

Starting a project/thesis

Table of contents

Official info	1
1 Preamble	2
2 Proposal format	2
2.1 Introduction (1–2 pages)	2
2.1.1 Data description	2
2.2 Background/Lit review(4+pages)	3
2.3 Methodology (2+ pages)	3
2.4 Timeline	4
2.4.1 Suggested deadlines (mandatory)	4
2.5 Bimonthly progress forms	4
2.5.1 Additional Thesis meetings	5
2.6 Supervisory Dissolution	5
3 Expectations For Writing	5
4 Proposal evaluation criteria	6

Official info

The official info about your proposal is available on moodle. The major items to think about during the proposal are:

- Committee (thesis) or second reader (project)
- Forms
- Proposal Document

1 Preamble

Your proposal should be written as if you are writing a project requirement for a good, but brand new data science student to execute for you. In the proposal, you will give an overview of the mathematics and algorithms that the student must learn through your document, with less relevant details available from original source documents. You should imagine that the student would need to learn the math as much as I require you to learn the math in courses, so the mathematical details will be there in the final document. Implementation details should focus on what the algorithm does and uses very limited analogies/metaphors unless used as an introduction mechanism. You should not discuss how the functions or libraries work, the parameters you pass, etc..

For project students, this is *your* proposal for your project, not a project that is assigned to you and designed by an instructor.

I will approve whether it is appropriate for a Master's degree and you will tell me the criteria for which I will hold you accountable during the terms you are completing your project. Only rarely should you get stuck or need help because you've given sufficient thought and research to your proposal. As you have designed the project, you have claimed to understand the terms and requirements (which has been approved).

2 Proposal format

After several issues, I now require all projects and theses to be written in LaTeX.

Your proposal should have the following document sections

2.1 Introduction (1–2 pages)

- If the main goal is to analyze a particular data set, motivate the context (a few sentences) of how that data is generated and what kind of problem will be solved or what insight you wish to gain from it..
- If the main goal is to analyze performance (e.g., for comparison) you need to motivate what the method(s) are designed to accomplish, and why there are different methods that exist in the first place.

2.1.1 Data description

- Identify/describe of all variables. You must identify the type of variable. You must also know what each variable is in the context of the problem; it can't just be just another feature you use to train an algorithm.

- Continuous variables should indicate units
- Ordinal and nominal variables should at least identify the number of levels, and all levels if there are only a few and an example of a few if there are more a - If relevant to the investigation, an approximate distribution of each variable, using an appropriate method of summarizing. This may be extended to correlations

2.2 Background/Lit review(4+pages)

The background/lit review will explain relevant methodological and algorithmic information to follow the next section (proposed methodology for the project/thesis). A bad habit of students is to rely on analogies of what's being calculated. You must understand and justify all of this. This section is usually the longest section of the proposal for this reason.

Methods should explain what any objective functions are, how optimization occurs, etc.. If a calculation (e.g., the objective function) is not the first, most intuitive calculation the brand new data science student would think of, you must justify why you use something less intuitive/more complicated to solve this problem.

If you elect to use a method, you are taking on the responsibility of explaining these choices at the Master's level. If you can't answer that, the topic/method is unsuitable for you and you must choose something you can understand and explain. Exceptions to this are rare. Many students underestimate how much they must understand the *math* behind algorithms to do a project, let alone a thesis.

What you discuss in your lit review should be required for the methodology. One notable exception is if you are doing a thesis and plan to provide an alternative approach to something, it is still worthwhile to explain what your approach is alternative to, here. This is especially important if you're performing some kind of comparison.

2.3 Methodology (2+ pages)

Assuming the theoretical new data science student understands the literature review, they should be able to implement the methodology set out in the proposal with ideally no other material necessary. A good lit review will mean the methodology will be very concise, direct, and almost like a checklist of instructions.

The methods should include the specific implementation of the algorithm. E.g., a lit review will explain multiple regression with dummy variables, while the methodology will literally write out the equation of the regression variables listing each coefficient for the particular data set. Algorithmic details for your particular data set go here as well such as if you need special initial values for your algorithm, parameteres you'll set in functions, or tuning parameters that you know ahead of time.

Simulation studies will identify exactly which parameters of the simulation will be investigated, which values will be fixed, and any corresponding assumptions.

A thesis which develops *new* methods will have a much longer methodology section because nothing exists to solve the problem at hand, so you propopose to develop it in in this section, and the development is part of the instructions to the student. But because this is a proposal, you don't actually perform the derivations here, you write a skeleton of how you plan to perform the derivation using elements of the lit review.

2.4 Timeline

Timelines should generally be performed backward. For thesis students, check the official documentation since there are some institutionally-dictated ones. For project students, this is slightly more flexible but the official documentation has the necessary deadlines. Include all of the pertinent deadlines in your proposal.

Dates should be set for the following (put them in chronological order but when you are brainstorming you'll do this backward):

2.4.1 Suggested deadlines (mandatory)

Plan for at least the following deadlines. You will have *many* more deadlines, however.

- Nov 15: implementation complete
- Jan 15: first complete draft*
- Feb 15: second draft
- Mar 15: last day to submit final draft to circulate to committee/second reader
- Apr 15: presentation/defence

2.5 Bimonthly progress forms

Both project and thesis students will conduct supervisor-only progress reports every 2 months on approximately the following dates

- Oct 01
- Dec 01
- Feb 01
- Mar 01

At these meetings, the supervisor evaluates if the student is making appropriate progress according to the timeline set out in the proposal. While infrequent setbacks can occur, the student is still generally expected to follow the timeline.

The results of progress reports are usually one of 3 or 4 categories:

- satisfactory progress
- concerning progress
- unsatisfactory progress

2.5.1 Additional Thesis meetings

The thesis students will have supervisory committee meetings at least once a term. The committee will decide whether it should be the first, second, or third month of each term at the inaugural committee meeting.

2.6 Supervisory Dissolution

The student will state in the proposal that they agree the thesis or project will be dissolved if any of the following happen:

- Two consecutive progress reports are unacceptable
- Three consecutive progress reports are concerning/unacceptable
- An academic integrity violation is suspected by the supervisor and suspected by at least one other faculty member.

3 Expectations For Writing

For documents like the proposal and the final document, you'll be given feedback. The feedback is usually in terms of content, or presentation/writing. Ideally you only need feedback for the former, and small amounts for the latter.

In practice, I find that some students underestimate the standard of writing expected for the document that ends their Master's degree and so the timeline I propose reflects that.

I will typically limit how much time I spend reading/annotating a document you submit (e.g., an hour for a proposal, three hours for final documents). If there are so many issues that I do not make it through your document due to poor writing, you've lost your opportunity for me to give you feedback on the later contents of the document, where I may catch additional major issues which themselves require additional revisions. If you require more revisions, they will simply occur *after* your original timeline, which means your degree completion will be

set back by a semester (or more), and you will have responsible for paying the corresponding extension fee.

4 Proposal evaluation criteria

While not a binding document, you're encouraged to look at the following (non-exhaustive) list of things that I look for when evaluating a proposal or final project or thesis.