Pandas For Everyone Python Data Analysis Informit

Pandas dataframe basics -- Pandas data structures -- Introduction to plotting -- Data assembly -- Missing data -- Tidy data -- Data types -- Strings and text data -- Apply -- Groupby operations : split-apply-combine -- The datetime data type -- Linear models -- Generalized linear models -- Model diagnostics -- Regularization -- Clustering -- Life outside of pandas -- Toward a self-directed learner.

The only way to master a skill is to practice. In Python Workout, author Reuven M. Lerner guides you through 50 carefully selected exercises that invite you to flex your programming muscles. As you take on each new challenge, you'll build programming skill and confidence. Summary The only way to master a skill is to practice. In Python Workout, author Reuven M. Lerner guides you through 50 carefully selected exercises that invite you to flex your programming muscles. As you take on each new challenge, you'll build programming skill and confidence. The thorough explanations help you lock in what you've learned and apply it to your own projects. Along the way, Python Workout provides over four hours of video instruction walking you through the solutions to each exercise and dozens of additional exercises for you to try on your own. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology To become a champion Python programmer you need to work out, building mental muscle with your hands on the keyboard. Each carefully selected exercise in this unique book adds to your Python prowess—one important skill at a time. About the book Python Workout presents 50 exercises that focus on key Python 3 features. In it, expert Python coach Reuven Lerner guides you through a series of small projects, practicing the skills you need to tackle everyday tasks. You'll appreciate the clear explanations of each technique, and you can watch Reuven solve each exercise in the accompanying videos. What's inside 50 hands-on exercises and solutions Coverage of all Python data types Dozens more bonus exercises for extra practice About the reader For readers with basic Python knowledge. About the author Reuven M. Lerner teaches Python and data science to companies around the world. Table of Contents 1 Numeric types 2 Strings 3 Lists and tuples 4 Dictionaries and sets 5 Files 6 Functions 7 Functional programming with comprehensions 8 Modules and packages 9 Objects 10 Iterators and generators

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.

Python Data Analytics will help you tackle the world of data acquisition and analysis using the power of the Python language. At the heart of this book lies the coverage of pandas, an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Author Fabio Nelli expertly shows the strength of the Python programming language when applied to processing, managing and retrieving information. Inside, you will see how intuitive and flexible it is to discover and communicate meaningful patterns of data using Python scripts, reporting systems, and data export. This book examines how to go about obtaining, processing, storing, managing and analyzing data using the Python programming language. You will use Python and other open source tools to wrangle data and tease out interesting and important trends in that data that will allow you to predict future patterns. Whether you are dealing with sales data, investment data (stocks, bonds, etc.), medical data, web page usage, or any other type of data set, Python can be used to interpret, analyze, and glean information from a pile of numbers and statistics. This book is an invaluable reference with its examples of storing and accessing data in a database; it walks you through the process of report generation; it provides three real world case studies or examples that you can take with you for your everyday analysis needs.

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

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 DaskWho this book is forThis 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.

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

Get a comprehensive, in-depth introduction to the core Python language with this hands-on book. Based on author Mark Lutz's popular training course, this updated fifth edition will help you quickly write efficient, high-quality code with Python. It's an ideal way to begin, whether you're new to programming or a professional developer versed in other languages. Complete with quizzes, exercises, and helpful illustrations, this easy-to-follow, self-paced tutorial gets you started with both Python 2.7 and 3.3— the latest releases in the 3.X and 2.X lines—plus all other releases in common use today. You'll also learn some advanced language features that recently have become more common in Python code. Explore Python's major built-in object types such as numbers, lists, and dictionaries Create and process objects with Python statements, and learn Python's general syntax model Use functions to avoid code redundancy and package code for reuse Organize statements, functions, and other tools into larger components with modules Dive into classes: Python's object-oriented programming tool for structuring code Write large programs with Python's exception-handling model and development tools Learn advanced Python tools, including decorators, descriptors, metaclasses, and Unicode processing

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

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

Enhance your data analysis and predictive modeling skills using popular Python tools Key Features Cover all fundamental libraries for operation and manipulation of Python for data analysis Implement real-world datasets to perform predictive analytics with Python Access modern data analysis techniques and detailed code with scikit-learn and SciPy Book Description Python is one of the most common and popular languages preferred by leading data analysts and statisticians for working with massive datasets and complex data visualizations. Become a Python Data Analyst introduces Python's most essential tools and libraries necessary to work with the data analysis process, right from preparing data to performing simple statistical analyses and creating meaningful data visualizations. In this book, we will cover Python libraries such as NumPy, pandas, matplotlib, seaborn, SciPy, and scikit-learn, and apply them in practical data analysis and statistics examples. As you make your way through the chapters, you will learn to efficiently use the Jupyter Notebook to operate and manipulate data using NumPy and the pandas library. In the concluding chapters, you will gain experience in building simple predictive models and carrying out statistical computation and analysis using rich Python tools and proven data analysis techniques. By the end of this book, you will have hands-on experience performing data analysis with Python. What you will learn Explore important Python libraries and learn to install Anaconda distribution Understand the basics of NumPy Produce informative and useful visualizations for analyzing data Perform common statistical calculations Build predictive models and understand the principles of predictive analytics Who this book is for Become a Python Data Analyst is for entry-level data analysts, data engineers, and BI professionals who want to make complete use of Python tools for performing efficient data analysis. Prior knowledge of Python programming is necessary to understand the concepts covered in this book

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.

Deal with data, build up financial formulas in code from scratch, and evaluate and think about money in your day-to-day life. This book is about Python and personal finance and how you can effectively mix the two together. In Personal Finance with Python you will learn Python and finance at the same time by creating a profit calculator, a currency converter, an amortization schedule, a budget, a portfolio rebalancer, and a purchase forecaster. Many of the examples use pandas, the main data manipulation tool in Python. Each chapter is hands-on, self-contained, and motivated by fun and interesting examples. Although this book assumes a minimal familiarity with programming and the Python language, if you don't have any, don't worry. Everything is built up piece-by-piece and the first chapters are conducted at a relaxed pace. You'll need Python 3.6 (or above) and all of the setup details are included. What You'll Learn Work with data in pandas Calculate Net Present Value and Internal Rate Return Query a third-party API with Requests Manage secrets Build efficient loops Parse English sentences with Recurrent Work with the YAML file format Fetch stock quotes and use Prophet to forecast the future Who This Book Is For Anyone interested in Python, personal finance, and/or both! This book is geared towards those who want to manage their money more effectively and to those who just want to learn or improve their Python.

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. Pandas in Action shows you how to master this versatile tool and take the next steps in your data science career. 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. Pandas in Action shows you how to master this versatile tool and take the next steps in your data science career. 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. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

Learn how to use JupyterLab, Numpy, pandas, Scipy, Matplotlib, and Seaborn for Data science KEY FEATURES ? Get familiar with different inbuilt Data structures, Functional programming, and Datetime objects. ? Handling heavy Datasets to optimize the data types for memory management, reading files in chunks, dask, and modin pandas. ? Time-series analysis to find trends, seasonality, and cyclic components. ?

Seaborn to build aesthetic plots with high-level interfaces and customized themes. ? Exploratory data analysis with real-time datasets to maximize the insights about data. DESCRIPTION The book will start with quick introductions to Python and its ecosystem libraries for data science such as JupyterLab, Numpy, Pandas, SciPy, Matplotlib, and Seaborn. This book will help in learning python data structures and essential concepts such as Functions, Lambdas, List comprehensions, Datetime objects, etc. required for data engineering. It also covers an in-depth understanding of Python data science packages where JupyterLab used as an IDE for writing, documenting, and executing the python code, Numpy used for computation of numerical operations, Pandas for cleaning and reorganizing the data, handling large datasets and merging the dataframes to get meaningful insights. You will go through the statistics to understand the relation between the variables using SciPy and building visualization charts using Matplotllib and Seaborn libraries. WHAT WILL YOU LEARN ? Learn about Python data containers, their methods, and attributes. ? Learn Numpy arrays for the computation of numerical data. ? Learn Pandas data structures, DataFrames, and Series. ? Learn statistics measures of central tendency, central limit theorem, confidence intervals, and hypothesis testing. ? A brief understanding of visualization, control, and draw different inbuilt charts to extract important variables, detect outliers, and anomalies using Matplotlib and Seaborn. WHO THIS BOOK IS FOR This book is for anyone who wants to use Python for Data Analysis and Visualization. This book is for novices as well as experienced readers with working knowledge of the pandas library. Basic knowledge of Python is a must. TABLE OF CONTENTS 1. Introduction to Data Analysis 2. Jupyter lab 3. Python overview 4. Introduction to Numpy 5. Introduction to Pandas 6. Data Analysis 7. Time-Series Analysis 8. Introduction to Statistics 9. Matplotlib 10. Seaborn 11. Exploratory Data Analysis

