Credit Card User Segmentation
Final Deliverable

12/09/2019

Cohort B Team 7
Kunpeng Huang, Yoki Liu, Lyufan Pan, Yunlei Zhou, Jiayuan Zou, Sherry Zuo

Business Problem

The purpose of this project is to explore the customer segmentation based on their purchasing behaviors with credit cards and to help the credit card company design different marketing strategies for various target customers based on characteristics explored of each cluster.

Dataset

Credit Card Dataset for Clustering
https://www.kaggle.com/arjunbhasin2013/ccdata

The sample Dataset summarizes the usage behavior of 8950 active credit card holders during a time length of 6 months. The file is at customer level with 18 variables.

Data Cleaning

In the dataset, there are 313 missing values in variable *minimum_payments*, and we used the mean of the *minimum_payments* to replace the missing values. Since *customer_id* is not useful in the analysis, we removed it from the dataset. After data cleaning, we have 17 numeric variables and 8950 observations.

Data Exploration

(Frequency: 0 = not frequent, 1 = very frequent)

- ***Balance:*** The remaining balance of majority users are below $5000 with an average of $1564.47.
- ***Balance frequency***: The majority of the cardholders are active and have a balance frequency around 1.
- ***Purchase frequency:*** Customer purchasing frequency is polarizing. Most credit card holders are concentrated on either not frequent or very frequent groups.
- ***One-off purchase***: The average of one-off purchase is $592.4, and most credit card holders use their credit cards on a routine purchasing basis.
- ***One-off purchase frequency***: Since these cards are widely used in daily life, the one-off purchase frequency is pretty low (mean = 0.083), which means customers were not tent to purchase expensive items.
- ***Installments purchase***: Credit card users are not relying on using their cards to purchase expensive items based on installment payment method (mean = $411.1).
- ***Purchases installments frequency***: 43.7% of cardholders are very inactive (frequency = 0) on using installments payment method, and 14.9% of cardholders are heavily relied on (frequency = 1) using the installments payment method.
- ***Cash advance***: Most users are not active in withdrawing cash. However, we find that a limited number of users have a large amount of money withdraw, and we need to care more about their purchasing behaviors because the primary goal of using credit is not for withdraw cash.
- ***Cash advance frequency:*** Most users are not frequently withdrawing cash using credit card.
- ***Credit limit:*** The money limit of credit card. Most users' credit limit is below $5,000.
- ***Percent full payment:*** The average percentage of full payment is pretty low (mean = 0.15), so we need to launch new strategies to encourage users to pay back the money as much as they can.

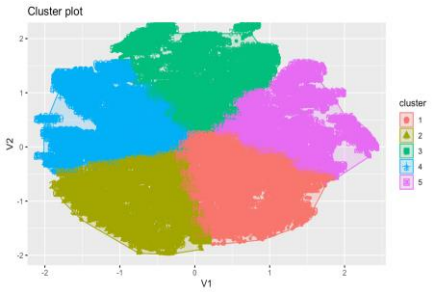
Analysis Methodology

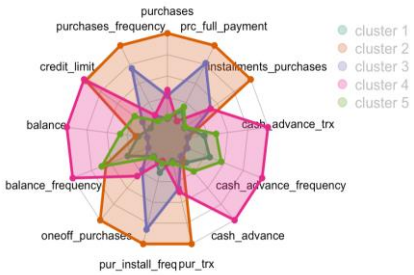
1. Baseline clustering Model (Kmeans)

- Silhouette score: choose cluster of 13 (7 and 9 also have very high scores, but it's similar trying different number of cluster)
- WSS: choose cluster of 2, 4, 7, 9, but the sum of square error is still too large.

Compare those clustering plots, we think k=9 is the best since each cluster has a similar size in the baseline clustering.

