INTRODUCTION

It is impossible to ignore the fact that we live in a data-driven civilization. Not only is the amount of data in the world doubling every two years, but the percentage of these data that are becoming valuable because of advanced analytics is also growing. The entire field of data capture and analysis is evolving so rapidly that organizations have difficulty keeping up. Yet data-driven business processes, competition, and the rewards of faster and more intelligent operations leave us with no other choice.

For a long time, our ability to capture data outpaced our ability to process it. This meant that large quantities of data were stored in data warehouses until some future time when tools would be available to find value in them or until they were discarded all together. Several things have happened in recent years to change this dynamic. One is the exponential growth in data; the other is the emergence of new platforms and technologies that make it possible to process data sets of almost unlimited size economically while lowering the cost and increasing the speed of analysis. These elements, combined with new analytic techniques and a growing use of machine learning to accelerate analytic methods, is changing almost every aspect of our lives.

To gain a fuller understanding of how modern analytical methods are being used in visible and not-so-visible ways, we approached data analytics experts from many fields and industries. I asked them to contribute essays about their experiences applying big data analytics. This e-book is a compilation of those essays. In it you will find discussions about new analytics technologies, how organizations can more effectively use their data assets, and many interesting use cases. The essays have been grouped into five sections:

- **Business Change.** Essays in this section speak to how advanced analytics are changing the way businesses operate. It is much more than a story about increased productivity and efficiency: it is a story about the complete transformation of traditional business models into something new and totally data driven.
- **Technology Platforms.** Essays in this section take a closer look at some of the tools and platforms that are making advanced analytics economical for organizations of all sizes.
• **Industry Examples.** This section continues the discussion of transformative analytics technologies in the context of specific business and public-sector use cases.

• **Research.** This section focuses on how new-age analytics are changing the way scientists are conducting research and how they are speeding knowledge acquisition.

• **Marketing.** This section focuses on advanced, analytics-driven marketing strategies and techniques. These techniques are being used for everything from brand marketing to personalization to public relations to attribution techniques that enable companies to analyze their most effective marketing activities in real time.

It is my hope that assembling knowledgeable insights and experiences from so many different perspectives will provide a valuable glimpse into this rapidly evolving technology. I have found many of these essays both eye-opening and thought provoking. There is no question that advanced analytics will continue to play an increasingly important role in business, government, health care, knowledge acquisition, and a broad spectrum of human endeavor.

All the best,
David Rogelberg
Publisher

Marketing

Shonali Burke  
Shonali Burke Consulting  

Francois Garillot  
Swisscom

Allison Lloyd  
DOCUMENT Strategy Media 
& Forum

Kirk Borne  
Booz Allen Hamilton

Chris Conrey  
Expanded.io

Andrew C. Sanderson  
Pawn Global Venture Capital 
Consultants

Russ Merz  
Eastern Michigan 
University

Greg Bonsib  
Zenith Products Corp
Application of big data analytics is beginning to disrupt the public relations (PR) industry. New analytics tools are an important part of this disruption, but it is also largely because of easy access to all those data. In PR, we focus on several categories of metrics. Output metrics measure messaging goals such as hits or impressions, but there are limits to what those metrics tell you. Outtake metrics measure what people are actually taking away from your messaging, and outcome metrics measure the ultimate outcome, such as a change in behavior. These data can tell so many stories through trends and patterns, but if you are not looking for them, you will not see them.

A simple example illustrates how data can change a PR strategy. I worked with a company that was rolling out a new online offering; its goal was to get people to sign up. The strategy consisted largely of a traditional outreach campaign with online ads, tracking URLs, and a limited reliance on social information.

“New analytics tools are an important part of this disruption, but it is also largely because of easy access to all those data.”

SHONALI BURKE
President and CEO, Shonali Burke Consulting

Shonali Burke was named to PRWeek’s inaugural top “40 Under 40” list of United States–based PR professionals, is one of 25 women who rock social media, and is the 2015 recipient of AWC-DC’s Matrix Award. As president and CEO of Shonali Burke Consulting, she uses measurable social PR to take business communications from corporate codswallop to community cool. Shonali teaches at The Johns Hopkins and Rutgers Universities, is founder and publisher of the popular PR community blog Waxing UnLyrical, and creator and curator of the #measurePR hashtag and Twitter chat.

