Project: Proposal

1 Introduction

As an initial step toward the semester project, you will develop a proposal for your work. This proposal should convene the core idea of your project, and frame it as a data science problem. *That is: it must be data-driven, and be structured to produce knowledge.* Your proposal will be reviewed by your classmates.

The use of peer review in this assignment serves multiple purposes that will improve the quality of your project. As a reviewer, you should try to provide feedback on the concept of the project. For example, you might: point out related ideas, suggest sources of data or potential methodology, provide commentary on the value of the work, etc. You should also help the author to determine the appropriateness and feasibility of the project. Although it is not typical for a scientific report, part of this assignment is for the reviewers to provide confidence to the author that they are embarking on a project that they can successfully complete. For example, you might: point out places where you think the project may run into trouble, encourage the author to focus their ideas, etc.

Note that is possible that your final project will differ from your proposal. Your proposal represents your current understanding of your problem domain and data science, and will likely be refined (as you learn new ideas from class), or adjusted (as you further investigate your data). The need for revision and adjustment is a fundamental part of conducting work on "the unknown".

2 Requirements

For this assignment, you will write a project proposal in markdown and submit it for peer review. The following nine items must be included:

1. Title: Create a concise title for your work.
2. Author: Indicate your name.
3. Date: Indicate the date of submission.
4. Keywords: Identify and list three keywords (may be phrases) which would help to categorize your work.
5. Description: Describe your project.
    - In general, this discusses the type of data you plan to analyze, and the types of questions you might ask about it.
6. Intellectual Merit: Describe the scientific discovery potential of your project.
    - That is, the potential to find some piece of knowledge (however small) which is new. Per the NSF, "[this] criterion encompasses the potential to advance knowledge"[1].
    - For undergraduates: you should be producing knowledge. For graduate students: you should be producing **NEW** knowledge. The "new" aspect will be evaluated in later milestones. **(UG VS GRAD DIFFERENCE)**

[1] https://www.nsf.gov/pubs/policydocs/pappg18_1/pappg_3.jsp

7. Data Sourcing: Identify and give a general methodology for retrieving data that is relevant to your project. The methodology proposed should be something you can conduct while following any restrictions/requirements on the use of the data.
    - For graduate students: the dataset must be relatively recent, that is, it MUST include data from the year 2023. **(UG VS GRAD DIFFERENCE)**
    - For graduate students: no Kaggle datasets. However! If you find something on Kaggle that then links to a (more trustworthy) original source, you are welcome to use that original source! **(UG VS GRAD DIFFERENCE)**
8. Background Knowledge: Identify and give three (3) sources that describe the background knowledge needed to accomplish your project. Any accurate source may be used (e.g., not just textbooks but blogs by reputable authors too).
    - Any citation format that can easily help the reader to retrieve the work is permissible. Including a DOI is recommended.
    - Note: you will not be required to cite/use these as a reference in your final written submission.
9. Related Work: Identify and give two (2) academic papers that is related to your proposed project.
    - Must come from a peer-reviewed venue.
    - Any citation format that can easily help the reader to retrieve the work is permissible. Including a DOI is recommended.
    - Note: you will not be required to cite/use these as a reference in your final written submission.

The entire file may contain a maximum of 500 words. There is no lower bound for size, but you must provide sufficiently detailed answers. See Canvas for a template file.

3 Evaluation Process

Please keep in mind that projects should be unique for all students in the class. If students submit the same core idea, then both will be required to make changes until they have distinct projects. Supposing that you came up with your idea in isolation (rather than googling: "data science projects"), your project is likely to be unique without any particular effort on your end.

After submitting your project proposal, we will conduct a single-pass peer review (see Appendix A for more information). Reviewers are expected not only to complete the rubric associated with the assignment (see Appendix B) but to also generate qualitative comments to the proposal author and justify the ratings given. For the "feedback on the concept of the project" mentioned in the introduction, Intellectual Merit is the most relevant criterion for leaving comments. Description or Data Sourcing may be used as well. As a rule, you must provide some justification for the assignment of level (Proficient, Competent, Novice).

- For this assignment, we will use the peer review functionality provided by Canvas.
- Peer reviews will be randomly assigned and will not be blinded.

For a later milestone: based on the evaluations by the reviewers, you will generate a revised version of your proposal and document that outlines your response to the peer feedback.

Be aware that we are doing real peer review. In previous classes, you may have asked to do peer review where you just assigned a number. That is not scientific peer review. In this class, you will be generating written feedback. Furthermore, this will be real feedback not just the rubric criterion restated.

4 Deliverables

For this assignment, you will receive two separate grades, which evaluate:

- Project Proposal Submission (Canvas, GitHub): see Section 2 for details. [10 points]
  - **Your proposal should be based on "project_proposal_template.md". The file should be renamed as "project_proposal_initial.md", and: 1) uploaded to your GitHub repository at root, 2) attached to the Canvas assignment.**
- Project Proposal Peer Reviews (Canvas): how you conduct the peer reviews for your assigned peer projects. Grading will be based on conducting accurate reviews and providing useful feedback/justification on the concept of the project. [10 points]

In rare cases, the proposal may be returned to the author for further revision. This will only occur with concerns of uniqueness or scope.

