

**CCT College Dublin Continuous Assessment**

|                                     |   |                      |   |
|-------------------------------------|---|----------------------|---|
| <b>Programme Title:</b>             | <i>BSc (Hons) in Computing and IT Y4</i>  |                      |   |
| <b>Cohort:</b>                      | <i>Y4</i>   |                      |   |
| <b>Module Title(s):</b>             | <i>Artificial Intelligence<br/>Data Visualisation &amp; Comms<br/>Data Exploration &amp; Prep</i>   |                      |   |
| <b>Assignment Type:</b>             | <i>Individual</i>   | <b>Weighting(s):</b> | <i>40% (AI)<br/>40% (Data Vis and Coms)<br/>40% (Data Exploration &amp; Prep)</i> |
| <b>Assignment Title:</b>            | <i>AI_DV_DP_Lv8_ICA_v1</i>  |                      |   |
| <b>Lecturer(s):</b>                 | <i>David McQuaid<br/>Marina Soledad Iantorno<br/>Muhammad Iqbal</i>   |                      |   |
| <b>Issue Date:</b>                  | <i>13th December 2021</i>   |                      |   |
| <b>Submission Deadline Date:</b>    | <i>16th January 2022</i>  |                      |   |
| <b>Late Submission Penalty:</b>     | Late submissions will be accepted up to <b>5</b> calendar days after the deadline. All late submissions are subject to a penalty of <b>10%</b> of the mark awarded.<br>Submissions received more than 5 calendar days after the deadline above <b>will not</b> be accepted and a mark of 0% will be awarded.  |                      |   |
| <b>Method of Submission:</b>        | <b>Moodle</b>   |                      |   |
| <b>Instructions for Submission:</b> | <i>Assessment must be submitted before 11.55pm 16th January 2022 as a Jupyter Notebook file and a CSV file containing your Dataset</i> <ul style="list-style-type: none"> <li><i>The Jupyter Notebook File Must be saved as "YourName_AI_DV_DP_ICA.ipynb"</i></li> <li><i>Word Document with report detailing rationale, including visualisations "YourName_AI_DV_DP_ICA.docx"</i></li> <li><i>If using R you must provide R script files or 2<sup>nd</sup> R based Jupyter Notebook</i></li> </ul> |                      |   |
| <b>Feedback Method:</b>             | <b>Results released on Moodle</b>   |                      |   |
| <b>Feedback Date:</b>               | <i>2 weeks after final submission Inc PMC</i>   |                      |   |

**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:

## Artificial Intelligence

MLO 2 - Distinguish the different agents and environments of current Artificial Intelligence, being aware of consideration to perception / action and potential changes to the environment.

(Linked to PLO 2 (Stage 4 SLO 2))

MLO 3 - Understand the differences and challenges involved in developing different levels of Artificial Intelligence (Linked to PLO 3 (Stage 4 SLO 3))

MLO 4 - Identify and apply an appropriate problem-solving strategy in relation to search, non-classical search, Adversarial Search, Constraint Satisfaction Problem.

(Linked to PLO 5 (Stage 4 SLO 5))

## Data Visualisation & Comms

MLO 1 - Explain the concepts, techniques and processes underlying data visualisation

(Linked to PLO 1 (Stage 4 SLO 1))

MLO 2 - Distinguish between visualisation approaches and techniques with respect to their suitability for different problem areas.

(Linked to PLO 5 (Stage 4 SLO 5))

MLO 3 - Select appropriate data visualisation techniques for a given use case and data characteristics.

(Linked to PLO 3 (Stage 4 SLO 3))

## Data Exploration & Prep

MLO 1 - Choose and implement suitable data-encoding techniques for a variety of machine learning algorithms. (Linked to PLO 1 (Stage 4 SLO 1))

MLO 2 - 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))

MLO 3 - Engineer new features selection in data with the goal of improving the performance of machine learning models (Linked to PLO 2 (Stage 4 SLO 2))

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 <b>most</b> 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 <b>many</b> respects is significantly beyond this                  |
| 50 – 59%         | Good                        | Achievement includes that required for a Pass and in <b>some</b> 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   |

## Assessment Task

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

Forests are prime sources of the Earth's capacity to establish a balance in the climate, by the global impact of their photosynthesis. They provided a strong and common defence against climate change, taking out the greenhouse gas carbon dioxide and generating oxygen for the planet. This is useful in purifying the atmosphere and controlling jumping up temperatures. However, forest fires create a big challenge to climate change. A data set is provided on Moodle along with source link is mentioned below

<https://archive.ics.uci.edu/ml/datasets/Forest+Fires>

Your responsibilities are to explore this data by considering the following questions for the AI, Data Visualization techniques and data exploration and preparation.

