National Unit Specification

General information

Unit title: Data Science (SCQF level 4)

Unit code: J8LW 44

Superclass: RB

Publication date: October 2024

Source: Scottish Qualifications Authority

Version: 01

Unit purpose

The purpose of this unit is to introduce learners to the basics of data science. The unit focuses on the fundamentals of data science including what it is, how it is used, and how to apply it to simple datasets. No previous knowledge or experience of data science is required. However, some computational and numerical competency is presumed.

Learners will be introduced to the basic ideas behind data science, what it is used for, basic skills in analysing simple datasets, and presenting information in a variety of simple visual formats. Statistical methods are introduced in context. The unit will permit learners to gain a familiarity with this emerging field and improve their appreciation of the growing importance of data science in contemporary society.

On completion of this unit, learners will understand the basics of data science and the importance of data in the world today as well as being able to manipulate and interpret simple datasets.

Learners may wish to progress to more advanced units in this field such as J8LW 45 Data Science at SCQF level 5, which will develop the knowledge and skills gained in this unit.

National Unit Specification: General information (continued)

Unit title: Data Science (SCQF level 4)

Outcomes

On successful completion of the unit the learner will be able to:

1. Describe data science.
2. State simple ways of analysing data.
3. Analyse simple data to communicate basic insights.

Credit points and level

1 National Unit credit at Scottish Credit and Qualifications Framework (SCQF) level 4: (6 SCQF credit points at SCQF level 4).

Recommended entry to the unit

Entry is at the discretion of the centre. No previous knowledge or experience is required. Some computational and numerical competency is assumed.

Core Skills

Opportunities to develop aspects of Core Skills are highlighted in the support notes for this unit specification.

There is no automatic certification of Core Skills or Core Skill components in this unit.

Context for delivery

If this unit is delivered as part of a group award, it is recommended that it should be taught and assessed within the subject area of the group award to which it contributes. For example, if this unit is delivered as part of the National Progression Award (NPA) in Data Science at SCQF level 4 there is overlap with other units within this award (particularly J890 44 Data Citizenship) and there will be opportunities to contextualise and integrate teaching, learning and assessment across component units.

The target cohort is learners seeking entry level data science skills.

National Unit Specification: General information (continued)

Unit title: Data Science (SCQF level 4)

The Assessment Support Pack (ASP) for this unit provides assessment and marking guidelines that exemplify the national standard for achievement. It is a valid, reliable and practicable assessment. Centres wishing to develop their own assessments should refer to the ASP to ensure a comparable standard. A list of existing ASPs is available to download from SQA's website: Internal Assessment Support Materials (www.sqa.org.uk/internal assessment support materials).

Equality and inclusion

This unit specification has been designed to ensure that there are no unnecessary barriers to learning or assessment. The individual needs of learners should be taken into account when planning learning experiences, selecting assessment methods or considering alternative evidence.

Further advice can be found on our website:
SQA Assessment Arrangements (www.sqa.org.uk/assessmentarrangements).

National Unit Specification: Statement of standards

Unit title: Data Science (SCQF level 4)

Acceptable performance in this unit will be the satisfactory achievement of the standards set out in this part of the unit specification. All sections of the statement of standards are mandatory and cannot be altered without reference to SQA.

Outcome 1

Describe data science.

Performance criteria

(a) Describe the applications of data science.
(b) Describe the benefits of data science.
(c) State the steps in solving a problem using data science.

Outcome 2

State simple ways of analysing data.

Performance criteria

(a) State common data types and data formats.
(b) State simple methods of cleaning and transforming data.
(c) State basic descriptive statistics used to summarise a dataset.
(d) Identify types of simple data visualisations.

Outcome 3

Analyse simple data to communicate basic insights.

Performance criteria

(a) Perform simple data cleaning and structuring.
(b) Perform basic analyses including sort, filter, and summarise.
(c) Visualise data to communicate basic insights.

National Unit Specification: Statement of standards (continued)

Unit title: Data Science (SCQF level 4)

Evidence requirements for this unit

Evidence is required to demonstrate that learners have achieved all outcomes and performance criteria.

Learners must provide **knowledge** and **product** evidence.

The knowledge evidence will relate to outcome 1 and outcome 2. The knowledge evidence may be sampled when testing is used and may be written or oral or a combination of these. The amount of evidence may be the minimum required to infer competence across both outcomes. The identifications, statements and descriptions may be straightforward but examples should be provided where appropriate.

