
Digital Marketing Analytics

Group Homework 1

Introduction

In this report, we inspect the dataset containing multi-channel sales campaigns and sales of a gifts company. The objective is to investigate the customer purchase behaviours pertaining to different marketing channels. In Part A, analytical questions will be raised and the dataset is investigated to identify interesting patterns. Part B segregates customers into traditional RFM segments and the response rates and revenues for each RFM cell are presented. Finally, Part C presents an alternative segmentation approach that could be used for more effective targeted marketing. For all parts of the report, we focus on comparing the effectiveness (and revenue) of catalogue and email marketing.

Response Rate Calculation

In this report, we will be comparing the expected response rates and expected revenue per customer order for each RFM cell. Their calculations involve a series of steps and are listed below:

- Individual Response Rate of each customer i : In this study, the response rate, also called effectiveness, for a customer i is defined as follows

$$\text{Individual Response Rate} = \frac{\text{numbers of orders made by customer } i \text{ within 30 days of contacts}}{\text{numbers of contacts made to customer } i}$$

- The individual response rate is capped at 1. If a customer has made more orders (within 30 days of contacts) than the total number of contacts, his/her individual response rate is capped at 1.
- To calculate the expected response rate for a group of people, the average of the individual response rates is taken across all the customers within the group.

$$\text{Expected Response Rate} = \frac{\sum \text{individual response rate of customer in the group}}{\text{number of customers in the group}}$$

Expected Catalogue and Email Response Rate Calculation

The formulae above can be extended to calculate the expected catalogue and email response rates. According to the dataset description document, catalogue channel orders are ordered through phone or mail whereas email channel orders are processed through Internet.

The calculation of the expected catalogue response rate is as below:

- Individual Catalogue Response Rate: For a customer i , his/her individual catalogue response rate is calculated by the below formula. Note that only the phone or mail orders placed within 30 days of contact (by catalogue), are included as the numerator. Product promotion typically lasts for 30 days and hence we attribute the channel orders to a particular contact within the 30-days timeframe.

$$\begin{aligned} &\text{Individual Catalogue Response Rate} \\ &= \frac{\text{numbers of phone/mail orders made within 30 days of catalogue contacts}}{\text{numbers of catalogue contacts made to customer } i} \end{aligned}$$

- After calculating the individual catalogue response rate, the average is taken for the entire pool of customers who have been contacted by catalogues, to derive the expected catalogue response rate:

$$\text{Expected Catalogue Response Rate} = \frac{\sum \text{Individual catalogue response rate}}{\text{number of customers contacted by catalogues}}$$

The calculation of the expected email response rate is similar:

- Individual Email Response Rate: For a customer i , his/her individual email response rate is calculated by the below formula. For the same reason, the Internet orders placed must be within 30-days of email contacts.

$$\begin{aligned} \text{Individual Email Response Rate} \\ = \frac{\text{numbers of internet orders made within 30 days of email contacts}}{\text{numbers of email contacts made to customer } i} \end{aligned}$$

- The average of all the individual email response rate is then taken to form the expected email response rate:

$$\text{Expected Email Response Rate} = \frac{\sum \text{Individual email response rate}}{\text{number of customers contacted by emails}}$$

Expected Revenue per Order Calculation

In this report, we will also be comparing the expected revenue per order for each RFM cell. The steps involved in calculating it is similar to response rate calculation:

- Individual average revenue per order of each customer i : A customer may make more than one order in a period and the average revenue per order, from a customer i is defined as follows:

$$\text{Individual Average Revenue Per Order} = \frac{\sum \text{value of orders made within 30 days of contacts}}{\text{numbers of orders by customer } i}$$

- To calculate expected revenue per customer order for a group of customers (say customers from the same RFM cell), the average of the individual values is taken across all the customers within that group.

$$\begin{aligned} \text{Expected Revenue Per Customer Order} \\ = \frac{\sum \text{Individual Average Revenue Per Order of customer in the group}}{\text{number of customers in the group}} \end{aligned}$$

Expected Catalogue and Email Revenue per Order Calculation

The calculations for expected catalogue and email revenue per order are given below.

- Individual Average Catalogue Revenue per Order: As before, only the orders that are placed within 30 days of contacts are included.

$$\begin{aligned} \text{Individual Average Catalogue Revenue Per Order} \\ = \frac{\sum \text{value of phone/mail orders made within 30 days of catalogue contacts}}{\text{numbers of phone/mail orders by customer } i} \end{aligned}$$

- Expected Catalogue Revenue Per Customer Order:

Expected Catalogue Revenue Per Customer Order

$$= \frac{\sum \text{Individual average catalogue revenue per order}}{\text{number of customers contacted by catalogues}}$$

- Email Average Revenue Per Order:

Individual Average Email Revenue Per Order

$$= \frac{\sum \text{value of Internet orders made within 30 days of email contacts}}{\text{numbers of Internet orders by customer } i}$$

- Expected Email Revenue Per Customer Order:

$$\text{Expected Email Revenue Per Customer Order} = \frac{\sum \text{Individual average email revenue per order}}{\text{number of customers contacted by emails}}$$

Part A: Analytical Questions

A total of five questions were looked into. Each question is divided into four parts: what the question aims to look into, what motivated us to look into that specific topic, a presentation of the data results, and a final part that summarises the main findings and recommendation.

Question 1

Which channel has better expected response rates and revenue per customer order, catalogue mailing or email?

Motivation

Both catalogue and Internet are important communication channels for the company. By inspecting the overall expected response rate and profitability of the respective channel, we can have an overview of the effectiveness of each channel in increasing the sales revenue. It also provides guidance to the marketing team on which marketing channel to prioritise going forward.

Data Results

	Expected Response Rate	Expected Revenue Per Customer Order
Catalogue	1.16%	\$104.09
Email	0.85%	\$80.60

Table 1 Overall Expected Response Rates and Revenue Per Customer Order

Preliminary Findings and Recommendations

- Catalogue channel orders have higher expected response rate and are also more lucrative in general. On the other hand, email channel orders tend to have a bit lower response rate and smaller in terms of expected revenue per customer order.
- Catalogue mailing should be prioritised due to its higher expected response rate and revenue per customer order. However, catalogue mailing typically has higher cost than email and it is not feasible to mail catalogues to every customer due to the limited marketing budget. For customers who typically place smaller order, we can target them with the less expensive marketing channel: email instead. The following section discusses the smart marketing approach on who to target for catalogue and email marketing.

