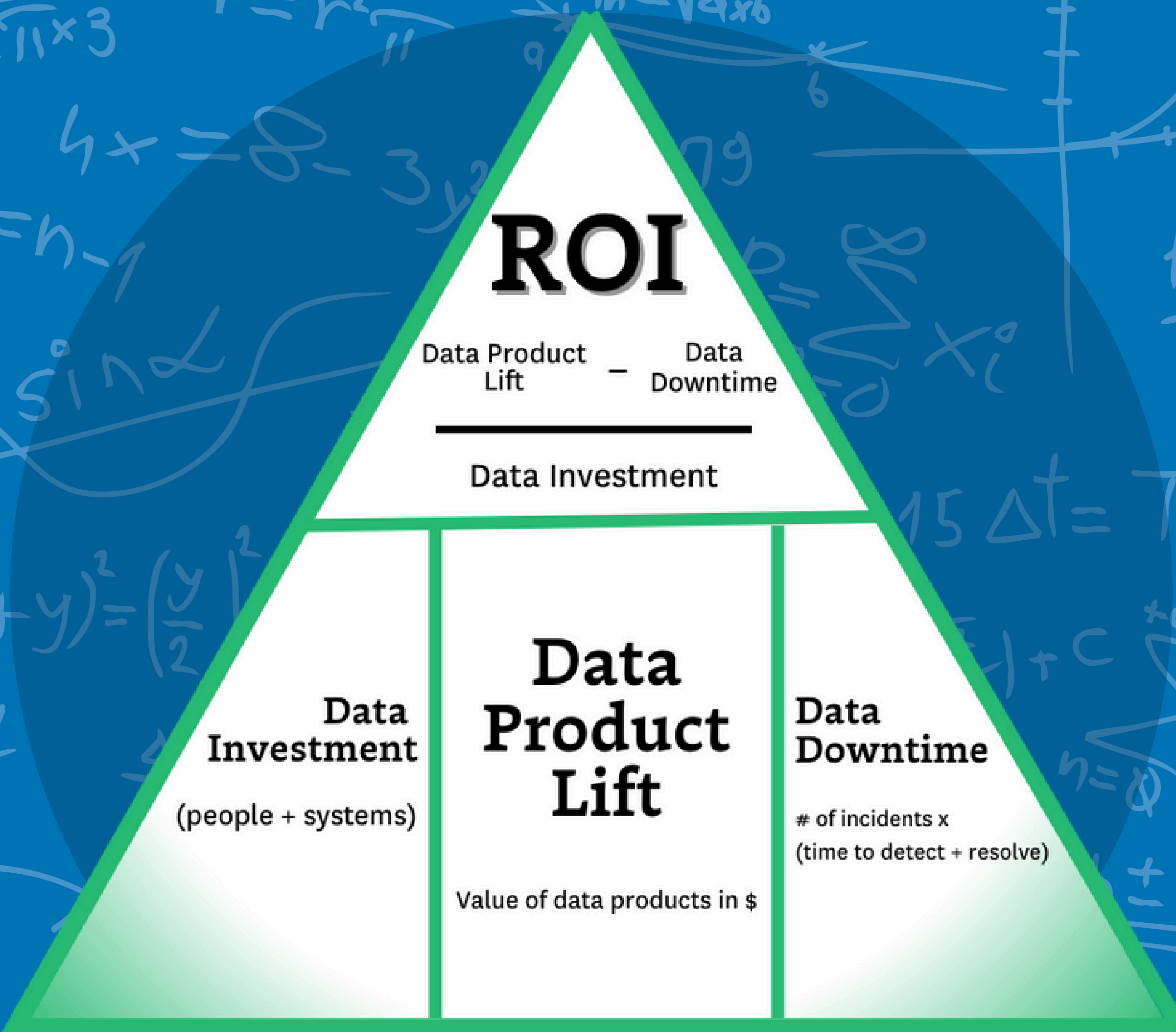
MONTE CARLO

Measuring Data Product ROI



MONTE CARLO

Measuring Data Product ROI

You can’t prove what you don’t measure.

Whether we’re talking about teams, tools, or products—investments that don’t deliver stakeholder value don’t last long in today’s economy.

As recently as a year ago, about half of the data leaders we spoke with felt the business value of their team sold itself. Today? Measuring and maximizing the value of their work is near the top of every data leader’s agenda.

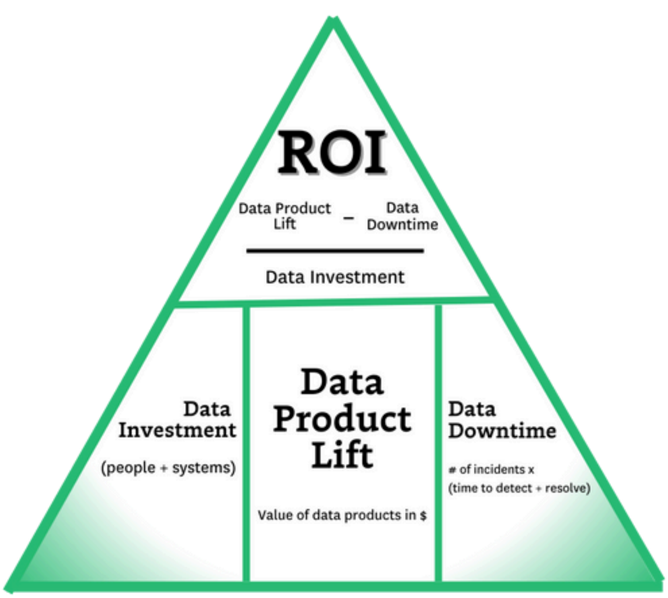
When it comes to evaluating data products, data teams need to take a quantitative approach to delivering value. And that mean’s measuring ROI.

