None of the factors given in either csv file can effectively predict user adoption. To determine which users were adopted users, I used pandas rolling windows and pd.merge to left join on the users dataframe on object ID. Then, I split the merged dataframe into two dataframes that contained adopted and non-adopted users. Most visitors are not adopted users. I used value counts to determine any noticeable patterns. Taking ratios of invited users to total visits yielded very similar results. Breaking down the timestamps of creation time into separate hour and month columns and plotting them as histograms with 24 and 12 bins respectively revealed very similar graphs. However, it is notable that in the May there is a spike of non-adopters creating accounts, but the month doesn’t explain why there is a spike in non-adopters.

Traditionally, I see tracking cookies used to collect more data on website users. The main data of interest may be what other sites users click on shortly before and after visiting the site. How are users finding the site? Are they going to a competitor shortly after? Knowing more about the customers’ browsing habits in the short time around their login would be beneficial.