



Data and Analytics in a Digital-First World

A global study of data workers' activities and the democratization of data and analytics in organizations

RESEARCH BY:



Stewart Bond
Research Director, Data Integration and
Data Intelligence Software, IDC





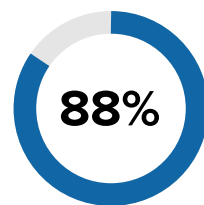
Navigating This InfoBrief

Click on titles or page numbers below to navigate to each.

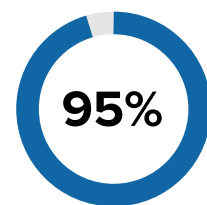
Executive Summary.....	3	Complexity and Multiple Tools Impede Productivity.....	12
Data and Analytics Power the Future Enterprise in a Digital-First World.....	4	Obstacles to Democratizing Data and Analytics.....	13
Generation Data: Data Workers Deliver Analytical Insights in a Digital-First World.....	5	Most Organizations Lack Advanced Skills in Data and Analytics.....	14
Tools Data Workers Use to Deliver Data and Analytics Outcomes.....	6	Skills and Technology Improvements Needed to Address Data and Analytics Complexity.....	15
78 Million Data Workers Are Advanced Spreadsheet Users.....	7	Regional Highlights.....	16
Time Spent in Spreadsheets Is Often Ineffective.....	8	Essential Guidance.....	17
Data Workers Expect to Spend More Future Time in High-Level Activities.....	9	Demographics.....	18
Time Is Wasted Every Week Because Activities Take Too Long.....	10	Industries.....	19
Productivity Suffers as Workers Lose 800 Hours Annually.....	11	Microsoft Excel Dominates the Spreadsheet Alternatives in Use for Advanced Analytics.....	20
		Definitions.....	21
		About the Analyst.....	22

Executive Summary

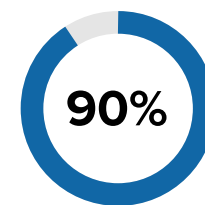
70% of organizations want to be more data-driven and improve employee data and analytics skills, but challenges exist.



88% of organizations face data and analytics technology constraints.



95% of organizations report being challenged by data in creating analytic outcomes.



90% of organizations use multiple tools in data and analytics activities.



The spreadsheet is ubiquitous in data and analytics activities.



Complexity, constraints, and lack of analytics automation and skills hurt productivity.

- ▶ **78 million data workers are advanced spreadsheet users.**
- ▶ Spreadsheets are dominant for gathering and preparing data but are used less often for data science and app development.
- ▶ **Users waste 7 hours per week** repeating work as data changes.

- ▶ **One-third of workers' time** is lost because activities cannot be completed in reasonable or expected time.
- ▶ **62.4 billion hours spent on data and analytics are lost annually worldwide.**
- ▶ Democratization requires unified data and analytics tools and upskilling of data workers.

Data and Analytics Power the Future Enterprise in a Digital-First World

A digital-first company, government, or person asks, “Is there some digital-based capability or enhancement that could improve our lives and desired outcomes?”

Leveraging data and analytics is key to success.

- ▶ 71% of respondents report their executives and top managers want to be more data driven.
- ▶ Democratizing data and analytics is required at digital-first scale, yet 88% of respondents face technology constraints, and 95% report data challenges.

The data and analytics experience needs to be improved.

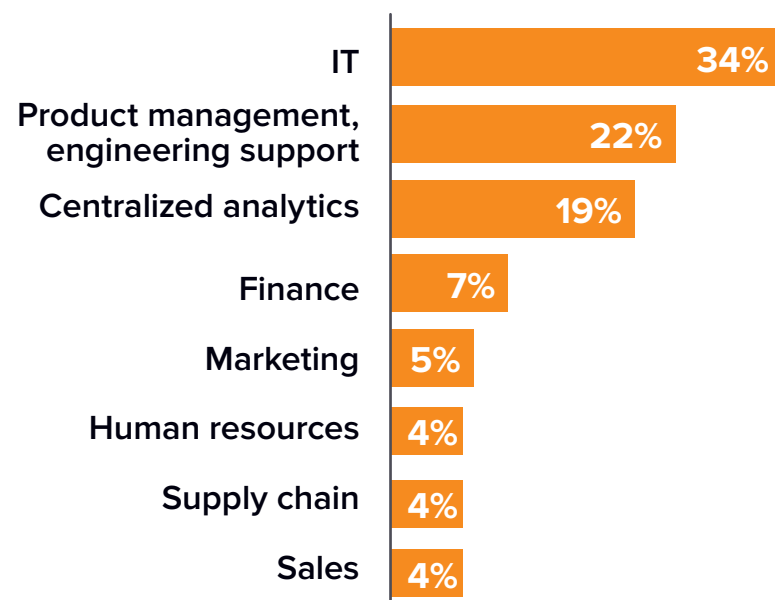
- ▶ Only 40% of people who work in data and analytics enjoy what they do.
- ▶ 91% admit to having skills gaps.

Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

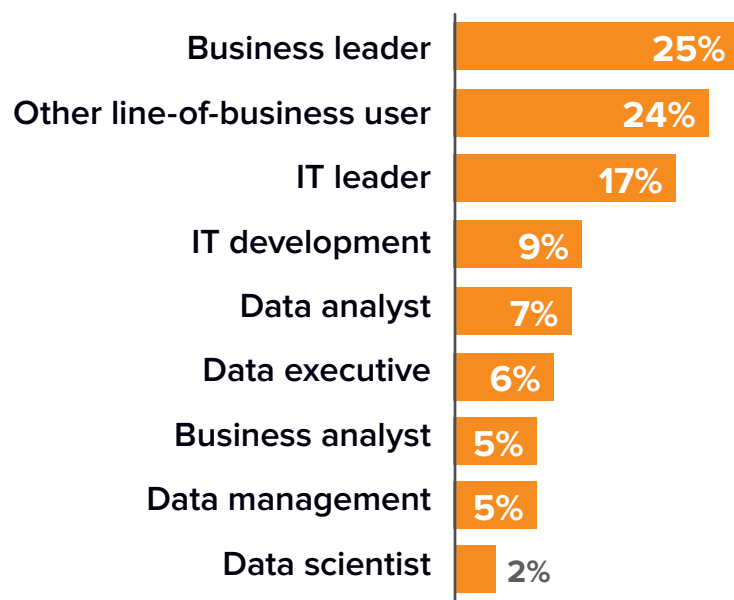
Generation Data: Data Workers Deliver Analytical Insights in a Digital-First World

Data workers are spread across functional areas, roles, and industries.

