Machine Learning

Practical Project 4
Unsupervised Learning

Consider the "*Credit Card Dataset*", available at Kaggle (https://www.kaggle.com/datasets/arjunbhasin2013/ccdata), and also at the course web page (https://www.di.ubi.pt/~hugomcp/ml/pratical_set_4.zip). It summarizes the usage behavior of about 9,000 active credit card holders for 6 months. The file is provided at a customer level, with 18 behavioral variables:

- **CUST_ID** : Identification of Credit Card holder (Categorical);
- **BALANCE** : Balance amount left in their account to make purchases;
- **BALANCE_FREQUENCY** : 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;
- **ONEOFF_PURCHASES** : Maximum purchase amount done in one-go;
- **INSTALLMENTS_PURCHASES** : Amount of purchase done in installment;
- **CASH_ADVANCE** : Cash in advance given by the user;
- **PURCHASES_FREQUENCY** : 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";
- **PURCHASES_TRX** : Numbe of purchase transactions made;
- **CREDIT_LIMIT** : 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.

Use different clustering strategies (e.g., K-Means, DBSCAN, SOM or Hierarchical clustering methods) to obtain "k" consistent groups of customers. The appropriate value of "k" should also be obtained by yourself.