SQL Case Study: Validate A/B Test Results

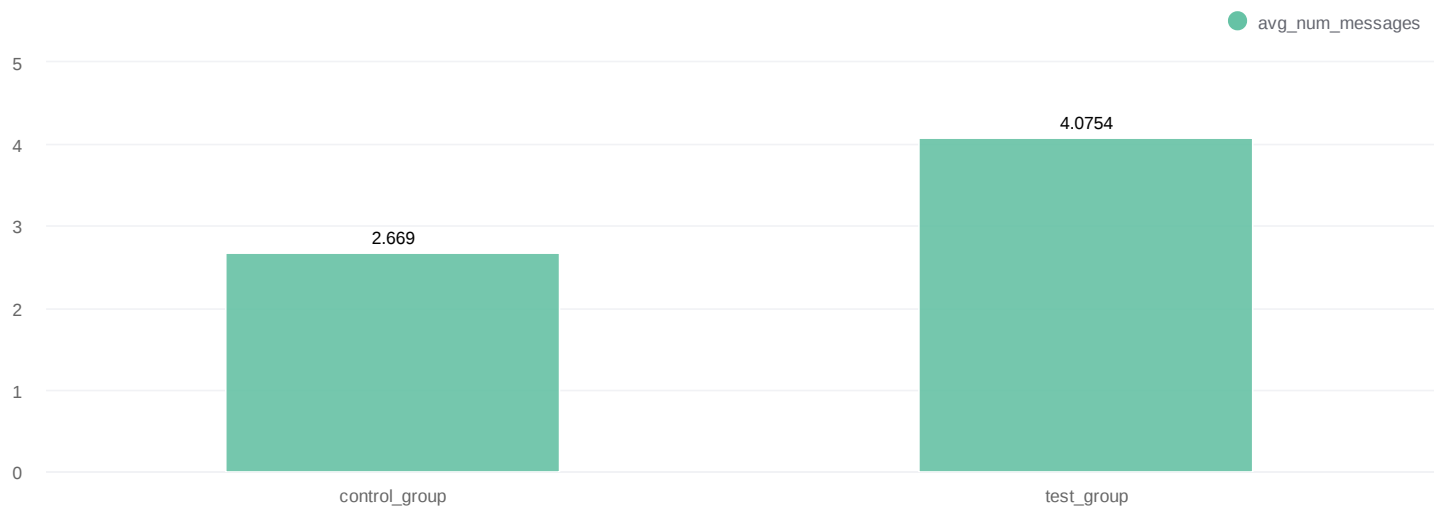
Background

Yammer is a social network for communicating with coworkers. The product team is continuously developing new features and looking for ways to improve existing ones. Recently, the product team improved Yammer's core "publisher", the module at the top of a Yammer feed where users type their messages. To test this feature, the product team ran an A/B test from June 1 through June 30. During this period, some users who logged into Yammer were shown the old version of the publisher (the “control group”), while other other users were shown the new version (the “treatment group”).

Problem

On July 1, the results from the A/B test came out. The test group's average message posting rate is 50% higher than that of the control group. Determine whether the result is valid and whether the feature is the real deal or too good to be true.