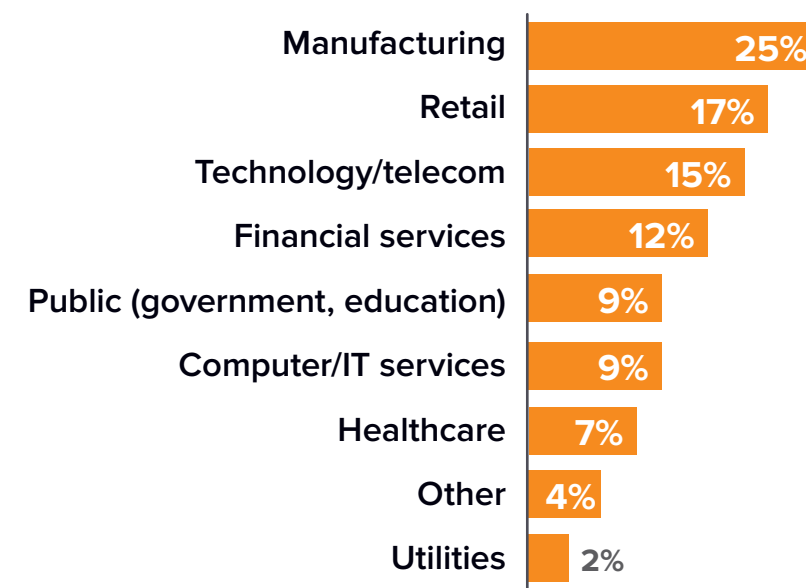
Functional Area



Organizational Role



Industry Distribution



Source: IDC's Worldwide Data Science and Analytics, November 2021, n = 1,117

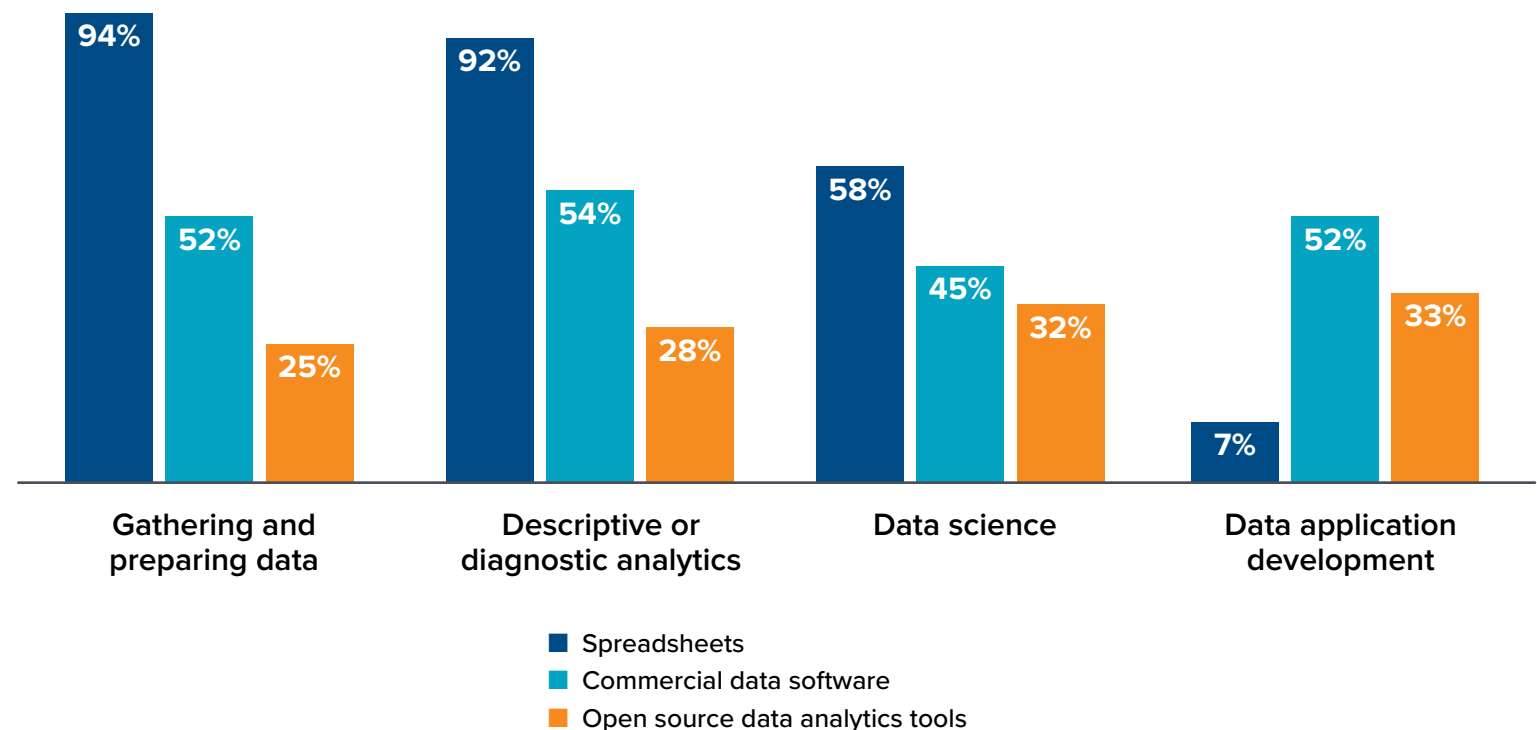
Tools Data Workers Use to Deliver Data and Analytics Outcomes

Data workers use spreadsheets and commercial and open source software tools. Spreadsheets are dominant for gathering data and descriptive analytics but are much less common for data science and rare for data application development.

On average, 3 or more commercial tools are being used in addition to spreadsheets and open-source software across all data and analytic activities.

Q. Which of the following software technology alternatives are available to you to perform these data activities?

(Percentage of respondents)



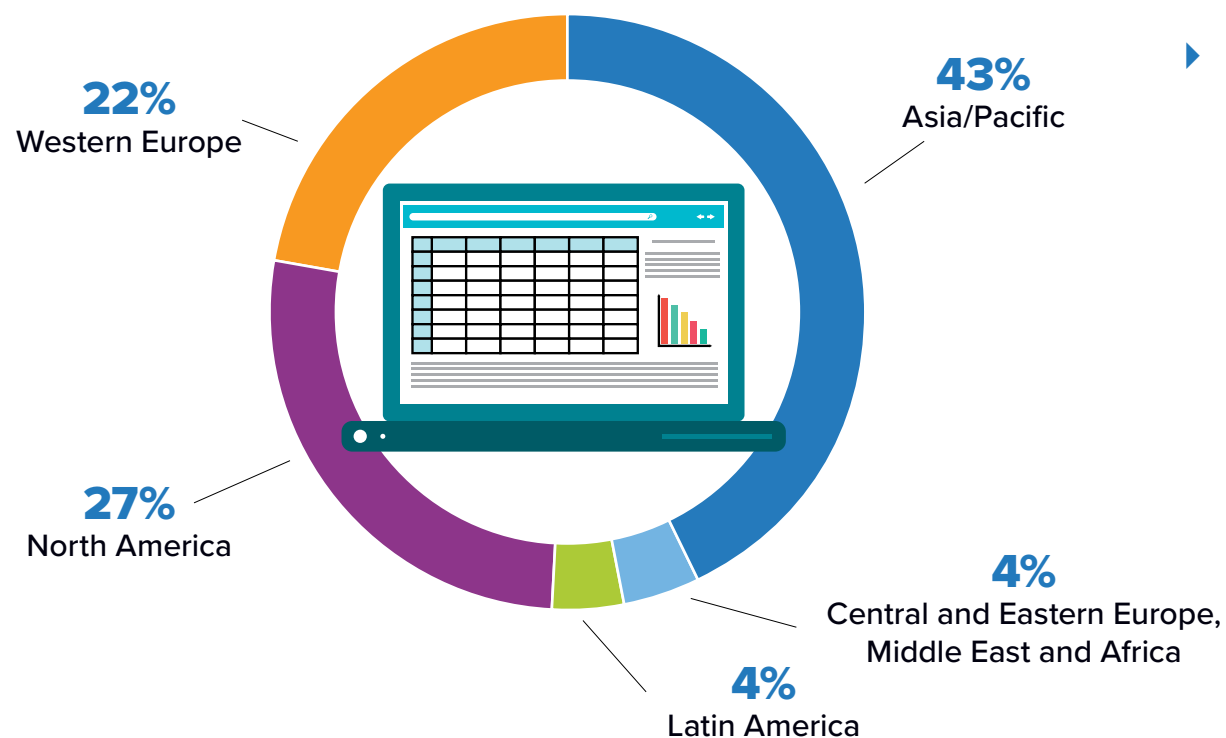
Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

