

Foundations For Analytics With Python From Non Programmer To Hacker

Foundations of Statistics for Data Scientists: With R and Python is designed as a textbook for a one- or two-term introduction to mathematical statistics for students training to become data scientists. It is an in-depth presentation of the topics in statistical science with which any data scientist should be familiar, including probability distributions, descriptive and inferential statistical methods, and linear modeling. The book assumes knowledge of basic calculus, so the presentation can focus on "why it works" as well as "how to do it." Compared to traditional "mathematical statistics" textbooks, however, the book has less emphasis on probability theory and more emphasis on using software to implement statistical methods and to conduct simulations to illustrate key concepts. All statistical analyses in the book use R software, with an appendix showing the same analyses with Python. The book also introduces modern topics that do not normally appear in mathematical statistics texts but are highly relevant for data scientists, such as Bayesian inference, generalized linear models for non-normal responses (e.g., logistic regression and Poisson loglinear models), and regularized model fitting. The nearly 500 exercises are grouped into "Data Analysis and Applications" and "Methods and Concepts." Appendices introduce R and Python and contain solutions for odd-numbered exercises. The book's website has expanded R, Python, and Matlab appendices and all data sets from the examples and exercises.

Summary Introducing Data Science teaches you how to accomplish the fundamental tasks that occupy data scientists. Using the Python language and common Python libraries, you'll experience firsthand the challenges of dealing with data at scale and gain a solid foundation in data science. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Many companies need developers with data science skills to work on projects ranging from social media marketing to machine learning. Discovering what you need to learn to begin a career as a data scientist can seem bewildering. This book is designed to help you get started. About the Book Introducing Data Science Introducing Data Science explains vital data science concepts and teaches you how to accomplish the fundamental tasks that occupy data scientists. You'll explore data visualization, graph databases, the use of NoSQL, and the data science process. You'll use the Python language and common Python libraries as you experience firsthand the challenges of dealing with data at scale. Discover how Python allows you to gain insights from data sets so big that they need to be stored on multiple machines, or from data moving so quickly that no single machine can handle it. This book gives you hands-on experience with the most popular Python data science libraries, Scikit-learn and StatsModels. After reading this book, you'll have the solid foundation you need to start a career in data science. What's Inside Handling large data Introduction to machine learning Using Python to work with data Writing data science algorithms About the Reader This book assumes you're comfortable reading code in Python or a similar language, such as C, Ruby, or JavaScript. No prior experience with data science is required. About the Authors Davy Cielen, Arno D. B. Meysman, and Mohamed Ali are the founders and managing partners of Optimately and Maiton, where they focus on developing data science projects and solutions in various sectors. Table of Contents Data science in a big data world The data science process Machine learning Handling large data on a single computer First steps in big data Join the NoSQL movement The rise of graph databases Text mining and text analytics Data visualization to the end user

* Covers low-level networking in Python —essential for writing a new networked application protocol. * Many working examples demonstrate concepts in action -- and can be used as starting points for new projects. * Networked application security is demystified. * Exhibits and explains multitasking network servers using several models, including forking, threading, and non-blocking sockets. * Features extensive coverage of Web and E-mail. Describes Python's database APIs.

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 Mining for Business Analytics: Concepts, Techniques, and Applications in Python presents an applied approach to data mining concepts and methods, using Python

software for illustration Readers will learn how to implement a variety of popular data mining algorithms in Python (a free and open-source software) to tackle business problems and opportunities. This is the sixth version of this successful text, and the first using Python. It covers both statistical and machine learning algorithms for prediction, classification, visualization, dimension reduction, recommender systems, clustering, text mining and network analysis. It also includes: A new co-author, Peter Gedeck, who brings both experience teaching business analytics courses using Python, and expertise in the application of machine learning methods to the drug-discovery process A new section on ethical issues in data mining Updates and new material based on feedback from instructors teaching MBA, undergraduate, diploma and executive courses, and from their students More than a dozen case studies demonstrating applications for the data mining techniques described End-of-chapter exercises that help readers gauge and expand their comprehension and competency of the material presented A companion website with more than two dozen data sets, and instructor materials including exercise solutions, PowerPoint slides, and case solutions Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python is an ideal textbook for graduate and upper-undergraduate level courses in data mining, predictive analytics, and business analytics. This new edition is also an excellent reference for analysts, researchers, and practitioners working with quantitative methods in the fields of business, finance, marketing, computer science, and information technology. "This book has by far the most comprehensive review of business analytics methods that I have ever seen, covering everything from classical approaches such as linear and logistic regression, through to modern methods like neural networks, bagging and boosting, and even much more business specific procedures such as social network analysis and text mining. If not the bible, it is at the least a definitive manual on the subject." —Gareth M. James, University of Southern California and co-author (with Witten, Hastie and Tibshirani) of the best-selling book An Introduction to Statistical Learning, with Applications in R

