

Compiling the Bright TV case study

Two table were joined on full outer join so that we can have the relevant information on the viewers watching. Data set had to be cleaned as it also contained user profile information but some of those viewers have not watched any channel. Decided to exclude the users name, surname, email and addresses as that wasn't going to provident any significant value to the insights.

Syntax for Full Outer Join to join the two tables (viewership and the user profile)

```
---JOINING USERPROFILE TABLE AND VIEWERS TABLE

SELECT A.USERID,
       GENDER,
       RACE,
       AGE,
       PROVINCE,
       B.CHANNEL,
       "RecordDate2 (UTC)",
       DAY,
       TIME,
       DURATION,
FROM USER_PROFILE AS A
FULL OUTER JOIN VIEWERS AS B
ON A.USERID = B.USERID;
```

Results										
Chart										
	# USERID	△ GENDER	△ RACE	# AGE	△ PROVINCE	△ CHANNEL	△ RecordDate2 (UTC)	△ DAY	⌚ TIME	⌚ DURATION
1	903994	male	coloured	33	Western Cape	Trace TV	1/4/2016 18:09	Monday	20:09:00	00:02:00
2	802352	None	None	0	None	ICC Cricket World C	3/30/2016 9:04	Wednesday	11:04:00	00:10:08
3	866125	male	indian_asian	30	Gauteng	ICC Cricket World C	3/30/2016 17:01	Wednesday	19:01:00	00:00:42
4	2425604	female	white	51	Gauteng	ICC Cricket World C	3/25/2016 15:23	Friday	17:23:00	00:00:10
5	793571	female	white	9	Eastern Cape	SawSee	2/2/2016 17:18	Tuesday	19:18:00	00:01:33
6	793571	female	white	9	Eastern Cape	Boomerang	3/18/2016 16:56	Friday	18:56:00	00:00:10
7	793571	female	white	9	Eastern Cape	Channel O	3/20/2016 4:25	Sunday	06:25:00	00:17:15
8	793571	female	white	9	Eastern Cape	Boomerang	2/3/2016 6:43	Wednesday	08:43:00	00:01:13
9	819248	None	None	0	None	Channel O	3/31/2016 10:01	Thursday	12:01:00	00:04:45
10	820321	male	indian_asian	26	Kwazulu Natal	Cartoon Network	1/9/2016 8:25	Saturday	10:25:00	00:00:40
11	892453	male	indian_asian	34	Gauteng	Supersport Live Ev	3/24/2016 9:09	Thursday	11:09:00	00:00:06
12	892453	male	indian_asian	34	Gauteng	ICC Cricket World C	3/26/2016 15:33	Saturday	17:33:00	05:01:12
13	828799	male	None	38	Mpumalanga	Channel O	2/28/2016 8:09	Sunday	10:09:00	00:00:00
14	787656	male	black	26	Gauteng	Channel O	2/15/2016 8:40	Monday	10:40:00	00:11:15
15	939566	male	None	29	Mpumalanga	Supersport Live Ev	1/9/2016 15:40	Saturday	17:40:00	00:22:02

To determine the total viewers subscribed to the Bright TV Channels

CASESTUDY.PUBLIC ▾ Settings ▾

Code Versions 🔍

```
24 FROM COMBTABLE;
25
26 --- Determine the total amount of viewers who have subscribed
27
28 SELECT COUNT (DISTINCT USERID) AS TOTAL_USERS
29 FROM COMBTABLE;
30
31
```

↩ Results Chart

🔍 📄 ⬇ 📋 ⌚

	# TOTAL_USERS
1	5375

Query Details ⋮

Query duration 781ms

Rows 1

Query ID 01bc3a8e-0000-87a0-0...

