STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Final Project:

Customer Segmentation using KMeans Clustering

Under the guidance of Professor M. Daneshmand

Course: ***MIS 637***

Course Name: ***Data Analytics and Machine Learning***

Presenter: ***Ankita Ramvir Singh Yadav***

Topics Covered

- Introduction
- Business Understanding Phase
  - Problem Statement
- Data Understanding Phase
- Data Preparation Phase
- Modeling Phase
  - Methodologies and Algorithms
  - Software
- Execution and Analysis
- Conclusion
- Reference

Introduction

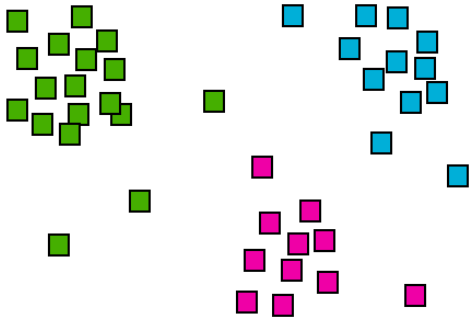
Customers are a key to any organization's success. Therefore, it is essential for enterprises and organizations to successfully acquire new customers and retain high-value customers.

As per Pareto's Principle, 80% of company sales come from about 20% of its customers hence it is important to retain the existing customers.

Companies need to market their product in the right manner for acquiring new customers and retaining the old ones. It is essential to understand the customers, their needs and wants for designing the right marketing campaign.

Spending too much on promotional events without understanding the requirement of the customers will not benefit the organization. This will just increase their expenses.

Also, at times companies do not have enough resources to target a mass market. Hence, the market should be segmented to understand the customers better as well as design tailor-made solutions for the existing and potential customers.



Introduction

Market Segmentation involves dividing the target market into smaller, well-defined categories. The segmented market meets certain criteria and share similar characteristics. These segments tend to respond identically to marketing strategies. Their interests, needs, wants, and demands are common.

Market segmentation enables an organization to gain a greater understanding of customer needs and wants by offering them products and services well suited for their needs thus increasing the likelihood of segments to purchase products.

The market segmentation process also gives marketers the ability to prioritize different products and services to different target audiences. The purpose of segmentation is the concentration of marketing energy and force on subdividing to gain a competitive advantage within the segment.

For a credit card company, the behavior of customers include:

- What is their credit card spending habits?
- How loyal are they to the credit card company?
- What is their usage rate of the credit card?
- What need is a consumer trying to satisfy?
- What is their credit, transaction, and payment history?

This information is relevant because it is directly related to how a consumer interacts with the credit card offerings. Therefore, marketers can market more effectively to customers by knowing their behaviors. They can design the credit card offers in such a manner that shall not only retain the customer but will also lead to an increase in usage of the credit card thereby increasing the profitability of the company.

Business Understanding Phase

The credit card company's objective is to launch an effective marketing campaign and offer better product solutions so that it can retain existing customers, and acquire new customers thereby increase the profitability of the company and reduce marketing expenditure through market segmentation based on the customer's behavior.

Problem Statement

Targeting the right customer is important for any organization's success. However, one of the pain point for marketers is identify the right needs of the customers and design marketing strategy accordingly. A credit card company can become more customer focused through a market segment-oriented approach.

The research shall answer the below questions:

- What is the customer's need and wants?
- What is the buying behavior of the customers?
- How to improve marketing campaign performance?
- Which are the areas the company can expand?
- What are the areas the product can be developed?

Marketers can create a personalized strategy based on key characteristics that have a higher chance of success than if they were to create a generic campaign and try to implement it across all segments. Companies can identify which segments are more likely to make more purchases through credit cards and can better channel their attention and resources effectively. Market segmentation shall enable the company to determine market opportunities, make best adjustments in marketing campaigns, and use resources effectively. Successful marketing usually occurs when the companies speak precisely to their target audience. Market segmentation helps achieve that precision.

Data Understanding Phase

Data Source

The data set has been downloaded from Kaggle. The link is shared below:

https://www.kaggle.com/rruchi/marketing-datacsv

Data Set Description

The data set consists of 18 columns and 8950 rows.

CUST_ID	PURCHASES_INSTALLMENTS_FREQUENCY
BALANCE	CASH_ADVANCE_FREQUENCY
BALANCE_FREQUENCY	CASH_ADVANCE_TRX
PURCHASES	PURCHASES_TRX
ONEOFF_PURCHASES	CREDIT_LIMIT
INSTALLMENTS_PURCHASES	PAYMENTS
CASH_ADVANCE	MINIMUM_PAYMENTS
PURCHASES_FREQUENCY	PRC_FULL_PAYMENT
ONEOFF_PURCHASES_FREQUENCY	TENURE

Data Understanding Phase

The data consists of 8950 entries. There are total 18 columns. The data type value for 14 columns is float64, 3 columns has int64 as data type, and 1 column as data type as object.

```
RangeIndex: 8950 entries, 0 to 8949
Data columns (total 18 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   CUST_ID                           8950 non-null   object
 1   BALANCE                           8950 non-null   float64
 2   BALANCE_FREQUENCY                 8950 non-null   float64
 3   PURCHASES                         8950 non-null   float64
 4   ONEOFF_PURCHASES                  8950 non-null   float64
 5   INSTALLMENTS_PURCHASES            8950 non-null   float64
 6   CASH_ADVANCE                      8950 non-null   float64
 7   PURCHASES_FREQUENCY               8950 non-null   float64
 8   ONEOFF_PURCHASES_FREQUENCY        8950 non-null   float64
 9   PURCHASES_INSTALLMENTS_FREQUENCY  8950 non-null   float64
 10  CASH_ADVANCE_FREQUENCY            8950 non-null   float64
 11  CASH_ADVANCE_TRX                  8950 non-null   int64
 12  PURCHASES_TRX                     8950 non-null   int64
 13  CREDIT_LIMIT                      8949 non-null   float64
 14  PAYMENTS                          8950 non-null   float64
 15  MINIMUM_PAYMENTS                  8637 non-null   float64
 16  PRC_FULL_PAYMENT                  8950 non-null   float64
 17  TENURE                            8950 non-null   int64
dtypes: float64(14), int64(3), object(1)
```

Data Understanding Phase

Attributes Description

Column Name	Column Description
CUST_ID	The unique identification of the customer (account holder)
BALANCE	The balance amount left in customers account for purchases
BALANCE_FREQUENCY	It mentions about how frequently the balance is updated The score is mentioned between 0 and 1 1 indicates the balance is frequently updated while 0 indicates the balance is not frequently updated
PURCHASES	The amount of purchases made by the account holder
ONEOFF_PURCHASES	Maximum amount spent by customer at once
INSTALLMENTS_PURCHASES	The amount of purchase done by the customer in installment
CASH_ADVANCE	Cash in advance used by the customer
PURCHASES_FREQUENCY	It mentions about how frequently the purchases are made The score is mentioned between 0 and 1 1 indicates the purchase is frequently made while 0 indicates the purchase is not frequent
ONEOFF_PURCHASES_FREQUENCY	It mentions how frequently the puchase is happening at one go The score is mentioned between 0 and 1 1 indicates the frequency is high and 0 indicates the frequency is low
PURCHASES_INSTALLMENTS_FREQUENCY	It mentions how frequently the puchase is happening in installment The score is mentioned between 0 and 1 1 indicates the frequency is high and 0 indicates the frequency is low
CASH_ADVANCE_FREQUENCY	It mentions how frequently the cash in advance is paid
CASH_ADVANCE_TRX	It mentions the number of transactions made in cash
PURCHASES_TRX	It mentions the number of purchase transactions done
CREDIT_LIMIT	It mentions the limit of credit card user
PAYMENTS	It mentions the amount of payment done by user
MINIMUM_PAYMENTS	It mentions the minimum amount of payment made by user
PRC_FULL_PAYMENT	It mentions the percent of full payment paid by user
TENURE	It mentions the tenure of the customer

Data Understanding Phase

Sample Data Set

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	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
2	C10001	40.900749	0.818182	95.4	0	95.4	0	0.166667	0	0.083333	0	0	2	1000	201.802084	139.509787	0	12
3	C10002	3202.46742	0.909091	0	0	0	6442.945483	0	0	0	0.25	4	0	7000	4103.0326	1072.340217	0.222222	12
4	C10003	2495.14886	1	773.17	773.17	0	0	1	1	0	0	0	12	7500	622.066742	627.284787	0	12
5	C10004	1666.67054	0.636364	1499	1499	0	205.788017	0.083333	0.083333	0	0.083333	1	1	7500	0		0	12
6	C10005	817.714335	1	16	16	0	0	0.083333	0.083333	0	0	0	1	1200	678.334763	244.791237	0	12
7	C10006	1809.82875	1	1333.28	0	1333.28	0	0.666667	0	0.583333	0	0	8	1800	1400.05777	2407.246035	0	12
8	C10007	627.260806	1	7091.01	6402.63	688.38	0	1	1	1	0	0	64	13500	6354.31433	198.065894	1	12
9	C10008	1823.65274	1	436.2	0	436.2	0	1	0	1	0	0	12	2300	679.065082	532.03399	0	12
10	C10009	1014.92647	1	861.49	661.49	200	0	0.333333	0.083333	0.25	0	0	5	7000	688.278568	311.963409	0	12
11	C10010	152.225975	0.545455	1281.6	1281.6	0	0	0.166667	0.166667	0	0	0	3	11000	1164.77059	100.302262	0	12
12	C10011	1293.12494	1	920.12	0	920.12	0	1	0	1	0	0	12	1200	1083.30101	2172.697765	0	12
13	C10012	630.794744	0.818182	1492.18	1492.18	0	0	0.25	0.25	0	0	0	6	2000	705.618627	155.549069	0	12
14	C10013	1516.92862	1	3217.99	2500.23	717.76	0	1	0.25	0.916667	0	0	26	3000	608.263689	490.207013	0.25	12
15	C10014	921.693369	1	2137.93	419.96	1717.97	0	0.75	0.166667	0.75	0	0	26	7500	1655.89144	251.137986	0.083333	12
16	C10015	2772.77273	1	0	0	0	346.81139	0	0	0	0.083333	1	0	3000	805.647974	989.962866	0	12
17	C10016	6886.21323	1	1611.7	0	1611.7	2301.491267	0.5	0	0.5	0.166667	4	11	8000	1993.43928	2109.90649	0	12
18	C10017	2072.07435	0.875	0	0	0	2784.274703	0	0	0	0.25	3	0	3000	391.974562	376.579631	0	8
19	C10018	41.089489	0.454545	519	0	519	0	0.416667	0	0.333333	0	0	8	2500	254.590662	73.203221	0.25	12
20	C10019	1989.07223	1	504.35	166	338.35	0	0.666667	0.083333	0.583333	0	0	9	13000	1720.83737	744.613395	0	12
21	C10020	3577.97093	1	398.64	0	398.64	0	1	0	1	0	0	12	4000	1053.98046	12465.54938	0	12
22	C10021	2016.68469	1	176.68	0	176.68	0	0.666667	0	0.666667	0	0	8	2000	223.0686	13557.29726	0	12
23	C10022	6369.53132	1	6359.95	5910.04	449.91	229.028245	1	0.916667	1	0.333333	6	92	11250	2077.95905	1659.775075	0	12
24	C10023	132.34224	0.636364	815.9	0	815.9	0	1	0	1	0	0	17	1800	2359.62996	86.329554	0	12
25	C10024	3800.15138	0.818182	4248.35	3454.56	793.79	7974.415626	1	0.083333	0.916667	0.333333	13	13	9000	9479.04384	1425.426525	0	12
26	C10025	5368.57122	1	0	0	0	798.949863	0	0	0	0.363636	4	0	6000	1422.72671	1657.002877	0	11
27	C10026	169.781679	1	399.6	0	399.6	0	1	0	1	0	0	12	3000	215.306142	163.005536	0	12
28	C10027	1615.96724	1	102	102	0	244.840485	0.166667	0.166667	0	0.25	5	2	1700	890.178845	971.183209	0	12
29	C10028	125.694817	1	233.28	0	233.28	0	1	0	1	0	0	12	1000	207.773715	164.748819	0	12
30	C10029	7152.86437	1	387.05	204.55	182.5	2236.145259	0.666667	0.166667	0.416667	0.833333	16	8	10500	1601.44835	1648.851345	0	12
31	C10030	22.06349	1	100	0	100	0	0.416667	0	0.416667	0	0	5	6900	160.767773	159.684442	0	12
32	C10031	12136.22	1	3038.01	1013.2	2024.81	3183.583301	1	0.166667	1	0.5	10	64	13000	4230.32349	3292.01503	0	12
33	C10032	1162.27332	1	1347.71	400	947.71	175.815755	1	0.083333	1	0.166667	2	45	6000	645.998299	256.848991	0	12
34	C10033	6732.82306	1	324.95	324.95	0	1189.533753	0.083333	0.083333	0	0.166667	3	3	10500	1720.11888	2506.745366	0	12
35	C10034	125.660453	1	636.79	636.79	0	0	0.916667	0.916667	0	0	0	11	1200	423.449696	196.301885	0.3	12
36	C10035	3517.10162	0.727273	547.28	0	547.28	0	1	0	1	0	0	14	11000	8735.60784	17862.55831	0	12

Data Preparation Phase

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE
count	8950.000000	8950.000000	8950.000000	8950.000000	8950.000000	8950.000000
mean	1564.474828	0.877271	1003.204834	592.437371	411.067645	978.871112
std	2081.531879	0.236904	2136.634782	1659.887917	904.338115	2097.163877
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	128.281915	0.888889	39.635000	0.000000	0.000000	0.000000
50%	873.385231	1.000000	361.280000	38.000000	89.000000	0.000000
75%	2054.140036	1.000000	1110.130000	577.405000	468.637500	1113.821139
max	19043.138560	1.000000	49039.570000	40761.250000	22500.000000	47137.211760

	PURCHASES_FREQUENCY	ONEOFF_PURCHASES_FREQUENCY	PURCHASES_INSTALLMENTS_FREQUENCY	CASH_ADVANCE_FREQUENCY
count	8950.000000	8950.000000	8950.000000	8950.000000
mean	0.490351	0.202458	0.364437	0.135144
std	0.401371	0.298336	0.397448	0.200121
min	0.000000	0.000000	0.000000	0.000000
25%	0.083333	0.000000	0.000000	0.000000
50%	0.500000	0.083333	0.166667	0.000000
75%	0.916667	0.300000	0.750000	0.222222
max	1.000000	1.000000	1.000000	1.500000

	CASH_ADVANCE_TRX	PURCHASES_TRX	CREDIT_LIMIT	PAYMENTS	MINIMUM_PAYMENTS	PRC_FULL_PAYMENT	TENURE
count	8950.000000	8950.000000	8949.000000	8950.000000	8637.000000	8950.000000	8950.000000
mean	3.248827	14.709832	4494.449450	1733.143852	864.206542	0.153715	11.517318
std	6.824647	24.857649	3638.815725	2895.063757	2372.446607	0.292499	1.338331
min	0.000000	0.000000	50.000000	0.000000	0.019163	0.000000	6.000000
25%	0.000000	1.000000	1600.000000	383.276166	169.123707	0.000000	12.000000
50%	0.000000	7.000000	3000.000000	856.901546	312.343947	0.000000	12.000000
75%	4.000000	17.000000	6500.000000	1901.134317	825.485459	0.142857	12.000000
max	123.000000	358.000000	30000.000000	50721.483360	76406.207520	1.000000	12.000000

The above table describes the count, mean, standard deviation, minimum value, quartile range, and maximum value of each column

Data Preparation Phase

Handling Missing Data

From the below table we can see that, there is 1 null value present in ‘CREDIT_LIMIT’ column, and 313 null value in ‘MINIMUM_PAYMENTS’.

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

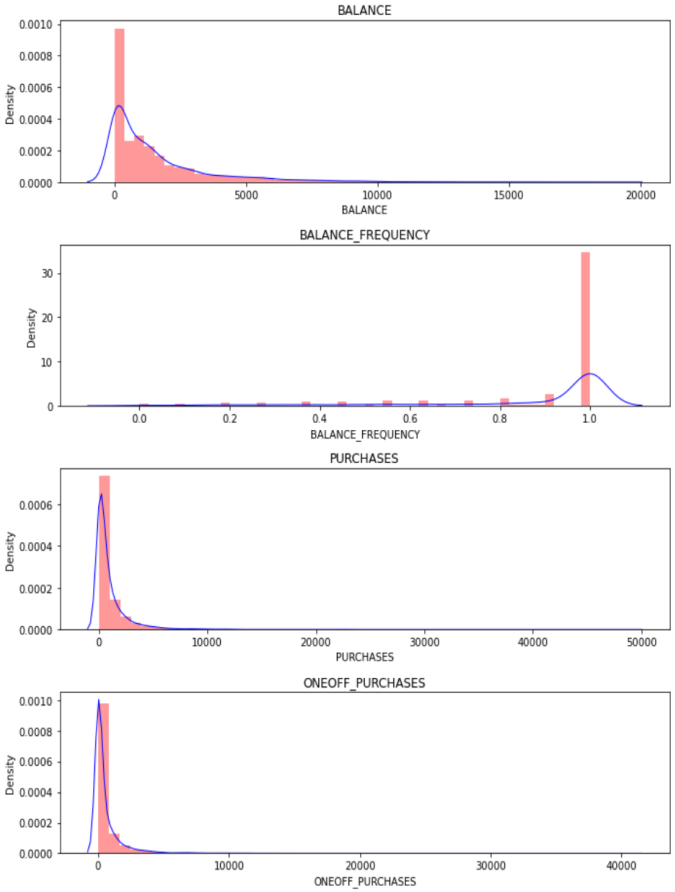
The missing value has been replaced with the field mean for both the numerical variables (CREDIT_LIMIT, MINIMUM_PAYMENTS).

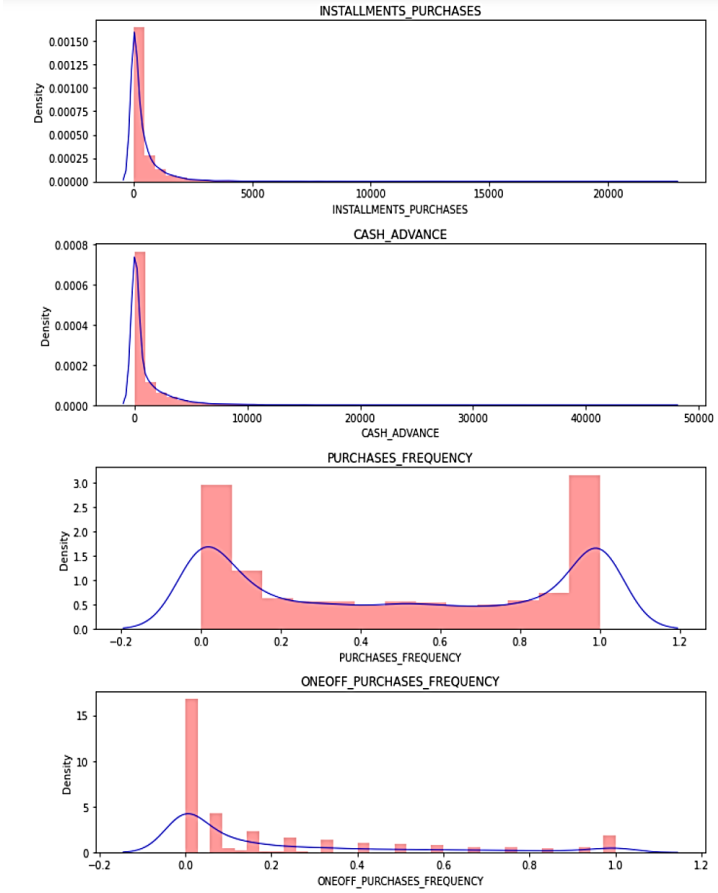
Obsolete Field

The field ‘CUST_ID’ is not required for the analysis hence the same has been dropped. So now we have 8950 rows and 17 columns for our analysis.

