**Relax Challenge**

1. To start with the analysis, the basic assumption made is takehome\_users.‘object\_id’ is referenced in takehome\_user\_engagement dataset as user\_id.
2. Both takehome\_user\_engagement and takehome\_users dataset can be joined with right outer join using takehome\_users.‘object\_id’ = takehome\_user\_engagement.user\_id
3. Invited\_by\_user\_id field can be converted into ‘invited’ field with 1 or 0 categories.
4. The categorical columns creation\_source, opted\_in\_to\_mailing\_list, enabled\_for\_marketing\_drip and invited can be one-hot encoded to get numerical features.
5. A correlation matrix can be plotted on the features to get the understanding on the correlation to each other. This can be used to identify the influential features.
6. Also, the feature importance can be derived by building a random forest classifier model.
7. Field org\_id, email\_id, email, creation\_time and invited\_by\_user\_id can be dropped as they don’t look contribute anything for the prediction and rest of the features can be used as the prediction factor of user adoption.
8. The target field ‘Adopted’ with 1 or 0 can be derived based on the business logic for ‘adopted user’ which is at least 3 separate days logins per 7 days. Count(visited)=> 3, window day difference on timestamp =7 group by user\_id.
9. The algorithm that can be used in the problem is supervised classifier algorithm.
10. Also, some data cleaning activities can be performed on features and the cleaned dataset can be split into train and test data for model training.