EXPLORIUM

AI is Making BI Obsolete, and Machine Learning is Leading the Way

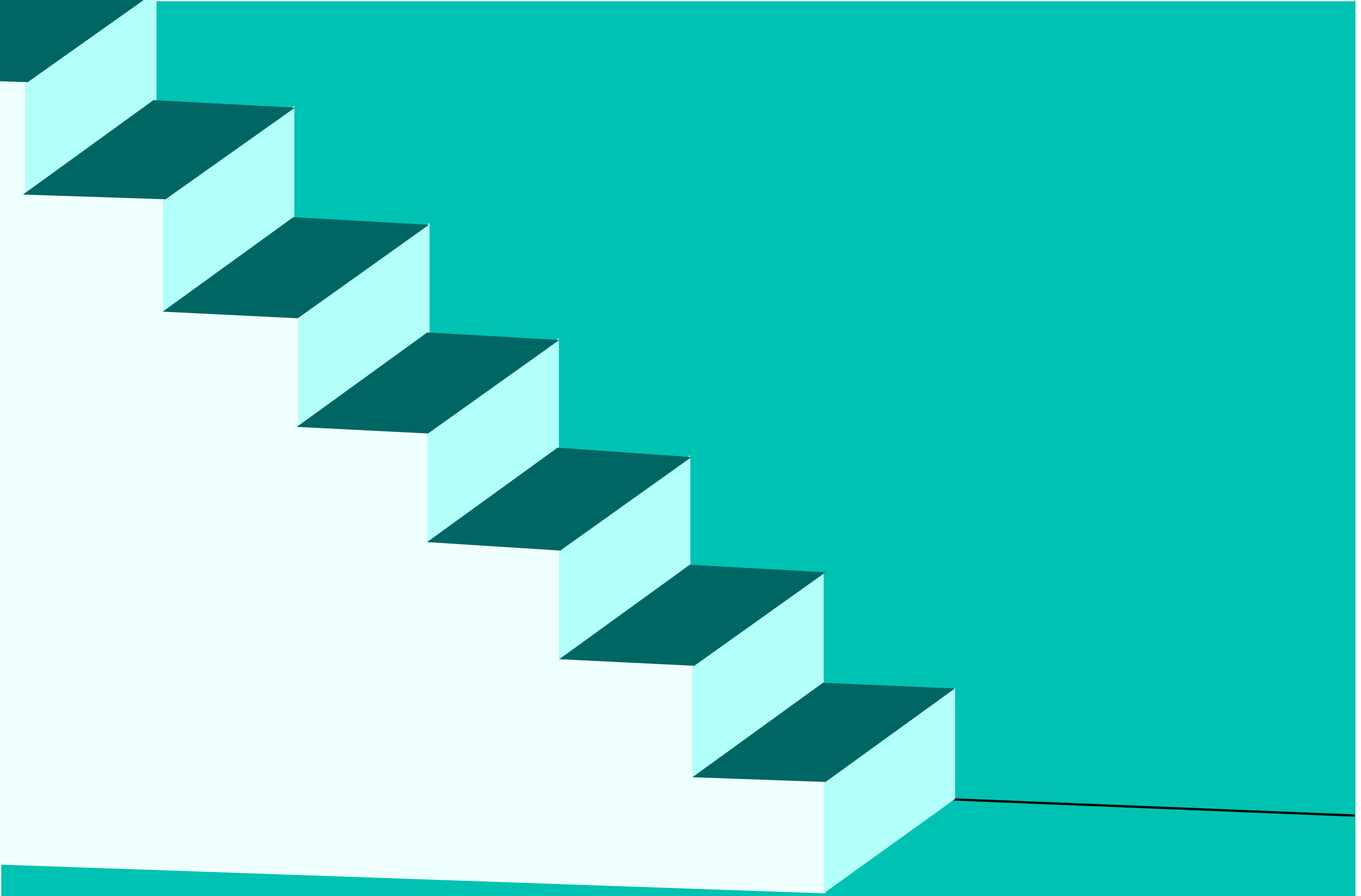
Evolution usually happens in one of two ways — it's either a slow crawl over a long period of time, or it's a bolt of lightning striking the ground. Technological innovation isn't much different, coming either in incremental improvements such as new versions or with game-changing inventions. Usually, the former follows the latter, but each massive step can mean the previous standard quickly becomes obsolete.

The emergence of machine learning (ML) models is giving us the power to extract significantly more and better insights from our data than ever before.

The early part of the past decade saw the rise of data-driven technologies and insights, going from basic analytics tools to the significantly more powerful business intelligence (BI) suite, which boasts the ability to aggregate an organization's data and display it in an easily digestible format.

The massive leap in analytics capacity meant that data became a valuable asset, and organizations are happy to invest millions into applications and tools if it means they can use it successfully. In the ensuing years, changes were incremental — basic visualization gave way to dynamic visualization, static BI gave way to real-time insights, and analytics chugged along.

Now, another lightning bolt has struck the ground, and BI may be looking at the face of the innovation that could make it obsolete. Data science isn't a new field — the term has existed since 2001, and the concepts behind it anywhere from 50 to over 100 years. However, automation has made data science exponentially more valuable, and the emergence of machine learning (ML) models is giving us the power to extract significantly more and better insights from our data than ever before.



Climbing uphill with BI

BI has become a mainstay of most organizations today thanks to its ability to convert massive raw datasets into easily digestible reports and visual dashboards. For companies looking for simpler predictive capabilities — extracting meaningful patterns, deriving actionable insights, and making faster decisions — BI provides an excellent and easy-to-use solution.

Additionally, BI tools are largely human-reliant. They may organize, visualize, and display your data in useful models, but the heavy lifting relies on your eyes and capabilities. Most importantly, business intelligence is focused on answers: solving those things you know you don't know.

But, what happens if you start without knowing the questions you need answered? We could certainly try to manually try to tease out better insights by working out the questions to ask based on our data. At a certain point, though, it simply becomes unfeasible with the toolkits most BI suites offer. On the other hand, machine learning is perfectly suited to the task of finding the right questions to ask. At this point, it's normal to wonder, well, aren't the patterns both humans and machines look for the same? The answer, predictably, is yes and no.

Machine learning - better questions for smarter answers

Like we mentioned above, BI is about answering known unknowns — the questions to which we don't have answers. This is great, but it limits how broadly we can use BI outside of its narrow scope. You can keep trying to climb uphill by manually looking for the right questions to ask, or you could build a machine learning model that does it in a few minutes, automatically.

Machine learning algorithms are mathematical models that leverage historical data to uncover patterns that can help predict the future to a certain degree of accuracy. When it comes to running a business, the ability to predict and make data-driven decisions (from the ability to identify customer churn before it happens to optimizing promotions and predicting loan defaults) based on those models is an invaluable tool to have in your arsenal.

> **Machine learning models, on the other hand, can infer more complex rules, deeper patterns, and interactions between different dimensions and different variables.**

Whereas BI may help us visualize historical data, we do so through a pair of human eyes, giving us a shallow understanding of the data given the cognitive bias we all have and the limited number of dimensions we're capable of looking at. Machine learning models, on the other hand, can infer more complex rules, deeper patterns, and interactions between different dimensions and different variables.

So, what if we analyze the model itself and not the data that it learns from?

Imagine it's your first day as a top executive in an industry with which you're not fully familiar. Even so, time is money, and in your first few hours you're already asked to make a crucial, time-sensitive decision. You don't have a real frame of reference, so you ask your researchers to get you a mountain of books and research to find the background you need.

This is a great start, but most of these books aren't specifically about your problem, so finding answers is like sifting through sand to find a few nuggets of gold. You can't read all the information and compare it to other data to help you make the decision in this short time.

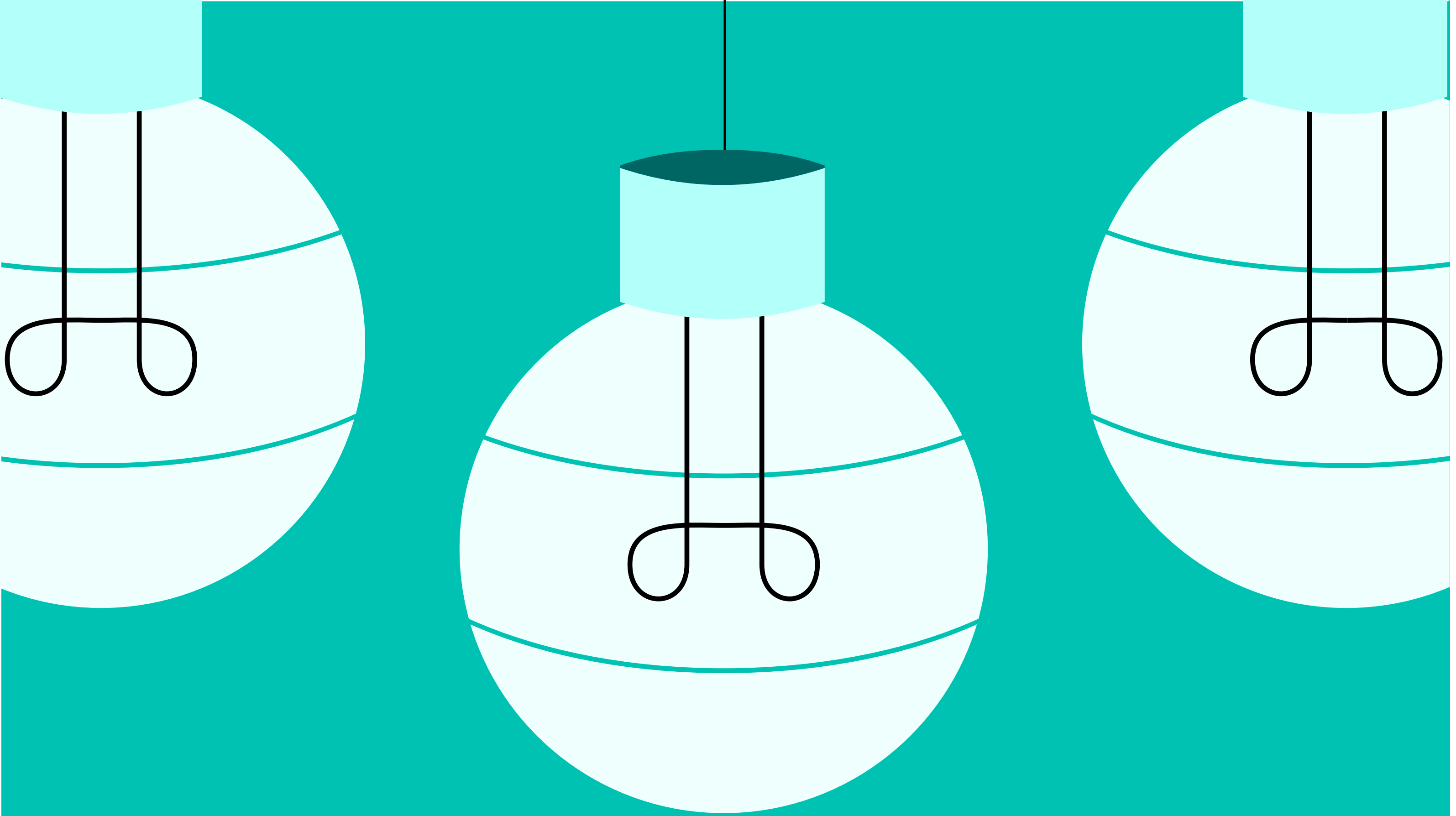
Think about your data-gathering process for a minute. When you're looking at your monthly revenue drivers, do you look at your CRM to identify clusters of good and loyal customers? Or, do you look to see if certain holidays and events cause peaks in sales? Maybe it's those marketing campaigns you're running. It isn't that we don't have the sources to find answers, but rather that we have too many possible sources to manually sort through them.

Let's go back to your hypothetical new job for one more minute.

You've been struggling to try to make sense of this pile of data, but what if you had a specialist to help you? What if instead of wading through mountains of documents, you could have someone extracting only the relevant information — an expert who has studied hundreds of documents before and knows exactly what to look for in every dataset? More importantly, what if on top of knowing what to look for, they can show you how each piece of the puzzle connects, how they interact, and what under-the-surface knowledge you can get from them?

If BI gives you hundreds of data sources and reports to look through, machine learning models take that a step further. ML algorithms give you the right data to focus on, but they also provide you with the most effective way to ask questions about it and derive valuable insights automatically.

Most crucially, ML improves your predictive capability by learning patterns that are most likely to explain a phenomenon and applying them to future problems. Hence, ML models give you not just the answers but show you which questions to ask. Once you need to go beyond the historic and surface levels of your analysis, BI becomes increasingly less valuable and capable.

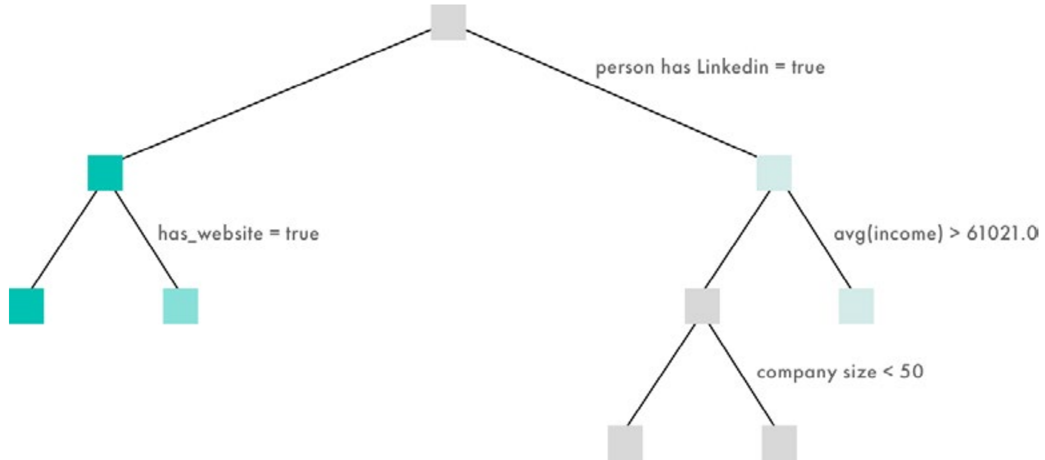


How can ML give us insights?

Without diving too deep under the hood, let's look at an example that can help illustrate this evolution.

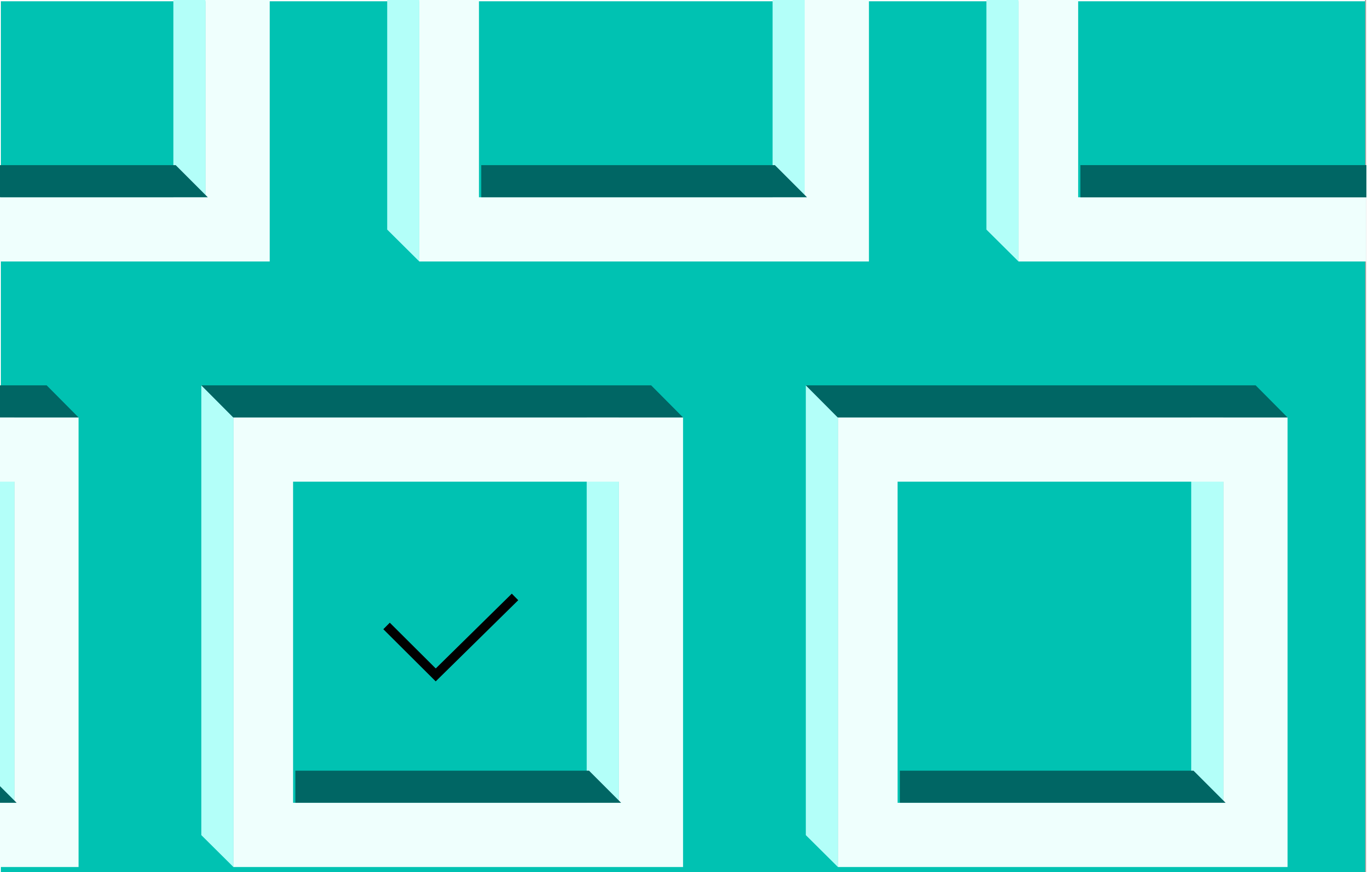
Let's say you're an online lender that has been having trouble planning ahead because it's hard to predict, month to month, which customers are likely to take a loan out in the next weeks, months, and even year. Your BI data can show you important historic data — people are less likely to take out loans during X month, or are more likely to apply on a Monday after a payday. These are great features, but they're still somewhat basic, may not give you all the answers you need, and might have taken you hours of looking to find.

Now imagine we want to go beyond these basic features. Maybe X month is a great month for loans for different reasons, or maybe your prediction that Mondays after payday are also loan days is not quite as accurate as you thought. We could start by building a decision tree, which creates and analyzes features that could better predict whether someone will or won't take a loan.



Starting from a simple question, we can break down a variety of possible factors that could indicate a loan taker, making it easier to determine who will borrow from you in the future. In the tree above, we can see that the ideal borrower has LinkedIn, an average income of over $61,000, and a company smaller than 50 people. Just like that, we've created a much better prediction tool to help us target customers. It sounds like alchemy, but it's rooted in data science.

