

### CCT College Dublin Continuous Assessment

<b>Programme Title:</b>	HDip in Data Analytics for Business		
<b>Cohort:</b>	FT		
<b>Module Title(s):</b>	Machine Learning for Business		
<b>Assignment Type:</b>	Individual	<b>Weighting(s):</b>	50%
<b>Assignment Title:</b>	CA1		
<b>Lecturer(s):</b>	Dr. Muhammad Iqbal		
<b>Issue Date:</b>	11 <sup>th</sup> September 2024		
<b>Submission Deadline Date:</b>	27 <sup>th</sup> October 2024		
<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% of the mark awarded</b> . 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>	Upload one or multiple files as word file, jupyter notebook, dataset and any supporting information. No pdfs are acceptable.		
<b>Feedback Method:</b>	<b>Results posted in Moodle gradebook</b>		
<b>Feedback Date:</b>	Three weeks after submission		

#### 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:

1. Critically evaluate and implement appropriate clustering algorithms and interpret and document their results. (Linked to PLO 1, PLO 5)
2. Apply modelling to time series data to facilitate business intelligence needs (Linked to PLO 1, PLO 2, PLO 3)

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

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 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.

## Assessment Task

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

This is an individual project using the PYTHON programming language. Develop and deploy machine learning models in any one of the following areas only and analyse the results.

- Agriculture, Fisheries and Forestry and Food
- Housing and Zoning
- Stock market dataset from only website: <https://finance.yahoo.com/quote/CSV/history/>

The dataset should have a decent number of rows and columns (for example, type of variables may be categorical, continuous, and discrete) after cleaning to justify the use of the machine learning modelling approach. The type of question(s) that you should formulate for the project will depend on the chosen domain of the dataset as mentioned above.

Project questions could be: (this is a small, suggested, sample of questions, other questions may be more appropriate to your project)

- How to measure similarity or dissimilarity between different clusters?
- Which clustering solution do you prefer, and why?
- How to analyse and investigate an inflation rate for a specific product from the provided time series dataset?

You will present their findings and defend the results in the report (MS Doc). Your report should capture the following aspects that are relevant to your project investigations. All implementations are carried in Jupyter notebook and no Python files will be accepted.

- i) A precise introduction, motivation, description of problem domain, project objectives and the rationale for the chosen dataset in the above-mentioned areas.  
(15 marks)
- ii) Which clustering algorithms would you consider for segmentation, and why? Implement the solutions and discuss the differences between silhouette score and Davies-Bouldin index in the context of clustering. Compare the results obtained from any two clustering algorithms from the chosen dataset.  
(30 marks)
- iii) Implement your solutions and discuss what insights can you derive from the initial exploration of the time series data based on the provided topics? Describe any trends, seasonality, or anomalies observed. How did you determine the appropriate parameters (p, d, q) for the ARIMA model. Evaluate the performance of the ARIMA model in forecasting future values, highlighting any strengths and limitations based on your chosen dataset.  
(30 marks)
- iv) Interpret and justify the results based on the problem specification or project objectives by using suitable visualizations. Comments and description of Python code, conclusions of the project should be specified in the report as well as Jupyter notebook. Citations and references should be in the Harvard Style. Record a video to explain and justify your report and the code in the Jupyter notebook and the duration of the video should not be more than 7 minutes.  
(25 marks)

**Note:** You can choose two different datasets for task (ii) or task (iii) separately or one dataset for both tasks

## Submission Requirements

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

- All files (MS word, Jupyter notebook, data files and any supporting document) should be uploaded separately on Moodle.
- Clearly detail the number of words used in the report.
- Number of Words in the report (1250 words +/-10%) excluding diagrams, code, references, citations and titles.
- You must use the Github classroom link provided (<https://classroom.github.com/a/gGcqo8Qz>) for version control. You should have at least 5 commits on Github before submission.
- The rubric is provided for the detailed breakdown of marks at the end of this CA1.
- No AI agent voice is allowed to use in the video recording.
- Use [Harvard Referencing](#) when citing third party material
- Be the student's own work.
- Include the CCT assessment cover page.
- Be submitted by the deadline date specified or be subject to late submission penalties
- Must be clearly specified the number of words used after each section in the report.