This accessible and classroom-tested textbook/reference presents an introduction to the fundamentals of the emerging and interdisciplinary field of data science. The coverage spans key concepts adopted from statistics and machine learning, useful techniques for graph analysis and parallel programming, and the practical application of data science for such tasks as building recommender systems or performing sentiment analysis. Topics and features: provides numerous practical case studies using real-world data throughout the book; supports understanding through hands-on experience of solving data science problems using Python; describes techniques and tools for statistical analysis, machine learning, graph analysis, and parallel programming; reviews a range of applications of data science, including recommender systems and sentiment analysis of text data; provides supplementary code resources and data at an associated website.

The Complete Beginner's Guide to Understanding and Building Machine Learning Systems with Python Machine Learning with Python for Everyone will help you master the processes, patterns, and strategies you need to build effective learning systems, even if you're an absolute beginner. If you can write some Python code, this book is for you, no matter how little college-level math you know. Principal instructor Mark E. Fenner relies on plain-English stories, pictures, and Python examples to communicate the ideas of machine learning. Mark begins by discussing machine learning and what it can do; introducing key mathematical and computational topics in an approachable manner; and walking you through the first steps in building, training, and evaluating learning systems. Step by step, you'll fill out the components of a practical learning system, broaden your toolbox, and explore some of the field's most sophisticated and exciting techniques. Whether you're a student, analyst, scientist, or hobbyist, this guide's insights will be applicable to every learning system you ever build or use. Understand machine learning algorithms, models, and core machine learning concepts Classify examples with classifiers, and quantify examples with regressors Realistically assess performance of machine learning systems Use feature engineering to smooth rough data into useful forms Chain multiple components into one system and tune its performance Apply machine learning techniques to images and text Connect the core concepts to neural networks and graphical models Leverage the Python scikit-learn library and other powerful tools Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

Data analytics may seem daunting, but if you're an experienced Excel user, you have a unique head start. With this hands-on guide, intermediate Excel users will gain a solid understanding of analytics and the data stack. By the time you complete this book, you'll be able to conduct exploratory data analysis and hypothesis testing using a programming language. Exploring and testing relationships are core to analytics. By using the tools and frameworks in this book, you'll be well positioned to continue learning more advanced data analysis techniques. Author George Mount, founder and CEO of Stringfest Analytics, demonstrates key statistical concepts with spreadsheets, then pivots your existing knowledge about data manipulation into R and Python programming. This practical book guides you through: Foundations of analytics in Excel: Use Excel to test relationships between variables and build compelling demonstrations of important concepts in statistics and analytics From Excel to R: Cleanly transfer what you've learned about working with data from Excel to R From Excel to Python: Learn how to pivot your Excel data chops into Python and conduct a complete data analysis

Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use

flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn how to solve real-world data analysis problems with thorough, detailed examples

"Data Science Foundations is most welcome and, indeed, a piece of literature that the field is very much in need of...quite different from most data analytics texts which largely ignore foundational concepts and simply present a cookbook of methods...a very useful text and I would certainly use it in my teaching." - Mark Girolami, Warwick University Data Science encompasses the traditional disciplines of mathematics, statistics, data analysis, machine learning, and pattern recognition. This book is designed to provide a new framework for Data Science, based on a solid foundation in mathematics and computational science. It is written in an accessible style, for readers who are engaged with the subject but not necessarily experts in all aspects. It includes a wide range of case studies from diverse fields, and seeks to inspire and motivate the reader with respect to data, associated information, and derived knowledge.

