1. Завантажте дані:

* Створіть схему pandemic у базі даних за допомогою SQL-команди.

CREATE SCHEMA IF NOT EXISTS pandemic;

* Оберіть її як схему за замовчуванням за допомогою SQL-команди.

use pandemic;

* Імпортуйте [дані](https://drive.google.com/file/d/1lHEXJvu2omYRgvSek6mHq-iQ3RmGAQ7e/view?usp=sharing) за допомогою Import wizard так, як ви вже робили це у темі 3.

[infectious\_cases.csv](https://prod-files-secure.s3.us-west-2.amazonaws.com/89fec302-e51c-45d8-a968-db8bdab456d2/74d1694c-7c06-403a-b8fa-301608033701/infectious_cases.csv)

* Продивіться дані, щоб бути у контексті.

2. Нормалізуйте таблицю infectious\_cases до 3ї нормальної форми. Збережіть у цій же схемі дві таблиці з нормалізованими даними.

Створимо таблицю entities.

CREATE TABLE entities(

id INT PRIMARY KEY auto\_increment,

entity VARCHAR(50) NOT NULL UNIQUE,

code VARCHAR(8) UNIQUE

);

Додамо в цю таблиці рядки з infectious\_cases. Не забуваємо, що в нас в оригінальній таблиці не null, а пусті рядки, тому використовуємо NULLIF.

INSERT INTO entities(entity, code)

SELECT DISTINCT entity, NULLIF(code, '') FROM infectious\_cases;

Скопіюємо дані з оригінальної таблиці в infectious\_cases\_new.

CREATE TABLE infectious\_cases\_new AS

SELECT \* FROM infectious\_cases;

Додамо 2 колонки id та entity\_id у нову таблицю.

ALTER TABLE infectious\_cases\_new

ADD id INT PRIMARY KEY auto\_increment first,

ADD entity\_id INT after id;

Створимо лукап на табличку entities в infectious\_cases\_new.

ALTER TABLE infectious\_cases\_new

ADD constraint cases\_to\_entities foreign key (entity\_id) references entities(id);

Заповнимо entity\_id колоночку. Це можна зробити, бо entities.entity NOT NULL UNIQUE.

UPDATE infectious\_cases\_new i, entities e

SET i.entity\_id = e.id WHERE i.entity = e.entity;

Приберемо зайві тепер колоночки entity та code з нової таблички.

alter table infectious\_cases\_new

drop column entity,

drop column code;

Також мені не подобається, що в нас рядкові колонки кейсів, хоча повинні бути double, тому ми це переробимо (чому одразу не поставила в імпорт візарді? Бо там були не null, а пісті рядки, черерз що візард падав).

ALTER TABLE infectious\_cases\_new

ADD COLUMN Number\_yaws\_double DOUBLE,

ADD COLUMN polio\_cases\_double DOUBLE,

ADD COLUMN cases\_guinea\_worm\_double DOUBLE,

ADD COLUMN Number\_rabies\_double DOUBLE,

ADD COLUMN Number\_malaria\_double DOUBLE,

ADD COLUMN Number\_hiv\_double DOUBLE,

ADD COLUMN Number\_tuberculosis\_double DOUBLE,

ADD COLUMN Number\_smallpox\_double DOUBLE,

ADD COLUMN Number\_cholera\_cases\_double DOUBLE;

UPDATE infectious\_cases\_new

SET Number\_yaws\_double = NULLIF(Number\_yaws, ''),

polio\_cases\_double = NULLIF(polio\_cases, ''),

cases\_guinea\_worm\_double = NULLIF(cases\_guinea\_worm, ''),

Number\_rabies\_double = NULLIF(Number\_rabies, ''),

Number\_malaria\_double = NULLIF(Number\_malaria, ''),

Number\_hiv\_double = NULLIF(Number\_hiv, ''),

Number\_tuberculosis\_double = NULLIF(Number\_tuberculosis, ''),

Number\_smallpox\_double = NULLIF(Number\_smallpox, ''),

Number\_cholera\_cases\_double = NULLIF(Number\_cholera\_cases, '');

ALTER TABLE infectious\_cases\_new

DROP COLUMN Number\_yaws,

DROP COLUMN polio\_cases,

DROP COLUMN cases\_guinea\_worm,

DROP COLUMN Number\_rabies,

DROP COLUMN Number\_malaria,

DROP COLUMN Number\_hiv,

DROP COLUMN Number\_tuberculosis,

DROP COLUMN Number\_smallpox,

DROP COLUMN Number\_cholera\_cases;

ALTER TABLE infectious\_cases\_new

CHANGE COLUMN Number\_yaws\_double Number\_yaws DOUBLE,

CHANGE COLUMN polio\_cases\_double polio\_cases DOUBLE,

CHANGE COLUMN cases\_guinea\_worm\_double cases\_guinea\_worm DOUBLE,

CHANGE COLUMN Number\_rabies\_double Number\_rabies DOUBLE,

CHANGE COLUMN Number\_malaria\_double Number\_malaria DOUBLE,

CHANGE COLUMN Number\_hiv\_double Number\_hiv DOUBLE,

CHANGE COLUMN Number\_tuberculosis\_double Number\_tuberculosis DOUBLE,

CHANGE COLUMN Number\_smallpox\_double Number\_smallpox DOUBLE,

CHANGE COLUMN Number\_cholera\_cases\_double Number\_cholera\_cases DOUBLE;

3. Проаналізуйте дані:

* Для кожної унікальної комбінації Entity та Code або їх id порахуйте середнє, мінімальне, максимальне значення та суму для атрибута Number\_rabies.
* Результат відсортуйте за порахованим середнім значенням у порядку спадання.
* Оберіть тільки 10 рядків для виведення на екран.

SELECT entity\_id,

AVG(Number\_rabies) as avg,

MIN(Number\_rabies) as min,

MAX(Number\_rabies) as max,

SUM(Number\_rabies) as sum

FROM infectious\_cases\_new

WHERE Number\_rabies is not NULL

GROUP BY entity\_id

ORDER BY avg DESC

LIMIT 10;

4. Побудуйте колонку різниці в роках.

Для оригінальної або нормованої таблиці для колонки Year побудуйте з використанням вбудованих SQL-функцій:

* атрибут, що створює дату першого січня відповідного року,
* атрибут, що дорівнює поточній даті,
* атрибут, що дорівнює різниці в роках двох вищезгаданих колонок.

SELECT Year,

MAKEDATE(Year, 1) as first\_day\_of\_year,

CURDATE() AS today,

TIMESTAMPDIFF(YEAR, MAKEDATE(Year, 1), CURDATE()) AS diff

FROM infectious\_cases\_new;

5. Побудуйте власну функцію.

* Створіть і використайте функцію, що будує такий же атрибут, як і в попередньому завданні: функція має приймати на вхід значення року, а повертати різницю в роках між поточною датою та датою, створеною з атрибута року (1996 рік → ‘1996-01-01’).

DROP FUNCTION IF EXISTS year\_difference;

DELIMITER //

CREATE FUNCTION year\_difference(input\_year INT)

RETURNS INT

deterministic

no sql

BEGIN

DECLARE diff\_years INT;

SET diff\_years = TIMESTAMPDIFF(YEAR, MAKEDATE(input\_year, 1), CURDATE());

RETURN diff\_years;

END

//

DELIMITER ;

SELECT Year,

year\_difference(Year) AS diff

FROM infectious\_cases\_new;