Unit title:	Transforming Data for Analysis
SQA code:	J6VE 48
SCQF level:	8
SCQF credit points:	8

History of changes

Publication date: September 2022

Version: 01

Version number	Date	Description	Authorised by

© Scottish Qualifications Authority 2022

This publication may be reproduced in whole or in part for educational purposes provided that no profit is derived from reproduction and that, if reproduced in part, the source is acknowledged.

Unit template: January 2022

Unit title:	Transforming Data for Analysis (SCQF level 8)

Unit purpose

This is about individuals performing data quality assessments and conducting data quality verification, identifying quality issues including missing values, duplicate data, inconsistencies, and the implications of data quality for data analysis. This also involves conducting data cleaning of noisy, incomplete data or data with data quality issues using relevant tools and programming languages.

Learning outcome

To prepare data and provide effective data quality assessment, identify potential data quality issues, cleanse, and standardise data ready for analysis in line with legislative, regulatory and organisational requirements.

Performance requirements	**Knowledge and understanding**
1. Collating and formatting data to facilitate quality reviews and processing ready for analysis. 2. Reshaping, restructuring, and combining datasets from various sources into required shape and format. 3. Assessing data quality using data quality toolsets to determine whether there is missing data, input issues, inconsistencies, or patterns which suggest the data is not a rightful representation. 4. Assessing the impact of the level of data quality on the data analysis requirements. 5. Identifying the options to remediate data quality issues. 6. Cleansing data to remove duplicates, typos, out of date data and other data defects. 7. Transforming data to formats required for analysis using appropriate tools and techniques. 8. Testing and assessing confidence in the data and its integrity. 9. Documenting data quality activities, issues, and resolutions.	1. The fundamentals of data structures. 2. The organisation's data architecture. 3. Data formats and their importance for analysis management. 4. How to collate and format data in line with industry standards. 5. Common data transformations including merging or appending data sources. 6. Software tools for data transformation. 7. The range of common data quality issues that can arise including misclassification, duplicate entries, spelling errors, obsolete data, compliance issues and interpretation of meaning. 8. The tools used to identify data quality issues. 9. The basic statistical methods and simple data modelling used to extract relevant data and normalise unstructured data. 10. The approaches used to deal with data quality issues. 11. How to avoid data quality issues in the future. 12. The different methods of validating data and the importance of taking corrective action. 13. The legal and regulatory requirements including Data Protection, Data Security, Intellectual Property Rights (IPR), Data sharing, marketing consent, personal data definition. 14. The ethical use of data.

Additional information about the unit

Assessment (evidence) requirements:

There should be a mix of **product**, **performance** and **knowledge** evidence for this unit.

The evidence should be produced as a by-product of normal work activities. The evidence should relate to the learner's work role. For example, the evidence relating to data quality standards should relate to the standards applicable in the workplace.

An example of product evidence is documentation describing data quality activities, issues and resolutions. An example of performance evidence is observation of the learner testing and assessing confidence in the data and its integrity. An example of knowledge evidence is written evidence that the learner understands data formats and their importance for analysis management.

Most evidence should be sourced from routine work activities. However, simulation is permissible when it is not possible to produce all the required evidence in this manner. For example, a limited number of data transformations may be routinely carried out in a workplace and additional transformations may be required.

Further guidance on the assessment (evidence) requirements for this unit, and other units, can be found in the Assessment Guidance document.

Guidance on instruments of assessment:

It is anticipated that most evidence will be generated as a by-product of routine workplace activities carried out as part of the learner's job role. However, there may be occasions when activities have to be undertaken to satisfy specific performance requirements or demonstrate specific knowledge and understanding.

In these circumstances, learners will be required to acquire the new knowledge and understanding, and apply this to a relevant work task. In exceptional circumstances, it is permissible for learners to undertake a simulated task using an artificial dataset to produce the required product or performance evidence.

It may be possible for one type of evidence to imply another. For example, data cleaning activities (performance evidence) may demonstrate a range of under-pinning knowledge and understanding (knowledge evidence). Combining assessment (evidence) in this way is encouraged. However, it is vital that each piece of evidence is clearly related to performance requirements and knowledge and understanding.

Professional recorded discussion and / or personal reflective statements can be used to demonstrate knowledge and understanding.

Additional information about the unit (continued)

Details of the relationship between the unit and relevant National Occupational Standards (if appropriate):

This unit is related to the following National Occupational Standards:

ESKITP801301: Assist in Delivering the Data Management Infrastructure to Support Data Analysis and Reporting.

ESKITP801401: Deliver Data Management Infrastructure Projects to Support Data Analysis and Reporting.

ESKITP802301: Assist in Delivering Routine Data Analysis Studies.

ESKITP802401: Design and Implement Data Analysis Studies.

ESKITP803301: Assist in Delivering Data Driven Business Insights.

https://www.ukstandards.org.uk/

Details of the relationship between the unit and other standards or curricula (if appropriate):

Assessment requirements specified by a sector or regulatory body (if appropriate):

Please refer to the Overarching Assessment Strategy for Competence-based Qualification and the associated sector-specific guidance relating to Digital Technology (Diploma in Digital Technology at SCQF Level 8) produced by SQA Accreditation (2022).

Meta-skills signposting:

This unit aligns to the following meta-skills:

- Critical thinking
- Focusing
- Integrity
- Sense making