The Research Behind Our New RFM Feature

Author: Nick Hartmann

Claps: 106

Date: Nov 6

At Klaviyo, when we set out to solve a problem for our users, we conduct research to help us consider what we should build and how we should build it. The product decisions we make when developing a new feature are based on many inputs, such as customer feedback and product usage data. In this blog post, I will discuss how we conducted an exploratory data analysis to help us make product decisions for a new segmentation and reporting tool within Klaviyo â€" the RFM Analysis feature.

Introduction

An e-commerce company using Klaviyo gets access to all of its conversion data. This data contains invaluable information that would inform how best to communicate with each customer, maximizing the potential value of each of them. However, actually extracting these insights can be challenging.

Within Klaviyo, we provide tools that give users the ability to create segments of customers based on their order history. We have seen users create segments such as â€œcustomers with 5+ ordersâ€� and â€œcustomers who have placed an order within the past 30 days.â€� The level of flexibility provided by Klaviyoâ€™s segmentation tool has many benefits, but it can also be overwhelming. Ultimately, users want to be able to answer questions like *Who are my high-value customers that I should be treating like VIPs?*, *Who are my customers I might need to re-engage?*, and *Who are my inactive customers that are unlikely to return to purchase?*

I am the engineering manager for Klaviyoâ€™s Predictive Analytics team. Our team recently released the RFM Analysis feature in Klaviyo. The feature categorizes a companyâ€™s customers into distinct cohorts based on their purchasing behavior, giving the company a framework through which to think about their customer base and communicate effectively with each type of buyer. Part of my role in building the feature was to lead the development of our methodology for grouping customers. Our objective was to design a framework that could intelligently identify distinct cohorts of customers for use in reporting and communication, but would also be interpretable and tunable by the user.

This blog post is a discussion of how we made the key product decisions for the feature: How should we define the different groups of purchasers, and how do we provide useful customizability that allows our users to derive the most value from the tool.

About RFM

RFM Analysis is an industry-standard technique used to group customers into different cohorts according to their spending habits. The cohorts are defined by three dimensions:

- **R**ecency â€" how recently did a customer place their latest order?
- **F**requency â€" how many total orders has the customer placed?

- **M**onetary Value â€” how much money has the customer spent?

Describing an individual customer by these three dimensions gives a well-rounded picture of who they are as a purchaser, and allows us to define customer cohorts that would need different types of communication. For example, we can have a cohort called â€œChampionsâ€� comprising customers who have made a lot of purchases, spent a lot, and purchased recently (high frequency, high monetary value, high recency). We can have another cohort called â€œNeeds Attentionâ€� to represent customers who have made a lot of purchases and spent a lot in the past, but have not made a purchase recently (high frequency, high monetary value, average recency). A marketer might then decide to have a different strategy for communicating with each of these two cohorts.

Even with the rise of predictive analytics and AI in the marketing industry, RFM analysis has remained a go-to tool for many marketers. In contrast to predictive analytics, RFM is a descriptive analysis; it can provide an organized and interpretable current or historical view of a companyâ€™s customer base. By understanding how their customers fall into different cohorts, as well as how customers are transitioning between cohorts, a marketer can strategize about how to focus their efforts. Maybe they observe that an increasing number of customers are moving from the â€œChampionsâ€� cohort to the â€œNeeds Attentionâ€� cohort, suggesting that the company might want to focus more on maintaining their highest value customers. RFM analysis can also be complementary to predictive analytics when tailoring communication to individual customers. Consider a customer who is showing a high predicted churn risk. A marketerâ€™s approach to mitigating that customerâ€™s risk might depend on whether the customer is currently in the â€œChampionsâ€� cohort or the â€œRecent (new customer)â€� cohort.

Although the RFM analysis technique is an industry standard, there is no standard way to define the ultimate cohorts of customers. While developing Klaviyoâ€™s RFM feature, we thought a lot about how to best structure our algorithm and provide options for customization to allow different users to tailor the analysis to their unique companyâ€™s needs.

Goals of the RFM Analysis Feature

The Klaviyo RFM dashboard groups a companyâ€™s customers into six mutually exclusive cohorts, which are described briefly below:

- **Champions**: Purchased recently, often, and spend the most money.
- **Loyal**: Purchased often or recently, and spend a good amount.
- **Recent**: Purchased recently, but not frequently.
- **Needs attention**: Frequent customers who have not purchased for some time.
- **At risk**: Frequent customers who spent a low or average amount but have not purchased for some time.
- **Inactive**: Customers who do not purchase frequently and have not purchased in a long time.

