Think Your Company Needs a Data Scientist? You're Probably Wrong.

Ask yourself these four questions before jumping on the data science bandwagon.

March 23, 2018 7 min read

A Note From The Editor

Think your company has what it takes to make our Top Company Cultures list? Apply now.

Apply now »

Opinions expressed by *Entrepreneur* contributors are their own.

When I started my career in data 15 years ago, I could never have envisioned a sexy rebranding of my work with the coining of the term "data scientist," let alone the immense popularity it's achieved in recent years. Widely considered one of the worlds hottest and most sought after positions, data scientists are re-writing what it means to be cool in the modern tech era. There has never been a better time for my fellow nerds. Jobs are overflowing with demand far exceeding supply. The industry has become so hot it's not uncommon for board members of startups to demand hiring of data scientists early in the product life cycle. It is in that capacity that I'm frequently brought in to meet with executives and more often than not, inform them that they do not need a data scientist.

Related: Who Owns the Data Your Business Uses? Not Knowing Could Hurt the Sale of Your Company.

How can a data evangelist such as myself argue this sudden interest in all things data science is on the verge of backfiring? Before I begin, let me start by saying that there are indeed many great uses for hiring a data scientist! I'm not going to argue that data science is not needed or is not useful, because when used correctly it's an incredibly powerful business weapon (yeah, I went there with "weapon"). I'm simply going to argue it's an overused term with little formal accreditation that refers to a large swath of data-related activities, not a tidy suite of skills that can be learned in a 12-month course. So, when it comes time to hire, organizations should put real thought and consideration into when and what kind of data scientist your organization needs.

When new prospective clients come to me, at least 50 percent of the time it's under the guise of "My CEO/board member/etc. told me I need to hire a data scientist." To which I generally ask the following four questions:

1. How much data do you have?

I say four questions, but many organizations never make it past the first. If you are a startup and you have not launched yet, you do not need a full-time data scientist. Full Stop. In fact, even if you are well-established but with a small customer/product/membership base, again you do not need a data scientist. Why, you ask? Because not surprisingly, data scientists need data. Not just any data will do. Many techniques require a minimum of tens of thousands, if not hundreds of thousands or even millions of data points to build.

Related: 5 Things to Keep in Mind When Using Data for Artificial Intelligence

Currently, there is a huge focus on deep learning. Job descriptions for data scientists are flooded with terms like neural networks, machine vision and natural language processing (NLP). The issue? These types of techniques rely on having massive amounts of training data. Consider the widely popular Google Translate, a type of neural network built on top of a lexicon of over 150 million words. The volume of data needed for successful deployment of these types of models exceeds what many companies own.

There are many techniques that use less data than deep learning, however, they still require reasonably large samples, not to mention a working knowledge of when to use which methodology. There is still valuable work to be done at this stage to create an environment where data science can thrive in the future, it just doesn't require a full-time, expensive resource to achieve.

2. Do you have established key performance indicators (KPIs) and regular business intelligence reporting?

Without basic understanding of what drives the organization, it's going to be very difficult to make use of advanced techniques. For example, a data scientist can use machine learning to make predictions such as which users will churn or become highly active, however, if the business does not have a definition for churn or highly active, that becomes a requirement prior to building the predictive models. Furthermore, it's difficult to validate models if you don't have sufficient metrics with which to evaluate them. Other techniques such as A/B testing require advanced selection of an overall evaluation criteria (OEC), which is typically a business-driven KPI.

Related: 6 Critical Questions to Help Businesses Cut Through the AI Hype

3. What do you imagine this data scientist will do once hired?

Perhaps the most subjective and interesting of the questions I ask, “What do you want this data scientist to do?” The most common answer I get is. “We don’t know, that’s why we need to hire one.” In that case, I gently tell the organization that they are setting up their data scientist to fail. There is no need to be an expert in data science to hire one, however you should have a good idea of what is and isn’t possible so that you don’t set unrealistic expectations.

Data science isn’t magic and it’s not even a traditional science. It’s just as much an art as it is a science, which means the variability in skills and ability is substantial. You may even have existing team members able to grow into many data science applications. An easy entry into data science for an existing analyst is to begin forecasting the KPIs they already report on. Here they have the opportunity to learn on data they are familiar with, which is not just good for employee morale; investing in your staff now means less need to recruit in a highly competitive market in the future.

Related: What Hiring Managers Don’t Understand About Hiring for Data Science

4. What support networks are available to your data scientist(s)?

If you don’t have the right support network for your data scientists, don’t bother investing in hiring them. In recent

years, there has been a huge surge in data science programs, however the graduates are for the most part simply not ready to tackle business problems without careful hand holding. The vast majority of programs have students solving pre-established problems on clean data. In the real world, you want your data scientist to help determine what problems are being solved, and clean data never exists.

Hiring a junior data scientist without a senior resource for guidance can not only lead to frustration on the part of the junior, it can often lead to bad analysis. Junior team members tend to struggle translating business problems into technical problems and the wrong translation could result in months of work on a product that misses the target.

This problem is not completely mitigated by hiring more senior, partly because certifying your senior hires are actually good and competent is extremely difficult. If you luck out and hire a talented and self-motivated data scientist, she will still need a lot of support at the executive level to succeed. Imagine a situation where models are created but never used because there is no buy-in from team leads. Or where A/B tests are conducted but the results ignored. Worse yet, the data tracking needed to analyze a problem isn't being collected at all.

Related: 4 Things a Data Scientist Can Do for Entrepreneurs

Frequently, a necessary first step is a robust data collection program, which is likely resourced by an engineer or database administrator, not a data scientist. At many organizations, the senior data scientist(s) spends exorbitant amounts of her time simply fighting for the data requirements and deployment of her team's work. That's a surefire way to lose that talented, self-motivated, senior data scientist.

The landscape for hiring and retaining good data science talent is competitive and expensive, but being smart and

conscientious on when, who and how to hire can mitigate the pain and cost. Don’t fall into the trap of job postings that are laundry lists of skills. Don’t expect magic pixie dust from your data scientist. Do take inventory of your true requirements and if possible, consult with a trusted professional prior to hiring. The success of your data program depends on it.

https://www.entrepreneur.com/article/310505