

Project Stage 1 FAQs

What is the point of the project?

The project is all about conceiving, formalising and executing an end-to-end data analytics project. In stage 1, you define the problem, explore/prepare the data and propose approach/evaluation. In stage 2, you will implement and validate the approach.

Do I need to solve a real business/research problem?

No. Any problem is fine, as long as you have access to relevant data and your solution includes some aspect(s) of data analytics: substantial exploratory analysis, development/comparison of statistical hypothesis testing, data mining or machine learning.

Can I work on a real business/research problem?

Yes. You are encouraged to use data from work or another project. Please be careful that you don't expose confidential or sensitive information.

How big should the project be?

The problem should be just big enough to carry out the project. It's completely OK if the proposed solution turns out to not actually be beneficial. Critical evaluation and characterisation errors/shortcomings is more important than positive results.

What code should we submit for Project Stage 1?

This will depend on the project. We'd like to see anything you used to explore, clean, transform, sample, clean, load, etc. If you use a spreadsheet or proprietary software, please include a description or screenshots to give us an idea how you did your exploration and preparation.

Do I have to use the tools we cover in INFO3406?

No problem if you prefer Weka, Spark ML, R, SAS, Tableau, etc. Feel free to use whatever makes you most productive for your project work. The goal is to formulate and execute a data analytics project, not to demonstrate mastery of any particular tools.