

Orit Hazzan
Koby Mike

Guide to Teaching Data Science

An Interdisciplinary Approach

 Springer


Guide to Teaching Data Science


Orit Hazzan · Koby Mike

Guide to Teaching Data Science

An Interdisciplinary Approach

 Springer

Orit Hazzan 
Department of Education in Science
and Technology
Technion—Israel Institute of Technology
Haifa, Israel

Koby Mike 
Department of Education in Science
and Technology
Technion—Israel Institute of Technology
Haifa, Israel

ISBN 978-3-031-24757-6 ISBN 978-3-031-24758-3 (eBook)
<https://doi.org/10.1007/978-3-031-24758-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To our families, students, and colleagues

Prologue

This *Guide to Teaching Data Science* can be used by all educators in all educational environments and settings: in formal education (from elementary schools through high schools to academia) and informal education, in industry, and in non-profit governmental and third sector organizations. Specifically, the guide can be used as a textbook for Methods of Teaching Data Science courses, in which prospective and in-service teachers learn the pedagogy of data science, which is currently emerging in parallel to the development of the discipline of data science. The guide can serve also other practitioners who are curious about data science, its main characteristics, the challenges it poses, and the opportunities it offers to a variety of populations.

To benefit all of its potential user populations, the guide is organized in a way that enables immediate application of its main ideas. This goal is achieved by presenting the rationale behind the inclusion of each topic presented in this guide, its background, development, and importance in the context of data science and data science education, as well as the details of the actual teaching process (including over 200 exercises, worksheets, topics for discussions, and more).

The writing of this guide is based on our five years of experience teaching and conducting research on data science education in a variety of frameworks (2018–2022 inclusive). Specifically, we have taught courses and facilitated workshops on data science and on data science education in different formats (from 2 h active-learning workshops to full, year-long academic courses) to a variety of populations: high school pupils, undergraduate and graduate students, practitioners in different sectors, researchers in a variety of domains, and pre-service and in-service data science teachers. In parallel, we researched a variety of data science education topics, such as teaching methods, learning processes, teacher preparation, and social and organizational aspects of data science education. We are also involved in various data science education initiatives and participate in national initiatives and policy-making committees. This guide enables us to share with the professional community of data science educators the professional knowledge that we have accumulated over

the years. In addition, supplementary pedagogical material is available on our website at <https://orithazzan.net.technion.ac.il/data-science-education/>.

We would like to thank all those who have contributed to our understanding of the nature of data science education and who have fostered the interdisciplinary approach to data science presented in this guide: These include all of the students in the various courses we have taught and many prospective and in-service high school computer science and data science teachers, as well as colleagues, researchers, and instructors who have collaborated with us throughout the years in a variety of teaching, research, and development initiatives. Over the past five years, they have all shared with us their knowledge, professional experience, thoughts, and attitudes with respect to data science education. We have learned from them all.

In addition, we thank [Tech.AI—Technion Artificial Intelligence Hub](#) and [The Bernard M. Gordon Center for Systems Engineering](#), also at the Technion, for their generous support of our research. Special thanks go to Ms. Susan Spira for her highly professional editing of many of our publications (including this guide).

Haifa, Israel
November 2022

Orit Hazzan
Koby Mike

Contents

1	Introduction—What is This Guide About?	1
1.1	Introduction	1
1.2	Motivation for Writing This Guide	3
1.3	Pedagogical Principles and Guidelines for Teaching Data Science	3
1.4	The Structure of the Guide to Teaching Data Science	4
1.4.1	The Five Parts of the Guide	5
1.4.2	The Chapters of the Guide	6
1.5	How to Use This Guide?	11
1.5.1	Data Science Instructors in Academia	12
1.5.2	K-12 Teachers	12
1.5.3	Instructors of the Methods of Teaching Data Science (MTDS) Course	12
1.6	Learning Environments for Data Science	13
1.6.1	Textual Programming Environments for Data Science	13
1.6.2	Visual Programming Environments for Data Science	14
1.7	Conclusion	14
	Reference	15

Part I Overview of Data Science and Data Science Education

2	What is Data Science?	19
2.1	The Interdisciplinary Development of Data Science	19
2.1.1	The Origins of Data Science in Statistics	20
2.1.2	The Origins of Data Science in Computer Science	21
2.1.3	The Origins of Data Science in Application Domains: The Case of Business Analytics	22
2.2	Data Science as a Science	22
2.3	Data Science as a Research Method	24

- 2.3.1 Exploratory Data Analysis 24
- 2.3.2 Machine Learning as a Research Method 24
- 2.4 Data Science as a Discipline 26
- 2.5 Data Science as a Workflow 28
- 2.6 Data Science as a Profession 30
- 2.7 Conclusion 32
- References 33
- 3 Data Science Thinking 35**
 - 3.1 Introduction 35
 - 3.2 Data Thinking and the Thinking Skills Associated with Its
Components 36
 - 3.2.1 Computational Thinking 37
 - 3.2.2 Statistical Thinking 39
 - 3.2.3 Mathematical Thinking 40
 - 3.2.4 Application Domain Thinking 44
 - 3.2.5 Data Thinking 52
 - 3.3 Thinking About Data Science Thinking 53
 - 3.4 Conclusion 55
 - References 55
- 4 The Birth of a New Discipline: Data Science Education 59**
 - 4.1 Introduction 59
 - 4.2 Undergraduate Data Science Curricula Initiatives 61
 - 4.2.1 Strengthening Data Science Education Through
Collaboration, 2015 62
 - 4.2.2 Curriculum Guidelines for Undergraduate
Programs in Data Science, 2016 63
 - 4.2.3 The EDISON Data Science Framework, 2017 64
 - 4.2.4 Envisioning the Data Science Discipline, 2018 65
 - 4.2.5 Computing Competencies for Undergraduate
Data Science Curricula, 2017–2021 67
 - 4.3 Data Science Curriculum for K-12 68
 - 4.4 Meta-Analysis of Data Science Curricula 69
 - 4.5 Conclusion 71
 - References 71

Part II Opportunities and Challenges of Data Science Education

- 5 Opportunities in Data Science Education 75**
 - 5.1 Introduction 75
 - 5.2 Teaching STEM in a Real-World Context 76
 - 5.3 Teaching STEM with Real-World Data 77
 - 5.4 Bridging Gender Gaps in STEM Education 78
 - 5.5 Teaching Twenty-First Century Skills 79
 - 5.6 Interdisciplinary Pedagogy 80

5.7	Professional Development for Teachers	81
5.8	Conclusion	82
	References	82
6	The Interdisciplinarity Challenge	85
6.1	Introduction	85
6.2	The Interdisciplinary Structure of Data Science	86
6.3	Is Data Science More About Computer Science or More About Statistics?	88
6.4	Integrating the Application Domain	91
6.4.1	Data Science Pedagogical Content Knowledge (PCK)	91
6.4.2	Developing Interdisciplinary Programs	92
6.4.3	Integrating the Application Domain into Courses in Computer Science, Mathematics, and Statistics	93
6.4.4	Mentoring Interdisciplinary Projects	94
6.5	Conclusion	97
	References	98
7	The Variety of Data Science Learners	101
7.1	Introduction	101
7.2	Data Science for K-12 Pupils	103
7.3	Data Science for High School Computer Science Pupils	105
7.4	Data Science for Undergraduate Students	108
7.5	Data Science for Graduate Students	109
7.6	Data Science for Researchers	111
7.7	Data Science for Data Science Educators	111
7.8	Data Science for Professional Practitioners in the Industry	112
7.9	Data Science for Policy Makers	114
7.10	Data Science for Users	115
7.11	Data Science for the General Public	116
7.12	Activities on Learning Environments for Data Science	117
7.13	Conclusion	119
	References	119
8	Data Science as a Research Method	121
8.1	Introduction	121
8.2	Data Science as a Research Method	122
8.2.1	Data Science Research as a Grounded Theory	124
8.2.2	The Application Domain Knowledge in Data Science Research	125
8.3	Research Skills	126
8.3.1	Cognitive Skills: Awareness of the Importance of Model Assessment—Explainability and Evaluation	127

- 8.3.2 Organizational Skills: Understanding the Field of the Organization 128
- 8.3.3 Technological Skills: Data Visualization 131
- 8.4 Pedagogical Challenges of Teaching Research Skills 133
- 8.5 Conclusion 134
- References 135
- 9 The Pedagogical Chasm in Data Science Education 137**
 - 9.1 The Diffusion of Innovation Theory 137
 - 9.2 The Crossing the Chasm Theory 140
 - 9.3 The Data Science Curriculum Case Study from the Diffusion of Innovation Perspective 141
 - 9.3.1 The Story of the New Program 141
 - 9.3.2 The Teachers’ Perspective 143
 - 9.4 The Pedagogical Chasm 145
 - 9.5 Conclusion 147
 - References 147

Part III Teaching Professional Aspects of Data Science

- 10 The Data Science Workflow 151**
 - 10.1 Data Workflow 151
 - 10.2 Data Collection 154
 - 10.3 Data Preparation 156
 - 10.4 Exploratory Data Analysis 157
 - 10.5 Modeling 160
 - 10.5.1 Data Quantity, Quality, and Coverage 160
 - 10.5.2 Feature Engineering 161
 - 10.6 Communication and Action 161
 - 10.7 Conclusion 162
 - References 163
- 11 Professional Skills and Soft Skills in Data Science 165**
 - 11.1 Introduction 165
 - 11.2 Professional Skills 167
 - 11.2.1 Cognitive Skills: Thinking on Different Levels of Abstraction 167
 - 11.2.2 Organizational Skills: Storytelling 169
 - 11.2.3 Technological Skills: Programming for Data Science 171
 - 11.3 Soft Skills 172
 - 11.3.1 Cognitive Skills: Learning 172
 - 11.3.2 Organizational Skills: Teamwork and Collaboration 173
 - 11.3.3 Technological Skills: Debugging Data and Models 174

- 11.4 Teaching Notes 176
- 11.5 Conclusion 177
- References 177
- 12 Social and Ethical Issues of Data Science 179**
 - 12.1 Introduction 179
 - 12.2 Data Science Ethics 180
 - 12.3 Methods of Teaching Social Aspects of Data Science 185
 - 12.3.1 Teaching Principles 186
 - 12.3.2 Kinds of Activities 188
 - 12.4 Conclusion 194
 - References 194

Part IV Machine Learning Education

- 13 The Pedagogical Challenge of Machine Learning Education 199**
 - 13.1 Introduction 199
 - 13.2 Black Box and White Box Understandings 200
 - 13.3 Teaching ML to a Variety of Populations 201
 - 13.3.1 Machine Learning for Data Science Majors
and Allied Majors 202
 - 13.3.2 Machine Learning for Non-major Students 202
 - 13.3.3 Machine Learning for ML Users 203
 - 13.4 Framework Remarks for ML Education 204
 - 13.4.1 Statistical Thinking 204
 - 13.4.2 Interdisciplinary Projects 205
 - 13.4.3 The Application Domain Knowledge 206
 - 13.5 Conclusion 207
 - References 207
- 14 Core Concepts of Machine Learning 209**
 - 14.1 Introduction 209
 - 14.2 Types of Machine Learning 210
 - 14.3 Machine Learning Parameters and Hyperparameters 211
 - 14.4 Model Training, Testing, and Validation 211
 - 14.5 Machine Learning Performance Indicators 214
 - 14.6 Bias and Variance 217
 - 14.7 Model Complexity 218
 - 14.8 Overfitting and Underfitting 219
 - 14.9 Loss Function Optimization and the Gradient Descent
Algorithm 222
 - 14.10 Regularization 222
 - 14.11 Conclusion 223
 - References 223

- 15 Machine Learning Algorithms** 225
 - 15.1 Introduction 225
 - 15.2 K-nearest Neighbors 226
 - 15.3 Decision Trees 229
 - 15.4 Perceptron 230
 - 15.5 Linear Regression 232
 - 15.6 Logistic Regression 232
 - 15.7 Neural Networks 233
 - 15.8 Conclusion 234
 - References 234

- 16 Teaching Methods for Machine Learning** 235
 - 16.1 Introduction 235
 - 16.2 Visualization 236
 - 16.3 Hand-On Tasks 237
 - 16.3.1 Hands-On Task for the KNN Algorithm 238
 - 16.3.2 Hands-On Task for the Perceptron Algorithm 240
 - 16.3.3 Hands-On Task for the Gradient Descent
Algorithm 243
 - 16.3.4 Hands-On Task for Neural Networks 246
 - 16.4 Programming Tasks 246
 - 16.5 Project-Based Learning 248
 - 16.6 Conclusion 248
 - References 249

- Part V Frameworks for Teaching Data Science**

- 17 Data Science for Managers and Policymakers** 253
 - 17.1 Introduction 253
 - 17.2 Workshop for Policymakers in National Education
Systems 255
 - 17.2.1 Workshop Rationale and Content 256
 - 17.2.2 Workshop Schedule 258
 - 17.2.3 Group Work Products 258
 - 17.2.4 Workshop Wrap-Up 261
 - 17.3 Conclusion 262
 - References 262

- 18 Data Science Teacher Preparation: The “Method for Teaching
Data Science” Course** 265
 - 18.1 Introduction 265
 - 18.2 The MTDS Course Environment 267
 - 18.3 The MTDS Course Design 267
 - 18.4 The Learning Targets and Structure of the MTDS Course 268
 - 18.5 Grading Policy and Submissions 269
 - 18.6 Teaching Principles of the MTDS Course 271

- 18.7 Lesson Descriptions 274
 - 18.7.1 Lesson 6 274
 - 18.7.2 Mid-Semester Questionnaire 277
 - 18.7.3 Lesson 7 279
- 18.8 Conclusion 280
- References 281
- 19 Data Science for Social Science and Digital Humanities**
- Research** 283
- 19.1 Introduction 283
- 19.2 Relevance of Data Science for Social Science and Digital Humanities Researchers 284
- 19.3 Data Science Bootcamps for Researchers in Social Sciences and Digital Humanities 286
 - 19.3.1 Applicants and Participants of Two 2020 Bootcamps for Researchers in Social Sciences and Digital Humanities 286
 - 19.3.2 The Design and Curriculum of the Data Science for Social Science and Digital Humanities Researchers Bootcamp 290
- 19.4 Data Science for Psychological Sciences 292
 - 19.4.1 The Computer Science for Psychological Science Course 293
 - 19.4.2 The Data Science for Psychology Science Course 295
- 19.5 Data Science for Social Sciences and Digital Humanities, from a Motivation Theory Perspective 296
 - 19.5.1 The Self-determination Theory 296
 - 19.5.2 Gender Perspective 297
- 19.6 Conclusion 299
- References 300
- 20 Data Science for Research on Human Aspects of Science and Engineering** 303
- 20.1 Introduction 303
- 20.2 Examples of Research Topics Related to Human Aspects of Science and Engineering that Can Use Data Science Methods 304
- 20.3 Workshop on Data Science Research on Human Aspects of Science and Engineering 306
 - 20.3.1 Workshop Rationale 307
 - 20.3.2 Workshop Contents 307
 - 20.3.3 Target Audience 307
 - 20.3.4 Workshop Framework (in Terms of weeks)—A Proposal 307
 - 20.3.5 Prerequisites 308
 - 20.3.6 Workshop Requirements and Assessment 308

20.3.7	Workshop Schedule and Detailed Contents	309
20.3.8	Literature (For the Workshop)	312
20.4	Conclusion	313
Epilogue	315
Index	317

List of Figures

Fig. 2.1	Map by John Snow showing the clusters of cholera cases in the London epidemic of 1854 (Source: https://en.wikipedia.org/wiki/John_Snow , image is public domain)	25
Fig. 2.2	The authors' version of the data science Venn diagram, as inspired by Conway (2010)	26
Fig. 2.3	Data science workflow (Authors' version)	28
Fig. 2.4	The data life cycle (Berman et al., 2016)	29
Fig. 3.1	Data science and data thinking	37
Fig. 3.2	Traffic light classification explanations: distribution of answer categories ($n = 98$)	48
Fig. 3.3	Carcinoma classification explanations: distribution of answer categories ($n = 88$)	48
Fig. 3.4	Traffic light question explanation types versus carcinoma classification question explanation types ($n = 88$)	50
Fig. 5.1	Proposed range of possibilities for data integration in STEM education (Mike & Hazzan, 2022a)	78
Fig. 6.1	The data science Venn diagram, as proposed by Conway (2010)	86
Fig. 6.2	The data science Venn diagram, as proposed by the authors of the guide to teaching data science	87
Fig. 6.3	Pedagogical content knowledge (PCK)	91
Fig. 8.1	Dr. Inbal Zafir-Lavie's slide on the knowledge required from life science graduates and from data science graduates (presented here with permission)	129
Fig. 9.1	Diffusion of innovation timeline (Rogers, 1962)	139
Fig. 10.1	The CRISP-DM workflow (based on Shearer, 2000)	152
Fig. 10.2	The agile data science workflow (based on Pfister et al., 2015)	153
Fig. 14.1	A general scheme of the ML model generation, testing, and prediction phases	212

Fig. 14.2 Presentation of the lion detection question using a confusion matrix 216

Fig. 14.3 Bias and variance in dart throwing (based on Domingos, 2012) 218

Fig. 14.4 Model complexity in a classification problem 219

Fig. 14.5 Model complexity in a regression problem 220

Fig. 14.6 Overfitting and underfitting—irrigation system analogy 221

Fig. 15.1 The classification problem of *Iris virginica* versus *Iris versicolor* 227

Fig. 15.2 KNN hyperparameter tuning: accuracy versus K 228

Fig. 15.3 The underfitting and overfitting phenomena in KNN 228

Fig. 15.4 Decision tree model for the classification of *Iris* flowers 229

Fig. 19.1 Bootcamp applicants’ research disciplines by gender 287

Fig. 19.2 Bootcamp applicants’ academic rank by gender 288

Fig. 19.3 Participants’ pre-bootcamp knowledge in programming, statistics and machine learning (on a 1–5 scale, $n = 38$) 289

Fig. 19.4 Participants’ interest in the different bootcamp topics (on a 1–5 scale, $n = 38$) 289

Fig. 19.5 Applicants’ computer science and statistics knowledge vs. the domain knowledge 290

List of Tables

Table 1.1	The five parts of the guide to teaching data science	5
Table 3.1	The traffic light classification question	45
Table 3.2	The carcinoma classification question	45
Table 3.3	The context and social consideration of the experiment questions	46
Table 3.4	The familiarity-driven domain neglect	49
Table 4.1	Initiatives of undergraduate data science curricula	61
Table 6.1	Questions about the relationships between data science, computer science, and statistics	88
Table 6.2	Students’ perceptions of the mutual relationships between data science and computer science and between data science and statistics	89
Table 6.3	Intervention program in biomedical signal processing project	95
Table 6.4	Students’ perception of the knowledge required to meet the project goal	96
Table 6.5	Students’ perception of the research project’s success factors	96
Table 7.1	Data science for high school curriculum—topics and number of hours	106
Table 7.2	Machine learning algorithms—the mathematical knowledge required to understand them and an alternative intuitive explanation	107
Table 8.1	Mapping data science research skills	122
Table 9.1	The adoption process of the 10th grade data science curriculum	143
Table 9.2	The schedule of the first data science teachers’ training course (teaching methods were added to the second course)	144
Table 11.1	Mapping data science skills	166
Table 12.1	Analysis of data science codes of ethics	182

Table 12.2	Ethical norms of different stakeholder in the different phases of the data science workflow	183
Table 15.1	The KNN algorithm	226
Table 15.2	Vector representation of the perceptron classification algorithm	231
Table 15.3	Non-vectorized representation of the perceptron classification algorithm	231
Table 15.4	The perceptron training algorithm	231
Table 15.5	Logistic regression classification algorithm	233
Table 15.6	Logistic loss function	233
Table 15.7	Logistic regression training algorithm	233
Table 16.1	KNN worksheet	238
Table 16.2	Perceptron worksheet	241
Table 16.3	The gradient descent algorithm worksheet	244
Table 16.4	Neural networks worksheet	247
Table 16.5	Worksheet—programming tasks on the distance function	248
Table 17.1	Topics for teamwork in the workshop for policymakers in the education system	259
Table 17.2	SWOT analysis of the integration of data science in the Israeli education system	260
Table 18.1	The course assignments	270
Table 18.2	List of asynchronous tasks	272
Table 18.3	Course schedule	275
Table 19.1	Bootcamp topics and hours	291
Table 19.2	Computer science for psychological science—topics and number of hours	294
Table 19.3	Data science for psychological science course—topics and number of hours	295
Table 20.1	Science and engineering research: Scientific and engineering aspects, human aspects, and human-related research topics	305
Table 20.2	Workshop schedule and content	309

List of Exercises

Exercise 1.1	Reflection	14
Exercise 2.1	Pedagogical implications of data science as a science	23
Exercise 2.2	Pedagogical implications of data science as a research method	26
Exercise 2.3	Pedagogical implications of data science as a discipline	28
Exercise 2.4	Pedagogical implications of data science as a workflow	29
Exercise 2.5	Environmental and social aspects of the data life cycle ...	29
Exercise 2.6	Types of data scientists	30
Exercise 2.7	Categories of data science skills	30
Exercise 2.8	Data science as a discipline and data science as a profession	31
Exercise 2.9	Pedagogical implications of data science as a profession	31
Exercise 2.10	Skills of the data scientists	31
Exercise 2.11	Characteristics of data scientists on LinkedIn	31
Exercise 2.12	Connections between the different facets of data science	32
Exercise 2.13	Connections between the different facets of data science and the main characteristics of data science	32
Exercise 2.14	Pedagogical implications of the multi-faceted analysis of data science	32
Exercise 2.15	Overview of the history of data science	33
Exercise 3.1	Computational thinking and data science online course	39
Exercise 3.2	Statistical thinking for everyone	40
Exercise 3.3	Concept formulation processes, objects, and procepts	42
Exercise 3.4	Data science from the perspective of the process-object duality theory	42

Exercise 3.5 The process-object duality theory from the constructivist perspective 42

Exercise 3.6 Reducing abstraction and thinking on different levels of abstraction 43

Exercise 3.7 Describing data science concepts on different levels of abstraction 43

Exercise 3.8 Comparing the performance of machine learning algorithms 46

Exercise 3.9 Analysis of the familiarity-driven domain neglect 49

Exercise 3.10 Developing questions that demonstrate the domain neglect 50

Exercise 3.11 Outcomes of cognitive biases 50

Exercise 3.12 Teaching possible effects of cognitive biases on the interpretation of machine learning models 52

Exercise 3.13 Cases in which data thinking is important 53

Exercise 3.14 Application domain knowledge and real-life data in data thinking 54

Exercise 3.15 Data thinking and other cognitive aspects of data science 54

Exercise 3.16 Additional modes of thinking required for data science 54

Exercise 3.17 Analytical thinking 55

Exercise 4.1 Interdisciplinary education 60

Exercise 4.2 Pioneering data science programs 61

Exercise 4.3 Comparison of undergraduate data science curriculum initiatives 69

Exercise 4.4 Rating the interdisciplinarity of the reports 69

Exercise 4.5 Analysis of the reports from additional prisms 69

Exercise 4.6 Multidisciplinary, interdisciplinary, and transdisciplinary education 70

Exercise 4.7 Multidisciplinary, interdisciplinary, and transdisciplinary education in schools 70

Exercise 4.8 Didactic transposition in data science 70

Exercise 5.1 Teaching the STEM subjects in a real-world context 76

Exercise 5.2 Real data in mathematics and computer science courses 78

Exercise 5.3 Gender gaps in STEM subjects 79

Exercise 5.4 Twenty-first century skills in data science education 80

Exercise 5.5 Interdisciplinary pedagogy 81

Exercise 5.6 Professional development for teachers 81

Exercise 6.1 The interdisciplinary structure of data science 87

Exercise 6.2 Is data science more about computer science or more about statistics? 88

Exercise 6.3 Does data science include computer science? Does it include statistics? 89

Exercise 6.4	Designing an undergraduate data science program	90
Exercise 6.5	Analysis of undergraduate data science programs by components	90
Exercise 6.6	Data science PCK	92
Exercise 6.7	The challenge of developing interdisciplinary programs	92
Exercise 6.8	Knowledge gaps in PBL	96
Exercise 6.9	Additional challenges of mentoring interdisciplinary projects	97
Exercise 6.10	Revisiting the questions about the interdisciplinary structure of data science	97
Exercise 7.1	Different data science professions	102
Exercise 7.2	Diversity in data science	103
Exercise 7.3	The AI + Ethics Curriculum for Middle School initiative	104
Exercise 7.4	Initiatives for K-12 data science education	104
Exercise 7.5	Learning environments for high school pupils	104
Exercise 7.6	Real data in data science education	106
Exercise 7.7	Didactic transposition of machine learning algorithms	107
Exercise 7.8	Discussions about the role of data science in admission requirements	108
Exercise 7.9	Analysis of undergraduate data science programs	109
Exercise 7.10	Graduate studies	110
Exercise 7.11	Data science and diversity	111
Exercise 7.12	Online learning platforms versus physical campuses	113
Exercise 7.13	Online data science courses	113
Exercise 7.14	Profiles of online learners of data science courses	113
Exercise 7.15	Analysis of data gathered by online learning platforms	114
Exercise 7.16	Policy making based on data science methods	114
Exercise 7.17	Data science knowledge for users	115
Exercise 7.18	Data science knowledge for the general public	116
Exercise 7.19	Designing a data science course for the general public	116
Exercise 7.20	Books about data science for the general public	117
Exercise 7.21	Storytelling from the general public's perspective	117
Exercise 7.22	Categorization of learning environments for data science according to learner groups	117
Exercise 7.23	Textual programing environments for data science	118
Exercise 7.24	Visual programing environments for data science	118
Exercise 7.25	Use of general data processing tools for data science education purposes	118
Exercise 7.26	Teachable Machine	119
Exercise 8.1	Mapping data science skills	122
Exercise 8.2	Technion Biomedical Informatics Research and COVID-19	123
Exercise 8.3	Sub-cycles of the data science workflow	124

Exercise 8.4 The challenge of leaning the application domain 125

Exercise 8.5 The key role of understanding the application domain 126

Exercise 8.6 Populations for which research skills are relevant 127

Exercise 8.7 Model assessment 128

Exercise 8.8 A metaphor for a discourse between a biologist and
a data scientist 129

Exercise 8.9 Building a data science team 130

Exercise 8.10 The importance of the application domain for the
data scientist 130

Exercise 8.11 Data science and anthropology 131

Exercise 8.12 Visualization tools of data analysis environments 131

Exercise 8.13 Visualization in the data science workflow 132

Exercise 8.14 Visual programming environments for data science 132

Exercise 8.15 Connections between data visualization and other
data science skills 132

Exercise 8.16 Reflective and analysis skills 133

Exercise 8.17 Challenges of data science from a research
perspective 134

Exercise 8.18 Additional data science skills 134

Exercise 9.1 Diffusion of educational programs 138

Exercise 9.2 Reflection on your experience with the adoption of
innovation 139

Exercise 9.3 Reflection on your experience with the adoption of
educational innovation 142

Exercise 9.4 Pedagogical chasms 147

Exercise 10.1 The agile data science workflows 152

Exercise 10.2 Additional data science workflows 153

Exercise 10.3 Data science workflow for learners 154

Exercise 10.4 Data collection 155

Exercise 10.5 Search for a dataset 155

Exercise 10.6 Disadvantages of data gathering by learners 155

Exercise 10.7 Other tools for collecting data from people 155

Exercise 10.8 Activities included in data preparation 156

Exercise 10.9 Data wrangling 156

Exercise 10.10 Data preparation 157

Exercise 10.11 Types of visualization methods 158

Exercise 10.12 Selecting an appropriate visualization method 158

Exercise 10.13 Interpreting graphs 159

Exercise 10.14 Analysis of the exploratory data analysis phase from
the perspective of abstraction levels 159

Exercise 10.15 Feature engineering 161

Exercise 10.16 Communication and actions 162

Exercise 11.1 Mapping data science skills 166

Exercise 11.2 Critical thinking 166

Exercise 11.3 Reflection on dataset exploration 168

Exercise 11.4	Analysis of unknown dataset	170
Exercise 11.5	The rhetorical triangle	170
Exercise 11.6	Anthropology and data science	171
Exercise 11.7	Programming tools used in the data science workflow	172
Exercise 11.8	Lifelong learning	173
Exercise 11.9	Coursera 4-week (14 h) course Learning How to Learn	173
Exercise 11.10	A metaphor for a discourse between a data scientist, you, and another employee in your organization	174
Exercise 11.11	Giving and receiving feedback on a presentation about data analysis	174
Exercise 11.12	Exploration of a dataset as a debugging process	176
Exercise 11.13	The expression of skills in the data science workflow	176
Exercise 11.14	Skills of the data science workflow stakeholders	177
Exercise 11.15	Additional data science skills	177
Exercise 11.16	An interdisciplinary perspective on data science skills	177
Exercise 12.1	Famous cases that illustrate the need for an ethical code for data science	180
Exercise 12.2	Generative adversarial networks from an ethical perspective	181
Exercise 12.3	Codes of ethics	181
Exercise 12.4	Comparisons of data science codes of ethics	182
Exercise 12.5	Stakeholders' behavior in the different phases of the data science workflow	183
Exercise 12.6	Responsible AI	183
Exercise 12.7	Recommendations of curriculum guidelines with respect to the inclusion of ethics in data science programs	184
Exercise 12.8	Ethical principles to be applied in the creation of image sets	184
Exercise 12.9	The ethical aspect of product development	185
Exercise 12.10	Analysis of a documentary movie	185
Exercise 12.11	Embedded ethics in data science	187
Exercise 12.12	Revisiting the AI + Ethics curriculum for middle school initiative	188
Exercise 12.13	Exploration of previously published data science case studies	189
Exercise 12.14	Development of data science case studies	189
Exercise 12.15	Developing scenarios development about data science	190
Exercise 12.16	Reflection on a lecture	191
Exercise 12.17	Reflection as a habit of mind	192
Exercise 12.18	PBL and data science	193
Exercise 12.19	Social issues of data science and interdisciplinarity	193
Exercise 13.1	The concepts of explainability and interpretability	204
Exercise 13.2	Machine learning and statistical thinking	205

Exercise 13.3 Machine learning and interdisciplinary projects 206

Exercise 13.4 Machine learning performance indicators 206

Exercise 14.1 Types of machine learning 210

Exercise 14.2 Machine learning performance indicators 214

Exercise 14.3 True or false 216

Exercise 14.4 Performance indicators 217

Exercise 14.5 Comparing Performance indicators 217

Exercise 14.6 Overfitting and underfitting 221

Exercise 14.7 Overfitting and underfitting analogies 221

Exercise 14.8 Regularization analogies 223

Exercise 15.1 K as hyperparameter 229

Exercise 17.1 Decision making in governmental offices 254

Exercise 17.2 Data culture 255

Exercise 17.3 The roles of the workshop facilitator 258

Exercise 17.4 Culture 260

Exercise 17.5 Working on the group task 261

Exercise 17.6 Data science in the service of government offices 261

Exercise 17.7 A follow-up workshop for the “Data science for education policymaking, governance, and operations” workshop 261

Exercise 17.8 The pedagogical chasm and policymaking 262

Exercise 18.1 Facilitation of an asynchronous activity 269

Exercise 18.2 Topics to be included in a Method of Teaching Data Science course 273

Exercise 18.3 Interdisciplinarity of and in the MTDS course 274

Exercise 18.4 Classification of the KNN description into process and object conceptions 276

Exercise 18.5 Analysis of students’ responses to the mid-semester questionnaire 278

Exercise 19.1 Data science applications that require knowledge in social sciences and digital humanities 285

Exercise 19.2 Data science job ads that require knowledge in social sciences and digital humanities 286

Exercise 19.3 Interviewing researchers in social sciences and digital humanities 290

Exercise 19.4 Uses of supervised and unsupervised machine learning algorithms in social sciences applications 296

Exercise 19.5 Theoretical perspectives on data science education for researchers in social sciences and digital humanities 298

Exercise 19.6 Data science education for undergraduate social sciences and digital humanities students 299

Exercise 20.1 Examples of human aspects of engineering research 305

Exercise 20.2 Data-driven research 306

Exercise 20.3 Data visualization 306

Exercise EP.1	Pedagogical chasms	315
Exercise EP.2	Connections between chapters	316
Exercise EP.3	Final reflection task	316