

Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

Recognizing the pretentiousness ways to acquire this books **pandas for everyone python data analysis addison wesley data analytics** is additionally useful. You have remained in right site to start getting this info. get the pandas for everyone python data analysis addison wesley data analytics connect that we manage to pay for here and check out the link.

You could purchase lead pandas for everyone python data analysis addison wesley data analytics or acquire it as soon as feasible. You could quickly download this pandas for everyone python data analysis addison wesley data analytics after getting deal. So, next you require the book swiftly, you can straight acquire it. Its appropriately very easy and so fats, isnt it? You have to favor to in this way of being

[Practical Statistics for Data Scientists](#) - Peter Bruce
2017-05-10

Statistical methods are a key part of of data science, yet very few data scientists have any formal statistics training.

Courses and books on basic statistics rarely cover the topic from a data science

perspective. This practical guide explains how to apply various statistical methods to data science, tells you how to avoid their misuse, and gives you advice on what's important and what's not. Many data science resources incorporate statistical methods but lack a deeper statistical perspective.

If you're familiar with the R programming language, and have some exposure to statistics, this quick reference bridges the gap in an accessible, readable format. With this book, you'll learn:

- Why exploratory data analysis is a key preliminary step in data science
- How random sampling can reduce bias and yield a higher quality dataset, even with big data
- How the principles of experimental design yield definitive answers to questions
- How to use regression to estimate outcomes and detect anomalies
- Key classification techniques for predicting which categories a record belongs to
- Statistical machine learning methods that "learn" from data
- Unsupervised learning methods for extracting meaning from unlabeled data

Pandas for Everyone - Daniel Y. Chen 2017-12-15

The Hands-On, Example-Rich Introduction to Pandas Data Analysis in Python Today, analysts must manage data characterized by extraordinary variety, velocity, and volume.

Using the open source Pandas library, you can use Python to rapidly automate and perform virtually any data analysis task, no matter how large or complex. Pandas can help you ensure the veracity of your data, visualize it for effective decision-making, and reliably reproduce analyses across multiple datasets. Pandas for Everyone brings together practical knowledge and insight for solving real problems with Pandas, even if you're new to Python data analysis. Daniel Y. Chen introduces key concepts through simple but practical examples, incrementally building on them to solve more difficult, real-world problems. Chen gives you a jumpstart on using Pandas with a realistic dataset and covers combining datasets, handling missing data, and structuring datasets for easier analysis and visualization. He demonstrates powerful data cleaning techniques, from basic string manipulation to applying functions simultaneously across dataframes. Once your

data is ready, Chen guides you through fitting models for prediction, clustering, inference, and exploration. He provides tips on performance and scalability, and introduces you to the wider Python data analysis ecosystem. Work with DataFrames and Series, and import or export data Create plots with matplotlib, seaborn, and pandas Combine datasets and handle missing data Reshape, tidy, and clean datasets so they're easier to work with Convert data types and manipulate text strings Apply functions to scale data manipulations Aggregate, transform, and filter large datasets with groupby Leverage Pandas' advanced date and time capabilities Fit linear models using statsmodels and scikit-learn libraries Use generalized linear modeling to fit models with different response variables Compare multiple models to select the "best" Regularize to overcome overfitting and improve performance Use clustering in unsupervised machine learning

Spark: The Definitive Guide - Bill Chambers 2018-02-08 Learn how to use, deploy, and maintain Apache Spark with this comprehensive guide, written by the creators of the open-source cluster-computing framework. With an emphasis on improvements and new features in Spark 2.0, authors Bill Chambers and Matei Zaharia break down Spark topics into distinct sections, each with unique goals. You'll explore the basic operations and common functions of Spark's structured APIs, as well as Structured Streaming, a new high-level API for building end-to-end streaming applications. Developers and system administrators will learn the fundamentals of monitoring, tuning, and debugging Spark, and explore machine learning techniques and scenarios for employing MLlib, Spark's scalable machine-learning library. Get a gentle overview of big data and Spark Learn about DataFrames, SQL, and Datasets—Spark's core APIs—through worked

examples Dive into Spark's low-level APIs, RDDs, and execution of SQL and DataFrames Understand how Spark runs on a cluster Debug, monitor, and tune Spark clusters and applications Learn the power of Structured Streaming, Spark's stream-processing engine Learn how you can apply MLlib to a variety of problems, including classification or recommendation

Marketing Data Science -

Thomas W. Miller 2015-05-02 Now , a leader of Northwestern University's prestigious analytics program presents a fully-integrated treatment of both the business and academic elements of marketing applications in predictive analytics. Writing for both managers and students, Thomas W. Miller explains essential concepts, principles, and theory in the context of real-world applications. Building on Miller's pioneering program, Marketing Data Science thoroughly addresses segmentation, target marketing, brand and product

positioning, new product development, choice modeling, recommender systems, pricing research, retail site selection, demand estimation, sales forecasting, customer retention, and lifetime value analysis. Starting where Miller's widely-praised Modeling Techniques in Predictive Analytics left off, he integrates crucial information and insights that were previously segregated in texts on web analytics, network science, information technology, and programming. Coverage includes: The role of analytics in delivering effective messages on the web Understanding the web by understanding its hidden structures Being recognized on the web - and watching your own competitors Visualizing networks and understanding communities within them Measuring sentiment and making recommendations Leveraging key data science methods: databases/data preparation, classical/Bayesian statistics, regression/classification,

machine learning, and text analytics Six complete case studies address exceptionally relevant issues such as: separating legitimate email from spam; identifying legally-relevant information for lawsuit discovery; gleaning insights from anonymous web surfing data, and more. This text's extensive set of web and network problems draw on rich public-domain data sources; many are accompanied by solutions in Python and/or R. Marketing Data Science will be an invaluable resource for all students, faculty, and professional marketers who want to use business analytics to improve marketing performance.

Python and HDF5 - Andrew Collette 2013-10-21

Gain hands-on experience with HDF5 for storing scientific data in Python. This practical guide quickly gets you up to speed on the details, best practices, and pitfalls of using HDF5 to archive and share numerical datasets ranging in size from gigabytes to terabytes.

Through real-world examples

and practical exercises, you'll explore topics such as scientific datasets, hierarchically organized groups, user-defined metadata, and interoperable files. Examples are applicable for users of both Python 2 and Python 3. If you're familiar with the basics of Python data analysis, this is an ideal introduction to HDF5. Get set up with HDF5 tools and create your first HDF5 file Work with datasets by learning the HDF5 Dataset object Understand advanced features like dataset chunking and compression Learn how to work with HDF5's hierarchical structure, using groups Create self-describing files by adding metadata with HDF5 attributes Take advantage of HDF5's type system to create interoperable files Express relationships among data with references, named types, and dimension scales Discover how Python mechanisms for writing parallel code interact with HDF5

R for Everyone - Jared P. Lander 2014

A guide to using and

understanding the 'R' computer programming language.

