

## UNIT 5

# BUSINESS ANALYTICS

### 5.1

#### BIG DATA EMERGING TECHNOLOGIES

Apart from the above mentioned big data technologies, there are several other emerging big data technologies. The following are some essential technologies among them:

##### 1. TensorFlow

TensorFlow combines multiple comprehensive libraries, flexible ecosystem tools, and community resources that help researchers implement the state-of-art in Machine Learning. Besides, this ultimately allows developers to build and deploy machine learning-powered applications in specific environments.

TensorFlow was introduced in 2019 by Google Brain Team. It is mainly based on C++, CUDA, and Python. Companies like Google, eBay, Intel, and Airbnb are using this technology for their business requirements.

##### 2. Beam

Apache Beam consists of a portable API layer that helps build and maintain sophisticated parallel-data processing pipelines. Apart from this, it also allows the execution of built pipelines across a diversity of execution engines or runners.

Apache Beam was introduced in June 2016 by the Apache Software Foundation. It is written in Python and Java. Some leading companies like Amazon, ORACLE, Cisco, and VerizonWireless are using this technology.

##### 3. Docker

Docker is defined as the special tool purposely developed to create, deploy, and execute applications easier by using containers. Containers usually help developers pack up applications properly, including all the required components like libraries and dependencies. Typically, containers bind all components and ship them all together as a package.

Docker was introduced in March 2003 by **Docker Inc.** It is based on the Go language. Companies like Business Insider, Quora, Paypal, and Splunk are using this technology.

#### 4. Airflow

Airflow is a technology that is defined as a workflow automation and scheduling system. This technology is mainly used to control, and maintain data pipelines. It contains workflows designed using the DAGs (Directed Acyclic Graphs) mechanism and consisting of different tasks. The developers can also define workflows in codes that help in easy testing, maintenance, and versioning.

Airflow was introduced in May 2019 by the **Apache Software Foundation**. It is based on a Python language. Companies like Checkr and Airbnb are using this leading technology.

#### 5. Kubernetes

Kubernetes is defined as a vendor-agnostic cluster and container management tool made open-source in 2014 by Google. It provides a platform for automation, deployment, scaling, and application container operations in the host clusters.

Kubernetes was introduced in July 2015 by the **Cloud Native Computing Foundation**. It is written in the Go language. Companies like American Express, Pear Deck, PeopleSource, and Northwestern Mutual are making good use of this technology.

These are emerging technologies. However, they are not limited because the ecosystem of big data is constantly emerging. That is why new technologies are coming at a very fast pace based on the demand and requirements of IT industries.

## 5.2

## BUSINESS ANALYTICS

Today, we are at an inflection point at which we have the intellectual and computational power to fully capitalize on and ride a wave to cost effective top-line revenue growth powered by Big Data analytics.

Predictive analytics in the past has been largely constrained so much so that many of the constraints became well-entrenched assumptions that were the shackles that constrained

the potential results. Even so, those who have used predictive analytics have realized significant, quantifiable business value. In this new age of Big Data analytics, the assumptions have been cast aside as Big Data is disrupting the well-known but not fully realized discipline of predictive analytics.

The term predictive analytics was coined to distinguish statistics from more advanced type of calculations that are used to predict likelihoods of future outcomes. Certain industries, such as banks, insurance companies, and digital advertising, have fully embraced predictive analytics but most other businesses are still at very early stages of adoption. While there is a full spectrum of analytics for Big Data analytics the emphasis will be to use the descriptive statistics in the Big Data exploratory stage when the value of the data isn't yet known and move toward the predictive and prescriptive analytics stages as value becomes known or discovered (see Figure)

Descriptive Analytics (Business Intelligence)	Predictive Analytics	Prescriptive Analytics
<ul style="list-style-type: none"> <li>o What and when did it happen?</li> <li>o How much is impacted and how often does it happen?</li> <li>o What is the problem?</li> </ul>	<ul style="list-style-type: none"> <li>o What is likely to happen next?</li> <li>o What if these trends continue?</li> <li>o What if?</li> </ul>	<ul style="list-style-type: none"> <li>o What is the best answer?</li> <li>o What is the best outcome given uncertainty?</li> <li>o What are significantly differing and better choices?</li> </ul>
Statistics	Data Mining Predictive Modeling Machine Learning Forecasting Simulation	Constraint-based optimization Multiobjective optimization Global optimization
Information Management		

Figure: Big Data Analytics

Big Data analytics uses predictive and prescriptive analytics and is changing the analytics landscape. While descriptive statistics, initially made popular by SAS and SPSS, describe what has happened in the past, predictive analytics uses the past information to predict future outcomes with some degree of likelihood. Prescriptive analytics takes that past information and uses it to direct future activities to achieve optimal or near optimal results. While each of these types of techniques has been used for decades there are major shifts underway when these techniques are combined with Big Data. The major shifts are:

- ✓ Using all or more of the data to create a predictive model.

- ✓ Combining multiple analytic models and techniques to improve the results
- ✓ Creating a closed loop where new learnings are used to adapt the production models
- ✓ Using the predictive models in as close to real time as possible
- ✓ Focusing on applying predictive model techniques (a.k.a. algorithms) rather than inventing new techniques

For those organizations that develop Big Data analytic solutions there is the potential to realize unprecedented business value but you need new skills to be able to capitalize on the shifts. You 'll need data scientists who are adept at visualizing large data and discerning between signal and noise.

These data scientists will need to have deep math and computer science skills. They 'll also have to be open minded and willing to try new things and fail quickly. They 'll need the ability to quickly incorporate learning's from previous failures to try new models and techniques. And most important, they 'll need to understand your business or have the listening skills to be able to learn from those in your business that knows your business and industry well.

Let 's take a look at some of those critical skills and how that relates to successfully and effectively deploying Big Data analytics. Let 's take a look at some of those critical skills and how that relates to successfully and effectively deploying Big Data analytics.

### 5.2.1 The Last Mile in Data Analysis

You cannot chat with Dr. Usama Fayyad for very long without the term "last mile" coming up in the conversation. Here 's what he means, in his own words:

Let 's say you process the data, you run the best algorithms on it using the best infrastructure and you do all sorts of acrobatics. Now you face the most diffi cult question: How do you use the results of your efforts to deliver value to the business unit?

The "last mile" is a group of people who are basically there to deliver the results of the analysis and put them in terms that the business can understand. This "last mile" group is made up of data analysts who know enough about the business to present to the CMO or the CEO. You need experienced data analysts who aren 't afraid to get their hands dirty.

It's not easy to find and hire those people. And in many companies, they are quickly turned into tactical resources. That's a mistake, because these are people who can help develop and guide strategy, move the needle, and grapple with big issues.

At Yahoo!, we created a lot of controversy. People accused us of being wasteful. They said, "How can you just sit there and just think strategy?" But we created some of the biggest value drivers for the company by addressing these big needles that never get addressed if you go about them tactically.

A very quick example that I can talk about publicly is Yahoo! Mail, which is used by 2501 million people or more, worldwide. One of the things you will see when you log in is a news preview module. The news preview module became important because it helped us retain active users. People would sign up for a Yahoo! Mail account, but then they wouldn't use it.

When we analyzed the data, we noticed something that I couldn't explain. The data was saying that new users like to read news when they read email. Don't ask me why. I can't explain it. By adding a news preview window to the new users of Yahoo! Mail, we were able to increase the return rate by 40%. And it turned out that in addition to new users, the entire population of users liked reading news while looking at their email. As a result, the news preview became one of the main features of Yahoo! Mail. To this day, I still can't explain why. But we know statistically it holds, and it made a huge difference to the business.

Another story I'll share with you is how we built the behavioral targeting business at Yahoo! This business was designed for targeting ads to users on what we believe are their interests as opposed to targeting ads the traditional way: based on the context of the pages the ads appear in. The power here is that pages where context is meaningful to ads is a small minority of page views (think about reading email or news). When I joined, it was a \$20 million business. When I left, it was a half a billion dollar business, with essentially the same inventory and almost zero investment.

Here's what it was: We had something like 2,500 categories of interest per user. We were very proud of it. For any user, I can tell you 2,500 dimensions of what they're interested in. Nobody took the time to look at it very carefully, other than me. I asked, "Okay guys, where do these categories come from? How do they die away? How do they get managed? How confident are we that we're computing them correctly?"

One of the biggest controversial things I did was to reduce the categories from 2,500 to 300. People cried foul. They said, "How could you do this to us? You're the data guy." I basically said, "Guys, I can reduce the number of categories and guarantee the quality is up to date, and make sure we're using the right algorithms. For example, after you buy a digital camera, you are no longer interested in seeing ads for digital cameras. You're done, you've already bought the camera, and you won't become interested again for another six months to a year."

But algorithms are going to continue targeting the hell out of you with digital camera ads because you showed us signs that you were interested in digital cameras—even though the truth is that you won't be looking for cameras again for the next 6 to 12 months. If you buy a car, you probably won't be looking at cars for the next two years, so targeting you with car ads during that time will be counterproductive. Shoes are different. Some people will wait months before looking at shoes again, other people will continue shopping. The point is that your attributes aren't fixed, they change over time. Data analysts are the people who figure these things out. They help companies save money and make money.

In one sense, what he's saying echoes Georges Clemenceau's famous phrase "War is too important to be left to the generals." But what we really believe he's saying is that data should be perceived and treated as a critical asset, not as a cost of doing business. Data analytics—and in particular Big Data analytics—should be elevated to the senior management level and become the responsibility of C-level executives. When that happens, data analytics will have traveled the "last" and most important mile.

Now that we have a working definition for Big Data, let's define analytics a bit and then relate both back to our examples in marketing. Analytics is another broad term that is used to define an entire spectrum of computations—see Figure 5.1 for examples of analytics in various categories along the spectrum.

Many of these techniques started in statistics and were used in business intelligence dashboards and scorecards. BI helped us aggregate information about events that took place in the past. As valuable as BI is, it essentially provides us with a look in the rearview mirror. The colloquial definition of analytics is using past information to provide insights into the future, which essentially is like using the bright lights in the dark to look out ahead and anticipate what's coming down the road.

Now, if we combine massive amounts of data (a.k.a. Big Data) along with powerful analytics (a.k.a. advanced analytics) there is a greater deal of texture or context to predict future outcomes, which allows us to then take it to the next level of using those insights to automate actions or to provide a short list of actions that are likely to result in significantly improved outcomes.

Not only can the Big Data collected by the iPad or Kindle be used to create lift in marketing campaigns and used to drive promotional activities to drive higher customer lifetime value but it can also be used to innovate products that will meet latent demand. Can you imagine actually creating a product that you know will be successfully adopted in the market before you spend \$1 on the research and development? Not only does that reduce your R&D costs but it also speeds your time to market.

### 5.3

### GEOSPATIAL INTELLIGENCE WILL MAKE YOUR LIFE BETTER

At this point, it's fair to say that we have assembled an all-star team of experts to share their thoughts on the most important and most relevant trends in Big Data analytics.

Even on an all-star team there are standout performers. Jeff Jonas certainly qualifies as a standout among all-stars, and we're gratified that he made the time to chat with us about a topic that's close to his heart. But first, here's some background about Jeff.

Jonas is an IBM Fellow and Chief Scientist of the IBM Entity Analytics Group. The IBM Entity Analytics Group was formed based on technologies developed by Systems Research & Development (SRD), founded by Jonas in 1984, and acquired by IBM in January, 2005. Prior to the acquisition, Jonas lead SRD through the design and development of a number of unique systems including technology used by the Las Vegas gaming industry to make cheating more difficult. You've probably seen

Prior to the acquisition, Jonas lead SRD through the design and development of a number of unique systems including technology used by the Las Vegas gaming industry to make cheating more difficult. You have probably seen Jonas' work featured in documentaries on the Discovery Channel and other networks.

Following an investment in 2001 by In-Q-Tel, the venture capital arm of the CIA, SRD began playing a role in America's national security and counterterrorism mission.

One such contribution includes an analysis of the connections between the individual 9/11 terrorists. That analysis is now taught in universities and has been widely cited by think tanks and the media.

Jonas is a member of the Markle Foundation Task Force on National Security in the Information Age, a Board Member of the U.S. Geospatial Intelligence Foundation (USGIF), on the EPIC Advisory Board, on the Privacy International Advisory Board, a Senior Associate at the Center for Strategic and International Studies (CSIS), and a Distinguished Engineer of Information Systems (adjunct) at Singapore Management University. He periodically testifies on privacy and counterterrorism in such venues as the Department of Homeland Security's Data Privacy and Integrity Advisory Committee, and other federally convened commissions.

Jonas was briefly a quadriplegic in 1988 following a car accident. Today, he competes in Ironman triathlons around the world. As you can see, Jonas' legendary status extends beyond the world of business data. Here's a summary of what he shared with us in a recent conversation:

More and more sensors are coming online generating higher and higher quality data about where things are when. As a result, there's a new form of analytics that is right now just beginning to mature. Those new analytics take advantage of data about where things are when (and how things move), and use that data for making extraordinarily high quality predictions.

The combined impact of those new analytics creates what Jeff describes as "geospatial intelligence." Simply put, geospatial intelligence is about using data about space and time to improve the quality of predictive analysis.

You pull up your Smartphone and you look at traffic and it shows you the streets highlighted in red and yellow. It's taking anonymized and aggregated motion data from people's phones and looking at their average speed of travel. Then it's helping you to avoid traffic, plain and simple. That's geospatial analytics today.

But this form of analytics is evolving very fast, says Jonas. Not too long from now, geospatial analytics will be considered a standard part of modern life:

Imagine searching for a drycleaner and getting a recommendation that isn't the closest drycleaner to your house, or even the closest to where you are right now, or the

most popular drycleaner by user ratings. Imagine that instead of the typical search results, you get the name and address of a drycleaner that is on the precise route that you travel most frequently, and it's on the correct side of the road. And if there are two drycleaners nearby, it's going to tell you which one tends to be the least busy at the predicted time of your arrival. That's the promise of geospatial intelligence.

For advertisers, geospatial intelligence will mean the ads we see will feel less like spam and more like relevant information.

It's going to come from weaving together data that has traditionally not been woven together. It will combine stuff about you on Facebook with comments you've made recently on Twitter. When you weave all of that data together, the quality of the predictions get better and better.

For example, Facebook knows I'm a triathlete. So I see an ad that says something like, "You're a triathlete. Do you want abs like this?" and it shows me a picture of some dude's perfect abs, and of course I want abs like that, so I might click on the ad.

But most likely, he will skip it. Now Jeff describes a scenario that's more likely to generate a positive outcome for the advertiser. Let's say that he's attending a business meeting in another part of the country, and he sees an ad for a triathlon coach near the hotel where he's staying. That ad is highly likely to grab his attention. What's the difference between the two ads? The ad for the triathlon coach is generated by a system that uses geospatial data and the ad for the great abs isn't.

From Jonas' perspective, geospatial data represents ultimate truth— it's based on what's really happening, where it's happening, and at the very moment that it's happening. Data doesn't get more real than that! Here's a brief story that Jeff tells to illustrate the point:

Imagine two ladies sitting in front of you. They're identical twins. But they both say, "We're one person." Of course, you're sitting there with them and you see two separate individuals. But they are adamant about being one person. And the data—at least most of it—is on their side. Their passports will show the same photo. Their fingerprints, their voiceprints, their irises, their DNA—all of the standard biometric tests—will support their version of the truth. Even their mother swears they are one person. But you're sitting there and you know they are lying, no matter what the data tells you.

By now, says Jonas, you're starting to feel a little confused. You might think, "Maybe they are one person, and I'm going crazy . . ." But then you remember there's a sure way to prove they are two people. Hand each of the twins a mobile phone with a GPS app. Each phone will show a slightly different location. Case closed. That's why proponents of geospatial intelligence call it the ultimate truth, the realest form of reality. For competitive organizations, this isn't mere sophistry. This represents incredible value. Let's end this section of the book with Jeff's take on the future of predictive analytics:

Real-time streaming analytics and geospatial intelligence will be essential for competitive businesses. When you make higher quality decisions faster and more consistently than your competitors, you win. You are more efficient, you deliver your products and services at lower cost, and your customers are happier. Jeff is an inspiration on many levels, and we are honored to include his perspective in our book.

