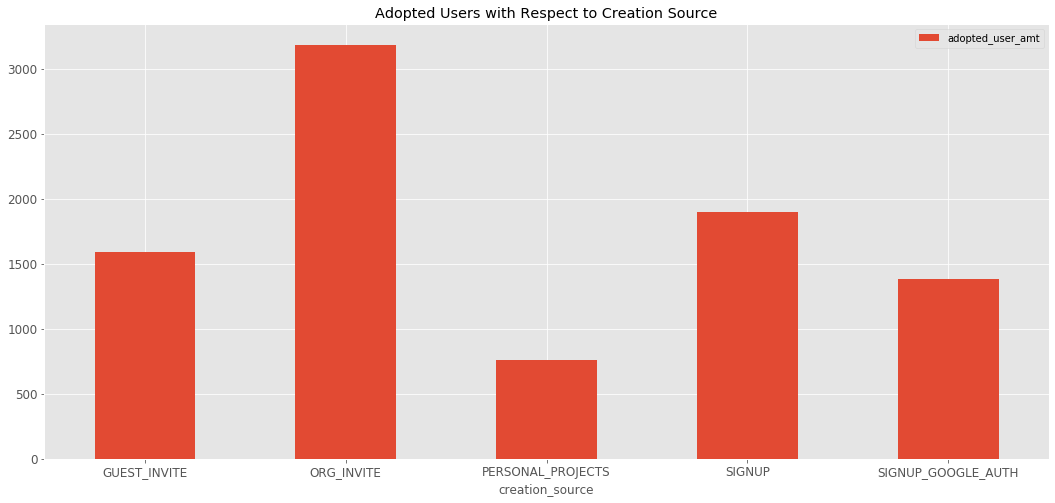
**17.1 Data Take Home Challenge: Relax Inc.**

**Approach**

The data was modeled using the random forest algorithm to identify which features were the most useful in predicting future user adoption.

**Discussion**

The first thing that was to be considered as the most important feature was the creation source feature, which identifies which channels their accounts were created. The figure below shows a bar plot of how the adopted user’s accounts were created. It is makes sense to see that users stuck around when their account was created through an invitation to an organization. The hypothesis that is formed from this is that there is a cost attached to this service that isn’t available in the data which limits permissions for the other creation sources.

After modeling the data, the variable importance method was used to obtain a numerical value for how useful a particular feature was and the table below is the result.

|  |  |
| --- | --- |
| **Feature** | **Variable Importance** |
| last\_session\_creation\_time | 0.636915 |
| org\_id | 0.278290 |
| opted\_in\_to\_mailing\_list | 0.013034 |
| enabled\_for\_marketing\_drip | 0.012151 |

**Conclusion**

It is a bit obvious that last\_session\_creation\_time would be a good predictor in the data, but it doesn’t create that much insight. What strikes as interesting is the org\_id column. Based upon the earlier hypothesis, it seems that when the organization that the user belongs to invites them to use the service, they are more likely to use it.