

Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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 efficiency 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 for convenient access to downloads, updates, and corrections as they become

available.

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.

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

Discover how to describe your data in detail, 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.

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.

Python Data Analytics

With Pandas, NumPy, and Matplotlib

**The Ultimate and Complete Guide for Beginners to
Master Data Science with Python Step By Step**

Python for Everybody

Murach's Python for Data Analysis

Essential Tools for Working with Data

Beginning Data Science with Python and Jupyter

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Discover techniques to summarize the characteristics of your data using PyPlot, NumPy, SciPy, and pandas Key

FeaturesUnderstand the fundamental concepts of exploratory data analysis using PythonFind missing values in your data and identify the correlation between different variablesPractice graphical exploratory analysis techniques using Matplotlib and the Seaborn Python packageBook 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.

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

analysis and synthesis. Fundamental knowledge of Python programming and statistical concepts is all you need to get started with this book.

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.

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

compatible) notebooks, so students can execute all coding examples interactively without installing or configuring any software.

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-

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

**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
Pandas in Action**

**Python Tools for Data Munging, Analysis,
and Visual**

Mastering Python Data Analysis

Pandas 1.x Cookbook

**A complete guide to pandas, from
installation to advanced data analysis
techniques, 2nd Edition**

**Automate the Boring Stuff with Python, 2nd
Edition**

Python Data Science Essentials

*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.*

*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*

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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 machine 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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. 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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.

Learn to use powerful Python libraries for effective data processing and analysis About This Book

Learn the basic processing steps in data analysis and how to use Python in this area through supported packages, especially Numpy, Pandas, and Matplotlib Create, manipulate, and analyze your data to extract useful information to optimize your system A hands-on guide to help you learn data analysis using Python Who This Book Is For If you are a Python developer who wants to get started with data analysis and you need a quick introductory guide to the python data analysis libraries, then this book is for you. What You Will Learn Understand the importance of data analysis and get familiar with its processing steps Get acquainted with Numpy to use with arrays and array-oriented computing in data analysis Create effective visualizations to present your data using Matplotlib Process and analyze data using the time

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

series capabilities of Pandas Interact with different kind of database systems, such as file, disk format, Mongo, and Redis Apply the supported Python package to data analysis applications through examples Explore predictive analytics and machine learning algorithms using Scikit-learn, a Python library In Detail Data analysis is the process of applying logical and analytical reasoning to study each component of data. Python is a multi-domain, high-level, programming language. It's often used as a scripting language because of its forgiving syntax and operability with a wide variety of different eco-systems. Python has powerful standard libraries or toolkits such as Pylearn2 and Hebel, which offers a fast, reliable, cross-platform environment for data analysis. With this book, we will get you started with Python data analysis and show you what its advantages are. The book starts by introducing the principles of data analysis and supported libraries, along with NumPy basics for statistic and data processing. Next it provides an overview of the Pandas package and uses its powerful features to solve data processing problems. Moving on, the book takes you through a brief overview of the Matplotlib API and some common plotting functions for DataFrame such as plot. Next, it will teach you to manipulate the time and data structure, and load and store data in a file or database using Python packages. The book

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

will also teach you how to apply powerful packages in Python to process raw data into pure and helpful data using examples. Finally, the book gives you a brief overview of machine learning algorithms, that is, applying data analysis results to make decisions or build helpful products, such as recommendations and predictions using scikit-learn. Style and approach This is an easy-to-follow, step-by-step guide to get you familiar with data analysis and the libraries supported by Python. Topics are explained with real-world examples wherever required.

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

data analysis concepts is all you need to get started with this book.

Python: Data Analytics and Visualization

Python Data Science

Mastering pandas

Become a Python Data Analyst

Hands-On Exploratory Data Analysis with Python

A Guide for Data Scientists

Python Data Analysis

If you want to learn more about Data Analysis or how to master it with the Python Programming Language, then keep reading.

Everyone talks about data today. You have probably come across the term "data" more times than you can remember in one day. Data as a concept is so wide. One thing that is true about data is that it can be used to tell a story. The story could be anything from explaining an event to predicting the future. Data is the future. Businesses, governments, organizations, criminals- everyone needs data for some reason. Entities are investing in different data approaches to help them understand their current situation, and use it to prepare for the unknown. The world of technology as we know it is evolving towards an open-source platform where people share ideas freely. This is seen as the first step towards the decentralization of ideas and eliminating unnecessary monopolies.

Therefore, the data, tools, and techniques used in the analysis are easily available for

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

anyone to interpret data sets and get relevant explanations. With Python for Data Analysis you will learn about the main steps that are needed to correctly implement Data Analysis and the procedures to help you extract the right insights from the right data. Some of the topics that we will discuss inside include: What Data Analysis is all about and why businesses are investing in this sector The 5 steps of a Data Analysis Pandas, Jupyter and PyTorch The 7 Python libraries that make Python one of the best choices for Data Analysis Neural Network How Data Visualization and Matplotlib can help you to understand the data you are working with. Some of the main industries that are using data to improve their business with 14 real-world applications And Much More! While most books focus on how to implement advanced predictive models, this book takes the time to explain the basic concepts and all the necessary steps to correctly implement Data Analysis, including Data Visualization and providing practical examples and simple coding scripts. Don't miss the opportunity to learn more about these topics. Even if you never used Data Analysis, learning it is easier than it looks, you just need the right guidance. This practical guide provides all the knowledge you need in a simple and practical way. Regardless of your previous experience, you will learn the steps of Data Analysis, how to implement them in Python, and the most important real-world

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

applications. Would You Like To Know More? Scroll Up And Click The BUY NOW Button To Get Your Copy!

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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. 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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.*

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

analysis is required to enjoy this book. 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

*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.
Data Structures and Algorithms in Python
Perform Data Collection, Data Processing,
Wrangling, Visualization, and Model Building
Using Python*

Data Science Bookcamp

*Introduction to Machine Learning with Python
Getting Started with Python Data Analysis
Data Analysis and Science using pandas,
matplotlib and the Python Programming
Language*

Hands-On Data Analysis with NumPy and Pandas

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.

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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
Pandas for Everyone Python Data Analysis Addison-Wesley Professional

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

will explore some applications of anomaly detection, regression, clustering, and classification using scikit-learn to make predictions based on past data. This updated edition will equip you with the skills you need to use pandas 1.x to efficiently perform various data manipulation tasks, reliably reproduce analyses, and visualize your data for effective decision making – valuable knowledge that can be applied across multiple domains. What you will learn

Understand how data analysts and scientists gather and analyze data
Perform data analysis and data wrangling using Python
Combine, group, and aggregate data from multiple sources
Create data visualizations with pandas, matplotlib, and seaborn
Apply machine learning algorithms to identify patterns and make predictions
Use Python data science libraries to analyze real-world datasets
Solve common data representation and analysis problems using pandas
Build Python scripts, modules, and packages for reusable analysis code

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

If you want to learn more about Data Science or how to master it with the Python Programming Language, then keep reading. Data Science is one of the biggest buzzwords in the business world nowadays. Many businesses know the importance of collecting information, but as they can collect so much data in a short period, the real question is: "what is the next step?" Data Science includes all the different steps that you take with the data: collecting and cleaning them if they come from more than one source, analyzing them, applying Machine Learning algorithms and models, and then presenting your findings from the analysis with some good Data Visualizations. And this is what you will

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

learn in Python Data Science. You will learn about the main steps that are needed to correctly implement Data Science techniques and the algorithms to help you sort through the data and see some amazing results. Some of the topics that we will discuss inside include: What data science is all about and why so many companies are using it to give them a competitive edge. Why Python and how to use it to implement Data Science What is the intersection between Machine Learning and Data Science and how to combine them The main Data Structures & Object-Oriented Python, with practical codes and exercises to use Python Functions and Modules in Python The 7 most important algorithms and models in Data Science Data Aggregation and Group Operations 9 important Data Mining techniques in Data Science Interaction with databases and data in the cloud And Much More! Where most books only focus on how collecting and cleaning the data, this book goes further, providing guidance on how to perform a proper analysis in order to extract precious information that may be vital for a business. Don't miss the opportunity to learn more about these topics. Even if you have never implemented Data Science techniques, learning them is easier than it looks. You just need the right guidance. And Python Data Science provides all the knowledge you need in a simple and practical way. Regardless of your previous experience, you will learn, the techniques to manipulate and process datasets, the principles of Python programming, and its most important real-world applications. Would You Like To Know More? Scroll up and click on the BUY NOW button to get your copy now! 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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.

Data Wrangling with Pandas, NumPy, and IPython

Five real-world Python projects

Data Analytics with Spark Using Python

Perform EDA techniques to understand, summarize, and investigate your data

Implement Python Packages from Data Manipulation to Processing Thinking in Pandas

R for Everyone

Based on the authors' market leading data structures books in Java and C++, this textbook offers a comprehensive, definitive introduction to data structures in Python by authoritative authors. Data Structures and Algorithms in Python is the first authoritative object-oriented book available for the Python data structures course. Designed to provide a comprehensive introduction to data structures and algorithms, including their design, analysis, and implementation, the text will maintain the same general structure as Data Structures and Algorithms in Java and Data Structures and Algorithms in C++. The Hands-On, Example-Rich Introduction to Pandas Data Analysis in Python Today, analysts must manage data

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

finished reading Learning the Pandas Library and I liked it... very useful and helpful tips even for people who use pandas regularly. -Tom Z.

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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. 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Use powerful industry-standard tools within Jupyter and the Python ecosystem to unlock new, actionable insights from your data

Exploring Data in Python 3

Foundations of Python

A Python data science handbook for data collection, wrangling, analysis, and visualization

Treading on Python Volume 1

Practical Programming for Total Beginners

Learning pandas

Learn data science with Python by building five real-world projects! Experiment with card game predictions, tracking disease outbreaks, and more, as you build a flexible and intuitive understanding of data science. In Data Science Bookcamp you will learn: - Techniques for computing and plotting probabilities - Statistical analysis using Scipy - How to organize datasets with clustering algorithms - How to visualize complex multi-variable datasets - How to train a decision tree machine learning algorithm In Data Science Bookcamp you'll test and build your knowledge of Python with the kind of open-ended problems that professional data scientists work on every day. Downloadable data sets and thoroughly-explained solutions help you lock in what you've learned, building your confidence and making you ready for an exciting new data science career. Purchase of the print book

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology A data science project has a lot of moving parts, and it takes practice and skill to get all the code, algorithms, datasets, formats, and visualizations working together harmoniously. This unique book guides you through five realistic projects, including tracking disease outbreaks from news headlines, analyzing social networks, and finding relevant patterns in ad click data. About the book Data Science Bookcamp doesn't stop with surface-level theory and toy examples. As you work through each project, you'll learn how to troubleshoot common problems like missing data, messy data, and algorithms that don't quite fit the model you're building. You'll appreciate the detailed setup instructions and the fully explained solutions that highlight common failure points. In the end, you'll be confident in your skills because you can see the results. What's inside - Web scraping - Organize datasets with clustering algorithms - Visualize complex multi-variable datasets - Train a decision tree machine learning algorithm About the reader For readers who know the basics of Python. No prior data science or machine learning skills required. About the author Leonard Apeltsin is the Head of Data Science at Anomaly, where his team applies advanced analytics to uncover healthcare fraud, waste, and abuse. Table of Contents CASE STUDY 1 FINDING THE WINNING STRATEGY IN A CARD GAME 1 Computing probabilities using Python 2 Plotting probabilities using Matplotlib 3 Running random simulations in NumPy 4 Case study 1 solution CASE STUDY 2 ASSESSING ONLINE AD CLICKS FOR SIGNIFICANCE 5 Basic probability and statistical analysis using SciPy 6 Making predictions using the central limit theorem and SciPy 7 Statistical hypothesis testing 8 Analyzing tables using Pandas 9 Case study 2 solution CASE STUDY 3 TRACKING DISEASE OUTBREAKS USING NEWS HEADLINES 10 Clustering data into groups 11 Geographic location visualization and analysis 12 Case study 3 solution CASE STUDY 4 USING

ONLINE JOB POSTINGS TO IMPROVE YOUR DATA

SCIENCE RESUME 13 *Measuring text similarities* 14 *Dimension reduction of matrix data* 15 *NLP analysis of large text datasets* 16 *Extracting text from web pages* 17 *Case study 4 solution* **CASE STUDY 5 PREDICTING FUTURE FRIENDSHIPS FROM SOCIAL NETWORK DATA** 18 *An introduction to graph theory and network analysis* 19 *Dynamic graph theory techniques for node ranking and social network analysis* 20 *Network-driven supervised machine learning* 21 *Training linear classifiers with logistic regression* 22 *Training nonlinear classifiers with decision tree techniques* 23 *Case study 5 solution*

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.

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*

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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 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.

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

A Step-By-Step Guide to Master the Basics of Data Science and Analysis in Python Using Pandas, Numpy And Ipython

Learning Python

How to Use the Python Data Analysis Library the Right Way

Learning the Pandas Library

Pandas for Everyone

Python Data Cleaning Cookbook

Managing Spark, YARN, and MapReduce

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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*.

"Optimizing and boosting your Python programming"--Cover.

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.

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Get to grips with pandas—a versatile and high-performance Python library for data manipulation, analysis, and discovery About This Book Get

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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,

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

of using Python and pandas using interactive and incremental 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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 Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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 explains 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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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 Modern techniques and Python tools to detect and remove dirty data and extract key insights Perform exploratory data analysis and gain insight into scientific computing using Python Hands-On Data Analysis with Pandas Foundational Python for Data Science Expert Hadoop 2 Administration Python Data Analysis - Third Edition

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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

Get Free Pandas For Everyone Python Data Analysis Addison Wesley Data Analytics

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.

SciPy and NumPy

*Powerful Object-Oriented Programming
Efficiently perform data collection, wrangling, analysis, and visualization using Python*

*Advanced Analytics and Graphics
Practical recipes for scientific computing, time series analysis, and exploratory data analysis using Python, 2nd Edition*

Python for Data Analysis

Python Data Science Handbook