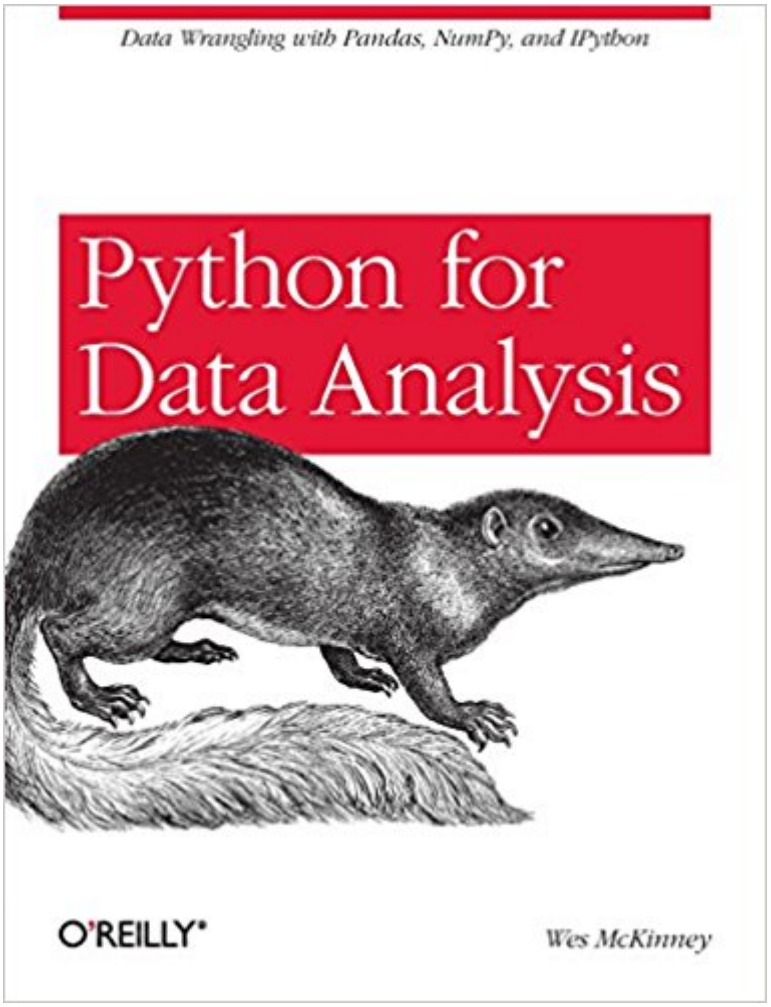
Python For Data Analysis: Data Wrangling With Pandas, NumPy, And IPython



Synopsis

Python for Data Analysis is concerned with the nuts and bolts of manipulating, processing, cleaning, and crunching data in Python. It is also a practical, modern introduction to scientific computing in Python, tailored for data-intensive applications. This is a book about the parts of the Python language and libraries youâ ™ll need to effectively solve a broad set of data analysis problems. This book is not an exposition on analytical methods using Python as the implementation language.Written by Wes McKinney, the main author of the pandas library, this hands-on book is packed with practical cases studies. Itâ ™s ideal for analysts new to Python and for Python programmers new to scientific computing.Use the IPython interactive shell as your primary development environmentLearn basic and advanced NumPy (Numerical Python) featuresGet started with data analysis tools in the pandas libraryUse high-performance tools to load, clean, transform, merge, and reshape dataCreate scatter plots and static or interactive visualizations with matplotlibApply the pandas groupby facility to slice, dice, and summarize datasetsMeasure data by points in time, whether itâ ™s specific instances, fixed periods, or intervalsLearn how to solve problems in web analytics, social sciences, finance, and economics, through detailed examples

Book Information

Paperback: 466 pages
Publisher: O'Reilly Media; 1 edition (November 1, 2012)
Language: English
ISBN-10: 1449319793
ISBN-13: 978-1449319793
Product Dimensions: 7 x 0.9 x 9.2 inches
Shipping Weight: 1.8 pounds (View shipping rates and policies)
Average Customer Review: 4.2 out of 5 stars 146 customer reviews
Best Sellers Rank: #4,995 in Books (See Top 100 in Books) #6 inÂ Books > Computers & Technology > Databases & Big Data > Data Processing #7 inÂ Books > Computers & Technology > Programming > Languages & Tools > Python #10 inÂ Books > Computers & Technology > Programming > Web Programming

Customer Reviews

View larger This is by no means a complete list. Even though it may not always be obvious, a large percentage of data sets can be transformed into a structured form that is more suitable for analysis and modeling. If not, it may be possible to extract features from a data set into a structured

form. As an example, a collection of news articles could be processed into a word frequency table which could then be used to perform sentiment analysis. Most users of spreadsheet programs like Microsoft Excel, perhaps the most widely used data analysis tool in the world, will not be strangers to these kinds of data. What Is This Book About? This book is concerned with the nuts and bolts of manipulating, processing, cleaning, and crunching data in Python. It is also a practical, modern introduction to scientific computing in Python, tailored for data-intensive applications. This is a book about the parts of the Python language and libraries youÃ¢Â Â™ll need to effectively solve a broad set of data analysis problems. This book is not an exposition on analytical methods using Python as the implementation language. When I say 'data', what am I referring to exactly? The primary focus is on structured data, a deliberately vague term that encompasses many different common forms of data, such as: Multidimensional arrays (matrices). Tabular or spreadsheet-like data in which each column may be a different type (string, numeric, date, or otherwise). This includes most kinds of data commonly stored in relational databases or tab- or comma-delimited text files. Multiple tables of data interrelated by key columns (what would be primary or foreign keys for a SQL user). Evenly or unevenly spaced time series.

