## Business problem:

Identify which factors predict future user adoption

Who are the adopted users?

Adopted user: as a user who has logged into the product on three separate days in at least one seven day period.

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:

- 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.
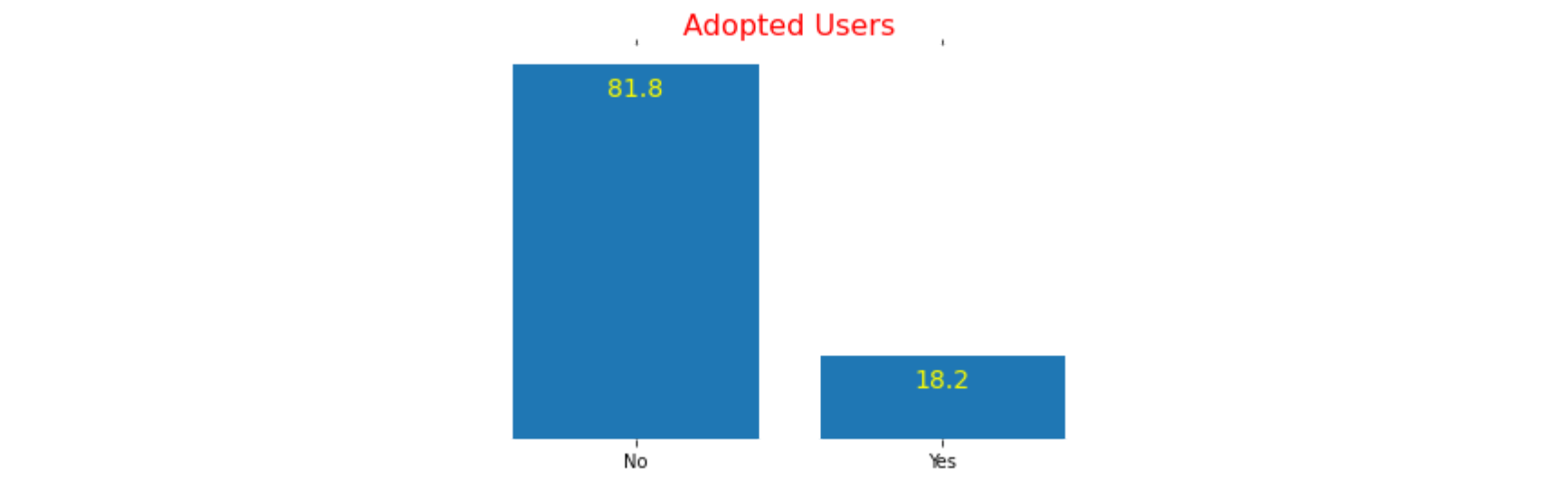

