

# Magic Quadrant for Data Science and Machine-Learning Platforms

**Published:** 22 February 2018    **ID:** G00326456

**Analyst(s):** Carlie J. Idoine, Peter Krensky, Erick Brethenoux, Jim Hare, Svetlana Sicular, Shubhangi Vashisth

## Summary

Data science and machine-learning platforms enable organizations to take an end-to-end approach to building and deploying data science models. This Magic Quadrant evaluates 16 vendors to help you identify the right one for your organization's needs.

## Market Definition/Description

*This document was revised on 26 February 2018 and again on 12 March 2018. The document you are viewing is the corrected version. For more information, see the Corrections ([http://www.gartner.com/technology/about/policies/current\\_corrections.jsp](http://www.gartner.com/technology/about/policies/current_corrections.jsp)) page on gartner.com.*

This Magic Quadrant evaluates vendors of data science and machine-learning platforms. These are software products that data scientists use to help them develop and deploy their own data science and machine-learning solutions. More precisely, Gartner defines a data science and machine-learning platform as:

A cohesive software application that offers a mixture of basic building blocks essential both for creating many kinds of data science solution and incorporating such solutions into business processes, surrounding infrastructure and products.

Machine learning is a popular subset of data science that warrants specific attention when evaluating these platforms.

Cohesiveness is an important attribute of a data science and machine-learning platform. The basic building blocks should be integrated into a single platform. The platform's components should have a consistent "look and feel" and be interoperable across the entire analytics "pipeline" – from accessing and analyzing data to operationalizing and managing content.

A data science and machine-learning platform supports data scientists in the performance of tasks across the entire data and analytics pipeline. These include tasks relating to data access and ingestion, data preparation, interactive exploration and visualization, feature engineering, advanced modeling, testing, training, deployment and performance engineering.

Building your own data science and machine-learning models internally is not the only option, however. For outsourcing options, see "Market Guide for Data Science and Machine-Learning Service Providers." In addition, some vendors offer analytic solutions focused on specific industries or types of analysis. This Magic Quadrant does not evaluate such vendors specifically, but it does assess whether the featured vendors offer prepackaged content to address specific needs.

This year's Magic Quadrant includes the term "machine learning" in its title (compare 2017's "Magic Quadrant for Data Science Platforms" ). Although data science and machine learning are slightly different, they are usually considered together and often thought to be synonymous. "Machine learning" is the term most often used in vendors' marketing materials, and it frequently appears in research on this market. The Magic Quadrant's new name reflects machine learning's momentum and its significant contribution to the broader discipline of data science.

Readers should know that:

Gartner invited a diverse mix of data science platform vendors to participate in the evaluation process for potential inclusion in this Magic Quadrant, as data scientists have different preferences for UIs and tools. Some prefer to code data science models in Python or R; others favor Scala or Apache Spark; some like to run data models in spreadsheets; others are more comfortable building models by creating visual pipelines via a point-and-click UI. Tool diversity is an important characteristic of this market.

The wide range of products available offers a breadth and depth of capability and varied approaches to developing and deploying models. It is therefore important to evaluate your specific needs when assessing vendors. A vendor in the Leaders quadrant, for example, may not be your best choice. For an extensive review of the functional capabilities of each platform, see "Critical Capabilities for Data Science and Machine-Learning Platforms."

Open-source platforms are excluded from this Magic Quadrant if they have no vendor supporting them as commercially licensable products. Commercially licensed open-source platforms are included. We also recognize the growing trend for commercial platforms to use open-source libraries and content. Open-source solutions represent an opportunity to get started with data science and machine learning with little upfront investment (see Note 1).

Artificial intelligence (AI) is the subject of considerable hype, but cannot be ignored. Data science is undoubtedly a core discipline for the development of AI. Machine learning is a core enabler of AI, but not the whole story. Machine learning is about creating and training models; AI is about using models to infer conclusions under certain conditions. A self-driving car, for example, has machine-learning capability, but its AI requires much more than that.

The diversity of data science platforms largely reflects the diverse types of data scientist who use them. This Magic Quadrant is therefore aimed at a variety of audiences:

Line of business (LOB) data science teams. Typically, these are sponsored by their LOB's executive and charged with addressing LOB-led initiatives in areas such as marketing, risk management and CRM. They focus on their own priorities. Levels of collaboration with other LOB data science teams vary.

Corporate data science teams. These have strong and broad executive sponsorship, and can take a cross-functional perspective from a position of enterprisewide visibility. In addition to supporting model building, they are often charged with defining and supporting an end-to-end process for building and deploying data science and machine-learning models. They often work in partnership with LOB data science teams in multitier organizations.

Adjacent, "maverick" data scientists. These are typically one-off scientists in various LOB units. They tend to work independently on "point" solutions and usually strongly favor open-source tools, such as Python, R and Apache Spark. They rarely collaborate much with other data scientists in their organization.

Holders of additional roles, such as citizen data scientist, data engineer and application developer. They need to understand the nature of the data science and machine-learning market, and how it differs from, but complements, the analytics and business intelligence (BI) market (see "Magic Quadrant for Analytics and Business Intelligence Platforms" ).

## Magic Quadrant

**Figure 1.** Magic Quadrant for Data Science and Machine-Learning Platforms



Source: Gartner (February 2018)

## Vendor Strengths and Cautions

### Alteryx

Alteryx (<https://www.alteryx.com/>) is based in Irvine, California, U.S. It offers a unified machine-learning platform, Alteryx Analytics, which enables citizen data scientists to build models in a single workflow. In mid-2017, Alteryx acquired Yhat, a data science vendor focused on model deployment and management. Alteryx issued an initial public offering (IPO) on the New York Stock Exchange in early 2017, which strengthened its ability to invest in expanding and enhancing its platform's capabilities.

Alteryx has progressed from the Challengers quadrant to the Leaders quadrant. This is thanks to strong execution (in terms of both revenue growth and customer acquisition), impressive customer satisfaction, and a product vision focused on helping organizations instill a data and analytics culture without needing to hire expert data scientists.

## STRENGTHS

**Focus on business users and citizen data scientists:** Customers often choose Alteryx because its platform is easy to learn and use. Alteryx has differentiated itself by offering a platform for business users and citizen data scientists. This approach addresses the skills shortage and enables people with the right domain experience and business acumen to build and run models.

**Platform vision:** Alteryx's vision is for its platform to serve different kinds of user with equal ease of use and confidence. The inclusion of more automation and rule-based recommendations, as outlined in the product roadmap, will make it easier to operationalize models. In early 2018, Alteryx plans to introduce a new product, Alteryx Promote, which will incorporate Yhat capabilities, to provide model management and deployment functionality.

**Customer satisfaction:** Alteryx focuses intently on ensuring that customers derive business value from its platform. Alteryx's surveyed reference customers scored it among the top vendors for overall customer experience and satisfaction. They reported that Alteryx's support and community provide excellent guidance, and that the vendor is extremely responsive.

## CAUTIONS

**Perception as only a data preparation solution vendor:** Although Alteryx has achieved solid revenue growth, until recently most of its customers had been using its platform primarily for its data preparation capabilities. Although data preparation is a key aspect of data science, Alteryx needs to market itself more aggressively as a vendor of a complete data science platform. To its credit, it is already shifting its go-to-market strategy to emphasize its complete platform capabilities, specifically for citizen data scientists, but it must keep doing so to improve its visibility.

**Automation, reporting and visualization gaps:** To accelerate its platform's adoption by citizen data scientists, Alteryx needs to add automated model-building and selection capabilities. In addition, its reporting and visualization capabilities remain comparatively weak. Customers reported that the reporting features are not as intuitive as the rest of the platform. Alteryx plans to address its shortcomings in reporting and visualization as part of an integration with Plotly during 2018, which will provide interactive in-line visualizations.

**Enterprise readiness:** Alteryx needs to continue making its platform more "enterprise-grade." For example, Alteryx Server does not support Linux OS and currently can run on only a single machine. And, while Alteryx strives to meet enterprise requirements, it must support backward compatibility to ensure seamless product migrations.

Anaconda

Anaconda (<https://www.anaconda.com/>) , formerly known as Continuum Analytics, is based in Austin, Texas, U.S. It sells Anaconda Enterprise 5.0, an open-source development environment based on the interactive-notebook concept. It also provides a loosely coupled distribution environment, giving access to a wide range of open-source development environments and open-source libraries, mainly Python-based.

Anaconda's strength lies in its ability to federate and provide a central access point for a very large number of Python developers who build machine-learning capabilities continuously. However, Anaconda has little or no control over those developers' efforts in terms of quality, dependability and sustainability. Anaconda nurtures a broad developer community through Anaconda Cloud. Anaconda's position as a Niche Player reflects its suitability for seasoned data scientists fluent in Python.

## STRENGTHS

**Python and open-source support:** Anaconda provides an open-source development center active through Anaconda Cloud. The growing popularity of Python among data scientists gives Anaconda excellent visibility to developers, for whom it provides a wide range of code, libraries, notebooks and shared projects. Anaconda is the only data science vendor not only supporting but also indemnifying and securing the Python open-source community, to make the platform suitable for enterprises.

**Flexible code integration:** Anaconda's environment has exceptionally complete and flexible capabilities for integrating Python code libraries. This is important because developers can use them to tailor Anaconda to available project and compute resources. Surveyed users praised these capabilities.

**Active ecosystem:** Reference customers praised Anaconda's extensive, active community engagement. The community fosters technological development, cutting-edge Python code libraries and integration with other open-source data science projects. Although most libraries are still in the early stages of development, some of the most advanced capabilities are available through those beta libraries.

## CAUTIONS

**Focus on experts and minimal automation:** Anaconda targets experienced data scientists familiar with Python and notebooks. Novice Anaconda users will have difficulty finding their way through the Python "jungle." The "do it yourself" skill and attitude exhibited by typical Anaconda users does not readily accept the imposition of machine-learning automation practices. Python developers tend to automate their practices themselves, rather than rely on automation mechanisms promoted by more business-oriented audiences.

**Lack of comprehensive sales strategy:** Although Anaconda has grown steadily, its sales team still takes the "classic" approach of converting open-source Python users to Anaconda Enterprise through online marketing and evangelism. To accelerate growth, Anaconda should tap into the community that it created and continues to foster.

**Poorly organized documentation:** Anaconda offers an excessive amount of documentation and training materials. These badly need better organization and clarity.

# Angoss

Angoss (<http://www.angoss.com/>) , which is based in Toronto, Canada, was acquired by Datawatch in January 2018. It still appears as Angoss in this document due to the acquisition's lateness, relative to the Magic Quadrant process, and uncertain impact. This evaluation covers the following products: KnowledgeSEEKER, the company's most basic offering, aimed at citizen data scientists in a desktop context; KnowledgeSTUDIO, which includes many more models and capabilities than KnowledgeSEEKER; and the newly launched KnowledgeENTERPRISE, a flagship product that includes the full range of capabilities.

Angoss has lengthy experience with banking customers. This underpins its ability to deliver to the banking sector and other sectors with similar data and analytical needs, such as insurance, transportation and utilities. Angoss has loyal customers, but remains a Niche Player as it is still perceived as a vendor for desktop environments. It recently added a series of enterprise functionalities to the platform (namely model management, cloud and open-source functionalities), but too recently for evaluation in this Magic Quadrant. Angoss focuses on nonindustrial clients and use cases.

## STRENGTHS

**Ease of use and well-rounded functionality:** The user-friendliness and ease of use of the company's core products, KnowledgeSEEKER and KnowledgeSTUDIO, earned Angoss solid scores for most of the critical capabilities assessed.

**Some strong features:** Angoss has significantly improved its open-source support by allowing users to integrate some cutting-edge capabilities into its visual composition framework (Spark ML, TensorFlow and H2O, for example). In terms of deployment, Angoss has better-than-average capabilities to export models into SAS, SQL, Predictive Model Markup Language (PMML) and, partially, into Java.

**Integration of other advanced analytics features:** Angoss has a good breadth of tightly integrated optimization capabilities (linear and nonlinear constrained optimization, for example) and text analytics (via an OEM relationship with Lexalytics). In addition, Angoss enables model training to be conducted seamlessly either on-premises or in the cloud.

## CAUTIONS

**Market traction:** With over 20 years' experience, Angoss is an industry veteran, but its market traction should be stronger than it is. With just over 300 loyal clients, the company's user community remains modest. Angoss' reputation is still that of an easy-to-use desktop tool vendor, even though many customers use Angoss' server environment. Without significant marketing and sales efforts, it remains to be seen whether the company can materially accelerate adoption of its technology.

**Customer concerns:** Our survey of reference customers shows that overall satisfaction with Angoss' platform is moderate, but some reference customers had concerns about aspects of product stability, flexibility and data access capability. We expect the company to address these concerns in upcoming releases.

**Innovation speed:** Angoss is lagging behind in a few important respects: for example, it has yet to embrace the trend for automating core machine-learning processing, and it has limited traction in the AI service market. Its upcoming model factory capabilities could help remedy these shortcomings.

## Databricks

Databricks (<https://databricks.com/>) is based in San Francisco, California, U.S. It offers the Apache Spark-based Databricks Unified Analytics Platform in the cloud. In addition to Spark, it provides proprietary features for security, reliability, operationalization, performance and real-time enablement on Amazon Web Services (AWS). Databricks announced a Microsoft Azure Databricks platform for preview in November 2017, which is not considered in this Magic Quadrant because it was not generally available at the time of evaluation.

Databricks is a new entrant to this Magic Quadrant. As a Visionary, it draws on the open-source community and its own Spark expertise to provide a platform that is easily accessible and familiar to many. In addition to data science and machine learning, Databricks focuses on data engineering. A 2017 Series D funding round of \$140 million gives Databricks substantial resources to expand its deployment options and fulfill its vision.

### STRENGTHS

**Center of the Spark ecosystem:** Founded by creators of Apache Spark, Databricks uses its key position in the Spark ecosystem to grow its customer base. The Spark user community is expanding, because Databricks spearheads numerous Spark meetups, Spark Summits and training courses integrated with the Databricks Community Edition. Many companies introduce machine learning by using Spark as a starting point. Experienced organizations often select the Spark ecosystem to further strengthen their business.

**Work with large datasets:** Databricks optimizes its infrastructure for performance and scalability, and pays special attention to large datasets. As a result, Databricks' platform outperforms plain Apache Spark. Reference customers are especially pleased with its SQL performance, and with the exposure of deep learning in SQL as a feature of sparkdl.

**Innovation:** Databricks' innovation in open-source software, streaming and the Internet of Things (IoT) accounts for its Visionary status. Reference customers like its turnkey notebooks for interactive collaboration and support of multiple languages (SQL, Python, R and Scala) on data from various sources. Databricks' innovative infrastructure approaches to cluster management and serverless capabilities enable execution of machine-learning models at scale.

### CAUTIONS

**Limited market awareness:** Despite the marketing strategy that underpins Databricks' impressive growth, much of the market is not aware of the fully managed Databricks platform built on Spark, and is instead buying Spark support from other cloud vendors or Hadoop distributors.

**Cost tracking:** The overall cost of running Databricks' platform consists of the cost of the underlying cloud capacity, which is paid to the cloud provider directly, and the native cost of the Databricks platform. Although Databricks reduces the Spark total cost of ownership (TCO) for comparable loads, reference customers identified difficulties with control, analysis and monitoring of third-party cloud expenses.

**Debugging capabilities:** Most customers use Databricks for "do it yourself" machine learning. In addition to the debugging capabilities that Databricks already offers, reference customers wish the vendor could provide debugging features better suited to the needs of data scientists. Databricks would also benefit from an integrated development environment (IDE) with comprehensive facilities for enterprise-grade debugging, development and version control, in addition to the currently offered IDE on GitHub that leaves many reference customers dissatisfied.

## Dataiku

Dataiku (<https://www.dataiku.com/>) is headquartered in New York City, U.S., and has a main office in Paris, France. It offers Data Science Studio (DSS) with a focus on cross-discipline collaboration and ease of use.

Dataiku remains a Visionary – and a popular choice for many data science needs – by enabling users to start machine-learning projects rapidly. Its position for Completeness of Vision is due to its collaboration and open-source support, which are also the focus of its product roadmap. Its overall Completeness of Vision score is lower than in the prior Magic Quadrant, due to comparatively poor breadth in terms of use cases and deficiencies in automation and data streaming. Dataiku's Ability to Execute has also decreased, due to some difficulties in operationalizing and scaling machine-learning models.

## STRENGTHS

**Collaboration across skill sets:** Dataiku DSS is differentiated by having multiple collaboration features across the machine-learning pipeline. It offers three profiles for users with different skill levels – from dashboard-type graphical UIs for less-skilled users to a visual pipelining tool for intermediate-level users. Data scientists can use coding features, including shells and notebooks that offer more flexibility but also require more in-depth knowledge. Many reference users observed that the collaborative nature of Dataiku DSS has democratized machine learning across their organization.

**Flexibility and openness:** Dataiku DSS enables machine-learning algorithms to be "plugged in" from the open-source community. Users can choose to create models using native machine-learning code recipes, leading machine-learning engines (such as those of H2O.ai and Apache Spark MLlib), notebooks (such as Jupyter Notebook), and language wrappers for R, Python and Scala. There is good integration with Hadoop and Spark.

**Ease of use and rapid prototyping:** An intuitive interface makes Dataiku DSS a popular tool for rapid prototyping. Reference customers reported significant time savings when using the tool for proofs of concept, rapid prototyping, and even a test-and-learn approach to get

those who are not data scientists started with machine-learning projects. Although the platform lacks certain advanced functionalities, many reference customers use it for product development and research and development (R&D).

## CAUTIONS

**Operationalization of machine-learning algorithms:** Reference customers pointed to some difficulties deploying models in production environments and migrating to the latest version. Several mentioned instabilities and bugs, although they also reported that Dataiku was quick to fix them. DSS received low scores for delivery and performance and scalability.

**Lack of model factory automation and advanced analytics functionality:** Dataiku's product scores for automation are low, with deficiencies in model factory capabilities, such as the use of machine learning to automatically propose or select features and for model optimization. DSS does not offer a full range of capabilities using advanced analytics techniques such as simulation and image analytics. However, the company's roadmap includes plans to offer native prebuilt deep-learning models for text and images, in addition to the existing integration with TensorFlow.

**Pricing:** High prices are a concern, with a significant number of reference customers identifying them as inhibitors of wider adoption. Reference customers also gave Dataiku low scores for end-user evaluation and contract negotiation experience.

## Domino

Domino (<https://www.dominodatalab.com/>) (Domino Data Lab), headquartered in San Francisco, California, U.S., offers the Domino Data Science Platform. This is an end-to-end solution for expert data science teams. The platform focuses on integrating tools from both the open-source and proprietary-tool ecosystems, collaboration, reproducibility, and centralization of model development and deployment. Founded in 2013, Domino is a recognized name in this market and continues to gain mind share among expert data scientists.

Domino maintains its position as a Visionary. Its Ability to Execute, though improved, is still hampered by weaker functionality at the beginning of the machine-learning life cycle (data access, data preparation, data exploration and visualization). Over the past year, however, Domino has demonstrated the ability to win new accounts and gain traction in a highly competitive market.

## STRENGTHS

**Open-source innovation and tool-agnostic approach:** Domino's offering is well-designed and positioned to capitalize on the popularity of open-source technologies by offering freedom of choice in terms of tools. Of the vendors in this Magic Quadrant, Domino earned the highest overall score for flexibility, extensibility and openness.

**Outstanding customer service and support:** Domino has maintained the high standard of customer satisfaction that we recognized in the prior Magic Quadrant. Reference customers gave favorable reviews to Domino's service and customer support (onboarding and troubleshooting). Reference customers also praised their overall experience with the vendor.

**Responsiveness to customers' requests for product improvements:** Domino received one of the highest overall scores for inclusion of requested product enhancements into subsequent product releases. Domino is proactive in supporting the latest open-source tools and has acted on requests for strong delivery and model management capabilities. Consequently, it received excellent scores for its overall capabilities and ability to meet clients' needs.

**Collaboration:** Domino has two years of excellent customer feedback about its collaboration functionality, which earned it the overall highest score for this critical capability among the vendors in this Magic Quadrant. Domino provides excellent features to enable data scientists to offer transparency and work effectively with nontechnical users.

## CAUTIONS

**High technical bar:** Domino's offering is not a good choice for citizen data scientists, as it offers neither visual pipelining nor a visual composition framework. However, some expert data scientists did recognize improvements to the interface's ease of use. Domino's scores for analytic support (training and technique selection) were below average.

**Beginning of the machine-learning life cycle and business exploration:** Domino's score for data access was in the bottom quartile. Its scores for data preparation, data exploration and visualization, and business exploration were mediocre.

**Few quick or "precanned" solutions:** Domino's offering is not a lightweight tool for "quick and easy" data science. Users must navigate a vast open-source ecosystem for precanned solutions for common use cases in marketing, sales, R&D, finance and other areas.

**Advanced enterprise-grade capabilities:** Some reference customers complained of a lack of functionality when, for example, working with cloud infrastructures and supporting large and complex hybrid cloud environments.

## H2O.ai

H2O.ai (<https://www.h2o.ai/>) , which is based in Mountain View, California, U.S., offers an open-source machine-learning platform. For this Magic Quadrant, we evaluated H2O Flow, its core component; H2O Steam; H2O Sparkling Water, for Spark integration; and H2O Deep Water, which provides deep-learning capabilities.

H2O.ai has progressed from Visionary in the prior Magic Quadrant to Leader. It continues to progress through significant commercial expansion, and has strengthened its position as a thought leader and an innovator.

## STRENGTHS

**Technology leader:** H2O.ai scored highly in categories such as deep-learning capability, automation capability, hybrid cloud support ("deploy anywhere") and open-source integration. H2O Deep Water offers a deep learning front end that abstracts away many of the details of back ends like TensorFlow, MXNet and Caffe. Its machine-learning automation capabilities (dubbed Driverless AI) are impressive and, although still developing, demonstrate the company's distinguished vision. In terms of flexibility and scalability,

reference customers considered H2O.ai to be first class. It has one of the best Spark integrations, and is ahead of all the other Magic Quadrant vendors in its graphics processing unit integration efforts.

**Mind share, partners and status as quasi-industry standard:** H2O.ai's platform is now used by almost 100,000 data scientists, and many partners (such as Angoss, IBM, Intel, RapidMiner and TIBCO Software) have integrated H2O.ai's technology platform and library. This shows the company's technical leadership, which especially derives from the highly scalable implementation of some core algorithms.

**Customer satisfaction:** H2O.ai's reference customers gave it the highest overall score for sales relationship and account management, evaluation and contract negotiation experience, customer support (including onboarding and troubleshooting), and overall service and support. They also gave it outstanding scores for analytic support (including training and technique selection), integration and deployment, inclusion of requested product enhancements in subsequent product releases, and overall experience with the vendor.

## CAUTIONS

**Ease of use:** H2O.ai's toolchain is primarily code-centric. Although this typically increases flexibility and scalability, it impedes ease of use and reuse.

**Data preparation and interactive visualization:** These capabilities are problematic for all code-centric platforms, of which H2O.ai's is one. Nonetheless, H2O.ai's platform will prove challenging for clients expecting more interactivity and better, easier-to-use data ingestion, preparation and visualization capabilities. Capabilities for the entire early part of the data pipeline are far less developed than the quantitative parts of H2O.ai's offerings.

**Business model:** H2O.ai is a full-stack open-source provider — even its most advanced products are free to download (except for the closed-source Driverless AI). H2O.ai derives nearly all its revenue from subscriptions to technical support. Given the development of H2O.ai's revenue lately, we are slightly less concerned than we would otherwise have been about this policy of "giving everything away." However, we still maintain that this policy is difficult to scale. In the long run, H2O.ai will need to consider a more scalable software-licensing model and support infrastructure.

## IBM

IBM (<http://www.ibm.com/>) , which is based in Armonk, New York, U.S., provides many analytic solutions. For this Magic Quadrant, we evaluated SPSS, including both SPSS Modeler and SPSS Statistics. Data Science Experience (DSX), a second data science and machine-learning offering, did not meet our criteria for evaluation on the Ability to Execute axis, but does contribute to IBM's Completeness of Vision.

IBM is now a Visionary, having lost ground in terms of both Completeness of Vision and Ability to Execute, relative to other vendors. IBM's DSX offering, however, has potential to inspire a more comprehensive and innovative vision. IBM has announced plans to deliver a new interface for its SPSS products in 2018, one that fully integrates SPSS Modeler into DSX.

## STRENGTHS

**Market understanding:** IBM remains a leader in terms of market share, with 9.5% of the data science market. Its strategy, focused on the complete analytic pipeline, enables both expert and novice data scientists to be productive.

**Innovative approach focused on key market trends:** With the inclusion of DSX, IBM's roadmap offers extensive openness, hybrid cloud support and strong analytic capabilities for both expert and novice data scientists across the full analytic pipeline.

**Data preparation and model management capabilities:** IBM SPSS is a trusted and vetted enterprise solution. IBM's robust data preparation capabilities and ability to operationalize and manage models are key strengths and differentiators.

**"Legacy" user base across all analytic capabilities:** IBM's strong user base in both the analytics and BI market and the data science and machine-learning market gives it an advantage in terms of not only maintaining usage but extending and increasing it. Customers often undertake due diligence with a view to maintaining or extending their existing investments before considering a move to a different vendor.

## CAUTIONS

**No single, comprehensive offering:** With the aging IBM SPSS Modeler undergoing renovation and DSX continuing to develop, IBM currently has no single, comprehensive, modern data science platform. A revamped UI for SPSS is planned to be generally available in 2018, however.

**Continuing transition to DSX:** DSX promises a new, more comprehensive approach to data science and machine learning, but development and rollout of this offering is still ongoing.

**Multipronged approach:** There remains confusion in the market about IBM's Watson branding. By now, we would have expected IBM SPSS Modeler to be fully integrated into the IBM Watson ecosystem. In addition, there is confusion about the relationship of, and distinction between, SPSS and DSX. The general availability in August 2017 of the Watson Machine Learning service, which assists operationalization, model management and workflow automation, exacerbates the confusion.

**Poor customer experience, operations and perceived high cost:** Feedback from reference customers on their customer experience with IBM was unfavorable, including low scores for inclusion of enhancements/requests into subsequent releases, overall rating of product capabilities and business value delivered. IBM's operations also scored poorly, with low scores for documentation, customer support and analytic support.

## KNIME

KNIME (<https://www.knime.com/>) is based in Zurich, Switzerland. It provides the fully open-source KNIME Analytics Platform, which is used by over 100,000 people worldwide. KNIME offers commercial support and commercial extensions to boost collaboration, security and performance for enterprise deployments.

In the past year, KNIME has introduced cloud versions of its platform for AWS and Microsoft Azure, paid more attention to data quality, expanded its deep-learning features, and converted some of its commercial capabilities to open source. Bolstered by a €20 million investment in 2017, KNIME is accelerating its product development and customer acquisition efforts.

KNIME's platform is used by most industries and in most regions of the world. The vendor demonstrates a deep understanding of the market, a robust product strategy and strength across all use cases. Together, these attributes have solidified its place as a Leader.

## **STRENGTHS**

**Low TCO:** KNIME's unwavering commitment to the open-source approach enables many users and organizations to minimize their data science software costs without compromising quality. The open-source nature of its data access, methods and techniques makes it a good choice for collective innovation. KNIME's pricing of its commercial offerings aims to make data science affordable. In particular, KNIME enables many nonprofit organizations to undertake data science.

**Cohesive platform for data scientists of all skills levels:** KNIME provides a single, consistent data science framework. It offers highly rated data access and manipulation capabilities, a breadth of algorithms, and a comprehensive machine-learning toolbox suitable for both beginners and expert data scientists. KNIME's platform integrates with other tools and platforms, such as R, Python, Spark, H2O.ai, Weka, DL4J and Keras. KNIME's contextual help with "what comes next" is more flexible than fixed "wizards." The UI and extensive examples provided with the platform appeal to citizen data scientists.

**Automation of model creation and deployment:** KNIME Model Process Factory offers automation of model creation and deployment, as well as of the modeling process, as per the Cross Industry Standard Process (CRISP). It also has automated approaches to data quality and feature generation. KNIME can trigger model retraining and supports automated data refresh and synchronization.

## **CAUTIONS**

**Lack of marketing and sales innovation:** Although KNIME has over 1,000 paying customers, many companies are unaware of it. The vendor has limited sales and marketing resources, and those it has focus on solution engineering and partner programs. This focus helps customers succeed with current tasks, but does not instill new ideas, which are needed in this rapidly changing market.

**Performance and scalability:** Reference customers reported issues with large-scale deployments and performance on large datasets. A KNIME Server deployment is currently limited to a single host. The KNIME Analytics Platform is designed to "mix and match" resources, but should do a better job of explaining resource recommendations. The outlook for KNIME's performance and scalability is positive, however, thanks to its newer offerings in the cloud.

**Limited commercial options:** Reference customers want more commercial options to match their specific needs. They also want better security, in-depth training and enterprise-grade platform management capabilities. Despite having one of the widest global customer populations, KNIME lacks truly global services and sometimes provides patchy, although competent, support.

## MathWorks

MathWorks (<https://www.mathworks.com/>) is a privately held company headquartered in Natick, Massachusetts, U.S. Its two major products are MATLAB and Simulink, but only MATLAB met the inclusion criteria for this Magic Quadrant.

MathWorks remains a Challenger. Its Ability to Execute is aided by its sustained visibility in the general advanced analytics field, a significant installed base and strong customer relations, but impaired by average scores from reference customers for critical capabilities. Its Completeness of Vision is limited by its focus on engineering and high-end financial use cases, largely to the exclusion of customer-facing use cases like marketing, sales and customer service.

### STRENGTHS

**Presence and user loyalty:** Despite the rapidly growing popularity of open-source technologies, MathWorks remains one of the most prominent vendors in the data science sector. The company has decades-long relationships with many customers, and its long-standing users are generally pleased with its stability and with MATLAB's reliability.

**Customer success stories and availability of toolboxes:** MathWorks can point to strong customer success stories in its areas of strategic focus, particularly engineering, manufacturing and quantitative finance. Toolbox extensions offer different types of user the functionality they need for individual projects, such as machine-learning, optimization, computer vision and robotics projects.

**Communication, account management and support:** Reference customers gave MathWorks excellent scores for sales relationship and account management. MathWorks maintains a strong global presence, and customers also gave it outstanding scores for onboarding and troubleshooting and overall service and support. They are also very pleased with their product evaluation and contract negotiation experience.

### CAUTIONS

**Customer-facing use cases:** MathWorks' vision for data science does not focus on marketing, sales or customer service. Data science teams using MATLAB for engineering and scientific work should consider alternatives for marketing, sales and customer service use cases.

**Lateness to include open-source innovations:** Although MATLAB remains a pre-eminent data science platform for engineering problems, and now has the ability to call and be called from Python, MathWorks has not joined the movement to provide first-class support for open-source languages. MathWorks remains committed to serving its large core of loyal users, but a new generation of data scientists and statisticians prefers to work with the

evolving open-source ecosystem. In addition, MathWorks does not natively support deep-learning packages such as Caffe and TensorFlow. Instead, it offers functionality to import those packages into its own deep-learning framework. Still, the benefits of native support should not be dismissed.

**Inclusion of product enhancements requested by customers:** MathWorks has added new capabilities relating to the industrial IoT, edge analytics and popular cloud performs. However, MathWorks' reference customers gave it among the lowest overall scores for inclusion of requested product enhancements into subsequent product releases. MathWorks needs to prioritize carefully the diverse and highly technical evolving needs of a demanding customer base.

## Microsoft

Microsoft (<http://www.microsoft.com/>) , which is based in Redmond, Washington, U.S., provides a number of software products for data science and machine learning. In the cloud, it offers Azure Machine Learning (including Azure Machine Learning Studio), Azure Data Factory, Azure Stream Analytics, Azure HDInsight, Azure Data Lake and Power BI. For on-premises workloads, Microsoft offers SQL Server with Machine Learning Services, which was released in September 2017 – after the cutoff date for consideration in this Magic Quadrant. Only Azure Machine Learning Studio fulfilled the inclusion criteria for this Magic Quadrant, although Microsoft's broader advanced analytics offerings did influence our assessment of its Completeness of Vision.

Microsoft remains a Visionary. Its position in this regard is attributable to low scores for market responsiveness and product viability, as Azure Machine Learning Studio's cloud-only nature limits its usability for the many advanced analytic use cases that require an on-premises option.

## STRENGTHS

**Cloud infrastructure approach:** Microsoft's cloud-only approach with Azure Machine Learning Studio enables strong capabilities in terms of flexibility, extensibility and openness. This approach also lends itself well to performance tuning and scalability. The cloud model enables frequent updates, so that users can take early advantage of improvements.

**Marketing execution and mind share:** Microsoft is a trusted brand. It has a presence in most organizations and its momentum continues to build. As such, Azure Machine Learning Studio has good visibility and resonance as a platform for providing comprehensive capabilities across the full range of descriptive, diagnostic, predictive and prescriptive analytic types.

**Solid operations and support:** Reference customers scored Microsoft's operational support for Azure Machine Learning Studio highly. They value its customer support (including onboarding and troubleshooting), analytic support, and overall service, integration and deployment support.

**Innovative, visionary roadmap:** Microsoft's roadmap stands out as visionary. Its innovative capabilities are reflected in Microsoft's commitment to open-source, deep-learning, streaming and IoT use cases, and its focus on the end-to-end analytic process, including operationalization of models.

## CAUTIONS

**Cloud-only applicability:** Microsoft's cloud-only approach with Azure Machine Learning Studio, though powerful in some respects, is limiting in others. Although cloud capability lends itself well to the advanced-prototyping use case, it causes issues for business exploration and production refinement. This is because most customers want to interact easily with on-premises data and deploy their models on-premises – especially larger enterprises looking for a hybrid cloud solution. Microsoft's recently announced preview services and capabilities to enable deployment anywhere via containers should help in this respect.

**Delivery capabilities:** Azure Machine Learning Studio received low scores from reference customers for its delivery capabilities, due to shortcomings in terms of code synthesis, containerization and embedded delivery options."

**Market responsiveness and traction:** Relative to other vendors in this market, Microsoft has achieved slower uptake, primarily due to its product's immaturity compared with other vendors' offerings.

**Work-in progress capabilities:** Azure Machine Learning Studio is used primarily by data scientists and application developers, and remains a relative newcomer. While providing some strong functionalities and capabilities, such as a solid UI, flexibility, extensibility and openness, many other capabilities, such as delivery, data preparation and coherence across the platform, are lacking.

## RapidMiner

RapidMiner (<https://rapidminer.com/>) is based in Boston, Massachusetts, U.S. Its platform includes RapidMiner Studio, RapidMiner Server and RapidMiner Radoop. RapidMiner Studio is the model development tool, available as both a free edition and a commercial edition; it is priced according to the number of logical processors and the amount of data used by a model. With the free edition, customers get one logical processor and 10,000 rows of data. RapidMiner Server is designed for sharing, collaborating on and maintaining models. RapidMiner Radoop extends RapidMiner's execution directly into a Hadoop environment.

RapidMiner remains a Leader by delivering a well-rounded and easy-to-use platform to the full spectrum of data scientists and data science teams. RapidMiner continues to emphasize core data science and speed of model development and execution by introducing new productivity and performance capabilities.

## STRENGTHS

**Positive market response:** RapidMiner focuses equally on data scientists and citizen data scientists – it aims for completeness and simplicity in its offerings. Reference customers expressed satisfaction with how RapidMiner's offerings meet their needs. Accordingly,

RapidMiner's revenue has grown impressively, year over year.

**Model factory and model development:** RapidMiner provides deep and broad modeling capabilities for automated end-to-end model development. For example, RapidMiner Studio includes a visual workflow designer and guided analytics. RapidMiner's process execution framework provides a flexible, extensible and scalable capability for running and monitoring large numbers of model processes. RapidMiner supports automatic retraining of models, based on newer data. Marketplace Extensions add further capabilities to its products, such as text mining, web crawling and integration.

**Ease of use:** Given the complexity and sophistication of data science endeavors, data scientists value ease of use highly, as it enables them to be more productive. RapidMiner delivers this value by providing an intuitive interface, easy access to data sources, simple programming for developing models and easily understood results. Reference customers reported a short learning curve.

## CAUTIONS

**Reduced attention to open-source software:** Although most of RapidMiner's offering remains open-source, the company continues to add new functionality that is available only with a commercial license. Its new approach to commercial pricing and license management are challenging to some customers and make pricing unpredictable in some cases. Reference customers observed that RapidMiner's licensing sometimes limits usage of its offerings where they would otherwise fit well.

**Data visualization:** Although RapidMiner offers standard and even advanced data visualizations, they are rigid and hard to work with. Data visualization enhancements could help provide the ease of use sought by data scientists.

**Documentation and training:** Reference customers are not fully satisfied with RapidMiner's documentation and online help. Although basic documentation is satisfactory, it lacks depth, best practices and advanced tutorials. Reference customers want better education options. RapidMiner is trying to mitigate many of its documentation and training problems with a free customer success program.

## SAP

SAP (<http://www.sap.com/>) is based in Walldorf, Germany. It has yet again rebranded its platform: SAP Business Objects Predictive Analytics is now simply SAP Predictive Analytics (PA). This platform has a number of components, such as Data Manager for dataset preparation and feature engineering, Automated Modeler for citizen data scientists, Predictive Composer for more sophisticated machine learning, and Predictive Factory for operationalization. SAP Leonardo Machine Learning and other components of the SAP Leonardo ecosystem did not contribute to SAP's Ability to Execute position in this Magic Quadrant.

Over the past year, SAP has made good progress in several respects, but still lags behind in others. It is a Niche Player due to low customer satisfaction scores, a lack of mind share, a fragmented toolchain, and significant technological weak spots (in relation to the cloud, deep

learning, Python and notebooks, for example), relative to others.

## STRENGTHS

**Collaboration across roles:** SAP has increased the integration of its two core machine-learning environments (Automated Modeler and Predictive Composer), to enable use by both expert and novice data scientists. The improved integration also encourages collaboration between roles.

**Business-integrated machine learning:** SAP's vision, evident in SAP Leonardo, of a unified machine-learning fabric across all its applications, is unique. SAP's new Predictive Analytics Integrator (PAI) is a good start. The first Leonardo applications demonstrated were SAP Fraud Management and SAP Customer Retention, but it remains to be seen whether they will scale and integrate with potential machine-learning deployment points (such as SAP SCM and SAP Forecasting and Replenishment).

**Some strong product capabilities:** SAP PA is especially good at automating many tasks and deploying across a range of business applications. It can also scale to handle very large datasets, especially via its tight integration with SAP HANA. The size of the datasets processed determines the licensing cost – a great simplification.

## CAUTIONS

**Customer experience and mind share:** SAP has one of the lowest overall customer satisfaction scores in this Magic Quadrant. Its reference customers indicated that their overall experience with SAP was poor, and that the ability of its products to meet their needs was low. SAP continues to struggle to gain mind share for PA across its traditional customer base. SAP is one of the most infrequently considered vendors, relative to other vendors in the Magic Quadrant, by those choosing a data science and machine-learning platform.

**Fragmented and ambiguous toolchain:** Multiple tools contribute to the SAP data science and machine-learning experience. Various role-based tools create a machine-learning pipeline or implement different sections of the process flow at different levels (for example, to implement and push down data preparation and data quality preprocessing to the database source in SAP HANA). SAP PA Automated Modeler and Expert Analytics target different roles. SAP HANA offers pipeline development and scripting capabilities to database developers or data scientists enabled to work in-database. This fragmentation results in confusion and cumbersome version management. In addition, the UI is cluttered and difficult to use.

**Technology delays:** SAP is lagging behind in key technology areas, such as capabilities for cognitive computing (in relation to vision, text, audio and video) and "deploy anywhere" cloud capabilities for its core machine-learning pipeline. SAP was one of the last vendors to integrate with Python and deep-learning capabilities, although it announced TensorFlow integration in August 2017. Integration with Python came with PA 3.3 in November 2017, after the cutoff date for evaluation in this Magic Quadrant. Additionally, it is still early days for SAP's Leonardo Machine Learning activities, and reference customers' feedback on SAP PAI was unavailable at the time of writing. Furthermore, SAP's vision for Leonardo Machine Learning seems rather decoupled from that for PA.

# SAS

SAS (<http://www.sas.com/>) is based in Cary, North Carolina, U.S. It provides many software products for analytics and data science. For this Magic Quadrant, we evaluated SAS Enterprise Miner (EM) and the SAS Visual Analytics suite of products, which includes Visual Statistics and Visual Data Mining and Machine Learning.

SAS remains a Leader, but has lost some ground in terms of both Completeness of Vision and Ability to Execute. The Visual Analytics suite shows promise because of its Viya cloud-ready architecture, which is more open than prior SAS architecture and makes analytics more accessible to a broad range of users. However, a confusing multiproduct approach has worsened SAS's Completeness of Vision, and a perception of high licensing costs has impaired its Ability to Execute. As the market's focus shifts to open-source software and flexibility, SAS's slowness to offer a cohesive, open platform has taken its toll.

## STRENGTHS

**Broad base and good visibility and mind share:** SAS again leads in terms of total revenue and number of paying clients. Customers are familiar with its brand and its extensive support for multiple use cases. Reference customers indicated that SAS is the vendor that most frequently appears on shortlists for product evaluation. Its partner network enhances its visibility and support.

**Modern architecture:** SAS Viya represents a modernized architecture and the foundation of SAS's technological developments. SAS EM can fully exploit the capabilities in SAS Viya architecture, which gives customers multiple deployment options. SAS's Visual Analytics suite is generally available.

**Appeal to a broad range of users:** SAS's offerings appeal to all types of user – from business analysts to citizen data scientists to expert data scientists. The Visual Analytics suite on the Viya architecture contributes to this appeal.

**Operational excellence:** SAS's comprehensive worldwide support infrastructure is unmatched. Customers choose SAS for its robust, enterprise-grade platform capabilities, from exploration to modeling to deployment. Reference customers gave high scores to SAS's documentation, customer and analytic support, and overall service and support.

## CAUTIONS

**Pricing and sales execution:** SAS's reference customers gave scores for product evaluation and contract negotiation experience that were in the bottom quartile. In addition, SAS's pricing remains a concern. Free open-source data science platforms are increasingly used along with SAS products as a way of controlling costs, especially for new projects.

**Complex and confusing multipronged approach:** Offering two platforms that are not fully interoperable and that have multiple components with different dependencies increases confusion and complexity in terms of managing, deploying and using SAS's products. The coexistence of SAS Viya and other SAS platform versions perpetuates the perception of a

lack of cohesion. Although SAS has made some progress in this regard, migration remains an issue for those that want to exploit Viya's capabilities but are not currently on that architecture.

**Product and sales strategy:** New entrants to this market have changed its landscape by offering open, innovative platforms and new approaches. The increased competition they bring requires "traditional" vendors, such as SAS, not only to respond but to proactively provide comprehensive, cohesive platforms. Some reference customers reported that SAS was slow to support new technologies and to act on requests for new features.

**Lack of capabilities across both platforms:** Both SAS EM and the SAS Visual Analytics suite received low scores, in comparison to other vendors, for data access capabilities and flexibility, extensibility and openness, and coherence and collaboration. Reference customers also gave SAS low scores for its lack of open-source support and deep-learning algorithm capabilities (although Viya partially addresses this issue).

## Teradata

Teradata (<https://www.teradata.com/>) is based in San Diego, California, U.S. The Teradata Unified Data Architecture (UDA) is an enterprise analytical ecosystem that combines open-source and commercial technologies to deliver analytic capabilities. The UDA includes Aster Analytics, a Teradata database, Hadoop and data management tools. Although Teradata has strong operationalization capabilities, it still lacks a unified end-to-end technology platform.

Teradata has maintained its intrinsic performance and reliability strengths, but its lack of cohesion and ease of use on the data science development side have impaired both its Ability to Execute and its progress on the Completeness of Vision axis. It remains a Niche Player.

## STRENGTHS

**Performance and scalability:** Teradata's main strength lies in its proven ability to operationalize, at scale, machine-learning solution deployments. Its reliability and sustainable performance are fundamental to its customers' loyalty. Its competitors' advances in cloud-based environments are starting to challenge that loyalty, however.

**Existing data infrastructure:** Teradata benefits from a long history in industrial-strength data warehousing for multiple business sectors and solutions. Teradata uses this experience to provide machine-learning capabilities via SQL. The company's deep understanding of complex data management is an important asset, given the increasingly wide variety of data that organizations have to handle.

**Internal experts and consultants:** Teradata's expertise in deploying advanced analytical solutions has fostered a corporate culture rich in experience and a deep reservoir of knowledge that customers can tap. The same culture is at the root of Teradata's consulting excellence. Teradata Think Big Analytics, a pure-play analytics consulting organization, is an important customer enabler with a good understanding of analytics for data lakes and data warehouses.

**Precanned solutions:** Teradata has continuously formalized and packaged its experience in the form of technology assets that integrate domain expertise. Advanced customer analytics, fraud and anomaly detection, risk management and IoT solutions are further evidence of Teradata's support for application domains.

## CAUTIONS

**Pricing:** Given Teradata's strength at scale and its deeply entrenched solutions at many large customers, it has been able to fend off competitors at the high end of the market. However, as Teradata's competitors are building scalable solutions, some prospective customers might be discouraged by the need for significant upfront investment. They should therefore explore Teradata's new flexible licensing options.

**Model development capabilities:** Teradata has lacked an appealing analytical asset development environment. Many of its customers have admitted to using an alternative model development environment, and it seems that this trend will persist in the short term. To mitigate this deficiency, Teradata recently introduced additional data science capabilities by integrating with solutions from vendors such as Dataiku and KNIME.

**Platform coherence:** The name "Teradata Unified Data Architecture" suggests a unified approach, but its components and configurations are multiple. Existing and prospective Teradata customers admit that it could be daunting to select, assemble and adjust the platform's various elements.

**Open-source acceptance:** Teradata is contributing to open-source environments, from both a development and a deployment perspective, but it remains to be seen how the open-source developer community will respond, and whether it will make full use of Teradata's power.

## TIBCO Software

TIBCO Software (<https://www.tibco.com/>) is based in Palo Alto, California, U.S. Building on its presence in the analytics and BI sector, TIBCO entered the data science and machine-learning market by acquiring the well-established Statistica platform from Quest Software in June 2017. Additionally, in November 2017, TIBCO announced the acquisition of Alpine Data, a Visionary in the prior Magic Quadrant. In terms of Ability to Execute, this Magic Quadrant evaluates only TIBCO's ability with the Statistica platform. Other acquisitions by TIBCO contribute only to its Completeness of Vision.

TIBCO enters this Magic Quadrant as a Challenger. The Statistica platform has a large and mature customer base, and received high scores for the three most typical use cases: business exploration, advanced prototyping and production refinement.

## STRENGTHS

**Aggressive expansion:** TIBCO acquired both Statistica and Alpine Data in 2017. Its vision for integrating these acquisitions into the TIBCO analytics ecosystem is developing rapidly, and TIBCO has many strong brands with well-tested functionality to weave together.

**Operationalization capabilities:** TIBCO's Statistica platform received excellent scores for aspects of its delivery (write-back, recoding and code synthesis, for example), automation (the model factory, for instance) and model management (metadata management, versioning and champion/challenger testing, for example). Together, these produced a top-quartile score for the production refinement use case. TIBCO also scored in the top quartile for the business exploration and advanced-prototyping use cases.

**"Connected Intelligence" strategy and IoT support:** TIBCO's Connected Intelligence strategy and accompanying products offer a strong ecosystem for IoT use cases. Statistica was already recognized for its excellent support for IoT and edge analytics, and the platform will benefit from TIBCO's middleware and event-processing expertise, as well as from the capabilities acquired from Alpine Data.

**Breadth of customer support:** TIBCO received high scores from reference customers for its support and service for onboarding, troubleshooting, analytic training and technique selection. They also gave it excellent scores for overall integration and deployment, and overall service and support.

## CAUTIONS

**Statistica's recent history of ownership change:** In the past four years, the Statistica platform has moved from Statsoft to Dell to Quest and on to TIBCO. The vision for the platform has understandably changed several times. It remains to be seen whether Statistica has finally found a permanent home that will support a long-term strategy.

**Performance and scalability:** The Statistica platform received one of the lowest overall scores from reference customers for performance and scalability. Further integration with TIBCO products and components could address this shortcoming, however.

**Cloud-native capabilities:** Statistica needs stronger cloud-native capabilities. With its acquisition of Alpine Data, TIBCO indicated its intent to merge Alpine's cloud capabilities with Statistica's functionality, but this work is still in its early stages.

**End-to-end platform integration:** Although TIBCO's acquisitions bring much-needed capabilities and strong assets, seamless integration of these assets will be a challenge for some organizations.

## Vendors Added and Dropped

We review and adjust our inclusion criteria for Magic Quadrants as markets change. As a result of these adjustments, the mix of vendors in any Magic Quadrant may change over time. A vendor's appearance in a Magic Quadrant one year and not the next does not necessarily indicate that we have changed our opinion of that vendor. It may be a reflection of a change in the market and, therefore, changed evaluation criteria, or of a change of focus by that vendor.

### Added

Anaconda

Databricks

## Dropped

FICO, which did not pass the inclusion gates.

Quest, because its platform, Statistica, has been acquired by TIBCO Software.

Alpine Data, because of TIBCO Software's announced intention to acquire it. Due to the late timing of this development, relative to the Magic Quadrant process, TIBCO is assessed only on its Statistica platform.

## Inclusion and Exclusion Criteria

The inclusion criteria have not changed substantially from those of the prior Magic Quadrant. The inclusion process included requirements for vendors to meet a revenue threshold and identify reference customers. A stack ranking process assessed how well products support the most typical use case scenarios for data science and machine-learning, namely:

**Business exploration:** This is the classic scenario of "exploring the unknown." It requires extensive data preparation, exploration and visualization capabilities for new and established data sources and types. This scenario may involve increased emphasis on the use of "smart" capabilities, to guide data preparation, use of visualizations and analysis, that incorporate machine-learning techniques "under the covers." Gartner refers to these capabilities collectively as "augmented analytics."

**Advanced prototyping:** This scenario covers the kinds of project in which data science and machine-learning solutions are employed to significantly improve on traditional approaches to addressing business problems. Traditional approaches can involve human judgment, exact solutions, heuristics and data mining. Projects typically involve some or all of the following:

- Many complex data sources, such as structured, unstructured and streaming sources

- Novel analytic approaches, such as deep neural nets, ensembles and natural-language processing

- Significant computing infrastructure requirements

- Specialized skills, such as in coding, SQL and statistics

**Production refinement:** In this scenario, the organization has several data science solutions delivered to and implemented by the business, but the focus is on refining, improving and updating existing models.

We used the following 15 critical capabilities when scoring vendors' data science platforms across the three use-case scenarios:

1. **Data access:** How well does the platform support the accessing of many types of data (such as tables, images, graphs, logs, time series, audio, texts)?

2. **Data preparation:** Does the platform have a significant array of noncoding or coding data preparation features?
3. **Data exploration and visualization:** Does the platform allow for a range of exploratory steps, including interactive visualization?
4. **Automation:** Does the platform facilitate automation of feature generation and hyperparameter tuning?
5. **User interface:** Does the product have a coherent "look and feel" and an intuitive UI, ideally with support for a visual pipelining component or visual composition framework?
6. **Machine learning:** How broad are the machine-learning approaches that are easily accessible from, or prepackaged and shipped with, the platform? Does the offering also include support for modern machine-learning approaches like ensemble techniques (boosting, bagging and random forests) and deep learning?
7. **Other advanced analytics:** How are other methods of analysis, from statistics, optimization, simulation, text analytics and image analytics, integrated into the development environment?
8. **Flexibility, extensibility and openness:** How can various open-source libraries be integrated into the platform? How can users create their own functions? How does the platform work with notebooks?
9. **Performance and scalability:** How can desktop, server and cloud deployments be controlled? How are multicore and multinode configurations used?
10. **Delivery:** How well does the platform support the ability to create APIs or containers (such as code, Predictive Model Markup Language [PMML], Portable Format for Analytics [PFA] and packaged apps) that can be used for faster deployment in business scenarios?
11. **Platform and project management:** What management capabilities does the platform provide (such as for security, compute resource management, governance, reuse and version management of projects, auditing lineage and reproducibility)?
12. **Model management:** What capabilities does the platform provide to monitor and recalibrate hundreds or thousands of models? This includes model-testing capabilities, such as K-fold cross-validation, training, validation and test splits, area under the curve (AUC), receiver operating characteristic (ROC), loss matrices, and testing models side-by-side (for example, champion/challenger [A/B] testing).
13. **Precanned solutions:** Does the platform offer "precanned" solutions (for example, for cross-selling, social network analysis, fraud detection, recommender systems, propensity to buy, failure prediction and anomaly detection) that can be integrated and imported via libraries, marketplaces and galleries?
14. **Collaboration:** How do users with different skills work together on the same workflows and projects? How can projects be archived, commented on and reused?

15. **Coherence:** How intuitive, consistent and integrated is the platform to support an entire data analytics pipeline? The platform itself must provide metadata and integration capabilities for the preceding 14 capabilities. It must also provide a seamless end-to-end experience to make data scientists more productive across the whole data and analytics pipeline, from accessing data to generating insight to recommending actions to measuring impact. This metacapability should ensure that data input/output formats are standardized, wherever possible, so that components have a consistent "look and feel" and terminology is unified across the platform.

The most significant change to the scoring process this year was in the alignment of specific subcriteria to each of the critical capabilities. This enabled more consistent and detailed evaluation of each of the critical capabilities' consistency across all platforms. In addition, we have adjusted the weighting criteria within the defined bands to place more emphasis on the platform and less on market presence, responsiveness and viability. This helps to "level the playing field" for some of the newer, less established vendors, which are demonstrating strong capabilities and innovative products in a way that is transforming the market. This adjustment reflects customers' increased focus on product capability and reduced emphasis on market performance. In addition, we have adjusted the weightings for each use case to reflect more accurately the percentage of time survey respondents indicated that they spend on each case.

To qualify for inclusion in the Magic Quadrant, each vendor had to pass the following assessment "gates":

### **Gate 1: Revenue and Number of Paying Customers**

Three common license models were assessed, and revenues (and/or customer adoption) from each were combined (if applicable):

**Perpetual license model:** Software license, maintenance and upgrade revenue (excluding revenue from hardware and professional services) for the calendar year 2016.

**SaaS subscription model:** Annual contract value (ACV) at year-end 2016, excluding any professional services included in annual contracts. For multiyear contracts, only the contract value for the first 12 months was used for this calculation.

**Customer adoption:** The number of active paying client organizations using the vendor's data science and machine-learning platform (excluding trials).

To progress to the next assessment gate, vendors had to have generated revenue from data science and machine-learning platform software licenses and technical support, for each platform under consideration, of:

At least \$5 million in 2016 (or the closest reporting year) in combined-revenue ACV, *or*

At least \$1 million in 2016 (or the closest reporting year) in combined-revenue ACV, *and either*

At least 150% year-over-year revenue growth for 2015 to 2016, *or*

A minimum of 200 paying end-user organizations

Only individual platforms that passed this initial revenue requirement were considered for the second inclusion gate.

## **Gate 2: Reference Customers**

Vendors that passed Gate 1 were then evaluated on the basis of the reference customers they identified. Vendors had to show significant cross-industry and cross-geographic traction for each platform under consideration. In addition, the reference customers had to be using the latest versions of the software packages being evaluated.

### *Cross-Industry Reference Customers*

Each vendor had to identify reference customers that use each of their platforms in production environments. For a platform to be considered, 25 unique reference customers were required that used the platform in a production environment, and they had to come from at least four of the following major industry segments:

- Banking, insurance and other financial services

- Education and government

- Healthcare

- Logistics and transportation

- Manufacturing and life sciences

- Mining, oil and gas, and agriculture

- Retail

- Telecommunications, communications and media

- Utilities

- Other industries

No more than 60 reference customers per platform were accepted.

### *Cross-Region Reference Customers*

Among the reference customers for each vendor, there had to be at least two customers from each of the following three areas:

- North America

- European Union

- Rest of the world

Only vendors that passed Gate 2 progressed to Gate 3.

## **Gate 3: Product Capability Scoring**

Vendors were next assessed by Gartner analysts using a scoring system to measure how well their platform(s) addressed the 15 critical capabilities defined above.

Platform capabilities were scored as follows:

- 0 = rudimentary capability or capability not supported
- 1 = capability partially supported
- 2 = capability fully supported

A product could achieve a maximum score of 30 points, given there are 15 critical capabilities.

Only products that achieved at least 20 points were considered for inclusion in this Magic Quadrant. Also, because the number of vendors that can be included is limited, only the top 16 products continued to the detailed evaluation phase.

If two or three platforms had tied, we would have included them, bringing the maximum number of vendors included up to 18. If more than three had platforms had tied, we would have used a composite metric of internet search, Gartner search and inquiry data to determine which vendors' products had the higher market traction. In no case would more than 18 vendors be included.

Over 70 vendors were considered for inclusion. Sixteen vendors (collectively offering 17 platforms) were selected for inclusion. Only SAS had more than one qualifying platform.

### **Honorable Mentions**

You may wish to consider vendors not featured in this Magic Quadrant. The following list includes notable vendors that either did not meet the inclusion criteria or whose eligibility for inclusion we were unable to verify due to a lack of information:

Amazon, which offers a largely automated platform that applies machine-learning algorithms to data stored in the AWS platform.

Big Squid, which uses automated machine learning to forecast the movement of key business metrics.

DataRobot, which designs its machine-learning platform to automate model building and presents users with a leaderboard of "best fit" models pulled from multiple sources, such as R, Python, H2O and Apache Spark.

DataScience.com, which offers a virtual-like environment that contains tools, libraries and languages to support model development and deployment.

FICO, which is a strong choice for organizations in the financial services sector and for those that depend on scorecard modeling for decision management.

Google, which, through Google Cloud, gives users access to state-of-the-art algorithms with pretrained models. There is also a service that enables users to generate their own models similar to the ones used by Google in its search and other applications, and that enables users to build proprietary models.

Megaputer, which provides a multipurpose collection of content analytics and supports various vertical solutions.

Pitney Bowes, which uses powerful data visualization and rapid model automation to deliver insights to customers.

## Evaluation Criteria

### Ability to Execute

The Ability to Execute criteria include:

**Product/service:** Core goods and services that compete in and/or serve the defined market. This criterion includes current product and service capabilities, quality, feature sets, skills and so on. This can be offered natively or through OEM agreements and partnerships, as defined in the market definition and detailed in the subcriteria.

**Overall viability (business unit, financial, strategy, organization):** This criterion includes an assessment of the organization's overall financial health, as well as the financial and practical success of the business unit. The criterion also assesses the likelihood of the organization continuing to offer and invest in the product, as well as the product's position in the current portfolio.

**Sales execution/pricing:** This criterion assesses the organization's capabilities in all presales activities and the structure that supports them. Included are deal management, pricing and negotiation, presales support and overall effectiveness of the sales channel.

**Market responsiveness and track record:** This criterion assesses the vendor's ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customers' needs evolve, and market dynamics change. This criterion also considers the vendor's history of responsiveness to changing market demands.

**Marketing execution:** This criterion assesses the clarity, quality, creativity and efficacy of programs designed to deliver the organization's message in order to influence the market, promote the brand, increase awareness of the products and establish a positive identification in the minds of customers. This mind share can be driven by a combination of publicity, promotional, thought leadership, social media, referrals and sales activities.

**Customer experience:** This criterion assesses products, services and/or programs that enable customers to achieve anticipated results with the products evaluated. Specifically, this criterion assesses the quality of supplier/buyer interactions, technical support and account support. Ancillary tools, customer support programs, availability of user groups and SLAs, among other things, may also be considered.

**Operations:** This criterion assesses the organization's ability to meet its goals and commitments. Factors include the quality of the vendor's organizational structure, skills, experiences, programs, systems and other vehicles that enable it to operate effectively and efficiently.

**Table 1.** Ability to Execute Evaluation Criteria

Evaluation Criteria

**Product or Service**

**Weighting**

High

**Overall Viability**

**Weighting**

Low

**Sales Execution/Pricing**

**Weighting**

Low

**Market Responsiveness/Record**

**Weighting**

Medium

**Marketing Execution**

**Weighting**

Medium

**Customer Experience**

**Weighting**

High

**Operations**

**Weighting**

Medium

**Table 1.** Ability to Execute Evaluation Criteria

<b>Evaluation Criteria</b>	
<b>Product or Service</b>	
Weighting	High
<b>Overall Viability</b>	
Weighting	Low
<b>Sales Execution/Pricing</b>	
Weighting	Low
<b>Market Responsiveness/Record</b>	
Weighting	Medium
<b>Marketing Execution</b>	
Weighting	Medium
<b>Customer Experience</b>	
Weighting	High
<b>Operations</b>	
Weighting	Medium

**Table 1.** Ability to Execute Evaluation Criteria

<b>Evaluation Criteria</b>	
<b>Product or Service</b>	

Weighting	High
<b>Overall Viability</b>	
Weighting	Low
<b>Sales Execution/Pricing</b>	
Weighting	Low
<b>Market Responsiveness/Record</b>	
Weighting	Medium
<b>Marketing Execution</b>	
Weighting	Medium
<b>Customer Experience</b>	
Weighting	High
<b>Operations</b>	
Weighting	Medium

Source: Gartner (February 2018)

## Completeness of Vision

The Completeness of Vision criteria include:

**Market understanding:** This criterion assesses the vendor's ability to understand customers' needs and to translate them into products and services. Vendors with a clear vision that listen to customers' demands and understand them can shape or enhance market changes.

**Marketing strategy:** This criterion looks for clear, differentiated messaging, consistently communicated internally, and externalized through social media, advertising, customer programs and positioning statements.

**Sales strategy:** This criterion looks for a sound strategy for selling that uses appropriate networks, including direct and indirect sales, marketing, service and communication networks. It also considers partners that extend the scope and depth of the vendor's market reach,

expertise, technologies, services and customer base.

**Offering (product) strategy:** This criterion looks for an approach to product development and delivery that emphasizes market differentiation, functionality, methodology and features as they map to current and future requirements.

**Innovation:** This criterion looks for direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or pre-emptive purposes.

**Table 2.** Completeness of Vision Evaluation Criteria

<b>Evaluation Criteria</b>	
<b>Market Understanding</b>	
<b>Weighting</b>	
	Medium
<b>Marketing Strategy</b>	
<b>Weighting</b>	
	Low
<b>Sales Strategy</b>	
<b>Weighting</b>	
	Low
<b>Offering (Product) Strategy</b>	
<b>Weighting</b>	
	High
<b>Business Model</b>	
<b>Weighting</b>	
	Not Rated

<b>Vertical/Industry Strategy</b>	
<b>Weighting</b>	
	Not Rated
<b>Innovation</b>	
<b>Weighting</b>	
	High
<b>Geographic Strategy</b>	
<b>Weighting</b>	
	Not Rated

**Table 2.** Completeness of Vision Evaluation Criteria

<b>Evaluation Criteria</b>	
<b>Market Understanding</b>	
Weighting	Medium
<b>Marketing Strategy</b>	
Weighting	Low
<b>Sales Strategy</b>	
Weighting	Low
<b>Offering (Product) Strategy</b>	
Weighting	High
<b>Business Model</b>	

Weighting	Not Rated
<b>Vertical/Industry Strategy</b>	
Weighting	Not Rated
<b>Innovation</b>	
Weighting	High
<b>Geographic Strategy</b>	
Weighting	Not Rated

**Table 2.** Completeness of Vision Evaluation Criteria

<b>Evaluation Criteria</b>	
<b>Market Understanding</b>	
Weighting	Medium
<b>Marketing Strategy</b>	
Weighting	Low
<b>Sales Strategy</b>	
Weighting	Low
<b>Offering (Product) Strategy</b>	
Weighting	High
<b>Business Model</b>	
Weighting	Not Rated

## Vertical/Industry Strategy

Weighting

Not Rated

## Innovation

Weighting

High

## Geographic Strategy

Weighting

Not Rated

Source: Gartner (February 2018)

## Quadrant Descriptions

### Leaders

Leaders have a strong presence and significant mind share in the data science and machine-learning market. They demonstrate strength in depth and breadth across a full exploration, model development and implementation process. While providing outstanding service and support, Leaders are also nimble in responding to rapidly changing market conditions. The number of data scientist professionals skilled in the use of Leaders' platforms is significant and growing.

Leaders are in the strongest position to influence the market's growth and direction. They address all industries, geographies, data domains and use cases and, thus, have a solid understanding of, and strategy for, this market. Not only are they able to focus on executing effectively, based on current market conditions, but they also have solid and robust roadmaps to take advantage of new developments and advancing technologies in this rapidly transforming sector. They provide thought leadership and innovative differentiation, often disrupting the market in the process.

Leaders are suitable vendors for most organizations to evaluate. They should not be the only vendors evaluated, but at least two are likely to be included in a typical shortlist of five to eight vendors. They provide a benchmark of high standards to which others should be compared.

### Challengers

Challengers have an established presence, credibility, viability and robust product capabilities. They may not, however, demonstrate thought leadership and innovation to the same degree as Leaders.

There are two main types of Challenger:

Long-established data science and machine-learning vendors that succeed because of their stability, predictability and long-term customer relationships. They need to revitalize their vision to stay abreast of market developments and become more broadly influential and innovative. If they simply continue doing what they have been doing, their growth and market presence may be impaired.

Vendors well-established in adjacent markets that are entering the data science and machine-learning market with solutions that extend their current platforms for existing customers but are also a reasonable option for many potential new customers. As these vendors prove they can influence this market and provide clear direction and vision, they may develop into Leaders. They must avoid the temptation to introduce new capabilities quickly but superficially.

Challengers are well-placed to succeed in this market as it is currently defined and are operating effectively within current market conditions. Their vision and roadmap, however, may be impaired by a lack of market understanding, excessive focus on short-term gains, and strategy- and product-related inertia and lack of innovation. Equally, their marketing efforts, geographic presence and visibility may be deficient, relative to Leaders.

## Visionaries

Visionaries are typically smaller vendors or newer entrants representative of trends that are shaping, or have the potential to shape, the market. There may, however, be concerns about these vendors' ability to keep executing effectively and to scale as they grow. They are typically not well-known in the market, and therefore often have low momentum, relative to Challengers and Leaders.

Visionaries have a strong vision and supporting roadmap. They are innovative in their approach to addressing the needs of the market. Although their offerings are typically innovative and solid in the capabilities they do provide, there are often gaps in the completeness and breadth of their offerings.

Visionaries are worth considering because they may:

- Represent an opportunity to jump-start an innovative initiative

- Provide some compelling, differentiating capability that offers a competitive advantage as either a complement to, or a substitute for, existing solutions

- Be more easily influenced with regard to their product roadmap and approach

Visionaries, however, also pose increased risk. In today's highly competitive data science and machine-learning market, they could be targets for acquisition. They may also struggle to gain momentum, develop a presence, increase their market share or fulfill their vision.

As Visionaries mature and prove their Ability to Execute, they may eventually become Leaders.

## Niche Players

Niche Players demonstrate strength in a particular industry or approach, or pair well with a specific technology stack.

Some Niche Players demonstrate a degree of vision, which suggests they could become Visionaries. They are, however, often struggling to make their vision compelling, relative to others in the market. They may also be struggling to develop a track record of innovation and thought leadership that could give them the momentum to become Visionaries.

Other Niche Players could become Challengers if they continue to execute in a way that increases their momentum and traction in the market.

## Context

Traditional data science and machine-learning vendors are being challenged by new entrants and smaller, nimbler competitors that have adapted more easily and responded more readily as the market has evolved. Other vendors are also staking claims. Google, for example, has a data science and machine-learning platform in development. Amazon released SageMaker in November 2017, which provides a fully managed service to build, train, and deploy machine-learning models in a secure and scalable hosted environment. Business software companies are also entering the market – for example, Salesforce with Einstein and Workday with its acquisition of Platfora. In addition, vendors that have traditionally focused on traditional descriptive and diagnostic analytics are extending their capabilities, through technological advances or acquisitions, to provide a range of analytic capability that includes not only descriptive and diagnostic analytics but also predictive and prescriptive analytics.

Acquisitions that have enabled vendors to enhance their vision and broaden their capabilities include DataRobot's acquisition of Nutonian, Progress' acquisition of DataRPM, and TIBCO Software's acquisition of Statistica (from Quest Software) and Alpine Data.

Many organizations are awaiting the arrival of new platforms that will enable them to extend their current infrastructure and work well with their other technology. Users must be aware of the rapid changes in this market and monitor how vendors and other organizations are responding to those changes. They should regularly assess the state of the market and the ability of their current vendor to respond and adapt. In addition, they should consider extending their analytic capability to include descriptive, diagnostic, predictive and prescriptive capabilities in a cohesive manner. They must familiarize themselves with, and assess, new entrants and disruptive vendors to understand these competitors' value propositions for data science and machine-learning technology. When evaluating new vendors and platforms, they must assess whether the platforms compete with or complement any analytic platforms they have already deployed. In addition, it is often good practice to consider getting help from an external service provider.

As data science and machine learning becomes more pervasive, it becomes increasingly important to have an enterprise capability to manage both data and analytics. The ability not only to manage but also to operationalize and manage data science and machine learning models throughout their life cycle and the ability to collaborate and share throughout the analytic life cycle also become increasingly important. This is an area that has not received much focus so far, but it is becoming critical to organizations' ability to maximize their ROI.

## Market Overview

The application of data science and machine-learning technology is a priority for many organizations. Increasingly sophisticated analytic procedures are available that enable them to exploit the massive amounts of data available to them.

Data science and machine-learning platforms are increasingly available for a broad spectrum of users. These range from operational workers who make day-to-day decisions based on sophisticated models working behind the scenes, to citizen data scientists who need data science and machine-learning capabilities but have minimal skills in advanced data science, to highly skilled engineers and data scientists who design experiments and deploy models to represent and optimize business decisions.

Revenue from data science and machine-learning platforms grew by 9.3% in 2016, to \$2.4 billion (in constant currency, 11.1% and \$2.7 billion, respectively). This growth is more than double that of the overall analytics and BI market (4.5%), and the \$2.4 billion represents 14.1% of the total worldwide analytics and BI revenue (up from 13.5% in 2015). The growth in the data science and machine-learning platform market is being driven by end-user organizations' desire to use more advanced analytics to improve decision-making across the business. For more details, see "Market Share Analysis: Analytics and BI Software, 2016."

The data science and machine-learning market continues to expand and remains in a state of flux. Key drivers of changes in this market include:

Analytics markets expanding across the analytics continuum. Analytics and BI platforms are gaining more advanced predictive and prescriptive capabilities. Data science and machine-learning platforms are gaining stronger data visualization and other descriptive and diagnostic capabilities. Gartner estimates that, by 2020, predictive and prescriptive analytics will attract 40% of enterprises' net-new investment in analytics and BI technologies.

An increased number of data sources from multiple locations requiring increased scalability, flexibility and a hybrid approach to combining data on-premises and in the cloud.

Enhanced AI technologies, based on deep learning, which are becoming more affordable, available and accessible through cloud services, APIs, new tools, and integration with existing software products and services.

A wider range of users, beyond traditional expert data scientists. New users increasingly include citizen data scientists and application developers.

Augmented analytics – techniques that incorporate machine learning to provide a "smart," guided approach to data preparation, business analytics and data science capabilities (such as feature selection). They extend the use of advanced capabilities to nontraditional roles.

A focus on operationalization, including the ability to manage models and collaborate at an enterprise level. This also includes the ability to orchestrate the complete analytical process, with a focus, ultimately, on productionizing and operationalizing models, increasing productivity and boosting business impact.

The availability of free and open-source options that are often the first to offer innovative analytic capabilities. They also represent easy-to-access, low-initial-investment options for starting data science and machine-learning projects.

The increasing demand for "servware." Analytic services and software are converging to form new solutions – "servware" – that disrupt vendors' established practices and create opportunities for organizations to differentiate themselves (see "Take Advantage of the Disruptive Convergence of Analytic Services and Software" ).

The driving forces behind, and the thought leaders within, the data science and machine-learning platform market today are no longer the traditional vendors who have consistently been Leaders in this series of Magic Quadrants. The definition of what constitutes a modern data science and machine-learning platform is changing, which results in the market's current state of flux. The changes are likely to become more revolutionary than evolutionary. Today's modern data science and machine-learning platform emphasizes many new capabilities, including:

Open-source support, which enables vendors to exploit the innovative solutions often available first as open-source options and to respond at a faster pace than is attainable through internal development.

Scalability, to cope with changing data demands, including the need to access all types of data, accommodate changing data volumes, and address data hosted both on-premises and in the cloud via a hybrid approach.

Support for and management of a large number of models across the analytic pipeline. Operationalization of models is key to turning analytic insights into action and, ultimately, impact. In addition, models must be reassessed, retuned and managed over time, which included managing collaboration and the sharing of models as analytic assets.

AI frameworks, which offer deep-learning capabilities that are easily accessible and usable.

Support for multiple roles, beyond the traditional expert data scientist, often via augmented analytics. New roles that increasingly need access to data science and machine-learning platforms include citizen data scientists and application developers.

Activity in the data science and machine-learning market is centering on a number of platforms that provide similar critical capabilities but are differentiated from other vendor offerings in some specific way. This differentiation could take the form of a focus on a specific audience or skill set, a specific part of the analytical pipeline (such as model development, as opposed to model deployment), or specific capabilities (such as cognitive capabilities), or the provision a framework for exploiting open-source offerings. As such, there is a danger of comparing "apples and oranges" or, worse, "apples, oranges and tomatoes" when assessing the relative merits of different offerings.

The challenges to using a data science and machine-learning platform effectively are not limited to choosing the right platform for an organization's analytical needs. Data, people and process must also be addressed. Data science and machine-learning approaches require

increasingly accurate data to build models representative of the real world. Information management therefore has an important role to play by helping to ensure that models are based on sound inputs and practices.

Providing data science and machine-learning capability to nontraditional users, such as citizen data scientists and application developers, has implications for ease of use and training. In addition, building the models themselves is not the end of the analytic process. To have impact, models must be deployed and managed over time. This requires enterprise-level management to enable the best use of models at scale and in a collaborative and consistent way. Organizations will continue to search for balance between convenience (ease of use) on the one hand and control on the other.

Gartner expects continued turbulence in the data science and machine-learning platform market over the next two to three years. Entirely new vendors, as well as existing vendors from adjacent markets (such as analytics and BI), will continue to enter it. For example, a recent noteworthy acquisition was Datawatch's purchase of Angoss in January 2018. Traditional vendors will strive to become more agile and responsive. We expect further acquisitions and extensions by vendors to gain descriptive, diagnostic, predictive and prescriptive analytic capabilities. Applied analytic solutions, designed to solve specific industry problems, will continue to represent alternative options to starting advanced analytics initiatives from scratch.

*Additional research contribution and review by Nigel Shen*

## Evidence

Gartner's assessments and commentary in this report are based on the following sources:

Evaluation of instruction manuals and documentation of selected vendors to verify platform functionality.

An online survey of vendors' reference customers conducted from July through August 2017. This survey elicited 458 responses evaluating the reference customers' experience with vendors' platforms. The list of survey participants derived from information supplied by the vendors.

A questionnaire completed by the vendors.

Vendors' briefings, including product demonstrations, about their strategy and operations.

An extensive RFP inquiring about how each vendor delivers the specific features that make up our 15 critical capabilities (see "Toolkit: RFP for Data Science and Machine-Learning Platforms" ).

A prepared video demonstration of how well vendors' data science and machine-learning platforms address specific functionality requirements across the 15 critical capabilities.

Interactions between Gartner analysts and Gartner clients making decisions about their evaluation criteria, and Gartner clients' opinions about how successfully vendors meet these criteria.

# Note 1

## Definition of Open-Source Platform

The open-source approach is becoming more common throughout the data science market. It enables people to innovate collaboratively, each contributing their own perspectives in a way that accelerates time to market. The most common examples in the data science and machine-learning market are open-source components. The open-source approach is quickly becoming a more mainstream way to introduce new capabilities. Many of these capabilities are evaluated in this Magic Quadrant.

Open-source components include open-source data, as introduced by vendors such as Databricks and Microsoft; open-source programming languages, such as R and Python; open-source algorithm libraries such as those found in DL4J and H2O; open-source visualizations such as D3 and Plotly; open-source notebooks like Jupyter and Zeppelin; open-source data management platforms, such as Spark and Hadoop; and open-source frameworks like SparkML and TensorFlow.

A platform is considered open (but not open-source) if it offers flexibility and extensibility for accessing open-source components. In addition, a platform itself can be open-source, which means that the source code is made available for use or modification. Open-source software is usually developed as a public collaboration and made freely available. Only open-source platforms that also have commercially licensable products can be included in this Magic Quadrant.

## Evaluation Criteria Definitions

### **Ability to Execute**

**Product/Service:** Core goods and services offered by the vendor for the defined market. This includes current product/service capabilities, quality, feature sets, skills and so on, whether offered natively or through OEM agreements/partnerships as defined in the market definition and detailed in the subcriteria.

**Overall Viability:** Viability includes an assessment of the overall organization's financial health, the financial and practical success of the business unit, and the likelihood that the individual business unit will continue investing in the product, will continue offering the product and will advance the state of the art within the organization's portfolio of products.

**Sales Execution/Pricing:** The vendor's capabilities in all presales activities and the structure that supports them. This includes deal management, pricing and negotiation, presales support, and the overall effectiveness of the sales channel.

**Market Responsiveness/Record:** Ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customer needs evolve and market dynamics change. This criterion also considers the vendor's history of responsiveness.

**Marketing Execution:** The clarity, quality, creativity and efficacy of programs designed to deliver the organization's message to influence the market, promote the brand and business, increase awareness of the products, and establish a positive identification with the product/brand and organization in the minds of buyers. This "mind share" can be driven by a combination of publicity, promotional initiatives, thought leadership, word of mouth and sales activities.

**Customer Experience:** Relationships, products and services/programs that enable clients to be successful with the products evaluated. Specifically, this includes the ways customers receive technical support or account support. This can also include ancillary tools, customer support programs (and the quality thereof), availability of user groups, service-level agreements and so on.

**Operations:** The ability of the organization to meet its goals and commitments. Factors include the quality of the organizational structure, including skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently on an ongoing basis.

### **Completeness of Vision**

**Market Understanding:** Ability of the vendor to understand buyers' wants and needs and to translate those into products and services. Vendors that show the highest degree of vision listen to and understand buyers' wants and needs, and can shape or enhance those with their added vision.

**Marketing Strategy:** A clear, differentiated set of messages consistently communicated throughout the organization and externalized through the website, advertising, customer programs and positioning statements.

**Sales Strategy:** The strategy for selling products that uses the appropriate network of direct and indirect sales, marketing, service, and communication affiliates that extend the scope and depth of market reach, skills, expertise, technologies, services and the customer base.

**Offering (Product) Strategy:** The vendor's approach to product development and delivery that emphasizes differentiation, functionality, methodology and feature sets as they map to current and future requirements.

**Business Model:** The soundness and logic of the vendor's underlying business proposition.

**Vertical/Industry Strategy:** The vendor's strategy to direct resources, skills and offerings to meet the specific needs of individual market segments, including vertical markets.

**Innovation:** Direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or pre-emptive purposes.

**Geographic Strategy:** The vendor's strategy to direct resources, skills and offerings to meet the specific needs of geographies outside the "home" or native geography, either directly or through partners, channels and subsidiaries as appropriate for that geography and market.

© 2018 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. or its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. If you are authorized to access this publication, your use of it is subject to the Usage Guidelines for Gartner Services ([/technology/about/policies/usage\\_guidelines.jsp](/technology/about/policies/usage_guidelines.jsp)) posted on gartner.com. The information contained in this publication has been obtained from sources believed to be reliable. Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information and shall have no liability for errors, omissions or inadequacies in such information. This publication consists of the opinions of Gartner's research organization and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice. Gartner provides information technology research and advisory services to a wide range of technology consumers, manufacturers and sellers, and may have client relationships with, and derive revenues from, companies discussed herein. Although Gartner research may include a discussion of related legal issues, Gartner does not provide legal advice or services and its research should not be construed or used as such. Gartner is a public company, and its shareholders may include firms and funds that have financial interests in entities covered in Gartner research. Gartner's Board of Directors may include senior managers of these firms or funds. Gartner research is produced independently by its research organization without input or influence from these firms, funds or their managers. For further information on the independence and integrity of Gartner research, see "Guiding Principles on Independence and Objectivity." ([/technology/about/ombudsman/omb\\_guide2.jsp](/technology/about/ombudsman/omb_guide2.jsp))"

---

About (<http://www.gartner.com/technology/about.jsp>)

Careers (<http://www.gartner.com/technology/careers/>)

Newsroom (<http://www.gartner.com/newsroom/>)

Policies ([http://www.gartner.com/technology/about/policies/guidelines\\_ov.jsp](http://www.gartner.com/technology/about/policies/guidelines_ov.jsp))

Privacy (<https://www.gartner.com/privacy>)

Site Index (<http://www.gartner.com/technology/site-index.jsp>)

IT Glossary (<http://www.gartner.com/it-glossary/>)

Contact Gartner ([http://www.gartner.com/technology/contact/contact\\_gartner.jsp](http://www.gartner.com/technology/contact/contact_gartner.jsp))