Project Update 10/25

Andrew Whittum

Giacomo Ieronutti

Jaewoo Kang

Presented data preprocessing and preliminary analysis to client on 10/20

Calculated number of contents received by user, and the average and standard deviation for completion percentage per user

Also calculated “Activity Delay” column, which shows the amount of time between when the user received the content and when they opened. Applied a log transformation to deal with scaling issues

Then we broke the data down by day of week, and plotted completion vs time for each day of the week to see visually whether there were some times during the week when the user was more likely to look at the content

Next, we will begin looking at the columns to see which columns may affect time and completion percentage, and we will attempt to apply a time series analysis to the data.

We are a little bit unsure of how to actually go about building any models with regression/clustering/NLP, so we have emailed to the PM for some insights

We are also preparing deliverable 1 for Wed 10/28