Question 2

Does age affect the response rate to catalogue and email communication and how much is spent on average by the customers from each age group?

Motivation

Customers of different age groups might respond differently to different communication channels; younger customer might be more responsive to email communication while older customers might prefer the more traditional approach of catalogues. Identifying which communication channel each group is more responsive to can help in targeting these groups with their preferred communication channel in order to optimise budget allocation. Average spend also allows the customization of what offerings are included in the communication.

Data Results

The two age groups “65 – 74” and “75+” have been merged into a single group “65+”. The people in this age range are predominantly retirees and therefore are likely to present similar consumption habits.

	Expected Catalogue Response Rate	Expected Catalogue Revenue Per Customer Order
18 – 24	0.53%	\$93.00
25 – 34	0.50%	\$105.65
35 – 44	0.79%	\$100.73
45 – 54	1.04%	\$97.78
55 – 64	1.33%	\$94.55
65+	2.29%	\$90.16

Table 2 Expected Catalogue Response Rates and Catalogue Revenue Per Customer Order (Breakdown by Age)

	Expected Email Response Rate	Expected Email Revenue Per Customer Order
18 – 24	0.35%	\$63.14
25 – 34	1.43%	\$67.64
35 – 44	1.06%	\$77.24
45 – 54	0.79%	\$76.45
55 – 64	0.79%	\$71.51
65+	0.49%	\$85.57

Table 3 Expected Email Response Rates and Email Revenue Per Customer Order (Breakdown by Age)

Preliminary Findings and Recommendations

- Catalogue: There is a huge difference between the expected catalogue response rates in different age groups. Generally, we can see a trend where the expected response rate increases with age. The age group “65+” has an expected catalogue response rate of 2.29% whereas “18-24” only has 0.53% (more than 4 times difference). As for the expected revenue per customer order, there is little difference between age groups. From the table, it can be seen that people aged more than 54 have higher expected response rate than the overall catalogue response rate. Therefore, they should be the target of the catalogue-based campaign. If the cost of catalogue mailing (which may include design and content cost, printing cost and distribution cost) is known, the Return of Investment (ROI) could be calculated in order to identify how to best optimise the marketing budget.
- E-mail: People who are between 25 and 44 years old are most responsive to the email marketing. The age group “25 – 34” achieves the highest expected email response rate of 1.43%, this is more than four times difference compared to the lowest group response rate 0.35% (Age Group “18 – 24”). The expected email revenue per customer order is around \$60 to \$80 for most of the age groups. This is considerably smaller than catalogue mailing. In spite of this, email marketing remains an important channel due to its considerably lower cost. Email campaigns should be targeted at the groups who are most responsive to email marketing. The design and the content of emails can be adapted to these age groups. Further study on the unsubscribing rates of customers can be used to determine a suitable frequency for email marketing. We would like to send out sufficient number of emails to prospects while minimising the unsubscribing rates in order to avoid losing prospective customers.

Question 3

Is response rate to catalogue and email communication affected by the occupation of the customer, and how does the amount spent by customer differ by occupation?

Motivation

Occupation can have an effect to response rates as it affects a lot of aspects of a customer’s life; for example, whether a lot of time is spent on their laptops. It can also affect how much the customer is willing to spend as it can give an indication about the level of income of the customer. This is especially useful since the offerings

of the company are not necessities, hence heavily affected by income. Given these effects, customers of different occupations might be more responsive to different communication channels, and will be more interested in products of certain value than others. Knowing how responsive a customer is to a certain communication channel will allow the company to target them through that specific channel, while average amount spent by customer with a given occupation allows the company to customize which products are included in the advertisements.

Data Results

There are fifteen different occupation groups in the dataset. We further group them into three categories based on their income characteristics:

White Collars	Blue Collars	No Income
<ul style="list-style-type: none"> Professional - Technical Doctors Lawyers - Judges Teachers - Librarians Administrative/Mgmt/Supervisors Sales – Service Clerical - White Collar 	<ul style="list-style-type: none"> Craftsman - Blue Collar Farm Unskilled Service Workers Military 	<ul style="list-style-type: none"> Student Retired Homemaker

Table 4 Further Grouping of Occupations

White collar workers generally have, on average, the highest incomes (compared to the other two groups) as the jobs often require specialised skills. Blue collar workers have lower incomes as their job requires less specialised skills and are generally associated with lower education qualification.

“No Income” refers to a group of people who are not actively earning an income. Students may have limited purchasing power whereas retirees may have substantial cash savings after a long career.

	Expected Catalogue Response Rate	Expected Catalogue Revenue Per Customer Order
White Collars	1.30%	\$95.31
Blue Collars	0.80%	\$81.11
Not Working	2.04%	\$83.38

Table 5 Expected Catalogue Response Rates and Revenue Per Customer Order (Breakdown by Occupation Groups)

	Expected Email Response Rate	Expected Email Revenue Per Customer Order
White Collars	0.68%	\$75.74
Blue Collars	0.10%	\$61.42
Not Working	0.82%	\$66.41

Table 6 Expected Email Response Rates and Revenue Per Customer Order (Breakdown by Occupation Groups)

It is important to note that, according to the dataset description document, the occupation information is provided by a data vendor company. Only around 15% of the customers have occupation records. The occupation information may also not be up-to-date (since people change jobs all the time). Hence, the analysis in this section only provides a rough indication whether income characteristic affects the purchase behaviours of customers.

Preliminary Findings and Recommendations

- White collars: White collar workers tend to spend more per customer order than other groups. Their expected response rates for both catalogue and email are moderately high compared to blue collar workers. One customised marketing approach could be to market premium products to this group of customers. The company could advertise relevant premium products which white collars workers tend to buy in the catalogue or email advertisements.

- Blue collars: Unsurprisingly, blue collar workers have low response rates to both catalogue and email based campaigns. Their expected revenue per customer order is relatively low as well. The company could potentially decrease marketing cost by reducing advertisements to this group of customers (and increase advertisements to other groups instead).
- No Income: Despite not earning an active income, this group of customers have the highest expected response rate and moderate expected revenue per customer order. Therefore, the company should continue to focus on advertising to this group of customers and market medium priced products in the advertisements.

