Pentathlon:
Next Product to Buy Modeling

The department director's meeting had not led to the outcome Anna Quintero wanted. Neither the CMO nor the department directors were fully convinced by Quintero's survey data. She would need some direct evidence that limiting promotional e-mails did not just make for more satisfied customers but also benefitted the bottom line. Quintero knew that the only way to create such evidence was to run a randomized test that carefully tracked customer behavior and profits over time.

The Test

Quintero knew that the department directors were focused on the incremental revenues that could be obtained by sending out more offers to e-mail recipients. Her primary concern, however, was that customers who received too many e-mails would choose to "unsubscribe." Then, these customers would not receive promotional e-mails at all, potentially lowering their profitability in the long run. To capture this short-run vs. long-run impact, the test needed to last six months, a period that was long enough to observe customer behavior over time but short enough that she could argue that the existing policy should be kept in place until the test was completed.

The analytics team would need to test four different e-mail frequencies: Four e-mails per week (which was close to the "decentralized" average) versus three more restrictive policies, namely three, two, or one e-mail(s) per week.

The team decided to assign 10,000 customers randomly to each of the four conditions. To make conditions as similar as possible, all the customers in the test sample would receive e- mails featuring only one of the product departments during any given week of the experiment, irrespective of whether they received 1, 2, 3, or 4 e-mails per week. In other words, test customers would receive different numbers of e-mails, but all from the same department in any given week.

The Results

At the end of the 6-month test, Quintero received a spreadsheet with the following information from her team:

1 e-mail per week	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Subscriber attrition for promotional e-mails	2.0%	3.1%	1.9%	3.1%	2.4%	2.4%
Average revenue of subscribed customer	$1.05	$1.02	$0.83	$1.27	$0.86	$0.98
Average revenue of unsubscribed customer	$0.87	$0.73	$0.64	$0.54	$0.50	$0.51
LTV						**$2.26**

2 e-mails per week	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Subscriber attrition for promotional e-mails	3.6%	3.9%	3.0%	3.3%	4.1%	2.5%
Average revenue of subscribed customer	$1.35	$1.47	$1.34	$1.65	$1.17	$1.08
Average revenue of unsubscribed customer	$0.95	$0.78	$0.63	$0.52	$0.50	$0.48
LTV						**$2.96**

3 e-mails per week	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Subscriber attrition for promotional e-mails	7.5%	8.3%	11.8%	11.2%	9.8%	8.8%
Average revenue of subscribed customer	$1.51	$1.51	$1.42	$1.39	$1.28	$1.43
Average revenue of unsubscribed customer	$1.02	$0.83	$0.80	$0.55	$0.68	$0.59
LTV						**$2.92**

4 e-mails per week	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Subscriber attrition for promotional e-mails	15.4%	21.0%	23.2%	16.5%	15.9%	12.4%
Average revenue of subscribed customer	$1.45	$1.34	$1.21	$1.53	$1.58	$1.54
Average revenue of unsubscribed customer	$0.89	$0.75	$0.66	$0.75	$0.70	$0.48
LTV						**$2.56**

Quintero quickly understood what was going on: Sending out more e-mail clearly increased the revenue from a customer who still received e-mails. However, it also dramatically increased how many customers unsubscribed. Once a customer had unsubscribed, they could still purchase on the website. However, the revenue was lower level after unsubscribing. Combining these different effects in a 6-month lifetime value number showed that sending two e-mails per week was best.

Moving Forward

The e-mail frequency test changed the dynamics of the monthly Product Department Director meetings.[1] The test had put to rest the question of whether limiting the number of promotional e-mails to customers was in the best interest of the company. The test had also increased Anna Quintero's credibility in digital marketing. Most surprising to Anna, however, was that the department directors had started to seek her advice on problems that required some customer analytics to solve.

