**Problem-2**

* The file "train\_data.tsv" contains training data. Note that in bioinformatics datasets, features (genes) are represented in the row and samples are in the column. Therefore, in the file, each row represents a different feature, and each column is a sample.
* The file "train\_label.tsv" contains information about the samples. Specifically, the column "X\_primary\_disease" tells you what kind of cancer a sample is (breast, kidney, or lung adenocarcinoma).
* Your task is to train a machine learning model using this data and predict a label for each sample in the file "test\_data.tsv".
* Bonus points if you perform dimensionality reduction, visualization, and other unsupervised learning analyses of the data as well.
* Write a report (maximum 1-2 pages) describing your approach and key information such as cross-validation accuracy, and algorithms you have used. The file should be named "ML\_analysis\_YourName.pdf".
* Also submit the predicted level for samples in the "test\_data.tsv" file. The predicted label file should be names "ML\_predict.csv"
* Send all the files by email