Arianna Lang Wang
Springboard Jan 2nd, 2018 Cohort

SQL Case Study - Investigating a Drop in User Engagement
Feb 28th, 2018

List of possible causes for the dip in user engagement:

-- Speculation #1: Is there a drop of active users in a particular country?

```
SELECT location,
    DATE_TRUNC('week', occurred_at) AS weekly_date,
    COUNT(DISTINCT user_id) AS weekly_active_users
  FROM tutorial.yammer_events
WHERE event_type = 'engagement'
  AND occurred_at < '2014-08-25'
  AND occurred_at > '2014-07-28'
GROUP BY 1, 2
ORDER BY 1
```

-- Speculation #2: Are all the device working? Is there a device bug?

```
SELECT device,
    DATE_TRUNC('week', occurred_at) AS weekly_date,
    COUNT(DISTINCT user_id) AS weekly_active_users
  FROM tutorial.yammer_events
WHERE event_type = 'engagement'
  AND occurred_at < '2014-08-25'
  AND occurred_at > '2014-07-01'
GROUP BY 1, 2
ORDER BY 1, 2
```

-- Speculation #3: People take vacations during the summer, so work can be slow around this time:

```
SELECT DATE_TRUNC('week', occurred_at) AS "number of events",
    COUNT(event_name) AS weekly_events
  FROM tutorial.yammer_events
WHERE event_type = 'engagement'
  AND occurred_at < '2014-08-25'
```

```
  AND occurred_at > '2014-04-25'
GROUP BY 1
ORDER BY 1
```

-- Yes, number of events decreased from 21,474 to 16,158 from the end of July to the end of August, 2014.

-- Speculation #4: To further verifying my second speculation, I want to check if there is a drop in number of active emails opened/clicked-through in August 2014.

```
SELECT DATE_TRUNC('week', event.occurred_at) AS weekly_date,
    COUNT(email.action) AS active_emails
FROM tutorial.yammer_events event
JOIN tutorial.yammer_emails email
  ON event.user_id = email.user_id
WHERE event.event_type = 'engagement'
  AND event.event_name = 'login'
  AND email.occurred_at < '2014-08-25'
  AND email.occurred_at > '2014-06-28'
  AND email.action = 'email_open'
  OR email.action = 'email_clickthrough'
GROUP BY 1
ORDER BY 1
```

-- Yes, there is indeed a drop in number of active emails opened/clicked-through in August 2014. But this drop can be caused by the summer slow-usage or it can be caused by the emails themselves. So further investigation may be needed by the engineering department.

Answer the following questions:

Do the answers to any of your original hypotheses lead you to further questions?
Yes. Speculation #3 led me to formulate speculation #4. If there is a drop in number of events from end of July to end of August 2014, then I want to check if there is a drop in number of active emails opened/clicked-through during this period of time as well.

If they are questions that you can't answer using data alone, how would you go about answering them (hypothetically, assuming you actually worked at this company)?
I would go talk to people from other departments and try to figure out if there are any changes made about the product.

What seems like the most likely cause of the engagement dip?

From my analysis, the cause of the engagement dip may either be the summer time of the year or something happened with the emails.