**Methodology**

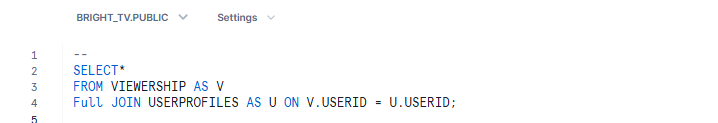
**1. Data Sources**

Two primary CSV files were utilized:

* **UserProfiles.csv** – Containing user demographic information such as USERID, Gender, Province, Age, and Race.
* **Viewership.csv** – Logging television consumption events with fields like USERID, Channel2, DURATION\_2, and RecordDate2.

**2. Data Integration**

To perform a comprehensive analysis, the datasets were integrated using a **FULL JOIN on USERID**, allowing for alignment between viewership behavior and demographic attributes.



**3. Data Cleaning & Filtering**

Several queries were constructed to filter out incomplete or invalid data entries:

* Excluding gender, province, or race marked as 'None', 'null', or empty.
* Excluding users with invalid or missing age groups.

**4. Descriptive Analytics**

The following core insights were generated using SQL:

* **Gender Distribution**:
* **Top Channels by Views**:
* **Views by Province and Race**
* **User Segmentation by Age Groups**
* **Day of the Week Viewership Trends**
* **Time Slot Analysis** *(morning, prime time, etc.)*
* **Repeat Viewers / Loyalty Measurement**

**5. Demographic Cross-Analysis**

Cross-tabulations were used to show combinations like:

* **Gender + Province**
* **Age Group + Gender**
* **Channel Preferences by Gender**

**6. Visualization & Dashboarding**

The cleaned and aggregated data was exported to **Power BI**, where various charts were designed, including:

* Bar Charts (Top Channels, Top Users)
* Donut Charts (Gender Split)
* Stacked Column Charts (Age Group by Gender)
* Heatmaps and Slicers (Time of Day, Race, Province)

**7. Outcome**

This methodology enabled a multi-dimensional view of television consumption behavior segmented by user demographics. It provides actionable insights for marketing, content personalization, and audience targeting strategies.