AI–Driven Analytics

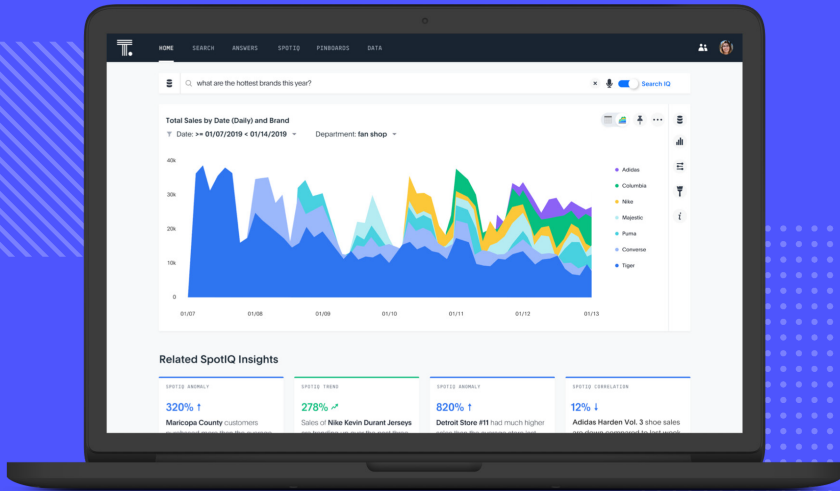## How Artificial Intelligence Is Creating a New Era of Analytics for Everyone

Sean Zinsmeister, Andrew Yeung & Ryan Garrett

REPORT

ThoughtSpot

Search & AI-Driven Analytics

Next-Generation Analytics Platform for the Enterprise
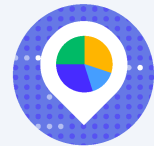

Simple

Designed for true
self-service for
everyone

Smart

Answers both
known and
unknown questions

Fast

Lightning fast
answers on 10s of
billions of rows

Insights at the speed of thought

Request a Demo

AI-Driven Analytics

How Artificial Intelligence Is Creating a
New Era of Analytics for Everyone

Sean Zinsmeister, Andrew Yeung,
and Ryan Garrett

AI-Driven Analytics

by Sean Zinsmeister, Andrew Yeung, and Ryan Garrett

Copyright © 2019 O'Reilly Media. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. *AI-Driven Analytics*, the cover image, and related trade dress are trademarks of O'Reilly Media, Inc.

Table of Contents

AI-Driven Analytics

Executive Summary

For hundreds of years, scientists and philosophers have dreamed of intelligent calculation machines that can perform work that is otherwise performed by humans. The advent, design, and development of computers moved this dream toward a reality, and in 1956, artificial intelligence (AI) became an academic discipline. But only recently has computing technology caught up to the scale of data and processing power to enable machines to intelligently "think."

Business intelligence (BI) has undergone its own evolution since the term was first coined. Beginning in the 1960s, enterprises used mainframes to support mission-critical applications such as reconciling the general ledger. In the 1980s and 1990s, BI software became an industry in its own right. In the late 1990s and early 2000s, new vendors emphasized usability and self-serve capabilities. Now, BI is being usurped by analytics software that uses larger scale and improved processing performance to enable search-based and AI-driven analytics capabilities.

For decades, AI was out of reach because the requisite compute scale and processing capabilities did not exist. Even when computational processing power advanced to adequate speed, costs kept AI development beyond the reach of many otherwise-interested parties.

Now in the age of big data and nanosecond processing, machines can rapidly mimic aspects of human reasoning and decision making across massive volumes of data. Through neural networks and deep learning, computers can even recognize speech and images.

The question for executives then becomes, "how can I implement AI to improve my business?" There are many advantages to using AI-driven analytics. AI can enable you to sort through mountains of data, even uncovering insights to questions that you didn't know to ask—revealing the proverbial needle in the haystack. It can increase data literacy, provide timely insights more quickly, and make analytics tools more user-friendly. These capabilities can help organizations grow revenue, improve customer service and loyalty, drive efficiencies, increase compliance, and reduce risk—all requirements for competing in the digital world.

Organizations dependent on traditional (pre-AI) BI increasingly struggle to meet these demands for two main reasons:

- Traditional BI establishes a publisher/consumer model in which a handful of well-trained specialists create reports and dashboards for potentially thousands of consumers. This creates significant bottlenecks. Business people end up waiting weeks or months for reports. And the minute a businessperson needs to dig deeper or ask a related question, the process begins again. In contrast, AI opens analytics to the entire population and can enable users to dig into and across datasets on their own.
- Data volumes are massive today. It is either impractical or impossible to hire enough resources to sort through all your data to uncover all of the valuable insights buried in it. And this challenge continues to grow more formidable. However, AI-driven analytics are powerful enough to scan tens of millions of rows of data and return interesting insights in seconds.

AI-driven analytics is already transforming a diverse group of industries, including healthcare, retail, financial services, and manufacturing. Though we are in the early days of AI-driven analytics, analytics infused with AI will generate greater benefits for the organizations that take advantage of this disruptive combination in their decision making.

The Origins of AI

For a concept and technology as game-changing and seemingly mystifying as AI, it can be a valuable grounding experience to take a few steps back to understand how we arrived at the capabilities of today.

The Evolution of AI

At its core, AI happens when machines use inputs to create a desired output or achieve a desired goal. Examples include the following:

- Amazon's Alexa understanding your voice and intent when you ask it to play a genre of music from a specific decade.
- Algorithms analyzing streams of machine data to predict when a component of the machine is about to fail. (Or, in the medical field, machines analyzing data from humans to predict serious medical issues.)
- A car's safety system scanning the environment around it to know when to slow down, change lanes, or stop backing up.

The idea of machines mimicking human intelligence has been around for hundreds of years, even in ancient Greek mythology. The field of AI research was officially founded when Dartmouth College held a workshop on the subject in 1956. Around the same time, computer scientists developed programs to compete with humans in checkers and chess. There was great optimism about the future of thinking computers, and governments poured billions of dollars into research around AI. However, the requisite computing power and scale did not exist at the time to turn such visions into reality.

In recent years, though, academics and engineers have made significant progress in both computational power and massively scalable data processing platforms. In this and the previous decade, founders have created thousands of companies to deliver AI-driven solutions, and large, established organizations have made AI an integral component of new and existing products. Now, AI is so ubiquitous in our daily lives that we seldom even notice it.

The Evolution of BI

BI, as we know it, also is relatively young. Organizations began to implement decision-support systems—the precursor to BI—in the 1960s, and these systems became an area of serious research in the 1970s, with academics and vendors investing considerably in the interactions and interface between the systems and users.

In parallel, many proponents of relational database systems proposed that these databases should be the platform for decision-

support systems. In fact, some experts have traced the common use of the term "BI" back to the mid-1980s, when Procter & Gamble hired Metaphor Computer Systems to build and integrate a user interface with a database.

BI would continue to be closely linked to data warehousing and relational databases in the following decades, though it would be many years before researchers and technology providers would connect AI and BI.

AI-driven Analytics

Today, AI is becoming a key driver of analytics. BI remains out of the technical reach of the average business person, and data volumes have exploded. When Teradata was born in 1979, most business leaders could never imagine amassing an entire terabyte of data. Today, many people store terabytes in their homes and the cloud. And we continue to create more data all the time with things as common as our phones, as well as with connected devices such as smart homes, cars, and planes and trains—the Internet of Things (IoT)—to name but a few data sources.

In a recent McKinsey analytics survey, nearly half of all respondents said "data and analytics have significantly or fundamentally changed business practices in their sales and marketing functions, and more than one-third say the same about R&D."

The challenge for traditional BI—in which data experts summarize and aggregate data from a data warehouse or data mart and then load it to a BI server for exploration and reporting—is that it cannot support the agility and deeper insights businesses require, nor the data volumes. Still, organizations recognize the need to be data-driven to keep up with existing competitors and fend off new digital natives.

This is where AI presents a significant opportunity. Thanks in part to the parallel explosions of data, affordable compute resources, and advanced algorithms, AI now can gather the amount of inputs necessary for it to make reasonable decisions and deliver the results of analyses in a timely fashion so that they are valuable.

AI-driven analytics can help users reveal insights in seconds in multiple ways. One example is the use of natural language processing (NLP). Analytics solutions with strong AI capabilities can under-

stand and translate queries such as, "What are sales for each category and region?" to identify the appropriate underlying data, calculate the sums, and visually present a best-fit chart, as shown in Figure 1-1. The user never needs to think about the rows and columns and calculations.
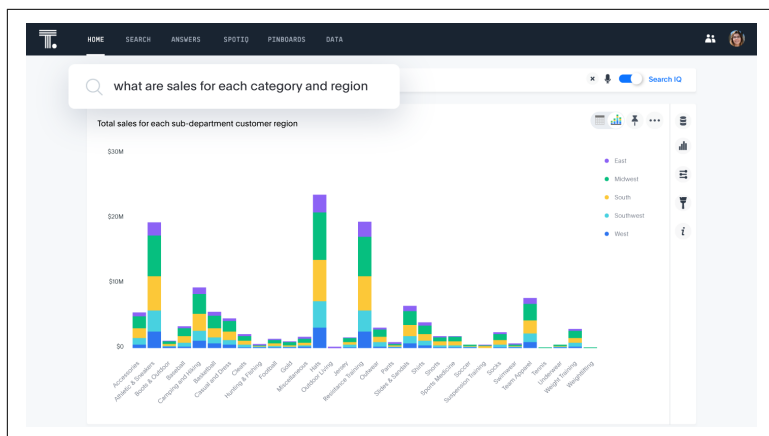

Figure 1-1. Modern analytic solutions support NLP to enable you to use everyday language to ask questions of your data

Automated analytics are another example of AI augmenting analytics to accelerate time-to-insight. In this case, a user can simply point their analytics solution at a dataset, a field, or even a specific data point and ask AI to identify key drivers of and anomalies within that data. Thanks to modern compute power and programming techniques, the AI can run thousands of analyses on billions of rows in seconds. Through natural language generation, the system can present the AI-driven insights to the user in an intuitive fashion—including results to questions that the user might not have thought to ask. With user feedback and machine learning, the AI can become more intelligent about which insights are most useful.

This notion of augmented analytics—applying AI techniques such as machine learning and natural language generation to analytics—presents such a disruption to the data and analytics market that industry thought leaders are encouraging their adoption. The opportunity is so significant that analyst firm Gartner, Inc. says that augmented analytics are "crucial for unbiased decisions, impartial contextual awareness and acting on insights".

Embracing AI Technologies

As with many new technologies, potential users and beneficiaries of AI must first consider whether to embrace it—and if they choose to do so, where and how to apply it. Fortunately, the technologies that enable AI are common and well understood, and the list of potential applications is broad.

AI Demystified

For AI—and its offshoots, machine learning and deep learning—to support real-world use cases requires massively scalable technology architectures. That's because AI is more "artificial" than "intelligent." AI requires massive amounts of data to train and learn so that it can deliver accurate (and relevant) results.

For example, consider a Google weather search. When you searched "weather" and some zip code or city seven years ago, Google would return links to multiple pages with current weather and forecasts for that locale.

Fast-forward to the present day. As soon as you type "weather" into your search bar, Google will return the current conditions based on your IP address. If you complete the search with a zip code, Google returns multiple details about current and forecasted weather conditions—and, depending on your search patterns, might also include links to relevant items like emails in your Gmail inbox that reference that locale, things to do there, and other interesting facts.

All of this is the result of Google's AI learning over multiple years and billions of searches what users are interested in when they search around "weather." Storing and processing all this information requires massive scale.

AI: uniting database and analytics technologies

Fundamentally, AI requires both database and analytics technologies that operate at massive volume and speed. AI requires significant storage to hold all of the data that its models require for training and learning. And AI needs analytics technology to do something useful with all that data, whether the end result is identifying a person by their face or predicting which products will be hot sellers in the next month. All of this must be combined with massive processing power to return results in a timely fashion.

Essentially, data is the crude oil of our digital economy. There is great value to be gained in data, but it requires very significant resources to turn massive volumes of dirty data into shiny insights. Data in its raw form is often useless. Like oil refinement, data refinement is difficult and expensive. As a portion of the population that can benefit from data insights, those who know how to process and analyze data are relatively small in number. And, like crude oil, there are millions of consumers waiting to use the completed data products.

This is where AI comes in—presenting accurate, relevant answers at the time that they matter to the business user.

AI requires an extremely tight integration between data storage and computation. Even though databases and analytics have long been closely connected (with database innovations often enabling new user interactions and analytic modeling techniques), there have been fewer efforts to jointly, inextricably develop storage, computation, analytics, and visualization together. Instead, enterprises have combined and integrated various components to build best-of-breed solutions based on their use cases (and existing vendor contracts and budgets).

In this paradigm, AI could never integrate with BI beyond the simplest use cases, as BI was not built for scale. Traditional BI relies on cubes and data aggregates loaded to a single BI server. The minute someone—or something, like AI—needs to learn more by drilling deeper into a detail that is more granular or outside the scope of the cube, the process breaks.

This is not to say that AI-driven analytics require every component and feature of traditional databases. But it does, at a minimum, require tighter integration between storage, compute, and analytics, along with a visualization layer or some other publication technique for the intelligence to be delivered in a timely enough fashion to be of value.

On a related note, in our always-connected world, we now expect that information always be available to us no matter where we are located or what we are doing. Therefore, the serving layer for AI-driven insights and results must be planet-scale. This was not possible prior to the widespread adoption of cloud technologies. Amazon Web Services (AWS), which holds the largest share of the cloud-based computing and storage market, is only a dozen years old.

Hence, AI-driven analytics is a relatively young, though already proven, technological advancement.

The role of memory

The evolving market for in-memory storage and processing also has played an important role in recent advances in AI-driven analytics. The second generation of BI tools were invented prior to the popularization of 64-bit computing and could only scale up to a few gigabytes of random-access memory (RAM). As the cost of RAM has decreased, enterprises are finding it more feasible to store and process increasingly large volumes of data in-memory rather than on less expensive but significantly slower disk drives.

"To become insight-driven or insight-centric, the goal is to get from data to analytics to action with a latency of only subseconds in the pipeline," writes Nadav Finish, CTO of GigaSpaces. "Businesses must advance beyond traditional analytics perspectives, which separate data inputs and transactional systems from the analytics systems."

Indeed, developers of memory-based, AI-driven analytics measure their code optimizations in nanoseconds—one billionth of a second. InfoWorld says that "nanosecond latency is at the bleeding edge of real-time computing," and "the value of time has never been higher and therefore speed has never been more critical to business applications."

Recent advances in AI

AI-driven analytics is a relatively young concept, but it is not the only area in which AI has made advances in recent years. Many organizations have actively embraced various forms of *machine learning*, the aforementioned subset of AI in which machines become progressively smarter or better at performing specific tasks. Essentially, machine learning is the use of algorithms for statistical analysis on input data to predict outputs. Machine learning is often broken into three categories: *supervised*, *unsupervised*, and *reinforcement*. Let's take a moment to look at each of these:

Supervised machine learning
    A data scientist or analyst provides both the inputs and a desired output, including feedback on the results to help the models "learn" so that they can make better predictions. The

expert iterates and the machine tweaks the models until there are ultimately no or very few wrong outputs.

A popular application occurs on social media websites in which users identify people in pictures. When a user loads a new photo, the site can make a very accurate suggestion of who should be tagged in the photo.

Unsupervised machine learning
Computers rely on deep learning similar to neural networks (rather than feedback from a data expert) to make their predictions. By looking at extremely large numbers of data points, machines can identify trends and correlations between variables on their own and then use this training to recognize new data points or make predictions.

Marketers use unsupervised machine learning algorithms such as clustering to identify similar groups of customers or prospects for targeted marketing campaigns.

Reinforcement machine learning
Machines take actions in an environment to maximize a "reward." This is typically done through a *Markov Decision Process* when there is no exact mathematical model of the environment and experts are not involved in providing the inputs or feedback on outputs. The goal is to maximize the reward based on existing knowledge while simultaneously acquiring new knowledge.

A popular example is that of a gambler with a row of slot machines from which to choose. Common applications include financial portfolio optimization, network routing, and clinical trials. Reinforcement machine learning is often applied in video games and robotics.

Many companies have invested heavily in *deep learning*, a subset of machine learning, which is itself a subset of AI. Deep learning and *artificial neural networks* enable image recognition, voice recognition, NLP, and other recent advancements. We have already come to take these for granted in our personal lives in the age of the internet and big data, but such features are hardly commonplace in analytics software.

Common AI algorithms used in analytics

Although AI-driven analytics is still too nascent to describe the algorithms behind it as "popular," there are algorithms that are becoming more widely used across the AI-for-analytics landscape. Let's examine a few of these here:

Linear regression

> *Linear regression* (Figure 1-2) models the response of a dependent to an independent variable or set of independent variables. The model is an equation with the dependent on one side and a weight for each variable on the other side. The equation can be used to generate insights on customer behavior or profitability.
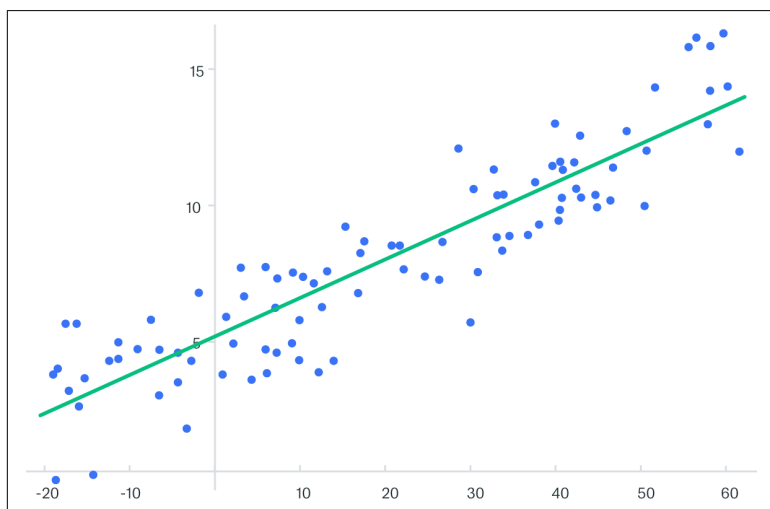

Figure 1-2. An example of linear regression

Logistic regression

> *Logistic regression* (Figure 1-3) is similar to linear regression in that it builds a linear model for an independent and a dependent variable. However, in a logistic regression, the dependent is binary—0 or 1, true or false, yes or no. It can be used for image segmentation and processing or categorical predictions.
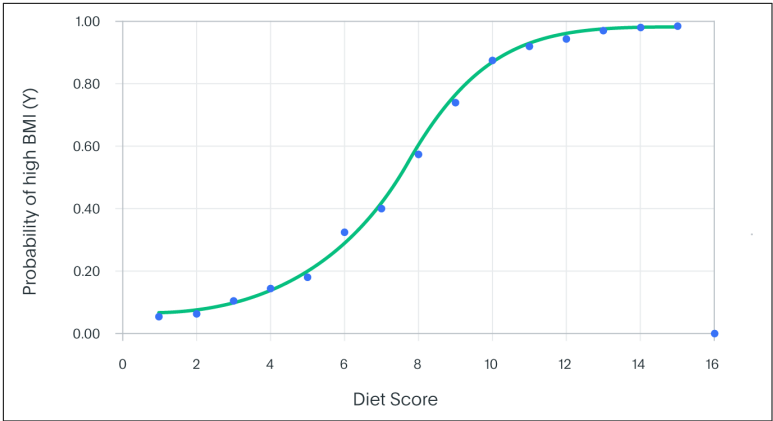
Figure 1-3. An example of logistic regression

Decision trees

    Decision trees are tree-like models of decisions and conse-
quences or outcomes, often with the likelihood of those out-
comes modeled as weights. They are popular in logistics, project
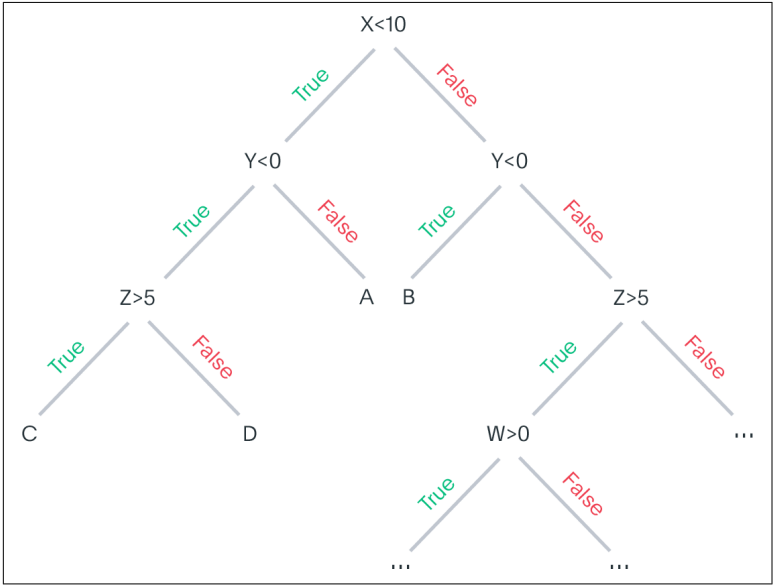management, health care, and finance. Figure 1-4 shows an
example.

Figure 1-4. An example of a decision tree

Naive Bayes Classification

> *Naive Bayes Classification* is a machine learning technique, shown in Figure 1-5, that assumes that features or predictors are independent of one another to calculate the likelihood that an item is classified into various categories. It is very popular in text analytics for use cases such as spam recognition and news category tagging.
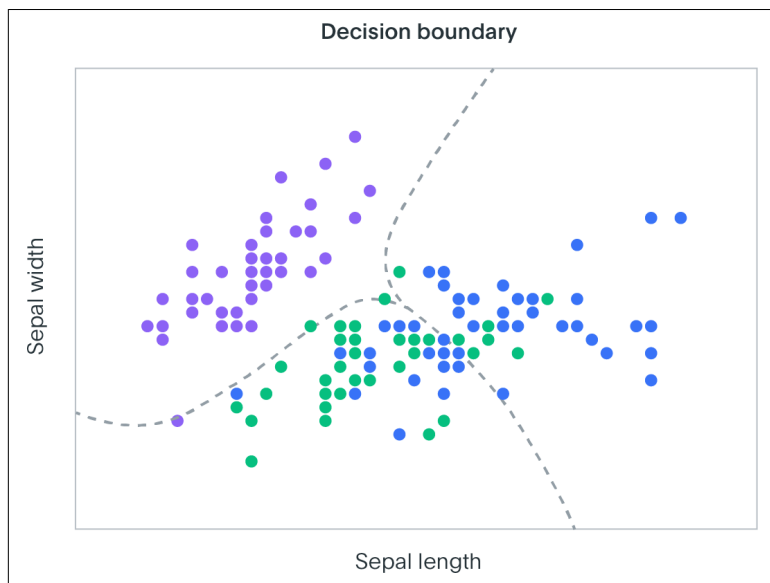

Figure 1-5. Results of a Naive Bayes Classification model

Clustering algorithms

> These types of algorithms attempt to group together items that are more similar to each other. *K*-means, depicted in Figure 1-6, is probably the most popular clustering algorithm. To begin, you select the number of classes or groups that you want to create and the centers of those groups. As the model trains, it will shift the center of the groups until ultimately it finds the center with the shortest distance between the members of its group and the farthest distance from members of the other group. This is a very fast method because there are few computations—you are only calculating the distance between data points and the center. Clustering algorithms are used in customer segmentation, bioinformatics, medical imaging, social network analysis, and web search.
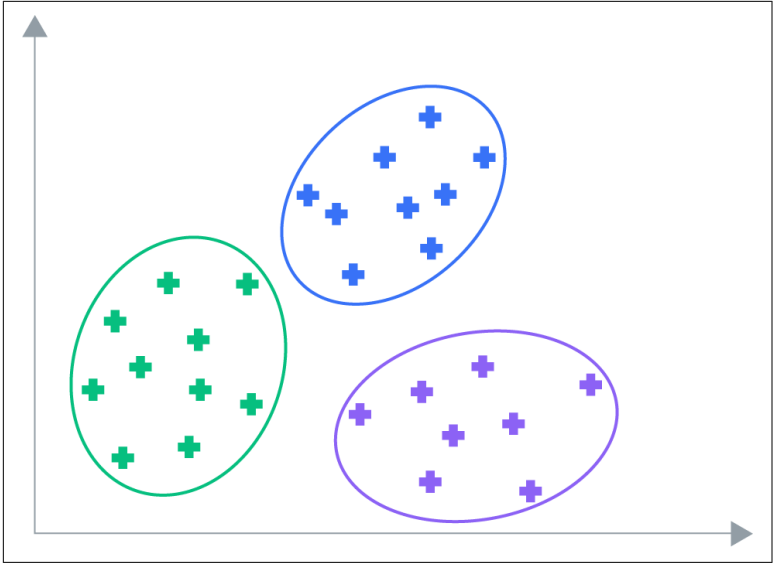
Figure 1-6. Results of a clustering model

Principal Component Analysis (PCA)

    PCA is most commonly used for dimension reduction. In this case, PCA measures the variation in each variable (or column in a table). If there is little variation, it throws the variable out, as illustrated in Figure 1-7, thus making the dataset easier to visualize. PCA is used in finance, neuroscience, and pharmacology.
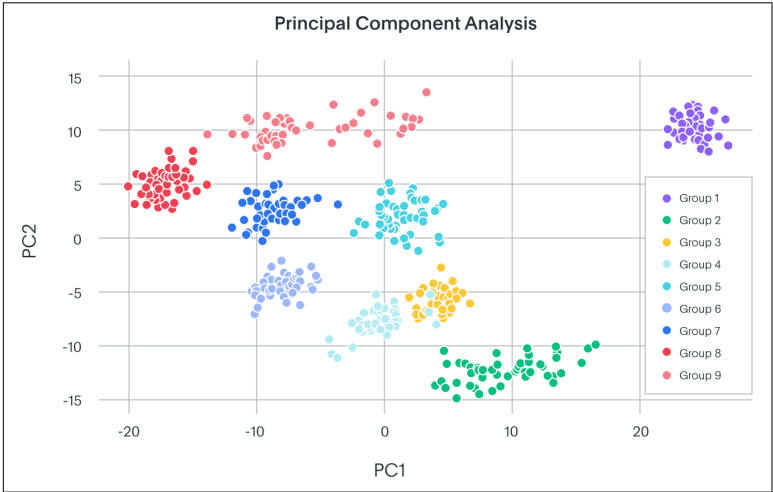

Figure 1-7. Results of a principal component analysis

Advances in processing

Though some of the algorithms outlined in the previous section require relatively few calculations, they will require massive volumes of data to effectively train. As discussed earlier, lack of significant volumes of compute power and data have historically been impediments to AI development. Large volumes of data are necessary for models to learn, and large compute resources are necessary to quickly perform any complex calculations.

Fortunately, developers have made significant advances in processing over the past few decades, and the cost to take advantage of these advancements has considerably decreased. Graphic processing units (GPUs) and RAM are now affordable enough that enterprises can incorporate them into mission-critical systems. Companies now store terabytes of their enterprise data in RAM for fast access and analytics. And thanks to massively parallel processing, software can intelligently combine the resources of many machines to multiply processing power.

This has led to new programming frameworks such as deep learning. In deep learning, computers learn how features are represented rather than relying on task-specific algorithms. Deep learning has risen to prominence in speech recognition, NLP, bioinformatics and other fields. This can be supervised or unsupervised.

Deep neural networks, perhaps the most popular subfield of deep learning, examine inputs and determine outputs. "Deep" means that a neural network has multiple hidden layers between the input and output, which deep neural networks use to learn features of the data to apply weights—the likelihood or probability of each output—to various factors to create a feature hierarchy. This requires both large amounts of data and large amounts of compute power. Through training, eventually the machine will settle on the appropriate weight for each feature and be able to correctly classify new inputs without any human assistance.

Implementing AI

With the technology that is available, how do you get started with AI?

One consideration should be your desire for customization and flexibility versus ease of development or use. For example, TensorFlow

is a free, open-source software library popular for its machine learning library. But for those looking for a quick path to get started or prove out a concept, Keras is a high-level neural networks library that acts as a wrapper to TensorFlow as well as Microsoft Cognitive Toolkit and Theano to enable fast experimentation. "Being able to go from idea to result with the least possible delay is key to doing good research," according to Keras' backers (as of April 2019).

Also, you should consider how AI and machine learning align with your cloud strategy. Various frameworks are more tightly coupled with the popular public clouds:

- Google Cloud Platform (GCP) provides tight integration with its Cloud Machine Learning Engine for "developers and data scientists to build and bring superior machine learning models to production."
- AWS provides pretrained services, frameworks for building, training and deploying machine learning, and open source frameworks for building custom models.
- Microsoft Azure offers a drag-and-drop authoring environment for machine learning applications.
- IBM Cloud includes Watson services for NLP, visual recognition, and machine learning.

Additionally, technology providers of solutions that take advantage of AI might choose to roll out features or even entire products on one or two clouds rather than all public clouds, based on a variety of strategic and market factors.

For many of the open source technologies, there are providers whose main purpose is to develop, implement, and service core AI technologies and their ecosystems. Databricks, for example, was founded by the creators of Apache Spark and provides an analytics platform powered by Spark for data science teams.

Transforming the way your organization operates with AI

Perhaps one of the most important considerations in implementing AI technologies is ensuring that teams have the skillsets to use something so advanced. In fact, 51% or organizations see staffing as their primary challenge regarding AI and machine learning, according to *Harvard Business Review*. In the public sector, AI is likely to free up significant hours—up to 30%—from mundane tasks such as paperwork. But it won't necessarily reduce the need for staff hours.

Rather, agencies likely will shift staff time to focus on tasks that require human judgement, according to Deloitte.

Fortunately, in the world of analytics, AI can actually reduce complexity. Thanks to advances in speed and scale as we discussed earlier, AI-driven analytics can traverse massive volumes of data and perform thousands of analyses to uncover outliers and other valuable data points and insights. But how does a machine distinguish between interesting insights and noise from the end user's perspective? Machine learning, the bedrock of AI, can learn what areas of a dataset are most commonly inspected by the user or their team in order to expose only the most relevant facts. Figure 1-8 shows that auto transactions account for the highest revenues. However, AI reveals the hidden pattern that in ZIP code 80017, dining transactions are 10 times the average.
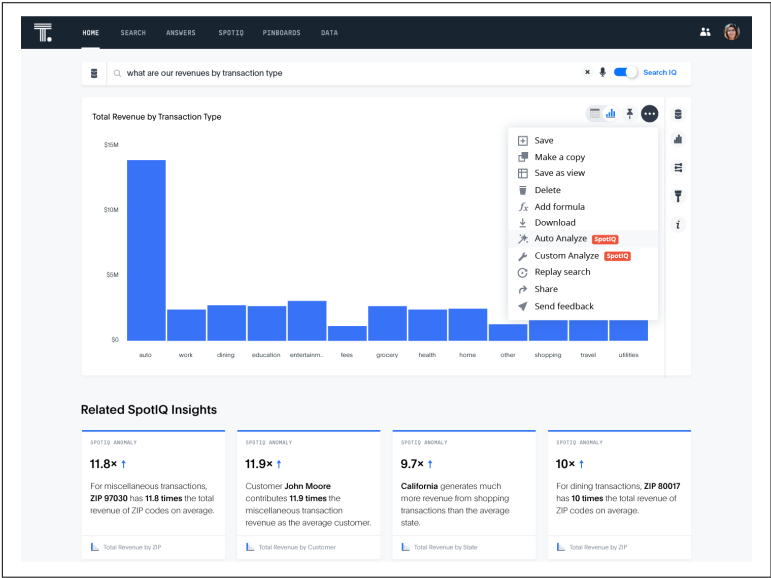
Figure 1-8. In analytics, AI can reduce complexity by exposing the most relevant facts such as the SpotIQ Insights

In addition, AI is improving data literacy across the organization. Historically, a data expert needed to understand the details of the data, determine the best way to visualize analytics results, and typically copy and paste those visualizations into a PowerPoint presentation, where they would explain what was being shown. AI-driven analytics solutions greatly reduce the need for human intervention

and intelligence throughout this process. Now, a user asks a question, the software determines how to best visually represent the results, and then the system uses natural language to explain those results.

In the traditional BI paradigm, business people struggled to get access to their data, but even if they crossed that hurdle, they still needed to know what questions to ask. AI can be an accelerator and educator for data literacy by helping business people know what to look for in their data in the first place.

This automation of analytics through AI can minimize the opportunity for errors and reduce human bias by presenting valuable insights to questions that users might not have considered asking in the first place. Because a machine can run thousands of correlation calculations on data points and even on the results of its own analysis, it can uncover key contributing factors that users have not yet considered.

By reducing complexity and the need for expert skills, AI-driven analytics can be a significant contributor to the growth of the new class of *citizen data scientists*. Gartner defines a citizen data scientist as "a person who creates or generates models that use advanced diagnostic analytics or predictive and prescriptive capabilities, but whose primary job function is outside the field of statistics and analytics."

"Citizen data science (CDS) fills the gap between mainstream self-service analytics by business users and the advanced analytics techniques of data scientists. Data and analytics leaders should use CDS to explore new data sources, apply new analytics capabilities and access a larger user audience," according to Gartner.

The opportunity for AI to transform your organization while enabling the new class of citizen data scientists cannot be overlooked.

Why AI for Analytics

It should be clear that the next evolution of analytics will be powered by AI. In the age of big data, static reports and dashboards no longer suffice to give us all of the insights we need to maximize the value of our data and stay ahead of competitors. AI is necessary to comb through the troves of data that businesses, customers, and marketplace forces constantly create, to present the insights that matter to our business in an intuitive manner.

Perhaps it is useful to consider all the benefits that AI provides for analytics through two lenses: efficiency and effectiveness.

We have covered many of the ways in which AI makes analytics more efficient. As highlighted in the previous sections, there is a shortage of data experts, which has contributed to the rise of the citizen data scientists. AI enables non-experts to benefit from complex analytics processes without knowing how to program every detailed component of the workflow.

Because AI-driven analytics can learn what is important to us, it can accelerate our speed to insights. Rather than slicing and dicing and drilling down for hours on end, users can simply ask the system to perform analyses and present the most relevant insights. With automation, we can even schedule such analyses. For example, consider a table in your database that is updated in near real time. You could schedule an AI-driven analysis to run every day or even every hour, and the system could then alert you if it spotted relevant changes or patterns in the new data.

Ultimately, all these AI-driven features make analytics and BI easier to use. Non-experts can use AI to conduct analyses that were once the purview of a handful of trained specialists. The experts can focus on higher-value tasks rather than wading through the backlog of requests that pile up in traditional BI scenarios.

Also, AI increases the effectiveness of analytics by revealing relevant insights without human interaction, as shown in Figure 1-9. AI-driven analytics increases analytical literacy and upgrades user skills, according to industry thought-leaders Wayne Eckerson and Julian Ereth of Eckerson Group: "AI-driven BI tools surface insights rather than force business users to hunt for them. With a click of a mouse, users can perform a root cause analysis of any metric in a chart or

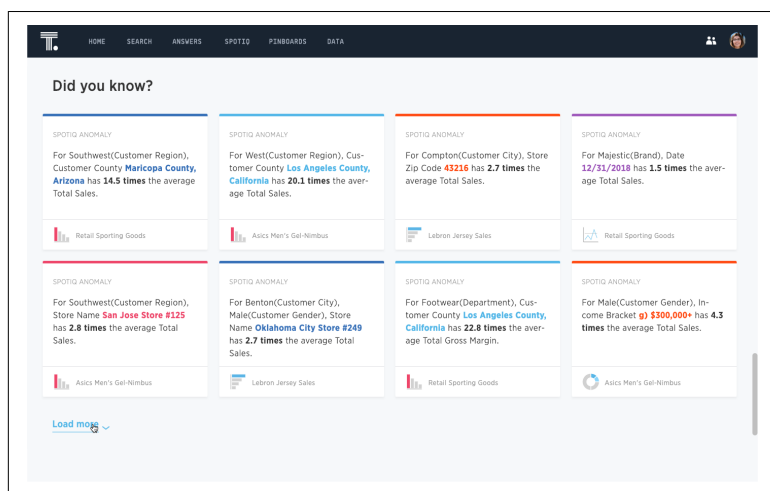dashboard and view related insights and reports. Some can even close the loop by recommending next steps and actions."

Figure 1-9. AI can increase analytics literacy by revealing insights without requiring human effort

As stated at the beginning of this section, the massive volumes of data available in our big data world render traditional BI solutions useful for known, simple questions, but ineffective for AI and revealing hidden patterns. AI-driven analytics solutions can highlight context for users to provide the full picture. Without AI, data volumes make it difficult to discover important insights that would otherwise remain buried in the data. With AI, analytics solutions can intelligently discover relationships between records and suggest new areas to explore.

Common Applications of AI in Analytics

Although it is difficult to overstate the potential for AI to reshape analytics, there are several applications for which AI has already developed a strong foothold:

Predictive analytics
  Predictive analytics use data on what has happened to predict what will likely happen next. AI is powerful here because it can examine massive amounts of historical data and test many possible predictions at the same time to find the best answer. This is important in marketing; for example, in use cases such as deter-

mining which offer to make next, how to price, and how significantly to discount. Also, AI-driven predictive analytics can help us predict churn so that we can in turn reduce it. One of the most popular use cases for industrial IoT is predictive maintenance, a specific type of predictive analytics. Here, systems collect streams or regular batches of data from machines and look for patterns to predict when equipment or components might fail.

Automated insight-generation

Insight-generation is a key application of AI in analytics. Rather than relying on a user to ask the right questions, AI-driven analytics solutions can traverse massive datasets at the click of a button—or via a scheduler—and find interesting insights on their own. With user feedback, as shown by the thumbs up and thumbs down icons in Figure 1-10, machine learning can help determine which insights are actually interesting and which are just noise to individual users and groups. Figure 1-10 shows the results of an in-depth automated analysis by transaction type with 21 uncovered insights.
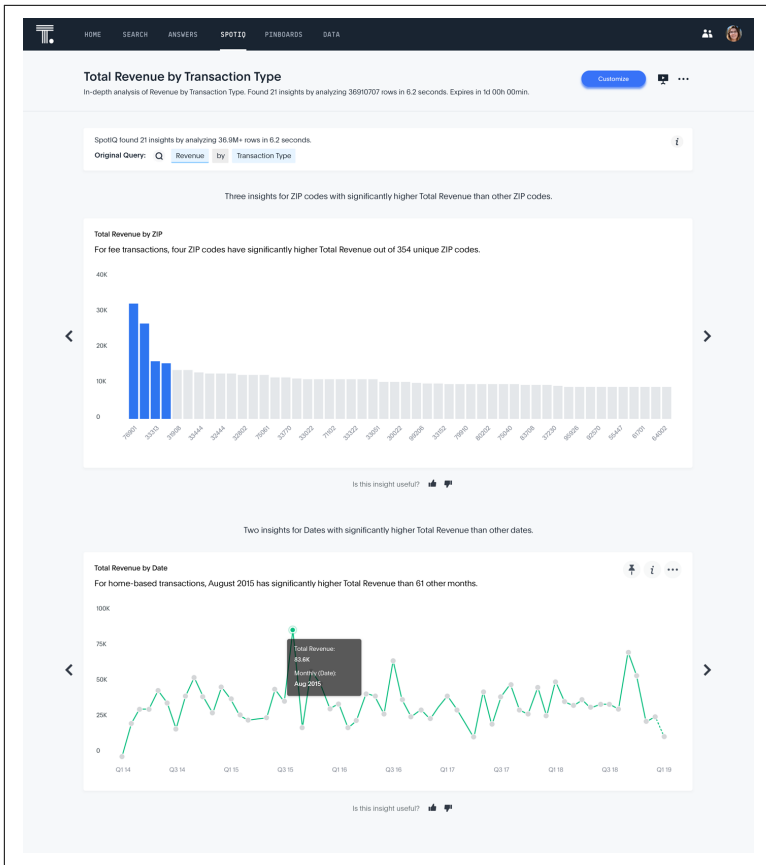
Figure 1-10. AI can generate insights on massive datasets and allow users to rate the relevance of the results

Natural search

Natural search relies on NLP, a field in AI. Historically, an analyst who was familiar with their data would write SQL to extract the necessary data from their data warehouse or mart and load it to their BI server. Then, they would log in to their BI interface and drag metrics and dimensions into a workspace, select a visualization to graphically represent the data, then massage and filter as necessary until they arrived at a means of answering their question. With natural search, analytics solutions can intelligently interpret human language queries such as "revenue by region by quarter last year." We can train NLP to be even more natural and conversational so that it understands, for

example, that "hottest items" means "top-selling products." In Figure 1-11, the user has searched for the "hottest selling nike products in new york," and the solution has interpreted this to identify the top Nike products by sales revenue and filtered to the state of New York.
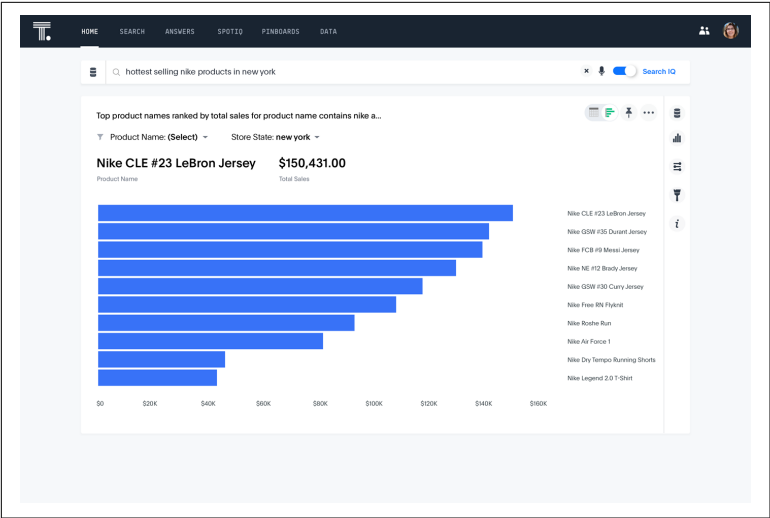

Figure 1-11. AI enables users of analytics to use natural language to ask questions of their data

Natural language generation

Natural language generation lowers the barrier to entry in analytics for nonspecialists. With natural language generation, solutions can explain analyses and results in a human manner. Rather than requiring a user to examine a table or visualization, AI-driven analytics solutions can simply state insights such as "Brand A sales are 23% higher in Cleveland than Columbus" or "Michael Jordan jersey sales jumped 15% year-over-year in Charlotte." Figure 1-12 demonstrates natural language generation that is easy for nontechnical business users to understand.
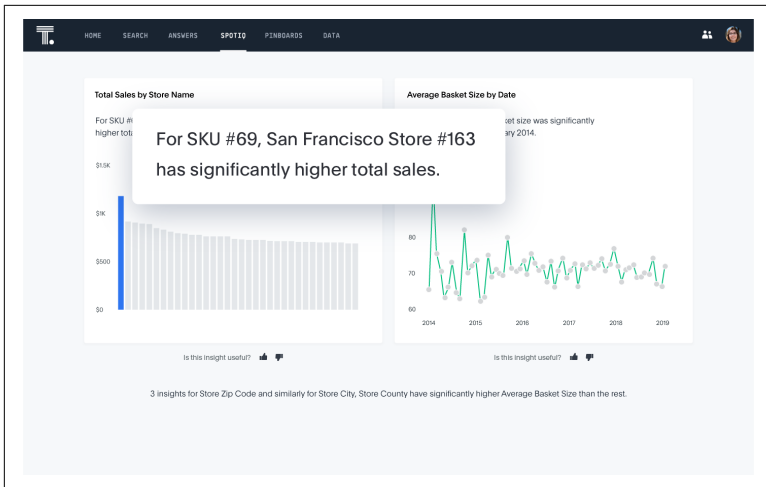
Figure 1-12. AI can generate natural language so that analytic results are easy for non-experts to understand

Self-driving analytics

Self-driving analytics take advantage of AI to provide users with more valuable outputs without requiring more inputs. For example, Figure 1-13 shows how an end user can simply "watch" an analysis in which they are interested, and the system will then regularly perform that analysis as new data enters the system, to monitor for anomalies and new trends. Like other advances in AI, self-driving analytics require massive scale to regularly take snapshots of data to actively track business-critical metrics. First, AI-driven analytics creates a time-series view of these metrics. It then regularly looks for significant changes. When the system detects something valuable, it alerts the user. The user can simply indicate the business areas they want to monitor, and from there, the system and AI take over.
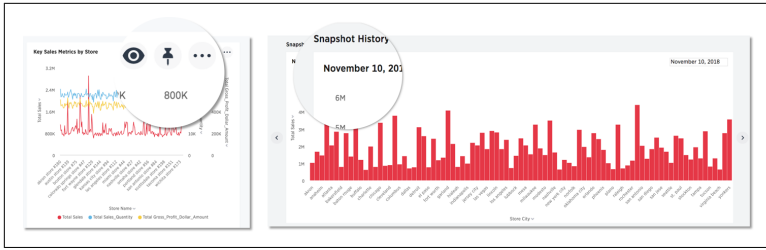
Figure 1-13. AI enables users to get significant, relevant analytic insights with very little human input

Data wrangling and preparation
> Most of the analytics life cycle involves preparing data for analytics. There are many tasks that can be involved, and they often require a skilled analyst or data scientist. However, machine learning and deep learning can help systems learn to handle some of these tasks independently, freeing up end users' time to focus on analytics, insights, and taking action.

These are a few of the areas where AI is becoming core to analytics technology.

Diagnostic Versus Predictive

When first introduced to AI-driven analytics, most people tend to jump right to the potential for predictive analytics. When it comes to making predictions, AI can help tell us "what is likely to happen next" by using the power of machine learning to select and accelerate the testing of predictive algorithms on historical data. However, there are far more potential use cases for front-line knowledge workers surrounding the diagnostic capabilities of AI-driven analytics—"why did something happen?"

Because data is constantly moving and growing, augmented analytics is crucial to more quickly diagnose root causes and understand trends. An analyst may need hours, days, or even weeks or months to develop and evaluate hypotheses and separate correlations from causality to explain sudden anomalies or nascent trends. With AI-driven analytics and machine learning techniques, they can sort through billions of rows in seconds to diagnose the indicators and root causes in their data, guiding and augmenting their work to deliver consistent, accurate, trustworthy results.

By leveraging AI-driven diagnostic analytics, you get valuable insights on the current state of the world faster. This yields a competitive advantage to businesses trying to stay ahead in dynamic marketplaces, enables medical researchers to save lives sooner, and provides the public sector an opportunity to more immediately optimize the allocation of resources.

AI-Driven Analytics in Practice

Several industries have rapidly adopted analytics solutions that embed AI directly into the systems. Retailers, financial services providers, manufacturers, and technology companies are all taking advantage of these new technologies.

Retail

In retail, merchandising teams have been some of the heaviest adopters of AI-driven analytics.

With traditional BI, business analysts at Haggar Clothing had their exploration restricted to the data contained in cubes and aggregates. They could not explore across subject areas such as inventory and sales, limiting their insights, and the static, canned reports they used couldn't keep pace with dynamic customer and store conditions. With their AI-driven analytics solution, they point the system at a specific dataset, product ID, or store ID, and the system will drill down to identify the best and worst selling items at any point in time so that Haggar can optimize its fixtures to improve gross margin. This led to a 5% increase in fill rates on top SKUs, resulting in a $10 million annual revenue increase.

> **NOTE** By optimizing store fixtures with AI-driven analytics, Haggar clothing is increasing fill rates on top SKUs by 5% for an additional $10 million in annual revenue.

On the business-to-business side of retail, De Beers Group sells diamonds to roughly 100 buyers about 10 times each year. Before adopting an AI-driven analytics solution, the pricing team at De Beers used a web of linked spreadsheets to determine the price for each diamond based on its characteristics and historical prices. The spreadsheets took minutes to open before the pricing team could

even begin calculating prices, and the team couldn't explore questions outside the spreadsheet. For example, it could not determine whether a buyer had rejected similar diamonds in the past. With its modern analytics solution, the team can automatically investigate a range of pricing scenarios to determine the optimal price for each diamond.

Back in the consumer world, merchandisers at a popular convenience store chain had very little access to point-of-sale data and loyalty data in its Microsoft Azure data lake with its traditional BI solution. With its AI-driven analytics solution, it is not only privy to this data—it is asking natural language questions like, "Who is signing up for the loyalty program but not using it?" "How are points being redeemed in various regions?" and "How frequently are different segments using the loyalty program?" Figure 1-14 shows an example of a natural language query and results that would be of interest to retailers. Merchandisers at the convenience store chain used insights from these types of questions to optimize loyalty incentives for a popular sports beverage, increasing the value of baskets containing the beverage by 21%.
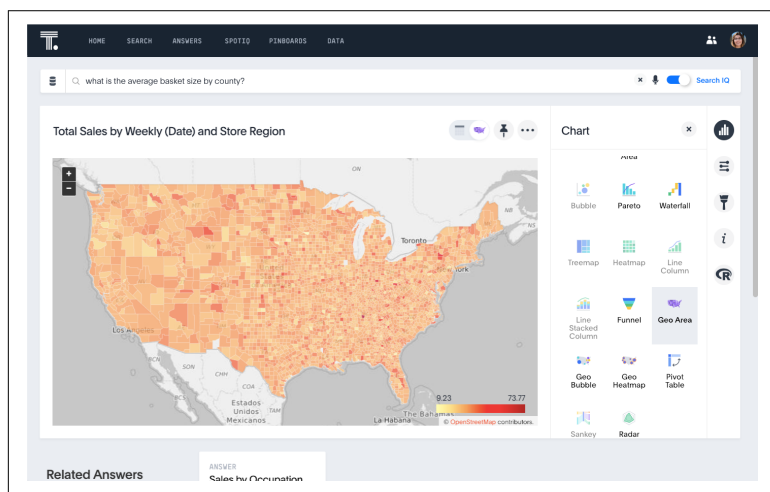

Figure 1-14. Retailers use AI-driven analytics to get fast insights that help them immediately improve revenues by tens of millions of dollars

With insights from AI-driven analytics, a popular convenience store chain increases the value of specific baskets by 21%.

Retail sales teams also are adopting AI-driven analytics. The Cellular Connection (TCC) operates more than 650 Verizon Wireless reseller stores. With its traditional BI solution, the 2,300 store associates requested an average of 14 reports per week that required three IT individuals to spend a combined 90 hours each week to build. After TCC selected an AI-driven analytics solution, those three individuals are working on much more strategic initiatives. Frontline store associates are able to use AI to answer their own questions, and they no longer rely on IT experts for new reports.

Financial services

Financial services providers face incredibly stiff competition in addition to very complex regulations. One focus of financial services regulations is transparency. Regardless how intelligent AI is or becomes, it must be auditable to stand up to regulatory scrutiny.

Fortunately, some technology providers that embed AI capabilities into their analytics solutions have taken this into consideration. When users take advantage of such AI capabilities, they are able to see exactly which types of analyses were run on which specific datasets and points.

At one of the world's largest investment management companies, sales and marketing teams measure and predict operational transactions associated with mutual funds sold by providers like Broadridge, so it needs visibility into a variety of data, including customer, asset, performance, and risk assessment data. The company's traditional BI solution, though positioned as self-service, was too complicated for business users, who waited days for the BI team to fulfill each report request. The IT team had to spend hours prepping, blending, and joining data for analysis. With AI-driven analytics, the IT team easily sets up all of its data sources in one place, and anyone can use AI for ad hoc analysis in seconds across all of the company's data. Figure 1-15 shows an example query that associates might investigate. The company has cut time-to-insight by 50%. Sales and marketing teams now have instant insights, allowing them to create targeted marketing campaigns and increase sales.

One of the world's largest investment management companies cuts time-to-insight by 50% with AI-driven analytics.
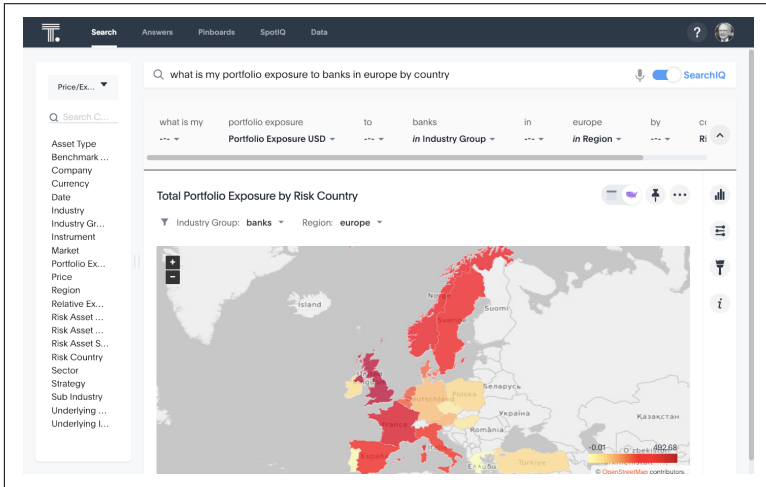

Figure 1-15. Financial services organizations get a complete picture of their business and customers via AI-driven analytics

Similarly, data analysts at Royal Bank of Scotland were bombarded with report requests. The bank estimated that 1,300 people are involved in reports and dashboards each day. The analyst team wanted to enable business users to perform analytics in a self-service manner, so it adopted an AI-driven analytics solutions. Now, managers can use AI to easily answer for themselves how individual team members are affecting key performance metrics such as customer net promoter score (NPS).

Manufacturing and high tech

Although they might at first seem like an unlikely pair, manufacturing firms and high-tech companies are both adopting BI solutions with AI capabilities to improve their procurement analytics.

Mondelez International used traditional BI for its reporting needs, and this sufficed for its global megabrands that had a plethora of operational and procurement support resources. However, the teams supporting its smaller brands lacked an easy way to access

procurement-related information. With AI-driven analytics, all teams have self-service access to insights on procurement data. They use AI to answer a variety of questions around purchase orders, such as which regions have the most open purchase orders for each brand; what brands have the largest total volume of purchase orders; and which purchase orders are not yet fulfilled.

Daimler AG's procurement system had a custom-built reporting interface that left the procurement team dependent on power users —business users could not create reports themselves. And the system was slowing as data volumes grew. At the same time, the procurement team needed to address increasing complexity and risk. As the amount and distribution of suppliers grew, the procurement team needed to manage risk of the entire supply chain in case of unexpected events. It also needed to have a flexible 360-degree supplier view prepared, in case of short notice opportunities to optimize procurement performance. With AI-driven analytics, the procurement team gains insights into all of its data without adding resources. The system scales to billions of rows of data and thousands of users. Figure 1-16 shows an example of a pinboard that a procurement team might create to answer multiple questions about purchase orders.
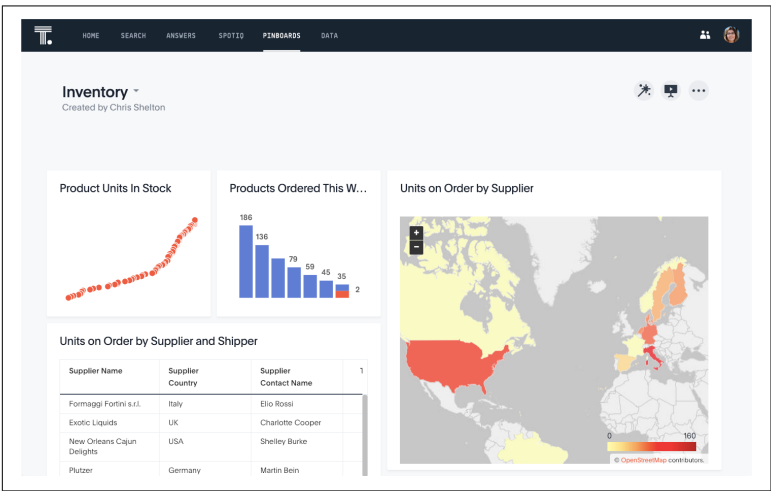

Figure 1-16. Manufacturing and technology companies use AI-driven insights to put their procurement teams in the best possible position for negotiations and ensure timely shipments from suppliers

NOTE Daimler AG looks to AI-driven analytics to manage risk in its supply chain and optimizes procurement procurement performance.

Conclusion

Although some recent notions and popular depictions of AI caused some to fear that machines would take over the world, at worst—or our jobs, at a minimum—the facts and examples highlighted in this report show that AI is truly a powerful enabler both for analytics and the knowledge workers who use them.

In fact, many of the companies who are broadly adopting AI are using it much more frequently in computer-to-computer activities rather than in automating human activities, according to *Harvard Business Review*. "'Machine-to-machine' transactions are the low-hanging fruit of AI, not people-displacement," writes HBR.

The benefits of AI-driven analytics are many, as we've covered here. Data analysts have more time to focus on deep data insights (that they might not have uncovered previously) rather than data prep and report development. Decision makers can explore much deeper and faster than they were ever able to with predefined dashboards.

Perhaps the most important ingredient to adopting AI as part of your analytics strategy is *trust*. For AI-driven analytics to gain people's trust, there are three key considerations:

- Accuracy
- Relevance
- Transparency

These are paramount concerns if business leaders are to make decisions and take action based on the results of AI-driven analyses. We have discussed how technology providers are addressing each of these concerns. With these bases covered, expect to see more and more companies adopting analytics strategies that take advantage of artificial intelligence as a significant component.

Are you interested in exploring how AI-driven analytics can enable your organization's digital transformation by driving more value from your data so you stay ahead of competition? Contact Thought-Spot.

About the Authors

Sean Zinsmeister is the head of product marketing at ThoughtSpot, where he is responsible for architecting the messaging and go-to-market strategy for ThoughtSpot's search and AI-driven analytics platform.

Prior to joining ThoughtSpot, he led product marketing at Infer, the leading predictive analytics platform for enterprise sales and marketing teams, through a successful acquisition in 2017. He's widely regarded as a thought leader in data, analytics, and AI, with his works appearing in publications like *VentureBeat*, *Forbes*, *Information Age*, and more. Sean holds advanced degrees from Suffolk Sawyer School of Business and Northeastern, respectively, in strategic marketing and project management.

Andrew Yeung is senior director of product marketing at ThoughtSpot. A technologist at heart, he is a 15-year marketing veteran with a passion for bringing disruptive innovations to market at both large public companies and high growth early-stage startups. Prior to ThoughtSpot, he held senior leadership roles in product management and marketing at ClearStory Data, CA Technologies, BEA Systems and Oracle. Andrew received his MBA from the Haas School of Business at the University of California at Berkeley, a MS in electrical engineering from Stanford University, and a BS in electrical engineering from Cornell University.

Ryan Garrett is a seasoned marketer in the data and analytics spaces. His goal is to help organizations derive value from data by making analytics more accessible, repeatable, and consumable. He has a decade of experience in big data and analytics at companies large and small, an MBA from Boston University, and a bachelor's degree in journalism from the University of Kentucky.