EBOOK

The Economics of AI

How to Shift Data Projects from Cost to Revenue Center

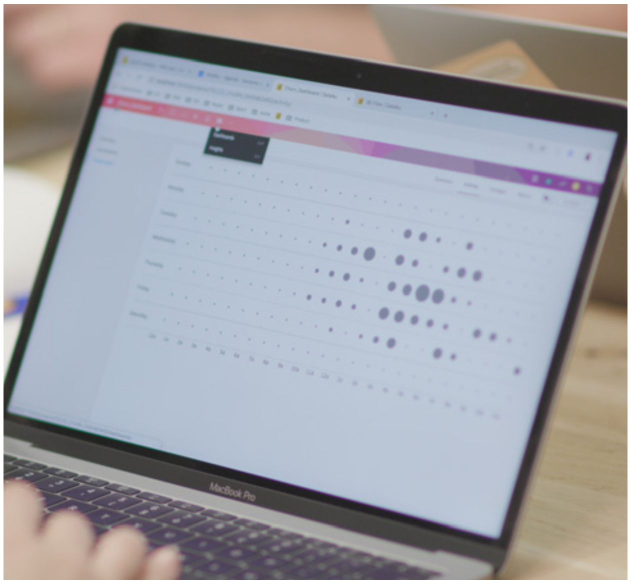
www.dataiku.com

Introduction

In the coming years, the ability of organizations to pivot their activities around Enterprise AI will fundamentally determine their fate. Those able to efficiently leverage data science and machine learning techniques to improve business operations and processes (as well as to find new business opportunities) will get ahead of the competition, while those unable to shift will fall behind, perhaps swept away with the tide of rising costs and diminishing revenue.

> *Three out of four C-suite executives believe that if they don't scale artificial intelligence (AI) in the next five years, they risk going out of business entirely.*
> *- Accenture Research Report, AI Built to Scale[1], November 2019*

Of course, the key word here is efficiently; it's not enough for organizations to simply leverage Enterprise AI techniques at any price. Eventually, in order for Enterprise AI strategy to be truly sustainable, one must consider the economics: not just gains, but cost.



This white paper will delve into:

- The economics of AI: Why scalability isn't simply a matter of more use cases.
- The ongoing costs of Enterprise AI: What they are and why they exist.
- Capitalization and reuse: How these techniques can be used to reduce costs and ultimately scale Enterprise AI efforts.

[1] https://www.accenture.com/ro-en/insights/artificial-intelligence/ai-investments

The Economics of AI

When organizations take their first steps into the Enterprise AI world, the most common technique is to start with a finite list of select use cases, ideally optimized for a balance between difficulty in execution vs. potential impact. The initial entry point might be costly in terms of technologies and change management, but assuming the use cases are operationalized, the economic value is generally positive.

In fact, Accenture's AI: Built to Scale[2] uncovered that companies in the study that use this multi-use case approach to get started report nearly three times the return from AI investments compared to companies pursuing siloed proof of concepts.

Real-World Example

GE Aviation focused specifically on their self-service data program as a first step in Enterprise AI with a goal of scaling and streamlining (i.e., making processes smoother and business operations more efficient overall). With these first efforts, they have been able to quantify their efficiencies and savings via the self-service data program to the tune of millions of dollars.

But what happens after companies find success with this first list of use cases? Well, they tend to repeat the process, adding more and more use cases. Getting to the tenth or twentieth AI project or use case usually still has a positive impact on the balance sheet, but eventually, the marginal value of the next use case is lower than the marginal costs.

[2] *https://www.accenture.com/ro-en/insights/artificial-intelligence/ai-investments*

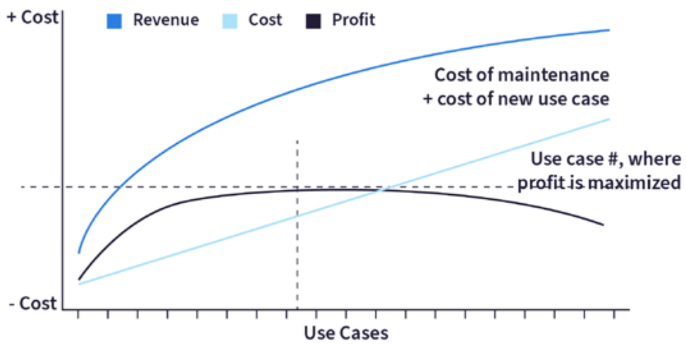


Figure 1: Cumulative revenues, costs, and profits over time (# of use cases); note that after use case #, profit is decreasing due to increased costs and stagnation of revenue.

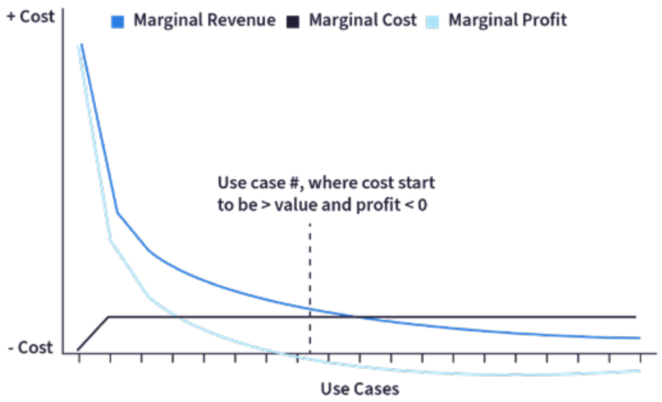


Figure 2: Marginal revenue, cost, and profit over time (# of use cases)

Figures 1 and 2 illustrate that there is a point in time where the economic value of Enterprise AI decreases because:

- *The marginal cost of the supplemental use cases is not decreasing.*
- *The marginal value of the supplemental use cases is decreasing (i.e., the first use case has more value than the Nth use case).*
- *The marginal profit of supplemental use cases quickly becomes negative.*

One might see this analysis and conclude that the most profitable way to approach Enterprise AI, then, is to only address the top five to 10 most valuable use cases and stop. But this does not take into account the continued cost of maintenance of this core of AI projects.

Adding marginal cost to the maintenance costs will generate negative value and negative numbers on the balance sheet. It is, therefore, economically impossible to scale use cases, and it's a big mistake to think that the business will be able to easily generalize Enterprise AI everywhere by simply taking on increasingly more AI projects throughout the company.

Ultimately, to continue seeing returns on investment (ROI) in AI projects at scale, taking on exponentially more use cases, companies must find ways to decrease both the marginal costs and incremental maintenance costs of Enterprise AI.

The Cost of Enterprise AI

Before looking at how organizations can reduce costs, it's important to understand what some of those costs of Enterprise AI are. Of course, there are obvious, tangible costs (like that of tools and technology), which should certainly be managed in order to successfully scale.

But it's the following less tangible costs that tend to bog down organizations' efforts by adding up over time, hampering their ability to scale and profit from Enterprise AI:

Data Cleaning and Preparation

By now, most have heard that the adage that data scientists spend about 80 percent of their time finding, cleaning, and preparing data. Indeed, 43 percent of overall respondents to an **AI Maturity survey by Dataiku** in 2019 said that data cleaning and wrangling is "the most difficult or time-consuming part of data processes at my organization," including 63 percent of the respondents in the C-suite.

What's perhaps most challenging about data cleaning and preparation is that it's a huge task - and, therefore, a huge cost in terms of employee time - that needs to be done for every single use case or AI project.

Data cleaning and preparation are critical parts of an AI project, and if not done well, can translate into poor quality models. So reducing this cost is not necessarily about simply discouraging time spent or outsourcing the work.

Rather, it's about ensuring efficiency, putting systems in place that allow data to be found, cleaned, and prepared once, then used a maximum number of times in different use cases. Instead what often happens for organizations today, **as Data Science Senior Director Chris Kakkanatt explains**, is costly repeated work across people, teams, and the wider organization.

Operationalizing and Pushing to Production

In the process of operationalization, there are multiple workflows: some internal flows correspond to production, while some external or referential flows relate to specific environments.

Moreover, data science projects are comprised of not only code, but also data (including code for data transformation, configuration and schema for data, public referential data, and internal referential data). That's why, to support the reliable transport of code and data from one environment to the next, they need to be packaged together.

Consistent packaging, release, and operationalization is complex, and without any way to do it consistently, it can be extremely time consuming. Dataiku survived more than 200 IT professionals, asking "On average, how long does it take to release the first version of a machine learning model in production?" More than half said "between three and six months."

That is a massive cost not only in man hours, but also in lost revenue for the amount of time the machine learning model is not in production and able to benefit the business. Multiply this not by one model but by hundreds, and the cost is debilitating.

Data Scientist Hiring and Retention

Data scientists by nature are curious and entered the field in order to make a difference and to have impact. They are also driven by efficiency, which means they don't like to do things twice if they don't need to.

If all data scientists at the organization are doing is spending time playing with data in a sandbox (never seeing their projects in production having impact on real data), spending time on data cleaning and prep instead of problem solving or cutting-edge technologies, or doing repetitive work, they are not going to be very happy in their jobs, and in turn, the company will spend a lot of money dealing with constant turnover. Reducing this cost is a matter of proper tooling: providing the resources for staff to capitalize on past projects and reuse work.

Cloud Costs and ROI

As organizations' data teams grow and as more staff outside of data teams starts working with data, having a modern approach to architecture that allows for scaling up and down of resources is critical. Indeed, elasticity will be the name of the game in 2020.

Even though the cloud is growing in popularity, most companies will take a hybrid approach, investing in AI platforms that sit on top of the underlying architecture to provide a consistent user experience for working with data no matter where it is stored. Yet in early 2019, The Information reported[3] that more and more companies find themselves surprised by their rising cloud costs.

Ultimately, companies that don't actively work on developing a larger cloud strategy and managing cloud costs in 2020 will face an uphill battle to prove positive ROI with AI projects, racking up a bill that isn't offset by the financial gains or savings from the projects themselves.

Model Maintenance

Machine learning models are not like software code where they can be put in production once and work, untouched, until something about the system fundamentally changes. Data is constantly changing by nature, which causes models to drift over time.

That means continual AI project maintenance cannot be ignored (or at least not without an effect on profit). Depending on the use case, the model can either become less and less effective in a best case scenario; in the worst case, it can become harmful to the business.

Maintenance becomes even more challenging the more use cases the company takes on, which even further drives up the costs. MLOps has emerged as a way of controlling the cost of maintenance, shifting from a one-off task handled by a different person - usually the original data scientist who worked on the project - for each model into a systematized, centralized task.

[3] https://www.theinformation.com/articles/as-aws-use-soars-companies-surprised-by-cloud-bills

One additional sub-dimension under model maintenance is the cost of maintaining infrastructure. That is, the maintenance of models necessarily requires maintaining the infrastructure on which they run. Given the speed at which the average organizations' infrastructure stack is evolving, this can become costly (both in terms of time and money) quickly.

Complex Technological Stacks

It's not just infrastructure that needs to be maintained; seemingly all technologies in the AI space are moving at the speed of light. That means switching from one to another happens often, and when it does, it can be costly. Also, in large organizations, different teams or different geographies might be using completely different technologies altogether.

