CCT College Dublin Continuous Assessment

Programme Title:	BSc (Hons) in Computing in IT (4th Yr)		
Delivery Mode:	FT		
Cohort Details:	*BSc (Hons) in Computing in IT - Sept 2021 cohort, Semester 7*		
Module Title(s):	*Data Exploration & Preparation*		
Assignment Type:	*Individual*	**Weighting(s):**	*40%*
Assignment Title:	*CA1*		
Lecturer(s):	*Dr Muhammad Iqbal*		
Issue Date:	*1st October 2024*		
Submission Deadline Date:	*24th November 2024 (23:59 IST)*		
Late Submission Penalty:	Late submissions will be accepted up to **5** calendar days after the deadline. All late submissions are subject to a penalty of **10%** of the mark awarded. Submissions received more than 5 calendar days after the deadline above **will not** be accepted and a mark of 0% will be awarded.		
Method of Submission:	**This assignment is submitted via Moodle.**		
Instructions for Submission:	*Upload all files MS word file, R code or jupyter notebook, dataset and any supporting information on Moodle separately.*		
Feedback Method:	**Results posted in Moodle gradebook**		
Feedback Date:	*15th December 2024*		

Assessment Outline

This is an individual assignment using the R programming language or any other language of your choice. Analyse a specific problem only in the one of following areas

- Environment
- Population and Society
- Economy and Finance

The dataset should have at least 3000 rows and 7 columns after cleaning and there is not any upper bound. The type of question(s) that you should formulate for the assignment will depend on the chosen domain of the dataset that you are considering for the Data Exploration and Preparation (DEP) assignment. The objectives of the DEP assignment are based on the domain knowledge of data. You should complete the following tasks during the development of this assignment.

You will present your findings and defend the results in the form of a PowerPoint poster presentation.

i. Description of problem domain, motivation, data set chosen and challenges faced during this assignment. You should provide the characterization, description and explanation of techniques used to prepare the data set (size / attributes / missing values / outliers) by exhibiting code snippets.

(25 marks)

ii. Find unusual patterns by identifying variations and covariation between the features in the dataset and perform Exploratory Data Analysis (EDA) to justify outcomes with supporting questions and visualizations.

(25 marks)

iii. Show the implementation of an encoding scheme, such as one-hot, Label etc. Apply Principal Component Analysis (PCA) for the dimensionality reduction on the chosen dataset. Interpret and explain the outcomes obtained using PCA.

(20 marks)

iv. Present your results and findings in the form of a PowerPoint poster (template provided on Moodle). Ensure that the poster effectively communicates the key aspects of this assignment. Submit the code developed for the assignment including comments to explain its functionality in detail. Conclusions, citations and references should be provided in the poster using the Harvard Style.

(20 marks)

v. Provide a video recording (maximum 5 minutes) explaining your key findings from the poster and offering an overview of the code developed for the assignment.

(10 marks)

Assessment Requirements

All assessment submissions must meet the following minimum requirements:

- Be submitted in the format outlined in the assignment summary table.

- The code and datasets should be provided and uploaded on Moodle.
- Number of Words in the poster presentation should be in between 400 to 500 excluding the citations and references.
- Be submitted by the deadline date specified or be subject to late submission penalties.
- Be submitted via Moodle upload.
- Use Harvard Referencing when citing third party material.
- Be the student's own work.
- Include the CCT assessment cover page.
- Students must use the classroom GitHub link (https://classroom.github.com/a/wEEY_D2y) and you should have more than 7 commits on GitHub on different dates and times.
- No use of Gen AI voice used for the recording.
- Use of Gen AI (chatGPT or Copilot or others) must be clearly stated for any part of your assignment in the Assignment submission declaration form.

Learning Outcomes:

This assessment addresses the following module learning outcomes for this module:

- Develop strategies for identifying and handling missing and out-of-range data, as well as feature engineering as part of the preparation phase of data analysis. (Linked to PLO 4 (Stage 4 SLO 4))
- Understand the purpose of and methods to achieve dimensionality reduction and the difference between dimensionality reduction and feature selection. (Linked to PLO 1 / PLO 3 (Stage 4 SLO 1 / SLO 3))
- Select and perform appropriate feature selection and/or dimensionality reduction techniques on a variety of wide datasets. (Linked to PLO 3 (Stage 4 SLO 3))

Statement of Acceptable Use of Artificial Intelligence

Acceptable and Unacceptable Use of AI

- The use of generative AI tools (e.g. ChatGPT, Dall-e, etc.) is permitted in this assignment for the following activities:
  - Brainstorming and refining your ideas;
  - Fine tuning your research questions;
  - Finding information on your topic;
  - Drafting an outline to organise your thoughts; and
  - Checking grammar and style.
- The use of generative AI tools is not permitted in this course for the following activities:
  - Impersonating you in classroom context
  - Completing group work that your group has assigned to you
  - Generating code for your assignment
  - Writing a draft of a writing assignment
  - Writing entire sentences, paragraphs or papers to complete class assignments.
- You are responsible for the information you submit based on an AI query. Your use of AI tools must be properly documented and cited.
- Any assignment that is found to have used generative AI tools in an unauthorised way will be subject to college disciplinary procedures as outlined in the QA Manual.
- When in doubt about permitted usage, please ask for clarification.

