Assignment A2

Professor: Jeremy Harper **TAs**: Bhavana Dhonvan, Latha Sree Rayala

Student Name: Deepak Rajput **Program**: MS Applied Data Science

Index Table:

Sr. No.	Assignment Requirement	Page number
1	Import the data into mysql	1-2
2	Draw the ERD for the database	3
3	Using the database design language to list all the tables, relationships, and attributes, including the PK, FK constraints.	4-5
4	Explain your design rationales.	5
5	query to return data on the state, update date, the case incidence rate, and case fatality ratio	6
6	Part 2.a	7-8
7	Part 2.b	9-10

Part 1: See 3 attached files, 01-19-2021.csv, Download 01-19-2021.csv,01-19-2022.csv and Download 01-19-2022.csv and 01-19-2023.csv Download 01-19-2023.csv and design a relational database that can store all three files worth of data in a new table structure and import your data into that structure.

MySQL code for create relational database tables from provided file.
Combine all 3 files and Load given file via table insert wizard, then created below tables.
Insert Entry_no as primary key for second table to secure unique values for all data. Also added year to bifurcate data further chronological.

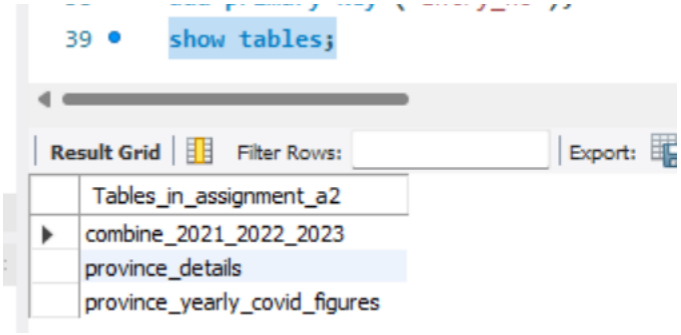
Mysql code for table generation.txt

Code file

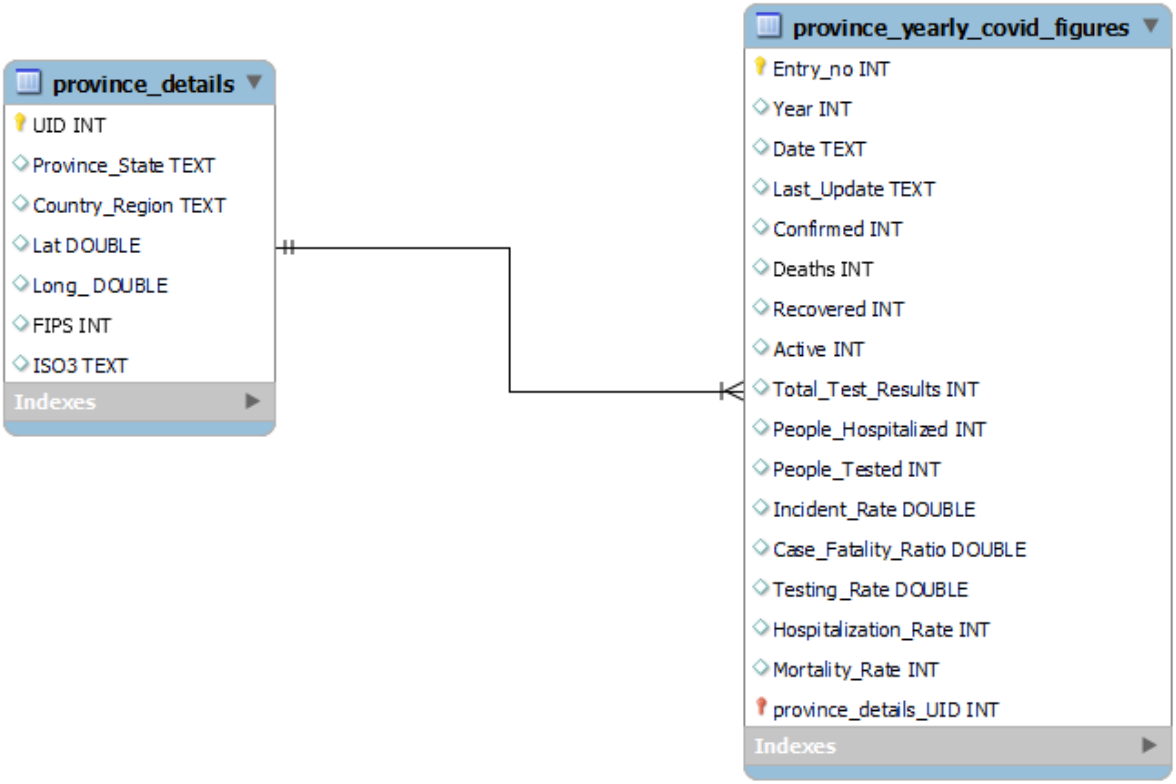
MySql code:

```
create schema assignment_A2;
use assignment_A2;
select * from combine_2021_2022_2023;
create table province_details
select distinct UID,
Province_State,
Country_Region,
Lat,
Long_,
FIPS,
ISO3
from combine_2021_2022_2023;
alter table `assignment_a2`.`province_details`
change column `UID` `UID` INT NOT NULL,
add primary key (`UID`);
```

```
select * from province_details;
create table province_yearly_covid_figures
select Entry_no,
Year,
Date,
Last_Update,
Confirmed,
Deaths,
Recovered,
Active,
Total_Test_Results,
People_Hospitalized,
People_Tested,
Incident_Rate,
Case_Fatality_Ratio,
Testing_Rate,
Hospitalization_Rate,
Mortality_Rate
from combine_2021_2022_2023;
select * from province_yearly_covid_figures;
alter table `assignment_a2`.`province_yearly_covid_figures`
change column `Entry_no` `Entry_no` INT NOT NULL,
add primary key (`Entry_no`);
show tables;
```

