Credit Card Customer Segmentation

Aditya Kapoor

Problem Statement

Objective

To develop a customer segmentation based on their credit card usage in order for the party to define a marketing strategy.

Data

We are given a Dataset of **8950 customers** along with **18 behavioural** features.

Variables:

1. CUST_ID: Unique Customer Identifier
2. BALANCE: Average Balance (Based on Daily Average Balances)
3. BALANCE FREQUENCY: Ratio of last 12 months with balance
4. PURCHASES : Net Dollar Value of Purchases in last 12 months
5. ONEOFF_PURCHASES: Amount of ONEOFF Purchases
6. INSTALLMENTS_PURCHASES: Total amount of instalment purchases
7. CASH_ADVANCE: Cash Withdrawn From Credit Card (Assume a period of 12 months)
8. PURCHASE_FREQUENCY: Percentage of months with at least one purchase
9. ONEOFF_PURCHASES_FREQUENCY: Freq of ONEOFF purchases
10. PURCHASES_INSTALLMENTS_FREQUENCY: Frequency of Instalment Purchases
11. CASH_ADVANCE_FREQUENCY: Cash-Advance frequency
12. CASH_ADVANCE_TRX: Average amount per cash-advance transaction
13. PURCHASES_TRX: Average amount per purchase transaction
14. CREDIT_LIMIT: Credit limit
15. PAYMENTS: Total payments (due amount paid by the customer to decrease their statement balance) in the period
16. MINIMUM_PAYMENTS: Total minimum payments due in the period.
17. PRC_FULL_PAYMENT: Percentage of months with full payment of the due statement balance
18. TENURE: Number of months as a customer

Approach

Data Exploration

We will first look at the structure of the dataset, its features and the information it conveys.

```
  CUST_ID      BALANCE  BALANCE_FREQUENCY  PURCHASES  ONEOFF_PURCHASES  \
0  C10001    40.900749           0.818182      95.40              0.00   
1  C10002  3202.467416           0.909091       0.00              0.00   
2  C10003  2495.148862           1.000000     773.17            773.17   
3  C10004  1666.670542           0.636364    1499.00           1499.00   
4  C10005   817.714335           1.000000      16.00             16.00   

   INSTALLMENTS_PURCHASES  CASH_ADVANCE  PURCHASES_FREQUENCY  \
0                    95.4      0.000000             0.166667   
1                     0.0   6442.945483             0.000000   
2                     0.0      0.000000             1.000000   
3                     0.0    205.788017             0.083333   
4                     0.0      0.000000             0.083333   

   ONEOFF_PURCHASES_FREQUENCY  PURCHASES_INSTALLMENTS_FREQUENCY  \
0                    0.000000                          0.083333   
1                    0.000000                          0.000000   
2                    1.000000                          0.000000   
3                    0.083333                          0.000000   
4                    0.083333                          0.000000   

   CASH_ADVANCE_FREQUENCY  CASH_ADVANCE_TRX  PURCHASES_TRX  CREDIT_LIMIT  
0                0.000000                 0              2        1000.0  
1                0.250000                 4              0        7000.0  
2                0.000000                 0             12        7500.0  
3                0.083333                 1              1        7500.0  
4                0.000000                 0              1        1200.0  

      PAYMENTS  MINIMUM_PAYMENTS  PRC_FULL_PAYMENT  TENURE  
0   201.802084        139.509787          0.000000      12  
1  4103.032597       1072.340217          0.222222      12  
```

First few records of the dataset

We would now like to check if any of the features has NAN values. If the number of NAN values is not much (<20%) we could easily impute them with mean/median.

On running the code to check for NAs, we get the following result:

```
CUST_ID                             0
BALANCE                             0
BALANCE_FREQUENCY                   0
PURCHASES                           0
ONEOFF_PURCHASES                    0
INSTALLMENTS_PURCHASES              0
CASH_ADVANCE                        0
PURCHASES_FREQUENCY                 0
ONEOFF_PURCHASES_FREQUENCY          0
PURCHASES_INSTALLMENTS_FREQUENCY    0
CASH_ADVANCE_FREQUENCY              0
CASH_ADVANCE_TRX                    0
PURCHASES_TRX                       0
CREDIT_LIMIT                        1
PAYMENTS                            0
MINIMUM_PAYMENTS                  313
PRC_FULL_PAYMENT                    0
TENURE                              0
dtype: int64
```

We see that Minimum Payments Column has 313 NaN values, which is a small percentage of total rows in the dataset. Hence, we go ahead and impute them with the median.

Let us now describe the data. We would like to know how values in each of the features is distributed.

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	\
count	8950.000000	8950.000000	8950.000000	8950.000000	
mean	1564.474828	0.877271	1003.204834	592.437371	
std	2081.531879	0.236904	2136.634782	1659.887917	
min	0.000000	0.000000	0.000000	0.000000	
25%	128.281915	0.888889	39.635000	0.000000	
50%	873.385231	1.000000	361.280000	38.000000	
75%	2054.140036	1.000000	1110.130000	577.405000	
max	19043.138560	1.000000	49039.570000	40761.250000	

	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	\
count	8950.000000	8950.000000	8950.000000	
mean	411.067645	978.871112	0.490351	
std	904.338115	2097.163877	0.401371	
min	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.083333	
50%	89.000000	0.000000	0.500000	
75%	468.637500	1113.821139	0.916667	
max	22500.000000	47137.211760	1.000000	

	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLMENTS_FREQUENCY
count	8950.000000	8950.000000
mean	0.202458	0.364437
std	0.298336	0.397448
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.083333	0.166667
75%	0.300000	0.750000
max	1.000000	1.000000

	CASH_ADVANCE_FREQUENCY	CASH_ADVANCE_TRX	PURCHASES_TRX	CREDIT_LIMIT
count	8950.000000	8950.000000	8950.000000	8950.000000
mean	0.135144	3.248827	14.709832	4494.282473
std	0.200121	6.824647	24.857649	3638.646702
min	0.000000	0.000000	0.000000	50.000000
25%	0.000000	0.000000	1.000000	1600.000000
50%	0.000000	0.000000	7.000000	3000.000000
75%	0.222222	4.000000	17.000000	6500.000000
max	1.500000	123.000000	358.000000	30000.000000

	PAYMENTS	MINIMUM_PAYMENTS	PRC_FULL_PAYMENT	TENURE
count	8950.000000	8950.000000	8950.000000	8950.000000
mean	1733.143852	844.906767	0.153715	11.517318
std	2895.063757	2332.792322	0.292499	1.338331
min	0.000000	0.019163	0.000000	6.000000
25%	383.276166	170.857654	0.000000	12.000000
50%	856.901546	312.343947	0.000000	12.000000
75%	1901.134317	788.713501	0.142857	12.000000
max	50721.483360	76406.207520	1.000000	12.000000

Distribution of features in the dataset

Deriving Useful KPIs

We would now like to derive certain key performance indicators (KPIs) which might explain our data in a better way. Here is a list of KPIs we have derived from the existing features.

KPIs we are interested in:

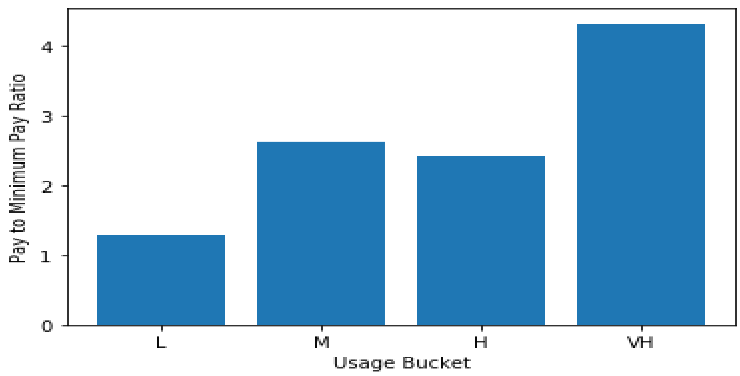
1. Usage By Type: We would like to know if the customer uses his card for cash, one off purchases, instalments or a combination of the three. It has 8 categories.
    a. Cash Only (C)
    b. One Off Purchases Only (O)
    c. Instalments Only (I)
    d. Instalments and One Off Purchases (OI)
    e. Instalments and Cash Purchases (IC)
    f. One off Purchases and Cash (OC)
    g. One off, Instalments and Cash (OIC)
    h. No Transaction (NONE)

2. Payments to Minimum Payments Ratio: Payments made divided by the Minimum Payment the user is required to make.
3. Balance to Credit Limit Ratio: At an average, what part of the given credit limit a user utilises.
4. Average Purchase Size: Size of an average purchase.
5. Average Cash Advance: Size of an Average Cash Advance Transaction
6. Average Cash Advance per Month: Average Cash Advance taken per 'active' cash months. Active cash month is nothing but cash frequency*tenure. This metric would be relevant if we want to see how much average cash advance is being taken in months when it was taken.
7. Average Purchase per Month: Average purchase per active month.
8. Average Usage per Month: Average usage per month (Not Active Month)
9. Usage: Net Transaction
10. One Off Purchases - Instalment Purchases Ratio: Ratio of One Off and Instalment Purchases.

Insights and Reporting

First we create buckets out of existing metrics to get valuable insights by them. For example, we could create a bucket on Purchase with factors like Low, Medium and high to see how the payment to min pay ratio of those with high purchases differs with customers with low purchases. Following are the groupings and insights derived from them.

1. Summary by Usage Bucket: Grouping data by usage bucket, we observe the following trends and behavioural patterns.
    a. Ratio of pay to min pay goes up with Usage. This indicates that customers who use the card more pay more than min balance.
    b. Also, Balance to Credit Limit ratio and Payment to MinPayment ratio goes up with usage.
    c. Metrics Like Average Monthly Purchase, Cash Advance, Transaction Size go up with usage as expected.



USAGE_BUCKET	CREDIT_LIMIT	PAY_MINPAY_RATIO	BALANCE_CREDITLIMIT_RATIO
VL	NaN	NaN	NaN
L	2000.0	1.286656	0.048470
M	3000.0	2.620626	0.279250
H	6000.0	2.405039	0.458589
VH	9000.0	4.316352	0.495762

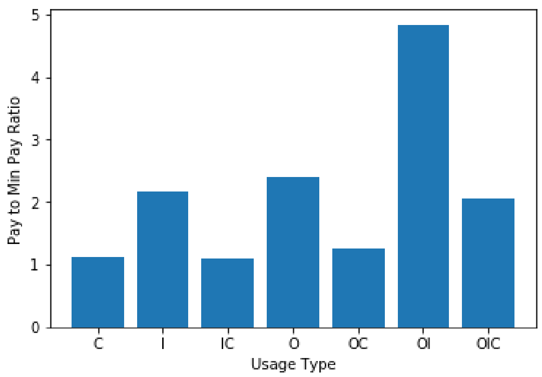
USAGE_BUCKET	AVERAGE_PURCHASE_SIZE	AVERAGE_CASH_SIZE	CASH_MONTH
VL	NaN	NaN	NaN
L	23.584167	0.000000	0.000000
M	45.922500	0.000000	0.000000
H	55.359512	247.484038	498.618913
VH	69.761864	365.004417	953.246889

USAGE_BUCKET	PURCHASE_MONTH	USAGE_MONTH	USAGE_MONTH_BALANCE_RATIO
VL	NaN	NaN	NaN
L	26.485058	14.503849	0.102771
M	78.733937	92.566574	0.164149
H	165.727143	308.982500	0.155584
VH	309.142500	702.794843	0.209071

USAGE_BUCKET	ONEOFF_INSTALLMENT_RATIO
VL	NaN
L	0.000000
M	0.000000
H	0.000000
VH	0.343277

Behaviour by Usage Bucket

2. Behaviour by Usage Type
    a. One of the most dramatic finding from this analysis is that OI Usage Type, that is, people who do not use this credit card for cash advance but use for Instalment and One Off payments is very high. If we assume this metric to represent a dimension of compliance, then we see that customer OI, O, OIC and I have a good compliance compared to OC, IC and C. Hence, customers who use their card more for cash are likely to have a lesser payment to min payment ratio.
    b. Also, Balance to Credit Limit ratio is quite high for C and IC. OC and OIC have moderately high ratio while I, OI and O have quite low ratios.
    c. Purchase Size is highest in O and OC and lowest in I and IC
    d. Usage per month is highest in case of OIC



USAGE_TYPE	ONEOFF_INSTALLMENT_RATIO
C	0.000000
I	0.000000
IC	0.000000
O	0.000000
OC	0.000000
OI	1.345084
OIC	1.331906

USAGE_TYPE	CREDIT_LIMIT	PAY_MINPAY_RATIO	BALANCE_CREDITLIMIT_RATIO	\
C	3000.0	1.118858	0.606633	
I	2500.0	2.170861	0.024842	
IC	3000.0	1.099800	0.668660	
O	3000.0	2.392582	0.077179	
OC	3500.0	1.261198	0.547982	
OI	5000.0	4.837576	0.132904	
OIC	5000.0	2.059641	0.519185	

USAGE_TYPE	AVERAGE_PURCHASE_SIZE	AVERAGE_CASH_SIZE	CASH_MONTH	\
C	0.000000	272.970591	466.388244	
I	33.350000	0.000000	0.000000	
IC	35.846000	292.507170	538.554978	
O	87.793333	0.000000	0.000000	
OC	80.000000	242.947798	408.484181	
OI	61.121200	0.000000	0.000000	
OIC	56.334000	225.473003	418.484173	

USAGE_TYPE	PURCHASE_MONTH	USAGE_MONTH	USAGE_MONTH_BALANCE_RATIO	\
C	0.000000	108.730274	0.084842	
I	42.000000	28.539000	0.677079	
IC	48.131667	163.415893	0.103605	
O	134.280537	38.412500	0.257547	
OC	127.600051	162.120823	0.103877	
OI	163.120833	129.638333	0.297296	
OIC	145.888333	248.436423	0.132096	

Behaviour by Usage Type

3. Behaviour by Cash Advance Bucket
    a. We see that q25 of Cash Advance is 0. Hence, nobody falls in 'low' category.
    b. Credit Limit of those bucketed 'VH' is more than twice that of others.
    c. We see that pay to min pay ratio decreases with increase in Cash Advance Amount.
    d. Purchase is low for High and Very High buckets

CASH_ADVANCE_BUCKET	CREDIT_LIMIT	PAY_MINPAY_RATIO
M	3000.0	2.257503
H	3000.0	1.173567
VH	7000.0	1.599023

CASH_ADVANCE_BUCKET	BALANCE_CREDITLIMIT_RATIO	AVERAGE_PURCHASE_SIZE
M	0.154367	45.869459
H	0.573772	10.393571
VH	0.578995	18.682961

CASH_ADVANCE_BUCKET	AVERAGE_CASH_SIZE	CASH_MONTH	PURCHASE_MONTH
M	0.000000	0.000000	74.100896
H	338.185919	594.083448	12.557497
VH	481.898037	1116.746687	24.276435

CASH_ADVANCE_BUCKET	USAGE_MONTH	USAGE_MONTH_BALANCE_RATIO
M	54.642083	0.203623
H	188.064500	0.115038
VH	452.192197	0.144487

4. Behaviour by One Off Purchase Bucket:
    a. Pay to Min Pay ratio is very high for those in 'VH' bucket. This means that those who have a higher number of OneOFF purchases than normal are paying way more than the min due.
    b. However, we notice that balance to credit limit ratio is less utilised even among those having very high One Off purchases compared to those who took cash advance.
    c. Those falling in 'M' bucket have significantly less One Off to Instalment Purchase Ratio. Hence, we could expect them to be having higher instalment purchases.

ONEOFF_PURCHASES_BUCKET	CREDIT_LIMIT	PAY_MINPAY_RATIO
M	3000.0	1.606346
H	4500.0	3.811627
VH	6500.0	7.596705

ONEOFF_PURCHASES_BUCKET	BALANCE_CREDITLIMIT_RATIO	AVERAGE_PURCHASE_SIZE
M	0.324903	28.375476
H	0.264571	73.072222
VH	0.260779	94.950000

ONEOFF_PURCHASES_BUCKET	AVERAGE_CASH_SIZE	CASH_MONTH	PURCHASE_MONTH
M	36.619299	44.052907	37.999991
H	0.000000	0.000000	174.624167
VH	0.000000	0.000000	367.289816

ONEOFF_PURCHASES_BUCKET	USAGE_MONTH	USAGE_MONTH_BALANCE_RATIO
M	56.111137	0.137803
H	129.776224	0.186366
VH	339.614167	0.283217

ONEOFF_PURCHASES_BUCKET	ONEOFF_INSTALLMENT_RATIO
M	0.000000
H	0.802377
VH	2.063239

Modelling

Data Preparation

Data Preparation is an important step of building the model. We observe that we have quite a lot of features in our dataset. We would like to remove some features and only take the valuable ones to build the model.

