

---

# File Type PDF Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

---

Getting the books **Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics** now is not type of challenging means. You could not lonely going afterward ebook collection or library or borrowing from your friends to door them. This is an definitely simple means to specifically get lead by on-line. This online message Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics can be one of the options to accompany you gone having additional time.

It will not waste your time. admit me, the e-book will unconditionally circulate you other matter to read. Just invest tiny time to gate this on-line broadcast **Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics** as with ease as review them wherever you are now.

---

## 6D3 - GIANNA ADELAIDE

---

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 Cre-

ate 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 Understand, evaluate, and visualize data About This Book Learn basic steps of data analysis and how to use Python and its packages A step-by-step guide to predictive modeling including tips, tricks, and best practices Effectively visualize a broad set of analyzed data and generate effective results Who This Book Is For This book is for Python Developers who are keen to get into data analysis and wish to visualize their analyzed data in a more efficient and insightful manner. What You Will Learn Get acquainted with NumPy and use arrays and array-oriented computing in data analysis Process and analyze data using the time-series capabilities of Pandas Understand the statistical and mathematical concepts behind predictive analytics algorithms Data

visualization with Matplotlib Interactive plotting with NumPy, Scipy, and MKL functions Build financial models using Monte-Carlo simulations Create directed graphs and multi-graphs Advanced visualization with D3 In Detail You will start the course with an introduction to the principles of data analysis and supported libraries, along with NumPy basics for statistics and data processing. Next, you will overview the Pandas package and use its powerful features to solve data-processing problems. Moving on, you will get a brief overview of the Matplotlib API .Next, you will learn to manipulate time and data structures, and load and store data in a file or database using Python packages. You will learn how to apply powerful packages in Python to process raw data into pure and helpful data using examples. You will also get a brief overview of machine learning algorithms, that is, applying data analysis results to make decisions or building helpful products such as recommendations and predictions using Scikit-learn. After this, you will move on to a data analytics specialization—predictive analytics. Social media and IOT have resulted in an avalanche of data. You will get started with predictive analytics using Python. You will see how to create predictive models from data. You will get balanced information on statistical and mathematical concepts, and implement them in Python using libraries such as Pandas, scikit-learn, and NumPy. You'll learn more about the best predictive modeling algorithms such as Linear Regression, Decision Tree, and Logistic Regression. Finally, you will master best practices in predictive modeling. After this, you will get all the practical guidance you need to help you on the journey to effective data visualization. Starting with a chapter on data frameworks, which ex-

plains the transformation of data into information and eventually knowledge, this path subsequently cover the complete visualization process using the most popular Python libraries with working examples This Learning Path combines some of the best that Packt has to offer in one complete, curated package. It includes content from the following Packt products: Getting Started with Python Data Analysis, Phuong Vo.T.H &Martin Czygan Learning Predictive Analytics with Python, Ashish Kumar Mastering Python Data Visualization, Kirthi Raman Style and approach The course acts as a step-by-step guide to get you familiar with data analysis and the libraries supported by Python with the help of real-world examples and datasets. It also helps you gain practical insights into predictive modeling by implementing predictive-analytics algorithms on public datasets with Python. The course offers a wealth of practical guidance to help you on this journey to data visualization

Summary Dask is a native parallel analytics tool designed to integrate seamlessly with the libraries you're already using, including Pandas, NumPy, and Scikit-Learn. With Dask you can crunch and work with huge datasets, using the tools you already have. And Data Science with Python and Dask is your guide to using Dask for your data projects without changing the way you work! Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. You'll find registration instructions inside the print book. About the Technology An efficient data pipeline means everything for the success of a data science project. Dask is a flexible library for parallel computing in Python that makes it easy to build intuitive workflows for ingesting and analyzing large, distributed datasets. Dask pro-