2. Dimension Reduction Model (t-SNE model + Kmeans)





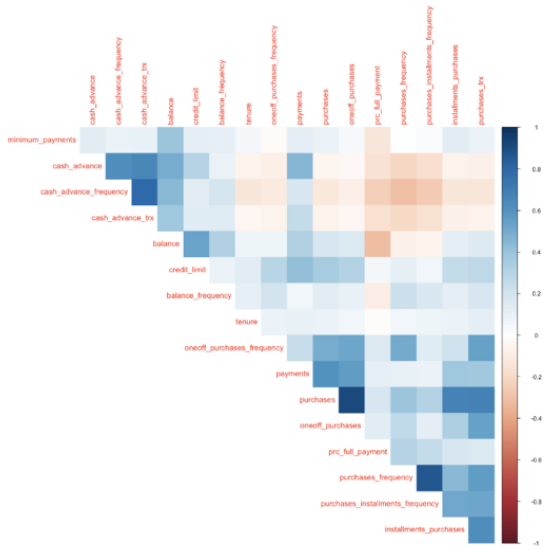
Based on the result of t-SNE model, we try k from 4 to 8 and finally get the clustering model with each of them.

Unfortunately, after getting the features of each cluster, we get the idea that features are not clear using t-SNE method, especially when we compare the features model with PCA later.

Here is the example of k=5. In the clustering plot, it is hard to find a clear division of groups, Kmeans just divides the entire two-dimensional space evenly. Plus, in the radar plot of each cluster, the unclear differences among groups make it difficult to describe customer characteristics in the real world.

3. Dimension Reduction Model (PCA model + Kmeans)

Based on the correlation plot colors, we think there are more than 4 groups that have high correlations in our dataset. Thus, we think we need to consider PCA dimension reduction.



Correlation plot of variables

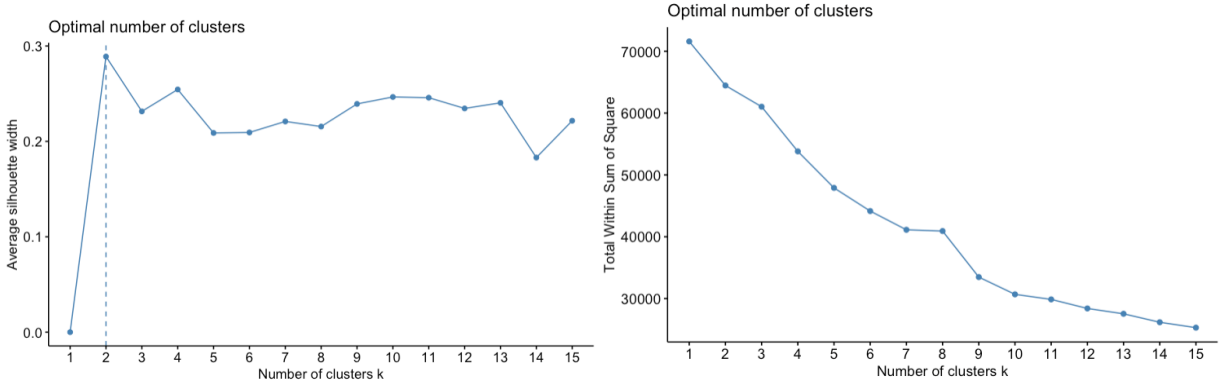
Compare **Eigenvalue** and **Cumulative Variance**

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.5	1.0646746466	6.262797e+00	70.12678
Dim.8	0.7237725420	4.257486e+00	85.02171

Based on eigenvalue, we first want to have all eigenvalues larger than 1, so Dimension with 5 meets the condition. However, dimension with 5 only gives us a low cumulative variance of 70%, so we choose **Dimension with 8**, which has a higher cumulative variance around 85%.

