EBOOK

Building Self-Service Analytics in the Age of AI

Introduction

Right now for many organizations, self-service analytics is a wasted opportunity. At one time, the idea of self-service analytics was attractive and people believed in its potential. But, somewhere over time, companies went off track. The promise and value exchange between data creators in the center and data self-servers out in the business was never actually fully negotiated (or fulfilled, for that matter). For true success, teams need to align and work together from the beginning in order to turn missed opportunities into untapped opportunities with data.

> ***Self-service analytics refers to decentralized ownership of the insight production process (and isn't just "pulling numbers from dashboards"). For example, line of business professionals or analysts would be able to work with data to generate insights and data visualizations with little direct support from data scientists, IT, or the larger data team (though the data product itself and its greater platform should be supported by these profiles).***

However, because of varying organizational maturity, there is no true one-size-fits-all for self-service analytics. So, to avoid making the concept something that is (even more) daunting to or wholly avoided by business teams, we put together this ebook to outline some of those nuances (many of which stem from semantics) and highlight the future trajectory of self-service analytics for data executives.

More importantly, though, organizations can't afford to waste money spinning up or supporting a self-service analytics program that people aren't actually using or that's not generating tangible value. This is always true, but it's especially true in challenging economic times when programs and costs are more closely scrutinized.

Here, we provide a helpful analogy for understanding self-service analytics, supporting reasons to believe that self-service analytics isn't going anywhere, shortcomings to look out for, and how the concept of Everyday AI fills those gaps to make self-service more scalable and more valuable.

Redefining Self-Service Analytics

As we mentioned in the introduction, self-service analytics is not a new concept, and is still tremendously relevant to today's enterprises. In fact, according to Gartner®, "Organizations are embracing self-service analytics to democratize analytics capabilities among all end users. Gartner inquiries reveal a seven-fold increase in self-service topics since 2019."[1]The world of analytics at large is changing, especially when it comes to how people are generating insights.

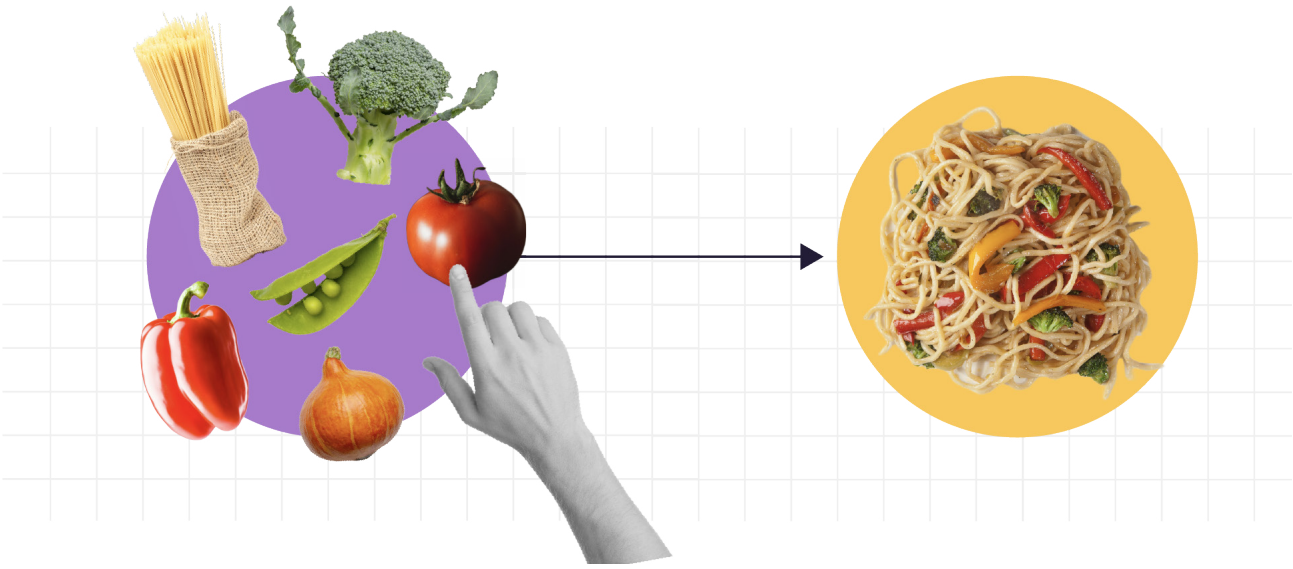
In the past, the challenge was that insight generation was limited to people in lines of business filtering and tweaking data products built by IT people, or perhaps combining two of those products together (e.g., "Show me sales split by region"). However, as we move toward what we call Everyday AI (more on this concept later), more sophisticated analyses and reporting will become table stakes — especially as more and more people are empowered and realize that they can and should build data products themselves instead of relying on a central team. They want to take more control and ownership of the end-to-end process of building data products.

While advanced analytics won't be critical for every person in every business unit, the tools now exist for a much broader group of people to take on these challenges themselves, whether they are an analyst who wants to generate basic reporting improvements or one who wants to go beyond that and fulfill a citizen data science role and co-build models with data experts.

To help illustrate this notion of data analysis as something that used to only be done by experts and can now be done by the masses (and has genuinely created more value despite the risks), let's take a cooking analogy.

1 *Gartner - How to Enable Self-Service Analytics to Ensure D&A Success, 16 June 2021, Anirudh Ganeshan, Carlie Idoine. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.*

Self-service analytics (when not done in the optimal way) is like a buffet experience — you can take whatever food has been prepared and fill yourself up, but you may not feel fully satisfied or like you had an "experience" (read: you can consume whatever data was given to you in the self-service app or dashboard by the data team, but you may still have questions or need more information). The data is over there, help yourself!



Cooking, on the other hand, is active for whoever is involved. When you're involved in the creation of the final deliverable (such as a dashboard, a web app, or a data marketplace), you'll be able to collaborate with the data team and ask your questions upfront instead of requesting the output to be reworked again and again because it doesn't actually satisfy your business objective or need.

Plus, you can customize it, see all the data sources (ingredients) that are used, and have full visibility into how the raw data gets transformed into the final data product. By comparison, most previous iterations of self-service analytics have stopped at the level of making your own changes to the final data product.

Let's face it: Nobody leaves the buffet entirely satisfied the way they do after they participate in the creation of the food. For IT and data teams, self-service analytics can be similarly disappointing and here's why:

- They think putting more and more food on the buffet (i.e., dashboards in the BI tool) is the answer, but that just confuses the business further.
- There are too many options that none of them actually taste good (i.e., none of the outputs actually answer the questions the business has, especially the new ones, because those can't be predicted well ahead of time).

What might be more successful for practitioners is if IT instead said, “Look, we know you’re behind the counter. Here’s a subset of ingredients — a mise en place if you will — and some pretty safe cooking tools, why don’t you give it a go yourself? And by the way, we’re expert cooks and we’ll be nearby when you need us.” Not only does this route give the business more autonomy to experiment on their own, but it does so while mitigating risks involved.

By equipping them with the right tools (that have instructions and documentation) as well as smoke alarms to prevent fire (or a burnt meal), they’re doing their due diligence to help them cook a successful meal without wholly eliminating guardrails and other protective measures.

They aren’t saying, “Here’s a full commercial kitchen, you have no training, good luck!” (i.e., we send you on a Python course) but rather are enabling the business user to cook something decent because they got help with the ingredients being measured and chopped already, tools that are collaborative and easy to use and, of course, oversight. The chef won’t leave them alone to their own devices, but rather help them upskill and make a dish they’re happy with. Similarly, IT and data teams will have fun inspiring new data people versus shipping unused dashboards into the ether (read: making dishes no one will consume).

To be clear — a buffet won’t always be bad, especially if you like the food that happens to be served, but very few people have an “aha” moment about a new cuisine, just as very few people generate truly new and meaningful insight by slicing and dicing others’ dashboards. Moreover, budding new chefs (business users experimenting with self-service analytics) will still go to a restaurant and admire an expert who creates the same dish they did (and did it twice as well and by the books). Not only does this demonstrate that self-service analytics isn’t to be used all the time or for everything, but rather hit home the point that sometimes, when we’re involved from the beginning, we prefer our own version as the final “meal.”

Why Self-Service Analytics Is Here to Stay

Shifting out of the kitchen, there are plenty of compelling reasons that self-service analytics — while its naming convention may continue to evolve over time — is not going away:

1. IT'S ESSENTIAL TO THE FUTURE OF WORKING.

> "Pervasive siloing of data and resistance to data sharing limit the value of data and analytics."
>
> "Data sharing is an essential business capability. Use of data and analytics must be aligned to digital business outcomes. However, most data capture is intended for a specific use case. Many organizations struggle to determine whether the additional investment and risk involved in broadening access to D&A assets and the facilitation of reuse of them will provide appropriate value to the enterprise."
>
> -Gartner, 4 Case Studies for Developing and Governing Enterprise Data Sharing, 2022[2]

For those working with data, running into roadblocks because of siloed data (like constantly having to ask for access or not even knowing what data exists to work with in the first place) is frustrating. And that frustration has a tangible cost:

- **Lost time:** The more data teams (or other staff) are held up tracking down data, the less time they have to work on business-impacting models. The bigger the team and the more siloed the data, the higher the costs will be to pay staff to hunt down the data they need. In fact, in a 2022 Dataiku survey, only 6% of respondents said that it's "super easy" in their day-to-day work to find and get access to data they trust when they need it.

2 *Gartner - 4 Case Studies for Developing and Governing Enterprise Data Sharing, 31 March 2022, Data and Analytics Practitioner Research Team*

- **Incomplete data projects:** WIthout a way to see what data is available, teams can work on entire data projects without the information they really need, which can, in turn, make models and insights less valuable than they could be.

- **Incorrect models:** Worse than a data project that isn't complete is one that is incorrect. When teams do get access to siloed data, often it's still challenging to understand what exactly that data means. Without a central system or ownership, teams are much more likely to work with data that they don't understand, ultimately misguiding the business and leading to decisions based on fundamentally flawed models or projects.

All of this is relevant in the context of self-service analytics, too. If organizations don't adopt a culture of data sharing (where one "route" of sharing is through self-service analytics), they are setting themselves up for failure as a future-proof organization. The organizations that are the most successful in the future when it comes to their data efforts will be the ones that don't share data just for ad hoc use cases but data sharing is a regular, everyday part of business processes.

They will be able to use data sharing to drive data discoverability across various profiles and teams, empower all employees to use data in innovative ways for faster and better decisions in their day-to-day work, foster a culture of reuse, and even unlock potential for new use cases from data that was previously locked up in silos.

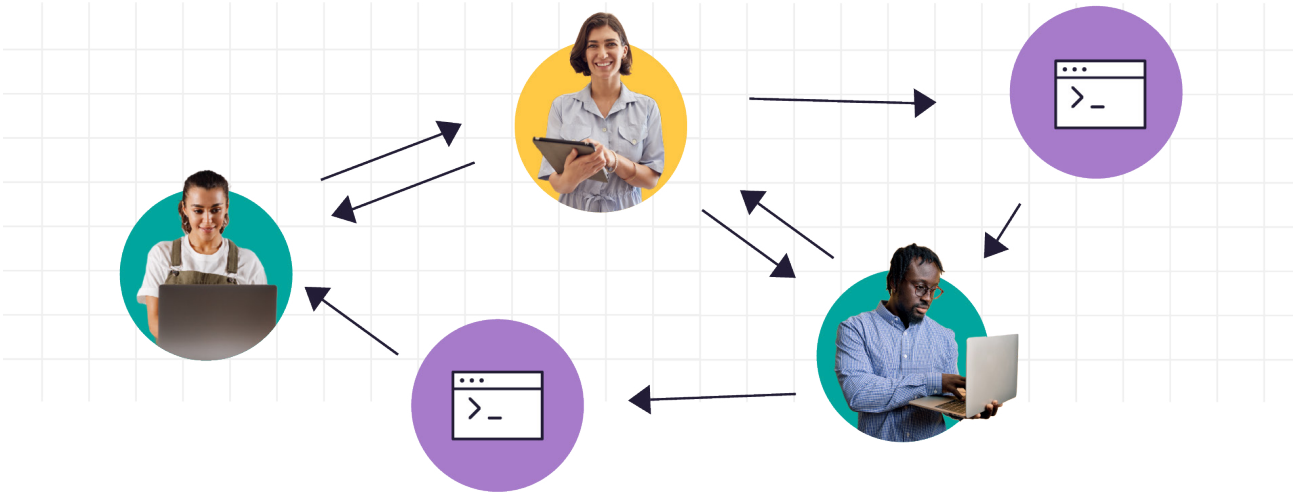
2. IT REAFFIRMS THE NARRATIVE THAT, TO TRULY SCALE AI EFFORTS, WE NEED MORE THAN JUST TECH EXPERTS.

> "Regular people, those without 'data' in their title, are central to all data-related work. Without buy-in and contributions from your company's rank and file, even the cleverest AI-derived model will sit idle and data-driven decision-making will just go around in circles.
>
> Conversely, costs go down and products get better when people help improve data quality, use small amounts of data to improve their team's processes, make better decisions, and contribute to larger data science and data monetization initiatives."
>
> -Harvard Business Review, 2022[3]

3 *https://hbr.org/2022/03/your-data-initiatives-cant-just-be-for-data-scientists*

In an ideal world, the self-service analytics data product is built by a combination of technical data people and the business people. It wouldn't be built in a one-sided, "Here's a mockup of what my team wants, deliver these requirements" way, but rather in a way where the product is co-built and is more robust to being taken apart, analyzed under the hood, and reused in other ways.

Moreover, according to Gartner, "By 2025, 70% of new applications developed by enterprises will use low-code or no-code technologies."[4] At Dataiku, we've always believed that data projects require involvement and alignment from both technical experts (who usually prefer to code) and business players (who might prefer to work with low or no code) and wholly support the notion that organizations won't scale AI without enlisting non-experts to the cause.



Increasingly, we're observing business users becoming empowered to co-build analytics workflows with experts because they have more access to data and the experts are more inclined to work with them to ensure the proper business subject matter expertise and context. This ability to collaborate on projects with coders and other technical practitioners — combined with upskilling and proper tooling (such as Dataiku) — can lead to faster time to impact.

With self-service analytics specifically, organizations can give everyone (with proper access rights) the ability to discover and use data, prepare that data, and create a data product. They can also enable data product creators to share their work with other colleagues across teams and departments as well as enable non-data teams to improve access to better data insights, understand key metrics, and streamline processes.

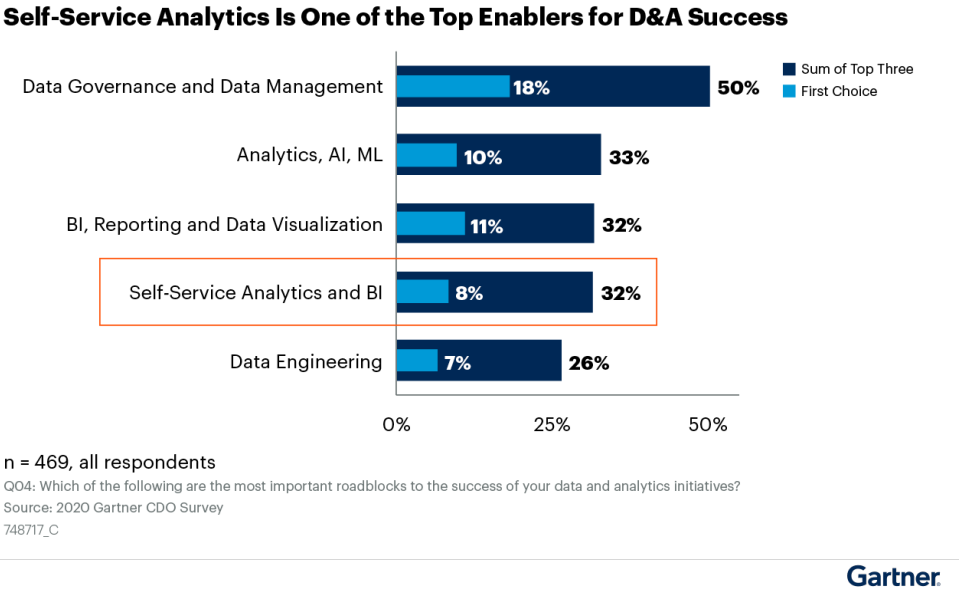
4 *Gartner - Scaling Analytics Requires Balance and Synergy Between Self-Service Analytics and Composable Analytics Applications, 26 July 2021, Julian Sun, Joao Tapadinhas, et al.*

3. THE NEXT LEVEL OF BUSINESS VALUE CAN ONLY BE GENERATED BY BRINGING IN NON-DATA EXPERTS, TOO.

> "Mastering self-service analytics at scale continues to evade organizations, leading to disconnected business value. Data and analytics leaders must showcase self-service value, foster collaborative development between IT and business, and adopt lightweight management and control."
>
> -Gartner, How to Enable Self-Service Analytics to Ensure D&A Success, 2021[5]

As highlighted in the figure below from Gartner, self-service analytics one of the key enablers for data and analytics success. However, we feel there's often a disconnect between *doing* self-service analytics and *driving business* value from self-service analytics.

It's fair to say that many people have a vision of what self-service analytics is at their organization and may even look at it as a failed experiment in their company or as an initiative that has limited value. The business may also feel like they were left high and dry (i.e., "I have to do the work and no one is here to verify I'm doing it correctly"), turning them off of the notion of self-service analytics as a whole.