Knowledge evidence may be sampled when testing is used. Testing must be carried out under supervised conditions and must be controlled in terms of location and time. Access to reference material is not permitted when testing is used. The sampling frame, on all occasions, must include outcome 1 and outcome 2 (but not every performance criterion within each outcome). The sampling frame must always include outcome 2, performance criterion (d).

The product evidence will relate to outcome 3. The product evidence will take the form of a completed analysis of a small dataset. The dataset will be supplied to the learner; some records will require cleaning and structuring. The dataset must be cleaned, structured, sorted, filtered and summarised. The analysis must include at least one visualisation, in the form of a chart or graph, which must illustrate patterns in the data.

The analysis will demonstrate that learners can:

- identify missing data and outliers
- group and summarise to get counts, totals, averages and minimum / maximum values
- sort on one column to find minimum / maximum values

The summary must include basic insights to answer at least one simple question.

The evidence may be produced by the learner with limited assistance. The analysis may be done in lightly controlled conditions, over an extended period of time, at times and places at the discretion of the learner, with access to reference materials.

The SCQF level of this unit (level 4) provides additional context on the nature of the required evidence and the associated standards. Appropriate level descriptors should be used when making judgements about the evidence.

When evidence is produced in loosely controlled conditions it must be authenticated. The guide to assessment provides further advice on methods of authentication.

The support notes section of this specification provides specific examples of instruments of assessment that will generate the required evidence.

National Unit Support Notes

Unit title: Data Science (SCQF level 4)

Unit support notes are offered as guidance and are not mandatory.

While the exact time allocated to this unit is at the discretion of the centre, the notional design length is 40 hours.

Guidance on the content and context for this unit

This unit is intended for beginners in data science. It provides a basic introduction to the subject for a wide range of learners. No previous knowledge of computer science, data science or statistics is assumed.

This unit is one in a series of units, with rising difficulty, that relate to data science. This is the first unit in that series and is the most accessible to learners. There is no requirement to undertake the units in sequence since each unit can be attempted without previous knowledge or experience of the subject. However, this unit serves as a gentle introduction to the subject and will serve as a foundation for progressing to higher level units.

The aim of the unit is to show learners what data science is, how it can be used, and how to perform simple analyses on small datasets using contemporary software.

It is suggested that the unit covers the following knowledge and skills:

Knowledge	Skills
Description of data science (how it's used and the benefits of using it). Introduce contemporary applications of data science and benefits (artificial intelligence (AI) applications): • Better / faster decision-making • Improved operations and processes • Creation of a data product • Understanding customer trends • Creating innovative products and services	Clean a simple dataset: • Rename columns • Remove unrequired columns • Reformat columns • Fix simple missing data (where values can be deduced from existing data) • Fix simple errors (for example, misspellings)

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 4)

Knowledge	Skills
Data analysis process:	Transform a simple dataset:
• The process of answering a question or solving a problem using data science (PPDAC).	• Select columns • Reorder columns • Reformat columns (display format) • Filter rows • Sort rows
Basic data types, display formats and categories of data:	Carry out data understanding of a simple dataset:
• Data types — integers, text, Boolean, floats • Display formats, for example dates, numeric and categorical (or quantitative and qualitative)	• Find the dimensions of dataset (that is, number of rows and columns). • Identify the data types in a dataset • Calculate simple descriptive statistics — summarise data to get counts, totals, averages (means) and minimum / maximum values
Simple methods to clean and transform data.	Perform an analysis on a simple dataset to solve a problem or answer a question:
Simple descriptive statistics:	
• Sum / totals • Averages (mean, median and mode) • Maximum / minimum values • Range • Frequency tables	• Identify missing data and obvious outliers (that can be easily fixed) • Group and summarise to get counts, totals, averages and minimum / maximum values • Sort on one column to find minimum / maximum values
Basic graphs and charts.	Create a simple visualisation:
• Bar chart • Line chart • Pie chart • Histogram	• Bar charts (vertical and horizontal) • Line graphs
Anatomy of each chart covered in skills section.	Communicate basic insights to answer a simple question.

Learners will require access to appropriate software to undertake this unit. It is recommended that, at this level, familiar software is used such as Microsoft Excel (trademark), or Google Sheets which provide the required functionality.

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 4)

The selection of appropriate data is important for teaching and learning. The datasets used at this level should be fairly small but varied, and include familiar and unfamiliar contexts. It is not appropriate to focus learning solely on familiar datasets. A critical objective of this unit is to introduce the size of contemporary datasets and the need for specialist tools to handle them. Familiar data will be easier for learners to understand and analyse but unfamiliar data should also be used to reinforce learning in unfamiliar contexts. It is recommended that learners use real data to improve the authenticity of learning.

