Module Code	Pre-requisite Module codes	Co-Requisite Modules code(s)	ISCED Code	Subject Code	ECTS Credits	NFQ Level (CPD)#
DATA9910	None	None	489		10	9
Module Title	Working with Data					

School Responsible:	School of Computing

Module Overview:

The module focuses on developing the key data specific skills necessary for the successful development and implementation of an analytic project. The module covers the key skills needed to manage, create, process and extract data located in an organisation's database using R and SQL programming. The module also develops skills in general data wrangling including data cleansing, pre-processing, conversion, parsing and data fusion and integration - tasks required to handle data from a variety of sources.

Learning Outcomes (LO):	
On Completion of this module, the learner will be able to	
1	Installing, setting and configuring an R environment, including managing R packages
2	Reading and writing data using a variety of different formats and data sources
3	How to explore your data and perform descriptive analytics
4	How to process data using functions and control loops (FOR, IF, etc.)
5	How to automate your R scripts
6	Merging, transforming, and creating new data sets
7	How to process and update numeric, string, factor and date data types
8	How and when to use the different types of SQL functions
9	Understand and create an ER diagram for some everyday scenarios
10	Implementing an ER diagram using a commercial tool and creating of objects in the database
11	Examine some of the 300+ statistical function available in the SQL language
12	How to create procedures and functions in the database to process you data
13	How to use R to access data in the database, extracting, querying and storing data
14	Examine how to use embedded R execution in a database

Indicative Syllabus:

Module content will be broadly as follows:

- R
  - Introduction to R
  - Processing data from different data sources and different data formats
  - How to use the various data data structures in R
  - Performing descriptive analytics using R
  - How to use the basic graphics in R
  - How to merge, subset, sample, etc data to create new or data subsets
  - How to use R packages such as Tidyverse, tidyr, dplyr, stringr, lubridate
  - How to create and use R function
  - Creating and using R scripts for automation
  - Timeseries, forecasting and Text Mining
- SQL
  - Introduction and overview of SQL and the role of Databases
  - How to use the various SQL DML functions for basic data opertations
  - Examine how data in a database is modelled
  - How to create an ER diagram (in third normal form)
  - How to generate a logical and physical data model using a commercial ER tool
  - Examine some of the typical statistical functions available in SQL
  - How to create and use procedures and functions in the database

Module Code	Pre-requisite Module codes	Co-Requisite Modules code(s)	ISCED Code	Subject Code	ECTS Credits	NFQ Level (CPD)#
DATA9910	None	None	489		10	9
Module Title	Working with Data					

- R & SQL
  - How to use R to connect to and use a databases
  - Examine the different methods of extracting data from a databases
  - Examine some of the data issues involved in extracting data
  - Examine the different methods of loading and processing data in the database
- R in the Databases
  - Examine how database vendors have integrated R into the database
  - Explore how the to use Database R from an R client
  - Explore how to use SQL to run R in the database

Learning and Teaching Methods:	
Lectures, tutorials and computer laboratory sessions	
Total Teaching Contact Hours	48
Total Self-Directed Learning Hours	152

Module Delivery Duration:
One Semester.

Assessment		
Assessment Type	**Weighting (%)**	**LO Assessment (No.)**
Continuous Assessment	100%	1-14
Module Specific Assessment Arrangements (if applicable)		
(a) Derogations from General Assessment Regulations		
(b) Module Assessment Thresholds		
(c) Special Repeat Assessment Arrangements		

Essential Reading: (author, date, title, publisher)

Wickam & Grolemund, R for Data Science, O'Reilly Press, 2016

Nanda, Tierney, Helskyaho, Widlake, Nuijten, Real World SQL and PL/SQL, McGraw-Hill/Oracle Press

Tierney, Oracle R Enterprise, McGraw-Hill/Oracle Press

Supplemental Reading: (author, date, title, publisher)
A range of supplemental reading will be used to support the topics being covered throughout the semester.

These will include various websites, white pages, research papers, reoirts by industrial analysts, and reports, tutorials, etc from data and database management companies.

Version No:		**Amended By**	
Commencement Date		**Associated Programme Codes**	

Modules that are to be offered as Stand-Alone CPD Programmes must have an NFQ level assigned
*Details of the assessment schedule should be contained in the student handbook for the programme stage.

Date of Academic Council approval ……