1495 UNIVERSITY OF ABERDEEN | CELEBRATING 525 YEARS 1495 – 2020

ABERDEEN 2040

Assignment, and Exam Information

JC3503: Data Mining & Visualisation

2025

Today…

- Request for Feedback
- Assignment Information
- Exam Information

Mid-Way Briefing

As of right now, we are beyond halfway through the course!

As such, it is a great time to have a quick discussion on all things course-related.

Note that this is an open discussion: Please feel free to raise any questions or thoughts as we go through!

Request for Feedback

Call For Feedback

There are two critical points for giving us feedback:

- **Giving feedback now (throughout the course)**
- **Giving feedback at the end of the course**

Both are incredibly important and useful for all of us!

Feedback Now…

Giving feedback now is a great way to help steer the course while we are delivering it to you!

What is working well? What is *not* working as well? How would you like things to change for the remaining two weeks?

We can discuss here, or you can raise specific points/thoughts with us in labs, office hours, via email, etc.

Feedback at the End of the Course…

Giving feedback at the end of the course is also incredibly valuable to us!

While at that point it is too late to adapt the course for yourselves, your input will ensure that future years have the best learning experience.

You will be invited to give feedback – please do so!

Assignment Information

Assignment Information

As part of this course, you will be required to demonstrate your data mining abilities through a programming assignment.

This is worth 25% of your overall course mark, and the details of this assessment are now available on MyAberdeen.

The deadline for this assignment is:

Sunday 25th May, 2025 at 23:59 (UTC+8; Beijing time)

Assignment Information – Datasets

In this assignment, the aim is to demonstrate your ability to conduct EDA, data visualization, and data mining techniques.

The assignment involves analysing a dataset: ***movies.csv***

This dataset has been added to MyAberdeen for you to download and start working on.

Assignment Information – Datasets

The movies dataset contains information about ~4,800 movies from IMDB.

It includes metadata and movie keywords, information about their budget, revenue, and ratings, and an overview (synopsis) about the film, as well as other interesting variables.

It contains quantitative, categorical, and text fields – offering numerous possibilities for data mining and analysis.

Assignment Information – Objective

In undertaking the assignment, your task is to analyse the data, however you see fit, using techniques from this course.

You will be expected to undertake EDA, data visualisation, and demonstrate your understanding of data mining approaches.

You will use code comments and markdown to explain your reasoning, justify choices, and add your own interpretation.

Assignment Information – Objective

Importantly, much like data mining itself, this is an exploratory task; there is no clear end-goal or natural end-point.

Rather, you need to determine what level of data analysis is appropriate, given the scale and the timing of the assignment.

This course has hopefully taught you a broad range of approaches to try, and you should use EDA to guide you.

Assignment Information – Instructions

In MyAberdeen, under **Assessments**, there is the movies.csv file, a .ipynb notebook, and an information document.

Download the notebook file and the dataset, and modify this notebook to undertake your analysis.

Make sure that your name, email address, and student ID are all included within the markdown at the top of the template.

Assignment Information – Submission

To submit, you rename your (completed) Jupyter notebook with your student ID (e.g. ‘50080001.ipynb’).

You should **not** zip this file, or include any other files with your submission (e.g. the dataset).

You then upload this notebook to MyAberdeen, under Assessments/Assignment, before the deadline.

Assignment Information – Marking

The aim of this assignment is to demonstrate your:

- ability to undertake exploratory data analysis on new data.
- depth and breadth of knowledge with relation to data mining and visualisation.
- communication skills (clear, technical contents and sound reasoning).

Assignment Information – Marking

The assignment will be worth 100 marks:

- Appropriate use of EDA and descriptive statistics (~30 marks)
- Appropriate use of data mining techniques (~30 marks)
- Appropriate use of data visualization (~20 marks)
- Appropriate documentation via code comments and markdown (~20 marks)

In reality, grading will be more flexible with mark allocation, but this gives you some idea of how you should be prioritizing!

Assignment Information – Marking

Important: There are many ways in which you can explore and analyse this dataset. Therefore, we want to see your processes and justifications for any analysis that you carry out.

Use what you have learned throughout this course to demonstrate your aptitude in exploring and analysing data.

Keep **all** code/markdown documenting the process of EDA!

Assignment Information – Marking

Markdown should be used to explain what you are doing (at a high-level), and to add your own interpretation of what the data and the results are showing, at each step of your analysis.

Document your code well, so that we can follow along with your thinking for undertaking particular analyses or processes.

All markdown / code comments should be in English.

Assignment Information – Advice

Important: This assignment is not an ML challenge – Don't just focus on *prediction for prediction's sake*.

Your main objective is towards inference and understanding of the dataset's variables (and their relationships).

Simply training ML models for the sake of prediction will not result in a good mark.

Assignment Information – Advice

The dataset has many interesting attributes, including categorical, quantitative, and text-based variables.

Don’t just focus on one or two. Explore them! Use them to understand the broader patterns and trends within the data!

We are looking for curiosity and creativity!

Assignment Information – Important Points

Assignments are to be undertaken individually:

- Do not share code/scripts, or discuss the steps you have taken with your fellow coursemates.
- Do not obtain analysis scripts or code blocks online (without clearly outlining what is obtained and attributing the source)
- The university plagiarism policy applies:
  - Assignments will be checked for plagiarism, between students and across various online resources/repositories.

Assignment Information – Important Points

Assignments are to be undertaken individually:

Be safe to prevent yourself from experiencing any issues!

Any concerns you might have with this, please discuss it with us in the first instance.

and across various online resources/repositories.

Assignment Information – Extensions

You should apply for an extension by emailing your request (using your UoA email address) to: **uoa-ji-enquiries@abdn.ac.uk**

Students should also include any supporting evidence where possible e.g medical letter.

Please read the Extensions and Late Submission of Work policy for full details. This is linked in the information sheet on MyAberdeen.

Assignment Information – Late Penalties

If a piece of work has been submitted late without an extension (or beyond the agreed extension date) penalties will be applied in the following way:

- For coursework submitted up to 24 hours late, **we will deduct 2 CGS points from your grade**;
- For each subsequent day, up to a maximum of seven days total, **a further one CGS point will be deducted for each day**, or part of a day, up to a maximum of seven days late;
- Over seven days late, **a grade of G3 will be awarded**.

Assignment Information – Late Penalties

It is therefore very important that you submit your assignment before the deadline.

Any time beyond 23:59 on the submission due date will be considered late.

Make sure to leave enough time (> 30 minutes) to upload and submit the assignment itself.

Exam Information

Exam Information

This course will also involve a written exam, which will be worth 75% of your overall mark for this course.

This exam will take place around the first week of June.

Exam Information

Half of the exam will consist of multiple choice questions, across all of the course's content (i.e. lectures).

The remaining half of the exam will involve 3-4 larger questions, which will require more effort to answer.

A calculator is not permitted in the exam hall, so all required calculations will be possible without one.

Exam Information – Past Papers

A number of past papers are being made available to you via MyAberdeen, to help you revise and practice for the exam.

Alongside the practicals, lab sessions are a great opportunity to prepare for the exam, and raise any questions you have.

Note: some of the older past papers include questions that are no longer covered by this course. If you're not sure whether you should know something, ask during the labs.

Open Q&A / Discussion