Group coursework 2

- Please upload your report in **one pdf file** and **one code script**.

- The report should be no more than 1000 words and 5 pages, with font size of at least 12, margin size of at least 1 inch and 1.5 line spacing.

- Please do not submit .ipynb or .rmd files as the report.

- This coursework will be assessed based on the quality of your report, e.g. presentation, quality of plots, details about the experiment settings, motivations to use the methods, analysis of the results, conclusions etc.

- Make sure that you have included **sufficient** comments in the codes to make them **readable** by other people. There should be **no error messages** shown when I run your scripts. You can assume that I have installed all required packages.

Download the `IMDB-Movies.csv` data from Moodle. This dataset contains information of 50 top-rated movies from IMDb.

1. Apply PCA to the numerical variables, excluding Year. Summarise and interpret your key findings.

2. Perform clustering on the movies using only the numerical variables, excluding Year.

   - Clearly explain why you chose the specific clustering method(s).

   - Report and interpret the clustering results.

3. Use ChatGPT-4o to perform the same tasks as in questions 1 and 2.

   - Compare the output from ChatGPT-4o with your own solution.

   - Provide a critical evaluation of the similarities and differences between the two, and comment on the reliability and usefulness of ChatGPT-4o in this context.

Note: For Question 3, please include both the prompts you used and the full output generated by ChatGPT-4o in the Appendix of your report.