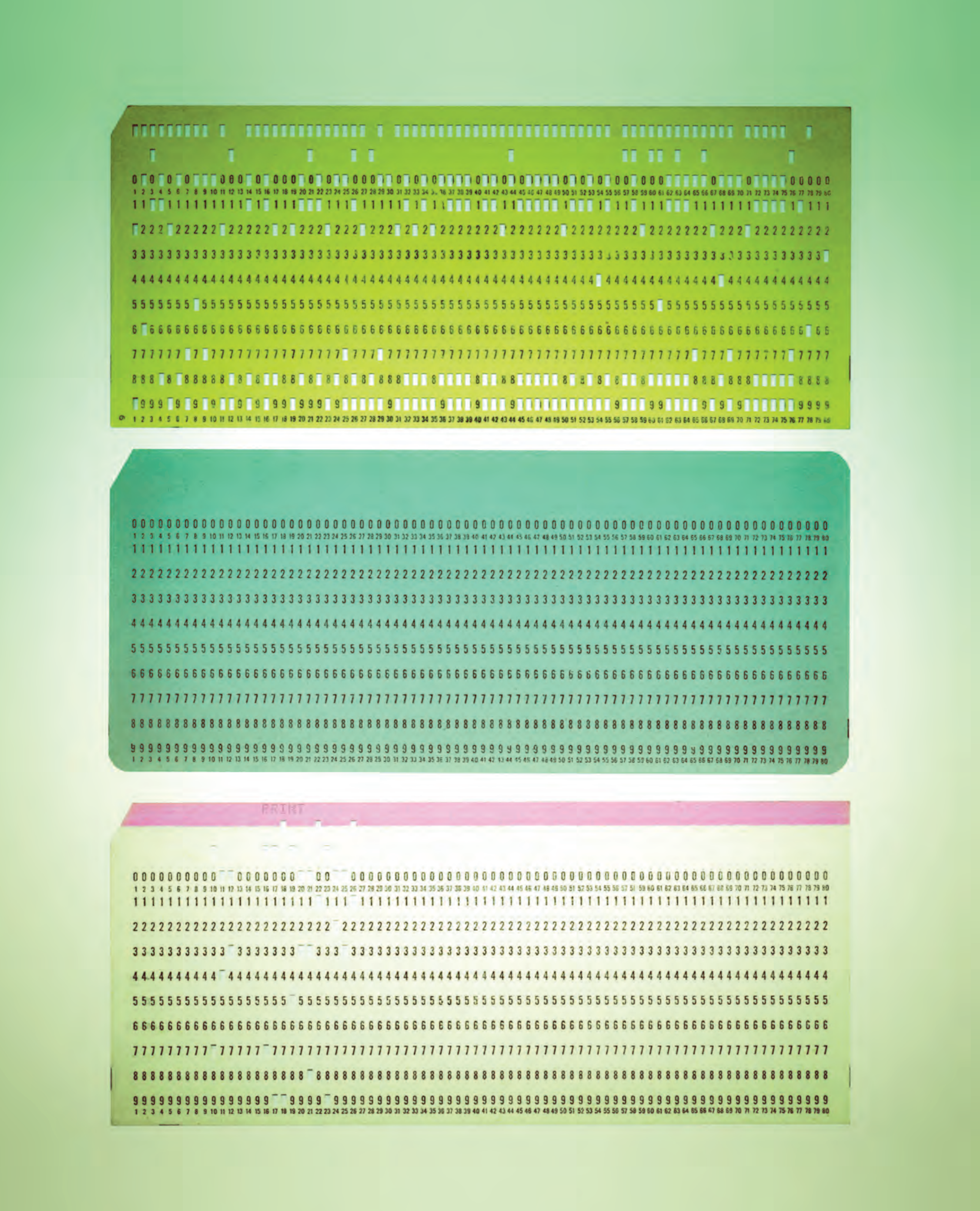

PRINT

THE PATH TO PRESCRIPTION:

Closing the Gap Between the Promise and the Reality of Big Data

To extract the full value of their data investments, organizations need to build tighter links between data analytics and everyday business conversations.