vides dynamic task scheduling and parallel collections that extend the functionality of NumPy, Pandas, and Scikit-learn, enabling users to scale their code from a single laptop to a cluster of hundreds of machines with ease. About the Book Data Science with Python and Dask teaches you to build scalable projects that can handle massive datasets. After meeting the Dask framework, you'll analyze data in the NYC Parking Ticket database and use DataFrames to streamline your process. Then, you'll create machine learning models using Dask-ML, build interactive visualizations, and build clusters using AWS and Docker. What's inside Working with large, structured and unstructured datasets Visualization with Seaborn and Datashader Implementing your own algorithms Building distributed apps with Dask Distributed Packaging and deploying Dask apps About the Reader For data scientists and developers with experience using Python and the PyData stack. About the Author Jesse Daniel is an experienced Python developer. He taught Python for Data Science at the University of Denver and leads a team of data scientists at a Denver-based media technology company. Table of Contents PART 1 - The Building Blocks of scalable computing Why scalable computing matters Introducing Dask PART 2 - Working with Structured Data using Dask DataFrames Introducing Dask DataFrames Loading data into DataFrames Cleaning and transforming DataFrames Summarizing and analyzing DataFrames Visualizing DataFrames with Seaborn Visualizing location data with Datashader PART 3 - Extending and deploying Dask Working with Bags and Arrays Machine learning with Dask-ML Scaling and deploying Dask Become an efficient data science practitioner by understanding Python's key

concepts About This Book Quickly get familiar with data science using Python 3.5 Save time (and effort) with all the essential tools explained Create effective data science projects and avoid common pitfalls with the help of examples and hints dictated by experience Who This Book Is For If you are an aspiring data scientist and you have at least a working knowledge of data analysis and Python, this book will get you started in data science. Data analysts with experience of R or MATLAB will also find the book to be a comprehensive reference to enhance their data manipulation and machine learning skills. What You Will Learn Set up your data science toolbox using a Python scientific environment on Windows, Mac, and Linux Get data ready for your data science project Manipulate, fix, and explore data in order to solve data science problems Set up an experimental pipeline to test your data science hypotheses Choose the most effective and scalable learning algorithm for your data science tasks Optimize your machine learning models to get the best performance Explore and cluster graphs, taking advantage of interconnections and links in your data In Detail Fully expanded and upgraded, the second edition of Python Data Science Essentials takes you through all you need to know to succeed in data science using Python. Get modern insight into the core of Python data, including the latest versions of Jupyter notebooks, NumPy, pandas and scikit-learn. Look beyond the fundamentals with beautiful data visualizations with Seaborn and ggplot, web development with Bottle, and even the new frontiers of deep learning with Theano and TensorFlow. Dive into building your essential Python 3.5 data science toolbox, using a single-source approach that will allow to to work with Python 2.7 as well.

Get to grips fast with data munging and preprocessing, and all the techniques you need to load, analyse, and process your data. Finally, get a complete overview of principal machine learning algorithms, graph analysis techniques, and all the visualization and deployment instruments that make it easier to present your results to an audience of both data science experts and business users. Style and approach The book is structured as a data science project. You will always benefit from clear code and simplified examples to help you understand the underlying mechanics and real-world datasets.

If you are interested in quantitative finance, financial modeling, and trading, or simply want to learn how Python and pandas can be applied to finance, then this book is ideal for you. Some knowledge of Python and pandas is assumed. Interest in financial concepts is helpful, but no prior knowledge is expected.

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.

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](http://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. 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.