Question 4

What is the average number of days between orders placed as a result of catalogue and email communication?

Motivation

The products offered by the company are seasonal ones, and hence customers are expected to be buying at certain points of the year. Knowing how often customers tend to buy from the company can help minimise communication cost by only sending catalogues or emails to customers when they are likely to be buying, therefore cutting costs by stopping communication at times when customers are not interested in buying.

Data Results

Order Method	Average Number of Days Between Orders
Mail	399.56
Phone	425.48
Internet	375.31

Table 7 Average Number of Days between Purchases

Two assumptions have been made in calculating the above figures:

- 1) All Internet orders are results of email marketing and all phone and mail orders are results of catalogue marketing.
- 2) Some customers only placed orders on a single day (could be a single order or multiple orders on the same day) for their first purchase. But they did not purchase from the company any more after that day. These customers were excluded from the calculation of average days between orders to avoid bringing the average number of days down by non-recurring customers.

Preliminary Findings and Recommendations

- On average, most customers only place an order once a year, regardless of the order method.
- A number of customers placed more than one order on the same day for their first purchase. However, they did not continue purchasing from the company since then. This suggests that the company is doing well in acquiring new customers, but facing trouble in retaining them.
- This may be due to the fact that the core business of the company consists of food products often purchased as gifts during Christmas season. These customers may purchase only once during festive season and switch to another provider in the following years. Additional marketing efforts (such as second purchase discount, loyalty programme) can therefore be directed towards reducing the company's churn rate. Additional cost-benefit analysis may need to be performed to analyse expected benefits of such programme.

Question 5

What is the average value of items bought as a result of catalogue and email communication?

Motivation

Knowing whether customers tend to buy higher value items through internet, phone and mail orders helps the company decide on which product to advertise in catalogue and email communication, assuming that internet orders are results of email communication and phone and mail orders are results of catalogue communication.

Data Results

Order Method	Average Item Value
Mail	\$53.85
Phone	\$53.83
Internet	\$48.88

Table 8 Average Line Item Value of Mail, Phone and Internet Orders

Preliminary Findings and Recommendations

- Items ordered as the result of catalogue advertising (through mail and phone) are on average of higher value than items ordered as the result of email advertising. However, the difference in value is not big enough to warrant the inclusion of only lower value products in email communication and higher value products in catalogue communication since it is a difference of only about \$5. Therefore, this does not warrant the company in advertising different products specific to the marketing channel.

Part B: RFM Segmentation

In this section, we will emulate the RFM exercise. In a practice, a RFM exercise is performed at a particular point in time. RFM segmentation is performed using the historical customer purchase data, and experimentation is then performed by sending sample customers advertisements and observing the response.

To better simulate this exercise, we split the dataset into two using time. They will then be used for RFM segmentation and experimentation.

Data Split

The dataset contains orders records from 2001-01-01 to 2008-01-01, and contacts records from 2005-01-06 to 2008-01-01. The goal is to find the relationship between RFM segmentation cells and response rates. Since contacts information is only available after 6 Jan 2005, the data prior to this date is discarded.

The dataset is then split into two sets. In this exercise, they are called training set (2005-01-06 to 2006-12-31), and validation set (2007-01-01 to 2008-01-01). The training set data is used for segmentation (clustering customers into different RFM cells) and initial estimation of expected response rate and revenue per customer order. The validation set is to simulate experimentation where the response rate and revenue per customer order are estimated using most recent data.

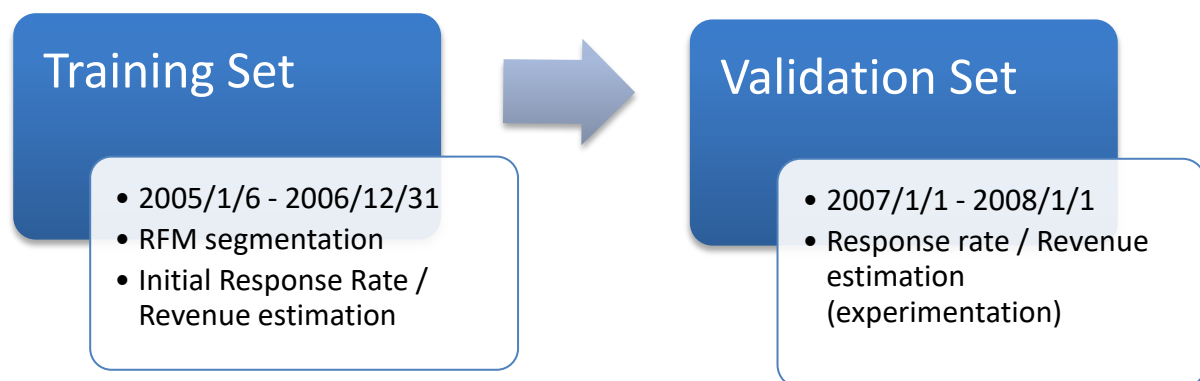


Figure 1 Data Splitting by Dates

Segmentation, Simulating Experimentation & Consistency Check

In this simulation, we separate the tasks into three different phases: segmentation, simulating experimentation and consistency check. Also note that as stated in the initial part of this report, a valid “response” is a customer order placed within 30 days of being contacted.

- **Segmentation**
 - RFM Segmentation is performed using orders data from the training set.
 - Out of the 42,576 distinct customers in the training set, only 11,597 of them exist in the validation set. Since we only have 11,597 distinct customers for RFM experimentation, RFM segmentation is performed to split these 11,597 customers into 125 RFM cells.

- Using the same training data, we obtain an initial estimate of the expected response rate and revenue per customer order for the 125 RFM cells. The purpose of collecting these results is to compare them with the validation set results to check for consistency.
- **Simulating Experimentation**
 - After segmenting the 11,597 customers into 125 RFM cells, we begin to estimate the expected response rates and revenue per customer order using the validation set. In the actual RFM exercise, we would do sample mailing. In this report, we will use the orders data between 1 Jan 2007 and 1 Jan 2008 for this experimentation simulation.
 - In the actual RFM exercise, the experimentation results will be used to formulate the targeted marketing strategies. Therefore, they will be used for drawing Lift and Gain charts in later parts of this report.
- **Consistency check**
 - We now have both training set response rate / revenue per order and the validation set estimates. We will perform an additional step in comparing them to inspect whether the purchase behaviours of each RFM cells change from time to time.