By Bernardo Blum, Avi Goldfarb and Mara Lederman

DESPITE ALL THE HOOPLA AROUND BIG DATA these days, the fact is, there is nothing dramatically new about having data at our fingertips. Businesses and governments have always collected data to varying degrees. What *is* new is that modern technology has increased the types of data that can be collected and made it cheaper, easier and faster to collect, store and analyze that data. An immediate and obvious implication is that with better data and faster technology, businesses can 'do analytics' better and faster. A less obvious, but more important implication, is that, with better data and new technologies, businesses should do analytics *differently*.

A Scientific Approach to Business

About 25 years ago, the health care industry adopted what is now known as 'evidence-based medicine'. This approach recognized that patient care decisions should be based on the results of rigorous scientific studies, rather than a physician's individual clinical experience and/or beliefs. To be sure, physician judgement remains a critical part of patient care: doctors interpret and evaluate the evidence, determine which studies are most relevant to a particular case, and help patients make difficult trade-offs. However, medicine, as a field, determined that *whenever possible*, decisions should be based on data and evidence rather than intuition or personal beliefs. In short, the medical profession adopted data-driven decision-making, and it did so because it led to better outcomes for patients, at lower costs.

So, our question is, why is the business world struggling to adopt data-driven decision-making? Why don't we talk about 'evidence-based business'?

For years, the business world has accepted — even celebrated — decisions based on intuition, gut instinct and lucky guesses. Perhaps this is because businesses didn't have the technological capabilities to systemically collect and analyze data; perhaps it is because they didn't have the infrastructure to carry out randomized experiments — the cornerstone of scientific research. Or, perhaps it is because the speed of decision-making is faster today, making it impossible to collect, analyze and interpret data at a relevant pace.

Big Data — and the technologies that enable it — changes all of this. Businesses now have access to enormous amounts of

Prescriptive analyses can deliver stark recommendations about particular courses of action that an organization should take.

data — in many cases, data that is generated and stored in real-time, and they have access to inexpensive (often free) software programs that can quickly analyze this data. With so much economic and social activity moving online, they also have access to a platform for running experiments. In short, Big Data allows companies to do analytics differently, enabling evidence-based business. Of course, it takes some work to get there.

While many different types of data and analytical techniques exist, it is helpful to consider three broad ways that organizations can utilize the data they collect.

DESCRIPTIVE ANALYSES. The simplest way that companies can use data is to describe, or 'take stock' of what is happening or has happened. Descriptive analyses keep track of your key performance indicators, typically consisting of simple statistical analyses — tables, graphs, and charts. What was our year-over-year change in sales? How many calls came into our call centre this week? What was the average handling time per call?

This form of data analysis is sometimes called 'dashboarding' because — like the dashboard of a car — it provides timely and easily accessible information on your most important parts and systems. Today, dashboarding is 'table-stakes' in any data-literate organization, and as with a car's dashboard, such analyses are largely used to identify where problems may exist or which areas need attention. While descriptive analyses may inform business decisions, on their own, they generally do not generate enough information to provide solutions.

PREDICTIVE ANALYSES. Predictive analyses use your organization's existing data — both structured and unstructured — to predict the value of variables that you don't possess, but would benefit from knowing. For example, beverage companies predict sales for each brand and package type in order to make production decisions; retailers predict the number of customers that will shop in their stores at Christmas time, to make hiring decisions; and hospitals predict the number of incoming emergency room patients so they can make staffing decisions.

Much of the enthusiasm around Big Data and advanced analytics has centered on the potential for carrying out faster, better and more reliable predictions. The popular press is awash with now well-known examples of predictive analyses: **Target** famously predicted which of its customers were likely to be pregnant; the **Obama** campaign predicted which voters were likely to vote democrat (if they made it to the polls); and the **Oakland Athletics** predicted which players were likely to generate the most wins.

While it is indeed true that Big Data and advanced analytics techniques allow for greater predictive power, the basic ideas behind predictive analytics haven't changed dramatically. Moreover, the critical issue for most managers is not which predictive algorithm to use; rather, it is identifying the metrics that are most valuable for your business to predict, and determining how to act on the output of the predictive exercise.

