Dataset review

Email Marketing Stats (Infographic)

This is a blog which has summary statistics showcasing the impact of email marketing by concerte numbers. However since it's an infographic I don't think we can extract enough data to work on the problem at hand.

Enron Email Dataset

The Enron email dataset contains text of the emails sent by employees of the Enron Corporation. It was obtained during investigation of systemic financial fraud in Enron. I think this dataset is also not suitable for us as it does not pertain to any kind of marketing but would rather be appropriate for some NLP problem.

Crossover Analysis between Archivist and Research Data Management

This data set contains analysis based on the the connections between archives professionals and research data management. These connections are established on the email correspondences. The dataset demonstrates how frequently archivists and records professionals discuss research data on the Archives-NRA list, the topics which are discussed, and an increase in these discussions over time. Anyhow as with earlier dataset this also doesn't present a problem which directly relates to our topic of research on email marketing.

Email Campaign Management for SME

This is the most interesting dataset as it has data from Small and medium-sized enterprises (SME) who use email marketing to target their customers. The dataset includes different aspects of emails to characterize it and also tracks if the mail is ignored, read or acknowledged by the recipient. This is of particular interest as we are striving to work on a similar solution.

Of the above, I think I will go with the SME dataset. There is potential to apply supervised machine learning approaches such as multi-class classification and/or regression analysis to establish relations between email attributes and the recipient's response. Moreover, I have generated my own dataset as discussed in last meeting and I am planning to either augment the SME dataset with it or somehow get value from it in discourse of this project.