Data Preparation Phase

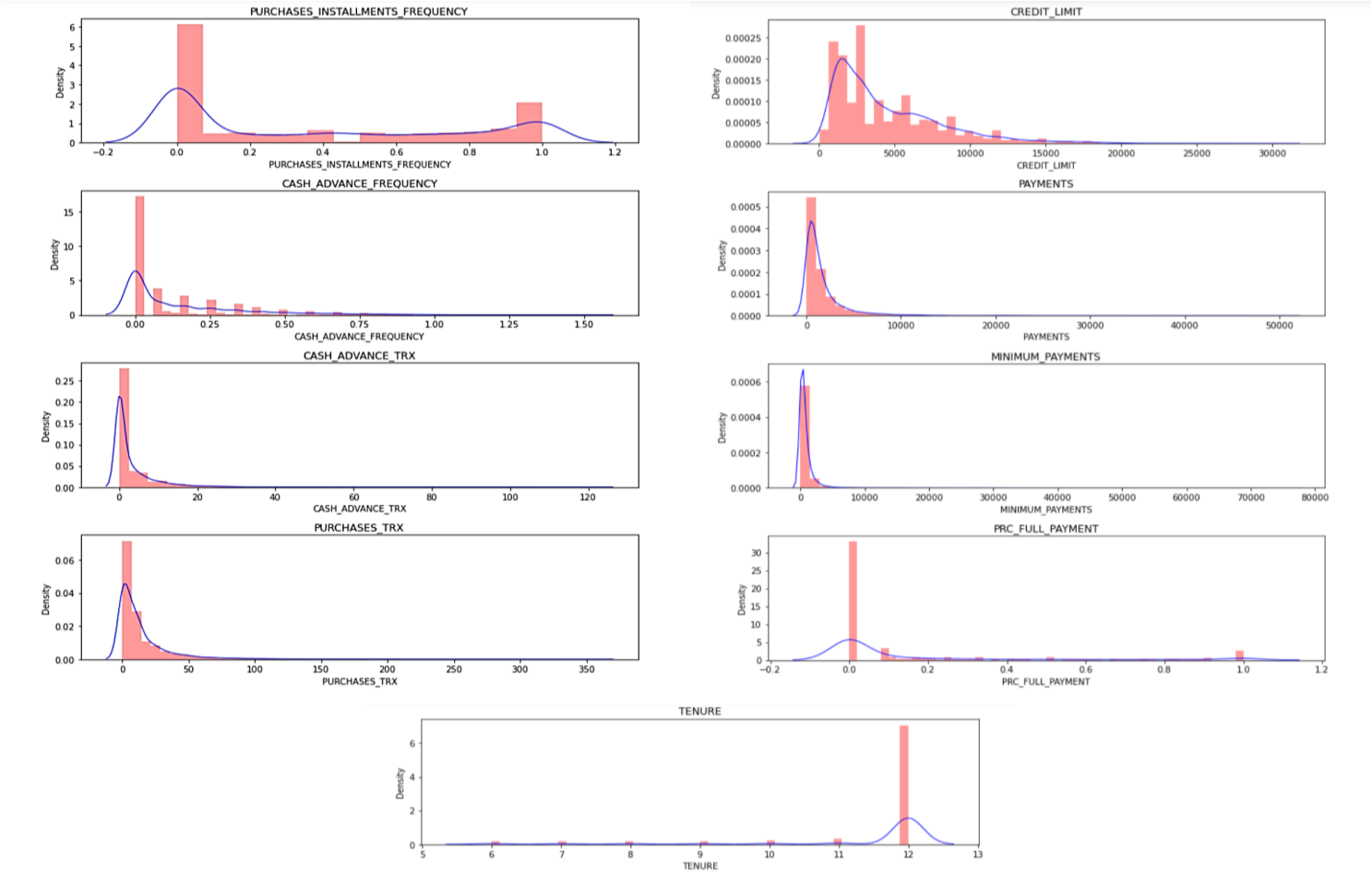
Distribution of data across variables





Data Preparation Phase

Distribution of data across variables



Data Preparation Phase

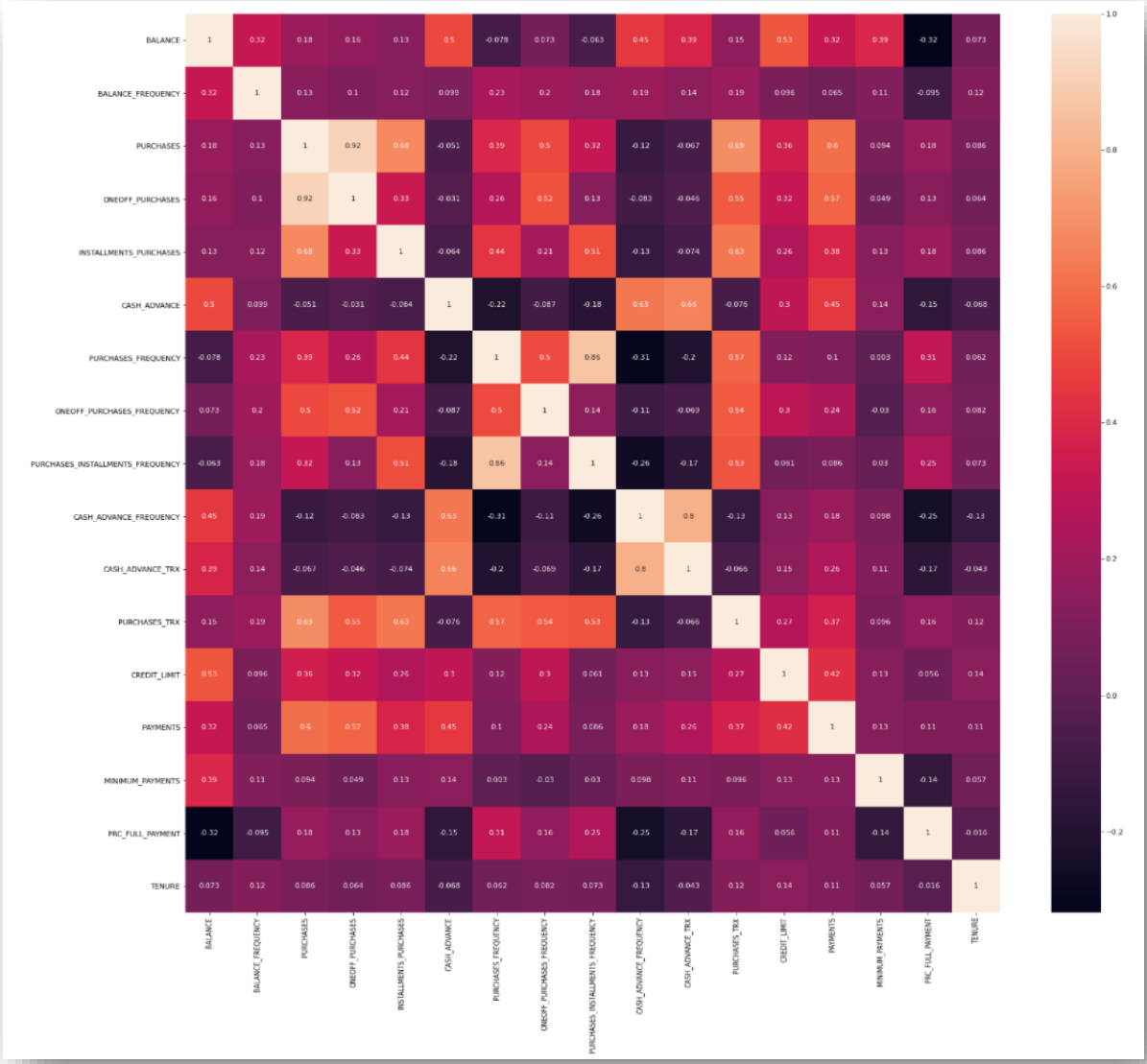
Correlation among Variables

As we can see from the diagram, there is a strong correlation present among variables.

PURCHASE and ONEOFF_PURCHASE has a strong correlation, PURCHASE_INSTALLMENT_ FREQUENCY and PURCHASE_FREQUENCY too share a strong correlation.

Many variables share correlation score above 0.40

Since skewness and correlation is present among variable, we shall do normalization and principal component analysis



Data Preparation Phase

Data Transformation through Min-Max normalization

Min–max normalization works by seeing how much greater the field value is than the minimum value min*(X)* and scaling this difference by the range.

min–max normalization values will range from zero to one.

Raw Data

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_I
0	C10001	40.900749	0.818182	95.40	0.00	95.40	0.000000	
1	C10002	3202.467416	0.909091	0.00	0.00	0.00	6442.945483	
2	C10003	2495.148862	1.000000	773.17	773.17	0.00	0.000000	
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.00	205.788017	
4	C10005	817.714335	1.000000	16.00	16.00	0.00	0.000000	
...	...	...	...	...	...	...	...	
8945	C19186	28.493517	1.000000	291.12	0.00	291.12	0.000000	
8946	C19187	19.183215	1.000000	300.00	0.00	300.00	0.000000	
8947	C19188	23.398673	0.833333	144.40	0.00	144.40	0.000000	
8948	C19189	13.457564	0.833333	0.00	0.00	0.00	36.558778	
8949	C19190	372.708075	0.666667	1093.25	1093.25	0.00	127.040008	

8950 rows × 18 columns

Cleaned Data

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENC
0	40.900749	0.818182	95.40	0.00	95.40	0.000000	0.16666
1	3202.467416	0.909091	0.00	0.00	0.00	6442.945483	0.0000(
2	2495.148862	1.000000	773.17	773.17	0.00	0.000000	1.0000(
3	1666.670542	0.636364	1499.00	1499.00	0.00	205.788017	0.0833
4	817.714335	1.000000	16.00	16.00	0.00	0.000000	0.0833
...	...	...	...	...	...	...	
8945	28.493517	1.000000	291.12	0.00	291.12	0.000000	1.0000(
8946	19.183215	1.000000	300.00	0.00	300.00	0.000000	1.0000(
8947	23.398673	0.833333	144.40	0.00	144.40	0.000000	0.8333
8948	13.457564	0.833333	0.00	0.00	0.00	36.558778	0.0000(
8949	372.708075	0.666667	1093.25	1093.25	0.00	127.040008	0.6666(

8950 rows × 17 columns

Normalized Data

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY
0	0.002148	0.818182	0.001945	0.000000	0.004240	0.000000	0.166667
1	0.168169	0.909091	0.000000	0.000000	0.000000	0.136685	0.000000
2	0.131026	1.000000	0.015766	0.018968	0.000000	0.000000	1.000000
3	0.087521	0.636364	0.030567	0.036775	0.000000	0.004366	0.083333
4	0.042940	1.000000	0.000326	0.000393	0.000000	0.000000	0.083333
...	...	...	...	...	...	...	...
8945	0.001496	1.000000	0.005936	0.000000	0.012939	0.000000	1.000000
8946	0.001007	1.000000	0.006118	0.000000	0.013333	0.000000	1.000000
8947	0.001229	0.833333	0.002945	0.000000	0.006418	0.000000	0.833333
8948	0.000707	0.833333	0.000000	0.000000	0.000000	0.000776	0.000000
8949	0.019572	0.666667	0.022293	0.026821	0.000000	0.002695	0.666667

8950 rows × 17 columns

Data Preparation Phase

Principal Component Analysis (PCA)

Is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

For PCA, number of components is considered as 2. Below components are generated post doing principal component analysis.

	0	1
0	-0.482164	-0.097656
1	-0.608577	-0.019379
2	0.304507	0.920946
3	-0.588622	-0.005605
4	-0.554354	0.052965
...	...	...
8945	0.646921	-0.463044
8946	0.549893	-0.406711
8947	0.360389	-0.422647
8948	-0.646095	-0.204332
8949	-0.094598	0.399578

8950 rows × 2 columns

Modeling Phase

Methodologies and Algorithms

For the project, K-means clustering shall be used. K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.

Unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes.

The objective of K-means is to group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (*k*) of clusters in a dataset.

A cluster refers to a collection of data points aggregated together because of certain similarities.

The K-means algorithm identifies *k* number of centroids (imaginary or real location representing the center of the cluster), and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

K-means clustering is an extensively used technique for data cluster analysis.

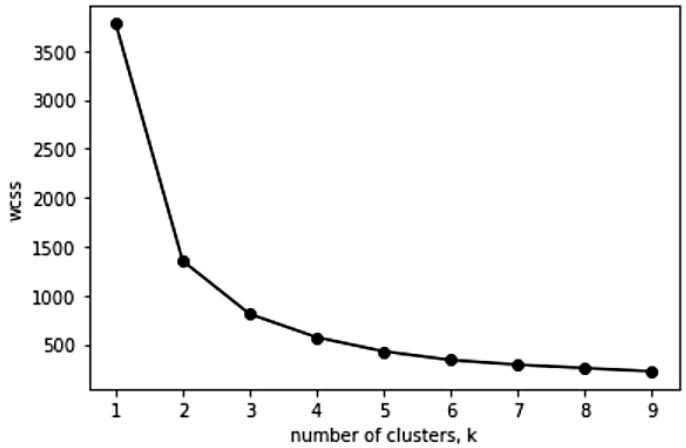
By using K-means clustering on the selected data set, we shall be able to create meaningful clusters of customers.

The market segmentation shall enable the financial institution to create specialized marketing campaigns for their customers.

Modeling Phase

Finding optimal number of clusters

Through within cluster sum of squares score we shall find the optimal number of clusters.



