Higher National Unit Specification

General information

Unit title: Working with Data (SCQF level 8)

Unit code: J4Y6 35

Superclass:	CA
Publication date:	December 2020
Source:	Scottish Qualifications Authority
Version:	02

Unit purpose

The purpose of this unit is to develop learners' data analysis skills using a variety of software tools. This unit is suitable for learners who want to develop their existing data analysis skills, as part of their current work role, and those who wish to gain specialist skills in this field. Previous experience of data analysis is recommended but not required. A familiarity with numerical software (such as spreadsheets) and well-developed numeracy skills are assumed. The focus of the unit is data preparation and analysis, rather than data visualisation and communication.

The unit covers: data extraction (including data quality), data transformation (including data cleaning), and data analysis (including a variety of statistical analyses). Learners will use one or more software tools to carry out analyses on large, complex datasets to gain insights and facilitate data-driven decision making. Statistical methods will be introduced in context, including correlation and probability.

This unit may be undertaken alongside, or before, J4Y7 35 *Communicating with Data* at SCQF level 8. At the completion of this unit, learners may progress to a variety of related units, such as J4YB 35 *Programming for Data* and/or J4Y8 35 *Statistics for Data* at SCQF level 8.

Outcomes

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

1 Plan an analysis to provide business intelligence.
2 Extract data from a variety of sources.
3 Transform data to prepare for analysis.
4 Analyse data to provide business intelligence.

Higher National Unit Specification: General information (cont)

Unit title: Working with Data (SCQF level 8)

Credit points and level

1 Higher National Unit credit at Scottish Credit and Qualifications Framework (SCQF) level 8: (8 SCQF credit points at SCQF level 8)

Recommended entry to the unit

It is recommended that learners have previous experience of using software to carry out analyses of familiar data. Learners are required to possess numeracy skills prior to undertaking this unit, which could be evidenced by possession of the Core Skill in *Numeracy* at SCQF level 6 (or equivalent).

Core Skills

Achievement of this Unit gives automatic certification of the following Core Skills component:

Core Skill component   Critical Thinking at SCQF level 6
                            Planning and Organising at SCQF level 6

Complete Core Skill   Information and Communication Technology at SCQF level 6

There are also opportunities to develop aspects of Core Skills which are highlighted in the support notes of this unit specification.

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.

This unit is complimentary to J4Y7 35 *Communicating with Data* at SCQF level 8, which focuses on the visualisation of analysed data. As such, it is recommended that teaching, learning and assessment are combined with that unit when possible.

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 **www.sqa.org.uk/assessmentarrangements**.

Higher National Unit Specification: Statement of standards

Unit title: Working with Data (SCQF level 8)

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.

Where evidence for outcomes is assessed on a sample basis, the whole of the content listed in the knowledge and/or skills section must be taught and available for assessment. Learners should not know in advance the items on which they will be assessed and different items should be sampled on each assessment occasion.

Outcome 1

Plan an analysis to provide business intelligence.

Knowledge and/or skills

♦   Business intelligence and data-driven decision making
♦   Domain knowledge and the business context
♦   Business processes and data flows
♦   Internal and external data sources
♦   Data quality
♦   Stages in the data analysis process
♦   Descriptive, diagnostic, predictive and prescriptive analysis
♦   Data management including security
♦   Ethical implications of business requirements
♦   Tools for data analysis

Outcome 2

Extract data from a variety of sources.

Knowledge and/or skills

♦   Tools for querying data sources
♦   Types of data (categorical and numerical data and their sub-types)
♦   Data formats
♦   Data structures including tables and databases
♦   Data models and types of table relationship
♦   Data quality including data bias
♦   Ethical issues in data sourcing and extraction

Higher National Unit Specification: Statement of standards (cont)

Unit title: Working with Data (SCQF level 8)

Outcome 3

Transform data to prepare for analysis.

Knowledge and/or skills

♦ Tools for data transformation including tools for automating data transformation
♦ Types of data transformation
♦ Transformations including joins
♦ String manipulation including extracting a substring, replacing part of a string, joining strings and splitting strings
♦ Data cleaning
♦ Data modelling including star schema
♦ Data loading
♦ Legal and ethical considerations for data storage

Outcome 4

Analyse data to provide business intelligence.

Knowledge and/or skills

♦ Exploratory data analysis and targeted data analysis (hypothesis testing)
♦ Data structures including tables and databases
♦ Probability and probability distributions
♦ Summary statistics including measures of centrality, spread and skewness
♦ Concept of correlation and correlation matrices
♦ Time series analysis
♦ Role of domain knowledge in interpreting analyses
♦ Documenting analyses

Higher National Unit Specification: Statement of standards (cont)

Unit title: Working with Data (SCQF level 8)

Evidence requirements for this unit

The evidence requirements for this unit will take **one** form: **product** evidence.

The product evidence will consist of **at least one** complete analysis of a large, unfamiliar dataset that requires significant transformation prior to analysis. The required transformations should include some complex transformations.

The dataset must relate to real data and comprise at least 5,000 multi-variate records; the data must be sourced from at least three different sources (including internal and external sources). The datasets must include a range of data types, including dates and text, some of which will require categorising and parsing.

The completed analysis must include the following.

1   Plan for the analysis including business requirements and data management.
2   Evidence that data quality and data bias have been considered.
3   Evidence of data modelling.
4   Results of data analysis including statistical summaries.
5   Documentation describing the analysis.
6   Evidence of collaboration with domain experts.

The analysis must include all relevant summaries and identify all significant relationships in the data.

The evidence may be produced over an extended period, under loosely controlled conditions.

The SCQF level of this unit provides additional context on the nature of the required evidence and the associated standards. The level descriptors should be used (explicitly or implicitly) 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 *Guidelines on approaches to assessment* (see the support notes section of this specification) provides specific examples of instruments of assessment.

Higher National Unit Support Notes

Unit title: Working with Data (SCQF level 8)

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 designed to develop learners' knowledge and skills in data analysis. Special attention will be required for learners who have no existing knowledge of this subject area.

A variety of software tools could be used in this unit. It is recommended that, when possible, learners are exposed to a range of software. For example, this unit could be delivered using a general-purpose programming language, such as Python, or a dedicated data analysis toolset, such as Excel™ and Power Query™, or a notebook, such as Noteable™. Ideally, learners should be exposed to a variety of analysis tools during this unit. Whatever tool(s) is used, some degree of automation is required so it is unlikely that spreadsheet software (alone) will be sufficient.

The scope of this unit does not include data visualisation. The outcomes relate to the ETL (Extract, Transform, Load) process only. For a more complete learning experience, learners should undertake this unit alongside J4Y7 35 *Communicating with Data* at SCQF level 8.

Whenever possible, learning should take place using real data. A distinguishing feature between this unit and the corresponding level 7 unit is the size, complexity and familiarity of the datasets that learners are expected to work with. At this level, learners are expected to become competent in analysing large, complex, unfamiliar datasets.

Please note that the following guidance, relating to specific outcomes, does not seek to explain each knowledge/skills statement, which is left to the professionalism of the teacher. It seeks to clarify the *Statement of Standards* where it is potentially ambiguous. It also focuses on non-apparent teaching and learning issues that may be over-looked, or not emphasised, during unit delivery. As such, it is not representative of the relative importance of each knowledge/skill.

Higher National Unit Support Notes (cont)

Unit title: Working with Data (SCQF level 8)

Outcome 1

This outcome relates to the planning of an analysis. The knowledge and skills statements are self-explanatory. This is a non-technical outcome, with the emphasis on business requirements, business processes and data flows. The importance of understanding business processes, and working with domain experts, should be emphasised. The importance of data quality should also be emphasised, particularly the time-consuming nature of data cleaning.

It is recommended that the outcome concludes with a review of the tools that can be employed in data analysis since that will lead to the more technical outcomes that follow. A range of tools, and tool types, should be introduced including general purpose packages (such as Excel™), dedicated software (such as Power BI™), notebooks (such as Noteable™) and programming languages (such as Python).

Outcome 2

This outcome relates to data extraction. The knowledge and skills statements are self-explanatory. The most complex part of this outcome is data modelling, which learners may find challenging. Modelling should be restricted to the most common data models such as star schema, using dimension and fact tables. At this level, learners are required to source large, relatively complex, unfamiliar datasets from a variety of internal and external sources. Whenever possible, learners should use real data, rather than centre-supplied artificial datasets.

The discussion about data quality should include data bias ('bias in/bias out') and the difficulty of eliminating bias from datasets (particularly historical datasets). Both data bias and model (algorithmic) bias should be introduced and the difference explained. The tension between fairness and accuracy should be explored. Learners may not appreciate the relationship between fairness (protecting privacy) and accuracy (the descriptive or predictive correctness of an analysis). For example, increasing anonymity in a dataset or removing some characteristics (such as race or gender) is likely to increase privacy at the expense of the accuracy and value of the analysis.

Outcome 3

This outcome relates to data transformation. Learners are expected to automate their transformations so that they can be stored and updated using new data. This could be accomplished in a number of ways such as writing code (in Python for example) or using applied steps (in Power Query™).

At this level, learners are expected to be able to carry out complex transformations on the source datasets. For example, learners are expected to understand (and be able to apply) different types of join to datasets. It is also expected that learners can apply transformations to large, complex, unfamiliar datasets, and be aware of the processor, storage and data management implications of working with large datasets.

Higher National Unit Support Notes (cont)

Unit title: Working with Data (SCQF level 8)

Outcome 4

This outcome takes sourced and transformed data and performs analyses on it. Although the context for this unit is data analysis to answer specific business questions, this outcome provides an opportunity to introduce exploratory data analysis. Learners are only required to appreciate the difference between these types of analysis.

