



The Democratization of Analytics: Are we all turning into Data Scientists?

June 2015

Email: info@damoconsulting.net
Website: www.damoconsulting.net

Executive Summary

- We are entering an era of democratization of analytics driven by the acceleration of technological innovation and the open source movement, among other factors.
- Analytics is becoming pervasive as complex analytical tools allow businesses to collect and process vast amounts of diverse data.
- Manual processes are being substantially replaced by automated processes. New tools that perform activities that were not possible earlier are now widely available.
- Advanced analytics and machine learning have shifted from CIOs and specialized units and are becoming pervasive across businesses. The envelope will get pushed even further in the coming year and advanced analytics capabilities will be available to individual users at a fraction of today's costs.
- Organizations still need experts to carefully review the analyses. The democratization of analytics can help leaders make better decisions by enabling more people across the enterprise to leverage the complete range of tools at their disposal.

What is Democratization of Analytics?

Analytics is cropping up everywhere across industries, organizational levels and even in our homes and personal lives. No longer are complex statistical analyses the realm of PhD statisticians, economists, engineers and other technical professionals. Today, you can find analytics used by everyone from the CEO to your local bartender. Analytic tactics are becoming pervasive as businesses collect and analyze vast amounts of data. Increasingly, more job responsibilities involve using analytics. Why is this happening and why is it important?

In recent years, expensive and complex analytical tools and technology have become accessible and affordable to people and business. For many years, data sets were so large and complex that it was all but impossible except for a select few to have access to the data and the computational horsepower needed to conduct analysis. Storage and computing capacity were too expensive to make big data projects viable. As innovations in data storage, processing and technology business models evolved, the ability for increasing numbers of businesses and individuals to analyze large volumes of data has steadily increased.

Even a year or two ago, analysis of such data required one to have the skills of a statistician, computer programmer and business analyst.

Today, we have companies that offer entire technology stacks that claim to deliver all the capabilities for end-to-end solutions. As a result, we are able to measure, collect and analyze all sorts of data using simple-to-use analytical tools to make business, policy and personal decisions. This process of spreading the availability of once complex information and technology to broader audiences is known as democratization.



In this paper, we propose that we are entering an era of democratization of analytics, driven by the following factors:

- Increasing importance of analytics for enterprise growth and profitability
- Falling prices for data storage and computing capacity
- Acceleration of technology innovation
- An unsustainable bubble in the data scientist talent market

The emergence of high-end cognitive computing capabilities for population health, best exemplified by IBM's Watson platform, promises to take principles of machine learning and artificial intelligence (AI) much further into corporate environments and democratize the process of developing and disseminating analytical insights to a wide range of users that goes beyond scientists and researchers!

The Talent Crunch and the Pursuit of Unicorns

In 2013, The McKinsey Global Institute (MGI), published a report² that by 2018 the United States will experience a shortage of 190,000 skilled data scientists, and 1.5 million managers and analysts capable of reaping actionable insights from the big data deluge. The term "data scientist", coined circa 2012, referred to individuals with a special combination of skills in technology, data management, and statistical modeling. This was a cue for many enterprises to go out and find these "unicorns" in as many numbers as they could, in anticipation of the coming talent crunch.

Competition for talent is likely to be robust for people with experience in healthcare data and analytics. A recent study³ mentions that 37% of respondents indicated a lack of qualified staff as a factor in adoption rates for analytics. Another study⁴ highlights some nuances to this talent market:

- Data scientists have a median of only six years of experience, but are highly educated (92 percent have at least a master's degree. 48 percent have a Ph.D.).
- Over one-third are employed on the West Coast (36 percent) and almost half work for firms in the technology and gaming industries (43 percent) making it difficult for firms in other locations and industries to find qualified data scientists.

There is a giant bottleneck today in the efforts to increase adoption rates for analytics, as one of the authors of this paper pointed out in an earlier publication⁵. In addition, it appears that younger workers and recent college grads prefer to work in smaller organizations that provide more challenging data analytics problems to solve⁶.

