Improving Financial Wellness With Banking Data

Shane Racey(str45), Joshua Clark(jlc489),
Maddi Klancher(mrk256), Oren Steingarten(oas26)

Table of Contents

Figures

Capital One transaction data can improve financial wellness

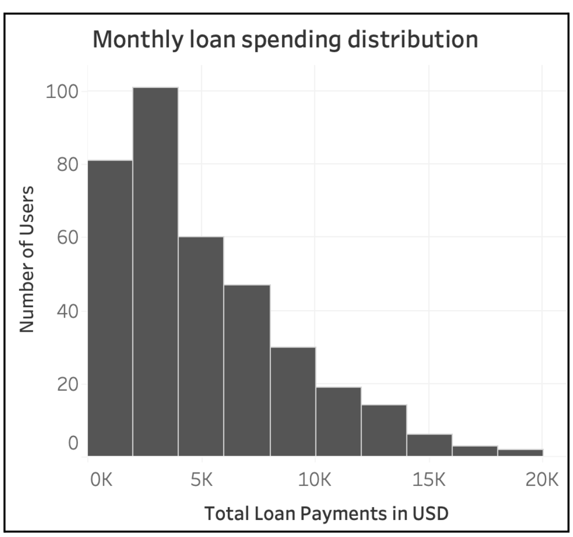
Financial data is among the most expensive and protected data available today, and for good reason. We make lots of transactions throughout our daily lives. Understanding what this data says about ourselves can help us identify ways to improve our own habits and elevate the quality of transaction management.

As a major retail bank, Capital One is in the unique position of having access to this data. However, a recent report by PricewaterhouseCoopers suggests that most banks leverage less than 1% of this data for their business needs. In this report, we will explore customer profiling, with the objective of providing better financial outcomes for customers through personalized rewards. We are using two datasets to emulate the data your clients generate: real transaction data, pulled from team members via the Plaid API (Application Programming Interface) , and public banking data from Kaggle. The Plaid API dataset is high fidelity transaction data from the debit and credit accounts of our team members, then standardized and cleaned by Plaid's platform. The public transaction data is stitched together from Czech CRM (Customer Relationship Management) and core banking data from 1999, converted into English in 2018 for use in data science projects. This data includes millions of transactions, as well as loan data, customer demographics, and account balances.

Our objective is to identify opportunities for your bank to go beyond the industry standard uses of financial data, such as credit approval and fraud detection, to focus on the ways that Capital One can leverage their insights to create financial wellness for their customers. We will also detail how these better outcomes for users improve the bottom line of the business. Finally, we will provide a recommendation for testing two experimental features that make use of these insights, and discuss how these tests might be implemented.

Predictive features in retail banking data

From spending patterns and transaction data, we identified factors that correlate with financial stability and trends. These factors can help us identify and predict profit, interest, and transaction fees for banking services. Determining how to convert this data into both usable and valuable categories for data models is essential to profiling customers and predicting trends.