Time will be required within this outcome to teach statistical techniques and methods. The treatment of probability may be basic but sufficient to understand the concept behind probability distributions, which should include discrete and continuous distributions (including Gaussian).

At this level, learners are expected to carry out relatively sophisticated analyses on datasets. For example, all significant summaries and relationships (including correlations) should be identified within a dataset.

The importance of working with domain experts to interpret the results of analysis should be emphasised.

Guidance on approaches to delivery of this unit

A suggested distribution of time, across the outcomes, is:

Outcome 1: 6 hours

Outcome 2: 8 hours

Outcome 3: 14 hours

Outcome 4: 12 hours

Learners will require access to appropriate hardware and software throughout this unit. A range of software could be used to facilitate learning (see *Guidance on the content and context of this unit*).

This is a practical unit and the focus should be on the acquisition of practical skills in data analysis. However, theory should be introduced in context. For example, learners will require to understand data types, data formats, data structures and data models in a theoretical sense before they can apply that knowledge to the extraction and transformation of data in outcome 2 and outcome 3.

When learners are in employment, they should be given the opportunity to work with datasets relevant to their work roles.

Higher National Unit Support Notes (cont)

Unit title: Working with Data (SCQF level 8)

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.

The evidence requirements for this unit may be satisfied in two ways. An assignment may be set, towards the end of the unit, requiring learners to carry out a specific analysis on a dataset. This single, practical activity would have to satisfy all the evidence requirements. Alternatively, learners could maintain a portfolio of analyses carried out during the life of the unit. The portfolio would collectively satisfy all the evidence requirements (but note that at least one analysis must comprise at least 5,000 records).

There are opportunities to carry out formative assessment at various stages in the unit. For example, formative assessment could be carried out on the completion of each outcome to ensure that learners have grasped the knowledge contained within it. This would provide assessors with an opportunity to diagnose misconceptions, and intervene to remedy them before progressing to the next outcome.

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.

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 **www.sqa.org.uk/e-assessment**.

Opportunities for developing Core and other essential skills

Opportunities to develop aspects of Core Skills occur in all outcomes. These are:

Information and Communication Technology (ICT) at SCQF level 6

Numeracy at SCQF level 6

Learners will learn, and apply, *ICT* skills throughout this unit through their use of computer hardware and software required to carry out analyses.

Learners will also develop their numeracy knowledge and skills in outcome 4 when they use a range of arithmetical and statistical functions to perform analyses.

Higher National Unit Support Notes (cont)

Unit title: Working with Data (SCQF level 8)

Guidance on approaches to assessment of this unit

The Critical Thinking and Planning and Organising component of Problem Solving at SCQF level 6 are embedded in this unit. When a learner achieves the unit, their Core Skills profile will also be updated to include this component.

The Core Skill of Information and Communication Technology at SCQF level 6 is embedded in this unit. When a learner achieves the unit, their Core Skills profile will also be updated to include this Core Skill.

History of changes to unit

Version	Description of change	Date
02	Core Skills Components Critical Thinking and Planning and Organising at SCQF level 6 embedded. Core Skill Information and Communication Technology at SCQF level 6 embedded	09/12/20

Unit template: June 2017

General information for learners

Unit title: Working with Data (SCQF level 8)

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.

This unit is designed to develop your data analysis knowledge and skills. It is suitable for a wide range of learners who wish to improve their existing skills in this area. It is also suitable for learners who are interested in pursuing a career in data analysis or data science.

You will use a variety of software to carry out your analyses. This might include Microsoft Excel™, Power Query™, Power BI™ or Tableau™, Jupyter Notebooks or Noteable™, or programming languages such as Python or R.

The unit covers a variety of topics including:

♦   importance of business intelligence and data-driven decision making
♦   business processes and data flows
♦   stages in data analysis
♦   data management and security
♦   data quality and data bias
♦   data transformations
♦   data cleaning
♦   data modelling
♦   statistics including probability distributions
♦   descriptive and diagnostic analytics
♦   using software to perform these analyses
♦   data ethics

During the unit, you will work with datasets relevant to your current work role. You will learn how to use data analysis to gain insights into these datasets and make data-driven decisions based on your analysis.

The unit may be assessed by a practical activity (involving you in carrying out a complete analysis) or by maintaining a portfolio of your work throughout the unit.

This unit may be taken alongside J4Y7 35 *Communicating with Data* at SCQF level 8, which focuses on the visualisation and communication of the results of your analyses.

The Critical Thinking and Planning and Organising components of Problem Solving at SCQF level 6 are embedded in this unit. When a learner achieves the unit, their Core Skills profile will also be updated to include this component.

The Core Skill of Information and Communication Technology SCQF at level 6 is embedded in this unit. When a learner achieves the unit, their Core Skills profile will also be updated to include this Core Skill.