Accelerated digital transformation

Research, perspective and guidance
from data scientists and thought leaders.

Check it out

Contents

Skip around just by clicking

Introduction

Prospects for the field of data science continue to grow around the world as digital transformation has accelerated due to the COVID-19 pandemic, driving whole swaths of the world to work, play and seek entertainment online. And multiple studies measurably confirmed that outcome – particularly a McKinsey survey taken well after the pandemic had swept the globe that found that the COVID crisis accelerated the digitization of customer interactions by several years.[1]

With all this increased digitization generating more data, simple logic points to a greater need to analyze and make good use of it, which translates to no shortage of work for data scientists. It has turned out that way, but the increases in work volumes and the opportunities you'd expect to come with them don't seem to have been consistent nor without complications because of the differences among organizations in leadership, systems, cultures, and even the nature of the business challenges.

Survey objectives

So we decided to look below the surface of industry trends and into the work environments of data scientists to see how the story might change, and also to formulate ideas about how to enhance the situation for them. We reached out directly to data professionals around the world with a survey to get their inputs across five dimensions of their daily work environments:

1. The impact of the pandemic on their work.
2. Work activities performed – the variety and relative time spent on them.
3. Obstacles they face in fulfilling their roles.
4. Satisfaction with aspects of the analytics environment they're in.
5. Skills needed to fulfill their roles and their relative proficiency at them.

Who we talked to

In order to get a mix of viewpoints, we targeted a cross-section of industries, education levels and geographies. We included a range of roles from scientists and analysts to supervisors and executive-level data professionals. This allowed us to have comparisons and contrasts between individual contributors and their bosses.

We also conducted interviews with both data professionals and data science experts. The in-depth interviews with Danielle Boyce,[2] Patrick Butler[3] and Nick Dowmon[4] — data professionals in three different fields — added dimension and specific examples to illuminate what the survey revealed. Interviews with Kirk Borne[5] and Sally Eaves[6], both internationally recognized experts in data science, provided an opportunity to layer in key validating points from their own research-informed perspectives, as well as practical takeaways for **how to make the most of this moment.**

Impact of the pandemic

The current global pandemic has changed the landscape of business forever, and one of the biggest changes has been the sudden shift to remote work across many fields. For some types of work, this shift has made it difficult to bring employees together to tackle problems and drive creative solutions. That particular aspect of collaboration did not seem to affect most aspects of data scientists' work.

Most survey respondents indicated that the direct impact of the COVID-19 pandemic on their work was neutral to positive. Regarding productivity, Figure 1 below shows nearly half reported getting more work done now, versus just about a quarter citing a decrease in the amount of work they get done now versus pre-pandemic levels. Senior software engineer Nick Dowman has definitely felt more productive, and opportunities to collaborate have improved in a remote work environment.

The bigger impact is on the importance of their work as shown in Figure 2 below, with 91% indicating their work being about the same or greater in importance compared to before the COVID-19 pandemic.

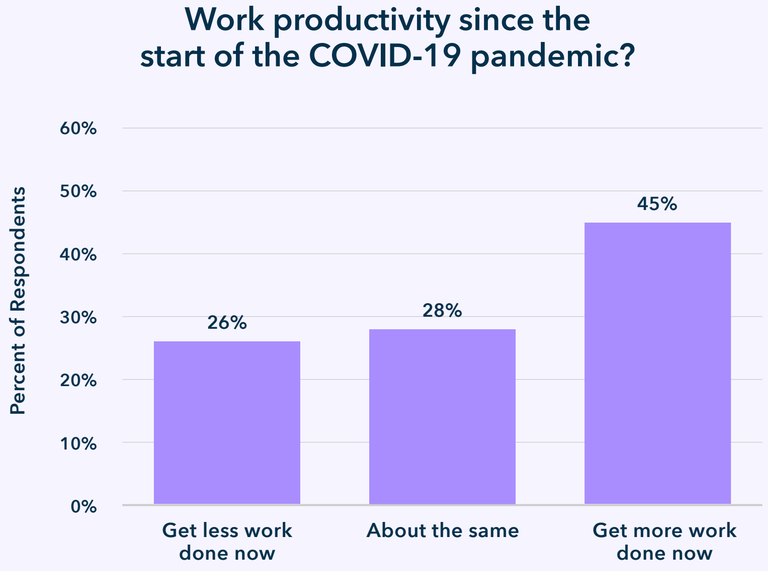


Figure 1. Change in work productivity since the start of the pandemic.

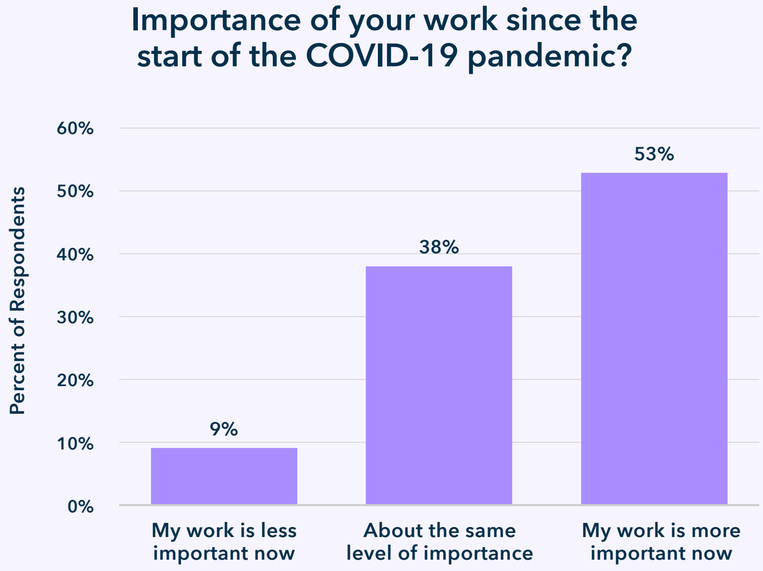


Figure 2. Change in importance of work since the start of the pandemic.

> **We might write our code and share it on GitHub, write pull requests, ask people to review it, and then do it at our own pace with multiple people on Slack.**

Nick Dowmon, Senior Software Engineer

The biggest impact of the pandemic on data scientists is likely due to the acceleration of digital transformation, which caused a ripple effect of adaptations in processes, practices, and operating parameters and assumptions. Everything that changed with the pandemic then moved the assumptions and variables in the models and predictive algorithms that were in effect throughout organizations worldwide, such as:

- A surge in e-commerce, while some stores were shuttered.
- Supply chain disruptions and changes in shipping costs.
- Staffing needs becoming unpredictable.

The resultant impact on collaboration and the correlated use of cloud services, as shown in **Figures 3 and 4** below, had a largely positive effect. The most marked result was the very low number indicating a lower use of cloud services, which just highlights how much of a shift there has been toward cloud services.

It's interesting that our study revealed statistically significant positive correlations between the use of cloud services and level of collaboration (r = .44) and productivity (r = .53). That is, data professionals who increased collaboration and increased their productivity also have used more cloud services since the start of the pandemic, suggesting that improving collaboration and productivity in a time of remote work is dependent on the use of cloud services.

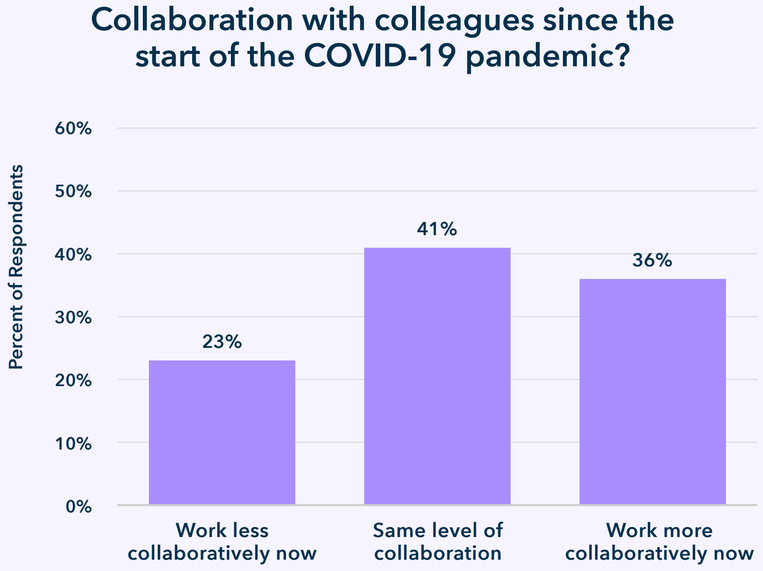


Figure 3. Change in collaboration since the start of the pandemic.

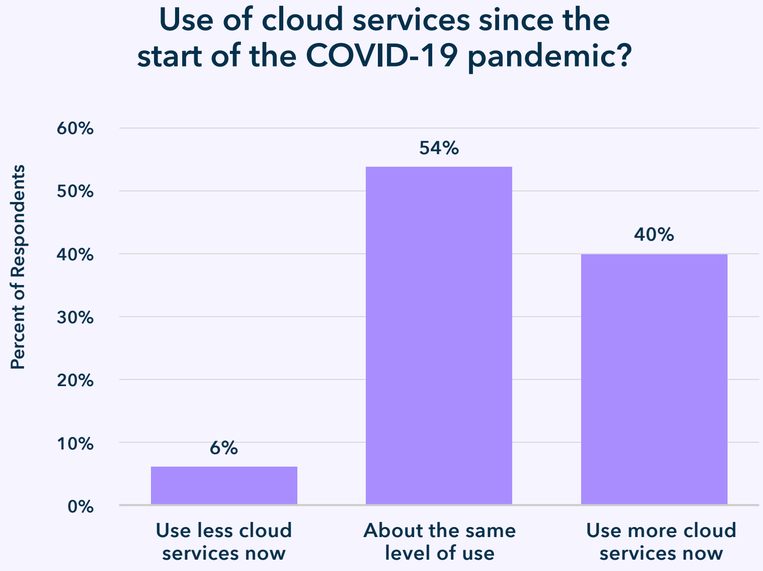


Figure 4. Change in use of cloud services since the start of the pandemic.

