Bright TV Viewership Review

SQL Analysis on Snowflake

Presented by: Lerato Motsusi

Combing User Profile Table and Viewership Table

I decided to join the two tables (Userprofile and Viewership) on a full outer join as I believe there are insights that can also be made from users who have profiles but have not watched any shows in the sample period.

It is also important to note that I did not include each users name, surname and email addresses as I do not believe that it will provide any significant insight to know the users personal details.

The Syntax used was:

CASESTUDY.PUBLIC Settings Code Versions

```
SELECT*FROM USERPROFILES

SELECT A.USERID,
       A.GENDER,
       A.RACE,
       A.AGE,
       A.PROVINCE,
       B.CHANNEL2,
       B.RECORDDATE2,
       B.DURATION2
 FROM USERPROFILES AS A
 FULL OUTER JOIN VIEWERSHIP AS B
 ON A.userid = B."userid";
```

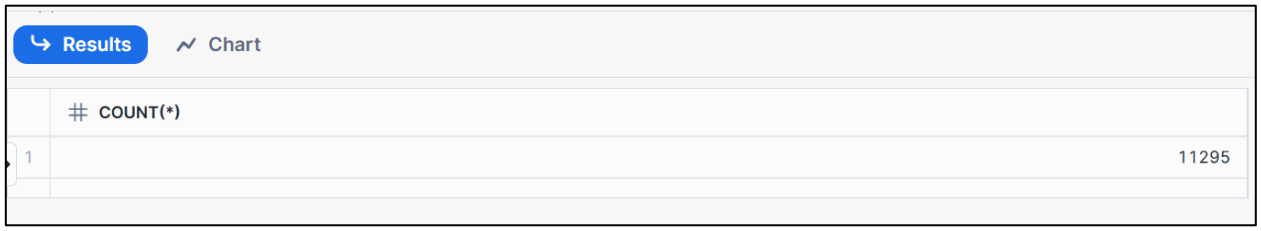
ACCOUNTADMIN • COMPUTE_WH (X-Small) Share

Results Chart

	USERID	GENDER	RACE	AGE	PROVINCE	CHANNEL2	RECORDDATE2	DURATION2
1	903994	male	coloured	33	Western Cape	Trace TV	1/4/2016 18:09	00:02:00
2	802352	None	None	0	None	ICC Cricket World Cup 2011	3/30/2016 9:04	00:10:06
3	866125	male	indian_asian	30	Gauteng	ICC Cricket World Cup 2011	3/30/2016 17:01	00:00:42
4	2384970	male	coloured	39	Kwazulu Natal	ICC Cricket World Cup 2011	3/25/2016 15:23	00:00:10
5	793571	female	white	9	Eastern Cape	SawSee	2/2/2016 17:18	00:01:33
6	793571	female	white	9	Eastern Cape	Boomerang	3/18/2016 16:56	00:00:10
7	793571	female	white	9	Eastern Cape	Channel O	3/20/2016 4:25	00:17:15
8	793571	female	white	9	Eastern Cape	Boomerang	2/3/2016 6:43	00:01:13
9	819248	None	None	0	None	Channel O	3/31/2016 10:01	00:04:45
10	820321	male	indian_asian	26	Kwazulu Natal	Cartoon Network	1/9/2016 8:25	00:00:40
11	892453	male	indian_asian	34	Gauteng	Supersport Live Events	3/24/2016 9:09	00:00:06
12	892453	male	indian_asian	34	Gauteng	ICC Cricket World Cup 2011	3/26/2016 15:33	05:01:12

Understanding the data

Syntax to count number of records to see what the size of the sample is. Based on the results, there are 11 295 records on the new combined tables.



Syntax to count number of unique User Profiles to see the number of profiles there are on record. Based on the results, there are 5375 unique User Profiles.

Replacing NULL Values and separating the Timestamp

In the syntax below, I wanted to achieve the below objectives:

- To make sure that I re-name all the 'NULL' or zero values in each column
- Round off the values in the 'User_ID' columns and the 'Age' columns
- Chang the content in the "Recorddate2" column from a string into a timestamp format
- Separate the timestamp columns into two separate columns - one containing just the date, the other just containing the time.

The reason for all the above is to clean up the data and to also make sure the date and time are in a format that is usable for me to draw insights from this information.