Number of Cluster	WCSS
1	3777.95
2	1353.02
3	809.47
4	571.95
5	427.83
6	339.95
7	291.42
8	257.28
9	223.55

As seen in the above graph, an elbow effect is formed at point 3 of number of clusters. The inertia score for 1 cluster is 3777.95, for 2 cluster is 1353.02, and for 3 cluster is 809.47. After that there is no significant change in the inertia score. Hence, we conclude that the optimal number of cluster is 3.

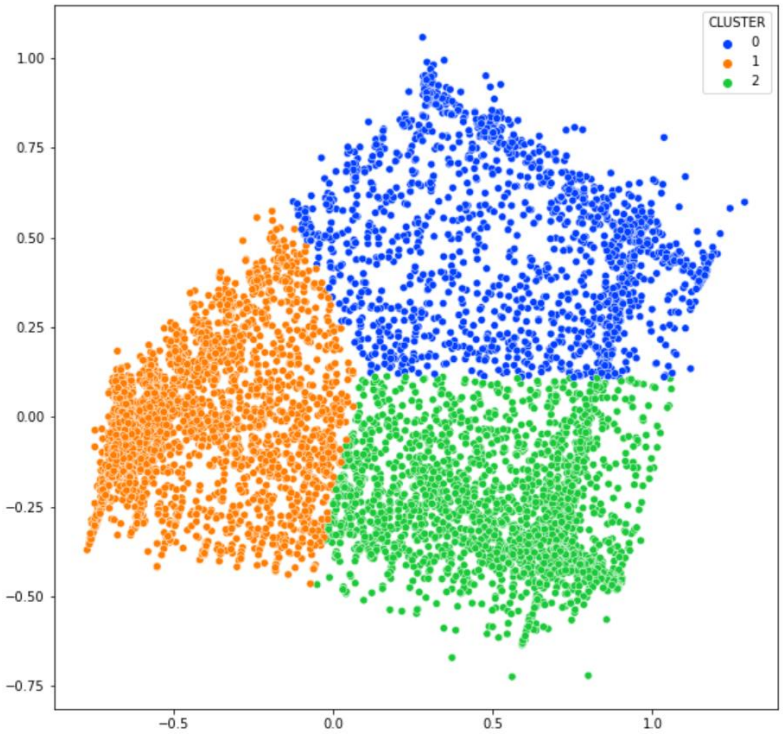
Modeling Phase

By applying K-means clustering, 3 distinct clusters are formed.

Cluster 0 consists of 4708 data points.

Cluster 1 consists of 1522 data points.

Cluster 2 consists of 2720 data points.

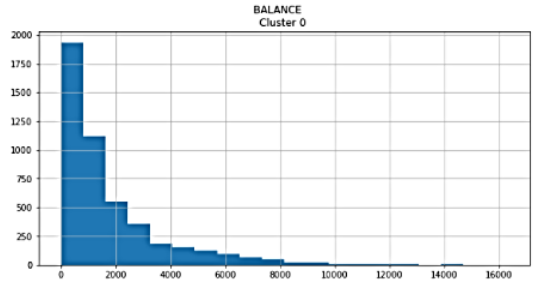


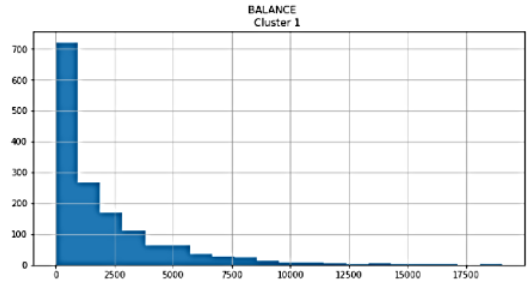
Modeling Phase

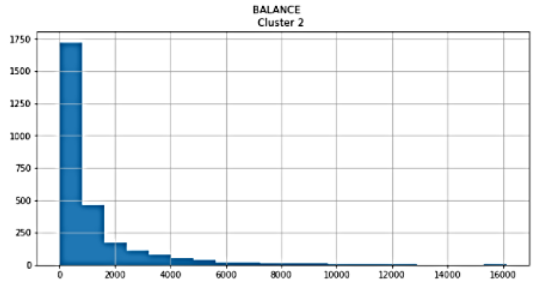
Software

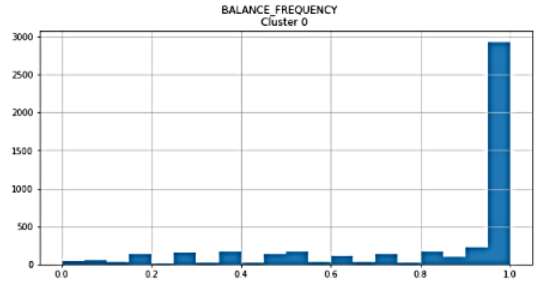
Python has been used for data processing steps such as data cleaning, data integration, data reduction, data transformation, modeling, execution, and analysis phase

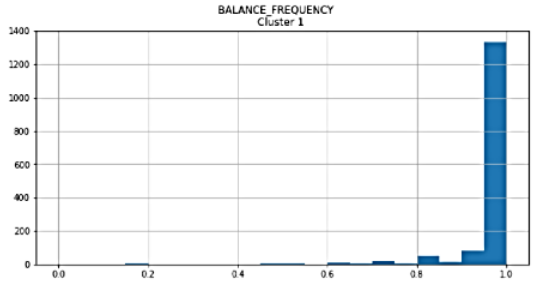
Execution and Analysis

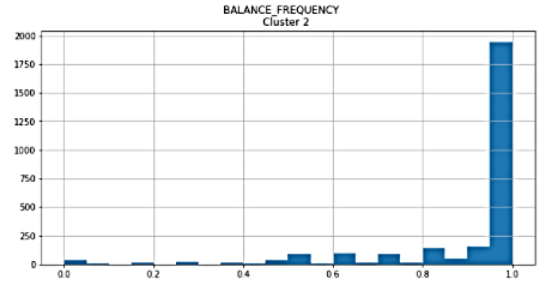










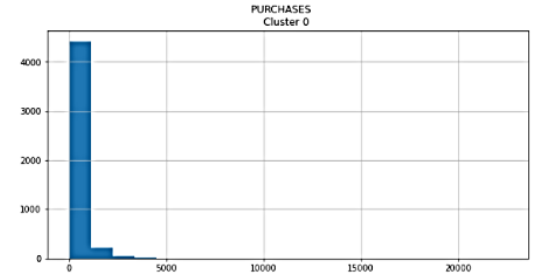


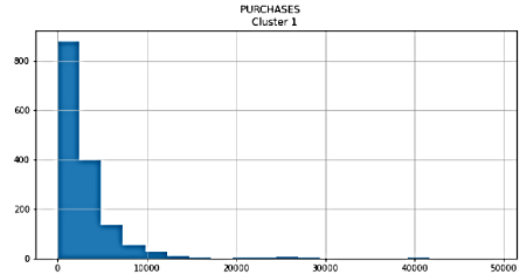
Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
BALANCE	1712.86	2014.64	1055.72
BALANCE_FREQUENCY	0.8277	0.977	0.9

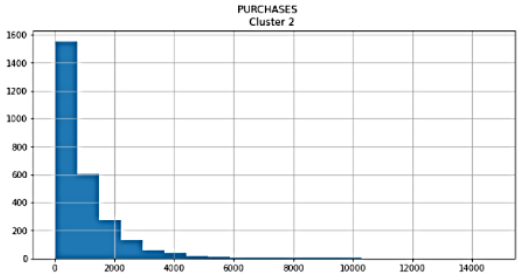
The balance amount left in the customers account is highest in the case of Cluster 1 and lowest in the case of Cluster 2.

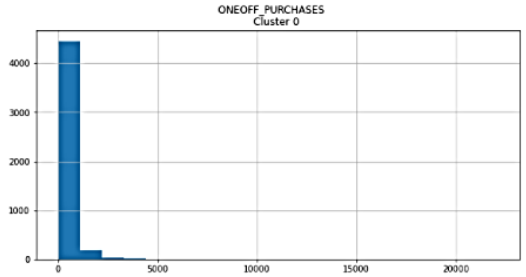
The balance frequency is highest in the case of Cluster 1 and lowest in the case of Cluster 0.

