# Clustering Credit Card Users

By Tom Ribaroff


# What is Clustering?

Clustering is a Machine Learning technique, which allows us to group similar objects/people together into ‘clusters’.

We use the power of the computer to make groups in which objects are as similar as possible.

By making these groups, we can define targeted market strategies for each, that will be more effective/profitable than implementing changes to the entire group or our own arbitrarily made ‘clusters’.

# Our Data and Preparations

| | CUST_ID | BALANCE | BALANCE_FREQUENCY | PURCHASES | ONEOFF_PURCHASES | INSTALLMENTS_PURCHASES | CASH_ADVANCE |
|---|---|---|---|---|---|---|---|
| 0 | C10001 | 40.900749 | 0.818182 | 95.40 | 0.00 | 95.4 | 0.000000 |
| 1 | C10002 | 3202.467416 | 0.909091 | 0.00 | 0.00 | 0.0 | 6442.945483 |
| 2 | C10003 | 2495.148862 | 1.000000 | 773.17 | 773.17 | 0.0 | 0.000000 |
| 3 | C10004 | 1666.670542 | 0.636364 | 1499.00 | 1499.00 | 0.0 | 205.788017 |
| 4 | C10005 | 817.714335 | 1.000000 | 16.00 | 16.00 | 0.0 | 0.000000 |

We have a dataset of ~9000 credit card users, containing information like spending history, credit scores, length of time with the bank and more.

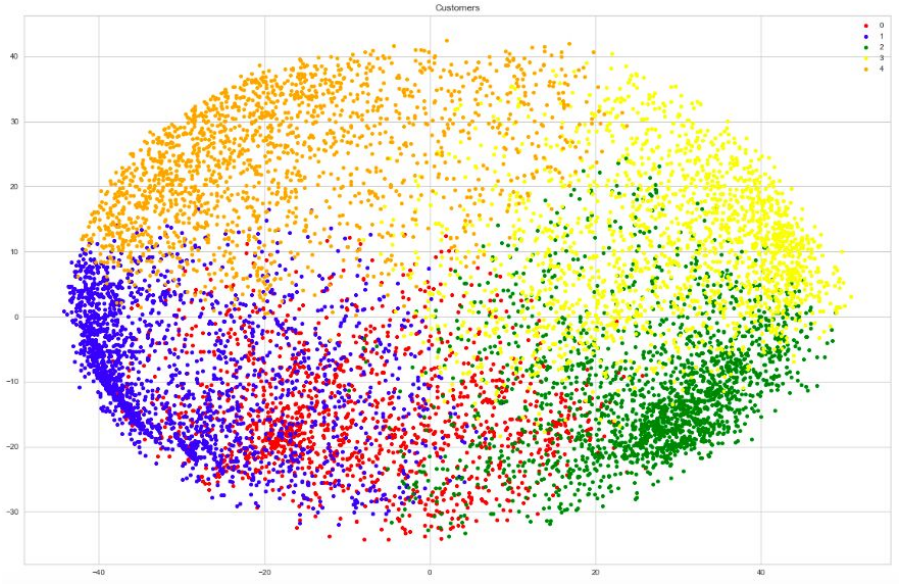
After fixing any missing errors in the data, we use the ‘K Means Algorithm’ to make the best possible clusters.

K Means is an algorithm that helps us to decide what is the best number of groups to make from our data.

Our definition of ‘best’ here is the making the number of groups where each groups has to lowest variance in the data inside the group, while still trying to make each group vary from each other.

We will therefore find similar groups of credit card users, where each group is distinct, as we have made the groups vary from each other as much as possible too.

# Results




We found the best way to group these users was to create 5 clusters.

Each of these clusters had defining features which we explored, so that we can better understand why the computer grouped them together.

So, what are the defining features of each cluster?

So, what are the defining features of each cluster?

0 - Very Low Balance Frequency (ie people who are least careful about their bills/debts. Don't plan their purchases)

1 - Smallest Number of Purchases and Smallest Sum of Purchases (ie least active users of the card)

2 - Highest Installment Purchases Frequency (ie likely low income/savings, living paycheck to paycheck)

3 - Largest Number of Purchases and Largest Sum of Purchases (ie most active users of the card)

4 - Highest Cash Advance Transactions (ie people who are most careful about their bills/debts. Planners)

# Business Recommendations

Now that we have the clusters, we can define market strategies for each!

For example, we could target advertising to whatever demographics are represented in Cluster 3, as this cluster is likely to be the customers which are most profitable.

Cluster 2 looks like it could very unprofitable in general, so it would be worth exploring what predictors there are for people in that group, and to try turn down more credit card applications for those who are likely to end up being part of that cluster down the line.