Clean Data - Data Science Strategies for Tackling Dirty Data

Megan Squire

Clean Data - Data Science Strategies for Tackling Dirty Data

Megan Squire

Clean Data - Data Science Strategies for Tackling Dirty Data Megan Squire

Key Features

- Grow your data science expertise by filling your toolbox with proven strategies for a wide variety of cleaning challenges
- Familiarize yourself with the crucial data cleaning processes, and share your own clean data sets with others
- Complete real-world projects using data from Twitter and Stack Overflow

Book Description

Is much of your time spent doing tedious tasks such as cleaning dirty data, accounting for lost data, and preparing data to be used by others? If so, then having the right tools makes a critical difference, and will be a great investment as you grow your data science expertise.

The book starts by highlighting the importance of data cleaning in data science, and will show you how to reap rewards from reforming your cleaning process. Next, you will cement your knowledge of the basic concepts that the rest of the book relies on: file formats, data types, and character encodings. You will also learn how to extract and clean data stored in RDBMS, web files, and PDF documents, through practical examples.

At the end of the book, you will be given a chance to tackle a couple of real-world projects.

What you will learn

- Understand the role of data cleaning in the overall data science process
- Learn the basics of file formats, data types, and character encodings to clean data properly
- Master critical features of the spreadsheet and text editor for organizing and manipulating data
- Convert data from one common format to another, including JSON, CSV, and some special-purpose formats
- Implement three different strategies for parsing and cleaning data found in HTML files on the Web
- Reveal the mysteries of PDF documents and learn how to pull out just the data you want
- Develop a range of solutions for detecting and cleaning bad data stored in an RDBMS
- Create your own clean data sets that can be packaged, licensed, and shared with others
- Use the tools from this book to complete two real-world projects using data from Twitter and Stack Overflow

About the Author

Megan Squire is a professor of computing sciences at Elon University. She has been collecting and cleaning dirty data for two decades. She is also the leader of FLOSSmole.org, a research project to collect data and analyze it in order to learn how free, libre, and open source software is made.

Table of Contents

[Download](#) Clean Data - Data Science Strategies for Tackling Dirty ...pdf

[Read Online](#) Clean Data - Data Science Strategies for Tackling Dir ...pdf

Download and Read Free Online Clean Data - Data Science Strategies for Tackling Dirty Data Megan Squire

Download and Read Free Online Clean Data - Data Science Strategies for Tackling Dirty Data Megan Squire

---

From reader reviews:

Belia Gillespie:

Do you have favorite book? For those who have, what is your favorite's book? E-book is very important thing for us to know everything in the world. Each reserve has different aim or goal; it means that publication has different type. Some people really feel enjoy to spend their time for you to read a book. These are reading whatever they acquire because their hobby is usually reading a book. Think about the person who don't like examining a book? Sometime, individual feel need book when they found difficult problem or perhaps exercise. Well, probably you will need this Clean Data - Data Science Strategies for Tackling Dirty Data.

Victor Brown:

Do you have something that you want such as book? The e-book lovers usually prefer to choose book like comic, small story and the biggest some may be novel. Now, why not trying Clean Data - Data Science Strategies for Tackling Dirty Data that give your fun preference will be satisfied by simply reading this book. Reading practice all over the world can be said as the method for people to know world far better then how they react when it comes to the world. It can't be stated constantly that reading routine only for the geeky person but for all of you who wants to be success person. So , for all of you who want to start reading as your good habit, you may pick Clean Data - Data Science Strategies for Tackling Dirty Data become your personal starter.

Paul Williams:

In this period of time globalization it is important to someone to receive information. The information will make professionals understand the condition of the world. The health of the world makes the information simpler to share. You can find a lot of personal references to get information example: internet, magazine, book, and soon. You will observe that now, a lot of publisher that will print many kinds of book. Typically the book that recommended to your account is Clean Data - Data Science Strategies for Tackling Dirty Data this guide consist a lot of the information in the condition of this world now. That book was represented how does the world has grown up. The vocabulary styles that writer use for explain it is easy to understand. Typically the writer made some exploration when he makes this book. Here is why this book acceptable all of you.

Michael Gage:

As we know that book is significant thing to add our information for everything. By a book we can know everything you want. A book is a list of written, printed, illustrated or maybe blank sheet. Every year was exactly added. This e-book Clean Data - Data Science Strategies for Tackling Dirty Data was filled with regards to science. Spend your extra time to add your knowledge about your scientific research competence. Some people has diverse feel when they reading the book. If you know how big selling point of a book, you

can really feel enjoy to read a e-book. In the modern era like now, many ways to get book that you simply wanted.

Download and Read Online Clean Data - Data Science Strategies for Tackling Dirty Data Megan Squire #0N3J4TEZCUI

Read Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire for online ebook

Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire Free PDF d0wnl0ad, audio books, books to read, good books to read, cheap books, good books, online books, books online, book reviews epub, read books online, books to read online, online library, greatbooks to read, PDF best books to read, top books to read Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire books to read online.

Online Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire ebook PDF download

Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire Doc

Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire Mobipocket

Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire EPub

Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire Ebook online

Clean Data - Data Science Strategies for Tackling Dirty Data by Megan Squire Ebook PDF