



Learn SQL From Scratch

Capstone Project:

Putting it all together with CoolTShirts.com

1. Get familiar with CoolTShirts

1.1 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`.

1.2 What pages are on their website?

2. What is the user journey?

2.1 How many first touches is each campaign responsible for?

2.2 How many last touches is each campaign responsible for?

2.3 How many visitors make a purchase?

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

2.5 What is the typical user journey?

3. Optimize the campaign budget

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

1.

**Get familiar with
CoolTShirts**

1.1 Get familiar with CoolTShirts



1.1) How many campaigns and sources does CoolTShirts use? Which source is used for each campaign?

- **Campaigns** are coordinated efforts by CoolTShirts to drive business, generate positive PR, and increase brand awareness.
 - a. There are **8** different campaigns represented in the dataset.
- **Sources** are the medium by which a campaign is delivered (e.g. email, nytimes, google.)
 - a. There are **6** different sources in the dataset.
- **Sources** can contain multiple different types of **Campaigns**.
- Theoretically, **Campaigns** could span several different types of **Sources**, though that does not appear in this dataset.

```
--1) Get Familiar with CoolTshirts
--- Get COUNT of distinct campaigns:
SELECT COUNT(DISTINCT utm_campaign) AS 'Campaign Count'
FROM page_visits;
--- Get COUNT of distinct sources:
SELECT COUNT(DISTINCT utm_source) AS 'Source Count'
FROM page_visits;
---List sources per campaign
SELECT DISTINCT utm_campaign AS Campaign, utm_source AS Source
FROM page_visits;
```

Campaign	Source
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

Campaign Count
8
Source Count
6

1.2 Get familiar with CoolTShirts Cont.

codecademy

1.2) What pages are on the CoolTShirts website?

- Find the distinct values of the **page_name** column.
- The DISTINCT query to the right returns 4 page names.
 - a. It is possible (likely) that these page names are roll-up categorizations of a multitude of pages.

```
---Find unique pages that are on website  
SELECT DISTINCT page_name AS 'Page Names'  
FROM page_visits;
```

Page Names

1 - landing_page

2 - shopping_cart

3 - checkout

4 - purchase

2.

**What is the user
journey?**

2.1 What is the user journey?

2.1) 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 the campaign and source which drove that user to the site for the first time (AKA 'first touch').
- Interesting to note that out of nearly 2,000 unique users, only 4 of 8 campaigns represented in the dataset were responsible for the first touch.

```
--2.1 COUNT first touches per campaign
--Create temporary table that finds first
--touches by user ID:
WITH first_touch AS (
  SELECT user_id,
         MIN(timestamp) AS first_touch_time
  FROM page_visits
  GROUP BY user_id),
--Create 2nd temporary table that adds
--source and campaign data from the
--page_visits table, joining on user
--ID and timestamp:
ft_attr AS (
  SELECT ft.user_id,
         ft.first_touch_time,
         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_time = pv.timestamp
)
--The below select statement counts the number
--of records from the above temporary table,
--which has been designed to get a reliable count
--of unique records (by user ID) that represent the
--number of first touches for each campaign (and by definition, source.)
SELECT ft_attr.utm_source AS Source,
       ft_attr.utm_campaign AS Campaign,
       COUNT(*) AS Count
FROM ft_attr
GROUP BY 1, 2
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: the prompt did not require the addition of 'Source', but the hint prompt suggested including it.

2.2 What is the user journey? Cont.

2.2) 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 the campaign and source which drove that user to the site for the last time (AKA 'last touch').
- All 8 campaigns are represented in the dataset when it comes to last touches.
- E-mail newsletters and facebook ads are responsible for the lion's share of last touches.
 - a. 890 of 1979, or roughly 45%.

```
---Count last touches per campaign
---Create temporary table that returns last touches
---by user ID:
WITH last_touch AS (
  SELECT user_id,
         MAX(timestamp) AS last_touch_at
  FROM page_visits
  GROUP BY user_id),
---Create 2nd temporary table that adds
---source and campaign data from the
---page_visits table, joining on user
---ID and timestamp:
lt_attr AS (
  SELECT lt.user_id,
         lt.last_touch_at,
         pv.utm_source,
         pv.utm_campaign,
         pv.page_name
  FROM last_touch lt
  JOIN page_visits pv
  ON lt.user_id = pv.user_id
  AND lt.last_touch_at = pv.timestamp
)
---The below select statement counts the number
---of records from the above temporary table,
---which has been designed to get a reliable count
---of unique records (by user ID) that represent the
---number of last touches for each campaign (and by definition, source.)
SELECT lt_attr.utm_source AS Source,
       lt_attr.utm_campaign AS Campaign,
       COUNT(*) AS Count
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

Source	Campaign	Count
email	weekly-newsletter	447
facebook	retargeting-ad	443
email	retargeting-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: the prompt did not require the addition of 'Source', but the hint prompt suggested including it.

2.3) How many visitors make a purchase?

- By using a simple count of unique users who reach the purchase page, we find that 361 (or 18%) of CoolTShirt's visitors are converted to consumers.