Show more ▾

TOTAL_USERS #

100% filled

How to change the hh:mm:ss format to seconds to calculate the total viewership time

CASESTUDY.PUBLIC Settings Code Versions

```
30  
31  
32 ---Converting duration to seconds  
33  
34 SELECT DURATION,  
35        extract(hour from duration)*3600+  
36        extract(minute from duration)*60+  
37        extract(second from duration) AS DURATION1  
38 FROM viewers;  
39
```

Results Chart

	DURATION	# DURATION1
1	00:02:00	120
2	00:10:06	606
3	00:00:42	42
4	00:00:10	10
5	00:01:33	93
6	00:00:10	10
7	00:17:15	1035
8	00:01:13	73
9	00:04:45	285
10	00:00:40	40
11	00:00:06	6
12	05:01:12	18072

Query Details

Query duration 1.4s

Rows 10K

Query ID 01bc3a92-0000-87a1-0...

Show more

DURATION 100% filled

DURATION1 #

0 41368

To establish which weekday has most viewers (split by gender), what channels they watching and how much time they spending on each channel.

CASESTUDY.PUBLIC Settings Code Versions

```
41  
42 ----Finding which day has most viewerships  
43 WITH join_tables AS (SELECT  
44     A.userid,  
45     A.channel,  
46     A.duration,  
47     EXTRACT(HOUR FROM A.duration) * 3600 +  
48     EXTRACT(MINUTE FROM A.duration) * 60 +  
49     EXTRACT(SECOND FROM A.duration) AS duration1,  
50     B.gender,  
51     A.day  
52 FROM viewers AS A  
53 FULL OUTER JOIN user_profile AS B  
54 ON A.USERID = B.USERID)  
55 SELECT  
56     DAY, CHANNEL, GENDER,  
57     COUNT(DURATION1) AS TOTAL_TIME,  
58 FROM join_tables  
59 GROUP BY day, channel, GENDER  
60 ORDER BY DAY;  
61
```

Results Chart

	DAY	CHANNEL	GENDER	# TOTAL_TIME
1	Friday	E! Entertainment	male	44
2	Friday	SuperSport Blitz	male	110
3	Friday	M-Net	None	1
4	Friday	ICC Cricket World Cup 2011	male	187
5	Friday	Boomerang	male	92

Query Details

Query duration154ms

Rows318

Query ID01bc3a98-0000-87a1-0...

Show more

Establishing weekday with most viewers limit it to top ten viewers who sent most time viewing

```
CASESTUDY.PUBLIC  Settings
61
62  ---Finding which day has most viewerships, LIMIT TO TEN
63  WITH join_tables AS (SELECT
64      A.userid,
65      A.channel,
66      A.duration,
67      EXTRACT(HOUR FROM A.duration) * 3600 +
68      EXTRACT(MINUTE FROM A.duration) * 60 +
69      EXTRACT(SECOND FROM A.duration) AS duration1,
70      B.gender,
71      A.day
72  FROM viewers AS A
73  FULL OUTER JOIN user_profile AS B
74  ON A.USERID = B.USERID)
75  SELECT
76      DAY, CHANNEL, GENDER,
77      COUNT(DURATION1) AS TOTAL_TIME,
78  FROM join_tables
79  GROUP BY day, channel, GENDER
80  ORDER BY TOTAL_TIME DESC
81  LIMIT 10;
82
83
```

OUTPUT TABLE

<div> Results Chart </div>					<div> Query Details ... </div>	
<div> DAY CHANNEL GENDER TOTAL_TIME </div>					<div> Query duration 51ms </div>	
1	Wednesday	ICC Cricket World Cup 2011	male	455	<div> Rows 10 </div>	
2	Friday	Supersport Live Events	male	383	<div> Query ID 01bc3aa2-0000-87a0-0... </div>	
3	Thursday	Supersport Live Events	male	346	<div> Show more </div>	
4	Saturday	Supersport Live Events	male	301	<div> <div> <div>DAY</div> <div>Friday</div> <div>Wednesday</div> <div>Saturday</div> </div> <div> <div>3</div> <div>2</div> <div>2</div> </div> </div>	
5	Saturday	ICC Cricket World Cup 2011	male	232		
6	Friday	ICC Cricket World Cup 2011	male	187		
7	Tuesday	ICC Cricket World Cup 2011	male	186		
8	Sunday	Supersport Live Events	male	183		
9	Friday	Channel O	male	152		
10	Wednesday	Channel O	male	148		

Determining viewership by gender and which gender spend most of they time watching

