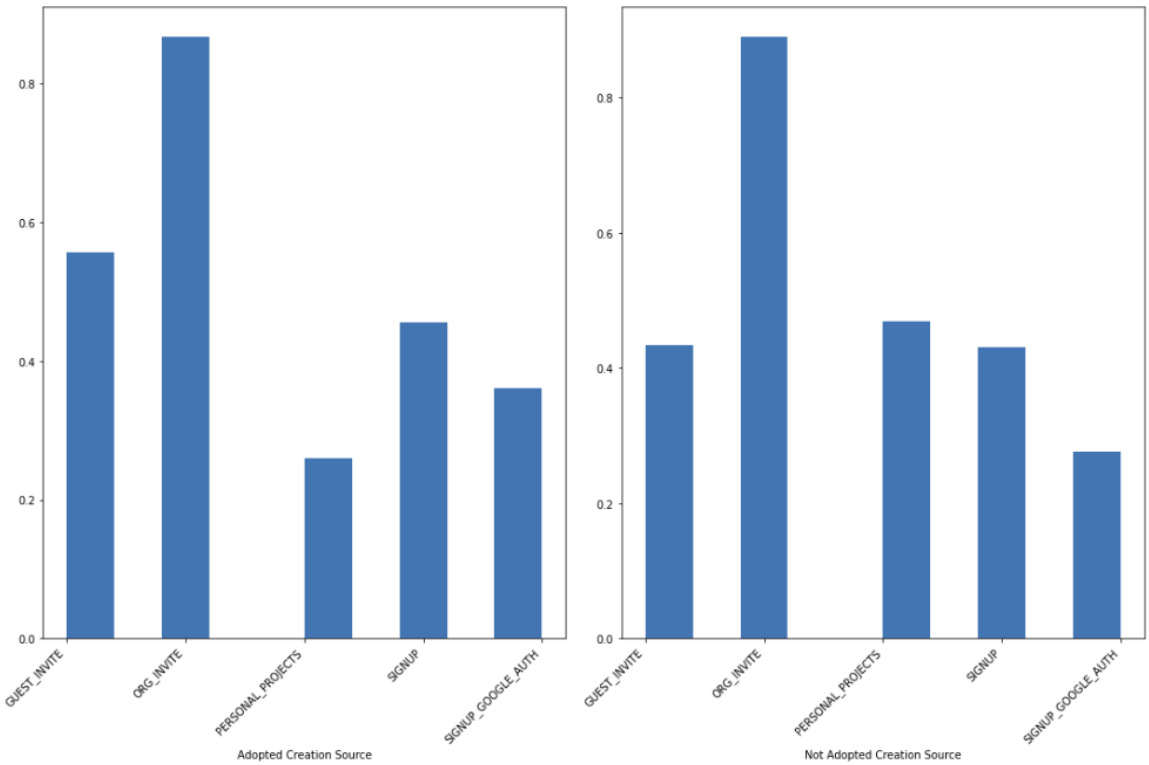
Rest Inc. Take-Home Challenge Report

Based on my analysis of the data given by Rest Inc., I am finding it challenging to discern any demographic information that would allow us to predict whether or not a user will become an “adopted” user. I looked at grouping the users first by whether or not they would be considered adopted. I then visually assessed the demographic data based on this information. For each histogram I explored, the distributions between the two groups were nearly identical. The only difference was their magnitude. The only difference I found was that between the not adopted and adopted groups, there was a higher ratio of accounts created through a “guest invite” that became adopted accounts and a higher proportion of accounts created for “personal projects” were not adopted (see Fig. 1) Perhaps those who create guest accounts are more likely to become adopted users.



I decided to attempt an XGBoost decision tree model after creating a column that designed a user as “adopted” or “not adopted”. We tried a few variations of the model but all of them struggled to classify the adopted group with anything higher than a 50% precision and almost always a 0% recall. A potential reason for this is due to the fact that the two groups are imbalanced with the adoption group taking up a much smaller proportion of the dataset than the unadopted group. For future approaches, I would probably do an upsampling method on the adopted group in order to increase the volume of data that we have for the adopted group. This way we can really begin to tease out some of the finer details for the model.

Perhaps looking at some data regarding the kinds of projects that people are working on when joining the website or their industry we can parse out some differences. For future research, I definitely recommend gathering a wider breadth of demographic data that can help better identify users. We could look at age and whether age of a user impacts their potential for adopting the website. There definitely is room for future improvement and analysis.