



Data, Data and Even More Data: Harvesting Insights From the Data Jungle

Bernd Skiera

■■■■■

KEYWORDS

*Data Science, Big Data,
Analytics, Growth Hacking,
Online Advertising*

•

THE AUTHOR

Bernd Skiera,
Chair of Electronic Commerce,
Faculty of Business and Economics,
Goethe-University,
Frankfurt am Main, Germany
skiera@wiwi.uni-frankfurt.de.

■■■■■

From data wasteland to data jungle /// Increasing global digitalization brings huge and ever-growing amounts of data. It all began with the invention of the browser that made access to the Internet via a desktop computer so much easier and faster. More and more consumers started to enjoy direct online interaction with each other and with companies... and started to leave their traces. The cost of observing these interactions fell to marginal costs that were very close to zero. It was, for example, possible for the first time to observe on a large scale not only that an advertisement was shown to a consumer, but also how the user reacted to that ad. So marketers were able to measure whether the consumer clicked and even purchased after clicking on the ad. Previously, a comparable measurement of advertising success was only possible for direct marketing activities but the cost of doing so was much higher and the quality of measurement much lower. For example, direct marketers could not even observe whether the consumer opened the letter they sent. Just compare this opportunity to the ones that email marketers have today.

The next major step forward came with the availability of affordable and powerful smartphones and mobile data plans. They now enable companies to target consumers everywhere, add location-based information to consumers' actions, and record consumers' reactions at the location and the time where the reaction occurs. Thus, instead of interacting with

FIGURE 1:
A flashlight on the data science jungle



consumers during the few hours per day that they use their desktop, companies can nowadays interact with consumers essentially 24/7. The availability of data exploded and Hal Varian, chief economist at Google and previously a well-known researcher in microeconomics, became famous for saying around 2005 that “the sexiest job in the next 10 years will be statisticians.” So, instead of a data wasteland we seem to be living in a data jungle full of ripe fruit. Can marketers simply pick it up now? Is all of it wholesome? Or is harvesting insights from a data jungle a more challenging task than anticipated and one that requires new skills?

What companies can gain from big data analysis

> **Insights from academic research** /// Many companies are convinced that the fruit of the data jungle is wholesome. Insights that arise from big data analyses are in high demand. In contrast to ten years ago, the number of company jobs for PhDs in marketing and economics is growing

and growing. Well executed academic studies attract huge interest among companies. Managers are, for instance, willing to use the insight of a study by the researchers Blake, Nosko and Tardelis in 2015 on the unprofitability of Google AdWords for branded keywords to reallocate millions of dollars of advertising budget to other advertising media. The Wharton Customer Analytics Initiative (WCAI) successfully draws the interest of Fortune 500 companies to sponsor competitions that attract academics to analyze the data of those companies and share the respective insights. The list of marketing problems that are analyzed on the analytics platform Kaggle is constantly increasing, and online classes on “machine learning” are among the most popular online courses.

> **Improving marketing decisions** /// Marketing can be much more effective if more and better information is available. With this special issue, we intend to help com-

{Box 1}

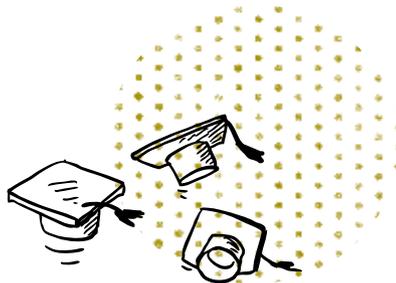
STRIKING A BLOW FOR MATH IN MARKETING

Loosely speaking, the biggest challenge for data science applications in marketing is that many people in marketing decided to work in marketing or study marketing because they no longer wanted to do mathematics. A good marketing campaign is doubtless creative, touches emotions, builds a brand and encourages consumers to talk about the product. Math alone does not allow for doing so but my prediction is that managers without good technical skills will not even get a chance to be responsible for marketing campaigns. To illustrate my point, let me share a very recent experience with you:

SENSITIZING STUDENTS FOR A DATA-LADEN MARKETING WORLD

A few weeks ago, I had the chance to run an undergraduate seminar together with a superb Business Development Manager at Amadeus, Sandro Cuzzolin, who is deeply involved in its online travel company Travelaudiance. He put together four data sets for this five day seminar. They ranged in size from 770,000 to more than 9,000,000 observations; sizes too large for popular spreadsheet programs. Our marketing students had to analyze the data sets to answer simple questions such as identifying the regions where the company makes largest profit and should spend more on advertising. Not an easy task, indeed:

- » The number one problem that most students had was to use a database engine such as SQLite to condense the data in such a way that they could further analyze it in a spreadsheet environment to calculate a return on investment.
- » Students stopped complaining about the size of the data sets after Sandro Cuzzolin reminded them that these data sets did not even cover a week.
- » He further asked the students to imagine how long such an analysis would probably take in reality if they always had to ask the IT department to properly select and condense the data.
- » Needless to say, the students' analysis did not come close to separate causality from correlation.

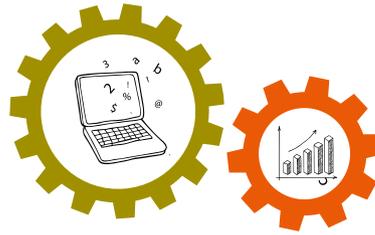


You might think that my student requirements and my predictions are exaggerated? Maybe, but keep in mind that the largest provider of advertising inventory in most Western countries, Google, started as a high tech company. Its competitive advantage is also due to using superb technology and analytics to provide free-of-charge insights into the success of clients' advertising campaigns. In addition, other companies such as Facebook or Instagram do not have their roots in marketing.

>>

The two fields, marketing and information systems, are moving closer together.

<<



panies participate in the gains in efficiencies. I am glad that some of the most prestigious researchers in marketing took their time to contribute and highlight how marketing decisions can be improved in our data-intensive environment. Pradeep Chintagunta, Mike Hanssens and John Hauser illustrate how data can be transformed into useful information for various marketing applications (pp. 18) Recently, they served as co-editors of a Special Issue on big data of the flagship journal *Marketing Science*, and they share some highlights with us. In their article they also predict a much stronger collaboration of marketers with data scientists and computer scientists and stress that marketers can not only benefit but also contribute by modeling structure and exploring causal relationships. Martin Spann and his coauthors show how to get location data, how to analyze it and how to use it to make customized marketing decisions (pp. 30). Including the situational variables that smartphone applications generate helps companies to design their offers so that consumers perceive them to be much more relevant. Martin Spann has a strong background in marketing and now conducts much of his research into information systems. Other researchers came the opposite way. These developments in research interests are fine examples of how the two fields, marketing and information systems, are moving closer together. Martin Schmidberger looks at the data jungle from a managerial perspective. For many years he has been the head of data and marketing at ING DiBa. Remarkably, this is the only bank whose market share in the German retail banking market has grown substantially over the last ten years. In my interview with him (pp. 50), he confirms that systematic use of data and machine learning techniques leads to a better understanding of customer behavior. Their systems generate customized recommendations with a better response than traditional marketing achieved and thus highlight again the fact that data science and marketing are a winning team.

Marketing challenges for harvesting insights from the data jungle /// However, fruit-picking in a data jungle is not automatically rewarding. It can be very tricky, sometimes even painful to harvest true insights. Mathe-

matical skills are a precondition (see Box 1), but marketers need to react to diverse challenges and our authors deal with other critical issues as well.

> **Drawing correct conclusions** /// Two of the strongest women in our field, Anja Lambrecht and Catherine Tucker, highlight the important difference between correlation and causality (pp. 24). On the one hand, this topic is an old one and every student hopefully learns that a regression analysis indicates correlation but not necessarily causation. On the other hand, the ability in a digital world to target individual consumers makes this problem so much more important. In Box 2 you find an example that illustrates how tempting it is to draw wrong conclusions. Hopefully you will see that only experiments like the ones Anja Lambrecht and Catherine Tucker describe are able to observe the causal effect of marketing actions and that only those causal effects should guide marketing allocation decisions.