## Acceptable Use of AI for Assignment at CCT

Acceptable and Unacceptable Use of AI	<ul style="list-style-type: none"><li>• The use of generative AI tools (e.g. ChatGPT, Dall-e, etc.) is permitted in this assignment for the following activities:<ul style="list-style-type: none"><li>◦ Brainstorming and refining your ideas;</li><li>◦ Fine tuning your research questions;</li><li>◦ Finding information on your topic;</li><li>◦ Drafting an outline to organise your thoughts; and</li><li>◦ Checking grammar and style.</li></ul></li><li>• The use of generative AI tools is not permitted in this course for the following activities:<ul style="list-style-type: none"><li>◦ Impersonating you in classroom context</li><li>◦ Completing group work that your group has assigned to you</li><li>◦ Writing a draft of a writing assignment</li><li>◦ Writing entire sentences, paragraphs or papers to complete class assignments.</li></ul></li><li>• You are responsible for the information you submit based on an AI query. Your use of AI tools must be properly documented and cited.</li><li>• 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.</li><li>• When in doubt about permitted usage, please ask for clarification.</li></ul>	<p>This statement is useful when you are allowing the use of AI tools for certain purposes, but not for others. Adjust this statement to reflect your particular parameters of acceptable use, and your discipline context.</p>
---------------------------------------	---	---

