



CAPSTONE: ATTRIBUTION

codecademy/PRO

Learn SQL From Scratch

Emily Newton

July 6, 2018

PRESENTATION CONTENTS

1. Get familiar with CoolTShirts

- a) CoolTShirts Schema
- b) CoolTShirts Campaign and Source Overview
- c) CoolTShirts Campaign and Source Detail
- d) CoolTShirts Campaign and Source Relationship

2. What is the user journey?

- a) CoolTShirts Website Pages
- b) CoolTShirts User Activity Detail
 - Unique user touches by page
 - Campaign-generated first touches
 - Campaign-generated purchase conversions

3. Optimize the campaign budget



GETTING FAMILIAR WITH COOLTSHIRTS

Campaigns, Sources
& Site Structure

COOLTSHIRTS SCHEMA

A SELECT * command limiting the return to the first 10 records of the table returns all fields in the page_visits schema. The table comprises a record of each “landing” on the CoolTShirts website:

page_name: the names of the website’s pages

timestamp: the time associated with each record in the table

user_id: the user associated with each record in the table

utm_campaign: the campaign triggering each record in the table

utm_source: the source (location) of origin of each record in the table

```
SELECT *  
FROM page_visits  
LIMIT 10;
```

page_name	timestamp	user_id	utm_campaign	utm_source
1 - landing_page	43124.13352	10006	getting-to-know-cool-tshirts	nytimes
2 - shopping_cart	43124.16963	10006	getting-to-know-cool-tshirts	nytimes
3 - checkout	43125.96546	10006	weekly-newsletter	email
1 - landing_page	43125.85558	10030	ten-crazy-cool-tshirts-facts	buzzfeed
2 - shopping_cart	43125.96183	10030	ten-crazy-cool-tshirts-facts	buzzfeed
3 - checkout	43128.55975	10030	retargetting-campaign	email
4 - purchase	43128.56808	10030	retargetting-campaign	email
1 - landing_page	43105.77172	10045	getting-to-know-cool-tshirts	nytimes
2 - shopping_cart	43105.88631	10045	getting-to-know-cool-tshirts	nytimes
3 - checkout	43109.12867	10045	retargetting-ad	facebook

COOLTSHIRTS CAMPAIGN AND SOURCE OVERVIEW

3 SELECT statements provide a clean overview of campaigns, sources and their relationships to each other

- 1) Count the distinct utm_campaign values in the page_visits table
- 2) Count the distinct utm_source values in the page_visits table
- 3) Generate a list of the distinct campaigns and their respective sources, grouping by and ordering by utm_campaign

```
1 SELECT COUNT(DISTINCT utm_campaign) AS num_campaigns
   FROM page_visits;

2 SELECT COUNT(DISTINCT utm_source) AS num_sources
   FROM page_visits;

3 SELECT utm_campaign
      ,utm_source
   FROM page_visits
  GROUP BY utm_campaign
 ORDER BY utm_campaign;
```

CAMPAIGN & SOURCE SELECT RESULTS

Source and Campaign Distinct Count Results:

Query Results	
1	num_campaigns 8
2	num_sources 6

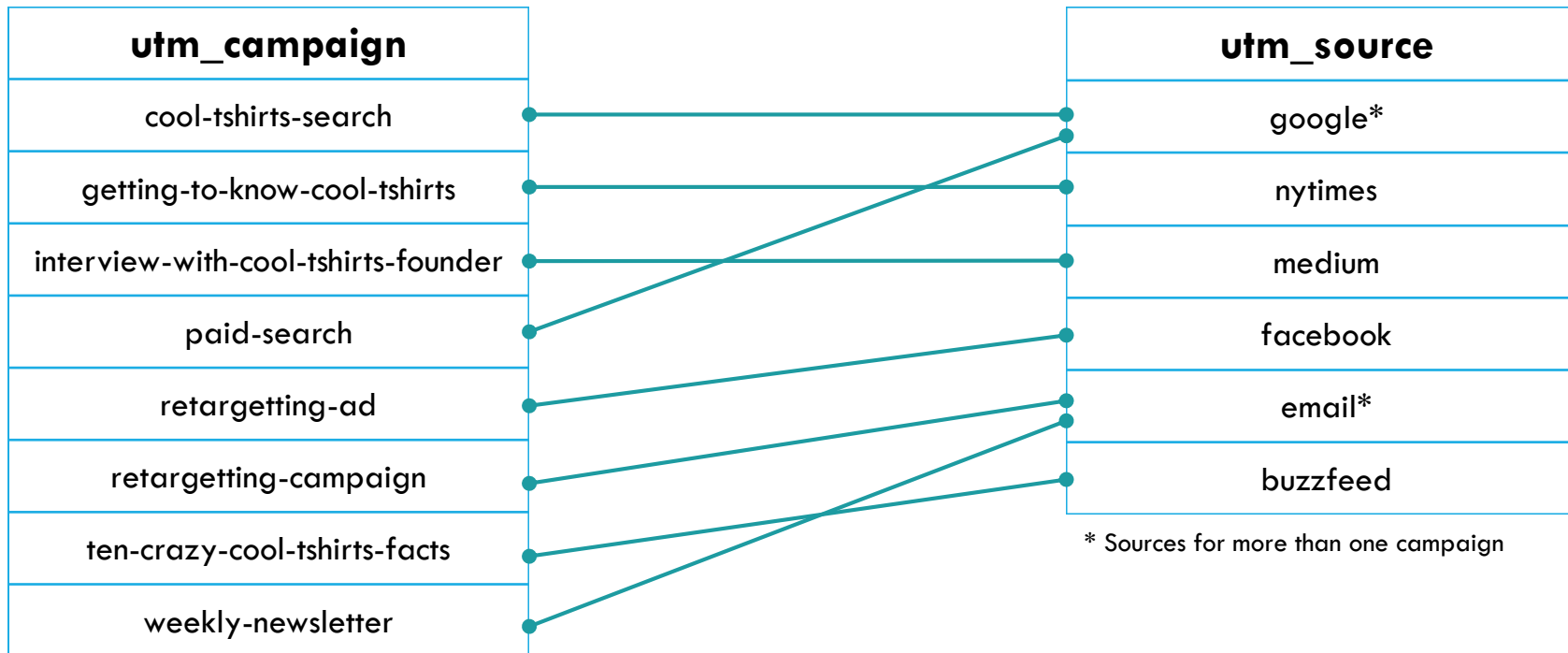
8 distinct campaigns originate from 6 distinct sources

Source and Campaign Results Grouped by Campaign Results:

Query Results																			
3	<table><tr><th>utm_campaign</th><th>utm_source</th></tr><tr><td>cool-tshirts-search</td><td>google</td></tr><tr><td>getting-to-know-cool-tshirts</td><td>nytimes</td></tr><tr><td>interview-with-cool-tshirts-founder</td><td>medium</td></tr><tr><td>paid-search</td><td>google</td></tr><tr><td>retargeting-ad</td><td>facebook</td></tr><tr><td>retargeting-campaign</td><td>email</td></tr><tr><td>ten-crazy-cool-tshirts-facts</td><td>buzzfeed</td></tr><tr><td>weekly-newsletter</td><td>email</td></tr></table>	utm_campaign	utm_source	cool-tshirts-search	google	getting-to-know-cool-tshirts	nytimes	interview-with-cool-tshirts-founder	medium	paid-search	google	retargeting-ad	facebook	retargeting-campaign	email	ten-crazy-cool-tshirts-facts	buzzfeed	weekly-newsletter	email
utm_campaign	utm_source																		
cool-tshirts-search	google																		
getting-to-know-cool-tshirts	nytimes																		
interview-with-cool-tshirts-founder	medium																		
paid-search	google																		
retargeting-ad	facebook																		
retargeting-campaign	email																		
ten-crazy-cool-tshirts-facts	buzzfeed																		
weekly-newsletter	email																		

HOW ARE CAMPAIGNS AND SOURCES RELATED?