> **Data quality and data pricing** /// Marc Grether, COO of Xaxis, focuses on the potential of big data use in the online advertising industry (pp. 38). The challenge here is to reduce wastage and target exactly those consumers that form a specific target group. While AdWords builds on search phrases and consumers' self-selection, successful display advertising relies on data about consumers (see Box 3). It is still quite difficult to provide such data in good quality and to find a reasonable price for the data, as Marc Grether outlines nicely. The data industry has yet to develop business models to successfully handle these challenges.

Data quality, or – to be more precise – missing data is also the biggest challenge for using big data in market research. Volker Bosch of GfK notes that it is fairly challenging to technically and methodologically handle “big data” (pp. 56). But the future of big data in marketing research is bright. One major reason is that new technologies allow for passively measuring consumers' attitudes and preferences. He also highlights the opportunities that data imputation provides for solving the problem of limited depth of data and calls for closer collaboration between data science and marketing science.

{Box 2}

THE REAL RETURN ON MARKETING

In Google AdWords or other forms of targeted advertising, you typically present advertising messages to consumers that have already expressed interest in your offering. In Google AdWords, for example, consumers search for a product that you offer. Let us assume that this group of consumers has a purchase probability of 90%. Does 90% represent the impact of your advertising? Most likely not. The reason is that you don't know whether the consumers' purchases were triggered by your ad or if they would have occurred anyway.

Comparing the purchase probability of 90% to the response rate of a group of consumers who have not seen the advertisement won't help neither. Even if such a group displays a response rate of only 5%, you can't attribute the difference of 85% to advertising alone. The reason is that you targeted consumers who are simply much more likely to purchase. Thus these consumers would have more likely purchased even if you did not send them any advertising message. The challenge is that you need to disentangle the observed difference (85%) into the causal effect of advertising and the systematic difference between the two groups of consumers: the ones who already expressed interest in your offering and received an ad and the others who did not express interest before and therefore did not receive an ad. The latter, however, can only be observed if you run an experiment in which you stop advertising for randomly selected consumers from the group that expressed an interest in your offer.

Let us assume that the experiment shows that you still observe an 80% purchase probability for consumers who expressed interest in your offering (e.g., searched on Google) but who you did not target with an ad. Then you can conclude that 80% would have bought anyway and the ad increases the purchase probability to 90%. Thus the causal effect of the ad is 10 percentage points. Consequently, in return of investment calculations, you should compare the cost of the ad to these 10 percentage points.



> **New skills** /// One of the key challenges will also be to hire specialists and to form teams that find their way around in the data jungle. Kay Lemon elaborates on the seven skills necessary to survive and thrive in the new market conditions (pp. 44). She currently serves as director of the Marketing Science Institute (MSI) and is therefore one of the key players at the interface between academic marketing theory and business practice. Marketing teams need diverse talents that will work closely together to deliver great experiences at the moments that matter. As a result, marketing stars of tomorrow will be the ones who understand the customer's decision journey and the critical moments in that journey – in real time, and in context.

The future of the data jungle: perish – or thrive?

> **Even more to gain** /// Will we observe even more data in the future? It is very hard to believe that this will not be the case. Devices such as watches, glasses, cameras, technologies like face recognition, thermal imaging, WiFi tracking or beacon communities like WhatsApp, WeChat, and Snapchat will generate even more data. Almost every area will be impacted but, given that consumers produce a huge amount of that data, marketing is the area in business that will experience the strongest impact from the availability of data. So the jungle will be even denser with even more to gain for those who have the necessary skills to harvest data and transform it into real insights. We now know so

{ Box 3 }

A SPOTLIGHT ON ONLINE ADVERTISING

The online advertising industry was essentially turned upside down when Google began in 2002 to use auctions to sell its AdWords. Remarkably, Yahoo! owned a patent for doing so and granted Google a license to use this patent. What happened is that search engine marketing became the most prominent online marketing instrument in most countries, with Google essentially having a monopoly in many Western countries. This change came along with a disruption of how online advertising was sold. Historically, most online advertisements were sold by sales representatives who usually negotiated a long-term contract with a uniform price for all ads. In Google AdWords, Bing and Yahoo! and the prominent search engines in China and Russia, Baidu and Yandex, each ad is sold in a real-time auction. We have an individual price for each ad that is sold on the spot. I am not aware of another industry

whose pricing mechanism changed so quickly. Currently we are observing a similar change for display advertisements, which are more and more often sold via real-time auctions, also known as real-time bidding. The remarkable difference between display advertisement and search engine marketing is the information that is available about a consumer. In search engine marketing, the search phrase essentially contains all the information. In display advertisement, however, data is collected about consumers' interests and preferences, frequently via third parties and data aggregators. In that sense, the kind of data to select customers is much closer to the kind of data that is used in TV advertising, with the important difference that we now have data about each individual consumer.

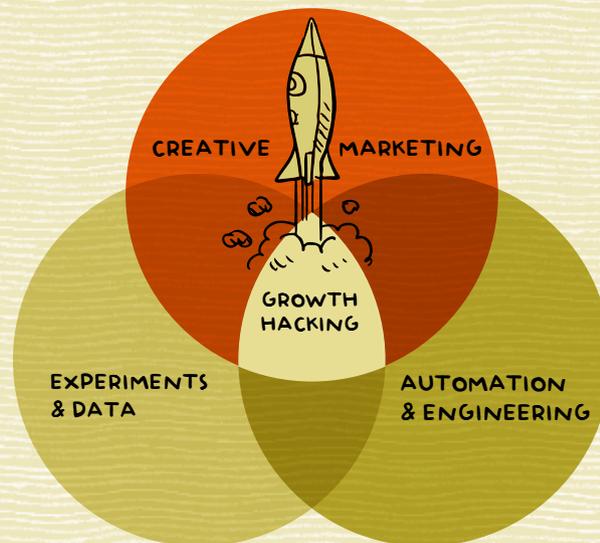
much better what consumers do, where they are, what they think, or how they react to the companies' messages. No one predicted 25 years ago how much information we would have available today and what opportunities this data provides.

> **Growth opportunities from technical expertise** /// Consumer insights have always been considered a major driver for growth, but in the digital world successful growth can come from new angles. "Growth hacking" is a term not too popular in marketing but is likely to be at the core of how marketing will develop in the future. Andrew Chen, an influential Silicon Valley-based manager and blogger, describes a growth hacker in a 2012 blog post as a hybrid of marketer and coder who looks at the traditional question of how to acquire customers and answer them with A/B tests, landing pages, viral factors, email deliverability, and Open Graph. He uses Airbnb and its remarkable integration into the classified advertisement website Craigslist as an example to demonstrate how technical expertise allows to better spread an offering. He concludes

by stating: "Let's be honest, a traditional marketer would not even be close to imagining the integration – there's too many technical details needed for it to happen". Even if there is an abundance of ripe fruit, the harvesting and processing equipment needs to be sophisticated and tuned for purpose to insure a successful harvest.

> **Prepare the field for talent to grow** /// So technical and methodological skills will be the key to success in a data-laden marketing environment. I am convinced that tomorrow's marketing curricula will include much more data science related topics. Students will learn how to collect data via crawlers or APIs and how to handle such data with sophisticated database software. They will be able to handle unstructured data with text-mining approaches and to analyze data with econometric and machine learning techniques. Students will know how to separate correlation and causation via properly conducting experiments and will understand techniques to visualize relations between many objects and how to automatically react to customers in real time.

FIGURE 2:
Growth Hacking: Skillful technical implementation
of marketing ideas to generate growth



With such experts, companies will be able to distinguish between wholesome and perished data fruit and harvest real insights that improve decisions and enable growth. But despite all my enthusiasm, let me conclude by stressing that one thing will never change. Companies, and in particular marketing, need to provide value for the customers. So data science in marketing is just a means to an end. Probably, however, a very powerful one.

1.

Blake, Thomas; Chris Nosko; Steven Tadelis (2015):
“Consumer Heterogeneity and Paid Search Effectiveness:
A Large-Scale Field Experiment,”
Econometrica, Vol. 83 (1), pp. 155 – 174.

Chen, Andrew (2012):
“Growth Hacker is the new VP Marketing,”
<http://andrewchen.co/how-to-be-a-growth-hacker-an-airbnb-craigslist-case-study/>.