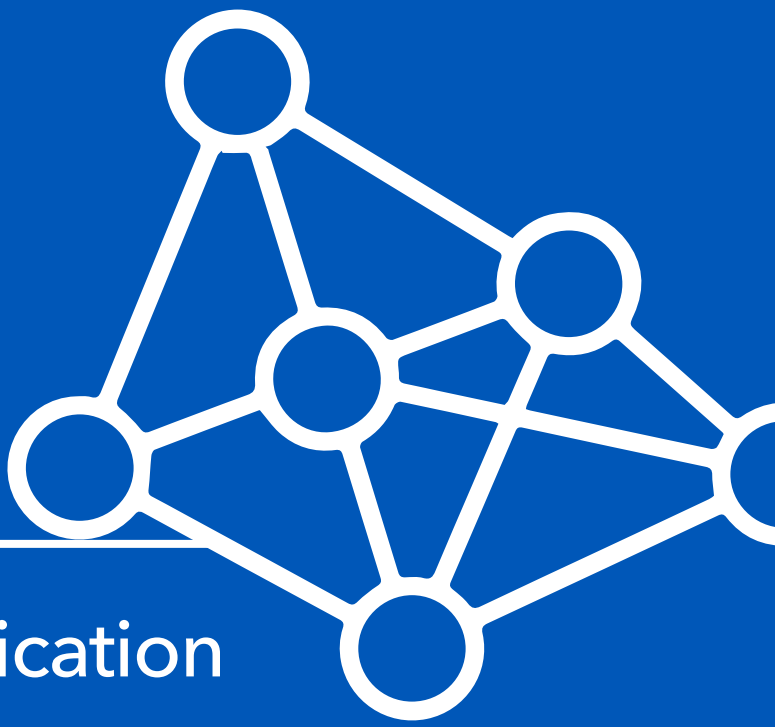
NCC education®
Awarding Great British Qualifications

SHORT COURSE IN DATA SCIENCE: Intermediate



Short Course Specification

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 Computing 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).

Overview and Objectives

This **Short Course in Data Science: Intermediate** focuses on Data Analytics for professionals who have 2 -3 years working experience in other domains such as finance, business planning, marketing or sales.

This course provides knowledge on all the key techniques such as Statistical Analysis, Probability Theory, Regression Analysis, Visualisation with Matplotlib, Python, Text Mining, Natural Language Processing, Text Clustering and many more.

Hardware and Software Requirements

Hardware: Learners need access to computers with Internet access.

Software: Learners must have access to Python, SQL and Matplotlib.

Short Course in Data Science: Intermediate

Title:	Short Course in Data Science: Intermediate

Guided Learning Hours	80 hours

Learning Outcomes; The Learner will:	**Objectives;** The Learner can:
1. Understand the skill sets needed to be a data scientist.	1.1 Explain the evaluation of data science. 1.2 Explain the skills needed to be a data scientist.
2. Be able to describe the data science process and how these components interact.	2.1 Explain the different fields interlinked with data science. 2.2 Explain the whole data science process and how its different components interact.
3. Understand probability theory and random variables.	3.1 Explain the use of probability in Data Science. 3.2 Find the probabilities for a given problem. 3.3 Explain conditional probability and Bayes theorem. 3.4 Explain different types of probability distributions (normal, binomial, and Poisson distribution).
4. Understand statistics concepts.	4.1 Explain the use of statistics in Data Science. 4.2 Differentiate quantitative data analysis and qualitative data analysis; 4.3 Explain Descriptive Statistics and Inferential Statistics. 4.4 Understand and use different measures of central tendency. 4.5 Understand and use different measures of dispersion.
5. Be able to perform linear regression analysis and statistical data modelling techniques.	5.1 Explain the different types of regression. 5.2 Explain the use cases of linear regression. 5.3 Explain the different statistical modelling techniques.
6. Be able to create and manipulation of databases with Python.	6.1 Explain the usage of Python with SQL. 6.2 Create and manipulate relational and SQL databases with Python.
7. Be able to create visualisation with Matplotlib.	7.1 Explain the architecture of Matplotlib. 7.2 Plotting with Matplotlib. 7.3 Generate different visualisation tools using Matplotlib (line plots, area plots, histograms, bar charts, box plots, and pie charts).
8. Be able to design and develop database system	8.1 Explain logical database design and Relational Database Management Systems. 8.2 Design and develop database system using SQL. 8.3 Query SQL commands on databases.

