

Quora New Design Analysis

Overview of Problem: Do the people using New Design show evidence of increased engagement? Do we have evidence that if there is increase, it is because of the New Design?

Issues with Data:

Majority of people participating in the experiment are males ~50%, and non readers are 75% of the population. Additionally, people in control group are 4 times those in treatment group. These people were monitored for 6 months before experiment start but only 5 months after experiment began. Additionally, new users came into the system very late (until January 2019), and we do not have sufficient data on their old site behavior.

```
before experiment began, # of males: (27592, 5)
before experiment began, total: (49644, 5)
post experiment, # of males: (26125, 5)
post experiment began, total: (46605, 5)
```

Figure 1. Disproportionate # of males in data (~50%)

```
before experiment began, # of non readers: (35973, 5)
before experiment began, total: (49644, 5)
post experiment, # of non readers: (34462, 5)
post experiment began, total: (46605, 5)
```

Figure 2. Disproportionate # of non readers in data (~75%)

Approach:

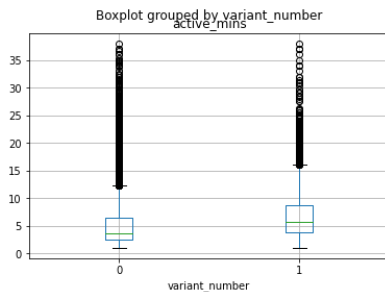


Figure 3. Active minutes of Treatment group seem higher

Average active minutes per user was taken for each variant group. Due to enormous (4:1) ratio of the size of the two groups, **Welch t-test** was performed & **Confidence Interval** found for testing validity of apparent increase in active minutes of Treatment group (Figure 3.)

Engagement data from 6 months before experiment began was combined to gauge an individual's change in activity better. We notice a considerable increase in mean active minutes and so **Paired t-test** was performed on the difference of average active minutes between old and new site for each user.

	active_mins_pre	active_mins_post
mean	4.870482	6.924268
max	29.000000	38.000000

Figure 4. Treatment group metrics

	active_mins_pre	active_mins_post
mean	5.184737	5.368229
max	31.000000	38.000000

Figure 5. Control group metrics

To check for lurking variables or confounding variables, data was split by gender and user types. Pre and post experiment combined data was adjusted to show average activity of users per day, creating a time-series over 11 months.

Results:

With a small p-value of $2e-180$ for Welch, p-value of 0 for paired test and a confidence interval of (1.99, 2.11), mathematically, we are safe (~100%) to accept that average active minutes per user is higher for treatment group on new design than old design and also higher than control group.