PRESCRIPTIVE ANALYSIS. Prescriptive analyses are different from descriptive or predictive analyses in that they provide direct insight into the consequences of different actions by uncovering the key cause-and-effect relationships that impact the outcomes your organization cares about. While they often involve similar analytical techniques as descriptive and predictive analyses, they require a more subtle and nuanced interpretation of the data.

Prescriptive analyses are about understanding what-causes-what, and why. While a predictive analysis aims to predict the value of an outcome of interest (sales, hospital wait-times), a prescriptive analysis aims to understand the factors that determine that outcome, so that it can be influenced in the organization's favour. For example, a beverage company might predict that one of its brands is unpopular with female buyers. But, to figure out how to increase its popularity with female customers, it needs to determine what is causing women not to like the product. Is it the taste? The marketing? The distribution channels?

While prescriptive analyses are the most difficult to carry out, they can deliver stark recommendations about particular courses of action that an organization should — or should not — take.

The Path to Prescription

Prescriptive analyses are the underpinnings of an evidence-based approach to business, but as indicated, they are difficult. That's because they require hard thinking — not just hard analytics. Following are three critical steps to getting started on the path to prescription.

1. FOCUS ON WHY AND HOW, NOT JUST ON WHAT, WHO AND WHICH

Descriptive and predictive analytics are about the 'what', 'who' and ‘which’ of a business. A good dashboard will tell you things like ‘what sales were last quarter’ or ‘which markets have grown most quickly over the last year’. A good predictive model might tell you ‘who is most likely to click on an online ad’, or ‘which of your trucks will require servicing next’. To be sure, these are valuable uses of data, but organizations today should not be satisfied just answering the what, who, and which: they should also be probing deeper into the *why* and the *how*.

Consider the fundraising department of a large academic hospital. The department likely has detailed data on all donations received to date, and some data on the demographics and other characteristics of its donors. It also likely has data on its own marketing, engagement and outreach activities, ideally in a way that can be matched to particular donors. So, how should the department use this data to grow its donor base and increase donation amounts?

The department may begin by describing donation activity over the past year: how much was raised? What was the average donation amount? What was the average donation frequency? It may build a profile of its donor base, including age, income, occupation and hospital visits. It may want to distinguish donor types — one-time versus monthly, and compare the characteristics of the different types. It could go a step further and build a model that predicts the likelihood that a one-time donor converts to a monthly donor, or a model that predicts the likelihood that a donor will increase her donation amount over time.

The output of these exercises would certainly help the team better target its marketing and engagement efforts. However, none of these analyses address the problem the fundraisers really want to solve: what actions can they take to increase donations?

For this, they need to know *why* some people donate and others don’t; *why* some people donate year after year, while others stop; and why some people increase their donations over time, while others decrease. More specifically, they want to know whether donor behaviour is influenced by actions they themselves have taken — or could take in the future. Did that recent TV campaign result in higher donations? Did offering tours of the hospital lead people to donate? Are email campaigns effective? Answers to each of these questions would prescribe a particular set of actions.

On the surface, these questions don’t seem that difficult to answer. Take the TV campaign. Couldn’t the team measure the volume of donations that came in the month before the campaign ran, and the month after? Unfortunately, it’s not that simple: perhaps the campaign ran during the holiday season, when people tend to be more charitable; or, perhaps it ran in the month before the deadline for tax receipts for charitable giving. If the data show higher donations after the campaign, can the fundraisers really conclude that the campaign caused the incremental donations?

Consider the case of hospital tours. A naïve analysis might compare donations by those who went on a tour and those who didn’t. Suppose the data indicate that those who went on the tour donated twice as much, and were more likely to convert from one-time donors to monthly donors: can the fundraisers conclude that *hospital tours increase donations*? Or, could it be that more committed donors are simply more interested in touring the hospital?

One might question whether all of this analysis is really necessary: couldn’t the fundraisers just ask people why they did or didn’t donate? Not that long ago, if you wanted to know why your customers were behaving in a particular way, you would run a focus group, or perhaps send out a survey, asking questions like, ‘Do you like our cars?’ If people answered affirmatively, you would then ask, ‘Why do you like our cars?’ Participants might say things like, ‘Your cars are safe, and I like safe cars. That is the most important thing to me’. You might then ask, ‘What else do you like?’ ‘They get great mileage and they’re environmentally friendly, and I like to know that my car is safe and environmentally friendly.’ Problems arise, however, when these seemingly-meaningful answers turn out to have come from an individual who owns a car that is neither safe nor environmentally friendly!

