

Yammer_Data_Analysis

1. Product/Service에 대한 이해



Yammer란? “기업용 소셜 네트워크 서비스, 협업 도구(공동 작업)”

▼ 일반 SNS와 유사한 기능

1. 최근 글 최상단에 표시 기능
2. 호응 (좋아요), 댓글, 공유 기능
3. 팔로잉/팔로우 기능
4. 프로필 표시 및 수정 기능 (사진, 이름, 부서, 직급 등)
5. 사용자 접속 상태 파악 기능 (온라인, 오프라인, 모바일)
6. 비공개 채팅 기능
7. 데스크톱용 Yammer Notifier, 모바일 Yammer 앱
8. 한글 및 다국어 지원

▼ 추가적인 기능

1. 투표, 칭찬하기, 공지사항 기능
2. 파일 및 메모 공유 기능
3. 사용자, 본문, 댓글, 파일 검색 기능
4. 메일로 게시글 . 댓글 작성 기능
5. Office Online 지원 - 웹에서 엑셀, 파워포인트, 워드 문서 읽기 및 편집 가능
6. 외부 네트워크 기능 - 고객, 파트너, 계약업체와 같은 외부 사람과도 커뮤니케이션 가능
7. 테마 변경 가능
8. 사내 테마 일괄 적용 가능

▼ 특징 및 장점

1. 사용법이 간단하여 별도의 교육이 필요없습니다.

2. 쉽게 이벤트를 공유할 수 있습니다.
3. 동료들의 근황을 손쉽게 알 수 있습니다.
4. 거리적으로 먼 직원(출장, 파견, 다른 근무지)과도 편리하게 소통이 가능합니다
5. 정보를 실시간으로 공유 가능 합니다.
6. 게시판보다 글을 작성하는 데 부담이 적습니다.
7. 여러 사람의 의견을 쉽게 얻을 수 있어 토론/토의, 아이디어 공유가 쉽습니다.
8. 업무나 특정 주제 단위로 협업이 가능 합니다.
9. 업무의 흐름을 알기가 쉽습니다.

▼ Yammer 분석 철학

Yammer 분석가는 각 개별 프로젝트의 가치를 지속적으로 고려하도록 교육을 받았습니다. 그들은 시간에 대한 수익을 극대화하려고 합니다. 분석가는 일반적으로 훨씬 적은 시간을 투자하는 것을 의미하는 경우 문제에 대한 덜 정확한 솔루션을 선택합니다.

그들은 또한 모든 것이 회사 전체에 미치는 영향을 고려하도록 배웁니다. 여기에는 우선 순위를 지정할 프로젝트를 선택하는 것과 같은 높은 수준의 의사 결정이 포함됩니다. 또한 분석가가 메트릭에 대해 생각하는 방식에도 영향을 미칩니다. 제품 관련 결정은 항상 제품별 사용 지표(예: 다른 사용자의 프로필을 본 횟수) 외에 핵심 참여, 유지 및 성장 지표에 대해 평가됩니다.

Yammer의 분석가는 발생하는 제품 및 비즈니스 문제를 분류하는 일을 담당합니다. 대부분의 경우 이러한 문제는 경영진과 관리자가 매일 확인하는 주요 메트릭 대시보드를 통해 표면화됩니다.

2. 데이터

살펴볼 테이블 4개

• Table 1: Users

사용자 계정에 대한 설명 정보와 함께 사용자당 하나의 행이 포함되어 있다.

user_id:	A unique ID per user. Can be joined to user_id in either of the other tables.
created_at:	The time the user was created (first signed up)
state:	The state of the user (active or pending)
activated_at:	The time the user was activated, if they are active
company_id:	The ID of the user's company
language:	The chosen language of the user

• Table 2: Events

이벤트당 하나의 행이 포함되며 여기서 event는 사용자가 Yammer에서 수행한 작업을 의미한다. 이러한 이벤트에는 로그인 이벤트, 메시징 이벤트, 검색 이벤트, 사용자가 가입 유입경로를 통해 진행하는 동안 기록되는 이벤트, 수신된 이메일 관련 이벤트가 포함되어 있다.

