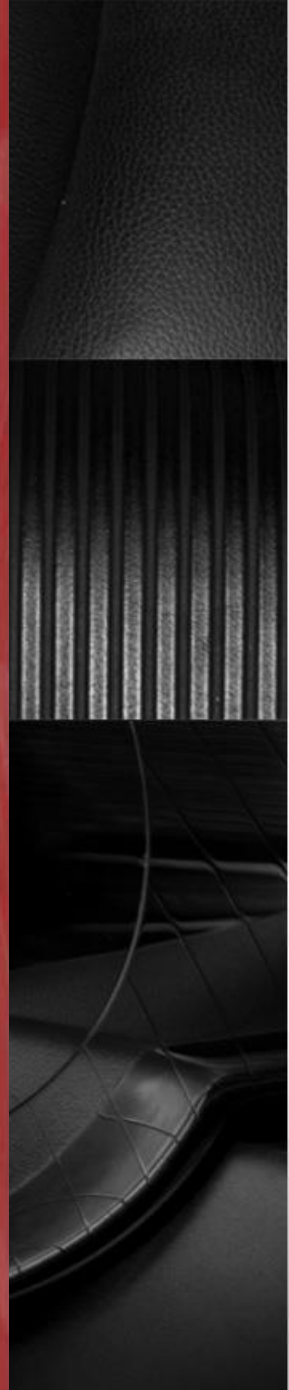# CODECADEMY
# LEARN SQL FROM SCRATCH

NICK NORTON

# Contents

**Get familiar with the company.**

- 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*.

- What pages are on their website?

**What is the 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 the typical user journey?

**Optimize the campaign budget**

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

# Getting familiar with the company

```sql
--The number of distinct campaigns
SELECT COUNT(DISTINCT utm_campaign) AS 'Campaign Count'
FROM page_visits;

-- The number of distinct sources
SELECT COUNT(DISTINCT utm_source) AS 'Source Count'
FROM page_visits;

--How they are related
SELECT DISTINCT utm_campaign AS 'Campaign', utm_source AS 'Source'
FROM page_visits;
```

**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*.**

A campaign in this database refers to a message or strategy to reach and convert customers. For example campaigns like "ten-crazy-cool-tshirts-facts" or "getting-to-know-cool-tshirts" cover specific messages while "paid-search" or "retargeting" are the strategies used to reach customers.

Source refers to where the customer traffic was driven from before reaching the CoolTShirts website. These sources can be specific websites, like "buzzfeed" or "nytimes", or broader mediums like "email".

Campaign and Source have a many to many relationship where multiple campaigns can run across multiple sources and vice versa. We can see an instance of this for the google Source which has 2 camapigns "paid-search" and "cool-tshirts-search".

| Campaign Count | Source Count |
|----------------|--------------|
| 8 | 6 |

| 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 |

# Getting familiar with the company

**What pages are on their website?**

There are 4 types of pages on the CoolTShirts website which are all listed on the right.

The page_name column in the database can be used to show the progression of customers through the sales funnel starting with the "1 - landing_page", where the customer starts their engagement with the website, and finishes with "4 – purchase" where a sale can be completed.

It can be expected that each of the steps listed would be a collection of webpages. For instance landing page could be made up of the homepage, campaign specific pages, product pages etc…

```
--What pages are on the CoolTShirts website?
SELECT DISTINCT page_name AS 'Pages'
FROM page_visits;
```

| Pages |
|---|
| 1 - landing_page |
| 2 - shopping_cart |
| 3 - checkout |
| 4 - purchase |

# What is the user journey?

**How many first touches is each campaign responsible for?**

In total 1,979 first touches can be attributed to the campaign activity, the breakdown across the 4 campaigns can be seen on the right.

This was accomplished by building off the previous query which counted first touches and creating a new temporary table "ft_at" (short for First Touch Attribution) which included the columns we were looking to connect to the first touch engagement.

