Data science challenge summary: Relax Inc.
Problem statement: Identify factors that predict future user adoption of the software. An adopted user is identified as have at least three active sessions within at least one week.
**Data:** user information (Table1) and user log-in time stamps (Table2).

Approach:
We considered six categories of factors that may affect user adoption:
1) Invited by other users or not
2) Creation source (guest invite, organization invite, etc.)
3) Whether opted in to mailing list or enabled for marketing drip (showing stronger interests)
4) Type of organizations the user works at (org_id)
5) Email domain (do users of certain domains have higher adoption rate?)
6) Being active first 30 days after sign-up and last 30 days before last session.
We used visualization, a logistic regression model and a random forest classifier to evaluate the importance of these factors.

Results:
1) Among the categorical variables, creation source and email domains are more predictive on user adoption than the others. Users who signed up via "google authentication" or "guest invite", and those who use Hotmail or Gmail are more likely to stay active.
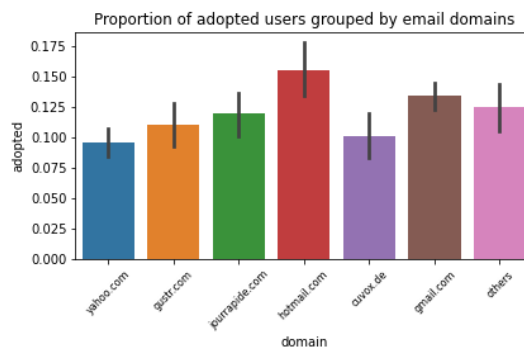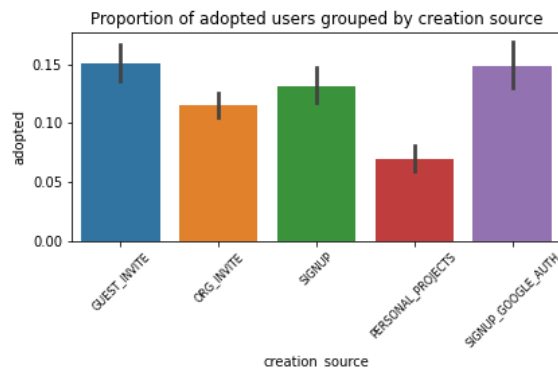
2) Logistic regression and random forest models had similar results, by which the two numeric variables are most predictive: being active first 30 days after sign-up and last 30 days before last session.
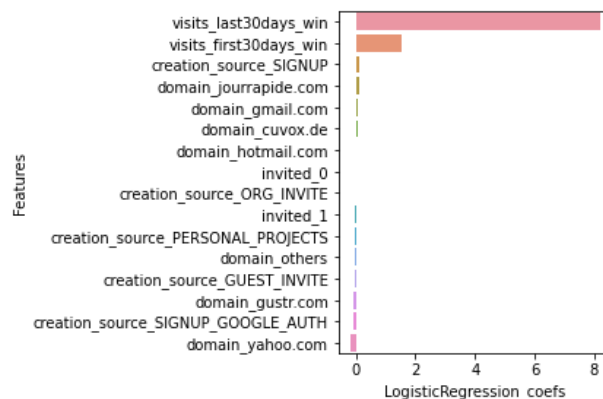
What's next:
1) Group organization types
2) Delve deeper into each creation source category and find trends of active sessions that link to adoption.
3) Investigate: does email functionality/convenience (such as google auth) help to keep users active?