

Data Science Create Teams That Ask The Right Questions And Deliver Real Value

In this concise book you will learn what you need to know to begin assembling and leading a data science enterprise, even if you have never worked in data science before. You'll get a crash course in data science so that you'll be conversant in the field and understand your role as a leader. You'll also learn how to recruit, assemble, evaluate, and develop a team with complementary skill sets and roles. You'll learn the structure of the data science pipeline, the goals of each stage, and how to keep your team on target throughout. Finally, you'll learn some down-to-earth practical skills that will help you overcome the common challenges that frequently derail data science projects.

At first glance, the skills required to work in the data science field appear to be self-explanatory. Do not be fooled. Impactful data science demands an interdisciplinary knowledge of business philosophy, project management, salesmanship, presentation, and more. In *Managing Your Data Science Projects*, author Robert de Graaf explores important concepts that are frequently overlooked in much of the instructional literature that is available to data scientists new to the field. If your completed models are to be used and maintained most effectively, you must be able to present and sell them within your organization in a compelling way. The value of data science within an organization cannot be overstated. Thus, it is vital that strategies and communication between teams are dexterously managed. Three main ways that data science strategy is used in a company is to research its customers, assess risk analytics, and log operational measurements. These all require different managerial instincts, backgrounds, and experiences, and de Graaf cogently breaks down the unique reasons behind each. They must align seamlessly to eventually be adopted as dynamic models. Data science is a relatively new discipline, and as such, internal processes for it are not as well-developed within an operational business as others. With *Managing Your Data Science Projects*, you will learn how to create products that solve important problems for your customers and ensure that the initial success is sustained throughout the product's intended life. Your users will trust you and your models, and most importantly, you will be a more well-rounded and effectual data scientist throughout your career. Who This Book Is For Early-career data scientists, managers of data scientists, and those interested in entering the field of data science

Putting predictive models into production is one of the most direct ways that data scientists can add value to an organization. By learning how to build and deploy scalable model pipelines, data scientists can own more of the model production process and more rapidly deliver data products. This book provides a hands-on approach to scaling up Python code to work in distributed environments in order to build robust pipelines. Readers will learn how to set up machine learning models as web endpoints, serverless functions, and streaming pipelines using multiple cloud environments. It is intended for analytics practitioners with hands-on experience with Python libraries such as Pandas and scikit-learn, and will focus on scaling up prototype models to production. From startups to trillion dollar companies, data science is playing an important role in helping organizations maximize the value of their data. This book helps data scientists to level up their careers by taking ownership of data products with applied examples that demonstrate how to: Translate models developed on a laptop to scalable deployments in the cloud Develop end-to-end systems that automate data science workflows Own a data product from conception to production The accompanying Jupyter notebooks provide examples of scalable pipelines across multiple cloud environments, tools, and libraries (github.com/bgweber/DS_Production). Book Contents Here are the topics covered by *Data Science in Production*: Chapter 1: Introduction - This chapter will motivate the use of Python and discuss the discipline of applied data science, present the data sets, models, and cloud environments used throughout the book, and provide an overview of automated feature engineering. Chapter 2: Models as Web Endpoints - This chapter shows how to use web endpoints for consuming data and hosting machine learning models as endpoints using the Flask and Gunicorn libraries. We'll start with scikit-learn models and also set up a deep learning endpoint with Keras. Chapter 3: Models as Serverless Functions - This chapter will build upon the previous chapter and show how to set up model endpoints as serverless functions using AWS Lambda and GCP Cloud Functions. Chapter 4: Containers for Reproducible Models - This chapter will show how to use containers for deploying models with Docker. We'll also explore scaling up with ECS and Kubernetes, and building web applications with Plotly Dash. Chapter 5: Workflow Tools for Model Pipelines - This chapter focuses on scheduling automated workflows using Apache Airflow. We'll set up a model that pulls data from BigQuery, applies a model, and saves the results. Chapter 6: PySpark for Batch Modeling - This chapter will introduce readers to PySpark using the community edition of Databricks. We'll build a batch model pipeline that pulls data from a data lake, generates features, applies a model, and stores the results to a No SQL database. Chapter 7: Cloud Dataflow for Batch Modeling - This chapter will introduce the core components of Cloud Dataflow and implement a batch model pipeline for reading data from BigQuery, applying an ML model, and saving the results to Cloud Datastore. Chapter 8: Streaming Model Workflows - This chapter will introduce readers to Kafka and PubSub for streaming messages in a cloud environment. After working through this material, readers will learn how to use these message brokers to create streaming model pipelines with PySpark and Dataflow that provide near real-time predictions. Excerpts of these chapters are available on Medium (@bgweber), and a book sample is available on Leanpub.

Hiring Data Scientists and Machine Learning Engineers is a concise, practical guide to help you hire the right people for your organization. The book will help you navigate the plethora of data science related roles and skills and help you create an effective hiring strategy to suit your organization's needs.

Data Scientists at Work is a collection of interviews with sixteen of the world's most influential and innovative data scientists from across the spectrum of this hot new profession. "Data scientist is the sexiest job in the 21st century," according to the Harvard Business Review. By 2018, the United States will experience a shortage of 190,000 skilled data scientists, according to a McKinsey report. Through incisive in-depth interviews, this book mines the what, how, and why of the practice of data science from the stories, ideas, shop talk, and forecasts of its preeminent practitioners across diverse industries: social network (Yann LeCun, Facebook); professional network (Daniel Tunkelang, LinkedIn); venture capital (Roger Ehrenberg, IA Ventures); enterprise cloud computing and neuroscience (Eric Jonas, formerly Salesforce.com); newspaper and media (Chris Wiggins, The New York Times); streaming television (Caitlin Smallwood, Netflix); music forecast (Victor Hu, Next Big Sound); strategic intelligence (Amy Heineike, Quid); environmental big data (André Karpištšenko, Planet OS); geospatial marketing intelligence (Jonathan Lenaghan, PlaceIQ); advertising (Claudia Perlich, Dstillery); fashion e-commerce (Anna Smith, Rent the Runway); specialty retail (Erin Shellman, Nordstrom); email marketing (John Foreman, MailChimp); predictive sales intelligence (Kira Radinsky, SalesPredict); and humanitarian nonprofit (Jake Porway, DataKind). The book features a stimulating foreword by Google's Director of Research, Peter Norvig. Each of these data scientists shares how he or she tailors the torrent-taming techniques of big data, data visualization, search, and statistics to specific jobs by dint of ingenuity, imagination, patience, and passion. *Data Scientists at Work* parts the curtain on the interviewees' earliest data projects, how they became data scientists, their discoveries and surprises in working with data, their thoughts on the past, present, and future of the profession, their experiences of team collaboration within their organizations, and the insights they have gained as they get their hands dirty refining mountains of raw data into objects of commercial, scientific, and educational value for their organizations and clients. This book provides an introduction to the mathematical and algorithmic foundations of data science, including machine learning, high-dimensional geometry, and analysis of large networks. Topics include the counterintuitive nature of data in high dimensions, important linear algebraic techniques such as singular value decomposition, the theory of random walks and Markov chains, the fundamentals of and

important algorithms for machine learning, algorithms and analysis for clustering, probabilistic models for large networks, representation learning including topic modelling and non-negative matrix factorization, wavelets and compressed sensing. Important probabilistic techniques are developed including the law of large numbers, tail inequalities, analysis of random projections, generalization guarantees in machine learning, and moment methods for analysis of phase transitions in large random graphs. Additionally, important structural and complexity measures are discussed such as matrix norms and VC-dimension. This book is suitable for both undergraduate and graduate courses in the design and analysis of algorithms for data.

Organize, plan, and build an exceptional data analytics team within your organization In *Minding the Machines: Building and Leading Data Science and Analytics Teams*, AI and analytics strategy expert Jeremy Adamson delivers an accessible and insightful roadmap to structuring and leading a successful analytics team. The book explores the tasks, strategies, methods, and frameworks necessary for an organization beginning their first foray into the analytics space or one that is rebooting its team for the umpteenth time in search of success. In this book, you'll discover: A focus on the three pillars of strategy, process, and people and their role in the iterative and ongoing effort of building an analytics team Repeated emphasis on three guiding principles followed by successful analytics teams: start early, go slow, and fully commit The importance of creating clear goals and objectives when creating a new analytics unit in an organization Perfect for executives, managers, team leads, and other business leaders tasked with structuring and leading a successful analytics team, *Minding the Machines* is also an indispensable resource for data scientists and analysts who seek to better understand how their individual efforts fit into their team's overall results.

Analyze your data and delve deep into the world of machine learning with the latest Spark version, 2.0 About This Book Perform data analysis and build predictive models on huge datasets that leverage Apache Spark Learn to integrate data science algorithms and techniques with the fast and scalable computing features of Spark to address big data challenges Work through practical examples on real-world problems with sample code snippets Who This Book Is For This book is for anyone who wants to leverage Apache Spark for data science and machine learning. If you are a technologist who wants to expand your knowledge to perform data science operations in Spark, or a data scientist who wants to understand how algorithms are implemented in Spark, or a newbie with minimal development experience who wants to learn about Big Data Analytics, this book is for you! What You Will Learn Consolidate, clean, and transform your data acquired from various data sources Perform statistical analysis of data to find hidden insights Explore graphical techniques to see what your data looks like Use machine learning techniques to build predictive models Build scalable data products and solutions Start programming using the RDD, DataFrame and Dataset APIs Become an expert by improving your data analytical skills In Detail This is the era of Big Data. The words 'Big Data' implies big innovation and enables a competitive advantage for businesses. Apache Spark was designed to perform Big Data analytics at scale, and so Spark is equipped with the necessary algorithms and supports multiple programming languages. Whether you are a technologist, a data scientist, or a beginner to Big Data analytics, this book will provide you with all the skills necessary to perform statistical data analysis, data visualization, predictive modeling, and build scalable data products or solutions using Python, Scala, and R. With ample case studies and real-world examples, *Spark for Data Science* will help you ensure the successful execution of your data science projects. Style and approach This book takes a step-by-step approach to statistical analysis and machine learning, and is explained in a conversational and easy-to-follow style. Each topic is explained sequentially with a focus on the fundamentals as well as the advanced concepts of algorithms and techniques. Real-world examples with sample code snippets are also included.

Publisher's Note: Products purchased from Third Party sellers are not guaranteed by the publisher for quality, authenticity, or access to any online entitlements included with the product. Use machine learning to understand your customers, frame decisions, and drive value The business analytics world has changed, and Data Scientists are taking over. *Business Data Science* takes you through the steps of using machine learning to implement best-in-class business data science. Whether you are a business leader with a desire to go deep on data, or an engineer who wants to learn how to apply Machine Learning to business problems, you'll find the information, insight, and tools you need to flourish in today's data-driven economy. You'll learn how to: •Use the key building blocks of Machine Learning: sparse regularization, out-of-sample validation, and latent factor and topic modeling•Understand how use ML tools in real world business problems, where causation matters more than correlation•Solve data science programs by scripting in the R programming language Today's business landscape is driven by data and constantly shifting. Companies live and die on their ability to make and implement the right decisions quickly and effectively. *Business Data Science* is about doing data science right. It's about the exciting things being done around Big Data to run a flourishing business. It's about the precepts, principals, and best practices that you need know for best-in-class business data science.

The design patterns in this book capture best practices and solutions to recurring problems in machine learning. The authors, three Google engineers, catalog proven methods to help data scientists tackle common problems throughout the ML process. These design patterns codify the experience of hundreds of experts into straightforward, approachable advice. In this book, you will find detailed explanations of 30 patterns for data and problem representation, operationalization, repeatability, reproducibility, flexibility, explainability, and fairness. Each pattern includes a description of the problem, a variety of potential solutions, and recommendations for choosing the best technique for your situation. You'll learn how to: Identify and mitigate common challenges when training, evaluating, and deploying ML models Represent data for different ML model types, including embeddings, feature crosses, and more Choose the right model type for specific problems Build a robust training loop that uses checkpoints, distribution strategy, and hyperparameter tuning Deploy scalable ML systems that you can retrain and update to reflect new data Interpret model predictions for stakeholders and ensure models are treating users fairly

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 in *DataSmart*, author and data scientist John Foreman will show you how that's done within the familiar environment of a spreadsheet. Why a spreadsheet? It's comfortable! You get to look at the data every step of the way, building confidence as you learn the tricks of the trade. Plus, spreadsheets are a vendor-neutral place to learn data science without the hype. But don't let the Excel sheets fool you. This is a book for those serious about learning the analytic techniques, the math and the magic, behind big data. Each chapter will cover a different technique in a spreadsheet so you can follow along: Mathematical optimization, including non-linear programming and genetic algorithms Clustering via k-means, spherical k-means, and graph modularity Data mining in graphs, such as outlier detection Supervised AI through logistic regression, ensemble models, and bag-of-words models Forecasting, seasonal adjustments, and prediction intervals through monte carlo simulation Moving from spreadsheets into the R programming language You get your hands dirty as you work alongside John through

each technique. But never fear, the topics are readily applicable and the author laces humor throughout. You'll even learn what a dead squirrel has to do with optimization modeling, which you no doubt are dying to know.

"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"--

An Introduction to Data Science by Jeffrey S. Saltz and Jeffrey M. Stanton is an easy-to-read, gentle introduction for people with a wide range of backgrounds into the world of data science. Needing no prior coding experience or a deep understanding of statistics, this book uses the R programming language and RStudio® platform to make data science welcoming and accessible for all learners. After introducing the basics of data science, the book builds on each previous concept to explain R programming from the ground up. Readers will learn essential skills in data science through demonstrations of how to use data to construct models, predict outcomes, and visualize data.

Data science libraries, frameworks, modules, and toolkits are great for doing data 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

Learn the techniques and math you need to start making sense of your data About This Book Enhance your knowledge of coding with data science theory for practical insight into data science and analysis More than just a math class, learn how to perform real-world data science tasks with R and Python Create actionable insights and transform raw data into tangible value Who This Book Is For You should be fairly well acquainted with basic algebra and should feel comfortable reading snippets of R/Python as well as pseudo code. You should have the urge to learn and apply the techniques put forth in this book on either your own data sets or those provided to you. If you have the basic math skills but want to apply them in data science or you have good programming skills but lack math, then this book is for you. What You Will Learn Get to know the five most important steps of data science Use your data intelligently and learn how to handle it with care Bridge the gap between mathematics and programming Learn about probability, calculus, and how to use statistical models to control and clean your data and drive actionable results Build and evaluate baseline machine learning models Explore the most effective metrics to determine the success of your machine learning models Create data visualizations that communicate actionable insights Read and apply machine learning concepts to your problems and make actual predictions In Detail Need to turn your skills at programming into effective data science skills? Principles of Data Science is created to help you join the dots between mathematics, programming, and business analysis. With this book, you'll feel confident about asking—and answering—complex and sophisticated questions of your data to move from abstract and raw statistics to actionable ideas. With a unique approach that bridges the gap between mathematics and computer science, this book takes you through the entire data science pipeline. Beginning with cleaning and preparing data, and effective data mining strategies and techniques, you'll move on to build a comprehensive picture of how every piece of the data science puzzle fits together. Learn the fundamentals of computational mathematics and statistics, as well as some pseudocode being used today by data scientists and analysts. You'll get to grips with machine learning, discover the statistical models that help you take control and navigate even the densest datasets, and find out how to create powerful visualizations that communicate what your data means. Style and approach This is an easy-to-understand and accessible tutorial. It is a step-by-step guide with use cases, examples, and illustrations to get you well-versed with the concepts of data science. Along with explaining the fundamentals, the book will also introduce you to slightly advanced concepts later on and will help you implement these techniques in the real world.

Master the skills necessary to hire and manage a team of highly skilled individuals to design, build, and implement applications and systems based on advanced analytics and AI Key Features Learn to create an operationally effective advanced analytics team in a corporate environment Select and undertake projects that have a high probability of success and deliver the improved top and bottom-line results Understand how to create relationships with executives, senior managers, peers, and subject matter experts that lead to team collaboration, increased funding, and long-term success for you and your team Book Description In Building Analytics Teams, John K. Thompson, with his 30+ years of experience and expertise, illustrates the fundamental concepts of building and managing a high-performance analytics team, including what to do, who to hire, projects to undertake, and what to avoid in the journey of building an analytically sound team. The core processes in creating an effective analytics team and the importance of the business decision-making life cycle are explored to help achieve initial and sustainable success. The book demonstrates the various traits of a successful and high-performing analytics team and then delineates the path to achieve this with insights on the mindset, advanced analytics models, and predictions based on data analytics. It also emphasizes the significance of the macro and micro processes required to evolve in response to rapidly changing business needs. The book dives into the methods and practices of managing, developing, and leading an analytics team. Once you've brought the team up to speed, the book explains how to govern executive expectations and select winning projects. By the end of this book, you will have acquired the knowledge to create an effective business analytics team and develop a production environment that delivers ongoing operational improvements for your organization. What you will learn Avoid organizational and technological pitfalls of moving from a defined project to a production environment Enable team members to focus on higher-value work and tasks Build Advanced Analytics and Artificial Intelligence (AA&AI) functions in an organization Outsource certain projects to competent and capable third parties Support the operational areas that intend to invest in business intelligence, descriptive statistics, and small-scale predictive analytics Analyze the operational area, the processes, the data, and the organizational resistance Who this book is for This book is for senior executives, senior and junior managers, and those who are working as part of a team that is accountable for designing, building, delivering and ensuring business success through advanced analytics and artificial intelligence systems and applications.

At least 5 to 10 years of experience in driving your organization to a higher level of efficiency will be helpful.

Learn how to embed data science, Big Data and AI in your organization's decision-making process and make your organization more data-driven, profitable, and intelligent in 10 steps. Book description This book covers every aspect of the implementation of data science, from the algorithms that make your decisions more refined, effective and faster to the people, skills, culture, and mindset required to make it happen. How do you set the right KPIs and targets? How are the best data-driven organizations structured? Why do you need a data warehouse or data lake? How do you manage a data science project? This book tackles every question relevant to implementing data science. Many organizations start by collecting data without a goal, but that data science approach is doomed to fail. This book takes you through the process of implementing data science from the ground floor all the way to the top. It all starts with the question: what do we want to achieve? It covers all the subsequent steps on a macro and micro level, from the process of registering data, to processing it, to the organization's response. All the relevant data science techniques and technologies are discussed, from algorithms and AI to the right management strategies. Based on many practical case studies and best practices, this book reveals what works and what doesn't. Benefit from the author's many years of experience in making organizations more intelligent and data-driven as a consultant and an educator. What you will learn - The most important benefits of data science. - The essential aspects of decision making and the role of data science. - How to determine the right KPIs and use them to manage effectively. - How to turn data into knowledge and information. - How to make your organization more agile. - The many types of algorithms that can be used to make more effective decisions on every level. - How to manage data science projects - who and what do you need to effectively implement data science? - How to design a data science roadmap. - And much, much more. Who is this book for This book is for every manager or professional, and all those who want to learn how to embed the effective use of data science in every facet of the organization. This comprehensive management handbook is a must-read for (business) consultants, business managers, Chief Data Officers (CDOs), CIOs, and other executives, project managers, Data Science consultants, Data Scientists, AI consultants, (business) controllers, quality managers, and BI consultants.

Learn how to build a data science team within your organization rather than hiring from the outside. Teach your team to ask the right questions to gain actionable insights into your business. Most organizations still focus on objectives and deliverables. Instead, a data science team is exploratory. They use the scientific method to ask interesting questions and run small experiments. Your team needs to see if the data illuminate their questions. Then, they have to use critical thinking techniques to justify their insights and reasoning. They should pivot their efforts to keep their insights aligned with business value. Finally, your team needs to deliver these insights as a compelling story. Insight!: How to Build Data Science Teams that Deliver Real Business Value shows that the most important thing you can do now is help your team think about data. Management coach Doug Rose walks you through the process of creating and managing effective data science teams. You will learn how to find the right people inside your organization and equip them with the right mindset. The book has three overarching concepts: You should mine your own company for talent. You can't change your organization by hiring a few data science superheroes. You should form small, agile-like data teams that focus on delivering valuable insights early and often. You can make real changes to your organization by telling compelling data stories. These stories are the best way to communicate your insights about your customers, challenges, and industry. What You Will Learn: Create data science teams from existing talent in your organization to cost-efficiently extract maximum business value from your organization's data Understand key data science terms and concepts Follow practical guidance to create and integrate an effective data science team with key roles and the responsibilities for each team member Utilize the data science life cycle (DSL) to model essential processes and practices for delivering value Use sprints and storytelling to help your team stay on track and adapt to new knowledge Who This Book Is For Data science project managers and team leaders. The secondary readership is data scientists, DBAs, analysts, senior management, HR managers, and performance specialists.

Written by renowned data science experts Foster Provost and Tom Fawcett, Data Science for Business introduces the fundamental principles of data science, and walks you through the "data-analytic thinking" necessary for extracting useful knowledge and business value from the data you collect. This guide also helps you understand the many data-mining techniques in use today. Based on an MBA course Provost has taught at New York University over the past ten years, Data Science for Business provides examples of real-world business problems to illustrate these principles. You'll not only learn how to improve communication between business stakeholders and data scientists, but also how participate intelligently in your company's data science projects. You'll also discover how to think data-analytically, and fully appreciate how data science methods can support business decision-making. Understand how data science fits in your organization—and how you can use it for competitive advantage Treat data as a business asset that requires careful investment if you're to gain real value Approach business problems data-analytically, using the data-mining process to gather good data in the most appropriate way Learn general concepts for actually extracting knowledge from data Apply data science principles when interviewing data science job candidates

Now that people are aware that data can make the difference in an election or a business model, data science as an occupation is gaining ground. But how can you get started working in a wide-ranging, interdisciplinary field that's so clouded in hype? This insightful book, based on Columbia University's Introduction to Data Science class, tells you what you need to know. In many of these chapter-long lectures, data scientists from companies such as Google, Microsoft, and eBay share new algorithms, methods, and models by presenting case studies and the code they use. If you're familiar with linear algebra, probability, and statistics, and have programming experience, this book is an ideal introduction to data science. Topics include: Statistical inference, exploratory data analysis, and the data science process Algorithms Spam filters, Naive Bayes, and data wrangling Logistic regression Financial modeling Recommendation engines and causality Data visualization Social networks and data journalism Data engineering, MapReduce, Pregel, and Hadoop Doing Data Science is collaboration between course instructor Rachel Schutt, Senior VP of Data Science at News Corp, and data science consultant Cathy O'Neil, a senior data scientist at Johnson Research Labs, who attended and blogged about the course.

All the answers to your data science questions Over half of all businesses are using data science to generate insights and value from big data. How are they doing it? Data Science Strategy For Dummies answers all your questions about how to build a data science capability from scratch, starting with the "what" and the "why" of data science and covering what it takes to lead and nurture a top-notch team of data scientists. With this book, you'll learn how to incorporate data science as a strategic function into any business, large or small. Find solutions to your real-life challenges as you uncover the stories and value hidden within data. Learn exactly what data science is and why it's important Adopt a data-driven mindset as the foundation to success Understand the processes and common roadblocks behind data science Keep your data science program focused on generating business value Nurture a top-quality data science team In non-technical language, Data Science Strategy For Dummies outlines new perspectives and strategies to effectively lead analytics and data science functions to create real value.

Data science teams looking to turn research into useful analytics applications require not only the right tools, but also the right approach if they're to succeed. With the revised second edition of this hands-on guide, up-and-coming data scientists will learn how to use the Agile Data Science development methodology to build data applications with Python, Apache Spark, Kafka, and other tools. Author Russell Journey demonstrates how to compose a data platform for building, deploying, and refining analytics applications with Apache Kafka, MongoDB, Elasticsearch, d3.js, scikit-learn, and Apache Airflow. You'll learn an iterative approach that lets you quickly change the kind of analysis you're doing, depending on what the data is telling you. Publish data science work as a web application, and affect meaningful

change in your organization. Build value from your data in a series of agile sprints, using the data-value pyramid Extract features for statistical models from a single dataset Visualize data with charts, and expose different aspects through interactive reports Use historical data to predict the future via classification and regression Translate predictions into actions Get feedback from users after each sprint to keep your project on track

Data Science in Education Using R is the go-to reference for learning data science in the education field. The book answers questions like: What does a data scientist in education do? How do I get started learning R, the popular open-source statistical programming language? And what does a data analysis project in education look like? If you're just getting started with R in an education job, this is the book you'll want with you. This book gets you started with R by teaching the building blocks of programming that you'll use many times in your career. The book takes a "learn by doing" approach and offers eight analysis walkthroughs that show you a data analysis from start to finish, complete with code for you to practice with. The book finishes with how to get involved in the data science community and how to integrate data science in your education job. This book will be an essential resource for education professionals and researchers looking to increase their data analysis skills as part of their professional and academic development.

Summary You are going to need more than technical knowledge to succeed as a data scientist. **Build a Career in Data Science** teaches you what school leaves out, from how to land your first job to the lifecycle of a data science project, and even how to become a manager. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology What are the keys to a data scientist's long-term success? Blending your technical know-how with the right "soft skills" turns out to be a central ingredient of a rewarding career. About the book **Build a Career in Data Science** is your guide to landing your first data science job and developing into a valued senior employee. By following clear and simple instructions, you'll learn to craft an amazing resume and ace your interviews. In this demanding, rapidly changing field, it can be challenging to keep projects on track, adapt to company needs, and manage tricky stakeholders. You'll love the insights on how to handle expectations, deal with failures, and plan your career path in the stories from seasoned data scientists included in the book. What's inside **Creating a portfolio of data science projects** Assessing and negotiating an offer Leaving gracefully and moving up the ladder Interviews with professional data scientists About the reader For readers who want to begin or advance a data science career. About the author Emily Robinson is a data scientist at Warby Parker. Jacqueline Nolis is a data science consultant and mentor. Table of Contents: PART 1 - GETTING STARTED WITH DATA SCIENCE 1. What is data science? 2. Data science companies 3. Getting the skills 4. Building a portfolio PART 2 - FINDING YOUR DATA SCIENCE JOB 5. The search: Identifying the right job for you 6. The application: Résumés and cover letters 7. The interview: What to expect and how to handle it 8. The offer: Knowing what to accept PART 3 - SETTLING INTO DATA SCIENCE 9. The first months on the job 10. Making an effective analysis 11. Deploying a model into production 12. Working with stakeholders PART 4 - GROWING IN YOUR DATA SCIENCE ROLE 13. When your data science project fails 14. Joining the data science community 15. Leaving your job gracefully 16. Moving up the ladder

The essential guide for data scientists and for leaders who must get more from their data science teams **The Economist** boldly claims that data are now "the world's most valuable resource." But, as Kenett and Redman so richly describe, unlocking that value requires far more than technical excellence. **The Real Work of Data Science** explores understanding the problems, dealing with quality issues, building trust with decision makers, putting data science teams in the right organizational spots, and helping companies become data-driven. This is the work that spells the difference between a good data scientist and a great one, between a team that makes marginal contributions and one that drives the business, between a company that gains some value from its data and one in which data truly is "the most valuable resource." "These two authors are world-class experts on analytics, data management, and data quality; they've forgotten more about these topics than most of us will ever know. Their book is pragmatic, understandable, and focused on what really counts. If you want to do data science in any capacity, you need to read it." —Thomas H. Davenport, Distinguished Professor, Babson College and Fellow, MIT Initiative on the Digital Economy "I like your book. The chapters address problems that have faced statisticians for generations, updated to reflect today's issues, such as computational Big Data." —Sir David Cox, Warden of Nuffield College and Professor of Statistics, Oxford University "Data science is critical for competitiveness, for good government, for correct decisions. But what is data science? Kenett and Redman give, by far, the best introduction to the subject I have seen anywhere. They address the critical questions of formulating the right problem, collecting the right data, doing the right analyses, making the right decisions, and measuring the actual impact of the decisions. This book should become required reading in statistics and computer science departments, business schools, analytics institutes and, most importantly, by all business managers." —A. Blanton Godfrey, Joseph D. Moore Distinguished University Professor, Wilson College of Textiles, North Carolina State University

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. · 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.

Gain a practical introduction to **DataOps**, a new discipline for delivering data science at scale inspired by practices at companies such as Facebook, Uber, LinkedIn, Twitter, and eBay. Organizations need more than the latest AI algorithms, hottest tools, and best people to turn data into insight-driven action and useful analytical data products. Processes and thinking employed to manage and use data in the 20th century are a bottleneck for working effectively with the variety of data and advanced analytical use cases that organizations have today. This book provides the approach and methods to ensure continuous rapid use of data to create analytical data products and steer decision making. **Practical DataOps** shows you how to optimize the data supply chain from diverse raw data sources to the final data product, whether the goal is a machine learning model or other data-orientated output. The book provides an approach to eliminate wasted effort and improve collaboration between data producers, data consumers, and the rest of the organization through the adoption of lean thinking and agile software development principles. This book helps you to improve the speed and accuracy of analytical application development through data management and DevOps practices that securely expand data access, and rapidly increase the number of reproducible data products through automation, testing, and integration. The book also shows how to collect feedback and monitor performance to manage and continuously improve your processes and output. **What You Will Learn** Develop a data strategy for your organization to help it reach its long-term goals Recognize and eliminate barriers to delivering data to users at scale Work on the right things for the right stakeholders through agile collaboration Create trust in data via rigorous testing and effective data management Build a culture of learning and continuous improvement through monitoring deployments and measuring outcomes Create cross-functional self-organizing teams focused on goals not reporting lines Build robust, trustworthy, data pipelines in support of AI, machine learning, and other analytical data products **Who This Book Is For** Data science and advanced analytics experts, CIOs, CDOs (chief data officers), chief analytics officers, business analysts, business team leaders, and IT professionals (data engineers, developers, architects, and DBAs) supporting data

teams who want to dramatically increase the value their organization derives from data. The book is ideal for data professionals who want to overcome challenges of long delivery time, poor data quality, high maintenance costs, and scaling difficulties in getting data science output and machine learning into customer-facing production.

As data science evolves to become a business necessity, the importance of assembling a strong and innovative data teams grows. In this in-depth report, data scientist DJ Patil explains the skills, perspectives, tools and processes that position data science teams for success. Topics include: What it means to be "data driven." The unique roles of data scientists. The four essential qualities of data scientists. Patil's first-hand experience building the LinkedIn data science team.

Understand data science concepts and methodologies to manage and deliver top-notch solutions for your organization
Key Features
Learn the basics of data science and explore its possibilities and limitations
Manage data science projects and assemble teams effectively even in the most challenging situations
Understand management principles and approaches for data science projects to streamline the innovation process
Book Description
Data science and machine learning can transform any organization and unlock new opportunities. However, employing the right management strategies is crucial to guide the solution from prototype to production. Traditional approaches often fail as they don't entirely meet the conditions and requirements necessary for current data science projects. In this book, you'll explore the right approach to data science project management, along with useful tips and best practices to guide you along the way. After understanding the practical applications of data science and artificial intelligence, you'll see how to incorporate them into your solutions. Next, you will go through the data science project life cycle, explore the common pitfalls encountered at each step, and learn how to avoid them. Any data science project requires a skilled team, and this book will offer the right advice for hiring and growing a data science team for your organization. Later, you'll be shown how to efficiently manage and improve your data science projects through the use of DevOps and ModelOps. By the end of this book, you will be well versed with various data science solutions and have gained practical insights into tackling the different challenges that you'll encounter on a daily basis. What you will learn
Understand the underlying problems of building a strong data science pipeline
Explore the different tools for building and deploying data science solutions
Hire, grow, and sustain a data science team
Manage data science projects through all stages, from prototype to production
Learn how to use ModelOps to improve your data science pipelines
Get up to speed with the model testing techniques used in both development and production stages
Who this book is for
This book is for data scientists, analysts, and program managers who want to use data science for business productivity by incorporating data science workflows efficiently. Some understanding of basic data science concepts will be useful to get the most out of this book.

Data science is emerging as a field that is revolutionizing science and industries alike. Work across nearly all domains is becoming more data driven, affecting both the jobs that are available and the skills that are required. As more data and ways of analyzing them become available, more aspects of the economy, society, and daily life will become dependent on data. It is imperative that educators, administrators, and students begin today to consider how to best prepare for and keep pace with this data-driven era of tomorrow. Undergraduate teaching, in particular, offers a critical link in offering more data science exposure to students and expanding the supply of data science talent. Data Science for Undergraduates: Opportunities and Options offers a vision for the emerging discipline of data science at the undergraduate level. This report outlines some considerations and approaches for academic institutions and others in the broader data science communities to help guide the ongoing transformation of this field.

Learn how easy it is to apply sophisticated statistical and machine learning methods to real-world problems when you build on top of the Google Cloud Platform (GCP). This hands-on guide shows developers entering the data science field how to implement an end-to-end data pipeline, using statistical and machine learning methods and tools on GCP. Through the course of the book, you'll work through a sample business decision by employing a variety of data science approaches. Follow along by implementing these statistical and machine learning solutions in your own project on GCP, and discover how this platform provides a transformative and more collaborative way of doing data science. You'll learn how to:
Automate and schedule data ingest, using an App Engine application
Create and populate a dashboard in Google Data Studio
Build a real-time analysis pipeline to carry out streaming analytics
Conduct interactive data exploration with Google BigQuery
Create a Bayesian model on a Cloud Dataproc cluster
Build a logistic regression machine-learning model with Spark
Compute time-aggregate features with a Cloud Dataflow pipeline
Create a high-performing prediction model with TensorFlow
Use your deployed model as a microservice you can access from both batch and real-time pipelines

Discover how data science can help you gain in-depth insight into your business - the easy way! Jobs in data science abound, but few people have the data science skills needed to fill these increasingly important roles. Data Science For Dummies is the perfect starting point for IT professionals and students who want a quick primer on all areas of the expansive data science space. With a focus on business cases, the book explores topics in big data, data science, and data engineering, and how these three areas are combined to produce tremendous value. If you want to pick-up the skills you need to begin a new career or initiate a new project, reading this book will help you understand what technologies, programming languages, and mathematical methods on which to focus. While this book serves as a wildly fantastic guide through the broad, sometimes intimidating field of big data and data science, it is not an instruction manual for hands-on implementation. Here's what to expect:
Provides a background in big data and data engineering before moving on to data science and how it's applied to generate value
Includes coverage of big data frameworks like Hadoop, MapReduce, Spark, MPP platforms, and NoSQL
Explains machine learning and many of its algorithms as well as artificial intelligence and the evolution of the Internet of Things
Details data visualization techniques that can be used to showcase, summarize, and communicate the data insights you generate
It's a big, big data world out there—let Data Science For Dummies help you harness its power and gain a competitive edge for your organization.

This edited collection brings together voices of the strongest thought leaders for diversity, equity and inclusion in the field of statistics and data science, with the goal of encouraging and steering the profession into the regular practice of inclusive and humanistic leadership. It provides both futuristic ideas for promoting opportunities for equitable leadership, as well as tested approaches that have already been found to make a difference. It speaks to the challenges and opportunities of leading successful research collaborations and making strong connections within research teams. Curated with a vision that leadership takes a myriad of forms, and that diversity has many dimensions, this volume examines the nuances of leadership within a workplace environment and promotes storytelling and other competencies as critical elements of effective leadership. It makes the case for inclusive and humanistic leadership in statistics and data science, where there often remains a dearth of women and members of certain racial communities among the employees. Titled and non-titled leaders will benefit from the planning, evaluation, and structural tools offered within to contribute inclusive excellence in workplace climate, environment, and culture.

Explains how Billy Beene, the general manager of the Oakland Athletics, is using a new kind of thinking to build a successful and winning baseball team without spending

enormous sums of money.

With this practical book, AI and machine learning practitioners will learn how to successfully build and deploy data science projects on Amazon Web Services. The Amazon AI and machine learning stack unifies data science, data engineering, and application development to help level up your skills. This guide shows you how to build and run pipelines in the cloud, then integrate the results into applications in minutes instead of days. Throughout the book, authors Chris Fregly and Antje Barth demonstrate how to reduce cost and improve performance. Apply the Amazon AI and ML stack to real-world use cases for natural language processing, computer vision, fraud detection, conversational devices, and more Use automated machine learning to implement a specific subset of use cases with SageMaker Autopilot Dive deep into the complete model development lifecycle for a BERT-based NLP use case including data ingestion, analysis, model training, and deployment Tie everything together into a repeatable machine learning operations pipeline Explore real-time ML, anomaly detection, and streaming analytics on data streams with Amazon Kinesis and Managed Streaming for Apache Kafka Learn security best practices for data science projects and workflows including identity and access management, authentication, authorization, and more

This book provides readers with a thorough understanding of various research areas within the field of data science. The book introduces readers to various techniques for data acquisition, extraction, and cleaning, data summarizing and modeling, data analysis and communication techniques, data science tools, deep learning, and various data science applications. Researchers can extract and conclude various future ideas and topics that could result in potential publications or thesis. Furthermore, this book contributes to Data Scientists' preparation and to enhancing their knowledge of the field. The book provides a rich collection of manuscripts in highly regarded data science topics, edited by professors with long experience in the field of data science. Introduces various techniques, methods, and algorithms adopted by Data Science experts Provides a detailed explanation of data science perceptions, reinforced by practical examples Presents a road map of future trends suitable for innovative data science research and practice Mining big data requires a deep investment in people and time. How can you be sure you're building the right models? With this hands-on book, you'll learn a flexible toolset and methodology for building effective analytics applications with Hadoop. Using lightweight tools such as Python, Apache Pig, and the D3.js library, your team will create an agile environment for exploring data, starting with an example application to mine your own email inboxes. You'll learn an iterative approach that enables you to quickly change the kind of analysis you're doing, depending on what the data is telling you. All example code in this book is available as working Heroku apps. Create analytics applications by using the agile big data development methodology Build value from your data in a series of agile sprints, using the data-value stack Gain insight by using several data structures to extract multiple features from a single dataset Visualize data with charts, and expose different aspects through interactive reports Use historical data to predict the future, and translate predictions into action Get feedback from users after each sprint to keep your project on track

Doing data science is difficult. Projects are typically very dynamic with requirements that change as data understanding grows. The data itself arrives piecemeal, is added to, replaced, contains undiscovered flaws and comes from a variety of sources. Teams also have mixed skill sets and tooling is often limited. Despite these disruptions, a data science team must get off the ground fast and begin demonstrating value with traceable, tested work products. This is when you need Guerrilla Analytics. In this book, you will learn about: The Guerrilla Analytics Principles: simple rules of thumb for maintaining data provenance across the entire analytics life cycle from data extraction, through analysis to reporting. Reproducible, traceable analytics: how to design and implement work products that are reproducible, testable and stand up to external scrutiny. Practice tips and war stories: 90 practice tips and 16 war stories based on real-world project challenges encountered in consulting, pre-sales and research. Preparing for battle: how to set up your team's analytics environment in terms of tooling, skill sets, workflows and conventions. Data gymnastics: over a dozen analytics patterns that your team will encounter again and again in projects The Guerrilla Analytics Principles: simple rules of thumb for maintaining data provenance across the entire analytics life cycle from data extraction, through analysis to reporting Reproducible, traceable analytics: how to design and implement work products that are reproducible, testable and stand up to external scrutiny Practice tips and war stories: 90 practice tips and 16 war stories based on real-world project challenges encountered in consulting, pre-sales and research Preparing for battle: how to set up your team's analytics environment in terms of tooling, skill sets, workflows and conventions Data gymnastics: over a dozen analytics patterns that your team will encounter again and again in projects

[Copyright: 15746431d2594b4f86e3a4f2abfada3b](https://www.amazon.com/dp/B078888888)