In this guide, we’ll show you how to calculate and improve the return on your data products, including identifying critical assets, understanding data product adoption, and evaluating reliability issues that may be affecting your bottom line.

The goal of this guide is aimed squarely at helping data leaders

- Get closer to the business
- Balance competing priorities
- And focus on the right metrics to generate value for their stakeholders.

Creating powerful and effective data products in today's modern landscape is no easy task. But with the right metrics in place, you can start to understand where you’re treating your data as a product—and what you can do today to drive more value for stakeholders tomorrow.



First things first, what is a data product?

Before you can calculate the return of a data product, you have to be able to identify it.

Just about any data asset *can* be a data product—from a Looker dashboard to a multi-layered data platform—but not every asset *will be* a data product. Waving a magic wand and declaring an executive dashboard a data product without any real changes to the data's underlying governance or reliability won't do your team—or return calculations—any favors.

Eric Weber, the Head of Data Experimentation at Yelp, suggests, "talking about data products in a generic way can produce generic results. Data product is a useful idea, but to make it really create value, we have to get into the specifics…."

So, let's get specific. Regardless of what data the product visualizes, crunches, or puts to work, a data product will generally provide value in one of five ways:

- Increase data accessibility (surface data where people need it when they need it)
- Increase data democratization (make it easier for people to manipulate the data)
- Deliver faster insights from the data
- Save time for the data team / data consumers
- Drive more precise insights (i.e., experimentation platforms)

Similarly, there are important characteristics or qualities a data product should have.

- **Reliability**. Acceptable downtime for a SaaS product is a discussion of "how many 9s?" As in 99.9% or 99.999% availability. Just as software engineers use products such as Datadog or New Relic to track SaaS product performance, data product managers need solutions to identify and solve data product performance issues in near real-time.
- **Scalability**. The data product should scale elasticity as the organization and demand grows.
- **Extensibility**. While the data product has likely been built from an integration of different solutions, it needs to maintain the ability to easily integrate with APIs and be versatile enough to be ingested in all the different ways end users like to consume data.
- **Usability**. Great SaaS products focus on providing a great user experience. They are easy to learn, fun to use, and quick to get work done.
- **Security and Compliance**. Data leaks are costly and painful, as are regulatory fines.
- **Release Discipline and Roadmap**. SaaS products continually evolve and improve. Roadmaps are built at least a year into the future with a strong quality assurance process for updates.

Now, let's take a look at how metrics can help us validate the value of our data products.

Calculating Data Product ROI

As a general rule of thumb, the higher you report into an organization, the fewer and more encompassing your metrics will become. The CEO doesn't care how many dashboards you support or your data freshness SLA adherence percentage.

They want to know what their investors want to know, "am I getting a return on my investment?"

Most data team ROI formulas focus on some version of the following calculation:

Lift / investment = ROI.

And while there's certainly value in its simplicity, it misses one critical component of a data product's value.

Instead of limiting the equation to a binary of lift/no lift and investment/no investment, a true data product ROI calculation will take into account the value of the data itself. With that perspective in mind, the basic equation for calculating Data Product ROI is as follows:

(Data product lift – data downtime) / data investment = ROI

While this equation is *similar* to the equation mentioned above, it includes one additional variable that directly impacts the return a data product is able to deliver:

Data downtime.

There's a direct correlation between the return of a given data product and the impact of data downtime. The more revenue a data product drives—via machine learning models, customer-facing apps, data democratization, and other initiatives—the more severe the consequences of data downtime will become in terms of lost time, revenue, and trust.

And as data downtime decreases, data product return increases.

So, now that we have a framework for calculating ROI, let's dive deeper into the specifics of each variable.

Data Product Investment

Determining the investment for a given data product is a pretty easy formula:

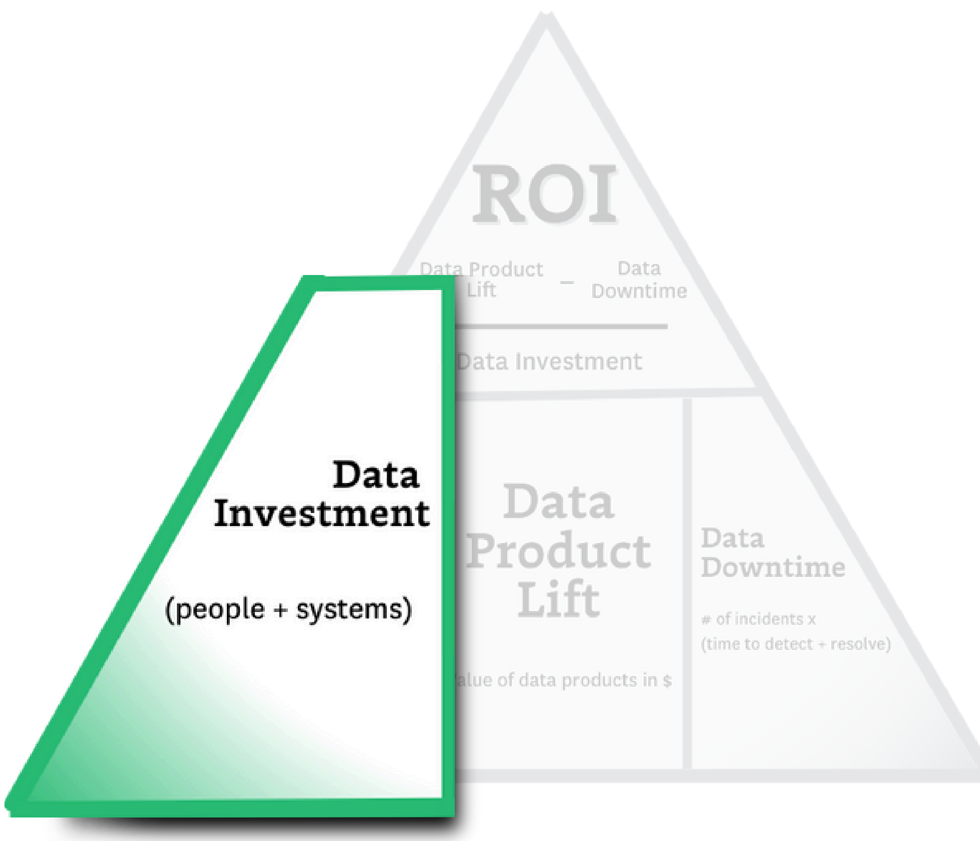
Investment = people + solutions.

But because this step in the formula is so simple on its face, it's also tempting to overcomplicate.

Some contracts are annual. Others are not. Some solutions charge based on usage. Others don't.

The best course of action is to keep this component relatively simple. Stick to an aggregate projection of costs divided evenly across a time period (typically a month or quarter). This should include the cost to stand up the data product and ongoing maintenance.

Note: If you haven't yet built your data product, you could also use cost to build and projected return to calculate pay back period and determine viability for a potential data product.

