Q1. Refer to the data below, apply the 2-means algorithm (i.e. with 2 clusters) to identify two segments in the customer base shown below. Use customers b and d as initial means for the two clusters. Describe each step, and show all relevant calculations (Hint: you don’t necessarily need to use calculators to compute the distances. You only need to tell which distance is bigger).

|  |  |  |
| --- | --- | --- |
| Customer ID | Visits frequency per month | Average order amount ($) |
| a | 2 | 2 |
| b | 6 | 4 |
| c | 2 | 0 |
| d | 4 | 4 |

Do not use computer to solve this question, just handwriting and take photos (pasted in a word document)/

A close-up of a notebook

Description automatically generatedA close-up of a notebook

Description automatically generatedA close-up of a paper

Description automatically generated

Q2.

Please download the dataset from Beachboard (Country-data.csv, data-dictionary.csv).

**Title: Clustering the Countries by using Unsupervised Learning for HELP International**

**Objective:**

To categorize the countries using socio-economic and health factors that determine the overall development of the country.

**About organization:**

HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities.

**Problem Statement:**

HELP International have been able to raise around $ 10 million. Now the CEO of the NGO needs to decide how to use this money strategically and effectively. So, CEO has to make decision to choose the countries that are in the direst need of aid. Hence, your Job as a Data scientist is to categorize the countries using some socio-economic and health factors that determine the overall development of the country. Then you need to suggest the countries which the CEO needs to focus on the most.

This problem is from <https://www.kaggle.com/rohan0301/unsupervised-learning-on-country-data>

Please try to analyze the data by yourself first before checking the answer of other people provided in Kaggle.

This is an open end questions. You many follow the car sales example (in file Chapter 15 - Cluster analysis.ipynb at Canvas) to analysis this problem. Please submit your .ipynb file to me.