# Why most big data analytics projects fail: How to succeed by engaging with your clients

Max Henrion, PhD,
CEO, Lumina Decision Systems
Los Gatos, California.

Despite the great excitement about big data, better analytics tools, and the vast resources that many organizations are investing in growing their teams and technology, multiple surveys of data analytics groups report that most analytics projects fail to provide real business value. In 2015, Gartner Research estimated that 60% of big data projects would fail over the next two years. Two years later, Gartner analyst Nick Heudecker admitted that they had been "too conservative": The actual failure rate was "closer to 85 percent".

A four-year study of major analytics initiatives in large companies reported in the Harvard Business Review [] that less than half of the 36 companies studied reported measurable results, and little more than a third had met their objectives of widespread adoption. A 2019 survey by New Vantage Partners [] found it "particularly striking that 77% of respondents say that 'business adoption' of big data and AI initiatives continues to represent a challenge for their organizations." Andrew White, a key Gartner analyst predicts [] that it will not get better any time soon "Through 2022, only 20% of analytic insights will deliver business outcomes." The excitement (the "hype cycle") appears to be peaking and about to enter the trough of disillusion.

## Why do so many projects fail?

According to the Gartner survey [], two of the main reasons for failure of analytics projects were: "management resistance, and internal politics." The HBR study [] reported similar findings: The biggest impediments to successful business adoption were "insufficient organizational alignment, lack of middle management adoption and understanding, and business resistance." In other words, most of the practitioners and leaders of data science and analytics groups, who responded to these surveys, blamed their managers for failing to recognize the value of what they were doing. In many cases, these managers were the same executives who had approved large investments in high-priced analysts and technology.

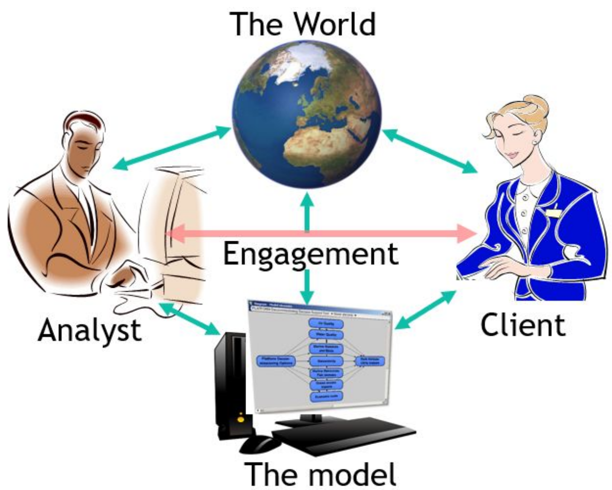
As a long-time practitioner of analytics, I too have found myself on occasion irritated at the obtuseness of my clients and their dysfunctional organizations. But then I remember: Our purpose as analysts is to bring greater clarity and insight to our clients and to improve the quality of their decisions. How effective is it for us to blame our managers or clients for their failure to understand and appreciate the results of our hard work? We are failing as analysts if our clients do not appreciate the value of our hard work.

Let's assume that we have mastered the technical skills to clean, and analyze the data, to fit statistical models, apply machine learning methods, and create compelling visualizations. There's something else many of us are missing. Often the most challenging task – and one for which many analysts have little or no training – is to engage effectively with our clients, to understand how they think and what they care about. We need to understand and clarify their view of their problems, and make sure that our analysis addresses that view if our client is going to appreciate our results.

## Who is your "client"?

By ***client***, I mean the person or, more often, group of people who you are trying to help make better decisions. It may be your immediate manager, a group of senior executives, your consulting client, or a set of stakeholders and subject matter experts from several organizations, end users of decision-support software, or it may even be you. Here I use "analyst" and "client" in singular form, recognizing that often one or both are actually a team with multiple participants.

Ultimately, the client should be responsible for making decisions, or at least have significant influence over important decisions. Otherwise, why bother? It is unsatisfying to do analysis, no matter how elegant and interesting, if it has no noticeable effect on significant decisions.




## It's all about the relationship

Effective analytics is all about relationships. Most obviously there is the relationship between the model you are constructing and a selected part of the World that your client needs to understand. The "model" might be simple descriptive statistics and visualization of the data, a sophisticated statistical model with predictions, a simulation of a dynamic system, or a decision analysis. In each case the value of the model depends on having sufficient realism and clarity to be a guide for effective decision making.

At least as important is the human relationship between analyst and client. Too often this relationship consists of just two kinds of interaction: The project starts with the client giving the analyst a brief -- sometimes little more than "see what insights you can find in this data". And it ends with the analyst presenting the client with the "results" – often a Powerpoint presentation and/or a written report. This approach is a recipe for failure. Success requires a much more extensive engagement between analyst and client.

## Geeks and empaths

The aptitude and training of most practitioners of analytics is about numbers, equations, and computational tools, the "hard skills". Not to put too fine a point on it, many of us are "geeks". Few of us start out with the soft skills to engage effectively with our clients so that we can really understand them:

How to ask probing questions in a non-threatening way, to listen effectively, to observe tacit body language, and to get inside the heads of our clients, and the way their organizations work.  In short, we need to become "empaths".

Fortunately for us geeks, it is usually possible to learn these soft skills once we realize how critical they are to our success and decide to put the time and effort.  We can learn some from classes by experienced practitioners, but most effective is to work on real applications for clients with such practitioners. There are also a set of tools and methods designed to support and encourage effective engagement between analyst and client. Most were developed by decision analysts, perhaps the subfield of analytics that has paid most attention to the relationship with the client. The rest of this article introduces several of the most useful of these tools and methods.

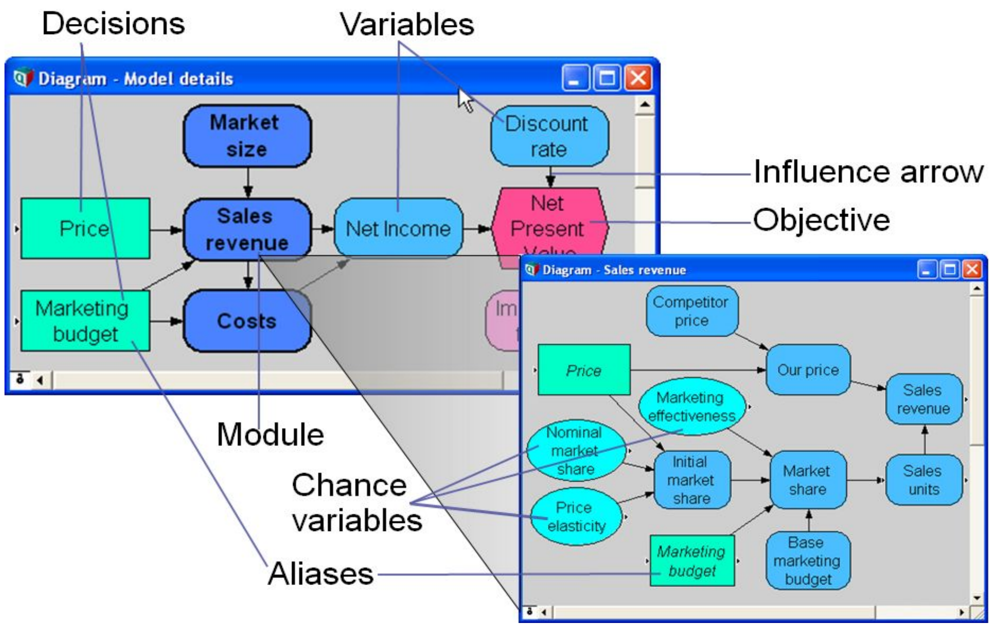
## Influence Diagrams

Often a client's decisions and objectives may at first appear confused and even contradictory. The analyst's job is not simply to ask and write down the answers, but to work with the client to help them clarify them into a well-structured form amenable to analysis.  Decision analysis is perhaps best known for its use of ***decision trees*** for structuring simple decisions under uncertainty. Decision trees are less tractable for complex problems because the number of branches is exponential in the number of decisions and uncertainties. Less widely known. but far more practical for structuring complex problems, are ***influence diagrams***, a complementary visual representation also developed by decision analysts (Jim Matheson and Ronald Howard, 1984).