Work activities performed

A typical data science project involves a variety of activities, almost always beginning with preparing data. One notable finding in the survey is that data professionals are spending much more time **(58%)** gathering, exploring, managing and cleaning data, as shown in **Figure 5** than they would like to **(42%)**. And that's in contrast with an average of **11%** of their time spent creating computer models, as shown in **Figure 6** would be closer to **21 - 24%.**

Proportion of your time typically devoted to these following tasks?

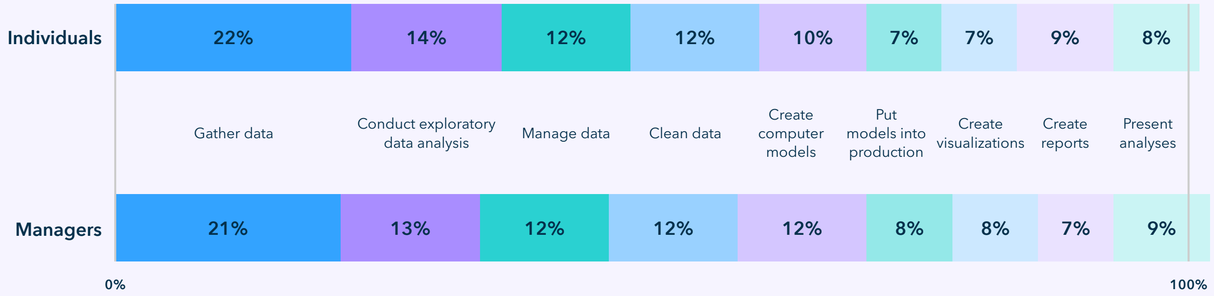
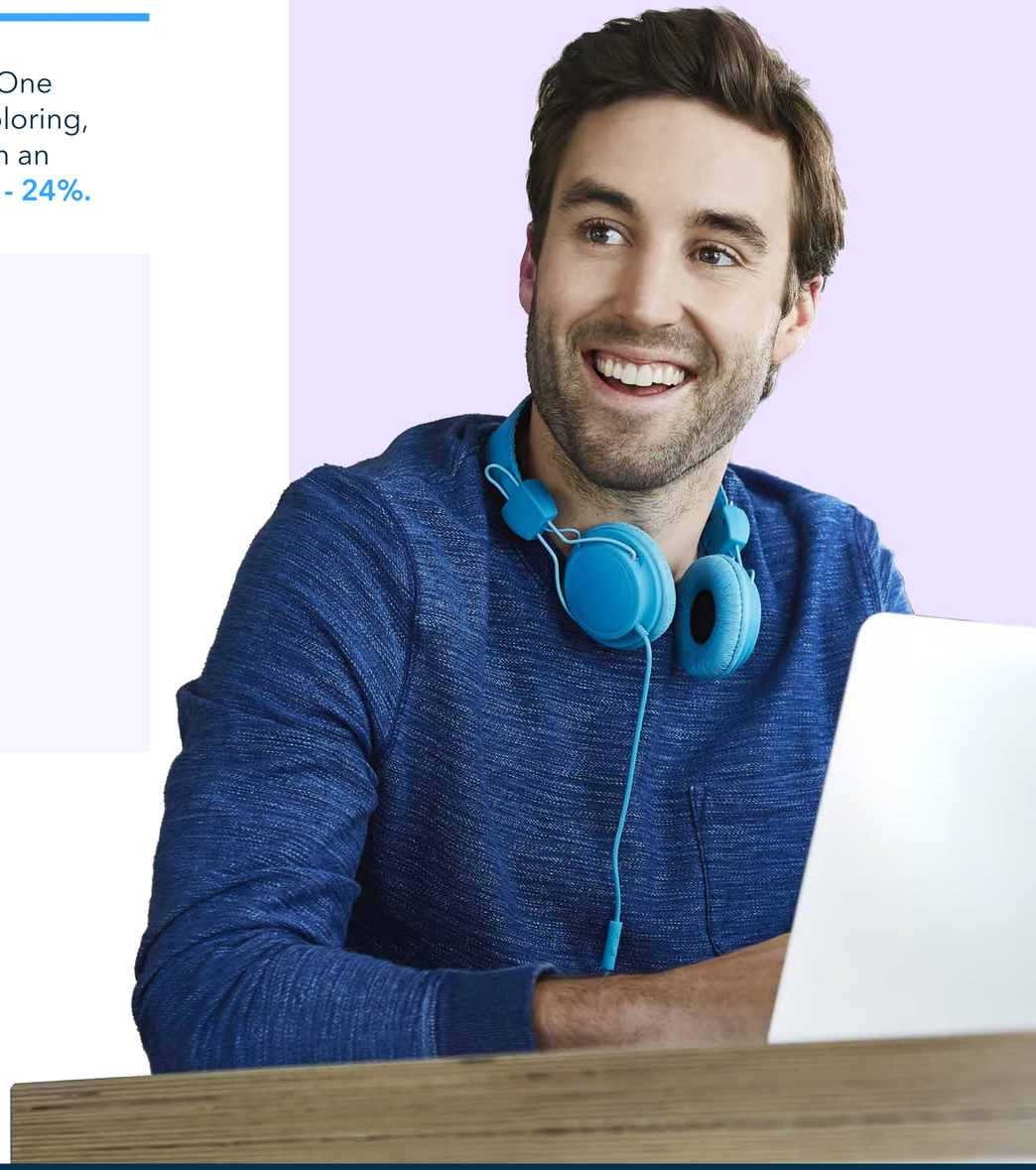


Figure 5. Proportion of time spent on tasks during a typical data science project.