We have long understood that people don’t necessarily say what they think, and that actions can be more revealing about true motivations. Indeed, the difference between what people say and what they do can be substantial. What prescriptive analysis does is infer, or uncover, why people behave in the way they do, based on the decisions they are observed to have made.

2. KNOW WHERE YOUR DATA COMES FROM

Knowing your data well is always important; but as our examples show, knowing the process — the set of decisions and behaviours that is generating the data — is critical to performing a prescriptive analysis.

Statistician **Abraham Wald**'s work provides a particularly salient example of the importance of understanding the behaviour that generates your data. During World War II, aircraft returning from bombing raids would often be riddled with bullets from anti-aircraft fire. Parts of the aircraft could be reinforced with stronger armor, but only up to a point: too much armor would make the aircraft too heavy. Thus, the objective was to figure out where to put the extra armor.

Wald offered a simple recommendation: fortify the planes in the areas *without* bullet holes. This recommendation was counterintuitive to the engineers who had been focusing on examining the parts of the planes that had been hit by bullets. But, it is in fact quite sensible (and obvious) once you consider the process that generated the data the engineers were studying.

How did the planes return to base? By staying in the air. Therefore if a plane had a bullet hole, and it returned, the location of that bullet hole did not need to be fortified. Since it is difficult to aim anti-aircraft fire, Wald assumed that the planes that did not return to base were shot in the places where the returning planes were not hit. By understanding the *process* that gave rise to the data in the first place, he was able to suggest an effective plan of action. Similarly, returning to the hospital fundraising example, understanding the process that led some donors to tour the hospital while others did not is critical to understanding how to interpret the two groups' behaviour.

These same ideas apply to modern businesses. Consider **eBay**, arguably one of the most successful and enduring internet companies. The site — which offers an online platform for buyers and sellers to interact — has established a reputation system whereby buyers can rate the sellers they purchase from. The resulting 'reputation score' is the fraction of reviews that are positive. New buyers can see the ratings that previous buyers have left, and are able to use this information to determine whether a seller is trustworthy. Such a mechanism is important, because without face-to-face interaction, trust is particularly difficult to establish online.

A few years ago, eBay managers noticed something curious: seller reputations were overwhelmingly positive. Indeed, most sellers had never received negative feedback, and less than half of one per cent of all feedback was negative. As a leading data-driven organization, eBay wondered whether these reputation scores were reliable. One possibility was that the buyers were overwhelmingly happy; but an alternative was that unhappy buyers were simply not giving reviews at all.

Digging further into its data, eBay discovered that formal disputes — whereby a buyer appeals to eBay about a seller — were *twice as common* as negative reviews. It also had some data on emails between buyers and sellers, and discovered that emails with negative or nasty words were *six times as common* as negative reviews. Together, this made eBay question whether its reputation system was working. If disputes between buyers and sellers were happening, why weren't they showing up in the reputation data? Did the reputation metric need to be redesigned?

In order to figure out how to proceed, eBay considered the *process* that generated the seller feedback data. It realized that while the metric measured the percentage of a seller's reviews that were positive, it missed an important step: the fact that buyers first decide whether to give the seller a review. eBay realized that an unhappy buyer actually had two choices: give a negative review, or give no review.

If the buyer gave no review, how would eBay know that they were unhappy? Given that eBay had data on everything that takes place on its site, it came up with a proxy for being a very unhappy buyer: it began to identify buyers who stopped coming to the site after a transaction. eBay tracked the behaviour of these users and discovered that a handful of sellers was generating non-returning buyers. Not surprisingly, it found that these were the same sellers who generated disputes and nasty emails; yet, their average feedback score was still high. Recognizing that these sellers were hurting their company, ebay set out to develop ways to direct buyers to the sellers that made people *return* to the site, and to direct them away from those who generated non-returning customers.

