CODEACADEMY

LEARN SQL FROM SCRATCH

JOHN JANENDA | 2018.05.31

code|cademy

1. Get familiar with the company.

    2. 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`.

    3. What pages are on their website?

4. What is the user journey?

    5. How many first touches is each campaign responsible for?

    6. How many last touches is each campaign responsible for?

    7. How many visitors make a purchase?

    8. How many last touches *on the purchase page* is each campaign responsible for?

    9. What is the typical user journey?

10. Optimize the campaign budget

    11. CoolTShirts can re-invest in 5 campaigns. Which should they pick and why?

code|cademy

TABLE OF CONTENTS

- GET FAMILIAR WITH COOLTSHIRTS
- WHAT IS THE USER JOURNEY?
- OPTIMIZE THE CAMPAIGN BUDGET

codecademy

2. **How many campaigns and sources does CoolTShirts use? Which source is used for each campaign?**

- A source is the type of communication, like email, or site on which a campaign is shared with prospective customers, like google, Facebook, Medium, Buzzfeed or the New York Times

- A campaign is a specific message designed to elicit a specific response from a specific group of prospective customers.

- A source, like email, can run multiple campaigns, like 'retargeting' or 'weekly newsletter'.

- The same campaign could be run across multiple sources, although that is not represented in this dataset.

```sql
1   --1. Count campaigns and sources, then match
2   -- get COUNT of campaigns
3   SELECT COUNT(DISTINCT utm_campaign) AS 'Campaign Count'
4   FROM page_visits;
5
6   --get COUNT of sources
7   SELECT COUNT(DISTINCT utm_source) AS 'Source Count'
8   FROM page_visits;
9
10  --list sources per campaign to show relationship
11  SELECT DISTINCT utm_campaign as Campaigns,
12              utm_source as Sources
13  FROM page_visits;
```

Campaign Count	Source Count
8	6

Campaigns	Sources
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

GET FAMILIAR WITH COOLTSHIRTS – CONT'D

code|cademy

3. What pages are on the CoolTShirts website?

   - Find the distinct values of the page_name column.

   - Pages names shown are presumably a subset of actual pages on site and denote the specific page the campaign directed the prospective customer to upon clicking the link in the campaign. For instance, there are no specific 'product_pages' or 'department_pages'.

```
1   --2. find unique pages that are on website
2   SELECT DISTINCT page_name AS 'Page Names'
3   FROM page_visits;
```

Page Names
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

4. WHAT IS THE USER JOURNEY?

code|cademy

5. How many first touches is each campaign responsible for?

- By identifying the timestamps with the MIN date for each user, it is possible to determine which source and campaign first captured the prospects attention and prompted them to click through to the site.

- Interesting to note that out of 1,979 unique visitors only 4 campaigns of the 8 active were responsible for first touch.

```
1   -- 5. COUNT first touches per campaign
2   -- Create temp table that finds first touches by user id
3   WITH first_touch AS (
4       SELECT user_id,
5               MIN(timestamp) AS first_touch_at
6       FROM page_visits
7       GROUP BY user_id),
8   --Create 2nd temp table that adds source and
9   --campaign FROM page_visits and joins them on
10  --user id and timestamp
11  ft_attr AS (
12    SELECT ft.user_id,
13            ft.first_touch_at,
14            pv.utm_source,
15            pv.utm_campaign
16    FROM first_touch ft
17    JOIN page_visits pv
18      ON ft.user_id = pv.user_id
19      AND ft.first_touch_at = pv.timestamp
20  )
21  --SELECT and COUNT Rows WHERE first touch is
22  --associated with a campaign and source
23  SELECT ft_attr.utm_source AS Source,
24          ft_attr.utm_campaign AS Campaign,
25          COUNT(*) AS COUNT
26  FROM ft_attr
27  GROUP BY 1, 2
28  ORDER BY 3 DESC;
```

Source*	Campaign	Count
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

*Note that while prompt did not require addition of 'Source' column, the 'Hint' under the prompt suggested including it.

code|cademy

6. How many last touches is each campaign responsible for?

- By identifying the timestamps with the MAX date for each user, it is possible to determine which source and campaign was responsible for driving the prospects *last* visit to the site.

