

# Analytic Executive Challenged to Create and Deploy Models



The analytic executive of today is faced with numerous challenges when aligning with corporate goals and finding ways to monetize the data assets of their enterprise. It's becoming a well-understood strategy that data and analytics can lead to significant marketplace advantages. Bain and Company, in this [brief](#), reported:

However, it's not as simple as deciding to implement analytics, and then watching the advantages roll in. This executive brief aims to summarize a few key trends in data science tool usage and the challenges faced when taking data science into production, as well as what can be done about it.

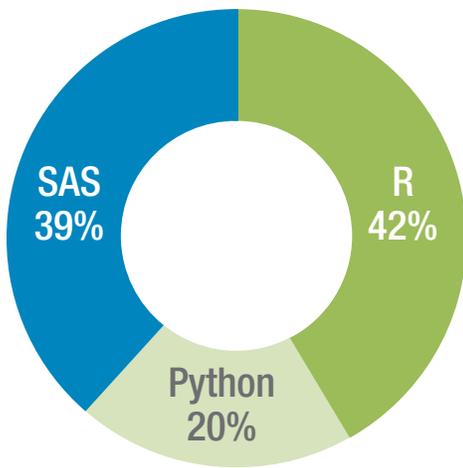


Figure 1 - Graph from Burtch Works Report on Tool Usage

*“[companies] with the most advanced analytics capabilities are outperforming their competitors by wide margins. The leaders are:*

- *Twice as likely to be in the top quartile of financial performance within their industries*
- *Five times as likely to make decisions much faster than market peers*
- *Three times as likely to execute decisions as intended”*

~ Bain and Company

## The rise of the heterogeneous data science tool box

Over the past several years, there has been a rise of a heterogeneous data science environment as seen in many surveys and reports. A [recent report](#) from Rexer Analytics, which surveyed over 1,200 analytic professionals, indicates the average data scientist reports using 5 tools in their daily job. The report states, “While R has established a dominant position in the world of analytic software, 64% of data scientists select other tools as their primary analytic software, and the average data scientist reports using five tools”. The latest Burtch Works [report](#) indicates SAS, R and Python are taking more equal shares in usage in recent years.

As a final data point, a recent [Dataiku report](#) indicates, “80% of people have a polyglot development environment”. No one can blame the data science team. They are reaching for more and more tools to solve a variety of problems. This creates a new problem for the analytic executive—and the organization as a whole—as they try to then leverage those data science assets into production. For each new tool or language being used by the analytics team, the implementation team for production—typically from the IT organization—must find a way to handle that new system. And an equally important trend is that the traditional large “end to end” players like SAS and SPSS appear to be

“Less than two thirds of respondents indicate that models are deployed “most of the time” or “always.” These numbers have not improved over time.”

~ Rexer Analytics

## Frequency of Model Deployment

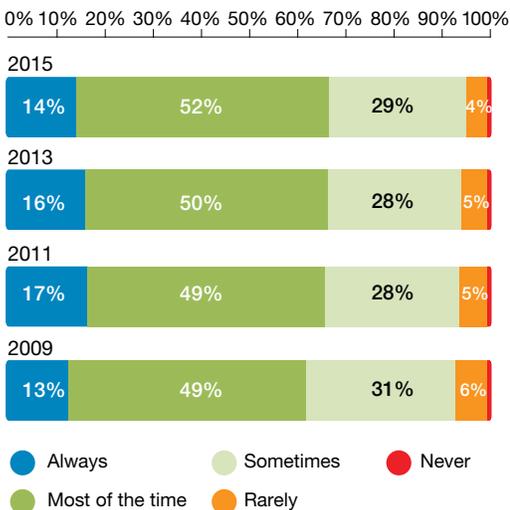


Figure 2 - Frequency of Model Deployment  
Graph from Rexer Analytics 2015  
Data Miner Survey Summary Report

losing dominance as open source tools are taking larger and larger portions of data science analyses. This disaggregation trend means there are fewer out of the box ways to take the data science design process and quickly move it into production.

