

Data Munging With Github

Create data mining algorithms About This Book Develop a strong strategy to solve predictive modeling problems using the most popular data mining algorithms Real-world case studies will take you from novice to intermediate to apply data mining techniques Deploy cutting-edge sentiment analysis techniques to real-world social media data using R Who This Book Is For This Learning Path is for R developers who are looking to making a career in data analysis or data mining. Those who come across data mining problems of different complexities from web, text, numerical, political, and social media domains will find all information in this single learning path. What You Will Learn Discover how to manipulate data in R Get to know top classification algorithms written in R Explore solutions written in R based on R Hadoop projects Apply data management skills in handling large data sets Acquire knowledge about neural network concepts and their applications in data mining Create predictive models for classification, prediction, and recommendation Use various libraries on R CRAN for data mining Discover more about data potential, the pitfalls, and inferencial gotchas Gain an insight into the concepts of supervised and unsupervised learning Delve into exploratory data analysis Understand the minute details of sentiment analysis In Detail Data mining is the first step to understanding data and making sense of heaps of data. Properly mined data forms the basis of all data analysis and computing performed on it.

Read Free Data Munging With Github

This learning path will take you from the very basics of data mining to advanced data mining techniques, and will end up with a specialized branch of data mining—social media mining. You will learn how to manipulate data with R using code snippets and how to mine frequent patterns, association, and correlation while working with R programs. You will discover how to write code for various predication models, stream data, and time-series data. You will also be introduced to solutions written in R based on R Hadoop projects. Now that you are comfortable with data mining with R, you will move on to implementing your knowledge with the help of end-to-end data mining projects. You will learn how to apply different mining concepts to various statistical and data applications in a wide range of fields. At this stage, you will be able to complete complex data mining cases and handle any issues you might encounter during projects. After this, you will gain hands-on experience of generating insights from social media data. You will get detailed instructions on how to obtain, process, and analyze a variety of socially-generated data while providing a theoretical background to accurately interpret your findings. You will be shown R code and examples of data that can be used as a springboard as you get the chance to undertake your own analyses of business, social, or political data. 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: Learning Data Mining with R by Biter Makhabel R Data Mining Blueprints by Pradeepta Mishra Social Media Mining with R by Nathan

Read Free Data Munging With Github

Danneman and Richard Heimann Style and approach A complete package with which will take you from the basics of data mining to advanced data mining techniques, and will end up with a specialized branch of data mining—social media mining.

"This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience"--

How do you take your data analysis skills beyond Excel to the next level? By learning just enough Python to get stuff done. This hands-on guide shows non-programmers like you how to process information that's initially too messy or difficult to access. You don't need to know a thing about the Python programming language to get started. Through various step-by-step exercises, you'll learn how to acquire, clean, analyze, and present data efficiently. You'll also discover how to automate your data process, schedule file-editing and clean-up tasks, process larger datasets, and create compelling stories with data you obtain. Quickly learn basic Python syntax, data types, and language concepts Work with both machine-readable and human-consumable data Scrape websites and APIs to find a bounty of useful information Clean and format data to eliminate duplicates and errors in your datasets Learn when to standardize data and when to test and script data cleanup Explore and analyze your datasets with new Python libraries and techniques Use Python solutions to automate your entire data-wrangling process Data science libraries, frameworks, modules, and toolkits are great for doing data

Read Free Data Munging With Github

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

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

Read Free Data Munging With Github

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

Read Free Data Munging With Github

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.

The book provides a one-stop solution for getting into data science with Python and teaches how to extract insights from data.

More physicists today are taking on the role of software developer as part of their research, but software development isn't always easy or obvious, even for physicists. This practical book teaches essential software development skills to help you automate and accomplish nearly any aspect of research in a physics-based field. Written by two PhDs in nuclear engineering, this book includes practical examples drawn from a working knowledge of physics concepts. You'll learn how to use the Python programming language to perform everything from collecting and analyzing data to building software and publishing your results. In four parts, this book includes: Getting Started: Jump into Python, the command line, data containers, functions, flow control

Read Free Data Munging With Github

and logic, and classes and objects Getting It Done: Learn about regular expressions, analysis and visualization, NumPy, storing data in files and HDF5, important data structures in physics, computing in parallel, and deploying software Getting It Right: Build pipelines and software, learn to use local and remote version control, and debug and test your code Getting It Out There: Document your code, process and publish your findings, and collaborate efficiently; dive into software licenses, ownership, and copyright procedures

"GitHub is one of the largest social coding platforms on the web. Its collaborative features allow GitHub users to follow each other's code developments, build off each other's work, and make it easier than ever to create open source software. Based on content from the book "Mining the Social Web" (O'Reilly Media) by Matthew Russell, this course shows you how to mine GitHub data for insight into the platform's projects and community of users. For example, you'll be able to trace open source project histories, the types of programming languages used in those projects, and the relative popularity of those languages. The course teaches you how to make API requests on GitHub's developer platform, use NetworkX to construct interest graphs from GitHub data, and create visualizations of graphs. To get the most out of the course, learners should have basic Python experience and a GitHub account."--Resource description page.

Read Free Data Munging With Github

A concise, hands-on guide with many practical examples and a detailed treatise on inference and social science research that will help you in mining data in the real world. Whether you are an undergraduate who wishes to get hands-on experience working with social data from the Web, a practitioner wishing to expand your competencies and learn unsupervised sentiment analysis, or you are simply interested in social data analysis, this book will prove to be an essential asset. No previous experience with R or statistics is required, though having knowledge of both will enrich your experience.

Summary Data Wrangling with JavaScript is hands-on guide that will teach you how to create a JavaScript-based data processing pipeline, handle common and exotic data, and master practical troubleshooting strategies. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Why not handle your data analysis in JavaScript? Modern libraries and data handling techniques mean you can collect, clean, process, store, visualize, and present web application data while enjoying the efficiency of a single-language pipeline and data-centric web applications that stay in JavaScript end to end. About the Book Data Wrangling with JavaScript promotes JavaScript to the center of the data analysis stage! With this hands-on guide, you'll create a JavaScript-based data processing pipeline, handle common

Read Free Data Munging With Github

and exotic data, and master practical troubleshooting strategies. You'll also build interactive visualizations and deploy your apps to production. Each valuable chapter provides a new component for your reusable data wrangling toolkit. What's inside

- Establishing a data pipeline
- Acquisition, storage, and retrieval
- Handling unusual data sets
- Cleaning and preparing raw data
- Interactive visualizations with D3
- About the Reader
- Written for intermediate JavaScript developers.

No data analysis experience required.

About the Author

Ashley Davis is a software developer, entrepreneur, author, and the creator of Data-Forge and Data-Forge Notebook, software for data transformation, analysis, and visualization in JavaScript.

Table of Contents

- Getting started: establishing your data pipeline
- Getting started with Node.js
- Acquisition, storage, and retrieval
- Working with unusual data
- Exploratory coding
- Clean and prepare
- Dealing with huge data files
- Working with a mountain of data
- Practical data analysis
- Browser-based visualization
- Server-side visualization
- Live data
- Advanced visualization with D3
- Getting to production

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

Read Free Data Munging With Github

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

Hundreds of grassroots groups have sprung up around the world to teach programming, web design, robotics, and other skills outside traditional classrooms. These groups exist so that people don't have to learn these things on their own, but ironically, their founders and instructors are often teaching themselves how to teach. There's a better way. This book presents evidence-based practices that will help you create and deliver lessons that work and build a teaching community around them. Topics include the differences between

Read Free Data Munging With Github

different kinds of learners, diagnosing and correcting misunderstandings, teaching as a performance art, what motivates and demotivates adult learners, how to be a good ally, fostering a healthy community, getting the word out, and building alliances with like-minded groups. The book includes over a hundred exercises that can be done individually or in groups, over 350 references, and a glossary to help you navigate educational jargon.

JavaScript is the native language of the Internet. Originally created to make web pages more dynamic, it is now used for software projects of all kinds, including scientific visualization and data services. However, most data scientists have little or no experience with JavaScript, and most introductions to the language are written for people who want to build shopping carts rather than share maps of coral reefs. This book will introduce you to JavaScript's power and idiosyncrasies and guide you through the key features of the language and its tools and libraries. The book places equal focus on client- and server-side programming, and shows readers how to create interactive web content, build and test data services, and visualize data in the browser. Topics include:

- The core features of modern JavaScript
- Creating templated web pages
- Making those pages interactive using React
- Data visualization using Vega-Lite
- Using Data-Forge to wrangle tabular data
- Building a data service with Express
- Unit testing with Mocha

Read Free Data Munging With Github

All of the material is covered by the Creative Commons Attribution-Noncommercial 4.0 International license (CC-BY-NC-4.0) and is included in the book's companion website at <http://js4ds.org> . Maya Gans is a freelance data scientist and front-end developer by way of quantitative biology. Toby Hodges is a bioinformatician turned community coordinator who works at the European Molecular Biology Laboratory. Greg Wilson co-founded Software Carpentry, and is now part of the education team at RStudio

