

Capstone: Attribution

Learn SQL from Scratch
Jacob Hoberg
10 July 2018



Project Goals

- Get Familiar
 - How many campaigns do CoolTShirts use and how are they related?
 - What pages are on their website?
- 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 a typical user journey?
- Final Insights
 - Which 5 campaigns should CoolTShirts re-invest in?

Get Familiar



How many campaigns do CoolTShirts use and how are they related?

- There are 8 distinct advertising campaigns (utm_campaign)
- There are 6 distinct internet sources from which users visited (utm_source)
- There are no campaigns with multiple sources, however there are 2 sources with multiple campaigns
 - email gets weekly-newsletter and retargeting-campaign
 - Google gets paid-search and cool-tshirts-search

Query Results	
COUNT(DISTINCT utm_campaign)	
8	
COUNT(DISTINCT utm_source)	
6	
utm_campaign	utm_source
ten-crazy-cool-tshirts-facts	buzzfeed
weekly-newsletter	email
retargeting-campaign	email
retargeting-ad	facebook
paid-search	google
cool-tshirts-search	google
interview-with-cool-tshirts-founder	medium
getting-to-know-cool-tshirts	nytimes

What pages are on their website?

- There are 4 web pages
 - 1 - landing_page
 - 2 - shopping_cart
 - 3 - checkout
 - 4 - purchase

```
test.sqlite
19 SELECT DISTINCT page_name
20 FROM page_visits;
21
22
```

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

User Journey



How many first touches is each campaign responsible for?

- There are 1,979 user visits associated with an advertising campaign
- There are 4 advertising campaigns that are responsible the initial user visits
- The top 3 advertising campaigns contribute nearly equally
- cool-tshirts-search is the least effective campaign for bringing users to the website



test.sqlite



```
18
19 WITH first_touch AS (
20     SELECT user_id,
21            MIN(timestamp) AS first_touch_at
22     FROM page_visits
23     GROUP BY user_id)
24 SELECT pv.utm_campaign,
25        COUNT(*) AS first_touch_count
26 FROM first_touch ft
27 JOIN page_visits pv
28     ON ft.user_id = pv.user_id
29     AND ft.first_touch_at = pv.timestamp
30 GROUP BY pv.utm_campaign
31 ORDER BY 2 DESC;
```

Query Results

utm_campaign	first_touch_count
interview-with-cool-tshirts-founder	622
getting-to-know-cool-tshirts	612
ten-crazy-cool-tshirts-facts	576
cool-tshirts-search	169

Database Schema

page_visits		5692 rows
page_name		TEXT
timestamp		TEXT
user_id		INTEGER
utm_campaign		TEXT
utm_source		TEXT

How many last touches is each campaign responsible for?

- There are 1,979 user visits associated with an advertising campaign
- All 8 advertising campaigns are responsible the final user 'touch' of the website
- Nearly half of all final interactions are a result of either weekly-newsletter or retargeting-ad
- cool-tshirts-search is the least effective campaign for the last users interaction

test.sqlite

```
33 WITH last_touch AS (  
34     SELECT user_id,  
35            MAX(timestamp) AS last_touch_at  
36     FROM page_visits  
37     GROUP BY user_id)  
38 SELECT pv.utm_campaign,  
39        COUNT(*) AS last_touch_count  
40 FROM last_touch lt  
41 JOIN page_visits pv  
42     ON lt.user_id = pv.user_id  
43     AND lt.last_touch_at = pv.timestamp  
44 GROUP BY pv.utm_campaign  
45 ORDER BY 2 DESC;
```

Query Results

utm_campaign	last_touch_count
weekly-newsletter	447
retargeting-ad	443
retargeting-campaign	245
getting-to-know-cool-tshirts	232
ten-crazy-cool-tshirts-facts	190
interview-with-cool-tshirts-founder	184
paid-search	178
cool-tshirts-search	60

Database Schema

page_visits		5692 rows
page_name	TEXT	
timestamp	TEXT	
user_id	INTEGER	
utm_campaign	TEXT	
utm_source	TEXT	

How many visitors make a purchase?

