Yammer Data Challenge

Yammer is a social network for communicating with coworkers. Individuals can **share documents,**  **updates,** and **ideas** by **posting them in groups.** It is free to use but companies pay license fees to access **administrative controls,** including integration with **user management systems** like **ActiveDirectory.**

My job: Analytics team in Engineering Org.

Problem: There is a **drop in user engagement** what do you **recommend to the Eng. Org.?**

The Data come from four tables:

**Users, Events, Email Events,** and **Rollup Periods**

Users: Unique user ID and descriptive info about the user’s account

Events: Each row is an event, or single action, by user (user ID)

Email Events: Similar to events but just related to the sending of emails

Rollup Periods: Less necessary, used to create rolling time periods

Some general initial ideas and thoughts.

We have user engagement time series. Can we overlay other behaviors? Like overlay Avg number of emails sent. Be careful with causation over correlation here..

Other ideas, label before the peak and after the peak as separate classes. Train a classification model and determine the features that correlate with user engagement increasing vs decreasing…

For example, is user engagement decreasing within a subset of the users? Etc.

Let’s create a notebook…

Idea: similar to Jeremy’s project, let’s use catboost to classify the data, train and determine the important features that distinguish the increasing user section to the decreasing.

Missing data: activation time. Could either mean the user was never activated, or that the data are missing.

One plot could be signup time vs time to activation for EDA Figure 1.

Activation rate flatlined after august….. Where did this come from?

Plot avg number of engagements per login

Plan ----

Limit the start of the time to 2013/9 to END

New hard limit if we use the events file  
user\_type is nan if the user never activated….

**Constant number of sign ups after the date too**

Label each user at before or after 2014/9 and whether they activated. Here we are going to focus on the user table, but we can fill in some info, like device, location, avg. engagement per login,

NEW QUESTION: What caused users to stop signing up?