

Customer Response Prediction



By: Bailey Paulk, Madhuparna Deb, Martin Vincent,
Bhavika Bhandari, Naymal Mehmood

The Grocery Industry



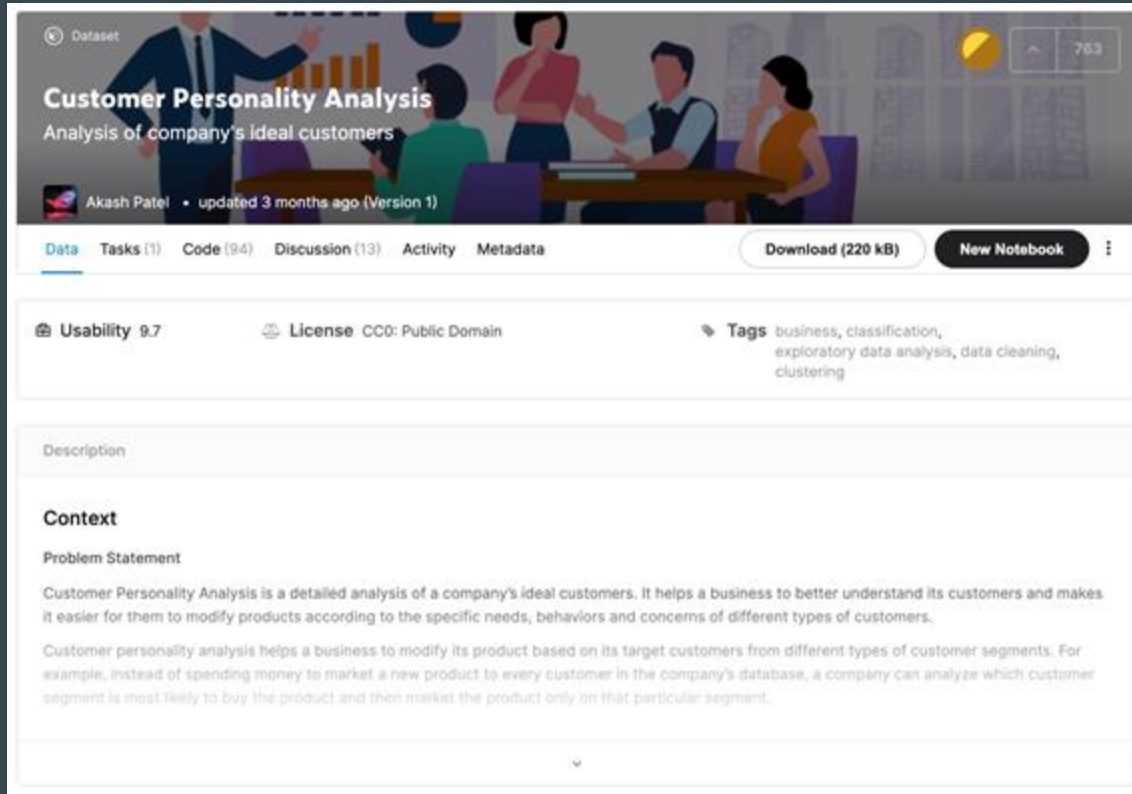
- Grocery stores rely heavily on ad campaigns and crafting a brand and reputation
- Ad campaigns can increase profits widely, but first we must answer one important question...

*Source



What customers should be sent an advertising campaign?

Where We Got Our Data



The screenshot shows the Kaggle dataset page for 'Customer Personality Analysis'. The header features a banner image of people in a meeting, the dataset title, a subtitle 'Analysis of company's ideal customers', and the creator 'Akash Patel' with a note 'updated 3 months ago (Version 1)'. Below the header is a navigation bar with tabs for 'Data', 'Tasks (1)', 'Code (94)', 'Discussion (13)', 'Activity', and 'Metadata'. To the right of the tabs are buttons for 'Download (220 kB)' and 'New Notebook'. Below the navigation bar is a section with 'Usability 9.7', 'License CC0: Public Domain', and 'Tags' including 'business', 'classification', 'exploratory data analysis', 'data cleaning', and 'clustering'. The main content area is titled 'Description' and contains a 'Context' section with a 'Problem Statement'. The problem statement explains that the dataset is a detailed analysis of a company's ideal customers, helping businesses understand their customers and modify products accordingly. It also mentions that the analysis helps in modifying products based on target customer segments, providing an example of how a company can analyze which customer segment is most likely to buy a product and then market it only to that segment.

