

CSC2141

Project Part 1

Finding Data and Planning
Your Database

Name: Duren Gouda

B00949586

Date: February 7, 2025

Section 1:

- a) The dataset consists of COVID-19-related data, including the total number of cases and deaths across different countries and continents. It also includes information on income classifications of countries, vaccine development, and the number of survivors who received specific vaccines. Each entity—such as **continents, countries, income classes, vaccines, and survivors**—is uniquely identified using primary keys. This structured dataset allows for effective tracking of how the pandemic impacted various regions and economic groups worldwide.
- b) This dataset is interesting because it provides insights into the global impact of the COVID-19 pandemic, allowing us to analyze trends based on geography, economic status, and vaccination efforts. By examining variations in case numbers, deaths, and vaccine distributions across different income classes, we can understand how healthcare resources and economic conditions influenced pandemic outcomes. Additionally, tracking survivors by vaccine type helps assess the effectiveness and reach of different vaccination programs worldwide.
- c) The database will enable us to answer key analytical questions, such as:
 - a. Which continent had the highest number of total deaths due to COVID-19?
 - b. How did the number of deaths vary among different income classes?
 - c. Which country developed the most widely used COVID-19 vaccine?

Section 2:

a) Description of the Dataset

The proposed database contains structured information about the global impact of the COVID-19 pandemic, categorized by continents, countries, income classifications, vaccines, and survivors. The database is designed to store details such as total cases, total deaths, vaccination efforts, and economic classifications of countries. Each table is connected using **primary and foreign keys**, allowing efficient querying and relational data analysis.

The database consists of **five main tables**:

1. **Continent** (1 Entity, 3 Attributes): Stores continent-level data, including continent ID, name, and total deaths.
2. **Country** (1 Entity, 3 Attributes): Contains country-level details, such as country ID, continent ID (FK), total cases, total deaths, first and last recorded cases, and income class ID (FK).
3. **Income Class** (1 Entity, 3 Attributes): Classifies countries into different income groups using income class ID, name (e.g., “Low”, “High”), and total deaths per class.
4. **Vaccine** (1 Entity, 3 Attributes): Stores vaccine-related data, including vaccine ID and name.
5. **Survivors** (1 Entity, 3 Attributes): Tracks the number of people who received specific vaccines, linking vaccine ID (FK) and country ID (FK) with the number of survivors.

The **logical relationships** between these tables are:

- a) **One-to-Many (1:M) between Continent and Country** (each continent has multiple countries).
- b) **Many-to-One (M:1) between Country and Income Class** (each country belongs to one income class).
- c) **One-to-Many (M:M) between Country and Vaccine** (a country can make many vaccines).

- d) **Many-to-Many (M:M) between Vaccine and Survivors** (many vaccines can be taken by many survivors, vice-versa)

b) Dataset generation:

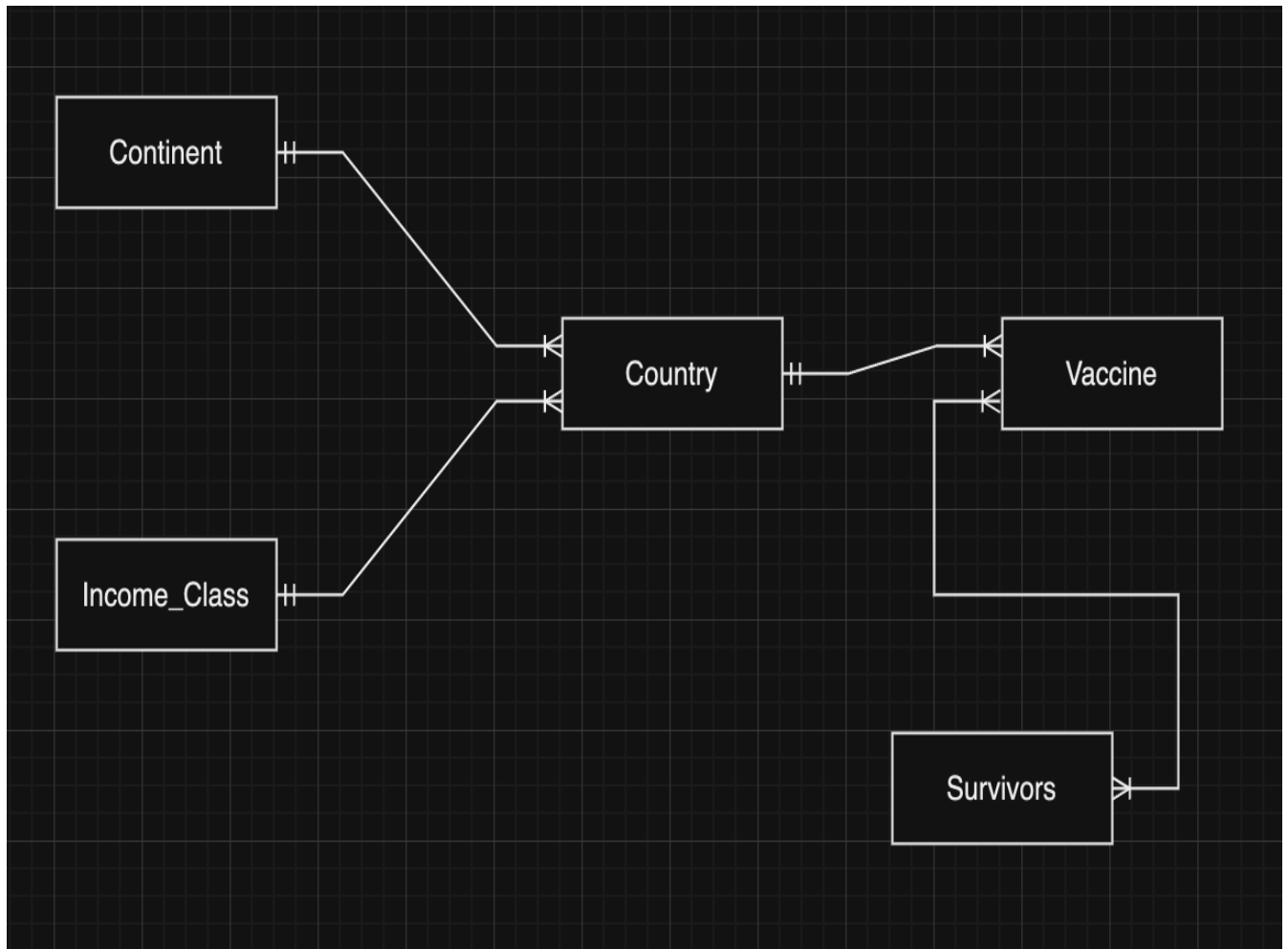
Dataset taken from:

URL: <https://www.kaggle.com/datasets/imdevskp/corona-virus-report>

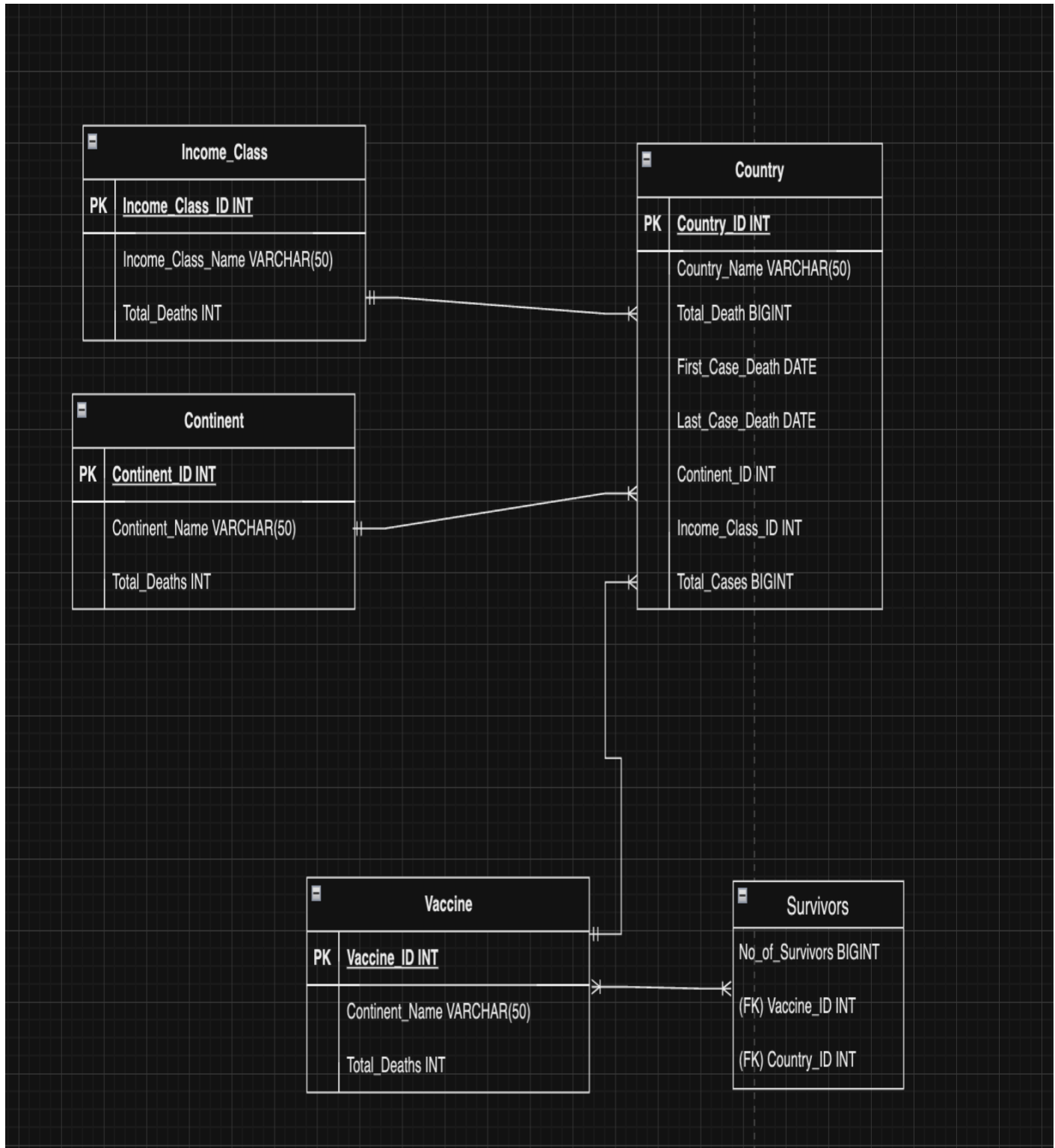
- Note: Some entities and attributes, I have made by my own, as per the guidelines, no GenAI used.
- Entities like Country, Deaths, Confirmed used from the above dataset.

