

CompSci 301 - Algorithms and Databases

Spring 2021, Session 4 - Database Project

Problem Description – Zovid-12 Infection Database

Currently, a city called Enrangel is suffering from Zovid-12. Your task is to investigate the data collected in Enrangel and make an exposure prediction, in order to implement an email warning system to alert the residents in Enrangel.

The data definition of each table is as follows:

Table Name (PK/FK)	Attributes and Data Types		Notes
audience	Name	VARCHAR	All the audience information who watched the movie on one day.
	id_Number	NUMERIC	
	row_Number	NUMERIC	
	seat_Number	NUMERIC	
	auditorium_Number	NUMERIC	
census	id_Number	NUMERIC	Residents' information in Enrangel.
	city	VARCHAR	
	latitude	NUMERIC	
	longitude	NUMERIC	
zovid12test	id_Number	NUMERIC	Records of the Zovid-12 test.
	result	VARCHAR	
	date	DATE	
hospital_info	hospital_ID	NUMERIC	Eight hospital information in major cities in Enrangel.
	city	VARCHAR	
	latitude	NUMERIC	
	longitude	NUMERIC	
assigned_hospital	city	VARCHAR	Residents in different cities are assigned to nearest hospitals in Enrangel.
	hospital_ID	NUMERIC	

patient_info	name	VARCHAR	Records of the patients information in Erangel.
	id_Number	NUMERIC	
	gender	VARCHAR	
	height_cm	NUMERIC	
	weight_kg	NUMERIC	
positive_cases	id_Number	NUMERIC	Positive cases happened in Erangel.
	case_Number	NUMERIC	
	discharge_Date	DATE	
positive_physical_exam	case_Number	NUMERIC	Physical exam information for all positive cases in Erangel.
	vital_cap	NUMERIC	
	metabolism	NUMERIC	
	exam_date	DATE	
positive_billing_info	case_Number	NUMERIC	Billing information for all positive cases in Erangel.
	bill_Amount	NUMERIC	
	bill_Status	VARCHAR	
patients_physical_exam	id_Number	NUMERIC	Physical exam information for all patients in Erangel.
	vital_cap	NUMERIC	
	metabolism	NUMERIC	
	exam_date	DATE	
city_policy	city	VARCHAR	The table contains the current city policy announced by the government in Erangel.
	level	NUMERIC	
level_description	level	NUMERIC	The table contains the city policy level description announced by the government in Erangel.
	description	VARCHAR	

Tasks

The assignment is divided into **Three** main tasks:

1. Preparation and understanding the database. [15%]

The output of this task is:

Design the E/R diagram of the operational database.

2. Create the short reports using queries. [30%]

After importing the files into MySQL. Based on your own understanding of the database. For each question, you ought to produce the SQL command and the screenshots of the query results (or part of the query results).

Question 1: Calculate the age of each resident and add a column “age” into the census table with their ages.

Hint: You can extract the year of birth from id_Number. For example, id_Number: 68526192306294851 => Age: 2020-1923 = 97

The outputs of this task are:

- (a) The SQL commands,
- (b) The screenshots of the query results (or part of the query results), including all attribute names.

Question 2: Calculate the BMI of each resident and add a column “bmi” into the census table with their BMI.

$$\text{BMI} = \frac{m}{h^2}$$

BMI = body mass index

m = mass (in kilograms)

h = height (in meters)

The outputs of this task are:

- (a) The SQL commands,
- (b) The screenshots of the query results (or part of the query results), including all attribute names.

Question 3: Use the result from the **closest** physical exam date and calculate the body fitness score. Hint: You need to check both physical_exam tables.

Then, add the score into column “score” into the census table.

$$\text{body fitness score} = \left[\frac{\text{vital_cap}}{5200} + \frac{\text{metabolism}}{2800} + \left(1 - \frac{|\text{age}-25|}{75}\right) + \left(1 - \frac{|\text{bmi}-23|}{30}\right) \right] \times 25$$

* Use the average body fitness score from all the residents for someone who missed some of the information required to calculate the body fitness score. **Do not** include the residents who don't have the body fitness score when you calculate the average.

The outputs of this task are:

- (a) The SQL commands,
- (b) The screenshots of the query results (or part of the query results), including all attribute names.

Question 4: Add the zovid-12 test result and body fitness score into column “test_result” and “score” into the audiences table. Then, export the audiences table as .csv.

The outputs of this task are:

- (a) The SQL commands,
- (b) The screenshots of the query results (or part of the query results), including all attribute names.
- (c) The audiences_info.csv file

Question 5: Use **LEFT JOIN** to add the city policy level description into column “description” into the city_policy table, and sort the city name (A-Z). Then, export the city_policy_2 table as .csv.

The outputs of this task are:

- (a) The SQL commands,
- (b) The screenshots of the query results (or part of the query results), including all attribute names.
- (c) The city_policy_2.csv file

Question 6: How many confirmed cases contain the letter “b” or “B” in their names?

The outputs of this task are:

- (a) The SQL commands,
- (b) The screenshots of the query results (or part of the query results), including all attribute names.

3. Integrated

You are required to use C++ to implement an alert system in order to make an exposure prediction.

$$exposure\ time\ (sec) = \lfloor body\ fitness\ score \times 0.06 \rfloor$$

Question 1 [30%]: There are a total of 6 auditoriums in the database. In an auditorium, each seat can have one of three situations:

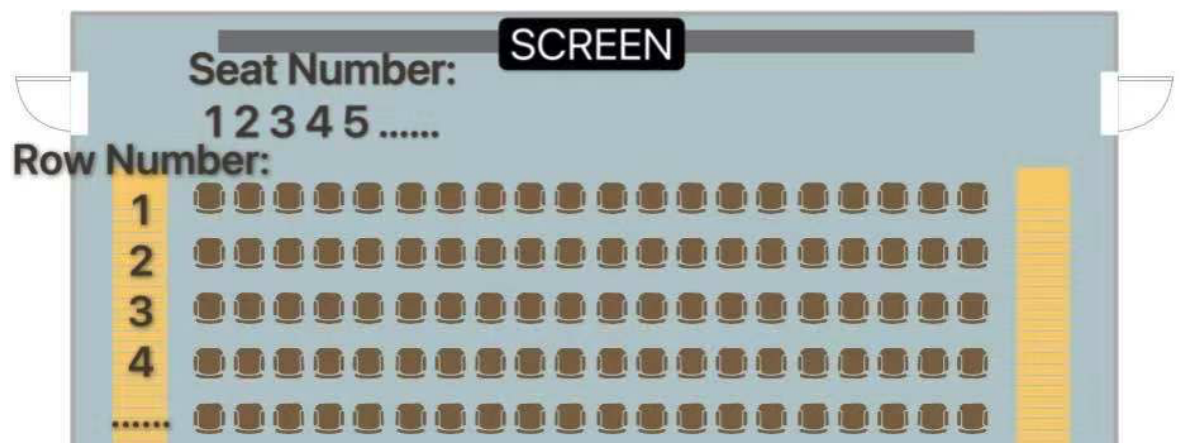
- An empty seat;
- An audience who is received negative in zovid-12 test;
- An audience who is received positive in zovid-12 test;

Every second, any audience who receives a positive result will start to infect another negative result audience that is adjacent (4-directionally).

Return the seconds when all the in danger audiences are infected.

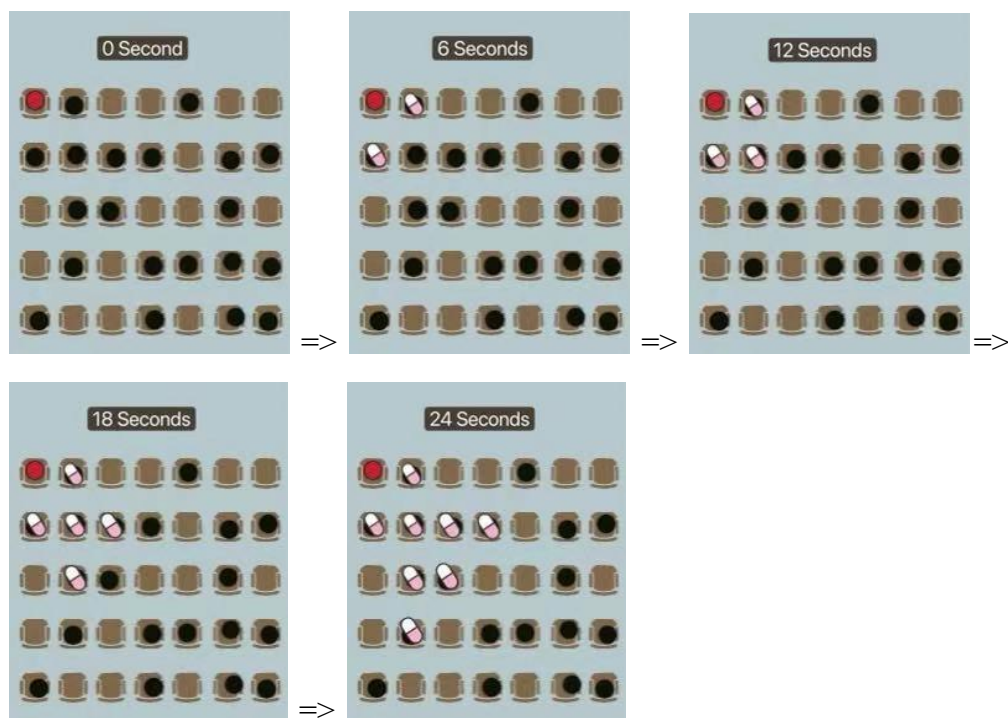
Return -1 if nobody will be infected.

