**Module Descriptor**

# Section A

1. **Module Title**

Data Transformation

1. **SITS Module Code**

UI108006

1. **SCQF Level**

8

1. **SCQF Credit Points**

20

1. **Module Leader, include staff ID and email address**

Philippe Gleizon – nwh1pg ([philippe.gleizon@uhi.ac.uk](mailto:philippe.gleizon@uhi.ac.uk))

1. **Module Team Members, include staff IDs and email addresses**

Charles McCrimmon – nwh21cm ([charles.mccrimmon@uhi.ac.uk](mailto:charles.mccrimmon@uhi.ac.uk))

1. **Faculty and Cognate Subject Group**

**Faculty: Science, Health and the Environment**

**CSG: Engineering, Computing and the Built Environment**

1. **Exam Board and Exam Board Module Sub-group**

**Exam Board: Science, Technology and the Environment**

**Sub-group: Computing**

1. **Date of Module Start / Most Recent Revision**

September 2021 / December 2024

1. **Semester**

SC

1. **Minimum / Maximum Student Numbers**

Minimum numbers: 10

Maximum numbers: n/a

1. **Pre-requisites**

n/a

1. **Co-requisites**

n/a

1. **Mode of Study**

Table 1: Proportions of mode of study

| **Mode of study** | **Percentage** | **Hours** |
| --- | --- | --- |
| Face to face | 0% | 0 |
| Video-conference (VC facilities on UHI campus or learning centre) | 15.0% | 30 |
| Video-conference (other video technologies accessed via Internet) | 0% | 0 |
| VLE (online, tutor-supported study) | 12.5% | 25 |
| Audio conference | 0% | 0 |
| Self-directed study | 35.0% | 70 |
| Other (please specify)  Team activities | 37.5% | 75 |
| **Total** | **100%** | **200 Hours** |

1. **Assessment**

Table 2: Assessment

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Assessment number** | **Type** | **Details** | **Weighting** | **Component**  **Threshold Mark** | **Submission week** | **Learning Outcome(s) assessed** |
| **1** | Practical | Analyse, organise and transform data to enhance software performance | 50% | 40% | 12 (S1) | LO1, LO4, LO5, LO6 |
| **2** | Group work | Portfolio of evidence, equivalent to 3000 - 3500 words in total. Evidence submitted in a variety of formats including project, practical, oral presentation, discussion board participation. | 50% | 40% | 14 (S2) | All |

1. **Experiential Education**

Highlight all that apply

Work placement

Case studies 🗸

Simulations

Field trip

Laboratory work

Research project

Internship

Guest lecture 🗸

Clinical practice

Community engagement

Service learning

Job shadowing

Study abroad

Summer school

Volunteering

Co-operative education 🗸

Capstone course

Other 🗸

Other detail: group project

1. **Specialist Learning Resources**

Students are expected to sign up to several online services that will facilitate their learning and team work. This includes but is not limited to:

* IBM Cloud,
* Atlassian Jira,
* Bitbucket.

1. **Additional Costs to Students**

Students are expected to have access to a computer that they have complete control over. The following table has the recommended minimum requirements for a system. This is slightly higher than the UHI minimum requirements found at [Information for Students - Buying your own device](https://www.uhi.ac.uk/en/lis/information-for-students/buying-your-own-device/).

| **Minimum System Requirements** | | |
| --- | --- | --- |
|  | Windows | Mac |
| Operating system | Windows 10 or 11 | MacOS 11 or newer |
| Processor | 2 GHz or better, INTL or AMD is recommended  (Must support virtual machines) | |
| Graphics | OpenGL version 1.2 or later compatible | |
| RAM | 8GB or more | |
| Monitor | 17" or larger (the bigger the better) (Laptop: 15" or larger screen) | |
| Microphone / headphones | USB headset with microphone | |
| Webcam | Built-in or external (it is easier to adjust the camera angle with an external) | |
| Broadband | Reliable connection required   1. Mbps (receive)   1.5 Mbps (send)  (Cisco recommendations for good quality video calls) | |

Students are expected to be willing to sign up to a range of industry standard tools located online. Students will not be required to pay for any software.

