

3

Business Impact of Big Data

Organizations are starting to realize that big data is more about business transformation than IT transformation. Big data is allowing companies to answer questions they could not previously answer, and make more timely decisions at a finer level of fidelity than before, yielding new insights that can deliver business differentiation and new operational efficiencies. Let's take a look at an example of how big data is transforming how we look at business.

For decades, leading organizations have been exploiting new data sources, plus new technologies, for business differentiation and competitive advantage. And for the most part, the questions that the business users are trying to ask, and answer, with these data sources and new technologies really haven't changed:

- Who are my most valuable customers?
- What are my most important products?
- What are my most successful campaigns?
- What are my best performing channels?
- What are my most effective employees?

The more I thought about these "simple" questions, the more I realized just how "not simple" these questions really were. Because of the new insights available from new big data sources, companies are able to take these types of "simple" questions to the next level of sophistication and understanding.

Let's look at the most valuable customer question. When you ask who your most valuable customers are, do you mean largest by revenue (which is how many companies today still define their most valuable customers)? Or do you mean the most profitable customers, contemplating more aspects of the customer engagement including marketing and sales costs, cost to service, returns, and payment history (which is how some of the more advanced companies think today)? Or by adding

2 Chapter 3

social media into the mix, do you now mean your most influential customers and the financial value associated with their circle of friends?

Companies are learning that their most profitable customers may not actually be their most valuable customers because of the net influencer or advocacy effect. Advocates can have significant influence and persuasive effect on a larger community of customers, and the profitability of the “baskets” associated with that community of customers. Same with the most important product question, which retailers have understood for quite a while (think loss leaders like milk that drive store traffic even though they don’t drive much in the form of profits), and consumer goods manufacturers understand as well (think category strategies and the use of flanking products to protect their premium-priced core products).

Those nebulous and hard-to-define words, like valuable, important, and successful, allow the business users to move beyond just financial measures and to consider the entirety of the contributions those customers, products, and campaigns make to the business. It is the basis for a more engaging business discussion about what data sources could be critical in defining “valuable” and what analytic models could be used to quantify “valuable.” It’s the basis for a wonderful conversation that you can have with your business users about defining those valuable, important, and successful words in light of what big data and advanced analytics can bring to the table.

Big Data Impacts: The Questions Business Users Can Answer

Big data has changed the nuances for defining and quantifying terms such as valuable, important, and successful. It is these nuances that fuel the insights that are the source of competitive advantage and business differentiation. New big data sources, plus new advanced analytic capabilities, enable higher fidelity answers to these questions, and provide a more complete understanding of your customers, products, and operations that can drive business impact across various business functions, such as:

- Merchandising to identify which marketing promotions and campaigns are the most effective in driving store or site traffic and sales.
- Marketing to optimize prices for perishable goods such as groceries, airline seats, and fashion merchandise.
- Sales to optimize the allocation of scarce sales resources against the best sales opportunities and most important or highest potential accounts.
- Procurement to identify which suppliers are most cost-effective in delivering high-quality products in a predictable and timely manner.

- Manufacturing to flag machine performance and process variances that might be indicators of manufacturing, processing, or quality problems.
- Human Resources to identify the characteristics and behaviors of your most successful and effective employees.

Managing Using the Right Metrics

Since baseball is one of my loves in life, and in honor of the enlightening book, *Moneyball: The Art of Winning an Unfair Game*, by Michael Lewis (Norton, 2004), I thought it was only appropriate to discuss how the pursuit and identification of the right metrics has not only changed how the game of baseball is managed, but has the same potential impact on how you manage your business.

In 2004, Lewis wrote the book *Moneyball*, which chronicled how the Oakland A's and Billy Beane, their general manager, were using new data and metrics in order to determine the value of any particular player. The A's were unique at that time in the use of *sabermetrics*, which is the application of statistical analysis to baseball data in order to evaluate and compare the performance of individual players. The results were that the A's had a demonstrable competitive advantage in determining how much to pay any particular player playing any specific position, especially in the costly era of free agency.

As a result, the A's enjoyed a significant cost advantage in what they were paying for wins versus a team like the Yankees. The comparison is shown in Figure 3-1.

