

INVESTIGANDO A QUEDA NO ENGAJAMENTO

RESUMO E PROBLEMA DE NEGÓCIO

O projeto a seguir é uma investigação da queda no engajamento da rede social corporativa Yammer.

O modelo de negócios da empresa é altamente influenciado pelo nível de engajamento dos usuários, já que ele é o motor da geração de receita por meio de propagandas e inscrições mensais.

Para realizar o projeto, seguirei as seguintes etapas:

- 1) Análise da queda de engajamento
- 2) Formulação de hipóteses que causas motoras
- 3) Recomendação de ações

ANÁLISE DA SITUAÇÃO ATUAL

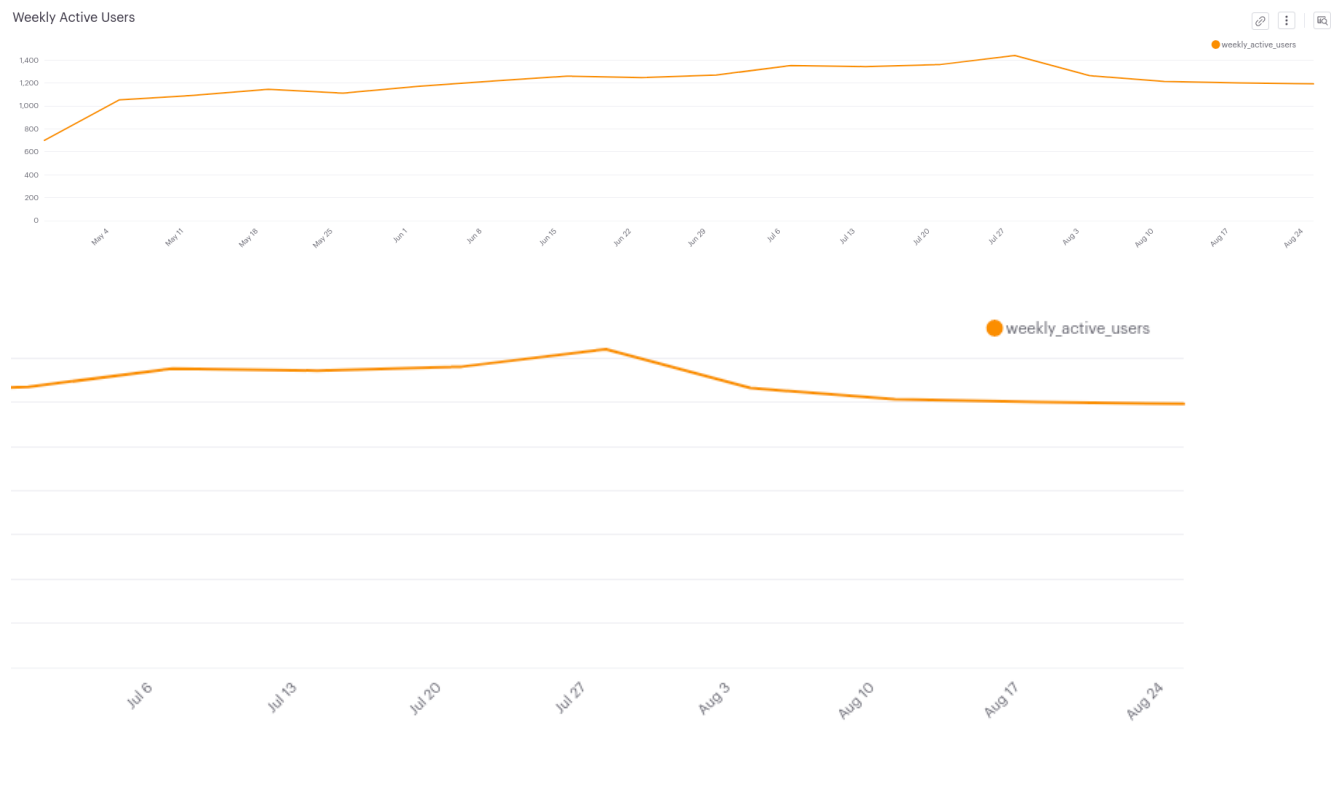


Imagem com zoom, compreendendo julho e agosto

Como se observa no gráfico acima, houve uma queda no engajamento da plataforma no mês de agosto. Para entender melhor o problema, a estratégia será colocar uma lupa em diferentes segmentações de clientes e de uso da plataforma.