78 Million Data Workers Are Advanced Spreadsheet Users

Approximated worldwide in organizations with 10 to 10,000+ employees

Worldwide Distribution of Advanced Spreadsheet Users

- ▶ Nearly half (~45%) of all spreadsheet users use advanced data and analytics features and functions.
- ▶ Microsoft Excel is used by **94%** of advanced spreadsheet users.

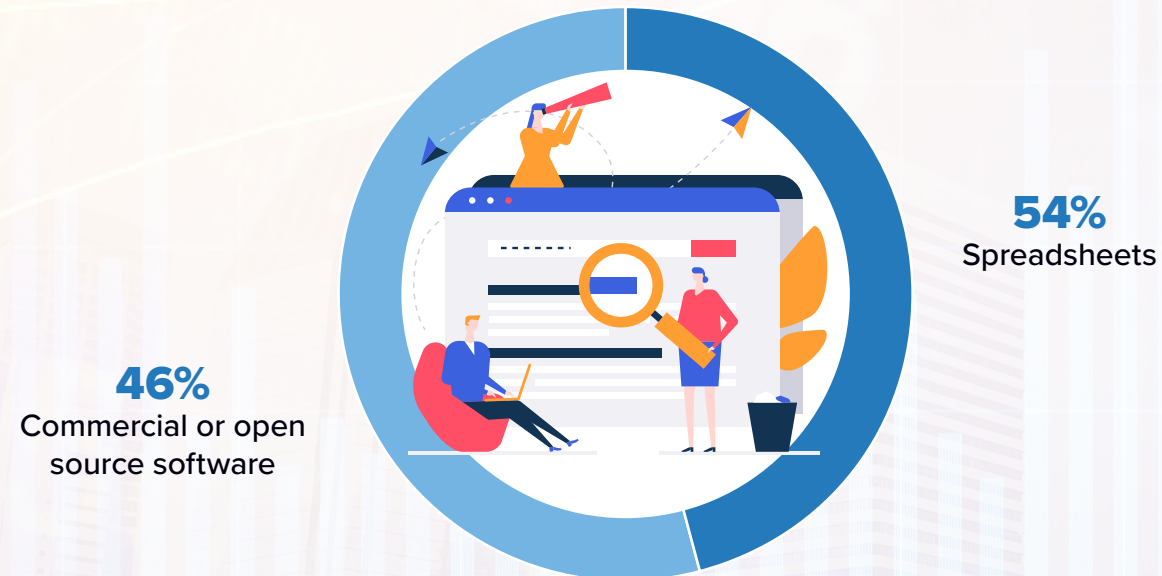


- ▶ The percentage of advanced spreadsheet users has **increased 15%** in two years (2019–2021), demonstrating a measured increase in the number of data workers.

Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117
IDC's *Worldwide Data Science and Analytics*, February 2019, n = 836

Time Spent in Spreadsheets Is Often Ineffective

Time Spent On Data by Tool



Data workers spend **27%** of their time in spreadsheets repeating the same or similar activity steps every time a data source has been updated or refreshed.

This is equivalent to an average of 7 hours per week.

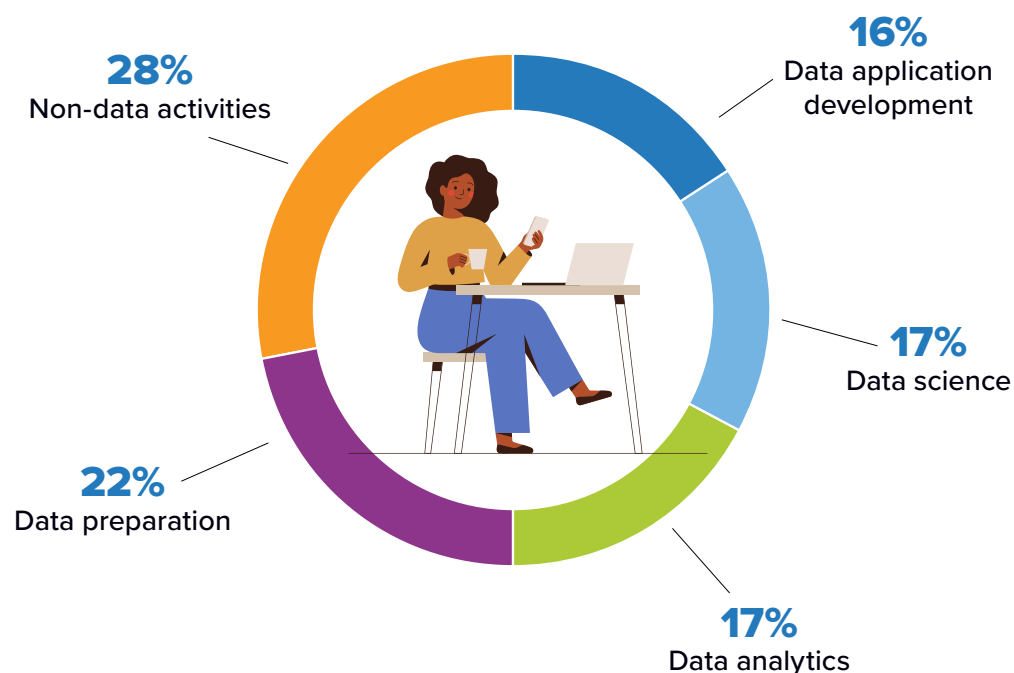
Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

Data Workers Expect to Spend More Future Time in High-Level Activities

Data workers anticipate spending more of their future time in data science and data application development activities rather than in prep and descriptive analytics. A large part (72%) of a data worker's weekly time is spent in data-related activities.

Distribution of Weekly Time by Activity

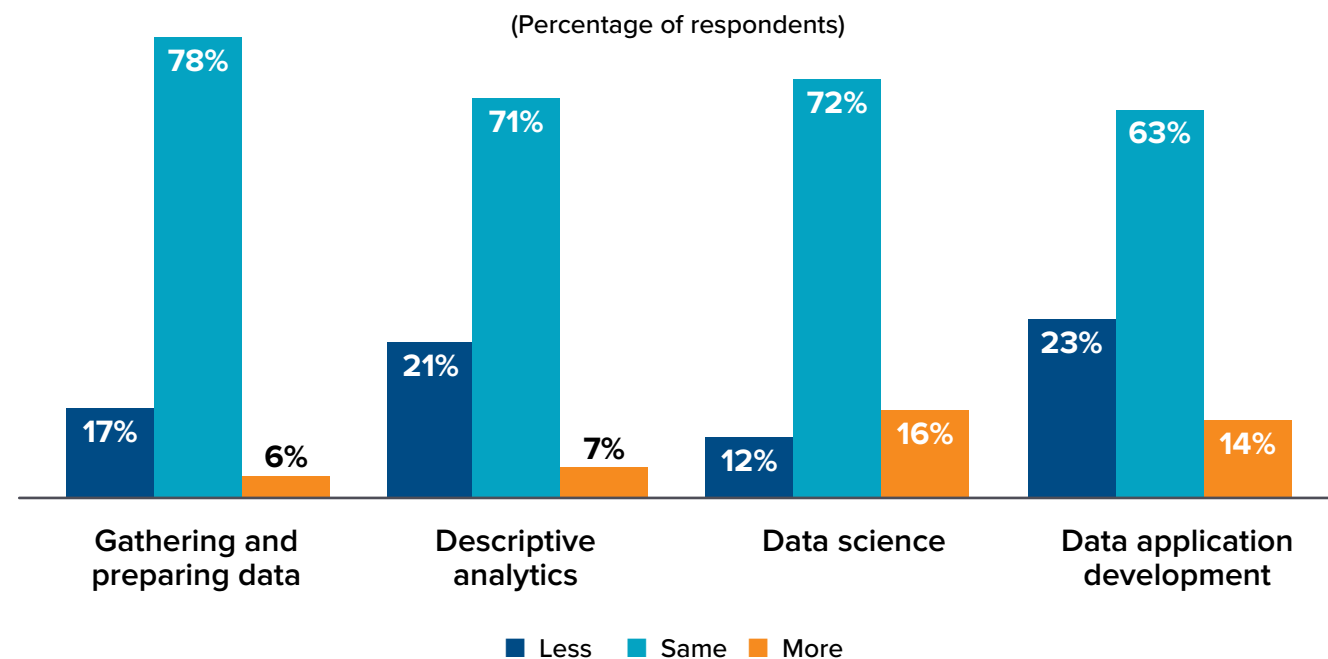
Q: Thinking about a typical work week, how many hours per week do you spend on the following activities?



Source: IDC's Worldwide Data Science and Analytics, November 2021, n = 1,117

Future Time Expectations

Q: Looking ahead, how do you expect your efforts to change for the following activities in the next 12-18 months?

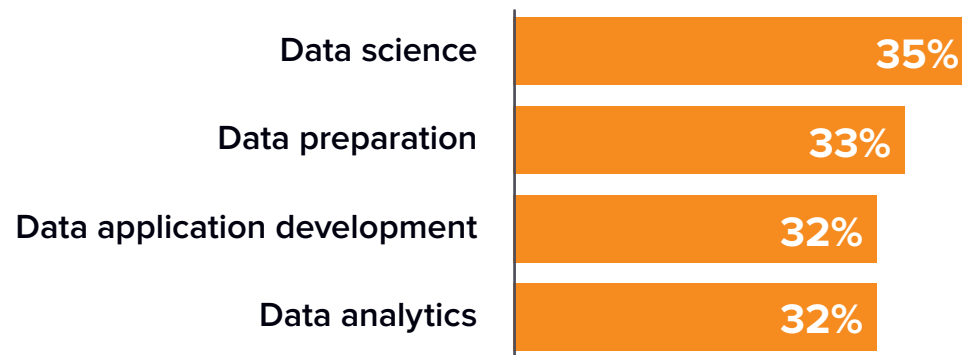


Time Is Wasted Every Week Because Activities Take Too Long

Data science activities require the most data exploration and preparation, resulting in more unsuccessful time.

Average Percentage of Unsuccessful Time Spent by Activity

Q: What percentage of the time are you unsuccessful in these activities?



Weekly Time Spent in Data Activities



61% of data workers perform activities in spreadsheets only. This includes 13% who have access to alternative tools.

On average, this is equivalent to nine hours every week per person.

Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

Productivity Suffers as Workers Lose 800 Hours Annually

Productivity loss is caused by complexity and challenges associated with data and analytics democratization.

More than 62 billion hours are wasted every year worldwide because workers are not able complete tasks in a reasonable or expected amount of time. Time is also lost from repeating work in spreadsheets every time an underlying data source changes.

Weekly hours spent repeating tasks in spreadsheets	7
Weekly hours of unsuccessful data activity effort*	9
Total weekly hours lost	16
Weeks per year	50
Number of advanced spreadsheet users worldwide	78 million
Worldwide hours wasted annually	62.4 billion

Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

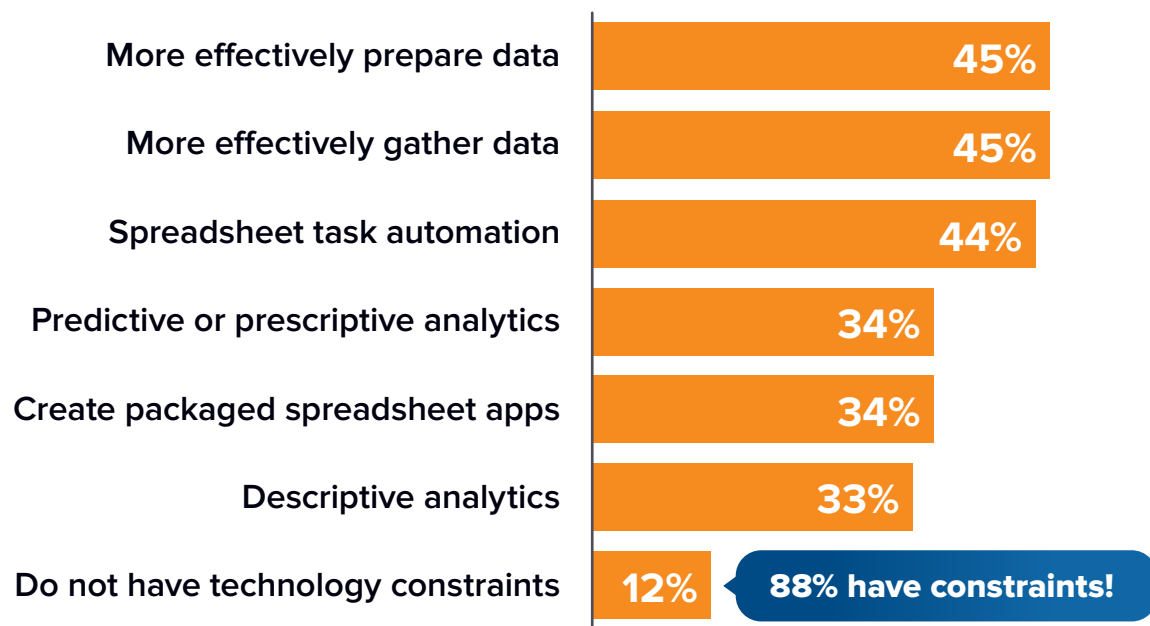
*Unsuccessful time may not be mutually exclusive from time spent repeating activities in spreadsheets.

Complexity and Multiple Tools Impede Productivity

In their efforts to democratize data and analytics, 88% of organizations report that technology constraints limit improvement.

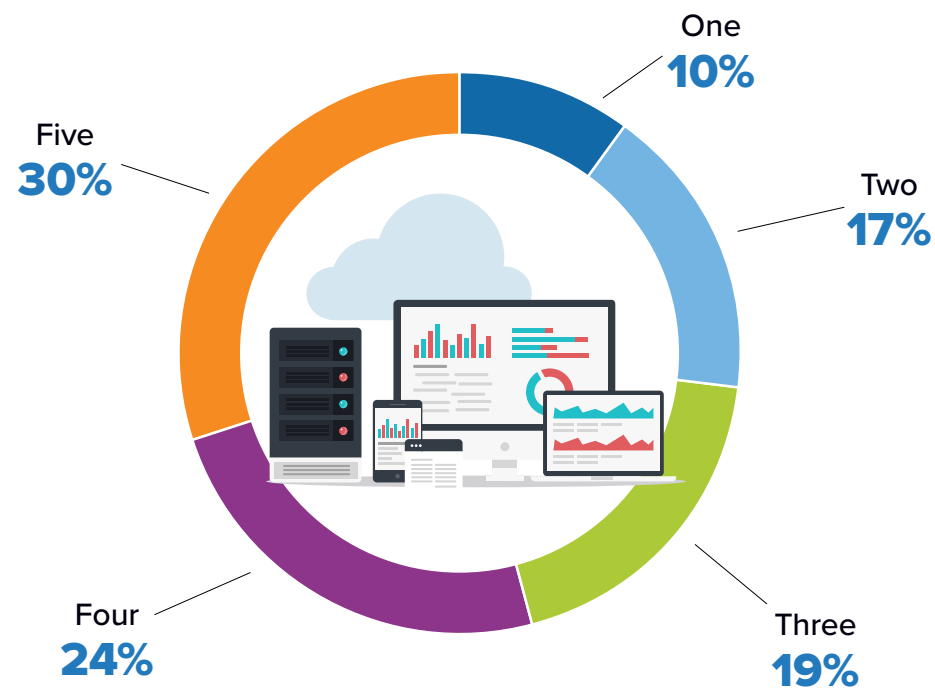
Technology Constraints on Improvement

Q. Which of the following activities would you like to improve on but are limited by technology constraints?



Number of Commercial Tools in Use

Q: How many different commercial data or analytics software vendors are used for each activity?



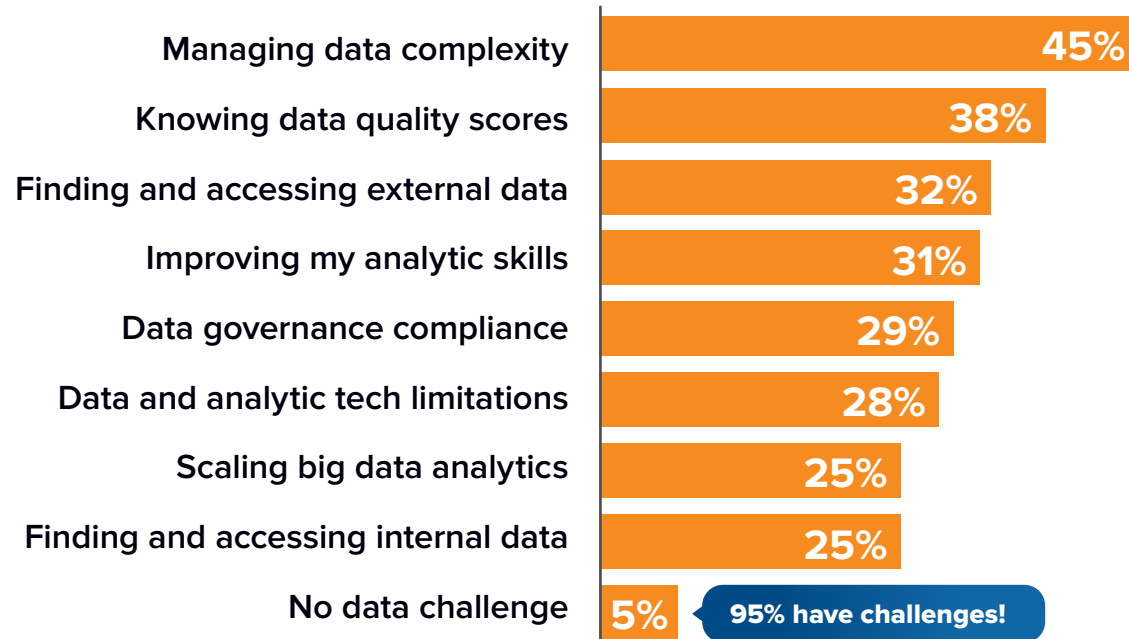
Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

Obstacles to Democratizing Data and Analytics

Almost all organizations (95%) report difficulties with data when performing analytic activities, mainly in complexity, data quality, data access, and analytics skills.

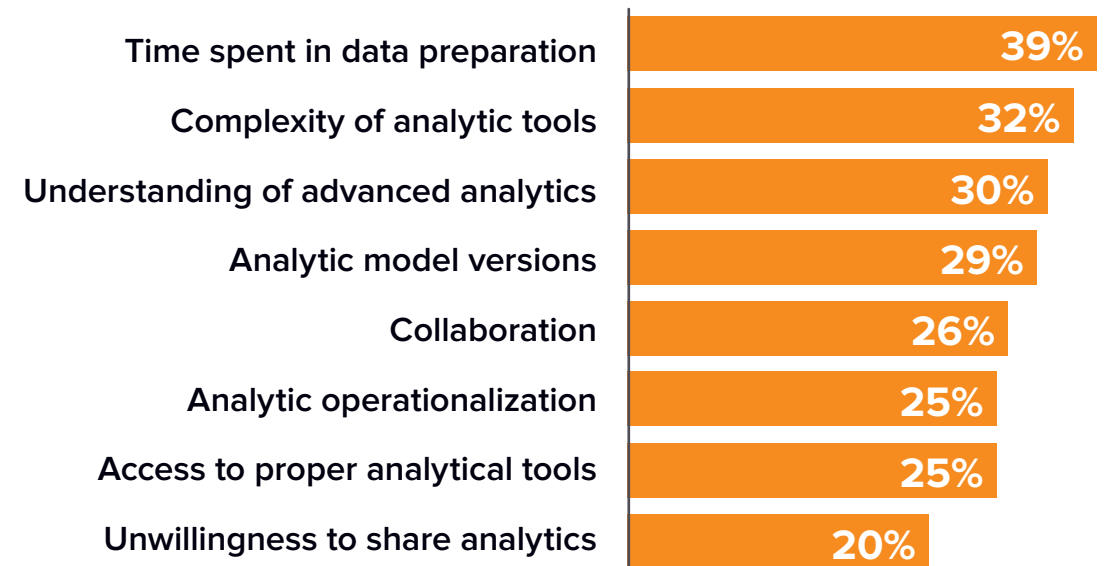
Data-Specific Challenges

Q. What are the top 3 data-specific challenges you have when performing data analytics activities?



Non-Data-Specific Challenges

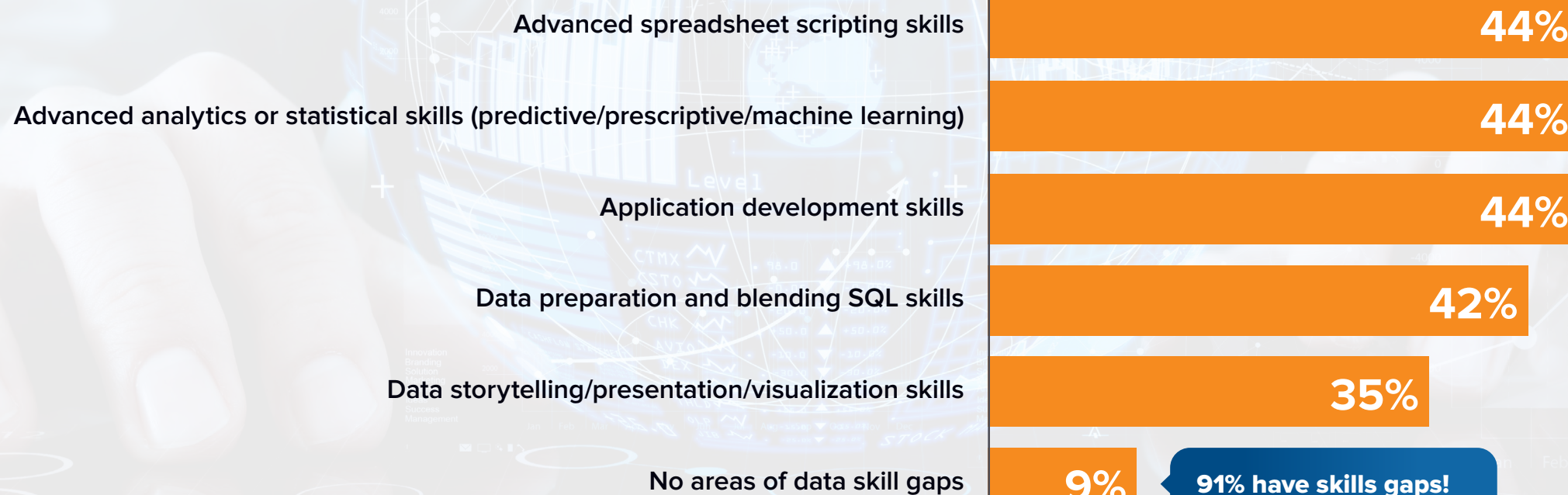
Q. What are the top 3 primary non-data-specific challenges you have when performing data analytics activities?



Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

Most Organizations Lack Advanced Skills in Data and Analytics

Q. Please select the top 3 areas of data skills gaps that you believe exist in your organization?

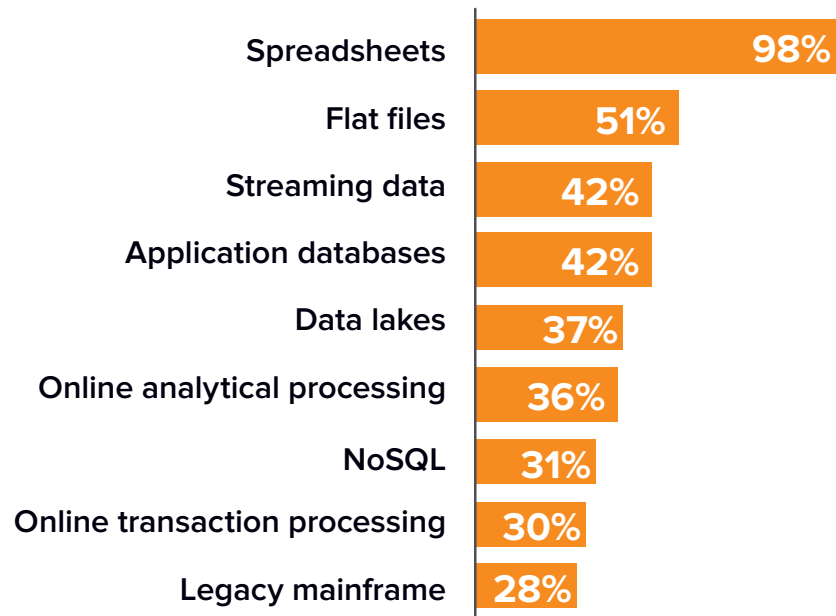


Skills and Technology Improvements Are In Demand to Address Data and Analytics Complexity

Data pipelines are processing highly distributed, diverse, and dynamic data at scale and delivering it to multiple assets for consumption.

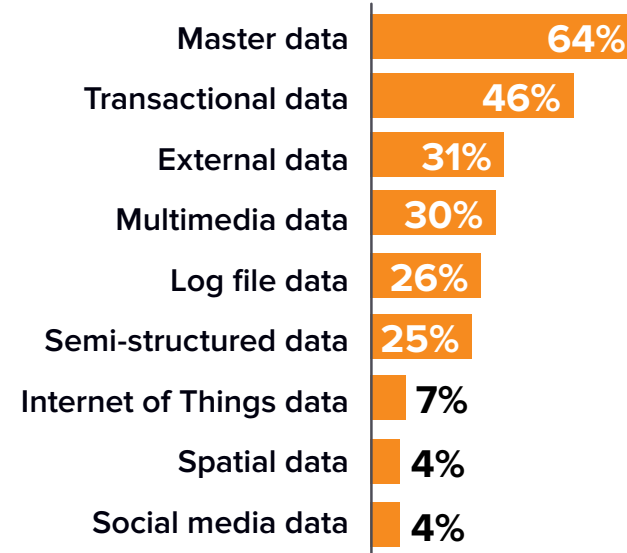
Data Sources

Q. Select all the data management technologies that are being used to store and/or process data, including cloud and on-premises implementations.



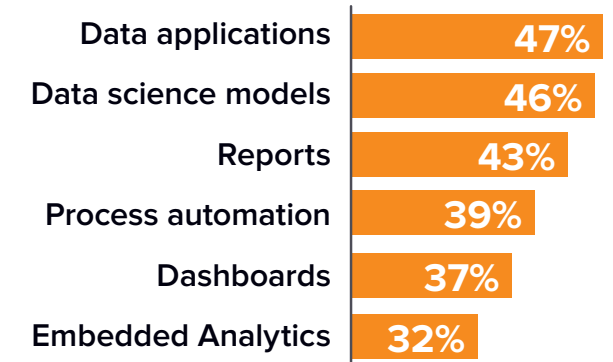
Data Types

Q. What types of data are being integrated or blended as part of your data preparation for analytics or data science?



Analytics Consumption

Q. Select the top three ways your data and analytics outputs are being operationalized.



Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

Regional Highlights

North America

- Lower-than-average ratio of advanced spreadsheet users
- Biggest employee challenge compared to other regions: improving data and analytics skills
- Largest barrier to implementing alternative data and analytic tools: corporate IT standards and policies
- More data and analytics workloads moved to the cloud compared to other regions
- Highest belief that a complete overhaul of data and analytics is required compared to other regions

Western Europe

- Second-lowest ratio of advanced spreadsheet users
- Lowest number of commercial tools in use
- Least significant part of data and analytics strategies now compared to before COVID-19: improving employee skills
- Least resistance to implementing alternative data and analytics tools compared to other regions
- Least desire to improve data and analytics compared to other regions

Central and Eastern Europe, Middle East, and Africa

- Lowest ratio of advanced spreadsheet users
- Least successful in completing data and analytics activities compared to other regions
- Big focus on improving employee data and analytic skills

Latin America

- Higher-than-average ratio of advanced spreadsheet users
- Most successful in completing data and analytics activities compared to other regions
- Mainframes a significant source of data for analytics

Asia/Pacific

- Highest ratio of advanced spreadsheet users
- Highest number of commercial tools in use
- Most time spent in data science activities compared to other regions
- Primary data and analytics strategy now versus pre-pandemic: improving employee skills
- Largest barrier to implementing alternative tools: compatibility with data technology

Essential Guidance

Democratize data and analytics with tool reconciliation, consolidation, and worker upskilling.

- Prioritize solutions that offer unified and automated data prep and analytic experiences.
- Seek solutions that enable upskilling to improve data literacy and analytic outcomes.

Understand the skill sets of data natives and implement solutions that enable upskilling.

- Data natives want to improve analytics and data application development skills.
- Data literacy continues to be a challenge but is now mandatory as decisions are increasingly data driven.

















Review your data and analytics environment and understand the challenges data natives face.

- A multiplicity of tools impact effective data operations and analytic processes.
- Data sources, formats, quality, scale, compliance, processing, and consumption are too complex.
- Manual and repetitive activities are used instead of analytics process automation.
- Technology and skills constraints inhibit self-service data preparation, analytics, and data science.
- Spreadsheets are inadequate to perform data science and app development.

Demographics

Methodology

IDC conducted a very international sample in five regions: North America; Western Europe; Asia/Pacific; Latin America; and Central and Eastern Europe, Middle East, and Africa.

 United States: 25%	 Australia: 3%
 Canada: 6%	 China: 3%
 Germany: 5%	 India: 6%
 France: 4%	 Singapore: 3%
 United Kingdom: 9%	 Japan: 4%
 Italy: 4%	 Russia: 4%
 Brazil: 8%	 South Africa: 4%
 Mexico: 7%	 UAE: 4%

Total sample size: 1,117

All respondents:

- Are currently employed
- Use a spreadsheet (Excel, Google, Zoho, etc.) to analyze data for work
- Regularly use advanced features (pivoting data, VLOOKUP, cross-tabulation, statistical modeling, etc.) to analyze data

Size of Organization

The sample included a significant number of large organizations, including about half with more than 1,000 employees. Data was collected from all five regions of the world.

Q: What is the number of employees in your worldwide organization?

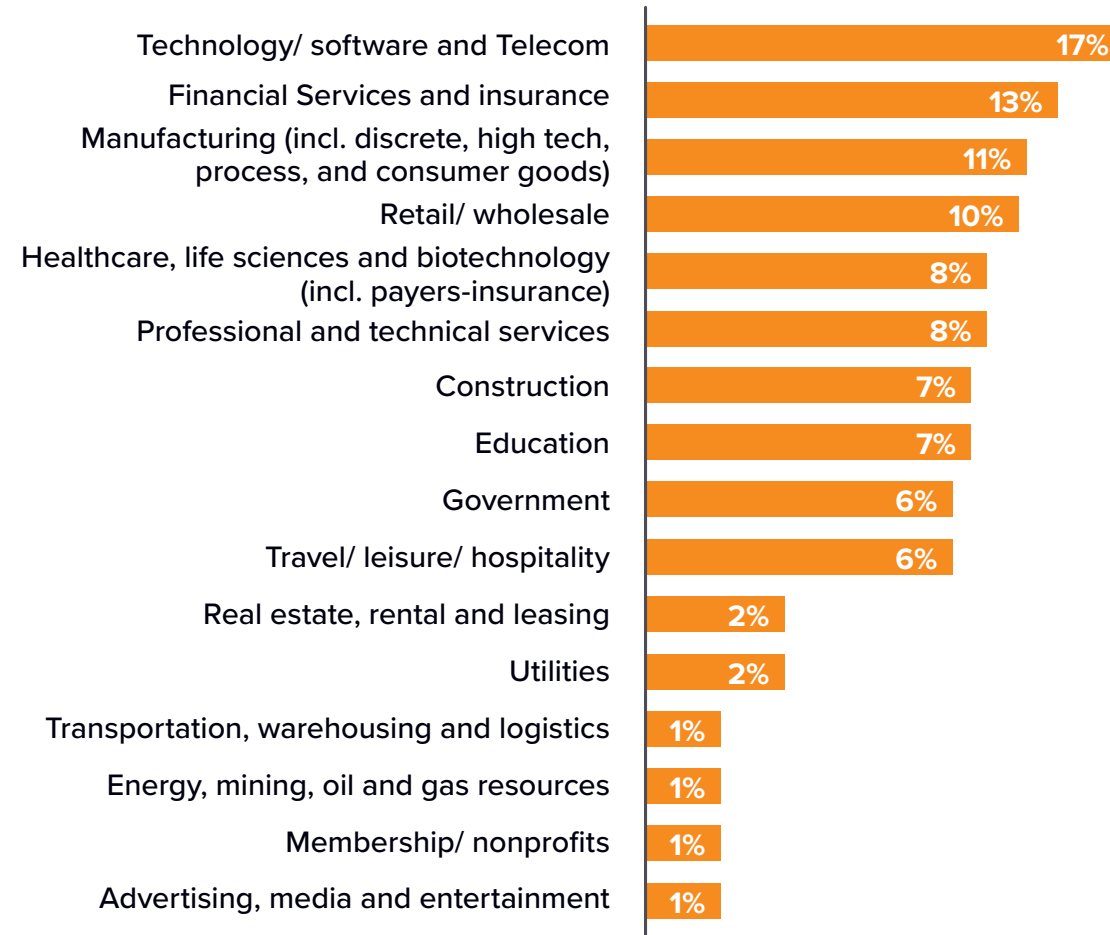
Less than 500	33%
500–999	17%
1,000–1,999	14%
2,000–9,999	23%
10,000–24,999	9%
More than 25,000	5%

	North America	Western Europe	Asia Pacific	Central and Eastern Europe, Middle East and Africa	Latin America
Small <i>(Less than 1,000 employees)</i>	50%	52%	40%	51%	58%
Large <i>(More than 1,000 employees)</i>	50%	48%	60%	49%	42%

Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

Industries

(Percentage of Respondents)

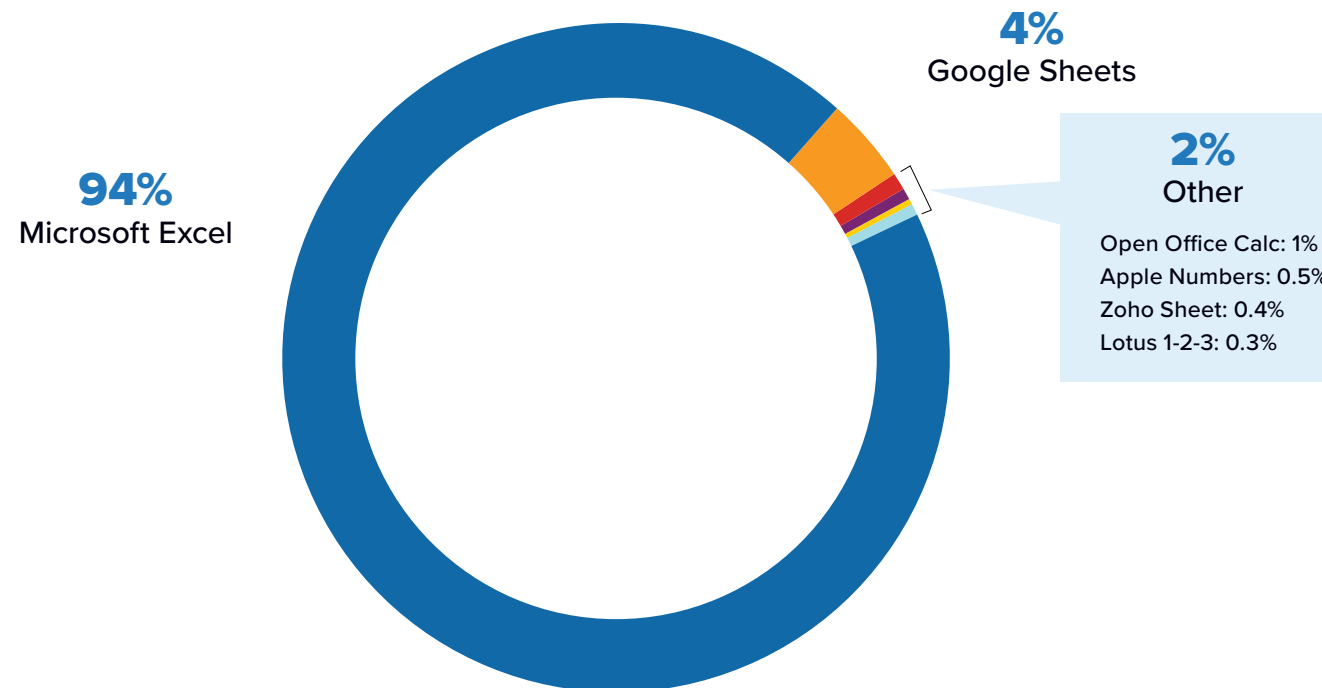


Grouping	Industry
Retail	<ul style="list-style-type: none"> • Travel/leisure/hospitality • Retail/wholesale
Manufacturing	<ul style="list-style-type: none"> • Construction • Manufacturing • Energy, mining, oil and gas • Transportation, warehousing and logistics
Public	<ul style="list-style-type: none"> • Education • Government
Financial services	<ul style="list-style-type: none"> • Financial services and insurance
Healthcare	<ul style="list-style-type: none"> • Healthcare, life science and biotechnology
Technology/telecom	<ul style="list-style-type: none"> • Advertising, media and entertainment • Technology/software and telecom
Computer services	<ul style="list-style-type: none"> • Professional and technical services
Utilities	<ul style="list-style-type: none"> • Utilities
Other	<ul style="list-style-type: none"> • Membership/nonprofits • Real estate, rental and leasing

Source: IDC's *Worldwide Data Science and Analytics*, November 2021, n = 1,117

Microsoft Excel Dominates the Spreadsheet Alternatives in Use for Advanced Analytics

Q. Which spreadsheet software do you use the most for advanced data and analytics activities?



Source: IDC's Worldwide Data Science and Analytics, November 2021, n = 1,117

Definitions

Data preparation: Shaping, cleansing, deduplicating, blending, filtering, and moving data from transaction systems to data lakes/warehouses for subsequent analysis.

Descriptive or diagnostic analytics: The act of aggregating, pivoting, and synthesizing results to determine what happened and why.

Data science: Methods include correlation, statistical modeling, regression analysis, predictive analysis, and data algorithm development.

Data application development: The process of building data-oriented applications and operationalizing data science models.

Advanced spreadsheet users: People who regularly use advanced features (pivoting data, VLOOKUP, cross-tabulation, statistical modeling, etc.) to analyze data for work.

Commercial data software tools: Any data software that companies license or sell to end users. Examples include Alteryx, Tableau Prep, Microsoft Power BI, SAS, SAP, etc.

Open source data analytics tools: Any data analytics tool that can be used, modified, enhanced, or distributed freely by users. Examples include Grafana, Redash, Pentaho, Apache Spark, etc.

About the Analyst



Stewart Bond

Research Director, Data Integration and Data Intelligence Software, IDC

Stewart Bond is Research Director of IDC's Data Integration and Intelligence Software service. Mr. Bond's core research coverage includes watching emerging trends that are shaping and changing data movement, ingestion, transformation, mastering, cleansing and consumption in the era of digital transformation. Having worked in the IT industry for over 25 years, from early experience in database and application development, through solution design and deployment, to strategic architectural consulting, Stewart has worked through some significant changes in the IT industry. His depth of field experience coupled with market insight gives him a unique perspective, valued by his customers and peers.

[More about Stewart Bond](#)

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. As a premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets, IDC's Custom Solutions group helps clients plan, market, sell and succeed in the global marketplace. We create actionable market intelligence and influential content marketing programs that yield measurable results.



 [@idc](#)

 [@idc](#)

[idc.com](#)

© 2022 IDC Research, Inc. IDC materials are licensed for external use, and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.

[Privacy Policy](#) | [CCPA](#)