CASESTUDY1.PUBLIC Settings Code Versions

```
SELECT
    IFNULL(ROUND(USERID, 0), '0') AS USER_ID,
    IFNULL(GENDER, 'No_Gender') AS GENDER,
    IFNULL(RACE, 'No_Race') AS RACE,
    IFNULL(ROUND(AGE, 0), '0') AS AGE,
    IFNULL(PROVINCE, 'No_Province') AS PROVINCE,
    IFNULL(CHANNEL2, 'No_Channel') AS CHANNEL2,
    IFNULL(RECORDDATE2, 'No_Recorddate') AS RECORDDATE2,
    TO_TIMESTAMP(RECORDDATE2, 'MM/DD/YYYY HH24:MI') AS TIMESTAMP_VALUE,
    CAST(TO_TIMESTAMP(RECORDDATE2, 'MM/DD/YYYY HH24:MI') AS DATE) AS DATE_PART,
    TO_CHAR(TO_TIMESTAMP(RECORDDATE2, 'MM/DD/YYYY HH24:MI'), 'HH24:MI:SS') AS TIME_PART
FROM BRIGHTTV_CS;
```

ACCOUNTADMIN • COMPUTE_WH (X-Small) Share

Results Chart

	RACE	AGE	PROVINCE	CHANNEL2	RECORDDATE2	TIMESTAMP_VALUE	DATE_PART	TIME_PART
1	coloured	33	Western Cape	Trace TV	1/4/2016 18:09	2016-01-04 18:09:00.000	2016-01-04	18:09:00
2	None	0	None	ICC Cricket Worlc	3/30/2016 9:04	2016-03-30 09:04:00.000	2016-03-30	09:04:00
3	indian_asian	30	Gauteng	ICC Cricket Worlc	3/30/2016 17:01	2016-03-30 17:01:00.000	2016-03-30	17:01:00
4	coloured	39	Kwazulu Natal	ICC Cricket Worlc	3/25/2016 15:23	2016-03-25 15:23:00.000	2016-03-25	15:23:00
5	white	9	Eastern Cape	SawSee	2/2/2016 17:18	2016-02-02 17:18:00.000	2016-02-02	17:18:00
6	white	9	Eastern Cape	Boomerang	3/18/2016 16:56	2016-03-18 16:56:00.000	2016-03-18	16:56:00
7	white	9	Eastern Cape	Channel O	3/20/2016 4:25	2016-03-20 04:25:00.000	2016-03-20	04:25:00
8	white	9	Eastern Cape	Boomerang	2/3/2016 6:43	2016-02-03 06:43:00.000	2016-02-03	06:43:00
9	None	0	None	Channel O	3/31/2016 10:01	2016-03-31 10:01:00.000	2016-03-31	10:01:00
10	indian_asian	26	Kwazulu Natal	Cartoon Network	1/9/2016 8:25	2016-01-09 08:25:00.000	2016-01-09	08:25:00
11	indian_asian	34	Gauteng	Supersport Live E	3/24/2016 9:09	2016-03-24 09:09:00.000	2016-03-24	09:09:00
12	indian_asian	34	Gauteng	ICC Cricket Worlc	3/26/2016 15:33	2016-03-26 15:33:00.000	2016-03-2	

Ask Copilot

Viewership by Province

The syntax below is to count number of users per province in descending order. Based on the results, Gauteng has the highest number of viewers , followed by the Western Cape and Kwa Zulu Natal. There are 221users who did not give information on the province they are in (NULL values).

CASESTUDY1.PUBLIC Settings Code Versions

```
50      COUNT(userid) AS User_count_by_Province
51    FROM brighttv_cs
52    GROUP BY PROVINCE
53    ORDER BY user_count_by_province DESC;
54
```

Results Chart

	PROVINCE	USER_COUNT_BY_PROVINCE
1	Gauteng	3786
2	Western Cape	1883
3	Kwazulu Natal	1217
4	Mpumalanga	964
5	None	828
6	Limpopo	776
7	Eastern Cape	718
8	North West	358
9	Free State	308
10	Northern Cape	236
11	null	221

Most watched shows by viewers

The syntax below is to count number of users per show on Channel 2 in descending order. Based on the results, Supersport Live Events were the most watched show on the channel, closely followed by the ICC Cricket World Cup 2011. We do not have information on what show was being watched by the third highest number of users. This will need to be further investigated.

CASESTUDY1.PUBLIC Settings Code Versions

```
58  SELECT
59    CHANNEL2,
60    COUNT(userid) AS Channel2_Viewership
61  FROM BRIGHTTV_CS
62  GROUP BY CHANNEL2
63  ORDER BY Channel2_Viewership DESC;
```

Results Chart

	CHANNEL2	CHANNEL2_VIEWERSHIP
1	Supersport Live Events	1638
2	ICC Cricket World Cup 2011	1465
3	No_Views	1295
4	Channel O	1050
5	Trace TV	952
6	SuperSport Blitz	896
7	Africa Magic	859
8	Cartoon Network	793
9	Boomerang	714
10	CNN	505
11	E! Entertainment	367
12	SawSee	
13	M-Net	116

Ask Copilot

Viewership Split By Age Buckets

The syntax below is to group users according to various age buckets. The results show us that the people who watch Channel 2 are adults between the ages of 30-50.s

CASESTUDY1.PUBLIC Settings Code Versions

```
92  SELECT
93      CASE
94          WHEN AGE IS NULL THEN 'Unknown'
95          WHEN AGE < 13 THEN 'Child'
96          WHEN AGE >= 13 AND AGE < 18 THEN 'Teen'
97          WHEN AGE >= 18 AND AGE < 30 THEN 'Young Adult'
98          WHEN AGE >= 30 AND AGE < 50 THEN 'Adult'
99          WHEN AGE >= 50 AND AGE < 65 THEN 'Middle Aged'
100         WHEN AGE >= 65 THEN 'Senior'
101         ELSE 'Other'
102     END AS AGE_GROUP,
103     COUNT(USERID) AS USER_COUNT
104 FROM BRIGHTTV_CS
105 GROUP BY
106     CASE
107         WHEN AGE IS NULL THEN 'Unknown'
108         WHEN AGE < 13 THEN 'Child'
109         WHEN AGE >= 13 AND AGE < 18 THEN 'Teen'
110         WHEN AGE >= 18 AND AGE < 30 THEN 'Young Adult'
111         WHEN AGE >= 30 AND AGE < 50 THEN 'Adult'
112         WHEN AGE >= 50 AND AGE < 65 THEN 'Middle Aged'
```

Ask Copilot

Results Chart

	AGE_GROUP	USER_COUNT
1	Senior	53
2	Teen	199
3	Middle Aged	529
4	Child	1151
5	Young Adult	3782
6	Adult	5581

Viewership Split By Time Slots

The below syntax buckets the record time of when the user started watching a show by time slots through out the day. The results show us that the most popular time slot that our sample of people would watch Channel 2 was in the afternoon time slot, which is between 12 noon to 6pm.

CASESTUDY1.PUBLIC Settings Code Versions

```
328
329  SELECT
330      CASE
331          WHEN TIME_CONVERTED BETWEEN '05:00:00' AND '11:59:59' THEN 'Morning_Viewing'
332          WHEN TIME_CONVERTED BETWEEN '12:00:00' AND '17:59:59' THEN 'Afternoon_Viewing'
333          WHEN TIME_CONVERTED BETWEEN '18:00:00' AND '23:59:59' THEN 'Evening_Viewing'
334          ELSE 'Midnight_Viewing'
335      END AS Viewing_slots,
336      COUNT(USERID) AS USER_COUNT
337  FROM BRIGHTTV_CS
338  GROUP BY
339      CASE
340          WHEN TIME_CONVERTED BETWEEN '05:00:00' AND '11:59:59' THEN 'Morning_Viewing'
341          WHEN TIME_CONVERTED BETWEEN '12:00:00' AND '17:59:59' THEN 'Afternoon_Viewing'
342          WHEN TIME_CONVERTED BETWEEN '18:00:00' AND '23:59:59' THEN 'Evening_Viewing'
343          ELSE 'Midnight_Viewing'
344      END
345      ORDER BY User_count DESC;
346
347
348
```

Ask Copilot

Results Chart

	VIEWING_SLOTS	USER_COUNT
1	Afternoon_Viewing	3807
2	Morning_Viewing	3383
3	Evening_Viewing	2352
4	Midnight_Viewing	1753

Viewership Split By Gender

The syntax below is to count number of males and females in the data set. From the results, we can see that most of the viewers are males (9190) which can be a possible reason as to why sports shows make up 60% of the shows viewed on Channel 2. Many men enjoy watching sports.

CASESTUDY1.PUBLIC Settings Code Versions

```
172  SELECT
173    GENDER,
174    COUNT(userid) AS Male_Female_Split
175  FROM BRIGHTTV_CS
176  GROUP BY GENDER
177  ORDER BY Male_Female_Split DESC;
```

Results Chart

	GENDER	MALE_FEMALE_SPLIT
1	male	9190
2	female	1056
3	None	828
4	No_Gender	221

Power BI Dashboard