	Salaries (\$M)		Wins		Cost per Win (\$M)		
	A's	Yankees	A's	Yankees	A's	Yankees	A's % of Yankees
2005	\$55.4	\$208.3	88	95	\$0.63	\$2.19	29%
2004	\$59.4	\$184.2	91	101	\$0.65	\$1.82	36%
2003	\$50.3	\$152.7	96	101	\$0.52	\$1.51	35%
2002	\$40.0	\$125.9	103	103	\$0.39	\$1.22	32%
2001	\$33.8	\$112.3	102	95	\$0.33	\$1.18	28%
2000	\$32.1	\$88.1	91	87	\$0.35	\$1.01	35%
	\$271.0	\$871.5	571	582	\$0.47	\$1.50	32%

<p>Yankee Batting KPIs:</p> <ul style="list-style-type: none"> • Batting average • RBIs • Fielding percentage • Steals 	<p>A's Batting KPIs:</p> <ul style="list-style-type: none"> • On-base percentage • Slugging percentage
---	---

Figure 3-1: Payroll cost per win: Athletics versus Yankees

4 Chapter 3

Unfortunately for Billy Beane and the Oakland A's, other teams (most notably the Boston Red Sox) copied this model and reduced the competitive advantage that the A's briefly enjoyed. But that's the nature of a competitive business isn't it, whether it's in sports, retail, banking, entertainment, telecommunications, or healthcare.

So how does one survive in a world where competitive advantage via analytics can be so short-lived? By constantly innovating, thinking differently, and looking at new sources of data and analytic tools to bring to light those significant, material, and actionable insights that can differentiate your business from that of your competitors.

One of the challenges with metrics is that eventually folks learn how to game the metrics for their own advantage. Sticking with our baseball scenario, let's take the Fielding Percentage metric as an example. The Fielding Percentage metric is calculated as the total number of plays (chances minus errors) divided by the number of total chances. Some players have learned that one of the ways to improve their Fielding Percentage is to stop trying to field balls that are outside of their fielding comfort zone. If you don't try hard for the ball, there can't be an error assessed. While that might be good for the individual's performance numbers, it is obviously less than ideal for the team who wants all of their players trying to make plays in the field. Let's see how that works.

Let's say that an outfielder has 1,000 fielding chances, and makes 20 errors out of those 1,000 fielding chances for a Fielding Percentage of 98 percent (see Figure 3-2). Now, if the fielder doesn't try to field the 100 hardest opportunities (resulting in only 900 Fielding Chances), he will likely cut down significantly on the number of errors (let's say, eliminating 10 errors) resulting in an increased Fielding Percentage of 98.9 percent.

- Example: Fielding Percentage
- **Fielding percentage:** total plays (chances minus errors) divided by the number of total chances

$$\text{Fielding Percentage} = \frac{(\text{Chances} - \text{Errors})}{\text{Number of total chances}}$$

- However, a player can "game" the system by not trying to catch difficult chances

	Tries	Doesn't Try
Number of Chances	1000	900
Errors (example)	20	10
Fielding Percentage	98.0%	98.9%

By not trying to catch the 100 most difficult chances, the player commits an estimated 10 fewer errors and improves their field percentage

- Note: In 2011 for Center Fielders, the #1 and #11 top-fielding percentages were separated by 0.9 basis points (100.0% to 99.1%).

Figure 3-2: Picking the wrong metrics can incent the wrong behaviors

While the 0.9 basis-point difference (98.9 minus 98.0) between the two efforts may not seem significant, suffice it to say that the difference between the #1 center fielder in Major League Baseball in 2011 and the #11 center fielder was only 0.9 basis points. The difference probably means millions of dollars to their playing contract.

So the bottom line is that some players have figured out that they will perform better by only trying to field those opportunities within their comfort zone. Not the sort of behavior that leads to very many World Series appearances.

So how does the world of big data change this measure? Baseball stadiums have installed video cameras ~~focused on the players~~ throughout the stadium to get a better idea as to actual game dynamics. One of the benefits of these cameras is a new set of metrics that are better predictors of players' performance. **feet**

For example, video cameras now can measure how much many ~~yards~~ a particular fielder can cover within a certain period of time in fielding their position. Ultimately, this will lead to the creation of an Effective Fielding Range metric which measures how much of the playing field the fielder can cover, and how **effectively** they cover the playing field (see Figure 3-3). This metric will allow baseball management to value players differently because Effective Fielding Range is a much better predictor of fielding performance than the traditional Fielding Percentage.

effectively
also make font/point size
consistent.

