

**Specification of Competency Standards**  
**for the Information & Communications Technology Industry**  
**Unit of Competency**

Functional Area - Data Science

Title	Conduct and review pre-processing and transformation of the data with data quality assessment
Code	111140L6
Range	This Uoc involves conducting pre-processing and transformation of the data and review data quality to ensure that it is in the optimal format, layout or shape for the project purposes using feature engineering, exploratory data analysis and a range of data audit techniques and approaches
Level	6
Credit	6 (For Reference Only)
Competency	<p>Performance Requirements</p> <ol style="list-style-type: none"> <li>1. Have knowledge of pre-processing and exploratory data analysis (See Remark 1)</li> <li>2. Appraise and manage pre-processing and transformation of the data to ensure that it is in the optimal format, layout or shape for the project purposes, i.e., feature engineering</li> <li>3. Conduct exploratory data analysis to perform initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations</li> <li>4. Conduct data quality assessment with extensive knowledge of the data, potential issues such as missing values, duplicates and inconsistent formats, and the implications for the data science/analytics process</li> </ol>
Assessment Criteria	The integrated outcome requirement of this UoCs are the abilities to have relevant knowledge and conduct the appropriate pre-processing, exploratory data analysis and data quality assessment to ensure the data in the optimal format and perform initial investigation on data quality (See Remark 2)
Remark	<ol style="list-style-type: none"> <li>1. Preprocessing of the data will be performed by analysing the data either categorical or numerical, visualizing them and some statistical decision. Exploratory data analysis (EDA) is a term for certain kinds of initial analysis and findings done with data sets, usually early on in an analytical process.</li> <li>2. Data should be valid, accurately interpreted, and applicable to the desired contexts</li> </ol>