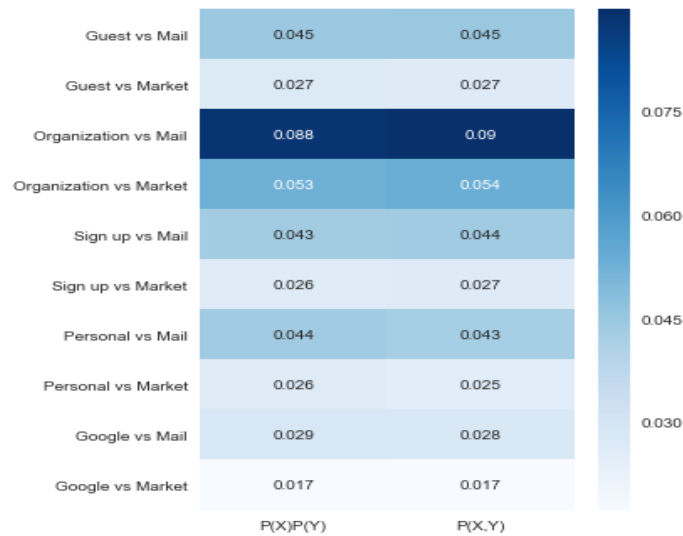# Relax Challenge Analysis

After processing the login data to get the True/False adopted label for each user, the irrelevant columns were stripped leaving only 'Creation Source', 'Mailing List', and 'Market Drip'. These remaining fields are categorical in nature. The below heatmap shows that there is almost no correlation between features.

This allows us to model the system using the Naïve Bayes generative approach where the $P(Y|X)$ is proportional to $P(Y|Xi)$ for each column multiplied together. Below is the graph showing the probability of adoption given that category. Total mean is the red line.
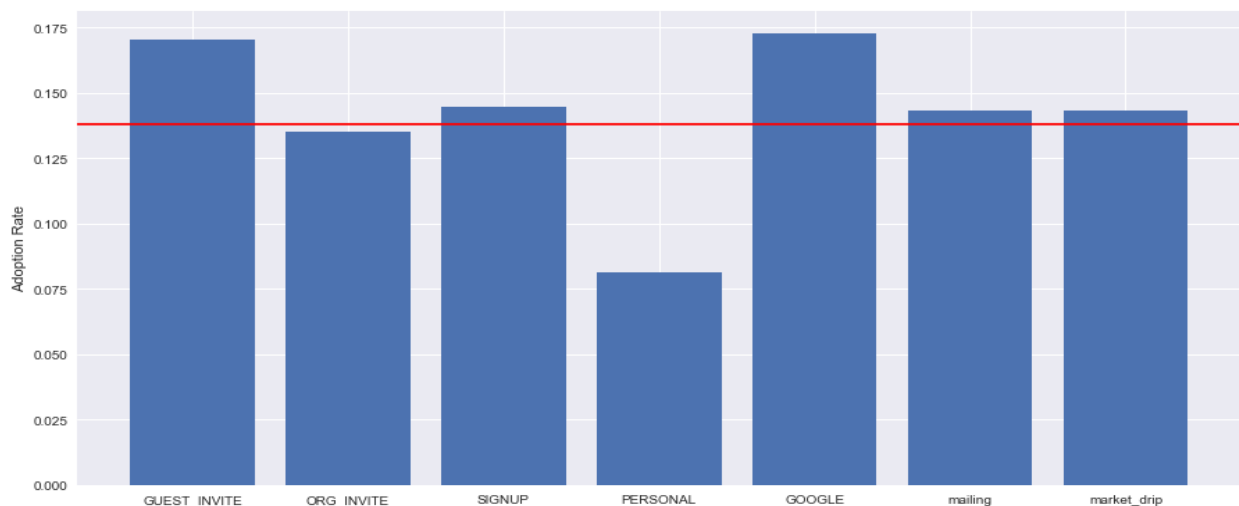
All deviations from the mean were statistically significant given an alpha of 5%. Based on this data Google authenticated and Guest invites had the highest adoption levels while Personal projects are the least likely. It may be useful to add meta data about organizations so that we can create categories based on the 'org_id' field in the data.