National Unit Specification

General information

Unit title: Data Science (SCQF level 5)

Unit code: J8LW 45

Superclass: RB

Publication date: October 2024

Source: Scottish Qualifications Authority

Version: 01

Unit purpose

The purpose of this unit is to introduce learners to data science in today's world. The unit focuses on the tools and techniques involved in data science, the main methods of data analysis, and provides an opportunity for learners to apply this knowledge in a practical context. No previous knowledge or experience of data science is required. However, computational and numerical competency is assumed.

Learners will be introduced to a variety of topics relating to data science including: the applications, benefits and challenges of data science as a discipline. Different sources of data are discussed along with methods of preparing and analysing data sets. Learners will also gain practical skills in using software to identify patterns and trends in data. Statistical knowledge will be introduced in context.

On completion of this unit, learners will appreciate the basic principles of data science and be able to apply this knowledge to solve routine problems using data analysis software. Learners may wish to undertake this unit alongside J890 45 Data Citizenship at SCQF level 5 or learners may focus on particular aspects of data science by undertaking specialist units alongside this unit, such as J891 45 Machine Learning at SCQF level 5 or J2G8 45 Data Science: Statistics at SCQF level 5.

Learners may wish to progress to more advanced units in this field such as J8LW 46 Data Science at SCQF level 6, which explores data science in the context of larger datasets.

National Unit Specification: General information (continued)

Unit title: Data Science (SCQF level 5)

Outcomes

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

1. Describe data science.
2. Describe methods of data analysis.
3. Analyse data to identify patterns and trends.

Credit points and level

1 National Unit credit at Scottish Credit and Qualifications Framework (SCQF) level 5: (6 SCQF credit points at SCQF level 5).

Recommended entry to the unit

Entry is at the discretion of the centre. No previous knowledge or experience of data science is required. However, competency in computing and numeracy is required. This may be evidenced by possession of the Core Skills units in Information and Communication Technology (ICT) and Numeracy at SCQF level 5.

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 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 in Data Science at SCQF level 5 there is overlap with other units within this award (particularly J890 45 Data Citizenship) and there will be opportunities to contextualise and integrate teaching, learning and assessment across component units. There is particular scope for integration with J2G8 45 Data Science: Statistics at SCQF level 5, which would permit learners to gain a deeper appreciation of the statistical techniques involved in data science.

National Unit Specification: General information (continued)

Unit title: Data Science (SCQF level 5)

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 5)

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, benefits and challenges of data science.
(b) Describe the steps in solving a problem using data science.
(c) Identify sources of public and private datasets.

Outcome 2

Describe methods of data analysis.

Performance criteria

(a) Describe common data types and data formats.
(b) Describe the composition of a structured dataset.
(c) Describe methods of cleaning and transforming data.
(d) Describe methods of keeping data secure.
(e) Describe descriptive statistics used to summarise a dataset.
(f) Describe commonly used data visualisations and give examples of appropriate use.

Outcome 3

Analyse a dataset to identify patterns and trends.

Performance criteria

(a) Import data from an external source.
(b) Describe a dataset.
(c) Perform data cleaning and transforming.
(d) Perform analyses including sort, filter, consolidate, group and summarise.
(e) Convey insights on patterns and trends using data visualisation.

National Unit Specification: Statement of standards (continued)

Unit title: Data Science (SCQF level 5)

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 written or oral or a combination of these. The amount of evidence may be the minimum required to infer competence across both outcomes. The descriptions may be straightforward but examples should be provided where appropriate. For outcome 2, the descriptive statistics should include measures of central tendency and dispersion.

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 for outcomes 1 and 2. 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 (e).

The product evidence will relate to outcome 3. The product evidence will take the form of a completed analysis of one or more datasets. The dataset will be created (captured) by the learner, sourced externally and will require some cleaning. The dataset must be cleaned, structured, sorted, filtered, grouped and summarised. The analysis must include at least three visualisations, presented attractively and be informative.

The analysis will demonstrate that learners can:

- identify missing data and outliers, and clean / transform data
- group and summarise to get counts, totals, measures of central tendency and dispersion
- use data visualisations to identify patterns and trends.

The summary must include basic insights to answer a simple problem.

The evidence must be produced by the learner, without assistance. The analysis may be done in lightly controlled, open book conditions, over an extended period of time, at times and places at the discretion of the learner.

The SCQF level of this unit (level 5) 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.

National Unit Specification: Statement of standards (continued)

Unit title: Data Science (SCQF level 5)

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 5)

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 learners who are new to data science and those who wish to develop existing knowledge and skills (such as learners who have completed the corresponding level 4 unit). No previous knowledge of computer science, data science or statistics is assumed. However, learners should possess computational and numerical skills, which will be required to calculate statistics and perform the data analysis.

This unit is one in a series of units, with rising difficulty, that relate to data science. This is the second unit in the series. There is no requirement to undertake the units in sequence since each unit can be attempted without previous knowledge or experience of the subject.

