

### Data Cleaning:

- I started out by cleaning up the data's missing values. I filled in the values with the mean of its corresponding column. Then I made a new column that was made to label whether a certain entry used to have a NaN value for a certain feature.
- Next Step was to begin one-hot encoding the categorical features.
- The next step was to look at the ***takehome\_user\_engagement*** table and begin exploring. I decided to group the ***user\_ids*** by their login periods. This helped me retrieve the data and then use it to figure out whether a user was an adoptive user or not.
- I found that out by checking each entry's next login datetimes and if they met the conditions of visiting the page three times in under 7 days, then they were marked with a 1.

### Exploratory Data Analysis:

After the data was cleaned up I started exploring which features were the most impactful to the adoptive user column. According to the heatmap, the most correlated Features like

***last\_session\_creation\_time*** and ***GUEST\_INVITE*** are one of the most correlated features.

It would mean that the time they last signed in would determine whether they'd come back and there are more guests on the product who are coming back to use the product.