codecademy

Marketing Attribution Using SQL

Analyze Data with SQL

Ant Erduman

1st October 2024

Project: Marketing Attribution Analysis Using SQL by Ant Erduman

In this representation, several questions, obtained through Codecademy's SQL education program, are answered.

Answers include queries of each finding through the data set, Codecademy provides during its course. As questions are about digital marketing some answers do include insights to the subjects of issue.

1. Getting familiar with the Company

Question 1, a

1a. 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.

CoolTshirts is using 8 campaigns which 2 of them are retargeting campaigns. While the former 6 campaigns have directed their users on the landing page, the latter 2 are directing them to checkout.

“Weekly Newsletter”, “Retargeting Campaign” provide sources through email marketing. “Paid Search” provides users through search engines with the difference that the ad is placed on top of search results, whereas “Cool T-shirt Search” is the source of a direct search. “Retargeting Ad” provides users from a social media channel which helps users to continue their journey and directs them to their before-visited page. “Getting To Know Cool T-shirts”, “Ten Crazy Cool T-shirts Facts” are sources using online article marketing, while the “Interview With Cool T-shirts Facts” can also be considered as part of online article marketing.

Data related, since they are all sources and part of a user journey, they all are bound to a specific “user_id” and reflect a unique “timestamp” which specifies users engagement with the product on its particular time.

“Utm_campaign” and “utm_source” are parameters with snippets attached in order to be able to understand where the traffic is driven from. While the former makes it possible to track from which campaign the user derived from, the latter indicates the medium from which the source was acquired from.

Question 1, a

The query and results received for the question are as;

UTM Campaigns	UTM Sources
Cool T-shirt Search	Google
Getting To Know Cool T-shirts NYTimes	NYTimes
Interview With Cool T-shirts Founder	Medium
Ten Crazy Cool T-shirts Facts	Buzzfeed
Paid Search	Google
Weekly Newsletter	E-Mail
Retargeting Ad	Facebook
Retargeting Campaign	E-Mail

```
SELECT utm_campaign, utm_source
FROM page_visits;
```

Question 1, b

1b. What pages are on their website?

1. Landing Page
2. Shopping Cart
3. Checkout
4. Purchase

2. What is the User Journey?

Question 2, a

2a. How many first touches is each campaign responsible for?

Data set provides with the given result;

UTM Sources	UTM Campaigns	Page Views
Medium	interview-with-cool-tshirts-founder	622
NYTimes	getting-to-know-cool-tshirts	612
BuzzFeed	ten-crazy-cool-tshirts-facts	576
Google	cool-tshirts-search	169

Question 2, a

The query and results received for the question are as;

```
WITH first_touch AS (
    SELECT user_id, MIN(timestamp) AS first_touch_at
    FROM page_visits
    GROUP BY 1
),
ft_attr AS (
    SELECT ft.user_id, ft.first_touch_at, pv.utm_source, pv.utm_campaign
    FROM first_touch ft
    JOIN page_visits pv
    ON ft.user_id = pv.user_id
    AND ft.first_touch_at = pv.timestamp
)
SELECT ft_attr.utm_source, ft_attr.utm_campaign, COUNT(*)
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3
DESC;
```

Question 2, a

Data, according to a similar query for different pages, acquired according to user journey, results as;

UTM Source	UTM Campaigns	Page Views	Page Name
Medium	interview-with-cool-tshirts-founder	622	2 – Shopping Cart
NYTimes	getting-to-know-cool-tshirts	612	2 – Shopping Cart
BuzzFeed	ten-crazy-cool-tshirts-facts	576	2 – Shopping Cart
Google	cool-tshirts-search	169	2 – Shopping Cart

UTM Source	Campaign Name	Page Views	Page Name
E-Mail	Weekly Newsletter	450	3 – Checkout
Facebook	Retargeting Ad	445	3 – Checkout
E-Mail	Retargeting Campaign	246	3 – Checkout
Google	Paid Search	179	3 – Checkout
NYTimes	Getting To Know Cool T –Shirts	41	3 – Checkout
BuzzFeed	Ten Crazy Cool T-Shirts Facts	32	3 – Checkout
Medium	Interviews With Cool T-Shirts	21	3 – Checkout
Google	Cool T-Shirts Search	7	3 – Checkout

UTM Source	Campaign Name	Page Views	Page Name
E-Mail	Weekly Newsletter	115	4 – Purchase
Facebook	Retargeting Ad	113	4 – Purchase
E-Mail	Retargeting Campaign	54	4 – Purchase
Google	Paid Search	52	4 – Purchase
NYTimes	Getting To Know Cool T –Shirts	9	4 – Purchase
BuzzFeed	Ten Crazy Cool T-Shirts Facts	9	4 – Purchase
Medium	Interviews With Cool T-Shirts	7	4 – Purchase
Google	Cool T-Shirts Search	2	4 – Purchase