5 *Gartner - How to Enable Self-Service Analytics to Ensure D&A Success, 16 June 2021, Anirudh Ganeshan, Carlie Idoine. © [2022] Gartner, Inc. and/or its affiliates. All rights reserved.*

In order to avoid those sentiments and drive tangible business value, organizations need to be sure that their self-service analytics efforts:

- Are driven by intimate business knowledge (i.e., they consider the specific problems or roadblocks the business is facing and explores how data can help solve them) and always tie back to business objectives and KPIs
- Don't exist in a vacuum but rather encourage collaborative discourse and engagement between IT and business people
- Are governed and have guardrails (more on this in the next section)
- Are sustainable and reproducible for other projects and requests down the road (and, notably, free up key resources to support the identification of key value creation opportunities for operationalization)

4. IT CAN HELP MITIGATE SILOS AND PROMOTE BEST PRACTICES AROUND TRUST AND CREDIBILITY.

> "Without careful planning, self-service analytics often leads to analytic silos, resulting in duplication of analytics content, security, and compliance issues."
>
> -Gartner, Quick Answer: 4 Easy Ways to Promote Trust in Self-Service Analytics, 2021[6]

There are a lot of different trust and credibility considerations when it comes to self-service analytics:

- The enterprise needs to trust employees' ability to use data in a self-service context.
- Business users working on self-service analytics need to trust the data that they're working with (and there needs to be someone continually responsible for its quality, making sure it's regularly updated, formatted, and being used appropriately).
- Managers and executives alike need to trust the insights delivered from self-service analytics projects.

6 *Gartner, Quick Answer: 4 Easy Ways to Promote Trust in Self-Service Analytics, 1 November 2021, Anirudh Ganeshan, Austin Kronz. © [2022] Gartner, Inc. and/or its affiliates. All rights reserved.*

In order to ensure strong governance practices, organizations need to strike the balance between control and agility. They need to build a solid (yet flexible) strategy that allows lines of business access to the data they need while also restricting any access they have no business need to access.

They also need to maintain a workable feedback strategy that enables users to gain access they don't have (but need) in order to avoid those data access problems from ultimately killing the data project. This balance is best struck in a centralized environment (like Dataiku) where roles and rights can be managed and updated easily as obligations and priorities evolve over time.

How can teams reach that middle ground in practice? They can be sure to do the due diligence to manage risk and ensure compliance at scale across the organization by:

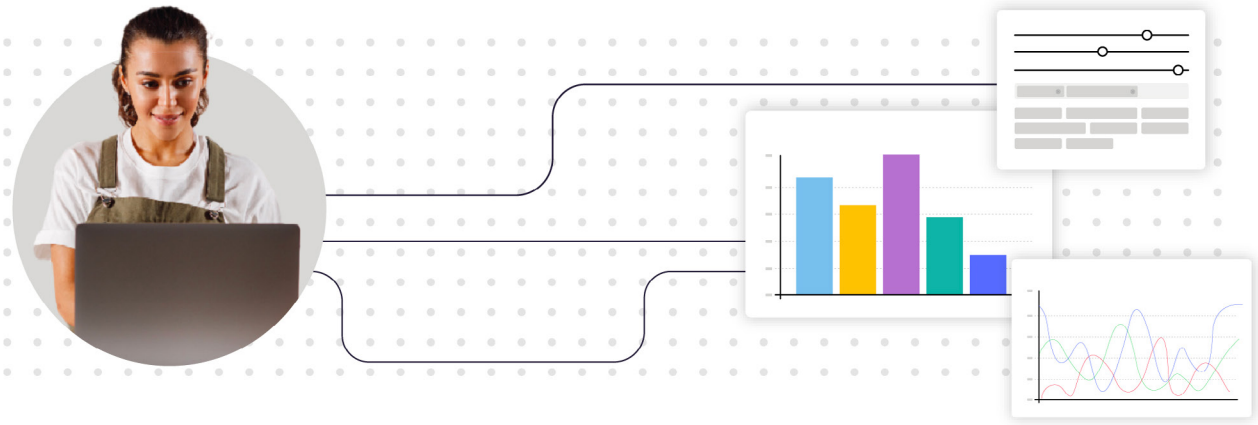
Enacting permissions management for who can access, read, or make changes to a project, access different datasets, or reuse various project steps for one of their own projects

