**Problem Statement**
To predict the importance of factors/features responsible for a user to be getting converted into an adopted user.

**Approach**

1. Extract the number of adopted user ids.
   Extract the list of user ids logging more than or equal to 3 times a week and then filter the data_users.csv to create adopted users dataset.
2. Prepare the non-adopted dataset.
   It includes removal/replacement of null values present in dataset.
   a. 'Invited-by-user-ids' col by a separate category 'Non-invited.'
   b. All other rows with Null values present (predominantly in last-session-creation time) are removed.
3. Combine the adopted and non-adopted dataset.
   a. Randomly sample data rows of the same size as adopted samples data to create a final balanced dataset.
4. Feature Engineering and EDA
   a. Extract day, month, and year from the 'last-session-creation-time' and 'creation-time'
   b. Categorizing invited-by-user-id to two classes 'Invited and Non-Invited' will result in a much better feature.
   c. Removal of un-relevant columns i.e. columns having unique values for each row or having same values for each row in dataset
   d. Maximum number of account creation happens in the second week of the month i.e. between 8-11 of the months.
5. Preprocess the data and create a machine learning model to predict if the users is adopted or not
   a. Create a pipeline for one-hot-encoding the categorical values.
   b. Use variety of ml algorithms like Logistic Regression's Classifier, Decision Tree, Random Forest and SVM.
   c. Hyperparameter tune the model which is performing among the best i.e. having highest F1 score.
6. Gather the feature importance as returned by best fit model and also aggregate it by the relevant columns.

**Results**

1. We observe the highest f1 score resulted by Random Forest Classifier model with F1 Score of 62.95

2. We see the feature importance as in the following order