Assignment 2
Advanced SQL for Data Analysis and Visualization

This assignment is to test your ability to write advanced SQL queries to address a variety of data analysis questions and visualize them in an appropriate way to understand the result.

Instruction:

- Use any relational database (e.g., SQL Server, MySQL, PostgreSQL) for database creation and relation insertion.
- Execute SQL queries to validate the accuracy of results on the same plateform.
- Populate some arbitrary data as instructed.
- Utilize visualization tools such as Tabulae, Power BI, or Streamlit for clear and informative presentation of the results.

1. Database Schema Design (1 marks)
   Create a database name "ADTAssignment2" with the following table:

   `program` (programID, name)
   `depCourse` (courseID, deptName, programID)
   `users` (userID, programID)
   `courseSiteVisit` (visitID, courseID, userID, date)

2. Data Population (1.5 mark)
   Populate each table to meet the specified criteria:

   Insert sample data,
   1. there are at least two departments (for example, Computer Science and Maths).
   2. there are at least 2 programs (for example, undergrad and Master).
   3. there are at least 2 courses in each program per department.
   4. there are at least 3 users in each program.
   5. populate the courseSiteVisit table with at least 100 records with following criteria:
      i. there are at least 2 users visited all courses.
      ii. all users should have visited at least 1 course.
      iii. each user should have visited at least 1 course multiple times on the same date.
      iv. each user should have visited multiple dates per a single course.

3. Data Analysis (2.5 mark) and Visualization (2.5 mark)
   a. The total number of times a course has been visited or accessed by all users.
   b. The total number of visits for each course, categorized by program.
   c. The total number of students or users enrolled in each program.
   d. The total number of unique visitors per department by program.
   e. The most recent date (or last date) on which a user visited each course.
   f. The number of times a user has visited each course.

g. The user who has visited a course the most (i.e., most frequent visitor per course), along with the visit count.
h. The user who visited a course the most times in a single day. (i.e., most frequent visited user in single day per course)
i. Longest visit streak days per user per course (i.e., the maximum number of days in a row that a user has accessed or engaged with the course site)
j. Longest gap between visit per user and number of days in single course.
k. The user who visited the most courses within a short duration.

Submission:

- Meet GA to demonstrate your work: they will check all the above tasks from your laptop.
- They will question you to explain the reason for using certain statements.
- Submit a pdf document with the following:
    - SQL file
    - Report to explain the results including the visualization.