INF6027 Coursework 2024-25
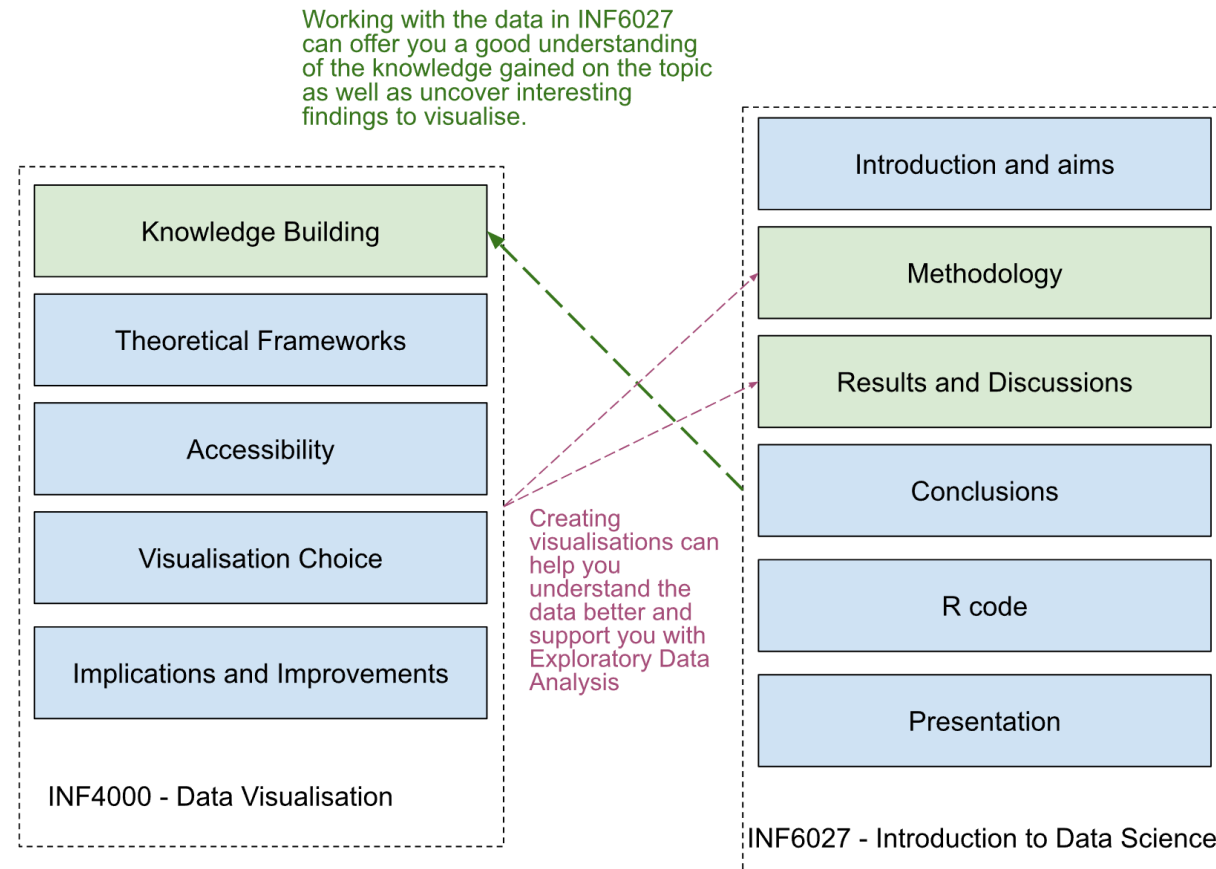
Dr. Suvodeep Mazumdar

Why this coursework?

- Being a data scientist isn't just taking data, processing it and creating a classification/regression/predictive/clustering model.

- Data science is not data analysis/software development

- It is also about thinking critically, engaging with previous research, setting up your own research context, and communicating your findings appropriately to different audiences.

