BIG DATA

HOW TO TURN INFORMATION INTO VALUE

JAVIER ZAMORA
Lecturer Information Systems, IESE Business School

We now have access to more data than our systems are designed to cope with. We can use this information to improve our decision making but only if we know what we're looking for and don't mistake correlation for causality.

nformation is power, but only if you know what to do with it. We have now entered into the era of what is called "big data," data that goes far beyond the transactional data that most companies' information technology is designed to deal with. Now we have vast volumes of data coming from social media as well as a variety of sensors, because mobile phones emit a huge quantity of data. However, it's a process analogous to panning for gold, because this information is of little value unless we find ways to sift through it in order to translate it into meaningful information from which we can make meaningful decisions.

UNSTRUCTURED DATA
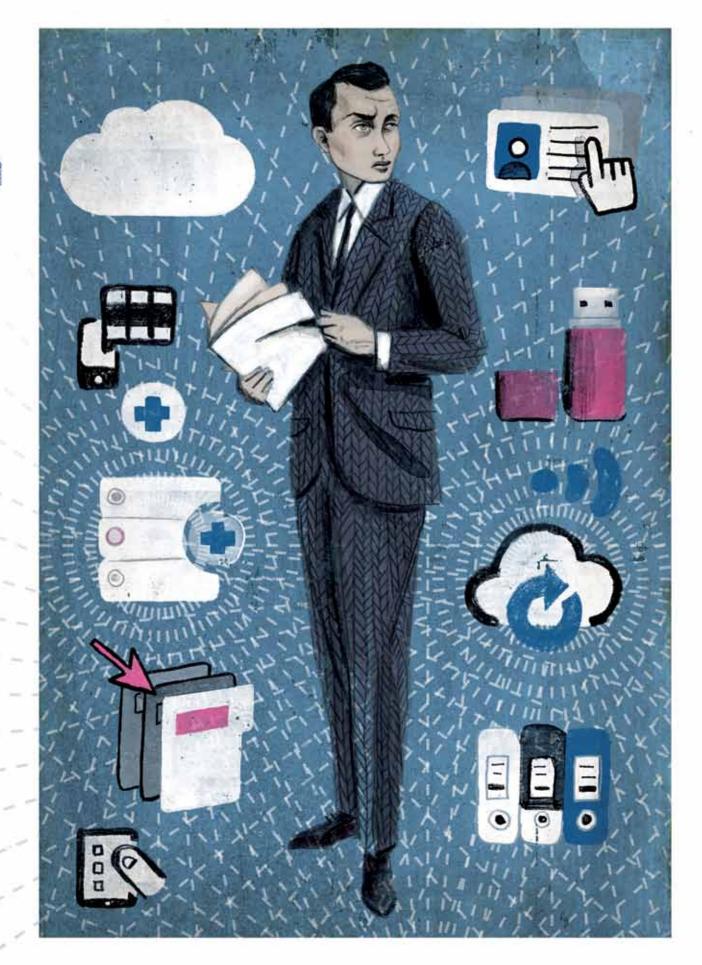
There are three main sources of big data. The first is structured, this is traditional data that companies hold or gather from customers' transactions. Then there's a lot of unstructured data, which includes the way customers behave when they're looking at a web page, as well as blogs and the tone and sentiment expressed in emails, and of course comments on social media about products and services. The third source is the information from sensors. Anyone using a smartphone is generating huge amounts of information. A transaction carried out using a mobile provides a lot of interesting information. First there is information about your location, but there is also information prior to the transaction, such as who you called beforehand or what kind of apps you used. The question is: how can we use all this data to enhance our own decision-making process?

When we talk about the characteristics of big data we talk about the three Vs: volume, velocity and variety. The

BIG DATA IS NOT A SUBSTITUTE FOR INTUITION BUT WE CAN USE BOTH GUT FEELING AND ANALYTICAL THINKING TO ARRIVE AT DECISIONS

fourth V would be how to get some value out of all this. Companies need to understand that it's not just an issue of technology. It involves changing the culture and the processes inside the company.

Traditionally, many business decisions are based on intuition. Big data is not a substitute for intuition but it can give us more insight. It's not a matter of either intuition or data, it's about using both. We can use gut feeling and analytical thinking. There's an interesting example of this in the recently released movie *Moneyball*. The film tells the true story of **Billy Beane**, owner of the Oakland Athletics baseball team, and his bid to build a winning team with a fraction of the budget of his competitors. Intuition – in the person of the talent scouts – says that the only way to win is to sign existing or rising stars, none of whom the club can afford.

WINNING WITH NUMBERS

Aided by the fact that baseball is a data-heavy sport, where every fan knows the players' batting averages, number of strike-outs and so on, a computer nerd in the team's office runs a statistical analysis on a string of second-rate players who are either near the end of their playing life or whose careers have been plagued by social problems and are therefore affordable. He convinces **Beane** that by aggregating their talents it is statistically possible to put together a winning team. Inevitably, the scouts dismiss the idea as preposterous, as does the manager, and almost as inevitably Oakland, after a disastrous start, go on to a record-breaking 20-match winning streak.

This is not to say that the intuition of the scouts was wrong, because it is equally true that great sports teams have been built around a few stars and a supporting cast of makeweights. However, the story is illustrative of how data can complement intuition and experience.

The volume of data that a computer can deal with also sometimes throws up correlations that we could not otherwise have seen. For example, a study was carried out in the public high schools of Gwinnett County, Atlanta (the 14th biggest county in the USA), where pupils were achieving poor test scores. Various approaches were tried, including increasing the number of teachers and reducing class size, but the results did not improve. Then they analyzed data from the school district and they discovered that the most successful graduates were successful in Algebra 1. Then they found a correlation in that the students who were doing well in Algebra 1 were also doing well in creative writing. This is not an obvious correlation but when you have a lot of data you discover these sorts of connections and they open up new approaches to problem solving.

The pioneer in making these sorts of connections based on a large amount of data is Google. Google hired **Hal Varian**, an emeritus professor of economics from Berkeley, who is Google's chief economist. Google is using the enormous amount of data it has access to in order to improve prediction models. For example, using a product called Google Flu they mapped the spread of people who were searching for information about flu symptoms. If somebody is searching Google for flu symptoms it's very likely that they or someone close to them has got the flu. And if you study the spread of flu through these searches it's more accurate than the Centers for Disease Control epidemiology. The same thing applies to house prices. Instead of looking at how house prices have behaved in the previous quarter, you look at how people are searching for houses via Google in different regions, or through different estate agents. The prediction based on the searches is better than a prediction based on any other economic model, which uses past data instead of present data.

This is called "now casting" because with so much data you can make better predictions. Now casting exploits data series from a range of sources in order to predict

economic trends. For example, by extrapolating from unemployment figures, trade data and so forth, it is possible to make accurate predictions the about current quarter's GDP before the official GDP figures are published.

To give you an example of the scale of big data, Wal-Mart has one million consumer transactions every hour and this translates into 2.5 petabytes of information, which is equivalent to producing 167 times the information in the Library of Congress every hour. If they can analyze this data then they can optimize their supply chain and decide when and where to put which products in the shops. But this highlights one of the problems with big data: you have to know what you are looking for. If you don't know what you're looking for you can easily fall into the trap of mistaking correlation for causality.

The data explosion is exponential. Imagine a city with a population of 300,000 that is growing according to Moore's Law. That means that within 20 years it would have the population of the entire earth. But it's not only the way that computer power is growing, it's the way that the algorithms are improving. In the mid-to long-term this new data is going to generate more wealth. But right now I think we are at Year Zero as far as big data is concerned. We are in a sort of Wild West scenario, particularly in regard to copyright and privacy issues.

DATA: A NEW ASSET

At the last meeting of the World Economic Forum in Davos, big data was considered as a new asset. They compared it to the new oil of the economy. But if big data is an asset and you want to establish a market then the first thing you have to do it is establish ownership. There will have to be legislation on what kind of data can be used and what kind companies can collect and use. What is needed is to create a market of big data but the issue of ownership remains complicated and controversial.

Big data will change the architecture of business just as the invention of the electric dynamo, which replaced the steam engine as the central power source, changed the layout of the factories from narrow high rise structures to low-rise horizontal ones and created different ways of working. It will change governance and process, but this requires thinking outside the box. It's like having an analog watch and saying you want it to become a digital watch. You can't do it, it's a completely different thing. With big data we are still thinking in analog terms and trying to reproduce an analog experience in this new digital world but it requires a new way of thinking. It also throws up important issues of privacy, security and risk.

TIME FOR A NEW MODEL

Sony Music approached the digital revolution with an army of lawyers but the fact is this technology is here to stay and so things have to change. It's better to start thinking of a new business model from which you can extract value than fighting to defend the old one. The problem is the old business models are threatened at a higher speed than we are able to generate new ones. This is where the tension lies.

Things are moving so fast they are destroying layers of the economy faster than we can think about how to adapt them. We've seen this in how, since the slump in 2008, the U.S. economy is recovering but employment is not.

Einstein said information is not knowledge. This is more true now than ever, not just because we have so much information but because information is becoming a commodity. So the question is how to transform this information into knowledge and decisions. This is a reverse of the situation we have been experiencing up until now, where information was scarce. Experts were filtering in information and now the challenge is to filter it out in order to reach meaningful decisions.

IT'S BETTER TO START THINKING OF A NEW BUSINESS MODEL FROM WHICH YOU CAN EXTRACT VALUE THAN FIGHTING TO DEFEND THE OLD ONE