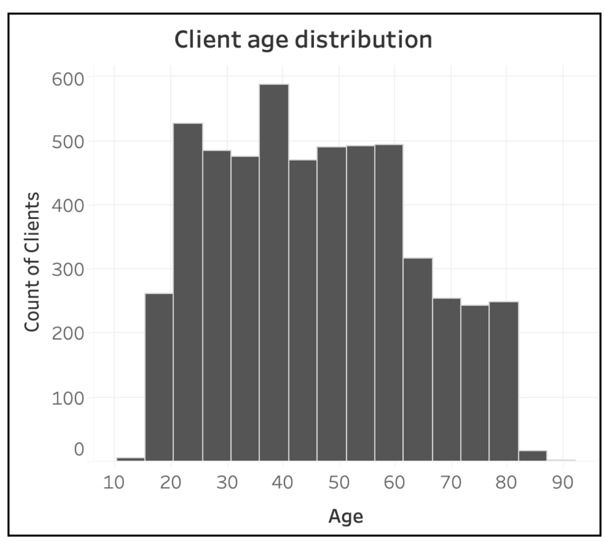


Figure 1.1: *The monthly loan spending distribution of clients*

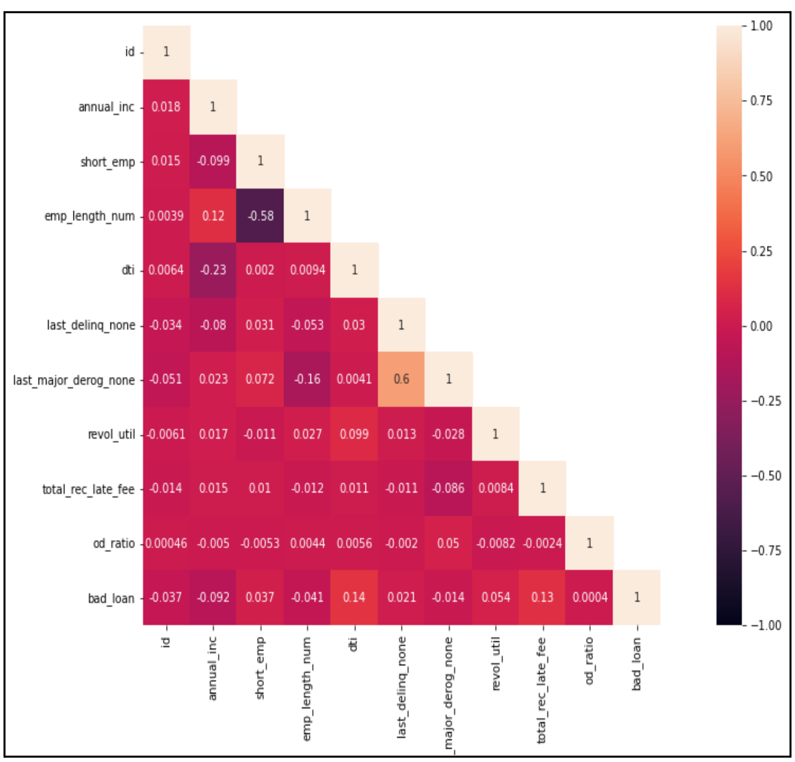
Figure 1.2: *A distribution of the clients ages*

Figures 1.1 and 1.2 represent two different ways to explore client demographic data. This will allow us to identify features of account holder personas and begin organizing users into clusters. In figure 1.1, we graph the distribution of monthly loan spending by number of clients. In figure 1.2, we chart the distribution of account holder ages. Loan payments and age are two variables that form a strong baseline to create features that determine how to profile Capital One account holders.

Figure 1.1 provides a lot of insight on a customer's loan payments. This chart's rightward skew reveals a significant variation in client payments. We are interested in how this correlates with loan payment history and reveals whether clients repay their loans on-time or behind schedule. Sorting account holders in this way will help us group them into categories that will identify their unique needs. For example, a client that struggles with repaying their loan on time could need help learning how to better budget their money so they don't fall behind. This insight can help your bank provide customers with a more personalized experience that can help keep them financially stable, in turn increasing customer satisfaction.

Figure 1.2 also sets a strong foundation to explore how age impacts an account holder's financial stability and needs. This chart demonstrates how our dataset includes a pretty balanced variety of ages. The gaps between these ages will allow us to explore which kind of account holder personas Capital One currently attracts and identify opportunities in other age markets. The distribution's smaller count of clients below 20 highlights how young individuals are typically less concerned and involved with their finances. On the other side, the drop in count of clients between ages 60 and 90 highlights how older users shift their spending habits as they are more likely to have retired. Nevertheless, for clients between these age ranges, it might be better to look at additional variables. Overall, knowing the age of an account holder may help Capital One identify their clients' unique economic priorities.

