Take-home Assignment

Use Case

A client has data on users for an application from the past two years. They define an "adopted user" as a user who has logged into the application on three separate days in at least one seven-day period**. They want to understand what variables contribute to a user converting into an adopted user. The assignment is to inspect the data and prepare an analysis that shows non-technical stakeholders what variables and conditions are associated with user adoption.**

Data Overview

The data is available as two attached CSV files:

- *takehome_user_engagement.csv*
- *takehome_users.csv*

The data has the following two tables:

1. A user table (*"takehome_users"*) with data on 12,000 users who signed up for the product in the last two years. This table includes:
   - **name:** the user's name
   - **object_id:** the user's id
   - **email:** email address
   - **creation_source:** how their account was created. This takes on one of 5 values:
     - **PERSONAL_PROJECTS:** invited to join another user's personal workspace
     - **GUEST_INVITE:** invited to an organization as a guest (limited permissions)
     - **ORG_INVITE:** invited to an organization (as a full member)
     - **SIGNUP:** signed up via the website
     - **SIGNUP_GOOGLE_AUTH:** signed up using Google Authentication (using a Google email account for their login id)
   - **creation_time:** when they created their account
   - **last_session_creation_time:** unix timestamp of last login
   - **opted_in_to_mailing_list:** whether they have opted into receiving marketing emails
   - **enabled_for_marketing_drip:** whether they are on the regular marketing email drip
   - **org_id:** the organization (group of users) they belong to
   - **invited_by_user_id:** which user invited them to join (if applicable)
2. A usage summary table (*"takehome_user_engagement"*) that has a row for each day that a user logged into the product.

Instructions

Defining an "adopted user" as a user who has logged into the application on three separate days in at least one seven day period**, identify which factors predict future user adoption**. Arriving at an answer may look something like this:

1. Merge, clean, and organize data as necessary
2. Define a transformation to evaluate which users are adopted users along with other feature engineering
3. Conduct exploratory data analysis
4. If necessary, develop a machine-learning model

5. Produce a report with findings about the influence of different variables with respect to adopted users.

We suggest spending 1-2 hours on this, but you're welcome to spend more or less. Please send us a brief writeup of your findings (the more concise, the better - no more than one page), along with any summary tables, graphs, code, or queries that can help us understand your approach. Please note any factors you considered or investigations you did, even if they did not pan out. Feel free to identify any further research or data you think would be valuable.

Task submission guidelines:

The candidate is expected to submit two things:

1. **Google Colab Notebook link**

   The applicant is expected to write the code and share it as a Google Colab Notebook Link. Regarding this, please make sure of the following:

   - The colab link should be made public.
   - Please write comment lines which are explanatory in nature.
   - In case you have to make a choice, mention the reason as a comment line.

2. **Report**

   The applicant is expected to write a short report of maximum of one-page length. The report should be given as a PDF document.

   The report should at least have the following:

   - Summary of the exploration (4 sentences maximum), mentioning the findings.
   - Preprocessing steps taken (and possible reasons for the choices made)
   - Feature engineering steps taken (if at all; and reasons for the choices made)
   - Model selection (if needed; and reasons for the choice made)
   - Conclusion.
   - Figures, graphs and tables (if necessary)
   - Reference link (if any)