code cademy

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

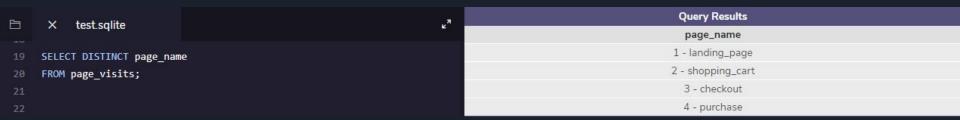
How many campaigns do CoolTShirts use and how are they related?

- 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
 - o email gets weekly-newsletter and retargeting-campaign
 - Google gets paid-search and cool-tshirts-search

n)
)
utm_source
buzzfeed
email
email
facebook
google
google
medium
nytimes

What pages are on their website?

- There are 4 web pages
 - o 1 landing_page
 - 2 shopping_cart
 - o 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

	X test.sqlite 27	Query Results	
	7 Cocoque	utm_campaign	first_touch_count
18		interview-with-cool-tshirts-founder	622
19	WITH first_touch AS (getting-to-know-cool-tshirts	612
	SELECT user_id,	ten-crazy-cool-tshirts-facts	576
21	MIN(timestamp) AS first_touch_at	cool-tshirts-search	169
22	FROM page_visits	Database Schema	
23	GROUP BY user_id)	page_visits	5692 rows
24	SELECT pv.utm_campaign,	page_name	TEXT
25	COUNT(*) AS first_touch_count	timestamp	TEXT
	FROM first_touch ft	user_id	INTEGER
27	JOIN page_visits pv	utm_campaign	TEXT
28	ON ft.user_id = pv.user_id	utm_source	TEXT
29	AND ft.first_touch_at = pv.timestamp	7	
	GROUP BY pv.utm_campaign		
31	ORDER BY 2 DESC;		

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 retargetting-ad
- cool-tshirts-search is the least effective campaign for the last users interaction

	X test.sqlite 2	Query Results	
	7 testaque	utm_campaign	last_touch_count
33	WITH last_touch AS (weekly-newsletter	447
34	SELECT user_id,	retargetting-ad	443
	MAX(timestamp) AS last_touch_at	retargetting-campaign	245
	FROM page_visits	getting-to-know-cool-tshirts	232
37	GROUP BY user_id)	ten-crazy-cool-tshirts-facts	190
	SELECT pv.utm_campaign,	interview-with-cool-tshirts-founder	184
	COUNT(*) AS last_touch_count	paid-search	178
	FROM last_touch lt	cool-tshirts-search	60
41	JOIN page_visits pv	Database Schem	aa
42	ON lt.user_id = pv.user_id	page_visits	5692 rows
43	AND lt.last_touch_at = pv.timestamp	page_name	TEXT
44	GROUP BY pv.utm_campaign	timestamp	TEXT
	ORDER BY 2 DESC;		
		user_id	INTEGER
47		utm_campaign	TEXT
42		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

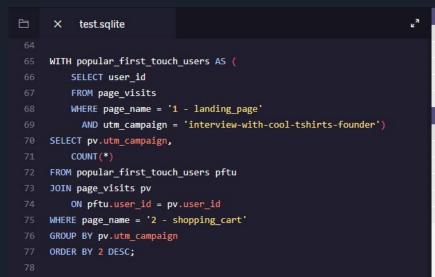
- ·	□ × test.sqlite	Query Res	sults
		page_name	COUNT(*)
56	SELECT page_name,	1 - landing_page	2000
57	COUNT(*)	2 - shopping_cart	1900
58	FROM page_visits	3 - checkout	1431
59	GROUP BY page_name;	4 - purchase	361

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)
 - o retargeting-ad = 31.3% (113 purchases)

	× test.sqlite	Query Resu	ılts
	* test-squite	utm_campaign	last_touch_count_at_purchase
	WITH last_touch AS (weekly-newsletter	115
	SELECT user_id,	retargetting-ad	113
	MAX(timestamp) AS last_touch_at	retargetting-campaign	54
71	FROM page_visits	paid-search	52
72	WHERE page_name = '4 - purchase'	getting-to-know-cool-tshirts	9
	GROUP BY user_id)	ten-crazy-cool-tshirts-facts	9
	SELECT pv.utm_campaign,	interview-with-cool-tshirts-founder	7
	COUNT(*) AS last_touch_count_at_purchase	cool-tshirts-search	2
	FROM last_touch lt	Database Sch	nema
77	JOIN page_visits pv	page_visit	
	ON lt.user_id = pv.user_id		
79	AND lt.last_touch_at = pv.timestamp	page_name	TEXT
80	GROUP BY pv.utm campaign	timestamp	TEXT
81	ORDER BY 2 DESC;	user_id	INTEGER
82		utm_campaign	TEXT
23		utm_source	TEXT

- 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



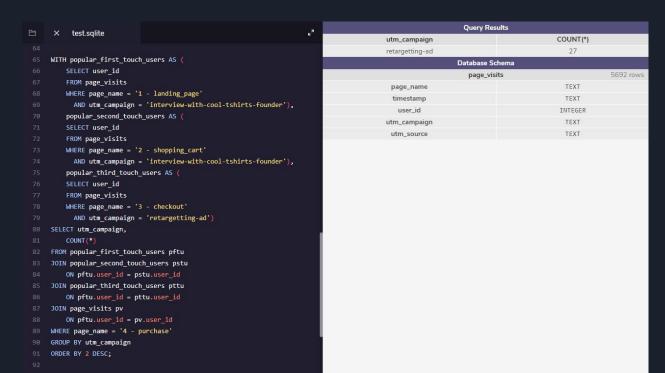
Query Results		
pv.utm_campaign	COUNT(*)	
interview-with-cool-tshirts-founder	515	
getting-to-know-cool-tshirts	60	
ten-crazy-cool-tshirts-facts	33	
cool-tshirts-search	1	

Database Schema			
page_vis	sits 5692 rows		
page_name	TEXT		
timestamp	TEXT		
user_id	INTEGER		
utm_campaign	TEXT		
utm_source	TEXT		

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

	× test.sqlite	, s	Query Results	
	A testsquite		utm_campaign	COUNT(*)
		BWW.	retargetting-ad	129
	WITH popular_first_touch	_users AS (weekly-newsletter	111
	SELECT user_id		retargetting-campaign	59
	FROM page_visits		paid-search	54
	WHERE page_name = '1		interview-with-cool-tshirts-founder	31
		'interview-with-cool-tshirts-founder'),	Database Schem	a
	popular_second_touch	_users AS (page_visits	5692 rows
	SELECT user_id		page_name	TEXT
72	FROM page_visits		timestamp	TEXT
	WHERE page_name = '2	- shopping_cart'	user_id	INTEGER
	AND utm_campaign =	'interview-with-cool-tshirts-founder')		
	SELECT utm_campaign,		utm_campaign	TEXT
	COUNT(*)		utm_source	TEXT
	FROM popular_first_touch	_users pftu		
	JOIN popular_second_touck	h_users pstu		
	ON pftu.user_id = pst	tu.user_id		
	JOIN page_visits pv			
	ON pftu.user_id = pv	.user_id		
82	WHERE page_name = '3 - cl	heckout'		
	GROUP BY utm_campaign			
	ORDER BY 2 DESC;			

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



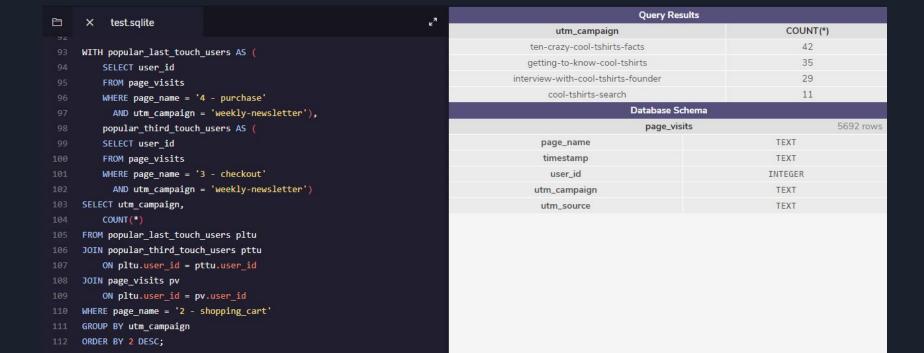
- Starting at the most typical 'first touch' visit, the typical user journey is:
 - Page '1 landing_page' through "interview-with-cool-tshirts-founder" campaign
 - Page '2 shopping_cart' through "interview-with-cool-tshirts-founder" campaign
 - Page '3 checkout' through "retargetting-ad" campaign
 -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

- 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

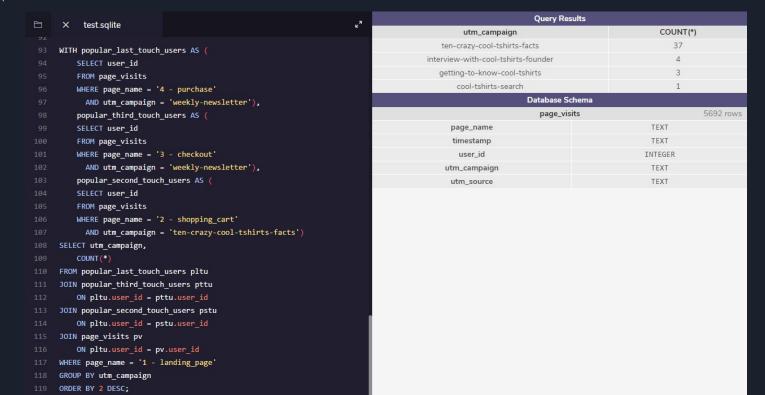
	× test.sqlite v ⁿ	Query Results		
92	X test.sqlite $ u^7 $	utm_campaign	COUNT(*)	
93	WITH popular last touch users AS (weekly-newsletter	115	
94	SELECT user id	Database Schema		
	FROM page visits	page_visits	5692 rows	
96	WHERE page_name = '4 - purchase'	page_name	TEXT	
97	AND utm_campaign = 'weekly-newsletter')	timestamp	TEXT	
	SELECT utm_campaign,	user_id	INTEGER	
	COUNT (*)	utm_campaign	TEXT	
100	FROM popular_last_touch_users pltu	utm_source	TEXT	
101	JOIN page_visits pv			
102	ON pltu.user_id = pv.user_id			
103	WHERE page_name = '3 - checkout'			
104	GROUP RV utm campaign			

ORDER BY 2 DESC;

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



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



- 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

- 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"
 - o 622 'first touch' users
 - o 119 purchases
 - ~19.1% purchase rate

	X test.sqlite 27	Query Results		
	↑ test.squite	COUNT(*)		
121	WITH popular_first_touch_users AS (119		
122	SELECT user_id	Database Schema		
123	FROM page_visits	page_visits	5692 rows	
124	WHERE page_name = '1 - landing_page'	page_name	TEXT	
125	AND utm_campaign = 'interview-with-cool-tshirts-founder')	timestamp	TEXT	
126	SELECT COUNT(*)	user_id	INTEGER	
127	FROM popular_first_touch_users pftu	utm_campaign	TEXT	
128	JOIN page_visits pv	utm_source	TEXT	
129	ON pftu.user_id = pv.user_id	-contactant 2571 00	- COA 10 (44)	
	The second secon			

- For "getting-to-know-cool-tshirts"
 - o 612 'first touch' users
 - o 106 purchases

WHERE page_name = '4 - purchase';

• ~17.3% purchase rate

	X test.sqlite v ^a	Query Results		
	7 test-squite	COUNT(*)		
121	WITH popular_first_touch_users AS (106		
122	SELECT user_id	Database Schema		
123	FROM page_visits	page_visits	5692 rows	
124	WHERE page_name = '1 - landing_page'	page_name	TEXT	
125	AND utm_campaign = 'getting-to-know-cool-tshirts')	timestamp	TEXT	
126	SELECT COUNT(*)	user_id	INTEGER	
127	FROM popular_first_touch_users pftu	utm_campaign	TEXT	
128	JOIN page_visits pv	utm_source	TEXT	
129	ON pftu.user_id = pv.user_id	50000000000000000000000000000000000000		

- For "ten-crazy-cool-tshirts-facts"
 - o 576 'first touch' users
 - o 111 purchases

WHERE page_name = '4 - purchase';

~19.3% purchase rate

	X test.sqlite	Query Results		
	7 testsquite	COUNT(*)		
121	WITH popular_first_touch_users AS (111		
122	SELECT user_id	Database Schema		
123	FROM page_visits	page_visits	5692 rows	
124	WHERE page_name = '1 - landing_page'	page_name	TEXT	
125	AND utm_campaign = 'ten-crazy-cool-tshirts-facts')	timestamp	TEXT	
126	SELECT COUNT(*)	user_id	INTEGER	
127	FROM popular_first_touch_users pftu	utm_campaign	TEXT	
128	JOIN page_visits pv	utm_source	TEXT	
129	ON pftu.user_id = pv.user_id	30 D 2000 March 2000 2000 2		

- For "cool-tshirts-search"
 - o 169 'first touch' users
 - o 31 purchases

WHERE page_name = '4 - purchase';

~18.3% purchase rate

	× test.sqlite	Query Results COUNT(*)	
	* test.squite		
121	WITH popular_first_touch_users AS (31	
122	SELECT user_id	Database Schema	
123	FROM page_visits	page_visits	5692 rows
124	WHERE page_name = '1 - landing_page'	page_name	TEXT
125	AND utm_campaign = 'cool-tshirts-search')	timestamp	TEXT
126	SELECT COUNT(*)	user_id	INTEGER
127	FROM popular_first_touch_users pftu	utm_campaign	TEXT
128	JOIN page_visits pv	utm_source	TEXT
129	ON pftu.user_id = pv.user_id		331000

Final Insights

Which 5 campaigns should CoolTShirts re-invest in?

- We've established that CoolTShirts utilizes 8 advertising campaigns
 - o ten-crazy-cool-tshirt-facts via buzzfeed
 - weekly-newsletter via email
 - o retargetting-campaign via email
 - retargetting-ad via facebook
 - o paid-search via google
 - o cool-tshirts-search via google
 - o interview-with-cool-tshirts-founder via medium
 - o 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
 - o interview-with-cool-tshirts-founder (31.4%)
 - o getting-to-know-cool-tshirts (30.9%)
 - ten-crazy-cool-tshirt-facts (29.1%)
- Each of these can be associated with a purchase rate
 - o interview-with-cool-tshirts-founder (19.1%)
 - o getting-to-know-cool-tshirts (17.3%)
 - ten-crazy-cool-tshirt-facts (19.3%)
- The large majority (92.5%) of last touches <u>at purchase</u> are associated with 4 campaigns
 - weekly-newsletter (31.9%)
 - retargetting-ad (31.3%)
 - retargetting-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