Hands-On Data Analysis with NumPy and Pandas - Miller Miller 2018-06-29

Get to grips with the most popular Python packages that make data analysis possible
Key Features Explore the tools you need to become a data analyst Discover practical examples to help you grasp data processing concepts Walk through hierarchical indexing and grouping for data analysis
Book Description Python, a multi-paradigm programming language, has become the language of choice for data scientists for visualization, data analysis, and machine learning. *Hands-On Data Analysis with NumPy and Pandas* starts by guiding you in setting up the right environment for data analysis with Python, along with helping you install the correct Python distribution. In addition to this, you will work with the Jupyter notebook and set up a database. Once you have covered Jupyter, you will dig deep into Python's NumPy package, a powerful extension

with advanced mathematical functions. You will then move on to creating NumPy arrays and employing different array methods and functions. You will explore Python's pandas extension which will help you get to grips with data mining and learn to subset your data. Last but not the least you will grasp how to manage your datasets by sorting and ranking them. By the end of this book, you will have learned to index and group your data for sophisticated data analysis and manipulation. What you will learn Understand how to install and manage Anaconda Read, sort, and map data using NumPy and pandas Find out how to create and slice data arrays using NumPy Discover how to subset your DataFrames using pandas Handle missing data in a pandas DataFrame Explore hierarchical indexing and plotting with pandas Who this book is for *Hands-On Data Analysis with NumPy and Pandas* is for you if you are a Python developer and want to take your first steps into the

world of data analysis. No previous experience of data analysis is required to enjoy this book.

Learn More Python 3 the Hard Way - Zed A. Shaw
2017-09-01

Transform Your Ideas into High-Quality Python Code! Zed Shaw has perfected the world's best system for becoming a truly effective Python 3.x developer. Follow it and you will succeed—just like the tens of millions of programmers he's already taught. You bring the discipline, commitment, and persistence; the author supplies everything else. In *Learn Python 3 the Hard Way*, Zed Shaw taught you the basics of Programming with Python 3. Now, in *Learn More Python 3 the Hard Way*, you'll go far beyond the basics by working through 52 brilliantly crafted projects. Each one helps you build a key practical skill, combining demos to get you started and challenges to deepen your understanding. Zed then teaches you even more in 12 hours of online videos, where he shows you

how to break, fix, and debug your code. First, you'll discover how to analyze a concept, idea, or problem to implement in software. Then, step by step, you'll learn to design solutions based on your analyses and implement them as simply and elegantly as possible.

Throughout, Shaw stresses process so you can get started and build momentum, creativity to solve new problems, and quality so you'll build code people can rely on. Manage complex projects with a programmer's text editor Leverage the immense power of data structures Apply algorithms to process your data structures Master indispensable text parsing and processing techniques Use SQL to efficiently and logically model stored data Learn powerful command-line tools and skills Combine multiple practices in complete projects It'll be hard at first. But soon, you'll just get it—and that will feel great! This course will reward you for every minute you put into it. Soon, you'll go beyond merely writing code

that runs: you'll craft high-quality Python code that solves real problems. You'll be a serious Python programmer. Perfect for Everyone Who's Already Started Working with Python, including Junior Developers and Seasoned Python Programmers Upgrading to Python 3.6+ Register your product at informit.com/register for convenient access to downloads, updates, and/or corrections as they become available.

Mastering pandas - Ashish Kumar 2019-10-25

Perform advanced data manipulation tasks using pandas and become an expert data analyst. Key Features Manipulate and analyze your data expertly using the power of pandas Work with missing data and time series data and become a true pandas expert Includes expert tips and techniques on making your data analysis tasks easier Book Description pandas is a popular Python library used by data scientists and analysts worldwide to

manipulate and analyze their data. This book presents useful data manipulation techniques in pandas to perform complex data analysis in various domains. An update to our highly successful previous edition with new features, examples, updated code, and more, this book is an in-depth guide to get the most out of pandas for data analysis. Designed for both intermediate users as well as seasoned practitioners, you will learn advanced data manipulation techniques, such as multi-indexing, modifying data structures, and sampling your data, which allow for powerful analysis and help you gain accurate insights from it. With the help of this book, you will apply pandas to different domains, such as Bayesian statistics, predictive analytics, and time series analysis using an example-based approach. And not just that; you will also learn how to prepare powerful, interactive business reports in pandas using the Jupyter notebook. By the end of this book, you will learn how to

perform efficient data analysis using pandas on complex data, and become an expert data analyst or data scientist in the process. What you will learn

Speed up your data analysis by importing data into pandas

Keep relevant data points by selecting subsets of your data

Create a high-quality dataset by cleaning data and fixing missing values

Compute actionable analytics with grouping and aggregation in pandas

Master time series data analysis in pandas

Make powerful reports in pandas using Jupyter notebooks

Who this book is for

This book is for data scientists, analysts and Python developers who wish to explore advanced data analysis and scientific computing techniques using pandas. Some fundamental understanding of Python programming and familiarity with the basic data analysis concepts is all you need to get started with this book.

Foundations of Deep Reinforcement Learning - Laura Graesser 2019-11-20
The Contemporary

Introduction to Deep Reinforcement Learning that Combines Theory and Practice

Deep reinforcement learning (deep RL) combines deep learning and reinforcement learning, in which artificial agents learn to solve sequential decision-making problems. In the past decade deep RL has achieved remarkable results on a range of problems, from single and multiplayer games—such as Go, Atari games, and DotA 2—to robotics.

Foundations of Deep Reinforcement Learning is an introduction to deep RL that uniquely combines both theory and implementation. It starts with intuition, then carefully explains the theory of deep RL algorithms, discusses implementations in its companion software library SLM Lab, and finishes with the practical details of getting deep RL to work. This guide is ideal for both computer science students and software engineers who are familiar with basic machine learning concepts and have a working understanding of Python.