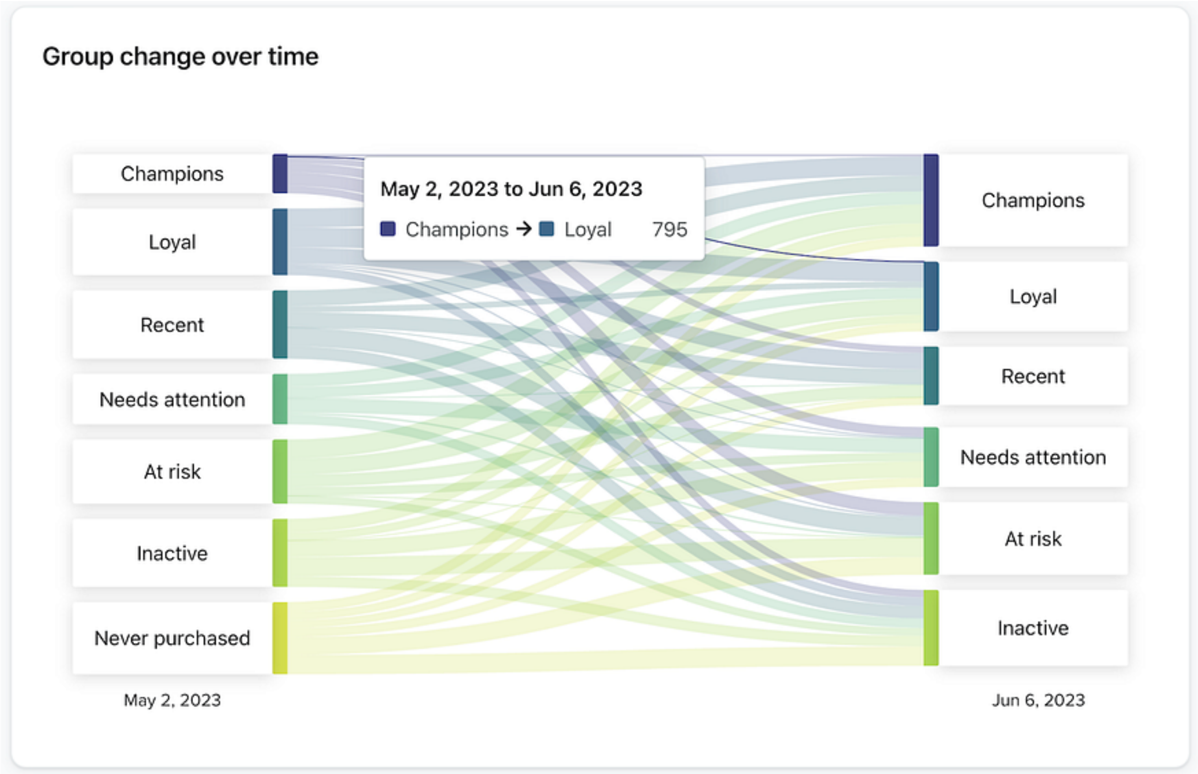
The decision to have six cohorts with these definitions was based primarily on industry-standard practices for RFM analysis, as well as conversations with our prospective users to learn how they think about their customer base.

When considering what we wanted to accomplish with our RFM feature, we established the following narratives from the perspective of a hypothetical user:

- When I review my business performance, I want to see how many low- / medium- / high-value profiles I have so that I can better plan my marketing strategy to move profiles to higher value groups.

- When I analyze the effect of my past marketing actions, I want to see how my customersâ€™ purchase behavior changed over time so that I can understand customer retention in high value groups, identify opportunities to mitigate churn risk, and demonstrate the value of my marketing activities.
- When I am planning my strategy for communicating with my customers, I want to be able to tailor communication based on each recipientâ€™s potential value and place in the customer lifecycle.

To address each of these goals, we created a two-part feature. The first part is a dashboard that shows a company how its customers fit into each of the six cohorts. The dashboard has several tools for visualization, including the Sankey diagram shown below, which illustrates the paths customers take between cohorts over time.



The second part of the feature is the ability to create audiences and targeted messaging based on each customerâ€™s current cohort and their movement between cohorts.

As we were developing our algorithm for grouping customers into cohorts, we aimed to choose methodologies that would be most appropriate to accomplish the ultimate goals of the RFM feature.

Our Scoring Framework

Before we came up with our system for translating a customerâ€™s Recency, Frequency, and Monetary Value into a cohort, we first wanted to talk to our customers to understand how they think about customer purchasing behavior, especially with respect to the R, F, and M dimensions. We came away from those conversations with several insights, including the following:

- **Every company has its own definition/intuition for what makes a â€œrecent,â€� â€œfrequent,â€� or â€œhigh-valueâ€� customer.** This is something that makes sense intuitively. Some companies see customers purchase more frequently than others, just due

to differences across industries. A â€œfrequentâ€� purchaser for a mattress brand might make one purchase every few years, whereas a â€œfrequentâ€� purchaser for a cereal brand might make one purchase every week. In addition to industry differences, some companies simply have different preferences for how they like to think about purchaser groups. For example, some companies like to categorize a group of VIP customers as the top 10% of their purchasers (by some metric). For others, their VIP group is more exclusive â€” maybe just the top 1% of purchasers. To account for differences across companies, we decided that it would be important to define cohorts using percentiles (e.g., â€œtop 10% of customers in terms of Frequencyâ€�), and to provide options for customization of the grouping algorithm.
- **Given a dimension (R, F, or M), companies like to think about their customer base as sorted into three buckets.** For example, considering Recency, we often heard companies say that they organize their customer base into three groups â€” active (high Recency), lapsed (medium Recency), and inactive (low Recency). This way of thinking helped inform how we should categorize customers for the purpose of the RFM analysis, when weâ€™re thinking about all three dimensions simultaneously.

Based on our customer conversations, we came up with a high-level framework for scoring customers and organizing them into RFM cohorts. Given a companyâ€™s customers, we take the following steps:

1. Assign each customer a **scaled score** from 1 to 3 on each dimension, where 3 represents high value and 1 represents low value. For example, consider a customer who has made a lot of purchases and spent a lot of money, but has not placed an order in a long time. That customer might be a 1 for Recency, a 3 for Frequency, and a 3 for Monetary Value. Ultimately, each customer can be represented as a **vector** of their scores for **R**, **F**, and **M**. The customer in the preceding example could be represented as (1, 3, 3). Since there are three possible scores for each of the three dimensions, there are 27 possible vectors to represent a customer.
2. Map each customer to the appropriate **cohort** based on their vector of scores. There are six cohorts â€” â€œChampions,â€™â€™ â€œLoyal,â€� â€œRecent,â€� â€œNeeds Attention,â€� â€œAt Risk,â€� and â€œInactive.â€� There is also a predefined mapping of vectors to cohorts. For example, a customer with a score vector of (3, 3, 3), (3, 2, 3), or (3, 3, 2) will be assigned the cohort of â€œChampion.â€�

The mappings of vectors to cohorts is defined in the following table, and is based on insights from customer conversations regarding how they think about their purchaser base:

Cohort	R	F	M
Champion	3	3	3
Champion	3	3	2
Champion	3	2	3
Loyal	3	3	1
Loyal	3	2	2
Loyal	3	2	1
Loyal	2	3	3
Loyal	2	3	2
Recent	3	1	3
Recent	3	1	2
Recent	3	1	1
Recent	2	2	3
Recent	2	2	2
Needs attention	2	2	1
Needs attention	2	1	3
Needs attention	1	3	3
Needs attention	1	3	2
Needs attention	1	2	3
At risk	2	3	1
At risk	2	1	2
At risk	2	1	1

Given this scoring framework, we still had a lot of nuances to consider. The remainder of this blog post will discuss the exploratory data analysis that helped us nail down the specifics of how scoring would be done and smart defaults would be determined.

