Attribution Sample

With SQL

By Martin MacFarlane
March 21, 2025

Summary

The following short project is intended to showcase SQL skills while gathering attribution insights from the "page_visits" table.

Questions

1. How many distinct campaigns are there?

2. How many distinct sources are there?

3. How are they related

4. What are the distinct page names?

5. When were the "first touches" for each campaign?

6. When were the "last touches" for each campaign for users who made a purchase?

page_name	timestamp	user_id	utm_campaign	utm_source
1 - landing_page	2018-01-24 03:12:16	10006	getting-to-know-cool-tshirts	nytimes
2 - shopping_cart	2018-01-24 04:04:16	10006	getting-to-know-cool-tshirts	nytimes
3 - checkout	2018-01-25 23:10:16	10006	weekly-newsletter	email
1 - landing_page	2018-01-25 20:32:02	10030	ten-crazy-cool-tshirts-facts	buzzfeed
2 - shopping_cart	2018-01-25 23:05:02	10030	ten-crazy-cool-tshirts-facts	buzzfeed
3 - checkout	2018-01-28 13:26:02	10030	retargetting-campaign	email
4 - purchase	2018-01-28 13:38:02	10030	retargetting-campaign	email
1 - landing_page	2018-01-05 18:31:17	10045	getting-to-know-cool-tshirts	nytimes
2 - shopping_cart	2018-01-05 21:16:17	10045	getting-to-know-cool-tshirts	nytimes
3 - checkout	2018-01-09 03:05:17	10045	retargetting-ad	facebook
1 - landing_page	2018-01-16 04:17:46	10048	interview-with-cool-tshirts-founder	medium
2 - shopping_cart	2018-01-16 07:28:46	10048	interview-with-cool-tshirts-founder	medium
3 - checkout	2018-01-19 00:00:46	10048	retargetting-campaign	email
1 - landing_page	2018-01-02 23:14:01	10069	ten-crazy-cool-tshirts-facts	buzzfeed
2 - shopping_cart	2018-01-02 23:55:01	10069	ten-crazy-cool-tshirts-facts	buzzfeed
3 - checkout	2018-01-04 08:12:01	10069	retargetting-ad	facebook
4 - purchase	2018-01-04 08:13:01	10069	retargetting-ad	facebook
1 - landing_page	2018-01-29 21:37:10	10162	getting-to-know-cool-tshirts	nytimes
2 - shopping_cart	2018-01-29 22:11:10	10162	ten-crazy-cool-tshirts-facts	buzzfeed
3 - checkout	2018-02-01 04:15:10	10162	weekly-newsletter	email

```
1    -- Preview the data
2    SELECT *
3    FROM page_visits
4    LIMIT 20;
```

Preview The Data

utm_campaign	utm_source
getting-to-know-cool-tshirts	nytimes
weekly-newsletter	email
ten-crazy-cool-tshirts-facts	buzzfeed
retargetting-campaign	email
retargetting-ad	facebook
interview-with-cool-tshirts-founder	medium
paid-search	google
cool-tshirts-search	google

```
14    -- How are they related?
15    SELECT DISTINCT utm_campaign, utm_source
16    FROM page_visits;
17
```

3. How are they related?

```
18     -- What are the distinct page names
19     SELECT DISTINCT page_name
20     FROM page_visits;
21
```

page_name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

4. What are the distinct page names?

```
22    -- Display first touches for each campaign
23    WITH first_touch AS (
24      SELECT user_id,
25        MIN(timestamp) AS 'first_touch_at'
26      FROM page_visits
27      GROUP BY 1
28    ),
29    ft_attr AS(
30    SELECT
31      ft.user_id, ft.first_touch_at,
32      pv.utm_campaign, pv.utm_source
33    FROM first_touch as ft
34    JOIN page_visits as pv
35      ON ft.user_id = pv.user_id
36      AND ft.first_touch_at = pv.timestamp
37    )
38    SELECT
39      ft_attr.utm_campaign,
40      ft_attr.utm_source,
41      COUNT(*)
42    FROM ft_attr
43    GROUP BY 1, 2
44    ORDER BY 3 DESC;
45
```

5. "First Touches" for each campaign

utm_campaign	utm_source	COUNT(*)
interview-with-cool-tshirts-founder	medium	622
getting-to-know-cool-tshirts	nytimes	612
ten-crazy-cool-tshirts-facts	buzzfeed	576
cool-tshirts-search	google	169

5. "Last Touches" for purchasers

```sql
46    -- Display last touches for each campaign for visitors who made a purchase
47    WITH last_touch AS (
48        SELECT user_id,
49            MAX(timestamp) as last_touch_at
50        FROM page_visits
51        WHERE page_name = '4 - purchase'
52        GROUP BY user_id),
53    lt_attr AS (
54    SELECT lt.user_id,
55        lt.last_touch_at,
56        pv.utm_source,
57                pv.utm_campaign
58    FROM last_touch lt
59    JOIN page_visits pv
60        ON lt.user_id = pv.user_id
61        AND lt.last_touch_at = pv.timestamp)
62    SELECT
63        lt_attr.utm_campaign, lt_attr.utm_source,
64        COUNT(*)
65        FROM lt_attr
66        GROUP BY 1,2
67        ORDER BY 3 DESC;
```

utm_campaign	utm_source	COUNT(*)
weekly-newsletter	email	115
retargetting-ad	facebook	113
retargetting-campaign	email	54
paid-search	google	52
getting-to-know-cool-tshirts	nytimes	9
ten-crazy-cool-tshirts-facts	buzzfeed	9
interview-with-cool-tshirts-founder	medium	7
cool-tshirts-search	google	2