CS- 483
HW 5
Due: Monday, November 1, 2021
[120 points]

You are going to practice clustering on two different scenarios:

1. **Customer Segmentation**

When it comes to marketing, customer segmentation is really important. In Customer Segmentation's project, we are looking for grouping customers to help the marketing team to make a better offer.
Use the "Bank Marketing Data Set "from UCI machine learning repository.
https://archive.ics.uci.edu/ml/datasets/Bank+Marketing
Each team needs to decide which attributes to use, and provide the reason for dropping or choosing specific attributes. You need to use the 10-fold cross-validation technique to make your train and test data set and use the K-means algorithm for clustering. Choose the number of clusters wisely and explain how did you decide (Elbow method or Silhouette method). Make sure to add the confusion matrix and correlation matrix's results as well. [1]

2. **Patient segmentation**

Patient segmentation is usually based on the elements such as the assessment, definition, and operationalization of population or patient characteristics that are related to healthcare needs, outcomes aimed at when patient needs or addressing population, and the segmentation logic expressing how subpopulations or patient groups are formed.
In this project, each team needs to segment the COVID-19 patients. Use the K-means clustering by holding 20% of the data for the test.
COVID19 dataset contains demographic information (e.g. number of patients in the cohort, age, and gender), treatments (including antibiotics, intubation, etc.), and outcomes (including discharge, hospital length of stay, death, etc.). Please use the link below to download the data:
https://www.covidanalytics.io/dataset

What do you need to submit for each scenario?

1. Clear explanation of the chosen data
2. Your code
3. Confusion matrix (analyze the result)
4. Correlation matrix (analyze the result)
5. Plot Elbow (or Silhouette) method
6. Plot the final clustering
7. Summarize steps that you have taken to derive at the result an explanation of the logic for each step

[1] Hint: https://data-flair.training/blogs/r-data-science-project-customer- segmentation