8 Campaigns Bring Users from 6 Sources



CAMPAIGN & SOURCE TRAFFIC COUNTS?

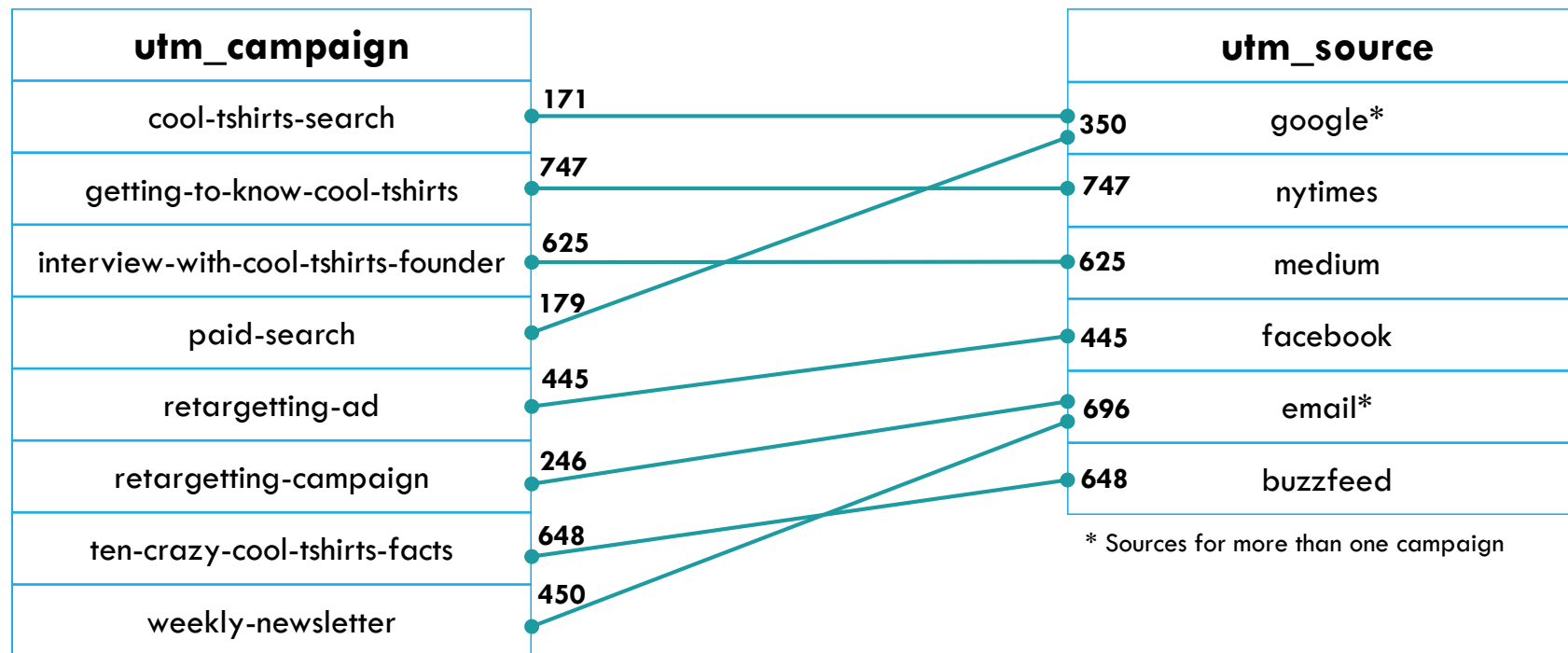
Selecting a unique user_id count along with campaign and source information, grouped by utm_campaign, will provide detail on touches generated by both:

```
SELECT utm_campaign AS campaign
      ,utm_source AS source
      ,COUNT(DISTINCT user_id) AS traffic
FROM page_visits
GROUP BY utm_campaign
ORDER BY 3 DESC;
```

campaign	source	traffic
getting-to-know-cool-tshirts	nytimes	747
ten-crazy-cool-tshirts-facts	buzzfeed	648
interview-with-cool-tshirts-founder	medium	625
weekly-newsletter	email	450
retargeting-ad	facebook	445
retargeting-campaign	email	246
paid-search	google	179
cool-tshirts-search	google	171