Creating logs for user access and activity (helpful for both troubleshooting and compliance with internal controls and external regulations)

Putting enterprise-level security measures in place (i.e., documentation, change management, rollback, and monitoring)

Managing how data and models are being used

Away From the Past & Into the Realm of Self-Service for AI



Let's say someone on your central data team is struggling to service a lot of requests and prioritize the highest value work. The idea that they can get other people involved is very attractive. So, for IT-driven buying and thinking, that's why self-service analytics keeps rearing its head — it's a huge win. The business users, though, need to find the time to do the work themselves without the expertise of the data experts, so it's a hard sell to those stakeholders, especially because "self serve" as a concept doesn't necessarily scream "positive value story" and might make them feel like they're on an island.

Traditional BI tools used for visualization, reporting, and presenting insights work when some data has already been prepared in a format and structure that allows someone like a non-technical, line-of-business user to answer questions about specific KPIs, prices, SLAs, etc. At the most basic level, they can look at someone else's data product (such as a dashboard), ask their question (such as, "How did sales perform last year?") and get an answer. The data product is well-built for questions, and the data is trustable. This is the smoothest type of self-service analytics (in the realm of BI, anyway). The business user had a question, asked it, and got an answer — they didn't need anyone in the middle to help.

SO, HOW CAN YOUR TEAMS MOVE FROM TRADITIONAL SELF-SERVICE ANALYTICS TO SELF-SERVICE IN THE REALM OF AI?

It is more than just moving from descriptive to predictive (and even prescriptive) analytics. Both approaches have the same basic principle: You're asking questions and then finding answers in data. Every data product ingests data, transforms it, and spits it back out.

So, it's less about understanding the past or predicting the future, and more about empowering people — because we've reached a point where people (such as an analyst or someone on the business side who is doing much more with data than ever before) want to drive *additional* insights. They've created a dataset, merged it, cleaned it, made data visualizations, and they still want to go further to answer a business objective (and probably ask new questions as well). Dataiku makes all of that possible and connects doers with data by:

- Bringing people of diverse skill sets together to work with data in a common ground
- Providing a remotely accessible platform that provides a centralized place to find data and previously built datasets and projects by other colleagues
- Enabling the users of data to make more of the consequential decisions about their data by encouraging them to help build the data project from the get-go
- Giving business users autonomy when it comes to creating said data projects (via best-in-class data exploration, data access, and data transformation abilities — all without having to learn to code or put additional strain on IT)
- Granting users automation capabilities (in the cases where an insight needs to be created repeatedly) to update the insight every hour/day/week without the need to stop their current day-to-day tasks
- Making it second nature to productize new insights and data preparation methods (in a way that is easily reusable) so that all business users will benefit from the expertise of the many as timely new data products are created and maintained across the whole business

Dataiku customer GE Aviation has implemented their own version of a self-service system that serves their specific needs and requirements and that allow them to use real-time data at scale to make better and faster decisions throughout the organization:

Engineering uses data from these tools to redesign parts and build jet engines more efficiently.

The supply chain team uses it to get better data insights into their shop floors and streamline supply chain processes.

Finance uses it to understand key metrics such as cost, cash, etc.

The commercial group (by leveraging data scientists) uses these tools to transform engine sensor data from customers and build analytics services for them.

Their data initiative is called Self-Service Data (SSD), but it encompasses both self-service analytics as well as an element of operationalization) for both business lines and IT users. At its core, the SSD equips everyone (with proper access rights) with the ability to discover and use data, prepare that data, and create a data product, including developing predictive models within Dataiku.

