Modern Data Management & Business Intelligence
Assignment #1 – Due Date: November 13th, 2023 – Groups of two students

You can use either MySQL or Microsoft's SQL Server

Send email to the TA (ssafras@gmail.com) if you are really stuck, first check Google!!!

Description of the Case:
A telecom provider (TelcoX) wants to develop a relational database to monitor customers, calls and plans. Customers of TelcoX are described through a unique identifier, first and last name, date of birth, gender ('male' or 'female') and live in a city. Cities are described by a unique identifier, name, population and mean income. A customer has one or more contracts with TelcoX. A contract has a unique identifier, phone number, starting date, ending date and a description. A contract is also associated to a plan offered by TelcoX (e.g. Red1 of Vodafone). A plan is described by a unique identifier, name, free-minutes, free-sms and free-MB attributes. Finally, calls made by a phone number have to be stored, along with a unique identifier, the date/time of the call (hour, minute, day, month, year), the called phone number and the duration of the call (in seconds).

Deliverables (in <u>one</u> word document):

1. (10%) Use the Entity-Relationship Diagram (ERD) to model entities, relationships, attributes, cardinalities, and all necessary constraints. Use any tool you like to draw the ERD.
2. (10%) Create the relational schema in MySQL/SQLServer and insert a few records into the tables to test your queries below. You will have to hand in the CREATE TABLE statements.
3. (60%) Write SQL code and test it to your data for the following queries

    a. Show the call id of all calls that were made between 8am and 10am on June 2022 having duration < 30
    b. Show the first and last name of customers that live in a city with population greater than 20000
    c. Show the customer id that have a contract in the plan with name LIKE 'Freedom' (use nested queries).
    d. For each contract that ends in less than sixty days from today, show the contract id, the phone number, the customer's id, his/her first name and his/her last name.
    e. For each contract id and each month of 2022, show the average duration of calls
    f. Show the total duration of calls in 2022 per plan id
    g. Show the top called number among TP's customers in 2022
    h. Show the contract ids and the months where the total duration of the calls was greater than the free minutes offered by the plan of the contract
    i. For each month of 2022, show the percentage change of the total duration of calls compared to the same month of 2021
    j. For each city id and calls made in 2022, show the average call duration by females and the average call duration by males (i.e. three columns)
    k. For each city id, show the city id, the ratio of the total duration of the calls made from customers staying in that city in 2022 over the total duration of all calls made in 2022, and the ratio of the city's population over the total population of all cities (i.e three columns)

4. (20%) Using the programming language of your choice, connect to the database and implement query (k) above – ***without using GROUP BY SQL statements***,