Inventory Analytics provides a comprehensive and accessible introduction to the theory and practice of inventory control – a significant research area central to supply chain planning. The book outlines the foundations of inventory systems and surveys prescriptive analytics models for deterministic inventory control. It further discusses predictive analytics techniques for demand forecasting in inventory control and also examines prescriptive analytics models for stochastic inventory control. Inventory Analytics is the first book of its kind to adopt a practicable, Python-driven approach to illustrating theories and concepts via computational examples, with each model covered in the book accompanied by its Python code. Originating as a collection of self-contained lectures, Inventory Analytics will be an indispensable resource for practitioners, researchers, teachers, and students alike.

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

This textbook, suitable for an early undergraduate up to a graduate course, provides an overview of many basic principles and techniques needed for modern data analysis. In particular, this book was designed and written as preparation for students planning to take rigorous Machine Learning and Data Mining courses. It introduces key conceptual tools necessary for data analysis, including concentration of measure and PAC bounds, cross validation, gradient descent, and principal component analysis. It also surveys basic techniques in supervised (regression and classification) and unsupervised learning (dimensionality reduction and clustering) through an accessible, simplified presentation. Students are recommended to have some background in calculus, probability, and linear algebra. Some familiarity with programming and algorithms is useful to understand advanced topics on computational techniques.

A new edition of a graduate-level machine learning textbook that focuses on the analysis and theory of algorithms. This book is a general introduction to machine learning that can serve as a textbook for graduate students and a reference for researchers. It covers fundamental modern topics in machine learning while providing the theoretical basis and conceptual tools needed for the discussion and justification of algorithms. It also describes several key aspects of the application of these algorithms. The authors aim to present novel theoretical tools and concepts while giving concise proofs even for relatively advanced topics. Foundations of Machine Learning is unique in its focus on the analysis and theory of algorithms. The first four chapters lay the theoretical foundation for what follows; subsequent chapters are mostly self-contained. Topics covered include the Probably Approximately Correct (PAC) learning framework; generalization bounds based on Rademacher complexity and VC-dimension; Support Vector Machines (SVMs); kernel methods; boosting; on-line learning; multi-class classification; ranking; regression; algorithmic stability; dimensionality reduction; learning automata and languages; and reinforcement learning. Each chapter ends with a set of exercises. Appendixes provide additional material including concise probability review. This second edition offers three new chapters, on model selection, maximum entropy models, and conditional entropy models. New material in the appendixes includes a major section on Fenchel duality, expanded coverage of concentration inequalities, and an entirely new entry on information theory. More than half of the exercises are new to this edition.

Data science libraries, frameworks, modules, and toolkits are great for doing data science, but they're also a good way to dive into the discipline without actually understanding

the Python ecosystem enable analysts to implement analytics tasks as performing as with C or C++, but using only about one-tenth of the code or even less. Derivatives Analytics with Python — Data Analysis, Models, Simulation, Calibration and Hedging shows you what you need to know to supercharge your derivatives and risk analytics efforts.

The fast and easy way to learn Python programming and statistics Python is a general-purpose programming language created in the late 1980s—and named after Monty Python—that's used by thousands of people to do things from testing microchips at Intel, to powering Instagram, to building video games with the PyGame library. Python For Data Science For Dummies is written for people who are new to data analysis, and discusses the basics of Python data analysis programming and statistics. The book also discusses Google Colab, which makes it possible to write Python code in the cloud. Get started with data science and Python Visualize information Wrangle data Learn from data The book provides the statistical background needed to get started in data science programming, including probability, random distributions, hypothesis testing, confidence intervals, and building regression models for prediction.

Drawing on the authors' two decades of experience in applied modeling and data mining, Foundations of Predictive Analytics presents the fundamental background required for analyzing data and building models for many practical applications, such as consumer behavior modeling, risk and marketing analytics, and other areas. It also discusses a variety of practical topics that are frequently missing from similar texts. The book begins with the statistical and linear algebra/matrix foundation of modeling methods, from distributions to cumulant and copula functions to Cornish–Fisher expansion and other useful but hard-to-find statistical techniques. It then describes common and unusual linear methods as well as popular nonlinear modeling approaches, including additive models, trees, support vector machine, fuzzy systems, clustering, naïve Bayes, and neural nets. The authors go on to cover methodologies used in time series and forecasting, such as ARIMA, GARCH, and survival analysis. They also present a range of optimization techniques and explore several special topics, such as Dempster–Shafer theory. An in-depth collection of the most important fundamental material on predictive analytics, this self-contained book provides the necessary information for understanding various techniques for exploratory data analysis and modeling. It explains the algorithmic details behind each technique (including underlying assumptions and mathematical formulations) and shows how to prepare and encode data, select variables, use model goodness measures, normalize odds, and perform reject inference. Web Resource The book's website at www.DataMinerXL.com offers the DataMinerXL software for building predictive models. The site also includes more examples and information on modeling.

