Group 40's Approach to Data Driven Credit Card Design with FiTech

Kieran Huffman, Aarushi Gaikwad, Charles Qian, Sakshi Gandhi

The first step taken in determining how to best approach maximizing customer lifetime value (CLV) for FiTech was understanding the data available and the methods we could undertake with the data. Quickly, it was understood that there was no historical data for individual, past customers. Instead, we had to work with two very small datasets, one of which contained details and performance metrics regarding past marketing campaigns, and the other of which included basic details on all 12 of the possible credit card offerings.

It was important that we conduct some basic descriptive analysis simply by reading over these small datasets in order to get an idea of how we might go about our further analysis. We looked at how a given bankruptcy (BK) score related to CLV, noticing that there was an inverse relationship between the two: as BK score increased, CLV decreased, undoubtedly due to the increase in credit risk that comes with customers who have higher BK scores. We also considered if there was a "best product" that would universally be preferred by all customers. We determined that there was, in fact, such a product: the third credit card in the dataset of the 12 possible credit cards had a fixed APR of 14.9% and no annual fee, making it the most affordable option for any customer. However, consequently, this option also had the lowest CLV values, with a CLV of only 2 for a customer with a BK score of 250. As a result, we had one offer which would likely lead to the highest customer response and satisfaction, but also the least CLV and revenue for FiTech.

Because we did not have a historical dataset regarding individual customers and their habits, we instead had to build our model based on a modified form of the small dataset of past marketing campaigns. Using the Pandas library in Python, we stacked and melted the dataset, modifying it to instead show customer response rates based on the different possible types of credit cards. From there, we trained a logistic regression model, using this modified dataset as its data. The response variable was whether or not the customer responded to the marketing campaign, and the explanatory variables were APR, whether the APR was fixed or variable, the annual fee for the card, whether the card was used Visa or MasterCard, how many customers were contacted in the marketing campaign, and the BK score of the customer targeted. The count of responses and non-responses was included as a weight in the model.

This resulted in a statistically significant model with a McFadden Pseudo-R-squared of 0.033. Our resulting area under the RO curve (AUC) was 0.661, and our resulting recall was 82%. Our coefficients were largely as expected, with higher APRs, variable APRs, and annual fees all lowering customer response probabilities. Cards using Visa also had higher customer response than cards using MasterCard. We decided to stick with our logistic regression model as opposed to a more complex machine learning model (such as random forest, XGBoost, etc.) due to logistic regression's much easier interpretability, as well as its general effectiveness and reliability.

Next, we used simulation to further investigate customer behavior and response and used our insights for a test marketing campaign. We wanted to ensure that each product offer was tested across the entire range of BK scores (150, 200, and 250). To do this, we used a partial factorial design that we created in Radiant. For the sake of fairness in our testing, we evenly distributed how many emails would be sent out to market each of the 12 credit cards. This allowed us to see which details of each credit card (APR amount and type, as well as annual fees) resulted in higher or lower customer response, grouped by different customer BK scores. Running our simulation, we found that our results aligned with the results of our logistic regression model, with preferable features for customers (lower APR, no annual fee, etc.) resulting in greater response rates.

For the final, roll-out round of the marketing campaign, we incorporated CLV into our data as well in order to maximize profitability for FiTech. Looking at response rates across BK score segments, we categorized each credit card as high-performing, moderate-performing, or low-performing. Credit cards with the lowest response rates of all were removed from the campaign, as they would bring in very few customers, resulting in little overall CLV for FiTech. In determining the allocation of email offers to be sent out for each possible credit card, we adjusted our distribution from the even distribution we used in the test round. For this round, we distributed it with a CLV-weighted response model, prioritizing customers with a BK score of 150 but still considering customers with BK scores of 200 or 250 where they previously showed high response rates.

The results of our roll-out round were successful: we achieved higher profits and CLV compared to the test round. This largely was driven by better response from high-value customers, which we attribute to our increased targeting of them in the roll-out round. This validated our data-driven approach to the campaign, resulting in a successful roll-out of the credit card offers to 750,000 potential customers.