Our Exploratory Analysis

The goal of our exploratory analysis was to use real customer data to help us decide on the nuances of how we should compute the cohorts. There are three big components here:

1. **How do we compute a customerâ€™s â€œraw scoresâ€� for each of the 3 dimensions?** We understood at a high level what â€œMonetary Valueâ€� should mean, but should we base the scores on average order value or on total spend across all orders? We wanted to explore questions like this for all three dimensions.
2. **How do we translate raw scores for each dimension into a score of 1, 2, or 3 (a â€œscaled scoreâ€�)?** How should we define a high-Frequency customer? Should it be someone who has made 5+ orders? 10+ orders? Maybe it should be different for every company. Should we base a customerâ€™s Frequency score on their number of orders relative to the rest of the customer base? For example, maybe we want a high-Frequency customer to be defined as someone in the top 10% or top 20% of number of orders when compared to the companyâ€™s whole customer base. These mappings are customizable by users, but we wanted to find smart defaults, so that the feature would be valuable right out of the box.
3. **How do we map vectors to cohorts?** Given a customer with a score vector of (1, 3, 2), how should we categorize that customer according to our six cohorts? Should they be in the same or a different cohort from a customer with a score vector of (1, 3, 3)? This type of question largely depends on how users would want to communicate differently with different vectors.

For the remainder of this section, I will refer to the above questions as Question 1, Question 2, and Question 3.

Data

To conduct our analysis, we produced an aggregated dataset of customer behavior that would allow us to explore how Klaviyo companiesâ€™ customers behave with respect to recency, frequency, and monetary value. We produced this dataset with respect to each of two dates â€” T1 and T2. Having data for two different dates allowed us to look at how customers move between RFM cohorts over time.

Our first goal for our analysis was to gather insights to determine how to answer Question 1 and Question 2 for each of our three dimensions.

Scoring Frequency

In an RFM analysis, the dimension of Frequency is meant to capture how often a customer makes purchases. The most straightforward way to compute each customerâ€™s raw score for Frequency (Question 1) would be to simply tally up their number of placed orders. However, we did consider an alternative â€” computing the number of unique dates on which a customer has placed an order. The argument for this method comes from the fact that, in most cases, if a customer places multiple orders on the same day, they could just as easily have been bundled into a single order. Consider a customer who places an order and then quickly realizes that they had forgotten to add something to their cart, so they place another follow-up order. Should that customer be treated as more â€œfrequentâ€� than a customer who ordered everything they wanted in a single transaction?

Ultimately, we decided that a given customerâ€™s raw score for Frequency should simply be defined as the number of orders they have placed. This decision was made for two main reasons:

1. This methodology felt more intuitive to customers in our interviews and more consistent with existing Klaviyo reporting functionality. Nowhere in our product do we provide a measurement of a customerâ€™s unique number of dates with a placed order.
2. Our analysis of our order data revealed that it is rare for a single customer to have multiple placed orders in a single day. The decision of how to compute raw scores for Frequency is unlikely to have a large impact on the final counts of number of customers in each cohort and the insights that the RFM dashboard would provide. We therefore felt that it was justifiable to choose the simpler and more explainable option.

Determining how these raw scores should be translated into scaled scores of 1, 2, and 3 (Question 2) is less straightforward. A natural starting point is to take a look at the distribution of the number of orders across customers. We first wanted to understand how many 1-time purchasers, 2-time purchasers, and 3+ time purchasers a company typically has. We realized that there was quite a bit of variability here. For some companies, the vast majority of their customers are 1- or 2-time purchasers, given the nature of their business or industry. For other companies, repeat purchasers are quite common. To accommodate this variability across companies, we developed the following scoring system for Frequency:

Compute the 33rd and 66th percentile for *number of orders* among customers within the company. Call those numbers a and b respectively. Then, the frequency scores for a given customer are computed as follows:

- If the customerâ€™s number of orders is less than or equal to a, assign a score of 1
- If the customerâ€™s number of orders is greater than a but less than or equal to max(a+1, b), assign a score of 2
- Otherwise, assign a score of 3

When we define the Frequency scores this way, many companies will have all of their customers with 1 order receiving a score of 1, all customers with 2 orders receiving a score of 2, and all customers with 3+ orders receiving a score of 3. However, for companies that have a lot of multi-time purchasers, the ranges are wider.

Conclusion

The scoring system for Frequency provides a fair way to distinguish customers within the context of their respective companies. For companies that rarely see multi-time purchasers, we still establish three distinct buckets of customers with respect to their Frequency, but the requirement to receive a top score of 3 is as lenient as possible. For companies that see many multi-time purchasers, customers are organized into â€œtop third,â€� â€œmiddle third,â€� and â€œbottom third,â€� making for an intuitive default grouping.