#### 5.4

#### CONSUMPTION OF ANALYTICS

The Communication Cycle provides a framework and process for making analytics consumable throughout an organization. Each of the stages in the cycle is key to enterprise synthesis of the analytics. Let's take a closer look at each of these stages :

1. **Communication.** Successful consumption of analytics is a collaborative endeavor. The first step in this process is to take your analytics intent beyond your core team and sell it to a wider group of decision makers—the prospective daily consumers of analytics in your organization. The current economic scenario gives you a compelling storyline and helps you create a convincing platform to evangelize analytics in your organization
2. **Implement.** Organizations that successfully consume analytics are driven by leadership, which builds consensus in the organization and allows for moving ahead without the need to have everyone on board every step of the way. Strong leadership has been found to be the most important trigger in the wider analytics adoption in organizations. The initial focus of implementation should be on getting all the right ingredients in place to create the basic human and technology infrastructures to help you pluck the low hanging fruit.
3. **Measure.** The true test of consumption is to use analytics to measure itself. However, the benefits need not always be translated in hard numbers. One often unanticipated

but profound change in organizations is the maturing of a culture of objective debates, arguments, and viewpoints driven by data and not just “gut feel.” Be forewarned against overestimating the impact of analytics by discounting the human element in it. A successful business decision is a healthy combination of business experience and analytics—both merit equal credit for successful analytics consumption.

- 4 **Align incentives.** Successful consumption of analytics mandates creation of more structured decision-making processes, which is driven by data and analysis. This puts constraints on free-flowing, experience-driven decision making. The implementation will also bring in new stakeholders in your employees’ decisions as well as higher levels of oversight. Sometimes a general tendency of status quo bias exists, and employees do not want to venture out of their comfort zone. You need to create robust incentives to overcome these barriers. However, creating well-aligned incentives is just one piece of the puzzle and is not fully sufficient to overcome this bias.
- 5 **Develop cognitive repairs.** Our everyday decision making is influenced by numerous biases intrinsic to human nature. Presence of data and analysis challenges these biases and drives us out of our comfort zones. This results in undesirable conflicts and dysfunctional behaviors. The onus is on top leadership to make the decision makers in the organization aware of the presence of these biases. Creation of counterintuitive business insights based on data and then going and proving it right for all to see is by far the most effective to both expose biases and create repairs.

## 5.5

## FROM CREATION TO CONSUMPTION

Various organizations in services as well as product-based industries have flawlessly executed on creating analytics but have failed miserably on consumption. Creating analytics does not automatically result in institutionalizing analytics. Whether or not your organization suffers from this challenge will be determined by your answers to these key questions:

1. Do you have experience in creating a lot of analytics but failing at consumption?
2. Does it make sense to ramp up/down analytics creation to maintain balance with consumption?

3. Human bias exists. Do you need to develop structures that push people toward healthy conflict and resolution?

If the answer to any of these questions is affirmative, then your organization suffers from the creation-consumption gap. To put it another way, it is the difference between doing analytics and being analytical. While doing analytics focuses on creating analytics, being analytical balances and integrates creation of analytics with consumption of analytics. Organizations able to bridge this creation-consumption gap will be able to capitalize on analytics as a source of competitive advantage. Indeed, a key challenge that companies face is:

How to convert thought into action and bridge the gap between analytics creation and consumption?

#### **Visualizing: How to Make It Consumable?**

Visualizing data is a technique to facilitate the identification of patterns in data and presenting data to make it more consumable. Charts, graphs, and dashboards have been used for decades to synthesize data into a cohesive and comprehensible format for business analysts, managers, and executives. These techniques have been used to differentiate the contexts and intents of the data to be visualized. Intents such as:

1. **Describing.** Attempting to explain the thing being described, for basic meaning
2. **Reporting.** Summarizing findings from the past as of a point-in-time

However, as we move into the next stage of visualization, we move beyond these initial intents into the realm of:

1. **Observing.** Viewing data to identify significance or patterns which unfold over a period of time.
2. **Discovering.** Interacting with data to explore, interact, and understand relationships between data.

