University of Aberdeen

South China Normal University

Aberdeen Institute of Data Science

& Artificial Intelligence.

UNIVERSITY OF ABERDEEN

Please read all the information below carefully

Assessment I Briefing Document – Individually Assessed (no teamwork)

Course: JC3503 – Data Mining & Visualisation (2023/2024)	***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/sls/onlinehttps://www.abdn.ac.uk/sls/online-resources/avoiding-plagiarism/resources/avoiding-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 two datasets. Students are to determine what constitutes appropriate analysis to be undertaken based on what they have learned during the Data Mining and Visualisation course.

Report Guidance & Requirements:

Your submission must contain two Jupyter notebooks (one for each dataset), 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 two datasets needed to fulfil the requirements of this assessment can be found in MyAberdeen, alongside two .ipynb Jupyter notebooks that are to be used to undertake and document your analysis.

The deadline for this assignment is: 23:59, Tue 28th May, 2024

Datasets:

The assignment involves analysing two datasets:

- Classifieds: Apartments for Rent
- Students' Dropout and Academic Success

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 **Assignments**, there are two .csv files, two Jupyter (.ipynb) notebooks, and documents with information and context of the variables. Download and modify the notebook files to undertake your analysis.

To submit, upload your two (completed) Jupyter notebooks containing all of your analysis and markdown to MyAberdeen.

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. 20 marks)
- Appropriate use of data visualization (approx. 20 marks)
- Appropriate use of data mining techniques (approx. 40 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 data. Therefore, we want to see your processes and justifications for any analysis that you carry out.

- 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. This will contribute to your marks.

- This assignment is not an ML challenge. While you may want to explore your ability to predict variables, inference and understanding of the dataset's variables (and their relationships) is equally (if not more) important.

- The datasets have many interesting variables – Explore them, and use them to understand the data more broadly.

- The assignment involves exploring, analysing, and mining both datasets – **Avoid just focusing on one at the expense of the other!**

Please read all the information below carefully

Additional Information / Questions:

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.

Any questions pertaining to any aspects of this assessment, please contact your lecturer in the first instance:

AI:	Chris Norval	< chris.norval@abdn.ac.uk >
BMIS:	Junfeng Gao	< Junfeng.gao@abdn.ac.uk >
CS:	Siwei Liu	< siwei.liu@abdn.ac.uk >