

All assumptions are false! 7 Lessons I wish I paid more attention to on every Predictive Analytics Project

How many times do you have to be told something before you believe it? Do you have to experience failure first hand before you really learn the lessons? When I first got involved with Predictive Analytics, I was lucky enough to receive lots of advice from peers. *“Start with the business question, not with the model”*; *“70% of your time is taken up with understanding the business and preparing data”* (CRISP-DM wasn't wrong!). But over the last 10 years there have been several lessons that I don't think I ever fully appreciated until I experienced them (time and again). Together with famous quotes I like, I've shared below 7 important lessons that you can never give too much attention to!

1. *“Progress is impossible without change, and those who cannot change their minds cannot change anything”* **George Bernard Shaw**

Buy-in is always far more important to a Predictive Analytics project than you first think. Without buy-in, changing processes, implementing predictions, monitoring effectiveness, and even measuring return on investment are all considerably more difficult. You may already have champions supporting the cause, but all too often a project will fail during implementation because of organisational resistance. It is just as important for Analysts, Line of Business managers, IT managers, Campaign managers and CxOs to get behind your project. Predictive Analytics can be as much about managing change within an organisation as it is about a model or prediction. [Take a look at this article](#) from my colleague Suzy Kell expanding on this.



2. *“An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question”* **John Tukey**

Understanding organisational processes should come before data. As early as possible, you should try to frame the customer experience. Understand any constraints, understand what it is that the organisation can or will do with the results. Understand what they may not be able to do or use. Will there be a return? How can it be measured? Choose modelling techniques that best fit with existing processes. It may even be the case that you need to pay special attention to how you define what it is that you are predicting. Can it actually be turned into an action? – ahead of time?



3. *“In God we trust, all others must bring data”* **W. Edwards Deming**

Perhaps an over-stated quote, but I've lost count of the number of times during business and data understanding meetings that I've heard a statement about data consistency or business processes that turned out to be untrue. It may seem suspicious, but without testing these assumptions, the quality of the end model, in



particular the interpretation, can be spurious. Whether it be appropriateness or uniqueness of a data table, or assumptions around payments and invoices, each should be tested. If proven true, then you've just been careful. If proven false, there's a challenging insight for the organisation to address.

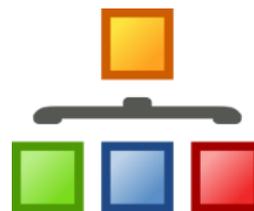
4. *"To expect the unexpected shows a thoroughly modern intellect"* **Oscar Wilde**

The last thing that you want when creating a model is surprises. A surprise that there are some data inconsistencies skewing your results; A surprise that an important field has not been used in the model. A surprise that the accuracy isn't what you were expecting once deployed. Extra time spent is never wasted when checking for Multi-collinearity, non-linear relationships, anomalous records and learning your data inside and out. Then, by the time you start modelling, interpretation is easier and the most apparent flaws can be quickly understood and explained.



5. *"If you can't explain it simply, then you don't understand it well enough"* **Albert Einstein**

What hope does the business have then? As important as it is to create meaningful insights, it is just as necessary that those responsible for implementing the results understand the reasoning and the action. Try to keep it simple and to tell a story. Look at insights from the perspective of the business, use communication techniques like top-down thinking. This will go a very long way toward making the insights essential to organisational processes.



According to my colleague Joel Elebiju, this is one of the 8 pointers that any analyst should strive to develop. [Take a look his article](#) to read more about what makes a good analyst.

6. *"All models are wrong, but some are useful"* **George Box**

Another common quote in statistics, and while it shouldn't be used to excuse a poor model it does tell an important story. To begin with, usability and understanding should be considered at least on par with accuracy when evaluating models. Further, if a model is performing badly, then this should be considered a valuable opportunity to learn something about why this is the case. Best practice predictive analytics is an iterative process. Very rarely would you get the best possible models first time around.



7. *"If there is no struggle, there is no progress"* **Frederick Douglas**

Sometimes after a predictive model is built, your job is still not even half done. Project Implementation, documentation, data loading, quality testing, knowledge transfer, and change management are all examples of post modelling tasks that take far more time than you ever hope or expect. Prepare for this as best you can. Rushing the implementation can often serve to introduce confusion and resistance. Without proper attention, the model may end up unused or forgotten.



In the end, I believe each of these lessons are well accepted and understood by most data scientists and statisticians. They may not come as a surprise even. The temptation though, is to hope each are not as critical or as impactful as they could be and often end up being. It is somewhat natural to minimise the focus on these lessons in favour of an easier, *seemingly* painless and speedy result. ***“I can resist everything except temptation”*** said *Oscar Wilde*. Maybe this is too true, but I can’t help but want to re-tell myself all these lessons over and over again for every project I’m involved in. Little by little, with each project, I’m resisting more the temptation to make assumptions and take these lessons for granted in the face of experience. So what’s my advice to people starting out in Predictive Analytics nowadays? (and even to myself for each new project...)

“All assumptions are false” *Alan McMahon*

About Presidion

Presidion have operated for over 20 years and have been the pioneers in implementing cutting edge predictive analytics solutions with top UK and Irish organisations. We specialise in helping organisations leverage their data to deliver tangible practical returns on investment, aligned with their strategies.

Presidion works with both government and commercial clients, currently partnering with hundreds of organisations enabling them to understand what has happened in the past, anticipate what may happen next to take appropriate and timely strategic decisions for their organisation.