

DECODING THE DATA UNIVERSE:

# The State of Data Science and Machine Learning

Mike Leone, *Principal Analyst*

ENTERPRISE STRATEGY GROUP

September 2023

## Research Objectives

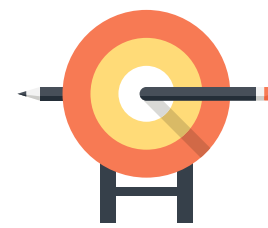
Several challenges are preventing organizations from successfully integrating machine learning (ML) models into their software development lifecycle. Bridging the gap between different skill sets, handling complex and large data sets, managing specialized hardware, and ensuring availability, scalability, and security in production collectively delay time to value and cause organizational bottlenecks.

Due to the increasing interest in and complexity of machine learning projects, organizations need improved agility, efficiency, and performance, with risk reduction through right-sized governance. Organizations recognize that they need clear data science and machine learning strategies. As part of these strategies, MLOps can provide a structured and standardized approach to developing, deploying, and maintaining ML models in production to see greater value. To gain further insight into these trends, TechTarget's Enterprise Strategy Group (ESG) surveyed 366 professionals at organizations in North America (US and Canada) involved with data science and machine learning technologies and processes, including potential responsibility for strategizing, evaluating, purchasing, building, and managing these technologies.

### This study sought to:



**Identify** investment plans, objectives, and challenges of data science and machine learning initiatives and projects.



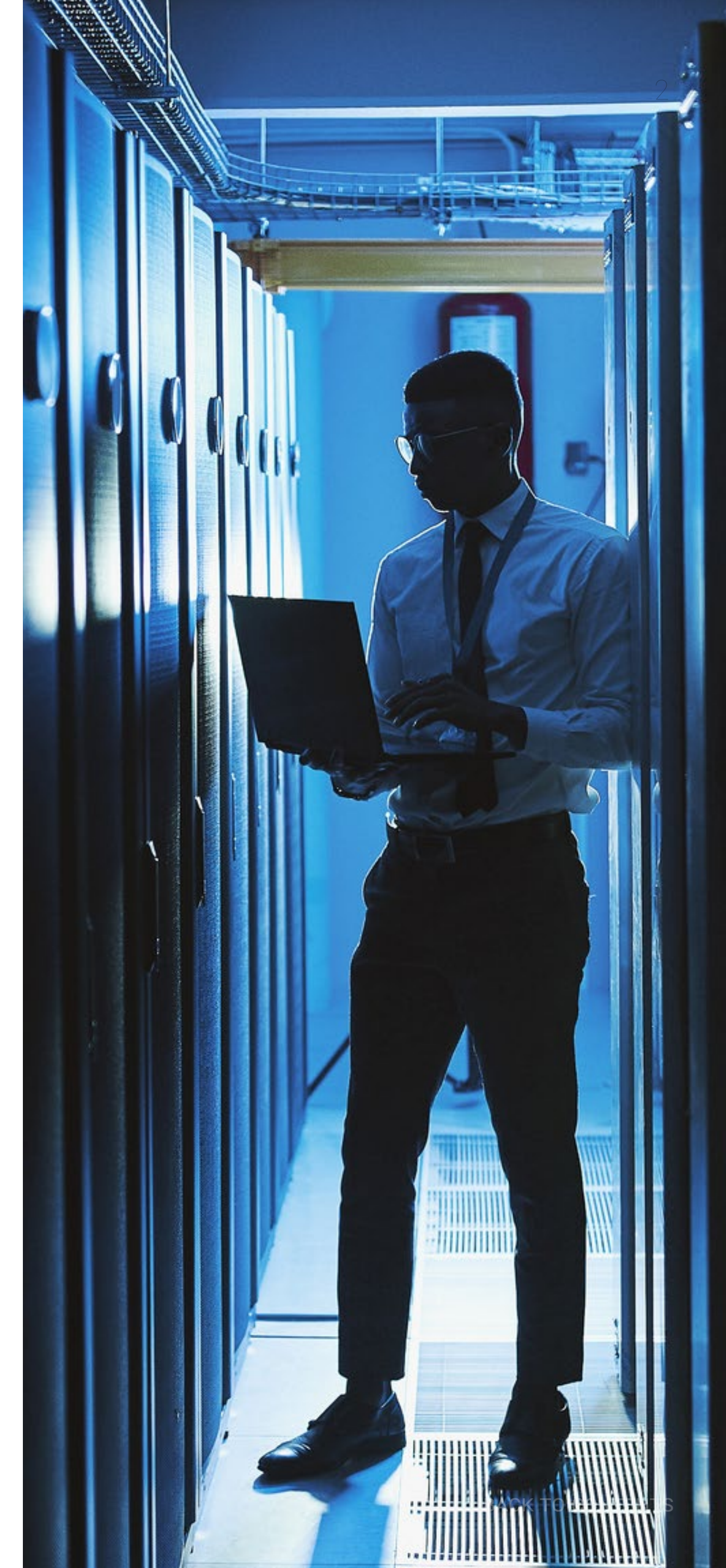
**Determine** how organizations are prioritizing solutions to best help them succeed.



**Establish** the current state of operationalizing AI through MLOps.

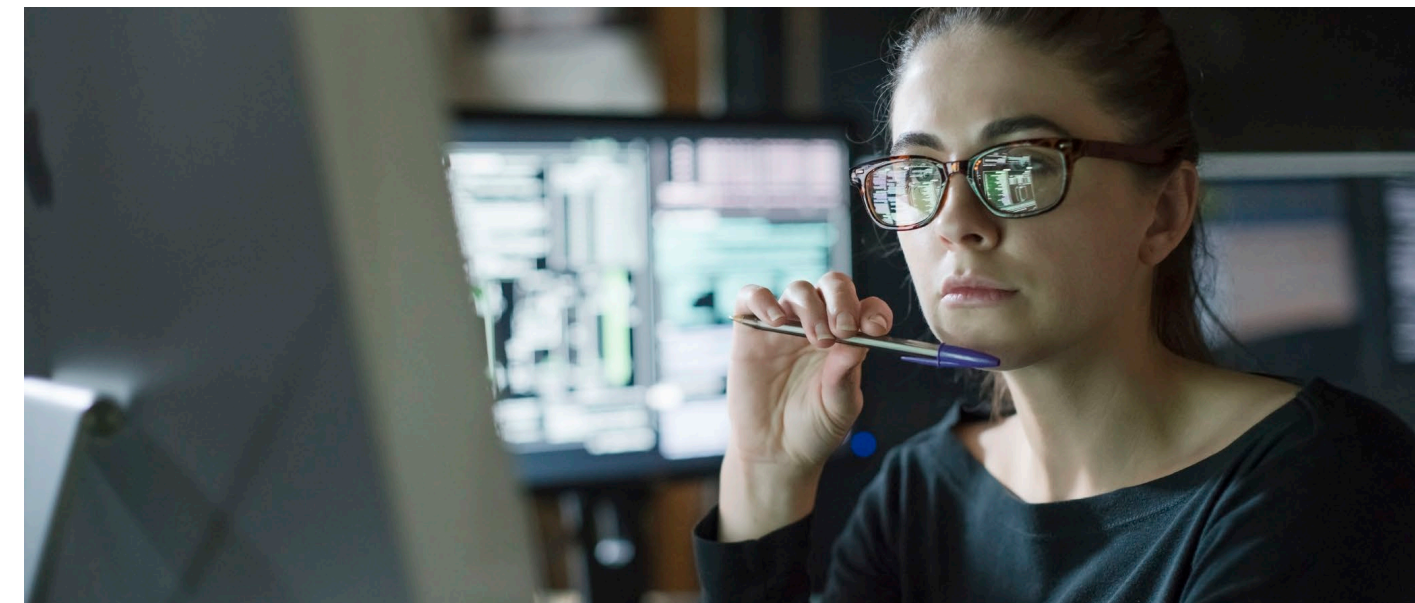


**Understand** the evolving stakeholder landscape, including team makeup, involvement, and growth opportunities.



# KEY FINDINGS

CLICK TO FOLLOW



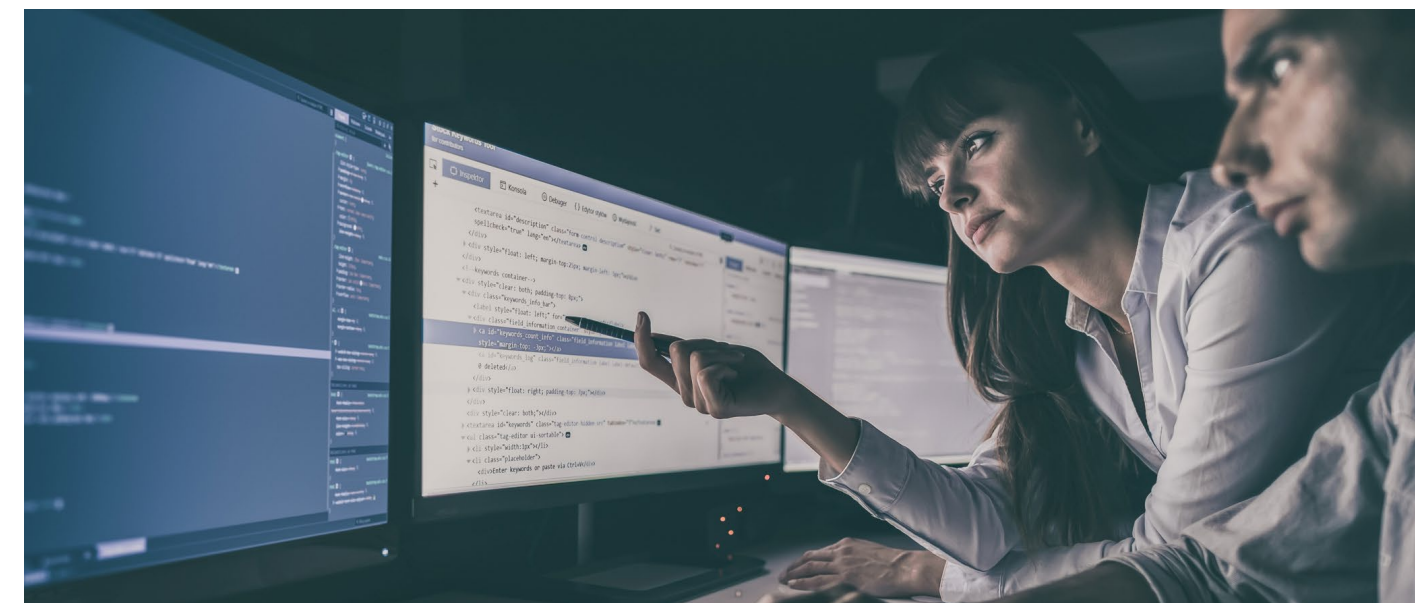
## Investments Point to Staggering Growth, But Challenges Loom Large

PAGE 4



## Focus Sharpens on Improving Early and Late Stages of Data Science Lifecycle

PAGE 10



## Organizations Improve Their Ability to Shift Models to Production But Need Further Efficiencies

PAGE 14



## Data Science and Machine Learning Become a Team Sport, With Vendors Focused on Enabling All Stakeholders

PAGE 17

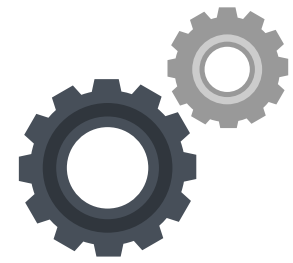
**Investments  
Point to  
Staggering  
Growth, But  
Challenges  
Loom Large**



## Primary Business Objectives Point Inward

Improving operational efficiency continues to be the lynchpin to most business objectives driving data science and machine learning initiatives. It not only empowers organizations to improve agility, cost-effectiveness, and customer centricity, but also lays the groundwork for sustainable growth and scale in an increasingly data-driven world. Once operations are performing at optimal levels, organizations can focus more on other business imperatives. However, data science and machine learning initiatives also are expected to improve product development, customer experience, risk management, and other areas.

| Primary business objectives of data science and machine learning initiatives.



**66%**

Improving operational efficiency



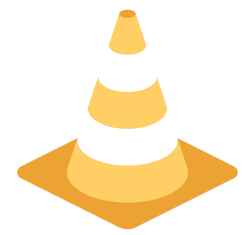
**60%**

Improving product development and innovation



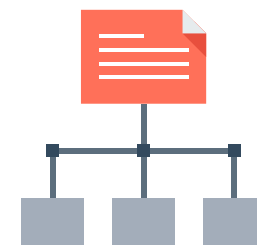
**52%**

Enhancing customer experience/ improving customer satisfaction



**49%**

Improving risk management



**47%**

Enhancing decision making

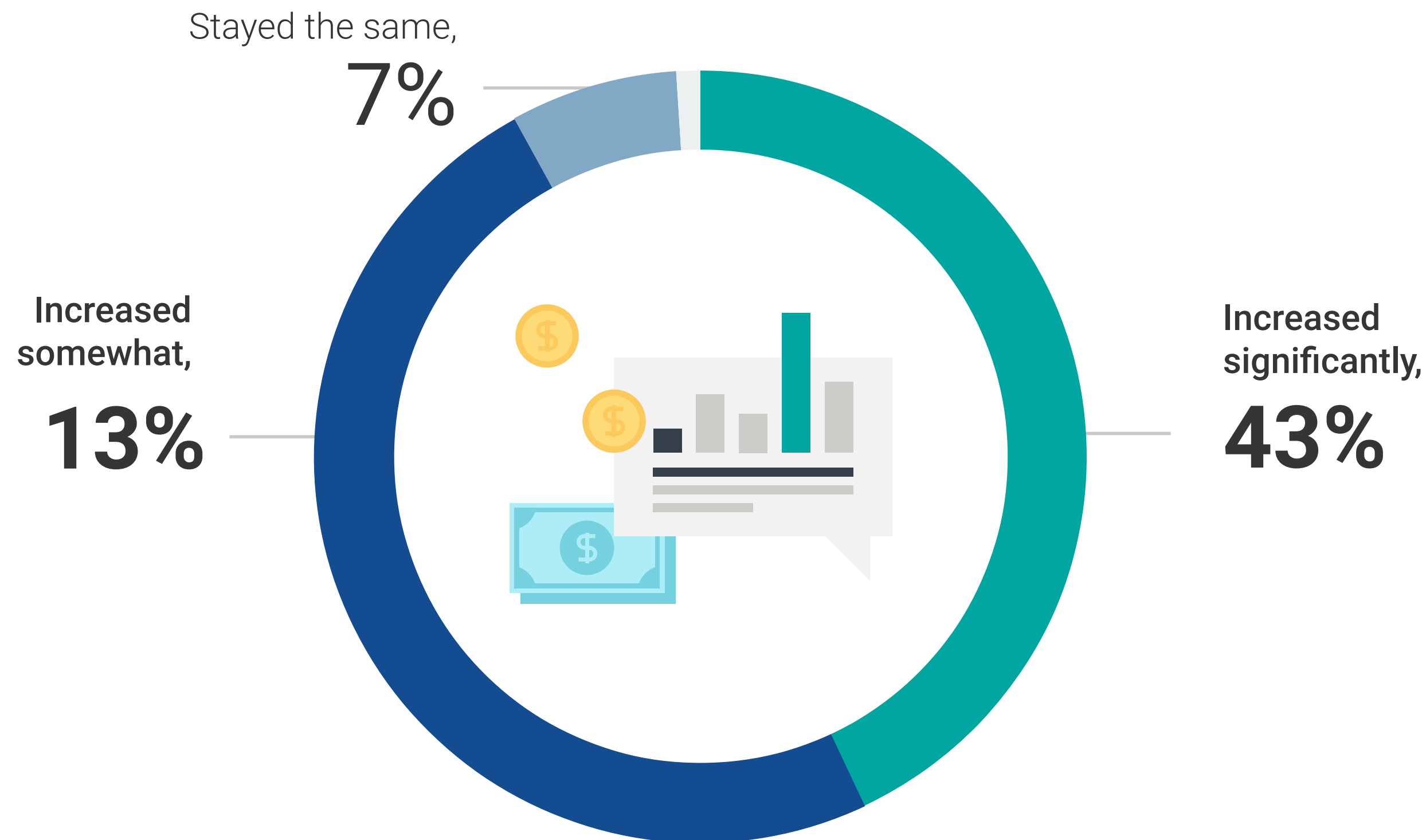


**43%**

Identifying new business opportunities and/or increasing revenue

“ This heightened investment reflects an understanding that **data science not only enhances operational efficiency but also enables informed decision making, predictive analytics, and innovative product development.**”

| Change in budget for data science and machine learning projects/initiatives compared with previous year.



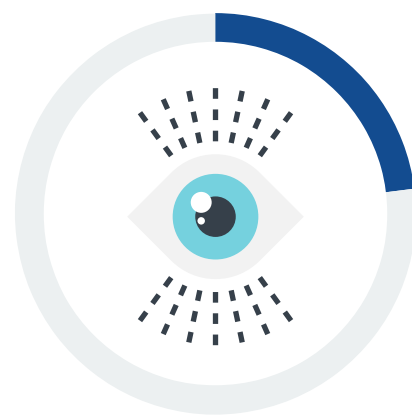
## Budgets Are on the Rise

Nearly all (92%) organizations saw a year-to-year increase in budget allocation for data science and machine learning projects/initiatives. These budgets are significant, with nearly one in four organizations (24%) planning to invest at least \$1 million in people, process, or technology in association with data science and machine learning over the next several years. This heightened investment reflects an understanding that data science not only enhances operational efficiency but also enables informed decision making, predictive analytics, and innovative product development. This financial support emphasizes the pivotal role that data science and machine learning play in enabling the business to extract valuable knowledge from vast and complex data sets, propelling organizations toward success in the digital age.

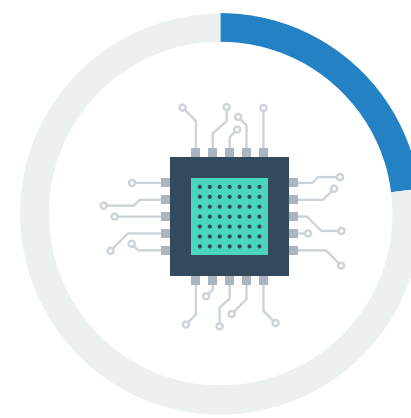
## Strategies Are Diverse When Prioritizing Data Science Projects

The willingness to sacrifice time to market and proceed with limited resources highlights the cautiously optimistic approach organizations are taking. They recognize they can't afford to wait but also that they must ensure robust model development, thorough testing, and accurate insights to avoid potential costly errors. This deliberate and calculated approach can enhance long-term performance, reliability, and stakeholder confidence, which far outweigh the initial time investment.

| Prioritized approach to data science-related projects.



**23%**  
**Business impact**  
(i.e., projects with highest potential business impact)



**23%**  
**Technical complexity**  
(i.e., projects with highest technical complexity)



**7%**  
**Time to market**  
(i.e., projects with shortest time to market)



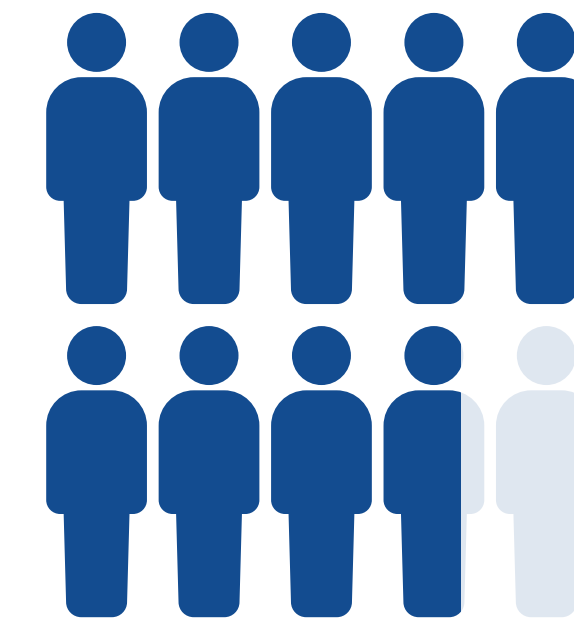
**13%**  
**Resource availability**  
(i.e., projects that can be completed with available resources)



**14%**  
**Customer feedback**  
(i.e., projects that address customer feedback)



**19%**  
**Executive leadership**  
(i.e., priorities are dictated by the executive leadership team)

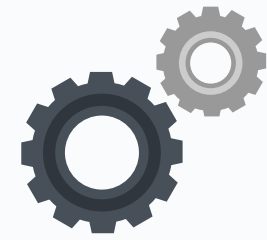


**88%**  
of organizations agree  
**that open source is  
critical to innovation  
in data science and  
machine learning.**

| Areas used to measure data science projects/initiatives.

## The Art of Measuring Data Science Project Impact

Each data science project brings a distinct dimension to measuring impact. The proximity of responses is a testament to the diversity of approaches and use cases that highlight the transformative power of data science across domains. Because operational efficiency is the most common business driver for data science initiatives, it follows that it is also the most common area measured to ensure the performance of those strategies. Customer satisfaction and cost saving are also commonly monitored to determine the impact of these initiatives.



**53%**

Improved operational efficiency



**48%**

Customer satisfaction



**45%**

Cost savings or revenue generation



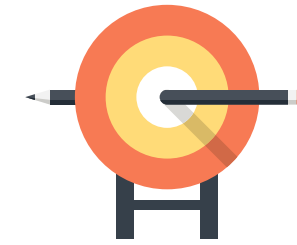
**39%**

Time savings



**37%**

Competitive advantage



**37%**

Predictive accuracy



**36%**

Innovation potential



**35%**

Employee satisfaction/happiness



**26%**

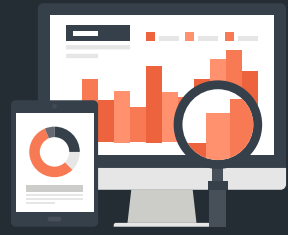
Social impact



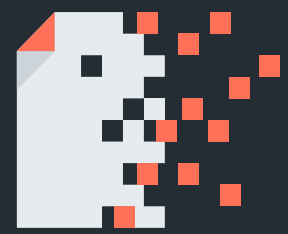
## Challenges Loom Large

Nearly all (94%) organizations face challenges in developing and implementing data science projects.

Challenges come in several shapes and sizes:



**Organizational:** skilled talent, budgets, defining objectives, and measuring outcomes.

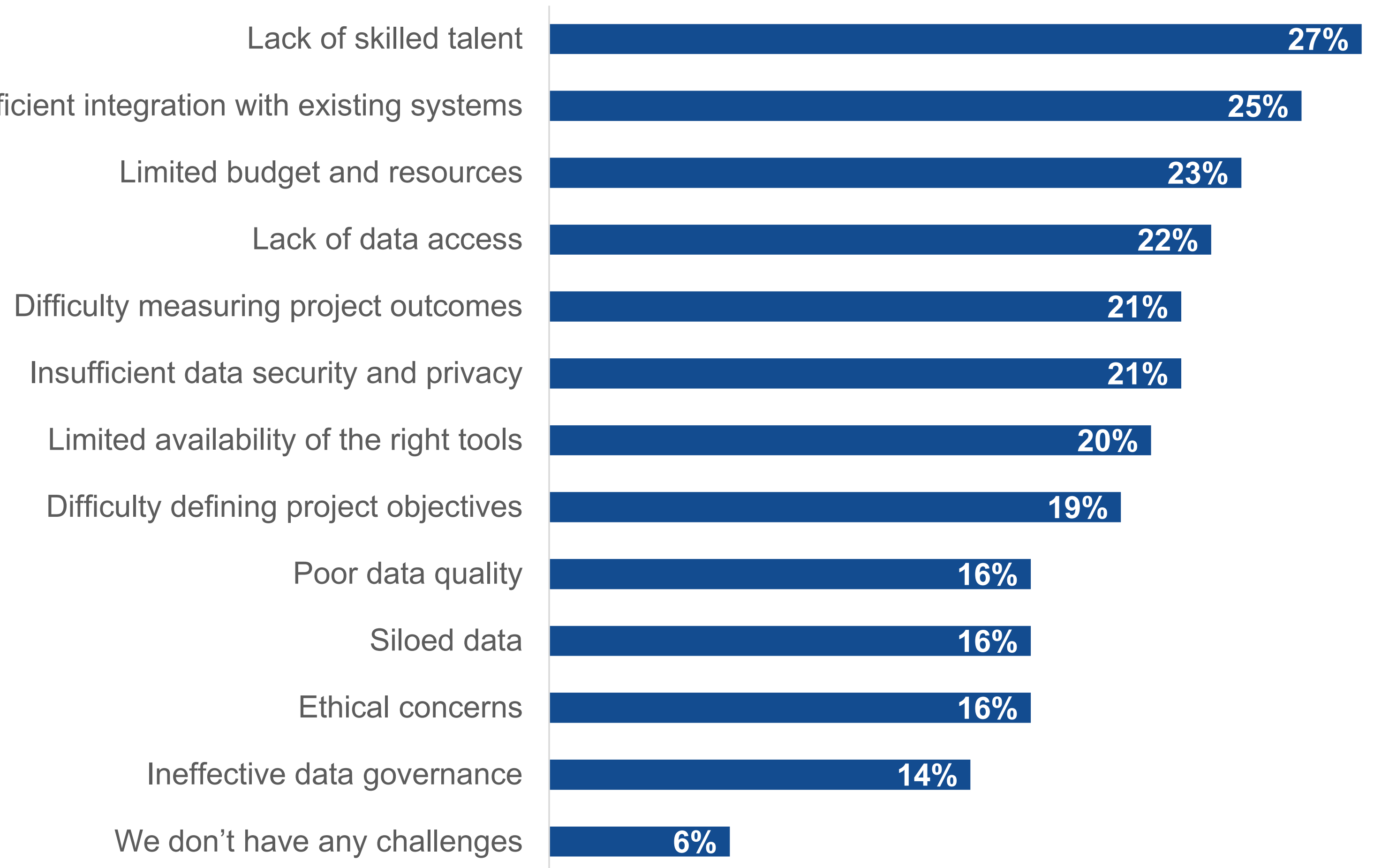


**Data/environment:** integrating with existing systems, data accessibility, limited tools, poor data quality, and siloed data.



**Trust:** data security/privacy, ethical concerns, and data governance.

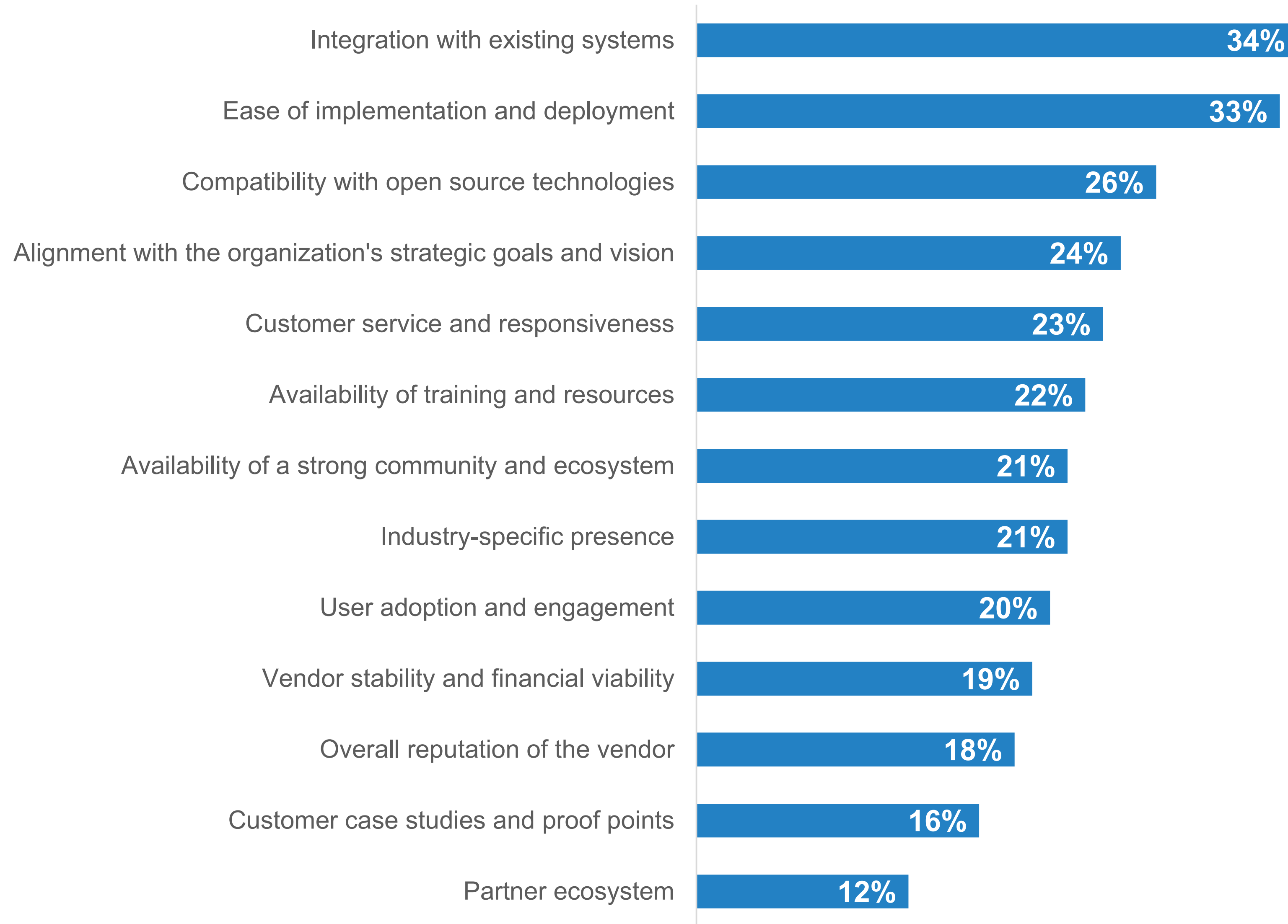
| Most significant challenges faced in developing and implementing data science projects.



# Focus Sharpens on Improving Early and Late Stages of Data Science Lifecycle



| Most important factors when considering purchases to support data science initiatives.



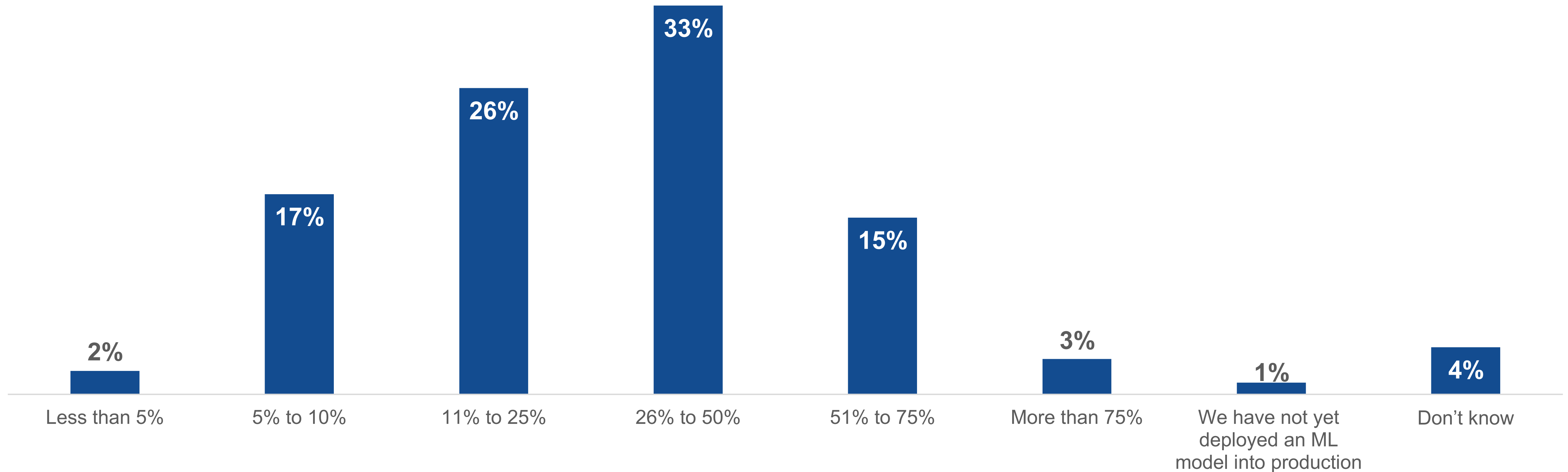
## Factors Weighed in Consideration of Data Science Purchases Highlight a Desire for Integration and Simplicity

Many organizations have already made massive investments in their data science and machine learning initiatives, so ensuring they still see value from those investments is critical. Simplifying implementation and deployment highlights the desire for organizations to ramp up quickly and improve the time between data generation and data insights. Note also that over a quarter (26%) of organizations consider compatibility with open source technologies, likely foreshadowing a larger open source deployment trend moving forward.

## Significant Room for Improvement Moving Models to Production

Within the last year, organizations have made great strides in improving the operationalization of machine learning models and transitioning them into production environments. Between robust frameworks and automated pipelines for model training, validation, and deployment, the industry has seen more seamless integration into existing systems, as well as streamlined processes that enable faster iterations. At the root of this improved success is the advent of MLOps practices to promote collaboration between data and IT stakeholders. However, despite these improvements, there is still significant room for improvement in the rate at which organizations deploy machine learning models into production environments. For example, 45% of organizations see less than 25% of their models make it into production. Challenges persist that require ongoing attention in managing the entire lifecycle of models, from initial development through continuous monitoring and maintenance to deal with model drift, performance degradation, interpretability issues, and more.

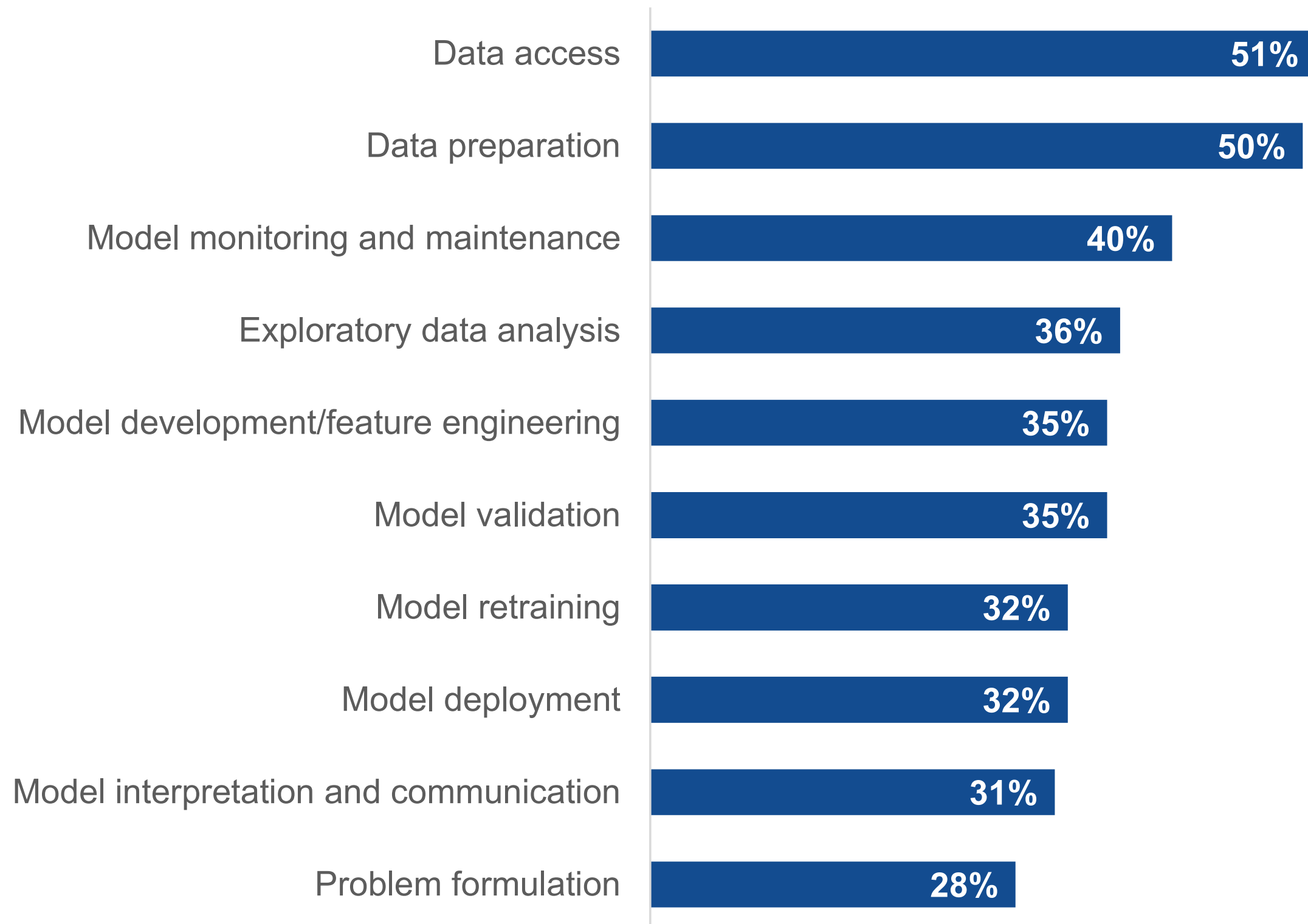
| Percentage of machine learning models deployed into production environments.



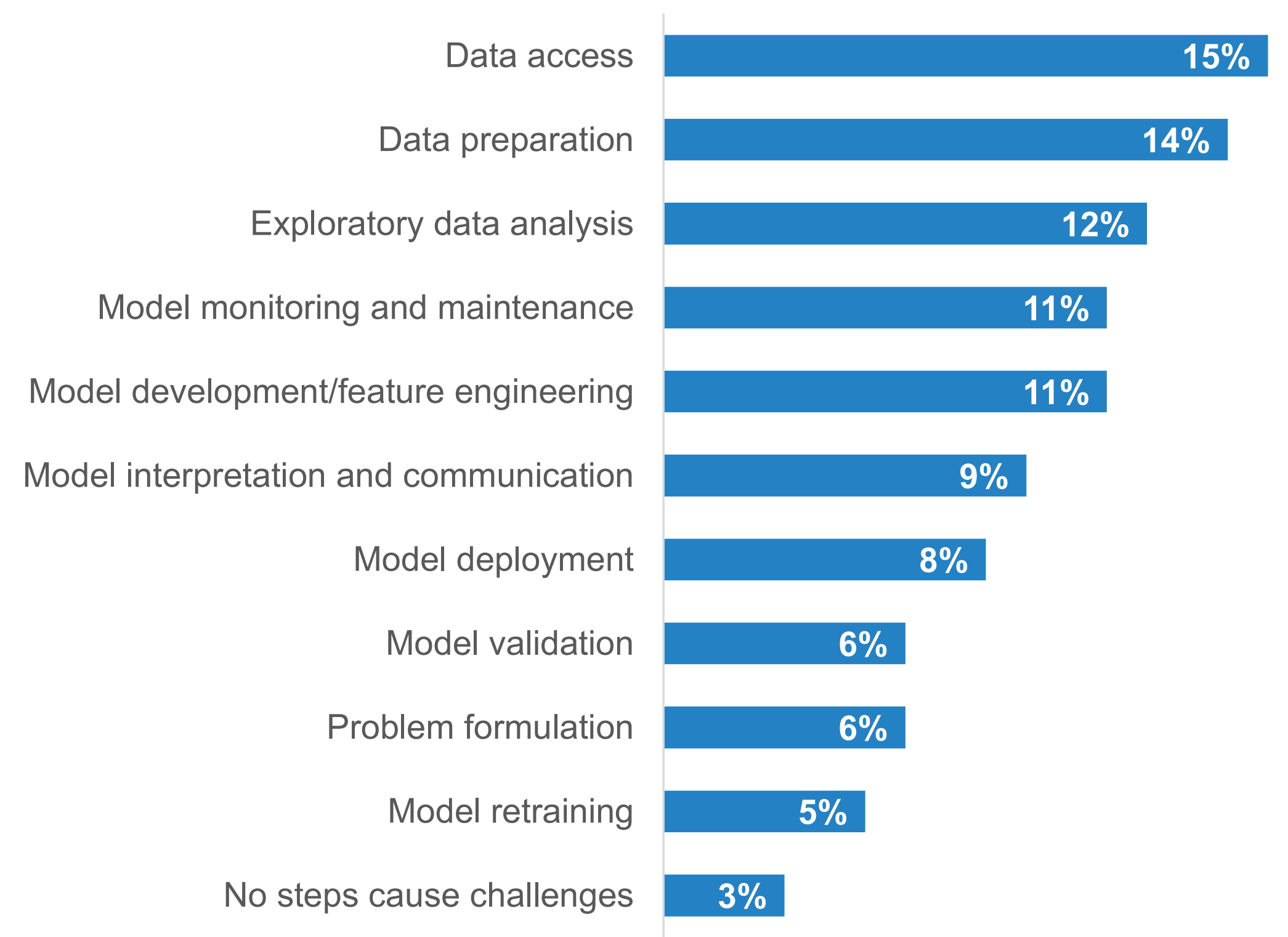
## The Importance of Data Cannot Be Overstated

Data accessibility and data preparation go hand in hand. Data accessibility forms the foundation for the entire data science lifecycle, highlighting not only why this is most commonly performed on a regular basis but also why it poses the largest challenge for organizations today. Data preparation, including cleansing, structuring, and transforming data, is a necessary step to ensure that subsequent analytical experiments are founded on a reliable and accurate basis.

| Data science lifecycle steps performed on a regular basis.



Most challenging data science lifecycle steps.



**Organizations  
Improve Their  
Ability to Shift  
Models to  
Production But  
Need Further  
Efficiencies**



## Unpacking Challenges in ML Deployment and Monitoring

Considering 58% of organizations have significant room to improve on their processes for moving models into production, it makes sense that even the most mature organizations run into challenges. Technical complexities arise when integrating models into existing infrastructure, ensuring compatibility with various systems, and encountering unexpected real-world data variability. Compliance and governance challenges impact reliability and trust as well as introduce risk. Operational complexities arise such as maintaining model performance over time and identifying/responding to failures. Continuous monitoring also poses challenges, such as addressing data drift and managing model dependencies such as model versioning.

| Challenges with deployment and monitoring of machine learning models.



**35%**

Difficulty managing multiple environments



**33%**

Difficulty ensuring compliance with corporate governance policies



**33%**

Difficulty detecting and responding to data drift



**29%**

Inconsistent model performance in production



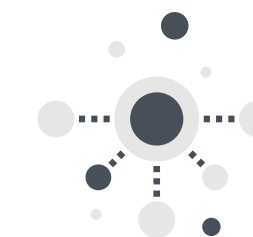
**29%**

Difficulty detecting and responding to model failures



**26%**

Inefficient retaining processes

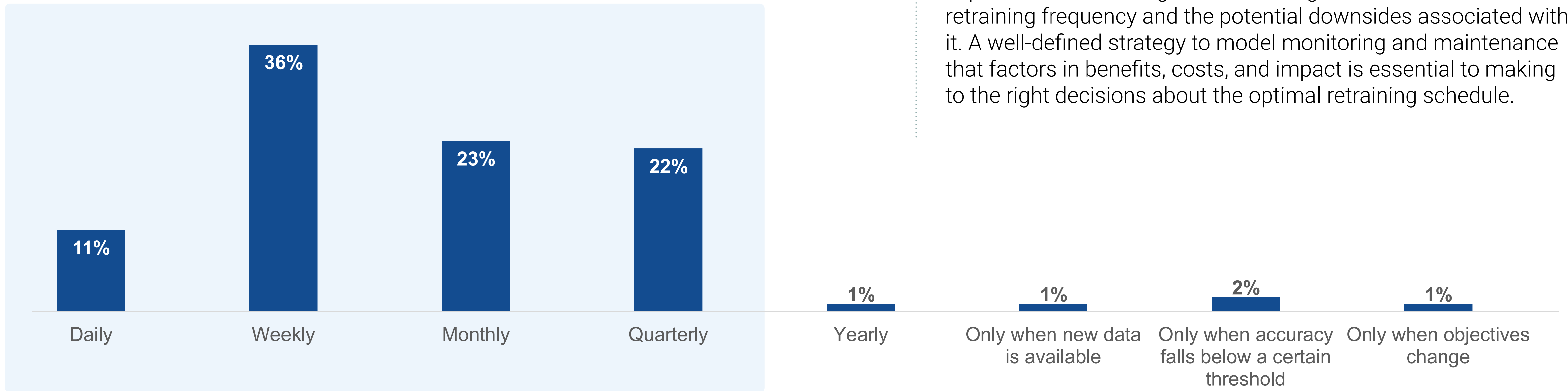


**26%**

Difficulty managing dependencies

“ A well-defined strategy to model monitoring and maintenance that factors in benefits, costs, and impact **is essential to making to the right decisions about the optimal retraining schedule.**”

| Frequency of retraining machine learning models in production.



## Striking a Balance Between Retraining and Maintaining

With 47% of organizations retraining models on at least a weekly basis, it is important to understand the impact frequent retraining can have on an organization, from resource strain and inefficiency to amplifying data noise and creating versioning complexities. While making changes via retraining based on data drift is important, doing so excessively can disrupt operations, confuse users, and hinder strategic focus on critical deployment aspects like monitoring and ethics. Organizations must balance retraining frequency and the potential downsides associated with it. A well-defined strategy to model monitoring and maintenance that factors in benefits, costs, and impact is essential to making to the right decisions about the optimal retraining schedule.



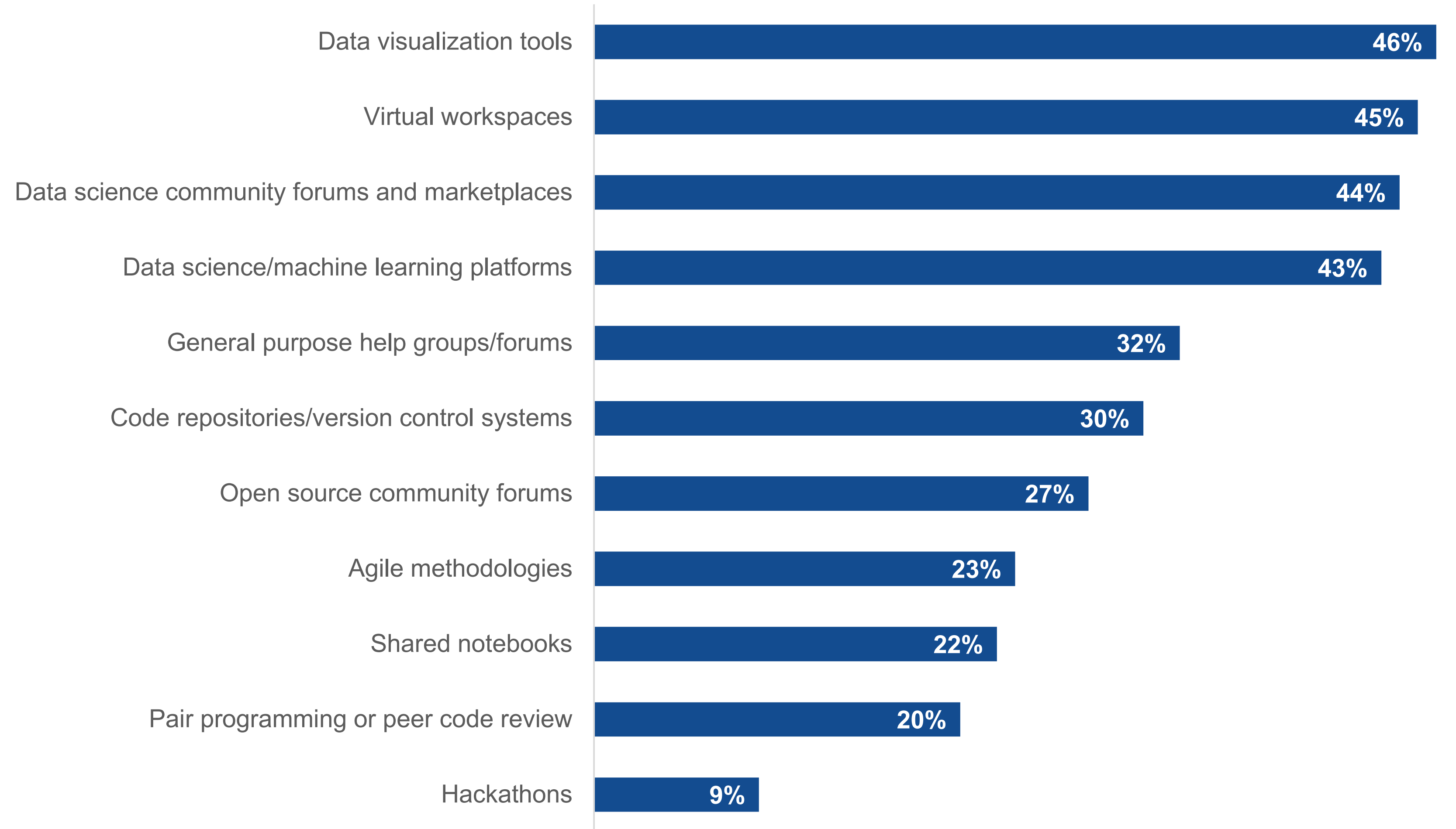
**Data Science and  
Machine Learning  
Become a Team  
Sport, With Vendors  
Focused on Enabling  
All Stakeholders**



## Building Bridges for Collaborative Data Science Success

Collaboration among stakeholders and team members is vital for successful data science initiatives. Organizations employ tools and methods to integrate expertise, fostering constructive dialogue, strategy refinement, and collective guidance. This open communication empowers diverse roles to shape outcomes, enhancing analysis quality and propelling organizations toward transformative insights and decisions.

| Sources used to ensure collaboration between stakeholders and other team members on data science initiatives.



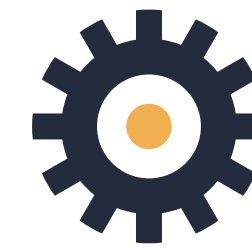
## Mapping Stakeholder Involvement Across the Data Science Lifecycle

Non-data science stakeholders play a significant role across the data science lifecycle, influencing various stages from data collection and preprocessing to model deployment and model management. This is a big reason why 92% of respondents rated the experience of business stakeholders involved in data science initiatives and working with data science teams as positive, if not very positive. Creating data science and machine learning solutions that cater to the non-data science community poses significant opportunities for vendors as organizations move forward in data science regardless of their levels of data science expertise.

| Machine learning model building areas that involve non-data science professionals (e.g., business analysts).



**44%**  
Data collection/  
supply



**40%**  
Data preprocessing



**39%**  
Model deployment



**38%**  
Model monitoring/  
maintenance



**36%**  
Model training



**36%**  
Model evaluation



**30%**  
Model selection



**28%**  
Logic building

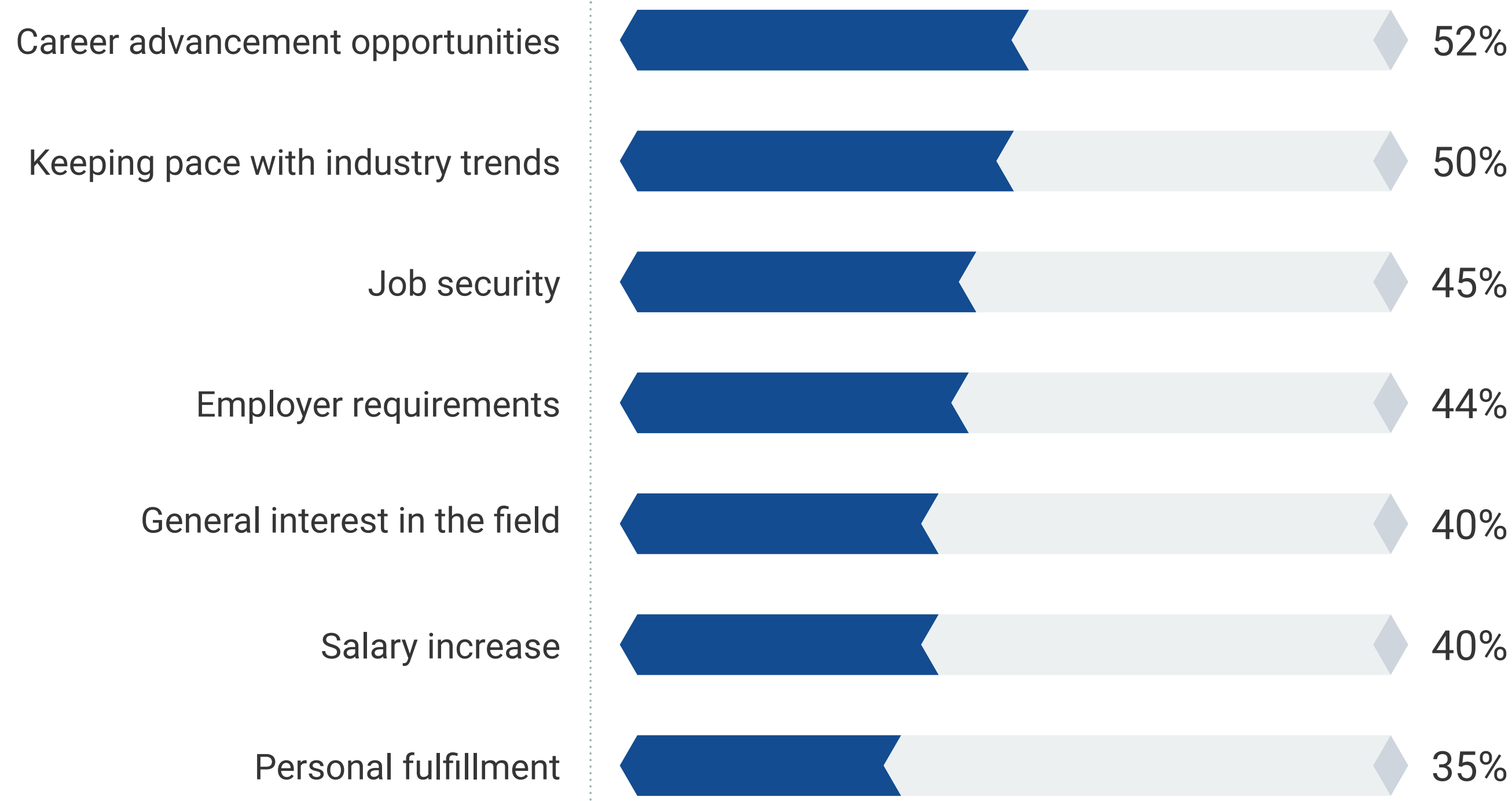


**27%**  
Use case/problem  
definition



# 99% of respondents are motivated to improve their data science and machine learning skills.

| Employees' drivers to improve skills in data science and machine learning.



## Unlocking Employee Potential

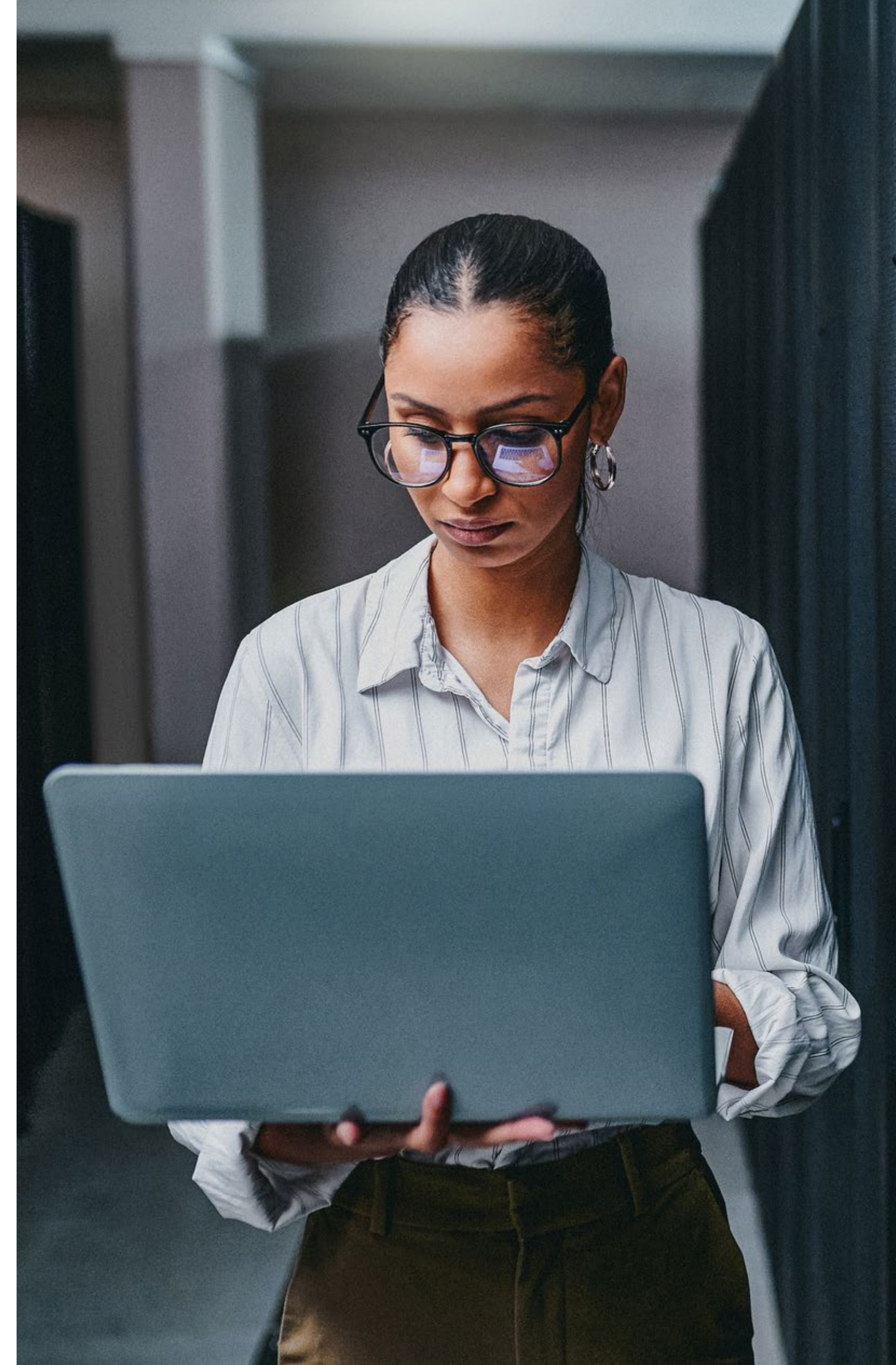
With 99% of people motivated to improve their data science and machine learning skills, the research highlights that improvements are fueled by a combination of intrinsic and extrinsic motivations. The prospects of career advancement, recognition, and salary increases, along with the promise of contributing meaningfully to cutting-edge projects, act as powerful external motivators. This combination of tangible rewards with intellectual curiosity creates an interesting dynamic within the work environment where employees are inspired to invest time (sometimes outside of work) to continue honing their skills.



**KNIME helps everybody make sense of data.**

Its free and open-source KNIME Analytics Platform enables anyone--whether they come from a business, technical or data background--to intuitively work with data, every day. KNIME Business Hub is the commercial complement to KNIME Analytics Platform and enables users to collaborate on data science and share insights across the organization. Together, the products support the complete data science lifecycle, allowing teams at all levels of analytics readiness to support the operationalization of data and to build a scalable data science practice.

[Learn More](#)

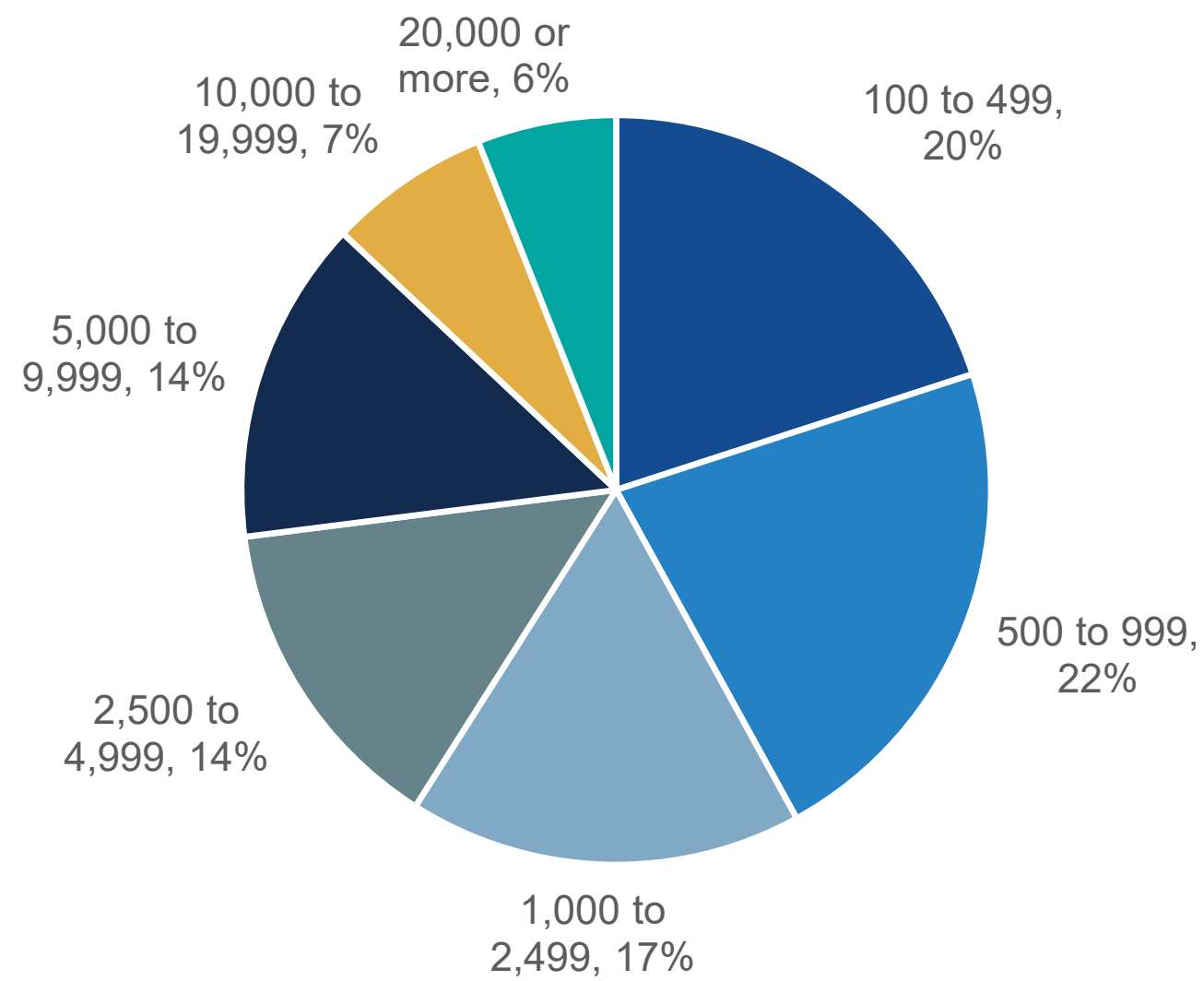


## Research Methodology and Demographics

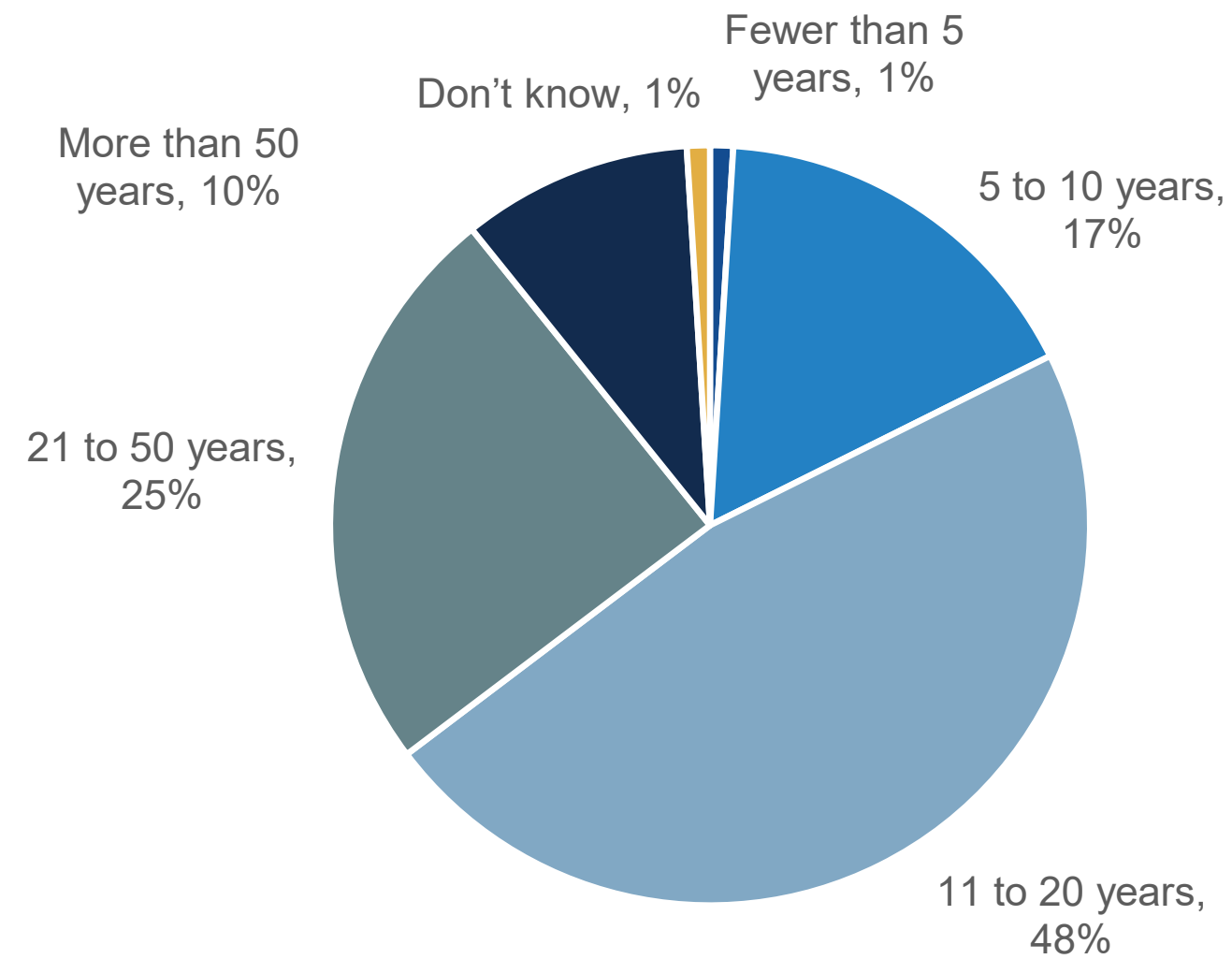
To gather data for this report, ESG conducted a comprehensive online survey of data professionals from private- and public-sector organizations in North America (United States and Canada) between June 5, 2023 and June 27, 2023. To qualify for this survey, respondents were required to be involved with data science and machine learning technologies and processes, including potential responsibility for strategizing, evaluating, purchasing, building, and managing these technologies. All respondents were provided an incentive to complete the survey in the form of cash awards and/or cash equivalents.

After filtering out unqualified respondents, removing duplicate responses, and screening the remaining completed responses (on a number of criteria) for data integrity, we were left with a final total sample of 366 data professionals.

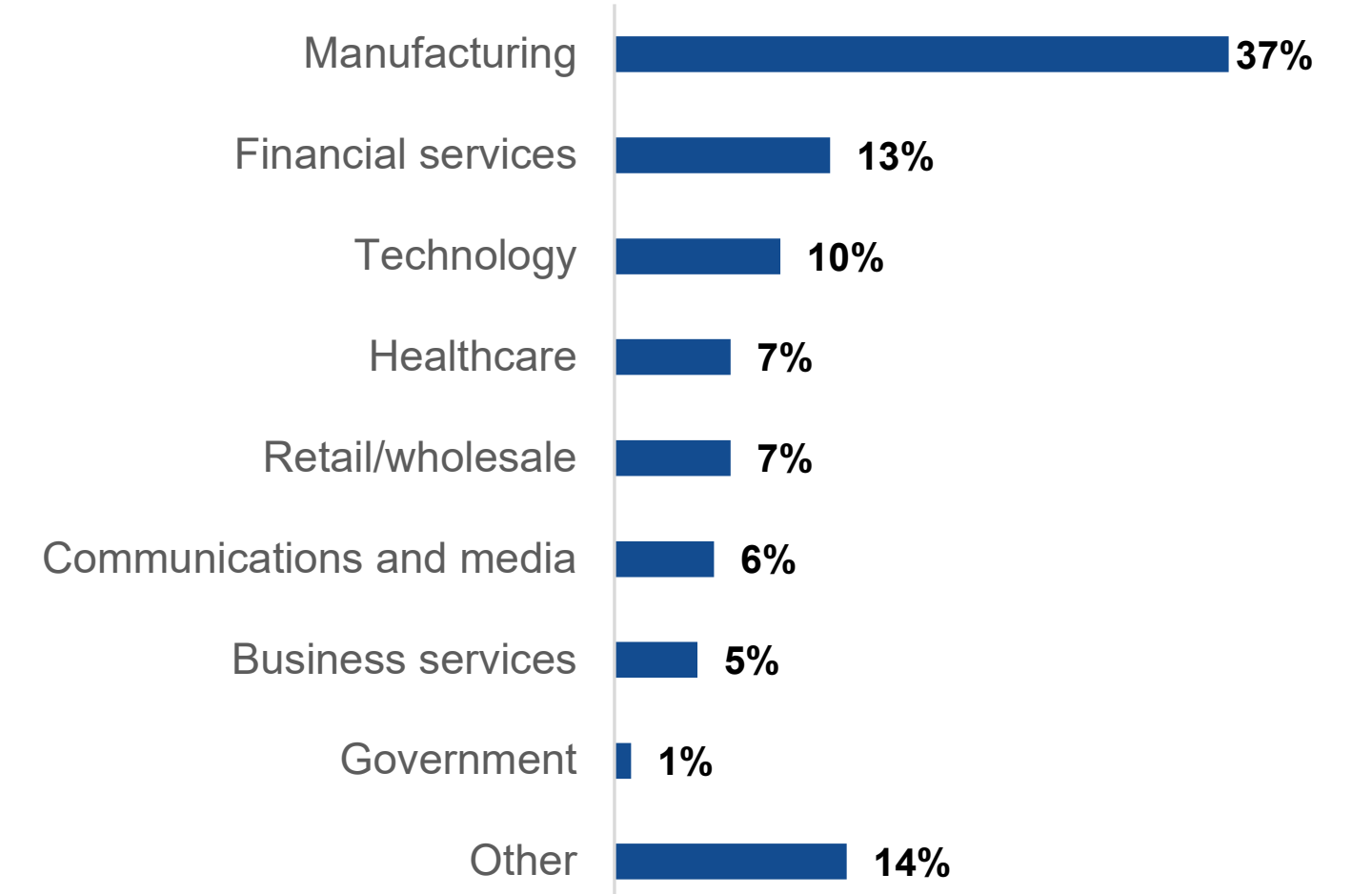
RESPONDENTS BY NUMBER OF EMPLOYEES



RESPONDENTS BY AGE OF COMPANY



RESPONDENTS BY INDUSTRY



All product names, logos, brands, and trademarks are the property of their respective owners. Information contained in this publication has been obtained by sources TechTarget, Inc. considers to be reliable but is not warranted by TechTarget, Inc. This publication may contain opinions of TechTarget, Inc., which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget, Inc.'s assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget, Inc. makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.

This publication is copyrighted by TechTarget, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of TechTarget, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at [cr@esg-global.com](mailto:cr@esg-global.com).



**Enterprise Strategy Group** is an integrated technology analysis, research, and strategy firm providing market intelligence, actionable insight, and go-to-market content services to the global technology community.

© 2023 TechTarget, Inc. All Rights Reserved.