Understand each key aspect of a deep RL problem Explore policy- and value-based algorithms, including REINFORCE, SARSA, DQN, Double DQN, and Prioritized Experience Replay (PER) Delve into combined algorithms, including Actor-Critic and Proximal Policy Optimization (PPO) Understand how algorithms can be parallelized synchronously and asynchronously Run algorithms in SLM Lab and learn the practical implementation details for getting deep RL to work Explore algorithm benchmark results with tuned hyperparameters Understand how deep RL environments are designed Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

[Hands-On Data Analysis with Pandas](#) - Stefanie Molin
2021-04-29

Get to grips with pandas by working with real datasets and master data discovery, data manipulation, data preparation,

and handling data for analytical tasks Key Features Perform efficient data analysis and manipulation tasks using pandas 1.x Apply pandas to different real-world domains with the help of step-by-step examples Make the most of pandas as an effective data exploration tool Book Description Extracting valuable business insights is no longer a 'nice-to-have', but an essential skill for anyone who handles data in their enterprise. Hands-On Data Analysis with Pandas is here to help beginners and those who are migrating their skills into data science get up to speed in no time. This book will show you how to analyze your data, get started with machine learning, and work effectively with the Python libraries often used for data science, such as pandas, NumPy, matplotlib, seaborn, and scikit-learn. Using real-world datasets, you will learn how to use the pandas library to perform data wrangling to reshape, clean, and aggregate your data. Then, you will learn how to conduct exploratory

data analysis by calculating summary statistics and visualizing the data to find patterns. In the concluding chapters, you will explore some applications of anomaly detection, regression, clustering, and classification using scikit-learn to make predictions based on past data. This updated edition will equip you with the skills you need to use pandas 1.x to efficiently perform various data manipulation tasks, reliably reproduce analyses, and visualize your data for effective decision making - valuable knowledge that can be applied across multiple domains. What you will learn Understand how data analysts and scientists gather and analyze data Perform data analysis and data wrangling using Python Combine, group, and aggregate data from multiple sources Create data visualizations with pandas, matplotlib, and seaborn Apply machine learning algorithms to identify patterns and make predictions Use Python data science libraries to analyze

real-world datasets Solve common data representation and analysis problems using pandas Build Python scripts, modules, and packages for reusable analysis code Who this book is for This book is for data science beginners, data analysts, and Python developers who want to explore each stage of data analysis and scientific computing using a wide range of datasets. Data scientists looking to implement pandas in their machine learning workflow will also find plenty of valuable know-how as they progress. You'll find it easier to follow along with this book if you have a working knowledge of the Python programming language, but a Python crash-course tutorial is provided in the code bundle for anyone who needs a refresher.

Hands-On Data Science and Python Machine Learning -

Frank Kane 2017-07-31

This book covers the fundamentals of machine learning with Python in a concise and dynamic manner. It covers data mining and

large-scale machine learning using Apache Spark. About This Book Take your first steps in the world of data science by understanding the tools and techniques of data analysis Train efficient Machine Learning models in Python using the supervised and unsupervised learning methods Learn how to use Apache Spark for processing Big Data efficiently Who This Book Is For If you are a budding data scientist or a data analyst who wants to analyze and gain actionable insights from data using Python, this book is for you. Programmers with some experience in Python who want to enter the lucrative world of Data Science will also find this book to be very useful, but you don't need to be an expert Python coder or mathematician to get the most from this book. What You Will Learn Learn how to clean your data and ready it for analysis Implement the popular clustering and regression methods in Python Train efficient machine learning models using decision trees and random forests

Visualize the results of your analysis using Python's Matplotlib library Use Apache Spark's MLlib package to perform machine learning on large datasets In Detail Join Frank Kane, who worked on Amazon and IMDb's machine learning algorithms, as he guides you on your first steps into the world of data science. Hands-On Data Science and Python Machine Learning gives you the tools that you need to understand and explore the core topics in the field, and the confidence and practice to build and analyze your own machine learning models. With the help of interesting and easy-to-follow practical examples, Frank Kane explains potentially complex topics such as Bayesian methods and K-means clustering in a way that anybody can understand them. Based on Frank's successful data science course, Hands-On Data Science and Python Machine Learning empowers you to conduct data analysis and perform efficient machine learning using Python. Let Frank help you unearth the

value in your data using the various data mining and data analysis techniques available in Python, and to develop efficient predictive models to predict future results. You will also learn how to perform large-scale machine learning on Big Data using Apache Spark. The book covers preparing your data for analysis, training machine learning models, and visualizing the final data analysis. Style and approach This comprehensive book is a perfect blend of theory and hands-on code examples in Python which can be used for your reference at any time.

Foundational Python for Data Science - Kennedy Behrman 2021

Data science and machine learning - two of the world's hottest fields - are attracting talent from a wide variety of technical, business, and liberal arts disciplines. Python, the world's #1 programming language, is also the most popular language for data science and machine learning. This is the first guide specifically designed to help

students with widely diverse backgrounds learn foundational Python so they can use it for data science and machine learning. This book is catered to introductory-level college courses on data science. Leading data science instructor and practitioner Kennedy Behrman first walks through the process of learning to code for the first time with Python and Jupyter notebook, then introduces key libraries every Python data science programmer needs to master. Once students have learned these foundations, Behrman introduces intermediate and applied Python techniques for real-world problem-solving. Throughout, Foundational Python for Data Science presents hands-on exercises, learning assessments, case studies, and more - all created with colab (jupyter compatible) notebooks, so students can execute all coding examples interactively without installing or configuring any software.

Pandas Cookbook - Theodore Petrou 2017-10-23

Over 95 hands-on recipes to

leverage the power of pandas for efficient scientific computation and data analysis

About This Book

Use the power of pandas to solve most complex scientific computing problems with ease Leverage fast, robust data structures in pandas to gain useful insights from your data Practical, easy to implement recipes for quick solutions to common problems in data using pandas

Who This Book Is For

This book is for data scientists, analysts and Python developers who wish to explore data analysis and scientific computing in a practical, hands-on manner. The recipes included in this book are suitable for both novice and advanced users, and contain helpful tips, tricks and caveats wherever necessary. Some understanding of pandas will be helpful, but not mandatory.

What You Will Learn

Master the fundamentals of pandas to quickly begin exploring any dataset Isolate any subset of data by properly selecting and querying the data Split data into independent groups before

applying aggregations and transformations to each group Restructure data into tidy form to make data analysis and visualization easier Prepare real-world messy datasets for machine learning Combine and merge data from different sources through pandas SQL-like operations Utilize pandas unparalleled time series functionality Create beautiful and insightful visualizations through pandas direct hooks to Matplotlib and Seaborn

In Detail

This book will provide you with unique, idiomatic, and fun recipes for both fundamental and advanced data manipulation tasks with pandas. Some recipes focus on achieving a deeper understanding of basic principles, or comparing and contrasting two similar operations. Other recipes will dive deep into a particular dataset, uncovering new and unexpected insights along the way. The pandas library is massive, and it's common for frequent users to be unaware of many of its more impressive features. The official pandas

documentation, while thorough, does not contain many useful examples of how to piece together multiple commands like one would do during an actual analysis. This book guides you, as if you were looking over the shoulder of an expert, through practical situations that you are highly likely to encounter. Many advanced recipes combine several different features across the pandas library to generate results. Style and approach The author relies on his vast experience teaching pandas in a professional setting to deliver very detailed explanations for each line of code in all of the recipes. All code and dataset explanations exist in Jupyter Notebooks, an excellent interface for exploring data.

