5 Things Every Business Leader Needs to Know About Data Strategy

Hugo Bowne-Anderson
@hugobowne

DataCamp
A bit about me

→ Hugo Bowne-Anderson, data scientist at DataCamp
  ◆ Undergrad in sciences/humanities (double math major)
  ◆ PhD in Pure Mathematics (UNSW, Sydney)
  ◆ Applied math research in cell biology (Yale University, Max Planck Institute)
  ◆ Python curriculum engineer at DataCamp
  ◆ Host of DataFramed, the DataCamp podcast
  ◆ Data & AI evangelist, strategy consultant
Our Mission

Our mission is to democratize data science education by building the best platform to learn and teach data skills and make data fluency accessible to millions of people and businesses around the world.
Learn by doing

- Short videos from expert instructors
- In-browser coding
- Real-time feedback

300+ Unmatched data science courses

- Languages: Python, R, SQL, Git, Shell, Spreadsheets
- Topics: Importing & Cleaning, Data Manipulation, Visualization, Probability & Statistics, Machine Learning, and more!

Industry-leading instructors

- Learn from the authors of renowned code packages and the organizations that understand data science innovation
The 80/20 rule for data science
Big data ain't all that (big)
The future of data work is point-and-click
Data strategy means data culture
Data strategy means empathy
Today’s topics of discussion

➔ The 80/20 rule for data science
➔ Big data aint all that (big)
➔ The future of data work is point-and-click
➔ Data strategy means data culture
➔ Data strategy means empathy
The Pareto Principle

- Many trivial tasks: 80% of time expended
- Few vital tasks: 20% of time
- 20% of results
- 80% of results
Large, unsexy wins for data

What can move the needle?

- Data unification
- Data views and dashboards
- Call center routing
- Customer churn models
- Sales propensity modeling (B2B)
- (Basic) Conversational AI
We can slice data science into 3 components:

1. Descriptive analytics (Business Intelligence)
2. Predictive analytics (Machine Learning)
3. Prescriptive Analytics (Decision Science)
1. **Business Intelligence (descriptive analytics)**

- Taking data company already has
- Getting that data to the right people
- In form of dashboards, reports, emails
Data scientists answer business questions & are one of several inputs into the decision-making process.

Renee Teate (Heliocampus, Becoming a Data Scientist):

Renee’s expectation for any data analyst or data scientist:

1. Business Question
2. Data Question
3. Data Answer
4. Business Answer
< 20% of your data work will generate >80% of the value.

Examples of data projects & initiatives.
→ List 5 projects based on descriptive analytics that could inform your decision making.
→ Order them in terms of what could move the needle for your business.
Today’s topics of discussion

➔ The 80/20 rule for data science
➔ Big data aint all that (big)
➔ The future of data work is point-and-click
➔ Data strategy means data culture
➔ Data strategy means empathy
Small data: shooting for the stars
How can a poll of only 1,004 Americans represent 260 million people with only a 3 percent margin of error?

Andrew Gelman, a professor in the departments of statistics and political science at Columbia University, explains.
The End of Theory: The Data Deluge Makes the Scientific Method Obsolete
Where is big data in the Gartner Hype cycle?
POLL: Where is big data in the Gartner Hype cycle?

1. Peak of Inflated Expectations
2. Trough of Disillusionment
3. Slope of Enlightenment
Where is big data in the Gartner Hype cycle?

The End of Theory, Wired

big data
Search term
This is the right way to think about AI. It’s not a magical black box — it’s a highly-specialized set of tools. It’s not about shooting for the moon — it’s about winning the ground wars. And it’s not about mountains of data — it’s about small, high-precision data.
The Power of 'Thick' Data

Businesses need to know how a product or service fits into the emotional lives of their customers

By Christian Madsbjerg and Mikkel B. Rasmussen
March 21, 2014 7:15 pm ET

Big Data Is Only Half the Data Marketers Need

by Mikkel B. Rasmussen and Andreas W. Hansen
Small data can be as powerful as big data.

Having a variety of data is more important than having big data.
Choose a data source that’s valuable to your business.

How much of this data do you really need to inform decision making?

What is the added value of increasing this amount 2x, 5x, and 10x?

What thick data could you use to enhance it?
The 80/20 rule for data science
Big data aint all that (big)
The future of data work is point-and-click
Data strategy means data culture
Data strategy means empathy
Google closes $2.6B Looker acquisition

Ron Miller  @ron_miller / 11:35 am EST • February 13, 2020

Salesforce Completes Acquisition of Tableau

PUBLISH DATE: AUGUST 1, 2019 - 5:30AM
Google AutoML makes adding AI as simple as drag and drop
Microsoft launches a drag-and-drop machine learning tool
Non-coder uses for ML tools

➔ Customer churn
➔ Hiring flow
➔ Marketing funnel
➔ Programmatic buying
➔ Supply chain optimization
Drag-and-drop reduces the barrier for entry: in house or vendor?

### Build vs. Buy: The Age-Old Question, for AI

<table>
<thead>
<tr>
<th>Build If...</th>
<th>Data Processing Infrastructure</th>
<th>Coordination of all Data Efforts (AI Platform)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Models / Applications</strong></td>
<td><strong>Internal skills available to support, maintain and evolve data infrastructure and processing technologies.</strong></td>
<td><strong>The use case is one-off and short term, focused on a limited team.</strong></td>
</tr>
<tr>
<td>Details of the use case (data, process, decision criteria) are specific to your organization. Short to medium ambition to develop broader in-house data capability and expertise.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Buy if...</strong></td>
<td><strong>No extended internal data IT team or hardware stack.</strong></td>
<td><strong>The enterprise is looking to scale up data competency to many use cases and teams over the long term.</strong></td>
</tr>
<tr>
<td>Data and processes for specific use case are generic. Limited/no further data projects expected in the short term.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

→ Disclaimer: dataiku is a data science platform vendor
Modzy accelerates the deployment of trusted AI at scale

The Modzy platform and model marketplace create the missing AI layer in today’s tech stack, accelerating the deployment of AI from the lab to the enterprise. With click-to-deploy access for a growing list of AI models from leading tech companies and open source communities, and an environment to upload, manage and reuse AI models, Modzy greatly reduces risk and barriers to adopting and scaling AI. Interested customers can request an invitation to the Modzy Early Access Program, beginning today.
Amazon scraps secret AI recruiting tool that showed bias against women

**Survival of the Best Fit**, an educational game about hiring bias in AI.
We’ll see more & more data work done in GUIs.
The space is increasingly competitive (Google/MSFT, and more).
Even if much of your data science is done in house, you’ll increasingly work with vendors.
You may be equipping people to be dangerous!
How much data work in your organization currently happens in code (e.g. in human hours)?

How much in GUIs?

How do you see this changing over the next 24 months?
Today’s topics of discussion

- The 80/20 rule for data science
- Big data ain't all that (big)
- The future of data work is point-and-click
- Data strategy means data culture
- Data strategy means empathy
Taras Gorishnyy identified the necessary moving parts (Episode 30, DataFramed):

- Executive support
- Analytics vision
- Build data foundations
- Distribution of skills & establish data culture
- Establish impact of analytics early on in the process
“Everyone at any level, C-level, entry level, should be looking and diving into data the same way that you were expected to start using email 20 years ago.”

TANYA CASHORALI
POLL: How would you rate your data culture (1-5)?
To what extent has your company taken (or plans to take) the following actions to become data fluent?

- Create a vision for analytics aligned with company strategy: 44% (Mature DF competencies, N=150), 6% (Immature DF competencies, N=95)
- Establish strong C-level support: 35% (Mature DF competencies, N=150), 7% (Immature DF competencies, N=95)
- Extract value early from at least several use cases: 32% (Mature DF competencies, N=150), 12% (Immature DF competencies, N=95)
- Implement process redesign and culture change: 39% (Mature DF competencies, N=150), 7% (Immature DF competencies, N=95)
- Build data infrastructure to provide integrated, trusted, and timely data: 43% (Mature DF competencies, N=150), 20% (Immature DF competencies, N=95)
Hiring top talent is the primary data fluency hurdle

To what extent do the following business challenges prevent your company from building/improving data fluency?

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Already Invested in DF, N=157</th>
<th>Future Investment in DF, N=114</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of digital strategy</td>
<td>18%</td>
<td>34%</td>
</tr>
<tr>
<td>Complexity of core business</td>
<td>30%</td>
<td>39%</td>
</tr>
<tr>
<td>Lack of C-level support</td>
<td>18%</td>
<td>36%</td>
</tr>
<tr>
<td>Difficulty hiring top talent</td>
<td>29%</td>
<td>46%</td>
</tr>
<tr>
<td>Instability of business plans</td>
<td>20%</td>
<td>34%</td>
</tr>
</tbody>
</table>
How to build data skills

What actions has your company taken to build data skills?