Such a problem had been at the center of the most recent meeting. In fact, it was a problem of Anna's own making. The company-wide agreement to limit promotional e-mails to twice a week meant that the different departments now had to coordinate their promotional e-mail activities

[1] Each department director runs one of the major product categories sold by Pentathlon: Endurance (e.g., running, cycling), Strength and Fitness (e.g., gymnastics, yoga), Water Sports (e.g., sailing, kayaking), Team Sports (e.g., soccer, basketball, rugby), Backcountry (e.g., hiking, climbing), and Racquet (e.g., tennis, badminton).

with each other. Having to coordinate was not in and of itself something that departments objected to. They were used to negotiating over scarce resources such as marketing budgets, retail space, and head counts. Instead, the problem was a lack of information. François Cabret, the department head of Endurance Sports put it this way:

“When we negotiate over budgets or space allocation in our stores there are a clear set of key metrics to focus on. For example, everyone agrees that sales per square foot of store space is important. However, when we negotiate over how to allocate promotional e-mails across departments, we just can't see eye-to-eye. For example, I have argued that endurance-themed e-mails should have high priority for women because running is just as popular among women as it is among men. But Patricia [the department head of Racquet Sports] says that, online, her category has particular sales appeal to women, even if this is not our experience in stores. Frankly, I don’t know how to resolve these questions. We can get reports about which customer segments buy which products online. But that is not what we need. What we really want to know is how effective different promotional e-mail messages are for different customer segments. But there are so many different customer types with different purchase histories that I don’t even know how to start thinking about finding out the answer.”

During the meeting. the department heads had discussed the idea of running another test in order to clarify the effectiveness of different promotional e-mail messages for different customer segments. But the idea had fizzled because no one felt that they could wait another six months, the length of the e-mail frequency test, to get an answer. Ten minutes after the idea of a test had been rejected Anna suddenly spoke up:

“We have been thinking about this the wrong way. I agree, we should not run a test – but not because it is going to take too long. We don’t need to run a test because we already have all the data we need to determine the effectiveness of promotional e-mail messages. In fact, my analytics team should be able determine which e-mail message should be sent to each individual customer. Give me a few days and I will get you some answers.”

The Idea

In the months since the decision to limit customer e-mails, the departments – unable to agree on an optimal allocation procedure – had used a simple random rule as an interim compromise. Anna’s sudden realization during the meeting was that the random allocation rule had created something close to experimental data that could be used to analyze the effect of different promotional messages.

The random allocation rule had been implemented as follows:

- Each week the digital marketing department split customers with valid e-mail addresses into randomly assigned e-mail groups.
- Each of the departments was allocated one of these e-mail groups for their exclusive use for one week, subject to the e-mail frequency restriction. An additional group received no promotional e-mails (i.e., the “control” condition).
- The e-mails sent were designed by each department and would feature products from that department. Of course, once customers clicked on the promotional e-mail and were on the Pentathlon website, they could buy products from any department they were interested in.

While this procedure had been chosen because it did not favor any department over another and it was easy to administer, Anna noticed that it was ideally suited to analyze how different customers reacted to different messages. The key was that customers were being allocated to departments – and therefore to differently-themed messages, including the control condition – randomly.

The Data

Anna asked her analytics team to pull the following data:

- The data pull should be based on the last e-mail sent to each customer. Hence, an observation would be a "customer–promotional e-mail" pair. In addition to the six message groups, a seventh group of customers received no promotional e-mails for the duration of the test ("control").
- The data should contain the basic demographic information available to Pentathlon:
  - age: Customer age (coded in 4 buckets: "1" < 30, "2" 30 to 44, "3" 45 to 59, "4" >= 60)
  - female: Gender identity coded as female "yes" or "no"
  - income: Income in Euro, rounded to the nearest EUR5,000
  - education: Percent of college graduates at the neighborhood level of the customer, coded from 0-100
  - children: Average number of children at the neighborhood level of the customer
- The data should contain some basic historical information about customer purchases, specifically, a department-specific frequency measure.
  - freq_endurance – freq_racquet: The number of purchases in each department over the last year, excluding the two test weeks.
- The key outcome variables should be:
  - Buyer: Did the customer complete a purchase within two days of receiving the e-mail ("yes" or "no")?
  - total_os: Total ordersize (in Euro) conditional on the customer having purchased (buyer == "yes"). This measured spending on all products, not just for the department that sent the message.

