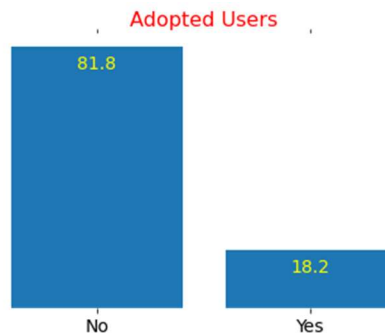


## **Business problem:** Identify which factors predict future user adoption.

**Who are the adopted users?** Adopted user is a user who has logged into the product on three separate days in at least one seven-day period.

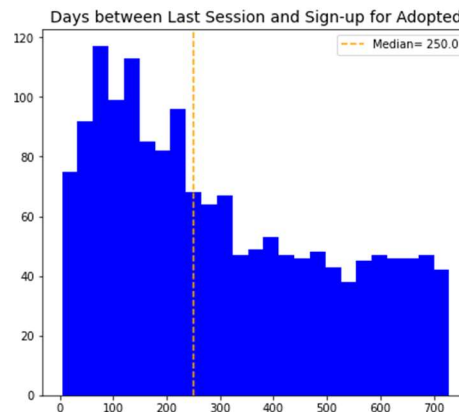
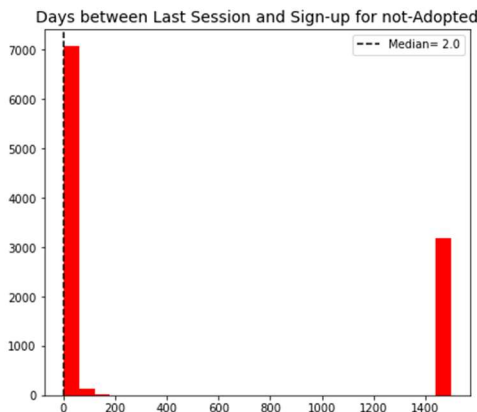
**Exploration Findings:** After merging and analyzing two tables about users who signed up to use a product from '2012-05-31' to '2014-06-06' I found:

1. From 12000 users who signed up, 3177 of them didn't have logged-in for any sessions. From the remaining 8823 users who created session at least one time, only 1602 has adopted.

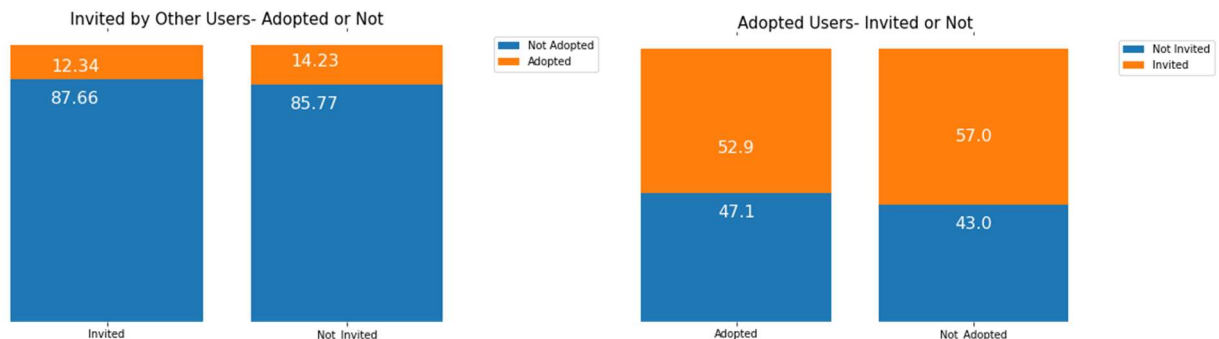


This is a highly imbalanced dataset.

