A picture containing icon

Description automatically generated

**CCT College Dublin Continuous Assessment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Programme Title:** | *HDip in AI Applications* | | |
| **Delivery Mode:** | Online | | |
| **Cohort Details:** | *Programming for AI (HDip in AI Applications - Sept 2024 - HCI cohort)* | | |
| **Module Title(s)**: | *Programming AI Concepts* | | |
| **Assignment Type:** | *Individual* | **Weighting(s):** | ***40%*** |
| **Assignment Title:** | *CA1* | | |
| **Lecturer(s)**: | *Taufique Ahmed* | | |
| **Issue Date:** | *03-10-2024* | | |
| **Submission Deadline Date:** | *Thursday, 31-10-2024 at 11:59pm* | | |
| **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:** | *All files must be submitted separately.*  *Expected files : Written report (word ONLY), Code files (Jupyter Notebook), recorded video of Jupyter notebook with explanation.* | | |
| **Feedback Method:** | **Results posted in Moodle gradebook** | | |
| **Feedback Date:** | *3 weeks after submission* | | |

# Assessment Outline

## Description of Assessment Task

Complete the following tasks involving Python, NumPy, pandas, SQL, and exploratory data analysis (EDA). Your solutions should include user-defined functions, database integration, and an EDA coding and report on the Iris dataset. Submit a 1500-word report interpreting your findings, discussing the significance of each task in artificial intelligence.

**Q1 (15 marks)**

Select two datasets and create a user defined function in Python that leverages the fundamental features of the Python without using Pandas built in functions such as

(numeric\_df =df.select\_dtypes(include=['number'])

categorical\_df = df.select\_dtypes(exclude=['number'])

Your function should accept a Data Frame as input and categorize its columns into numeric and categorical types, then display the lists of numeric and categorical columns. Include a section in your report where you discuss your interpretation of this task and its significance within the field of artificial intelligence and make sure to run the code and attach a screenshot of your machine in the appendix of your report.

**Q2 (15 marks)**

Write a function, the function should achieve the following objectives:

1. In pandas:
   * Concatenate the two datasets, which you have used in Question1, along the rows.
   * Remove any duplicate rows.
   * Print the number of rows and columns in the resulting DataFrame.
2. In NumPy:
   * Calculate the correlation matrix for all numeric columns.
   * Identify the pair of columns with the highest correlation coefficient.
   * Print the names of these columns along with their correlation coefficient.

Include a section in your report where you discuss your comprehension of this task and its relevance in the field of data pre-processing and analysis using pandas and NumPy libraries. Make sure to run the code and attach a screenshot of your machine in the appendix of your report.

**Q3 (10 marks)**

Writ a python program to implement the below algorithm:

* Create a NumPy array with 1000 random elements (numbers) and take the mean of every 5 sample window.
* [datamean] < -- mean ( [ numpy\_array (1 : 5 : end ) ] )
* [data]min < -- min ( [ datamean ] )
* [data]max < -- max ( [ datamean ] )
* [ value ] max < -- max ( abs ( [data]max), abs([ [data]min])

Include a section in your report where you discuss your interpretation and significance of this task and make sure to run the code and attach a screenshot of your machine in the appendix of your report.

**Q4 (10 marks)**

Write a code for database integration and manipulating the data using SQL clause. Explain the significance of RDBMS such as SQL in the field of artificial intelligence (AI). Discuss why they are essential for tasks such as data storage, retrieval, pre-processing, and integration in AI applications. Describe the process of connecting to a MySQL database, creating tables, and storing values. Highlight how proficient utilization of databases and SQL enhances the efficiency and effectiveness of AI systems. Make sure to run the code and attach a screenshot of your machine in the appendix of your report.

**Q5 (15 marks)**

Conduct an exploratory data analysis (EDA) on the Iris dataset from scikit learn using Python. The objective is to provide a comprehensive overview of the dataset, including an examination of its features and target variables, as well as understanding the data structure—specifically its shape, data types, and any missing values. The analysis should include descriptive statistics for each feature, along with visualizations such as histograms and boxplots to explore feature distributions and relationships among the species. Additionally, a correlation matrix should be generated to identify significant correlations between features. Finally, the report should include interpretations and conclusions drawn from the analyses and visualizations, effectively summarizing insights gained from the EDA process. Make sure to run the code and attach a screenshot of your machine in the appendix of your report.

**Report (25 marks)**

A report (up to 1500 words excluding appendix section and conclusion) that provides a critical analysis and interpretation of the findings derived from all the previous questions.

## Screen Recording (10 marks):

## You are required to submit a video, no longer than 7 minutes, demonstrating your Jupyter Notebook in action, with a step-by-step explanation of the code.