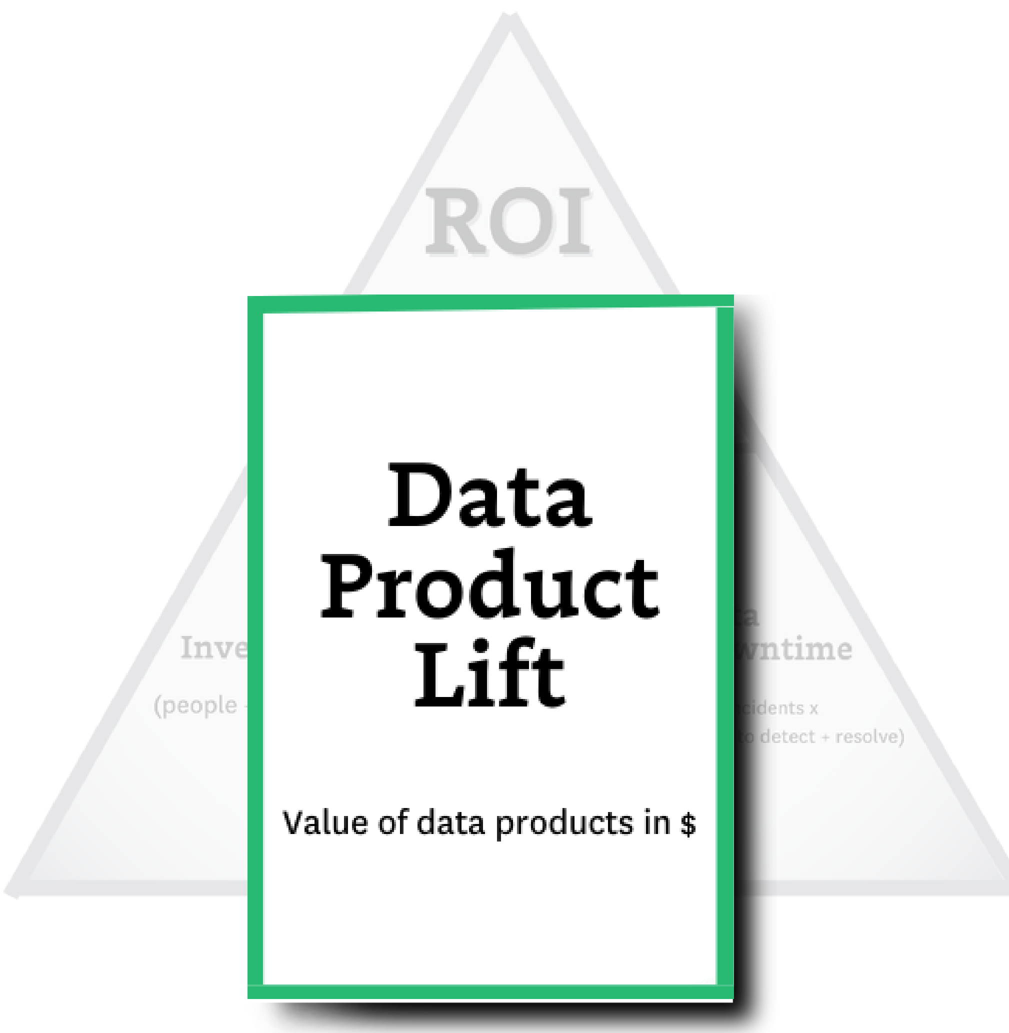


Data Product Lift

Calculating data product return is by far the most complex step of the ROI calculation. That's because as the data industry continues to advance, the variety and complexity of data product use cases has advanced right along with it.

Fortunately, data products can generally be split into one of three primary categories: analytical data products, operational data products, and customer-facing data products. These can exist in the form of dashboards, ML models, experiments leveraging data insights, and—of course—generative AI. While the latter may be a bit more complex to build, genAI is still a data product at its core and its value can still be calculated using the methodologies we'll outline below.

On the next page, we'll outline two simple calculations you can use for Data Product Lift, followed by a couple nuances based on the type of data product you're evaluating.



Two methods for calculating lift

Unlike user satisfaction which is a subjective sentiment from your downstream consumers, “lift” presents a quantitative metric for how your data products are impacting your organization. The key to making this metric valuable is anchoring the result in real business outcomes: in other words, converting your lift calculation into a dollar value.

For example, if the value of your lift is calculated in *time saved,* determine the approximate cost of an hour based on the team that was impacted and multiply that dollar value by the total number of hours saved. Here are two very basic lift formulas.

Metric	Definition And Calculation	A Note From Monte Carlo
Performance Lift versus Control	Performance lift is measured by subtracting the baseline performance for a given KPI from performance for teams leveraging a given data product. For example, if time-to-value is 3 months for a team that’s adopted a given data product and 6 months for your control, the lift in time-to-value would be 3 months [(3 months)-(6 months)= 3 months saved]	This metric can demonstrate the impact of data products across teams and offer critical data to encourage adoption. This calculation is best used to compare two or more teams or individuals with similar goals and KPIs. Key metrics to consider here are impact in dollars, time-to-value, and impact of insights.
Performance Lift Period-Over-Period	Another way to measure lift is period-over period. For this lift calculation, you would subtract performance from the previous period before releasing your data product to performance since release. For example, if closed-won sales were 15% pre-data product and 25% post data product, the performance lift period-over-period would be 10%. [(25% closed-won)—(15% closed-won)=10% lift in closed-won sales]	Understanding the period-over-period impact is a great way to understand the impact of a data product *within* adopting teams. Unlike baseline performance, this metric is best used when attempting to understand how a product has performed against a stakeholder’s own benchmarks.

Unfortunately, no two data products are the same. And there’s a good chance your data product use case won’t be as simple or straightforward as these two calculations. So, let’s take a look at how lift calculations might work themselves out in a few different scenarios.

Analytical lift

Analytical data products are a mix of the critical dashboards, ML models, and experiments conducted and supported by your data team to deliver insights for decision making.

Whether we're talking about the marketing dashboards or important metrics like customer LTV, analytical data products play a fundamental role in the day-to-day operations of any business.

Measuring incremental impact

Measuring absolute impact is one of the simplest ways to understand return from any data product. And calculating the value of an experiment conducted as a result of your data team's research and analytical insights can be quickly estimated by understanding the delta between test versus control and translating those numbers into dollars earned/saved each month.

For a more conservative approach, you could estimate value by computing the return against a random or average decision to better represent decisions made without support from your data team.

Combining tens or hundreds of these experiments per year will give you a ballpark figure for the incremental value delivered by the experimentation platform and the analytical work surrounding those experiments.

Measuring value-to-stakeholders

But what about dashboards? Rarely are these initiatives so easily measured by a controlled or natural experiment.

To account for these data products, we'll need to estimate value. In this case, we'll be translating qualitative data into something representative by tapping into the consumers themselves.

Your business users and data consumers are actually quite knowledgeable about how valuable your dashboard is (or isn't) to them, and those opinions can be collated and quantified. For example, MIT economists asked respondents how much they would have to be paid not to use Facebook or Google Maps for a year and then used that as their value metric.

For the most important dashboards, data teams can go a step further by creating a benchmark for respondents, like "We estimate the cost of maintaining this dashboard to be about $5,000 last quarter. In your estimation did it add that level of value to your work over that period of time?" For a baseline, here is how our survey of 200 data professionals revealed they judged their data consumers would value their dashboards:

- Less than $500k: 5%
- 500k-1m: 11%
- 1m-10m: 49%
- 10m-25m: 32%
- 25m+: 5%

Customer-facing lift

Here I'm referring specifically to *data* that's customer facing, not the ML models powered by data. This data use case generally comes in two flavors.

