CCT College Dublin Continuous Assessment

Programme Title:	HDIP in Sci in Data Analytics for Business		
Delivery Mode:	FT/ PT		
Cohort Details:	*MLBus_HDipData_Feb25_FT, Semester 2*		
Module Title(s):	*Machine Learning for Business* *Programme schedules are all published on the CCT IQR Provider Profile*		
Assignment Type:	*Individual*	**Weighting(s):**	*50%*
Assignment Title:	*CA1*		
Lecturer(s):	*Dr. Muhammad Iqbal*		
Issue Date:	*Wednesday 10rd September 2025*		
Submission Deadline Date:	*Friday 31st October 2025@11:59*		
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:	*Ensure that all required files, including Word documents, Jupyter notebooks, datasets, and any supporting materials, are uploaded as separate files. Multiple files can be submitted if necessary. However, PDFs are not accepted.*		
Feedback Method:	**Results posted in Moodle gradebook**		
Feedback Date:	*Three weeks after submission, Approximately Monday 24th November 2025*		

Assessment Outline

Github Link for CA1 (50%):

- https://classroom.github.com/a/GkZLmocH

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.

- Housing and zoning
- Population and Society
- Stock market dataset (only from last 10 years) 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. (**Note:** No marks will be awarded if the dataset is not based on the above-mentioned guidelines.)

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 practical implementations are carried in the Jupyter notebook (**Note:** The code in the form of Python files will not 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 in the jupyter notebook. Citations and references should follow

the Harvard style. You are required to record a video presentation to explain and justify your report and the code in the Jupyter Notebook. The video should not exceed 7 minutes in duration, and the use of AI-generated voices or agents is not permitted.

(25 marks)

Note: You can choose ***two different datasets*** for task (ii) or task (iii) separately or one dataset for both tasks based on your justification for this project.

Assessment Requirements

All assessment submissions must meet the following minimum requirements:

- All files (MS word, Jupyter notebook, data files and any supporting document) should be uploaded separately on Moodle.
- Students are strictly prohibited from generating synthetic data using any GenAI tool (chatGPT, etc.). All datasets utilized in CA1 must be authentic, real-world and accompanied by a valid source, provided either through a web link or a shared Google drive link.
- 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/GkZLmocH) 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.

Learning Outcomes:

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

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)

Use of Artificial Intelligence

Artificial Intelligence (AI) can be used in this assignment in the following way(s):

- Brainstorming and refining your ideas;
- Fine tuning your research questions;
- Finding information on your topic;
- Drafting an outline to organise your thoughts;
- Checking grammar and style
- Other

Note for students:

- You are required to include all AI prompt(s) and outputs in an appendix, including a short reflection (approx. 250 words) on how AI enhanced your work, or otherwise.
- 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.
- Cite all AI references used in your assignment.
- You must declare how you used AI on your assignment cover sheet.
- In the case of any suspicious work submitted by a student, the student will be called for a Q&A session.
- When in doubt about permitted usage, please ask for clarification to your lecturer.

Grading Criteria

Criteria	Criteria 1: Introduction & Project Objectives (15 marks – 20%)	Criteria 2: Clustering Algorithms & Comparison (30 marks – 30%)	Criteria 3: Time Series Data Exploration & ARIMA Model (30 marks – 25%)	Criteria 4: Results Interpretation, Documentation & Video (25 marks – 25%)
Excellent (+70%)	Provides a precise introduction with strong motivation, clear description of the problem domain, well-defined project objectives, and a compelling rationale for the chosen dataset.	Clearly justifies choice of clustering algorithms for segmentation; implements solutions effectively; provides insightful discussion of silhouette score vs. Davies-Bouldin index; offers a detailed and well-justified comparison of results from two clustering algorithms.	Highly insightful initial exploration of time series with clear identification of trends, seasonality, and anomalies; accurate determination of ARIMA parameters (p, d, q); comprehensive evaluation of forecasting performance with strengths and limitations well highlighted.	Results are interpreted and justified with strong alignment to objectives; visualizations effectively support conclusions; Python code is meticulously commented and documented; report and notebook include clear conclusions; full Harvard-style referencing; video is professional, aligned, and within the required time and format.
Very Good (60–69%)	Very good introduction with clear motivation and objectives, but minor gaps in detail or rationale.	Very good explanation of algorithm choices; implementations correct; comparison of silhouette vs. Davies-Bouldin is sound but could be more detailed; comparison of results between two algorithms is satisfactory.	Very good exploration of time series with identification of key patterns; reasonable determination of ARIMA parameters; strong but requires exhaustive evaluation of model performance.	Results interpretation is strong with good visualizations; Python code documentation is clear and can be improved; conclusions are sound; good attempt at Harvard-style referencing; video presentation is clear but may need minor improvements.
Good (50–59%)	Good introduction, and a more depth in precision or clarity in motivation, objectives, or rationale is required.	Good explanation of algorithm choices, but justification is limited; implementations are correct but analysis of silhouette vs. Davies-Bouldin requires more depth; results comparison is basic, and more insight is required.	Good exploration of time series, but identification of trends/seasonality/anomalies is limited; ARIMA parameters chosen but justification weak; evaluation of forecasting performance is limited.	Interpretation is adequate but conclusions are basic; visualizations present but required improvements; Python code documentation is limited; some referencing issues; video is good but inconsistencies with report/notebook exist.
Acceptable (40–49%)	Acceptable introduction giving only a basic idea of the problem, objectives, or dataset rationale; lacks details or clarity.	Acceptable explanation of algorithm choices; limited or unclear discussion of clustering metrics; weak comparison of results between algorithms.	Acceptable exploration of time series with minimal discussion of trends or anomalies; weak or basic ARIMA parameter selection discussed; More depth in the evaluation of forecasting performance is required.	Interpretation is weak or unclear; limited visualizations; code documentation minimal; conclusions underdeveloped; referencing inconsistent; video acceptable but with major issues in quality or alignment.
Fail (<39%)	Introduction has major deficiencies: unclear objectives, missing rationale, or insufficient description of dataset/problem.	Serious deficiencies in algorithm explanation; incorrect or unclear discussion of silhouette vs. Davies-Bouldin; no meaningful comparison of results.	Poor or missing exploration of time series; incorrect or unclear ARIMA parameter selection; little or no evaluation of forecasting performance.	Interpretation of results are missed or incorrect; code poorly documented or absent; no meaningful conclusions; no or very poor referencing; video missing or unacceptable.

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.

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
- For additional support with IT subject content, students are advised to access the CCT Support Hub.