SCS2211 - LABORATORY II
R Lab Practical Sheet - 12

Instructions

- **Do the Activities and save in a .ipynb file**
- **File name should be <Index number>.ipynb (Eg: 2000000.ipynb) and upload to the given link.**
- **Any form of plagiarism or collusion is not allowed**

Objective:

The goal of this assignment is to analyze a given dataset using R, apply appropriate statistical and machine learning techniques, and generate a detailed report on your findings.

Instructions:

1. **Dataset Selection:**
    - Choose a dataset from an open-source repository (e.g., Kaggle, UCI Machine Learning Repository) or use a dataset provided by your instructor.
    - Ensure the dataset has multiple variables and at least 500 observations.
2. **Data Preprocessing:**
    - Load the dataset into R.
    - Handle missing values (e.g., imputation or removal).
    - Perform exploratory data analysis (EDA) with summary statistics, visualizations, and correlation analysis.
3. **Data Visualization:**
    - Use ggplot2 or base R to create meaningful visualizations (e.g., histograms, scatter plots, box plots, and bar charts).
    - Identify trends, patterns, or outliers in the data.
4. **Statistical Analysis:**
    - Perform descriptive statistics (mean, median, variance, standard deviation).
    - Conduct hypothesis testing if applicable (e.g., t-tests, ANOVA, chi-square tests).

5. **Interpretation & Discussion:**
    - Summarize the key findings from your analysis.
    - Discuss the implications of your results.
    - Highlight any limitations or challenges faced during the analysis.
6. **Report Writing:**
    - Write a structured report (1500-2500 words) including the following sections:
        - **Title Page**: Assignment title, your name, date.
        - **Introduction**: Overview of the dataset and objectives.
        - **Methodology**: Steps taken in analysis and justification.
        - **Results**: Findings with visualizations and tables.
        - **Discussion**: Interpretation and significance of results.
        - **Conclusion**: Summary and potential future work.
        - **References**: Cite sources (if any).
7. **Submission Requirements:**
    - Submit the R script (.R file) and the report (PDF or Word document).
    - Ensure your R code is well-commented and organized.