CAMPAIGNS AND SOURCES TRAFFIC COUNTS

8 Campaigns Bring Users from 6 Sources





THE USER JOURNEY

Tracing origins, touches, and transactions in the user's journey

COOLTSHIRTS WEBSITE PAGES

The website comprises four pages, which can be returned selecting distinct page names from `page_visits`

```
SELECT DISTINCT page_name  
FROM page_visits;
```

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

UNIQUE PAGE TOUCHES BY USER

How many unique touches does each page have? Are users visiting every page on the site? A count of distinct `user_id` values grouped by `page_name` will show us the distribution of page touches across the site:

```
SELECT COUNT(DISTINCT user_id) AS user_touches
      ,page_name
FROM page_visits
GROUP BY page_name;
```

Query Results	
user_touches	page_name
1979	1 - landing_page
1881	2 - shopping_cart
1431	3 - checkout
361	4 - purchase

Page visit counts decrease as users move through the search, selection, checkout, and purchase processes represented by pages 1, 2, 3, and 4, respectively.

CAMPAIGN-GENERATED FIRST TOUCHES

```
WITH first_touch AS (  
  SELECT user_id  
    ,MIN(timestamp) as first_touch_at  
  FROM page_visits  
  GROUP BY user_id),  
campaignsource_landing AS  
(SELECT ft.user_id  
  ,ft.first_touch_at  
  ,pv.page_name  
  ,pv.utm_campaign  
  ,pv.utm_source  
FROM first_touch as ft  
JOIN page_visits as pv  
  ON ft.user_id = pv.user_id AND  
  ft.first_touch_at = pv.timestamp  
ORDER BY ft.user_id, ft.first_touch_at)  
SELECT utm_campaign as campaign  
  ,utm_source as source  
  ,COUNT(user_id) as landings  
FROM campaignsource_landing  
GROUP BY utm_campaign  
ORDER BY 3 DESC;
```

A query breaking out first touches by campaign shows where first-time site visits originated

campaign	source	landings
interview-with-cool-tshirts-founder	medium	622
getting-to-know-cool-tshirts	nytimes	612
ten-crazy-cool-tshirts-facts	buzzfeed	576
cool-tshirts-search	google	169

CAMPAIGN-GENERATED PURCHASE CONVERSIONS

```
WITH last_touch AS (  
  SELECT user_id  
    ,MAX(timestamp) as last_touch_at  
  FROM page_visits  
  GROUP BY user_id),  
campaignsource_purchase AS (  
  SELECT lt.user_id  
    ,lt.last_touch_at  
    ,pv.page_name  
    ,pv.utm_campaign  
    ,pv.utm_source  
  FROM last_touch as lt  
  JOIN page_visits as pv  
    ON lt.user_id = pv.user_id AND  
    lt.last_touch_at = pv.timestamp  
  WHERE pv.page_name = '4 - purchase')  
SELECT utm_campaign  
  ,utm_source  
  ,COUNT(user_id) as purchases  
FROM campaignsource_purchase  
GROUP BY utm_campaign  
ORDER BY 3 DESC;
```

A query breaking out last touches on '4 – purchase' by campaign shows where purchase transactions to the website originated

Query Results		
utm_campaign	utm_source	purchases
weekly-newsletter	email	114
retargeting-ad	facebook	112
retargeting-campaign	email	53
paid-search	google	52
getting-to-know-cool-tshirts	nytimes	9
ten-crazy-cool-tshirts-facts	buzzfeed	9
interview-with-cool-tshirts-founder	medium	7
cool-tshirts-search	google	2



