



<b>Title:</b>	Programming for Data Analytics <b>APPROVED</b>
<b>Long Title:</b>	Programming for Data Analytics
<b>Module Code:</b>	SOFT8032
<b>Duration:</b>	1 Semester
<b>Credits:</b>	5
<b>NFQ Level:</b>	Advanced
<b>Field of Study:</b>	Computer Software
<b>Valid From:</b>	Semester 1 - 2017/18 ( September 2017 )
<b>Module Delivered in</b>	<a href="#">4 programme(s)</a>
<b>Module Coordinator:</b>	Sean McSweeney
<b>Module Author:</b>	Ted Scully
<b>Module Description:</b>	Data analytics is a set of techniques and processes that can be used to provide identify and analyze patterns from data. In this module the learner will be provided with the skills to import and manipulate various forms of data and perform exploratory data analysis. The learner will also be equipped with the skills to effectively visualize different aspects of data and perform basic classification and clustering techniques.
<b>Learning Outcomes</b>	
<i>On successful completion of this module the learner will be able to:</i>	
LO1	Apply programming techniques to facilitate the importation, manipulation and cleaning of data.
LO2	Implement exploratory data analysis techniques and interpret results.
LO3	Choose and employ appropriate visualization techniques for depicting data.
LO4	Select and apply basic classification and clustering techniques to a range of datasets.
LO5	Evaluate the accuracy and interpret the results of classification algorithms.
<b>Pre-requisite learning</b>	
<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 CIT 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>	
<b>Incompatible Modules</b>	
<i>These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.</i>	
No incompatible modules listed	
<b>Co-requisite Modules</b>	
No Co-requisite modules listed	
<b>Requirements</b>	
<i>This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.</i>	
No requirements listed	

**Module Content & Assessment**

**Indicative Content**

**Introduction**

Overview of the terminology and applications in the area of data science and analytics. Importance of data analytics in industry.

**Data Extraction and Manipulation**

Importing data from different sources in different formats. Applying data manipulation techniques such as reshaping, pivoting, array-based indexing, joining, cleaning and munging, grouping, aggregation.

**Visualization and Exploratory Data Analysis**

Overview of a range of visualization techniques such as histograms, scatter plots, heatmaps, clustered matrices, boxplots, regression plots. Critically evaluate and assess the applicability of different visualization techniques for a specific task. Obtaining basic summary statistics such as mean, median, standard deviation, variance and range. Using programming techniques to perform more advanced analysis such as multi-collinearity analysis.

**Clustering Analysis**

Introduction to the concept and motivation for clustering. Examine the operation of a standard clustering algorithm such as k-means. Case study illustrating the advantages and limitations of applying k-means to a dataset from a specific application domain.

**Classification**

Overview of the concept of classification and its role solving real-world problems. Introduction to a standard category of classification algorithm such as decision trees. Case study to illustrate the application of classification to a dataset. Overview of k-fold cross validation as an appropriate method of evaluation.

**Assessment Breakdown**

	%
Course Work	100.00%

**Course Work**

Assessment Type	Assessment Description	Outcome addressed	% of total	Assessment Date
Open-book Examination	Perform importation, cleaning and manipulation of a dataset and perform exploratory data analysis.	1,2	20.0	Week 6
Project	Complete a comprehensive analysis of a real-world dataset and produce a report documenting findings and incorporating appropriate visualizations.	1,2,3	30.0	Week 8
Project	Select and apply appropriate classification techniques to a dataset from a specific application domain. Findings should be documented and supported with appropriate visualisations.	3,4,5	50.0	Sem End

No End of Module Formal Examination

**Reassessment Requirement**

**Coursework Only**

*This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.*

**The institute reserves the right to alter the nature and timings of assessment**

**Module Workload**

<b>Workload: Full Time</b>				
<i>Workload Type</i>	<i>Workload Description</i>	<i>Hours</i>	<i>Frequency</i>	<i>Average Weekly Learner Workload</i>
Lecture	Lecture delivering theory underpinning learning outcomes.	2.0	Every Week	2.00
Lab	Practical computer-based lab supporting learning outcomes.	2.0	Every Week	2.00
Independent Learning	Independent Student Learning.	3.0	Every Week	3.00
Total Hours				7.00
Total Weekly Learner Workload				7.00
Total Weekly Contact Hours				4.00

<b>Workload: Part Time</b>				
<i>Workload Type</i>	<i>Workload Description</i>	<i>Hours</i>	<i>Frequency</i>	<i>Average Weekly Learner Workload</i>
Lab	Practical computer-based lab supporting learning outcomes.	2.0	Every Week	2.00
Lecture	Lecture delivering theory underpinning learning outcomes.	2.0	Every Week	2.00
Independent Learning	Independent Student Learning.	3.0	Every Week	3.00
Total Hours				7.00
Total Weekly Learner Workload				7.00
Total Weekly Contact Hours				4.00

## Module Resources

### Recommended Book Resources

- Wes McKinney 2012, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*, 1st Ed., O'Reilly Media [ISBN: 9781449319793]
- John V. Guttag 2016, *Introduction to Computation and Programming Using Python: With Application to Understanding Data*, 2nd Ed., MIT Press [ISBN: 9780262529624]
- Joel Grus 2015, *Data Science from Scratch: First Principles with Python*, 1st Ed., O'Reilly Media [ISBN: 9781491901427]

### Supplementary Book Resources

- Clinton W. Brownley 2015, *Foundations for Analytics with Python*, 1st Ed., O'Reilly Media [ISBN: 9781491922538]
- Fabio Nelli 2015, *Python Data Analytics: Data Analysis and Science using pandas, matplotlib and the Python Programming Language*, 1st Ed., Apress [ISBN: 9781484209592]

*This module does not have any article/paper resources*

### Other Resources

- Website: *Python Documentation*  
<http://www.python.org/doc/>
- Website: *Python Data Analysis Library*  
<http://pandas.pydata.org/>
- Website: *Numerical Python Library*  
<http://www.numpy.org/>
- Website: *Python Interactive Computing*  
<http://ipython.org/>
- Website: *Seaborn Visualization*  
<http://seaborn.pydata.org/>
- Website: *Matplotlib Visualization*  
<http://matplotlib.org/>

**Module Delivered in**

<b>Programme Code</b>	<b>Programme</b>	<b>Semester</b>	<b>Delivery</b>
CR_ESMPR_8	<a href="#"><u>Bachelor of Engineering (Honours) in Smart Product Engineering</u></a>	8	Group Elective 2
CR_KSDEV_8	<a href="#"><u>Bachelor of Science (Honours) in Software Development</u></a>	5	Mandatory
CR_KDNET_8	<a href="#"><u>Bachelor of Science (Honours) in Computer Systems</u></a>	5	Elective
CR_KCOMP_7	<a href="#"><u>Bachelor of Science in Software Development</u></a>	5	Mandatory