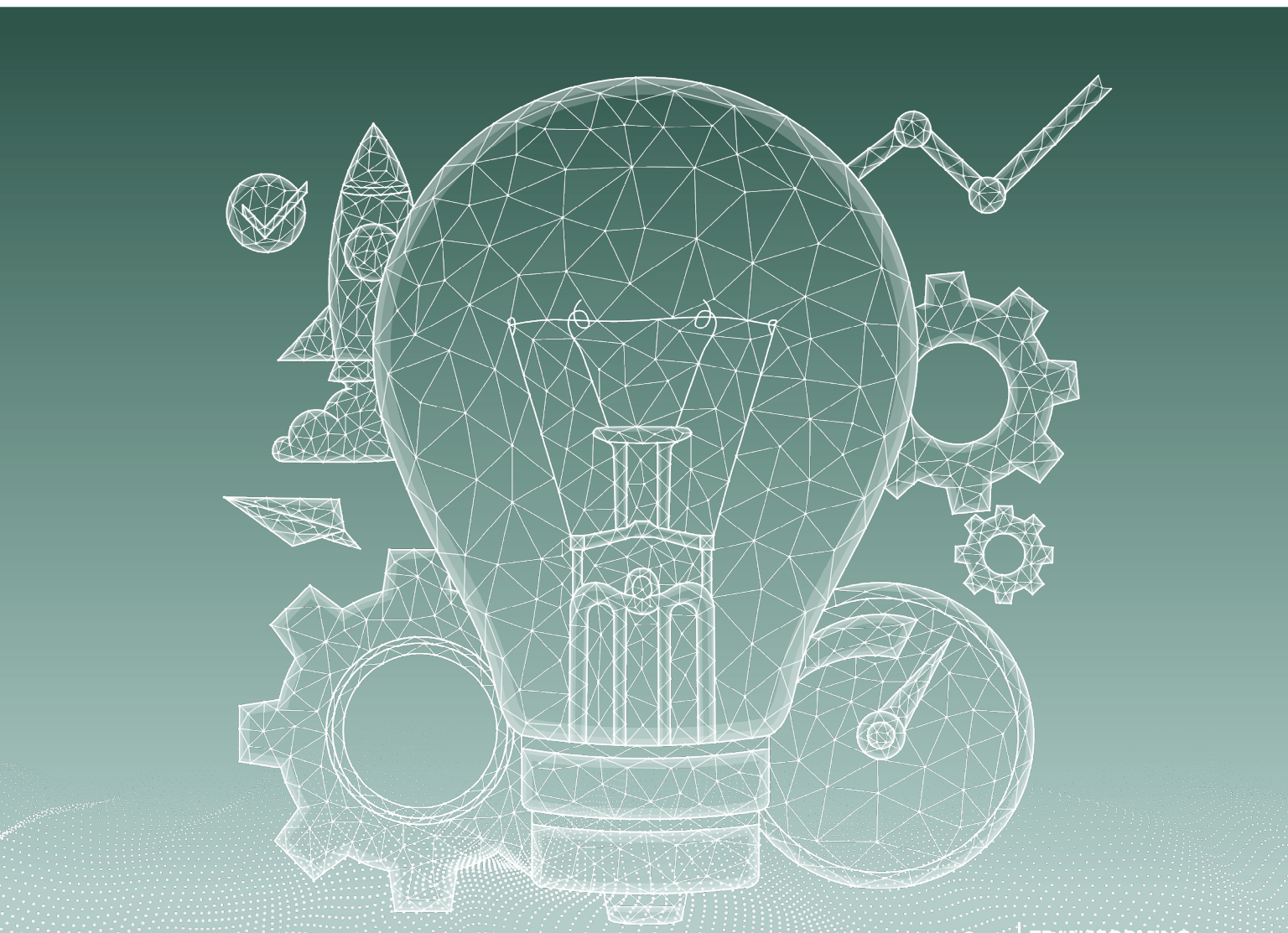
Co-sponsored by:

StreamSets
A SOFTWARE AG COMPANY

BEST PRACTICES REPORT

Q2 2023

Achieving Success with Modern Analytics



By Fern Halper, Ph.D.

tdwi TRANSFORMING DATA WITH INTELLIGENCE

Achieving Success with Modern Analytics

By Fern Halper, Ph.D.

Table of Contents

Executive Summary 4

The Move Forward: The Need for Advancing Analytics 5

- What Is Modern Analytics? 6
- What Are the Drivers for Modernizing Analytics? 7

State of Modern Analytics 7

- The Current State of Analytics Adoption . . . 7
- Data, Tools, and Platforms 8
- Use Cases for Modern Analytics 10
- Platforms for Modern Analytics 10

Challenges with Modern Analytics 12

New Technologies for Data and Analytics 16

- Automation and Augmentation 16
- Other Tool Types 17

Characteristics of Successful Organizations . . . 18

Recommendations 23

Research Co-sponsor: StreamSets 25

About the Author

FERN HALPER, Ph.D., is vice president and senior director of TDWI Research for advanced analytics. She is well known in the analytics community, having been published hundreds of times on data mining and information technology over the past 20 years. Halper is also coauthor of several Dummies books on cloud computing and big data. She focuses on advanced analytics, including predictive analytics, machine learning, AI, cognitive computing, and big data analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead data analyst for Bell Labs. She has taught at both Colgate University and Bentley University. Her Ph.D. is from Texas A&M University. You can reach her by email (fhalper@tdwi.org), on Twitter (twitter.com/fhalper), and on LinkedIn (linkedin.com/in/fbhalper).

About TDWI Research

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, onsite courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resource-filled publications, an in-depth research program, and a comprehensive website: tdwi.org.

About the TDWI Best Practices Reports Series

This series is designed to educate technical and business professionals about new data and analytics technologies, concepts, or approaches that address a significant problem or issue. Research for the reports is conducted via interviews with industry experts and leading-edge user companies, and it is supplemented by surveys of data professionals.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving data and analytics problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical business intelligence issues. To suggest a topic that meets these requirements, please contact TDWI Senior Research Directors David Stodder (dstodder@tdwi.org), James Kobielus (jkobielus@tdwi.org), or Fern Halper (fhalper@tdwi.org).

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who agreed to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI's production team: James Powell, Peter Considine, Lindsay Stares, and Rod Gosser.

Sponsors

SAP, Snowflake, and StreamSets sponsored the research and writing of this report.

Research Methodology and Demographics

Report purpose. As data becomes more diverse and volumes increase and as analytics becomes more complex, organizations are struggling to gain insights and put those insights into action. There are numerous reasons for this including cultural issues, skills issues, and technology issues. This Best Practices Report examines these factors as well as the difference between successful and unsuccessful companies.

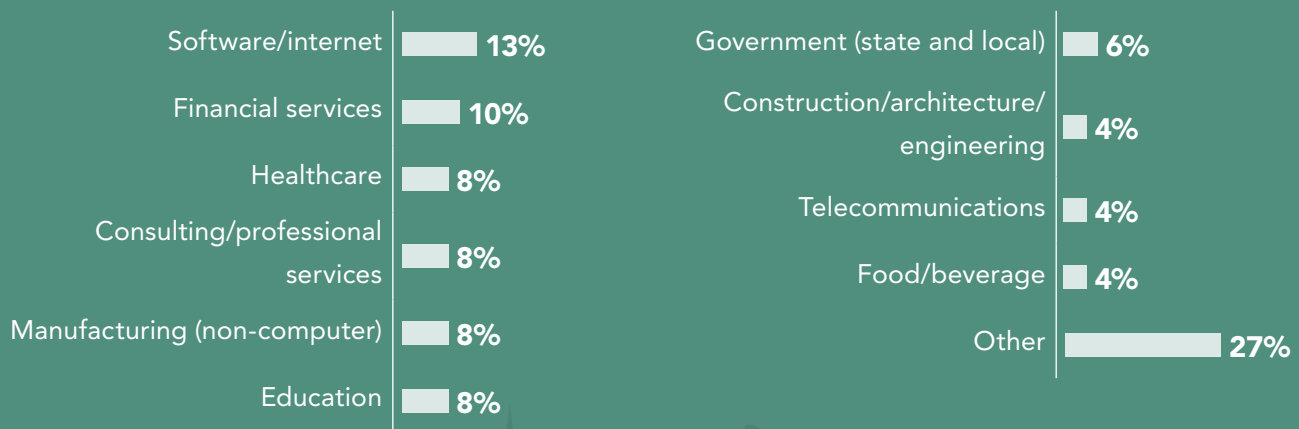
Survey methodology. In March 2023, TDWI sent an invitation via email to the analytics and data professionals in our database, asking them to complete an online survey. Three hundred and ninety-four respondents completed the entire survey and met quality standards. This group is used for analysis.

Research methods. In addition to the survey, TDWI conducted interviews with technical users, business sponsors, and analytics experts. TDWI also received briefings from vendors that offer products and services related to these technologies.

Survey demographics. Respondents act in a variety of roles. These include corporate executives/VPs/directors (IT and business), data analysts, managers, architects, data scientists, and others.

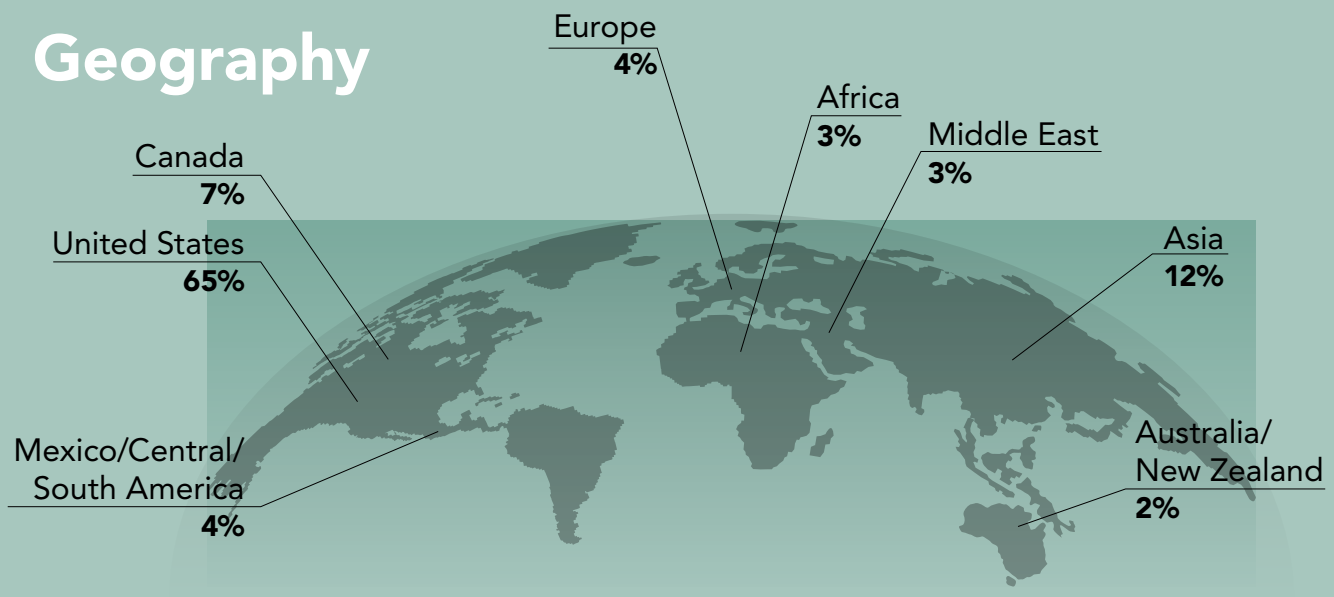
Respondents came from a range of industries including software/internet (13%); financial services (10%); healthcare, consulting, manufacturing, and education (all 8%); and others. Most survey respondents reside in the U.S. (65%), Asia (12%), and Canada (7%). Respondents come from enterprises of all sizes.

Industry

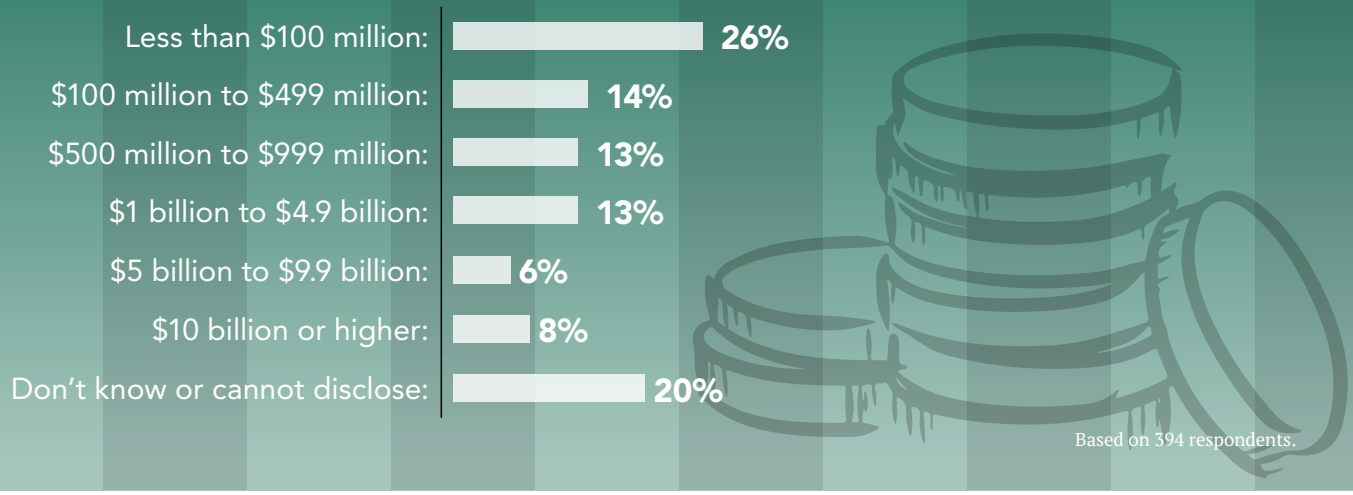


("Other" consists of multiple industries, each represented by less than 4% of respondents.)

Geography



Company Size by Revenue



Based on 394 respondents.

Executive Summary

Modern analytics can provide a significant path to value for organizations. Although many companies are still analyzing structured data, newer data sources such as machine data or text or image data, together with newer analytics approaches, including automation and new techniques, are becoming part of an evolving data and analytics landscape.

This report examines the drivers for modern analytics and the current state of analytics adoption (including tools and platforms) and explores the differences between those who are using modern analytics successfully and those who are not.

TDWI research finds that organizations are embracing modern technologies in their efforts to support a modern data infrastructure for analytics and a complex data landscape. These include:

- Cloud platforms and technologies
- Automated and augmented tools that help manage and govern data

- Data fabrics to address hybrid environments
- Pipelines that can support new data types such as streaming data
- New governance tools including automated data quality and lineage tools

On the analytics front, about a third of survey respondents are utilizing more advanced analytics techniques such as machine learning and deep learning. Some are embracing self-service tools that may automate the surfacing of insights for business users. A small group is already starting to embrace generative AI models that create new outputs—such as images, music, text, or other forms of media—based on the training data. Organizations are using both open source and commercial products; they are using first-party data as well as data from partners and third-party providers, including marketplace data.

> *This report examines the characteristics of companies that are successful with modern analytics.*

A unique feature of this report is its examination of the characteristics of companies that self-identify as being successful with modern analytics. In other words, it explores how those companies compare to those that are not successful. Some highlights include differences in leadership, organizational style, data platforms, and tools used, as well as how they approach issues such as data literacy and governance. For instance:

- Those who are successful are more likely to have a committed analytics leader in place (82% versus 51%)
- Data literacy is a priority (78% versus 44%)
- They are more likely to state that cloud platforms are a priority (69% versus 44%)

Despite some organizations' success, many businesses still struggle to implement or see benefits from modern analytics. Part of the issue is with technology. Additionally, there are organizational challenges that must be addressed, which often include building cultures, funding, hiring the right people, and organizing to execute. This report concludes with recommended best practices that can guide your organization to overcome these difficulties and be successful with modern analytics.

The Move Forward: The Need for Advancing Analytics

Today's enterprises need meaningful, data-driven insights to answer the strategic business questions that will help them succeed. This requires the use of self-service analytics as well as more advanced analytics techniques such as predictive analytics and machine learning (PA/ML), natural language processing, and other types of artificial intelligence. Many organizations also plan to develop data products that include analytics to drive value. They are working to embed analytics. To accomplish these goals and provide value to customers and partners, organizations must move forward with analytics and advanced analytics.

Unfortunately, adoption appears to be stalled. For instance, comparing TDWI survey results over the past seven years indicates that organizations are not adopting machine learning and other advanced analytics technologies at the rate they planned.

In a recent TDWI survey, 75% of respondents said they were struggling to get to the next stage of analytics maturity.[1] Even self-service visual analytics is not pervasive throughout companies.

Why are these companies struggling and what can we learn from companies that have been successful in moving forward with modern analytics?

What Is Modern Analytics?

Modern analytics refers to the use of advanced data analysis techniques, such as machine learning and natural language processing, to extract insights from large and complex data sets. It also includes newer self-service tools that promote democratization—tools that surface insights or use natural language query to make it easy to ask questions of data. Modern analytics involves using sophisticated tools and technologies to process, transform, and analyze data in real time. Examples of modern analytics include use cases we've been talking about for years at TDWI, such as fraud detection in banking, predictive maintenance in manufacturing, hyper-personalization in retail, and supply chain optimization in logistics. It includes newer use cases involving new data products and applications and real-time analytics against streaming data.

Modern analytics also includes the use of automated and augmented approaches—tools that infuse AI and machine learning techniques to perform tasks such as deriving features for machine learning models, suggesting queries, and surfacing insights from data. Vendors are embedding these tools into products, and we've seen organizations begin to adopt them.

[1] Unpublished 2022 TDWI survey.

A glossary of popular modern analytics techniques

Machine learning (ML): Systems learn from data to identify patterns with minimal human intervention. Machine learning originated in the field of computer science. Popular machine learning algorithms include decision trees, neural networks, and Naïve Bayes classification.

Deep learning: A subfield of machine learning where algorithms learn functions that can classify complex patterns, such as images. Often uses deep neural networks.

Natural language processing (NLP): Systems that can read, analyze, and understand human language with the goal of human/computer interaction. Many rely on machine learning algorithms, although other techniques, such as rule-based approaches, are used. A popular example is text analytics, which is used to extract entities, concepts, and sentiments from text data.

Natural language generation (NLG): Systems that can automatically generate a narrative from data. Considered a type of NLP/computational linguistics.

Speech recognition: Also a subfield of NLP and computational linguistics that enables computers to recognize and translate spoken language into text (e.g., speech to text).

Computer vision: A field of computer science where computers can obtain information from images or other multidimensional data.

Generative AI: A type of artificial intelligence that involves generating new and original content based on patterns and data inputs.

Finally, modern analytics includes very new approaches, such as generative AI (systems that are designed to create new outputs, such as images, music, text, or other forms of media, based on their training data). This includes large language models such as ChatGPT or BERT that are trained using a corpus of text with millions of parameters to be able to predict the next word based on a user's prompt. Uses for these models include text generation, producing code, and analyzing unstructured data. Generative AI also includes other tools, such as DALL-E or MidJourney, that utilize deep learning models to generate images from natural language descriptions.

What Are the Drivers for Modernizing Analytics?

Today's environment of inflation, increasingly complex customer demands, supply chain disruptions, and other drivers highlights the need for analytics. On the self-service and democratization front, analytics can be used by business users, who understand the business, to drive timely decisions on everything from sales forecasts to when to restock inventory. On the ML and AI front, analytics can help organizations become more proactive by predicting churn, targeting the most valuable customers, and building smart products. In other words, modern analytics helps organizations be more productive and competitive.

> *Modern analytics helps organizations be more productive and competitive.*

We asked survey respondents about what's driving them to employ modern analytics. Many respondents said they need deeper insights and actionable intelligence to become more competitive. Others cited improved business efficiency as a primary driver. Some respondents mentioned specific business needs such as churn analysis and fraud detection. They talked about specific use cases including population health, understanding product life cycles, and voice of the customer in social media analysis, as well as predicting early termination of contracts, use cases for computer vision, and the like. The bottom line is that organizations are realizing that they need data and analytics for decision making and to grow.

State of Modern Analytics

To understand the current state of modern analytics, we asked respondents questions about the kinds of data and analytics they are using, where analytics is used, the kinds of tools (commercial versus open source) used, and the data infrastructure to support modern analytics.

The Current State of Analytics Adoption

We asked respondents about their organization's current state of analytics adoption. Respondents were split (see Figure 1). Twenty-one percent are still using spreadsheets for analytics. Forty-nine percent (the majority) are utilizing dashboards and visual analytics tools including self-service. Only 8% chose "self-service visual analytics is used by the majority of business users who need it" as the best description of their state of analytics adoption. We've seen that self-service has not widely penetrated throughout many organizations in other surveys.

The rest of the respondents (29%) were either on the cusp of using more advanced tools such as machine learning or were already building and/or deploying models into production. Although we've seen higher percentages (about 35-40%) in other surveys in terms of advanced analytics adoption, this figure jibes with the struggle we've seen in

other research to adopt machine learning or other advanced techniques. As a point of reference, when TDWI published a predictive analytics Best Practices Report in 2007, 21% of the respondents had fully or partially implemented the technology. In other words, 15 years later, adoption rates (measured only by these groups of respondents) may have increased by 10–20%. Over the years, we've seen that although large percentages of respondents say they are planning to implement the technology, they have not yet done so.

That is to say, we've seen that some companies are moving forward to utilize analytics—such as machine learning and other advanced and modern techniques—but most companies are not. During this same period, the amount and kinds of data companies have had to deal with has increased dramatically. Whereas 15 years ago companies may have been dealing with structured data only, today they are collecting text, machine-generated, geospatial, clickstream, and a host of other types of data. This has driven the need for new platforms to support modern approaches as well as the need to analyze this data for business advantage.

In the sections that follow, we use the term *modern analytics* to refer to self-service analytics used by business users as well as the other technologies mentioned above.

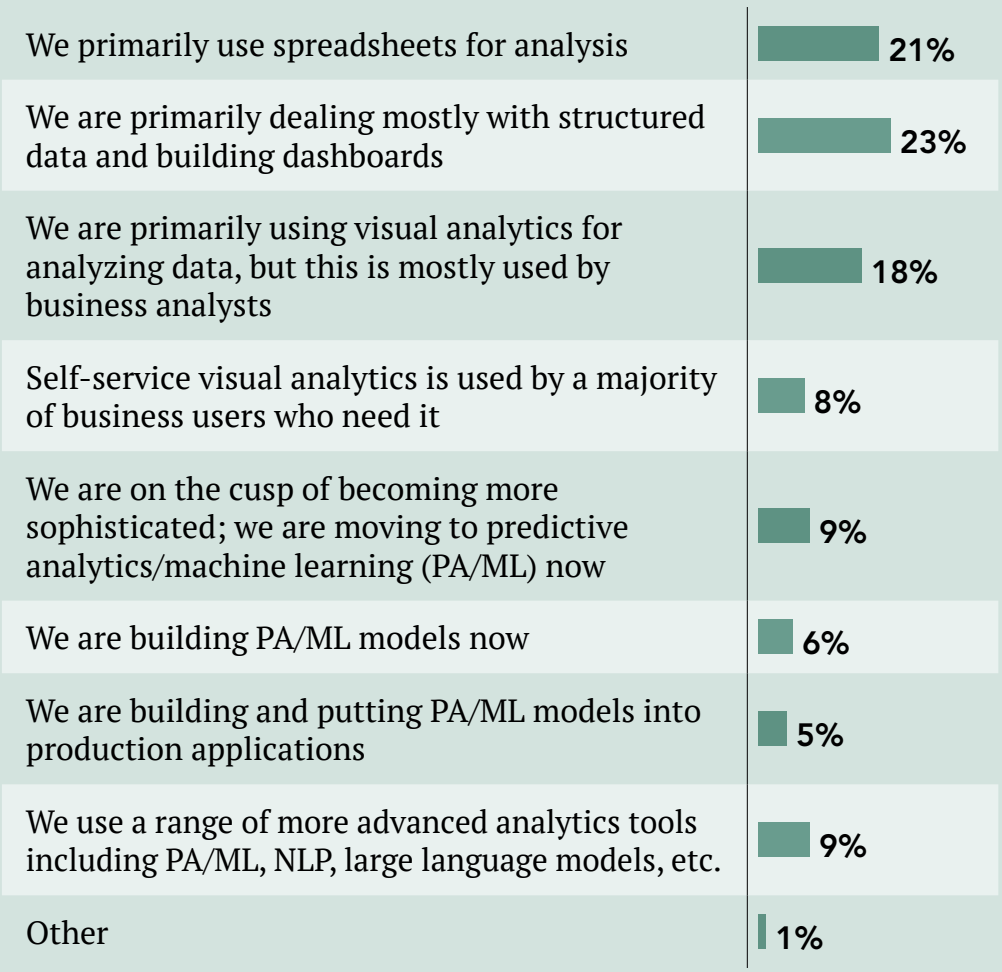
Data, Tools, and Platforms

In the realm of advanced analytics tools, survey respondents who were already performing

Figure 1

Which statement BEST describes your organization's current state of analytics adoption?

Based on 394 respondents.



modern analytics (i.e., self-service and beyond) were primarily using machine learning (Figure 2). Fifty-nine percent of those using modern analytics were deploying machine learning. Natural language understanding was used by about a third of these respondents. This technology is often used in chatbots, sentiment analysis, language translation, or classifying text.

Figure 2 also illustrates the strong and varied interest of respondents using modern analytics in other kinds of modern approaches, including deep learning, computer vision (which often uses deep learning), and generative approaches (which may also use deep learning).

We also asked whether the respondents were utilizing open source or commercial products. Open source analytics programming languages, such as Python or R, are free to use and many organizations view them as a low-cost entry point for advanced analytics. The open source community development model appeals to those who enjoy coding and getting access to the latest algorithms. They also like the fact that there is no vendor lock-in. On the other hand, some organizations prefer commercial products because they are often easier to use, provide support, and are well-documented.

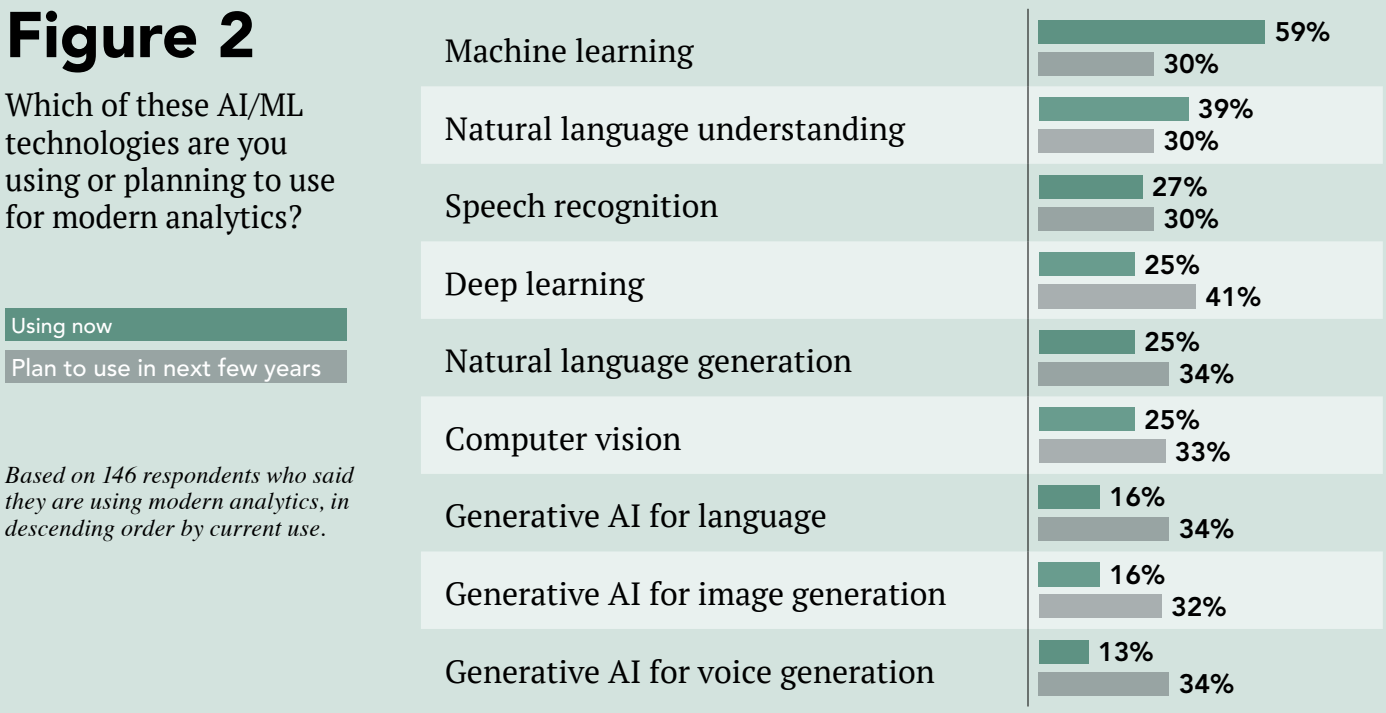
In this survey, approximately 25% of those using modern analytics were using only commercial tools, and another 25% were only using open source. The other 50% were making use of both open source and commercial products (not shown). This makes sense, as often organizations utilize many tool types depending on the persona using them and their preference.

Modern analytics is also used across the organization. Fifty percent of those using it are using it in operations, 38% are using it in marketing, 35% are using it in sales, and 33% are using it in finance (not shown).

Data from both internal and external sources is used to enrich data sets for modern analytics. External data such as demographic, firmographic,

Figure 2

Which of these AI/ML technologies are you using or planning to use for modern analytics?



Based on 146 respondents who said they are using modern analytics, in descending order by current use.

weather, sensor, and industry-specific data can be obtained through an online data marketplace, which can help streamline access to external data. TDWI is seeing companies increasingly interested in using new data sources, such as marketplaces.

In this survey, the overwhelming majority (over 80%) of those using modern analytics are using first-party data, 73% were using second-party/partner data, and 45% were using third-party/marketplace data (not shown).

Use Cases for Modern Analytics

As previously noted, respondents stated that modern analytics is being used across the organization, especially in operations, marketing, sales, and finance. We asked respondents about their most exciting use case for modern analytics. For most respondents, the use cases cited were either industry-specific or involved operational or sales and marketing purposes. These are summarized in the table below.

Platforms for Modern Analytics

Once there is a business need for modern analytics, modernizing the data infrastructure is the next step for some companies. To support modern analytics and handle massive amounts of diverse data, companies (regardless of analytics maturity) are using different tools on different platforms to support all kinds of data.

In this survey, for instance, about 30% of respondents were collecting internal or external text data, with slightly less (about 25%) analyzing it. Likewise, about 25–30% of respondents stated that they are collecting machine-generated data, video data, image data, and other more complex data types (all not shown). To use these new data types, either to enrich data sets for analytics or on their own, the platform must be able to support them.

> *More respondents are utilizing cloud data warehouses than on-premises data warehouses.*

Cloud wins the day. As Figure 3 illustrates, overall, more respondents are utilizing cloud data warehouses than on-premises warehouses (46% versus 39%). Likewise, more respondents are using a cloud data lake than one on premises (26% versus 21%). The use of cloud data warehouses and cloud data lakes is more prevalent in respondents who were using modern analytics. We have seen upwards of 55% of respondents using cloud data

Type of use case	Examples
Vertical	Disease detection, disease prediction, surgical procedures, network performance, fraud prevention, threat assessment, determining the right premiums
Operational	Predicting no-shows, predicting maintenance, predicting field conditions, employee sentiment, application use, system uptime
Sales and Marketing	Predicting customer behavior/purchases, customer churn, forecasting future sales, likelihood to respond, next best action, recommendation engines, targeted communications for certain products

Summary of respondents' frequently cited use cases for modern analytics.

warehouses in other surveys and 40% using cloud data lakes.[2] What is interesting here is this is the first time we've seen cloud data warehouses and cloud data lakes deployed by more respondents than those platforms on premises. Additionally, 21% of respondents claimed to have deployed a unified platform that can serve as a data warehouse or data lake (sometimes called a cloud data lakehouse). In previous surveys, we've seen this at about 10%. There is, of course, overlap between those using these platforms; organizations rarely use only one platform. For instance, in this survey, close to 15% were utilizing a data warehouse in the cloud as well as one on premises (not shown). We've seen higher overlap in other surveys.

Data fabrics and semantic layers are important. Although cloud platforms are becoming more popular, organizations are also using other data management practices. For instance, in this survey, 16% of respondents were using a data fabric approach. The term *data fabric* has been used to describe a way to bring together disparate data in an intelligent fashion. The data fabric maps and connects relevant application data stores with metadata to describe data assets and their relationships. This approach can help unify data that might be on different platforms.

One approach to the data fabric design is accomplished with data virtualization, a method that integrates heterogeneous and distributed data across multiple platforms without replicating it. The approach creates a single virtual data layer that unifies data and supports multiple applications and users. Another approach is a semantic layer. Early semantic layers originated as functionality embedded and buried within larger tools, typically for reporting, enterprise business intelligence, data integration, and database management. A modern semantic layer is a standalone tool type that provides data semantics services for multiple tools within a multitool and multiplatform data architecture. It uses tools such as data virtualization or federation or newer approaches such as data catalogs, data lineages, or automated generation of data descriptions via knowledge graphs.

> *Data catalogs play a key role in data management by making it easier for users to search for and find data from diverse sources.*

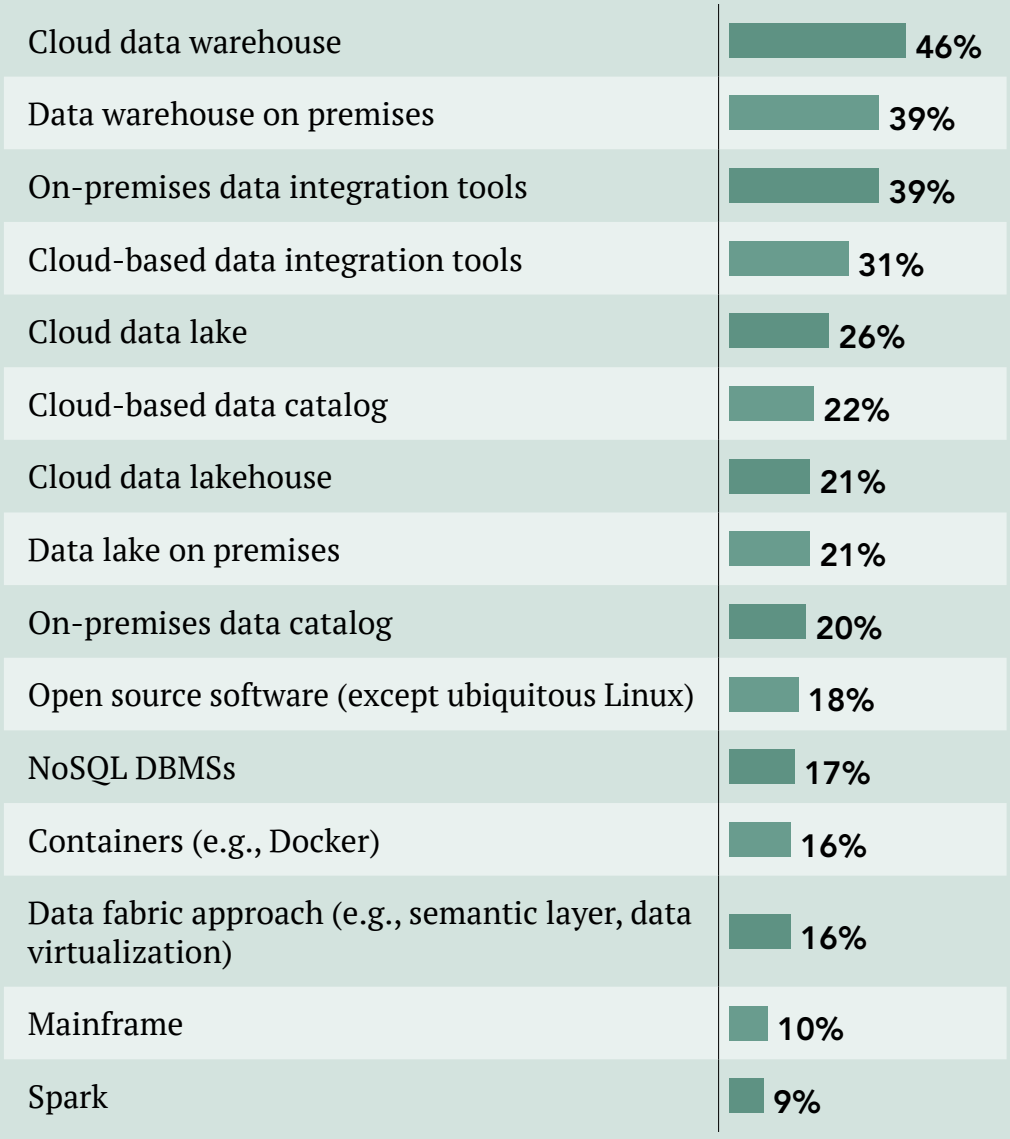
Data catalogs are gaining steam. Twenty-two percent of respondents to this survey are using a cloud data catalog; about the same percentage are using one on premises. Data catalogs, business glossaries, and metadata repositories collect information about how data is defined and modeled, its location, and how models and schema may have changed. Data catalogs play a key role in data management by making it easier for users to search for and find data from diverse sources. They help in semantic layers by providing descriptions about data. Some modern data catalogs include features for automated data cleansing, classifying sensitive data, and certification of data sets by their owners. Other solutions keep track of changes to data schema or structure. Data catalogs help provide trusted data for analytics, which is critical on the modern analytics journey.

[2] Unpublished 2022 TDWI survey.

Figure 3

What data management technologies is your organization using to support modern analytics? (Select all that apply.)

Based on 394 respondents.



Challenges with Modern Analytics

As noted, we see organizations often struggling to get to the next phase of analytics maturity for many reasons, both organizational and technical. In this survey we asked all respondents, "What challenges is your organization facing with modern analytics?" We asked the question two different ways: one way for those respondents not using modern analytics (illustrated in Figure 4) and another for those using modern analytics (Figure 5).

Organizational challenges lead the way. For those respondents who hadn't yet made the move to more modern analytics, the top reason cited was lack of skills (39%). A lack of data literacy (i.e., the ability for individuals across the organization to understand data, derive insights from it, and communicate results effectively) is a key roadblock in terms of analytics maturity. It blocks business users from using self-service tools and it prevents data analysts from using more sophisticated tools.

Lack of funding (38%) was the second most common challenge. About a third of respondents in this group (28%) stated that there is simply no

interest in technologies such as machine learning or natural language processing at their companies. Although only 16% stated that they don't have the support of executive management, this is typically because although executives say that analytics is valuable, they don't necessarily support the effort with funding, walk the walk, or create a culture where more advanced tools could be used.

In those organizations where more modern analytics is in use, talent is still an issue for analytics and for maintenance. Thirty-eight percent of this group said it was a challenge (Figure 5), second only to high complexity. For instance, modern analytics, such as machine learning, often increases the need for data engineers to gather and process the data as well as an MLOps team that puts models into production and manages them in production. Many organizations start with their data scientists supporting MLOps, but this isn't sustainable once many models are operationalized. Of these respondents, 15% face issues putting models into production and making sure that models stay fresh.

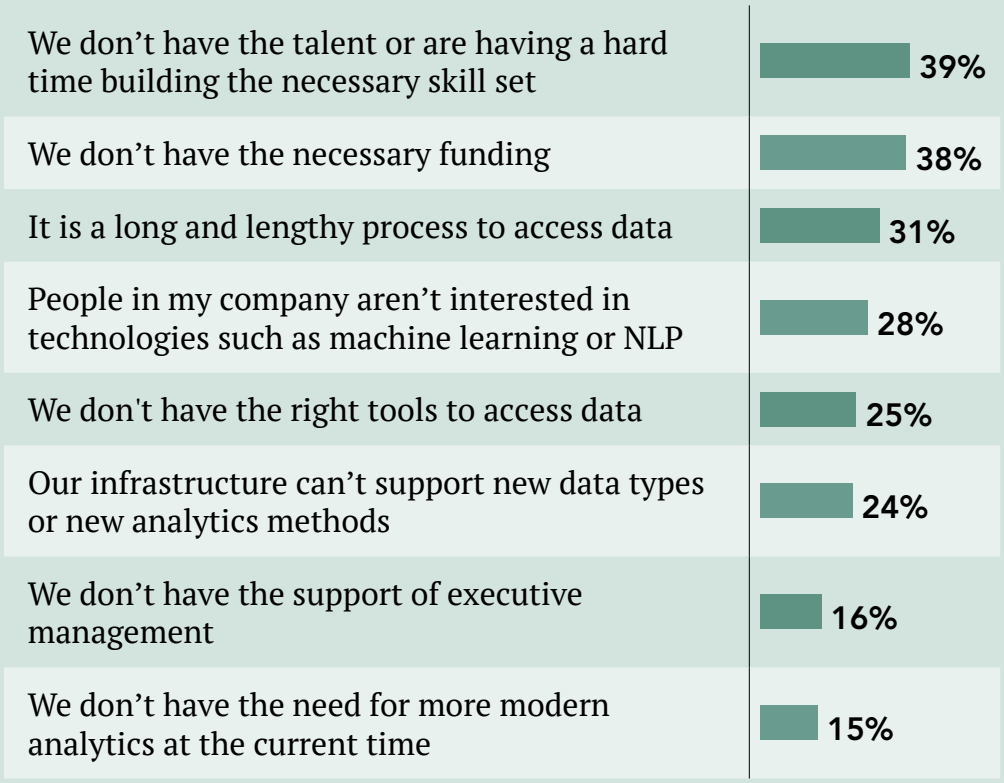
Technology challenges are also critical. Although there are numerous organizational and cultural challenges with adopting modern analytics, there are also technical challenges. This is the case regardless of where organizations are in terms of adopting modern analytics. For instance, both groups cited access to data as a problem (Figures 4 and 5). Twenty-four percent of those not using modern analytics currently state that their data infrastructure is an issue because it can't support new data types. In fact, more organizations utilizing modern analytics are making use of cloud platforms than on-premises platforms, as we saw earlier. (By comparison, only 10% of those using modern analytics cite infrastructure as a challenge.)

However, those enterprises currently using modern analytics still face technology challenges.

Figure 4

What challenges is your organization facing in trying to modernize its analytics? (Select all that apply.)

Based on 244 respondents who are not using modern analytics.



For instance, 39% state that the complexity of analytics is a problem. Additionally, 26% note they waste time on maintenance activities.

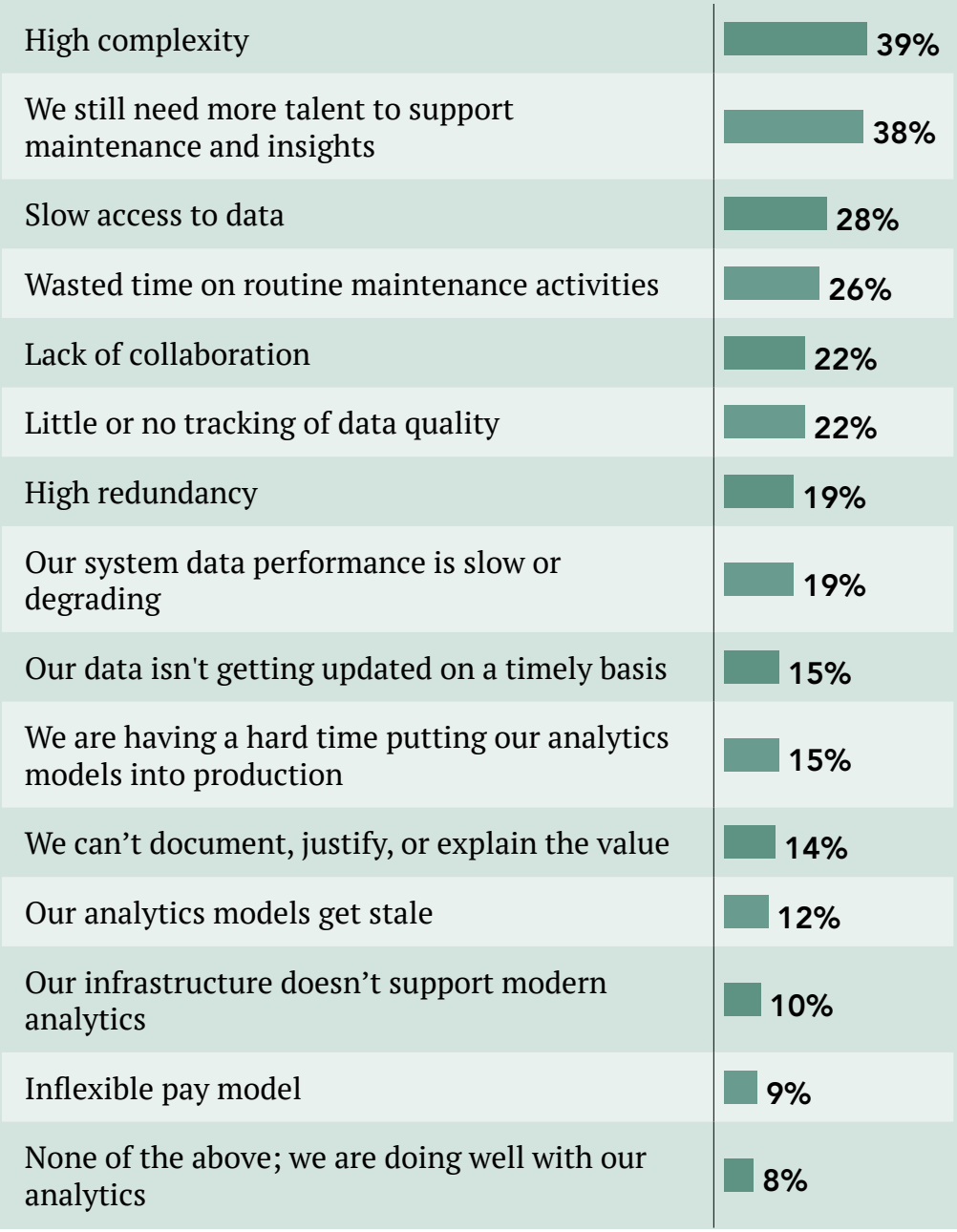
Close to 20% of respondents deploying modern analytics stated that complex data pipelines are a challenge (not shown). As organizations move to more modern analytics, pipelines will become even more critical. For instance, they will feed machine learning and other models. If the pipeline breaks, that will cause a problem. Organizations need a way to monitor these pipelines in real time and be alerted if the pipeline breaks.

Cloud-related challenges. Many organizations utilizing modern analytics are deploying cloud data platforms to support them. We asked, "What opportunities have you unlocked using cloud data platforms for analytics?" The top answer was scalability to support modern analytics (46%).

Figure 5

What challenges is your organization facing with modern analytics? Please select your top five challenges.

Based on 146 respondents using modern analytics.



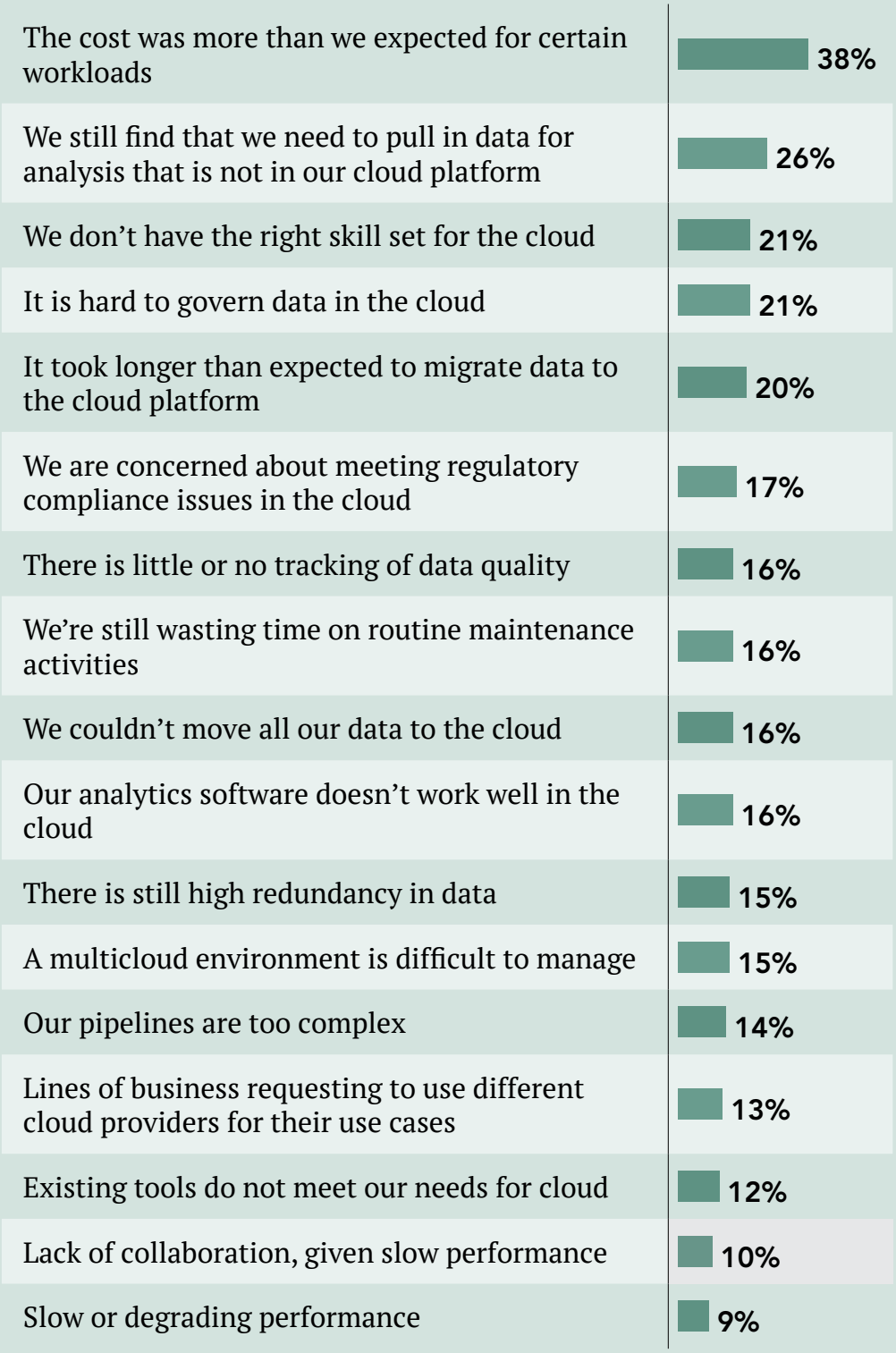
Other top opportunities include easier access to data (29%), easier to share and collaborate with data (25%), and greater agility for lines of business to explore new ideas (23%) (all not shown).

Of course, there are challenges with the cloud, but that does not mean these challenges outweigh the benefits. We will see later in this report that organizations using the cloud for modern analytics are more likely to be successful than those that

Figure 6

What issues have you had with the cloud? Please select your top five responses.

Based on 394 respondents.



do not. However, the cost associated with certain workloads was the top challenge cited by all respondents (Figure 6). We often hear from our audience that, especially when first starting out with the cloud, organizations may not monitor cloud usage and costs can get out of control. Additionally, 26% of all respondents noted that they often need to pull data that isn't on their cloud platform for analytics. At TDWI, we see that most organizations maintain some sort of hybrid environment, either intentionally (e.g., for certain kinds of data) or as a path to move to the cloud. Twenty-one percent stated that it is hard to govern data in the cloud.

Some organizations are using a data fabric approach to address the hybrid environment. As we saw, a small group is using this approach. Some of them have faced issues including time to get it off the ground, complexity, and cost.

New Technologies for Data and Analytics

In this research, we also learned more about how organizations are making use of new technologies and tools for data governance and analytics.

Automation and Augmentation

These new tools are often automated and augmented to help improve productivity and reduce complexity. Augmented tools are infused with advanced analytics such as machine learning to find patterns in data and perform tasks. Figure 7 illustrates some areas where automation is being used.

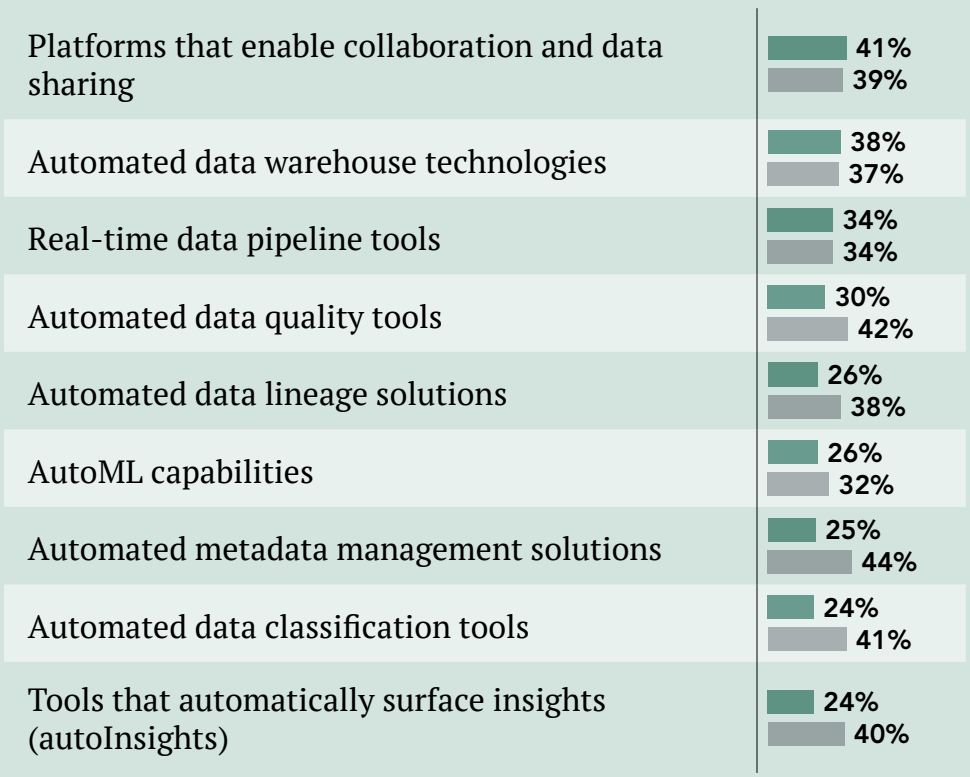
Figure 7

What data management and analytics tools are a priority as part of your company's analytics modernization effort?

Using now

Plan to use in next few years

Based on 394 respondents, in descending order by current use.



Cloud data warehouse automation. Data warehouse automation is a strategy for DW modernization. It's mostly about modern tools for DW design, development, and administration. In this survey, 38% of respondents were already using data warehouse automation today and that percentage is set to double if users stick to their plans. DW automation often involves automated data discovery and profiling, design and testing, documentation, and data update (often through change data capture).

Data governance automation. Over the past five or more years, as environments became more complex, we've seen the emergence of data governance tools that automate and augment data governance. These include data quality tools to automatically surface problems in data quality and help produce trusted information. In this survey, 30% were already using these tools; another 42% were planning to use them in the next few years. Likewise, organizations are utilizing tools for automated data lineage (26%), metadata management (25%), and data classification (24%).

Machine learning automation. Automation often occurs at different points in the machine learning life cycle. For instance, AutoML automates many aspects of machine learning workflow, which include data preprocessing, feature engineering, model selection, and sometimes model deployment. In some cases, all the user needs to do is provide the target and the system does the rest. It creates features, runs through many models and determines the best one, and then presents it to the user. In this survey, about 26% of respondents were utilizing AutoML for feature engineering and model building. Another 32% planned to use this technology in the next few years.

Other Tool Types

In this survey, 34% were already using real-time pipeline tools to help manage streaming data. When it comes to modern analytics, streaming data is important for use cases such as predictive maintenance via IoT devices or smart appliances. Once a model is built, it will also require fresh data for scoring.

Respondents are also making use of data sharing and collaboration tools. As we saw earlier, some companies believe that the cloud helps in these tasks. This is an important concept for modern analytics. Organizations need to be able to internally collaborate on data science projects. In external situations, supply chain partners should be sharing supply chain analytics. Organizations are sharing data and analytics with their customers on everything from inventory levels to alerts when the temperature gets too hot for certain produce on a freight train. In this survey, 41% of respondents were employing platforms that enable collaboration and data sharing tools. Thirty-nine percent were planning to use them in the next few years.

Respondents also noted their use of tools to help operationalize models in production. For example, in this survey, about a third of those utilizing modern analytics were monitoring models in production and slightly more were planning to do it over the next few years (not shown). Models get stale and degrade over time. It is important to monitor these models to see if this is occurring. Likewise, as machine learning use increases, it will also be important to ensure that models are explainable and do not contain bias.

USER STORY

Utilizing modern pipelines for analytics success

TDWI spoke to a senior director of cloud and DevOps at a company that provides a customer data platform that enables consumer marketers and brands to manage their first-party data from online and offline sources. The company collects large amounts of customer data that they need to put into their cloud platform. According to the senior director, "This data could be received in various ways including files, databases, FTP, and so on. The company is running thousands of pipelines at one time. We operate across cloud providers. We need to be able to predict how much data is ingested on a daily basis."

Along with managing a scalable pipeline process, he said there are other important factors when dealing with large numbers of pipelines. For instance, "Data quality is important; the quality check should be done in the pipeline. Validation rules are important. If you haven't created business rules for ingestion, the data will fall into tables and lead to errors." Additionally, he noted that, "Scheduling should be part of the kickoff plan."

This senior director also painted a picture of a monitoring platform to help with modern pipelines. He said, "It is important for pipeline tools to provide alerts if a pipeline breaks. What is needed is a unified dashboard tool to show a real-time pipeline [view of those] that are running and active, passive, stalled, and failed. It is important to show where it is failing; alerts that show that the pipeline took too long to run are important as well."

Characteristics of Successful Organizations

To explore best practices further, we wanted to understand the characteristics of those companies that were successful with modern analytics and if they differed from companies that were not reporting value.

We separated the respondents into three groups—those who self-identified as successful with modern analytics (19%), those who were somewhat successful (49%), and those who were not successful (23%). The rest didn't know (and were excluded from this analysis).

We compared the three groups across several dimensions, including demographics, organizational support, and technology (Figure 8). The sample size is small, but some interesting findings emerged that support points made earlier in this report. Of course, some of the findings may be a result of becoming more advanced, but the analysis is telling, nonetheless, and can help guide those who want to use modern analytics.

> *A committed analytics leader is key for achieving success with modern analytics.*

Those in the "successful" group have a committed analytics leader. Respondents who claim they are successful with modern analytics are more likely to have a committed analytics leader in place (82%) than those who are unsuccessful (51%). This leader must be someone—ideally in the C-suite—who is a champion for analytics and advanced analytics. As noted, that person needs to walk the walk, ensure a strategy, and evangelize and fund analytics, including more advanced analytics.

Data literacy is a priority. Respondents who claim they are successful with modern analytics are more likely to agree with the statement that data literacy is a priority in their organizations (78%) versus those who are unsuccessful (39%) and even those who are somewhat successful (44%). The difference was statistically significant. Data literacy training helps to democratize analytics. There can be a range of training types. For instance, training might consist of helping business users better understand the output of analytics on their ERP data. It might consist of putting a formal program in place. The point is that successful companies will recognize the importance of data literacy training and fund the effort. Once business users are data literate, they can analyze their own data and release business analysts to do more sophisticated work. Data scientists benefit too because they can ask better questions.

"Successful" enterprises are more likely to view the move to the cloud as a priority. Although moving to the cloud is important for all organizations, those respondents who were successful with modern analytics were more likely to state that a move to the cloud was a priority (69%) versus those who were unsuccessful (47%). Those who were somewhat successful were not significantly different from those who were successful. We saw earlier that many respondents were utilizing cloud data warehouses and data lakes. These platforms can support massive amounts of data. Data lakes handle unstructured data and other new data types. These capabilities are necessary for modern analytics. Examples include the massive amounts of structured data being fed into machine learning models, understanding unstructured data such as customer sentiment using NLP, and using deep learning to classify images.

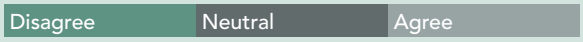
Those in the "successful" group are more likely to use automated tools. It isn't surprising that those respondents who are successful with modern analytics are more likely to use automated tools to handle the complexity. It would be difficult to do this manually and is something to which other organizations should pay attention. These automated tools are for data management and data governance as well as analytics. In this survey, respondents who claimed they are successful in modern analytics were statistically more likely to be using automated tools (81%)—for data warehouse automation, data lineage, data quality, AutoML, and autoInsights—than those who were not successful (42%) or those who were somewhat successful (51%). They are also more likely to be using modern pipeline tools (56%), such as real-time data pipeline tools, than those who were unsuccessful (30%) (not shown).

> *Successful organizations were more likely to have some sort of analytics governance in place.*

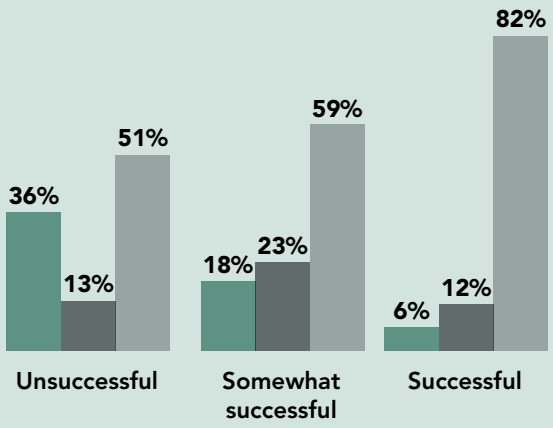
They are more likely to have analytics governance in place. Data governance is critical for modern analytics. Not only is it important for meeting compliance needs, it is also necessary to ensure that data is trustworthy. The old saying, "garbage in, garbage out" applies here. There may be a limited number of chances you have to ensure that members of the organizations want to use data for analysis. In order for this to happen, they need to trust the data. As one respondent to the survey said, "Modern analytics is only as good as the data that is used to fuel it. To be successful, organizations need to focus on data quality and governance, ensuring that data is accurate, complete, and compliant with relevant regulations." Those who have had success are more likely to put data quality standards in place than those who have not.

Figure 8

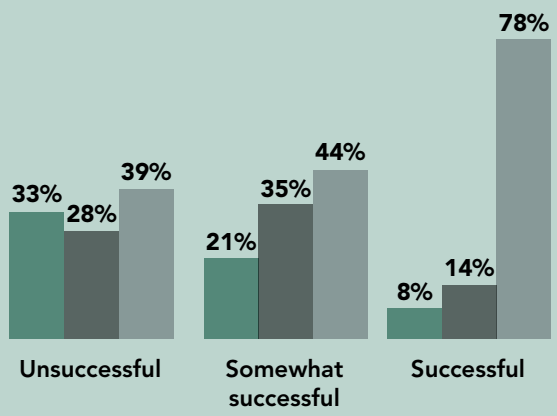
Do you agree or disagree with each of the following statements?



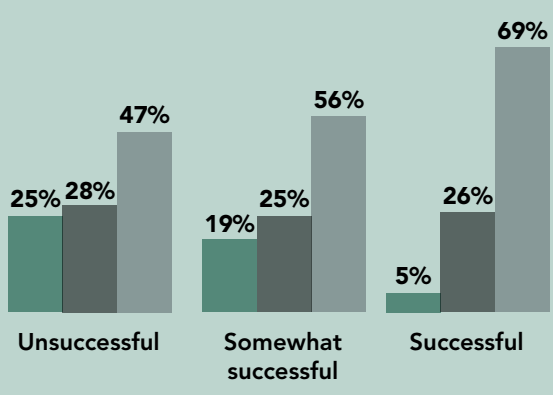
A committed analytics leader is present in the company



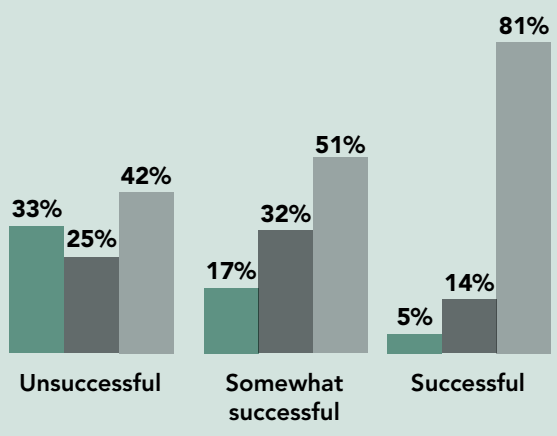
Data literacy is a priority at my company



Moving to a cloud data platform is a priority at my company



We have automated many data processes at my organization



Selected results based on 364 respondents who self-identified as successful, somewhat successful, or unsuccessful with modern analytics.

Additionally, those respondents who were successful were also more likely to have deployed analytics governance in addition to data governance. Analytics governance is a relatively new concept that deals with issues related to analytics such as how analytics is used as well as registering, versioning, documenting, and explaining models. This may involve extending data governance practices to include analytics governance. In some cases, it involves new offices such as the data office, headed by the CDO. In other cases, the MLOps team is responsible for analytics governance.

The point is that there is a group in charge. Those respondents who claimed to succeed with analytics were more likely to have some sort of analytics governance in place than those who were unsuccessful. Approximately 80% of those respondents who were unsuccessful in analytics did not have any formal documented procedures for governing their analytics initiatives (not shown).

> *Respondents who said they are successful with analytics are more likely to have some sort of organizational structure in place for analytics.*

"Successful" organizations have an organizational structure for analytics. At TDWI, we've been examining how companies are organized to execute for modern analytics, and we've been looking at three models. The first is the hub-and-spoke model, where a central office might be responsible for coordinating the kinds of tools used, training for those tools, governance, and the data infrastructure to support the effort. In a decentralized model, business units own their own analytics. There is also the centralized model, where data management and analytics happen in a central organization, such as IT.

Although more organizations may use the decentralized model, in this survey, no one model was more successful than another. We've seen this in other research, as well.[3] However, those respondents who said they are successful appear to be more likely to have *some* sort of organizational structure in place for analytics than those who are not successful.

It matters what you measure. Interestingly, KPIs were considered a priority by over 60% of each group; there was no significant difference. However, it often depends on what is measured. For instance, many respondents cited adoption of a solution as a metric for success, and it might well be, if the goal is democratizing analytics. There weren't very many KPIs mentioned that tied specifically back to business goals—and they should. As one respondent said, "I think that the most important part is to understand the need for the analysis. Sometimes the analysis is great, but it provides an answer to a problem that is not relevant."

Other factors. There are additional factors that are associated with successful modern analytics. For instance, larger companies (with over $1B of annual revenue) are more likely to be successful with analytics versus smaller companies. Second, collaboration and data sharing platforms are more likely to be in use at successful enterprises. Finally, those who claim to be successful are more likely to view monetizing data and analytics as a priority than those who are not successful.

[3] See for instance, the 2022 *TDWI Best Practices Report: Modernizing the Organization to Support Data and Analytics*, online at tdwi.org/bpreports.

USER STORY

Dealing with company culture on the road to becoming data-driven

A data management manager at a midsize retail company said that a goal of his company is to become more data-driven. As part of this effort, he has a unique role, saying, "I am the bridge between IT and the business, so I understand what the business is requesting."

Part of the plan was to empower business users in self-service analytics. "We didn't want IT being bottlenecked, and we wanted business users to be more self-sufficient. We guided the business users, showed them how to use the analytics tools, showed them where the data was. However, many business users prefer simply consuming the data in reports and dashboards. They studied retail management, not data joins. About 10% of the people we trained are power users."

The team decided that one way to get data to the people who needed it was to create KPIs and exception dashboards that go out daily to executives as well as store managers. All of the data is in a data warehouse that is updated nightly. According to this manager, "The dashboards are generally going to have the most impact on the organization. If someone has a more specific question, they go to a power user or go to IT."

The company has also been working on their data science effort. The most recent project has been to use machine learning for future inventory forecasting, allowing sales to set inventory levels accordingly. "We've been working on this for a few years," said the data manager. "There is hesitancy by some. Many people have been at the company for 30 or more years. Change can be a challenge at times. We stress that these enhancements aren't meant to take jobs away but to help them produce more effectively. We want to handle routine scenarios so SMEs can solve more interesting problems. It is an incremental process. However, the company is growing with more stores and products, so it is impractical to do everything manually." Currently, the company is testing the model at one store, and it is going well.

According to this manager, one best practice he has found is to "really immerse yourself in the line of business. I have a BS and MS in computer science, but I need to understand the needs of the business to do bridge building and for them to trust me. I need to understand how to translate the business problem into a data problem, so really understanding the business is essential. This can mean spending a few hours with a replenishment buyer or going to the store and talking to the employees. It is important to put yourself in their shoes to get an adequate solution."

Recommendations

This report has examined the state of and best practices for modern analytics. In closing, we explain the top best practices that can guide your organization to be successful with modern analytics. Remember, however, that becoming successful with modern analytics is a journey and your roles and processes will evolve.

Put people first. Being successful with modern analytics isn't just about the technology; it is also about the people. It will be important to have an executive who is a champion for the cause, including funding for it. Make sure to include people in the process, as they impact the culture that your organization is trying to build. As one respondent stated, "People and collaboration make the difference, not magic technology." Involve all stakeholders in the process early.

Put the right roles in place. Along with understanding the importance of the organization, it will be important to put the right roles in place and organize to execute (whatever the model—hub-and-spoke, centralized, or decentralized). This will include teams such as DataOps (to engineer data for modern analytics) as well as MLOps (to put models into production). It will include trainers (see below) as well as data scientists. It may even include data product managers if your company plans to sell data products. Plan for these roles early.

Measure the right thing. Although companies across the success spectrum claim to use KPIs, many times they don't measure the right thing. Think about your KPIs tied to business goals. Develop metrics from these KPIs. This will help you know whether your team is meeting its goals as well as help you illustrate success. That success can then build on itself.

Deploy data literacy enablement teams. Some organizations are investing in data literacy enablement teams, a relatively new but important role because data literacy is so critical for success in modern analytics. These teams determine what roles need what training, devise the strategy, and execute the plan. Well-oiled data literacy teams will assess data literacy and put together training programs (which may involve scheduled formal sessions or self-directed online practice) based on role and experience levels.

Think through your infrastructure. Modern analytics will require high volumes of diverse data. This most likely means a move to a platform that can scale and perform to manage these new data types. If you haven't already, plan to move to a new platform or architecture. This may involve a move to a cloud platform or deploying a data fabric approach to unify your data. When companies first move to the cloud, they often are surprised that the cost may be more than expected for certain workloads. We're seeing growing interest in FinOps, short for "financial operations," which is a set of practices and principles that enable organizations to manage their cloud computing costs effectively. In this approach, people from different teams, such as finance, operations, and engineering, come together to optimize cloud spending, increase accountability, and ensure the efficient use of cloud resources.

Remember that pipelines can break. Data pipelines are a critical component of your infrastructure, especially in the cloud. As your company deploys more analytics against more diverse data types, it will have more pipelines. However, pipelines can break (for instance, if a source system changes upstream). When a model is running and a pipeline breaks, your enterprise faces a problem—for instance, your recommendation engine might go down. Think about pipelines and their exposure. Employ pipelines that enable monitoring and alerts.

Don't forget that data quality matters for modern analytics success. As we've seen in this report, respondents who are successful have established data quality standards. Data quality matters, and it is going to matter even more as your organization modernizes. You will have to develop new standards for quality for new kinds of data you're using to enrich data sets for modern analytics. Additionally, your organization will need to ensure quality as you introduce new data types or risk turning off potential users of that data.

Plan for data and analytics governance. In line with data quality, your organization will also need to plan for advancing how it will ensure both data and analytics governance. This will include how to govern new models that might be put into production. Start planning for that sooner rather than later.

Make sure automated tools are on your road map. As data and analytics environments become larger and more complex, automation will be critical. Whether you use that automation to automate pipelines, identify sensitive data, or address data quality, these tools can help your organization become more productive. Additionally, automated and augmented tools in analytics (such as those that surface insights or build machine learning models) can help data scientists, business analysts, and business users all reach insights more quickly.

Don't forget about the importance of operationalizing analytics. Analytics provide value when they are actionable. This often involves putting analytics into production. Once your organization has enough models, it will be important to put some an operations group in place, such as an MLOps team that can treat the models like software—which includes versioning, monitoring, and updates.

Consider analytics products and monetizing analytics. As companies mature, they start to think more about the concept of data products. Data products are derivative assets created from data. These products can run the gamut, from enriched data sets provided to a customer to a dashboard that provides the output of machine learning models to external partners—or apps that use derived data for a specific industry. Some companies even enable their core ERP systems to provide data products. These can provide value to customers and partners and can also be monetized by organizations. Start thinking about your data as an asset early and thoughtfully explore business opportunities as your company matures.

How Resilient, Governed Pipelines Drive Modern Analytics Success

The demand for analytics from stakeholders across the enterprise has never been higher. With volatile markets, constant competitive pressure, and continually increasing customer expectations, every organization needs all its employees to make data-driven decisions. However, as TDWI shared, most businesses (75%) struggle with getting to modern analytics.

StreamSets provides the critical data integration component that lets data teams deliver on modern analytics promises. Here's how:

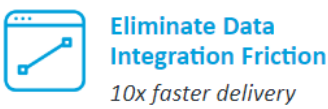
Getting to Modern Analytics	How StreamSets Makes It Easy
Get connected across environments and data types	StreamSets data integration platform supports complex hybrid and multicloud environments (even mainframes) and supports all data types. It provides flexibility for data engineers to get data from various sources, formats, and databases and lets them easily move it to a wide variety of destinations.
Use resilient data pipelines	StreamSets data pipelines resist data drift—the constant, unexpected changes to data structure, semantics, and infrastructure. Decoupled, schema-light data pipelines are split into independent ingest, store, process, and consume layers so data teams can keep up with constant change.
Improve data trust and automate monitoring with data governance	StreamSets mission control panel and topologies show how systems are connected and data flows across the enterprise. Data SLAs and rules enable you to expose hidden problems in your data flows, create guardrails and quality checks, and then manage by exception. Automatically see when a new integration point is being created and if there is a more direct route for the data. And you can know where data comes from to help understand and explain outcomes, for example, in AI/ML models.
Support data fabric	StreamSets prioritizes making managed data available through an open metadata-sharing model. Metadata produced through StreamSets data pipelines is made available for data fabric initiatives such that you can collate platform metrics, data, schemas, and transform logic automatically and analyze it in-house or send it to tools like Collibra and Apache Atlas.
Enable self-service	StreamSets low-code/no-code interface lets the people who know the data best build their own pipelines. Teams can innovate at their own pace without additional development time from the data engineering team.
Unify data integration options to increase efficiency and reduce silos	StreamSets also lets you build batch, CDC, ETL, ELT, and ML pipelines—and all from a single UI. That mission control panel mentioned earlier also lets you seamlessly integrate data from any system—on premises, cloud, hybrid, or even mainframe.

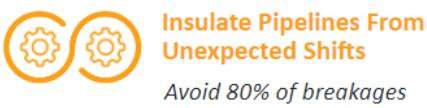
Leading Enterprises Trust StreamSets

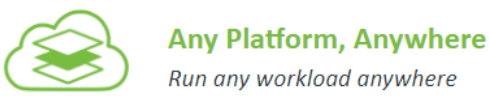
The largest companies in the world trust StreamSets to power millions of data pipelines for modern analytics, data science, smart applications, and hybrid integration.

See how these and other leading companies have modernized data integration with StreamSets

StreamSets' data integration platform provides these valuable benefits:







Request a demo to see StreamSets in action

TDWI Research provides research and advice for data professionals worldwide. TDWI Research focuses exclusively on data management and analytics issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of data management and analytics solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

A Division of 1105 Media
6300 Canoga Avenue, Suite 1150
Woodland Hills, CA 91367

E info@tdwi.org

tdwi.org