



The 5 Myths of Advanced Analytics

“Only 4% of enterprises are able to attribute better decision-making to the use of big data & analytics.”

— Bain & Co

Advanced Analytics refers to the practice of extracting information from existing datasets to identify patterns and predict future outcomes and trends. According to Gartner, Advanced Analytics is becoming a key strategy for enterprises that want to take advantage of the massive insights buried in their data marts, data warehouses, Apache™ Hadoop® lakes, and spreadsheets. But, despite the millions of dollars invested in analytics technologies, the majority of companies still struggle to establish an efficient and programmatic way to do analytics at scale. According to Bain & Co., only 4% of enterprises have been able to attribute better decision-making to the use of analytics. Why are these investments failing to meet expectations? In this paper, we delve into today's most common Advanced Analytics myths and offer potential solutions.

MYTH #1: DATA SCIENTISTS ARE DATA JANITORS

A typical enterprise has 6-10 databases and thousands of employees with great ideas on how to leverage their data. However, at least 70% of this information is siloed and difficult for advanced analytics teams to discover and work with. It's been an accepted requirement that 80% of a data scientist's work is spent on data acquisition, movement, and transformation. These are the most tedious steps that an advanced analytics team must go through in order to start extracting value from their organization's data.

The issue with “data janitor” work is that it’s an extremely linear and poorly integrated process. Data teams traditionally rely on data engineers to extract data. The extracted data is then moved into a new location, at which point data scientists use their tools and methods to clean and format the data in preparation for analysis. Finally, the data scientists present the results to business analysts who use yet another set of tools to transform, blend, and analyze the outputs. This lack of integration between people, processes, and tools results in a terribly inefficient approach to analytics, requiring constant back and forth communication, delays, and errors.

Modern approaches have transformed data logistics, transformation, and enrichment into a highly collaborative and integrated process. New solutions enable engineers, scientists, and analysts to engage early in the process so that each role can iteratively determine what data needs to be cleaned and analyzed, as well as which role along the analytics continuum is most appropriate for each task.

“Data scientists spend up to 80% of their time mired in the mundane labor of collecting and preparing unruly data, before it can be explored for useful nuggets.”

— New York Times

TRUTH #1

Data scientists don’t have to be isolated data janitors. They are constantly forced to do janitor work today because legacy technologies require them to sample data and package it into formats that are accessible by the subsequent groups in the traditional data pipeline. Modern approaches to data science include all roles (engineers, scientists, and analysts) earlier in the data pipeline, each participating in the process of cleaning, formatting, and enriching data. With various roles collaborating on the most time-consuming parts of the data analytics process, like data blending and processing, businesses can ramp up the productivity of their analytics teams.

MYTH #2: SOFTWARE WILL REPLACE DATA SCIENTISTS

Data scientists have been at the center of a controversial debate over the past few years. First, Google’s Chief Economist, Hal Varian, called the job the “sexiest job on earth” in a New York Times interview. Second, McKinsey released a report predicting that the US would face a shortage of 140,000 to 190,000 people with deep analytical skills.

Soon enough, software vendors leveraged this industry buzz to turn against data scientists and declared that their tools were so simple to use that “data scientists would not be needed anymore.” The fate of data scientists is not that simple, however. Data scientists are the foundation of the predictive enterprise, and companies that want to win with data shouldn’t compromise their data science values.

Although industry studies have shown that true data scientists are rare, no one ever suggested that organizations should use data scientists alone as their analytical backbones. For each data scientist, there are another 10 or 100 business analysts involved in the data process. And for each business analyst team, there is another set of executives driven to use data to make better decisions.

“Embedded analytics and solutions provide better scaling of data science resources, but it doesn’t remove the need for data scientists.”

– Gartner

TRUTH #2

Data scientists are not going away. They are the nucleus of the predictive enterprise and should be supported by tools, technologies, and practices that help them scale their knowledge and talent. Over the next 10 years, we predict that data scientists will be supported by technological innovations as well as the natural evolution of our data-driven culture. Here are some interesting trends:

MORE DATA-NATIVES

Data science is now available in more schools for undergraduates and soon will be made broadly available to younger generations. Similar to chemistry, Data Science is built on math, data interpretation, and experimentation.