The first is when data IS the product. There are a significant number of businesses that ingest, transform, and then sell data to other companies. It could be a data mining company compiling insights from web scraping ecommerce sites or a television manufacturer that sells viewership data to advertisers.

In this case, calculating is pretty straightforward: the revenue of the data product is the revenue of the sale. When you find ways to enrich this data, you make it more valuable and thus increase the sale price.

However, what about cases where data is only part of the product being offered? For example, a point of sale system providing insights back to the merchant on their foot traffic patterns? Or a video player that breaks down views across audience segments by time?

In some cases, the data will be a nice to have. In other cases, it will be a significant factor on customer acquisition and retention. Luckily, data teams have been experimenting and measuring the impact of features on retention for a while now.

Operational lift

Operational data use cases are activities that MUST take place. Examples would include reporting to the board or an airline re-accommodating passengers of a delayed flight.

If the data systems went down, these activities would still happen, but they would be considerably more painful. An organization may have to manually collect and aggregate data from across the business for its report or passengers may need to go to the customer service desk rather than have an app automatically present their options for re-accommodation.

In these cases the value is typically best determined by the hours saved between the more automated and more painful process. In some situations, alternative impacts such as fine avoidance or poor customer satisfaction could be calculated as well.

The lift of machine learning applications is an important component of this. These applications often support the user experience (e.g. recommenders, ad targeting) and directly drive revenue. It's worth noting that precision is not the objective here – you should be aiming to get an understanding of the approximate model lift, and how that translates into dollars made or saved each month.

Data Downtime

Finally, we need to understand how data downtime impacts ROI.

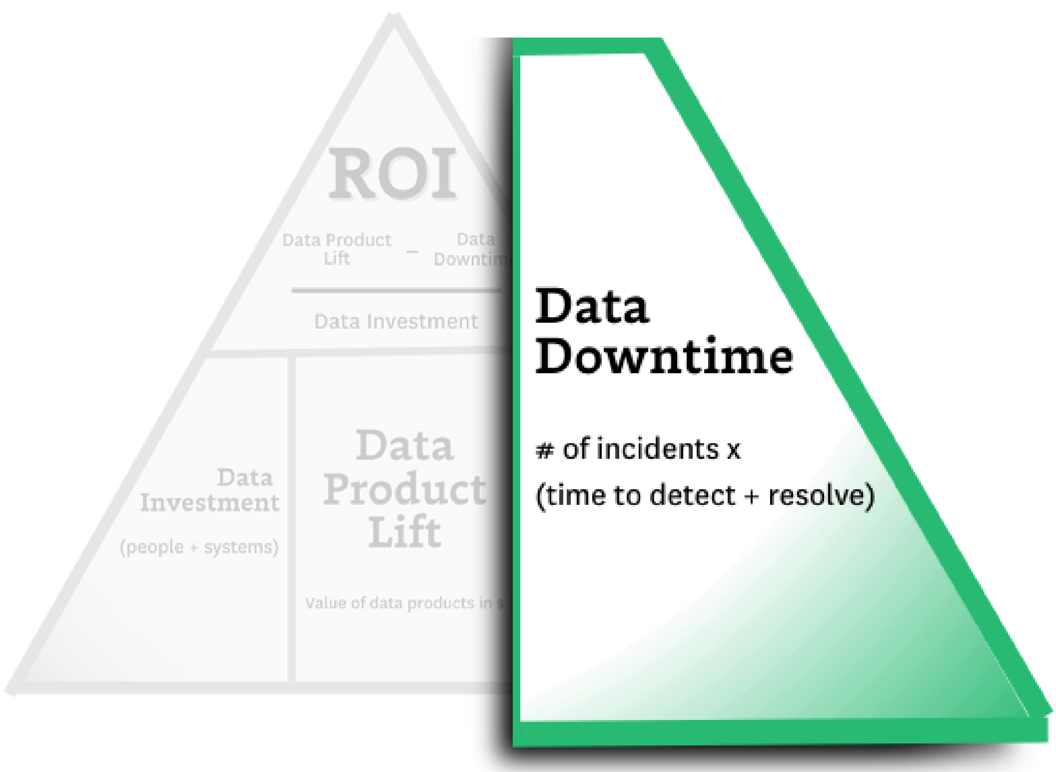
Here's how to calculate data downtime using your incident and response times:

Number of incidents x (average time to detection + average time to resolution)

This is helpful in measuring how your overall data product reliability is trending. But in this case, we aren't as interested in the aggregate data downtime or the efficiency of the team (yet). What we want to find out here is the operational cost for the data downtime of specific data products.

Since we've already calculated the revenue from our data product, we can now subtract the operational cost of that downtime from the revenue.

For this component of the ROI calculation, I recommend only focusing on downtime that violates your data SLAs. If a dashboard that's checked daily has a data freshness issue that only persists for a few hours before being resolved, that downtime is unlikely to have an operational impact on the organization (and your data team shouldn't be penalized for it).



MONTE CARLO

How to Improve Data Product ROI

If your Data Product ROI calculations leave your stakeholders wanting, don't worry! There are steps you can take to get those data products back on track and aligned to the right business outcomes.