- **Established a data university/Center of Excellence**: 37% (Mature DF competencies, N=150) 15% (Immature DF competencies, N=95)
- **Provide on-site/in-person training**: 60% (Mature DF competencies, N=150) 57% (Immature DF competencies, N=95)
- **Provided access to online learning platforms**: 64% (Mature DF competencies, N=150) 62% (Immature DF competencies, N=95)
- **Launched internal campaigns to promote/incentivize learning**: 51% (Mature DF competencies, N=150) 37% (Immature DF competencies, N=95)
- **Recruited employees with existing, relevant skills in data fluency**: 49% (Mature DF competencies, N=150) 48% (Immature DF competencies, N=95)
- **Sourced solutions from a vendor to bridge gaps in data fluency**: 35% (Mature DF competencies, N=150) 23% (Immature DF competencies, N=95)
Establishing data culture with IPTOP

**Infrastructure**
- Set up a data lake
- Enable data discovery

**People**
- Map out roles and skills
- Identify skill gaps
- Personalize learning path

**Tools**
- Build tools to encapsulate.
- Build frameworks to automate.

**Organization**
- Embrace a hybrid model
- Build flexibility

**Processes**
- Standardize project structure
- Embrace version control
- Embrace notebooks
Communication and data flows are key to success.
Everybody needs to be data literate.
Establish a data culture instead of hiring unicorns.
→ List 3-5 outputs of data work in your organization.
→ Audit them with respect to how much they are *actually* used to inform decision making.
POLL: What percentage of your data work is actually used?

1. 0-25%
2. 26-50%
3. 51-75%
4. 76-100%
Today’s topics of discussion

- The 80/20 rule for data science
- Big data ain't all that (big)
- The future of data work is point-and-click
- Data strategy means data culture
- Data strategy means empathy
At-scale bias in data science

The @AppleCard is such a [redacted] sexist program. My wife and I filed joint tax returns, live in a community-property state, and have been married for a long time. Yet Apple’s black box algorithm thinks I deserve 20x the credit limit she does. No appeals work.

3:34 PM · Nov 7, 2019 · Twitter for iPhone
8.8K Retweets 26.2K Likes

Building a fairer and more inclusive financial services industry for everyone

By Linda A. Lacewell, Superintendent of New York State Department of Financial Services

Future Finance

Apple Co-Founder Says Goldman’s Apple Card Algorithm Discriminates

By Shahien Nasiripour and Sridhar Natarajan

- Steve Wozniak urges tougher regulation on credit algorithms
- Regulator opens probe after sexism was alleged in viral tweets

“These sorts of unfairnesses bother me and go against the principle of truth. We don’t have transparency on how these companies set these things up and operate,” Wozniak said in an interview on Sunday. “Our government isn’t strong enough on the issues of regulation. Consumers can only be represented by the government because the big corporations only represent themselves.”

Wozniak said he can borrow 10 times as much as his wife on their Apple Cards even though they share bank and other credit card accounts, and that other lenders treat them equally.

“Algos obviously have flaws,” Wozniak said. “A huge number of people would say, ‘We love our technology but we are no longer in control.’ I think that’s the case.”
Consider your stakeholders

Amazon scraps secret AI recruiting tool that showed bias against women

The Algorithmic Colonization of Africa

Startups are importing and imposing AI systems founded on individualistic and capitalist drives

Machine Bias

There’s software used across the country to predict future criminals. And it’s biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchne, ProPublica
May 23, 2016
Old definition of data scientist: “A data savvy, quantitatively minded coding literate problem solver.”

New definition: “Data science doesn't just predict the future. It causes the future.”

Algorithmic audits, including a sensitivity analysis

Ethical matrix: “rows are the stakeholders, the columns are the concerns.”

<table>
<thead>
<tr>
<th></th>
<th>Efficiency</th>
<th>Fairness</th>
<th>False +'s</th>
<th>False -'s</th>
<th>Transparency</th>
<th>Predictive Parity</th>
<th>Consistency</th>
<th>Data quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Court</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black Defendants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Defendants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northpointe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How facial recognition became the most feared technology in the US

Two lawmakers are drafting a new bipartisan bill that could seriously limit the use of the technology across the US.

By Shirin Ghaffary | Aug 9, 2019, 4:00pm EDT

Democrats Propose Federal Privacy Legislation That’s Tougher Than CCPA

by Allison Schiff // Wednesday, November 6th, 2019 - 12:35 am

RENEE DIRESTA IDEAS 07.24.2018 09:00 AM

A New Law Makes Bots Identify Themselves—That's the Problem
Employees as stakeholders

I got fired from work today

They replaced me with a human

OH really? What happened?
Recap

→ Consider your key stakeholders & assess your impact across them.
→ Consider your blind spots!
→ Have honest, transparent conversations around what may & may not be automated & career paths.
Choose a current data initiative & list the stakeholders (those impacted).

List 2 ways that each stakeholder could be positively impacted and two ways they could be negatively impacted.
What’s next?
Now you know the basics, next time we'll delve deeper into key aspects of implementing data strategy.

➔ “Building a scalable data strategy with IPTOP”
➔ 11am ET, March 19 (the third Thursday of the month)
Thank you!

Hugo Bowne-Anderson
Data Scientist
@hugobowne