Dataset

Customer Personality Analysis

Analysis of company's ideal customers

Akash Patel • updated 3 months ago (Version 1)

Data Tasks (1) Code (94) Discussion (13) Activity Metadata

Download (220 kB) New Notebook

Usability 9.7 License CC0: Public Domain Tags business, classification, exploratory data analysis, data cleaning, clustering

Description

Context

Problem Statement

Customer Personality Analysis is a detailed analysis of a company's ideal customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviors and concerns of different types of customers.

Customer personality analysis helps a business to modify its product based on its target customers from different types of customer segments. For example, instead of spending money to market a new product to every customer in the company's database, a company can analyze which customer segment is most likely to buy the product and then market the product only on that particular segment.

*Source

Predicting Likelihood of Campaign Acceptance

- Creating a Model based on Logistic Regression
- Why Logistic Regression?
- We had to decide what variables were important



*Source

Variables Being Used

Recency : Number of days since customer's last purchase

MntWines : Amount spent on wine in last 2 years

MntFruits : Amount spent on fruits in last 2 years

MntMeatProducts : Amount spent on meat in last 2 years

MntFishProducts : Amount spent on fish in last 2 years

MntSweetProducts : Amount spent on sweets in last 2 years

MntGoldProds : Amount spent on gold in last 2 years

NumDealsPurchases : Number of purchases made with a discount

Variables Being Used

NumWebPurchases : Number of purchases made through the company's web site

NumStorePurchases : Number of purchases made directly in stores

NumWebVisitsMonth : Number of visits to company's web site in the last month

AcceptedCmp1 : 1 if customer accepted the offer in the 1st campaign, 0 otherwise

AcceptedCmp2 : 1 if customer accepted the offer in the 2nd campaign, 0 otherwise

AcceptedCmp3 : 1 if customer accepted the offer in the 3rd campaign, 0 otherwise

AcceptedCmp4 : 1 if customer accepted the offer in the 4th campaign, 0 otherwise

AcceptedCmp5 : 1 if customer accepted the offer in the 5th campaign, 0 otherwise

Response : 1 if customer accepted the offer in the last campaign, 0 otherwise

Which Variables were Important

- Interpretation of Logistic Regression
 - From the model output we can see which of our predictors are statistically significant at the 5% level
 - All the significant predictors have positive coefficients except for number of purchases made directly in stores.



Determining Customer Profitability

- Average amount spend per customer is about **\$33**
- Average profit margin for grocery store is **4%***
- Average profit per customer is $\$33 \times .04 = \mathbf{\$1.33}$
- Cost of sending each flyer is **\$0.46***
- Breakeven response rate = $\$0.46/\$1.33 = \mathbf{0.35}$

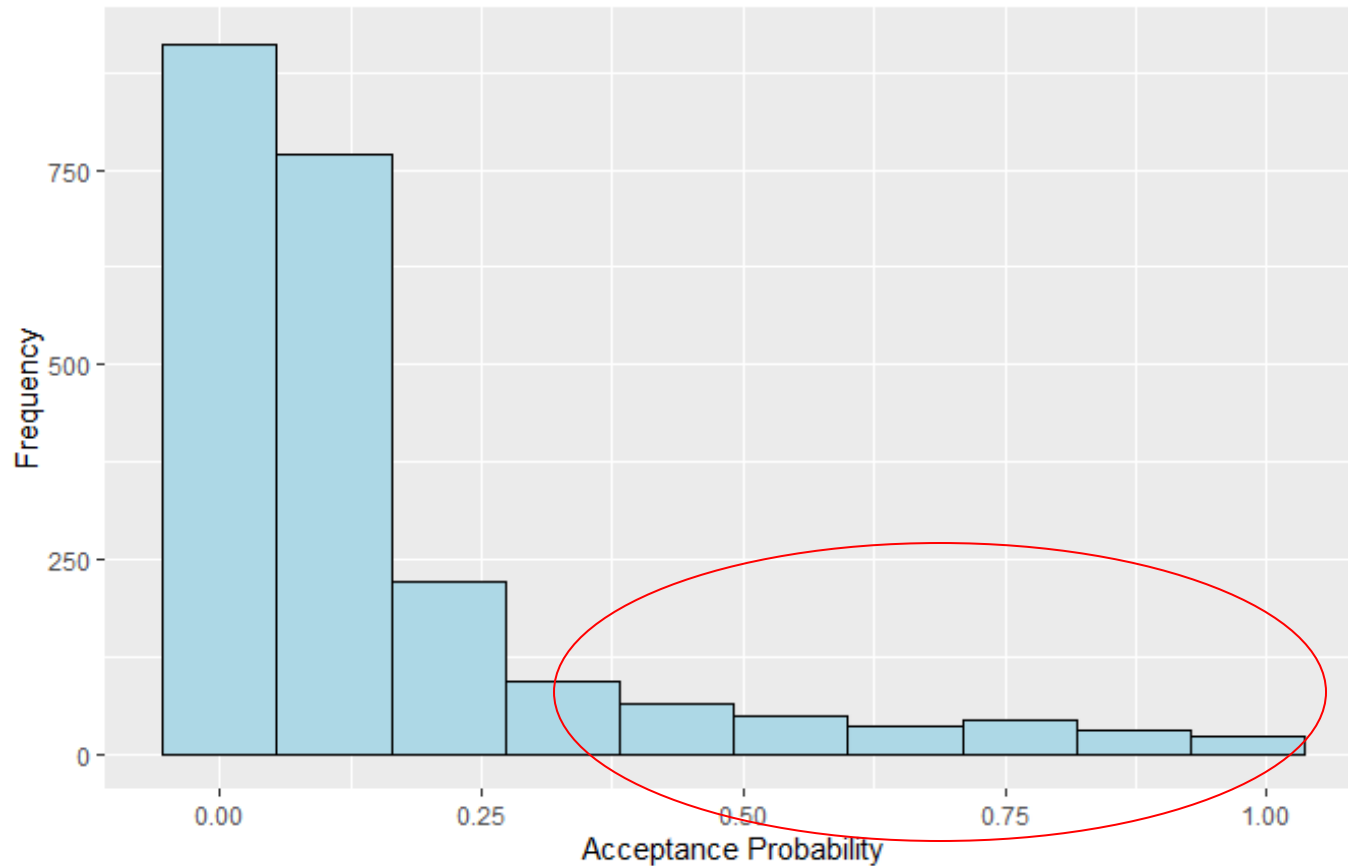


Who should be sent a campaign?

- Cutoff rate was 35%
- We filtered campaign based on acceptance probability $>$ cutoff rate
- We were left with 268 customers out of 2,240 original customers - 12% of our customers

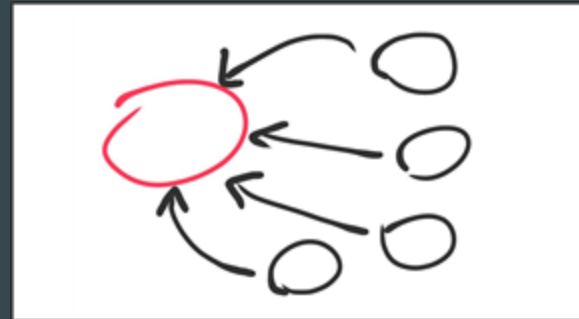


Predicted Acceptance Probabilities



In Summary

- Objective: Determine who to direct our marketing towards
- Data on customers includes demographic and Recency-Frequency-Monetary information, from which we created our model
- Assuming a 4% profit margin, we created a predicted profit per customer and a cutoff rate of 35%
- We should send advertisements to 268 customers, 12% of the customers we analyzed



Thank You!