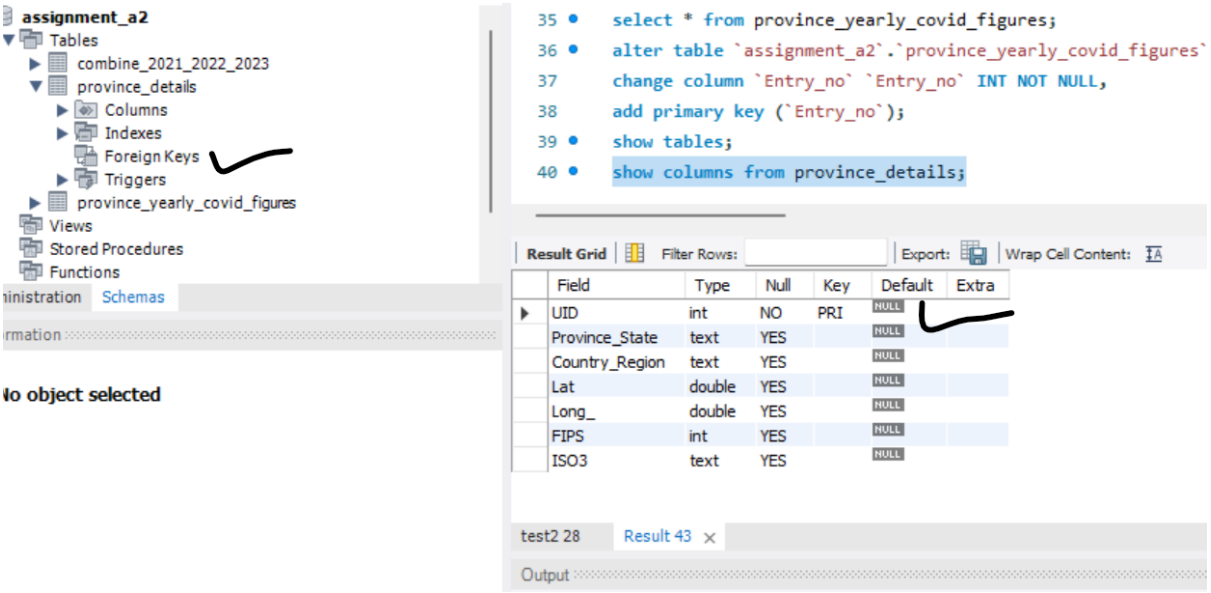


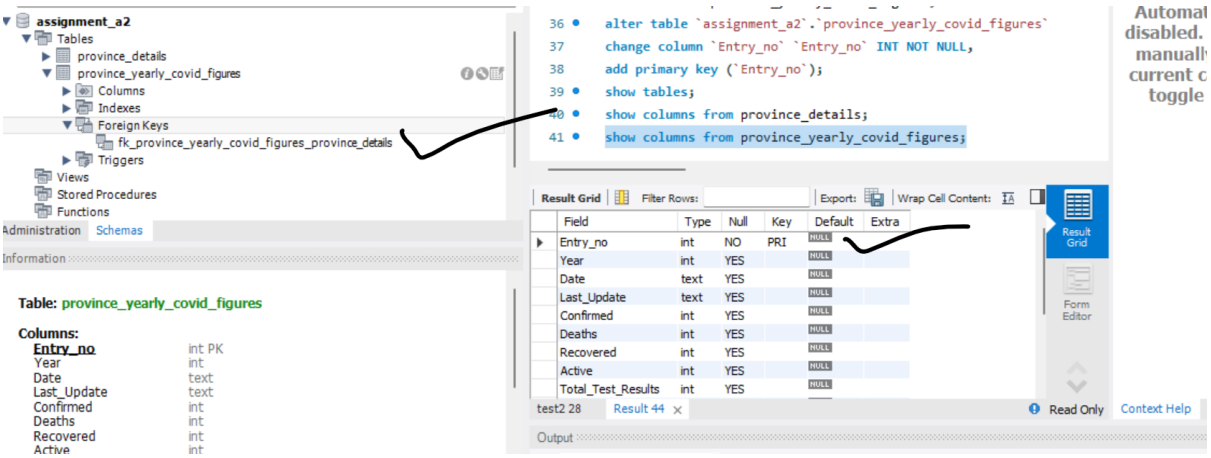
1. Draw an ERD for the database.



2. **Using the database design language to list all the tables, relationships, and attributes, including primary and foreign key constraints.**

   1) **Tables:** Segregated given data in two different tables name as below:
   province_details and province_yearly_covid_figures
   segregation is design according to the static and dynamic nature of attributes which define both entities very well and intact their uniqueness. It also saves storage and avoid duplication of entries in the data.
   Codes:
   show columns from province_details;
   show columns from province_yearly_covid_figures;





   2) **Relationship for the database designs given below**
   One to many identifying
   province_details and province_yearly_covid_figures

3) **Attributes including PK, FK constraints are given in below excel diagram Please find all required information from below diagram**

Entity name : province_details		
Attributes	UID	PK
	Province_State	
	Country_Region	
	Lat	
	Long_	
	FIPS	
	ISO3	

One to Many

Entity name : province_yearly_covid_figures		
Attributes	Entry_no	PK
	Year	
	Last_Update	
	Confirmed	
	Deaths	
	Recovered	
	Active	
	Total_Test_Results	
	People_Hospitalized	
	People_Tested	
	Incident_Rate	
	Case_Fatality_Ratio	
	Testing_Rate	
	Hospitalization_Rate	
	Mortality_Rate	
	province_details_UID	FK

3. **Explain your design rationales.**

Student response:

Segregated given database in 2 different Tables and Entities province_details and province_yearly_covid_figures to have a clear understanding and relationship between all these entities as per their static and dynamic nature.

