**Final Project Report: Team 6**

Team Members: Xiaoqi Hu, Fernanda Lin, Haolan Ma, Chenhang Niu, Yi Yu

Date: 2019-12-09

## Problem Statement

Over time, credit card companies have received support data about the various customers they own. The company collects data on facts related to customers, such as their balances, purchases, cash advances, credit lines, and more. Our team's mission is to get meaningful insights from the data and then develop strategies that the company can target customers and increase the amount of credit card sales, which in turn increases revenue with the application of dimension reduction technique and k-means clustering.

## Data set Introduction

## The dataset[[1]](#footnote-0) summarizes the usage behavior of 8950 active credit card holders within 6 months period. It has 8950 rows(observations) and 18 columns(variables). Please refer to the appendix 1 for that detail explanation for the variables.

1. Findings

To continue our progress since the mid report, we did Principal Component Analysis (PCA) on the dataset to find out how we could divide the variables, and the optimal result is 5 principal components that cover 75 percent of the cumulative variance, which is sufficient to represent the variables significantly. Then, we applied these 5 principal components to explore the number of clusters using k-means, and we got the optimal number of clusters is 6, as suggested by the silhouette plot. To illustrate the clusters, we used box plots and radar charts (Appendix 3) to show the characteristics of each variable for each cluster to facilitate our cluster analysis since graphics can clearly show the magnitude of variables, so it is easy and intuitive to read and interpret. From the cluster analysis, we want to dig into the variables first to see how does each cluster performs on each variable. Additionally, to get an idea that how does each cluster looks like, we summarize each cluster’s performance on each variable to form our profiling that will be used to design the marketing plan respectively.

1. ***Variable Analysis***

The variables in the dataset can be divided into several groups according to their functions. For instance, there are three variables that describes payments, so we can put these variables together to explore and analyze. In this way, we have the groups for balance, payment, purchase, tenure, credit limit, and cash in advance.

* *Balance*:

Balance is reflected from two variables: Balance amount and the frequency of balance amount change. Most cluster have low balance amount except cluster. However, the frequency of balance amount changed for cluster 4 is similar with most of cluster which have low credit balance, which means customer in cluster 4 didn’t use a credit card to make normal transaction. Besides, they use it for other purposes, such as cash in advance. What’s more, customer in cluster 2 has the lowest balance and the frequency of balance amount changed, which shows they are the most inactive customer compare with other customers in the other clusters.

* *Cash in Advance:*

Cash Advance was reflected in three variables: “Cash in Advance”, “Frequency of Cash in Advance being paid” and “Number of transactions made with Cash in Advance”. For Cash in advance we can see most of the cluster has low cash in advance, only cluster 4 pops out with a higher value, it can be a potential sign for a cash borrower figure. Continue with cluster 4 we can observe that this cluster also high on both cash advance transaction numbers and frequency of cash in advance being paid while other clusters stays low relatively. So we can conclude that cluster 4 are group of customers who needs cash very frequently and heavily. However, they also have ability to pay back the cash they borrowed on time which means they have very good turnover rate on cash. So they might be small business’s owner. Other than that only cluster 3 seems like take relatively more cash than other four clusters, it may depend on their neighborhood and daily situations, they may need cash for stores or restaurants only accept cash more than people who live in other neighborhoods.

* *Credit Limit:*

There is only one variable reflects Credit limit which is itself. From the graph we can conclude that cluster 6, cluster 4 and cluster 5 are clusters with higher credit limit increasingly. Both clusters 4 and cluster 5 has outliers with highest credit limit $30000 per month, however in average cluster 5 has higher credit limit than cluster 4. So we might be able to conclude that cluster 5 is the richest cluster clients and host truthful customers we have in our portfolio, therefore to permanently or temporarily increase credit limit for group 5 can be considered as a way to incentive them spend more. For cluster 4 might be considered as small cooperation clients who needs borrow cash to turnover frequently, lastly cluster 6 has slightly higher credit limit than cluster 1,2 and3, which can be considered as a potential high value customers for further market events and promotion events.

* *Purchase:*
  + *Regular Purchase:*

Regular purchase is reflected from three variables: purchase amount made, purchase frequency scores between 0 and 1, and the number of purchase transactions made. The benchmark for the amount of purchase made is 8,855,647 dollars, which is the average total amount purchased of all accounts. Along with the reference, we found that clusters 5 clearly a big spender, spend 1,265,993 dollars collectively (14.3% of 8,855,647 ), which is significantly higher than the total amount purchased by the other clusters. Cluster 6, as shown in the box plot (Appendix 2), outperforms than the rest of the groups, which seems like spending very conservatively overall. Expectedly, cluster 5 and cluster 6 that score 0.925 and 0.918 on purchase frequency spend in a very frequent manner to accomplish their overall high spending. When it comes to the rest of the clusters, all of them score below the average purchase frequency of around 0.5. The cluster 1, similar to cluster 6, compared to cluster 2, 3, and 4 clearly spends more frequently. The rest groups are either not using a credit card to purchase or not using our service (credit card) to spend. Additionally, the total number of purchases transactions made reveals the consistent pattern as the purchase frequency.

* + *Installment Purchase & One-off Purchase:*

Installment purchase and one-off purchase are two different types of purchase methods. One-off purchase is that people pay all money when they make purchases, while installment purchase is that people rather paid the money in small parts in a fixed period of time than pay at one time. Usually people make installments when they are short of money but they want to buy expensive items. Cluster 5 has the highest value in both installment and one-off purchases, which may indicate this cluster has very high purchase ability. For frequency, cluster 1 prefer installments more. Cluster 5 and 6 love both installments and one-off purchases.

* *Tenure:*

The tenure for all cluster is on similar level, except cluster 1. Significantly, customer in cluster 1 is younger than in other clusters.