Discover techniques to summarize the characteristics of your data using PyPlot, NumPy, SciPy, and pandas Key Features Understand the fundamental concepts of exploratory data analysis using Python Find missing values in your data and identify the correlation between different variables Practice graphical exploratory analysis techniques using Matplotlib and the Seaborn Python package Book Description Exploratory Data Analysis (EDA) is an approach to data analysis that involves the application of diverse techniques to gain insights into a dataset. This book will help you gain practical knowledge of the main pillars of EDA - data cleaning, data preparation, data exploration, and data visualization. You'll start by performing EDA using open source datasets and perform simple to advanced analyses to turn data into meaningful insights. You'll then learn various descriptive statistical techniques to describe the basic characteristics of data and progress to performing EDA on time-series data. As you advance, you'll learn how to implement EDA techniques for model development and evaluation and build predictive models to visualize results. Using Python for data analysis, you'll work with real-world datasets, understand data, summarize its characteristics, and visualize it for business intelligence. By the end of this EDA book, you'll have developed the skills required to carry out a preliminary investigation on any dataset, yield insights into data,

present your results with visual aids, and build a model that correctly predicts future outcomes. What you will learn

- Import, clean, and explore data to perform preliminary analysis using powerful Python packages
- Identify and transform erroneous data using different data wrangling techniques
- Explore the use of multiple regression to describe non-linear relationships
- Discover hypothesis testing and explore techniques of time-series analysis
- Understand and interpret results obtained from graphical analysis
- Build, train, and optimize predictive models to estimate results
- Perform complex EDA techniques on open source datasets

Who this book is for This EDA book is for anyone interested in data analysis, especially students, statisticians, data analysts, and data scientists. The practical concepts presented in this book can be applied in various disciplines to enhance decision-making processes with data analysis and synthesis. Fundamental knowledge of Python programming and statistical concepts is all you need to get started with 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.

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 de-

veloper 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 ad-

vanced 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.

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

Practice makes perfect pandas. Work out your pandas skills against dozens of real-world challenges, each carefully designed to build an intuitive knowledge of essential pandas tasks. Practice makes perfect pandas. Work out your pandas skills against dozens of real-world challenges, each carefully designed to build an intuitive knowledge of essential pandas tasks. In *Pandas Workout*, discover 50 exercises that will strengthen your pandas skills to a level of automatic fluency. You'll test yourself against common pandas challenges such as data cleaning, and explore real-world datasets such as New York Taxis, Kickstarter projects, and global tourist spending. Detailed explanations help guide your success and make your new skills stick. You'll even get a big boost to productivity, with tasks that used to mean a trip to StackOverflow now a natural part of your skillset. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

Getting started with data science doesn't have to be an uphill battle. This step-by-step guide is ideal for beginners who know a little Python and are looking for a quick, fast-paced introduction. Key Features Get up and running with the Jupyter ecosystem and some example datasets Learn about key machine learning concepts like SVM, KNN classifiers and Random Forests Discover how you can use web scraping to gather and parse your own bespoke datasets **Book Description** Get to grips with the skills you need for entry-level data science in this hands-on Python and Jupyter course. You'll learn about some of the most commonly used libraries that are part of the Anaconda distribution, and then explore ma-

chine learning models with real datasets to give you the skills and exposure you need for the real world. We'll finish up by showing you how easy it can be to scrape and gather your own data from the open web, so that you can apply your new skills in an actionable context. What you will learn Get up and running with the Jupyter ecosystem and some example datasets Learn about key machine learning concepts like SVM, KNN classifiers, and Random Forests Plan a machine learning classification strategy and train classification, models Use validation curves and dimensionality reduction to tune and enhance your models Discover how you can use web scraping to gather and parse your own bespoke datasets Scrape tabular data from web pages and transform them into Pandas DataFrames Create interactive, web-friendly visualizations to clearly communicate your findings **Who this book is for** This book is ideal for professionals with a variety of job descriptions across large range of industries, given the rising popularity and accessibility of data science. You'll need some prior experience with Python, with any prior work with libraries like Pandas, Matplotlib and Pandas providing you a useful head start.

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 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 Discover how to describe your data in de-

tail, identify data issues, and find out how to solve them using commonly used techniques and tips and tricks Key Features Get well-versed with various data cleaning techniques to reveal key insights Manipulate data of different complexities to shape them into the right form as per your business needs Clean, monitor, and validate large data volumes to diagnose problems before moving on to data analysis Book Description Getting clean data to reveal insights is essential, as directly jumping into data analysis without proper data cleaning may lead to incorrect results. This book shows you tools and techniques that you can apply to clean and handle data with Python. You'll begin by getting familiar with the shape of data by using practices that can be deployed routinely with most data sources. Then, the book teaches you how to manipulate data to get it into a useful form. You'll also learn how to filter and summarize data to gain insights and better understand what makes sense and what does not, along with discovering how to operate on data to address the issues you've identified. Moving on, you'll perform key tasks, such as handling missing values, validating errors, removing duplicate data, monitoring high volumes of data, and handling outliers and invalid dates. Next, you'll cover recipes on using supervised learning and Naive Bayes analysis to identify unexpected values and classification errors, and generate visualizations for exploratory data analysis (EDA) to visualize unexpected values. Finally, you'll build functions and classes that you can reuse without modification when you have new data. By the end of this Python book, you'll be equipped with all the key skills that you need to clean data and diagnose problems within it. What you will learn Find out how to read and analyze

data from a variety of sources Produce summaries of the attributes of data frames, columns, and rows Filter data and select columns of interest that satisfy given criteria Address messy data issues, including working with dates and missing values Improve your productivity in Python pandas by using method chaining Use visualizations to gain additional insights and identify potential data issues Enhance your ability to learn what is going on in your data Build user-defined functions and classes to automate data cleaning Who this book is for This book is for anyone looking for ways to handle messy, duplicate, and poor data using different Python tools and techniques. The book takes a recipe-based approach to help you to learn how to clean and manage data. Working knowledge of Python programming is all you need to get the most out of the book.

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 Da-

ta Visualization 9. Hands-on Examples and Exercises, Case Studies, and Further Resources

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.

In this "important and comprehensive" guide to statistical thinking (New Yorker), discover how data literacy is changing the world and gives you a better understanding of life's biggest problems. Statistics are everywhere, as integral to science as they are to business, and in the popular media hundreds of times a day. In this age of big data, a basic grasp of statistical literacy is more important than ever if we want to separate the fact from the fiction, the ostentatious embellishments from the raw evidence -- and even more so if we hope to participate in the future, rather than being simple bystanders. In *The Art of Statistics*, world-renowned statistician David Spiegelhalter shows readers how to derive knowledge from raw data by focusing on the concepts and connections behind the math. Drawing on real world examples to introduce complex issues, he shows us how statistics can help us determine the luckiest passenger on the Titanic, whether a notorious serial killer could have been caught earlier, and if screening for ovarian cancer is beneficial. *The Art of Statistics* not only shows us how mathematicians have used statistical science to solve these problems -- it teaches us how we too can think like statisticians. We learn how to clarify our questions, assumptions, and expectations

when approaching a problem, and -- perhaps even more importantly -- we learn how to responsibly interpret the answers we receive. Combining the incomparable insight of an expert with the playful enthusiasm of an aficionado, *The Art of Statistics* is the definitive guide to stats that every modern person needs.

*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.

Statistical Computation for Programmers, Scientists, Quants, Excel Users, and Other Professionals Using the open source R language, you can build powerful statistical models to answer many of your most challenging questions. R has traditionally been difficult for non-statisticians to learn, and most R books assume far too much knowledge to be of help. R for Everyone, Second Edition, is the solution. Drawing on his unsurpassed experience teaching new users, professional data scientist Jared P. Lander has written the perfect tutorial for anyone new to statistical programming and modeling. Organized to make learning easy and intuitive, this guide focuses on the 20 percent of R functionality you'll need to accomplish 80 percent of modern data tasks. Lander's self-contained chapters start with the absolute basics, offering extensive hands-on practice and sample code. You'll download and install R; navigate and use the R environment; master basic program control, data import, manipulation, and visualization; and walk through several essential tests. Then, building on this foundation, you'll construct several complete models, both linear and nonlinear, and use some data mining techniques. After all this you'll make your code reproducible with LaTeX, RMarkdown, and Shiny. By the time you're done, you won't just know how to write R programs, you'll be ready to tackle the statistical problems you care about most. Coverage includes Explore R, RStudio, and R packages Use R for math: variable types, vectors, calling functions, and more Exploit data structures, including data.frames, matrices, and lists Read many different types of data Create attractive, intuitive statistical graphics Write user-defined functions Control program flow with if, ifelse, and complex checks Improve program effi-

ciency with group manipulations Combine and reshape multiple datasets Manipulate strings using R's facilities and regular expressions Create normal, binomial, and Poisson probability distributions Build linear, generalized linear, and nonlinear models Program basic statistics: mean, standard deviation, and t-tests Train machine learning models Assess the quality of models and variable selection Prevent overfitting and perform variable selection, using the Elastic Net and Bayesian methods Analyze univariate and multivariate time series data Group data via K-means and hierarchical clustering Prepare reports, slideshows, and web pages with knitr Display interactive data with RMarkdown and htmlwidgets Implement dashboards with Shiny Build reusable R packages with devtools and Rcpp Register your product at [informit.com/register](http://informit.com/register) for convenient access to downloads, updates, and 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.

The second edition of this best-selling Python book (over 500,000 copies sold!) uses Python 3 to teach even the technically uninclined how to write programs that do in minutes what would take hours to do by hand. There is no prior programming experience required and the book is loved by liberal arts majors and geeks alike. If you've ever spent hours renaming files or updating hundreds of spreadsheet cells, you know how tedious tasks like these can be. But what if you could have your computer do them for you? In this fully revised second edition of the best-selling classic Automate the Boring Stuff with Python, you'll learn how to use Python to write programs that do in minutes what would take you hours to do by hand--no prior

programming experience required. You'll learn the basics of Python and explore Python's rich library of modules for performing specific tasks, like scraping data off websites, reading PDF and Word documents, and automating clicking and typing tasks. The second edition of this international fan favorite includes a brand-new chapter on input validation, as well as tutorials on automating Gmail and Google Sheets, plus tips on automatically updating CSV files. You'll learn how to create programs that effortlessly perform useful feats of automation to:

- Search for text in a file or across multiple files
- Create, update, move, and rename files and folders
- Search the Web and download online content
- Update and format data in Excel spreadsheets of any size
- Split, merge, watermark, and encrypt PDFs
- Send email responses and text notifications
- Fill out online forms

Step-by-step instructions walk you through each program, and updated practice projects at the end of each chapter challenge you to improve those programs and use your newfound skills to automate similar tasks. Don't spend your time doing work a well-trained monkey could do. Even if you've never written a line of code, you can make your computer do the grunt work. Learn how in Automate the Boring Stuff with Python, 2nd Edition.

Explore the fundamentals of data analysis, and statistics with case studies using Python. This book will show you how to confidently write code in Python, and use various Python libraries and functions for analyzing any dataset. The code is presented in Jupyter notebooks that can further be adapted and extended. This book is divided into three parts - programming with Python, data analysis and visualization, and statistics. You'll start with an introduction to Python - the syntax,

functions, conditional statements, data types, and different types of containers. You'll then review more advanced concepts like regular expressions, handling of files, and solving mathematical problems with Python. The second part of the book, will cover Python libraries used for data analysis. There will be an introductory chapter covering basic concepts and terminology, and one chapter each on NumPy(the scientific computation library), Pandas (the data wrangling library) and visualization libraries like Matplotlib and Seaborn. Case studies will be included as examples to help readers understand some real-world applications of data analysis. The final chapters of book focus on statistics, elucidating important principles in statistics that are relevant to data science. These topics include probability, Bayes theorem, permutations and combinations, and hypothesis testing (ANOVA, Chi-squared test, z-test, and t-test), and how the Scipy library enables simplification of tedious calculations involved in statistics. What You'll Learn Further your programming and analytical skills with Python Solve mathematical problems in calculus, and set theory and algebra with Python Work with various libraries in Python to structure, analyze, and visualize data Tackle real-life case studies using Python Review essential statistical concepts and use the Scipy library to solve problems in statistics Who This Book Is For Professionals working in the field of data science interested in enhancing skills in Python, data analysis and statistics.

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 Date-time 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, Date-time 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 Matplotlib 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 Da-

ta 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

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

The goal of this hands-on book is to teach you, the reader, the ins and outs of data manipulation and analytics using Pandas. It introduces comprehensive and in-depth data analytics techniques with Pandas, and helps you quickly master data manipulation and analytical skills, whether you are new to Python or an experience data scientist using other languages. Throughout this book, three small, carefully designed data sets are used to explain all of Pandas' functions. At the end, a detailed use case is provided to demonstrate how Pandas works with Python in real-life applications. If you are new to data analytics, you will find that this book explains complex concepts in a simple yet effective way, assisted with visual explanations. As for those who are experienced data engineers or data scientists, this book will be your best friend when working on projects requiring Pandas. In short, this book is for everyone, regardless of skill level.

This is the eBook of the printed book and may not include any media, website ac-

cess codes, or print supplements that may come packaged with the bound book. The Comprehensive, Up-to-Date Apache Hadoop Administration Handbook and Reference “Sam Alapati has worked with production Hadoop clusters for six years. His unique depth of experience has enabled him to write the go-to resource for all administrators looking to spec, size, expand, and secure production Hadoop clusters of any size.” —Paul Dix, Series Editor In Expert Hadoop® Administration, leading Hadoop administrator Sam R. Alapati brings together authoritative knowledge for creating, configuring, securing, managing, and optimizing production Hadoop clusters in any environment. Drawing on his experience with large-scale Hadoop administration, Alapati integrates action-oriented advice with carefully researched explanations of both problems and solutions. He covers an unmatched range of topics and offers an unparalleled collection of realistic examples. Alapati demystifies complex Hadoop environments, helping you understand exactly what happens behind the scenes when you administer your cluster. You’ll gain unprecedented insight as you walk through building clusters from scratch and configuring high availability, performance, security, encryption, and other key attributes. The high-value administration skills you learn here will be indispensable no matter what Hadoop distribution you use or what Hadoop applications you run. Understand Hadoop’s architecture from an administrator’s standpoint Create simple and fully distributed clusters Run MapReduce and Spark applications in a Hadoop cluster Manage and protect Hadoop data and high availability Work with HDFS commands, file permissions, and storage management Move data, and use YARN to allocate resources and schedule jobs Ma-

nage job workflows with Oozie and Hue Secure, monitor, log, and optimize Hadoop Benchmark and troubleshoot Hadoop

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

The Hands-On, Example-Rich Introduction to Pandas Data Analysis in Python Today, analysts must manage data char-



acterized 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 Com-

pare 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](http://informit.com/register) for convenient access to downloads, updates, and/or corrections as they become available.

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.  
 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.

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

Understand data analysis pipelines using

machine learning algorithms and techniques with this practical guide  
 Key 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 analysis  
 Book Description- Data 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 AR-MA 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. This book primarily targets Python developers who want to learn and use Python's machine learning capabilities and gain valuable insights from data to develop effective solutions for business problems.

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 analy-

sis. 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

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. 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 perfor-

mance of pandas operations with other Python libraries Who This Book Is For Software engineers with basic programming skills in Python keen on using pandas for a big data analysis project. Python software developers interested in big 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 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 da-

tasetsSolve common data representation and analysis problems using pandasBuild Python scripts, modules, and packages for reusable analysis codeWho 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.