In the next couple of pages, we'll look at a handful of metrics you can monitor to make sure your data products are aligned to stakeholders and needs and delivering on their promises.

We'll also arm you with some tips and tricks from other data leaders that will help you shore up your data products for the long haul.

Reduce Data Downtime

When your data products aren't reliable, they aren't valuable. So, if you notice that a particular data product is underperforming relative to its importance, it's time to take a second look at reliability.

Improvements in data downtime can have a dramatic impact on your data team's ROI calculation, especially for use cases when the data is so central that data downtime is equivalent to operational downtime. Here are some quick calculations you can use to dig a little deeper into your data product health:

Quality Metric	Definition And Calculation	How It Relates to ROI
SLA Adherence	Minutes the data product met SLA standards divided by the minutes it was in violation of that SLA. You won't achieve the 99.999% uptime of software applications, but the high 90s are a good North Star.	SLA adherence, the amount of time the SLA has been breached divided by the amount of time it has been upheld, can help data teams maintain a granular understanding of how many data quality incidents are having an adverse impact on specific data assets—and what measures to take top protect the value of those products. For example, a financial analytical data product may require high precision but medium data freshness, where a machine learning model may be the opposite. See how Red Ventures leveraged data SLAs for data quality.
Docu-mentation	Percentage of data products with documentation covering ownership, SLAs, business logic, and incident history.	Without strong documentation, the only people that can build with, maintain, or fix data assets are the data engineers that built them in the first place. Better documentation wil have an indirect but noticeable impact on data product ROI.
Quality Issues Detected	Total number of data quality issues detected for a given data product.	Data quality issues are tantamount to data downtime, so reducing this number should have a direct impact on data product return and will decrease as your team invests in the right data quality solutions.
Status Updates	The total number of alerts generated for a data product divided by the number of alerts that have a resolved status (fixed, no action needed, expected, false alarm).	Each data quality incident SHOULD be documented and kick off a root cause analysis investigation. Resolving data quality issues faster will limit the impact of any one issue and improve value-generating uptime for your critical assets—so the closer this number approaches to 100% for any one data product, the better. Note: Make sure you have formalized processes and tooling in place to detect and resolve incidents quickly.
Relative Data Quality Cost	Total hours spent on detection and resolution for a data product divided by the total amount of time spent on data quality across all data products.	This metric covers time spent on avoidable tasks like writing and maintaining data tests. Assuming your data team is investing time to resolve data quality issues, this number will decrease as your data products become more reliable. However, this metric is always relative to the health of your data in aggregate. For example, a data team who spends 40% of its time on data quality with 25% on a single data product will have more data quality pain than a team that spends 15% of its time on data quality with 30% on a single product. See how Prefect saved 50% of their data engineering time.

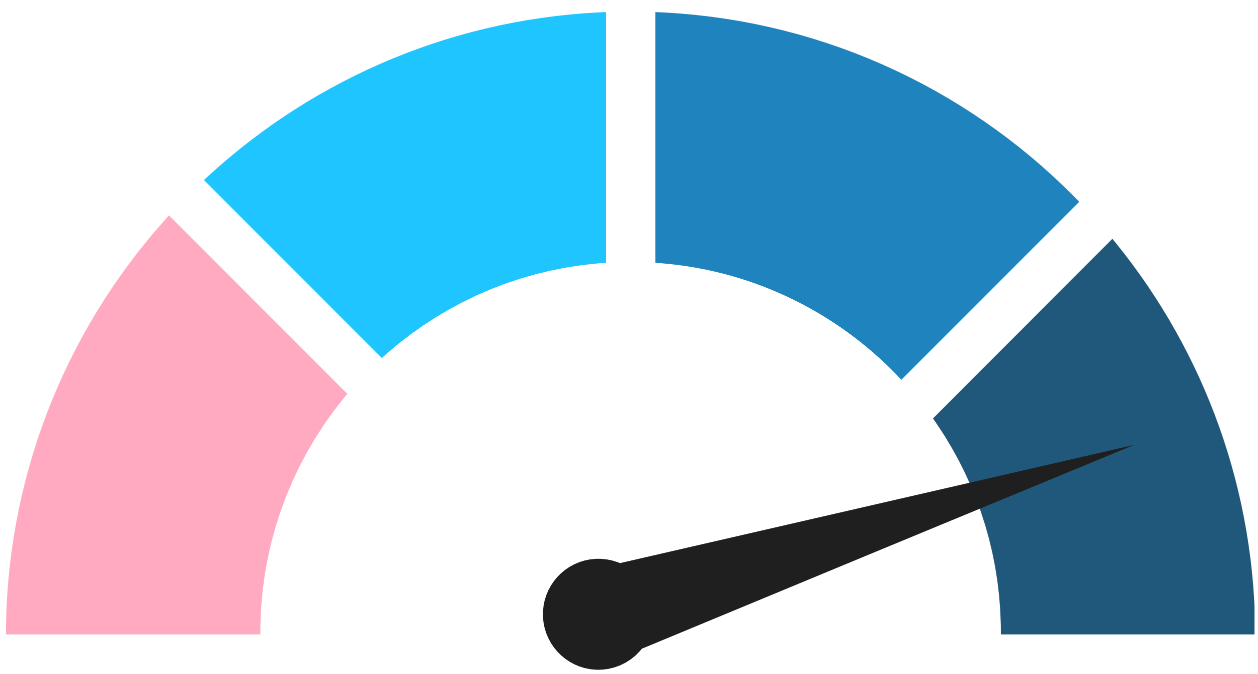
Improve Satisfaction and Adoption

Know a data product is important, but not seeing the returns you expected? You might have a data product adoption problem.

