Mini Project

This project aims to apply the analytical skills you have gained in **DSC4033/STA4053 - Multivariate Methods II** to a real-world problem. Your task is to apply various multivariate techniques to analyze and interpret patterns, relationships, and structures within the data.

Due Date of submission: 10th May 2025

Analysis

Dataset Selection

- Select a suitable dataset with more than 10 variables and at least 500 observations. The dataset should be relevant to a real-world application, such as finance, healthcare, social sciences, or business analytics.
- Ensure the dataset contains both continuous and categorical variables to facilitate a broader analysis.

Statistical Analysis

Apply the multivariate techniques learned in class to analyze your dataset. If necessary, you may also consider techniques not covered in class.

Software Usage

Use R or Python to perform the analysis. Provide clear documentation of code and outputs.

Format of the Report

It is necessary to submit a fully referenced report with the following sections.

1. **Introduction**: Provide a clear statement introducing the major question, the study's purpose,
2. **Methodology**: Describe the dataset and the statistical methods employed.
3. **Results and discussion**: Present the results clearly with appropriate visualizations and interpretations.
4. **Conclusion and recommendation**: Summarize the main findings and discuss any limitations.
5. **References**: Use APA style for writing references.

6. **Appendices** (Mandatory):
    - Include a part of the dataset (or link if publicly available).
    - Provide all R/Python codes in Markdown format with proper documentation.

Report Length

The report should not exceed 15 pages, including appendices. Use clear headings, bullet points, and visuals to enhance readability.