**Problem-1**

**If you have a bioinformatics background:**

* The file "TCGA-BRCA.htseq\_counts\_gene\_name.tsv" contains gene expression data. These are RNAseq count values in the form of log2(count+1).
* Perform differential gene expression analysis using the Deseq2 package. The aim is to find the differentially expressed genes between "Tumor" and "Normal". Sample labels are present in the "TCGA-BRCA.pheno.tsv" file under the column "type".
* Also, perform pathway enrichment analysis using the KEGG pathway database.
* Note that not all samples are present in both files. You will need to first subset the samples that are present in both files.
* Write a report (maximum 2 pages), briefly describing your approach and the key genes and pathways that resulted from your analysis. Put figures and references in the report. The file should be named "diff\_analysis\_YourName.pdf".
* Also submit the output table (in csv format) of differential gene expression and pathway analysis. The files should be named "diff\_result\_YourName.csv" and "pathway\_result\_YourName.csv".
* Send all the files by email

**If you do not have a bioinformatics background:**

* The file "TCGA-BRCA.htseq\_fpkm-uq\_gene\_name.tsv" contains cancer gene expression 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 "TCGA-BRCA.pheno.tsv" contains information about the samples. Specifically, the column "type" tells you if a sample is "Tumor" or "Normal".
* Your task is to find features that are associated with the "type". Select the top 50 features that are highly associated with the "Tumor" or "Normal" label, perform clustering using these features, and create a heatmap.
* Note that not all samples are present in both files. You will need to first subset the samples that are present in both files.
* Write a report (maximum 2 pages), briefly describing your approach. Also, cite research papers that support your findings. Put figures and references in the report. The file should be named "feature\_analysis\_YourName.pdf".
* Also submit the output table (in csv format) of your top 100 features. The files should be named "feature\_result\_YourName.csv"
* Send all the files by email