The Data ROI Toolkit:
How to Determine the Value of Your Data Initiatives

A WHITE PAPER BY DATAIKU

www.dataiku.com

INTRODUCTION

SPENDING & THE HYPE CYCLE

There is no sign that data science is on the decline. Businesses have jumped on the bandwagon, investing in expensive data teams, infrastructure, and tools in hopes of improving or increasing the velocity and output of the team, but there is little understanding of how to calculate the impact of that data spending.

At the time, **experts claimed** that these investments would take years to pay off. Here we are years later, still with billions and billions of dollars in big data-related costs (from staff to tools), still investing and putting faith in data teams. And yet many enterprises report not seeing a return on their investment and are left unable to prove that any of it is really worth the time and the money.

Looking at some of the key areas driving spending over the next few years, Gartner forecasts $2.9 trillion in new business value opportunities attributable to AI by 2021, as well as the ability to recover 6.2 billion hours of worker productivity.

Yet without hard numbers pointing to success, it is difficult for executives to continue to invest thousands (or millions) of dollars on the latest data efforts. Indeed, any enterprise that has a data team, it seems, is reevaluating their productivity and return on investment (ROI). And anyone looking to spin up a new data division is doing their homework first, closely examining costs and potential ROI before diving in.

REAL-WORLD ROI CALCULATION: FACTS & FIGURES

Many companies are, of course, reluctant to disclose the ROI they have received from their data science efforts (especially for specific projects that provide a competitive advantage). However, in February 2018, Forbes reported that Netflix credits its content recommendation engine with reducing customer churn to the tune of $1 billion annually.

In a study sponsored by Teradata conducted by Forbes and McKinsey, large enterprises reported that data efforts improved company growth by just 1 to 3 percent on average. But still, in this study, only 37 percent of respondents could quantify the business case for big data analytics, while 47 percent could not and 9 percent reported they had "no clear vision." And yet another study by BCG estimated 20 to 30 percent EBITDA gains for data-driven companies. So clearly this is something that today, most companies struggle with (you are not alone!) - and that analysts struggle to define it too.

So the bottom line is that while most companies (and analysts) struggle with calculating big data ROI, including what exactly they need to measure to get there, larger data suggests that investing in big data, data science, machine learning, etc., is well worth it - for example, this study by the International Institute for Analytics (IIA).

But how can your business concretely show ROI?

CHALLENGES TO CALCULATING DATA SCIENCE ROI

Of course, when it comes to ROI, it always seems to be easier said than done. The reality of measuring the return on investments in data teams and projects, nevermind data tools and technologies, can be particularly challenging because:

- It's often difficult to isolate the contribution of data alone to improvements, especially larger business outcomes (like higher profit margins, lower costs, etc.).
- The calculation is complicated because the value isn't all in one number - it can be spread across multiple departments and teams.

For these reasons, measuring ROI for data projects can end up being a data project in and of itself, which is often difficult to justify.

The goal of this white paper is to address and alleviate some of these challenges for an easier path to providing a clear ROI for your data efforts. Specifically we focus on investing in tools, which can ultimately unlock additional support from leaders and other teams in the company that will provide continued fodder for a thriving data culture.

WHAT DOES ROI ACTUALLY MEAN?

The question isn't so much what does ROI mean, but rather what can it mean - businesses invest in data teams, infrastructures, and tools for all different reasons and are executing different projects at various stages of maturity. So it makes sense that when it comes to figuring out ROI, it's not a uniform calculation.

Therefore, the first step in calculating ROI is to define "success" for the particular business and considering all the ways - directly and indirectly - that data, or a data department, has made contributions. Value can come in many different forms, so the biggest part of the work involved is often determining all the possible ways that data could be bringing success. The list of factors are different for each business, but the universal contributions are:

POSITIVES

Time Saved

- This is the easiest to quantify. The salaried-time saved through data initiatives is easy money you can put straight in the bank (or plug back into new initiatives). If data cleaning used to take you 10 hours per data scientist per week, and through a new initiative you can expect that to drop to 1 hour per week, you will immediately save money.
- However, this isn't as straightforward as it first appears. The time saved by data initiatives is rarely constrained to within the data team; you may need to look across the entire company to fully realize the ramifications of each project. If you would prefer to be conservative in your calculations, you may choose to limit this metric to the data team only. This also increases the ease of analysis prep, saving time for whoever is doing the ROI calculations!

Reduced Internal Churn

- Data Scientists and Analysts are hot commodities right now; according to Accenture, upwards of 80% of new data science jobs are not being filled. Thus, keeping data scientist morale high is critical to the continued success of your business. This ties into time saved because when data initiatives save data scientists from drudgery, they are happier in the workplace and free to work on more complicated, revolutionary tasks.
- While it is difficult to put a monetary value on this, we can extrapolate. According to Glassdoor, the average data scientist salary is $117,000. If a data scientist quits for a more interesting or better paying position, it could take up to a year of recruiting efforts to refill the role, considering there are over 20,000 open data scientist job postings. So while you would save $117,000 (plus ~$10,000 in benefits), you lose a significant portion of a recruiter's time (priced at $51,000 a year) while slowing down your existing data flow and increasing stress on the remaining data team. If the data team could complete 6 projects a year, each averaging $1 million ROI, and without this data scientist are now able to complete only 5.5 projects in a year, then that's a loss of $500,000 and a feedback loop is created as there is more stress on the existing data team, who then are liable to quit, which is liable to slow down projects even more, etc.
- Approximate value: priceless
- (Okay, but really): [**Avg annual rate of DS churn**](**500,000 - 127,000 + 0.05(51000)**) **= $375,550 x** [**Avg annual rate of DS churn**]

Reduced Production Costs

This value is entirely dependent on the individual project. Data Initiatives can improve accuracy and efficiency across industries. A few common projects and their impacts are

- Predictive maintenance
  Money saved from not having to replace machines or systems that have catastrophic breakdown, decreased customer churn due to failures, fewer staff required for maintenance, and fewer redundant parts/contingencies purchased
- Anomaly detection
  - Healthcare: faster successful patient turnover and fewer unnecessary medical procedures.
  - Finance: reduced fraud
  - Retail: decreased staffing costs and improved speed of launch
- Recommendation Engine
  Increased customer engagement and brand loyalty.
- Customer Service Chatbot
  Decreased support operating costs and improved customer experience

More Customers

- This is a twofold benefit. When data initiatives demonstrate your company's commitment to their customers, they are more likely to choose you over competitors, giving you an edge. Additionally, customers are less likely to churn when their experience is tailored to their individual needs. Data initiatives can make this sort of luxury experience possible without radically increased staffing costs.
- To calculate this benefit, you need to track the rate of customer engagement in the period following the initiative, and compare it with the previous rate. Subtract the cost of marketing the data initiative to prospective customers from the increase in customer engagement (or a percentage thereof if other launches or factors may have contributed).
- Customer trust is also a factor to consider, along the lines of predictive maintenance. If you suffer from a serious data problem, you may never recover the trust of your consumer base. This is impossible to calculate, but should figure into your project prioritization.

CONS - THE COSTS

While data initiatives (when realistic goals that are operationalized) have financial benefits, there are some short term losses that can factor into the overall ROI.

Workflow Disruption

- If your data initiative has internal ramifications, there will be a short term in productivity as users adapt to the change. This may also cause a drop in morale if users do not fully understand the benefits of the change. However, taking the time to get buy-in can help mitigate this.
- Depending on the severity of the shift, budget for a reasonable percentage of time lost to change inefficiencies.

Education

- If you are working on cutting edge data science, you may need to educate your team into this new terrain. While this also has a benefit of decreasing data scientist churn,—when the company invests in them, they notice—in the short term, this will take time and funding.

> [**Number of hours spent**] x [**Number of team members learning**] x [**salary of team**] + [**cost of tuition/resources**]

Tools

- If you leverage a data science tool to execute on your data initiatives, you must factor in the subscription price into your ROI calculation.

ROI CALCULATION WORKSHEET EXAMPLE

Often one of the biggest costs when it comes to really ramping up the velocity of data projects or building a data team are the tools and technology required to thrive. This can be a difficult investment because it's inherently difficult to calculate ROI for tools of any kind, but on top of that, data science tools are relatively new, so it's not necessarily globally accepted that they're critical (unlike, say, a CRM, which most businesses would use without question). Add that on top of the aforementioned challenges of calculating the ROI of data projects more broadly, and things get complicated quickly.

To clear the waters, we created a baseline ROI calculator for you to determine the ROI for data science tools investment. Test out the Excel worksheet that came with this whitepaper for yourself.

LESS TANGIBLE ROI FACTORS TO CONSIDER

It's worth mentioning that there are other factors businesses might want to consider when looking at the ROI of data science outside of the standard success definitions. These tend to be much less quantifiable, but can be equally important when it comes to the longer-term image of a company and brand. For example:

The Risk Of Non-Compliance

Not investing in data science more broadly or a data science platform specifically can increases the chances of a catastrophic public relations issue (such as a large data breach) or fines due to regulatory non-compliance (like GDPR - for example).