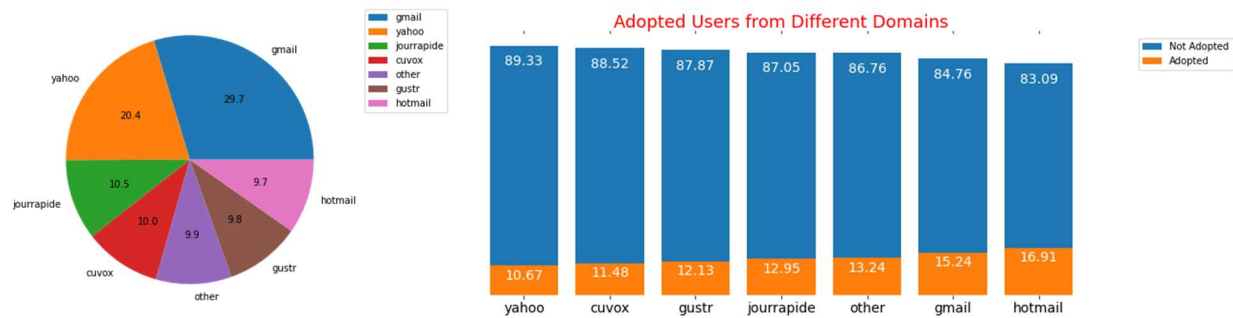
2. 50% of not-adopted users logged-in for the last time within first 2 days after their sign up. But, last logged-in of 50% of adopted users happened within 250 days after their sign-up date.



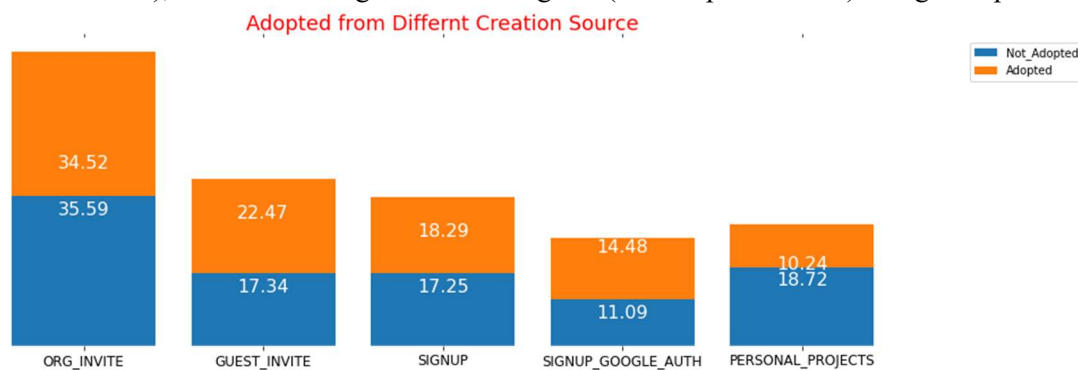
3. Self-motivated users were more adopted than users invited by other users. Since, only 12% of invited users were adopted, while almost 14% of not-invited users were adopted. Also, almost 57% of not adopted users were invited by other users.



4. About 51% of users had Gmail or Yahoo. About one third of adopted users used Hotmail and Gmail.



5. More than 75% of adopted users created their accounts when invited to an organization (as a full member), invited to an organization as a guest (limited permissions) or signed up via the website.



6. Only 8 organization id had one percent or more frequency (12000/100= 120, 120 or higher frequency). Those were [0, 1, 2, 3, 4, 6, 5, 9].
7. I used **Boruta** method for feature engineering using Random Forest Classifier. The Result was:

	Feature	Rank
2	timedelta_lastsession_signup	1
3	invited	2
0	opted_in_to_mailing_list	3
16	freq_org_id	3
12	creation_source_ORG_INVITE	5
1	enabled_for_marketing_drip	6
5	email_domain_gmail	7
11	creation_source_GUEST_INVITE	8
14	creation_source_SIGNUP	9
13	creation_source_PERSONAL_PROJECTS	10
15	creation_source_SIGNUP_GOOGLE_AUTH	11
10	email_domain_yahoo	12
6	email_domain_gustr	13
9	email_domain_other	13
4	email_domain_cuvax	15
7	email_domain_hotmail	16
8	email_domain_jourrapide	17

### What do selected features tell?

If number of days between the signed up and last logged-in is high (median of 250 days), the user most probably converted into adopted user. If user has been invited by other user, it would increase the probability that the user would be adopted. If organization invited a user to create an account, especially organization with id in [0, 1, 2, 3, 4, 6, 5, 9], the user most probably will be adopted. If user has been opted to receive marketing email, or signed up from Gmail domain the probability of being adopted would be increased.