- This coursework is aimed to assess all these skills, not just writing code for modelling/analysing data

Link with other modules

- Closely linked with INF4000

- Why?
  - Data viz is a key component of data science
  - Exploratory data analysis and communication – Data viz
  - Both modules use R – hopefully less work for you

- But assessment is different
  - INF4000: your ability to effectively present information using visualisations
  - INF6027: how you identify a problem, design a data science project, …

Working with the data in INF6027 can offer you a good understanding of the knowledge gained on the topic as well as uncover interesting findings to visualise.

INF4000 - Data Visualisation

- Knowledge Building
- Theoretical Frameworks
- Accessibility
- Visualisation Choice
- Implications and Improvements

Creating visualisations can help you understand the data better and support you with Exploratory Data Analysis

INF6027 - Introduction to Data Science

- Introduction and aims
- Methodology
- Results and Discussions
- Conclusions
- R code
- Presentation

The domains

- The coursework will require you to engage with one primary dataset, which can be combined with other datasets from the list of the provided datasets

- Datasets are from one of the following 5 domains: Football, Open Governmental data, News (text), Stock market (temporal), Music and films – these are very standard datasets, but may come with issues typically representative of real-world data

- You can combine datasets from other domain, but I doubt they will lead to much interesting insights

The task

- All the R practical sessions we have done should be sufficient for you to do this coursework well. However, it is likely you will need to go beyond just the practical worksheets and build on the knowledge you gain in the practical sessions.

- Minimum requirement :
  - Perform at least one type of data analysis (e.g. clustering, prediction, time-series etc.)
  - Include at least two visualisations (e.g. charts, maps etc.)
  - Provide a reasonable discussion of the motivations, research context, methodology and experimental design, results and discussions

What you should provide

- One main document (report), with links to a GitHub profile
- The Report is meant for an academic audience – where you detail your research, your methodology, your findings and communicate the results. There is a clear template which should help explain what exactly is needed
  - The report must have a link to the GitHub Profile, with a description of how you arranged your GitHub pages
- The GitHub link is meant for a non-academic audience. It could be a potential client or a future employer. It should have
  - Your profile, with your professional skills and interests
  - A page on the INF6027 coursework (you don't need to call it INF6027 explicitly, but it should be obvious what the project is)
    - Brief introduction (few lines)
    - Research Questions
    - Key Findings
    - (link to the) R code
    - Instructions for downloading and running the R code
  - A page on INF4000 coursework (this will be detailed in the INF4000 coursework document)
  - There will be support for the GitHub part of the coursework later on in the semester

How to do well

- Read the coursework carefully – read it multiple times
- Use the coursework template provided – it should hopefully help you avoid missing important information (New for 2024/25)
- Start working on the assessment from now – don't leave it to the last minute. Remember deadlines can easily pile up later on
- Start going through the datasets – try doing some initial analysis to study the features and characteristics
- Start reading the literature – this is a starting point but you will then need to engage with further literature
- Set the research context – what is the gap in knowledge, why choosing this topic, what are the research questions
- Justify your methodology and make sure you link with your research questions throughout

Become an Academic Representative

Become the voice of the students on your course, enhance your CV and develop your skills.

As a Student Academic Rep, you'll be part of the School's Student Staff Committee.

What will it involve?

Attending 4 x 1 hour Student Staff Committee pre-meetings and 4 x 1 hour Student Staff Committee Meetings. These will be scheduled around your timetable across the year so that they don't impact on your studies. Each hour long meeting is a forum for students, academic staff and support staff to discuss important matters on teaching and student support.

What will you do?

You will bring your fellow students' issues to the meetings, and our discussions will feed directly into the School's policies and development, so you'll be making a real difference to maximising the student experience. You will also have the opportunity to chair meetings

8

University of Sheffield

Are you interested in becoming a Faculty of Social Sciences Student Ambassador?

Represent your subject area at our various recruitment events and share your experiences as a current student with prospective students through this flexible role… and get paid £12.30 an hour!

foss-ambassadors@sheffield.ac.uk

RUSSELL GROUP

Frequently Asked Questions

- Please first check if your question has been asked and responded before.

- If not, please send an email to the module coordinator with the subject 'INF6027 Question'