5 Next Step

Once the list of projects has been finalized for the class, there will be a short assignment where you:

- Identify other students in the class who appear to be working on a related project, and who may serve as peer reviewers for future project assignments.

Appendix A: Single Pass Peer Review

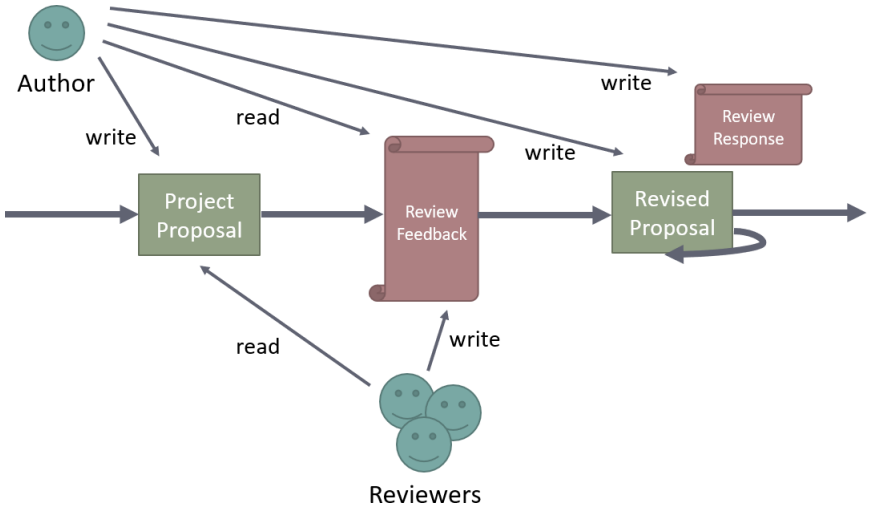

Figure 1: Graphical overview of peer review process.

In this assignment, we will conduct what is called a single-pass peer review. Let's break this down. First, let's borrow Wikipedia's definition for peer review: "[...] the evaluation of work by one or more people with similar competencies as the producers of the work (peers). It functions as a form of self-regulation by qualified members of a profession within the relevant field. Peer review methods are used to maintain quality standards, improve performance, and provide credibility."[2] Hopefully, that sounds useful! Peer review IS stressful but at the end of the day, the reviewer's job is really to help you. *Single pass* means that only one cycle of reviews will be conducted.

5.1 Process Outline

1. Initially, the **author** will submit the project proposal that they have written.
2. Next, and this is invisible, several students will be assigned as peer reviewers of each proposal.
3. The reviewers will each read and evaluate the submitted work against a rubric (see Appendix B).
    - A written justification for the rubric evaluation is required, to help the author to make corrections.
    - As you complete your review, try to be accurate. If you are unsure of something, state it.
4. The review feedback generated by the **reviewers** will be returned to the **author**.
5. Based on the peer reviews they received, the **author** will determine a series of changes that need to be made to the proposal.
    - Not all all reviewer concerns must be fixed! Reviewers may make mistakes.

In a future milestone, you will assemble a revised proposal that reflects changes based on review feedback and also changes due to learning more about your data after munging and visualization.

[2] https://en.wikipedia.org/wiki/Peer_review

Appendix B: Peer Review Rubric (tentative)

The final version of this rubric will be provided on Canvas where you fill it out when evaluating peer submissions.

Criteria	Proficient	Competent	Novice
Basic Information	A concise title, author, date, and at least three keywords are provided.	A title, author, date, and some keywords are provided, but title is not concise or less than three keywords are provided.	At least one piece of basic information is missing.
Description	Described your project sufficiently for its intellectual merit, appropriateness, and feasibility, to be evaluated.	Described your project for some but not all of its intellectual merit, appropriateness, and feasibility, to be evaluated.	Described your project insufficiently for its intellectual merit, appropriateness, and feasibility, to be evaluated.
Intellectual Merit	Described the scientific discovery potential of your project.	Described the scientific discovery potential of your project, however potentially unclear or reasoning incomplete.	Did not describe any potential for scientific discovery.
Data Sourcing	Identified a general methodology for retrieving data which is relevant to the project.	Identified a general methodology for retrieving data which was relevant to the project but is incomplete or unreasonable.	Identified a general methodology for retrieving data which was not relevant to the project.
Related Work	Identified an academic paper related to the proposed project.	Identified work did not meet academic requirements.	Identified work was not related to proposed project.
Uniqueness	To the reviewer's knowledge, this project uses different data sources and/or asks different questions than other projects being conducted.	DO NOT USE (This row will not be worth any points. If there is a uniqueness issue, revision will be required.)	This project appears to be duplicate with another project in the course.
Appropriateness	The project is data driven, is structured as to produce knowledge, and can level leverage techniques from data science.	The project is data driven, or is structured as to produce knowledge, but not both.	The project cannot level leverage techniques from data science.
Feasibility	The scope of the project aligns a single semester, solo student, project.	The project is too large in scope for a single semester, solo student, project.	The project is too small in scope for a single semester, solo student, project.