Project Assignment: Attribution Queries using SQL

CoolTShirts, an innovative apparel shop, is running a bunch of marketing campaigns. In this project, you'll be helping them answer these questions about their campaigns:

1. Get familiar with the company.
   - How many campaigns and sources does CoolTShirts use and how are they related? Be sure to explain the difference between **utm_campaign** and **utm_source**.
   - What pages are on their website?

2. What is the user journey?
   - How many first touches is each campaign responsible for?
   - How many last touches is each campaign responsible for?
   - How many visitors make a purchase?
   - How many last touches *on the purchase page* is each campaign responsible for?
   - What is the typical user journey?

3. Optimize the campaign budget.
   - CoolTShirts can re-invest in 5 campaigns. Which should they pick and why?

codecademy

Marketing Attribution

Analyze Data with SQL
Delphine Vincent
17/07/2024

Table of Contents

1. Inspect data
   - Whole data
   - Campaigns and sources
   - Pages
2. User journey
   - First touches
   - Last touches
   - Purchasing customers
   - Campaign leading to purchase
   - Campaign investment

1. Inspect data

1.1 Inspect page_visits table

```
/* Inspect page_visits table */
SELECT * FROM page_visits LIMIT 7;
```

There are 5 columns. Page_names represent the customer (user_id) journey. Timestamp captures when those pages were visited. Campaigns (utm_campaigns) are advertisement events from CoolTSHirts assigned to specific sources (utm_sources).

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

1.2 Investigate campaigns and sources

```
/* Task 1 - identify campaigns and sources */
/*  number of distinct campaigns */
SELECT COUNT(DISTINCT utm_campaign) AS
'campaign_count'
FROM page_visits;
-- there are 8 campaigns
/* number of distinct sources */
SELECT COUNT(DISTINCT utm_source) AS
'source_count'
FROM page_visits;
-- there are 6 sources
/* relationship between campaigns and sources
*/
SELECT DISTINCT utm_campaign, utm_source
FROM page_visits
ORDER BY 2;
```

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

There are 8 campaigns and 6 sources. Most campaigns are run on unique sources, with the exception of google and email which have been assigned 2 different campaigns each.

1.3 CoolTShirts pages

```
/* Task 2 - pages on the CoolTShirts website */
SELECT DISTINCT page_name
FROM page_visits;
```

There are 4 different pages:
- landing_page
- shopping_cart
- checkout
- purchase

They represent the customer's journey, from browsing, to selecting items to buy, then reviewing the list of chosen items, and ultimately finalizing payment to purchase items.

2. User journey

2.1 First touch per campaign

```
/* number of unique users */
SELECT COUNT(DISTINCT user_id) AS
'user_total_count'
FROM page_visits;

/* Task 3 - User first journey: first
touches per campaign */
WITH first_touch AS (
    SELECT user_id,
            MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id
)
SELECT pv.utm_campaign, utm_source,
        COUNT(DISTINCT ft.user_id) AS
ft_user_count
FROM first_touch ft
JOIN page_visits pv ON ft.user_id =
pv.user_id AND ft.first_touch_at =
pv.timestamp
GROUP BY pv.utm_campaign
ORDER BY ft_user_count DESC;
```

utm_campaign	utm_source	ft_user_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

There are 1979 unique customers.
Four (out of 8) campaigns trigger first touches, with 'interview-with-cool-tshirts-founder' eliciting most of the touches (622/1979=31%) and 'cool-tshirts-search' being the least successful (169/1979=8.5%).
4 campaigns are not listed.

2.2 Last touch per campaign

```
/* Task 4 - User last journey: last
touches per campaign */
WITH last_touch AS (
    SELECT user_id,
         MAX(timestamp) as
last_touch_at
    FROM page_visits
    GROUP BY user_id
)
SELECT pv.utm_campaign, utm_source,
     COUNT(DISTINCT lt.user_id) AS
lt_user_count
FROM last_touch lt
JOIN page_visits pv ON lt.user_id =
pv.user_id AND lt.last_touch_at =
pv.timestamp
GROUP BY pv.utm_campaign
ORDER BY lt_user_count DESC;
```

utm_campaign	utm_source	lt_user_count
weekly-newsletter	email	447
retargetting-ad	facebook	443
retargetting-campaign	email	245
getting-to-know-cool-tshirts	nytimes	232
ten-crazy-cool-tshirts-facts	buzzfeed	190
interview-with-cool-tshirts-founder	medium	184
paid-search	google	178
cool-tshirts-search	google	60

All campaigns trigger last touches, with 'weekly-newsletter' eliciting most of the touches (447=23%) and 'cool-tshirts-search' being the least successful (60=3%).

2.3 Purchasing customers in total

```sql
/* Task 5 - number of purchasing customers */
SELECT COUNT(DISTINCT user_id) AS 'purchasing_user'
FROM page_visits
WHERE page_name = '4 - purchase';
```

361 customers made a purchase, which represents 18% of all visitors.

2.4 Purchasing customers per campaign

```
/* Task 6 - User last journey: last
touches per campaign leading to
purchase*/
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) as
last_touch_at
    FROM page_visits
    WHERE page_name = '4 - purchase' --
Move the filter here to only include
visits to the purchase page
    GROUP BY user_id
)
SELECT pv.utm_campaign, utm_source,
        COUNT(DISTINCT lt.user_id) AS
purchasing_user
FROM last_touch lt
JOIN page_visits pv ON lt.user_id =
pv.user_id AND lt.last_touch_at =
pv.timestamp
WHERE pv.page_name = '4 - purchase' --
This line is optional if the filter is
already applied in the CTE
GROUP BY pv.utm_campaign, pv.utm_source
ORDER BY purchasing_user DESC;
```

utm_campaign	utm_source	purchasing_user
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

The most successful campaign leading to purchase is 'weekly newsletter' by email accounting for 32% of all purchase, closely followed by 'retargeting-ad' via Facebook (31%). A 'cool-tshirts-search' via Google only elicited 2 purchases (0.6%).

2.5 Campaign investments

Our analysis showed that whilst the 'interview-with-cool-tshirts-founder' campaign via Medium enticed most of the visitors' first touches (31%), it only resulted in 184 last touches (9%) and 7 purchases (2%).

Both the weekly newsletter via email and retargeting-ad via Facebook triggered the largest numbers of last touches, respectively 447 (23%) and 443 (22%), and ultimately the largest sales (32% and 31%, respectively).

The next best 2 campaigns are: retargeting-campaign by email and paid-search via google.

I would advice CoolTShirts to renew their investments with these 4 campaign-mode/provider. As there is enough money to invest into a 5th campaign but none of the other campaigns explored so far are successful, I'd recommend exploring alternative options (tik tok, twitter, youtube, etc…).

First touch/campaign

utm_campaign	utm_source	ft_user_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

Last touch/campaign

utm_campaign	utm_source	lt_user_count
weekly-newsletter	email	447
retargetting-ad	facebook	443
retargetting-campaign	email	245
getting-to-know-cool-tshirts	nytimes	232
ten-crazy-cool-tshirts-facts	buzzfeed	190
interview-with-cool-tshirts-founder	medium	184
paid-search	google	178
cool-tshirts-search	google	60

Purchase/campaign

utm_campaign	utm_source	purchasing_user
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