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

- By narrowing our query to return only last touch results from the purchase page, we can identify the sources and campaigns that were most directly responsible for converting visitors to customers.
- The data shows that the weekly newsletter e-mail and facebook ads account for 228 of the 361 (63.2%) purchase page last touch records.

```
---Count the distinct users who end up making a purchase:
SELECT COUNT(DISTINCT user_id) AS 'Purchasing Customer Count'
FROM page_visits
WHERE page_name = '4 - purchase';
```

```
---Count last touches per campaign that led to a purchase
---Create temporary table that returns last touches
---by user ID:
WITH last_touch AS (
  SELECT user_id,
         MAX(timestamp) AS last_touch_at
  FROM page_visits
  ---add WHERE clause to narrow results to purchase page only
  WHERE page_name = '4 - purchase'
  GROUP BY user_id),
---Create 2nd temporary table that adds
---source and campaign data from the
---page_visits table, joining on user
---ID and timestamp:
lt_attr AS (
  SELECT lt.user_id,
         lt.last_touch_at,
         pv.utm_source,
         pv.utm_campaign,
         pv.page_name
  FROM last_touch lt
  JOIN page_visits pv
  ON lt.user_id = pv.user_id
  AND lt.last_touch_at = pv.timestamp
)
---The below select statement counts the number
---of records from the above temporary table,
---which has been designed to get a reliable count
---of unique records (by user ID) that represent the
---number of last touches for each campaign (and by definition, source.)
SELECT lt_attr.utm_source AS Source,
       lt_attr.utm_campaign AS Campaign,
       COUNT(*) AS Count
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

Purchasing Customer Count

361

Source	Campaign	Count
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

Note: the prompt did not require the addition of 'Source', but the hint prompt suggested including it.

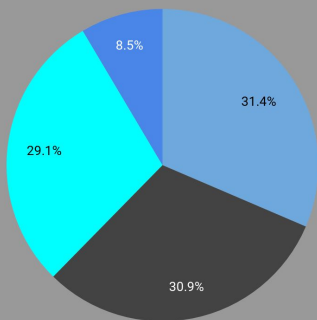
2.5 What is the user journey? Cont.

2.5) What is the typical user journey?

- **Users are drawn to the site** initially by one of only four campaigns.
 - a. **91% of users** initially visit the site via one of these narrative-style campaigns (as opposed to sales-driven communications.)
 - i. **Interview with CoolTShirts Founder**
 - ii. **Getting to Know CoolTShirts**
 - iii. **Ten Crazy CoolTShirts Facts**

First Touches by Campaign

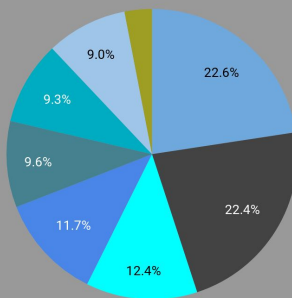
interview-with-cool-tshirts-founder getting-to-know-cool-tshirts
ten-crazy-cool-tshirts-facts cool-tshirts-search



- **Users need additional touch-points to stay engaged:**
 - a. **31% of users** who came via one of the narrative-style campaigns don't return, and these campaigns only account for 6% of purchases.
 - b. The vast majority of visitors required a 2nd (sales-oriented) campaign to become customers.

Last Touches by Campaign

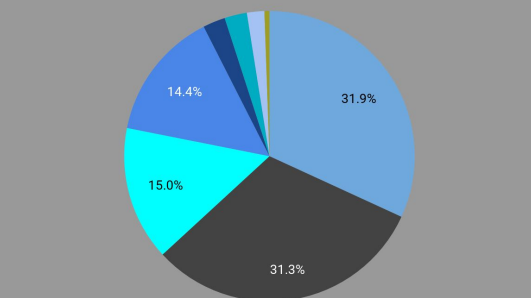
weekly-newsletter retargeting-ad retargeting-campaign
getting-to-know-cool-tshirts ten-crazy-cool-tshirts-facts
interview-with-cool-tshirts-founder paid-search cool-tshirts-search



- **Sales-oriented campaigns generate the most revenue:**
 - a. 93% of purchases were generated via a Sales-oriented campaign:
 - i. Weekly newsletter (email)
 - ii. Retargeting ad (facebook)
 - iii. Retargeting campaign (email)
 - iv. Paid search (google)

Purchases by Campaign

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



3

Optimize the Campaign Budget



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

The table below has green cell shading for any values representing at least 10% of the total within each metric. By reinvesting in the 5 campaigns recommended in the table to the right, CoolTShirts will be retaining over 95% of sales sources.

If assumption #1 to the right is not correct, we recommend swapping the paid search (least effective sales-driving campaign) for the ten crazy cool tshirts facts campaign. This more balanced approach would ensure CoolTShirts would continue to drive a robust number of new customers (1st touches) as well as sales (% purchases.)

CAMPAIGN REINVESTMENT

RECOMMENDATIONS:

We recommend that the following 5 campaigns be reinvested in:

- 1) Getting to know cool tshirts
- 2) Weekly newsletter
- 3) Retargeting Campaign
- 4) Retargeting Ad
- 5) Paid Search

DOCUMENTED ASSUMPTIONS:

- 1) Purchases (driving revenue) is a much higher priority than 1st touches (driving new potential customers).
- 2) All campaigns have an equal cost to CoolTShirts.

Campaign	1st Touches	% 1st Touches	Last Touches	% Last Touches	Purchases	% Purchases
getting-to-know-cool-tshirts	612	31%	232	12%	9	2.49%
weekly-newsletter	-	-	447	23%	115	31.86%
retargeting-campaign	-	-	245	12%	54	14.96%
retargeting-ad	-	-	443	22%	113	31.30%
paid-search	-	-	178	9%	52	14.40%
REINVESTMENT SUBTOTAL:	612	31%	1,545	78%	343	95.01%
ten-crazy-cool-tshirts-facts	576	29%	190	10%	9	2.49%
interview-with-cool-tshirts-founder	622	31%	184	9%	7	1.94%
cool-tshirts-searchz	169	9%	60	3%	2	0.55%
DISCONTINUE SUBTOTAL:	1,367	69%	434	22%	18	4.99%
TOTAL:	1,979	100%	1,979	100%	361	100%