The training set and validation set results for RFM are both included in the submitted Excel file *RFMEstimates.xlsx*.

RFM Methodology

In the RFM segmentation phase, customers are clustered into 125 RFM cells based on the following criteria:

- **Recency** captures how recently customers purchase; in this case, it is measured as the minimum length of dates between 2006/12/31 and order dates for each customer. Most recent purchase will be placed under the highest quantile in R criterion.
- **Frequently** literally means how frequent do customers purchase (between 2005/1/6 and 2006/12/31); it is measured as the count of orders for each customer.
- **Monetary** captures how much customers purchase; it is measured as the average order value for each customer.

Based on the value of three attributes, customers are segmented into 5 quantiles of each attribute. 5 is the highest and 1 is the lowest. Eventually 125 segments are created of which each segment has a unique RFM code, for example, segment “555” contains the most recent, frequent and monetarily- valuable customers.

Training Set Estimate of Expected Response Rate and Revenue per Customer Order

The initial estimates are obtained using training set and the top 5 segments are displayed below (sorted by catalogue response rates).

R	F	M	Catalogue Response Rate	Catalogue Revenue Per Customer Order
4	5	5	5.54%	\$247.10
5	5	5	5.28%	\$249.70
3	5	5	4.82%	\$184.83
4	3	2	4.32%	\$38.59
4	3	3	4.29%	\$56.46

Table 9 Top 5 RFM Segments for Catalogue Response Rate (Training Set)

The top 5 segments for email response rates are:

R	F	M	Email Response Rate	Email Revenue Per Customer Order
1	5	3	6.38%	\$31.67
3	1	4	5.94%	\$79.92
1	5	5	4.06%	\$299.91
3	5	2	2.96%	\$39.93
3	5	4	2.61%	\$75.58

Table 10 Top 5 RFM Segments for Email Response Rate (Training Set)

For the catalogue response rates, the segments with high 'Monetary' value tend to have high response rates while segments with lower monetary value have comparatively low response rates.

Simulating Experimentation and Consistency check

Expected response rates and revenue per customer order are calculated using the validation set. The top 5 segments (for catalogue and email respectively) are tabulated below:

R	F	M	Catalogue Response Rate	Catalogue Revenue Per Customer Order
5	1	2	6.09%	\$74.68
4	2	3	5.21%	\$60.75
4	1	1	5.05%	\$41.69
4	1	3	4.93%	\$81.30
4	2	2	4.76%	\$67.99

Table 11 Top 5 RFM Segments for Catalogue Response Rate (Validation Set)

R	F	M	Email Response Rate	Email Revenue Per Customer Order
1	5	3	20.17%	\$48.45
2	1	5	6.25%	\$57.90
2	1	4	6.13%	\$54.44
1	2	2	5.27%	\$51.68
2	2	1	4.24%	\$59.06

Table 12 Top 5 RFM Segments for Email Response Rate (Validation Set)

As seen from the above, the validation set results are quite different from the training set results. There could be two that lead to this:

- Insufficient data used for RFM segmentation and experimentation, since we only have two-year-long data for training and one-year-long data for validation.
- The purchase behaviours of the different RFM cells change dynamically with time. Hence a top segment in year 2005 & 2006 may no longer be a top segment in year 2007. This highlights the importance where we should obtain RFM experimentation regularly to obtain the latest estimates (response rate and revenue per customer order) to tailor the marketing strategies.

Note that RFM segment 153 has the highest email response rates in both training and validation set. This shows that it is a resilient segment with constantly high receptiveness to email marketing. This segment should be highly prioritised while planning email advertising.

Lift Charts

Since the experimentation results are the baseline in formulating the marketing strategies, it will be useful to plot the lift and gain charts using the validation set results. Using the RFM clustering, the smart marketing strategy would be to mail advertisements to the cells with higher response rate first.

The lift chart shows the increase in response due to the targeted (smart) marketing versus random mailing to random set of customers. The lift charts for catalogue and email are as below:

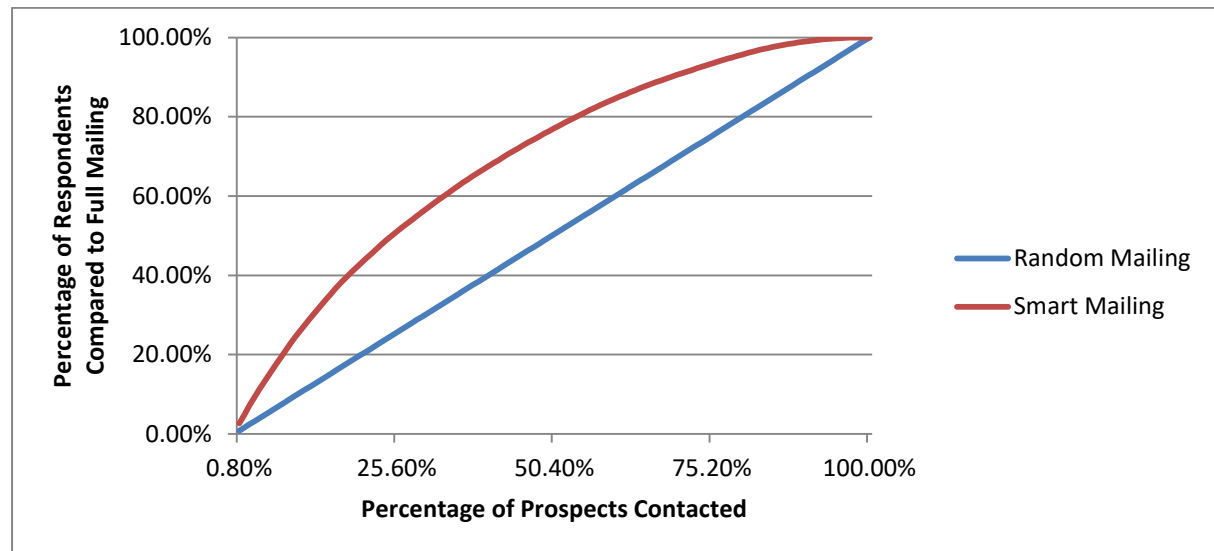


Figure 2 RFM Lift Chart (Catalogue)

