

SEVERN BUSINESS COLLEGE

Qualifi Level 7 Diploma in Data Science

Course Handbook

Qualification
Qualifi Level 7 Diploma in Data Science

Ofqual Number
610/3596/8

Level
7

Total Qualification Time
1200

Credit Value
120

Aim of the Course
The Level 7 Diploma in Data Science has been developed to enable learners to formulate research hypotheses, organise analyse and visualise data to uncover hidden solutions that can challenge traditional business assumptions and support organisations and their managers in key decision making through the application of a wide range of techniques and models used in data science.

Assessment
Assessment is through practical assignments, with no exams - to more accurately reflect the real working environment.

Course Structure

Qualifi Level 7 Diploma in Data Science			
Unit number	**Units**	**Unit level**	**Unit credit**
J/618/4970	Exploratory Data Analysis	7	8
L/618/4971	Statistical Inference	7	12
R/618/4972	Fundamentals of Predictive Modelling	7	15
Y/618/4973	Advanced Predictive Modelling	7	15
D/618/4974	Time Series Analysis	7	15
H/618/4975	Unsupervised Multivariate Methods	7	15
K/618/4976	Machine Learning	7	15
M/618/4977	Further Topics in Data Science	7	15
T/618/4978	Contemporary Themes in Business Strategy	7	10

Assessment Grades

Grade	Marking Criteria
Pass	All learning outcomes are achieved. All assessment criteria are met.
Fail	All learning outcomes are not achieved. All assessment criteria are not met.
No Marks	Plagiarism

UNIT SPECIFICATIONS

Unit Title
Exploratory Data Analysis

Level
7

Learning Time Hours
80

Credit Value
8

Unit aim
This unit provides learners with an in-depth understanding of R and Python programming and the fundamentals of statistics. This includes writing R and Python commands for data management and basic statistical analysis. The unit will help the learner to understand and perform descriptive statistics and present the data using appropriate graphs/diagrams and serves as a

foundation for advanced analytics. Most industry analysis starts with Exploratory Data Analysis and a thorough study of this will help learners to perform data health checks and provide initial business insights.

Learning outcomes and assessment criteria

In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1. Handle and manage multiple datasets within R and Python environments.	1.1 Work smoothly in R and Python development environments. 1.2 Import and export data sets and create data frames within R and Python in accordance with instructions. 1.3 Sort, merge, aggregate and append data sets in accordance with instructions.
2. Use measures of central tendency to summarize data and assess both the symmetry and variation in the data.	2.1 Differentiate between variable types and measurement scales. 2.2 Calculate the most appropriate (mean, median or mode etc.) measure of central tendency based on variable type. 2.3 Compare variation in two datasets using the coefficient of variation. 2.4 Assess symmetry of data using measures of skewness.
3. Present and summarise distributions of data and the relationships between variables graphically.	3.1 Select the most appropriate graph to present the data. 3.2 Assess distribution using Box-Plot and Histogram. 3.3 Visualize bivariate relationships using scatter-plots. 3.4 Present time-series data using motion charts.

Suggested Resources

- Exploratory Data Analysis with R Radhika Datar, Harish Garg Publisher: Packt Publishing (31 May 2019) ISBN: 178980437X
- Exploratory Data Analysis in Python Suresh Kumar Mukhiya Publisher: Packt Publishing (27 Mar. 2020) ISBN-10: 1789537258, ISBN-13: 978-1789537253

UNIT SPECIFICATIONS

Unit Title

Statistical Inference

Level

7

Learning Time Hours

120

Credit Value

12

Unit aim

This unit provides learners with an in-depth understanding of statistical distribution and hypothesis testing. Statistical distributions include Binomial, Poisson, Normal, Log Normal, Exponential, t, F and Chi Square. Parametric and non-parametric tests used in research problems are covered in this unit. The unit will help learners to formulate research hypotheses, select appropriate tests of hypothesis, write mainly R programs to perform hypothesis testing and to draw inferences using the output generated. Learners will also study planned experiments as part of the unit.

Learning outcomes and assessment criteria

In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1. Evaluate standard discrete and standard continuous distributions.	1.1 Analyse the statistical distribution of a discrete random variable. 1.2 Calculate probabilities using R for Binomial and Poisson Distribution. 1.3 Fit Binomial and Poisson distributions to observed data 1.4 Evaluate the properties of Normal and Log Normal distributions. 1.5 Calculate probabilities using R for normal and Log normal distributions. 1.6 Fit normal, Log normal and exponential distributions to observed data. 1.7 Evaluate the concept of sampling distribution (t, F and Chi Square).
2. Formulate research hypotheses and perform hypothesis testing.	2.1 Write R and Python programmes that evaluate appropriate hypothesis tests 2.2 Draw statistical inference using output in R

	2.3 Translate research problems into statistical hypotheses 2.4 Assess the most appropriate statistical test for a hypothesis
3. Analyse the concept of variance (ANOVA) and an select an appropriate ANOVA or ANCOVA model.	3.1 Define variable, factor and level for a given research problem. 3.2 Evaluate the sources of variation, explained variation and unexplained variation. 3.3 Define a linear model for ANOVA/ANCOVA. 3.4 Confirm the validity of assumption based on definitions and analysis of variation. 3.5 Perform analysis using R and Python programs to confirm validity of assumptions. 3.6 Draw inferences from statistical analysis of the research problem.

Suggested Resources

- Publisher: Packt Open Textbook Library Statistical Inference For Everyone Copyright Year: 2017 Brian Blais, Bryant University

UNIT SPECIFICATIONS

Unit Title

Fundamentals of Predictive Modelling

Level

7

Learning Time Hours

150

Credit Value

15

Unit aim

This unit provides a strong foundation for predictive modelling. Its objective is to define the entire modelling process with the help of real life case studies. Many concepts in predictive modelling methods are common and therefore, these concepts will be discussed in detail in this unit. A good understanding of predictive modelling leads to a smart data scientist as many business problems are related to successfully predicting future outcomes.

Learning outcomes and assessment criteria

In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1 Carry out global and individual testing of parameters used in defining predictive models.	1.1 Evaluate dependent variables and predictors. 1.2 Develop linear models using the lm function in R and the .ols function in Python. 1.3 Interpret signs and values of estimated regression coefficients. 1.4 Interpret output of global testing using F distributions. 1.5 Identify significant and insignificant variables.
2. Validate assumptions in multiple linear regression.	2.1 Resolve multicollinearity problems. 2.2 Revise a model after resolving the problem. 2.3 Assess the performance of the ridge regression model. 2.4 Perform residual analysis – graphically & using statistical tests to analyse results. 2.5 Resolve problems of non-normality of errors and heteroscedasticity.
3. Validate models via data partitioning, out of sample testing and cross-validation.	3.1 Develop models and implement them on testing data in accordance with the specification. 3.2 Evaluate the stability of the models using k-fold cross validation. 3.3 Evaluate influential observations using Cook's distance and hat matrix.

Suggested Resources

- Mastering Predictive Analytics with R - Second Edition James D. Miller, Rui Miguel Forte Publisher Packt Publication date: August 2017 ISBN 9781787121393

UNIT SPECIFICATIONS

Unit Title

Advanced Predictive Modelling

Level
7

Learning Time Hours
150

Credit Value
15

Unit aim
In this unit, learners are introduced to model development for categorical dependent variables. Binary dependent variables are encountered in many domains such as risk management, marketing and clinical research and this unit covers detailed model building processes for binary dependent variables. In addition, multinomial models and ordinal scaled variables will also be discussed.

Learning outcomes and assessment criteria
In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1 Develop models using binary logistic regression and assess their performance.	1.1 Evaluate when to use Binary Linear Regression correctly. 1.2 Develop realistic models using functions in R and Python. 1.3 Interpret output of global testing using Linear Regression Testing in order to assess the results. 1.4 Perform out of sample validation that tests predictive quality of the model.
2. Develop applications of multinomial logistic regression and ordinal logistic regression.	2.1 Select method for modelling categorical variable. 2.2 Develop models for nominal and ordinal scaled dependent variable in R and Python correctly.
3. Develop generalised linear models and carry out survival analysis and Cox regression.	3.1 Evaluate the concept of generalised linear models. 3.2 Apply the Poisson regression model and negative binomial regression to count data correctly. 3.3. Model 'time to event' variable using cox regression.

Suggested Resources
- Mastering Predictive Analytics with R - Second Edition James D. Miller, Rui Miguel Forte Publisher Packt Publication date: August 2017 ISBN 9781787121393

UNIT SPECIFICATIONS

Unit Title
Time Series Analysis

Level
7

Learning Time Hours
150

Credit Value
15

Unit aim
The objective of this unit is to discuss time series forecasting methods. Learners will analyse and forecast macroeconomic variables such as GDP and inflation. Panel data regression methods will also be discussed in this unit.

Learning outcomes and assessment criteria
In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1: Assess the concepts and uses of time	1.1 Create time series object in R and Python correctly including decomposing time

series analysis and test for stationarity in time series data.	series and assessing different components. 1.2 Assess whether a time series is stationary. 1.3 Transform non-stationary time series data into stationary time series data.
2: Validate ARIMA (Auto Regressive Integrated Moving Average) models and use estimation.	2.1 Identify p, d and q of ARIMA model using ACF (autocorrelation function) and a PACF (partial auto-correlation function) to describe how well values are related. 2.2 Develop ARIMA models using R and python and evaluate whether errors follow the white noise process. 2.3 Finalize the model and forecast n-period ahead to make accurate predictions.
3. Implement panel data regression methods.	3.1 Evaluate the concept of panel data regression. 3.2 Analyse the features of panel data. 3.3 Build panel data regression models in a range of contexts. 3.4 Evaluate the difference between fixed effect and random effect models.

Suggested Resources

- Publisher: Hands on Time Series Analysis with R Rami Krispin Publisher: Packt Copyright Year: May 2019

UNIT SPECIFICATIONS

Unit Title

Unsupervised Multivariate Methods

Level

7

Learning Time Hours

150

Credit Value

15

Unit aim

Data reduction is a key process in business analytics projects. In this unit, learners will learn data reduction methods such as PCA, factor analysis and MDS. They will also learn to form segments using cluster analysis methods. Forming segments and then analysing is a key technique for large groups of data and their intrinsic information comes out in detail once segmented thoughtfully.

Learning outcomes and assessment criteria

In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1: Define Principal Component Analysis (PCA) and its derivations and assess their application.	1.1 Evaluate the need for data reduction. 1.2 Perform principal component analysis and develop scoring models using R and python to minimise data loss and improve interpretability of data. 1.3 Resolve multi-collinearity using Principal Component Regression.
2: Understand hierarchical and non-hierarchical cluster analysis and assess their outputs.	2.1 Perform data reduction and derive interpretable factors and use factor scores to interpret the data set. 2.2 Obtain a brand perception map using multi-dimensional scaling.
3. Evaluate the concept of panel data regression and implement panel data methods.	3.1 Evaluate the need for cluster analysis. 3.2 Obtain clusters using suitable methods. 3.3 Interpret cluster solutions and analyse the use of clusters for business strategies.

Suggested Resources

- Applied Unsupervised Learning with R Publisher: Packt R Copyright Year: March 2019 ISBN 9781789956399 Alok Malik, Bradford Tuckfield

UNIT SPECIFICATIONS

Unit Title

Machine Learning

Level
7

Learning Time Hours
150

Credit Value
15

Unit aim
Machine learning algorithms are new generation algorithms used in conjunction with classical predictive modelling methods. In this unit, learners will understand applications of various machine learning algorithms for classification problems.

Learning outcomes and assessment criteria
In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1: Appraise classification methods including Naïve Bayes and the support vector machine algorithm.	1.1 Evaluate different methods of classification and the performance of classifiers. 1.2 Design optimum classification rules to achieve minimum error rates.
2: Apply decision tree and random forest algorithms to classification and regression problems.	2.1 Use decision trees for classification and regression problems in comparison with classical methodologies. 2.2 Analyse concepts of bootstrapping and bagging. 2.3 Apply the random forest method in a range of business and social contexts .
3. Analyse Market Baskets and apply neural networks to classification problems.	3.1 Analyse transactions data for possible associations and derive baskets of associated products. 3.2 Apply neural networks to a classification problem in domains such as speech recognition, image recognition and document categorisation.

