Social Media Posting Model User Analysis

Adam Bruce

2024-02-18

Data Background

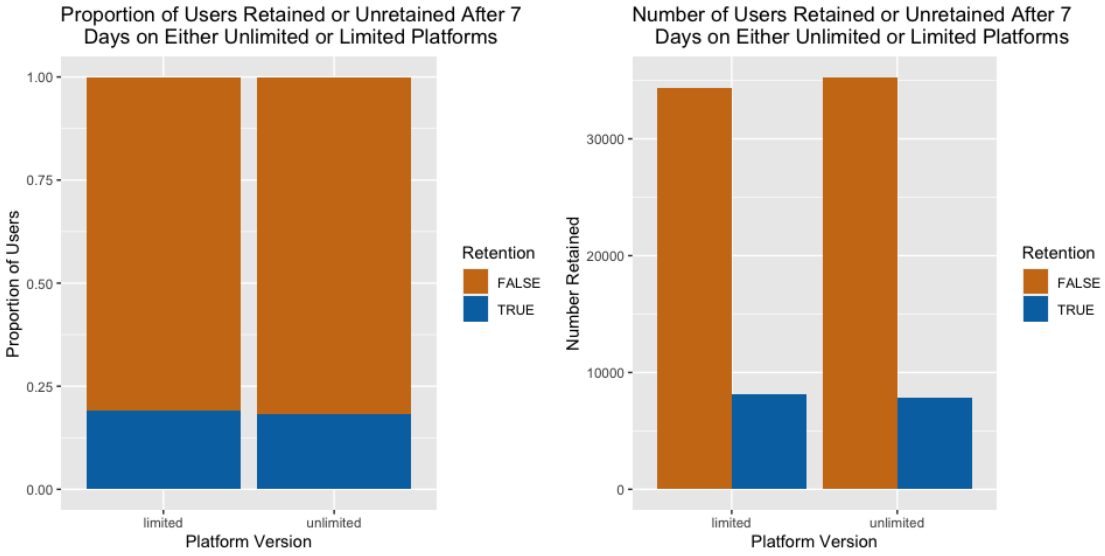
The data set `platform_retention.csv` used in this analysis contains data on a sample of new users of the social media platform. Collectively, users were assigned to one of two user engagement model groups. These included either unlimited, the companies current model, or limited posting groups.

Primarily, this study aims to identify whether limiting user posts increases 7-day retention when compared to the unlimited posting model currently used. However, we also investigate if, between each model, site visits by day 7 vary, 1-day retention vary, or if other relationships among recorded variables exist and warrant further testing going forward.

Exploratory Analysis

In total, 4,516 users were missing data on their retention after one and seven days. It is impossible to infer retention for these users, so they were dropped from final analysis. The result was 85,673 users to study.

Investigation: Does limiting user posts increases 7-day retention when compared to the unlimited posting model?

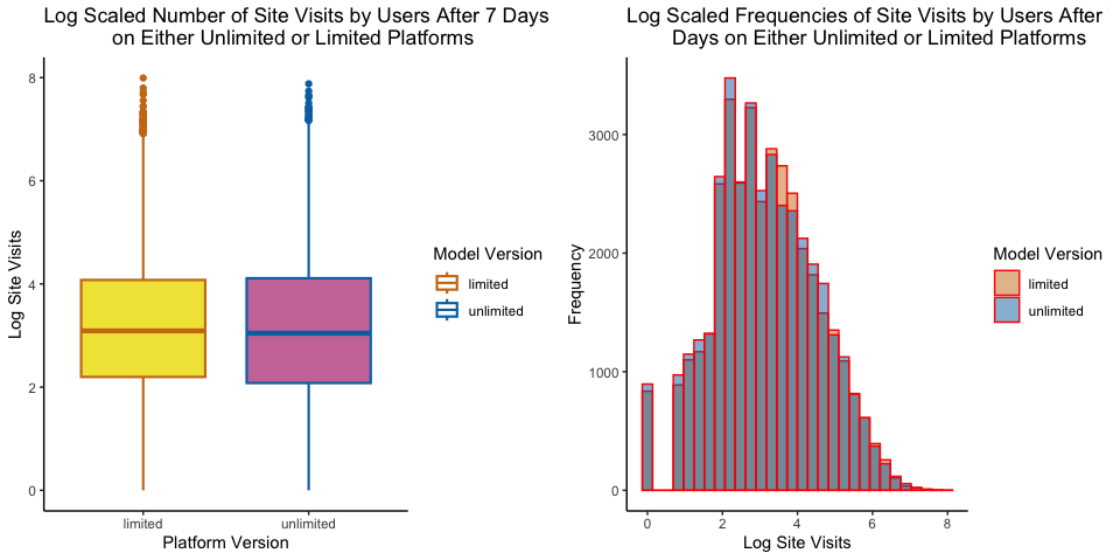



```
##
##         limited unlimited
##   FALSE   80.89     81.76
##   TRUE    19.11     18.24
```

