

# THE DATA SCIENTIST'S SURVIVAL GUIDE

How To Win Over Business  
Leaders By Asking The  
Right Questions



prevedere



# TABLE OF CONTENTS

The Challenges of the Modern Data Scientist	<b>4-5</b>
The Opportunity for a New Era in Data Science	<b>6</b>
Step 1: Understand the Big Picture	<b>7</b>
Step 2: Align Tasks with the Big Picture	<b>8</b>
Step 3: Have the Right Data	<b>9-10</b>
Step 4: Identify and Analyze Leading Indicators	<b>11-12</b>
Step 5: Present Business Drivers to Leaders	<b>13-14</b>
Prevedere Empowers Today's Data Scientists with Tomorrow's Insights	<b>15</b>

# THE CHALLENGES OF THE MODERN DATA SCIENTIST

The rapid investment in analytics and the nearly insatiable appetite for business insights are causing unanticipated challenges for data science teams.

Kaggle, a platform for data scientists to connect, learn, find and explore data, surveyed more than 17,000 people in data science roles, to find common barriers to performance.

## Top 6 Data Science Challenges:

- 1 Dirty Data**
- 2 Lack of Data Science Talent**
- 3 Lack of Management/Financial Support**
- 4 Lack of Clear Question to Answer**
- 5 Unavailable/Inaccessible Data**
- 6 Results Not Used to Drive Business**

# THE CHALLENGES OF THE MODERN DATA SCIENTIST

Three trends are clear in these top challenges:

-  **Data scientists need access to clean, quality data to make their work meaningful.** Currently, data scientists spend almost 80% of their time gathering and preparing data. That means data scientists are left with only 20% of their time to actually analyze and glean insights from data. The most valuable aspect of their jobs is, unfortunately, often given the least amount of time.
-  **Data scientists should have clear visibility into business needs, decisions and results.** Data teams are frequently barraged by last minute, ad hoc data requests. In fact, a PricewaterhouseCoopers study shows that only 20% of work done by most analytics teams focused on foresight and were truly predictive. That means 80% of an analytics team's time is spent on ad hoc work that cannot be leveraged for additional requests. To have a strategic impact, data scientists' time should be protected to focus on challenges that will truly drive business value. They need clarity on the questions to be answered and resulting decisions to do so.
-  **Lack of available data science talent leave data science teams stretched thin.** Job postings for data science roles rose 75% from January 2015 to January 2018 at Indeed.com. The demand for data scientists and advanced analysts is projected to grow another 39% by 2020, according to IBM. Companies are also increasing their investments in tools to help power these jobs, with almost half of CIOs planning to increase their budgets for predictive analytics, according to International Data Group.

## THE OPPORTUNITY FOR A NEW ERA IN DATA SCIENCE

With all of this demand for data science, analytics teams have the opportunity to truly impact business performance. Given the fact that these teams are stretched so thin, they must focus on the problems that have the most impact on the entire company. **Understanding what will happen in the future with enough time to react is nearly always a top priority to executives.**

An analytics team has an opportunity to answer questions such as:

- ✔ **What trends will drive business performance in the coming year?**
- ✔ **What markets are most ripe for entry and exit?**
- ✔ **What will our strongest consumer segments be for new products?**

With these answers, data scientists can help decision makers deliver relevant, forward looking plans. **Following are five steps to help data scientists drive such decisions.**

*"CEOs are up at night anticipating things that are 3-5 years down the road. Decisions based on their view of the future can make or break not just the company, but their career."* -Craig Morrison, Retired, Former CEO, Hexion, Inc.



# STEP 1

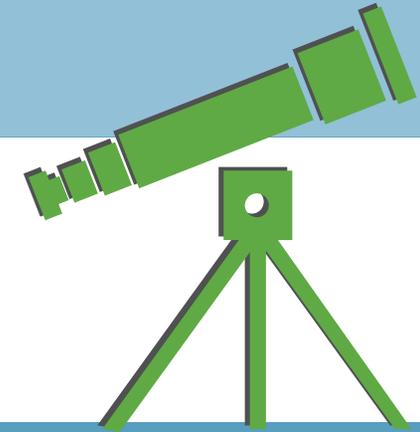
## Understand the big picture.