FOUNDATIONS FOR ANALYTICS WITH PYTHON. Foundations for Analytics with Python From Non-Programmer to Hacker "O'Reilly Media, Inc."

In this textbook, basic mathematical models used in Big Data Analytics are presented and application-oriented references to relevant practical issues are made. Necessary mathematical tools are examined and applied to current problems of data analysis, such as brand loyalty, portfolio selection, credit investigation, quality control, product clustering, asset pricing etc. - mainly in an economic context. In addition, we discuss interdisciplinary applications to biology, linguistics, sociology, electrical engineering, computer science and artificial intelligence. For the models, we make use of a wide range of mathematics - from basic disciplines of numerical linear algebra, statistics and optimization to more specialized game, graph and even complexity theories. By doing so, we cover all relevant techniques commonly used in Big Data Analytics. Each chapter starts with a concrete practical problem whose primary aim is to motivate the study of a particular Big Data Analytics technique. Next, mathematical results follow - including important definitions, auxiliary statements and conclusions arising. Case-studies help to deepen the acquired knowledge by applying it in an interdisciplinary context. Exercises serve to improve understanding of the underlying theory. Complete solutions for exercises can be consulted by the interested reader at the end of the textbook; for some which have to be solved numerically, we provide descriptions of algorithms in Python code as supplementary material. This textbook has been recommended and developed for university courses in Germany, Austria and Switzerland. The authors Vladimir Shikhman is a professor of Econometrics at Chemnitz University of Technology. David Müller is one of his doctoral students.

This second edition of Foundations of Python Network Programming targets Python 2.5 through Python 2.7, the most popular production versions of the language. Python has made great strides since Apress released the first edition of this book back in the days of Python 2.3. The advances required new chapters to be written from the ground up, and others to be extensively revised. You will learn fundamentals like IP, TCP, DNS and SSL by using working Python programs; you will also be able to familiarize yourself with infrastructure components like memcached and message queues. You can also delve into network server designs, and compare threaded approaches with asynchronous event-based solutions. But the biggest change is this edition's expanded treatment of the web. The HTTP protocol is covered in extensive detail, with each feature accompanied by sample Python code. You can use your HTTP protocol expertise by studying an entire chapter on screen scraping and you can then test lxml and BeautifulSoup against a real-world web site. The chapter on web application programming now covers both the WSGI standard for component interoperability, as well as modern web frameworks like Django. Finally, all of the old favorites from the first edition are back: E-mail protocols like SMTP, POP, and IMAP get full treatment, as does XML-RPC. You can still learn how to code Python network programs using the Telnet and FTP protocols, but you are likely to appreciate the power of more modern alternatives like the paramiko SSH2 library. If you are a Python programmer who needs to learn the network, this is the book that you want by your side.

Learn data science by doing data science! Data Science Using Python and R will get you plugged into the world's two most widespread open-source platforms for data science: Python and R. Data science is hot. Bloomberg called data scientist "the hottest job in America." Python and R are the top two open-source data science tools in the world. In Data Science Using Python and R, you will learn step-by-step how to produce hands-on solutions to real-world business problems, using state-of-the-art techniques. Data Science Using Python and R is written for the general reader with no previous analytics or programming experience. An entire chapter is dedicated to learning the basics of Python and R. Then, each chapter presents step-by-step instructions and walkthroughs for solving data science problems using Python and R. Those with analytics experience will appreciate having a one-stop shop for learning how to do data science using Python and R. Topics covered include data preparation, exploratory data analysis, preparing to model the data, decision trees, model evaluation, misclassification costs, naïve Bayes classification, neural networks, clustering, regression modeling, dimension reduction, and association rules mining. Further, exciting new topics such as random forests and general linear models are also included. The book emphasizes data-driven error costs to enhance profitability, which avoids the common pitfalls that may cost a company millions of dollars. Data Science Using Python and R provides exercises at the end of every chapter, totaling over 500 exercises in the book. Readers will therefore have plenty of opportunity to test their newfound data science skills and expertise. In the Hands-on Analysis exercises, readers are challenged to solve interesting business problems using real-world data sets.

