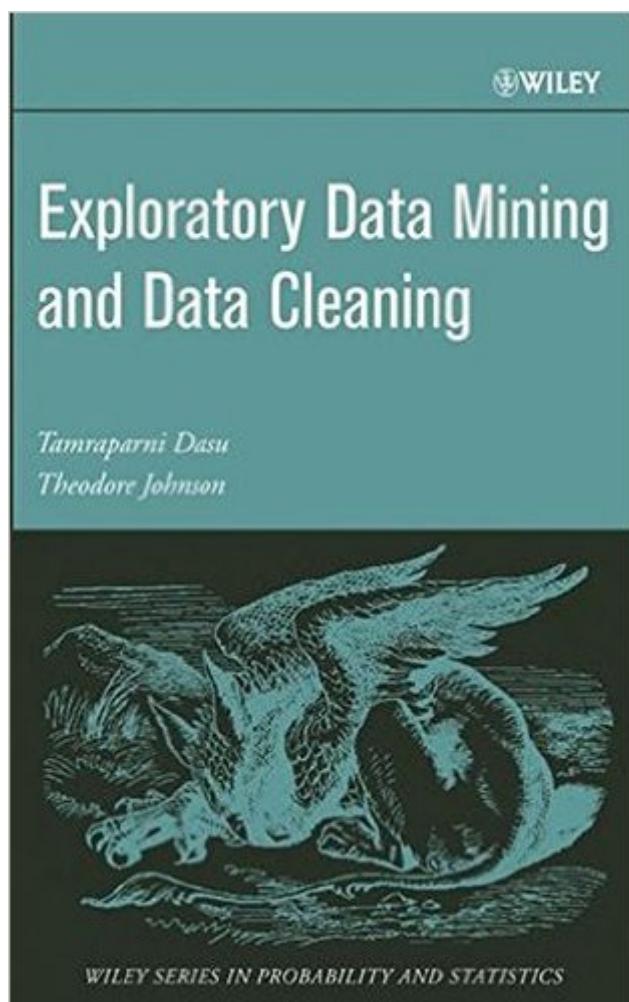


The book was found

# Exploratory Data Mining And Data Cleaning



## **Synopsis**

Written for practitioners of data mining, data cleaning and database management. Presents a technical treatment of data quality including process, metrics, tools and algorithms. Focuses on developing an evolving modeling strategy through an iterative data exploration loop and incorporation of domain knowledge. Addresses methods of detecting, quantifying and correcting data quality issues that can have a significant impact on findings and decisions, using commercially available tools as well as new algorithmic approaches. Uses case studies to illustrate applications in real life scenarios. Highlights new approaches and methodologies, such as the DataSphere space partitioning and summary based analysis techniques. Exploratory Data Mining and Data Cleaning will serve as an important reference for serious data analysts who need to analyze large amounts of unfamiliar data, managers of operations databases, and students in undergraduate or graduate level courses dealing with large scale data analysis and data mining.

## **Book Information**

Hardcover: 224 pages

Publisher: Wiley-Interscience; 1 edition (May 9, 2003)

Language: English

ISBN-10: 0471268518

ISBN-13: 978-0471268512

Product Dimensions: 6.4 x 0.8 x 9.6 inches

Shipping Weight: 1.6 pounds (View shipping rates and policies)

Average Customer Review: 5.0 out of 5 stars See all reviews (1 customer review)

Best Sellers Rank: #1,634,183 in Books (See Top 100 in Books) #243 in Books > Computers & Technology > Computer Science > AI & Machine Learning > Machine Theory #514 in Books > Textbooks > Computer Science > Artificial Intelligence #1030 in Books > Computers & Technology > Computer Science > AI & Machine Learning > Intelligence & Semantics

## **Customer Reviews**

This is the best deep and practical introduction to data cleaning that I have seen. It provides an excellent overview of the practical problems in data cleaning, gives a good intuitive feeling for the core issues of outliers and robust statistics, and overviews of a good set of techniques for addressing data cleaning issues in a practical but relatively deep manner. It doesn't try to provide cookbook solutions, and instead points out the complexities and leaves the reader with a toolbox to work on tackling them. The really interested reader will want to augment the book with some other

reading, including (on the practical side) a book or website of tips on how to express robust statistics in SQL (the O'Reilly book on TransactSQL has good stuff), and (on the more statistical side) a deeper introduction to robust statistics (e.g. Rousseeuw and Leroy's Robust Regression and Outlier Detection). In a future edition it would be nice to see more discussion of timeseries outliers, as well as an SQL cookbook that will run on commodity databases of modest size (which is the common case in practice, as opposed to the massive TelCo databases that the authors discuss).

[Download to continue reading...](#)

Exploratory Data Mining and Data Cleaning RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications) Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business. Leveraging the Power of Data Analytics, Data ... (Hacking Freedom and Data Driven) (Volume 2) Yellowcake Towns - Uranium Mining Communities in the American West (Mining the American West) Bitcoin Mining: The Bitcoin Beginner's Guide (Proven, Step-By-Step Guide To Making Money With Bitcoins) (Bitcoin Mining, Online Business, Investing for ... Beginner, Bitcoin Guide, Bitcoin Trading) Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data) Collage Lab: Experiments, Investigations, and Exploratory Projects (Lab Series) Exploratory Programming for the Arts and Humanities (MIT Press) Data Analysis and Data Mining using Microsoft Business Intelligence Tools: Excel 2010, Access 2010, and Report Builder 3.0 with SQL Server Healthcare Data Analytics (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series) Data Mining and Analysis: Fundamental Concepts and Algorithms Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) Transforming Health Care: The Financial Impact of Technology, Electronic Tools and Data Mining Data Mining for Business Analytics: Concepts, Techniques, and Applications with JMP Pro Dark Web: Exploring and Data Mining the Dark Side of the Web (Integrated Series in Information Systems)

[Dmca](#)