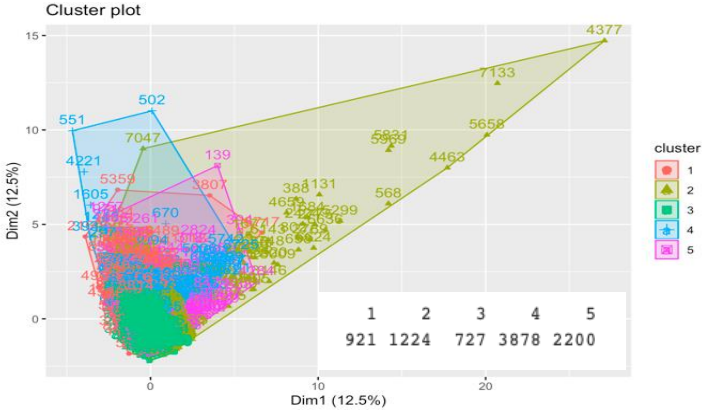
Clustering for PCA model

- Silhouette score - choose cluster of 2
- WSS - choose cluster of 2, 5, 7, 9



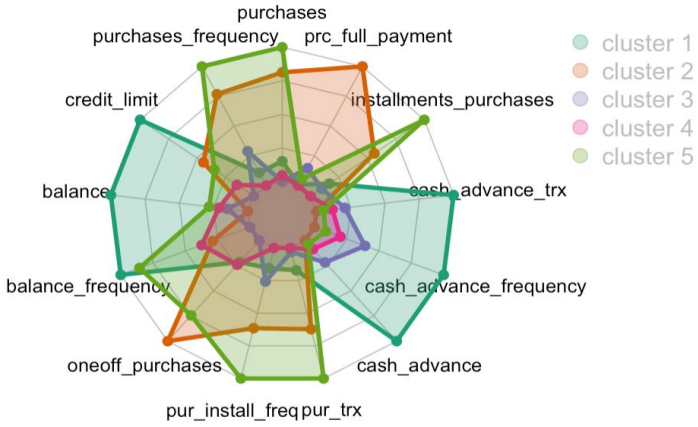
Since we don't want the size of cluster to be too small or large, so we try to average the size of clusters and we think **the best k is 5 for the PCA model**.

Our best model overall:



Thus, our best model is in **Dimension 8 and 5 Clusters**.

Performance Evaluation



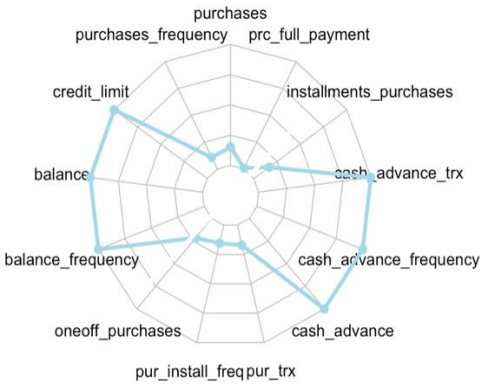
Add clustering back to the original dataset, compare variables for clusters based on different **boxplots** and **radar charts**, and summarize the features of each cluster and define the characteristics of each cluster of card users.

- **Features:**
    - Cluster 1: high credit, low purchase, no installment, high cash advance
    - Cluster 2: median credit, no cash advance, high purchase, high purchase frequency, high installment and like to run out of credit but then pay full amount back
    - Cluster 3: low credit, try every feature of the card
    - Cluster 4: low credit, low activities in every feature
    - Cluster 5: median credit, high purchase, high purchase frequency, high installments, but rarely pay full amount back

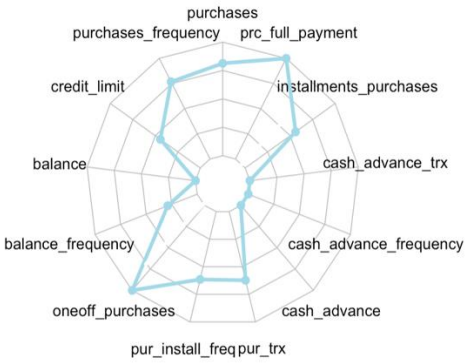
- **Characteristics:**
    - Cluster 1 (921): long-time users for daily purchases and like cash in advance
    - Cluster 2 (1224): users live from paycheck to paycheck
    - Cluster 3 (727): new card users
    - Cluster 4 (3878): inactive cardholders
    - Cluster 5 (2200): installment users with high spending behaviors

