INTERNATIONAL QUALIFICATIONS

AND ASSESSMENT CENTRE (IQAC)

«Approved»

2025 y. «16» 12

Programme	LEVEL 5 EXTENDED DIPLOMA IN DATA SCIENCE		
Unit Number/ Unit Title	UNIT 11 DATA SCIENCE CAPSTONE PROJECT		
Cohort Code:	L05DSCP-U11		
Unit Level	Level 5		
Total GLH	Total qualification time 200/ Total Guided learning hours 90/ Self-guided learning hours 110		
Credits	20 CATS/ 10 ECTS		
Lecturer			
Start Date		**End Date**	

Unit Aims	This module aims to integrate and apply knowledge and skills acquired throughout the programme to a substantial data science project.
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** • O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Penguin Books. • Wickham, H., & Grolemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media. • VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.

Learning Outcome		Assessment Criteria
LO1.	**1. Plan and scope a data science project.**	1.1 Define project objectives and scope based on stakeholder requirements. 1.2 Develop a project plan outlining tasks, timelines, and resources.
LO2.	**2. Execute data analysis and interpretation.**	2.1 Collect, clean, and preprocess data for analysis. 2.2 Apply appropriate data science techniques and algorithms to analyze data.
LO3.	**3. Communicate project findings effectively.**	3.1 Present project findings and insights to stakeholders using clear and persuasive communication . 3.2 Reflect on the project outcomes and propose recommendations for future work.

No	Learning Outcome / Topic	Learning and Teaching Activities	Which assessment criteria does the session relate to?	Day/month/year/ signature
1.	**Capstone Introduction & Expectations**	**Capstone Introduction & Expectations** Overview, learning outcomes, and success criteria	LO1: Project Planning & Scoping	
2.	**Identifying Stakeholders & Requirements**	**Identifying Stakeholders & Requirements** Business needs, problem definition, success metrics	LO1: Project Planning & Scoping	
3.	**Data Science Project Lifecycle**	**Data Science Project Lifecycle** CRISP-DM, Agile for data science	LO1: Project Planning & Scoping	
4.	**Project Proposal Development**	**Project Proposal Development** Problem statement, objectives, scope, constraints	LO1: Project Planning & Scoping	
5.	**Project Planning & Timeline**	**Project Planning & Timeline** Gantt charts, milestones, risk assessment	LO1: Project Planning & Scoping	
6.	**Data Sourcing Strategies**	**Data Sourcing Strategies** APIs, web scraping, public datasets, partnerships	LO2: Data Collection & Preprocessing	
7.	**Data Cleaning & Validation**	**Data Cleaning & Validation** Handling missing values, outliers, inconsistencies	LO2: Data Collection & Preprocessing	
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)** Summary statistics, visualizations, hypothesis generation	LO2: Data Collection & Preprocessing	
10.	**Feature Engineering**	**Feature Engineering** Transformations, encoding, dimensionality reduction	LO2: Data Collection & Preprocessing	

11.	**Data Storage & Management**	**Data Storage & Management** Databases (SQL/NoSQL), cloud storage (AWS S3, Google Cloud)	LO2: Data Collection & Preprocessing	
12.	**Algorithm Selection**	**Algorithm Selection** Supervised vs. unsupervised, regression vs. classification	LO3: Model Development & Analysis	
13.	**Baseline Model Development**	**Baseline Model Development** Simple models (linear regression, decision trees)	LO3: Model Development & Analysis	
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.	**Advanced Model Implementation**	**Advanced Model Implementation** Ensemble methods, neural networks, time-series models	LO3: Model Development & Analysis	
18.	**Model Evaluation & Validation**	**Model Evaluation & Validation** Cross-validation, metrics (RMSE, AUC-ROC, F1-score)	LO3: Model Development & Analysis	
19.	**Hyperparameter Tuning**	**Hyperparameter Tuning** Grid search, random search, Bayesian optimization	LO3: Model Development & Analysis	
20.	**Results Interpretation**	**Results Interpretation** Feature importance, SHAP values, model explainability	LO4: Interpretation & Communication	

21.	**Visual Storytelling**	**Visual Storytelling** Dashboards (Tableau, Power BI), interactive plots (Plotly)	LO4: Interpretation & Communication	
22.	**Report Writing**	**Report Writing** Structure (executive summary, methodology, findings)	LO4: Interpretation & Communication	
23.	Half-Term Exam	**Final Presentation & Demo** Live demo, Q&A, peer feedback		
24.	**Stakeholder Presentation Skills**	**Stakeholder Presentation Skills** Tailoring messages for technical vs. non-technical audiences	LO4: Interpretation & Communication	
25.	**Ethical Considerations & Bias Mitigation**	**Ethical Considerations & Bias Mitigation** Fairness audits, privacy compliance	LO4: Interpretation & Communication	
26.	**Model Deployment Basics**	**Model Deployment Basics** Flask/Django APIs, Docker, cloud deployment (AWS SageMaker)	LO5: Deployment & Future Work	
27.	**Monitoring & Maintenance** **Lessons Learned & Reflection**	**Monitoring & Maintenance** Drift detection, retraining strategies **Lessons Learned & Reflection** Challenges, successes, and improvements	LO5: Deployment & Future Work	
28.	**Future Recommendations**	**Future Recommendations** Scalability, additional data sources, model enhancements **Final Presentation & Demo** Live demo, Q&A, peer feedback	LO5: Deployment & Future Work	
29.	Final Exam Preparation & Review	LO1, LO2, LO3, LO4	LO1, LO2, LO3, LO4	
30.	Final Exam		LO1, LO2, LO3, LO4	