INTERNATIONAL QUALIFICATIONS

AND ASSESSMENT CENTRE (IQAC)

«Approved»

2025 y. «16» 06

Programme	Level 6 Diploma in Data Science		
Unit Number/ Unit Title	UNIT 6 CAPSTONE PROJECT: APPLIED DATA SCIENCE SOLUTIONS		
Cohort Code:	L06CPDSS-U6		
Unit Level	Level 6		
Total GLH	Total qualification time 200/ Total Guided learning hours 90/ Self-guided learning hours 110		
Credits	10 CATS/5 ECTS		
Lecturer			
Start Date		**End Date**	

Unit Aims	This unit enables students to design, implement, and present a full-scale data science project addressing a real-world problem. It integrates technical, analytical, ethical, and communication skills developed across all modules.
Differentiation Strategies *(e.g. planned activities or support for individual learners according to their needs)*	The total number of students to be in the lesson is approximately 20. This is a multicultural group of students predominantly between the ages of 24 – 45, with numerous ethnic, gender, and creed background. These are UK academic level 5 students; hence it is assumed that they have practical, theoretical, or technological knowledge and understanding of a subject or field of work to find ways forward in broadly defined, complex contexts. These students must be able to generate information, evaluate, synthesise the use information from a variety of sources. Various approaches to addressing the various identified students needs will be adopted throughout the lesson. Such will include:- 1. Progressive tasks 2. Digital resources 3. Verbal support 4. Variable outcomes

	5. Collaborative learning 6. Ongoing assessment 7. Flexible-pace learning
Equality & Diversity	Variety of teaching techniques will be employed to ensure that the needs of each individual learner are met.
Safeguarding & Prevent	Safeguarding policies and the Prevent duty are strictly observed to ensure the safety, well-being, and inclusivity of all students and staff.
Health & Safety	SIRM H&S policies will be maintained.
Learning Resources	**Teaching and Learning Materials** • Wickham, H., & Grolemund, G. (2017). R for Data Science. O'Reilly Media. • Provost, F., & Fawcett, T. (2013). Data Science for Business. O'Reilly Media. • UKRI Good Research Conduct Guidelines. • Examples from Kaggle and real-world data challenges.

Learning Outcome	Assessment Criteria
LO1. 1. Identify and define a real-world data problem.	Project Proposal: 1.1 Develop a research question and objectives. 1.2 Conduct feasibility analysis.
LO2. 2. Design and execute a data science pipeline.	Final Report: 2.1 Collect and clean relevant data. 2.2 Apply analytical and ML models. 2.3 Interpret and visualise results.
LO3. 3. Demonstrate ethical and governance awareness.	Ethics Section: 3.1 Ensure GDPR compliance. 3.2 Address potential biases.
LO4. 4. Communicate findings to technical and non-technical audiences.	Presentation: 4.1 Deliver a professional oral presentation. 4.2 Prepare a technical poster or slide deck.

No	Learning Outcome / Topic	Learning and Teaching Activities	Which assessment criteria does the session relate to?	Day/month/year/ signature
1.	**Selecting a Real-World Problem**	**Selecting a Real-World Problem** Industry domains (healthcare, finance, climate, etc.)	LO1: Problem Identification & Feasibility	
2.	**Stakeholder Analysis**	**Stakeholder Analysis** Identifying end-users, beneficiaries, and constraints	LO1: Problem Identification & Feasibility	
3.	**Research Question Formulation**	**Research Question Formulation** SMART criteria (Specific, Measurable, Achievable, Relevant, Time-bound)	LO1: Problem Identification & Feasibility	
4.	**Feasibility Assessment**	**Feasibility Assessment** Data availability, technical resources, ethical considerations	LO1: Problem Identification & Feasibility	
5.	**Project Proposal Writing**	**Project Proposal Writing** Structure: Problem statement, objectives, methodology	LO1: Problem Identification & Feasibility	
6.	**Data Collection Strategies**	**Data Collection Strategies** APIs (REST, GraphQL), web scraping, public datasets (Kaggle, UN Data)	LO2: Data Science Pipeline Design	
7.	**Data Cleaning & Preprocessing**	**Data Cleaning & Preprocessing** Handling missing values, outliers, normalization	LO2: Data Science Pipeline Design	
8.	Half-Term Exam	- Review of LO1 topics - Practice questions and mock assessment - **Half-term assessment** based on LO1 (theory)	LO1 LO2	