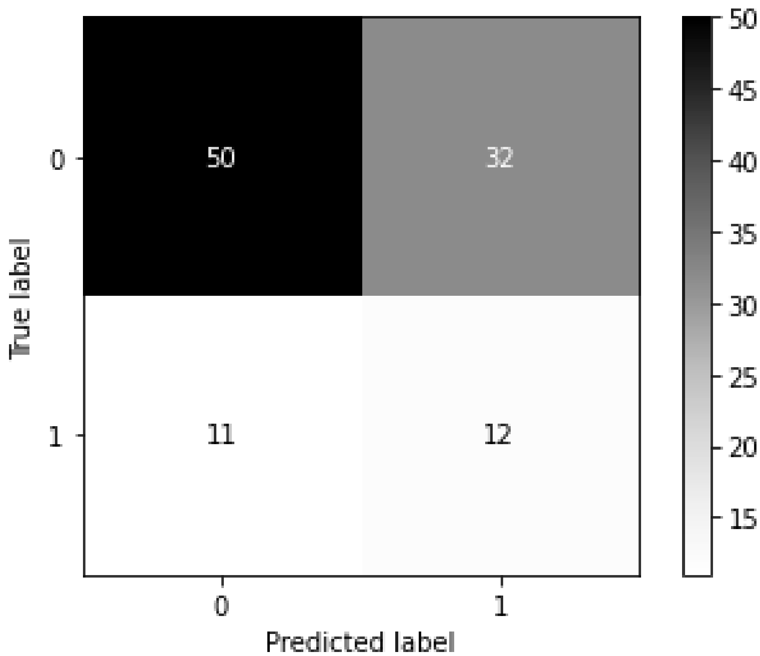
Logistic regression identifies missed payments and balance as the greatest contributing factors to default



In our next step, we wanted to use user past credit data to model the risk of future loan defaults. Because defaulting is a binary dependent variable, we opted for logistic regression to model missed loan repayments. That being said, what were the main independent variables to consider and incorporate into this linear regression? To identify the variables that are most closely correlated with loan defaults, we constructed a heat map. Heat maps are an effective tool to find positive

Figure 2.1: *Heat map for showing positive and negative variable correlation*

and negative correlations among variables. With regard to bad_loan, a binary variable indicating whether or not not a user defaulted (1 if yes, 0 if no), we identified short_emp (binary variable, 1 when employed for 1 year or less), dti (debt-to-income ratio), last_deliq_none (binary variable, 1 when a user had at least one circumstance of delinquency), revol_util (the amount of credit a borrower is using compared to all other available credit), and total_rec_late_fee (late fees received to date) to have the greatest effect.

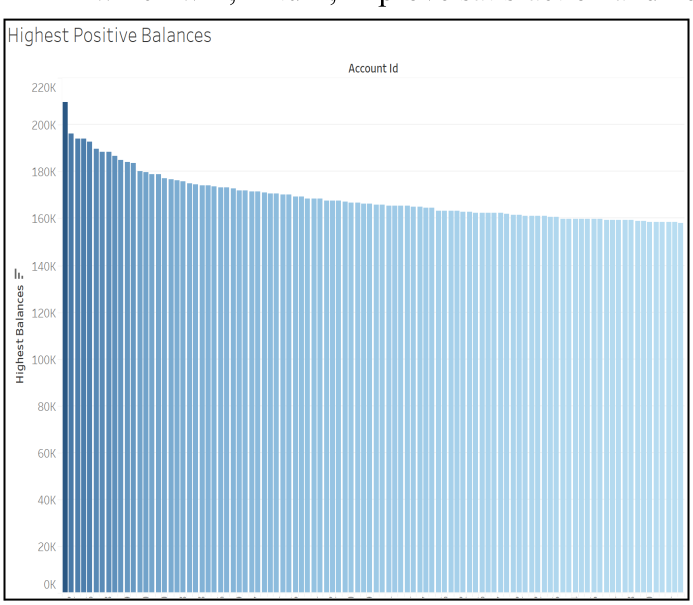


Based on the features identified in Figure 2.1, a logistic model was trained to predict which loans would be defaulted on based on each user's financial data. A test set was held in reserve, and was compared against the predictions of the model. The logistic model tends to overestimate the number of properly repaid loans, which suggests that other variables need to be incorporated for a better model. For example, with the help of CapitalOne's clean banking data, taking user credit score as a variable in the future could help address this false positive issue.

Figure 2.2: *Confusion matrix for logistic model predicting loan default*

How can banking data be leveraged to allocate rewards or conduct interventions?

How should Capital One allocate rewards credit to best drive sales? Alternatively, when should an intervention be sent to a user to warn about excessive spending? With respect to Capital One, customers will benefit by being able to better understand their spending habits, which will, in turn, improve satisfaction and help the businesses.