- There are 2,000 user visits to the landing_page
- 95% of initial visitors add an item to their shopping_cart
- 75.3% of visitors with an item in their cart proceed to the checkout page
- Only 25.2% of users proceed from the checkout page to actual purchase
 - This means that only 18% of initial visitors make a purchase

```
56 SELECT page_name,  
57        COUNT(*)  
58 FROM page_visits  
59 GROUP BY page_name;
```

Query Results

page_name	COUNT(*)
1 - landing_page	2000
2 - shopping_cart	1900
3 - checkout	1431
4 - purchase	361

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

- Of the 2,000 initial visits, 361 users make a purchase
- All 8 campaigns lead to some amount of user purchase
- Over 63% of last touches on the purchase page can be attributed to 2 advertising campaigns
 - weekly-newsletter = 31.9% (115 purchases)
 - retargeting-ad = 31.3% (113 purchases)

```
testsqlite
68 WITH last_touch AS (
69     SELECT user_id,
70            MAX(timestamp) AS last_touch_at
71     FROM page_visits
72     WHERE page_name = '4 - purchase'
73     GROUP BY user_id)
74 SELECT pv.utm_campaign,
75        COUNT(*) AS last_touch_count_at_purchase
76 FROM last_touch lt
77 JOIN page_visits pv
78     ON lt.user_id = pv.user_id
79     AND lt.last_touch_at = pv.timestamp
80 GROUP BY pv.utm_campaign
81 ORDER BY 2 DESC;
82
83
```

Query Results	
utm_campaign	last_touch_count_at_purchase
weekly-newsletter	115
retargeting-ad	113
retargeting-campaign	54
paid-search	52
getting-to-know-cool-tshirts	9
ten-crazy-cool-tshirts-facts	9
interview-with-cool-tshirts-founder	7
cool-tshirts-search	2
Database Schema	
page_visits	
	5692 rows
page_name	TEXT
timestamp	TEXT
user_id	INTEGER
utm_campaign	TEXT
utm_source	TEXT

What is a typical user journey?

- We've established that there have been 2,000 user visits to the landing_page
- 1,979 of those visits are associated with a campaign
- The most popular 'first touch' campaign is "interview-with-cool-tshirts-founder"
- Starting with that information, we can see which campaigns lead to the second page of the site
- Of the 622 'first touch' visits associated with "interview-with-cool-tshirts-founder"
 - Only 609 users proceeded to page 2 - shopping_cart
 - 515 of those users did so through the "interview-with-cool-tshirts-founder" campaign

```
64
65 WITH popular_first_touch_users AS (
66     SELECT user_id
67     FROM page_visits
68     WHERE page_name = '1 - landing_page'
69         AND utm_campaign = 'interview-with-cool-tshirts-founder')
70 SELECT pv.utm_campaign,
71        COUNT(*)
72 FROM popular_first_touch_users pftu
73 JOIN page_visits pv
74     ON pftu.user_id = pv.user_id
75 WHERE page_name = '2 - shopping_cart'
76 GROUP BY pv.utm_campaign
77 ORDER BY 2 DESC;
78
```

Query Results	
pv.utm_campaign	COUNT(*)
interview-with-cool-tshirts-founder	515
getting-to-know-cool-tshirts	60
ten-crazy-cool-tshirts-facts	33
cool-tshirts-search	1
Database Schema	
page_visits	
5692 rows	
page_name	TEXT
timestamp	TEXT
user_id	INTEGER
utm_campaign	TEXT
utm_source	TEXT

What is a typical user journey?

- Of those 515 users
 - 384 users visited page 3 - checkout
 - The most popular campaign for page 3 is “retargetting-ad”

```
test.sqlite
64
65 WITH popular_first_touch_users AS (
66     SELECT user_id
67     FROM page_visits
68     WHERE page_name = '1 - landing_page'
69         AND utm_campaign = 'interview-with-cool-tshirts-founder'),
70     popular_second_touch_users AS (
71     SELECT user_id
72     FROM page_visits
73     WHERE page_name = '2 - shopping_cart'
74         AND utm_campaign = 'interview-with-cool-tshirts-founder')
75 SELECT utm_campaign,
76        COUNT(*)
77 FROM popular_first_touch_users pftu
78 JOIN popular_second_touch_users pstu
79     ON pftu.user_id = pstu.user_id
80 JOIN page_visits pv
81     ON pftu.user_id = pv.user_id
82 WHERE page_name = '3 - checkout'
83 GROUP BY utm_campaign
84 ORDER BY 2 DESC;
85
```

Query Results	
utm_campaign	COUNT(*)
retargetting-ad	129
weekly-newsletter	111
retargetting-campaign	59
paid-search	54
interview-with-cool-tshirts-founder	31
Database Schema	
page_visits	
5692 rows	
page_name	TEXT
timestamp	TEXT
user_id	INTEGER
utm_campaign	TEXT
utm_source	TEXT

What is a typical user journey?