We have selected the following features.

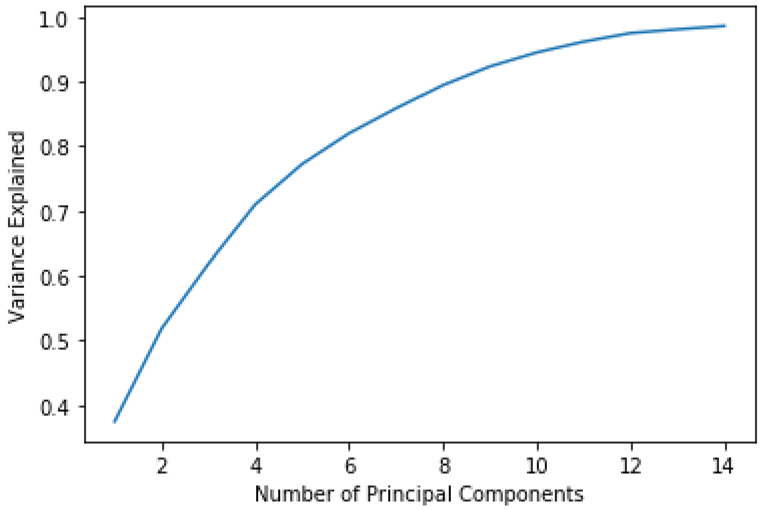
```
Index(['CUST_ID', 'BALANCE', 'BALANCE_FREQUENCY', 'PURCHASES',
       'ONEOFF_PURCHASES', 'INSTALLMENTS_PURCHASES', 'CASH_ADVANCE',
       'PURCHASES_FREQUENCY', 'ONEOFF_PURCHASES_FREQUENCY',
       'PURCHASES_INSTALLMENTS_FREQUENCY', 'CASH_ADVANCE_FREQUENCY',
       'CASH_ADVANCE_TRX', 'PURCHASES_TRX', 'CREDIT_LIMIT', 'PAYMENTS',
       'MINIMUM_PAYMENTS', 'PRC_FULL_PAYMENT', 'TENURE',
       'ACTIVE_PURCHASE_MONTHS', 'USAGE', 'PAY_MINPAY_RATIO',
       'BALANCE_CREDITLIMIT_RATIO', 'AVERAGE_PURCHASE_SIZE',
       'AVERAGE_CASH_SIZE', 'CASH_MONTH', 'PURCHASE_MONTH', 'USAGE_MONTH',
       'USAGE_MONTH_BALANCE_RATIO', 'ONEOFF_INSTALLMENT_RATIO', 'C', 'I', 'IC',
       'O', 'OC', 'OI', 'OIC'],
      dtype='object')
```

Please note, we have derived dummies (flags) out of USAGE_TYPE column, which may be seen as C, I, IC, O, OC, OI, OIC columns.

Dimensionality Reduction

We see that we have over 30 features in our dataset. We would like to reduce the number of features without losing information embedded in them. Hence, we use something called Principal Component Analysis (PCA) to reduce the number of features. Before proceeding, we would like to use the MinMaxScaler to convert all value between 0 and 1.

Let us now pass this dataset for PCA to determine the optimum number of principal components for our dataset. Consider the following graph.



In the figure, we see that just 10 principal components are explaining ~95% of the variance in the data. Hence, we select 10 as the number of principal components and transform the data into these 10 components.

The data look something like this

	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8	PC_9	PC_10
0	-0.316566	0.764652	0.200325	-0.210027	0.165876	-0.100471	0.296373	-0.263952	-0.052643	-0.125270
1	-1.052899	-0.012752	-0.425606	0.008489	-0.155245	0.196935	-0.149643	-0.027715	-0.129321	0.033665
2	0.433795	-0.588645	0.920680	0.003717	-0.493269	0.009155	-0.514656	-0.435381	0.287251	0.250737
3	-0.651205	-0.098186	0.436726	0.001038	0.884011	0.052639	0.165085	0.027500	-0.134569	0.171669
4	-0.632399	-0.162102	0.783121	-0.192600	-0.332333	-0.331651	-0.026644	-0.097972	-0.099073	-0.351612

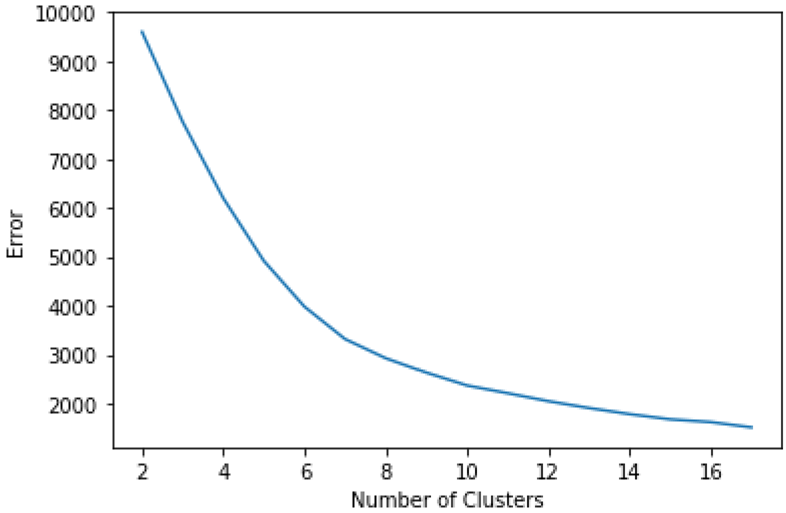
Clustering

The first thing we want to do before building a KMeans model is to figure out the number clusters. Ideally, intercluster variance should be high and intracluster variance should be as low as possible. Adding more clusters always reduces the error, but adding too many clusters can result in over fitting and meaningless clusters. Since this is a segmentation project, we want our clusters to make good business sense.