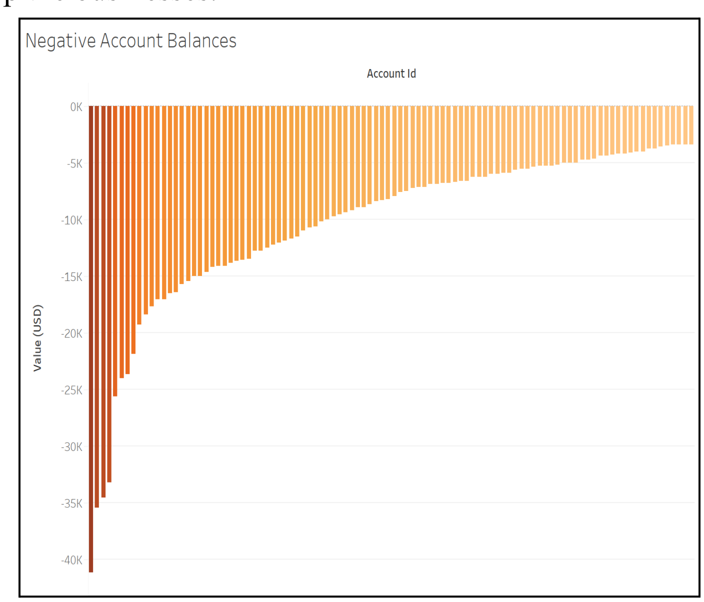


Figure 3: *Highest and lowest customer account balances*

These visualizations plot a series of positive and negative account balances of varying balances and debt respectively. This allows us to identify which account id's correspond to high value accounts or concerning accounts in debt. Accounts between either side of this spectrum will better be organized by other variables.

An account with a higher balance suggests that the account owner can be more affluent and would benefit from rewards programs more than other clients. Therefore, this data can help us identify which clients to target when creating and marketing rewards programs. We can join this data with the spending categories–such as travel, shopping, groceries, rent–that each account spends the most on. Combining this data provides Capital One with insights to create a more personalized experience for their users that increases their likelihood to engage with rewards and offers. This data-driven business strategy will encourage these high value accounts to spend more using Capital One, thereby increasing both income and customer satisfaction.

With debt closely tied to financial instability and an inability to make interest repayments, this data allows us to identify customers who may be at financial risk. Accounts exceeding a certain amount of debt should be high-priority targets for interventions and/or other programs. We can join this data with loan payment history using account ids associated with these concerning accounts to identify which of these accounts are at risk of remaining in debt and which are proactively working on moving towards a higher balance with regular payments. Similarly, we can join this data with other insights to learn more about how to best help these struggling clients. It is safe to assume that clients with the most concerning account balances have a higher need for education on budgeting and financial literacy overall. By providing these resources and intervening with these clients, Capital One can encourage them to pay regularly and increase their financial stability. As their clients see positive results from this initiative, they will have a better customer experience with Capital One. On the other hand, Capital One will see increased profits as these clients move away from debt.

Forecasting user spending enables interventions to prevent overdraft and incentivize spending

We used a Holt-Winters' seasonal forecasting model to forecast the spending trends of a Capital One customer. We chose this exponential smoothing method to fit common seasonal trends of user spending. Figure 4 helps us identify trends in each user's spending, which in turn allows Capital One to personalize their initiatives and offers. To use this model, we focused on one user's account transaction data. First, we cleaned the transaction data by filtering out withdrawals from the account made to pay off the monthly credit balances. After removing these outlying data points, we then flattened the transaction curve.

Using transaction data from April 2020 to January 2022, we forecasted the number of transactions a user makes per week as well as transaction amount per day. In Figures 4.1 and 4.2 below, the orange line represents the model's forecast from January 2022 to July 2022. To test the accuracy of our forecast, we plotted our real transaction data from January 2022 to July 2022. Here, we can see the comparisons of our forecast in orange with the real transaction data in blue.