1. **Employability / Graduate Attributes**

Employability attributes (meta-skills) have been aligned with [Skills Development Scotland’s *Skills 4.0*](https://www.skillsdevelopmentscotland.co.uk/media/pgkgrzlf/skills-4-0_a-model-to-drive-scotlands-future.pdf), published in 2018.

|  |  |
| --- | --- |
| The ability to focus on the present and deflect/avoid distractions | 🗸 |
| The ability to exercise control over your own impulses, emotions and desires | 🗸 |
| Being open to new ideas and approaches – having a growth mindset | 🗸 |
| The ability to self educate without the guidance of others | 🗸 |
| The ability to actively understand information provided by the speaker, and display interest in the topic discussed | 🗸 |
| The ability to take the perspective of others in order to understand their feelings and motivations | 🗸 |
| A sense of responsibility and concern for wider society | 🗸 |
| Being aware of others’ reactions and understanding why they react as they do | 🗸 |
| The ability to energise and create a sense of direction, purpose, excitement and momentum | 🗸 |
| Working to gain the agreement of others to a particular course of action | 🗸 |
| Encouraging others to achieve goals, accomplish tasks, and complete objectives | 🗸 |
| The ability to coach and constructively review the work of others to improve and advance their skills, knowledge and performance level | 🗸 |
| Having the ability to ignite change | 🗸 |
| Breaking down a complex problem or system into smaller, more manageable parts before developing a new way of addressing the problem | 🗸 |

Aligned with **SDS Skills 4.0. GA** students will demonstrate:

* **Self-Managemen**t: planning tasks (e.g., data wrangling sprints) with minimal external guidance.
* **Social Intelligence**: active listening, stakeholder engagement, short code review sessions.
* **Innovation**: exploring new data libraries, advanced automation for data cleaning, and sustainability considerations in data pipelines.

# Section B

1. **Module Summary**

This module aims to provide students with the knowledge and skills used for securing and manipulating data. Many successful products now use external and open data sets within them. Software developers need to understand how that data can be stored, but also how to manipulate it successfully using languages such as R and Python. In a business context this might include working alongside a Data Scientist or a member of the Marketing team. Alongside the technical goals you will be learning to develop your active listening skills as well as starting to look at case studies of products in a marketing context.

1. **Module Keywords**

Data, data science, data wrangling, databases, R, python, panda, numpy, dashboards, visualisation, animation, charts

1. **Module Learning Outcomes**

On successful completion of this module, students should be able to:

|  |  |  |
| --- | --- | --- |
| **Number** | **Theme** | **Learning Outcome** |
| 1 | Agile | Apply industry-standard processes, methods, techniques, and tools to execute projects |
| 2 | Business | Differentiate between features and benefits, giving appropriate examples, and show how to use them effectively in a marketing |
| 3 | Meta-skills\*\* | Explain what active listening is; demonstrate how to use it to appreciate others’ views and contributions |
| 4 | Security | Examine technical aspects of info security (client data protection, data protection act |
| 5 | Technical | Explain data mining algorithms for various mining goals; relate them to real-world problems including big data |
| 6 | Sustainability | Analyse andoptimise data for storage and processing. |

1. **Indicative Content**

**Skills that will be practiced and developed:**

* Continuing to grow your experience and understanding of Agile
* Develop active listening skills
* Building an understanding of legal requirements for storing and handling data
* Practice manipulating data in an appropriate language
* Building the foundation required to use machine learning algorithms effectively

**Syllabus Content**

* Teamworking and review
* Implications of information security law on technical methods
* Active listening
* Handling data in languages commonly used in Data Science
* Cleaning data
* Converting data into features
* Foundation in Statistics and Probability
  + Working with probabilities
  + Calculating conditional probabilities
  + Modelling a discrete random variable
  + Using the laws of expectation and variance
  + Using discrete probability distributions
  + Using continuous probability distributions
  + Using the normal approximation to discrete probability distributions
  + Produce appropriate charts and diagrams depending on data
  + Using a variety of methods to produce predictions

Floating point calculations and precision vs accuracy

1. **Library Resources**

[Talis Library Resource List](https://uhi.rl.talis.com/index.html)

Recommended readings:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author(s)** | **Title** | **Editor** | **Year** | **ISBN** |
| **Matthias Plaue** | **Data science – An introduction to statistics and machine learning** | **Springer** | **2023** | ISBN 978-3-662-67881-7  [ISBN 978-3-662-67882-4](https://doi.org/10.1007/978-3-662-67882-4) (eBook) |
| **Ian H Witten**  **Eibe Frank**  **Mark A Hall**  **Christopher J. Pal** | **Data Mining: practical machine learning tools and techniques** | **Morgan Kaufmann** | **2017**  **(4th ed)** | [ISBN 978-0-12-804291-5](https://www.vlebooks.com/Product/Index/888510?page=0&startBookmarkId=-1) |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |