# Relaxed Data Science Challenge

## Logic for defining ‘adopted user’

The following code creates a flag to determine whether a user was adopted or not – For every user, and for each time timestamp, if the different between the current time stamp and third time stamp from current time stamp is <= 7 days, then it’s an adopted user otherwise it’s not an adopted user.

user\_engagement['third\_time\_stamp'] = user\_engagement.groupby(['user\_id'])['time\_stamp'].shift(-2)

user\_engagement['first\_third\_day\_diff'] = (user\_engagement.third\_time\_stamp - user\_engagement.time\_stamp).astype('timedelta64[D]')

user\_engagement['ind\_adopted'] = np.where((user\_engagement['first\_third\_day\_diff'] <= 7),1, 0)

## Additional factors

1. Ind\_invited = Contains value 1 if the user was invited by another user otherwise 0

users['ind\_invited'] = np.where((users['invited\_by\_user\_id'].isnull()), 0, 1)

1. Tenure = Difference in days between the time of the last login and time when the account was created

users['tenure'] = (users['last\_session\_creation\_time'] - users['creation\_time']).astype('timedelta64[D]')

## Factors predicting future user adoption

Predicted model used: Logistic Regression was used a prediction model to determine relevant factors impacting user adoption as the dependent variable is binary



## Conclusion [HOW DO YOU DETERMINE IMPACT AND IMPACTFUL VARIABLES]

The factors which are most relevant in predicting user adoption are: opted\_in\_to\_mailing\_list, creation\_source and tenure

Note: There’s a Python Jupyter Notebook that contains the code for this analysis