9.	**Exploratory Data Analysis (EDA)**	**Exploratory Data Analysis (EDA)** Statistical summaries, correlation analysis, visualizations (Matplotlib, Seaborn)	LO2: Data Science Pipeline Design	
10.	**Feature Engineering**	**Feature Engineering** Dimensionality reduction (PCA), feature scaling, encoding	LO2: Data Science Pipeline Design	
11.	**Model Selection & Justification**	**Model Selection & Justification** Supervised vs. unsupervised, baseline models	LO2: Data Science Pipeline Design	
12.	**Machine Learning Implementation**	**Machine Learning Implementation** Regression, classification, clustering (scikit-learn, TensorFlow)	LO3: Model Development & Evaluation	
13.	**Hyperparameter Tuning** Grid search, random search, Bayesian optimization	**Hyperparameter Tuning** Grid search, random search, Bayesian optimization	LO3: Model Development & Evaluation	
14.	Final Exam Preparation & Review	- Comprehensive review of all learning outcomes - Practice questions and revision of key topics		
15.	Final Exam	- **Final-term assessment** covering all learning outcomes (theory and practical elements)		
16.	Feedback & Reflection	- Review of final exam - Individual feedback on performance - Reflective discussion on key learning points		
17.	**Model Evaluation Metrics**	**Model Evaluation Metrics** Accuracy, precision, recall, F1, RMSE, AUC-ROC	LO3: Model Development & Evaluation	
18.	**Bias & Fairness Auditing**	**Bias & Fairness Auditing** Fairlearn, AIF360, SHAP for interpretability	LO3: Model Development & Evaluation	
19.	**Ethical Compliance Check**	**Ethical Compliance Check** GDPR, HIPAA, anonymization techniques	LO3: Model Development & Evaluation	

20.	**Model Deployment Options**	**Model Deployment Options** Flask/Django API, Docker containers, cloud (AWS SageMaker, GCP AI Platform)	LO4: Deployment & Scalability	
21.	**Scalability Considerations**	**Scalability Considerations** Batch vs. real-time processing, load testing	LO4: Deployment & Scalability	
22.	**Monitoring & Maintenance**	**Monitoring & Maintenance** Drift detection, model retraining pipelines	LO4: Deployment & Scalability	
23.	Half-Term Exam	**Technical Report Writing** Structure: Abstract, methodology, results, conclusion	LO5: Communication & Presentation	
24.	**Data Visualization Best Practices**	**Data Visualization Best Practices** Interactive dashboards (Plotly, Tableau), storytelling techniques	LO5: Communication & Presentation	
25.	**Oral Presentation Skills**	**Oral Presentation Skills** Audience adaptation, handling Q&A, time management	LO5: Communication & Presentation	
26.	**Poster/Slide Deck Design**	**Poster/Slide Deck Design** Key elements: Problem, solution, impact, visuals	LO5: Communication & Presentation	
27.	**Peer Review & Feedback**	**Peer Review & Feedback** Iterative improvements based on critiques	LO5: Communication & Presentation	
28.	**Final Submission & Demo**	**Final Submission & Demo** Code repository (GitHub), live demo (if applicable)	LO5: Communication & Presentation	
29.	Final Exam Preparation & Review	LO1, LO2, LO3, LO4	LO1, LO2, LO3, LO4	
30.	Final Exam		LO1, LO2, LO3, LO4	