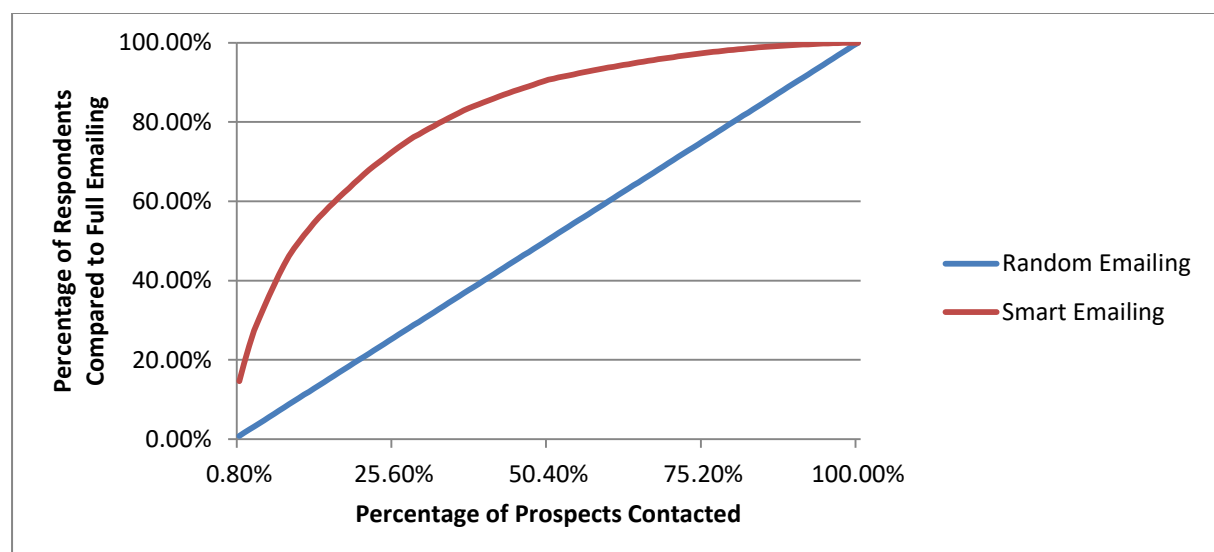


Figure 3 RFM Lift Chart (Email)

As seen from the above charts, the smart marketing curves are entirely above the random mailing curves. In other words, the targeted marketing techniques (using RFM) outperform random mailing such that there will be more respondents given a particular percentage of prospects contacted. In practice, marketing budget is limited and the company usually only contacts a small percentage of the prospective customers (say 25%). It can be seen that for 25% of prospects contacted, the smart marketing response rates more than double the random mailing strategy (for both catalogue and email).

Also note that email has a better lift than catalogue (the curve is higher), this implies that the RFM strategy is more effective in email campaigns than catalogue campaigns.

Gain Charts

Another useful chart to look at is the Gain Chart, which depicts the increase in profit due to the targeted (smart) marketing versus random mailing to random set of customers. Our smart marketing strategy is the same as before: mail to the RFM cell with highest response rate first.

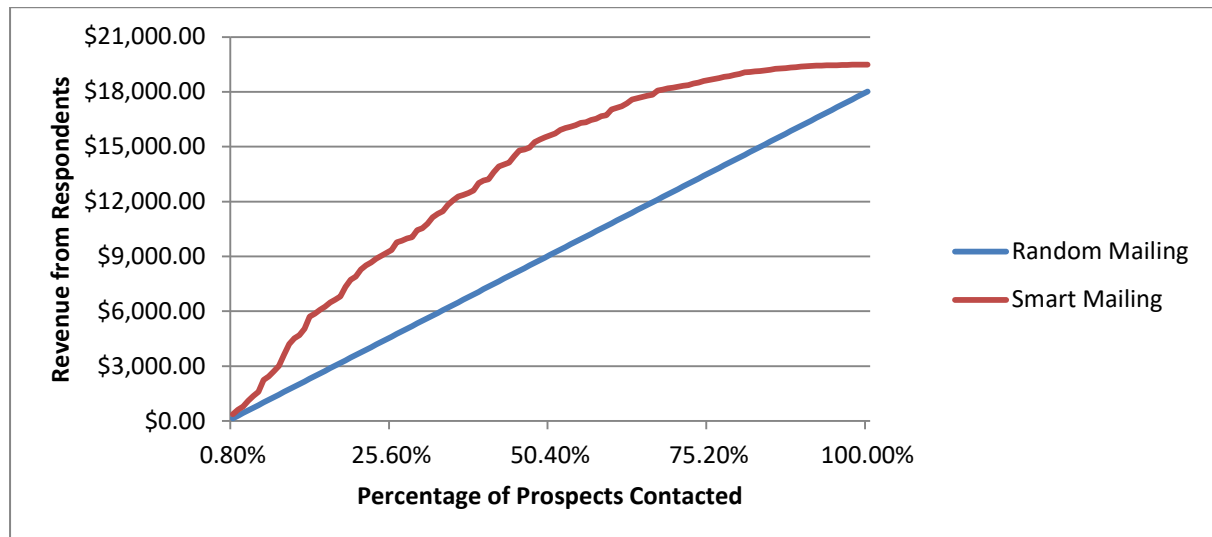


Figure 4 RFM Gain Chart (Catalogue)