To understand the big picture of business challenges and opportunities, data scientists must first be afforded a direct line of communication with business leaders to understand their KPIs, the overall business landscape and major industry challenges.

With input from these business leaders, data scientists should review recent performance and KPIs with an eye toward:

- ✓ **Missed forecasts**
- ✓ **Results from new markets/markets exited**
- ✓ **Spikes in performance**
- ✓ **Changes in consumer demographics**

Often the biggest concern of CEOs is not why these events happened in the past, but what will happen next. Identifying and agreeing on the right questions in this initial audit will help business leaders prepare for similar changes and gain smarter decisions for the future.



### Common Industry Challenges:

- ✓ **Digital disruption**
- ✓ **Changing consumer preferences**
- ✓ **Complex supply chains**
- ✓ **Global economic threats**
- ✓ **Business model disruption**

**TIP:** Ask questions... even if they are not asked of you.

# STEP 2

## Align tasks with the big picture.

Data scientists' time is disproportionately spent on siloed work, decreasing focus on overall corporate challenges and directives. After analyzing the big picture (e.g. major industry trends, corporate performance and strategies), data scientists should then evaluate whether their current and upcoming projects are related to those themes.

### Are current tasks aligned with overarching business goals?

YES

NO

**If the answer is yes**, the next question is how these projects will drive decisions, so that data scientists can focus on the answers that will propel real change. For example, would explaining what caused a strong third quarter, which historically was the weakest, allow marketing and sales to replicate those campaigns/offers/price changes next year? Would understanding the factors that led to a market exit allow executives to test the health of other markets to anticipate future challenges? If so, then these projects should continue to be pursued. If not, re-evaluate the project list to focus on finding the answers that truly do matter.

**If current tasks are not aligned with business goals**, revisit the purpose behind those tasks. Sometimes, business leaders have difficulty articulating the answers they really need. It is a best practice to ask how the outcome of the analysis will drive decisions and how much impact those decisions will have. In many cases, the tasks really are aligned with overall business goals, but were not initially well defined. For instance, a marketing manager may request a breakdown of customers by income segment by all regions, which could be a time-consuming endeavor depending on how data is tagged and stored. In reality, the real question may be why a particular promotion did not meet expectations in a certain region. By identifying the real purpose, data scientists could find more effective and efficient ways to deliver insights to the business.

## STEP 3 Have the right data.

Most organizations have no lack of data available. Instead, the challenges lie in identifying, formatting and preparing that data. Often, corporate infrastructure makes it difficult to gather information before analysis begins. Acquisitions and disparate or legacy systems mean data may live in multiple places and in inconsistent formats. Before mining and cleaning begins, analytics teams should determine whether data has the potential to provide useful, forward-looking insights. Information that does not do so may be ignored, saving data scientists valuable time by eliminating the effort to clean and standardize data.

### **Companies are, however, frequently lacking data from outside**

**the organization.** Often, companies factor in few external insights into their data science processes. However, data, such as global economic trends and consumer preferences, can provide valuable insights about where demand and organizational performance is heading. By incorporating external insights, data scientists can begin to identify leading indicators of business.

A leading indicator is a related data series that changes before the data series that is being analyzed. Leading indicators are commonly used by economists to predict certain industry trends, but they can also be used to effectively predict future business performance. For example, if the number of hours worked by hourly wage employees steadily increases, that means they have more spending ability to eat out, go on vacation or make larger purchases. This helps decision makers predict and plan for future changes in demand, consumer preferences and supply chain performance.