FORMULAÇÃO DE HIPÓTESES

Primeiro, é necessário criar algumas hipóteses sobre o motivo da queda.

As mais prováveis são:

- Hipótese 1: A queda no engajamento ocorre por alguma característica de sazonalidade;
- Hipótese 2: A queda no engajamento ocorre em alguma região específica;
- Hipótese 3: A queda no engajamento ocorre por algum canal de marketing ineficiente;
- Hipótese 4: A queda no engajamento ocorre por algum defeito na plataforma;
- Hipótese 5: A queda no engajamento ocorre nos cliente mais novos;
- Hipótese 6: A queda no engajamento ocorre por problema em algum dispositivo.

Vamos usar as bases de dados para visualizar em gráficos cada uma das hipóteses.

Os dados que temos são:

- Tabela de usuários
 - mostra desde quando o usuário tem a conta, quando ela foi ativada e a linguagem de preferência
- Tabela de eventos
 - mostra as ações que os usuários executam na plataforma, como login, mensagem, cliques, pesquisas; são um bom indicativo de engajamento.
 - Há também informações complementares, como o local de login e o tipo de dispositivo usado
- Tabela de e-mails
 - mostra se e quando o usuário abriu o e-mail, assim como se clicou em algum link.

Hipótese 1: A queda no engajamento ocorre por alguma característica de sazonalidade

Uma primeira hipótese é que há algum fenômeno semanal ou mensal que possa estar influenciando a queda.

Código SQL

```

SELECT DATE_TRUNC('day',created_at) AS day,
       COUNT(CASE WHEN activated_at IS NOT NULL THEN user_id ELSE NULL END)
AS activated_users
FROM tutorial.yammer_users
WHERE activated_at < '2014-08-26'
GROUP BY 1
ORDER BY 1

```

Gráfico gerado no Mode - usuários ativos por mês desde 2013:



Código SQL

```

SELECT DATE_TRUNC('day',created_at) AS day,
       COUNT(CASE WHEN activated_at IS NOT NULL THEN user_id ELSE NULL END)
AS activated_users
FROM tutorial.yammer_users
WHERE activated_at BETWEEN '2014-08-01' AND '2014-08-26'
GROUP BY 1
ORDER BY 1

```

Gráfico gerado no Mode - usuários ativos por dia em agosto de 2014:

Activated users per day



Olhando para o engajamento desde 2013, após constante evolução, é possível perceber a maior queda de engajamento da série histórica em agosto de 2014. No entanto, não parece haver indícios de algum fenômeno sazonal analisando os meses. Há uma leve queda nos finais de semana, o que é de se esperar pela natureza do negócio.

Hipótese 2: A queda no engajamento ocorre em alguma região específica

Como a base de dados dá o país de acesso, é possível fazer a análise por continente, visualizando o padrão de cada um.