KEY LESSONS

1. FEW PUBLIC RELATIONS AGENCIES THINK OF USING DATA ANALYTICS TO REFINE THEIR STRATEGY.

2. DATA SHOW WHAT IS WORKING AND WHAT IS NOT, AND IT PROVIDES CLEAR INSIGHTS INTO WHAT STRATEGIES YOU SHOULD USE.
Data showed us early on that the traditional outreach was not accomplishing its goal of moving traffic to the website. Realizing that there was no online discussion related to our customer, we decided to try Facebook and Twitter ads along with social engagement. We worked with our customer to start an online discussion of the topic. Data showed that practically all the traffic coming to the new website was coming from Twitter and the online initiatives. We decided to cut back on traditional outreach and focus almost entirely on online events that got people talking. The analytics showed that that is what would work—and it did, much to our client’s delight.

Many PR agencies do not think of using data analytics to refine their strategy. Some big creative agencies are building data capabilities, but many large PR firms are not, partly due because big agencies find it difficult to invest the time to build analytics skills and use analytics in their process. When you work in an agency, you are a slave to your utilization and availability, and the larger the agency gets, the harder it must work to maximize utilization. These large PR organizations rely more heavily on where their bread-and-butter projects come from, which is traditional media, and that makes it more difficult to develop an analytics capability.

The industry is changing, however, as companies spend more on digital strategies and recognize how important analytical measurements are to seeing their effectiveness. Data show what is working and what is not, and they provide clear insights into what strategies you should use. Businesses must map that knowledge back to business objectives and the story they need their audience to react to amidst everything else happening around them.

“Data can tell many stories—if you know how to listen.”
There are two primary ways to advertise on the Internet. One is to analyze browsing behavior, often with other personal information, and based on that analysis pop up a product-specific advertisement in real time that is appropriate for that person. This technique, performance advertising, has become the most common form on the Internet.

Another approach to Internet advertising is brand advertising. In that process, the goal is to reach every possible person with a brand ad that tells a story—something difficult to do on the Internet because there are so many outlets where to reach someone. Attempting to saturate the Internet with an ad would be prohibitively expensive. However, new analytics models are making Internet brand advertising across large populations both possible and effective. Here is an example.

In 2014, I worked with a team tasked with creating a brand advertising strategy that would work, say, for a coffee machine manufacturer. The Internet advertising company wanted to offer this branding strategy to customers in the United Kingdom. But because there are far more tea drinkers than coffee drinkers in Britain, it would be a waste of money to present coffee brand advertising to everyone. So, the challenge was how to identify coffee drinkers in a predominately tea-drinking population, and then figure out how to reach them with coffee machine brand advertising.

“"The challenge was how to identify coffee drinkers in a predominately tea-drinking population.""
We devised an unsupervised clustering model that would analyze the whole UK population to define groups with shared browsing patterns. The model is unsupervised in that we use an algorithm to train it, with which it analyzes browsing data in a way that is completely guided by the data, with no input from our particular desires or biases. In this way, the model quickly found many population groupings, including some we did not expect to find. For instance, one group that popped up quickly consisted of singles. We had had no idea in advance that we would find this large cohort appear in mobile browsing data.

Another practical effect of this kind of modeling is the identification of more groups than we know what to do with. It is up to the company to decide whether a group we discovered is a suitable target for its brand advertising, but we can select certain groups based on criteria such as group size, location, or combinations of characteristics members of the group shared. It even becomes possible to segment groups. For example, in defining coffee drinkers, it’s possible, based on data, to witness a group that is interested in coffee; but within that group, we could see sub-segments pop up within people who primarily like to drink coffee, such as those who are extreme coffee connoisseurs. The vanilla coffee drinkers will be more interested in elegant, easy-to-use machines that might use coffee packets. The coffee connoisseurs are more interested in the esthetics of coffee making and are likely to be interested in high-end espresso machines. The interesting part is that finding the segments as revealed in the data results in a more effective brand advertising strategy than hoping for those segments to appear.

It took us just five weeks to develop a solution scalable to a population of millions of people. We did it by using Scala as our development language combined with open source libraries such as Apache Spark and an interactive shell. This foundation enabled us to develop and test routines quickly for analyzing and comparing subsets of data, writing only a few lines of code at a time. The entire development process was fast, and the result was a solution that changed brand advertising from advertising to everyone to advertising only to relevant — though previously unidentified — groups.

"The result was a solution that changed brand advertising from advertising to everyone to advertising only to relevant — though previously unidentified — groups."

""
A leading Fortune 100 company I work with has built a highly engaged online community comprising both current and prospective customers, who provide feedback on the company’s marketing collateral. The company does a lot of data collection and segmentation and a ton of testing—both qualitative and quantitative—through its online channel, which gives it focused marketing intelligence about both its offline and online marketing touchpoints. Data from the company’s offline touchpoints, however, such as newspaper and TV ads, call centers, and direct marketing, are more difficult to quantify and use because they are not as trackable as online data.

