Springboard – Unit 17.1 Relax Inc, Take Home Challenge – 2019-08-22
Tom Preston

The Relax Inc analysis focuses on identifying factors which predict user adoption. An "adopted user" has logged into the product on 3 separate days within one seven-day period. The two main data sources are the user sign up data and user activity data from May 2012 – April 2014.

Analysis: The user access freq. (logins) were summarized by user id and grouped into bins. As shown in the table below, 10,174 users have either never logged into their account (3,177) or only once (6,997). Another 923 users (5-9 times & 10 – 49 times) are considered low volume users.

Access Freq	Number of Users
0	3177
1	6997
5-9	368
10-49	555
50-99	243
100-199	283
200-299	165
300-399	113
400-499	76
500-599	21
600-649	2

The remaining 903 users (50 – 99 thru 600-649) would be considered the "adopted" user. This is only 7.5% of the users. The user activity data only shows login dates therefore we will review the user signup data for any insights.

The summarized user activity information was merged with the user sign up data. The user sign up data has three main attributes for analysis (creation source, opted into mailing lists, on the regular marketing drip, etc). This data was summarized (via pandas cross tab) for each of the three attributes and the percentages were calculated to identify any data trends.

The table below shows the three attributes (signup group, mailing list, and marketing) by the access freq bins. The highest signup group value is highlighted in red. The org invite has the highest values for the active users and the personal projects is the highest for the low volume users. The mailing list and marketing information is pretty consistent across all groups. Both have a majority (70 – 80%) declining both options.

	Signup Group					Mailing List		Marketing	
Access Freq	GUEST INVITE	ORG INVITE	PERSONAL PROJECTS	SIGNUP	SIGNUP GOOGLE_AUTH	No_mail_list	Yes_mail_list	No_marketing	Yes_marketing
0	0.18	0.34	0.42	0.06	0	0.76	0.24	0.86	0.14
1	0.17	0.36	0.08	0.22	0.16	0.75	0.25	0.85	0.15
5-9	0.22	0.38	0.11	0.15	0.14	0.77	0.23	0.83	0.17
10-49	0.23	0.35	0.11	0.17	0.14	0.76	0.24	0.85	0.15
50-99	0.24	0.31	0.11	0.2	0.14	0.7	0.3	0.83	0.17
100-199	0.23	0.37	0.09	0.2	0.12	0.71	0.29	0.84	0.16
200-299	0.16	0.38	0.14	0.2	0.13	0.73	0.27	0.85	0.15
300-399	0.22	0.3	0.07	0.17	0.24	0.75	0.25	0.85	0.15
400-499	0.25	0.29	0.07	0.18	0.21	0.8	0.2	0.89	0.11
500-599	0.43	0.33	0.05	0.05	0.14	0.67	0.33	0.71	0.29
600-649	0	0.5	0	0	0.5	1	0	1	0

This limited set of data does not provide much insight. Additional analysis could be performed on the creation time and last login date for the user however given only 7.5% of the 12,000 users are active, this would probably just further reduce the number of active users.

Additional information around user activity (what they do when they log on, how long they are active on the system, etc) is needed to identify active users and to predict when active users may become inactive