**THE UNIVERSITY OF HONG KONG**

**DEPARTMENT OF COMPUTER SCIENCE**

**FITE7410 Financial Fraud Analytics**

**First Semester, 2023-2024**

**Assignment 1**

**(Due Date: 15 Oct, 2023 (Sun) 23:59)**

**Assessment Criteria:**

* Plagiarism: Please follow the guidelines laid down by our department.
* You are allowed to discuss the assignment with your classmates, however, direct copy and paste is PROHIBITED and would be considered as PLAGIARISM.
* Assignments would be marked based on the logic, presentation and understanding of the problem; not only on accuracy.
* LATE PENALTIES: 50% of assignment marks will be deducted for late submissions. 0 marks if the submission is later than 2 weeks.

1. (50%) Exploratory Data Analysis
   1. Download the ENRON Dataset from Moodle.
   2. Using the R package, conduct exploratory analysis of the dataset downloaded

NOTE: A sample R script is provided, but you still need to complete the program. Or you can write the script by yourselves and use whatever library you like.

1. (50%) Write a 2-3 pages essay on the following:
   1. Describe the dataset based on the exploratory analysis result, including:
      * Summary description of the dataset
      * Univariate analysis
      * Bi-/Multi-variate analysis
      * Missing data/Outlier analysis
2. Submission on Moodle:
   1. R language script
   2. A **pdf** version report