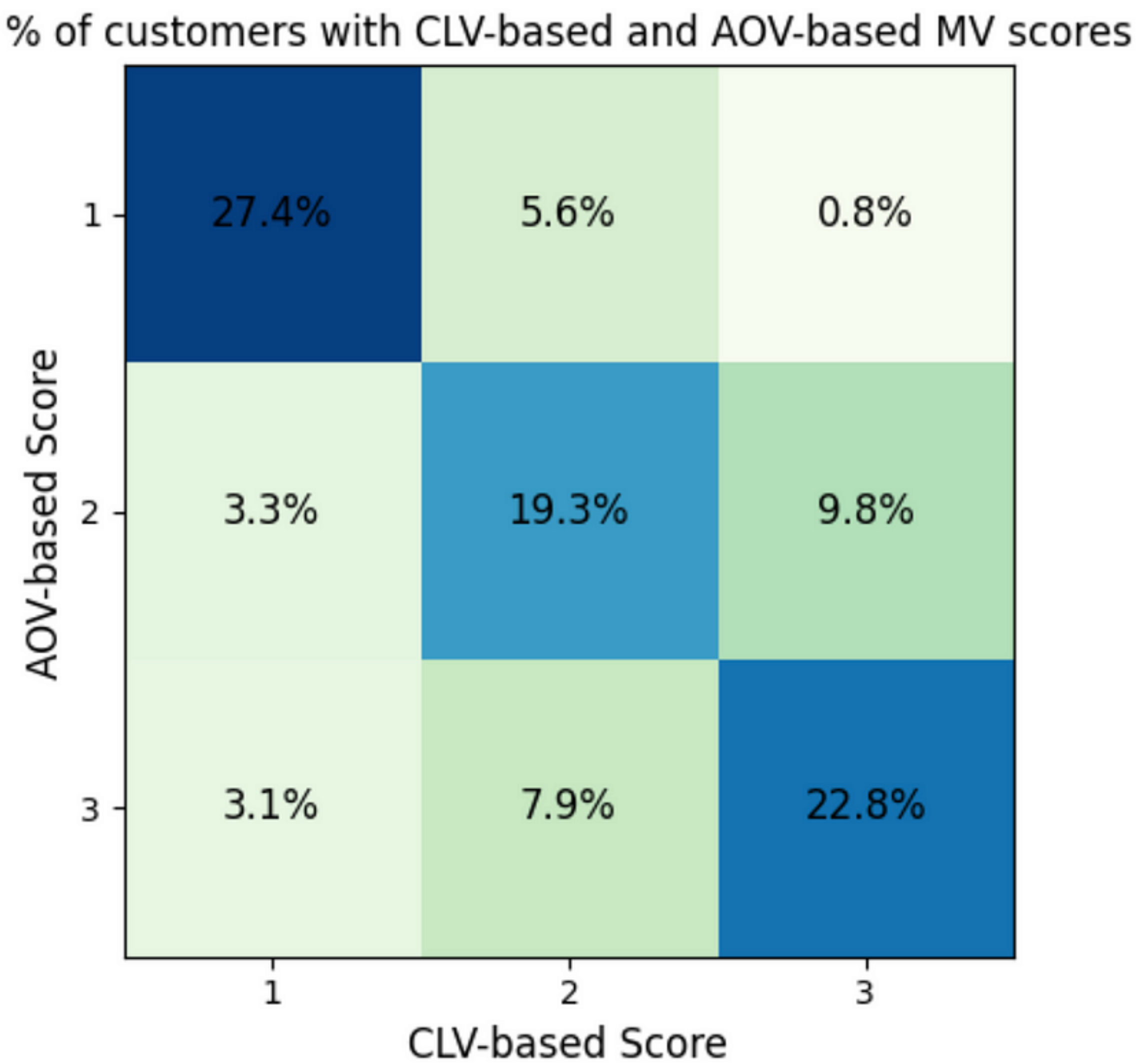
For the dimension of Frequency, as well as for Recency and Monetary Value, the user is ultimately able to adjust these default scaled score boundaries if they like.

Scoring Monetary Value

For the dimension of Monetary Value, the big question was Question 1 â€” how we should define a customerâ€™s raw score. There are two standard ways to compute the raw M component for a customer in an RFM analysis:

1. **AOV** â€” the average order value across all of a customerâ€™s past purchases
2. **Historic CLV** â€” the total lifetime spend of a customer

We explored the pros and cons of using each for the M component of our RFM model. First, we were curious to see whether there would be any correlation between a CLV-based Monetary Value score and an AOV-based Monetary Value score. Letâ€™s assume that we assign scaled scores for Monetary Value based on terciles (i.e., bottom 33% gets a score of 1, middle 33% gets a score of 2, and top 33% gets a score of 3). The following heat map shows how many customers in our dataset would get each scaled score, depending on which methodology for computing raw scores we chose. For example, 3.3% of all customers in our sample would receive a scaled score of 2 if the raw score were based on AOV and a scaled score of 1 if the raw score were based on CLV.