Código SQL

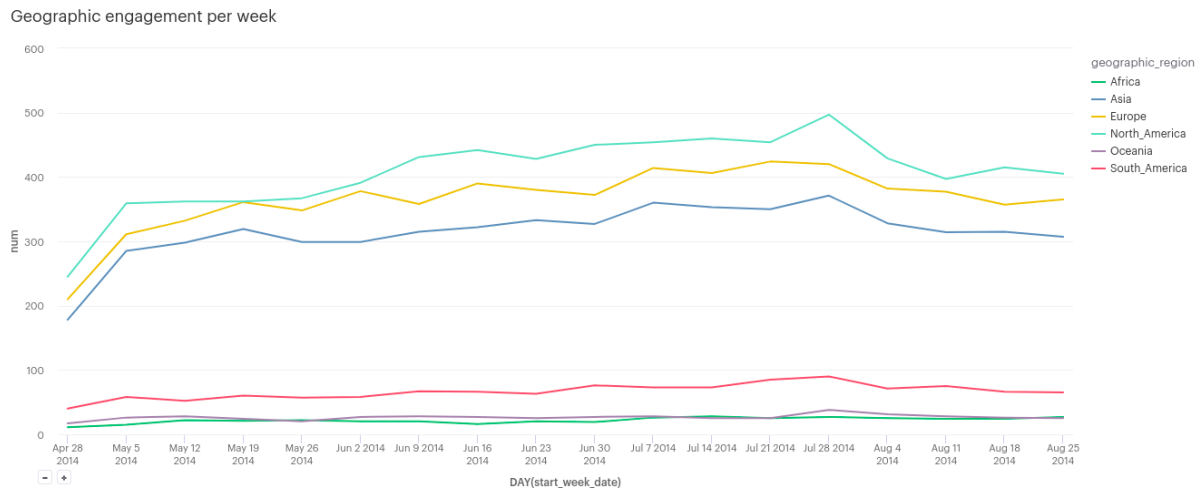
```
WITH geographic_regions AS(
  SELECT user_id,
    CASE
      WHEN location IN ('Indonesia', 'Korea', 'Singapore', 'Israel', 'Malaysia', 'Hong Kong',
'Saudi Arabia', 'Philippines', 'Turkey', 'United Arab Emirates', 'Taiwan', 'Thailand', 'India',
'Iran', 'Japan', 'Iraq', 'Russia', 'Pakistan')
      THEN 'Asia'
      WHEN location IN ('Venezuela', 'Colombia', 'Argentina', 'Chile', 'Brazil')
      THEN 'South_America'
      WHEN location IN ('Sweden', 'Ireland', 'Portugal', 'Finland', 'France', 'Netherlands',
'Spain', 'Belgium', 'Italy', 'United Kingdom', 'Germany', 'Greece', 'Denmark', 'Switzerland',
'Norway', 'Austria', 'Poland')
      THEN 'Europe'
      WHEN location IN ('United States', 'Canada', 'Mexico')
      THEN 'North_America'
      WHEN location IN ('Nigeria', 'Egypt', 'South Africa')
      THEN 'Africa'
      WHEN location IN ('Australia')
```

```

    THEN 'Oceania'
    ELSE null end as geographic_region,
    COUNT(*)
FROM tutorial.yammer_events
WHERE event_type = 'engagement'
GROUP BY 1, 2

```

Gráfico gerado no Mode - engajamento por continente ao longo do tempo:



Não parece haver diferenças significativas entre as regiões. Portanto, passamos para a próxima hipótese.

Hipótese 3: A queda no engajamento ocorre em um dispositivo específico

É possível que haja alguma incompatibilidade entre Iphone, Android ou Tablet com a plataforma. Por isso, vamos agrupar os tipos de dispositivos e tentar encontrar um padrão. Para captar a mudança, optou-se por medir a diferença de engajamento entre os meses de julho e agosto.

Código SQL

```

WITH device_type AS(
SELECT device,
CASE WHEN device in ('iphone 5','iphone 4s','iphone 5s','nexus 5','samsung galaxy s4','htc one','nokia lumia 635','samsung galaxy note','amazon fire phone') THEN 'Phone'
WHEN device in ('windows surface','kindle fire','ipad mini','samsung galaxy tablet','ipad air','nexus 7','nexus 10') THEN 'Tablet'
WHEN device in ('dell inspiron notebook','macbook air','macbook pro','lenovo thinkpad','acer aspire notebook','asus chromebook','acer aspire desktop','mac mini','hp pavilion desktop','dell inspiron desktop') THEN 'Desktop'
END AS device_type
FROM tutorial.yammer_events),

```

```

device_count AS(
  SELECT device,
    COUNT(CASE WHEN date_trunc('month', occurred_at) BETWEEN '2014-07-01' AND
'2014-07-31' THEN user_id ELSE NULL END) AS July_Sales,
    COUNT(CASE WHEN date_trunc('month', occurred_at) BETWEEN '2014-08-01' AND
'2014-08-31' THEN user_id ELSE NULL END) AS August_Sales
  FROM tutorial.yammer_events
  WHERE event_type='engagement'
  AND event_name = 'login'
  AND occurred_at BETWEEN '2014-07-01' AND '2014-08-31'
  GROUP BY 1
)
SELECT device_count.device, device_type.device_type, August_Sales, July_Sales,
(August_Sales - July_Sales) as diff
FROM device_count
JOIN device_type
ON device_count.device = device_type.device
GROUP BY 1, 2, 3, 4, 5
ORDER BY 5

```

Gráfico gerado no Mode - engajamento por dispositivo:

Engagement drop per device



Aqui já encontramos um efeito significativo. Os usuários via celular apresentaram a maior queda no engajamento, indicando que os problemas podem ser causados por algum defeito do aplicativo.

Hipótese 4: A queda no engajamento ocorre nos usuários novos

Vamos realizar uma análise cohort para entender se há alguma diferença de engajamento significativa entre antigos e novos clientes.

Código SQL

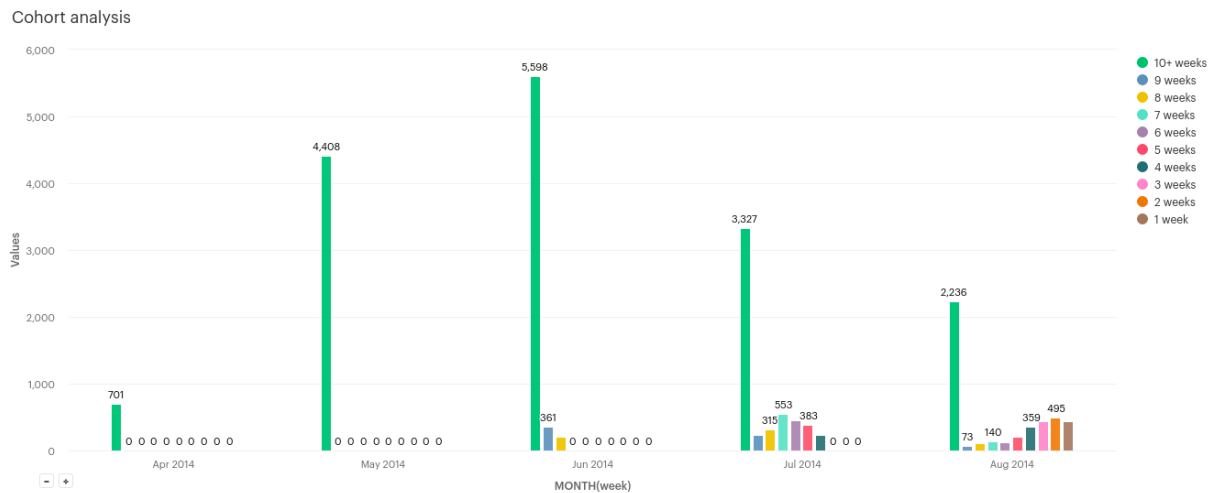
```

SELECT DATE_TRUNC('week',z.occurred_at) AS "week",
       AVG(z.age_at_event) AS "Average age during week",
       COUNT(DISTINCT CASE WHEN z.user_age > 70 THEN z.user_id ELSE NULL END)
AS "10+ weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 70 AND z.user_age >= 63 THEN
z.user_id ELSE NULL END) AS "9 weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 63 AND z.user_age >= 56 THEN
z.user_id ELSE NULL END) AS "8 weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 56 AND z.user_age >= 49 THEN
z.user_id ELSE NULL END) AS "7 weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 49 AND z.user_age >= 42 THEN
z.user_id ELSE NULL END) AS "6 weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 42 AND z.user_age >= 35 THEN
z.user_id ELSE NULL END) AS "5 weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 35 AND z.user_age >= 28 THEN
z.user_id ELSE NULL END) AS "4 weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 28 AND z.user_age >= 21 THEN
z.user_id ELSE NULL END) AS "3 weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 21 AND z.user_age >= 14 THEN
z.user_id ELSE NULL END) AS "2 weeks",
       COUNT(DISTINCT CASE WHEN z.user_age < 14 AND z.user_age >= 7 THEN
z.user_id ELSE NULL END) AS "1 week",
       COUNT(DISTINCT CASE WHEN z.user_age < 7 THEN z.user_id ELSE NULL END)
AS "Less than a week"
FROM (
    SELECT e.occurred_at,
           u.user_id,
           DATE_TRUNC('week',u.activated_at) AS activation_week,
           EXTRACT('day' FROM e.occurred_at - u.activated_at) AS age_at_event,
           EXTRACT('day' FROM '2014-09-01'::TIMESTAMP - u.activated_at) AS user_age
    FROM tutorial.yammer_users u
    JOIN tutorial.yammer_events e
      ON e.user_id = u.user_id
     AND e.event_type = 'engagement'
     AND e.event_name = 'login'
     AND e.occurred_at >= '2014-05-01'
     AND e.occurred_at < '2014-09-01'
    WHERE u.activated_at IS NOT NULL
) z

GROUP BY 1
ORDER BY 1

```

Gráfico gerado no Mode - retenção de usuários ativos desde o início do engajamento:



Também é possível visualizar que a retenção é um grande problema, talvez o maior encontrado. Quanto mais antigos os usuários, maior é a queda no engajamento. Isso também afasta a hipótese de que a queda seria gerada por alguma disfunção no marketing ou na atração de novos usuários.

