Augmented Analytics Is the Future of Analytics

Published 30 October 2019 - ID G00444837 - 50 min read

By Analysts Rita Sallam, Carlie Idoine

Augmented analytics leverages machine learning and AI techniques to transform how analytics content is developed, consumed and shared. Data and analytics leaders should plan to adopt augmented analytics as platform capabilities mature.

Overview

Key Findings

- Augmented analytics in ABI platforms (augmented ABI) transforms and democratizes how business people explore, analyze and act on insights by surfacing key insights in data through ML and AI assisted data preparation, insight generation and insight explanation.
- By automating many aspects of DSML development, management and deployment augmented analytics in DSML platforms (augmented DSML), expert data scientists become more productive. This also extends DSML model building to a broader range of less skilled users including new citizen data science roles (business analysts, developers and others).
- Explainability is increasingly becoming an important capability to give users confidence in using machine-generated insights and recommendations.
- Natural language interfaces such as search via text and voice combined with augmented analytics make analytics more accessible and consumable by a broader set of users.

Recommendations

As a data and analytics leader planning to use augmented analytics to modernize solutions, you should:

- Find opportunities to complement existing data and analytics initiatives by piloting augmented analytics for high value, but time-consuming, manual analysis.
- Build trust in augmented DSML models by fostering collaboration between expert and citizen data scientists to back test and prove value. Understand the limitations of machine-assisted models, which work best with proven algorithms versus cutting-edge techniques.
- Monitor the augmented analytics roadmaps of established data and analytics providers, enterprise application vendors and startups. Assess upfront setup, data preparation, openness and explainability of insights and models, range of algorithms, and model accuracy.
- Plan for new roles and expand data literacy skills to support wider adoption due to augmented analytics by people who don't currently make decisions based on insights from analytics and BI platforms or from data science and ML models.

Strategic Planning Assumptions

By 2021, augmented analytics will be a dominant driver of new purchases of analytics and BI as well as data science and machine learning platforms, and of embedded analytics.

By 2021, conversational analytics and natural language processing (NLP) will boost analytics and BI adoption from 32% of employees to more than 50% of an organization's employees, to include new classes of users — particularly in front offices.

By 2021, automation of data science tasks will enable citizen data scientists to produce a higher volume of advanced analysis, than specialized data scientists.

By 2025, a scarcity of data scientists will no longer hinder the adoption of data science and machine learning in organizations.

Analysis

Analytics has become a strategic component of how value is created in most businesses. However, it is at a critical inflection point. As data complexity increases, business people across the enterprise are awash in data, struggling to identify what is most important and what best actions to take. Larger and more varied dataset combinations also mean more variables and relationships to analyze, explore and test.

Across the analytics stack, tools have become easier to use and more agile, enabling greater access and self-service. However, many processes remain largely manual and prone to bias. These include managing data, preparing data for analysis, analyzing data, building data science and ML/AI models, interpreting results, telling stories with data, and making insights actionable. Using current approaches, it is not possible for users to explore every possible combination and pattern, let alone determine whether their findings are the most relevant, significant and actionable.

Relying on business users to find patterns and data scientists to build and manage models manually may result in them exploring their own biased hypotheses, missing key findings and drawing their own incorrect or incomplete conclusions. This will adversely affect decisions, actions and outcomes.

To address this growing challenge, augmented analytics is rapidly gaining traction. Central to this development is the use of ML automation/AI techniques to augment human intelligence and contextual awareness, and to transform data management, analytics and BI as well as many aspects of data science and ML/AI model development and consumption. As businesses become inundated with data, augmented analytics becomes crucial for presenting to users across the business only what is important for them in their context to act upon at that moment.

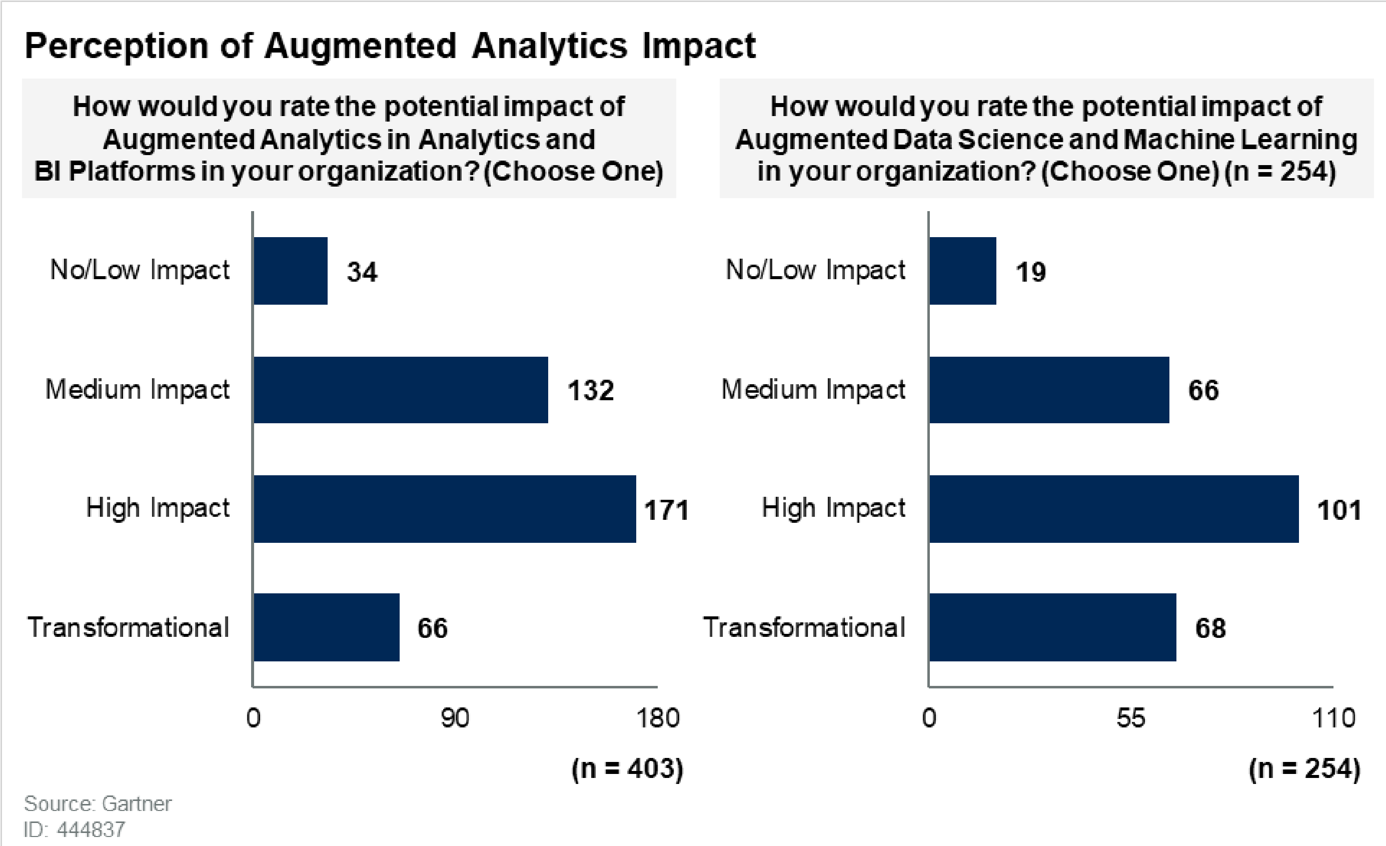
Having to manually find patterns in the data is like looking for a needle in the haystack. Augmented analytics helps find the needle faster by acting like a giant magnet hovering over the hay.

As more organizations digitally transform, they want to expand the use of data science and ML/AI. They want to leverage it to create new differentiated analytic applications, and embed ML/AI into existing applications. However, the scarcity of expert data science skills has become a significant barrier. By automating many time-consuming and bias-prone tasks, augmented analytics expands the capabilities of those with more widely available skill sets — the business analyst and the application developer (a new breed of citizen data scientists). It allows them — in addition to expert data scientists — to generate insights and create augmented analytics-assisted models to embed in applications. Because augmented analytics automates feature engineering, model selection and increasingly model management and model operationalization (MLops) tasks, augmented analytics also makes expert data scientists more productive.

Due to the perception of high potential impact, augmented analytics capabilities will advance rapidly to mainstream adoption, as a key feature of modern analytics and BI as well as data science and ML platforms. More importantly, automated insights from augmented analytics will also be embedded in enterprise applications and NLP-enabled applications and conversational chatbots for analytics to make insights available to more people across the organization.

More than 60% of respondents to a Gartner Data and Analytics Summit poll said they believe augmented analytics will have a high or transformational impact on their ability to scale the value of analytics in their organization (see Figure 1). [2]

Figure 1. Perception of Augmented Analytics Impact



Innovative startups and large vendors now offering augmented analytics capabilities are transforming how people generate and consume analytics insights. They are disrupting data and analytics markets, spanning analytics and BI and data science and ML platforms and analytics embedded in enterprise applications.

Given the business impact and market potential, remaining vendors will quickly follow and adoption will accelerate quickly as the technology matures.

Purpose: This document explores augmented analytics and its ramifications for organizational and market disruption. It provides guidance to data and analytics leaders planning to adopt these capabilities in order to modernize and to drive digital transformation and innovation (see Figure 2).

Figure 2. The Path to Augmented Analytics Adoption

The Path to Augmented Analytics Adoption

Source: Gartner
ID: 444837

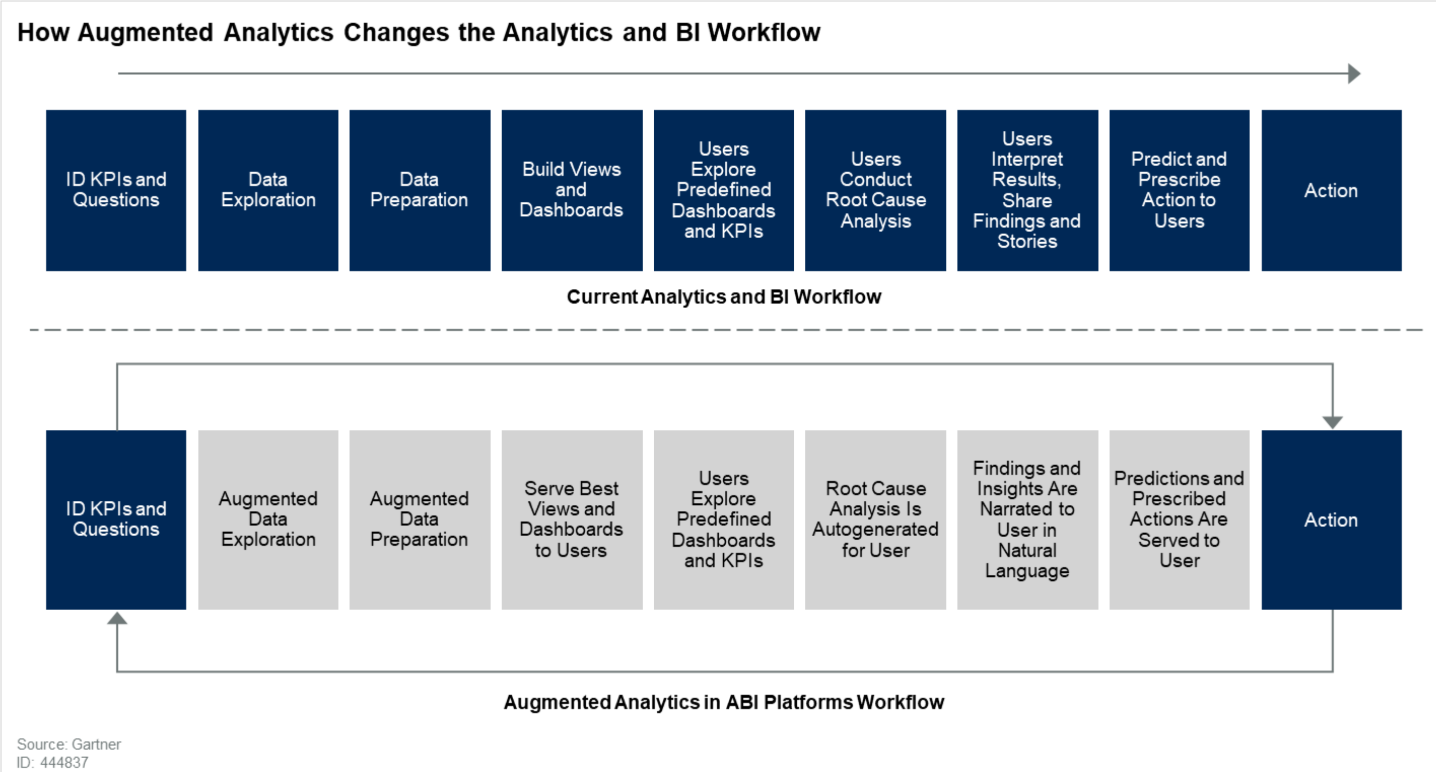
Definition

Augmented analytics includes:

Augmented data preparation — Uses ML/AI automation to augment and accelerate data preparation activities. These include data profiling and data quality, harmonization, data modeling, manipulation, enrichment/inference, metadata development, and data cataloging in analytics, BI and data science and ML platforms, and in stand-alone data preparation tools. The same dynamics affecting analytics and BI and data science and machine learning are also transforming data management. AI and ML techniques are automating many aspects of data preparation/integration, data quality, master data management, metadata management, data cataloging, and database management. (See "Top 10 Data and Analytics Technology Trends That Will Change Your Business," "Rebalance Your Integration Effort With a Mix of Human and Artificial Intelligence," and "Market Guide for Data Preparation.")

