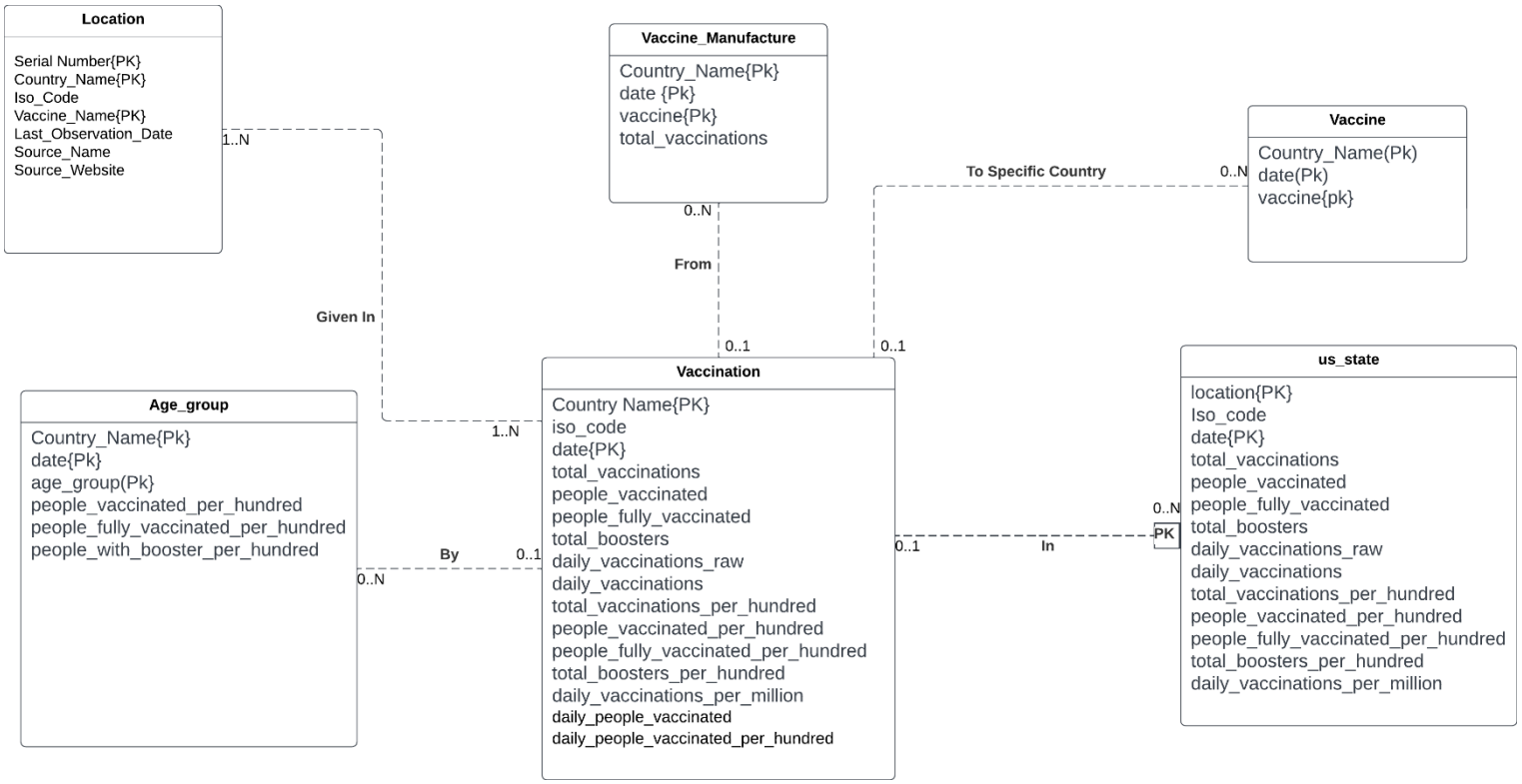
Modelling The Database

ER DIAGRAM: -



Assumptions: -

1. Vaccination Table – Vaccine_Manufacture
    a. Vaccination table contains the number of vaccinations happening in a particular country on a Particular day. Where as "Vaccine_Manufacture" contains the information about what all vaccines were used on the particular day in the particular country. So, the sum of different vaccines which were administered in the particular country on a particular day will be equal to the number of Vaccination in the "Vaccination Table". Hence the relation among these table is 1: N.
2. Vaccination Table – Vaccine
    a. Vaccine Table contains the merged information of the data from the countries like Australia, New Zealand, England, USA.
    b. Vaccine Table provide the name of the vaccines which were used on a particular day in the country.
    c. As for a particular day there may be more then one vaccine used in the country hence the relation between them is 1: N
3. Vaccination Table – US state Table
    a. Although US_State table has its own primary key which makes every row unique but it has no direct relation to the database hence making it a weak entity.
    b. The weak entity borrows the primary key from the strong entity to give it a relation and in this case the strong entity is "Vaccination" Table.
    c. As for a particular day in US the number of vaccines administered can be taken out from the "Vaccination" table but to see how these Vaccinations are divided into different state we can use the "US_State" table.
    d. Hence number of vaccinations on a particular day will be equal to the sum of vaccinations in different state of the USA on the particular day.
    e. Hence the relation between them is 1: N.

4. Vaccination Table – Age Group Table
    a. "Age Group" table contains bi furcation of the vaccination which are given to different age group on the particular day.
    b. So, the number of vaccinations can be derived from the vaccination table.
5. Location Table - Vaccination Table
    a. The relation between these table is N: M.
    b. The Vaccination Table provide the vaccination on a particular day where as the Location table gives the information about the name of the vaccine used in the particular country.

Normalisation: -

1. Normalisation and Resulting Changes in the dataset.
    a. The location dataset contained name of the different vaccine in the single column separated by ",". So, the values were separated to different column using the excel power query function.
    b. The country data of Australia, New Zealand, England and USA contained name of the vaccine separated by ",". So, the names were separated to different columns using Excel function.
    c. The data from the different countries namely Aus, Nz, Eng and Usa were loaded to the single table "Vaccine" as the they all had the same column name. Hence, carrying same information from different countries.
    d. Some of the columns in the "Vaccine" table were redundant like Total_Vaccination, People_Vaccinated, etc. as the same information for the country can be taken out from the vaccination table.

Database Schema: -

1. Strong Entity
    a. Location (serial_number, country_name, Iso_Code, Vaccine_Name, Last_Observation_Date, Source_Name, Source_Website)
    b. Vaccine_Manufacture (Country_Name, Date, Vaccine, Total_Vaccinations)
    c. Vaccine (Country_Name, Date, Vaccine)
    d. Age_Group(Country_Name, Date, Age_Group, People_Vaccinated_Per_Hundred, People_Fully_Vaccinated_Per_Hundred, People_With_Booster_Per_Hundred)
    e. Vaccination (Country_Name, 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)
2. Weak Entity
    a. Us_State(Location, 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,Vaccination.Date*)
3. 1: N Relation
    a. Vaccine (Country_Name, Date, Vaccine, Vaccination.Country_Name*,Vaccination.Date*)
    b. Vaccine_Manufacture (Country_Name, Date, Vaccine, Total_Vaccinations,Vaccination.Country_Name*,Vaccination.Date*)
    c. Age_Group(Country_Name, Date, Age_Group, People_Vaccinated_Per_Hundred, People_Fully_Vaccinated_Per_Hundred, People_With_Booster_Per_Hundred, Vaccination.Country_Name*,Vaccination.Date*)
4. N:M
    a. Given_in(Vaccination.Country_Name*,Location.Vaccine_Name*)

References: -

1. GITHUB *Location* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/locations.csv.
2. GITHUB *vaccinations-by-manufacturer* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/vaccinations-by-manufacturer.csv.
3. GITHUB *vaccinations* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/vaccinations.csv
4. GITHUB *vaccinations-by-age-group* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/vaccinations-by-age-group.csv
5. GITHUB *us_state_vaccinations* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/us_state_vaccinations.csv.
6. GITHUB *Australia* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country_data/Australia.csv
7. GITHUB *England* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country_data/England.csv
8. GITHUB *New Zealand* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country_data/New%20Zealand.csv
9. GITHUB *United States* [Data Set], GitHub, accessed 6/1/2023. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country_data/United%20States.csv