* *Payments:*

For the different clusters, those in group 5 have the highest median followed by 4, 6, 1, 3 and 2 respectively. As it was described before, those in cluster 5 tend to spend more than others, because of this, they are also required to pay more credit card bills, leading this cluster to have a high payment amounts. Cluster 4 makes major monetary loans, which also forces them to score higher on payment as well. Although 2 has a lower median, the results vary across the board, as this cluster has many outliers that have much greater payment values than the rest.

Minimum payments tend to be low for all clusters, but there are many outliers lying above the 3rd quartiles. We can assume that on average most customers usually pay smaller amounts regularly, or their required minimum payment amount is low, so their minimum payments remain low.

Although cluster 4 makes more minimum payments than most of the other clusters, when it comes to paying off their credit card bills, it seems like it has a tendency of not paying them in full. Cluster 3 usually pays off a small portion of its credit card bills, but this group does it frequently. Cluster 5 prefers to pay most of their credit card balances all at once, however, the majority pays partially in smaller portions throughout the period. Cluster 2 has a similar median to cluster 6, but based on its balance frequency, it takes them longer to pay their credit card bills. Clusters 1 and 6 are similar to cluster 3 which, pay regularly but in smaller amounts.

1. ***Cluster Analysis***

* *Cluster 1: The Subprime Customer*

From the visualization of cluster 1 and analysis of each variable, we define them as subprime customers in our portfolio. As we can observe from the radar chart, customers from cluster 1 has two highest values among all the clusters. Highest purchase frequency and highest purchase installments frequency. With a low installment purchase amount, and a low one off purchase frequency, we can know this group of customers like to shop with installment payment method for goods and services in normal price range. Also, customers in this cluster has a very low credit balance left each month, in addition with a low minimum payment needs to be made each month and a medium percentage of full amount paid, we can know that customers in this cluster has debt on their bank balance sheet each month. Furthermore, customers in this cluster tend to have larger tenure index and lower credit limit than other clusters’ customer which indicates they are relatively new customers, the combination of all these indexes tells us that customers in cluster 1 could be quite risky for the bank, once they don’t have the ability to pay back their debt, especially when they don’t have insurance to back their life up while they facing an uncontrollable situation, they will go bankrupt. All these indicates us that these customers are “moonlite” who live from paycheck to paycheck or “subprime borrowers”, which is a group of customers who will be considered as a relatively high credit risk for banks or lenders.

* *Cluster 2: Self-aware avoiders*

Members of this cluster Since they have a low number of purchases, purchase frequency, one-off purchase, one of purchase frequency, and almost all factors, except tenure, prc full payment and credit limit. They avoid using credit cards, although they blame themselves rather than issuers for their debt problems and worry about the damage they could do to themselves with a credit card. They use debit cards and cash for their purchases, but still carry a low revolving debt. They are slightly better off than the Average Joe and, for that matter, average cardholders—in terms of income and wealth. These consumers are likely to respond to simplicity and transparency in fees, rates and terms. Bolstering their confidence that they can use credit cards without mishap for certain purchases, particularly for short-term borrowing, will make a credit card more viable.

* *Cluster 3: The Average Joe*

Compared to most of the other clusters, this group of people tends to score lower in most of the variables. Cluster 3 pays a small portion of their credit card bill in full and has a low credit limit. However, the users update their balances very often. From this information we could assume that this group mostly uses their credit card for necessities, such as grocery shopping or other daily goods and rarely make big purchases. This group makes small regular purchases and pays back their credit card in small portions.

* *Cluster 4: The Cash Borrower*

From the visualization of cluster 4, it is interesting to find out that the distribution of variables for cluster 4 is extreme. That is, some variables in cluster 4 have extremely high value, while others have extremely low value. To be specific, cluster 4 has high value in balance, tenure, payments, minimum payments, credit limit, cash in advance, cash in advance frequency, and number of cash in advance type of transactions, meanwhile they have low value in all variables about purchases and purchases frequency as well as the percentage of full payment. In other words, the credit card users in cluster 4 have high credit limits, and they use their credit cards frequently. However, it seems that they hardly use the cards for purchasing. Instead, they use their credit cards to cash advance. According to Wikipedia, a cash advance is a service provided by most credit card and charge card issuers. The service allows cardholders to withdraw cash, either through an ATM or over the counter at a bank or other financial agency, up to a certain limit. For a credit card, this will be the credit limit (or some percentage of it). In this way, card issuers can gain from a higher interest and fee that is 3 to 5 percent of the amount being borrowed. When made on a credit card, the interest is often higher than other credit card transactions. The interest compounds daily starting from the day cash is borrowed. Other statistics can prove that there users are likely to be “cash advancers”: they almost never pay back with the full payment. Additionally, they also have relatively high credit limit and payments, which may indicate that they normally can pay back on time.

* *Cluster 5: The Big Spenders*

The Big Spenders. This group is by far the most interesting to analyze. Most of the features are widely varied in values. Since they do not only have a high number of purchases, purchase frequency, one-off purchase, and one off purchase frequency, but they also have high payments, credit limit, tenure, and low balance, cash advance, cash advance frequency. They use a credit card to make transactions frequently, do not carry a balance from month to month, pay their credit card bills in full by the due date, and hardly use any cash advance function. We learned cluster5 is significantly transactor, not revolver. As the nature of "Big Spenders", there are many outliers in this cluster: people who have/make abnormally high balance, purchases, cash advances, and payment. People in this segment keep their finances in order, they dislike revolving debt. They look for convenience and minimum effort, preferring to “set it and forget it” rather than get closely involved with banks or card issuers. This is the wealthiest segment in our analysis.

