Part A: Understanding Data

The most hopeful alternative for putting a stop to the epidemic is the efficient use of COVID-19 vaccinations. We receive a vaccination dataset. This regularly updated dataset contains information on the total number of vaccinations given, the first, second, and booster doses given, daily vaccination rates. It will be maintained throughout the ongoing global immunization campaign. The analysis of the current and anticipated vaccination deployment rates is made easier by this tool for policymakers and researchers. interactions with anti-vaccination measures, the potential impact of vaccinations on pandemic outcomes like disease, mortality, and transmission, and unequal access to vaccines globally.

In this project we are using following dataset:

1. vaccinations.csv – detail distribution of number of vaccine doses taken by people with respect to location and date. It has 235 unique countries with data recorded between February 2021 to October 2022
2. location.csv – vaccines available in a particular country and the source from where this information is retrieved. It has 223 unique countries. It also has the last date when the vaccine was administered
3. vaccinations-by-age-group.csv - number of vaccine doses taken by different age groups with respect to location and date. It has data recorded between January2020 to October 2022
4. country_data - number of vaccine doses taken by people, the source, and the vaccines for four countries, Australia, Germany, Italy, and United States. Data recorded between December 2020 to October 2022
5. us_state_vaccinations.csv - detail distribution of number of vaccine doses taken by people with respect to US states and date. It has 65 unique states with data recorded between January 2021 to October 2022
6. vaccinations-by-manufacturer.csv – it has data of 44 countries on the typed of vaccines produced by these countries and the quantity produced on each day
7. available-vaccine.csv – has list of all unique vaccine names

Part B: Designing the Database:

1. Entity-Relationship (ER) Diagram: -

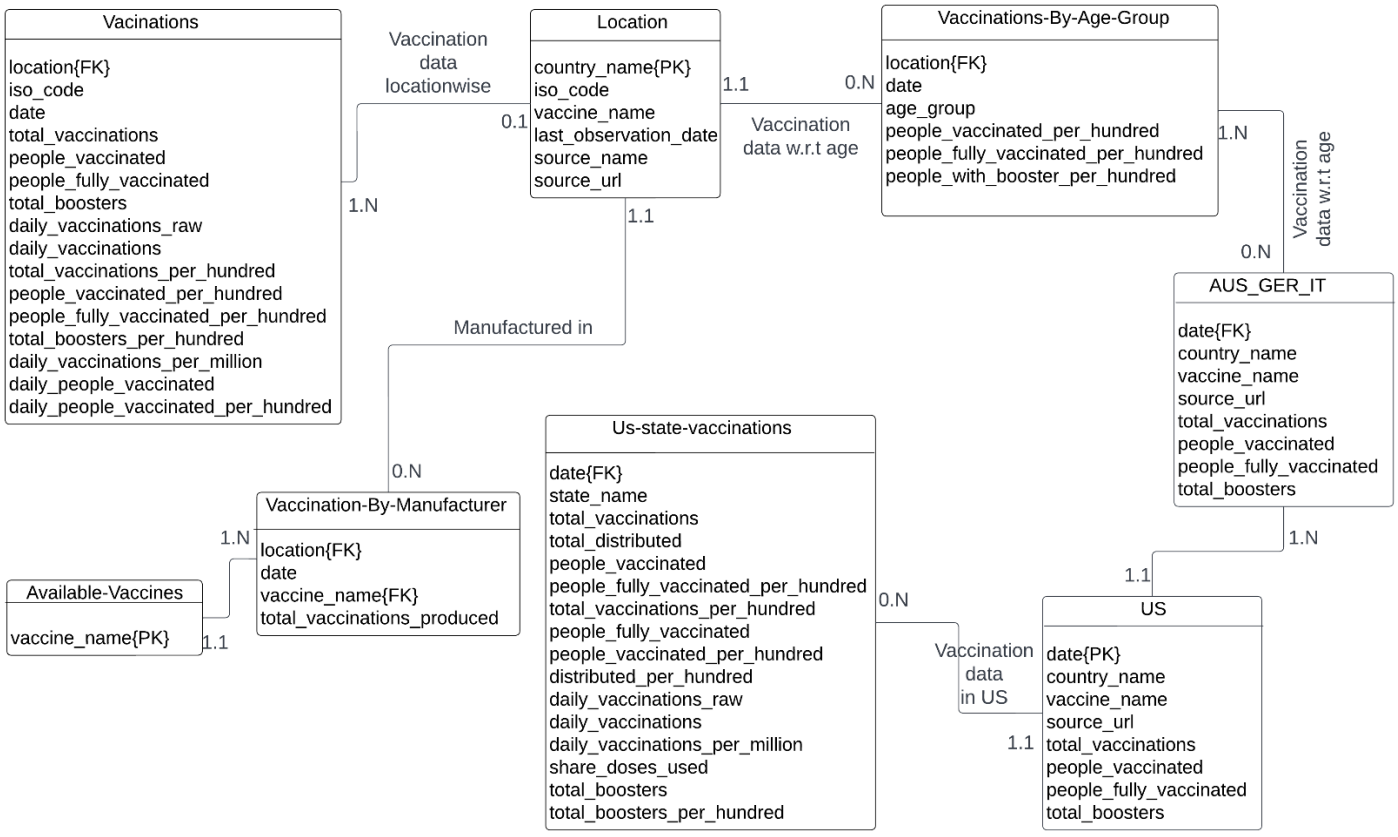

Fig 1.1 *ER-Diagram*

Assumptions:

- Between 2 entities, if the number of locations in one entity is greater than the other, then it is assumed that all locations of smaller entity are present in the higher entity.

For example; *Location* entity has 223 unique countries and *Vaccinations* entity has 235. Then all 223 countries of entity *Location* are present in entity *Vaccinations*.

- In entity Available-Vaccines, it is assumed that it contains all vaccine name that has ever manufactured and are unique.

2. Normalization Forms: -

→ **1NF** – (First Normal Form)
Because all of the values are atomic and no columns are left hanging or unrelated, every table is in the first normal form. Therefore, the First Normal form requirements are satisfied by our original database architecture.

→ **2NF** – (Second Normal Form)
There is a second normal form for each table, due to the fact that none of the properties functionally depend on any one component of the composite primary key. The Second Normal Form restrictions are satisfied by our initial database architecture.

→ **3NF** – (Third Normal Form)
Because no non-primary key attribute is functionally dependent on any other non-primary key attribute, every table is in the third normal form. If we included the population of each country as a property, some of the variables, such as total vaccinations per hundred, may be seen as functionally dependent on other attributes, such as total vaccinations.
As a result, they satisfy the third normal form and are independent in their current form.

3. Database Schema: -

Location(country_name, iso_code, vaccine_name, last_observation_date, source_name, source_url)

Vaccinations(location*,date, iso_code, 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)

Vaccinations-By-Age-Group(location*,date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

US(date, country_name, vaccine_name, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters)

AUS_GER_IT(date*,country_name, vaccine_name, source_url, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters)

US-state-vaccinations(date*,state_name ,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)

Available-Vaccines(vaccine_name)

Vaccination-By-Manufacturer(location*,date, vaccine_name*, total_vaccinations)