Data products begin and end with your downstream consumers. A product in search of a consumer does not a product make. So, a critical component of delivering ROI from your data products is understanding how your products align to your customers.

An NPS survey is a great way to quantify the subjective sentiment of your business stakeholders. If your data product alignment is high, you should see correspondingly strong scores across trust, value, ease of use, and satisfaction. You should also see improvement in your data adoption numbers as well.

Since this type of product research is a relatively new practice for data teams, you won't find many industry benchmarks to compare your results to just yet. But as you make the NPS survey a more standardized step in your data product management process, you'll be able to easily compare your past and present performance to track progress as your data products evolve.



Data Product Satisfaction and Adoption

Metric	Definition And Calculation	How It Relates to ROI
Data Product Utility	Ask , "On a scale of 1-10 how well would you say this data product does at answering your specific questions or solving for your use case?"	A data product may drive value, but is it driving the value your consumers need right now? Understanding how well your data product addresses your consumers use-case will go a long way towards estimating the value it's likely to drive for that audience.
Trust	Ask data consumers, "On a scale of 1-10 how likely would you be to describe this data product as accurate and reliable?" All responses with a 9 or 10 score count as +100. All responses with a 1-6 count as -100. Total and average.	Ideally you should see trust levels rise as you invest more time and resources into your data products. JetBlue's data team for example regularly conducts a survey of their data consumers to measure their level of trust in the data products they provide.
Value	Ask, "On a scale of 1-10 how likely would you be to describe this data product as essential to the performance of a business operation or toward optimizing performance?"	Understanding the value your consumers place on a given data product is the first step to understanding how much time you should invest. Machine learning models offer hard returns that are easy to quantify, but understanding the "pay off" across dozens or even hundreds of analytical data products is a bit more challenging. For teams with more diverse data products, a survey like this is essential.
Ease of Use	Ask , "On a scale of 1-10 how likely would you be to say I can easily obtain the data I need?"	A data product may be useful, but if it's not also easy to use, it's unlikely to be adopted by downstream consumers . Can data consumers get the information they need? Can your data power-users (data scientists, analysts, etc) easily navigate the data warehouse or is it a data swamp? How well documented are your assets? All these questions factor into how easily your consumers can realize the value of your data products.
Satisfaction	Average your trust, value, and ease of use Net Promoter Scores for a top-line satisfaction number.	It can be helpful to roll up value, ease of use, and trust into a top-line metric like "data product satisfaction" to report up to higher levels of the organization.
Adoption	The overall adoption from users internal to the organization as measured by monthly active users of dashboards, ML models, and other data products. To find this metric, divide the number of employees at your organization by the number of monthly active users for a given data product.	It's difficult to treat your data-as-a-product if you aren't monitoring one of the most important metrics for product teams: user adoption. This metric will go a long way towards telling you if you've made something useful or if you've got some more work to do. If you aren't seeing adoption, look at the dimensions of data trust, ease of use, and value for potential direction to improve your data products. If these scores are high, but data product adoption still remains low, you might just have a messaging problem—consider how to improve your internal communications to educate consumers about the product.

Other Data Product Tips & Tricks

Curious how other data teams are handling their data products? Check out some of the insights we've collected from data leaders at companies like Meta and Thoughtworks.

Create a vision for positive change

According to Afua Bruce, Former Data Strategy Lead for the FBI, when you're building a new data product you need to create a vision that helps stakeholders understand *why* they should care. "You may not spend a lot of time reading those wonderfully worded, often beautifully designed pages on your company's website, but you'll want to refer to them as you're building a business case for the work that you're doing."

By leading this conversation internally, you can more effectively evangelize your product to your organization and improve adoption across relevant stakeholders.

Define and enforce the attributes of your data products

As you've probably seen from the metrics above, consistency is key to improving your data products for stakeholders. And the first step to consistency is good documentation.

According to the team at Roche, data products should be produced with a description, transformation script/code, and access policy. This looks and operates similar to data contracts with the goal of defining and enforcing the meaning and structure of a data product.

At Roche, this information is automatically added to the data catalog whenever a product is published with a custom front end portal for searchability and findability. The catalog entry includes policies, tags and data sources that were previously defined and fed into their access management tool related to the product.

More Tips & Tricks...

Identify reusable data products

The two best ways to fail at creating valuable, reusable data products are to develop them without any sense of who they're for and to make them more complicated than they need to be. It's likely that your data pipelines follow some sort of Pareto distribution where 20% of your tables receive 80% of the queries. The team at Thoughtworks suggests that by discovering the tables with the most overlap in usage, data teams can deliver value more efficiently by creating data products that serve multiple teams at once.

For example, if the table customer_accounts is constantly being queried by marketing, finance, support and other domains, that can be taken as a signal that building a data product that consolidates the necessary information into a full 360 view may have shared utility.

Employ a data product manager

Wouldn't it be great if there was a role that could be responsible for wrangling all this disparate information into the development of your data products? Well, good news—there is!

Instacart Product Lead Atul Gupte recommends electing a data product manager to identify holes in your data products and then working with data and analytics teams to bridge the gap.

According to Atul, data product managers should be responsible for setting the vision, prioritizing projects, and leading the charge toward enabling data consumers to effectively operationalize their data.

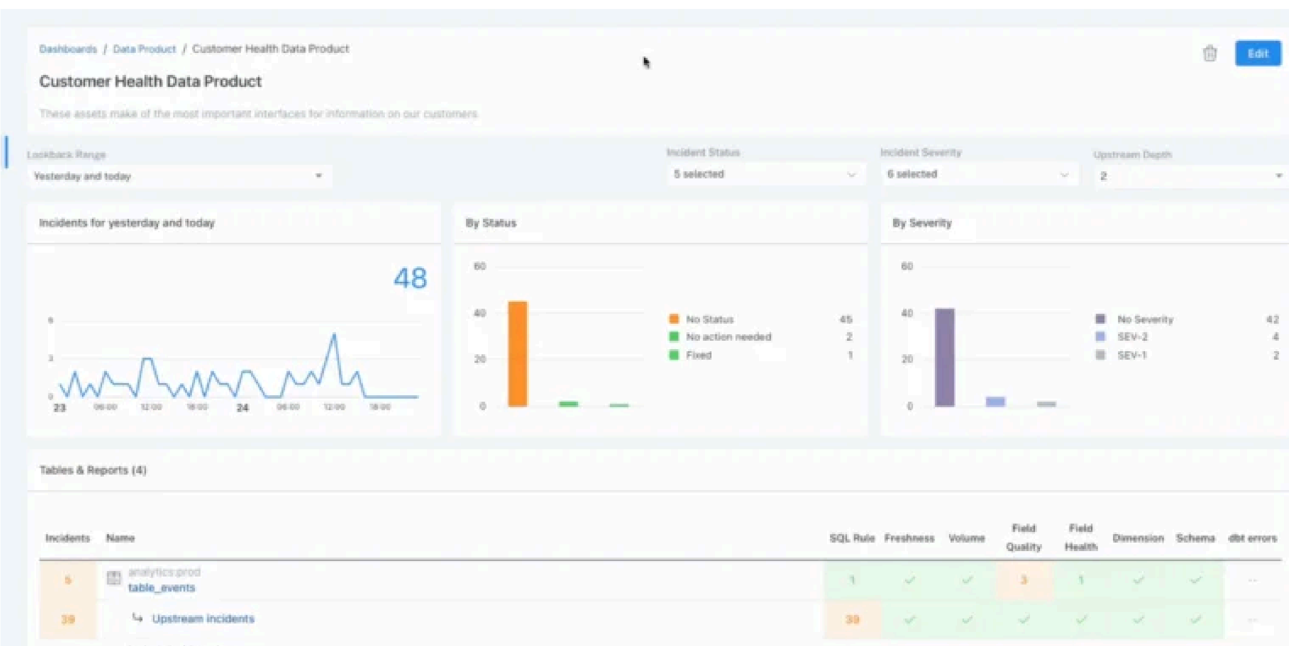
A data product manager is arguably one of the single greatest factors in your ability to effectively optimize your data products. So, put that hire at the top of your list.

Next Steps And How To Get Started

While calculating the ROI of a data product will never be an exact science—or even a quick win—it's absolutely a worthwhile endeavor. And perhaps most importantly, it's an achievable destination.

There are few activities more existentially critical for data leaders. By better quantifying and optimizing the value of your data products, you'll be better equipped to not only cultivate the trust of your colleagues, but deliver recognition for the profit-driving work your team adds to the business.

Modern data solutions like a [data observability] platform and Monte Carlo's Data Product Dashboard (pictured below) can also go a long way towards improving the return of your data products by promoting data engineering efficiency, mitigating risk, and improving the health of their underlying data.



And remember, building great data products that deliver ROI is a marathon, not a sprint. That means measurement should be a regular part of your data team's process. So, build a sustainable data product management motion that you can stick to for the long haul—because the more value your data team drives, the more data products you'll find yourself supporting.

Interested in learning about how Monte Carlo can help you improve the health of your data products? [Schedule a call] with one of our data product experts.

MONTE CARLO

Build better data products with data observability.

Ready to unlock the real secret to better data products? Request a demo to see how Monte Carlo can help you deliver data products your data consumers can trust.

Request a demo

Looking for more tips & tricks? Check out some of our other resources.

Data Downtime Blog

Data Quality Fundamentals

Follow us on LinkedIn

Subscribe to our YouTube Channel

Monte Carlo is the world's #1 Data Observability Platform