UNIVERSITY OF ABERDEEN

University of Aberdeen

South China Normal University

Aberdeen Institute of Data Science

& Artificial Intelligence.

Please read all the information below carefully
Assignment Briefing Document – Individually Assessed (no teamwork)

Course: JC3503 – Data Mining & Visualisation (2024/2025)	***Note: This part of assessment accounts for 25% of your total mark of the course.***

Learning Outcomes

On successful completion of this component, a student will have demonstrated competence in the following areas:

- Ability to manipulate, format, prepare, and clean data sets prior to analysis.
- Ability to analyse complex datasets by applying data pre-processing, exploration, clustering and classification, time series analysis, and others.
- Ability to design appropriate visualisation solutions for different applications, scenarios, and audiences.

Information for Plagiarism and Conduct: Your submitted report may be submitted for plagiarism check (e.g., Turnitin). Please refer to the slides available at MyAberdeen for more information about avoiding plagiarism before you start working on the assessment. Please also read the following information provided by the university:

https://www.abdn.ac.uk/students/academic-life/academic-integrity/

https://www.abdn.ac.uk/staffnet/education/assessment/academic-integrity/plagiarism/

In addition, please familiarise yourselves with the following document "code of practice on student discipline (Academic)": https://tinyurl.com/y92xgkq6

Please read all the information below carefully

Overview:

This assignment tasks you with undertaking appropriate exploratory data analysis, data mining, and data visualisation techniques on a dataset. You are to determine what constitutes appropriate data analysis based on what you have learned during the Data Mining and Visualisation course.

Report Guidance & Requirements:

Your submission must contain a Jupyter notebook, with the corresponding code and comments/markdown, containing a full critical and reflective account of all of the processes undertaken, including during the process of exploratory data analysis (EDA).

The dataset needed to fulfil the requirements of this assessment can be found in MyAberdeen, alongside a Jupyter notebook, which you should use to undertake and document your analysis.

The deadline for this assignment is: Sunday 25th May, 2025, at 23:59 (UTC+8; Beijing time).

Dataset:

The assignment involves analysing the *movie.csv* dataset, which can be found on MyAberdeen (under Assessments).

Objectives:

The main purpose of this assignment is the following:

1. Demonstrate your ability to determine appropriate methods for exploratory data analysis (EDA), data mining, and data visualisation.
2. Undertake these appropriate methods in order to explore, understand, and analyse the data at hand.

Please read all the information below carefully

Submission Instructions:

In MyAberdeen, under **Assessments**, there is a .csv file (the dataset), and a Jupyter (.ipynb) notebook (the template file). Download and modify the notebook file to undertake your analysis. Make sure that you add your name, email address, and student ID to the markdown at the top of the template.

Prior to submission, you should rename the notebook to your student ID, e.g. '50080001.ipynb'. You should not zip your notebook, or upload any other files.

To submit, upload your (completed) Jupyter notebook, containing all of your analysis and markdown, to MyAberdeen (Assessments/Assignment).

Marking Criteria:

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).

The assignment will be worth 100 marks:

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

Important Points:

- 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.

- 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. Keep **all** code/markdown documenting the process of EDA, as this will contribute to your mark. **All markdown / code comments should be in English.**

- This assignment is not an ML challenge. While you may want to explore your ability to predict variables, you should do so with the goal of inference and understanding of the dataset's variables (and their relationships). Simply training ML models will not result in a good mark.

- The dataset has many interesting variables – Explore them all, and use them to understand the broader patterns and trends within the data.

Please read all the information below carefully

Extension Requests:

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 policy at Extensions and Late Submission of Work for full details.

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

It is therefore very important that you submit your assignment before the deadline, leaving enough time to upload and submit the assignment itself.

Policy on Group Working:

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.

Additional Information / Questions:

Any questions pertaining to any aspects of this assessment, please contact your lecturer:

AI: Chris Norval < chris.norval@abdn.ac.uk >

BMIS: Jiangtian Nie < jiangtian.nie@abdn.ac.uk >

CS: Siwei Liu < siwei.liu@abdn.ac.uk >