A segmented bar plot (left) shows the proportion of the users within each platform group. The results show that a majority of individuals were not retained in both groups (retention = FALSE). Most importantly, the plot shows little difference in the limited and unlimited model retention after seven days. The same outcome is observed in the grouped bar plot (right). However, this plot shows the total number of users retained or unretained after seven days for each model. This plot is arguably better for explaining the small difference observed as a small gap between the "FALSE", unretained users can be seen.

Overall, the unlimited model does have a slightly higher percentage of users unretained at 81.76% compared to the limited model at 80.89%. Therefore, it is plausible that the new, limited model, could increase retention (TRUE = 19.11%) over the unlimited model (TRUE = 18.23%). Though, it seems unlikely given the large sample size and small difference in the groups.

Investigation: Does the number of site visits by day 7 differ between the two posting models?



version	Version Sample Mean	Version Sample SD	Upper 95% CI	Lower 95% CI	sample
limited	51.85121	102.6825	52.84733	50.85509	42504
unlimited	51.85093	104.2465	52.85442	50.84745	43168

One user had a purported 49,859 site visits after seven days on the limited posting model. This would average to over 7,100 visits per day, which is unreasonable, especially considering the next highest user was less than 3,000. Therefore, this user was removed before investigating this question.

Plotting revealed it was necessary to log transform `site_visits_7` in order to get useful visual results. Boxplots (left) of site visits after seven days show almost no difference between the limited and unlimited models. Median visits, depicted by bold lines in the center of each plot, are identical between groups at 17 (log(17 visits) = 2.83). Histograms (right) further show nearly no difference between groups. Each curve represents

frequencies of user site visits after seven days. Nearly complete overlap between models (gray hue) indicates the users in each group visited the site nearly the same number of times after seven days. These histograms are arguably a better insight into the question at hand because they are an easy visual of the nearly identical overlap between the two interest groups. No single color (model) sticks out, as the whole curve is almost entirely gray.

Finally, the table (below) shows the mean site visits after seven days between groups are nearly identical, shown by `Version Sample Mean`, as discussed (limited = 51.8512, unlimited = 51.8509). These means deviate, `Sample SD`, among models about the same, and the 95% confidence intervals overlap. Confidence intervals tell the range of average site visits after seven days of all users of the app for both models. We are 95% confident all users of the limited model will visit the site between 50.86 and 52.85 times after seven days on average, while all users on the original unlimited model will visit the site between 50.85 and 52.85 times after seven days on average. The fact that both 95% confidence intervals overlap greatly supports the idea that the two models do not vary in user visits to the site after seven days.

Testing: Is the number of users retained after 7 days different between the two posting models?

Model	Users Retained 7 Days	Total Users	Percent Retained	Conclusion
Limited	8121	42504	19.11	Greater
Unlimited	7872	43168	18.24	Less

Difference in proportions testing indicates that there is evidence of a difference in site retention after seven days between the limited and unlimited models. Particularly, there is evidence to indicate that site retention is greater in the limited posting model compared with the unlimited model (See Table). This provides support for the implementation of a limited posting model on the current platform. Overall, this evidence was obtained at a significance level of 0.05, indicating 95% confidence in the results of this finding (P = 0.0005374). However, there is a 5% our test output is incorrect, but this would have relatively little impact on the company if true.