Examine the techniques and Java tools supporting the growing field of data science About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples Make your Java applications more capable using machine learning Who This Book Is For This book is for Java developers who are comfortable developing applications in Java. Those who now want to enter the world of data science or wish to build intelligent applications will find this book ideal. Aspiring data scientists will also find this book very helpful. What You Will Learn Understand the nature and key concepts used in the field of data science Grasp how data is collected, cleaned, and processed Become comfortable with key data analysis techniques See specialized analysis techniques centered on machine learning Master the effective visualization of your data Work with the

Java APIs and techniques used to perform data analysis In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this book, we cover the important data science concepts and how they are supported by Java, as well as the often statistically challenging techniques, to provide you with an understanding of their purpose and application. The book starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. The next section examines the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. The final chapter illustrates an in-depth data science problem and provides a comprehensive, Java-based solution. Due to the nature of the topic, simple examples of techniques are presented early followed by a more detailed treatment later in the book. This permits a more natural introduction to the techniques and concepts presented in the book. Style and approach This book follows a tutorial approach, providing examples of each of

Read Free Data Munging With Github

the major concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly. For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you'll learn how to use:

- IPython and Jupyter: provide computational environments for data scientists
- using Python NumPy: includes the ndarray for efficient storage and manipulation of dense data arrays in Python
- Pandas: features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python
- Matplotlib: includes capabilities for a flexible range of data visualizations in Python
- Scikit-Learn: for efficient and clean Python implementations of the most important and established machine learning algorithms

Read Free Data Munging With Github

Deep learning is often viewed as the exclusive domain of math PhDs and big tech companies. But as this hands-on guide demonstrates, programmers comfortable with Python can achieve impressive results in deep learning with little math background, small amounts of data, and minimal code. How? With fastai, the first library to provide a consistent interface to the most frequently used deep learning applications. Authors Jeremy Howard and Sylvain Gugger, the creators of fastai, show you how to train a model on a wide range of tasks using fastai and PyTorch. You'll also dive progressively further into deep learning theory to gain a complete understanding of the algorithms behind the scenes. Train models in computer vision, natural language processing, tabular data, and collaborative filtering Learn the latest deep learning techniques that matter most in practice Improve accuracy, speed, and reliability by understanding how deep learning models work Discover how to turn your models into web applications Implement deep learning algorithms from scratch Consider the ethical implications of your work Gain insight from the foreword by PyTorch cofounder, Soumith Chintala

The Practice of Reproducible Research presents concrete examples of how researchers in the data-intensive sciences are working to improve the reproducibility of their research projects. In each of the thirty-one case studies in

Read Free Data Munging With Github

this volume, the author or team describes the workflow that they used to complete a real-world research project. Authors highlight how they utilized particular tools, ideas, and practices to support reproducibility, emphasizing the very practical how, rather than the why or what, of conducting reproducible research. Part 1 provides an accessible introduction to reproducible research, a basic reproducible research project template, and a synthesis of lessons learned from across the thirty-one case studies. Parts 2 and 3 focus on the case studies themselves. The Practice of Reproducible Research is an invaluable resource for students and researchers who wish to better understand the practice of data-intensive sciences and learn how to make their own research more reproducible. Impossible ideas, invisible patterns, hidden connections—visualized Deepen your understanding of the world with these mind-blowing infographics from the bestselling author of The Visual Miscellaneum

This practical guide provides nearly 200 self-contained recipes to help you solve machine learning challenges you may encounter in your daily work. If you're comfortable with Python and its libraries, including pandas and scikit-learn, you'll be able to address specific problems such as loading data, handling text or numerical data, model selection, and dimensionality reduction and many other topics. Each recipe includes code that you can copy and paste into a toy dataset

Read Free Data Munging With Github

to ensure that it actually works. From there, you can insert, combine, or adapt the code to help construct your application. Recipes also include a discussion that explains the solution and provides meaningful context. This cookbook takes you beyond theory and concepts by providing the nuts and bolts you need to construct working machine learning applications. You'll find recipes for: Vectors, matrices, and arrays Handling numerical and categorical data, text, images, and dates and times Dimensionality reduction using feature extraction or feature selection Model evaluation and selection Linear and logical regression, trees and forests, and k-nearest neighbors Support vector machines (SVM), naïve Bayes, clustering, and neural networks Saving and loading trained models

Introduction to Data Science: Data Analysis and Prediction Algorithms with R introduces concepts and skills that can help you tackle real-world data analysis challenges. It covers concepts from probability, statistical inference, linear regression, and machine learning. It also helps you develop skills such as R programming, data wrangling, data visualization, predictive algorithm building, file organization with UNIX/Linux shell, version control with Git and GitHub, and reproducible document preparation. This book is a textbook for a first course in data science. No previous knowledge of R is necessary, although some experience with programming may be helpful. The book is divided into six parts:

Read Free Data Munging With Github

R, data visualization, statistics with R, data wrangling, machine learning, and productivity tools. Each part has several chapters meant to be presented as one lecture. The author uses motivating case studies that realistically mimic a data scientist's experience. He starts by asking specific questions and answers these through data analysis so concepts are learned as a means to answering the questions. Examples of the case studies included are: US murder rates by state, self-reported student heights, trends in world health and economics, the impact of vaccines on infectious disease rates, the financial crisis of 2007-2008, election forecasting, building a baseball team, image processing of hand-written digits, and movie recommendation systems. The statistical concepts used to answer the case study questions are only briefly introduced, so complementing with a probability and statistics textbook is highly recommended for in-depth understanding of these concepts. If you read and understand the chapters and complete the exercises, you will be prepared to learn the more advanced concepts and skills needed to become an expert.

Master the Shiny web framework—and take your R skills to a whole new level. By letting you move beyond static reports, Shiny helps you create fully interactive web apps for data analyses. Users will be able to jump between datasets, explore different subsets or facets of the data, run models with parameter values

Read Free Data Munging With Github

of their choosing, customize visualizations, and much more. Hadley Wickham from RStudio shows data scientists, data analysts, statisticians, and scientific researchers with no knowledge of HTML, CSS, or JavaScript how to create rich web apps from R. This in-depth guide provides a learning path that you can follow with confidence, as you go from a Shiny beginner to an expert developer who can write large, complex apps that are maintainable and performant. Get started: Discover how the major pieces of a Shiny app fit together Put Shiny in action: Explore Shiny functionality with a focus on code samples, example apps, and useful techniques Master reactivity: Go deep into the theory and practice of reactive programming and examine reactive graph components Apply best practices: Examine useful techniques for making your Shiny apps work well in production

Mine the rich data tucked away in popular social websites such as Twitter, Facebook, LinkedIn, and Instagram. With the third edition of this popular guide, data scientists, analysts, and programmers will learn how to glean insights from social media—including who's connecting with whom, what they're talking about, and where they're located—using Python code examples, Jupyter notebooks, or Docker containers. In part one, each standalone chapter focuses on one aspect of the social landscape, including each of the major social sites, as well as web

Read Free Data Munging With Github

pages, blogs and feeds, mailboxes, GitHub, and a newly added chapter covering Instagram. Part two provides a cookbook with two dozen bite-size recipes for solving particular issues with Twitter. Get a straightforward synopsis of the social web landscape Use Docker to easily run each chapter's example code, packaged as a Jupyter notebook Adapt and contribute to the code's open source GitHub repository Learn how to employ best-in-class Python 3 tools to slice and dice the data you collect Apply advanced mining techniques such as TFIDF, cosine similarity, collocation analysis, clique detection, and image recognition Build beautiful data visualizations with Python and JavaScript toolkits How can you tap into the wealth of social web data to discover who's making connections with whom, what they're talking about, and where they're located? With this expanded and thoroughly revised edition, you'll learn how to acquire, analyze, and summarize data from all corners of the social web, including Facebook, Twitter, LinkedIn, Google+, GitHub, email, websites, and blogs. Employ the Natural Language Toolkit, NetworkX, and other scientific computing tools to mine popular social web sites Apply advanced text-mining techniques, such as clustering and TF-IDF, to extract meaning from human language data Bootstrap interest graphs from GitHub by discovering affinities among people, programming languages, and coding projects Build interactive visualizations with

Read Free Data Munging With Github

D3.js, an extraordinarily flexible HTML5 and JavaScript toolkit Take advantage of more than two-dozen Twitter recipes, presented in O'Reilly's popular "problem/solution/discussion" cookbook format The example code for this unique data science book is maintained in a public GitHub repository. It's designed to be easily accessible through a turnkey virtual machine that facilitates interactive learning with an easy-to-use collection of IPython Notebooks.

Social media marketing is the process of gaining website traffic or attention through social media sites. Social media marketing programs usually center on efforts to create content that attracts attention and encourages readers to share it across their social networks. The resulting electronic word of mouth (eWoM) refers to any statement consumers share via the Internet (e.g., web sites, social networks, instant messages, news feeds) about an event, product, service, brand or company. When the underlying message spreads from user to user and presumably resonates because it appears to come from a trusted, third-party source, as opposed to the brand or company itself, this form of marketing results in earned media rather than paid media.