Comparing individual contributor respondents to their managers reveals little difference in attitudes about how time is spent. So it seems that career progression won't necessarily change the amount of time you spend on data management. In contrast to our respondent pool, our interviewees had strong opinions about the importance of the data preparation stage.



Kirk Borne from DataPrime suggested you think of data prep as your first date in your long-term relationship with your data. Senior Clinical Data Analyst Danielle Boyce from Johns Hopkins Medicine says that in her eyes the analysis starts the minute she begins sizing up the data she's working with. The whole front-end managing and cleaning data process is an intrinsic part of the modeling process for corporate Data Scientist and Data Science Bootcamp Leader Patrick Butler – without it, he says, all the modeling that follows is truly **"just math."** And the way Dowmon sees it, "As a data scientist, you need to prioritize having really clean data sets and making sure you have a really solid set of training data."

Proportion of your time you would rather spend on each of the following tasks?

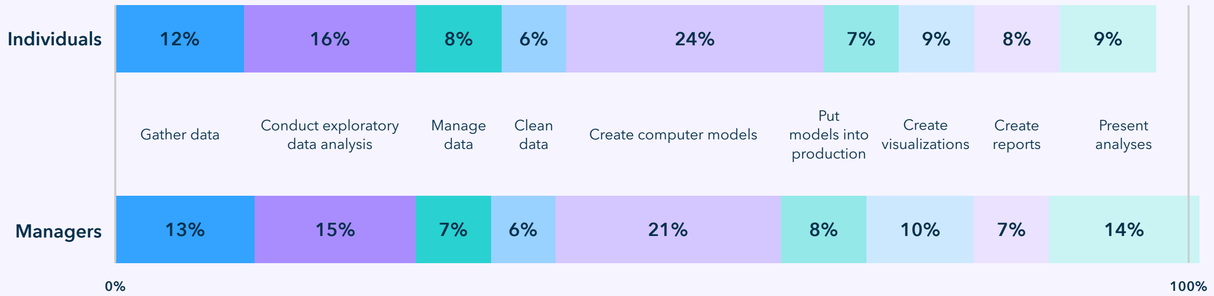


Figure 6. Proportion of time data professionals would like to spend on tasks during a typical data science project.

The fact is that, regardless of your level in the organization, data management will probably take a large share of your time, even with the development of low code/no code tools and AI and machine learning algorithms being written for it. The likely reason is that the data you have and how you decide what's relevant is probably specific to your industry and organization. As is the case for how you approach your model-building, knowing which data is relevant and why has a lot to do with the issues you are trying to solve. Sally Eaves had an even broader view of possibilities for data scientists, including aspects of modeling and getting involved in shaping the guiding principles of working with data in your organization.

Data cleaning is one of the most important data science roles. You are the hero every single time because nobody could get any good information out of that data set without you.

Danielle Boyce,
Senior Clinical Data Analyst

Obstacles faced in data science projects

As with any job function, data professionals face obstacles and challenges as part of the normal course of their work. When given the ability to identify multiple challenges, on average, data professionals indicated they have experienced around five challenges in the past 12 months, with the responses shown in **Figure 7** below.

Borne underscored the universality of all the barriers cited by the respondents – lack of talent, friction from decision makers in deployments, insufficient data fluency in the organization and so on. His advice to avoid getting frustrated is simply to **"think big, and start small"** — but to get started nonetheless.