Figure 4*: Modeling and predicting user spending patterns for alert systems*

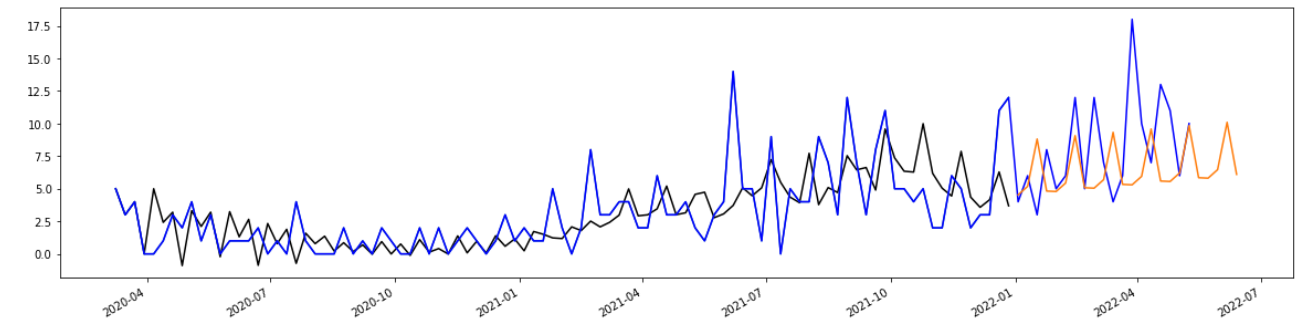


Figure 4.1: Forecasting the count of transactions per week

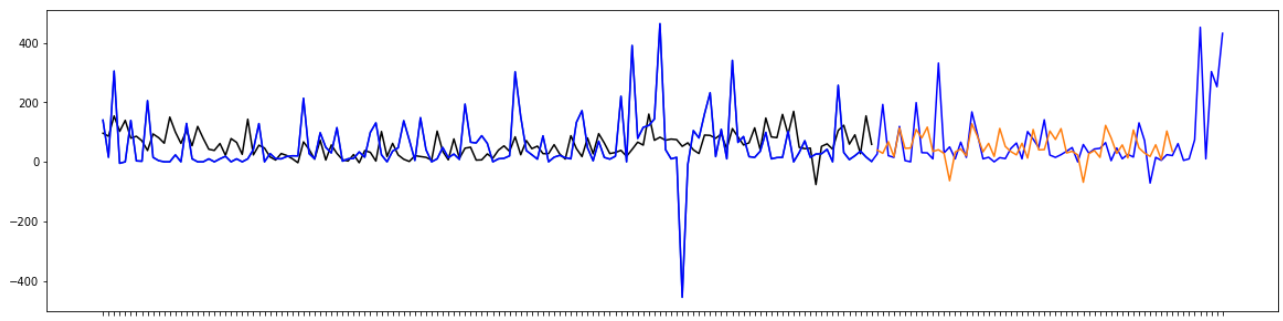


Figure 4.2: Forecasting the transaction amount (USD) per day

These predictions allow Capital One to access each user's spending habits by forecasting future user behavior. Predicting when users are more likely to spend will help Capital One plan when to offer rewards according to each user's spending trends. An even more essential application for this model would be comparing expected spending and earning with the current balance. After using this approach to identify which accounts are high or low spending, the user's current balance will suggest whether they are projected to overdraft or maintain a stable balance. Using these data insights, Capital One can take initiatives such as sending an alert to the user if their projected spending would result in an overdraft, or spending above what they would be able to pay off. Another initiative Capital One could take is to offer rewards for users with lower spending and higher balances to incentivize spending.

By leveraging an ongoing understanding of the spending patterns of users, Capital One can create better economic outcomes for their users, and help protect themselves against overdraft or default risk. We recommend that Capital One generates simple spending predictions, either via forecasting or aggregation, for users to power alert systems.

What impact do banking interventions, such as loan refinancing, payment alerts, and debt forgiveness, have on credit card repayment?

Given the objective of identifying opportunities to improve the financial outcomes of users, we selected an account that experienced such an intervention, in the form of refinancing a homeowners loan. From this user's credit card spending, we are looking for indicators of financial insecurity and the impact of the refinance.