CAMPAIGN SPEND OPTIMIZATION

Campaign investment
opportunities with highest
projected ROI

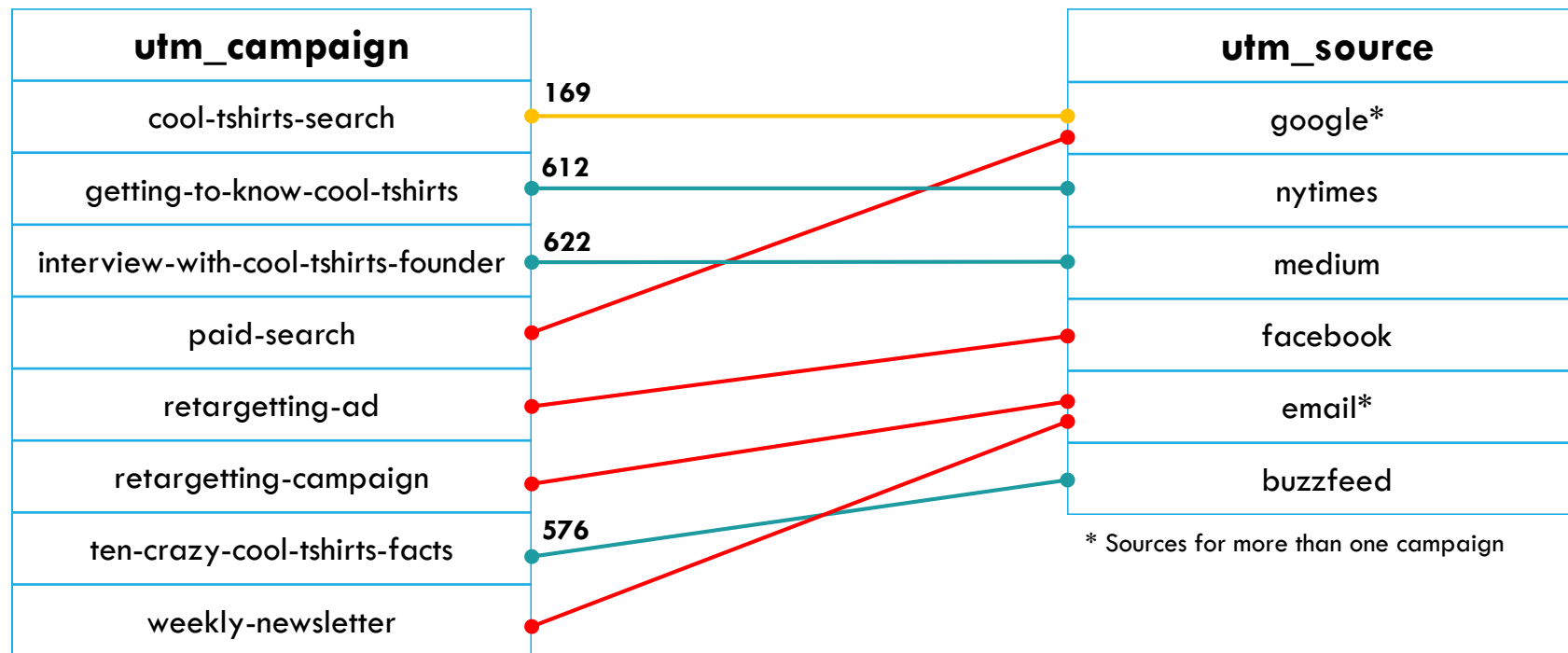
OPTIMIZED BUDGET FOR 5 CAMPAIGNS

STRATEGY:

Optimize spend through efficiency: select top traffic drivers to site and balance with top conversion drivers to CoolTShirts.com, utilizing as many sources of traffic as possible

