**Investigating a Drop in User Engagement**

**The problem**

You show up to work Tuesday morning, September 2, 2014. The head of the Product team walks over to your desk and asks you what you think about the latest activity on the user engagement dashboards.

**Initial list of possible causes:**

* Tracking system error – identify if capturing growth/usage
* Region issues – identify specific(s) region that drove decline
* Major Vacation/Holiday period – Not a high usage time
* System issues – During the period the system was down, slow, application failures, …

**Summary initial analysis**

* Tracking system error – identify if capturing growth/usage

Could not identify any daily gaps in the system for tracking usage.

* Region issues – identify specific(s) region that drove decline

Used language to maybe identify geographic regions with problems. English is by far the dominate language. Even though there were a few languages, other than English that did not follow the trend, the majority did. So, at this time will follow up with this.

* Major Vacation/Holiday period – Not a high usage time

Since this happening at the July and early August want to see if drop off was historically consistent, since in many parts of world it could be high season for summer vacation. The data did exist for previous year so could not validate if this was true.

* System issues – During the period the system was down, slow, application failures, …

Look at usage by device and found decline with phone and tablet users. But mainly phone users.

Look at Email activity and found decline with the 'email\_clickthrough'

Will follow up with the product team on these two items to see if they are aware of any issues or changes corresponding to these two items and if I can help further identify more details or other potential problem areas

**Tables used for analysis:**

**Tables Counts**

* Users 19066
* Events 340832
* Email Events 90389
* Rollup Period 56002

**Initial data exploration**

select count(\*) from tutorial.yammer\_users

select count(\*) from tutorial.yammer\_events

select count(\*) from tutorial.yammer\_emails

select count(\*) from benn.dimension\_rollup\_periods

**Tracking system error –** identify if capturing growth/usage

Capture daily usage during the period '2014-07-01' and '2014-08-31'.

*On a daily basis no system issues are highlighted*

SELECT DATE\_TRUNC('day',created\_at) AS day,

COUNT(\*) AS all\_users,

COUNT(CASE WHEN activated\_at IS NOT NULL THEN u.user\_id ELSE NULL END) AS activated\_users

FROM tutorial.yammer\_users u

WHERE created\_at between '2014-07-01' AND '2014-08-31'

GROUP BY 1

ORDER BY 1

[Link to SQL and Chart in Mode - Daily capture of usage](https://modeanalytics.com/editor/rivasjmr/reports/3796d8711733)

**Region issues – Look at languages to maybe identify regions that drove decline**

Look at languages to associate with geographical areas during the period '2014-07-21' and '2014-08-31'

SELECT

u.language, COUNT(DISTINCT e.user\_id)

FROM tutorial.yammer\_events e

join tutorial.yammer\_users u

on e.user\_id = u.user\_id

WHERE e.occurred\_at between '2014-07-21' and '2014-08-31'

GROUP BY 1

ORDER BY 1

Compare by language for same period if one language (which could dictate geographical region) drove decline

SELECT DATE\_TRUNC('week', e.occurred\_at) week,

COUNT(DISTINCT e.user\_id) AS weekly\_active\_users,

COUNT(CASE WHEN u.language = 'arabic' THEN u.language ELSE NULL END) AS arabic,

COUNT(CASE WHEN u.language = 'chinese' THEN u.language ELSE NULL END) AS chinese,

COUNT(CASE WHEN u.language = 'english' THEN u.language ELSE NULL END) AS english,

COUNT(CASE WHEN u.language = 'french' THEN u.language ELSE NULL END) AS french,

COUNT(CASE WHEN u.language = 'german' THEN u.language ELSE NULL END) AS german,

COUNT(CASE WHEN u.language = 'indian' THEN u.language ELSE NULL END) AS indian,

COUNT(CASE WHEN u.language = 'italian' THEN u.language ELSE NULL END) AS italian,

COUNT(CASE WHEN u.language = 'japanese' THEN u.language ELSE NULL END) AS japanese,

COUNT(CASE WHEN u.language = 'korean' THEN u.language ELSE NULL END) AS korean,

COUNT(CASE WHEN u.language = 'portugese' THEN u.language ELSE NULL END) AS portuguese,

COUNT(CASE WHEN u.language = 'russian' THEN u.language ELSE NULL END) AS russian,

COUNT(CASE WHEN u.language = 'spanish' THEN u.language ELSE NULL END) AS spanish,

COUNT(CASE WHEN u.language <> 'english' THEN u.language ELSE NULL END) AS non\_english

FROM tutorial.yammer\_events e

join tutorial.yammer\_users u

on e.user\_id = u.user\_id

WHERE e.occurred\_at between '2014-07-21' and '2014-08-31'

GROUP BY 1

ORDER BY 1

*English is the dominate language, so initially compared English and grouped all the non-English languages together. Noticed that the non-English language category didn’t follow the same trend. Look closer at each language and most did follow the trend with a few exceptions:*

* + *Chinese and Portuguese actually trended upwards during that period*
  + *French and Spanish declined initially like the trend, but started upwards earlier*

[Link to SQL and Chart in Mode -By Language](https://modeanalytics.com/editor/rivasjmr/reports/3796d8711733)

**Major Vacation/Holiday period:**

Compare to same period previous year to see if trend is the same, but data does not exist.