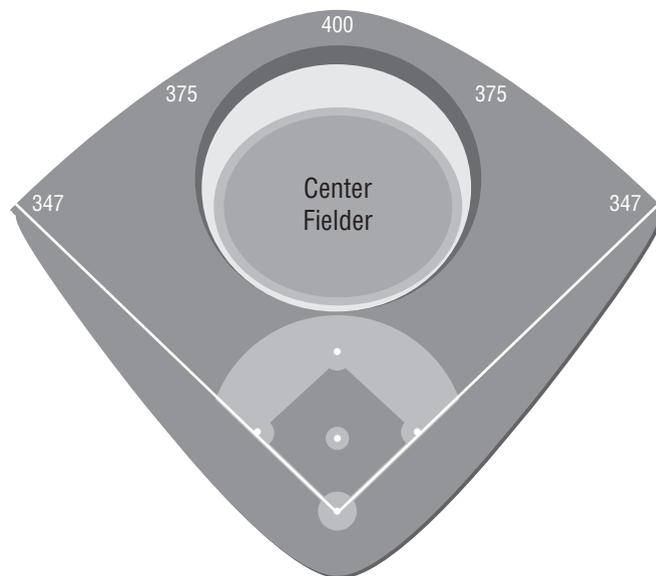


Figure 3-3: Big data hits baseball

6 Chapter 3

As illustrated in the figure, the Center Fielder is very efficient in covering the outfield going left, right, or forward (indicated by the green coverage area), but is less efficient going backwards (indicated by the yellow and red coverage areas).

Much like the world of baseball, organizations must be constantly vigilant in search of metrics that are better predictors of business performance. The new data sources and capabilities enabled by big data hold huge potential to be the first mover in uncovering those significant, measurable, and actionable insights that can lead to competitive advantage—on the baseball field or in the corporate battlefields.

Data Monetization Opportunities

Data monetization is certainly the holy grail of the big data discussion: How do I leverage my vast wealth of customer, product, and operational insights to provide new revenue-generating products and services, enhance product performance and the product experience, and create a more compelling and “sticky” customer relationship?

But how does one even start thinking about this data monetization discussion? Let me take a data monetization example from the digital media world and present a process that other industries can use to uncover and capitalize on potential data monetization opportunities.

Digital Media Data Monetization Example

Digital media companies like Yahoo!, Google, Facebook, and Twitter have worked to master the data monetization process. They must because their entire business model is built on monetizing data. These companies work with bytes to create services, unlike most other companies who work with atoms to build physical products like shoes, tractors, houses, and burrito bowls with double chicken and guacamole.

So what process do these digital media companies go through to identify how to monetize their data assets? The data monetization process starts with two key understandings:

1. Who are my target customers (targeted personas) and what business solutions do they need for which they are willing to pay?
2. What data assets do I have (or could I have)?

Once you have a solid understanding of these two questions, then you are in a position to start the data monetization process.

Digital Media Data Assets and Understanding Target Users

First, digital media companies need to identify and really (and I mean really!) understand their target customers—that is, who is making the million dollar marketing and campaign decisions, and what information and insights do they need to make those decisions? Digital media companies target the following three customers or personas: Media Planners and Buyers, Campaign Managers, and Digital Media Executives. These digital media decision-makers buy the following “solutions”:

- Audiences, such as soccer moms, country squires, gray power, and weekend warriors
- Inventory (like sports, finance, news, and entertainment) available on certain days and times of days
- Results or measures, such as Cost per Thousands of Impressions (CPM), Cost Per Acquisition (CPA), product sales, or conversions (where conversions could include getting a visitor to share their e-mail address, request a quote, or schedule a reservation)

(CPM)

impressions

For each of these targeted personas, the digital media company needs to understand what questions they are trying to answer, what decisions they are trying to make, under what circumstances they are making these decisions, and within what sort of environment or user experience they are typically working when they have to answer their questions and make their decisions.

Next, digital media companies assess the breadth, depth, and quality of their data assets, including:

- Visitors and their associated demographic, psycho-demographic, and behavioral insights
- Properties and the type of content and advertising real estate (e.g., full banner, pop-under, skyscraper, leaderboard, half-page) that is provided on properties (like Yahoo! Finance, Yahoo! Sports, or Yahoo! Entertainment)
- Activities that visitors perform on those properties (for example, they viewed a display impression, clicked a display ad, entered a keyword search) including how often, how recent, and in what sequence

moused over a display ad,

This data assessment process should also include what additional data could be captured through data acquisition, as well as through more robust instrumentation and experimentation techniques.

Data Monetization Transformations and Enrichments

The key challenge is then to transform, augment, enrich, and repackage the data assets into the solutions that the target digital media customers want to buy. For example, digital media companies instrument or set up their sites and tag their visitors (via cookies) to capture visitors' web site and search activities in order to determine or ascertain additional visitor insights, including:

- Geographic information such as ZIP code, city, state, and country
- Demographic information such as gender, age, income, social class, religion, race, and family lifecycle
- Psycho-demographic information such as lifestyle, personality, and values
- Behavioral attributes such as consumption behaviors, lifestyles, patterns of buying and using, patterns of spending money and time, and similar factors
- Product categories of interest (Schmarzo likes Chipotle, Starbucks, the Cubs and the Giants, and all things basketball)
- Social influences such as interests, passions, associations, and affiliations

With this information in hand, the digital media company needs the data processing capacity and advanced analytical skills to profile, segment, and package those visitors into the audiences that advertisers and advertising agencies want to buy.