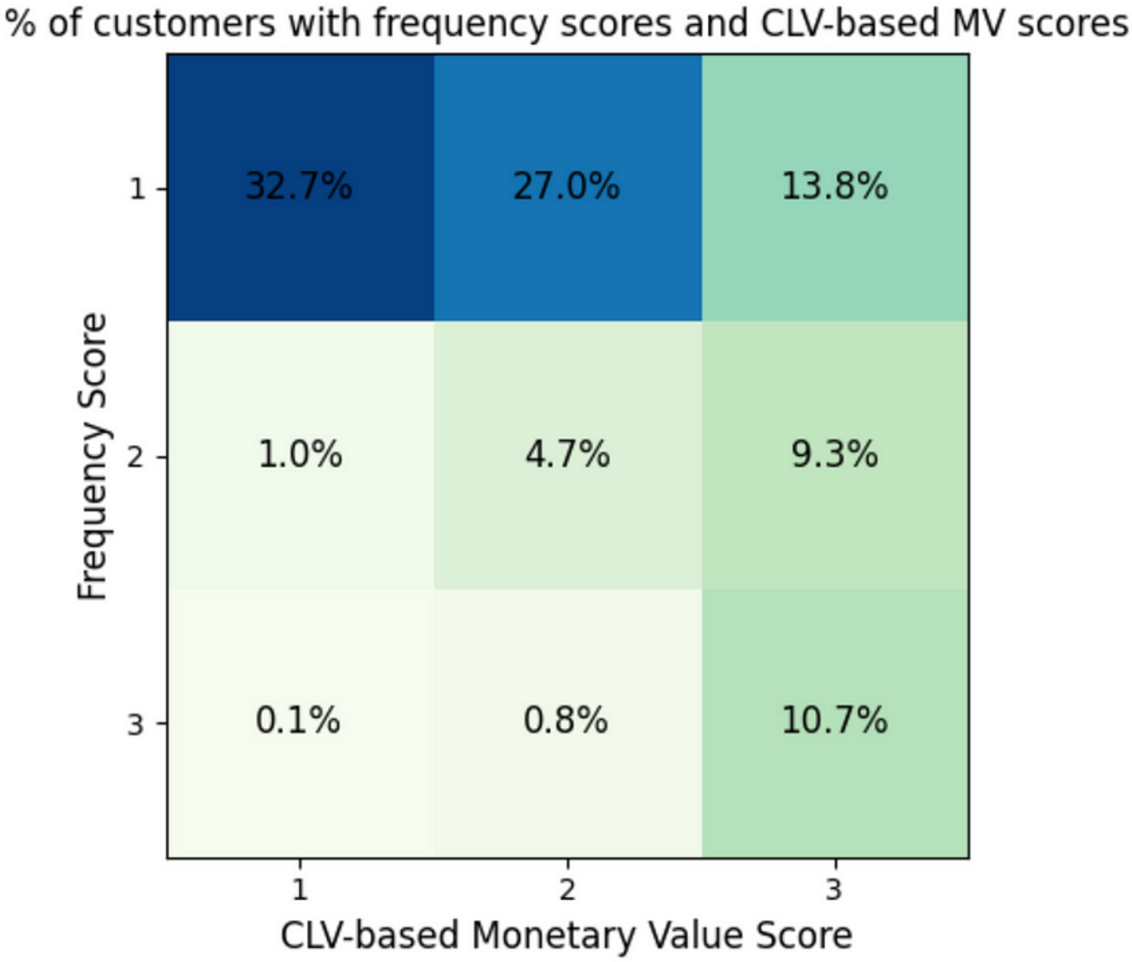
(chart produced using the T2 data from our dataset)

As we might have expected, in a large proportion of the cases, a customerâ€™s CLV-based scaled score is equivalent to their AOV-based scaled score. This is probably in part due to the fact that so many customers are one-time customers, so their CLV is equivalent to their AOV.

However, there are some customers who have a 3 for the CLV-based score and a 1 for the AOV-based score. This group would include customers who make frequent purchases of inexpensive items. We kept this distribution in mind as we weighed the considerations for our raw scoring methodology for Monetary Value.

Correlation with Frequency

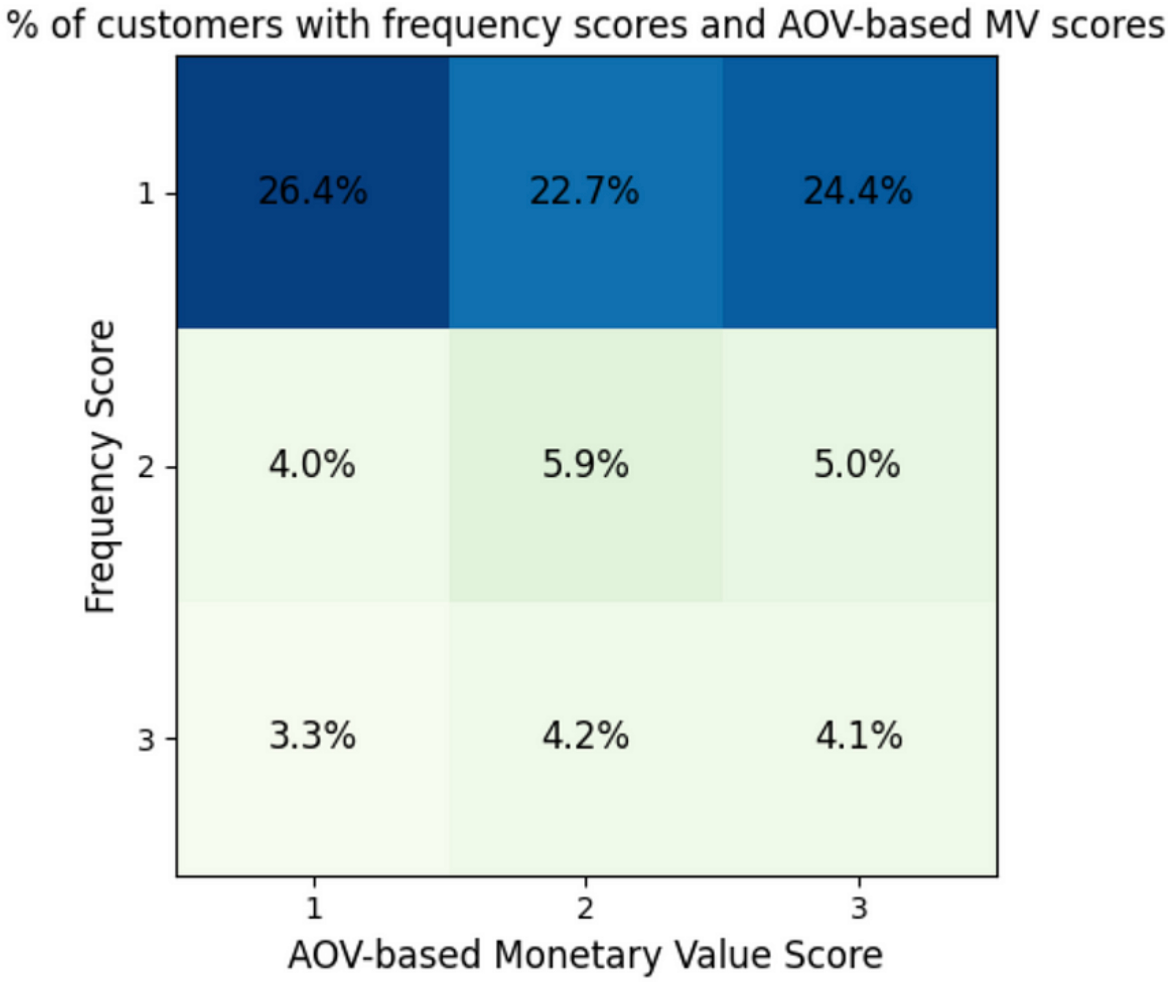
One consideration for the Monetary Value score was whether it would be too highly correlated with the Frequency score. If we use a CLV-based Monetary Value score, then it is very likely that the high-Frequency customers are also going to have a high Monetary Value score. This is not necessarily a concern, but might mean that the Monetary Value score is not really providing much information that we donâ€™t already have from the Frequency score. The heat map below shows the number of customers who would have each combination of Frequency scaled scores and CLV-based Monetary Value scaled scores. (Frequency scaled scores are being calculated according to the methodology described in the previous section.)