Once this was set up all that was required was calling on the campaign name and first touch count using the "ft_at" table and grouping/ordering to make it easier to analyse.

```sql
--How many first touches is each campaign responsible for?
WITH first_touch AS (
    SELECT user_id,
        MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id),

ft_at AS (
  SELECT ft.user_id,
        ft.first_touch_at,
        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_at = pv.timestamp
        )
SELECT ft_at.utm_campaign AS 'Campaign',
        COUNT(*) AS 'First Touch'
FROM ft_at
GROUP BY 1
ORDER BY 2 DESC;
```

| Campaign | First Touch |
|---|---|
| interview-with-cool-tshirts-founder | 622 |
| getting-to-know-cool-tshirts | 612 |
| ten-crazy-cool-tshirts-facts | 576 |
| cool-tshirts-search | 169 |

# What is the user journey?

**How many last touches is each campaign responsible for?**

In total 1,979 first touches can be attributed to the campaign activity, the breakdown across the 8 campaigns can be seen on the right.

It's interesting to note that while there were only 4 campaigns that saw first touch attributions all 8 campaigns saw a last touch.

The 4 new campaigns are the newsletter, paid search and two retargeting which makes sense as they are campaigns which rely on an initial engagement and are used to move customers further down the sales funnel.

To find these results the same process was followed from the first touch query however the MAX calculation was used instead of MIN.

```sql
--How many last touches is each campaign responsible for?
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) AS last_touch_at
    FROM page_visits
    GROUP BY user_id),

lt_at AS (
  SELECT lt.user_id,
        lt.last_touch_at,
        pv.utm_source,
        pv.utm_campaign
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
    )
SELECT lt_at.utm_campaign AS 'Campaign',
        COUNT(*) AS 'Last Touch'
FROM lt_at
GROUP BY 1
ORDER BY 2 DESC;
```

| Campaign | Last Touch |
|---|---|
| weekly-newsletter | 447 |
| retargetting-ad | 443 |
| retargetting-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 |

# What is the user journey?

**How many visitors make a purchase?**

By querying the distinct number of users who have reached the step 4 purchase page we can see that there have been 361 purchasers to date.

While this query tells us how many customers have made a purchase it wouldn't necessarily tell us the total number of purchases there have been, ie. If one customer bought two items over 2 weeks. However after performing a 2nd query it doesn't look as though there have been any repeat customers to date.

Looking at our previous queries we can also calculate the conversion rate 361(purchasers)/ 1,979 (total first touches) = 18%.

```sql
--How many visitors make a purchase?
SELECT COUNT(DISTINCT user_id) AS 'purchasers'
FROM page_visits
WHERE page_name = '4 - purchase';
```

| Purchasers |
|:----------:|
| 361 |

```sql
--How many visitors make a purchase?
SELECT COUNT(user_id) AS 'purchases'
FROM page_visits
WHERE page_name = '4 - purchase';
```

| Purchases |
|:---------:|
| 361 |

# What is the user journey?

```sql
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) AS last_touch_at
    FROM page_visits
    WHERE page_name = '4 - purchase'
    GROUP BY user_id),

lt_at AS (
  SELECT lt.user_id,
        lt.last_touch_at,
        pv.utm_source,
        pv.utm_campaign
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
        )
SELECT lt_at.utm_campaign AS 'Campaign',
        COUNT(*) AS 'Last Touch'
FROM lt_at
GROUP BY 1
ORDER BY 2 DESC;
```

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

To get this result we can use the query from one of the previous questions *"How many last touches is each campaign responsible for?"* and add a WHERE clause to filter the results for the last_touch temporary table and only pull results for the final "4 – purchase" page.

Looking at these results we can see that the top purchase drivers are the newsletter, retargeting and paid search campaigns which account for 93% of the sales.

Looking at these results we can see that there are 2 clear campaign types; performance, which are used to drive purchases which are the top 4 campaigns on the right, and Brand campaigns which drive traffic to the site and get that initial engagement which the performance campaigns capitalize off.

| Campaign | Last Touch |
|---|---|
| weekly-newsletter | 115 |
| retargetting-ad | 113 |
| retargetting-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 |

# What is the user journey?

**What is the typical user journey?**

By adding source to the existing queries we can get a fully detailed picture of the first, last and purchase milestones that customers of CoolTShirts have reached.

Using this and a new query "Customer Journey" which looks at how many unique users are reaching each page we can see that customers are being driven to the site by the branding campaigns like "interview-with-cool-tshirts-founder." Looking at the "Customer Journey" results we can see that while we're seeing a lot of traffic being driven to the site, 1,979 customers, only 361 completing a purchase.

The greatest drop off along this conversion path seems to be coming from the checkout page where only 25% of customers who reach this page complete a purchase which indicates that maybe the user experience on this last step isn't very good which is why there's such a drop off.

Once the customer has reached the site about 18% complete a purchase and the rest are then followed up with the performance campaigns like the "weekly-newsletter" which brings them back to the site and serves as a reminder to complete their purchase.

```sql
--Last Touch 4 - Purchse by Source and Campaign
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) AS last_touch_at
    FROM page_visits
    WHERE page_name = '4 - purchase'
    GROUP BY user_id),

lt_at AS (
  SELECT lt.user_id,
        lt.last_touch_at,
        pv.utm_source,
        pv.utm_campaign
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
        )

SELECT lt_at.utm_campaign AS 'Campaign',
        lt_at.utm_source AS 'Source',
        COUNT(*) AS 'Purchase'
FROM lt_at
GROUP BY 1, 2
ORDER BY 3 DESC;
```

```sql
--Last Touch by Source and Campaign
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) AS last_touch_at
    FROM page_visits
    GROUP BY user_id),

lt_at AS (
  SELECT lt.user_id,
        lt.last_touch_at,
        pv.utm_source,
        pv.utm_campaign
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
        )

SELECT lt_at.utm_campaign AS 'Campaign',
        lt_at.utm_source AS 'Source',
        COUNT(*) AS 'Last Touch'
FROM lt_at
GROUP BY 1, 2
ORDER BY 3 DESC;
```

```sql
--First Touch by Source and Campaign
WITH first_touch AS (
    SELECT user_id,
        MIN(timestamp) AS first_touch_at
    FROM page_visits
    GROUP BY user_id),

ft_at AS (
  SELECT ft.user_id,
        ft.first_touch_at,
        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_at = pv.timestamp
        )

SELECT ft_at.utm_campaign AS 'Campaign',
        ft_at.utm_source AS 'Source',
        COUNT(*) AS 'First Touch'
FROM ft_at
GROUP BY 1, 2
ORDER BY 3 DESC;
```

```sql
--Customer Journey
SELECT page_name AS 'Page Name',
        COUNT(DISTINCT user_id) 'Page View'
FROM page_visits
GROUP BY page_name;
```

*Please see Results for the above queries on the next page

# What is the user journey?
## The Results

### First Touch

| Campaign | Source | First Touch |
|---|---|---|
| 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 |

### Customer Journey

| Page Name | Page View | Drop Off % |
|---|---|---|
| 1 - landing_page | 1,979 | - |
| 2 - shopping_cart | 1,881 | 95% |
| 3 – checkout | 1,431 | 76% |
| 4 - purchase | 361 | 25% |

### Last Touch

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

### Purchases

| Campaign | Source | Purchase |
|---|---|---|
| weekly-newsletter | email | 115 |
| retargetting-ad | facebook | 113 |
| retargetting-campaign | email | 54 |
| 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 |

# Optimize the campaign budget

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

Based on the previous queries the key area for growth and increased sales would be improving the conversion rate from Checkout to Purchase. The best campaigns for this would be the weekly –newsletter, Facebook retargeting and paid search campaigns which have had the greatest success as delivering purchases. With the weekly newsletter and Facebook retargeting activity driving the most sales they make the best options to drive more volume of sales. Assuming as well that the email used for the weekly newsletter would be collected at the checkout step the weekly newsletter would make for the best campaign to drive customers from the checkout to complete that final step.

The reason the paid-search activity was picked over email retargeting , which drove similar amounts of purchases, is because the paid search activity was seeing a much higher conversion rate between last touch and purchase. With additional budget moved into paid search we would hope to see an increase in volume of conversions as it's already shown itself to be the best driver of quality traffic with the highest conversion rate of 29% (178 last touch/52 purchases)

On top of the 3 performance campaigns suggest above, which are all purchase focused, there would also need to be a constant supply of new customers which the branding campaigns (the campaigns which are listed under first touch) can deliver. The two most successful campaigns for driving traffic to date have been the "interview-with-cool-tshirts-founder" and "getting-to-know-cool-tshirts". With 62% of traffic driven by these 2 campaigns there would be a good supply of new customers which can then be fully utilised across the 3 performance campaigns which would work to drive sales.

On top of reinvesting in the above campaigns I'd also look at the "3 – checkout page" to see why there's such a noticeable drop off ie. Too many steps, unfriendly user interface etc... If there's any changes that can be made to the website to help the campaigns drive sales this would also be highly valuable and cost effective.