NCC education®
Awarding Great British Qualifications

SHORT COURSE IN DATA SCIENCE: Advanced

Qualification Unit Specification

2020/21

Modification History

Version	Revision Description
V1.0	For release

Published by: NCC Education Limited, The Towers, Towers Business Park, Wilmslow Road, Didsbury, Manchester M20 2EZ UK

Tel: +44 (0) 161 438 6200 Fax: +44 (0) 161 438 6240 Email: info@nccedu.com
http://www.nccedu.com

1. About NCC Education

NCC Education is a UK-based awarding body, active in the UK and internationally. Originally part of the National Computing Centre, NCC Education started offering IT qualifications in 1976 and from 1997 developed its Higher Education portfolio to include Business qualifications, IT qualifications for school children and a range of Foundation qualifications.

With Centres in over forty countries, four international offices and academic managers worldwide, NCC Education strives to employ the latest technologies for learning, assessment and support. NCC Education is regulated and quality assured by Ofqual (the Office of Qualifications and Examinations Regulation, see www.ofqual.gov.uk) in England and Northern Ireland.

Overview and Objectives

This **Short Course in Data Science: Advanced** focuses on Analytics practitioners who have more than 3 years experience operating in a data analytics or analysis team looking to increase their knowledge in data science.

This course provides knowledge on all the key techniques such as ETL, Linear Algebra, Matrices, R programming, Visualisation with R's packages, Automated Knowledge Acquisition, Storytelling and Dashboard design and many more.

Hardware and Software Requirements

Hardware: Learners need access to computers with Internet access.

Software: Learners must have access to R programming and R graphic packages.

Short Course in Data Science: Advanced

Title:	Short Course in Data Science: Advanced

Guided Learning Hours	120 hours

Learning Outcomes; The Learner will:	**Objectives;** The Learner can:
1. Be able to extract, transform and load (ETL) data.	1.1 Explain what ETL is and why ETL is needed. 1.2 Understand and explain the ETL process: extraction steps, transformation steps, loading steps.
2. Be able to explain the ETL testing process and types of ETL testing.	2.1 Explain the ETL testing process, test scenarios, and test cases. 2.2 Understand and explain the responsibilities of an ETL tester.
3. Understand the basic properties of matrices and vectors, scalar multiplication, linear transformation, transpose, conjugate, rank, determinant.	3.1 Explain and find the transpose of a matrix. 3.2 Explain matrix addition and its properties. 3.3 Explain and find the conjugate and rank of a matrix. 3.4 Transform matrices to echelon and reduced echelon forms. 3.5 Establish the determinant of a matrix. 3.6 Explain the spare of vectors.
4. Understand inner and outer products, matrix inverse, matrix multiplication rules and algorithms.	4.1 Explain and find the inner and outer product. 4.2 Explain and find the inverse of matrices. 4.3 Explain matrix multiplication and various algorithms used for matrix multiplication.
5. Undertake data analysis with R.	5.1 Explain data analysis processes and types of data analysis. 5.2 Understand and program in R programming.
6. Become familiar with R data structures.	6.1 Explain different data types and data structures used in R the Programming Language. 6.2 Understand and use vectors, lists, matrices, arrays, data frame and factors in R programming.
7. Discover R's graphic packages.	7.1 Explain different visualisation packages used in R Programming Language. 7.2 Understand different types of R Graphics and graphic devices used in R programming.

8. Be able to create different visualisations using R.	8.1 Create the different type of plot graphs in R Programming language i.e. scatterplots, line plots, box plots, histogram and bar charts etc. 8.2 Explain the lattice plotting system and ggplot2 package in R Programming language.
9. Understand supervised and unsupervised learning and automated knowledge acquisition.	9.1 Explain the techniques used for supervised and unsupervised learning. 9.2 Explain knowledge engineering. 9.3 Explain the development cycle of a knowledge-based system. 9.4 Explain manual, automated and semi-automated acquisition methods. 9.5 Explain artificial intelligence rules and the development process of expert system.
10. Be able to analyse scenarios for different industries using machine learning.	10.1 Explain the evolution of machine learning. 10.2 Understand and explain different uses of machine learning in different fields.
11. Understand design principle for data visualisation.	11.1 Explain data visualisation. 11.2 Explain the bar charts, line graphs, scatterplot, sparkline, pie chart, waterfall, gauge, funnel, tables and maps. 11.3 Explain design principles for effective visualisation. 11.4 Explain the steps for designing an information visualisation
12. Be able to explain EDA and perform a data story presentation and dashboard design for communication.	12.1 Explain exploratory, descriptive and diagnostic analysis 12.2 Explain data storytelling and its benefits. 12.3 Explain dashboard and the design of dashboard for communication.