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

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

Knowledge	**Skills**
Applications of data science. • How data science is being used in a variety of human endeavours • Value of data to individuals, groups and organisations Data science problem solving steps (PPDAC: problem, plan, data, analyse, communicate). Sources of public and private datasets.	Clean a dataset. • Remove metadata from a dataset • Rename columns using a naming convention • Fix missing data and outliers • Remove duplicates • Fix strings • Convert between different data types

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 5)

Knowledge	Skills
Structured data. Data types and data formats [and categories of data]: • Numeric (discrete and continuous) and categorical (nominal and ordinal) • L4 data types + dates, times, date-times Methods to clean and transform data: • Understand the reasons why there may be missing or outlying values, and how these reasons affect the ways in which we handle them Descriptive statistics. • Sum / totals • Averages (mean, median and mode) • Range • Percentiles • Frequency tables Graphs and charts. • Bar chart • Line chart • Pie chart • Histogram • Scatter plot • Select an appropriate graph / chart • Make a chart accessible	Transform a dataset: • Subset data • Create new variables by extracting and combining existing data • Create new variables by calculation • Reshape a dataset from wide to long or vice versa. • Combine datasets (appending and joining by single key column) Carry out Data Understanding of a dataset: • Calculate descriptive statistics — averages (mode and median and when to use) Perform an analysis on a dataset: • As level 4 • Use a data dictionary to understand a dataset • Identify patterns and trends in a dataset, • for example, genuine outliers • using data visualisations to identify patterns and trends (rather than simple counts or averages) Create a visualisation. • Select an appropriate graph / chart • Bar charts (grouped and stacked) • Scatter plots • Pie charts, including why alternative charts are better Communicate insights to answer a question or solve a problem using a data visualisation.

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 5)

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™ or Google Sheets which provide all of the required functionality.

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 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).

There is a Trello board of 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.

A Trello board of suitable data visualisations available at Data Visualisations grouped by visualisation type, and 'bad graphs' at Bad Graphs. 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).

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

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.

It is recommended that the ethical implications of data science are emphasised throughout the unit and may be discussed when examining the sources of data sets.

The unit comprises three outcomes. Outcome 1 and outcome 2 are theoretical, and outcome 3 is practical. However, the outcomes may be delivered holistically, without a clear delineation between them (see guidance on the content and context for this unit).

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 5)

Outcome 1: This outcome explores benefits and applications of data science along with the kinds of problems it may be used to solve. For those new to the subject, this will be their first exposure to the field so topics should be introduced with care. This outcome sets the scene for subsequent outcomes. Emphasis should be placed on the volume of data that is generated and the need for tools to harness this data.

The treatment of each topic should be relatively light. Breadth is more important than depth at this level. For example, in performance criterion (a) a wide range of applications of data science should be described in relatively shallow detail, rather than a narrow range of applications in depth. Emphasis should be placed on the application of data science in a range of fields, spanning astronomy to crime fighting.

Outcome 2: This outcome relates to methods of data analysis. Once again, given the level of this unit, treatment of any single topic should be light. For example, methods of cleaning and transforming data (performance criterion (c)) should be limited to the most common techniques for cleaning and structuring data. The key learning outcome is that learners appreciate the need for data cleaning and transforming, rather than an in-depth knowledge of the techniques for doing so.

The descriptive statistics (performance criterion (e)) should cover a range of basic descriptive statistics including percentiles. Learners are expected to know how to calculate these statistics manually and using software.

Learners should be introduced to a range of visualisations (performance criterion (f)) and be able to select appropriate visualisations for different types of data, including a suitable chart from the level 5. For example, learners should know the best types of visualisation to find out about the variability of a group of data points, compare two or more groups, to see if two numeric factors are correlated, or to see how a total is proportioned.

Outcome 3: This outcome applies the knowledge gained in outcome 1 and outcome 2. Learners are required to perform routines analyses on datasets.

If this is learners first exposure to analytical software (or their first exposure to the analytical features of familiar software such as Excel™ or Sheets) some time will be required to gain familiarity with the software and its analytical features. The terminology of data analytics (‘clean’, ‘transform’, ‘visualise’, etcetera) will require careful introduction.

At this level, learners are expected to understand and use specific analytical features of the software.

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 5)

Guidance on approaches to delivery of this unit

The following distribution of time is suggested.

Outcome 1 8 hours
Outcome 2 12 hours
Outcome 3 20 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 a small dataset.

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 (e)), learners can use software to calculate these statistics for a variety of small datasets (outcome 3, performance criterion (d)).

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. For example, learners could be supplied with a (fictitious) dataset comprising 500 car journeys taken from Edinburgh and attempt to answer specific questions relating to that dataset (such as the most popular destinations, the average age of cars being driven, the average journey time).

Learners will require access to computing resources, including software capable of analysing data.

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™ or Sheets) rather than an entirely new product.

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 5)

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 learners to source and analyse a dataset of appropriate size and complexity.

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.

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 particularly 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 2 and outcome 3. Numeracy skills will be developed in outcome 2, when learners are introduced to descriptive statistics, and outcome 3, when learners are introduced to visualisations.

National Unit Support Notes (continued)

Unit title: Data Science (SCQF level 5)

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 when learners gain skills in the use of software to analyse data. For example, skills in using spreadsheet software are valued by employers.

History of changes to unit

Version	Description of change	Date

Unit template: February 2024

General information for learners

Unit title: Data Science (SCQF level 5)

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 excellent job opportunities to keen, motivated students. This unit provides an opportunity to learn how data is sourced and how it is used in order to provide insights or solve problems. You do not require any previous knowledge or experience of data science. You will also learn the basic skills, used by data scientists, in preparing and analysing datasets and presenting information.

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 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 and challenges of data science and why this field of computing has become so important to companies, governments and academics.

You will learn the meaning of public and private datasets along with sources of each. You will be able to describe 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 common data types and formats along with how structured datasets are composed. You will learn about some techniques used to clean and transform a dataset along with how to keep datasets secure.

This outcome also discusses how statistics can be used to describe a dataset and how that information can be presented in appropriate ways.

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. Analysis may include querying, sorting, filtering, consolidating, grouping and summarising data. 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.