

Methodology: Bright TV

Data Sources

User Profile Table

Contains demographic and social data for each viewer, including (User ID, Name, Surname, Email, Gender, Race, Age, Province, Social Media Handle)

Viewership Table

Captures user-level viewing behavior, including (User ID, Channel2, Date, Time, Duration 2)

Data Cleaning and Preparation

- Verified consistency in User ID across both tables to ensure relational integrity.
- Removed duplicate records and incomplete entries (e.g., missing Duration or User ID values).
- Standardized time (from UTC to SA Time) and date formats for chronological analysis.

Data Integration

Joined the two tables using User ID as the primary key to form a unified dataset for analysis.

1. Full outer Join on user id

```
4  --Full outer join on userid
5  SELECT
6  *
7  FROM
8  VIEWERSHIP AS V
9  FULL OUTER JOIN
10 USER_PROFILE AS U
11 ON
12 V.userid = U.userid;
```

Results Chart

	# USERID	CHANNEL2	DATE	TIME_	DURATION_2	# USERID	NAME	SURNAME
1	903994	Trace TV	0016-01-04	08:09:00	00:02:00	903994	Bernardo	Hier
2	802352	ICC Cricket Work	0016-03-30	11:04:00	00:10:06	802352	None	None
3	866125	ICC Cricket Work	0016-03-30	07:01:00	00:00:42	866125	Emerson	Mccollum
4	2425604	ICC Cricket Work	0016-03-25	05:23:00	00:00:10	2425604	Tiffani	Pilot
5	793571	SawSee	0016-02-02	07:18:00	00:01:33	793571	Shelley	Reisinger
6	793571	Boomerang	0016-03-18	06:56:00	00:00:10	793571	Shelley	Reisinger

2. Total viewership by Channel

```
15 SELECT
16 V.channel2,
17 COUNT(U.userid) AS total_viewership
18 FROM
19 USER_PROFILE AS U
20 LEFT JOIN
21 VIEWERSHIP AS V
22 ON
23 U.userid = V.userid
```

Results Chart

	CHANNEL2	TOTAL_VIEWERSHIP
1	SuperSport Live Events	24
2	Sawsee	4
3	Boomerang	714
4	Break in transmission	66
5	CNN	505
6	SawSee	251
7	kykNET	45
8	DStv Events 1	107
9	null	989

3. Viewership by gender

```
28 SELECT
29   u.userid,
30   u.gender,
31   v.channel2
32 FROM
33   user_profile AS u
34 JOIN
35   viewership AS v
36 ON
37   u.userid = v.userid
38 ORDER BY
39   u.gender;
```

	# USERID	GENDER	CHANNEL2
9844	821083	male	Supersport Live Events
9845	893076	male	SawSee
9846	789220	male	ICC Cricket World Cup 2011
9847	789220	male	SuperSport Blitz
9848	819656	male	Africa Magic
9849	789220	male	Supersport Live Events
9850	789220	male	Supersport Live Events

4. Daily interval

```
127 SELECT
128   CASE DAYOFWEEK(v.date)
129     WHEN 1 THEN 'Sunday'
130     WHEN 2 THEN 'Monday'
131     WHEN 3 THEN 'Tuesday'
132     WHEN 4 THEN 'Wednesday'
133     WHEN 5 THEN 'Thursday'
134     WHEN 6 THEN 'Friday'
135     WHEN 7 THEN 'Saturday'
```

	DAY_OF_WEEK	TOTAL_VIEWS
1	Thursday	
2	Friday	
3	Tuesday	
4	Wednesday	
5	null	
6	Monday	
7	Sunday	

5. Converting duration time to seconds

```
156 --converting duration to seconds
157 SELECT
158     EXTRACT(HOUR FROM duration_2) * 3600 +
159     EXTRACT(MINUTE FROM duration_2) * 60 +
160     EXTRACT(SECOND FROM duration_2) AS total_seconds
161 FROM viewership;
```

Results Chart

	# TOTAL_SECONDS
1	120
2	606
3	42
4	10
5	93
6	10
7	1035
8	73
9	285
10	40

6. Time intervals, age grouping and Month conversion

```
232 CASE
233 WHEN Time_ BETWEEN '00:00:01'::TIME AND '04:00:00'::TIME THEN '12AM-4AM'
234 WHEN Time_ BETWEEN '04:00:01'::TIME AND '08:00:00'::TIME THEN '4AM-8AM'
235 WHEN Time_ BETWEEN '08:00:01'::TIME AND '12:00:00'::TIME THEN '8AM-12PM'
236 WHEN Time_ BETWEEN '12:00:01'::TIME AND '16:00:00'::TIME THEN '12PM-4PM'
237 WHEN Time_ BETWEEN '16:00:01'::TIME AND '20:00:00'::TIME THEN '4PM-8PM'
238 ELSE '8PM-12AM'
239 END AS TIME_BUCKET,
240
241 CASE
242 WHEN age BETWEEN 0 AND 12 THEN 'Children'
243 WHEN age BETWEEN 13 AND 19 THEN 'Teenager'
244 WHEN age BETWEEN 20 AND 34 THEN 'Youth'
245 WHEN age BETWEEN 35 AND 39 THEN 'Old youth'
246 WHEN age BETWEEN 40 AND 59 THEN 'Adults'
247 ELSE 'Pensioner'
```

Results Chart

	TIME_BUCKET	AGE_GROUP	MONTH_
1	8AM-12PM	Youth	APR
2	8AM-12PM	Children	APR
3	4AM-8AM	Youth	APR
4	4AM-8AM	Adults	APR
5	4AM-8AM	Children	APR

```
243 WHEN age BETWEEN 13 AND 19 THEN 'Teenager'
244 WHEN age BETWEEN 20 AND 34 THEN 'Youth'
245 WHEN age BETWEEN 35 AND 39 THEN 'Old youth'
246 WHEN age BETWEEN 40 AND 59 THEN 'Adults'
247 ELSE 'Pensioner'
248 END AS AGE_GROUP,
249
250 CASE
251 WHEN DATE BETWEEN '2016-01-01' AND '2016-01-31' THEN 'JAN'
252 WHEN DATE BETWEEN '2016-02-01' AND '2016-02-29' THEN 'FEB'
253 WHEN DATE BETWEEN '2016-03-01' AND '2016-03-31' THEN 'MAR'
254 ELSE 'APR'
255 END AS MONTH_
256
257 FROM BRIGHTTV_;
```

Results Chart

	TIME_BUCKET	AGE_GROUP	MONTH_
1	8AM-12PM	Youth	APR
2	8AM-12PM	Children	APR
3	4AM-8AM	Youth	APR
4	4AM-8AM	Adults	APR
5	4AM-8AM	Children	APR