Figure 5 compares the daily account balance for a credit card account of a single user over the one year period before refinancing a homeowners loan and the one year period after. In the year before refinancing, this user would spend months unable to repay their credit card debt between deposits, resulting in high card utilization and interest payments – all negatively impacting their credit score. This can be seen in the data at the high peaks, as well as long periods between negative slopes or returns to a $0 balance. The user refinanced in May of 2019, and immediately paid off their credit card. For the next year, they never used more than $1000 of their $10,000 limit, and never missed a minimum payment, as demonstrated by the frequent downward slopes of considerably reduced magnitude.

This result could suggest that banking interventions, such as offering to refinance a loan for accounts with reliable earnings and equity but low cash reserves, have a strong opportunity to positively impact the future financial outcomes of users; This user reduced their annual interest payments from $214.20 during the year before the refinance to $8.59 in the year after. However, since this is an account of a single user, in the future we may compare refinanced loans across multiple accounts to see if this trend holds for a larger population. Nevertheless, our overall goal moving forward will be to identify similar emergent opportunities, and recommend this or other interventions.

In terms of turning this single-user profile into a generalizable feature, credit card spending is only an issue when the account balance rolls over from one month to the next, or when payments are not being made. There is an opportunity to extract this result via integration over the account balance, comparing monthly interest payments to spending, or checking average account balances over some period.

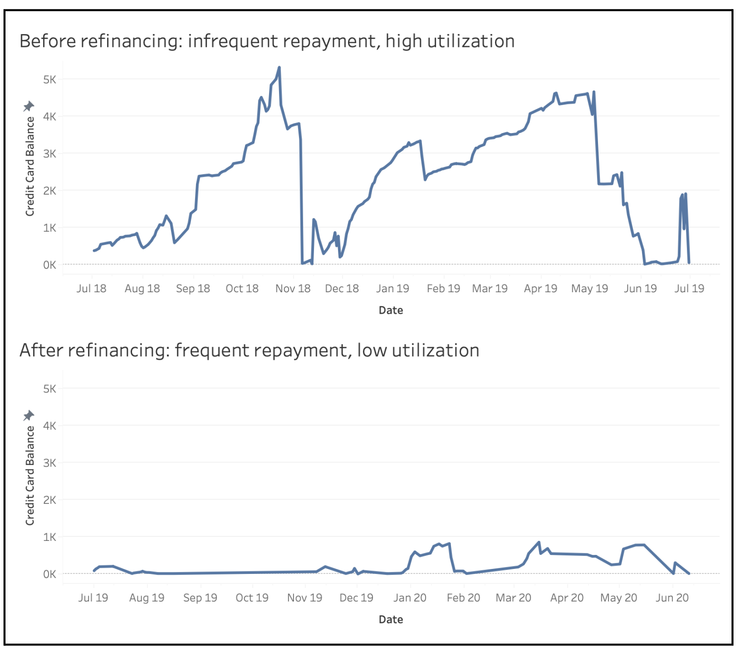


***Figure 5**: Effect of refinancing on credit card balance history*

Simulation identifies programs that increase the likelihood of payments as effective methods to reduce credit card default rate

Given the logistic model developed to identify users at risk of default, we wanted to test how a few different intervention methods might change the outcomes for Capital One users. To test this, we defined two different programs Capital One might employ in an effort to reduce the number of defaults: The Make-A-Payment Program, where users were reminded via banking app notification to make a minimum payment on their credit card, and the Debt Forgiveness Program, where a percentage of credit card debt was forgiven and credited to the user's account. To simulate these programs, the test data for the logistic regression model for user defaults was used as a baseline, and the test data was treated on a range of treatment effects for each program.

In the Make-A-Payment program, the treatment effect represents what percent of users respond to the notification and make an extra payment. For the Debt Forgiveness program, the treatment effect represents the amount the Balance for the most recent month was reduced. The new features were then used to generate new default rates, and compared across increasing treatment effectiveness rates.

