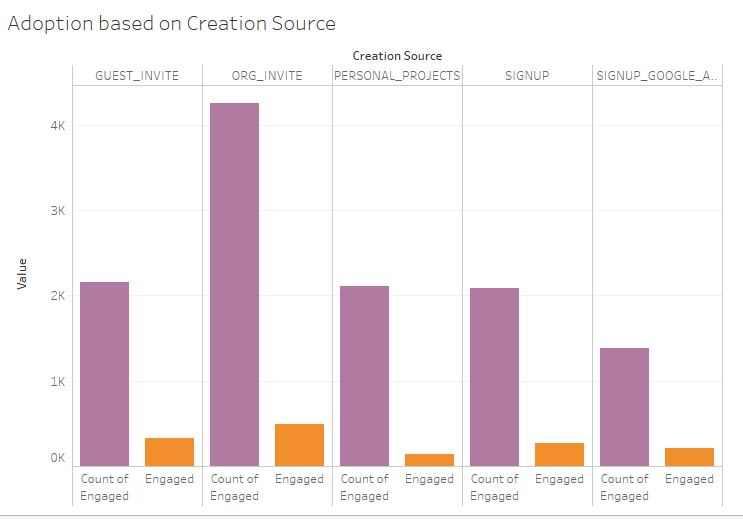
**How likely are users to be adopted on the productivity platform- A brief study**

**Objective:**Relax Inc makes a productivity and project management software that’s used by individuals as well as teams. The objective is to find out the likelihood of a customer being engaged in the product and the factors that influence the same.

**Methodology:**The adoption was calculated based on the login dates of the customers available. If there was any week where the customer was logged in for more than 3 times, the customer has been considered to be engaged.  
Below is an analysis done on the relationship of adoption with the remaining parameters.

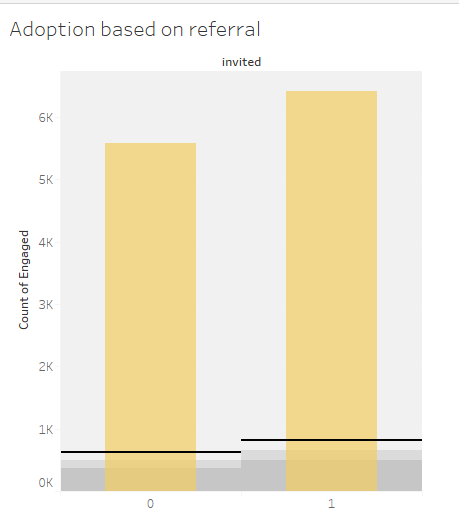


It can be seen that the adoption rate is consistent throughout the sources.   
No clear correlation between the Source and the adoption can be pointed out.



It was seen that out of the 10,208 customers who hadn’t opted for Matketing Drip, 12% i.e 1226 were adopted.

However 219 of the ones who had opted for it(12% of the 1792 who had opted for it) were adopted.   
Again there is no clear correlation is seen between the Marketing drip and the adoption of the customer.



There were 6417 customers who were referred to by other customers of whom 823 were adopted(11.1%).

However out of 5583 pupils who were acquired through other sources, 622 were adopted (11.1%).  
We again see no clear correlation between adoption and referral of customers.

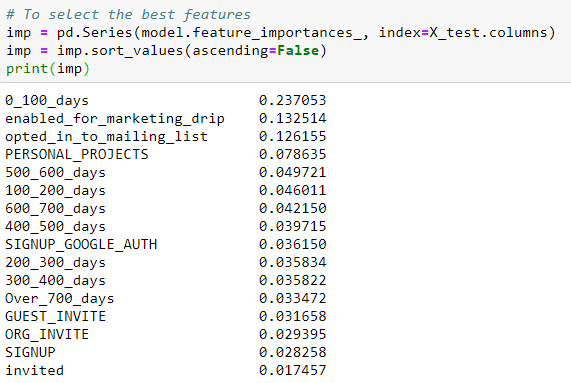
**Model performance:**A random forest classification model was built using the independent features Months on Books, Referral, Opted In to mailing list, Enabled for marketing Drip, Channel for signing up and Engaged as the dependent feature.

The model was trained on different number of estimators from 10 to 150 and features varying from 2 to 10.   
It was seen to perform the best with 10 parameters and 100 decision trees.  
The performance of the models was not seen to be very good as the best model gave an AUC of 0.615

The accuracy was seen to be 87.6% while the precision was seen to be very low at 0.45%. Recall was also quite low at 25%

|  |  |  |
| --- | --- | --- |
| **CONFUSION MATRIX** | Predicted Adopted | Predicted Not Adopted |
| Actual adopted | 3149 | 447 |
| Actual not adopted | 3 | 1 |

An imbalance is seen in the dataset. The number of Not adopted is very less in comparison the Adopted customers. This is resulting in a low precision and recall.

**Feature selection:**  
Customers who have been recently onboarded are seen to have a slightly higher possibility of adopting.  
The scores of feature importance are however seen to be insignificant.

**Scope for improvement:**The data is seen to be imbalanced. Hence if there is a possibility of collection of more data points, the model can trained better.

If more parameters can be collected, we can get a better picture of the factors impacting human behavior. Thus more research can be done on the data that was collected in order to build a robust model and reach better conclusions.