**A logo for college computing

Description automatically generated**

**Assessment Cover Page**

|  |  |
| --- | --- |
| ***Student Full Name: Conor Kirwan*** |  |
| ***Student Number: 24130*** |  |
| ***Module Title: Programming for AI*** |  |
| ***Assessment Title: CA1*** |  |
| ***Assessment Due Date: 31/Oct/2024*** |  |
| ***Date of Submission: 31/Oct/2024*** |  |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

**Abstract**

**[NOTE: This section is designated for the abstract. Abstracts are not assigned page numbers and should precede the table of contents. If an abstract is unnecessary for your work, please delete this page.]**

**Attention: All notes must be removed from the document before submission!!**

**Contents**

[Question 1 4](#_Toc181020356)

[Question 2 4](#_Toc181020357)

[Question 3 4](#_Toc181020358)

[Question 4 4](#_Toc181020359)

[Question 5 4](#_Toc181020360)

[References 4](#_Toc181020361)

[Appendix 4](#_Toc181020362)

[**Assessment Outline** 5](#_Toc181020363)

[**Description of Assessment Task** 5](#_Toc181020364)

[**Screen Recording (10 marks):** 7](#_Toc181020365)

[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. 7](#_Toc181020366)

[**Assessment Requirements** 7](#_Toc181020367)

[**Learning Outcomes:** 8](#_Toc181020368)

# Introduction

# Question 1

# Question 2

# Question 3

# Question 4

# Question 5

# Conclusion

# References

# Appendix

**A screenshot of a computer

Description automatically generated**

**[NOTE: For the table of contents to function properly, you must use the correct headings for all your chapters and subchapters.**

**Heading 1: This is the main heading and should be employed for the primary title or chapter. For example: CHAPTER 1.**

**Heading 2: Use Heading 2 as a subheading. For instance: Chapter 1.1.**

**Heading 3: Heading 3 provides a more detailed breakdown, such as Chapter 1.1.1.**

**By adhering to this hierarchical structure, you ensure an organized and effective document outline, enhancing readability and navigation. However, you are not forced to use all 3 headings, usually heading 1 and 2 are sufficient.**

**The remainder of your text should be written using a normal font.]**

**References**

**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)