CONTENTS

List of figures and tables ix
Authors xi
Foreword xii
Acknowledgements xiv
Abbreviations xv

1. **INTRODUCTION** **1**
Practical tips 3

2. **STAKEHOLDERS** **4**
Identifying stakeholders 4
Investigating stakeholders 8
Managing stakeholders 9
Practical tips 12

3. **PROJECT DELIVERY** **13**
Data analysis lifecycle 13
Managing the lifecycle 15
Practical tips 17

4. **ETHICS AND LAWFULNESS** **18**
When to consider? 18
Who might be impacted? 20
Ethics 20
Legislation and regulation 21
Internal policies 23
Other governance 23
Commercial constraints 24
Intellectual property 24
Risk mitigation strategies 24
Practical tips 26

5. **DISCOVERY** **27**
The team 27
Understanding the domain 29
Resources 29
Problem definition 30
Developing the hypothesis 31

Effective enquiry 31
Practical tips 32

6. PROPERTIES OF DATA 33
The importance of the right properties 33
Bias or skewed data? 35
Quantitative research 37
Qualitative research 37
Things to note 39

7. SOURCING 40
Big Data 40
Data collection 41
Principles of data: data types 42
Finding data 43
Storing data 44
Moving data 45
Data governance 45
Data quality 46
Metadata 47
Practical tips 48

8. PREPARATION 49
Form 49
Properties of data sets 53
Quality risks 54
Quality checks 55
Mitigation strategies 56
Practical tips 57

9. BASIC CONCEPTS 58
Averages 58
Measures of spread 60
Probability 60
Navigating the Cartesian plane and measuring distances 62
Conclusion 65

10. MODEL SELECTION 66
Which model to use? 66
Descriptive analytics: understanding the past 67
Predictive analytics: peeking into the future 68
Prescriptive analytics: deciding what to do 68
How they work together 69
Understanding the data for your model 69
Understanding the hypothesis 71
How to choose between different models 72
Predicting value over time: time series analysis 80
Classification 85
Clustering 89
Association 94

11.	**VISUALISATIONS**	**99**
	Considerations	99
	User experience in data visualisation	100
	Bar chart	102
	Histogram	103
	Line graph	104
	Area chart	105
	Scatter plot	106
	Box and whisker plot	106
	Heat map	107
	Stem-and-leaf plot	108
	Combining charts	109
	Practical tips	112
12.	**MODEL EVALUATION**	**113**
	What is probability?	113
	What is a *p*-value?	114
	Hypothesis testing: a step-by-step guide	114
	Is this chance? Understanding significance	116
	How much does the model explain? Comparing train and test	116
	How to interpret the model	118
13.	**COMMUNICATION**	**119**
	Principles of user experience	120
	Improving user experience	121
	Know your audience	121
	Domain context	122
	Information to knowledge	122
	Storytelling	123
	Risks in storytelling	125
	Visualising a story	125
	Recommendations	127
	Operationalise	127
	Practical tips	127
14.	**MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE**	**128**
	Machine learning	129
	Data science and machine learning	130
	Neural networks	130
	Artificial intelligence	131
	Foundational models	132
	Practical tips	133
15.	**CASE STUDIES**	**134**
	Innovation Factory	134
	Smart Container Co	135
	Cognitive Business	136
	Good With	137
	SMARTabg	139

16. CONCLUSION **140**
The power of data science 140
Ethical considerations in data science 146
The future of data science 147
A final thought 147

Glossary 148
References 155
Index 157