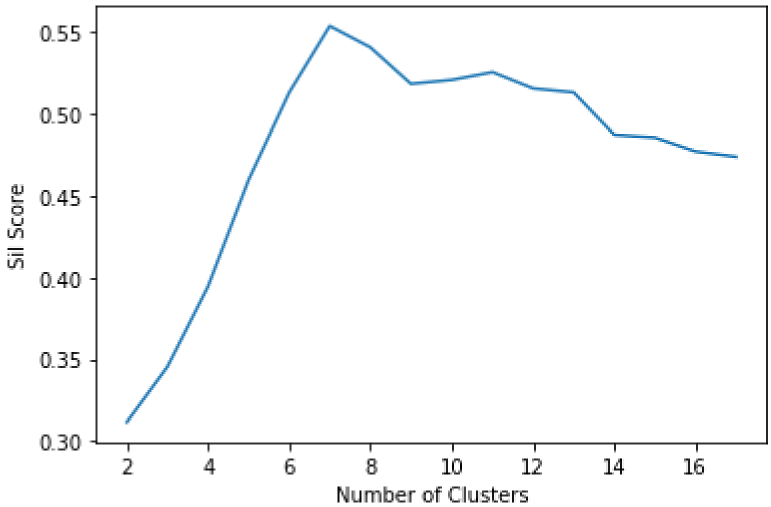
In order to determine the number of clusters, we use two methods called Elbow Method and Silhouette Method. Elbow Method gives a declining curve between number of clusters and error.

Consider the figure below. It is evident that error would go down with the numbers of clusters. We see that there is an ‘elbow’ between 6-9, (clutters) which means that the curve starts flattening in this range. This means that with the increase in number of clusters the decrease in error isn’t that much. Hence, the incentive of adding more clusters is low.

However, 6-9 is a broad range. How do we decide how many clusters do we use?



We now use something called Silhouette method. Higher the silhouette score, better the model. Consider the figure below. It has silhouette scores plotted against number of clusters.

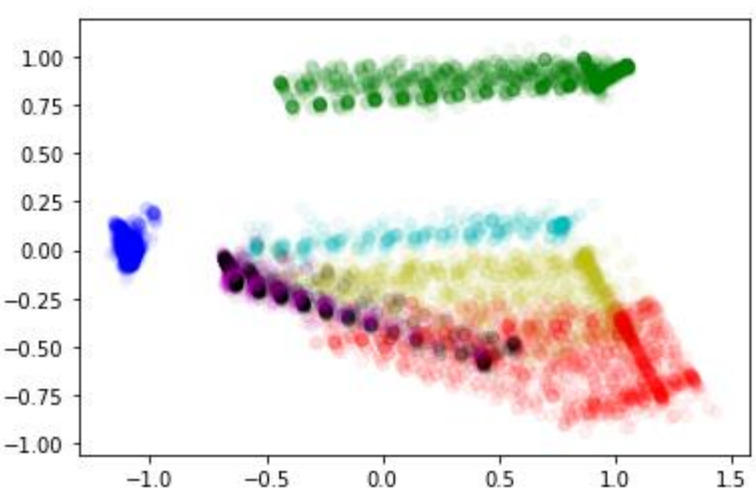


We observe the peak of this curve at 7 clusters. However, 8 too has a comparable silhouette score, and both fall within the range of 6-9 obtained from elbow method.

Hence, we would build two models C7 and C8 with 7 and 8 clusters respectively.

C7 Model

Consider the following scatter plot between PC_1 and PC_2. Different colors represent different clusters.



We see that clusters are well partitioned. We'll now get the labels from the model and merge them with the original dataset to derive detailed insights.

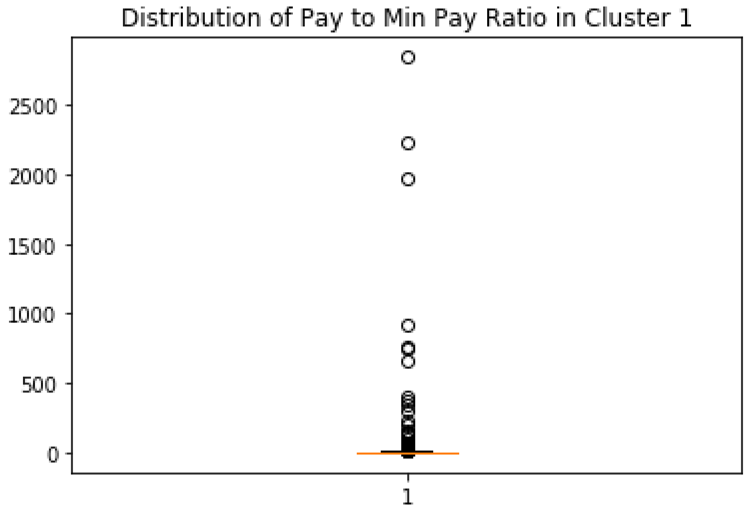
Insights and Observations

Cluster 0:

1. Cluster 0 has a very high median Credit Limit and the highest pay to min pay ratio among all the clusters.
2. Median Purchase Size is moderately high while mean purchase size is moderate. One Interesting observation about Cluster 0 is that customers in this segment have not taken any cash advance.
3. Their median purchases/month are high, with one-off purchases getting a greater share than instalments.
4. Balance of cluster 0 customers is low as compared to other clusters.
5. Balance to credit limit ratio is moderately low.

