Objectives of the Task:
- To uncover trends in **TV engagement and user profiles.**
- To identify **viewership behaviour** by day, time, and channel.
- To segment insights by **age group, gender, and province.**
- To highlight **key audience segments** and periods of high activity.
- To recommend potential focus areas for **content and engagement strategy.**

Data Preparation
- Cleaned user profile data by removing spaces from fields like Name, Email, and Social_Media_Handle.
- Created Full_Name, Age_Group, and Age_Category columns to enable demographic segmentation.
- Categorised TV channels into **broader groups** (News, Sports, Music, Kids, Entertainment, Other) for easier trend analysis.
- Extracted **date and time dimensions** from viewership timestamps: Record_Date, Month_Name, DayOfWeek, Day_Type (Weekday/Weekend) Time_Category (Morning, Afternoon, Evening, Night)
- Split duration into hours, minutes, and seconds to analyse viewing length.

Data Integration
- Merged **User Profiles** and **Viewership Data** using a **FULL OUTER JOIN**.
- Introduced Account_Status (Active vs. Inactive) based on whether the user had viewership activity.
- Retained NULL profiles since they still represent registered users (some may have incomplete data but valid UserIDs).

The BrightTV dashboard visualises audience behaviour across **age, gender, location, channel, and time.**
On the **Overview page**, most viewers are **Young Adults**, mainly **male**, with strong audiences in **Gauteng** and **Western Cape.** **Sports** and **Entertainment** dominate engagement.
On the **Engagement Insights page**, viewing peaks in the **Afternoon** and **Evening**, especially on **weekends. Sports** content has the **longest watch time**, and **Young Adults** show the widest channel engagement, while **Children** favour **Cartoon Network** and **Boomerang.**

Databricks: Database Assembly

Data Source: Mapping out relevant data source CSV

Load Data into Catalogue

Created Bright TV workspace

Identify data types.

EDA: Explore the uploaded data structure.

INT

DATE/TIME

DECIMAL

VARCHAR

USER_PROFILES.CSV
- UserID

VIEWERSHIP.CSV
- UserID

VIEWERSHIP.CSV
- Record_Date
- Month_Name
- DayOfWeek
- Day_Type
- Record_Time
- Time_Category
- Duration_Clean
- Hours
- Minutes
- Seconds

USER_PROFILE.CSV
- Name
- Surname
- Email
- Gender
- Race
- Age
- Province
- Social Media Handle

VIEWERSHIP.CSV
- Channel2
- RecordDate2

VIEWERSHIP.CSV
- Channel Category

Viewership Channel2 & RecordDate were transformed into SAST from UTC time as a string to actual SAST time and date standards, including Days of the week, month names, record dates, time categories through CASE Statements, hours, minutes and seconds.