A picture containing icon

Description automatically generated

**CCT College Dublin Continuous Assessment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Programme Title:** | *HDIP Data Analytics* | | |
| **Cohort:** | *HDipData\_Feb24\_FT/SB+/HCI* | | |
| **Module Title(s)**: | *Full approved module title* | | |
| **Assignment Type:** | *Data Preparation* | Weighting(s): | *50%* |
| **Assignment Title:** | *CA1\_DataPrep\_HDip* | | |
| **Lecturer(s)**: | *David McQuaid* | | |
| **Issue Date:** | *08/03/2024* | | |
| **Submission Deadline Date:** | *05/04/2024* | | |
| **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:** | **Moodle** | | |
| **Instructions for Submission:** | Assessment must be submitted to Moodle before 05/04/2024 11:59pm as a Jupyter Notebook file  The Jupyter Notebook File Must be saved as “YourName\_DPrepHDip\_CA1.ipynb”  CA Cover Page as Word Document  NO PDF’s!  NO PYTHON FILES (.py) | | |
| **Feedback Method:** | **Results posted in Moodle gradebook** | | |
| **Feedback Date:** | Approx. 3 weeks after FINAL submission (inc PMC cases) | | |

**Learning Outcomes:**

Please note this is not the assessment task. The task to be completed is detailed on the next page.

This CA will assess student attainment of the following minimum intended learning outcomes:

* Develop strategies, incorporating basic programming skills (input / output and basic data structures) for identifying and handling missing and out-of-range data. (linked to PLO 4)
* Programmatically implement graphical methods to identify issues within a data set (missing, out of range, dirty data) (linked to PLO 2, PLO 3)
* Perform a critical analysis of a data set to optimise the data for a given problem space.
* Document the rationale behind the decisions to peers and stakeholders. (linked to PLO 1,PLO 3, PLO 6)

Attainment of the learning outcomes is the minimum requirement to achieve a Pass mark (40%). Higher marks are awarded where there is evidence of achievement beyond this, in accordance with QQI *Assessment and Standards, Revised 2013*, and summarised in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| **Percentage Range** | **CCT Performance Description** | **QQI Description of Attainment** | |
| **Level 6, 7 & 8 awards** |
| 90% + | Exceptional | Achievement includes that required for a Pass and in **most** respects is significantly and consistently beyond this |
| 80 – 89% | Outstanding |
| 70 – 79% | Excellent |
| 60 – 69% | Very Good | Achievement includes that required for a Pass and in **many** respects is significantly beyond this |
| 50 – 59% | Good | Achievement includes that required for a Pass and in **some** respects is significantly beyond this |
| 40 – 49% | Acceptable | Attains all the minimum intended programme learning outcomes |
| 35 – 39% | Fail | Nearly (but not quite) attains the relevant minimum intended learning outcomes |
| 0 – 34% | Fail | Does not attain some or all of the minimum intended learning outcomes |

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.

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 experience of 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.

**CA1 NOTE DO NOT ZIP YOUR SUBMISSION FILES, ALL FILES MUST BE SUBMITTED INDIVIDUALLY**

**Acceptable and Unacceptable Use of AI**

|  |  |
| --- | --- |
| **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   + Writing a draft of a writing assignment   + Writing entire sentences, paragraphs, papers, code fragments, functions, scripts 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. |

**Note ALL Students are required to use Git for any Assignments that they are working on.**

This means that ALL changes must be committed to Git during your assignment. (Not just a single commit at the end!) This is to allow you to display your incremental progress throughout the assessments, allows you to create an online portfolio that can be used to showcase your work and to ensure that there are no problems with final uploads (as all your work will be available on GitHub). It is expected that there will be a minimum of 10 commits (with many of you making very many more).

You may Only use your CCT email for your git account, private/work email-based accounts will not be accepted. You must also include your lecturer's CCT email as a collaborator on your account.

**Assessment Task**

**Students are advised to review and adhere to the submission requirements documented after the assessment task.**

All data wrangling, analysis, and visualizations must be generated using python.

Minimum Requirements

You are required to use the dataset contained within the file “student\_data.csv”, conduct the following analysis and report on your findings:

* Characterisation of the data set: size; number of attributes; has/does not have missing values, number of observations etc. and what these Characterisations mean. [0-10]
* Application of Data preparation/evaluation methods (Cleaning, renaming, etc) and EDA visualizations (plural), including a clear and concise explanation of your rationale for what you are doing with the data and why you are doing it. [0-40]
* Use PCA to establish the minimum number of features needed for retaining 99.5% variance in the data and then implement PCA to dimensionally reduce the data to the number of features that you have discovered. Include a clear and concise explanation of your rationale for what you are doing with the data and why you are doing it. [0-30]
* Explain **in your own words** what the “Curse of Dimensionality ” is and how it may affect your analysis going forward. [0-10]
* Conclusions, Findings of data set and references (HARVARD style). [0-10]

Note that all written work MUST be completed in Jupyter Notebook Markdown (please review “Jupyter Notebook Tutorial” Notes in Moodle if you are unsure of this).

All Code must be included in code blocks (As normal). No other upload will be accepted.

All written work MUST be detailed in your Jupyter Markdown (NOT in code comments).

**Data Dictionary**

Please Review the Excel File “student\_data\_dictionary.xlsx”

**NOTE your dependant feature is called “Target” and has 3 categories: dropout, enrolled, and graduate.**

**Submission Requirements**

All assessment submissions must meet the minimum requirements listed below. Failure to do so may have implications for the mark awarded.

All assessment submissions must:

* Be submitted before 05/04/2024 11:59pm as a Jupyter Notebook file.
* The Jupyter Notebook File Must be saved as “YourName\_DPrepHDip\_CA1.ipynb”
* 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.

**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 or access the [CCT Learning Space](http://learningspace.cct.ie/subjects/index.php).
* For additional support with subject matter content students are advised to contact the [CCT Student Mentoring Academy](https://moodle.cct.ie/mod/forum/view.php?id=55148)
* For additional support with IT subject content, students are advised to access the [CCT Support Hub](https://moodle.cct.ie/course/view.php?id=1861).