Write up of Findings for Relax Challenge

The user engagement file was used to determined user_id whom fulfilled the criteria for adopted user. Those users were mapped onto the users dataframe. For feature engineering, month and year of the creation time was mapped to two new features. Null values in invited_by_user_id were mapped to not_invited column as Boolean values. Same method was conducted for last_session_creation_time to never_on column. Its found that 13.6% of users are adopted users.

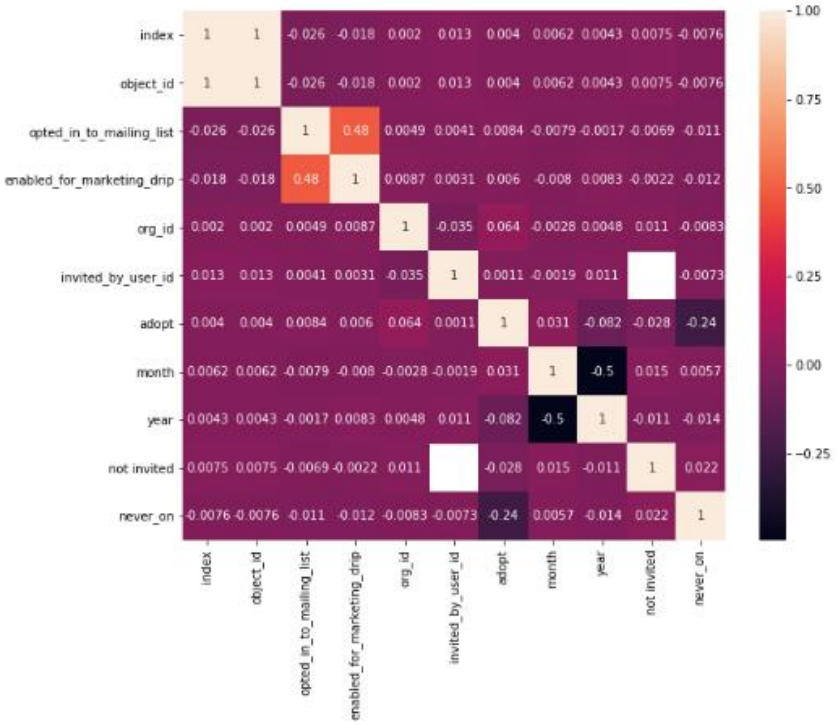

Figure 1: Heatmap based on Kendall Correlation

The heatmap based on the kendall correlation was performed (fig1). Kendall in general do perform well on data with ranks. The data generally shows weak correlation between feature variable and the target variable ('adopt'). However, 'never on' shows some correlation to 'adopt' at 0.24. Year and org_id shows very weak correlation at -0.082 and 0.064 respectively. It should be noted that org_id is randomized number for organization that user belongs to and may not play a role. A more thorough examination of the 400+ organizations and it's relationship with adopted users should be done to establish relationship.

The dataframe was then grouped by 'creation_source' and 'adopt' features and counted. Count values were normalized using the percent of adopted user and unadopted users to get a relative count. The bar graph of the relative count and it's relationship to creation source and adopted users are shown (fig2). Fundamentally for some creation source, it proves to show little correlation like ORG_INVITE but people that signs up based PERSONAL_PROJECTS, tends to not adopt the product

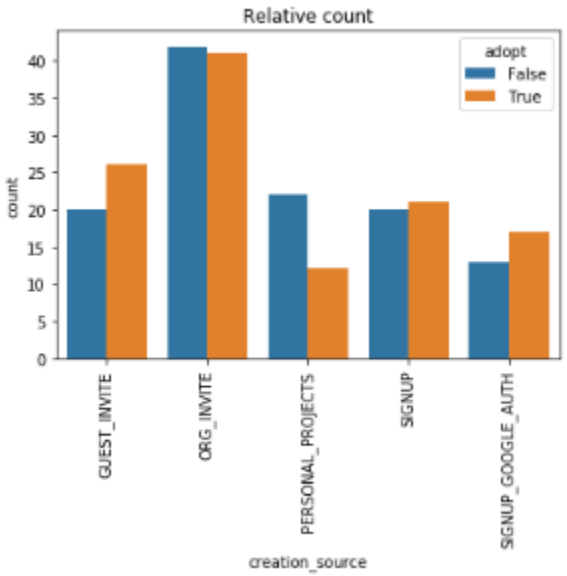

Figure 2:Bar Plot of Relative Count of Creation Source grouped by Adopted