27.2.3 Relax Take-Home Challenge Report

Where Are We?

Up to this point, I was not able to pinpoint any major factors that impact the adoption rate more than the other however we still made some observations and progress.

I have seen the variables from the user dataframe do not have a great predictive power on the user adoption (based on the definition given in text) by themselves.

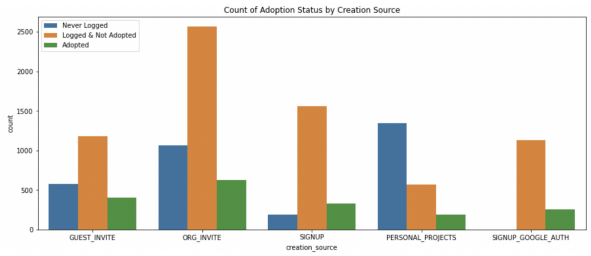
Some observations that I can make are

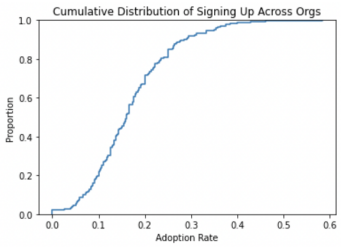
i) **Personal Projects may not be the best way** to gain adoptees

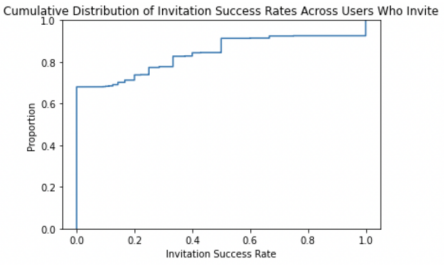
ii) **10% of companies have a larger than 0.30 rate of adoption**

iii) **There are a set of people who invite others that have an approximately 50% rate of success.**

Besides, expanding upon the last two paths, I believe I should make better use of the time-based engagement data provided. [Better graphs are in the notebook]







.

Where to Go from Here?

1) Further Exploration: Making Better Use of Engagement/Login (Time-based) Data

What this analysis seems to miss is capturing the dynamics of events - both on a per person and per organization basis.

On a per person basis, one could use the creation_time of a users account as a normalizing point for all users and proceed to ask some conditional questions across people cross-sectionally. Questions such as:

-**What is the probability (looking from creation_date) that a person will adopt conditional on the fact that there were 1 {2, or 3 or more} people already adopted across the organization?** If the person eventually adopts, what is the average number of days? And does this number change if there were more people signed up at the organization? If we see some sort of a network effect here, we could make efforts to accelerate the network effect.

-What has been the average number of logins (or logins per week) per person before they adopted? This sum/average could be compared to the group that didn't adopt. At the end of the day, if the difference is meaningful, one could imagine an A/B test to increase the logins in the population (taking care not to alienate people)

-What is the probability of adoption conditional on the person being close to adoption before (defined by logging say 2 times per week)? Then, one could run an A/B test on that population to see if the conversion probability will move up.

2) A Modelling Path - Ensemble Models / Logistic Regression

A logistic regression or an ensemble model could be used incorporating some more extra features (perhaps some of the ones used above) to give us the feature importances associated with each of those factors. Results of those importances could be further used to guide the company to generate better hypothesis regarding what might positively improve adoption