Explore the latest Python tools and techniques to help you tackle the world of data acquisition and analysis. You'll review scientific computing with NumPy, visualization with matplotlib, and machine learning with scikit-learn. This revision is fully updated with new content on social media data analysis, image analysis with OpenCV, and deep learning libraries. Each chapter includes multiple examples demonstrating how to work with each library. At its heart lies the coverage of pandas, for high-performance, easy-to-use data structures and tools for data manipulation Author Fabio Nelli expertly demonstrates using Python for data processing, management, and information retrieval. Later chapters apply what you've learned to handwriting recognition and extending graphical capabilities with the JavaScript D3 library. Whether you are dealing with sales data, investment data, medical data, web page usage, or other data sets, Python Data Analytics, Second Edition is an invaluable reference with its examples of storing, accessing, and analyzing data. What You'll Learn Understand the core concepts of data analysis and the Python ecosystem Go in depth with pandas for reading, writing, and processing data Use tools and techniques for data visualization and image analysis Examine popular deep learning libraries Keras, Theano,TensorFlow, and PyTorch Who This Book Is For Experienced Python developers who need to learn about Pythonic tools for data analysis

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.

Understand and implement big data analysis solutions in pandas with an emphasis on performance. This book strengthens your intuition for working with pandas, the Python data analysis library, by exploring its underlying implementation and data structures. Thinking in Pandas introduces the topic of big data and demonstrates concepts by looking at exciting and impactful projects that pandas helped to solve. From there, you will learn to assess your own projects by size and type to see if pandas is the appropriate library for your needs. Author Hannah Stepanek explains how to load and normalize data in pandas efficiently, and reviews some of the most commonly used loaders and several of their most powerful options. You will then learn how to access and transform data efficiently, what methods to avoid, and when to employ more advanced performance techniques. You will also go over basic data access and munging in pandas and the intuitive dictionary syntax. Choosing the right DataFrame format, working with multi-level DataFrames, and how pandas might be improved upon in the future are also covered. By the end of the book, you will have a solid understanding of how the pandas library works under the hood. Get ready to make confident decisions in your own projects by utilizing pandas—the right way. What You Will Learn Understand the underlying data structure of pandas and why it performs the way it does under certain circumstances Discover how to use pandas to extract, transform, and load data correctly with an emphasis on performance Choose the right DataFrame so that the data analysis is simple and efficient. Improve performance of pandas operations with other Python libraries Who This Book Is ForSoftware engineers with basic programming skills in Python keen on using pandas for a big data analysis project. Python software developers interested in big data.

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

Data is collected everywhere these days, in massive quantities. But data alone does not do you much good. That is why data analysis -- making sense of the data -- has become a must-have skill in the fields of business, science, and social science. But it is a tough skill to acquire. The concepts are challenging, and too many books and online tutorials treat only parts of the total skillset needed. Now, though, this book draws all the essential skills together and presents them in a clear and example-packed way. So you will soon be applying your programming skills to complex data analysis

problems, more easily than you ever thought possible. In terms of content, this book gets you started the right way by using Pandas for data analysis and Seaborn for data visualisation, with JupyterLab as your IDE. Then, it helps you master descriptive analysis by teaching you how to get, clean, prepare, and analyse data, including time-series data. Next, it gets you started with predictive analysis by showing you how to use linear regression models to predict unknown and future values. And to tie everything together, it gives you 4 real-world case studies that show you how to apply your new skills to political, environmental, social, and sports analysis. At the end, you will have a solid set of the professional skills that can lead to all sorts of new career opportunities. Sound too good to be true? Download a sample chapter for free from the Murach website and see for yourself how this book can turn you into the data analyst that companies are looking for.

While Excel remains ubiquitous in the business world, recent Microsoft feedback forums are full of requests to include Python as an Excel scripting language. In fact, it's the top feature requested. What makes this combination so compelling? In this hands-on guide, Felix Zumstein--creator of xlwings, a popular open source package for automating Excel with Python--shows experienced Excel users how to integrate these two worlds efficiently. Excel has added quite a few new capabilities over the past couple of years, but its automation language, VBA, stopped evolving a long time ago. Many Excel power users have already adopted Python for daily automation tasks. This guide gets you started. Use Python without extensive programming knowledge Get started with modern tools, including Jupyter notebooks and Visual Studio code Use pandas to acquire, clean, and analyze data and replace typical Excel calculations Automate tedious tasks like consolidation of Excel workbooks and production of Excel reports Use xlwings to build interactive Excel tools that use Python as a calculation engine Connect Excel to databases and CSV files and fetch data from the internet using Python code Use Python as a single tool to replace VBA, Power Query, and Power Pivot

Knowing how to work with data to extract insights generates significant value. This book will help you to develop data analysis skills using a hands-on approach and real-world data. You'll get up to speed with pandas 1.x in no time and build some software engineering skills in the process, vastly expanding your data science toolbox.

