Title of Module: Data Mining and Visualisation			
Code: COMP11069	**SCQF Level: 11** (Scottish Credit and Qualifications Framework)	**Credit Points: 20**	**ECTS: 10** (European Credit Transfer Scheme)
School:	School of Computing, Engineering and Physical Sciences		
Module Co-ordinator:	Naeem Ramzan		

Summary of Module
This module covers the main analytical skills and tools a data scientist/analyst requires to extract useful information from data. It will build on the foundations on how to properly communicate this information to the relevant audience. Students will be progressively guided through the world of data analytics starting from an introduction to quantitative research methods and related techniques. The journey will take them to the world of machine learning and pattern recognition by presenting the key ideas in this area enabling informed decisions when applying these concepts to real-world problems. Additional key advanced concepts and research trends in the field will also be presented as well as the basic needed principles to communicate clearly and effectively the patterns found on the data. A special attention will thus be paid to data visualisation so that the students are able to answer research questions, inform their decisions, and convey results with data supported evidence. This module has been specifically designed considering the UWS Graduate Attributes of Universal, Work ready, and Successful. Details to these attributes is available at UWS Graduate Attributes webpage.

Module Delivery Method					
Face-To-Face	**Blended**	**Fully Online**	**HybridC**	**HybridO**	**Work-based Learning**
	✓				

Face-To-Face
Term used to describe the traditional classroom environment where the students and the lecturer meet synchronously in the same room for the whole provision.
Blended
A mode of delivery of a module or a programme that involves online and face-to-face delivery of learning, teaching and assessment activities, student support and feedback. A programme may be considered "blended" if it includes a combination of face-to-face, online and blended modules. If an online programme has any compulsory face-to-face and campus elements it must be described as blended with clearly articulated delivery information to manage student expectations
Fully Online
Instruction that is solely delivered by web-based or internet-based technologies. This term is used to describe the previously used terms distance learning and e learning.
HybridC
Online with mandatory face-to-face learning on Campus
HybridO
Online with optional face-to-face learning on Campus
Work-based Learning
Learning activities where the main location for the learning experience is in the workplace.

Campus(es) for Module Delivery
The module will **normally** be offered on the following campuses / or by Distance/Online Learning: (Provided viable student numbers permit)

Paisley:	Ayr:	Dumfries:	Lanarkshire:	London:	Distance/Online Learning:	Other:
✓						

Term(s) for Module Delivery

(Provided viable student numbers permit).

Term 1		Term 2	✓	Term 3	

Learning Outcomes: (maximum of 5 statements)

On successful completion of this module the student will be able to:
L1. LO1: Gain deep Knowledge and comprehensive understanding of the main methods and tools available for data analysis, including the underlying theoretical concepts.
L2. LO2: Apply and evaluate different data analysis methods and visualisation tools to real problems and make an informed decision on their suitability for specific situations.
L3. LO3: Design data analysis methodologies for specific problems including effective communication of main findings to relevant audiences, and critically appraise the results.

Employability Skills and Personal Development Planning (PDP) Skills

SCQF Headings	During completion of this module, there will be an opportunity to achieve core skills in:
Knowledge and Understanding (K and U)	SCQF Level 11. Comprehensive knowledge and understanding of the main methods and tools employed for data analysis including the underlying theory and principles.
Practice: Applied Knowledge and Understanding	SCQF Level 11. Apply a variety of data analysis methodologies to real situations and communicate the results using available visualisation tools. Gather relevant information from the data. Make informed decisions on the most suitable methods to analyse a specific dataset. Interpret results and communicate their findings using the appropriate visualisation technique.
Generic Cognitive skills	SCQF Level 11. Critical knowledge of the state-of-the-art in data analysis. Identify most suitable methods/tools to make informed decisions in a real situation. Design a complete data analysis methodology and communicate findings in real problems.
Communication, ICT and Numeracy Skills	SCQF Level 11. Effectively communicate the information extracted from data analysis using a variety of available tools, both from report writing and presentations. Critically appraise numerical results gathered from the analysed data.
Autonomy, Accountability and Working with others	SCQF Level 11. Initiative and autonomy working in lab assignments. Students will also work as part of a team in project assignments and must develop a sense of accountability to others.
Pre-requisites:	Before undertaking this module the student should have undertaken the following:

	Module Code:	Module Title:
	Other:	No specific module is required, but a basic knowledge on statistics and maths involved is expected.
Co-requisites	Module Code:	Module Title:

* Indicates that module descriptor is not published.