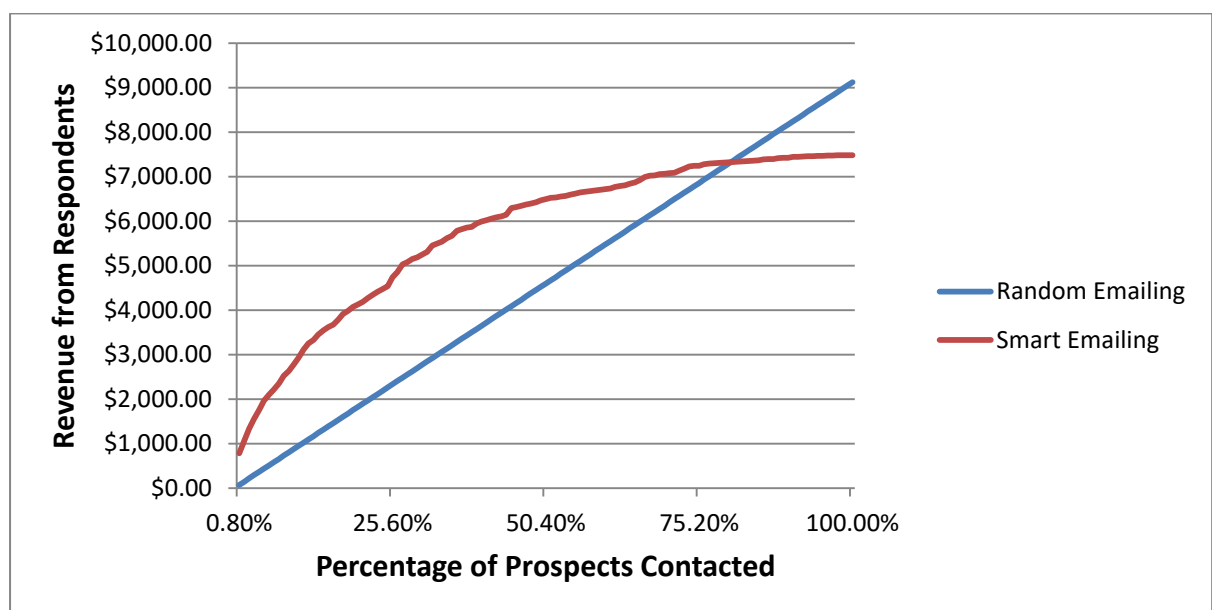


Figure 5 RFM Gain Chart (Email)

The Gain chart looks similar to the Lift chart. The only difference is that the red curves are not as “smooth”. This is because each RFM cell has its own expected revenue per customer order, and the individual values are used to calculate the final “gain” in terms of profits. In the blue curve (random mailing), a fixed average revenue per customer order is used. The different revenue per customer order used (in the two curves) also explains why the two marketing strategies achieve different revenue when 100% of the prospects are contacted.

Note that in this dataset, there is no cost associated with catalogue mailing and emails and is assumed to be zero; hence the gain chart is monotonically increasing. If the cost is accounted for in the Gain chart, we will

observe a peak in the red curve (smart mailing curve) where we could achieve the highest profits. In practice, this peak will be used to determine how many prospects we would target our marketing efforts to.

Recall that from the Lift chart perspective, email campaign fares better than catalogue. While looking at the Gain chart, we can see that catalogue achieves significantly higher profits than emails (even though email response rate is higher). This is due to the fact that catalogue orders tend to be of larger value.

Therefore, the company should use catalogue campaign to target customers who make huge purchases and email campaign to target customers who are not as responsive to catalogue campaigns.

Part C: Alternative Segmentation (RDA)

The success of RFM in the area of targeted marketing is that it segments customers into different groups where certain groups are more responsive to the company's marketing channels. The company can therefore focus their marketing efforts on the groups that are more responsive.

We could potentially segment the customers using different criteria. If these criteria are able to better segregate the customer base such that highly responsive customers tend to be grouped together, this segmentation strategy could achieve better results than RFM. In this assignment, we propose an alternative segmentation approach, RDA.

RDA Methodology & Business Justification

Recall that there are only 11,597 distinct customers who exist in both training and validation set. The three dimensions in RFM segmentation segregate the customers into 125 cells. Adding another dimension would increase the number of cells to 625. Given that there are only 11,597 distinct customers, we do not want too many cells or else the response rate / revenue estimates will not be accurate. Due to this reason, we propose an alternative segmentation strategy, RDA, which also consists of three dimensions. Nevertheless, a higher-dimensional segmentation strategy could potentially improve the performance metrics of targeted marketing. This can potentially be validated if we have a larger set of customer data.

Recency, distance and age are taken to form a similar segmentation approach (like Part B) to segregate customers into 125 cells:

- **Recency** is the same recency criterion used in RFM. Recency segregates customers based on how recently a customer makes a purchase. A more recent purchase may indicate that the customer likes the company's products and likely to respond to the communication channels.
- **Distance** measures how far a customer stays from the nearest company's store. If a customer lives far away from the company's store, he might not be motivated to purchase in-store and more inclined to purchase using catalogue or Internet. In this model, customers with smaller store distance will be placed under higher quantile of 'D'.
- **Age** captures age group of the customers. As discovered in Part A, elderly people tend to have high response rates in catalogue. Considering that the company's products are often treated as gifts, it might be more suitable to target elderly customers than young customers. In this model, customers of higher age will be placed under higher quantile of 'A'

Similar to RFM, RDA groups customers into 125 segments with unique RDA codes. For example, RDA segment "555" contains the customers who make most recent purchase, live near to company's store and of highest age quantile.

In this exercise, we perform the same RDA simulation as we did for RFM:

- Split data into two sets: training set and validation set using the same splitting strategy
- Perform RDA segmentation using the criteria discussed above
- Obtain training set response rate and revenue per order for consistency check
- Simulate experimentation using validation set
- Compare training set and validation set results for consistency check

The training set and validation set results for RDA are both included in the submitted Excel file *RDAEstimates.xlsx*.

Training Set Estimate of Expected Response Rate and Revenue per Customer Order

The initial estimates are obtained using training set and the top 5 segments are displayed below (sorted by catalogue response rates).

R	D	A	Catalogue Response Rate	Catalogue Revenue Per Customer Order
3	1	5	5.80%	\$123.75
4	1	5	4.77%	\$98.79
4	3	5	4.55%	\$108.40
5	1	5	4.41%	\$107.03
4	4	5	4.15%	\$96.12

Table 13 Top 5 RDA Segments for Catalogue Response Rate (Training Set)

The top 5 segments for email response rates are:

R	D	A	Email Response Rate	Email Revenue Per Customer Order
3	2	5	4.25%	\$58.59
3	4	1	3.86%	\$67.30
1	3	4	3.47%	\$86.23
3	2	2	1.82%	\$72.23
3	4	2	1.49%	\$68.58

Table 14 Top 5 RDA Segments for Email Response Rate (Training Set)

As we have expected, elderly people tend to be responsive to catalogue communication channel. The top 5 segments which have the highest catalogue response rate all belong to the highest A quantile: 5. Furthermore, customers who live far away from company's store seem to be more likely to order through catalogue. This is supported by the fact that 3 out of the top 5 RDA segment have their D value equal to 1.