Practical Artificial Intelligence with Swift -

Mars Geldard 2019-09-03
Create and implement AI-based features in your Swift apps for iOS, macOS, tvOS, and watchOS. With this practical book, programmers and developers of all kinds will find

a one-stop shop for AI and machine learning with Swift. Taking a task-based approach, you'll learn how to build features that use powerful AI features to identify images, make predictions, generate content, recommend things, and more. AI is increasingly essential for every developer—and you don't need to be a data scientist or mathematician to take advantage of it in your apps. Explore Swift-based AI and ML techniques for building applications. Learn where and how AI-driven features make sense. Inspect tools such as Apple's Python-powered Turi Create and Google's Swift for TensorFlow to train and build models. I: Fundamentals and Tools—Learn AI basics, our task-based approach, and discover how to build or find a dataset. II: Task Based AI—Build vision, audio, text, motion, and augmentation-related features; learn how to convert preexisting models. III: Beyond—Discover the theory behind task-based practice, explore AI and ML methods,

and learn how you can build it all from scratch... if you want to

Pragmatic AI - Noah Gift

2018-07-12

Master Powerful Off-the-Shelf Business Solutions for AI and Machine Learning Pragmatic AI will help you solve real-world problems with contemporary machine learning, artificial intelligence, and cloud computing tools. Noah Gift demystifies all the concepts and tools you need to get results—even if you don't have a strong background in math or data science. Gift illuminates powerful off-the-shelf cloud offerings from Amazon, Google, and Microsoft, and demonstrates proven techniques using the Python data science ecosystem. His workflows and examples help you streamline and simplify every step, from deployment to production, and build exceptionally scalable solutions. As you learn how machine language (ML) solutions work, you'll gain a more intuitive understanding of what you can achieve with

them and how to maximize their value. Building on these fundamentals, you'll walk step-by-step through building cloud-based AI/ML applications to address realistic issues in sports marketing, project management, product pricing, real estate, and beyond. Whether you're a business professional, decision-maker, student, or programmer, Gift's expert guidance and wide-ranging case studies will prepare you to solve data science problems in virtually any environment. Get and configure all the tools you'll need Quickly review all the Python you need to start building machine learning applications Master the AI and ML toolchain and project lifecycle Work with Python data science tools such as IPython, Pandas, Numpy, Jupyter Notebook, and Sklearn Incorporate a pragmatic feedback loop that continually improves the efficiency of your workflows and systems Develop cloud AI solutions with Google Cloud Platform, including TPU, Colaboratory,

and Datalab services Define Amazon Web Services cloud AI workflows, including spot instances, code pipelines, boto, and more Work with Microsoft Azure AI APIs Walk through building six real-world AI applications, from start to finish Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

Effective Computation in

Physics - Anthony Scopatz

2015-06-25

More physicists today are taking on the role of software developer as part of their research, but software development isn't always easy or obvious, even for physicists. This practical book teaches essential software development skills to help you automate and accomplish nearly any aspect of research in a physics-based field. Written by two PhDs in nuclear engineering, this book includes practical examples drawn from a working knowledge of physics concepts. You'll learn how to use the

Python programming language to perform everything from collecting and analyzing data to building software and publishing your results. In four parts, this book includes: Getting Started: Jump into Python, the command line, data containers, functions, flow control and logic, and classes and objects Getting It Done: Learn about regular expressions, analysis and visualization, NumPy, storing data in files and HDF5, important data structures in physics, computing in parallel, and deploying software Getting It Right: Build pipelines and software, learn to use local and remote version control, and debug and test your code Getting It Out There: Document your code, process and publish your findings, and collaborate efficiently; dive into software licenses, ownership, and copyright procedures **Pandas in Action** - Boris Paskhaver 2021-10-05 Pandas has rapidly become one of Python's most popular data analysis libraries. With pandas you can efficiently sort,

analyze, filter and munge almost any type of data. In *Pandas in Action*, a friendly and example-rich introduction, author Boris Paskhaver shows you how to master this versatile tool and take the next steps in your data science career. about the technology Anyone who's used spreadsheet software will find pandas familiar. While its column-based grids might remind you of Excel or Google Sheets, pandas is more flexible and far more powerful. It can efficiently perform operations on millions of rows and be used in tandem with other Python libraries for statistics, machine learning, and more. And best of all, using pandas doesn't mean sacrificing user productivity or needing to write tons of complex code. It's clean, intuitive, and fast. about the book *Pandas in Action* makes it easy to dive into Python-based data analysis. You'll learn to use pandas to automate repetitive spreadsheet functionality and derive insight from data by sorting columns, filtering data subsets, and

creating multi-leveled indices. Each chapter is a self-contained tutorial, letting you dip in when you need to troubleshoot tricky problems. Best of all, you won't be learning from sterile or randomly created data. You'll start with a variety of datasets that are big, small, incomplete, broken, and messy and learn how to clean and format them for proper analysis. what's inside Import a CSV, identify issues with its data structures, and convert it to the proper format Sort, filter, pivot, and draw conclusions from a dataset and its subsets Identify trends from text-based and time-based data Organize, group, merge, and join separate datasets Real-world datasets that are easy to download and explore about the reader For readers experienced with spreadsheet software who know the basics of Python. about the author Boris Paskhaver is a software engineer, Agile consultant, and educator. His six programming courses on Udemy have amassed 236,000 students,

with an average course rating of 4.59 out of 5. He first used Python and the pandas library to derive a variety of business insights at the world's #1 jobs site, Indeed.com.

[Cleaning Data for Effective Data Science](#) - David Mertz
2021-03-31

Think about your data intelligently and ask the right questions. Key Features: Master data cleaning techniques necessary to perform real-world data science and machine learning tasks. Spot common problems with dirty data and develop flexible solutions from first principles. Test and refine your newly acquired skills through detailed exercises at the end of each chapter. Book Description: Data cleaning is the all-important first step to successful data science, data analysis, and machine learning. If you work with any kind of data, this book is your go-to resource, arming you with the insights and heuristics experienced data scientists had to learn the hard way. In a light-hearted and engaging

exploration of different tools, techniques, and datasets real and fictitious, Python veteran David Mertz teaches you the ins and outs of data preparation and the essential questions you should be asking of every piece of data you work with. Using a mixture of Python, R, and common command-line tools, *Cleaning Data for Effective Data Science* follows the data cleaning pipeline from start to end, focusing on helping you understand the principles underlying each step of the process. You'll look at data ingestion of a vast range of tabular, hierarchical, and other data formats, impute missing values, detect unreliable data and statistical anomalies, and generate synthetic features. The long-form exercises at the end of each chapter let you get hands-on with the skills you've acquired along the way, also providing a valuable resource for academic courses. What you will learn: Ingest and work with common data formats like JSON, CSV, SQL and NoSQL databases, PDF, and binary