Cluster 1:

1. The mean of pay to min pay ratio is high while the median is low. This suggests that there clearly are outliers with very high pay to min pay ratio within this cluster. The same is illustrated in the following boxplot.
2. Purchases are zero while average cash size and purchase a month is quite high for this segment.
3. They have a moderate balance and a high balance to credit limit ratio.



MEDIAN:

C7	CREDIT_LIMIT	PAY_MINPAY_RATIO	AVERAGE_PURCHASE_SIZE	BALANCE
0	5000.0	4.837576	61.121200	532.467828
1	3000.0	1.118858	0.000000	1455.213589
2	2500.0	2.170861	33.350000	56.091512
3	5000.0	2.059641	56.334000	2006.680789
4	3000.0	2.392582	87.793333	245.199649
5	3000.0	1.097855	35.744667	1629.791663
6	3500.0	1.261198	80.000000	1601.393425

C7	AVERAGE_CASH_SIZE	CASH_MONTH	PURCHASE_MONTH	ONEOFF_PURCHASES
0	0.000000	0.000000	163.120833	762.50
1	272.970591	466.388244	0.000000	0.00
2	0.000000	0.000000	42.000000	0.00
3	225.473003	418.484173	145.888333	611.65
4	0.000000	0.000000	134.280537	443.25
5	292.026577	534.105899	48.100831	0.00
6	242.947798	408.484181	127.600051	293.26

C7	INSTALLMENTS_PURCHASES	BALANCE_CREDITLIMIT_RATIO
0	534.520	0.132904
1	0.000	0.606633
2	322.290	0.024842
3	444.960	0.519185
4	0.000	0.077179
5	335.995	0.668180
6	0.000	0.547982

Cluster 2

1. Cluster 2 has the lowest credit limit among all the clusters.
2. Just like cluster 1, it has a low median pay to min pay ratio but a much higher mean.
3. Just like cluster 0, customers in this cluster do not have a cash advance transaction. However, the mean and median of purchases is less than that of cluster 0.
4. Also, unlike cluster zero which has both One Off and Instalment transactions, this cluster has no One Off Transactions.
5. This cluster has lowest average balance and balance to credit limit ratio

Cluster 3

1. Cluster 3 customers have a high credit limit and pay to min pay ratio.
2. They also have a high cash, one-off and instalment transactions as compared to other clusters.
3. Customers in this clusters seem to be using their credit card more heavily as compared to other customers.
4. They have the highest balance and a high balance to credit limit ratio.

```
MEAN:
    CREDIT_LIMIT  PAY_MINPAY_RATIO  AVERAGE_PURCHASE_SIZE      BALANCE  \
C7
0    5702.960625          8.484759              79.504401  1214.121554
1    4026.000327         10.091979               0.000000  2148.620515
2    3114.849018         15.804855              52.177027   399.101290
3    5800.797553          5.081285              71.966157  2828.460343
4    4424.473022          6.160784             200.376333   762.978246
5    4374.025974          3.318745              55.297057  2581.681901
6    4637.888600          4.784567             142.275261  2340.086599

    AVERAGE_CASH_SIZE  CASH_MONTH  PURCHASE_MONTH  ONEOFF_PURCHASES  \
C7
0            0.000000    0.000000      243.849884       1506.717467
1          480.790312  769.896142        0.000000          0.000000
2            0.000000    0.000000       66.735444          0.000000
3          348.878375  632.776347      208.110985       1160.763825
4            0.000000    0.000000      275.667801        881.972801
5          488.584394  804.076762       74.850732          0.000000
6          383.402828  642.124787      233.034956        659.928020

    INSTALLMENTS_PURCHASES  BALANCE_CREDITLIMIT_RATIO
C7
0               956.203802                   0.255332
1                 0.000000                   0.574004
2               538.818155                   0.181346
3               770.304818                   0.523230
4                 0.000000                   0.258283
5               533.055152                   0.623051
6                 0.000000                   0.544846
```

Cluster 4

1. They have an average credit limit and an average pay to min pay ratio.
2. Highest average purchase size and low balance.
3. They have high purchase/month, zero instalment purchases, zero cash advance and moderate one off purchase. Basically, this segment used credit card only for One-Off purchases.
4. Balance to credit limit ratio is moderately low.

Cluster 5

1. Low Pay to Min Pay ratio, average credit limit and low average purchase size.
2. High Balance
3. Average Cash and Cash/Month metrics are high and purchase metrics are low.
4. They have zero One off Purchases and comparatively low instalment purchases.
5. Balance to credit limit ratio is the highest.

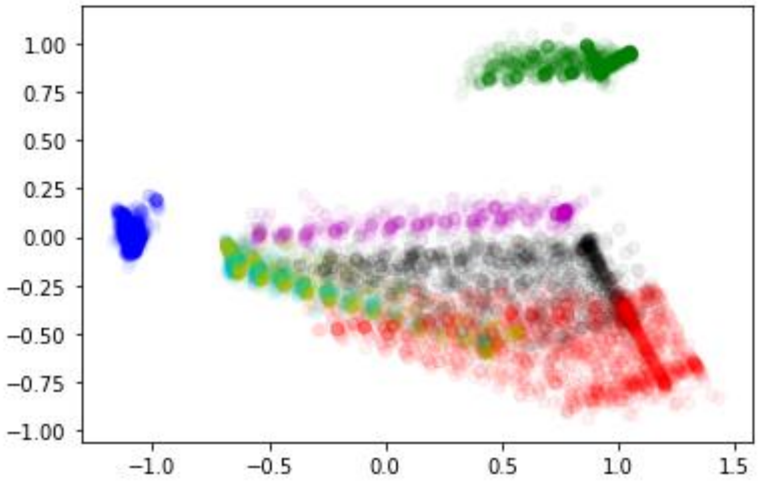
Cluster 6

1. Pay to Min Pay ratio is moderately low.
2. Average Purchase Size and Balance is high.
3. Moderate Cash per month and low purchase and one off purchases a month.
4. Zero Instalment Purchases and High Balance to Credit Limit Ratio.