There are many sources of authentic data including services such as Kaggle (Kaggle). For formative purposes, artificially generated data may be useful and can be found from sources such as Mockeroo (Mockaroo)

A Trello board of suitable datasets available at Data Sets grouped by topic. These are collated by Data Education in Schools (Data Education in Schools) and contributions from educators and learners are welcomed.

The development of learners' technical vocabulary is important. Terminology should be introduced, in context, throughout the unit. Learners should be encouraged to use the correct technical terms at all times.

The unit comprises three outcomes. Outcome 1 and outcome 2 are theoretical, and outcome 3 is practical.

Outcome 1: This outcome is a basic introduction to the field of data science. It should be assumed that this is the learner's first exposure to the subject. As such, the descriptions should be simple and cover the fundamentals. For example, the descriptions of the benefits and applications of data science (performance criterion (a)) should be high level and broad, rather than narrow and deep. When introducing the applications of data science, it is important to illustrate the breadth which spans astronomy, business, cosmology, crime, education, healthcare and science. Detailed descriptions of the application of data science in these fields are not required.

Performance criterion (b) (the steps involved in data science) should avoid technicalities. A simple: capture, clean, structure, analyse, visualise and report process description is sufficient. There is scope to introduce common sources of data.

Outcome 2: This outcome relates to simple ways of analysing data. Most of the performance criteria are self-evident. The key aspect of this outcome is the level of treatment, which should be basic (for all performance criteria).

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 4)

It is recommended that delivery of this theoretical outcome is done in the context of an actual software product. For example, the description of data types and formats (performance criterion (a)) could be related to the data types and formats supported by Excel™ (trademark) or Google Sheets.

When discussing methods of data cleaning and transformation (performance criterion (b)), the reasons for needing to clean and transform data should be emphasised, as should the time-consuming nature of data cleaning, which often constitutes the longest, and most labour intensive, part of the data analysis process. At this level, the methods of data cleaning and transformation introduced to learners should be simple. For example, cleaning may be limited to renaming / removing / reformatting columns along with fixing simple errors and missing data. Transformations might be limited to some (not much) restructuring of the data in preparation for its subsequent analysis.

The basic descriptive statistics (performance criterion (c)) should be limited to describing the size of the data set and performing such actions as grouping / summarising to find totals, averages, maximum and minimum values.

Learners should be exposed to different types of simple data visualisations (performance criterion (d)). It would be natural to combine learning with the Data Citizenship unit which requires interpretation of data at this point.

A Trello board of suitable data visualisation available at [Data Visualisations](#) grouped by visualisation type and complexity, and 'Bad Graphs' available at [Bad Graphs](#). These are collated by Data Education in Schools [(Data Education in Schools)](#) and contributions from educators and learners are welcomed.

Outcome 3: This outcome applies the knowledge gained in outcome 1 and outcome 2. Learners are required to analyse a simple dataset. The dataset should be large enough to encourage use of data analysis skills and discourage manual analysis.

Since this will be learners first exposure to analytical software (or their first exposure to the analytical features of familiar software such as Excel (trademark)), some time will be required to gain basic familiarity with the software and its analytical features. The terminology of data analytics ('clean', 'transform', etcetera) will require careful introduction.

Learners are not required to capture data. Data will be supplied to learners. To motivate learners, it is recommended that the data relates to topics of interest to learners.

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 4)

At this level, it is sufficient to frame learning around small, familiar datasets that learners use to practice basic analytical techniques using appropriate software. A significant part of this outcome could be spent exploring data visualisations (performance criterion (c)), describing best practice in how to communicate their insights.

There are lessons are available online at Learn Data Science covering some of the topics in this unit. The lessons come with Powerpoint presentations, lesson plans, activities in a choice of format (either Excel workbook or interactive Python notebook, although for level 4 Excel would be highly recommended rather than Python).

There is a Trello board of teaching resources on for the NPA Data Science available at Data Science Resources and links.

Guidance on approaches to delivery of this unit

The following distribution of time is suggested.

Outcome 1   10 hours
Outcome 2   15 hours
Outcome 3   15 hours

This unit is a mixture of theory and practice. Outcome 1 and outcome 2 relate to theory and outcome 3 relates to practice.

It is recommended that the outcomes are taught in sequence. Outcome 1 provides a broad introduction to the field, outcome 2 introduces basic analytical methods, and outcome 3 applies this knowledge to the analysis of simple datasets.

However, there is scope to combine outcome 2 and outcome 3 so that learners are introduced to methods in outcome 2 and immediately practice those methods, using appropriate software, in outcome 3. For example, once basic descriptive statistics are described in outcome 2 (performance criterion (c)), learners can use software to calculate these statistics for a variety of simple datasets (outcome 3, performance criterion (b)).