So why are people still using BI more than the modeling alternatives?



When you do (and don't) need machine learning

All that said, building a predictive ML model is not always the most feasible option.

For one, you'll likely have to hire a data scientist or specialist who can build relevant models. Moreover, results aren't instant — building models and training them takes time and costs money for some time before it gives you real value in return.

Why?

Because building machine learning models that actually learn and make accurate, robust predictions with the right data is an involved process. You can't just take an algorithm, point it to the relevant data table and say *"Go do your magic thing with the patterns."*

For instance, feeding your ML model a massive raw dataset may give you suboptimal results. Instead, you should ideally spend some time preprocessing the data (scaling numerical columns, converting categorical variables, harmonizing your data), choosing the right algorithm, and calibrating it. Fortunately, this whole process doesn't have to be manual thanks to tools that can automate the modeling work. This is especially true if you're working from smaller, single-source data sets.

The reality is that if you're looking for insights, and you have data that is constantly streaming in from multiple sources, investing in ML can give you a better long-term ROI.

Even so, the real world is rarely so kind, and we're rarely dealing with single sources or perfectly prepared datasets.

What happens when you want to get insights derived from multiple data sources? What if our online lender, instead of simply looking at financial data, wants to understand how different locations may affect borrowing. Or, maybe they want to see how Facebook interactions translate into potential clients.

These queries require you to combine data from various places and mean that your ML model has to aggregate these sources into a unified and coherent output. They also raise significant questions, as well as complications. If you're simply looking for basic insights, this is the point where you may find that ML is too much. This is the sweet spot for BI — where you need quick insights, but don't require heavy prediction capabilities.

You may find that a dashboard is all you really need, and it may even fit your budget. However, the minute you need something deeper, or with powerful prediction capabilities, you'll be back at square one. The reality is that if you're looking for insights, and you have data that is constantly streaming in from multiple sources, investing in ML can give you a better long-term ROI. This is more so the case if you focus on not just building automated ML models, but if you automate your data science in the process.

Automate your data

Most tools and platforms built for data scientists in the past few years have focused on perfecting the algorithmic layer — selecting and tuning the right models. This is valuable, to a degree, but gives you diminishing returns after some time.

Instead, the new focus should be on automating data discovery and feature generation, which can simplify data discovery, aggregation, and preparation. These new algorithms can quickly generate and test features, tune your ML pipeline, and create models using optimal features to provide more accurate results.

Unleashing an ML model that can explore thousands of data points and millions of possible features while aggregating data can redefine how we view insights across siloed data sources. An ML model that prioritizes data over algorithmic improvements can help you speed up your time to insight.

In a few seconds, our online lender could quickly discover that:

- → Customers between the age of 17-24 (CRM) who are college-educated (professional profile enrichment), and have seen the video-ad (online marketing data) are a key driver to your revenue
- → Small businesses in the beauty industry (e.g. hair salons) who have a lot of competition in the area around them (location-based enrichment) and low reviews in online business reviews websites like Yelp (business web-based data enrichments) are at higher risk of defaulting on a loan
- → Stores with a 1+1 promotion (historical promotion data) located in an area with a higher percentage of young families (census and demographic data) are selling more products and are at a higher risk to experience out of stock

You could have found the same results with your BI, but it could have taken hours or even days. More importantly, your inability to connect seemingly random datasets and data points (as compared to ML tools) means that you could miss out on valuable features that could be crucial.

Automated insights could mean breaking the human barrier of thinking. Nowadays, when we use our BI tools, we're basically searching where we already know we will find answers. Instead, prioritizing the right questions and predictive models could disrupt our way of thinking. Instead of a highly commoditized BI industry, we could be looking at a smarter way to generate insights that's faster and broader. In a few short years, we'll see a tectonic shift from BI to AI, with machine learning leading the way.