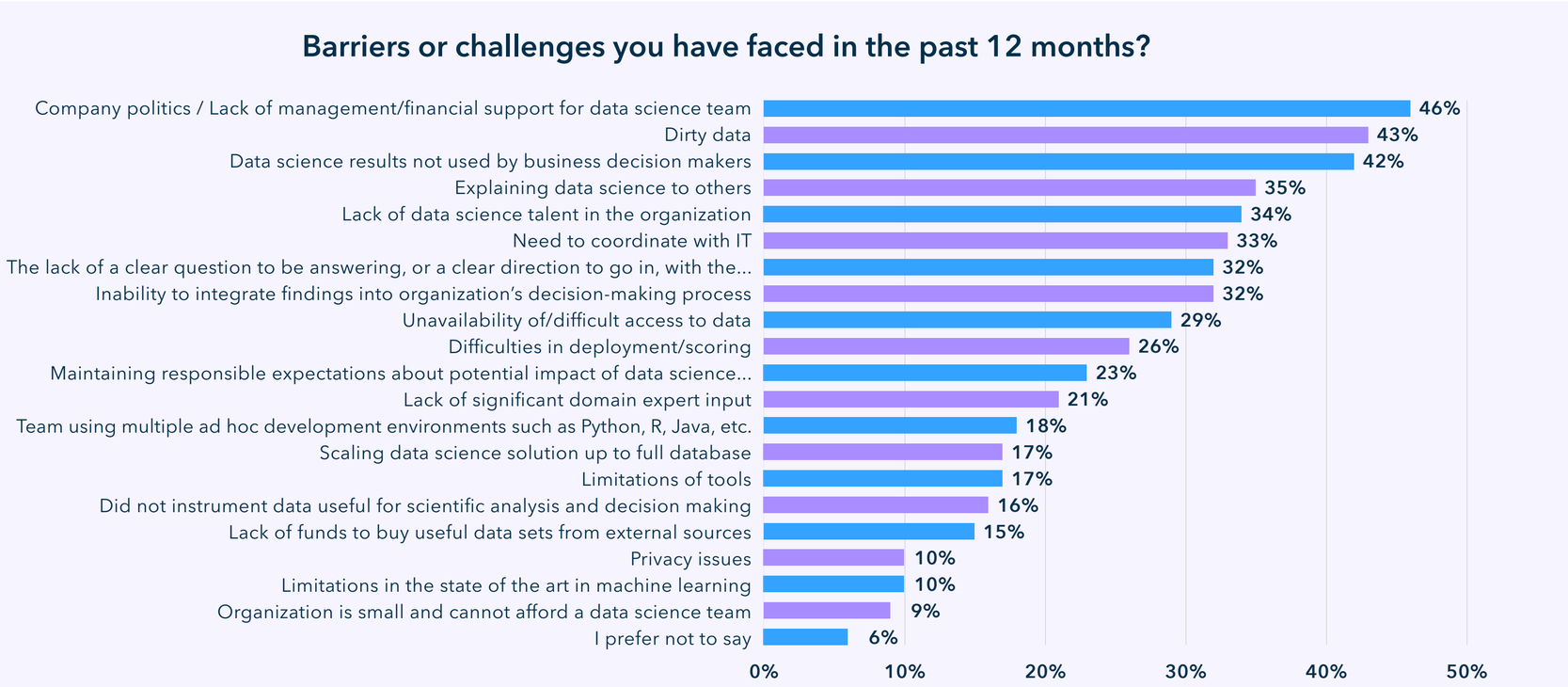


Figure 7. Barriers or challenges faced by data professionals in the previous 12 months

AI and ethics – a major roadblock?

As part of this focus on obstacles, we looked at ethics practices in relation to AI to see if there were any roadblocks to efforts at limiting undesirable, unethical or biased applications of the technology. While this is a huge topic, worthy of a study in its own right, here, **43%** of respondents indicated that their organization does not conduct specific reviews of its analytical processes with respect to bias and discrimination. At the same time, responsible AI doesn't seem to have the same level of priority universally – just **26%** of respondents indicated that unfair bias is used as a measure of model success in their organization.

Borne suggests that the field of clinical science may be able to help, by thinking of applied AI as a form of a grand experiment on humanity. In clinical studies that involve human subjects there is usually an internal review board process that weighs the risks, benefits and efficacies and ensures there's fair and equitable distribution of risks and benefits to the participants.

> **"Data scientists can lend their expertise to craft working guidelines for data access, usage security and broader issues, such as sustainability and data ethics and bias."**
>
> Sally Eaves, PhD

> In the case of one lender, they tested if model outputs could be used to infer the input values they intentionally left out of the training, such as gender or race. If they could make inferences from the outputs, then input values leaked to the output solution and bias could have leaked in, despite their best intentions to fix that.

Kirk Borne, PhD

Figure 8 shows the biggest roadblocks to ensuring the outputs of analytics processes are fair and unbiased, with respondents indicating an average of one challenge in this area in the past 12 months. The top two roadblocks were a lack of communication between those who collect the data and those who analyze it, and difficulty in collecting data about groups that may be unfairly targeted.

The biggest opportunity for any data scientist that feels they're facing obstacles to mitigating bias and discrimination in their models, is to follow Borne's advice **"Get started with your own models,"** document your measures, record your outcomes and then present it to management.

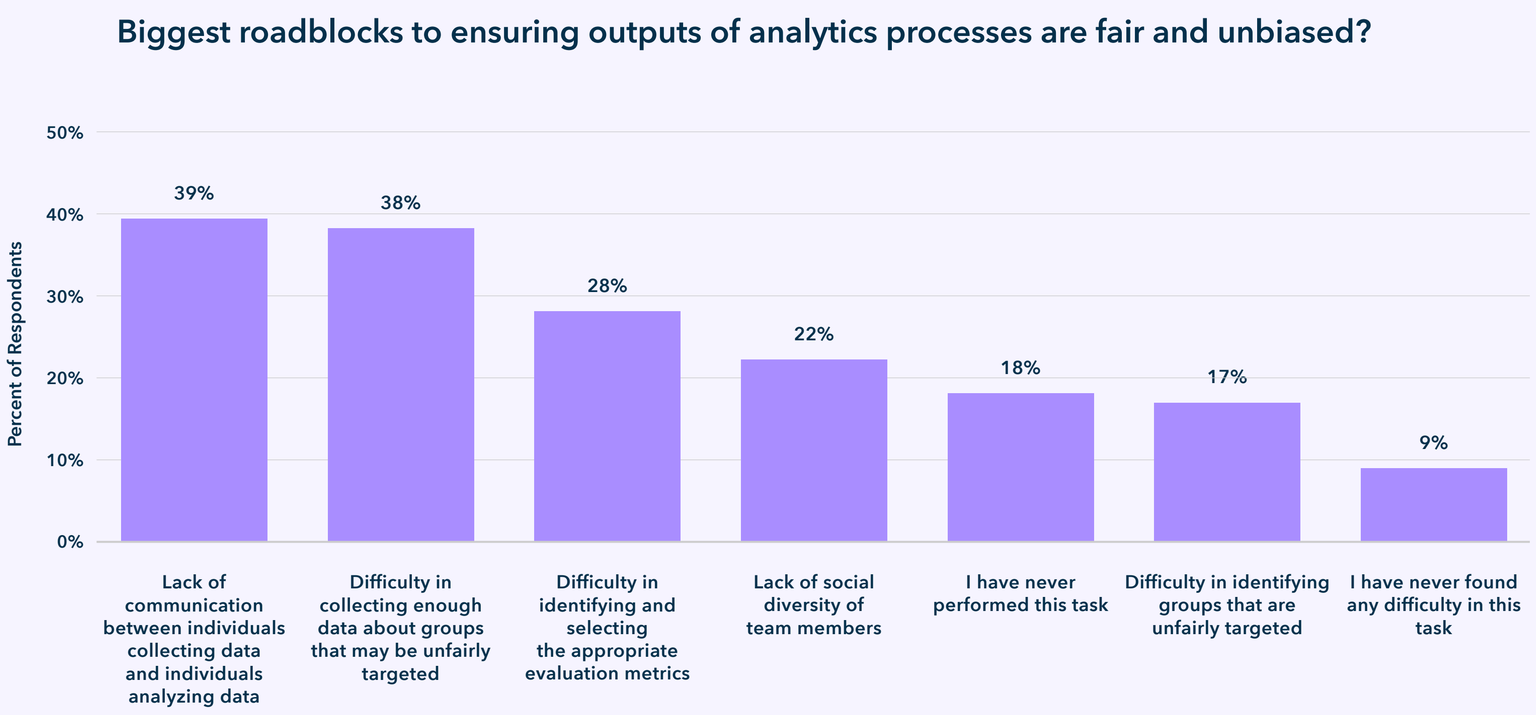


Figure 8. Roadblocks to ensuring outputs of analytics processes are fair and unbiased.

Satisfaction with analytics environment

Multiple studies connect employee satisfaction with lower turnover, better customer service, greater productivity and higher profits. We checked our respondents' satisfaction levels regarding three aspects of the analytics environment in their organization:

1. **The outcomes of the projects they work on.**
2. **The deployment of their models.**
3. **How their company makes use of analytics.**

The respondents showed the highest level of satisfaction and lowest levels of dissatisfaction with the outcome of the projects they work on among both individuals and managers. Those results seem to be related to the higher levels of control they may have in those situations. The other two categories' responses are markedly different – see Figures 9 and 10 below.

Satisfaction with analytics environment

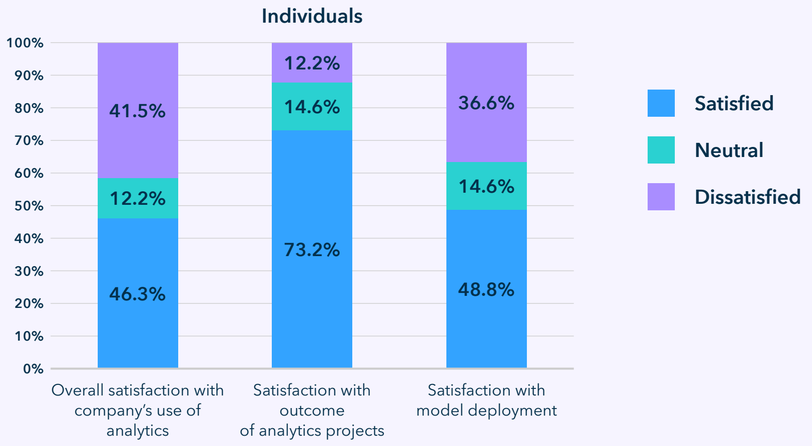


Figure 9. What is your satisfaction level with various analytics-related efforts in your organization?

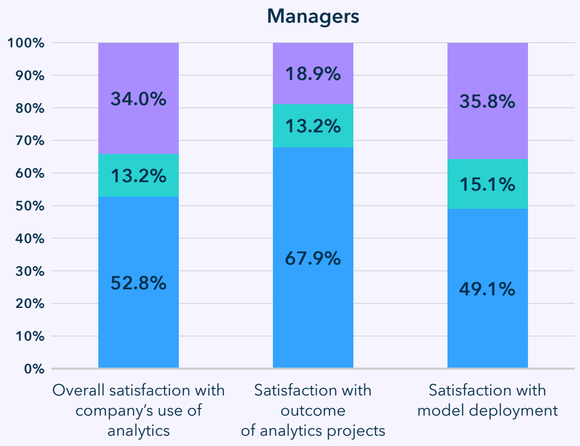


Figure 10. What is your satisfaction level with various analytics-related efforts in your organization?

Possible red flags for bigger issues?

The main factor driving the difference is dissatisfaction with model deployment and with the overall use of analytics – both registering well over a third of respondents. Disconnects of that magnitude are often red flags for broader operational issues and may also reflect a lag between the speed of technology evolution and ability for the organization to keep up with additional training for skill development and new process changes.

To sum up: Managers are generally more satisfied with the company's use of analytics compared to individuals; however, individuals seem more satisfied with the outcome of analytics projects. That difference echoes the possible difference between satisfaction with their own projects' outcomes versus how they're deployed, and that data science as a whole is more than siloed, individual efforts.

One observation Borne made points the way to an opportunity to reduce dissatisfaction – work to make data science a team sport in your organization and to give yourself "a seat at the table" if you are not already there.
In other words — don't limit your efforts to model building and deployment; try to become a part of the decision-making process that uses your models' outputs.

> **…don't limit your efforts to model building and deployment, try to become a part of the decision-making process that uses your models' outputs.**
>
> - Kirk Borne

Skills proficiency

Given the complexity of the world we live in, there should be no surprise in the wide variety of skills that today's data scientists need to be successful. Our survey included 25 different skills for which we asked respondents to indicate their level of proficiency. A few interesting patterns pop out of the range of proficiency levels shown in **Figure 11:**

Over half of the respondents indicated advanced or expert proficiency in these skills:

- Science/scientific method.
- Managing structured data.
- Content knowledge in specialized domains.
- Statistics and statistical modeling.
- Communication.

By contrast, less than a third of the respondents reported having advanced or expert proficiency in program-heavy skills, such as:

- Cloud management.
- Database administration.
- Front- and back-end programming.
- NLP and text mining.
- Systems administration and design.

These findings point to possibilities for data scientists to develop skills in data engineering. The survey findings were validated by a media report indicating robust job growth for data engineers[7], and by input from Boyce, who cited programming capabilities as "the best place to start" when looking to improve the chances of success in a data science project.

Managers showed higher levels of proficiency at identifying/defining business problems and stakeholder management – both of which are necessary for successfully selecting and overseeing projects. These seem like good opportunities for data scientists looking to move up in the organization.

Even if you're interested in remaining an individual contributor, those might even be skills to consider sharpening. Independently of the survey, Butler frequently cites business acumen as the No. 1 competency that data scientists should develop for very specific reasons in his data science bootcamps and in the classes he teaches at the University of Chicago. In his view, **"data science without being connected to business value is just math."**