SELECT min(e.occurred\_at),

max(e.occurred\_at)

FROM tutorial.yammer\_events e

WHERE e.event\_type = 'engagement'

AND e.event\_name = 'login'

**min max**

2014-05-01 02:27:15 2014-08-31 23:01:15

SELECT DATE\_TRUNC('week', e.occurred\_at),

COUNT(DISTINCT e.user\_id) AS weekly\_active\_users

FROM tutorial.yammer\_events e

WHERE e.occurred\_at between '2014-07-21' and '2014-08-31'

And e.event\_type = 'engagement'

AND e.event\_name = 'login'

GROUP BY 1

ORDER BY 1

SELECT DATE\_TRUNC('week', e.occurred\_at),

COUNT(DISTINCT e.user\_id) AS weekly\_active\_users

FROM tutorial.yammer\_events e

WHERE e.occurred\_at between '2013-07-21' and '2013-08-31'

And e.event\_type = 'engagement'

AND e.event\_name = 'login'

GROUP BY 1

ORDER BY 1

**System issues – During the period the system was down, slow, application failures, …**

Look at type of devices used during the period '2014-07-01' and '2014-08-31'

SELECT

e.device as Device,

count(\*)

FROM tutorial.yammer\_events e

WHERE e.occurred\_at between '2014-07-01' and '2014-08-31'

AND e.event\_type = 'engagement'

AND e.event\_name = 'login'

group by 1

order by 1

Look at usage by device. Phone and Tablet users show decline.

SELECT DATE\_TRUNC('week', occurred\_at) AS week,

COUNT(DISTINCT e.user\_id) AS weekly\_active\_users,

COUNT(DISTINCT CASE WHEN e.device IN ('macbook pro','lenovo thinkpad','macbook air','dell inspiron notebook',

'asus chromebook','dell inspiron desktop','acer aspire notebook','hp pavilion desktop','acer aspire desktop','mac mini')

THEN e.user\_id ELSE NULL END) AS computer,

COUNT(DISTINCT CASE WHEN e.device IN ('iphone 5','samsung galaxy s4','nexus 5','iphone 5s','iphone 4s','nokia lumia 635',

'htc one','samsung galaxy note','amazon fire phone') THEN e.user\_id ELSE NULL END) AS phone,

COUNT(DISTINCT CASE WHEN e.device IN ('ipad air','nexus 7','ipad mini','nexus 10','kindle fire','windows surface',

'samsumg galaxy tablet') THEN e.user\_id ELSE NULL END) AS tablet

FROM tutorial.yammer\_events e

WHERE e.occurred\_at between '2014-07-01' and '2014-08-31'

AND e.event\_type = 'engagement'

AND e.event\_name = 'login'

GROUP BY 1

ORDER BY 1

[Link to SQL and Chart in Mode -System Issues - devices](https://modeanalytics.com/editor/rivasjmr/reports/3796d8711733)

Looking at Email activity during the period '2014-07-01' and '2014-08-31'

SELECT

e.action as Action,

count(\*)

FROM tutorial.yammer\_emails e

WHERE e.occurred\_at between '2014-07-01' and '2014-08-31'

GROUP BY 1

ORDER BY 1

*Looking at by individual email activities. Decline in 'email\_clickthrough'*

SELECT DATE\_TRUNC('week', occurred\_at) AS week,

COUNT(CASE WHEN e.action = 'sent\_weekly\_digest' THEN e.user\_id ELSE NULL END) AS weekly,

COUNT(CASE WHEN e.action = 'sent\_reengagement\_email' THEN e.user\_id ELSE NULL END) AS reengagement,

COUNT(CASE WHEN e.action = 'email\_open' THEN e.user\_id ELSE NULL END) AS opens,

COUNT(CASE WHEN e.action = 'email\_clickthrough' THEN e.user\_id ELSE NULL END) AS clickthroughs

FROM tutorial.yammer\_emails e

WHERE e.occurred\_at between '2014-07-01' and '2014-08-31'

GROUP BY 1

ORDER BY 1

[Link to SQL and Chart in Mode -System Issues - Email](https://modeanalytics.com/editor/rivasjmr/reports/3796d8711733)