9. Understand text mining and natural language processing techniques.	9.1 Explain the areas of text mining. 9.2 Explain text mining process and methods. 9.3 Explain Natural Language Processing and NLP pipeline. 9.4 Explain the different types of techniques used in NLP.
10. Be able to classify text to categories.	10.1 Explain text classification and the hand-coded rules for text classification. 10.2 Explain the supervised learning techniques used for text classification. 10.3 Explain the difference between clustering and classification. 10.4 Explain different clustering algorithms.

Syllabus

Topic No	Title	Content
1	Introduction to Data Science	• What is Data Science? • What Data Scientists do? • How to be a data scientist? • What should you learn? • Data Science Process ***Learning Outcome: 1, 2***
2	Probability Theory and Random Variables	• Probability Concepts • Random Variables • Types of Random Variables • Properties of Random Variable • Conditional Probability • Bayes Theorem • Probability Distributions • Types of Probability Distributions ***Learning Outcome: 3***
3	Statistics Concepts	• Statistics for data science • Quantitative Analysis • Qualitative Analysis • Categories in Statistics • Descriptive Statistics • Inferential Statistics • Measures of Central Tendency • Measures of Dispersion ***Learning Outcome: 4***

4	Linear Regression Analysis and Statistical Data Modelling Techniques	• Types of regression • Linear Regression – Use cases • Statistical Data Modelling Techniques • Logistic Regression • Classification • K Nearest Neighbors • SVM (Support Vector Machine) • Naïve Bayes Classifier • Decision Trees • Random Forests • Neural Networks • Unsupervised Learning ***Learning Outcome: 5***
5	Introduction to Python	• Introduction to Python • Why Python Programming? • What can Python do? • Getting Started • Interpreter • Some Fundamental Concepts • Creating variables • Data Types • Control Flow Tools • Functions • Structured & Unstructured Data • Semi Structured Data ***Learning Outcome: 6***
6	Python with SQL	• SQL with Python • Creating a Database • Creating a Table ***Learning Outcome: 6***

7	Visualisation with Matplotlib	• Introduction to Matplotlib • Backend Layer • Artist Layer • Scripting Layer • Basic Visualisation Tools • Specialized Visualisation Tools • Advanced Visualisation Tools • Plotting Libraries ***Learning Outcome: 7***
8	Database System	• Introduction to database • Database Management Systems • E-R model construct • Basic E-R Notations • Entities • Weak & Strong Entities • Attributes • Relationships • Degree of Relationships • Cardinality of relationships ***Learning Outcome: 8***
9	Design and Develop Database System using SQL	• Integrity Constraints • Transforming E-R diagrams into Relations • Normalization • Dependency Diagram • SQL Data Definition Commands • SQL Data Manipulation Commands • Creating the database • Arithmetic Operators in SQL • Advanced Data Definition Commands • Joining Database Tables ***Learning Outcome: 8***

10	Text Mining and Natural Language Processing	• Areas of Text Mining • Text Mining Process and Methods • Natural Language Processing • Building an NLP Pipeline • Techniques Used in NLP ***Learning Outcome: 9***
11	Text Classification	• Text Classification • Text Classification: Hand Coded Rules • Text Classification: Supervised Learning • Naïve Bayes Classifier • K-Nearest Neighbors • Logistics Regression ***Learning Outcome: 10***
12	Text Clustering	• Text Clustering • Classification vs Clustering • Clustering Problem in general • Clustering in text mining • Applications of text mining • Distance metrics • Two basic categories of clustering algorithms • Evaluation metrics ***Learning Outcome: 10***