We can see that there is a fairly high correlation between the Frequency score and the CLV-based Monetary Value score. If we focus on customers with a Frequency score of 3, we can see that 92% of them have a 3 for their CLV-based Monetary Value score. This should not come as much of a surprise. Given that the vast majority of customers are one-time purchasers, we should expect that if a customer has 3+ purchases, theyâ€™re likely going to be in the top 33% of customers in terms of lifetime spend. This means that when it comes time to sort high-frequency customers into cohorts, they will be distinguished from one another only by their Recency (in most cases), because almost all of them will get the same Monetary Value score â€” a 3.

On the other hand, if we look at the customers who got a Frequency score of 1, not all of them get a CLV-based Monetary Value score of 1. Some of them made expensive purchases to earn themselves a Monetary Value score of 2 or 3. For these customers, the CLV-based Monetary Value score is adding information that Frequency alone does not provide.

Letâ€™s also take a look at the correlation between the frequency score and an AOV-based monetary value score.



There is some correlation between the Frequency score and the AOV-based monetary value score, but things look much more even. In some ways, an AOV-based monetary value score seems to do well where a CLV-based score sometimes falls short â€” by providing more information about a customer with a top Frequency score. That could be considered a point in favor of the AOV-based scoring method, but we had other considerations to weigh before making that decision.

A potential downside for AOV-based scoring

The ultimate goal of the RFM analysis is to categorize customers in groups that describe their current and potential value and their place in the customer lifecycle. As we considered some of the potential grouping outcomes of an AOV-based Monetary Value score, we started to doubt that the AOV methodology was aligned with that RFM goal.

Consider an active customer Alice, who has made 3 purchases with an average order value of \$100. If we assume that \$100 is a high value for this company, then Aliceâ€™s score vector might be a (3, 3, 3). Now consider another active customer Bob, who has made 50 purchases with an average order value of \$20. Maybe \$20 is a low order value for this company, so Bob gets a score of 1 for Monetary Value, and his vector is (3, 3, 1). According to our current mapping of vectors to cohorts, Alice would be put in the â€œChampionsâ€� cohort, and Bob would be put in the â€œLoyalâ€� cohort. This categorization might imply to the user that Alice should be treated as an VIP in ways that Bob should not be, when in reality, Bob has brought in much more money to the company than Alice has.

We considered reworking our mapping of score vectors to cohorts with this type of scenario in mind. However, in order to â€œfixâ€� the situation described in the previous paragraph, we would have to make the vectors (3, 3, 3), (3, 3, 2), and (3, 3, 1) all map to â€œChampions.â€� We would have to make similar adjustments throughout our mapping, ultimately resulting in the M dimension having little or no effect on the mapping of a vector to a cohort. Even though AOV provides a different interesting piece of information about customers, we wondered whether it might not be a piece of information that belonged in a system for identifying a customerâ€™s value.

We kept this potential drawback of AOV in mind as we spoke with customers during our development of the RFM feature.

Conclusion

Taking into account our findings from our analysis together with our insights from conversations with prospective users, we ultimately chose to define raw Monetary Value scores using Historic CLV. That methodology is more consistent with the goals of the RFM analysis; users agreed that Historic CLV was a better metric for helping to identify potential value of a customer.

Scoring Recency

For the dimension of Recency, the standard way to compute the raw score for a customer is as the number of days since the customerâ€™s last order. That meant that Question 1 was already answered for us. However, finding a reasonable answer for Question 2 was not as straightforward.

Experimenting with the tercile approach

We first considered the simplest way to compute scaled scores â€” dividing the population into terciles. With this approach, we would first compute the 33rd and 66th percentiles for â€œdays since last orderâ€� for a companyâ€™s customers. Letâ€™s imagine that these values are 90 days and 360 days for the company. For a given customer, if they had made a purchase within the past 90 days, they would get a scaled Recency score of 3. If not, they would get a scaled score of 2 if they had made a purchase within the past 360 days, or a score of 1 otherwise. We took a look at the distribution of scaled score thresholds that this strategy would create across our sample companies:

	Threshold for Scaled Score = 3 (33rd percentile)	Threshold for Scaled Score = 2 (66th percentile)
Mean	366	684
10th percentile	103	227
20th percentile	149	323
30th percentile	202	417
40th percentile	254	513
50th percentile	305	591
60th percentile	355	679
70th percentile	438	800
80th percentile	523	914
90th percentile	673	1,270

One thing that stood out when looking at these numbers was that for most companies, the cutoff to get the top score for Recency would be pretty lenient. For the median company, a customer would have had to place an order within the past 305 days to be in the top 33%. Of course, the idea of a â€œrecentâ€� customer will mean something different to different companies, depending on the nature of their business and how frequently people place orders, but 305 days seemed like a long time for the average company. We hypothesized that this leniency was being caused by companies having large populations of churned customers.

Composition of churned customers

When we were evaluating the dimension of Frequency, we learned that for most companies, a very large percentage of their customer base is made up of one-time purchasers. This is a fact that we needed to keep in mind when we determined our strategy for computing scaled scores for Recency.

Suppose greater than 66% of a companyâ€™s customers have churned, and will never purchase again; then, the 33rd percentile and 66th percentiles for â€œdays since last orderâ€� are going to become more and more lenient as time goes on and the amount of time that has passed since customers churned increases. This is probably not a property that we would want for our thresholds for scaled Recency scores. Suppose at one point the threshold to get a scaled score of 3 for Recency was 90 days, but over time it has relaxed to 300 days, due to churned customers â€œweighing downâ€� the percentile-based thresholds. That would make it very difficult for the company to compare RFM distributions of customers across different time periods.