For emails, the segregation is less distinct. However, the topmost segment "325" contains customers from the highest age group. This is interesting since it implies that elderly customers are starting to purchase via Internet.

Simulating Experimentation and Consistency check

Expected response rates and revenue per customer order are calculated using the validation set. The top 5 RDA segments (for catalogue and email respectively) are tabulated below:

R	D	A	Catalogue Response Rate	Catalogue Revenue Per Customer Order
4	2	5	5.39%	\$88.69
4	1	5	4.77%	\$123.38
4	3	5	4.53%	\$117.16
4	4	5	4.32%	\$117.32
5	2	4	3.86%	\$104.32

Table 15 Top 5 RDA Segments for Catalogue Response Rate (Validation Set)

R	F	M	Email Response Rate	Email Revenue Per Customer Order
3	5	3	6.18%	\$63.70
4	5	1	4.98%	\$72.10
1	4	1	4.79%	\$71.24
1	2	4	4.35%	\$57.69

2	5	1	4.08%	\$43.87
---	---	---	-------	---------

Table 16 Top 5 RDA Segments for Email Response Rate (Validation Set)

This time round, the catalogue response rate (and revenue per customer order) looks very similar across the two data sets. RDA segments 415, 435 and 445 are the top five segments most responsive to catalogue mailing (in both sets). This shows that our proposed RDA segmentation is resilient across timeframes and is able to identify customers who are highly responsive to catalogue marketing.

For email campaign, the top 5 segments are a lot less similar. Again, this might be attributed to the change in customers' purchase behaviours. It emphasises the importance of performing RDA segmentation regularly, in order to capture the change in customer purchase behavioural trends.

Lift Charts

Same as Part B, the lift charts of RDA is plotted using the validation set results.

The lift chart shows the increase in response due to the targeted (smart) marketing versus random mailing to random set of customers. The lift charts for catalogue and email are as below:

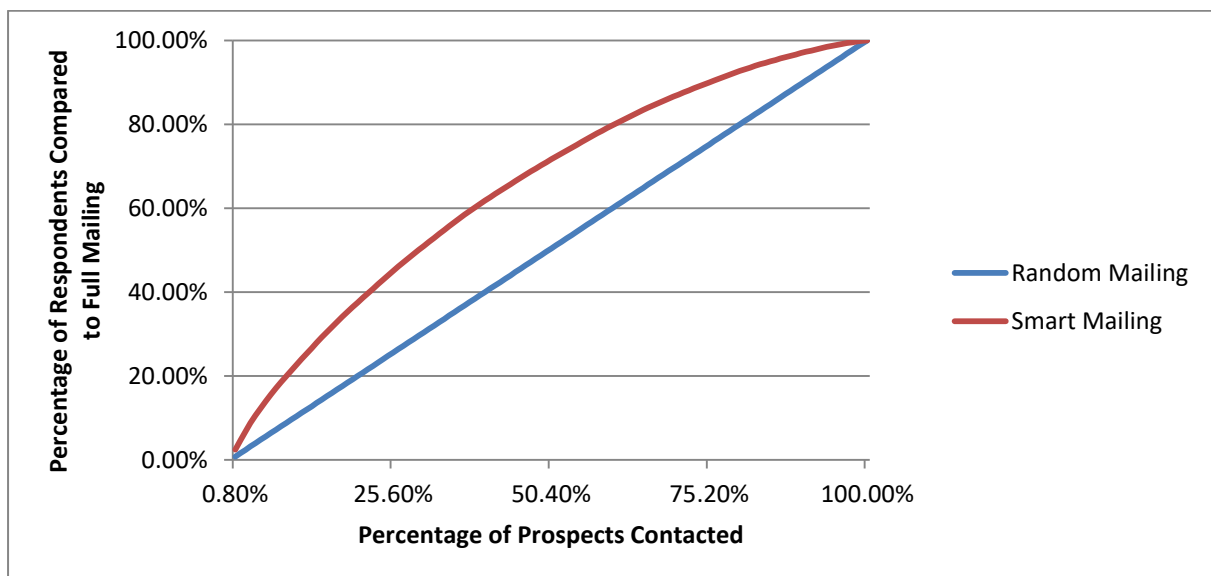


Figure 6 RDA Lift Chart (Catalogue)

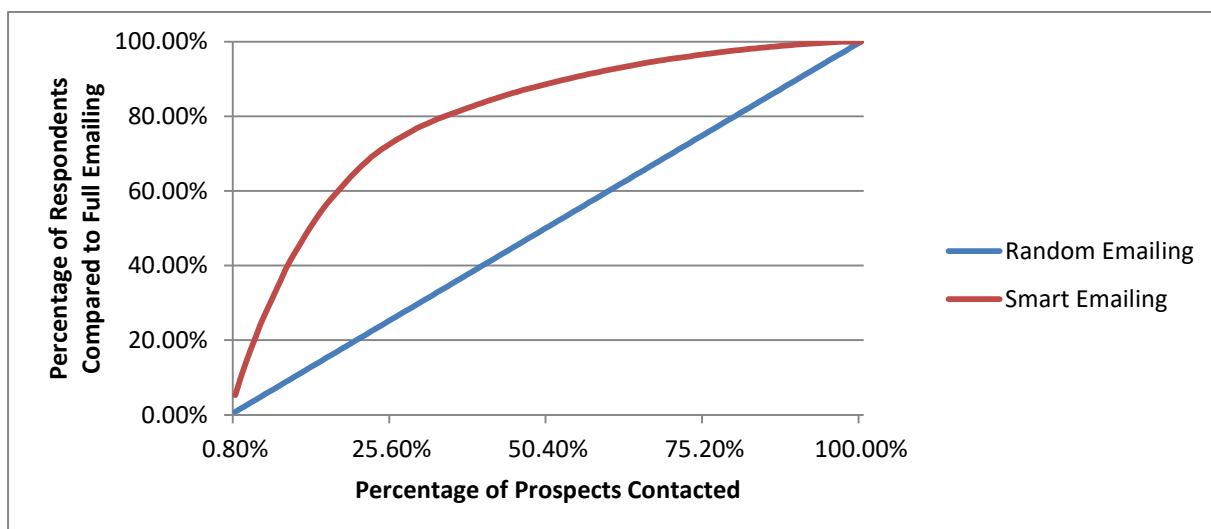


Figure 7 RDA Lift Chart (Email)

It can be seen that the RDA lift charts are extremely similar to the RFM lift charts. In other words, RDA achieves similar performance as RFM in terms of response rates. The RDA lift chart for email is also slightly higher than the RDA lift chart for catalogue. Again, this implies that RDA fares better in drawing responses using email communication channel (compared to catalogue).

