UGANDA CHRISTIAN UNIVERSITY
A Centre of Excellence in the Heart of Africa

FACULTY OF ENGINEERING, DESIGN AND TECHNOLOGY
DEPARTMENT OF COMPUTING AND TECHNOLOGY
EASTER 2025 SEMESTER **TEST**

PROGRAM: *BACHELOR OF SCIENCE IN DATA SCIENCE*

YEAR: 2 SEMESTER: 2

COURSE NAME: *WEB MINIG AND WRANGLING*

TIME ALLOWED: *1.5 weeks*

Examination Instructions

1. The general Uganda Christian University examination guidelines and academic & financial policies apply to this examination. Violating any of the policies by the student automatically makes this examination attempt void, even if you have completed and submitted the answer booklet.
2. This exam consists of a project to be executed in *one and half* weeks.
    a. Assessment of the project shall be based on five milestones, evaluated during the duration of the project. Each milestone shall be evaluated out of 20 marks.
    b. At the end of the project, the following SHALL be submitted on Moodle.
        i. A well-written project (Font: Trebuchet MS, 12Pts, 1.5 spacing, justified aligned), IEEE Referencing style.
3. Every student has a responsibility to prove their contribution towards every milestone, and marks may be awarded to every student individually.

PART A:

Instructions: Your task is to scrape movie data from a publicly accessible movie website that does not have anti-scraping mechanisms.

GOAL: To extract various movie details, store them in a structured format, and analyze them.

You are required to:

1. Scrape movie data from an identified website and store the extracted data in a Pandas DataFrame. Ensure you include at least five different movie parameters. Save your output as a CSV file labelled with your last name (e.g., Groupname_movies.csv).

   Example Movie Parameters: You might consider scraping parameters such as:

   - Movie Title
   - Genre(s)
   - Director
   - Cast (Main Actors)
   - Year of Release
   - IMDb Rating (or similar rating)
   - Runtime
   - Synopsis/Plot Summary
   - Budget
   - Box Office Gross
   - Awards (if easily available)

2. Explore and analyze the scraped dataset. Answer the following:
   1) What are the most common movie genres in your dataset? [2 MARKS]
   2) What is the average IMDb rating of the movies scraped? [2 MARKS]
   3) What trends or patterns can be identified from the data?

3. Identify which movie parameters have the highest correlation with "IMDb Rating" (or a similar rating parameter you scraped) using statistical analysis and visualization techniques.

~END~