It may well be that the scarce pool of highly talented data scientists is being sucked up by hot Silicon Valley start-ups funded by deep-pocketed VC's. Since the data scientist pool cannot scale quickly or adequately to meet demand, we can assume that these data scientists are presumably contributing to the development of analytics platforms that will provide scalable options for enterprises and reduce the need for large teams of very expensive data scientists.



At the same time, large enterprises have to develop creative solutions such as a "team approach" that tries to fulfill a data scientist's job requirement with two or three \$ 80,000 individuals instead of one \$ 250,000 rock star. There is a growing movement to outsource analytics services to India-based firms that have seized the opportunity to address this talent gap by grooming college grads to become data scientists by setting up in-house academies that teach them the basics of statistics, data, and technology.

However, such efforts will be limited by data privacy laws that restrict access to protected information in sectors such as healthcare. The solution lies in emerging and innovative technologies.

In our view, the big trend that will ultimately democratize advanced analytics is the emergence of technology and machine-learning solutions that are being packaged into out-of-the-box solutions, mainly cloud-based analytics.

There are many start-ups in Silicon Valley that are attempting to reduce the need for data scientists in enterprises by offering a complete solution, including pre-built machine learning algorithms in an end-to-end technology stack.

Microsoft Azure's machine learning platform is a cloud-based predictive analytics engine that comes in a freemium model with a basic free version that comes with detailed tutorials on how to select and deploy models. Amazon's machine learning as-a-service provides visualization tools and wizards that guide you through the process of creating machine learning (ML) models without having to learn complex ML algorithms and technology. Amazon has basically put all its internally developed algorithms on their storefront for anyone to deploy.

Some firms choose to focus on specific aspects of the analytics value chain. Alteryx focuses on data integration. Tableau has developed a best-in-class platform that provides extremely intuitive visualization capabilities and a set of pre-built templates for specific industries and use cases that can be easily deployed by non-technical users with little knowledge of programming or coding.

In May 2015, Google announced a price drop of 30% across the board for its cloud platform⁷. This was just one more step towards democratizing industrial-grade cloud computing capacity and removing constraints imposed by storage and computing capacity in the adoption of big data analytics.



The "Open" Movement – Open Source, Open Data, Open Courses, Open Research

An additional hallmark of the democratization of analytics has been the increase in accessibility to data, software, hardware, education and even research results for free or low cost. Today you can learn a programming language for free from a variety of online services and use free or inexpensive hardware and software for your analysis.

Open source software, such as R and Python, are rapidly becoming the de facto standard tools for analysis by many, particularly in the Big Data Analytics disciplines. Having the advantage of being both flexible and free they have become a popular option to more expensive enterprise software solutions. The software to run these tools can be downloaded easily from the Internet.

If you want to do some complex analysis on more data than your computer can hold, look no further than to cloud services offered by Amazon, Microsoft and Google to name a few. Need a cluster of 100 processors for your 500 terabytes of data? No problem. Just sign up for an account, upload your data, specify your needs and you are ready to go.

If you don't know how to do this, watch free videos on online services such as YouTube or perhaps take a more in-depth learning path using MOOCs (Massive Online Open Courses) such as those provided by Coursera. MOOCs offer courses that teach a wide variety of subjects including how to develop predictive models using R.

Academic medical institutions, such as the Mayo Clinic, make their research open to the larger community by publishing and sharing their algorithms with healthcare enterprises to improve patient outcomes in a range of areas such as cardiovascular health. The unlocking of this knowledge, hitherto residing within research departments of academic medical centers, could have significant impact on the widespread adoption of predictive modeling in healthcare if other medical institutions follow suit.

The emergence of predictive modeling markup language (PMML), an XML-based file format provides a way for applications to describe and exchange models produced by data mining and machine learning algorithms, makes it easier to operationalize analytics and integrate them into workflow and applications.