CASESTUDY.PUBLIC

Settings

Code Versions

```
84 ---Viewership by Gender
85
86 WITH join_tables AS (SELECT channel, gender, B.userid
87 FROM viewers AS A
88 FULL OUTER JOIN user_profile AS B
89 ON A.userid =B.userid)
90 SELECT Gender,channel,count(userid)
91 FROM join_tables
92 Group by gender,channel
93 order by gender;
94
```

Results

Chart

	DAY	CHANNEL	GENDER	TOTAL_TIME
1	Wednesday	ICC Cricket World Cup 2011	male	455
2	Friday	Supersport Live Events	male	383
3	Thursday	Supersport Live Events	male	346
4	Saturday	Supersport Live Events	male	301
5	Saturday	ICC Cricket World Cup 2011	male	232
6	Friday	ICC Cricket World Cup 2011	male	187
7	Tuesday	ICC Cricket World Cup 2011	male	186
8	Sunday	Supersport Live Events	male	183
9	Friday	Channel O	male	152
10	Wednesday	Channel O	male	148

Query Details

...

Query duration25ms

Rows10

Query ID01bc3aa3-0000-87a0-0...

Show more

DAY

Friday3

Wednesday2

Saturday2

Group the information by channel to see which channel are watched by the viewers

CASESTUDY.PUBLIC ▾ Settings ▾

```
94  
95 ----GROUP BY CHANNEL  
96  
97 SELECT count(USERID), CHANNEL  
98 FROM (SELECT B.userid,channel  
99 FROM VIEWERS AS A  
100 FULL OUTER JOIN user_profile AS B  
101 ON A.userid=B.userid)  
102 GROUP BY channel;  
103  
104
```

Results Chart

#	COUNT(USERID)	CHANNEL
1	859	Africa Magic
2	714	Boomerang
3	66	Break in transmission
4	505	CNN
5	793	Cartoon Network
6	1050	Channel O
7	107	DStv Events 1
8	367	E! Entertainment
9	1465	ICC Cricket World Cup 2011
10	2	Live on SuperSport
11	116	M-Net

How many Null values are in this data set

```
104  
105 ---- How many items have NULL or no value  
106  
107 SELECT COUNT(*) AS Missing_Count  
108 FROM COMBTABLE  
109 WHERE GENDER IS NULL OR GENDER = 'None' OR GENDER = 'Other'  
110      OR RACE IS NULL OR RACE = 'None' OR RACE = 'Other'  
111      OR AGE IS NULL OR AGE = 0  
112      OR PROVINCE IS NULL OR PROVINCE = 'None' OR PROVINCE = 'Province';  
113  
114  
115 ----AGE BUCKET OF VIEWERS
```

Results Chart

MISSING_COUNT
1829

Query Details

Query duration 89ms

Rows 1

Query ID 01bc3aaa-0000-87a0-0...

Show more

Group by the viewer per age bucket

CASESTUDY.PUBLIC ▾ Settings ▾

Code Versions 🔍

```
114
115 ----AGE BUCKET OF VIEWERS
116 SELECT UserID, Gender, Race,
117 CASE
118 WHEN AGE between 3 AND 12 THEN 'Youngster'
119 WHEN AGE between 13 and 19 THEN 'Teenager'
120 WHEN AGE between 20 and 59 THEN 'Adult'
121 WHEN AGE between 60 and 70 THEN 'Senior'
122 ELSE 'Golden Ager'
123 END AS AGE_CATEGORY
124 FROM COMBTABLE;
125
126
```

Results Chart

🔍 📄 ⬇️ 📋 ⌛

#	USERID	GENDER	RACE	AGE_CATEGORY
1	903994	male	coloured	Adult
2	802352	None	None	Golden Ager
3	866125	male	indian_asian	Adult
4	2425604	female	white	Adult
5	793571	female	white	Youngster
6	793571	female	white	Youngster
7	793571	female	white	Youngster
8	793571	female	white	Youngster
9	819248	None	None	Golden Ager
10	820321	male	indian_asian	Adult

Query Details ⋮

Query duration 48ms

Rows 11.0K

Query ID 01bc3aac-0000-87a1-0...

Show more ▾

USERID #

34 2429528

Total viewers per province

