Data Analysis Interview Challenge - Relax Inc.

We wish to predict future user adoption:

```
adopted                       1.000000
usage_length                  0.086207
org_id                        0.066995
opted_in_to_mailing_list      0.008838
enabled_for_marketing_drip    0.006578
invited_by_user_id            0.000890
creation_source              -0.003465
email_provider               -0.030662
```

Fig. 1. Linear correlation

The factors which predict future user adoption are those factors which are correlated with “adopted” or “adopted user.” Adopted user is found by "taking a user’s input and returns whether or not they have been active within any 7-day period.” It turns out that usage_length,

users['usage_length'] = users['last_session_creation_time'] - users[‘creation_time’],

‘last_session_creation_time,’ which is unix timestamp of last login, and ‘creation_time’, which is when they created their account, and org_id are positively correlated with adopted user, so will best predict future user adoption.

The accuracy, according to the confusion matrix of the Keras neural network, is 86%.

Reference:
olsenben/relax-data-science-challenge