user_id:	The ID of the user logging the event. Can be joined to user_id in either of the other tables.
occurred_at:	The time the event occurred.
event_type:	The general event type. There are two values in this dataset: "signup_flow", which refers to anything occurring during the process of a user's authentication, and "engagement", which refers to general product usage after the user has signed up for the first time.
event_name:	The specific action the user took. Possible values include: create_user : User is added to Yammer's database during signup process enter_email : User begins the signup process by entering her email address enter_info : User enters her name and personal information during signup process complete_signup : User completes the entire signup/authentication process home_page : User loads the home page like_message : User likes another user's message login : User logs into Yammer search_autocomplete : User selects a search result from the autocomplete list search_run : User runs a search query and is taken to the search results page search_click_result_X : User clicks search result X on the results page, where X is a number from 1 through 10. send_message : User posts a message view_inbox : User views messages in her inbox
location:	The country from which the event was logged (collected through IP address).
device:	The type of device used to log the event.

• Table 3: Email Events

이메일 전송과 관련된 이벤트가 포함되어 있다. 이벤트 테이블과 구조가 유사하다.

user_id:	The ID of the user to whom the event relates. Can be joined to user_id in either of the other tables.
occurred_at:	The time the event occurred.
action:	The name of the event that occurred. "sent_weekly_digest" means that the user was delivered a digest email showing relevant conversations from the previous day. "email_open" means that the user opened the email. "email_clickthrough" means that the user clicked a link in the email.

• Table 4: Rollup Periods

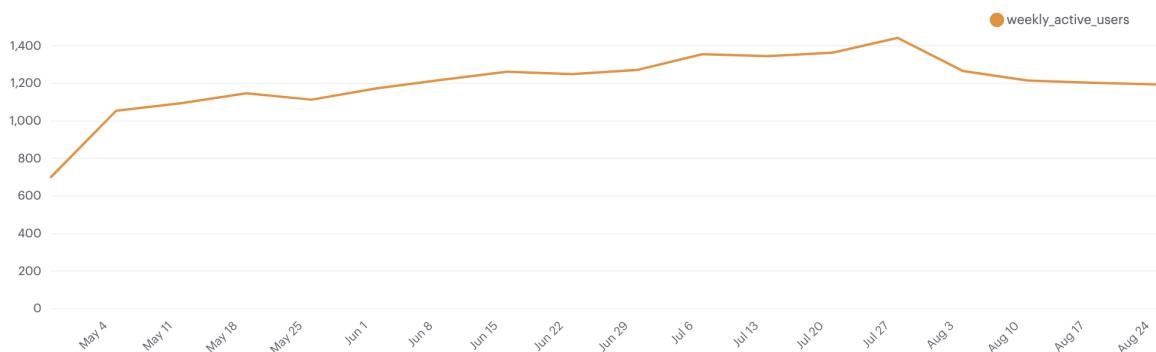
롤링 기간을 만드는 데 사용되는 조희 테이블로, INTERVAL() 함수를 사용할 수 있지만 롤링 기간을 만드는 것은 종종 이와 같은 테이블에서 가장 쉽다. 작성하는 쿼리에서 반드시 이 테이블을 사용할 필요는 없다.

period_id:	This identifies the type of rollup period. The above dashboard uses period 1007, which is rolling 7-day periods.
time_id:	This is the identifier for any given data point — it's what you would put on a chart axis. If time_id is 2014-08-01, that means that it represents the rolling 7-day period leading up to 2014-08-01.
pst_start:	The start time of the period in PST. For 2014-08-01, you'll notice that this is 2014-07-25 — one week prior. Use this to join events to the table.
pst_end:	The start time of the period in PST. For 2014-08-01, the end time is 2014-08-01. You can see how this is used in conjunction with pst_start to join events to this table in the query that produces the above chart .
utc_start:	The same as pst_start, but in UTC time.
utc_end:	The same as pst_end, but in UTC time.

3. 문제 정의

문제 상황

Weekly Active Users



2014년 9월 2일 화요일을 현시점으로 볼 때, 위의 WAU 그래프에서 7월 27일을 기점으로 하락세를 보이고 있다. 조금씩 상승하던 그래프의 하락 원인에 대한 분석이 필요하다.

4. 가설 설정

1. 8월 여름 휴가 시즌이라 리텐션이 줄었을 것이다.
2. 어딘가 서비스 기능이 잘못/손상되었을 것이다.
3. 로그가 기록이 안된 것일 것이다.