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

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—IPython, 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

Become an expert at using Python for advanced statistical analysis of data using real-world examples About This Book Clean, format, and explore data using graphical and numerical summaries Leverage the IPython environment to efficiently analyze data with Python Packed with easy-to-follow examples to develop advanced computational skills for the analysis of complex data Who This Book Is For If you are a competent Python developer who wants to take your data analysis skills to the next level by solving complex problems, then this advanced guide is for you. Familiarity with the basics of applying Python libraries to data sets is assumed. What You Will Learn Read, sort, and map various data into Python and Pandas Recognise patterns so you can understand and explore data Use statistical models to discover patterns in data Review classical statistical inference using Python, Pandas, and SciPy Detect similarities and differences in data with clustering Clean your data to make it useful Work in Jupyter Notebook to produce publication ready figures to be included in reports In Detail Python, a multi-paradigm programming language, has become the language of choice for data scientists for data analysis, visualization, and machine learning. Ever imagined how to become an expert at effectively approaching data analysis problems, solving them, and extracting all of the available information from your data? Well, look no further, this is the book you want! Through this comprehensive guide, you will explore data and present results and conclusions from statistical analysis in a meaningful way. You'll be able to quickly and accurately perform the hands-on sorting, reduction, and subsequent analysis, and fully appreciate how data analysis methods can support business

decision-making. You'll start off by learning about the tools available for data analysis in Python and will then explore the statistical models that are used to identify patterns in data. Gradually, you'll move on to review statistical inference using Python, Pandas, and SciPy. After that, we'll focus on performing regression using computational tools and you'll get to understand the problem of identifying clusters in data in an algorithmic way. Finally, we delve into advanced techniques to quantify cause and effect using Bayesian methods and you'll discover how to use Python's tools for supervised machine learning. Style and approach This book takes a step-by-step approach to reading, processing, and analyzing data in Python using various methods and tools. Rich in examples, each topic connects to real-world examples and retrieves data directly online where possible. With this book, you are given the knowledge and tools to explore any data on your own, encouraging a curiosity befitting all data scientists.

The book shows you how to view data from multiple perspectives, including data frame and column attributes. You will cover common and not-so-common challenges that are faced while cleaning messy data for complex situations. You will learn to manipulate data and get them down to a form that can be useful for making the right decisions.

Python for Everybody is designed to introduce students to programming and software development through the lens of exploring data. You can think of the Python programming language as your tool to solve data problems that are beyond the capability of a spreadsheet.Python is an easy to use and easy to learn programming language that is freely available on Macintosh, Windows, or Linux computers. So once you learn Python you can use it for the rest of your career without needing to purchase any software.This book uses the Python 3 language. The earlier Python 2 version of this book is titled "Python for Informatics: Exploring Information".There are free downloadable electronic copies of this book in various formats and supporting materials for the book at www.pythonlearn.com. The course materials are available to you under a Creative Commons License so you can adapt them to teach your own Python course.

Treading on Python is designed to bring developers and others who are anxious to learn Python up to speed quickly. Not only does it teach the basics of syntax, but it condenses years of experience. You will learn warts, gotchas, best practices and hints that have been gleaned through the years in days. You will hit the ground running and running in the right way.

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

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. This book is a guide for you on how to use Pandas and Numpy in Python programming language for data analysis. The author begins by helping you familiarize yourself with the basics of data science, Numpy and Pandas. You are guided on how to work with Numpy arrays and how to manipulate them. The various operations that you can perform on your data via the Pandas library have been discussed. You will also know how to create various data structures in Pandas for data storage. Data from the environment is dirty. The process of cleaning such data has been discussed. This involves handling outliers, missing values etc. The author guides you on how to work with data in various types of storage formats. Examples include MS Excel, CSV files, JSON, etc. You are also guided on how to calculate various measures for your data. The process of visualizing data has been explored in detail. About this book: Getting Started with Python for Data Science Working with Numpy Working with Pandas Cleansing Data Working with CSV Data Working with XLS Data Data Wrangling Measures of Central Tendency Calculating Variance Normal Distribution Working with JSON Data Data Visualization Tags: data science with python, python, pandas programming, numpy, pandas, pandas python, pandas in python, numpy in python, numpy python, numpy pandas, data science, ms excel books, json, python for data science, pivot tables, excel pivot tables, data visualisation, data visualisation python, data visualisation for dummies, data visualisation excel, algorithms for data science.

