Module 2: Big Data Analytics

Stage				1			
Semester				1			
Module Title				Big Data Analytics			
Module Number				2			
Module Status				Mandatory			
Module ECTS Credits				5			
Module NFQ level				9			
Pre-Requisite Module Titles				None			
Co-Requisite Module Titles				None			
Capstone Module				No			
List of Module Teaching Personnel				Mr Paddy Fahy Guest Lecturers			
Contact Hours				**Non-contact Hours**			**Total Effort (hours)**
42				**58**			**100**
Lecture	**Practical**	**Tutorial**	**Seminar**	**Assignment**	**Placement**	**Independent Work**	
24	18			30		28	
Allocation of Marks (Within the Module)							
	Continuous Assessment	**Project**	**Practical**	**Final Examination**		**Total**	
Percentage Contribution	**100**					**100**	

Intended Module Learning Outcomes

On successful completion of this module the learner will be able to:

1. Perform data gathering of large data from a range of data sources.
2. Critically analyse existing Big Data datasets and implementations, taking practicality, and usefulness metrics into consideration.
3. Understand and demonstrate the role of statistics in the analysis of large of datasets
4. Select and apply suitable statistical measures and analyses techniques for data of various structure and content and present summary statistics
5. Understand and demonstrate advanced knowledge of statistical data analytics as applied to large data sets
6. Employ advanced statistical analytical skills to test assumptions, and to generate and present new information and insights from large datasets

Module Objectives

This module aims to equip the learner with a range of most relevant topics that pertain to contemporary analysis practices, and are foundational to the emerging field of big data analytics. Learners are guided through the theoretical and practical differences between traditional datasets and Big Data datasets. An overview of the initial collection of data will be explored for multiple data sources. A formal grounding in analytical statistics is a major part of the module curriculum. Learners are expected to apply principles of statistical analytics to solve problems and inform decision making. Learners achieve this through developing knowledge and understanding of statistical analytics techniques and principles while applying these techniques and principles in typical real world scenarios.

Module Curriculum

- **Overview of Big Data Analysis**
  Types of data sources and data gathering: streaming, time-dependent data, appropriate time scales, time lag / frequency-based data, granularity, dimensionality, accessing Data from a Big Data System

- **Exploratory Data Analysis**
  Detection of mistake, checking assumptions, preliminary selection of appropriate models, determining relationships among explanatory variables, assessing the direction and size of relationships between explanatory and outcome variables

- **Role of Statistics in Data Analytics**
  Origin and growth of statistics, definitions, functions of statistics: condensation, comparison, forecasting, estimating, test of hypothesis

- **Review of Fundamental Statistical Concepts**
  Overview of statistical attributes (e.g. variance, standard deviation), probability and probability distributions, illustrating probability distributions, typical scenarios where these distributions are helpful and how to apply them, Confidence Intervals and Interval Estimation,

- **Hypothesis Testing**
  Formulating the Null and Alternative hypotheses, critical region and significance level, test procedures, Type I and Type II errors

- **Statistical Analytics**
  Goodness of fit, regression and correlation, getting the measure of relationship between two variables, analysis of variance, correlation co-efficient experimental design and the analysis of variance, permutations and combinations

- **Reporting, Graphing and Plotting**
  Frequency Curves, Histograms, 2d and 3d Scatter Diagrams, graphical representation of regression and correlation information, contour plots, 3d plots, tools e.g. R, iPhython, Tableau

Reading Lists and other learning materials

Recommended Reading

Lakin S., 2011, *How to Use Statistics*, Pearson Education (Study Skills Series)

Brase C H., Brase C P., 2001, *Understanding Basic Statistics*, Houghton Mifflin

Granville V., 2014, *Developing Analytic Talent: Becoming a Data Scientist*, Wiley

Secondary Reading

Lander J., 2013, *R For Everyone: Advanced Analytics and Graphics*, Addison Wesley Data & Analytic Series

Mendenhall W., Beaver R J., Beaver B M., 2006, *Introduction to Probability & Statistics,* Thomson Higher Education

Davenport T., 2014, *Big Data at Work*, Harvard Business Review

Miller T W., 2014, *Modelling Techniques in Predictive Analytics: Business Problems and Solutions with R*, Pearson Education

Additional reading material will include peer reviewed research papers and most recent and relevant case studies.

Module Learning Environment

Accommodation

Lectures are carried out in class rooms / lecture halls in the College. Lab tutorials are carried out in computer labs throughout the Campus. All have the software required to deliver the programme.

Library

All learners have access to an extensive range of physical and electronic (remotely accessible) library resources. The library monitors and updates its resources on an on-going basis, in line with the College's Library Acquisition Policy. Lecturers update reading lists for this course on an annual basis as is the norm with all courses run by Griffith College.

Module Teaching and Learning Strategy

Each week involves both classes and practical laboratory sessions. Classes are used to deliver theoretical content and may be supported by online delivery of notes, examples, and web resources. Introduction of statistical analysis techniques are

provided in-class theory and reinforced via worked example. All work is implemented on real-world data.

Laboratory Practicals are used to provide continuous progression of theory presented in lectures with each session building upon ideas of the previous lectures and laboratory sessions.

Module Assessment Strategy

The module assessment consists of a series of tutorials, a class test, an applied statistics project and a final examination.

Element	Weighting	Type	Description	Learning Outcomes Assessed
1	40%	Weekly Work Submissions	Each worksheet will relate to the current topic covered in lecturers.	Cumulatively these worksheets will enforce learning outcomes 1,3-6
2	60%	Exam	Closed book end of the term exam	Learning outcomes 2-6