Hackathon

You are given a health care data of patient containing their Demographical parameters, Diet parameters and lab-tested parameters for 2019-2020.Based on the given data we want to patient with similar kind and group them together. Your task is to identify the similar patient and provide same group number for them.

The data parameters were collected from the National Health and Nutrition Examination Survey of National Center for Health Statistics organised by Centers for Disease Control and Prevention.

For more information about the parameter visit https://wwwn.cdc.gov/Nchs/Nhanes/

Data Description :

The **Data_Description.csv** file contains the description of the dataset

Metric to measure :

The metric to measure is the performance is **silhouette** score. In this Hackathon, you will get the silhouette score across all the clusters for the appropriate number of cluster labels submitted by the participants. .Please visit the link for more details on silhouette score-
https://en.wikipedia.org/wiki/Silhouette_(clustering)

If a student performs a data reduction technique on the dataset, the utmost variation of the dataset should be captured

Marking Procedure :

The score of the student will be converted into marks taking into consideration the relative score in the leaderboard and converting that into 20.

Submission :

Participant can submit their final file using

https://docs.google.com/forms/d/e/1FAIpQLScTZVlg_iuW2QH9MEm5gAt4SkG6o6MNaBxf E-QGIOet6lot7g/viewform?usp=sf_link

A student can submit a maximum of 10 times.

File Format: You should submit a CSV file with exactly **9831 entries** plus a header row. **The file should be saved with the participant name.** For example, a student with name Sutithi Chakraborty will submit their 4th submission file as SutithiChakraborty_4.csv

The file should have exactly two columns ● SEQN ● Cluster

Score Board

You can see the LeaderBoard using the link below

https://drive.google.com/file/d/1eVVYfMFesB--8IXo_CWhEDrklaER6Jdq/view

The score board will be updated every 10 minutes.