**GRADING RUBRIC – Machine Learning for Business - 2024**

GRADE	90-100%	80-90%	70-79%	60-69%	50-59%	40-49%	35-39%	<35%
Performance	Exceptional	Outstanding	Excellent	Very Good	Good	Acceptable	Fail	Fail
<b>Introduction and Project Objectives (15 marks)</b>	The introduction is exceptionally precise and provides a thorough motivation, a clear description of the problem domain, well-defined project objectives, and a compelling rationale for the chosen dataset.	The introduction is outstandingly precise and provides a nice motivation, a clear description of the problem domain, well-defined project objectives, and a logical rationale for the chosen dataset.	The introduction is excellent and precise and provides an excellent motivation, a clear description of the problem domain, well-defined project objectives, and a logical rationale for the chosen dataset.	The introduction is very good, but there may be some gaps or require more details in certain aspects.	The introduction is good, but it may lack precision, depth, or clarity in some key areas.	The introduction is acceptable, providing a basic understanding but with notable gaps or lack of clarity.	The introduction contains major deficiencies, such as a lack of clarity on project objectives or a weak rationale for the chosen dataset.	The introduction is insufficient or completely lacks key elements.
<b>Clustering Algorithms and Comparison (30 marks)</b>	Implement and provide an insightful explanation of clustering algorithm choices, a nice comparison of silhouette score and Davies-Bouldin index, and a detailed comparison of results from two clustering algorithms with strong justification.	Implement and provide an outstanding explanation of clustering algorithm choices, a good comparison of silhouette score and Davies-Bouldin index, and a solid comparison of results from two clustering algorithms.	Implement and provide an excellent explanation of clustering algorithm choices, a reasonable comparison of silhouette score and Davies-Bouldin index, and a competent comparison of results.	Implement and provide a very good explanation of clustering algorithm choices, a basic comparison of silhouette score and Davies-Bouldin index, and a satisfactory comparison of results.	Implement and provide a good explanation of clustering algorithm choices, but the comparison of silhouette score and Davies-Bouldin index may lack depth or clarity, and the comparison of results may be limited.	The explanation of clustering algorithm choices is acceptable, but the comparison of silhouette score and Davies-Bouldin index, as well as the comparison of results, may be weak or unclear.	Major deficiencies in the explanation of clustering algorithm choices, lack of clarity in comparing silhouette score and Davies-Bouldin index, and significant issues in the comparison of results.	Serious inadequacies in explaining clustering algorithm choices, unclear or incorrect comparison of silhouette score and Davies-Bouldin index, and a lack of meaningful comparison of results.
<b>Time Series Data Exploration and ARIMA Model (30 marks)</b>	Implement and provide a highly insightful exploration of time series data, a clear determination of ARIMA parameters, and a comprehensive evaluation of the ARIMA model's performance, highlighting both strengths and limitations.	Implement and provide an outstanding exploration of time series data, a sound determination of ARIMA parameters, and a strong evaluation of the ARIMA model's performance.	Implement and provide an excellent exploration of time series data, an excellent determination of ARIMA parameters, and a competent evaluation of the ARIMA model's performance.	Implement and provide a very good exploration of time series data, a basic determination of ARIMA parameters, and a very good evaluation of the ARIMA model's performance.	Implement and provide a good exploration of time series data, but the determination of ARIMA parameters and the evaluation of the ARIMA model requires a depth or clarity.	The exploration of time series data is acceptable, but the determination of ARIMA parameters and the evaluation of the ARIMA model may be weak or unclear.	Major deficiencies in the exploration of time series data, lack of clarity in determining ARIMA parameters, and significant issues in the evaluation of the ARIMA model.	Serious inadequacies in exploring time series data, unclear or incorrect determination of ARIMA parameters, and a lack of meaningful evaluation of the ARIMA model.
<b>Results Interpretation, Project Documentation and video recording (25 marks)</b>	Demonstrates exceptional interpretation and justification of results, with thorough and insightful conclusions. Python code is meticulously documented, ensuring clarity and reproducibility. There is full adherence to the Harvard Style for citations and references. The video recording is of an exceptional standard, completely aligned with the report and Jupyter notebook, displaying exceptional technical and presentation skills.	Demonstrate outstanding interpretation and justification of results, with strong, well-reasoned conclusions. Python code is well-documented, though with minor gaps. There is strong adherence to the Harvard Style for citations and references. The video recording is of an outstanding quality, well-supported by the report and Jupyter notebook, with only minor areas for improvement.	Demonstrates an excellent interpretation of results, with reasonable and well-justified conclusions. Python code documentation is clear and understandable. There is consistent adherence to the Harvard Style for citations and references. The video recording is of an excellent standard, effectively complementing the report and Jupyter notebook.	Demonstrates a very good interpretation of results, with satisfactory justifications and conclusions. Python code documentation is present but lacks depth in certain areas. There is an attempt to follow the Harvard Style for citations and references. The video recording is of a very good standard, supporting the report and Jupyter notebook, though some improvements are needed.	Demonstrates a good interpretation of results, though some clarity or depth may be lacking. Python code documentation is present but limited. The conclusions are basic, and there is some adherence to the Harvard Style for citations and references. The video recording is of a good standard, though inconsistencies with the report and Jupyter notebook are evident..	Demonstrates an acceptable interpretation of results, but with noticeable weaknesses or lack of clarity. Python code documentation is limited, and conclusions may be weak or underdeveloped. Adherence to the Harvard Style for citations and references is inconsistent. The video recording is of an acceptable quality, but significant improvements are required to align it with the report and Jupyter notebook.	Demonstrates major deficiencies in interpreting and justifying results, with a lack of clarity or accuracy. Python code documentation is poor, and conclusions are unclear or insufficient. Significant issues with adherence to the Harvard Style for citations and references. The video recording is well below the standard, failing to adequately support the report and Jupyter notebook.	Shows serious inadequacies in interpreting and justifying results, with minimal or no meaningful Python code documentation. Conclusions are incorrect or entirely unclear, with little to no adherence to the Harvard Style for citations and references. The video recording is of an unacceptable standard, demonstrating a lack of understanding or effort.

**CCT College Dublin**

Assessment Cover Page

To be provided separately as a word doc for students to include with every submission

Module Title:	
Assessment Title:	
Lecturer Name:	
Student Full Name:	
Student Number:	
Assessment Due Date:	
Date of Submission:	

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.