- Not adopted users were mostly logged-in in the first 3 months, but majority of adopted users created sessions in the first year.
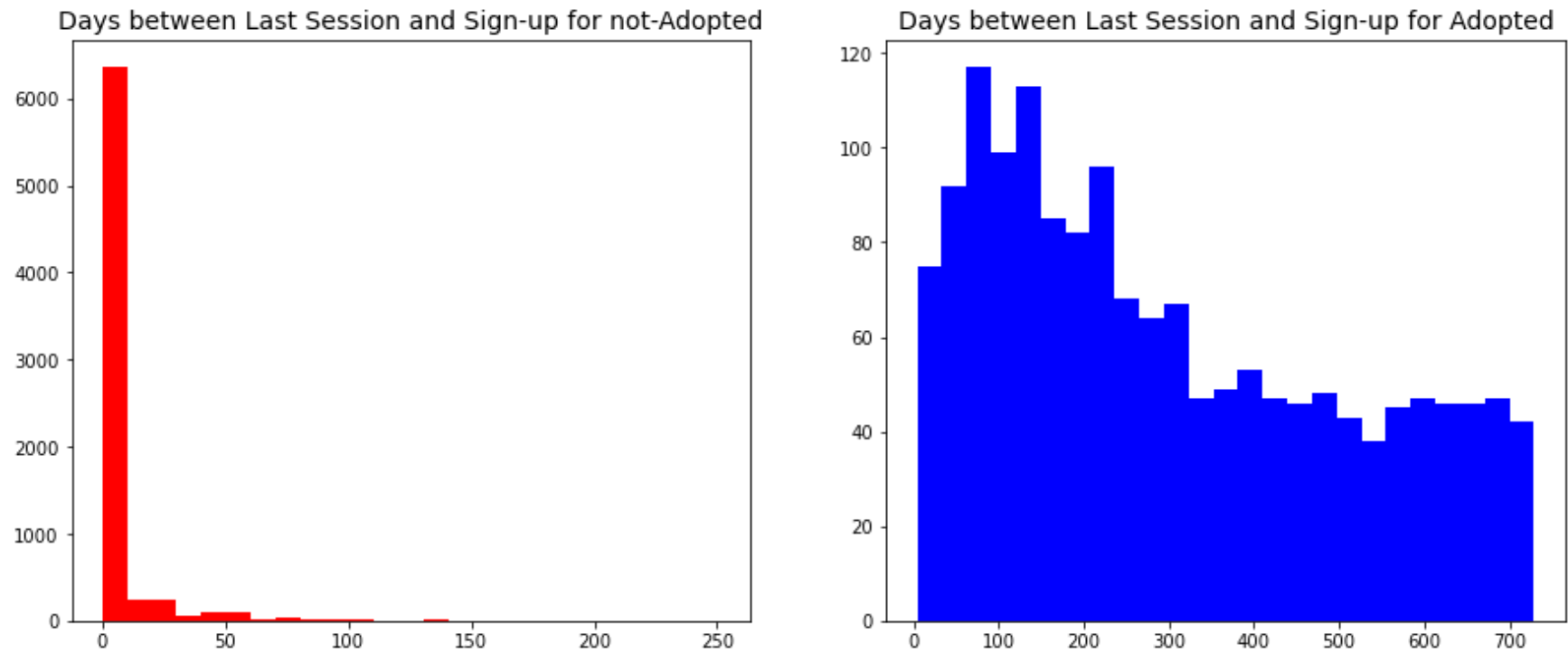
- Majority of users signed up by their gmail and yahoo email address. Percentage of adopted users signed up with gmail and hotmail domains were higher than percentage of not adopted users.
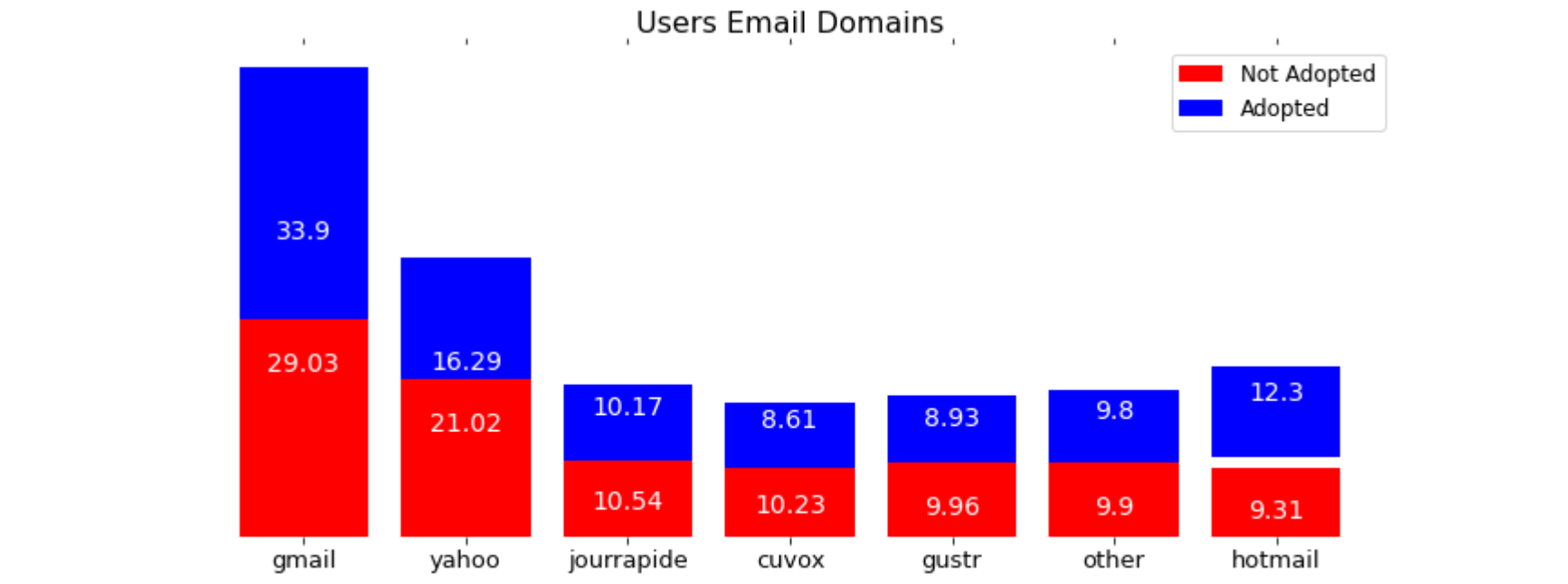
- About 43% of users invited by other users to sign up and use the product were adopted, while about 48% of users not invited by other users were adopted. **Users who waere not invited by other users, significantly adopted (57%).**
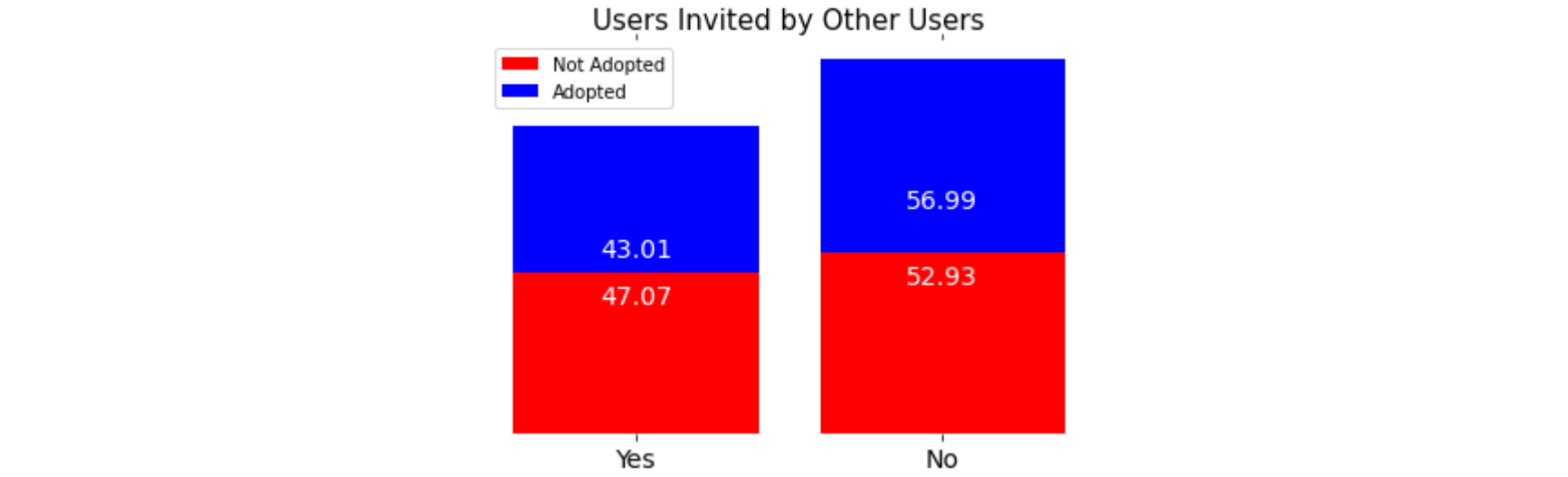
Boruta method with Random Forest Classifier as its estimator was used for feature selection. The result was:

```
In [22]: from IPython.display import Image
         Image('E:/Springboard/Unit 26_Effective Interviewing for Data Scientsts/26.2 Take home challenge/relax
         _challenge/Ranking.png')
Out[22]:
```

	Feature	Rank
3	timedelta_lastsession_signup	1
11	org_id_hash1	2
12	org_id_hash2	3
18	org_id_hash8	4
2	invited	5
13	org_id_hash3	6
20	org_id_hash10	7
19	org_id_hash9	8
14	org_id_hash4	9
0	opted_in_to_mailing_list	10
5	email_domain_gmail	11
1	enabled_for_marketing_drip	12
10	email_domain_yahoo	13
4	email_domain_cuvox	14
7	email_domain_hotmail	15
8	email_domain_jourrapide	16
6	email_domain_gustr	16
9	email_domain_other	18
15	org_id_hash5	20
16	org_id_hash6	20
17	org_id_hash7	20

About selected features:

'timedelta_lastsession_signup', time between the signed up and last logged-in is the most important feature. If user has been invited by other user, it would increase the probability that the user would be adopted. If user has been opted to receive marketing email, or signed up from gmail domain the probability of being adopted would be incresed.