Finally, Anna requested that her team pull a total of 600,000 observations and divide the data into a training sample and a test sample using a 70-30 split.

The Analysis

After compiling the data, the digital marketing analytics team began to work through the instructions Anna had e-mailed them:

"Please perform all estimation using the training sample. Use the test sample to assess model performance for the binary decision of whether a customer buys after receiving a particular message or no-message. Please use the following steps:

1. **For each customer** determine the message (i.e., endurance, strength, water, team, backcountry, racquet, or no-message) predicted to lead to the highest **probability of purchase**. Describe your approach.

2. For each message, report the percentage of customers for whom that message or no-message maximizes their **probability of purchase**. Comment on the distribution of expected response across messages.
3. For each customer, determine the message (i.e., endurance, strength, water, team, backcountry, racquet, or no-message) predicted to lead to the highest **expected profit** (COGS is 60%). Describe your approach to predict order size and how you calculated expected profit.
4. Report for each message, i.e., endurance, racket, etc., and no-message, the percentage of customers for whom that (no) message maximizes their **expected profit**. Comment on the distribution of expected profit across messages.
5. What expected profit can we obtain, on average, per customer if we customize the message to each customer? Include no-message as an option as well.
6. What is the expected profit per e-mailed customer if every customer receives the same message? Answer this question for each of the possible messages (i.e., endurance, strength, water, team, backcountry, racquet) and the no-message option. Comment on the distribution of expected profit across messages and the no-message option.
7. What is the expected profit per e-mailed customer if every customer is assigned randomly to one of the messages or the no-message condition?
8. For the typical promotional e-mail blast to 5,000,000 customers, what improvement (in percent and in total Euros) could Pentathlon achieve by customizing the message (or no-message) to each customer. Compare the performance predictions from this personalized approach to scenarios where (1) each customer is sent the same message selected based on average performance, (2) a random message assignment approach is used (i.e., the status quo), and (3) no message is sent (i.e., the control condition)?

A New Policy Proposal

In addition to presenting the results of the analysis during the next monthly department director meeting, Anna Quintero decided to propose a new process for allocating promotional e-mails across departments that was based on her team's analytical results. She wrote a draft for a new e-mail policy:

1. Promotional e-mails will be allocated to departments on a monthly basis
2. During a month, e-mails will be assigned to departments as follows:
    a. For each customer, the analytics team determines the two messages that yield the highest expected profits
    b. The two departments whose messages yield the highest expected profit for a customer each control ½ of the allowed e-mail messages to that customer during that month
3. During the last week of each month the analytics team uses the data from e-mails sent during the first three weeks of that month and repeats the analysis described in step 2

Case Questions

1. Perform the analysis following the instruction e-mailed by Anna to the analytics team (Step 1 to 8 above). Use logistic regression, neural networks, random forests, and

XGBoost. Each ML model must be tuned using at least two hyper parameters **(40 points)**

2. Comment on the draft for a new e-mail policy proposal. Are there any weaknesses? Can you suggest at least one improvement? **(5 points)**
3. Generative AI (**5 points**): Describe **in detail** how your team used Generative AI-tools like ChatGPT to support your work on this case. Provide pdfs and/or screenshots of your "discussions" with these tools and comment on what things did and did not go well. Make sure to add discussion about your thought process and how you tried to maximize the benefits from using these tools. Also add any questions you may have about the assignment and the support you received from GenAI so we can discuss these topics in class.

   Note: No matter how you used Generative AI-tools, you are expected to fully understand all elements of the case solution submitted by your group. Any group member may be called on in class to walk us through your thought process and how different parts of your code work and how you arrived at your solution.

Hints

- Your priority in this case should be to generate (customized) predictions for each customer (step 1-8). Spend most of your time getting this working for each of the 4 model types.
- In this case you should select your final model, first and foremost, based on how well it fits the data (i.e., evaluate fit in the test data). The better the model fits, the better it will be able to predict the profit that can be achieved by customizing which message to send to each individual customer (or no-message). Use model performance measures such as AUC, R-squared, RMSE, etc. and Gains Curves to compare model performance.
- Please work in the NPTB framework. You do not need to use uplift modeling.