



DATA SCIENCE

By Sarah Walker-Leptich

walkrinthecloud.com | sarah@walkrinthecloud.com

An interpretation of thoughts covered by authors of the Harvard Business Review series on data science.

CHAPTERS

- 01** The difference between human-driven vs AI-driven data decision making.

- 02** Data literacy for leaders starts with understanding five key points about data.

- 03** The most significant challenges of data scientist.

- 04** The importance of hiring statisticians and analysts.

- 05** The lost art of visual information storytelling.

- 06** A dangerous lesson in history of misinterpreting data visuals.

- 07** Closing the gap between algorithms and executives.

- 08** Building a world-class data science team.

- 09** The best practices for your first strategic analysis project with your data.

- 10** Ethical collection and use of employee data.

“Companies recognize that data is one of their most important assets, yet many are not equipped to effectively manage it.”

INTRODUCTION

Google believes that every company will eventually become a data company, and data has been termed the “gold” of this time. Most companies also recognize that data is one of their most important assets, yet many are not equipped to manage their data or put it to good use effectively. They lack the talent, they assign the wrong people to deal with quality, and their organization silos make sharing difficult.

By the end of 2020, 44 zettabytes made up the entire digital universe¹, and the current estimate of how much data is collected daily is 1.145 trillion megabytes. Many modern businesses are taking advantage of cloud computing regarding infrastructure modernization and processing large amounts of data more cost-effectively.

The companies most actively taking advantage of cloud computing are small businesses and gigantic enterprises. Enterprises typically have data lakes worth of data, and smaller companies are often startups who are cloud-native. Many of these companies responded to the analytics boom by hiring the best data scientists, but many of them haven’t gotten the value they expected from their data teams.

To understand why companies are not yet taking advantage of their data analytics or why companies who have invested in data science teams are not seeing a return in investment, this paper will dive deeper into the trends and challenges with data science and strategic analytics.

¹ <https://techjury.net/blog/how-much-data-is-created-every-day/#gref>

01 THE DIFFERENCE BETWEEN HUMAN-DRIVEN VS AI DRIVEN DECISION MAKING

Both humans and artificial intelligence (AI) process data, but each has different strengths and weaknesses in analyzing data. AI's most prominent strength over human processors is the ability to process large batches of data and determine an outcome efficiently. For example, AI is more successful at determining which creative ad will be more effective¹ or the optimal inventory levels during peak seasons.

Processing and remembering petabytes of data and coming to a single conclusion is impossible for humans. Data reduction is necessary, which is costly and often prone to cognitive biases.

The best way to process data is a hybrid approach between data-driven and AI-driven, in the form of data-supported decision making. Utilizing the power of cloud-enabled data processors like Google's BigQuery that sorts and analyzes data with built-in machine learning (ML) algorithms helps leaders make more informed decisions. Implementing a business intelligence tool like Looker or PowerBI provides clear visuals assisting non-technical people to make confident decisions using real-time analytics.

¹ <https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like>

02 DATA LITERACY FOR LEADERS STARTS WITH UNDERSTANDING DATA

According to Hugo Bowne-Anderson, a data science and marketing evangelist, understanding and communicating about data is an increasingly important skill. Bowne-Anderson touches on key critical points to data literacy in his article on the Harvard Business Review¹.

Understanding your data logistics: getting a clear idea of how data is collected and stored in your organization will help leaders understand what data you have, what you are currently utilizing, and what data you could make use of for future projects.

Different types of data: businesses collect two types of data: structured and unstructured data. Think of structured data as an Excel file. The opposite of structured data is unstructured data, mainly text, HTML, and rich media such as images or videos.

Statistical intuition and common pitfalls: Data analytics can tell you what is happening, but it rarely tells you why. To fully benefit from your data, you need to know what problems you can solve and, most importantly, what problems are worth solving.

Building algorithms with data for AI/ML: The main goal of data analytics should be training AI and ML algorithms to provide insight and automate admin tasks, freeing up time for your employees to work on more impactful projects.

Being conscious of ethics and privacy: An essential part of building a strategy around your data is ensuring that the data of customers, employees, and the business is protected and only used for ethical projects.