Average Number of Messages Sent per User

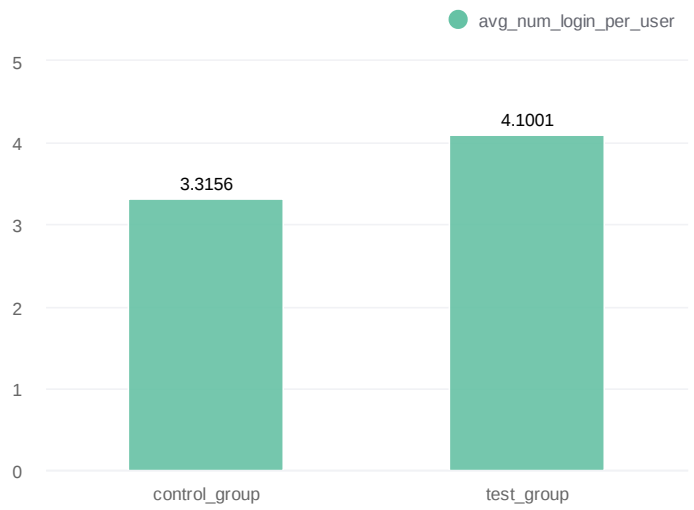


Measure the A/B Test using other metrics

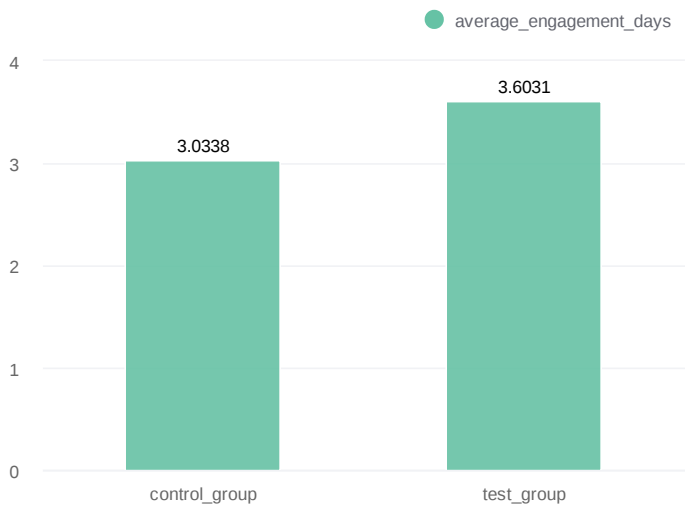
The test’s group’s average number of login per user is higher than that of the control group. In addition, the test group’s average length of the engagements is also higher than that of the control group. This suggests that not only are users sending more messages, but they are also engaging (i.e. logging-in) with the product longer and more frequently.

SQL Case Study: Validate A/B Test Results

Average Number of Login per User



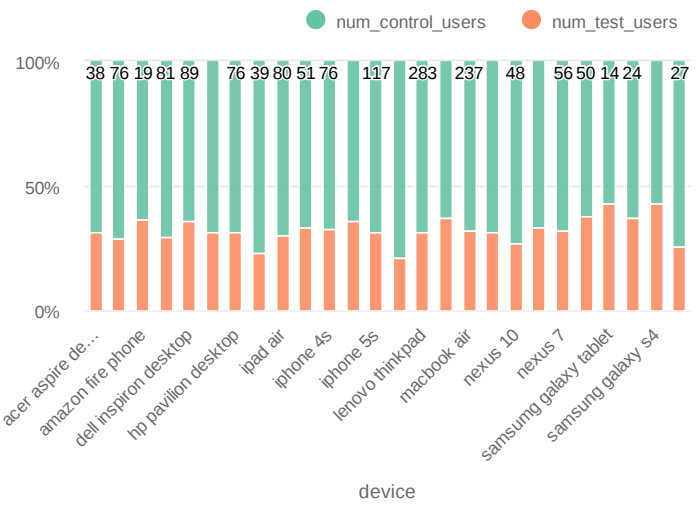
Average Days of Engagement



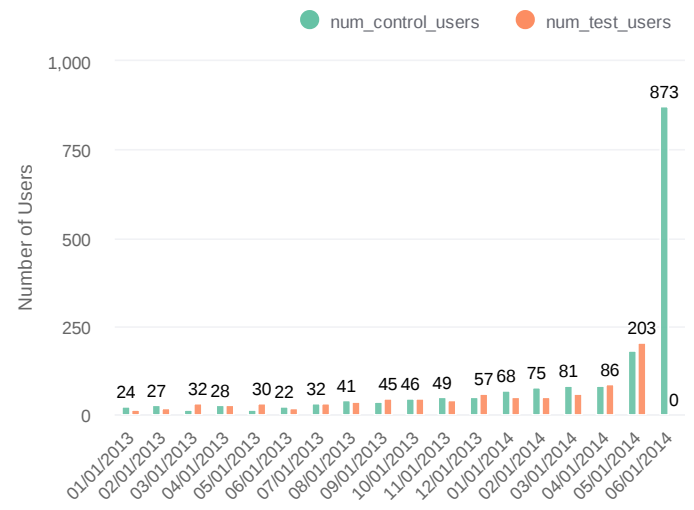
Experiment Group

The A/B test suffers from a methodological error. The test assigned all of the new users who activated in June 2014 into the control group while the test group only consists of existing users. The problem is that the test was conducted during June. The new users, especially those who joined later during the month, would be expected to post less than existing users given their shorter exposure to Yammer. Including the new users in the control group would lower the group's overall posting rate.