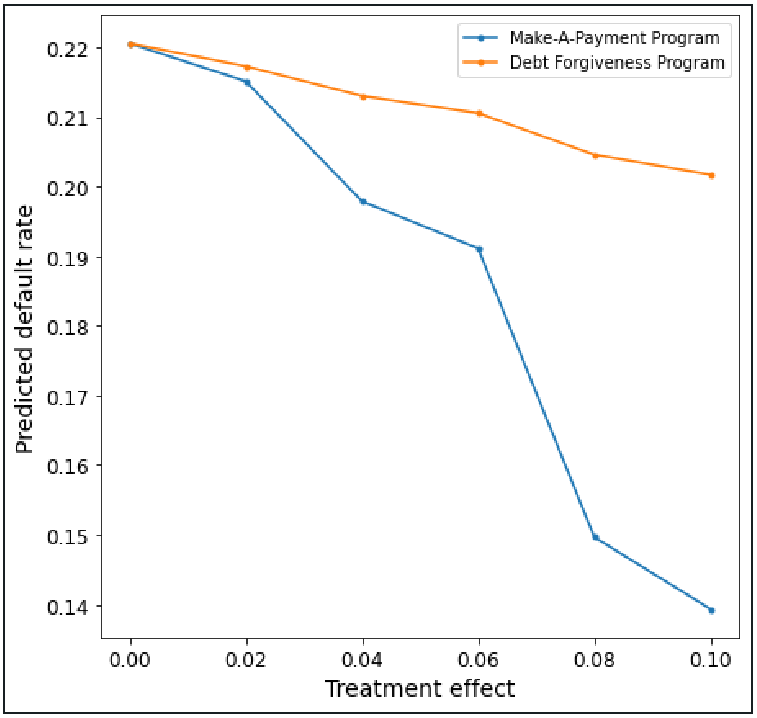


Figure 6: *Testing the effectiveness of default reduction programs*

This experiment determined that the Make-A-Payment program was more effective than debt forgiveness as a strategy to reduce defaults. Though lower account balances do correlate with lower default rates, making a payment pushes back the time-to-default by at least another 9 months. Interestingly, there is a large spike in effectiveness in the payment program between 6% and 8%, while the debt forgiveness program remains relatively linear in its effect.

The fact that a program encouraging payments is more effective than reducing debt in lowering the average default rate is a promising one for Capital One. A 10% debt reduction strategy corresponds to a massive expense for the issuing bank, while encouraging payments via app notification, one-tap integration for quick minimum balance payments, or other methods of increasing the rate of repayment in users are comparatively inexpensive.

Next steps for Capital One: Using casualty to test intervention methods for improve banking outcomes in practice

It was our goal with this project to identify a few different vectors for how Capital One might improve the financial wellbeing of their customers and reduce the capital risks of their business. While the Make-A-Payment program seems promising in simulation, the best way for

Capital One to identify the scope and value proposition of this opportunity is through A/B testing of real users within their ecosystem, rather than relying on simulation data or modeling.

Two experimental programs that A/B testing would work well on are predictive overdraft notifications and the Make-A-Payment program. In both cases, we recommended that Capital One develop the appropriate feature for their banking app, then roll out a pilot program to a subset of their users, then record the financial outcomes of the pilot group and a control group.

In the case of the Make-A-Payment program, exact numbers on how effective the "make a payment" notifications are in driving user payments and the relative increase or decrease in defaults could be determined. These numbers could then be used to identify how much should be invested in these programs; for example, if each default costs the bank $0.04 for each dollar of credit offered, a value per percent reduction in defaults could be calculated, and educational programs, rewards programs, or other expense-driven solutions could be used to augment the Make-A-Payment program, so long as the costs of these programs were offset by commensurate reduction in defaults.

Given the opportunity represented by the wealth of financial data Capital One has available, we hope that we have encouraged you to look beyond credit scores and fraud detection to ways that your bank can improve the financial stability of your users while also reducing default risk. We strongly believe that using this financial data to improve the financial well being of your users can be mutually beneficial. A great extension to this project would be to try to identify financial interventions that improve a user's chance to move to a higher income bracket, higher credit rating, or some other measure of financial security. A different direction could be identifying business accounts primed for expansion if given a loan, or recommending alternative merchants to reduce spending.

References

PricewaterhouseCoopers. (n.d.). Data and analytics in financial services. Retrieved April 17, 2022, from https://www.pwc.com/us/en/industries/financial-services/research-institute/top-issues/data-analytics.html

Petrocelli, L. (2020). Retail Banking Demo Data. Retrieved April 17, 2022, from https://data.world/lpetrocelli/retail-banking-demo-data.