**ISOM 2600 In-class Exercise (Class 3)**

A commercial bank is reviewing their market strategy for their credit card. Based on the customer information, they would like to divide their customers into groups, then develop a specific strategy for the different groups.

The marketing team would like to find and interpret the label of customers, then suggest an appropriate strategy for each group.

In the dataset, there are 4 variables, including:

* Months\_on\_book:   
  Customer relationship with bank in months
* Months\_Inactive\_12\_mon:   
  Number of months inactive in the last 12 months.
* Total\_Trans\_Amt:  
  Total transaction amount in last 12 months
* Total\_Trans\_Ct:  
  Total counts of transaction in last 12 months

The marketing team uses k-means clustering for the segmentation. They first standardize the variables by the mean and SD.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Total\_Trans\_Ct** | **Total\_Trans\_Amt** | **Months\_Inactive\_12\_mon** | **Months\_on\_book** |
| **count** | 8500 | 8500 | 8500 | 8500 |
| **mean** | 69 | 4655 | 2 | 36 |
| **std** | 23 | 3513 | 1 | 8 |

Question 1

Which of the following description about k-means clustering is **false**?

A. K-means clustering involves randomness in the process, and therefore data might belong to different clusters every time it runs.

B. Elbow plot can tell the best number of clusters unambiguously.

C. It is necessary to first decide the number of k before running the k-means clustering.

D. K-means clustering does not require the presence of response variable.

Answer: B

Question 2

The team would like to start clustering analysis based on these two variables first, then extends it later. Consider the months on book and total transaction amount only, they constructed the elbow plot:

Shape

Description automatically generated

Based on the elbow plot, what is the suggested value of K (the number of group)?

A. 2 B. 3

C. 4 D. 5

Answer: C

Question 3

Assume the optimal value of k is used. The team would like to identify the “loyal customer”, who has long relationship with the bank, uses the bank’s card for most of the expense. Recall that the overall average month on book is 36 months, the average total transaction amount is $4655.

Which kind of observations we expect to see for loyal customer?

1. Months\_on\_book < 36, total\_trans\_amt < 4655
2. Months\_on\_book < 36, total\_trans\_amt > 4655
3. Months\_on\_book > 36, total\_trans\_amt < 4655
4. Months\_on\_book > 36, total\_trans\_amt > 4655

Answer: D

Question 4

Then we consider all variables. The team decide to use k = 5 as the optimal number and then run the k-means clustering. Some results are given below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Total\_Trans\_Ct** | **Total\_Trans\_Amt** | **Months\_Inactive\_12\_mon** | **Months\_on\_book** | **Label** |
| **1** | -0.203886 | -0.466797 | 0.714319 | -2.104460 | 3 |
| **2** | 1.105148 | -0.100968 | -1.252865 | -0.982453 | 4 |
| **3** | -0.509327 | -0.559607 | 0.714319 | -1.356456 | 3 |
| **4** | 1.672396 | 3.291424 | 0.714319 | -1.481123 | 1 |
| **5** | 0.319728 | -0.188938 | -1.252865 | 1.136894 | 4 |

The team would like study the shared attributes among the group. They compute the average of all attributes by groups as below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Total\_Trans\_Ct** | **Total\_Trans\_Amt** | **Months\_Inactive\_12\_mon** | **Months\_on\_book** |
| **0** | 0.24 | -0.07 | 1.08 | 0.25 |
| **1** | 1.81 | 2.81 | -0.06 | -0.11 |
| **2** | -1.38 | -0.85 | -0.12 | 0.56 |
| **3** | -0.32 | -0.41 | -0.14 | -1.42 |
| **4** | 0.33 | -0.01 | -0.73 | 0.28 |

The team would like to interpret the ‘label 0’. Which of the following interpretation suits the best, based on the shared attributes?

A. Inactive customer B. Loyal customer

C. New customer D. Usual customer

Answer: A

Question 5

Given a customer with the standardized attributes:

|  |  |  |  |
| --- | --- | --- | --- |
| **Total\_Trans\_Ct** | **Total\_Trans\_Amt** | **Months\_Inactive\_12\_mon** | **Months\_on\_book** |
| 1.3 | 2 | -0.1 | -0.1 |

Which group should the customer belong to?

A. Group 0 B. Group 1

C. Group 2 D. Group 3

E. Group 4

Answer: B