Influence diagrams are a graphic facilitation tool for structuring and analyzing complex decision problems. Even people with limited quantitative skills find them intuitive. Each node depicts a variable, with its role indicated by shape and color[1]:

- A **decision** (green rectangle) is under the direct control of the decision maker, such as price or marketing budget for a company that sells products or services.
- A **chance** variable (light blue oval) has an uncertain outcome that the decision maker cannot control and may be expressed as a range of possible values. Decision analysts typically represent uncertainty as a probability distribution.
- An **objective** (red hexagon) expresses a quantity that the decision maker desires to maximize or minimize, such as expected net present value. Decision analysts like to define this as utility, including risk attitude and often multiple attributes – or sub-objectives.
- A simple **variable** (blue oblong) depicts an input or calculated variable that is a deterministic function of the variables that influence it.
- A **module** (dark blue oblong with thick border) depicts a sub-model with its own diagram, providing a way to organize a large model as a hierarchy of comprehensible elements.

[1] All influence diagrams depict decisions as rectangles and chance nodes as ovals. Notations vary slightly in the choice of shapes and colors for other roles. These examples are from Analytica software.




Two levels of influence diagram for a simple pricing and marketing model

Typically, the analyst starts by interviewing the client asking about key objectives, decisions, and uncertainties, and adds each element as a node on the diagram. You may also ask about key sources of relevant data, estimates, and uncertainties that may help analyze how alternative decision strategies might help achieve the objectives. You draw in influence arrows to show how the variables affect each other. As analyst and client collaborate to develop an influence diagram, they must identify the role of each variable: Is it a decision, a chance variable, objective, or something else? This process naturally focuses attention on these key elements. The effort to distinguish decisions and chance variables often generates a valuable discussion about what is or is not under the control of decision makers. Asking the client to articulate objectives often leads to a deep and fruitful conversation about the aims of the organization and its attitude to risk.

When possible, it's helpful to work face-to-face with clients so that you can be alert to tacit nonverbal communication and body language for closer communication. Influence diagrams also work well in web-meetings, when your clients and analyst teams are geographically remote, as a visual representation to develop a shared understanding of the problem. The goal is to develop a shared understanding of the decision problem between analysts and clients.

The initial goal is just to develop a simple qualitative representation. Often, the analysts will work later separately from their clients to develop the underlying numbers and formula to quantify each influence. They may be deterministic or probabilistic. Some maybe simple accounting relationships (Earnings = Revenues – Costs). Others may be empirical relationships, with statistical models estimated from data, or based on expert judgment of the conditional probabilities. Often analysts add further variables to add structure and refine an influence relationship. It is helpful to move these additional variables into a module or sub-diagram to avoid making the original diagram too complex**. In this way, you can organize a complex model with hundreds or even thousands of variables as a hierarchy of modules, each with its own diagram with few enough nodes to be easily comprehensible.**

## Agile Analytics

The traditional *waterfall process* for developing software, borrowed from physical systems engineering, consists of a sequence of steps: Needs analysis, specification, design, implementation, quality assurance, documentation, user testing, and deployment. More popular nowadays is an *agile process*: You start with a brief exploration of user needs. You build a simple proof of concept or prototype. Users try it out and give feedback. In response, you refine and extend the software. You repeat the process until the users are happy (or you run out of time and money). An agile approach is almost always best for analytics. Even after your initial work with the client to clarify their view of the problem, it will can be unclear exactly what you can obtain from the data, analytics, and modeling to address that view. So, it pays to start out with a simple proof of concept or prototype to get feedback from the client. You can then refine the analysis, focusing on extending those aspects the client sees as likely to be helpful.

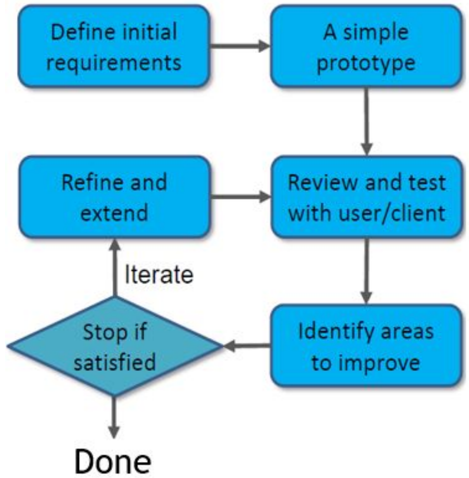
This iterative process results in a series of engagements with the client during which you may help them refine their decision and objectives as well as your analytic results. It enables a more rapid convergence between the two that is much more likely to provide real value to the client. It also means that they are likely to get some insights early in the process, rather than having to wait until the end. Given inevitable uncertainty about what is in the data and the needed level of effort to conduct the analysis, it greatly reduces the risk of running out of time or funds before you provide the client with real insights. In some cases, after interim results will inspire the client to identify new questions, decisions, or objectives, and propose project extensions to generate additional value.

### Uncertainty and sensitivity analysis

No matter how much data we have, a decision-focused analysis will have to deal with uncertainties, both from fit to past data, missing data, and the need to use expert judgment. Any data has a huge limitation: It is always from the past. So how do we predict how our decisions will affect the future? We can build predictive models that are informed by past data, from simple time-series forecasting to complex systems dynamics simulations. Due to the limitations of our understanding and the chaotic behavior of natural and economic systems, there will always be substantial uncertainty in forecasts. Our models always need a good dose of judgment from informed human experts.

We should represent each of these sources of uncertainty explicitly in our analysis, initially as a range, and later, where appropriate, as a probability distribution. At each iteration, you can perform sensitivity analysis to see which assumptions or uncertainties could make a significant difference to the recommended decisions and objective. Exploring why each uncertainty makes such a difference is often a potent source of insights. It also gives valuable guidance for the next iteration of the analysis to make sure that you focus your further data gathering, model fitting, or model extensions on the issues that make most difference. **It helps you avoid the common trap on spending most of your time on analysis of the quantities for which you have most data rather than the issues that make a difference to your client.**

Sometimes you will find that a decision is relatively robust and performs well in the face of a wide variety of possible futures. If not, perhaps you can work with your client to design a more robust decision strategy. These kinds of explorations are especially effective if you can do them interactively with your client. Often you may find results that are unexpected or surprising – a sign that it requires deeper investigation. Is there a bug in the model or faulty data – or is it there something wrong with their intuition about what to expect? That is often the source of a valuable new insight.




Agile analytics process

## Interactive Decision Support

We learn best about how things work by playing with them directly – rather than being told about them. Often the most effective way for you and your client to understand a model and get real insights is to explore it interactively. Together you can examine the effects of changing assumptions and compare alternative scenarios to get a visceral sense of the model behavior. It helps if you can design compelling and insightful visualizations to understand the model behavior and sensitivities. If your analytics is so computationally intensive that it takes hours or more to recompute, it often helps to precompute results for multiple scenarios with a wide variety of assumptions. You can then create a decision-support tool to explore and compare these scenarios interactively even if the full model would take too long to rerun.

## Conclusions

Successful analysts already know that close engagement with the client is a key to their success. They understand that effective analytics, like any kind of consulting, relies on a deep conversation between analyst and client. I have introduced some key tools and methods that to support this kind of engagement:

- **Ask questions and listen** carefully to understand who your real client is, how they think, and what decisions they can make.
- Use **influence diagrams** to help your clients identify their decisions, objectives, and uncertainties, as well as how available data could inform better decisions.
- Represent **risks and uncertainty** in your assumptions explicitly, including the (in)accuracy of models fitted to data and expert judgment. Use sensitivity analysis to find out which uncertainties really make a difference to the recommended decisions and why.
- Use **agile development**: Start from a simple proof-of-concept, then iteratively refine and extend the analysis or model guided by feedback from your client and sensitivity analysis.
- Provide your client an **interactive decision tool** so that they can get a visceral experience of the model behavior with visual insights into the effects of key assumptions and tradeoffs.

Your success as an analyst in producing effective results that are appreciated by your clients will depend as much on the quality of engagement with your client as on your skills with analytics techniques and tools. A few analytic software tools support these methods, including influence diagrams, sensitivity analysis, agile development, and building interactive decision tools. But, to become truly proficient at engaging with your client there is no substitute for practice on real applications with experienced analysts.

## References

[1] 85% of big data projects fail, Matt Asay, TechRepublic, Nov, 2017
https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/

[2] Big Data Executive Survey 2017: NewVantage Partners LLC,
http://newvantage.com/wp-content/uploads/2017/01/Big-Data-Executive-Survey-2017-Executive-Summary.pdf

[3] *The Reason So Many Analytics Efforts Fall Short,* Chris McShea, Dan Oakley, & Chris Mazzei, Harvard Business Review, August 29, 2016
https://hbr.org/2016/08/the-reason-so-many-analytics-efforts-fall-short

[4] *Our Top Data and Analytics Predictions for 2019,* by Andrew White, January, 2019.
https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/