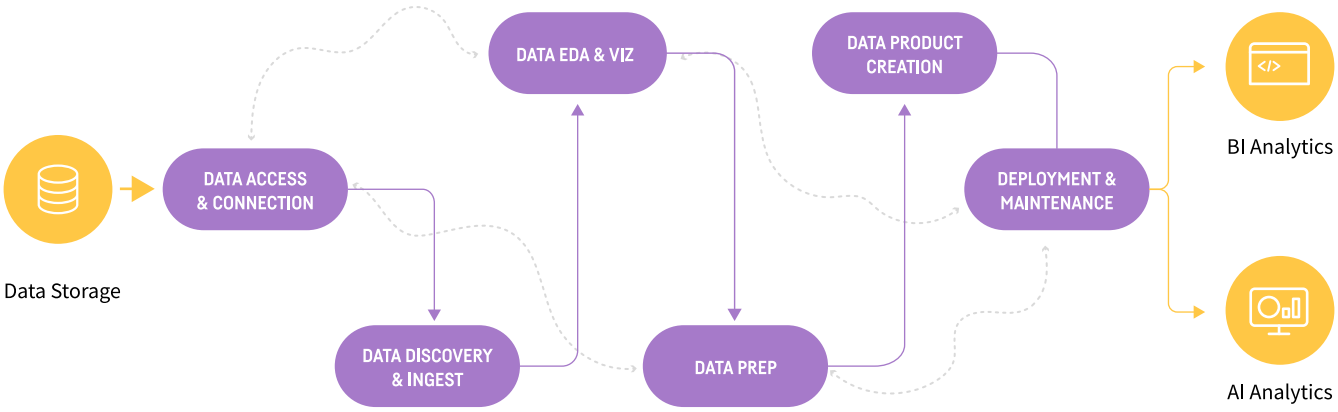
It also grants data product creators the ability to share their work with other colleagues and deploy data pipelines in production using macros developed in Dataiku. Lastly, to avoid pure chaos in the kitchen (to go back to our cooking metaphor), the SSD team has a process to ensure that anything users want to put in production passes a set of checks and balances and governance measures.

Don’t Forget About Governance

It’s important that these guardrails for data quality and governance aren’t overlooked. Taking another example, the FP&A team at Standard Chartered Bank has developed their own brand of self-service analytics with Dataiku. The idea isn’t that individuals can do and build whatever they want with data (which would lead to data chaos), but rather that a center of excellence owns the core structured intelligence of the bank. That is, enterprise-level data connected to a homogenous pool with product owners for every dataset and defined governance which connects to the entire organization.

From there, the team builds specific experiences on top of that data that can deliver answers through core apps, and the ultimate self-serve flexibility comes from how people around the organization use those apps to solve business problems day to day. On top of it all, Dataiku enables all of this work through visually legible data pipelines and not on desktops, which helps IT teams sleep better at night.

Further, when it comes to time savings and efficiency, analysts and line of business managers can capitalize on their existing BI data pipelines and turn them into AI data pipelines by creating AI data products. To transition from building BI data products to AI data products, staff can reuse the data assets and infrastructure they’ve already built and understand well, shorten the learning curve by leveraging an identical user experience in the data pipeline, and scale faster and more economically by doing both BI and AI projects in one environment.



Data and Analytics Pipeline

Now, while BI and AI analytics can certainly build off each other, it's only part of the story. It goes beyond using data to just prove what you already know. For Dataiku customer Rabobank, one of the biggest advantages of Dataiku is that data projects progress. Often, the business starts out with a simple insights question, but those insights then lead to new initiatives. For example, if you know certain customers have certain risks, that starts out as a dashboard but quickly gets into predictive analytics and machine learning.

> ***"If you start with a BI tool, then you have to do all kinds of work to set up a new environment once the project progresses. Dataiku allows us to start out with relatively simple insights questions and grow toward a more specific predictive question, developing a model all in the same tool."***
>
> ***-Roel Dirks, Product Manager Big Data Lab, Rabobank***

Why We Call All of This Everyday AI

There are people in the business who have the ambition to go on their own data journey and will do it if the points of friction are reduced and they are enabled to do so. And for those on the business side who aren't ready to jump into AI, they can start somewhere and work their way up to it — but they might as well do it in one place that's collaborative and future proof, like Dataiku.

To give our cooking example one last bit of airtime, these teams don't need to be in the kitchen and ready to go right now. They just need to believe that one day they might want to cook something really impressive, so they have the tools and processes in place for when that day comes. When people of different skill sets and backgrounds get in our kitchen and start cooking (i.e., using Dataiku), they do get inspired. For example:

Unilever designed a responsible, self-service tool for natural language processing.

Unilever's People Data Centre (PDC) teams across the globe deal with vast amounts of unstructured text data on a daily basis to gain insight into their customers. The answers that the company's marketers, product research and development, and supply chain specialists need also require analytics approaches tailored to the business. Data scientists and software engineers in PDC built a range of NLP methodologies, including a plugin. With Dataiku, they can easily assess how and where people across the company use it across different research projects.

Westpac has created a collaborative, self-service operating model to upskill and drive a closer alignment between the business and tech teams.

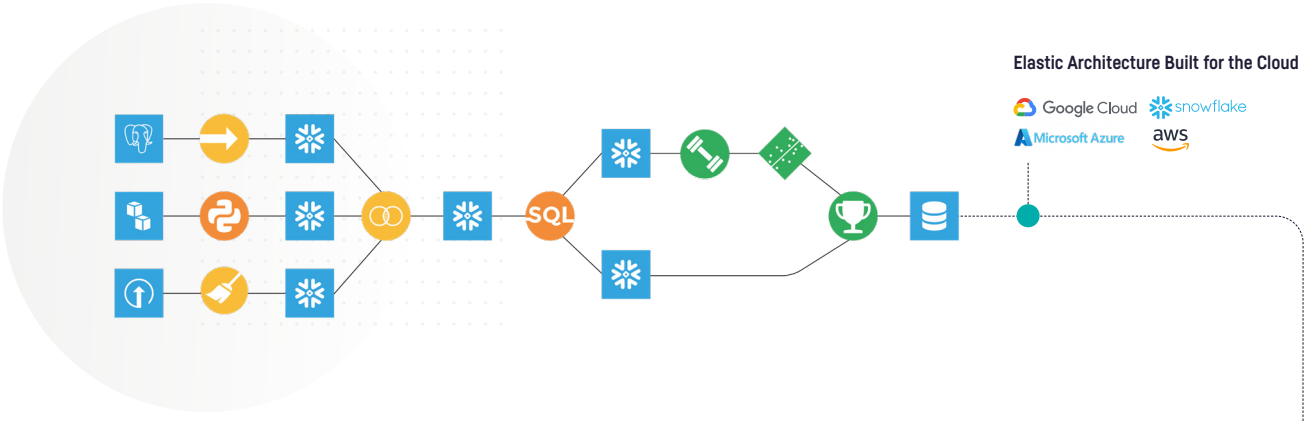
Westpac developed a new operating model and new processes to ensure a strong alignment between the Discovery Lab (consisting of technical players) and the business teams, while enabling them to broaden their understanding and gain new data skills throughout the project. When a new team member joins, they attend an orientation session so they understand how they can leverage Dataiku and go through an assessment to further define their use case, the objectives and expected impact, and how they visualize the outputs and outcomes of the initiative. The tech team is enabled to serve the business, help them upskill, and work together to drive innovation at scale.

- **AstraZeneca has created a self-service capability to put the power of AI and analytics into the hands of employees.**

  AstraZeneca provides core data science capabilities to 120+ users and 90+ data science projects within one year. They span across R&D, operations, and commercial teams and the users range from technical data scientists and ML engineers to business analysts who, with the aid of Dataiku, have been able to produce their first ML model. The team has also created macros for automating all of the steps in the project creation process, including group and connection creation, and helped SMEs and non-technical users produce valuable insights from their data.

At the end of the day, Everyday AI brings down the temperature of self-service analytics and makes it more consumable. It involves enabling cross-functional teams and people with diverse skill sets and levels to work together, understand what each other is working on, understand what has been published by whom, and how they can rebuild data products on top of what already exists.

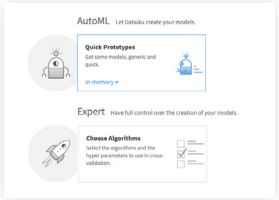
Everyday AI,
Extraordinary People



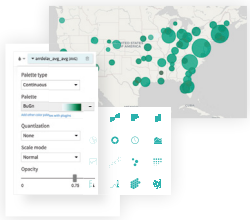
Elastic Architecture Built for the Cloud

Google Cloud snowflake
Microsoft Azure aws

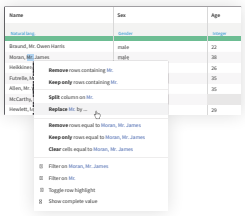
Machine Learning



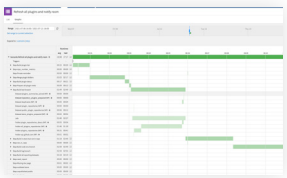
Visualization



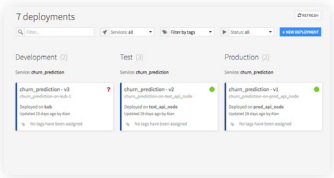
Data Preparation



DataOps



Governance & MLOps



Applications

450+
CUSTOMERS

45,000+
ACTIVE USERS

Dataiku is the world's leading platform for Everyday AI, systemizing the use of data for exceptional business results. Organizations that use Dataiku elevate their people (whether technical and working in code or on the business side and low- or no-code) to extraordinary, arming them with the ability to make better day-to-day decisions with data.