The Cost Of Disruption

Data-savvy companies are shaking up every industry, so it's important to consider what the business might be losing by not being the first one and the most innovative one in the space when it comes to machine learning, deep learning, artificial intelligence, etc. Or, at the very least, it's important to start considering ways to become data-driven so that even if your business isn't the first in the industry to do so, it won't be the last to jump on board. Disruption also influences staff turnover - talented staff might move on to other more cutting-edge companies.

The Value Of Digital Maturity

Today, customers often place more trust in companies that have a high digital maturity, providing more personal information or making more purchases as a result of this trust. The use of data science platforms put enterprises at the forefront of the big data era when it comes to both innovation and data governance, which is more likely to result in this consumer trust (though customer fear of becoming a "product" by providing their data is also a reality).

These consideration factors are tightly related to long-term company performance, which is, in turn, intertwines with analytics maturity (*as the International Institute for Analytics found*).

HOW TO ENSURE THAT YOU CAN TRUST YOUR ROI

Even the best equation can't solve the question of ROI for you. In order to make sure your calculations are robust, make sure that you are:

- **Conservative In Your Estimates:**
  It is better to undervalue a project than to promise the moon. This ties into making sure your estimates are as accurate as possible, not just guesswork. Talk to the teams impacted to better understand their time limitations and morale levels. This will also help decrease potential resentment leading from workflow disruption.
- **Solving The Base Issues First**
  If you can't trust the accuracy of the data you are working with, then the best predictive maintenance model in the world will still fail. While high-tech POC projects can help drive company buy in, you need to make sure your data is coming from trusted sources, consistent, and updates automatically, otherwise no data project will succeed.