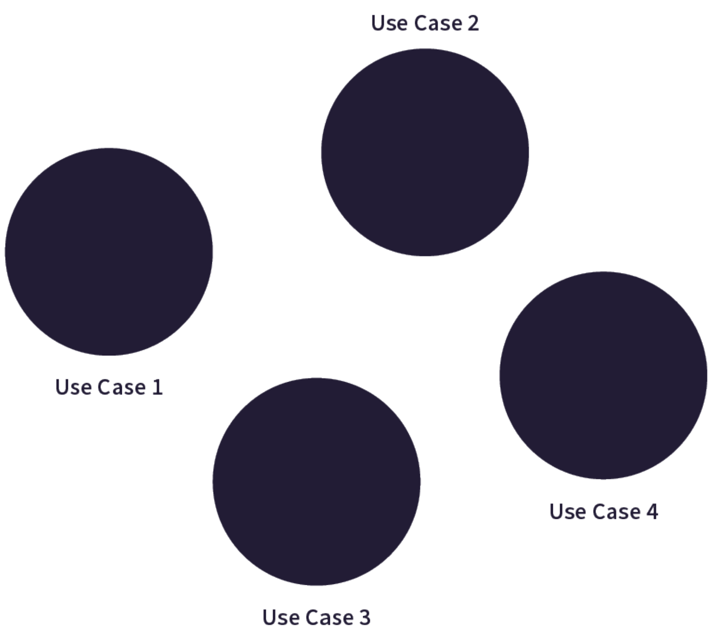
Without the ability to stitch together the larger technology picture - and allow reuse and sharing of knowledge across these teams or geographical - things can get even more expensive at scale.

Capitalization and Reuse

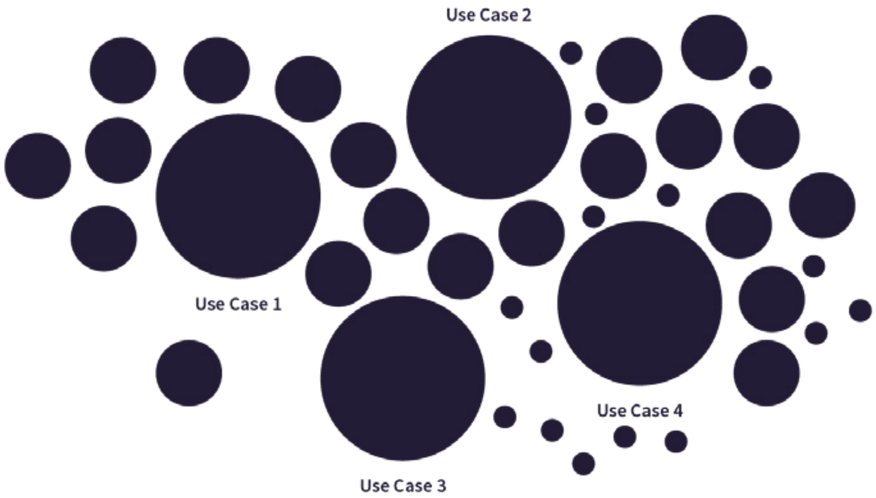
Common sense and economics tell us not to start from scratch every time, and that is exactly the principal behind reducing costs associated with data cleaning, preparation, operationalizing, model maintenance, and even hiring woes.

Reuse is the simple concept of avoiding rework in AI projects, from small details (like code snippets that can be shared to speed up data preparation) to the macro level (like ensuring two data scientists from different parts of the company aren't working on the same project). Capitalization in Enterprise AI takes reuse to another level - it's about sharing the cost incurred from an initial AI project (most commonly the cost of finding, cleaning, and preparing data) across other projects, resulting in many use cases for the price of one, so to speak.

But how exactly do reuse and capitalization ensure scale? By increasing the number of use cases addressed with AI projects while reducing the impact of the costs outlined above. For example, say the business has a list of four uses cases in mind to start their Enterprise AI efforts:



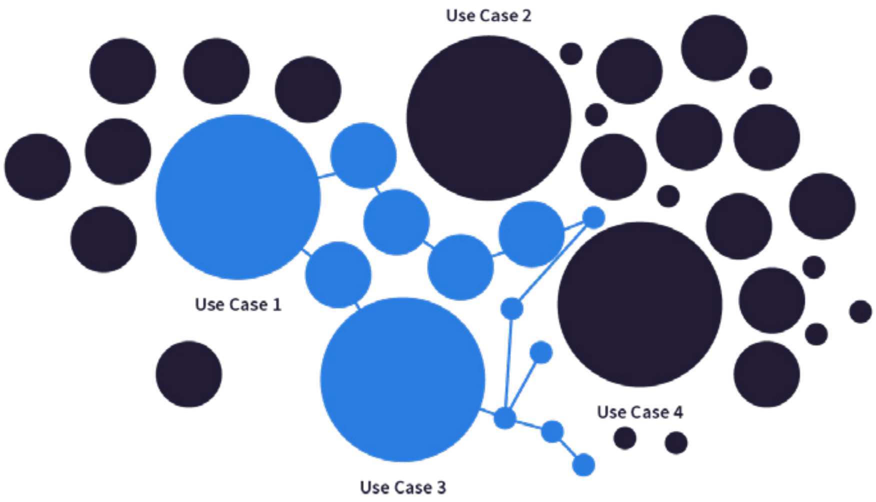
In addition to these four, of course, there are lots of other potential use cases across the business:



Capitalization means that while tackling these larger, high-priority use cases, the organization can also take on lots of other smaller use cases by reusing bits and pieces, eliminating the need to reinvent the wheel with data cleaning and prep, operationalization, monitoring and - in doing all of that - ensuring that data scientists are happy, spending their time on high-value tasks.

It can also spurr the discovery of hidden use cases. By capitalizing on the work of existing projects to spin up new ones, teams might find previously untapped use cases that bring a lot more value than expected, opening up businesses to new possibilities (and sources of profit or savings).

In addition, the surfacing of these hidden use cases via reuse often comes from the work of analysts or business users; i.e., it is one of the keys to unlocking data democratization, where it's not just data scientists that are bringing value from data. This idea often goes hand-in-hand with self-service data initiatives.



Capitalization and reuse sound easy in principal, but in practice, they require strong, enterprise-wide, centralized processes where:

- People can easily access information, including who is working on what projects.
- People can transparently consume things done by others (including seeing data transformation, models, etc.)
- People can take, reuse, and adapt work done by others.
- Data experts can capitalize (and monitor) a portfolio of data treatments to be used across the organization.
- Data experts can easily build and share projects to be used by others.
- People - whether coders or not - can work efficiently in his or her preferred way.
- Data leaders can ensure the quality of AI projects, ensuring that capitalization and reuse are being used properly.

To many organizations, this is a scary list - it's a level of transparency that many find uncomfortable. Some industries are hampered by regulations that make transparency more difficult, but certainly not impossible (see: **Executing Data Privacy-Compliant Data Projects**). But ultimately, it's the level of transparency one needs in order to execute on capitalization and reuse and profit from Enterprise AI efforts.

It's also worth noting that, in order to build a solid and scalable Enterprise AI foundation, these principals must be considered and built into the strategy from the start - not after the nth use case when costs are already impacting profits.

Capitalization and Reuse with Dataiku

Data science, machine learning, and AI platforms like Dataiku are tools to enable Enterprise AI by allowing people within the organization to scale, providing transparency and reproducibility throughout - and across - teams and all of the dimensions previously touched upon in the costs section:

Cost	Mitigated with Dataiku via...
Data Cleaning and Preparation	Reuse of already cleaned and prepared data across projects as well as between personas (i.e., data scientists can use data prepared by analysts).
Operationalizing and Pushing to Production	Reuse from design to production (i.e., without the need to recode models and pipelines from scratch to operationalize).
Data Scientist Hiring and Retention	Reuse of project elements across users, allowing data scientists to spend more time on higher-value tasks (which also happen to be more interesting for them, which means lower risk of brain drain) as well as ability to code in preferred languages and leverage open-source tools.
Model Maintenance	Automated scenarios, monitoring, and reuse of infrastructure across technology stacks.
Complex Technological Stacks	Freedom to reuse and adapt even across changes in technology with Dataiku as an abstraction layer, freeing people from the underlying technology.

A tool that can enable capitalization and reuse should provide at a minimum:

- Robust documentation so that contributors can explain what has been done in a specific project via wikis, to-do lists, versioning, activity logs, etc.
- A built-in, centralized, and structured calolog of data treatments (from data sources to data preparation, algorithms, and more) for easy consumption.
- The possibility to grab parts of data projects and input them in new projects or mix components of two different projects together.
- The possibility for advanced users to package data products as plugins to be used by others without the need to understand all the underlying complexities.
- An advanced console to monitor usage, versions, and quality to ensure easy and efficient operationalization.
- The possibility to automate scenarios using complex triggers as well as automate test production and deployment.