This issue is largely a problem of big data. Executives tell me all the time that they are data rich and information poor. They collect a lot of data, but that data does not always translate to actionable information. A customer’s buying pattern is complex, and the journey from awareness to conversion can be the result of a marketing effort’s many channels, messages, and touchpoints. Therefore, choosing the right method of attribution when using advanced analytics and data modeling can power intelligent decision making on your marketing spend. It can help you choose how to prioritize engagement channels, allocating credit percentages to those channels that lead to eventual sales.

“Executives tell me all the time that they are data rich and information poor.”
Marketers need to move away from last-click attribution to multi-dimensional attribution models that can provide better context. In other words, how can marketing touches along the buying journey eventually work together to trigger the desired action? Taking that holistic view requires a shift in mindset, a move from product sales to customer relationships.

To that end, it is essential that marketing organizations build an effective business case that clearly outlines return on investment, risk, and cost. It must build up its IT infrastructure and processes to classify and rapidly retrieve information that propels decision-making. It must make investments in data-collection strategies and analytics to create an optimal mix of channels and touchpoints and to gain real insight in the long term. To me, attribution is really about understanding the channels in which to invest. Which are the most interactive? How are those interactions propelling action?

Many companies struggle to develop effective data-collection methodologies that provide the granularity and segmentation necessary to show clearly whether messaging in a given channel is actually triggering the desired customer actions. Therefore, a good place to start is determining the data-collection methodology that best suits the organization's needs and stated goals. All data collected must be connected to actual decision-making.

“\nIt is essential that marketing organizations build an effective business case that clearly outlines return on investment, risk, and cost.\n\n”
I’m a white guy, over 50 years of age, living in a particular ZIP code with a lot of other people who answer to similar descriptions. A unified marketing campaign should reach all of us with equal effectiveness, and we should all be equally responsive. Right? Well, maybe not.

That plan breaks down as soon as you look past the first demographic layer. For instance, I’m a big college football fan. The guy next door loves indie rock concerts. A guy down the street loves football, too, but he’s an NFL guy. His wife couldn’t care less. She’s a fan of ‘The Real Housewives of Atlanta.’ Suddenly, these people aren’t so identical. That uniform marketing campaign is running into trouble.

Enter divisive clustering. You start with whole populations, and then separate groups into smaller and smaller clusters that share common characteristics. If you continue drilling all the way down to individuals, you will have erased the concept of demographics—the lifeblood of traditional marketing.

The problem? Some businesses have millions of customers. None can afford to conduct millions of personalized marketing campaigns, but they can cluster those individuals into groups of prototype “personas.” For instance, “this person likes to buy sporting goods on the weekends.”

“Your company is already capturing all the data it needs to define personas.”
Say your company wants to target 100 million American adults. How many personas does the company need to define? My estimate for that scenario is roughly 17. Splitting each dimension of an individual’s data into two distinct segments (Positive/Negative, True/False, or High/Low), we find that 2 to the 17th power equals about 100 million unique combinations in 17 dimensions. So, 17 campaigns (based on one unique choice in each of the 17 dimensions) times 2 (to distinguish both Positive and Negative segments) is all you need. That becomes doable but still highly personalized.

Your company is already capturing all the data it needs to define personas, by the way. Call it mouse clicks, impressions, click-through rate, all of the above and then some—whatever you like. You’re tracking it. Most likely, however, you’re not using it.

You can mine clicks and web histories, drilling into what individuals look at, what they’re typing into your site’s search box, which pages they view and in what sequence. All that is evidence you can leverage to develop a model of that person’s interest. Then, you can make your pitch directly, based on the persona model.

Marketing is a forensic science: it should be based on evidence, not just the opinion of the highest-paid person in the room.

Here is my advice for transforming a marketing organization into a data-driven organization:

• Use all the data you collect.
• Take advantage of open, publicly available data for context. Lots of cities, counties, and states are collecting and presenting publicly a rich mine of deep-dive demographics and census data that reach all the way down to the household level.
• Race recklessly to the end of demographics.

It’s science—but it is not rocket science. We just need to use what’s already in front of us.
First, start with analytics and start small. That gives you an action item, a place to begin without being overwhelmed. Take one piece of information, one metric, and track everything in relation to it. You can determine whether it’s wrong later, but you have to start somewhere.

When you find the one thing you know you can measure right now, track it relentlessly. When you have some data, you can then go through and figure out what the “whys” are. Look for patterns. Even if you can’t explain why the pattern is what it is just yet, you’ll get there. For example, say you always do a lot of business right around the second week of the month: What does that tell you about your customers?