Section 3:

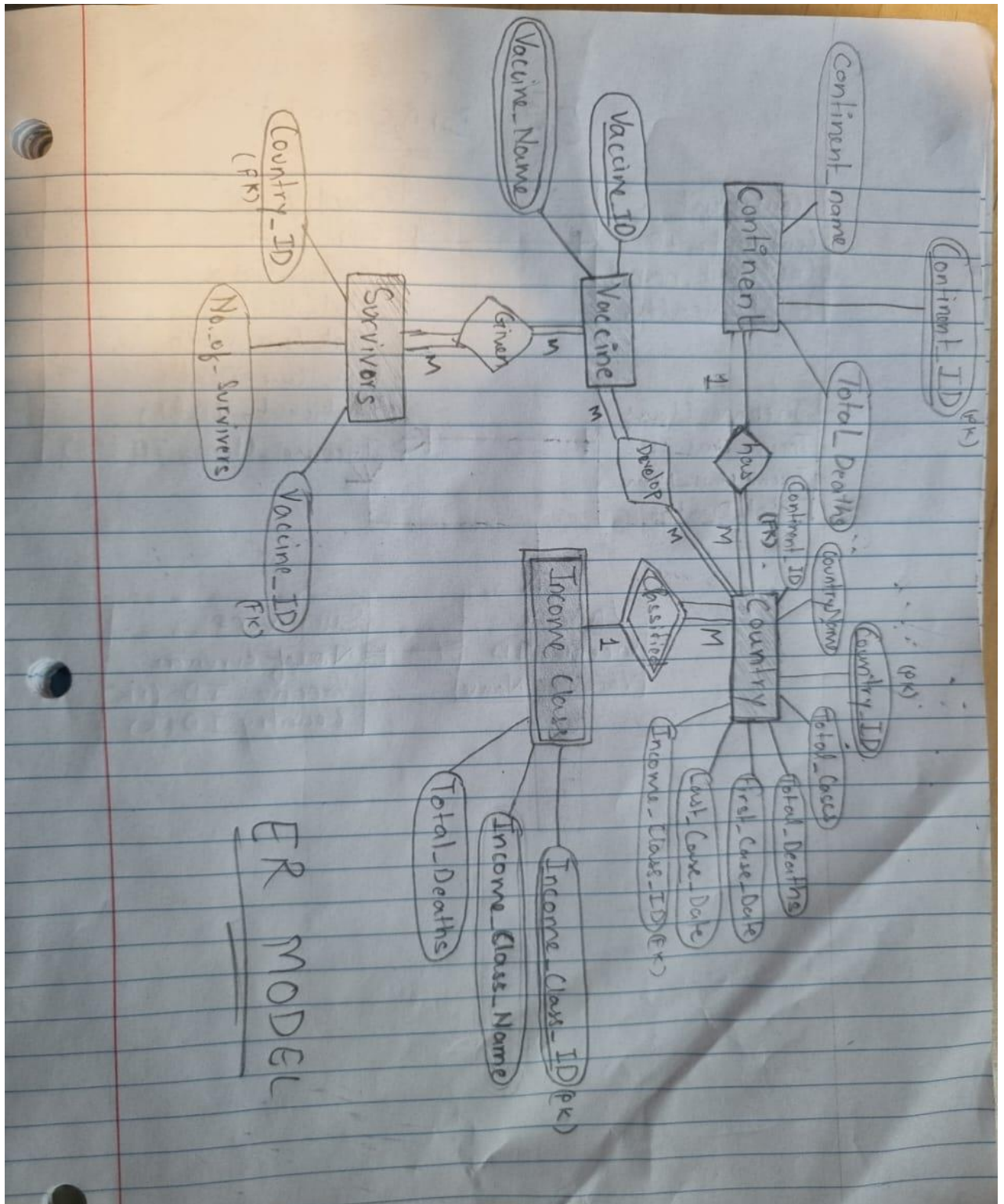
a) Conceptual Model



b) Internal Model



- **NOTE:** I have also made ER_Model and ER_Diagram in pen and paper, and I am attaching those as well. For the ER_Model I have made it UML based.



ER DIAGRAM