This data transformation, augmentation, and enrichment process is then repeated in converting properties into inventory, visitor activities into digital treatments, and campaigns into results such as sales and conversions (see Table 3-1).

NOTE The table below has been organized with step 1 at the far right, as it represents the end solutions that we are trying to deliver. Step 2 is on the far left as it represents the key data assets, which will go through step 3 to be transformed and enriched into our targeted solutions.

Table 3-1: Data Monetization Example—Digital Media Company

Step 2: Assess Data Assets	Step 3: Identifying Transformation, Enrichment, and Analytic Requirements	Step 1: Define Digital Advertiser Solutions
Visitor	Demographics Insights Psycho-Demographics Insights Behavioral Insights Social and Mobile Insights	Audiences What audiences am I reaching? Who is my most engaged audience? What similar audiences could I target?

Step 2: Assess Data Assets	Step 3: Identifying Transformation, Enrichment, and Analytic Requirements	Step 1: Define Digital Advertiser Solutions
Properties (Sites)	Product categories (Sports, Finance) Audiences Premium vs. Remnant	Inventory What inventories are most effective? What product categories are most effective? What other product categories should I use?
Web Activities	Impressions Clicks Keyword Searches Social Posts and Activities Mobile Tracking	Marketing Treatments What marketing treatments are most effective? What are minimum frequency/recency levels? What is the optimal sequencing of treatments?
Campaigns	Instrumentation Analytics (Attribution, Audience Insights, Benchmarking) Optimizations and Predictions Recommendations User Experience	Sales/Conversions/CPM Will I achieve campaign objectives (predict)? What will be the impact if I re-allocate spending? What recommended changes will improve performance? How can I optimize inflight cross-media spending?

Based on this digital media example, here are the steps that your company needs to go through in order to better understand how to monetize your data assets.

1. Identify your target customers and their desired solutions (solution capabilities and required insights) in order to optimize their performance and simplify their jobs. Identify and profile the target business customers or personas for those solutions, and internalize how those customers will use that solution within their existing work environment. Quantify the business value of those solutions, and document the business questions the users need to answer and business decisions the business users need to make as part of the desired solution.
2. Inventory and assess your data assets; that is, identify the most important and valuable “nouns” of your business. Understand what additional data could be

gathered to enrich your data asset base via data acquisition and a more robust instrumentation and experimentation strategy.

3. Understand the aggregation, transformation, cleansing, alignment, data enrichment, and analytic processes necessary to transform your data assets into business solutions. Document what insights and analytics you can package that meets your customers' needs for a solution that optimizes business performance and simplifies their jobs. Identify the data enrichment and analytic processes necessary to transform data into actionable insights and understand how those insights manifest themselves within the customers' user experience.

There are numerous opportunities for organizations to improve product performance, enhance product design and development, preempt product failure, and enhance the overall user (shopper, driver, patient, subscriber, member) experience. More and more, the data and the resulting insights teased out of the data will become a key component, and potentially a differentiator, in the products and services that companies provide.

Summary

This chapter covered how asking the right questions is one of the key starting points in your big data journey. You learned how big data has changed the nuances for defining and quantifying terms, such as *valuable*, *important*, and *successful*, and saw some examples of how big data is helping various business functions ask the right questions.

Then I reviewed how big data is enabling organizations to identify new measures and metrics that are better predictors of business performance. I discussed the impact that the book *Moneyball* and the world of sabermetrics has had on helping baseball teams, particularly the Oakland A's, exploit a superior understanding of the "right" metrics to optimize baseball success on the baseball field. I also provided an example of how big data is taking the world of baseball analytics to the next level of predictive excellence with new insights about baseball player performance that are better predictors of in-game success.

The chapter concluded with a discussion on how you can monetize your data assets. I reviewed how your organization can leverage data assets to deliver new revenue opportunities and a more compelling, differentiated business relationship through superior customer, product, and market insights. I used the world of digital media marketing as an example and provided a "How To" framework to help your

at a finer level of fidelity

organization explore data monetization opportunities by understanding your target customers (personas) and their desired solutions, understanding your data assets, and by identifying the data transformation, enrichment, and analytic processes necessary to transform your data assets into business solutions.