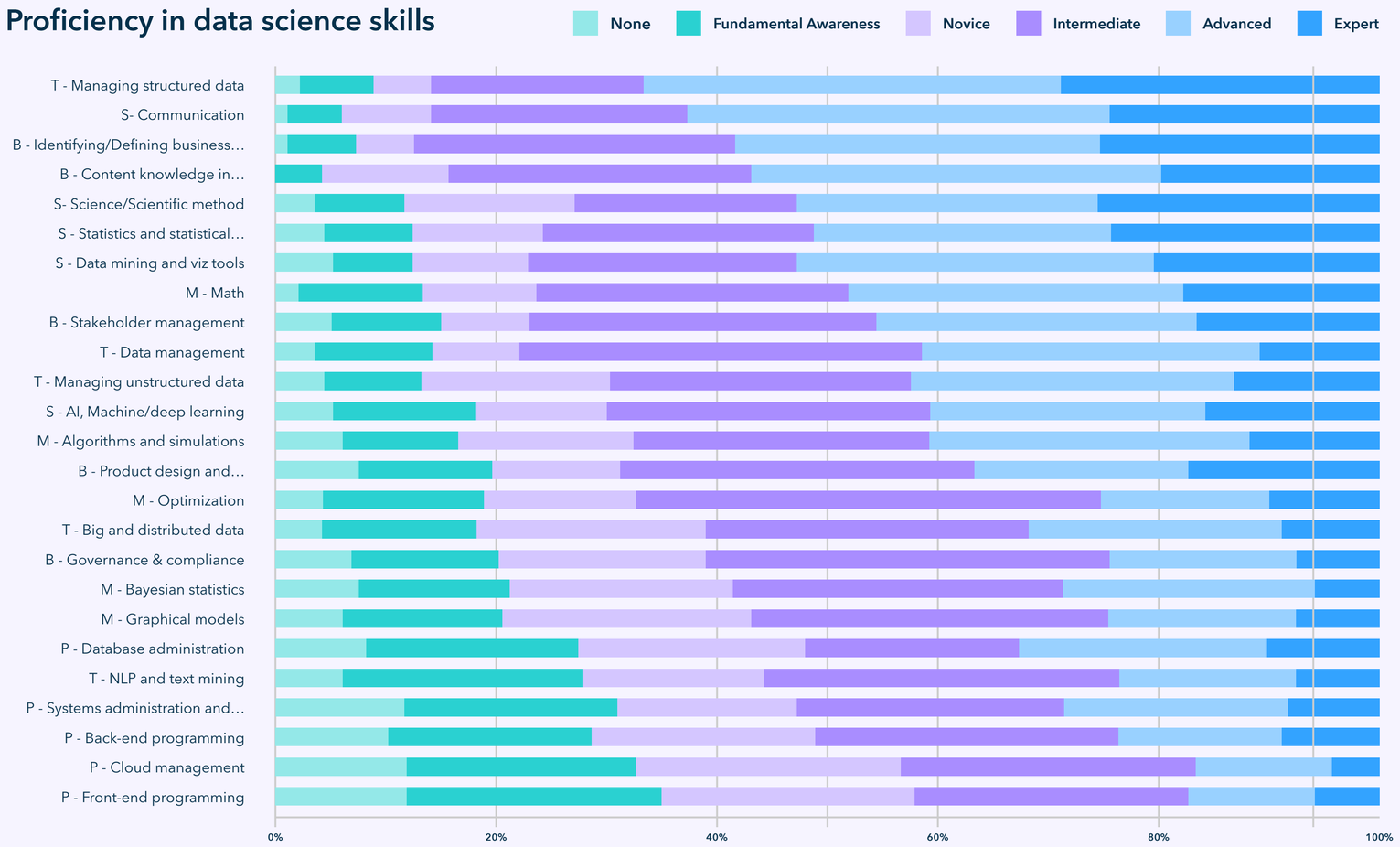


Figure 11. Data professionals' rating of their proficiency for 25 data science skills

Conclusions

The big-picture view of the COVID-19 pandemic causing the acceleration of digital transformation is undeniable. And what this research revealed is that big-picture views don't always tell the full story.

By directly polling and interviewing data professionals and experts, we have explored both the challenges and opportunities. The latter fall into five cohesive thoughts about how to make the most of this moment as a data scientist:

1. **Embrace data prep as your first step in modeling**

   The constant need for data wrangling will spur development of ever-improving AI and machine learning applications to handle those tasks. However, remember our experts' advice, to use the time you save to really get to know your data by cleaning and preparing it yourself.

2. **Pick one obstacle and find a viable work-around**

   Pick one obstacle and get started with finding a solution – it doesn't matter which, but pick one and get started. There is no shortage of obstacles, and they are mostly universal, so find one you can tackle with the time and resources available to you. Then move on to the next.

3. **Be the change you want to see with responsible AI**

   If your organization hasn't yet started down the path of responsible AI, get started with your own project and find ways to add in the means to detect and measure bias. Document your work and present it to management. Sometimes a working example of success is what's needed to get started – there's no reason it can't be your work that sparks the beginning.

4. **Find yourself a seat at the table**

   Work with your internal clients collaboratively and find ways to be included in strategy discussions at the planning stages as well as the decision stages. Borne was emphatic about your need to be integrated into the core of the business and a part of the business conversation, and it meshes with Butler's advice about the No. 1 skill to pick up.

5. **Fine-tune your skills**

   Keep your skills sharp, but two kinds of skills jumped out of our survey and interviews to focus on above others – business skills and program-heavy/data engineering skills. The former will help you keep that seat at the table – and it will help you ask the questions you didn't know to ask once you start discovering what's hidden in the data. The latter are where fewer data scientists have developed expertise – that's a simple matter of supply and demand.

Find pathways to reach new heights as a data scientist »

Methodology and scope

We invited data professionals to answer survey questions during 2021 about their work via multiple sources, including TechTarget community members and through social media platforms. In all, **277 data professionals** globally completed survey responses and the quantitative portion of the research was augmented with qualitative interviews with the data professionals and experts.

A majority of respondents worked for companies with more than **1,000 employees,** and a majority of the respondents held a master's degree or higher, with roughly an even split of individual contributors and managerial-level respondents.

Endnotes

1.https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever

2. Danielle Boyce, MPH, DPA: https://www.linkedin.com/in/data-danielle

3. Patrick Butler: https://www.linkedin.com/in/patricknicholasbutler/

4. Nick Dowmon: https://www.linkedin.com/in/ndowmon/

5. Kirk D. Borne, PhD: https://www.linkedin.com/in/kirkdborne

6. Sally Eaves, PhD: https://www.linkedin.com/in/sally-eaves/

7. https://analyticsindiamag.com/why-data-engineering-is-the-fastest-growing-tech-job-in-2021/

112560_G204922.0122