- Worth noting that the campaigns which drove the majority of first touches drop significantly down the list when examining last touches.

```
1   -- 6. COUNT last touches per campaign
2   -- Create temp table that finds LAST touches by user id
3   WITH last_touch AS (
4       SELECT user_id,
5           MAX(timestamp) AS last_touch_at
6       FROM page_visits
7       GROUP BY user_id),
8   --Create 2nd temp table that adds source and
9   --campaign FROM page_visits and joins them on
10  --user id and timestamp
11  ft_attr AS (
12      SELECT lt.user_id,
13          lt.last_touch_at,
14          pv.utm_source,
15          pv.utm_campaign
16      FROM last_touch lt
17      JOIN page_visits pv
18          ON lt.user_id = pv.user_id
19          AND lt.last_touch_at = pv.timestamp
20  )
21  --Select and count rows where first touch is
22  --associated with a campaign and source
23  SELECT ft_attr.utm_source AS Source,
24      ft_attr.utm_campaign AS Campaign,
25      COUNT(*) AS Count
26  FROM ft_attr
27  GROUP BY 1, 2
28  ORDER BY 3 DESC;
```

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

*Note that while prompt did not require addition of 'Source' column, the 'Hint' under the prompt suggested including it.

4. WHAT IS THE USER JOURNEY? – CONT'D

code_cademy

7. How many visitors make a purchase?

   - Conversion rate from visitor to 'customer' is a solid 18% (361 / 1,979 = 18%)

8. How many last touches *on the purchase page* is each campaign responsible for?

   - By isolating MAX timestamp where the page visited was the 'purchase page' we can identify the sources and campaigns that were responsible for conversion and ultimately revenue for the company.

   - Again, initial traffic generating campaigns continue to drop in effectiveness

```
1  -- 7. COUNT DISTINCT users who made a purchase
2  SELECT COUNT(DISTINCT user_id) AS 'Customers that Purchase'
3  FROM page_visits
4  WHERE page_name = '4 - purchase';
```

```
1   -- 8. COUNT last touches per campaign that led to a purchase
2   -- Create temp table that finds LAST touches by user id
3   WITH last_touch AS (
4       SELECT user_id,
5           MAX(timestamp) AS last_touch_at
6       FROM page_visits
7   --add WHERE clause
8     WHERE page_name = '4 - purchase'
9       GROUP BY user_id),
10  --Create 2nd temp table that adds source and
11  --campaign FROM page_visits and joins them on
12  --user id and timestamp
13  ft_attr AS (
14    SELECT lt.user_id,
15          lt.last_touch_at,
16          pv.utm_source,
17          pv.utm_campaign
18    FROM last_touch lt
19    JOIN page_visits pv
20      ON lt.user_id = pv.user_id
21      AND lt.last_touch_at = pv.timestamp
22  )
23  --SELECT and COUNT Rows WHERE first touch is
24  --associated with a campaign and source
25  SELECT ft_attr.utm_source AS Source,
26         ft_attr.utm_campaign AS Campaign,
27         COUNT(*) AS COUNT
28  FROM ft_attr
29  GROUP BY 1, 2
30  ORDER BY 3 DESC;
```

Customers that Purchase		
361		

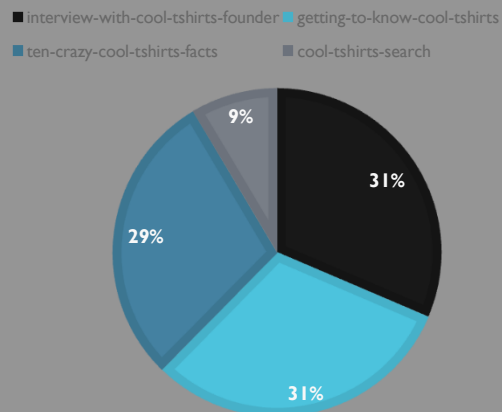| Source*	Campaign	Count
email	weekly-newsletter	115
facebook	retargetting-ad	113
email	retargetting-campaign	54
google	paid-search	52
buzzfeed	ten-crazy-cool-tshirts-facts	9
nytimes	getting-to-know-cool-tshirts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