Data Wrangling with Pandas, NumPy, and IPython

Wes McKinney is the main author of pandas, the popular open sourcePython library for data analysis. Wes is an active speaker andparticipant in the Python and open source communities. He worked as aquantitative analyst at AQR Capital Management and Python consultantbefore founding DataPad, a data analytics company, in 2013. Hegraduated from MIT with an S.B. in Mathematics.

IMHO, this is the most important book for someone trying to become a data scientist, or do data analysis with Python. I learnt a lot by working through the book, and still keep it for reference during work

Pretty good resource for a beginner to Python's data analysis libraries. This book is not for a Python beginner. Although Wes McKinney wrote Pandas, I feel that the Pandas part of the book is somewhat outdated or redundant in 2016. There is an excellent online resource maintained by Wes and other key contributors that is up-to-date, and, in my opinion, has better content for beginners learning Pandas. Because of several different contributors' perspectives, I find the examples and explanations better than those in the book. In 2012, when this book was published, that online

resource may not have been as good. I found the Pandas Time Series and Financial Applications chapters interesting, but they are also replicated on the website. So the book was improved upon by the author's own website (with the help of other contributors). :-) See[...] In particular, see sections: Tutorials, Intro to Data Structures - Series and DataFrame, and Essential Basic Functionality.The remaining 1/4 of the book had very useful concentrated intro to NumPy, Advanced NumPy, and Python Essentials reference. This book does not cover the newer development of R function calls from Python. In my opinion, R is winning the R vs Pandas argument due to ggplot2 and statistical learning professors publishing code first in R. Since R is now easy to use from within Python, Pandas might not get as much use. But it's still useful to know how to use Pandas as part of a data analyst's toolkit.I also want to warn buyers about faint printing on several physical copies of this book. I bought from AND directly from O'Reilly Media in trying to get a physical book that had good, solid printing on all pages. This was not possible. The physical book from O'Reilly had even fainter/worse printing than the version I got from . Better to save your money and just get with the eBook version if you are OK with that, which you can usually find cheaper online. O'Reilly puts on excellent conferences, but may be getting out of the printed book business. I guess most programmers buy eBooks now. I just find eBooks difficult to deal with when it comes to dense, technical books. I am fine with eBooks for fiction or more narrative non-fiction such as economics, popular science, or history.

Excellent coverage of using Python for data analysis. I liked the diversity of examples which motivated learning about the potent capabilities of Panda, Numpy and IPython for data analysis. I think the length of the book and level of detail was just right.I would have liked to see more examples about groupby and multi-level indexing. I'm finding this to be a very powerful aspect of Pandas and had to re-read those sections multiple times as well as scour the internet for more documentation and examples.

The author does a great job introducing the reader to Pandas, NumPy, and IPython. I am a novice in Python, recently switching from Java. I had bought "Programming in Python 3: A Complete Introduction to the Python Language" by Mark Summerfield and had a lot of trouble understanding how to install Python 3 and other libraries due to the fact my Mac already had Python 2.7. Wes McKinney does a phenomenal job walking the user through setting up IPython, Pandas, and NumPy.The only reason I am not giving the author 5 stars is because some of the datasets provided as examples could be better. I work in an environment where I am working with structured

data extracted out of SAP. Using Python to analyze 20k+ lines in multiple spreadsheets will simplify my work drastically.

Print quality is abysmal - horribly faded. I'm really surprised a publisher like O'Reilly can't managed to print a decent quality book - a home inkjet looks far better than this.

Pandas is a great package, and this book serves a good beginner's book to pandas and also to python.However, it would be better if concept and conventions in pandas can be introduced first, and later comes with more specifics, applications of using pandas to solve data tasks, it would be better. Some examples are good, but feels taking too long to get the understanding the concepts.

This book is a fantastic guide to familiarize yourself with pandas and numpy for python data analysis/modeling. I taught myself R a few years ago when I began working with statistics and data analysis and now that python is the "hot tool" to us, I began reading this book. I keep it on my desk at work so anytime an issue comes up in my code or I forget the correct syntax to use, it's a great book to reference.

Mining, Big Data) Pandas and Other Endangered Species: A Nonfiction Companion to Magic Tree House Merlin Mission #20: A Perfect Time for Pandas Principles of Data Wrangling: Practical Techniques for Data Preparation 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) Maya Python for Games and Film: A Complete Reference for Maya Python and the Maya Python API Analytics: Business Intelligence, Algorithms and Statistical Analysis (Predictive Analytics, Data Visualization, Data Analytics, Business Analytics, Decision Analysis, Big Data, Statistical Analysis) Python: Learn Python in a Day and Master It Well: The Only Essential Book You Need to Start Programming in Python Now 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. 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 Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (Princeton Series in Modern Observational Astronomy)