

## PROG C9001: Programming for Data Analytics

Module Details	
Module Code:	PROG C9001
Full Title:	Programming for Data Analytics <b>APPROVED</b>
Valid From::	Semester 1 - 2019/20 ( June 2019 )
Language of Instruction:	English
Duration:	1 Semester
Credits::	10
Module Owner::	John Loane
Departments:	Unknown
Module Description:	This module will teach students about data structures and programming techniques which will allow them to gather, manipulate, store and graph data sets.

Module Learning Outcome	
On successful completion of this module the learner will be able to:	
#	Module Learning Outcome Description
MLO1	Analyse and evaluate the effectiveness of programming technologies for data analysis.
MLO2	Assess the most appropriate data structure to store data sets.
MLO3	Review and select libraries based on the processing of datasets.
MLO4	Design and develop programs to scrap data from the web.
MLO5	Design and prepare datasets for consumption over computer networks.
MLO6	Design and develop RESTful APIs.
Pre-requisite learning	
<b>Module Recommendations</b> <i>This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named DkIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).</i>	
No recommendations listed	

Module Indicative Content
<b>Learning Python</b> Installing, Whitespace, Basic constructs, Functions, Modules, Packages, Third-party libraries.
<b>Working with in-memory data</b> Ordered/unordered data, lists, tuples, dictionaries, sets.
<b>Working with persistent data</b> TXT, CSV, Pickles, Binaries, JSON, XLSX, Local Databases.
<b>Manipulating data</b> Curation, Sorting, Searching, Transforming, Mapping, Filtering, Comprehensions.
<b>Working with web data</b> Scraping, HTML, XML, NLTK.
<b>Working with large numerical datasets</b> Numpy and Scipy.
<b>Working with data frames, time series, financial and economic data</b> Pandas.
<b>Producing graphs and plots from your data</b> Matplotlib, Jupyter notebooks, Bokeh.
<b>Working in the cloud</b> Accessing datasets via a REST based API and publishing data programmatically on the web.
<b>Other programming technologies</b> R

Module Assessment	
Assessment Breakdown	%
Course Work	100.00%
Module Special Regulation	

## Assessments

Full Time On Campus			
Course Work			
<b>Assessment Type</b>	Continuous Assessment	<b>% of Total Mark</b>	25
<b>Marks Out Of</b>	100	<b>Pass Mark</b>	40
<b>Timing</b>	Week 3	<b>Learning Outcome</b>	1,2
<b>Duration in minutes</b>	0		
<b>Assessment Description</b> Given "dirty" data devise a series of automated cleansing operations and then save the data for later processing.			
<b>Assessment Type</b>	Continuous Assessment	<b>% of Total Mark</b>	25
<b>Marks Out Of</b>	100	<b>Pass Mark</b>	40
<b>Timing</b>	Week 6	<b>Learning Outcome</b>	2,3,4
<b>Duration in minutes</b>	0		
<b>Assessment Description</b> Devise an automated scraping strategy for web-based data, provide code that scraps, cleans, curates and stores the "clean" web-scraped data in a database.			
<b>Assessment Type</b>	Continuous Assessment	<b>% of Total Mark</b>	25
<b>Marks Out Of</b>	100	<b>Pass Mark</b>	40
<b>Timing</b>	Week 9	<b>Learning Outcome</b>	1,2,3,4
<b>Duration in minutes</b>	0		
<b>Assessment Description</b> Redo all of the work for Assessments 1 and 2 to take advantage of existing software libraries for data manipulation and analysis. Compare this approach with the previous manual approach.			
<b>Assessment Type</b>	Continuous Assessment	<b>% of Total Mark</b>	25
<b>Marks Out Of</b>	100	<b>Pass Mark</b>	40
<b>Timing</b>	End-of-Semester	<b>Learning Outcome</b>	3,5,6
<b>Duration in minutes</b>	0		
<b>Assessment Description</b> Integrate classroom-developed visualisations into a webapp and deploy to the cloud. Make sure that if backend data changes, so too do the visualizations. Provide API access to the data. This assessment will be linked with Data Project 1 which is a joint project with Research Process for Data Analytics and Advanced Statistics.			
No Project			
No Practical			
No Final Examination			