To confirm our suspicions that this phenomena would be common, we looked at how the thresholds for Recency would change for our sample of companies from T1 to T2:

- The requirement for a Recency score of 3 would get more lenient over time for 93% of companies

- The requirement for a Recency score of 2 would get more lenient over time for 96% of companies

We can see from these proportions that itâ€™s almost a guarantee that the requirements for high Recency scores will become more lenient over time for a given company. We should accept that the thresholds for Recency scores might change over time as the companyâ€™s business changes, but the effect weâ€™re observing here is not due to business changes, but instead due to the churn a company naturally experiences over time. So, we considered ways to prevent this from happening.

Option 1: When calculating thresholds, restrict to customers active in the past 2 years

We believed that if we restricted our threshold computations to only customers who were active in the past *n* days, then the problem of the inactive / churned customers unfairly affecting the thresholds over time might be mitigated. To test out this hypothesis, we recomputed the Recency thresholds for T1 and T2 for each company, but this time, we only included customers who had made a purchase within 2 years of the time from which each threshold is being computed:

- The requirement for a recency score of 3 would get more lenient over time for 81% of companies
- The requirement for a recency score of 2 would get more lenient over time for 85% of companies

This did not work as well as we instinctively thought it would. After we looked more closely at some individual companies, the reason for this pattern was illuminated. Consider a hypothetical Company X, who is new to e-commerce. In January of 2020, Company X was acquiring 10 new customers per day, but by January of 2022, they have grown steadily to a rate of 100 new customers per day. Letâ€™s say that we compute the 33rd percentile â€œdays since last orderâ€� among all customers who were active within the past two years, as of January 2022, and use that to represent the threshold for a scaled Recency score of 3. Among the customer pool weâ€™re looking at, most of them have made their first purchase very recently, due to Company Xâ€™s exponential growth, so we expect that threshold number of days to be fairly low. Now imagine that Company X continues to grow over the next year, but the growth rate levels off. By January of 2023, they are acquiring 150 new customers per day. If in January 2023 we take a look at the pool of customers who were active over the preceding two years, the proportion represented by very newly acquired customers will not be quite as high as when we looked in January 2022. As a result, the 33rd and 66th percentiles for â€œdays since last purchaseâ€� are likely to be slightly higher in terms of number of days.

The growth trajectory of Company X is one that is pretty common for brands. Indeed, the numbers of companies in our sample who would still see their Recency scoring thresholds become more lenient over time even when using this method suggested that we should find a different methodology.

Option 2: Base thresholds on average days between orders

Given that defining Recency scoring thresholds by percentiles of Recency raw scores was not working as weâ€™d hoped, we decided to think outside the box.

What if we were to base frequency cutoffs on the typical gaps between orders for active customers. Suppose that for Company Xâ€™s multi-time purchasers the median days between orders is 90 days. Then, maybe it makes sense that in order to get a Recency score of 3, a customer should have placed an order within the past 180 days. Similarly, letâ€™s say that the 95th percentile of days between orders is 360 days. Maybe then we should use 360 days as the threshold between a score of a 1 and a 2 for Recency â€” if a customer hasnâ€™t purchased in 360 days, they are likely not to purchase again given the patterns in this companyâ€™s data, and itâ€™s probably fair to consider them lapsed. We hypothesized that the average number of days between orders should also theoretically provide us with relatively consistent thresholds over time â€” not automatically becoming more and more lenient.

To test this out, we first filtered our dataset to only multi-purchase customers who had made a purchase within 2 years of T2, and we computed the average time between orders for each of those customers. From there, we grouped by company to find the 50th and 95th percentiles for â€œaverage number of days between ordersâ€� for each companyâ€™s multi-purchase customers. That distribution looked like this:

	50th Percentile Average Days Between Orders	95th Percentile Average Days Between Orders
Mean	89	481
10th percentile	24	170
20th percentile	36	242
30th percentile	50	308
40th percentile	63	360
50th percentile	75	415
60th percentile	88	481
70th percentile	106	562
80th percentile	127	700
90th percentile	171	859

For the median company, their median days between orders is 75, but some companies will have a higher or lower value depending on the nature of their business.

Using the 50th percentile and 95th percentile for â€œaverage days between ordersâ€� as the thresholds for Recency scaled scores, we found the following trends:

- The requirement for a Recency score of 3 would get more lenient over time for 56% of companies
- The requirement for a Recency score of 2 would get more lenient over time for 52% of companies

We were pleased to see numbers around 50% using this methodology. We should expect that there will be a little variability in each companyâ€™s thresholds over time, and we found here that for roughly 50% of companies the thresholds got slightly more lenient over time, and for the other 50% the thresholds got slightly less lenient.

Conclusion

Given the strengths outlined in the previous section, we decided to define the Recency scoring thresholds based on the 50th and 95th percentiles for *average days between orders*. The exact percentile cutoffs we chose were intuitive, but rather arbitrary. We might decide to revisit this decision based on user feedback.

