

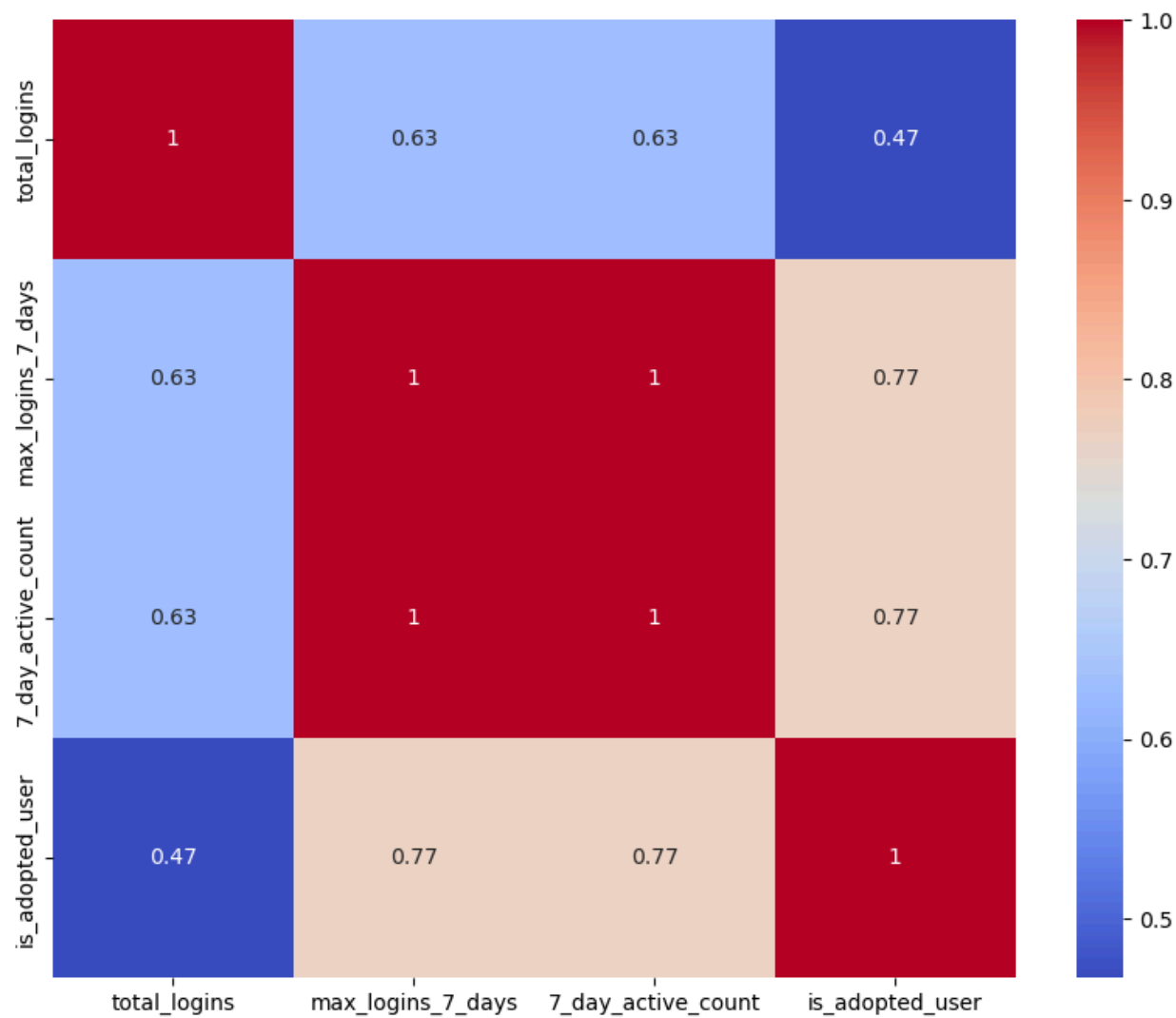
Relax Inc.

Data was imported and date and time stamps were converted to date time objects to aid in aggregation.

'Total\_logins' and '7\_day\_active\_count' counts were calculated by using groupby on a 7 day rolling period.

Adopted users were identified and a column was added with a boolean for each user.

The correlation plot of the numerical columns shows a strong correlation between 'is\_adopted' with 'Total\_logins' and '7\_day\_active\_count'



A random forest classifier was selected for classification.  
Creation source was encoded as the only categorical variable.

```
creation_source      int64
opted_in_to_mailing_list  int64
org_id               int64
invited_by_user_id    float64
total_logins          int64
7_day_active_count    float64
```

precision recall f1-score support

False	1.00	1.00	1.00	9523
True	1.00	1.00	1.00	32061

accuracy			1.00	41584
macro avg	1.00	1.00	1.00	41584
weighted avg	1.00	1.00	1.00	41584

It achieve a 100% accuracy and was cross fold validated , cv=5

Redundant or linearly correlated features were removed to prevent overfitting but accuracy remained at 100%

The plot of rf importance shows highest influence due to 7 day active count.

