**🎬 Project 2: Streaming Service Content Performance Analysis**

**Scenario:**A streaming service wants to analyze the performance of its content across different regions and genres. The goal is to understand viewer preferences and content popularity over time.

**Datasets:**

1. User Viewing Data: User ID, content ID, watch time (in minutes), and device type (some rows may have missing device types or incorrect watch times).
2. Content Library: Content ID, title, genre, release date, and region availability (some duplicate entries and inconsistent region tags).
3. Subscription Data: User ID, subscription type, start date, renewal status (some missing renewal statuses and irregular date formats).

**Tasks:**

* Python: Clean the datasets by fixing incorrect timestamps and removing duplicates.
* SQL: Set up tables, import the cleaned data, and write queries to analyze genre popularity, top-watched shows per region, and subscriber retention rates.
* Power BI: Create dashboards to visualize user engagement, genre popularity by region, and subscription trends over time.

**Phases & Steps:**

**Phase 1: Data Cleaning with Python**

1. Import datasets using Pandas.
2. Identify and correct invalid timestamps.
3. Remove duplicate content entries.
4. Normalize inconsistent genre labels and region tags.
5. Save cleaned data to new CSV files.

**Phase 2: SQL Database Integration**

1. Set up a database.
2. Create tables for User Viewing Data, Content Library, and Subscription Data.
3. Import cleaned data into the database.
4. Run SQL queries:
   * Top-watched genres by region.
   * Viewer retention analysis.
   * Monthly subscriber growth.

**Phase 3: Data Visualization with Power BI**

1. Connect Power BI to the database.
2. Visualize:
   * Viewer engagement across devices.
   * Regional genre popularity.
   * Subscription trends and churn rates.