Gain Charts

We will again inspect the Gain Chart. The smart marketing strategy is the same: target RDA cell with the highest response rate first, then the second highest, so forth and so on.

Part B has explained why the red curve in the Gain chart is not “smooth” and why it is monotonically increasing in this case. If we have the costs associated with catalogue mailing and emails, we could observe a peak in the red curve and formulate our marketing strategy to achieve highest possible profits.

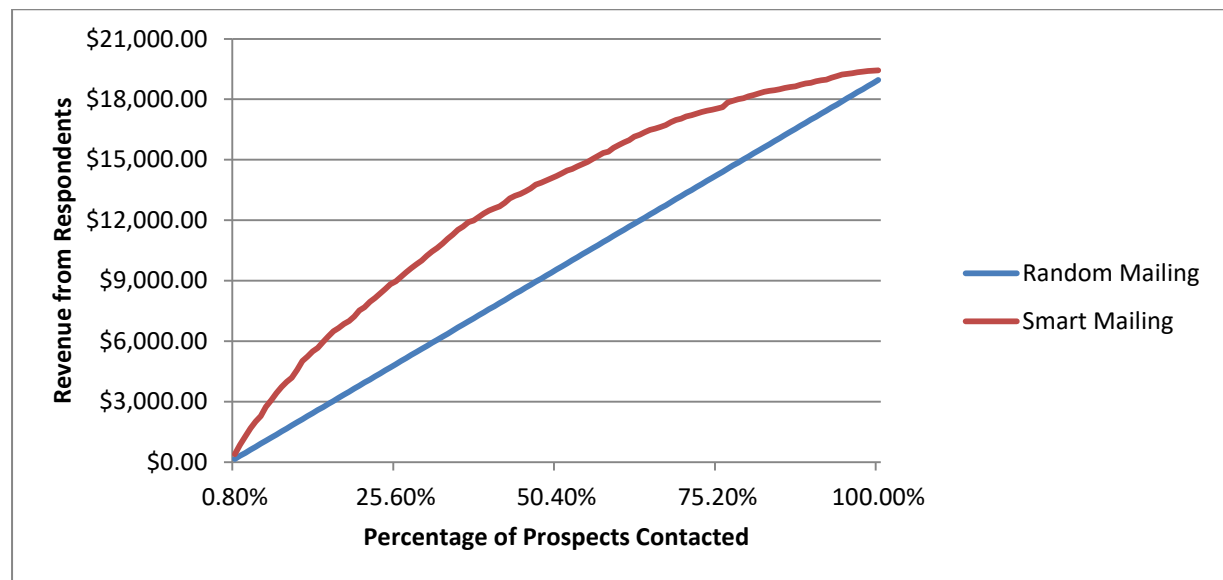


Figure 8 RDA Gain Chart (Catalogue)

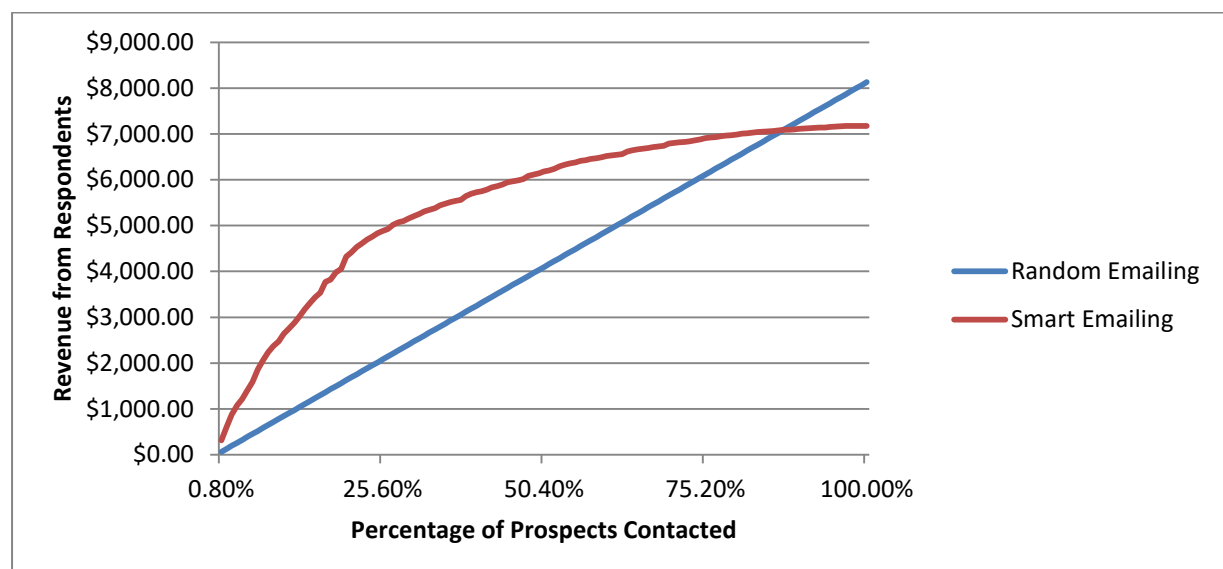


Figure 9 RDA Gain Chart (Email)

The Gain charts of RDA exhibit similar patterns as the RFM Gain charts. The achievable revenues for catalogue is between \$0 and \$20,000 and for email, it is between \$0 and \$7500. These figures are very close to the

figures obtained from RFM Gain charts. This shows that RDA is able to achieve similarly good performance as RFM in terms of response rate and revenues.