As the field of visualization matures, data artisans are using many different dimensions to represent and/or evaluate data. A sampling of such dimensions and attributes include:

1. Spatial, geospatial: position, direction, velocity

2. Temporal, periodicity: state, cycle, phase
3. Scale, granularity: weight, size, count
4. Relativity, proximity
5. Value, priority
6. Resources: energy, temperature, matter
7. Constraints

A static visual representation can never address multiple dimensions effectively, nor can it effectively show change over time. A series of static representations can only approximate change through periodic snapshots.

But looking at and observing data through visualization, even complex animations, isn't the same as interacting with it to uncover deeper meaning. It requires an effort to traverse and explore the data to uncover these various dimensions. Dr. Cox suggests that such mechanisms assist in our ability to explore and engage in wayfaring for discovery, interpretation, and deeper understanding.

Visualization provides a powerful means of making sense of data. By mapping data attributes to visual properties such as position, size, shape, and colour, visualization designers leverage perceptual skills to help users discern and interpret patterns within data. A single image, however, typically provides answers to, at best, a handful of questions. Instead, visual analysis typically progresses in an iterative process of view creation, exploration, and refinement. Meaningful analysis consists of repeated explorations as users develop insights about significant relationships, domain-specific contextual influences, and causal patterns.



Interactive data visualization and discovery tools, such as Tableau, are shifting insight discovery from a handful of specialists to everyone in the business. Ravi Bandaru, Nokia's Product Manager for Advanced Data Visualizations and Data Analytics, has been using Tableau since July 2010. Within his organization, he said about 350 to 400 people use Tableau – either in desktop or interactive form – and that it has brought people together. "It's letting the analyst do more analysis himself or herself without IT coming between them and their data," he said. "Using this kind of in-memory capability, I do see this being useful in exploring more complex and largish data sets, which were inaccessible before."

## 5.6

## DATA PRIVACY AND ETHICS

The current Big Data surge has predominantly been driven by a never-ending deluge of online clicks, queries, and paths: the artifacts of online activity. Debates have been going on for years and will continue as we work out the wrinkles between "rights" – the individual's rights to determine what personal information he or she is willing to "barter" in exchange for free services versus the service provider's rights to determine how to continue to provide free services.

It's a basic matter of quid pro quo – a mutual agreement of exchange of value. As with any relationship, there will always be squabbles around everyone's interpretation of said agreements.

In a personal one-on-one relationship, two parties mutually work out the details of an exchange and agreement. In the case of online (digital product/ service) agreements, the individual has only one immediate binary choice: accept the terms as stated or forego the service. There's generally no negotiation or setting conditional exchanges. Then again, this is not much different than product agreements in the past.

The fact of the matter is most people don't comprehend the computational power and possibilities that Big Data brings. Some think it would be impossible for someone to glean deep insight into their private lives. Big Data makes it harder to keep secrets. Jeff Jonas, IBM Chief Scientist, posted in his blog "Using Transparency as a Mask":

Unlike two decades ago, humans are now creating huge volumes of extraordinarily useful data as they self-annotate their relationships and yours, their photographs and yours, their thoughts and their thoughts about you ... and more. With more data comes better understanding and prediction. The convergence of data might reveal your "discreet" rendezvous or the fact that you are no longer on speaking terms with your best friend. Pity ... you thought that all of this information was secret.

## 5.6.1 The Privacy Landscape

There are four main constituents involved in the privacy landscape. Table shows how they are impacted.