Importantly, only by understanding *the process that generated the reputation score* did eBay realize that the score was not serving its intended purpose.

3. THE BURDEN OF PROOF

Adopting an evidence-based approach to business requires you to determine what, in fact, is admissible as evidence. It is widely accepted that the best evidence comes from controlled experiments, which allow one to isolate the impact of a single variable on the outcome of interest. As economic and social activity continues to migrate online, businesses can, at relatively low-cost, run experiments that can deliver stark conclusions about which actions impact key organizational outcomes. **Google**, **Facebook**,

Data and analytics are only valuable if they are used to generate insights, learning and evidence that inform business decisions.

Amazon and many other software-focused companies run thousands of experiments each year. As examples, Google experimentally tests improvements to its search algorithm before implementing them system-wide, Facebook continually tests ad campaigns to determine the most effective types of advertising on its site, and Amazon uses experiments to test changes to its webpage design.

Such experiments are not limited to online businesses. Consider our hospital fundraising group, and their assessment of whether hospital tours generate donations. The finding that *people who went on tours donated twice as much as those who didn't* should not pass the burden of proof, because it is possible — even probable — that the individuals who chose to take the tour were already more engaged with the hospital and more likely to donate. Suppose, instead, that the fundraisers invited potential donors to sign-up for tours, and then randomly chose half of those who signed up to go on a tour, while the other half was told that there was insufficient space that year. Now, the fundraisers could determine whether the tour was the reason why those individuals donated more — since both those who went on the tour and those who didn't had expressed interested in the tour. This set-up would provide clear evidence on the value of offering tours for potential donors.

Unfortunately, organizations can't always run such experiments. Much of the data companies have comes from regular day-to-day operations, and much of the analytics companies carry out is based on this type of data. Such data can still be incredibly useful. The key question managers need to ask is whether the relationships uncovered in these data can be interpreted as if they were generated from an experiment. Often they can, and a few pointed questions will help to uncover key issues that need to be considered.

For example, in the case of hospital tours and donations, the fundraisers should be asking questions like, Did the people who went on the tour also give more in the previous year? Was the tour only offered to the hospital's largest donors? Did the timing of the tours coincide with other marketing efforts by the hospital? Are the demographics of those who went on the tour different than those who didn't? If the answers to these questions — all of which can be uncovered in the hospital's data — suggest that those who took the tour are otherwise pretty similar to those who didn't take the tour, the case can be made that *tours positively impact donations*.

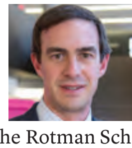
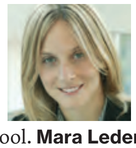
Closing the Gap

Big Data has the potential to revolutionize the way organizations make decisions. But data, on its own — or even combined with cutting-edge technologies and highly skilled data scientists — isn't enough. Data and analytics are only valuable if they are used to generate insights, learning and evidence that inform business decisions — and this means that managers have a critical role to play.

It is up to managers to identify where analytics can add value and improve decision-making. This means that analytics need to start as business questions that are relevant to the decisions faced by managers. For this to happen, managers must have a basic level of data literacy so that they can determine which types of decisions could be improved by data analytics. In addition, as we've seen throughout our examples, interpreting data requires a deep understanding of the behaviours and decisions that generated the data in the first place.

Because of their deep domain expertise and their involvement with day-to-day decision making, managers are uniquely positioned to help data analysts understand the nuances and details of where a company's data comes from. Managers and analysts must learn to work together, marrying business expertise with data expertise, so that analytics are done purposefully and usefully.

If analytics are not driven by business decisions, the role of analytics in your business will be underwhelming at best, and destructive at worst. On the other hand, if your analytics are motivated by the key decisions and challenges facing your managers, then Big Data truly has the potential to transform your organization for the better. ***RM***

Bernardo Blum is an Associate Professor of Business Economics at the Rotman School of Management. **Avi Goldfarb** is the Ellison Professor of Marketing at the Rotman School. **Mara Lederman** is an Associate Professor of Strategic Management at the Rotman School. Together, they teach the Rotman School's Executive Program in Data Literacy. Rotman faculty research is ranked #4 globally by the *Financial Times*.