* *Cluster 6: The Platinum Class*

Cluster *6*, similar to cluster 5, has large installment purchase, purchase very frequently and also spend a lot compared to the rest of clusters. Unlike cluster 5, who tend to spend a lot at once, people in cluster 6 shop very conservatively but tend to pay all at once when it comes to one-off purchase. Moreover, they are also the group who shows the least interest to cash in advance. Consistent with this pattern, their payment, minimum payment, and percent of full payment fit with their purchase activity as well. The conclusion for this group is that they are a group of people who spend less than our biggest spender, cluster 5, and cluster 4, the cash browser. Given their purchase ability, they tend to shop responsibly. Besides, their little need for cash in advance suggest their stable life and less ability to default. However, their credit limit, which is only slightly higher than the newcomer, cluster 1, did not correctly reflect that their purchasability and predictive payment pattern. This also can be confirmed by the little balance amount left. Therefore, we name this cluster as the platinum class because this group of people has the potential to become our big spender and requires reevaluation on their credit limit.

IV. Recommendations

After the analysis of clusters, we draft out the marketing plan for each group based on the profiling.

* *Cluster 1: The Subprime Customer*

According to the analysis above, for this group of customers, we should increase the interest rate on the amount they did not paid on time, in order to encourage them to pay back loans on time. Also, the credit limit promoted for customers who falls into this cluster should be lower than normal customers in order to decrease the default risk. However, once customers being clustered into another clusters, we should provide them other services according to their ability to pay back, usually according to their actual income and assets owned. Furthermore, if some customers want to perpetually or temporarily increase their credit limit, we should first check their credit background, then according to their credit score, actual income and bank balance sheet, we can increase their credit limit and inform them they will get charge of extra interest rate with increased credit limit. Lastly, for new customers who falls into this cluster we may wants to provide them unsecured credit cards with higher interest rates in order to attract them from rival while preventing their default rate.

* *Cluster 2: Self-aware avoiders*

Cards that could instantly calculate the scale and duration of monthly payments for a given purchase at the point of sale would better meet this group’s needs. At the same time, we could provide more higher percentage of cash back to encourage customer make purchase by credit card, especially in the holiday season.

* *Cluster 3: The Average Joe*

To attract more people to apply for this credit card and incentivize the current users to spend more. With this, we should try giving them options for cash back or points to go to certain restaurants. We could offer 1-5% cash back depending on the frequency that the customer uses their card, the time they have had their card for and the total sum of their transactions. We would have to look at the frequency of their credit card payments as well. The customers could choose to get cash back from one type of expenditure such as groceries, gas, transportation, online purchases or restaurants, one of their choices. We could also incentivize them to get people to apply for the credit card by offering $50 credit for each of the people they introduce to apply for a credit card and get accepted. Eliminating annual fees to make it more affordable would help since they do not have high balances would also encourage users. While applying a 10% fee on their purchases when they do not pay the minimum monthly payment of $25, if they make purchases equal to or greater to that amount, would help decrease the number of people who pay their credit card bills late.

* *Cluster 4: The Cash Borrower*

Generally, people use credit cards for cash advance for some reasons: one is that they need money in a hurry, but they do not have that much amount of cash in their account, so they have to withdraw some cash from credit card to meet the imminent need. As for this type of user, the card issuer do not need to worry about too much because the users just do cash advance once in a while and they can usually pay back. However, since the users in cluster 4 have long tenure for credit card, they are more likely to use credit cards for just doing cash advance. They always need money but they do not have money at the same time, so they use cash advance on credit card as a way to make loan. Credit card issuers should pay more attention on these users because they can potentially make huge profits and risks at the same time. On one hand, most of these users can pay back on time, so issuers can make profits from charging high interests and fees. On the other hand, issuers are taking the risks of losing principals. Although there are only few payments that cannot be repaid, one non-repaid payment could cause big loss to issuers. Therefore, we think that the optimal choice for users who do cash advance is setting a maximum credit limit. With this strategy, users can still increase the credit limit with strong repayment ability, but if they do cash advance for many times, the credit limit will not be raised once it reaches the maximum credit limit. Credit card issuers can gain profits from cash advance with less risks in this way.

* *Cluster 5: The Big Spender*

Getting a card into this segment’s wallet involves offering rewards. To differentiate their

armed with enhanced segmentation, we can not only craft better value propositions but also identify groups that are not well served by current offers. New frontiers in credit card segmentation: Tapping unmet consumer needs, we should position it not merely as a spending instrument but as a tool that facilitates financial success through ease of use. For instance, if a credit card could offer a means of steering a large purchase straight into a low-rate installment loan, it could meet this segment’s occasional borrowing needs without the stigma or higher interest rates associated with revolving credit.

* *Cluster 6: The Platinum Class*

As mentioned in the analysis for cluster 6, we conclude that they are a group of people who have the potential to spend more with a stable payback ability but their credit limit are underestimated. In general, the credit limit determined by the credit card company with a series of complex evaluation such as tenure, the income level and their purchase power. Additionally, the clients who accumulated a record of making regular purchases on their card each month and repaid their balance in full on time will receive approve more easily when applying for raising the credit limit. People in cluster 6 have been our clients for a long time as shown in the tenure record and clearly demonstrates their potentials on purchasing power over time. Their profile indicates that they could be the clients who move to a higher stage on their life development such as promotion and marriage. Furthermore, their independent of cash and stable payback pattern indicates that they are eligible for a raise of credit limit. Therefore, the next plan for this cluster is aim on promoting them our platinum card which include various benefits such as reward program and personal concierge service[[2]](#footnote-1) after reevaluating their profile.