“You want to get to the point where you can start figuring out attribution—which touch points are driving conversions.”
You want to get to the point where you can start understanding which touch points are driving conversions. Does your ad placement here make more sense than another option? The best way to start proper attribution is to tag your assets so that your analytics platform will do the heavy lifting. UTM tagging is a simple process, and there are plenty of tools to help you build such tags. Building correct links allows you to analyze properly and accurately identify which sources in your strategy are working and which are ineffective.

We had a client that was selling chocolates and had planned ramp-ups in known chocolate holidays—Valentine's Day, Easter, Christmas, Mother's Day. We were seeing ads driving sales at the beginning of the week. So, we split the ads up and tried to run a few a little heavier toward the front of the week. It worked: those ads created more conversions. That's the type of pattern for which you are looking.

Adopting a data-driven process requires a bit of change from the way you’re used to doing things. You have to get comfortable with your analytics platform of choice, get to know more than just the stuff you can figure out in the first hour on your own. Really dig into it. Your data will be running on these analytics; your analytics program allows you to see conversions, and then grow from there.

Then, it's a matter of creating a routine and building a process that requires you to look at the data regularly. All the analytics in the world aren't going to do you any good if you set them and forget them.

A once-a-week report won't tell you what you need to know. If that's what you’re using, then you’re missing a lot of opportunities to learn more about your needs and behaviors. So, look at a better solution for your clients, and look like a better solution for your clients.

The only way you're going become a data-driven organization is to take the time to collect the data and look at it yourself. Noticing a pattern that leads to better conversions goes a long way toward moving your organization forward.
To turn your marketing organization into a data-driven marketing organization, you must first understand the value of attribution as it relates to your business. Attribution is assigning or ascribing a value or category to something like a number or data point. Ask yourself this simple question: “What are the numbers that I’m currently tracking telling me right now, and why do I attribute it to that specific information?” If you can’t answer that question, then you’re not truly living in the data-driven marketplace. Most small and midsized businesses in the United States are using few metrics other than their own sales numbers to try to gain a better understanding of the market and industries they’re involved in. For those who are looking to gain greater insight into their customer base and engage with them to establish those metrics worth tracking, I suggest starting out by using the social media platforms out there that allow the paid user to view data points.

First, the free model: How are you using social media channels that don’t charge for publication, fan connection, or customer interaction? What are you doing with those free channels that provide metrics and analytics that you can use? Then the premium model: If you are a LinkedIn member, are you a premium member? Do you pay to have access to those statistics and the data that allow you to analyze your current marketing?

"For those who are looking to gain greater insight into their customer base . . . I suggest using the social media platforms out there that allow the paid user to view data points."
The prepaid model is pay-per-click advertising—the paid leads, if you will. If I search for your company name, what comes up at the top of the search results? Who else is there besides your company? When you know whom you want to reach and the message you want to send them, it becomes imperative that you then use those same platforms to purchase specific pay-per-click ads into the market you want.

Attribution of the data you gather and the assignment of values and categories you choose to put them in are an important issue in a data-driven environment. By assigning each data point or metric to a specific category, you are able to categorize like values or metrics into a more discernable chart. It requires analyzing the differences in performance among the various pipelines you have in place so that you can make adjustments in your strategy, shifting the budget to higher-performing channels if need be.

Sometimes, you encounter surprises that make you reevaluate your approach to attribution from an entirely different angle. I once had an experience like that where an industry influencer was concerned. My company was, as a best practice, regularly looking at metrics on our Twitter performance, such as impressions and re-tweets, to make sure our strategy was as finely tuned as possible. Then, one day at a meeting, an executive told us that he had analyzed the interactions we were having on Twitter with his competitors and other executives in the field. He perceived us as a potential ally when he looked at the data we were gathering and the people we were collaborating with, and so he proposed a strategic alliance with us.

It’s difficult to calculate the value of a single LinkedIn update or Twitter tweet until you gather the market responses that reaffirm that your marketing is attracting new revenue through found business. The tricky part of the data, viewed in aggregate, is the anonymity that comes with it. Unless you’re carefully looking at the influencers within your audience, you might not know if one of the 1,000 people who saw your tweet was the CEO of your competition. That was the one impression it took to convert that person to a potential ally or partner, and it was powerful for us.
Recently, I worked with a packaged goods company that had an active social media presence. Its goal was to figure out what effect its social media programs had in driving traffic to its website. To quantify this, we analyzed page views, time onsite, clicks, and other metrics with the aid of an analytics engine. We demonstrated positive, empirical linkages between the client’s social media activity and changes in website patterns. We even categorized social media conversation types and showed which ones most strongly influenced web activity.