Cautionary Notes

As innovation and discovery accelerate, business and policy insights can be made available at a lower cost and in a more timely fashion. What's not to like?

There is more to conducting a sound analysis than being able to operate the mechanics. Just because the computer can make some calculations and no error appears does not mean that the results are meaningful. A good analysis requires the analyst to understand what the tool is doing, whether it is appropriate or not, its limits and how to frame the data and analysis properly.

At the same time a cautionary note for enterprises

would be that the results of automated analysis are not automatically valid. The need for expert human intervention will be greater than ever before.

Enterprises also have to be cautious about getting seduced by resumes that claim "data scientist" qualifications, especially when the hiring managers themselves do not understand clearly what the term means.

A final, and important, consideration for the rapid democratization of advanced analytics, especially in healthcare, is the need to manage user data responsibly and follow HIPAA regulations.

Conclusion

Analytics is finding its way into more of what we do at home and at work. In many ways, the evolution of technology is turning us all into data scientists. Wearable technologies provide us with an immediate look at our health habits and monitoring of trends, allowing us to make decisions and course corrections the right way. The technology is practically effortless. The same is true for businesses and government agencies. It is far easier to examine work processes and identify and fix problems early.

While technology and tools are undoubtedly

democratizing analytics and putting it in the hands of more and more individuals, the potential for the misinterpretation or misapplication of these tools exists.

Before we apply software-generated insights blindly we need to consider taking human expert advice. This will ensure that problems arising from incorrect or misleading insights, whether they are business or personal, do not become much larger and harder to fix. The democratization of analytics is here and it will continue. And we'll be all the better for it.



References:

1. How Fitbit and IBM Watson are transferring the power of healthcare data from one hand to another, [CIO Online](#) , 05/14/2015
2. [McKinsey Report](#) highlights impending data shortage, July 2013
3. [CDW Research](#) on Analytics in Healthcare, March 2015
4. The [Burtch Works study](#) : Salaries of Data Scientists, April 2015
5. Healthcare Analytics: [4 things impacting adoption rates](#), Paddy Padmanabhan, CIO Online, March 31, 2015
6. Despite choices, new data scientists still gravitate to small firms, Tom DiChristopher, [CNBC](#) , May 22, 2015
7. Google Cloud Platform [blog](#) : May 18, 2015

Authors

Paddy Padmanabhan



Paddy is CEO of Damo Consulting Inc., a management consulting firm that provides advanced analytics solutions to healthcare enterprises. Paddy is an accomplished business leader and entrepreneur with over 15 years of experience in the healthcare sector. He has held global leadership positions in top-tier consulting firms, led large global operations teams in India, Philippines, and South America. You can learn more about Paddy at <https://www.linkedin.com/in/paddyadmanabhan99> and follow him @paddyadmanabha

Larry Boyer



Larry Boyer is an analytics leader who enjoys working with the interplay of people, processes and technology. He is experienced in leading cross functional teams through the analytics life cycle. His 20 years of analytics experience spans industries and disciplines. He is an analytics translator who can bring people from across a business together to solve complex problems. You can learn more about Larry at <http://www.linkedin.com/in/LarryBoyer> and follow him @LarryBoyer.

Damo Consulting Inc is an advanced analytics, digital solutions and global sourcing advisory firm focused on the healthcare sector. We help healthcare enterprises successfully implement and manage big data and advanced analytics solutions, digital transformation programs, and global sourcing strategies.

USA

One Lincoln Center 18W140 Butterfield Road Oakbrook Terrace
Suite 1500 Oak Brook, Illinois, 60181

INDIA

603/11/A, 19th Main Road, 2nd Floor, Sector 2,
HSR Layout, Bangalore 560 102 (India)

Email: info@damoconsulting.net • website: www.damoconsulting.net

Copyright © 2015 Damo Consulting, All Rights Reserved

