

1. Record Nr.	UNINA9910973988803321
Autore	Foreman John W.
Titolo	Data smart : using data science to transform information into insight / / John W. Foreman
Pubbl/distr/stampa	Indianapolis : , : Wiley, , [2014] ©2014
ISBN	1-118-83986-2 1-118-66148-6
Edizione	[1st ed.]
Descrizione fisica	1 online resource (434 p.)
Disciplina	006.312
Soggetti	Data mining Web sites - Design Web usage mining
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Cover; Title Page; Copyright; Contents; Chapter 1 Everything You Ever Needed to Know about Spreadsheets but Were Too Afraid to Ask; Some Sample Data; Moving Quickly with the Control Button; Copying Formulas and Data Quickly; Formatting Cells; Paste Special Values; Inserting Charts; Locating the Find and Replace Menus; Formulas for Locating and Pulling Values; Using VLOOKUP to Merge Data; Filtering and Sorting; Using PivotTables; Using Array Formulas; Solving Stuff with Solver; OpenSolver: I Wish We Didn't Need This, but We Do; Wrapping Up Chapter 2 Cluster Analysis Part I: Using K-Means to Segment Your Customer Base Girls Dance with Girls, Boys Scratch Their Elbows; Getting Real: K-Means Clustering Subscribers in E-mail Marketing; Joey Bag O' Donuts Wholesale Wine Emporium; The Initial Dataset; Determining What to Measure; Start with Four Clusters; Euclidean Distance: Measuring Distances as the Crow Flies; Distances and Cluster Assignments for Everybody!; Solving for the Cluster Centers; Making Sense of the Results; Getting the Top Deals by Cluster; The Silhouette: A Good Way to Let Different K Values Duke It Out How about Five Clusters? Solving for Five Clusters; Getting the Top

Deals for All Five Clusters; Computing the Silhouette for 5-Means Clustering; K-Medians Clustering and Asymmetric Distance Measurements; Using K-Medians Clustering; Getting a More Appropriate Distance Metric; Putting It All in Excel; The Top Deals for the 5-Medians Clusters; Wrapping Up; Chapter 3 Naive Bayes and the Incredible Lightness of Being an Idiot; When You Name a Product Mandrill, You're Going to Get Some Signal and Some Noise; The World's Fastest Intro to Probability Theory; Totaling Conditional Probabilities Joint Probability, the Chain Rule, and Independence What Happens in a Dependent Situation?; Bayes Rule; Using Bayes Rule to Create an AI Model; High-Level Class Probabilities Are Often Assumed to Be Equal; A Couple More Odds and Ends; Let's Get This Excel Party Started; Removing Extraneous Punctuation; Splitting on Spaces; Counting Tokens and Calculating Probabilities; And We Have a Model! Let's Use It; Wrapping Up; Chapter 4 Optimization Modeling: Because That "Fresh Squeezed" Orange Juice Ain't Gonna Blend Itself; Why Should Data Scientists Know Optimization? Starting with a Simple Trade-Off Representing the Problem as a Polytope; Solving by Sliding the Level Set; The Simplex Method: Rooting around the Corners; Working in Excel; There's a Monster at the End of This Chapter; Fresh from the Grove to Your Glass...with a Pit Stop Through a Blending Model; You Use a Blending Model; Let's Start with Some Specs; Coming Back to Consistency; Putting the Data into Excel; Setting Up the Problem in Solver; Lowering Your Standards; Dead Squirrel Removal: The Minimax Formulation; If-Then and the "Big M" Constraint Multiplying Variables: Cranking Up the Volume to 11

Sommario/riassunto

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.