Question 2, a

The query describes the “first_touch” temporary table first, then describes the “ft_attr” which in it defines “first_touch” as “ft” and “page_visits” as “pv”, where it also joins both tables by claiming “ft.user_id” and “pv.user_id” as an equation to generate the combination on. Finally, it calls for the “ft_attr.utm_source” and “ft_attr.utm_campaign” columns, where also the integer count of the number of each user id these columns when put together.

```
WITH last_touch AS (
  SELECT user_id, MAX(timestamp) AS last_touch_at
  FROM page_visits
  GROUP BY 1
),
lt_attr AS (
  SELECT lt.user_id, lt.first_touch_at, pv.utm_source,
pv.utm_campaign
  FROM last_touch ft
  JOIN page_visits pv
  ON lt.user_id = pv.user_id
  AND lt.last_touch_at = pv.timestamp
)
SELECT lt_attr.utm_source, lt_attr.utm_campaign, COUNT(*)
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3
DESC;
```

Question 2, a

The query describes the “first_touch” temporary table first, then describes the “ft_attr” which in it defines “first_touch” as “ft” and “page_visits” as “pv”, where it also joins both tables by claiming “ft.user_id” and “pv.user_id” as an equation to generate the combination on. Finally, it calls for the “ft_attr.utm_source” and “ft_attr.utm_campaign” columns, where also the integer count of the number of each user id these columns when put together.

Question 2, b

2b. How many last touches is each campaign responsible for?

This equation is basically the same as the one before, which actually also just gives the “last touch”, instead of the “first_touch”.

```
WITH last_touch AS (
  SELECT user_id, MAX(timestamp) AS last_touch_at
  FROM page_visits
  GROUP
),
lt_attr AS (
  SELECT lt.user_id, lt.last_touch_at, pv.utm_source, pv.utm_campaign
  FROM last_touch ft
  JOIN page_visits pv
  ON lt.user_id = pv.user_id
  AND lt.last_touch_at = pv.timestamp
)
SELECT lt_attr.utm_source, lt_attr.utm_campaign, COUNT(*)
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3
DESC;
```

Question 2, c

2c. How many visitors make a purchase?

Page_name	COUNT(user_id)
4 - Purchase	361

```
SELECT DISTINCT page_name, COUNT(user_id)
FROM page_visits pv
WHERE pv.page_name = '4 – purchase'
GROUP BY 1
ORDER BY1;
```

Question 2, d

2d. How many last touches on the purchase page is each campaign responsible for?

UTM Sources	UTM Campaigns	Page Views
Email	weekly-newsletter	115
Facebook	Retargeting-ad	113
Email	Retargeting-campaign	54
Google	Paid-search	52
BuzzFeed	ten-crazy-cool-tshirts-facts	9
NYTimes	getting-to-know-cool-tshirts	9
Medium	interview-with-cool-tshirts-founder	7
Google	cool-tshirts-search	2

Question 2, d

2d. How many last touches is each campaign responsible for?

As last touches in this case are from the ‘purchase’ page, any touches are all last touches. There are no first_touches because users have to pass through the checkout page before arriving at ‘purchase’. Adding a WHERE clause to the previous query of “last touch”, and a COUNT(*) which would indicate the user_id with the relevant campaigns, it would look like the query beside.

```
WITH last_touch AS (
   SELECT user_id, MAX(timestamp) AS last_touch_at
   FROM page_visits pv
   WHERE pv.page_name = ‘4 – purchase’
   GROUP BY 1),
lt_attr AS (
   SELECT lt.user_id, lt.last_touch_at, pv.utm_source, pv.utm_campaign
   FROM last_touch ft
   JOIN page_visits pv
   ON lt.user_id = pv.user_id
   AND lt.last_touch_at = pv.timestamp
)
SELECT lt_attr.utm_source, lt_attr.utm_campaign, COUNT(*)
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3
DESC;
```

Question 2, e

2e. What is the typical user journey?

The typical user journey, though campaigns and sources differ, is a direct one on the web site, which can be seen through the number of visits of each journey step.

While users acquired by article marketing campaigns leave the journey after the shopping cart, those of retargeting, paid search, and especially those from the email marketing campaign “weekly newsletter” due to continuation of purchasing process, seem to be the right targeted segment.

It makes therefor sense to continue in those segments by reinvesting in the latter campaigns mentioned.

3. Optimising the Campaign Budget

Question 3, a

3a. CoolTShirts can reinvest in 5 campaigns. Which should they pick and why?

It makes sense to continue in those segments by reinvesting in Weekly Newsletter, Retargeting Ad, Retargeting Campaign, Paid Search. Lastly I would add the Interview With Cool T-shirts Founder article because it acquires the most users in comparison to the others.

It is possible to have a clearer idea with a user journey chart according to journey steps and page views.

Thank you,
Ant Erduman

1. Landing Page	Step 1	1979
	Step 2	2
	Step 3	11
	Step 4	5
	Step 5	3
2. Shopping Cart	Step 2	1879
	Step 3	3
	Step 4	11
	Step 5	4
	Step 6	3
3. Checkout	Step 3	1419
	Step 4	2
	Step 5	10
4. Purchase	Step 1	358
	Step 1	3