

Please write a 1-2 page proposal describing your thoughtful plan for working on your personal Data Science Project. Your proposal is meant to clearly articulate that you have carefully considered the various aspects of the Data Science Project Lifecycle (e.g. [OSEMN](#)) and have identified a clear question, related work, and relevant datasets to work with.

Please be as specific as possible in your descriptions to indicate your project plans now (with the understanding that these details may change as the project evolves) to demonstrate that you have a coherent and thoughtful plan for working through your project.

Your proposal should include the question you are trying to explore/answer, related work to consider, the data sets you plan to use to address it, your plan for using this data, description of the types of analyses you plan to perform (i.e. types of models, data transformations, etc.), the types of results you expect, and how you are planning to explain/share your results. Please be as specific as possible in your descriptions to indicate your project plans now (with the understanding that these details may change as the project evolves) to demonstrate that you have a coherent and thoughtful plan for working through your project.

Please be sure to include the following elements in your proposal under separate headings:

1. **Overview** -- a high brief level description of the project including the main question, type of analysis, and possible results/outcomes.
(Examples: <https://www.purdue.edu/data-science/research/winning-proposals.php>)
2. **Related Work** -- please include and describe at least two references to specific work that is related to your question and may help you to give context and additional meaning to your analysis. If you believe that there are no such references, please indicate your search process and how you came to that conclusion.
3. **Relevant Data** -- please include and describe at least two concrete datasets that you have identified that you plan to analyze to answer your main question, being sure to carefully explain how/why this data is appropriate and adequate for your analysis. This may include discussing the scope of the data, its timeliness, and interesting features/fields related to your question, as well as any exploratory data analysis you have already performed.
4. **Analysis/Modelling** -- please describe the types of transformations, analyses and models you are planning to perform on these datasets to allow you to gain insight into your main question. This may include both general types of analysis (i.e. regression/classification/clustering) and specific models (e.g. Regularized Regression, Random Forests, Word2Vec), to explain your plans.
(Examples of Data Science Methods/Models: <https://scikit-learn.org/stable/index.html>)

5. **Results** -- please describe the types of results you expect your analysis and modelling to produce, including the types of insights that might arise and also how you would plan to effectively communicate them (e.g. table, scatter plot, choropleth, heatmap, histogram, etc.). Here it's very ok to imagine a result your project might reasonably produce, and then explain how to effectively communicate it!

(Galleries of Data Visualization Examples:

<https://observablehq.com/@d3/gallery>

https://en.wikipedia.org/wiki/Data_visualization)

If you have any questions about the project proposal (or your project in general!), feel free to ask on our class slack channel **#research-projects-in-data-science**, during office hours, in class, or send an e-mail to jonhanke@princeton.edu.

Hope you have a great time thinking about your personal Data Science project and the steps to make it a reality!