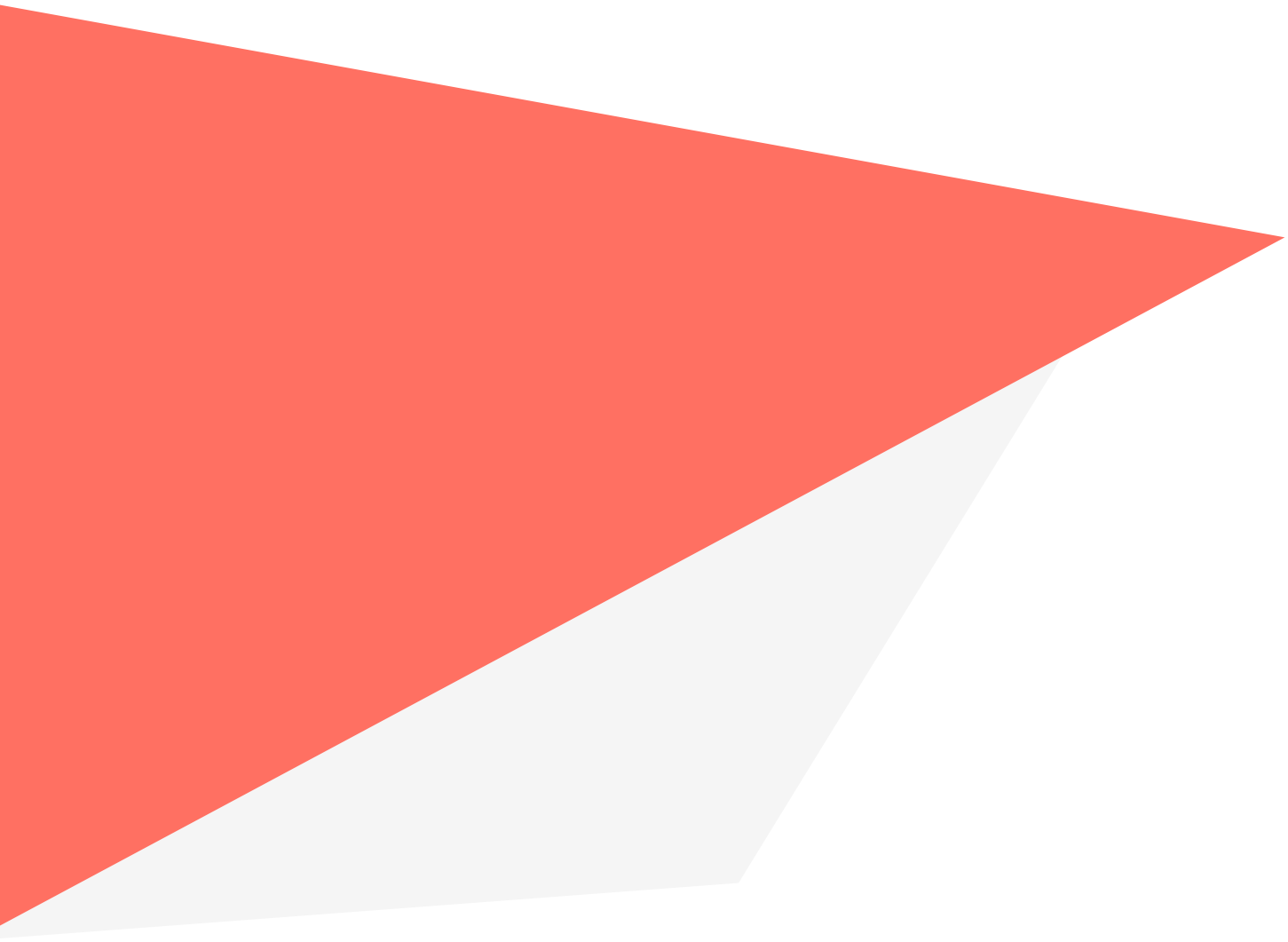
HOW TO INCREASE ROI

The next question to ask after calculating ROI is: how can I improve the ROI I get from my data and data science initiatives? This may mean gathering new data from untapped sources, or it may mean improving the productivity of the data team as a whole. Simply purchasing a tool or hiring a team to do data science will not magically bring ROI - there is no silver bullet. It takes organizational change throughout (from high-level management down to each individual contributor) to get value from data. One way to make it easier to calculate the ROI of data science projects, technologies, and tools moving forward is by setting clear KPIs that are put in place and tracked over time. At a minimum, this means tracking:

- The number of projects delivered per month
- Time from start to prototype and from start to production
- Project completion - that is, the ratio of projects deployed to production vs. requests from business teams (or other teams across the enterprise)

CONCLUSION

Most businesses investing in data science are unable to accurately calculate the ROI on that investment. Because of this challenge, companies who have not yet fully taken the plunge in leveraging their data for machine learning projects may be reluctant to invest in things like data science tools that make it happen. But by taking the time to consider and quantify the potential gains, data teams can be well set up and positioned to prioritize the right projects to deliver on the promised value that data science holds.



data iku

Your Path to **Enterprise AI**

Dataiku is the platform democratizing access to data and enabling enterprises to build their own path to AI. To make this vision of Enterprise AI a reality, Dataiku is the only platform on the market that provides one simple UI for data wrangling, mining, visualization, machine learning, and deployment based on a collaborative and team-based user interface accessible to anyone on a data team, from data scientist to beginner analyst.

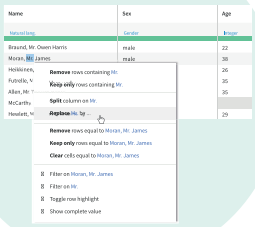
300+
CUSTOMERS

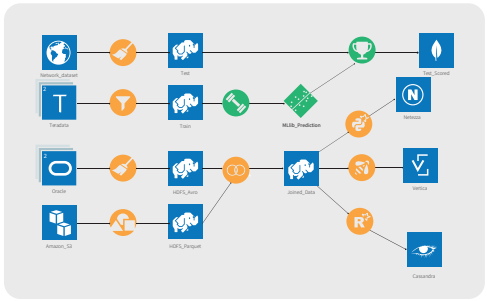
30,000+
ACTIVE USERS

*data scientists, analysts, engineers, & more

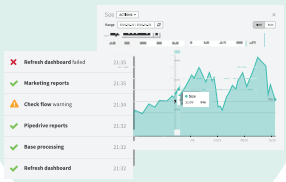
GE · Santander · FOX NETWORKS GROUP · Unilever · KUKA · BNP PARIBAS · paloalto NETWORKS · PREMERA BLUE CROSS · NXP · SEPHORA

1. Clean & Wrangle



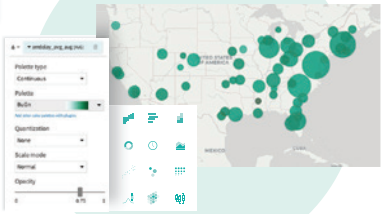


5. Monitor & Adjust

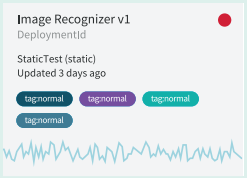


2. Build + Apply Machine Learning

3. Mining & Visualization



4. Deploy to production



WHITE PAPER

www.dataiku.com