Grading Criteria

This grading rubric sets out the marking criteria for your assignment.

Criteria	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5
Marking Criteria	**Problem Description, Motivation, Dataset Characterization and Challenges**	**Exploratory Data Analysis (EDA) and Unusual Pattern Identification**	**Encoding Schemes & PCA Implementation**	**PowerPoint Poster Presentation and Code Explanation**	**Video Presentation**
Weighting per Criteria	**25 marks**	**25 marks**	**20 marks**	**20 marks**	**10 marks**
Excellent (+70%)	Clear and thorough problem description, motivation, and dataset characterization with detailed handling of challenges and data preparation.	Comprehensive EDA with well-justified patterns and strong visualizations identifying variations and covariations between features.	Accurate implementation of encoding and PCA with thorough interpretation and explanation of outcomes aligned to problem objectives.	Exceptional poster with clear, concise communication of key aspects; code fully commented, and conclusions supported with Harvard-style citations/ references.	Excellent video with clear, structured explanation of findings in the poster and code.
Very Good (60-69%)	Convincing problem description, motivation, and dataset characterization with strong data preparation efforts.	Strong EDA with well-identified patterns, supported by good visualizations.	Well-executed encoding and PCA with strong interpretation and alignment with objectives.	Very good poster with clear communication and well-commented code; conclusions, citations and references well-supported.	Very good video with structured explanation of findings and code.
Good (50-59%)	Satisfactory problem description and motivation; dataset characterization and preparation are present but lack depth.	Good EDA with clear identification of patterns; visualizations are useful but could be stronger.	Adequate encoding and PCA; interpretation of results is acceptable and required more depth.	Good poster with sufficient communication; code comments and conclusions are present, though may lack detail.	Good video with clear, but somewhat incomplete explanation of findings and code.
Acceptable (40-49%)	Partial Satisfactory problem description and motivation; dataset characterization is partially complete or required more detail.	Basic EDA with partial identification of patterns; limited or weak visualizations.	Encoding and PCA implementation is partially complete or required more depth and interpretation.	Acceptable poster with limited communication of key aspects; code comments and conclusions are minimal. Citations and references required improvements.	Acceptable video and required more clarity and detailed explanation.
Fail (<39%)	Inadequate or unclear problem description and dataset characterization; little or no data preparation is evident.	Poor or missing parts of EDA; no meaningful pattern identification or visualizations.	Encoding and PCA are either missing or poorly executed with no clear interpretation.	Poor poster, little communication of key aspects; code comments and conclusions are missing or inadequate.	Poor or missing video, unclear or no explanation of findings and code.

The Irish Grading System

The grading system in CCT is the QQI percentage grading system and is in common use in higher education institutions in Ireland. The pass mark and thresholds for different grade bands may be different from what you have experienced in the higher education system in other countries. CCT grades must be considered in the context of the grading system in Irish higher education and not assumed to represent the same standard the percentage grade reflects when awarded in an international context.

Please review the CCT Grade Descriptor available on the module Moodle page for a detailed description of the standard of work required for each grade band, and review the marking criteria outlined in this assignment brief for a breakdown of the marking criteria for this specific assignment.

Additional Information

- Lecturers are not required to review draft assessment submissions. This may be offered at the lecturer's discretion.
- In accordance with CCT policy, feedback to learners may be provided in written, audio or video format and can be provided as individual learner feedback, small group feedback or whole class feedback.
- Results and feedback will only be issued when assessments have been marked and moderated / reviewed by a second examiner.
- Additional feedback may be provided as individual, small group or whole class feedback. Lecturers are not obliged to respond to email requests for additional feedback where this is not the specified process or to respond to further requests for feedback following the additional feedback.
- Following receipt of feedback, where a student believes there has been an error in the marks or feedback received, they should avail of the recheck and review process and should not attempt to get a revised mark / feedback by directly approaching the lecturer. Lecturers are not authorised to amend published marks outside of the recheck and review process or the Board of Examiners process.
- Students are advised that disagreement with an academic judgement is not grounds for review.
- For additional support with academic writing and referencing students are advised to contact the CCT Library Service.
- For additional support with subject matter content students are advised to contact the CCT Student Mentoring Academy
- For additional support with IT subject content, students are advised to access the CCT Support Hub.