serialized data structures Understand how and why we use tools such as pandas, SciPy, scikit-learn, Tidyverse, and BashApply useful rules and heuristics for assessing data quality and detecting bias, like Benford's law and the 68-95-99.7 rule Identify and handle unreliable data and outliers, examining z-score and other statistical properties Impute sensible values into missing data and use sampling to fix imbalances Use dimensionality reduction, quantization, one-hot encoding, and other feature engineering techniques to draw out patterns in your data Work carefully with time series data, performing detrending and interpolation Who this book is for This book is designed to benefit software developers, data scientists, aspiring data scientists, teachers, and students who work with data. If you want to improve your rigor in data hygiene or are looking for a refresher, this book is for you. Basic familiarity with statistics, general concepts in machine

learning, knowledge of a programming language (Python or R), and some exposure to data science are helpful. *Python Cookbook* - David Beazley 2013-05-10 If you need help writing programs in Python 3, or want to update older Python 2 code, this book is just the ticket. Packed with practical recipes written and tested with Python 3.3, this unique cookbook is for experienced Python programmers who want to focus on modern tools and idioms. Inside, you'll find complete recipes for more than a dozen topics, covering the core Python language as well as tasks common to a wide variety of application domains. Each recipe contains code samples you can use in your projects right away, along with a discussion about how and why the solution works. Topics include: Data Structures and Algorithms Strings and Text Numbers, Dates, and Times Iterators and Generators Files and I/O Data Encoding and Processing Functions Classes and Objects Metaprogramming

Modules and Packages
Network and Web
Programming Concurrency
Utility Scripting and System
Administration Testing,
Debugging, and Exceptions C
Extensions

Effective Python - Brett
Slatkin 2015

Effective Python will help
students harness the full power
of Python to write exceptionally
robust, efficient, maintainable,
and well-performing code.

Utilizing the concise, scenario-
driven style pioneered in Scott
Meyers's best-selling Effective
C++, Brett Slatkin brings
together 53 Python best
practices, tips, shortcuts, and
realistic code examples from
expert programmers. Each
section contains specific,
actionable guidelines organized
into items, each with carefully
worded advice supported by
detailed technical arguments
and illuminating examples.

**Learn Python 3 the Hard
Way** - Zed A. Shaw 2017-06-26

You Will Learn Python 3! Zed
Shaw has perfected the world's
best system for learning Python
3. Follow it and you will

succeed—just like the millions
of beginners Zed has taught to
date! You bring the discipline,
commitment, and persistence;
the author supplies everything
else. In Learn Python 3 the
Hard Way, you'll learn Python
by working through 52
brilliantly crafted exercises.
Read them. Type their code
precisely. (No copying and
pasting!) Fix your mistakes.
Watch the programs run. As
you do, you'll learn how a
computer works; what good
programs look like; and how to
read, write, and think about
code. Zed then teaches you
even more in 5+ hours of video
where he shows you how to
break, fix, and debug your
code—live, as he's doing the
exercises. Install a complete
Python environment Organize
and write code Fix and break
code Basic mathematics
Variables Strings and text
Interact with users Work with
files Looping and logic Data
structures using lists and
dictionaries Program design
Object-oriented programming
Inheritance and composition
Modules, classes, and objects

Python packaging Automated testing Basic game development Basic web development It'll be hard at first. But soon, you'll just get it—and that will feel great! This course will reward you for every minute you put into it. Soon, you'll know one of the world's most powerful, popular programming languages. You'll be a Python programmer. This Book Is Perfect For Total beginners with zero programming experience Junior developers who know one or two languages Returning professionals who haven't written code in years Seasoned professionals looking for a fast, simple, crash course in Python 3

Data Science for Marketing Analytics - Tommy Blanchard
2019-03-30

Explore new and more sophisticated tools that reduce your marketing analytics efforts and give you precise results Key Features Study new techniques for marketing analytics Explore uses of machine learning to power your marketing analyses Work

through each stage of data analytics with the help of multiple examples and exercises Book Description Data Science for Marketing Analytics covers every stage of data analytics, from working with a raw dataset to segmenting a population and modeling different parts of the population based on the segments. The book starts by teaching you how to use Python libraries, such as pandas and Matplotlib, to read data from Python, manipulate it, and create plots, using both categorical and continuous variables. Then, you'll learn how to segment a population into groups and use different clustering techniques to evaluate customer segmentation. As you make your way through the chapters, you'll explore ways to evaluate and select the best segmentation approach, and go on to create a linear regression model on customer value data to predict lifetime value. In the concluding chapters, you'll gain an understanding of regression techniques and

tools for evaluating regression models, and explore ways to predict customer choice using classification algorithms. Finally, you'll apply these techniques to create a churn model for modeling customer product choices. By the end of this book, you will be able to build your own marketing reporting and interactive dashboard solutions. What you will learn

Analyze and visualize data in Python using pandas and Matplotlib

Study clustering techniques, such as hierarchical and k-means clustering

Create customer segments based on manipulated data

Predict customer lifetime value using linear regression

Use classification algorithms to understand customer choice

Optimize classification algorithms to extract maximal information

Who this book is for

Data Science for Marketing Analytics is designed for developers and marketing analysts looking to use new, more sophisticated tools in their marketing analytics efforts. It'll help if you have

prior experience of coding in Python and knowledge of high school level mathematics. Some experience with databases, Excel, statistics, or Tableau is useful but not necessary.

Think Stats - Allen B. Downey
2014-10-16

If you know how to program, you have the skills to turn data into knowledge, using tools of probability and statistics. This concise introduction shows you how to perform statistical analysis computationally, rather than mathematically, with programs written in Python. By working with a single case study throughout this thoroughly revised book, you'll learn the entire process of exploratory data analysis—from collecting data and generating statistics to identifying patterns and testing hypotheses. You'll explore distributions, rules of probability, visualization, and many other tools and concepts. New chapters on regression, time series analysis, survival analysis, and analytic methods will enrich your discoveries.

Develop an understanding of probability and statistics by writing and testing code Run experiments to test statistical behavior, such as generating samples from several distributions Use simulations to understand concepts that are hard to grasp mathematically Import data from most sources with Python, rather than rely on data that's cleaned and formatted for statistics tools Use statistical inference to answer questions about real-world data

Python for Programmers -

Paul J. Deitel 2019-03-15

The professional programmer's Deitel® guide to Python® with introductory artificial intelligence case studies

Written for programmers with a background in another high-level language, Python for Programmers uses hands-on instruction to teach today's most compelling, leading-edge computing technologies and programming in Python—one of the world's most popular and fastest-growing languages.

