Database for FHI360 health facility profile

The Excel file contains three sheets: "Woredas," "PHCU," and "HPs." To guide you in designing the database, let me provide a brief overview of what each sheet contains and how best to structure them into a PostgreSQL database.

Let me first inspect the data from these sheets to give you a more informed recommendation. ​​

The three sheets contain the following data:

1. **Woredas**:
   * This sheet seems to represent regional administrative units (Woredas), including population and demographic data such as pregnant women, lactating women, children, and adolescents.
   * Key fields: FtFWoredaID, Cohort, Region, Zone, Woreda, COPYear, and various demographic statistics.
2. **PHCU (Primary Health Care Units)**:
   * This sheet contains health centers linked to Woredas, including population and demographic statistics for each health facility.
   * Key fields: FtFWoredaID, HFName (Health Facility), COPYear, and demographic statistics similar to the Woredas sheet.
3. **HPs (Health Posts)**:
   * Health posts are linked to health centers and contain detailed population and demographic data.
   * Key fields: HFGUID, HPGUID, HFName (Health Facility), HPName (Health Post), COPYear, and demographic statistics.

**Database Design Recommendation (Relational Model using PostgreSQL)**

Given the hierarchical nature of the data (Woredas, Health Facilities, Health Posts), a relational database design in PostgreSQL would be most suitable. Here's a suggested schema:

**1. Woredas Table:**

* **Columns**:
  + FtFWoredaID: Primary key, representing the unique ID for each Woreda.
  + Cohort, Region, Zone, Woreda: Attributes defining administrative and regional characteristics.
  + COPYear: A date or string field to capture the year of data collection.
  + Various demographic statistics (e.g., PregnantW, LactatingW, etc.).
* **Primary Key**: FtFWoredaID.

**2. PHCU Table (Health Centers):**

* **Columns**:
  + HFID: Auto-generated primary key.
  + FtFWoredaID: Foreign key linking to the Woredas table.
  + HFName: Name of the health center.
  + COPYear: A date or string field to capture the year of data collection.
  + Population and demographic fields similar to those in the Woreda table.
* **Primary Key**: HFID.
* **Foreign Key**: FtFWoredaID references Woredas(FtFWoredaID).

**3. HPs Table (Health Posts):**

* **Columns**:
  + HPGUID: Primary key, representing the unique identifier for each health post.
  + HFGUID: Foreign key linking to the PHCU (health center) table.
  + HFName: Health center name (for readability and redundancy).
  + HPName: Health post name.
  + COPYear: Year of data collection.
  + Population and demographic fields (similar to Woredas and PHCU).
* **Primary Key**: HPGUID.
* **Foreign Key**: HFGUID references PHCU(HFGUID).

**Relationships:**

* **One Woreda can have multiple Health Centers (PHCU)**.
* **One Health Center can have multiple Health Posts (HPs)**.

**Steps to Implement in PostgreSQL:**

1. **Create the Woredas, PHCU, and HPs tables** with the necessary columns and relationships using CREATE TABLE statements.
2. **Ensure referential integrity** by defining foreign key constraints.
3. **Import the data** from the Excel file into PostgreSQL tables using the COPY command or an ETL tool such as pgAdmin or psql command-line utility.

This approach will allow you to query, join, and analyze data efficiently while maintaining the integrity of the relationships between Woredas, PHCUs, and Health Posts.