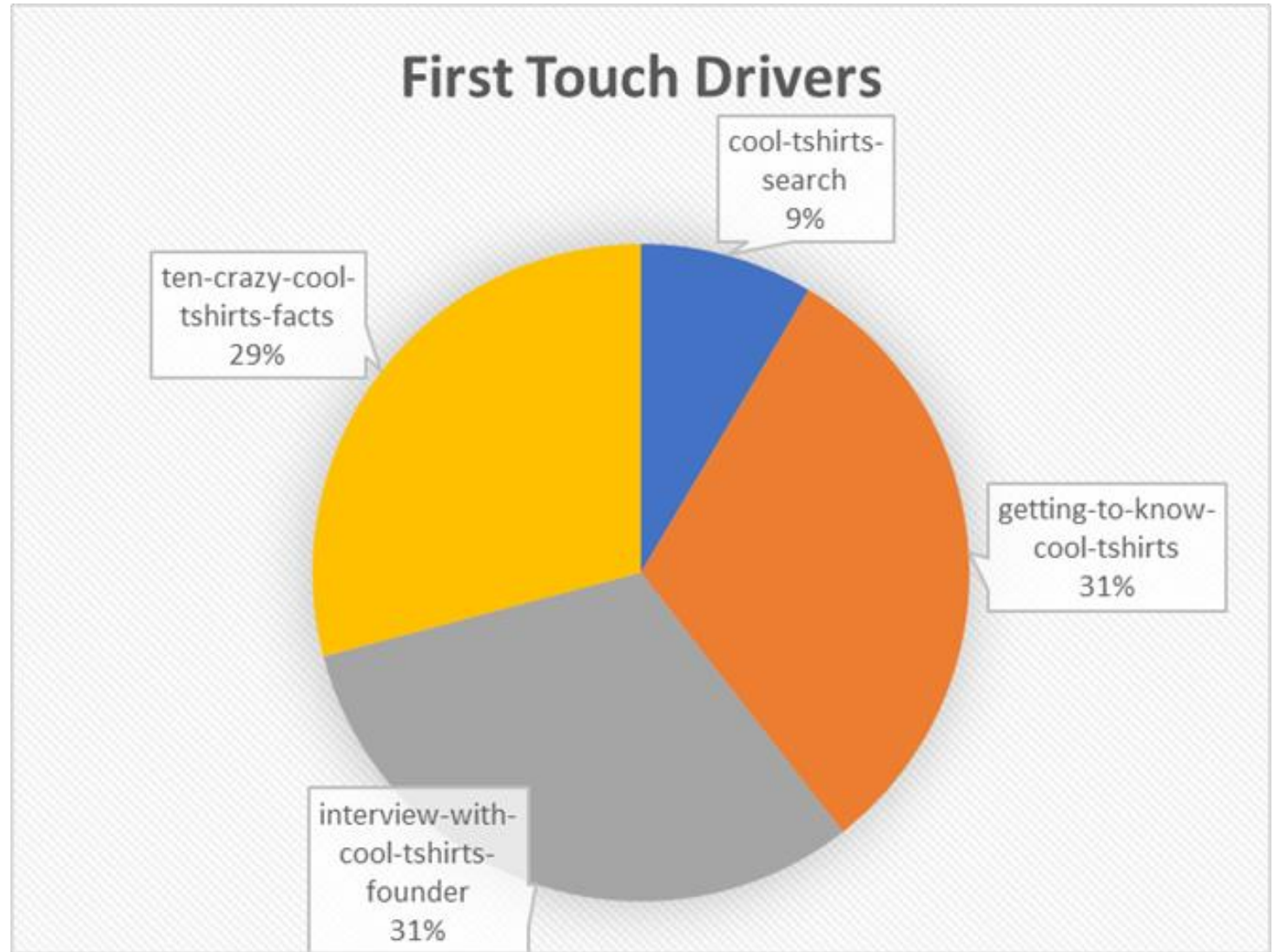
CAMPAIGN AND SOURCE FIRST TOUCHES

4 Campaigns Bring ALL First-Time Users From 4 Sources



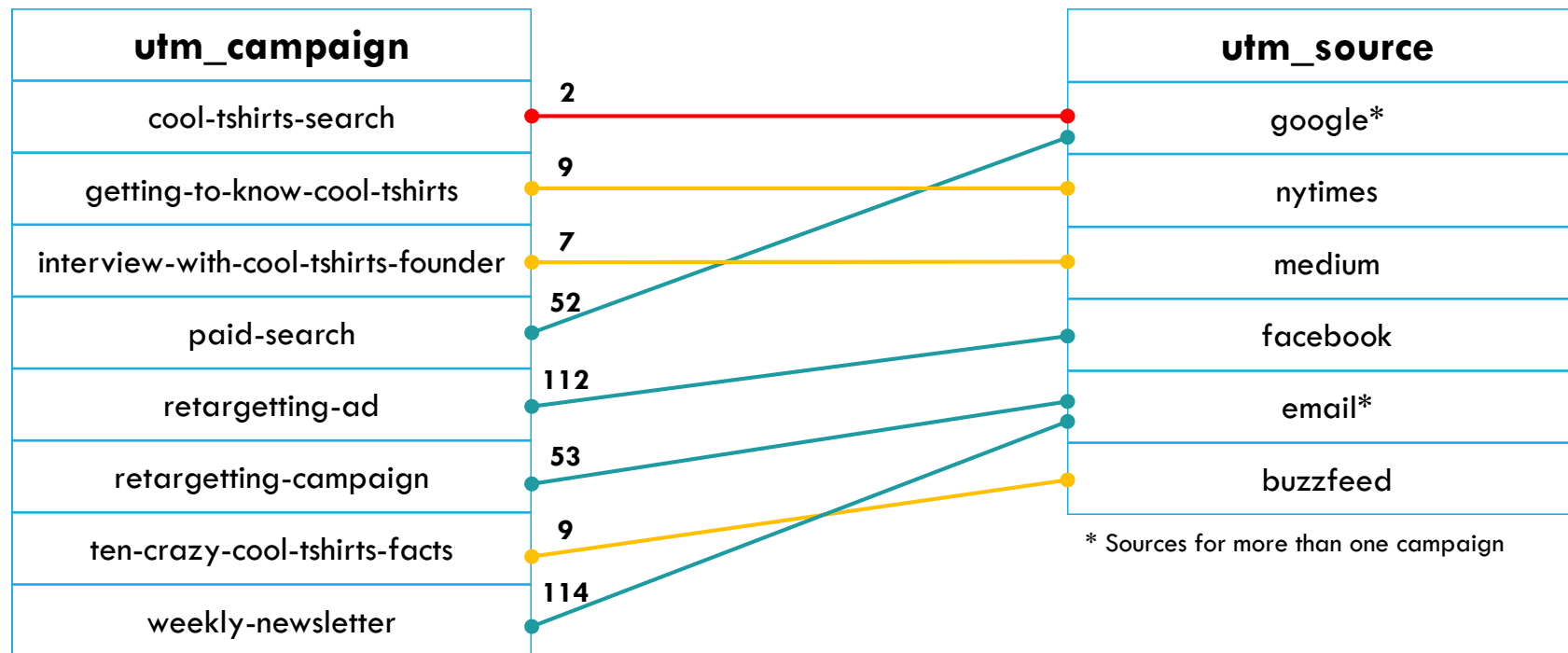
CAMPAIGNS DRIVING TRAFFIC TO WEBSITE

1. **Getting to know cooltshirts/
nytimes**
2. **Interview with cooltshirts
founder / medium**
3. **Ten crazy cool thirts facts /
buzzfeed**