V. Conclusion

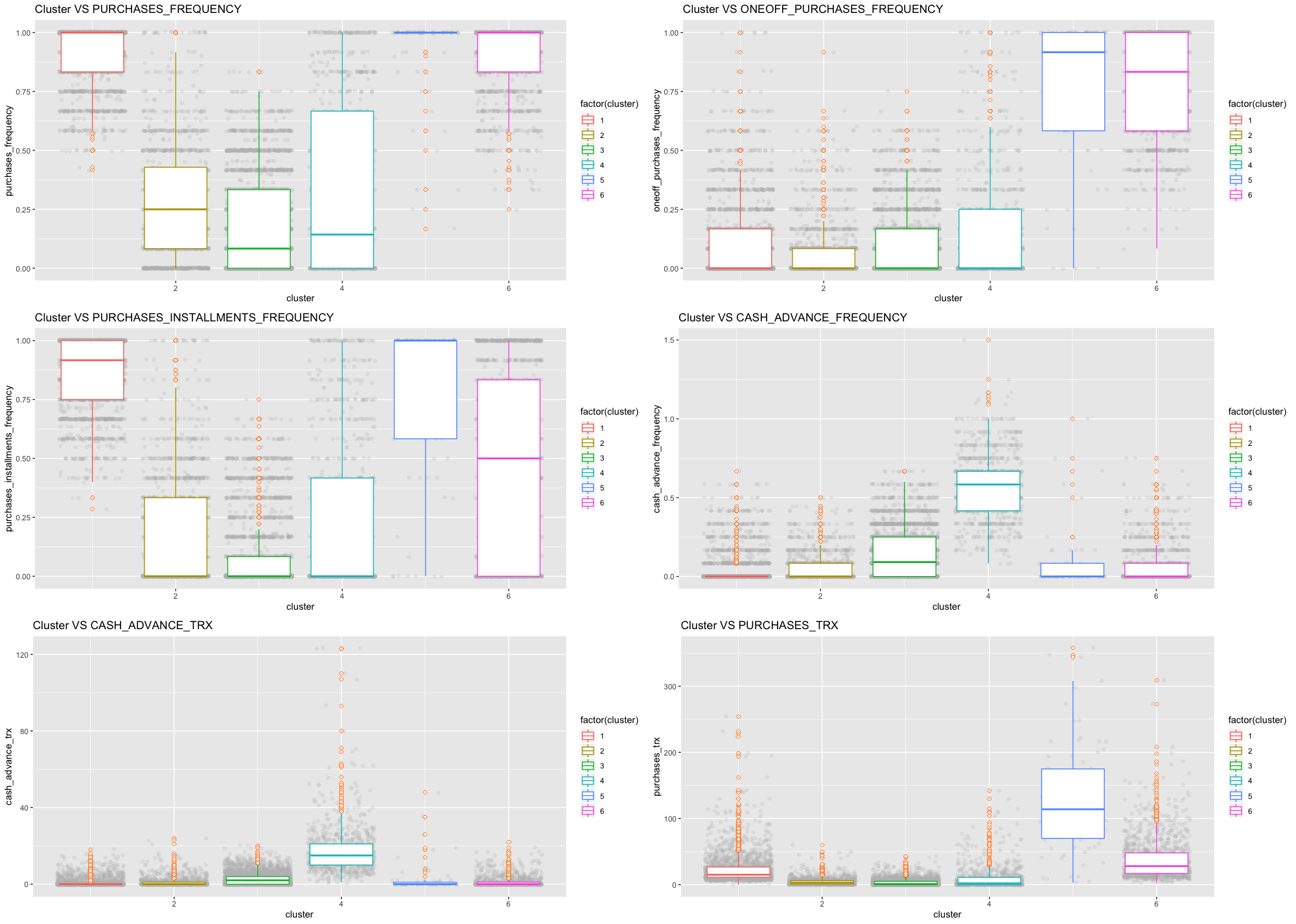
We reach alignment on our business objectives: what problem are we trying to solve with the segmentation, and how should we use the segmentation? When our objective is to attract new customers and drive revenues and growth, we should use the segmentation to structure not just product design but the whole acquisition process from customer targeting to segment positioning to delivery. That means starting with a needs-based segment, differentiated service with features to meet clients in each cluster. However, if we are more concerned with revenue leakage, it could increase the ratio of behavioral to attitudinal measures so as to optimize scoring efficiency rather than needs-based differentiation among the customers it is trying to retain. By building a richer, deeper view of customer segments, we can sharpen our value proposition, integrate cards with other lending products, reduce channel confusion and clarify positioning and promotions.