¹ <https://hbr.org/2018/10/your-data-literacy-depends-on-understanding-the-types-of-data-and-how-theyre-captured>

03 THE MOST SIGNIFICANT CHALLENGES OF DATA SCIENTIST

Companies have invested billions of dollars into hiring the most talented data scientist and implementing the latest analysis tools but still face significant challenges in providing actionable outcomes and meeting objectives. According to Scott Berinato in his article, “Data Science and The Art of Persuasion¹,” the biggest challenge of data scientists is communication and explaining the results in a meaningful way to non-technical stakeholders or executives.

Kaggle’s annual survey on ML and data science in 2017² captured similar barriers faced at work by data scientists, including lack of management buy-in or financial support, no defined questions to answer, decision-makers not utilizing results, and explaining data science to others.

“The biggest challenge of data scientist is communication and explaining the results in a meaningful way”

¹ <https://hbr.org/2019/01/data-science-and-the-art-of-persuasion>
² <https://www.kaggle.com/kaggle/kaggle-survey-2017>

*“If you only hire data scientist,
you will fail to deliver
meaningful projects”*

04 THE IMPORTANCE OF HIRING STATISTICIANS AND ANALYSTS

If you only hire the best data scientists and ML engineers you will fail to deliver meaningful projects and waste your experts time.

It is crucial to hire a data science team with multiple talents alongside top engineers, including analysts and statisticians. Analysts shape the data and summarize the facts, often identifying what problems to solve. Statisticians then poke holes at the data and discover which issues are worth solving. Adding talented statisticians and analysts to your team provides clarity and direction, ensuring your engineers are not wasting their time on projects that will suffer low returns.

05 THE LOST ART OF VISUAL INFORMATION STORYTELLING

More than a century ago, Willard Brinton published *Graphic Methods for Presenting Facts*¹ and wrote, “time after time it happens that some ignorant or presumptions member of a committee of a board of directors will upset the carefully thought-out plan of a man who knows the facts, simply because the man with the facts cannot present his facts ready enough to overcome the opposition. As the cathedral is to its foundation, so is an effective presentation of data to data”.

Over the years, the advancements in PowerPoint and Excel have considerably changed how we communicate visual data by

automating beautiful charts and graphs. We now see another significant change in data graphics with business intelligence tools. With the help of marketing, charts and graphs are eloquently placed into slides with animations or on internal websites for everyone to access.

This automated advancement removed the need for the technical draftsman, one whose main task is mastering the visual display of quantitative information.

However, there is a risk associated with removing this key element when presenting data findings. Edward Tufte, who has been named

the “Leonardo de Vinci of Data²” by the New York Times, stated that “allowing artist-illustrators to control the design and content of statistical graphics is almost like allowing typographers to control the content, style, and editing of prose.”

Tufte meant that unless you fully understand the data from a technical standpoint, graphic designers might miss what the data is telling them, visually suggest their own biases, and not highlight the critical data points.

¹ https://www.google.ca/books/edition/Graphic_methods_for_presenting_facts/9JHO1yc5ynQC?hl=en&gbpv=1&pg=PR1&printsec=frontcover

² <https://www.nytimes.com/1998/03/30/business/the-da-vinci-of-data.html>

06 A DANGEROUS LESSON IN HISTORY OF MISINTERPRETING DATA VISUALS

A heartbreaking yet crucial reminder of executives misinterpreting visually displayed data as technical data facts is the tragic story of NASA's 2003 spaceflight, Columbia. Directly after the Columbia's shuttle's liftoff, a piece of foam insulation broke off, hit the left-wing, and broke through the wing's thermal protection.

At the time, NASA's engineers and executives were unaware of the extent of the damage. Multiple teams started to analyze the data they had from the incident, preparing 3 reports with a total of 28 PowerPoint slides dealing with the debris impact. The slides had different styles of text hierarchy, large bullet points with small bullet points beneath, which made the report confusing and ultimately led to a decision that had a tragic outcome.

Edward Tufte was one of the leads asked to present to the Columbia Accident Investigation Board. His findings can be found in his article, "PowerPoint Does Rocket Science — and Better Techniques for Technical Reports¹." One of the main arguments he uncovered was how PowerPoint's cognitive style affects the quality of engineering analysis and how design can hinder or hide the crucial facts.