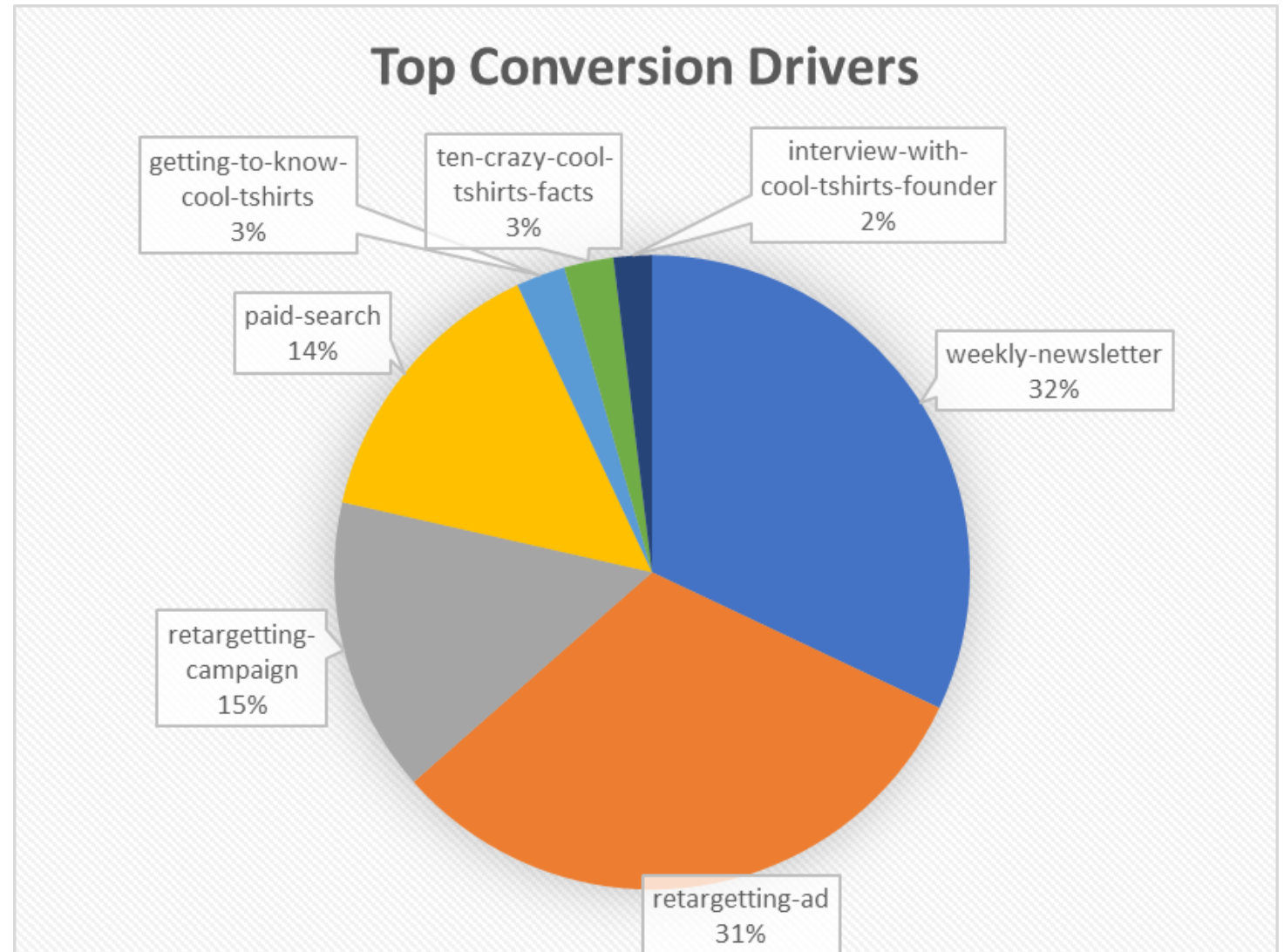
CAMPAIGN AND SOURCE PURCHASE CONVERSIONS

8 Campaigns Bring Users from 6 Sources



CAMPAIGNS DRIVING PURCHASE CONVERSION

1. Weekly-Newsletter / email
2. Retargeting-ad / facebook
3. Regargetting Campaign / email



5 CAMPAIGNS FOR TARGET SPEND

The selected campaigns account for 78% of all purchase conversions and 62% of all traffic to the CoolTShirts.com website. Four source channels are covered by the five campaigns, with email serving as the focused driver returning shoppers to the site to complete their transactions.

1. Weekly Newsletter
2. Retargeting Ad
3. Retargeting Campaign
4. Getting to know CoolTShirts
5. Interview with CoolTShirts Founder