**Higher National Unit Specification**

**General information**

**Unit title:** Data Science Project (SCQF level 9)

**Unit code:** J4Y3 36

**Superclass:** CA

**Publication date:** September 2020

**Source:** Scottish Qualifications Authority

**Version:** 01

## Unit purpose

This **specialist** unit is designed for learners who wish to undertake a detailed analysis of a large, complex dataset that will produce actionable insights into the data. Learners will be expected to use a range of analysis tools, such as programming and visualisation tools, and analysis techniques, such as machine and/or statistical learning.

Learners will undertake a real analysis, using real data, to produce insights, which they will critically evaluate. The analysis will have real-world value and produce insights of value to an organisation or client. Learners will work alone to perform the analysis.

The topics covered in this unit include: project management, source control, data modelling, data analysis using machine and/or statistical learning, data visualisation, and presentation skills. At the completion of the project, learners will evaluate their work and suggest improvements. Learners will also consider ethical and security factors, as they relate to their project.

At the completion of this unit, learners may progress to undergraduate studies in data science.

## Outcomes

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

1    Manage a data science project.
2    Engineer data architectures and data pipelines.
3    Perform a statistical analysis on a large, complex dataset.
4    Evaluate the analysis.

# Higher National Unit Specification: General information (cont)

**Unit title:**    Data Science Project (SCQF level 9)

## Credit points and level

2 Higher National Unit credits at level 9 of the *Scottish Credit and Qualifications Framework*: (16 SCQF credit points at SCQF level 9)

## Recommended entry to the unit

Learners are presumed to possess knowledge and skills in statistics, programming, data engineering and the analysis and communication of data. This could be evidenced by possession of J4Y6 35 *Working with Data* at SCQF level 8, *Communicating with Data* at SCQF level 8, J4Y7 35 *Statistics for Data* at SCQF level 8 (or higher), J4YB 35 *Programming for Data* at SCQF level 8, and J4YC 36 *Data Engineering* at SCQF level 9. Knowledge of machine learning is desirable.

This unit is a mandatory unit within the Professional Development Award (PDA) Data Science at SCQF level 9. It should be done after the other mandatory unit (*Data Engineering*) and the optional units (selected from a range of units). The choice of optional unit will influence the nature of the project. For example, if *Machine Learning* (ML) is selected, the project may have an ML focus; if *Statistics for Data* is chosen, the project may have a statistical focus.

## 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.

This unit is part of the Professional Development Award (PDA) in Data Science at SCQF level 9. It should be delivered after the other units in this award have been completed.

Learners are required to carry out the activities in this project without assistance from others. Autonomous working is an important part of this unit.

The scale and scope of the project should be commensurate with the level of this unit (SCQF level 9). Learners are expected to identify their own projects, which must involve the use of real data to solve real business problems.

## 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:     Data Science Project (SCQF level 9)

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

Manage a data science project.

### Knowledge and/or skills

♦   Project management tools and techniques
♦   Stages in a data science project
♦   Business analysis including business processes, data flows and business requirements
♦   Source control systems
♦   Data security and privacy practices
♦   Domain knowledge
♦   Problems encountered in data science projects including the synthesis of competing requirements
♦   Data quality and data bias
♦   Ethical issues in data science

## Outcome 2

Engineer data architectures and data pipelines.

### Knowledge and/or skills

♦   Algorithms and data structures
♦   Conceptual logical and physical data architectures
♦   Data modelling
♦   Data stores
♦   Distributed systems
♦   Data sources
♦   Data pipelines and data automation
♦   Data processing frameworks
♦   System testing

## Higher National Unit Specification: Statement of standards (cont)

**Unit title:** Data Science Project (SCQF level 9)

## Outcome 3

Perform a statistical analysis on a large, complex dataset.

### Knowledge and/or skills

♦ Transforming and cleaning data, as appropriate to the project
♦ Selecting and applying statistical models and/or machine learning techniques
♦ Interpreting the results of statistical models and/or machine learning techniques
♦ Evaluation of the statistical models and/or machine learning techniques
♦ Creating visualisations that are clear and tailored to the problem
♦ Creating interactive dashboards and reports

## Outcome 4

Evaluate the analysis.

### Knowledge and/or skills

♦ Presenting the results of the analysis to a technical and non-technical audience
♦ Justifying the methods and models used in the analysis
♦ Identifying strengths and weaknesses of the analysis
♦ Identifying ethical issues including data bias
♦ Identifying improvements to the analysis

### Evidence requirements for this unit

Learners will need to provide evidence to demonstrate their knowledge and/or skills across all outcomes. The evidence requirements for this unit will consist of two types of evidence:

1 Product evidence
2 Performance evidence

This evidence will be produced by each learner, without assistance.

The **product evidence** relates to outcome 1, outcome 2 and outcome 3. The product evidence will have two parts.

1 Description of the project, which references all of the knowledge and skills specified in outcome 1 and outcome 2.
2 Results of the analysis, which references all of the knowledge and skills in outcome 3.

The description will be written and will be contextualised. For example, the 'project management tools and techniques' will describe the project management tools and techniques used in each project.

# Higher National Unit Specification: Statement of standards (cont)

## Unit title:    Data Science Project (SCQF level 9)

The scope of the project and the scale of the dataset should reflect the level of this unit. The dataset will consist of real data, which may be familiar or unfamiliar to the learner. The dataset must be large and complex, created from multiple sources, and require significant transformation prior to analysis. The analysis must be diagnostic, predictive or prescriptive; purely descriptive analyses are not appropriate at this level. The results of the analysis may be presented in a number of digital formats, including a dashboard and/or a report.

The **performance evidence** will relate to outcome 4. It will take the form of a Viva, when learners will present, explain and justify their methods, techniques and results, under oral questioning from two or more assessors, at least one of whom must have technical knowledge and one of whom should have knowledge of the business area. The performance evidence must be observed and recorded. Learners may access reference material during the oral examination.

The product evidence may be produced over an extended period of time, under loosely controlled conditions. When evidence is produced in loosely controlled conditions it must be authenticated. The *Guide to assessment* provides further advice on methods of authentication. The performance evidence must be produced under controlled conditions in terms of location, duration and invigilation.

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.

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:**     Data Science Project (SCQF level 9)

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 80 hours.

The following guidance 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.

## Guidance on the content and context for this unit

This unit is designed to permit learners to undertake a significant piece of project work, which will permit them to apply their existing knowledge of data analysis, data engineering, statistical modelling and/or machine learning.

The project will be done individually, without assistance from other learners. The selection of the project should be at the discretion of the learner, but the project must involve real data and the analysis should address real problems. Ideally, learners should select a relevant work-related project, involving data from their workplace. The results of the analysis must have real business value.

There is not time within this unit for teaching or learning technical skills in data analysis, data engineering or any other technical aspect of the unit. The focus of this unit is on applying existing knowledge and skills to a real business problem to provide real business solutions using a range of techniques, methods and technologies.

Datasets should be realistic in terms of their size and complexity. The analysis must be more than descriptive; it must provide diagnostic, predictive or prescriptive insights.

## Guidance on approaches to delivery of this unit

This is a project-based unit, which aims to apply learners' existing knowledge and skills in the context of a data science project.

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

Outcome 1: 10 hours
Outcome 2: 20 hours
Outcome 3: 30 hours
Outcome 4: 20 hours

# Higher National Unit Support Notes (cont)

**Unit title:**     Data Science Project (SCQF level 9)

Given that the project must be done individually, the role of the teacher will be to guide learners, rather than actively participate in the project. Autonomous working and problem solving are important learning outcomes in this unit. However, it is permissible to provide advice when required.

Learners may need particular assistance when they are selecting a project and defining the requirements (and scope) of the analysis. Every project brief should be approved by the teacher before learners proceed to development.

## 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.

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.

The assessment for this unit will have two parts. The first part is project work. The second part is the evaluation of that work, which will be done by Viva.

A project brief should be written (and approved) before learners commence work. This should define the following.

1    The business problem to be analysed.
2    Scope and scale of the analysis.
3    Sources of data.
4    Data engineering approaches.
5    Planned analyses.
6    Planned visualisations, dashboards and reports.
7    Ethical considerations.

The second part involves learners evaluating their results. This will likely be done by a presentation of their methodologies and results, and a subsequent oral examination by two or more assessors. This might typically take one to two hours in controlled conditions. The assessment of learner performance could be done using a checklist, which would define specific criteria for acceptable performance. Although presentation skills are an important part of this assessment activity, more important is the quality of analysis and the justification and evaluation of the methods and models used.

Self-assessment could be considered for inclusion in the overall assessment scheme. Inflated self-assessment scores could be countered by lower scores in other parts of the marking rubric (such as those relating to the quality of evaluation).

Formative assessment should be used at various stages in this unit. For example, for many learners, participating in a Viva will be a new (and challenging) experience, and may require practice and refinement after feedback.

## Higher National Unit Support Notes (cont)

**Unit title:**     Data Science Project (SCQF level 9)

## 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

The unit provides opportunities to develop some of the following Core Skills:

♦ Information and Communication Technology (ICT) at SCQF level 6
♦ Problem Solving at SCQF level 6
♦ Numeracy at SCQF level 6

## History of changes to unit

| Version | Description of change | Date |
|---------|----------------------|------|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## General information for learners

## Unit title:     Data Science Project (SCQF level 9)

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 involves carrying out a large-scale data science project in an area of your choice. If you are in employment, you will be encouraged to select a work-related project with real value to your employer.

You are assumed to already possess a wide range of technical skills in data analysis, data engineering, statistical modelling and/or machine learning. The focus of this unit is applying your knowledge and skills in these areas to a real-world problem.

During the unit you will learn (or apply):

♦   how to manage a data science project
♦   how to use project management software
♦   how to analyse business problems
♦   how to work with domain experts
♦   how to identify and resolve ethical issues
♦   how to build data architectures and data pipelines
♦   how to test systems
♦   how to perform statistical analyses
♦   how to create reports and dashboards
♦   how to evaluate your work

The assessment will take the form of a project, which will involve you in creating a report, carrying out an analysis and presenting your findings. The assessment will involve a Viva, when you will be examined on your methods, techniques and results.

The unit will provide you with opportunities to develop aspects, at SCQF level 6, of the Core Skills in *Information and Communication Technology* (ICT), *Problem Solving*, and *Numeracy*.