E-R Diagram:-

Assumptions:-

I am connecting my entire database design with locations iso_code.

I have date, state_name, vaccine_name, age_group as primary keys in my entire design.

us_state table holds US State names

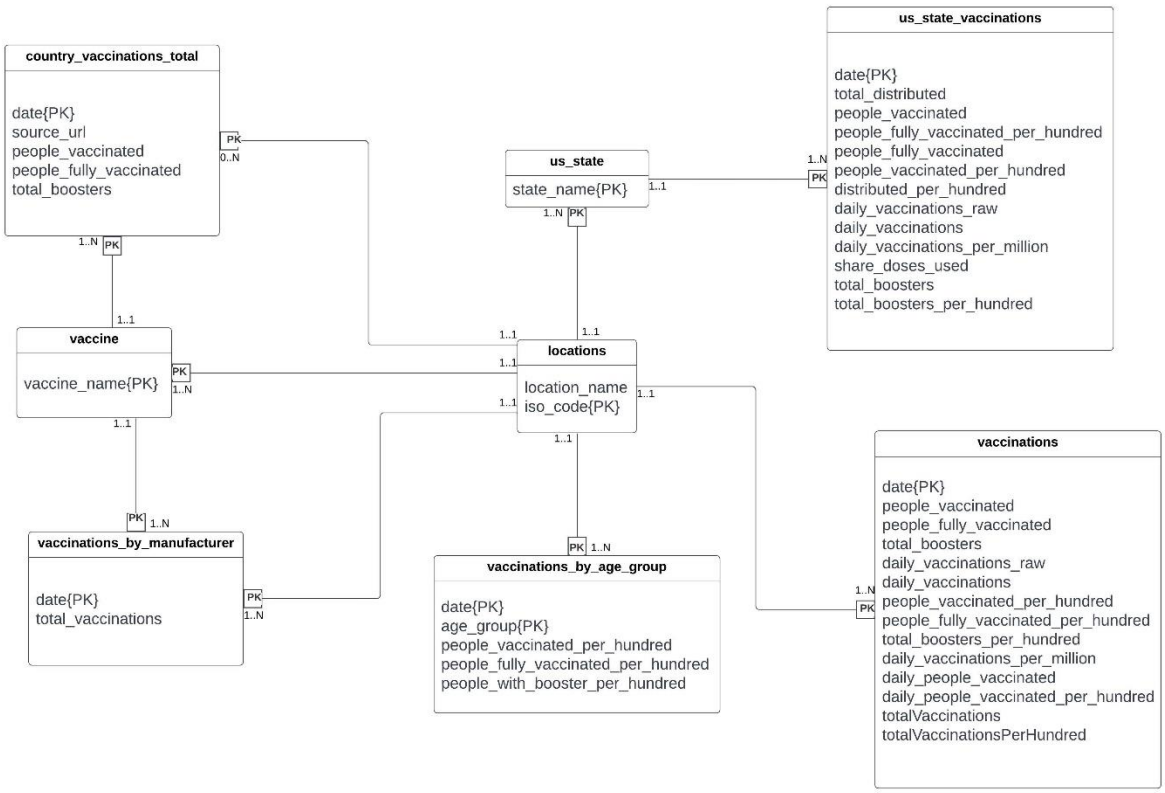
country_vaccinations_tool holds all country vaccination details

vaccine contains various vaccine name

vaccintions_by_age_group contains information about people with various age_group

us_state_vaccinations contain us_state people data.

vaccinations holds various people's data



Relational Schema:-

Step 1:- Strong Entities

locations(location_name, iso_code)

Step 2:- Weak Entities

~~**vaccinations_by_age_group** (date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)~~

~~**vaccinations_by_manufacturer** (date, vaccine, total_vaccinations)~~

~~**vaccinations** (date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred)~~

~~**us_state_vaccinations** (date, total_distributed, total_vaccinations, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_fully_vaccinated, people_vaccinated_per_hundred, distributed_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, share_doses_used, total_boosters, total_boosters_per_hundred)~~

~~**country_vaccinations_total** (date, total_vaccinations, source_url, people_vaccinated, people_fully_vaccinated, total_boosters)~~

~~**us_state** (state_name)~~

~~**vaccine**(vaccine_name)~~

Step 3:- 1 - 1 Relationship

vaccinations_by_age_group (iso_code*, age_group, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

vaccinations_by_manufacturer (iso_code*, vaccine_name*, date, total_vaccinations)

~~**vaccinations** (iso_code*, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred)~~

us_state_vaccinations (state_name*, iso_code*, date, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, people_fully_vaccinated, people_vaccinated_per_hundred, distributed_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, share_doses_used, total_boosters, total_boosters_per_hundred)