Suggested Resources
- Introduction to Machine Learning with Python: A Guide for Data Scientists by Andreas Muller Publisher: O'Reilly 2016 ISBN10 1449369413 ISBN13 9781449369415

UNIT SPECIFICATIONS

Unit Title
Further Topics in Data Science

Level
7

Learning Time Hours
150

Credit Value
15

Unit aim
In this module, learners will learn how to analyse unstructured data using text mining. The focus will be on sentiment analysis of text data, including data available on social media. For building interactive web apps straight from R, the concept of the "SHINY" package will be introduced. Big Data concepts and artificial Intelligence will be covered in the unit, as well as an introduction to SQL programming and how it is used to handle data.

Learning outcomes and assessment criteria
In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1. Perform text mining on social media data.	1.1 Appraise the concepts and techniques used in text mining. 1.2 Analyse unstructured data and perform sentiment analysis of Twitter data to identify the positive, negative or neutral tone of the text.

2: Develop web pages using the SHINY package	2.1 Build interpretable dashboards using the SHINY package. 2.2 Host standalone applications on a web page to present the results of data analysis.
3. Apply the Hadoop framework in Big Data Analytics.	3.1 Evaluate core concepts of Hadoop. 3.2 Appraise applications of Big Data Analytics in various industries. 3.3 Evaluate the use of the HADOOP platform for performing Big Data Analytics.
4. Evaluate the fundamental concepts of artificial intelligence.	4.1 Build a simple AI model using common machine learning algorithms that support business analysis and decisionmaking. In comparison with traditional assumptions from business theory.
5. Use SQL programming for data analysis.	5.1 Evaluate core SQL for data analytics. 5.2 Carry out data wrangling and analysis in SQL to uncover insights in underutilized data.

Suggested Resources

- Applied Multivariate Statistical Analysis: Pearson New International Edition Richard Johnson, By (author) Dean Wichern Publication date 24 Jul 2013 Publisher Pearson Education Limited ISBN10 1292024941 ISBN13 9781292024943

UNIT SPECIFICATIONS

Unit Title

Contemporary Themes in Business Strategy

Level

7

Learning Time Hours

100

Credit Value

10

Unit aim

The convergence of Cloud computing, Big Data, Artificial Intelligence and The Internet of Things will see organisations of all shapes and sizes either survive and thrive or face extinction. New operational and strategic norms, types of organisations, the nature of work and employment are changing fundamentally across vast parts of the global economy. This unit introduces learners to the strategic and managerial challenges generated by the impact of digital technology on business and organisations.

Learning outcomes and assessment criteria

In order to pass this unit, the evidence that the learner presents for assessment needs to demonstrate that they can meet all the learning outcomes for the unit. The assessment criteria determine the standard required to achieve the unit.

Learning Outcome	Assessment Criteria
1. Evaluate the concept of transformation and the key technologies that drive it.	1.1 Analyse the technologies that underpin digital transformation. 1.2 Assess the managerial challenges associated with implementing digital transformation successfully.
2. Assess the strategic impact of the application of Big Data and Artificial Intelligence on business organisations.	2.1 Evaluate theories of strategy and their application to the digital economy and business. 2.2 Analyse examples of the application of Artificial intelligence on business operations or strategy.
3. Appraise theories of innovation and distinguish between disruptive and incremental change.	3.1 Evaluate theories of disruptive innovation and how they explain the impact of innovation on industries. 3.2 Evaluate the managerial challenges of promoting and implementing innovation within organizations.
4. Evaluate ethics practices within organisations and how they relate to issues in Data Science.	4.1 Assess the role that codes of ethics play in the operation and sustainability of organisations. 4.2 Evaluate the importance of reporting and disclosure for ethical practice.

Suggested Resources

- The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail (Management of Innovation and Change) ISBN-10: 1633691780 Clayton Christensen Harvard Business Review Press Jan. 2004

- Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World by Marco Iansiti (Author), Karim R. Lakhani (Author) Harvard Business Review Press Jan. 2020 ISBN-10: 1633697622 ISBN-13: 978-1633697621
- Digital Transformation Tom Siebel Rosetta Books 2019 SBN-13: 978-1948122481 ISBN-10: 1948122480

Version:2507a