

---

## Book Review

---

Reviewed by Zhongxian Wang

Email: wangz@mail.montclair.edu

**The Decision Maker's Handbook to Data Science: A Guide for  
Non-Technical Executives, Managers, and Founders  
by: Styllanos Kampakis  
Published 2019  
by PublishDrive, Budapest, Gogol u. 13, 1131 Hungary, 141pp  
ISBN: 6610000098170**

Chapter 1 defines data science and its different aspects. In 2017 alone, we generated more data than we did over the previous 5000 years of our history. All data has use, we just need to find a better way of doing so. In short, data science is all about using data to do something useful. There are three core fields of data science: artificial intelligence, machine learning, and statistics. The author analyses the different subfields of data science. These subfields include cybermetrics, artificial neural networks, computational intelligence, machine learning, and data mining and knowledge discovery in databases, etc.

Chapter 2 displays data management and all of the disciplines involved in working with data. The author defines data management as the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of an enterprise. Understanding where data comes from is very important. Companies collect data from internal sources like transaction, log data, and emails. Also, they collect data from external sources like social media, audio sources, as well as photos and ideas. The different sources of data can be classified into data collection methods. These methods include observational and experimental. When acquiring data, you must take into consideration the appropriateness of the data, nature of the domain, time requirements, and the cost. Qualitative and quantitative researches are very important when figuring out how the data should be collected but also the types of data and how it is analysed.

The author explains how data collection has its share of problems that can potentially lead to issues in the future in Chapter 3. One issue is data not being recorded correctly. Another issue is the data collection is not supervised by the appropriate principal investigator. Another problem can be when data is kept only by a specific person rather than being in a centralised system. The chapter also provides six examples of data collection, including B2C apps, sales, retail, finance, sports, and social media.

Chapter 4 deals with both good and bad data management practices. Some good practices include establishing a goal first, having awareness of how data collection affects the rest of your business, and establishing a data standard. Some bad practices include not having a clear objective, ignoring the connection between data collection and rest of your business and having no documentation or data standard.

Chapter 5 looks at some examples of the impact of data science and collection on your business. It can really affect your user interface and can lead to completely different business models. For example, sites like Tinder and OkCupid. Tinder does not take in a lot of data so the recommendations for potential dates are not as accurate as OkCupid, which asks over 100 questions during the sign up process in order to get accurate recommendations. Also, Tinder is significantly easier to understand than OkCupid is. This can affect the entire user interface, the user experience, and the whole business model itself.

Chapter 6 presents the data science process. It shows how to think like a data scientist without being one. The author lists the definition of the data science from a pragmatic point of view and puts five points. First, you collect data. Second, you organise the data. Third, you analyse the data. Fourth, you interpret the data. And lastly, you communicate the finding from the data. The actual data science process has 4 steps. Step 1, problem definition. Step 2, choosing the right data. Step 3, Solution of the problem. And step 4, creating value through actionable insights. The process involves two main actors, them being the domain expert, and the data scientist.

Chapter 7 is a short introduction to statistics. Statistics has two branches to it, descriptive statistics and inferential statistics. Descriptive statistics is what most people associate with the work statistics. They include things like collecting data, using summary metrics such as the mean, and visualising data. Inferential statistics is what most statisticians think of when they think of the word statistics. It involves more advanced concepts such as sampling and inferring the values of the parameters of the population. The author also goes in depth about how to properly use statistics for data science purposes as well as how statistics can sometimes be misleading.

Chapter 8 introduces machine learning. It gave some definitions from different researchers. One of the main advantages of machine learning is implementation and automation. Once you have the algorithm necessary for whatever you're doing, it is easy for a machine to go through thousands or even millions of images without experiencing any fatigue, which is impossible for a human brain to do. It also covers the different types of machine learning, those being supervised and unsupervised learning.

Chapter 9 focuses on problem solving. Understanding whether a problem has the potential to be solved is very important to begin with. You then need to see if the problem can be phrased as: a statistical modelling problem, a hypothesis test, a supervised learning problem, or an unsupervised learning problem. If the heuristics fail however, it can be because of lack of the right data, the quality of the data, the data volume in terms of number of variables, and data size.

Chapter 10 demonstrates hiring and managing data scientists. It discovers how a data scientist thinks. This is basically referring to all the different things they need to be knowledgeable in to be a successful data scientist. One thing they need to have is hacking skills. They also need to be advance in mathematical and statistical skills. Domain knowledge is also very important for them to have. The different motivations that data scientists have are also explained in this chapter as well as what makes them disengage from their work.

Chapter 11 guides when a data scientist is looking for a job. First of all, what do they want? Well, they want compensation of course. However that is not enough. They need a good team to work on, they want to know the specific problems they will be handling, the technology that they will be using, and the relationship to academia. It details the hiring process of finding data scientists and how they like to avoid traditional limitations.

One of the main issues when deciding where to work as a data scientist is choosing to work for a startup or a larger company. But both have their pros and cons to consider.

Chapter 12 classifies the different types of data scientists. The author describes them as tribes. In terms of the major tribes, there are computer scientists, statisticians, and the quantitative specialist. The smaller tribes consist of the self-taught data scientist, the software platform user, and the domain specialist. This leads into Chapters 13 and 14 where the author concentrates building a data-science culture. First, there must be an understanding of what a data-science culture is about. They use a great example of establishing a culture with the movie *Moneyball*, and how the entire focus of the team was to build a data centric culture. There are three levels of a data-science culture which are the management level, the employee level, and the organisational level. You then need to create a friendly environment for data scientists and make sure that it is data centric on an organisational level. Culture is extremely important in any workplace because it helps establish the identity of the organisation.

In conclusion, this book gives great insight on all the different aspects of data science and all the different nuances of it that make it tick. It is a very valuable field and reading this book made me realise just how important it is. It is very informative and if data science is something that sparks your interest, I strongly recommend reading this book.