*Note that while prompt did not require addition of 'Source' column, the 'Hint' under the prompt suggested including it.
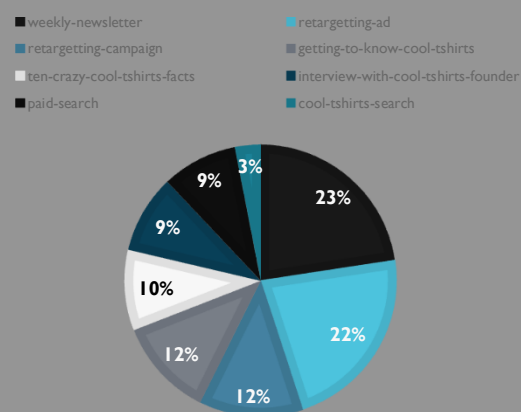
9. WHAT IS THE 'TYPICAL' USER JOURNEY?

code|cademy

- ■ 'Story' campaigns draw users to site
  - ■ 91% of users first come to the site via one of these 'story' campaigns:
    - ■ Interview with CoolTShirts Founder
    - ■ Getting to Know CoolTShirts
    - ■ Ten Crazy CoolTShirts Facts

- ■ But, users need additional reminders to stay engaged
  - ■ 31% of users that came to site via a 'story' campaign don't return and 'story' campaigns only account for 6% of purchases
  - ■ The majority of prospects require a 2nd campaign to become purchasing customers.

- ■ And those reminders generate most of the revenue
  - ■ 78% of purchases were generated via email newsletter, email retargeting, or Facebook retargeting ads.
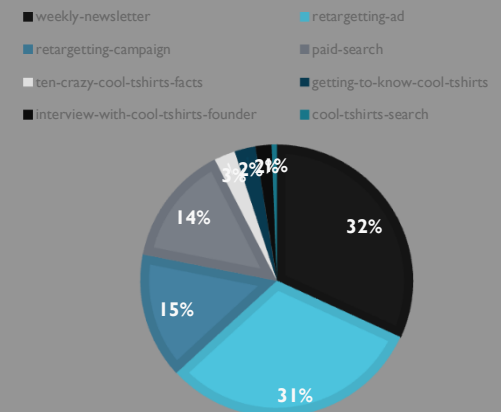
FIRST TOUCHES BY CAMPAIGN

- ■ interview-with-cool-tshirts-founder
- ■ getting-to-know-cool-tshirts
- ■ ten-crazy-cool-tshirts-facts
- ■ cool-tshirts-search

9%
31%
29%
31%

LAST TOUCHES BY CAMPAIGN

- ■ weekly-newsletter
- ■ retargetting-ad
- ■ retargetting-campaign
- ■ getting-to-know-cool-tshirts
- □ ten-crazy-cool-tshirts-facts
- ■ interview-with-cool-tshirts-founder
- ■ paid-search
- ■ cool-tshirts-search

3%
23%
9%
9%
22%
10%
12%
12%

PURCHASES BY CAMPAIGN

- ■ weekly-newsletter
- ■ retargetting-ad
- ■ retargetting-campaign
- ■ paid-search
- □ ten-crazy-cool-tshirts-facts
- ■ getting-to-know-cool-tshirts
- ■ interview-with-cool-tshirts-founder
- ■ cool-tshirts-search

3% 2% 1%
32%
14%
15%
31%

9. WHAT IS THE 'TYPICAL' USER JOURNEY? CONT'D

code|cademy

FIRST TOUCHES BY CAMPAIGN

91%
- interview-with-cool-tshirts-founder
- getting-to-know-cool-tshirts
- ten-crazy-cool-tshirts-facts
- cool-tshirts-search

(Pie chart: 31%, 31%, 29%, 9%)

LAST TOUCHES BY CAMPAIGN

57%
- weekly-newsletter
- retargetting-ad
- retargetting-campaign

31%
- getting-to-know-cool-tshirts
- ten-crazy-cool-tshirts-facts
- interview-with-cool-tshirts-founder
- paid-search
- cool-tshirts-search

(Pie chart: 23%, 22%, 12%, 12%, 10%, 9%, 9%, 3%)

PURCHASES BY CAMPAIGN

78%
- weekly-newsletter
- retargetting-ad
- retargetting-campaign
- paid-search

6%
- ten-crazy-cool-tshirts-facts
- getting-to-know-cool-tshirts
- interview-with-cool-tshirts-founder
- cool-tshirts-search

(Pie chart: 32%, 31%, 15%, 14%, 2%, 2%, 2%)

Story Campaigns

Retargeting Campaigns

JOHN JANENDA | LEARN SQL FROM SCRATCH

OPTIMIZING THE CAMPAIGN BUDGET
Which 5 campaigns should CoolTShirts re-invest in?

code|cademy

- Limiting re-investment to 5 campaigns will likely leave money on the table as there are at least 6 high performing campaigns.