Master these effective techniques to reduce costs and run times, handle huge datasets, and implement complex machine learning applications efficiently in Python. Fast Python for Data Science is a hands-on guide to writing Python code that can process more data, faster, and with less resources. It takes a holistic approach to Python performance, showing you how your code, libraries, and computing architecture interact and can be optimized together. Written for experienced practitioners, Fast Python for Data Science dives right into practical solutions for improving computation and storage efficiency. You'll experiment with fun and interesting examples such as rewriting games in lower-level Cython and implementing a MapReduce framework from scratch. Finally, you'll go deep into Python GPU computing and learn how modern hardware has rehabilitated some former antipatterns and made counterintuitive ideas the most efficient way of working. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

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.

Pandas for EveryonePython Data AnalysisAddison-Wesley Professional

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 Register your product at informit.com/register for convenient access to downloads, updates, and/or corrections as they become available.

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.

Data science libraries, frameworks, modules, and toolkits are great for doing data science, but they're also a good way to dive into

the discipline without actually understanding data science. In this book, you'll learn how many of the most fundamental data science tools and algorithms work by implementing them from scratch. If you have an aptitude for mathematics and some programming skills, author Joel Grus will help you get comfortable with the math and statistics at the core of data science, and with hacking skills you need to get started as a data scientist. Today's messy glut of data holds answers to questions no one's even thought to ask. This book provides you with the know-how to dig those answers out. Get a crash course in Python Learn the basics of linear algebra, statistics, and probability—and understand how and when they're used in data science Collect, explore, clean, munge, and manipulate data Dive into the fundamentals of machine learning Implement models such as k-nearest Neighbors, Naive Bayes, linear and logistic regression, decision trees, neural networks, and clustering Explore recommender systems, natural language processing, network analysis, MapReduce, and databases

"Eric and Russell were early adopters of Cassandra at SimpleReach. In Practical Cassandra, you benefit from their experience in the trenches administering Cassandra, developing against it, and building one of the first CQL drivers. If you are deploying Cassandra soon, or you inherited a Cassandra cluster to tend, spend some time with the deployment, performance tuning, and maintenance chapters… If you are new to Cassandra, I highly recommend the chapters on data modeling and CQL." –From the Foreword by Jonathon Ellis, Apache Cassandra Chair Build and Deploy Massively Scalable, Super-fast Data Management Applications with Apache Cassandra Practical Cassandra is the first hands-on developer's guide to building Cassandra systems and applications that deliver breakthrough speed, scalability, reliability, and performance. Fully up to date, it reflects the latest versions of Cassandra–including Cassandra Query Language (CQL), which dramatically lowers the learning curve for Cassandra developers. Pioneering Cassandra developers and Datastax MVPs Russell Bradberry and Eric Lubow walk you through every step of building a real production application that can store enormous amounts of structured, semi-structured, and unstructured data. Drawing on their exceptional expertise, Bradberry and Lubow share practical insights into issues ranging from querying to deployment, management, maintenance, monitoring, and troubleshooting. The authors cover key issues, from architecture to migration, and guide you through crucial decisions about configuration and data modeling. They provide tested sample code, detailed explanations of how Cassandra works "under the covers," and new case studies from three cutting-edge users: Ooyala, Hailo, and eBay. Coverage includes Understanding Cassandra's approach, architecture, key concepts, and primary use cases–and why it's so blazingly fast Getting Cassandra up and running on single nodes and large clusters Applying the new design patterns, philosophies, and features that make Cassandra such a powerful data store Leveraging CQL to simplify your transition from SQL-based RDBMSes Deploying and provisioning through the cloud or on bare-metal hardware Choosing the right configuration options for each type of workload Tweaking Cassandra to get maximum performance from your hardware, OS, and JVM Mastering Cassandra's essential tools for maintenance and monitoring Efficiently solving the most common problems with Cassandra deployment, operation, and application development

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.

Machine learning has become an integral part of many commercial applications and research projects, but this field is not exclusive to large companies with extensive research teams. If you use Python, even as a beginner, this book will teach you practical ways to build your own machine learning solutions. With all the data available today, machine learning applications are limited only by your imagination. You'll learn the steps necessary to create a successful machine-learning application with Python and the scikit-learn library. Authors Andreas Müller and Sarah Guido focus on the practical aspects of using machine learning algorithms, rather than the math behind them. Familiarity with the NumPy and matplotlib libraries will help you get even more from this book. With this book, you'll learn: Fundamental concepts and applications of machine learning Advantages and shortcomings of widely used machine learning algorithms How to represent data processed by machine learning, including which data aspects to focus on Advanced methods for model evaluation and parameter tuning The concept of pipelines for chaining models and encapsulating your workflow Methods for working with text data, including text-specific processing techniques Suggestions for improving your machine learning and data science skills