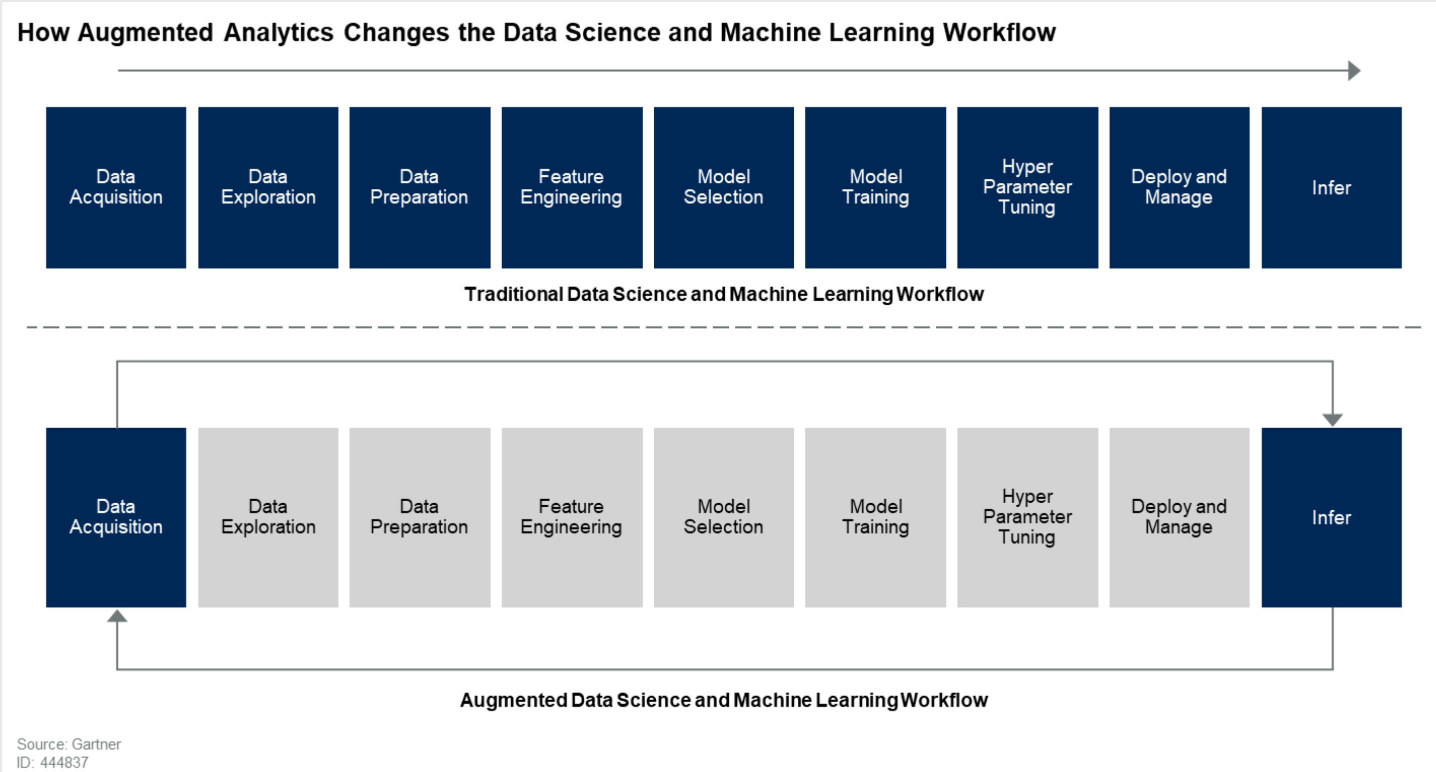
Augmented analytics as part of analytics and BI platforms (Augmented ABI) — Enables machine learning and AI assisted data preparation, insight generation, and insight explanation to augment how business people and analysts explore and analyze data in analytics and BI platforms. With augmented analytics, business people and citizen data scientists automatically find, visualize and narrate relevant findings (such as correlations, exceptions, clusters, drivers and predictions), without having to build models or write algorithms. It can also be used by business analysts and citizen data scientists to analyze data without preconceived notions of relationships among variables in data, for early prototyping and hypothesis development with less manual experimentation. Augmented analytics will transform how all people interact with analytics content. Instead of accessing analytics content on static dashboards, users will go to dynamic, autogenerated dashboards or sets of stories that will be delivered to them, showing the most important insights for them in their context. It will also automatically generate drivers and explanations changes with prescriptive recommendations. See Figure 3.

Figure 3. How Augmented Analytics Changes the Analytics and BI Workflow



Augmented data science and machine learning (Augmented DSML) — Uses ML/AI to automate key aspects of data science and ML/AI modeling, such as feature engineering and model selection, as well as model operationalization, model explanation and, ultimately, model tuning and management. (See: "How Augmented Machine Learning is Democratizing Data Science.") This reduces the requirement for specialized skills to generate, operationalize and manage an advanced analytics model. It opens up data science and ML content creation to citizen data scientists and (application) developers who must embed ML/AI into applications. Operationalizing models into applications is critical to deriving value from them. Highly-skilled data scientists can also be more productive and have more time to focus on building and operationalizing the most relevant models. See Figure 4.

Figure 4. How Augmented Analytics Changes the Data Science and Machine Learning Workflow



Many autogenerated and human-augmented ML models created through augmented analytics are being embedded in customer facing and enterprise applications; for example, those of the HR, finance, sales, marketing, customer service, procurement and asset management departments. This helps to optimize the decisions and actions of all employees as well as customers and partners, not just those of analysts and data scientists.

Augmented analytics is becoming a key feature of conversational chatbots for analytics. This is an emerging approach that enables business people to generate queries, explore data and receive and act on insights in natural language (voice or text) via mobile devices, personal assistants and embedded enterprise application chatbots. For example, instead of accessing a daily dashboard, a decision maker with access to Amazon Alexa or a conversational chatbot for analytics integrated with Slack might say:

> *"Alexa, analyze my sales results for the northeast region for the past three months this year versus last!" or*
>
> *"What are the top three things I can do to improve my close rate today?"*

NLP-driven conversational analytics (or conversational chatbots for analytics) applications are emerging with early integrations among vendors. Analytics and enterprise application vendors are using APIs, acquiring capabilities, and building integrations (in some cases with the help of partners) to make these applications easier to deploy. We expect out-of-the-box and enterprise-ready instances to appear over the next two to five years (see "Hype Cycle for Analytics and Business Intelligence, 2019" and "Hype Cycle for Data Science and Machine Learning, 2019").

Augmented Analytics in ABI Platforms: The Next Wave of Analytics Market Disruption

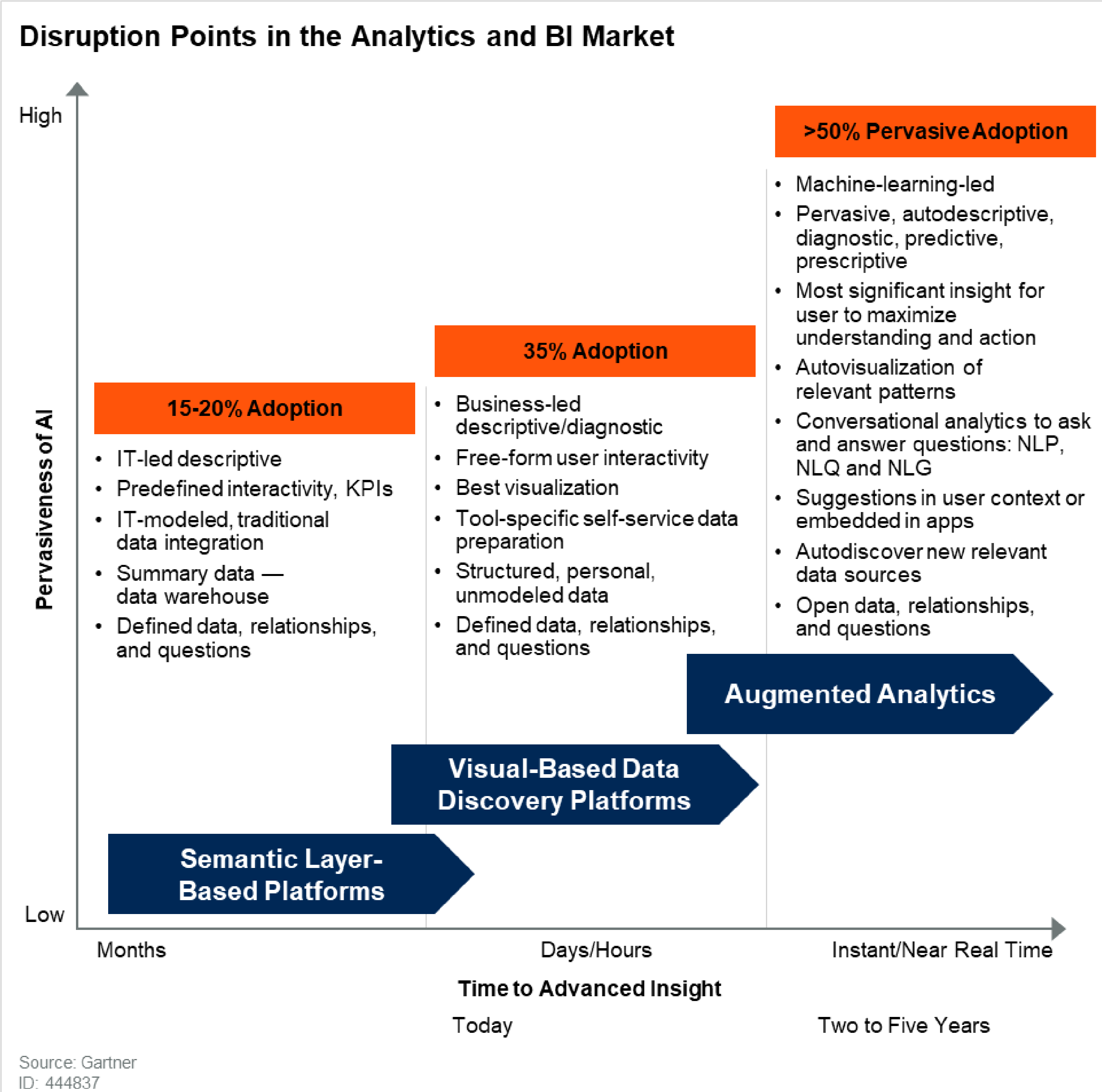
Over the past 10 years, visual-based data discovery capabilities have become the defining feature of modern analytics and BI platforms that disrupted the traditional enterprise reporting-centric BI market. These easy-to-use tools enable users to assemble data rapidly, explore hypotheses visually and find new insights in data. They have transformed how business users explore data, in comparison with the IT-centric, semantic-layer-based approach of traditional BI platforms.

Visual-based data discovery features — now standard capabilities of modern analytics and BI platforms — are easy to use, because users analyze data by creating visual queries to investigate hypotheses. But when data sizes and the number of variables are large, it becomes impossible for users to explore every possible pattern and combination, let alone determine whether their findings are the most relevant, Moreover, “a picture is worth a thousand words” has long been assumed in the field of data and analytics, and rightly so. Visualizations are a powerful and consumable way to find and communicate patterns in data (more so than tables or lists). However, they only highlight visual relationships; they do not identify statistically significant findings. For example, a user can see visual differences between bars on a bar chart, but unless the size differences are stark, it would require user interpretation or further statistical analysis to determine whether the differences are relevant, significant and actionable. Moreover, finding insights from advanced analytics — a key aspirational goal for most companies, particularly those facing rapid digital and industry change — requires expert data science skills, which are extremely scarce.

> **Visual-based data discovery with manual interactive exploration using visualizations has been the defining feature of modern analytics and BI platforms. But now augmented analytics characterized by ML/AI automation of the insight discovery, exploration, explanation, prediction and prescription process is a defining feature of new-generation analytics and BI platforms (see Figure 5).**

Augmented analytics can reduce time-consuming exploration and the identification of false or less-relevant insights. Applying a range of algorithms and ensemble learning to data in parallel, and explaining actionable findings to users, reduces the risk of missing important insights in the data in comparison to manual exploration. It also optimizes resulting decisions and actions. This shift requires investment in data literacy throughout organizations, as insights are distributed to all employees (see “Fostering Data Literacy and Information as a Second Language: A Gartner Trend Insight Report”).

Figure 5. Disruption Points in the Analytics and BI Market



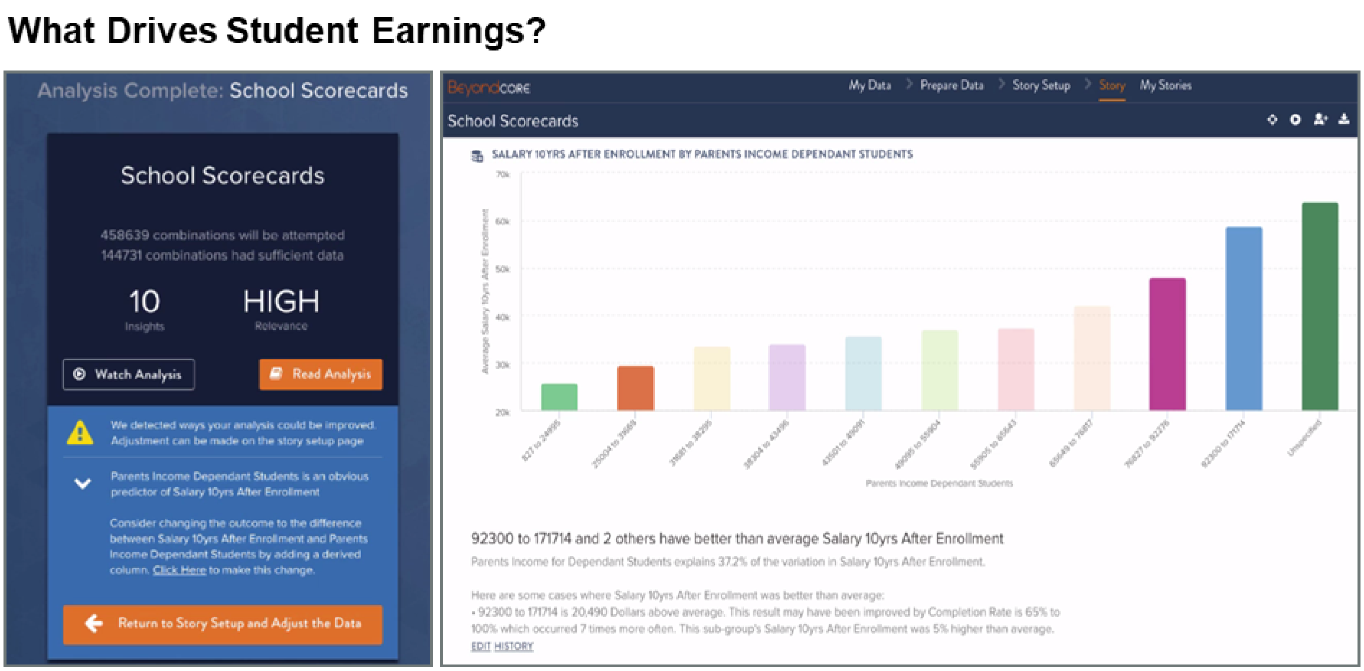
Example 1. *How Salesforce Einstein Analytics showed that attendance at a top university is not the main predictor of high earning power:*

- At Gartner's 2016 "BI Bake-Off" at the Data and Analytics Summit in Dallas, Texas, we gave representatives of several modern analytics and BI platform vendors, university and college student demographic data, payroll data and a demo script. In addition to showcasing functional differences across Critical Capabilities, we asked them to combine the datasets and derive insights about which university graduates would have the most earning power 10 years after graduation.
- Given the number of variables and combinations available to explore manually, the representatives did what expert analysts typically do — explored their own hypotheses first. In this case, it was the "usual suspects" of leading universities — "*because going to Harvard means you out-earn those going to state universities, right?*"
- While there was a relationship in the data between attendance at top universities and earning power, all missed the most important driver — one that is not intuitive. The biggest indicator of students' future earning power in the data was not their university. It was their parents'

means it takes five or six years? Or is it that socioeconomic advantages translate to great job opportunities and accelerated career progression? We can, however, say that parental income was not a driver that the participants in the "BI Bake-Off" knew to look for.

- By contrast, although we gave all the vendors in the vendor exhibit hall the same dataset, only Salesforce Einstein Discovery uncovered the main driver. After just a few seconds of ingesting the data, it automatically analyzed more than 458,000 combinations in the data and generated the top 10 insights, with a narrative about the results (see Figure 6).

Figure 6. What Drives Student Earnings?



Source: Salesforce
ID: 444837

Example 2. *How DataRobot found a link between diversity of student population and earnings power in the same dataset.*

- At Gartner's 2019 "Data Science and Machine Learning Bake-Off," DSML platform vendors were given the same dataset as in Example 1. In this case, they were asked to find insights about which university graduates would have the most earning power 10 years after graduation. DataRobot found strong relationship between the diversity of the student population of the university attended and the student's earning power 10 years out. Similarly, no other vendor found this non-obvious insight.

Example 3. *How Salesforce Einstein Analytics uncovered new drivers of traffic fatalities*

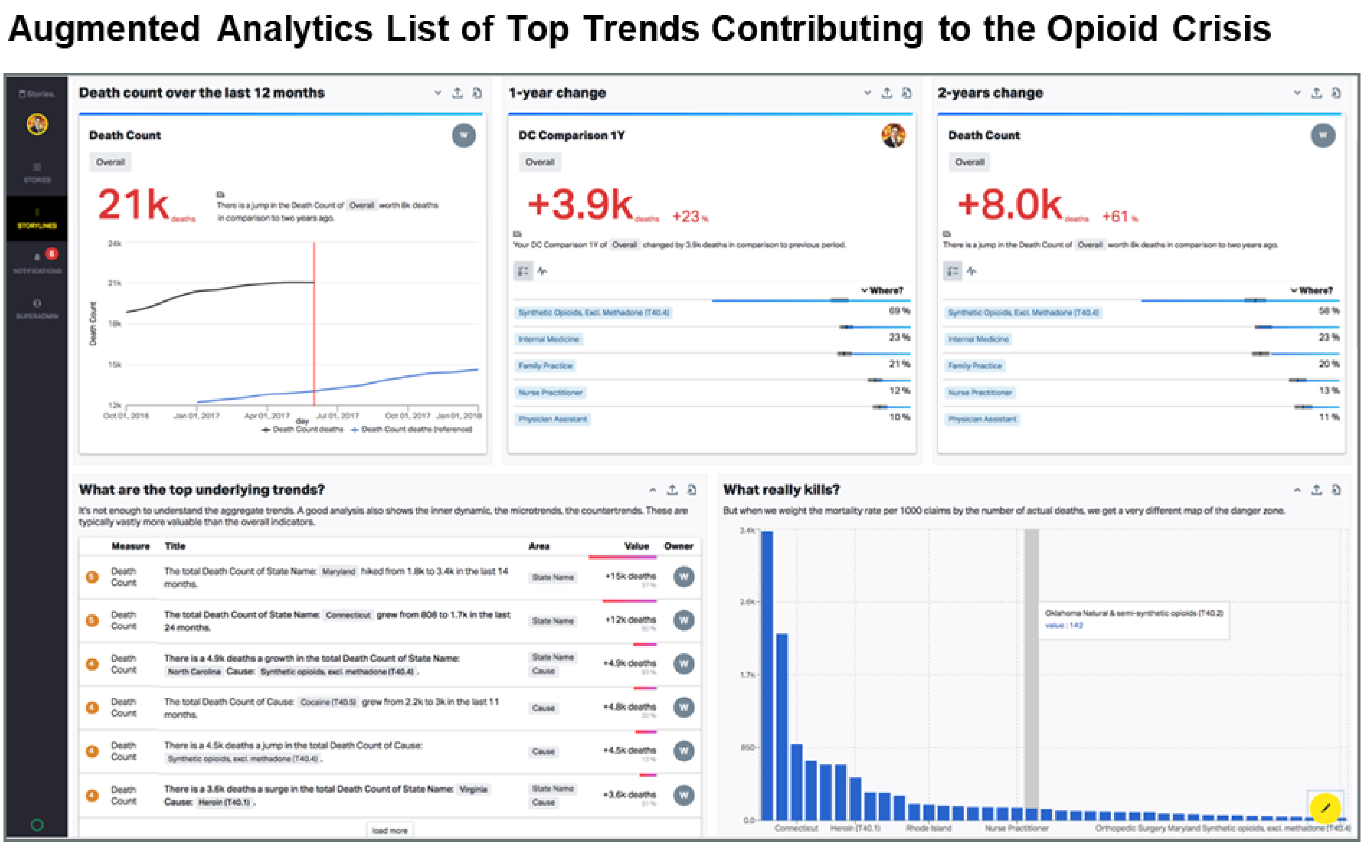
- At Gartner's 2017 "BI Bake-Off," analytics and BI vendors were asked to find unique patterns in traffic fatality data. Insights around age, gender, where someone sits in a car, drink- and drug-related driving and distracted driving were all previously known factors linked to traffic fatalities. These factors were found by all participating vendors in this dataset.

However, Salesforce Einstein Discovery found that young people — less-experienced drivers — have almost the same elevated accident risk when driving in dark lighting conditions (e.g., in the early morning when driving to school), as when they are driving drunk or on drugs .This is an actionable finding for school administrators to consider when assessing school start times.

Example 4. *How Stories.bi (since acquired by Workday) found a specific positive trend in opioid addiction treatment that was impossible to identify from aggregate data analysis:*

- At Gartner's 2018 "BI Bake-Off," opioid epidemic data was given to the four participating vendors and all the vendors with a booth on the

Figure 7. Augmented Analytics List of Top Trends Contributing to the Opioid Crisis

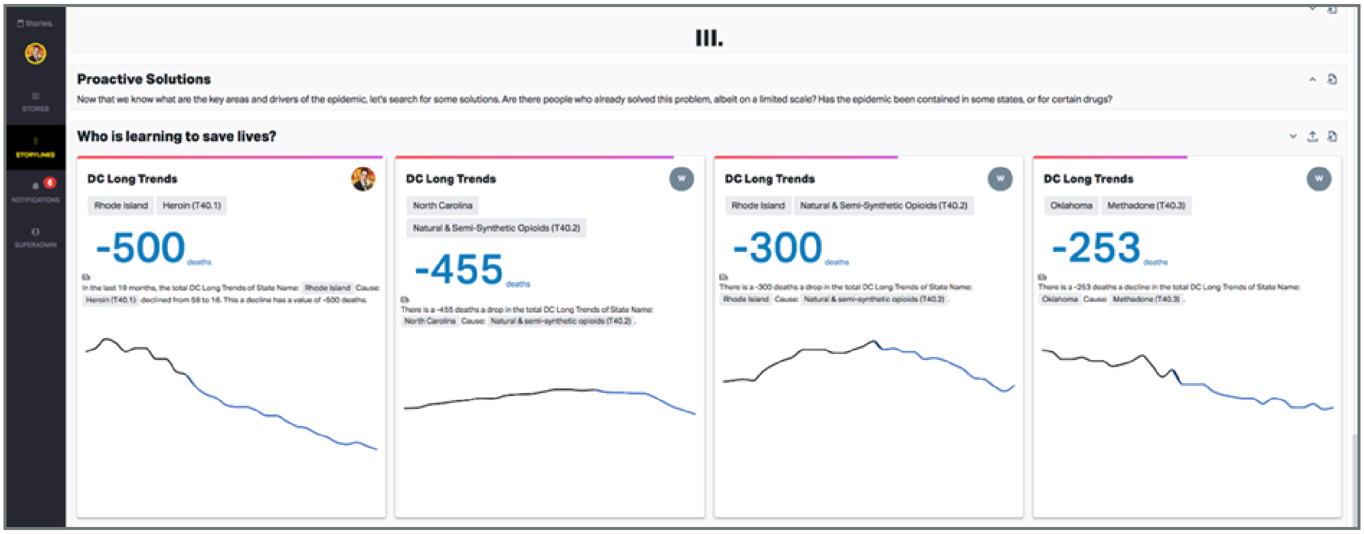


Source: Workday (Stories.bi)
ID: 444837

Workday (Stories.bi) also automatically detected a positive (downward) trend in overdose — essentially which locations are learning to save lives. It found pockets of excellence around the country where people have learned how to reverse the crisis. “Heroin in Rhode Island” came in No. 1 (see Figure 8). This trend would have been almost impossible to find manually as the overall Rhode Island trend was still upward. It didn’t come up in any report by state or by overdose cause.

Figure 8. Where Are People Learning to Save Lives in the Opioid Epidemic?

Where Are People Learning to Save Lives in the Opioid Epidemic?



Source: Workday (Stories.bi)
ID: 444837

- When the Stories.bi team did an online search for heroin in Rhode Island, they found that a unique program for treating heroin addictions in prisons had been started just prior to the trend reversal. [1]
- The program, launched in 2016 and the only one of its kind in the nation, screens all Rhode Island inmates for opioid use disorder and provides medications for addiction treatment (MAT) for those who need it. The program is proving effective.
- This trend, found in a few seconds by Stories.bi, was not found by any other vendor with access to the dataset.

Think of one of the most common cases of analytics: Business people monitoring revenue and profit in their business by a number of dimensions. Typically, known sales and profit drivers are built into dashboards that business people access every day. When sales trends change, users have to explore those relationships to find the root cause of the change. Is it pricing, promotions, packaging, product quality, the weather, a combination of these factors, or something else? What if the change is caused by new, unknown factors — factors not included in the dashboard or that the business person didn't think of or have time to fully explore?

How often do business people draw suboptimal conclusions from their data? How often do they explore what they think are the key drivers or attributes of an outcome variable, and stop when they confirm their hypotheses? How many times might there be other, more important factors affecting the outcome variable that they have not thought to explore?

This is the root of the challenge with the current manual process for exploring data and building data science and ML models. The desire to overcome it will drive the transformational nature of the next wave of market disruption. Namely, this is automation of all aspects of the analytics workflow. It will improve the accuracy and timeliness of advanced analysis (in light of the human context), reduce bias and elevate the skills of more users to the level of citizen data scientists.

Since automation will enable expert data scientists to focus on specialized problems and on operationalizing and embedding enterprise-grade models into applications, only the most accurate and significant insights will be acted on by users. Expanded use of automation should also translate into fewer errors from the bias inherent in manual exploration.

Augmented Analytics Is Transforming Each Part of the Analytics Workflow

Currently in analytics, content authors such as analysts, citizen data scientists and expert data scientists perform the following three data-to-insight-to-action activities iteratively to find meaningful insights:

1. Preparing the data

3. Sharing and operationalizing findings from the data

The augmented analytics paradigm accelerates the time it takes to get accurate insights for business users. It augments their analysis by using ML algorithms to automate those three main parts of the data and analytics workflow (see Figures 9 and 10). How augmented analytics affects each component of the workflow is explored below.

Figure 9. Current Data Analytics Workflow

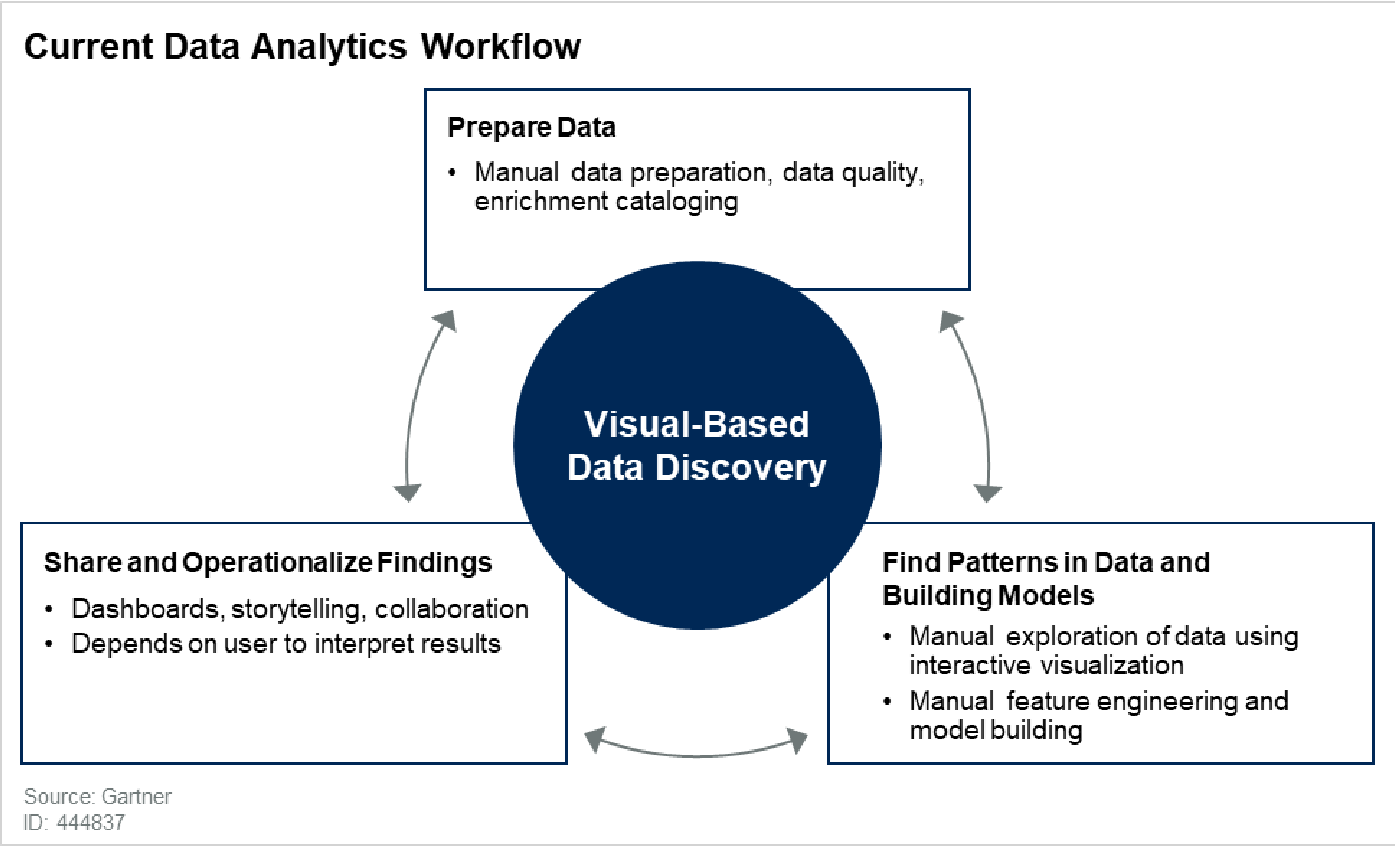
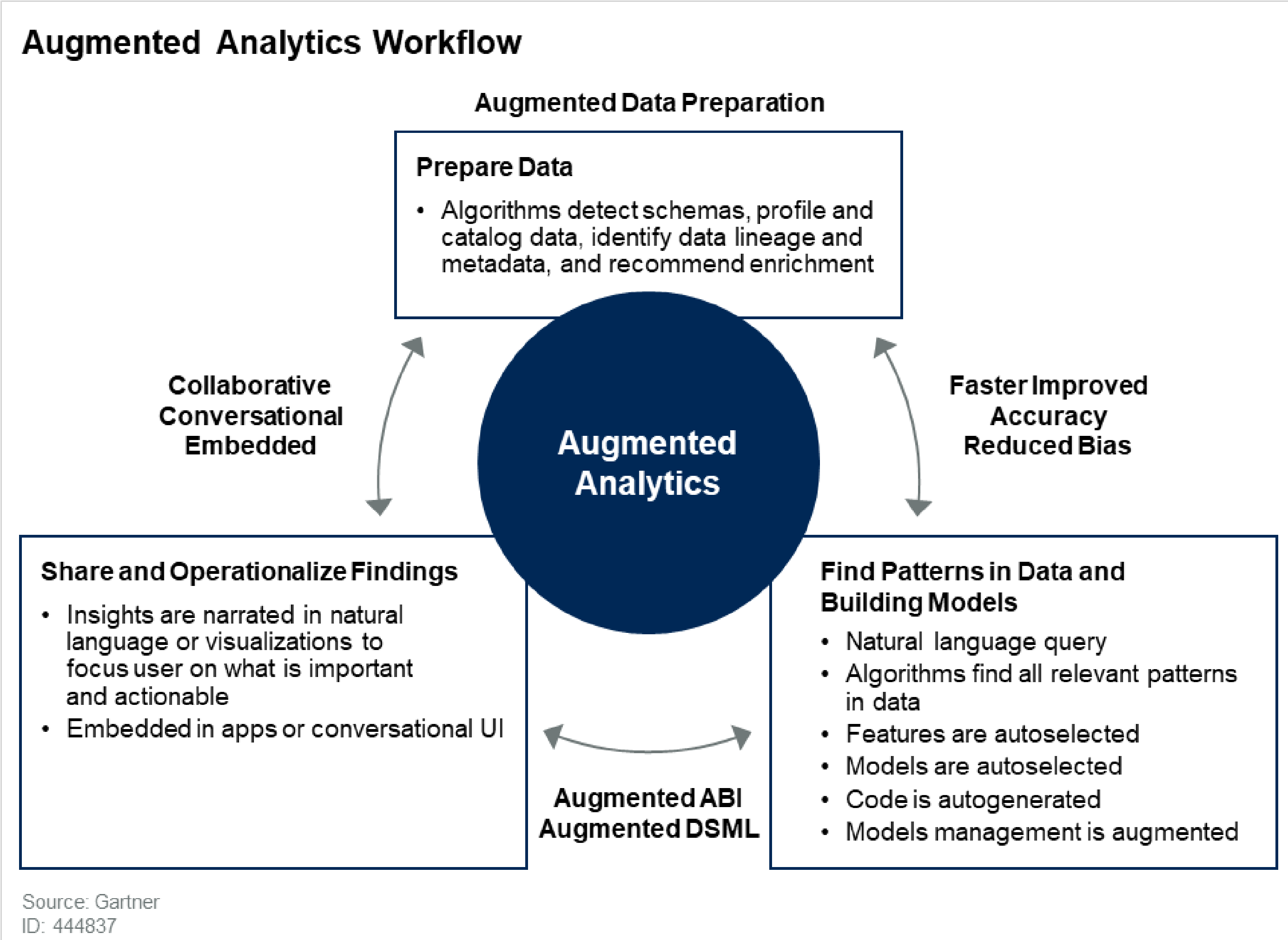


Figure 10. Augmented Analytics Workflow



Prepare Data: Augmented Data Preparation

Preparing data for analysis is the most time-consuming task facing data and analytics users. Most modern analytics and BI platforms offer basic data preparation capabilities for joining, data manipulation and transformation. Data science and ML platforms offer some data pipelining capabilities that are often incomplete and difficult to use. This leaves much of the data profiling, quality, modeling, manipulation, enrichment, metadata development and harmonization work to the business user or data scientist (if self-service), or to IT staff (if data enablement is centralized).

Either way, this creates a bottleneck for business users and data scientists. It also creates business risk due to a lack of governance, as organizations give more users the ability to build analytic content. Augmented data preparation — a component of augmented analytics — uses algorithms to find relationships in data. It also profiles and recommends the best approaches for cleaning, reconciling, enriching, manipulating and modeling data, with capabilities to capture metadata and lineage for reuse and governance. More broadly, automated ML/AI is transforming all of data management, including data cataloging, metadata management, data quality and database/data lake management.

Most analytics and BI platforms as well as data science and ML platform vendors are making data preparation — with varying degrees of ML automation — an investment priority, due to the major time-to-insight and governance improvements it can produce. The stand-alone data preparation market is crowded with venture-funded startups and established data integration players. Many of these are using ML to streamline and accelerate the data preparation process. They are making curated and described data accessible to all analytics content authors in data catalogs via easy-to-use interfaces such as search. (See "Market Guide for Data Preparation" and "Rebalance Your Integration Effort With a Mix of Human and Artificial Intelligence.")

Augmented data preparation enables users to combine more data sources from both trusted sources and external or ad hoc sources, faster than using traditional data integration approaches. It supports the growing requirement to deploy pervasive and trusted analytics to

Table 1 below shows two industry examples of the impacts of augmented data preparation. The section that follows details specific examples from different vendors.

Table 1: Examples of Augmented Data Preparation Impact

Industry	Augmented Data Preparation Impact
CPG	■ Data harmonization involved five people taking four to five weeks to access, clean, blend, harmonize, model and reconcile a company's retail POS, Nielsen, pricing and brand/category data. ■ Augmented data preparation reduced this to one person taking one hour, with one-click updates.
Banking	■ Regulatory compliance and anti-money-laundering processes from harmonizing data took 21 days. ■ Involved harmonizing data from internal transactions across the bank's network, ATMs, tellers, deposits, withdrawals, accounts and credit cards, across more than 40 countries, along with data originating from external sources such as money transfers and PEP data sources. ■ Augmented data preparation reduced the time taken by 95% to one day.

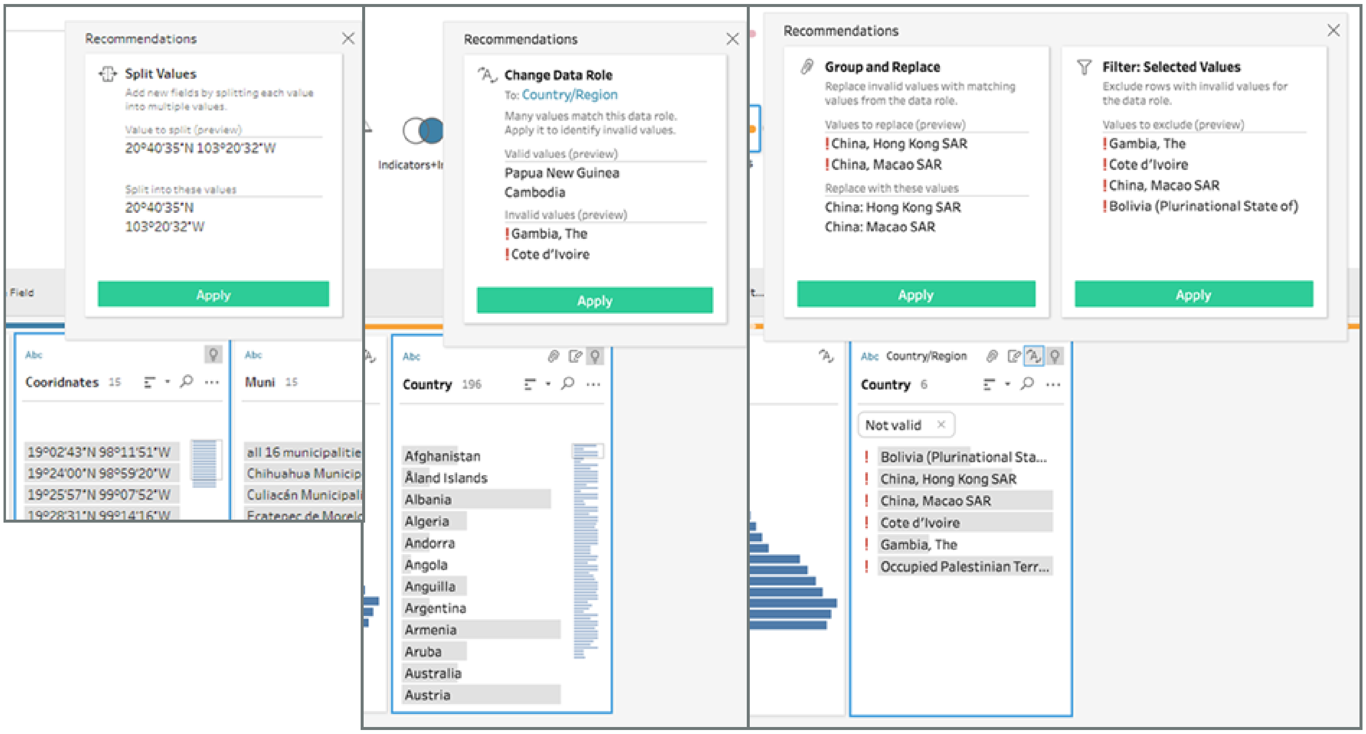
Source: Gartner (October 2019)

For augmented data preparation, Tableau Prep Builder from Tableau (Salesforce), a free product for Creator license customers and Tableau Data Prep, an optional component to Tableau Server, identifies if there is consistent pattern in data values in a field. It recommends ways to split and break out the field into useful ones. It detects that the field might have known semantics and then recommends changing the data field's role to the one detected. Once a field is tagged with semantics, the system will identify invalid values and automatically fix them by mapping them to the closest valid value. Tableau Prep Builder Data Prep also identifies some data quality issues, such as missing values, and provides the user with basic recommendations for how to deal with those missing values (not shown in Figure 11).

Some stand-alone augmented data preparation vendors such as Paxata and Trifacta have more extensive data profiling and recommendation features. Augmented analytics vendor, Outlier, for example, includes auto generated daily data quality "data stories" alongside autogenerated business stories that are consumed by business users as well as data engineers.

Figure 11. Using Augmented Analytics to Harmonize Data

Using Augmented Analytics to Harmonize Data



Source: Tableau (Salesforce)
ID: 444837

Sample Vendors: Augmented Data Preparation

Stand-alone augmented data preparation: Altair Datawatch, Paxata, Trifacta, Unifi.

Analytics and BI platforms with some tool specific augmented data prep features: Tableau, Oracle Analytics Cloud, TIBCO Spotfire, Salesforce Einstein Analytics, Qlik, IBM Cognos Analytics, SAS.

Find Patterns in Data: Analytics in Analytics and BI Platforms

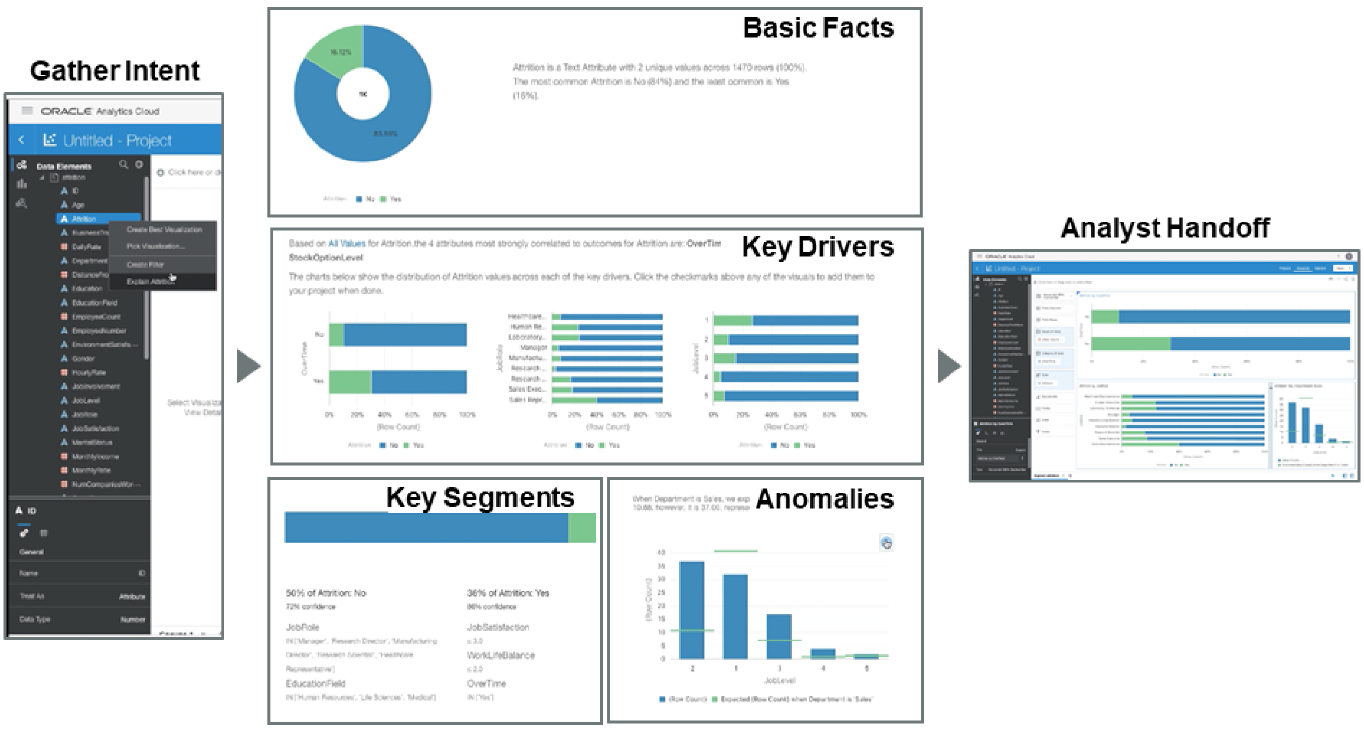
Current visual-based data discovery approaches used in modern analytics and BI platforms enable business people to visually explore relationships and patterns in data, using interactive techniques such as filtering, sorting, pivoting, linking, grouping and user-defined calculation.

With augmented analytics, instead of an analyst manually testing all the combinations of data, algorithms for detecting correlations, segments, clusters, outliers and relationships are automatically applied to the data. Only the most statistically significant and relevant results are presented to the user, in the form of smart visualizations that are optimized for the user's interpretation. Applying a range of algorithms to the data in parallel reduces the risk of missing important insights in the data.

Augmented analytics platforms should make the underlying models open for inspection, testing and validation by expert data scientists. This is important for building trust and confirming the accuracy of automated insights. See Figure 12 for an example from Oracle of how a user could automatically find patterns in data impacting attrition, instead of having to search for them. NLG is coupled with autogenerated visualizations to explain the findings to users. This is an evolution of the current visual-based exploration drag and drop paradigm.

Figure 12. Automated Insights Driving Attrition

Automated Insights Driving Attrition

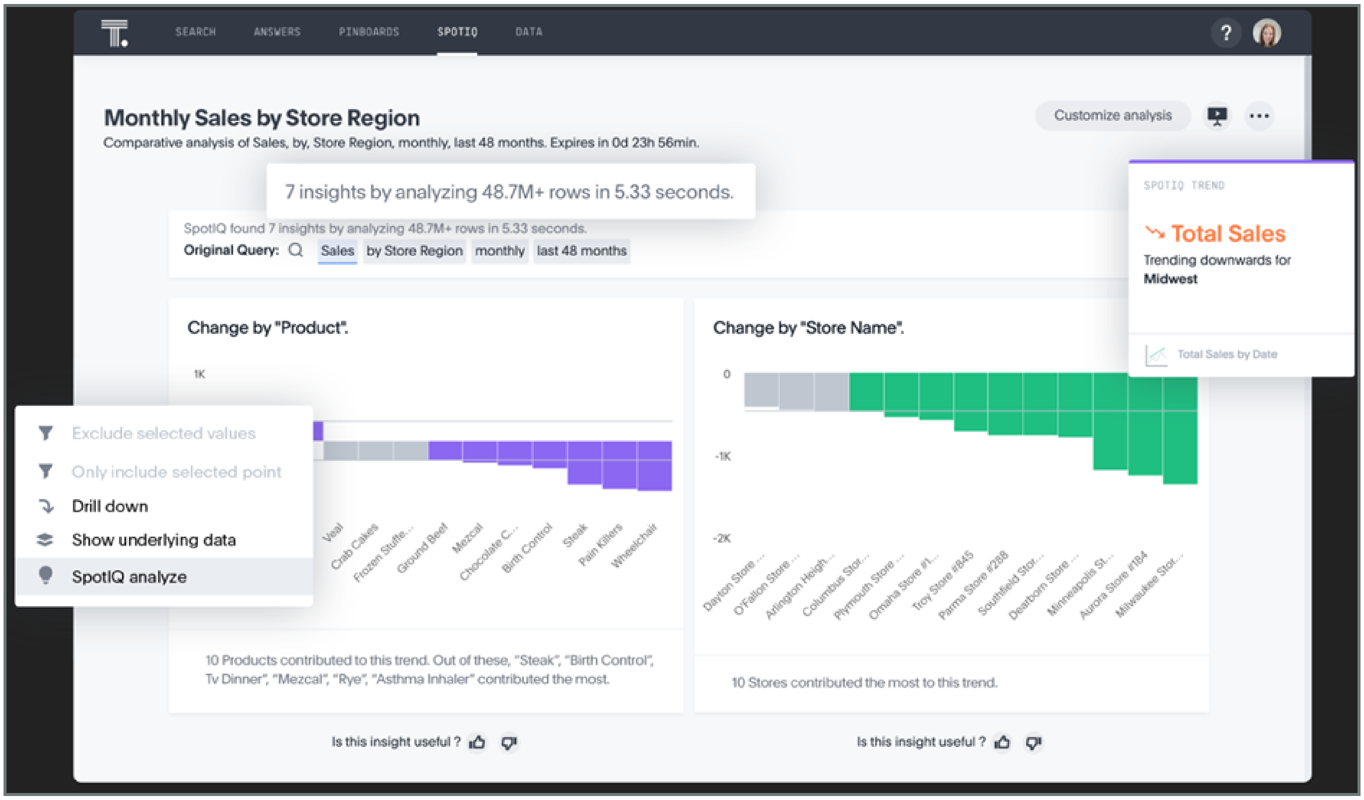


Source: Oracle
ID: 444837

Figure 13 shows how a ThoughtSpot platform user, asking an ad hoc question about their sales performance, can select an autoanalysis to generate 23 insights in order of importance and explained in natural language. This marries augmented analytics with a natural language query experience rather than a visual-based exploration approach to building queries.

Figure 13. Automated Insights Find and Explain Insights Driving Sales Changes

Automated Insights Find and Explain Insights Driving Sales Changes

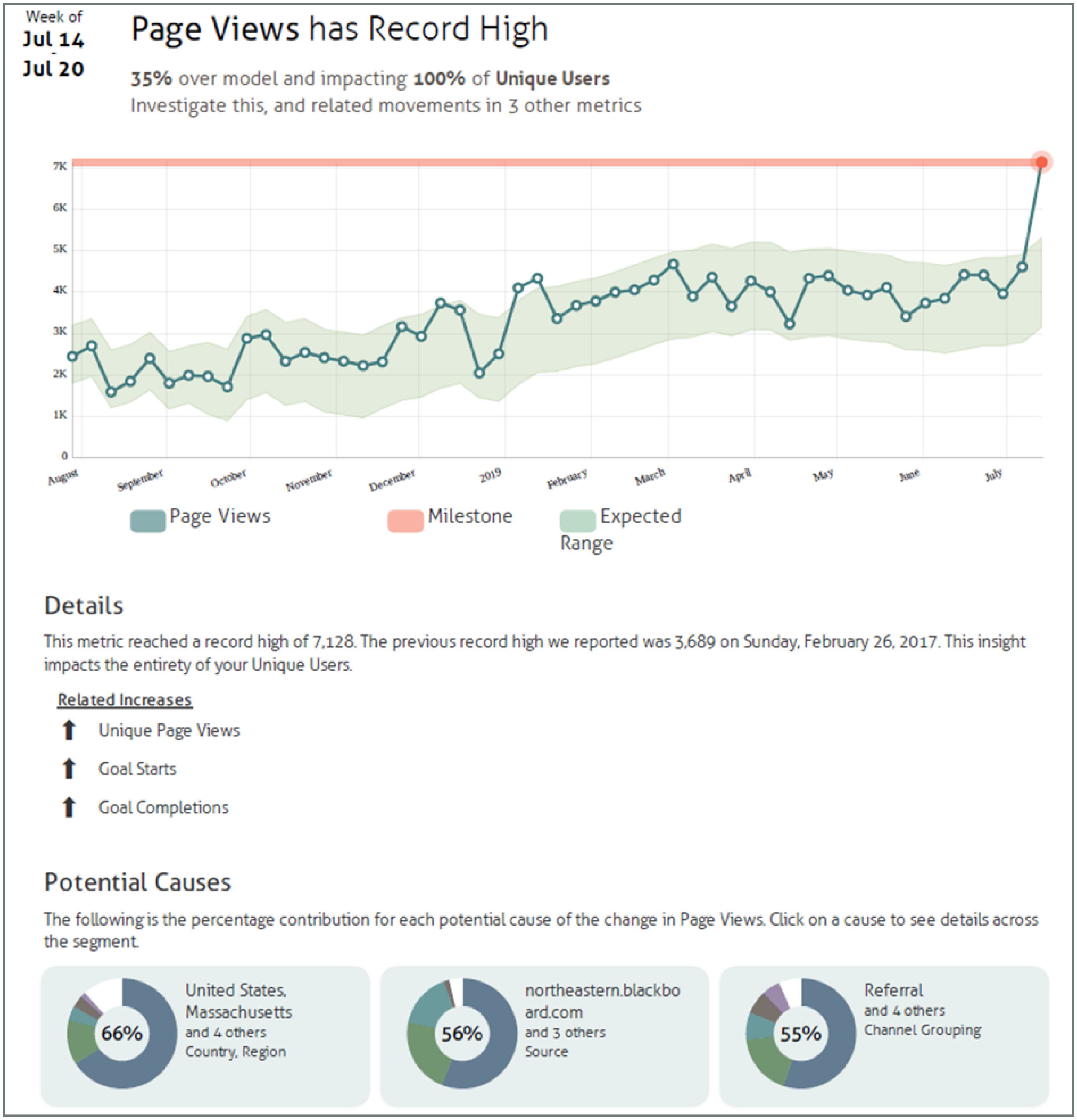


Source: ThoughtSpot
ID: 444837

Figure 14 shows how Outlier automatically creates stories in visualizations and a natural language narrative for the user showing statistically significant changes in key KPIs, potential causes and what to explore further.

Figure 14. Automated Stories Showing What Is Happening in Key KPIs and Why

Automated Stories Showing What is Happening in Key KPIs and Why



Source: Outlier
ID: 444837

Establishing processes that encourage, or even require, collaboration between citizen data scientists and expert data scientists is important to broaden adoption of and build trust in augmented analytics deployments.

AnswerRocket, IBM (Cognos Analytics, 2018 version), Microsoft (Power BI), Oracle (Analytics Cloud), Outlier, Salesforce (Einstein Discovery), SAP (Analytics Cloud), ThoughtSpot, Tellius, Qlik, Outlier.

Augmented DSML: Automate Aspects of Data Science and ML/AI Models

Augmented analytics has emerged as one of the most transformational innovations in data science and machine learning. It helps expert and citizen data scientists (business analysts or application developers who need to add ML/AI to their applications) quickly build and deploy models. Augmented data science and machine learning (augmented DSML) uses automation and embedded ML to help expert and citizen data scientists more efficiently generate, deploy and manage models, to generate inferences and predictive insights. Augmented DSML streamlines the feature engineering, model selection, model management and model operationalization process.

Autofeature engineering and model selection as a capability of augmented data science, has gained traction over the past year or so. Augmented analytics extends automation beyond the feature engineering and model selection to model operationalization (one-click deployments/code generation), and ultimately over time to model management and tuning life cycle (auto-monitoring and auto-alerting of model health and accuracy degradation). For augmented analytics in an analytics and BI platform, the user is a businessperson or citizen data scientist, and the output is an insight, both visual and narrated in natural language. A model is also generated. In more mature tools, the underlying model is made available for validation, to build trust and for further modification or operationalization by an expert data scientist, if required.