Please read the Table of Contents diagram inside the

front cover and the Preface for more details. In the context of 500+, real-world examples ranging from individual snippets to 40 large scripts and full implementation case studies, you'll use the interactive IPython interpreter with code in Jupyter Notebooks to quickly master the latest Python coding idioms. After covering Python Chapters 1-5 and a few key parts of Chapters 6-7, you'll be able to handle significant portions of the hands-on introductory AI case studies in Chapters 11-16, which are loaded with cool, powerful, contemporary examples. These include natural language processing, data mining Twitter® for sentiment analysis, cognitive computing with IBM® Watson™, supervised machine learning with classification and regression, unsupervised machine learning with clustering, computer vision through deep learning and convolutional neural networks, deep learning with recurrent neural networks, big data with Hadoop®, Spark™ and NoSQL

databases, the Internet of Things and more. You'll also work directly or indirectly with cloud-based services, including Twitter, Google Translate™, IBM Watson, Microsoft® Azure®, OpenMapQuest, PubNub and more. Features 500+ hands-on, real-world, live-code examples from snippets to case studies IPython + code in Jupyter® Notebooks Library-focused: Uses Python Standard Library and data science libraries to accomplish significant tasks with minimal code Rich Python coverage: Control statements, functions, strings, files, JSON serialization, CSV, exceptions Procedural, functional-style and object-oriented programming Collections: Lists, tuples, dictionaries, sets, NumPy arrays, pandas Series & DataFrames Static, dynamic and interactive visualizations Data experiences with real-world datasets and data sources Intro to Data Science sections: AI, basic stats, simulation, animation, random variables, data wrangling, regression AI, big data and

cloud data science case studies: NLP, data mining Twitter®, IBM® Watson™, machine learning, deep learning, computer vision, Hadoop®, Spark™, NoSQL, IoT Open-source libraries: NumPy, pandas, Matplotlib, Seaborn, Folium, SciPy, NLTK, TextBlob, spaCy, Textastic, Tweepy, scikit-learn®, Keras and more Accompanying code examples are available here: http://ptgmedia.pearsoncmg.com/imprint_downloads/informit/bookreg/9780135224335/9780135224335_examples.zip. Register your product for convenient access to downloads, updates, and/or corrections as they become available. See inside book for more information. *Hands-On Data Analysis with Pandas* - Stefanie Molin 2019-07-26 Get to grips with pandas—a versatile and high-performance Python library for data manipulation, analysis, and discovery Key Features Perform efficient data analysis and manipulation tasks using pandas Apply pandas to

different real-world domains using step-by-step demonstrations. Get accustomed to using pandas as an effective data exploration tool. Book Description Data analysis has become a necessary skill in a variety of positions where knowing how to work with data and extract insights can generate significant value. Hands-On Data Analysis with Pandas will show you how to analyze your data, get started with machine learning, and work effectively with Python libraries often used for data science, such as pandas, NumPy, matplotlib, seaborn, and scikit-learn. Using real-world datasets, you will learn how to use the powerful pandas library to perform data wrangling to reshape, clean, and aggregate your data. Then, you will learn how to conduct exploratory data analysis by calculating summary statistics and visualizing the data to find patterns. In the concluding chapters, you will explore some applications of anomaly detection, regression, clustering, and classification,

using scikit-learn, to make predictions based on past data. By the end of this book, you will be equipped with the skills you need to use pandas to ensure the veracity of your data, visualize it for effective decision-making, and reliably reproduce analyses across multiple datasets. What you will learn. Understand how data analysts and scientists gather and analyze data. Perform data analysis and data wrangling in Python. Combine, group, and aggregate data from multiple sources. Create data visualizations with pandas, matplotlib, and seaborn. Apply machine learning (ML) algorithms to identify patterns and make predictions. Use Python data science libraries to analyze real-world datasets. Use pandas to solve common data representation and analysis problems. Build Python scripts, modules, and packages for reusable analysis code. Who this book is for This book is for data analysts, data science beginners, and Python developers who want to explore each stage of data

analysis and scientific computing using a wide range of datasets. You will also find this book useful if you are a data scientist who is looking to implement pandas in machine learning. Working knowledge of Python programming language will be beneficial.

A Hands-On Introduction to Data Science - Chirag Shah
2020-04-02

An introductory textbook offering a low barrier entry to data science; the hands-on approach will appeal to students from a range of disciplines.

Python for Data Analysis - Wes McKinney
2017-09-25

Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the

creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing. Learn basic and advanced features in NumPy (Numerical Python). Get started with data analysis tools in the pandas library. Use flexible tools to load, clean, transform, merge, and reshape data. Create informative visualizations with matplotlib. Apply the pandas groupby facility to slice, dice, and summarize datasets. Analyze and manipulate regular and irregular time series data. Learn how to solve real-world data analysis problems with thorough, detailed examples. *Data Analytics with Spark Using Python* - Jeffrey Aven
2018-06-18
Solve Data Analytics Problems with Spark, PySpark, and

Related Open Source Tools
Spark is at the heart of today's Big Data revolution, helping data professionals supercharge efficiency and performance in a wide range of data processing and analytics tasks. In this guide, Big Data expert Jeffrey Aven covers all you need to know to leverage Spark, together with its extensions, subprojects, and wider ecosystem. Aven combines a language-agnostic introduction to foundational Spark concepts with extensive programming examples utilizing the popular and intuitive PySpark development environment. This guide's focus on Python makes it widely accessible to large audiences of data professionals, analysts, and developers—even those with little Hadoop or Spark experience. Aven's broad coverage ranges from basic to advanced Spark programming, and Spark SQL to machine learning. You'll learn how to efficiently manage all forms of data with Spark: streaming, structured, semi-structured, and unstructured. Throughout,

concise topic overviews quickly get you up to speed, and extensive hands-on exercises prepare you to solve real problems. Coverage includes:

- Understand Spark's evolving role in the Big Data and Hadoop ecosystems
- Create Spark clusters using various deployment modes
- Control and optimize the operation of Spark clusters and applications
- Master Spark Core RDD API programming techniques
- Extend, accelerate, and optimize Spark routines with advanced API platform constructs, including shared variables, RDD storage, and partitioning
- Efficiently integrate Spark with both SQL and nonrelational data stores
- Perform stream processing and messaging with Spark Streaming and Apache Kafka
- Implement predictive modeling with SparkR and Spark MLlib

