Attribution Queries

1. Get familiar with CoolTShirts
2. What is the user journey?
3. Optimize the campaign budget

1. Getting familiar with the company

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

I've used the count distinct function to get the number of campaigns and sources. The third query shows how the campaigns and sources are related.

Num of Campaigns	Num of Sources
8	6

```
-- You can put your query here
SELECT
COUNT(DISTINCT(utm_campaign))  AS "Num of Campaigns"
FROM page_visits;


SELECT
COUNT(DISTINCT(utm_source))  AS "Num of Sources"
FROM page_visits;


SELECT
utm_source, utm_campaign
FROM page_visits
GROUP BY 2;
```

1.2 What pages are on their website?

I've used the distinct function for page names from the table.

```
-- You can put your query here
SELECT
DISTINCT(page_name) AS Pages
FROM page_visits;
```

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

2. What is the user journey?

2.1 How many first touches is each campaign responsible for?

I've joined tables in order to get the first touch campaign applying the MIN function in the timestamp.

Campaigns	Num of First Touches
interview-with-cool-tshirts-founder	622
getting-to-know-cool-tshirts	612
ten-crazy-cool-tshirts-facts	576
cool-tshirts-search	169

```sql
WITH first_touch AS (
  SELECT
    user_id,
    MIN(timestamp) AS first_touch_at
  FROM
    page_visits
  GROUP BY
    user_id),
  ft_attr 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_attr.utm_campaign AS "Campaigns",
  COUNT(*) AS "Num of First Touches"
FROM
  ft_attr
GROUP BY
  1
ORDER BY
  2 DESC;
```

2.2 How many last touches is each campaign responsible for?

Similar to previous query, but this time with the MAX in timestamp.

Campaigns	Num of Last Touches
weekly-newsletter	447
retargetting-ad	443
retargetting-campaign	245
getting-to-know-cool-tshirts	232

```
WITH last_touch AS (
  SELECT
    user_id,
    MAX(timestamp) AS last_touch_at
  FROM
    page_visits
  GROUP BY
    user_id),
lt_attr 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_attr.utm_campaign AS "Campaigns",
  COUNT(*) AS "Num of Last Touches"
FROM
  lt_attr
GROUP BY
  1
ORDER BY
  2 DESC;
```

2.3 How many visitors make a purchase?

I've used the count distinct to get the number of user_ids filtered by page name equals 4 - purchase

```sql
SELECT
  COUNT(DISTINCT(user_id)) AS "Users who Purchased"
FROM
  page_visits
WHERE
  page_name = "4 - purchase";
```

Users who Purchased
361

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

Similar to the previous queries this one includes the page name as evidence of purchase.

Num of Campaigns	Num of Sources
weekly-newsletter	115
retargetting-ad	113
retargetting-campaign	54
paid-search	52

```
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_attr 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_attr.utm_campaign AS "Campaigns",
  COUNT(*) AS "Num of Last Touches"
FROM
  lt_attr
GROUP BY
  1
ORDER BY
  2 DESC;
```

2.5 What is the typical user journey?

I've used the distinct function to get the page names and the sort by timestamp. In theory this would demonstrate a typical sequence of page visits.

```
-- You can put your query here
SELECT
  DISTINCT page_name
FROM
  page_visits
ORDER BY
  timestamp ASC;
```

Page_name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

3. Optimize the campaign budget

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

I've used the distinct function to get the campaigns that have been attributed to conversions (based on the page name "purchase") and sort them to get the top 5.

```sql
SELECT
  DISTINCT utm_campaign AS "Top 5 Campaigns"
FROM
  page_visits
WHERE
  page_name = "4 - purchase"
ORDER BY
  1 DESC
LIMIT
  5;
```

Top 5 Campaigns
Weekly-newsletter
ten-crazy-cool-tshirts-facts
retargetting-campaign
retargetting-ad
paid-search