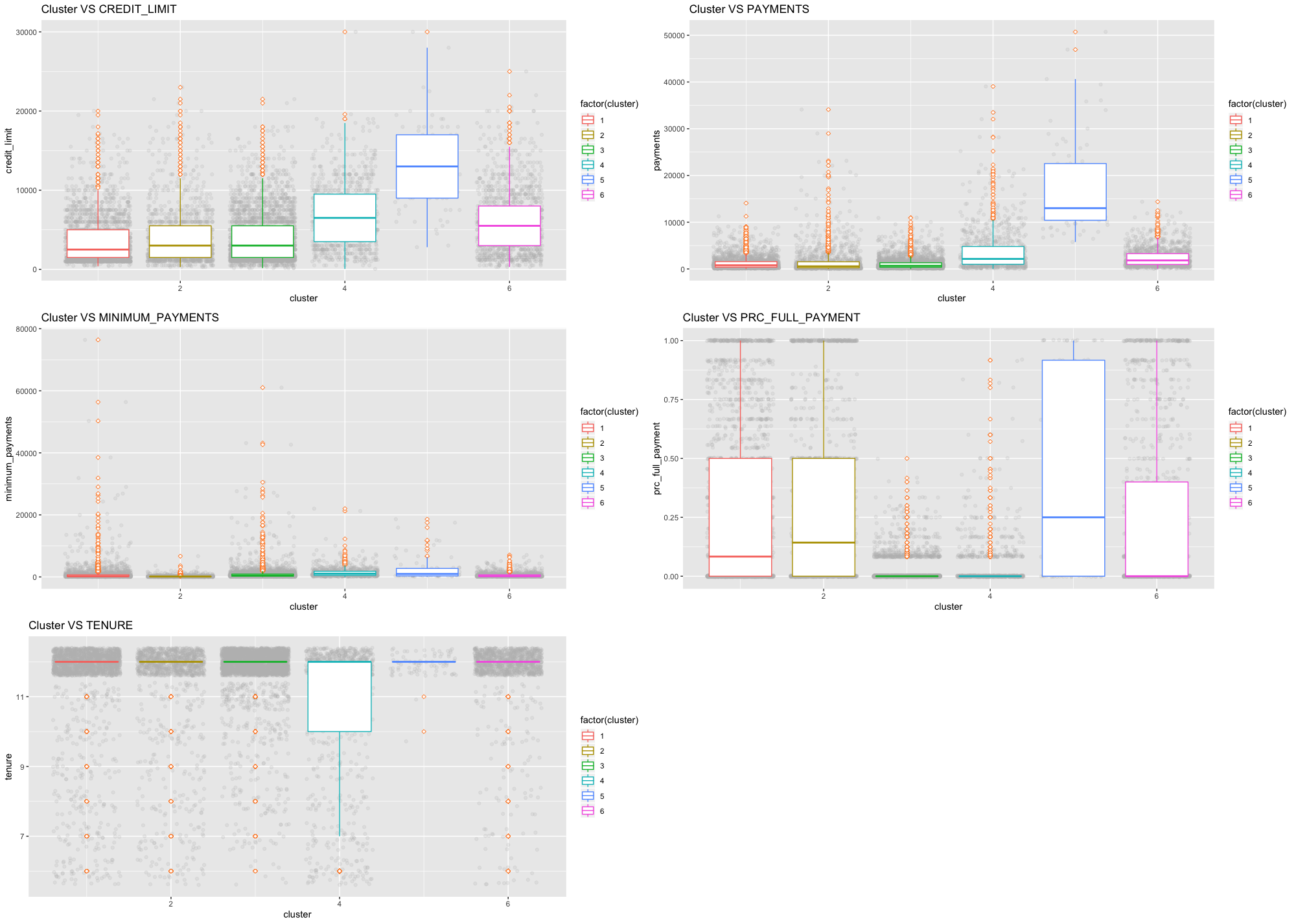
V. Appendix

Appendix 1: Data Dictionary

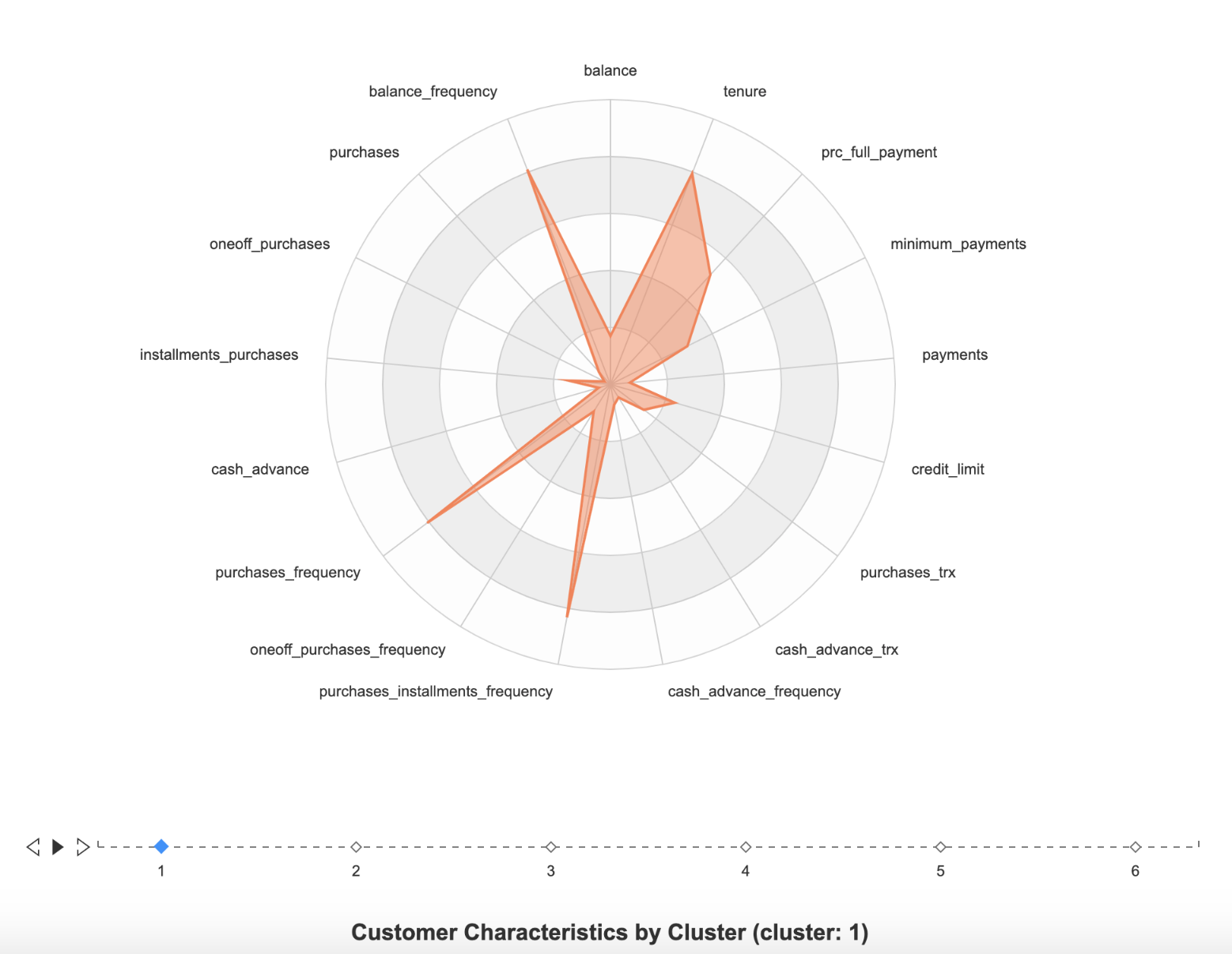
|  |  |  |
| --- | --- | --- |
| **Variable Names** | **Explanation** | **Data Format** |
| CUST\_ID | Identification of Credit Card holder | (character) |
| BALANCE | Balance amount left in their account to make purchases | (numeric) |
| BALANCE\_FREQUENCY | How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated) | (numeric) |
| TENURE | Tenure of credit card service for user | (numeric) |
| PURCHASES | Amount of purchases made from account | (numeric) |
| PURCHASES\_FREQUENCY | How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased) | (numeric) |
| PURCHASES\_TRX | Number of purchase transactions made | (numeric) |
| ONEOFF\_PURCHASES | Maximum purchase amount done in one-go | (numeric) |
| PURCHASES\_INSTALLMENTS\_FREQUENCY | How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done) | (numeric) |
| INSTALLMENTS\_PURCHASES | Amount of purchase done in installment | (numeric) |
| ONEOFF\_PURCHASES\_FREQUENCY | How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased) |  |
| CASH\_ADVANCE | Cash in advance given by the user | (numeric) |
| CASH\_ADVANCE\_FREQUENCY | How frequently the cash in advance being paid | (numeric) |
| CASH\_ADVANCE\_TRX | Number of Transactions made with "Cash in Advance" | (numeric) |
| CREDIT\_LIMIT | Limit of Credit Card for user | (numeric) |
| PAYMENTS | Amount of Payment done by user | (numeric) |
| MINIMUM\_PAYMENTS | Minimum amount of payments made by user | (numeric) |
| PRC\_FULL\_PAYMENT | Percent of full payment paid by user | (numeric) |

