

Preface	19
Before You Begin	45
I Introduction to Computers and Python	49
1.1 Introduction	50
1.2 Hardware and Software	51
1.2.1 Moore's Law	52
1.2.2 Computer Organization	52
1.3 Data Hierarchy	54
1.4 Machine Languages, Assembly Languages and High-Level Languages	57
1.5 Introduction to Object Technology	58
1.6 Operating Systems	61
1.7 Python	64
1.8 It's the Libraries!	66
1.8.1 Python Standard Library	66
1.8.2 Data-Science Libraries	66
1.9 Other Popular Programming Languages	68
1.10 Test-Drives: Using IPython and Jupyter Notebooks	69
1.10.1 Using IPython Interactive Mode as a Calculator	69
1.10.2 Executing a Python Program Using the IPython Interpreter	71
1.10.3 Writing and Executing Code in a Jupyter Notebook	72
1.11 Internet and World Wide Web	77
1.11.1 Internet: A Network of Networks	77
1.11.2 World Wide Web: Making the Internet User-Friendly	78
1.11.3 The Cloud	78
1.11.4 Internet of Things	79
1.12 Software Technologies	80
1.13 How Big Is Big Data?	81
1.13.1 Big Data Analytics	86
1.13.2 Data Science and Big Data Are Making a Difference: Use Cases	87
1.14 Case Study—A Big-Data Mobile Application	88
1.15 Intro to Data Science: Artificial Intelligence—at the Intersection of CS and Data Science	90

2	Introduction to Python Programming	97
2.1	Introduction	98
2.2	Variables and Assignment Statements	100
2.3	Arithmetic	104
2.4	Function and an Intro to Single- and Double-Quoted Strings	106
2.5	Triple-Quoted Strings	107
2.6	Getting Input from the User	109
2.7	Decision Making: The Statement and Comparison Operators	114
2.8	Objects and Dynamic Typing	116
2.9	Intro to Data Science: Basic Descriptive Statistics	118
2.10	Wrap-Up	
3	Control Statements and Program Development	121
3.1	Introduction	122
3.2	Algorithms	123
3.3	Pseudocode	123
3.4	Control Statements	126
3.5	Statement	128
3.6	... and ... Statements	133
3.7	Statement	134
3.8	Statement	136
3.8.1	Iterables, Lists and Iterators	136
3.8.2	Built-In Function	137
3.9	Augmented Assignments	138
3.10	Program Development: Sequence-Controlled Repetition	138
3.10.1	Requirements Statement	138
3.10.2	Pseudocode for the Algorithm	139
3.10.3	Coding the Algorithm in Python	140
3.10.4	Introduction to Formatted Strings	141
3.11	Program Development: Sentinel-Controlled Repetition	145
3.12	Program Development: Nested Control Statements	149
3.13	Built-In Function : A Deeper Look	150
3.14	Using Type for Monetary Amounts	153
3.15	and Statements	154
3.16	Boolean Operators , and	157
3.17	Intro to Data Science: Measures of Central Tendency— Mean, Median and Mode	159
3.18	Wrap-Up	
4	Functions	167
4.1	Introduction	168
4.2	Defining Functions	171
4.3	Functions with Multiple Parameters	173
4.4	Random-Number Generation	

	15
	17
4.5 Case Study: A Game of Chance	18
4.6 Python Standard Library	18
4.7 Module Functions	18
4.8 Using IPython Tab Completion for Discovery	18
4.9 Default Parameter Values	18
4.10 Keyword Arguments	18
4.11 Arbitrary Argument Lists	18
4.12 Methods: Functions That Belong to Objects	18
4.13 Scope Rules	18
4.14 : A Deeper Look	19
4.15 Passing Arguments to Functions: A Deeper Look	19
4.16 Function-Call Stack	19
4.17 Functional-Style Programming	19
4.18 Intro to Data Science: Measures of Dispersion	19
4.19 Wrap-Up	19
5 Sequences: Lists and Tuples	203
5.1 Introduction	204
5.2 Lists	204
5.3 Tuples	209
5.4 Unpacking Sequences	211
5.5 Sequence Slicing	214
5.6 Statement	217
5.7 Passing Lists to Functions	219
5.8 Sorting Lists	220
5.9 Searching Sequences	222
5.10 Other List Methods	224
5.11 Simulating Stacks with Lists	226
5.12 List Comprehensions	227
5.13 Generator Expressions	229
5.14 Filter, Map and Reduce	230
5.15 Other Sequence Processing Functions	233
5.16 Two-Dimensional Lists	235
5.17 Intro to Data Science: Simulation and Static Visualizations	239
5.17.1 Sample Graphs for 600, 60,000 and 6,000,000 Die Rolls	239
5.17.2 Visualizing Die-Roll Frequencies and Percentages	241
5.18 Wrap-Up	247
6 Dictionaries and Sets	257
6.1 Introduction	258
6.2 Dictionaries	258
6.2.1 Creating a Dictionary	258
6.2.2 Iterating through a Dictionary	260
6.2.3 Basic Dictionary Operations	260

6.2.4	Dictionary Methods	and	262
6.2.5	Dictionary Comparisons		264
6.2.6	Example: Dictionary of Student Grades		265
6.2.7	Example: Word Counts		266
6.2.8	Dictionary Method		268
6.2.9	Dictionary Comprehensions		268
6.3	Sets		269
6.3.1	Comparing Sets		271
6.3.2	Mathematical Set Operations		273
6.3.3	Mutable Set Operators and Methods		274
6.3.4	Set Comprehensions		276
6.4	Intro to Data Science: Dynamic Visualizations		276
6.4.1	How Dynamic Visualization Works		276
6.4.2	Implementing a Dynamic Visualization		279
6.5	Wrap-Up		282

7 Array-Oriented Programming with NumPy 287

7.1	Introduction		288
7.2	Creating s from Existing Data		289
7.3	Attributes		290
7.4	Filling s with Specific Values		292
7.5	Creating s from Ranges		292
7.6	List vs. Performance: Introducing		294
7.7	Operators		296
7.8	NumPy Calculation Methods		298
7.9	Universal Functions		300
7.10	Indexing and Slicing		302
7.11	Views: Shallow Copies		304
7.12	Deep Copies		306
7.13	Reshaping and Transposing		307
7.14	Intro to Data Science: pandas and s		310
7.14.1	pandas		310
7.14.2	s		315
7.15	Wrap-Up		323

8 Strings: A Deeper Look 331

8.1	Introduction		332
8.2	Formatting Strings		333
8.2.1	Presentation Types		333
8.2.2	Field Widths and Alignment		334
8.2.3	Numeric Formatting		335
8.2.4	String's Method		336
8.3	Concatenating and Repeating Strings		337
8.4	Stripping Whitespace from Strings		338

8.5	Changing Character Case	339
8.6	Comparison Operators for Strings	340
8.7	Searching for Substrings	340
8.8	Replacing Substrings	342
8.9	Splitting and Joining Strings	342
8.10	Characters and Character-Testing Methods	345
8.11	Raw Strings	346
8.12	Introduction to Regular Expressions	347
8.12.1	Module and Function	348
8.12.2	Replacing Substrings and Splitting Strings	351
8.12.3	Other Search Functions; Accessing Matches	352
8.13	Intro to Data Science: Pandas, Regular Expressions and Data Munging	355
8.14	Wrap-Up	360

9 Files and Exceptions 367

9.1	Introduction	368
9.2	Files	369
9.3	Text-File Processing	369
9.3.1	Writing to a Text File: Introducing the Statement	370
9.3.2	Reading Data from a Text File	371
9.4	Updating Text Files	373
9.5	Serialization with JSON	375
9.6	Focus on Security: Serialization and Deserialization	378
9.7	Additional Notes Regarding Files	378
9.8	Handling Exceptions	379
9.8.1	Division by Zero and Invalid Input	380
9.8.2	Statements	380
9.8.3	Catching Multiple Exceptions in One Clause	383
9.8.4	What Exceptions Does a Function or Method Raise?	384
9.8.5	What Code Should Be Placed in a Suite?	384
9.9	Clause	384
9.10	Explicitly Raising an Exception	387
9.11	(Optional) Stack Unwinding and Tracebacks	387
9.12	Intro to Data Science: Working with CSV Files	390
9.12.1	Python Standard Library Module	390
9.12.2	Reading CSV Files into Pandas	392
9.12.3	Reading the Titanic Disaster Dataset	394
9.12.4	Simple Data Analysis with the Titanic Disaster Dataset	395
9.12.5	Passenger Age Histogram	396
9.13	Wrap-Up	397

403

10 Object-Oriented Programming

10.1	Introduction	404
10.2	Custom Class	406

10.2.1	Test-Driving Class	406
10.2.2	Class Definition	408
10.2.3	Composition: Object References as Members of Classes	409
10.3	Controlling Access to Attributes	411
10.4	Properties for Data Access	412
10.4.1	Test-Driving Class	414
10.4.2	Class Definition	418
10.4.3	Class Definition Design Notes	419
10.5	Simulating "Private" Attributes	421
10.6	Case Study: Card Shuffling and Dealing Simulation	421
10.6.1	Test-Driving Classes and	423
10.6.2	Class —Introducing Class Attributes	425
10.6.3	Class	426
10.6.4	Displaying Card Images with Matplotlib	430
10.7	Inheritance: Base Classes and Subclasses	432
10.8	Building an Inheritance Hierarchy; Introducing Polymorphism	432
10.8.1	Base Class	435
10.8.2	Subclass	
10.8.3	Processing s and s Polymorphically	439
10.8.4	A Note About Object-Based and Object-Oriented Programming	439
10.9	Duck Typing and Polymorphism	440
10.10	Operator Overloading	441
10.10.1	Test-Driving Class	442
10.10.2	Class Definition	443
10.11	Exception Class Hierarchy and Custom Exceptions	445
10.12	Named Tuples	447
10.13	A Brief Intro to Python 3.7's New Data Classes	448
10.13.1	Creating a Data Class	449
10.13.2	Using the Data Class	451
10.13.3	Data Class Advantages over Named Tuples	453
10.13.4	Data Class Advantages over Traditional Classes	454
10.14	Unit Testing with Docstrings and	454
10.15	Namespaces and Scopes	459
10.16	Intro to Data Science: Time Series and Simple Linear Regression	462
10.17	Wrap-Up	471

II Computer Science Thinking: Recursion, Searching, Sorting and Big O 479

11.1	Introduction	480
11.2	Factorials	481
11.3	Recursive Factorial Example	481
11.4	Recursive Fibonacci Series Example	484
11.5	Recursion vs. Iteration	487
11.6	Searching and Sorting	488

11.7	Linear Search	488
11.8	Efficiency of Algorithms: Big O	490
11.9	Binary Search	492
11.9.1	Binary Search Implementation	493
11.9.2	Big O of the Binary Search	495
11.10	Sorting Algorithms	496
11.11	Selection Sort	496
11.11.1	Selection Sort Implementation	497
11.11.2	Utility Function	498
11.11.3	Big O of the Selection Sort	499
11.12	Insertion Sort	499
11.12.1	Insertion Sort Implementation	500
11.12.2	Big O of the Insertion Sort	501
11.13	Merge Sort	502
11.13.1	Merge Sort Implementation	502
11.13.2	Big O of the Merge Sort	507
11.14	Big O Summary for This Chapter's Searching and Sorting Algorithms	507
11.15	Visualizing Algorithms	508
11.15.1	Generator Functions	510
11.15.2	Implementing the Selection Sort Animation	511
11.16	Wrap-Up	516

12 Natural Language Processing (NLP) **525**

12.1	Introduction	526
12.2	TextBlob	527
12.2.1	Create a TextBlob	529
12.2.2	Tokenizing Text into Sentences and Words	530
12.2.3	Parts-of-Speech Tagging	530
12.2.4	Extracting Noun Phrases	531
12.2.5	Sentiment Analysis with TextBlob's Default Sentiment Analyzer	532
12.2.6	Sentiment Analysis with the	534
12.2.7	Language Detection and Translation	535
12.2.8	Inflection: Pluralization and Singularization	537
12.2.9	Spell Checking and Correction	537
12.2.10	Normalization: Stemming and Lemmatization	538
12.2.11	Word Frequencies	539
12.2.12	Getting Definitions, Synonyms and Antonyms from WordNet	540
12.2.13	Deleting Stop Words	542
12.2.14	n-grams	544
12.3	Visualizing Word Frequencies with Bar Charts and Word Clouds	545
12.3.1	Visualizing Word Frequencies with Pandas	548
12.3.2	Visualizing Word Frequencies with Word Clouds	551
12.4	Readability Assessment with Textstatistic	553
12.5	Named Entity Recognition with spaCy	555
12.6	Similarity Detection with spaCy	

12.7	Other NLP Libraries and Tools	557
12.8	Machine Learning and Deep Learning Natural Language Applications	557
12.9	Natural Language Datasets	558
12.10	Wrap-Up	558

13 Data Mining Twitter 563

13.1	Introduction	564
13.2	Overview of the Twitter APIs	566
13.3	Creating a Twitter Account	567
13.4	Getting Twitter Credentials—Creating an App	568
13.5	What's in a Tweet?	569
13.6	Tweepy	573
13.7	Authenticating with Twitter Via Tweepy	573
13.8	Getting Information About a Twitter Account	575
13.9	Introduction to Tweepy Cursors: Getting an Account's Followers and Friends	577
13.9.1	Determining an Account's Followers	577
13.9.2	Determining Whom an Account Follows	580
13.9.3	Getting a User's Recent Tweets	580
13.10	Searching Recent Tweets	582
13.11	Spotting Trends: Twitter Trends API	584
13.11.1	Places with Trending Topics	584
13.11.2	Getting a List of Trending Topics	585
13.11.3	Create a Word Cloud from Trending Topics	587
13.12	Cleaning/Preprocessing Tweets for Analysis	589
13.13	Twitter Streaming API	590
13.13.1	Creating a Subclass of	591
13.13.2	Initiating Stream Processing	593
13.14	Tweet Sentiment Analysis	595
13.15	Geocoding and Mapping	599
13.15.1	Getting and Mapping the Tweets	600
13.15.2	Utility Functions in	604
13.15.3	Class	606
13.16	Ways to Store Tweets	607
13.17	Twitter and Time Series	608
13.18	Wrap-Up	608

14 IBM Watson and Cognitive Computing 613

14.1	Introduction: IBM Watson and Cognitive Computing	614
14.2	IBM Cloud Account and Cloud Console	616
14.3	Watson Services	616
14.4	Additional Services and Tools	620
14.5	Watson Developer Cloud Python SDK	621
14.6	Case Study: Traveler's Companion Translation App	622

14.6.1	Before You Run the App	623
14.6.2	Test-Driving the App	624
14.6.3	Script Walkthrough	625
14.7	Watson Resources	635
14.8	Wrap-Up	637

15 Machine Learning: Classification, Regression and Clustering

641

15.1	Introduction to Machine Learning	642
15.1.1	Scikit-Learn	643
15.1.2	Types of Machine Learning	644
15.1.3	Datasets Bundled with Scikit-Learn	646
15.1.4	Steps in a Typical Data Science Study	647
15.2	Case Study: Classification with k-Nearest Neighbors and the Digits Dataset, Part 1	647
15.2.1	k-Nearest Neighbors Algorithm	649
15.2.2	Loading the Dataset	650
15.2.3	Visualizing the Data	654
15.2.4	Splitting the Data for Training and Testing	656
15.2.5	Creating the Model	657
15.2.6	Training the Model	658
15.2.7	Predicting Digit Classes	658
15.3	Case Study: Classification with k-Nearest Neighbors and the Digits Dataset, Part 2	660
15.3.1	Metrics for Model Accuracy	660
15.3.2	K-Fold Cross-Validation	664
15.3.3	Running Multiple Models to Find the Best One	665
15.3.4	Hyperparameter Tuning	667
15.4	Case Study: Time Series and Simple Linear Regression	668
15.5	Case Study: Multiple Linear Regression with the California Housing Dataset	673
15.5.1	Loading the Dataset	674
15.5.2	Exploring the Data with Pandas	676
15.5.3	Visualizing the Features	678
15.5.4	Splitting the Data for Training and Testing	682
15.5.5	Training the Model	682
15.5.6	Testing the Model	683
15.5.7	Visualizing the Expected vs. Predicted Prices	684
15.5.8	Regression Model Metrics	685
15.5.9	Choosing the Best Model	686
15.6	Case Study: Unsupervised Machine Learning, Part 1—Dimensionality Reduction	687
15.7	Case Study: Unsupervised Machine Learning, Part 2—k-Means Clustering	690
15.7.1	Loading the Iris Dataset	692

15.7.2	Exploring the Iris Dataset: Descriptive Statistics with Pandas	694
15.7.3	Visualizing the Dataset with a Seaborn	695
15.7.4	Using a Estimator	698
15.7.5	Dimensionality Reduction with Principal Component Analysis	700
15.7.6	Choosing the Best Clustering Estimator	703
15.8	Wrap-Up	704

16 Deep Learning 713

16.1	Introduction	714
16.1.1	Deep Learning Applications	716
16.1.2	Deep Learning Demos	717
16.1.3	Keras Resources	717
16.2	Keras Built-In Datasets	718
16.3	Custom Anaconda Environments	720
16.4	Neural Networks	722
16.5	Tensors	722
16.6	Convolutional Neural Networks for Vision; Multi-Classification with the MNIST Dataset	724
16.6.1	Loading the MNIST Dataset	725
16.6.2	Data Exploration	726
16.6.3	Data Preparation	728
16.6.4	Creating the Neural Network	730
16.6.5	Training and Evaluating the Model	739
16.6.6	Saving and Loading a Model	744
16.7	Visualizing Neural Network Training with TensorBoard	745
16.8	ConvnetJS: Browser-Based Deep-Learning Training and Visualization	748
16.9	Recurrent Neural Networks for Sequences; Sentiment Analysis with the IMDb Dataset	749
16.9.1	Loading the IMDb Movie Reviews Dataset	750
16.9.2	Data Exploration	751
16.9.3	Data Preparation	753
16.9.4	Creating the Neural Network	754
16.9.5	Training and Evaluating the Model	757
16.10	Tuning Deep Learning Models	758
16.11	Convnet Models Pretrained on ImageNet	759
16.12	Reinforcement Learning	760
16.12.1	Deep Q-Learning	761
16.12.2	OpenAI Gym	761
16.13	Wrap-Up	762

17 Big Data: Hadoop, Spark, NoSQL and IoT 771

17.1	Introduction	772
17.2	Relational Databases and Structured Query Language (SQL)	776
17.2.1	A Database	778

17.2.2	Queries	782
17.2.3	Clause	782
17.2.4	Clause	784
17.2.5	Merging Data from Multiple Tables:	785
17.2.6	Statement	786
17.2.7	Statement	787
17.2.8	Statement	787
17.3	NoSQL and NewSQL Big-Data Databases: A Brief Tour	789
17.3.1	NoSQL Key-Value Databases	789
17.3.2	NoSQL Document Databases	790
17.3.3	NoSQL Columnar Databases	790
17.3.4	NoSQL Graph Databases	791
17.3.5	NewSQL Databases	791
17.4	Case Study: A MongoDB JSON Document Database	792
17.4.1	Creating the MongoDB Atlas Cluster	793
17.4.2	Streaming Tweets into MongoDB	794
17.5	Hadoop	803
17.5.1	Hadoop Overview	803
17.5.2	Summarizing Word Lengths in <i>Romeo and Juliet</i> via MapReduce	806
17.5.3	Creating an Apache Hadoop Cluster in Microsoft Azure HDInsight	806
17.5.4	Hadoop Streaming	808
17.5.5	Implementing the Mapper	808
17.5.6	Implementing the Reducer	809
17.5.7	Preparing to Run the MapReduce Example	810
17.5.8	Running the MapReduce Job	811
17.6	Spark	814
17.6.1	Spark Overview	814
17.6.2	Docker and the Jupyter Docker Stacks	815
17.6.3	Word Count with Spark	818
17.6.4	Spark Word Count on Microsoft Azure	821
17.7	Spark Streaming: Counting Twitter Hashtags Using the Docker Stack	825
17.7.1	Streaming Tweets to a Socket	825
17.7.2	Summarizing Tweet Hashtags: Introducing Spark SQL	828
17.8	Internet of Things and Dashboards	834
17.8.1	Publish and Subscribe	836
17.8.2	Visualizing a PubNub Sample Live Stream with a Freeboard Dashboard	836
17.8.3	Simulating an Internet-Connected Thermostat in Python	840
17.8.4	Creating the Dashboard with Freeboard.io	842
17.8.5	Creating a Python PubNub Subscriber	846
17.9	Wrap-Up	