High Performance Python - Micha Gorelick 2020-04-30
Your Python code may run correctly, but you need it to run faster. Updated for Python 3, this expanded edition shows you how to locate performance

bottlenecks and significantly speed up your code in high-data-volume programs. By exploring the fundamental theory behind design choices, High Performance Python helps you gain a deeper understanding of Python's implementation. How do you take advantage of multicore architectures or clusters? Or build a system that scales up and down without losing reliability? Experienced Python programmers will learn concrete solutions to many issues, along with war stories from companies that use high-performance Python for social media analytics, productionized machine learning, and more. Get a better grasp of NumPy, Cython, and profilers Learn how Python abstracts the underlying computer architecture Use profiling to find bottlenecks in CPU time and memory usage Write efficient programs by choosing appropriate data structures Speed up matrix and vector computations Use tools to compile Python down to machine code Manage multiple

I/O and computational operations concurrently Convert multiprocessing code to run on local or remote clusters Deploy code faster using tools like Docker

Machine Learning with Python for Everyone - Mark Fenner 2019-07-30

The Complete Beginner's Guide to Understanding and Building Machine Learning Systems with Python Machine Learning with Python for Everyone will help you master the processes, patterns, and strategies you need to build effective learning systems, even if you're an absolute beginner. If you can write some Python code, this book is for you, no matter how little college-level math you know. Principal instructor Mark E. Fenner relies on plain-English stories, pictures, and Python examples to communicate the ideas of machine learning. Mark begins by discussing machine learning and what it can do; introducing key mathematical and computational topics in an approachable manner; and

walking you through the first steps in building, training, and evaluating learning systems. Step by step, you'll fill out the components of a practical learning system, broaden your toolbox, and explore some of the field's most sophisticated and exciting techniques. Whether you're a student, analyst, scientist, or hobbyist, this guide's insights will be applicable to every learning system you ever build or use. Understand machine learning algorithms, models, and core machine learning concepts Classify examples with classifiers, and quantify examples with regressors Realistically assess performance of machine learning systems Use feature engineering to smooth rough data into useful forms Chain multiple components into one system and tune its performance Apply machine learning techniques to images and text Connect the core concepts to neural networks and graphical models Leverage the Python scikit-learn library and other powerful tools

Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

SciPy and NumPy - Eli Bressert 2012

"Optimizing and boosting your Python programming"--Cover. Python Data Analysis - Third Edition - Avinash Navlani 2021-02-05

Understand data analysis pipelines using machine learning algorithms and techniques with this practical guideKey Features* Prepare and clean your data to use it for exploratory analysis, data manipulation, and data wrangling* Discover supervised, unsupervised, probabilistic, and Bayesian machine learning methods* Get to grips with graph processing and sentiment analysisBook DescriptionData analysis enables you to generate value from small and big data by discovering new patterns and trends, and Python is one of the most popular tools for analyzing a wide variety of

data. With this book, you'll get up and running using Python for data analysis by exploring the different phases and methodologies used in data analysis and learning how to use modern libraries from the Python ecosystem to create efficient data pipelines. Starting with the essential statistical and data analysis fundamentals using Python, you'll perform complex data analysis and modeling, data manipulation, data cleaning, and data visualization using easy-to-follow examples. You'll then understand how to conduct time series analysis and signal processing using ARMA models. As you advance, you'll get to grips with smart processing and data analytics using machine learning algorithms such as regression, classification, Principal Component Analysis (PCA), and clustering. In the concluding chapters, you'll work on real-world examples to analyze textual and image data using natural language processing (NLP) and image analytics techniques, respectively.

Finally, the book will demonstrate parallel computing using Dask. By the end of this data analysis book, you'll be equipped with the skills you need to prepare data for analysis and create meaningful data visualizations for forecasting values from data. What you will learn*

- Explore data science and its various process models*
- Perform data manipulation using NumPy and pandas for aggregating, cleaning, and handling missing values*
- Create interactive visualizations using Matplotlib, Seaborn, and Bokeh*
- Retrieve, process, and store data in a wide range of formats*
- Understand data preprocessing and feature engineering using pandas and scikit-learn*
- Perform time series analysis and signal processing using sunspot cycle data*
- Analyze textual data and image data to perform advanced analysis*
- Get up to speed with parallel computing using Dask

Who this book is for: This book is for data analysts, business analysts, statisticians, and data

scientists looking to learn how to use Python for data analysis. Students and academic faculties will also find this book useful for learning and teaching Python data analysis using a hands-on approach. A basic understanding of math and working knowledge of the Python programming language will help you get started with this book.

Pandas 1.x Cookbook - Matt Harrison 2020-02-27

Use the power of pandas to solve most complex scientific computing problems with ease. Revised for pandas 1.x. Key Features This is the first book on pandas 1.x Practical, easy to implement recipes for quick solutions to common problems in data using pandas Master the fundamentals of pandas to quickly begin exploring any dataset Book Description The pandas library is massive, and it's common for frequent users to be unaware of many of its more impressive features. The official pandas documentation, while thorough, does not contain many useful examples of how to piece together

multiple commands as one would do during an actual analysis. This book guides you, as if you were looking over the shoulder of an expert, through situations that you are highly likely to encounter. This new updated and revised edition provides you with unique, idiomatic, and fun recipes for both fundamental and advanced data manipulation tasks with pandas. Some recipes focus on achieving a deeper understanding of basic principles, or comparing and contrasting two similar operations. Other recipes will dive deep into a particular dataset, uncovering new and unexpected insights along the way. Many advanced recipes combine several different features across the pandas library to generate results. What you will learn Master data exploration in pandas through dozens of practice problems Group, aggregate, transform, reshape, and filter data Merge data from different sources through pandas SQL-like operations Create visualizations via pandas hooks

to matplotlib and seaborn Use pandas, time series functionality to perform powerful analyses Import, clean, and prepare real-world datasets for machine learning Create workflows for processing big data that doesn't fit in memory Who this book is for This book is for Python developers, data scientists, engineers, and analysts. Pandas is the ideal tool for manipulating structured data with Python and this book provides ample instruction and examples. Not only does it cover the basics required to be proficient, but it goes into the details of idiomatic pandas.