▼ 더 많은 가설

• 휴일:

Yammer와 같은 업무용 응용 프로그램을 사용하는 사람들은 휴일에 참여하는 비율이 낮을 수 있습니다. 한 국가의 참여도가 다른 국가보다 훨씬 낮은 경우 이것이 원인일 수 있습니다.

- **손상된 기능:**

응용 프로그램의 일부가 손상되어 사람들이 사용하지 못할 수 있습니다. 애플리케이션의 다른 부분이 메트릭에서 다르게 표시되기 때문에 정확히 지적하기가 조금 더 어렵습니다. 예를 들어 가입 흐름의 문제가 발생하여 새 사용자가 Yammer에 가입하지 못하는 경우 성장도 저하됩니다. 모바일 앱이 불안정하고 충돌하는 경우 해당 장치 유형에 대해서만 참여가 중단됩니다.

- **손상된 추적 코드:**

이벤트를 기록하는 코드 자체가 손상되었을 수 있습니다. 특정 유형의 이벤트가 완전히 0으로 떨어지고 손상된 기능을 배제한다면 이것이 가능합니다.

- **봇의 이상 트래픽:**

대부분의 주요 웹사이트는 봇의 많은 활동을 확인합니다. 봇이 사이트와 상호 작용하기 어렵게 만들 수 있는 제품 또는 인프라의 변경은 참여를 감소시킬 수 있습니다(봇이 실제 사용자처럼 보인다고 가정). 패턴이나 특정 이벤트를 통해 봇과 유사한 동작을 식별해야 하기 때문에 결정하기가 까다롭습니다.

- **귀하의 사이트에 대한 트래픽 차단:**

인터넷 서비스 제공업체가 귀하의 사이트를 차단할 수 있습니다. 이것은 전문적인 응용 프로그램에서는 매우 드물지만 그럼에도 불구하고 가능합니다.

- **마케팅 이벤트:**

예를 들어 슈퍼볼 광고는 제품 가입을 엄청나게 증가시킬 수 있습니다. 그러나 일회성 마케팅 공세를 통해 입력한 사용자는 예를 들어 친구가 추천한 사용자보다 낮은 비율로 유지하는 경우가 많습니다. 차트는 롤링 7일 기간을 사용하기 때문에 1주일 동안 높은 참여로 등록되고 그 다음 주에는 참여가 크게 감소한 것처럼 보일 것입니다. 대부분의 경우 이를 확인하는 가장 좋은 방법은 마케팅 부서의 누군가에게 최근에 큰 일이 있었는지 물어보는 것입니다.

- **잘못된 데이터: 잘못된 데이터**

를 기록하는 방법에는 여러 가지가 있습니다. 예를 들어 대부분의 대형 웹 앱은 QA 데이터를 프로덕션 데이터와 분리합니다. 어떤 식으로든 QA 데이터는 프로덕션 데이터베이스로 들어갈 수 있습니다. 극소수의 사용자로부터 기록된 추가 데이터로 나타날 가능성이 높기 때문에 이 특별한 경우에는 문제가 되지 않을 것입니다.

- **검색 크롤러 변경:**

많은 트래픽을 받는 웹 사이트의 경우 검색 엔진이 색인을 생성하는 방식이 변경되면 트래픽이 크게 변동될 수 있습니다.

5. 실험 설계 및 검증

1. 신규가입자 분석
2. 가입 코호트별 WAU 분석
3. 디바이스별 WAU 분석
4. 이메일 인게이지먼트 분석

6. 결과

1. 신규 가입자는 WAU 차트에서 7월 27일에 소폭 감소 후 평소 수준으로 복귀했다.
2. 10주 이전 가입 코호트에서 WAU 감소가 확인되었다.
3. 디바이스 별로는 전체적으로 감소가 확인되었으나 태블릿과 모바일을 통한 engagement가 크게 감소했다. 태블릿과 모바일 환경에서 14년 8월 당시 업데이트 또는 변경에 대한 확인이 필요하다.
4. 이메일 분석을 통해 weekly digest 이메일 내 링크 클릭 수, 이메일 수신 5분 이내 링크 클릭 유저 비율 모두 직전 주차 대비 감소한 것이 확인되므로 이 간격을 낮추는 방향으로 조정할 필요가 있다.