**Identifying factors that help predicting future user adoption**

The available data includes information of 12,000 users who signed up for the product in the last two years such as email, how they created their account, if they opted into the mailing list, if they enabled marketing, the organization they belonged to, and the user who invited them. We also have login information of all users and the time showing when they logged in.

Defining an "adopted user" as a user who has logged into the product on three separate days in at least one seven­ day period**,** the company is interested in identifying which factors can help predict future user adoption.

To understand the effect of each feature on the target (adoption) I first performed exploratory data analysis (EDA) and plotted the relation between each feature and the target. The results showed that some features such as “enabled for marketing drips” and “opting into the mailing list” have no effect on user adoption. EDA showed that for some organisation IDs there is a higher user adoption compared to the others. I separated those organizations and put the rest in another category.

I also created a few new features and investigated their effect on predicting the user adoption. I extracted the domain information from each user's email address. I used login events for each user to extract information such as the duration between each user’s first and last login, the average time between each log, the maximum time between each login, and the total number of logins for each user. I also looked into the number of logins for each month, day of the week, and hour of the day by all users to see if any of those can help in predicting user adoption.

I finally used all the features that looked important based on EDA to predict user adoption using a few different models and compared their performance.