1.25
10	586	102	0.163	0.185	17.4%	24031	16778	80.1%	98.0%	69.8%	1.22
9	477	66	0.142	0.163	13.8%	24508	16844	81.7%	98.3%	68.7%	1.20
8	565	69	0.120	0.142	12.2%	25073	16913	83.6%	98.7%	67.5%	1.18
7	705	71	0.099	0.120	10.1%	25778	16984	85.9%	99.2%	65.9%	1.15
6	875	68	0.077	0.099	7.8%	26653	17052	88.8%	99.6%	64.0%	1.12
5	814	49	0.056	0.077	6.0%	27467	17101	91.6%	99.8%	62.3%	1.09
4	529	11	0.034	0.056	2.1%	27996	17112	93.3%	99.9%	61.1%	1.07
3	671	10	0.013	0.034	1.5%	28667	17122	95.6%	100.0%	59.7%	1.05
2	646	4	-0.009	0.013	0.6%	29313	17126	97.7%	100.0%	58.4%	1.02
1	470	2	-0.030	-0.009	0.4%	29783	17128	99.3%	100.0%	57.5%	1.01
0	217	0	-0.099	-0.030	0.0%	30000	17128	100.0%	100.0%	57.1%	1.00

ANSWERS TO THE THREE QUESTIONS

Now I am ready to answer the three questions. Due to the cost of delivering email ACME decides to send email to only 25% of it's subscriber base for week 27.
Given the data provided:
(1) Which subscribers would you send email to?
(2) Which campaign(s) would you deliver to them?
(3) What do you expect the response rate to be?

For each subscriber, I calculated ten scores that represented ten campaign IDs in week 27 using gradient boosting tree. Then I selected the campaign and score corresponding to the maximum score among all of ten scores. These are the best campaign and score for that subscriber. The following table gives the gradient boosting tree scores for subscriber ID 3 and 27 for 10 campaign IDs. There best campaign IDs for subscriber 3 and 147 are campaign 10 and 1, respectively.

Table 6. Scores for Subscribers 3 and 147

(sorted by descending order of score)

Week	Subscriber ID	Campaign ID	GB Tree Score	Rank
27	3	10	0.885656	1
27	3	4	0.879846	2
27	3	7	0.874074	3
27	3	1	0.873177	4
27	3	6	0.847155	5
27	3	2	0.79252	6
27	3	5	0.749152	7
27	3	9	0.683055	8
27	3	3	0.643667	9
27	3	8	0.612419	10
27	147	1	1.10412	1
27	147	4	1.09089	2
27	147	7	1.079032	3
27	147	6	1.065686	4
27	147	10	1.060525	5
27	147	2	0.994183	6
27	147	5	0.949915	7
27	147	9	0.889205	8
27	147	3	0.850059	9
27	147	8	0.797806	10

Once I got the best campaign ID and best score for each subscriber, I ranked best scores in descending order and select the top 25% subscribers. These are the subscribers we should send email to. Table 7 gives the distribution of campaign IDs in the top 25% subscribers.

Table 7. Distribution of Campaigns in the Best 25% Subscribers

Campaign ID	# of Campaign	% of Campaigns
10	703	28.1%
1	644	25.8%
4	543	21.7%
7	376	15.0%
6	234	9.4%

The table 8 gives the number of subscribers for these 2,500 in each score band. The expected number of response for each score bucket is the product of number of subscriber and response rate (Table 5) which were derived based on the testing data set, i.e., week of 24, 25, and 26. The total expected number of responses is 2451 and the expected response rate for the 2500 is 98.0%, i.e., 2451 divided by 2500.

Table 8. Expected Response for Each Score Bucket for the Best 25% Subscribers

Bucket	Low Score	High Score	# of Subscribers	Response Rate	Expected # of Responses
49	1.000	1.141	611	99.5%	607.7
48	0.979	1.000	330	99.4%	328.0
47	0.957	0.979	336	97.9%	329.0
46	0.936	0.957	453	98.3%	445.4
45	0.914	0.936	452	96.5%	436.0
44	0.901	0.914	318	95.7%	304.5

However, there is always an estimation error around these response rates (Table 5). To solve this problem, I created 20 bootstrapping sample sets of the test data and calculated the response rates for each bootstrapping sample. Then I applied 20 response rates for each score bucket to the 2500 subscribers and found out that the average expected response rate is 97.97% with a standard error 0.22%. And these are the best answers I got from this study given the time constraint.