Auditorium Seating Map:



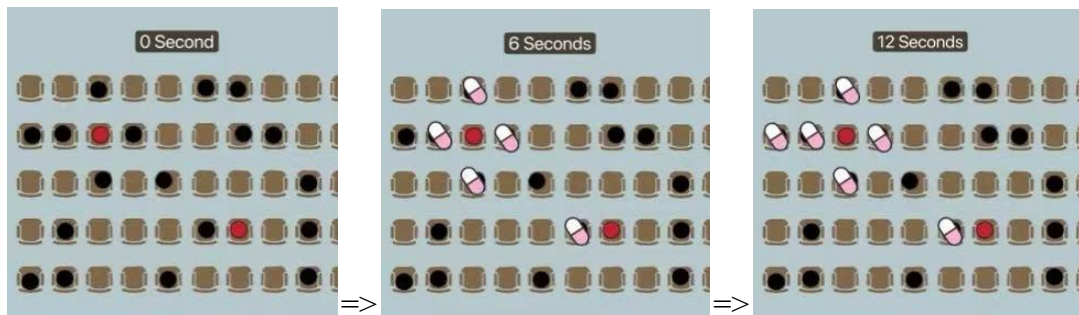
Assume all audiences' body fitness score is 100 in both of the examples.

Example 1:



Output: 24

Example 2:



Output: 12

The outputs of this task are:

- (a) The code of your implementation.
- (b) The screenshots of the results (or part of the results).
- (c) A brief description with screenshots about how you solve this problem.

Question 2 [12%]: Generate a .csv file called **alert_info** that includes three columns which have all the audience's information that are **potentially in danger**. Sort the ExposureTime in a **decreasing** order before you submit.

Output for Example 1: (You need to fill in the Name and id_Number according to the information in the Database.)

	A	B	C
1	Name	id_Number	ExposureTime
2	Aaron Eytcn1	xxxxxxxxxxxxxxxx	24
3	Aaron Eytcn2	xxxxxxxxxxxxxxxx	24
4	Aaron Eytcn3	xxxxxxxxxxxxxxxx	24
5	Aaron Eytcn4	xxxxxxxxxxxxxxxx	18
6	Aaron Eytcn5	xxxxxxxxxxxxxxxx	18
7	Aaron Eytcn6	xxxxxxxxxxxxxxxx	12
8	Aaron Eytcn7	xxxxxxxxxxxxxxxx	6
9	Aaron Eytcn8	xxxxxxxxxxxxxxxx	6

Output for Example 2: (You need to fill in the Name and id_Number according to the information in the Database.)

	A	B	C
1	Name	id_Number	ExposureTime
2	Aaron Eytcn1	xxxxxxxxxxxxxxxx	12
3	Aaron Eytcn2	xxxxxxxxxxxxxxxx	6
4	Aaron Eytcn3	xxxxxxxxxxxxxxxx	6
5	Aaron Eytcn4	xxxxxxxxxxxxxxxx	6
6	Aaron Eytcn5	xxxxxxxxxxxxxxxx	6
7	Aaron Eytcn6	xxxxxxxxxxxxxxxx	6

The outputs of this task are:

- (d) The code of your implementation.
- (e) The screenshots of the results (or part of the results).
- (f) A brief description with screenshots about how you integrated the SQL and C++ together.
- (g) The alert_info.csv file.

4. Report [10%]

Include all the required screenshots and descriptions in the report. Treat your report as a user manual that helps people who will use your system later.

Submission Checklist

- ☐ One **ERD_netID.jpg** file from Task 1
- ☐ Eight **Q1/Q2/Q3/Q4/Q5/Q6.sql** files from Task 2
- ☐ One **audiences_info.csv** file from Task 2 Q4
- ☐ One **city_policy_2.csv** file from Task 2 Q5
- ☐ One **infect_time.cpp** file from Task 3 Q1
- ☐ One **alert_info.cpp** file from Task 3 Q2
- ☐ One **alert_info.csv** file from Task 3 Q2
- ☐ One **Zovid_Report_netID.pdf** file as a report

Name all the required files properly [3%], zip all the files above, and upload this zip file to Sakai.

Due date: Week 7, Sunday, May 9th 11:59PM

Late submission: 10% deduction for each day

You must ensure that you have all the files listed in this checklist before submitting your homework to Sakai. Failure to submit a complete list of files will lead to mark penalties.