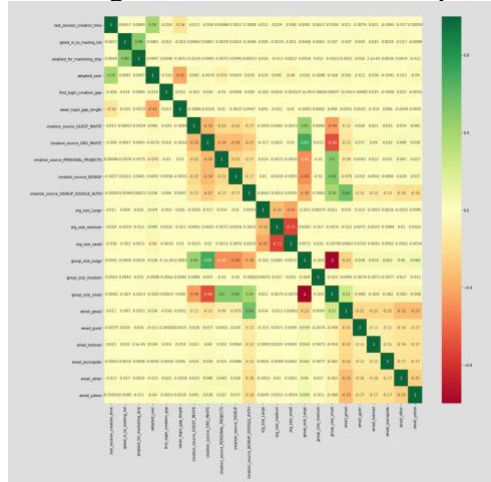


# Asana Data Challenge Report

## Analysis:

To do a whole picture analysis, I joined two tables together and cleaned the data. To see which feature play import role in predicting adopted user. I used the feature selection method in machine learning. The first is to use heatmap to plot correlation matrix. The second is to use univariate selection to the statistical test to find out the import features. The third is used the random forest tree-based model to discover the import features.

Used correlation matrix, from the heatmap 1a, we can see the “last\_session\_creation\_time”, “mean\_login\_gap\_length” have strong correlation with the adopted user value.



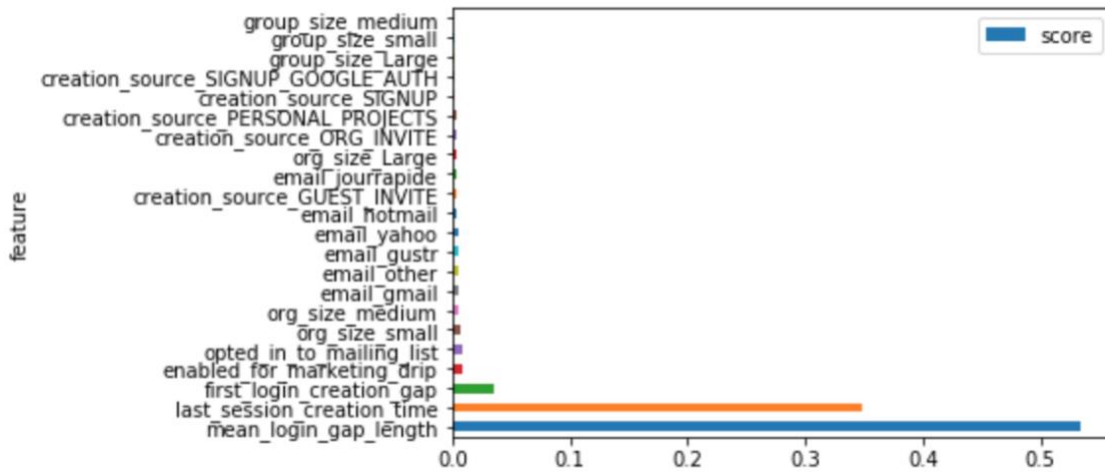
1a

The second is to use univariate selection to select the feature have the strongest relationship with output. From 1b, we can see that “mean\_login\_gap\_length”, “creation\_source\_PERSONAL\_PROJECTS ” and “last\_session\_creation\_time” these features play more import role on effecting the adopted-users

	Specs	Score
4	mean_login_gap_length	2716.483567
7	creation_source_PERSONAL_PROJECTS	56.844560
0	last_session_creation_time	36.239314
12	org_size_small	22.581597
5	creation_source_GUEST_INVITE	20.283149
21	email_yahoo	15.237361
11	org_size_medium	15.133797
9	creation_source_SIGNUP_GOOGLE_AUTH	13.848040
18	email_hotmail	12.762762
16	email_gmail	11.048830

1b

The third method I used is to extract the feature importance from Tree based classifier. I used Extra Tree Classifier to extra top import features. From 1c, the “mean\_login\_gap\_length”, “last\_session\_creation\_time”, “first\_login\_creation\_gap”are import features effecting the adopted users.



1c

	feature	score
0	mean_login_gap_length	0.533762
1	last_session_creation_time	0.348580
2	first_login_creation_gap	0.035305
3	enabled_for_marketing_drip	0.008656
4	opted_in_to_mailing_list	0.008292
5	org_size_small	0.006540
6	org_size_medium	0.005723
7	email_gmail	0.005516
8	email_other	0.005073
9	email_gustr	0.005039
10	email_yahoo	0.004949
11	email_hotmail	0.004476
12	creation_source_GUEST_INVITE	0.004133
13	email_jourrapide	0.004053
14	org_size_Large	0.003960
15	creation_source_ORG_INVITE	0.003762
16	creation_source_PERSONAL_PROJECTS	0.003514
17	creation_source_SIGNUP	0.002526
18	creation_source_SIGNUP_GOOGLE_AUTH	0.002272
19	group_size_Large	0.001955
20	group_size_small	0.001907
21	group_size_medium	0.000007

1d

In conclusion, “mean\_login\_gap\_length”, “creation\_source\_PERSONAL\_PROJECTS”, “last\_session\_creation\_time” and “first\_login\_creation\_gap” these four features played import role on the adopted-user. Other feature like email and “enabled\_for\_marketing-drip”, “opted\_in\_to\_mailing\_list” and org\_size also are less important, also have an influence on adopted\_user.