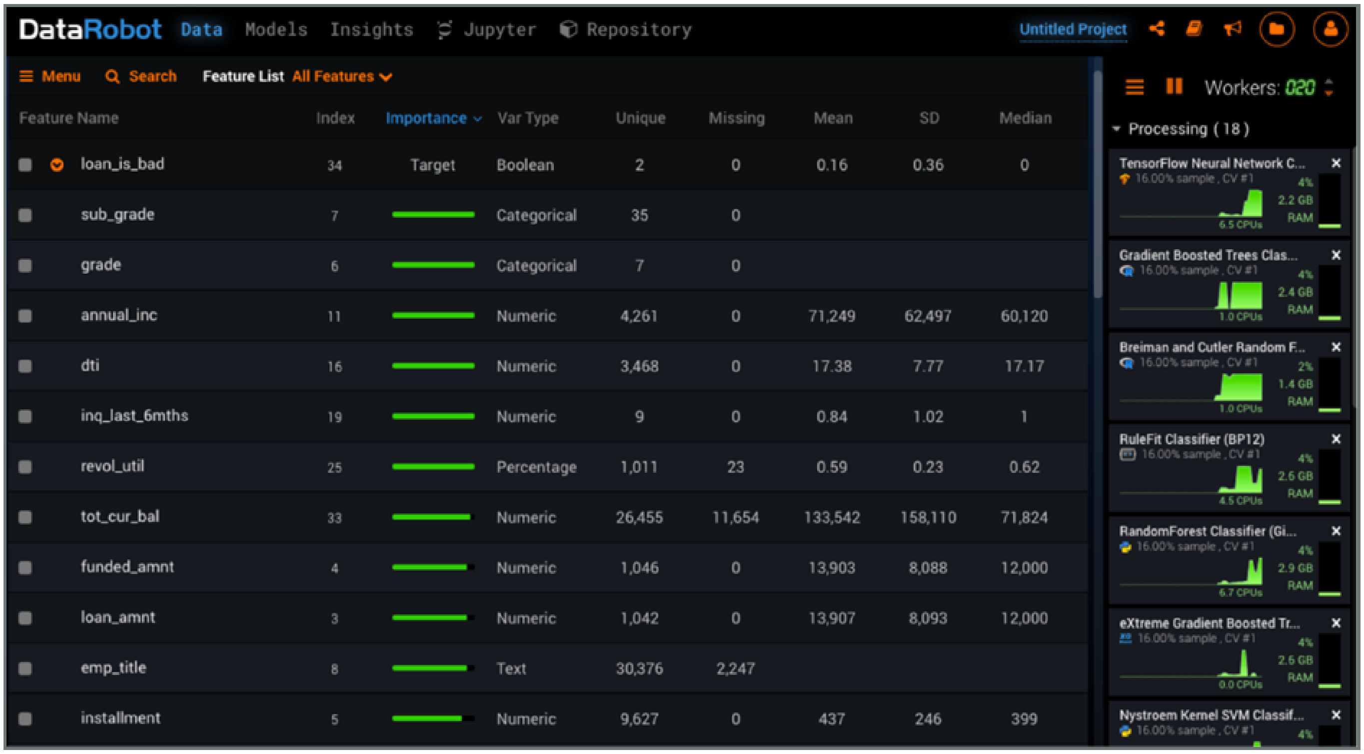
For augmented data science, the user is an expert or citizen data scientist, developer data scientist or expert data scientist, and the output is a model. The intent of augmented data science is to make the expert data scientist more productive, so they can focus their efforts on high-value tasks and making the enterprise-grade models they build scalable and less prone to bias.

Augmented data science also makes it possible to empower citizen data scientists with the tools they need to supplement the model building and integration done by expert data scientists. Given the scarcity of expert data scientists and the ever-increasing demand for their skills, even higher productivity will be expected and more analytics work needed of this new class of citizen data scientist (developer) who embed ML/AI into applications. These disruptive capabilities will help to ease the expert data science and ML/AI talent and skills shortage facing the market.

Figure 15 shows models on the right that were autogenerated in DataRobot’s platform based on the features on the left to predict the likelihood of loan default.

Figure 15. Automated DSML Uncovers Loan Default Drivers

Automated DSML Uncovers Loan Default Drivers



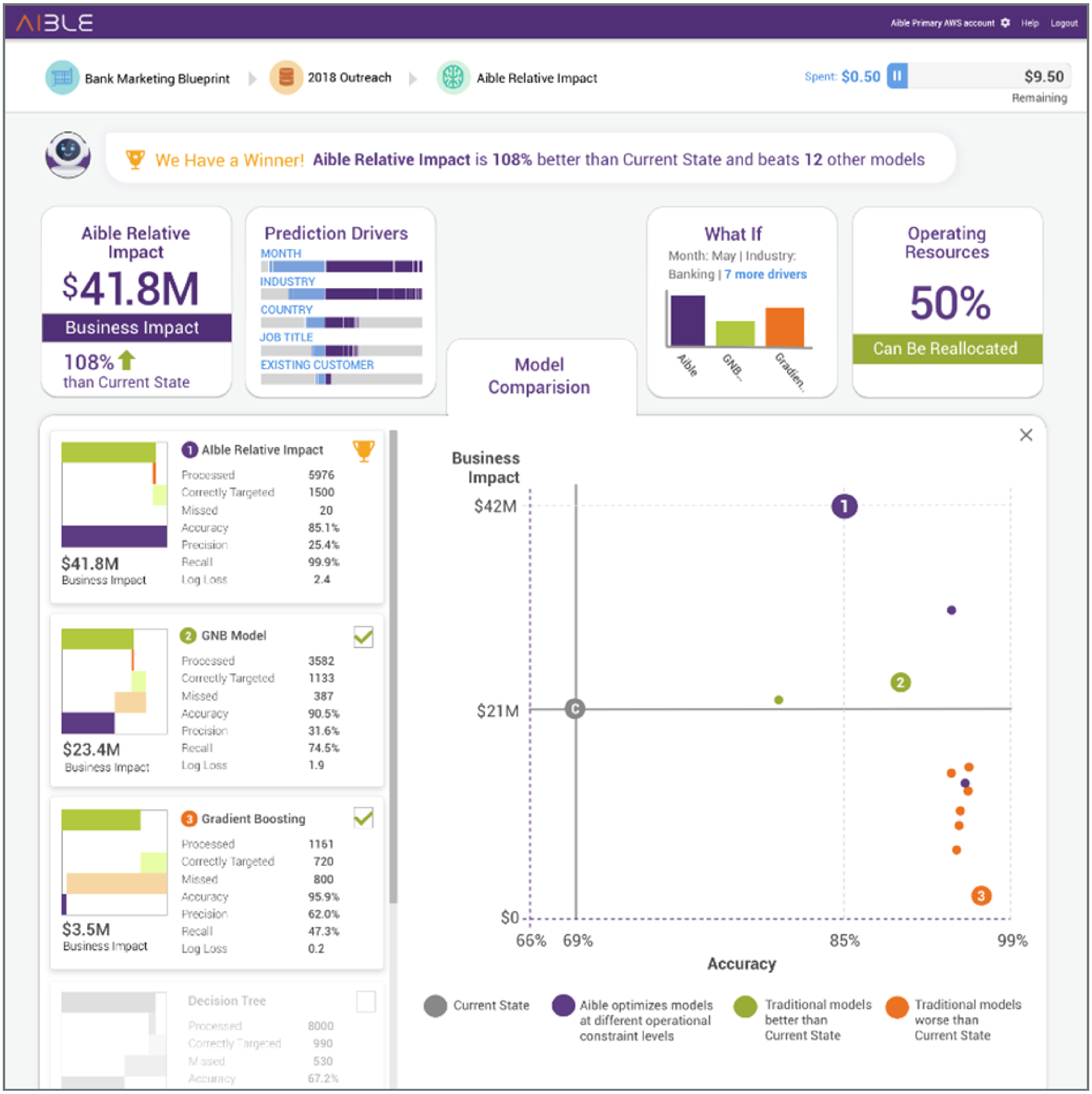
Source: DataRobot
ID: 444837

Figure 16 shows a different approach to augmented DSML. Aible automatically selects the best model based on expected business impact subject to user input. The model is based on (a) cost-benefit trade-offs of correct and incorrect predictions; (b) operational constraints such as how many customers they can pursue with their existing sales team; and (c) what they are doing today before adopting the model.

In this case the value of correctly targeting a prospect was 50 times greater than the cost of incorrectly pursuing a prospect that wouldn't have become a customer. The Aible Relative Impact model is far more aggressive and catches most of the prospects that would have bought. However, it has far worse accuracy and Log Loss ratings because many of the prospects it suggests pursuing don't actually buy. Business users, however, care about the net business impact and are willing to pursue far more leads to get more wins as long as they have the operating resources to pursue these leads. In this case, they had the capacity to pursue 12,000 leads and thus even the aggressive model uses only 50% of the available resources.

Figure 16. Augmented DSML Model Selection Optimized for Business Impact

Augmented DSML Model Selection Optimized for Business Impact



Source: Aible
ID: 444837

Sample Vendors: Augmented Data Science

Big Squid, DataRobot, dotData (spin off of NEC), Firefly, Google, H2O.ai, Prevedere, RapidMiner, SparkBeyond, Aible, DataStories.

Examples of the Business Impact of Augmented Analytics in Analytics and BI Platforms and Augmented Data Science Platforms

Table 2 shows examples of how augmented analytics has had an impact across industry use cases. Also, see: "Four Real World Case Studies: Implement Augmented DSML to Enable Expert and Citizen Data Scientists."

Industry ↓	Before Augmented Analytics ↓	After Augmented Analytics ↓
Banking	Targeted older customers for wealth management services.	Found that wealthy young clients between the ages of 20 and 35 are more likely to transition into wealth management.
Agriculture	Data scientists took months to build models to find the best handful of hybrid seed combinations out of thousands to sell to farmers.	Domain specialist geneticists took over the process and reduced process duration to days.
Agriculture	Expert data sciences could not find source of crop yield decline.	Identified a combination of previously unknown factors that changed physical asset planning.
Healthcare	Tracked patient sickness measures as a key driver of transportation (ambulance) costs.	Found main cost driver in segment of under 12 year-olds. Investigation found contracts with vendors allowed charging by person, and parents accompany sick children to treatments.
Insurance	Thought that case rate eliminated a financial incentive for insurance companies to have patients go home sooner than they should, or for hospitals to keep them in the hospital longer than they should.	Discovered that, for the same procedures, the day of the week on which the member was admitted to some hospitals made a difference in the overall hospital care cost. Occurring despite case rate, which is a “one size fits all” rate designed to be fair to both the insurance company and the hospital.
Healthcare	Total cost of care differed by patient demographics, geographies, treatment modalities and providers.	Found that some providers are more efficient than others at handling certain kinds of population. Enabling members who fit a certain profile to get their primary preventative care in the best setting for them made patients healthier and kept them out of hospital, which also reduced costs.
Insurance	Looked for the drivers affecting life insurance purchase.	Found a previously unknown purchasing driver — time from a person’s birthday — that resulted in a change in marketing and upsell/cross-sell rate.
Fashion	Merchandisers set seasonal discounting timing and amounts based on experience.	Margin improved by 50% by automating discount recommendations for each product.
Food and beverage	Location of more-profitable fountain drinks in a fast-food restaurant relative to bottle drinks was not considered.	After a coincidental change in store location due to remodeling, the augmented analytics system picked up a 20% increase in sales of fountain drinks and profitability — changing plans for future store design.
Higher education	Targeting higher-revenue student segment was previously based on application numbers by region to focus marketing funds.	Uncovered unknown patterns that doubled the chance that the student would accept the placement offer.
Not for profit (microfinance lender)	Traditional risk assessment model accuracy for microfinance lending risks resulting in loan fees of as much as 33% of the value of each loan, to compensate for high risk. This resulted in a microloan product that was too expensive for broad market adoption. The models also took months of specialist data scientist time to build.	Business people created two models in less than two weeks that were trained to identify new opportunities using ML, to minimize risk and create new loan products that could be priced more attractively. These models improved the percentage of loans repaid to lenders by reducing loan defaults by 5%.
Healthcare	Business people across payers and providers needed to understand their market’s population cohort, track progress and assess performance. However, users without advanced analytical skills found it difficult to quickly understand and interpret the visualizations.	NLG explanations were added to interactive visualizations to explain important findings to payers and providers, so that they could have a common understanding of whether reimbursement rates correlate with high-quality outcomes. Findings around shared savings and trends in the reimbursement rate can be communicated to people with different levels of analytics skills based on a single interpretation of what is most statistically significant in the data.

Industry	Before Augmented Analytics	After Augmented Analytics
Banking	"Thin file" credit scores took weeks.	Built credit scores for thin-file consumers in less than an hour by microsegmenting customers into granular buckets, testing 10 models per hour (as opposed to one every two to four weeks using traditional techniques).

Source: Gartner (October 2019)

Differences Between Augmented Analytics and BI and Augmented Data Science and ML Platforms

Augmented analytics within modern analytics and BI platforms deliver insights to citizen data scientists (see Figure 17). A model is generated and can be embedded in an application, after further vetting by an expert data scientist. But the goal or deliverable is insight. NLQ and NLG are important user experience features.

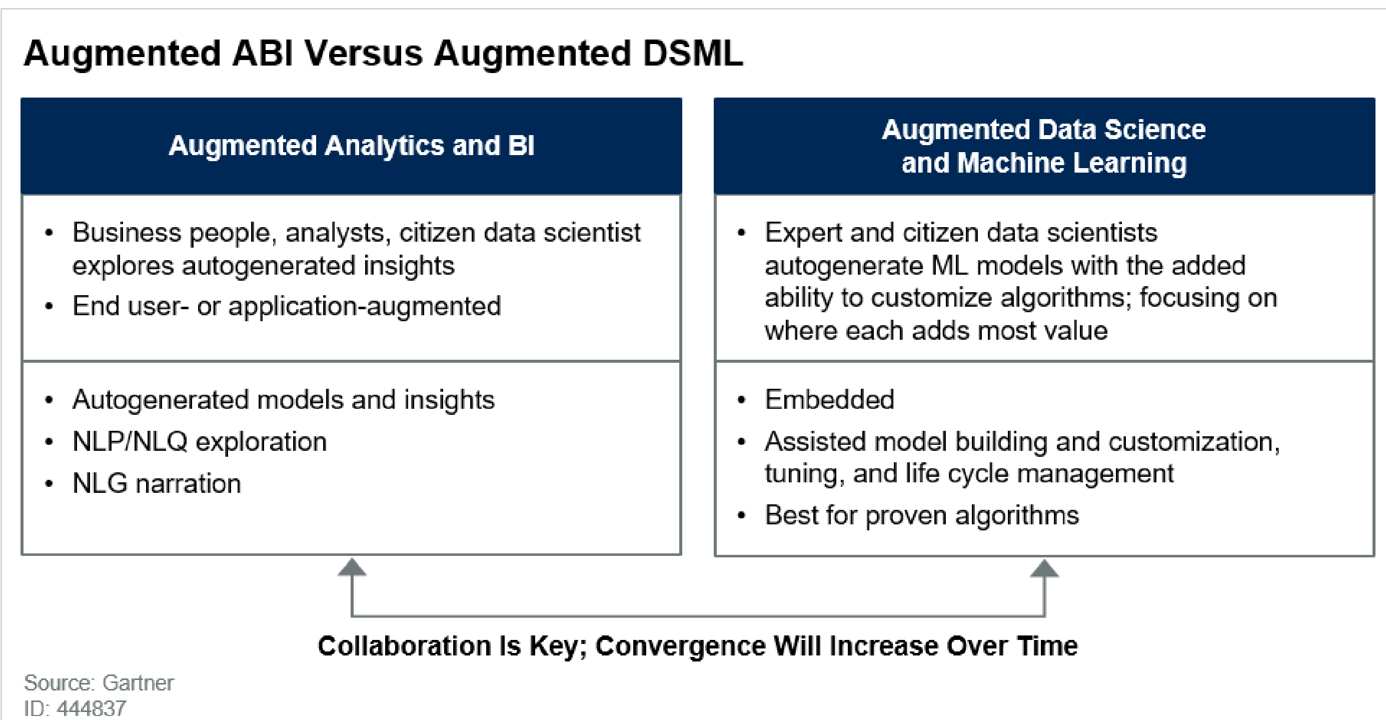
Augmented data science platforms, by contrast, automatically generate a model for either a citizen data scientist (analyst or application developer) or an expert data scientist, or for embedding. These platforms assist in model building, life cycle management and governance.

The differences between the two types of platforms are subtle and narrowing to the point where greater convergence over time is likely.

Augmented ABI as well as augmented DSML both reduce time-consuming exploration and the identification of false or less-relevant insights. They require a collaborative process that focuses a citizen data scientist on what is important, and provides an expert data scientist with a starting point or early prototype to explore and operationalize models for only relevant patterns. Both the citizen and expert data scientists become more productive by reducing the experimentation/initial exploration phase. Augmented DSML platforms can also help reduce bias for expert data scientists via their automated feature generation and algorithm selection. This ultimately results in faster times to insight and action.

Currently, machine-assisted models work best with proven algorithms versus cutting-edge techniques.

Figure 17. How Augmented Analytics and Augmented Data Science and Machine Learning Platforms Differ



Sharing and Operationalizing Findings

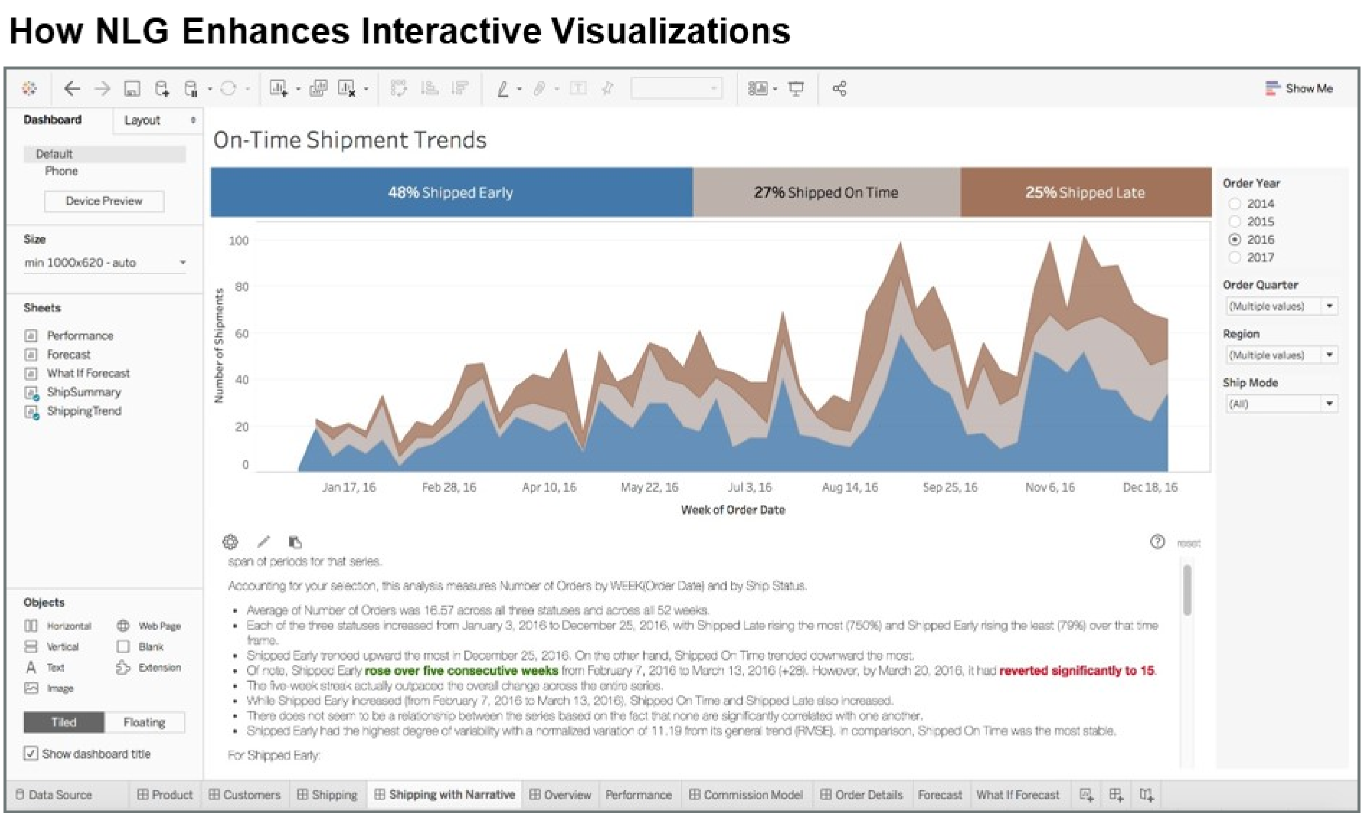
Conversational Interfaces

significant in the data, and many users lack the ability to fully interpret statistically significant visual-based insights.

With the addition of NLG, augmented analytics platforms automatically present a written or spoken context-based narrative or explanation of findings in the data. Alongside the visualization, this informs the user about what is most important for them to act on the data.

Figure 18 shows an example of how NLG, when combined with an interactive visualization, can enhance the finding and communication of important insights to users. The stacked line chart from Tableau (Salesforce) shows the breakdown of shipment timing. The data behind the chart powers the story, provided by the Narrative Science platform. The story highlights key insights such as highs/lows, trends, streaks, correlations and predictions that are not readily apparent from the visualization alone.

Figure 18. How NLG Enhances Interactive Visualizations



Source: Narrative Science and Tableau
ID: 444837

While integration between augmented analytics and conversational chatbots for analytics is still immature and most vendors require some custom integration and set up, Qlik has acquired and integrated its Crunchbot acquisition with Qlik Sense Enterprise. In Figure 19, a Qlik Insight Bot is running within Qlik Sense Enterprise and it shows the Bot responding to a question "What are my total sales?." The smart Bot not only answers the question but then, without prompting, shows everything it knows about sales by producing a chart on the fly and providing a natural language analysis of the data that makes up total sales. It will also understand the user's context. If the user's next question is "What about Newark?" it will answer knowing that the user is still interested in Sales.

Figure 19. Finding All Patterns in Sales Using Conversational Interfaces

Finding All Patterns in Sales Using Conversational Interfaces

Source: Qlik
ID: 444837

Sample Vendors: NLG/Conversational Chatbots for Analytics

Arria NLG, Automated Insights, Qlik, Marlabs, Narrative Science, Unscrambl, Yseop.

Augmented Analytics Embedded in Applications

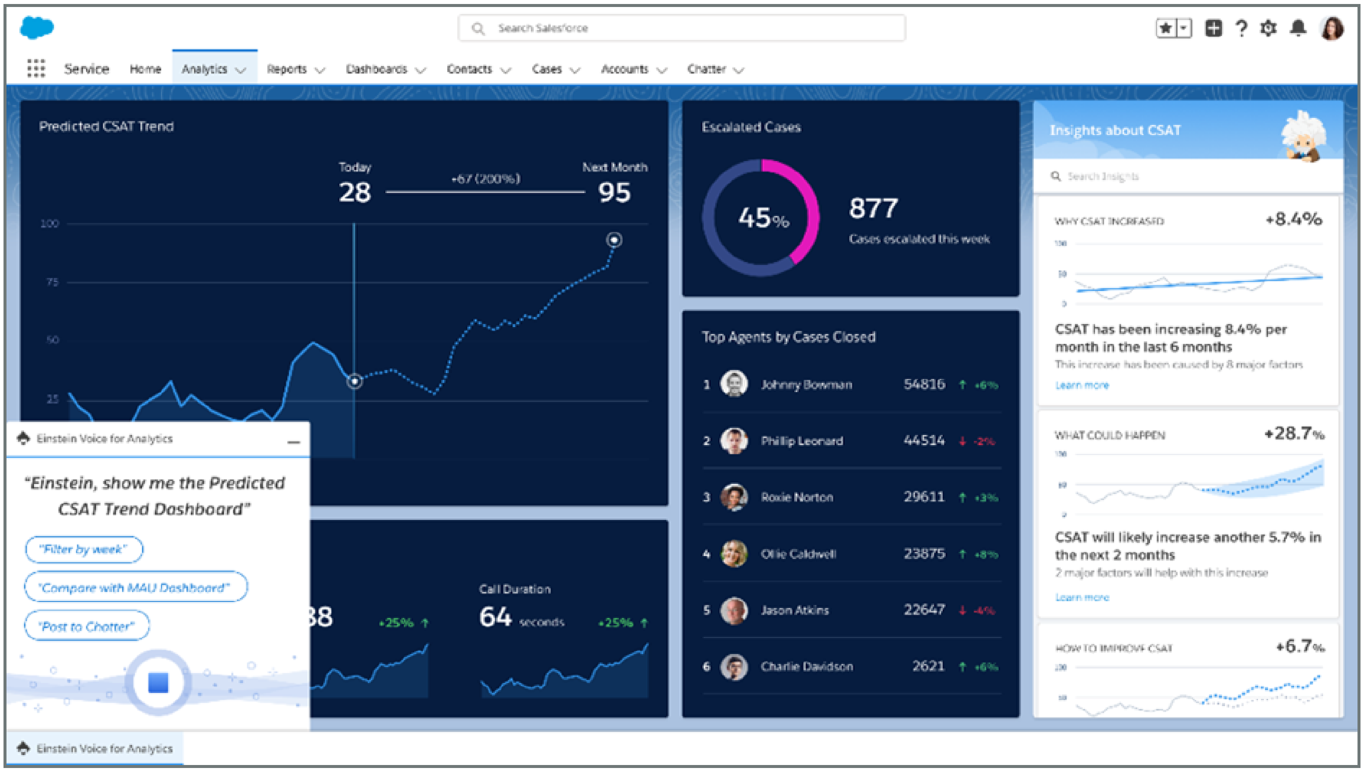
Expanding access to insights from analytics to all workers will be key to driving transformative business impact. However, access to analytics content from analytics and BI as well as data science platforms has mostly been limited to power users, business analytics users and expert data scientists with varying degrees of analytical and technical skills. A lot of root cause analysis is also done by people in IT operations or business operations or factory operations who are not data scientists or even citizen data scientists. Gartner surveys show that only around 35% of employees have access to analytics and BI tools (see “Survey Analysis: Why BI and Analytics Adoption Remains Low and How to Expand Its Reach”).

Increasingly, automated models are being embedded in enterprise applications for sales, marketing, HR and finance teams. Workforce analytics, supply chain analytics and CRM analytics will be the largest functional domains to benefit from augmented analytics (see “Market Opportunity Map: Analytics and Business Intelligence, Worldwide”).

Salesforce is embedding its Einstein Discovery capability within its sales, service and marketing applications. Figure 20 shows a customer satisfaction dashboard generated using NLQ that reveals the most relevant insights and predictions with an explanation of findings. Conversational analytics has the potential to address analytics adoption challenges. It will do so by enabling any employee to interact with data using natural language when on a mobile device or outside a dashboard, to gain optimized insights for their role and context. For example, instead of logging into a dashboard, any user — from C-level to analysts and operational workers — can interact with personal digital (analytics) assistants (e.g., Amazon Alexa, Google Home) to ask for a relevant analysis. Sales managers might ask for an analysis of sales or sales pipeline, based on their role. They will be served an explanation or narrative of statically important drivers of change, and might be sent visualizations to show important trends, patterns or outliers, based on their role.

Figure 20. Augmented Analytics Within an Embedded Customer Service Application

Augmented Analytics Within an Embedded Customer Service Application



Source: Salesforce
ID: 444837

Sample Vendors: Enterprise Applications With Embedded Augmented Analytics Capabilities

Adobe, Salesforce, Workday

Update Roles and Invest in Data Literacy

New roles are emerging to support more-agile and pervasive analytics, which will be more widely adopted due to augmented analytics.

Data engineers may sit in IT or in the line of business, but they are charged with curating data for distributed content authors in the organization. (See "Toolkit: Job Description for the Role of a Data Engineer" and "Market Guide for Data Preparation.")

Citizen data scientists are emerging to help fill the gap (see Glossary Terms). The gap between business analyst and citizen data scientist is not wide — a few weeks of training on the proper tools, analytical concepts and best practices might be all that is required to transform the former into the latter. By contrast, it could take years of intensive training and real-world experience to turn a business analyst into a highly skilled data scientist. This is time that few organizations can afford. (See "Pursue Citizen Data Science to Expand Analytics Use Cases," "How Citizen Data Science Can Maximize Self-Service Analytics and Extend Data Science" and "Citizen Data Scientists and Why They Matter.")

Incorporating citizen data scientists into the data science and ML process also facilitates collaboration with expert data scientists, enabling each role to focus on the parts of the process where they can add the most value. This means better allocation of resources. For example, the citizen data scientists can focus on data ingestion, exploration, feature engineering and initial model development; the expert data scientists can focus on model validation, testing, delivery and operationalization.

Developer data scientists, a type of citizen data scientist, is a significant development driven by augmented analytics. These are application developers armed with augmented data science tools, who can build ML and AI models to embed in their applications. They will relieve intense demand for expert data science skills and offer opportunities to upskill existing application developers who are in more abundant

Data and analytics leaders should put in place a formal upskilling program to retrain and enhance the skills of existing people who can leverage augmented analytics capabilities to meet the growing demand.

Adoption: Scale Augmented Analytics Across the Enterprise

Efforts to incorporate augmented analytics will likely encounter resistance, for several reasons:

- The perception that augmented analytics tools are not transparent and represent a “black box” approach to decision-making
- Users not trusting recommendations or insights without seeing the factors that drove a specific recommendation
- The threat to job security, as redivision of workloads and redefinition of work processes upsets the status quo
- Business leaders’ reliance on intuition and traditional decision-making practices, and their resistance to change
- The mistaken belief that analytics maturity follows a linear progression and maturation process, such that predictive and prescriptive analytics can be considered only once a solid data foundation has been established

These challenges will require data and analytics leaders to challenge the current processes and approaches to analytics, demonstrate the gaps and limitations in current approaches, and create an environment and culture that supports change. An effective way to demonstrate the potential value of augmented analytics is to identify business problems or specific business decisions where efforts to use current data and analytics approaches have failed to deliver results in timely, relevant or actionable ways. This approach can help business users form a relevant, contextual connection between a specific problem and a technical solution. By identifying gaps, errors or issues in the historical analytic processes, you can lay the foundation for new, more automated approaches.

AI Governance: A Focus on Explainability

Beyond adoption risk, findings and insights that surface through the use of augmented analytics, must be verified or tested for relevance, accuracy and bias. Findings should inform/augment a decision-maker who can interpret the discovered insights using experience and human intuition in order to decide on a course of action.

> **Augmented analytics represents a new approach to problem-solving that supports humans in the decision-making process, not replace them.**

Acquiring the necessary data literacy skills will be a challenge to the responsible use of augmented analytics. As augmented analytics makes actionable insights available to users, it must also provide context, interpret findings and recommend how best to act on the discoveries. This requires everyone in the organization — not just managers — to have some knowledge of data analysis, statistics and interpretation.

Organizations need to recruit people with analytics skills across all job categories, and invest in data literacy as an ongoing priority. According to our research and TalentNeuron data service, demand (as reflected in job descriptions) for analytics-related skills from 2012 to 2016 was 4.3 times higher in non-IT jobs than in IT jobs (see “The Talent Implications of Digitization”). Also according to TalentNeuron, jobs posted for AI and as data and analytics specific jobs in the top 12 countries — by GDP between July 2015-March 2019 — was 2.2 times higher outside IT than inside. These numbers reflect the fact that analytics will become a necessary component of every job. And augmented analytics will accelerate this trend.

The enablement of citizen data scientists through mainstream adoption of augmented analytics will require data and analytics leaders to further emphasize the need for AI and analytics governance and collaboration between analysts and data scientists.

> **Be aware of the limitations of machine-assisted models as they work best with proven algorithms versus cutting-edge techniques.**

Existing organizational models will need to evolve in order to support adoption of augmented analytics capabilities and a growing footprint of citizen data scientists embedded within business units. The rise of self-service visual-based data discovery stimulated the first wave of transition from centrally provisioned traditional BI to decentralized analysis. However, the emergence of augmented analytics represents an entirely new level of business user autonomy, which could not only yield sizable returns but, if left unchecked, also have adverse results.

Data and analytics leaders must develop guidelines outlining the "rules of the road" with respect to where primary responsibility lies for accessing, preparing, provisioning and validating data accessed by augmented analytics tools. You must outline similar rules to govern the use of analytic content and insights created as outputs of augmented analytics as well as data science and ML tools, to ensure the accuracy, validity and bias levels of findings and recommendations. (See "How to balance control and agility with your self-service analytics.") Enabling citizen data scientists within an organization to use such tools will promote widespread use of higher-value analytics within business processes. However, inputs and outputs need to be validated, which will require processes and collaboration between IT staff, business users and data science teams.

Assessing Vendors: Evaluation Factors

Startups and some large vendors are offering a range of augmented analytics capabilities that have the potential to disrupt current modern analytics and BI vendors in the long term. This will force data and analytics leaders to reevaluate investments. Initially, most organizations should complement modern analytics and BI platforms with augmented analytics tools or leverage the augmented features being made available in their existing ABI platforms as they become available.

Many of the normal considerations when evaluating vendors, apply to vendors of augmented analytics tools. They include strategic factors (vendor viability, global presence, support and pricing) and functionality factors.

In terms of functionality, distinct considerations for the evaluation of augmented analytics vendors include:

- **Data access and preparation.** The breadth of data sources to analyze differentiates in augmented analytics tools. Most initially support only structured data sources. More advanced products support ingestion of messy, multistructured data sources stored in a variety of formats, both on-premises and in the cloud. A product should profile the data and make intelligent recommendations and inferences about how to cleanse and enrich it. Advanced features include recommending datasets to combine. For example, when analyzing traffic fatalities, it may make sense to combine data about these fatalities with public data about population density.

- **Algorithms, transparency and interoperability.** Consider the range of algorithms supported. Ideally, the product should allow additional/custom algorithms to be added to the out-of-the-box libraries, or refined using data science languages such as Python and R. Are autogenerated insights and models open for examination and further customization? This is important to building trust.

- **Embeddability/operationalization.** Can autogenerated insights and models be embedded in other applications? This is important to scaling the value and business impact. What capabilities are there for model management and tuning?

- **Business-friendly explanation of models.** AI governance is a critical competency as more and more AI models are autogenerated. Building trust in model results is key to adoption. Augmented analytics platforms should offer mechanisms for explaining models and insights generated in business-friendly terms. This may also support regulatory requirements such as GDPR.

- **Identification of bias and privacy violation risk in models.** Part of AI governance, emerging augmented analytics vendors should offer capabilities to identify when additional data is needed to mitigate bias in data. Privacy violation risk is another emerging capability to support AI governance.

- **Explainability of models.** An important part of AI governance is also being able to explain in plain language how a model works, and its drivers, strengths and limitations. While this will be required as part of compliance with privacy legislation, such as GDPR, it also is critical to building trust in models and protection brand reputation.

- **Accuracy of models.** Accuracy of models is key to building trust. A best practice is to back-test platforms on historical data to test accuracy with actual outcomes, and to run them in parallel with existing models if they exist.

- **Integration with NLQ, NLG and visual-based exploration.** Users should be able to ask questions using NLQ search, either entered via a search box or via a conversational chatbot for analytics. Relevant results should be narrated by text or voice. Vendors may develop their own NLQ and NLG interfaces or partner with third-party providers. Do augmented analytics platforms support integration with

- **Number of variables supported.** How many variables can be autoanalyzed at one time? The more the better. Check for limitations so that models do not need to be iterated as this can be time-consuming.
- **Upfront setup or configuration/modeling required.** Assess how much upfront set up is required prior to insights and models being autogenerated. What skills are needed for this set-up? How many ongoing services from the vendor are required? Are these skills widely available and/or easily taught?
- **Deployment options.** Are solutions available in cloud, on-premises and hybrid deployment modes? Importantly, does data need to be duplicated into the vendor's data mart in the cloud or on-premises for automated analysis, or can processing be pushed down to the original data source?

Augmented analytics is a critical capability evaluated for vendors that qualify for inclusion in the "Market Guide for Augmented Analytics Tools," "Magic Quadrant for Analytics and Business Intelligence Platforms," "Critical Capabilities for Analytics and Business Intelligence Platforms" and "Toolkit: Visual Guide to Analytics and Business Intelligence Platform Capabilities.")

Recommendations: The Path to Augmented Analytics Adoption

Data and analytics leaders planning to modernize using next-generation analytics should also see Figure 2:

1. Pilot and validate:

- Identify where automating algorithms to detect patterns in data could reduce the exploration phase of analysis and improve highly skilled data scientists' productivity, while recognizing that they still need to validate models, findings and applications.
- Launch an augmented analytics pilot to assess viability and prove value. Start with a small list of specific business problems that are currently tackled manually and are time-intensive or prone to bias.
- Evaluate the extent of augmented analytics features and their role in overall data management, as well as specifically in the data preparation and cataloging process, when evaluating data preparation and data cataloging vendors.
- Factor in the limitations of machine-assisted models as they work best with proven algorithms versus cutting-edge techniques.

2. Update roles and invest in data literacy:

- Recognize that augmented analytics is here and will be rolled out in the platforms you already have deployed over the next one to three years. Plan and invest in upskilling business people with data literacy.
- Identify and upskill citizen data scientists (analysts, developers) in your organization.

3. Scale across the enterprise:

- Educate business leaders and decision-makers about the potential transformational impact that augmented analytics can have if used by a wider audience. Stress the need for responsible use and governance to avoid unintended consequences.
- Develop guidelines for appropriate use of augmented analytics tools and capabilities, with an emphasis on people and process.
- Modify current deployment models, emphasizing the need for analytic/AI governance and providing incentives for collaboration between citizen data scientists, data engineers and expert data scientists.
- Run augmented analytics initiatives in parallel with existing analytics and decision processes to prove their value and build trust in augmented analytics. Expert data scientists are often wary of black-box approaches. They are also likely to be wary of someone less skilled doing what has historically been their job.
- Encourage citizen data scientists to work collaboratively and iteratively with expert data scientists — both internal, if available, and external.

4. Mitigate expert pushback and user misinterpretation:

- Use augmented analytics tools to confirm or challenge findings surfaced by human interpretation of manual data discovery exercises. Validate findings with expert data scientists.
- Engage both business analysts and data scientists in learning about and incorporating augmented analytics tools into the analytic process, in an effort to identify the best division of responsibility between the roles.
- Recruit people with analytics skills across all job functions, and expand investments in companywide data literacy.
- Incorporate augmented analytics using a collaborative approach between citizen data scientists and expert data scientists, having each focus on the steps of the analytic process where they add the most value.
- Focus on explainability of insights and models as a key feature of augmented analytics platforms.

5. Assess vendors:

- Familiarize yourself with and monitor the augmented analytics capabilities and roadmaps of your analytics and BI, data science and ML, and self-service data preparation platform vendors, as well as emerging startups as they mature. Do so particularly in terms of the upfront setup required, the data preparation required, the types of data that can be analyzed, the types and range of algorithms supported, and the accuracy of the findings.
- Look for opportunities to use sandboxing and free trials to test and explore how augmented analytics complements existing data integration, analytics and BI, and data science initiatives.
- Recognize that these tools will mature and evolve over the next couple of years. Consult "Cool Vendors in Analytics," "Cool Vendors in Data Science and Machine Learning" and "Other Vendors to Consider for Modern Analytics and BI." Also monitor vendors' moves in "Magic Quadrant for Analytics and Business Intelligence Platforms," "Critical Capabilities in Analytics in BI Platforms," "Magic Quadrant for Data Science and Machine Learning Platforms" and "Critical Capabilities for Data Science and Machine Learning Platforms."

Acronym Key and Glossary Terms

Citizen data scientist	Gartner definition: A person who creates or generates models that use predictive or prescriptive analytics but whose primary job function is outside the field of statistics and analytics. The person is not typically a member of an analytics team (for example, an analytics center of excellence) and does not necessarily have a job description that lists analytics as his or her primary role. This person is typically in a line of business, outside IT and outside a BI team. However, an IT or BI professional may be a citizen data scientist if the professional's work on analytics is secondary to his or her primary role. Citizen data scientists are "power users" who are able to use simple and moderately sophisticated analytic applications that would previously have required more expertise.

Evidence

This document draws on Gartner analysts' research; surveys of vendors' reference customers, vendor briefings and hands-on testing of platforms conducted for 2018's Magic Quadrant and Critical Capabilities reports; and discussions with users of Gartner's client inquiry service.

[1] "The Rhode Island Heroin User Who Came Back From the Dead," GQ. 2016. Also, "Opioid Addiction Treatment Behind Bars Reduced Post-Incarceration Overdose Deaths in Rhode Island," Brown University. 2018.

[2] The Innovative Analytics in Action session from Data and Analytics Summit in Orlando, March 2019.