Experiment Group by Device Type



Experiment Group by Activation Month

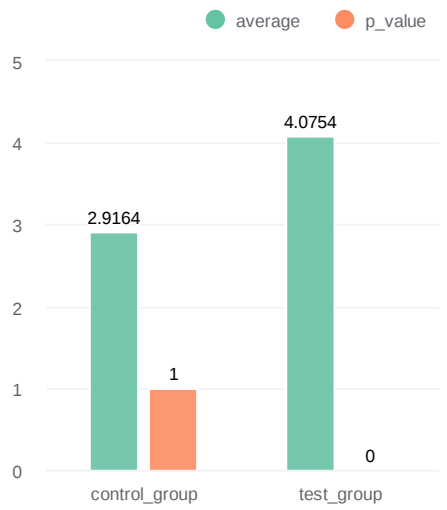


Metric, Excluding New Users

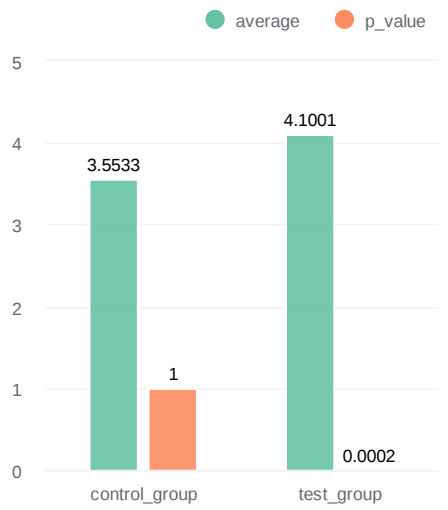
By excluding the new users who joined during the A/B testing period, I am able to test for novelty effects. Users familiar with Yammer might try out a new feature just because it's new, temporarily boosting their overall engagement. For new users, the feature isn't "new," so they are much less likely to use it just because it is different.

SQL Case Study: Validate A/B Test Results

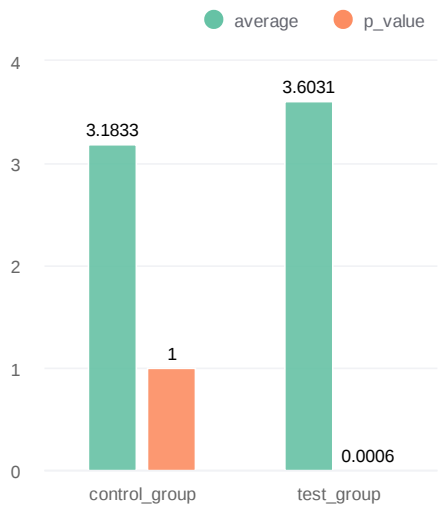
Avg Messages Sent per Existin…



Avg Login per Existing User



Avg Days of Engagement per E…



The test group's average number of messages sent per existing user is 28% higher than that of the control group.
The p-value of 3.199999998315661e-8 is less than 0.05, which suggests there is a statistically significant difference between the test group and the control

The test group's average number of login per existing user is 13% higher than that of the control group.
The p-value of 0.0002 is less than 0.05, which suggests there is a statistically significant difference between the test group and the control group.

The test group's average number of days of engagement per existing user is 11.65% higher than that of the control group.
The p-value of 0.0006 is less than 0.05, which suggests there is a statistically significant difference between the test group and the control group.

Conclusion

Overall, the A/B test results stayed strong even after excluding the new users. The p-value from all three metrics suggests there is a statistically significant difference between the control and test group, using the 0.05 threshold. The product feature improvement produced statistically significant results, e.g. higher average posting rate, login, and days of login.