Learning pandas - Michael Heydt 2017-06-30

Get to grips with pandas—a versatile and high-performance Python library for data manipulation, analysis, and discovery About This Book Get comfortable using pandas and Python as an effective data exploration and analysis tool Explore pandas through a framework of data analysis, with an explanation of how

pandas is well suited for the various stages in a data analysis process A comprehensive guide to pandas with many of clear and practical examples to help you get up and using pandas Who This Book Is For This book is ideal for data scientists, data analysts, Python programmers who want to plunge into data analysis using pandas, and anyone with a curiosity about analyzing data. Some knowledge of statistics and programming will be helpful to get the most out of this book but not strictly required. Prior exposure to pandas is also not required. What You Will Learn Understand how data analysts and scientists think about of the processes of gathering and understanding data Learn how pandas can be used to support the end-to-end process of data analysis Use pandas Series and DataFrame objects to represent single and multivariate data Slicing and dicing data with pandas, as well as combining, grouping, and aggregating data from multiple sources How to access

data from external sources such as files, databases, and web services Represent and manipulate time-series data and the many of the intricacies involved with this type of data How to visualize statistical information How to use pandas to solve several common data representation and analysis problems within finance In Detail You will learn how to use pandas to perform data analysis in Python. You will start with an overview of data analysis and iteratively progress from modeling data, to accessing data from remote sources, performing numeric and statistical analysis, through indexing and performing aggregate analysis, and finally to visualizing statistical data and applying pandas to finance. With the knowledge you gain from this book, you will quickly learn pandas and how it can empower you in the exciting world of data manipulation, analysis and science. Style and approach Step-by-step instruction on using pandas within an end-to-end

framework of performing data analysis Practical demonstration of using Python and pandas using interactive and incremental examples **Python Data Visualization Essentials Guide** - Kallur Rahman 2021-07-30

Build your data science skills. Start data visualization Using Python. Right away. Become a good data analyst by creating quality data visualizations using Python. KEY FEATURES ● Exciting coverage on loads of Python libraries, including Matplotlib, Seaborn, Pandas, and Plotly. ● Tons of examples, illustrations, and use-cases to demonstrate visual storytelling of varied datasets. ● Covers a strong fundamental understanding of exploratory data analysis (EDA), statistical modeling, and data mining. DESCRIPTION Data visualization plays a major role in solving data science challenges with various capabilities it offers. This book aims to equip you with a sound knowledge of Python in conjunction with the concepts you need to master to succeed

as a data visualization expert. The book starts with a brief introduction to the world of data visualization and talks about why it is important, the history of visualization, and the capabilities it offers. You will learn how to do simple Python-based visualization with examples with progressive complexity of key features. The book starts with Matplotlib and explores the power of data visualization with over 50 examples. It then explores the power of data visualization using one of the popular exploratory data analysis-oriented libraries, Pandas. The book talks about statistically inclined data visualization libraries such as Seaborn. The book also teaches how we can leverage bokeh and Plotly for interactive data visualization. Each chapter is enriched and loaded with 30+ examples that will guide you in learning everything about data visualization and storytelling of mixed datasets. WHAT YOU WILL LEARN ● Learn to work with popular Python libraries and frameworks, including

Seaborn, Bokeh, and Plotly. ● Practice your data visualization understanding across numerous datasets and real examples. ● Learn to visualize geospatial and time-series datasets. ● Perform correlation and EDA analysis using Pandas and Matplotlib. ● Get to know storytelling of complex and unstructured data using Bokeh and Pandas. ● Learn best practices in writing clean and short python scripts for a quicker visual summary of datasets. WHO THIS BOOK IS FOR This book is for all data analytics professionals, data scientists, and data mining hobbyists who want to be strong data visualizers by learning all the popular Python data visualization libraries. Prior working knowledge of Python is assumed. TABLE OF CONTENTS 1. Introduction to Data Visualization 2. Why Data Visualization 3. Various Data Visualization Elements and Tools 4. Using Matplotlib with Python 5. Using NumPy and Pandas for Plotting 6. Using Seaborn for Visualization 7. Using Bokeh with Python 8.

Using Plotly, Folium, and Other Tools for Data Visualization 9. Hands-on Examples and Exercises, Case Studies, and Further Resources

Data Just Right - Michael Manoochchri 2013

Presents an introduction to data analytics, describing the management of multi-terabyte datasets, such query tools as Hadoop, Hive, and Google BigQuery, the use of R to perform statistical analysis, and advanced data visualization tools.

Data Structures and Algorithms in Python -

Michael T. Goodrich
2013-03-08

Based on the authors' market leading data structures books in Java and C++, this textbook offers a comprehensive, definitive introduction to data structures in Python by authoritative authors. *Data Structures and Algorithms in Python* is the first authoritative object-oriented book available for the Python data structures course. Designed to provide a comprehensive introduction to data structures and algorithms,

including their design, analysis, and implementation, the text will maintain the same general structure as *Data Structures and Algorithms in Java* and *Data Structures and Algorithms in C++*.

Learning the Pandas Library - Matt Harrison 2016-06

Python is one of the top 3 tools that Data Scientists use. One of the tools in their arsenal is the Pandas library. This tool is popular because it gives you so much functionality out of the box. In addition, you can use all the power of Python to make the hard stuff easy! *Learning the Pandas Library* is designed to bring developers and aspiring data scientists who are anxious to learn Pandas up to speed quickly. It starts with the fundamentals of the data structures. Then, it covers the essential functionality. It includes many examples, graphics, code samples, and plots from real world examples. The Content Covers:
Installation
Data Structures Series
CRUD Series
Indexing Series
Methods Series
Plotting Series
Examples
DataFrame

Methods DataFrame Statistics
Grouping, Pivoting, and
Reshaping Dealing with
Missing Data Joining
DataFrames DataFrame
Examples Preliminary Reviews
This is an excellent
introduction benefitting from
clear writing and simple
examples. The pandas
documentation itself is large
and sometimes assumes too
much knowledge, in my
opinion. Learning the Pandas
Library bridges this gap for
new users and even for those
with some pandas experience
such as me. -Garry C. I have
finished reading Learning the
Pandas Library and I liked it...
very useful and helpful tips
even for people who use
pandas regularly. -Tom Z.

**Feature Engineering for
Machine Learning** - Alice
Zheng 2018-03-23

Feature engineering is a
crucial step in the machine-
learning pipeline, yet this topic
is rarely examined on its own.
With this practical book, you'll
learn techniques for extracting
and transforming features—the
numeric representations of raw

data—into formats for machine-
learning models. Each chapter
guides you through a single
data problem, such as how to
represent text or image data.
Together, these examples
illustrate the main principles of
feature engineering. Rather
than simply teach these
principles, authors Alice Zheng
and Amanda Casari focus on
practical application with
exercises throughout the book.
The closing chapter brings
everything together by tackling
a real-world, structured dataset
with several feature-
engineering techniques. Python
packages including numpy,
Pandas, Scikit-learn, and
Matplotlib are used in code
examples. You'll examine:
Feature engineering for
numeric data: filtering,
binning, scaling, log
transforms, and power
transforms Natural text
techniques: bag-of-words, n-
grams, and phrase detection
Frequency-based filtering and
feature scaling for eliminating
uninformative features
Encoding techniques of
categorical variables, including

feature hashing and bin-
counting Model-based feature
engineering with principal
component analysis The

concept of model stacking,
using k-means as a
featurization technique Image
feature extraction with manual
and deep-learning techniques