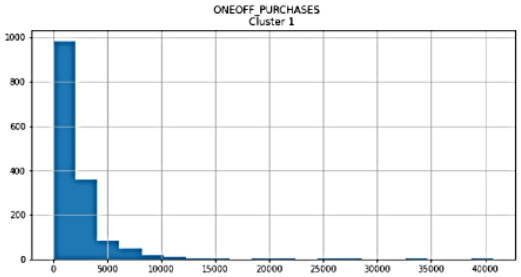
Execution and Analysis

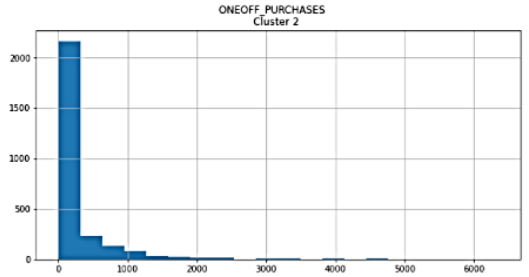










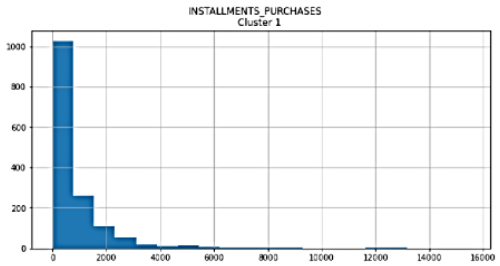


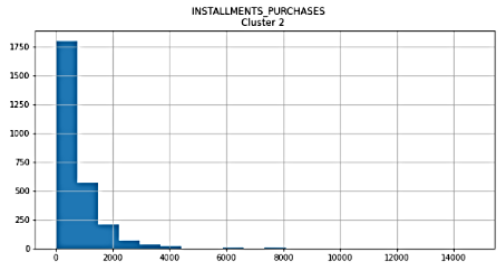
Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
PURCHASES	297.01	3164.99	1015.88
ONEOFF_PURCHASES	235.757	2332.3	236.25

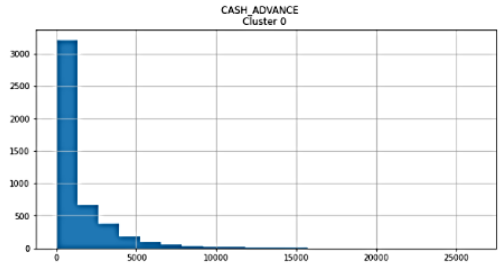
The purchases is highest in the case of Cluster 1 and lowest in the case of Cluster 0.

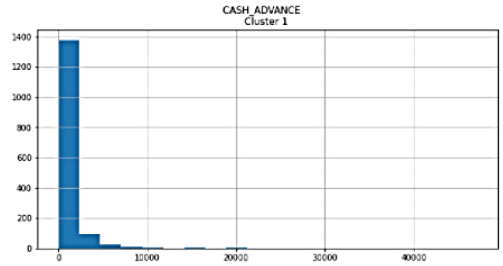
The one-off purchase is highest of Cluster 1 and lowest of Cluster 0.

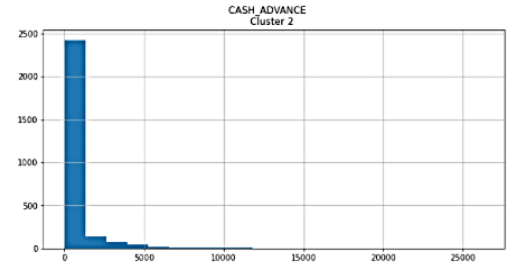
Execution and Analysis









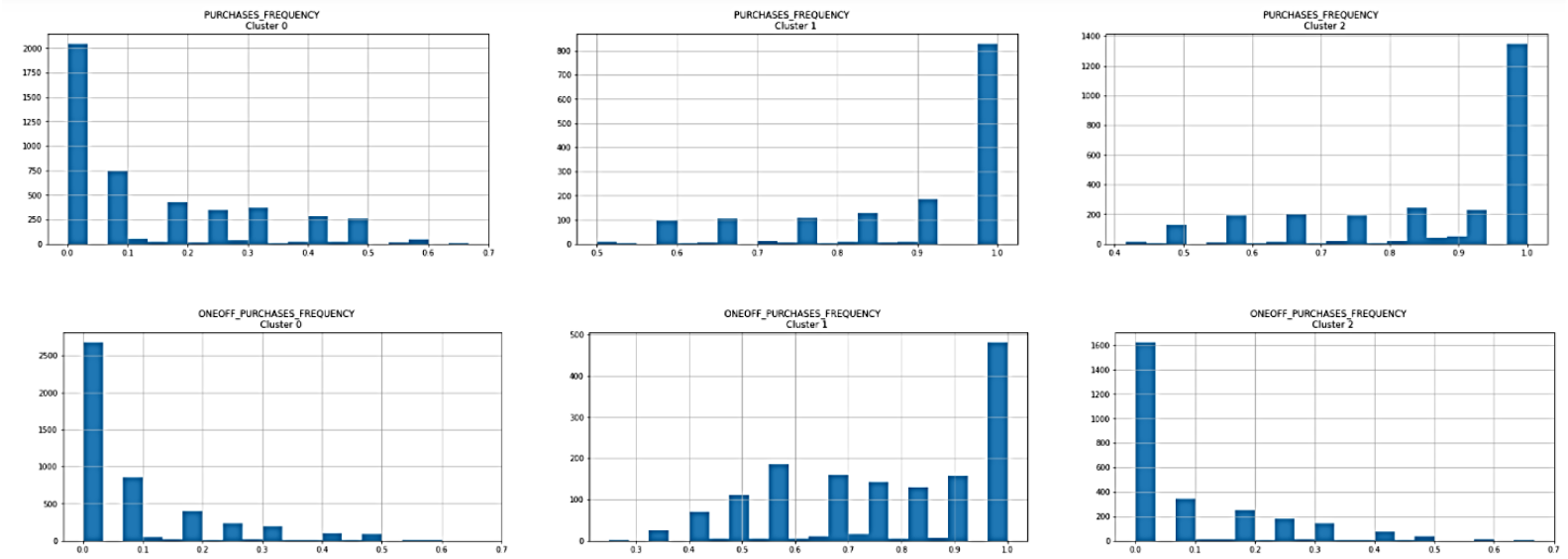


Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
INSTALLMENTS_PURCHASES	61	832.68	780.37
CASH_ADVANCE	1354.83	723.31	471.12

The installments purchase is highest in the case of Cluster 1 and lowest in the case of Cluster 0.

The cash advance is taken highest by Cluster 0 and lowest in the case of Cluster 2.

Execution and Analysis

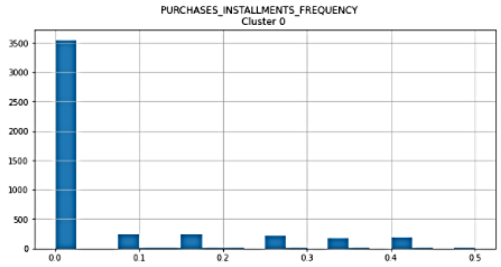


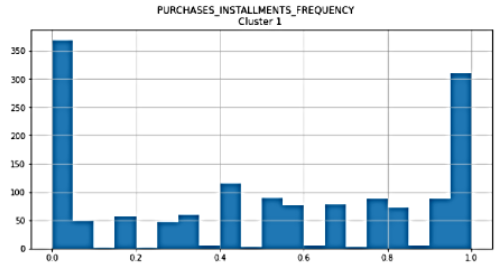
Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
PURCHASES_FREQUENCY	0.14	0.89	0.86
ONEOFF_PURCHASES_FREQUENCY	0.08	0.78	0.08

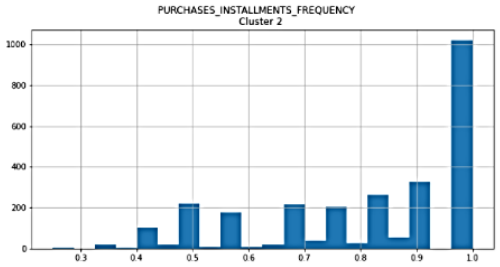
The purchase frequency is highest in the case of Cluster 1 and lowest in the case of Cluster 0.

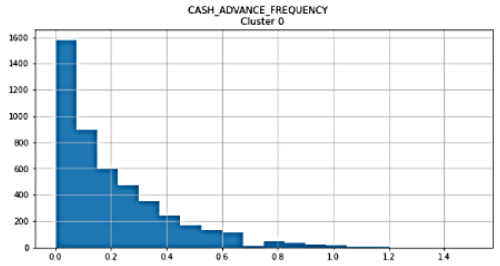
The one-off purchase frequency is highest of Cluster 1 and lowest of Cluster 0 and Cluster 2.