# STEP 3

## Have the right data.

### EXAMPLE LEADING INDICATORS TO CONSIDER



1. **Consumer Sentiment Index** - a monthly report published by the University of Michigan that measures consumers' confidence in the economy. If sentiment improves over time, consumers may feel more comfortable making larger purchases (e.g., car, home, major appliance).



2. **Personal Savings Rate** - measures the percentage of an individual's disposable income that is set aside for retirement. Typically, consumers save less and spend more in a stronger economy.



3. **Foreign Exchange Rates** - a strengthening U.S. dollar can encourage U.S. consumers to travel internationally. It can also encourage companies to shift production outside the United States to improve profitability.



4. **Architecture Billing Index** - measures how many hours architects work on a monthly basis. More work for architects means more construction work for builders. This indicator helps determine new home demand, as well as wage changes for the many workers involved in construction.



5. **Average Retail Weekly Hours** - measures the number of hourly-wage retail employees. A decline in this indicator can signal weaker consumer spending ability, as stores do not need to staff as much.

# STEP 4

## Identify and analyze leading indicators.

Traditional analysis techniques were built upon the age-old adage: history repeats itself. To find leading indicators, however, data scientists need a process that accounts for the constant ebb and flow of the economy.

### Consider two of the most popular forecasting techniques:

1. **ARIMA** – in this method, which stands for autoregressive integrated moving average, models are built on time series data sets, hence “autoregressive”, based on a previous state.
2. **Holt-Winters** – this method, a version of Exponential Trend Smoothing (ETS), takes seasonality into account when forecasting.

At their core, these tools are based on projecting recent past performance into the near future. In today's global business environment, techniques like this are not powerful enough and are unable to identify a major turning point, a sharp trajectory change or a random shift in momentum.

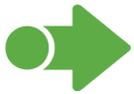
More powerful forecasting models, such segmented multivariate analysis fueled by machine learning based software, outperform traditional processes in longer-term forecasts and can be used to identify leading indicators.

Working with an analytics partner, like Prevedere, can also greatly speed up project timelines, allowing data science teams to continue to focus on insights and day-to-day work. **By rapidly analyzing millions of external data sets against internal metrics, machine learning techniques and cloud computing can calculate lead time and predictive confidence, creating models in minutes rather than days.**

# STEP 4

## Identify and analyze leading indicators.

### 4 CHARACTERISTICS OF A GOOD LEADING INDICATOR



1. **It leads.** It may sound obvious, but a leading indicator should provide insights into where a business is heading. It should provide enough advance notice that there is actually time to act upon the data.



2. **It has a strong correlation.** Meaningful leading indicators have a higher r-squared value and have strong pro or counter-cyclic relationships.



3. **It is based on enough historical data.** It is best practice to have at least 3 years of monthly data for back tests to be reliable.



4. **It makes sense.** An explainable, intuitive leading indicator is oftentimes more believable by the business than an unrelated indicator with a higher r-squared value.

# STEP 5

## Present business drivers to leaders.

With the right data and forward-looking indicators, data scientists can unlock specific, relevant insights on future business performance. Naturally the next step is to share these projections with business leaders, so they can make decisions about how to address future challenges and opportunities.

**The guidelines below will help to ensure executives understand the information being presented and actually use it to make decisions.**

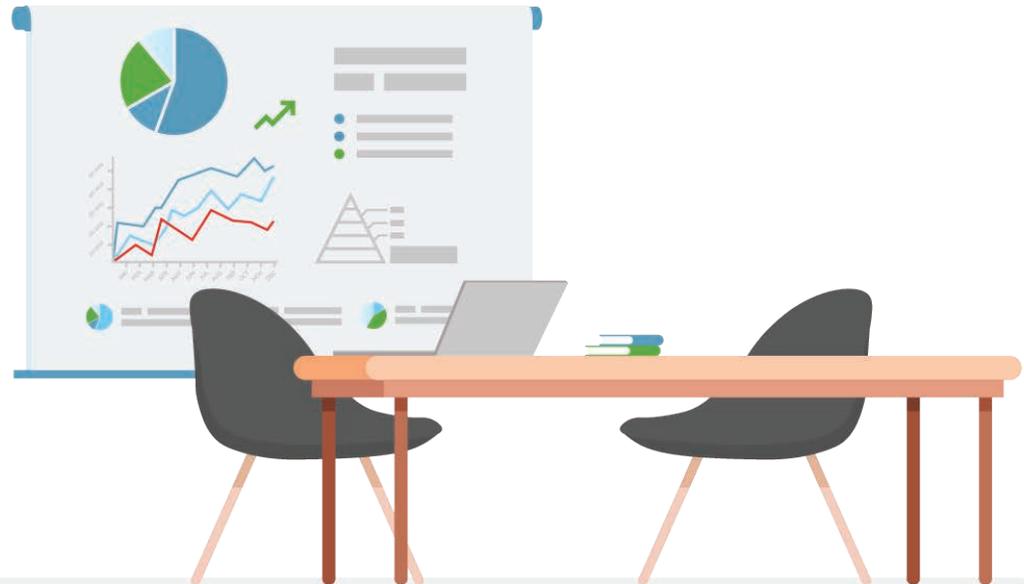
- 1. Use simple charts and graphs.** Executives do not need to see all of the analysis details. Instead, use compelling charts that display only the most relevant insights.
- 2. Focus on outcomes, not history.** Executives cannot act on historical information, they need to understand what is going to happen. Focus the presentation on the future, not the past.
- 3. Discuss the business value.** Extrapolate insights to provide specific, measurable impacts showcased by a specific analysis or indicator. For example, if the analysis relates to a given business unit that accounts for \$50 million in annual revenue, conclude the potential increase or decrease based on the analysis.
- 4. Have more detailed data as back up.** Be able to easily address any questions that might come up by having deeper data accessible, but do not dilute the message by including it in the presentation.

# STEP 5

## Present business drivers to leaders.

- 5. Ask questions.** One of the most important aspects of the presentation is to ask questions of business leaders to get their perspective and explain how the information could be used. If executives are uncertain it may mean that they were not clear on the data provided and analytics teams should revise the way insights are presented.
- 6. Close the loop.** A few weeks after the presentation, follow up with decision makers to determine exactly how they used the insights, such as how plans changed and how forecasts were altered. This is the best way to get a tangible view of how data scientists are adding value to the organization.

As data scientists seek not just to survive, but to thrive, in today's volatile business climate, clear communication and seeking feedback is the best way to ensure analytics teams are driving real business results.

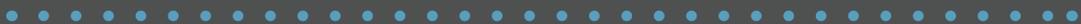


# PREVEDERE EMPOWERS TODAY'S DATA SCIENTISTS WITH TOMORROW'S INSIGHTS

Prevedere is an industry insights and predictive analytics company helping business leaders make better decisions by providing a real-time view of their company's future. Prevedere offers both global data and powerful analytics and modeling software.

**ERIN (External Real-Time Insights) Engine** is the first analysis engine that intelligently monitors external economic and consumer activity and notifies analytics teams of future impact on performance. The ERIN Engine determines the best combination of leading indicators, to develop predictive models that can easily be integrated into existing analytics or reporting software. **Learn more at [www.prevedere.com/erin](http://www.prevedere.com/erin)**

**ERIN Indicators** is a global data subscription offering that provides millions of economic, consumer, financial, climate and more directly into a company's existing data lake, database or reporting platform. ERIN Indicators are clean, comprehensive, continuously updated, categorized by industry, and proven to be leading indicators by their use in thousands of predictive models across a range of industries. **Learn more at [www.prevedere.com/erin-indicators](http://www.prevedere.com/erin-indicators)**



To see how Prevedere can augment your analytics capabilities and improve your business performance, contact us for an assessment at [inquiries@prevedere.com](mailto:inquiries@prevedere.com) or call 888-686-7746.

prevedere