CONCLUSIONS

In this study, I analyzed 10,000 subscribers' responses to 260,000 campaigns. More than 10 derived variables were developed. The data from first 23 weeks were used as training set based on which 10 models were created. The models included simple cell models, logistic regressions, decision tree, and general gradient boosting tree. The remaining 3 weeks were used as testing set to compare the performances of these models. Gain charts and lift curves were calculated to compare the models. The gradient boosting tree provided the best performance and was used as the final model to provide answers to the above questions. Appendix B gives the top 30 subscribers and the campaigns that should be sent to them. The list of full 25% subscribers and their campaigns is included in a separate text file. The expected response rate for the best 25% subscribes in week 27 is 97.97% with a standard error 0.22%

Further studies Suggested

More models such as neural nets should be tried. The splitting of the training and testing sets based on subscriber IDs also makes sense if we intend to apply the models to new subscribers that are not included in the data. Other data sources could be added to improve the prediction accuracy. We may also use all 26 weeks data to develop model and apply it to week 27. Instead of predicting response/non response, the actual money amount that each subscriber spent may also be used as target variable. It is worthy trying more feature calculation algorithms.

APPENDIX A. DISTRIBUTION OF THE DATA

```
------------------------------
frequency var1
   10000 01
   10000 02
   10000 03
   10000 04
   10000 05
   10000 06
   10000 07
   10000 08
   10000 09
   10000 10
   10000 11
   10000 12
   10000 13
   10000 14
   10000 15
   10000 16
   10000 17
   10000 18
   10000 19
   10000 20
   10000 21
   10000 22
   10000 23
   10000 24
   10000 25
   10000 26
------------------------------
frequency var3
    53586 A
    77116 B
    64974 C
    64324 D
------------------------------
frequency var4
     5538 1
     4862 2
     5278 3
     5226 4
     4706 5
     5616 6
     4992 7
     5824 8
     5382 9
    5044 10
    5408 11
    5486 12
    5746 13
    5070 14
    5694 15
    5356 16
    4966 17
    5122 18
    5252 19
    5148 20
    5408 21
    5356 22
    4784 23
    5434 24
    5772 25
    5330 26
    5200 27
    4888 28
    5746 29
    5096 30
    4576 31
    4966 32
    5200 33
    4472 34
    5018 35
    4914 36
    5148 37
    5070 38
    5200 39
```

```
 5122 40
 4914 41
 5252 42
 5408 43
 4810 44
 5226 45
 5330 46
 5486 47
 5512 48
 5304 49
 4342 50
-------------------------------
frequency var5
143182 F
116818 M
-------------------------------
frequency var6
 26000 1
 26000 2
 26000 3
 26000 4
 26000 5
 26000 6
 26000 7
 26000 8
 26000 9
26000 10
-------------------------------
frequency var7
149642 0
110358 1
```

APPENDIX B. TOP 30 SUBSCRIBERS AND THEIR CAMPAIGNS FOR WEEK 27

Subscriber ID	Campaign ID	Score
877	1	1.140979
4178	6	1.137126
5238	6	1.134669
4345	10	1.130894
7213	6	1.123973
6963	6	1.122848
7747	6	1.121901
650	6	1.116828
12658	10	1.114781
2788	10	1.113119
597	7	1.110993
2085	6	1.109465
9148	7	1.108347
147	1	1.10412
7710	7	1.099032
5618	10	1.098568
8082	6	1.097883
9007	7	1.097683
13120	10	1.094874
5821	6	1.093732
3878	6	1.091608
5055	10	1.091507
5006	6	1.090796
4049	1	1.088167
5514	6	1.086374
12678	1	1.086036
5904	7	1.085526
3061	10	1.085353
960	1	1.080537
7592	10	1.080259