- Of those 119 users
 - Only 27 made a purchase
 - All 27 were associated with the “retargetting-ad” campaign

```
64
65 WITH popular_first_touch_users AS (
66   SELECT user_id
67   FROM page_visits
68   WHERE page_name = '1 - landing_page'
69   AND utm_campaign = 'interview-with-cool-tshirts-founder'),
70 popular_second_touch_users AS (
71   SELECT user_id
72   FROM page_visits
73   WHERE page_name = '2 - shopping_cart'
74   AND utm_campaign = 'interview-with-cool-tshirts-founder'),
75 popular_third_touch_users AS (
76   SELECT user_id
77   FROM page_visits
78   WHERE page_name = '3 - checkout'
79   AND utm_campaign = 'retargetting-ad')
80 SELECT utm_campaign,
81        COUNT(*)
82 FROM popular_first_touch_users pftu
83 JOIN popular_second_touch_users pstu
84   ON pftu.user_id = pstu.user_id
85 JOIN popular_third_touch_users ptu
86   ON pftu.user_id = ptu.user_id
87 JOIN page_visits pv
88   ON pftu.user_id = pv.user_id
89 WHERE page_name = '4 - purchase'
90 GROUP BY utm_campaign
91 ORDER BY 2 DESC;
92
```

Query Results	
utm_campaign	COUNT(*)
retargetting-ad	27
Database Schema	
page_visits	
page_name	TEXT
timestamp	TEXT
user_id	INTEGER
utm_campaign	TEXT
utm_source	TEXT



What is a typical user journey?

- Starting at the most typical 'first touch' visit, the typical user journey is:
 - Page '1 - landing_page' through "interview-with-cool-tshirts-founder" campaign
 - Page '2 - shopping_cart' through "interview-with-cool-tshirts-founder" campaign
 - Page '3 - checkout' through "retargeting-ad" campaign
 - Page '4 - purchase' through "retargeting-ad" campaign

What is a typical user journey?

- The same process can be reversed to identify the journey most likely to lead to a purchase
- The most common campaign associated with a purchase is “weekly-newsletter” (115 users)
 - Of those users, all of them reached page ‘3 - checkout’ through the same campaign

test.sqlite

```
93 WITH popular_last_touch_users AS (  
94     SELECT user_id  
95     FROM page_visits  
96     WHERE page_name = '4 - purchase'  
97         AND utm_campaign = 'weekly-newsletter')  
98 SELECT utm_campaign,  
99     COUNT(*)  
100 FROM popular_last_touch_users pltu  
101 JOIN page_visits pv  
102     ON pltu.user_id = pv.user_id  
103 WHERE page_name = '3 - checkout'  
104 GROUP BY utm_campaign  
105 ORDER BY 2 DESC;
```

Query Results

utm_campaign	COUNT(*)
weekly-newsletter	115

Database Schema

page_visits		5692 rows
page_name	TEXT	
timestamp	TEXT	
user_id	INTEGER	
utm_campaign	TEXT	
utm_source	TEXT	

What is a typical user journey?