Let us now look at the C8 Model

C8 Model

Consider the following figure. This doesn't look very different from C7 model



Insights and Observations from C8 Model

Cluster 0:

1. Credit Limit, Average Purchase Size and Pay to Min Pay Ratio is quite high for this group is high.
2. Average Balance for this group is moderately low and average cash advance is 0.
3. Purchase per month is high with one-off purchases getting more share than instalments.
4. Balance to Cred Limit ratio is low.

Cluster 1:

1. Credit Limit is moderately low. Mean of pay to min pay ratio is high while the median is low. Indicates that there are a lot of outliers in this cluster with regards to pay to min pay ratio.
2. Balance is High and there are no purchase transactions. Cash Transactions are quite high.
3. Balance to credit limit ratio is quite high.

Cluster 2:

1. This cluster has low credit limit. Like Cluster 1, they have a moderate pay to min pay median but a very high mean.
2. Average Purchase size is low and so is purchase per month and balance.
3. There are no cash advance transactions.
4. One Off purchases are zero and instalment purchases are moderately high.
5. Balance to Credit Limit Ratio is low.
6. These are customers who only use their credit card for instalment purchases. Their overall usage is quite low.

MEDIAN:

C8	CREDIT_LIMIT	PAY_MINPAY_RATIO	AVERAGE_PURCHASE_SIZE	BALANCE
0	5000.0	4.837576	61.121200	532.467828
1	3000.0	1.118858	0.000000	1455.213589
2	2000.0	2.304347	33.003333	73.945889
3	3000.0	2.392582	87.793333	245.199649
4	5000.0	2.059641	56.334000	2006.680789
5	3500.0	1.261198	80.000000	1601.393425
6	3000.0	1.099800	35.846000	1630.991993
7	2500.0	1.996425	34.858667	30.898879

C8	AVERAGE_CASH_SIZE	CASH_MONTH	PURCHASE_MONTH	ONEOFF_PURCHASES
0	0.000000	0.000000	163.120833	762.50
1	272.970591	466.388244	0.000000	0.00
2	0.000000	0.000000	42.823174	0.00
3	0.000000	0.000000	134.280537	443.25
4	225.473003	418.484173	145.888333	611.65
5	242.947798	408.484181	127.600051	293.26
6	292.507170	538.554978	48.131667	0.00
7	0.000000	0.000000	40.399995	0.00

C8	INSTALLMENTS_PURCHASES	BALANCE_CREDITLIMIT_RATIO
0	534.520	0.132904
1	0.000	0.606633
2	456.835	0.035694
3	0.000	0.077179
4	444.960	0.519185
5	0.000	0.547982
6	336.000	0.668660
7	159.220	0.012194

Cluster 3:

1. Credit Limit for customers in this segment is moderately low.
2. Average purchase size and purchases per month metrics are very high, all which are one off purchases.
3. No cash advance/instalment transaction.

4. Balance is moderately low.
5. Balance to credit limit ratio is low.

Cluster 4:

1. Credit Limit is high and pay to min pay ratio is moderate.
2. Average purchase size is moderate and balance is very high.
3. Metrics like Average Cash Size, Cash per Month and purchase per month are moderate.
4. One off Purchases and instalment purchases are both high.
5. High Balance to Credit Limit Ratio.

```
MEAN:
    CREDIT_LIMIT  PAY_MINPAY_RATIO  AVERAGE_PURCHASE_SIZE      BALANCE  \
C8
0    5702.960625          8.484759              79.504401  1214.121554
1    4026.000327         10.091979               0.000000  2148.620515
2    2900.444388         18.594152              48.334694   424.839589
3    4424.473022          6.160784             200.376333   762.978246
4    5800.797553          5.081285              71.966157  2828.460343
5    4637.888600          4.784567             142.275261  2340.086599
6    4374.837310          3.324109              55.417008  2586.433087
7    3433.245332         11.657934              57.789271   361.014995

    AVERAGE_CASH_SIZE  CASH_MONTH  PURCHASE_MONTH  ONEOFF_PURCHASES  \
C8
0            0.000000    0.000000      243.849884       1506.717467
1          480.790312  769.896142        0.000000          0.000000
2            0.000000    0.000000       64.971937          0.000000
3            0.000000    0.000000      275.667801        881.972801
4          348.878375  632.776347      208.110985       1160.763825
5          383.402828  642.124787      233.034956        659.928020
6          489.644230  805.820963       75.013098          0.000000
7            0.000000    0.000000       69.252347          0.000000

    INSTALLMENTS_PURCHASES  BALANCE_CREDITLIMIT_RATIO
C8
0               956.203802                   0.255332
1                 0.000000                   0.574004
2               691.747123                   0.200547
3                 0.000000                   0.258283
4               770.304818                   0.523230
5                 0.000000                   0.544846
6               534.211453                   0.624190
7               311.842218                   0.152826
```

Cluster 5:

1. Credit Limit is moderately high and pay to min pay ratio is low.
2. Average Purchase Size, Balance and purchase per month metrics are high.
3. Average Cash Size and Cash per month are moderate.
4. One off purchases are moderately low and there are no instalment purchases.
5. Balance to credit limit ratio is high.

