FINAL SUBMISSION for Alison Wellsfry (wellsfry36@gmail.com)
Project : Attribution Queries for CoolTShirts

1a. 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.
      utm_campaign shows 8 campaigns (listed below)
      utm_source shows 6 sources
      related as shown
1b. What pages are on their website?
      page_name shows 4 pages, with the last indicating purchase

Code 1a
SELECT DISTINCT utm_campaign
FROM page_visits;
SELECT DISTINCT utm_source
FROM page_visits;
SELECT DISTINCT utm_campaign, utm_source
FROM page_visits;

Code 1b
SELECT DISTINCT page_name
FROM page_visits;

1. QUERY RESULTS:

utm_source
nytimes
email
buzzfeed
facebook
medium
google

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

utm_campaign	utm_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

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

2. What is the user journey?
2a. How many first touches is each campaign responsible for? 622, 612, 576, 169
2b. How many last touches is each campaign responsible for? 447 down to 60 (see table)
2c. How many visitors make a purchase? total 361
2d. How many last touches on the purchase page is each campaign responsible for? 115 down to 2 (see table)
2e. What is the typical user journey? Most users are introduce from an article and end up purchasing after the weekly newsletter or facebook retargeting ad

Code 2a
```
WITH first_touch AS (
    SELECT user_id,
       MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id)
SELECT ft.user_id,
    ft.first_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    COUNT (utm_campaign)
FROM first_touch ft
JOIN page_visits pv
    ON ft.user_id = pv.user_id
    AND ft.first_touch_at = pv.timestamp
GROUP BY utm_campaign
ORDER BY 5 DESC;
```

Code 2b
```
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) as last_touch_at
    FROM page_visits
    GROUP BY user_id)
SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    COUNT (utm_campaign)
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
GROUP BY utm_campaign
ORDER BY 5 DESC;
```

Code 2c and 2d
```
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)
SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    COUNT (utm_campaign)
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
GROUP BY utm_campaign
ORDER BY 5 DESC;
```

2. QUERY RESULTS

user_id	first_touch_at	utm_source	utm_campaign	COUNT (utm_campaign)
99990	2018-01-13 23:30:09	medium	interview-with-cool-tshirts-founder	622
99933	2018-01-25 00:04:39	nytimes	getting-to-know-cool-tshirts	612
99765	2018-01-04 05:59:46	buzzfeed	ten-crazy-cool-tshirts-facts	576
99684	2018-01-13 13:20:49	google	cool-tshirts-search	169

user_id	last_touch_at	utm_source	utm_campaign	COUNT (utm_campaign)
99933	2018-01-26 06:18:39	email	weekly-newsletter	447
99928	2018-01-24 05:26:09	facebook	retargetting-ad	443
99990	2018-01-16 11:35:09	email	retargetting-campaign	245
99589	2018-01-15 04:55:43	nytimes	getting-to-know-cool-tshirts	232

99765	2018-01-04 05:59:47	buzzfeed	ten-crazy-cool-tshirts-facts	190
99838	2018-01-02 07:40:34	medium	interview-with-cool-tshirts-founder	184
98840	2018-01-10 04:58:48	google	paid-search	178
99344	2018-01-18 21:36:32	google	cool-tshirts-search	60

user_id	last_touch_at	utm_source	utm_campaign	COUNT (utm_campaign)
99933	2018-01-26 06:18:39	email	weekly-newsletter	115
99897	2018-01-06 09:41:19	facebook	retargetting-ad	113
99285	2018-01-24 09:00:58	email	retargetting-campaign	54
94567	2018-01-19 16:37:58	google	paid-search	52

92172	2018-01-16 15:15:29	nytimes	getting-to-know-cool-tshirts	9
98651	2018-01-15 04:17:36	buzzfeed	ten-crazy-cool-tshirts-facts	9
83547	2018-01-10 18:20:21	medium	interview-with-cool-tshirts-founder	7
95650	2018-01-18 00:25:00	google	cool-tshirts-search	2

3. Optimize the campaign budget
CoolTShirts can re-invest in 5 campaigns. Which should they pick and why?
The email weekly-newletter resulted in the most purchases and may be the least expensive since it may be internally controlled. If only looking at top 5 causes of purchases, reinvest in those. But if there is a campaign cost which can be divided by the purchase numbers, others may rise to the top.