*“Power Points
cognitive style could
affect the quality of
engineering analysis
and hide crucial facts”*

¹ https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001yB

“ Business leaders expect data scientists to be unicorns; one who can wrangle data, analyze it, and present a compelling story”

07 CLOSING THE GAP BETWEEN ALGORITHMS AND EXECUTIVES

Business leaders expect a data scientist to be a unicorn. Someone who wrangles the data, analyses it, knows the business and its strategy, can design charts (on brand), and present a compelling story. To have this level of talent in one single person is extremely rare. Most situations are not as critical as spaceflight, but it could be why your executive team misses crucial information, including potentially catastrophic cyber breaches.

One of the most challenging positions to fill, and yet mission-critical, is a liaison between the technical team and the executive team. Someone who is both skilled at information design and presenting the potential return on investment to executives. This person needs to grasp data science principles, understand effective statistical design (not company branding), and be well versed in delivering rhetoric communication.

08 BUILDING A WORLD-CLASS DATA TEAM

Understand the strengths of your current team and note the gaps that you need to fill. Scott Berinato¹ recommends doing an internal talent audit. Performing a talent audit helps managers do a better job of planning for projects, configuring teams, and identifying what positions to hire. Berinato breaks down a team into six different categories; some employees may have more than one of these qualities:

Project Management: Highly organized with great people management skills. Knowledge of project management like Agile or Scrum.

Data Wrangling: Expert in coding and systems architectures with a passion for finding, cleaning, and structuring data.

Data Analytics: Advanced critical thinking and experience with statistics. Must also understand the business's strategy and communicate effectively.

Subject Expertise: High level of functional knowledge and strategy development to identify project goals. Ability to communicate to non-technical stakeholders.

Design: Attention to detail and expertise in information design and presentation design.

Storytelling: Skilled in persuasive storytelling with a background in information design and strong writing skills.

Even those who are not skilled in certain areas should have a general knowledge of each skill set because it creates an appreciation of what each other brings to the table as a team.

“There should be at least six crucial positions on every data science team.”

¹ <https://www.linkedin.com/in/scott-berinato-6330ba54>

09 BEST PRACTICES FOR YOUR FIRST STRATEGIC ANALYTICS PROJECT

First, focus on the right internal customer by narrowing your efforts to the departments that will have the most impact.

Second, identify high-impact problems — special projects with the highest-value outcomes with a defined problem and a large volume of quality data.

Third, place a deadline on execution. The deadline can be flexible, but it's essential to give the team the right to cancel if they discover it is not yielding results.

Quick failures will lead to more successes because you won't have a team focusing long-term on a project that will be unsuccessful.

Lastly, consider other factors: who will be the project's executive support? Does the project need outside expertise for a fresh set of eyes, such as a consultant?

10 ETHICAL COLLECTION AND USE OF EMPLOYEE DATA

Many employees find their data insights from collaboration tools such as Microsoft O365 and Google Workspace beneficial. The analytics provided to users on how many hours they are in meetings, whom they meet with most frequently, and knowing their most productive hours can help employees manage their time better.

There is also a general understanding that if an employee can see that level of insight into their daily productivity, there is a high chance their managers can also have insight into it, which can pose some trust issues between employees and employers.

If a company plans to use the analytics insights in productivity, they should avoid making employees feel they are being watched for mistakes. Insights should be used responsibly by focusing on motivation and promotion rather than punishment. It is also important to consider allowing employees to opt-out of the data collection and manage their data, including deleting it.

“If a company plans to use employee data for insight, they should avoid making employees feel like they are being watched.”

CONCLUSION

In conclusion, the Harvard Business Review on Strategic Analytics was a captivating read. When building an analytics strategy, essential takeaways are understanding the different analytical skills of your team and communicating effectively to non-technical stakeholders. However, more importantly, always be respectful of the privacy of users' data and consider the ethical limitations of each project.

ABOUT THE AUTHOR



Sarah Walker-Leptich

www.walkrinthecloud.com
sarah@walkrinthecloud.com



I'm Sarah Walker-Leptich – and I'm obsessed with the business impacts of data analytics, blockchain and AI & ML.

For 10 years, I've been a technical product and partner marketing expert, where I've focused my time on Google Cloud, AWS, and Cisco. While at Softchoice, my most favored accomplishment has been hosting the first ever customer facing AWS DeepRacer event in Canada, which taught our largest financial customers hands-on technical experience with reinforcement learning.

Outside of work, you can find me building Lego, exploring the night sky with a telescope, or with my head in a book.