Cluster 6:

1. Credit Limit and Pay to Min Pay ratio are moderately low.
2. Average purchase size and purchase per month are low.
3. Balance, Average Cash Size and Cash Advance amounts are high.
4. No one off transactions and instalment purchases are moderately low.
5. Balance to credit limit ratio is high.

Cluster 7:

1. Credit Limit is low.
2. Mean of Pay to Min Pay ratio is high while the median is low because of outliers within the cluster.
3. Average Purchase Size, purchase a month, balance and instalment purchases are low. No one off purchases.
4. Balance to Credit Limit Ratio is low
5. No Cash Transactions.

Comparison between C7 and C8

Consider the following crosstab.

col_0	0	1	2	3	4	5	6	7
row_0								
0	1757	0	0	0	0	0	0	0
1	0	2041	0	0	0	0	0	0
2	0	0	1074	0	0	0	0	725
3	0	0	0	0	1017	0	0	0
4	0	0	0	1071	0	0	0	0
5	0	0	0	0	0	0	461	1
6	0	0	0	0	0	803	0	0

This means that cluster 2 in C7 broke up to make clusters 2 and 7 in C8. We observe that there isn't much difference between Cluster 2 and 7 of C8 model. Both the clusters are similar to Cluster 2 of C7 model. Hence, we will go ahead with C7 Model.

Summary of C7 Cluster

Consider the following two tables that summarise the clusters.

C7	CREDIT_LIMIT	PAY_MINPAY_RATIO	AVERAGE_PURCHASE_SIZE	BALANCE	AVERAGE_CASH_SIZE	CASH_MONTH	PURCHASE_MONTH	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	BALANCE_CREDITLIMIT_RATIO
0	High	High	Moderate	Moderate	NULL	NULL	High	Very High	Very High	Moderate
1	Moderate	High	NULL	High	High	High	NULL	NULL	NULL	Very High
2	Very Low	High	Very Low	Very Low	NULL	NULL	Very Low	NULL	Low	Low
3	Very High	High	Low	Very High	Low	Low	Moderate	High	High	High
4	Moderate	High	Very High	Low	NULL	NULL	High	Moderate	NULL	Low
5	Moderate	Low	Very Low	High	Very High	Very High	Very Low	NULL	Low	Very High
6	Moderate	Low	High	High	Low	Low	Moderate	Low	NULL	High

C7	Comments
0	Purchase per month is high. Both One Off and Installment purchases are high. NO CASH, Credit Limit and Pay Min Pay Ratio is High
1	Moderate Credit Limit, High Pay to min pay ratio. No purchase. High cash advance. Balance to credit limit ratio is very high. (some have ay to min pay very high)
2	Low Credit Limit, very low usage. Only for installment. That too is low
3	Very High Credit Limit. They use for all 3 modes. Cash, Installment and Purchases. Cash is low. Rest is high/moderate. Balance to credit limit ratio is also high. The diference in 0 and
4	They have high value purchases, but they seem to be using the card less. No cash/installment purchases. Pay to min pay ratio is high. Balance to credit limit ratio is high
5	Moderate Cred limit, Low pay ratio, purchase size is very low,balance is high, cash is very high,purchase per month is very low,no one off purchases, low installment purchases, very
6	Moderate CL,Low pay to min pay ratio, high purchase size, balance,low cash,moderate purchase a month,low one off purchase,high cred limit ratio

Segmentation and Recommendations

As per their behaviour, we have named our clusters and also given recommendations based on the same. Please find them below.

1. **High Flyers (Cluster 0)**: They are important customers. They are very active credit card users and do not use their card for cash advance. Also, their balance to credit limit ratio is moderate. They must be given good offers so that they are encouraged to use their card more. They could be given offers on Cash Advance as their profile seems strong and they are less likely to default. Their credit limit could also be increased.
2. **Non Swiping (Cluster 1)**: They are high risk customers as they only use their cards for Cash Advance. They have never used their credit card for purchases. Their Balance to Credit Limit Ratio is also quite high. They should be given more information on benefits of using credit card for purchases and if possible, also encouraged to do so through good offers.
3. **Inactive (Cluster 2)**: These are highly inactive customers. Their purchase size/ purchase per month is very low. They only use their credit cards for instalment purchases and that too isn't much. Also, we observe that their credit limit is quite low. This could be because of a poor credit score. They must be made aware about the benefits of a good credit score, and how swiping their card more and timely payments can earn them one. This group of customers must be targeted for awareness and financial education.
4. **Heavy Users (Cluster 3)**: They are as good as High Flyers. The only difference is that they use cash advance facility too. However, that should not be a concern as they tend to use their card more for purchases. However, they have less one off and instalment purchases compared to high flyers. They should be targeted with better offers on one off Payments.
5. **High Potential (Cluster 4)**: They have a high pay to min pay ratio and their purchase size is good. Despite that, they lag behind the high flyers and Heavy Users in total purchase. Marketing Campaign for them could focus on offers that motivate them to do more one off transactions. If required, their credit limit could be enhanced too. Communications regarding offers on instalment purchases should also be sent to them. Focus should be to convert them to Heavy Users/High Flyers.
6. **Cash Affine (Cluster 5)**: They are much like Non Swiping customers. However, they have better credit limit, so they seem to be at lesser risk to default. Their cash advance transactions are higher than Non Swiping customers and pay to min pay ratio is low. They must be targeted to use their cad more for One Off purchases through offers/communication.
7. **Low Potential (Cluster 6)**: They differ with inactive customers in a way that their balance to credit limit ratio is high and pay to min pay ratio is low. They have no instalment purchases and most of their transactions are either cash/one off. This, with a low pay to min pay puts them at a risk of defaulting. They must be encouraged to make full payments through extra card points or discounts.