Syllabus

Topic No	Title	Content
1	Introduction to Extract, Transform and Load Process	• ETL Introduction • ETL Process • ETL Tools • ETL Testing Process • ETL Test Scenarios and Test Cases ***Learning Outcome: 1, 2***
2	Linear Algebra I	• Matrices • Matrix Transpose • Matrix Addition and Subtraction • Scalar Multiplication • Conjugation of a Matrix • Ranks of Matrices • Echelon Form • Full Rank Matrices • Linear Transformation ***Learning Outcome: 3***
3	Linear Algebra II	• Vector Inner Product • Vector Outer Product • Matrix Multiplication • Properties of Matrix Multiplication • Matrix Multiplication Algorithms • Inverse of a Matrix ***Learning Outcome: 4***

4	Data Analysis with R	• What is Data Analysis? • Types of Data Analysis • Data Analysis Process • Tools Used for Data Analysis • Introduction to R Programming • Why R programming? • Applications of R Programming • How R works? • Basic R Programming Data Types • Conditional Statements in R • Loops in R ***Learning Outcome: 5***
5	Understanding R Data Structures	• Data Types and Data Structures in R • Vectors and Properties of Vectors • Types of Atomic Vectors and accessing Vector Elements • Vectors Arithmetic • Recycling Vector Elements • Sorting a Vector • Lists and accessing elements of a list • Add and Delete Elements of List • Update Elements of Lists • Matrices, Arrays, Data Frame, Factors • Creating a Data Frame and extracting columns/rows from data frame ***Learning Outcome: 6***

6	R's Graphic packages & Visualisations	• R Visualisation Packages • R Graphics & Graphics Devices • Advantages and Disadvantages of Data Visualisation in R • Data Visualisation in R • Base R graphics • Scatterplot, Bar Charts, Box Plots, Histogram in R programming • Lattice Plotting System • Lattice Functions • Lattice ideas in a Nutshell • Ggplot2 package ***Learning Outcome: 7 & 8***
7	Automated Knowledge Acquisition	• Knowledge Engineering Process • Development cycle of a knowledge-based system • Elicitation Methods • Manual Methods • Semi-Automated Methods • Automated Methods • Artificial Intelligence Rules • Semantic Networks • Frames • Knowledge Relationship Representations • Reasoning Program • Expert System Development ***Learning Outcome: 9***

8	Supervised and Unsupervised Learning	• Machine Learning Techniques • Clustering – Why? • Stereotypical Clustering • Cluster Bias • K-means Clustering • K-means properties • Choosing clustering dimension • DBSCAN, DBSCAN Clusters and DBSCAN Algorithm • DBSCAN Performance • Matrix Factorisation – Motivation • Matrix Factorisation with SGD and • Matrix Factorisation with MCMC • Alternating Least Squares (ALS) • Performance - Offline/Online Performance ***Learning Outcome: 9***
9	Analytics Scenarios for Different Industries and Machine Learning Models	• Evolution of Machine Learning • What's required to create good machine learning systems? • Popular Machine Learning Use Cases • Supervised learning • Unsupervised learning ***Learning Outcome: 9 &10***
10	Design Principles for Data Visualisation	• Data Visualisation • What makes data visualisation effective? • Types of Big Data Visualisation Categories • Best Practices for Bar Chart Visualisation • Best Practices for Line Chart Visualisation • Best Practices for Scatterplot Visualisation • Best Practices for Sparkline Visualisation • Best Practices for Pie Chart Visualisation • Steps to Designing an Information Visualisation ***Learning Outcome: 11***

11	Exploratory, Descriptive and Diagnostic Analysis	• What is EDA and Aim of EDA • Exploratory vs Confirmatory Data Analysis • EDA and Visualisation • Steps of EDA • Classification of EDA ***Learning Outcome: 12***
12	Data Story Presentation & Dashboard Design for Communication	• Visualisation Definitions • Uses for Data Visualisations • Data Storytelling • Why Data Storytelling? • Why data storytelling the future? • Key elements of Storytelling • A basic recipe for storytelling • Dashboards • Dashboards for different departments • Dashboard Tab Types • Tips for designing a dashboard • Why should your business visualise data on a dashboard? ***Learning Outcome: 12***