..Continued

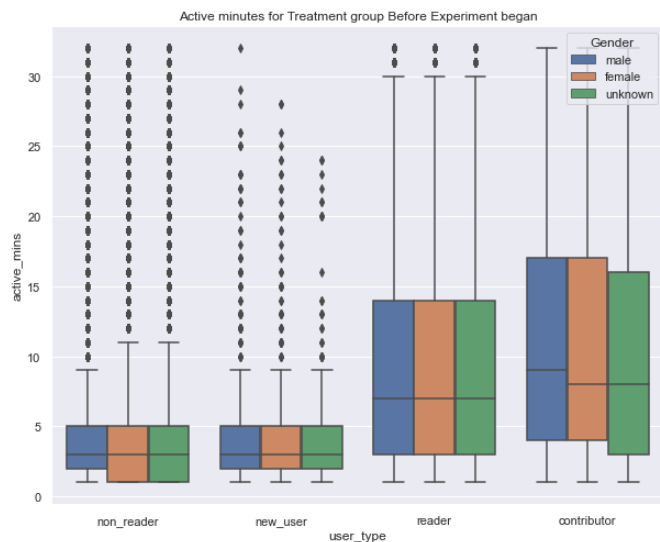


Figure 6. Active minutes for Treatment group on old design

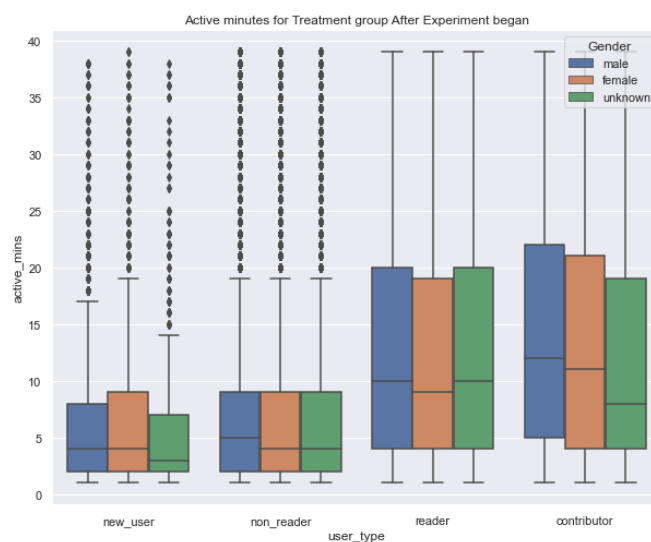


Figure 7. Active minutes for Treatment group on new design. The average treatment population shows increase in engagement (not seen in control). **Epecially females** have shown an increase among new users. Contributors show increase in reading times even in Control group so, we cannot associate the apparent increase in Fig. 7 to new design

Conclusion

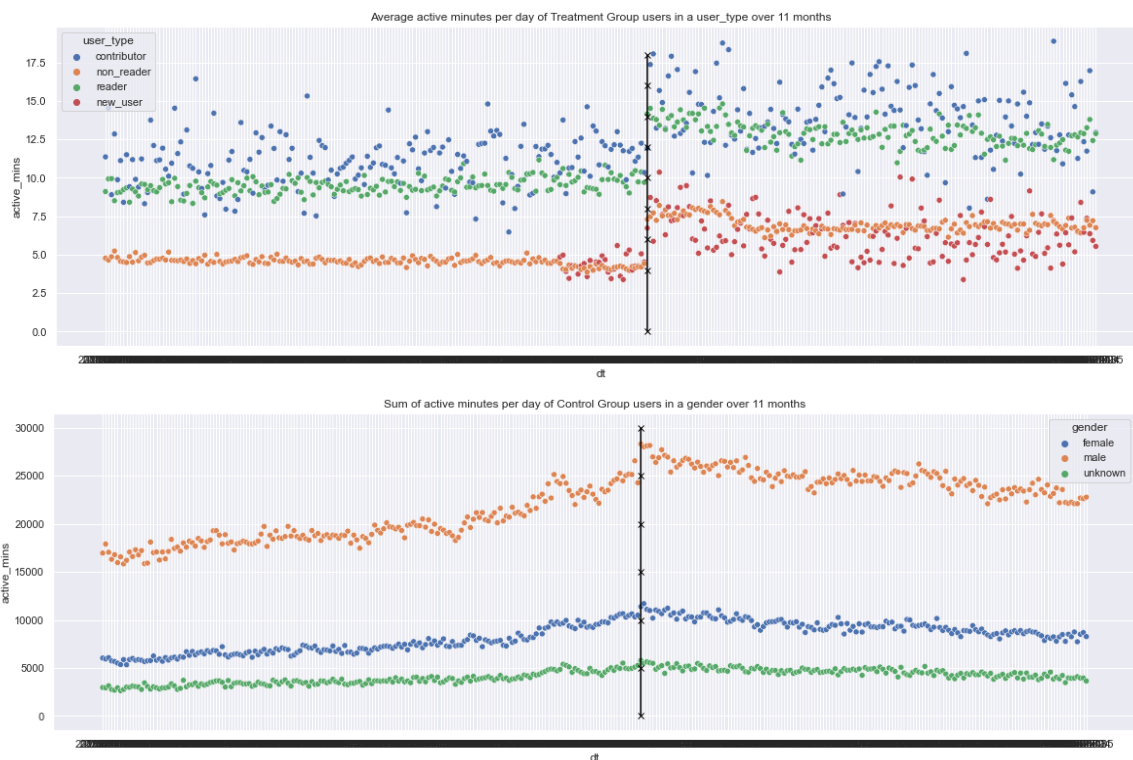


Figure 8. Average Active minutes per day of Treatment group -by user type - over 11 months. (Split: Experiment began at Feb 6th 2019)

Figure 9. Average Active minutes per day of Treatment group - By Gender - over 11 months. (Split: Experiment began at Feb 6th 2019)

The sudden spike (whether for a user type or for gender) when experiment began is followed by a steady decrease, so was this spike because people had trouble navigating and HAD to spend more time unwantedly(another lurking variable here)?

Due to highly unbalanced data (males vs females, new users vs other types), our predictions may fail in production. So I would recommend to get more people onboard for the experiment. I also would like to get feedback data from these participants finding the reason for their increased activity.

I can not recommend production deployment with this data.

--- X ---