It is recommended that a problem-solving approach is taken to teaching and learning. Learners should develop their knowledge and skills in the context of different problems, with varying complexity, relating to a variety of datasets. Learners will require access to computing resources, including software capable of analysing data.

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 4)

There are many sources of engaging content about data science that will aid the delivery of outcome 1. For example, there are many case studies relating to the applications of data science, describing how it can be used in a wide range of fields. Outcome 2 provides learners first exposure to data analysis and will require care in the way that it is taught. Learning can be enlivened through the use of videos and real-world examples.

Outcome 3 is likely to be learners first experience of applying data analysis. The learning curve will be significantly reduced if this software is already familiar to learners (such as Excel™ (trademark) or Sheets)) rather than an entirely new product.

Guidance on approaches to assessment of this unit

Evidence can be generated using different types of assessment.

The following are suggestions only. There may be other methods that would be more suitable to learners and the type of learner assessment activities will vary depending on the resources available.

Centres are reminded that prior verification of centre-devised assessments would help to ensure that the national standard is being met. Where learners experience a range of assessment methods, this helps them to develop different skills that should be transferable to work or further and higher education.

A traditional approach to assessment might involve the use of a test (for knowledge evidence) and a practical exercise (for product evidence). The test could take the form of a selected response test, comprising 25 questions, with an appropriate pass mark. The practical exercise would require analysis of a simple, supplied dataset of interest or relevance to learners.

An alternative approach to assessment could involve the use of a portfolio, which would contain knowledge and product evidence. If this approach is taken, evidence for all performance criteria would be required.

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 4)

Opportunities for e-assessment

E-assessment may be appropriate for some assessments in this unit. By e-assessment we mean assessment which is supported by Information and Communication Technology (ICT), such as e-testing or the use of e-portfolios or social software. Centres which wish to use e-assessment must ensure that the national standard is applied to all learner evidence and that conditions of assessment as specified in the evidence requirements are met, regardless of the mode of gathering evidence. The most up-to-date guidance on the use of e-assessment to support SQA's qualifications is available at SQA e-Assessment. (www.sqa.org.uk/Guide to best practice.pdf).

Opportunities for developing Core and other essential skills

The unit is well suited to developing the Core Skills of Numeracy and Information and Communication Technology (ICT). ICT skills will be used throughout the unit, particularly outcome 3. Numeracy skills will be developed in outcome 2, when learners are introduced to descriptive statistics and visualisations.

The computational thinking skills of abstraction and automation will be developed in this unit when learners create models (abstraction) and perform analyses (automation) using software tools.

Employability skills will be developed in outcome 3, when learners gain skills in the use of software to analyse data. For example, skills in using spreadsheets are valued by employers.

History of changes to unit

Version	Description of change	Date

Additional copies of this unit specification can be purchased from the Scottish Qualifications Authority. Please contact the Business Development and Customer Support team, telephone 0303 333 0330.

Unit template: February 2024

General information for learners

Unit title: Data Science (SCQF level 4)

This section will help you decide whether this is the unit for you by explaining what the unit is about, what you should know or be able to do before you start, what you will need to do during the unit and opportunities for further learning and employment.

Data science is growing field which offers good job opportunities to motivated learners. This unit provides an opportunity to learn the basics of data science including what it is, how it is used, and how to apply it to simple datasets. You do not require any previous knowledge or experience of data science. You will also learn the basic skills in analysing simple datasets and presenting information as used by data scientists.

On completion of this unit, learners will understand the basics of data science and the importance of data in the world today as well as being able to manipulate and interpret simple datasets.

In outcome 1 you will learn about data science and some of the uses it is put to in real life. You will learn about the benefits of data science and why this field of computing has become so important to companies, governments and academics.

You will learn about the steps data scientists follow to understand a dataset, plan and implement a solution to a problem and then present their results.

In outcome 2 you will learn about the type of data that is found in simple data sets. You will learn about some techniques used to clean a dataset like removing errors or unwanted data.

This outcome also discusses the information that may be discovered in data and how that information can be presented.

Outcome 3 is a practical outcome where you will put the theory you learned in outcomes 1 and 2 into practise. You will be given datasets to work on throughout this outcome. You will be expected to solve simple problems by cleaning, transforming and then analysing each dataset. Once you have found the information you need you will present it in a way that clearly communicates what you have found.

You can be assessed in a variety of ways such as writing short answers to questions, making a poster or presentation, speaking to answer questions, making a short video or audio recording or working in a group with other learners.