Linking social media to website behavior was a necessary building block for developing algorithms to predict consumer behavior. It’s an example of applying attribution measurement methods that go well beyond classic last-click attribution.

Algorithmic attribution models are only as good as the data injected into them: algorithms need a verified foundation of accurate data. Without the right model, it’s difficult to establish advanced algorithmic attribution.

“Without good models, it’s virtually impossible to allocate resources in a way that will have any true impact on ROI.”

KEY LESSONS

1. Algorithmic attribution models are only as good as the data injected into them: algorithms need a verified foundation of accurate data.

2. Apply the four Ms—make a strategy, move on it, measure it, monetize it.

Russ Merz is an experienced research scientist, analytics consultant, and professor. He has subject matter expertise in market research methods, as well as in the development and application of customer experience analytics to marketing management problems in the areas of advertising, public relations, branding, retailing, social media, and e-commerce.

Download the full e-book: Data Disruption
To that end, my consulting team and I have developed what we call exposure models, which empirically assess the effect of various experiential influences on targeted outcomes, like customers’ intent to purchase. Often, we rely on survey data linked to web metrics, but other data might work equally well.

Clearly, sound empirical data are crucial for understanding and shaping the marketing decisions that drive attribution. Without good models, it’s virtually impossible to allocate resources in a way that will have any true impact on ROI. You have to understand what you’re trying to accomplish.

I’ve got a heuristic that I use with clients, a framework I call The Four Ms. It goes like this: make a strategy, move on it, measure it, and monetize it. It’s simple but effective.

Working through a framework like that helps you understand how your strategy and tactics might change your key performance indicators. With a framework in place, you can measure how those changes affect sales, profit, market share—whatever metrics are most important to you. The key to this approach is building an integrated measurement system that links all these elements.

In simplest terms, here’s my advice:

• Create a framework.
• Quantify the various pieces of the framework.
• Look for empirical linkages between the pieces.

"With a framework in place, you can measure how those changes affect sales, profit, market share—whatever metrics are most important to you."
Data has always played an important role in marketing. What has changed in recent years is the sheer volume of data now available from so many different sources. If we were once challenged to see the forest through the trees, we're now challenged to see the forest through the leaves.

Here at Zenith, we primarily sell our products through our network of retail partners. We do have a product website, but it is not a major source of our sales. In my role at Zenith Products Corp, I look at data strategically, with the essential premise that the customer is the ultimate focus of interest. With that in mind, I have a four-step approach to using data in customer marketing:

1. **Start small.** For example, we often use Amazon Marketing Services (AMS) ads to test the effectiveness of promotional strategies. We will focus on a product or a phrase or an idea and watch how it performs. By doing so, we can create a test with a focused objective, and we can keep the cost of our tests very low. Amazon Webstore provides an excellent data environment in which to try something small at low cost and capture meaningful performance data around it.

   "I look at data strategically, with the essential premise that the customer is the ultimate focus of interest."
2. Test and learn. By keeping the test small and the cost low, we can try many different variations. For instance, I used AMS to test category ads, product ads, brand ads, ads based on search words, ads based on competitors, even ad placement that AMS recommended. The good thing about this method is that Amazon provides data that allow us to track ad performance directly to Amazon Webstore sales. In this way, we can identify ad strategies that perform well and those that are total duds. Failures are as important as successes. The best-performing ads returned $5 in sales for every $1 spent on marketing—a 500 percent return on investment (ROI) on marketing spend.

3. Build on success. Using the proof point of a 500 percent marketing ROI, I can then build that ad concept into a larger marketing strategy that involves product packaging, in-store promotions, and many other things. Then, I start looking at other kinds of performance data. For instance, if a store is running a promotion on a product, I pay close attention to the halo effect of that product on the sales of other products.

4. Document your knowledge. It is incredibly important to document what you do. Documentation enables you to build a body of knowledge about what works and also to share that knowledge through your organization.

The power of this kind of data-driven marketing is that when you have data that prove your strategy, you can gain credibility and resources to do it again.

“
In this way we can identify ad strategies that perform well and those that are total duds. Failures are as important as successes.”
Mighty Guides make you stronger.

These authoritative and diverse guides provide a full view of a topic. They help you explore, compare, and contrast a variety of viewpoints so that you can determine what will work best for you. Reading a Mighty Guide is kind of like having your own team of experts. Each heartfelt and sincere piece of advice in this guide sits right next to the contributor's name, biography, and links so that you can learn more about their work. This background information gives you the proper context for each expert’s independent perspective.

Credible advice from top experts helps you make strong decisions. Strong decisions make you mighty.