**EdWisor Project-2**

**Credit Card Segmentation**

**By- Ayush Raj**

**Aim**- To divide the customers into different groups or clusters and define different marketing strategy for each group. Each group would be approached in a different way to increase the conversion rate/sales.

Understanding the Features

1. The train dataset consists of credit card based behavioural data of approx. 9000 credit card holders in the last 6 months.
2. The dataset is at the customer level with 18 behavioural features

The features are discussed below-

1. CUST\_ID- Identification of the credit card holder
2. BALANCE- Amount left in the credit card balance to make purchases
3. BALANCE\_FREQUENCY- It is the ratio of the number of days the account had credit balance to make further purchase.
4. PURCHASES- Total money spent by the customer in purchasing something using the credit card in the last 12 months.
5. ONEOFF\_PURCHASES- Maximum money spent on a purchase in one go or at one time.
6. INSTALLMENTS\_PURCHASES- Amount of purchase done in instalments.
7. CASH\_ADVANCE- Total cash in advance given by the user before the billing date
8. PURCHASES\_FREQUENCY- How frequently the purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)
9. ONEOFF\_PURCHASES\_FREQUENCY- How frequently Purchases are happening in one-go (and not in instalments) (1 = frequently purchased, 0 = not frequently purchased)
10. PURCHASES\_INSTALLMENTS\_FREQUENCY- How frequently purchases in instalments are being done (1 = frequently done, 0 = not frequently done)
11. CASH\_ADVANCE\_FREQUENCY- How frequently the cash in advance is being paid
12. CASH\_ADVANCE\_TRX- Number of Transactions made with "Cash in Advanced"
13. PURCHASES\_TRX- Number of purchase transactions made
14. CREDIT\_LIMIT- Limit of Credit Card for the user
15. PAYMENTS- Total Payments (due amount paid by the customer to decrease their statement balance) within the time frame
16. MINIMUM\_PAYMENTS – Minimum amount of payment made by the user (minimum value of the payment) (EdWisor- **Total minimum payment due in the period**)
17. PRC\_FULL\_PAYMENT- Percent of months with full payment of the due statement balance.
18. TENURE- Number of months as a customer

Hypothesis and questions

Hypothesis:

Our aim as a credit card provider is to increase use and spending through credit cards. **Our best customer is the one who is in debt and pays off the debt incrementally, only the minimum amount required per month is paid and carries the debt forward to the next month**.

We want our customers to

1. Use up the complete credit card BALANCE.
2. Spend as much amount on PURCHASES as possible.
3. Make more INSTALLMENTS\_PURCHASES.
4. Reduce the CASH\_ADVANCE payments (Such customers can spend more)
5. Increase the PURCHASE\_FREQUENCY ()
6. Increase the ONEOFF\_PURCHASE\_FREQUENCY and PURCHASE\_INSTALLMENT\_FREQUENCY (So that the customer uses the complete balance available to him/her)
7. Reduce CASH\_ADVANCE\_FREQUENCY
8. Reduce CASH\_ADVANCE\_TRX (a high value means that a person is eager to reduce debt. We want to lure him to spend more rather than paying the debt)
9. Increase the PURCHASE\_TRX
10. Have a PAYMENTS value equal to MINIMUM\_PAYMENTS
11. Never make full payments i.e. have a PRC\_FULL\_PAYMENT value equal to zero.

A customer with all he above indicators is an idea customer.

(Source: <https://money.howstuffworks.com/credit-card-debt.htm> <https://money.howstuffworks.com/personal-finance/debt-management/revolving-credit.htm> )

**Marketing Strategy**:

First, we would want to divide our customers into three groups/segments:

1. An ideal customer (as listed above). We will be calling them as “Customer Category-1” (CC-1)
2. A customer who is paying more money in advance, having a higher balance, making less purchases, paying more than minimum payment and having a high value for PRC\_FULL\_PAYMENT and paying much more than the MINIMUM\_PAYMENTS should belong to “CUSTOMER CATEGORY-2” (CC-2)
3. A person who is making PAYMENTS less than the MINIMUM\_PAYMENTS “CUSTOMER CATEGORY-3” (CC-3)

**CC-1** customers are the ideal customers. Because of their good performance we can target them to increase their CREDIT\_LIMIT and give them some more good offers so that they spend more.

