**CSCI 341 Assignment 2**

You are given a description of the database tables below. You are supposed to create a relational database (**part 1**) and write and execute queries that will extract and update some data from the database (**part 2**).

**Part 1 (30 points)**

Construct a physical database according to the given specifications (schema) below. To create the physical database, it is required to use a database management system. You may use one of the free DBMSs -- PostgreSQL or MySQL.

*The schema is:*

* **DiseaseType**(id:integer, description:varchar(140))
* **Country**(cname:varchar(50), population:bigint)
* **Disease**(disease code:varchar(50), pathogen:varchar(20), description:varchar(140), id:integer) References **DiseaseType** (id)
* **Discover**(cname:varchar(50), disease code:varchar(50), first enc date:date) References **Disease** (disease code), Country (cname)
* **Users**(email:varchar(60), name:varchar(30), surname:varchar(40), salary:integer, phone:varchar(20), cname:varchar(50)) References **Country** (cname)
* **PublicServant**(email:varchar(60), department:varchar(50)) References **Users** (email)
* **Doctor**(email:varchar(60), degree:varchar(20)) References **Users** (email)
* **Specialize**(id:integer, email:varchar(60)) References **DiseaseType** (id), **Doctor** (email)
* **Record**(email:varchar(60), cname:varchar(50), disease code:varchar(50), total deaths:integer, total patients:integer) References **Disease** (disease code), **Country** (cname), **PublicServant** (email)

After constructing the database model, you should create tables and insert data instances in the tables (at least 10 instances for each table). Please note that the result set of the queries from **part 2** should be non-empty. (This means that you must think of data instances you are going to insert).

**Part 2 (70 points)**

In this part, you are going to connect to your (PostgreSQL or MySQL) database and execute some queries and updates. You are asked to use Python3 and SQLAlchemy library to connect and interact with your database. Please refer to the SQLAlchemy documentation (<https://docs.sqlalchemy.org/en/14/intro.html>). You can use SQLAlchemy ORM or execute the raw SQL scripts using Textual SQL (<https://docs.sqlalchemy.org/en/13/core/tutorial.html#using-textual-sql>). You may choose any approach you prefer as long as it is easy to demonstrate. You can also use IDEs to help you organize your work. Some of the options are DataGrip and Pycharm with Database extensions.

**SQL Queries:**

1. List the disease code and the description of diseases that are caused by “bacteria” (pathogen) and were discovered before 1990.
2. List the name, surname and degree of doctors who are not specialized in “infectious diseases”.
3. List the name, surname and degree of doctors who are specialized in more than 2 disease types.
4. For each country list the cname and average salary of doctors who are specialized in “virology”.
5. List the departments of public servants who report “covid-19” cases in more than one country and the number of such public servants who work in these departments. (i.e “Dept1 3” means that in the “Dept1” department there are 3 such employees.)
6. Double the salary of public servants who have recorded covid-19 patients more than 3 times.
7. Delete the users whose name contain the substring “bek” or “gul” (e.g. Alibek, Gulsim)
8. Create an index namely “idx pathogen” on the “pathogen” field.
9. List the email, name, and department of public servants who have created records where the number of patients is between 100000 and 999999
10. List the top 5 counties with the highest number of total patients recorded.
11. Group the diseases by disease type and the total number of patients treated.

**Submission**

Export your database in .sql format (you can use pg\_dump in order to do this).

Submit a zip file on Moodle that will consist of the following files:

1. .sql file of your database.
2. .py file that will connect to your database and execute all the queries.

**Bonus assignment - 60 points**

This is the bonus part of the assignment. Note that you must complete the first two parts of the assignment if you want to (be eligible to) work on the bonus assignment. Note that you can earn extra points for the course grade.

We ask you to develop a web application (website) that will use the previously constructed database. The web application should have at least basic CRUD (CREATE, READ, UPDATE, DELETE) functionalities. You may use web development frameworks such as Django, Flask, Spring Boot, etc. You would need to deploy your web application on the internet, and we should be able to access it (there are free platforms that you can use such as Heroku, pythonanywhere, etc. In addition, we think that Amazon offers free servers for education purposes (register with @nu.edu.kz emails) where you can host/deploy your web app).

**Submission and Grading**

You need to prepare a video where you demonstrate your implementation, clearly separating Part 1, Part 2, and the bonus part (if you have), in that order. In the video, you need to show how you created your tables, how you populated them, and the code of your queries and their results. Moreover, you need to submit an executive summary of your project (at most one page and it should include a brief description of what you managed to complete and what you could not do). The assignment will be graded live via Zoom or in-class.

Good luck!