At the same time, no single tool appears poised to win the battle. Neither R, nor Python are set to dominate the world, muddled by many articles that discuss why / how to use one tool or the other depending on the use case. This has driven even further tool choice growth with “combo” offerings like Rpy2. It seems safe to assume, for the time being, that the data science team will continue to—as it absolutely should—use as many tools as required to achieve the business objective. Thus the analytic executive must prepare for a multi-tool world, and what that implies for downstream workflows and systems.

## Analytic deployment is not easy

Adding to the increasing fragmentation in the tool set of a data scientist, the analytic executive must also consider how to deploy the data science assets once they are created. It's not a simple task. The *Rexer Analytics* reports that less than 60% of respondents in their survey indicate that their data science deployments (once the model is complete) take weeks or longer to deploy, with a shocking 30% indicating it can take months or longer. What's worse is “less than two thirds of respondents indicate that models are deployed “most of the time” or “always.” These numbers have not improved over time.”

A recent *Dataiku* report confirms there has been, historically, a potential lack of focus on data science production. The report finds that “50% of all respondents do not have a specific data science production procedure”. In addition, the report astutely notes that data science production is an IT-led problem. This is one of the great challenges for an organization that wishes to leverage analytics and data to its fullest potential.

## Analytic Creation is not the same as Deployment

At the heart of the issue is combining two disparate systems. Today's data science tooling was not built for large-scale deployment in enterprise IT systems. In fact, many of the standard IT stacks lack the capability to safely handle complex analytics at scale. At the same time, both the data science tool kit, and the underlying data systems continue to evolve and churn – resulting in continuous change – which is at odds with building stable production systems in IT.

Enter the idea of an analytic deployment environment. In order to safely and efficiently leverage data science assets, the analytic executive must internalize that creating data science assets is not the same as deploying them. These are two, complementary proficiencies that must be developed in the organization. Further, the analytic executive must look for solutions that comprehend the IT department will likely own large portions of the deployment responsibility. Data science and IT teams must work together to solve this challenge.

“50% of all respondents do not have a specific data science production procedure”

~ Dataiku

**Three key attitudes must be developed, and tooling to support these ideas must be put in place:**

- 1. Enable Alignment** - for the organization and its processes. Both data science and IT teams must be aligned to common standards, interfaces and hand-offs. If tools ignore one side or the other, or don't create agnostic interfaces that allow natural churn, those projects may fail.
- 2. Co-ordination** - technology and systems for both data science and IT teams must be coordinated. No solution that ignores the needs of a particular team can survive. Likewise, no initiatives that strive to “do it without IT” or “force data science to use tool X” will succeed long term.
- 3. Collaboration** – Each team must understand they alone cannot achieve the enterprises' business objectives. The analytic alone does not make the decision. A perfect data store that goes untapped provides no value. A complete system view, including how to manage assets in deployment, is required to solve the business objectives.

These three attributes combine to demand an environment, both organizationally and operationally, that can be leveraged to successfully deploy analytics at scale. Broken down one step further, a useful deployment environment must:

- 1.** Enable the data scientist to continue to leverage their heterogeneous tool kit to achieve the best results possible.
- 2.** Give appropriate tooling to the IT / DevOps team that supports their journey for continuous integration, operation and agile development methodologies.
- 3.** Support a modern, microservices based architecture to future proof the enterprise stack
- 4.** Enable continuous tuning and iteration – both for the data science team (Can I quickly change a model type or add a factor?) and the IT team (how much memory and CPU will this take and how to efficiently change a model in production?).
- 5.** Provide input to any business process that can leverage the analytic outcomes

A system that can achieve these 5 goals will take a place in the enterprise stack as a truly fundamental building block – an analytic deployment environment.

## Summary

There are many trends in data science today – from tooling to potential use cases. The analytic executive faces a myriad of challenges and opportunities to help their organization fully monetize their data strategy. The landscape is continuously changing, but data shows that the data science team will have heterogeneous tools sets, and that data science deployment is challenging. One critical piece of this puzzle must be the acknowledgement that analytic creation and deployment are two complementary but unique capabilities, and the executive should make a plan for implementation of an analytic deployment environment. The analytic executive must be part of the bridge that has data science and IT working together to solve business problems.



111 W St John Street, Suite 530  
San Jose, CA 95113  
info@opendatagroup.com  
1-708-488-8660  
<http://www.opendatagroup.com>