

Summer 2024: CSEN 5303-600 Programming Assignment #2

DUE: Saturday, June 29, 2024 (**Softcopy @ 11:59 pm via Blackboard**)

Instructions

- ❑ **All work must be your own other than the instructor provided data/code and hints to be used. You are not to work in teams on this assignment.**

Assignment Description

Dataset: You are required to work on the dataset provided accompanying this assignment. The dataset provided is a credit card dataset (can be found under dataset folder), which can be utilized to perform clustering to define marketing strategy. The dataset summarizes the usage behavior of about 9000 active card holders during the last 6 months. The data file consists of 18 behavioral variables.

Following is the Data Dictionary for credit card dataset:-

CUSTID : Identification of Credit Card holder (Categorical)

BALANCE : Balance amount left in their account to make purchases

BALANCEFREQUENCY : How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated)

PURCHASES : Amount of purchases made from account

ONEOFFPURCHASES : Maximum purchase amount done in one-go

INSTALLMENTSPURCHASES : Amount of purchase done in installment

CASHADVANCE : Cash in advance given by the user

PURCHASESFREQUENCY : How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)

ONEOFFPURCHASESFREQUENCY : How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased)

PURCHASESINSTALLMENTSFREQUENCY : How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done)

CASHADVANCEFREQUENCY : How frequently the cash in advance being paid

CASHADVANCETRX : Number of Transactions made with "Cash in Advanced"

PURCHASESTRX : Number of purchase transactions made

CREDITLIMIT : Limit of Credit Card for user

PAYMENTS : Amount of Payment done by user

MINIMUM_PAYMENTS : Minimum amount of payments made by user

PRCFULLPAYMENT : Percent of full payment paid by user

TENURE : Tenure of credit card service for user

PART (A) [30 points]

K-means Clustering: In this part, you will be utilizing K-means clustering algorithm to identify the appropriate number of clusters. You may use any language and libraries to implement K-mean clustering algorithm. Your K-mean clustering algorithm should look for appropriate values of K at least in the range of 0 to 15 and show their corresponding sum-of-squared errors (SSE).

PART (B) [40 points]

Hierarchical Clustering: In this part, you will apply hierarchical clustering algorithm (agglomerative or divisive) to the provided credit card dataset. You may use any language and libraries to implement hierarchical clustering algorithm.

PART (C) [30 points]

Density-based Clustering: In this part, you will apply density-based clustering algorithm to the provided credit card dataset. You may use any language and libraries to implement density-based clustering algorithm.

PART (D)

Submission: You will need to submit two files: (a) Programming code files, and (b) a report file consisting a detail explanation of your approaches and the results obtained for all of the clustering algorithms required in Parts A, B, and C. Put all the files in a folder: PA1_<Name_ID> and then compress it and submit the compressed file via Blackboard. Your compressed file should look like PA1_<Your_name_id>.zip.

Inside the PA1_<Your_name_id>.zip, you must submit a single paged document declaring that the submitted work is **completely yours and you have not either partially or fully copies someone else's assignments directly or indirectly.**

Without part (D) being submitted, the assignment will NOT be graded, and the default score will be '0' (zero).

--- X ---