Appendix 2: Box Plots

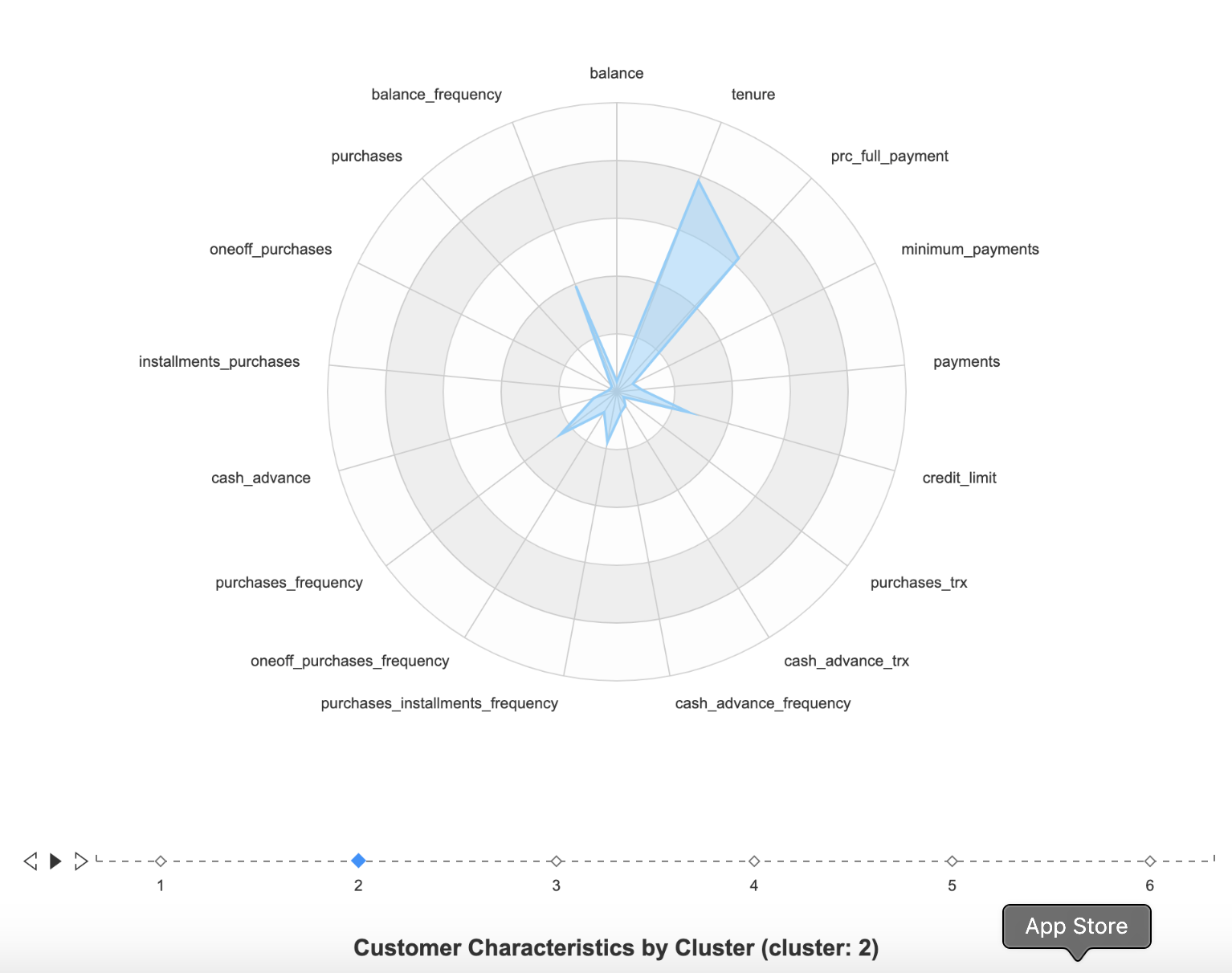




Appendix 3: Radar Plots



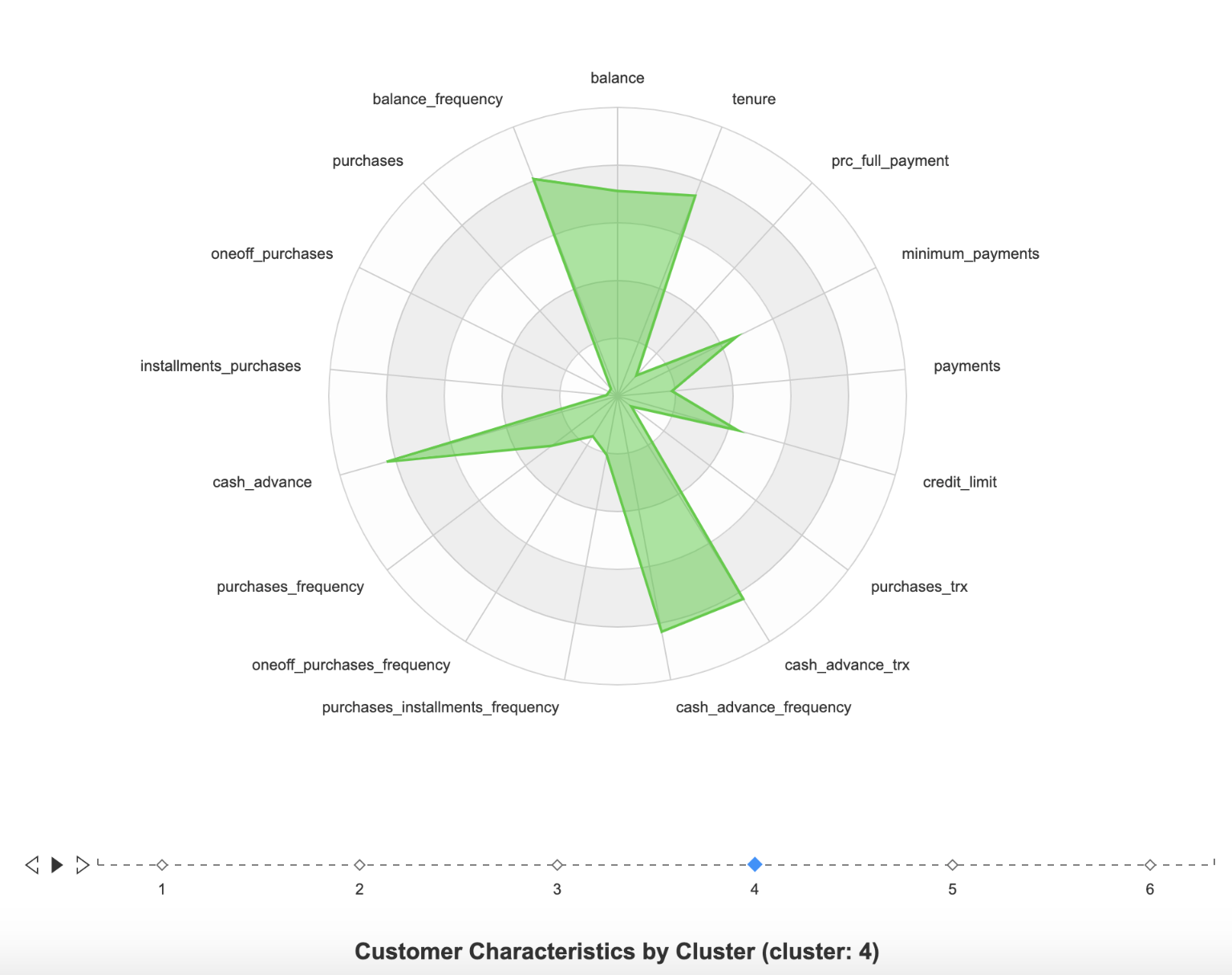
Cluster 1



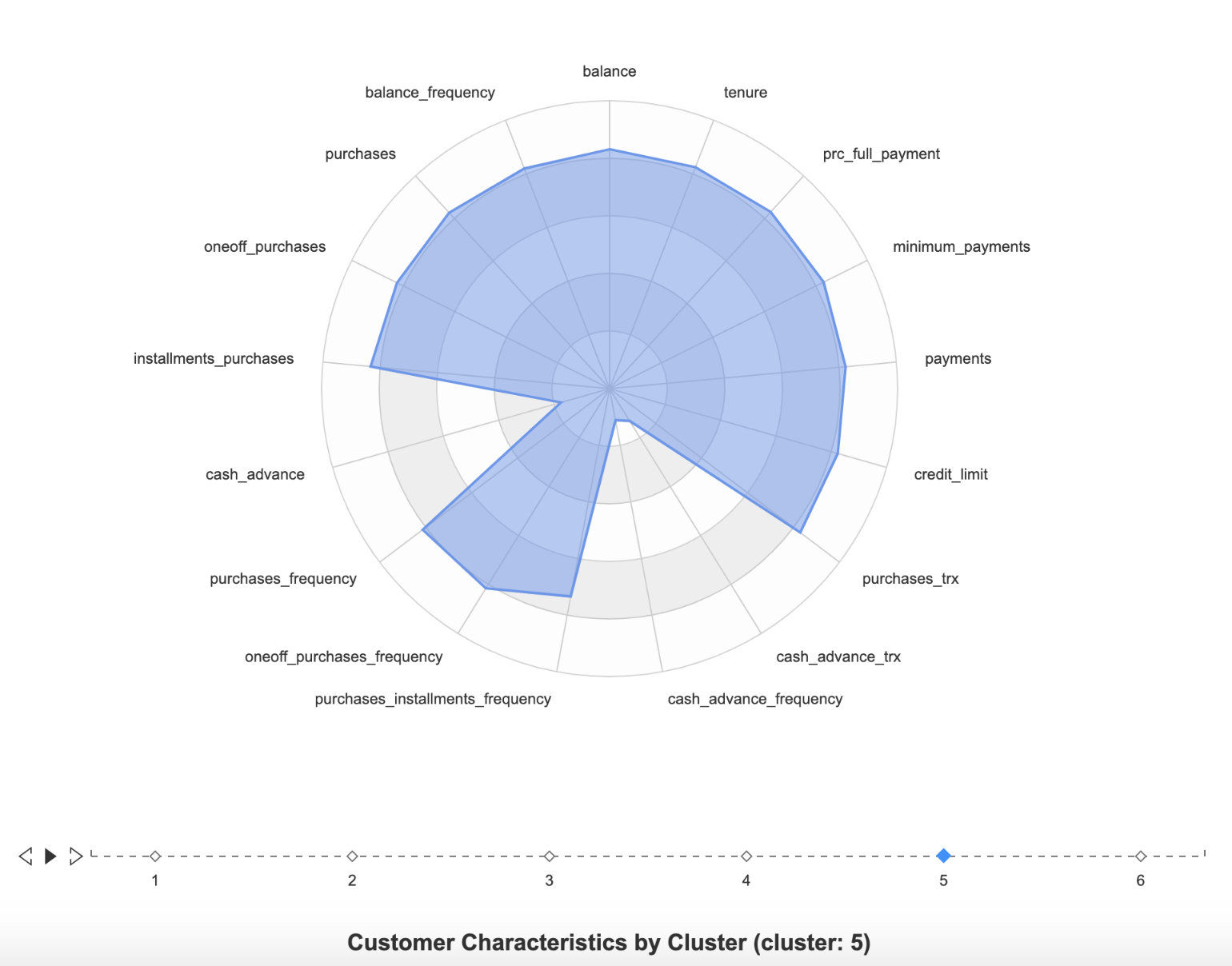
Cluster 2



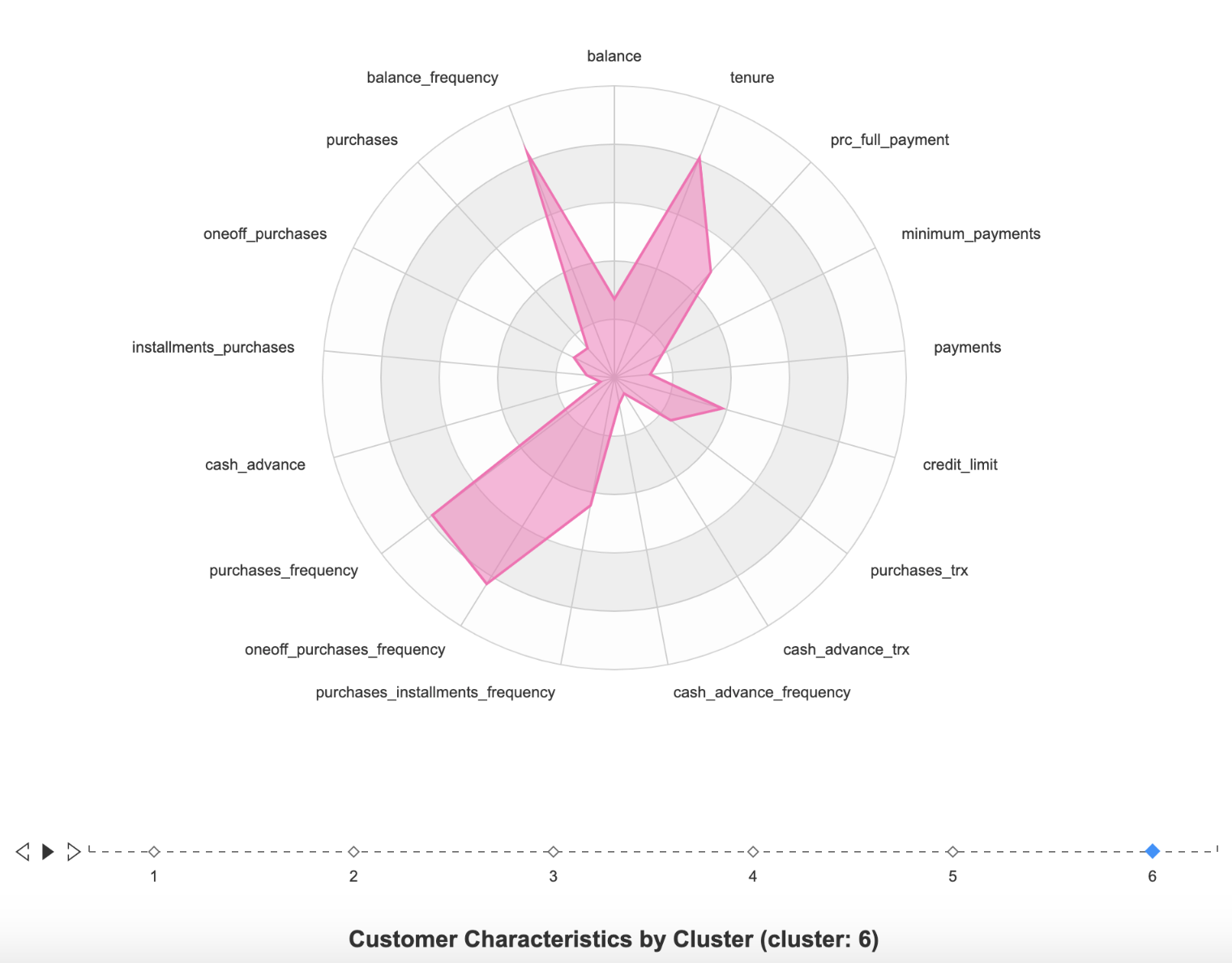
Cluster 3



Cluster 4



Cluster 5



Cluster 6

1. *Data Source:*[*https://www.kaggle.com/arjunbhasin2013/ccdata*](https://www.kaggle.com/arjunbhasin2013/ccdata) [↑](#footnote-ref-0)
2. https://www.finder.com/what-is-a-platinum-credit-card [↑](#footnote-ref-1)