Expand your computer and IT skills and earn more money by learning the world's most popular programming language - Python! Become even more computer savvy and rise above the competition when applying to jobs with proficient Python programming skills. Python programming provides you with a sustainable foundation in computer programming that is easy to build upon and specialize your skills. This

results in becoming a better candidate for job openings and increasing your salary! With this guide in your hands, you will: Learn the Python programming language from scratch with little to no experience required Specialize in a computer language and make yourself more valuable to a company Open the door to new job opportunities after learning and implementing Python Study 3 complete books in one to build on your skills Become more desirable when applying for jobs, especially in the startup community Plus Much More! Right now Python is one of the most popular and useful languages programmers should know. With absolutely no experience required, you could learn the foundations of this language and easily build on your skills to increase your income and open the door to incredible job opportunities. Are you ready to make more money and learn an essential programming language from scratch? ...Then Order Your Complete Guide and Start Learning Today!

A Primer with MATLAB® and Python™ present important information on the emergence of the use of Python, a more general purpose option to MATLAB, the preferred computation language for scientific computing and analysis in neuroscience. This book addresses the snake in the room by providing a beginner's introduction to the principles of computation and data analysis in neuroscience, using both Python and MATLAB, giving readers the ability to transcend platform tribalism and enable coding versatility. Includes discussions of both MATLAB and Python in parallel Introduces the canonical data analysis cascade, standardizing the data analysis flow Presents tactics that strategically, tactically, and algorithmically help improve the organization of code

Despite the recent rapid growth in machine learning and predictive analytics, many of the statistical questions that are faced by researchers and practitioners still involve explaining why something is happening. Regression analysis is the best 'swiss army knife' we have for answering these kinds of questions. This book is a learning resource on inferential statistics and regression analysis. It teaches how to do a wide range of statistical analyses in both R and in Python, ranging from simple hypothesis testing to advanced multivariate modelling. Although it is primarily focused on examples related to the analysis of people and talent, the methods easily transfer to any discipline. The book hits a 'sweet spot' where there is just enough mathematical theory to support a strong understanding of the methods, but with a step-by-step guide and easily reproducible examples and code, so that the methods can be put into practice immediately. This makes the book accessible to a wide readership, from public and private sector analysts and practitioners to students and researchers. Key Features: • 16 accompanying datasets across a wide range of contexts (e.g. academic, corporate, sports, marketing) • Clear step-by-step instructions on executing the analyses. • Clear guidance on how to interpret results. • Primary instruction in R but added sections for Python coders. • Discussion exercises and data exercises for each of the main chapters. • Final chapter of practice material and datasets ideal for class homework or project work.

Learn how to apply powerful data analysis techniques with popular open source Python modules About This Book Find, manipulate, and analyze your data using the Python 3.5 libraries Perform advanced, high-performance linear algebra and mathematical calculations with clean and efficient Python code An easy-to-follow guide with realistic examples that are frequently used in real-world data analysis projects. Who This Book Is For This book is for programmers, scientists, and engineers who have the knowledge of Python and know the basics of data science. It is for those who wish to learn different data analysis methods using Python 3.5 and its libraries. This book contains all the basic ingredients you need to become an expert data analyst. What You Will Learn Install open source Python modules such as NumPy, SciPy, Pandas, statsmodels, scikit-learn, theano, keras, and tensorflow on various platforms Prepare and clean your data, and use it for exploratory analysis Manipulate your data with Pandas Retrieve and store your data from RDBMS, NoSQL, and distributed filesystems such as HDFS and HDF5 Visualize your data with open source libraries such as matplotlib, bokeh, and plotly Learn about various machine learning methods such as supervised, unsupervised, probabilistic, and Bayesian Understand signal processing and time series data analysis Get to grips with graph processing and social network analysis In Detail Data analysis techniques generate useful insights from small and large volumes of data. Python, with its strong set of libraries, has become a popular platform to conduct various data analysis and predictive modeling tasks. With this book, you will learn how to process and manipulate data with Python for complex analysis and modeling. We learn data manipulations such as aggregating, concatenating, appending, cleaning, and handling missing values, with NumPy and Pandas. The book covers how to store and retrieve data from various data sources such as SQL and NoSQL, CSV files, and HDF5. We learn how to visualize data using visualization libraries, along with advanced topics such as signal processing, time series, textual data analysis, machine learning, and social media analysis. The book covers a plethora of Python modules, such as matplotlib, statsmodels, scikit-learn, and NLTK. It also covers using Python with external environments such as R, Fortran, C/C++, and Boost libraries. Style and approach The book takes a very comprehensive approach to enhance your understanding of data analysis. Sufficient real-world examples and use cases are included in the book to help you grasp the concepts quickly and apply them easily in your day-to-day work. Packed with clear, easy to follow examples, this book will turn you into an ace data analyst in no time.

Data science and machine learning - two of the world's hottest fields - are attracting talent from a wide variety of technical, business, and liberal arts disciplines. Python, the world's #1 programming language, is also the most popular language for data science and machine learning. This is the first guide specifically designed to help students with widely diverse backgrounds learn foundational Python so they can use it for data science and machine learning. This book is catered to introductory-level college courses on data science. Leading data science instructor and practitioner Kennedy Behrman first walks through the process of learning to code for the first time with Python and Jupyter notebook, then introduces key libraries every Python data science programmer needs to master. Once students have learned these foundations, Behrman introduces intermediate and applied Python techniques for real-world problem-solving. Throughout, Foundational Python for Data Science presents hands-on exercises, learning assessments, case studies, and more - all created with colab (jupyter compatible) notebooks, so students can execute all coding examples interactively without installing or configuring any software.

Machine learning and analytics have been widely utilized across the healthcare sector of late. This book will bridge the gap between practicing doctors and you as a data scientist. You will learn how to work with healthcare data and gain better insight from this data to improve healthcare outcomes.

The Foundational Hands-On Skills You Need to Dive into Data Science "Freeman and Ross have created the definitive resource for new and aspiring data scientists to learn foundational programming skills." –From the foreword by Jared Lander, series editor Using data science techniques, you can transform raw data into actionable insights for domains ranging from urban planning to precision medicine.

Programming Skills for Data Science brings together all the foundational skills you need to get started, even if you have no programming or data science experience. Leading instructors Michael Freeman and Joel Ross guide you through installing and configuring the tools you need to solve professional-level data science problems, including the widely used R language and Git version-control system. They explain how to wrangle your data into a form where it can be easily used, analyzed, and visualized so others can see the patterns you've uncovered. Step by step, you'll master powerful R programming techniques and troubleshooting skills for probing data in new ways, and at larger scales. Freeman and Ross teach through practical examples and exercises that can be combined into complete data science projects. Everything's focused on real-world application, so you can quickly start analyzing your own data and getting answers you can act upon. Learn to Install your complete data science environment, including R and RStudio Manage projects efficiently, from version tracking to documentation Host, manage, and collaborate on data science projects with GitHub Master R language fundamentals: syntax, programming concepts, and data structures Load, format, explore, and restructure data for successful analysis Interact with databases and web APIs Master key principles for visualizing data accurately and intuitively Produce engaging, interactive visualizations with ggplot and other R packages Transform analyses into sharable documents and sites with R Markdown Create interactive web data science applications with Shiny Collaborate smoothly as part of a data science team Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

The Hands-On, Example-Rich Introduction to Pandas Data Analysis in Python Today, analysts must manage data characterized by extraordinary variety, velocity, and volume. Using the open source Pandas library, you can use Python to rapidly automate and perform virtually any data analysis task, no matter how large or complex. Pandas can help you ensure the veracity of your data, visualize it for effective decision-making, and reliably reproduce analyses across multiple datasets. Pandas for Everyone brings together practical knowledge and insight for solving real problems with Pandas, even if you're new to Python data analysis. Daniel Y. Chen introduces key concepts through simple but practical examples, incrementally building on them to solve more difficult, real-world problems. Chen gives you a jumpstart on using Pandas with a realistic dataset and covers combining datasets, handling missing data, and structuring datasets for easier analysis and visualization. He demonstrates powerful data cleaning techniques, from basic string manipulation to applying functions simultaneously across dataframes. Once your data is ready, Chen guides you through fitting models for prediction, clustering, inference, and exploration. He provides tips on performance and scalability, and introduces you to the wider Python data analysis ecosystem. Work with DataFrames and Series, and import or export data Create plots with matplotlib, seaborn, and pandas Combine datasets and handle missing data Reshape, tidy, and clean datasets so they're easier to work with Convert data types and manipulate text strings Apply functions to scale data manipulations Aggregate, transform, and filter large datasets with groupby Leverage Pandas' advanced date and time capabilities Fit linear models using statsmodels and scikit-learn libraries Use generalized linear modeling to fit models with different response variables Compare multiple models to select the "best" Regularize to overcome overfitting and improve performance Use clustering in unsupervised machine learning

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

Data analytics may seem daunting, but if you're familiar with Excel, you have a head start that can help you make the leap into analytics. Advancing into Analytics will lower your learning curve. Author George Mount, founder and CEO of Stringfest Analytics, clearly and gently guides intermediate Excel users to a solid understanding of analytics and the data stack. This book demonstrates key statistical concepts from spreadsheets and pivots your existing knowledge about data manipulation into R and Python programming. With this practical book at your side, you'll learn how to: Explore a dataset for potential research questions to check assumptions and to build hypotheses Make compelling business recommendations using inferential statistics Load, view, and write datasets using R and Python Perform common data wrangling tasks such as sorting, filtering, and aggregating using R and Python Navigate and execute code in Jupyter notebooks Identify, install, and implement the most useful open source packages for your needs And more

If you're like many of Excel's 750 million users, you want to do more with your data—like repeating similar analyses over hundreds of files, or combining data in many files for analysis at one time. This practical guide shows ambitious non-programmers how to automate and scale the processing and analysis of data in different formats—by using Python. After author Clinton Brownley takes you through Python basics, you'll be able to write simple scripts for processing data in spreadsheets as well as databases. You'll also learn how to use several Python modules for parsing files, grouping data, and producing statistics. No programming experience is necessary. Create and run your own Python scripts by learning basic syntax Use Python's csv module to read and parse CSV files Read multiple Excel worksheets and workbooks with the xlrd module Perform database operations in MySQL or with the mysqlclient module Create Python applications to find specific records, group data, and parse text files Build statistical graphs and plots with matplotlib, pandas, ggplot, and seaborn Produce summary statistics, and estimate regression and classification models Schedule your scripts to run automatically in both Windows and Mac environments

The Contemporary Introduction to Deep Reinforcement Learning that Combines Theory and Practice Deep reinforcement learning (deep RL) combines deep learning and reinforcement learning, in which artificial agents learn to solve sequential decision-making problems. In the past decade deep RL has achieved remarkable results on a range of problems, from single and multiplayer games—such as Go, Atari games, and DotA 2—to robotics. Foundations of Deep Reinforcement Learning is an introduction to deep RL that uniquely combines both theory and implementation. It starts with intuition, then carefully explains the theory of deep RL algorithms, discusses implementations in its companion software library SLM Lab, and finishes with the practical details of getting deep RL to work. This guide is ideal for both computer science students and software engineers who are familiar with basic machine learning concepts and have a working understanding of Python. Understand each key aspect of a deep RL problem Explore policy- and value-based algorithms, including REINFORCE, SARSA, DQN, Double DQN, and Prioritized Experience Replay (PER) Delve into combined algorithms, including Actor-Critic and Proximal Policy Optimization (PPO) Understand how algorithms can be parallelized synchronously and asynchronously Run algorithms in SLM Lab and learn the practical implementation details for getting deep RL to work Explore algorithm benchmark results with tuned hyperparameters Understand how deep RL environments are designed Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

[Copyright: 970967db39bef9a7425a8946cfb026d5](https://www.pdfdrive.com/970967db39bef9a7425a8946cfb026d5)