**CC-2** customers are the ones who have a capacity to spend significantly more than what they are currently doing. Such customers should be given more generous offers and their credit limits can also be increased by a larger value

**CC-3** customers are the ones who are more likely to default. Offers available to them should be made less generous or stopped. In case or poor performance their CREDIT\_LIMIT should be decreased so that they are able to pay the MINIMUM\_PAYMENTS required each month.

**Data Pre-processing**

1. CREDIT\_LIMIT has one null value and ‘MINIMUM\_PAYMENTS’ has 313 null values.
2. These values are imputed by the mean
3. Different features have values spread across different scales which can impact the clustered model so, we scaled the values of all the features so that the mean is ‘0’ and the standard deviation is ‘1’

**Model-1:** **Applying K-Means Algorithm on scaled data**

1. K-means algorithm is applied. Within Cluster Sum of Squares is calculated for clusters ranging from 1 to 15. The Elbow curve obtained is shown below

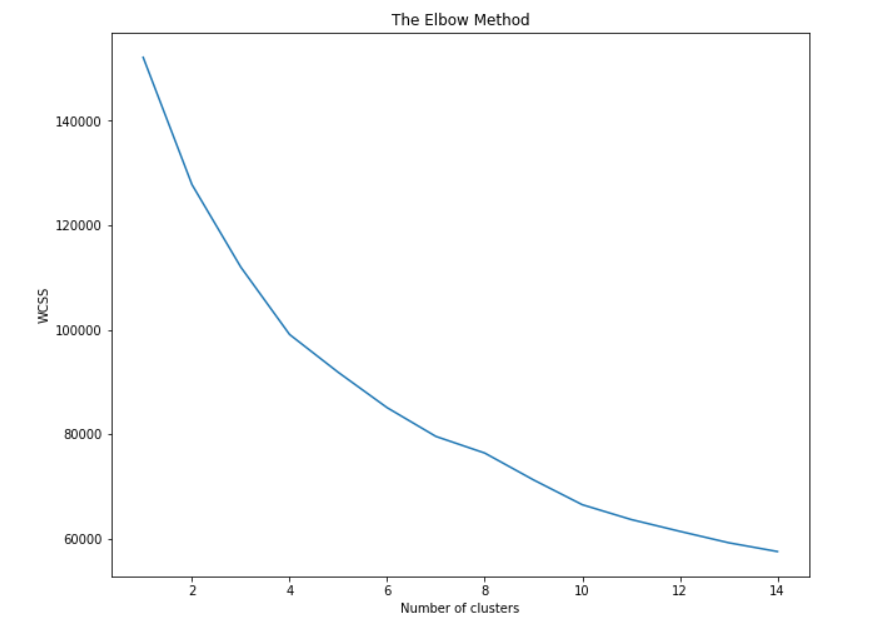


Fig1: Elbow curve showing change in ‘WCSS’ with the number of clusters

1. As we can see, it is difficult to estimate the optimal number of clusters from the above curve. So, we use the ‘Silhouette Method’ to estimate the optimal number of clusters.
2. As can be seen from the figure-2, the ‘Silhouette Method’ shows that the when three clusters are chosen, the silhouette coefficient has the maximum value.
3. So, we move ahead with ‘3’ Clusters.

