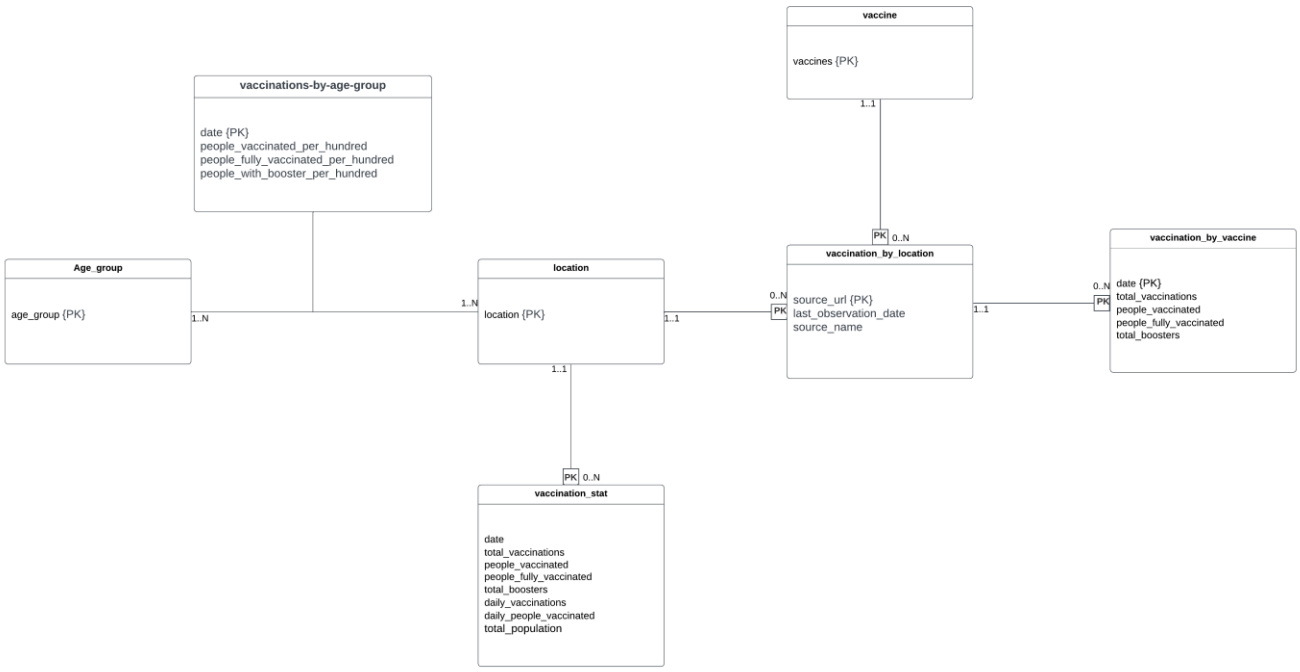
ER Diagram



Normalization Challenges and Changes:

Normalization is the process of organizing data in a database to reduce data redundancy and improve data integrity. Here are some normalization challenges you might encounter:

- Vaccine Data: The same vaccine name can appear multiple times for a single country in the dataset. This could lead to data redundancy. To normalize this, created a "Vaccine" table and link it to the "vaccination_stat" table through foreign keys. Same method applied for location data and age group data. Separate tables are created and linked this information to other required table using foreign key.
- Vaccination Statistic Data: The "Vaccination Event" table contains both daily and per-hundred metrics. Normalized this by removing data that can be derive from the raw data.

Database schema

- location (location)
- vaccine (vaccines)
- age_group (age_group)
- vaccination_stat (location*, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations, daily_people_vaccinated, total_population)
- vaccinations-by-age-group (location*, age_group*, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)
- vaccination_by_location (location*, vaccine*, source_url, last_observation_date, source_name)
- vaccination_by_vaccine (location*, vaccine*, source_url*, date, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters)