# Data and AI Survey Report 2021/2022

The Rise of Data Democratization and the Growing Need for (Data) Translation





### **Table of Contents**

- 02 General Information
- 03 Executive Summary
- 04 Data and AI Strategy
- 04 Challenges to Creating a Successful Data Strategy Within Organization
- 06 Significance of Data and AI to Overall Strategy
- 07 The Potential of Data and Al
- 08 Team and Organization
- 08 Impact of COVID
- 09 Biggest Challenges to Introducing a Data-Driven Way of Working
- 11 Biggest Challenges to Implementing Data-Driven Solutions
- 12 The Need for Analytics Translators
- 12 Fully Automated Data Ingestion and Transfer
- 13 The Use of Data for Dashboards and Reports and to Develop Predictive Models
- 14 Organizing, Developing and Creating Opportunities from Data and AI Projects
- 14 Using External Consultants to Develop Data-Driven Use Cases
- 14 Attracting Talented Data Professionals
- 16 Data and Technology Trends
- 16 The Most Popular Data Technologies
- 18 The Most Popular Data Visualization Tools
- 18 Cloud Security
- 19 Cloud Solutions
- 20 Data Trends: Taking Data intoProduction Ranked Most Important

#### **General Information**

For the sixth consecutive year, GoDataDriven has collaborated with Big Data Expo and Computable to conduct its annual Data Survey. Around 400 professionals with various positions from more than ten different industries shared their insights and experiences on topics such as data strategy and technology, data science implementation, the cloud, and more efficient talent acquisition.

### **Executive Summary**

#### Data Democratization and the Data Translation Gap

- **01.** Collecting data is becoming easier, but improving data quality and preparing data sets is one of the biggest challenges. Making data broadly available across the organization is challenging, therefore, digitizing processes remains difficult.
- **02.** Many organizations have not automated data ingestion and want to optimize this process, including how to translate raw data into structured data available for the business. Otherwise, a gap forms between data and business, obstructing the business.
- **03.** Data translators can streamline this process. Together with another position to bridge that gap and translate abstract concepts into usability: the analytics engineer.

#### The Challenge of Implementing Data Applications

- **01.** Taking data applications into production is the most important data trend.
- **02.** Limited technical know-how is the main obstacle, because organizations still have difficulty increasing the level of knowledge of their workforce in data engineering and data science.

So finding the right skills to implement data applications and bring automation to important decisions remains a critical and complex challenge.

#### The Effects of COVID and Geographically Dispersed Teams

- **01.** COVID-19 accelerated digital and data transformations.
- **02.** Across industries, the move to online pushed the implementation of data-driven applications.
- **03.** Working from home did not hinder this transition. On the contrary, many organizations saw it increase their data-related activities and most professionals were optimistic about the positive effect on their productivity.

#### **Evolving Cloud Trends**

- **01.** Complete on-premise data storage is a thing of the past. Not a single organization doesn't use cloud in some way or another.
- **02.** Private clouds are losing ground, with more trust placed in public cloud offerings.
- **03.** Amazon, Google, and Microsoft still hold the biggest chunk of market share for public cloud services. Microsoft Azure leads the pack, with 50% market share amongst the survey participants.



# Data and Al Strategy Challenges to Creating a Successful Data Strategy Within Organizations

This year, unlike previous years, we asked professionals the most difficult elements of a successful data strategy instead of the most important.

This year's results show that embedding data processes within the organization are considered the most difficult across most industries, such as a data workflow and Agile development. Embedding is especially difficult for professionals working within the organization's C suite (54%) and managers responsible for data applications, overseeing teams (45%). Also, data scientists and data engineers who develop data applications indicate embedding data processes to be the most difficult for creating a successful data strategy (43%).

Challenges to Creating a Successful Data Strategy Within Organizations

- Professionals agree that embedding processes within the organization and creating a vision are the most significant challenges to creating a successful data strategy.
- Data and AI are predominantly seen as an essential part of an organization's strategy, regardless of its size.
- Most respondents clearly see the potential of data and Al, especially those working in larger corporations.
- Although respondents indicated that data and Al are essential for realizing a successful strategy, organizations in some industries have inadequate budgets for data projects.

Creating a vision and getting support from management are the second and third biggest challenges to creating a successful data strategy within the organization. We found that, especially within the utilities (64%), healthcare (48%), and media and entertainment (55%) sectors, it is challenging to establish a strong vision around data and Al.

ewer professionals saw getting support from management, making budget available, and putting support systems in place as the most difficult aspect to attaining a successful data strategy. Logically, larger companies (>1000 employees) have more difficulty getting support from management (36%) and putting support systems in place (34%) due to more complex hierarchical structures and legacy IT landscapes. Especially within sectors with similar structures, like finance and retail, a high percentage of professionals (38% and 50% have difficulty getting suitable support systems in place.

Although last year's respondents viewed training talent (33%) and attracting talent (55% as crucial aspects of a successful data strategy, this year's survey shows that both of these are now the least difficult. Only organizations within the telecom sector still struggle significantly more with training their staff (55%). Interestingly, these organizations also have the most difficulty with growing data science knowledge (67%). This fact suggests that training others becomes increasingly difficult without existing knowledge in the workforce, especially while implementing a data-driven way of working. Data and Al Strategy / Challenges to creating a successful data strategy within organizations.

Embedding processes in the organization (e.g. the data workflow or Agile) >



#### Significance of Data and AI to Overall Strategy Data and AI are considered essential elements in an organization's strategy. However, budget is often industry dependent.

Data and AI Strategy / Significance of Data and AI to Overall Strategy.





Regardless of size or industry, most organizations agree that data and Al are absolutely essential to their strategy. The travel and tourism sectors agree most with this statement (63% fully agree). In contrast, the education and telecom industries agree the least, with partial or complete disagreement combined at 37% and 25%, respectively. Larger organizations (1000+ employees) with more available data have slightly higher agreement: 64% partial and total agreement combined.

Results were similar across different roles, except for professionals working within the organization's C-suite, who more often indicated that the budget available for data and AI is adequate (53%). Professionals working with data (applications) report less often that there is enough budget for them to propose new data and AI projects (34%).

Although professionals agree that data and AI are essential to their strategies, a quarter still finds available budgets insufficient for these projects. This number has decreased from last year's 29%, meaning organizations are investing more in data and AI projects. However, survey results suggest that healthcare (39%) and retail (25%) sectors are investing the least in data and AI projects, even though these same sectors ranked highest in indicating data and AI as essential to their strategies (56% healthcare, 76% retail).

Data and AI Strategy / Sufficient Budget is Available for Data and AI Projects.



#### The Potential of Data and AI

This year's survey showed a predominantly positive view on the potential of data and AI. Across all industries, 89% of respondents indicated that their organization has untapped data and AI potential, and even more so as the company grows. These organizations often have more data at hand, so see more opportunities to use its potential. Of professionals working at organizations with less than ten employees, 88% see data and AI potential. At the same time, 96% of professionals working at organizations with over 1000 employees see its potential.

This view was most apparent among professionals with a position in the C suite (65% agree), managers responsible for data applications and overseeing teams (72% agree), as well as data scientists and data engineers who develop data applications (70% agree). However, only 40% of professionals who use data for marketing or business intelligence fully agree that their organization could do more with data and Al.

Data and AI Strategy / There is a lot of potential for Data and AI in the organization where I work.



1%	< Totally disagree
2%	<ul> <li>Partially disagree</li> </ul>
8%	< Neutral >
26%	< Partially agree >
63%	< Totally agree >



### Team and Organization Impact of COVID-19

In the last 1.5 years, online remote work has become the norm for most organizations due to the COVID-19 outbreak. Only 14% of survey respondents indicated that they are doing less with data, and not a single one was working in the travel and tourism industry. Not surprisingly, 91% of those working in sectors forced to resort to online alternatives to their products are now leveraging data more in their work and offerings. In the retail industry, 37% are doing more with data, and only 19% are doing less. Similarly, 40% are doing more with data in the media and entertainment industry, and only 20% are doing less.

- The COVID-19 pandemic has increased the level at which organizations work with data. It also forced many professionals to work remotely, although this predominantly had a positive effect on their productivity.
- More effort has been put into attracting new talent professionals since last year. Being forced to work from home did not diminish the perceived importance of a company car or commuting time.
- Overall, this year's results show a decrease in the number of challenges when implementing a data-driven way of working. Making time available for experiments and developing knowledge of big data and data science are still at the top of the list.
- Data translators might not be able to bridge the gap between collected data and usable data. An analytics translator might be more suitable.
- Data ingestion and transfer is still a challenge for many organizations, with 39% of organizations stating they do not have this automation in place.



Team and Organization / Due to the Covid-19 outbreak, our organization is doing more with data.



While the forced shift to remote work increased their use of data, it did not hinder organizations' data operations. On the contrary, 28% reported that remote work had a positive effect on their data operations. Professionals working in industries with more physical machinery reported a slightly more negative view towards working with data remotely: 11% in healthcare, 14% in manufacturing, and 9% in the utilities sector.

### Team and Organization / What's the effect of working from home on your data operations?



### Biggest Challenges to Introducing a Data-Driven Way of Working

Results from this year's survey suggest that being forced to work from home has given professionals fewer challenges when introducing a data-driven way of working. Respondents perceived gaining support from management as less challenging (34% this year vs. 41% last year). The same goes for developing knowledge of big data and data science. Only 34% of respondents indicated this as the most significant challenge, compared to 46% last year. Organizations have had to adapt to a rapidly changing environment since last year's report, which hints at a more embedded data-driven way of working. In the face of tremendous turmoil, they have become more dynamic, thus decreasing the challenge.

Among the lowest challenges are adopting an agile way of working (down from 16% last year to 15% this year), recruiting data engineers (down from 25% last year to 18%), and working in multidisciplinary teams (19% compared to 23% last year).

One challenge that increased among respondents in this year's survey report is making automated decisions for business-critical processes (29% compared to last year's 25%). The increasing availability of data across organizations, also called data-democratization, is one explanation. Although there is a declining trend in the number of challenges organizations face when introducing a data-driven way of working, 34% of organizations still lack the knowledge to use the data. This limited know-how creates a gap between the amount of data available and its usability to make automated decisions across the business.



Team and Organization / What are the biggest challenges of introducing a data-driven way of working?

Making time available for experiments >



#### Biggest Challenges to Implementing Data-Driven Solutions

The challenge to implement data-driven solutions is also declining, except when implementing security and compliance (GDPR). Security and compliance were slightly more complicated when implementing these solutions (up from 21% to 22%), especially in the financial sector (36%).

Despite declining challenges overall, improving data quality is still the number one challenge in implementing data-driven solutions, and it increases as companies grow larger. Only 26% of smaller organizations (<10 employees indicated improving data quality as one of its greatest difficulties. In comparison, 50% of corporates (1000+ employees) indicated struggling with improving data quality. Interestingly, the challenge of collecting data is decreasing (down from 26% last year to 19% this year). Again, this creates a gap between available raw data and organization-wide usable data, suggesting a need for more analytics translators within organizations to bridge it.



#### **The Need for Analytics Translators**

Analytics translators form the liaison between senior management, business operations, and data experts. The analytics translator can be both a gatekeeper and a host for Al projects, brainstorm ideas with executives, and work with the data experts to prioritize the backlog of viable concepts.

Although last year's data survey suggested that the demand for analytics translators would exceed that of data scientists and data engineers, we find that this year the role is deemed slightly less valuable than last year. Still, however, only a small percentage of respondents (17%) see no value in having an analytics translator within their organization. Larger companies (1000+ employees = 39% valuable and only 7% not) ) benefit more from such a role than smaller organizations (10-50 employees = 25% valuable and 34% not).

So, analytics translators are still seen as valuable in the struggle to implement data-driven solutions. However, we see the rise of another role that is becoming more successful in bringing data and analytics to the business.

Team and Organization / We have fully automated data ingestion and transfer within our organization.



Team and Organization / How valuable would an analytics translator be within your organization?



#### **Fully Automated Data Ingestion and Transfer**

When it comes to increasing data quality and the automation of data pipelines, we see a trend towards a new role: the analytics engineer. This role brings the principles of software development to the more traditional data and analytics space. Results from this year's survey suggest that going from raw data to usable data is still a huge challenge for most organizations. The analytics engineer role could form the much-needed bridge between data engineering and the business.

Fully automated data ingestion and transfer is not the norm across all organizations, as indicated by the survey results. Only 15% of all respondents agree that their organization's data ingestion and transfer is fully automated, whereas 39% partially or fully disagree with this statement. A similar trend exists across all company sizes. The utilities sector seems to be most ahead in this case, with 54% indicating their data transfer and ingestion is automated.

The sectors next in line are utilities (55%), healthcare (47%), finance (48%), and retail (48%). Sectors most behind in automated data ingestion and transfer are manufacturing (38%) and telecom (37%). These sectors require a lot of work to make high-quality data readily available for their businesses.

### The Use of Data for Dashboards and Reports and to Develop Predictive Models

Many professionals agree that they use data for dashboards and reports (85% either partially or fully agree). Slightly less, but still a majority indicate that they also use data to develop predictive models (54%).

Note the interesting difference between professionals with a position in the C-suite and data professionals. Although 53% of professionals responsible for data applications fully agree that they are using data for dashboards and reports, only 37% of professionals with a management position state the same. When asked if they also use data to develop predictive models, this gap increases. Here, 27% of professionals responsible for data applications and 36% of professionals developing these applications fully agree with this statement, compared to only 18% of C-level professionals. These results suggest a knowledge gap between data professionals and management in how they are utilizing available data in their organization.



Team and Organization / We use data to develop predictive models.



Team and Organization / We use data for dashboards and reports.

### Organizing, Developing and Creating Opportunities from Data and AI Projects

Results from this year's survey show that most experiments are run together with the business. Less frequently, organizations will run experiments mainly in their innovation hub. When it comes to developing and implementing AI applications with the business, the leading industries include utilities (46%), telecom (37%), and healthcare (32%). The manufacturing industry and media and entertainment are lagging the furthest behind in data science activities; 30% of both sectors do not engage in data science activities at all. Although more than a quarter of organizations with less than 1000 employees indicate not implementing data science and advanced analytics projects, it is very rare for larger organizations not to engage in these activities (9%).

#### Using External Consultants to Develop Data-Driven Use Cases

Organizations do not always develop data-driven use cases in-house, with 44% of respondents indicating that their organization uses external consultants. However, 45% do not use external consultants. Mainly organizations within the financial (55%), telecom (50%), retail (50%), and travel and leisure (56%) sectors are using external consultants for developing these data-driven use-cases. The use of external consultants rises as companies grow in size, with 58% of large companies (1000+ employees) stating they are using external consultants. In contrast, in the smallest organizations (<10 employees), only 25% of respondents indicate hiring external consultants to develop data-driven use cases.





#### **Attracting Talented Data Professionals**

This year's survey asked respondents what they deemed most important for attracting new, talented data professionals. In line with previous years, freedom to experiment ranked highest with team expertise and flexible working hours, coming in second and third. Overall, all scores have increased compared to last year, suggesting that organizations are putting more effort into attracting new talent. Overall, all scores have increased compared to last year, suggesting that organizations are putting more effort into attracting new talent. The increase in the importance of flexible working hours could be linked to remote working being the norm across many organizations due to the COVID-19 pandemic. Compared to last year's survey, the most significant increase is that of salary, this year scored a 7.6 compared to the previous year's 6.7.

The same industries hit most by the pandemic also tend to score higher across many aspects around attracting talent. Within the tourism sector, freedom to experiment ranks 9.3, flexible working hours 8.9, and a transparent organization 8.6. When it comes to salary, organizations that scored the highest were mainly within the retail (8.3), Telecom (8.3), and utilities (8.2) sectors.

< Benefit > < Score 1 to 10 >	Score 2018	Score 2019	Score 2020	Score 2021
< Freedom to experiment >	8.4	8.4	8.2	8.2
< Team expertise >	7.8	8.0	7.3	8.0
< Flexible working hours >	7.8	7.9	7.3	8.0
< Transparent organization >	7.6	8.0	7.1	7.8
< Salary >	7.4	7.5	6.7	7.6
< Training budget >	7.4	7.5	6.7	7.4
< Agile way of working >	7.0	6.8	5.9	6.8
< Commuting time >	6.2	6.4	5.3	6.2
< Company car >	4.6	4.5	3.6	4.0





# Data and Technology Trends The Most Popular Data Technologies

Unchanged from last year, Python and R are still the most popular data programming and analysis technologies. Next in line are IBM's SPSS (18%, down from 25% last year), SAS (15%, down from 21% last year), and Apache Spark (14%, down from 18% last year). This year, the only technology that has risen in popularity has been Databricks (14%, up from 12% last year). Databricks is being utilized especially often within corporations with over 1000 employees (25%). We also see above-average popularity for Python (68%), R (53%), SAS (28%), Spark (23%), SPSS (23%), and Apache Kafka (18%) among these larger corporations this year.

- The most popular data technologies remain Python (58%) and R (42%). The use of programming language differs with company size and industry.
- PowerBI and Tableau remain the top two data visualization tools regardless of company size and industry. Google Data Studio is gaining popularity quickly, mainly among SMEs.
- Increasingly, more trust is being placed in cloud solutions.
   The use of private clouds is decreasing, with Amazon, Microsoft, and Google sharing most of the public cloud market share.
- Taking data applications into production and deep learning and neural networks rank as the most promising data trends. Blockchain, however, has lost some perceived potential due to a lack of successful use cases.

< Python >	<b>→</b> 1	58%	2021	1 71%	2020
< R >	<b>→</b> 2	42%	2021	<sup>2</sup> 56%	2020
< SPSS >	→ 3	18%	2021	<sup>3</sup> 25%	2020
<sas></sas>	<b>→</b> 4	15%	2021	4 21%	2020
∢Apache Spark∍	→ 5	14%	2021	5 18%	2020
< Databricks >	16	14%	2021	10 12%	2020
< Spark >	↓ 7	12%	2021	6 17%	2020
< Elastic >	18	12%	2021	<b>7</b> 17%	2020
<kafka⇒< th=""><th>19</th><th>11%</th><th>2021</th><th>8 16%</th><th>2020</th></kafka⇒<>	19	11%	2021	8 16%	2020
< Matlab >	↓ 10	10%	2021	9 15%	2020
< Airflow >	↑ 11	8%	2021	12 10%	2020
< Teradata >	↑ 12	6%	2021	13 7%	2020
< Cloudera >	↑ 13	5%	2021	16 6%	2020

< IBM Watson >	↓ 14	5%	2021	11 12%	2020
< Apache Cassandra >	↑ 15	5%	2021	17 6%	2020
<neo4j></neo4j>	↓ 16	5%	2021	14 6%	2020
< Rapidminer >	<sup>↑</sup> 17	4%	2021	19 5%	2020
< H2O >	† 18	3%	2021	22 3%	2020
< MapR >	↓ 19	3%	2021	15 6%	2020
< Dataiku >	î 20	3%	2021	21 3%	2020
< Pentaho >	↓ 21	2%	2021	18 5%	2020
< Talend >	↓ 22	2%	2021	20 5%	2020
< Datameer >	↑ 23	2%	2021	24 2%	2020
< Apache Flink >	↓ 24	2%	2021	23 2%	2020
< DataStax >	→ 25	2%	2021	25	2020



#### The Most Popular Data Visualization Tools

The right tool stack is essential for creating data science products, but being able to visualize and analyze the increasing amounts of data organizations are dealing with is often equally important. Adequately visualizing data allows organizations to reveal new insights and further improves collaboration and understanding within and between business units. This year, results show two prominent visualization tools organizations use most often: Microsoft Power BI and Tableau. Among larger corporations (1000+ employees), PowerBI is becoming even more popular (PowerBI 58% and Tableau 30%). Three additional tools that stand out as well within larger corporations are Qlik (21%), SAS (16%), and Ggplot (15%). In this year's survey, an interesting newcomer is Google Data Studio (10% across all respondents). After coming out of beta in 2018, the visualization tool by Google has rapidly gained popularity, especially in small and medium-sized enterprises. Of respondents working at organizations with 51-100 employees, 18% indicate they use Google Data Studio for their data visualization.

Industry differences are showing particular popularity in some specific tools. Within the education industry, 15% of respondents indicate using IBM's BI solution for their data visualization. Plotly is especially popular among professionals working in the manufacturing industry (20%).

However, Plotly remains unused within the financial sector (0%). Although Looker is rarely used across most industries, professionals working in the media and entertainment sector have adopted it more regularly in their work (18%).

#### **Cloud Security**

Increasingly more data is being stored and processed in the cloud. This trend reflects this year's results when it comes to cloud security. Overall, respondents are putting more trust in the cloud when comparing it to on-premise solutions. Last year, 17% of respondents totally agreed that storing data in the cloud is just as secure as storing data on-premise. This year, we've seen an increase by 10% percent, with 27% of respondents completely agreeing with this statement. The amount of respondents disagreeing with this has also decreased from 17% to 9%. Data and Technology Trends / Storing data in the cloud is just as secure as storing data on-premise.



Across most industries, there's a predominantly positive view towards cloud security, except within the media entertainment, and utilities sectors, where 10% of respondents completely disagreed. Organizations operating within the tourism industry put the most trust in cloud solutions, with 100% of respondents in this sector indicating that they agree with this statement. Finally, results show the highest skepticism amongst C-level professionals, with 15% disagreeing that cloud solutions are as secure as on-premise solutions. Contrastingly, the professionals developing data applications were the most positive, with 48% agreeing that the cloud is just as secure.

#### **Cloud Solutions**

The 2020-2021 data survey shows a noticeable decrease in the use of private clouds. Although they can be seen as more secure and make organizations less dependent on other parties, they can be extremely costly and inconvenient. The largest portion of market share is, like last year, being shared by Microsoft Azure, Amazon Web Services, and Google's Cloud Platform. Noticeable is Oracle Cloud (9%), which was not present in last year's survey. Although Alibaba's Cloud is the most widely used cloud platform solution in Asia-Pacific, the platform still hasn't gained much popularity in the Netherlands, staying at 1% like the previous year.



#### Data Trends: Taking Data into Production Ranked Most Important

Although professionals collectively find taking data applications into production one of the biggest challenges, they also see it as one of the most important data trends in this year's survey. They deemed deep learning and neural networks as an almost equally important data trend.

Compared to last year, blockchain declined the most, with only half of the survey respondents perceiving it as one of the most important trends, going from 18% to 9%. This could be due to a lack of promised use cases that are yet to come to fruition over the last few years. Only in the education industry is there a slightly higher perceived importance (15%). Moreover, blockchain was seen as a less important trend the smaller the organization gets. Where 10% of corporates (1000+ employees) see blockchain as an important trend, only 4% of organizations with 10-50 employees acknowledge the potential of this technology.

Cloud technology is seen as a more promising trend this year, going from 19% last year to 29%. This trend is seen as especially important for creating future opportunities in the finance (45%) and media and entertainment industries (36%).



35%	< Taking data applications into production
33%	< Deep learning and neural networks >
29%	< Cloud technology >
17%	< Internet of things >
9%	< Blockchain >
6%	< Quantum Computing >
5%	< Edge Computing >
4%	< Other >

Taking Data Applications in Production Seen As The Most Important Trend



# The Data and Al Survey Contributors

The Big Data Survey is an initiative of the leading Dutch trade show on data, Big Data Expo, and the data consultants from GoDataDriven.

#### GoDataDriven

Increase your AI innovation speed and operate faster and smarter with ease. Since 2009, GoDataDriven has provided services and solutions for modern organizations to realize true business value from big data and AI. GoDataDriven clients include Heineken, Tele2, ING, NS, eBay, RTL, and Unilever.

> godatadriven.com

#### **Big Data Expo**

The Big Data Expo is a platform where people meet and knowledge and data converge. It recognizes and advances big data as a leading technology trend, with scientific studies that show a clear link between the use of big data and significantly improved efficiency. It's no longer debatable—companies must use big data. The question now is, how?

The Big Data Expo brings the brightest minds on this topic together for two days for the most prominent big data event in the Benelux. Get inspired by global-leading companies, gain ready-to-use tools, discover the added value of big data for your organization, and boost your competitive advantage.

> bigdata-expo.nl

#### Computable

Computable.nl is the platform for ICT professionals. It provides independent daily ICT news, articles and expert opinions, career information, white papers, product updates, discussions, trending topics, and case studies, as well as an extensive archive. > computable.nl



Xebia

Creating Digital Leaders.