Table : Privacy Landscape

Businesses	<ul style="list-style-type: none"> <li>■ Increased need to leverage personally identifiable and sensitive information for competitive advantage</li> <li>■ Significant investment in data sources and data analytics</li> </ul>
Criminals	<ul style="list-style-type: none"> <li>■ Dramatic surge in identity theft</li> <li>■ Sophisticated technology to exploit data security vulnerabilities</li> </ul>
Consumers	<ul style="list-style-type: none"> <li>■ Increased awareness and concern about collection, use, and disclosure of their personal information</li> </ul>
Legislators	<ul style="list-style-type: none"> <li>■ Responding to consumer concern by restricting access to and use of personal information</li> <li>■ Significant impact and restriction for business</li> </ul>

**5.7****THE GREAT DATA GRAB ISN'T NEW**

While there are critical differences between digital product agreements and their “analog” counterparts, there are historical similarities worth considering briefly. While data volumes are clearly greater today, the great data grab has been going on for decades. One primary means of past data exchange was via the warranty card that came with products. Under the guise of registering a product to ensure warranty rights, companies gathered personal information about individuals. These cards were not managed by the companies themselves but were outsourced to processing centers, which then had access to data across products. They gladly bundled and sold this data to other companies.

What began as list brokering evolved into the field of database marketing, which added analytics and attempted to address the differences between channels to better “target” exchanges with consumers. Over time, the marketplace of physical products was dwarfed by “services as products,” such as banking, telecommunications, and the like. These services industries relied even more on such data to not only offer the right products to consumers but to change their products to better meet consumer need.

For years, big companies paid millions to brokers to provide segmented lists (created using data mining scoring algorithms, assessing specific behavioral tendencies). The process was long and tedious: the turn cycle for new data was often six months or more.

Then in the mid-1990s, data warehousing made it feasible for such companies to "roll their own." MCI, which managed nearly 300 million individuals in their Friends & Family database, was able to bring their data in-house, on their own data warehouse for the same cost as their annual list broker (in this case, Epsilon) outsourcing fees—and they were able to update the data monthly rather than every six months. This meant that data going into the call centers was far more accurate (not to mention that in the process of the transition, it was discovered that one of the twenty algorithmic scores being used by the outsourcing vendor had a flaw in it and had been providing inaccurate results).

In the mid-1990s the sources of data that MCI leveraged went well beyond the lone warranty card. While they still used data from the list brokers they also used: updates from the U.S. Postal Service, "partner" relationships with other companies that offered affinity exchanges (airlines offering miles for minutes), and credit data (although the latter was dropped when new regulations were put in place that made it difficult to "use" the data). Unfortunately, neither their databases nor their operations were designed to truly leverage the data they could have gotten directly from their own consumers. They spent inordinate time and money using algorithms to "predict" behaviors rather than simply allowing their own customers to share their preferences.

## 5.8

## PREFERENCES, PERSONALIZATION, AND RELATIONSHIPS

As database marketing gave way to Customer Relationship Management (CRM) and one-to-one marketing, the focus shifted from segmentation to personalizing relationships: The more you knew about a consumer, the better you could meet their needs.

There are many reasons why we as consumers are perfectly willing to provide personal data (including specifying preferences) in doing business: Used responsibly, it can enhance our exchange. Indeed, we are often frustrated by dealing with companies that don't "know" us as individuals, that don't combine our information across products, across organizational boundaries, and across channels, and that don't remember anything about the last interaction we had with them.

When attempting to resolve issues, we tire quickly of having to repeat ourselves and re-create yet again the details of our relationship with the company. This is the flipside of the coin—companies not collecting "enough" information about us.

If we are doing business with a company, there is a relationship that we've inherently agreed to. But we also expect that companies will respect and not abuse the information we share with them. And we're particularly leery of companies asking us for personal information before we've decided that we want to do business with them at all (flash back to bad online site designs asking for "registration" before you've even seen what they're all about). That's like asking for a marriage license before the first date.

And yet, like with a marriage relationship we often feel like we've woken up next to a stranger—that the relationship we thought we had seems foreign to what it is we're now experiencing. As with relationships, this alienation is often due to three key factors:

1. Non-sustained or non-differentiating memory (not remembering what's important to an individual)
2. Lagging empathy (missing consideration for needs)
3. Inability to "listen" and process what's been heard (insufficient mechanisms for individuals to engage in conversations and see corresponding change/results)

These concerns go way beyond the growing industry focus on privacy. We raise them here just as a matter of setting a larger scope for the realm in which privacy policies fit and to reiterate that they're in support of something larger and more meaningful: a relationship. Privacy is the fallback for the "low bar to entry" in relationships—it's more about the "have to do" rather than the "need to do."