## Assessment Requirements

All assessment submissions must meet the following minimum requirements:

* Be submitted in the format outlined in the assignment summary table.
* Report (maximum 1500 words), and screen recorded video with explanation of code.
* You are required to create a GitHub account, upload your CA work, and grant access to my email address. Ensure to include the GitHub account link in your CA Word document.
* Be submitted by the deadline date specified or be subject to late submission penalties.
* Be submitted via Moodle upload.
* Use [Harvard Referencing](http://40.115.124.2/sp/subjects/guide.php?subject=harvardref) when citing third party material.
* Be the student’s own work.
* Include the CCT assessment cover page.

## Learning Outcomes:

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

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:

1. Demonstrate knowledge of fundamental programming concepts.

(Linked to PLO 1)

1. Develop a requirements specification to Integrate file handling and database integration as part of given cognitive system proposal. (help / chat bot, game of life ) (Linked to PLO 4, PLO 5)

**Statement of Acceptable Use of Artificial Intelligence**

|  |
| --- |
| **Use Prohibited** |
| * The use of generative AI tools (such as ChatGPT, DALL-E, etc.) is not permitted in this assignment. * 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](https://www.cct.ie/wp-content/uploads/QA-Manual-2022-23-Revised-definitive-3-1.pdf). * 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***  *Use of fundamental Python concepts to answer the first question* | ***Criteria 2***  *Use of the Pandas and NumPy libraries to answer the second question* | ***Criteria 3***  Use of basic Python concepts to write the Python code for the given pseudocode | ***Criteria 4***  Integrating Python  with MySQL and manipulating the data using SQL clause | ***Criteria 5***  Conducting exploratory data analysis and providing a critical analysis of the obtained results. | ***Criteria 6***  *Technical report of 1500 words* | ***Criteria 7***  *Screen recorded video of in action Jupyter notebook* |
| **Weighting per criteria** | **15 marks** | **15 marks** | **10 marks** | **10 marks** | **15 marks** | **25 marks** | **10 marks** |
| *Excellent (+70%)* | Comprehensive application of Python fundamentals with efficient, error-free code and clear categorization. | Effective and advanced use of Pandas and NumPy for data manipulation, producing accurate results with optimized code. | Precise implementation of pseudocode; optimized and logically structured code producing accurate results. | Seamless integration with clear knowledge of database connectivity and efficient SQL manipulation. | Comprehensive EDA with insightful visualizations and critical interpretations identifying key trends and correlations. | Detailed, well-structured, and insightful report with critical analysis and clear explanations of findings. | Clear, concise explanation of code execution; no errors, demonstrating deep understanding. |
| *Very Good (60 - 69%)* | Solid understanding with minor mistakes; categorization mostly correct but could be optimized. | Good usage with minor mistakes; proper handling of datasets with some inefficiencies. | Correct implementation with minor errors or inefficiencies in logic or execution. | Good integration with minor issues; proper usage of SQL clauses but some inefficiencies present. | Good analysis with a solid range of visualizations; minor gaps in insights or interpretations. | Good report with minor inconsistencies; adequate discussion of findings. | Well-explained video with few minor mistakes; overall clarity maintained. |
| *Good (50 - 59%)* | Basic understanding with functional code but some inefficiencies and errors present. | Satisfactory usage, but code is inefficient or contains minor errors in functionality. | Basic implementation with some errors; functional but lacking clarity and efficiency. | Basic integration with inefficient use of SQL; functional but not optimized. | Basic EDA with sufficient visualizations but limited depth in analysis; important insights missed. | Satisfactory report lacking depth in analysis and structure; some important points missing. | Adequate explanation but lacks depth or contains some gaps in understanding. |
| *Acceptable (40 - 49%)* | Limited understanding; multiple errors and unclear categorization. | Limited application, significant mistakes in handling data and correlations. | Significant errors; incomplete logic resulting in incorrect or incomplete code. | Partial integration with many errors; struggles with proper data manipulation. | Minimal analysis; incomplete visualizations or inaccurate interpretations. | Poorly structured report with limited discussion; fails to critically analyze findings. | Poor explanation with significant code errors; lacks clarity in demonstration. |
| *Fail (< 39%)* | Lacks fundamental knowledge, significant errors, or incomplete task. | Incorrect or no use of libraries, major issues in functionality. | Major errors or incomplete implementation, lacking basic understanding. | No or incorrect integration, major issues in SQL usage or functionality. | EDA is absent or contains severe errors; no meaningful insights or visualizations. | Missing or severely incomplete report with no relevant discussion. | Missing or unclear video; fails to demonstrate understanding of concepts. |

**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](https://moodle.cct.ie/course/view.php?id=827)
* 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).