**Relax Inc. Take-Home Challenge**

**Objective:**

Identify the factors that predict future user adoption, considering that an “adopted user” is a user who has logged into the product on three separate days in at least one seven-day period.

**Methodology:**

1. From the provided usage summary table “takehome\_user\_engagement.csv,” I identified the adopted users. Each row represents each time that a user logged into the product. I used this table to count the number of times that a user logged into a table in a 7-day-period window.
2. The second table that was provided was the user table “takehome\_users.csv,” which contains the data on the 12,000 users that signed up for the product in the last two years.
3. Once the “adopted” users were identified, I added this column as my target variable

**Exploratory Data Analysis and Feature Engineering:**

I first considered the following relevant features to determine user adoption: how the account was created (e.g., invited as a guest, by signing up, using Google Authentication, invited to an organization, and invited to join another user’s personal workspace), the time the account was created, the time of the last login, and whether the user opted into receiving marketing emails. Additional features that I added are the number of days since the account was created, the season of the year in which the account was created, and whether the user was invited by a user that is considered adopted.

**Hypotheses and Results:**

I hypothesized that the there would be a difference between the way “adopted” users and “not adopted” users created their accounts and whether they were are more likely to be adopted users if they are invited by users that were considered adopted. However, the chi square tests resulted in p-values greater than 0.05, which suggest that there is no correlation between how the users created their accounts, as well as the type of user (adopted or not) that invited them to join. I also hypothesized that the season of the year in which the account was created would make a difference in whether a user would be adopted or not. However, I obtained a p-value less than 0.05 for the corresponding chi square test as well.

I also hypothesized that there would be a difference between the average number of days since the account was created for “adopted” and “not adopted” users, and after testing for normality in the data, I performed a Wilcoxon Rank Sum test to test this hypothesis. However, the p-value was greater than 0.05, suggesting that there is no statistical significant difference between the number of days since creating the account for “adopted” and “not adopted users”.

As a last attempt to identify important factors for determining user adoption, I used logistic regression and decision tree models. Unfortunately, the results are inconclusive since the models poorly predicted user adoption. For example, out of the 1249 "adopted" users in the training set, the decision tree only identified 7 users as “adopted” users.

**Future Directions:**

Unfortunately, I didn’t find any correlation between the features I explored and user adoption. This indicates that I need to do better feature engineering or more data needs to be gathered. It seems that the data provided is more related to how the account was created and when. What may be a better measure for user adoption is data related to user engagement. How are they using the product when they login? How many projects are they working on, how many organizations are they part of, what are the other users they engage with? This could give us more insight into how to retain and adopt the users.