All selected attributes well defined the entity and maintains uniqueness among all entities.

Role of primary key as UID and Entry no are selected for their uniqueness and not null property.

Foreign elements are defined in table mentioned in ERD diagram and table given above provide required relationship to overlook all require data among tables & database.

By designing the same we can store, analysis, retrieve and represent data efficiently and save storage spare.

4. Write a query to return data on the state, update date, the case incidence rate, and case fatality ratio. Group by each state and filter to the following states Indiana, Illinois, Michigan, & Ohio

Query code:

```
select Province_State,
Last_Update,
Incident_Rate,
Case_Fatality_Ratio
from combine_2021_2022_2023
where Province_State = 'Indiana'
or Province_State = 'Illinois'
or Province_State = 'Michigan'
or Province_State = 'Ohio';
order by Province_State;
```

mysql assignment 2 query.csv

Output:

```
49      or Province_State = 'Michigan'
50      or Province_State = 'Ohio'
51      order by Province_State;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Province_State	Last_Update	Incident_Rate	Case_Fatality_Ratio
Illinois	20-01-2021 05:30	8495.479853	1.872029814
Illinois	20-01-2022 04:31	21381.88347	1.178937314
Illinois	20-01-2023 04:31	31635.89511	1.022239828
Indiana	20-01-2021 05:30	8844.572644	1.800361416
Indiana	20-01-2022 04:31	22139.74323	1.375381081
Indiana	20-01-2023 04:31	30018.89867	1.276436981
Michigan	20-01-2021 05:30	5874.450791	2.505995674
Michigan	20-01-2022 04:31	20857.24267	1.501452484
Michigan	20-01-2023 04:31	30219.19709	1.364668974
Ohio	20-01-2021 05:30	7152.381278	1.938761963
Ohio	20-01-2022 04:31	20867.3465	1.418043994
Ohio	20-01-2023 04:31	28570.30909	1.235143484

Result Grid | Form Editor | Field Types | Query Stats

Part 2: Attached is what is known as a time series data source. (Note all the date specific columns, you do not need to import these files to successfully complete this assignment)

Part 2.a>> We segregated table data in three different relational table such as

Province_details>> Contains details of province_id, Province_State, Country_Region, iso2, iso3, code3 here we added province_id like serial number to give unique id for primary key definition.

Province_region>> contains details of various region of Combined_Key, UID, FIPS, Admin2, lat, long_, Population. Here choosed Combined_Key as primary key as it will always remain unique to data entry

Datewise_death_columns>>>contains details of death count in date wise details

Add Sr_no columns to give unique primary key identifier for all entries.

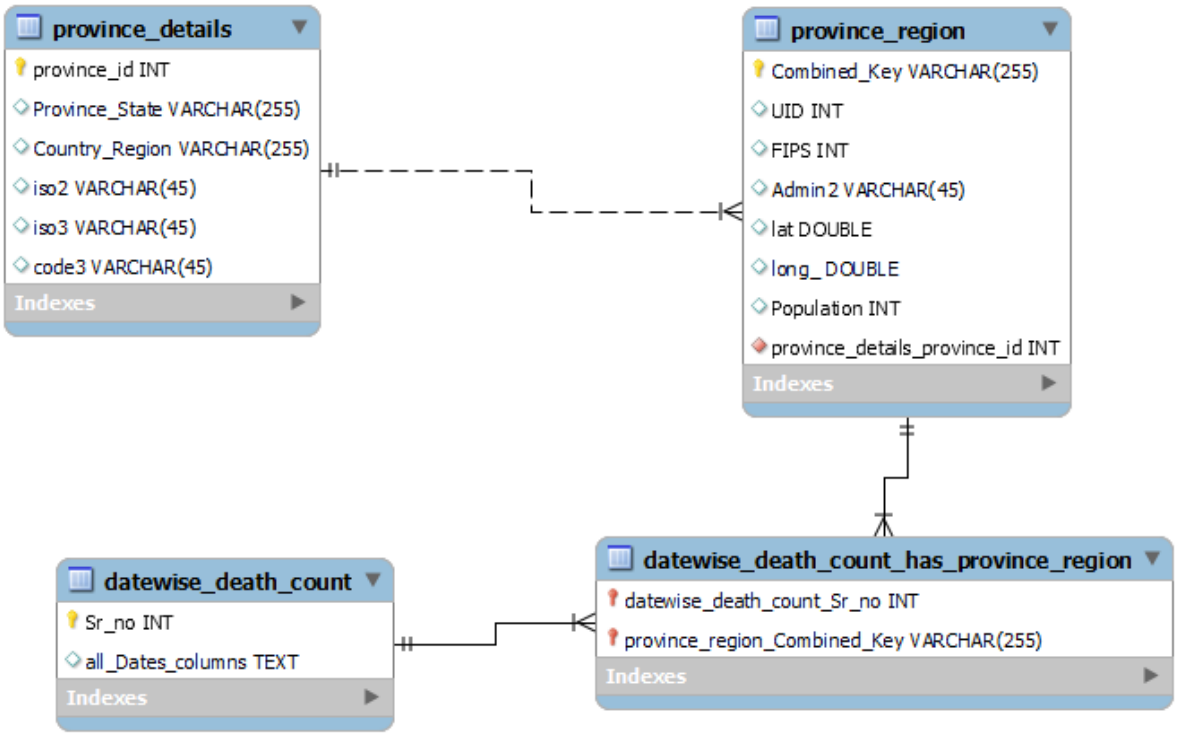
All dates Columns are large in count so specific as columns.

Code:

```
create schema assign2_part2;

use assign2_part2;

create table Province_details(

province_id int,

Province_State varchar(255),

Country_Region varchar(255),

iso2 varchar(45),

iso3 varchar(45),

code3 varchar(45));

alter table `assign2_part2`.`province_details`

change column `province_id` `province_id` int not null,

add primary key (`province_id`);

create table Province_region(

Combined_Key varchar(255),

UID int,

FIPS int,

Admin2 varchar(45),
```

```
lat double,
long_ double,
Population int);
alter table `assign2_part2`.`province_region`
change column `Combined_Key` `Combined_Key` varchar(255) not null,
add primary key (`Combined_Key`);
create table Datewise_death_count(
Sr_no int,
all_Dates_columns text);
alter table `assign2_part2`.`Datewise_death_count`
change column `Sr_no` `Sr_no` int not null,
add primary key (`Sr_no`);
```

ERD diagram:

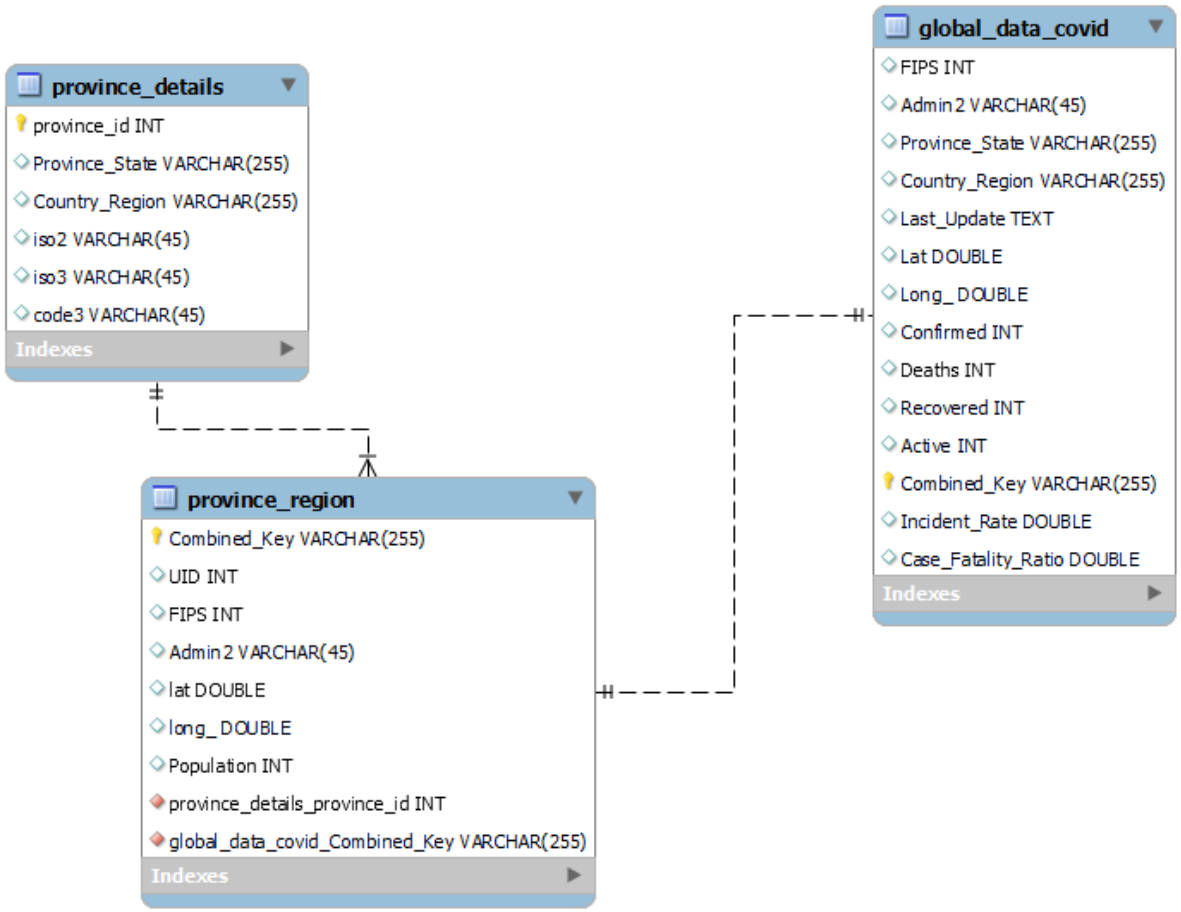


Part 2.b Link the following file into your new schema and show how it is related to what you built in 2.a

