**HW1 (Group K)**

**Seongmin Lee 01247436, Otto Godwin: 01422474,**

**Qiuyue Zhang: 01378527, Samantha Tia: 01405785**

**Part A**

We raised 2 questions. We examined which channel had better response rates, catalogue mailing or email and then we split it further by analysing the data by month.

**Better Response Rate By Channel**

There are two direct marketing methods in the order tables as email and catalogue. Based on the below assumption that

1. One catalogue/email can affect the multiple orders
2. Multiple catalogs on the same date are not regarded as a single catalogue

we could also consider the maximum duration influenced by direct marketing. Hence, the date differences are divided into 7, 10, 14 days because most of the food products are re-purchased in at least a week. If an order is made for a gift, then the duration could be longer than 7 days. However, after 14 days, it is hard to linger on the consumer's memory.

The response rate for email is slightly better if the date difference between order date and contact date is within 7 days. This could be caused by the fact that catalogues have longer delivery lead time than emails, and that catalogues are real objects that play a role of reminder while emails are virtual and can be deleted or buried by other new emails. However, when we set the date differences as 10 or 14 days, the response rate by catalogues is a bit higher than the email.

**Catalog Response**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date difference** | **Total Catalog** | **Responded Catalog** | **Response Rate** |
| 7 days | 1,021,014 | 19,744 | **1.93%** |
| 10 days | 27,153 | **2.66%** |
| 14 days | 38,080 | **3.73%** |

**Email Response**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date difference** | **Total Email** | **Responded Email** | **Response Rate** |
| 7 days | 2,368,315 | 46,142 | **1.95%** |
| 10 days | 61,145 | **2.58%** |
| 14 days | 80,293 | **3.39%** |

**Analysis of Response Rate By Channel** **when split by Month**

During the process, we ranked by customer ID, contact date and order number and then considered only rows with rank 1 to remove cases where multiple catalogs were sent to the same customer in 1 day (Example: Customer ID 40193406 was sent two catalogs on 24th November 2005 and we take that as sending one catalog). The reason for this is because we assume that catalogs sent on the same day would be the same catalog and so we only considered this as one catalog sent. We would apply the same argument when analyzing emails. We considered any order made within 10 days of getting a contact as a response. This is as we assume that it takes 3 days to send a catalog and then gave customers a week to respond to it.

For response rate, we did three types of calculations based on different assumptions.

1. Multiple catalogs can contribute to one order and all are weighted as 1
2. Multiple catalogs can contribute to one order but we weighted each order as the inverse of the difference between the order and catalog date
3. Only one catalog can contribute to one order and we took the most recent catalog sent as the effective one (Prevent counting an order number multiple times as responses)

Overall, for all three ways of calculating, the performance of both catalogs and emails response rates has been pretty dismal with all response rates being less than 2% if we exclude months November and December.

The number of orders made in the top 2 months, November and December, is 32926 and 82443 respectively. The response rate of these two months are also usually the highest for both catalog and emails. Using the lines table, we found that 76.8% of the items were bought as gifts. This can be attributed to the fact that people would buy food gifts for the Christmas season. This result is backed by Packaged Facts (2016) who found that “almost 50% of people who buy food gifts for others have done so for the winter holidays, while 32% of those who have purchased food gifts for themselves have done so for these holidays”. This is also why the number of catalogs and emails sent by the firm is also very high during this time period. However, this means that we are unable to gauge whether the higher response rate for the month of November and December is due to effectiveness of the catalogs and emails sent or whether it is because customers would usually buy during this time period.

In another study by Packaged Facts (2018) found that 81% of adults celebrated Easter in 2017 and this number is almost eight times as many as those who had purchased an Easter food gift for someone. This indicates that Easter is a huge market untapped. The number of orders made in the month of April (13131) is only slightly higher than the average of that when we remove months November and December for the firm. If we could improve the effectiveness of the marketing campaigns for Easter, more profits could be made.

We also calculated transition probabilities and average purchase basket size by purchase occasion. We did not set a time frame where a customer is considered churned because food gifts tend to be bought during holidays, which would result in a range of time periods. With this, we see that the probability that a customer would buy a second time is 45.11% and the transition probability increases thereafter. This indicates a lack of brand loyalty. We should focus our efforts on retaining customers who have just bought our products.

**Part B**

We split the train and test set as those before 1st Jan 2017 and those after respectively. The RFM values were calculated using the orders table with data before 1st Jan 2017. We took the contacts data and filtered those where contact was made via catalog as mailing catalogs would cost money and not email.

Again, as in part A, we calculated based on assumptions 1 and 3.

With that, we get the following table sorted by decreasing response rates. We then made a decision to mail to all the respondents that were identified in the groups that had a response rate of more than 6%, 7%, and 8% respectively and mailed catalogs to them from 1st Jan 2017 onwards.

**Using Assumption 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Response Rate Groups** | **Number of Contacts** | **Number of Orders** | **Profits** | **ROI** |
| 8% | 59558 | 2721 | 22072 | 37.06% |
| 7% | 63936 | 2853 | 21654 | 33.87% |
| 6% | 76684 | 3203 | 19406 | 25.31% |

**Using Assumption 3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Response Rate Groups** | **Number of Contacts** | **Number of Orders** | **Profits** | **ROI** |
| 8% | 36693 | 1584 | 10827 | 29.51% |
| 7% | 63098 | 2646 | 16282 | 25.80% |
| 6% | 69236 | 2808 | 15004 | 21.67% |

If we use assumption 1, our ROI for each of the three groups is 25.31%, 33.87% and 37.06%. If we use assumption 3, our ROI for each of the three groups is 21.67%, 25.8% and 29.51%. We can see that if we want to maximize our ROI, we should only mail to customers in the RFM group that had a response rate of 8% and more, but if we want to maximize profits then for assumption 3, mailing to customers in the RFM group that had a response rate of 7% and more would be better.

**Part C**

For part C, we added age and income in addition to RFM. According to a report by Packaged Facts (2018), “the food gifting market remains buoyed by higher-household income consumers and is heavily dependent on older consumers”. Hence, age and income are dimensions to target because it is usually middle-aged customers who would buy food gifts for their friends or are in charge in the family to buy them. Also, those with higher income would have a higher propensity to spend on food gifts.

For both assumptions, we adjusted our minimum response rates for each of the three groups to send similar number of catalogs that we sent to the three groups in part B within the same time frame stated.

**Using Assumption 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Response Rate** | **Number of Contacts** | **Number of Orders** | **Profits** | **ROI** |
| 6.8% | 59841 | 2533 | 16149 | 26.99% |
| 6.5% | 63437 | 2667 | 16573 | 26.13% |
| 5.5% | 76441 | 3055 | 15209 | 19.90% |

**Using Assumption 3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Response Rate** | **Number of Contacts** | **Number of Orders** | **Profits** | **ROI** |
| 7.74% | 37094 | 1594 | 10726 | 28.92% |
| 6.50% | 63768 | 2675 | 16482 | 25.85% |
| 5.62% | 69600 | 2859 | 16170 | 23.23% |

We can see that the ROI decreased or remained similiar for part C compared to part B on the test set for all groups and both assumptions, which indicates that using both dimensions did a worse job overall. This could be due to overfitting on the training set.

# Bibliography

Packaged Facts, 2018. *Food Gifting in the U.S.: Consumer and Corporate, 6th Edition,* s.l.: Packaged Facts.

Packaged Facts, 2016. *Food Gifting in the U.S., 5th Edition,* s.l.: Packaged Facts.