~~**country_vaccinations_total** (vaccine_name*, total_vaccinations, date, source_url, people_vaccinated, people_fully_vaccinated, total_boosters)~~

us_state (state_name, iso_code*)

vaccine (vaccine_name, iso_code*)

Step 4:- 1-N Relationship:-

country_vaccinations_total (iso_code*, vaccine_name*, source_url, date, people_vaccinated, people_fully_vaccinated, total_boosters)

vaccinations_by_age_group – Same as 1 - 1 Relationship

vaccinations (iso_code*, date, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, totalVaccinations, totalVaccinationsPerHundred)

Step 5:- N-N Relationship:-

None

Step 6 :- Multivalued Attributes

None

Step 7:- Higher-Order Relationship

None

Final Schema:-

locations (iso_code, location_name)

vaccinations_by_age_group (iso_code*, age_group, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

vaccinations_by_manufacturer (iso_code*, vaccine_name*, date, total_vaccinations)

us_state_vaccinations (state_name*, iso_code*, date, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, people_fully_vaccinated, people_vaccinated_per_hundred, distributed_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, share_doses_used, total_boosters, total_boosters_per_hundred)

us_state (state_name, iso_code*)

country_vaccinations_total (iso_code*, vaccine_name*, source_url, date, source_url, people_vaccinated, people_fully_vaccinated, total_boosters)

vaccinations (iso_code*, date, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, totalVaccinations, totalVaccinationsPerHundred)

vaccine (vaccine_name, iso_code*)

Normalisation:

locations (iso_code, location_name)

Functional Dependencies

iso_code -> location_name

vaccine (vaccine_name, iso_code*)

Functional Dependencies

None

us_state (state_name, iso_code*)

Functional Dependencies

None

vaccinations_by_age_group (iso_code*, age_group, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

Functional Dependencies

iso_code, age_group, date -> people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred

vaccinations (iso_code*, date, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, totalVaccinations, totalVaccinationsPerHundred)

Functional Dependencies

iso_code, date -> people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, totalVaccinations, totalVaccinationsPerHundred)

vaccinations_by_manufacturer (iso_code*, vaccine_name*, date, total_vaccinations)

Functional Dependencies

iso_code, vaccine_name, date -> total_vaccinations

country_vaccinations_total (iso_code*, vaccine_name*, source_url, date, people_vaccinated, people_fully_vaccinated, total_boosters)

Functional Dependencies

vaccine_name, iso_code, date -> source_url, people_vaccinated, people_fully_vaccinated, total_boosters

us_state_vaccinations (state_name*, date, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations_per_hundred, people_fully_vaccinated,

people_vaccinated_per_hundred, distributed_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, share_doses_used, total_boosters, total_boosters_per_hundred)

In us_state_vaccinations table, there is no common attribute to link to entire database, I include iso_code to this table.

us_state_vaccinations (state_name*, iso_code*, date, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, people_fully_vaccinated, people_vaccinated_per_hundred, distributed_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, share_doses_used, total_boosters, total_boosters_per_hundred)

Functional Dependencies

state_name, iso_code, date -> total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, people_fully_vaccinated, people_vaccinated_per_hundred, distributed_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, share_doses_used, total_boosters, total_boosters_per_hundred

Final Schema:-

locations (iso_code, location_name)

vaccine (vaccine_name, iso_code*)

us_state (state_name, iso_code*)

vaccinations_by_age_group (iso_code*, age_group, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

vaccinations (iso_code*, date, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, totalVaccinations, totalVaccinationsPerHundred)

vaccinations_by_manufacturer (iso_code*, vaccine_name*, date, total_vaccinations)

country_vaccinations_total (iso_code*, vaccine_name*, source_url, date, people_vaccinated, people_fully_vaccinated, total_boosters)

us_state_vaccinations (state_name*, iso_code*, date, total_distributed, people_vaccinated, people_fully_vaccinated_per_hundred, people_fully_vaccinated, people_vaccinated_per_hundred, distributed_per_hundred, daily_vaccinations_raw, daily_vaccinations, daily_vaccinations_per_million, share_doses_used, total_boosters, total_boosters_per_hundred)

References:

1.‘SQL Server COALESCE() Function’
<https://www.w3schools.com/sql/func_sqlserver_coalesce.asp>.

2.‘Data on COVID-19 (coronavirus) vaccinations by Our World in Data’
<https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/README.md>

3.‘covid-19-data/public/data/vaccinations at master · owid/covid-19-data’ *GitHub*, <https://github.com/owid/covid-19-data>.

4.‘owid/covid-19-data’,

<https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations>