



Ebook Directory
the best source of ebook

The book was found

Python Data Science Handbook: Essential Tools For Working With Data



Synopsis

For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—Python, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you'll learn how to use: IPython and Jupyter: provide computational environments for data scientists using Python NumPy: includes the ndarray for efficient storage and manipulation of dense data arrays in Python Pandas: features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python Matplotlib: includes capabilities for a flexible range of data visualizations in Python Scikit-Learn: for efficient and clean Python implementations of the most important and established machine learning algorithms

Book Information

Paperback: 548 pages

Publisher: O'Reilly Media; 1 edition (December 10, 2016)

Language: English

ISBN-10: 1491912057

ISBN-13: 978-1491912058

Product Dimensions: 7 x 1 x 9.1 inches

Shipping Weight: 1.8 pounds (View shipping rates and policies)

Average Customer Review: 4.3 out of 5 stars 27 customer reviews

Best Sellers Rank: #10,380 in Books (See Top 100 in Books) #6 in Books > Computers & Technology > Databases & Big Data > Data Modeling & Design #10 in Books > Computers & Technology > Databases & Big Data > Data Processing #15 in Books > Science & Math > Experiments, Instruments & Measurement > Methodology & Statistics

Customer Reviews

Jake VanderPlas is a long-time user and developer of the Python scientific stack. He currently works as an interdisciplinary research director at the University of Washington, conducts his own astronomy research, and spends time advising and consulting with local scientists from a wide

range of fields.

Awesome book.

The figures were generated in color, but printed black and white, so they are often unintelligible. It's hard to tell the red dots from the blue when they are both grey. Apart from that major oversight, the book is ok. If you want to learn data science, this is not for you; it doesn't get into the fundamentals much at all. If you are an experienced R user looking for how to translate into python, this will get you started. The rest of my review comes from this perspective. The book spends far too much time on low-level ipython, numpy, and matplotlib functionality (chapters 1, 2, and 4). You are rarely going to use this stuff. The pandas section (chapter 3) is fine, but I was a little disappointed in the treatment of the grouping/aggregation functions. The book mentions the split-apply-combine paradigm of Hadley Wickham, but doesn't cover the topic in nearly as much detail as the paper of the same name. I was hoping to learn how to translate the dplyr verbs (`group_by`, `filter`, `select`, `mutate`, `summarize`, `arrange`) into pandas, but this book doesn't provide that. You will learn the basics of grouping and aggregation, but your code is going to be a lot more verbose than it was in R. The machine learning case studies in chapter 5 are pretty nice - probably the only reason I would recommend this book. The chapter provides a good overview of the scikit-learn API and effective patterns for machine learning problems.

I have been reading and recommending this book from the early-release stage. The first half of the book is dedicated to introduce the basic Python libraries for data analysis (and scientific computing in general). The second part deals with Machine Learning from a practical point of view using the Python library scikit-learn. The book is ideally suited to those that already know the basic Python stuff (or know how to program in a language like R/Julia, etc) and want to learn how to use Python for data-analysis. Even if you already know Python and how to use it for data analysis you could still find some gems here and there in the form of very clear examples or comments. In summary, if you want to do data-analysis and you already know Python read this book, if you do not know Python read Think Python by Allen B. Downey first and then read this book.

I am currently taking a Machine Learning course from Udacity and this book has proven to be a great reference guide for several projects and quizzes. Although it does not go in depth in regards to machine learning (although almost half of the book is dedicated to it), it does give an understanding

of essential concepts. For those interested in machine learning I would recommend bying "Hands-On Machine Learning with Scikit-Learn and TensorFlow" by Geron as well as this book. There is no one book for data science, and this one is no exception. Just keep that in mind before buying it. Other than that, I am really happy with my purchase. P.S. For those complaining about black and white graphs and diagrams - check the author's GitHub.

This book is not bad, but not great either in my opinion. The author obviously knows his stuff and covers the material in an accurate and competent manner. However there are a couple of flaws in my opinion. Firstly this book is quite dull, I've read other data science books that are more engaging. For example he might discuss matplotlib and go exhaustively through many options. That is fine, but boring, in my opinion it might be better to do things in not quite as much depth but more in the context of a realistic analysis. Secondly this book can't decide if it is a reference or a tutorial. The author gets a bit carried away showing too many features, I often found myself nodding off or losing my concentration. With so many online references, it might be better to concentrate on being a tutorial and not try to show so many features. Or perhaps separate each chapter into a tutorial and then a reference. On the other hand I realize that some readers might want this extra depth, so I'm just saying what I personally would have preferred. A related problem is that the material can quickly go out of date, I already found some options to be deprecated when running code. Thirdly, I question some of the organization of material, he often introduces some aspect, doesn't explain it properly, and then returns to it later on to explain it in more depth. An example is the Scikit-learn pipeline object, he starts using this leaving me puzzled and only later returned to explain it. This kind of issue was relatively common. In conclusion this is a decent book and certainly not a bad book, but more suited for particular audiences. This book would be good for those looking for a reference and relatively detailed information on a particular topic. Not so suited for beginners, I think they would be confused, or overwhelmed. People with some experience who are looking for more of a tutorial could be bored. It is worth noting as well that there are so many python data science books, but nearly all of them not very good. Relative to the other books, this is probably one of the best. In contrast there are a number of excellent books that use R.

This is a fantastic book. You definitely want to read it with one hand on your keyboard, following along with the examples.

When I first received this book, I was surprised that it didn't get to scikit-learn until the last third of

the book. The first third is about numpy and pandas, and the middle third is about matplotlib. Now that I've been applying it at work, however, I've found that the items covered in the first two thirds were really essential. I wouldn't be nearly as productive if I had just jumped straight to the sections on scikit-learn. The author does an excellent job covering broad terrain with enough detail that you are able to apply it to your problems. You will find yourself going back to use this book as a reference.

[Download to continue reading...](#)

Python: The Complete Python Quickstart Guide (For Beginner's) (Python, Python Programming, Python for Dummies, Python for Beginners) Python: Programming: Your Step By Step Guide To Easily Learn Python in 7 Days (Python for Beginners, Python Programming for Beginners, Learn Python, Python Language) Hacking with Python: Beginner's Guide to Ethical Hacking, Basic Security, Penetration Testing, and Python Hacking (Python Programming, Hacking, Python Coding, Python and Hacking Book 3) PYTHON: PYTHON'S COMPANION, A STEP BY STEP GUIDE FOR BEGINNERS TO START CODING TODAY! (INCLUDES A 6 PAGE PRINTABLE CHEAT SHEET)(PYTHON FOR BEGINNERS, PYTHON FOR DUMMIES, PYTHON PROGRAMMING) Python Data Science Handbook: Essential Tools for Working with Data PYTHON: LEARN PYTHON in A Day and MASTER IT WELL. The Only Essential Book You Need To Start Programming in Python Now. Hands On Challenges INCLUDED! (Programming for Beginners, Python) Python Programming: Python Programming for Beginners, Python Programming for Intermediates, Python Programming for Advanced Data Analytics and Python Programming: 2 Bundle Manuscript: Beginners Guide to Learn Data Analytics, Predictive Analytics and Data Science with Python Programming Python: Learn Python in a Day and Master It Well: The Only Essential Book You Need to Start Programming in Python Now Maya Python for Games and Film: A Complete Reference for Maya Python and the Maya Python API Python Programming: An In-Depth Guide Into The Essentials Of Python Programming (Included: 30+ Exercises To Master Python in No Time!) Python: The Fundamentals Of Python Programming: A Complete Beginners Guide To Python Mastery. Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data Book 1) Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics (Big Data Analytics, Data Science, Business Intelligence Book 6) Big Data For Business: Your Comprehensive Guide to Understand Data Science, Data Analytics and Data Mining to Boost More Growth and Improve Business - Data Analytics Book, Series 2 Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis,

Business Analytics, Data Mining, Big Data) Programming for Computations - Python: A Gentle Introduction to Numerical Simulations with Python (Texts in Computational Science and Engineering) Unsupervised Machine Learning in Python: How to Find Distinct Patterns in Your Data Without Being at the Mercy of Data Labeling by Third-Party Workers Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (Princeton Series in Modern Observational Astronomy) Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython

[Contact Us](#)

[DMCA](#)

[Privacy](#)

[FAQ & Help](#)