Table of Contents: Opening Words 6 Why Social Media and Business Intelligence? 6 Maturity 6 Cost & variety 6 User Friendly 6 Integration 7 The Time is Now 7 Preface: Social Media and Business Intelligence 8 1 Introduction 9 2 What is 'Out' in Social Media and what is 'In' 12 3 The 5 Pillars Of Social Media and Business Intelligence 14 3.1 Top CRM Vendors 2012 (via CIO Magazine): 18 Biggest Companies in Marketing Management and Automation: 18 3.3 Top Business Intelligence Products: 21 4 The 7 Key Reasons You Need Social Media And Enterprise Marketing 24 5 8 Key Social Media Metrics and Their

Read Free Data Munging With Github

Actionable Responses 28 6 The Nine Step Enterprise And Social Media Marketing Work Flow 31 7 Linking Enterprise Marketing To Internal Business Processes 35 8 7 Social Media Metrics That Drive Industry Development 39 9 Test The Social Media/Enterprise Marketing Waters For Free 41 10 Available Solutions and 17 Essential Questions to ask a potential vendor 50 11 Index - 5 Pillars, Top 5 White Papers on SMBI 54 12 About the author 55

If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions

Do you want to use R to tell stories? This book was written for you—whether you already know some R or have never coded before. Most R texts focus only on programming or statistical theory. Practical R for Mass Communication and Journalism gives you ideas, tools, and

Read Free Data Munging With Github

techniques for incorporating data and visualizations into your narratives. You'll see step by step how to: Analyze airport flight delays, restaurant inspections, and election results Map bank locations, median incomes, and new voting districts Compare campaign contributions to final election results Extract data from PDFs Whip messy data into shape for analysis Scrape data from a website Create graphics ranging from simple, static charts to interactive visualizations for the Web If you work or plan to work in a newsroom, government office, non-profit policy organization, or PR office, Practical R for Mass Communication and Journalism will help you use R in your world. This book has a companion website with code, links to additional resources, and searchable tables by function and task. Sharon Machlis is the author of Computerworld's Beginner's Guide to R, host of InfoWorld's Do More With R video screencast series, admin for the R for Journalists Google Group, and is well known among Twitter users who follow the #rstats hashtag. She is Director of Editorial Data and Analytics at IDG Communications (parent company of Computerworld, InfoWorld, PC World and Macworld, among others) and a frequent speaker at data journalism and R conferences.

Development Research in Practice leads the reader through a complete empirical research project, providing links to continuously updated resources on the DIME Wiki as well as illustrative examples from the Demand for Safe Spaces study. The handbook is intended to train users of development data how to handle data effectively, efficiently, and ethically. "In the DIME Analytics Data Handbook, the DIME team has produced an extraordinary public good: a detailed, comprehensive, yet easy-to-read manual for how to manage a data-oriented research project from beginning to end. It offers everything from big-picture guidance on the determinants of high-quality empirical research, to specific practical guidance on how to

Read Free Data Munging With Github

implement specific workflows—and includes computer code! I think it will prove durably useful to a broad range of researchers in international development and beyond, and I learned new practices that I plan on adopting in my own research group.†? —Marshall Burke, Associate Professor, Department of Earth System Science, and Deputy Director, Center on Food Security and the Environment, Stanford University “Data are the essential ingredient in any research or evaluation project, yet there has been too little attention to standardized practices to ensure high-quality data collection, handling, documentation, and exchange. Development Research in Practice: The DIME Analytics Data Handbook seeks to fill that gap with practical guidance and tools, grounded in ethics and efficiency, for data management at every stage in a research project. This excellent resource sets a new standard for the field and is an essential reference for all empirical researchers.†? —Ruth E. Levine, PhD, CEO, IDinsight “Development Research in Practice: The DIME Analytics Data Handbook is an important resource and a must-read for all development economists, empirical social scientists, and public policy analysts. Based on decades of pioneering work at the World Bank on data collection, measurement, and analysis, the handbook provides valuable tools to allow research teams to more efficiently and transparently manage their work flows—yielding more credible analytical conclusions as a result.†? —Edward Miguel, Oxfam Professor in Environmental and Resource Economics and Faculty Director of the Center for Effective Global Action, University of California, Berkeley “The DIME Analytics Data Handbook is a must-read for any data-driven researcher looking to create credible research outcomes and policy advice. By meticulously describing detailed steps, from project planning via ethical and responsible code and data practices to the publication of research papers and associated replication packages, the DIME handbook

Read Free Data Munging With Github

makes the complexities of transparent and credible research easier.†? —Lars Vilhuber, Data Editor, American Economic Association, and Executive Director, Labor Dynamics Institute, Cornell University

Develop, Implement and Tuneup your Machine Learning applications using the power of Java programming About This Book Detailed coverage on key machine learning topics with an emphasis on both theoretical and practical aspects Address predictive modeling problems using the most popular machine learning Java libraries A comprehensive course covering a wide spectrum of topics such as machine learning and natural language through practical use-cases Who This Book Is For This course is the right resource for anyone with some knowledge of Java programming who wants to get started with Data Science and Machine learning as quickly as possible. If you want to gain meaningful insights from big data and develop intelligent applications using Java, this course is also a must-have. What You Will Learn Understand key data analysis techniques centered around machine learning Implement Java APIs and various techniques such as classification, clustering, anomaly detection, and more Master key Java machine learning libraries, their functionality, and various kinds of problems that can be addressed using each of them Apply machine learning to real-world data for fraud detection, recommendation engines, text classification, and human activity recognition Experiment with semi-supervised learning and stream-based data mining, building high-performing and real-time predictive models Develop intelligent systems centered around various domains such as security, Internet of Things, social networking, and more In Detail Machine Learning is one of the core area of Artificial Intelligence where computers are trained to self-learn, grow, change, and develop on their own without being explicitly programmed. In

Read Free Data Munging With Github

this course, we cover how Java is employed to build powerful machine learning models to address the problems being faced in the world of Data Science. The course demonstrates complex data extraction and statistical analysis techniques supported by Java, applying various machine learning methods, exploring machine learning sub-domains, and exploring real-world use cases such as recommendation systems, fraud detection, natural language processing, and more, using Java programming. The course begins with an introduction to data science and basic data science tasks such as data collection, data cleaning, data analysis, and data visualization. The next section has a detailed overview of statistical techniques, covering machine learning, neural networks, and deep learning. The next couple of sections cover applying machine learning methods using Java to a variety of chores including classifying, predicting, forecasting, market basket analysis, clustering stream learning, active learning, semi-supervised learning, probabilistic graph modeling, text mining, and deep learning. The last section highlights real-world test cases such as performing activity recognition, developing image recognition, text classification, and anomaly detection. The course includes premium content from three of our most popular books: Java for Data Science Machine Learning in Java Mastering Java Machine Learning On completion of this course, you will understand various machine learning techniques, different machine learning java algorithms you can use to gain data insights, building data models to analyze larger complex data sets, and incubating applications using Java and machine learning algorithms in the field of artificial intelligence. Style and approach This comprehensive course proceeds from being a tutorial to a practical guide, providing an introduction to machine learning and different machine learning techniques, exploring machine learning with Java libraries, and demonstrating real-

Read Free Data Munging With Github

world machine learning use cases using the Java platform.

Data is the new oil, but it comes crude. To do anything meaningful - modeling, visualization, machine learning, for predictive analysis – you first need to wrestle and wrangle with data.

Data Wrangling with Python teaches you the essentials that will get you up and running with data wrangling in no time.

Provides information on data analysis from a variety of social networking sites, including Facebook, Twitter, and LinkedIn.

Written in Ron Cody's signature informal, tutorial style, this book develops and demonstrates data cleaning programs and macros that you can use as written or modify which will make your job of data cleaning easier, faster, and more efficient. --

Using data from one season of NBA games, *Basketball Data Science: With Applications in R* is the perfect book for anyone interested in learning and applying data analytics in basketball. Whether assessing the spatial performance of an NBA player's shots or doing an analysis of the impact of high pressure game situations on the probability of scoring, this book discusses a variety of case studies and hands-on examples using a custom R package. The codes are supplied so readers can reproduce the analyses themselves or create their own. Assuming a basic statistical knowledge, *Basketball Data Science with R* is suitable for students, technicians, coaches, data analysts and applied researchers. Features: · One of the first books to provide statistical and data mining methods for the growing field of analytics in basketball. · Presents tools for modelling graphs and figures to visualize the data. ·

Read Free Data Munging With Github

Includes real world case studies and examples, such as estimations of scoring probability using the Golden State Warriors as a test case. · Provides the source code and data so readers can do their own analyses on NBA teams and players.

This engaging and clearly written textbook/reference provides a must-have introduction to the rapidly emerging interdisciplinary field of data science. It focuses on the principles fundamental to becoming a good data scientist and the key skills needed to build systems for collecting, analyzing, and interpreting data. The Data Science Design Manual is a source of practical insights that highlights what really matters in analyzing data, and provides an intuitive understanding of how these core concepts can be used. The book does not emphasize any particular programming language or suite of data-analysis tools, focusing instead on high-level discussion of important design principles. This easy-to-read text ideally serves the needs of undergraduate and early graduate students embarking on an “Introduction to Data Science” course. It reveals how this discipline sits at the intersection of statistics, computer science, and machine learning, with a distinct heft and character of its own. Practitioners in these and related fields will find this book perfect for self-study as well. Additional learning tools: Contains “War Stories,” offering perspectives on how data science applies in the real world Includes “Homework Problems,” providing a wide range of exercises and projects for self-study Provides a complete set of lecture slides and online video lectures at www.data-manual.com Provides “Take-Home Lessons,” emphasizing the big-picture concepts to

Read Free Data Munging With Github

learn from each chapter Recommends exciting “Kaggle Challenges” from the online platform Kaggle Highlights “False Starts,” revealing the subtle reasons why certain approaches fail Offers examples taken from the data science television show “The Quant Shop” (www.quant-shop.com)

Much of the data available today is unstructured and text-heavy, making it challenging for analysts to apply their usual data wrangling and visualization tools. With this practical book, you’ll explore text-mining techniques with tidytext, a package that authors Julia Silge and David Robinson developed using the tidy principles behind R packages like ggraph and dplyr. You’ll learn how tidytext and other tidy tools in R can make text analysis easier and more effective. The authors demonstrate how treating text as data frames enables you to manipulate, summarize, and visualize characteristics of text. You’ll also learn how to integrate natural language processing (NLP) into effective workflows. Practical code examples and data explorations will help you generate real insights from literature, news, and social media. Learn how to apply the tidy text format to NLP Use sentiment analysis to mine the emotional content of text Identify a document’s most important terms with frequency measurements Explore relationships and connections between words with the ggraph and widyr packages Convert back and forth between R’s tidy and non-tidy text formats Use topic modeling to classify document collections into natural groups Examine case studies that compare Twitter archives, dig into NASA metadata, and analyze thousands of Usenet messages

Read Free Data Munging With Github

This guide for practicing statisticians, data scientists, and R users and programmers will teach the essentials of preprocessing: data leveraging the R programming language to easily and quickly turn noisy data into usable pieces of information. Data wrangling, which is also commonly referred to as data munging, transformation, manipulation, janitor work, etc., can be a painstakingly laborious process. Roughly 80% of data analysis is spent on cleaning and preparing data; however, being a prerequisite to the rest of the data analysis workflow (visualization, analysis, reporting), it is essential that one become fluent and efficient in data wrangling techniques. This book will guide the user through the data wrangling process via a step-by-step tutorial approach and provide a solid foundation for working with data in R. The author's goal is to teach the user how to easily wrangle data in order to spend more time on understanding the content of the data. By the end of the book, the user will have learned: How to work with different types of data such as numerics, characters, regular expressions, factors, and dates The difference between different data structures and how to create, add additional components to, and subset each data structure How to acquire and parse data from locations previously inaccessible How to develop functions and use loop control structures to reduce code redundancy How to use pipe operators to simplify code and make it more readable How to reshape the layout of data and manipulate, summarize, and join data sets

Summary Beyond Spreadsheets with R shows you how to take raw data and transform

Read Free Data Munging With Github

it for use in computations, tables, graphs, and more. You'll build on simple programming techniques like loops and conditionals to create your own custom functions. You'll come away with a toolkit of strategies for analyzing and visualizing data of all sorts using R and RStudio. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Spreadsheets are powerful tools for many tasks, but if you need to interpret, interrogate, and present data, they can feel like the wrong tools for the task. That's when R programming is the way to go. The R programming language provides a comfortable environment to properly handle all types of data. And within the open source RStudio development suite, you have at your fingertips easy-to-use ways to simplify complex manipulations and create reproducible processes for analysis and reporting. About the Book With Beyond Spreadsheets with R you'll learn how to go from raw data to meaningful insights using R and RStudio. Each carefully crafted chapter covers a unique way to wrangle data, from understanding individual values to interacting with complex collections of data, including data you scrape from the web. You'll build on simple programming techniques like loops and conditionals to create your own custom functions. You'll come away with a toolkit of strategies for analyzing and visualizing data of all sorts. What's inside How to start programming with R and RStudio Understanding and implementing important R structures and operators Installing and working with R packages Tidying, refining, and plotting your data About

Read Free Data Munging With Github

the Reader If you're comfortable writing formulas in Excel, you're ready for this book. About the Author Dr Jonathan Carroll is a data science consultant providing R programming services. He holds a PhD in theoretical physics. Table of Contents

Introducing data and the R language Getting to know R data types Making new data values Understanding the tools you'll use: Functions Combining data values Selecting data values Doing things with lots of data Doing things conditionally: Control structures Visualizing data: Plotting Doing more with your data with extensions

Data collection, processing, analysis, and more About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples A highly practical course covering a broad set of topics - from the basics of Machine Learning to Deep Learning and Big Data frameworks. Who This Book Is For This course is meant for Java developers who are comfortable developing applications in Java, and now want to enter the world of data science or wish to build intelligent applications. Aspiring data scientists with some understanding of the Java programming language will also find this book to be very helpful. If you are willing to build efficient data science applications and bring them in the enterprise environment without changing your existing Java stack, this book is for you! What You Will Learn Understand the key concepts of data science Explore the data science ecosystem available in Java Work with the Java APIs and techniques used to perform efficient data analysis Find out how to approach different

Read Free Data Munging With Github

machine learning problems with Java Process unstructured information such as natural language text or images, and create your own search Learn how to build deep neural networks with DeepLearning4j Build data science applications that scale and process large amounts of data Deploy data science models to production and evaluate their performance In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this course, we cover the basic as well as advanced data science concepts and how they are implemented using the popular Java tools and libraries. The course starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. You will examine the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. Throughout this course, the chapters will illustrate a challenging data science problem, and then go on to present a comprehensive, Java-based solution to tackle that problem. You will cover a wide range of topics – from classification and regression, to dimensionality reduction and clustering, deep learning and working with Big Data. Finally, you will see the different ways to deploy the model and evaluate it in

Read Free Data Munging With Github

production settings. By the end of this course, you will be up and running with various facets of data science using Java, in no time at all. This course contains premium content from two of our recently published popular titles: Java for Data Science Mastering Java for Data Science Style and approach This course follows a tutorial approach, providing examples of each of the concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly.

Social media marketing is the process of gaining website traffic or attention through social media sites. Social media marketing programs usually center on efforts to create content that attracts attention and encourages readers to share it across their social networks. The resulting electronic word of mouth (eWoM) refers to any statement consumers share via the Internet (e.g., web sites, social networks, instant messages, news feeds) about an event, product, service, brand or company. When the underlying message spreads from user to user and presumably resonates because it appears to come from a trusted, third-party source, as opposed to the brand or company itself, this form of marketing results in earned media rather than paid media.

Table of Contents
:Opening Words 6Why Social Media and Business Intelligence? 6Maturity 6Cost & variety 6User Friendly 6Integration 7The Time is Now 7Preface: Social Media and Business Intelligence 81 Introduction 92 What is 'Out' in Social Media and what is 'In' 123 The 5 Pillars Of Social Media and Business Intelligence 143.1 Top CRM Vendors

Read Free Data Munging With Github

2012 (via CIO Magazine): 18Biggest Companies in Marketing Management and Automation: 183.3 Top Business Intelligence Products: 214 The 7 Key Reasons You Need Social Media And Enterprise Marketing 245 8 Key Social Media Metrics and Their Actionable Responses 286 The Nine Step Enterprise And Social Media Marketing Work Flow 317 Linking Enterprise Marketing To Internal Business Processes 358 7 Social Media Metrics That Drive Industry Development 399 Test The Social Media/Enterprise Marketing Waters For Free 4110 Available Solutions and 17 Essential Questions to ask a potential vendor 5011 Index - 5 Pillars, Top 5 White Papers on SMBI 5412 About the author 55

Machine learning has become an integral part of many commercial applications and research projects, but this field is not exclusive to large companies with extensive research teams. If you use Python, even as a beginner, this book will teach you practical ways to build your own machine learning solutions. With all the data available today, machine learning applications are limited only by your imagination. You'll learn the steps necessary to create a successful machine-learning application with Python and the scikit-learn library. Authors Andreas Müller and Sarah Guido focus on the practical aspects of using machine learning algorithms, rather than the math behind them. Familiarity with the NumPy and matplotlib libraries will help you get even more from this book. With this book, you'll learn: Fundamental concepts and applications of machine learning Advantages and shortcomings of widely used machine learning

Read Free Data Munging With Github

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

[Copyright: 074f28ac4b66e55963b9d729f86cab6c](#)