Relationship will be one to one between global_data_covid and part 2.a province_region table

Code:

```
create table Global_data_Covid(
FIPS int,
Admin2 varchar(45),
Province_State varchar(255),
Country_Region varchar(255),
Last_Update text,
Lat     double,
Long_ double,
Confirmed int,
Deaths int,
Recovered int,
Active int,
Combined_Key varchar(255),
Incident_Rate double,
Case_Fatality_Ratio double);
```


province_details
province_id INT
Province_State VARCHAR(255)
Country_Region VARCHAR(255)
iso2 VARCHAR(45)
iso3 VARCHAR(45)
code3 VARCHAR(45)
Indexes
global_data_covid
FIPS INT
Admin2 VARCHAR(45)
Province_State VARCHAR(255)
Country_Region VARCHAR(255)
Last_Update TEXT
Lat DOUBLE
Long_ DOUBLE
Confirmed INT
Deaths INT
Recovered INT
Active INT
Combined_Key VARCHAR(255)
Incident_Rate DOUBLE
Case_Fatality_Ratio DOUBLE
Indexes
province_region
Combined_Key VARCHAR(255)
UID INT
FIPS INT
Admin2 VARCHAR(45)
lat DOUBLE
long_ DOUBLE
Population INT
province_details_province_id INT
global_data_covid_Combined_Key VARCHAR(255)
Indexes