INTERNATIONAL QUALIFICATIONS

AND ASSESSMENT CENTRE (IQAC)

«Approved»

2025 y. «16» июл

Programme	**LEVEL 4 EXTENDED DIPLOMA IN ARTIFICIAL INTELLIGENCE**
Unit Number/ Unit Title	UNIT 6 INTRODUCTION TO DATA SCIENCE
Cohort Code:	L04IDS-U6
Unit Level	4
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	The unit aims to provide students with a fundamental understanding of the data science process, including the key steps involved in transforming raw data into meaningful insights. By the end of the unit, students will be equipped with the necessary skills and knowledge to tackle real-world data challenges, leveraging a variety of data analysis tools and techniques.
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** • "Data Science for Business" by Foster Provost and Tom Fawcett • "Python Data Science Handbook" by Jake VanderPlas • "Doing Data Science" by Cathy O'Neil and Rachel Schutt.

Learning Outcome	Assessment Criteria
LO1. Understand the data science process.	1.1 Describe the stages of the data science lifecycle. 1.2 Explain the importance of data cleaning and pre-processing.
LO2. Collect and analyze data.	2.1 Perform data collection using various methods and tools. 2.2 Conduct exploratory data analysis to identify patterns and insights.
LO3. Visualize data effectively.	3.1 Create visualizations using libraries such as Matplotlib and Seaborn. 3.2 Interpret and present data findings through visualizations.
LO4. Apply data science techniques to real-world problems	4.1 Use data science methods to solve business problems. 4.2 Develop data-driven decision-making skills.

No	Learning Outcome / Topic	Learning and Teaching Activities	Which assessment criteria does the session relate to?	Day/month/year/ signature
1.	**Introduction to Data Science**	**Introduction to Data Science** Definition, importance, and applications across industries.	LO1: Understand the Data Science Process	
2.	**Data Science Lifecycle (Stages)**	**Data Science Lifecycle (Stages)** Problem definition, data collection, cleaning, exploration, modeling, deployment.	LO1: Understand the Data Science Process	
3.	**Role of a Data Scientist**	**Role of a Data Scientist** Skills, tools, and responsibilities in the industry.	LO1: Understand the Data Science Process	
4.	**Data Sources & Types**	**Data Sources & Types** Structured vs. unstructured data (CSV, SQL, NoSQL, APIs, web scraping).	LO1: Understand the Data Science Process	
5.	**Data Cleaning & Preprocessing**	**Data Cleaning & Preprocessing** Handling missing values, outliers, duplicates, and normalization.	LO1: Understand the Data Science Process	
6.	**Data Collection Methods**	**Data Collection Methods** Surveys, sensors, web scraping, APIs, public datasets (Kaggle, UCI).	LO2: Collect and Analyze Data	
7.	**Data Storage & Management**	**Data Storage & Management** Databases (SQL vs. NoSQL), data lakes, cloud storage (AWS S3, Google BigDrive).	LO2: Collect and Analyze Data	
8.	Half-Term Exam	- Review of LO1 LO2 topics - Practice questions and mock assessment - **Half-term assessment** based on LO1 (theory)		
9.	**Introduction to Exploratory Data Analysis (EDA)**	**Introduction to Exploratory Data Analysis (EDA)** Objectives and key techniques (summary stats, distributions).	LO2: Collect and Analyze Data	

10.	**Descriptive Statistics for EDA**	**Descriptive Statistics for EDA** Mean, median, variance, skewness, kurtosis.	LO2: Collect and Analyze Data	
11.	**Data Correlation & Hypothesis Testing**	**Data Correlation & Hypothesis Testing** Pearson, Spearman correlation, p-values.	LO2: Collect and Analyze Data	
12.	**Handling Categorical & Numerical Data**	**Handling Categorical & Numerical Data** Encoding (one-hot, label encoding), binning, scaling.	LO2: Collect and Analyze Data	
13.	**Feature Engineering Basics**	**Feature Engineering Basics** Creating new features, datetime handling, text preprocessing.	LO2: Collect and Analyze Data	
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)	LO1 LO2	
16.	Feedback & Reflection	- Review of final exam - Individual feedback on performance - Reflective discussion on key learning points		
17.	**Principles of Data Visualization**	**Principles of Data Visualization** Importance of visuals, choosing the right chart types.	LO3: Visualize Data Effectively	
18.	**Matplotlib for Basic Plots**	**Matplotlib for Basic Plots** Line plots, bar charts, histograms, box plots.	LO3: Visualize Data Effectively	
19.	**Advanced Visualizations with Seaborn**	**Advanced Visualizations with Seaborn** Heatmaps, pair plots, violin plots, FacetGrid.	LO3: Visualize Data Effectively	
20.	**Interactive Visualizations (Plotly/Dash)**	**Interactive Visualizations (Plotly/Dash)** Dynamic charts for web-based dashboards.	LO3: Visualize Data Effectively	
21.	**Geospatial Data Visualization**	**Geospatial Data Visualization** Folium/Geopandas for maps and location-based insights.	LO3: Visualize Data Effectively	
22.	**Storytelling with Data**	**Storytelling with Data** Designing dashboards (Tableau/Power BI), presenting insights.	LO3: Visualize Data Effectively	

23.	Half-Term Exam	**Case Study: Business Problem-Solving** Framing problems (e.g., customer churn, sales forecasting).		
24.	**Predictive Modeling Overview**	**Predictive Modeling Overview** Regression/classification applications in business.	LO4: Apply Data Science Techniques to Real-World Problems	
25.	**A/B Testing & Experimentation**	**A/B Testing & Experimentation** Designing tests, interpreting results for decision-making.	LO4: Apply Data Science Techniques to Real-World Problems	
26.	**Ethics in Data Science**	**Ethics in Data Science** Bias, privacy (GDPR), fairness in AI/ML models.	LO4: Apply Data Science Techniques to Real-World Problems	
27.	**Data-Driven Decision Making**	**Data-Driven Decision Making** From insights to actionable strategies (e.g., marketing, operations).	LO4: Apply Data Science Techniques to Real-World Problems	
28.	**End-to-End Data Science Project**	**End-to-End Data Science Project** Walkthrough: From dataset to deployment (e.g., Flask/Streamlit).	LO4: Apply Data Science Techniques to Real-World Problems	
29.	Final Exam Preparation & Review	LO1, LO2, LO3, LO4	LO1, LO2, LO3, LO4	
30.	Final Exam		LO1, LO2, LO3, LO4	