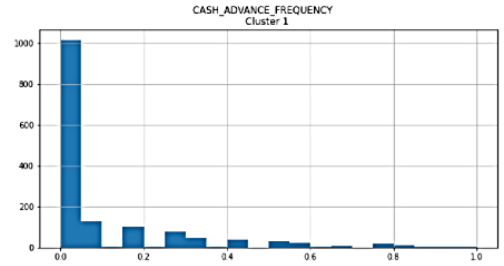
Execution and Analysis

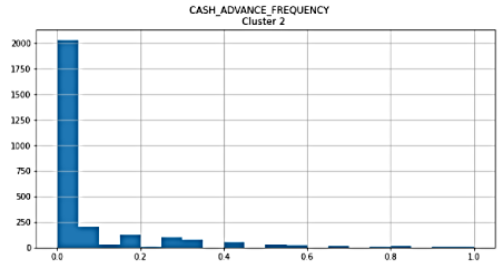










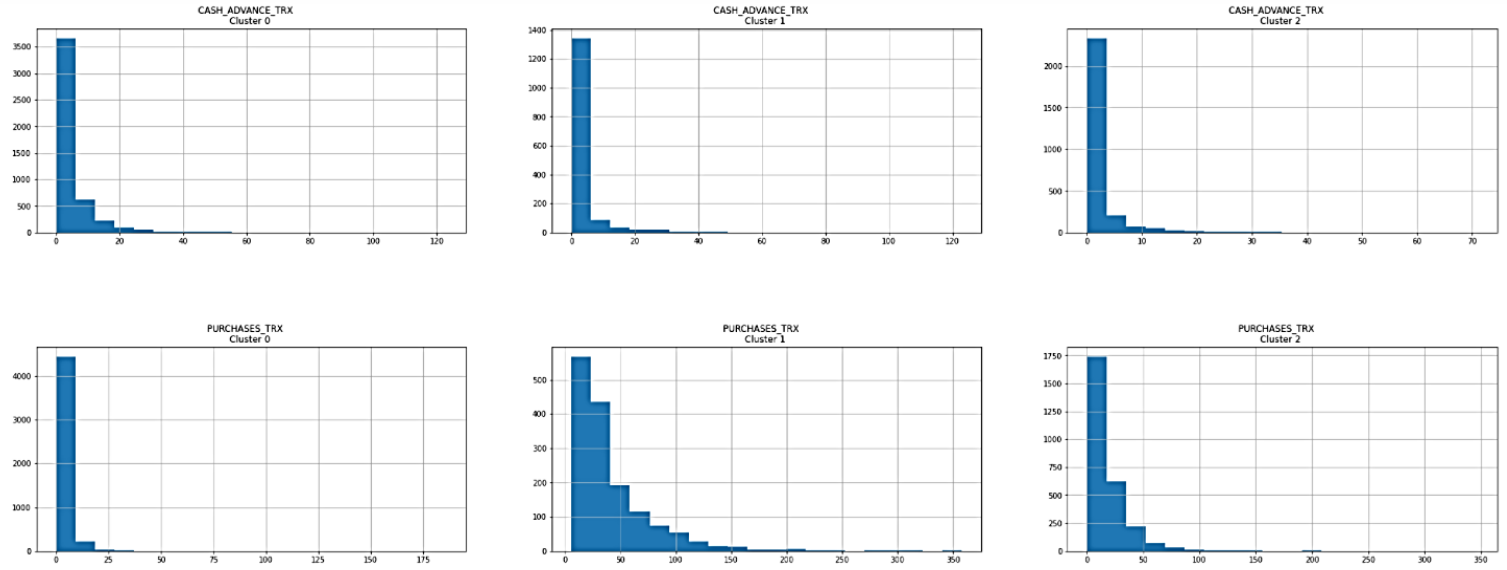


Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
PURCHASES_INSTALLMENTS_FREQUENCY	0.05	0.5	0.81
CASH_ADVANCE_FREQUENCY	0.18	0.09	0.06

The purchase installments frequency is highest in the case of Cluster 2 and lowest in the case of Cluster 0.

The cash advance frequency is highest of Cluster 0 and lowest of Cluster 2.

Execution and Analysis

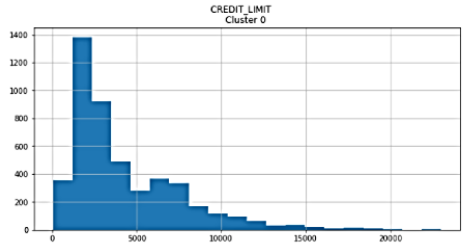


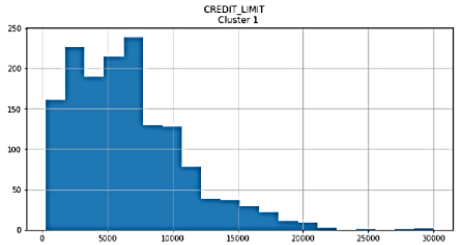
Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
CASH_ADVANCE_TRX	4.42	2.58	1.58
PURCHASES_TRX	2.67	42.43	20.03

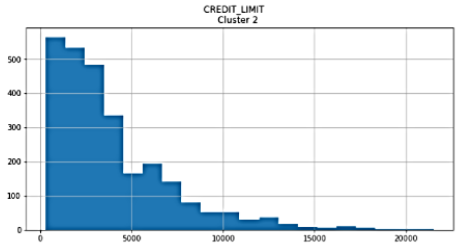
The cash advance transaction is highest of Cluster 0 and lowest of Cluster 2.

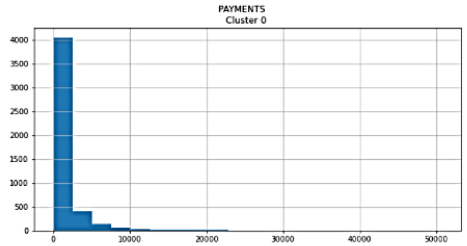
The purchase transaction is highest of Cluster 1 and lowest of Cluster 0.

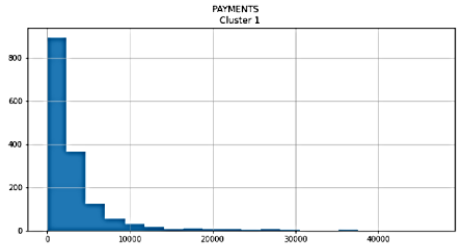
Execution and Analysis

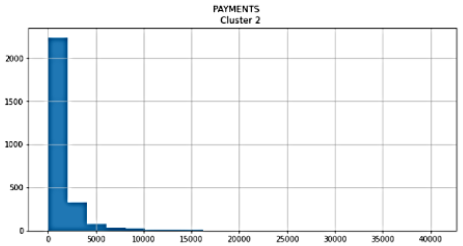










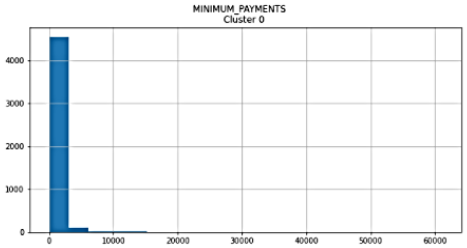


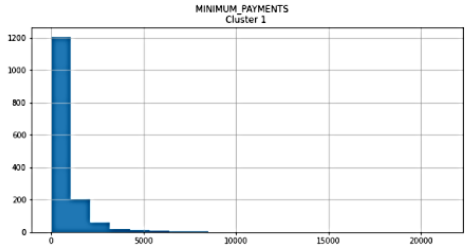
Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
CREDIT_LIMIT	4146.45	6696.44	3864.5
PAYMENTS	1474.31	3235.65	1340.39

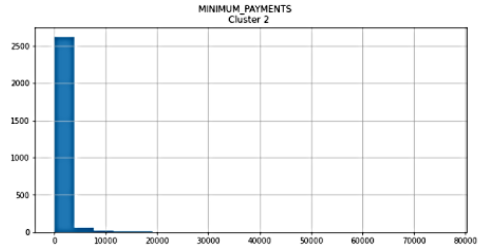
The credit limit is highest of Cluster 1 and lowest of Cluster 2.

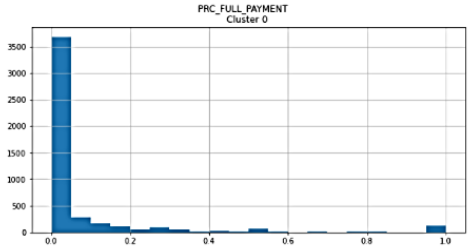
The amount of payment done by user is highest of Cluster 1 and lowest by Cluster 2.

