

Assessment: Fundamentals of Data Analysis

ian.mcloughlin@atu.ie

Winter 22/23

These are the instructions for the assessment of Fundamentals of Data Analysis in Winter 22/23. These cover the full 100% of the assessment for this module.

Submission

- The deadline for submission is January 7th, 2023.
- Your whole submission must be in a single GitHub repository.
- Use the form on the Moodle page to submit your repository.
- All you need to do is submit the repository URL.
- You should submit the URL as soon as possible.
- Commits in GitHub on or before the deadline will be considered.¹

¹ Once you have submitted your URL, you do not need to do anything other than commit to your repository and push the changes to GitHub.

What to submit

This assessment has three overlapping components, as follows.

GITHUB REPOSITORY (20%):

- Your repository should be presentable in job interviews.
- Include an informative and concise README.
- Organize your repository — no unnecessary files or clutter.
- Work regularly, adding regular reasonably-sized commits.

SHANNON ENTROPY NOTEBOOK (40%):

- Create a notebook about Shannon entropy.
- Define and explain the main concepts.
- Pitch the notebook at your classmates.
- Use visuals like plots to explain concepts.

OUTLIER IDENTIFICATION NOTEBOOK (40%):

- Create a notebook about outlier identification in data sets.
- Describe the different kinds of outliers that might be present in a data set.
- Source an example data set online and identify and deal with any outliers that might be present.
- Explain the role bias might play in outlier identification.
- Well-written explanations².

² Pitch all your work at your classmates as the audience.

Marking Scheme

Each component will be marked using the four categories below. To receive a good mark in a category, your submission needs to provide evidence of meeting each of the criteria listed under it³.

Research (25%): evidence of research on topics; appropriate referencing; building on work of others; comparison to similar work.

Development (25%): clear, concise, and correct code; appropriate tests; demonstrable knowledge of different approaches and algorithms; clean architecture.

Documentation (25%): clear explanations of concepts in notebooks; concise comments in code and elsewhere; appropriate, standard README for a GitHub repository.

Consistency (25%): tens of commits, each representing a reasonable amount of work; literature, documentation, and code evidencing work on the assessment; evidence of reviewing and refactoring.

Advice

- Students sometimes struggle with the freedom given in an open-style assessment.
- You must decide where and how to start, what is relevant content for your submission, how much is enough, and how to make the submission your own.
- This is by design — we assume you have a reasonable knowledge of programming and an ability to source your own information.
- Companies tell us they want graduates who can (within reason) take initiative, work independently, source information, and make design decisions without needing to ask for help.
- The point of this assessment is to demonstrate you can do that.
- You need a plan, you cannot just start coding straight away.

Policies

- You are bound by all ATU policies and any GMIT policies that have not yet been replaced by new ATU policies.
- Review the GMIT Quality Assurance Framework⁴.
- Pay particular attention to the Policy on Plagiarism and the Code of Student Conduct.
- If you have any doubts about what is permissible, email me to ask⁵.

³ In line with ATU policy, the examiners' overall impression of the submission may affect individual marks in each category.

⁴ GMIT. Quality Assurance Framework.
<https://www.gmit.ie/general/quality-assurance-framework>

⁵ ian.mcloughlin@atu.ie