Learning and Teaching

The module will be delivered by means of lectures and supervised hands-on lab work. Lectures will cover the theoretical background and practical applicability in real life problems. Concepts will be introduced by posing a practical problem and working out the needed theoretical knowledge to solve them. The delivery will encourage student participation to ensure an active learning experience. Group discussions will be held to promote critical thinking and boost informed decisions on the suitability of different state-of-the-art methods. Lab exercises will help student develop their knowledge in incremental fashion using a learning-by-doing approach. This will support the development of knowledge and understanding of the topics. In addition, labs will develop their skills to carry out a full data analysis project and communicate the results, which will be part of the final assessment.

Indicative summary of the topics covered by this module:
Unit 1-Introduction
• Overview of module and topics covered
• Introduction to data analytics and data mining
• Definitions
• Data mining process and framework

Unit 2-Exploratory data analysis and visualisation
• Review of basic concepts of probability theory and random variables
• Descriptive statistics
• Visualising single variables
• Visualising multiple variables
• Dirty data
• Data exploration vs. presentation

Unit 3-Visualisation tools
• Libraries and toolboxes in data analysis packages
• Specific packages
• Libraries and APIs
• Module toolset

Unit 4-Basic analytics and preprocessing
• Statistical methods for evaluation
• Data sources and types of data
• Data sampling
• Missing data
• Outlier detection and handling
• Normalization
• Categorization

Unit 5-Data classification I
• Introduction to statistical pattern recognition
• Unsupervised vs. supervised classification
• Feature extraction and selection
• Principal component analysis
• Linear discriminant analysis

• Dimensionality reduction

Unit 6-Data classification II
• Classifier design
• Model evaluation and error estimation
• Decision trees
• Classifiers based on Bayes decision theory

Unit 7-Advanced data classification
• Classifiers based on decision boundaries
• Neural Networks
• Support Vector Machines
• Ensembles
• Clustering

Unit 8-Regression, correlation, and time series analysis
• Correlation analysis
• Linear regression
• Logistic regression
• Introduction to time series analysis

Unit 9 – Advanced data analytics
• Decision rules
• Analysis of textual data and social media
• Association rules
• Introduction to deep learning

Unit 10- Communication and data visualisation
• Tips on data presentation
• Representation methods
• Evolution of a graph
• Cleaning up a graphic
• Case studies and examples

Learning Activities During completion of this module, the learning activities undertaken to achieve the module learning outcomes are stated below:	**Student Learning Hours** (Normally totalling 200 hours): (Note: Learning hours include both contact hours and hours spent on other learning activities)
Lecture/Core Content Delivery	20
Laboratory/Practical Demonstration/Workshop	20
Asynchronous Class Activity	50
Independent Study	106
Personal Development Plan	4
	200 Hours Total

**Indicative Resources: (eg. Core text, journals, internet access)

The following materials form essential underpinning for the module content and ultimately for the learning outcomes:
[1] C. Long (ed.). Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services; John Wiley & Sons

[2] V. Kotu and B. Deshpande. Predictive Analytics and Data Mining. Concepts and Practice with RapidMiner. Morgan Kauffman

[3] Bart Baesens. Analystics in a Big Data World. The Essential Guide to Data Science and its Applications. John Wiley & Sons

[4] I. H. Witten and E. Frank. Data Mining. Practical Machine Learning Tools and Techniques: Morgan Kauffman

[5] S. Theodoridis and K. Koutrumbas. Pattern Recognition. Academic Press.

