APPROVED

PROG C9001: Programming for Data Analytics

Module Details

Module Code:	PROG C9001
Full Title:	Programming for Data Analytics APPROVED
Valid From::	Semester 2 - 2023/24 (January 2024)
Language of Instruction:	English
Duration:	1 Semester
Credits::	10
Module Owner::	• John Loane • Jack McDonnell
Departments:	Computing Science & Mathematics
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	Evaluate and compare the effectiveness of programming technologies for data analysis.
MLO2	Assess the most appropriate data structures for storing and analysing data sets.
MLO3	Review and select libraries for the processing of datasets.
MLO4	Create and develop programs to scrape data from the web.
MLO5	Design and prepare datasets for consumption over computer networks.
MLO6	Design and develop RESTful APIs.

Pre-requisite learning
Module Recommendations *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).*
No recommendations listed

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

Module Assessment

Assessment Breakdown	%
Course Work	100.00%

Module Special Regulation

Assessments

Full-time

Course Work

Assessment Type	Continuous Assessment	**% of Total Mark**	10
Marks Out Of	100	**Pass Mark**	40
Timing	S1 Week 3	**Learning Outcome**	1,2
Duration in minutes	0		

Assessment Description
Given "dirty" data devise a series of automated cleansing operations and then save the data for later processing.

Assessment Type	Continuous Assessment	**% of Total Mark**	30
Marks Out Of	100	**Pass Mark**	40
Timing	S1 Week 6	**Learning Outcome**	2,3,4
Duration in minutes	0		

Assessment Description
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.

Assessment Type	Continuous Assessment	**% of Total Mark**	30
Marks Out Of	100	**Pass Mark**	40
Timing	S1 Week 9	**Learning Outcome**	1,2,3,4
Duration in minutes	0		

Assessment Description
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.

Assessment Type	Continuous Assessment	**% of Total Mark**	30
Marks Out Of	100	**Pass Mark**	40
Timing	End-of-Semester	**Learning Outcome**	3,5,6
Duration in minutes	0		

Assessment Description
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 Statistics.

No Project

No Practical

No Final Examination

Part-time

Course Work

Assessment Type	Continuous Assessment	**% of Total Mark**	10
Marks Out Of	100	**Pass Mark**	40
Timing	S1 Week 6	**Learning Outcome**	1,2
Duration in minutes	0		

Assessment Description
CA1 (Semester 1) - Given "dirty" data devise a series of automated cleansing operations and then save the data for later processing.

Assessment Type	Continuous Assessment	**% of Total Mark**	30
Marks Out Of	100	**Pass Mark**	40
Timing	End-of-Semester	**Learning Outcome**	2,3,4
Duration in minutes	0		

Assessment Description
CA2 (Semester 1) -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.

Assessment Type	Continuous Assessment	**% of Total Mark**	30
Marks Out Of	100	**Pass Mark**	40
Timing	S1 Week 6	**Learning Outcome**	1,2,3,4
Duration in minutes	0		

Assessment Description
CA3 (Semester 2) - 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.

Assessment Type	Continuous Assessment	**% of Total Mark**	30
Marks Out Of	100	**Pass Mark**	40
Timing	End-of-Semester	**Learning Outcome**	3,5,6
Duration in minutes	0		

Assessment Description
CA 4 (Semester 2) - 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 Statistics.

No Project

No Practical

No Final Examination

Reassessment Requirement

No repeat examination
Reassessment of this module will be offered solely on the basis of coursework and a repeat examination will not be offered.

DKIT reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full-time					
Workload Type	*Contact Type*	*Workload Description*	*Frequency*	*Average Weekly Learner Workload*	*Hours*
Practical	Contact	Practical lab session using an appropriate mix of synchronous and asynchronous delivery	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					
Workload Type	*Contact Type*	*Workload Description*	*Frequency*	*Average Weekly Learner Workload*	*Hours*
Practical	Contact	2.5 hours of Practical lab session per week (when delivered over 2 semesters) using an appropriate mix of synchronous and asynchronous delivery.	Every Week	2.50	2.5
Directed Reading	Non Contact	Reading lecturer recommended texts	Every Week	1.50	1.5
Independent Study	Non Contact	Trying practical tasks	Every Week	4.00	4
				Total Weekly Learner Workload	8.00
				Total Weekly Contact Hours	2.50

Module Resources

Recommended Book Resources

Wes McKinney. (2017), Python for Data Analysis, O'Reilly Media, p.550, [ISBN: 1491957662].

Jake VanderPlas. (2023), Python Data Science Handbook, O'Reilly Media, [ISBN: 1098121228].

Joel Grus. (2019), Data Science from Scratch, O'Reilly Media, p.500, [ISBN: 1492041130].

Dorian Pyle. (1999), Data Preparation for Data Mining, Morgan Kaufmann, p.566, [ISBN: 1558605290].

Ryan Mitchell. (2018), Web Scraping with Python, [ISBN: 1491985577].

Recommended Article/Paper Resources

CODATA Data Science Journal. CODATA Data Science Journal,
http://datascience.codata.org

JDS Journal of Data Science. JDS Journal of Data Science,
http://www.jds-online.com

Other Resources

Website, Data Camp,
http://datacamp.com

Website:, Matplotlib Visualization,
http://matplotlib.org/

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:, Data Carpentry,
https://datacarpentry.org/