- Of those 115 users, the most popular campaign (42 users) reached page '2 - shopping_cart' through the "ten-crazy-cool-tshirts-facts" campaign

```
testsqlite
93 WITH popular_last_touch_users AS (
94     SELECT user_id
95     FROM page_visits
96     WHERE page_name = '4 - purchase'
97         AND utm_campaign = 'weekly-newsletter'),
98     popular_third_touch_users AS (
99     SELECT user_id
100    FROM page_visits
101    WHERE page_name = '3 - checkout'
102        AND utm_campaign = 'weekly-newsletter')
103 SELECT utm_campaign,
104        COUNT(*)
105 FROM popular_last_touch_users pltu
106 JOIN popular_third_touch_users ptu
107     ON pltu.user_id = ptu.user_id
108 JOIN page_visits pv
109     ON pltu.user_id = pv.user_id
110 WHERE page_name = '2 - shopping_cart'
111 GROUP BY utm_campaign
112 ORDER BY 2 DESC;
```

Query Results	
utm_campaign	COUNT(*)
ten-crazy-cool-tshirts-facts	42
getting-to-know-cool-tshirts	35
interview-with-cool-tshirts-founder	29
cool-tshirts-search	11
Database Schema	
page_visits	
5692 rows	
page_name	TEXT
timestamp	TEXT
user_id	INTEGER
utm_campaign	TEXT
utm_source	TEXT

What is a typical user journey?

- Of those 42 users, almost all (37) had their 'first touch' of the website through the "ten-crazy-cool-tshirts-facts" campaign.

```
testsqlite
93 WITH popular_last_touch_users AS (
94   SELECT user_id
95   FROM page_visits
96   WHERE page_name = '4 - purchase'
97   AND utm_campaign = 'weekly-newsletter'),
98   popular_third_touch_users AS (
99   SELECT user_id
100  FROM page_visits
101  WHERE page_name = '3 - checkout'
102  AND utm_campaign = 'weekly-newsletter'),
103   popular_second_touch_users AS (
104   SELECT user_id
105   FROM page_visits
106   WHERE page_name = '2 - shopping_cart'
107   AND utm_campaign = 'ten-crazy-cool-tshirts-facts')
108 SELECT utm_campaign,
109        COUNT(*)
110 FROM popular_last_touch_users pltu
111 JOIN popular_third_touch_users pttu
112   ON pltu.user_id = pttu.user_id
113 JOIN popular_second_touch_users pstu
114   ON pltu.user_id = pstu.user_id
115 JOIN page_visits pv
116   ON pltu.user_id = pv.user_id
117 WHERE page_name = '1 - landing_page'
118 GROUP BY utm_campaign
119 ORDER BY 2 DESC;
```

Query Results	
utm_campaign	COUNT(*)
ten-crazy-cool-tshirts-facts	37
interview-with-cool-tshirts-founder	4
getting-to-know-cool-tshirts	3
cool-tshirts-search	1
Database Schema	
page_visits 5692 rows	
page_name	TEXT
timestamp	TEXT
user_id	INTEGER
utm_campaign	TEXT
utm_source	TEXT



What is a typical user journey?

- This means that the user journey most likely to lead to a purchase is:
 - Page '1 - landing_page' through "ten-crazy-cool-tshirts-facts" campaign
 - Page '2 - shopping_cart' through "ten-crazy-cool-tshirts-facts" campaign
 - Page '3 - checkout' through "weekly-newsletter" campaign
 - Page '4 - purchase' through "weekly-newsletter" campaign

What is a typical user journey?

- Each of the 4 campaigns responsible for 'first touch' can also be evaluated for what percentage of purchases result.
- For "interview-with-cool-tshirts-founder"
 - 622 'first touch' users
 - 119 purchases
 - ~19.1% purchase rate

```
test.sqlite
121 WITH popular_first_touch_users AS (
122     SELECT user_id
123     FROM page_visits
124     WHERE page_name = '1 - landing_page'
125         AND utm_campaign = 'interview-with-cool-tshirts-founder')
126 SELECT COUNT(*)
127 FROM popular_first_touch_users pftu
128 JOIN page_visits pv
129     ON pftu.user_id = pv.user_id
130 WHERE page_name = '4 - purchase';
```

Query Results

COUNT(*)

119

Database Schema

page_visits		5692 rows
page_name	TEXT	
timestamp	TEXT	
user_id	INTEGER	
utm_campaign	TEXT	
utm_source	TEXT	

What is a typical user journey?