...

1. Database
2. Creating Total Visits and Total Spent Variables
3. Logistic Regression - Model + Output
4. Calculating Acceptance Probability
5. Acceptance Probability Distribution
6. Filtering Table based on Visits, Avg Profit, and Cutoff/
Breakeven Response Rate
7. Filtered Customers List

```

10
11 #Import database
12 mkt_campaign <- read.csv("marketing_campaign.csv", sep = '\t')
13

```

mkt_camp... 2240 obs.

Final project showing which variables

campaign_customers

mkt_campaign

ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Di_Customer	Recency	NumVisits	MedFruits	MedWineProducts	MedFishProducts	MedSweetProducts
1	1974	Graduation	Single	18138	0	0	04-09-2012	18	033	88	146	172	88
2	2174	1954	Graduation	46344	1	1	08-03-2014	18	11	1	6	2	1
3	4141	1955	Graduation	71811	0	0	21-08-2012	26	426	49	127	111	21
4	6182	1984	Graduation	26846	1	0	10-02-2014	26	11	4	20	19	9
5	5124	1981	PhD	16210	1	0	15-01-2014	94	173	40	118	46	27
6	7446	1917	Master	62513	0	1	09-09-2013	18	120	42	98	0	42
7	865	1971	Graduation	59425	0	1	13-11-2012	14	231	65	164	50	49
8	6177	1981	PhD	13414	1	0	08-05-2013	32	76	10	16	1	1
9	4855	1974	PhD	10311	1	0	06-06-2013	19	18	0	24	3	3
10	5899	1950	PhD	5048	1	1	13-03-2014	68	28	0	8	1	1
11	1994	1983	Graduation	100	1	0	15-11-2013	12	9	5	6	0	2
12	167	1976	Reinc	7500	0	0	13-11-2012	19	6	18	13	13	1
13	2125	1959	Graduation	63013	0	0	15-11-2013	82	156	63	480	225	112
14	8180	1972	Master	19318	1	1	15-11-2013	53	219	2	12	3	5
15	2569	1987	Graduation	17923	0	0	10-10-2012	18	3	14	17	6	1
16	2114	1946	PhD	82900	0	0	24-11-2012	25	1006	22	115	59	68
17	9736	1980	Graduation	41810	1	1	24-12-2012	51	53	5	19	2	13
18	4939	1946	Graduation	37760	0	0	10-08-2012	29	84	5	38	158	17
19	6561	1949	Master	76995	0	1	28-03-2013	91	1012	80	498	0	16
20	2278	1981	2n Cycle	11812	1	0	03-11-2012	86	4	17	19	30	24
21	9160	1982	Graduation	57640	0	0	08-08-2012	41	86	2	71	69	58
22	5176	1979	Graduation	2447	1	0	06-01-2013	42	1	1	1725	1	1
23	1983	1949	PhD	18607	0	1	23-12-2012	63	867	0	86	0	0
24	4047	1954	PhD	45124	0	1	11-01-2014	8	184	0	163	21	52
25	1409	1951	Graduation	40849	0	1	18-03-2013	68	270	3	27	28	6
26	7882	1969	Graduation	18189	0	0	02-01-2013	89	6	4	25	11	12
27	2404	1976	Graduation	11319	1	1	17-05-2013	4	173	4	30	3	6
28	5251	1986	Graduation	100	1	0	20-02-2013	19	5	3	3	3	263
29	9422	1989	Graduation	18360	1	0	11-01-2013	26	38	2	42	20	21
30	1966	1983	PhD	84818	0	0	22-11-2013	96	884	100	801	21	66
31	1884	1989	Master	10979	0	0	22-05-2014	14	8	4	10	2	2
32	8913	1963	Master	18620	0	0	11-05-2013	16	112	17	44	34	22
33	1730	1970	Graduation	40548	0	1	10-10-2012	52	110	0	9	2	0
34	7373	1952	PhD	46610	0	1	29-10-2012	8	96	12	96	31	27
35	8755	1946	Master	68617	0	0	25-02-2013	4	482	34	471	119	68
36	10738	1951	Master	49089	1	1	29-08-2013	55	40	0	10	2	1
37	4339	1970	PhD	67513	0	1	11-12-2012	17	702	17	533	0	0
38	10751	1976	2n Cycle	23718	1	0	02-08-2013	76	6	3	14	15	7