Part Time On Campus			
Course Work			
<b>Assessment Type</b>	Continuous Assessment	<b>% of Total Mark</b>	25
<b>Marks Out Of</b>	100	<b>Pass Mark</b>	40
<b>Timing</b>	Week 3	<b>Learning Outcome</b>	1,2
<b>Duration in minutes</b>	0		
<b>Assessment Description</b> Given "dirty" data devise a series of automated cleansing operations and then save the data for later processing.			
<b>Assessment Type</b>	Continuous Assessment	<b>% of Total Mark</b>	25
<b>Marks Out Of</b>	100	<b>Pass Mark</b>	40
<b>Timing</b>	Week 6	<b>Learning Outcome</b>	2,3,4
<b>Duration in minutes</b>	0		
<b>Assessment Description</b> Devise an automated scraping strategy for web-based data, provide code that scraps, cleans, curates and stores the "clean" web-scraped data in a database.			
<b>Assessment Type</b>	Continuous Assessment	<b>% of Total Mark</b>	25
<b>Marks Out Of</b>	100	<b>Pass Mark</b>	40
<b>Timing</b>	Week 9	<b>Learning Outcome</b>	1,2,3,4
<b>Duration in minutes</b>	0		

<b>Assessment Description</b>			
Redo all of the work for Assessments 1 and 2 to take advantage of existing software libraries for data manipulation and analysis. Compare this approach with the previous manual approach.			
<b>Assessment Type</b>	Continuous Assessment	<b>% of Total Mark</b>	25
<b>Marks Out Of</b>	100	<b>Pass Mark</b>	40
<b>Timing</b>	End-of-Semester	<b>Learning Outcome</b>	3,5,6
<b>Duration in minutes</b>	0		
<b>Assessment Description</b>			
Integrate classroom-developed visualisations into a webapp and deploy to the cloud. Make sure that if backend data changes, so too do the visualisations. Provide API access to the data. This assessment will be linked with Data Project 1 which is a joint project with Research Process for Data Analytics and Advanced Statistics.			
No Project			
No Practical			
No Final Examination			
<b>Reassessment Requirement</b>			
<b>No repeat examination</b>			
Reassessment of this module will be offered solely on the basis of coursework and a repeat examination will not be offered.			

## Module Workload

### Workload: Full Time On Campus

<i>Workload Type</i>	<i>Contact Type</i>	<i>Workload Description</i>	<i>Frequency</i>	<i>Average Weekly Learner Workload</i>	<i>Hours</i>
Practical	Contact	Practical lab session	Every Week	5.00	5
Directed Reading	Non Contact	Reading lecturer recommended texts	Every Week	3.00	3
Independent Study	Non Contact	Trying practical tasks	Every Week	8.00	8
				Total Weekly Learner Workload	16.00
				Total Weekly Contact Hours	5.00

### Workload: Part Time On Campus

<i>Workload Type</i>	<i>Contact Type</i>	<i>Workload Description</i>	<i>Frequency</i>	<i>Average Weekly Learner Workload</i>	<i>Hours</i>
Practical	Contact	Practical lab session	Every Week	5.00	5
Directed Reading	Non Contact	Reading lecturer recommended texts	Every Week	3.00	3
Independent Study	Non Contact	Trying practical tasks	Every Week	8.00	8
				Total Weekly Learner Workload	16.00
				Total Weekly Contact Hours	5.00

## Module Resources

### Recommended Book Resources

Grus, J.. (2015), Data Science From Scratch, 1. O'Reilly Media.  
Dorian Pyle. (1999), Data Preparation for Data Mining, Morgan Kaufman.  
McKinney W.. (2013), Python for Data Analysis, 1. O'Reilly Media.  
Lawson R.. (2015), Web scraping with Python, Packt.

### Recommended Article/Paper Resources

CODATA Data Science Journal,  
<http://datascience.codata.org>  
JDS Journal of Data Science,  
<http://www.jds-online.com>

### Other Resources

Website:, PyData,  
<http://pydata.org/>  
Website:, The R Project for Statistical Computing,  
<https://www.r-project.org/>  
Website:, Data Show Podcast,  
<https://www.oreilly.com/topics/oreilly-d-ata-show-podcast>  
Website:, Python Data Analysis Library,  
<http://pandas.pydata.org/>  
Website:, Matplotlib Visualization,  
<http://matplotlib.org/>  
Website:, Data Carpentry,  
<https://datacarpentry.org/>