Marketing Strategies Recommendation

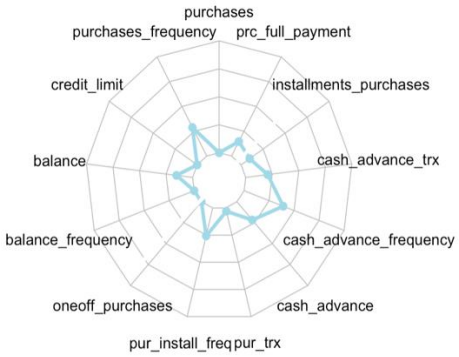
Based on those 5 features, we design different market strategies for various target customers.



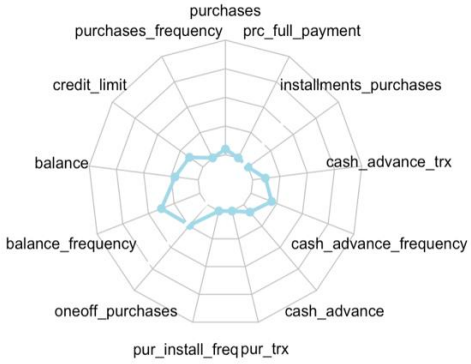
Cluster1:
long-time users for daily purchases and like cash in advance
Generalize cashback focused credit card. Having policies like unlimited 1% cashback on every purchase or on up to a certain amount in combined purchases in bonus categories (like gas stations, grocery stores, and etc.) each quarter. Customers in cluster1 that use credit cards for small purchases and daily purchases are usually very economical and will be attracted by adding on benefits.



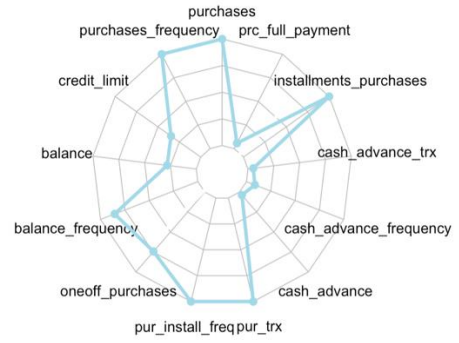
Cluster2:
users live from paycheck to paycheck
Generalize bonus point focused credit card with a certain amount of annual fee. Having policies like earning a certain amount of bonus points after the customer spends $3,000 on purchases in the first three months from account opening, and that allows them to redeem the points for shopping, travel, and etc. Furthermore, if the customers spend more on traveling, we can also have policies like $300 annual travel credit as reimbursement or double points on travel expenses. Customers in cluster 2 with high spending behaviors care more about their living quality and are willing to spend money on enjoyment.



Cluster3:
new card users
Generalize special incentives for new card users. Sending out introduction mails to give them a brief summary of how credit cards work and target them with titles like “pre-approved”, “benefits cards” and lead them to keep using the cards. Offer 3% cashback on selected category and let the customers choose which category they would like to activate. The majority of the new card users are students and targeting their own spending interest will be a good way to make them form the habit of using credit cards.



Cluster4:
inactive cardholders
Generalize a comprehensive credit card to target inactive cardholders. People who don’t use credit cards might prefer to use debit cards to prevent from spending too much money or have lots of credit cards to choose from. An all-sided credit card may qualify their needs and attract them to use our credit card more often than others. For example, like no annual fee, free or low annual percentage rate, and unlimited 2-3% cashback on all purchases. Customers in cluster 4 are difficult to perceive and offering an all-sided strategy will be an appropriate way to target them.



Cluster5:
installment users with high spending behavior
Generalize benefits like priority airport lounges, room upgrades, and special gifts for purchases focused credit cards but with a certain amount of annual percentage rate required. We may target installment users with free APR for the first six months, and offering free balance transfer functions to satisfy their needs. Customers in cluster 5 usually spend a lot but prefer to do installments, thus premium benefits with limited time of free APR will be exactly what they need when applying for credit cards.