Hipótese 5: A queda no engajamento ocorre por alguma disfuncionalidade na plataforma

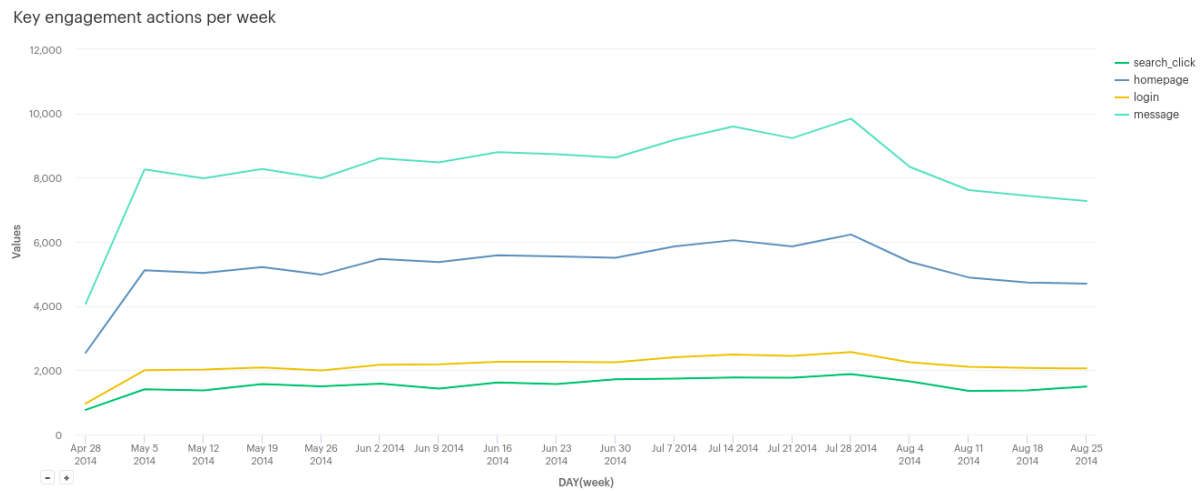
Há vários tipos de interações que os usuários podem fazer na plataforma: login, pesquisa, envio de mensagem...

Vamos investigar se alguma dessas formas de interação com a plataforma teve uma queda repentina. Se sim, pode ser um indicativo de defeito no site/app.

Código SQL

```
SELECT cast(date_trunc('week', occurred_at) as date) AS week,
COUNT(CASE WHEN event_name IN ('login') THEN user_id ELSE NULL END) AS login,
COUNT(CASE WHEN event_name IN ('home_page') THEN user_id ELSE NULL END)
AS homepage,
COUNT(CASE WHEN event_name IN ('view_inbox','like_message','send_message')
THEN user_id ELSE NULL END) AS message,
COUNT(CASE WHEN event_name IN ('search_autocomplete', 'search_run',
'search_click_result_1','search_click_result_2','search_click_result_3','search_click_result
_4','search_click_result_5','search_click_result_6','search_click_result_7','search_click_res
ult_8','search_click_result_9','search_click_result_10') THEN user_id ELSE NULL END)
AS search_click
FROM tutorial.yammer_events
GROUP BY 1
```


Gráfico gerado no Mode - interações com a plataforma por semana:



Todas as ações apresentaram queda no período de julho para agosto, no entanto, é possível observar uma queda proporcionalmente maior nas mensagens. Isso sugere que o problema pode ter ligação com algum defeito na plataforma, como tempo de carregamento ou defeito de alguma funcionalidade

CONCLUSÕES

Os insights do análise foram:

- Os maiores problemas estão concentrados em usuários antigos e nas pessoas que usam celular.
- Como ação, é interessante realizar uma pesquisa qualitativa com esses usuários para entender quais os maiores problemas.
- Ações que visem a retenção também são indicadas, como estratégias de gamificação e novas funcionalidades específicas para o público.