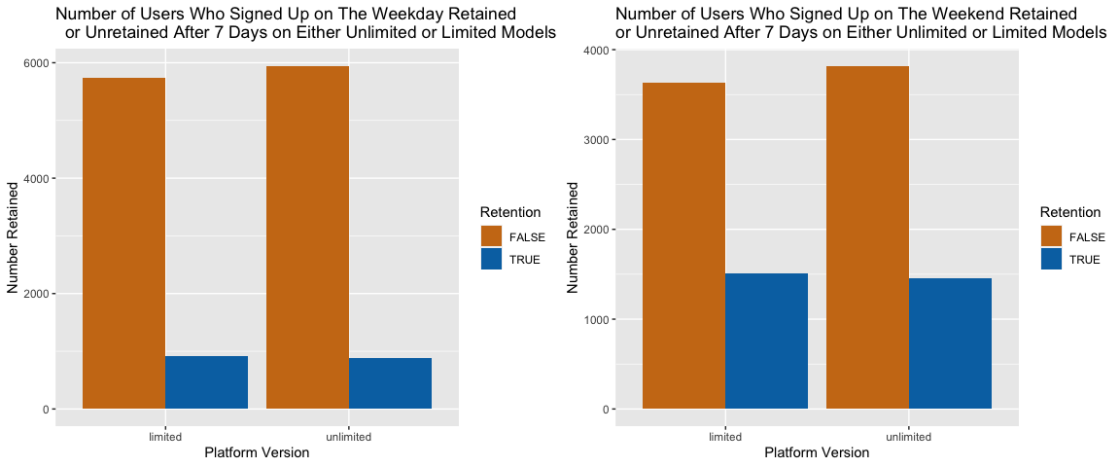
Additionally, we are 95% confident that the true proportion of seven day retention between the limited and unlimited posting models is between 0.003488 and 0.01393. This 95% confidence interval gives a reasonable range of possibilities for the true difference between versions, and is therefore a justifiable test level for this analysis.

Final Analysis of Secondary Research Questions and Potential Future Pathways

Model	Retained After 1 Day	Total Users	Percent Retained	Conclusion
Limited	19063	42504	44.85	No Difference
Unlimited	19117	43168	44.29	No Difference

A difference in proportions test for retention after one day indicates there is no evidence of a difference in site retention after one day between the limited and unlimited models. This provides insights regarding how users initially use the site. Particularly, it indicates that

regardless of the model, a large amount of users will be unretained initially, which is not surprising. Overall, this evidence was obtained at a significance level of 0.05, indicating 95% confidence in the results of this finding (P = 0.09633). However, there is a 5% chance our test results are incorrect. Additionally, we are 95% confident that the true proportion of one day retention between the limited and unlimited posting models is approximately zero.



Finally, an investigation into whether retention after seven days depended on if the user signed up on a weekday (Monday through Thursday) versus a traditional weekend (Friday through Sunday) yielded some intriguing results. Grouped bar plots (above) However, show the total number of users retained is higher in the limited model for both the weekend and weekday. For users who sign up during the weekday, 13.74 percent were retained in the limited model compared to 12.85 percent in the unlimited group. Meanwhile, for users signing up on the weekends, 29.31 percent were retained in the limited model and 27.61 were retained in the unlimited model. However, the difference is greater in the limited model for users signing up on the weekend at 1.7% compared to users signing up during weekdays at only 0.89%.

Difference in proportion tests formally indicate that for users who sign up on the weekends, there is evidence to indicate that site retention after seven days is greater in the limited posting model compared with the unlimited model (P = 0.02724). On the other hand, testing for users signing up during the weekday found no significant evidence that the limited model retention was greater after seven days than the limited approach (P = 0.06499). Both tests were performed at a significance level of 0.05, indicating 95% confidence in their results.

Overall, this investigation indicates that while introducing the new limited model on our platform would increase user retention in the long term. However, going forward, we may want to allocate more resources towards marketing to potential users on weekends over weekdays to further increase retention. How we attack this marketing would be a great subject for further research, as perhaps there are other factors influencing retention that we could utilize in advertising campaigns going forward.