Showing 1 to 38 of 2,240 entries, 34 total columns

Console

Environment History Connections

Global Environment

data

- campaign, 268 obs. of 2 variables
- logit_pk, list of 38
- mkt_camp, 2240 obs. of 34 variables
- mkt_camp, 2236 obs. of 35 variables

values

- cutoff @.345824364639596
- profit_p_ 1.3332391775245

Files Plots Packages Help View

Export

```

34
35 ###profit analysis - Martin
36 #Find total spend divided by number of visits for all customers to get avg spent per visit
37 mkt_campaign$total_visits <- mkt_campaign$NumCatalogPurchases + mkt_campaign$NumDealsPurchases +
38   mkt_campaign$NumStorePurchases + mkt_campaign$NumWebPurchases
39
40 mkt_campaign$total_spent <- mkt_campaign$MntWines + mkt_campaign$MntFruits + mkt_campaign$MntMeatProducts +
41   mkt_campaign$MntFishProducts + mkt_campaign$MntSweetProducts + mkt_campaign$MntGoldProds
42

```

accept_prob	total_visits	total_spent
0.348184304	25	1617
0.050702702	6	27
0.059503637	21	776
0.071550767	8	53
0.017592396	19	422
0.092308593	22	716
0.117026295	21	590
0.119782956	10	169
0.188723453	6	46
0.800145254	2	49
0.136094944	4	19
0.051083502	6	61
0.020745933	16	1102
0.053029289	15	310
0.079821962	5	46
0.414000775	26	1315
0.073828840	9	96
0.102081309	13	317
0.224213943	26	1782
0.019326475	8	133
0.089041148	12	316
0.825108134	43	1730
0.082398626	17	972
0.095384950	20	544
0.075665697	20	444
0.020435160	8	75
0.217639677	14	257
0.396304930	27	637
0.043054834	9	131
0.162310884	26	1672


```
#Logistic Regression, selecting them for business purposes
logit_mktcmp <- glm(Response ~ Recency +
  MntWines + MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
  MntGoldProds + NumDealsPurchases + NumWebPurchases + NumStorePurchases +
  NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 +
  AcceptedCmp1 + AcceptedCmp2, data = mkt_campaign, family = "binomial")
summary(logit_mktcmp)
```

```
Call:
glm(formula = Response ~ Recency + MntWines + MntFruits + MntMeatProducts +
  MntFishProducts + MntSweetProducts + MntGoldProds + NumDealsPurchases +
  NumWebPurchases + NumStorePurchases + NumWebVisitsMonth +
  AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 + AcceptedCmp1 +
  AcceptedCmp2, family = "binomial", data = mkt_campaign)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3384	-0.4712	-0.3098	-0.1884	2.9563

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.8846568	0.3258079	-8.854	< 2e-16 ***
Recency	-0.0267394	0.0027120	-9.860	< 2e-16 ***
MntWines	0.0004356	0.0003064	1.422	0.15510
MntFruits	0.0025313	0.0022497	1.125	0.26052
MntMeatProducts	0.0026322	0.0003996	6.587	4.49e-11 ***
MntFishProducts	0.0002013	0.0016801	0.120	0.90461
MntSweetProducts	0.0005038	0.0020912	0.241	0.80964
MntGoldProds	0.0017214	0.0014459	1.191	0.23384
NumDealsPurchases	0.0537274	0.0363850	1.477	0.13977
NumWebPurchases	0.0739172	0.0286359	2.581	0.00984 **
NumStorePurchases	-0.1403464	0.0317461	-4.421	9.83e-06 ***
NumWebVisitsMonth	0.2071958	0.0387848	5.342	9.18e-08 ***
AcceptedCmp3	1.7675006	0.2063526	8.565	< 2e-16 ***
AcceptedCmp4	0.8188212	0.2589397	3.162	0.00157 **
AcceptedCmp5	1.6053765	0.2594646	6.187	6.12e-10 ***
AcceptedCmp1	1.1180354	0.2531659	4.416	1.00e-05 ***
AcceptedCmp2	1.2938719	0.5151584	2.512	0.01202 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1886.8 on 2239 degrees of freedom
Residual deviance: 1338.8 on 2223 degrees of freedom
AIC: 1372.8

Number of Fisher Scoring iterations: 6

> |

29

30 #Prediction

31 mkt_campaign\$accept_prob<- predict(logit_mktcmp, newdata = mkt_campaign, type = "response")

32 mkt_campaign\$accept_prob

33

> mkt_campaign\$accept_prob