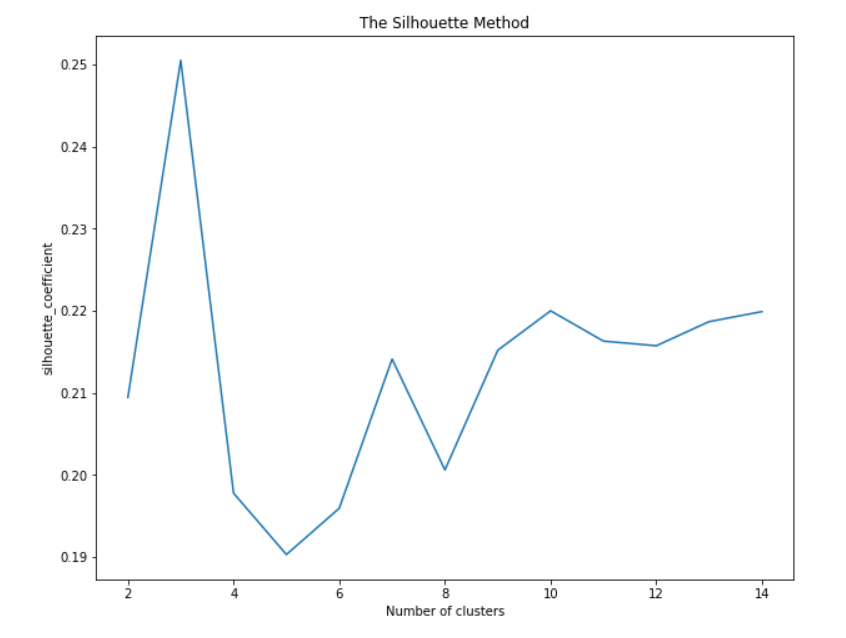


Fig-2: Silhouette Method showing the variation in silhouette coefficient with the number of clusters

1. The clusters created are visualised using t-SNE. The data is reduced to two dimensions



Fig-3: Visualizing the clustered data in two dimensions using t-SNE

1. Figure-3 shows the clustered data that has been reduced to two dimensions using t-SNE
2. We observe that the clusters formed overlap in some areas
3. The same clustered data is visualised using PCA after being reduced into two dimensions. This can be seen in figure-4



Fig-4: Visualizing clustered data using PCA

1. After performing PCA we observe that as was concluded using t-SNE, the clusters seem to overlap.

### Model-2: Reducing the dimensionality of data and then applying K-means

1. We reduce the dimensions of data using PCA such that about 85% of the variance of the original dataset is preserved. This leads us to reducing the data to 8 dimensions.
2. This data is then clustered into three clusters using K-means algorithm.
3. The clustered data is then visualised using t-SNE as shown in the figure-5