Producing Score Vectors

Once we had our methodologies for producing scaled scores for each dimension, we could identify the scoring thresholds for each company in our dataset, and then sort each customer into the appropriate cohort. The distribution below shows how all customers within our sample (across all of our sample companies) were organized when we computed cohorts as of T2:

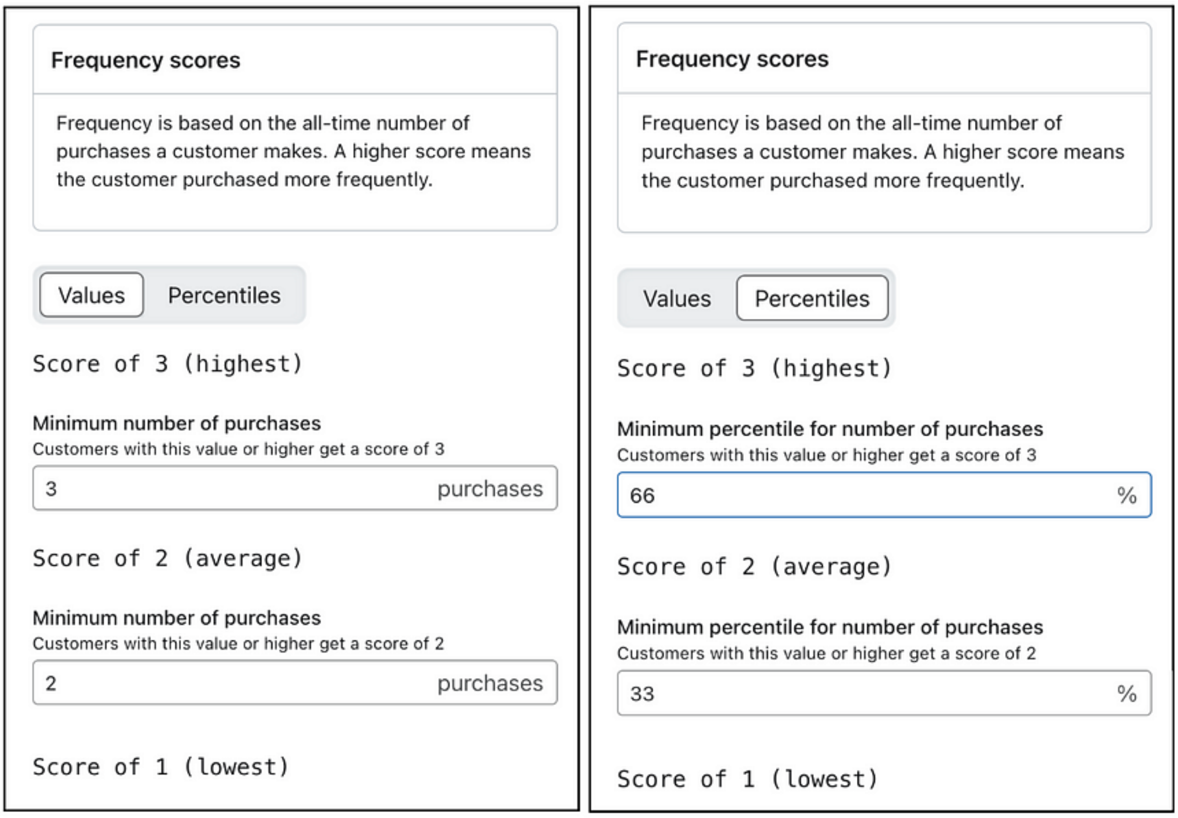
R	F	M	Cohort	% of all customers
1	1	1	Inactive	25.86%
1	1	2	Inactive	18.41%
1	1	3	Inactive	7.00%
2	1	2	At risk	6.88%
2	1	1	At risk	6.83%
1	2	3	Needs attention	4.73%
1	3	3	Needs attention	4.43%
1	2	2	At risk	3.73%
2	3	3	Loyal	3.35%
2	1	3	Needs attention	3.15%
2	2	3	Recent	2.57%
3	3	3	Champion	2.54%
3	1	2	Recent	2.25%
3	1	1	Recent	2.10%
2	2	2	Recent	1.22%
3	1	3	Recent	1.05%
3	2	3	Champion	1.03%
1	2	1	Inactive	0.86%
1	3	2	Needs attention	0.64%
3	2	2	Loyal	0.59%
2	3	2	Loyal	0.23%
2	2	1	Needs attention	0.18%
3	3	2	Champion	0.17%
3	2	1	Loyal	0.11%
1	3	1	Needs attention	0.05%
2	3	1	At risk	0.02%
3	3	1	Loyal	0.01%

It was not surprising that this distribution was not even, for reasons described in the previous sections. For example, the (3, 3, 1) vector is going to be very rare because most customers who qualified for a score of 3 in Frequency would also meet the bar to qualify for at least a 2 in Monetary Value.

Customization Options

Although we did our best to set smart defaults so that the RFM Analysis feature would work out-of-the-box for our users, we understood that some companies would want to be able to tweak the parameters to match how they think about their own business. For example, maybe a company wants to be stricter about which customers earn a 3 for Frequency, or more specific about what constitutes a â€œrecentâ€� customer.

At our initial launch, we decided to provide users with the option to decide the thresholds that map raw scores to scaled scores for each dimension. For example, as shown in the below screenshot, the user can define the percentile thresholds to define the Frequency scaled scores, or can choose to define thresholds in terms of raw numbers of purchases.



When deciding how much flexibility in customization to give, we had to strike a balance between enabling the user to tailor the analysis to their needs and stopping ourselves from overwhelming the user by providing too many customization options. We made two decisions specifically to prevent confusion:

1. For the dimension of Recency, a customer can define the thresholds only with raw numbers of days, and not with percentiles. The default raw values for a company are still based on the *average days between orders* methodology described in the Recency section, but we have chosen not to allow users to tweak our choice of the 50th and 95th percentiles representing the thresholds. This decision is based on our belief that explaining how the thresholds are calculated concisely would be challenging, and we do not want users tweaking a parameter that they might not completely understand.

2. We are currently not allowing users to tweak the mapping of score vectors to thresholds. A vector of (3, 3, 3), for example, will always map to the Champion cohort.

Our decisions about customization were based primarily on interviews we conducted with our prospective users before launching the feature. We may provide additional customization capabilities in the future.

Closing Remarks

Klaviyo empowers customers to be data-driven decision makers. Similarly, as developers of the Klaviyo application, we strive to make data-driven decisions, to provide our users with the best possible product. Although we are happy with the decisions we have made for the RFM Analysis feature, we recognize that we have only scratched the surface of the value that this type of tool will provide. Now having released the feature, we look forward to gathering user feedback and iteratively improving our scoring system and user experience.