```

[1] 0.348184304 0.050702702 0.059503637 0.071550767 0.017592396 0.092308593 0.117026295 0.119782956 0.188723453 0.800145254
[11] 0.136094944 0.051083502 0.020745933 0.053029289 0.079821962 0.414000775 0.073828840 0.102081309 0.224213943 0.019326475
[21] 0.0890041148 0.825108134 0.082398626 0.095384950 0.075665697 0.020435160 0.217639677 0.396304930 0.043054834 0.162310884
[31] 0.057247463 0.026566562 0.109269233 0.159385694 0.289668501 0.045177720 0.024245574 0.037375071 0.009116585 0.380841722
[41] 0.482350388 0.025364280 0.091374974 0.008328836 0.081936731 0.015750345 0.180073775 0.036755636 0.013618392 0.156364544
[51] 0.022963316 0.255547496 0.049303774 0.639036441 0.021358503 0.122050745 0.082723591 0.029612576 0.037350523 0.053861104
[61] 0.526921258 0.032786859 0.139069667 0.096505253 0.190950716 0.028208564 0.170746533 0.778790207 0.094278940 0.077489626
[71] 0.255765993 0.145877424 0.148383037 0.063609227 0.013499825 0.026831434 0.550089089 0.664324487 0.051142030 0.629766814
[81] 0.031433228 0.154948130 0.024502925 0.026566562 0.035794776 0.013642357 0.134400447 0.631464977 0.239110669 0.007496877
[91] 0.436386353 0.059279947 0.009682525 0.028542916 0.016639203 0.068882331 0.084216844 0.073118707 0.113242502 0.061251638
[101] 0.208178662 0.051571864 0.056393957 0.053811159 0.136122002 0.009812439 0.151390520 0.053246237 0.140748217 0.116044067
[111] 0.041624162 0.554537796 0.178470771 0.061916660 0.179783842 0.083438080 0.848920095 0.120521427 0.014466826 0.059937650
[121] 0.021854597 0.100621009 0.034840326 0.019909887 0.237149463 0.007704165 0.311990188 0.037986720 0.084556761 0.015607482
[131] 0.163751900 0.158636444 0.130385943 0.016789094 0.080808578 0.180069685 0.094087229 0.063756296 0.248613491 0.028491833
[141] 0.131500486 0.238024429 0.051006275 0.064687038 0.211532047 0.134090670 0.046096918 0.065895965 0.035711529 0.014780467
[151] 0.017124637 0.044285110 0.156845076 0.123607682 0.027918563 0.793414751 0.207984991 0.023926068 0.046153859 0.01609858
[161] 0.172408628 0.225190742 0.026252126 0.071558653 0.373238704 0.018565770 0.300537369 0.035863614 0.047685081 0.016620732
[171] 0.040167276 0.031873838 0.150883693 0.082595750 0.092245695 0.161716291 0.247557681 0.018823389 0.004548496 0.113242502
[181] 0.016628512 0.023410099 0.037263868 0.090854127 0.019193646 0.090638121 0.020049116 0.068547454 0.137581776 0.133316149
[191] 0.212021562 0.013861085 0.011142698 0.024488761 0.065844215 0.127380659 0.007527216 0.276734284 0.256604083 0.276005071
[201] 0.096948961 0.044782857 0.098140362 0.532261110 0.027916578 0.452386318 0.139690318 0.067126010 0.011437866 0.849798704
[211] 0.083041926 0.055650197 0.056924858 0.161666026 0.025226600 0.081855448 0.195267537 0.369411803 0.110587535 0.070222778
[221] 0.006435147 0.038329739 0.065460014 0.090059854 0.095602143 0.082066005 0.065259030 0.109579077 0.065142848 0.042437084
[231] 0.124373727 0.787615511 0.043695956 0.028904144 0.029068233 0.041534290 0.050325854 0.024607038 0.058078616 0.057193069
[241] 0.093006242 0.028119531 0.044159013 0.441035950 0.199777701 0.063542512 0.871225063 0.063252966 0.125712484 0.123822442
[251] 0.053859966 0.038260476 0.946305439 0.015654176 0.052130413 0.058206291 0.223063178 0.006749493 0.010511779 0.118875779
[261] 0.210948627 0.067406445 0.014941103 0.122149277 0.262679939 0.109095287 0.260206934 0.243689522 0.096798835 0.135853834
[271] 0.362451001 0.088779344 0.135885657 0.037284184 0.072658305 0.007297173 0.063844390 0.082725260 0.166486136 0.043685592
[281] 0.061703524 0.065259030 0.102081309 0.010724371 0.078841827 0.034461843 0.053954971 0.026313156 0.315531340 0.076624810

```