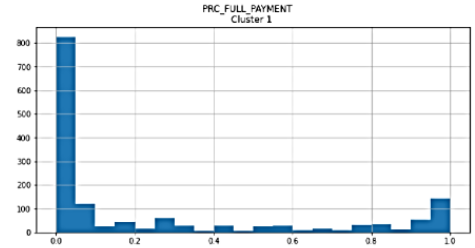
Execution and Analysis

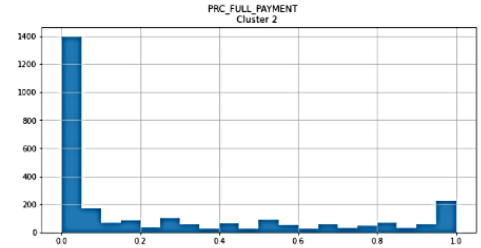










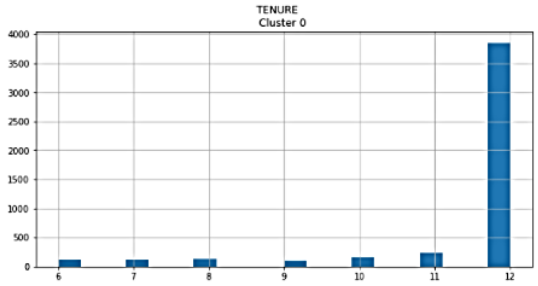


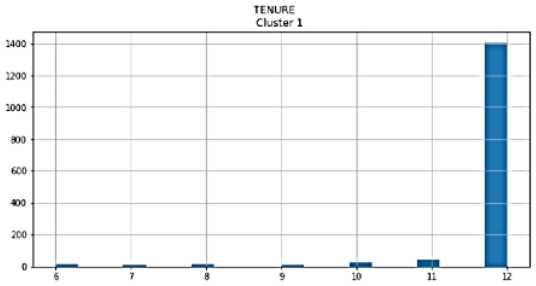
Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
MINIMUM_PAYMENTS	854.22	806.34	913.85
PRC_FULL_PAYMENT	0.06	0.24	0.25

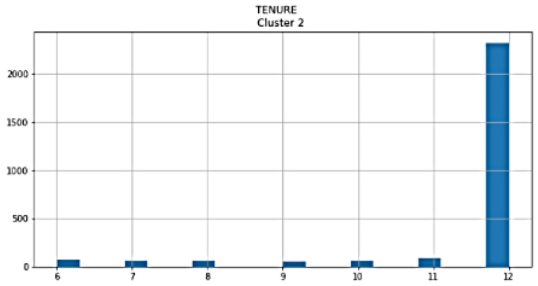
The minimum payments is done highest by Cluster 2 and lowest by Cluster 1.

The percentage of full payment is made highest by Cluster 1 and lowest by Cluster 0.

Execution and Analysis







Column Name	Mean		
	Cluster 0	Cluster 1	Cluster 2
TENURE	11.44	11.78	11.5

The tenure is approximately similar of all the three clusters. The mean tenure of Cluster 1 is 11.78 years, Cluster 2 is 11.5 years, and Cluster 0 is 11.44 years.

Execution and Analysis

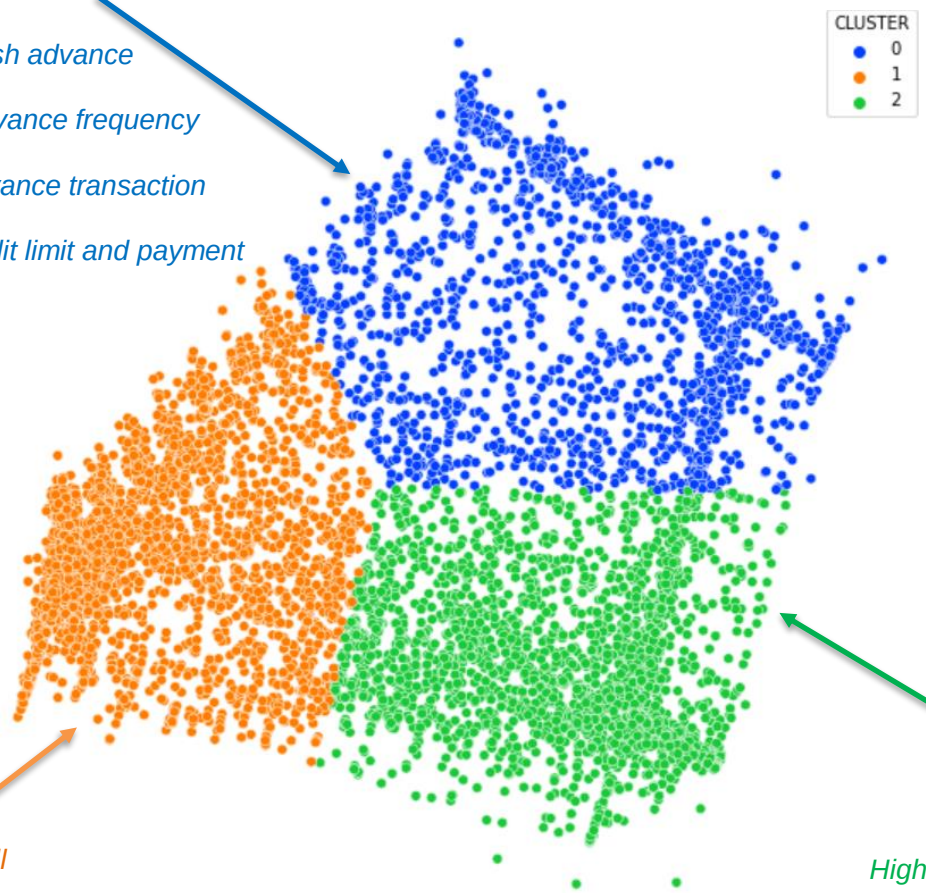
Cluster 0

Highest cash advance

Highest cash advance frequency

Highest cash advance transaction

Second highest credit limit and payment



Cluster 1

Highest purchases

Second highest percentage of full payment

Highest credit limit, purchase frequency, installment purchases

Cluster 2

Highest minimum payment

Lowest credit limit

Second highest installment purchases

Execution and Analysis

Cluster 0 are the cluster of customers who tend to borrow money in the form of cash from the credit card company. Their cash advance transaction, frequency, and amount are the highest.

Cluster 1 are the cluster consisting of customers who are the most frequent user of credit card. They make the highest purchases and purchases transaction. These customers have high credit limit and highest installment purchases.

Cluster 2 are the cluster of customer involving the highest minimum payment, lowest credit limit, and second highest installment purchases customer.

Conclusion

Through K-means clustering we have discovered three type of customers of the credit card company. Cluster 0 also known as the borrower customer, Cluster 1 also known as the star customer, and Cluster 3 known as the question mark customer.

The company should apply proper marketing strategy towards the star customer (cluster 1). These are the active user of the credit card hence more offers should be given to them to further increase their purchase transaction. Also, they can be given offers such as percentage of discount if they spend a certain amount in a particular month. This will increase their purchase and keep them loyal towards the company.

The company should focus their attention on the borrower customer (cluster 0). These customers do give higher interest application charges to the credit card company however the customers are not managing their funds properly. There is a risk involve with such customers, The marketing strategy for such customers should focus on converting them towards the star customer (cluster 1).

The last cluster, cluster 2 also known as the question mark cluster. These are high risk customers. Though the company has given these customers lowest credit limit still marketing strategy should focus on attracting such customers to make purchases on installment and make timely payments.

Reference

- wp.kntu.ac.ir/mohammadi/54.pdf
- www.kaggle.com/rruchi/marketing-datacsv
- towardsdatascience.com/understanding-k-means-clustering-in-machine-learning 6a6e67336aa1
- www.analyticsvidhya.com/blog/2020/12/a-detailed-introduction-to-k-means-clustering-in-python/
- www.imsmarketing.ie/business-strategy/the-importance-of-market-segmentation/#:~:text=Segmentation%20helps%20marketers%20to%20be,most%20likely%20to%20purchase%20products
- www.lotame.com/what-is-market-segmentation/
- ventureandgrow.com/understanding-the-importance-of-segmentation-in-marketing/
- Discovering Knowledge in Data – An introduction to Data Mining by Daniel T. Larose

Thank You