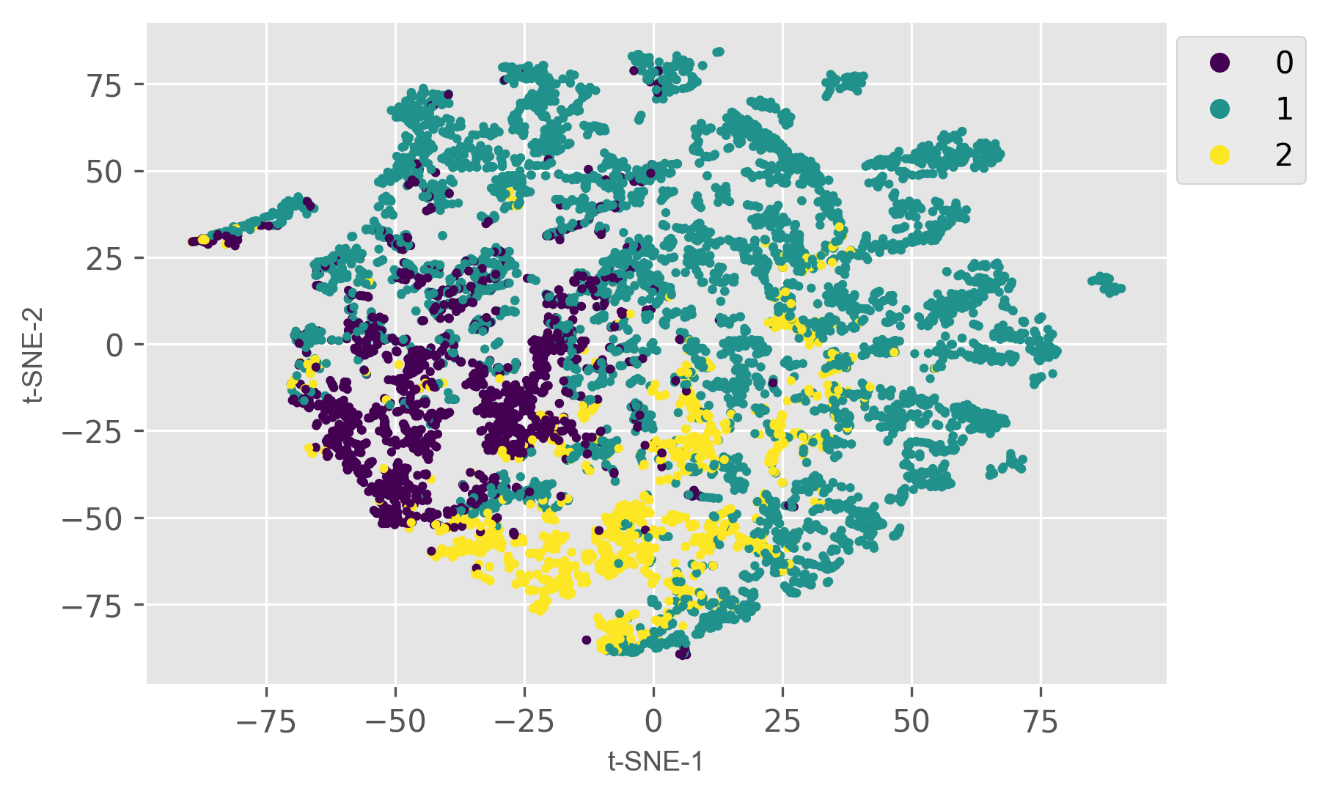


Fig-5: Visualizing the clustered data in two dimensions using t-SNE

1. As can be seen from the figure above these clusters have better defined boundaries than the ones created using the first method.
2. The same data is visualised using PCA where the data is reduced to two dimensions. This can be seen in figure-6.