```
43 mkt_campaign_filtered <- filter(mkt_campaign, total_visits!=0)
44
45 mkt_campaign_filtered$spent_per_visit<- mkt_campaign_filtered$total_spent/mkt_campaign_filtered$total_visits
46
47 mean(mkt_campaign_filtered$spent_per_visit)
48
49 #avg profit assuming a 4% profit margin -
50 profit_per_person <- mean(mkt_campaign_filtered$spent_per_visit) * .04
51
52 #assuming cost per letter sent to a person is $0.46, then breakeven response rate is:
53 cutoff <- .46/profit_per_person
54
```

```
> mean(mkt_campaign_filtered$spent_per_visit)
[1] 33.33098
> profit_per_person <- mean(mkt_campaign_filtered$spent_per_visit) * .04
> profit_per_person
[1] 1.333239
```

```
> cutoff <- .46/profit_per_person
> cutoff
[1] 0.3450244
```

```
#filter out customers who have accept probability greater than breakeven rate
campaign_customers <- filter(mkt_campaign, mkt_campaign$accept_prob > cutoff)
campaign_customers <- campaign_customers[,c("ID", "accept_prob")]
```

Final project showing which custo... campaign_customers

Filter

	ID	accept_prob
1	5524	0.3481843
2	5899	0.8001453
3	2114	0.4140008
4	5376	0.8251081
5	5255	0.3963049
6	2968	0.3808417
7	8601	0.4823504
8	2225	0.6390364
9	6853	0.5269213
10	9369	0.7787902
11	1859	0.5500891
12	7503	0.6643245
13	1618	0.6297668
14	4452	0.6314650
15	8996	0.4363864
16	7431	0.5545378
17	1592	0.8489201
18	10240	0.7934148
19	8475	0.3732387
20	2798	0.5322611
21	624	0.4523863
22	380	0.8497987
23	6274	0.3694118
24	5341	0.7876155
25	9529	0.4410360
26	2176	0.8712251
27	10089	0.9463054
28	2379	0.3624510
29	9262	0.4521319
30	4543	0.5520798

Showing 1 to 30 of 268 entries, 2 total columns