- For “getting-to-know-cool-tshirts”
 - 612 ‘first touch’ users
 - 106 purchases
 - ~17.3% purchase rate

```
test.sqlite
121 WITH popular_first_touch_users AS (
122     SELECT user_id
123     FROM page_visits
124     WHERE page_name = '1 - landing_page'
125         AND utm_campaign = 'getting-to-know-cool-tshirts')
126 SELECT COUNT(*)
127 FROM popular_first_touch_users pftu
128 JOIN page_visits pv
129     ON pftu.user_id = pv.user_id
130 WHERE page_name = '4 - purchase';
```

Query Results		
COUNT(*)		
106		
Database Schema		
page_visits		5692 rows
page_name	TEXT	
timestamp	TEXT	
user_id	INTEGER	
utm_campaign	TEXT	
utm_source	TEXT	

What is a typical user journey?

- For “ten-crazy-cool-tshirts-facts”
 - 576 ‘first touch’ users
 - 111 purchases
 - ~19.3% purchase rate

```
test.sqlite
121 WITH popular_first_touch_users AS (
122     SELECT user_id
123     FROM page_visits
124     WHERE page_name = '1 - landing_page'
125           AND utm_campaign = 'ten-crazy-cool-tshirts-facts')
126 SELECT COUNT(*)
127 FROM popular_first_touch_users pftu
128 JOIN page_visits pv
129     ON pftu.user_id = pv.user_id
130 WHERE page_name = '4 - purchase';
131
```

Query Results		
COUNT(*)		
111		
Database Schema		
page_visits		5692 rows
page_name		TEXT
timestamp		TEXT
user_id		INTEGER
utm_campaign		TEXT
utm_source		TEXT

What is a typical user journey?

- For “cool-tshirts-search”
 - 169 ‘first touch’ users
 - 31 purchases
 - ~18.3% purchase rate

```
test.sqlite
121 WITH popular_first_touch_users AS (
122     SELECT user_id
123     FROM page_visits
124     WHERE page_name = '1 - landing_page'
125           AND utm_campaign = 'cool-tshirts-search')
126 SELECT COUNT(*)
127 FROM popular_first_touch_users pftu
128 JOIN page_visits pv
129     ON pftu.user_id = pv.user_id
130 WHERE page_name = '4 - purchase';
131
```

Query Results		
COUNT(*)		
31		
Database Schema		
page_visits		5692 rows
page_name	TEXT	
timestamp	TEXT	
user_id	INTEGER	
utm_campaign	TEXT	
utm_source	TEXT	

Final Insights





Which 5 campaigns should CoolTShirts re-invest in?

- We've established that CoolTShirts utilizes 8 advertising campaigns
 - ten-crazy-cool-tshirt-facts via buzzfeed
 - weekly-newsletter via email
 - retargeting-campaign via email
 - retargeting-ad via facebook
 - paid-search via google
 - cool-tshirts-search via google
 - interview-with-cool-tshirts-founder via medium
 - getting-to-know-cool-tshirts via nytimes



Which 5 campaigns should CoolTShirts re-invest in?

- The large majority (91.5%) of first touches are associated with only 3 campaigns
 - interview-with-cool-tshirts-founder (31.4%)
 - getting-to-know-cool-tshirts (30.9%)
 - ten-crazy-cool-tshirt-facts (29.1%)
- Each of these can be associated with a purchase rate
 - interview-with-cool-tshirts-founder (19.1%)
 - getting-to-know-cool-tshirts (17.3%)
 - ten-crazy-cool-tshirt-facts (19.3%)
- The large majority (92.5%) of last touches at purchase are associated with 4 campaigns
 - weekly-newsletter (31.9%)
 - retargeting-ad (31.3%)
 - retargeting-campaign (15.0%)
 - paid-search (14.4%)
- Additionally, we've established that the most likely user journey to lead to a purchase is:
 - ten-crazy-cool-tshirt-facts -> weekly-newsletter
- Also note that nearly half (46.9%) of all purchases are associated with email campaigns



Which 5 campaigns should CoolTShirts re-invest in?

- Given the preceding information, reinvestment in the following 5 campaigns are recommended:

- *Note that campaigns are listed according to their associated purchase rate.

1. weekly-newsletter (31.9%)
2. retargetting-ad (31.3%)
3. interview-with-cool-tshirts-founder (19.1%)
4. ten-crazy-cool-tshirt-facts (19.3%)
5. getting-to-know-cool-tshirts (17.3%)

THANK YOU

Learn SQL from Scratch

Jacob Hoberg

10 July 2018