## 5.9

## RIGHTS AND RESPONSIBILITY

A consumer privacy bill of rights is a critical public policy proposal. I have long advocated for consumer privacy protections and will continue to push for legislation because I believe consumer privacy is a right, not a luxury. Above all else, it is absolutely crucial that, as we move forward in an evolving online and mobile world, consumer choice is protected and preserved. I will work to do just that.

While the Federal Trade Commission (FTC) has a set of online privacy guidelines that it enforces, online privacy is largely a matter of exchange between consumers and those they do business with—they're effectively self-regulated, although there is some

state regulation such as the California Office of Privacy Protection's requirement for an agency, person, or business that conducts business in California and owns or licenses computerized "personal information" to disclose any breach of security.

The truth is that most companies develop their own privacy policies as a matter of establishing a modicum of "trust" with consumers. There are several variations of seven principles outlined in the "EU-US Safe Harbor Principles," which most companies have engrained into their self-regulation for data privacy:

### Seven Global Privacy Principles

1. **Notice (Transparency):** Inform individuals about the purposes for which information is collected
2. **Choice:** Offer individuals the opportunity to choose (or opt out) whether and how personal information they provide is used or disclosed
3. **Consent:** Only disclose personal data information to third parties consistent with the principles of notice and choice
4. **Security:** Take responsible measures to protect personal information from loss, misuse, and unauthorized access, disclosure, alteration, and destruction
5. **Data Integrity:** Assure the reliability of personal information for its intended use and reasonable precautions and ensure information is accurate, complete, and current
6. **Access:** Provide individuals with access to personal information data about them
7. **Accountability:** A firm must be accountable for following the principles and must include mechanisms for assuring compliance Self-governance may soon change.

In February 2012, the White House announced an initiative to protect privacy under the "Consumer Privacy Bill of Rights." To be initiated by the Commerce Department, and developed in conjunction with Internet companies and consumer advocate groups, it will be enforced by the FTC:

American Internet users should have the right to control personal information about themselves. Based on globally accepted privacy principles originally developed in the United States, the Consumer Privacy Bill of Rights is a comprehensive statement of the rights

consumers should expect and the obligations to which companies handling personal data should commit. These rights include the right to control how personal data is used, the right to avoid having information collected in one context and then used for an unrelated purpose, the right to have information held securely, and the right to know who is accountable for the use or misuse of an individual's personal data.

That said, this is still a voluntary act, not legislation. According to CNN Money, "The Administration supports Federal legislation that adopts the principles of the Consumer Privacy Bill of Rights. Even without legislation, the Administration is convening multi-stakeholder processes that use these rights as a template for codes of conduct that are enforceable by the Federal Trade Commission.

Other than HIPAA and GLB, the only real legislative action that can be taken (in 2012) is where the FTC steps in to require a company to honor its stated privacy policy.

Attention to privacy policies took a spike in March 2012 when Google announced plans to merge their sixty different privacy policies (separate ones for various services) into one. What appeared to raise the most concern was not so much the merging of the policies but the implied merging of the data associated with the different services.

This rang vaguely familiar to a point in history over a decade earlier when Doubleclick (which had a lot of click stream data) bought Abacus Direct (a company that knew a lot about our offline buying habits). The potential for merging such data got the attention of watchdog groups such as the Electronic Privacy Information Center, which then got the FTC involved ending in a settlement.

## **IMPORTANT QUESTIONS**

---

1. Discuss briefly about big data emerging technologies.
2. What is business analytics? Explain about the last mile in data analysis.
3. Explain about geospatial intelligence will make your life better.
4. Discuss briefly about consumption of analytics and from creation to consumption
5. What is data privacy and ethics? Explain about the privacy landscape and the great data grab isn't new.
6. Explain about preferences, personalization, and relationships in data privacy and ethics.
7. What are the rights and responsibility of data privacy and ethics.