MORE DECISIONS VIA AUTOMATION

In his book, *Think Twice*, Michael J. Mauboussin makes a great case for the compartmentalization of decision-making. Some decision tasks are better operated by humans. Some are better automated via machines. As humanity surrounds itself with more intelligent devices, we will grow more comfortable with empowering machines to make decisions on our behalf.

MORE PEOPLE, MORE DECISIONS

In his Strata presentation, Steven Hillion, former chief product office of Alpine Data (now part of TIBCO Software), presented a controversial view of the data science professional makeup. For example, his data indicated that women will take a far more prominent role in the world of data science over the next 10 years, which is great for the data field.

Predicting the “death of the data scientist” is misguided. They are important today, and they will be important tomorrow.

MYTH #3: MOVING DATA IS A REQUIREMENT TO DO ADVANCED ANALYTICS

Historically, companies have provisioned data into siloed analytics environments that were separated from their central analytics environments. This practice created multiple copies of the same information across the organization. The proliferation of new data silos across the enterprise brought along delays and errors, amplifying the already existing challenge of working with big data.

It also forced teams to work with sample datasets because the entirety of the data they wanted to analyze, although available, was often too large to move into their analytics environment. Leveraging subsets of data ended up limiting companies’ perspectives and the utility of their insights.

The introduction of new technological paradigms allows teams to work with data at scale, without ever having to move it or compromise the depth of insight because of data size.

TRUTH #3

Moving data into siloed analytics environments is a thing of the past. It introduces delays, errors, and potential security issues in the data pipeline. The companies that succeed with big data have created an open and secure environment that allows users to send instructions to the data where it sits. A key technological innovation in this field is “in-cluster processing.” Instead of pulling the data into separate analytics environments, in-cluster processing allows for instructions to be pushed down into databases or Hadoop clusters. This capability allows data analysts and scientists to run algorithms at scale without having to worry about moving data or optimizing their work based on the data source that their data sits in.

“70 to 80% of the world’s data is trapped in silos with no secure and reliable way to access it.”

“The desired business impact must drive an integrated approach to data sourcing, model building, and organizational transformation.”

— Gartner

MYTH #4: YOU CAN’T SHARE & PROTECT DATA AT THE SAME TIME

Data is arguably the most valuable asset that the modern enterprise has, regardless of the industry. Data must be guarded and treated with the highest level of security. Typically, database administrators and big data leads are charged with the difficult tasks of making data accessible within their enterprise, while making sure this access adheres to their company’s governance and security policies.

But in order to be successful, analytics teams need fast and easy access to their organization’s data. Provisioning data to analytics teams has become a daunting and time-consuming task, resulting in constant back-and-forth communication for data access. The consequence is a reduction in the organization’s speed of discovery and execution.

TRUTH #4

Enterprise leaders have learned to build processes and deploy tools that enable fluid communication between data engineers, data scientists, analysts, and line-of-business leaders. Business employees should be able to search and explore datasets immediately. Items such as data, people, models, and projects should be at the fingertips of each employee to promote creativity and optimize business practices.

MYTH #5: ADVANCED ANALYTICS IS JUST ABOUT THE ALGORITHMS

When investing in Advanced Analytics solutions, organizations are primarily focused on the quality and complexity of their algorithms. While this is certainly an important and necessary step, neglecting the other aspects of creating a data-driven culture has led to poor insight adoption.

The ways in which data is consumed and implemented across the enterprise often dictates the success or failure of an analytics initiative. Here are three key questions that every advanced analytics project leader should address when planning analytics initiatives:

- What and who influences decision-making within your business?
- What sort of data analytics would be required to influence each decision area?
- Who needs to be involved in each phase of the analytical process for the most fluid implementation?

TRUTH #5

While adhering to data governance and security policies, business leaders need to shift their organization’s focus towards sharing data. Analytical insights need to be made transparent throughout each area of an organization, from the most senior executive to the most junior associate. Moreover, analytical insights should be embedded into tools through a visual and collaborative interface, so they can be digested by decision-making business executives.

The companies that drive sustainable change understand that while they must take advantage of data by ingesting large amounts of it, they must use advanced analytics to filter the noise and surface only the insights that are relevant at decision time. Bottom line: it’s not just about the algorithm. It’s about getting the power of the algorithm into the hands of more people, at the point of decision.