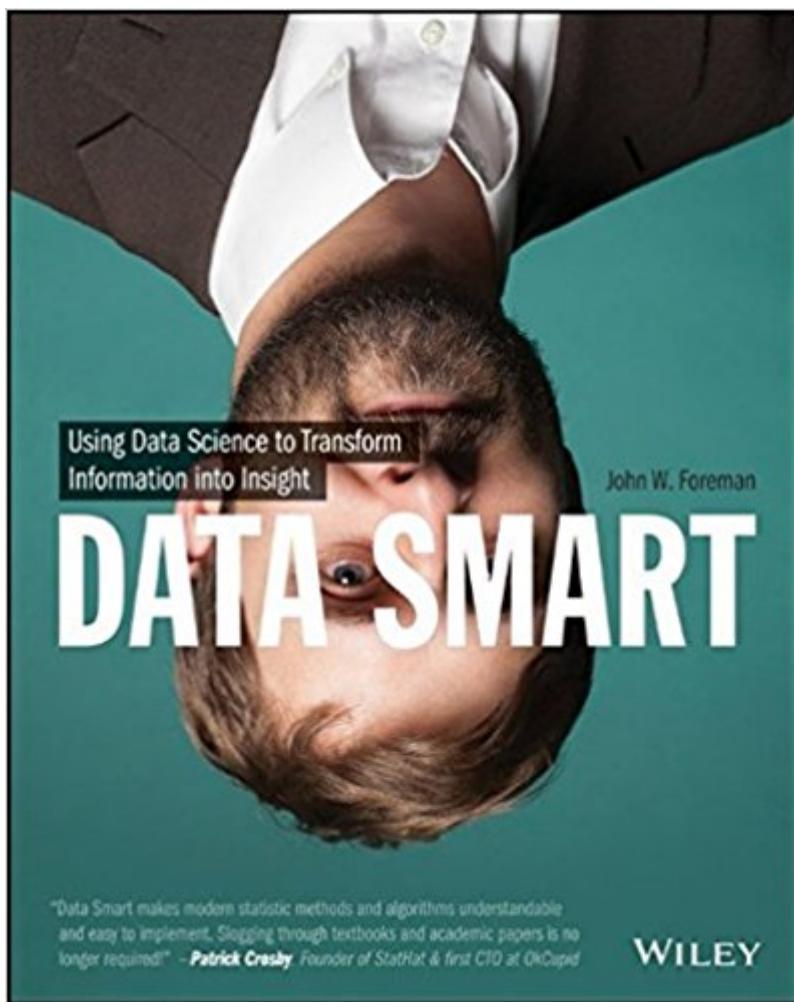## The book was found

# Data Smart: Using Data Science To Transform Information Into Insight

## Synopsis

Data Science gets thrown around in the press like it's magic. Major retailers are predicting everything from when their customers are pregnant to when they want a new pair of Chuck Taylors. It's a brave new world where seemingly meaningless data can be transformed into valuable insight to drive smart business decisions.But how does one exactly do data science? Do you have to hire one of these priests of the dark arts, the "data scientist," to extract this gold from your data? Nope.Data science is little more than using straight-forward steps to process raw data into actionable insight. And inÃ Â Data Smart, author and data scientist John Foreman will show you how that's done within the familiar environment of a spreadsheet.

## Book Information

Paperback: 432 pages

Publisher: Wiley; 1 edition (November 4, 2013)

Language: English

ISBN-10: 111866146X

ISBN-13: 978-1118661468

Product Dimensions:  7.2 x 1 x 9.2 inches

Shipping Weight: 1.4 pounds (View shipping rates and policies)

Average Customer Review:    4.7 out of 5 stars     125 customer reviews

Best Sellers Rank: #14,010 in Books (See Top 100 in Books)   #1 inÃ Â Books > Computers & Technology > Computer Science > Computer Simulation   #10 inÃ Â Books > Textbooks > Computer Science > Artificial Intelligence   #12 inÃ Â Books > Computers & Technology > Databases & Big Data > Data Mining

## Customer Reviews

"Data Smart makes modern statistic methods and algorithms understandable and easy to implement. Slogging through textbooks and academic papers is no longer required!" Ã¢â ¬â •Patrick Crosby, Founder of StatHat & first CTO at OkCupid "When Mr. Foreman interviewed for a job at my company, he arrived dressed in a 'Kentucky Colonel' kind of suit and spoke about nonsensical things like barbecue, lasers, and orange juice pulp. Then, he explained how to de-mystify and solve just about any complex 'big data' problem in our company with simple spreadsheets. No server clusters, mainframes, or Hadoop-a-ma-jigs. Just Excel. I hired him on the spot. After reading this book, you too will learn how to use math and basic spreadsheet formulas to improve your business or, at the very least, how to trick senior executives into hiring you as their

data scientist." Ã¢â ¬â •Ben Chestnut, Founder & CEO of MailChimp "You need a John Foreman on your analytics team. But if you can't have John, then reading this book is the next best thing." Ã¢â ¬â •Patrick Lennon, Director of Analytics, The Coca-Cola Company Most people are approaching data science all wrong. Here's how to do it right. Not to disillusion you, but data scientists are not mystical practitioners of magical arts. Data science is something you can do. Really. This book shows you the significant data science techniques, how they work, how to use them, and how they benefit your business, large or small. It's not about coding or database technologies. It's about turning raw data into insight you can act upon, and doing it as quickly and painlessly as possible. Roll up your sleeves and let's get going. Relax Ã¢â ¬â • it's just a spreadsheet Visit the companion website at www.wiley.com/go/datasmart to download spreadsheets for each chapter, and follow them as you learn about:  Artificial intelligence using the general linear model, ensemble methods, and naive Bayes Clustering via k-means, spherical k-means, and graph modularity Mathematical optimization, including non-linear programming and genetic algorithms Working with time series data and forecasting with exponential smoothing Using Monte Carlo simulation to quantify and address risk Detecting outliers in single or multiple dimensions Exploring the data-science-focused R language

John W. Foreman is Chief Data Scientist for MailChimp.com, where he leads a data science product development effort called the Email Genome Project. As an analytics consultant, John has created data science solutions for The Coca-Cola Company, Royal Caribbean International, Intercontinental Hotels Group, Dell, the Department of Defense, the IRS, and the FBI.

I would go as far as to say that the book is brilliant.First, a drop about me from the standpoint of this book. I have been an IT professional for many years specializing in programming, database, and MS Office add-ons. Part of my job entails self enrichment, that is, expand my working knowledge in areas potentially important for my job. I chose Foreman's book to help with this task for a number of reasons: a) Data Science is a hot area and my company does have a Data Science group, b) I have lots of data experience under my belt - I felt that it would be nice for once to get some useful information from the data, and c) I have a really good Excel background - so I figured that Foreman's approach would be perfect for me - little did I know that I would seriously add to my Excel bag of tricks.The author makes the assumptions that: a) the reader is somewhat technical, b) he knows nothing about Data Science, and c) he is relatively comfortable working in Excel.Reading the book is a joy because Foreman has a cozy, chummy style. He definitely doesn't throw all the

technical stuff at the reader rat-tat-tat machine gun style like many other authors. Instead, Foreman gently introduces his topics and then ramps up technical details carefully. This most definitely helps the learning process.Speaking of learning, by the end of the you will have learned important concepts in "machine learning" and I believe that you will be ready for the next step. I sure was. I found the topics interesting and I wanted to learn more. This is where the book's only problem area comes into play - the next step. Foreman has 3 references - one good, but minor, one terrible, and the other is inappropriate. Let me explain.Foreman recommends a free resource as a follow-on to his Forecasting Chapter. This is a good reference, but I believe that Forecasting is a minor topic in Data Science, unless, of course, Forecasting becomes your thing.Foreman's main reference is: "Data Mining with R" by Luis Torgo. Foreman recommends this as the next step after his book.I tried to read this several times, but couldn't. It certainly wasn't my next step.The other reference, "The Elements of Statistical Learning" by Trevor Hastie, et. al, is totally inappropriate for Data Science newbies. You can checkout the  reviews for this book and you'll see that you need a pretty serious background in statistics to get anything out of that reference. In fact, the author Hastie says as much in his next book "An Introduction to Statistical Learning- with Applications in R". This is the appropriate next step, but I'll get to that in a moment.Here are my recommendations:A. Read Foreman's book and follow along with him in working through the Excel spreadsheets. This is a first step in getting comfortable with Machine Learning.B. Take the Coursera courses: 1) Machine Learning Foundations: A Case Study Approach, and 2) Machine Learning: Regression. The courses are free unless you want completion certificates, in which case there is nominal cost.C. Now you are ready for:Ã  Â An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)Ã  Â This book is also available for free by the authors - check online.

Disclaimer: I served as a paid technical editor for Data Smart. I am not affiliated with the publisher, but I did receive a small fee for double-checking the book's mathematical content before it went to press. I also went to elementary school with the author. So as you read the rest of the review, keep in mind that this reviewer's judgment could be clouded by my lifelong allegiance to Lookout Mountain Elementary School, as well as the Scarface-esque pile of one dollar bills currently sitting on my kitchen table.Anyway, books about "Data" seem to fit into one of the following categories:* Extremely technical gradate-level mathematics books with lots of Greek letters and summation signs* Pie-in-the-sky business bestsellers about how "Data" is going to revolutionize the world as we know it. (I call these "Moneyball" books)* Technical books about the hottest new "Big Data" technology such as R and HadoopData Smart is none of these. Unlike "Moneyball" books, Data

Smart contains enough practical information to actually start performing analyses. Unlike most textbooks, it doesn't get bogged down in mathematical notation. And unlike books about R or the distributed data blah-blah du jour, all the examples use good old Microsoft Excel. It's geared toward competent analysts who are comfortable with Excel and aren't afraid of thinking about problems in a mathematical way. It's goal isn't to "revolutionize" your business with million-dollar software, but rather to make incremental improvements to processes with accessible analytic techniques.I don't work at a big company, so I can't attest to the number of dollars your company will save by applying the book's methods. But I can attest that the author makes difficult mathematical concepts accessible with his quirky sense of humor and gift for metaphor. For example, I previously had not been exposed to the nitty-gritty of clustering techniques. After a couple of hours with the clustering chapters, which include illuminating diagrams and spreadsheet formulas, I felt like I had a good handle on the concepts, and would feel comfortable implementing the ideas in Excel -- or any other language, for that matter.What I like most about the book is that it doesn't try to wave a magic data wand to cure all of your company's ills. Instead it focuses on a few areas where data and analytic techniques can deliver a concrete benefit, and gives you just enough to get started. In particular:* Optimization techniques (Ch. 4) can systematically reduce the cost of manufacturing inputs* Clustering techniques (Ch. 2 and 5) can deliver insights into customer behavior* Predictive techniques (Ch. 3, 6, and 7) can increase margins with better predictions of uncertain outcomes* Forecasting techniques (Ch. 8) can reduce waste with better demand planningIt may take some creativity to figure out how to apply the methods to your own business processes, but all of the techniques are "tried and true" in the sense of being widely deployed at large companies with big analytics budgets and teams of Ph.D.'s on staff. This book's contribution is to make these techniques available to anyone with a little background in applied mathematics and a copy of Excel. For that reason, despite the absence of glitter and/or Jack Welch on the book's cover, I think Data Smart is an important business book.I had a few criticisms of the book as I was reading drafts, but almost all of them were addressed before the final revision. For the sake of completeness, I'll tell you what they were. Some of the chapters ran on a bit long, but these have been split up into manageable pieces. The Optimization chapter is a bit of a doozie, and used to be at the very beginning, but the reader can now "warm up" with some easier chapters on clustering and simple Bayesian techniques. The Regression chapter originally didn't discuss Receiver Operating Characteristic curves, which are important for evaluating predictive models visually, but now ROC curves are abundant.Only one real criticism from me remains: I would have liked to see more on quantile regression, which is only mentioned in passing. It's a great technique for dealing with

outlier-heavy data. The book by Koenker has good but highly mathematical coverage, and I would have loved to see this subject given the Foreman treatment. But, you can't have everything, and I suppose John needs to leave some material for Data Smart 2: The Spreadsheet of Doom.In sum, Data Smart is a well-written and engaging guide to getting new insights from data using familiar tools. The techniques aren't really cutting-edge -- in fact, most have been around for decades -- but to my knowledge this is the first time they've been presented in a way that Excel-slinging business analysts can apply the methods without needing her own team of operations researchers and data scientists. If you're not sure whether the book's sophistication is on par with your own skills, you can download a complete sample chapter (as well as example spreadsheets) from the author's website.One last thing: unlike many books with a technical bent, the prose is engaging and extremely clear. I think this can be traced to John's childhood. When John misbehaved, his father (who is a professor of English) would punish John by forcing him to read a novel by Charles Dickens. Minor infractions resulted in A Christmas Carol being meted out, and when he was really bad he had to read Great Expectations. This is a true story which you should ask John about if you see him at a book-signing event.

Data Smart: Using Data Science to Transform Information into Insight Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics (Big Data Analytics, Data Science, Business Intelligence Book 6) Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data Book 1) Big Data For Business: Your Comprehensive Guide to Understand Data Science, Data Analytics and Data Mining to Boost More Growth and Improve Business - Data Analytics Book, Series 2 Insight Guides: St Petersburg Smart Guide (Insight Smart Guide) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) R for Data Science: Import, Tidy, Transform, Visualize, and Model Data Insight Travel Map: Norway (Insight Travel Maps) 4th (fourth) Edition published by Insight Maps (2013) Barbados Insight Guide (Insight Guides) (Insight Guides) Dusseldorf Insight Guide (Insight Guides) (Insight City Guides) Insight Guide Las Vegas Smartguide (Insight Smart Guide Las Vegas) Insight Map Frankfurt: Fleximap Plus Travel Information (Insight Fleximaps) Insight Map Us & British Virgin Islands: Fleximap Plus Travel Information (Insight Fleximaps) Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data Data Analytics and Python Programming: 2 Bundle Manuscript:

Beginners Guide to Learn Data Analytics, Predictive Analytics and Data Science with Python Programming My Mueller Spiral-Ultra Vegetable Spiralizer Cookbook: 101 Recipes to Turn Zucchini into Pasta, Cauliflower into Rice, Potatoes into Lasagna, Beets into ... Slicer! (Vegetable Spiralizer Cookbooks) Information Sources in Science and Technology, 3rd Edition (Library and Information Science Text (Paperback)) Information Sources in Science and Technology (Library and Information Science Text) Science and Technology Resources: A Guide for Information Professionals and Researchers (Library and Information Science Text)