[6] B. Rosner. Fundamentals of Biostatistics. Cengage Learning.

Additional Resources can be found online and in scientific databases.

Software packages:
• Matlab with relevant toolboxes (Statistics and Machine Learning, Neural Networks, etc.)
• R with relevant packages (ggplot2, etc.)

(**N.B. Although reading lists should include current publications, students are advised (particularly for material marked with an asterisk*) to wait until the start of session for confirmation of the most up-to-date material)

Engagement Requirements

In line with the Academic Engagement Procedure, Students are defined as academically engaged if they are regularly engaged with timetabled teaching sessions, course-related learning resources including those in the Library and on the relevant learning platform, and complete assessments and submit these on time. Please refer to the Academic Engagement Procedure at the following link: Academic engagement procedure

Supplemental Information

Programme Board	Computing
Assessment Results (Pass/Fail)	No
Subject Panel	Business & Applied Computing
Moderator	Keshav Dahal
External Examiner	C Luo
Accreditation Details	Not required at present
Version Number	1.07

Assessment: (also refer to Assessment Outcomes Grids below)

Exam (40%). Students will be required to take a formal written on-campus exam. The exam will both contain theoretical and practical questions aiming at assessing the achievement of LO1 and

LO2.

Coursework(60%). Students will work in groups to develop a specific data analysis project over a selected dataset. Coursework will assess achievement of LO2 and LO3 by means of a written report justifying the selection of methods, detailing the analysis, and presenting the results. Lab implementations will also be submitted and the main findings presented to the class.

Prior to this assessment, students lab work and progress will be assessed and formative feedback provided in the lab.

(N.B. (i) **Assessment Outcomes Grids** for the module (one for each component) can be found below which clearly demonstrate how the learning outcomes of the module will be assessed.
(ii) An **indicative schedule** listing approximate times within the academic calendar when assessment is likely to feature will be provided within the Student Handbook.)

Assessment Outcome Grids (Footnote A.)

Component 1

Assessment Type (Footnote B.)	Learning Outcome (1)	Learning Outcome (2)	Learning Outcome (3)	Weighting (%) of Assessment Element	Timetabled Contact Hours	
Unseen open book exam	✓	✓		40	2	

Component 2

Assessment Type (Footnote B.)	Learning Outcome (1)	Learning Outcome (2)	Learning Outcome (3)	Weighting (%) of Assessment Element	Timetabled Contact Hours	
Dissertation/ Project report/ Thesis		✓	✓	30	0	
Creative output/ Audiotapes/ Videotapes/ Games/ Simulations		✓		20	0	
Presentation			✓	10	2	
Combined Total For All Components				100%	4 hours	

Footnotes
A. Referred to within Assessment Section above
B. Identified in the Learning Outcome Section above

Note(s):

1. More than one assessment method can be used to assess individual learning outcomes.
2. Schools are responsible for determining student contact hours. Please refer to University Policy on contact hours (extract contained within section 10 of the Module Descriptor guidance note).
   This will normally be variable across Schools, dependent on Programmes &/or

Professional requirements.

Equality and Diversity
The University policies on equality and diversity will apply to this module. In order for the student to complete this module the student will be required to view photographic image materials. Students whose vision and hearing is substantially impaired should be assessed and counselled prior to them selecting courses requiring this module.   When a student discloses a disability a special needs advisor will agree the appropriate adjustments to be made, consulting with the module coordinator if necessary.   Diversity in cultures, backgrounds, abilities, learning and cognitive styles and individual differences are valued and appreciated. The assessments have taken this into account.   (N.B. Every effort will be made by the University to accommodate any equality and diversity issues brought to the attention of the   School.)   UWS Equality and Diversity Policy
(N.B. Every effort will be made by the University to accommodate any equality and diversity issues brought to the attention of the School)