Relax Data Science

**November 9, 2020**

# OVERVIEW

Relax Data Set summary and evaluation. Attempting to use the two tables to generate a predictive model to estimate users that will actually utilize the software. 12,000 records of user information and a log table of user sessions.

# GOALS

1. Is there a predictive set of features that will positively reflect the adoption of the software by a user?

# Analysis

The identification of an ‘adopted’ user is the one-time fact of the user logging into the system three times within a single 7 day period. Evaluation of the log using a rolling 7 day period found 1602 users met or exceeded that threshold.

Utilizing that list of ‘adopted users’ and inserting into the user information table provided an opportunity to evaluate the user information to predict the ‘adoption’ rate or the possibility of adoption by a user.

Simple counts of user information between the users signing up for the mailing list or the marketing drip campaigns appeared to be a similar fraction of the whole dataset as the number of adopted users. The number of users signed up for the mailing list was 2994, the number signing up for marketing was 1792 and the number of adopted users was 1602.

# Features

However, utilizing correlation analysis found that there was little correlation between any of the user information features.

Doing a cluster analysis found only one cluster.

Doing PCA analysis on the user features found nothing very interesting.

Finally, doing a cross-validation check of the features using LassoCV found that there was only one negatively correlated ‘last\_session\_creation\_time’, which makes sense since ‘adopted’ users would be the ones most frequently using the system.

# Further Options: