

HUMAN BARRIERS TO SCALING A DATA TEAM

Challenges

CONTRIBUTION



Data analysts feel that they can't contribute to data projects in meaningful ways

DATA PREP



Data preparation takes too long, feels draining for the team, and takes up valuable data science resources

COLLABORATION



Data analysts, data scientists, IT, and business teams aren't empowered to work together collaboratively

Needs



The capability to work with data either in a visual interface or using SQL, Python, etc.



Faster data prep and the ability for analysts to contribute to data preparation



The ability for multiple profiles to work together on different project components

Benefits



Cost savings in getting the right profiles to do the right tasks to match their skills



Better retention of talent when individuals are able to work on challenging tasks and projects



Faster, more scalable process to go from raw data to prediction

Dataiku Provides



A customizable drag-and-drop visual interface at any step of the predictive dataflow prototyping process – from wrangling to analysis to modeling.



The ability for all skill levels (from beginner data analyst to expert data scientist) to make meaningful contributions to data projects.



Easy access to over 80 built-in visual processors for code-free data wrangling.



A fast way to perform mass actions on data via automatically suggested contextual transformations. The ability to use formulas and Python macros for custom processing.



The ability to use either code or visual interface for effective contributions from all skill levels at all stages of a data project.



Integrated documentation and knowledge sharing, including project to-do lists, commenting, and sharing.



Robust project change management and rollback.



Advanced team activity monitoring for effective data project management.

TECHNOLOGICAL BARRIERS TO SCALE A DATA TEAM

Challenges

GOVERNANCE



Lack of centralized data access and user controls

ACCESS



Difficulty connecting to and combining siloed data sources

DISCONNECT



Use of different tools/platforms for data preparation, model development, and deployment

Needs



One central repository for all data science work with robust access controls and monitoring



Ability to connect to any data source regardless of where or how it's stored



One system that handles everything from raw data to deployment

Benefits



Team can use the latest and best data science technologies



No loss of productivity on in-progress projects when team members leave



Overall centralization and organization allows for multiple data projects to happen simultaneously (better scalability)



Easier and quicker onboarding since each team member can use languages and techniques with which (s)he's familiar

Dataiku Provides



The ability to organize data tasks into clearly identified projects, document actions/datasets, and search across and within projects.



Fine-grain access rights and detailed monitoring through logs and dedicated dashboards.



Connections to more than 25 data storage systems plus the ability to connect to nearly any data thanks to a custom API connector.



Automatic format and schema detection for instant data access without the need to write fastidious formatting settings before reading a dataset.



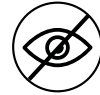
A single interface for transforming raw data into real-time predictions from beginning to end (including fully staged deployment models from dev to test to preproduction to production, all within a single UI).

PRACTICAL BUSINESS & OPERATIONS BARRIERS TO SCALE A DATA TEAM
Challenges
PRODUCTION


Lack of process to quickly put models in production

ITERATION


Inability to easily make changes to models already in production

MONITORING


No way to monitor models that are in production

Needs


Automatically push into production using an API



Easy model deployment as a reproducible package



Monitoring models' performance and get alerts

Benefits


Business sees more ROI faster on data projects, meaning more resources and better funding for the team, allowing it to grow and take on more exciting projects

Dataiku Provides


The ability to package an entire workflow, optionally including data and models, as a single deployable and reproducible package.



A REST API for easy and fast model deployment in production for real-time predictions.



Safe versioning and rollback to easily review changes, activate new models, and revert if necessary.

A/B

Automated model A/B testing.



A robust feedback loop with access to history of logs queries and predictions to check that model performance is not drifting with time.



Access history of logs queries and predictions at any time to check that model performance is not drifting with time.



Run multiple versions of the same model at the same time for automated A/B testing.

ABOUT DATAIKU

Dataiku is the advanced analytics leader and preferred software solution in helping organizations succeed in the world's rapidly evolving data-driven business ecosystem. Guided by the belief that true innovation comes from the effective combination of diversity of cultures, of mindsets, and of technologies, Dataiku's purpose is to enable all enterprises to imagine and deliver the data innovations of tomorrow.

Try Dataiku Data Science Studio (DSS) for free by visiting www.dataiku.com/try