- If CoolTShirts can only invest in 5, data recommends that all 3 'story' campaigns highlighted in blue are exceptional at generating interest, they should be continued

- In addition, retargeting is critical, so the two remaining campaigns should be based on a retargeting strategy.

  - Perhaps slightly counter-intuitive, but suggest that weekly newsletter be discontinued, despite it's success, under the assumption that all 'touches' that remind customer of company 'may' be equal

  - Suggest continuing the email retargeting campaign and place a premium on email capture as part of journey from story campaign to site as email newsletters and retargeting campaigns account for 47% of sales. Having customer email address is extremely important to conversion. Would have been interesting to see how many email addresses were in the database as compared to visitors.

  - Further suggest extending retargeting ad campaigns as they have the potential to reach 100% of prospects that interact with site, INCLUDING those that do not provide an email address. Would strongly consider purchasing retargeting ads on additional sites like NYT, Buzzfeed and Medium, if possible, as they generate excellent traffic from 'story' campaigns.

Campaign	1st Touches	% 1st Touches	Last Touches	% Last Touches	Purchases	% Purchases
interview-with-cool-tshirts-founder	622	31%	184	9%	7	2%
getting-to-know-cool-tshirts	612	31%	232	12%	9	2%
ten-crazy-cool-tshirts-facts	576	29%	190	10%	9	2%
cool-tshirts-search	169	9%	60	3%	2	1%
weekly-newsletter	0	0%	447	23%	115	32%
retargetting-ad	0	0%	443	22%	113	31%
retargetting-campaign	0	0%	245	12%	54	15%
paid-search	0	0%	178	9%	52	14%

OPTIMIZING THE CAMPAIGN BUDGET
Rationale for discontinued campaigns

codecademy

- A note on the discontinued campaigns

  - As noted previously regarding weekly-newsletter, would prefer to keep it active, but assuming that possibly *any* appearance in customer's inbox will trigger a response and since a newsletter costs 'something' to produce, perhaps that money can be spent on increasing retargeting ads to additional sites beyond Facebook (see note on 'Extra's' page regarding traffic by source.

  - Organic search, presumably generated via SEO activities generates 9% of traffic – not insignificant, but if budget needs to be cut, this is a less effective means of reaching prospects for initial capture, presumably because brand recognition for CoolTShirts is relatively low and there are likely hundreds of competitors in the t-shirt space.

  - Paid search accounts for 14% of actual purchases, so it is valuable, but only after customers have been introduced to site via 'story' campaigns as it does not appear in initial visit data. If something needs to be cut, this is a candidate.

Campaign	1st Touches	% 1st Touches	Last Touches	% Last Touches	Purchases	% Purchases
interview-with-cool-tshirts-founder	622	31%	184	9%	7	2%
getting-to-know-cool-tshirts	612	31%	232	12%	9	2%
ten-crazy-cool-tshirts-facts	576	29%	190	10%	9	2%
cool-tshirts-search	169	9%	60	3%	2	1%
weekly-newsletter	0	0%	447	23%	115	32%
retargetting-ad	0	0%	443	22%	113	31%
retargetting-campaign	0	0%	245	12%	54	15%
paid-search	0	0%	178	9%	52	14%

EXTRAS

code|cademy

- Retargeting ads should be considered for New York Times, Buzzfeed and Medium as they all appear better at generating traffic than Facebook or Google

```
1   --Extra: determine if a specific source is better than others for
    generating traffic
2   SELECT utm_source as Source,
3       COUNT(DISTINCT user_id) as 'Visit Count'
4   FROM page_visits
5   GROUP BY 1
6   ORDER BY 2 DESC;
```

Source	Visit Count
nytimes	747
email	696
buzzfeed	648
medium	625
facebook	445
google	339