Dataiku provides all of these things while adding one extra layer: user interface accessible to anyone on a data team, from data scientist to beginner analyst. True inclusivity and democratization of data efforts brings capitalization and reuse to another level, as it's no longer just a question of data scientists, but of reuse by everyone across the organization.

Conclusion

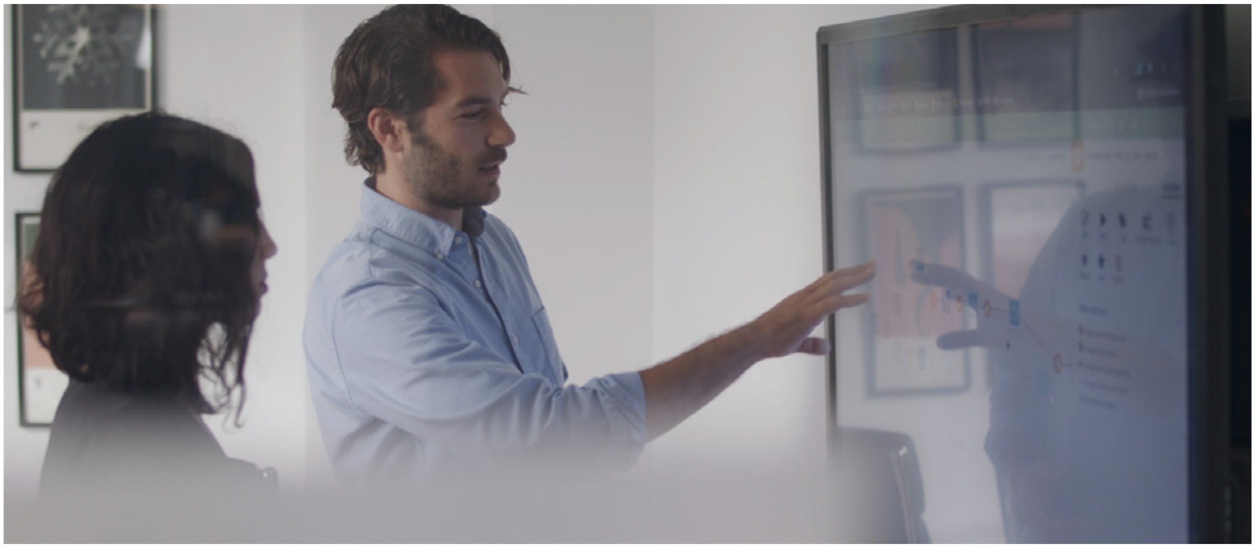
Indeed, as previously mentioned and worth reiterating: capitalization and reuse sound easy in principal, but are difficult in practice. Besides the previously discussed transparency strategies for building a strong foundation of reuse, what are the next steps? What does it take to really execute?

It comes down not only to having the right technology in place, but also to the right foundation of people and processes. Organizations that see real success from reuse and capitalization start with a strong data science center of excellence, which can help enormously at the start by establishing best practices and ensuring that people across the business - whether data scientists, analysts, or business users - are following them.

For example:

- GE Aviation found a lot of success with implementing processes through a combination of the right technology paired with thorough training and gamification (**read the full white paper** for a breakdown of what their programs look like).
- Pfizer found success through reuse, massively scaling their data science efforts via collaboration, transparency, and explainability (**watch the video** featuring their Data Science Senior Director to learn more).

For more tips on how to set up a data science center of excellence, **watch Nicholas Bignell**, Director of Data Science at UBS. Ultimately, organizations need to find the right combination of training, initiatives, and tooling that allows for maximum reuse and collaboration to bring the economics of AI into balance.



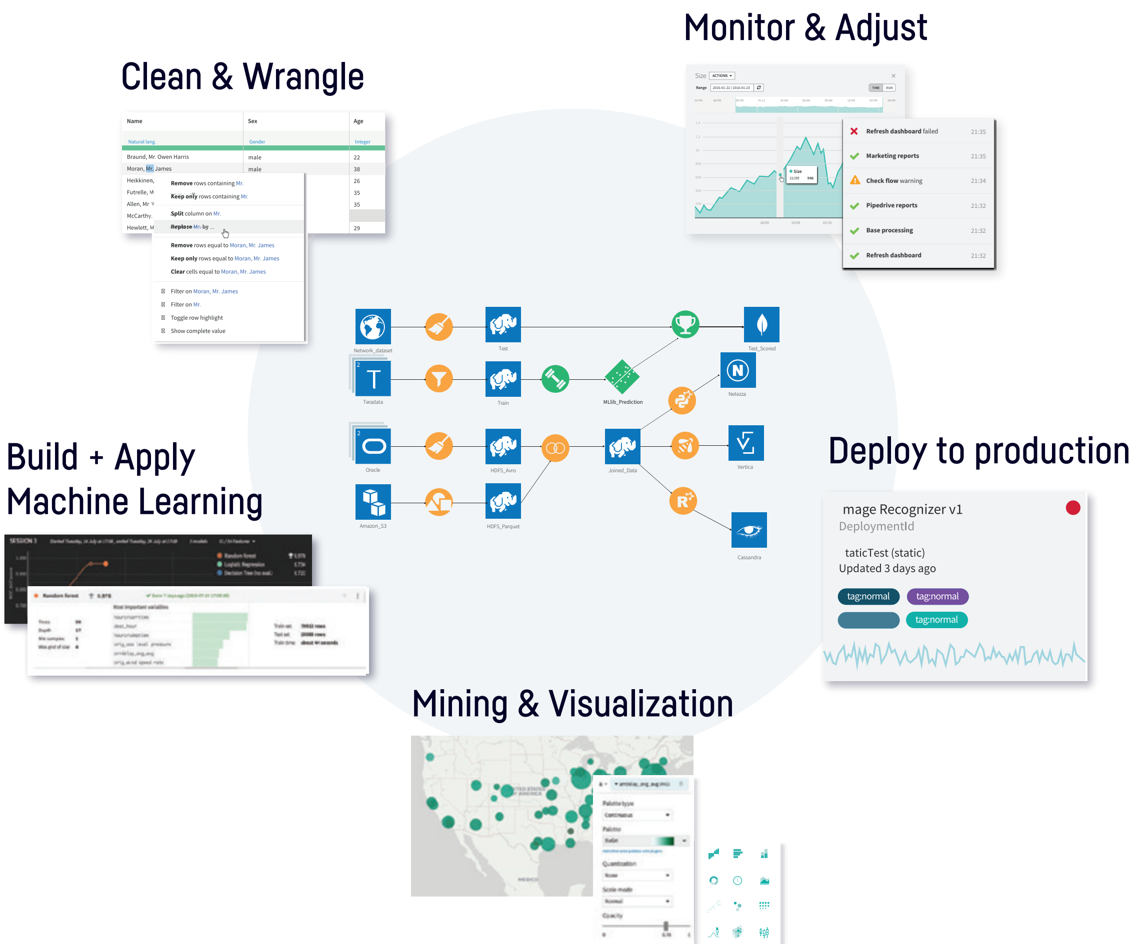
About the Author

Alexis Fournier

Director of AI strategy
at Dataiku

Alexis began his career as a data scientist in the telecommunications industry before joining an international organization, where he applied this expertise to economic research. Following this international experience, Alexis worked at SAP to support many customers in their innovation journey around data science. Today, alongside Dataiku's customer teams, Alexis supports organizations on the understanding of the value of AI in the enterprise and its processes, as well as on the different paths to Enterprise AI.

Your Path to Enterprise AI



400+
CUSTOMERS

40,000+
ACTIVE USERS*

*data scientists, analysts, engineers, & more

Dataiku is one of the world's leading AI and machine learning platforms, supporting agility in organizations' data efforts via collaborative, elastic, and responsible AI, all at enterprise scale. Hundreds of companies use Dataiku to underpin their essential business operations and ensure they stay relevant in a changing world.

EBOOK

www.dataiku.com