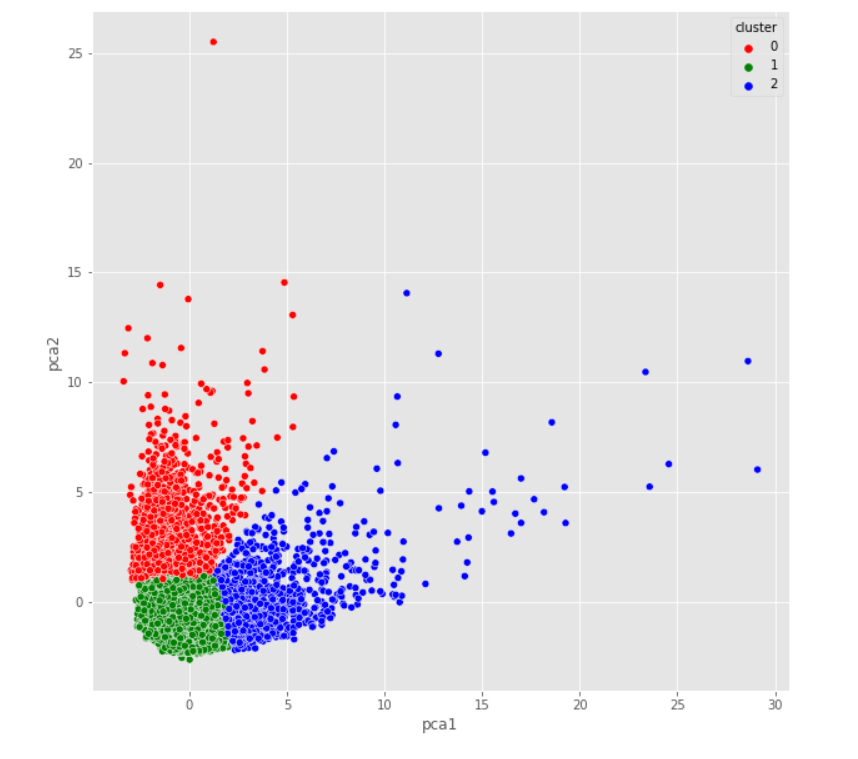


Fig 6: Visualizing clustered data using PCA

1. As can be seen from the figure above, this clustering approach led to better, well defined clusters. This model is accepted as it gives the best results.

### Model-3: Using Gaussian Mixture Clustering Algorithm on scaled data

### The scaled data is clustered into three clusters using the Gaussian Clustering Algorithm.

### We visualise the clustered data using t-SNE as shown in figure-7.

### We see that the clusters created are not as well defined as was in method-2.

### Figure-8 shows the visualisation of the clusters created using PCA. It can be observed that the datapoints from different clusters do not have a very well defined boundry

### 

Fig 7: Visualizing the clustered data in two dimensions using t-SNE

### 

### Fig-8: Visualizing clustered data using PCA

**Model-4: Using Gaussian Clustering Algorithm on data that has been reduced in dimensions using PCA**

1. The data is reduced in dimension using PCA so as to retain 85% of the variance of the original dataset
2. Gaussian Clustering Algorithm is applied to create 3 clusters.

