

STA 360: Project Description

This is a comprehensive Bayesian data analysis project worth 20% of your final grade.

- Due to time constraints and the virtual learning environment, this project will be completed individually. You are more than welcome to work in groups if you wish, but keep in mind that we expect a greater depth of modeling and analysis for group projects.
- Please sign up for a project topic in the following [Google document](#), or feel free to add your own topic to the spreadsheet. The only requirements are that the topic must tackle an important real-world problem, and interpretable conclusions and meaningful decisions can be made from Bayesian analysis. Note that there is a wide spectrum in difficulty for the provided topics; some are quite straight-forward, while others are more difficult. We will be more lenient in grading projects on more difficult topics, provided that the analysis is done well.
- We strongly encourage everyone to actively seek feedback on Piazza and during office hours. *Use this time wisely!* Come prepared with ideas, questions, and analysis, and the TAs and I will give useful feedback and advice on how to improve your project.
- Some useful tips:
 - *Start early!* Meaningful and interpretable predictive modeling requires careful thought and analysis, and is difficult to do well when crammed in a couple of days. Starting early will give you ample opportunities to seek feedback and improve your project.
 - *Get feedback!* You are strongly encouraged to proactively seek feedback from me and the TAs during *all* stages of your project. Evaluations aside, we are here to *help* you develop this into a project you can be proud to present to potential employers. Of course, seeking and implementing feedback will improve the quality of your project, which may result in better grades.

Deliverables:

- *Final report* (due May 1st 11:59pm EST on Sakai): You will submit a final report (no more than 5 pages single-spaced, or 10 pages double-spaced) with the following sections:

- *Introduction*: A few paragraphs which (i) motivate problem importance & relevance (supported by any pertinent literature), (ii) describe project goals and how such goals address the problem, (iii) a high-level roadmap of the proposed Bayesian modeling framework, and (iv) other relevant information for the reader. See project rubric for details.
- *Data*: A couple of paragraphs describing the data to be used. You may wish to discuss: (i) data sources – where are you getting the data? (ii) data description – what data / variables will be used for modeling? (iii) data type – ordinal discrete, nominal discrete, continuous, etc., and (iv) data scraping / wrangling – how to extract and clean data for modeling? See project rubric for details.
- *Model*: Discussion & justification of the proposed Bayesian model framework (prior and sampling model). This discussion should elicit prior information on the problem, the data sources available, and relevant project goals. Any “downstream” uses of the model (e.g., for prediction, optimization, ranking) should be discussed in detail here. See project rubric for details.
- *Results*: Posterior analyses from the fitted Bayesian model, and a translation of such findings into meaningful & understandable conclusions for the target audience (e.g., engineers, business managers, policy-makers, etc). See project rubric for details.
- *Conclusion*: A summary of key findings and potential impacts of your project.

We will be following the project rubric quite closely for grading. If any parts of the rubric are unclear, you should clarify with the instructor or TAs – we’re here to help!