The main method I used for exploring and modeling this data was to first create an altered version of the user engagement table to sort by user_id and sum the number of times visited while using the minimum of the timestamp column. I could then use this aggregated table to merge with the users table, matching on the creation time and timestamp columns. Doing this allowed me to see which user belongs to which user id, and thus the number of times they've visited as well. I then wrote a function to use on the original user engagement table to then add an adopted column to indicate if a user had indeed logged in 3 times within the span of a week.

With the goal of predicting whether a new user will be adopted or not, I decided to use a logistic regression model since we're working with a boolean outcome. I broke down the creation date into its temporal components, made dummies for creation source, and turned the invited by existing user column into booleans, since they are mostly an indicator of if someone was invited by another user.

If the column representing the number of times a user has visited is allowed as a predictor variable, it is easily the most defining column of whether or not a user will become adopted, as it is the only predictor variable with a coefficient outside of -1:1 at 5.303. This makes sense given that some users have visited hundreds of times, and some less than 3 times. Without this column, I have yet to tune the model in a way that it avoids guessing that everyone will be adopted. Signup is the best creation source for being indicative of adoption, and signup with Google authentication is the least. Finally, being added to the mailing list, being invited by another user, and being enabled for market drip all slightly indicate adoption in order of most influential to least.