**Tasks for Data Exploration and Preparation ( You may use R or Python for this Section , If using R you must provide R script files or 2nd R based Jupyter Notebook)**

- Discuss and describe the purpose and need for data exploration and mining within data analytics projects. Describe the possible consequences of relying on completely automatic data analysis tools rather than using manual statistical tools. Use illustrations to justify your answer.
- Are outliers present in the above data set? Provide a strategy of removing outliers from the data set.
- Choose a relevant encoding technique to transform the categorical data to numeric format.
- Formulate the questions for EDA (Exploratory Data Analysis) and perform EDA based on the data provided and address the issues to understand the impact of fire and the resulting burned area.
- Explain the purpose of PCA (Principal Component Analysis) and how It is effective to reduce dimensionality of the dataset. Find two components of the above data set.
- Analyse the benefits and drawbacks for feature selection and extraction techniques. Which features are important from the provided dataset? (Feature Selection Method)
- Prepare the data for the machine learning model after encoding and feature selection, predict the burned area of forest fires using the machine learning model.
- You are free to perform any process to enhance an understanding of data preparation and exploration of the Forest data set.

**Tasks for Data Visualisation**

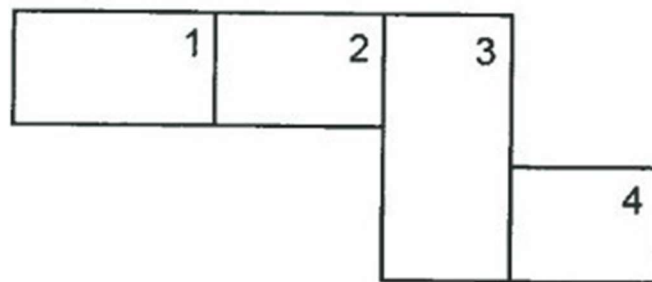
- Analyse the data, select appropriate visualisations to summarise the data before cleaning it and perform the same visualisations after data cleaning. Summarise any differences that you discover. Briefly explain the advantages of preparing the data for further analysis in terms of graphic presentation.
- Use more than one interactive visualisation to highlight any impact that forest fire has on the environment. You are required to use additional data and merge it with the dataset provided to complete this task.
- Use statistical visualisations to describe and identify different aspects of the data (e.g., median, outliers, probabilistic histogram, etc...).
- Compare and contrast at least two different visualisation libraries that you used to complete the tasks of this assignment. Detail the differences and/or similarities between these libraries. Rationalise your design choices such as colour selection, font size, use of white spaces, etc...
- Pick a target audience of your choice, design your visualizations with this audience in mind.
- Include final visualisations relevant to the fire officers to illustrate your findings

## Tasks for Artificial Intelligence

- 1) In honour of the hard work that the fire officers are doing, the community has decided to build a new fire station for them. There are seven fire officers and four offices. Because there are more fire officers than offices, some fire officers have to be in the same offices as others. However, the fire officers are very picky about who they share with. The community administration is having trouble assigning fire officers to offices. They have asked you to plan where each fire officer goes.

The fire officers are **Phylis, Ann, Henry, Eva, Bill, Mark, and Bob**.

They have given you the plans of the fire station.



Each numbered area is an office in the fire station. Multiple officers can go into the same office, and not all offices have to be filled.

Each fire person has restrictions about where they can be placed.

1. **Phylis** and **Eva** don't get on, and do not want to be in the same office.
2. **Mark** and **Bob** are best friends, and have to be in the same office.
3. **Henry** listens to loud music. Only **Eva** will share his office.
4. **Eva** doesn't talk to **Mark, Bob, and Bill**.
5. **Ann** is always late but is a great fire officer. To hide that **Ann** is always late, **Ann** cannot be in either the same office or in an office adjacent to **Phylis** or **Eva**.
6. **Phylis** annoys **Bill**, so **Bill** doesn't want to be in **Phylis**'s office.
7. **Phylis** is the fire chief, so she wants to be in office 1.

Using any CSP (Constraint Satisfaction Problem) framework, discover if the above problem can be solved and if so detail who would be in each office.

- 2) Discuss in detail how using Constraint Satisfaction finds an answer or finds no solution to the problem in Tasks for Artificial Intelligence part 1. How does this differ from standard algorithmic solutions?
- 3) Can this problem be solved using any other algorithm we have studied in the module? Discuss your answer in detail including a proof of your hypothesis.
- 4) Support your arguments with citations/references in Harvard Style

## 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 **11.55pm 16th January 2022** as a Jupyter Notebook file.
- The Python Jupyter Notebook File Must be saved as “YourName\_AI\_DV\_DP\_ICA.ipynb”, and the dataset you have used.
- If using R you must provide R script files or 2nd R based Jupyter Notebook
- Word Document with report detailing rationale, including visualisations  
“YourName\_AI\_DV\_DP\_ICA.docx”
- 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.
- 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 requested by contacting the Lecturer, 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](#).
- 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](#).