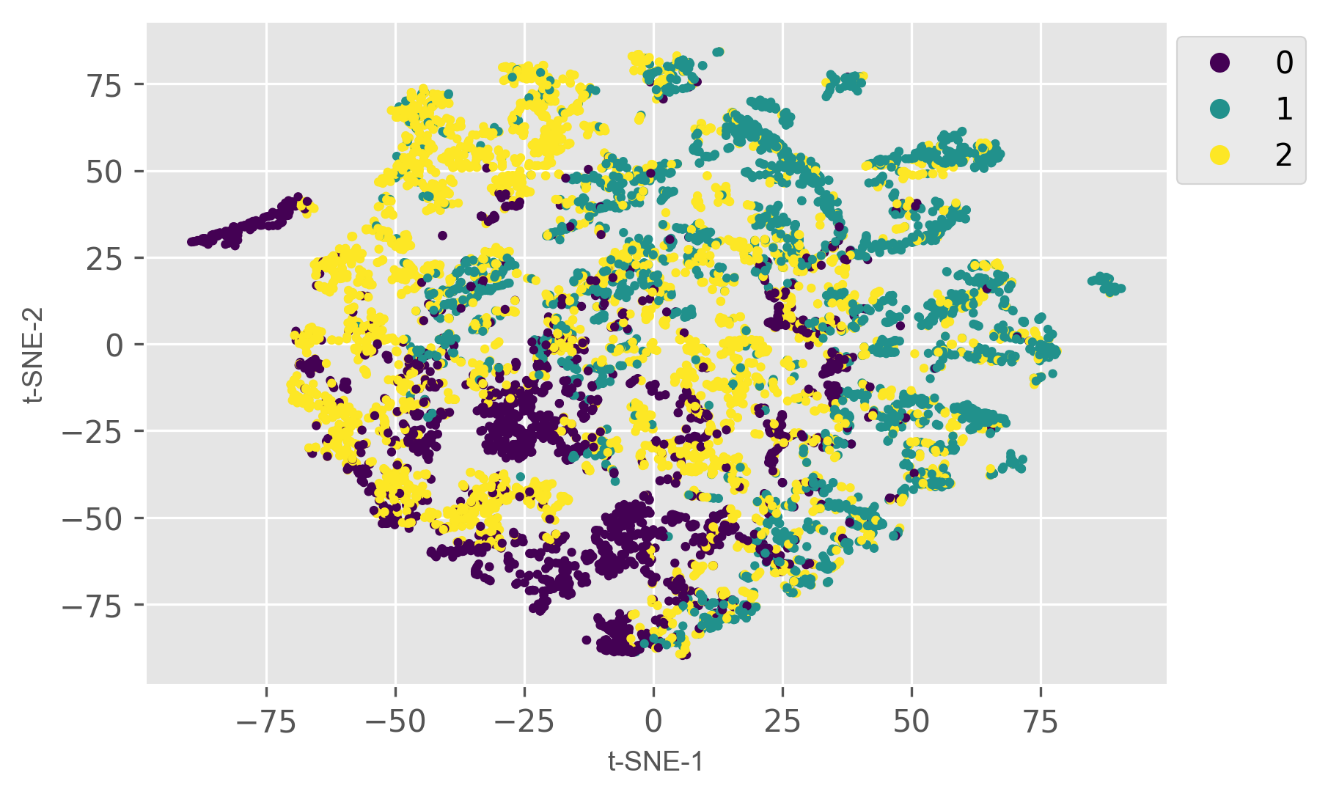


Fig-9: Visualizing the clustered data in two dimensions using t-SNE

1. As can be seen in figure-9, the clusters created do not have defined, distinct boundaries.
2. The same data is visualised using PCA after reducing it to two dimensions.
3. As can be seen in figure-10, the clusters are not well defined

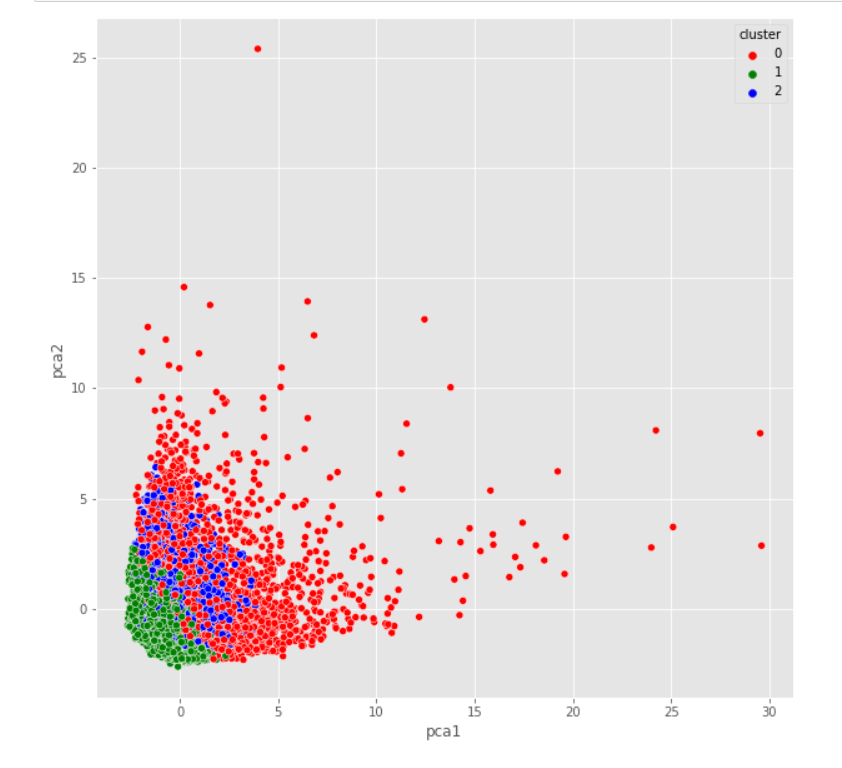


Fig-10: Visualizing the clustered data using PCA